В.Л.Бонч-Бруевич, С.Г.Калашников

ФИЗИКА ПОЛУПРОВОДНИКОВ

ОГЛАВЛЕНИЕ

Предисловие 10

Глава I

НЕКОТОРЫЕ СВОЙСТВА ПОЛУПРОВОДНИКОВ

§ 1. Кинетические явления в полупроводниках 13

а. Электропроводность (13).

б. Эффект Холла (15).

в. Изменение сопротивления в магнитном поле (17).

г. Термоэдс (18).

д. Эффект Томсона (19).

е. Эффект Пельтье (20).

ж. Эффект Нернста—Эттингсгаузена (21).

з. Эффект Риги—Ледюка (22).

и. Продольные термомагнитные эффекты (23).

§ 2. Время релаксации 23

§ 3. Элементарная теория гальваномагнитных явлений 25

а. Тензор электропроводности в магнитном поле (25).

б. Угол Холла и постоянная Холла (28).

в. Магнетосопротивление (30).

§ 4. Смешанная проводимость 32

а. Эффект Холла (32).

б. Магнетосопротивление (35).

§ 5. Некоторые экспериментальные результаты 35

а. Электронная и дырочная проводимость (35).

б. Собственная и примесная проводимость (36).

в. Запрещенная энергетическая зона (40).

г. Удельная электропроводность (42).

д. Подвижности (43).

е. Собственная концентрация электронов (45).

ж. Магнетосопротивление (47).

Глава II

ХИМИЧЕСКИЕ СВЯЗИ В ПОЛУПРОВОДНИКАХ

§ 1. Кристаллические решетки 50

§ 2. Электронная конфигурация атомов 53

§ 3. Типы химической связи 56

а. Ионная связь (57).

б. Гомеополярная связь (57).

в. Ван-дер-ваальсовская связь (60).

§ 4. Строение некоторых полупроводниковых кристаллов 60

а. Ионные кристаллы (60).

б. Гомеополярные кристаллы (61).

в. Кристаллы со смешанными связями (62).

§ 5. Некристаллические полупроводники 64

а. Аморфные полупроводники (64).

б. Жидкие полупроводники (65).

в. Стеклообразные полупроводники (68).

§ 6. Запрещенная зона энергий 69

§ 7. Полупроводниковые свойства и химическая связь 70

§ 8. Полупроводники с малой подвижностью 71

§ 9. Примесные атомы 73

§ 10. Вакансии и междоузельные атомы 78

§ 11. Дислокации 81

Глава III

ЭЛЕМЕНТЫ ЗОННОЙ ТЕОРИИ

ТВЕРДОГО ТЕЛА
I. ИДЕАЛЬНАЯ РЕШЕТКА
§ 1. Основные предположения 87
§ 2. Волновая функция электрона в периодическом поле 89
§ 3. Зоны Бриллюэна 94
§ 4. Энергетические зоны 100
§ 5. Метод сильно связанных электронов 103
§ 6. Закон дисперсии. Изоэнергетические поверхности 112
§ 7. Металлы и полупроводники 114
§ 8. Эффективная масса 116
§ 9. Зонная структура некоторых полупроводников 122
Глава IV
ЭЛЕМЕНТЫ ЗОННОЙ ТЕОРИИ ТВЕРДОГО ТЕЛА
II. КРИСТАЛЛЫ ВО ВНЕШНИХ ПОЛЯХ. НЕИДЕАЛЬНЫЕ КРИСТАЛЛЫ
§ 1. Средние значения скорости и ускорения электрона в кристаллической решетке 129
§ 2. Электроны и дырки 133
§ 3. Движение носителей заряда в постоянном и однородном магнитном поле (классическая теория). Диамагнитный резонанс 137
§ 4. Метод эффективной массы 145
§ 5. Энергетический спектр носителя заряда в постоянном и однородном магнитном поле (квантовая теория) 149
§ 6. Движение и энергетический спектр носителей заряда в постоянном электрическом поле 154
§ 7. Мелкие примесные уровни в гомеополярном кристалле 160
Глава V
СТАТИСТИКА ЭЛЕКТРОНОВ И ДЫРОК В ПОЛУПРОВОДНИКАХ
§ 1. Введение 167
§ 2. Распределение квантовых состояний в зонах 168
§ 3. Распределение Ферми — Дирака 169
§ 4. Концентрации электронов и дырок в зонах 170
§ 5. Невырожденные полупроводники 172
§ 6. Случай сильного вырождения 174
§ 7. Эффективная масса плотности состояний 175
§ 8. Плотность состояний в квантующем магнитном поле 179
§ 9. Концентрации электронов и дырок на локальных уровнях. Простые центры 181
§ 10. Многозарядные центры 185
§ 11. Распределение. Гиббса 186
§ 12. Частные случаи 189
§ 13. Определение положения уровня Ферми 191
§ 14. Уровень Ферми в собственном полупроводнике 193
§ 15. Полупроводник с примесью одного типа 194
§ 16. Взаимная компенсация доноров и акцепторов 195
§ 17. Компенсированные полупроводники 197
§ 18. Определение энергетических уровней примесных атомов 199
а. Многозарядные акцепторы в полупроводнике n-типа (199).
б. Многозарядные акцепторы в полупроводнике p-типа (201).
в. Многозарядные доноры в полупроводнике n-типа (202).
г. Многозарядные доноры в полупроводнике p-типа (202).
Глава VI

ЯВЛЕНИЯ В КОНТАКТАХ (МОНОПОЛЯРНАЯ ПРОВОДИМОСТЬ)
§ 1. Потенциальные барьеры 205
§ 2. Плотность тока. Соотношение Эйнштейна 207
§ 3. Условия равновесия контактирующих тел 209
§ 4. Термоэлектронная работа выхода 210
§ 5. Контактная разность потенциалов 213
§ 6. Распределение концентрации электронов и потенциала в слое объемного заряда 216
§ 7. Длина экранирования 218
§ 8. Обогащенный контактный слой в отсутствие тока 220
§ 9. Истощенный контактный слой 222
§ 10. Токи, ограниченные пространственным зарядом 226
§ 11. Выпрямление в контакте металл — полупроводник 232
§ 12. Диффузионная теория 236
§ 13. Сравнение с экспериментом 239
Глава VII
НЕРАВНОВЕСНЫЕ ЭЛЕКТРОНЫ И ДЫРКИ
§ 1. Неравновесные носители заряда 243
§ 2. Время жизни неравновесных носителей заряда 244
§ 3. Уравнения непрерывности 247
§ 4. Фотопроводимость 250
§ 5. Квазиуровни Ферми 255
§ 6. Электронно-дырочные переходы 258
§ 7. Обнаружение неравновесных носителей заряда 261
§ 8. Амбиполяркая диффузия и амбиполярный дрейф 264
§ 9. Длины диффузии и дрейфа 268
§ 10. n^+ — n^- и p^+ — p^- переходы 271
Глава VIII
ВЫПРЯМЛЕНИЕ И УСИЛЕНИЕ ПЕРЕМЕННЫХ ТОКОВ С ПОМОЩЬЮ p-n-ПЕРЕХОДОВ
§ 1. Статическая вольтамперная характеристика p —n-перехода 274
§ 2. p — n-переход при переменном напряжении 277
§ 3. Туннельный эффект в p — n-переходах. Туннельные диоды 282
§ 4. Биполярный полупроводниковый триод 285
§ 5. Гетеропереходы 289
Глава IX
СТАТИСТИКА РЕКОМБИНАЦИИ ЭЛЕКТРОНОВ И ДЫРОК
§ 1. Различные типы процессов рекомбинации 294
§ 2. Темп рекомбинации зона — зона 295
§ 3. Время жизни при излучательной рекомбинации 298
§ 4. Рекомбинация через примеси и дефекты 303
§ 5. Нестационарные процессы 307
а. Монополярное возбуждение (307).
б. Биполярное возбуждение (308)
§ 6. Стационарные состояния 310
§ 7. Многозарядные ловушки 314
Глава X
ПОВЕРХНОСТНЫЕ ЭЛЕКТРОННЫЕ СОСТОЯНИЯ
§ 1. Происхождение поверхностных состояний 317
§ 2. Влияние поверхностного потенциала на электропроводность 322
§ 3. Эффект поля 325

§ 4. Некоторые эффекты, связанные с поверхностными состояниями 332
§ 5. Скорость поверхностной рекомбинации 335
§ 6. Влияние поверхностной рекомбинации на фотопроводимость 338
а. Стационарная фотопроводимость при объемной однородной генерации (338).
б. Стационарная фотопроводимость при поверхностной генерации (340).
§ 7. Затухание фотопроводимости в тонких пластинках и нитевидных образцах 341
§ 8. Зависимость поверхностной рекомбинации от поверхностного потенциала 343
§ 9. Ток насыщения диодов 345
Глава XI
ФОТОЭЛЕКТРОДВИЖУЩИЕ СИЛЫ
§ 1. Роль неосновных носителей 347
§ 2. Фотоэдс в однородных полупроводниках 350
§ 3. Объемная фотоэдс 351
§ 4. Вентильная фотоэдс 355
§ 5. Вентильные фотоэлементы 359
§ 6. Поверхностная фотоэдс 365
§ 7. Фотоэлектромагнитный эффект 366
Глава XII
КОЛЕБАНИЯ КРИСТАЛЛИЧЕСКОЙ РЕШЕТКИ
§ 1. Малые колебания 374
§ 2. Нормальные координаты 376
§ 3. Частоты нормальных колебаний. Акустические и оптические ветви 384
§ 4. Вектор смещения 389
§ 5. Квантовомеханическое рассмотрение колебаний решетки 391
§ 6. Фононы 394
Глава XIII
ЭЛЕМЕНТЫ КИНЕТИЧЕСКОЙ ТЕОРИИ ЯВЛЕНИЙ ПЕРЕНОСА
§ 1. Феноменологические соотношения 399
а. Носители заряда в постоянном и однородном слабом электрическом поле (399).
б. Носители заряда в постоянном и однородном слабом температурном поле (400).
в. Носители заряда в постоянных и однородных электрическом и магнитном полях (403).
§ 2. Кинетические коэффициенты и функция распределения 406
§ 3. Кинетическое уравнение 408
§ 4. Термодинамическое равновесие. Принцип детального равновесия 413
§ 5. Малые отклонения от равновесия 416
§ 6. Интеграл столкновений в случае упругого рассеяния и изотропных изоэнергетических поверхностей. Время релаксации импульса 418
§ 7. Элементарные стационарные решения кинетического уравнения в случае малых отклонений от равновесия 424

а. Статическая электропроводность (424).
б. Термоэдс и коэффициент Пельтье (429).
в. Постоянная Холла и магнетосопротивление (434).
§ 8. Носители заряда в слабом переменном электрическом поле 439
§ 9. Плазменные волны 444
Глава XIV
РАССЕЯНИЕ НОСИТЕЛЕЙ ЗАРЯДА В НЕИДЕАЛЬНОЙ КРИСТАЛЛИЧЕСКОЙ РЕШЕТКЕ
§ 1. Постановка задачи. Теория возмущений 447
§ 2. Вероятность перехода. Условие применимости кинетического уравнения 448
§ 3. Энергия взаимодействия носителей заряда с фононами 453
а. Общие соображения (453).
б. Взаимодействие носителей заряда с акустическими фононами; метод потенциала деформации (454).
в. Взаимодействие носителей заряда с оптическими фононами в гомеополярном кристалле; метод потенциала деформации (457).
г. Взаимодействие носителей заряда с оптическими фононами в гетерополярном кристалле (459).
д. Взаимодействие носителей заряда с пьезоэлектрическими колебаниями решетки (461).
е. Сводка формул (464).
§ 4. Рассеяние носителей заряда фононами 465
§ 5. Рассеяние носителей заряда примесными атомами 474
§ 6. Подвижность, холл-фактор и термоэдс при различных механизмах рассеяния 481
§ 7. Одновременное действие нескольких механизмов рассеяния 484
Глава XV
АКУСТО-ЭЛЕКТРОННЫЕ ЯВЛЕНИЯ
§ 1. Предварительные замечания 488
§ 2. Взаимодействие упругих волн с электронами проводимости 490
§ 3. Упругие волны в пьезодиэлектриках 493
§ 4. Упругие волны в пьезоэлектрических полупроводниках 494
§ 5. Электронное поглощение и усиление ультразвуковых волн 500
§ 6. Акусто-электрический эффект 502
§ 7. Случай $ql \gg 1$ 505
§ 8. Усиление тепловых флуктуаций 508
§ 9. Заключительные замечания 511
Глава XVI
ГОРЯЧИЕ ЭЛЕКТРОНЫ
§ 1. Нагрев электронного газа 513
§ 2. Симметричная и антисимметричная части функции распределения 518
§ 3. Уравнения баланса 520
§ 4. Электронная температура 521
§ 5. Роль неупругости рассеяния 527
§ 6. Зависимость подвижности и концентрации носителей заряда от напряженности поля. 529
§ 7. Дифференциальная проводимость 537
§ 8. Флуктуационная неустойчивость 540
§ 9. Электрические домены и 542

токовые шнуры
§ 10. Движущиеся и статические домены 545
Глава XVII
ПРОБЛЕМЫ ОБОСНОВАНИЯ ЗОННОЙ ТЕОРИИ И ЗАДАЧИ, ВЫХОДЯЩИЕ ЗА ЕЕ РАМКИ
§ 1. Три вопроса к зонной теории 547
§ 2. Адиабатическое приближение 548
§ 3. Приближение малых колебаний 552
§ 4. Роль колебаний решетки. Полярон 553
§ 5. Метод самосогласованного поля 556
§ 6. Электроны и дырки как элементарные возбуждения многоэлектронной системы в полупроводнике 560
§ 7. Экситон 563.
а. Решения, принадлежащие непрерывному спектру (565). б. Решения, принадлежащие дискретному спектру (565).
§ 8. Мелкие локальные уровни при учете экранирования примесных центров 567
§ 9. Механизмы рекомбинации 569
Глава XVIII
ОПТИКА ПОЛУПРОВОДНИКОВ
§ 1. Поглощение и испускание света полупроводниками. Феноменологические соотношения 576
а. Непоглощающая среда (579).
б. Слабое поглощение волны достаточно большой частоты (580).
§ 2. Механизмы поглощения 581
§ 3. Поглощение и отражение электромагнитных волн газом свободных носителей заряда 583
§ 4. Коэффициенты поглощения и излучения при оптических переходах зона — зона 587
§ 5. Прямые и непрямые переходы 591
§ 6. Полупроводниковые лазеры 594
§ 7. Коэффициенты поглощения при прямых переходах. Комбинированная плотность состояний 599
§ 8. Критические точки 604
§ 9. Непрямые переходы 610
§ 10. Электрооптика 613
§11. Модуляционная спектроскопия 614
§ 12. Магнетооптика 615
а. Магнетоплазменные эффекты (615).
б. Междузонные переходы В квантующем магнитном поле (616).
Глава XIX
СИЛЬНО ЛЕГИРОВАННЫЕ ПОЛУПРОВОДНИКИ
§ 1. Примесные уровни и примесные зоны 617
§ 2. Особенности сильно легированных полупроводников 620
§ 3. Иерархия энергий 626
§ 4. Плотность состояний 628
§ 5. Хвост плотности состояний 632
§ 6. Междузонные оптические переходы в сильно легированных полупроводниках 635
§ 7. Некристаллические полупроводники 639
ПРИЛОЖЕНИЯ
I. К доказательству теоремы Блоха 643
11. Интегралы с функциями Блоха 644

III. Таблица значений интеграла $\Phi_{1/2}$; 646
IV. Дельта-функция 647
V. Рекомбинация через многозарядные ловушки 648
VI. Интеграл поверхностной проводимости 650
VII. Диффузия неравновесных носителей заряда в магнитом поле 652
VIII. Вычисление суммы (XII. 2.6) 655
IX. Вывод условия ортогональности (XII.2.11) 655
X. Переход от суммирования по дискретным компонентам квазиимпульса к интегрированию 656
XI. Гамильтониан взаимодействия электронов с акустическими фононами 657
XII. Потенциал заряженного центра при учете экранирования свободными носителями заряда 659
XIII. Усреднение по координатам примесных атомов 66)
XIV. Теорема об интеграле от периодической функции 664
XV. Интегралы с функцией Ферми в условиях сильного вырождения 664
Литература 666
Основные обозначения 670

ОГЛАВЛЕНИЕ

Предисловие 10

Глава I

НЕКОТОРЫЕ СВОЙСТВА ПОЛУПРОВОДНИКОВ

§ 1. Кинетические явления в полупроводниках 13
а. Электропроводность (13). б. Эффект Холла (15). в. Изменение сопротивления в магнитном поле (17). г. Термоэдс (18). д. Эффект Томсона (19). е. Эффект Пельтье (20). ж. Эффект Нернста — Эттингсгаузена (21). з. Эффект Риги — Ледюка (22). и. Продольные термомагнитные эффекты (23).
§ 2. Время релаксации 23
§ 3. Элементарная теория гальваномагнитных явлений 25
а. Тензор электропроводности в магнитном поле (25). б. Угол Холла и постоянная Холла (28). в. Магнетосопротивление (30).
§ 4. Смешанная проводимость 32
а. Эффект Холла (32). б. Магнетосопротивление (35).
§ 5. Некоторые экспериментальные результаты 35
а. Электронная и дырочная проводимость (35). б. Собственная и примесная проводимость (36). в. Запрещенная энергетическая зона (40). г. Удельная электропроводность (42). д. Подвижности (43). е. Собственная концентрация электронов (45). ж. Магнетосопротивление (47).

Глава II

ХИМИЧЕСКИЕ СВЯЗИ В ПОЛУПРОВОДНИКАХ

§ 1. Кристаллические решетки 50
§ 2. Электронная конфигурация атомов 53
§ 3. Типы химической связи 56
а. Ионная связь (57). б. Гомеополярная связь (57). в. Ван-дер-ваальсовская связь (60).
§ 4. Строение некоторых полупроводниковых кристаллов 60
а. Ионные кристаллы (60). б. Гомеополярные кристаллы (61). в. Кристаллы со смешанными связями (62).
§ 5. Некристаллические полупроводники 64
а. Аморфные полупроводники (64). б. Жидкие полупроводники (65). в. Стеклообразные полупроводники (68).
§ 6. Запрещенная зона энергий 69
§ 7. Полупроводниковые свойства и химическая связь 70
§ 8. Полупроводники с малой подвижностью 71
§ 9. Примесные атомы 73
§ 10. Вакансии и междоузельные атомы 78
§ 11. Дислокации 81

Глава III

ЭЛЕМЕНТЫ ЗОННОЙ ТЕОРИИ ТВЕРДОГО ТЕЛА
I. ИДЕАЛЬНАЯ РЕШЕТКА

§ 1. Основные предположения 87
§ 2. Волновая функция электрона в периодическом поле 89
§ 3. Зоны Бриллюэна 94
§ 4. Энергетические зоны 100
§ 5. Метод сильно связанных электронов 103
§ 6. Закон дисперсии. Изоэнергетические поверхности 112
§ 7. Металлы и полупроводники 114
§ 8. Эффективная масса 116
§ 9. Зонная структура некоторых полупроводников 122

Глава IV

ЭЛЕМЕНТЫ ЗОННОЙ ТЕОРИИ ТВЕРДОГО ТЕЛА
II. КРИСТАЛЛЫ ВО ВНЕШНИХ ПОЛЯХ. НЕИДЕАЛЬНЫЕ КРИСТАЛЛЫ

§ 1. Средние значения скорости и ускорения электрона в кристаллической решетке 129
§ 2. Электроны и дырки 133
§ 3. Движение носителей заряда в постоянном и однородном магнитном поле (классическая теория). Диамагнитный резонанс 137
§ 4. Метод эффективной массы 145
§ 5. Энергетический спектр носителя заряда в постоянном и однородном магнитном поле (квантовая теория) 149
§ 6. Движение и энергетический спектр носителей заряда в постоянном электрическом поле 154
§ 7. Мелкие примесные уровни в гомеополярном кристалле 160

Глава V

СТАТИСТИКА ЭЛЕКТРОНОВ И ДЫРОК В ПОЛУПРОВОДНИКАХ

§ 1. Введение 167
§ 2. Распределение квантовых состояний в зонах 168
§ 3. Распределение Ферми — Дирака 169
§ 4. Концентрации электронов и дырок в зонах 170
§ 5. Невырожденные полупроводники 172
§ 6. Случай сильного вырождения 174
§ 7. Эффективная масса плотности состояний 175
§ 8. Плотность состояний в квантующем магнитном поле 179
§ 9. Концентрации электронов и дырок на локальных уровнях. Простые центры 181
§ 10. Многозарядные центры 185
§ 11. Распределение Гиббса 186
§ 12. Частные случаи 189
§ 13. Определение положения уровня Ферми 191
§ 14. Уровень Ферми в собственном полупроводнике 193
§ 15. Полупроводник с примесью одного типа 194
§ 16. Взаимная компенсация доноров и акцепторов 195
§ 17. Компенсированные полупроводники 197
§ 18. Определение энергетических уровней примесных атомов 199
а. Многозарядные акцепторы в полупроводнике *n*-типа (199). б. Многозарядные акцепторы в полупроводнике *p*-типа (201). в. Многозарядные доноры в полупроводнике *n*-типа (202). г. Многозарядные доноры в полупроводнике *p*-типа (202).

Глава VI

ЯВЛЕНИЯ В КОНТАКТАХ (МОНОПОЛЯРНАЯ ПРОВОДИМОСТЬ)

§ 1. Потенциальные барьеры 205
§ 2. Плотность тока. Соотношение Эйнштейна 207
§ 3. Условия равновесия контактирующих тел 209
§ 4. Термоэлектронная работа выхода 210
§ 5. Контактная разность потенциалов 213
§ 6. Распределение концентрации электронов и потенциала в слое объемного заряда 216
§ 7. Длина экранирования 218
§ 8. Обогащенный контактный слой в отсутствие тока 220
§ 9. Истощенный контактный слой 222
§ 10. Токи, ограниченные пространственным зарядом 226
§ 11. Выпрямление в контакте металл — полупроводник 232
§ 12. Диффузионная теория 236
§ 13. Сравнение с экспериментом 239

Глава VII

НЕРАВНОВЕСНЫЕ ЭЛЕКТРОНЫ И ДЫРКИ

§ 1. Неравновесные носители заряда 243
§ 2. Время жизни неравновесных носителей заряда 244
§ 3. Уравнения непрерывности 247
§ 4. Фотопроводимость 250
§ 5. Квазиуровни Ферми 255
§ 6. Электронно-дырочные переходы 258
§ 7. Обнаружение неравновесных носителей заряда 261
§ 8. Амбиполярная диффузия и амбиполярный дрейф 264
§ 9. Длины диффузии и дрейфа 268
§ 10. $n^+ - n$- и $p^+ - p$- переходы 271

Глава VIII

ВЫПРЯМЛЕНИЕ И УСИЛЕНИЕ ПЕРЕМЕННЫХ ТОКОВ С ПОМОЩЬЮ p—n-ПЕРЕХОДОВ

§ 1. Статическая вольтамперная характеристика $p - n$-перехода 274
§ 2. $p - n$-переход при переменном напряжении 277
§ 3. Туннельный эффект в $p - n$-переходах. Туннельные диоды 282
§ 4. Биполярный полупроводниковый триод 285
§ 5. Гетеропереходы 289

Глава IX

СТАТИСТИКА РЕКОМБИНАЦИИ ЭЛЕКТРОНОВ И ДЫРОК

§ 1. Различные типы процессов рекомбинации 294
§ 2. Темп рекомбинации зона — зона 295
§ 3. Время жизни при излучательной рекомбинации 298
§ 4. Рекомбинация через примеси и дефекты 303
§ 5. Нестационарные процессы 307
а. Монополярное возбуждение (307). б. Биполярное возбуждение (308)
§ 6. Стационарные состояния 310
§ 7. Многозарядные ловушки 314

Глава X

ПОВЕРХНОСТНЫЕ ЭЛЕКТРОННЫЕ СОСТОЯНИЯ

§ 1. Происхождение поверхностных состояний 317
§ 2. Влияние поверхностного потенциала на электропроводность 322
§ 3. Эффект поля 325
§ 4. Некоторые эффекты, связанные с поверхностными состояниями 332
§ 5. Скорость поверхностной рекомбинации 335
§ 6. Влияние поверхностной рекомбинации на фотопроводимость 338
а. Стационарная фотопроводимость при объемной однородной генерации (338). б. Стационарная фотопроводимость при поверхностной генерации (340).
§ 7. Затухание фотопроводимости в тонких пластинках и нитевидных образцах 341
§ 8. Зависимость поверхностной рекомбинации от поверхностного потенциала 343
§ 9. Ток насыщения диодов 345

Глава XI

ФОТОЭЛЕКТРОДВИЖУЩИЕ СИЛЫ

§ 1. Роль неосновных носителей 347
§ 2. Фотоэдс в однородных полупроводниках 350
§ 3. Объемная фотоэдс 351
§ 4. Вентильная фотоэдс 355
§ 5. Вентильные фотоэлементы 359
§ 6. Поверхностная фотоэдс 365
§ 7. Фотоэлектромагнитный эффект 366

Глава XII

КОЛЕБАНИЯ КРИСТАЛЛИЧЕСКОЙ РЕШЕТКИ

§ 1. Малые колебания 374
§ 2. Нормальные координаты 376
§ 3. Частоты нормальных колебаний. Акустические и оптические ветви . . . 384
§ 4. Вектор смещения 389
§ 5. Квантовомеханическое рассмотрение колебаний решетки 391
§ 6. Фононы 394

Глава XIII

ЭЛЕМЕНТЫ КИНЕТИЧЕСКОЙ ТЕОРИИ ЯВЛЕНИЙ ПЕРЕНОСА

§ 1. Феноменологические соотношения 399
а. Носители заряда в постоянном и однородном слабом электрическом поле (399). б. Носители заряда в постоянном и однородном слабом температурном поле (400). в. Носители заряда в постоянных и однородных электрическом и магнитном полях (403).
§ 2. Кинетические коэффициенты и функция распределения 406
§ 3. Кинетическое уравнение 408
§ 4. Термодинамическое равновесие. Принцип детального равновесия 413
§ 5. Малые отклонения от равновесия 416
§ 6. Интеграл столкновений в случае упругого рассеяния и изотропных изоэнергетических поверхностей. Время релаксации импульса 418

§ 7. Элементарные стационарные решения кинетического уравнения в случае малых отклонений от равновесия 424
а. Статическая электропроводность (424). б. Термоэдс и коэффициент Пельтье (429). в. Постоянная Холла и магнетосопротивление (434).
§ 8. Носители заряда в слабом переменном электрическом поле 439
§ 9. Плазменные волны 444

Глава XIV

РАССЕЯНИЕ НОСИТЕЛЕЙ ЗАРЯДА В НЕИДЕАЛЬНОЙ КРИСТАЛЛИЧЕСКОЙ РЕШЕТКЕ

§ 1. Постановка задачи. Теория возмущений 447
§ 2. Вероятность перехода. Условие применимости кинетического уравнения 448
§ 3. Энергия взаимодействия носителей заряда с фононами 453
а. Общие соображения (453). б. Взаимодействие носителей заряда с акустическими фононами; метод потенциала деформации (454). в. Взаимодействие носителей заряда с оптическими фононами в гомеополярном кристалле; метод потенциала деформации (457). г. Взаимодействие носителей заряда с оптическими фононами в гетерополярном кристалле (459). д. Взаимодействие носителей заряда с пьезоэлектрическими колебаниями решетки (461). е. Сводка формул (464).
§ 4. Рассеяние носителей заряда фононами 465
§ 5. Рассеяние носителей заряда примесными атомами 474
§ 6. Подвижность, холл-фактор и термоэдс при различных механизмах рассеяния 481
§ 7. Одновременное действие нескольких механизмов рассеяния 484

Глава XV

АКУСТО-ЭЛЕКТРОННЫЕ ЯВЛЕНИЯ

§ 1. Предварительные замечания 488
§ 2. Взаимодействие упругих волн с электронами проводимости 490
§ 3. Упругие волны в пьезодиэлектриках 493
§ 4. Упругие волны в пьезоэлектрических полупроводниках 494
§ 5. Электронное поглощение и усиление ультразвуковых волн 500
§ 6. Акусто-электрический эффект 502
§ 7. Случай $ql \gg 1$ 505
§ 8. Усиление тепловых флуктуаций 508
§ 9. Заключительные замечания 511

Глава XVI

ГОРЯЧИЕ ЭЛЕКТРОНЫ

§ 1. Нагрев электронного газа 513
§ 2. Симметричная и антисимметричная части функции распределения . . . 518
§ 3. Уравнения баланса 520
§ 4. Электронная температура 521
§ 5. Роль неупругости рассеяния 527
§ 6. Зависимость подвижности и концентрации носителей заряда от напряженности поля. 529
§ 7. Дифференциальная проводимость 537
§ 8. Флуктуационная неустойчивость 540
§ 9. Электрические домены и токовые шнуры 542
§ 10. Движущиеся и статические домены 545

Глава XVII

ПРОБЛЕМЫ ОБОСНОВАНИЯ ЗОННОЙ ТЕОРИИ И ЗАДАЧИ, ВЫХОДЯЩИЕ ЗА ЕЕ РАМКИ

§ 1. Три вопроса к зонной теории 547
§ 2. Адиабатическое приближение 548
§ 3. Приближение малых колебаний 552
§ 4. Роль колебаний решетки. Полярон 553
§ 5. Метод самосогласованного поля 556
§ 6. Электроны и дырки как элементарные возбуждения многоэлектронной системы в полупроводнике 560
§ 7. Экситон 563
а. Решения, принадлежащие непрерывному спектру (565). б. Решения, принадлежащие дискретному спектру (565).
§ 8. Мелкие локальные уровни при учете экранирования примесных центров 567
§ 9. Механизмы рекомбинации 569

Глава XVIII

ОПТИКА ПОЛУПРОВОДНИКОВ

§ 1. Поглощение и испускание света полупроводниками. Феноменологические соотношения 576
а. Непоглощающая среда (579). б. Слабое поглощение волны достаточно большой частоты (580).
§ 2. Механизмы поглощения 581
§ 3. Поглощение и отражение электромагнитных волн газом свободных носителей заряда 583
§ 4. Коэффициенты поглощения и излучения при оптических переходах зона — зона 587
§ 5. Прямые и непрямые переходы 591
§ 6. Полупроводниковые лазеры 594
§ 7. Коэффициенты поглощения при прямых переходах. Комбинированная плотность состояний 599
§ 8. Критические точки 604
§ 9. Непрямые переходы 610
§ 10. Электрооптика 613
§ 11. Модуляционная спектроскопия 614
§ 12. Магнетооптика 615
а. Магнетоплазменные эффекты (615). б. Междузонные переходы в квантующем магнитном поле (616).

Глава XIX

СИЛЬНО ЛЕГИРОВАННЫЕ ПОЛУПРОВОДНИКИ

§ 1. Примесные уровни и примесные зоны 617
§ 2. Особенности сильно легированных полупроводников 620
§ 3. Иерархия энергий 626
§ 4. Плотность состояний 628
§ 5. Хвост плотности состояний 632
§ 6. Междузонные оптические переходы в сильно легированных полупроводниках 635
§ 7. Некристаллические полупроводники 639

ПРИЛОЖЕНИЯ

I. К доказательству теоремы Блоха 643
II. Интегралы с функциями Блоха 644
III. Таблица значений интеграла $\Phi_{1/2}$ 646
IV. Дельта-функция 647
V. Рекомбинация через многозарядные ловушки 648
VI. Интеграл поверхностной проводимости 650
VII. Диффузия неравновесных носителей заряда в магнитном поле 652
VIII. Вычисление суммы (XII. 2.6) 655
IX. Вывод условия ортогональности (XII.2.11) 655
X. Переход от суммирования по дискретным компонентам квазиимпульса к интегрированию 656
XI. Гамильтониан взаимодействия электронов с акустическими фононами 657
XII. Потенциал заряженного центра при учете экранирования свободными носителями заряда 659
XIII. Усреднение по координатам примесных атомов 661
XIV. Теорема об интеграле от периодической функции 664
XV. Интегралы с функцией Ферми в условиях сильного вырождения 664

Литература 666

Основные обозначения 670

ПРЕДИСЛОВИЕ

Эта книга написана на основе лекций, в течение ряда лет читавшихся авторами для студентов физического факультета Московского университета и факультета физической и квантовой электроники Московского физико-технического института. Вместе с тем, она не является конспектом лекций, а задумана как учебное пособие и поэтому содержит дополнительный материал для желающих изучить предмет более подробно. Первая половина книги (гл. I—XI) примерно соответствует содержанию лекций, читавшихся на 4-м курсе, а вторая предназначена для студентов 5-го курса. В соответствии с этим постепенно изменяется стиль и характер изложения материала.

Книга рассчитана на лиц, владеющих материалом общефизических и математических курсов в объеме программ, принятых на физических факультетах университетов и в физико-технических институтах. Для изучения глав теоретического характера необходимо знать статистическую физику и квантовую механику в объеме обычных университетских курсов.

Авторам всех учебников хорошо известны трудности, связанные с отбором материала. Эти трудности особенно возрастают, когда речь идет о такой широкой и быстро развивающейся области, как физика полупроводников. При этом часто нелегко бывает даже решить, следует ли относить те или иные вопросы к собственно физике полупроводников или уже к другим, смежным, дисциплинам. Далее, в учебник естественно включать лишь материал, достаточно устоявшийся. Однако при этом есть опасность, что книга устареет, еще не успев выйти из печати. Наконец, учебник по объему должен быть таким, чтобы его можно было без труда поднять одной рукой. Единственный выход, который нам удалось найти (и который, по-видимому, не оригинален), состоит в безжалостном отсечении многих частных, хотя и важных, вопросов в сочетании с достаточно подробным изложением задач принципиального характера. Мы опустили также все расчетные (в том числе теоретико-групповые) методы зонной теории твердого тела, методы решения кинетического уравнения в случае анизотропных изоэнергетических поверхностей и некоторые другие, ограничившись изложением лишь тех вопросов теории, с которыми, как нам кажется, должен быть знаком экспериментатор, чтобы сознательно относиться к своей

деятельности. Здесь задачу авторов весьма облегчило существование превосходных теоретических монографий, указанных в списке основной литературы. Дублировать эти книги нам казалось нецелесообразным.

Таким образом, настоящая книга не предназначена служить исчерпывающим руководством по экспериментальным и теоретическим вопросам физики полупроводников. Ее задача — дать представление об основных понятиях и идеях современной физики полупроводников и подготовить читателя к изучению, в случае необходимости, специальных обзоров и оригинальных работ по отдельным вопросам данной области.

В ряде мест мы указываем монографии, в которых содержится более подробное рассмотрение отдельных проблем, лишь затронутых в книге (это особенно относится к сравнительно новым и трудным задачам). Вместе с тем в соответствии со стилем, принятым во многих учебниках, мы старались избегать ссылок на не всегда доступные студентам многочисленные оригинальные статьи (в том числе и на работы авторов). Исключение составляют лишь ссылки на источники, из которых непосредственно взяты воспроизводимые в книге экспериментальные данные, а также отдельные случаи, когда другие литературные источники отсутствуют. Мы надеемся, что наши коллеги, интересные и важные работы которых используются в этой книге, правильно поймут намерения авторов и не будут чувствовать себя ограбленными. Тем же из них, кто останется недоволен этим объяснением, мы можем лишь напомнить, что при изложении законов Ньютона или описании опытов Фарадея ссылки на оригинальные работы этих авторов делаются очень редко.

В книгу не включены задачи и упражнения в виде отдельных приложений, так как уже имеются соответствующие задачники *). Приведенные в тексте расчетные примеры имеют целью только проиллюстрировать применение тех или иных принципиальных положений.

Один из «больных вопросов» современной научной литературы — вопрос об обозначениях. Понятий, требующих количественного описания, больше, чем букв в обычно употребляемых алфавитах. Поэтому некоторые символы пришлось употреблять в разных смыслах. Для частичной компенсации этого недостатка к книге приложен список основных обозначений. Отметим здесь, что векторные индексы обозначаются в книге чаще всего малыми греческими буквами, а по повторяющимся индексам подразумевается суммирование, если не оговорено исключение из этого правила. Формулы в пределах каждой главы имеют двойную нумерацию. Первая цифра указывает номер параграфа, вторая — номер формулы в данном

*) *В. Л. Бонч-Бруевич, И. П. Звягин, И. В. Карпенко, А. Г. Миронов,* Сборник задач по физике полупроводников, «Наука», М., 1968.

параграфе (например, (4.7)). При ссылках на формулы из других глав ставится еще римская цифра, указывающая номер главы (например, (IV. 4.7)).

Все формулы даны в системе единиц Гаусса. Однако численные значения некоторых полупроводниковых характеристик приведены в практических единицах, которые обычно употребляются для этого в литературе (например, подвижности — в единицах $см^2/сек \cdot В$).

Мы выражаем глубокую благодарность М. И. Каганову, взявшему на себя труд рецензирования книги и сделавшему очень много ценных замечаний. Разумеется, он не несет никакой ответственности за возможные неудачные места книги.

Мы благодарны также В. Д. Егорову, П. Е. Зильберману, В. Ф. Киселеву, И. А. Куровой, А. Г. Миронову, Ю. Ф. Новотоцкому-Власову и А. Н. Темчину, прочитавшим некоторые главы и сделавшим ряд полезных замечаний. Один из нас (В. Л. Б.-Б.) весьма признателен О. А. Головиной, В. В. Бонч-Бруевичу и А. К. Куприяновой за помощь при подготовке рукописи к печати.

Апрель 1977 г.

В. Л. Бонч-Бруевич,
С. Г. Калашников

Глава I

НЕКОТОРЫЕ СВОЙСТВА ПОЛУПРОВОДНИКОВ

§ 1. Кинетические явления в полупроводниках

В настоящей главе мы рассмотрим некоторые результаты экспериментального исследования свойств полупроводников. Эти данные существенно способствовали пониманию физических процессов в полупроводниках и явились базой для построения теории.

При исследовании полупроводников большую роль играет изучение так называемых явлений переноса или, иначе, кинетических явлений. Общая причина этих явлений заключается в том, что электроны проводимости при своем движении переносят связанные с ними физические величины: массу, электрический заряд, энергию и др. Вследствие этого при определенных условиях возникают направленные потоки этих величин, приводящие к ряду электрических и тепловых эффектов. Укажем кратко важнейшие кинетические явления, особенно интересные для исследования полупроводников.

а. Электропроводность. В отсутствие внешнего электрического поля электроны в кристалле совершают только тепловое движение со скоростями $\mathbf{v}_T$. Характер этого движения таков, что электрон некоторое время (время свободного пробега) движется приблизительно прямолинейно и равномерно, но затем, в результате взаимодействия с кристаллической решеткой, резко изменяет направление своего движения. Такие процессы изменения импульса мы будем называть в дальнейшем процессами рассеяния импульса. Они аналогичны соударениям атомов, рассматриваемым в кинетической теории газов. Вследствие беспорядочности теплового движения в электронном газе в состоянии теплового равновесия нет преимущественных направлений движения, и поэтому среднее значение тепловой скорости равно нулю. Это значит, что средний поток частиц, а следовательно, и средняя плотность тока для любого направления равны нулю.

При наложении внешнего электрического поля электроны получают дополнительные скорости $\mathbf{v}$ под действием поля. В этом случае результирующее движение электронов уже не является совершенно беспорядочным и возникает направленный поток электрического заряда (электрический ток). Среднее значение скорости упорядоченного движения для одного электрона (вычисленное для

промежутка времени, охватывающего большое число соударений) мы будем обозначать через $\bar{\mathbf{v}}$, а среднее значение этой скорости для всей совокупности электронов — через $\langle \mathbf{v} \rangle \equiv \mathbf{v}_d$. Среднюю скорость упорядоченного движения $\mathbf{v}_d$ называют дрейфовой.

Во многих случаях оказывается, что дрейфовая скорость пропорциональна напряженности электрического поля $\mathfrak{E}$. Поэтому весьма полезно понятие дрейфовой подвижности заряженных частиц μ, которая, по определению, есть дрейфовая скорость, приобретаемая частицей в поле с напряженностью единица. Или, иначе,

$$\mathbf{v}_d = \mu \mathfrak{E}. \tag{1.1}$$

Для электронов подвижность μ отрицательна, для положительных частиц — положительна.

Если имеются заряженные частицы только одного типа, то плотность электрического тока равна

$$\mathbf{j} = en\mathbf{v}_d = en\mu\mathfrak{E}, \tag{1.2}$$

где e — заряд одной частицы *), а n — концентрация подвижных частиц. С другой стороны, согласно закону Ома

$$\mathbf{j} = \sigma\mathfrak{E}, \tag{1.3}$$

где σ — удельная электропроводность вещества. Отсюда

$$\sigma = en\mu. \tag{1.4}$$

Так как при изменении знака e одновременно изменяется и знак подвижности, то σ не зависит от знака заряда частиц.

В изотропных веществах дрейфовая скорость направлена либо параллельно полю (у положительных частиц), либо противоположно полю (у отрицательных частиц), поэтому μ и σ суть скаляры и, соответственно, векторы $\mathbf{j}$ и $\mathfrak{E}$ совпадают по направлению. В анизотропных веществах это уже не имеет места и соотношение между $\mathbf{j}$ и $\mathfrak{E}$ имеет более общий вид:

$$\begin{aligned} j_x &= \sigma_{xx}\mathfrak{E}_x + \sigma_{xy}\mathfrak{E}_y + \sigma_{xz}\mathfrak{E}_z, \\ j_y &= \sigma_{yx}\mathfrak{E}_x + \sigma_{yy}\mathfrak{E}_y + \sigma_{yz}\mathfrak{E}_z, \\ j_z &= \sigma_{zx}\mathfrak{E}_x + \sigma_{zy}\mathfrak{E}_y + \sigma_{zz}\mathfrak{E}_z. \end{aligned} \tag{1.5}$$

Или, в сокращенной записи,

$$j_\alpha = \sigma_{\alpha\beta}\mathfrak{E}_\beta \qquad (\alpha, \beta = x, y, z), \tag{1.5а}$$

где подразумевается суммирование по повторяющемуся индексу. Следовательно, в этом случае явление переноса заряда определяется уже не единственным кинетическим коэффициентом, а совокупностью коэффициентов $\sigma_{\alpha\beta}$, которые являются компонентами тензора 2-го ранга — тензора электропроводности.

*) В этой главе e обозначает алгебраический заряд частицы.

б. Эффект Холла. При наличии внешнего магнитного поля на движущиеся электроны действует сила Лоренца, перпендикулярная к направлениям их скорости и магнитной индукции. Поэтому движение электронов в различных направлениях происходит по-разному и даже полупроводник, изотропный в отсутствие магнитного поля, становится анизотропным. Это обстоятельство приводит к возникновению гальваномагнитных явлений. Важнейшими из них являются эффект Холла и изменение сопротивления в магнитном поле.

Эффект Холла заключается в том, что в проводнике с током, помещенном в магнитное поле, появляются электродвижущие силы и, как следствие, возникает дополнительное электрическое поле. Сущность этого эффекта в простейшем и наиболее важном случае поясняет рис. 1.1. Рассмотрим однородный и изотропный проводник в виде прямоугольного параллелепипеда с электродами на его торцах. Направим прямоугольные оси координат вдоль ребер параллелепипеда и положим, что вектор плотности тока $\mathbf{j}$ параллелен оси X, а магнитная индукция $\mathfrak{B}$ направлена по оси Z. В отсутствие магнитного поля напряженность электрического поля в проводнике $\mathfrak{E}$ совпадает с направлением $\mathbf{j}$ и между двумя поперечными контактами a и b, расположенными в плоскости, перпендикулярной к $\mathbf{j}$, разность потенциалов равна нулю. При включении поперечного магнитного поля между разомкнутыми контактами a и b появляется разность потенциалов, которая изменяет знак при изменении направления тока или магнитного поля.

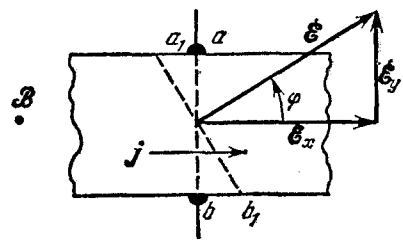


Рис. 1.1. Эффект Холла.

Появление этой разности потенциалов показывает, что в присутствии магнитного поля в проводнике возникает дополнительное электрическое поле $\mathfrak{E}_y$. Направление результирующего электрического поля $\mathfrak{E}$ теперь не совпадает с направлением $\mathbf{j}$, а повернуто относительно $\mathbf{j}$ на некоторый угол φ, который получил название угла Холла. Эквипотенциальные поверхности, которые в отсутствие магнитного поля были плоскостями, перпендикулярными к $\mathfrak{E}_x$ (одна из них проходила через точки a и b), теперь перпендикулярны $\mathfrak{E}$, т. е. повернуты тоже на угол φ (a_1b_1 на рис. 1.1.).

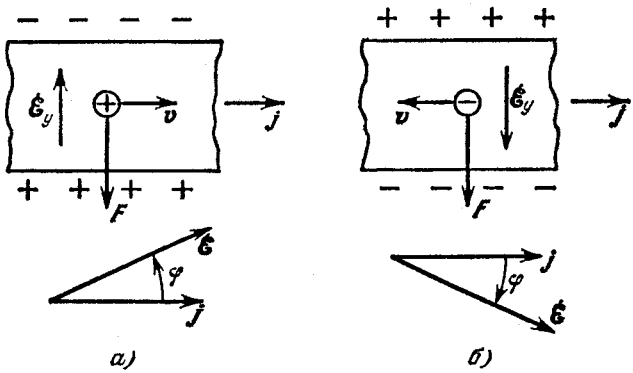
Опыт показывает, что напряженность поля эффекта Холла $\mathfrak{E}_y$ и напряжение Холла u можно выразить формулами

$$\mathfrak{E}_y = \frac{u}{d} = R\mathfrak{B}j = R\mathfrak{B}\frac{i}{ad}. \qquad (1.6)$$

Здесь d — толщина образца, a — его ширина (в направлении магнитного поля), i — полная сила тока, а R — коэффициент пропор-

циональности, который в слабых магнитных полях не зависит от магнитной индукции и характеризует только свойства вещества. Он получил название постоянной Холла.

Знаки угла Холла и постоянной Холла зависят от знака заряда подвижных частиц, обуславливающих электропроводность. Это поясняет рис. 1.2. Если магнитная индукция направлена от плоскости рисунка к читателю и подвижные частицы несут положительный заряд, то при указанном направлении тока сила Лоренца $\mathbf{F}$ будет направлена вниз и нижняя грань кристалла будет заряжаться положительно, а верхняя — отрицательно. Результирующее электрическое поле $\mathcal{E}$ будет повернуто относительно тока $\mathbf{j}$ против часовой стрелки. В этом случае условились считать угол Холла и постоянную Холла положительными.



а) *б)*

Рис. 1.2. Знак угла Холла: *а)* $\varphi > 0$; *б)* $\varphi < 0$.

При отрицательно заряженных частицах сила $\mathbf{F}$ направлена тоже вниз, однако в этом случае нижняя грань кристалла будет заряжаться отрицательно и холловское поле $\mathcal{E}_y$ изменит знак. Соответственно поле $\mathcal{E}$ окажется повернутым по часовой стрелке и φ и R будут отрицательны.

Эффект Холла находит себе различные технические применения. Его можно использовать для измерения напряженности магнитного поля или, если последнее известно, для измерения силы тока и мощности. С помощью эффекта Холла можно генерировать, модулировать и демодулировать электрические колебания, осуществлять квадратичное детектирование колебаний, усиливать электрические сигналы и решать другие технические задачи.

Угол Холла и постоянная Холла выражаются непосредственно через компоненты тензора электропроводности в магнитном поле $\sigma_{\alpha\beta}$. Будем считать сначала, что есть носители заряда только одного типа (обобщение на случай нескольких типов см. в § 4). Тогда их скорость дрейфа направлена вдоль тока (оси X, рис. 1.1.), а сила

Лоренца и поле Холла лежат в плоскости XY, и в соотношениях (1.5) $\mathcal{E}_z = 0$. Далее, мы сразу учтем, что

$$\sigma_{xy} = -\sigma_{yx}, \qquad \sigma_{xx} = \sigma_{yy}. \tag{1.7}$$

В справедливости первого из этих соотношений мы убедимся прямым расчетом (§ 3). Второе очевидно без расчета, так как оси X и Y равноправны по отношению к $\mathscr{B}$. Тогда, полагая в (1.5) $j_y = 0$ (разомкнутые потенциальные зонды a и b), мы имеем

$$\operatorname{tg}\varphi = \frac{\mathcal{E}_y}{\mathcal{E}_x} = -\frac{\sigma_{yx}}{\sigma_{yy}} = \frac{\sigma_{xy}}{\sigma_{xx}}. \tag{1.8}$$

Далее, исключая $\mathcal{E}_x$ из первого соотношения (1.5) и формулы (1.8), находим

$$\mathcal{E}_y = \frac{\sigma_{xy}}{\sigma_{xx}^2 + \sigma_{xy}^2} j_x.$$

Сравнивая это с формулой (1.6), получаем для постоянной Холла

$$R = \frac{1}{\mathscr{B}} \frac{\sigma_{xy}}{\sigma_{xx}^2 + \sigma_{xy}^2}. \tag{1.9}$$

в. Изменение сопротивления в магнитном поле. Внешнее магнитное поле вызывает не только появление холловского поля $\mathcal{E}_y$, но изменяет также ток j_x. Это значит, что сопротивление проводника в поперечном магнитном поле изменяется. Опыт показывает, что в некотором интервале достаточно слабых магнитных полей изменения электропроводности $\Delta\sigma_\perp$ и удельного сопротивления $\Delta\rho_\perp$ подчиняются закону

$$-\frac{\Delta\sigma_\perp}{\sigma} = \frac{\Delta\rho_\perp}{\rho} = \varkappa_\perp \mathscr{B}^2. \tag{1.10}$$

Здесь σ и ρ — значения при $\mathscr{B} = 0$, а $\varkappa_\perp$ — «коэффициент поперечного магнетосопротивления», зависящий от свойств материала.

Эффект магнетосопротивления есть непосредственный результат того, что в магнитном поле электропроводность становится тензором. Поэтому магнетосопротивление можно выразить, как и угол Холла, через компоненты этого тензора. Полагая в соотношениях (1.5), как и раньше, $\mathcal{E}_z = j_y = 0$ и исключая из первых двух соотношений поле $\mathcal{E}_y$, получаем

$$j_x = \left(\sigma_{xx} + \frac{\sigma_{xy}^2}{\sigma_{xx}}\right) \mathcal{E}_x.$$

Удельная электропроводность в поперечном магнитном поле равна

$$\sigma_\perp(\mathscr{B}) = \frac{j_x}{\mathcal{E}_x} = \frac{\sigma_{xx}^2 + \sigma_{xy}^2}{\sigma_{xx}}. \tag{1.11}$$

Отсюда можно выразить и коэффициент магнетосопротивления $\varkappa_\perp$ через компоненты тензора $\sigma_{\alpha\beta}$. Однако мы отложим это до § 3.

Если магнитное поле параллельно току, то в рассматриваемой модели сила Лоренца не возникает, $\sigma_\parallel(\mathscr{H}) = \sigma$, и продольное магнетосопротивление $\Delta\rho_\parallel/\rho = 0$.

Отметим, что формулы (1.8), (1.9) и (1.11) строго справедливы лишь для безграничной среды. В реальных образцах конечных размеров из-за наличия токоподводящих металлических контактов возникают искажения линий тока и линий поля, которые нужно учитывать при измерениях. Чтобы влияние краевых эффектов было мало, длина образцов (в направлении тока) должна быть намного больше их поперечных размеров.

г. Термоэдс. Физически выделенное направление появляется не только во внешнем магнитном поле, но и при наличии градиента температуры, даже если магнитного поля нет. Это приводит к возникновению группы термоэлектрических явлений. Важнейшими термоэлектрическими эффектами в однородных полупроводниках являются термоэлектродвижущая сила (термоэдс) и так называемое явление Томсона.

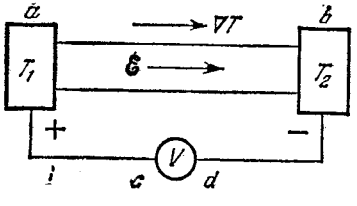


Рис. I.3. Термоэдс. Указанный знак напряжения соответствует положительным носителям заряда и $T_2 > T_1$.

Первый из них заключается в том, что между концами разомкнутого проводника, которые имеют различную температуру, возникает разность потенциалов, а, значит, внутри проводника появляется электродвижущая сила (рис. 1.3).

Происхождение термоэдс объясняется тем, что поток диффузии заряженных частиц от нагретого конца к холодному больше, чем в обратном направлении. Поэтому на концах проводника (и на его поверхности) появляются электрические заряды, а внутри проводника — электрическое поле. В стационарном состоянии разомкнутого проводника это поле таково, что вызываемый им ток дрейфа как раз компенсирует результирующий поток диффузии. Однако полная величина напряжения, регистрируемого вольтметром между металлическими электродами a и b (рис. 1.3), складывается не только из падения напряжения внутри проводника, но еще и из скачков потенциала, существующих на границах полупроводник—металл (контактная разность потенциалов, см. § VI.5), которые зависят от температуры. Если оба металлических электрода одинаковы и разности температур нет, то сумма обоих скачков потенциала равна нулю. Если же температура концов различна, то величина обоих скачков потенциала уже неодинакова, что вносит дополнительный вклад в термоэдс.

Величину эдс dV_0 при малой разности температур между концами dT принято выражать соотношением

$$dV_0 = \alpha \cdot dT. \tag{1.12}$$

Здесь α, по определению, есть дифференциальная термоэдс. Если оба металлических электрода и соединительные провода сделаны из одинакового материала и точки c и d (рис. 1.3) находятся при одинаковой температуре, то α зависит только от свойств полупроводника. Так как эти последние изменяются при изменении температуры, то α тоже зависит от температуры. Величина α у металлов невелика и обычно имеет порядок $1 \div 10$ мкВ/град. У полупроводников она может быть во много сотен и тысяч раз больше.

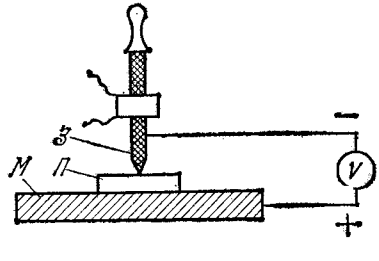


Рис. 1.4. Термозонд. $З$ — нагретый стержень, $П$ — полупроводник, $М$ — холодная металлическая пластина. Знак напряжения показан для положительных частиц.

Определяя полярность термоэлектрического напряжения, можно просто установить знак подвижных носителей заряда. Для этого используют «термозонд» — нагретый металлический стержень, например обычный электрический паяльник, который прижимают к исследуемому полупроводнику. Милливольтметр включают между стержнем и холодной стороной образца (рис. 1.4).

д. Эффект Томсона. Если в однородном проводнике имеется градиент температуры в направлении оси X и в том же направлении течет электрический ток с плотностью j, то в каждой единице объема в единицу времени выделяется, кроме джоулева тепла j^2/σ, еще дополнительное тепло

$$-\alpha_T \cdot j \frac{dT}{dx}. \tag{1.13}$$

Здесь α_T — коэффициент Томсона, зависящий от рода вещества и его состояния. В отличие от джоулева тепла, тепло Томсона пропорционально первой степени j. Поэтому при изменении направления тока на обратное тепло Томсона изменяет знак: вместо поглощения тепла наблюдается его выделение, и наоборот.

При наличии градиента температуры в проводнике имеется еще тепловой поток, обусловленный теплопроводностью вещества. При этом количество тепла, проходящее через единицу поверхности в единицу времени в направлении X, есть

$$-\varkappa \frac{dT}{dx}, \tag{1.14}$$

где $\varkappa$ — коэффициент теплопроводности. Если этот поток изменяется в пространстве (вследствие изменения $\varkappa$ или dT/dx), то в объеме проводника тоже выделяется тепло

$$\frac{d}{dx}\left(\varkappa \frac{dT}{dx}\right).$$

В общем случае, когда направления $\mathbf{j}$ и ∇T не совпадают, полная генерация тепла в единицу времени и в единице объема проводника равна

$$Q_V = \frac{j^2}{\sigma} - \alpha_T (\mathbf{j} \cdot \nabla T) + \operatorname{div} (\varkappa \cdot \nabla T). \tag{1.15}$$

Здесь первое слагаемое дает джоулево тепло, второе — тепло Томсона, а третье обусловлено теплопроводностью. В стационарном состоянии $Q_V = 0$. Поэтому в проводнике устанавливается такое пространственное распределение температуры, при котором тепло, отводимое теплопроводностью, как раз равно сумме тепла Джоуля и тепла Томсона.

е. Эффект Пельтье. Обратимое выделение тепла наблюдается также у границы контакта двух различных проводников. Количество тепла, выделяемое на единице площади контакта в единицу времени, Q_S, равно

$$Q_S = \Pi_{12} \cdot j. \tag{1.16}$$

Здесь j — плотность тока через контакт, а Π_{12} — коэффициент Пельтье, зависящий от свойств контактирующих проводников. Индексами 1, 2 обозначено, что ток предполагается направленным от проводника 1 к проводнику 2. При изменении направления тока на обратное вместо выделения тепла наблюдается его поглощение, и наоборот. Следовательно, $\Pi_{12} = -\Pi_{21}$.

Причина выделения (поглощения) тепла Пельтье заключается в том, что средние энергии электронов E_1 и E_2 в различных проводниках 1 и 2 неодинаковы, даже если оба проводника имеют одну и ту же температуру. При переходе из одного проводника в другой изменяется, во-первых, потенциальная энергия электронов $-e\varphi$, так как на границе имеется скачок электростатического потенциала и поэтому $\varphi_1 \neq \varphi_2$. И, во-вторых, может меняться их средняя кинетическая энергия E_k. Последнее есть результат того, что электроны подчиняются не классической статистике Максвелла—Больцмана, а квантовой статистике Ферми—Дирака, согласно которой E_k зависит не только от температуры, но еще и от концентрации электронов. Поэтому при наличии тока для поддержания температуры контакта постоянной от него необходимо либо отводить энергию, если $E_1 > E_2$ (выделение тепла Пельтье), либо подводить ее к контакту, когда $E_1 < E_2$ (поглощение тепла Пельтье).

Теория явлений переноса энергии электронами (см. гл. XIII) позволяет заключить, что Π_{12} можно представить в виде разности:

$$\Pi_{12} = \Pi_1 - \Pi_2, \tag{1.17}$$

где Π_1 и Π_2 — коэффициенты Пельтье для проводника 1 и, соответственно, проводника 2. Она показывает также, что термоэлектрические кинетические коэффициенты α, α_T и Π не являются независимыми, а связаны соотношениями

$$\Pi = \alpha T, \tag{1.18}$$

$$\alpha_T = T \frac{d\alpha}{dT}. \tag{1.19}$$

Из формулы (1.18) видно, что Π пропорционален α. Поэтому у полупроводников коэффициент Пельтье, так же как и термоэдс, гораздо больше, чем у металлов. С другой стороны, соотношение (1.19) показывает, что тепло Томсона обусловлено температурной зависимостью термоэдс α. Если α не зависит от температуры, то и тепло Томсона равно нулю.

Термоэлектрические явления имеют интересные перспективы технического применения. Термоэдс можно использовать для устройства термоэлектрических генераторов небольшой мощности, предназначенных для прямого преобразования тепловой энергии в электрическую. Применяя эффект Пельтье, можно осуществить различные термоэлектрические охлаждающие устройства. Однако на этих вопросах мы не будем останавливаться и отсылаем читателя к специальным монографиям.

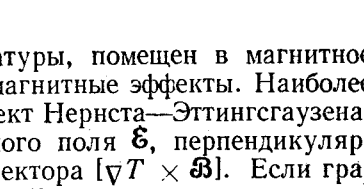


Рис. 1.5. Поперечный термомагнитный эффект Нернста — Эттингсгаузена.

ж. Эффект Нернста—Эттингсгаузена. Когда полупроводник, в котором имеется градиент температуры, помещен в магнитное поле, то возникают различные термомагнитные эффекты. Наиболее интересным является поперечный эффект Нернста—Эттингсгаузена, состоящий в появлении электрического поля $\mathcal{E}$, перпендикулярного к ∇T и $\mathcal{B}$, т. е. в направлении вектора $[\nabla T \times \mathcal{B}]$. Если градиент температуры направлен вдоль оси X, а магнитная индукция — вдоль Z, то электрическое поле параллельно оси Y. Поэтому между поперечными зондами a и b (рис. 1.5) возникает разность электрических потенциалов u. Величину $\mathcal{E}_y$ можно выразить формулой

$$\mathcal{E}_y = \frac{u}{d} = q_\perp \mathcal{B}_z \frac{dT}{dx}. \tag{1.20}$$

Здесь $q_\perp$ — так называемая постоянная Нернста—Эттингсгаузена, которая зависит от свойств полупроводника и может быть как положительной, так и отрицательной.

Чтобы получить представление о величине этого эффекта, укажем, что, например, в германии с удельным сопротивлением ~ 1 Ом·см при комнатных температурах, при $\mathcal{B} \sim 10^3$ Гс и

$dT/dx \sim 10^2$ град/см наблюдаемое $\mathcal{E}_y \sim 10^{-2}$ В/см. Значения $q_\perp$, а следовательно и $\mathcal{E}_y$, сильно зависят от температуры образца и от магнитного поля и при изменении этих величин могут даже изменять знак.

Рассматриваемый эффект возникает по той же причине, что и эффект Холла, т. е. в результате отклонения потока заряженных частиц силой Лоренца. Различие, однако, заключается в том, что при эффекте Холла направленный поток частиц возникает в результате их дрейфа в электрическом поле, а в данном случае — в результате диффузии.

Легко видеть, что, в отличие от постоянной Холла, знак $q_\perp$ не зависит от знака носителей заряда. Действительно, при дрейфе в электрическом поле изменение знака заряда приводит к изменению направления дрейфа, что и дает изменение знака поля Холла (ср. рис. 1.2). В данном же случае поток диффузии всегда направлен от нагретого конца образца к холодному, независимо от знака заряда частиц. Поэтому направления силы Лоренца для положительных и отрицательных частиц взаимно противоположны, однако направление потоков электрического заряда в обоих случаях одно и то же.

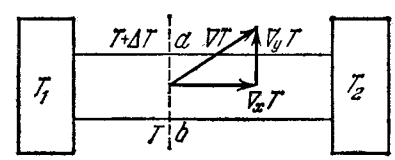


Рис. 1.6. Поперечный термомагнитный эффект Риги — Ледюка.

з. Эффект Риги—Ледюка. В полупроводнике, в котором имеется градиент температуры, при включении магнитного поля появляется также поперечная (по отношению к первоначальному тепловому потоку и направлению $\mathfrak{B}$) разность температур (рис. 1.6). Величину поперечного градиента температуры можно записать в виде

$$\frac{\partial T}{\partial y} = S \cdot \mathfrak{B}_z \frac{\partial T}{\partial x}, \tag{1.21}$$

где S — постоянная Риги—Ледюка, характеризующая свойства данного вещества.

Эффект Риги—Ледюка связан с тем, что диффундирующие носители заряда переносят с собой тепло (теплопроводность). Без магнитного поля поток тепла направлен от горячего конца образца к холодному, т. е. параллельно $-\nabla_x T$. В магнитном поле потоки диффузии и тепла поворачиваются силой Лоренца на некоторый угол. Поэтому возникает составляющая теплового потока вдоль оси Y, что и приводит к появлению составляющей градиента температуры $\nabla_y T$. Так как направление силы Лоренца при данном направлении диффузии зависит от знака заряженных частиц, то угол поворота теплового потока, а значит, и постоянная S имеют разные знаки для положительных и отрицательных носителей заряда.

и. Продольные термомагнитные эффекты. Помимо рассмотренных выше поперечных эффектов, в полупроводниках наблюдаются также продольные термомагнитные эффекты. Если существует первоначальный поток тепла в направлении оси X, то при включении поперечного магнитного поля появляется не только поперечная разность потенциалов, но еще и продольная (продольный эффект Нернста—Эттингсгаузена). Этот эффект можно истолковать как изменение термоэдс в поперечном магнитном поле.

Если после включения магнитного поля тепловой поток вдоль оси X (рис. 1.6) поддерживается постоянным, то, кроме поперечной разности температур (вдоль оси Y), возникает дополнительная продольная разность температур вдоль оси X (продольный эффект Риги—Ледюка). Она появляется вследствие изменения теплопроводности в магнитном поле.

Отметим, что значения кинетических коэффициентов могут изменяться при изменении теплообмена образца с окружающей средой. Поэтому различают изотермические и адиабатические эффекты. Изотермическими называют эффекты, возникающие при условии, что поперечные градиенты температуры (в направлении осей Y и Z) равны нулю. Адиабатическими называются эффекты при условии, что поперечные потоки тепла равны нулю. Исключением является, очевидно, лишь эффект Риги—Ледюка, который, по определению, не может быть изотермическим.

Помимо указанных эффектов, к явлениям переноса относятся, конечно, и процессы диффузии и теплопроводности.

Величины различных кинетических коэффициентов — электропроводности, постоянной Холла, термоэдс и др. — существенно зависят от свойств подвижных носителей заряда: их заряда, массы, энергетического спектра в кристалле, а также от особенностей их взаимодействия с кристаллической решеткой. Поэтому исследование кинетических явлений дает обширную информацию об электронных процессах в полупроводниках.

Теория кинетических явлений будет дана в гл. XIV. Однако уже сейчас целесообразно остановиться на элементарном рассмотрении гальваномагнитных явлений, так как это сразу позволит нам интерпретировать ряд важных экспериментальных фактов.

§ 2. Время релаксации

Движение электронов в кристалле подчиняется законам квантовой механики. Однако в ряде случаев уравнения их движения можно представить в классической форме, если только приписать электрону вместо массы в изолированном состоянии m_0 некоторую другую величину, так называемую эффективную массу (см. гл. IV). Поэтому в настоящей главе мы будем пользоваться корпускулярным способом описания и рассматривать электроны как классиче-

ские частицы. Кроме того, мы сначала будем считать эффективную массу скалярной величиной.

Чтобы найти среднюю скорость упорядоченного движения электронов, рассмотрим закон распределения времен свободного пробега. Положим, что большое число электронов N_0 одновременно испытали соударение в момент времени $t = 0$. Обозначим, далее, через $N(t)$ количество электронов из этой совокупности, не испытавших соударений за последующий промежуток времени от 0 до t. Тогда можно положить, что число электронов dN, которые испытают столкновения в интервале времени $(t, t + dt)$, будет пропорционально N и величине интервала dt, а число электронов N уменьшится на dN. Следовательно,

$$-dN = N\frac{dt}{\tau}, \tag{2.1}$$

где коэффициент пропорциональности обозначен через $1/\tau$. Эта величина, по смыслу, есть вероятность того, что один электрон испытает соударение за единицу времени. Интегрируя написанное уравнение и учитывая, что в начальный момент времени $t = 0$ ни один электрон не успел еще испытать соударение, мы имеем

$$N = N_0 \exp\left(-\frac{t}{\tau}\right).$$

Поэтому для числа соударений, происходящих в интервале времени от t до $t + dt$, получается

$$dN = N_0 \exp\left(-\frac{t}{\tau}\right)\cdot\frac{dt}{\tau}.$$

Вместо того, чтобы рассматривать один свободный пробег у каждого из большой совокупности электронов, мы могли бы следить за движением одного электрона за время, охватывающее большое число соударений. Поэтому, полученные результаты можно применить и к движению одного электрона, если под N_0 понимать полное число свободных пробегов. Разумеется, это число должно быть достаточно большим, чтобы можно было применять законы статистики. Вероятность какому-то одному электрону иметь время свободного пробега в пределах от t до $t + dt$ есть

$$f(t)\,dt = \frac{dt}{\tau}\exp\left(-\frac{t}{\tau}\right). \tag{2.2}$$

При этом, как легко убедиться,

$$\int_0^\infty f(t)\,dt = 1. \tag{2.3}$$

Входящая сюда постоянная τ имеет простой физический смысл. Вычислим среднее время свободного пробега. Последнее, по

определению среднего значения любой величины, есть

$$\bar{t} = \int_0^\infty t f(t)\, dt,$$

где $f(t)$ — нормированная к единице вероятность рассматриваемой величине иметь значение t. Подставляя сюда для $f(t)$ выражение (2.2), мы находим

$$\bar{t} = \tau. \tag{2.4}$$

Таким образом, постоянная τ есть среднее время свободного пробега.

Выше мы считали τ постоянным. Однако это, вообще говоря, неверно. Дело в том, что, говоря о соударениях электронов, мы имеем в виду процессы изменения их импульса, которые возникают вследствие силового взаимодействия электронов с различными нерегулярностями решетки (вызванными тепловым движением, примесными атомами и структурными дефектами). Но результаты этого взаимодействия зависят от состояния движения электрона и, в частности, от его полной энергии. Поэтому τ может быть разным для разных электронов. Одно из простейших предположений заключается в том, что τ зависит только от полной энергии электрона.

Когда мы рассматривали выше распределение времен пробега у совокупности электронов, мы молчаливо предполагали, что либо τ не зависит от энергии, либо энергии всех электронов близки по величине, так что τ можно считать постоянным. Точно так же, применяя распределение (2.2) к одному электрону, мы должны считать, что полная энергия электрона существенно не меняется при его движении.

Для отдельного электрона это предположение во многих случаях справедливо. Это имеет место, если электрические поля не слишком сильны, так что энергия, приобретаемая электроном за один свободный пробег в его упорядоченном движении, мала по сравнению с энергией теплового движения, и если эта энергия не накапливается, а передается решетке при соударениях. При нахождении же среднего для всей совокупности электронов $\langle\tau\rangle$ нужно учитывать, что электроны в твердом теле могут иметь весьма различную энергию, и, вообще говоря, нужно считать τ зависящим от энергии. Величину $\langle\tau\rangle \equiv \tau_p$ называют временем релаксации импульса. Более подробно смысл этой величины будет рассмотрен в гл. XIII.

§ 3. Элементарная теория гальваномагнитных явлений

Тензор электропроводности в магнитном поле. Пользуясь понятием времени релаксации, можно вычислить тензор электропроводности в магнитном поле $\sigma_{\alpha\beta}$, который, как мы видели, опре-

деляет все гальваномагнитные эффекты. Для этого мы сначала рассмотрим движение одной частицы между двумя последовательными соударениями, а затем усредним полученный результат по всем свободным пробегам. Это позволит нам найти скорость дрейфа $\langle \mathbf{v} \rangle$, а следовательно, и плотность тока, откуда непосредственно определяются и компоненты тензора $\sigma_{\alpha\beta}$.

Положим, что магнитная индукция $\mathfrak{B}$ направлена вдоль оси Z прямоугольной системы координат. Тогда составляющие силы Лоренца, действующей на частицу, равны

$$F_x = e\mathfrak{E}_x + \frac{1}{c} e\dot{y}\mathfrak{B}, \quad F_y = e\mathfrak{E}_y - \frac{1}{c} e\dot{x}\mathfrak{B}, \quad F_z = e\mathfrak{E}_z,$$

где e — заряд частицы, а $\dot{x}$ и $\dot{y}$ — составляющие скорости по осям X и Y. Уравнения движения между последовательными соударениями имеют вид

$$\ddot{x} = \frac{e}{m}\mathfrak{E}_x + \omega_c \dot{y}, \quad \ddot{y} = \frac{e}{m}\mathfrak{E}_y - \omega_c \dot{x}, \quad \ddot{z} = \frac{e}{m}\mathfrak{E}_z, \tag{3.1}$$

где через ω_c обозначена «циклотронная» частота:

$$\omega_c = \frac{e\mathfrak{B}}{cm}, \tag{3.2}$$

т. е. частота равномерного вращения частицы в магнитном поле, которая не зависит от радиуса орбиты и энергии частицы.

В качестве начальных условий, как и раньше, примем:

$$t = 0: \quad x = y = z = \dot{x} = \dot{y} = \dot{z} = 0. \tag{3.3}$$

Решение последнего из уравнений (3.1) получается сразу:

$$z = \frac{e}{m}\mathfrak{E}_z t^2. \tag{3.4}$$

Так как при выбранном направлении $\mathfrak{B}$ сила Лоренца лежит в плоскости XY, то движение вдоль Z не изменяется в магнитном поле. Решения же первых двух уравнений (3.1), удовлетворяющие граничным условиям, как нетрудно убедиться непосредственной подстановкой, имеют вид

$$\begin{aligned} x - a &= -a\cos\omega_c t - b\sin\omega_c t + b\omega_c t, \\ y - b &= -b\cos\omega_c t + a\sin\omega_c t - a\omega_c t. \end{aligned} \tag{3.5}$$

Здесь a и b — постоянные размерности длины, равные

$$\begin{aligned} a &= \frac{1}{\omega_c^2}\frac{e}{m}\mathfrak{E}_x, \\ b &= \frac{1}{\omega_c^2}\frac{e}{m}\mathfrak{E}_y. \end{aligned} \tag{3.6}$$

Это движение имеет простой характер. Последние слагаемые в (3.5) описывают переносное движение в плоскости XY с постоянной скоростью v_t. Ее составляющие по осям равны

$$v_{tx} = b\omega_c = \frac{\mathcal{E}_y}{\mathcal{B}}, \quad v_{ty} = -a\omega_c = -\frac{\mathcal{E}_x}{\mathcal{B}}. \quad (3.7)$$

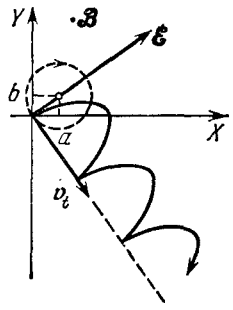


Рис. 1.7. Движение заряженной частицы в скрещенных электрическом и магнитном полях.

При этом скалярное произведение $(\mathbf{v}_t\mathcal{E}) = v_{tx}\mathcal{E}_x + v_{ty}\mathcal{E}_y = 0$, а следовательно, $\mathbf{v}_t$ перпендикулярно $\mathcal{E}$ (и перпендикулярно $\mathcal{B}$). Остальные слагаемые в (3.5) описывают равномерное вращение с частотой ω_c по окружности $(x-a)^2 + (y-b)^2 = a^2 + b^2$, центр которой расположен в точке (a, b). В результате сложения обоих этих движений частица движется по циклоиде, изображенной на рис. 1.7.

Умножая x, y и z на вероятность (2.2) частице иметь время свободного пробега t и интегрируя по t от 0 до ∞, мы получим средние перемещения $\bar{x}$, $\bar{y}$ и $\bar{z}$ для одной частицы, а деля полученные средние на τ, найдем средние скорости $\bar{v}_x$, $\bar{v}_y$, $\bar{v}_z$. Учитывая, что

$$\int_0^\infty \exp\left(-\frac{t}{\tau}\right) \cdot \sin\omega_c t \cdot \frac{dt}{\tau} = \frac{\omega_c\tau}{1+\omega_c^2\tau^2},$$

$$\int_0^\infty \exp\left(-\frac{t}{\tau}\right) \cdot \cos\omega_c t \cdot \frac{dt}{\tau} = \frac{1}{1+\omega_c^2\tau^2},$$

мы получаем

$$\bar{v}_x = \frac{e}{m}\mathcal{E}_x \frac{\tau}{1+\omega_c^2\tau^2} + \omega_c \frac{e}{m}\mathcal{E}_y \frac{\tau^2}{1+\omega_c^2\tau^2},$$

$$\bar{v}_y = \frac{e}{m}\mathcal{E}_y \frac{\tau}{1+\omega_c^2\tau^2} - \omega_c \frac{e}{m}\mathcal{E}_x \frac{\tau^2}{1+\omega_c^2\tau^2}, \quad (3.8)$$

$$\bar{v}_z = \frac{e}{m}\mathcal{E}_z\tau.$$

Чтобы получить дрейфовые скорости, эти значения нужно еще усреднить по всем частицам. Введем для сокращения записи следующие обозначения:

$$\zeta_1 = \left\langle \frac{\tau}{1+\omega_c^2\tau^2} \right\rangle, \quad \zeta_2 = \left\langle \frac{\tau^2}{1+\omega_c^2\tau^2} \right\rangle, \quad (3.9)$$

где $\langle \ldots \rangle$, как и раньше, обозначает усреднение по всем частицам. Тогда для плотности тока получаем

$$j_x = en\langle v_x\rangle = \frac{e^2n}{m}(\zeta_1 \mathcal{E}_x + \omega_c\zeta_2 \mathcal{E}_y),$$
$$j_y = en\langle v_y\rangle = \frac{e^2n}{m}(-\omega_c\zeta_2 \mathcal{E}_x + \zeta_1 \mathcal{E}_y), \qquad (3.10)$$
$$j_z = en\langle v_z\rangle = \frac{e^2n}{m}\langle\tau\rangle \mathcal{E}_z.$$

Сравнивая эти выражения с соотношениями (1.5), находим компоненты тензора электропроводности в виде

$$\sigma_{xx} = \sigma_{yy} = \frac{e^2n}{m}\zeta_1, \qquad \sigma_{xy} = -\sigma_{yx} = \omega_c\frac{e^2n}{m}\zeta_2,$$
$$\sigma_{zz} = \frac{e^2n}{m}\langle\tau\rangle, \qquad \sigma_{xz} = \sigma_{zx} = \sigma_{yz} = \sigma_{zy} = 0. \qquad (3.11)$$

Если магнитного поля нет, то $\omega_c = 0$ и недиагональные компоненты $\sigma_{\alpha\beta}$ равны нулю. При этом $\zeta_1 = \langle\tau\rangle$ и все диагональные компоненты становятся равными. Электропроводность превращается в скалярную величину

$$\sigma = \frac{e^2n}{m}\langle\tau\rangle. \qquad (3.12)$$

Подвижность носителей заряда в отсутствие магнитного поля равна

$$\mu = \frac{e}{m}\langle\tau\rangle. \qquad (3.13)$$

Она определяется величиной (и знаком) удельного заряда и средним временем релаксации.

Для получения численных значений $\sigma_{\alpha\beta}$ необходимо провести усреднение в соответствии с формулами (3.9) для ζ_1 и ζ_2. Для этого надо явно определить правило усреднения, обозначаемого символом $\langle \ldots \rangle$. К этому мы вернемся в гл. XIII, а сейчас ограничимся результатами, которые не требуют фактического проведения указанного усреднения.

б. Угол Холла и постоянная Холла. Подставляя найденные значения (3.11) для σ_{xx} и σ_{xy} в формулу (1.8), находим для угла Холла выражение

$$\operatorname{tg}\varphi = \frac{\sigma_{xy}}{\sigma_{xx}} = \omega_c\frac{\zeta_2}{\zeta_1}.$$

Или, подставляя для ω_c ее значение (3.2),

$$\operatorname{tg}\varphi = \frac{1}{c}\mu_H\mathcal{B}, \qquad (3.14)$$

где введено обозначение

$$\mu_H = \frac{e}{m}\frac{\zeta_2}{\zeta_1}. \qquad (3.15)$$

Так как отношение ζ_2/ζ_1 имеет размерность времени, то μ_H имеет размерность подвижности. Однако она, вообще говоря, не равна дрейфовой подвижности μ, выражаемой формулой (3.13). Подвижность μ_H, определяемая из эффекта Холла, получила название холловской подвижности.

В слабых магнитных полях, определяемых условием

$$\operatorname{tg}\varphi \simeq \omega_c\tau \ll 1, \tag{3.16}$$

мы имеем $\zeta_1 \simeq \langle\tau\rangle$, $\zeta_2 \simeq \langle\tau^2\rangle$. В этом случае

$$\mu_H = \frac{e}{m}\frac{\langle\tau^2\rangle}{\langle\tau\rangle}. \tag{3.15а}$$

Постоянная Холла получается непосредственно из формулы (1.9). Ограничиваясь случаем не очень сильных магнитных полей, удовлетворяющих условию (3.16), мы имеем $\sigma_{xy} \ll \sigma_{xx}$ и, кроме того, $\sigma_{xx} \simeq \sigma = en\mu$. Тогда

$$R \simeq \frac{\sigma_{xy}}{\mathscr{B}\sigma^2} = \frac{\operatorname{tg}\varphi}{\mathscr{B}\sigma} = \gamma\frac{1}{cen}, \tag{3.17}$$

где $\gamma = \mu_H/\mu$. Отсюда видно, что при известном «холловском факторе» γ из измерений постоянной Холла можно определить концентрацию носителей заряда n. Произведение же постоянной Холла на удельную электропроводность равно

$$R\sigma = \frac{1}{c}\mu_H \tag{3.18}$$

и дает холловскую подвижность μ_H.

Остановимся теперь на факторе γ. Для слабых магнитных полей из формул (3.15а) и (3.13) имеем

$$\gamma = \frac{\langle\tau^2\rangle}{\langle\tau\rangle^2}. \tag{3.19}$$

Так как среднее значение квадрата всегда больше (или равно) квадрата среднего значения, то всегда $\gamma \geqslant 1$. Если τ не зависит от энергии, то $\langle\tau\rangle = \tau$, $\langle\tau^2\rangle = \tau^2$ и поэтому $\gamma = 1$. Для определения значения γ необходимо знать зависимость времени релаксации от энергии. Она определяется тем, какие типы процессов рассеяния импульса играют главную роль в рассматриваемом полупроводнике при данной температуре, и поэтому γ имеет различное значение в разных случаях. Сейчас мы укажем без вывода (см. гл. XIV) некоторые наиболее важные случаи.

Если полупроводник относительно чист (т. е. не содержит примесей в больших концентрациях), его температура достаточно высока и, кроме того, концентрация подвижных частиц не слишком велика (так называемый невырожденный полупроводник), то главную роль играют процессы рассеяния на тепловых колебаниях решетки. В этом случае $\gamma = 3\pi/8 = 1{,}18$.

Если, напротив, полупроводник находится при низкой температуре, когда решеточное рассеяние мало, и содержит значительное количество примесей, атомы которых заряжены, то главным процессом является рассеяние на заряженных примесях. Тогда расчет показывает, что $\gamma = 315\pi/512 \simeq 1{,}93$.

Наконец, если мы имеем проводник с очень большой концентрацией подвижных частиц, как это имеет место, например, в металлах (так называемые вырожденные проводники), то $\gamma = 1$.

Однако во всех указанных случаях γ оказывается порядка единицы и, более того, различные его значения лежат в сравнительно узком интервале от 1 до 2. Поэтому во многих случаях, где не требуется большая точность, вопрос о типе главного процесса рассеяния не очень существен и приближенно можно считать $\gamma \simeq 1$.

в. Магнетосопротивление. Подставляя в соотношение (1.11) значения σ_{xx} и σ_{xy} из формул (3.11), получаем

$$\sigma_\perp(\mathscr{B}) = \frac{e^2 n}{m}\left(\zeta_1 + \omega_c^2 \frac{\zeta_2^2}{\zeta_1}\right).$$

Найдем теперь относительное изменение электропроводности $\Delta\sigma_\perp/\sigma$, где $\Delta\sigma_\perp = \sigma_\perp(\mathscr{B}) - \sigma$, а $\sigma = e^2 n\langle\tau\rangle/m$ — электропроводность без магнитного поля. Учитывая, что

$$\zeta_1 - \langle\tau\rangle = \left\langle \frac{\tau}{1+\omega_c^2\tau^2} - \tau \right\rangle = -\omega_c^2 \zeta_3,$$

где для краткости введено обозначение

$$\zeta_3 = \left\langle \frac{\tau^3}{1+\omega_c^2\tau^2} \right\rangle, \tag{3.20}$$

получаем

$$-\frac{\Delta\sigma_\perp}{\sigma} = \frac{\Delta\rho_\perp}{\rho} = \omega_c^2 \frac{\zeta_3 - \zeta_2^2/\zeta_1}{\langle\tau\rangle}. \tag{3.21}$$

Так как ω_c пропорциональна магнитной индукции $\mathscr{B}$, а все величины в (3.21) содержат только ω_c^2, то изменение сопротивления в магнитном поле (как и следовало ожидать) есть четный эффект, т. е. не зависит от направления магнитного поля. Однако зависимость $\Delta\sigma$ от $\mathscr{B}$, вообще говоря, не квадратичная, а более сложная.

Как и для эффекта Холла, важным предельным случаем являются слабые магнитные поля, определяемые условием (3.16). Тогда выражения для ζ_1, ζ_2 и ζ_3 можно разложить в ряд по степеням $(\omega_c\tau)^2$ и удержать в (3.21) только члены разложения со степенью не выше $(\omega_c\tau)^2$. После этого (3.21) дает

$$-\frac{\Delta\sigma_\perp}{\sigma} = \frac{\Delta\rho_\perp}{\rho} = \varkappa_\perp^2 \mathscr{B}^2, \tag{3.22}$$

где

$$\varkappa_\perp^2 = \left(\frac{e}{cm}\right)^2 \frac{\langle\tau^3\rangle\langle\tau\rangle - \langle\tau^2\rangle^2}{\langle\tau\rangle^2}. \tag{3.23}$$

В слабых полях изменение сопротивления в магнитном поле пропорционально квадрату магнитной индукции.

Если бы τ не зависело от энергии, то мы имели бы $\varkappa_1^3 = 0$, а следовательно, изменения сопротивления (в слабых полях) не было бы вовсе. Отметим, однако, что при наличии двух (или больше) различных типов частиц изменение сопротивления уже не равно нулю, даже если τ не зависит от энергии.

Рассмотрим теперь другой предельный случай — очень сильных магнитных полей,

$$(\omega_c\tau)^2 \gg 1. \tag{3.24}$$

При этом, однако, мы должны сделать важное ограничение. Так как заряженные частицы совершают в магнитном поле прецессионное движение, то это движение квантуется (подробнее см. § IV.5). Чтобы используемый нами классический способ описания был применим, необходимо, чтобы кванты энергии $\hbar\omega_c$, связанные с этим движением, были малы по сравнению со средней энергией теплового движения kT. Поэтому под сильными полями мы будем понимать такие поля, которые, с одной стороны, удовлетворяют условию (3.24), но, с другой стороны, подчиняются условию

$$\frac{\hbar\omega_c}{kT} \ll 1. \tag{3.25}$$

(Для сильно вырожденных полупроводников вместо kT входит энергия Ферми, см. ниже.)

В случае сильного поля удобнее рассматривать не изменение электропроводности, а само значение электропроводности. Разлагая теперь ζ_1, ζ_2 и ζ_3 в ряд по степеням малой величины $1/\omega_c^2\tau^2$ и пренебрегая малыми второго порядка и выше, после несложных преобразований получаем

$$\sigma_\infty = \frac{e^2 n}{m}\left\langle\frac{1}{\tau}\right\rangle^{-1}.$$

Электропроводность в очень сильном магнитном поле перестает зависеть от магнитного поля и достигает постоянного значения σ_∞. Отношение предельных значений электропроводности или, соответственно, удельного сопротивления равно

$$\frac{\sigma}{\sigma_\infty} = \frac{\rho_\infty}{\rho} = \left\langle\frac{1}{\tau}\right\rangle\langle\tau\rangle. \tag{3.26}$$

Этот результат, как и все предыдущие, был получен в предположении, что эффективная масса и время релаксации не зависят от направления движения. Однако в действительности эффективная масса может сильно зависеть от направления и является не скаляром, а тензором 2-го ранга (см. гл. III). Выражения для постоянной Холла и для изменения сопротивления в магнитном поле становятся

при этом более сложными. Это обстоятельство не вносит принципиальных изменений в закономерности эффекта Холла, но может иметь решающее значение в явлениях магнетосопротивления. В частности, в этом случае магнитное поле, параллельное электрическому, тоже может привести к изменению сопротивления; отношение ρ_∞/ρ в сильных магнитных полях может оказаться зависящим от магнитной индукции (см. гл. XIII).

§ 4. Смешанная проводимость

Как мы увидим ниже, в полупроводниках встречаются часто случаи, когда необходимо считать, что перенос электрического заряда производится частицами разных типов. Поэтому для дальнейшего необходимо обобщить полученные результаты на случай одновременного присутствия различных носителей заряда.

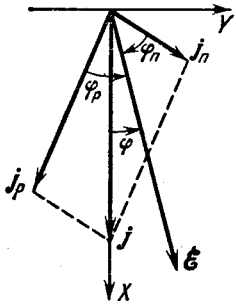


Рис. 1.8. Эффект Холла при смешанной проводимости.

а. Эффект Холла. Качественная картина эффекта Холла при двух разных типах частиц показана на рис. 1.8. Для определенности принято, что частицы одного типа заряжены положительно, а другого—отрицательно. Тогда векторы плотности тока положительных частиц $\mathbf{j}_p$ и отрицательных частиц $\mathbf{j}_n$ будут направлены так, что вектор результирующего электрического поля $\mathcal{E}$ окажется повернутым относительно $\mathbf{j}_p$ на угол Холла φ_p в положительном направлении и на угол φ_n относительно $\mathbf{j}_n$ в отрицательном направлении. Результирующая плотность тока равна $\mathbf{j} = \mathbf{j}_p + \mathbf{j}_n$. Наблюдаемый на опыте результирующий угол Холла φ есть угол, на который повернут вектор $\mathcal{E}$ относительно вектора $\mathbf{j}$. Этот угол может быть как положительным, так и отрицательным, в зависимости от величин и направлений $\mathbf{j}_p$ и $\mathbf{j}_n$, т. е. в зависимости от соотношения между концентрациями и подвижностями частиц обоих типов.

Для вычисления результирующего угла Холла вернемся опять к выражениям для плотности тока, создаваемой частицами сорта «i», которые, учитывая соотношение (1.8), запишем в виде

$$j_x^{(i)} = \sigma_{xx}^{(i)} (\mathcal{E}_x + \mathcal{E}_y \operatorname{tg} \varphi^{(i)}), \qquad j_y^{(i)} = \sigma_{yy}^{(i)} (\mathcal{E}_y - \mathcal{E}_x \operatorname{tg} \varphi^{(i)}). \tag{4.1}$$

Составляющие полного тока, с учетом (3.11), можно записать в виде

$$\begin{aligned} j_x &= \sum_i j_x^{(i)} = \mathcal{E}_x \sum \sigma_{xx}^{(i)} + \mathcal{E}_y \sum \sigma_{xx}^{(i)} \operatorname{tg} \varphi^{(i)}, \\ j_y &= \sum_i j_y^{(i)} = -\mathcal{E}_x \sum \sigma_{xx}^{(i)} \operatorname{tg} \varphi^{(i)} + \mathcal{E}_y \sum \sigma_{xx}^{(i)}. \end{aligned} \tag{4.2}$$

Если, как и раньше, холловские электроды разомкнуты, то $j_y = 0$. Отсюда находим результирующий угол Холла:

$$\operatorname{tg}\varphi = \frac{\mathcal{E}_y}{\mathcal{E}_x} = \frac{\sum \sigma_{xx}^{(i)} \operatorname{tg}\varphi^{(i)}}{\sum \sigma_{xx}^{(i)}}. \tag{4.3}$$

Так как $\varphi^{(i)}$ выражаются формулой (3.14), то отсюда можно найти $\operatorname{tg}\varphi$, если известны концентрации (входящие в σ_{xx}) и холловские подвижности для частиц каждого сорта.

Чтобы найти постоянную Холла, исключим из соотношений (4.2) поле $\mathcal{E}_x$. Тогда получим

$$\mathcal{E}_y = \frac{\sum \sigma_{xx}^{(i)} \operatorname{tg}\varphi^{(i)}}{\left(\sum \sigma_{xx}^{(i)}\right)^2 + \left(\sum \sigma_{xx}^{(i)} \operatorname{tg}\varphi^{(i)}\right)^2} j_x.$$

Сравнивая это с формулой (1.6), находим

$$R\mathcal{B} = \frac{\sum \sigma_{xx}^{(i)} \operatorname{tg}\varphi^{(i)}}{\left(\sum \sigma_{xx}^{(i)}\right)^2 + \left(\sum \sigma_{xx}^{(i)} \operatorname{tg}\varphi^{(i)}\right)^2}. \tag{4.4}$$

Или, вводя в числителе холловские подвижности,

$$R = \frac{1}{c} \frac{\sum \sigma_{xx}^{(i)} \mu_H^{(i)}}{\left(\sum \sigma_{xx}^{(i)}\right)^2 + \left(\sum \sigma_{xx}^{(i)} \operatorname{tg}\varphi^{(i)}\right)^2}. \tag{4.4a}$$

В этой формуле все величины в общем случае зависят от магнитной индукции и поэтому «постоянная» Холла не является постоянной, а тоже зависит от $\mathcal{B}$. Однако в слабом магнитном поле она принимает постоянное значение.

Если имеется только два различных типа частиц и магнитное поле слабое, то

$$\operatorname{tg}\varphi^{(i)} \ll 1, \qquad \sigma_{xx}^{(i)} \simeq en^{(i)}\mu^{(i)}$$

и формула (4.4a) принимает вид

$$R = \frac{1}{ce} \frac{n_1\mu_1\mu_{1H} + n_2\mu_2\mu_{2H}}{(n_1\mu_1 + n_2\mu_2)^2}. \tag{4.5}$$

Если еще считать, что холловский фактор γ одинаков для частиц каждого типа:

$$\gamma = \frac{\mu_{1H}}{\mu_1} = \frac{\mu_{2H}}{\mu_2},$$

то полученный результат можно представить в таком виде:

$$R = \frac{\gamma}{cen_1} \frac{1 + \frac{n_2}{n_1}\left(\frac{\mu_2}{\mu_1}\right)^2}{\left(1 + \frac{n_2}{n_1}\frac{\mu_2}{\mu_1}\right)^2}, \tag{4.5a}$$

где второй множитель в правой части представляет поправку, вносимую частицами типа 2.

В случае очень сильного магнитного поля, удовлетворяющего условиям (3.24) и (3.25), мы имеем $\operatorname{tg}\varphi^{(i)} \gg 1$, и поэтому в формуле (4.4) можно положить

$$\sum \sigma_{xx}^{(i)} \operatorname{tg}\varphi^{(i)} \gg \sum \sigma_{xx}^{(i)}, \qquad \zeta_2^{(i)} \simeq \frac{1}{\omega_{ci}^2},$$

$$\sigma_{xx}^{(i)} \operatorname{tg}\varphi^{(i)} = \frac{e^2 n^{(i)}}{m^{(i)}} \zeta_1^{(i)} \cdot \omega_{ci} \frac{\zeta_2^{(i)}}{\zeta_1^{(i)}} \simeq \frac{cen^{(i)}}{\mathscr{H}}.$$

Тогда получается

$$R_\infty = \frac{1}{ce(n_1 + n_2)}. \tag{4.6}$$

Следовательно, в этом предельном случае постоянная Холла тоже становится не зависящей от магнитного поля. Она не зависит вовсе и от отношения подвижностей γ и определяется только суммарной концентрацией частиц. Поэтому в принципе определение концентрации заряженных частиц с помощью измерения постоянной Холла в сильном магнитном поле имеет большие преимущества. Однако, к сожалению, на опыте не часто удается осуществить условия, в которых одновременно выполняются оба неравенства (3.24) и (3.25).

Рассмотренный случай двух типов частиц, заряженных одинаково, встречается довольно часто. Мы имеем его, например, в германии и кремнии, в которых существуют так называемые «легкие» и «тяжелые» положительные дырки (см. гл. IV). Он реализуется также во многих полупроводниках, содержащих примеси, при низких температурах, когда, кроме электропроводности обычного типа, играет заметную роль еще так называемая проводимость в примесной зоне (см. § IV.7).

Еще более часто мы имеем случай, когда электропроводность осуществляется частицами двух различных типов, заряженными разноименно (отрицательные электроны и положительные дырки, см. ниже). В этом случае мы должны везде приписать углу Холла для отрицательных частиц знак минус. Поэтому, если p и n — концентрации положительных и, соответственно, отрицательных частиц, в формулу (4.6) вместо $(n_1 + n_2)$ войдет $(p - n)$. Выражение для постоянной Холла в слабых полях будет

$$R = \frac{1}{ce} \frac{p\mu_p\mu_{pH} - n\mu_n\mu_{nH}}{(p\mu_p + n\mu_n)^2}, \tag{4.7}$$

где индексы p и n указывают, к частицам какого типа относятся значения подвижностей. Если не требуется большая точность, то можно еще приближенно положить $\mu_{pH} \simeq \mu_p$, $\mu_{nH} \simeq \mu_n$, и тогда

$$R = \frac{1}{ce} \frac{p - nb^2}{(p + nb)^2}, \tag{4.7a}$$

где $b = \mu_n/\mu_p$. В этой форме выражение для R используется наиболее часто. Из (4.7а) видно, что в случае смешанной проводимости постоянная Холла может быть как положительна, так и отрицательна, в зависимости от соотношения между концентрациями частиц обоих типов. В частности, если $p = nb^2$, постоянная Холла обращается в нуль.

б. Магнетосопротивление. Аналогично обобщаются на случай смешанной проводимости и формулы для магнетосопротивления. Полагая в соотношениях (4.2) $j_y = 0$ и исключая из них $\mathcal{E}_y$, находим

$$j_x = \mathcal{E}_x \left\{ \sum \sigma_{xx}^{(i)} + \frac{\left(\sum \sigma_{xx}^{(i)} \operatorname{tg} \varphi^{(i)}\right)^2}{\sum \sigma_{xx}^{(i)}} \right\}.$$

Здесь множитель, стоящий в скобках, есть электропроводность в магнитном поле $\sigma_\perp(\mathscr{B})$. Поэтому, учитывая еще формулу (4.3), получаем

$$\sigma_\perp(\mathscr{B}) = (1 + \operatorname{tg}^2 \varphi) \sum \sigma_{xx}^{(i)}. \tag{4.8}$$

Таким образом, электропроводность в магнитном поле можно выразить через результирующий угол Холла и диагональные члены тензора электропроводности. На исследовании различных частных случаев этого соотношения мы останавливаться не будем.

§ 5. Некоторые экспериментальные результаты

а. Электронная и дырочная проводимость. Опыт показывает, что знак постоянной Холла и термоэдс может быть как отрицательным, так и положительным. Случай R, $\alpha < 0$ соответствует отрицательно заряженным подвижным частицам, определяющим электропроводность, т. е. электронам. Полупроводники такого типа получили название электронных *) или полупроводников n-типа (от negative — отрицательный).

Однако ничуть не реже в полупроводниках, в которых заведомо нет ионной проводимости и в которых, следовательно, ток обусловлен тоже отрицательными электронами, наблюдается R, $\alpha > 0$. Следовательно, в этих случаях эффект Холла соответствует движению положительных частиц. С классической точки зрения эта особенность не имеет объяснения. Она является следствием того, что движение электронов в кристалле подчиняется законам квантовой механики, которые показывают, что при определенных условиях некоторые явления при движении электронов в электрических и магнитных полях происходят так, как если бы вместо отрицатель-

*) Термин «электронный полупроводник» употребляется в двух смыслах. Им, во-первых, обозначают то обстоятельство, что в данном полупроводнике отсутствует ионная проводимость. Во-вторых, он обозначает, что постоянная Холла $R < 0$.

ных электронов имелись положительно заряженные частицы (см. гл. IV). Такие частицы получили название положительных дырок, а полупроводники, в которых R и $\alpha > 0$, называются дырочными или полупроводниками p-типа (от positive — положительный).

Тип электропроводности (электронная или дырочная) зависит от рода примесей, содержащихся в полупроводнике. Так, например, добавляя к германию или кремнию мышьяк, сурьму или любой другой элемент V группы, мы создаем электронную проводимость. Напротив, добавление галлия, бора или любого иного элемента III группы сообщает германию и кремнию дырочную проводимость. Избыток кислорода в закиси меди Cu_2O вызывает дырочную проводимость, а избыток меди — электронную, и т. д. Примеси, сообщающие данному полупроводнику электронную проводимость, называют *донорами*, а примеси, вызывающие дырочную проводимость,— *акцепторами*. Отметим, что один и тот же химический элемент может быть в одних полупроводниках донором, а в других — акцептором.

Таким образом, электропроводность в полупроводниках может осуществляться не только движением отрицательных электронов, но и движением положительных частиц — дырок. В зависимости от количества и рода примесей, а также от температуры соотношение между концентрациями электронов и дырок может быть весьма различным. Частицы, представленные в большинстве, называют основными носителями заряда (электроны в полупроводнике n-типа, дырки в полупроводнике p-типа), а представленные в меньшинстве — неосновными носителями заряда (электроны в полупроводнике p-типа, дырки в полупроводнике n-типа).

б. Собственная и примесная проводимость. Важные результаты дает исследование зависимости постоянной Холла от температуры. Такие типичные зависимости показаны на рис. 1.9. для случая кристаллов германия, в которые были добавлены различные количества мышьяка*). В области высоких температур постоянная Холла очень сильно зависит от температуры. При этом кривые для всех кристаллов при достаточном повышении температуры (тем больше, чем больше концентрация мышьяка) сливаются в одну кривую, которая в координатах $\lg R$ и $1/T$ имеет вид приблизительно прямой линии. Это показывает, что основная часть подвижных носителей заряда в этой области температур возникает не за счет ионизации примесей, а вследствие отщепления электронов от атомов основной решетки, и поэтому данная область обозначается как область *собственной проводимости*.

Как будет видно ниже, в этой области мы имеем и отрицательные электроны, и положительные дырки, и притом в равном количестве (§ II.6). Поэтому, согласно формуле (4.7а) постоянная Холла

*) По данным работы *P. P. Debye, E. M. Conwell,* Phys. Rev. 93, 693 (1954).

здесь равна ($n = p$)

$$R = -\frac{1}{ce n_i}\frac{b-1}{b+1}, \tag{5.1}$$

где через n_i обозначена концентрация электронов (дырок) в области собственной проводимости (*собственная концентрация*) в данном

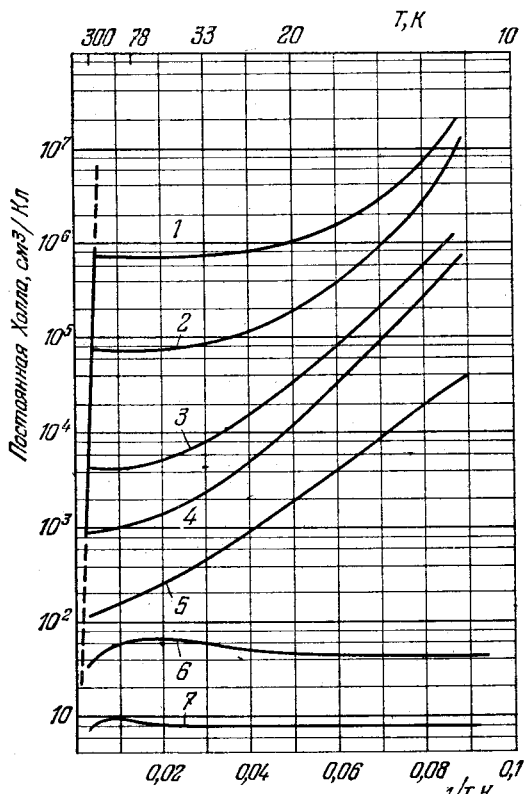


Рис. 1.9. Зависимость постоянной Холла от температуры для кристаллов германия с примесью мышьяка. Концентрация примеси возрастает от кристалла *1* к кристаллу *7*.

полупроводнике. Так как для германия подвижность электронов больше подвижности дырок, то $b > 1$ и постоянная Холла отрицательна.

В области более низких температур кривые зависимости R от $1/T$ существенно зависят от концентрации примеси, причем с увеличением количества примеси R уменьшается, а следовательно, концентрация электронов n увеличивается. В этой области температур подвижные электроны возникают за счет тепловой ионизации

атомов примесей, и мы имеем область *примесной проводимости*. В случае примесной проводимости практически имеется один тип подвижных заряженных частиц (в данном случае — электроны) и постоянная Холла выражается формулой (3.17), причем $R < 0$.

Пользуясь указанными соотношениями, по значениям постоянной Холла можно найти зависимость концентрации электронов проводимости от температуры. Она показана на рис. 1.10 (см. прим. на стр. 36), на котором также отчетливо видны области собственной и примесной проводимости. В области собственной проводимости концентрация n_i чрезвычайно сильно зависит от температуры. Напротив, в примесной области имеется некоторый интервал температур, в котором концентрация электронов практически не зависит от температуры вовсе. Это объясняется тем, что в этой области температур все атомы мышьяка ионизованы, причем каждый атом дает один электрон проводимости. При дальнейшем понижении температуры ионизация атомов примеси становится неполной, и концентрация электронов проводимости снова уменьшается, однако теперь по другому закону, нежели в области собственной проводимости.

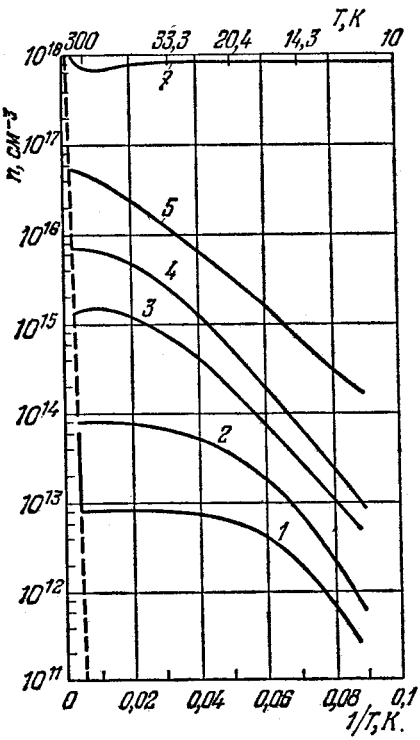


Рис. 1.10. Зависимость концентрации электронов от температуры для кристаллов германия с примесью мышьяка.

Кривые рис. 1.9 и рис. 1.10 также показывают, что при больших концентрациях мышьяка зависимость R и n от температуры ослабляется. При концентрациях, больших $\sim 10^{17}$ см$^{-3}$, они вообще перестают зависеть от температуры и электропроводность приобретает металлический характер. Отсюда следует, что при больших концентрациях примеси энергия ионизации примесных атомов обращается в нуль. Причина этого будет рассмотрена в § IV.7.

Аналогичные зависимости наблюдаются в дырочных полупроводниках, например в германии, содержащем какой-либо акцептор III группы периодической системы.

В § 4 мы отмечали, что в случае смешанной проводимости постоянная Холла может обращаться в нуль при некотором определенном соотношении между концентрациями положительных и отрицательных частиц. Это действительно наблюдается на опыте. На рис. 1.11 показана зависимость постоянной Холла от температуры для сурьмянистого индия InSb *). В области примесной проводимости концентрация электронов мала по сравнению с концентрацией дырок и согласно формуле (4.7а) $R > 0$. Но при повышении температуры n увеличивается и при $nb^2 = p$ постоянная Холла обращается в нуль. При дальнейшем повышении температуры концентрация электронов увеличивается еще больше и постоянная Холла изменяет знак. В кристаллах сурьмянистого индия $b > 1$ и велико (~ 100), и поэтому изменение знака R наблюдается в материале p-типа, и притом особенно отчетливо. Если бы подвижность электронов была меньше подвижности дырок ($b < 1$), то инверсия знака R происходила бы в материале n-типа.

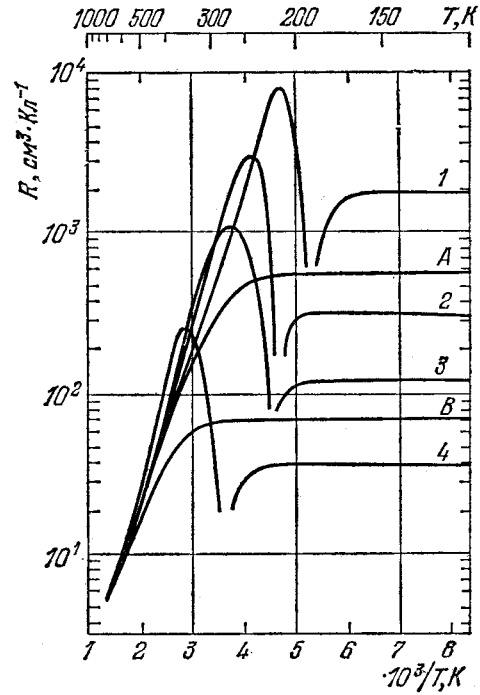


Рис. 1.11. Зависимость постоянной Холла от температуры для кристаллов сурьмянистого индия. *1, 2, 3, 4* — p-тип; *A, B* — n-тип.

Приведенные результаты показывают, что в полупроводниках наблюдается характерная температурная зависимость концентрации подвижных заряженных частиц, отличающая их от металлов. В металлах концентрация электронов проводимости практически не зависит от температуры. Напротив, в типичных полупроводниках всегда имеется такая область температур (область собственной проводимости), где концентрация подвижных частиц n_i существенно зависит от температуры и притом увеличивается при нагревании.

*) Кривые взяты из книги: *У. Данлэп*, Введение в физику полупроводников, ИЛ, 1959.

в. Запрещенная энергетическая зона. Наличие определенной энергии ионизации в идеальном беспримесном полупроводнике можно изобразить с помощью энергетической диаграммы рис. 1.12, где по вертикали снизу вверх отложены значения полной энергии электронов в кристалле. Нижняя заштрихованная полоса, или зона энергий, содержит различные уровни энергии валентных электронов, связанных с решеткой и не участвующих в электропроводности. Наивысшая возможная энергия связанных электронов изображается верхним краем этой «валентной» зоны E_v. Верхняя зона («зона проводимости») содержит различные возможные значения энергии отщепленных электронов, или электронов проводимости, обусловливающих электропроводность. Наинизшее значение их полной энергии изображается нижним краем зоны проводимости E_c. Тогда наименьшая энергия отщепления электрона от решетки, или энергия ионизации, изобразится расстоянием между краями зон $E_g = E_c - E_v$. Промежуточные значения энергии, лежащие между E_c и E_v, не соответствуют никаким возможным стационарным движениям электрона («запрещенная» зона энергий). В металлах энергия ионизации равна нулю, и поэтому $E_g = 0$.

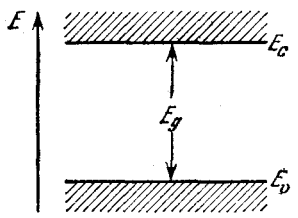


Рис. 1.12. К понятию запрещенной зоны энергии.

Существование энергетических зон, введенных нами здесь только в связи с энергией ионизации, обосновывается теоретически при решении квантовомеханической задачи о движении электронов в периодическом поле кристалла (гл. III). Однако эти представления полезно ввести уже сейчас для истолкования экспериментальных фактов.

Температурная зависимость концентрации электронов в области собственной проводимости n_i позволяет определить ширину запрещенной зоны E_g.

Поясним это сначала простыми, хотя и нестрогими рассуждениями. Предположим, что электроны в кристалле подчиняются статистике Больцмана (что имеет место при достаточно высоких температурах). Обозначим, далее, число возможных состояний электрона в зоне проводимости (пока нам неизвестное) через N_c. Тогда число незанятых состояний в зоне проводимости будет $(N_c - n_i)$, а число вакансий в валентной зоне — n_i. Число электронов проводимости, возникающих в единице объема и в единицу времени вследствие тепловой ионизации, должно быть пропорционально $(N_c - n_i)$ и больцмановскому фактору $\exp(-E_g/kT)$. С другой стороны, частота обратных процессов — переходов электронов проводимости в связанное состояние — должна быть пропорциональна наличной концентрации электронов n_i и числу

вакансий n_i, т. е. n_i^2. Поэтому в состоянии равновесия

$$a(N_c - n_i)\exp\left(-\frac{E_g}{kT}\right) = bn_i^2,$$

где a и b — некоторые коэффициенты пропорциональности. Если $n_i \ll N_c$, то для n_i получается экспоненциальная зависимость от температуры:

$$n_i = C\exp\left(-\frac{E_g}{2kT}\right).$$

Более строгий расчет (гл. V) приводит к несколько измененной формуле:

$$n_i = AT^{3/2}\exp\left(-\frac{E_g}{2kT}\right), \tag{5.2}$$

которая отличается от предыдущего выражения множителем $T^{3/2}$. Однако влияние этого множителя обычно мало́ по сравнению с экспоненциальным множителем.

Соотношение (5.2) позволяет определить ширину запрещенной зоны E_g по экспериментальным данным о зависимости n_i от T в области собственной проводимости. Однако ширина запрещенной зоны сама зависит от температуры и обычно увеличивается при охлаждении. Поэтому возникает вопрос, какое именно ее значение мы определяем из температурных измерений. Для выяснения этого представим ширину запрещенной зоны разложенной в ряд по степеням T:

$$E_g = E_{g0} - \alpha T + \ldots,$$

где E_{g0} — ее ширина при абсолютном нуле. Если в этом разложении ограничиться только линейным членом, то из (5.2) мы определим E_{g0}, так как при подстановке члена αT мы получим постоянный множитель $\exp(-\alpha/2k)$, который будет включен в постоянную A. Если же эту зависимость нельзя считать линейной во всей области температур, то E_g, определенная из температурных измерений, не будет равна фактической ширине запрещенной зоны при $T = 0$. Это поясняет рис. 1.13. Она будет равна E_{gT}, которая получается экстраполяцией к $T = 0$ прямой, приближенно описывающей зависимость E_g от T в исследуемом интервале температур. Величину E_{gT} иногда называют термической шириной запрещенной зоны.

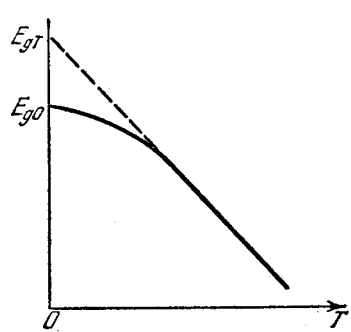


Рис. 1.13. К определению ширины запрещенной зоны.

Фактическую ширину запрещенной зоны (и ее зависимость от температуры) можно определить из оптических измерений, исследуя зависимость коэффициента поглощения света от энергии фотонов. Оптическая ширина запрещенной зоны может несколько отличаться от термической ширины.

Значения ширины запрещенной зоны для некоторых важных полупроводников приведены в таблице 1.1.

Таблица 1.1

Ширина запрещенной зоны E_g (эВ) и подвижности электронов μ_n и дырок μ_p (см²/с · В) при 300 К для некоторых полупроводников (*М. Родо*, Полупроводниковые материалы, «Металлургия», 1971).

Цифры в скобках обозначают температуру, если она отличается от 300 К.

Группы периодической системы	Вещество	E_g	μ_n	μ_p
IV	Алмаз	5,4	1800	1400
IV	Si	1,1	1300	500
IV	Ge	0,67	3800	1820
IV—IV	α-SiC	3,1	220	48
III—V	GaP	2,32 (77)	300	100
III—V	GaAs	1,52 (77)	8800	400
III—V	InP	1,40 (77)	4600	150
III—V	GaSb	0,80 (77)	4000	1400
III—V	InAs	0,43 (77)	33000	460
III—V	InSb	0,22 (77)	78000	750
II—VI	ZnO	3,2	180	—
II—VI	ZnSe	2,80	260	15
II—VI	α-CdS	2,42	295	—
II—VI	β-CdSe	1,85	500	—
IV—VI	PbS	0,39	550	600
IV—VI	PbSe	0,27	1000	900
IV—VI	PbTe	0,33	1600	550

г. Удельная электропроводность. Температурная зависимость удельной электропроводности полупроводников, так же как и постоянная Холла, обнаруживает характерные особенности. На рис. 1.14 (см. прим. на стр. 36) показана зависимость удельного сопротивления от температуры для тех же кристаллов, к которым относится рис. 1.10. И здесь отчетливо можно видеть область собственной проводимости (пунктирная прямая) и, при более низких температурах, область примесной проводимости.

В области собственной проводимости ($n = p = n_i$) удельная электропроводность равна

$$\sigma = en_i(\mu_n + \mu_p). \qquad (5.3)$$

Так как подвижности зависят от температуры сравнительно слабо, то сильное изменение σ и ρ в этой области обусловлено сильной зависимостью n_i от температуры.

В примесной области концентрация дырок p пренебрежимо мала по сравнению с концентрацией электронов n и

$$\sigma \simeq en\mu_n.$$

В этой области σ и ρ зависят от концентрации мышьяка: чем больше концентрация этой примеси, тем больше σ и тем меньше ρ. Однако, в отличие от кривых рис. 1.9 и 1.10, вместо плато в области примесной проводимости здесь наблюдается более сложное, немонотонное изменение ρ. Это объясняется тем, что ρ зависит от произведения концентрации электронов на их подвижность, а подвижность тоже зависит от температуры. При дальнейшем понижении температуры сопротивление монотонно увеличивается. Однако при очень низких температурах (для рассматриваемого примера — ниже ~ 4 К) ρ перестает зависеть от температуры. Причина этого заключается в существовании проводимости по так называемой примесной зоне (§ IV.7).

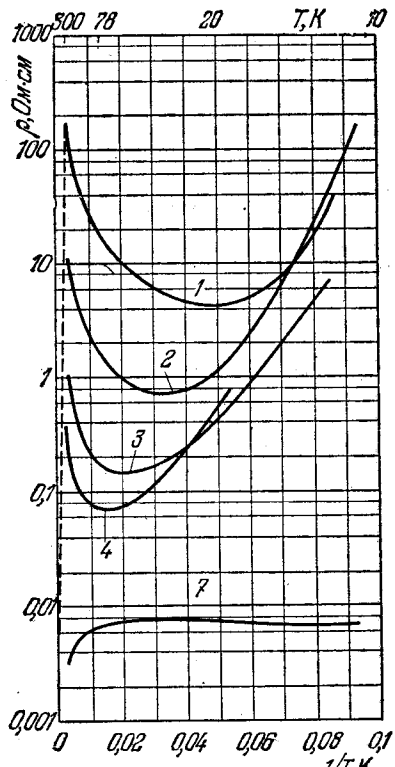


Рис. 1.14. Зависимость удельного сопротивления от температуры в кристаллах германия с примесью мышьяка.

д. Подвижности. В таблице 1.1 даны значения подвижностей электронов и дырок для ряда полупроводников. Они относятся к чистым (по возможности) кристаллам при 300 К. При сопоставлении этих данных следует иметь в виду, что подвижности зависят от чистоты кристаллов и их структурного совершенства. Поэтому для таких кристаллов, которые в настоящее время получают в очень чистом виде (германий, кремний, антимонид индия), подвижности хорошо известны. Для многих же других кристаллов, технология получения которых менее разработана, значения подвижностей характеризуют не только свойства данного полупроводника, но, в известной степени, и качество исследованных кристаллов. В таблице приведены наибольшие измеренные значения.

Из таблицы видно, что подвижности в различных кристаллах весьма различны. Так, в GaP подвижность дырок имеет порядок только 100 см^2/с·В (у некоторых веществ она еще значительно ниже). У сурьмянистого индия подвижность электронов при 300 К достигает $\sim$ 80 000 см^2/с·В, а при понижении температуры может превышать 10^6 см^2/с·В.

Измеряя R и σ при различных температурах, можно найти температурную зависимость холловской подвижности основных носителей. Однако из таких измерений можно при некоторых условиях найти и зависимость дрейфовых подвижностей от температуры (тоже для основных носителей). Последнее можно сделать, например, для того интервала температур в области примесной проводимости, где концентрация электронов (или дырок) не зависит от температуры (плато на рис. 1.9 и 1.10). Тогда температурная зависимость электропроводности дает температурную зависимость дрейфовой подвижности основных носителей.

Опыт показывает, что в полупроводниках с большой подвижностью (например, в германии и кремнии) зависимость μ от T можно удовлетворительно представить в виде

$$\mu = AT^p,$$

где A и p зависят от рода полупроводника, рассматриваемого интервала температур и типа частиц (электроны или дырки). В кристаллах с малой концентрацией примесей и при достаточно высоких температурах $p < 0$, т. е. μ уменьшается при увеличении температуры.

При увеличении концентрации примесей и холловская, и дрейфовая подвижности уменьшаются. При большой концентрации примесей и в области низких температур, в отличие от первого случая, μ увеличивается при повышении температуры, что соответствует $p > 0$. Примеры зависимостей μ от температуры и концентрации примеси будут приведены позже (см. рис. 14.1 и 14.2 на стр. 485, 486).

Общий характер наблюдаемых закономерностей удовлетворительно объясняется теорией рассеяния, которая будет подробнее рассмотрена в связи с явлениями переноса в гл. XIII, XIV. Сейчас же мы ограничимся несколькими общими замечаниями.

Квантовая механика показывает, что в идеальном безграничном кристалле, т. е. в строго периодическом поле, электроны движутся с постоянной энергией и импульсом (см. гл. IV). Они не испытывают «соударений», и поэтому время релаксации τ, а следовательно, и подвижность $\mu = e\tau/m$ равны бесконечности. Конечная подвижность обусловлена процессами рассеяния импульса, возникающими в результате искажений идеальной решетки. Наиболее существенными из них являются тепловое движение и ионизованные примесные атомы, каждое из которых обусловливает некоторые определенные времена релаксации τ_T (тепловые колебания) и τ_I (ионизованные

примеси) и определенные вероятности рассеяния $1/\tau_T$ и $1/\tau_I$. Как будет показано в гл. XIV, результирующая вероятность рассеяния в хорошем приближении равна сумме этих вероятностей, т. е.

$$\frac{1}{\tau} = \frac{1}{\tau_T} + \frac{1}{\tau_I},$$

где τ — результирующее время релаксации, обусловленное этими двумя процессами.

Величины τ_T и τ_I зависят от температуры, и притом различно. С увеличением температуры вероятность решеточного рассеяния увеличивается и поэтому τ_T уменьшается. Напротив, τ_I с повышением температуры увеличивается. Это качественно ясно уже из классических представлений, так как чем больше тепловая скорость частицы, тем меньшие изменения в ее движении будет вызывать электрическое поле ионизованного атома.

В соответствии со сказанным выше, наблюдаемая зависимость $\mu(T)$ в кристаллах, содержащих примеси в значительных концентрациях, в широком интервале температур имеет довольно сложный характер. При низких температурах преобладает примесное рассеяние и μ сначала увеличивается при увеличении температуры. Однако при дальнейшем повышении температуры основную роль начинает играть рассеяние на колебаниях решетки и μ уменьшается при нагревании. При некоторой температуре μ проходит через максимум (ср. рис. 14.1). Температура, соответствующая этому максимуму, тем выше, чем больше концентрация данной примеси.

В полупроводниках с очень малой подвижностью (Se, B, NiO_2, Fe_2O_3, In_2Te_3 и др.) зависимость μ от T имеет другой характер. А именно, μ увеличивается при повышении температуры приблизительно по экспоненциальному закону $\mu \sim \exp(-w/kT)$, где w — некоторая «энергия активации». Такой закон, совместно с малой величиной подвижности, позволяет заключить, что в этих веществах движение носителей заряда имеет скачкообразный характер. Носители заряда находятся преимущественно в связанных состояниях и только время от времени переходят из одного положения в другое. Так как для такого перехода необходимо преодоление некоторого потенциального барьера, то частота скачков увеличивается с возрастанием температуры, и притом по указанному экспоненциальному закону, а w имеет смысл энергии, необходимой для преодоления барьера.

е. Собственная концентрация электронов. Собственная концентрация n_i является одной из важнейших характеристик полупроводника. Она существенно входит во многие теоретические соотношения, и поэтому точное знание этой величины и ее зависимости от температуры необходимо для самых различных расчетов.

О зависимости n_i от температуры уже говорилось выше, она выражается формулой (5.2). Остановимся теперь на величине n_i.

Ее можно найти, например, из постоянной Холла в области собственной проводимости с помощью соотношения (5.1). Однако при этом нужно еще знать хотя и слабую, но все же существующую зависимость $b = \mu_n/\mu_p$ от температуры. Кроме того, в формуле (5.1) предположено, что холловские подвижности равны дрейфовым, и поэтому для точного определения n_i нужно было бы пользоваться точной формулой (4.7) и знать отношение этих подвижностей.

Более точно величину n_i можно найти из электропроводности. Как уже говорилось выше, дрейфовые подвижности μ_n и μ_p и их зависимость от температуры можно найти из данных об электропроводности и постоянной Холла в кристаллах n- и p-типов в примесной области. Если затем проэкстраполировать эти значения в область более высоких температур, то из значения электропроводности в собственной области (5.3) можно определить n_i. В таблице 1.2 в качестве примера приведена величина n_i^2 (которая обычно и входит в теоретические формулы) при 300 К для некоторых важных полупроводников. Там же указаны величина собственного удельного сопротивления ρ_i для 300 К и экспериментальное значение постоянной A в формуле (5.2).

Таблица 1.2

Собственные концентрации n_i и удельные сопротивления ρ_i некоторых полупроводников при 300 К

Вещество	A^2, см$^{-6}\cdot$К$^{-3}$	n_i^2, см$^{-6}$	ρ_i, Ом·см
Ge	$3{,}10\cdot10^{32}$	$5{,}61\cdot10^{26}$	47
Si	$1{,}5\cdot10^{33}$	$1{,}9\cdot10^{20}$	$2{,}3\cdot10^{5}$
InSb	$2{,}0\cdot10^{29}$	$2{,}9\cdot10^{32}$	$4{,}5\cdot10^{-3}$
GaAs	$1{,}2\cdot10^{29}$	$1{,}2\cdot10^{14}$	$6{,}4\cdot10^{7}$

Отметим, что эффект Холла и электропроводность не являются единственными способами определения n_i. В частности, пользуясь соотношениями статистики (гл. V) постоянную A можно выразить через универсальные постоянные и эффективные массы электронов и дырок, и поэтому n_i можно вычислить, если точно известны эффективные массы и ширина запрещенной зоны E_g. При этом оказывается, что для хорошо изученных полупроводников полученные таким образом значения n_i близки к найденным из эффекта Холла и электропроводности.

В заключение еще укажем, что измерения концентрации электронов (или дырок) и удельного сопротивления широко используют для оценки степени чистоты полупроводниковых материалов. Выше мы видели, что в области примесной проводимости даже ничтожные

количества некоторых примесей уже заметно изменяют концентрацию электронов (дырок) и электропроводность. Так, концентрация электронов в очень чистых кристаллах германия при 300 К равна $\sim 10^{13}$ см$^{-3}$. С другой стороны, в области плато на рис. 1.10 каждый атом мышьяка дает один электрон проводимости, и поэтому при концентрации мышьяка $\sim 10^{13}$ см$^{-3}$ уже наблюдается заметное изменение электропроводности. Так как число атомов в см^3 германия $\sim 10^{22}$, то указанная концентрация соответствует одному атому мышьяка приблизительно на 10^9 атомов германия, или 10^{-7} атомного процента. Это имеет место и в других полупроводниках. Так, например, в закиси меди Cu_2O при 300 К содержание кислорода в 0,1 ат.% изменяет электропроводность примерно в 10^5 раз. Эти примеры также поясняют, почему в ряде случаев к чистоте полупроводниковых материалов предъявляют исключительные требования.

Однако при этом следует иметь в виду, что определения только сопротивления, так же как и определения только концентрации носителей заряда, может оказаться недостаточным для суждения о чистоте материала. Если в полупроводнике имеются одновременно донорные и акцепторные примеси и притом приблизительно в одинаковых количествах, то примеси могут компенсировать друг друга. При этом концентрация электронов может быть мала, а удельное сопротивление может даже превышать ρ_i вследствие уменьшения подвижности из-за примесного рассеяния. Поэтому такие измерения всегда желательно дополнять определением подвижности, которая в случае компенсированных примесей будет всегда заниженной вследствие дополнительного примесного рассеяния.

С другой стороны, имеются и такие примеси, которые слабо влияют на электропроводность. Примером может служить кислород в германии и кремнии, содержание которого может достигать сотых долей ат.% без существенного влияния на электропроводность.

Из сказанного также ясно, что влияние примесей на электропроводность зависит еще от области температур. Поэтому один и тот же материал в области собственной проводимости (высокие температуры) может рассматриваться как очень чистый, а в области примесной проводимости (низкие температуры) — как очень грязный.

ж. Магнетосопротивление. В § 3 мы видели, что если имеется один тип носителей заряда и m и τ не зависят от направления движения, то изменение сопротивления в поперечном магнитном поле $\Delta\rho_\perp/\rho$ должно быть в слабых магнитных полях $\sim \mathscr{B}^2$. Такой же результат получается и при учете нескольких типов носителей заряда. Опыт показывает, что это действительно имеет место.

Однако другие выводы простой теории не согласуются с опытом. Выше мы уже говорили, что в изотропной модели продольное магнетосопротивление $\Delta\rho_\parallel/\rho$ должно быть равно нулю, так как при параллельных магнитном поле и скорости дрейфа сила Лоренца

равна нулю. На рис. 1.15 показана зависимость $\Delta\rho/\rho$ от угла θ между направлениями тока и магнитного поля в германии и кремнии. Из рисунка видно, что продольное магнетосопротивление $\Delta\rho_{\parallel}/\rho$ (при $\theta = 0$ и $\theta = 180°$) отнюдь не равно нулю, а имеет тот же порядок величины, что и $\Delta\rho_{\perp}/\rho$. Кроме того, величина $\Delta\rho_{\perp}/\rho$ оказывается значительно больше, чем вычисляемая из изотропной модели. Имеются и другие расхождения простой теории с опытом. Это показывает, что изотропная модель недостаточна для количественного описания магнетосопротивления.

Наблюдаемые закономерности удовлетворительно объясняются полной теорией, учитывающей анизотропию эффективных масс и времен релаксации, наличие носителей заряда разных типов и зависимость времен релаксации от энергии. Частично эти вопросы рассматриваются в гл. XIII.

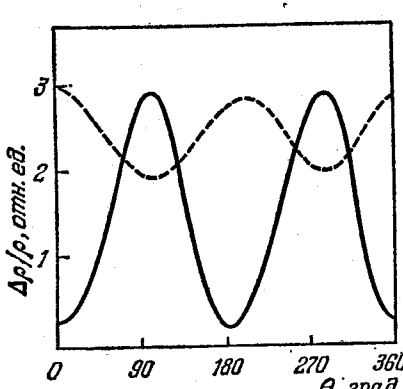


Рис. 1.15. Зависимость магнетосопротивления $\Delta\rho/\rho$ от угла θ между направлениями тока и магнитного поля в германии (пунктир) и кремнии (сплошная кривая) n-типа. Ток направлен по [100], магнитное поле лежит в плоскости (010).

Суммируя рассмотренные экспериментальные данные, мы приходим к следующим важным выводам:

а) Образование электронов проводимости в полупроводниках, в отличие от металлов, требует определенной энергии активации.

б) Электропроводность в полупроводниках может осуществляться не только электронами, но и положительно заряженными частицами — дырками.

в) В полупроводниках существует собственная проводимость, определяемая только свойствами основной решетки кристалла, и примесная проводимость, обусловленная содержащимися в кристалле примесями.

г) При определенных условиях примеси могут очень сильно влиять на электропроводность; при этом различные примеси можно разделить на две группы: вызывающие электронную проводимость (донорные) и дырочную проводимость (акцепторные).

д) Подвижности носителей заряда в различных полупроводниках могут отличаться на много порядков, что указывает на сильное различие эффективных масс и времен релаксации.

е) Время релаксации зависит от энергии носителей заряда. Время релаксации и эффективная масса могут зависеть также от направления движения.

Глава II

ХИМИЧЕСКИЕ СВЯЗИ В ПОЛУПРОВОДНИКАХ

Полная теория твердого тела должна основываться на квантовомеханическом рассмотрении системы из многих электронов и атомных ядер, взаимодействующих между собой. Такой строгий подход связан с необычайными математическими трудностями, и поэтому при практическом решении задачи приходится прибегать к тем или иным радикальным упрощениям. Так поступают, например, в зонной теории твердого тела, которая будет рассмотрена в следующей главе.

Однако, наряду с этим, весьма полезными оказываются также полуфеноменологические способы анализа свойств полупроводников, при которых часть микроскопических характеристик вещества не вычисляется, а заимствуется из опыта. При этом наиболее существенные особенности взаимодействия между электронами и ядрами учитываются автоматически. Один из таких способов, который можно назвать кристаллохимическим, заключается в анализе особенностей химических связей в полупроводниках. Он основан на том, что энергия химической связи атомов внутри кристалла значительно меньше энергии ионизации внутренних оболочек атомов. По этой причине оказывается, что сопоставляя между собой: а) расположение атомов в структуре данного вещества, б) электронную конфигурацию атомов, из которых построен полупроводник, и в) тип химической связи, удается ответить на многие важные вопросы. Таким способом можно объяснить, почему одни вещества оказываются полупроводниками или изоляторами, а другие — металлами, и в известной степени предсказывать наличие полупроводниковых свойств. Можно понять, каким образом изменяются основные полупроводниковые характеристики (ширина запрещенной зоны, подвижность и др.) внутри данного класса аналогичных веществ, а также и ряд других особенностей полупроводников. Такой анализ позволяет также объяснить происхождение полупроводниковых свойств у некоторых некристаллических веществ (аморфные и жидкие полупроводники), что пока гораздо труднее сделать другими способами. Недостаток этого метода заключается в том, что большинство получаемых результатов имеет качественный характер. Однако его достоинством является большая наглядность, и поэтому мы сначала остановимся на этом методе.

§ 1. Кристаллические решетки

В дальнейшем мы будем учитывать, что атомы внутри кристаллов расположены в правильном порядке и поэтому кристаллы обладают свойством *трансляционной симметрии*. Это значит, что для любой кристаллической структуры можно указать три некомпланарных вектора $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, обладающих тем свойством, что любое перемещение структуры на вектор

$$\mathbf{a}_n = \mathbf{a}_1 n_1 + \mathbf{a}_2 n_2 + \mathbf{a}_3 n_3, \tag{1.1}$$

где n_1, n_2, n_3 — любые целые числа, оставляет эту структуру неизменной. Наименьшие значения векторов $\mathbf{a}_1$, $\mathbf{a}_2$ и $\mathbf{a}_3$, в соответствии с принятой терминологией, мы будем называть *основными векторами* кристаллической решетки, а вектор $\mathbf{a}_n$ — *вектором решетки*, обозначая индексом n совокупность целых чисел (n_1, n_2, n_3).

Разумеется, говоря о трансляционной симметрии, мы допускаем известную идеализацию, так как отвлекаемся от теплового движения атомов и не учитываем конечные размеры кристалла. Однако для многих вопросов эта идеализация несущественна.

Если из центра какого-либо атома решетки (узла решетки) отложить основные векторы $\mathbf{a}_1$, $\mathbf{a}_2$ и $\mathbf{a}_3$, мы получим некоторый параллелепипед — *элементарную ячейку* данной решетки. При этом весь кристалл можно представить как последовательное повторение его элементарных ячеек.

Элементарная ячейка может содержать один или несколько атомов. В первом случае мы будем говорить о простой решетке, а во втором — о сложной решетке. Всякую сложную решетку, очевидно, можно рассматривать как несколько простых решеток, смещенных определенным образом друг относительно друга. Совокупность координат $\mathbf{r}_{10}$, $\mathbf{r}_{20}$, ..., $\mathbf{r}_{n0}$ всех атомов (отсчитанных от какого-либо узла решетки), входящих в состав одной элементарной ячейки, образует *базис* данной сложной решетки.

Геометрические свойства различных возможных кристаллических структур подробно исследуются и классифицируются в кристаллографии. В зависимости от соотношения между углами, составленными основными векторами решетки друг с другом, и их длинами различные структуры подразделяются на семь кристаллических систем, или сингоний. В зависимости же от набора элементов симметрии, которым обладает кристалл в целом и его элементарные ячейки, структуры характеризуются различными кристаллическими классами и пространственными группами (см., например, [М4]).

Для дальнейшего существенно иметь в виду, что выбор основных векторов, а следовательно и элементарной ячейки, в известной степени произволен. Рассмотрим двумерную простую прямоугольную решетку с основными векторами $\mathbf{a}_x$ и $\mathbf{a}_y$ (рис. 2.1,*а*). Ее элементарная ячейка есть прямоугольник, вершины которого можно по-

местить в узлы решетки. Так как каждый из узлов одновременно принадлежит четырем соседним ячейкам, то число атомов в ячейке равно $\frac{1}{4} \cdot 4 = 1$. Очевидно, что и любой другой подобный прямоугольник, полученный из первого параллельным его смещением

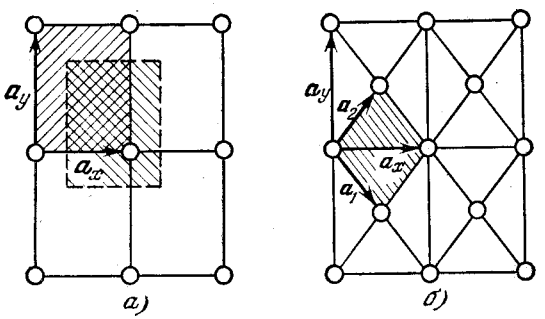


Рис. 2.1. Двумерная прямоугольная решетка: *а*) простая; *б*) центрированная. $\mathbf{a}_1$, $\mathbf{a}_2$ — основные векторы решетки Бравэ.

(рис. 2.1,*а*), будет также элементарной ячейкой. Однако можно изменять и форму элементарной ячейки. Это иллюстрирует рис. 2.1*б*, на котором изображена двумерная прямоугольная центрированная решетка. В качестве ее основных векторов мы по-прежнему можем выбрать взаимно перпендикулярные векторы $\mathbf{a}_x$ и $\mathbf{a}_y$. Тогда ее элементарная ячейка будет прямоугольником, содержащим два атома, а ее базис будет $\mathbf{r}_{10} = 0$ и $\mathbf{r}_{20} = {}^1/_2(\mathbf{a}_x + \mathbf{a}_y)$. Однако в качестве основных можно выбрать также векторы $\mathbf{a}_1$ и $\mathbf{a}_2$, проведенные из любого узла к центрам прилегающих прямоугольников. В этом случае мы получим элементарную ячейку в форме ромба, содержащую лишь один атом.

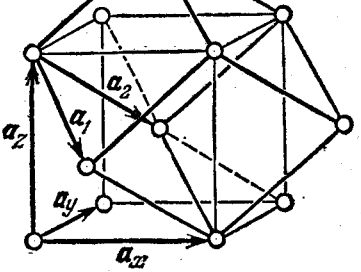


Рис. 2.2. Прямоугольная элементарная ячейка объемноцентрированной кубической решетки (основные векторы $\mathbf{a}_x$, $\mathbf{a}_y$, $\mathbf{a}_z$) и ромбоэдрическая ячейка решетки Бравэ (основные векторы $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$).

То же самое имеет место и для трехмерных структур. На рис. 2.2 показано расположение атомов в кубической центрированной решетке. Если направить основные векторы $\mathbf{a}_x$, $\mathbf{a}_y$, $\mathbf{a}_z$ вдоль ребер куба, мы получим кубическую элементарную ячейку. Так как каждый атом в вершинах принадлежит одновременно восьми соседним

кубам, то число атомов в ячейке равно $\frac{1}{8}\cdot 8+1=2$. Ее базис есть

$$\mathbf{r}_{10}=0, \quad \mathbf{r}_{20}={}^1/_2\mathbf{a}_x+{}^1/_2\mathbf{a}_y+{}^1/_2\mathbf{a}_z,$$

а координаты узлов равны

$$\mathbf{r}_1=\mathbf{a}_x l_x+\mathbf{a}_y l_y+\mathbf{a}_z l_z, \tag{1.2а}$$

$$\mathbf{r}_2=\mathbf{a}_x(l_x+{}^1/_2)+\mathbf{a}_y(l_y+{}^1/_2)+\mathbf{a}_z(l_z+{}^1/_2), \tag{1.2б}$$

где l_x, l_y и l_z — любые целые числа. Такую решетку можно рассматривать как две простые кубические решетки, вдвинутые друг в друга. Но за основные векторы можно также принять и векторы

$$\begin{aligned}\mathbf{a}_1&=-{}^1/_2\mathbf{a}_x+{}^1/_2\mathbf{a}_y+{}^1/_2\mathbf{a}_z,\\ \mathbf{a}_2&={}^1/_2\mathbf{a}_x-{}^1/_2\mathbf{a}_y+{}^1/_2\mathbf{a}_z,\\ \mathbf{a}_3&={}^1/_2\mathbf{a}_x+{}^1/_2\mathbf{a}_y-{}^1/_2\mathbf{a}_z,\end{aligned} \tag{1.3}$$

которые получаются, если из любого узла решетки провести отрезки к центрам прилегающих кубов. Тогда получится ромбоэдрическая элементарная ячейка, показанная на рис. 2.2. Так как атомы в таких ячейках расположены только в их вершинах, все ячейки содержат лишь по одному атому, а решетка, построенная из этих ячеек, оказывается простой. При этом координаты всех узлов будут выражаться формулой

$$\mathbf{r}=\mathbf{a}_1 l_1+\mathbf{a}_2 l_2+\mathbf{a}_3 l_3, \tag{1.4}$$

где l_1, l_2, l_3 — по-прежнему целые числа. Подставляя в формулу (1.4) вместо $\mathbf{a}_1$, $\mathbf{a}_2$ и $\mathbf{a}_3$ их выражения (1.3) и сопоставляя полученное выражение с формулами (1.2), легко видеть, что при $(l_1+l_2+l_3)$ четном мы получаем вершины кубов, а при $(l_1+l_2+l_3)$ нечетном — их центры.

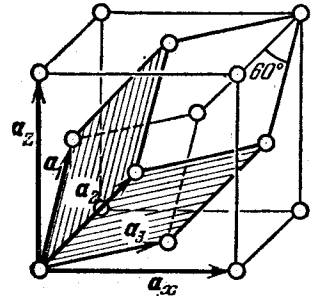


Рис. 2.3. Прямоугольная элементарная ячейка кубической решетки с центрированными гранями и ромбоэдрическая ячейка решетки Бравэ.

Второй важный пример показан на рис. 2.3, где изображена кубическая решетка с центрированными гранями. Выбирая за основные векторы $\mathbf{a}_x$, $\mathbf{a}_y$, $\mathbf{a}_z$, направленные вдоль ребер куба, мы получаем сложную решетку. Так как каждый атом в центрах граней принадлежит двум соседним ячейкам, то число атомов в одной ячейке равно $\frac{1}{8}\cdot 8+\frac{1}{2}\cdot 6=4$ и базис ее есть

$$\mathbf{r}_{10}=0, \quad \mathbf{r}_{20}={}^1/_2(\mathbf{a}_x+\mathbf{a}_y), \quad \mathbf{r}_{30}={}^1/_2(\mathbf{a}_y+\mathbf{a}_z), \quad \mathbf{r}_{40}={}^1/_2(\mathbf{a}_z+\mathbf{a}_x).$$

Рассматриваемую сложную решетку можно представить как смещенные друг относительно друга четыре простые решетки. Однако

она обращается в простую решетку, если в качестве основных выбрать векторы, направленные из любого узла к центрам прилегающих граней:

$$\mathbf{a}_1 = {}^1/_2\,(\mathbf{a}_y + \mathbf{a}_z), \quad \mathbf{a}_2 = {}^1/_2\,(\mathbf{a}_z + \mathbf{a}_x), \quad \mathbf{a}_3 = {}^1/_2\,(\mathbf{a}_x + \mathbf{a}_y). \qquad (1.5)$$

Тогда получается ромбоэдрическая ячейка, показанная на рис. 2.3. Из рисунка видно, что атомы в ячейке расположены только в восьми ее вершинах, так что на одну ячейку приходится один атом. При выборе такой ромбоэдрической ячейки координаты всех узлов гранецентрированной решетки опять выражаются формулой (1.4).

В кристаллографии показывают, что всего существует 14 типов кристаллических решеток, которые могут быть сведены к простым. Они получили название решеток Бравэ.

Из сказанного видно, что число атомов, принадлежащих одной ячейке, зависит от выбора формы (и размеров) ячейки. Отметим уже сейчас, что знание наименьшего количества атомов в одной ячейке существенно для понимания некоторых явлений в твердом теле. В частности, от этого числа зависят возможные типы колебаний решетки, которые будут рассмотрены в гл. XII.

Важной для дальнейшего характеристикой структуры является координационное число k. Оно указывает, сколько ближайших соседей имеет атом в данной структуре. Так, например, в простой кубической решетке каждый атом имеет шесть ближайших соседей и, соответственно, $k = 6$. В объемноцентрированной кубической решетке каждый атом (находящийся, например, в центре кубической элементарной ячейки) имеет восемь ближайших соседей (в вершинах ячейки) и $k = 8$. Для кубической гранецентрированной решетки $k = 6$.

В кристаллах химических соединений, решетки которых содержат неодинаковые атомы, координационные числа для атомов разного типа могут быть различны. С таким примером мы встретимся ниже.

§ 2. Электронная конфигурация атомов

Как известно из атомной физики, состояние электрона в изолированном атоме определяется четырьмя квантовыми числами: главным квантовым числом n, азимутальным квантовым числом l, магнитным квантовым числом m и спиновым квантовым числом s. Главное квантовое число может принимать положительные целые значения $n = 1, 2, 3, \ldots$ Совокупность электронов, характеризуемых одинаковым главным квантовым числом, образует электронную оболочку атома.

Квантовое число l (называемое иногда квантовым числом орбитального импульса) определяет величину момента количества движения электрона и может иметь n различных значений:

$l = 0, 1, 2, \ldots, (n-1)$. В отсутствие внешних магнитных и электрических полей энергия электрона $E = E(n, l)$.

Электроны с одинаковым квантовым числом l образуют группу электронов. Электронная группа есть часть электронной оболочки. В силу исторических традиций, связанных с наименованием оптических спектральных серий (и установленных задолго до развития современной квантовой теории атома), различные электронные группы принято обозначать буквами по следующей схеме:

$$\begin{array}{l} l = 0,\ 1,\ 2,\ 3,\ \ldots \\ \quad\ s,\ p,\ d,\ f,\ \ldots \end{array}$$

В соответствии с этим совокупность электронов, принадлежащих к определенной оболочке и определенной группе, мы будем обозначать указанием главного квантового числа n и буквенного обозначения группы, а индексом сверху указывать фактическое число имеющихся в атоме таких электронов. Так, например, символ $3p^2$ обозначает, что в атоме имеется 2 электрона в квантовых состояниях, характеризующихся квантовыми числами $n = 3$ и $l = 1$.

Третье квантовое число m устанавливает возможные пространственные направления вектора момента количества движения при наличии какого-либо физически выделенного направления и может принимать $(2l + 1)$ значений: $m = -l, -(l-1), \ldots, 0, \ldots, (l-1), l$.

Четвертое квантовое число s обусловлено наличием у самого электрона (независимо от его движения внутри атома) определенного механического момента количества движения («спина») (и связанного с ним магнитного момента). Для собственного механического момента электрона также имеет место пространственное квантование. Однако спиновое квантовое число может принимать только два различных значения: $s = \pm 1/2$.

Согласно принципу Паули в атоме (как и в любой другой системе электронов) в одном и том же квантовом состоянии не может быть более одного электрона. Этот принцип приводит к определенной конечной вместимости электронных групп и электронных оболочек. Согласно сказанному вместимость группы ($l = \text{const}$) равна числу различных возможных значений m, умноженному на 2 (вследствие двух возможных значений s) или $2(2l + 1)$. Это дает для разных групп:

группа:	s	p	d	f	...
l:	0	1	2	3	...
вместимость:	2	6	10	14	...

В каждой заполненной группе всегда имеется четное число электронов, и при этом каждому электрону с определенными квантовыми числами n, l и m всегда соответствует другой электрон с теми же квантовыми числами, но с противоположно направленным спином.

Вместимость оболочек ($n = \text{const}$) равна

$$\sum_{l=0}^{n-1} 2(2l+1) = 2n^2.$$

Это дает:

оболочка:	K	L	M	N	O	P
n:	1	2	3	4	5	6
вместимость:	2	8	18	32	50	72

Напомним, наконец, что заполненные электронные группы обладают особой устойчивостью. Наличие таких заполненных групп характерно для атомов благородных газов. Химическая валентность атомов определяется электронами внешних незаполненных групп (валентными электронами). Атомы металлов характеризуются небольшим числом валентных электронов, которые сравнительно легко могут отщепляться от атома. Так, например, атом типичного металла натрия, имеющий порядковый номер $Z = 11$ и обладающий, следовательно, 11 электронами, имеет электронную структуру:

$$\text{Na}^{11}: \quad (1s)^2\,(2s)^2\,(2p)^6\,(3s)^1 = (\text{Ne})^{10}\,(3s)^1.$$

У атома натрия имеется только один валентный электрон $3s$, и поэтому натрий является одновалентным.

У атомов металлоидов, напротив, одна из электронных групп почти заполнена. Так, например, у атома типичного металлоида хлора электронная структура имеет вид:

$$\text{Cl}^{17}: \quad (\text{Ne})^{10}\,(3s)^2\,(3p)^5.$$

Здесь для заполнения группы $3p$ (вместимость которой 6 электронов) не хватает всего одного электрона. Поэтому атом хлора тоже одновалентен. Однако, в отличие от натрия, он не отдает, а легко присоединяет один электрон, чем обусловлены его металлоидные свойства.

Приведем еще электронные структуры атомов углерода, кремния и германия, особенно интересных в связи с полупроводниковыми свойствами кристаллов этих элементов:

$$\text{C}^{6}: \quad (\text{He})^2\,(2s)^2\,(2p)^2,$$

$$\text{Si}^{14}: \quad (\text{Ne})^{10}\,(3s)^2\,(3p)^2,$$

$$\text{Ge}^{32}: \quad (\text{Ar})^{18}\,(3d)^{10}\,(4s)^2\,(4p)^2.$$

Все эти элементы расположены в одной и той же подгруппе IV B периодической системы и являются четырехвалентными. Их четырьмя валентными электронами являются два s-электрона и два p-электрона.

Отметим, что у атомов этих элементов в незаполненной последней группе имеется только два электрона: $(2p)^2$ у углерода, $(3p)^2$ у кремния и $(4p)^2$ у германия. Поэтому можно было бы ожидать, что эти

элементы будут не четырехвалентными, а двухвалентными. В действительности, однако, эти атомы вступают в химические соединения, находясь не в основном, а в возбужденном состоянии, в котором один из s-электронов переходит в группу p. Так, например, для углерода это соответствует конфигурации $(He)^2\ (2s)^1\ (2p)^3$ и аналогичным конфигурациям для других элементов этой подгруппы. Поэтому указанные четыре s- и p-электрона оказываются принадлежащими к незаполненным группам и являются валентными электронами, способными образовывать так называемую «гибридную s — p-связь».

§ 3. Типы химической связи

Остановимся теперь кратко на основных типах взаимодействия между атомами (подробнее см. например, [1]). Для этого рассмотрим сначала простейший случай двух атомов.

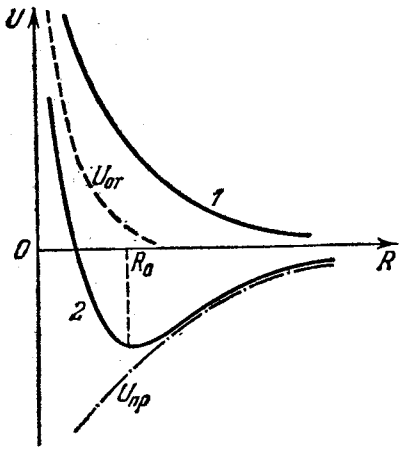


Рис. 2.4. Зависимость потенциальной энергии U двух атомов от расстояния R между их ядрами. Кривая 2 разложена на части, соответствующие силам притяжения (—·—) и отталкивания (— — —).

На рис. 2.4 схематически показана зависимость потенциальной энергии U двух атомов от расстояния R между их ядрами для двух важных типичных случаев. При этом энергия атомов в отсутствие взаимодействия ($R = \infty$) принята за начало отсчета. В случае *1* энергия везде положительна и увеличивается при уменьшении расстояния. Это означает, что между атомами действует сила отталкивания при любом значении R, а значит, образование молекулы невозможно. В случае *2* потенциальная энергия имеет минимум при некотором расстоянии R_0. Здесь возможно образование устойчивой двухатомной молекулы.

Кривую потенциальной энергии типа *2* можно истолковать как результат существования двух сил: притяжения и отталкивания. Полная потенциальная энергия состоит из суммы двух слагаемых: положительного $U_{от}$ (отталкивание), быстро убывающего при увеличении расстояния, и отрицательного $U_{пр}$ (притяжение), уменьшающегося с расстоянием более медленно (рис. 2.4).

Различают следующие основные типы химической связи: а) ионную, или гетерополярную, б) ковалентную, или гомеополярную,

в) ван-дер-ваальсовскую и г) металлическую. В типичных полупроводниках существенную роль играют только два первых типа связи.

а. Ионная связь. В случае ионной связи силы притяжения — кулоновские электростатические силы. Такой тип связи мы имеем в двухатомных молекулах щелочно-галоидных соединений, например в молекуле NaCl. При образовании такой молекулы единственный валентный электрон натрия $3s$ переходит на атом хлора (у которого не хватает как раз одного электрона для завершения группы $3p$), вследствие чего образуются два иона Na^+ и Cl^-, притягивающиеся друг к другу. Отметим, что такой переход электрона происходит потому, что он энергетически выгоден, ибо вследствие притяжения между возникающими ионами Na^+ и Cl^- общая энергия этой системы оказывается меньше, нежели в случае нейтральных атомов Na и Cl.

На больших расстояниях взаимодействие ионов можно рассматривать приближенно как взаимодействие точечных зарядов и считать, что часть энергии взаимодействия, связанная с силами притяжения, равна $-e^2/R$. С другой стороны, при малых R должны проявляться силы отталкивания. Они обусловлены взаимным проникновением друг в друга электронных оболочек атомов, а также отталкиванием одноименно заряженных ядер. Во многих случаях с достаточным приближением можно положить, что величина силы отталкивания обратно пропорциональна некоторой степени расстояния между ядрами. Поэтому полную энергию взаимодействия можно представить в виде интерполяционной формулы:

$$U = \frac{a}{R^m} - \frac{e^2}{R}, \tag{3.1}$$

где a и m — некоторые постоянные, характеризующие силы отталкивания. При $m > 1$ зависимость $U(R)$ изображается кривой типа *2* на рис. 2.4 и имеет минимум при некотором равновесном расстоянии R_0 между ионами.

б. Гомеополярная связь. Такая связь в наиболее чистом виде осуществляется в случае одинаковых, незаряженных (в целом) атомов. Классическим примером является молекула водорода.

Возникновение гомеополярной связи можно пояснить следующим образом. Рассмотрим два атома водорода в основном состоянии и предположим, что сначала оба атома удалены на большое расстояние, так что никакого взаимодействия между ними нет. Из решения задачи об атоме водорода известно, что в этом случае распределение плотности объемного заряда в каждом из атомов обладает сферической симметрией. Оно схематически изображено на рис. 2.5, где прочерчены линии постоянной плотности заряда, густота которых приблизительно пропорциональна плотности заряда. Каждый атом имеет в этом состоянии некоторую энергию E_0, так что полная энергия системы есть $2E_0$.

Представим себе теперь, что расстояние R между ядрами атомов уменьшается. Вследствие взаимодействия между электронами и ядрами, входящими в состав каждого из атомов, изменится и распределение электронной плотности, и полная энергия системы, которая теперь будет $E = 2E_0 + u(R)$. Величина $u(R)$ есть средняя энергия кулоновского взаимодействия зарядов. При классическом рассмотрении задачи она оказывается всегда положительной и увеличивается при уменьшении R, так что мы имеем всегда отталкивание двух атомов водорода. Поэтому факт существования молекулы H_2 в классической теории не находит никакого объяснения. Принципиально другой результат получается в квантовой механике (см., например, [M1]). Благодаря специфическим особенностям поведения тождественных частиц — в данном случае электронов — средняя энергия кулоновского взаимодействия может оказаться и отрицательной. Ее можно представить в виде суммы двух членов, один из которых имеет чисто классический вид (и всегда положителен), второй же имеет различные знаки (и различную величину) в зависимости от относительной ориентации спинов электронов. Этот второй член

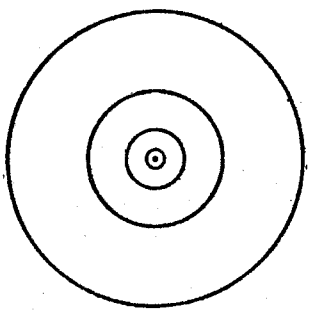


Рис. 2.5. Распределение электронной плотности в основном состоянии атома водорода.

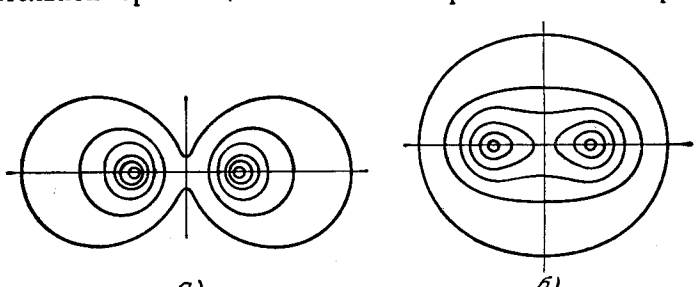


Рис. 2.6. Распределение электронной плотности у двух взаимодействующих атомов водорода: *а*) электронные спины параллельны (отталкивание); *б*) электронные спины антипараллельны (молекула H_2).

называется обменной энергией. Как показывает расчет, если спины электронов параллельны друг другу, то для полной энергии взаимодействия $U(R)$ получается кривая типа *1* на рис. 2.4. Кривые равной плотности заряда в электронном облаке такой системы схематически изображены на рис. 2.6,*а*. В этом случае плотность заряда между атомами имеет минимум и отрицательный заряд локализован возле

каждого из ядер. Энергия взаимодействия всегда положительна, а следовательно, два атома H в таком состоянии отталкиваются.

Если оба спина антипараллельны, то зависимость $U(R)$ изображается кривой типа *2* на рис. 2.4; $U(R)$ имеет минимум, объясняющий возможность образования устойчивой молекулы H_2 при $R = R_0$. Кривые равной плотности заряда для этого случая показаны на рис. 2.6,*б*. Здесь электронные облака обоих атомов как бы сливаются вместе и образуют единое облако объемного заряда, охватывающее оба ядра.

Важным свойством гомеополярной связи является ее насыщаемость. При сближении с молекулой H_2 третьего атома H энергия взаимодействия изменяется только за счет классического взаимодействия, обменная же энергия при этом не возникает. Поэтому третий атом H будет отталкиваться от молекулы H_2.

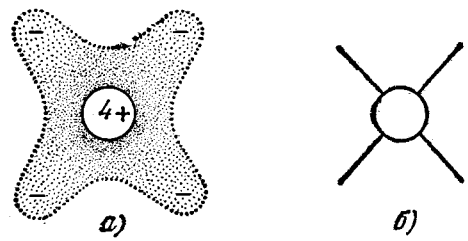


Рис. 2.7. Электронное облако 2*s*- и 2*p*-электронов атома углерода (схематически) и четыре его направленные валентности.

Другая важная особенность гомеополярных связей заключается в их пространственной направленности. Она возникает потому, что результирующее электронное облако, образуемое валентными атомными электронами, вообще говоря, не сферично, а вытянуто вдоль определенных направлений. С другой стороны, для образования гомеополярной связи необходимо перекрытие электронных облаков взаимодействующих атомов. Поэтому связываемые атомы ориентированы определенным образом друг относительно друга, что изображается в химии направленными черточками — валентностями.

Сказанное схематически поясняет рис. 2.7, где в двумерном представлении показано электронное облако атома углерода (*а*). Оно характеризует некоторое результирующее электронное состояние, образованное суперпозицией одного состояния 2*s* и трех состояний 2*p*. Там же изображены соответствующие ему четыре направленные валентности (*б*). В действительности эти четыре валентности лежат не в одной плоскости, а образуют тетраэдрическое расположение, составляя друг с другом углы 109,5°.

Отметим еще, что четкое разграничение между гомеополярной и ионной связями не всегда возможно. Гомеополярную связь в чистом виде мы имеем в молекулах из двух одинаковых атомов, например в молекуле водорода, где распределение электронной плотности симметрично относительно обоих ядер. Если оба атома различны, то и результирующее электронное облако в молекуле будет асимметрично и оба атома нельзя уже считать незаряженными. Ионную

связь можно рассматривать как предельный случай гомеополярной связи, когда электронное облако валентных электронов сосредоточивается вокруг одного из ядер.

в. Ван-дер-ваальсовская связь. Изучение поведения реальных газов показывает, что между нейтральными атомами даже в случае, когда гомеополярные (валентные) силы полностью отсутствуют, возникают силы притяжения, быстро убывающие с расстоянием. Происхождение этих так называемых ван-дер-ваальсовских или, иначе, дисперсионных сил тоже в основном квантовомеханическое.

Ван-дер-ваальсовские силы играют существенную роль в кристаллах благородных газов, а также в так называемых молекулярных кристаллах. Однако в типичных полупроводниковых материалах влияние этих сил ничтожно по сравнению с другими типами химической связи, и поэтому мы на них не будем останавливаться.

§ 4. Строение некоторых полупроводниковых кристаллов

Анализируя различные полупроводниковые материалы, можно установить, что их полупроводниковые свойства тесно связаны с типом химической связи.

Рассмотрим с этой точки зрения некоторые типичные примеры.

а. Ионные кристаллы. В наиболее чистом виде ионная связь имеется в кристаллах щелочно-галоидных соединений. На рис. 2.8 показана кубическая элементарная ячейка решетки NaCl. Ее можно рассматривать как две кубические решетки с центрированными гранями, состоящие из положительных ионов натрия и, соответственно, из отрицательных ионов хлора, вдвинутые друг в друга. Так как каждый ион в вершинах куба принадлежит одновременно восьми соседним ячейкам, а каждый ион в центре грани относится одновременно к двум ячейкам, то каждая элементарная ячейка содержит $\frac{1}{8}\cdot 8+\frac{1}{2}\cdot 6=4$ иона Na^+ и 4 иона Cl^+.

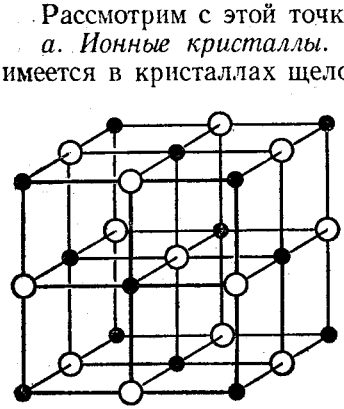


Рис. 2.8. Элементарная ячейка кристалла NaCl.

Каждый положительный ион имеет 6 ближайших отрицательных ионов, и наоборот. Следовательно, в данном случае координационные числа для ионов обоих типов одинаковы: $k_{Na} = k_{Cl} = 6$.

Другим примером кристаллической решетки щелочно-галоидных соединений является решетка CsI. Ее элементарная ячейка есть центрированный куб, в вершинах которого расположены ионы одного типа, а в центре — другого. Число ионов в ячейке равно 2, а координационные числа опять равны друг другу: $k_{Cs} = k_I = 8$.

Силами притяжения, удерживающими ионы в решетке, здесь, как и в молекуле NaCl, являются кулоновские силы. Энергию взаимодействия ионов в кристалле (энергию решетки) можно получить, складывая энергии взаимодействия различных пар ионов. Можно показать, что результат такого суммирования имеет вид

$$U = N\left(\frac{A}{R^m} - \alpha\frac{e^2}{R}\right), \tag{4.1}$$

где R — расстояние между соседними (разноименными) ионами, N — число пар ионов в кристалле, а A — некоторая новая постоянная. Входящий сюда множитель α есть безразмерная величина (постоянная Маделунга), зависящая от типа кристаллической решетки. Число α можно выразить в виде простых рядов. Так, например, для решетки типа NaCl $\alpha = 1{,}748$, а для CsI $\alpha = 1{,}763$. И в этом случае при $m > 1$ потенциальная энергия имеет минимум.

До сих пор мы ничего не говорили о постоянных A и m, характеризующих силы отталкивания. Их можно найти следующим образом. В равновесии

$$\left(\frac{dU}{dR}\right)_{R_0} = 0, \tag{4.2}$$

где R_0 — известное из опыта расстояние между ближайшими ионами в данном кристалле. Поэтому, подставляя (4.1) в (4.2), мы получим уравнение, из которого можно выразить постоянную A через показатель m. Значение же m подбирают таким образом, чтобы вычисление различных физических свойств кристалла с помощью выражения (4.1) приводило к согласию с опытными данными. Для этого, например, можно воспользоваться модулем всестороннего сжатия при низких температурах, который можно непосредственно вычислить из (4.1). Такое сравнение показывает, что m выражается большим числом, лежащим для исследованных ионных кристаллов в пределах $\sim 9 \div 11$. Таким образом, силы отталкивания очень быстро уменьшаются при увеличении расстояния между ионами.

Кристаллы щелочно-галоидных соединений при комнатных температурах являются изоляторами. Валентные s-электроны атомов щелочных металлов, переходя на атомы галоидов, оказываются в кристалле очень прочно связанными с последними, и тепловое движение недостаточно для их отщепления. При повышении температуры в таких кристаллах появляется заметная электропроводность. Однако она имеет не электронную природу, а ионную.

б. Гомеополярные кристаллы. Гомеополярная связь в чистом виде существует в кристаллах химических элементов. В частности, сюда относятся полупроводниковые кристаллы элементов подгруппы IVB — алмаз, кремний, германий. Они имеют одинаковую решетку кубического типа. Элементарная ячейка алмаза в прямоугольных основных векторах показана на рис. 2.9. Она представляет собой куб с центрированными гранями, внутри которого имеются еще

четыре атома. Эти внутренние атомы можно получить смещением внешних атомов ячейки в направлении пространственной диагонали куба на расстояние $^1/_4$ длины диагонали, и поэтому они также образуют гранецентрированную решетку. Иными словами, решетку алмаза можно представить как две гранецентрированные кубические решетки, вдвинутые друг в друга. Так как в одной гранецентрированной ячейке число атомов равно 4, то в кубической элементарной ячейке структуры алмаза содержится 8 атомов.

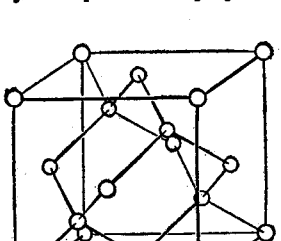


Рис. 2.9. Структура алмаза. Постоянная решетки a равна: алмаз — 3,56Å, кремний — 5,42Å, германий — 5,62Å, серое олово — 6,46 Å.

Напомним, однако, что выбор элементарной ячейки неоднозначен. В § 1 мы видели, что гранецентрированная решетка есть одна из решеток Бравэ и что при выборе ромбоэдрической ячейки в каждой ячейке имеется только один атом. Поэтому наименьшее возможное число атомов в одной ячейке кристаллов со структурой алмаза есть два.

Из сказанного выше ясно, что для образования гомеополярной связи необходимо, чтобы у взаимодействующих атомов были электроны, которые могли бы образовывать пары с антипараллельными спинами. Легко видеть, что это как раз имеет место в кристаллах подгруппы IVB. Из рис. 2.9 видно, что координационное число структуры алмаза равно 4. Тетраэдрическое расположение четырех ближайших соседей каждого атома можно изобразить двумерной схемой рис. 2.10, которая более удобна для дальнейшего. С другой стороны, каждый атом этой подгруппы имеет четыре валентных электрона: $(2s)^2$ и $(2p)^2$ — у алмаза, $(3s)^2$ и $(3p)^2$ — у кремния и т. д. Поэтому имеется как раз такое количество электронов, которое необходимо для образования парно-электронных (гомеополярных) связей между двумя любыми ближайшими атомами.

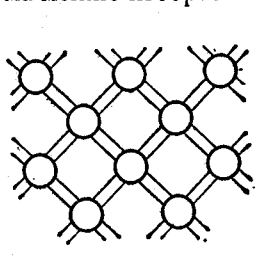


Рис. 2.10. Двумерная схема гомеополярных связей в решетке типа алмаза.

в. Кристаллы со смешанными связями. Между двумя рассмотренными крайними случаями чистой ионной и чистой гомеополярной связи находятся промежуточные случаи смешанной связи, которые имеют место в кристаллах многих химических соединений. Важным примером являются полупроводниковые кристаллы соединений элементов третьей и пятой групп ($A^{III}B^{V}$): InSb, GaAs, GaP и др. Расположение атомов в таких кристаллах, которые, как и алмаз, относятся к кубической системе, показано на рис. 2.11. И здесь

каждый атом имеет четырех ближайших соседей. С другой стороны, атом In (III группа) имеет 3 валентных электрона, а атом Sb (V группа) — 5 валентных электронов, так что всего на каждую пару атомов имеется 8 электронов, что как раз и нужно для образования всех парно-электронных связей. Однако здесь, вследствие различия атомов обоих компонентов, электронные облака оказываются несимметричными. Поэтому атомы In и Sb приобретают отрицательный и, соответственно, положительный заряды и химическая связь оказывается частично ионной и частично гомеополярной. Кристаллы соединений $A^{III}B^{V}$ являются тоже типичными полупроводниками.

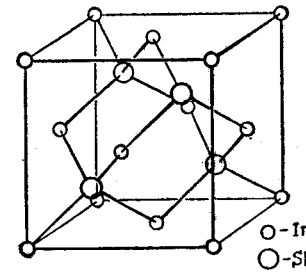


Рис. 2.11. Элементарная ячейка кристалла сурьмянистого индия InSb.

Другим примером полупроводниковых кристаллов со смешанной связью может служить закись меди Cu_2O. Она имеет кубическую решетку, элементарная ячейка которой показана на рис. 2.12. Ячейка содержит 4 атома меди и 2 атома кислорода. Взаимное расположение атомов (число ближайших соседей) в ней можно изобразить плоской схемой рис. 2.13. Каждый атом кислорода непосредственно связан с четырьмя атомами меди, а каждый атом меди — с двумя атомами кислорода. С другой стороны, каждый атом кислорода имеет 6 валентных электронов, а каждый атом меди — 1.

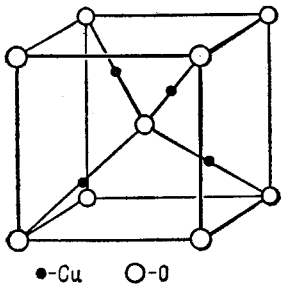


Рис. 2.12. Элементарная ячейка куприта Cu_2O.

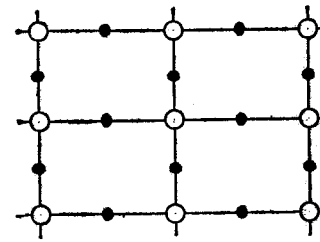


Рис. 2.13. Двумерная схема координации атомов в кристалле Cu_2O: O — кислород, • — медь.

Поэтому на каждую группу Cu_2O приходится 8 валентных электронов, которые как раз и необходимы для образования четырех гомеополярных связей между каждым атомом кислорода и непосредственно с ним связанными четырьмя атомами меди (сравнить

рис. 2.12). Однако и здесь, как и в соединениях $A^{III}B^{V}$, электронные облака несимметричны относительно O и Cu, и поэтому связи являются смешанными.

Приведенные выше данные, которые являются типичными, позволяют сделать некоторые общие выводы. Мы уже отмечали, что чисто ионные кристаллы являются изоляторами. Напротив, во всех других рассмотренных кристаллах, являющихся полупроводниками, существенную роль играла гомеополярная (ковалентная) связь. Это обстоятельство проявляется при анализе и других полупроводниковых кристаллов. Поэтому можно заключить, что полупроводниковые свойства особенно выражены в тех веществах, в которых преобладающим типом связи является ковалентная связь. При этом, как правило, чем сильнее выражена ионная составляющая сил связи, тем ближе данное вещество к изоляторам, и наоборот.

§ 5. Некристаллические полупроводники

Рассмотренные выше примеры полупроводников относились к кристаллам различных веществ. В этих случаях мы имеем, во-первых, правильное расположение ближайших соседей, связанных непосредственно с каждым атомом или ионом в решетке (ближний порядок), и, во-вторых, кристаллическую решетку, т. е. правильное периодическое чередование таких групп в большой области кристалла (дальний порядок). Поэтому возникает вопрос, какое именно из этих обстоятельств существенно для возникновения полупроводниковых свойств. Ответ на этот вопрос дает исследование электрических свойств некристаллических веществ. А именно, опыт показывает, что многие вещества в таком состоянии, когда они заведомо не имеют кристаллической решетки, обладают электронной проводимостью (т. е. доля ионной проводимости в них мала) и притом обнаруживают характерные признаки типичных полупроводников. Эти вещества можно разделить на три группы.

а. Аморфные полупроводники. Примером может служить селен. Он имеет две кристаллические модификации: серый, или гексагональный, селен (который обычно имеют в виду, рассматривая селен как полупроводник) и моноклинный селен, имеющий красный цвет и являющийся изолятором. Однако при быстром охлаждении расплавленного селена можно получить еще аморфную его модификацию, которая только в течение очень длительного времени постепенно переходит в гексагональную модификацию. Исследование электропроводности аморфного селена показывает, что его сопротивление, так же как и у кристаллического селена, уменьшается при повышении температуры, что типично для полупроводников. В аморфном селене наблюдается также изменение сопротивления при освещении (фотопроводимость), что является другим доказательством наличия у него определенной запрещенной зоны энергий.

Аморфное состояние у многих веществ можно получить, конденсируя их в вакууме в виде тонких пленок, если температура подложки не превышает некоторого критического значения, разного для различных веществ. При этом оказывается, что тонкие (несколько сот Å) аморфные пленки некоторых веществ обладают типичными полупроводниковыми свойствами (положительный температурный коэффициент электропроводности, наличие длинноволнового порога поглощения света, фотопроводимость). Полупроводниковые свойства, кроме селена, наблюдаются в аморфном состоянии у сурьмы, германия, кремния, теллура и некоторых химических соединений.

Приведенные данные показывают, что возникновение полупроводниковых свойств, по крайней мере у некоторых веществ, обусловлено ближним порядком структуры. Дальнейшие подтверждения этого заключения мы находим в исследованиях электропроводности расплавов.

б. Жидкие полупроводники. Если при переходе в жидкое состояние ковалентные связи разрушаются, то возникает химическая связь другого типа. Если свойства атомов и средние расстояния между ними таковы, что все валентные электроны находятся в отщепленном состоянии, то возникает металлическая связь и электропроводность сильно увеличивается при плавлении. При этом электропроводность в жидком состоянии, как у металлов, уменьшается при увеличении температуры (вследствие уменьшения подвижности). Однако может быть и другой случай, когда валентные электроны после разрыва валентных связей окажутся локализованными у отдельных атомов. В этом случае электропроводность может резко уменьшиться при плавлении, а ее температурный коэффициент может быть положительным.

Во всех случаях при разрыве насыщенных и направленных валентных связей будет изменяться ближний порядок и координационное число, что можно обнаружить, например, по изменению плотности вещества при плавлении.

Таким образом, при разрушении валентных связей мы должны ожидать во всех случаях изменения электропроводности, плотности и других физических свойств, связанных с ближним порядком. Однако эти изменения могут быть не обязательно резкими и скачкообразными. При температуре плавления скачкообразно разрушается лишь кристаллическая решетка, т. е. исчезает дальний порядок. Расположение же ближайших соседей может и не сделаться вполне хаотическим, т. е. ближний порядок может частично сохраниться и в жидком состоянии, исчезая лишь постепенно при достаточном повышении температуры. Разумеется, при этом координационное число приобретает статистический смысл и может оказаться даже дробным. При сохранении ближнего порядка мы должны ожидать, что вещество и после плавления будет иметь полупроводниковые свойства, т. е. будет жидким полупроводником.

Эти различные случаи действительно наблюдаются на опыте. Примером веществ, у которых ковалентные связи разрушаются сразу же после плавления, могут служить атомные полупроводники подгруппы IVB (структура алмаза) и интерметаллические соединения $A^{III}B^{V}$ (структура цинковой обманки).

В противоположность этому, многие вещества сохраняют полупроводниковые свойства после плавления. Сюда относятся, например, селен и многие оксиды и сульфиды металлов.

Интересным примером таких веществ является теллур. Кристаллы теллура (как и селена) имеют гексагональную решетку, а атомы расположены в ней в виде спиральных цепочек, изображенных на рис. 2.14 (показана половина элементарной ячейки). Одинаково заштрихованные атомы в разных цепочках лежат в одинаковых плоскостях. Каждый третий атом в одной и той же цепочке лежит над другим атомом той же цепочки в направлении оси C. Координационное число равно 2. Каждая пара атомов в одной и той же цепочке (расположенных близко) связана парно-электронными ковалентными силами. Разные же цепочки (расположенные гораздо дальше) связаны силами Ван-дер-Ваальса. Вследствие такого расположения атомов электрические свойства теллура сильно анизотропны. Так, его удельное сопротивление при 20°C для направления вдоль оси C равно $\rho_{\parallel} = 0{,}26$ Ом·см, в то время как $\rho_{\perp} = 0{,}51$ Ом·см. Теллур является типичным полупроводником.

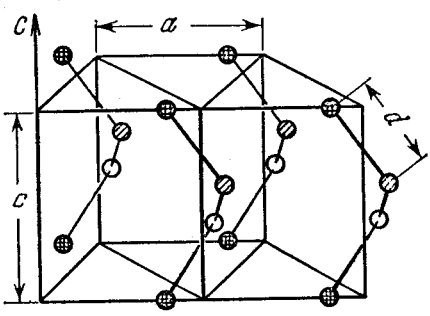


Рис. 2.14. Структура теллура. $a = 4{,}44$ Å, $c = 5{,}92$ Å, $d = 2{,}86$ Å.

Зависимость удельного сопротивления теллура от температуры показана на рис. 2.15 *) (направление тока в твердом состоянии перпендикулярно оси C). Ниже температуры плавления температурный коэффициент сопротивления отрицателен, что, как мы знаем, характерно для полупроводников. Однако и после плавления сопротивление продолжает еще уменьшаться при нагревании в некотором интервале температур, т. е. жидкий теллур ведет себя тоже как полупроводник.

Особенно интересной оказывается температурная зависимость постоянной Холла в теллуре. Она показана на рис. 2.16 *). При низких температурах, в области примесной проводимости, теллур

*) Кривые приведены по данным работы *A. S. Epstein, H. Fritsche, H. Lark-Horovitz*, Phys. Rev. **107**, 412 (1957).

обычно имеет дырочную проводимость и постоянная Холла $R > 0$ (эта область не показана на рисунке). При повышении температуры, когда проводимость теллура становится собственной, постоянная Холла меняет знак. Это показывает, что подвижность электронов в теллуре больше подвижности дырок. При дальнейшем повышении температуры, приблизительно при 230 °С, постоянная Холла второй раз меняет знак, и вплоть до температуры плавления $R > 0$. Такое

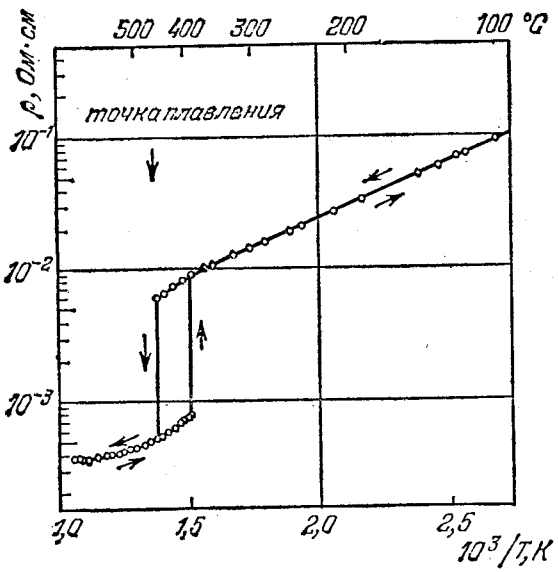


Рис. 2.15. Зависимость удельного сопротивления теллура от температуры. Смещение скачка сопротивления при охлаждении обусловлено переохлаждением расплава.

двукратное изменение знака R в твердом состоянии является результатом сложного строения энергетических зон теллура. Однако мы на этом обстоятельстве не будем останавливаться, так как оно не связано с существом рассматриваемого сейчас вопроса. При плавлении постоянная Холла резко уменьшается и при нагревании расплава продолжает уменьшаться, оставаясь положительной. При ~575 °С постоянная Холла снова обращается в нуль и в третий раз изменяет знак ($R < 0$). Последнее обстоятельство представляет особый интерес. В § I.4 мы видели, что постоянная Холла может обращаться в нуль только в случае смешанной проводимости. Поэтому можно заключить, что в жидком теллуре электропроводность обусловлена и отрицательными электронами, и положительными дырками.

Наблюдаемые явления приводят к следующим заключениям. Полупроводниковые свойства теллура в твердом состоянии обусловлены ковалентными связями, удерживающими атомы в цепочках. При плавлении разрушаются ван-дер-ваальсовские силы, связывающие цепочки друг с другом, ковалентные же связи в пределах отдельных цепочек при этом в значительной степени еще сохраняются, чем и объясняются полупроводниковые свойства жидкого теллура. Вблизи температуры плавления электроны проводимости возникают за счет ионизации цепочек (разрыва некоторых

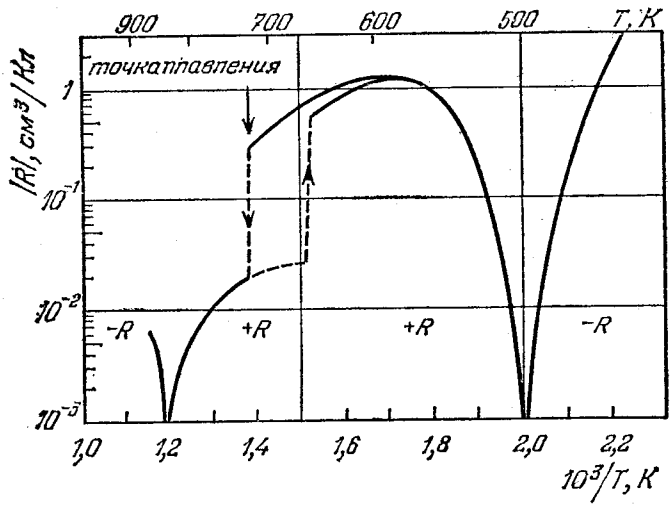


Рис. 2.16. Зависимость постоянной Холла от температуры в твердом и жидком теллуре.

ковалентных связей), а остающиеся незаполненные связи играют роль положительных дырок (ср. § 6). То обстоятельство, что в этой области температур $R > 0$, показывает, что подвижность дырок в жидком теллуре больше подвижности электронов. При дальнейшем повышении температуры начинается разрыв также ковалентных связей, и химическая связь переходит в металлическую. Это приводит к увеличению концентрации электронов, отчего постоянная Холла опять становится отрицательной.

Помимо приведенных примеров, в настоящее время известно большое число других полупроводниковых материалов, сохраняющих в определенном температурном интервале в жидком состоянии полупроводниковые свойства.

в. Стеклообразные полупроводники. К числу полупроводников относится также большая группа стеклообразных химических соединений. Сюда входят двойные соединения P, As, Sb и Bi с одним

из элементов S, Se и Te (халькогениды) и трехкомпонентные соединения этих элементов, такие, как $m\mathrm{As_2Se_3}\cdot n\mathrm{As_2Te_3}$ (где m и n — целые числа), и др. Хотя принципиально стеклообразные полупроводники не отличаются от жидких, так как стекла суть сильно переохлажденные жидкости с очень большой вязкостью, эта группа полупроводников является весьма своеобразной и интересной. В таких веществах наблюдаются: типичный для полупроводников отрицательный температурный коэффициент сопротивления; характерная спектральная зависимость коэффициента поглощения света (наличие пороговой длины волны, зависящей от состава вещества); фотопроводимость и другие полупроводниковые свойства. Подробное изложение вопроса о некристаллических полупроводниках читатель может найти в монографиях [3, 4].

Суммируя сказанное выше, можно заключить, что во многих веществах решающую роль в возникновении полупроводниковых свойств играет ближний порядок. Изложенные представления о важной роли ближнего порядка были высказаны впервые А. Ф. Иоффе и впоследствии развивались многими исследователями.

§ 6. Запрещенная зона энергий

Исходя из особенностей химической связи в полупроводниках, можно непосредственно объяснить существование запрещенной зоны энергий. Рассмотрим в качестве примера атомарные полупроводники подгруппы IVB и будем пока считать, что в них не имеется никаких химических примесей и структурных дефектов. Тогда из рис. 2.10 видно, что при ненарушенных связях в кристалле все валентные электроны каждого атома (два s-электрона и два p-электрона) участвуют в образовании ковалентных связей. Поэтому все валентные электроны являются в известном смысле структурными элементами и находятся в связанном состоянии. В таком состоянии (температура абсолютного нуля и отсутствие внешних ионизующих воздействий) кристалл является изолятором.

Для создания подвижных электронов необходим разрыв некоторого количества связей. Это происходит при повышении температуры и под действием подходящих ионизующих излучений (свет, быстрые электроны и т. п.). При разрыве каждой связи возникает один электрон проводимости и одно вакантное квантовое состояние электрона. Наименьшее приращение энергии электрона при его переходе из связанного состояния в состояние проводимости (работа разрыва связи) есть ширина запрещенной зоны энергий E_g.

Если в полупроводнике имеется электрическое поле, то электроны проводимости будут двигаться против поля, и возникнет электронный ток с некоторой плотностью $\mathbf{j}_n$. Однако, кроме этого процесса, возможен еще и другой механизм электропроводности. А именно, при нарушенных валентных связях (наличие вакантных

мест) какой-либо из электронов связи может перейти в одну из этих вакансий. В результате на месте вакансии будет восстановлена нормальная ковалентная связь, но зато появится вакансия в другом месте. На эту вакансию в свою очередь может перейти какой-либо другой из электронов связи и т. д. При этом, разумеется, возможны переходы электронов связи во всех направлениях. Однако переходы в направлении действующей силы (против поля) будут преимущественными, и поэтому возникнет некоторый дополнительный ток j_p, обусловленный перемещением электронов связи. Сами же вакансии будут перемещаться в направлении поля, т. е. так, как двигались бы положительно заряженные частицы. Этот второй механизм электропроводности и есть процесс дырочной проводимости. Сами же вакантные квантовые состояния электронов, локализованные у нарушенных ковалентных связей и способные перемещаться под действием поля, суть положительные дырки. Из сказанного ясно, что в беспримесном полупроводнике концентрация дырок p всегда равна концентрации электронов проводимости n. В гл. IV будет показано, что именно для дырок (а не для электронов связи) получаются простые уравнения движения, аналогичные уравнениям движения положительно заряженных частиц в классической механике, чем и оправдывается введение этого понятия.

Эти представления о механизме дырочной проводимости, рассмотренные нами качественно на примере полупроводников подгруппы IVB, справедливы для полупроводников любого типа.

§ 7. Полупроводниковые свойства и химическая связь

Из сказанного выше следует, что ширина запрещенной зоны E_g непосредственно зависит от прочности химических связей. К сожалению, эта последняя не может быть прямо измерена количественно. Однако она связана с другими физическими характеристиками вещества, допускающими непосредственное измерение, и поэтому можно ожидать, что между этими характеристиками и величиной E_g для кристаллов с одинаковой структурой должна существовать определенная связь.

В качестве величины, характеризующей прочность связей, некоторые исследователи выбирали теплоту образования кристалла Q. При этом оказывается, что между E_g и Q действительно обнаруживается определенная корелляция: как правило, E_g увеличивается при увеличении Q.

Аналогичные зависимости имеются также для подвижности электронов. Так, например, в ряду веществ с одинаковой кристаллографической структурой подвижность μ_n закономерно изменяется с изменением теплоты образования по закону

$$\mu_n = cQ^{-n},$$

где c и n — постоянные, зависящие от типа кристаллической структуры. На рис. 2.17 показана такая зависимость для ряда двойных соединений со структурой NaCl *). Подобная зависимость имеет место и для двойных соединений с другими структурами.

В кристаллах двойных соединений элементов II и VI групп $A^{II}B^{VI}$, имеющих одинаковую кристаллическую структуру (ZnS, ZnTe, HgTe и др.), наблюдается четкая зависимость E_g от среднего порядкового номера элементов $\bar{Z} = {}^1/_2 (\bar{Z}_A + Z_B)$: E_g уменьшается при увеличении $\bar{Z}$.

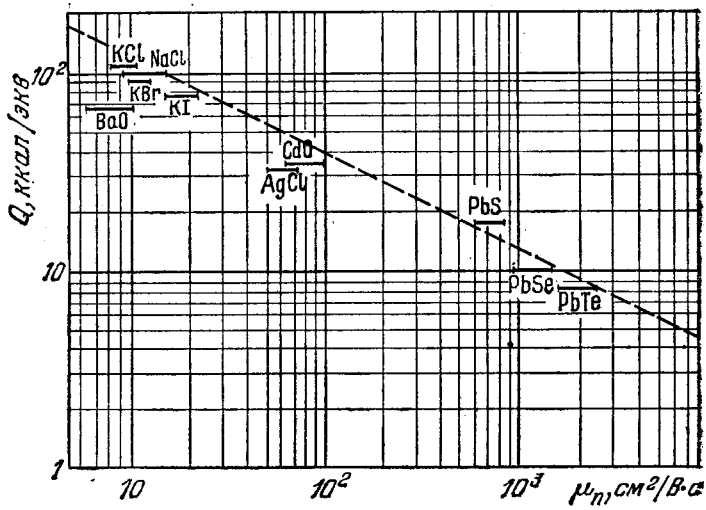


Рис. 2.17. Зависимость подвижности электронов от теплоты образования кристалла для двойных соединений со структурой NaCl.

Существуют также зависимости E_g и μ от других макроскопических величин, отражающих характер химических связей. Хотя такие закономерности носят полуэмпирический характер, они тем не менее представляют большой интерес, так как, во-первых, являются хорошим подтверждением правильности общих представлений о тесной зависимости между типом химической связи и полупроводниковыми свойствами и, во-вторых, могут быть полезны для суждения об электрических свойствах новых полупроводниковых материалов.

§ 8. Полупроводники с малой подвижностью

В различных полупроводниках движение электронов может иметь весьма различный характер. Для разъяснения этого обстоятельства оценим величину средней длины свободного пробега элек-

*) Из работы *В. П. Жузе*, ЖТФ 25, 2079 (1955).

тронов $l = v_T\tau$. Значение τ можно найти из величины подвижности μ (формула (I.3.13)). Полагая для оценки порядка величин $m = m_0 = 9 \cdot 10^{-28}$ г и $v_T \simeq 1 \cdot 10^7$ см $\cdot$ с$^{-1}$ (300 К), находим

$$l \simeq 5 \cdot 10^{-9}\mu \text{ (см)},$$

где μ — в см$^2 \cdot$с$^{-1} \cdot$В$^{-1}$.

Сравним теперь эту величину, во-первых, с длиной волны де Бройля, которая при 300 К ($m = m_0$) равна $\lambda = 7 \cdot 10^{-7}$ см, и, во-вторых, с постоянной решетки кристалла a, которая для большинства кристаллов лежит в пределах $(3—6) \cdot 10^{-8}$ см. Мы видим, что в полупроводниках, в которых $\mu \lesssim 100$ см$^2 \cdot$с$^{-1} \cdot$В$^{-1}$, значения $l < \lambda$, а при подвижности $\mu \lesssim 10$ см$^2 \cdot$с$^{-1} \cdot$В$^{-1}$ величина l становится даже меньше постоянной решетки a.

Приведенные выше оценки показывают, что следует различать два случая. В полупроводниках с большой подвижностью ($\mu \gtrsim 100$ см$^2 \cdot$с$^{-1} \cdot$В$^{-1}$) электроны за время свободного пробега успевают пройти пути, содержащие много постоянных решетки. Это движение с хорошим приближением можно рассматривать как движение в периодическом поле кристалла, которое лишь от времени до времени нарушается процессами рассеяния, скачкообразно изменяющими энергию и импульс электрона. Именно к таким полупроводникам с большой подвижностью (германий, кремний, соединения $A^{III}B^V$ и др.) применима зонная теория, рассматриваемая в гл. III, IV.

Напротив, в полупроводниках с малой подвижностью ($\mu \lesssim 10$ см$^2 \cdot$с$^{-1} \cdot$В$^{-1}$) движение электронов имеет совсем другой характер. В этом случае гораздо правильнее представлять, что электроны преимущественно связаны с определенными атомами решетки и сравнительно редко перескакивают между соседними узлами решетки. Существование такого «прыжкового» механизма электропроводности было впервые указано А. Ф. Иоффе и, независимо, Н. Моттом. К полупроводникам с малой подвижностью зонная теория неприменима.

Впрочем, как показывают оптические и электрические измерения, и в таких полупроводниках существуют зоны разрешенных и запрещенных энергий. Отсюда можно заключить, что представление об энергетических зонах является гораздо более общим, чем это следует из их обоснования в зонной теории (ср. гл. XVII, гл. XIX).

Оба механизма электропроводности характеризуются различной зависимостью подвижности от температуры, о чем уже упоминалось в § I.5. В чистых полупроводниках с большой подвижностью повышение температуры всегда уменьшает величину подвижности. В полупроводниках с малой подвижностью при повышении температуры увеличивается вероятность перескоков электронов, и поэтому подвижность тоже увеличивается.

Отметим также, что прыжковый механизм электропроводности может проявляться при известных условиях и в полупроводниках

с большой подвижностью при очень низких температурах. Так, например, в германии, содержащем доноры и акцепторы (при концентрациях $<10^{15}$ см$^{-3}$, когда еще не проявляется так называемая проводимость в примесной зоне, ср. гл. IV), при гелиевых температурах концентрация подвижных электронов или, соответственно, дырок в зонах ничтожно мала, так как практически все они локализованы на примесных атомах. Однако и в этом случае все же наблюдается очень малая электропроводность, которая, по-видимому, обусловлена прыжковым механизмом.

§ 9. Примесные атомы

Кристалл, в котором все узлы решетки заполнены только атомами данного вещества и решетка которого безгранично простирается во всех направлениях, мы будем называть идеальным кристаллом. Однако все реальные кристаллы имеют различные несовершенства, искажающие строгую периодичность кристаллической решетки.

Исследование особенностей дифракции рентгеновских лучей показывает, что в реальных кристаллах правильное расположение атомов в решетке сохраняется лишь в пределах небольших областей с линейными размерами $\sim 0{,}1 \div 1$ мкм. Сами же эти области слегка разориентированы друг относительно друга (повернуты на малые углы), образуя так называемую мозаичную структуру.

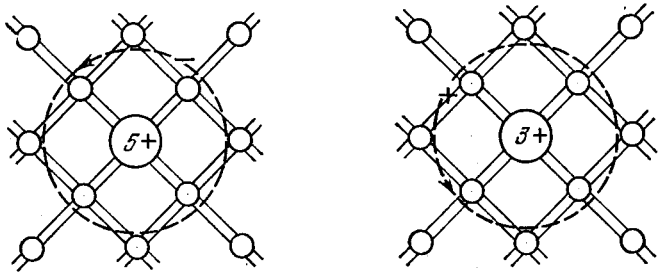
Другим типом несовершенств являются примесные атомы. Последние могут быть в решетке кристалла в двух состояниях: либо занимать узлы решетки, замещая некоторые атомы основного вещества кристалла (твердый раствор замещения), либо внедряться между узлами решетки (твердый раствор внедрения). В обоих случаях каждая примесь характеризуется определенной максимальной растворимостью, т. е. максимальной концентрацией примесных атомов, которую можно создать в решетке при данной температуре при термодинамическом равновесии. Отметим также, что оба типа твердых растворов сильно отличаются по величине коэффициентов диффузии примесей: как правило, при одной и той же температуре коэффициенты диффузии междоузельных атомов на несколько порядков больше, чем коэффициенты диффузии узельных атомов.

Несовершенствами в реальных кристаллах являются также пустые узлы решетки или вакансии, линейные дислокации и винтовые дислокации (см. ниже). Указанные несовершенства могут создавать дополнительные уровни энергии электронов, которые влияют на многие физические процессы в полупроводниках.

Рассмотрим сначала поведение примесных атомов. Наиболее простой и ясный случай мы имеем для примесей элементов III и V групп периодической системы в полупроводниках подгруппы IVB

(кремний, германий). Такие примеси образуют твердый раствор замещения. На это, в частности, указывают очень малые значения их коэффициентов диффузии.

Атом любого элемента V группы имеет 5 валентных электронов, и, соответственно, его ионный остаток несет положительный заряд $+5e$. Однако для образования тетраэдрических валентных связей в решетке типа алмаза необходимо только 4 электрона. Поэтому при замене атома основной решетки на атом примеси мы получим один «лишний» электрон. Последний будет двигаться в поле ионного остатка и остальных валентных электронов, т. е. в поле эффективного заряда $+e$, образуя систему, подобную атому водорода (рис. 2.18, *а*). Этот дополнительный электрон может быть



а) *б)*

Рис. 2.18. Узельный атом элемента V группы (*а*) и III группы (*б*) в решетке типа алмаза.

отщеплен от своего узла решетки под действием тепловых колебаний, освещения и т. п., и тогда он превратится в электрон проводимости. При этом образования свободной положительной дырки (как в случае разрыва связей в собственном полупроводнике) не будет. Из сказанного видно, что примеси элементов V группы в полупроводниках с решеткой алмаза являются донорами.

В случае узельного атома элемента III группы (B, Al и др.) имеется всего 3 валентных электрона, т. е. не хватает одного электрона для завершения валентных связей. Этот недостающий электрон может быть заимствован из числа соседних электронов связи. При этом образуется вакантное место среди соседних электронов связи, или положительная дырка, которая окажется в поле заряда ионного остатка ($+3e$) и четырех электронов связи ($-4e$), т. е. в поле эффективного заряда $-e$ (рис. 2.18, *б*). Так как в образовавшуюся вакансию могут переходить другие электроны связи, то дырка будет двигаться вокруг примесного центра, и мы опять получим водородоподобную систему, но только с неподвижным отрицательным зарядом и подвижным положительным зарядом.

Под влиянием теплового движения, освещения и иных внешних воздействий в образовавшуюся вакансию могут быть заброшены

более удаленные электроны связи. Тогда вместо дырки, связанной с данным примесным центром, появится дырка в другом месте, и эта дырка, вследствие последовательного заброса в нее других электронов связи, будет перемещаться в кристалле. Этот процесс аналогичен отрыву связанного электрона от донора V группы и может быть назван освобождением связанной дырки и переводом ее в свободное состояние. Образования же электрона проводимости при этом не происходит. Из сказанного видно, что атомы элементов III группы в решетке типа алмаза являются акцепторами.

Примесные центры, которые могут отдавать или, соответственно, принимать только один электрон и, следовательно, находиться только в двух различных зарядных состояниях, мы будем в дальнейшем называть простыми. Из сказанного выше следует, что простой донор есть такой примесный центр, с которым при завершенных валентных связях связан один электрон. И, аналогично, простой акцептор есть такой примесный центр, с которым при завершенных валентных связях связана одна дырка.

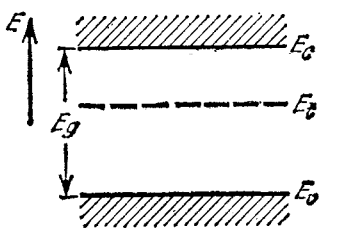


Рис. 2.19. Уровни энергии электронов в кристалле при наличии простых доноров или акцепторов.

Простые доноры можно охарактеризовать заданием наинизшего уровня энергии E_t неотщепленного электрона (основное состояние). Аналогично, для простого акцептора можно задать один уровень энергии захваченного электрона. Такие уровни, однако, в отличие от уровней электронов проводимости, являются локальными, так как электроны, их занимающие, расположены в непосредственной близости от примесных центров (рис. 2.19). Энергия ионизации донора $\mathscr{I}_d$ на этой диаграмме равна $(E_c - E_t)$. Аналогично, энергия, необходимая для заброса электрона из валентной зоны на акцептор (или, другими словами, энергия отрыва связанной дырки от акцептора $\mathscr{I}_a$), равна $(E_t - E_v)$.

Энергию ионизации примесных атомов в кристалле можно экспериментально определить либо из температурной зависимости постоянной Холла («термическая» энергия ионизации), либо из спектральных зависимостей коэффициента поглощения света и фотопроводимости («оптическая» энергия ионизации).

Значения энергии ионизации атомов III и V групп в германии и кремнии, определенные из оптических измерений при гелиевых температурах, приведены в таблице 2.1. Энергии ионизации в германии мало отличаются друг от друга и близки к 0,01 эВ. Этот результат хорошо объясняется в теории «водородоподобных» примесных атомов, которую мы рассмотрим в гл. IV. Вследствие малой энергии ионизации атомы этих элементов в германии практически

Таблица 2.1

Энергия ионизации (в миллиэлектронвольтах) атомов элементов III и V групп в основном состоянии в германии и кремнии *)

Элемент	Ge	Si	Тип	Элемент	Ge	Si	Тип
B	10,6	44,3	Акцепторы, $\mathscr{J}_a = E_t - E_v$	P	12,7	45,3	Доноры, $\mathscr{J}_d = E_c - E_t$
Al	10,9	68,4		As	14,0	53,5	
Ga	11,1	72,3		Sb	10,2	42,5	
In	11,7	155		Bi	12,7	71,0	
Tl	13,1	—					

*) Обзор *Ш. М. Коган, Т. М. Лифшиц*, phys, stat. sol. (a) 39, 11 (1977).

полностью ионизованы, уже начиная с температур $\sim$10K и выше. Энергии ионизации этих элементов в кремнии, особенно акцепторов III группы, отличаются значительно сильнее. Однако и здесь (за исключением In) энергия их ионизации невелика, и поэтому при температурах, близких к комнатным (и выше), атомы этих элементов тоже ионизованы почти полностью. С другой стороны, растворимость большинства элементов III и V групп (кроме Bi и Tl) в германии и кремнии очень велика (например, для In, Ga, P в германии она приближается к 10^{21} атомов/см^3). Поэтому, вводя элементы III и V групп в германий и кремний, можно в широких пределах менять концентрацию электронов и дырок и, соответственно, электропроводность. Подобные примеси, создающие мелкие уровни энергии и способные входить в решетку полупроводника в больших концентрациях, мы в дальнейшем будем называть *легирующими* примесями.

Если примесные атомы принадлежат группе периодической системы, которая отличается больше нежели на единицу от группы основного вещества полупроводника, то система локальных уровней энергии оказывается сложнее. При этом, во-первых, появляются два или несколько разных уровней энергии для одного и того же атома. И, во-вторых, возникают глубокие уровни энергии.

На рис. 2.20 в качестве примера приведены примесные уровни энергии в германии, который изучен наиболее подробно. Остановимся на смысле этой диаграммы.

Рассмотрим, например, медь, которая создает в германии три акцепторных уровня, удаленных на 0,04 и 0,33 эВ от верха валентной зоны и на 0,26 эВ от дна зоны проводимости. Это значит, что атом меди может присоединить к себе три электрона. Присоединение первого электрона (из числа электронов, образующих валентные связи в кристалле) требует наименьшей энергии $E_1 - E_v = 0{,}04$ эВ. При этом атом меди превращается

в отрицательный ион Cu^- и одновременно образуется подвижная положительная дырка. Поэтому мы можем также сказать, что этот электронный переход есть отрыв связанной дырки от акцептора, на что необходима энергия ионизации 0,04 эВ. Для отрыва захваченного электрона от иона Cu^- и перевода его в подвижное состояние необходима наименьшая энергия 0,79—0,04 = 0,75 эВ, где 0,79 эВ есть ширина запрещенной зоны германия (при 0К). Аналогично, энергия $E_2 - E_v = 0{,}33$ эВ на этой диаграмме есть приращение энергии центра при присоединении к иону Cu^- второго электрона (из числа электронов связи) и превращении его в ион Cu^{2-}. При этом среди электронов связи (в валентной зоне) образуется вторая

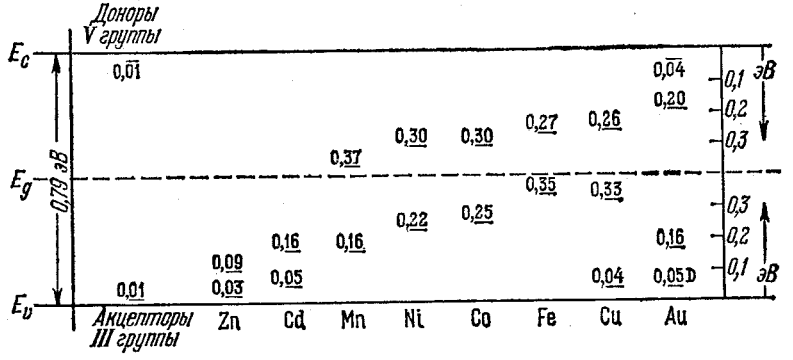


Рис. 2.20. Энергетические уровни некоторых примесей в германии при температуре абсолютного нуля. Все уровни — акцепторные, за исключением наинизшего уровня золота.

подвижная дырка, и мы можем сказать, что энергия 0,33 эВ есть энергия отрыва второй дырки, связанной с ионом Cu^-. Энергия отрыва электрона от иона Cu^{2-} есть 0,79 — 0,33 = 0,46 эВ. И, наконец, энергия присоединения третьего электрона, теперь уже к иону Cu^{2-} (или, что то же, энергия отрыва третьей дырки), равна $E_3 - E_v = 0{,}79 - 0{,}26 = 0{,}53$ эВ. При этом образуется ион Cu^{3-}. Энергия отрыва захваченного третьего электрона есть $E_c - E_3 = 0{,}26$ эВ. Таким образом, узельные атомы меди в германии являются тройными акцепторами и могут существовать в четырех зарядных состояниях: Cu^0, Cu^-, Cu^{2-} и Cu^{3-}.

Число различных локальных уровней энергии для узельных примесных атомов в германии в ряде случаев согласуется с ожидаемым на основании электронной структуры атома и тетраэдрического характера валентных связей в германии. Рассмотрим, например, элементы II группы Zn и Cd. Их атомы имеют по два валентных электрона: $(4s)^2$ и, соответственно, $(5s)^2$. Однако для образования завершенных тетраэдрических связей в германии необходимо четыре электрона. Недостаток двух электронов приводит к образованию

двух дырок, связанных с этими атомами, и поэтому они являются двойными акцепторами.

Элементы Mn (VII группа) и Fe, Co, Ni (VIII группа) все имеют на самой внешней оболочке по два электрона $(4s)^2$. По-видимому, именно эта пара электронов и участвует в образовании валентных связей. Но тогда опять для завершения валентных связей не хватает двух электронов, и атомы этих элементов тоже являются двойными акцепторами.

Элементы I группы Cu, Ag и Au имеют по одному валентному электрону: (4s), (5s) и (6s). Поэтому при узельном положении этих атомов с ними связаны три дырки, и они являются тройными акцепторами. Однако у Au обнаруживается еще и донорный уровень, связанный с возможностью отщепления валентного электрона.

Рассмотрим, наконец, еще пример Te (VI группа). Он имеет шесть валентных электронов $(5s)^2$ $(5p)^4$, в то время как для завершения валентных связей их необходимо только четыре. В соответствии с этим узельные атомы теллура создают в германии два донорных уровня.

§ 10. Вакансии и междоузельные атомы

Нарушения правильной периодичности кристалла вызывают не только атомы примесей, но и собственно структурные дефекты, т. е. различные неправильности в расположении атомов основного вещества кристалла. Такие структурные дефекты, как и атомы примесей, создают в отдельных узлах решетки нарушения правильного чередования химических связей, что может приводить к появлению в запрещенной зоне локальных уровней энергии.

Простейшим типом структурных дефектов являются вакансии (дефекты по Шоттки), которые представляют собой просто пустые узлы решетки. Они схематически показаны на примере решетки соединения типа AB на рис. 2.21, *а*. Здесь большие кружки изображают атомы A, малые кружки — атомы B. Вакансии могут быть как в простой решетке A, так и в простой решетке B. Процесс их образования можно представить как переход сначала одного из приповерхностных атомов на саму поверхность кристалла. В образовавшуюся таким образом приповерхностную вакансию тепловым движением может быть заброшен один из более глубоких атомов, отчего возникнет вакансия, более удаленная от поверхности, и т. д. С помощью такого процесса, который можно рассматривать как диффузию вакансий с поверхности в глубь кристалла, вакансии могут образоваться в любом месте кристалла.

Другой тип дефекта, тоже образованный смещением атомов решетки, получается, если смещенный атом остается в непосредственной близости от своего узла, так что дефект состоит из вакансии

и междоузельного атома (рис. 2.21, *б*). Представление о дефектах такого рода было введено Я. И. Френкелем.

Рассмотренные дефекты существуют во всех кристаллах, даже если они находятся в термодинамическом равновесии. В этом случае дефекты возникают вследствие флуктуаций теплового движения, при которых отдельные узлы решетки могут получить энергию, достаточную для образования дефекта. Концентрация вакансий и дефектов по Френкелю зависит от энергии их образования и температуры и очень быстро увеличивается при повышении температуры.

Большая концентрация вакансий может возникнуть в нестехиометрических кристаллах химических соединений, т. е. в кристаллах, имеющих избыток или недостаток одного из компонентов. Хорошим примером вакансий являются так называемые центры окраски

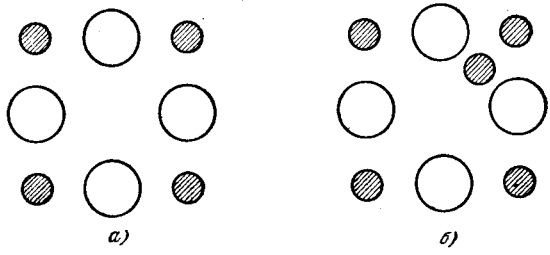


Рис. 2.21. *а*) Вакансия. *б*) Вакансия с междоузельным атомом.

в щелочно-галоидных кристаллах. Опыт показывает, что при нагревании кристаллов LiCl, NaCl, KCl и т. д. в парах своего щелочного металла кристаллы приобретают густую окраску: желтую в случае NaCl, синюю в случае KCl и др. Такую же окраску можно вызвать в щелочно-галоидных кристаллах, облучая их рентгеновскими лучами. Химический анализ показывает, что такие окрашенные кристаллы имеют избыток щелочного металла. Их окраска обусловлена возникновением в спектре поглощения света определенной полосы поглощения, которая различна для разных кристаллов. Эти полосы поглощения были названы Р. В. Полем *F*-полосами, а вызывающие их центры поглощения — *F*-центрами (от немецкого Farbzentren).

Объяснение природы *F*-центров было дано Н. Моттом и Р. Герни [2]. В кристаллах с избытком щелочного металла имеется недостаток галоида, что приводит к возникновению вакансий в решетке анионов. При этом положительный заряд катионов, окружающих данную вакансию, оказывается нескомпенсированным, и в кристалле возникает положительно заряженный центр, способный присоединить к себе избыточный электрон (рис. 2.22). Таким

образом, образуется центр донорного типа. Под действием фотонов достаточно большой энергии этот электрон может быть отщеплен от F-центра. При этом происходит поглощение света и возникает проводимость кристаллов (фотопроводимость). Энергия ионизации F-центра приблизительно равна энергии фотонов, соответствующей середине полосы оптического поглощения.

Опыт показывает, что после длительного поглощения света в F-полосе в окрашенных кристаллах эта полоса ослабевает, но одновременно появляется новая полоса поглощения в более длинных волнах (F'-полоса). Последующее поглощение света в F'-полосе тоже вызывает фотопроводимость. При этом F'-полоса поглощения ослабляется, а первоначальная F-полоса восстанавливается. Это показывает, что, наряду с F-центрами, в щелочно-галоидных кристаллах могут возникать еще центры другого типа (F'-центры). Имеющиеся экспериментальные данные приводят к вероятному заключению, что F'-центры представляют собой вакансию, захватившую два электрона, и что возникновение F'-полосы поглощения и связанной с ней фотопроводимости обусловлено ионизацией второго захваченного электрона.

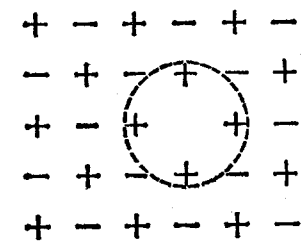


Рис. 2.22. Модель F-центра по Мотту и Герни.

Аналогично этому, имеются указания на то, что вакансия в катионной решетке ионных кристаллов приводит к образованию отрицательно заряженного центра, с которым связана положительная дырка, т. е. к акцепторному центру. Центры такого типа получили название V-центров.

Хорошим примером дефектов по Френкелю могут служить так называемые радиационные дефекты в германии и кремнии. Они возникают при бомбардировке кристаллов быстрыми электронами с энергией порядка сотен килоэлектронвольт и выше, которые при соударении с атомами кристалла смещают их из узельного положения в междоузельное. Опыт показывает, что при этом на образование одного дефекта необходима в среднем энергия около 3,6 эВ в германии и около 4,2 эВ в кремнии.

Исследования эффекта Холла, оптических и фотоэлектрических свойств кристаллов кремния и германия с радиационными дефектами показывают, что эти дефекты создают сложную структуру энергетических уровней. Установлено также, что по крайней мере часть из этих уровней обусловлена не простыми дефектами Френкеля, а более сложными центрами, образованными из дефектов Френкеля и присоединенных к ним атомов кислорода. Однако на этих вопросах мы не будем останавливаться и отсылаем читателя к специальной литературе [5].

§ 11. Дислокации

Рассмотренные выше структурные дефекты — примесные атомы, вакансии и комбинации вакансий с междоузельными атомами — нарушают правильное расположение атомов в пределах небольшого объема, порядка нескольких элементарных ячеек, и поэтому могут быть названы точечными дефектами. Кроме того, в кристаллах могут быть еще линейные структурные дефекты в виде так называемых дислокаций. Они образуются при деформации сдвига в результате скольжения атомных плоскостей друг относительно друга.

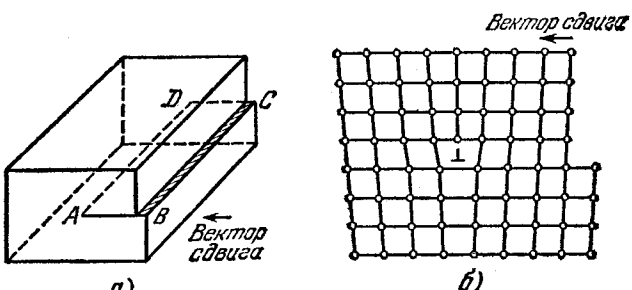


Рис. 2.23. *а*) Образование линейной дислокации при сдвиге. Линейная дислокация *AD* перпендикулярна к вектору сдвига. *б*) Расположение атомов в плоскости, перпендикулярной к линейной дислокации, в простом кубическом кристалле.

Если бы скольжение происходило по всей атомной плоскости кристалла, то все атомы этой плоскости передвинулись бы на одинаковые расстояния в одну или несколько постоянных решетки в направлении сдвига и заняли бы новые положения, идентичные прежним. Однако в действительности скольжение происходит только на части атомной плоскости. Это показано схематически на рис. 2.23, *а*, где *ABCD* есть площадь сдвига. В результате вокруг линии *AD*, отделяющей зону сдвига, образуется характерное искажение решетки. Линия *AD* называется линейной (или краевой) дислокацией и обозначается значком $\perp$. Линейная дислокация всегда перпендикулярна к направлению сдвига.

Расположение атомов в плоскости, перпендикулярной к дислокации, для случая простой кубической решетки показано на рис. 2.23, *б*. В результате сдвига образовалась одна неполная атомная плоскость, край которой и есть линейная дислокация. Из рисунка также видно, что вблизи дислокации решетка упруго деформирована, причем имеются как области сжатия, так и области растяжения.

Эти деформированные области возникают и у поверхности кристалла вблизи точек выхода дислокаций на поверхность. Поэтому скорость растворения кристалла в различных травителях изменяется вблизи этих точек. На этом обстоятельстве основан широко распространенный способ обнаружения дислокаций по «ямкам травления». Разумеется, для получения четкой картины нужны специально подобранные травители, разные для разных кристаллов (а иногда и для разных граней). Дислокационные ямки травления отличаются от других макроскопических дефектов (например, посторонних включений, раковин и т. п.) правильной формой, отражающей симметрию расположения атомов на данной грани.

Вследствие механических напряжений вокруг дислокаций потенциальная энергия примесных атомов вблизи дислокаций меньше, нежели вдали от них. Поэтому легко диффундирующие примеси могут накапливаться вокруг дислокаций, образуя примесную «атмосферу» вокруг них. Пользуясь этим обстоятельством, можно сделать дислокации видимыми. Так, например, можно видеть дислокации в кремнии, вводя в него диффузией медь и получая затем изображение внутренности кристалла в близком инфракрасном свете с помощью микроскопа и электронно-оптического преобразователя. Так как сам кремний хорошо прозрачен для длин волн $\lambda \gtrsim 1$ мкм, то при этом получается четкое изображение дислокаций, сделавшихся непрозрачными вследствие образования примесной атмосферы меди.

Дислокации могут возникать самопроизвольно в процессе роста кристалла. Такие «врожденные» дислокации могут образоваться вследствие деформаций при неравномерном охлаждении кристалла, вследствие неравномерности скорости затвердевания кристалла, а также в результате распространения дислокаций из затравки и других причин.

Если, напротив, в кристалле необходимо создать дислокации, то для этого удобно использовать пластическую деформацию и, в особенности, изгиб кристалла. При простом изгибе плотность дислокаций (т. е. число дислокаций на 1 см^2, перпендикулярный направлению дислокаций) обратно пропорциональна радиусу кривизны нейтральной плоскости и может быть теоретически рассчитана. Кроме того, при изгибе можно получить чисто линейные дислокации, одинаково направленные и параллельные оси изгиба.

Интересный случай упорядоченного расположения дислокаций наблюдается на границе двух сросшихся монокристальных блоков, разориентированных на малый угол. Эти так называемые «границы зерен под малым углом» могут возникать самопроизвольно в процессе кристаллизации или могут быть созданы преднамеренно, если вести кристаллизацию с помощью двух идентичных затравок,

составляющих малый угол. В этом случае соединение двух частей бикристалла осуществляется с помощью ряда параллельных дислокаций, образующих «дислокационную стенку» (рис. 2.24). Из рисунка видно, что среднее расстояние между соседними дислокациями равно $D = b/\theta$, где θ — угол между зернами (в радианах), а b — расстояние между атомами в направлении скольжения. При травлении поверхности кристалла, на которую выходят дислокации, получается упорядоченное расположение ямок травления вдоль определенных линий, указывающее границу двух слабо разориентированных областей кристалла.

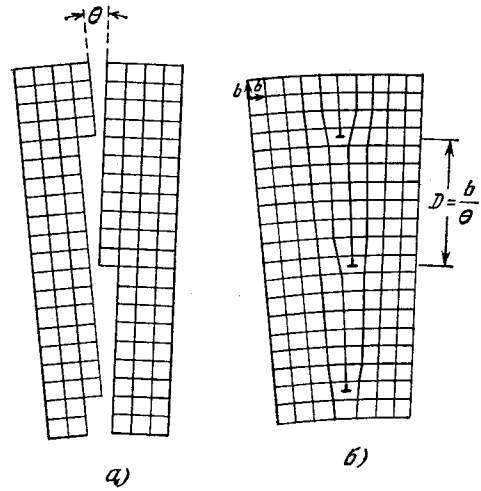


Рис. 2.24. Граница зерен под малым углом (схематически): *а*) расположение атомных плоскостей в обоих зернах; *б*) соединение зерен с помощью ряда параллельных дислокаций.

При термической обработке кристалла, в котором имеются механические напряжения, в нем происходит пластическая деформация. Это осуществляется путем перемещения дислокаций. При этом часто образуется целая система дислокационных стенок, которая превращает кристалл в собрание отдельных зерен, разориентированных на малые углы (явление «полигонизации»).

Если линейная плотность дислокаций $1/D$ на границах зерен становится очень большой, то становится большим и угол разориентации θ. При этом кристалл уже нельзя рассматривать как единый монокристалл с искаженной структурой, и он превращается в поликристаллическое вещество.

Помимо рассмотренных выше линейных (или краевых) дислокаций, существует еще другой тип структурных дефектов, называемых винтовыми дислокациями (см., например, [6]).

Дислокации, подобно примесным атомам и точечным структурным дефектам, создают дополнительные электронные состояния в запрещенной зоне энергий. В § 9 мы видели, что, анализируя особенности химических связей вблизи примесных атомов, можно сделать определенные заключения о характере возникающих локальных уровней. Попытаемся теперь подойти с этой точки зрения к дислокации и рассмотрим, следуя В. Шокли, линейную дислокацию в решетке типа алмаза (например, в германии или кремнии). На рис. 2.25, *а* изображено расположение атомов в решетке такого

типа *). Оно соответствует элементарной ячейке алмаза, изображенной на рис. 2.9, но здесь показано большее число атомов. Как и раньше, линиями показаны направленные валентные связи, которые соединяют каждый атом с четырьмя его ближайшими соседями. На рис. 2.25, *б* показана та же группа атомов при наличии

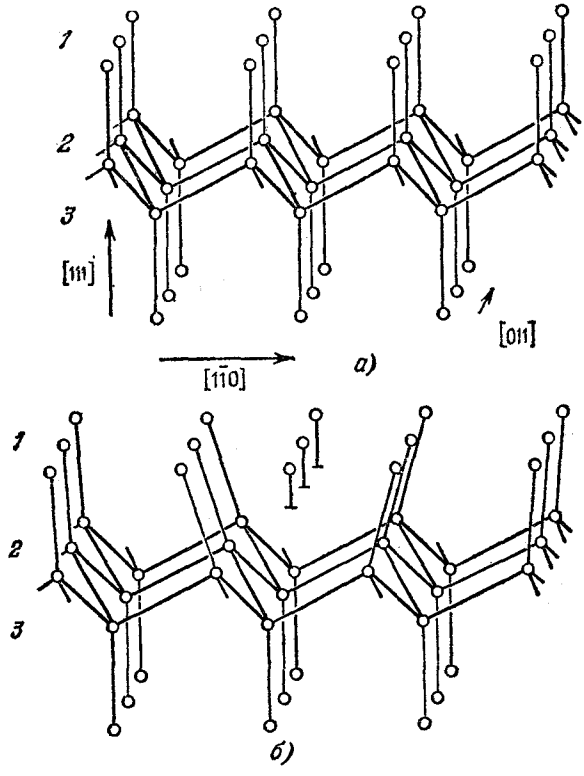


Рис. 2.25. *а*) Расположение атомов в структуре алмаза. *б*) Один из возможных типов линейной дислокации в структуре алмаза (по В. Шокли). Дислокация возникла вследствие частичного сдвига плоскости *1* относительно плоскости *2*. Показаны также три атома на крае неполной атомной плоскости и их неспаренные, или «болтающиеся», связи.

линейной дислокации, возникающей в результате частичного сдвига атомной плоскости *1* относительно плоскости *2*. На рисунке отмечены также основные кристаллографические направления. Сдвиг произошел в плоскости (*111*) в направлении [*1$\bar{1}$0*]. Дислокация

*) *W. T. Read*, Phil. Mag. 45, 775 (1954).

направлена вдоль оси $[01\bar{1}]$. Над плоскостью сдвига имеется ряд атомов (на рисунке показаны три из них), лежащих на крае неполной атомной плоскости, которые не имеют соответствующих соседей снизу. Поэтому каждый из этих атомов имеет по одному неспаренному электрону, которые образуют ненасыщенные, или (следуя терминологии Шокли) «болтающиеся», связи.

Отметим, что указанная конфигурация ненасыщенных связей не является единственно возможной. Так, например, при частичном сдвиге плоскостей *2* и *3* относительно друг друга структура связей оказалась бы другой.

Неспаренные электроны на крае неполной атомной плоскости и приводят к появлению дополнительных состояний. При этом, в принципе, следует учитывать две возможности. Во-первых, на ненасыщенную связь может быть захвачен второй электрон, отчего возникнет парно-электронная связь. Если энергия этого электрона окажется меньше, чем в зоне проводимости, но больше, чем в валентной зоне, то в запрещенной зоне будет разрешенное состояние для электронов. Дислокация окажется способной присоединять к себе дополнительный электрон и будет проявлять себя подобно акцепторам. Во-вторых, ненасыщенная связь может быть разорвана, и тогда неспаренный электрон превратится в электрон проводимости. Если энергия неспаренного электрона имеет промежуточное значение между энергиями электронов в зоне проводимости и в валентной зоне, то дислокация будет подобна донорам с энергетическими уровнями внутри запрещенной зоны.

Таким образом, линейные дислокации по электрическим свойствам имеют сходство с примесными центрами. Однако между ними имеются и существенные различия. В случае примесных центров локальные уровни в среднем распределены приблизительно изотропно в каждой макроскопической области кристалла. В случае же дислокаций распределение отдельных «центров» крайне анизотропно. Среднее расстояние между ними очень мало вдоль дислокаций (расстояние между соседними болтающимися связями). Поэтому в отсутствие примесных атмосфер дислокации следует представлять скорее как одномерные кристаллы, нежели как неравномерно распределенные точечные центры. Соответственно вместо локальных уровней энергии точечных дефектов (рис. 2.19) в этом случае правильнее говорить об энергетических зонах, создаваемых дислокациями в запрещенной зоне («дислокационные зоны»). При наличии примесных атмосфер вблизи дислокаций могут возникать и дискретные локальные уровни.

Пользуясь изложенными представлениями, можно приближенно рассматривать линейную дислокацию как заряженную линию. Если проявляются акцепторные свойства дислокации, то в материале *n*-типа она будет заряжена отрицательно. При этом, вслед-

ствие отталкивания от нее электронов проводимости, дислокация будет окружена цилиндрическим слоем индуцированного положительного заряда.

Из этой модели, далее, следует, что дислокации, как и примесные центры, должны вызывать дополнительное рассеяние электронов и, следовательно, уменьшать их подвижность. При этом можно ожидать, что если ток течет параллельно дислокациям, то их влияние будет мало, так как в этом случае основная часть электронов будет двигаться вне цилиндрических областей объемного заряда, за пределами которых электрическое поле заряженных дислокаций равно нулю. Если же ток течет перпендикулярно дислокациям, то траектории электронов должны существенно изменяться и подвижность электронов должна уменьшаться сильнее.

Влияние линейных дислокаций на электрические свойства полупроводников действительно наблюдается на опыте. Так, например, создание дислокаций в германии *n*-типа приводит к уменьшению концентрации электронов проводимости. Это показывает, что в германии дислокации создают локальные состояния акцепторного типа. При этом уменьшается и подвижность электронов.

Дислокации могут действовать также как эффективные центры рекомбинации (см. гл. IX) и существенно изменять кинетику электронных процессов.

Глава III

ЭЛЕМЕНТЫ ЗОННОЙ ТЕОРИИ ТВЕРДОГО ТЕЛА. I. ИДЕАЛЬНАЯ РЕШЕТКА

§ 1. Основные предположения

Как известно из квантовой механики, для теоретического исследования любой системы частиц, в частности для вычисления возможных значений ее энергии, надо решить соответствующее уравнение Шредингера. Последнее представляет собой дифференциальное уравнение в частных производных, содержащее столько переменных, сколько степеней свободы имеет рассматриваемая система. В физике твердого тела в эту систему входят, строго говоря, все электроны и атомные ядра атомов, составляющих кристалл. Таким образом, число степеней свободы, а с ним и число переменных в уравнении Шредингера, оказывается макроскопически большим — порядка $10^{22} \div 10^{23}$. В результате взаимодействия между частицами переменные не разделяются, и мы приходим к математической задаче исключительной трудности. Прямое решение ее в настоящее время невозможно. Более того, даже если бы решение поставленной задачи удалось найти, физическая интерпретация его представила бы, видимо, не меньше трудностей, чем сам процесс решения, ибо объем информации, таким путем полученной, был бы необычайно велик.

По этим причинам современная квантовая теория твердого тела вынуждена основываться на ряде упрощений. Последние выбираются с таким расчетом, чтобы сохранить лишь наиболее характерные черты системы, исключив все, сравнительно мало существенное. Зонная теория (смысл названия будет ясен из дальнейшего) представляет собой, по-видимому, самую простую квантовомеханическую схему, учитывающую наиболее важные особенности движения электронов во многих кристаллах. Как мы увидим, она позволяет успешно интерпретировать ряд экспериментальных данных.

Зонная теория базируется на следующих основных предположениях, составляющих в своей совокупности так называемое «зонное приближение»:

1) При изучении движения электронов атомные ядра, ввиду их большой массы, рассматриваются как неподвижные источники поля, действующего на электроны.

2) Расположение ядер в пространстве считается точно периодическим: они размещаются в узлах идеальной решетки данного кристалла.

3) Взаимодействие электронов друг с другом заменяется некоторым эффективным внешним полем. Иначе говоря, система электронов, взаимодействующих с атомными ядрами и друг с другом по закону Кулона, заменяется системой независимых электронов, движущихся в некотором заданном поле. Последнее складывается из поля атомных ядер и эффективного поля, приближенно описывающего взаимодействие между электронами.

Первое предположение позволяет рассматривать поведение электронов, не интересуясь движением тяжелых частиц. Эта возможность не является самоочевидной, ибо в результате взаимодействия между электронами и ядрами движения их не независимы: строго говоря, расположение ядер не задано, а изменяется с изменением состояния электронов. Смысл предположения 1) заключается в утверждении, что последний эффект мал.

Второе предположение ограничивает класс рассматриваемых систем: речь идет о кристаллических твердых телах, а не о жидкостях, стеклах и т. п.

Наконец, третье предположение сводит многоэлектронную задачу к одноэлектронной. Вместо одного уравнения Шредингера для всей системы атомных ядер и электронов мы получаем теперь совокупность идентичных, не связанных между собой уравнений Шредингера для каждого электрона в отдельности. Тот факт, что рассматриваемая система состоит из многих частиц, найдет свое явное отражение лишь позднее, когда мы будем интересоваться статистическим распределением электронов по состояниям (гл. V). Иначе говоря, вместо электронной жидкости — системы взаимодействующих друг с другом частиц — мы рассматриваем идеальный электронный газ в эффективном внешнем поле. Разумеется, такая трактовка может быть только приближенной и качество аппроксимации зависит от выбора эффективного поля. Метод подбора и явное вычисление его, равно как и общий вопрос об обосновании и установлении пределов применимости сделанных предположений, рассматриваются в гл. XVII. Там показывается, что ряд основных результатов зонной теории действительно удается обосновать с довольно общих позиций. Вместе с тем есть и важные задачи, выходящие за рамки зонного приближения. Они также обсуждаются в гл. XVII.

В настоящей главе мы примем предположения 1)—3) в качестве постулатов.

Заметим, что в большинстве электрических, магнитных и оптических явлений в твердых телах электроны внутренних атомных оболочек не играют активной роли. Действительно, энергия связи этих электронов со «своими» ядрами — порядка нескольких десятков или даже сотен электронвольт (на электрон). Это значительно

больше средней энергии взаимодействия их со многими внешними полями, равно как и энергии квантов электромагнитного поля в видимой и более длинноволновых областях. Поэтому во многих задачах оказывается возможным другое (также приближенное) разделение частиц на тяжелые и легкие. Именно, в «систему электронов», рассматриваемых явно, можно включить только валентные электроны атомов, составляющих решетку; электроны же внутренних оболочек вместе с ядрами образуют атомные остовы, состояния которых практически не изменяются в рассматриваемых явлениях. При этом роль неподвижных источников поля играют уже не ядра, а атомные остовы. Соответственно предположения 1) и 2) надо переформулировать, заменив в них слова «атомные ядра» на «атомные остовы».

§ 2. Волновая функция электрона в периодическом поле

Как мы видели в предыдущем параграфе, в рамках зонного приближения квантовомеханическая задача о системе электронов в твердом теле сводится к задаче об одном электроне, движущемся в заданном внешнем поле. Обозначим потенциальную энергию электрона в нем через $U(\mathbf{r})$, где $\mathbf{r}(x, y, z)$ — радиус-вектор данной точки пространства. Явный вид функции $U(\mathbf{r})$ нам пока неизвестен. В дальнейшем (гл. XVII) выяснится, что даже приближенное вычисление $U(\mathbf{r})$ связано с большими математическими трудностями. Однако многие важные особенности рассматриваемой системы можно выяснить, не задавая явного вида $U(\mathbf{r})$, а пользуясь лишь некоторыми общими свойствами этой функции. Именно этим, по существу, и объясняется успех зонного приближения при интерпретации экспериментальных данных.

Для выяснения свойств функции $U(\mathbf{r})$ заметим, прежде всего, что наша система зарядов в целом нейтральна: полный заряд валентных электронов равен по величине и противоположен по знаку заряду всех атомных остовов. Далее, следует ожидать, что не только остовы, но и электроны будут расположены в пространстве в среднем периодически. (В дальнейшем это будет показано явно.) Следовательно, и поле, создаваемое данной системой зарядов, должно быть периодично в пространстве:

$$U(\mathbf{r}) = U(\mathbf{r} + \mathbf{a}_n), \tag{2.1}$$

т. е. потенциальная энергия электрона в кристалле инвариантна относительно сдвига на вектор решетки $\mathbf{a}_n$. Итак, мы пришли к задаче о движении электрона в периодическом поле.

В этой главе будут рассматриваться только стационарные состояния электронов. Соответственно уравнение Шредингера имеет вид

$$H\psi = E\psi. \tag{2.2}$$

Здесь H — оператор энергии (гамильтониан) рассматриваемой системы, E — его собственные значения, а ψ есть волновая функция — собственная функция оператора H, принадлежащая собственному значению E. В рассматриваемой нами задаче оператор H представляет собой сумму операторов потенциальной энергии электрона U и кинетической энергии $\hat{\mathbf{p}}^2/2m_0$. Здесь m_0 есть масса свободного электрона, а $\hat{\mathbf{p}}$ — оператор импульса:

$$\hat{p}_x = -i\hbar\frac{\partial}{\partial x},\quad \hat{p}_y = -i\hbar\frac{\partial}{\partial y},\quad \hat{p}_z = -i\hbar\frac{\partial}{\partial z}, \tag{2.3}$$

т. е.

$$\hat{\mathbf{p}} = -i\hbar\nabla. \tag{2.3'}$$

Через $\hbar$, как обычно, обозначена постоянная Планка h, деленная на 2π.

В соответствии со сказанным уравнение (2.2) применительно к задаче об электроне в периодическом поле принимает вид

$$-\frac{\hbar^2}{2m_0}\nabla^2\psi + U\psi = E\psi, \tag{2.4}$$

причем функция $U(\mathbf{r})$ обладает свойством (2.1).

Коль скоро потенциальная энергия $U(\mathbf{r})$ не зависит от спина электрона, каждым собственному значению энергии E и координатной волновой функции ψ отвечают два состояния, соответствующие двум возможным ориентациям спина. Об этом говорят как о спиновом вырождении.

При наложении магнитного поля $\mathfrak{B}$ спиновое вырождение снимается, ибо энергия электрона становится зависящей от проекции магнитного момента на направление $\mathfrak{B}$. В отсутствие магнитного поля спиновое вырождение играет роль только при подсчете общего числа дозволенных квантовых состояний, приводя к появлению дополнительного множителя 2.

Наличие двух спиновых состояний можно отразить, рассматривая функцию ψ как матрицу:

$$\psi = \begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}. \tag{2.5}$$

Функции ψ_1 и ψ_2 отвечают двум разным значениям z-компоненты спина. В рассматриваемом нами случае $\psi_1 = \psi_2 = \psi$. Представление (2.5), однако, остается в силе и в более общем случае, когда принимается во внимание и спиновая зависимость энергии электрона. Эта зависимость возникает как во внешнем магнитном поле, так и в магнитном поле, обусловленном орбитальным движением самого электрона. Для учета последнего эффекта в гамильтониан электрона вводят оператор энергии спин-орбитального взаимодействия H_{so}. Явный вид его выводится в релятивистской квантовой механике [М2, М3]. Оказывается, что

$$H_{so} = \frac{\hbar}{4m_0^2c^2}(\boldsymbol{\sigma}[\nabla U \times \hat{\mathbf{p}}]). \tag{2.6}$$

Здесь c — скорость света в вакууме, $\boldsymbol{\sigma}$ — спиновый (матричный) вектор Паули.

В условиях (2.1) правая часть (2.6) периодически зависит от пространственных координат. Соответственно под U в уравнении (2.4) можно формально понимать и сумму потенциальной энергии и оператора (2.6). По этой причине учет спин-орбитального взаимодействия не влияет на последующие качественные выводы этого параграфа.

Определив из уравнения (2.4) волновую функцию ψ, мы можем найти средние значения любых физических величин, характеризующих поведение электрона. В частности, выражение

$$|\psi(\mathbf{r})|^2 d\mathbf{r} \tag{2.7}$$

дает нам вероятность обнаружить электрон в элементе объема $d\mathbf{r} = dx\,dy\,dz$ около точки $\mathbf{r}$. По смыслу понятия вероятности должно выполняться условие нормировки

$$\int |\psi(\mathbf{r})|^2 d\mathbf{r} = 1, \tag{2.8}$$

где интеграл берется по всему объему системы.

Условие сходимости этого интеграла играет роль одного из дополнительных условий к уравнению (2.4): физически допустимы только те решения, для которых интеграл (2.8) сходится.

Произведем в уравнении (2.4) замену аргумента

$$\mathbf{r} = \mathbf{r}' + \mathbf{a}_n.$$

Принимая во внимание равенство (2.1) и опуская штрих у переменной $\mathbf{r}'$, получаем

$$-\frac{\hbar^2}{2m_0}\nabla^2\psi(\mathbf{r}+\mathbf{a}_n) + U(\mathbf{r})\psi(\mathbf{r}+\mathbf{a}_n) = E\psi(\mathbf{r}+\mathbf{a}_n). \tag{2.4'}$$

Сравнивая это с (2.4), видим, что функции $\psi(\mathbf{r})$ и $\psi(\mathbf{r}+\mathbf{a}_n)$ удовлетворяют одному и тому же уравнению Шредингера с одним и тем же собственным значением энергии E. Если этому собственному значению принадлежит одна собственная функция (т. е. если это собственное значение не вырождено), то функции $\psi(\mathbf{r})$ и $\psi(\mathbf{r}+\mathbf{a}_n)$ могут отличаться лишь постоянным множителем:

$$\psi(\mathbf{r}+\mathbf{a}_n) = c_n\psi(\mathbf{r}). \tag{2.9}$$

Поскольку обе они должны быть нормированы, абсолютная величина c_n должна быть равна единице:

$$|c_n| = 1. \tag{2.10}$$

Таким образом,

$$|\psi(\mathbf{r}+\mathbf{a}_n)|^2 = |\psi(\mathbf{r})|^2 \tag{2.11}$$

— электрон с одинаковой вероятностью может быть обнаружен в элементе объема $d\mathbf{r}$ как около точки $\mathbf{r}$, так и около любой эквивалентной ей точки $\mathbf{r}+\mathbf{a}_n$. Иначе говоря, как мы и ожидали, среднее (в квантовомеханическом смысле) распределение электронов в решетке обладает пространственной периодичностью последней.

Если данному собственному значению энергии принадлежит несколько собственных функций (т. е. если оно вырождено), то рассуждения несколько усложняются: в правой части равенства (2.9) следует, вообще говоря, написать линейную комбинацию всех этих функций. Однако выбор собственных функций, принадлежащих данному вырожденному собственному значению E, не однозначен: любая линейная их комбинация тоже есть собственная функция, принадлежащая тому же собственному значению. В частности, можно выбрать собственные функции так, чтобы равенство (2.9) выполнялось для каждой из них в отдельности (см. Приложение I).

Легко найти явный вид зависимости c_n от вектора $\mathbf{a}_n$. Для этой цели произведем в уравнении (2.4) два последовательных сдвига аргумента — на $\mathbf{a}_n$ и $\mathbf{a}_{n'}$, где вектор $\mathbf{a}_{n'}$ отличается от $\mathbf{a}_n$ заменой чисел n_1, n_2, n_3 на n'_1, n'_2, n'_3. Применяя дважды соотношение (2.9), получим

$$\psi(\mathbf{r}+\mathbf{a}_n+\mathbf{a}_{n'}) = c_n c_{n'} \psi(\mathbf{r}). \tag{2.12'}$$

С другой стороны, по определению

$$\mathbf{a}_n + \mathbf{a}_{n'} = \mathbf{a}_{n+n'},$$

и, следовательно,

$$\psi(\mathbf{r}+\mathbf{a}_n+\mathbf{a}_{n'}) \equiv \psi(\mathbf{r}+\mathbf{a}_{n+n'}) = c_{n+n'}\psi(\mathbf{r}). \tag{2.12''}$$

Сравнивая равенства (2.12') и (2.12''), находим

$$c_n c_{n'} = c_{n+n'}. \tag{2.13}$$

Прямой подстановкой легко убедиться, что это функциональное уравнение имеет решение

$$c_n = e^{i\mathbf{k}\mathbf{a}_n}, \tag{2.14}$$

где $\mathbf{k}$ — произвольный вектор. В силу равенства (2.10) компоненты $\mathbf{k}$ должны быть вещественными.

Равенство (2.14) с сочетании с (2.11) позволяет в известной мере раскрыть вид функций $\psi(\mathbf{r})$, удовлетворяющих уравнению Шредингера с периодическим потенциалом $U(\mathbf{r})$. Именно, легко убедиться, что

$$\psi(\mathbf{r}) = e^{i\mathbf{k}\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}), \tag{2.15}$$

где $u_{\mathbf{k}}(\mathbf{r})$ — функция, периодическая с периодом решетки:

$$u_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r}+\mathbf{a}_n). \tag{2.16}$$

Действительно, в силу (2.11) $\psi(\mathbf{r})$ может отличаться от функции, периодической с периодом решетки, только фазовым множителем вида $e^{if(\mathbf{r})}$, где f — вещественная функция. В силу (2.14) она должна быть линейной.

Равенства (2.15), (2.16) составляют содержание *теоремы Блоха*: волновая функция электрона, движущегося в периодическом поле,

представляет собой модулированную плоскую волну. Иначе говоря, это есть произведение экспоненциальной функции $e^{i\mathbf{kr}}$ на функцию, периодическую с периодом решетки. Сами функции вида (2.15), (2.16) иногда называют *функциями Блоха* *).

Вектор $\mathbf{k}$ называют *квазиволновым*. Очевидно, его компоненты имеют размерность [см$^{-1}$]. Обычно $\mathbf{k}$ пишут в виде

$$\mathbf{k} = \frac{\mathbf{p}}{\hbar}, \tag{2.17}$$

где $\mathbf{p}$ — вектор размерности импульса. Он называется *квазиимпульсом*.

Соответственно

$$\psi(\mathbf{r}) = e^{\frac{i}{\hbar}\mathbf{pr}} u_{\mathbf{p}}(\mathbf{r}). \tag{2.15'}$$

Названия «квазиволновой вектор» и «квазиимпульс» указывают на известную аналогию между рассматриваемой задачей и случаем свободно движущегося электрона. Действительно, при свободном движении, когда потенциальная энергия электрона постоянна (и может быть положена равной нулю), уравнение Шредингера (2.2) имеет вид

$$-\frac{\hbar^2}{2m_0}\nabla^2\psi = E\psi. \tag{2.18}$$

Решения (2.18) суть плоские волны

$$\psi = Ne^{\frac{i}{\hbar}\mathbf{pr}}, \tag{2.19}$$

где N — нормировочный множитель. Вектор $\mathbf{p}$ в данном случае есть обычный импульс, связанный с энергией E равенством

$$E = \frac{\mathbf{p}^2}{2m_0}. \tag{2.20}$$

Из формулы (2.17) непосредственно вытекает соотношение де Бройля, связывающее импульс с волновым вектором:

$$\mathbf{p} = \hbar\mathbf{k}.$$

Формула (2.15') несколько напоминает (2.19), причем квазиимпульс играет роль, в известной мере аналогичную импульсу. Формально (2.19) есть частный случай (2.15') при $u_{\mathbf{p}} = \text{const} \equiv N$. И действительно, выражение (2.19) можно получить с помощью тех же рассуждений, что и (2.15'), требуя лишь, чтобы вероятность обнаружить электрон не изменялась при переходе из любой точки

*) Задолго до применения квантовой механики к задачам теории твердого тела одномерный аналог уравнения (2.4) был исследован в теории колебаний. Равенство (2.15) (в одномерном случае) было установлено Хиллом; в связи с этим функции такого вида называют также функциями Хилла.

пространства в любую другую. Это условие означает, что вектор $\mathbf{a}_n$ становится произвольным. Тогда в силу (2.11) должно иметь место равенство $|\psi|^2 = \text{const}$ — в соответствии с (2.19). Физически эквивалентность всех без исключения точек пространства означает отсутствие потенциальных сил, действующих на электрон. Действительно, при наличии таких сил потенциальная энергия электрона была бы различной в разных точках, т. е. по крайней мере некоторые из них были бы не эквивалентны другим.

Так обстоит дело, в частности, в периодическом поле, когда функция $u_k(\mathbf{r})$ отнюдь не сводится к константе и, следовательно, $|\psi|^2 \neq \text{const}$. При этом аналогия между импульсом и квазиимпульсом (или, соответственно, волновым и квазиволновым векторами) оказывается не полной. Так, например, легко проверить, что выражение (2.19) есть собственная функция оператора импульса (2.3′), принадлежащая собственному значению $\mathbf{p}$:

$$-i\hbar\nabla e^{\frac{i}{\hbar}\mathbf{pr}} = \mathbf{p}e^{\frac{i}{\hbar}\mathbf{pr}}.$$

С другой стороны, выражение (2.15′) этим свойством не обладает и фигурирующие там компоненты квазиимпульса $\mathbf{p}$ не являются собственными значениями операторов (2.3). Иначе говоря, значения компонент импульса в состоянии, описываемом волновой функцией (2.15′), точно не заданы.

Легко понять причины сходства и различия между импульсом и квазиимпульсом. Первый характеризует движение свободного электрона, когда система обладает инвариантностью относительно сдвига на любой вектор (все точки пространства эквивалентны). Второй характеризует движение в периодическом силовом поле, когда система обладает инвариантностью относительно сдвига на векторы решетки $\mathbf{a}_n$ (эквивалентны только точки, отстоящие друг от друга на векторы $\mathbf{a}_n$). Наличие инвариантности относительно сдвига в обоих случаях приводит к возможности охарактеризовать состояние электрона некоторым постоянным вектором — импульсом или квазиимпульсом. Различие в допустимых сдвигах приводит к тому, что импульс и квазиимпульс представляет собой существенно разные физические величины.

§ 3. Зоны Бриллюэна

Введенный в предыдущем параграфе вектор квазиимпульса определяется равенствами (2.14) и (2.17). Комбинируя их, мы получаем

$$\psi(\mathbf{r}+\mathbf{a}_n) = e^{\frac{i}{\hbar}\mathbf{pa}_n}\psi(\mathbf{r}). \tag{3.1}$$

Таким образом, вектор $\mathbf{p}$ (или $\mathbf{k} = \mathbf{p}/\hbar$) характеризует закон преобразования волновой функции электрона при сдвиге ее аргу-

мента на какой-либо вектор решетки. Разным собственным функциям соответствуют, вообще говоря, различные значения квазиимпульса (квазиволнового вектора). Поэтому компоненты его (как и компоненты импульса в случае свободного электрона) следует рассматривать как квантовые числа, характеризующие данное стационарное состояние. Однако, в отличие от компонент импульса и от квантовых чисел, встречающихся в теории атома, квазиимпульс определяется в принципе неоднозначно. Действительно, обозначим через $\mathbf{c}$ вектор, скалярное произведение которого на $\mathbf{a}_n$ есть целое кратное $2\pi\hbar$:

$$\mathbf{a}_n\mathbf{c} = 2\pi\hbar \times (\text{целое число}). \tag{3.2}$$

Очевидно, векторы $\mathbf{p}$ и $\mathbf{p} + \mathbf{c}$, будучи подставлены в правую часть (3.1), дают один и тот же результат. Но равенство (3.1) есть единственное условие, определяющее квазиимпульс. Следовательно, векторы $\mathbf{p}$ и $\mathbf{p} + \mathbf{c}$ физически эквивалентны: оба они определяют одно и то же преобразование волновой функции.

Нетрудно найти явный вид вектора $\mathbf{c}$. Для этого следует лишь ввести понятие обратной решетки. Основные векторы последней $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ определяются равенствами

$$\mathbf{b}_1 = 2\pi\frac{[\mathbf{a}_2\times\mathbf{a}_3]}{V_0}, \quad \mathbf{b}_2 = 2\pi\frac{[\mathbf{a}_3\times\mathbf{a}_1]}{V_0}, \quad \mathbf{b}_3 = 2\pi\frac{[\mathbf{a}_1\times\mathbf{a}_2]}{V_0}, \tag{3.3}$$

где $V_0 = |(\mathbf{a}_1[\mathbf{a}_2\times\mathbf{a}_3])|$ есть объем параллелепипеда, построенного на векторах $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ (объем элементарной ячейки). В частности, в простой кубической решетке, когда $a_1 = a_2 = a_3 = a$ и $b_1 = b_2 = b_3 = b$, мы имеем

$$b = \frac{2\pi}{a}. \tag{3.3'}$$

Очевидно, векторы $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ имеют размерность обратной длины. На основных векторах $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ можно построить периодическую решетку. Она и называется обратной (по отношению к прямой решетке данного кристалла).

Произвольный вектор обратной решетки имеет вид

$$\mathbf{b}_m = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3, \tag{3.4}$$

где m_1, m_2, m_3 — положительные или отрицательные целые числа или нули (при этом m_1, m_2 и m_3 не равны нулю одновременно), $m = \{m_1, m_2, m_3\}$.

Элементарная ячейка обратной решетки представляет собой параллелепипед, построенный на векторах $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$. «Объем» этого параллелепипеда равен $|(\mathbf{b}_1[\mathbf{b}_2\times\mathbf{b}_3])|$ (разумеется, он имеет размерность обратного объема). Подставляя сюда формулы (3.3) для $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ и раскрывая получающееся произведение, находим

$$|(\mathbf{b}_1[\mathbf{b}_2\times\mathbf{b}_3])| = (2\pi)^3/V_0. \tag{3.5}$$

Как и в случае прямой решетки, выбор элементарной ячейки в обратной решетке неоднозначен и определяется соображениями удобства.

Другой способ построения элементарной ячейки состоит в следующем. Какой-то узел обратной решетки выбирают в качестве начала координат и соединяют его прямыми линиями с ближайшими к нему узлами. Через середины этих линий перпендикулярно к ним проводят плоскости. В качестве элементарной ячейки обратной решетки можно выбрать наименьший многогранник, ограниченный так построенными плоскостями и содержащий внутри себя начало координат. Этот многогранник называется *ячейкой Вигнера — Зейтца*.

Такие многогранники можно построить около любого узла решетки; при этом они не перекрываются и совокупность их заполняет все обратное пространство. Отсюда следует, что объем одного многогранника действительно равен $(2\pi)^3/V_0$, как это и должно быть. В отличие от параллелепипеда, построенного на векторах $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$, элементарная ячейка, выбранная указанным только что образом, обладает всеми свойствами симметрии обратной решетки.

Из определения (3.3) вытекают равенства

$$\begin{aligned} &\mathbf{a}_1\mathbf{b}_1 = \mathbf{a}_2\mathbf{b}_2 = \mathbf{a}_3\mathbf{b}_3 = 2\pi, \\ &a_\alpha b_\beta = 0, \quad \alpha \neq \beta \qquad (\alpha,\ \beta = 1, 2, 3). \end{aligned} \tag{3.6}$$

Умножим теперь произвольный вектор решетки (II.1.1) на вектор обратной решетки (3.4). Пользуясь соотношениями (3.6), мы получаем

$$\mathbf{a}_n\mathbf{b}_m = (n_1m_1 + n_2m_2 + n_3m_3)\,2\pi.$$

В скобках в правой части этого равенства стоит целое число, и, следовательно, вектор $\mathbf{c}$, удовлетворяющий условию (3.2), можно записать в виде

$$\mathbf{c} = \hbar\mathbf{b}_m. \tag{3.7}$$

Итак, квазиимпульс определен лишь с точностью до вектора обратной решетки, умноженного на $\hbar$. Это обстоятельство позволяет ограничить изменение компонент квазиимпульса конечной областью, исчерпывающей все физически неэквивалентные их значения. Такая область — совокупность всех физически неэквивалентных значений квазиимпульса — называется *зоной Бриллюэна*. В силу произвольности вектора $\mathbf{b}_m$ в (3.7) выбор ее неоднозначен. Так, можно выбрать в качестве зоны Бриллюэна область, определяемую неравенствами

$$\begin{aligned} -\pi\hbar < \mathbf{pa}_1 \leqslant \pi\hbar, \\ -\pi\hbar < \mathbf{pa}_2 \leqslant \pi\hbar, \\ -\pi\hbar < \mathbf{pa}_3 \leqslant \pi\hbar. \end{aligned} \tag{3.8}$$

Эти неравенства определяют некоторый параллелепипед в **p**-пространстве, содержащий в себе начало координат. Его называют *первой зоной Бриллюэна*.

Можно определить первую зону Бриллюэна и для компонент квазиволнового вектора **k**: надо лишь заменить **p** на **k** в неравенствах (3.8), опустив множители $\hbar$. Квазиимпульс (или квазиволновой вектор), изменяющийся в пределах первой зоны Бриллюэна, называется приведенным. В частности, в простой кубической решётке векторы $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ одинаковы по величине (равной постоянной решетки a) и направлены по трем взаимно перпендикулярным осям куба. Выбирая эти оси в качестве координатных, получаем из (3.8) для данного частного случая

$$-\frac{\pi\hbar}{a} < p_\alpha \leqslant \frac{\pi\hbar}{a}, \qquad \alpha = x,\ y,\ z. \tag{3.8'}$$

Первая зона Бриллюэна здесь представляет собой куб объема $\frac{(2\pi\hbar)^3}{V_0} = \frac{(2\pi\hbar)^3}{a^3}$. В **k**-пространстве соответствующий объем равен $(2\pi)^3/V_0$.

Полученное только что выражение для объема первой зоны Бриллюэна справедливо и для произвольной решетки. Действительно, интересующий нас объем дается интегралом

$$\mathscr{J} = \int dp_x\, dp_y\, dp_z,$$

взятым по области, определяемой неравенствами (3.8). Для вычисления этого интеграла удобно ввести переменные

$$\varphi_1 = \frac{1}{\hbar}(\mathbf{pa}_1), \quad \varphi_2 = \frac{1}{\hbar}(\mathbf{pa}_2), \quad \varphi_3 = \frac{1}{\hbar}(\mathbf{pa}_3).$$

Якобиан перехода от переменных φ_1, φ_2, φ_3 к переменным p_x, p_y, p_z представляет собой детерминант [1]

$$\begin{vmatrix} \frac{\partial\varphi_1}{\partial p_x} & \frac{\partial\varphi_2}{\partial p_x} & \frac{\partial\varphi_3}{\partial p_x} \\ \frac{\partial\varphi_1}{\partial p_y} & \frac{\partial\varphi_2}{\partial p_y} & \frac{\partial\varphi_3}{\partial p_y} \\ \frac{\partial\varphi_1}{\partial p_z} & \frac{\partial\varphi_2}{\partial p_z} & \frac{\partial\varphi_3}{\partial p_z} \end{vmatrix} = \frac{1}{\hbar^3} \begin{vmatrix} a_{1x} & a_{2x} & a_{3x} \\ a_{1y} & a_{2y} & a_{3y} \\ a_{1z} & a_{2z} & a_{3z} \end{vmatrix} = \frac{(\mathbf{a}_1[\mathbf{a}_2 \times \mathbf{a}_3])}{\hbar^3} = \frac{V_0}{\hbar^3}.$$

Таким образом,

$$d\varphi_1\, d\varphi_2\, d\varphi_3 = \frac{V_0}{\hbar^3}\, dp_x\, dp_y\, dp_z,$$

и, следовательно,

$$\mathscr{J} = \frac{\hbar^3}{V_0} \int_{-\pi}^{+\pi} d\varphi_1 \int_{-\pi}^{+\pi} d\varphi_2 \int_{-\pi}^{+\pi} d\varphi_3 = \frac{(2\pi\hbar)^3}{V_0}.$$

Выбирая другие периоды p_x, p_y, p_z, мы получим вторую, третью и т. д. зоны Бриллюэна. Можно доказать [М6], что объемы их всех одинаковы и равны $(2\pi\hbar)^3/V_0$.

Вспоминая теперь равенство (3.5), видим, что объем зоны Бриллюэна равен объему элементарной ячейки в обратной решетке (умноженному на $\hbar^3$, если речь идет о **p**-пространстве).

Как и элементарную ячейку, первую зону Бриллюэна можно выбрать и другим способом. Действительно, по определению она должна обладать следующими свойствами: а) внутри нее должна содержаться точка $\mathbf{k} = 0$; б) любые два вектора **k**, входящие в нее, могут отличаться друг от друга не более чем на $(\mathbf{b}_1 + \mathbf{b}_2 + \mathbf{b}_3)$, и в) объем ее равен $(2\pi)^3/V_0$. Этими свойствами обладает ячейка Вигнера — Зейтца. Она и представляет собой первую зону Бриллюэна. В дальнейшем мы будем пользоваться этим определением.

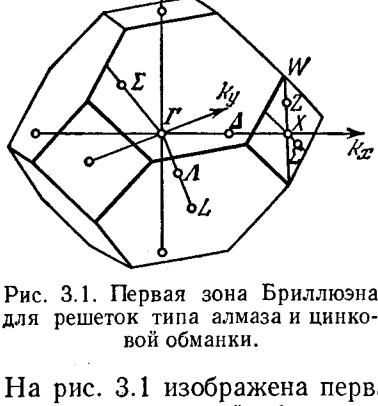


Рис. 3.1. Первая зона Бриллюэна для решеток типа алмаза и цинковой обманки.

Таким образом, зона Бриллюэна есть чисто геометрическое понятие: форма ее зависит только от структуры решетки, но не от природы действующих в ней сил. Более того, как видно из предыдущего, зона Бриллюэна определяется только основными векторами решетки. Следовательно, она одна и та же как для простых, так и для базисных решеток одной и той же сингонии, например для простой гранецентрированной решетки и для решетки типа алмаза. На рис. 3.1 изображена первая зона Бриллюэна для решеток типа алмаза и цинковой обманки. Некоторые точки в ней представляют особый интерес при исследовании поведения электронов и дырок. Это — «точки симметрии», обладающие тем свойством, что они переходят сами в себя при некоторых преобразованиях симметрии, допускаемых в данной решетке. К числу названных точек относятся центр первой зоны Бриллюэна (т. е. начало координат в **k**-пространстве), центры ее граней или ребер, точки на осях — линиях, соединяющих центр зоны с центрами граней или ребер, и т. д. Их принято обозначать большими греческими или латинскими буквами, которые и указаны на рисунках. Так, точка Γ есть не что иное, как центр первой зоны Бриллюэна, и т. д.

Неоднозначность квазиимпульса есть специфическое свойство электрона, движущегося в периодическом поле. Она составляет наиболее резкое различие между квазиимпульсом и импульсом: последний определен однозначно, и компоненты его p_x, p_y, p_z изменяются в пределах от $-\infty$ до $+\infty$. Заметим, что мы можем формально убедиться в этом, требуя, как и в конце предыдущего параграфа, инвариантности системы относительно сдвига на любой вектор. Действительно, в этом случае точки, бесконечно близкие

друг к другу, также являются эквивалентными, т. е. «постоянная решетки» a сколь угодно мала. Устремляя a к нулю, видим из (3.8′), что компоненты p_x, p_y, p_z изменяются в пределах от $-\infty$ др $+\infty$. Объем «зоны Бриллюэна» при этом также оказывается бесконечным: она представляет собой просто все импульсное пространство.

Неравенства (3.8) определяют пределы изменения компонент квазиимпульса, но ничего не говорят о физически дозволенных их значениях. Таковые определяются граничными условиями, накладываемыми на волновую функцию ψ. Строго говоря, граничные условия призваны отражать физическую ситуацию на поверхности образца. Эта ситуация определяется характером сил, действующих на электроны. Однако все силы взаимодействия, с которыми мы имеем дело, более или менее быстро убывают с расстоянием. Следовательно, условия на поверхности не могут сколько-нибудь заметно влиять на поведение электронов в глубине кристалла, если размеры его достаточно велики. Поэтому для определения возможных значений квазиимпульса в большом образце нет необходимости рассматривать истинную (далеко не простую) картину поверхностных явлений, а можно воспользоваться следующим искусственным приемом. Выделим прежде всего «внутреннюю область» кристалла, определив ее как объем, в котором не сказываются специфические эффекты, связанные с наличием поверхности. Внутреннюю область разобьем на ряд достаточно больших частей, например кубов со стороной L. Длина L должна значительно превышать все «физические» длины, фигурирующие в задаче, — постоянную решетки, длину волны электрона, длину свободного пробега и т. д.; в остальном значение L произвольно. Поскольку различные кубы ничем не выделены, естественно потребовать, чтобы значения волновой функции в соответственных (отстоящих друг от друга на расстояние L) точках соседних кубов были одинаковыми:

$$\psi(x, y, z)=\psi(x+L, y, z)=\ldots=\psi(x+L, y+L, z+L). \qquad (3.9)$$

Равенства (3.9) позволяют ограничиться изучением движения электронов в пределах только одного куба, рассматривая его как «кристалл»; все, что в нем происходит, будет повторяться и в других кубах. Таким образом, оказывается возможным формально ввести в задачу «размеры кристалла», не интересуясь в то же время явлениями на его поверхности *).

Условия (3.9) означают, что волновая функция должна быть периодична с периодом L. Соответственно и сами условия (3.9) называются условиями периодичности (или условиями Кармана — Борна). Сам куб со стороной L называется кубом периодичности

*) Можно показать [2], что при достаточно большой длине L распределение собственных значений компонент $\mathbf{p}$ с точностью до величин порядка L^{-1} не зависит от граничных условий.

(употребляют также названия «фундаментальный объем», «основной объем» или «основная область»).

Поскольку длина L сколь угодно велика, мы вправе считать L целым кратным a, где a — постоянная решетки по соответствующему направлению. Поэтому, накладывая условия (3.9) на функцию (2.15) и принимая во внимание свойство периодичности $u_k(\mathbf{r})$ (2.16), мы получаем

$$e^{i(k_x x + k_y y + k_z z)} u_k(x, y, z) = e^{ik_x(x+L) + ik_y(y+L) + ik_z(z+L)} u_k(x+L, y+L, z+L) = e^{ik_x(x+L) + ik_y(y+L) + ik_z(z+L)} u_k(x, y, z).$$

Отсюда

$$k_x = \frac{2\pi}{L} n_x, \quad k_y = \frac{2\pi}{L} n_y, \quad k_z = \frac{2\pi}{L} n_z, \tag{3.10}$$

где n_x, n_y, n_z — положительные или отрицательные целые числа (ограниченные условиями (3.8)) или нули. Соответственно

$$p_x = \frac{2\pi\hbar}{L} n_x, \quad p_y = \frac{2\pi\hbar}{L} n_y, \quad p_z = \frac{2\pi\hbar}{L} n_z. \tag{3.10'}$$

Таким образом, значения компонент квазиволнового вектора (или квазиимпульса) образуют дискретную совокупность. Разности между соседними значениями их, однако, весьма малы (равны соответственно $2\pi/L$ и $2\pi\hbar/L$). По этой причине указанная дискретность не проявляется в наблюдаемых на опыте электрических и оптических явлениях. Спектр такого типа называется квазинепрерывным.

§ 4. Энергетические зоны

Обратимся к вопросу об энергетическом спектре электрона, движущегося в периодическом поле. Подставляя волновую функцию (2.15′) в уравнение (2.4), получим

$$-\frac{\hbar^2}{2m_0}\nabla^2 u_{\mathbf{p}}(\mathbf{r}) - \frac{i\hbar}{m_0}(\mathbf{p}, \nabla u_{\mathbf{p}}) + \frac{\mathbf{p}^2}{2m_0} u_{\mathbf{p}} + U(\mathbf{r})\, u_{\mathbf{p}} = E u_{\mathbf{p}}. \tag{4.1}$$

Под U здесь понимается потенциальная энергия электрона в периодическом поле *).

*) При учете спин-орбитального взаимодействия в левой части (4.1) появились бы два дополнительных слагаемых:

$$\frac{\hbar}{4m_0^2c^2}(\boldsymbol{\sigma}[\nabla U \times \mathbf{p}])\, u_{\mathbf{p}} - \frac{i\hbar^2}{4m_0^2c^2}(\boldsymbol{\sigma}[\nabla U \times \nabla u_{\mathbf{p}}]).$$

Дальнейшие рассуждения при этом несколько усложнились бы, но качественные результаты (4.4), (4.5) и др. остались бы в силе (см., например, [M7]). При фактическом расчете функций $u_{\mathbf{p}}$ и собственных значений E учет спин-орбитального взаимодействия может, однако, оказаться существенным.

Уравнение (4.1) представляет собой обычную задачу на собственные значения, в которой компоненты вектора $\mathbf{p}$ играют роль параметров. Разумеется, от них зависят не только собственные функции, но и собственные значения энергии: $E = E(\mathbf{p})$. При заданных величинах p_x, p_y и p_z уравнение (4.1) имеет, вообще говоря, много собственных значений и принадлежащих им собственных функций. Пронумеруем их индексом l, расположив значения E в возрастающем порядке *):

$$E = E_l(\mathbf{p}), \qquad u_{\mathbf{p}}(\mathbf{r}) = u_{\mathbf{p}l}(\mathbf{r}),$$

причем

$$E_1(\mathbf{p}) \leqslant E_2(\mathbf{p}) \leqslant \ldots \tag{4.2}$$

Такой способ нумерации не обязателен, но удобен для дальнейшего.

Как показано в Приложении II, собственные функции уравнения (4.1), отвечающие различным значениям l при фиксированных компонентах $\mathbf{p}$, взаимно ортогональны **):

$$\int u^*_{\mathbf{p}l} u_{\mathbf{p}l'}\, d\mathbf{r} = \delta_{ll'}. \tag{4.3}$$

Здесь $\delta_{ll'}$ — символ Кронеккера (П.II.9); интеграл берется по основному объему.

Зафиксируем временно индекс l и рассмотрим свойства энергии как функции квазиимпульса при данном l.

Заменим, прежде всего, функцию $u_{\mathbf{p}}$ на $u^*_{\mathbf{p}}$ в уравнении (4.1), а также изменим там знак у $\mathbf{p}$. Последняя операция эквивалентна замене i на $-i$ в (4.1). Иначе говоря, указанные замены превращают уравнение (4.1) в сопряженное ему. Собственное значение E, будучи вещественным, при этом измениться не может. Следовательно,

$$E_l(-\mathbf{p}) = E_l(\mathbf{p}); \tag{4.4}$$

энергия электрона в периодическом поле есть четная функция квазиимпульса. Далее, мы знаем, что значения квазиимпульса, отличающиеся на $\hbar\mathbf{b}$, физически эквивалентны (для краткости мы не пишем теперь индекс m у вектора обратной решетки $\mathbf{b}$, если это не вызывается необходимостью). Следовательно, соответствующие значения энергии должны совпадать:

$$E_l(\mathbf{p}) = E_l(\mathbf{p} + \hbar\mathbf{b}). \tag{4.5}$$

Таким образом, энергия электрона в идеальной решетке есть периодическая функция квазиимпульса с периодом обратной решетки, умноженным на $\hbar$.

*) Не следует путать l с орбитальным квантовым числом, упоминавшимся в гл. II. Последнее нигде более нам не встретится.

**) Обратим внимание на то, что в левой части (4.1) фигурируют функции, принадлежащие одному и тому же вектору $\mathbf{p}$: при разных $\mathbf{p}$ это были бы собственные функции разных операторов, не обязанные удовлетворять каким-либо условиям типа (4.3).

То же самое относится и к функции $u_{\mathbf{p}l}(\mathbf{r})$:

$$u_{\mathbf{p}l}(\mathbf{r}) = u_{\mathbf{p}+\hbar\mathbf{b},\,l}(\mathbf{r}). \tag{4.5'}$$

Рассматривая зависимость энергии от квазиимпульса, мы можем ограничиться изменением компонент $\mathbf{p}$ в пределах одного периода (т. е. одной какой-нибудь зоны Бриллюэна). Пусть квазиимпульс пробегает зону Бриллюэна. Можно доказать [М6, М7], что при этом энергия есть непрерывная функция $\mathbf{p}$. Иначе говоря, ее значения квазинепрерывно заполняют некоторый интервал

$$E_{l,\,\min} \leqslant E_l(\mathbf{p}) \leqslant E_{l,\,\max}. \tag{4.6}$$

В соответствии с формулами (3.10') разности между последовательными уровнями энергии в пределах данного интервала убывают как $1/L$. При увеличении размеров куба периодичности они становятся сколько угодно малыми, в силу чего энергетический спектр электрона в пределах интервала (4.6) практически не отличается от непрерывного. По этой причине приставку «квази» часто опускают, говоря просто об области непрерывного спектра.

Интервал энергий, в пределах которого значения $E_l(\mathbf{p})$ изменяются непрерывно, носит название *энергетической зоны* (или полосы). Число различных зон равно числу значений, принимаемых индексом l. По этой причине последний называют номером зоны (или зонным индексом).

Иногда удобно рассматривать все функции $E_l(\mathbf{p})$ как различные ветви одной многозначной функции. Тогда индекс l есть номер ветви.

Величины $E_{l,\,\min}$ и $E_{l,\,\max}$ — минимум и максимум энергии в l-той зоне — называют, соответственно, дном и потолком зоны. Разность между ними есть ширина l-той зоны.

Обратимся теперь к зависимости собственных значений энергии от зонного индекса l. Рассмотрим две соседние зоны, характеризуемые функциями $E_l(\mathbf{p})$ и $E_{l+1}(\mathbf{p})$. Согласно (4.2) может реализоваться одна из двух возможностей:

либо

$$E_{l,\,\max} < E_{l+1,\,\min}, \tag{4.7а}$$

либо

$$E_{l,\,\max} \geqslant E_{l+1,\,\min}. \tag{4.7б}$$

Последнее неравенство, разумеется, не противоречит (4.2), ибо дно l-той и потолок $(l+1)$-й зоны могут располагаться в различных точках зоны Бриллюэна.

В случае (4.7а) мы будем говорить, что l-тая и $(l+1)$-я зоны разделены *запрещенной зоной*. Ширина последней есть $E_{l,\,l+1} = E_{l+1,\,\min} - E_{l,\,\max}$. Значения энергии, лежащие в запрещенной зоне, не относятся к собственным значениям гамильтониана. Иначе говоря, в отсутствие внешних полей электрон в идеальном кристалле такую энергию иметь не может, с чем и связано название «запре-

щенная зона». В противоположность запрещенным, энергетические зоны, о которых мы говорили выше, называются *разрешенными*.

Ситуация, соответствующая случаю (4.7а), изображена на рис. 3.2, *а*. Разрешенные зоны, в которых лежат собственные значения энергии, заштрихованы; между ними расположена запрещенная зона.

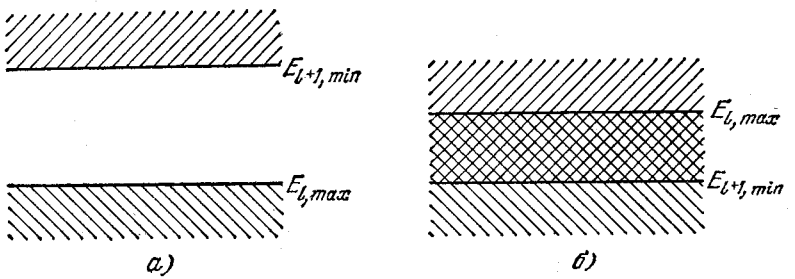


Рис. 3.2. Энергетические зоны: *а*) неперекрывающиеся; *б*) перекрывающиеся.

В случае (4.7б) мы будем говорить, что l-тая и $(l+1)$-я зоны перекрываются (или смыкаются, если имеет место знак точного равенства). Запрещенной зоны здесь, очевидно, нет. Такая ситуация изображена на рис. 3.2, *б*. Разрешенные зоны (l-тая и $(l+1)$-я) заштрихованы там в разных направлениях; значения энергии из области, покрытой двойной штриховкой, принадлежат обеим зонам.

Характеристики зонного спектра, т. е. ширины разрешенных и запрещенных зон, значения квазиимпульса, при которых функции $E_l(\mathbf{p})$ достигают максимумов и минимумов, и т. п. определяются конкретными особенностями данного кристалла. Их можно найти, решая уравнение Шредингера для данной конкретной системы. Практически, однако, проще определять необходимые величины непосредственно из опыта (см. § IV.3 и гл. XVIII).

§ 5. Метод сильно связанных электронов

Доказанная в § 2 теорема Блоха приводит к выводу о зонном характере энергетического спектра идеального кристалла. Она, однако, не раскрывает физического механизма образования зон. Существо дела можно понять, рассматривая, как дискретные энергетические уровни, характерные для свободных атомов, превращаются в зоны при постепенном сближении атомов и объединении их в кристаллическую решетку. Для явного решения этой задачи удобно воспользоваться приближенным методом, предложенным впервые также Ф. Блохом и применимым, когда расстояния между соседними атомами велики по сравнению с их размерами. Для упрощения выкладок будем рассматривать не трехмерный кристалл,

а одномерную бесконечную цепочку одинаковых периодически расположенных одновалентных атомов. Обобщение на случай трех измерений и атомов большей валентности требует — в рамках данного метода — лишь не принципиального усложнения вычислений.

Пусть нам известна волновая функция валентного электрона в изолированном атоме. Обозначим ее через $\varphi_g(\mathbf{r} - \mathbf{R}_g)$, где g — номер атома в цепочке, $\mathbf{r}$ и $\mathbf{R}_g$ — радиус-векторы электрона и g-го атомного остова *). Функция φ_g удовлетворяет уравнению

$$-\frac{\hbar^2}{2m_0}\nabla^2\varphi_g + U_g(\mathbf{r})\,\varphi_g = E_a\varphi_g. \tag{5.1}$$

Здесь $U_g(\mathbf{r})$ — потенциальная энергия взаимодействия электрона с g-м атомом, E_a — собственное значение энергии, принадлежащее собственной функции φ_g, т. е. дискретный энергетический уровень валентного электрона в данном атоме. Будем считать, что функция φ_g описывает s-состояние. Тогда уровень с энергией E_a вырожден (двукратно) только по спину.

Волновые функции, описывающие дискретные атомные уровни, быстро убывают по мере удаления электрона от ядра. Асимптотически на больших расстояниях φ_g ведет себя, в основном, как показательная функция:

$$\varphi_g \sim \exp\left\{-\frac{|\mathbf{r} - \mathbf{R}_g|}{r_a}\right\}, \tag{5.2}$$

где r_a — константа размерности длины. Она играет роль «эффективного радиуса» атома и часто называется радиусом валентной орбиты. Мы будем считать ее малой по сравнению с расстоянием между соседними атомами d:

$$r_a \ll d. \tag{5.3}$$

В реальных кристаллах это условие, как правило, не выполняется; поэтому излагаемый ниже метод расчета, основанный, по существу, на неравенстве (5.3), может иметь только ориентировочное значение. Тем не менее он позволяет понять физическую причину образования энергетических зон и разобраться в ряде их характеристик.

В силу предположения об идентичности атомов цепочки, функции φ_g с различными номерами g отличаются друг от друга лишь тем, что они «центрированы» около различных атомов; уровни энергии, которые им принадлежат, разумеется, одинаковы, почему мы и не снабдили E_a значком g.

При постепенном сближении атомов электрон начинает взаимодействовать не только со «своим», но и с «чужими» атомными остовами. Соответствующее уравнение Шредингера имеет вид (2.4).

*) Если направить, например, ось x вдоль цепочки, то $\mathbf{R}_g$ есть вектор с компонентами $\{dg, 0, 0\}$.

Условие (5.3) позволяет ожидать, что вблизи каждого атомного остова валентный электрон движется в основном так же, как и в изолированном атоме. Иначе говоря, вблизи g-го атомного остова волновая функция электрона в цепочке ψ должна быть близка к φ_g. В соответствии со сказанным будем искать решение уравнения (2.4) в виде

$$\psi(\mathbf{r}) = \sum_g a_g \varphi_g, \tag{5.4}$$

где a_g — коэффициенты, подлежащие определению. Подставляя (5.4) в (2.4), мы имеем

$$\sum_{g=-\infty}^{+\infty} a_g \left\{ -\frac{\hbar^2}{2m_0} \nabla^2 \varphi_g + (U - U_g)\varphi_g + U_g \varphi_g - E\varphi_g \right\} = 0. \tag{5.5}$$

С учетом уравнения (5.1) равенство (5.5) принимает вид

$$\sum_{g=-\infty}^{+\infty} a_g \{(E_a - E)\varphi_g + (U - U_g)\varphi_g\} = 0. \tag{5.6}$$

Для определения коэффициентов a_g умножим обе части уравнения (5.6) на комплексно сопряженную волновую функцию $\varphi^*_{g'}$ электрона в изолированном атоме g' и проинтегрируем по координатам электрона $\mathbf{r}$. Введем обозначения

$$\int \varphi^*_{g'} \varphi_g \, d\mathbf{r} = S_{g'g}, \quad \int \varphi^*_{g'} (U - U_g)\varphi_g \, d\mathbf{r} = U_{g'g}. \tag{5.7}$$

Получим следующую систему уравнений:

$$\sum_{g=-\infty}^{+\infty} a_g \{(E_a - E) S_{g'g} + U_{g'g}\} = 0. \tag{5.8}$$

Очевидно, $S_{gg} = 1$, так как это просто нормировочный интеграл для функции φ_g. Однако при $g' \neq g$ интеграл $S_{g'g}$ не равен нулю. Можно лишь утверждать, что он мал по сравнению с единицей: из условия (5.3) и асимптотического вида волновых функций (5.2) вытекает, что с увеличением расстояния между атомами интегралы $S_{g'g}$ и $U_{g'g}$ убывают, в основном, как $\exp\left(-\frac{d}{r_a}|g' - g|\right)$. Величину $S_{g'g}$ при $g' \neq g$ иногда называют *интегралом неортогональности* или интегралом перекрытия. Величина $U_{g'g}$ при $g' \neq g$ называется *интегралом переноса*.

Поскольку все атомы цепочки одинаковы, интегралы $S_{g'g}$ и $U_{g'g}$ не могут зависеть от того, где именно расположены g-й и g'-й атомы: существенно только расстояние между ними. Это означает, что $S_{g'g}$ и $U_{g'g}$ зависят только от абсолютного значения разности аргументов:

$$S_{g'g} = S(|g' - g|), \quad U_{g'g} = U(|g' - g|). \tag{5.9}$$

Равенства (5.9) представляют собой математическое выражение условия идентичности всех атомов решетки и постоянства расстояния между соседними атомами. Иначе говоря, они выражают условие трансляционной инвариантности рассматриваемой нами системы.

Принимая во внимание равенства (5.9), легко в явном виде найти решение системы (5.8). Действительно, положим

$$a_g = Ne^{i\lambda g}, \tag{5.10}$$

где нормировочный множитель N и параметр λ не зависят от g.

Поскольку волновая функция должна быть квадратично интегрируема, параметр λ должен быть вещественным. Действительно, при комплексных (или чисто мнимых) значениях λ коэффициент a_g (а с ним и вся волновая функция ψ) обращался бы в бесконечность при $g \to \infty$ или при $g \to -\infty$.

Выбор решения в виде (5.10) с вещественными значениями λ имеет ясный физический смысл. В самом деле, квадраты модулей отдельных членов ряда в (5.4) характеризуют вероятность найти электрон вблизи g-го атомного остова. Чтобы убедиться в этом, вспомним, что вероятность найти электрон в некотором элементе объема ΔV дается интегралом по этому объему $\int\limits_{\Delta V} |\psi|^2 d\mathbf{r}$. Подставляя сюда ψ в виде (5.4) и замечая, что в силу (5.3) функции φ_g и $\varphi_{g'}$ при $g' \neq g$ почти не перекрываются, мы получаем

$$\int\limits_{\Delta V} |\psi|^2 d\mathbf{r} \simeq \sum_{g'=-\infty}^{+\infty} |a_{g'}|^2 \int\limits_{\Delta V} |\varphi_{g'}|^2 d\mathbf{r} \simeq |a_g|^2,$$

если ΔV есть малый элемент объема, окружающий g-й атомный остов.

Поскольку, согласно (5.10), $|a_g|^2 = \text{const}$, электрон в состоянии (5.4) с равной вероятностью может быть обнаружен около любого атома цепочки — в полном согласии с трансляционной инвариантностью системы.

Подставляя (5.10) в (5.8) и принимая во внимание (5.9), мы получаем

$$N \sum_{g=-\infty}^{+\infty} \{(E_a - E) S(|g' - g|) + U(|g' - g|)\} e^{i\lambda g} = 0.$$

Произведем здесь замену переменной суммирования, полагая $g' - g = g''$. Сокращая затем на $e^{i\lambda g'}$, находим

$$N \sum_{g''=-\infty}^{+\infty} \{(E_a - E) S(|g''|) + U(|g''|)\} e^{-i\lambda g''} = 0. \tag{5.11}$$

Это равенство уже не содержит аргумента g', чем и доказывается правильность выбора решения в виде (5.10).

Поскольку $N \neq 0$, равенство (5.11) означает, что в нуль должна обращаться сумма по g''. Это есть условие для определения собственных значений энергии электрона. Заменяя индекс суммирования g'' на g, имеем:

$$E = E_a + \frac{\sum\limits_{g=-\infty}^{+\infty} U(|g|)\, e^{-i\lambda g}}{\sum\limits_{g=-\infty}^{+\infty} S(|g|)\, e^{-i\lambda g}}. \tag{5.12}$$

Равенство (5.12) можно переписать в виде

$$E = E_a + \frac{U(0) + 2\sum\limits_{g=1}^{\infty} U(g)\cos\lambda g}{1 + 2\sum\limits_{g=1}^{\infty} S(g)\cos\lambda g}. \tag{5.12'}$$

Суммы по g в (5.12) или (5.12′) представляют собой периодические функции λ с периодом 2π.

Легко видеть, что с точностью до множителя d^{-1} параметр λ представляет собой не что иное, как одномерный аналог квазиволнового вектора $\mathbf{k}$, введенного в § 2 из общих соображений. Действительно, волновую функцию (5.4) с учетом (5.10) можно переписать в виде

$$\Psi = N \sum_{g=-\infty}^{+\infty} e^{i\lambda g} \varphi_g(\mathbf{r} - \mathbf{R}_g). \tag{5.13}$$

Введем вектор $\mathbf{k}$ с компонентами $\{\lambda d^{-1}, 0, 0\}$ и заметим, что компоненты вектора $\mathbf{R}_g$ суть, по определению, $\{dg, 0, 0\}$. Тогда формулу (5.13) можно переписать в виде

$$\Psi = N \sum_{g=-\infty}^{+\infty} \varphi_g(\mathbf{r} - \mathbf{R}_g)\, e^{i(\mathbf{k}, \mathbf{R}_g - \mathbf{r}) + i\mathbf{k}\mathbf{r}} = e^{i\mathbf{k}\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}). \tag{5.14}$$

Очевидно, определяемая этим соотношением функция

$$u_{\mathbf{k}}(\mathbf{r}) = N \sum_{g=-\infty}^{+\infty} e^{i(\mathbf{k}, \mathbf{R}_g - \mathbf{r})} \varphi_g(\mathbf{r} - \mathbf{R}_g) \tag{5.15}$$

периодична с периодом нашей одномерной решетки d. Действительно, пусть $\mathbf{a}$ есть вектор с компонентами $\{d, 0, 0\}$. Для функции $u_{\mathbf{k}}(\mathbf{r} + \mathbf{a})$ мы имеем

$$u_{\mathbf{k}}(\mathbf{r} + \mathbf{a}) = N \sum_{g=-\infty}^{+\infty} e^{i(\mathbf{k}, \mathbf{R}_g - \mathbf{r} - \mathbf{a})} \varphi_g(\mathbf{r} - \mathbf{R}_g + \mathbf{a}). \tag{5.16}$$

Производя здесь замену переменной суммирования $g = g' + 1$, убеждаемся, что правые части равенств (5.15) и (5.16) одинаковы.

Итак, выражение (5.14) есть частный случай общей формулы (2.15): волновая функция ψ представляет собой произведение плоской волны $e^{i\mathbf{kr}}$ на функцию, периодическую с периодом решетки.

Формулы (5.12), (5.12′) также находятся в полном соответствии с § 4. В самом деле, период обратной решетки в данном случае равен $b = 2\pi/d$. Следовательно, собственные значения энергии E периодически зависят от $\mathbf{k}$ с периодом b, как это и должно быть. Интервал $(-b/2, +b/2)$ определяет первую (одномерную) зону Бриллюэна. Видно, далее, что E есть четная функция $\mathbf{k}$ — опять в соответствии с результатами § 4.

Аналогичные формулы (в принятых выше предположениях) справедливы и для двух- и трехмерного случая. Так, для энергии электрона в простой трехмерной кубической решетке одновалентных атомов мы получаем

$$E = E_a + \frac{\sum\limits_{\mathbf{g}} U(|\mathbf{g}|)\, e^{-i\boldsymbol{\lambda}\mathbf{g}}}{\sum\limits_{\mathbf{g}} S(|\mathbf{g}|)\, e^{-i\boldsymbol{\lambda}\mathbf{g}}}, \tag{5.17}$$

где $\mathbf{g}$ — вектор, компоненты которого суть целые числа (или нули). Вектор $\boldsymbol{\lambda}$ очевидным образом связан с квазиволновым вектором $\mathbf{k}$:

$$\lambda_x = ak_x \quad \text{и т. д.}$$

В силу неравенства (5.3) величины $U(|g|)$ и $S(|g|)$ быстро убывают с ростом аргумента $|g|$. По этой причине в рядах, фигурирующих в правых частях (5.12), (5.12′) и (5.17), можно оставить только первые члены. Тогда равенство (5.12′) принимает вид

$$E = E_a + U(0) + 2[U(1) - U(0)\, S(1)] \cos\lambda. \tag{5.12″}$$

Интегралы $S_{g'g}$ и $U_{g'g}$ можно вычислить, зная волновые функции электронов в изолированных атомах. Иногда, однако, эти величины (или подходящие их комбинации) рассматривают просто как параметры, подлежащие определению из опыта. При этом недостатки рассматриваемого метода расчета исчезают, но зато теряется возможность непосредственно связать структуру зон с химическими свойствами атомов, образующих кристалл.

Формула (5.12″), как, впрочем, и (5.12′) и (5.17), ясно показывает, как влияет поле кристаллической решетки на энергетический спектр электронов.

Во-первых, атомный уровень E_a сдвигается на постоянную — не зависящую от λ — величину $U(0)$. Смысл ее очевиден: согласно (5.9) и (5.7)

$$U(0) = \int \varphi_g^*(\mathbf{r})\,(U - U_g)\,\varphi_g(\mathbf{r})\, d\mathbf{r}. \tag{5.18}$$

Таким образом, $U(0)$ есть средняя энергия электрона, локализованного на каком-нибудь одном (g-м) атоме решетки, в поле всех

остальных атомов (результат, разумеется, не зависит от g — сдвиг энергии для всех атомов один и тот же).

Во-вторых, дискретный атомный уровень «размывается» в зону: при изменении параметра λ от $-\pi$ до $+\pi$ последнее слагаемое в правой части (5.12″) непрерывно изменяется в интервале от $-2\,|\,U(1)-U(0)\,S(1)\,|$ до $+2\,|\,U(1)-U(0)\,S(1)\,|$ (знак разности $U(1)-U(0)\,S(1)$ в принципе может быть любым; для дальнейших выводов он несуществен). Все значения энергии в этом интервале дозволены: им соответствуют всюду ограниченные волновые функции (5.14), характеризуемые (как и энергия E) различными значениями параметра λ. Ширина этого интервала, т. е. «ширина разрешенной зоны», равна

$$4\,|\,U(1)-U(0)\,S(1)\,|.$$

Согласно (5.9) и (5.7) она определяется степенью перекрытия волновых функций соседних атомов φ_g и $\varphi_{g\pm1}$. С уменьшением перекрытия (т. е. с увеличением расстояния между атомами d) ширина зоны быстро (экспоненциально) стремится к нулю — зона «стягивается» в дискретный уровень (возмущенный, может быть, присутствием других атомов согласно (5.18)). Зависимость E от параметра λ при этом исчезает.

Знак разности $U(1)-U(0)\,S(1)$ определяет, лежит ли минимум энергии (5.12″) (т. е. дно зоны) в точке $\lambda=0$ ($\mathbf{k}=0$) или в точке $\lambda=\pi$, т. е. в центре или на границе зоны Бриллюэна *). Вблизи минимума функция (5.12″) имеет вид

$$E=E_0+d^2\,|\,U(1)-U(0)\,S(1)\,|\,k^2, \tag{5.19}$$

где E_0 есть минимальное значение энергии, а величина k отсчитывается от точки минимума. Выражение (5.19) формально совпадает с формулой для энергии свободной частицы $\hbar^2k^2/2m_0$, если массу последней m_0 заменить выражением

$$m=\frac{\hbar^2}{2d^2\,|\,U(1)-U(0)\,S(1)\,|}. \tag{5.20}$$

Величина m носит название *эффективной массы*. Заметим, что она убывает вместе с увеличением ширины разрешенной зоны.

Аналогично, вблизи максимума энергии (потолка зоны) мы имеем

$$E=\text{const}+\frac{\hbar^2k^2}{2m'},\qquad m'<0, \tag{5.21}$$

где число k отсчитывается от точки максимума. Константы здесь и в формуле (5.19), разумеется, различны; разность их равна ширине разрешенной зоны.

*) Точки $\lambda=\pi$ и $\lambda=-\pi$ физически эквивалентны, и, как указывалось в § 4, их следует рассматривать как одну точку.

Правая часть (5.21) также похожа на формулу для кинетической энергии свободной частицы, но теперь эффективная масса m' отрицательна (в случае (5.12′) она по абсолютной величине совпадает с m; вообще же величины m и $|m'|$ могут быть и различными).

Как уже отмечалось, электрон, описываемый волновой функцией (5.13), с равной вероятностью может быть обнаружен около любого атома цепочки. Это обстоятельство на первый взгляд представляется парадоксальным. Действительно, график потенциальной энергии электрона в рассматриваемой цепочке представляет собой совокупность чередующихся потенциальных ям, связанных с каждым атомным остовом, и «горбов» (рис. 3.3). Казалось бы, можно «запереть» электрон внутри какой-либо одной ямы, если полная энергия его E меньше потенциальной энергии на «горбе» (пунктирная линия на рис. 3.3). Это рассуждение, однако, целиком основано на классических представлениях, которые в данном случае недостаточны. В квантовой механике показывается, что «запереть» электрон в потенциальной яме конечной глубины и ширины невозможно: существует конечная вероятность обнаружить электрон в области барьера, окружающего яму (заштрихованные участки на рис. 3.3). Если, как это и имеет место в нашем случае, рядом с данной потенциальной ямой есть другие (где электрон может иметь ту же энергию), то возникает возможность изоэнергетического туннельного перехода электрона из одной ямы в другую. При этом он уже не связан с каким-либо отдельным атомом решетки, а обобществляется между ними. Именно этот туннельный эффект и позволяет электронам беспрепятственно перемещаться через кристалл без предварительной термической активации.

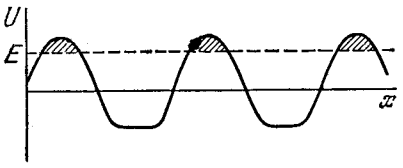


Рис. 3.3. Потенциальная энергия электрона в одномерной цепочке одинаковых атомов как функция координаты (схематически).

Вероятность туннельного перехода определяется степенью перекрытия волновых функций электрона в соседних потенциальных ямах, т. е., в нашем случае, интегралами $S_{g'g}$ и $U_{g'g}$. При безграничном раздвижении атомов вероятность туннельного эффекта обращается в нуль. И действительно, в неионизованном газе электроны локализованы у «своих» атомов.

В условиях, нами рассматриваемых (размеры атомов малы по сравнению с расстоянием между ними), вероятность туннельного перехода оказывается небольшой: электроны довольно сильно связаны со «своими» атомами *).

*) Отсюда — название «метод сильно связанных электронов».

Связь между вероятностью туннельного перехода и расстоянием между атомами позволяет наглядно интерпретировать зависимость эффективной массы от интегралов неортогональности и переноса (5.7). Именно уменьшение вероятности перехода (т. е. уменьшение $S_{g'g}$ и $U_{g'g}$) затрудняет перемещение электрона по кристаллу, что выражается, в частности, в «утяжелении» его: $m \to \infty$ при S (1) и U (1) $\to 0$ *). Тот факт, что вероятность туннельного перехода приближенно пропорциональна ширине разрешенной зоны, не должен вызывать удивления. Действительно, образование зон и туннельное просачивание электронов с одного атома на другой представляют собой два аспекта одного и того же явления — обобществления электронов между всеми атомами решетки, — наступающего благодаря перекрытию атомных волновых функций.

Итак, энергетические зоны в кристаллах возникают в результате «размытия» дискретных атомных уровней, обусловленного взаимодействием электронов с соседними (и, может быть, более далекими) атомами решетки и перекрытием волновых функций соседних атомов. Этот вывод, к которому мы пришли на основании приближенного расчета простейшей модели, оказывается совершенно общим. С другой стороны, как уже отмечалось, полученные в настоящем параграфе количественные соотношения на большую точность претендовать не могут **).

Тем не менее метод сильно связанных электронов иногда оказывается полезным. Главное его достоинство состоит в том, что он позволяет установить связь между характеристиками энергетических зон и химическими свойствами атомов, образующих кристалл, в частности — типом химической связи в данном материале. Действительно, расчет, изложенный в настоящем параграфе, нетрудно обобщить на более сложные случаи, принимая во внимание трехмерность задачи, наличие нескольких атомных уровней и т. д. (см., например, [М7]). При этом оказывается, что каждому атомному уровню соответствует столько зон, какова его кратность вырождения, причем различные зоны могут и перекрываться. Иначе говоря, вырожденным атомным уровням обычно соответствуют и вырожденные зоны. Аналогичным образом обстоит дело и в тех случаях, когда уровни энергии валентных электронов в атомах не вырождены, но достаточно близки по энергии.

Соответствующие волновые функции можно представить в виде сумм типа (5.4), с той лишь разницей, что теперь суммирование ведется не только по номерам атомов, но и по атомным состояниям, принимаемым нами в расчет. Иначе говоря, при образовании зон в результате размытия вырожденных или близких по энергии уровней роль «атомных функций» φ_g играют линейные комбинации

*) С этим связано и само название «интеграл переноса».

**) Обзор современных методов расчета зонной структуры твердого тела, претендующих на количественную точность, можно найти в книге [3].

атомных функций, принадлежащих названным уровням. Об этом говорят как о гибридизации соответствующих атомных состояний.

Эти качественные выводы не связаны непременно с условием (5.3); соответственно они дают известную ориентацию и в условиях, когда метод сильно связанных электронов в буквальной своей форме уже не применим. То же относится и к выводам о свойствах симметрии блоховских волновых функций; эта информация оказывается существенной во многих задачах, например в задаче о поглощении или испускании света при переходах электронов из одной энергетической зоны в другую (гл. XVIII).

§ 6. Закон дисперсии. Изоэнергетические поверхности

Ситуация, с которой мы встретились в двух последних параграфах, характерна для современной теории твердого тела. Именно, часто оказывается возможным, исходя из довольно общих физических соображений, выяснить некоторые общие особенности и свойства энергетического спектра системы, однако попытка явного расчета, например зависимости энергии от квазиимпульса, встречается с большими математическими трудностями. Как правило, при этом приходится вводить ряд упрощающих предположений и аппроксимаций. По этой причине важную роль играет экспериментальное исследование зонной структуры твердого тела. Чтобы правильно поставить его, надо прежде всего выяснить, с помощью каких понятий и величин удобно описывать энергетический спектр электронов в кристаллической решетке.

Введем прежде всего представление о *законе дисперсии*. Так называется зависимость энергии от квазиимпульса в разрешенной зоне *).

Закон дисперсии удобно изображать графически. Обычно пользуются одним из двух представлений.

Во-первых, можно фиксировать две из трех компонент квазиимпульса. Пусть, например, фиксированы компоненты p_y и p_z. Тогда зависимость $E_l(p_x)$ определяет некоторую кривую на плоскости (E_l, p_x). Кривые такого типа называются *дисперсионными*. Совокупность их, соответствующая различным значениям p_y и p_z, полностью характеризует закон дисперсии.

Во-вторых, можно фиксировать значение энергии в l-той зоне, полагая

$$E_l(\mathbf{p}) = \text{const}. \tag{6.1}$$

Уравнение (6.1) определяет поверхность в трехмерном пространстве квазиимпульсов. Ее называют изоэнергетической (или поверхностью

*) Название связано с волновыми представлениями об электроне. Действительно, в силу соотношений де Бройля закон дисперсии есть не что иное, как зависимость частоты электронной волны от квазиволнового вектора.

равных энергий). Придавая различные значения константе, стоящей в правой части (6.1), и задавая форму соответствующей изоэнергетической поверхности, мы полностью описываем закон дисперсии.

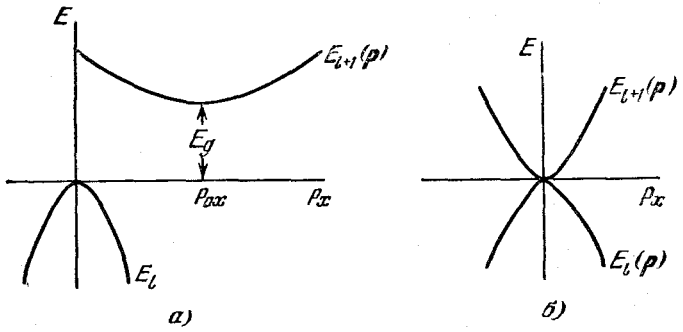


Рис. 3.4. Дисперсионные кривые (схематически): *а*) для двух неперекрывающихся зон; *б*) для смыкающихся зон.

Изоэнергетические поверхности обладают некоторыми свойствами симметрии. Прежде всего, согласно (4.4) они имеют центр симметрии, расположенный в точке $\mathbf{p} = 0$: при изменении знаков у всех компонент $\mathbf{p}$ изоэнергетическая поверхность переходит сама в себя. Этот факт является совершенно общим — он не связан ни со структурой данной кристаллической решетки, ни с видом силового поля в ней. Дальнейшие свойства симметрии навязываются структурой решетки: помимо центра симметрии, изоэнергетическая поверхность обладает и всеми элементами симметрии данной кристаллической решетки.

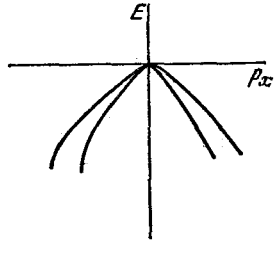


Рис. 3.5. Пример вырождения зон в точке максимума.

Рассмотрим возможный вид дисперсионных кривых и изоэнергетических поверхностей вблизи точек экстремума в зонах. На рис. 3.4, *а* схематически изображены кривые дисперсии для случая неперекрывающихся зон. Для определенности принято, что минимум $(l+1)$-й и максимум l-той зоны достигаются, соответственно, в точках $(p_{x0}, 0, 0)$ и $(0, 0, 0)$, причем, вообще говоря, $p_{0x} \neq 0$; энергия отсчитывается от потолка нижней зоны.

Дисперсионные кривые для случая, когда зоны смыкаются в центре зоны Бриллюэна, схематически представлены на рис. 3.4, *б*. Здесь одно и то же значение энергии $E_l(0) = E_{l+1}(0) = 0$ попадает сразу в две зоны. Иначе говоря, одному и тому же собственному значению энергии $E = 0$ принадлежат две собственные функции: $u_{0,l}$ и $u_{0,l+1}$, т. е. имеет место вырождение. Соответственно

точка, в которой смыкаются зоны, называется *точкой вырождения*, а сами зоны — *вырожденными* в этой точке. Точка вырождения есть особая точка функции E (**p**): в ней смыкаются две ветви E_l (**p**) и E_{l+1} (**p**), т. е. это есть точка ветвления 2-го порядка.

Возможен и более сложный случай, когда в точке вырождения смыкается большее число зон. Мы имеем тогда точку ветвления более высокого порядка. Может случиться также, что в точке вырождения обе зоны (или большее их число) имеют максимум (или минимум). Пример такой ситуации представлен на рис. 3.5.

Из рассмотрения вырожденных зон (рис. 3.4, *б*) ясен смысл условия (4.2): как видно из рис. 3.4, *б*, только в этом случае равенство (4.4) оказывается справедливым для каждой зоны в отдельности *).

§ 7. Металлы и полупроводники

Представление о разрешенных и запрещенных зонах в сочетании с принципом Паули позволяет выяснить причину глубокого различия между металлами и полупроводниками, влекущую за собой различие в наблюдаемых электрических и других их свойствах. Для этой цели рассмотрим, как распределяются электроны по зонам в предельном случае температуры, стремящейся к абсолютному нулю.

Как мы знаем, объем зоны Бриллюэна составляет $(2\pi\hbar)^3/V_0$, где V_0 — объем элементарной ячейки. С другой стороны, дозволенные значения компонент квазиимпульса даются формулами (3.10′). Из них следует, что каждому дозволенному вектору **p** — точке в зоне Бриллюэна — в обратной решетке соответствует ячейка объема $(2\pi\hbar)^3/V$, где V — объем кристалла. Таким образом, зона Бриллюэна содержит всего

$$\frac{(2\pi\hbar)^3}{V_0} \Big/ \frac{(2\pi\hbar)^3}{V} = \frac{V}{V_0} = G$$

состояний с разными квазиимпульсами — ровно столько, сколько элементарных ячеек имеется в данном кристалле. С учетом спина это означает, что полное число состояний в зоне Бриллюэна равно $2G$. Согласно § 4 это же есть и число дозволенных квантовых состояний, содержащихся в каждой отдельной энергетической зоне. По принципу Паули каждое из этих состояний может быть занято только одним электроном. Пусть в элементарной ячейке данного кристалла содержится z электронов. Тогда общее число электронов в решетке есть zG. При достаточно низкой температуре они займут

*) При доказательстве соотношения (4.4) мы воспользовались тем обстоятельством, что вещественная величина E_l (**p**) не изменяется при переходе к сопряженному уравнению. В точке вырождения, однако, заранее не ясно, к каким именно зонам следует относить одинаковые значения энергии E (**p**) и E (—**p**).

первые zG энергетических уровней. При этом имеются три возможности:

а) самая верхняя зона, содержащая электроны, занята не полностью, а содержит и вакантные уровни;

б) верхняя зона, содержащая электроны, заполнена целиком, но смыкается (или перекрывается) со следующей — более высокой — зоной;

в) верхняя, целиком заполненная зона отделена от следующей — пустой — зоны запрещенным участком конечной величины E_g.

Случай а) заведомо реализуется, если число z нечетное. Действительно, число заполненных зон, очевидно, равно $\frac{zG}{2G} = \frac{z}{2}$. При нечетном z это число не целое.

Из сказанного, однако, не следует, что при четном z непременно должен иметь место случай в): различные энергетические зоны могут перекрываться. Выяснить (не апеллируя к опыту), будет ли это иметь место фактически, можно, только явно решая уравнение (4.1). При этом, однако, не обязательно рассматривать все электроны кристалла. Действительно, электроны полностью заполненных внутренних атомных оболочек заведомо заполняют зоны, возникающие в результате размытия соответствующих атомных уровней. Поэтому под z часто можно понимать число только валентных электронов.

Электрические свойства вещества в первых двух и третьем случаях оказываются принципиально различными. Действительно, поместим кристалл в постоянное внешнее электрическое поле. Чтобы вызвать электрический ток, поле должно сообщить электронам ускорение, т. е. произвести над ними некоторую работу. Иначе говоря, поле должно перевести электроны на высшие энергетические уровни. Очевидно, в слабом поле эти уровни должны быть близки к заполненным. С другой стороны, сами они должны быть свободны: иначе, по принципу Паули, переход электронов на них окажется невозможным. Из сказанного явствует, что в случаях а) и б) сколь угодно слабое внешнее поле создаст электрический ток — вещество в этом смысле ведет себя как металл. Ясно также, что в случае в) слабое поле тока вызвать не может: электронам некуда переходить *). Здесь для создания тока нужна предварительная «активация»: часть электронов должна быть переброшена в незаполненную зону за счет энергии теплового движения, света и т. д. Такой кристалл обладает свойствами полупроводника (в предельном случае очень больших значений E_g — изолятора). Наивысшая из заполненных (при $T = 0$) зон называется валентной (ей приписывают индекс $l = v$), наинизшая из незанятых — зоной проводимости (индекс $l = c$). Верхнюю границу (потолок) валентной зоны принято

*) В достаточно сильном поле оказывается возможным «просачивание» электронов через запрещенную зону (см. § IV.6). Для этого, однако, нужны поля с напряженностью порядка 10^6 В/см.

обозначать через E_v, а нижнюю границу (дно) зоны проводимости — через E_c.

Пример, когда заведомо реализуется случай а), дают нам кристаллы щелочных металлов, содержащих по одному валентному s-электрону на атом. Менее тривиальный пример представляют кристаллы щелочноземельных элементов, содержащих по два валентных s-электрона на атом. Металлические свойства здесь обусловлены перекрытием зон, возникающих в результате размытия s- и p-уровней электронов в атомах (реализуется случай б)).

Наконец, в кристаллах германия и кремния реализуется случай в). Валентные оболочки свободных атомов этих элементов обладают электронной конфигурацией s^2p^2 (четыре валентных электрона — по два в s- и p-состояниях).

Разности энергий s- и p-уровней сравнительно невелики; поэтому при образовании кристаллов происходит гибридизация соответствующих состояний (ср. § II.2 и § 5). В результате волновые функции валентной зоны (в приближении сильной связи) строятся из одной s- и трех p-функций. Такая зона содержит по четыре состояния на атом — ровно столько, сколько там есть валентных электронов. При $T=0$ она полностью заполнена. Следующая разрешенная зона отделена от валентной запрещенным участком. Так же обстоит дело и в алмазе.

Заметим, что наличие запрещенной зоны в этом и в других случаях обусловлено не геометрическими (как форма зоны Бриллюэна), а динамическими особенностями данного кристалла — конкретным видом и величиной потенциальной энергии электрона. В принципе можно, изменяя ее путем, например, изменения межатомных расстояний (под давлением), добиться перекрытия валентной зоны и зоны проводимости. При этом рассматриваемое вещество превратится из полупроводника в металл. Возможен и обратный случай — возникновение запрещенной зоны и переход металла в полупроводник при достаточно сильном изменении давления на образец.

Из сказанного выше ясно, что электроны заполненных зон не принимают непосредственного участия в переносе постоянного тока, а также в переносе энергий. Они образуют фон, на котором движутся электроны проводимости и дырки в валентной зоне. В дальнейшем мы будем называть электроны проводимости просто электронами (если это не поведет к путанице).

§ 8. Эффективная масса

Закон дисперсии дает довольно полную информацию о поведении электрона в данной решетке. Однако его трудно определять из опыта, ибо в общем случае речь идет здесь о нахождении явного вида функции трех практически непрерывно изменяющихся аргу-

ментов — компонент квазиимпульса. К счастью, в большинстве задач физики полупроводников играет роль лишь сравнительно малая часть зоны Бриллюэна, отвечающая энергиям у дна или у потолка той или иной энергетической зоны. Как мы сейчас увидим, для описания закона дисперсии вблизи экстремумов зоны нет необходимости определять всю функцию $E_l(\mathbf{p})$ — достаточно задать лишь несколько констант.

Пусть экстремум l-той зоны лежит в точке $\mathbf{p} = \mathbf{p}_{0l}$ и зона в ней не вырождена. Тогда вблизи точки экстремума функцию $E_l(\mathbf{p})$ можно разложить в ряд Тэйлора. Принимая обычное соглашение о суммировании по дважды повторяющимся векторным индексам, мы имеем (опуская для краткости индекс l при p_0)

$$E_l(\mathbf{p}) = E_l(\mathbf{p}_0) + \frac{1}{2} m^{-1}_{\alpha\beta,\, l}(p_\alpha - p_{\alpha,\,0})(p_\beta - p_{\beta,\,0}) + \ldots \qquad (8.1)$$

Здесь

$$m^{-1}_{\alpha\beta,\, l} = \frac{\partial^2 E_l(\mathbf{p})}{\partial p_\alpha \partial p_\beta}\Big|_{\mathbf{p}=\mathbf{p}_0} \qquad (8.2)$$

— величины размерности обратной массы, а многоточием обозначены члены более высокого порядка, которыми в окрестности экстремума можно пренебречь. Слагаемые, линейные по $\mathbf{p} - \mathbf{p}_0$, отсутствуют в правой части (8.1) по определению точки $\mathbf{p}_0$.

Так как значение второй производной не зависит от порядка дифференцирования, величины $m^{-1}_{\alpha\beta,\, l}$ образуют симметричный тензор второго ранга. Значения $m_{\alpha\beta,\, l}$ зависят от выбора системы координат в $\mathbf{p}$-пространстве. В частности, можно выбрать ее так [М5], чтобы компоненты $m^{-1}_{\alpha\beta,\, l}$ при $\alpha \neq \beta$ обратились в нуль. Соответствующие оси координат называются *главными*, а компоненты $m^{-1}_{\alpha\beta,\, l}$ в этой системе координат — главными значениями данного тензора. Для краткости мы будем писать при них лишь один значок α; следует, однако, помнить, что это — компоненты тензора, а не вектора.

Равенство (8.1), будучи отнесено к главным осям, принимает вид

$$E_l(\mathbf{p}) = E_l(\mathbf{p}_0) + {}^1/_2 \sum_{\alpha = x,\, y,\, z} m^{-1}_\alpha (p_\alpha - p_{\alpha,\,0})^2. \qquad (8.1')$$

При этом $m_\alpha > 0$ или $m_\alpha < 0$, если рассматриваемый экстремум есть, соответственно, минимум или максимум.

Поскольку величина $E_l(\mathbf{p}_0)$ при заданном индексе l есть константа, уравнение (6.1) в этом случае принимает вид

$$\frac{(p_x - p_{x,\,0})^2}{2m_x} + \frac{(p_y - p_{y,\,0})^2}{2m_y} + \frac{(p_z - p_{z,\,0})^2}{2m_z} = \text{const}. \qquad (8.3)$$

(Здесь и в дальнейшем мы опускаем для краткости зонный индекс l.) Таким образом, изоэнергетические поверхности вблизи невырож-

денного экстремума представляют собой эллипсоиды с полуосями, пропорциональными $\sqrt{|m_x|}$, $\sqrt{|m_y|}$ и $\sqrt{|m_z|}$ *), и с центром в точке $\mathbf{p} = \mathbf{p}_0$. Выбранные нами оси координат суть главные оси этого эллипсоида.

Константы m_x, m_y и m_z не всегда независимы: они могут быть связаны условиями симметрии изоэнергетической поверхности. Рассмотрим, например, кристалл кубической системы. Тогда, как указано в § 6, изоэнергетическая поверхность должна обладать симметрией куба. В частности, она должна переходить сама в себя при поворотах на 90° вокруг каждой из взаимно перпендикулярных осей, проходящих через центр куба и центры его граней; сами эти оси физически эквивалентны.

Здесь могут представиться два случая:

1) Экстремум функции $E(\mathbf{p})$ лежит в центре зоны Бриллюэна: $\mathbf{p}_0 = 0$. Таковы, например, изоэнергетические поверхности в зоне проводимости InSb, GaSb, GaAs и в некоторых других соединениях: дно этой зоны отвечает там точке $\mathbf{p} = 0$. Из соображений симметрии ясно, что главные оси тензора $m^{-1}_{\alpha\beta}$ в данном случае суть указанные выше оси куба. При этом $m_x = m_y = m_z \equiv m$ и изоэнергетические поверхности у дна зоны представляют собой сферы с центром в точке $\mathbf{p} = 0$:

$$E(\mathbf{p}) = E_c + \frac{\mathbf{p}^2}{2m}. \tag{8.4}$$

Закон дисперсии (8.4) называют параболическим изотропным.

Для частного случая одномерной цепочки в приближении сильной связи равенство (8.4) уже было получено в § 5 (формула (5.19)). Как и там, величину m в (8.4) называют эффективной массой. В дальнейшем (§ IV.1) мы увидим, что это название имеет не только формальный смысл: величина m определяет и среднее ускорение электрона во внешнем электрическом или магнитном поле.

Величины m_x, m_y, m_z, даже когда они неодинаковы, также называются эффективными массами, а совокупность величин $m^{-1}_{\alpha\beta}$ — *тензором обратной эффективной массы.*

2) Экстремум лежит в точке $\mathbf{p}_0 \neq 0$, при этом вектор $\mathbf{p}_0$, соединяющий данный экстремум с центром зоны Бриллюэна, направлен вдоль одной из осей симметрии кристалла. Из свойств симметрии изоэнергетической поверхности следует, что в этом случае должно существовать несколько физически эквивалентных экстремумов, расположенных в соответствующих точках на эквивалентных осях симметрии. Соответственно надо говорить о нескольких векторах $\mathbf{p}_0$ и нескольких эллипсоидах вида (8.3); мы будем нумеровать их верхним латинским индексом i. Изоэнергетическая поверхность

*) В случае максимума, когда $m_\alpha < 0$, константа в правой части (8.3) отрицательна: это есть энергия, отсчитанная от своего максимума.

в данном случае представляет собой совокупность всех этих эллипсоидов. Таковы, например, изоэнергетические поверхности в зонах проводимости германия и кремния. В первом из указанных материалов дну зоны проводимости отвечают точки, лежащие на объемных диагоналях куба (на оси [*111*] и ей эквивалентных в обратной решетке); во втором — точки, расположенные на оси [*100*] и ей эквивалентных.

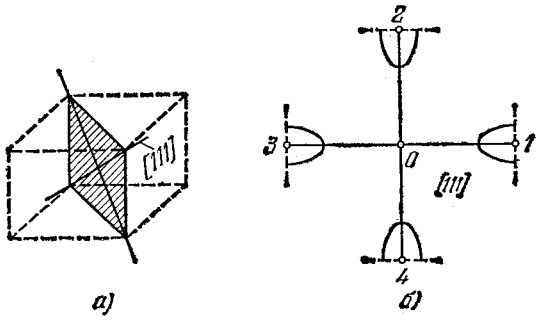


Рис. 3.6. *а*) Две (из четырех) пространственные диагонали куба [*111*]. *б*) Положение изоэнергетических поверхностей электронов в зоне Бриллюэна (Ge).Сечение плоскостью (*110*). Пунктир — границы зоны Бриллюэна.

На рис. 3.6 и 3.7 схематически изображены сечения электронных изоэнергетических поверхностей в германии и кремнии плоскостями $p_z = 0$. Вблизи каждого минимума изоэнергетическая поверхность здесь имеет вид эллипсоида (8.3). Заметим, что это не противоречит кубической симметрии кристалла. Действительно, при поворотах на 90° вокруг трех взаимно перпендикулярных осей, проходящих через какую-нибудь из точек $\mathbf{p}_0^i$, вся изоэнергетическая поверхность отнюдь не обязана переходить сама в себя. Иначе говоря, симметрия вблизи каждой такой точки ниже кубической. Из рис. 3.6 и 3.7 видно, однако, что в рассматриваемых случаях физически эквивалентными оказываются любые две взаимно перпендикулярные оси, лежащие в плоскостях, перпендикулярных

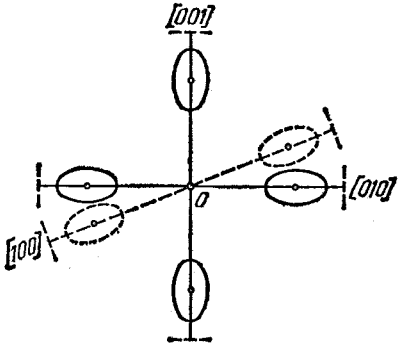


Рис. 3.7. Положение изоэнергетических поверхностей электронов в зоне Бриллюэна (Si). Пунктир — границы зоны.

векторам $\mathbf{p}_0^i$. Очевидно, эти оси вместе с $\mathbf{p}_0^i$ и будут главными осями тензора $m_{\alpha\beta}^i$ для данного эллипсоида. Примем ось, вдоль которой направлен вектор $\mathbf{p}_0^i$, за ось z, а остальные две — за оси x, y. В соответствии со сказанным выше

$$m_x = m_y \equiv m_\perp, \qquad m_z \equiv m_\parallel \neq m_\perp.$$

Таким образом, в рассматриваемом случае изоэнергетическая поверхность вблизи каждого экстремума имеет вид эллипсоида вращения:

$$E(\mathbf{p}) = E_c + \frac{(p_x - p_{x,0}^i)^2}{2m_\perp} + \frac{(p_y - p_{y,0}^i)^2}{2m_\perp} + \frac{(p_z - p_{z,0}^i)^2}{2m_\parallel}. \qquad (8.5)$$

Подчеркнем, что выражение (8.5) имеет смысл лишь вблизи каждого отдельного экстремума. Закон дисперсии вида (8.5) (или, общее, (8.1′) при не совпадающих друг с другом эффективных массах) называют параболическим анизотропным.

Области энергии вблизи каждого минимума иногда называют долинами, а полупроводники с несколькими эквивалентными минимумами энергии — многодолинными. В этих материалах имеется несколько разных групп носителей заряда. Все такие группы имеют одинаковые минимумы энергии, и в этом смысле они эквивалентны. В условиях термодинамического равновесия носители заряда распределяются поровну между разными долинами. Однако эффективные массы, а следовательно, и подвижности носителей заряда при данном направлении движения, вообще говоря, различны для разных долин. Наблюдаемая на опыте подвижность (и другие подобные величины) есть некоторая результирующая величина. В кристаллах кубической симметрии она оказывается изотропной вследствие кубической симметрии расположения эллипсоидов и равного числа электронов вблизи каждого из них.

Эквивалентность различных долин может быть нарушена при некоторых внешних воздействиях. Это имеет место, например, при одностороннем сжатии, когда энергия носителей заряда в одних долинах увеличивается, а в других — уменьшается. В результате изменяется распределение носителей между долинами, что приводит к изменению результирующей подвижности и проявляется в зависимости электрического сопротивления от давления (пьезосопротивление). Эквивалентность различных долин нарушается также при воздействии сильных внешних электрических и магнитных полей, что приводит к ряду интересных явлений. Некоторые из них будут рассмотрены ниже.

Экстремумы функции $E(\mathbf{p})$ могут располагаться как внутри зоны Бриллюэна, так и на ее границах. В германии, видимо, реализуется вторая возможность: дну зоны проводимости отвечают

точки L (рис. 3.1). При этом точки *1* и *3*, *2* и *4* на рис. 3.6 отвечают одному и тому же состоянию, и соответствующие им минимумы надо рассматривать как один: в каждом из эллипсоидов, сечения которых изображены на рис. 3.6, сохранена только половина, лежащая в пределах первой зоны Бриллюэна (ср. ниже рис. 3.9). Таким образом, фактически на рис. 3.6 изображены два, а не четыре эквивалентных экстремума. На трехмерной картине, отвечающей кристаллу германия, имеются четыре эквивалентных экстремума (по числу диагоналей куба), а не восемь. С другой стороны, в кремнии, видимо, реализуется первая возможность: дну зоны проводимости отвечают точки Δ. Соответственно на рис. 3.7 изображены четыре эквивалентных экстремума. На трехмерной картине, отвечающей кристаллу кремния, их имеется шесть.

Итак, вид изоэнергетической поверхности вблизи невырожденной точки экстремума определяется не более чем тремя константами. В частных случаях, когда две или все три эффективные массы оказываются одинаковыми в силу свойств симметрии изоэнергетической поверхности, число констант соответственно уменьшается.

Если точка экстремума зоны одновременно является и точкой вырождения, то положение несколько усложняется. Изоэнергетические поверхности тогда уже не являются эллипсоидами; форма их зависит, в частности, от кратности вырождения и от свойств симметрии решетки. Нетрудно написать, однако, общее выражение для функции $E(\mathbf{p})$ вблизи экстремума в практически важном случае, когда вырождение двукратное и экстремум энергетической зоны лежит в центре зоны Бриллюэна. Действительно, искомое выражение должно:

а) обладать двумя ветвями, которые смыкаются в точке $\mathbf{p} = 0$;

б) иметь нулевые первые производные в точке $\mathbf{p} = 0$;

в) содержать $|\mathbf{p}|$ в степени не выше второй.

В кристалле кубической системы единственную непрерывную функцию, удовлетворяющую поставленным условиям и обладающую должными свойствами симметрии, можно записать в виде

$$E(\mathbf{p}) = E(0) + \frac{1}{2m_0}\left\{Ap^2 \pm \sqrt{B^2p^4 + C^2\left(p_x^2p_y^2 + p_y^2p_z^2 + p_z^2p_x^2\right)}\right\}, \qquad (8.6)$$

где A, B, C — безразмерные скаляры *); при этом A и B вещественны, а C может быть и мнимым (но $C^2 > -3B^2$).

Тензор обратной эффективной массы $m_{\alpha\beta}^{-1}$ можно ввести и в случае закона дисперсии (8.6). Однако теперь его компоненты уже не постоянны, а зависят от направления вектора $\mathbf{p}$. Двум ветвям функ-

*) Условие того, чтобы точка $\mathbf{p} = 0$ была экстремальной, а не седловой, сводится к неравенству $A^2 > B^2 + {}^1/_3\, C^2$ (если $C^2 > 0$) или $A^2 > B^2$ (если $C^2 < 0$).

ции (8.6) отвечают разные значения компонент $m^{-1}_{\alpha\beta}$. Соответственно говорят о зонах «легких» и «тяжелых» носителей заряда (обычно дырок — см. ниже § 9). Изоэнергетические поверхности, изображаемые формулой (8.6), называются *гофрированными сферами*. На рис. 3.8 изображено их сечение плоскостью (*100*).

Приводить тензор $m^{-1}_{\alpha\beta}$ к главным осям в данном случае не имеет смысла: поскольку компоненты его зависят от $\mathbf{p}$, такое приведение можно сделать, вообще говоря, лишь в нескольких точках зоны Бриллюэна, переходящих друг в друга при преобразованиях симметрии.

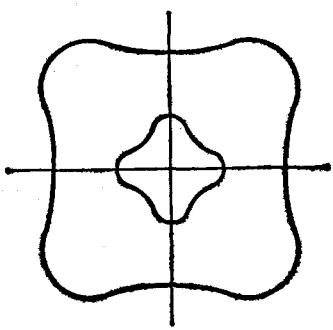


Рис. 3.8. Сечение гофрированной сферы (8.6) плоскостью (*100*). Внешний контур соответствует «тяжелым» носителям заряда, внутренний — «легким».

В ряде случаев отклонение поверхности (8.6) от сферической оказывается не очень большим. Тогда правую часть (8.6) можно приближенно заменить средним ее значением на сфере радиуса $|\mathbf{p}|$. При этом мы получаем две параболические зоны с изотропными законами дисперсии и разными эффективными массами m_1 и m_2.

Вырождение зон встречается довольно часто. Как указывалось в § 5, его можно ожидать, если зоны получаются в результате размытия вырожденных энергетических уровней электронов в атомах. Согласно § 7 такая картина может иметь место, например, в валентных зонах германия и кремния.

По определению, выражения вида (8.1) и т. п. справедливы лишь в окрестности данной точки экстремума. При удалении от нее начинают играть роль и более высокие члены разложения функции $E(\mathbf{p})$ по степеням $p_\alpha - p_{\alpha,0}$. Об этом говорят как о непараболичности закона дисперсии.

Для оценки условий, при которых закон дисперсии начинает заметно отличаться от параболического, следует решить уравнение (4.1). Оказывается, что непараболичность особенно заметна в полупроводниках с достаточно узкой запрещенной зоной. Законы дисперсии для некоторых таких материалов выписаны в § 9.

§ 9. Зонная структура некоторых полупроводников

Структуру энергетических зон в полупроводнике можно было бы рассчитать с помощью уравнения (4.1). Фактически, однако, при этом возникают весьма серьезные вычислительные трудности, для преодоления которых необходимы самые мощные современные электронно-вычислительные машины. Практически чаще используется комбинированный подход: такие параметры, как ширина

запрещенной зоны, эффективные массы и т. д., определяют из опыта, а методы квантовой механики применяют для того, чтобы получить предварительную ориентировочную информацию о возможном расположении максимумов и минимумов функций $E_l(\mathbf{p})$ в зоне Бриллюэна. Важную роль при этом играют соображения, связанные с типом химической связи, а также соображения симметрии.

Действительно, в точке экстремума $\mathbf{p} = \mathbf{p}_0$ с необходимостью должно выполняться условие

$$\nabla_{\mathbf{p}} E_l(\mathbf{p}) = 0. \tag{9.1}$$

Обращение левой части (9.1) в нуль может произойти либо случайно — в результате взаимной компенсации слагаемых в выражении типа (5.17), — либо в силу симметрии. Так как энергия, согласно (4.4), есть четная функция $\mathbf{p}$, то $\nabla_{\mathbf{p}} E_l(\mathbf{p})$ есть функция нечетная. Следовательно, она, будучи ограниченной, должна обращаться в нуль при $\mathbf{p} = 0$: в центре зоны Бриллюэна энергия электрона в кристалле любого класса непременно имеет либо максимум, либо минимум, либо точку перегиба. В зависимости от типа симметрии кристалла функция $E_l(\mathbf{p})$ может оказаться четной (а $\nabla_{\mathbf{p}} E_l(\mathbf{p})$ — нечетной) и относительно некоторых других точек в зоне Бриллюэна.

Таким образом, исследование свойств симметрии изоэнергетических поверхностей позволяет заранее определить, где именно надо искать экстремумы зон. При этом, как мы видели в § 8, удается определить направления главных осей тензора обратной эффективной массы и число различных главных его значений.

Способы экспериментального определения параметров зонного спектра основаны на изучении взаимодействия полупроводников с электромагнитными волнами, а также на исследовании явлений переноса в магнитном или упругом поле (измеряется изменение электрического сопротивления в магнитном поле или при деформации образца) (см. § IV.3 и гл. XVIII).

Исследуя особенности этих явлений, зависящие от расположения и формы изоэнергетических поверхностей, и сопоставляя затем результаты расчетов с опытом, можно определить искомые параметры: величины m_x, m_y, m_z (или A, B, C), точки $\mathbf{p}_0^i$ и т. д. Ниже суммируются экспериментальные данные для некоторых важных полупроводниковых материалов *).

На рис. 3.9 схематически изображены основные черты зонной структуры германия, кремния, арсенида галлия и антимонида индия **). В германии и кремнии в центре зоны Бриллюэна смы-

*) Пользуясь этими данными, следует помнить, что они могут изменяться с течением времени — по мере уточнения эксперимента.

**) По книге *O. Madelung*, Grundlagen der Halbleiterphysik, Springer-Verlag, New York, 1970.

каются две валентные зоны (с энергиями E_1 (p) и E_2 (p)); на величину Δ (при $p = 0$) ниже располагается (не показанная на рисунке) третья валентная зона (с энергией E_3 (p)), отщепленная от первых двух за счет спин-орбитального взаимодействия. Каждая из этих зон двукратно вырождена. Законы дисперсии в двух верхних даются формулой (8.6), закон дисперсии в нижней зоне близок к изотропному параболическому.

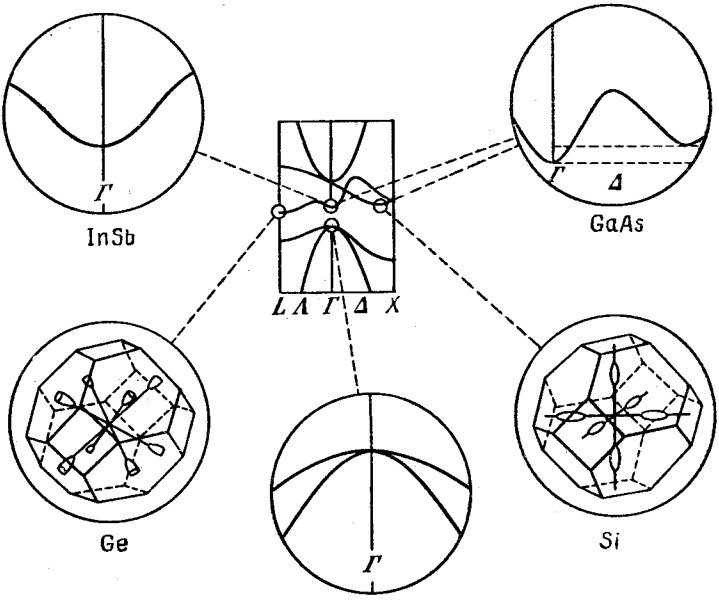


Рис. 3.9. Зонная структура некоторых полупроводников. В центре схематически изображены дисперсионные кривые; кружками на них отмечены точки, отвечающие экстремумам зон. Вверху слева и справа схематически представлены дисперсионные кривые, описывающие зоны проводимости в InSb и GaAs, внизу в центре — дисперсионные кривые для валентных зон в германии и кремнии. Внизу слева и справа схематически изображены изоэнергетические поверхности в зонах проводимости германия и кремния. Пунктирные линии, проведенные от кружков на центральном рисунке к боковым рисункам, указывают, какому материалу соответствует данный экстремум.

Как мы увидим в гл. IV, при рассмотрении валентных зон удобно пользоваться представлением о положительно заряженных свободных дырках. Энергии последних отсчитываются в направлении, противоположном направлению отсчета энергии электронов (на рис. 3.10 (см. ниже) и ему подобных — вниз). Таким образом, в рассматриваемом случае функции $-E_1$ (p) и $-E_2$ (p) описывают зоны тяжелых и легких дырок.

Законы дисперсии вблизи дна зоны проводимости в германии и кремнии рассматривались в § 8. Мы имеем здесь совокупность

эллипсоидов вида (8.5) с центрами в точках L (германий) и Δ (кремний). Заметим, однако, что при удалении от указанных экстремумов зависимость энергии от квазиимпульса носит не монотонный характер: в обоих материалах имеются побочные минимумы в центрах зоны Бриллюэна. В германии расстояние между побочным минимумом и потолком валентной зоны составляет 0,80 эВ (при комнатной температуре).

В таблице 3.1 приведены некоторые параметры энергетических зон германия и кремния *).

В арсениде галлия, антимониде индия и в других соединениях типа $A^{III}B^{V}$ валентные зоны в отсутствие спин-орбитального взаимодействия тоже смыкались бы в центре зоны Бриллюэна. Однако, в отличие от германия и кремния, спин-орбитальное взаимодействие приводит в этих соединениях не только к отщеплению дважды вырожденной подзоны E_3 (р), но и к снятию вырождения во всех трех подзонах. Исключения составляют направления [*100*] и [*111*] в зонах E_2 и E_3 и направление [*100*] в зоне E_1. На рис. 3.10 показано, во что превращаются дисперсионные кривые E_1 (р) и E_2 (р) в InSb. Видно, в частности, что максимум энергии в верхней подзоне смещается из центра зоны Бриллюэна в точки, расположенные на осях [*111*] и ей эквивалентных. Соответствующие изоэнергетические поверхности должны, строго говоря, иметь вид эллипсоидов типа (8.5) с осью [*111*] (или ей эквивалентными) в качестве оси вращения. Заметим

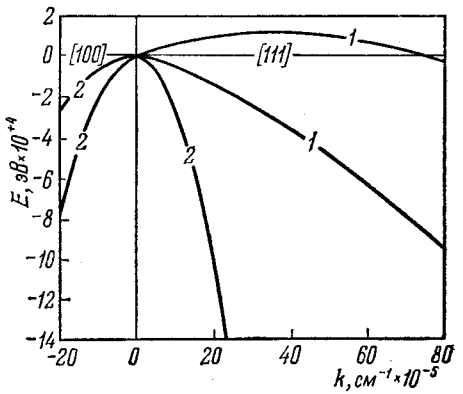


Рис. 3.10. Законы дисперсии для верхних валентных подзон в InSb. По оси абсцисс отложены значения квазиволнового вектора $k = p/\hbar$ в направлениях [*111*] (справа) и [*100*] (слева). Цифры у кривых указывают степени вырождения зон. Две верхние кривые в правой части рисунка отвечают зоне тяжелых дырок, нижние — зоне легких дырок.

*) Все данные относятся к гелиевым температурам. При повышении температуры ширина запрещенной зоны обычно уменьшается; в не слишком широкой области температуры это уменьшение аппроксимируется линейным законом (с угловым коэффициентом порядка $2 \div 3 \cdot 10^{-4}$ эВ/град). При $T = 300$ К в германии $E_g = 0{,}67$ эВ.

Значения эффективных масс и параметров A, B и C для германия взяты из работы *B. Levinger, D. Frankl,* Journ. Phys. Chem. Sol. **20**, 281 (1961), для кремния — из работ *J. Hensel, G. Feher,* Phys, Rev. **129**, 1041 (1963) (верхняя строчка) и *J. Balslev, P. Lawaetz,* Phys. Lett. **19**, 6 (1955) (нижняя строчка). Через m_1 и m_2 обозначены средние эффективные массы «тяжелых» и «легких» дырок (§ 8).

Таблица 3.1

Зонные параметры германия и кремния

Материал	E_g, эВ	Ширина запрещенной зоны в точке $\mathbf{p}=0$	$\frac{m_\parallel}{m_0}$	$\frac{m_\perp}{m_0}$	A	B	C	$\frac{m_1}{m_0}$	$\frac{m_2}{m_0}$
Германий	0,744	0,898	1,588	0,0815	−10,27	+8,63	12,4	0,30	0,044
Кремний	1,15	3,4	0,9163	0,1905	−4,98 −4,27	+0,75 +0,63	5,25 5,03	0,50	0,17

все же, что максимум $E_1(\mathbf{p}_0)$ очень близок к $E_1(0)$:

$$|E_1(\mathbf{p}_0)-E_1(0)|\simeq 1{,}0\cdot 10^{-4}\text{ эВ (при 1,5К) *).}$$

Величина $\mathbf{p}_0$ также составляет лишь малую (примерно 0,007) долю расстояния от центра зоны Бриллюэна до ее границы в данном направлении. Подобная же картина наблюдается и в других материалах этой группы.

По этой причине в них часто можно пренебречь снятием вырождения валентных подзон. При этом для описания двух верхних валентных подзон можно с разумным приближением воспользоваться формулой (8.6). Следует, однако, иметь в виду, что в числе соединений типа $A^{III}B^{V}$ есть и сравнительно узкозонные — антимонид и арсенид индия (см. табл. 3.2) **). Эффекты непараболичности в них проявляются уже при сравнительно небольших значениях квазиимпульса. То же относится и к зонам проводимости этих материалов.

Последовательный учет особенностей полупроводника с узкой запрещенной зоной, разделяющей две невырожденные разрешенные зоны, приводит к следующему выражению для закона дисперсии в зонах проводимости InSb и InAs ***):

$$E(\mathbf{p})=E(0)+\frac{p^2}{2m_0}-\frac{E_g}{2}\left\{1-\left[1+\frac{4p^2\mathscr{P}^2}{3m_0E_g^2}\left(2+\frac{E_g}{E_g+\Delta}\right)\right]^{1/2}\right\}. \quad (9.2)$$

Здесь $\mathscr{P}$ и Δ — параметры размерности импульса и энергии. Роль нижней зоны играет здесь зона легких дырок. Модель зонной структуры полупроводника, для которой справедлива формула (9.2),

*) Число взято из работы *C. Pidgeon, S. Groves,* Phys. Rev. Lett. **20**, 1003 (1968).

) Таблица составлена по данным работ *D. Bagguley, M. Robinson, R. Stradling,* Phys, Rev. Lett. **6, 143 (1963); *R. Stradling,* Phys. Lett. **20**, 217 (1966).

***) Вывод формулы (9.2) можно найти в книгах [M8] и [4].

Таблица 3.2

Зонные параметры материалов типа $A^{III}B^{V}$

	InSb	InAs	InP	GaSb	GaAs	GaP	AlSb
Точка в зоне Бриллюэна, в которой расположен основной минимум	Г	Г	Г	Г	Г	на оси [100]	на оси [100]
Эффективная масса у дна зоны проводимости (в единицах m_0)	0,012 (300 К)	0,023—0,027	0,077	0,047	0,043—0,071	0,34	0,39
A	$-25\pm2{,}5$			$-11{,}0\pm0{,}6$			
B	$+25\pm2{,}1$			$+6{,}0\pm1{,}5$			
C	$16\pm1{,}6$			11 ± 4			

называется моделью Кейна. Формула (9.2) справедлива при $E(p) - E(0) \ll E_g + {}^2/_3\,\Delta$; последнее условие имеет довольно широкую область применимости. Заметим, что закон дисперсии (9.2) изотропен: $E = E(|\mathbf{p}|)$. Соответствующие изоэнергетические поверхности суть сферы. При

$$\frac{p^2\mathscr{P}^2}{3m_0^2E_g^2} \ll 1 \tag{9.3}$$

соотношение (9.2) принимает вид (8.4), причем эффективная масса на дне зоны проводимости дается выражением

$$m = \frac{m_0}{1 + \dfrac{4\mathscr{P}^2}{3m_0E_g}\left(1 + \dfrac{E_g}{2(E_g+\Delta)}\right)}. \tag{9.4}$$

В более широкозонных материалах — арсениде и антимониде галлия, фосфиде индия и др. — эффекты непараболичности менее существенны, и закон дисперсии у дна зоны проводимости хорошо аппроксимируется формулой (8.4). Как и в германии и кремнии, в зонах проводимости этих веществ имеются и побочные минимумы. Так, в арсениде галлия имеется шесть эквивалентных минимумов, расположенных в точках Δ, т. е. на оси [100] и ей эквивалентных. Они отстоят от дна зоны на 0,34 эВ.

В таблице 3.2 приведены некоторые параметры энергетических зон в соединениях типа $A^{III}B^{V}$. Данные относятся к гелиевым температурам (если не приведена другая).

Параметр E_g в формуле (9.2) не обязательно строго положителен. Он может оказаться и равным нулю или даже отрицательным.

При $E_g = 0$ зоны проводимости и валентная смыкаются в точке $p = 0$, и для возникновения пары свободных носителей заряда тепловая активация не требуется. Соответственно концентрация носителей (а с ней и электропроводность вещества) оказывается отличной от нуля при сколь угодно низких температурах, как в металлах. Однако эта концентрация все же гораздо меньше, чем в обычных металлах. Поэтому такие вещества называются *полуметаллами*. К числу их относится, например, серое олово. При $E_g < 0$ валентная зона и зона проводимости перекрываются. Пока это перекрытие не слишком велико, рассматриваемое вещество также оказывается полуметаллом. Видимо, так обстоит дело в теллуриде и селениде ртути, а также в ряде других соединений.

Глава IV

ЭЛЕМЕНТЫ ЗОННОЙ ТЕОРИИ ТВЕРДОГО ТЕЛА II. КРИСТАЛЛЫ ВО ВНЕШНИХ ПОЛЯХ. НЕИДЕАЛЬНЫЕ КРИСТАЛЛЫ

§ 1. Средние значения скорости и ускорения электрона в кристаллической решетке

Вычислим квантовомеханические средние значения скорости и ускорения электрона в решетке. По общему правилу квантовой механики средняя скорость дается выражением

$$\mathbf{v}_l(\mathbf{p}) = \int \psi^*_{\mathbf{p}l}(\mathbf{r})\, \hat{\mathbf{v}}\, \psi_{\mathbf{p}l}(\mathbf{r})\, d\mathbf{r}, \tag{1.1}$$

где $\hat{\mathbf{v}}$ — оператор скорости и интегрирование ведется по основному объему. По определению

$$\hat{\mathbf{v}} = -i\frac{\hbar}{m_0}\nabla, \tag{1.2}$$

где m_0 — истинная масса электрона, а $-i\hbar\nabla$ — оператор импульса. В качестве $\psi_{\mathbf{p}l}$ в формулу (1.1) надлежит подставить функции Блоха (III.2.15′):

$$\psi_{\mathbf{p}l} = u_{\mathbf{p}l} \exp\left(\frac{i}{\hbar}\mathbf{pr}\right).$$

Таким образом, равенство (1.1) принимает вид

$$\mathbf{v}_l(\mathbf{p}) = \frac{\mathbf{p}}{m_0} - \frac{i\hbar}{m_0}\int u^*_{\mathbf{p}l}\nabla u_{\mathbf{p}l}\, d\mathbf{r}. \tag{1.1′}$$

Здесь принято во внимание условие нормировки (III.2.8).

Как видно из (1.1′), вообще говоря, $\mathbf{v}_l \neq \mathbf{p}/m_0$: как уже отмечалось в § III.2, квазиимпульс $\mathbf{p}$ не совпадает с импульсом. Второе слагаемсе в правой части (1.1′) вычислено в Приложении II. Подставляя формулу (П.II.10) в (1.1′), находим окончательно:

$$\mathbf{v}_l(\mathbf{p}) = \nabla_{\mathbf{p}} E_l(\mathbf{p}). \tag{1.3}$$

В правой части (1.3) стоит градиент от энергии электрона $E_l(\mathbf{p})$ в пространстве квазиимпульсов. Вообще говоря, этот вектор отличен от нуля. Исключение составляют лишь границы энергетических зон, где функция $E_l(\mathbf{p})$ имеет экстремумы, а также точки перегиба функции $E_l(\mathbf{p})$. Следовательно, отлична от нуля и средняя плотность электрического тока $\mathbf{j}_l(\mathbf{p})$, создаваемого одним электроном с квазиимпульсом $\mathbf{p}$ в l-той зоне. Действительно,

$$\mathbf{j}_l(\mathbf{p}) = -e\mathbf{v}_l(\mathbf{p}) = -e\nabla_{\mathbf{p}} E_l(\mathbf{p}). \tag{1.4}$$

Подчеркнем, что отличные от нуля значения $\mathbf{v}_l$ и $\mathbf{j}_l(\mathbf{p})$ получены здесь в предположении об отсутствии каких-либо внешних электрических полей. Это означает, что электрическое сопротивление идеального кристалла равно нулю. Конечное значение сопротивления обусловлено только отклонениями силового поля от идеально периодического в результате тепловых колебаний или наличия каких-либо структурных дефектов решетки.

Формула (1.3) становится особенно наглядной, если для энергии электрона воспользоваться формулой (III.8.4). Тогда (мы вновь опускаем индекс l)

$$\mathbf{v} = \frac{\mathbf{p}}{m}. \tag{1.5}$$

Ясно видны как формальная аналогия между импульсом и квазиимпульсом, так и глубокое принципиальное различие между ними: m — не истинная масса, а эффективная, и формула (1.5) справедлива не при всех $\mathbf{p}$, а лишь в некоторой области зоны Бриллюэна.

Равенство (1.3) можно переписать в несколько ином виде, если воспользоваться соотношениями де Бройля

$$\mathbf{p} = \hbar\mathbf{k}, \quad E = \hbar\omega,$$

где ω — частота электронной волны, а $\mathbf{k}$, как и в предыдущей главе, обозначает квазиволновой вектор. Тогда

$$\mathbf{v} = \nabla_{\mathbf{k}}\omega_{\mathbf{k}}. \tag{1.3'}$$

В теории волновых движений доказывается, что производные от частоты волны по компонентам волнового вектора определяют групповую скорость волны (см. [М1], § 7; [М2], § 3; [М3], § 3). По этой причине среднюю скорость электрона в идеальной кристаллической решетке можно рассматривать как групповую скорость волнового пакета, составленного из функций Блоха.

Если поместить кристалл во внешнее поле, электрическое или магнитное, то состояние с данным квазиимпульсом станет нестационарным: квазиимпульс — а с ним и средняя скорость — будет меняться под действием поля. Ограничимся достаточно слабыми электрическими полями. Именно, будем считать, что напряженность внешнего электрического поля $\mathcal{E}$, действующего на электрон, удовлетворяет неравенству

$$e\mathcal{E}a \ll E_g. \tag{1.6}$$

При этом междузонные переходы под влиянием электрического поля маловероятны (см. § 6) и основная роль поля сводится к изменению квазиимпульса электрона в пределах данной зоны.

Вычислим производную от квазиимпульса по времени в присутствии внешнего поля. Для этой цели вспомним, что для средних значений квантовомеханических величин остаются в силе классические уравнения движения ([М1], § 32; [М3], § 8).

Следовательно, изменение энергии со временем (в пределах данной зоны) должно определяться равенством

$$\frac{dE_l(\mathbf{p})}{dt} = \mathbf{F}\mathbf{v}_l(\mathbf{p}). \tag{1.7}$$

Далее,

$$\frac{dE_l(\mathbf{p})}{dt} = \frac{dE_l[\mathbf{p}(t)]}{dt} = \left(\nabla_{\mathbf{p}} E_l(\mathbf{p}),\ \frac{d\mathbf{p}}{dt}\right), \tag{1.8}$$

и, следовательно, равенство (1.7) можно переписать в виде

$$\left(\mathbf{v},\ \frac{d\mathbf{p}}{dt} - \mathbf{F}\right) = 0.$$

Это соотношение должно быть справедливо при любой ориентации вектора $\mathbf{v}$. Поэтому должно удовлетворяться уравнение движения

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}. \tag{1.9}$$

По форме это соотношение совпадает со вторым законом Ньютона (если под $\mathbf{p}$ понимать импульс!). Фактически, однако, здесь имеется глубокая разница: в уравнении (1.9) вектор $\mathbf{p}$ — не импульс, а квазиимпульс, а $\mathbf{F}$ — не полная сила, а только сила, действующая на электрон со стороны внешних полей. Сила, действующая на электрон со стороны регулярных атомов решетки (создающих периодическое поле), не входит в правую часть (1.9): она уже учтена в виде закона дисперсии $E_l(\mathbf{p})$ и, тем самым, в формуле для средней скорости.

Приведенный выше вывод уравнения (1.9) справедлив лишь для сил, способных совершать работу над электроном. В противном случае (например, для классической силы Лоренца, действующей на электрон в магнитном поле) равенство (1.7) обращается в тождество «0 = 0». Само уравнение (1.9), однако, остается в силе и для электрона, движущегося в магнитном поле. Доказательство этого утверждения оказывается более сложным; его можно найти в книге [1], гл. 6.

Отличие уравнения (1.9) от второго закона Ньютона проявляется при вычислении среднего ускорения электрона $\mathbf{a}$. Действительно, согласно (1.3) мы имеем

$$a_\alpha = \frac{dv_\alpha}{dt} = \frac{d}{dt}\frac{\partial E_l(\mathbf{p})}{\partial p_\alpha}.$$

Далее, как и при выводе уравнения (1.9), воспользуемся соотношением (1.8). Получим

$$a_\alpha = \frac{\partial^2 E_l(\mathbf{p})}{\partial p_\alpha\, \partial p_\beta}\frac{dp_\beta}{dt},$$

откуда

$$\frac{dv_\alpha(\mathbf{p}, l)}{dt} = \frac{\partial^2 E_l(\mathbf{p})}{\partial p_\alpha\, \partial p_\beta} F_\beta. \tag{1.10}$$

Величины

$$\frac{\partial^2 E_l(\mathbf{p})}{\partial p_\alpha \partial p_\beta} \equiv m^{-1}_{\alpha\beta}(\mathbf{p}, l) \tag{1.11}$$

называют компонентами тензора обратной эффективной массы электрона в l-той зоне в точке $\mathbf{p}$. Вообще говоря, они зависят от квазиимпульса; вблизи границ зон они совпадают с величинами (III.8.2). Видно, таким образом, что название «эффективная масса» имеет глубокий смысл: совокупность компонент $m^{-1}_{\alpha\beta, l}$ определяет в среднем всю динамику электрона вблизи границ зоны. Заметим, что именно эти области энергии обычно играют главную роль в большинстве электронных явлений в полупроводниках: в условиях термодинамического равновесия носители заряда сосредоточены в основном вблизи дна соответствующей зоны.

Уравнения (1.10) приобретают особенно простой вид вблизи границ невырожденных зон, когда для энергии можно воспользоваться выражением (III.8.5). Тогда, направляя оси координат вдоль главных осей эллипсоида энергии, мы имеем

$$\frac{dv_\alpha}{dt} = \frac{F_\alpha}{m_\alpha} \tag{1.10'}$$

(без суммирования по α!).

В частности, при одинаковых эффективных массах ($m_x = m_y = m_z = m$) мы имеем

$$\frac{d\mathbf{v}}{dt} = \frac{\mathbf{F}}{m}. \tag{1.10''}$$

Уравнения (1.10) — (1.10'') оправдывают введенное в гл. I представление об эффективной массе носителя заряда, устанавливая вместе с тем и пределы его применимости (неравенство (1.6)).

Заметим, что уравнения (1.10) — (1.10'') определяют лишь усредненное (в квантовомеханическом смысле) движение электрона. Для полного квантового описания требуется значительно бо́льшая информация, содержащаяся в волновой функции. В физике твердого тела, однако, часто встречается ситуация, когда квантовые поправки к движению электронов во внешних полях достаточно малы. Действительно, характерная длина волны де Бройля в электронном газе составляет $\bar{\lambda} = \frac{2\pi\hbar}{\bar{p}}$, где $\bar{p}$ — характерное значение квазиимпульса. В качестве такового для газа, подчиняющегося статистике Больцмана, следует взять «тепловое» значение $\bar{p} = \sqrt{mkT}$ (в случае электронного газа в металлах оценка меняется, но окончательный результат остается в силе). Таким образом, $\bar{\lambda} = \frac{2\pi\hbar}{\sqrt{mkT}}$, что при $T = 300$ К и $m = m_0$ составляет около 10^{-6} см. Как известно, одно из условий «классичности» движения состоит в том, чтобы потенциальная энергия электрона медленно менялась на расстоянии $\bar{\lambda}$. Видно, что это условие действительно часто выполняется.

Подчеркнем, что речь идет в данном случае только о внешних полях. Движение электронов в периодическом поле идеальной решетки, безусловно, должно рассматриваться квантовомеханически. Все нужные нам характеристики этого движения, однако, сводятся к закону дисперсии $E_l(\mathbf{p})$ (в частном случае (1.10″) — к эффективной массе m). Коль скоро закон дисперсии задан, дальнейшие рассуждения, касающиеся поведения системы во внешних полях, могут быть классическими. При этом уравнения (1.10) — (1.10″) дают практически всю необходимую информацию.

§ 2. Электроны и дырки

Уравнения (1.10) — (1.10″) в сочетании с формулой для тока (1.4) позволяют придать ясный смысл представлению о «положительных дырках», введенному в гл. I на основании опытных данных о знаке постоянной Холла и термоэдс. Для этой цели заметим прежде всего, что, как мы знаем, эффективные массы, входящие в (1.10) — (1.10″), могут быть и отрицательными: так обстоит дело вблизи потолка энергетической зоны. Легко видеть, что это можно интерпретировать как изменение знака заряда носителя. Действительно, пусть $\mathbf{F}$ есть сила Лоренца

$$\mathbf{F} = e\mathfrak{E} + \frac{e}{c}[\mathbf{v}\times\mathfrak{B}].$$

Видно, что в уравнения движения (1.10) — (1.10″) входят не заряд электрона и компоненты тензора $m^{-1}_{\alpha\beta}$ по отдельности, а только определенная комбинация их: $em^{-1}_{\alpha\beta}$ или, в случае (1.10″), e/m. Отсюда явствует, что электрон с отрицательной эффективной массой ускоряется электрическим и магнитным полями как частица с положительной массой и положительным зарядом.

Отрицательные значения эффективной массы не должны вызывать удивления. Действительно, уравнения (1.10′), (1.10″) описывают среднее ускорение электрона под действием силы Лоренца. Кроме нее, однако, есть еще сила, действующая на электрон со стороны атомов решетки и учитываемая видом закона дисперсии $E_l(\mathbf{p})$. Именно она и ответственна за возникновение отрицательной эффективной массы.

Направление макроскопических потоков заряда, энергии и т. д. в системе электронов зависит от того, какие именно электроны дают главный вклад в них. Здесь следует выделить предельные случаи почти свободной и почти заполненной зоны.

Первый из них соответствует зоне проводимости в полупроводнике обычного типа. В условиях, близких к термодинамическому равновесию, все электроны находятся у дна зоны, где эффективные массы положительны. Только эти электроны и могут участвовать в явлениях переноса — мы имеем здесь систему отрицательно заряженных носителей.

Второй случай отвечает, например, валентной зоне, из которой удалено некоторое (небольшое) число электронов (они могут быть, в частности, переброшены в зону проводимости за счет энергии теплового движения решетки). В условиях, близких к термодинамически равновесным, не заняты только состояния у потолка зоны. Нижние, полностью заполненные состояния, не дают вклада в потоки заряда, энергии и т. д. Действительно, в силу (III.4.4) каждому электрону с квазиимпульсом $\mathbf{p}$ там можно поставить в соответствие другой электрон с квазиимпульсом $-\mathbf{p}$ и той же энергией (III.4.4). Согласно (1.4) вклады этих электронов в ток взаимно компенсируются. Отличный от нуля вклад в поток электронов могут дать только те из них, которые не имеют «партнеров» с той же энергией и противоположно направленным квазиимпульсом. Это — электроны, находящиеся в не полностью занятых состояниях, т. е. у потолка зоны, где эффективные массы отрицательны. Иначе говоря, мы имеем здесь как бы систему положительно заряженных носителей.

При этом весьма удобным оказывается представление о «дырках» в не полностью заполненной зоне. Его легко ввести, замечая, что, как было выше показано, плотность тока $\mathbf{j}_l$, создаваемая всеми электронами целиком заполненной зоны, равна нулю:

$$\mathbf{j}_l = 2\sum_{\mathbf{p}} \mathbf{j}_l(\mathbf{p}) = 0. \tag{2.1}$$

Здесь множитель 2 возник из-за двух возможных значений проекции спина (при заданном квазиимпульсе $\mathbf{p}$); суммирование по $\mathbf{p}$ охватывает первую зону Бриллюэна. Пусть теперь один электрон удален из зоны, в результате чего освободилось состояние с квазиимпульсом $\mathbf{p}'$ и, скажем, положительной проекцией спина *) (состояние с квазиимпульсом $\mathbf{p}'$ и отрицательной проекцией спина при этом остается занятым). Тогда плотность тока будет

$$\mathbf{j}_l = 2\sum_{\mathbf{p}\neq\mathbf{p}'} \mathbf{j}_l(\mathbf{p}) + \mathbf{j}_l(\mathbf{p}') = 2\sum_{\mathbf{p}} \mathbf{j}_l(\mathbf{p}) - \mathbf{j}_l(\mathbf{p}') = -\mathbf{j}_l(\mathbf{p}'). \tag{2.2}$$

Плотность тока, создаваемая электронами зоны с одним незанятым состоянием, равна по величине и противоположна по направлению плотности тока, которая создавалась бы недостающим электроном.

Согласно (1.4) и (I.3.13) при $\mathfrak{B}=0$ плотность тока, связанная с одним электроном в единице объема, есть

$$\mathbf{j}_l(\mathbf{p}') = \frac{e^2}{m}\langle\tau\rangle\,\mathcal{E}.$$

*) Направление, на которое проектируется спин, здесь может быть выбрано произвольно.

В рассматриваемом случае $m < 0$. Следовательно,

$$\mathbf{j}_l = -\, j_l(\mathbf{p}') = \frac{e^2}{|m|}\langle\tau\rangle\,\mathcal{E}.$$

Это та плотность тока, которая создавалась бы частицей с массой $|m|$ и зарядом e, движущимся со скоростью $\frac{e}{|m|}\langle\tau\rangle\,\mathcal{E}$.

Таким образом, дело обстоит так, как если бы носителями заряда были частицы с положительной массой $|m|$ и положительным зарядом e. Фактически перемещаются, разумеется, электроны. Однако электрон у потолка зоны движется против действующей на него силы. По этой причине удаление такого электрона приводит к увеличению плотности тока — такому, как если бы появилась «частица» с положительной массой $|m|$ и положительным зарядом. Иначе говоря, коллективное движение всех оставшихся электронов валентной зоны эквивалентно движению одной «частицы».

Аналогично обстоит дело и с плотностью потока энергии. Она дается выражением

$$\mathbf{I}_l = \sum_{\mathbf{p},\,\sigma} E_l(\mathbf{p})\,\mathbf{v}(\mathbf{p})\,n(\mathbf{p},\,\sigma). \tag{2.3}$$

Здесь $n(\mathbf{p},\,\sigma)$ — концентрация электронов с квазиимпульсом $\mathbf{p}$ и проекцией спина σ. В полностью заполненной зоне $n(\mathbf{p},\,\sigma) = 1/V$, где V — объем системы.

При одном недостающем электроне в единичном объеме, плотность потока энергии будет, подобно (2.2),

$$\mathbf{I}_l = 2\sum_{\mathbf{p}\neq\mathbf{p}'} E_l(\mathbf{p})\,\mathbf{v}_l(\mathbf{p}) + E_l(\mathbf{p}')\,\mathbf{v}_l(\mathbf{p}') = -\,E_l(\mathbf{p}')\,\mathbf{v}_l(\mathbf{p}'). \tag{2.4}$$

Поток энергии, создаваемый электронами зоны с одним незанятым состоянием, равен по величине и противоположен по направлению потоку энергии, который создавался бы недостающим электроном. Иначе говоря, он равен потоку энергии, создаваемому одной «частицей» с энергией $-E_l(\mathbf{p}')$.

Совершенно так же можно было бы рассмотреть и случай нескольких отсутствующих электронов (с квазиимпульсами $\mathbf{p}'$, $\mathbf{p}''$, ...). Потоки заряда и энергии здесь оказываются такими же, как если бы они переносились положительно заряженными «частицами» (с энергиями $-E_l(\mathbf{p}')$, $-E_l(\mathbf{p}'')$, ...), взятыми в таком числе, сколько электронов отсутствует.

Эти «частицы» получили название «дырок», ибо происхождение их связано с наличием незаполненных состояний («мест») в зоне *). При описании явлений переноса «электронный» и «дырочный» языки полностью эквивалентны. Последний язык особенно удобен, если речь идет о почти заполненной зоне. Здесь, как мы видели,

*) Представление о дырках было введено Я. И. Френкелем (1928 г.)

свободны состояния у потолка зоны E_v и энергии дырок, отсчитанные от него, равны $-E_l(\mathbf{p})$. В частности, при изотропном параболическом законе дисперсии мы имеем

$$-E_l(\mathbf{p}) = -E_v - \frac{p^2}{2m} = -E_v + \frac{p^2}{2|m|}.$$

Как и следовало ожидать, это есть энергия частицы с положительной эффективной массой $|m|$. Энергия дырки у потолка валентной зоны равна $-E_v$. Эффективные массы дырок обычно снабжают индексом p (не путать с квазиимпульсом), а электронов — n. В данном случае $m_p = |m_n|$.

В такой ситуации, интересуясь только явлениями переноса заряда и энергии, мы можем вообще «забыть» о существовании валентной зоны и электронов в ней. Взамен вводится представление о газе дырок и о дырочной зоне. Минимум энергии в последней совпадает с потолком валентной зоны, а энергия дырок на зонной схеме отсчитывается вниз (рис. 4.1). Квазиимпульс дырки равен квазиимпульсу отсутствующего электрона с обратным знаком. Иначе говоря, мы вправе рассматривать газ носителей заряда двух знаков с положительными эффективными массами.

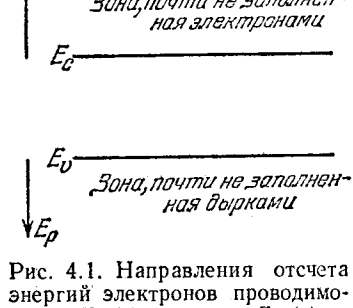


Рис. 4.1. Направления отсчета энергий электронов проводимости $E_n(\mathbf{p})$ и дырок $E_p(\mathbf{p})$.

В связи с этим результатом следует обратить внимание на три обстоятельства.

Во-первых, аналогия между дырками и электронами не полна. Действительно, электроны могут существовать и в вакууме, в то время как представление о дырках имеет смысл лишь постольку, поскольку существуют валентные оболочки, в которых могут оказаться незаполненные места. В отсутствие коллектива электронов представление о дырках лишается смысла. По этой причине дырки часто называют квазичастицами.

Во-вторых, эквивалентность описания системы с помощью представлений об электронах или дырках доказана выше только для потоков заряда и энергии, или, общее, с точки зрения поведения системы во внешних полях. Отсюда, однако, не следует, что эти картины эквивалентны и с точки зрения поля, создаваемого самими этими электронами и дырками. Иначе говоря, «электронная» и «дырочная» картины эквивалентны кинематически, но не всегда — динамически. Вопрос о том, какая из двух картин «правильнее», будет обсужден позднее, в гл. XVII.

Наконец, в-третьих, предыдущие рассуждения относились к поведению системы в электрическом и магнитном полях, но отнюдь

не в гравитационном (и, следовательно, инерционном) поле. Понятие об эффективной массе, а с ним и понятие о положительных дырках не имеют никакого отношения к гравитации (и инерции). Суть дела состоит в том, что силы инерции определяются через производную по времени от импульса, а не от квазиимпульса. Поэтому, например, в известных опытах Толмэна и Стюарта (определение удельного заряда электронов по измерению импульса тока при торможении вращающейся проволочной катушки) отношение заряда частиц к массе всегда будет получаться равным $-e/m_0$, т. е. будет соответствовать отрицательным электронам с истинной (а не эффективной) массой m_0. Такой результат получится и в том случае, когда вместо металла будет взят дырочный полупроводник. Далее, в § III.5 мы видели, что эффективная масса зависит от степени перекрытия атомных волновых функций. Она сильно увеличивается по мере увеличения глубины энергетической зоны; для глубоких узких зон $m \gg m_0$ (внутренние электроны атомов практически локализованы около своих ядер). Однако это обстоятельство никак не сказывается на весе тела, который определяется истинными массами всех ядер и электронов.

§ 3. Движение носителей заряда в постоянном и однородном магнитном поле (классическая теория). Диамагнитный резонанс

Рассмотрим задачу о движении электрона или дырки в постоянном и однородном магнитном поле. Решение этой задачи приведет нас к важному методу экспериментального определения эффективной массы. Дабы не затемнять существа дела, пренебрежем сначала рассеянием носителей заряда. Позднее мы учтем и этот эффект и выясним, в чем состоит его роль и в какой мере она существенна в рассматриваемой задаче.

Направим оси координат вдоль главных осей тензора обратной эффективной массы. Тогда, принимая во внимание формулу для силы Лоренца

$$\mathbf{F} = -\frac{e}{c}[\mathbf{v} \times \mathfrak{B}],$$

мы можем, применительно к рассматриваемой задаче, переписать уравнения (1.10′) в виде

$$\begin{aligned} \frac{dv_x}{dt} &= -\frac{\mathfrak{B}e}{cm_x}(\alpha_z v_y - \alpha_y v_z), \\ \frac{dv_y}{dt} &= -\frac{\mathfrak{B}e}{cm_y}(\alpha_x v_z - \alpha_z v_x), \\ \frac{dv_z}{dt} &= -\frac{\mathfrak{B}e}{cm_z}(\alpha_y v_x - \alpha_x v_y). \end{aligned} \tag{3.1}$$

Здесь α_x, α_y, α_z — косинусы углов между вектором $\mathfrak{B}$ и осями координат. Это есть система линейных однородных уравнений. Будем искать ее решение в виде гармонических колебаний:

$$v_x = v_1 e^{-i\omega t}, \quad v_y = v_2 e^{-i\omega t}, \quad v_z = v_3 e^{-i\omega t}, \tag{3.2}$$

где v_1, v_2, v_3 — постоянные коэффициенты (амплитуды колебаний), а ω — частота. Подставляя (3.2) в уравнения (3.1), получаем систему линейных алгебраических уравнений относительно амплитуд v_1, v_2, v_3:

$$\begin{aligned}
-i\omega v_1 + \frac{\mathscr{B}e}{m_x c}\alpha_z v_2 - \frac{\mathscr{B}e}{m_x c}\alpha_y v_3 &= 0,\\
-\frac{\mathscr{B}e}{m_y c}\alpha_z v_1 - i\omega v_2 + \frac{\mathscr{B}e}{m_y c}\alpha_x v_3 &= 0,\\
\frac{\mathscr{B}e}{m_z c}\alpha_y v_1 - \frac{\mathscr{B}e}{m_z c}\alpha_x v_2 - i\omega v_3 &= 0.
\end{aligned}$$

Условие разрешимости этой системы состоит в обращении в нуль ее детерминанта. Этим определяются возможные значения частоты колебаний ω:

$$\omega = 0, \tag{3.3а}$$

$$\omega^2 = \omega_c^2 \equiv \left(\frac{\mathscr{B}e}{c}\right)^2 \left(\frac{\alpha_x^2}{m_y m_z} + \frac{\alpha_y^2}{m_z m_x} + \frac{\alpha_z^2}{m_x m_y}\right), \tag{3.3б}$$

т. е.

$$\omega_c = \frac{\mathscr{B}e}{m_c c}, \tag{3.4}$$

где

$$\frac{1}{m_c} = \left(\frac{\alpha_x^2}{m_y m_z} + \frac{\alpha_y^2}{m_z m_x} + \frac{\alpha_z^2}{m_x m_y}\right)^{1/2}. \tag{3.5}$$

Величину m_c иногда называют *циклотронной эффективной массой*.

При $m_x = m_y = m_z = m$ выражение (3.5) переходит в формулу (I.3.2). Оно упрощается также, если изоэнергетическая поверхность представляет собой эллипсоид вращения (III.8.5). Действительно, обозначим через φ угол между направлением магнитной индукции $\mathscr{B}$ и осью вращения эллипсоида (осью Z). Тогда

$$\alpha_z^2 = \cos^2\varphi, \quad \alpha_x^2 + \alpha_y^2 = \sin^2\varphi, \quad m_z = m_\parallel, \quad m_x = m_y = m_\perp$$

и выражение (3.5) принимает вид

$$\frac{1}{m_c} = \left(\frac{\cos^2\varphi}{m_\perp^2} + \frac{\sin^2\varphi}{m_\perp m_\parallel}\right)^{1/2}. \tag{3.5′}$$

Корень (3.3а) соответствует движению с постоянной скоростью — равномерному движению вдоль направления поля. Два корня (3.3б) описывают гармонические колебания, в совокупности своей сводящиеся к равномерному вращению электрона в плоскости, перпен-

дикулярной магнитному полю. Угловая скорость этого вращения равна ω_c.

Можно обобщить формулу (3.5) и на случай невырожденных зон при непараболическом законе дисперсии. Для этой цели удобно воспользоваться уравнением (1.9), совмещая ось Z с направлением магнитного поля:

$$\frac{d\mathbf{p}}{dt} = -\frac{e}{c}[\mathbf{v}\times\mathcal{B}],$$

т. е.

$$\frac{dp_x}{dt} = -\frac{e}{c}\mathcal{B}v_y, \quad \frac{dp_y}{dt} = \frac{e}{c}\mathcal{B}v_x, \quad \frac{dp_z}{dt} = 0. \tag{3.6}$$

Поскольку сила Лоренца не совершает работы, видим, что в рассматриваемой задаче имеются два интеграла движения:

$$p_z = \text{const}, \quad E(\mathbf{p}) \equiv E = \text{const}. \tag{3.7}$$

Эти уравнения описывают траекторию электрона в пространстве квазиимпульсов. Будем считать ее замкнутой и несамопересекающейся. Согласно (III.8.3) в случае параболического закона дисперсии это есть эллипс

$$\frac{(p_x - p_x^0)^2}{2m_x} + \frac{(p_y - p_y^0)^2}{2m_y} = E - \frac{(p_z - p_z^0)^2}{2m_z} = \text{const}. \tag{3.7'}$$

Отклонения от параболичности приводят к деформации эллипса; когда они становятся очень большими, траектория может (но не обязана) стать и незамкнутой. Исследование показывает, однако, что это возможно лишь в области энергий, достаточно далеких от дна зоны.

Обозначим через $dp_\perp$ элемент дуги траектории и будем рассматривать время t как параметр, определяющий положение точки на ней. Тогда

$$\frac{dp_\perp}{dt} = \left[\left(\frac{dp_x}{dt}\right)^2 + \left(\frac{dp_y}{dt}\right)^2\right]^{1/2}. \tag{3.8}$$

Введем, далее, нормальную к кривой (3.7) составляющую вектора $\mathbf{v}$: $\mathbf{v}_\perp = \{v_x, v_y\}$. Абсолютная величина ее есть $v_\perp = \sqrt{v_x^2 + v_y^2}$. Пользуясь равенством (3.8), можем переписать уравнения (3.6) в виде

$$\frac{dp_\perp}{dt} = -\frac{e}{c}\mathcal{B}v_\perp. \tag{3.6'}$$

Отсюда

$$t = -\frac{c}{e\mathcal{B}}\int \frac{dp_\perp}{v_\perp}.$$

Период обращения электрона по орбите $2\pi/\omega_c$ получится, если взять здесь интеграл вдоль всей траектории (3.7):

$$\frac{2\pi}{\omega_c} = -\frac{c}{e\mathcal{B}}\oint \frac{dp_\perp}{v_\perp}. \tag{3.9}$$

Фигурирующий в (3.9) интеграл можно выразить через площадь $S(E, p_z)$, ограниченную кривой (3.7). Действительно,

$$S(E, p_z) = \int dp_x\, dp_y,$$

причем пределы интегрирования определяются равенствами (3.7).

Введем вместо p_x и p_y переменные E' и $p_\perp$, где E' принимает все значения от E_c до E. При этом элемент площади $dp_x\,dp_y$ дается выражением

$$dp_x\,dp_y = \frac{dE\,dp_\perp}{v_\perp}.$$

Таким образом,

$$S(E,\ p_z) = \int_{E_c}^{E} dE' \oint \frac{dp_\perp}{v_\perp},$$

и, следовательно,

$$\oint \frac{dp_\perp}{v_\perp} = \frac{\partial S(E,\ p_z)}{\partial E}.$$

Подставляя это выражение в (3.9), получаем

$$\omega_c = \frac{e\mathscr{B}}{m_c c}, \tag{3.4'}$$

где

$$m_c = \frac{1}{2\pi}\frac{\partial S(E,\ p_z)}{\partial E}. \tag{3.10}$$

В условиях (3.7') отсюда вновь получается формула (3.5). В более общем случае циклотронная эффективная масса (3.10) зависит от E и p_z.

В случае вырожденных зон изложенные выше общие соображения остаются в силе, однако для циклотронной эффективной массы вместо (3.5) получается более сложное выражение.

Из формулы (3.5) видно, что, измеряя частоту вращения ω_c электронов (дырок) при различных направлениях магнитной индукции, можно определить компоненты тензора эффективных масс, а следовательно, и форму изоэнергетической поверхности *).

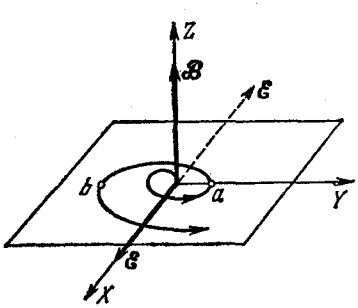


Рис. 4.2. К объяснению диамагнитного (циклотронного) резонанса.

Частоту ω_c можно определить, исследуя поглощение электромагнитных волн в полупроводнике. Эту возможность иллюстрирует рис. 4.2, на котором показана траектория электрона, движущегося в постоянном и однородном магнитном поле с индукцией $\mathscr{B}$. Линейно поляризованная электромагнитная волна распространяется вдоль оси Y, и вектор ее электрического поля $\mathscr{E}$ параллелен плоскости орбиты электрона. Длина электромагнитной волны намного

*) Строгое квантовомеханическое рассмотрение (см. ниже, § 5) показывает, что принятая в настоящем параграфе классическая трактовка движения носителей заряда в магнитном поле имеет определенные пределы применимости. Формула (3.5), однако, остается в силе и в точной теории.

больше диаметра орбиты, так что мгновенные значения $\mathfrak{E}$ одинаковы во всех ее точках. Если, например, электрон находится в точке a и поле $\mathfrak{E}$ направлено так, как показано на рисунке, то поле будет ускорять движение электрона. Когда электрон совершит половину оборота и попадет в точку b, фаза поля $\mathfrak{E}$ изменится. Если частота волны ω совпадает с ω_c, то фаза $\mathfrak{E}$ изменится как раз на π, и поле $\mathfrak{E}$ будет по-прежнему ускорять движение электрона. Иначе говоря, при $\omega = \omega_c$ будет иметь место резонансное поглощение электромагнитной волны системой электронов. Это явление получило название *диамагнитного* или *циклотронного резонанса*.

При непараболическом законе дисперсии условия резонанса оказываются различными для электронов разной энергии. Это приводит к «размазке» резонанса, затрудняя его наблюдение. Исключение составляет лишь случай сильно вырожденного газа носителей заряда, когда, как показано в гл. XIII, существенна лишь энергия, равная энергии Ферми.

В действительности прецессионное движение носителей заряда нарушается соударениями, которые уменьшают резонансное поглощение и могут даже сделать вообще невозможным его наблюдение. Остановимся на этом подробнее, пользуясь, как и выше, полуклассическим описанием. Рассмотрим для определенности случай постоянной скалярной эффективной массы m и положим, что индукция $\mathfrak{B}$ направлена по оси Z, а электрическое поле волны $\mathfrak{E}$ — по оси X (рис. 4.2). Магнитным полем волны будем пренебрегать по сравнению с сильным постоянным полем $\mathfrak{B}$. Влияние соударений будем учитывать введением некоторой силы трения

$$\mathbf{F}_{\text{тр}} = -\frac{m\mathbf{v}}{\tau_p},$$

где τ_p — время релаксации импульса, определяющее подвижность частиц (ср. гл. I, XIII). Тогда уравнения движения частицы будут

$$\begin{aligned} m\frac{dv_x}{dt} &= -e\mathfrak{E} - \frac{e}{c}v_y\mathfrak{B} - \frac{m}{\tau_p}v_x, \\ m\frac{dv_y}{dt} &= \frac{e}{c}v_x\mathfrak{B} - \frac{m}{\tau_p}v_y. \end{aligned} \tag{3.11}$$

Интересуясь, как и выше, установившимися колебаниями вида (3.2) и подставляя эти выражения в уравнения (3.11), мы получим

$$\begin{aligned} m(-i\omega + \tau_p^{-1})v_x &= -e\mathfrak{E} - \frac{e}{c}\mathfrak{B}v_y, \\ m(-i\omega + \tau_p^{-1})v_y &= \frac{e}{c}\mathfrak{B}v_x. \end{aligned}$$

Исключая отсюда v_y, мы найдем среднюю дрейфовую скорость v_x, а следовательно, и электропроводность σ для переменного поля

волны $\mathcal{E}$:

$$\sigma = -\frac{env_x}{\mathcal{E}} = \sigma_0 \frac{1 - i\omega\tau_p}{(1 - i\omega\tau_p)^2 + (\omega_c\tau_p)^2}. \qquad (3.12)$$

Здесь n — концентрация частиц, а через $\sigma_0 = en\mu = \frac{e^2 n}{m}\tau_p$ обозначена электропроводность для постоянного электрического поля. Величина σ оказывается зависящей от частоты ω и, кроме того, комплексной, что означает наличие разности фаз между колебаниями тока и поля. Поглощение электромагнитной волны пропорционально вещественной части σ. Из формулы (3.12) мы имеем

$$\frac{\mathrm{Re}\,\sigma}{\sigma_0} = \frac{1 + (\omega\tau_p)^2 + (\omega_c\tau_p)^2}{[1 + (\omega_c\tau_p)^2 - (\omega\tau_p)^2]^2 + 4(\omega\tau_p)^2}. \qquad (3.13)$$

Зависимость $\frac{\mathrm{Re}\,\sigma}{\sigma_0}$ от $\frac{\omega_c}{\omega}$ для различных значений $\omega\tau_p$ показана на рис. 4.3. При $\omega\tau_p \ll 1$ она монотонна. Это и понятно, так как в данном случае электрон не успевает сделать ни одного оборота без столкновений; скорость его после столкновения может быть направлена под любым углом к электрическому полю волны. Иначе говоря, столкновения нарушают синфазный характер движения электронов относительно колебаний поля в волне, необходимый для возникновения резонанса. Резонансное поглощение становится заметным при $\omega\tau_p \gtrsim 1$; оно тем заметнее, чем больше $\omega\tau_p$.

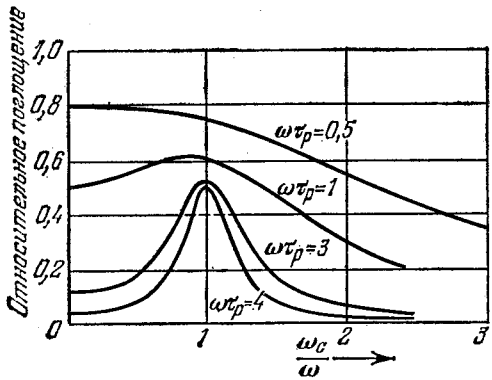


Рис. 4.3. Резонансное поглощение электромагнитной волны в зависимости от отношения ω_c/ω.

При экспериментальном осуществлении диамагнитного резонанса необходимо прежде всего обеспечить выполнение условия $\omega_c\tau_p > 1$. Пользуясь формулой (3.4) и выражая τ_p через подвижность μ, мы имеем

$$\omega_c\tau_p = \frac{\mathscr{B}e}{mc}\tau_p = \frac{1}{c}\mu\mathscr{B} > 1$$

или, если μ дано в практических единицах (см^2/с·В), а $\mathscr{B}$ — в гауссах,

$$\mu > 10^8/\mathscr{B}.$$

В типичных экспериментах используют электромагниты, создающие индукцию от нескольких тысяч до нескольких десятков тысяч гаусс.

Полагая $\mathscr{B} = 1 \cdot 10^3$ Гс, мы находим, что μ должно быть $> 10^5 см^2/с \cdot В$. В большинстве известных полупроводников при комнатных температурах подвижность заметно меньше, поэтому оказывается необходимым работать при низких температурах (обычно гелиевых) и использовать очень чистые кристаллы (чтобы примесное рассеяние не уменьшало подвижность).

Однако при этом возникает другая трудность: при очень низких температурах все остаточные примеси не ионизованы и в зонах практически не имеется носителей. Поэтому приходится прибегать к вспомогательным приемам создания подвижных электронов (дырок). Один из них заключается в ударной ионизации мелких примесей электрическим полем самой электромагнитной волны, для

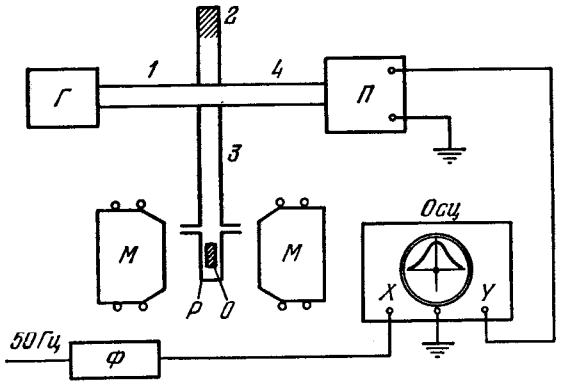


Рис. 4.4. Схема наблюдения диамагнитного резонанса. *Г* — генератор СВЧ колебаний, *П* — приемник, *Осц* — осциллограф, *Р* — резонансная полость с образцом *О*, *ММ* — электромагнит, *Ф* — фазовращатель.

чего мощность волны должна быть больше некоторого предела, различного для разных полупроводников. Подвижные носители заряда можно также создать с помощью слабого освещения кристалла.

Частоты ω_c, определяемые формулой (3.4), при $\mathscr{B} \sim 10^4$ Гс имеют порядок $10^{10} \div 10^{12}$ $с^{-1}$ и соответствуют электромагнитным волнам сантиметрового или миллиметрового диапазона. Поэтому для наблюдения диамагнитного резонанса чаще всего используют СВЧ технику.

На рис. 4.4 показана принципиальная схема одного из типичных опытов (в действительности такие установки содержат ряд дополнительных узлов и блоков; они опущены, как не имеющие принципиального значения). СВЧ колебания от генератора *Г* ослабляются аттенюатором и попадают в плечо *1* балансового моста. Последний представляет собой двойной волноводный тройник, имеющий четыре

плеча. Плечи *2* и *3* присоединяются к согласованным нагрузкам, одной из которых является резонансная полость P с исследуемым образцом O. При равенстве этих нагрузок сигнала в плече *4* нет. При возникновении резонанса изменяется поглощение волны в резонаторе с образцом, баланс моста нарушается и в плече *4* появляется сигнал. Последний регистрируется обычным приемником. На опыте удобно изменять не частоту волны ω, а величину магнитной индукции (т. е. ω_c); для чего в электромагните MM устраивают дополнительную обмотку, модулирующую магнитную индукцию с низкой частотой. Подавая выходной сигнал приемника на одну из пар отклоняющих пластин осциллографа *Осц*, а на другую пару пластин — сигнал, пропорциональный изменению магнитной индукции, можно получить на экране осциллографа всю резонансную кривую. На рис. 4.5 в качестве примера показана форма кривых поглощения в германии n-типа *). Из рисунка видно, что при повороте кристалла изменяются и число наблюдаемых максимумов, и их положение, что свидетельствует об анизотропии эффективной массы.

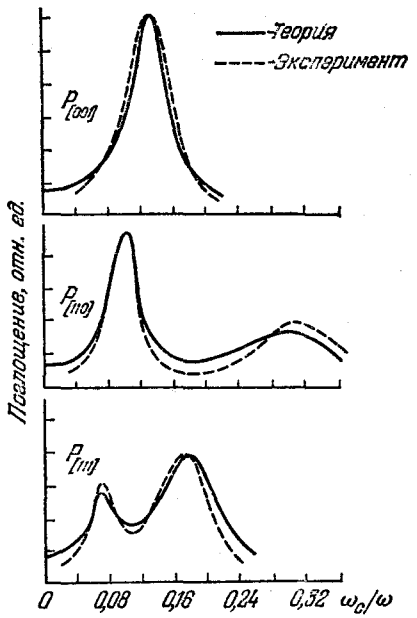


Рис. 4.5. Зависимость поглощения P СВЧ колебаний от магнитной индукции в германии n-типа. Индексы у P указывают направление магнитной индукции. $T=4{,}2$ К, частота 8,9 ГГц.

Зная значение $\mathscr{B}$, соответствующее резонансу ($\omega_c = \omega$), из формулы (3.4) можно непосредственно найти соответствующее значение m_c.

Укажем еще, что опыты с диамагнитным резонансом позволяют определить знак основных носителей заряда. Для этого вместо прямоугольного резонатора употребляют цилиндрический и возбуждают в нем колебания, поляризованные по кругу. Если магнитная индукция $\mathfrak{B}$ направлена параллельно оси резонатора, то круговые орбиты носителей будут перпендикулярны оси резонатора, т. е. будут лежать в той же плоскости, что и вектор $\mathfrak{E}$ волны. Легко сообразить, что поглощение волны должно наблюдаться только в том случае, когда направление вращения $\mathfrak{E}$ в волне совпадает

*) *B. Lax, H. J. Zeiger, R. N. Dexter*, Physica **20**, 818 (1954).

с направлением вращения частиц. А это последнее, при заданном направлении $\mathfrak{B}$, различно для положительных и отрицательных частиц. Такие опыты позволили получить прямое экспериментальное доказательство существования положительных дырок.

§ 4. Метод эффективной массы

Как указывалось в гл. II, атомы примеси, как и другие структурные дефекты решетки, могут создавать дискретные уровни в запрещенной зоне. Расчет их составляет одну из задач электронной теории реальных кристаллов. Потенциальную энергию электрона при этом можно представить в виде суммы двух слагаемых: $U(\mathbf{r}) + \delta U(\mathbf{r})$. Здесь $U(\mathbf{r})$ обозначает по-прежнему потенциальную энергию электрона в идеальном кристалле, а $\delta U(\mathbf{r})$ есть непериодическая функция, описывающая взаимодействие электрона с несовершенствами решетки. Так же обстоит дело и при квантовомеханическом рассмотрении движения электронов в идеальном кристалле при наличии внешнего электрического поля.

Во многих случаях функция $\delta U(\mathbf{r})$ сравнительно плавно изменяется в пространстве, оставаясь практически постоянной на протяжении постоянной решетки a. При этом задачу о движении электрона удобно решать с помощью приближенного приема, называемого *методом эффективной массы*. Идея его состоит в том, чтобы, пользуясь плавностью функции δU, свести уравнение Шредингера с потенциалом $U + \delta U$ к более простому виду, содержащему явно только δU. Роль периодического потенциала состоит при этом в изменении оператора кинетической энергии: вместо массы свободного электрона в нем появляются эффективные массы, описывающие поведение носителя заряда в соответствующем идеальном кристалле (с этим связано название метода). Зонную структуру идеального кристалла при этом следует считать известной.

Для выполнения намеченной только что программы необходимо, чтобы в области, где, в основном, движется электрон, удовлетворялось условие плавности поля δU *):

$$a\frac{|\nabla\delta U|}{|\delta U|} \ll 1. \tag{4.1}$$

Другие условия применимости метода эффективной массы будут указаны ниже. Подробное изложение метода можно найти в книгах [М7] и [2]. Здесь мы сформулируем только окончательные результаты.

Рассмотрим сначала поведение электронов вблизи дна невырожденной зоны с параболическим изотропным законом дисперсии.

*) Если кристалл не кубический, то в качестве a в (4.1) можно подставить любую из постоянных решетки, ибо все они — одного порядка величины. Метод эффективной массы был предложен С. И. Пекаром в 1948 г.

В этом случае волновую функцию электрона, движущегося в рассматриваемом силовом поле, можно представить в виде

$$\psi(\mathbf{r}) = \chi(\mathbf{r})\,\psi_c(\mathbf{r}). \tag{4.2}$$

Здесь $\psi_c(\mathbf{r})$ — нормированная функция Блоха, соответствующая дну рассматриваемой зоны, а $\chi(\mathbf{r})$ — «сглаженная» волновая функция. Уравнение для нее можно найти, подставляя выражение (4.2) в уравнение Шредингера с потенциальной энергией $U + \delta U$. Считая функцию χ достаточно плавной (см. ниже), мы получаем

$$\frac{\hat{p}^2}{2m}\chi + \delta U \cdot \chi = E\chi, \tag{4.3}$$

где $\hat{p} = -i\hbar\nabla$. Собственные значения уравнения (4.3) суть возможные значения энергии, отсчитанные от дна зоны E_c. В частности, область $E < 0$ соответствует энергии, меньшей E_c. При $\delta U = 0$ уравнение (4.3) дает

$$\chi = V^{-1/2} e^{\frac{i}{\hbar}\mathbf{pr}}, \quad E = \frac{p^2}{2m}. \tag{4.4}$$

Множитель $V^{-1/2}$ в (4.4) обеспечивает нормировку χ на единицу. Как и следовало ожидать, здесь просто восстанавливается энергетический спектр электрона в зоне проводимости.

Уравнение (4.3) и формула (4.2) справедливы, если функция χ мало изменяется на протяжении постоянной решетки, а собственные значения E лежат достаточно близко к дну зоны. Именно, должно выполняться неравенство

$$|E| \ll E_g, \tag{4.5}$$

где E_g — ширина запрещенной зоны или, общее, расстояние от дна рассматриваемой зоны до ближайшего к нему края какой-либо другой зоны.

Аналогично обстоит дело и в случае электронов, движущихся у потолка зоны (по-прежнему невырожденной и со скалярной эффективной массой). При этом эффективная масса отрицательна; соответственно аналог уравнения (4.3) имеет вид

$$\frac{\hat{p}^2}{2|m|}\chi - \delta U \cdot \chi = (-E)\chi, \tag{4.3'}$$

где, как и в § 2, $|m| = m_p$. Поскольку потенциальная энергия δU имеет электрическое происхождение, изменение знака при ней можно интерпретировать как изменение знака заряда носителя. Иначе говоря, уравнение (4.3) описывает поведение дырки с положительной эффективной массой $|m|$ вблизи дна дырочной (т. е. потолка валентной) зоны. Величина $-E$ есть энергия дырки, отсчитываемая, как и в § 2, от потолка валентной зоны E_v вниз. Отрицательные значения $-E$ (т. е. положительные значения E) соответствуют энергии, большей E_v.

При наличии магнитного поля уравнения (4.3), (4.3′) видоизменяются. Как известно из электродинамики, магнитное поле можно описывать вектор-потенциалом $\mathcal{A}$, удовлетворяющим условию

$$\operatorname{rot} \mathcal{A} = \mathcal{B}. \tag{4.6}$$

Фигурирующий в (4.3) оператор импульса при этом принимает известный из квантовой механики канонический вид:

$$\hat{\mathbf{p}} = -i\hbar\nabla + \frac{e}{c}\mathcal{A}.$$

Далее, к потенциальной энергии δU добавляется энергия спинового магнитного момента электрона

$$\beta g (\sigma, \mathcal{B}).$$

Здесь $\beta = e\hbar/2m_0c$ — магнетон Бора, g — гиромагнитное отношение. Для электронов и дырок в полупроводниках оно может отличаться от двух за счет спин-орбитального взаимодействия. Функции ψ и χ теперь представляют собой двухрядные матрицы.

Таким образом, вместо (4.3) или (4.3′) мы получаем

$$\frac{1}{2m}\left(-i\hbar\nabla \pm \frac{e}{c}\mathcal{A}\right)^2 \chi \pm \delta U \cdot \chi \pm g\beta (\sigma, \mathcal{B})\chi = \pm E\chi. \tag{4.3″}$$

Здесь верхние знаки относятся к электрону у дна зоны проводимости, нижние — к дырке у потолка валентной зоны.

При возведении в квадрат оператора

$$\left(-i\hbar\nabla \pm \frac{e}{c}\mathcal{A}\right)$$

в уравнении (4.3″) возникают выражения типа $(\mathcal{A}\nabla)\chi$ и $(\nabla\mathcal{A})\chi$. Поскольку операторы $\mathcal{A}$ и ∇, вообще говоря, не перестановочны друг с другом, уравнение (4.3″) следует дополнить правилом, указывающим порядок их следования. Тщательное исследование [1] показывает, что во всех таких случаях надлежит брать симметризованное выражение $(\mathcal{A}\nabla + \nabla\mathcal{A})\chi$.

Для вычисления различных средних значений также достаточно знать только сглаженные функции χ (их иногда называют «огибающими»). Именно, пусть $L(\mathbf{r})$ — любая функция, плавно изменяющаяся в пространстве, и пусть $\psi_1 = \chi_1\psi_c$, $\psi_2 = \chi_2\psi_c$, где χ_1 и χ_2 — какие-либо решения уравнения (4.3). Тогда, как можно показать [1, 2],

$$\int \psi_1^* L\psi_2\, d\mathbf{r} \simeq \int \chi_1^* L\chi_2\, d\mathbf{r}. \tag{4.7}$$

Иначе говоря, $\chi(\mathbf{r})$ играет роль «эффективной волновой функции» электрона.

Уравнения (4.3) — (4.3″) можно обобщить и на случай анизотропного параболического закона дисперсии, когда изоэнергетические поверхности вблизи дна зоны представляют собой эллипсоиды. Следует лишь заменить m^{-1} тензором

обратной эффективной массы. При этом первое слагаемое в левой части (4.3) примет вид

$$\frac{1}{2} m^{-1}_{\alpha\beta} \hat{p}_\alpha \hat{r}_\beta.$$

В частности, если оси координат совпадают с главными осями эллипсоида энергии, мы получаем вместо (4.3″)

$$\sum_{\alpha=x,\ y,\ z} \left\{ \frac{1}{2m_\alpha} \left(-i\hbar \frac{\partial}{\partial r_\alpha} \pm \frac{e}{c} \mathscr{A}_\alpha \right)^2 \chi \right\} \pm \delta U \cdot \chi \pm g\beta\, (\sigma \mathfrak{H})\, \chi = \pm E\chi. \qquad (4.8)$$

Как мы видели в § III.8, в кристаллах кубической системы анизотропия эффективной массы может иметь место, только если минимум энергии расположен не в центре зоны Бриллюэна. При этом имеется несколько эквивалентных минимумов, соответствующих квазиимпульсам $\mathbf{p}_i$ ($i = 1, 2, \ldots$). Уравнение (4.8) следует писать для каждого минимума в отдельности. Соответственно получится столько (вообще говоря, различных) решений, сколько есть эквивалентных минимумов. Каждое такое решение имеет вид

$$\psi_i = \chi_i \psi_{c,\, i}, \qquad (4.2')$$

где индекс i нумерует минимумы и под $\psi_{c,\, i}$ следует понимать функцию Блоха отвечающую квазиимпульсу $\mathbf{p}_i$ в i-м минимуме.

Методом эффективной массы можно пользоваться также и при непараболическом законе дисперсии. Для этой цели надо заменить в (4.3″) оператор «кинетической энергии» $\frac{1}{2m}\left(-i\hbar\nabla \pm \frac{e}{c}\mathfrak{A}\right)^2$ на $E_l(\hat{\mathbf{p}})$, где $\hat{\mathbf{p}} = -i\hbar\nabla \pm \frac{e}{c}\mathfrak{A}$. Уравнение (4.8) представляет собой частный случай этого общего правила.

Несколько более сложно обобщение на случай вырожденных зон. Тогда нельзя ограничиваться рассмотрением только одной из них, ибо условие (4.5) заведомо не будет выполняться. Физически это означает, что возмущение, описываемое потенциальной энергией δU (или вектор-потенциалом), «перемешивает» различные зоны, если расстояние между ними сравнимо со средней энергией возмущения. Пусть точка вырождения соответствует краю зоны. Тогда волновую функцию системы (в условиях плавности возмущения) можно представить в виде

$$\psi(\mathbf{r}) = \sum_{j=1}^{r} \chi_j(\mathbf{r})\, \psi_j(\mathbf{r}), \qquad (4.9)$$

где индекс j нумерует зоны, смыкающиеся в точке вырождения, r — число таких зон (кратность вырождения), ψ_j — функция Блоха, соответствующая рассматриваемому краю j-й зоны. Для коэффициентов разложения χ_j получается система r дифференциальных уравнений второго порядка

$$\sum_{j'=1}^{r} D^{jj'}_{\alpha\beta} \hat{p}_\alpha \hat{p}_\beta \chi_{j'} + \delta U \cdot \chi_j = E\chi_j. \qquad (4.10)$$

Здесь $D^{jj'}_{\alpha\beta}$ — постоянные, определяющиеся структурой зон и играющие ту же роль, что и компоненты тензора обратной эффективной массы для невырожденных зон. При $\delta U = 0$ уравнение (4.10) должно приводить к закону дисперсии в идеальном кристалле, что и позволяет выразить коэффициенты $D^{jj'}_{\alpha\beta}$ через параметры энергетического спектра, определяемые из опыта.

Рассмотрим, например, случай двукратно вырожденных зон в кубическом кристалле. Изоэнергетические поверхности здесь описываются формулой (III.8.6).

Полагая в (4.10) $\delta U=0$ и $\chi_j=C_j\exp\left(\frac{i}{\hbar}\mathbf{pr}\right)$, где C_j — постоянные, получаем для C_j систему однородных уравнений

$$\sum_{j'=1,\ 2} D^{jj'}_{\alpha\beta}p_\alpha p_\beta C_{j'}-EC_j=0. \tag{4.11}$$

Условие разрешимости системы (4.11) определяет E как функцию вектора $\mathbf{p}$. Введя для краткости обозначение

$$D^{jj'}_{\alpha\beta}\,p_\alpha p_\beta=\gamma_{jj'}, \tag{4.12}$$

легко находим

$$E=\frac{\gamma_{11}+\gamma_{22}}{2}\pm\left[\frac{(\gamma_{11}-\gamma_{22})^2}{4}+\gamma_{12}\gamma_{21}\right]^{1/2}. \tag{4.13}$$

Это есть не что иное, как формула (III.8.6), причем коэффициенты A, B, C легко выразить через $D^{jj'}_{\alpha\beta}$.

Аналогичным образом обстоит дело, если зоны не вырождены, но края их расположены достаточно близко друг от друга, так что условие (4.5) в интересной области энергий все же не выполняется (о таких зонах говорят как о «почти вырожденных»). Такой случай реализуется в полупроводниках с очень малой шириной запрещенной зоны. Роль «почти вырожденных» играют здесь зоны проводимости и валентная. При этом также надо пользоваться не простыми равенствами (4.2) и (4.3), а более сложными (4.9) и (4.10).

Подчеркнем, что деление на «невырожденные» и «почти вырожденные» зоны не является абсолютным. Как видно из предыдущего, оно определяется интервалом энергий, представляющим интерес в том или ином случае. Ориентир здесь дает неравенство (4.5).

§ 5. Энергетический спектр носителя заряда в постоянном и однородном магнитном поле (квантовая теория)

Воспользуемся методом эффективной массы для решения квантовомеханической задачи о движении электрона в постоянном и однородном магнитном поле. Ограничимся для простоты движением электрона со скалярной эффективной массой у дна невырожденной зоны и будем отсчитывать энергию от произвольного начала отсчета. В отсутствие каких-либо структурных дефектов уравнение (4.3″) принимает вид

$$\frac{1}{2m}\left(-i\hbar\nabla+\frac{e}{c}\mathcal{A}\right)^2\chi+\beta g\,(\sigma\mathcal{B})\,\chi=(E-E_c)\,\chi. \tag{5.1}$$

Направим ось Z параллельно напряженности магнитного поля. Тогда, в согласии с (4.6), вектор-потенциал можно выбрать в виде

$$\mathcal{A}_x=\mathcal{A}_z=0,\qquad \mathcal{A}_y=\mathcal{B}x. \tag{5.2}$$

При этом коэффициенты в уравнении (5.1) не зависят от координат y и z и можно искать решение в виде

$$\chi=e^{ik_2y+ik_3z}f(x), \tag{5.3}$$

где k_2 и k_3 — вещественные волновые числа, $f = \begin{pmatrix} f_+ \\ f_- \end{pmatrix}$ — матричная функция; составляющие ее f_+ и f_- соответствуют электрону с положительной и отрицательной z-компонентами спина. Подставляя это выражение в уравнение (5.1), раскрывая оператор $\left(-i\hbar\nabla + \frac{e}{c}\mathcal{A}\right)^2$ и учитывая, что, согласно (5.2), $\operatorname{div}\mathcal{A} = 0$, находим два уравнения для функций $f_\pm$. Их можно записать единым образом, приписывая и энергии E индекс «$\pm$»:

$$-\frac{\hbar^2}{2m}\frac{d^2 f_\pm}{dx^2} + \frac{1}{2m}\left(\hbar k_2 + \frac{e}{c}\mathcal{B}x\right)^2 f_\pm = \left(E_\pm - E_c - \frac{\hbar^2 k_3^2}{2m} \mp \beta g \mathcal{B}\right) f_\pm. \tag{5.4}$$

Удобно ввести обозначение

$$\nu_\pm = E_\pm - E_c - \frac{\hbar^2 k_3^2}{2m} \mp \beta g \mathcal{B} \tag{5.5}$$

и отсчитывать координату x от точки

$$x_0 = -\frac{c\hbar k_2}{\mathcal{B}e}, \tag{5.6}$$

полагая $x = x_0 + x'$.

Тогда уравнения (5.4) примут вид

$$-\frac{\hbar^2}{2m}\frac{d^2 f_\pm}{dx'^2} + \frac{e^2\mathcal{B}^2}{2mc^2}x'^2 f_\pm = \nu_\pm f_\pm. \tag{5.7}$$

Формально это есть не что иное, как уравнение Шредингера для гармонического осциллятора с «коэффициентом упругости» $k = \frac{e^2\mathcal{B}^2}{mc^2}$ и собственной частотой

$$\omega = \sqrt{\frac{k}{m}} = \frac{\mathcal{B}e}{mc} \equiv \omega_c. \tag{5.8}$$

Решение уравнения (5.7) хорошо известно из квантовой механики. Собственные значения $\nu_\pm$ равны

$$\nu_\pm = \hbar\omega_c (n + 1/2), \tag{5.9}$$

где $n = 0, 1, \ldots$ — целое число или нуль, а соответствующие собственные функции суть

$$f_\pm = \gamma^{-1/2} \exp\left\{-\frac{1}{2}\left(\frac{x - x_0}{\gamma}\right)^2\right\} H_n\left(\frac{x - x_0}{\gamma}\right). \tag{5.10}$$

Здесь

$$\gamma = \left(\frac{\hbar c}{\mathcal{B}e}\right)^{1/2} \tag{5.11}$$

есть постоянная размерности длины (ее называют магнитной длиной), H_n — полином Эрмита n-го порядка.

Комбинируя теперь равенства (5.5) и (5.9), находим

$$E_{\pm} - E_c = \frac{\hbar^2 k_3^2}{2m} \pm \beta g \mathscr{B} + \hbar\omega_c (n + 1/2). \qquad (5.12)$$

Первое слагаемое в правой части (5.12) представляет собой кинетическую энергию электрона с импульсом $\hbar k_3$, свободно движущегося вдоль направления магнитного поля. Такой характер движения вполне согласуется с классическими представлениями: компонента силы Лоренца по оси Z равна нулю.

Второе слагаемое в (5.12) описывает энергию спинового магнитного момента в магнитном поле. Она может принимать два значения соответственно двум возможным ориентациям спина.

Наконец, третье слагаемое отвечает энергии движения электрона в плоскости, перпендикулярной магнитному полю. Видим, что эта энергия оказывается квантованной. При заданных значениях проекции спина на ось Z и волнового числа k_3 возможные значения энергии $E_{\pm}$ образуют ряд эквидистантных дискретных уровней (рис. 4.6). Расстояние между ними равно $\hbar\omega_c$. Этот результат был получен в 1930 г. Л. Д. Ландау; соответственно говорят об *уровнях Ландау* и о *квантовании Ландау*. При учете движения электрона вдоль оси Z уровни Ландау размываются в одномерные «подзоны». Каждая из них соответствует определенному значению квантового числа n и той или иной проекции спина на направление магнитного поля. Нижняя из этих подзон отвечает значению $n = 0$, а нижний край ее — дно зоны проводимости в магнитном поле — дается выражением

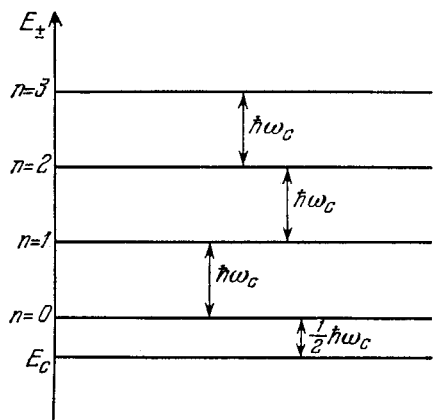


Рис. 4.6. Уровни Ландау при одной определенной ориентации спина (слагаемое $\beta g \mathscr{B}$ в (5.12) опущено).

$$E_{\min} = E_c - |\beta| g \mathscr{B} + \frac{1}{2}\hbar\omega_c. \qquad (5.13)$$

При классической трактовке орбитального движения электрона последнее слагаемое в (5.13) не появилось бы. Учет квантования приводит к повышению границы зоны на величину $^1/_2\hbar\omega_c$ — нулевую энергию осциллятора.

Заметим, что энергия электрона, движущегося в магнитном поле, не зависит от волнового числа k_2. Последнее входит только в выражение для координаты x'. Согласно (5.6) точка x_0 есть «центр

колебаний». В силу произвольности числа k_z при одном и том же значении энергии электрона центр колебаний может находиться где угодно в образце: имеет место вырождение по координате центра осциллятора. Этого и следовало ожидать: поскольку магнитная индукция не зависит от координат, в образце нет физически выделенных точек.

Вычислим теперь степень вырождения каждого уровня в подзоне Ландау, т. е. число квантовых состояний, отвечающих заданному значению энергии E и заданному значению k_z. Для этой цели рассмотрим образец конечного объема V. Как и в отсутствие магнитного поля, форма образца не играет роли, коль скоро размеры его достаточно велики. Удобно взять образец в виде куба со стороной L. По направлениям Y и Z подчиним волновую функцию условиям периодичности. Тогда, подобно (III.3.10),

$$k_y = \frac{2\pi}{L} n_y, \qquad k_z = \frac{2\pi}{L} n_z, \tag{5.14}$$

где n_y и n_z — положительные или отрицательные целые числа или нули.

Поскольку функция (5.10) локализована в области порядка γ около точки x_0, граничные условия по оси x вообще несущественны, коль скоро длина L значительно превышает магнитную длину γ. Достаточно ограничить возможные значения x_0 условием

$$-\frac{L}{2} \leqslant x_0 \leqslant \frac{L}{2}. \tag{5.15}$$

Согласно (5.6), (5.8) и (5.14) отсюда следует, что

$$-\frac{1}{2} m\omega_c L \leqslant \hbar k_y \leqslant \frac{1}{2} m\omega_c L \tag{5.16}$$

и

$$-\frac{1}{2} m\omega_c \frac{L^2}{2\pi\hbar} \leqslant n_y \leqslant \frac{1}{2} m\omega_c \frac{L^2}{2\pi\hbar}. \tag{5.17}$$

Таким образом, при заданных энергии и числе n_z квантовое число n_y может принимать

$$g = m\omega_c \frac{L^2}{2\pi\hbar} \tag{5.18}$$

значений. Это и есть искомая кратность вырождения.

Формулу (5.12) можно обобщить на случай эллипсоидальных изоэнергетических поверхностей. Для этой цели надо лишь, в соответствии с (4.8), заменить оператор кинетической энергии в (5.1) более сложным выражением. Пусть, например, изоэнергетические поверхности представляют собой эллипсоиды вращения (III.8.5). Тогда энергия носителя заряда по-прежнему дается формулой (5.12), но частота ω_c теперь оказывается зависящей от угла между осью магнитного поля и главной осью эллипсоида энергии: она

дается формулой (5.8), в которую вместо m надо подставить величину m_c из (3.5) или (3.5′).

Совпадение результатов квантовомеханического и классического расчетов частоты ω_c здесь не должно вызывать удивления. Действительно, классическое движение заряженной частицы массы m в однородном магнитном поле складывается из свободного перемещения вдоль оси поля и вращения в плоскости, перпендикулярной $\mathcal{B}$. Последнее можно представить как сумму двух гармонических колебаний с частотой ω_c. Тот факт, что при квантовом рассмотрении мы получили лишь один осциллятор, не имеет значения: при другом выборе вектор-потенциала мы могли бы, вместо (5.12), получить выражение, формально содержащее сумму энергий двух гармонических осцилляторов; при этом значения E фактически не изменились бы.

Классическая трактовка задачи о поведении электрона в магнитном поле страдает, однако, двумя недостатками.

Во-первых, она логически непоследовательна: классическое движение заряда по круговой орбите не может быть устойчивым из-за неизбежного — в рамках классической механики — излучения электромагнитных волн ускоренно движущимся зарядом.

Во-вторых, классическая формула для частоты колебаний ω_c совпадает с квантовой только при параболическом законе дисперсии. В более сложных условиях дело обстоит иначе: представление об уровнях и подзонах Ландау сохраняет силу, но формула (5.9) уже не имеет места; спектр уровней Ландау оказывается неэквидистантным. Так, например, обстоит дело в случае вырожденных зон, характерных для дырок в германии, кремнии и ряде других материалов.

Таким образом, магнитное поле существенно изменяет энергетический спектр свободных носителей заряда. Зона проводимости, понимаемая как совокупность значений энергии, характеризуемых тремя непрерывно изменяющимися компонентами квазиволнового вектора, здесь исчезает. Вместо нее появляется совокупность одномерных подзон, каждая из которых отвечает одному из уровней Ландау, т. е. определенному значению квантового числа n, и описывается лишь одной компонентой k_z. Классическая трактовка задачи о движении свободного электрона или дырки в постоянном и однородном магнитном поле может быть оправдана, только если расстояние между соседними уровнями Ландау мало по сравнению с характерной энергией носителей заряда $\bar{E}$:

$$\hbar\omega_c \equiv \frac{e\hbar\mathcal{B}}{mc} \ll \bar{E}. \tag{5.19}$$

Роль $\bar{E}$ обычно играет либо величина kT (если электронный газ подчиняется статистике Больцмана), либо уровень Ферми ζ, отсчитанный от дна соответствующей зоны (если имеет место фер-

миевское вырождение). Магнитные поля, для которых неравенство (5.19) не выполняется и надо учитывать квантование Ландау, называются квантующими.

Следует, однако, помнить, что формулы, полученные в настоящем параграфе, тоже имеют определенные пределы применимости. Действительно, мы рассматривали здесь поведение носителей заряда в идеальном кристалле. В применении к реальному кристаллу полученные таким путем результаты имеют смысл, если можно пренебречь процессами рассеяния. Последние приводят к нестационарности рассматриваемых нами состояний: на каждом из уровней электрон может находиться лишь конечное время τ_p — характерное время свободного пробега. Согласно принципу неопределенности между энергией и временем это означает, что каждый уровень приобретает конечную ширину: $\Delta E \sim \hbar/\tau_p$. Формулы (5.9), (5.12) и т. д. имеют смысл, лишь если эта ширина мала по сравнению с расстоянием между уровнями:

$$\hbar/\tau_p \ll \hbar\omega_c. \tag{5.20}$$

Как и следовало ожидать (в случае параболического закона дисперсии!), это есть не что иное, как классическое неравенство $\omega_c\tau_p \gg 1$.

Далее, применение в данной задаче метода эффективной массы оправдано, коль скоро магнитная длина, определяемая формулой (5.11), велика по сравнению с постоянной решетки. При $\mathscr{B} = 10^4$ Гс мы имеем, согласно (5.11), $\gamma = 0{,}25 \cdot 10^{-6}$ см.

Квантовомеханическое рассмотрение позволяет особенно ясно понять сущность явления диамагнитного резонанса. Очевидно, мы имеем здесь просто переходы между соседними уровнями Ландау, вызываемые фотонами частоты ω_c.

§ 6. Движение и энергетический спектр носителей заряда в постоянном электрическом поле

Пусть в кристалле создано постоянное и однородное в пространстве электрическое поле напряженности $\mathscr{E}$. Этого можно добиться, например, используя достаточно тонкий образец в качестве «диэлектрической прослойки» в конденсаторе. Будем считать, что величина $\mathscr{E}$ удовлетворяет условию (1.6).

В отличие от магнитного, электрическое поле уже с точки зрения классической механики может производить работу над зарядом. Соответственно следует ожидать, что при наложении постоянного и однородного электрического поля носитель заряда будет менять свою энергию, т. е. перемещаться по зоне. Для исследования этого движения воспользуемся уравнением движения (1.9). Для определенности будем говорить об электронах; все результаты для дырок получаются из приводимых ниже формул изменением знака при e.

Замечая, что в рассматриваемом случае сила $\mathbf{F} = -e\mathcal{E}$, мы имеем

$$\frac{d\mathbf{p}}{dt} = -e\mathcal{E}. \tag{6.1}$$

Интегрируя это уравнение с начальным условием $\mathbf{p} = \mathbf{p}_0$ при $t = 0$, находим

$$\mathbf{p} = \mathbf{p}_0 - e\mathcal{E}t. \tag{6.2}$$

По виду формула (6.2) описывает неограниченное возрастание квазиимпульса. Следует, однако, помнить, что по определению последний изменяется только в пределах первой зоны Бриллюэна. Поэтому слова «неограниченное возрастание» в данном случае лишены смысла. Характер движения электрона легко выяснить,

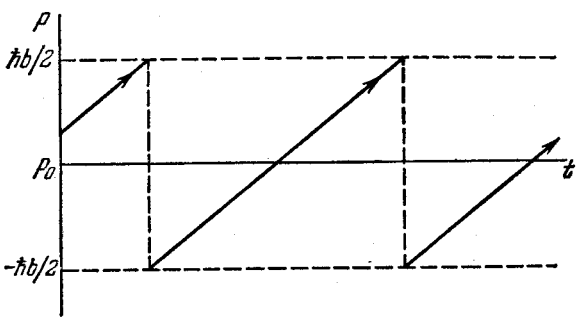


Рис. 4.7. Периодическое движение носителя заряда в зоне Бриллюэна во внешнем электрическом поле (одномерный случай). Сплошные линии со стрелками изображают траекторию носителя в пространстве квазиимпульсов.

вспоминая, что точки, разделенные вектором $\hbar\mathbf{b}$, описывают одно и то же состояние электрона. Следовательно, попав на границу зоны Бриллюэна, электрон, тем самым, попадает и на противоположную ее границу, откуда вновь начинается увеличение квазиимпульса. Если векторы $\mathcal{E}$ и $\mathbf{b}$ параллельны, то квазиимпульс будет изменяться периодически — так, как это показано (для одномерного случая) на рис. 4.7. Период осцилляций t_0, т. е. время, за которое электрон проходит всю зону Бриллюэна и возвращается в исходное состояние, легко найти из условия

$$\hbar\mathbf{b} = e\mathcal{E}t_0.$$

Отсюда

$$t_0 = \frac{\hbar}{e}\frac{(\mathbf{b}, \mathcal{E})}{\mathcal{E}^2}. \tag{6.3}$$

В частности, в кубическом кристалле, когда вектор $\mathcal{E}$ параллелен одной из главных его осей, мы имеем

$$t_0 = \frac{\hbar}{e\mathcal{E}a}. \tag{6.3'}$$

Периодическому движению электрона по зоне Бриллюэна должны соответствовать и осцилляции в энергетической зоне. В этом легко убедиться, полагая, в соответствии с (1.7) и (1.3),

$$\frac{\partial E(\mathbf{p})}{\partial t} = -e(\mathcal{E}\nabla_{\mathbf{p}} E(\mathbf{p})). \tag{6.4}$$

Общее решение этого дифференциального уравнения есть произвольная функция аргумента $\mathbf{p} - e\mathcal{E}t$. Поскольку при $\mathcal{E} = 0$ должна получиться просто энергия электрона в идеальной решетке, мы имеем

$$E(\mathbf{p}, t) = E(\mathbf{p} - e\mathcal{E}t). \tag{6.5}$$

Как мы знаем, $E(\mathbf{p})$ есть периодическая функция с периодом $\hbar\mathbf{b}$. Отсюда сразу следует, что электрон в постоянном и однородном электрическом поле, параллельном одному из векторов обратной решетки, осциллирует по энергетической зоне с периодом, определяемым формулой (6.3).

Обратим внимание на приближённый характер проделанного только что расчета. Именно, мы использовали формулу (1.7), подставив в нее выражение (1.3) для средней скорости электрона. Последнее, однако, было получено для электрона, движущегося в чисто периодическом поле. Отсюда следует, что равенство (6.5) (как и само представление об осцилляциях электрона по зоне) приближенно справедливо, лишь если напряженность «кристаллического» поля, формирующего зонный спектр, велика по сравнению с $\mathcal{E}$ *).

Далее, мы не принимали во внимание рассеяние носителей заряда неизбежно существующими несовершенствами решетки. При наличии его представление об осцилляциях электронов и дырок по зонам сохраняет смысл лишь при условии

$$t_0 \ll \tau_{\mathbf{p}}. \tag{6.6}$$

До сих пор мы рассматривали нестационарное движение носителей заряда. Можно поставить и задачу об их стационарных состояниях, т. е. об энергетическом спектре носителей заряда в постоянном и однородном электрическом поле. В рассматриваемом случае потенциальная энергия электрона имеет вид

$$U(\mathbf{r}) + e\mathcal{E}z, \tag{6.7}$$

*) По порядку величины напряженность «кристаллического» поля составляет в среднем около 10^7 В/см.

где $U(\mathbf{r})$, как и раньше, соответствует электрону в решетке, а поле $\mathcal{E}$ предполагается направленным вдоль оси Z. Коль скоро выполняется неравенство (1.6), второе слагаемое в (6.7) меняется весьма медленно по сравнению с первым. Следовательно, локально — в каждой не слишком большой области кристалла — потенциальную энергию электрона во внешнем поле можно рассматривать как почти постоянную, и мы можем написать собственные значения энергии в виде

$$E_l = E_l(\mathbf{p}) + e\mathcal{E}z. \tag{6.8}$$

Таким образом, локально здесь сохраняется представление о зонном энергетическом спектре электрона. Однако на больших интервалах z (порядка $E_g/e\mathcal{E}$) изменение второго слагаемого в (6.8) становится заметным. Это означает, что зоны, оставаясь практически неизменными по форме, должны наклониться так, как показано на рис. 4.8 (для простоты изображены только границы двух зон — проводимости и валентной).

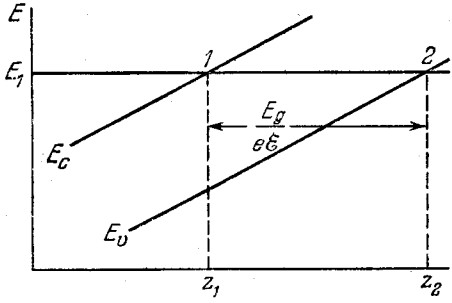


Рис. 4.8. Наклонные энергетические зоны в постоянном и однородном электрическом поле. Ось Z направлена вдоль вектора $\mathcal{E}$.

Эти рассуждения справедливы и в случае неоднородного электрического поля, если только оно достаточно плавное: напряженность поля должна медленно меняться на протяжении характерной длины волны электрона. Действительно, в этом случае потенциальную энергию электрона в рассматриваемом поле можно по-прежнему рассматривать как величину, локально постоянную. На достаточно больших расстояниях, однако, различие между однородным и неоднородным полями становится заметным: в последнем случае границы зон оказываются непрямолинейными. Об этом говорят как об искривлении зон. Представление об искривленных зонах весьма часто используется в задачах, в которых играют роль пространственно неоднородные электрические поля (гл. VI — XI). Оно позволяет, в частности, пользоваться уравнениями движения (1.9), (1.10)—(1.10″) не только в пространственно однородных, но и в плавных неоднородных полях.

Из рассмотрения рис. 4.8 вытекает, что запрещенной зоны в том смысле, в каком этот термин употреблялся раньше, здесь, строго говоря, нет. Действительно, для любого значения энергии можно найти область пространства, в которой оно попадает, например, в зону проводимости; в другой области изменения z оно же

попадает в валентную зону. Представление о границах зон тем не менее сохраняет известный смысл.

Чтобы выяснить его, вернемся к одномерному примеру, рассмотренному в § III.5 (результаты будут иметь и общее значение). В этом случае верхняя и нижняя границы энергии электрона в данной зоне (в чисто периодическом поле) соответствуют максимальному и минимальному значениям выражения (III.5.12″). При вещественных значениях $\lambda = kd$ выражение (III.5.12″) ограничено и зона имеет конечную ширину. Легко видеть, однако, что выражение (III.5.10) удовлетворяет системе уравнений (III.5.8) и при мнимых λ. При этом, очевидно, $\cos\lambda$ заменяется на $\operatorname{ch}|\lambda|$, т. е. абсолютная величина функции (III.5.12″) может быть сколь угодно большой. Следовательно, если бы были допустимы мнимые значения квазиволнового числа k, то ширина разрешенной зоны была бы бесконечно велика.

Мы знаем, однако, что в силу условия нормировки (III.2.8) компоненты квазиволнового вектора должны быть вещественными. Действительно, в противном случае волновая функция электрона содержала бы множитель

$$\exp(-z\,\mathrm{Im}\,k). \tag{6.9}$$

При любом знаке $\mathrm{Im}\,k$ выражение (6.9) расходится на бесконечности — либо при $z \to +\infty$, либо при $z \to -\infty$. Единственный способ избежать этой расходимости состоит в том, чтобы выбрать в данной области энергий тривиальное решение: $\psi = 0$. Это и означает, что электрон в запрещенной зоне находиться не может: вероятность обнаружить его там равна нулю.

Итак, значениям энергии, лежащим в запрещенной зоне, формально можно сопоставить ненулевые решения уравнения Шредингера. Эти решения, однако, не осциллируют, а возрастают при одном из способов стремления координаты z к бесконечности, почему и исключаются условием (III.2.8), заменяясь тривиальным $\psi = 0$. Иначе говоря, понятие «запрещенная зона» можно определить двояко:

а) как совокупность значений энергии, не допускаемых условием (III.2.8) и потому в идеальном кристалле физически не реализующихся;

б) как совокупность значений энергии, которым отвечают ненулевые решения с мнимым квазиимпульсом — затухающие или возрастающие на бесконечности.

Для идеального кристалла в отсутствие внешних полей эти два определения идентичны. При наложении электрического поля, однако, положение меняется. Первое определение теперь отпадает, второе же — остается в силе. Действительно, коль скоро потенциальная энергия электрона во внешнем поле остается почти

постоянной на протяжении многих постоянных решетки, волновые функции его локально будут иметь почти такой же вид, как и в отсутствие поля. Следовательно, сохраняется деление волновых функций на осциллирующие («разрешенная» зона) и затухающие или возрастающие («запрещенная» зона). Границы зон при этом будут зависеть от координат. (Для разрешенной зоны, понимаемой как совокупность дозволенных значений энергии, это было бы бессмысленно: собственные значения уравнения Шредингера, по определению, от координат не зависят.)

Из рис. 4.8 видно, почему в рассматриваемом случае условие ограниченности решения на бесконечности не исключает состояний с мнимым квазиимпульсом. Дело в том, что благодаря наклону зон электрон находится в запрещенной зоне лишь в ограниченной области пространства (так, электрон с энергией E_1 находится в запрещенной зоне на участке $z_1 < z < z_2$). Далее, для любой точки на границе дозволенной зоны запрещенная зона лежит со стороны только меньших или только бóльших значений z (см., например, точки *2* и *1* на рис. 4.8). Следовательно, существует физически осмысленное решение, при котором $z \operatorname{Im} k > 0$, т. е., согласно (6.9), волновая функция убывает по мере проникновения в глубь запрещенной зоны.

Иначе говоря, при наложении электрического поля вероятность обнаружить электрон в запрещенной зоне становится отличной от нуля, но она все же сравнительно невелика. В этом и состоит смысл термина «запрещенная зона» при наличии внешнего электрического поля.

Из сказанного явствует, что электрон во внешнем поле может, не совершая работы, перейти из одной разрешенной зоны в другую. Пусть, например, в начальный момент времени электрон с энергией E_1 находится на границе валентной зоны (в точке *2* рис. 4.8). Согласно (6.9) волновая функция его в запрещенной зоне теперь не равна нулю тождественно, а пропорциональна $\exp\{-(z_2 - z) \operatorname{Im} k\}$. Следовательно, существует конечная вероятность «просачивания» электрона через запрещенную зону в точку *1*, расположенную уже на границе зоны проводимости. Это явление в известной мере аналогично хорошо известному из квантовой механики туннельному переходу электрона из одной потенциальной ямы в другую сквозь разделяющий их потенциальный барьер. Соответственно оно получило название *туннельного эффекта* *). Вероятность его $\mathcal{P}$ тем больше, чем меньше расстояние $z_2 - z_1$, т. е. чем сильнее наклонены зоны. Вычисления, на которых мы не будем останавливаться (см. [1, 3]), показывают, что с точностью

*) Следует, однако, помнить, что аналогия с обычным туннельным эффектом не полна: потенциального барьера в обычном смысле слова запрещенная зона не представляет.

до несущественного предэкспоненциального множителя

$$\mathscr{P} \sim \exp\left(-\frac{\mathscr{E}_0}{\mathscr{E}}\right), \tag{6.10}$$

где $\mathscr{E}_0$ — постоянная, зависящая от параметров кристалла (от ширины запрещенной зоны, расположения экстремумов в зоне Бриллюэна, эффективных масс электрона в зонах проводимости и валентной и т. д.).

Пусть, например, изоэнергетические поверхности вблизи дна зоны проводимости и потолка валентной зоны суть сферы с центрами в центре зоны Бриллюэна; соответствующие эффективные массы обозначим через m_n и $-m_p$ ($m_p > 0$). Тогда

$$\mathscr{E}_0 = \frac{\pi E_g^{3/2} m_r^{1/2}}{2\hbar e}, \tag{6.11}$$

где $m_r = \dfrac{m_n m_p}{m_n + m_p}$ есть приведенная эффективная масса электрона и дырки. При $E_g = 1$ эВ и $m_r = 0{,}1 m_0$ формула (6.11) дает $\mathscr{E}_0 \simeq \simeq 10^7$ В/см.

Видно, что в слабых полях ($\mathscr{E} \ll \mathscr{E}_0$) вероятность туннельного эффекта совершенно ничтожна и им можно пренебречь. При этом достаточно рассматривать движение электрона только в одной зоне, т. е. справедливы рассуждения, изложенные в начале этого параграфа. С увеличением напряженности поля вероятность туннельного эффекта быстро возрастает, и при $\mathscr{E} \simeq 0{,}1\,\mathscr{E}_0$ его уже надо принимать во внимание.

Туннельный эффект может иметь место и в неоднородных электрических полях. С ним связан ряд явлений, наблюдаемых в полупроводниках. Так, он обусловливает один из возможных механизмов пробоя беспримесного образца: в достаточно сильном поле происходит спонтанное образование электронов проводимости и дырок в валентной зоне *). Далее, на этом эффекте основано действие туннельных диодов (см. § VIII. 3).

Наконец, некоторые особенности поглощения света полупроводниками в сильном поле также объясняются туннельным эффектом (§ XVIII.10).

§ 7. Мелкие примесные уровни в гомеополярном кристалле

Рассмотрим простейшую задачу теории реальных кристаллов — задачу о мелких уровнях, создаваемых в запрещенной зоне изолированными атомами посторонней примеси. Мелкими будем называть

*) Этот механизм — не единственный. По-видимому, в ряде случаев более существенную роль играет механизм ударной ионизации; электрон проводимости, разгоняемый электрическим полем, приобретает энергию, достаточную для того, чтобы «вышибить» другой электрон из валентной зоны.

уровни, удовлетворяющие неравенству (4.5): расстояние от уровня до границы ближайшей к нему зоны должно быть мало по сравнению с шириной запрещенной зоны. Такие уровни возникают, например, в германии и кремнии при введении туда примесей элементов V и III групп периодической системы (§ II.9). Для расчета этих уровней можно воспользоваться уравнением (4.3) или его обобщениями (4.8) и (4.10). В качестве $\delta U(\mathbf{r})$ здесь следует взять потенциальную энергию взаимодействия электрона с атомным остатком донора (или дырки — с акцептором). Точный вид δU зависит от химической природы примеси, и в общем случае его трудно определить. Для мелких уровней, однако, это обстоятельство не очень существенно, ибо, как мы увидим, главную роль в задаче играют расстояния $\mathbf{r}$ между электроном и ядром донора, значительно превышающие линейный размер атомного остатка. Как известно из электростатики, при этом детали, связанные с конкретным видом распределения заряда в атомном остатке, уже не играют роли, и функция $\delta U(\mathbf{r})$ дается обычным законом Кулона:

$$\delta U = -\frac{Ze^2}{\varepsilon r}. \tag{7.1}$$

Здесь Z — заряд атомного остатка в единицах e (в интересующем нас случае $Z = 1$), ε — диэлектрическая проницаемость решетки *).

Мы получаем, таким образом, задачу об атоме водорода — с той лишь разницей, что в рассматриваемом случае в выражении (7.1) фигурирует диэлектрическая проницаемость среды, а оператор кинетической энергии, вместо истинной массы электрона, содержит эффективную (или, общее, тензор обратной эффективной массы). Соответственно о модели примесного центра, основанной на использовании выражения (7.1) и метода эффективной массы, говорят как о водородной модели.

Отметим сразу же одно качественное следствие из этой модели, справедливое при любой анизотропии зон: расположение уровней не должно зависеть от химической природы примеси (разумеется, в пределах данной группы периодической системы — например, V группы для доноров и III группы для акцепторов в германии и кремнии). Таким образом, относительная разность между энергиями ионизации различных примесей этого типа в данном кристалле может служить мерой точности водородной модели. Данные таблицы 2.1 показывают, что для примесей III и V групп в Ge водородная модель дает неплохое приближение.

---

*) В рассматриваемом нами случае гомеополярных кристаллов вся поляризуемость безынерционная. Для мелких уровней область дисперсии ε соответствует частотам, значительно превышающим характерное значение $\mathscr{J}/\hbar$, где $\mathscr{J}$ — энергия ионизации примеси. Это позволяет считать ε просто константой, равной своему статическому значению.

Конкретный расчет особенно прост в случае изотропной эффективной массы. Подставляя выражение (7.1) (при $Z = 1$) в уравнении (4.3), мы имеем

$$-\frac{\hbar^2}{2m}\nabla^2\chi - \frac{e^2}{\varepsilon r}\chi = E\chi. \tag{7.2}$$

Введем характерные единицы длины и энергии:

$$a_B = \frac{\varepsilon\hbar^2}{me^2}, \qquad E_B = \frac{me^4}{2\varepsilon^2\hbar^2}. \tag{7.3}$$

Величины a_B и E_B обычно называют боровским радиусом и боровской энергией в кристалле.

Будем измерять длину в единицах a_B, а энергию — в единицах E_B. Тогда уравнение (7.2) примет вид (мы сохраняем прежние обозначения для безразмерных переменных)

$$-\nabla^2\chi - \frac{2}{r}\chi = E\chi. \tag{7.2'}$$

Решение уравнения (7.2') можно найти в курсах квантовой механики.

Имеются два участка спектра, соответствующие тому или другому знаку энергии E.

а) Непрерывный спектр: $E > 0$.

Здесь все значения E дозволены; функция χ есть кулоновская волновая функция непрерывного спектра.

Согласно сказанному в § 4 эта область энергий соответствует зоне проводимости. Отличие функции χ от плоской волны (4.4) описывает влияние примеси на движение свободных носителей заряда.

б) Дискретный спектр: $E < 0$.

Дозволенные значения энергии даются выражением

$$E_n = -\frac{E_B}{n^2}, \qquad n = 1, 2, \ldots \tag{7.4}$$

«Сглаженная» функция χ в данном случае есть кулоновская волновая функция дискретного спектра. В частности, при $n = 1$ (низший, или основной, уровень) мы имеем (в справедливости этого решения можно убедиться и простой подстановкой)

$$\chi = (\pi a_B^3)^{-1/2}\exp(-r/a_B). \tag{7.5}$$

Постоянный множитель здесь выбран в соответствии с обычным условием нормировки

$$\int|\chi|^2\,d\mathbf{r} = 1.$$

Поскольку начало отсчета энергии у нас совмещено с дном зоны проводимости E_c, дискретные уровни (7.4) лежат в запрещен-

ной зоне. Электрон, занимающий один из таких уровней, локализован вблизи атома примеси (соответственно чему эти уровни иногда называют локальными). Наибольшая степень локализации соответствует основному состоянию ($n = 1$). Тогда, согласно (7.5), «радиус орбиты», т. е. наиболее вероятное расстояние между атомом примеси и электроном, равен a_B. При бо́льших значениях n «радиус орбиты» есть na_B. Из предыдущего ясно, что весь расчет имеет смысл лишь при достаточно слабой локализации электрона. Именно, формула (7.1) оправдана, если

$$a_B \gg a. \tag{7.6}$$

Это же условие обеспечивает и возможность использовать метод эффективной массы и представление о диэлектрической проницаемости.

Энергия ионизации $\mathscr{I}$, т. е. работа, которую надо совершить, чтобы перевести электрон на дно зоны проводимости, есть, очевидно, $-E_1$:

$$\mathscr{I} = E_B = \frac{e^2}{2\varepsilon a_B} = 13{,}7\,\frac{m}{m_0}\,\varepsilon^{-2}\ \text{эВ}. \tag{7.7}$$

Для германия $\varepsilon = 16$. Полагая еще $m = 0{,}2m_0$, мы получаем из (7.7) $\mathscr{I} = 1{,}07 \cdot 10^{-2}$ эВ, что близко к экспериментальным данным для мелких доноров (табл. 2.1). По этой причине для ориентировочной порядковой оценки величин, характеризующих электроны проводимости в германии, иногда пользуются изотропной моделью с эффективной массой $m = 0{,}2m_0$. Согласно (7.3) при этом $a_B = 40$ Å, что удовлетворяет неравенству (7.6).

Строгий расчет мелких донорных и акцепторных уровней в германии и кремнии должен основываться на уравнениях (4.8) и (4.10). Здесь приходится прибегать к приближенным методам расчета, детали которых можно найти в оригинальных статьях [4, 5, 6].

До сих пор мы явно предполагали, что носитель заряда взаимодействует лишь с одним примесным атомом. Фактически атомов примеси всегда много, и, строго говоря, вместо (7.1) надо было бы писать

$$\delta U(\mathbf{r}) = -\sum_i \frac{Z_i e^2}{\varepsilon\,|\mathbf{r} - \mathbf{R}_i|}, \tag{7.1'}$$

где индекс i нумерует атомы примеси, а $\mathbf{R}_i$ есть радиус-вектор i-го примесного атома. Переход от (7.1') к (7.1) оправдан лишь при достаточно малой концентрации примеси N_i. Именно, пусть носитель заряда локализован на n-м уровне около какого-то примесного атома. Тогда энергия его взаимодействия со всеми остальными атомами примеси должна быть (по абсолютной величине) мала по сравнению с величиной $|E_n|$ (7.4). По порядку величины названная энергия составляет $e^2 N_i^{1/3} \varepsilon^{-1}$ (так как среднее расстояние

между примесными атомами — порядка $N_t^{-1/3}$). Таким образом, должно выполняться неравенство $\varepsilon E_B n^{-2} \geqslant e^2 N_t^{1/3}$ или, в силу (7.3)*),

$$a_B n^2 \ll N_t^{-1/3}. \tag{7.8}$$

При нарушении неравенства (7.8) взаимодействие между различными атомами примеси приводит к сдвигу и расщеплению уровней (7.4).

Особый интерес представляют явления, происходящие при перекрытии волновых функций носителей, локализованных у различных атомов примеси. Такое перекрытие становится существенным, когда среднее расстояние между атомами оказывается сравнимым с радиусом соответствующей орбиты:

$$N_t^{-1/3} \sim n a_B. \tag{7.9}$$

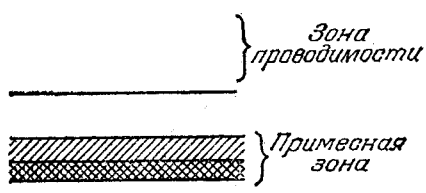


Рис. 4.9. Примесная зона доноров заштрихована. Двойной штриховкой показаны заполненные уровни в примесной зоне.

При этом происходит в качественном отношении то же самое, что и при образовании кристалла из атомов (ср. § III.5): появляется возможность непосредственного (без участия зоны проводимости) туннельного перехода электрона (дырки) с одного донора (акцептора) на другой. В результате дискретные примесные уровни «размываются», образуя дозволенную область, лежащую в запрещенной зоне (подобно тому как «размываются» уровни регулярных атомов). Эта область называется *примесной зоной* (рис. 4.9).

Заметим, однако, что аналогия с образованием «собственных» зон кристалла не полна, ибо атомы примеси, вообще говоря, не образуют периодической решетки. По этой причине было бы, например, неправильно приписывать электрону в примесной зоне строго определенный квазиимпульс.

При образовании примесной зоны становится возможным новый механизм электропроводности — по примесной зоне. Для этого примесная зона не должна быть целиком заполненной. Так обстоит дело, если кроме основной примеси, скажем, мелких доноров (концентрация N_d), в полупроводнике имеются еще и компенсирующие акцепторы (с концентрацией $N_a < N_d$), уровни энергии которых расположены ниже уровней доноров. Тогда при низких температурах все акцепторы будут заполнены электронами, доноры же частично (с концентрацией N_a) окажутся пустыми и станут

*) Фактически неравенство (7.8) может быть несколько ослаблено благодаря учету взаимодействия электронов друг с другом. Этот эффект приводит, в частности, к экранированию взаимодействия электронов с донорами, в результате чего поля последних эффективно ослабляются (см. гл. XVII).

возможными переходы электронов с заполненных доноров на пустые. Так как среднее расстояние между примесными центрами больше расстояния между атомами основной решетки, то и вероятности туннельных переходов между атомами примеси меньше, чем между атомами основной решетки. Поэтому подвижность электронов (дырок) в примесной зоне гораздо меньше, чем в «собственных» зонах.

Существование проводимости по примесной зоне можно обнаружить экспериментально, исследуя электрические свойства кристаллов при низких температурах. На рис. 4.10 *) даны температурные зависимости удельного сопротивления ρ и постоянной Холла R в кристаллах дырочного германия с примесями элементов III и V групп. При понижении температуры ρ и R сначала увеличиваются по экспоненциальному закону. Однако в дальнейшем зависимость ρ от T становится слабой, а R проходит через максимум (в рассматриваемом примере около 5,5 К).

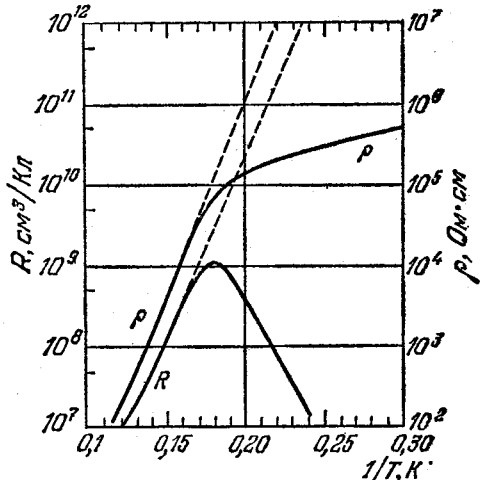


Рис. 4.10. Удельное сопротивление ρ и постоянная Холла R при низких температурах в германии. $N_a = 4{,}4 \cdot 10^{15}$ см$^{-3}$, $N_d = 1{,}0 \cdot 10^{14}$ см$^{-3}$.

Такое поведение ρ и R объясняется тем, что полная концентрация электронов на донорах ($N_d - N_a$) (для определенности мы говорим о полупроводнике n-типа) распределяется между зоной проводимости (концентрацию в которой обозначим через n_1) и примесной зоной (концентрация n_2). Тогда $\rho = 1/e\,(n_1\mu_1 + n_2\mu_2)$, а постоянная Холла выражается формулой (I.4.5). При сравнительно высоких температурах $n_1\mu_1 > n_2\mu_2$ и доминирует электропроводность в зоне проводимости. Если бы примесные уровни были дискретными ($\mu_2|_{T=0} = 0$), то при $T \to 0$ мы имели бы $\rho \to \infty$, $R \to \infty$, так как $n_1 \to 0$ (пунктир на рис. 4.10). При очень низких температурах, когда $n_1\mu_1 \ll n_2\mu_2$, главный вклад в электропроводность вносит примесная зона, концентрация электронов (дырок) в которой практически не зависит от температуры. При $T \to 0$ величины ρ и R должны стремиться к конечным значениям:

$$\rho = \frac{1}{e\,(N_d - N_a)\,\mu_2}, \qquad R = -\frac{\mu_{2H}}{\mu_2}\,\frac{1}{ce\,(N_d - N_a)}.$$

*) По данным работы *H. Fritzsche*, Phys. Rev. 99, № 2, 406 (1955).

Существование максимума постоянной Холла следует из формулы (I.4.5) при учете условия $n_1 + n_2 = \text{const}$. Так, например, если бы подвижности не зависели от температуры, а холловские факторы в обеих зонах были равны: $\mu_{1H}/\mu_1 = \mu_{2H}/\mu_2$, то выражение (I.4.5) достигало бы максимума при $n_1\mu_1 = n_2\mu_2$.

Примесная зона возникает сначала в результате перекрытия волновых функций возбужденных состояний. По мере увеличения концентрации примеси «размываются» все более и более низколежащие уровни — вплоть до основного. Наконец, при достаточно больших значениях N_t примесная зона перекрывается с ближайшей «собственной» зоной кристалла — зоной проводимости или валентной. В германии, легированном донорами V группы, такое перекрытие наступает при $N_t \approx 10^{17}$ см$^{-3}$. При этом деление на примесную и собственную зоны теряет смысл — возникает единая область непрерывного спектра, и вещество ведет себя в электрическом отношении как плохой металл. Полупроводники такого типа называют сильно легированными. Они рассматриваются в гл. XIX.

Глава V

СТАТИСТИКА ЭЛЕКТРОНОВ И ДЫРОК В ПОЛУПРОВОДНИКАХ

§ 1. Введение

В предыдущих главах были рассмотрены различные возможные квантовые состояния электронов в кристалле, т. е. различные типы стационарных движений электронов. В состоянии термодинамического равновесия для данного образца кристалла при заданной температуре существует определенное распределение электронов по различным квантовым состояниям. В результате в кристалле устанавливается определенная концентрация свободных электронов n в зоне проводимости и концентрация свободных дырок p в дырочной зоне. Кроме того, в кристалле, содержащем локальные уровни энергии (примесные атомы и структурные дефекты), могут быть еще отрицательно заряженные акцепторы, на которых в каждой единице объема находится некоторое количество связанных электронов n_t, и положительно заряженные доноры, содержащие некоторую концентрацию связанных дырок p_t.

Вычисление этих концентраций подвижных и связанных носителей заряда составляет основную задачу статистики электронов и дырок в кристаллах.

Решение указанной задачи необходимо для понимания многих электрических и оптических явлений в полупроводниках. В частности, оно позволяет выяснить зависимость основных электрических свойств полупроводника (электропроводности, подвижности и др.) от количества и состава примесей, и от температуры. И, наоборот, анализируя с помощью теоретических соотношений статистики экспериментальные данные о температурной зависимости концентраций электронов и дырок, оказывается возможным найти энергетические уровни, создаваемые примесными атомами и структурными дефектами, а также их концентрации.

Рассматриваемая задача распадается на две части: чисто квантовомеханическую — нахождение числа возможных квантовых состояний электронов и статистическую — определение фактического распределения электронов по этим квантовым состояниям при термодинамическом равновесии.

§ 2. Распределение квантовых состояний в зонах

Как показано в гл. III, стационарные состояния электрона в идеальном кристалле характеризуются квазиимпульсом $\mathbf{p} = \hbar\mathbf{k}$ и номером зоны l. При этом объем в зоне Бриллюэна, приходящийся на каждое значение $\mathbf{p}$, равен $(2\pi\hbar)^3/V$, где V — объем кристалла. Поэтому число квантовых состояний, соответствующее элементу объема в зоне Бриллюэна $d\mathbf{p}$ и рассчитанное на единицу объема кристалла, есть

$$2\,\frac{d\mathbf{p}}{(2\pi\hbar)^3/V}\,\frac{1}{V} = 2\,\frac{d\mathbf{p}}{(2\pi\hbar)^3}. \tag{2.1}$$

Здесь множитель 2 учитывает две возможные ориентации спина.

В термодинамическом равновесии вероятность заполнения квантовых состояний зависит только от их энергии E и температуры, поэтому в (2.1) в качестве $d\mathbf{p}$ часто удобно выбирать элемент объема, заключенный между двумя бесконечно близкими изоэнергетическими поверхностями (рис. 5.1). Число квантовых состояний, соответствующих заданному интервалу энергий E, $E + dE$ и по-прежнему рассчитанное на единицу объема кристалла, можно представить в виде $N(E)\,dE$, где $N(E)$ есть плотность состояний.

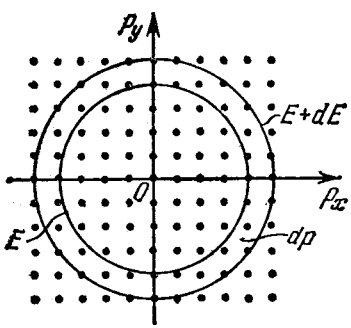


Рис. 5.1. К определению плотности квантовых состояний в зоне.

Точное вычисление функции $N(E)$ в общем случае не просто, так как поверхности постоянной энергии имеют сложную форму. Однако во многих случаях оказывается достаточным точно знать функцию $N(E)$ только вблизи краев зон, отчего задача существенно упрощается.

Рассмотрим сначала простейший случай изотропного параболического закона дисперсии (обобщение результатов на случай анизотропного закона дисперсии будет дано в § 7). Тогда для полной энергии электрона можно написать

$$E = E_c + \frac{p^2}{2m_n}, \tag{2.2}$$

при этом квазиимпульс $\mathbf{p}$ отсчитывается от его значения, соответствующего E_c. Далее $d\mathbf{p}$ есть объем сферического слоя $4\pi p^2 dp$. Из (2.2) имеем

$$p = (2m_n)^{1/2}(E - E_c)^{1/2}, \qquad d\mathbf{p} = \frac{1}{2}(2m_n)^{1/2}(E - E_c)^{-1/2}\,dE\,.$$

Поэтому

$$N_c(E) = \frac{1}{2\pi^2\hbar^3}(2m_n)^{3/2}(E - E_c)^{1/2}. \tag{2.3}$$

Распределение квантовых состояний в валентной (дырочной) зоне получается с помощью аналогичных рассуждений. Рассматривая по-прежнему случай изотропного параболического закона дисперсии, мы имеем

$$E = E_v - \frac{1}{2m_p} p'^2. \tag{2.4}$$

Поэтому для плотности состояний вблизи края валентной зоны вместо (2.3) получается

$$N_v(E) = \frac{1}{2\pi^2\hbar^3} (2m_p)^{3/2} (E_v - E)^{1/2}. \tag{2.5}$$

Более общие выражения для плотности состояний приводятся в § 7.

§ 3. Распределение Ферми — Дирака

Электроны, как частицы, обладающие полуцелым спином, подчиняются, как известно, статистике Ферми — Дирака. Вероятность того, что электрон будет находиться в квантовом состоянии с энергией E, выражается функцией Ферми — Дирака

$$f(E, T) = \left(1 + \exp \frac{E - F}{kT}\right)^{-1}. \tag{3.1}$$

Здесь F — электрохимический потенциал, или уровень Ферми, который, в общем случае, тоже зависит от температуры.

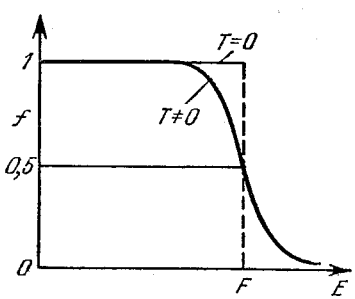


Рис. 5.2. Функция Ферми — Дирака.

Уровень Ферми тесно связан с термодинамическими характеристиками системы. Так, например, его можно определить как приращение свободной энергии системы при добавлении к ней одного электрона при условии постоянства объема и температуры (см. § 11). Однако мы не будем пока входить подробнее в обсуждение термодинамического смысла F, а будем рассматривать ее как некоторую характерную энергию, зависящую от типа полупроводника, его состава и состояния (температуры, давления и др.), значение которой подлежит дальнейшему определению. Из выражения (3.1) видно, что уровень Ферми можно также определить как энергию такого квантового состояния, вероятность заполнения которого при данных условиях равна $^1/_2$.

Вид функции Ферми — Дирака схематически показан на рис. 5.2. При $T = 0$ она имеет вид разрывной функции. Для $E < F$ она равна 1, а, значит, все квантовые состояния с такими энергиями заполнены электронами. Для $E > F$ функция $f = 0$ и соответствую-

щие квантовые состояния совершенно не заполнены. При $T \neq 0$ функция Ферми изображается непрерывной кривой и в узкой области энергий, порядка нескольких kT, в окрестности точки $E = F$ быстро изменяется от 1 до 0. Размытие функции Ферми тем больше, чем выше температура.

Вероятность того, что данное квантовое состояние с энергией E не занято электроном, f_p, есть

$$f_p = (1 - f) = \left(1 + \exp \frac{F - E}{kT}\right)^{-1}. \tag{3.1а}$$

Вычисление различных статистических величин значительно упрощается, если F лежит в запрещенной зоне энергий и удален от края зоны E_c хотя бы на $(2 \div 3)kT$. Тогда в распределении (3.1) можно считать

$$\exp \frac{E - F}{kT} \gg 1$$

и оно переходит в распределение Максвелла — Больцмана классической статистики

$$f = C \exp\left(-\frac{E}{kT}\right), \tag{3.1б}$$

где C — постоянная. В этом случае мы говорим, что электронный газ не вырожден. Аналогично, в полупроводнике p-типа для отсутствия вырождения дырочного газа необходимо, чтобы уровень Ферми F тоже лежал внутри запрещенной зоны и был расположен выше энергии E_v хотя бы на $(2 \div 3)kT$. Противоположный случай, когда уровень Ферми расположен внутри зоны проводимости или внутри валентной зоны, есть случай вырожденного электронного или, соответственно, дырочного газа. В этом случае существенно необходимо пользоваться распределением Ферми — Дирака (3.1) или (3.1а).

В дальнейшем при вычислении концентраций подвижных и связанных носителей заряда мы разделим задачу на две части. А именно, сначала мы выразим эти концентрации через уровень Ферми, а затем рассмотрим, как определить положение самого уровня Ферми.

§ 4. Концентрации электронов и дырок в зонах

Согласно сказанному выше концентрация электронов в зоне проводимости равна

$$n = \int_{E_c}^{\infty} N_c(E) f(E, T)\, dE. \tag{4.1}$$

Отметим, что в качестве верхнего предела в написанном интеграле мы должны были бы взять энергию верхнего края зоны проводи-

мости. Однако так как функция f для значений энергий $E > F$ быстро убывает с увеличением E, то замена верхнего предела на ∞ практически не меняет значения интеграла.

Интеграл (4.1) удобно представить в следующем виде. Выберем новую переменную интегрирования

$$x = \frac{E - E_c}{kT} \qquad (0 \leqslant x \leqslant \infty).$$

Обозначим, далее,

$$\zeta = F - E_c, \quad \zeta^* = \frac{\zeta}{kT}. \tag{4.2}$$

Величина ζ носит название *химического потенциала для электронов*, а ζ^* есть его безразмерное значение. Обозначим, наконец, для сокращения

$$N_c = 2\left(\frac{2\pi m_n kT}{(2\pi\hbar)^2}\right)^{3/2}. \tag{4.3}$$

Эта величина получила название *эффективной плотности состояний в зоне проводимости* (смысл этого названия будет выяснен ниже). Тогда нетрудно убедиться, что интеграл (4.1) принимает вид

$$n = N_c \Phi_{1/2}(\zeta^*), \tag{4.4}$$

где

$$\Phi_{1/2}(\zeta^*) = \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{x^{1/2}\, dx}{1 + \exp(x - \zeta^*)}. \tag{4.5}$$

Значение интеграла (4.5) зависит только от параметра ζ^*, т. е. от химического потенциала, и температуры. Этот интеграл известен как интеграл Ферми — Дирака (или, точнее, интеграл Ферми — Дирака с индексом $^1/_2$) и в общем случае не выражается через элементарные функции. Таблица его значений приведена в Приложении III.

Рассуждая аналогично, легко найти и выражение для концентрации положительных дырок в валентной зоне. Отличия от предыдущего случая заключаются лишь в том, что мы, во-первых, должны использовать выражение (2.5) для плотности состояний в валентной зоне. Во-вторых, мы должны подсчитывать не число заполненных, а число незанятых состояний. И, наконец, интегрирование по энергиям нужно производить в пределах валентной зоны. Поэтому

$$p = \int_{-\infty}^{E_v} N_v(E) f_p(E, T)\, dE, \tag{4.6}$$

где f_p выражается формулой (3.1а). Вводя, как и выше, безразмерную переменную

$$y = \frac{E_v - E}{kT} \qquad (0 \leqslant y \leqslant \infty)$$

и безразмерный параметр

$$\eta^* = \frac{E_v - F}{kT}, \tag{4.7}$$

мы приходим к формуле, аналогичной соотношению (4.4):

$$p = N_v \Phi_{1/2}(\eta^*). \tag{4.8}$$

Здесь N_v есть эффективная плотность состояний в валентной зоне:

$$N_v = 2\left(\frac{2\pi m_p kT}{(2\pi\hbar)^2}\right)^{3/2}, \tag{4.9}$$

а $\Phi_{1/2}(\eta^*)$ — прежний интеграл Ферми (4.5). Однако теперь он содержит другой параметр η^*, характеризующий положение уровня Ферми относительно края валентной зоны. Разность

$$\eta = E_v - F = -\zeta - E_g$$

есть *химический потенциал для дырок*.

Формулы (4.4) и (4.8) справедливы для полупроводника, однородного по составу и в отсутствие внешних полей. Легко обобщить их на случай, когда в образце имеется электрическое поле с потенциалом φ, плавно (в смысле, указанном в § IV.6) изменяющимся в пространстве. Действительно, в этом случае применимо представление об искривленных зонах и, следовательно, величины E_c и E_v надо заменить на $E_c - e\varphi$ и $E_v - e\varphi$, где E_c и E_v — энергии краев зон при $\varphi = 0$. Поэтому, если ζ_0^* есть безразмерный химический потенциал при $\varphi = 0$, то выражение для концентрации электронов принимает вид

$$n = N_c \Phi_{1/2}\left(\zeta_0^* + \frac{e\varphi}{kT}\right). \tag{4.4а}$$

Аналогично, для концентрации дырок получаем

$$p = N_v \Phi_{1/2}\left(\eta_0^* - \frac{e\varphi}{kT}\right). \tag{4.8а}$$

§ 5. Невырожденные полупроводники

Выражения (4.4) и (4.8) для концентраций значительно упрощаются для невырожденных полупроводников. Такой случай показан на рис. 5.3, где даны плотность квантовых состояний $N_c(E)$, функция Ферми $f(E, T)$, а также их произведение, равное dn/dE (здесь dn — концентрация электронов с энергией в интервале $E, E + dE$). Полное количество электронов в зоне определяется пло-

щадью, ограниченной кривой $N_c(E) f(E, T)$ и осью E. При этом существен только «хвост» распределения Ферми, который может быть аппроксимирован распределением Максвелла — Больцмана.

Для невырожденного полупроводника в формуле (4.5) мы имеем $\exp(x - \zeta^*) \gg 1$, и поэтому интеграл Ферми принимает вид

$$\Phi_{1/2}(\zeta^*) = \exp\zeta^* \cdot \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-x} x^{1/2}\, dx.$$

Входящий сюда интеграл хорошо известен и равен

$$\int_0^\infty e^{-x} x^{1/2}\, dx = 2\int_0^\infty e^{-z^2} z^2\, dz = \frac{1}{2}\sqrt{\pi}.$$

Поэтому

$$\Phi_{1/2}(\zeta^*) = \exp\zeta^* = \exp\frac{F - E_c}{kT},$$

и, следовательно,

$$n = N_c \exp\frac{F - E_c}{kT}. \qquad (5.1)$$

Аналогично упрощается выражение для концентрации дырок в невырожденном полупроводнике. Здесь в интеграле Ферми можно положить $\exp(y - \eta^*) \gg 1$, и поэтому, поступая, как и выше, мы имеем

$$p = N_v \exp\frac{E_v - F}{kT}. \qquad (5.2)$$

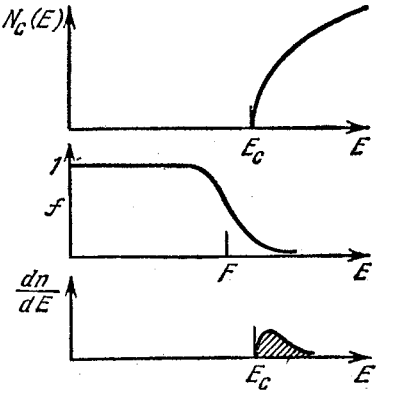


Рис. 5.3. Схематический ход функций $N_c(E)$ и $f(E, T)$ и dn/dE для невырожденного полупроводника n-типа.

Полученные выражения (5.1) и (5.2) разъясняют смысл названия «эффективная плотность состояний» в зонах для величин N_c и N_v. Экспоненциальный множитель в выражении (5.1) для невырожденного полупроводника (распределение Максвелла — Больцмана) дает вероятность заполнения квантового состояния с энергией E_c. Поэтому формула (5.1) обозначает, что для невырожденного полупроводника концентрация подвижных электронов получается такой же, как если бы, вместо непрерывного распределения состояний в зоне, в каждой единице объема было N_c состояний с одинаковой энергией E_c. Аналогично, экспоненциальный множитель в (5.2) выражает вероятность того, что состояние с энергией E_v не занято электроном, и потому формула (5.2) показывает, что при подсчете концентрации дырок валентную зону можно заменить

совокупностью состояний с одинаковой энергией E_v, число которых в каждой единице объема есть N_v.

Полагая в выражении (4.3) m_n равным массе изолированного электрона m_0 и $T = 300$ К, мы получаем $N_v = 2{,}510 \cdot 10^{19}$ см$^{-3}$. Для какой-либо другой температуры и иной эффективной массы мы имеем

$$N_{c\,(v)} = 2{,}510 \cdot 10^{19} \left(\frac{m_{n\,(p)}}{m_0}\right)^{3/2} \left(\frac{T}{300}\right)^{3/2} \text{см}^{-3}, \tag{5.3}$$

где $m_{n(p)}$ — эффективная масса электронов или, соответственно, дырок. В невырожденных полупроводниках концентрация основных носителей мала по сравнению с N_c и N_v. В вырожденных полупроводниках имеет место обратное. Поэтому, сопоставляя измеренные значения концентрации электронов и дырок со значениями N_c и N_v, определяемыми соотношением (5.3), можно сразу установить, является ли данный полупроводник вырожденным или нет.

Так как эффективная масса входит в выражения для n и p только в виде множителя $m^{3/2}$, в то время как уровень Ферми F входит в показатель степени, то отношение n/p зависит главным образом от положения F относительно краев зон. Выражения (5.1) и (5.2) показывают, что концентрация подвижных носителей заряда будет больше в той зоне, к которой ближе расположен уровень Ферми. Носители заряда в этой ближайшей зоне будут основными, поэтому в полупроводниках n-типа уровень Ферми расположен в верхней половине запрещенной зоны, а в полупроводниках p-типа — в нижней половине. Исключение могут составить узкозонные полупроводники, в которых m_n и m_p могут сильно отличаться.

Однако произведение концентраций электронов и дырок для невырожденного полупроводника не зависит от положения уровня Ферми. Согласно (5.1) и (5.2) оно равно

$$np = n_i^2 = N_c N_v \exp\left(-\frac{E_g}{kT}\right). \tag{5.4}$$

Здесь n_i есть концентрация электронов в условиях, когда $n = p$, т. е. в собственном полупроводнике. Соотношение (5.4) широко используется для определения (термической) ширины запрещенной зоны E_g по экспериментальным данным о зависимости собственной концентрации n_i от температуры.

§ 6. Случай сильного вырождения

Другой крайний случай мы имеем при сильном вырождении электронного газа, когда

$$\exp\frac{E_c - F}{kT} \ll 1.$$

В этом случае уровень Ферми лежит внутри зоны проводимости,

а концентрация электронов в зоне $n \gg N_c$. Взаимное расположение кривых $N_c(E)$, $f(E, T)$ и dn/dE для вырожденного полупроводника n-типа схематически показано на рис. 5.4. В этом случае в интеграле Ферми (4.5) $\exp(x - \zeta^*) \ll 1$. Далее, в качестве верхнего предела интеграла можно положить $x_m = (F - E_c)/kT$. Это совершенно точно при $T = 0$, однако справедливо с хорошим приближением и при $T \neq 0$ вследствие быстрого убывания функции Ферми—Дирака при $E > F$. Тогда опять интеграл Ферми вычисляется непосредственно и мы имеем

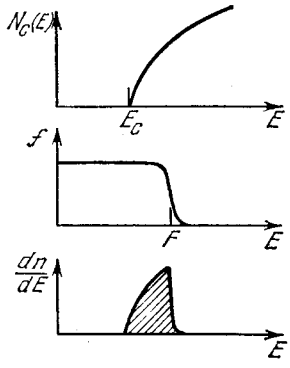


Рис. 5.4. Схематический ход функций $N_c(E)$, $f(E, T)$ и dn/dE в сильно вырожденном полупроводнике n-типа.

$$n = N_c \frac{2}{\sqrt{\pi}} \int_0^{x_m} x^{1/2}\, dx = \frac{4}{3\sqrt{\pi}} N_c x_m^{3/2} = \frac{4}{3\sqrt{\pi}} N_c \left(\frac{F - E_c}{kT}\right)^{3/2}. \qquad (6.1)$$

При температуре абсолютного нуля все состояния в зоне, энергия которых $E > F$, свободны, а все состояния с $E < F$ заняты электронами. Поэтому химический потенциал электронов $\zeta = F - E_c$ есть максимальная энергия электронов при $T = 0$. Эту величину, играющую важную роль в теории металлов, часто называют *энергией Ферми*. Подставляя в (6.1) для N_c его выражение (4.3), находим

$$\zeta = \frac{(3\pi^2)^{2/3} \hbar^2 n^{2/3}}{2m_n}. \qquad (6.2)$$

В случае вырожденного полупроводника p-типа, аналогично, в интеграле Ферми $\Phi_{1/2}(\eta^*)$ можно положить $\exp(y - \eta^*) \ll 1$, а в качестве верхнего предела интеграла можно выбрать $y_m = (E_v - F)/kT$. Тогда вместо (6.2) мы получим

$$\eta = \frac{(3\pi^2)^{2/3} \hbar^2 p^{2/3}}{2m_p}. \qquad (6.3)$$

§ 7. Эффективная масса плотности состояний

Во всех предыдущих параграфах мы считали закон дисперсии изотропным и параболическим. Весьма часто, однако, приходится иметь дело и с более сложными случаями. Так, например, для зоны проводимости в германии и кремнии изоэнергетические поверхности суть эллипсоиды вращения (см. § III.9). Центры этих эллипсоидов не совпадают с центром зоны Бриллюэна, и поэтому в ней имеется несколько эквивалентных минимумов энергии. Изоэнерге-

тические поверхности для дырок в этих полупроводниках имеют еще более сложную форму. Поэтому мы рассмотрим, что́ именно нужно понимать под величинами m_n и m_p, входящими в выражения эффективных плотностей состояний N_c и N_v при анизотропном или непараболическом законе дисперсии.

Рассмотрим случай эллиптических изоэнергетических поверхностей

$$E(\mathbf{p}) = E_c + \frac{p_x^2}{2m_x} + \frac{p_y^2}{2m_y} + \frac{p_z^2}{2m_z}, \tag{7.1}$$

где составляющие квазиимпульса p_x, p_y и p_z отсчитываются от их значений в одном из эквивалентных минимумов в зоне Бриллюэна. Тогда, согласно формуле (2.1), концентрация электронов в зоне, обусловленная одним эквивалентным минимумом, равна

$$n = \frac{2}{(2\pi\hbar)^3} \int \frac{dp_x\, dp_y\, dp_z}{1 + \exp\dfrac{E - F}{kT}}, \tag{7.2}$$

где E теперь выражается формулой (7.1).

Написанное выражение можно просто вычислить для невырожденных полупроводников. В этом случае

$$n = \frac{2}{(2\pi\hbar)^3} \exp\frac{F - E_c}{kT} \cdot \int \exp\left(-\frac{E - E_c}{kT}\right) dp_x\, dp_y\, dp_z.$$

При подстановке вместо E его значения из (7.1) написанный интеграл распадается на произведения трех интегралов типа

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{p_x^2}{2m_x kT}\right) \cdot dp_x = (2\pi m_x kT)^{1/2}.$$

Поэтому

$$n = 2\left(\frac{2\pi kT}{(2\pi\hbar)^2}\right)^{3/2} (m_x m_y m_z)^{1/2} \exp\frac{F - E_c}{kT}.$$

Если в зоне Бриллюэна имеется ν эквивалентных минимумов, то плотность квантовых состояний будет в ν раз больше и во столько же раз увеличится концентрация электронов при заданном F.

Сравнивая полученный результат с формулой (5.1) и выражением (4.3) для N_c, видим, что в данном случае роль скалярной эффективной массы играет величина

$$m_{nd} = \nu^{2/3} (m_x m_y m_z)^{1/3}. \tag{7.3}$$

Она получила название *эффективной массы плотности состояний*.

Для вычисления эффективной массы m_d плотности состояний в общем случае удобно ввести в интеграле (7.2) новые переменные интегрирования. Именно, пусть

$$E(\mathbf{p}) = E = \text{const} \tag{7.4}$$

есть уравнение изоэнергетической поверхности. Очевидно, все пространство квазиимпульсов можно исчерпать, интегрируя сначала по поверхности (7.4), а затем — по всем возможным значениям E. Обозначим через dS элемент площади на поверхности (7.4), а через η — координату (в пространстве квазиимпульсов), постоянную на этой поверхности. Тогда для элемента объема в пространстве квазиимпульсов будем иметь

$$d\mathbf{p} = dS \cdot d\eta. \tag{7.5}$$

Очевидно, изменению координаты η соответствует изменение энергии

$$dE = (\mathbf{v},\ \nabla_{\mathbf{p}} E)\, d\eta,$$

где $\mathbf{v}$ — единичный вектор внешней нормали к поверхности (7.4). По определению градиента $(\mathbf{v},\ \nabla_{\mathbf{p}} E) = |\nabla_{\mathbf{p}} E|$. Следовательно,

$$d\eta = \frac{dE}{|\nabla_{\mathbf{p}} E|}, \tag{7.6}$$

причем, выполнив здесь дифференцирование по компонентам вектора $\mathbf{p}$, надо затем выразить их через E и какие-нибудь удобные переменные, характеризующие положение точки на поверхности (7.4). Таким образом,

$$n = \frac{2}{(2\pi\hbar)^3} \int_0^\infty dE\, f(E,\ T) \int_S \frac{dS}{|\nabla_{\mathbf{p}} E|}. \tag{7.7}$$

Отсюда следует, что при произвольном законе дисперсии плотность состояний в зоне проводимости определяется равенством

$$N_c(E) = \frac{2}{(2\pi\hbar)^3} \int_S \frac{dS}{|\nabla_{\mathbf{p}} E|}. \tag{7.8}$$

Заметим, что $|\nabla_{\mathbf{p}} E|$ есть не что иное, как абсолютная величина средней скорости электрона в данном состоянии $\mathbf{v}(\mathbf{p})$.

В частности, в применении к полупроводникам с узкой запрещенной зоной интересен случай изотропного, но непараболического закона дисперсии типа (III.9.2). В этом случае поверхность (7.4) есть сфера $dS = p^2 \sin\theta\, d\theta\, d\varphi$, где θ и φ — полярные углы, и, следовательно,

$$N_c(E) = \frac{p^2(E)}{\pi^2\hbar^3 v(E)}. \tag{7.8а}$$

Еще одно удобное выражение для плотности состояний можно получить, пользуясь правилами интегрирования с δ-функцией (Приложение IV). Действительно, рассмотрим интеграл

$$\int d\mathbf{p} \cdot \delta(E - E(\mathbf{p})).$$

Пользуясь формулами (7.5), (7.6) и правилами Приложения IV, мы можем переписать его в виде

$$\int dS \int \frac{dE}{|\nabla_{\mathbf{p}} E(\mathbf{p})|} \delta(E - E(\mathbf{p})) = \int \frac{dS}{|\nabla_{\mathbf{p}} E(\mathbf{p})|_{E(\mathbf{p})=E}}.$$

С точностью до множителя $2/(2\pi\hbar)^3$ это есть не что иное, как правая часть (7.8). Таким образом,

$$N_c(E) = \frac{2}{(2\pi\hbar)^3} \int d\mathbf{p} \cdot \delta(E - E_c(\mathbf{p})),$$
$$N_v(E) = \frac{2}{(2\pi\hbar)^3} \int d\mathbf{p} \cdot \delta(E - E_v(\mathbf{p})). \qquad (7.8б)$$

Заметим, что, согласно (7.8б), $N_c(E) = 0$ при $E < E_c$ и $N_v(E) = 0$ при $E > E_v$.

Согласно (7.7) общее выражение для эффективной плотности состояний в случае невырожденного электронного газа можно записать в виде

$$N_c = \int_0^\infty N_c(E) \exp\left(-\frac{E}{kT}\right) \cdot dE, \qquad (7.9)$$

где $N_c(E)$ выражается формулой (7.8). Приравнивая правые части выражений (7.9) и (4.3), находим эффективную массу плотности состояний:

$$m_{nd}^{3/2} = \frac{1}{(2\pi kT)^{3/2}} \int dE \cdot \exp\left(-\frac{E}{kT}\right) \int \frac{dS}{|\nabla_{\mathbf{p}} E(\mathbf{p})|}. \qquad (7.10)$$

Вообще говоря, она зависит от температуры.

Аналогичные формулы справедливы и для N_v и $m_{pd}^{3/2}$. Интегралы, фигурирующие в формуле (7.10), при непараболическом законе дисперсии обычно удается вычислить только численными методами.

Таким образом, мы видим, что полученные нами ранее результаты можно применять и к полупроводникам с более сложными законами дисперсии. Однако в этом случае нужно найти подходящее значение для эффективной массы, зависящее от формы изоэнергетических поверхностей, от числа эквивалентных минимумов в зоне Бриллюэна и, может быть, от температуры.

Отметим в заключение еще одно существенное обстоятельство. В случае параболического изотропного закона дисперсии эффективная масса, определяющая подвижность носителей заряда (гл. XIII), и эффективная масса плотности состояний равны друг другу. В более сложных случаях эти величины, вообще говоря, не совпадают. Результирующее значение эффективной массы, определяющее подвижность, получило название *эффективной массы электропроводности,* которую нужно отличать от эффективной массы плотности состояний.

Исходя из общей теории явлений переноса, основанной на использовании кинетического уравнения Больцмана, и зная форму изоэнергетических поверхностей и зависимость времени релаксации от энергии, эффективную массу электропроводности в ряде случаев можно вычислить. В таблице 5.1 приведены вычисленные значения эффективных масс плотности состояний m_d и электропроводности m_σ для двух важных полупроводников — германия и кремния.

Таблица 5.1

Эффективные массы в германии и кремнии *)

Вещество	Эффективная масса плотности состояний					Эффективная масса электропроводности			
	электроны m_{nd}/m_0	тяжелые дырки, m_{pd}^H/m_0	легкие дырки, m_{pd}^L/m_0	результирующая, m_{pd}/m_0	собственный полупроводник, m_{id}/m_0	электроны, $m_{n\sigma}/m_0$	тяжелые дырки, $m_{p\sigma}^H/m_0$	легкие дырки, $m_{p\sigma}^L/m_0$	результирующая, $m_{p\sigma}/m_0$
Германий	0,57	0,36	0,043	0,37	0,46	0,12	0,31	0,044	0,25
Кремний	1,08	0,53	0,14	0,59	0,80	0,26	0,49	0,16	0,38

*) По данным работы *B. Lax, J. G. Mavroides*, Phys, Rev. 100, 1950 (1955).

§ 8. Плотность состояний в квантующем магнитном поле

Формулы (2.3), (2.5) и (7.8) описывают плотности состояний носителей заряда, свободно движущихся в периодическом поле идеальной решетки. В достаточно сильных электрическом или магнитном полях энергетический спектр свободных электронов и дырок претерпевает серьезные изменения, что отражается и на плотности состояний. Вычислим плотность состояний электрона, движущегося в постоянном и однородном магнитном поле индукции $\mathscr{B}$. Для этой цели заметим, что по правилам статистической физики общее число свободных электронов N в объеме V дается выражением

$$N = \sum_\lambda g_\lambda f(E_\lambda, T). \tag{8.1}$$

Здесь λ — совокупность квантовых чисел, от которых зависит энергия электрона E_λ, g_λ — кратность вырождения энергетического уровня.

В рассматриваемой задаче (§ IV.5) роль квантовых чисел играют величины n, $k_z = \frac{2\pi}{L} n_z$ и проекция спина на ось Z, а кратность

вырождения дается формулой (IV.5.18). Таким образом,

$$N = N_+ + N_-, \tag{8.2}$$

где $N_\pm$ — числа электронов с положительной и отрицательной проекциями спина:

$$N_\pm = \frac{L^2 m\omega_c}{2\pi\hbar} \sum_{n,\, n_3} f(E_\lambda,\ T), \tag{8.3}$$

а значения энергии $E_\pm$ даются формулой (IV.5.12).

Расстояние между двумя соседними значениями k_3 составляет $\Delta k_z = 2\pi/L$. При $L \to \infty$ эта величина сколь угодно мала, что позволяет перейти от суммирования по n_z к интегрированию по k_z. Действительно,

$$\sum_{n_3} f(E_\pm,\ T) = \frac{L}{2\pi} \sum_{n_3} f(E_\pm,\ T)\, \Delta k_z \to \frac{L}{2\pi} \int f(E_\pm,\ T)\, dk_z. \tag{8.4}$$

Пользуясь тем, что $E_\pm$ есть четная функция k_z, мы можем заменить здесь интеграл по k_z в бесконечных пределах удвоенным интегралом от нуля до бесконечности. Тогда получаем

$$N_\pm = L^3 \frac{m\omega_c}{(2\pi)^2 \hbar}\, 2 \int_0^\infty dk_z \sum_n f(E_\pm,\ T). \tag{8.5}$$

Очевидно, множитель при $L^3 = V$ есть концентрация электронов с данной проекцией спина $n_\pm$.

Наконец, введем вместо k_z переменную интегрирования $E_\pm$. Получим, опуская значок «$\pm$» у переменной интегрирования,

$$n_\pm = \int_{{}^1/_2 \hbar\omega_c \pm \beta g \mathscr{B}}^\infty N_\pm(E) f(E,\ T)\, dE, \tag{8.6}$$

где

$$N_\pm(E) = \frac{e\mathscr{B}\sqrt{2m}}{(2\pi\hbar)^2 c} \sum_{n \geqslant 0} \left[E - E_c - \hbar\omega_c \left(n + \frac{1}{2}\right) \pm \beta g \mathscr{B}\right]^{-1/2}. \tag{8.7}$$

Сумма берется здесь по всем значениям n, удовлетворяющим условию

$$E - E_c - \hbar\omega_c\,(n + {}^1/_2) \pm \beta g \mathscr{B} \geqslant 0. \tag{8.8}$$

Величины $N_\pm(E)$ и представляют собой искомые плотности состояний в магнитном поле, относящиеся к электронам с положительной и отрицательной проекцией спина. Сумма их

$$N(E) = N_+(E) + N_-(E)$$

есть аналог величины $N_c(E)$.

Формула (8.7) применима и к дыркам (если можно пренебречь эффектами, связанными с вырождением валентной зоны). Следует лишь заменить в ней $E - E_c$ на $E_v - E$ и изменить знак ω_c.

Из выражений (8.7) и (8.8) видно, что плотность состояний электронов в магнитном поле есть осциллирующая функция магнитной индукции. Она имеет острые максимумы при тех значениях магнитной индукции, при которых энергия $(E - E_c \pm \beta g \mathscr{B})$ близка к одному из уровней Ландау $\hbar\omega_c\,(n + {}^1/_2)$. Это обстоятельство проявляется в том, что в вырожденном электронном газе (в металлах, полуметаллах и вырожденных полупроводниках) многие термодинамические, электрические и оптические величины, выражающиеся через плотность состояний, при определенных условиях осциллируют при изменении магнитной индукции. Так как в вырожденном электронном газе существенную роль играют только электроны на поверхности Ферми, то осцилляции свойств происходят при прохождении какого-либо из уровней Ландау через уровень Ферми. Впервые такие квантовые осцилляции были обнаружены в величине магнетосопротивления (эффект Шубникова—де Гааза) и магнитной восприимчивости (эффект де Гааза — ван Альфена) при низких температурах. Количественная теория квантовых осцилляций довольно громоздка (ее можно найти в книге [М8]). Экспериментальное исследование квантовых осцилляций позволяет получить информацию о форме поверхностей Ферми.

Наряду с указанными выше, существуют также квантовые осцилляции различных физических величин, имеющие другое происхождение. Они тоже связаны с уровнями Ландау, но, в отличие от эффектов Шубникова—де Гааза и де Гааза—ван Альфена, не требуют вырождения электронного газа и наблюдаются при более высоких температурах. Когда энергетический интервал между какими-либо уровнями Ландау $n\hbar\omega_c$ $(n = 1, 2, \ldots)$ становится равным энергии оптических фононов (см. § XII.6), резко возрастает вероятность неупругого рассеяния электронов на колебаниях решетки. Поэтому все эффекты, зависящие от характера рассеяния электронов, такие, как магнетосопротивление, дифференциальная термоэдс и др., испытывают осцилляции при изменении магнитного поля. Эти явления получили название магнитофононных осцилляций (подробнее см., например, [М8]).

§ 9. Концентрации электронов и дырок на локальных уровнях. Простые центры

Обратимся теперь к вычислению концентраций электронов и дырок, связанных на локальных уровнях энергии. Рассмотрим сначала простейший случай, когда примесный атом или структурный дефект может либо иметь один электрон, либо быть пустым, т. е. может находиться только в двух различных зарядовых состояниях. Если не рассматривать пока возможные возбужденные состояния, то такой центр будет характеризоваться одним уровнем энергии E_1. Если, кроме того, каждому зарядовому состоянию центра соответствует только одно квантовое состояние электрона, то концентрация

центров N_1, имеющих электрон, будет $N_t f$, где N_t — полная концентрация центров, а f — функция Ферми—Дирака (3.1). Аналогично, концентрация пустых центров $N_0 = N_t(1-f)$. Отношение этих концентраций равно

$$\frac{N_1}{N_0} = \frac{f}{1-f} = \exp\frac{F-E_1}{kT}.$$

Однако в действительности одному и тому же уровню энергии связанного электрона соответствует несколько квантовых состояний (квантовомеханическое вырождение энергетического уровня). Если число таких состояний для заполненного уровня есть g_1, а для пустого — g_0, то мы, очевидно, будем иметь *)

$$\frac{N_1}{N_0} = \frac{g_1}{g_0}\exp\frac{F-E_1}{kT}. \tag{9.1}$$

Так как, кроме того, всегда

$$N_0 + N_1 = N_t, \tag{9.2}$$

то отсюда получается

$$f^{(1)} = \frac{N_1}{N_t} = \left(1+\frac{g_0}{g_1}\exp\frac{E_1-F}{kT}\right)^{-1}, \quad f^{(0)} = \frac{N_0}{N_t} = \left(1+\frac{g_1}{g_0}\exp\frac{F-E_1}{kT}\right)^{-1}. \tag{9.3}$$

Для случая невырожденных полупроводников в эти выражения удобно ввести концентрацию электронов в зоне n. Учитывая формулу (5.1), мы имеем

$$\frac{N_1}{N_t} = \frac{n}{n+n_1}, \quad \frac{N_0}{N_t} = \frac{n_1}{n+n_1}. \tag{9.4}$$

Здесь введено обозначение

$$n_1 = \frac{g_0}{g_1}N_c\exp\left(-\frac{\mathscr{I}}{kT}\right), \quad \mathscr{I} = E_c - E_1. \tag{9.5}$$

Очевидно, $\mathscr{I}$ есть не что иное, как энергия ионизации данного центра. Отметим еще, что, с точностью до факторов вырождения, n_1 равно концентрации электронов в зоне, когда уровень Ферми совпадает с уровнем энергии центра E_1.

Полученные результаты можно обобщить учетом возбужденных состояний центра. Рассмотрим сначала полупроводник n-типа и будем интересоваться связанными электронами. Обозначим энергию захваченного электрона в основном состоянии через E_1, а энергии возбужденных состояний через E_k ($k = 2, 3, \ldots$). Энергетический спектр такого центра схематически изображен на рис. 5.5. Уровни возбужденных состояний электрона расположены выше уровня основного состояния ($k = 1$).

*) При этом предполагается, что из-за кулоновского отталкивания при любой степени вырождения уровень может быть занят лишь одним электроном. О многозарядных центрах см. § 10.

Обозначим через $f_k^{(1)}$ вероятность нахождения электрона на центре и притом в k-м возбужденном состоянии, через β_k — кратность вырождения этого возбужденного состояния, а через $f^{(0)}$—вероятность найти центр не имеющим электрона. Тогда, аналогично формуле (9.1), можно написать

$$\frac{f_k^{(1)}}{f^{(0)}} = \frac{\beta_k}{g_0} \exp \frac{F - E_k}{kT}. \qquad (9.6)$$

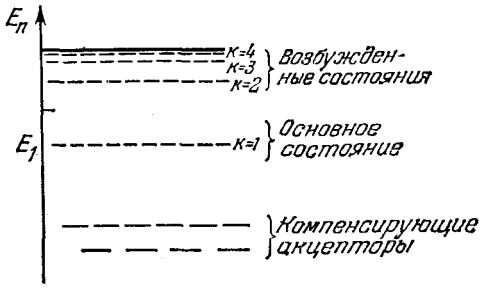


Рис. 5.5. Энергетический спектр связанного электрона при наличии возбужденных состояний.

Полная вероятность того, что электрон захвачен центром (и находится в каком-либо, основном или возбужденном, состоянии), есть

$$f^{(1)} = \sum_{k=1,\,2,\,\ldots} f_k^{(1)},$$

где суммирование производится по всем состояниям, энергии которых лежат в запрещенной зоне. Поэтому отношение концентрации центров N_1, имеющих электрон, к концентрации пустых центров N_0 равно

$$\frac{N_1}{N_0} = \frac{f^{(1)}}{f^{(0)}} = \frac{1}{g_0} \sum_{k=1,\,2,\,\ldots} \beta_k \exp \frac{F - E_k}{kT}. \qquad (9.7)$$

Полученное выражение можно опять записать в виде (9.1), где E_1 — энергия основного состояния электрона. Однако при этом g_1 имеет другой смысл и выражается формулой

$$g_1 = \beta_1 + \sum_{k=2,\,3,\,\ldots} \beta_k \exp\left(-\frac{E_k - E_1}{kT}\right). \qquad (9.8)$$

Так как и в этом случае, разумеется, справедливо нормировочное соотношение (9.2), то для $f^{(1)}$ и $f^{(0)}$ получаются формально те же выражения (9.3), что и без учета возбужденных состояний. Однако при этом «g-фактор» (9.8) может быть равен уже не целому числу и, что особенно важно, может сам зависеть от температуры.

Рассмотрим теперь полупроводник p-типа. Как уже многократно разъяснялось выше, в этом случае в качестве носителей заряда мы должны рассматривать не электроны, а положительные дырки. Поэтому мы должны и вычислять концентрацию связанных дырок, а не электронов и рассматривать энергетический спектр связанных дырок. Такой спектр схематически изображен на рис. 5.6, где отмечено основное состояние захваченной дырки ($k = 1$) и различные возбужденные состояния ($k = 2, 3, \ldots$). Если по-прежнему откладывать снизу вверх энергию электрона, то уровни возбужденных со-

стояний дырок будут расположены ниже уровня основного состояния. Энергия ионизации дырки, находящейся в k-м состоянии, на этой диаграмме есть $E_k - E_v$, она также уменьшается при увеличении номера возбужденного состояния. Диаграмма рис. 5.6 становится подобной диаграмме рис. 5.5, если изменить направление отсчета

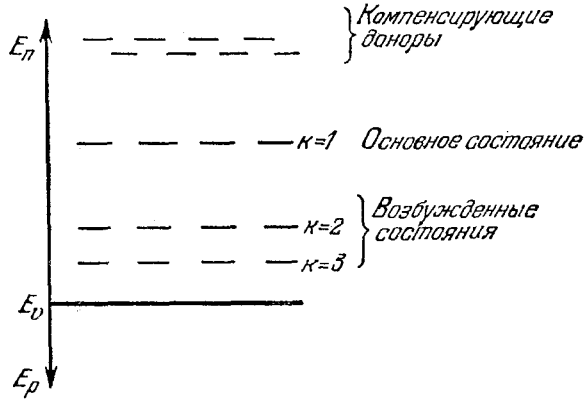


Рис. 5.6. Энергетический спектр связанной дырки при наличии возбужденных состояний.

энергии (т. е. повернуть диаграмму на 180°). При этом «пустым» мы должны считать центр, не захвативший дырку (т. е. заполненный электроном).

Легко понять, что в этом случае вместо соотношения (9.6) мы получим

$$\frac{f_{k,p}^{(1)}}{f_p^{(0)}} = \frac{\beta_{k,p}}{g_{0,p}} \exp\left(\frac{E_k - F}{kT}\right), \tag{9.9}$$

где $f_{k,p}^{(1)}$ — вероятность того, что на центре имеется дырка в k-м возбужденном состоянии, $f_p^{(0)}$ — вероятность отсутствия дырки, E_k — энергия связанной дырки, $\beta_{k,p}$ — кратность вырождения k-го возбужденного состояния (целое число), $g_{0,p}$ — кратность вырождения пустого (заполненного электроном) центра. Поступая, как и раньше, мы получим вместо формул (9.3) соотношения

$$f_p^{(1)} = \left(1 + \frac{g_{0,p}}{g_{1,p}} \exp\frac{F - E_1}{kT}\right)^{-1}, \quad f_p^{(0)} = 1 - f_p^{(1)}, \tag{9.10}$$

где $f_p^{(1)}$ — вероятность нахождения на центре дырки (в любом состоянии), E_1 — энергия основного состояния захваченной дырки. При этом обобщенный g-фактор дается выражением

$$g_{1,p} = \beta_{1,p} + \sum_{k=2,3,\ldots} \beta_{k,p} \exp\left(-\frac{E_1 - E_k}{kT}\right). \tag{9.11}$$

Так же, как и выше, для невырожденных полупроводников в формулы (9.10) удобно ввести концентрацию дырок в зоне p. Учитывая формулу (5.2), имеем

$$f_p^{(1)} = \frac{p}{p+p_1}, \quad f_p^{(0)} = \frac{p_1}{p+p_1}, \tag{9.12}$$

где

$$p_1 = \frac{g_{0,\,p}}{g_{1,\,p}} N_v \exp\left(-\frac{\mathscr{I}_p}{kT}\right), \tag{9.13}$$

а величина $\mathscr{I}_p = E_1 - E_v$ равна энергии отщепления дырки. По-прежнему, с точностью до множителя $g_{0,\,p}/g_{1,\,p}$, p_1 равно концентрации дырок в зоне, когда уровень Ферми совпадает с уровнем энергии центра E_1.

Во всех предыдущих рассуждениях мы ничего не говорили о характере рассматриваемых центров, т. е. о том, являются ли они донорами или акцепторами. Это обстоятельство действительно не играет никакой роли при вычислении вероятностей заполнения центров электронами или, соответственно, дырками, так как вероятности заполнения зависят только от энергетического спектра центров и положения уровня Ферми. Однако донорный или акцепторный характер центров должен учитываться при вычислении связанного заряда, т. е. концентраций заряженных центров. В случае, если центры являются донорами, они нейтральны, если с ними связан электрон (т. е. нет дырки), и положительно заряжены, если на них не имеется электрона (захвачена дырка). И наоборот, если центры являются акцепторами, они нейтральны, если с ними связана дырка (не имеется электрона), и отрицательно заряжены, когда не имеется дырки (захвачен электрон).

§ 10. Многозарядные центры

В § II.9 мы уже говорили, что многие примесные атомы создают в полупроводниках «многозарядные» центры, способные присоединять или отдавать не один, а несколько носителей заряда. Поэтому мы обобщим полученные выше результаты на этот более сложный случай.

Положим, что рассматриваемые центры могут содержать 0, 1, 2, ..., M электронов и, соответственно, находиться в $(M+1)$ различных зарядовых состояниях. Энергетическая диаграмма полупроводника с такими центрами для случая $M = 3$ показана на рис. 5.7. Если центр пустой, то для электрона существует некоторое вакантное основное квантовое состояние, характеризуемое локальным уровнем энергии E_1. Для этого уровень Ферми F должен быть расположен ниже уровня E_1 на несколько kT (рис. 5.7, *а*). Если $E^{(0)}$ есть энергия пустого центра, то $(E_1 - E^{(0)})$ равно приращению энергии центра при захвате первого электрона. Однако сама величина

$E^{(0)}$ для нас не играет никакой роли. Это обстоятельство будет проявляться в последующих формулах в том, что в них нигде не будет входить $E^{(0)}$, а будут содержаться только разности энергетических уровней E_1, E_2 и т. д. и уровня Ферми F.

Помимо основного состояния с энергией E_1, захваченный электрон может находиться в различных возбужденных состояниях, уровни энергии которых будут расположены выше уровня E_1. Эти уровни не изображены на рис. 5.7.

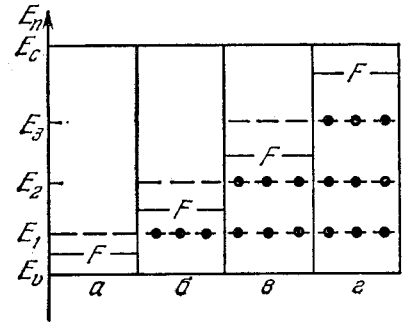


Рис. 5.7. Энергетическая диаграмма полупроводника с многозарядными центрами, создающими три локальных уровня энергии.

При заполнении уровня E_1 электронами (например, за счет мелких доноров, электроны с которых «проваливаются» на уровень E_1) для электронов возникает новое основное квантовое состояние с энергией E_2, которое может быть занято вторым захваченным электроном (рис. 5.7, *б*). Подчеркнем, что уровня энергии E_2 не существует, пока уровень E_1 не занят первым электроном. В этом заключается характерная особенность энергетического спектра многозарядных центров («альтернативные» уровни), отличающая его от системы уровней, создаваемых несколькими простыми центрами разных типов. В последнем случае вся совокупность локальных уровней всегда существует полностью, независимо от степени заполнения отдельных уровней.

Ситуация, изображенная на рис. 5.7, *б*, возникает тогда, когда уровень Ферми F расположен выше уровня E_1, но ниже уровня E_2 (см. ниже). При заполнении электронами уровня E_2 возникают новые квантовые состояния с энергией E_3 и т. д., которые могут быть либо заполненными, либо пустыми в зависимости от положения уровня Ферми (рис. 5.7, *в*, *г*).

§ 11. Распределение Гиббса

Для вычисления вероятности заполнения электронами многозарядных центров уже нельзя использовать функцию Ферми (3.1), а необходимо исходить из более общего принципа статистической физики — канонического распределения Гиббса для системы с переменным числом частиц. Рассмотрим изолированную систему, содержащую очень большое число одинаковых частиц $N_0 = \text{const}$ и обладающую определенной внутренней энергией $E_0 = \text{const}$. Выделим внутри этой системы малую ее часть, ограниченную по-

стоянным объемом («тело»), находящуюся в термодинамическом равновесии с остальной частью системы («средой»). Среда и тело могут обмениваться частицами и изменять при этом свои энергии. Обозначим число частиц, перешедших в тело, через j, а изменение энергии тела, получившего j частиц и находящегося в определенном квантовом состоянии m, — через $E_m^{(j)}$. При таком переходе энергия среды станет $(E_0 - E_m^{(j)})$, а число частиц в ней будет $(N_0 - j)$, и поэтому энтропия среды изменится на некоторую величину $\Delta S_m^{(j)}$. Тогда, согласно общим принципам статистической физики *), вероятность $f_m^{(j)}$ того, что в теле будет j частиц, а тело будет в m-м квантовом состоянии, есть

$$f_m^{(j)} = A \exp \frac{\Delta S_m^{(j)}}{k}. \tag{11.1}$$

Здесь k — постоянная Больцмана, а A — нормировочный множитель.

Величина $\Delta S_m^{(j)}$ связана с другими термодинамическими величинами. Так, например, рассматривая свободную энергию

$$\mathcal{F} = E - TS$$

и учитывая, что при термодинамическом равновесии температура везде (и в теле, и в среде) одинакова, имеем, что изменение свободной энергии среды равно

$$\Delta \mathcal{F}_m^{(j)} = -E_m^{(j)} - T \Delta S_m^{(j)}. \tag{11.2}$$

Учтем теперь, что свободная энергия при неизменных значениях переменных, определяющих состояние системы (в данном случае температуры и объема), есть функция аддитивная, т. е. пропорциональная числу частиц, содержащихся в системе. В соответствии с этим введем свободную энергию, рассчитанную на одну частицу:

$$\left(\frac{\partial \mathcal{F}}{\partial j}\right)_{V,\,T} = F. \tag{11.3}$$

Отметим, что мы умышленно выбрали для этой величины и уровня Ферми одно и то же обозначение, так как обе эти величины совпадают (см. § 12). Тогда в соотношении (11.2) мы имеем $\Delta \mathcal{F}_m^{(j)} = -jF$, и поэтому изменение энтропии среды равно

$$\Delta S_m^{(j)} = \frac{jF - E_m^{(j)}}{T}.$$

После этого формула (11.1) дает

$$f_m^{(j)} = A \exp \frac{jF - E_m^{(j)}}{kT}. \tag{11.4}$$

*) См. [М9], § 35.

Это соотношение выражает распределение Гиббса с переменным числом частиц.

Применим теперь распределение Гиббса к интересующему нас случаю многозарядных центров. Тогда под телом следует подразумевать центр, под средой — кристалл, а $f_m^{(j)}$ будет давать вероятность того, что на центре имеется j электронов, причем энергия центра равна $E_m^{(j)}$. Постоянная A определяется из условия нормировки вероятности на единицу

$$\sum_{j=0}^{M} \sum_{m} f_m^{(j)} = 1, \tag{11.5}$$

где суммирование по m учитывает все возбужденные состояния центра, а M обозначает максимальное число электронов, которые может захватить центр. Это дает

$$\frac{1}{A} = \sum_{j=0}^{M} \sum_{m} \exp \frac{jF - E_m^{(j)}}{kT}. \tag{11.6}$$

Полученный результат, выражаемый формулами (11.4) и (11.6), для дальнейших применений удобно представить в несколько ином виде. А именно, положим

$$E_m^{(j)} = E^{(j)} + \varepsilon_{jm} \qquad (m = 1, 2, \ldots), \tag{11.7}$$

где $E^{(j)}$ — энергия центра с j электронами в основном состоянии, а ε_{jm} — энергия m-го возбужденного состояния относительно основного состояния.

Далее, будем интересоваться вероятностью $f^{(j)}$ найти центр с j электронами (т. е. в данном зарядовом состоянии и не важно в каком возбужденном состоянии):

$$f^{(j)} = \sum_{m} f_m^{(j)}.$$

И, наконец, как и раньше, введем кратности вырождения β_{jm} для различных возбужденных состояний (целые числа). Тогда из (11.4), (11.6) и (11.7) получается

$$f^{(j)} = \frac{g_j \exp \frac{jF - E^{(j)}}{kT}}{\sum_{j=0}^{M} g_j \exp \frac{jF - E^{(j)}}{kT}}, \tag{11.8}$$

где через g_j обозначено

$$g_j = \beta_j + \sum_{m=1, 2, \ldots} \beta_{jm} \exp\left(-\frac{\varepsilon_{jm}}{kT}\right). \tag{11.9}$$

В этих выражениях β_j есть кратность вырождения основного состояния центра с j электронами, а g_j — обобщенные кратности вырождения j-го зарядового состояния. Так как по смыслу $E^{(j)}$ есть приращение энергии центра, то в формуле (11.8) следует полагать $E^{(0)} = 0$.

Укажем еще на связь энергий $E^{(j)}$, входящих в распределение Гиббса (11.8), с уровнями энергии E_j обычных энергетических диаграмм (рис. 5.7). Так как уровни энергии E_j изображают приращение энергии центра при добавлении к нему одного электрона, то

$$E^{(j)} = E_1 + E_2 + \ldots + E_j. \tag{11.10}$$

При этом уровень отсчета энергий здесь безразличен, если только E_j и F отсчитываются от одинакового начального уровня, так как начальный уровень из выражения ($jF - E^{(j)}$) выпадает.

Определяя уровень Ферми формулой (11.3), мы пользовались свободной энергией и, соответственно, в качестве термодинамических переменных выбирали объем и температуру. Однако мы могли бы выбрать и другие термодинамические величины: тепловую функцию W или термодинамический потенциал Φ и соответствующие им переменные. Тогда мы получили бы равнозначные определения:

$$F = \left(\frac{\partial \mathcal{F}}{\partial j}\right)_{V,\,T} = \left(\frac{\partial W}{\partial j}\right)_{S,\,p} = \left(\frac{\partial \Phi}{\partial j}\right)_{p,\,T}. \tag{11.11}$$

Подчеркнем, что при наличии внешнего поля уровень F, определяемый формулой (11.3) (или формулами (11.11)), должен содержать и потенциальную энергию электрона в этом поле $-e\varphi$. По этой причине его и называют электрохимическим потенциалом. Если условиться отсчитывать F от края зоны проводимости E_{c0} при $\varphi = 0$, то электрохимический потенциал электронов есть

$$F = \zeta - e\varphi, \tag{11.12}$$

где $\zeta = F - E_c$ — химический потенциал в поле с потенциалом φ.

§ 12. Частные случаи

Рассмотрим простой примесный центр, имеющий один невырожденный уровень энергии, без возбужденных состояний. Тогда в формуле (11.8) j может принимать только два значения: 0 и 1. При этом, согласно (11.10), $E^{(j)}$ есть единственный локальный уровень энергии центра E_1, а $g_0 = g_1 = 1$. Поэтому вероятность заполнения центра электроном получается равной

$$f = \frac{\exp\dfrac{F - E_1}{kT}}{1 + \exp\dfrac{F - E_1}{kT}} = \frac{1}{1 + \exp\dfrac{E_1 - F}{kT}}.$$

Таким образом, распределение Гиббса для простого, невырожденного энергетического уровня переходит в функцию Ферми—Ди-

рака (3.1). Учитывая формулу (11.3), отсюда видно, что введенный нами ранее уровень Ферми F есть не что иное, как свободная энергия, рассчитанная на один электрон.

Если имеется простой центр, способный захватывать или отдавать единственный электрон, но уровни энергии вырождены, то формула (11.8) при $M = 1$ переходит в формулы (9.3), полученные ранее с помощью менее строгих, интуитивных рассуждений.

Посмотрим теперь, как будет изменяться средний заряд многоэлектронного центра при изменении положения уровня Ферми. Для этого удобно исследовать поведение отношения вероятностей нахождения на центре j и, соответственно, $(j-1)$ электронов. Из (11.8) и (11.10) следует, что

$$\frac{j^{(j)}}{j^{(j-1)}} = \frac{N_j}{N_{j-1}} = \frac{g_j}{g_{j-1}} \exp \frac{F - E_j}{kT}. \tag{12.1}$$

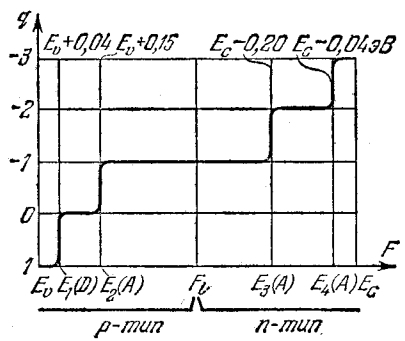


Рис. 5.8. Зависимость среднего заряда q (в элементарных зарядах) атомов золота в германии от положения уровня Ферми.

В дальнейшем мы будем предполагать, что различные уровни энергии E_j достаточно удалены друг от друга (по крайней мере на несколько kT). Тогда из формулы (12.1) видно, что если уровень Ферми лежит между двумя какими-либо уровнями E_{j-1} и E_j и удален от этих последних хотя бы на $(2 \div 3)\,kT$, то лежащий ниже уровень $(j-1)$ будет практически целиком заполнен, а лежащий выше уровень j будет практически пустым. Поэтому среднее число захваченных электронов на один центр будет очень близко к $(j-1)$. Оно быстро изменяется, когда уровень Ферми приближается к какому-либо уровню энергии центра. Когда уровень Ферми совпадает с каким-либо уровнем энергии E_j, то отношение концентраций центров в j-м и $(j-1)$-м зарядовых состояниях равно $N_j/N_{j-1} = g_j/g_{j-1}$.

Таким образом, если уровень Ферми лежит между двумя какими-либо соседними уровнями энергии и при своем перемещении не пересекает эти уровни, то практически центры будут находиться только в одном зарядовом состоянии, а их концентрация в других зарядовых состояниях будет исчезающе мала.

Для иллюстрации сказанного рассмотрим конкретный случай германия, содержащего атомы золота. Последние создают в германии четыре альтернативных уровня энергии, отмеченных на рис. 5.8, из которых нижний уровень является донорным, а остальные — акцепторными. Положим, далее, что уровень Ферми постепенно перемещается от края дырочной зоны E_v к краю зоны проводимости E_c (например, вследствие изменения температуры или концентрации других мелких доноров и акцепторов). Тогда средний заряд атомов золота будет

изменяться так, как показано на рис. 5.8. Когда F лежит ниже донорного уровня E_1, последний пуст и, следовательно, атомы золота заряжены положительно. Когда F лежит между уровнями E_1 и E_2, то уровень E_1 практически весь заполнен, а уровень E_2 — пустой. Так как уровень E_1 донорный, то подавляющее большинство атомов будет в нейтральном состоянии и средний их заряд будет близок к нулю. После перехода уровня Ферми через второй (акцепторный) уровень энергии E_2 практически все атомы будут содержать два электрона, а, следовательно, средний заряд будет близок к $-e$, и т. д. При последовательном повышении уровня Ферми атомы золота будут проходить через все возможные для них зарядовые состояния e, 0, $-e$, $-2e$, $-3e$.

§ 13. Определение положения уровня Ферми

В предыдущих рассуждениях мы считали, что уровень Ферми задан. Посмотрим теперь, как можно найти положение уровня Ферми.

Ответ на этот вопрос зависит от того, какие другие величины заданы. Если известны концентрации носителей заряда в зонах n и p, то значение F можно определить из формул §§ 4—6. Так, например, для невырожденного полупроводника n-типа из (5.1) мы имеем

$$F = E_c - kT \ln \frac{N_c}{n}.$$

Аналогично, для невырожденного полупроводника p-типа из (5.2) получаем

$$F = E_v + kT \ln \frac{N_v}{p}.$$

Эти выражения дают уже известный нам результат, что чем больше концентрация основных носителей, тем ближе уровень Ферми к краю соответствующей зоны.

Однако чаще мы встречаемся с задачей, когда задан состав кристалла, т. е. концентрации и типы содержащихся в нем примесей (их энергетические уровни), а концентрации свободных и связанных носителей заряда, напротив, должны быть вычислены. В этом случае положение уровня Ферми можно найти из условия электрической нейтральности образца.

Из классической электродинамики известно, что в однородной проводящей среде в состоянии равновесия объемный заряд в любой точке равен нулю. Возникший же по тем или иным причинам заряд рассасывается под действием кулоновских сил за время порядка максвелловского времени релаксации

$$\tau_M = \frac{\varepsilon}{4\pi\sigma}, \tag{13.1}$$

где ε — диэлектрическая проницаемость, σ — удельная электропроводность. Для типичных полупроводников τ_M весьма мало. Так, например, при $\varepsilon \sim 10$ и $\sigma \sim 1$ Ом$^{-1}\cdot$см$^{-1}$ $= 9\cdot 10^{11}$ ед. СГСЭ

мы имеем $\tau_M \sim 10^{-12}$ с. Поэтому в стационарном состоянии или даже при периодическом изменении состояния с частотой ω, удовлетворяющей условию $\omega\tau_M \ll 1$, внутренность полупроводника можно считать нейтральной.

Отсюда следует, что в равновесии концентрации положительно заряженных частиц должны быть всегда равны концентрациям отрицательных частиц. Положим, что в полупроводнике имеются доноры одного и того же типа с концентрацией N_d и, кроме того, акцепторы (тоже одинакового типа) с концентрацией N_a. Положительными частицами являются, во-первых, подвижные дырки с концентрацией p. Во-вторых, мы имеем еще связанные положительные заряды p_t в виде положительно заряженных доноров.

В случае простых доноров

$$p_t = N_d(1-f) \quad \text{(простые доноры)}, \tag{13.2}$$

где, как и раньше, f — вероятность заполнения донора электроном, выражаемая первой формулой (9.3) (где индекс 1 опущен).

Соответственно для многозарядных доноров, имеющих в нейтральном состоянии M избыточных электронов,

$$p_t = N_d \sum_{j=0}^{M-1} (M-j) f^{(j)} = N_d \sum_{j=1}^{M} j f_p^{(j)} \quad \text{(многозарядные доноры)}, \tag{13.2а}$$

где $f^{(j)}$ — вероятность донору иметь j электронов (формула (11.8)).

Концентрация отрицательных частиц складывается из концентрации электронов в зоне n и концентрации отрицательно заряженных акцепторов. Если n_t есть концентрация связанных элементарных зарядов, то для случая простых акцепторов

$$n_t = N_a f \quad \text{(простые акцепторы)}. \tag{13.3}$$

Аналогично, для многозарядных акцепторов

$$n_t = N_a \sum_{j=1}^{M} j f^{(j)} \quad \text{(многозарядные акцепторы)}. \tag{13.3а}$$

После этого условие электрической нейтральности можно записать в виде

$$p + p_t - n - n_t = 0. \tag{13.4}$$

Если в полупроводнике имеются доноры и акцепторы разных типов, то под p_t и n_t нужно понимать суммарные концентрации связанных элементарных зарядов.

Каждая из величин p, p_t, n и n_t зависит от значения уровня Ферми, и поэтому уравнение (13.4) можно использовать для его опре-

деления. Однако это уравнение является трансцендентным, и поэтому для его решения приходится либо пользоваться либо численными методами, либо исследовать различные частные случаи, к которым мы и обратимся.

§ 14. Уровень Ферми в собственном полупроводнике

Для собственного полупроводника p_t, $n_t \ll n$, p и условие нейтральности принимает вид $n = p$. Если ширина запрещенной зоны полупроводника достаточно велика, так что она охватывает много kT, и если эффективные массы электронов и дырок m_n и m_p одного порядка, то уровень Ферми будет достаточно удален от краев зон и полупроводник будет невырожденным. Поэтому, пользуясь для n и p выражениями (5.1) и (5.2), имеем

$$N_c \exp \frac{F-E_c}{kT} = N_v \exp \frac{E_v - F}{kT}.$$

Это дает

$$F = E_i - \frac{1}{2} kT \ln \frac{N_c}{N_v} = E_i - \frac{3}{4} kT \ln \frac{m_n}{m_p}, \qquad (14.1)$$

где через $E_i = {}^1/_2 (E_v + E_c)$ обозначена энергия середины запрещенной зоны.

Эта зависимость показана схематически на рис. 5.9. Там же отмечены края зон E_v и E_c и учтено, что ширина запрещенной зоны $E_g = E_c - E_v$ сама изменяется с температурой. При $T = 0$ уровень Ферми располагается точно в середине запрещенной зоны. При повышении температуры он удаляется от зоны более тяжелых носителей заряда и приближается к зоне более легких.

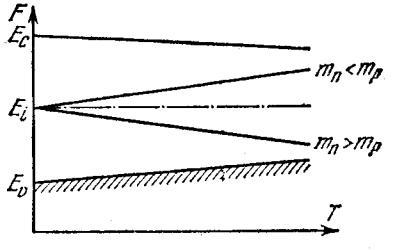


Рис. 5.9. Зависимость положения уровня Ферми от температуры в собственном полупроводнике (схематически).

Для полупроводников с узкой запрещенной зоной (HgSe, HgTe, серое олово и др.) даже при комнатной температуре приходится уже учитывать вырождение, и поэтому для n и p нужно брать общие выражения (4.4) и (4.6).

Из выражения (14.1) и рис. 5.9 видно, что если m_n и m_p весьма различны, то при повышении температуры уровень Ферми может приблизиться к зоне легких носителей на расстояние порядка kT и даже оказаться внутри этой зоны. Поэтому такие полупроводники при нагревании могут стать вырожденными. Этот случай мы имеем, например, в InSb, где $m_n \ll m_p$. При этом оказывается, что уровень Ферми попадает в зону проводимости при температурах $T \gtrsim 440$ К.

§ 15. Полупроводник с примесью одного типа

Рассмотрим полупроводник, содержащий простые доноры с энергетическим уровнем E_d. Далее, будем считать, что температура не слишком велика, так что собственной проводимостью можно пренебречь. В этом случае электроны в зоне проводимости возникают только за счет тепловой ионизации доноров. Найдем концентрацию электронов в зоне и положение уровня Ферми в зависимости от температуры.

Условие нейтральности (13.4) для этого случая ($p \ll n$, $p_t = 0$), при учете выражений (9.3) и (4.4), дает

$$\frac{N_d}{1+\frac{g_1}{g_0}\exp\frac{F-E_d}{kT}} = N_c\Phi_{1/2}\left(\frac{F-E_c}{kT}\right). \tag{15.1}$$

Уравнение (15.1) позволяет определить положение уровня Ферми F. Однако для общего случая решение этой задачи требует численных методов расчета. Поэтому мы рассмотрим только случай невырожденного полупроводника, когда

$$\Phi_{1/2}\left(\frac{F-E_c}{kT}\right) \simeq \exp\frac{F-E_c}{kT}.$$

Замечая, что $\exp\frac{F-E_d}{kT}$ можно представить в виде

$$\exp\frac{F-E_d}{kT} = \frac{n}{N_c}\exp\frac{\mathscr{J}}{kT},$$

где $\mathscr{J} = E_c - E_d$ есть энергия ионизации донора, условие нейтральности (13.4) можно переписать в виде

$$\frac{n^2}{N_d - n} = n_1(T). \tag{15.2}$$

Здесь n_1 — введенная нами ранее величина, определяемая формулой (9.5). Соотношение (15.2) приводит к квадратному уравнению относительно n, положительный корень которого есть

$$n = \frac{n_1}{2}\left(\sqrt{1+\frac{4N_d}{n_1}} - 1\right). \tag{15.3}$$

При достаточно низких температурах, определяемых условием $(4N_d/n_1)^{1/2} \gg 1$,

$$n = \left(\frac{g_0}{g_1}N_dN_c\right)^{1/2}\exp\left(-\frac{\mathscr{J}}{2kT}\right). \tag{15.3а}$$

В этом случае мы имеем частичную ионизацию доноров и концентрация электронов в зоне уменьшается по экспоненциальному закону

с понижением температуры. Изображая графически зависимость $\ln nT^{-3/4}$ от $1/T$, мы получим прямую линию, наклон которой равен $\mathscr{J}/2k$, т. е. отвечает половине энергии ионизации доноров $\mathscr{J}$.

При достаточно высоких температурах ($4N_d/n_1 \ll 1$) из (15.3) получается

$$n = N_d. \qquad (15.3б)$$

Этот случай соответствует полной ионизации доноров. Зависимость $n(T)$ для одного конкретного случая показана ниже на рис. 5.11 (кривая *1*).

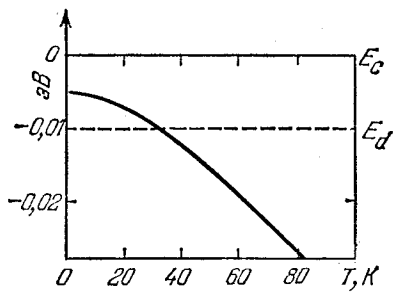


Рис. 5.10. Зависимость уровня Ферми от температуры в Ge с некомпенсированными донорами V группы (для данных кривой *1* на рис. 5.11).

Чтобы найти зависимость положения уровня Ферми F от температуры, нам не нужно заново решать уравнение (15.1), а можно воспользоваться соотношением (5.1), справедливым для невырожденных полупроводников. Это дает

$$F - E_c = kT \ln \frac{n_1}{2N_c}\left(\sqrt{1+\frac{4N_d}{n_1}} - 1\right). \qquad (15.4)$$

Эта зависимость показана на рис. 5.10.

При низких температурах (определяемых тем же условием, что и выше) формула (15.4) дает

$$F - E_c = \frac{1}{2}(E_d - E_c) + \frac{1}{2} kT \ln\left(\frac{g_0}{g_1}\frac{N_d}{N_c}\right).$$

При $T \to 0$ уровень Ферми F располагается посередине между E_c и E_d.

В случае некомпенсированных акцепторов справедливы аналогичные соотношения.

§ 16. Взаимная компенсация доноров и акцепторов

Случай примеси одного типа, когда влиянием других примесей можно пренебречь, встречается сравнительно редко. Дело в том, что современная техника очистки полупроводниковых материалов, несмотря на очень высокий уровень разработки, даже для такого хорошо освоенного полупроводника, как германий, позволяет снизить концентрацию остаточных примесей до $\sim 10^{10} \div 10^9$ см$^{-3}$, но не устранить их вовсе. Поэтому в реальных полупроводниках мы обычно имеем, кроме умышленно введенных доноров, некоторую концентрацию компенсирующих их акцепторов (или наоборот). Наличие же даже малой концентрации компенсирующей примеси

может при известных условиях существенно изменить температурную зависимость концентрации носителей заряда и положения уровня Ферми. Поэтому мы рассмотрим случай полупроводника, содержащего одновременно доноры и акцепторы.

Однако, прежде чем рассматривать общий случай, мы остановимся на одном простом, но практически важном примере, когда в полупроводнике имеются мелкие доноры (с уровнями вблизи зоны проводимости) и мелкие акцепторы (с уровнями вблизи дырочной зоны). Кроме того, мы предположим, что рассматриваемые интервалы температур и концентраций таковы, что уровень Ферми лежит между уровнями доноров и акцепторов и удален от тех и других по крайней мере на несколько kT. Такое положение, в частности, мы имеем в германии и кремнии, содержащих доноры V группы и акцепторы III группы, при концентрациях, меньших $\sim 10^{17}$ см$^{-3}$, и в области температур примерно от комнатной и до температуры твердого азота (~ 60 К). В этом случае все доноры практически полностью лишены электронов (так как уровень Ферми лежит ниже уровня доноров) и заряжены положительно, а все акцепторы практически полностью заполнены электронами (так как уровень Ферми расположен выше их уровней) и заряжены отрицательно. Поэтому условие нейтральности принимает простой вид:

$$n - p = N_d - N_a. \tag{16.1}$$

Если $N_d > N_a$, то $n > p$ и мы будем иметь полупроводник n-типа. Если при этом еще температура не слишком велика, так что концентрацией неосновных носителей можно пренебречь, то

$$n \simeq N_d - N_a. \tag{16.1а}$$

Концентрация электронов в зоне оказывается такой, как если бы в полупроводнике имелись одни доноры, но с меньшей концентрацией, так как часть N_a полной концентрации доноров компенсирована акцепторами.

Если концентрация акцепторов больше концентрации доноров, то мы будем иметь полупроводник p-типа и концентрация дырок в примесной области будет $p \simeq N_a - N_d$.

Если, наконец, концентрации доноров и акцепторов равны друг другу, то равенство (16.1) дает $n = p$. Так как, кроме того, для невырожденного полупроводника всегда $np = n_i^2$, то

$$n = p = n_i, \tag{16.2}$$

т. е. концентрация электронов и дырок будет такая же, как и при полном отсутствии каких бы то ни было примесей. В этом случае носители возникают только за счет возбуждений зона — зона.

Полученные результаты показывают, что при суждении о степени чистоты полупроводниковых материалов по измерению концентрации носителей необходимо соблюдать осторожность, так как

эта концентрация может быть мала не вследствие чистоты материала, а в результате взаимной компенсации доноров и акцепторов. Поэтому для окончательного решения вопроса приходится прибегать к дополнительным исследованиям, например к измерению подвижности.

§ 17. Компенсированные полупроводники

Рассмотрим опять полупроводник n-типа, содержащий простые доноры с концентрацией N_d и компенсирующие акцепторы с концентрацией $N_a < N_d$. Будем по-прежнему считать полупроводник невырожденным и рассмотрим область примесной проводимости, однако будем теперь рассматривать широкую область температур, включая и очень низкие температуры, когда ионизация доноров может быть неполной.

Условие нейтральности (13.4) для этого случая принимает вид

$$\frac{N_d}{1+\frac{g_1}{g_0}\exp\frac{F-E_d}{kT}} = n + N_a. \tag{17.1}$$

Выражая, как и выше, экспоненту через концентрацию электронов n, это условие можно представить в виде

$$\frac{n(n+N_a)}{N_d-N_a-n} = n_1(T), \tag{17.2}$$

где $n_1(T)$ по-прежнему выражается формулой (9.5). При $N_a = 0$ это уравнение переходит в ранее полученное (15.2). Оно опять приводит к квадратному уравнению относительно n. Для очень низких температур, когда $n \ll N_a$, $N_d - N_a$, уравнение (17.2) дает

$$n = \frac{N_d-N_a}{N_a}\frac{g_0}{g_1}N_c\exp\left(-\frac{\mathscr{J}_d}{kT}\right). \tag{17.3}$$

Следовательно, в координатах $\ln(nT^{-3/2})$ и $1/T$ зависимость $n(T)$ имеет опять вид прямой линии. Однако, в отличие от случая одних некомпенсированных доноров, наклон этой прямой равен $\mathscr{J}_d/k$, т. е. соответствует не половине, а полной энергии ионизации $\mathscr{J}_d$. Из выражения (17.3) также видно, что концентрация компенсирующих акцепторов сильно влияет на концентрацию электронов в зоне и может изменять ее на много порядков.

В общем случае примесной проводимости квадратное уравнение (17.2) дает

$$n = \frac{1}{2}(N_a+n_1)\left(\sqrt{1+\frac{4(N_d-N_a)n_1}{(N_a+n_1)^2}}-1\right). \tag{17.4}$$

При достаточно высоких температурах, когда $\frac{(N_d-N_a)n_1}{(N_a+n_1)^2} \ll 1$ и, кроме того, $n_1 \gg N_a$, мы имеем

$$n \simeq N_d - N_a,$$

т. е. полученную ранее формулу (16.1а). Эту область температур иногда называют «областью истощения» доноров.

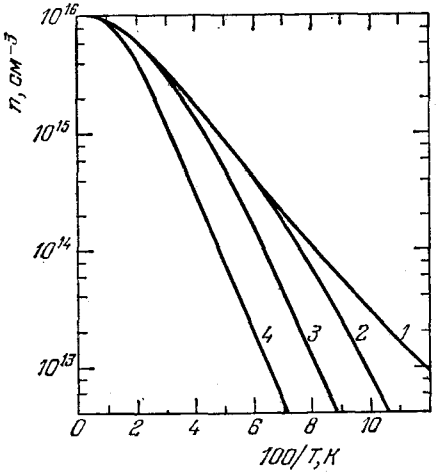


Рис. 5.11. Зависимость концентрации электронов от температуры в германии, содержащем доноры V группы (принято $E_c - E_d = 0{,}01$ эВ), частично компенсированные акцепторами III группы [1]. Для всех кривых $N_d - N_a = 10^{16}$ см$^{-3}$. *1* — $N_a = 0$; *2* — $N_a = 10^{14}$ см$^{-3}$; *3* — $N_a = 10^{15}$ см$^{-3}$; *4* — $N_a = 10^{16}$ см$^{-3}$. При расчетах положено $m_n = 0{,}25\, m_0$, $g_0/g_1 = \frac{1}{2}$.

Зависимость n от T для конкретного случая германия с донорами V группы, частично компенсированными акцепторами III группы, показана на рис. 5.11. Для случая некомпенсированных доноров кривая *1* при низких температурах имеет наклон, соответствующий половине энергии ионизации доноров (ср. § 15). При наличии компенсации наклон соответствует полной энергии ионизации. Следует подчеркнуть, однако, что при малой степени компенсации (кривая *2*) имеется область температур (она соответствует условию $N_a \ll n \ll N_d$), в которой наклон отвечает тоже половине энергии ионизации и лишь при достаточном понижении температуры этот наклон удваивается.

Чтобы найти температурный ход уровня Ферми, мы, как и раньше, воспользуемся соотношением (5.1). Это дает

$$F - E_c = kT \ln \frac{N_a + n_1}{2N_c} \left(\sqrt{1 + \frac{4(N_d - N_a) n_1}{(N_a + n_1)^2}} - 1 \right). \qquad (17.5)$$

Зависимость $F - E_c$ от T для тех же четырех степеней компенсации, что на рис. 5.11, показана на рис. 5.12. Из рисунка видно, что в частично компенсированном полупроводнике при низких температурах ход уровня Ферми существенно отличается от такового в некомпенсированном полупроводнике. При низких температурах (количественное условие см. выше) формула (17.5) дает

$$F = E_d - kT \ln \left(\frac{N_a}{N_d - N_a} \frac{g_1}{g_0} \right).$$

При $T \to 0$ F стремится к E_d, в то время как в некомпенсированном полупроводнике F располагается посередине между уровнями E_c и E_d.

Для полупроводника p-типа с акцепторами, частично компенсированными донорами, мы получили бы

$$\frac{p(N_d+p)}{N_a-N_d-p}=p_1(T), \tag{17.6}$$

где $p_1(T)$ дается формулой (9.13). В частности, при низких температурах концентрация дырок выражается формулой, аналогичной (17.3):

$$p=\frac{g_1}{g_0}\frac{N_a-N_d}{N_d}N_v\exp\left(-\frac{\mathscr{I}_a}{kT}\right). \tag{17.7}$$

Здесь $\mathscr{I}_a=E_a-E_v$ есть энергия ионизации акцептора (энергия отщепления дырки от него). При очень низких температурах уровень Ферми стремится к уровню энергии основной примеси, т. е. к E_a.

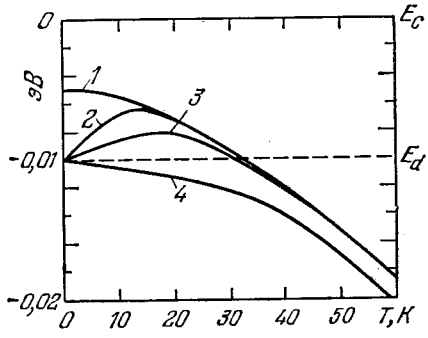


Рис. 5.12. Зависимость уровня Ферми от температуры в германии с частично компенсированными донорами V группы для тех же образцов, что на рис. 5.11.

Рассуждая, как и выше, легко найти температурную зависимость концентрации электронов (дырок) для двух других возможных случаев: частично компенсированных акцепторов в полупроводнике n-типа и частично компенсированных доноров в полупроводнике p-типа.

§ 18. Определение энергетических уровней примесных атомов

Результаты предыдущего параграфа лежат в основе важного метода определения локальных энергетических уровней, создаваемых примесными атомами и структурными дефектами. Для этого изготовляют образцы, содержащие исследуемую примесь и, кроме того, компенсирующую примесь, концентрацию последней подбирают так, чтобы исследуемый уровень энергии был компенсирован частично. В этом случае при достаточном понижении температуры уровень Ферми располагается на частично компенсированном уровне (ср. рис. 5.12), а зависимость концентрации основных носителей от температуры в координатах $\ln(nT^{-3/2})$ и $1/T$ описывается прямой линией, наклон которой дает энергию ионизации.

Подобным образом можно исследовать и многозарядные центры. Рассмотрим возможные при этом случаи.

а. Многозарядные акцепторы в полупроводнике n-типа. Создаваемые ими энергетические уровни E_1, E_2, ..., E_j, ..., E_M здесь и везде в дальнейшем мы будем нумеровать в порядке следования уровней от валентной зоны к зоне прово-

димости. В образце содержатся еще легко ионизуемые компенсирующие доноры с концентрацией N_d. Последняя такова, что

$$(j-1)\,N_a < N_d < jN_a. \tag{18.1}$$

Это значит, что при температуре абсолютного нуля N_d электронов-доноров полностью заселяют все $(j-1)$ нижних акцепторных уровней и лишь частично заполняют уровень j. Положим, далее, что частично компенсированный j-й уровень лежит выше, чем середина запрещенной зоны E_i. Тогда при низких температурах мы будем иметь некоторую концентрацию электронов в зоне проводимости за счет теплового возбуждения электронов с уровня j, образования же дырок практически происходить не будет, и мы будем иметь полупроводник n-типа. Такой случай изображен на рис. 5.13. В этом случае в некоторой области пониженных температур (области примесной проводимости) мы будем иметь в качестве положительных частиц только ионизованные доноры, а в качестве отрицательных — электроны в зоне проводимости и отрицательно заряженные акцепторы. Условие нейтральности будет иметь вид

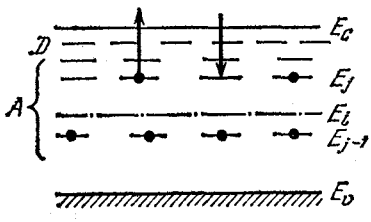


Рис. 5.13. Частично компенсированные многозарядные акцепторы. Полупроводник n-типа.

$$N_d = n + N_a \sum_{j=1}^{M} j f^{(j)}, \tag{18.2}$$

где $f^{(j)}$ в общем случае выражается формулой (11.8).

Задачу можно существенно упростить, если рассматривать лишь такую область температур, где возбуждение электронов происходит только с частично компенсированного уровня j. Это значит, что акцепторные центры практически будут находиться только в двух различных зарядных состояниях $(j-1)$ и j, и поэтому условие нормировки (11.5) примет простой вид:

$$f^{(j-1)} + f^{(j)} = 1. \tag{18.3}$$

Отсюда и из формулы (12.1) следует, что

$$f^{(j)} = \frac{1}{1 + \frac{g_{j-1}}{g_j} \exp \frac{E_j - F}{kT}}, \qquad f^{(j-1)} = \frac{1}{1 + \frac{g_j}{g_{j-1}} \exp \frac{F - E_j}{kT}}. \tag{18.4}$$

Условие нейтральности (18.2) в этом случае принимает простой вид:

$$N_d = n + (j-1)\,N_a f^{(j-1)} + jN_a f^{(j)},$$

или, с учетом (18.3),

$$jN_a - N_d + n = N_a f^{(j-1)} = \frac{N_a}{1 + \frac{g_j}{g_{j-1}} \exp \frac{F - E_j}{kT}}. \tag{18.5}$$

Поэтому, поступая, как в § 17, и используя соотношение

$$\exp \frac{F - E_j}{kT} = \frac{n}{N_c} \exp \frac{\mathscr{J}_j}{kT},$$

где $\mathscr{J}_j = E_c - E_j$ есть энергия отрыва электрона с уровня j, мы можем представить соотношение (18.5) в виде

$$\frac{n\,(jN_a - N_d + n)}{N_d - (j-1)\,N_a - n} = n_j\,(T). \tag{18.6}$$

Здесь введено обозначение

$$n_j\,(T) = \frac{g_{j-1}}{g_j}\,N_c \exp\left(-\frac{\mathscr{J}_j}{kT}\right). \tag{18.7}$$

Мы получили соотношение того же типа, что и (17.2). В частности, при достаточно низких температурах, когда $n \ll (jN_a - N_d)$ и $n \ll N_d - (j-1)\,N_a$, (18.6) дает

$$\frac{N_d - (j-1)\,N_a}{jN_a - N_d}\,n_j\,(T). \tag{18.8}$$

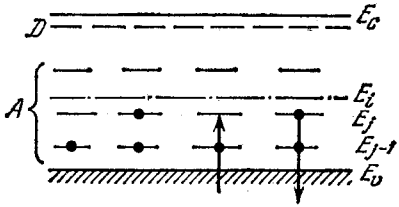


Рис. 5.14. Частично компенсированные многозарядные акцепторы. Полупроводник p-типа.

б. Многозарядные акцепторы в полупроводнике p-типа. Положим теперь, что частично заселенный электронами j-й уровень лежит в нижней половине запрещенной зоны ($E_j < E_i$, рис. 5.14). Тогда в некоторой области пониженных температур мы будем иметь только заброс электронов из валентной зоны на пустые уровни j (т. е. генерацию дырок и обратные переходы), а переходов электронов с заполненных уровней j в зону проводимости практически не будет, т. е. мы будем иметь полупроводник p-типа. В тех же предположениях, что и выше, условие нейтральности в этом случае будет

$$N_d + p = (j-1)\,N_a f^{(j-1)} + jN_a f^{(j)},$$

или, с учетом (18.3),

$$jN_a - N_d - p = \frac{N_a}{1 + \dfrac{g_j}{g_{j-1}} \exp \dfrac{F - E_j}{kT}}. \tag{18.9}$$

Так как основными носителями сейчас являются дырки, то мы выразим экспоненту через концентрацию дырок p:

$$\exp\frac{F - E_j}{kT} = \frac{N_v}{p} \exp\left(-\frac{\mathscr{J}_j}{kT}\right),$$

где $\mathscr{J}_j = E_j - E_v$ есть энергия отщепления дырки от акцептора. Учитывая это соотношение, (18.9) можно представить в виде

$$\frac{p\,(N_d - (j-1)\,N_a + p)}{jN_a - N_d - p} = p_j\,(T), \tag{18.10}$$

где

$$p_j\,(T) = \frac{g_j}{g_{j-1}}\,N_v \exp\left(-\frac{\mathscr{J}_j}{kT}\right). \tag{18.11}$$

Эти выражения определяют зависимость концентрации дырок от температуры в области примесной проводимости. Если температуры достаточно низки, так что $p \ll N_d - (j-1)\,N_a$ и $p \ll jN_a - N_d$, то уравнение (18.10) приводит

к экспоненциальной зависимости

$$p=\frac{jN_a-N_d}{N_d-(j-1)N_a}\,p_j(T). \tag{18.12}$$

в. Многозарядные доноры в полупроводнике n-типа. Рассмотрим теперь случай многоэлектронных доноров, частично компенсированных мелкими акцепторами, и притом таким образом, что частично заселен электронами j-й уровень (нумерация уровней, как и в случае акцепторов, производится от E_v к E_c). Для этого, очевидно, концентрации доноров и акцепторов должны удовлетворять условию

$$(M-j)N_d<N_a<(M-j+1)N_d. \tag{18.13}$$

При этом положим, что $E_j>E_i$, так что в примесной области мы будем иметь полупроводник n-типа. Тогда условие нейтральности будет

$$n+N_a=N_d\sum_{j=0}^{M-1}(M-j)f^{(j)}. \tag{18.14}$$

Рассматривая опять только такую область примесной проводимости, в которой практически представлены доноры в двух зарядных состояниях $e(M-j)$ и $e(M-j+1)$, мы можем записать уравнение (18.14) в виде

$$N_a-(M-j)N_d+n=N_df^{(j-1)}=\frac{N_d}{1+\dfrac{g_j}{g_{j-1}}\exp\dfrac{F-E_j}{kT}}. \tag{18.15}$$

Выражая опять $\exp\dfrac{F-E_j}{kT}$ через концентрацию электронов n (полупроводник n-типа), мы получаем из (18.15) температурную зависимость концентрации электронов:

$$\frac{n\,(N_a-(M-j)N_d+n)}{(M-j+1)N_d-N_a-n}=n_j(T), \tag{18.16}$$

где $n_j(T)$ выражается опять формулой (18.7). При достаточном понижении температуры (18.16) дает

$$n=\frac{(M-j+1)N_d-N_a}{N_a-(M-j)N_d}\,n_j(T). \tag{18.17}$$

г. Многозарядные доноры в полупроводнике p-типа. Обратимся теперь к последнему случаю, когда многоэлектронные доноры, как и в предыдущем примере, частично компенсированы мелкими акцепторами, причем опять выполняется условие (18.13), но частично заполненный электронами j-й уровень доноров лежит в нижней половине запрещенной зоны ($E_j<E_i$). В этом случае в области примесной проводимости мы будем иметь полупроводник p-типа. Считая опять, что доноры существуют только в двух зарядных состояниях, и поступая, как и раньше, получим температурную зависимость концентрации дырок:

$$\frac{p\,[(M-j+1)N_d-N_a+p]}{N_a-(M-j)N_d-p}=p_j(T), \tag{18.18}$$

где $p_j(T)$ дается формулой (18.11). При низких температурах (18.18) дает

$$p=\frac{N_a-(M-j)N_d}{(M-j+1)N_d-N_a}\,p_j(T). \tag{18.19}$$

Таким образом, в любом из рассмотренных случаев при достаточном понижении температуры зависимость концентрации основных носителей от обратной температуры становится экспоненциальной, а энергия активации в выражении под знаком экспоненты равна энергии ионизации того из уровней, который заселен носителями заряда частично.

При практическом осуществлении таких измерений приготовляют образцы полупроводника, содержащего кроме исследуемой примеси еще компенсирующую примесь противоположного типа (контрлегирование), и отношение концентраций примесей обоего типа варьируют от образца к образцу. Если монокристаллические образцы приготовляют вытягиванием затравки из расплава, то обе эти примеси вводят в расплав. При этом весьма полезно то обстоятельство, что отношение концентраций данной примеси в твердой фазе $c_т$ и в жидкой фазе $c_ж$ у границы раздела жидкой и твердой фаз растущего кристалла, или, иначе, так называемый коэффициент распределения примеси

$$K = c_т/c_ж,$$

оказывается различным для разных примесей. Это приводит к тому, что распределение разных примесей по длине слитка оказывается различным, а следовательно, вырезая пластинки из разных частей слитка, можно получить образцы с различной степенью компенсации исследуемой примеси. Варьируя концентрацию компенсирующей примеси в достаточно широких пределах, можно определить число и положение всех альтернативных уровней энергии многоэлектронных центров, а учитывая еще и тип примеси (доноры или акцепторы), необходимой для компенсации, можно установить, являются ли исследуемые уровни донорными или акцепторными.

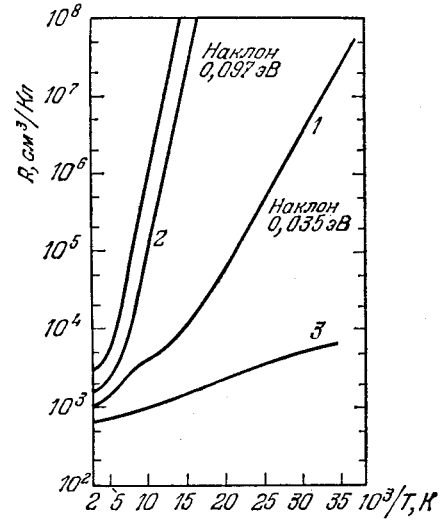


Рис. 5.15. Зависимость постоянной Холла от температуры для различных образцов германия, содержащих цинк, частично компенсированный сурьмой. *1* — $N_{Sb} < N_{Zn}$ (p-тип); *2* — $N_{Zn} < N_{Sb} < 2N_{Zn}$ (p-тип); *3* — $N_{Sb} > 2N_{Zn}$ (n-тип).

На рис. 5.15 приведен пример определения энергетических уровней атомов цинка в германии, которые являются двойными акцепторами *). Компенсирующие мелкие доноры создавались

*) По данным работы *W. W. Tyler, H. H. Woodbury,* Phys. Rev. 102, 647 (1956).

введением сурьмы. При $N_{Sb} < N_{Zn}$ зависимость логарифма постоянной Холла от $1/T$ в области низких температур становится прямолинейной, с наклоном, соответствующим энергии ионизации нижнего уровня цинка $E_1 - E_v$. Когда концентрация компенсирующих доноров возрастает и попадает в интервал $N_{Zn} < N_{Sb} < 2N_{Zn}$, наклон прямых становится бо́льшим и соответствует энергии ионизации второго уровня цинка $E_2 - E_v$. Если концентрация доноров делается еще больше ($N_{Sb} > 2N_{Zn}$), образцы становятся электронными (а не дырочными), концентрация основных носителей (электронов) определяется теперь уже возбуждением с избыточных атомов сурьмы и ее зависимость от температуры становится слабой.

Отметим в заключение, что в исследованиях такого рода часто измеряют не концентрацию основных носителей (постоянную Холла), а удельное сопротивление, что экспериментально проще. Так как в частично компенсированных полупроводниках подвижность носителей заряда зависит от температуры гораздо слабее, чем их концентрация, то это приводит лишь к небольшой ошибке, которой часто можно пренебречь.

Глава VI

ЯВЛЕНИЯ В КОНТАКТАХ (МОНОПОЛЯРНАЯ ПРОВОДИМОСТЬ)

§ 1. Потенциальные барьеры

При контакте двух различных полупроводников или полупроводника с металлом в пограничных слоях возникают потенциальные барьеры, а концентрации носителей заряда внутри этих слоев могут сильно изменяться по сравнению с их значениями в объеме. Свойства приконтактных слоев зависят от приложенного внешнего напряжения, что приводит в ряде случаев к сильно нелинейной вольтамперной характеристике контакта. Нелинейные свойства контактов используют для выпрямления электрического тока, для преобразования, усиления и генерации электрических колебаний и для других целей.

В настоящей главе мы будем предполагать, что ток образован только носителями заряда одного типа (монополярная проводимость). Явления в контактах в случае биполярной проводимости будут рассмотрены в гл. VIII.

Поясним образование потенциальных барьеров на примере контакта металл — электронный полупроводник.

На рис. 6.1, *а* показана энергетическая диаграмма для обоих тел до контакта. Отмечены: уровень энергии покоящегося электрона в вакууме E_0, края зон E_c и E_v в полупроводнике, а также положения уровней Ферми в металле $F_м$ и в полупроводнике $F_п$. Разность $E_0 - E_c = \chi$ есть электронное сродство полупроводника.

На рис. 6.1, *б* изображена ситуация сразу после образования контакта. Электроны из полупроводника получают возможность переходить в металл, создавая некоторую плотность тока j_1, а электроны металла — в полупроводник, образуя плотность тока $-j_2$. Эти токи, вообще говоря, не равны по величине. Если, например, $j_1 < j_2$, то полупроводник будет заряжаться отрицательно, а металл — положительно до тех пор, пока оба тока не скомпенсируют друг друга.

В установившемся состоянии мы получим энергетическую диаграмму рис. 6.1, *в*. Края энергетических зон окажутся изогнутыми вниз, а концентрация электронов в приконтактном слое — больше, чем в объеме (обогащенный слой).

В противоположном случае $j_1 > j_2$ установившееся искривление зон будет изображаться рис. 6.1, *г* и мы получим обедненный прикон-

тактный слой. Высота потенциального барьера (со стороны полупроводника) равна $-eu_k$, где u_k — разность потенциалов в глубине полупроводника и в контактной плоскости, а e — абсолютная величина заряда электрона.

На рис. 6.1 мы показывали искривление зон только в полупроводнике. Оно, конечно, существует и в металле. Однако толщина слоев объемного заряда и падение напряжения на них уменьшаются с увеличением концентрации носителей заряда (см. § 9 и § VIII.5). Поэтому в металле они пренебрежимо малы по сравнению с их значениями в полупроводнике.

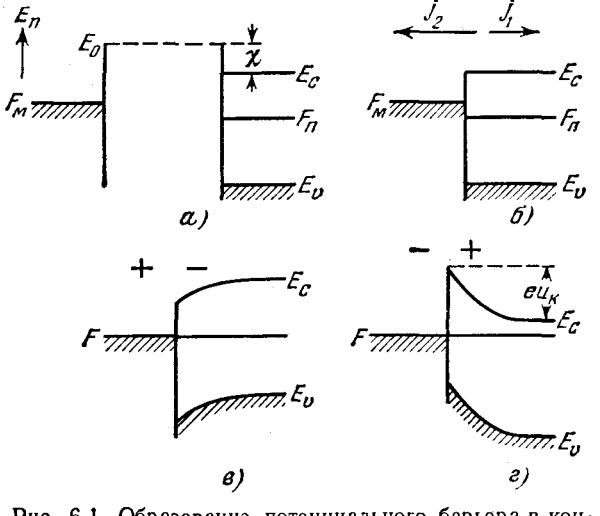


Рис. 6.1. Образование потенциального барьера в контакте металл — полупроводник: *а*) до контакта; *б*) при контакте, нет равновесия; *в*) равновесие, обогащенный слой; *г*) равновесие, обедненный слой.

При контакте двух полупроводников качественная картина явлений аналогична. Однако здесь уже необходимо учитывать существование барьеров в обоих контактирующих телах (см. § 9).

Отметим, что влияние потенциальных барьеров на электрический ток существенно проявляется только в контактах невырожденных полупроводников. Это связано с тем, что концентрация электронов в полупроводниках намного меньше, чем в металлах в силу особенностей энергетической структуры полупроводников (существования запрещенной зоны энергий и положения уровня Ферми внутри этой зоны). Вследствие этого, во-первых, толщина потенциальных барьеров оказывается больше длины волны де Бройля для электронов, так что квантовомеханический туннельный эффект в них не имеет места. И, во-вторых, основная часть электро-

нов сосредоточена у дна зоны проводимости, и поэтому электроны «чувствуют» потенциальный барьер. В случае же контакта двух металлов, так как концентрация электронов в них велика, толщина барьеров становится меньше длины волны де Бройля и электроны свободно проходят сквозь барьеры в результате туннельного эффекта. По этим причинам контакты именно полупроводников (с металлами или другими полупроводниками) оказываются особенно важными для технического использования.

§ 2. Плотность тока. Соотношение Эйнштейна

Если концентрация носителей заряда изменяется в пространстве, то плотность тока определяется не только дрейфом частиц в электрическом поле, но и их диффузией. Если коэффициент диффузии электронов есть D_n, то плотность конвекционного тока электронов равна

$$\mathbf{j}_n = \mathbf{j}_{\text{др}} + \mathbf{j}_{\text{диф}} = \epsilon n \mu_n \mathfrak{E} + eD_n \nabla n. \tag{2.1}$$

Здесь μ_n — абсолютная величина подвижности электронов, а ток диффузии записан со знаком «+», так как для отрицательных частиц направление потока диффузии противоположно направлению тока.

Аналогично, для плотности конвекционного тока дырок имеем

$$\mathbf{j}_p = ep\mu_p \mathfrak{E} - eD_p \nabla p. \tag{2.2}$$

Отметим, что выражения для токов диффузии и само понятие коэффициента диффузии имеют смысл, если изменение концентрации на длине свободного пробега l достаточно мало:

$$|\nabla n|\, l \ll n. \tag{2.3}$$

Для невырожденного полупроводника, согласно закону Больцмана,

$$n = n_0 \exp \frac{e\varphi}{kT} \tag{2.4}$$

и, следовательно,

$$\nabla n = n \frac{e}{kT} \nabla \varphi = -n \frac{e}{kT} \mathfrak{E}. \tag{2.5}$$

Тогда условие (2.3) принимает вид

$$\left| \frac{e\mathfrak{E} l}{kT} \right| \ll 1. \tag{2.6}$$

Применительно к потенциальному барьеру под $\mathfrak{E}$ следует понимать некоторое среднее поле внутри барьера $\overline{\mathfrak{E}}$. По порядку величины $\mathfrak{E} \sim u_k / L_э$, где $L_э$ — длина экранирования электрического поля (см. § 7). Поэтому условие (2.6) можно записать также в виде

$$\left| \frac{eu_k}{kT} \right| \frac{l}{L_э} \ll 1. \tag{2.7}$$

Подвижность и коэффициент диффузии не являются независимыми между собой величинами. Действительно, для данного типа частиц с заданной эффективной массой подвижность зависит только от среднего времени свободного пробега $\langle\tau\rangle$. Но и коэффициент диффузии частиц определяется той же самой величиной. Поэтому между обеими величинами существует связь. Она особенно проста для случая, когда электронный или, соответственно, дырочный газы можно считать невырожденными.

Рассмотрим электроны в полупроводнике при наличии градиента концентрации и в состоянии термодинамического равновесия. Это может быть, например, один из приповерхностных слоев, изображенных на рис. 6.1, в отсутствие тока. Тогда, подставляя ∇n из формулы (2.5) в правую часть выражения (2.1) и полагая $j_n = 0$, находим

$$\frac{\mu_n}{D_n} = \frac{e}{kT}. \tag{2.8}$$

Такую же формулу мы получили бы и для μ_p/D_p, рассматривая невырожденный газ дырок. Соотношение (2.8) было получено впервые Эйнштейном в теории броуновского движения. Однако оно имеет универсальный характер и применимо к любым частицам, если они образуют невырожденный газ. Оно позволяет по известной подвижности непосредственно найти коэффициент диффузии (который экспериментально определяется гораздо труднее), и обратно.

Полученный результат легко обобщить на случай произвольно вырожденного газа. Для этой цели нужно лишь воспользоваться формулой (V.4.4а), согласно которой концентрация электронов n есть функция безразмерного химического потенциала

$$\zeta^* = \zeta_0^* + \frac{e\varphi}{kT}. \tag{2.9}$$

В этом случае

$$\nabla n = \frac{dn}{d\zeta^*}\frac{e}{kT}\nabla\varphi = -\frac{dn}{d\zeta^*}\frac{e}{kT}\mathscr{E}. \tag{2.10}$$

Подставляя это выражение опять в правую часть формулы (2.1) при $j_n = 0$, мы получаем

$$n\mu_n - \frac{e}{kT}D_n\frac{dn}{d\zeta^*} = 0.$$

Отсюда

$$\frac{\mu_n}{D_n} = \frac{e}{kT}\frac{d(\ln n)}{d\zeta^*}. \tag{2.11}$$

Рассуждая аналогично, мы имеем для дырок

$$\frac{\mu_p}{D_p} = \frac{e}{kT}\frac{d(\ln p)}{d\eta^*}. \tag{2.12}$$

Для невырожденного электронного газа

$$n = N_c \exp \zeta^*, \qquad \frac{d(\ln n)}{d\zeta^*} = 1$$

и формула (2.11) переходит в (2.8). То же самое справедливо и для дырок.

Мы получили соотношение Эйнштейна, предполагая термодинамическое равновесие. Однако этим соотношением можно пользоваться и при наличии тока, если только плотность тока не становится настолько большой, что она приводит к существенному нарушению функции распределения электронов (подробнее см. гл. XVI).

§ 3. Условия равновесия контактирующих тел

Рассмотрим теперь, от чего зависит высота потенциального барьера. Ответ на этот вопрос непосредственно следует из общих условий термодинамического равновесия. А именно, из статистической физики известно, что если два (или несколько) тел способны обмениваться друг с другом частицами, то в состоянии термодинамического равновесия электрохимический потенциал, т. е. уровень Ферми (ср. § V.11), отсчитанный от произвольного, но одинакового уровня энергии, имеет одно и то же значение во всех частях системы:

$$F = \text{const}. \qquad (3.1)$$

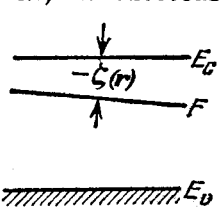


Рис. 6.2. К условию равновесия электронного газа.

Это условие имеет простой физический смысл. Допустим временно, что оно не выполняется и F изменяется в пространстве. Положим, далее, что в полупроводнике имеется электрическое поле, так что потенциал φ и концентрация электронов n зависят от координат. В этом случае энергетические зоны будут искривлены и химический потенциал $F - E_c = \zeta(\mathbf{r})$ будет изменяться в пространстве (рис. 6.2). Найдем теперь полную плотность конвекционного тока электронов. Так как n есть функция безразмерного химического потенциала $\zeta^* = \zeta/kT$, то можно написать

$$\nabla n = \frac{dn}{d\zeta^*} \nabla \zeta^* = \frac{1}{kT} \frac{dn}{d\zeta^*} \nabla \zeta.$$

Тогда выражение для плотности тока электронов (2.1) можно представить в виде

$$\mathbf{j}_n = \mu_n n \left(-e \nabla \varphi + \frac{e}{kT} \frac{D_n}{\mu_n} \frac{d(\ln n)}{d\zeta^*} \nabla \zeta \right).$$

Рассматривая общий случай произвольного вырождения электронного газа и выражая здесь D_n/μ_n по формуле (2.11), находим

$$\mathbf{j}_n = \mu_n n \nabla (\zeta - e\varphi).$$

Но $(\zeta - e\varphi)$ есть не что иное, как электрохимический потенциал F (ср. формулу (V.11.12)). Поэтому

$$\mathbf{j}_n = \mu_n n \nabla F. \tag{3.2}$$

Если бы мы вместо электронов рассматривали ток дырок, то получили бы совершенно аналогичное выражение:

$$\mathbf{j}_p = \mu_p p \nabla (\zeta - e\varphi) = \mu_p p \nabla F. \tag{3.2a}$$

Полученные результаты показывают, что условие равновесия (3.1) означает просто отсутствие тока. Это и понятно, так как ток есть нарушение термодинамического равновесия. Из формул (3.2) и (3.2а) также видно, что полная плотность тока пропорциональна градиенту уровня Ферми. Поэтому уровень Ферми изменяется в пространстве особенно сильно там, где концентрация носителей заряда мала (например, в обедненных приконтактных слоях), в то время как в областях с большими значениями $\mu_n n$ и $\mu_p p$ изменение F может быть очень малым.

Возвращаясь теперь к контакту металл—полупроводник, мы имеем, что в отсутствие тока $F_м = F_п$. Это и было отмечено на рис. 6.1. Поэтому высота барьера для электронов (со стороны полупроводника) равна

$$eu_k = F_п - F_м, \tag{3.3}$$

где $F_п$ и $F_м$ — положение уровней Ферми в полупроводнике и, соответственно, в металле до контакта.

§ 4. Термоэлектронная работа выхода

Для нахождения глубины залегания уровней Ферми, которые, согласно формуле (3.3), определяют высоту потенциального барьера в контакте, удобно пользоваться значениями термоэлектронной работы выхода. Термоэлектронная работа может быть непосредственно измерена на опыте и для многих материалов является известной характеристикой. Чтобы выяснить интересующую нас связь, вычислим плотность тока насыщения термоэлектронной эмиссии j_s, т. е. заряд, переносимый электронами, испаряющимися в вакууме в 1 с с 1 см^2 поверхности проводника, находящегося при температуре T. Для этого представим себе, что наш проводник заключен в адиабатическую оболочку, поддерживаемую при температуре T. Тогда над проводником будет электронный газ с некоторой концентрацией электронов $n_в$ и этот газ будет находиться в термодинамическом равновесии с проводником. Отсюда следует, что количество электронов, испаряющихся из проводника в вакуум, должно быть равно количеству электронов, приходящих из вакуума в проводник. Так как электронный газ над проводником не вырожден, то скорости электронов в нем распределены по закону

Максвелла. Поэтому мы воспользуемся хорошо известным из кинетической теории газов выражением для числа частиц, встречающих 1 см^2 стенки в 1 с: $^1/_4 nv_T$, где n — концентрация частиц, а v_T — среднее значение абсолютной величины тепловой скорости. Следовательно, для электронного тока насыщения можно написать

$$j_s = {}^1/_4 n_в v_T e, \tag{4.1}$$

где v_T, согласно закону Максвелла, равна

$$v_T = \sqrt{\frac{8kT}{\pi m_0}}. \tag{4.2}$$

Величину $n_в$ мы можем непосредственно найти из формул гл. V для концентрации электронов в зоне проводимости. Действительно, мы видели, что в случае параболического изотропного закона дисперсии для невырожденного полупроводника концентрация электронов в зоне выражается формулой (V.5.1). Для электронного газа в вакууме закон дисперсии точно выражается формулой

$$E = E_0 + \frac{p^2}{2m_0}, \tag{4.3}$$

т. е. имеет тот же вид, с тем отличием, что вместо энергии E_c сейчас входит E_0 — энергия покоящегося электрона в вакууме, а вместо эффективной массы — масса изолированного электрона m_0. Поэтому, заменяя в формуле (V.5.1) E_c на E_0 и в формуле (V.4.3) m на m_0, находим

$$n_в = 2\left(\frac{2\pi m_0 kT}{(2\pi\hbar)^2}\right)^{3/2} \exp\left(-\frac{\Phi}{kT}\right). \tag{4.4}$$

Здесь введено обозначение

$$\Phi = E_0 - F. \tag{4.5}$$

Подставляя (4.4) и (4.2) в (4.1), мы находим окончательно

$$j_s = AT^2 \exp\left(-\frac{\Phi}{kT}\right), \tag{4.6}$$

где

$$A = \frac{4\pi m_0 ek^2}{(2\pi\hbar)^3} \tag{4.7}$$

есть универсальная постоянная, которая должна быть одинакова для всех веществ.

Формула (4.6) известна в вакуумной электронике как формула Ричардсона—Дэшмана. При этом, по определению, энергия Φ есть *термоэлектронная работа выхода* для данного вещества. Из формулы (4.5) видно, что Φ является разностью между энергией покоящегося электрона в вакууме у поверхности образца и уровнем Ферми в данном веществе (рис. 6.3).

Остановимся подробнее на физическом смысле выражения (4.5). Рассмотрим сначала металлы. В них уровень Ферми совпадает с максимальной энергией электронов проводимости (если отвлечься от слабой «размазки» распределения Ферми при $T \neq 0$). При этом концентрация электронов в зоне практически не зависит от температуры (ср. § V.6). Поэтому Ф в металлах совпадает с работой, необходимой для удаления электрона с максимальной энергией из металла в вакуум (рис. 6.3). Однако при применении формулы (4.5) к невырожденным полупроводникам может возникнуть недоумение, так как в этом случае уровень Ферми F лежит внутри запрещенной зоны энергий (рис. 6.4) и поэтому Ф не соответствует работа удаления какого-либо реально существующего электрона. Причина этого, кажущегося странным, результата заключается в том, что j_s зависит не только от энергии электронов, но еще и от их концентрации в зоне, а эта последняя в полупроводниках, в отличие от металлов, может сильно зависеть от температуры. Формулы (4.5) и (4.6) учитывают оба эти обстоятельства, и именно поэтому, как показывает приведенный выше расчет, в формулу (4.5) входит уровень Ферми.

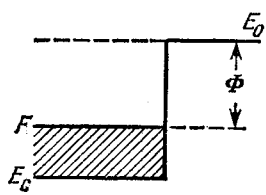


Рис. 6.3. Термоэлектронная работа выхода из металла.

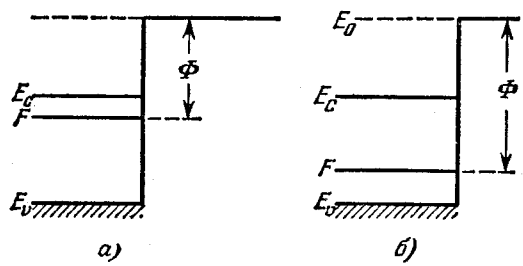


Рис. 6.4. Термоэлектронная работа выхода из полупроводника: *а*) n-тип; *б*) p-тип.

Из сказанного также следует, что в полупроводниках работа выхода может сильно зависеть от введенных примесей. При преобладании доноров мы будем иметь проводимость n-типа и уровень Ферми будет лежать в верхней половине запрещенной зоны (рис. 6.4, *а*). Напротив, в случае преобладания акцепторов уровень Ферми будет в нижней половине этой зоны (рис. 6.4, *б*), и поэтому работа выхода во втором случае будет больше.

Отметим еще, что работа выхода может очень сильно зависеть от самых ничтожных загрязнений поверхности, даже если чужеродные атомы образуют единицы атомных слоев. Причина такого влия-

ния состоит в том, что чужеродные атомы на поверхности могут обмениваться электронами с объемом кристалла и поэтому, как правило, существуют в виде положительно или отрицательно заряженных ионов, а это приводит к искривлению энергетических зон у поверхности. При отрицательном заряде поверхности зоны искривляются вверх (рис. 6.5), при положительном — вниз. При этом электронное сродство $(E_0 - E_c)$, зависящее только от структуры кристалла, у поверхности не изменяется. Поэтому работа выхода $(E_0 - F)$ увеличивается при искривлении зон вверх и уменьшается при искривлении вниз. Изменение работы выхода есть $\Delta\Phi = -e\varphi_s$, где φ_s — потенциал поверхности (относительно объема кристалла).

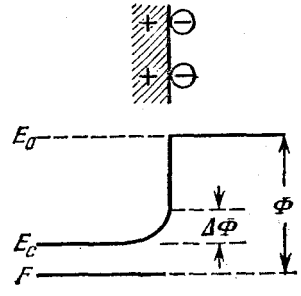


Рис. 6.5. Изменение работы выхода вследствие искривления зон у поверхности.

Вследствие сильного влияния загрязнения поверхности, определение термоэлектронных работ выхода требует очень тщательных опытов. Однако в настоящее время они для многих веществ достаточно хорошо известны. Пользуясь термоэлектронными работами выхода, формулу (3.3) для высоты потенциального барьера в контакте можно представить в удобном виде:

$$eu_k = \Phi_м - \Phi_п. \tag{4.8}$$

§ 5. Контактная разность потенциалов

Для определения термоэлектронной работы выхода по термоэлектронному току исследуемое вещество нужно нагревать до высокой температуры, чтобы термоэлектронные токи были достаточно сильны. А для многих веществ с невысокой температурой плавления и большой работой выхода это не всегда возможно. Однако работу выхода можно определить и другими методами, не требующими нагревания тела. В частности, это удобно сделать, измеряя контактную разность потенциалов, т. е. разность потенциалов между несоприкасающимися поверхностями двух различных проводников, находящихся в электронном равновесии. Последнее можно осуществить, соединяя оба проводника металлической проволокой (однако это может быть и не обязательным, так как равновесие может установиться и за счет слабого обмена электронами через вакуум).

Рассмотрим сначала случай двух металлов. Если оба тела в целом не заряжены, то в отсутствие контакта между ними электрического поля нет, а следовательно, энергия электрона в вакууме E_0

и электрический потенциал φ постоянны (рис. 6.6, *а*). Термоэлектронная работа выхода металла *I* есть $\Phi_1 = E_0 - F_1$, а металла *II* $\Phi_2 = E_0 - F_2$. При соединении обоих тел проводником устанавливается электронное равновесие и уровни Ферми F_1 и F_2 уравниваются. Но при этом вершины потенциальных ям уже оказываются

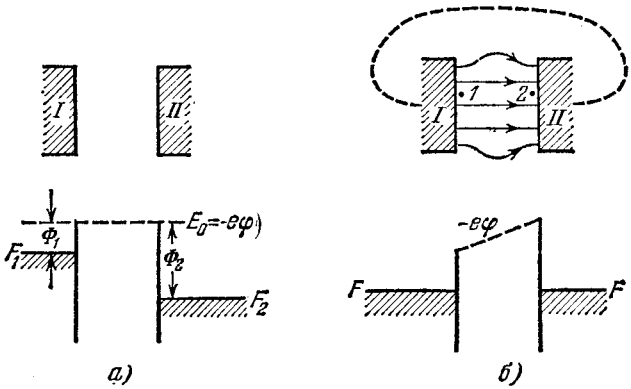


Рис. 6.6. К определению контактной разности потенциалов двух металлов.

на разной высоте, и поэтому потенциальная энергия электрона $-e\varphi$ как функция координат уже не изображается горизонтальной прямой. Это значит, что между телами появляется электрическое поле, а на поверхности тел возникают заряды. Контактная разность потенциалов u_k между телами *I* и *II*, по определению, есть разность потенциалов между любыми точками *1* и *2* (рис. 6.6, *б*), расположенными в непосредственной близости от поверхностей тел *I* и *II*, но находящимися вне этих тел. Из рис. 6.6 видно, что

$$-eu_k = -e(\varphi_1 - \varphi_2) = F_2 - F_1 = \Phi_1 - \Phi_2. \quad (5.1)$$

Таким образом, контактная разность потенциалов определяется разностью термоэлектронных работ выхода обоих тел.

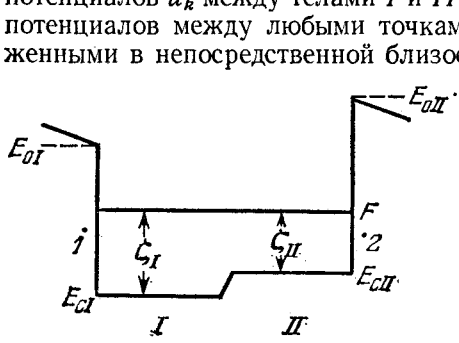


Рис. 6.7. Энергетическая диаграмма двух контактирующих металлов.

В связи со сказанным полезно более подробно рассмотреть энергетическую диаграмму двух контактирующих тел (рис. 6.7). На границе вакуум—металл *I* имеется скачок электростатического потенциала, и поэтому при переходе из вакуума в металл потенциальная энергия

электрона изменяется на величину

$$-e(\varphi_I - \varphi_1) = E_{cI} - E_{0I}. \tag{5.2}$$

Здесь φ_I — потенциал внутри металла, а φ_1 — его значение у поверхности металла в вакууме. На границе металл *I* — металл *II* также существует скачок потенциала $(\varphi_{II} - \varphi_I)$, определяемый условием равенства электрохимических потенциалов в обоих металлах (§ 3)

$$\zeta_I - e\varphi_I = \zeta_{II} - e\varphi_{II}.$$

При переходе электрона из металла *II* в вакуум его потенциальная энергия меняется на величину

$$-e(\varphi_2 - \varphi_{II}) = E_{0II} - E_{cII}. \tag{5.3}$$

И, наконец, при переходе электрона в вакууме из точки *2* в точку *1* изменение его энергии равно $-e(\varphi_1 - \varphi_2)$.

Рассмотрим теперь замкнутый контур, проходящий внутри контактирующих металлов и замыкающийся в вакуумном промежутке. Очевидно, что сумма всех скачков потенциала, встречающихся на этом пути, равна нулю, так как после обхода по этому пути мы возвращаемся вновь в исходную точку. Это дает

$$(E_{cI} - E_{0I}) + (\zeta_I - \zeta_{II}) + (E_{0II} - E_{cII}) - e(\varphi_1 - \varphi_2) = 0. \tag{5.4}$$

Но (см. рис. 6.7)

$$E_{0I} - E_{cI} - \zeta_I = E_{0I} - F = \Phi_1, \quad E_{0II} - E_{cII} - \zeta_{II} = E_{0II} - F = \Phi_2, \tag{5.5}$$

где Φ_1 и Φ_2 — термоэлектронные работы выхода металлов *I* и *II* соответственно. Поэтому соотношение (5.4) переходит в формулу (5.1).

Рассмотрим теперь случай, когда одно из тел есть металл, а другое — полупроводник. Соответствующие энергетические диаграммы показаны на рис. 6.8, *а* (до установления равновесия) и 6.8, *б* (при равновесии). Отличие от предыдущего случая двух металлов заключается лишь в том, что электрическое поле здесь частично проникает в полупроводник на некоторую глубину $L_э$ (длину экранирования, см. § 8). Поэтому полная разность потенциалов $(\Phi_1 - \Phi_2)/e$ распределяется между зазором (ширины d) и слоем объемного заряда. Если $d \gg L_э$ (что обычно всегда бывает при измерении контактной разности), то падением напряжения в слое объемного заряда можно пренебречь и мы опять получаем формулу (5.1). Это соотношение будет также справедливо и для двух полупроводников, если ширина зазора между ними достаточно велика (по сравнению с длинами экранирования в обоих полупроводниках).

Таким образом, высота потенциального барьера в контакте металл—полупроводник равна контактной разности потенциалов u_k. Поэтому в принципе, измеряя u_k, можно непосредственно определить и высоту барьера. Однако при этом нужно иметь в виду, что, измеряя u_k, мы определяем разность работ выхода при свободных поверхностях, которые могут иметь слои адсорбированных газов и другие загрязнения, влияющие на работу выхода. При

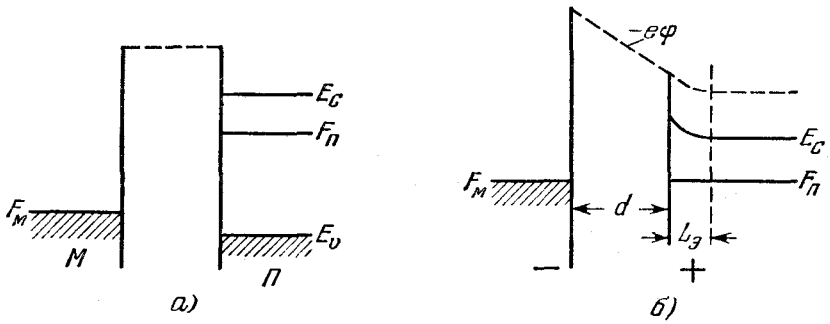


Рис. 6.8. Контактная разность потенциалов металла и полупроводника.

образовании же контакта состояния поверхностей, а следовательно и работы выхода, могут изменяться, и поэтому контактная разность u_k, измеренная в реальных опытах без надлежащих предосторожностей, может и не совпадать с высотой барьера при последующем образовании контакта.

§ 6. Распределение концентрации электронов и потенциала в слое объемного заряда

Рассмотрим теперь, от чего зависит толщина потенциального барьера в контактах. Положим, что все величины зависят лишь от одной координаты x, отсчитываемой вдоль нормали к плоскости контакта (рис. 6.9). Полупроводник будем считать невырожденным и, для определенности, n-типа. Тогда для нахождения распределения потенциала и концентрации электронов мы имеем следующую систему уравнений:

выражение для плотности тока

$$j = en\mu\mathscr{E} + \mu kT\frac{dn}{dx}, \tag{6.1}$$

уравнение Пуассона

$$\frac{\partial\mathscr{E}}{\partial x} = \frac{4\pi\rho}{\varepsilon} \tag{6.2}$$

и уравнение непрерывности

$$\frac{\partial\rho}{\partial t} = -\frac{\partial j}{\partial x}. \tag{6.3}$$

В уравнении (6.1) мы пренебрегли током смещения и, кроме того, выразили коэффициент диффузии через подвижность с помощью соотношения Эйнштейна (2.8). Пусть, далее, полупроводник содержит доноры с концентрацией N_d и акцепторы с концентрацией $N_a < N_d$. Тогда плотность объемного заряда ρ есть

$$\rho = e(N_d^+ - N_a^- - n), \qquad (6.4)$$

где концентрации заряженных доноров N_d^+ и акцепторов N_a^- в общем случае являются функциями координат и времени.

В настоящей главе мы ограничимся только стационарными состояниями. Тогда уравнение непрерывности дает

$$j = \text{const}.$$

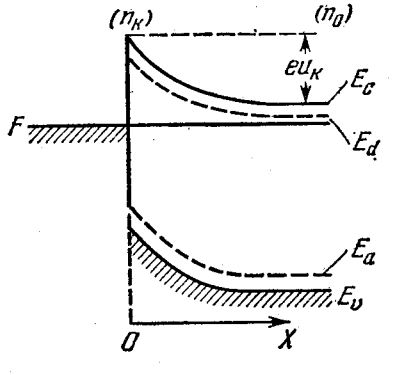


Рис. 6.9. Контакт полупроводника, содержащего мелкие доноры и акцепторы, с металлом.

Далее, будем сначала считать, что доноры и акцепторы достаточно мелкие, так что уровни доноров везде расположены выше уровня Ферми, а уровни акцепторов — ниже уровня Ферми хотя бы на несколько kT (рис. 6.9). Тогда доноры и акцепторы будут полностью ионизованы и $N_d^+ = N_d$, $N_a^- = N_a$. В этом случае исходная система уравнений приобретает простой вид:

$$j = -en\mu\frac{d\varphi}{dx} + \mu kT\frac{dn}{dx}, \qquad (6.5)$$

$$\frac{d^2\varphi}{dx^2} = \frac{4\pi e}{\varepsilon}(n - n_0). \qquad (6.6)$$

Здесь через n_0 обозначена постоянная концентрация электронов в глубине полупроводника, где $\rho = 0$.

Положим, что плоскость контакта расположена при $x = 0$, и будем отсчитывать потенциал от его значения в этой плоскости. Тогда граничные условия будут

$$\begin{aligned} x = 0: &\quad \varphi = 0, &\quad n = n_k; \\ x = \infty: &\quad \varphi = u_k + u, &\quad n = n_0. \end{aligned} \qquad (6.7)$$

Здесь u — внешняя приложенная разность потенциалов.

Граничная концентрация n_k в отсутствие тока есть заданная характеристика контакта, зависящая от природы металла и полупроводника:

$$n_k = n_0 \exp\left(-\frac{eu_k}{kT}\right). \qquad (6.8)$$

При наличии тока концентрация на границе $n(0)$, вообще говоря, уже не равна равновесной n_k, и точное определение этой концентра-

ции сложно. Однако мы положим, что это изменение невелико, так что n (0) можно считать приближенно не зависящим от тока и равным его равновесному значению n_k.

Чтобы оценить величину токов, при которых это предположение еще допустимо, можно воспользоваться следующими соображениями. Электроны в полупроводнике, уже достигшие плоскости контакта, далее не встречают потенциального барьера и все проходят в металл. Без внешнего напряжения число таких электронов, пересекающих 1 см^2 контактной плоскости в 1 с, есть $\frac{1}{4} n_k v_T$. Такое же число электронов переходит в обратном направлении из металла в полупроводник. Поэтому плотность тока через контакт можно выразить через граничные концентрации n (0) и n_k:

$$j = {}^1/_4 e v_T [n(0) - n_k]. \tag{6.9}$$

Отсюда видно, что n (0) $\simeq n_k$, если j мало по сравнению с каждым из слагаемых в (6.9). Полагая, например, $n_k \sim 10^{13}$ см$^{-3}$ и $v_T \sim$ $\sim 10^7$ см/с, мы получим для второго слагаемого в правой части (6.9) величину ~ 10 А/см^2. Поэтому, если плотность тока через контакт не превышает, скажем, 0,1 А/см^2, то разность между n (0) и n_k будет не более 1% и сделанное допущение будет хорошим приближением.

Мы рассмотрим сначала контакт в отсутствие тока. Тогда из уравнения (6.5) имеем

$$n = n_k \exp \frac{e\varphi}{kT}. \tag{6.10}$$

Подставляя это в (6.6), получаем уравнение

$$\frac{d^2\varphi}{dx^2} = \frac{4\pi e}{\varepsilon} n_k \left(\exp \frac{e\varphi}{kT} - \exp \frac{eu_k}{kT} \right) = \frac{4\pi e}{\varepsilon} n_0 \left(\exp \frac{e(\varphi - u_k)}{kT} - 1 \right), \tag{6.11}$$

из которого можно найти распределение потенциала φ (x).

Рассмотрим некоторые важные частные случаи.

§ 7. Длина экранирования

Положим, что искривление зон у поверхности мало, так что $eu_k/kT \ll 1$. Тогда экспоненты в уравнении (6.11) можно разложить в ряд и ограничиться первыми двумя членами разложения. После этого уравнение (6.11) принимает простой вид:

$$\frac{d^2\varphi}{dx^2} = \frac{\varphi - u_k}{L_D^2}, \tag{7.1}$$

где обозначено

$$L_D^2 = \frac{\varepsilon kT}{4\pi e^2 n_0}. \tag{7.2}$$

Его решение, удовлетворяющее граничным условиям (6.7), есть

$$\varphi = u_k\left[1 - \exp\left(-\frac{x}{L_D}\right)\right]. \tag{7.3}$$

Потенциал экспоненциально изменяется по мере удаления от границы контакта. Очевидно, что по такому же экспоненциальному закону будут изменяться напряженность поля и концентрация электронов, так как эти величины могут быть получены дифференцированием потенциала по координате, что опять приводит к экспоненциальной функции. Характеристическая длина L_D есть длина, на которой эти величины изменяются в e раз. Она получила название *длины экранирования Дебая*.

В полупроводниках с большой концентрацией электронов длина Дебая мала. Полагая, например, в формуле (7.2) $n_0 \sim 10^{15}$см^3 и $\varepsilon \sim 10$, мы находим для комнатной температуры $L_D \sim 10^{-5}$ см. При уменьшении n_0 длина Дебая L_D увеличивается.

Однако было бы ошибкой считать, что, достаточно уменьшая концентрацию электронов в зоне, можно сделать длину экранирования как угодно большой. При уменьшении n_0 длина Дебая L_D сначала увеличивается, но затем перестает зависеть от n_0 и остается постоянной. Это происходит потому, что при очень малой концентрации электронов экранирование определяется уже не электронами в зоне, а имеющимися в полупроводнике заряженными примесями. Чтобы выяснить, когда это имеет место, откажемся от предположения, что основная примесь (доноры) полностью ионизована, и согласно формуле (V.9.4) положим

$$N_d^+ = N_d\frac{n_1}{n+n_1}. \tag{7.4}$$

Здесь n_1 дается формулой (V.9.5). Тогда выражение для плотности объемного заряда будет

$$\rho = e\left(N_d\frac{n_1}{n+n_1} - N_a - n\right). \tag{7.5}$$

При этом в глубине полупроводника ($\rho = 0$)

$$N_d\frac{n_1}{n_0+n_1} - N_a - n_0 = 0. \tag{7.6}$$

Ограничиваясь, как и выше, слабым искривлением зон, имеем

$$n \simeq n_0\left[1+\frac{e}{kT}(\varphi - u_k)\right]. \tag{7.7}$$

Тогда

$$\frac{n_1}{n+n_1} \simeq \frac{n_1}{n_0+n_1}\left[1-\frac{e}{kT}\frac{n_0}{n_0+n_1}(\varphi-u_k)\right],$$

и, комбинируя это соотношение с формулами (7.5) и (7.6), получаем

$$\rho = -\frac{e^2 n_0}{kT\psi^2}(\varphi - u_k), \tag{7.8}$$

где через ψ^2 обозначена величина

$$\psi^2 = \frac{n_0+n_1}{N_a+n_1+2n_0}. \tag{7.9}$$

После этого уравнение Пуассона принимает вид

$$\frac{d^2\varphi}{dx^2} = \frac{\varphi - u_k}{L_D^2 \psi^2}. \tag{7.10}$$

Сравнивая это уравнение с уравнением (7.1), мы видим, что теперь длина экранирования равна

$$L_э = L_D \psi, \tag{7.11}$$

где L_D — по-прежнему есть длина Дебая, определяемая формулой (7.2).

Положим, что выполняются неравенства

$$N_a \ll n_1, \qquad n_0 \ll n_1. \tag{7.12}$$

Первое из них требует, чтобы концентрация компенсирующей примеси была достаточно мала. Второе неравенство, согласно формуле (7.4), обозначает, что доноры можно считать полностью ионизованными. Тогда $\psi \simeq 1$ и мы получаем рассмотренный раньше случай $L_э = L_D$.

Другой крайний случай мы имеем, когда n_0 становится малым по сравнению с N_a. Это происходит, например, в компенсированных полупроводниках в области примесной проводимости при достаточном понижении температуры. Тогда уровень Ферми локализован вблизи уровня энергии доноров E_d (ср. § V.17), а следовательно, $n_1 \simeq n_0$ и

$$\psi^2 \simeq \frac{2n_0}{N_a + 3n_0}.$$

Если при этом концентрация электронов становится настолько малой, что $n_0 \ll N_a$, то

$$\psi^2 \simeq \frac{2n_0}{N_a}, \qquad L_э^2 \simeq \frac{\varepsilon kT}{2\pi e^2 N_a}. \tag{7.13}$$

В этих условиях длина экранирования перестает зависеть от концентрации свободных электронов, а определяется концентрацией компенсирующей примеси. Физический смысл этого результата заключается в том, что при неполной ионизации примесей объемный заряд, экранирующий электрическое поле, создается не только перераспределением электронов, но и пространственным изменением заряда примесей. При малом n_0 последний эффект становится главным.

§ 8. Обогащенный контактный слой в отсутствие тока

Рассмотрим теперь распределение потенциала в случае обогащенного контактного слоя ($eu_k < 0$ и в несколько раз превышает kT) (рис. 6.10). При этом удобно раздельно рассматривать область вблизи объемного заряда контакта *1* и остальную толщу полупроводника *2*, где зоны можно считать уже неискривленными. В первом случае мы имеем

$$\exp\frac{eu_k}{kT} \ll \exp\frac{e\varphi}{kT} \tag{8.1}$$

и уравнение (6.11) упрощается:

$$\frac{d^2\varphi}{dx^2} = \frac{4\pi e}{\varepsilon} n_k \exp\frac{e\varphi}{kT}.$$

Умножая обе части этого уравнения на $d\varphi/dx$ и интегрируя по φ, получаем

$$\left(\frac{d\varphi}{dx}\right)^2 = \frac{8\pi n_k kT}{\varepsilon} \exp\frac{e\varphi}{kT} + C.$$

Постоянная интегрирования C определяется из условия, что на границе обеих областей

$$\varphi = u_k, \qquad \frac{d\varphi}{dx} = 0.$$

Поэтому

$$C = -\frac{8\pi n_k kT}{\varepsilon} \exp\frac{eu_k}{kT}.$$

Отсюда видно, что, вследствие условия (8.1), для области вблизи контакта постоянной C можно пренебречь по сравнению с первым слагаемым. Поэтому

$$\frac{d\varphi}{dx} = \pm\left(\frac{8\pi n_k kT}{\varepsilon}\right)^{1/2} \exp\frac{e\varphi}{2kT}.$$

Так как мы рассматриваем обогащенный слой в электронном полупроводнике, то $\varphi < 0$ и увеличивается по абсолютной величине с увеличением x, а следовательно, нашей задаче соответствует знак минус. Интегрируя это уравнение еще раз по x в пределах от 0 до x, находим распределение потенциала в виде

$$-e\varphi = 2kT \ln\left(1 + \frac{x}{a}\right), \qquad (8.2)$$

где a есть характеристическая длина:

$$a = \left(\frac{\varepsilon kT}{2\pi n_k e^2}\right)^{1/2}. \qquad (8.3)$$

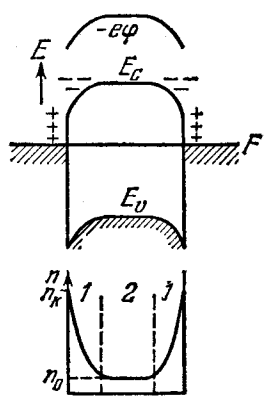


Рис. 6.10. Распределение электрического потенциала φ и концентрации электронов n в пластинке полупроводника с двумя металлическими контактами в отсутствие тока.

С точностью до множителя $2^{-1/2}$ это есть не что иное, как длина экранирования (7.2), в которой, однако, концентрация электронов в глубине образца n_0 заменена ее значением на контакте n_k. Таким образом, потенциал вблизи контакта изменяется по логарифмическому закону. Распределение концентрации электронов выражается соотношением

$$n = n_k \exp\frac{e\varphi}{kT} = n_k\left(\frac{a}{a+x}\right)^2. \qquad (8.4)$$

Вдали от контакта (область 2)

$$\varphi = u_k, \qquad n = n_k \exp\frac{eu_k}{kT} = n_0. \qquad (8.5)$$

Распределение потенциала и концентрации электронов в слое полупроводника между двумя одинаковыми металлическими электродами с обогащенными слоями схематически показано на рис. 6.10.

Таким образом, прилегающие к металлическим электродам слои полупроводника, толщина которых $\sim a$, могут «заливаться» носителями заряда. При этом концентрация носителей вблизи контактов, как показывает формула (8.4), не зависит от их концентрации в глубине полупроводника n_0, которая может быть как угодно мала (изолятор). Поэтому электропроводность слоистых (пленочных) структур металл — тонкий слой диэлектрика — металл (структура МДМ) может быть велика, даже если удельная электропроводность диэлектрика (в отсутствие контакта) ничтожно мала.

§ 9. Истощенный контактный слой

Рассмотрим теперь противоположный случай, когда приконтактный слой в полупроводнике обеднен основными носителями (запорный, или блокирующий, контакт). Мы ограничимся предельным случаем очень сильного обеднения, который особенно легко поддается расчету и в то же время часто встречается в практике.

Контакт металл—полупроводник. Положим, что к контакту приложено внешнее напряжение u, создающее обедненный слой. Если напряжение достаточно велико, то, следуя Шоттки, можно приближенно считать, что в некотором слое полупроводника толщиной d электронов нет вовсе («полностью истощенный слой»), так что объемный заряд обусловлен только заряженными донорами и акцепторами. В этом случае имеем $eu_k > 0$ (запорный слой), $\exp \frac{e(\varphi - u_k)}{kT} \ll 1$ и уравнение Пуассона (6.11) принимает простой вид:

$$\frac{d^2\varphi}{dx^2} = -\frac{4\pi e n_0}{\varepsilon} = \text{const}. \tag{9.1}$$

Интегрируя это уравнение два раза и принимая во внимание граничные условия (при d, подлежащем определению):

$$x = 0: \quad \varphi = 0; \quad x = d: \quad \varphi = u + u_k, \quad \frac{d\varphi}{dx} = 0, \tag{9.2}$$

мы получаем

$$-\frac{d\varphi}{dx} = \mathcal{E} = -\frac{4\pi e n_0}{\varepsilon}(d - x), \tag{9.3}$$

$$\varphi = u_k + u - \frac{2\pi e n_0}{\varepsilon}(d - x)^2. \tag{9.4}$$

Полагая в формуле (9.4) $\varphi = 0$, $x = 0$, находим толщину запорного слоя:

$$d = \sqrt{\frac{\varepsilon(u_k + u)}{2\pi e n_0}}. \tag{9.5}$$

Аналогично можно найти распределение электрического поля и потенциала в контакте двух полупроводников. При этом особый интерес представляет случай, когда один из полупроводников имеет проводимость p-типа, а другой — n-типа. Такие электронно-дырочные переходы, или, иначе, p—n-переходы, широко применяются в современной полупроводниковой электронике (см. гл. VII и VIII). Для получения истощенных слоев, очевидно, необходимо, чтобы к n-части перехода был приложен плюс источника, а к p-части — минус (рис. 6.11, *а*), или, при обратной полярности, $|u| < |u_k|$.

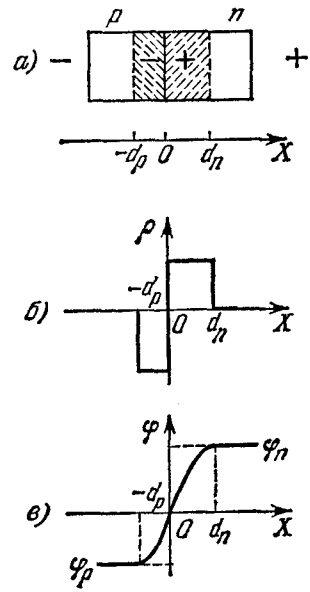


Рис. 6.11. Резкий p — n-переход при обратном смещении. Распределение объемного заряда ρ и потенциала φ.

Толщина запорных слоев d_p и d_n в обеих частях перехода и распределение потенциала в них зависят от закона распределения доноров и акцепторов. Мы рассмотрим два практически важных случая.

Резкий p—n-переход. Разность концентраций доноров N_d и акцепторов N_a постоянна в каждой области перехода и скачком меняется в плоскости контакта. Доноры и акцепторы будем считать полностью ионизованными. Тогда концентрации электронов и дырок вдали от контакта, где нет объемного заряда, равны:

$$n_0 \simeq N_d - N_a \qquad (x > d_n),$$
$$p_0 \simeq N_a - N_d \qquad (x < -d_p).$$

Поэтому для объемного заряда в приконтактных слоях мы имеем:

$$\begin{aligned} &p\text{-область:} && \rho = \begin{cases} -ep_0 & (-d_p < x < 0), \\ 0 & (x < -d_p); \end{cases} \\ &n\text{-область:} && \rho = \begin{cases} en_0 & (0 < x < d_n), \\ 0 & (x > d_n). \end{cases} \end{aligned} \tag{9.6}$$

Это распределение $\rho(x)$ изображено на рис. 6.11, *б*.

Будем отсчитывать, как и раньше, потенциал от его значения в плоскости $x = 0$. Тогда граничные условия задачи имеют вид:

$$\begin{aligned} &x = -d_p: && \varphi = u_p, && \frac{d\varphi}{dx} = 0; \\ &x = d_n: && \varphi = u_n, && \frac{d\varphi}{dx} = 0, \end{aligned} \tag{9.7}$$

где u_p и u_n — падения напряжения на слоях объемного заряда в p- и, соответственно, в n-области. Подставляя выражения (9.6) для ρ в уравнение Пуассона и интегрируя последнее при учете (9.7), получаем:

$$\begin{aligned} &p\text{-область:} \quad \varphi_p = u_p + \frac{2\pi e p_0}{\varepsilon}(x + d_p)^2; \\ &n\text{-область:} \quad \varphi_n = u_n - \frac{2\pi e n_0}{\varepsilon}(x - d_n)^2. \end{aligned} \tag{9.8}$$

Потенциал в слоях объемного заряда изменяется по параболическому закону, изображенному на рис. 6.11, *в*. При $x = 0$ оба выражения (9.8) должны давать одно и то же значение потенциала в плоскости контакта. Отсюда

$$u_n - u_p = \frac{2\pi e}{\varepsilon}(n_0 d_n^2 + p_0 d_p^2), \tag{9.9}$$

где

$$u_n - u_p = u + u_k$$

есть полное напряжение на переходе, складывающееся из контактной разности потенциалов u_k и напряжения источника u.

Так как электрическая индукция $\varepsilon \mathscr{E}$ должна быть везде непрерывна (а диэлектрическую проницаемость ε мы считаем одинаковой в p- и n-областях), то

$$\mathscr{E}_p(0) = -\left.\frac{d\varphi_p}{dx}\right|_{x=0} = \mathscr{E}_n(0) = -\left.\frac{d\varphi_n}{dx}\right|_{x=0}.$$

Это дает

$$n_0 d_n = p_0 d_p. \tag{9.10}$$

Соотношения (9.9) и (9.10) позволяют найти толщины слоев объемного заряда d_n и d_p.

Полагая в формулах (9.8) $x = 0$ и, соответственно, $\varphi_p = \varphi_n = 0$ и учитывая (9.10), находим, что падения напряжений в p- и n-областях относятся, как

$$\left|\frac{u_p}{u_n}\right| = \frac{n_0}{p_0}.$$

Поэтому, если, например, $p_0 \gg n_0$, то будем иметь $d_p \ll d_n$ и, кроме того, все приложенное напряжение будет сосредоточено в n-области. Если, напротив, $n_0 \gg p_0$, то все падение напряжения будет сосредоточено в p-области и при этом будет $d_p \gg d_n$.

При произвольных концентрациях n_0 и p_0 из формул (9.9) и (9.10) легко получить, что полная толщина слоя объемного заряда равна

$$d = d_n + d_p = \left(\frac{\varepsilon (u + u_k)}{2\pi e} \frac{p_0 + n_0}{p_0 n_0}\right)^{1/2}. \tag{9.11}$$

Если $n_0 \ll p_0$, то

$$d \simeq d_n = \left(\frac{\varepsilon (u+u_k)}{2\pi e n_0}\right)^{1/2}, \qquad (9.11')$$

что совпадает с формулой (9.5) для контакта металл—электронный полупроводник. При $p_0 \ll n_0$ в формулу (9.11′) вместо n_0 входит p_0.

Плавный p—n-переход. Если разность $(N_d - N_a)$ изменяется в пространстве непрерывно, то распределения поля и потенциала будут другими и будут зависеть от закона изменения $(N_d - N_a)$. Рассмотрим простейший случай, когда на протяжении слоя объемного заряда (толщина которого определяется формулой (9.15)) этот закон можно считать линейным:

$$N_d - N_a = ax,$$

где a — постоянная. Тогда распределение объемного заряда будет (рис. 6.12)

$$\rho = eax. \qquad (9.12)$$

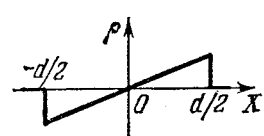


Рис. 6.12. Плавный p —n-переход при обратном смещении. Распределение объемного заряда.

При этом из симметрии задачи очевидно, что падения напряжения на каждом из слоев одинаковы и равны половине полного напряжения $(u + u_k)$. Тогда, подставляя (9.12) в уравнение Пуассона и учитывая граничные условия:

$$x = 0: \quad \varphi = 0; \qquad x = \pm {}^1/_2 d: \quad \frac{d\varphi}{dx} = 0, \qquad (9.13)$$

находим

$$\varphi = \frac{2\pi ae}{\varepsilon} ({}^1/_4 d^2 x - {}^1/_3 x^3). \qquad (9.14)$$

В этом случае потенциал изменяется по кубическому закону. Полагая в формуле (9.14) $x = d/2$ и, соответственно, $\varphi = (u + u_k)/2$, получаем для толщины двойного слоя объемного заряда

$$d = \left[\frac{3\varepsilon (u+u_k)}{\pi ea}\right]^{1/3}. \qquad (9.15)$$

Поступая аналогично, можно найти толщину слоя объемного заряда и для других законов распределения $(N_d - N_a)$.

Зарядная емкость. Толщина слоя объемного заряда возрастает с увеличением приложенного обратного напряжения, и при этом растет и полная величина заряда, сосредоточенного в слое. Отсюда следует, что контакт обладает определенной емкостью. Она получила название *зарядной емкости*.

Рассмотрим в качестве примера резкий p—n-переход с сильно легированной p-областью (или контакт металл—полупроводник n-типа). Тогда величина заряда в слое, рассчитанная на единицу поверхности, есть

$$q = en_0 d_n.$$

Подставляя сюда для d_n выражение (9.11′), находим для зарядной емкости

$$C = \frac{dq}{du} = \left(\frac{\varepsilon e n_0}{8\pi (u + u_k)}\right)^{1/2} = \frac{\varepsilon}{4\pi d_n}. \qquad (9.16)$$

Это выражение совпадает с формулой для емкости плоского конденсатора с толщиной зазора между обкладками, равной d_n.

Оценим порядок величины этой емкости. Рассмотрим германий ($\varepsilon = 16$) с концентрацией электронов $n_0 = 1 \cdot 10^{14}$ см$^{-3}$ и положим, что $u + u_k = 1$ В $= (1/300)$ ед. СГСЭ. Тогда из (9.11′) имеем $d_n \simeq 4 \cdot 10^{-4}$ см $= 4$ мкм и (9.16) дает $C \simeq 10^4/\pi$ см/см^2, или $\sim 10^3$ пФ/см^2.

Для плавного перехода с линейным распределением доноров и акцепторов имеем

$$q = ea \int_0^{d/2} x\, dx = \frac{1}{8} ead^2.$$

Подставляя сюда для d выражение (9.15) и дифференцируя по u, находим для емкости (опять на единицу поверхности)

$$C = \frac{\varepsilon}{4\pi}\left[\frac{\pi ea}{3\varepsilon (u + u_k)}\right]^{1/3} = \frac{\varepsilon}{4\pi d}. \qquad (9.17)$$

Из выражений (9.16) и (9.17) видно, что зарядная емкость зависит от приложенного обратного напряжения, и притом различно при различных законах распределения примесей. Для резкого перехода $C \sim u^{-1/2}$, для линейного закона распределения $C \sim u^{-1/3}$, а при другом законе распределения эта зависимость была бы иная. Поэтому, исследуя на опыте зависимость C от u, можно сделать выводы о структуре p—n-перехода, т. е. о законе распределения примесей в нем.

§ 10. Токи, ограниченные пространственным зарядом

Обратимся теперь к зависимости тока сквозь контакт от приложенного внешнего напряжения. Вольтамперные характеристики контактов оказываются различными в зависимости от того, имеем ли мы контакты с обогащенным слоем или с обедненным слоем.

Рассмотрим сначала случай обогащенного слоя.

Энергетическая диаграмма такого контакта изображена на рис. 6.13. В отсутствие внешнего напряжения потенциальная энергия электрона $-e\varphi$ в области объемного заряда вблизи металлического электрода увеличивается при удалении от электрода по логарифмическому закону (8.2). За областью объемного заряда она постоянна. Поэтому энергия дна зоны проводимости E_c, которая изменяется так же, как и $-e\varphi$, изображается кривой *1*. В области объемного заряда имеются ток диффузии и ток дрейфа, которые на-

правлены противоположно и в точности компенсируют друг друга. При наложении на второй электрод положительного напряжения потенциальная энергия электрона понижается, и тем больше, чем ближе расположена рассматриваемая плоскость к аноду. Поэтому для E_c мы получаем кривую *2*. В этом случае в некоторой плоскости x' потенциальная энергия имеет максимум, напряженность результирующего электрического поля равна нулю, а ток определяется только диффузией электронов. С увеличением внешнего напряжения плоскость x' приближается к катоду, а высота барьера понижается (кривая *3*). При этом всякий раз устанавливается такое распределение потенциала, что ток диффузии в плоскости x' как раз равняется току, протекающему через кристалл. Такие контакты, представляющие своеобразный резервуар электронов, в литературе называют еще *антизапорными* или *омическими* контактами.

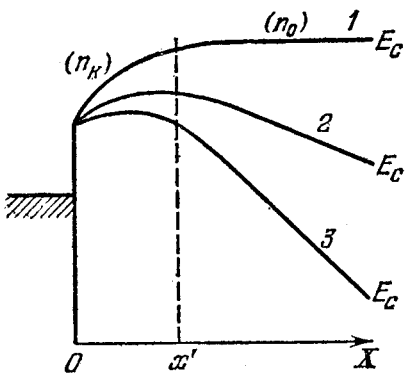


Рис. 6.13. Энергетическая диаграмма контакта с обогащенным слоем без тока (*1*) и при наличии тока (*2*, *3*).

Отметим, что диаграммы рис. 6.13 аналогичны кривым распределения потенциальной энергии электронов в вакуумном диоде при учете начальных скоростей термоэлектронов. Пользуясь терминологией вакуумной электроники, можно сказать, что в плоскости x' мы имеем «виртуальный» катод, имея под этим в виду такой катод, который способен эмиттировать любое количество электронов при условии, что на его поверхности напряженность электрического поля равна нулю.

При достаточно большом приложенном напряжении током диффузии можно пренебречь по сравнению с током дрейфа для любой плоскости внутри полупроводника. В этом случае максимум на кривых рис. 6.13 исчезает вовсе и распределение потенциала становится монотонным.

При положительном потенциале на втором электроде (аноде) электроны затягиваются в глубь полупроводника и полупроводник обогащается электронами. Это явление называется *инжекцией основных носителей*. При этом в полупроводнике возникает пространственный заряд и, в соответствии с уравнением Пуассона, появляется дополнительное электрическое поле, препятствующее вхождению электронов из контакта в полупроводник. Поэтому плотность протекающего тока определяется условием, что падение напряжения внутри полупроводника, обусловленное пространственным зарядом,

уравновешивается внешним напряжением. Мы имеем *токи, ограниченные пространственным зарядом* (ТОПЗ). Они аналогичны качественно токам в вакуумных диодах при напряжениях на аноде, меньших напряжения насыщения, хотя количественно между этими случаями имеются различия. Вследствие влияния объемного заряда такие контакты не подчиняются закону Ома, и плотность тока растет быстрее, чем приложенное напряжение.

Найдем теперь вольтамперную характеристику. Будем по-прежнему исходить из основных уравнений (6.5) и (6.6), считая, что доноры и акцепторы полностью ионизованы. Тогда, исключая из этих уравнений концентрацию электронов n, получаем для определения электрического поля $\mathscr{E}(x)$ нелинейное уравнение второго порядка

$$\frac{d^2\mathscr{E}}{dx^2} + \frac{e}{kT}\mathscr{E}\frac{d\mathscr{E}}{dx} - \frac{4\pi\sigma_0}{\varepsilon\mu}\frac{e}{kT}\mathscr{E} + \frac{4\pi j}{\varepsilon\mu}\frac{e}{kT} = 0. \qquad (10.1)$$

Здесь плотность тока $j = \text{const}$ входит как параметр, а $\sigma_0 = e\mu n_0$ есть электропроводность в глубине пластинки полупроводника в отсутствие инжекции.

Задача существенно упрощается, если можно пренебречь током диффузии (который приводит к первому слагаемому в уравнении (10.1)) по сравнению с током дрейфа. Тогда уравнение (10.1) переходит в уравнение первого порядка и его решение можно получить в аналитической форме.

Следуя Мотту и Герни [1], условие возможности пренебречь током диффузии можно оценить следующим образом. Интегрируя почленно уравнение (10.1) один раз, мы получим

$$\frac{d\mathscr{E}}{dx} - \frac{e}{2kT}\mathscr{E}^2 - \ldots = \text{const}.$$

Для грубой оценки порядка величин можно положить

$$\frac{d\mathscr{E}}{dx} \simeq \frac{\bar{\mathscr{E}}}{L}, \qquad \mathscr{E}^2 \simeq \bar{\mathscr{E}}^2,$$

где $\bar{\mathscr{E}} = u/L$ есть среднее значение поля, u — приложенное напряжение, а L — толщина рассматриваемого слоя полупроводника. Тогда условие возможности пренебрежения диффузией

$$\left|\frac{d\mathscr{E}}{dx}\right| \ll \left|\frac{e}{kT}\mathscr{E}^2\right|$$

принимает вид

$$u \simeq \bar{\mathscr{E}}L \gg \frac{kT}{e}.$$

При $T = 300$ К имеем $kT/e = 0{,}025$ В. Однако эта оценка иногда оказывается слишком грубой и для пренебрежения диффузией необходимы значительно большие напряжения.

Для исследования уравнения (10.1) удобно ввести безразмерные величины координаты x_1, поля $\mathcal{E}_1$, потенциала φ_1, а также максвелловское время релаксации τ_M, определяемые соотношениями

$$x_1 = \frac{en_0}{j\tau_M} x, \qquad \mathcal{E}_1 = \frac{\sigma_0}{j} \mathcal{E} = \frac{n_0}{n},$$
$$\varphi_1 = \frac{en_0\sigma_0}{j^2\tau_M} \varphi, \qquad \tau_M = \frac{\varepsilon}{4\pi\sigma_0}. \tag{10.2}$$

Тогда уравнение (10.1) (в пренебрежении диффузией) принимает вид

$$\mathcal{E}_1 \frac{d\mathcal{E}_1}{dx_1} - \mathcal{E}_1 + 1 = 0 \tag{10.3}$$

при граничном условии

$$x_1 = 0: \quad \mathcal{E}_1(0) = \frac{n_0}{n_k}. \tag{10.4}$$

Решение уравнения (10.3) есть

$$\mathcal{E}_1 + \ln(1 - \mathcal{E}_1) = x_1 + C, \tag{10.5}$$

где постоянная интегрирования C определяется из граничного условия

$$C = \mathcal{E}_1(0) + \ln[1 - \mathcal{E}_1(0)].$$

Так как, далее, для антизапорных контактов $\mathcal{E}_1(0) = n_0/n_k \ll 1$, то, разлагая логарифм в последнем соотношении в ряд, мы имеем

$$C = \mathcal{E}_1(0) + \left[-\mathcal{E}_1(0) - \frac{1}{2}\mathcal{E}_1^2(0) - \ldots\right] = -\frac{1}{2}\mathcal{E}_1^2(0) = -\frac{1}{2}\left(\frac{n_0}{n_k}\right)^2. \tag{10.6}$$

Падение напряжения φ_1 на толщине пластинки полупроводника x_1 равно

$$\varphi_1 = \int_0^{x_1} \mathcal{E}_1\, dx = \int_{\mathcal{E}_1(0)}^{\mathcal{E}_1} \mathcal{E}_1 \frac{dx_1}{d\mathcal{E}_1}\, d\mathcal{E}_1 = -\int_{\mathcal{E}_1(0)}^{\mathcal{E}_1} \frac{\mathcal{E}_1^2\, d\mathcal{E}_1}{1 - \mathcal{E}_1}.$$

В этом выражении можно приближенно положить нижний предел интегрирования $\mathcal{E}_1(0) = 0$, и тогда

$$\varphi_1 = \frac{1}{2}\mathcal{E}_1^2 + \mathcal{E}_1 + \ln(1 - \mathcal{E}_1) = \frac{1}{2}\mathcal{E}_1^2 + x_1 + C. \tag{10.7}$$

Так как безразмерные величины φ_1, $\mathcal{E}_1$ и x_1, согласно определениям (10.2), зависят от плотности тока j, то соотношение (10.7) представляет собой уравнение вольтамперной характеристики, заданное в параметрической форме. При этом для вычисления φ_1 на толщине пластинки x_1 при заданном токе j значение $\mathcal{E}_1(x_1)$ должно быть найдено из уравнения (10.5). Так как это уравнение трансцендентно, то в общем случае для этого требуются численные расчеты. Однако можно получить простые аналитические выражения, если рассматривать отдельные области токов и напряжений.

а. Слабые токи. Положим, что токи настолько слабы, что практически везде внутри полупроводника $|x_1| \gg 1$. Так как $|C| \ll 1$ и, согласно (10.2), $\mathcal{E}_1 \leqslant 1$, то уравнение (10.5) удовлетворяется только при отрицательных x_1. Это значит, что направление тока j противоположно положительному направлению оси X (от металла к полупроводнику), т. е. что при инжекции электронов в электронный полупроводник ток течет от полупроводника к металлу. Далее, из уравнения (10.5) следует, что при этом $\mathcal{E}_1 \simeq 1$. Или, в размерных величинах,

$$\mathcal{E} = \frac{j}{\sigma_0}$$

(область *1*). Напряжение на пластинке полупроводника толщиной L равно

$$u = \int_0^L \mathcal{E}\, dx = \frac{L}{\sigma_0} j. \tag{10.8}$$

При очень слабых токах (критерий указан ниже) выполняется закон Ома, а электропроводность равна ее значению σ_0 в отсутствие инжекции.

б. Сильные токи. Этот случай соответствует условию $\mathcal{E}_1 \ll 1$. Тогда, разлагая в (10.5) ln в ряд и удерживая два первых члена разложения, имеем

$$-\frac{1}{2}\mathcal{E}_1^2 = x_1 + C,$$

или

$$\mathcal{E}_1 = \sqrt{2}\,(|x_1| + |C|)^{1/2}. \tag{10.9}$$

Здесь, в свою очередь, удобно различать два случая. Если токи не очень сильны, то $|x_1| \gg C$ (область *2*) и

$$\mathcal{E}_1 = \sqrt{2|x_1|}.$$

Возвращаясь с помощью соотношений (10.2) к размерным величинам, получаем

$$\mathcal{E} = \left(\frac{2j}{\tau_M \sigma_0 \mu}\right)^{1/2} x^{1/2}.$$

Для напряжения имеем

$$u = \int_0^L \mathcal{E}\, dx = \left(\frac{8jL^3}{9\tau_M \sigma_0 \mu}\right)^{1/2},$$

или

$$j = \frac{9}{8}\tau_M \sigma_0 \mu \frac{u^2}{L^3} \tag{10.10}$$

(закон Мотта). В этом случае ток пропорционален квадрату напряжения (в отличие от вакуумных диодов, где $j \sim u^{3/2}$).

Порядок величины напряжений, при которых закон Ома переходит в квадратичный закон Мотта, можно оценить, приравнивая величины токов, получаемые из соотношений (10.8) и (10.10).) Это приводит к простому условию:

$$t_{\text{пр}} \simeq \tau_M, \tag{10.11}$$

где $t_{\text{пр}} = L^2/\mu u$ есть время пролета электрона между электродами. Отклонения от закона Ома (10.8) становятся заметными при таких напряжениях, при которых пролетное время делается меньше максвелловского времени релаксации.

Если токи становятся настолько сильными, что $|x_1| \ll C$, то (10.9) дает

$$\mathcal{E}_1 \simeq \sqrt{2|C|}.$$

Для размерных величин при этом получается

$$\mathcal{E} = \frac{j}{\sigma_0}\frac{n_0}{n_k}, \qquad j = e\mu n_k \frac{u}{L}. \tag{10.12}$$

При очень больших напряжениях квадратичный закон для тока опять переходит в закон Ома (область *3*), однако теперь с другой, гораздо большей электропроводностью $e\mu n_k$. Физически это обозначает, что вся пластинка заполняется инжектированными электронами с практически постоянной по объему концентрацией n_k.

Условие перехода из области *2* в область *3* можно оценить, приравнивая выражения для j из (10.10) и (10.12). Это дает

$$t_{\text{пр}} \simeq \tau'_M,$$

где $\tau'_M = \varepsilon/4\pi e\mu n_k$ — максвелловское время релаксации, но теперь соответствующее концентрации n_k. Омический режим восстанавливается, когда $t_{\text{пр}}$ уменьшается до τ'_M.

Найденная зависимость тока от напряжения схематически показана на рис. 6.14 линией *abcd* в двойном логарифмическом масштабе. Области *1* соответствует прямая *ab*, области *2* — прямая *bc* и области *3* — прямая *cd*. В действительности, конечно, переход между отдельными областями имеет плавный характер.

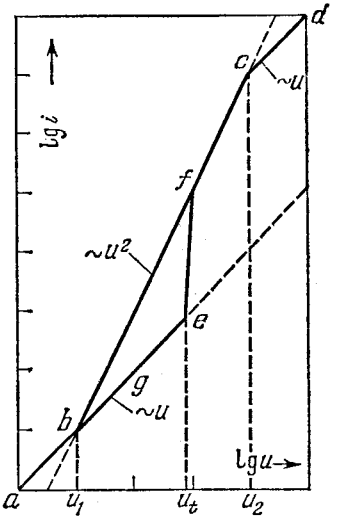


Рис. 6.14. Зависимость тока от напряжения (произвольные единицы) для ТОПЗ. *abfcd* — без ловушек; *abefcd* — при равномерном распределении глубоких ловушек.

Выше мы везде предполагали, что в полупроводнике (или диэлектрике) не имеется ловушек для электронов и что поэтому весь

объемный заряд образован избыточными подвижными электронами в зоне. При наличии ловушек задача становится гораздо сложнее. Качественно явления изменяются следующим образом. Положим, что ловушки создают глубокие уровни энергии (так что обратным тепловым выбросом электронов можно пренебречь) и что ловушки равномерно распределены по объему. Концентрацию ловушек будем считать намного большей концентрации равновесных электронов n_0. Тогда с увеличением напряжения инжектированные электроны будут сначала практически все захватываться на ловушки, а концентрация электронов в зоне увеличиваться не будет. Это приведет к тому, что первый омический участок вольтамперной характеристики не закончится при напряжениях $\simeq u_1$ (рис. 6.14), а будет продолжен в область больших напряжений (кривая *abe*). При некотором напряжении u_t все ловушки окажутся заполненными электронами, и тогда концентрация электронов в зоне начнет сильно увеличиваться с напряжением. Это приводит к очень крутому участку характеристики *ef* (рис. 6.14). И, наконец, когда величина тока достигнет значения, соответствующего квадратичному закону, сила тока будет изменяться по кривой *fcd*, т. е. так же, как и в отсутствие ловушек. Разумеется, форма реальных характеристик зависит от свойств ловушек и их концентрации и поэтому может в деталях отличаться от приведенных схематических зависимостей.

Отметим еще, что время установления объемного заряда определяется максвелловским временем релаксации τ_M. С другой стороны, установление электронного равновесия между ловушками и зоной характеризуется другим временем — «временем захвата» τ. Если $\tau_M < \tau$ (что часто имеет место), то после наложения внешнего напряжения в течение времени $t < \tau$ характеристика имеет такой же вид, как и в отсутствие ловушек. И лишь в последующее время $t > \tau$ устанавливается стационарная характеристика, определяемая ловушками. При этом значения токов в импульсном и стационарном режимах могут отличаться на много порядков. Исследуя кинетику установления ТОПЗ, а также характерные особенности стационарных характеристик, оказывается возможным оценить параметры ловушек (сечения захвата для электронов, концентрацию и др.).

§ 11. Выпрямление в контакте металл — полупроводник

Обратимся теперь к блокирующим, или запорным, контактам. Вольтамперная характеристика таких контактов может оказаться сильно нелинейной, и поэтому контакты соответственно подобранных пар металла и полупроводника (или двух полупроводников) широко используются как выпрямители переменного тока.

Влияние потенциального барьера на прохождение тока существенно зависит от соотношения между шириной барьера $L_э$ и длиной волны электронов λ. Если барьер достаточно тонкий, так что $L_э < \lambda$, то электроны любой энергии могут проходить сквозь барьер вследствие квантовомеханического туннельного эффекта. Напротив, если $L_э > \lambda$, то возможность преодоления потенциального барьера

определяется классическим условием: энергия электрона должна быть больше высоты барьера, т. е. электроны могут проходить только над барьером. Так, например, при концентрации электронов $n_0 \sim 10^{15}$ см$^{-3}$ при комнатной температуре, как мы видели в § 7, $L_э \sim 10^{-5}$ см. Для электронов с тепловыми скоростями при той же

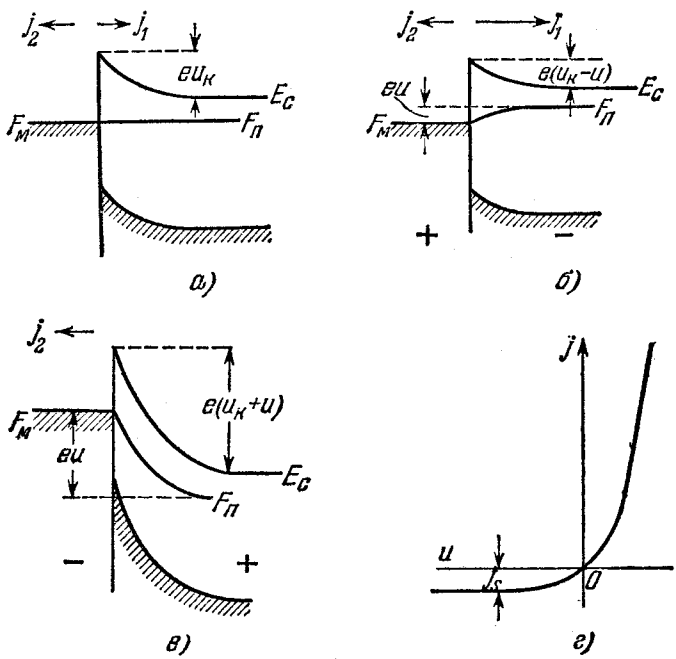


Рис. 6.15. Выпрямление в контакте металл — полупроводник: *а*) внешнее напряжение $u = 0$; *б*) проходное направление тока; *в*) запорное направление тока; *г*) вольтамперная характеристика (схематически).

температуре $\lambda \sim 10^{-6}$ см, и мы имеем классический случай прохождения над барьером. Однако $L_э \sim 1/\sqrt{n_0}$, и поэтому в сильно легированных полупроводниках с $n_0 \sim 10^{18} \div 10^{19}$ см$^{-3}$ (или больше) $L_э < \lambda$ и существенную роль будет играть туннельное просачивание электронов.

Сейчас мы рассмотрим случай не очень сильно легированных полупроводников (случай туннельного эффекта будет рассмотрен в § VIII.3). Мы будем также предполагать, что участием дырок в образовании тока можно пренебречь (монополярная проводимость).

Качественное объяснение выпрямления тока для этого случая дано на рис. 6.15. В отсутствие внешнего напряжения (рис. 6.15, *а*) энергия электронов в глубине полупроводника понижается отно-

сительно металла на $-eu_k$, где u_k — контактная разность, а следовательно, для электронов, движущихся из полупроводника в металл, существует энергетический барьер eu_k. В равновесии ток j_1 электронов, переходящих из полупроводника в металл, равен току j_2 электронов, переходящих из металла в полупроводник, и поэтому полный ток $j = j_1 - j_2 = 0$. В соответствии с этим уровень Ферми везде одинаков.

Рис. 6.15, *б* соответствует приложенному внешнему напряжению такого знака, что изменение энергии электронов в полупроводнике — $eu > 0$. Для этого электронный полупроводник должен быть соединен с минусом источника; для полупроводника *p*-типа знак напряжения должен быть обратным. В этом случае потенциальный барьер в полупроводнике уменьшается на величину eu и ток j_1 увеличивается, в то время как барьер со стороны металла, а следовательно, и ток j_2 не изменяются. Поэтому возникает ток $j = j_1 - j_2$, идущий от металла к полупроводнику, который быстро увеличивается при увеличении u («проходное» направление тока, возрастающая ветвь характеристики на рис. 6.15, *г*).

При напряжении противоположного знака (рис. 6.15, *в*) потенциальный барьер в полупроводнике, напротив, увеличивается и ток j_1 уменьшается. При eu порядка нескольких kT электроны из полупроводника практически вовсе не могут преодолеть потенциальный барьер, и поэтому обратный ток достигает некоторого значения $j_s = j_2$, которое не зависит от внешнего напряжения (ток насыщения). Для этого направления тока сопротивление контакта велико («запорное» направление).

Разумеется, вольтамперная характеристика, показанная на рис. 6.15, *г*, имеет только схематический характер. В частности, при достаточно большом обратном напряжении возникают явления «электрического пробоя» контакта и обратный ток резко возрастает.

При расчете вольтамперной характеристики необходимо учитывать соотношение между шириной запорного слоя $L_э$ и длиной свободного пробега l. Если $l \gg L_э$, то электроны внутри запорного слоя практически не испытывают соударений и поэтому токи j_1 и j_2 можно рассматривать как токи термоэлектронной эмиссии (диодная теория выпрямления Бете). Отличие от обычной термоэлектронной эмиссии (§ 4) заключается в том, что здесь эмиссия происходит не в вакуум, а из полупроводника в металл (и обратно). При этом, вследствие меньшей высоты потенциального барьера в контакте (равной разности работ выхода контактирующих тел), токи термоэмиссии могут быть очень сильны уже при комнатных (и даже более низких) температурах.

В другом предельном случае, когда электроны испытывают многочисленные соударения внутри запорного слоя, ток через контакт нужно рассчитывать с учетом диффузионного и дрейфового движения электронов (диффузионная теория выпрямления; основы

ее были заложены в работах Б. И. Давыдова, С. И. Пекара, Н. Мотта, В. Шоттки). Условие применимости диффузионной теории выражается неравенством (2.7).

В зависимости от типа полупроводника применимой может быть и та, и другая теория. Так, например, если полупроводником является германий с $n_0 \sim 10^{15}$ см$^{-3}$ при комнатной температуре, то, как мы видели, $L_э \sim 10^{-5}$ см (а при меньших n_0 она еще больше). С другой стороны, l можно оценить по соотношению $l \simeq 5 \cdot 10^{-9}\ \mu$, где μ — подвижность в см^2/В·с, а l — в см (§ II.8). Для чистого (некомпенсированного) германия μ может достигать $3{,}9 \cdot 10^3$ см^2/В·с и поэтому $l \sim 10^{-5}$ см. Таким образом, в этом случае число соударений в запорном слое будет невелико и удовлетворительным приближением будет диодная теория. Напротив, для таких полупроводников, как Cu_2O и Se, у которых μ гораздо меньше (а следовательно, и l меньше) и концентрация носителей заряда тоже меньше ($L_э$ больше), применима диффузионная теория.

Диодная теория. Ток j_1, создаваемый электронами из полупроводника, можно непосредственно найти, учитывая, что в невырожденном полупроводнике скорости электронов распределены по закону Максвелла и что преодолеть барьер могут только те электроны, энергия которых удовлетворяет условию

$$^1/_2 m v_x^2 \geqslant e(u_k + u),$$

где v_x — нормальная к плоскости контакта составляющая тепловой скорости. Однако мы воспользуемся уже полученным выражением (4.6) для плотности тока термоэлектронной эмиссии. При этом под работой выхода в данном случае нужно понимать разность между вершиной барьера и уровнем Ферми в глубине полупроводника, которая равна (ср. рис. 6.15)

$$e(u_k + u) + E_c - F_n.$$

Тогда

$$j_1 = \frac{4\pi e m k^2}{(2\pi\hbar)^3} T^2 \exp\left[-\frac{e(u_k+u)+E_c-F}{kT}\right]. \qquad (11.1)$$

Это выражение можно представить в более простом виде. Учитывая, что концентрация электронов n_0 в глубине полупроводника есть (ср. формулы (V.5.1) и (V. 4.9))

$$n_0 = 2\left(\frac{2\pi m k T}{(2\pi\hbar)^2}\right)^{3/2} \exp\frac{F - E_c}{kT},$$

и пользуясь выражением (4.2) для средней тепловой скорости электронов v_T (с заменой m_0 на m), (11.1) можно написать

$$j_1 = \frac{1}{4} e n_0 v_T \exp\left(-\frac{e(u_k+u)}{kT}\right). \qquad (11.1а)$$

Величина тока j_2 получается непосредственно из этого выражения при $u = 0$, так как в отсутствие внешнего напряжения величины токов j_1 и j_2 одинаковы:

$$j_2 = \frac{1}{4} en_0 v_T \exp\left(-\frac{eu_k}{kT}\right). \tag{11.2}$$

Поэтому для полной плотности тока получается

$$j = j_1 - j_2 = j_s[\exp(-\alpha u) - 1], \tag{11.3}$$

где введены сокращенные обозначения

$$j_s = \frac{1}{4} en_0 v_T \exp(-\alpha u_k), \quad \alpha = \frac{e}{kT}. \tag{11.4}$$

Формула (11.3) показывает, что при отрицательном потенциале полупроводника относительно металла ($u < 0$) ток быстро увеличивается при возрастании напряжения. Уже при $e \mid u \mid$, равном нескольким kT, единицей можно пренебречь по сравнению с первым членом и закон нарастания тока становится экспоненциальным.

При обратных напряжениях ($u > 0$) первый (экспоненциальный) член быстро уменьшается с увеличением напряжения. При $eu \gtrsim kT$ он становится пренебрежимо малым по сравнению с единицей и ток достигает насыщения. Плотность тока насыщения равна j_s.

В заключение подчеркнем, что во всех предыдущих рассуждениях u обозначало напряжение, падающее на запорном слое. В реальном выпрямительном диоде всегда имеется еще некоторое сопротивление r самого кристалла полупроводника, включенное последовательно с запорным слоем. Поэтому для получения зависимости тока от полного напряжения на диоде V в предыдущих формулах везде нужно заменить u на $(V - ir)$, где i — сила тока через диод. Это приведет к горизонтальному смещению всех точек характеристики на переменный отрезок ir, отчего прямая ветвь характеристики окажется более пологой.

§ 12. Диффузионная теория

При строгом решении задачи в диффузионной теории мы должны исходить из системы уравнений (6.5) и (6.6). Однако ниже мы увидим, что вольтамперная характеристика контакта слабо зависит от вида функции $\varphi(x)$. Поэтому мы выберем приближенный, но гораздо более простой путь [M7] и рассмотрим только одно уравнение (6.5), которое запишем в виде

$$\frac{dn}{dx} - \alpha \frac{d\varphi}{dx} n - \frac{j}{\mu kT} = 0. \tag{12.1}$$

Будем считать $\varphi(x)$ заданной и посмотрим, какие выводы можно сделать без детального знания вида этой функции.

Поместим начало оси X в плоскости контакта и условимся отсчитывать, как и раньше, потенциал от его значения при $x = 0$. Тогда граничное условие будет

$$x = 0: \quad \varphi = 0, \qquad n = n_k. \tag{12.2}$$

Кроме этого, для любой плоскости $x_1 = \text{const}$, лежащей за пределами слоя объемного заряда,

$$x = x_1: \quad \varphi = u_k + u, \qquad \frac{d\varphi}{dx} = 0, \qquad n = n_0. \tag{12.3}$$

Уравнение (12.1) — 1-го порядка относительно $n(x)$. Его решение, удовлетворяющее граничному условию (12.2), есть

$$n(x) = e^{\alpha\varphi(x)} \left\{ n_k + \frac{j}{\mu kT} \int_0^x e^{-\alpha\varphi(y)}\, dy \right\}, \tag{12.4}$$

в чем легко убедиться непосредственной подстановкой. При этом, в соответствии со сказанным в § 6, мы будем считать n_k не зависящим от внешнего напряжения и равным его равновесному значению для невырожденного полупроводника:

$$n_k = n_0 e^{-\alpha u_k}. \tag{12.5}$$

Учет зависимости u_k от приложенного напряжения мы рассмотрим дополнительно в § 13.

Применим теперь решение (12.4) к плоскости x_1 и учтем условия (12.3) и (12.5). Это дает

$$n_0 = e^{\alpha(u_k+u)} n_0 e^{-\alpha u_k} + \frac{j}{\mu kT} e^{\alpha(u_k+u)} \int_0^{x_1} e^{-\alpha\varphi(y)}\, dy. \tag{12.6}$$

Разрешая это уравнение относительно j, мы получаем выражение для вольтамперной характеристики в виде

$$j = j_s [\exp(-\alpha u) - 1], \tag{12.7}$$

где введено обозначение

$$j_s = \frac{\mu kT \exp(-\alpha u_k)}{\int_0^{x_1} \exp(-\alpha\varphi(y))\, dy}. \tag{12.8}$$

Значение входящего сюда интеграла можно приближенно найти следующим образом. Так как для электронов в полупроводнике существует потенциальный барьер, то $\varphi(y)$ положительно. Вследствие этого подынтегральная функция быстро убывает с увеличением y и величина интеграла определяется только областью y возле плоскости $y = 0$. Поэтому можно положить

$$\varphi(y) = \varphi(0) + \left.\frac{d\varphi}{dy}\right|_{y=0} y + \ldots = -\mathscr{E}(0) \cdot y,$$

где $\mathscr{E}(0)$ — напряженность электрического поля в полупроводнике у контактной плоскости. При этом $\mathscr{E}(0)$ направлено антипараллельно оси y, т. е. $\mathscr{E}(0) < 0$. Поэтому

$$\int_0^{x_1} \exp[-\alpha\varphi(y)\,dy] \simeq \int_0^{\infty} \exp[-\alpha|\mathscr{E}(0)|y]\,dy = \frac{1}{\alpha|\mathscr{E}(0)|}.$$

Тогда для тока насыщения окончательно получается

$$j_s = en_0\mu|\mathscr{E}(0)|\exp(-\alpha u_k). \qquad (12.9)$$

Здесь $\mathscr{E}(0)$, строго говоря, зависит от приложенного внешнего напряжения u. Однако этой зависимостью можно пренебречь, принимая во внимание экспоненциальный множитель в формуле (12.7), и приближенно понимать под $\mathscr{E}(0)$ ее значение при $u = 0$.

Выражение (12.7) для вольтамперной характеристики в диффузионной теории имеет такой же вид, как и формула (11.3) в диодной теории. Однако различие заключается в токе насыщения, который в обеих теориях зависит от разных величин. При этом величина тока насыщения в диффузионной теории получается гораздо меньшей.

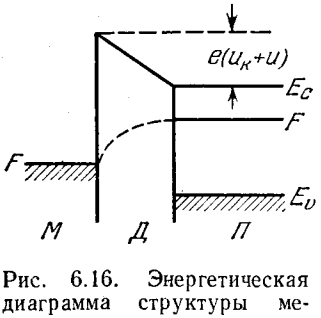


Рис. 6.16. Энергетическая диаграмма структуры металл — диэлектрик — полупроводник.

В настоящее время широко применяют также диоды, имеющие структуру металл — диэлектрик — полупроводник (МДП), в которых между металлом и полупроводником находится тонкий слой плохо проводящего вещества (искусственный, или «химический», запорный слой). Такие слои диэлектрика могут быть созданы конденсацией в вакууме, окислением или другими способами. Если толщина слоя диэлектрика d много больше длины экранирования $L_э$ в полупроводнике, то все контактное поле будет сосредоточено практически в диэлектрике (рис. 6.16). При этом во многих случаях можно пренебречь влиянием объемного заряда внутри диэлектрика и считать электрическое поле в нем однородным, отчего расчет вольтамперной характеристики сильно упрощается. Так, в уравнении (12.1) коэффициент $d\varphi/dx = -\mathscr{E}$ не будет зависеть от x и, рассуждая так же, как и выше, мы получим вместо формул (12.7) и (12.8) соотношение

$$j = en_0\mu|\mathscr{E}|\exp(-\alpha u_k)\frac{\exp(-\alpha u)-1}{1-\exp[-\alpha(u_k+u)]}. \qquad (12.10)$$

При всех напряжениях u, при которых еще существует потенциальный барьер $\gtrsim kT$, мы можем считать, что $\exp[-\alpha(u_k+u)] \ll 1$,

и поэтому предыдущая формула принимает вид

$$j = en_0\mu\,|\,\mathcal{E}\,|\exp(-\alpha u_k)\,[\exp(-\alpha u) - 1]. \qquad (12.11)$$

Мы получили ту же формулу, что и раньше, с тем отличием, что здесь $|\,\mathcal{E}\,| = (u_k + u)/d$ обозначает напряженность поля внутри диэлектрического слоя (а n_0, по-прежнему, относится к полупроводнику).

В заключение подчеркнем, что во всех предыдущих формулах внешнее напряжение u обозначало потенциал в глубине полупроводника относительно металла. Поэтому для электронного полупроводника проходному направлению тока соответствует $u < 0$. Если, как это часто встречается в литературе, считать проходное напряжение положительным, то формулы (11.3) и (12.7) будут иметь вид

$$j = j_s(\exp \alpha u - 1), \qquad (12.12)$$

где j_s по-прежнему выражается формулами (11.4) или, соответственно, (12.9).

§ 13. Сравнение с экспериментом

Одностороннюю проводимость контактов металл — полупроводник используют для устройства полупроводниковых выпрямителей переменного тока. Для выпрямления технических токов низкой частоты широко применяют селеновые выпрямители, в которых запирающий слой образуется у границы слоя Se и одного из металлических электродов («вентильного» электрода). Последний состоит обычно из сплава различных металлов (например, Bi, Cd и Sn). В меднозакисных выпрямителях запирающий слой возникает на границе между медной пластиной и слоем закиси меди Cu_2O, получающейся при окислении меди в атмосфере кислорода.

Для выпрямления токов высокой частоты применяют германиевые и кремниевые «точечные» СВЧ детекторы. Они содержат монокристалл полупроводника (германия n-типа или кремния p-типа), базовый (невыпрямляющий) металлический электрод большой площади и прижимной или приваренный металлический электрод (проволока) малого диаметра (микроны).

Контакты металл — полупроводник различных других конфигураций в настоящее время широко используют для создания быстродействующих нелинейных элементов, которые часто обозначают как «диоды Шоттки».

Опыт показывает, что направление выпрямленного тока хорошо согласуется с изложенной выше теорией (§§ 11, 12).

Далее, мы видели, что зависимость прямого тока от напряжения должна выражаться универсальной формулой (12.12), в которой постоянная $\alpha = e/kT$ не зависит от рода полупроводника и металла. Опыт дает, что эта экспоненциальная зависимость хорошо выпол-

няется. Однако значения α иногда получаются меньше теоретического (в 2 ÷ 3 раза) и различными для разных контактов. Причина этого кажущегося расхождения заключается в том, что геометрия реальных выпрямляющих контактов часто отличается от идеальной плоской структуры. При этом возникают краевые эффекты, приводящие к существенному увеличению электрического поля у периметра электродов. Кроме того, в реальных контактах могут быть еще значительные утечки вдоль поверхности полупроводника.

Эти обстоятельства можно практически исключить, выбирая рациональную конструкцию диодов [5]. Тогда прямая ветвь вольтамперной характеристики в широкой области изменения токов хорошо описывается формулой (12.12), а значение α с точностью до нескольких процентов совпадает с теоретической величиной (39 1/В при 300 К).

Обратная ветвь вольтамперной характеристики хуже согласуется с простой теорией. А именно, ток насыщения j_s не остается точно постоянным, а медленно увеличивается с увеличением обратного напряжения. Основная причина этого несоответствия состоит в том, что при расчетах в §§ 11, 12 мы считали граничную концентрацию у контакта n_k не зависящей от приложенного напряжения. Это допущение пригодно для прямого напряжения, так как в этом случае напряжение на контакте и напряженность поля в слое объемного заряда малы. При обратных же напряжениях напряженность поля в контакте становится большой, что приводит к понижению потенциального барьера для электронов и увеличению граничной концентрации электронов n_k.

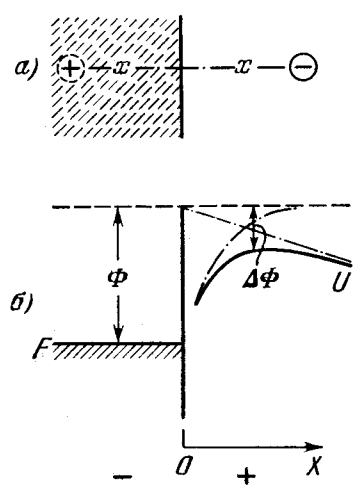


Рис. 6.17. Понижение потенциального барьера в контакте металл — полупроводник при обратном смещении (эффект Шоттки).

Рассматриваемый эффект аналогичен зависимости тока насыщения в вакуумных диодах от напряжения. Следуя Шоттки (но применительно к контактам металл—полупроводник), его можно приближенно учесть следующим образом. Электрон, переходящий из металла в полупроводник, испытывает силу притяжения к металлу от индуцированного им самим заряда на металле. Эту силу можно найти по методу зеркального изображения (рис. 6.17, *а*), а возникающее вследствие этого уменьшение потенциальной энергии электрона равно $-e^2/4\varepsilon x$. Здесь x — расстояние электрона до металла, а ε — диэлектрическая проницаемость полупроводника. Если, далее, $\mathscr{E}$ есть напряженность электрического поля внутри полупроводника (которую сначала мы будем считать постоянной), то вызываемое этим полем изменение потенциальной энергии есть $-e\mathscr{E}x$. Оно, как и в случае силы зеркального

изображения, отрицательно, так как при обратном смещении («+» на электронном полупроводнике) энергия уменьшается с увеличением x. Поэтому полное понижение потенциальной энергии электрона равно

$$U = -\frac{e^2}{4\varepsilon x} - e\mathscr{E}x.$$

Зависимость U от x показана на рис. 6.17, *б* (сплошная кривая). U имеет максимум при $x_m = \frac{1}{2}(e/\varepsilon\mathscr{E})^{1/2}$.

Подставляя это в предыдущее выражение, находим величину понижения потенциального барьера:

$$\Delta\Phi = -U_m = \left(\frac{e^3\mathscr{E}}{\varepsilon}\right)^{1/2}. \tag{13.1}$$

Выше мы считали, что электрическое поле $\mathscr{E}$ не зависит от x. В действительности же $\mathscr{E}$ изменяется в обедненном слое объемного заряда в соответствии с формулами (9.3) и (9.5). Однако расстояние x_m, на котором расположен максимум потенциального барьера, обычно намного меньше толщины слоя объемного заряда. Поэтому за $\mathscr{E}$ в формуле (13.1) приближенно можно принять максимальную напряженность поля, которая, в отсутствие рассматриваемого эффекта, существует в плоскости контакта $x = 0$. Тогда из формул (9.3) и (9.5) получаем

$$|\mathscr{E}| = |\mathscr{E}(0)| = \left[\frac{8\pi e n_0 (u_k + u)}{\varepsilon}\right]^{1/2}. \tag{13.2}$$

Таким образом, в результате совместного действия электрического поля в слое объемного заряда и силы зеркального изображения высота потенциального барьера со стороны металла уменьшается на $\Delta\Phi$. На такую же величину уменьшается высота потенциального барьера и со стороны полупроводника. Поэтому в формулах (11.4) и (12.9) мы должны заменить u_k на $\left(u_k - \frac{\Delta\Phi}{e}\right)$. Это дает

$$j_s = j_{s0} \exp\left(\frac{\Delta\Phi}{kT}\right), \tag{13.3}$$

где j_{s0} — ток насыщения без учета понижения барьера. Так как $\Delta\Phi \sim \mathscr{E}(0)^{1/2}$, а $\mathscr{E}(0) \sim (u_k + u_e)^{1/2}$, то отсюда следует, что $\ln j_s$ должен быть пропорционален $(u_k + u)^{1/4}$. Зависимость, близкая к этой, действительно наблюдается на опыте в диодах Шоттки при устранении краевых эффектов и утечек по поверхности.

Остановимся еще на зависимости сопротивления контактов от температуры. Удобно измерять дифференциальное сопротивление du/dj при внешнем напряжении $u = 0$ («нулевое сопротивление»). Согласно формуле (12.12) нулевое сопротивление r_0 на единицу площади контакта определяется соотношением

$$\frac{1}{r_0} = \left(\frac{dj}{du}\right)_{u=0} = j_s\alpha.$$

Так как и в диодной, и в диффузионной теориях $j_s \sim \exp(-\alpha u_k)$, а все остальные величины зависят от температуры гораздо слабее, то мы должны ожидать зависимость от температуры в виде

$$r_0 \sim \exp\frac{eu_k}{kT}. \tag{13.4}$$

Опыт показывает, что во многих случаях зависимость $\lg r_0$ от $1/T$ действительно выражается прямой линией. При этом r_0, в согласии с формулой (13.4), сильно увеличивается при понижении температуры.

Из наклона этих прямых можно найти контактную разность потенциалов u_k и сопоставить ее с работами выхода полупроводника и металла. Подробные опыты такого рода проводились, например, с селеновыми выпрямителями, в которых контактный электрод изготавливался из различных металлов. Так как селен обладает дырочной проводимостью, то для образования в нем потенциального барьера работа выхода из металла $\Phi_м$ должна быть меньше Φ_{Se}. При этом высота потенциального барьера для дырок будет $eu_k = \Phi_{Se} - \Phi_м$. Согласно формуле (13.4) можно ожидать, что с увеличением $\Phi_м$ величина r_0 будет сильно уменьшаться. Это действительно и наблюдается на опыте. Однако для германиевых и кремниевых выпрямителей такая зависимость между r_0 и $\Phi_м$ не соблюдается.

Причина этого отклонения от теории заключается в том, что в германии и кремнии большую роль играют так называемые поверхностные электронные состояния (см. гл. X). В результате их существования поверхность полупроводника оказывается заряженной, что вызывает искривление энергетических зон независимо от контактного поля. Поэтому и корреляция между работами выхода и высотой потенциального барьера в таких контактах уже не имеет места.

Глава VII

НЕРАВНОВЕСНЫЕ ЭЛЕКТРОНЫ И ДЫРКИ

§ 1. Неравновесные носители заряда

В полупроводниках, в отличие от металлов, под влиянием внешних воздействий (освещения, электрического тока в неоднородных структурах и других) концентрации электронов и дырок могут изменяться на много порядков. Это приводит к ряду специфических явлений, которые лежат в основе действия многих полупроводниковых приборов.

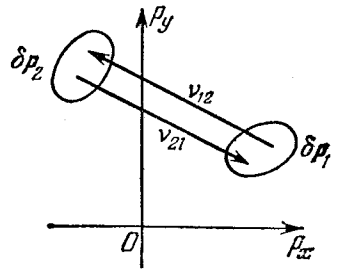


Рис. 7.1. Прямые и обратные электронные переходы.

Чтобы пояснить, почему изменяются концентрации электронов и дырок, рассмотрим группу состояний электронов, которые изображаются точками в **p**-пространстве, заключенными внутри малого объема $\delta\mathbf{p}_1$ (рис. 7.1). Часть из этих состояний фактически заполнена электронами, которые под влиянием теплового движения (сюда включается и тепловое излучение, находящееся в равновесии с кристаллом) могут переходить в другие состояния, например, принадлежащие объему $\delta\mathbf{p}_2$. При этом энергия электронов может изменяться. Обозначим через ν_{12} темп таких переходов, т. е. число переходов в единицу времени, рассчитанное на единицу объема кристалла, а темп обратных переходов из состояния *2* в состояние *1*— через ν_{21}. Тогда при термодинамическом равновесии устанавливается такое распределение электронов по квантовым состояниям, что

$$\nu_{12} = \nu_{21}. \tag{1.1}$$

Суммарные концентрации электронов и дырок равны при этом их равновесным значениям n_0 и p_0. Соотношение (1.1) выражает *принцип детального равновесия*, который играет большую роль при статистических расчетах и более подробно будет рассмотрен в гл. XIII.

Существенно подчеркнуть, что взаимное уравновешивание прямых и обратных переходов справедливо не только для всей совокупности состояний, но и для отдельных их физически малых групп. Так, например, рассматривая тепловые возбуждения электронов

из валентной зоны в зону проводимости, можно непосредственно заключить, что темп этих переходов (тепловой генерации электронно-дырочных пар) равен темпу обратных выпадений электронов из зоны проводимости в валентную зону (или рекомбинации электронно-дырочных пар), что вытекает просто из условия постоянства концентраций электронов и дырок в зонах. Однако согласно принципу детального равновесия можно утверждать гораздо больше, а именно что это равенство справедливо и для отдельных групп энергетических уровней в зонах (рис. 7.2).

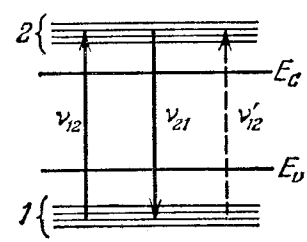


Рис. 7.2. К иллюстрации принципа детального равновесия. Тепловая генерация электронно-дырочных пар ($1 \to 2$) и обратный процесс рекомбинации электронов и дырок ($2 \to 1$).

При наличии внешних воздействий на полупроводник к тепловым переходам ν_{12} добавляются переходы ν'_{12} нетепловой природы, и при этом частота обратных переходов ν_{21} тоже изменяется. Состояние полупроводника в таких условиях мы будем называть *термодинамически неравновесным*. При этом принцип детального равновесия, вообще говоря, уже не выполняется.

При нарушении термодинамического равновесия концентрации электронов и дырок в зонах n и p изменяются по сравнению с их равновесными значениями n_0 и p_0, т. е. в зонах появляются *неравновесные* носители заряда с концентрациями $\delta n = n - n_0$ и $\delta p = p - p_0$. При этом изменяются и концентрации связанных носителей n_t и p_t.

Сказанное справедливо, конечно, и для металлов. Однако концентрация электронов в металлах гораздо больше, чем в полупроводниках, и ее относительное изменение обычно ничтожно мало.

Отметим, что уже в гл. VI мы имели дело, по существу, с неравновесными состояниями, так как наличие тока нарушает термодинамическое равновесие (ср. § VI. 3). Однако там можно было считать, что по отношению к концентрации носителей заряда равновесие сохраняется. Это значит, что для данного полупроводника при заданной температуре установившаяся концентрация электронов определяется только значением электростатического потенциала (ср., например, § VI.12). Теперь мы рассмотрим явления, в которых нарушается равновесие и по концентрациям. Именно такие состояния мы будем называть в дальнейшем неравновесными.

§ 2. Время жизни неравновесных носителей заряда

В настоящей главе мы будем рассматривать поведение неравновесных носителей заряда только в объеме полупроводника, отвлекаясь от возможного влияния его поверхности. Это можно сделать, если

отношение поверхности к объему образца достаточно мало (строго говоря, для бесконечно протяженного полупроводника). Однако основные понятия и соотношения, вводимые в настоящей главе, сохраняются и в тех случаях, когда процессами генерации и рекомбинации на самой поверхности пренебрегать нельзя. Влияние поверхности на неравновесные состояния будет рассмотрено дополнительно в гл. X, где будут выяснены и условия, при которых этим влиянием можно пренебречь.

Положим, что под влиянием внешнего воздействия в каждой единице объема полупроводника в единицу времени возникает g_n электронов проводимости и, соответственно, g_p дырок в валентной зоне. Темпы генерации g_n и g_p будем считать сначала одинаковыми по всему объему, хотя и не обязательно равными друг другу. Пусть, далее, R_n есть темп обратного процесса исчезновения свободных электронов вследствие процессов рекомбинации с дырками (свободными и связанными на локальных уровнях энергии) и, соответственно, R_p — темп рекомбинации свободных дырок. Если в полупроводнике нет электрического тока, то изменение во времени неравновесных концентраций электронов и дырок в зонах определяется уравнениями

$$\frac{d\delta n}{dt} = g_n - R_n, \qquad \frac{d\delta p}{dt} = g_p - R_p. \tag{2.1}$$

Подчеркнем, что g_n и g_p обозначают генерацию, обусловленную только *внешними* воздействиями, и не включают переходы, вызванные тепловым движением. Последние мы учитываем в величинах R_n и R_p, которые поэтому представляют результирующие темпы рекомбинации, т. е. разности между темпом захвата носителей из соответствующей зоны и темпом обратной тепловой генерации носителей в зону:

$$R_n = r_n - g_{nT}, \qquad R_p = r_p - g_{pT}. \tag{2.2}$$

Для количественного описания кинетики неравновесных электронных процессов широко применяют понятия *среднего времени жизни неравновесных электронов* в зоне проводимости τ_n и, соответственно, *дырок* в валентной зоне τ_p, которые определяются формулами

$$R_n = \frac{n - n_0}{\tau_n}, \qquad R_p = \frac{p - p_0}{\tau_p}. \tag{2.3}$$

Или иначе: $1/\tau_n$ есть вероятность исчезновения одного избыточного электрона из зоны проводимости в единицу времени вследствие рекомбинации (со свободными и связанными дырками). Аналогично, $1/\tau_p$ есть вероятность рекомбинации одной избыточной дырки, тоже за единицу времени.

Так как r_n и r_p в формулах (2.2) по физическому смыслу не могут иметь слагаемых, не зависящих от n и, соответственно, p, то из

сравнения формул (2.2) и (2.3) следует, что

$$g_{nT}=\frac{n_0}{\tau_n}, \quad g_{pT}=\frac{p_0}{\tau_p}. \tag{2.4}$$

Следовательно, времена τ_n и τ_p определяют не только темп суммарной рекомбинации, но и темп тепловой генерации электронов и дырок.

Пользуясь понятиями времен жизни, уравнения кинетики (2.2) для однородного образца без тока можно записать в виде

$$\frac{d\delta n}{dt}=g_n-\frac{\delta n}{\tau_n}, \quad \frac{d\delta p}{dt}=g_p-\frac{\delta p}{\tau_p}. \tag{2.5}$$

Стационарные концентрации неравновесных носителей заряда, устанавливающиеся после длительного воздействия внешней генерации, равны

$$(\delta n)_s=g_n\tau_n, \quad (\delta p)_s=g_p\tau_p. \tag{2.6}$$

Величины τ_n и τ_p зависят от физических особенностей элементарных актов рекомбинации электронов и дырок. При этом τ_n и τ_p, вообще говоря, могут сами зависеть от неравновесных концентраций δ_n и δ_p, а также от температуры. Поэтому τ_n и τ_p не являются характеристиками данного полупроводника, но зависят еще от условий опыта. Эти вопросы будут подробнее рассмотрены в гл. IX (см. также § XVII.9). Сейчас же мы будем считать τ_n и τ_p заданными феноменологическими величинами, определяющими кинетику электронных процессов.

В простейшем случае, когда τ_n и τ_p не зависят от n и p, интегрирование уравнений (2.5) дает

$$\delta n=g_n\tau_n-C\exp(-t/\tau_n)$$

и такое же выражение для концентрации неравновесных дырок δp. Здесь C — постоянная интегрирования, определяемая начальными условиями. Так, если вначале полупроводник находился в термодинамическом равновесии и затем в момент времени $t=0$ было включено внешнее воздействие, мы имеем $t=0$, $\delta n=0$. Это дает

$$C=g_n\tau_n=(\delta n)_s, \quad \delta n=(\delta n)_s[1-\exp(-t/\tau_n)]. \tag{2.7а}$$

Если в некоторый момент времени $t=t_1$ генерация выключается, то для $t\geqslant t_1$, $g_n=0$, а начальное условие есть $t=t_1$, $\delta n=(\delta n)_1$. Тогда

$$C=-(\delta n)_1\exp\frac{t_1}{\tau_n}, \quad \delta n=(\delta n)_1\exp(-(t-t_1)/\tau_n). \tag{2.7б}$$

При постоянном τ_n установление избыточной концентрации электронов и ее исчезновение описываются экспоненциальным законом, а среднее время жизни τ_n есть время, в течение которого концентрация неравновесных электронов изменяется в e раз. То же справедливо и для неравновесных дырок.

В общем случае τ_n и τ_p, изменяющихся вместе с n и p, соотношения (2.3) определяют *мгновенные* времена жизни. Величины $1/\tau_n$ и $1/\tau_p$ в этом случае дают вероятности рекомбинации одной частицы в единицу времени при данных значениях n и p, а τ_n и τ_p определяют времена, за которые неравновесные концентрации электронов или, соответственно, дырок изменились бы в e раз, если бы вероятность их рекомбинации сохранялась постоянной и равной данному мгновенному значению.

§ 3. Уравнения непрерывности

Если в полупроводнике имеются электрические токи, то изменение концентрации носителей определяется не только генерацией и рекомбинацией, но и движением частиц. Вклад последнего легко найти, рассматривая элементарный объем, например, в виде прямоугольного параллелепипеда с ребрами, параллельными прямоугольным осям координат, и вычисляя втекающие и вытекающие из него потоки частиц. Это дает

$$\left(\frac{\partial p}{\partial t}\right)_{\text{дв}} = -\frac{1}{e}\left(\frac{\partial j_{px}}{\partial x}+\frac{\partial j_{py}}{\partial y}+\frac{\partial j_{pz}}{\partial z}\right) = -\frac{1}{e}\operatorname{div}\mathbf{j}_p,$$

$$\left(\frac{\partial n}{\partial t}\right)_{\text{дв}} = \frac{1}{e}\left(\frac{\partial j_{nx}}{\partial x}+\frac{\partial j_{ny}}{\partial y}+\frac{\partial j_{nz}}{\partial z}\right) = \frac{1}{e}\operatorname{div}\mathbf{j}_n.$$

Здесь $\mathbf{j}_p$ и $\mathbf{j}_n$ — плотности конвекционного тока, обусловленного движением дырок и, соответственно, электронов:

$$\mathbf{j}_p = \sigma_p\mathcal{E} - eD_p\nabla p, \qquad \sigma_p = ep\mu_p, \tag{3.1}$$

$$\mathbf{j}_n = \sigma_n\mathcal{E} + eD_n\nabla n, \qquad \sigma_n = en\mu_n. \tag{3.2}$$

Поэтому полные темпы изменения концентраций в зонах выражаются уравнениями

$$\frac{\partial p}{\partial t} = g_p - \frac{1}{e}\operatorname{div}\mathbf{j}_p - \frac{\delta p}{\tau_p}, \quad \frac{\partial n}{\partial t} = g_n + \frac{1}{e}\operatorname{div}\mathbf{j}_n - \frac{\delta_n}{\tau_n}. \tag{3.3}$$

Уравнения (3.3) — это уравнения непрерывности, записанные порознь для дырок и электронов.

При нарушении термодинамического равновесия изменяются также концентрации связанных дырок p_t (т. е. концентрация положительно заряженных доноров) и связанных электронов n_t (концентрация отрицательно заряженных акцепторов). Поэтому в общем случае возникает объемный заряд с плотностью

$$\rho = e(p + p_t - n - n_t).$$

Электрическое поле $\mathcal{E}$, входящее в выражения для плотностей тока (3.1) и (3.2), определяется уравнением Пуассона

$$\operatorname{div}\mathcal{E} = \frac{4\pi e}{\varepsilon}(p + p_t - n - n_t) \tag{3.4}$$

и граничными условиями. Величины p и n, с одной стороны, и p_t и n_t, с другой стороны, не являются независимыми, а связаны уравнениями кинетики рекомбинации, которые будут нами рассмотрены в гл. IX. Система уравнений (3.1) — (3.4) совместно с рекомбинационными уравнениями полностью определяет изменение избыточных концентраций носителей заряда в пространстве и времени. Однако уже сейчас мы рассмотрим некоторые выводы, которые могут быть получены без детального разбора процессов рекомбинации.

Прежде всего отметим, что уравнения (3.3) дают для суммарных величин ρ и $\mathbf{j}$ уравнение непрерывности в его обычной форме, которой, в частности, мы пользовались в гл. VI. Для этого учтем очевидное равенство

$$g_p - \frac{\delta p}{\tau_p} - \left(g_n - \frac{\delta n}{\tau_n}\right) = -\frac{\partial}{\partial t}(p_t - n_t), \tag{3.5}$$

которое справедливо при любом типе процессов рекомбинации. Тогда, вычитая друг из друга уравнения (3.3) и учитывая (3.5), получаем

$$\frac{\partial \rho}{\partial t} = -\operatorname{div} \mathbf{j}, \tag{3.6}$$

что для одномерного случая совпадает с уравнением (VI. 6.3).

Далее, укажем, что во многих важных случаях использование уравнения Пуассона (3.4) оказывается излишним. Это связано с тем, что при релаксации неравновесных электронных состояний существуют два различных процесса. Когда в полупроводнике нарушается термодинамическое равновесие, в нем возникают токи диффузии и дрейфа, стремящиеся уничтожить изменения объемного заряда. В случае неоднородного полупроводника или приповерхностного слоя в результате появления этих токов восстанавливается равновесное распределение объемного заряда, при котором ток диффузии уравновешивается током дрейфа. В однородном полупроводнике эти токи стремятся обратить объемный заряд в нуль. При неизменных связанных зарядах в зонах установилось бы *диффузионно-дрейфовое равновесие*, характеризуемое определенным распределением носителей заряда в образце. Если скорости дрейфа носителей заряда не слишком велики, так что время их пролета через образец $t_{пр} \gg \tau_M = \varepsilon/4\pi\sigma$, то в подавляющей части объема однородного образца объемный заряд успеет обратиться в нуль. При этом и токи, ограниченные объемным зарядом, не имеют места (ср. условие (VI.10.11)). Быстрота установления диффузионно-дрейфового равновесия в этом случае определяется максвелловским временем релаксации τ_M. Однако для установления полного термодинамического равновесия необходимо еще равновесие между электронами в зоне проводимости, дырками в валентной зоне и зарядами, связанными на ловушках. Быстрота установления такого *рекомбинационного равновесия* определяется временами жизни τ_n и τ_p.

В зависимости от соотношения между τ_M и рекомбинационными временами электронные процессы могут протекать весьма различно. Однако в технически важных полупроводниках обычно $\tau_M \ll \tau_n, \tau_p$, и мы ограничимся в дальнейшем только этим случаем. Тогда при частоте процессов $\omega \ll 1/\tau_M$ для объема однородного полупроводника можно считать

$$\rho \simeq 0 \qquad (\tau_M \ll t_{\text{пр}},\ \tau_M \ll \tau_n,\ \tau_p)$$

(*условие квазинейтральности*). При этом согласно (3.6)

$$\operatorname{div} \mathbf{j} = 0. \tag{3.7}$$

В связи с последним уравнением отметим следующее. Выражая ρ из уравнения Пуассона (3.4), подставляя его в уравнение непрерывности (3.6) и изменяя в этом последнем порядок дифференцирования по координатам и времени, мы получаем, что

$$\operatorname{div}\left(\mathbf{j} + \frac{\varepsilon}{4\pi}\frac{\partial \mathfrak{E}}{\partial t}\right) = 0. \tag{3.8}$$

Второе слагаемое в скобках есть ток смещения Максвелла, и поэтому уравнение (3.8) выражает известное положение теории Максвелла, что линии полного тока, т. е. суммы конвекционного тока $\mathbf{j}$ и тока смещения, непрерывны. Сравнивая уравнения (3.8) и (3.7), мы видим, что равенство (3.7) справедливо, если током смещения можно пренебречь по сравнению с конвекционным током.

Из написанных уравнений получаются два важных следствия. Положим, что изменения концентраций связанных зарядов малы, так что $\delta p_t,\ \delta n_t \ll \delta p,\ \delta n$. Это, в частности, справедливо, если суммарная концентрация примесей мала по сравнению с избыточными концентрациями носителей в зонах, т. е. либо при малом содержании примесей, либо при высоком уровне возбуждения. Тогда из условия $\delta\rho = 0$, следует, что $\delta p = \delta n$. В этом случае концентрации избыточных электронов и дырок в зонах одинаковы.

Положим теперь, что генерация происходит вследствие электронных переходов зона — зона ($g_p = g_n$). Тогда, вычитая уравнения (3.3) друг из друга и учитывая (3.7), получаем

$$\frac{\partial p}{\partial t} - \frac{\partial n}{\partial t} = -\frac{\delta p}{\tau_p} + \frac{\delta n}{\tau_n} = 0.$$

Поэтому, если $\delta p = \delta n$, то оба времени τ_p и τ_n равны и мы имеем единое время жизни электронно-дырочных пар $\tau = \tau_p = \tau_n$. Этот результат не зависит от особенностей процессов рекомбинации и справедлив как для прямой рекомбинации зона—зона, так и для рекомбинации через ловушки (см. гл. IX).

§ 4. Фотопроводимость

Простейший способ создания неравновесных носителей состоит в освещении полупроводника. Возникновение неравновесных носителей проявляется в изменении электропроводности полупроводника (фотопроводимость).

Электронные переходы при оптической генерации могут быть различными. Если энергия фотонов $\hbar\omega \geqslant E_g$, то неравновесные электроны и дырки образуются вследствие возбуждения электронов из валентной зоны в зону проводимости (*собственная* оптическая генерация и, соответственно, *собственная фотопроводимость*, рис. 7.3, *а*). Обратный процесс есть прямая рекомбинация свободный электрон — свободная дырка. Однако при наличии примесей фотопроводимость может возникать и при $\hbar\omega < E_g$. На рис. 7.3, *б*

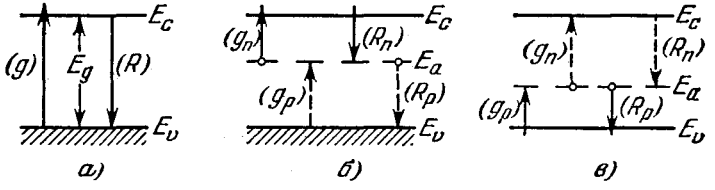


Рис. 7.3. Различные типы оптической генерации: *а*) собственная генерация пар; *б*) примесная генерация электронов; *в*) примесная генерация дырок.

показана энергетическая схема для конкретного случая полупроводника, содержащего глубокие акцепторы, частично компенсированные мелкими донорами. При этом уровни акцепторов расположены в верхней половине запрещенной зоны. Тогда при $\hbar\omega \geqslant E_c - E_a$ происходит возбуждение электронов, первоначально связанных на примеси, в зону проводимости и мы имеем случай *примесной* оптической генерации и, соответственно, *примесной фотопроводимости*. Если при этом $\hbar\omega < E_a - E_v$, то генерации свободных дырок не происходит вовсе и фотопроводимость оказывается монополярной. При $\hbar\omega > E_a - E_v$ генерируются и электроны, и дырки и фотопроводимость становится биполярной. Наконец, на рис. 7.3, *в* показан тот же случай частично компенсированных акцепторов, для которых, однако, уровень энергии расположен в нижней половине запрещенной зоны. Здесь при $\hbar\omega \geqslant E_a - E_v$ становится возможным оптический заброс электронов из валентной зоны на незанятые акцепторные уровни, т. е. генерация дырок (а не электронов), и возникает примесная монополярная фотопроводимость p-типа. При достаточном увеличении $\hbar\omega$ фотопроводимость опять становится биполярной.

Сказанное справедливо не только для акцепторных, но в равной мере и для донорных примесей.

Оптическая генерация электронов и дырок обязательно сопровождается дополнительным поглощением света. Целесообразно различать *собственное поглощение* света, наблюдаемое при $\hbar\omega > E_g$ и связанное с электронными переходами зона—зона и образованием пар, и *примесное поглощение*, связанное с возбуждением электронов и дырок с примесных уровней в зоны. Поглощение в собственной полосе частот обычно на много порядков больше поглощения в примесной области.

Темп оптической генерации связан с коэффициентом поглощения света. Пусть $I(x)$ есть монохроматический световой поток, рассчитанный на единицу поверхности, на расстоянии x от освещаемой поверхности полупроводника (эту величину мы будем называть интенсивностью света и измерять числом фотонов, проходящих через единицу поверхности в единицу времени), а γ — коэффициент поглощения (интенсивности) света. Тогда число фотонов, поглощаемых в единицу времени в слое с единичной площадью, расположенном между плоскостями x и $x + dx$, есть — $dI = I(x) \cdot \gamma\, dx$. А следовательно, число поглощенных фотонов, рассчитанное на единицу времени и единицу объема, равно $I(x) \cdot \gamma$. Поэтому темп оптической генерации g можно представить в виде

$$g = \nu(\omega)\,\gamma(\omega)\,I(x). \tag{4.1}$$

Здесь $\nu(\omega)$ — *квантовый выход* внутреннего фотоэффекта, который учитывает, что часть энергии поглощенных фотонов может быть передана непосредственно решетке или свободным носителям заряда в зонах без образования новых электронов и дырок.

В общем случае g различно в разных точках полупроводника (неоднородная генерация). Однако если $\gamma d \ll 1$ (где d — толщина пластинки), то $I(x) \simeq \text{const}$ (слабо поглощаемый свет) и g можно считать постоянным по объему (однородная генерация).

Изменение проводимости полупроводника обусловлено тем, что при освещении изменяются как концентрации электронов и дырок, так и их подвижности. Однако относительное влияние обеих этих причин может быть весьма различным. Действительно, возникающая в результате поглощения фотона пара электрон — дырка получает некоторый квазиимпульс и энергию ($\hbar\omega - E_g$). Пусть, для простоты, энергия передается только одному из фотоносителей, скажем электрону (что имеет место при сильном отличии масс m_n и m_p). Эта избыточная энергия затем растрачивается вследствие взаимодействия фотоэлектрона с решеткой, и через некоторое время, порядка времени релаксации энергии τ_E (см. § XVI.1), средняя энергия фотоэлектронов принимает значение, соответствующее температуре решетки. Аналогично, равновесное распределение квазиимпульса фотоэлектронов устанавливается за время порядка времени релаксации импульса τ_p (причем обычно $\tau_p \ll \tau_E$). Если $\tau_E \ll T_n$, где T_n — время существования фотоэлектронов в зоне,

то фотоэлектроны успевают «термолизоваться», т. е. приобрести такое же распределение по энергиям и квазиимпульсам, как и равновесные электроны. В этом случае подвижности не изменяются, а фотопроводимость обусловлена только изменением концентрации электронов и дырок и равна

$$\delta\sigma = e(\mu_p \delta p + \mu_n \delta n). \tag{4.2}$$

Если, напротив, $\tau_E \gtrsim T_n$, то за время своего существования фотоэлектроны не успевают термолизоваться и при освещении изменяются и концентрации фотоносителей, и их подвижности.

Влияние изменения подвижности проявляется обычно лишь при низких температурах (как правило, при водородных и гелиевых). В дальнейшем мы будем считать, что фотопроводимость вызвана только изменением концентрации электронов и дырок.

Будем считать, что происходит однородная генерация пар ($g_n = g_p = g = \text{const}$) и что в полупроводнике нет тока (или токи настолько слабы, что членами с $\operatorname{div} \mathbf{j}_p$ и $\operatorname{div} \mathbf{j}_n$ в уравнениях непрерывности (3.3) можно пренебречь). Тогда, умножая первое из уравнений (3.3) на $e\mu_p$, второе — на $e\mu_n$, складывая почленно оба уравнения, получаем

$$\frac{d\delta\sigma}{dt} = e(\mu_p + \mu_n) g - \frac{\delta\sigma}{\tau_{\text{фп}}}. \tag{4.3}$$

Здесь введено обозначение

$$\tau_{\text{фп}} = \frac{\mu_p \delta p + \mu_n \delta n}{\frac{\mu_p}{\tau_p} \delta p + \frac{\mu_n}{\tau_n} \delta n}. \tag{4.4}$$

Из уравнения (4.3) видно, что характерное время $\tau_{\text{фп}}$ есть *время релаксации фотопроводимости*, которое определяет темп установления и затухания $\delta\sigma$.

В стационарном состоянии фотопроводимость $(\delta\sigma)_s$ равна

$$(\delta\sigma)_s = e(\mu_p + \mu_n) g \cdot \tau_{\text{фп}}. \tag{4.5}$$

Здесь под $\tau_{\text{фп}}$ следует понимать его значение в стационарном состоянии, т. е. при данных установившихся значениях δp и δn. Отсюда видно, что чем больше $\tau_{\text{фп}}$, тем больше и $(\delta\sigma)_s$, т. е. тем выше чувствительность фотопроводника. Однако при этом будет больше и время установления (или затухания) фотопроводимости, т. е. будет больше инерционность фотопроводника. С этим противоречием между чувствительностью и быстродействием всегда приходится считаться при разработке фотосопротивлений для технических применений.

Для измерения фотопроводимости и ее кинетики разработано большое число различных методов, описание которых можно найти в специальной литературе [1]. Мы рассмотрим только один типичный

пример, показанный на рис. 7.4. Фотопроводник R, включенный последовательно с источником тока $Б$ и нагрузочным сопротивлением r, освещается прерывистым светом. Модуляция интенсивности света может быть получена, например, с помощью вращающегося непрозрачного диска $Д$ со щелями. Если ширина щелей равна ширине непрозрачных промежутков между ними, то полупроводник освещается в течение некоторого времени T, затем такое же время T интенсивность подсветки равна нулю и т. д. Вследствие изменения проводимости полупроводника в цепи возникает переменный ток, а на сопротивлении r — переменное напряжение. Последнее усиливается широкополосным усилителем $У$ и регистрируется осциллоскопом $О$. Зная параметры цепи (эдс источника, r) и темновое сопротивление фотопроводника R), отсюда можно легко найти фотопроводимость $\delta\sigma$ и ее изменение во времени.

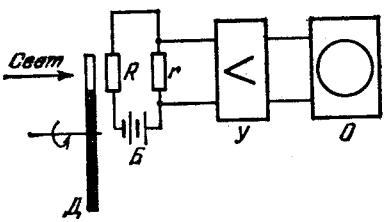


Рис. 7.4. Схема для наблюдения фотопроводимости.

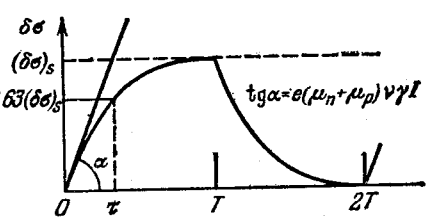


Рис. 7.5. Изменение фотопроводимости при прямоугольной модуляции освещенности ($\tau_{фп}$ = const).

Если бы $\tau_{фп}$ было постоянным, то $\delta\sigma$ изменялась бы во времени так, как показано на рис. 7.5. После включения освещения она возрастала бы по закону

$$\delta\sigma = (\delta\sigma)_s\left[1 - \exp\left(-\frac{t}{\tau_{фп}}\right)\right]$$

и при достаточно длинных импульсах света $T \gg \tau_{фп}$ достигала бы постоянного стационарного значения $(\delta\sigma)_s$. В последующий промежуток времени $(T - 2T)$ она уменьшалась бы от значения $(\delta\sigma)_s$ до 0 тоже по экспоненциальному закону:

$$\delta\sigma = (\delta\sigma)_s \exp\left(-\frac{t}{\tau_{фп}}\right).$$

Однако кривые рис. 7.5 наблюдаются в сравнительно редких случаях, а обычно зависимость $\delta\sigma$ (t) имеет более сложный характер. В этих случаях кинетика фотопроводимости не может быть охарактеризована единым временем релаксации и не может быть описана какими-либо универсальными соотношениями, так как она существенно зависит от особенностей процессов рекомбинации. Положение, однако, упрощается для двух предельных случаев: стационарного

состояния и начального участка процесса установления. В первом случае мы имеем

$$\delta p = g\tau_p, \quad \delta n = g\tau_n$$

и поэтому формула (4.4) дает

$$\tau_{\text{фп}} = \frac{\mu_p\tau_p + \mu_n\tau_n}{\mu_p + \mu_n}. \tag{4.6}$$

В этом случае стационарное время релаксации фотопроводимости (иногда называемое «временем жизни фотопроводимости») можно непосредственно выразить через времена жизни τ_n и τ_p. Из формулы (4.6) видно, что если, например, $\tau_p \gg \tau_n$, а подвижности одного порядка, то $\tau_{\text{фп}} \simeq \mu_p\tau_p/(\mu_p + \mu_n)$. При обратном неравенстве $\tau_{\text{фп}} \simeq \mu_n\tau_n/(\mu_p + \mu_n)$. Поэтому, измеряя $\tau_{\text{фп}}$, можно определить время жизни долгоживущих носителей заряда.

Для начального участка нарастания $\delta\sigma$ в уравнении (4.3) можно пренебречь вторым слагаемым в правой части (так как $\delta\sigma$ весьма мала). Тогда получается

$$\delta\sigma = e(\mu_p + \mu_n)gt. \tag{4.7}$$

Начальный участок кривой нарастания $\delta\sigma$ описывается прямой линией, наклон которой есть (рис. 7.5)

$$\operatorname{tg}\alpha = e(\mu_p + \mu_n)\nu\gamma I. \tag{4.8}$$

Этот участок вообще не зависит от процессов рекомбинации, а определяется только процессами генерации.

Исследование начальной стадии нарастания фотопроводимости часто используют для экспериментального определения квантового выхода ν. Такие опыты показывают, что ν может быть весьма различен [2]. Вблизи края собственного поглощения ($\hbar\omega \simeq E_g$) для многих полупроводниковых соединений $\nu \ll 1$. Малое значение ν в таких веществах объясняется тем, что фотоэлектроны и фотодырки в них могут связываться друг с другом, образуя так называемые экситоны, суммарный заряд которых равен нулю и которые поэтому не вносят вклада в электропроводность (см. гл. XVII). Однако известны и такие полупроводники (германий и кремний при не очень низких температурах, некоторые соединения $A^{III}B^{V}$), в которых $\nu \simeq 1$ в области края собственного поглощения. При достаточном увеличении энергии фотонов в таких веществах может быть $\nu > 1$. Это имеет место при комнатных температурах для германия при $\hbar\omega \gtrsim 4{,}3E_g$ и для кремния при $\hbar\omega \gtrsim 3E_g$. Причина увеличения ν состоит в том, что фотоэлектроны и фотодырки получают большую кинетическую энергию и способны поэтому сами генерировать дополнительные носители заряда вследствие ударной ионизации атомов решетки. Если энергия фотонов $\hbar\omega \gg E_g$, то $\nu = \hbar\omega/w_i$, где w_i — средняя энергия, необходимая для создания одной пары

электрон — дырка. Последняя имеет порядок немногих электронвольт (например, для германия $w_i \simeq 2{,}5$ эВ). Поэтому для жесткого рентгеновского и γ-излучений значения ν могут быть очень велики. То же имеет место и для быстрых электронов и других корпускулярных излучений. Это обстоятельство в настоящее время используют в полупроводниковых счетчиках элементарных частиц.

§ 5. Квазиуровни Ферми

В термодинамически неравновесном состоянии (например, при освещении полупроводника) уже не существует единого уровня Ферми для всей системы и поэтому выражения для концентраций электронов и дырок, полученные нами в гл. V, уже несправедливы. При этом, в частности, не выполняется и соотношение $np = n_i^2$.

Однако, следуя Шокли, можно обобщить соотношения статистики на неравновесные состояния, если вместо уровня Ферми формально ввести новые величины — *квазиуровни Ферми*. Положим, что вероятность заполнения электроном состояния с энергией E в зоне проводимости можно представить в виде, по форме совпадающем с распределением Ферми — Дирака (V.3.1):

$$f_n = \left(1 + \exp\frac{E - F_n}{kT}\right)^{-1}. \tag{5.1}$$

Тогда, по определению, величина F_n есть квазиуровень Ферми для электронов.

Аналогично, для вероятности нахождения вакансии (дырки) на уровне энергии E в валентной зоне положим

$$f_p = \left(1 + \exp\frac{F_v - E}{kT}\right)^{-1}, \tag{5.2}$$

где, по определению, F_p есть квазиуровень Ферми для дырок. Тогда очевидно, что для n и p мы получим те же соотношения, что и в гл. V, в которые, однако, вместо уровня Ферми F будут входить квазиуровни F_n и, соответственно, F_p. В частности, в невырожденных полупроводниках для неравновесных электронов будет по-прежнему справедливо распределение Больцмана

$$f_n = \exp\frac{F_n - E}{kT}. \tag{5.1а}$$

Соответственно вместо формул (V.5.1) и (V.5.2) мы получим

$$n = n_0 + \delta n = N_c \exp\frac{F_n - E_c}{kT}, \tag{5.3}$$

$$p = p_0 + \delta p = N_v \exp\frac{E_v - F_p}{kT}. \tag{5.4}$$

Вместо формулы (V.5.4) мы будем иметь

$$pn = n_i^2 \exp\frac{F_n - F_p}{kT}. \tag{5.5}$$

Таким образом, появление в зонах неравновесных электронов и дырок можно описать как «расщепление» первоначального уровня Ферми F на два квазиуровня F_n и F_p, каждый из которых смещается по направлению к своей зоне (рис. 7.6). Совершенно аналогично, для нахождения концентраций связанных носителей на ловушках n_t и p_t мы могли бы воспользоваться функцией распределения (V.9.3) (или более общей формулой (V.11.8)), заменив в ней F на некоторый квазиуровень Ферми для ловушек F_t, который, вообще говоря, отличается от F_n и F_p.

Введение квазиуровней Ферми физически обозначает предположение, что времена релаксации импульса τ_p и энергии τ_E для электронов и дырок намного меньше времени их существования в зонах. Как уже говорилось в § 4, в этом случае можно считать, что в электронном и дырочном газах устанавливается равновесное фермиевское распределение, и притом с одной и той же температурой для всей системы. Однако равновесия по отношению к концентрациям электронного и дырочного газов при этом может и не быть. Именно это и учитывается введением различных квазиуровней Ферми для дырок и электронов. Отметим, что указанное предположение a priori не очевидно. Однако имеются эксперименты, позволяющие его проверить (см. § 9). Они показывают, что, по крайней мере для некоторых полупроводников, оно действительно оправдывается.

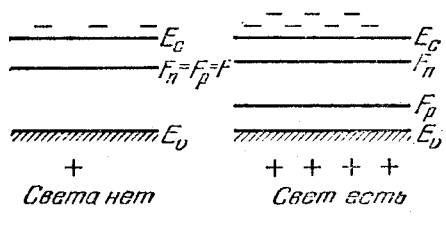


Рис. 7.6. К понятию квазиуровней Ферми.

Отметим теперь два важных свойства квазиуровней Ферми. В § VI.3 мы видели, что плотность тока в невырожденных полупроводниках пропорциональна градиенту уровня Ферми. Повторяя те же рассуждения для рассматриваемого случая биполярной неравновесной проводимости и используя вместо соотношений (V.5.1) и (V.5.2) обобщенные формулы (5.3) и (5.4), мы получим для электронной и дырочной составляющих плотности тока $\mathbf{j}_n$ и $\mathbf{j}_p$ формулы, аналогичные (VI.3.2):

$$\mathbf{j}_n = \mu_n n(\mathbf{r}) \nabla F_n, \tag{5.6}$$

$$\mathbf{j}_p = \mu_p p(\mathbf{r}) \nabla F_p. \tag{5.7}$$

Отсюда видно, что при наличии тока квазиуровни Ферми изменяются в пространстве, и тем быстрее, чем меньше локальные значения концентраций электронов n или, соответственно, дырок p.

Далее, легко видеть, что разность квазиуровней Ферми на концах полупроводника непосредственно связана с разностью электрических потенциалов, существующей между этими концами. Поло-

жим сначала, что в полупроводнике имеются только электроны и будем считать для простоты, что все величины зависят от одной координаты x. Тогда, вводя в формулу (5.6) электропроводность $\sigma_n(x) = e\mu_n n(x)$, имеем

$$j_{nx} = \frac{1}{e}\sigma_n(x)\frac{dF_n}{dx}.$$

С другой стороны, согласно общей форме дифференциального закона Ома, можно написать

$$j_{nx} = \sigma_n(x)(\mathcal{E}_x + \mathcal{E}^*_{nx}),$$

где $\mathcal{E}_x$ — напряженность кулоновского электрического поля, а $\mathcal{E}^*_{nx}$ — напряженность поля сторонних сил, действующих на электроны. (В рассматриваемом случае это есть сила давления электронного газа, обусловленная градиентом концентрации электронов). Сравнивая эти соотношения, получаем

$$\frac{dF_n}{dx} = e(\mathcal{E}_x + \mathcal{E}^*_{nx}).$$

Поэтому разность значений квазиуровней Ферми для двух каких-либо сечений полупроводника A и B (рис. 7.7) равна

$$F_{nB} - F_{nA} = e\left(\int_A^B \mathcal{E}_x\,dx + \int_A^B \mathcal{E}^*_{nx}\,dx\right).$$

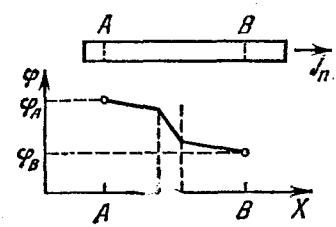


Рис. 7.7. К установлению связи между квазиуровнями Ферми и внешним напряжением.

Первый интеграл в правой части есть уменьшение потенциала при переходе от A к B, обусловленное током (или ir, где i — сила тока, а r — сопротивление участка). Второй интеграл есть электродвижущая сила V_0. Она вызывает увеличение потенциала при переходе от A к B на V_0. Поэтому все выражение в круглых скобках есть

$ir - V_0$ = разности потенциалов на концах участка $(\varphi_A - \varphi_B)$

и

$$F_{nB} - F_{nA} = -e(\varphi_B - \varphi_A).$$

Распределение потенциала показано на рис. 7.7 снизу, где для ясности чертежа предположено, что эдс имеется только в некотором слое.

Рассуждая аналогично, мы найдем, что если бы были только дырки, то разница квазиуровней для дырок была бы

$$F_{pB} - F_{pA} = -e(\varphi_B - \varphi_A)',$$

где разность потенциалов теперь была бы, вообще говоря, другой, так как распределения неравновесных электронов и дырок различны и сторонние силы $\mathcal{E}^*_{nx}$ и $\mathcal{E}^*_{px}$ неодинаковы.

Выберем теперь сечения A и B в таких частях полупроводника, где концентрации избыточных носителей $\delta p = \delta n = 0$. Тогда $F_{nB} = F_{pB} = F_B$, $F_{nA} = F_{pA} = F_A$ и, следовательно, $(\varphi_B - \varphi_A) = (\varphi_B - \varphi_A)'$. Поэтому

$$F_B - F_A = -e(\varphi_B - \varphi_A), \tag{5.8}$$

где $(\varphi_B - \varphi_A)$ есть приложенное внешнее напряжение. Таким образом, разность квазиуровней Ферми есть та величина, которую мы непосредственно измеряем вольтметром.

§ 6. Электронно-дырочные переходы

Для создания неравновесных носителей заряда широко применяют электронно-дырочные переходы (p—n-переходы), о которых мы уже говорили в § VI.9. Чтобы избежать сложного и неконтролируемого влияния микрогеометрии поверхности, такие переходы осуществляют не механическим соединением двух полупроводников, а внутри единого монокристалла, в котором создают подходящее распределение донорной и акцепторной примесей, например показанное на рис. 7.8. Если эти примеси полностью ионизованы (например, элементы III и V групп в германии и кремнии при комнатных температурах), то в левой части кристалла будет дырочная проводимость с концентрацией основных носителей $p \simeq N_a - N_d$, а в правой части — электронная ($n \simeq N_d - N_a$). Между ними расположен переходный слой («технологический» переход), в котором концентрация примесей быстро изменяется. В некоторой тонкой области этого слоя доноры и акцепторы компенсируют друг друга ($N_d \simeq N_a$) и имеет место собственная проводимость (i).

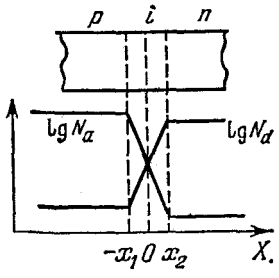


Рис. 7.8. p — n-переход.

Нужное распределение доноров и акцепторов можно осуществить различными технологическими приемами: добавлением одной из примесей в расплав в процессе роста кристалла; диффузией из газовой фазы одной из примесей (например, донорной в кристалл, уже имеющий акцепторы); сплавлением полупроводника p- (или n-) типа с металлом, являющимся донором (или, соответственно, акцептором), и другими, на которых мы не будем останавливаться.

Распределение концентраций электронов и дырок в p—n-переходе в отсутствие тока показано на рис. 7.9 (внизу) сплошными линиями *1* и *2*. При обозначении концентраций мы будем использо-

вать дополнительный индекс (p или n), указывающий, к какой области перехода относится данная концентрация. Так, концентрация дырок в дырочной области p_p (основные носители) постоянна и велика, она уменьшается (на много порядков) в переходной области и принимает малое значение p_n в n-области (неосновные носители). Аналогично, концентрация электронов изменяется от большого значения n_n в n-области до малой величины n_p в p-области.

Положим теперь, что к переходу приложено внешнее напряжение так, что на p-области находится «+» источника (как и в гл. VI, напряжение такого знака мы будем считать положительным). При этом дырки из p-области устремятся в n-область, где они станут неосновными носителями. Так как $p_p \gg p_n$, эти дырки будут рекомбинировать с электронами. Однако, вследствие конечности времени жизни дырок τ_p, рекомбинация произойдет не сразу, и поэтому в некоторой области за пределами перехода концентрация дырок будет оставаться больше p_n. Одновременно с этим увеличится и концентрация электронов в n-области, так как дополнительные электроны войдут из электрода для компенсации объемного заряда пришедших дырок. Аналогично, основные носители в n-области — электроны — будут переходить в p-область, становясь там неосновными носителями, и постепенно рекомбинировать с основными носителями — дырками. Поэтому и слева от перехода концентрация электронов увеличится, а также увеличится и концентрация дырок, которые войдут из левого электрода для компенсации объемного заряда электронов. Это явление получило название *инжекции неосновных носителей*. Однако, как мы видим, в действительности оно заключается в увеличении концентрации носителей обоего знака по обе стороны от перехода, т. е. в возникновении *квазинейтральных областей повышенной проводимости*. Распределение концентрации неосновных носителей при инжекции показано на рис. 7.9 пунктиром (кривые *3* и *4*). Из сказанного ясно, что инжекция неосновных носителей возможна только в полупроводниках со *смешанной* (биполярной) проводимостью.

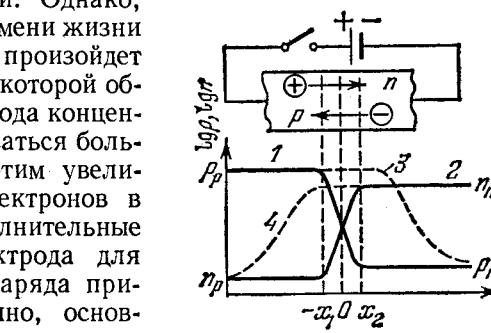


Рис. 7.9. Инжекция в $p-n$-переходе.

Для количественных расчетов электронных явлений, связанных с инжекцией, необходимо знать неравновесные концентрации дырок и электронов на границах перехода x_2 и $-x_1$. Они выражаются особенно просто, если а) полупроводники не вырождены и, кроме того, б) толщина переходной области $d = x_1 + x_2$ достаточно мала, так что рекомбинацией в этой области можно пренебречь. Количественно

второе условие означает, что длины диффузии (см. § 9) дырок L_p и электронов $L_n \gg d$. Тогда, следуя Шокли, искомые концентрации можно непосредственно получить из сопоставления хода квазиуровней Ферми F_n и F_p и краев энергетических зон. Если внешнее напряжение $u = 0$, то $F_n = F_p = F$ (рис. 7.10, *а*) и концентрации неосновных носителей на границах имеют равновесные значения

$$p(x_2) = p_n, \quad n(-x_1) = n_p \qquad (u = 0).$$

При наличии внешнего напряжения $u > 0$, во-первых, края зон E_c и E_v в n-области поднимаются относительно p-области на величину eu и, во-вторых, в окрестности перехода $F_n \neq F_p$ (рис. 7.10, *б*). Однако вдали от переходной области мы имеем:

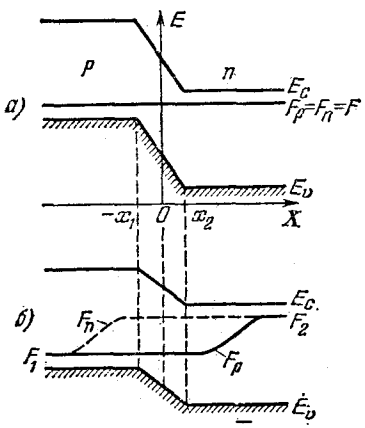


Рис. 7.10. Ход квазиуровней Ферми в $p-n$-переходе: *а*) $u = 0$; *б*) $u > 0$.

$$x \to \infty: \quad F_p = F_n = F_2,$$
$$p = p_n, \quad n = n_n;$$
$$x \to -\infty: \quad F_p = F_n = F_1,$$
$$p = p_p, \quad n = n_p.$$

При этом, согласно формуле (5.8),

$$F_1 = F_2 - eu.$$

Теперь учтем соотношения (5.6) и (5.7), которые показывают, что если концентрация носителей велика, то F_p и F_n меняются в пространстве слабо. Если рекомбинация в переходной области мала, то p имеет в ней большое значение p_p, и поэтому можно считать, что уровень F_p в этой области горизонтален. Или, иначе,

$$F_p(x_2) \simeq F_1 = F_2 - eu.$$

Аналогично, для электронов

$$F_n(-x_1) \simeq F_2 = F_1 + eu.$$

Тогда, подставляя эти выражения для F_p и F_n в формулы (5.3) и (5.4), получаем окончательно

$$p(x_2) = N_v \exp\frac{E_v - F_p(x_2)}{kT} = N_v \exp\frac{E_v - F_2}{kT} \cdot \exp\frac{eu}{kT} = p_n \exp\frac{eu}{kT},$$

$$n(-x_1) = N_c \exp\frac{F_n(-x_1) - E_c}{kT} = n_p \exp\frac{eu}{kT}. \qquad (6.1)$$

Отметим, что при $T = 300$ К, $e/kT = 39$ В$^{-1}$. Поэтому, прикладывая даже небольшое напряжение, можно очень сильно менять концентрацию неосновных носителей на границах. Например, при $u = 0{,}2$В она изменяется в $e^8 \simeq 10^3$ раз.

§ 7. Обнаружение неравновесных носителей заряда

Избыточные электроны и дырки могут быть обнаружены по изменению электропроводности. Они могут быть также зарегистрированы с помощью p—n-перехода. Когда вблизи перехода появляются избыточные носители, то электрическое поле, существующее внутри перехода, затягивает электроны в n-область, а дырки — в p-область (рис. 7.11). Поэтому при разомкнутой цепи n-область заряжается отрицательно, а p-область — положительно, т. е. в p—n-переходе возникает эдс. При замыкании цепи в ней возникает ток.

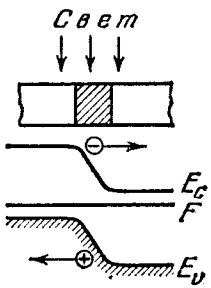


Рис. 7.11. При возникновении в окрестности p — n-перехода неравновесных электронов и дырок появляется эдс.

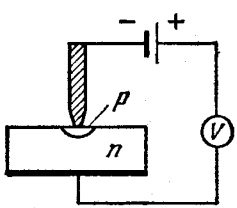


Рис. 7.12. p — n-переход в точечном контакте металл — полупроводник.

Микроскопические p—n-переходы часто образуются в контактах металлической заостренной проволоки с полупроводником, в особенности после «формовки» контакта импульсами сильного тока. Причины образования переходов могут быть различны: диффузия материала проволоки внутрь полупроводника, изменение концентрации электрически активных примесей при местном разогреве вследствие формовки (образование так называемых «термических» донорных или акцепторных центров) и другие. Для увеличения потенциального барьера внутри перехода на контакт иногда накладывают отрицательное напряжение (рис. 7.12).

На рис. 7.13, *а* показана одна из схем для исследования явления инжекции. Образец полупроводника (который для определенности положен n-типа) имеет форму тонкого и длинного стержня («нитевидный» образец). Инжекция осуществляется в точечном контакте $Э$ («эмиттер»), к которому прикладывают короткие импульсы напряже-

ния от импульсного генератора *Г*. Инжектированные носители регистрируются вторым точечным контактом *К* («коллектор»). Об относительных изменениях концентрации избыточных носителей можно судить по напряжению *u* на нагрузочном сопротивлении *r* в цепи коллектора. Это напряжение может быть усилено и зарегистрировано осциллоскопом *О* (развертка которого запускается генератором *Г*). Если к торцам образца не приложено внешнее напряжение или это напряжение имеет знак, показанный на рисунке, то сигнал коллектора имеет вид, изображенный на рис. 7.13, *б*. Когда генератор *Г* создает импульс напряжения, изменение потенциала эмиттера передается по цепи практически мгновенно (со скоростью

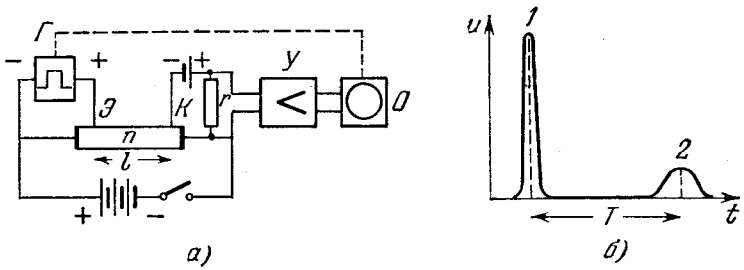


Рис. 7.13. Измерение скорости движения пакета инжектированных носителей.

света), и на экране виден острый пик *1* («наводка»). При этом в образце вблизи *Э* возникает пакет инжектированных носителей, который движется от *Э* к *К*. Когда пакет достигает коллектора, появляется пик второго сигнала *2* (который более размыт вследствие диффузионного расплывания пакета). При изменении полярности внешнего напряжения сигнал *2* не появляется.

Этот опыт (и ему подобные) обнаруживает неожиданную на первый взгляд особенность движения инжектированных носителей: пакет, будучи электрически нейтральным, управляется электрическим полем. Более того, направление его движения совпадает с направлением движения неосновных носителей, хотя их концентрация может быть на много порядков меньше концентрации основных носителей. Эта особенность, однако, становится вполне понятной, если учесть кулоновское взаимодействие между дырками и электронами (см. § 8).

Измеряя время *T* движения пакета и зная расстояние *l* между *Э* и *К*, можно найти скорость пакета $v = l/T$. Она оказывается пропорциональной полю $\mathscr{E}$ внутри образца. Отсюда можно определить подвижность инжектированного пакета:

$$\mu = \frac{v}{\mathscr{E}}.$$

Подобные опыты показывают, что если только материал не близок к собственному, то μ не только по знаку, но и по величине практически совпадает с подвижностью неосновных носителей.

На рис. 7.14, *а* показан принцип другой типичной схемы, в которой избыточные носители создаются светом, а роль эмиттера играет изображение I' ярко освещенной узкой щели I. Концентрация избыточных носителей (в относительных единицах) измеряется опять при помощи точечного коллектора K. Для повышения чувствительности схемы интенсивность света можно модулировать с низкой

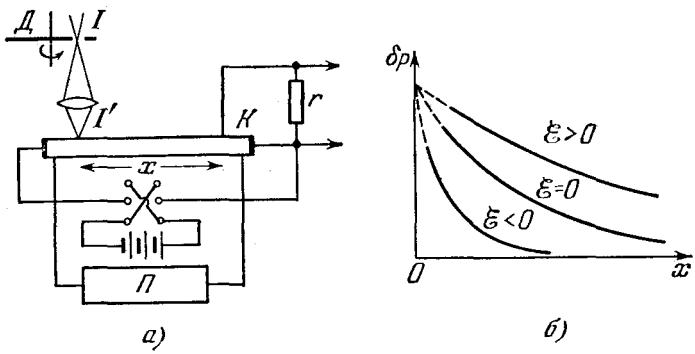


Рис. 7.14. Измерение длины диффузии и длины затягивания электрическим полем.

частотой (например, вращающимся диском $Д$ со щелями), что позволяет легко усиливать сигнал коллектора. Для точного определения тянущего поля показан потенциометр $П$, позволяющий измерять падение напряжения в образце компенсационным методом и этим исключить возможные скачки потенциала в зондах.

Изменяя в этой схеме расстояние x между световой щелью и коллектором, можно определить стационарное распределение концентрации неравновесных (избыточных) носителей в пространстве. Если концентрация избыточных носителей δp достаточно мала $(\delta p/(n_0+p_0) \ll 1)$, то δp убывает по мере удаления от освещенной части по экспоненциальному закону (рис. 7.14, *б*). Длина L, на которой δp уменьшается в e раз в отсутствие электрического поля, есть, по определению, *длина диффузии* неравновесных носителей.

В присутствии электрического поля распределение концентрации определяется другой характерной длиной $l(\mathscr{E})$, зависящей от поля $\mathscr{E}$ *(длиной затягивания)*. Если направление поля соответствует движению неосновных носителей вглубь образца, то $l(\mathscr{E}) > L$ и нейтральная область с избыточными носителями растягивается. При противоположном направлении поля $l(\mathscr{E}) < L$ и область повышенной концентрации носителей сжимается (рис. 7.14, *б*).

§ 8. Амбиполярная диффузия и амбиполярный дрейф

Особенности движения инжектированного пакета носителей легко понять, если учесть, что электроны и дырки суть заряженные частицы и что при их перераспределении возникает электрическое поле, которое в свою очередь воздействует на их движение. Поэтому диффузия избыточных носителей будет характеризоваться некоторым общим, *амбиполярным коэффициентом диффузии*. Так, если поверхность полупроводника освещать сильно поглощаемым светом, то в тонком приповерхностном слое возникнет повышенная концентрация электронов и дырок, которые будут диффундировать вглубь образца (рис. 7.15). Если, например, $D_n > D_p$, то электроны будут опережать дырки и поэтому в полупроводнике появятся заряды и возникнет электрическое поле (*поле амбиполярной диффузии*), которое будет тормозить электроны и ускорять дырки. При $D_n < D_p$ направление этого поля будет противоположным. В установившемся состоянии в каждой точке полупроводника будет такое поле, при котором потоки дырок и электронов равны друг другу. Совершенно аналогично, вследствие неравенства подвижностей электронов и дырок μ_n и μ_p во внешнем электрическом поле пакет инжектированных носителей приобретает некоторую общую, *амбиполярную дрейфовую скорость*.

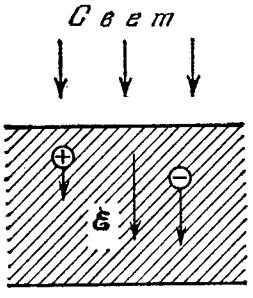


Рис. 7.15. Амбиполярная диффузия.

Величину коэффициента амбиполярной диффузии можно найти непосредственно из уравнений (3.1) и (3.2). Складывая почленно эти уравнения, имеем

$$\mathbf{j} = \mathbf{j}_p + \mathbf{j}_n = \sigma\mathcal{E} + eD_n \nabla n - eD_p \nabla p.$$

Отсюда полная напряженность поля

$$\mathcal{E} = \frac{\mathbf{j}}{\sigma} + \frac{e}{\sigma}(D_p \nabla p - D_n \nabla n). \tag{8.1}$$

Первое слагаемое в этом выражении есть поле, которое существует в данном полупроводнике при токе $\mathbf{j}$ в отсутствие избыточных носителей. Второе слагаемое есть поле амбиполярной диффузии

$$\mathcal{E}_a = \frac{e}{\sigma}(D_p \nabla p - D_n \nabla n). \tag{8.2}$$

В дальнейшем мы будем предполагать, что связанными зарядами на ловушках можно пренебречь и считать $\delta p = \delta n$, $\nabla p = \nabla n$. Тогда, подставляя выражение (8.1) для $\mathcal{E}$ в формулы для плотностей токов

(3.1), (3.2), мы получаем

$$\mathbf{j}_p = \frac{\sigma_p}{\sigma}\mathbf{j} - eD\nabla p, \quad \mathbf{j}_n = \frac{\sigma_n}{\sigma}\mathbf{j} + eD\nabla n, \tag{8.3}$$

где через D обозначено

$$D = \frac{\sigma_p D_n + \sigma_n D_p}{\sigma}. \tag{8.4}$$

Вторые слагаемые в формулах (8.3), пропорциональные $\nabla p = \nabla n$, дают токи диффузии дырок и электронов, которые, как и следовало ожидать, равны друг другу. Они определяются одним и тем же коэффициентом амбиполярной диффузии D. Полагая в формуле (8.4) $\sigma = \sigma_p + \sigma_n = ep\mu_p + en\mu_n$ и используя еще соотношение Эйнштейна $\mu_p = (e/kT)D_p$, $\mu_n = (e/kT)D_n$, получаем

$$D = \frac{p+n}{\dfrac{p}{D_n} + \dfrac{n}{D_p}}. \tag{8.4а}$$

При малой концентрации избыточных носителей в этой формуле можно положить $n \simeq n_0$, $p \simeq p_0$.

Полученные результаты показывают, что в материале n-типа ($n \gg p$) $D \simeq D_p$. Для материала p-типа ($p \gg n$) $D \simeq D_n$. Следовательно, в обоих случаях D совпадает с коэффициентом диффузии *неосновных* носителей. Для собственного полупроводника ($n = p = n_i$) мы имеем промежуточное значение $D = 2D_pD_n/(D_p + D_n)$. Общий характер зависимости D от концентрации электронов показан на рис. 7.16.

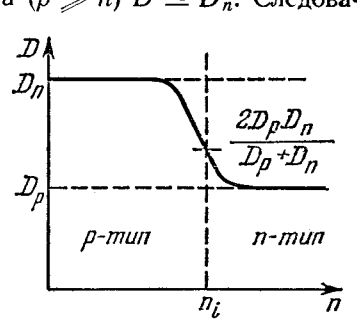


Рис. 7.16. Зависимость коэффициента амбиполярной диффузии от концентрации электронов.

Рассмотрим теперь движение пакета избыточных носителей во внешнем электрическом поле. Для этого мы используем не только выражения (3.1), (3.2) для плотностей токов, но еще и уравнения непрерывности (3.3). Как и выше, будем считать, что полупроводник однороден, $\delta p = \delta n$, и, соответственно, времена жизни электронов и дырок одинаковы: $\tau_p = \tau_n = \tau$ (ср. § 3).

Остановимся сначала на очень наглядном случае, когда поля настолько сильны, что влиянием диффузии и рекомбинации можно пренебречь по сравнению с влиянием дрейфа. Кроме того, положим, что все величины зависят от одной координаты x. Тогда уравнения

(3.1) — (3.3) принимают простой вид:

$$\mathbf{j}_p = \sigma_p(x,\ t)\,\mathfrak{E}(x,\ t), \qquad \mathbf{j}_n = \sigma_n(x,\ t)\,\mathfrak{E}(x,\ t), \tag{8.5}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{e}\operatorname{div}\mathbf{j}_p = -\frac{1}{e}\sigma_p\frac{\partial \mathfrak{E}}{\partial x} - \frac{1}{e}\mathfrak{E}\frac{\partial \sigma_p}{\partial x},$$

$$\frac{\partial n}{\partial t} = \frac{1}{e}\operatorname{div}\mathbf{j}_n = \frac{1}{e}\sigma_n\frac{\partial \mathfrak{E}}{\partial x} + \frac{1}{e}\mathfrak{E}\frac{\partial \sigma_n}{\partial x}, \tag{8.6}$$

$$\sigma_p = ep\mu_p, \qquad \sigma_n = en\mu_n, \qquad \sigma = \sigma_n + \sigma_p. \tag{8.7}$$

Мы положили внешнюю генерацию $g_p = g_n = 0$, так как она несущественна для последующих рассуждений. Умножая первое уравнение (8.6) на σ_n, второе уравнение на σ_p и складывая оба уравнения, исключим $\partial\mathfrak{E}/\partial x$. Тогда, учитывая еще, что $\partial p/\partial t = \partial n/\partial t$, получим

$$-\frac{\partial p}{\partial t} = \frac{\mu_p\mu_n(n-p)}{\mu_p p + \mu_n n}\mathfrak{E}\frac{\partial p}{\partial x}. \tag{8.8}$$

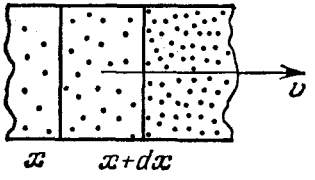


Рис. 7.17. К вопросу об амбиполярной скорости дрейфа.

Легко видеть, что множитель при $\partial p/\partial x$ есть скорость, с которой перемещается пакет неравновесных электронов и дырок в электрическом поле (амбиполярная скорость дрейфа). Действительно, рассмотрим газ частиц с концентрацией $p(x)$, движущихся с одинаковой скоростью v вдоль оси X (рис. 7.17). Тогда уменьшение числа частиц за единицу времени в бесконечно тонком слое между плоскостями x и $x + dx$ (рассчитанное на единичное сечение) есть

$$-dx\frac{\partial p}{\partial t} = vp(x+dx) - vp(x) = v\frac{\partial p}{\partial x}dx,$$

или

$$-\frac{\partial p}{\partial t} = v\frac{\partial p}{\partial x}.$$

Сравнивая это соотношение с формулой (8.8), мы видим, что амбиполярная скорость дрейфа пакета избыточных носителей равна

$$v = \frac{\mu_p\mu_n(n-p)}{\mu_p p + \mu_n n}\mathfrak{E}. \tag{8.9}$$

Она пропорциональна $\mathfrak{E}$, а следовательно, можно ввести *амбиполярную подвижность пакета* μ, которая есть

$$\mu = \frac{v}{\mathfrak{E}} = \frac{n-p}{\dfrac{p}{\mu_n} + \dfrac{n}{\mu_p}}. \tag{8.10}$$

При малых нарушениях равновесия в этой формуле можно заменить p и n их равновесными значениями p_0 и n_0.

Полученные результаты объясняют опытные факты по влиянию электрического поля на движение пакета инжектированных носителей. Из формулы (8.10) видно, что если только $n \neq p$, то $\mu \neq 0$, а значит, поле сообщает пакету определенную скорость, несмотря на то что эта область в целом электрически нейтральна. В материале n-типа $n \gg p$ и формула (8.10) дает $\mu \simeq \mu_p$. В этом случае $\mu > 0$ и пакет движется в направлении поля, т. е. в том же направлении, что и неосновные носители — положительные дырки. В материале p-типа мы имеем $p \gg n$ и поэтому $\mu \simeq -\mu_n < 0$. Это обозначает, что пакет перемещается против поля, т. е. тоже как неосновные носители, которыми здесь являются отрицательные электроны. В материале с собственной проводимостью $n = p$, и формула (8.10) дает $\mu = 0$, а следовательно, пакет не управляется полем вовсе.

В более общем случае, когда диффузией и рекомбинацией пренебречь нельзя, амбиполярная скорость дрейфа вычисляется аналогично. Подставляя выражения (3.1) и (3.2) для $\mathbf{j}_p$ и $\mathbf{j}_n$ в уравнения непрерывности (3.3), мы имеем

$$\frac{\partial p}{\partial t} = g - \frac{1}{e}\mathfrak{E}\cdot\nabla\sigma_p - \frac{1}{e}\sigma_p \operatorname{div}\mathfrak{E} + D_p \operatorname{div}(\nabla p) - \frac{\delta p}{\tau},$$

$$\frac{\partial n}{\partial t} = g + \frac{1}{e}\mathfrak{E}\cdot\nabla\sigma_n + \frac{1}{e}\sigma_n \operatorname{div}\mathfrak{E} + D_n \operatorname{div}(\nabla n) - \frac{\delta n}{\tau}.$$

Исключая, как и выше, из этих уравнений члены с $\operatorname{div}\mathfrak{E}$, мы найдем

$$\frac{\partial p}{\partial t} = g - \frac{1}{e\sigma}(\sigma_n\cdot\nabla\sigma_p - \sigma_p\cdot\nabla\sigma_n)\mathfrak{E} + \frac{\sigma_n D_p + \sigma_p D_n}{\sigma}\operatorname{div}(\nabla p) - \frac{\delta p}{\tau}. \tag{8.11}$$

Подставляя для σ_n, σ_p и σ их значения (8.7), имеем

$$\frac{1}{e\sigma}(\sigma_n\nabla\sigma_p - \sigma_p\nabla\sigma_n) = \frac{n-p}{\dfrac{p}{\mu_n} + \dfrac{n}{\mu_p}}\nabla p.$$

Далее, учитывая формулу (8.4), третье слагаемое в правой части (8.11) можно записать в виде

$$D\operatorname{div}(\nabla p) = \operatorname{div}(D\cdot\nabla p) - \nabla D\cdot\nabla p.$$

После этого мы получаем

$$\frac{\partial p}{\partial t} = g - \mathbf{v}\cdot\nabla p + \operatorname{div}(D\cdot\nabla p) - \frac{\delta p}{\tau}, \tag{8.12}$$

где

$$\mathbf{v} = \frac{n-p}{\dfrac{p}{\mu_n} + \dfrac{n}{\mu_p}}\mathfrak{E} + \nabla D. \tag{8.13}$$

Очевидно, что второе слагаемое в правой части (8.12) есть быстрота изменения концентрации во времени, обусловленная движением

пакета. В этом можно убедиться, рассуждая, как и выше, но рассматривая вместо бесконечно тонкого слоя бесконечно малый параллелепипед с ребрами, параллельными осям координат. Скорость дрейфа пакета $\mathbf{v}$ выражается формулой (8.13), которая есть обобщение формулы (8.9). Третье слагаемое в правой части (8.12) дает изменение концентрации вследствие диффузии, а последнее — вследствие рекомбинации. Уравнение (8.12) есть уравнение непрерывности в *амбиполярной форме*, которое в ряде случаев удобнее уравнений (3.3). В частности, так как мы уже учли амбиполярное электрическое поле введением амбиполярного коэффициента диффузии D, то под $\mathcal{E}$ в формуле (8.13) нужно понимать только поле, создаваемое внешними источниками.

§ 9. Длины диффузии и дрейфа

Применим теперь уравнение непрерывности (8.12) к исследованию пространственного распределения избыточных носителей.

Рассмотрим нитевидный образец полупроводника, для определенности — n-типа, в одной из частей которого (при $x<0$, рис. 7.18) генерируются электроны и дырки (например, светом), и найдем стационарное распределение концентрации избыточных носителей в области $x>0$, где генерации нет. Избыточные концентрации будем считать малыми и, соответственно, D, μ и $\mathcal{E}$ не зависящими от координаты. Тогда, полагая в уравнении (8.12) $v=\mu\mathcal{E}$ и $\partial p/\partial t=0$, получаем для определения δp уравнение

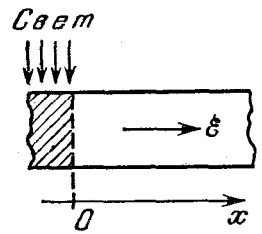


Рис. 7.18. К вычислению длины диффузии.

$$(\delta p)''-\frac{2}{\lambda}(\delta p)'-\frac{\delta p}{L^2}=0, \qquad (9.1)$$

где штрих обозначает дифференцирование по x, а через λ и L обозначены некоторые характерные длины:

$$\lambda=\frac{2D}{\mu\mathcal{E}}, \qquad L=\sqrt{D\tau}. \qquad (9.2)$$

Граничные условия задачи имеют вид:

$$x=0: \quad \delta p=(\delta p)_0; \qquad x\to\infty: \quad \delta p\to 0.$$

Решение уравнения (9.1) есть

$$\delta p=a\exp k_1x+b\exp k_2x,$$

где

$$k_{1,2}=\frac{1}{\lambda}\pm\sqrt{\frac{1}{\lambda^2}+\frac{1}{L^2}}=\frac{1}{\lambda}\left(1\pm\sqrt{1+\frac{\lambda^2}{L^2}}\right).$$

Положим, что $\mathscr{E} > 0$ и, соответственно, $\lambda > 0$. Это соответствует такому направлению поля, которое затягивает в образец *неосновные* носители. Тогда $k_1 > 0$, $k_2 < 0$ и граничные условия дают $a = 0$, $b = (\delta p)_0$. Следовательно, δp убывает с увеличением x по экспоненциальному закону:

$$\delta p = (\delta p)_0 \exp\left(-\frac{x}{l}\right), \tag{9.3}$$

где $l = l(\mathscr{E})$ есть длина, на которой δp уменьшается в e раз. Она равна

$$\frac{1}{l} = -k_2 = \frac{1}{\lambda}\left(\sqrt{1+\frac{\lambda^2}{L^2}} - 1\right). \tag{9.4}$$

Если электрическое поле очень слабое ($\lambda^2/L^2 \gg 1$), то формула (9.4) дает

$$l = L = \sqrt{D\tau}. \tag{9.5}$$

Здесь L, по определению, есть *длина диффузии* неравновесных носителей (ср. § 7).

В противоположном случае достаточно сильного поля ($\lambda^2/L^2 \ll 1$) можно положить $\sqrt{1+\lambda^2/L^2} \simeq 1 + \lambda^2/2L^2$ и из формулы (9.4) получается

$$l(\mathscr{E}) \simeq \frac{2L^2}{\lambda} = \mu\mathscr{E}\tau. \tag{9.6}$$

В сильных электрических полях длина затягивания равна расстоянию, которое проходит пакет инжектированных носителей за среднее время их жизни (*длина дрейфа*).

Рассмотрим теперь случай $\mathscr{E} < 0$ ($\lambda < 0$), т. е. такого направления поля, которое соответствовало бы затягиванию *основных* носителей. Тогда $k_1 < 0$, $k_2 > 0$ и граничные условия дают $b = 0$, $a = (\delta p)_0$. И в этом случае δp экспоненциально уменьшается с увеличением x, но длина затягивания оказывается равной

$$\frac{1}{l} = -k_1 = \frac{1}{|\lambda|}\left(1 + \sqrt{1+\lambda^2/L^2}\right). \tag{9.7}$$

Отсюда видно, что $l < L$, т. е. электрическое поле *препятствует* распространению избыточных носителей. В сильных электрических полях

$$l \simeq \frac{2}{|\lambda|} = \frac{D}{\mu\mathscr{E}} \tag{9.8}$$

и при $\mathscr{E} \to \infty$, $l \to 0$.

Полученные результаты показывают, что вследствие амбиполярности процессов диффузии и дрейфа пакет инжектированных носителей движется во внешнем электрическом поле так, как

двигались бы *неосновные* носители, и объясняют типичные опыты, рассмотренные в § 7 *).

Отметим, что граничную концентрацию $(\delta p)_0$, значение которой здесь для нас было несущественно, мы считали заданной. Эта концентрация зависит от интенсивности генерации g в освещенной области и от напряженности электрического поля. Ее значение можно найти, рассматривая решение уравнения непрерывности в освещенной области (при $g \neq 0$) и сшивая его с рассмотренным решением для неосвещенной области ($g = 0$), на чем, однако, мы не будем останавливаться.

В предыдущих рассуждениях мы пренебрегали зарядами, связанными на ловушках, и соответственно считали $\delta p = \delta n$ и $\tau_p = \tau_n$. Однако все приведенные рассуждения могут быть обобщены и для более сложного случая $\delta p \neq \delta n$ (см., например, [3]). При этом основные принципиальные выводы не изменяются.

В § 5 мы говорили, что введение квазиуровней Ферми нуждается в специальном экспериментальном обосновании. Исследования диффузии и дрейфа неравновесных электронов и дырок дают такую возможность. Действительно, в § VI.2 мы видели, что соотношение Эйнштейна

$$\frac{D}{\mu} = \frac{kT}{e},$$

связывающее коэффициент диффузии D и подвижность μ, справедливо только в том случае, когда рассматриваемые частицы подчиняются распределению Больцмана. С другой стороны, если возможно введение квазиуровней Ферми, то и для неравновесных носителей заряда в невырожденных полупроводниках, согласно формуле (5.1а), тоже должно быть справедливо распределение Больцмана, а следовательно, и для них должно выполняться соотношение Эйнштейна.

Отношение D/μ для неравновесных носителей заряда можно найти из опыта, измеряя длину диффузии $L = \sqrt{D\tau}$ и длину дрейфа $l(\mathscr{E}) = \mu \mathscr{E} \tau$:

$$\frac{D}{\mu} = \frac{L^2 \mathscr{E}}{l(\mathscr{E})}.$$

При этом, согласно § 8, в примесных полупроводниках D и μ соответствуют неосновным носителям. Поэтому можно проверить экспериментально, выполняется ли для неравновесных носителей заряда соотношение Эйнштейна, а значит, и распределение Больцмана. Такие измерения производились для германия многочислен-

*) Особенности диффузии и дрейфа квазинейтральных пакетов неравновесных носителей заряда, лежащие в основе действия многих полупроводниковых приборов, были открыты независимо друг от друга В. Е. Лашкаревым в СССР (при исследовании фотопроводимости в закиси меди) и Дж. Бардином и В. Браттейном в США (в явлениях инжекции в германии) в конце 40-х годов.

ными авторами. Они показали, что формула Эйнштейна хорошо выполняется и для неравновесных носителей, что является экспериментальным подтверждением возможности введения квазиуровней Ферми.

§ 10. $n^+ - n$- и $p^+ - p$-переходы

При инжекции в p—n-переходах концентрация неравновесных носителей на границе $(\delta p)_0 > 0$ и полупроводник обогащается электронами и дырками. Однако нарушение равновесных концентраций носителей при наличии тока через контакт может быть и таким, что $(\delta p)_0 < 0$.

Рассмотрим контакт двух полупроводников одинакового типа проводимости, но различающихся по степени легирования примесями. Такие контакты получили название n^+—n- и, соответственно,

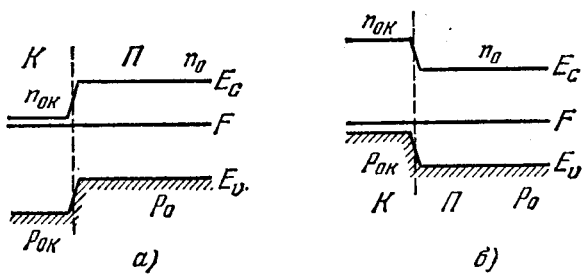


Рис. 7.19. Контакты $n^+ - n$ (*а*) и $p^+ - p$ (*б*).

p^+—p-контактов. Их энергетическая диаграмма в отсутствие тока показана на рис. 7.19. Такие переходы иногда образуются и при изготовлении контакта полупроводника с металлами за счет диффузии металла (например, при сплавлении), термической обработки и т. п. Будем условно называть одну из областей «контактным электродом», а другую — «полупроводником» и будем говорить, для определенности, о переходе n^+—n. Участие неосновных носителей (дырок) в образовании тока различно в обеих областях. Его можно охарактеризовать коэффициентами ξ (в полупроводнике) и ξ_k (в контактном электроде):

$$\xi = \frac{j_p}{j} = \frac{\mu_p p_0}{\mu_0 p_0 + \mu_n n_0}, \qquad \xi_k = \frac{j_{pk}}{j} = \frac{\mu_p p_{0k}}{\mu_p p_{0k} + \mu_n n_{0k}}. \tag{10.1}$$

Решающую роль для знака $(\delta p)_0$ играет разность между «коэффициентом инжекции» ξ_k и коэффициентом ξ. Если $\xi_k > \xi$, то $(\delta p)_0 > 0$. Если при этом внешнее напряжение положительно (минус на полупроводнике), то обогащенная область будет затягиваться в глубь полупроводника. При наложении прямоугольного импульса напряжения эта область будет перемещаться со скоростью $\mu \mathscr{E}$. Через

время порядка пролетного времени $d/\mu\mathscr{E}$, где d — длина образца, установится стационарное экспоненциальное распределение концентрации с длиной затягивания $l(\mathscr{E})$, определяемой формулой (9.4). Мы получим инжекцию неосновных носителей, которая принципиально не отличается от инжекции в p—n-переходах.

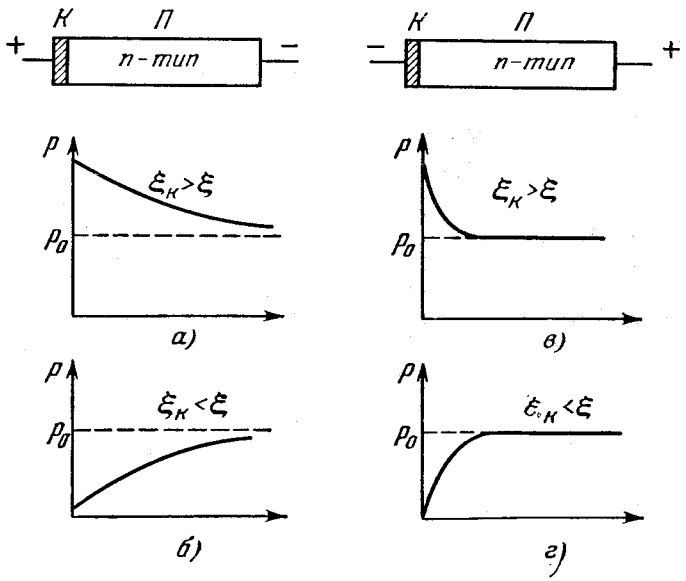


Рис. 7.20. Инжекция (*а*), эксклюзия (*б*), аккумуляция (*в*) и экстракция (*г*) носителей заряда.

При $\xi_k > \xi$, но отрицательном внешнем напряжении, образуется также обогащенная область. Она, однако, будет локализована в тонком слое у электрода, с толщиной, определяемой формулой (9.7). Это явление известно как *аккумуляция* неосновных носителей.

Если $\xi_k < \xi$, то возникает обедненная область с $(\delta p)_0 < 0$. И здесь возможны два случая. Если внешнее напряжение положительно, то обедненная область затягивается в полупроводник и распределение «дефицита» носителей δp определяется той же длиной затягивания (9.4), что и при инжекции. При длине дрейфа $\gtrsim$ длины образца обедненная область охватывает весь образец. Этот тип нарушения равновесия получил название *эксклюзии* неосновных носителей.

Однако, в отличие от инжекции, изменение электропроводности образца при эксклюзии ограничено. Действительно, в равновесии электропроводность

$$\sigma_0 = ep_0\mu_p + en_0\mu_n.$$

В предельном случае эксклюзии мы имеем $p_{\min} = 0$. Однако $n_{\min} \neq 0$, так как часть электронов не может быть удалена из образца без нарушения компенсации заряженных примесей и электрической нейтральности. Для случая полностью ионизованных доноров и акцепторов

$$p + N_d = n + N_a.$$

Поэтому

$$\sigma_{\min} = e\mu_n n_{\min} = e\mu_n (N_d - N_a). \tag{10.2}$$

Отсюда

$$\frac{(\Delta\sigma)_{\max}}{\sigma_0} = -\frac{p_0 (b+1)}{p_0 + bn_0}, \tag{10.3}$$

где $b = \mu_n/\mu_p$.

И, наконец, при $\xi_k < \xi$, но отрицательном внешнем напряжении возникает обедненный слой, локализованный у контакта. Его характерная длина, как и в случае аккумуляции, выражается формулой (9.7). Этот тип нарушения равновесия известен как *экстракция* неосновных носителей. На рис. 7.20 показаны для сравнения все четыре типа возможных нарушений равновесия по концентрации у контакта.

Глава VIII

ВЫПРЯМЛЕНИЕ И УСИЛЕНИЕ ПЕРЕМЕННЫХ ТОКОВ С ПОМОЩЬЮ $p-n$-ПЕРЕХОДОВ

§ 1. Статическая вольтамперная характеристика $p-n$-перехода

Вследствие инжекции неосновных носителей вольтамперная характеристика p—n-перехода становится очень нелинейной. При положительном внешнем напряжении («+» источника на p-области) дырки из p-области инжектируются в n-область и становятся там неосновными и неравновесными носителями. Они интенсивно рекомбинируют с электронами, концентрация которых в n-области велика, вследствие чего новые дырки могут легко входить из p- в n-область. То же самое имеет место для электронов, инжектированных из n- в p-область. Эти токи рекомбинации могут быть велики даже при малых внешних напряжениях, так как они не сопровождаются появлением заметного объемного заряда: заряд инжектированных дырок легко компенсируется перераспределением электронов в n-области, а заряд электронов — перераспределением дырок в p-области. Для такого направления тока сопротивление перехода мало (*проходное*, или *прямое*, направление). Напротив, при отрицательном напряжении через переход поступают только неосновные носители: электроны — из p- в n-область и дырки — из n- в p-область. Образуемый ими ток очень мал, и поэтому сопротивление перехода оказывается большим (*запорное*, или *обратное*, направление тока).

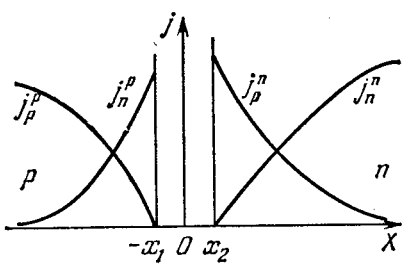


Рис. 8.1. Распределение дырочных и электронных токов в $p-n$-переходе.

Распределение плотностей тока дырок и электронов в обеих областях показано на рис. 8.1, где различные токи отмечены двумя индексами: нижний индекс показывает, какими частицами образован ток, верхний индекс — к какой из областей перехода (p или n) он относится. Ток j_p^n (образованный дырками, инжектированными в n-область) максимален на границе n-области x_2 и затухает по мере удаления от границы вследствие рекомбинации дырок. При этом

в n-области возникает еще и ток электронов j_n^n, так как для поддержания рекомбинации дырок электроны должны все время поступать из электрода в n-область. Он минимален на границе x_2 и увеличивается с удалением от границы. Суммарный ток, возникающий от инжекции дырок в n-область, есть

$$j^{(p)} = j_p^n(x) + j_n^n(x) = \text{const}.$$

Аналогично, в p-области появляются ток инжектированных электронов j_n^p и ток дырок j_p^p, а суммарный ток от инжекции электронов в p-область равен

$$j^{(n)} = j_n^p(x) + j_p^p(x) = \text{const}.$$

Если рекомбинация в самом переходном слое пренебрежимо мала, то $j_n^n(x_2) = j_p^p(-x_1) = 0$ (как показано на рис. 8.1). Если она значительна, то

$$j_n^n(x_2) + j_p^p(-x_1) = j_r$$

есть ток от рекомбинации внутри переходного слоя. Он равен числу рекомбинирующих электронов и дырок в единицу времени внутри переходного слоя с толщиной $d = x_1 + x_2$ и единичной площадью, умноженному на элементарный заряд e.

Из сказанного видно, что полную плотность тока через переход можно представить в виде

$$j = j^{(p)} + j^{(n)} = j_p^n(x_2) + j_n^p(-x_1) + i_r. \tag{1.1}$$

Поэтому определение вольтамперной характеристики сводится к вычислению токов неосновных носителей только на границах перехода и тока рекомбинации j_r в самом переходном слое.

Расчет особенно прост при следующих условиях:

1) Малый уровень инжекции. При этом, в частности, время жизни τ_p и длину диффузии L_p дырок в n-области и, соответственно, τ_n и L_n в p-области можно считать постоянными.

2) Длины n- и p-областей больше длин диффузии L_p, L_n, так что инжектированные носители успевают рекомбинировать, не доходя до металлических электродов.

3) Толщина перехода $d \ll L_p$, L_n, так что в формуле (1.1) можно пренебречь током рекомбинации j_r в самом переходном слое.

Если токи через переход не слишком велики (оценка дана ниже), то при вычислении $j_p(x_2)$ и $j_n(-x_1)$ можно пренебречь дрейфом по сравнению с диффузией. Действительно, рассмотрим, например, распределение дырок в n-области. Оно выражается формулой вида (VII.9.3):

$$\delta p = [p(x_2) - p_n] \exp\left(-\frac{x - x_2}{l_p}\right),$$

где l_p есть длина затягивания дырок, определяемая формулой (VII.9.4). Поэтому для тока диффузии дырок в плоскости x_2 имеем

$$j_{\text{диф}}(x_2) = -eD_p \frac{d\delta p}{dx}\Big|_{x=x_2} = \frac{eD_p}{l_p}[p(x_2) - \bar{p}_n]. \tag{1.2}$$

Ток дрейфа в той же плоскости равен

$$j_{\text{др}}(x_2) = ep(x_2)\mu_p \mathscr{E}.$$

Если $p(x_2)$ значительно превышает равновесное значение p_n, то отношение обоих токов есть

$$\frac{j_{\text{диф}}}{j_{\text{др}}} \simeq \frac{D_p}{l_p \mu_p \mathscr{E}} = \frac{1}{2}\frac{\lambda_p}{l_p} = \frac{1}{2}\left(\sqrt{1+\frac{\lambda_p^2}{L_p^2}} - 1\right).$$

Отсюда видно, что $j_{\text{диф}} \gg j_{\text{др}}$, если $\lambda_p/L_p \gg 1$. При этом $l_p \simeq L_p$. Отсюда получаем условие для поля:

$$\mathscr{E} \ll \frac{D}{\mu L_p} = \frac{kT}{eL_p}. \tag{1.3}$$

Для тонких p—n-переходов концентрация $p(x_2)$ связана с напряжением на переходе u соотношением (VII.6.1):

$$p(x_2) - p_n = p_n(\exp \alpha u - 1),$$

где $\alpha = e/kT$. Поэтому из формулы (1.2) имеем

$$j_p = j_p(x_2) = \frac{eD_p p_n}{L_p}(\exp \alpha u - 1).$$

Для тока $j_n(-x_1)$ получается аналогичное выражение:

$$j_n(-x_1) \simeq \frac{eD_n n_p}{L_n}(\exp \alpha u - 1).$$

Полная плотность тока через переход равна

$$j = j_s(\exp \alpha u - 1), \tag{1.4}$$

где

$$j_s = \frac{eD_p p_n}{L_p} + \frac{eD_n n_p}{L_n}. \tag{1.5}$$

Формула (1.4) выражает вольтамперную характеристику тонкого p—n-перехода при малых напряжениях. Зависимость j от u оказывается такой же самой, как и для случая контактов металл—полупроводник (ср. формулу (VI.12.12)), и не содержит вовсе параметров p- и n-областей. Однако свойства этих областей существенно влияют на величину тока насыщения j_s.

Происхождение тока насыщения в p—n-переходах имеет простой физический смысл. При отрицательном напряжении в p—n-переходе имеется потенциальный барьер для основных носителей,

так что дырки из p-области не могут переходить в n-область, а электроны из n-области не могут переходить в p-область (рис. 8.2). Напротив, для неосновных носителей барьера нет, и поэтому все электроны, генерируемые в p-области и доходящие без рекомбинации до плоскости $-x_1$, будут затянуты в n-область, а дырки из n-области, достигающие плоскости x_2, будут затянуты в p-область. Так как длина диффузии есть среднее расстояние, проходимое носителем заряда без рекомбинации, то число электронов, переходящих через единицу поверхности перехода в единицу времени, равно числу электронов, генерируемых в единицу времени в объеме цилиндра с единичной площадью основания и длиной образующей L_n. То же относится и к дыркам в n-области. Плотность образованного ими тока есть

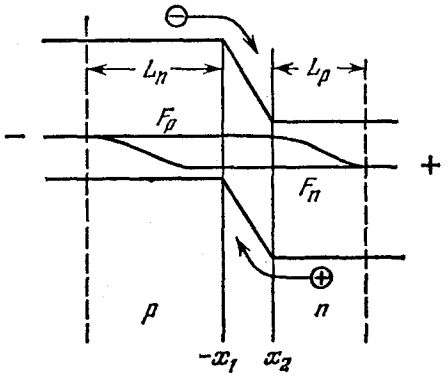


Рис. 8.2. $p-n$-переход при обратном напряжении.

$$j_s = e\,(L_n g_{nT} + L_p g_{pT}),$$

где g_{nT} и g_{pT} — темпы тепловой генерации электронов в p-области и, соответственно, дырок в n-области. Подставляя сюда

$$g_{nT} = \frac{n_p}{\tau_n}, \quad g_{pT} = \frac{p_n}{\tau_p}$$

(ср. (VII.2.4)) и учитывая, что $L_n = \sqrt{D_n \tau_n}$, $L_p = \sqrt{D_p \tau_p}$, мы получаем формулу (1.5). Мы видим, что ток насыщения в тонком p—n-переходе есть ток, обусловленный тепловой генерацией неосновных носителей заряда в p- и n-областях.

Если уровень инжекции нельзя считать малым или если ток от рекомбинации внутри перехода j_r не мал по сравнению с током насыщения j_s, то приведенные выше формулы становятся неприменимыми и должны быть заменены другими [1].

§ 2. $p-n$-переход при переменном напряжении

Положим, что к p—n-переходу приложено переменное напряжение $u(t)$, и найдем плотность тока, текущего через переход. Расчет проведем в тех же предположениях, что и в § 1. Для упрощения будем еще считать, что p-область сильно легирована ($p_p \gg n_n$), так что полный ток определяется только инжекцией дырок в n-область.

Концентрация избыточных дырок на границе n-области $x = x_2$ устанавливается за время порядка времени пролета дырок через переходный слой $t_{\text{пр}} = d/v$, которое мало. Так, если $d \sim 10^{-4}$ см и $v \sim 10^7$ см/с (тепловая скорость при 300 К), то $t_{\text{пр}} \sim 10^{-11}$ с. Поэтому, если частота переменного напряжения $\omega < 1/t_{\text{пр}}$, то можно считать, что концентрация дырок на границе следит за изменением напряжения и по-прежнему выражается формулой (VII.6.1). Следовательно,

$$\text{при } x = x_2: \quad \delta p = p_n (\exp \alpha u(t) - 1) = f(t). \tag{2.1}$$

Это соотношение дает первое граничное условие задачи. Второе граничное условие есть

$$x \to \infty: \quad \delta p \to 0. \tag{2.2}$$

Однако концентрация в глубине n-области устанавливается посредством гораздо более медленного процесса диффузии, и поэтому колебания концентрации δp, а следовательно, и плотности тока j_p отличаются от колебаний напряжения $u(t)$. Если даже $u(t)$ изменяется по чисто синусоидальному закону, то плотность тока, вообще говоря, содержит высшие гармоники.

Плотность тока при произвольной форме напряжения $u(t)$ (но при ограничениях, указанных в § 1) можно найти следующим образом. Концентрация избыточных дырок $\delta p(x, t)$, как и раньше, определяется уравнением непрерывности (VII.3.3), которое для нашего случая ($j_{\text{др}} \ll j_{\text{диф}}$) имеет вид

$$\frac{\partial \delta p}{\partial t} = D_p \frac{\partial^2 \delta p}{\partial x^2} - \frac{\delta p}{\tau_p}. \tag{2.3}$$

Его решение удобно искать в виде ряда Фурье

$$\delta p = \sum_{k=-\infty}^{\infty} c_k(x)\, e^{ik\omega t}. \tag{2.4}$$

Подставляя это в (2.3), мы получаем для определения c_k уравнения

$$\frac{d^2 c_k}{dx^2} - \frac{1 + ik\omega\tau_p}{L_p^2} c_k = 0. \tag{2.5}$$

Их решения, удовлетворяющие граничному условию (2.2), суть

$$c_k = A_k \exp\left(-\frac{x - x_2}{l_k}\right), \tag{2.6}$$

где A_k — постоянные, а

$$\frac{1}{l_k} = \frac{1}{L_p}(1 + ik\omega\tau_p)^{1/2}. \tag{2.7}$$

Постоянные A_k определяются из граничного условия (2.1):

$$\sum_{-\infty}^{\infty} A_k e^{ik\omega t} = p_n (e^{\alpha u(t)} - 1) = f(t).$$

Отсюда видно, что A_k суть коэффициенты разложения функции $f(t)$ в ряд Фурье:

$$A_k = \frac{\omega}{2\pi} \int_{-\pi/\omega}^{+\pi/\omega} f(\zeta)\, e^{-ik\omega\zeta}\, d\zeta. \tag{2.8}$$

Зная $\delta p(x, t)$, плотность тока можно найти по формуле

$$j = j_p = -eD_p \left(\frac{\partial \delta p}{\partial x}\right)_{x=x_2} = eD_p \sum_{-\infty}^{\infty} \frac{A_k}{l_k} e^{ik\omega t}. \tag{2.9}$$

Если бы инжекция электронов в p-область играла тоже заметную роль, то формула (2.9) давала бы только дырочную составляющую тока. Совершенно аналогичную формулу мы получили бы и для электронной составляющей. Полная плотность тока была бы равна сумме обеих составляющих.

Из сказанного видно, что вычисление тока сводится к нахождению коэффициентов Фурье A_k для функции $f(t)$. При этом член с $k = 0$ в сумме (2.9) дает постоянную составляющую тока

$$j_0 = \frac{eD_p}{L_p} A_0. \tag{2.10}$$

Рассмотрим теперь специальный важный случай, когда приложенное напряжение имеет постоянную составляющую u_0 и переменную составляющую $\tilde{u}$, изменяющуюся по чисто гармоническому закону

$$u(t) = u_0 + u_1 e^{i\omega t}.$$

Амплитуду u_1 будем считать малой, так что $\alpha u_1 \ll 1$. Тогда $f(t)$ можно разложить в ряд по степеням $\tilde{u}$ и ограничиться первыми двумя членами разложения:

$$f(t) = p_n [e^{\alpha u_0}(1 + \alpha u_1 e^{i\omega t} + \ldots) - 1] = p_n (e^{\alpha u_0} - 1) + p_n \alpha e^{\alpha u_0} u_1 e^{i\omega t}.$$

Подставляя это выражение в формулу (2.8), находим коэффициенты Фурье:

$$A_0 = p_n (e^{\alpha u_0} - 1), \quad A_1 = p_n \alpha e^{\alpha u_0} u_1, \quad A_k = 0 \qquad (k \neq 0, 1).$$

После этого формулы (2.9), (2.4) и (2.7) дают для плотности тока

$$j = \frac{eD_p p_n}{L_p} (e^{\alpha u_0} - 1) + \frac{eD_p p_n}{L_p} e^{\alpha u_0} (1 + i\omega\tau_p)^{1/2} u_1 \alpha e^{i\omega t}. \tag{2.11}$$

Первое слагаемое в этом выражении совпадает с формулой (1.4) и дает постоянную составляющую плотности тока j_0, вызываемого постоянным смещением u_0. Второе слагаемое выражает переменную составляющую плотности тока $\tilde{j}$. В рассматриваемом случае

малых u_1 составляющая $\tilde{j}$ изменяется гармонически с частотой ω, а высшие гармоники в первом приближении не возникают.

Полная проводимость p—n-перехода для малого переменного напряжения $\tilde{u}$ равна

$$A \equiv \frac{1}{Z} = S\frac{\tilde{j}}{\tilde{u}} = Sj_s\alpha e^{\alpha u_0}(1+i\omega\tau_p)^{1/2}. \qquad (2.12)$$

Здесь Z — импеданс перехода, S — площадь p—n-перехода, а $j_s = eD_p p_n/L_p$ есть плотность тока насыщения обратной ветви статической характеристики. Это выражение можно представить в более удобном виде, если воспользоваться тождеством

$$(1+i\omega\tau_p)^{1/2} = \frac{1}{\sqrt{2}}\left(\sqrt{\rho+1} + i\sqrt{\rho-1}\right), \quad \rho = \sqrt{1+\omega^2\tau_p^2}.$$

Тогда

$$A = \frac{S}{\sqrt{2}} j_s\alpha e^{\alpha u_0}\left(\sqrt{\rho+1} + i\sqrt{\rho-1}\right). \qquad (2.12а)$$

Так как мнимая часть проводимости $\operatorname{Im} A > 0$, то колебания $\tilde{j}$ опережают колебания $\tilde{u}$ и, следовательно, проводимость p—n-перехода имеет емкостный характер.

Из формулы (2.12а) следует, что p—n-переход при малом переменном напряжении представляет собой активное сопротивление R_D (называемое иногда «диффузионным» сопротивлением) и включенную параллельно ему емкость C_D («диффузионная» емкость) (рис. 8.3, *а*), равные

$$\frac{1}{R_D} = \operatorname{Re} A = \frac{S}{\sqrt{2}} j_s\alpha e^{\alpha u_0}\sqrt{\rho+1}, \qquad (2.13а)$$

$$C_D = \frac{1}{\omega}\operatorname{Im} A = \frac{S}{\sqrt{2}\,\omega} j_s\alpha e^{\alpha u_0}\sqrt{\rho-1}. \qquad (2.13б)$$

Физическая причина возникновения емкости C_D заключается в том, что при инжекции в n-области содержатся дополнительные (по сравнению с равновесием) дырки, а в p-области — дополнительные электроны, количество которых зависит от приложенного напряжения. Поэтому для всякого изменения напряжения Δu необходим переход определенного заряда Δq, что и обозначает наличие емкости. Аналогично, диффузионное сопротивление появляется потому, что для перераспределения носителей заряда при инжекции всегда необходимо дополнительное внешнее напряжение. При $\omega\tau_p \gg 1$ инжекция вообще не успевает осуществляться и, в соответствии с (2.13а), $C_D, R_D \to 0$.

Во всех предыдущих рассуждениях мы не учитывали нескольких важных обстоятельств. Во-первых, мы везде пренебрегали током смещения по сравнению с конвекционным током. Однако на очень

высоких частотах им пренебрегать не всегда возможно. Наличие тока смещения можно учесть, если приписать p—n-переходу дополнительную, «зарядную» емкость, уже рассмотренную нами в § VI.11. Во-вторых, мы не учитывали обычное сопротивление p- и n-областей (базы диода), обусловленное соударениями электронов и дырок с решеткой. Поэтому эквивалентная схема реального диода с $p-n$-переходом имеет вид, показанный на рис. 8.3, *б*. Здесь C — зарядная емкость, а r — активное сопротивление базы.

В-третьих, мы считали, что толщины p- и n-областей d_p и d_n достаточно велики: $d_p > L_n$, $d_n > L_p$, так что все избыточные носители заряда успевают рекомбинировать, не достигая электродов. Если это не так, то полученные формулы уже неприменимы. В этом

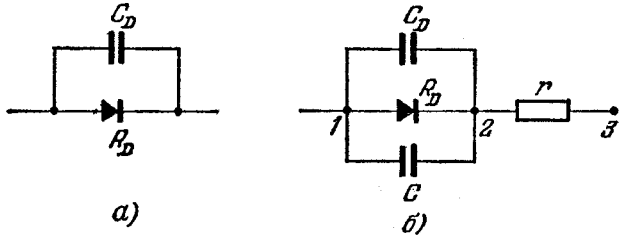


Рис. 8.3. Эквивалентная схема идеального (*а*) и реального (*б*) $p-n$-перехода при малом переменном напряжении.

случае статическая характеристика и полная проводимость на переменном токе существенно зависят от свойств электродов (от темпа рекомбинации на границе база — металлический электрод).

Наконец, в-четвертых, полученные результаты теряют силу, если амплитуда переменного напряжения не мала (нарушается неравенство $\alpha u_1 \ll 1$). Мы не будем останавливаться на этих случаях (см., например, [1]), а только отметим один интересный результат. При большом уровне инжекции, $\delta p/(n_n + p_n) \gtrsim 1$, сопротивление базы уже не остается постоянным, а существенно модулируется заливающими ее инжектированными носителями. Этот эффект может привести к тому, что колебания напряжения будут не отставать от колебаний тока (как при малых напряжениях), а их опережать. Иначе говоря, p—n-переход будет вести себя как некоторая индуктивность.

Вследствие сильной нелинейности вольтамперной характеристики p—n-переходы широко применяются для выпрямления технических переменных токов и для детектирования и преобразования частоты сигналов высокой и сверхвысокой частот в радиоприемных устройствах. Формулы (2.8) и (2.9) позволяют исследовать зависимость величины выпрямленного тока от параметров материала,

а также определить предельную рабочую частоту диода. Однако изложение этих вопросов относится уже к теории полупроводниковых приборов и выходит за рамки этой книги. Отметим лишь, что в высокочастотных диодах площадь контакта следует делать по возможности малой и уменьшать сопротивление базы.

Напротив, в мощных выпрямителях переменного тока низкой частоты площадь p—n-перехода должна быть большой для увеличения силы тока. Такие выпрямители должны еще выдерживать большие обратные напряжения без наступления явлений пробоя. Для этого оказывается выгодным увеличивать толщину переходного слоя (разумеется, до известного предела). Это реализуется, например, в структурах p-область — собственный полупроводник — n-область (p—i—n-структуры). Для создания мощных выпрямителей широко применяют также более сложные системы, содержащие несколько p—n-переходов.

§ 3. Туннельный эффект в $p-n$-переходах. Туннельные диоды

Если в p- и n-областях $p-n$-перехода увеличивать концентрацию мелких примесей, то при концентрациях электронов и, соответственно, дырок $\sim 10^{18} \div 10^{19}$ см$^{-3}$ возникают новые явления. А именно, начальная часть прямой ветви вольтамперной характеристики становится немонотонной и на ней возникает падающий участок (ток уменьшается при увеличении напряжения). В этой области напряжений дифференциальная проводимость перехода $\sigma_d = di/du$ отрицательна. Причина такой зависимости тока от напряжения заключается в том, что при увеличении концентрации носителей толщина потенциального барьера в p—n-переходе уменьшается (ср. § VI.9) и при указанных концентрациях становится сравнимой с длиной волны де Бройля ($\sim 10^{-6}$ см при комнатной температуре). Вместе с тем возрастает напряженность электрического поля в переходе. В результате становится заметной вероятность туннельных переходов электронов и дырок из одной зоны в другую (§ IV.6). Соответственно диоды с p—n-переходами рассматриваемого типа называют *туннельными*.

Происходящие в них явления качественно поясняет рис. 8.4, на котором изображены энергетические диаграммы сильно легированного p—n-перехода при различных напряжениях; p- и n-области вырождены, так что уровень Ферми в n-области $F_{(n)}$ лежит внутри зоны проводимости и удален от ее края на некоторую величину Δ_n, а уровень Ферми в p-области $F_{(p)}$ — внутри валентной зоны, ниже ее края на Δ_p (рис. 8.4, *а*). В отсутствие внешнего напряжения $F_{(n)} = F_{(p)}$ и результирующий ток равен нулю. При отрицательном внешнем напряжении (рис. 8,4, *б*, «+» на n-области) все энергии в n-области понижаются относительно p-области и электроны

из p-области туннелируют в n-область. Возникающий ток быстро увеличивается с увеличением напряжения, так как в p-области есть большое число заполненных состояний, а в n-области — пустых состояний. При небольших положительных напряжениях электроны переходят из n- в p-область. Однако в этом случае ток нарастает с напряжением лишь до некоторого предела, так как электроны из n-области не могут туннелировать внутрь запрещенной зоны в p-области. Поэтому при некотором напряжении ток достигает максимума (рис. 8.4, *в*) и затем уменьшается. Когда E_c совпадает с E'_v, туннельный ток зона — зона обращается в нуль (рис. 8.4, *г*). При дальнейшем увеличении напряжения высота потенциального барьера понижается настолько, что становятся возможными «надбарьерные» электронные переходы, и тогда появляется обычная диффузионная составляющая тока, рассмотренная ранее. Вольтамперная характеристика для этой модели показана на рис. 8.4, *д*.

Отметим, что в случае заметной роли туннельного эффекта в p—n-переходах большие токи возникают при отрицательных напряжениях и поэтому знак выпрямления оказывается обратным по сравнению с обычными «толстыми» переходами.

Расчет вольтамперной характеристики тонкого p—n-перехода очень осложняется тем, что электрическое поле внутри перехода не постоянно: оно максимально в некоторой плоскости внутри барьера и плавно убывает к его краям. Хотя распределение поля можно вычислить, количественная теория туннельного диода оказывается очень сложной и, как правило, требует численных расчетов.

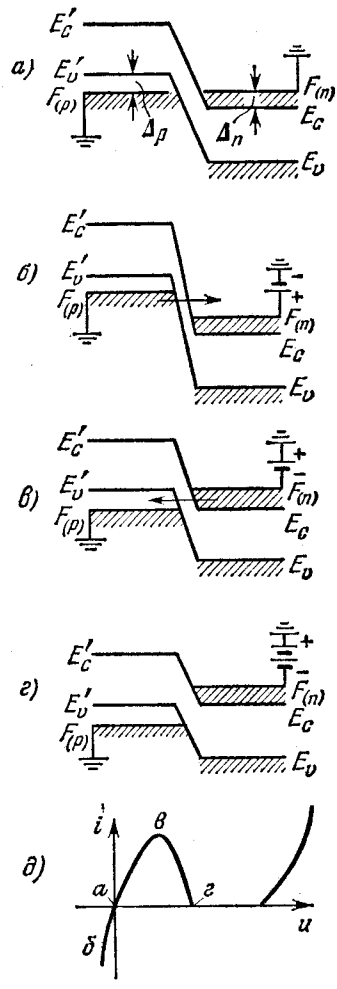


Рис. 8.4. К объяснению вольтамперной характеристики идеального туннельного диода (схематически).

Сопоставление теоретических вольтамперных характеристик туннельных диодов с экспериментальными показывает, что рассмотрен-

ная модель только приближенно описывает наблюдаемые закономерности (рис. 8.5). Наиболее важное расхождение состоит в том, что при напряжениях, отвечающих падающей ветви вольтамперной характеристики, плотность тока оказывается большей, чем предсказываемая теорией, и не обращается в нуль, когда края энергетических зон совпадают (рис. 8.4, *г*). Разность ($j_{эксп} - j_{теор}$) получила название *избыточного* тока туннельного диода.

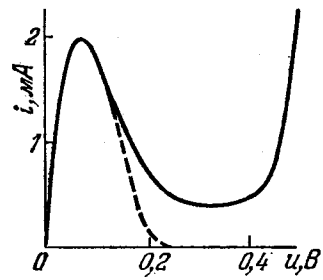


Рис. 8.5. Экспериментальная вольтамперная характеристика германиевого туннельного диода. 300 К, $\Delta_p = 6\kappa T$, $\Delta_n = 8\kappa T$. (Пунктирная кривая соответствует схеме рис. 8.4.)

Опыты показывают, что избыточный ток, при прочих равных условиях, увеличивается при увеличении концентрации примесей, создающих глубокие уровни энергии для электронов. Отсюда можно заключить, что одна из причин этого тока состоит в дополнительных туннельных переходах при участии примесей. На рис. 8.6 показаны два типа таких переходов. Так, электрон из зоны проводимости может быть сначала захвачен на пустой локальный уровень энергии E_{t1} (переход *1*), а затем с помощью изоэнергетического туннельного перехода *2* попасть в валентную зону. Другая возможность состоит в том, что электрон сначала может совершить туннельный переход из зоны проводимости на подходяще расположенный локальный уровень E_{t2} (переход *3*), а затем опуститься в валентную зону (переход *4*).

Возможны и другие, более сложные типы переходов.

Другая причина, ответственная за возникновение избыточного тока, связана с изменением самого характера энергетического спектра полупроводника при сильном его легировании. Она рассматривается в § XIX. 5.

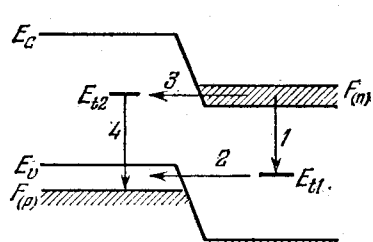


Рис. 8.6. Туннельные переходы с участием примесей.

Из принципа действия туннельных диодов видно, что процессы в них обусловлены *основными* носителями заряда, а рекомбинация (как в толстых p—n-переходах) не играет принципиальной роли. Поэтому характерное время в туннельных переходах есть не время жизни неравновесных носителей τ, а время релаксации Максвелла $\tau_M = \varepsilon/4\pi\sigma$. Последнее обычно намного меньше τ и, например, при $\sigma \sim 1$ (Ом·см)$^{-1}$ имеет порядок 10^{-12} с. Вследствие этого теоретический предел частот, до которых могут работать туннельные диоды,

$\omega_{пр} \sim 1/\tau_M$, намного больше, чем у толстых (диффузионных) p—n-переходов. Однако предельная частота реальных туннельных диодов понижается вследствие влияния паразитных емкости и индуктивности оправы диода, хотя у современных хороших диодов она достигает $\sim 10^{10}$ Гц.

Различные применения туннельных диодов основаны на использовании падающего участка их вольтамперной характеристики. При этом важно отношение токов i_{max}/i_{min} и максимальная величина отрицательной дифференциальной проводимости $|di/du|_{max}$. Чем больше эти величины и чем больше предельная частота, тем лучше качество диода.

Наличие отрицательной дифференциальной проводимости позволяет использовать туннельные диоды для усиления и генерации СВЧ электрических колебаний, для смешения частот в супергетеродинных приемных устройствах (при этом, в отличие от обычных смесителей, здесь возможна компенсация потерь на преобразование), в качестве быстродействующих переключателей для счетно-решающих устройств и для других целей.

§ 4. Биполярный полупроводниковый триод

Одно из самых важных технических применений полупроводников, которое существенно стимулировало развитие современной физики полупроводников, заключается в их использовании для усиления и генерации электрических колебаний. Приборы, предназначенные для этих целей, получили общее название *транзисторов*.

Примером очень распространенного транзистора может служить биполярный полупроводниковый триод *). Его рабочая часть состоит из пластинки полупроводника (обычно — монокристаллического), в которой путем надлежащего распределения примесей созданы два близко расположенных p—n-перехода (см. рис. 8.7). Область между обоими переходами принято называть *базой триода*, а оконечные области — *эмиттером* и, соответственно, *коллектором*. Такую систему из двух переходов можно осуществить двумя способами: создавая у эмиттера и коллектора дырочную проводимость, а у базы — электронную, т. е. структуру типа p—n—p (рис. 8.7, *а*), либо структуру типа n—p—n (рис. 8.7, *б*). Физические процессы в обоих случаях совершенно аналогичны и поэтому в дальнейшем для определенности мы будем говорить о p—n—p-структуре.

Для разъяснения принципа действия триода удобно сначала остановиться на одной из возможных схем его включения, показанной

*) Название «транзистор» произошло от соединения слогов английских слов transfer (передавать) и resistor (сопротивление). «Биполярный» указывает на то, что в данном приборе используется явление инжекции, а следовательно, необходима биполярная проводимость полупроводника.

на рис. 8.7, в которой базовый электрод является общим для обеих цепей — эмиттера и коллектора («схема с общей базой»). Токи будем считать постоянными или изменяющимися медленно. Между эмиттером и базой приложено небольшое прямое напряжение, так что дырки из эмиттера инжектируются в область базы. Если толщина базы (расстояние между границами эмиттерного и коллекторного p—n-переходов) $w < L_p$, где L_p — длина диффузии дырок в базе, то значительная часть этих дырок не успеет рекомбинировать

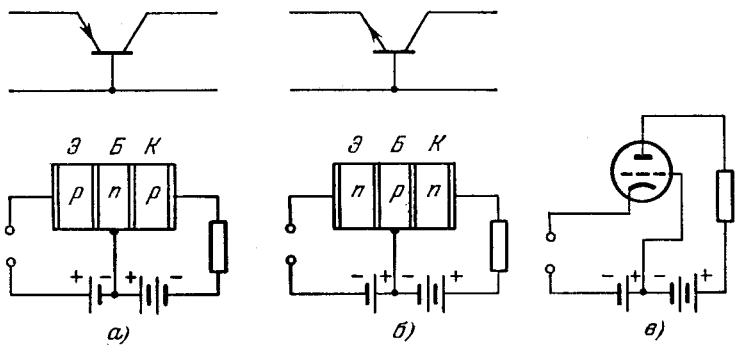


Рис. 8.7. Биполярный триод $p-n-p$ (*а*) и $n-p-n$ (*б*) типов в схеме с общей базой и аналогия с вакуумным триодом (*в*). Сверху — условное изображение на схемах.

в области базы и дойдет до коллекторного перехода. Так как к коллектору приложено большое обратное напряжение (потенциал коллектора отрицателен относительно базы), то в коллекторном переходе существует сильное электрическое поле, затягивающее дырки в коллекторную область. Поэтому все дырки, достигшие коллекторного перехода, будут входить в коллектор и участвовать в образовании тока в цепи коллектора. Обе цепи оказываются связанными друг с другом с помощью инжектированных дырок, так что всякое изменение тока в цепи эмиттера вызывает определенное изменение тока в цепи коллектора. В этом смысле полупроводниковый триод аналогичен вакуумному триоду, причем роль катода играет эмиттер, роль анода — коллектор, а роль управляющей сетки — база (рис. 8.7, *в*).

Так же как и в случае вакуумных триодов, усиление колебаний осуществляется за счет использования энергии источников постоянного напряжения, включенных в схему.

Транзисторный эффект описанного типа был открыт в 1948 г. Дж. Бардином, В. Браттейном и В. Шокли.

Важной характеристикой полупроводникового триода является коэффициент усиления тока в схеме с общей базой, который, по

определению, есть

$$\alpha = \left(\frac{\Delta i_c}{\Delta i_e}\right)_{u_c = \text{const}}. \tag{4.1}$$

Здесь Δi_c — абсолютная величина малого изменения тока коллектора, вызванного малым изменением тока эмиттера Δi_e, при условии, что напряжение коллектор — база u_c поддерживается постоянным. Этот коэффициент существенно определяет параметры триода в различных схемах его включения.

Величина α определяется свойствами эмиттерной, базовой и коллекторной областей транзистора (в первую очередь — концентрациями равновесных носителей заряда в них), а также процессами рекомбинации в области базы и в эмиттерном p—n-переходе. Действительно, полный ток эмиттера i_e, в соответствии с формулой (1.1), можно представить в виде

$$i_e = i_{pe} + i_{ne} + i_r.$$

Здесь $i_{pe} = i_p^n\ (x_2)$ есть ток дырок, вступающих из эмиттера в базу, на границе эмиттерный переход — база при $x = x_2$ (ср. рис. 8.1); $i_{ne} = i_n^p\ (-x_1)$ — ток электронов, идущих из базы в эмиттер, на границе перехода при $x = -x_1$; i_r — ток, возникающий вследствие рекомбинации, внутри эмиттерного p—n-перехода. Так как к коллектору движутся только дырки, то составляющие тока i_{ne} и i_r не изменяют тока коллектора и являются бесполезными. Отношение

$$\xi_e = \frac{i_{pe}}{i_e} = \frac{i_{pe}}{i_{pe} + i_{ne} + i_r} \tag{4.2}$$

называется *эффективностью эмиттера*. Качество эмиттера тем лучше, чем ближе ξ_e к единице.

Вследствие рекомбинации в базовой области, дырочный ток, достигающий коллекторного перехода, $i_{pc} < i_{pe}$. Влияние рекомбинации в базе можно охарактеризовать *коэффициентом передачи* дырок

$$\beta = \frac{i_{pc}}{i_{pe}}. \tag{4.3}$$

Он тем ближе к единице, чем меньше отношение w/L_p.

И, наконец, следует учесть, что полный ток коллектора тоже имеет дырочную и электронную составляющие. Поэтому, если

$$\xi_c = \frac{i_{pc}}{i_c} = \frac{i_{pc}}{i_{pc} + i_{nc}}, \tag{4.4}$$

то полный ток коллектора $i_c = i_{pc}/\xi_c$. Величину $1/\xi_c$ иногда называют «собственным коэффициентом усиления тока коллектора». Пользуясь этими понятиями, коэффициент α можно представить в виде

$$\alpha = \frac{\Delta i_{pc}}{\Delta i_{pe}} \frac{\xi_e}{\xi_c} = \beta \frac{\xi_e}{\xi_c}. \tag{4.5}$$

Отметим, что в биполярных полупроводниковых триодах толщину базы стараются сделать по возможности малой по сравнению с длиной диффузии, так что при положительном потенциале эмиттера концентрация дырок в базе бывает намного больше равновесной их концентрации. Поэтому обычно $i_{pc} \gg i_{nc}$ и, соответственно, можно положить $\xi_c \simeq 1$.

Так как $\xi_e < 1$ и $\beta < 1$, то и $\alpha < 1$. Таким образом, в рассматриваемой схеме с общей базой получить усиление тока нельзя. Однако изменения напряжения на нагрузочном сопротивлении в цепи коллектора $\Delta u_c = R_c \Delta i_c$ при большой величине R_c могут быть значительно больше изменений напряжения эмиттера Δu_e, т. е. усиление по напряжению может быть значительно больше единицы.

Через базовый электрод триода протекает также некоторый ток i_b, равный

$$i_b = i_e - i_c. \tag{4.6}$$

Этот ток обусловлен электронными составляющими токов i_{ne} и i_{nc}, а также рекомбинацией дырок и электронов в базе, так как для

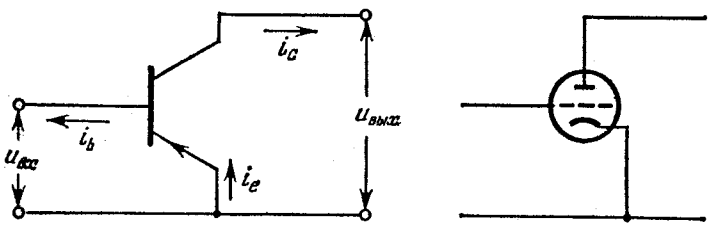


Рис. 8.8. Полупроводниковый триод в схеме с общим эмиттером.

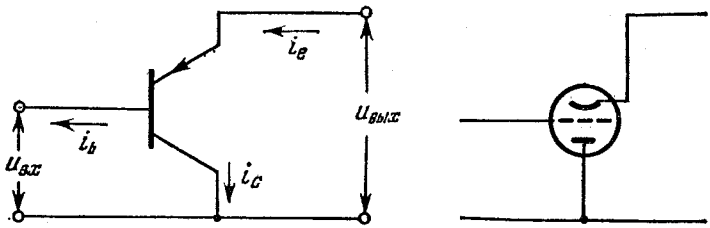


Рис. 8.9. Полупроводниковый триод в схеме с общим коллектором.

поддержания этой рекомбинации необходим постоянный приток электронов в базовую область. При $\xi_e = \xi_c = \beta = 1$ мы имели бы $i_b = 0$. В реальных триодах $i_b \neq 0$, однако $i_b \ll i_e$, i_c.

На рис. 8.8 и рис. 8.9 показаны две другие возможные схемы включения триода.

В схеме с общим эмиттером (рис. 8.8) управляющим током является ток базы i_b, а током в выходной цепи — ток i_c. Соответственно

коэффициент усиления тока в этой схеме α_e определяется как

$$\alpha_e = \left(\frac{\Delta i_c}{\Delta i_b}\right)_{u_{\text{вых}}=\text{const}}. \tag{4.7}$$

Для схемы с общим коллектором (рис. 8.9) коэффициент усиления тока есть

$$\alpha_c = \left(\frac{\Delta i_e}{\Delta i_b}\right)_{u_{\text{вых}}=\text{const}}. \tag{4.8}$$

В отличие от схемы с общей базой, усиление по току в двух последних схемах может быть значительно больше единицы. Это происходит потому, что управляющий ток в этих схемах i_b намного меньше выходного тока i_c или, соответственно, i_e. Поэтому значительным изменениям выходного тока соответствуют только малые изменения управляющего тока.

Количественная теория биполярного триода получается при использовании основных уравнений, приведенных в § VII.3 (уравнений непрерывности и уравнения Пуассона), и условий на границах каждого из p—n-переходов. Для тонких переходов эти граничные условия имеют вид (VII.6.1). Отсюда можно найти распределение дырок (электронов) в области базы и вычислить токи i_e, i_c как функции напряжений на эмиттере и коллекторе u_e, u_c относительно базы. А это позволяет определить все параметры полупроводникового триода как активного четырехполюсника. Однако мы не будем останавливаться на этих довольно длинных расчетах, которые можно найти в руководствах по полупроводниковым приборам [1, 2].

Рассмотренный выше качественно биполярный триод представляет собой, конечно, лишь один из типов полупроводниковых усилителей, основанных на использовании явления инжекции. Имеются и другие полупроводниковые устройства, предназначенные для тех же целей. Некоторые из них основаны на других принципах и не используют явление инжекции вовсе. Помимо рассмотренных выше туннельных диодов, сюда относятся так называемые полевые транзисторы (в которых для управления током в выходной цепи используют модуляцию толщины слоя объемного заряда с помощью входного напряжения) и другие приборы.

§ 5. Гетеропереходы

До сих пор мы рассматривали p—n-переходы, образованные в одном и том же монокристалле путем соответствующего распределения примесей. В таких переходах (часто называемых гомопереходами) по обе стороны от границы раздела p- и n-областей мы имели одну и ту же решетку кристалла с одинаковыми характеристиками (шириной запрещенной зоны, электронным сродством и т. п.). Сейчас мы коротко остановимся на контактах двух различных полупроводников (гетеропереходах).

Гетеропереходы можно получить, наращивая монокристальный слой одного из полупроводников на монокристальной же подложке другого полупроводника с помощью специальных методов. Такое наращивание без существенного нарушения монокристальной структуры возможно, разумеется, не для всякой пары полупроводников, так как для этого необходимо определенное соответствие между кристаллическими решетками. Гетеропереходы можно создать, используя пары полупроводников Ge—GaAs, GaAs—$Ga_xAl_{1-x}As$, GaAs—$GaAs_xP_{1-x}$, CdTe—CdSe и др. В зависимости от содержащихся примесей оба полупроводника могут иметь как одинаковый тип проводимости («изотипные гетеропереходы», например, структуры n—n^+, p—p^+ и т. д.), так и разный («анизотипные» переходы p—n, p—n^+ и др.).

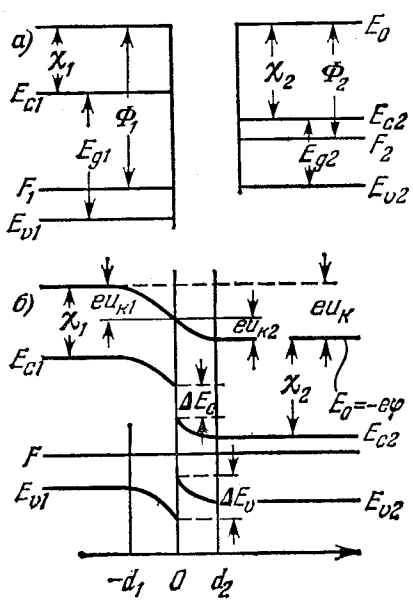


Рис. 8.10. Энергетическая диаграмма двух различных полупроводников до контакта (*а*) и $p-n$-гетероперехода (*б*). E_g — ширина запрещенной зоны, χ — электронное сродство, Φ — работа выхода, u_k — контактная разность потенциалов, E_0 — энергия покоящегося электрона в вакууме.

Энергетические диаграммы гетеропереходов, описывающие изгиб энергетических зон и возникающие потенциальные барьеры, имеют особенности по сравнению с таковыми для гомопереходов. Эти диаграммы можно построить следующим образом.

Рассмотрим, для определенности, анизотипный переход, образованный широкозонным полупроводником p-типа и узкозонным n-типа. Энергетическая диаграмма обоих полупроводников до образования перехода показана на рис. 8.10, *а*. После создания гетероперехода получается энергетическая диаграмма, изображенная на рис. 8.10, *б*. В отсутствие тока, как всегда, уровень Ферми в обоих полупроводниках становится одинаковым и между ними возникает контактная разность потенциалов $u_k = (\Phi_1 - \Phi_2)/e$. Уровень энергии в вакууме теперь изображается кривой $E_0 = -e\varphi(x)$, где $\varphi(x)$ — электростатический потенциал, создаваемый слоями объемного заряда у границы. Откладывая от уровня E_0 вниз отрезки χ_1 и, соответственно, χ_2, мы получим энергию дна зоны проводимости в обоих полупроводниках. Так как χ_1 и χ_2 в общем случае различны, то на границе перехода $x = 0$, в отличие от го-

моперехода, возникает разрыв в зоне проводимости $\Delta E_c = E_c(+0) - E_c(-0)$. Аналогично, откладывая в левой и правой частях диаграммы отрезки E_{g1} и, соответственно, E_{g2} от уровня $E_c(x)$, мы найдем края дырочных зон $E_v(x)$. И здесь в плоскости $x = 0$ образуется разрыв $\Delta E_v = E_v(+0) - E_v(-0)$. В зависимости от соотношения между электронным сродством χ_1 и χ_2, с одной стороны, и шириной запрещенных зон E_{g1} и E_{g2}, с другой, эти разрывы могут иметь либо вид «стенки» (ΔE_c на рис. 8.10, *б*), либо вид «крюка» (ΔE_v на рис. 8.10, *б*).

На рис. 8.11 показан другой пример — изотипного n—n-гетероперехода (тоже в равновесии). Здесь «крюк» возникает в зоне проводимости, а «стенка» — в валентной зоне. При этом уровень Ферми в области разрыва ΔE_c попадает в зону проводимости, так что в этом слое электронный газ оказывается вырожденным.

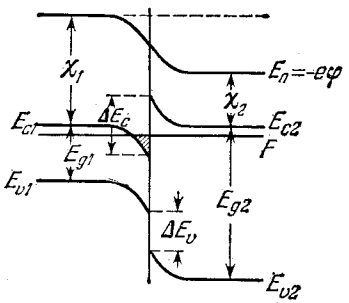


Рис. 8.11. Пример энергетической диаграммы изотипного n—n-гетероперехода. Заштрихованная часть вырождена.

Ход краев зон $E_c(x)$ и $E_v(x)$ в областях объемного заряда, а следовательно, и разрывы зон ΔE_c и ΔE_v можно определить из следующих соображений. Полная контактная разность u_k распределяется между обоими полупроводниками на части u_{k1} и u_{k2} (рис. 8.10). Если известны примеси в обоих полупроводниках (их концентрации и энергетические уровни), то можно вычислить объемный заряд $\rho(\varphi)$ как функцию потенциала φ и затем, интегрируя уравнение Пуассона, найти пространственное распределение потенциала $\varphi_1(x, u_{k1})$ и $\varphi_2(x, u_{k2})$ в каждом из полупроводников. Эти распределения будут зависеть также от u_{k1} и, соответственно, u_{k2}, входящих через граничные условия. Тогда из условия непрерывности нормальной составляющей электрической индукции на границе раздела

$$\varepsilon_1 \frac{d\varphi_1}{dx}\bigg|_{x=0} = \varepsilon_2 \frac{d\varphi_2}{dx}\bigg|_{x=0} \qquad (5.1)$$

можно найти u_{k1} и u_{k2}, а следовательно, и распределение потенциала, и изгиб энергетических зон.

Поясним сказанное на примере p — n-гетероперехода, изображенного на рис. 8.10. Положим для простоты, что в полупроводниках имеются мелкие, полностью ионизованные доноры и акцепторы, и обозначим концентрации дырок и электронов в глубине полупроводников через p_0 и, соответственно, n_0. В рассматриваемом случае в обоих полупроводниках возникают обедненные слои

объемного заряда, и поэтому мы используем приближение полностью истощенного слоя. Тогда мы можем сразу воспользоваться результатами в § VI.9, однако с учетом того, что в данном случае диэлектрические проницаемости в обеих областях различны. Соответственно вместо формул (VI.9.8) мы получим

$$\begin{aligned} \varphi_1(x) &= u_1 + \frac{2\pi e p_0}{\varepsilon_1}(x+d_1)^2 \qquad (p\text{-область}), \\ \varphi_2(x) &= u_2 - \frac{2\pi e n_0}{\varepsilon_2}(x-d_2)^2 \qquad (n\text{-область}), \end{aligned} \tag{5.2}$$

где u_1 и u_2 — значения потенциала при $x=-d_1$ и $x=d_2$. Подставляя (5.2) в условие (5.1), находим

$$\frac{d_1}{d_2} = \frac{n_0}{p_0}. \tag{5.3}$$

Далее, полагая в формулах (5.2) $x=0$, имеем

$$\begin{aligned} \varphi_1(0) - u_1 &= u_{k1} = \frac{2\pi e p_0}{\varepsilon_1} d_1^2, \\ u_2 - \varphi_2(0) &= u_{k2} = \frac{2\pi e n_0}{\varepsilon_2} d_2^2. \end{aligned} \tag{5.4}$$

Отсюда, с учетом формулы (5.3), получаем

$$\frac{u_{k1}}{u_{k2}} = \frac{n_0}{p_0}\frac{\varepsilon_2}{\varepsilon_1}. \tag{5.5}$$

Полная контактная разность потенциалов равна

$$u_k = u_{k1} + u_{k2} = 2\pi e\left(\frac{n_0 d_2^2}{\varepsilon_2} + \frac{p_0 d_1^2}{\varepsilon_1}\right). \tag{5.6}$$

Из формул (5.6) и (5.3) находим толщину слоев объемного заряда:

$$\begin{aligned} d_1 &= \sqrt{\frac{\varepsilon_1\varepsilon_2 n_0 u_k}{2\pi e p_0 (n_0\varepsilon_2 + p_0\varepsilon_1)}}, \\ d_2 &= \sqrt{\frac{\varepsilon_1\varepsilon_2 p_0 u_k}{2\pi e n_0 (n_0\varepsilon_2 + p_0\varepsilon_1)}}. \end{aligned} \tag{5.7}$$

Соотношения (5.2) — (5.7) полностью определяют распределение потенциала $\varphi(x)$, изгиб энергетических зон $-e\varphi(x)$, толщину слоев объемного заряда d_1 и d_2 и отношение контактных разностей потенциалов u_{k1}/u_{k2} в обоих полупроводниках.

Рассмотренные энергетические диаграммы являются идеализированными, так как не учитывают локальных уровней энергии для электронов на границе. Такие уровни возникают во всех случаях, когда нарушается периодичность потенциала решетки. Последнее имеет место, например, на свободной поверхности кристалла (см. гл. X), а также на границе гетероперехода вследствие неодинаковости кристаллических решеток обоих полупроводников. Если

концентрация поверхностных уровней велика, то на поверхности раздела может возникнуть значительный электрический заряд и энергетический профиль гетероперехода может быть существенно искажен.

Поверхностные уровни энергии создают возможность дополнительных электронных переходов в гетеропереходе. На рис. 8.12 схематически приведен пример энергетической диаграммы n — p-гетероперехода при учете поверхностных уровней E_s. При положительном внешнем напряжении и в отсутствие поверхностных уровней электроны переходят в полупроводник *2* только над барьером (переход *1*), где рекомбинируют со свободными дырками. Дырки переходят в обратном направлении, тоже над барьером (переход *1'*), и рекомбинируют в полупроводнике *1* со свободными электронами. При наличии поверхностных уровней электроны и дырки могут перейти на эти уровни с помощью туннельного эффекта (переходы *2* и *2'*). Опускаясь затем по цепочке поверхностных уровней, электроны получают возможность дополнительной рекомбинации с дырками. Разумеется, возможны и более сложные переходы с участием поверхностных уровней.

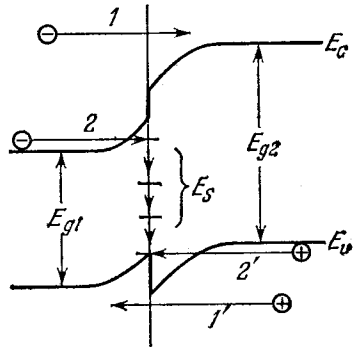


Рис. 8.12. Гетеропереход при наличии поверхностных уровней на границе.

Применение гетеропереходов в некоторых полупроводниковых приборах может оказаться более выгодным, нежели использование гомопереходов. Так, в гетеропереходах можно осуществить одностороннюю инжекцию, при которой только одна из областей гетероперехода будет обогащаться носителями заряда. Эффективность таких переходов $\xi = j_{но}/(j_{но} + j_{осн})$ можно сделать близкой к единице.

Создавая гетеропереход типа широкозонный полупроводник n-типа — тонкий слой узкозонного полупроводника — широкозонный полупроводник p-типа и прикладывая к нему большое положительное смещение, оказывается возможным легче осуществить высокий уровень инжекции в среднем слое, нежели в обычных гомопереходах. Это обстоятельство важно для создания полупроводниковых квантовых генераторов (лазеров).

Гетеропереходы позволяют создать фотоэлементы с резко ограниченной спектральной полосой чувствительности (и повысить их коэффициент полезного действия (§ XI.5)), а также и другие полупроводниковые приборы [3].

Недостатком гетеропереходов является гораздо более сложная технология их изготовления по сравнению с гомопереходами.

Глава IX

СТАТИСТИКА РЕКОМБИНАЦИИ ЭЛЕКТРОНОВ И ДЫРОК

§ 1. Различные типы процессов рекомбинации

Рассматривая неравновесные электронные состояния в гл. VII, мы ввели понятия о средних временах жизни избыточных электронов и дырок в зонах τ_n и τ_p. Однако мы не интересовались тем, от чего зависят эти времена. Между тем опыт показывает, что τ_n и τ_p могут изменяться в чрезвычайно широких пределах — от многих часов до 10^{-8} с и меньше. В настоящей главе мы рассмотрим подробнее процессы рекомбинации и генерации неравновесных электронов и дырок и выясним физические причины, влияющие на времена их жизни.

Процессы рекомбинации удобно разделить на два класса: 1) прямую рекомбинацию зона — зона и 2) рекомбинацию с участием примесей и дефектов. В первом случае свободный электрон из зоны проводимости рекомбинирует со свободной дыркой в валентной зоне в одном элементарном акте. В процессах второго класса свободные электроны рекомбинируют со связанными дырками на примесях или дефектах, а свободные дырки — со связанными электронами. Мы остановимся сначала на прямой рекомбинации зона — зона.

При переходе электрона из зоны проводимости в валентную зону всегда должны выполняться законы сохранения энергии и квазиимпульса. Поэтому если E' и $\mathbf{k}'$ — энергия и квазиволновой вектор электрона в начальном состоянии в зоне проводимости, а E и $\mathbf{k}$ — их значения в конечном состоянии в валентной зоне, то должны выполняться соотношения

$$E'(\mathbf{k}') = E(\mathbf{k}) + \Delta E, \tag{1.1}$$

$$\hbar\mathbf{k}' = \hbar\mathbf{k} + \Delta\mathbf{p}. \tag{1.2}$$

Здесь ΔE и $\Delta\mathbf{p}$ — энергия, освобождаемая в элементарном акте рекомбинации и, соответственно, изменение квазиимпульса электрона при переходе. Поскольку квазиимпульсы в начальном и конечном состояниях системы должны лежать в пределах первой зоны Бриллюэна (§ III.3), правая часть уравнения (1.2) может содержать еще слагаемое $\hbar\mathbf{b}$, которое мы не выписываем. Прямая рекомбина-

ция электрона и дырки возможна только в тех случаях, когда от электрона может быть отведена освобождающаяся энергия ΔE и обеспечено изменение квазиимпульса $\Delta \mathbf{p}$, определяемые законами сохранения (1.1) и (1.2).

Процессы, обеспечивающие выполнение законов сохранения, могут быть различны. Это может быть испускание кванта электромагнитного излучения — фотона определенной частоты ω, тогда *)

$$\Delta E = \hbar\omega, \quad |\Delta \mathbf{p}| = \frac{\hbar\omega}{c}. \tag{1.3}$$

В этом случае мы говорим об *излучательной рекомбинации*. Избыточные энергия и квазиимпульс могут быть переданы колебаниям решетки, или фононам. Известны и процессы, при которых энергия и квазиимпульс, освобождаемые при рекомбинации, передаются третьей свободной частице — электрону (в электронных полупроводниках) или дырке (в дырочных полупроводниках). Процессы такого типа получили название *ударной рекомбинации* (или *Оже-рекомбинации*). Очевидно, что вероятность таких процессов возрастает при увеличении концентрации носителей заряда, и поэтому ударная рекомбинация проявляется обычно в сильно легированных полупроводниках. Принципиально возможны и другие типы элементарных актов рекомбинации (см. § XVII.9). В случаях, когда фотоны не участвуют в обмене энергией и квазиимпульсом, говорят о *безызлучательной рекомбинации*.

Вероятность различных типов элементарных актов зависит от энергетической структуры кристалла, а также от концентрации электронов и дырок. При этом процессы разного типа могут сосуществовать.

В настоящей главе мы будем рассматривать только статистику рекомбинации и получим соотношения, справедливые при любом механизме рекомбинации. Конкретная природа рекомбинационных процессов будет определять лишь значения фигурирующих в дальнейшем вероятностей электронных переходов (эффективных сечений рекомбинации).

§ 2. Темп рекомбинации зона — зона

Рассмотрим группу состояний *1* (рис. 9.1) в зоне проводимости с энергией в интервале $(E', E' + dE')$ и группу состояний *2* в валентной зоне с энергией $(E, E + dE)$. Число переходов $1 \to 2$, рассчитанное на единицу времени и единицу объема, должно быть пропорционально количеству электронов в состояниях *1*, т. е.

*) Строго говоря, второе из соотношений (1.3) нуждается в дополнительном доказательстве, ибо $\hbar\omega/c$ есть импульс, а не квазиимпульс фотона (см. гл. XVIII).

$N_c(E')\,dE'\,f(E')$, и количеству незаполненных состояний в группе 2, которое есть $N_v(E)\,dE f_p(E)$. Поэтому число таких переходов в единице объема и в единицу времени можно записать в виде

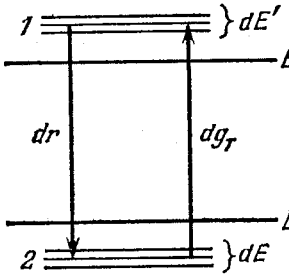


Рис. 9.1. Электронные переходы между двумя группами состояний.

$$dr = W(E',\ E)\,N_c(E')\times \\ \times N_v(E)\,f(E')\,f_p(E)\,dE'\,dE. \quad (2.1)$$

Здесь $W(E',\ E)$ — вероятность перехода, отнесенная к единице времени. В принципе ее можно вычислить методами квантовой механики (см. § XVIII.4). Чтобы найти полный темп рекомбинации, нужно просуммировать выражение (2.1) по всем энергиям E' и E.

Темп рекомбинации выражается особенно просто при выполнении двух условий: 1) существуют квазиуровни Ферми для электронов и дырок F_n и F_p и 2) полупроводник не вырожден. В этом случае (§ VII.5)

$$f(E') = \exp\frac{F_n - E'}{kT}, \qquad f_p(E) = \exp\frac{E - F_p}{kT}, \qquad np = n_0 p_0 \exp\frac{F_n - F_p}{kT},$$

где n и p — полные концентрации электронов и, соответственно, дырок, а n_0 и p_0 — их равновесные значения. Пользуясь еще соотношением (V.5.4), мы получаем

$$f(E')f_p(E) = \exp\frac{F_n - F_p}{kT}\cdot\exp\frac{E - E'}{kT} = \\ = \frac{np}{n_0 p_0}\exp\frac{E - E'}{kT} = \frac{np}{N_c N_v}\exp\left(-\frac{E' - E_c}{kT}\right)\cdot\exp\left(-\frac{E_v - E}{kT}\right). \quad (2.2)$$

Результат подстановки этого выражения в формулу (2.1) и последующего интегрирования по E' и E будет, очевидно, иметь вид

$$r = \alpha np, \quad (2.3)$$

где

$$\alpha = \frac{1}{N_c N_v}\int_{E'=E_c}^{\infty}\int_{E=-\infty}^{E_v} W(E',\ E)\,N_c(E')\,N_v(E)\times \\ \times\exp\left(-\frac{E' - E_c}{kT}\right)\cdot\exp\left(-\frac{E_v - E}{kT}\right)dE'\,dE. \quad (2.4)$$

Формула (2.3) показывает, что темп прямой рекомбинации зона — зона пропорционален произведению концентраций n и p, т. е. подчиняется тому же закону, что и скорость бимолекулярных химических реакций. Коэффициент α, по определению, есть коэффициент рекомбинации. Из формулы (2.4) видно, что он зависит

от вероятности элементарного акта рекомбинации $W(E', E)$ и распределения квантовых состояний в зонах $N_c(E')$ и $N_v(E)$, т. е. от структуры энергетических зон. Кроме того, он зависит от температуры.

В гл. VII мы определили суммарный темп рекомбинации как разность

$$R = r - g_T = \alpha np - g_T,$$

где g_T есть темп тепловой генерации электронов и дырок в условиях равновесия. Так как в равновесии (при $n = n_0$ и $p = p_0$) $R = 0$, то

$$g_T = \alpha n_0 p_0 = \alpha n_i^2. \tag{2.5}$$

Поэтому мы получаем окончательно

$$R = \alpha (np - n_0 p_0). \tag{2.6}$$

Именно эта величина должна входить во введенные ранее уравнения непрерывности.

Вероятность элементарного акта рекомбинации часто выражают при помощи эффективного сечения рекомбинации. При этом, по аналогии с кинетической теорией газов, каждый акт рекомбинации электрона с дыркой чисто формально уподобляют акту соударения электрона и частицы с некоторым сечением S. Тогда число «соударений» электронов с одной дыркой за единицу времени будет равно, очевидно, числу электронов, заключенных внутри цилиндра с основанием S и длиной образующей, равной средней тепловой скорости электронов v_T (рис. 9.2), т. е. $Sv_T n$. Полное число искомых элементарных актов с учетом не одной, а всех имеющихся дырок будет в p раз больше, т. е. $Sv_T np$. Сравнивая это с выражением (2.3), мы имеем

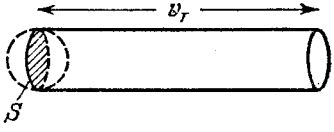


Рис. 9.2. К определению эффективного сечения рекомбинации.

$$\alpha = Sv_T, \tag{2.7}$$

где S, по определению, есть эффективное сечение рекомбинации. При этом в определении (2.7) под v_T обычно понимают скорость изолированного электрона (с массой m_0).

Обратимся теперь к времени жизни. Так как при переходах зона — зона каждое появление электрона в зоне проводимости сопровождается появлением одной дырки в валентной зоне, то $\delta n = \delta p$. Полагая, далее, $n = n_0 + \delta n$, $p = p_0 + \delta n$, мы имеем

$$R = \alpha (np - n_0 p_0) = \alpha (p_0 + n_0 + \delta n)\, \delta n.$$

Поэтому для времени жизни неравновесных пар, по определению, мы получаем

$$\tau = \frac{\delta n}{R} = \frac{1}{\alpha (n_0 + p_0 + \delta n)}. \tag{2.8}$$

Таким образом, время жизни пар зависит от концентрации равновесных носителей n_0, p_0 и концентрации неравновесных пар δn. Поэтому время жизни пар не является характеристикой данного типа полупроводника, а отражает только свойства конкретного образца при определенном его состоянии. Напротив, коэффициент рекомбинации или, соответственно, сечение рекомбинации являются характеристиками данного типа полупроводника. Однако они, конечно, зависят от температуры (ср. формулу (2.4)).

Отметим, наконец, что зависимость τ от n_0 и p_0 имеет резкий максимум. Учитывая, что равновесные концентрации n_0 и p_0 связаны соотношением $n_0 p_0 = n_i^2$, и приравнивая производную от выражения (2.8) по n_0 (или p_0) нулю, мы находим, что τ достигает максимума при $n_0 = p_0$, т. е. в собственном полупроводнике. При малом уровне возбуждения ($\delta n < n_0 + p_0$) этот максимум равен

$$\tau_{\max} = \tau_i = \frac{1}{2\alpha n_i}. \tag{2.9}$$

§ 3. Время жизни при излучательной рекомбинации

Непосредственное вычисление коэффициента рекомбинации по формуле (2.4) трудно. Однако его величину можно найти, используя принцип детального равновесия (§ VII.1) и связывая α с другими известными, макроскопическими, величинами.

Рассмотрим полупроводник в состоянии термодинамического равновесия. В нем имеются электроны с равновесной концентрацией n_0, дырки с концентрацией p_0 и равновесное (черное) излучение. Согласно принципу детального равновесия темп равновесной излучательной рекомбинации r должен быть равен темпу оптической генерации пар g_T, обусловленной поглощением равновесных фотонов в полупроводнике. Последняя величина выражается формулой (2.5). Поэтому, вычисляя g_T и зная собственную концентрацию носителей в данном полупроводнике n_i, можно найти α.

Равновесную оптическую генерацию g_T можно найти следующим образом. Пусть $\rho_0(\hbar\omega)\, d(\hbar\omega)$ есть концентрация равновесных фотонов, энергия которых лежит в интервале $[\hbar\omega, \hbar\omega + d(\hbar\omega)]$. Величина $\rho_0(\hbar\omega)$ выражается формулой Планка:

$$\rho_0(\hbar\omega) = \frac{\omega^2 n^3}{\pi^2 c^3 \hbar} \frac{1}{e^{\hbar\omega/kT} - 1}, \tag{3.1}$$

где n — показатель преломления среды.

Обозначим, далее, через $P(\hbar\omega)$ вероятность поглощения одного фотона с энергией $\hbar\omega$ в единицу времени, а через $\nu(\hbar\omega)$ — квантовый выход внутреннего фотоэффекта (число возникающих электронно-дырочных пар при поглощении одного фотона, § VII.4).

Тогда, очевидно,

$$g_T = \int_0^\infty \rho_0(\hbar\omega)\, P(\hbar\omega)\, \nu(\hbar\omega)\, d(\hbar\omega). \tag{3.2}$$

Вероятность $P(\hbar\omega)$ можно выразить через коэффициент поглощения света и показатель преломления. Действительно, описывая поглощение света как уменьшение концентрации фотонов, $\rho(\hbar\omega)$, мы можем написать

$$\frac{d\rho}{dx} = -\gamma\rho,$$

где, по определению, γ есть коэффициент поглощения света. Здесь вместо координаты x можно ввести время распространения пакета t, так как

$$x = vt = \frac{ct}{n}.$$

Поэтому можно также написать

$$\frac{d\rho}{dt} = -\frac{\gamma c}{n}\rho.$$

Отметим, что здесь под v мы должны были бы подразумевать групповую скорость пакета фотонов. Однако мы приближенно заменили ее фазовой скоростью c/n по причине, которая будет указана ниже. С другой стороны, по определению вероятности $P(\hbar\omega)$ мы имеем

$$\frac{d\rho}{dt} = -P(\hbar\omega)\,\rho.$$

Сравнивая оба соотношения, находим

$$P(\hbar\omega) = \gamma(\hbar\omega)\,\frac{c}{n}. \tag{3.3}$$

Для фактических расчетов удобно ввести безразмерную энергию фотонов

$$u = \frac{\hbar\omega}{kT}.$$

Тогда, подставляя в формулу (3.2) ρ_0 и P по формулам (3.1) и (3.3) и переходя к переменной u, получаем

$$g_T = \frac{(kT)^3}{\pi^2 c^2 \hbar^3} \int_0^\infty \frac{n^2\gamma(u)\,\nu(u)\,u^2\,du}{e^u - 1}. \tag{3.4}$$

Эта формула позволяет по известным из эксперимента спектральным зависимостям коэффициента поглощения $\gamma(\omega)$, квантового выхода $\nu(\omega)$ и значению показателя преломления n найти g_T и, следовательно, коэффициент излучательной рекомбинации α. Так как

зависимости $\gamma(\omega)$ и $\nu(\omega)$ обычно задаются в виде кривых, то такие расчеты требуют численных методов.

В предыдущих рассуждениях мы считали показатель преломления постоянным. При учете дисперсии $n(\omega)$, во-первых, изменилось бы выражение (3.3) для $P(\hbar\omega)$ и, во-вторых, вместо формулы (3.1) мы имели бы более сложное выражение для $\rho_0(\hbar\omega)$. Однако, как показывает расчет [3], входящие при этом дополнительные множители в окончательном результате взаимно сокращаются и мы опять получили бы ту же формулу (3.4), где теперь только нужно было бы считать $n = n(u)$. Учет дисперсии света вносит малую поправку, так как n изменяется с частотой очень мало по сравнению с сильной зависимостью коэффициента поглощения.

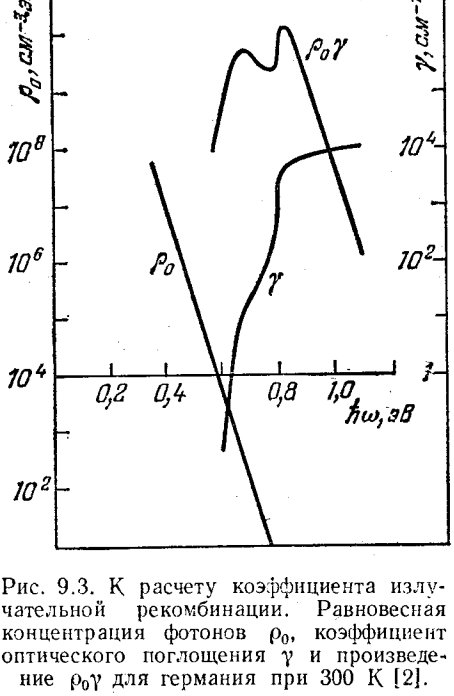


Рис. 9.3. К расчету коэффициента излучательной рекомбинации. Равновесная концентрация фотонов ρ_0, коэффициент оптического поглощения γ и произведение $\rho_0\gamma$ для германия при 300 К [2].

Отметим еще, что в значения интеграла (3.4) основной вклад вносят фотоны с энергией $\hbar\omega \sim E_g$, так как при меньших энергиях коэффициент поглощения очень мал, а при бо́льших энергиях фотонов экспонента в знаменателе становится очень большой. Поэтому для вычисления g_T нужно знать оптические постоянные полупроводника только в некотором спектральном интервале, в области края собственного поглощения.

На рис. 9.3 приведен пример такого расчета для германия при $T = 300$ К. На нем, во-первых, показана спектральная зависимость концентрации равновесных фотонов ρ_0, вычисленная по формуле (3.1). Так как в области края собственного поглощения $\exp u \gg 1$, то эта зависимость близка к экспоненциальной. Далее, приведены коэффициент поглощения γ (в см$^{-1}$) и произведение $\rho_0\gamma$. Так как для германия в области края собственного поглощения $\nu \simeq 1$, то это произведение пропорционально темпу равновесной генерации (и равному ему темпу рекомбинации) для фотонов с энергией u, т. е. dg_T/du. На последней кривой ясно выражены два максимума. Первый из них ($\hbar\omega \simeq 0{,}65$ эВ) соответствует переходам электронов из основного минимума в зоне проводимости («непрямые», или «косые», переходы), а второй — переходам из центрального минимума («прямые», или «вертикальные», переходы, ср. § XVIII.5). Полный темп генерации (рекомбинации) пропорционален площади под кривой $\rho_0\gamma$.

С учетом постоянных множителей, входящих в формулу (3.4), это дает при 300 К

$$g_T = 2{,}8 \cdot 10^{13} \text{ см}^{-3}\text{с}^{-1}.$$

Отсюда для коэффициента излучательной рекомбинации получается

$$\alpha = \frac{g_T}{n_i^2} = \frac{2{,}8 \cdot 10^{13}}{5{,}6 \cdot 10^{26}} = 0{,}50 \cdot 10^{-13} \text{ см}^3\text{с}^{-1}.$$

Так как средняя тепловая скорость электронов при 300 К имеет порядок 10^7 см/с, то эффективное сечение рекомбинации оказывается порядка 10^{-21} см^2. Максимальное время жизни пар (в собственном материале) по формуле (2.9) равно

$$\tau_i = \frac{1}{2\alpha n_i} = 0{,}43 \text{ с}.$$

Таблица 9.1

Время жизни электронно-дырочных пар при излучательной рекомбинации (300 К)

Полупроводник	Ширина запрещенной зоны E_g, эВ	Собственная концентрация n_i, см$^{-3}$	Эффективное сечение рекомбинации S, см^2	Время жизни τ_i
Si	1,1	$1{,}4 \cdot 10^{10}$	$\sim 10^{-22}$	~ 3 ч
Ge	0,65	$2{,}4 \cdot 10^{13}$	$\sim 10^{-21}$	0,43 с
InSb	0,17	$1{,}7 \cdot 10^{16}$	$\sim 10^{-18}$	$0{,}6 \cdot 10^{-6}$ с

В таблице 9.1 сопоставлены излучательные времена жизни для трех важных полупроводниковых материалов: Ge, Si (многодолинные полупроводники) и InSb (однодолинный полупроводник). Из таблицы видно, что с уменьшением ширины запрещенной зоны (увеличением n_i) τ_i катастрофически уменьшается, а, следовательно, темп излучательной рекомбинации в такой же мере увеличивается.

Излучательная рекомбинация зона — зона есть процесс, не устранимый никакими способами, и поэтому она принципиально ограничивает время жизни.

Остановимся еще на смысле времени жизни пар при излучательной рекомбинации. Согласно формуле (2.8) при данной концентрации неравновесных пар δn время жизни τ определяет темп рекомбинации R. В частности, при малом нарушении равновесия ($\delta n / (n_0 + p_0) \ll 1$) τ не зависит от δn. В этом случае при выключении внешней генерации δn должно уменьшаться по экспоненциальному закону, а τ будет определять то время, в течение которого δn уменьшается в e раз. В действительности, однако, фотоны, излучаемые при рекомбинации, могут вновь поглощаться в полупроводнике, вызывая дополнительную генерацию пар, причем

указанный процесс может повторяться многократно. В этом случае кинетика рекомбинации будет другой и более сложной. Следовательно, определенное нами выше излучательное время жизни имеет указанный простой смысл только при отсутствии вторичного образования пар. А для этого излучаемые фотоны должны достаточно быстро исчезать без образования пар (например, вследствие поглощения на свободных носителях, см. § XVIII. 2) или размеры образца должны быть достаточно малы, чтобы фотоны могли свободно выходить наружу.

Пока полупроводник находится в состоянии термодинамического равновесия, рекомбинационное излучение, конечно, наблюдать нельзя, так как наряду с излучением фотонов при рекомбинации происходит и их поглощение, и притом с таким же самым

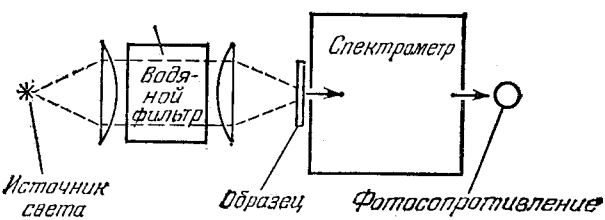


Рис. 9.4. Наблюдение рекомбинационного излучения при оптическом возбуждении.

темпом. Поэтому распределение фотонов по энергии остается равновесным. Однако если в полупроводнике каким-либо способом создать избыточные носители заряда, то темп рекомбинации будет превышать темп генерации на величину $R = \alpha (np - n_0 p_0)$ и столько же фотонов будет излучаться из каждой единицы объема в единицу времени.

На рис. 9.4 показана схема типичного опыта для наблюдения рекомбинационного излучения из германия при оптической генерации неравновесных носителей. Свет от внешнего источника (например, от лампы накаливания) проходит сначала через водяной фильтр и затем фокусируется на пластинке германия. Излучение, выходящее из германия, регистрируется спектрометром. Так как вода практически совершенно непрозрачна для длин волн $\lambda > 1{,}4$ мкм, а германий, наоборот, совершенно непрозрачен для излучения с $\lambda < 1{,}7$ мкм, то комбинация вода — германий нацело задерживает возбуждающий свет внешнего источника. Поэтому в спектрометр попадает только рекомбинационное излучение, возникающее в исследуемой пластинке германия. Чтобы это излучение не было поглощено в самой пластинке, толщина пластинки должна быть мала ($10^{-2} \div 10^{-3}$ см). На рис. 9.5 *) приве-

*) По данным работы *J. R. Haynes,* Phys. Rev. 98, 1867 (1955).

дено спектральное распределение рекомбинационного излучения из германия, на котором отчетливо видны два максимума. Один из них ($\lambda \simeq 1{,}75$ мкм) обусловлен излучательными переходами электронов (с участием фононов) из основного минимума в валентную зону. Другой максимум ($\lambda \simeq 1{,}52$ мкм) соответствует переходам центральный минимум — валентная зона (без участия фононов). Каждый из этих максимумов размыт вследствие того, что рекомбинирующие электроны в зоне проводимости, так же как и дырки в валентной зоне, имеют не одну определенную энергию, а характеризуются определенным распределением энергии вследствие их теплового движения.

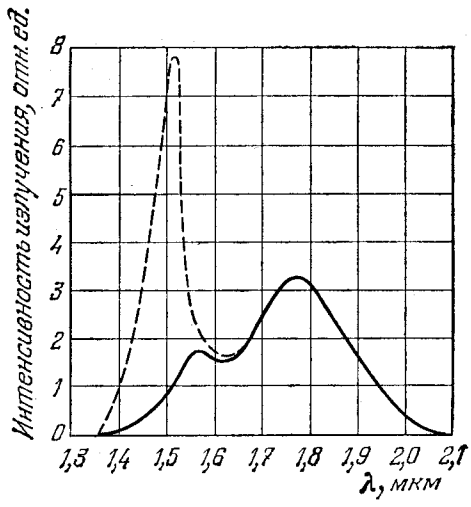


Рис. 9.5. Спектральное распределение рекомбинационного излучения из германия. Сплошная кривая — выходящее излучение. Пунктир — исправленное с учетом поглощения внутри образца.

Рекомбинационное излучение возникает также в диодах с p—n-переходами, где неравновесные электроны и дырки создаются с помощью инжекции. Это явление используют в полупроводниковых источниках света (светодиодах). Выбирая полупроводниковые материалы с различной шириной запрещенной зоны, оказывается возможным создавать источники света разного цвета.

§ 4. Рекомбинация через примеси и дефекты

Темп излучательной рекомбинации зона — зона велик в узкозонных полупроводниках. Она может быть преобладающей и в широкозонных полупроводниках с прямыми переходами (GaAs, CdS и др.). Однако очень часто в полупроводниках с широкой запрещенной зоной основную роль играют процессы рекомбинации с участием примесей и структурных дефектов. На это указывают многочисленные экспериментальные факты. Так, например, в § 3 мы видели, что если бы в кремнии излучательная рекомбинация была основным процессом, то время жизни пар должно было бы быть около 3 часов. Однако максимальные наблюдаемые времена жизни в кремнии при комнатных температурах не превышают нескольких миллисекунд. Опыт также показывает, что время

жизни чрезвычайно чувствительно даже к ничтожному содержанию некоторых примесей. Если, например, в чистейшие кристаллы германия n-типа с $\tau \sim 10^{-3}$ с ввести никель или золото (создающие глубокие локальные уровни энергии) с концентрацией $\sim 10^{15}$ см$^{-3}$ (т. е. всего 10^{-5} атомных %!), то время жизни электронов падает до $10^{-8} \div 10^{-9}$ с, т. е. уменьшается на шесть порядков.

Процесс рекомбинации с участием локальных уровней (ловушек) является двухступенчатым и состоит из захвата электрона из зоны проводимости на пустой центр рекомбинации (ловушку) и захвата дырки из валентной зоны на заполненный центр рекомбинации. При этом сами центры рекомбинации не изменяются и играют роль своеобразного катализатора процесса рекомбинации.

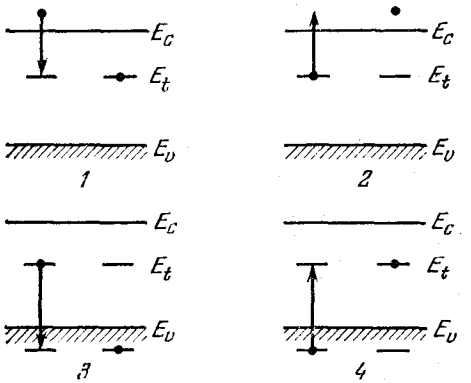


Рис. 9.6. Четыре типа элементарных процессов при рекомбинации электронов и дырок через локальные уровни.

Будем рассматривать сначала простые центры рекомбинации, которые могут захватывать или, соответственно, отдавать один-единственный электрон, а следовательно, находиться только в двух различных зарядных состояниях. Отвлекаясь от возбужденных состояний, такой центр можно охарактеризовать единственным локальным уровнем энергии для электронов E_t в запрещенной зоне (рис. 9.6). Если N_t есть полная концентрация ловушек (и равная ей концентрация локальных уровней), то в состоянии равновесия концентрация заполненных ловушек есть $N_t f_0$, а концентрация пустых равна $N_t (1 - f_0)$, где f_0 — функция Ферми. При нарушении равновесия концентрации заполненных и пустых ловушек изменяются и становятся $N_t f$ и $N_t (1 - f)$. Неравновесная вероятность заполнения ловушки f уже не выражается функцией Ферми и должна быть определена дополнительно (см. ниже).

В такой модели результирующая кинетика изменения концентрации электронов и дырок определяется четырьмя процессами: захватом электронов на ловушки (рис. 9.6, переходы *1*), тепловой эмиссией электронов с ловушек в зону проводимости (переходы *2*), захватом дырок на ловушки (переходы *3*) и эмиссией дырок с ловушек в валентную зону (переходы *4*). Найдем темпы каждого из этих переходов.

Рассмотрим группу электронов в зоне проводимости, энергия которых заключена в интервале $(E, E + dE)$. Число переходов

$(E, E + dE) \to E_t$ в единице объема и в единицу времени пропорционально, во-первых, концентрации рассматриваемых электронов $N_c(E)\,dE \cdot f(E)$ и, во-вторых, концентрации пустых ловушек $N_t[1 - f(E_t)]$. Поэтому темп захвата электронов из зоны проводимости на ловушки есть

$$r_n = N_t[1 - f(E_t)] \int_{E_c}^{\infty} c_n(E)\, N_c(E)\, f(E)\, dE. \tag{4.1}$$

Здесь коэффициент $c_n(E)$ пропорционален отнесенной к единице времени вероятности элементарного акта перехода электрона с уровня E на уровень E_t. Выражение (4.1) гораздо удобнее переписать в виде

$$r_n = \alpha_n N_t[1 - f(E_t)]\, n, \tag{4.2}$$

где n — полная концентрация электронов в зоне, а α_n есть, по определению, коэффициент захвата электронов на ловушки. Сравнивая оба выражения для r_n и учитывая еще, что

$$n = \int_{E_c}^{\infty} N_c(E)\, f(E)\, dE,$$

мы видим, что

$$\alpha_n = \frac{\int_{E_c}^{\infty} c_n(E)\, N_c(E)\, f(E)\, dE}{\int_{E_c}^{\infty} N_c(E)\, f(E)\, dE}. \tag{4.3}$$

Или, иначе,

$$\alpha_n = \langle c_n \rangle, \tag{4.3а}$$

где символ $\langle \ldots \rangle$ обозначает усреднение по всем состояниям в зоне проводимости.

Так же, как и в § 2, вместо коэффициента захвата α_n мы могли бы ввести понятие об эффективном сечении захвата электронов на ловушки S_n. Оно, по определению, связано с α_n соотношением

$$\alpha_n = S_n v_T. \tag{4.4}$$

Обратимся теперь к процессам *2* (рис. 9.6) — тепловой эмиссии электронов с ловушек в зону. Очевидно, что темп таких переходов g_{nT} пропорционален концентрации ловушек, заполненных электронами, $N_t f(E_t)$. Поэтому можно положить

$$g_{nT} = \beta_n N_t f(E_t), \tag{4.5}$$

где β_n есть некоторый коэффициент, зависящий от природы ловушек и температуры.

Коэффициент реэмиссии β_n просто связан с коэффициентом захвата α_n. Эту связь можно найти с помощью принципа детального равновесия. Действительно, в состоянии термодинамического равновесия должно быть $r_n = g_{nT}$. Это дает

$$\alpha_n N_t (1 - f_0) n_0 = \beta_n N_t f_0, \tag{4.6}$$

где f_0 — равновесная вероятность заполнения ловушек электронами. Последнюю для наших целей удобно выразить не обычной функцией Ферми, а по формуле (V.9.4) записать в виде

$$f_0 = \frac{n_0}{n_0 + n_1}.$$

Здесь n_1 зависит от энергии ионизации ловушки $E_c - E_t$ и температуры и выражается формулой (V.9.5). После этого соотношение (4.6) дает

$$\beta_n = \alpha_n \frac{1 - f_0}{f_0} n_0 = \alpha_n n_1.$$

Следовательно, коэффициент реэмиссии вполне определяется через коэффициент захвата, глубину залегания энергетического уровня ловушек и, естественно, зависит еще от температуры. Подставляя найденное выражение для β_n в формулу (4.5), получаем

$$g_{nT} = \alpha_n n_1 \cdot N_t f.$$

Суммарный темп захвата электронов на ловушки равен

$$R_n = r_n - g_{nT} = \alpha_n N_t [n (1 - f) - n_1 f]. \tag{4.7}$$

Рассуждения для дырок совершенно аналогичны. Отличие заключается в том, что темп захвата дырок из валентной зоны пропорционален концентрации ловушек, заполненных электронами (т. е. свободных от дырок), $N_t f$. Поэтому вместо выражения (4.2) мы будем иметь

$$r_p = \alpha_p N_t f p,$$

где α_p — коэффициент захвата для дырок. Соответственно темп тепловой реэмиссии дырок с ловушек будет

$$g_{pT} = \beta_p N_t (1 - f).$$

Приравнивая опять r_p и g_{pT} в случае равновесия, мы получим связь между β_p и α_p в виде

$$\beta_p = \alpha_p p_1,$$

где p_1 выражается формулой (V.9.13). Поэтому для суммарного темпа захвата дырок из валентной зоны на ловушки получаем

$$R_p = r_p - g_{pT} = \alpha_p N_t [pf - p_1 (1 - f)]. \tag{4.7a}$$

Именно эти величины, R_n и R_p, мы должны подставлять в уравнения непрерывности (VII.3.3) для учета рекомбинации. Если изменения концентраций вследствие движения частиц малы, то эти уравнения принимают вид

$$\frac{dn}{dt} = g_n - R_n = g_n - \alpha_n N_t [n(1-f) - n_1 f], \tag{4.8}$$

$$\frac{dp}{dt} = g_p - R_p = g_p - \alpha_p N_t [pf - p_1(1-f)]. \tag{4.9}$$

Здесь g_n и g_p — темпы внешней генерации электронов и, соответственно, дырок.

Так как при рекомбинации через ловушки каждый исчезающий свободный электрон оседает на ловушку (то же относится и к дыркам), то изменение концентрации заполненных ловушек $N_t f$ определяется уравнением

$$N_t \frac{df}{dt} = R_n - R_p + (g_p - g_n). \tag{4.10}$$

Кроме того, мы имеем еще уравнение квазинейтральности (ср. § VII.3)

$$p - p_0 = n - n_0 + N_t (f - f_0). \tag{4.11}$$

Уравнения (4.8) — (4.11) вполне определяют кинетику изменения концентрации неравновесных электронов и дырок вследствие рекомбинации.

Отметим, что одно из четырех указанных уравнений может быть получено как следствие из других. Так, например, дифференцируя уравнение квазинейтральности (4.11), мы имеем

$$\frac{dp}{dt} = \frac{dn}{dt} + N_t \frac{df}{dt}.$$

Подставляя сюда для каждого из слагаемых их выражения (4.8) — (4.10), мы получаем тождество. Остающиеся же три независимых уравнения определяют три искомые величины $n(t)$, $p(t)$ и $f(t)$.

Так как уравнения (4.8) и (4.9) нелинейны, то в общем случае закон изменения концентраций n и p не выражается экспонентой. Однако в специальных случаях существуют постоянные характерные времена процесса. Ниже приведены некоторые примеры.

§ 5. Нестационарные процессы

а. Монополярное возбуждение. Положим, что в полупроводнике генерируются неравновесные носители только одного знака, скажем только электроны (это имеет место, например, при примесном возбуждении фотопроводимости). Тогда мы имеем только одно дифференциальное уравнение (4.8). Полагая в нем

$$n = n_0 + \delta n, \qquad f = f_0 + \delta f$$

и учитывая, что

$$n_0(1-f_0)-n_1f_0=0,$$

находим

$$\frac{d\delta n}{dt}=g_n-\alpha_nN_t(1-f_0)\,\delta n-\alpha_nN_t(n_0+n_1+\delta n)\,\delta f.$$

Исключая отсюда δf с помощью уравнения квазинейтральности

$$\delta n+N_t\,\delta f=0,$$

получаем

$$\frac{d\delta n}{dt}=g_n-\alpha_n\,(N_t^0+n_0+n_1+\delta n)\,\delta n.$$

Здесь через $N_t^0=N_t(1-f_0)$ обозначена равновесная концентрация пустых ловушек. Отсюда видно, что мгновенное время жизни электронов (ср. § VII.2) равно

$$\tau_n=\frac{1}{\alpha_n\,(N_t^0+n_0+n_1+\delta n)}. \tag{5.1}$$

Оно само зависит от δn и изменяется в процессе релаксации концентрации электронов. Однако если неравновесная концентрация электронов не слишком велика, так что

$$\delta n\ll N_t^0+n_0+n_1,$$

то δn изменяется по экспоненциальному закону, рассмотренному нами в § VII.2, а время жизни оказывается постоянным.

В случае генерации дырок мы получили бы аналогичное выражение:

$$\tau_p=\frac{1}{\alpha_p\,(N_t^-+p_0+p_1+\delta p)}. \tag{5.2}$$

б. Биполярное возбуждение. Рассмотрим теперь более сложный случай, когда генерируются и электроны и дырки. При этом будем считать, что $g_n=g_p=g$ (как, например, при возбуждении светом в собственной полосе поглощения). Выделяя по-прежнему в величинах n, p и f их равновесные части n_0, p_0 и f_0 и неравновесные приращения δn, δp и δf, мы найдем для темпа рекомбинации электронов

$$R_n=\alpha_nN_t\,[n(1-f)-n_1f]=\alpha_nN_t\,[(1-f_0)\,\delta n-(n_0+n_1+\delta n)\,\delta f]. \tag{5.3}$$

Исключая отсюда δf с помощью уравнения квазинейтральности

$$\delta p=\delta n+N_t\,\delta f, \tag{5.4}$$

получаем

$$R_n=a\,\delta n+b\,\delta p, \tag{5.5}$$

где

$$a=\alpha_n\,(N_t^0+n_0+n_1+\delta n),\qquad b=-\,\alpha_n\,(n_0+n_1+\delta n).$$

При биполярном возбуждении темп рекомбинации электронов определяется не только избыточной концентрацией электронов δn, но еще и концентрацией избыточных дырок δp. Это происходит по той причине, что неравновесные дырки, так же как и электроны, изменяют степень заполнения ловушек, а эта последняя влияет на темп рекомбинации.

Рассуждая совершенно аналогично, мы получили бы для темпа рекомбинации дырок выражение

$$R_p = \alpha_p N_t [pf - p_1 (1-f)] = c\,\delta p + d\,\delta n, \tag{5.6}$$

где

$$c = \alpha_p (N_t^- + p_0 + p_1 + \delta p), \qquad d = -\alpha_p (p_0 + p_1 + \delta p).$$

Здесь $N_t^- = N_t f_0$ есть равновесная концентрация заполненных ловушек.

Таким образом, для определения неравновесных концентраций $\delta n\,(t)$ и $\delta p\,(t)$ мы получаем систему двух связанных дифференциальных уравнений первого порядка:

$$\frac{d\,\delta n}{dt} = g - a\,\delta n - b\,\delta p, \qquad \frac{d\,\delta p}{dt} = g - c\,\delta p - d\,\delta n. \tag{5.7}$$

Так как a, b, c и d зависят от δn и δp, то эта система нелинейна.

Уравнения кинетики становятся линейными, если выполняются условия

$$\delta n \ll n_0 + n_1, \qquad \delta p \ll p_0 + p_1. \tag{5.8}$$

Отметим, что одно из этих условий относится к основным носителям и соблюдается довольно часто. Однако условию для неосновных носителей удовлетворить гораздо труднее, и поэтому при биполярном возбуждении кинетика рекомбинации оказывается, как правило, неэкспоненциальной.

Если неравенства (5.8) выполняются, то, исключая из системы (5.7) концентрацию δp, мы получаем для δn уравнение

$$\frac{d^2(\delta n)}{dt^2} + (a+c)\frac{d(\delta n)}{dt} + (ac - bd)\,\delta n + g(b-c) = 0. \tag{5.9}$$

Для концентрации избыточных дырок получается аналогичное уравнение:

$$\frac{d^2(\delta p)}{dt^2} + (a+c)\frac{d(\delta p)}{dt} + (ac - bd)\,\delta p + g(d-a) = 0. \tag{5.10}$$

Пользуясь стандартным методом решения дифференциальных уравнений 2-го порядка с постоянными коэффициентами, мы находим, что концентрации неравновесных носителей изменяются по закону

$$\delta n\,(\delta p) = A_{n\,(p)} + B_{n\,(p)}\, e^{-t/\tau_1} + C_{n\,(p)}\, e^{-t/\tau_2}. \tag{5.11}$$

Здесь

$$A_n = g\frac{c-b}{ac-bd}, \qquad A_p = g\frac{a-d}{ac-bd},$$

а постоянные B и C определяются из начальных условий. Величины τ_1 и τ_2 дают времена релаксации избыточной концентрации и являются корнями характеристического уравнения

$$\left(\frac{1}{\tau}\right)^2 - (a+c)\frac{1}{\tau} + (ac - bd) = 0.$$

Следовательно,

$$\frac{1}{\tau_{1,2}} = \frac{1}{2}(a+c) \pm \sqrt{\frac{1}{4}(a+c)^2 - (ac-bd)}. \tag{5.12}$$

Легко убедиться, что эти корни всегда вещественны и положительны.

Таким образом, при биполярном возбуждении, даже при выполнении неравенств (5.8) (линеаризованная задача), изменение неравновесных концентраций электронов и дырок описывается двумя экспонентами и существуют два различ-

ных времени релаксации τ_1 и τ_2. Эти времена одинаковы для электронов и дырок. Из формулы (5.12) также видно, что времена релаксации неравновесных концентраций довольно сложным образом зависят от параметров ловушек — их коэффициентов захвата α_n и α_p и энергетических уровней (входящих в величины n_1 и p_1).

Соотношения, однако, упрощаются при некоторых специальных условиях. Так, например, если концентрация ловушек достаточно велика и температура кристалла низка, так что выполняются условия

$$N_t^0 \gg n_0 + n_1, \qquad N_t^- \gg p_0 + p_1, \tag{5.13}$$

то мы имеем $a \gg |b|$, $c \gg |d|$ и, следовательно, $bd \ll ac$. Тогда получается

$$\frac{1}{\tau_{1,2}} = \frac{1}{2}(a+c) \pm \frac{1}{2}(a-c),$$

т. е.

$$\frac{1}{\tau_1} = a \simeq \alpha_n N_t^0, \qquad \frac{1}{\tau_2} = c \simeq \alpha_p N_t^-. \tag{5.14}$$

В этом случае время τ_1 определяется только темпом захвата электронов, а время τ_2 — только темпом захвата дырок.

§ 6. Стационарные состояния

Остановимся теперь подробнее на стационарных (но неравновесных) состояниях, которые устанавливаются в полупроводнике через достаточное время после включения внешней постоянной генерации носителей заряда. В этом случае все производные по времени в уравнениях кинетики рекомбинации равны нулю и вычисление установившихся концентраций $(\delta n)_s$ и $(\delta p)_s$ сильно упрощается. Эти концентрации всегда можно выразить соотношениями (VII.2.6), и поэтому задача сводится к нахождению времени жизни τ_n и τ_p в стационарном состоянии. Отметим, что эти времена могут и не совпадать с временами релаксации τ_1 и τ_2 неравновесных концентраций в нестационарных процессах.

Для монополярного возбуждения результат получается непосредственно и искомые времена жизни выражаются формулами (5.1) или, соответственно, (5.2). Поэтому мы рассмотрим только биполярное возбуждение. Темпы генерации электронов и дырок будем считать одинаковыми.

Так как в стационарном состоянии концентрации заполненных и пустых ловушек не изменяются, то

$$R_n = R_p = R. \tag{6.1}$$

Это условие определяет неравновесную степень заполнения ловушек электронами f. Подставляя эту величину в формулу (4.7) или (4.7а), можно найти общий темп захвата электронов и дырок R и, пользуясь еще условием квазинейтральности, выразить его как функцию либо только концентрации неравновесных электронов $R(\delta n)$, либо только концентрации неравновесных дырок $R(\delta p)$.

Тогда искомые времена жизни определяются непосредственно как

$$\tau_n = \frac{\delta n}{R(\delta n)}, \qquad \tau_p = \frac{\delta p}{R(\delta p)}. \tag{6.2}$$

Рассмотрим некоторые важные примеры.

Положим, что уровень возбуждения неограниченно увеличивается. Тогда в выражениях (4.7) и (4.7а) для R_n и R_p можно пренебречь реэмиссией в зоны (которая ограничена, так как $f \leqslant 1$) по сравнению с захватом; кроме этого $n \simeq p$ и условие (6.1) дает

$$\alpha_n (1-f) = \alpha_p f.$$

Отсюда

$$f = \frac{\alpha_n}{\alpha_n + \alpha_p}, \qquad 1-f = \frac{\alpha_p}{\alpha_n + \alpha_p}.$$

При $\alpha_n \gg \alpha_p$ практически все ловушки становятся заполненными электронами, независимо от первоначального равновесного их заполнения, а при $\alpha_n \ll \alpha_p$ все ловушки опустошаются. Общий темп захвата электронов и дырок оказывается равным

$$R = \alpha_n N_t (1-f)\, n = \alpha_p N_t f p = N_t \frac{\alpha_n \alpha_p}{\alpha_n + \alpha_p} n,$$

а время жизни при очень высоком уровне возбуждения

$$\tau_\infty = \frac{n}{R} = \frac{1}{N_t} \frac{\alpha_n + \alpha_p}{\alpha_n \alpha_p}. \tag{6.3}$$

Для дальнейшего удобно ввести некоторые характеристические времена:

$$\tau_{n0} = \frac{1}{\alpha_n N_t}, \qquad \tau_{p0} = \frac{1}{\alpha_p N_t}, \tag{6.4}$$

смысл которых будет выяснен ниже. Тогда результат (6.3) можно представить в виде

$$\tau_\infty = \tau_{n0} + \tau_{p0}. \tag{6.3а}$$

Время жизни при высоком уровне возбуждения одинаково для электронов и дырок и не зависит от концентрации неравновесных носителей заряда. Оно определяется наибольшим из времен τ_{n0}, τ_{p0}, т. е. наименьшим коэффициентом захвата (наиболее «узким» каналом рекомбинации).

Рассмотренный случай часто реализуется в высокоомных фотопроводниках (фотосопротивлениях) с глубокими уровнями энергии ловушек. Так как в них темновые концентрации n_0 и p_0 очень малы (а также малы концентрации n_1 и p_1), то уже при умеренных интенсивностях света уровень возбуждения бывает высоким.

В общем случае условие (6.1) дает

$$\alpha_n [n(1-f) - n_1 f] = \alpha_p [pf - p_1(1-f)].$$

Отсюда получается

$$f = \frac{\alpha_n n + \alpha_p p_1}{\alpha_n (n+n_1) + \alpha_p (p+p_1)}, \quad 1-f = \frac{\alpha_n n_1 + \alpha_p p}{\alpha_n (n+n_1) + \alpha_p (p+p_1)}. \tag{6.5}$$

Подставляя эти выражения в формулу (4.7) или (4.7а), находим

$$R = \frac{np - n_0 p_0}{\tau_{n0}(p+p_1) + \tau_{p0}(n+n_1)}. \tag{6.6}$$

Рассмотрим теперь случай малой концентрации ловушек, когда можно считать избыточные концентрации $\delta n = \delta p$. Тогда, полагая $n = n_0 + \delta n$, $p = p_0 + \delta n$, находим

$$R = \frac{n_0 + p_0 + \delta n}{\tau_{n0}(p_0 + p_1 + \delta n) + \tau_{p0}(n_0 + n_1 + \delta n)} \delta n$$

и, соответственно,

$$\tau_n = \tau_p = \tau = \frac{\tau_{p0}(n_0 + n_1 + \delta n) + \tau_{n0}(p_0 + p_1 + \delta n)}{n_0 + p_0 + \delta n}. \tag{6.7}$$

При малой концентрации ловушек времена жизни электронов и дырок оказываются тоже одинаковыми. Однако теперь это общее время жизни зависит от уровня возбуждения. Если уровень возбуждения достаточно мал, так что $\delta n \ll (n_0 + n_1)$, $(p_0 + p_1)$, то τ опять перестает зависеть от δn и становится равным

$$\tau_0 = \frac{\tau_{p0}(n_0 + n_1) + \tau_{n0}(p_0 + p_1)}{n_0 + p_0}. \tag{6.8}$$

Это выражение называют формулой Шокли — Рида — Холла.

Полученные результаты могут быть легко обобщены на случай произвольной концентрации ловушек. Тогда во всех предыдущих рассуждениях мы должны были бы выразить δp через δn с помощью уравнения квазинейтральности (при вычислении τ_n) или, соответственно, выразить δn через δp (для нахождения τ_p). При этом мы получили бы более сложные формулы, которые для малого уровня возбуждения имеют вид

$$\tau_n = \frac{\tau_{p0}(n_0 + n_1) + \tau_{n0}(p_0 + p_1) + \tau_{n0} N_t \left(\frac{p_0 + p_1}{p_1}\right)^{-1}}{n_0 + p_0 + N_t \left(\frac{p_0 + p_1}{p_0}\right)^{-1} \left(\frac{p_0 + p_1}{p_1}\right)^{-1}},$$

$$\tau_p = \frac{\tau_{p0}(n_0 + n_1) + \tau_{n0}(p_0 + p_1) + \tau_{p0} N_t \left(\frac{p_0 + p_1}{p_0}\right)^{-1}}{n_0 + p_0 + N_t \left(\frac{p_0 + p_1}{p_0}\right)^{-1} \left(\frac{p_0 + p_1}{p_1}\right)^{-1}}. \tag{6.9}$$

При произвольной концентрации ловушек времена жизни электронов и дырок, даже при малом уровне возбуждения, оказываются различными.

Из сказанного видно, что влияние ловушек (примесей и дефектов структуры) на темп рекомбинации определяется их концентрацией, коэффициентами захвата α_n и α_p (входящими в τ_{n0} и τ_{p0}) и энергетическим уровнем ловушек E_t (через величины n_1 и p_1). При прочих равных условиях это влияние тем меньше, чем мельче энергетический уровень ловушки (относительно одной из зон). Это непосредственно видно, например, из формулы (6.7): при уменьшении $(E_c - E_t)$ или $(E_t - E_v)$ сильно возрастает n_1 или, соответственно, p_1, что и приводит к увеличению τ. Физически это обозначает, что большинство носителей заряда, захваченных из ближайшей зоны, сразу же выбрасываются обратно в зону. Такие примеси являются хорошими поставщиками равновесных электронов или дырок и сильно изменяют электропроводность (легирующие примеси). Напротив, примеси с глубокими уровнями энергии слабо влияют на равновесную электропроводность, но могут очень сильно изменять времена жизни (*рекомбинационные примеси*).

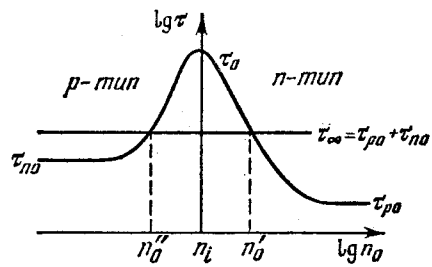


Рис. 9.7. Время жизни неравновесных электронов и дырок при бесконечно малом (τ_0) и бесконечно большом (τ_∞) уровнях возбуждения.

Однако при заданной концентрации ловушек определенного типа время жизни может быть еще очень разным, так как оно зависит также от уровня возбуждения, температуры (входящей в n_1 и p_1; кроме того, α_n и α_p тоже могут зависеть от температуры) и равновесных концентраций n_0 и p_0 (равновесной электропроводности кристалла).

Характер зависимости τ_0 от равновесной концентрации электронов показан на рис. 9.7. Так как n_0 и p_0 в невырожденных полупроводниках связаны между собой соотношением $n_0 p_0 = n_i^2$, то, как нетрудно убедиться по формуле (6.8), τ_0 достигает резкого максимума при $n_0 = n_i$, т. е. в материале с собственной проводимостью. При увеличении концентрации мелких доноров мы имеем $n_0 \gg p_0$. Если уровень ловушек лежит в верхней половине запрещенной зоны ($n_1 \gg p_1$), то формула (6.8) дает

$$\tau_0 \simeq \tau_{p0}\left(1 + \frac{n_1}{n_0}\right). \tag{6.10}$$

Если уровень расположен в нижней половине зоны, мы имеем

$$\tau_0 \simeq \tau_{p0} + \tau_{n0}\,\frac{p_1}{n_0}. \tag{6.10а}$$

В обоих случаях при увеличении степени легирования τ_0 уменьшается и стремится к значению τ_{p0}. Аналогично, в полупроводнике p-типа τ_0 уменьшается при увеличении концентрации мелких акцепторов и стремится к τ_{n0}. Таким образом, введенные нами ранее времена τ_{p0} и τ_{n0} суть предельные времена жизни в сильно легированных материалах, n- и, соответственно, p-типов.

Чтобы выяснить зависимость τ от уровня инжекции $\delta n/(n_0 + p_0)$, перепишем формулу (6.7) в виде

$$\tau = \tau_0 \frac{1 + \frac{\tau_\infty}{\tau_0} \frac{\delta n}{n_0 + p_0}}{1 + \frac{\delta n}{n_0 + p_0}}. \qquad (6.7a)$$

Отсюда видно, что с возрастанием уровня инжекции τ увеличивается, если $\tau_\infty > \tau_0$, и уменьшается при $\tau_\infty < \tau_0$. Так как τ_0 зависит от n_0 и p_0, а также от температуры, то при изменении температуры и равновесной электропроводности могут наблюдаться оба эти случая. Сказанное иллюстрирует рис. 9.7, где, помимо зависимости τ_0 (n_0), показана еще горизонталь $\tau_\infty = \tau_{p0} + \tau_{n0}$. Обе линии пересекаются в двух точках, одна из которых определяет некоторую равновесную концентрацию электронов n_0' в материале n-типа (основные носители), и концентрацию электронов n_0'' в материале p-типа (неосновные носители; соответствующая концентрация основных носителей есть $p_0'' = n_i^2/n_0''$). При этих концентрациях $\tau_0 = \tau_\infty$ и τ вообще не зависит от уровня инжекции. Если фактическая концентрация $n_0 > n_0'$ (в материале n-типа) или, соответственно, $p_0 > p_0''$ (p-тип), то τ будет увеличиваться при возрастании уровня инжекции. При $n_0 < n_0'$ или $p_0 < p_0''$ время τ будет уменьшаться.

§ 7. Многозарядные ловушки

В случае многозарядных примесных атомов (или дефектов) (ср. § II.9), создающих несколько энергетических уровней, результирующий темп рекомбинации будет равен сумме темпов рекомбинации через каждый из этих уровней. Если известны положения всех уровней и известны коэффициенты захвата электронов и, соответственно, дырок для каждого из уровней, то, рассуждая, как и выше, можно определить неравновесные степени заполнения каждого уровня и найти результирующий темп рекомбинации (а следовательно, и времена жизни электронов и дырок). Более детально путь такого расчета приведен в Приложении V.

Существенной особенностью рекомбинации через многозарядные ловушки является то, что при изменении температуры или равновесной концентрации электронов может происходить изменение зарядового состояния ловушек, что равносильно изменению

природы центров рекомбинации. Поясним это на конкретном примере атомов никеля в германии. Они создают два акцепторных уровня E_1 и E_2 (рис. 9.8) и могут, следовательно, находиться в трех зарядовых состояниях: Ni^0, Ni^- и Ni^{2-}. Положим, что в равновесии уровень Ферми (F_1 на рис. 9.8) расположен ниже уровня E_1 на несколько kT. Тогда (ср. § V.11) подавляющее большинство центров будет находиться в виде Ni^0. Это будет справедливо и при нарушении равновесия, если нарушение достаточно мало (например, при слабом освещении кристалла). Поэтому электроны будут захватываться только центрами Ni^0 (что соответствует пустому уровню E_1), а дырки — только центрами Ni^- (возникающими после захвата электрона на Ni^0), т. е. на заполненный уровень E_1, а следовательно, в рекомбинации будет участвовать только один энергетический уровень E_1. Если при этом концентрация центров мала ($\delta p \simeq \delta n$), то при увеличении равновесной концентрации основных носителей — дырок время жизни электронно-дырочных пар будет стремиться к $\tau_{n0} = (\alpha_n^0 N_t)^{-1}$, где α_n^0 — коэффициент захвата электронов нейтральными атомами никеля.

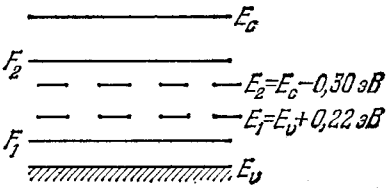


Рис. 9.8. Энергетические уровни никеля в германии.

Однако если мы будем смещать уровень Ферми вверх (например, увеличивая концентрацию мелких доноров или повышая температуру), то после пересечения последним уровня E_1 начнут появляться ионы Ni^{2-}, которые тоже будут участвовать в рекомбинации. Когда уровень Ферми окажется выше E_2 (F_2 на рис. 9.8), практически все центры будут в состоянии Ni^{2-}. Поэтому дырки будут захватываться только этими центрами (т. е. на заполненный уровень E_2), а электроны — центрами Ni^- (на пустой уровень E_2). И в этом случае рекомбинация будет происходить через один энергетический уровень, однако теперь уже через уровень E_2. При увеличении концентрации равновесных электронов время жизни будет стремиться к $\tau_{p0} = (\alpha_p^{2-} N_t)^{-1}$, где α_p^{2-} — коэффициент захвата дырок дважды отрицательными ионами никеля.

Если нарушение равновесия не мало, то концентрации центров в различных зарядовых состояниях могут сильно отличаться от равновесных (перезарядка уровней). Это может быть и при малых нарушениях равновесия, если величины коэффициентов захвата дырок и электронов сильно отличаются друг от друга. При увеличении уровня возбуждения неравновесные концентрации центров перестают зависеть от δp и δn и достигают предельных значений, зависящих от соотношения между коэффициентами захвата для разных уровней и глубины залегания уровней.

Исследуя на опыте, как зависят времена жизни τ_n и τ_p от концентраций электронов и дырок и от температуры, и зная из независимых измерений концентрацию ловушек (рекомбинационной примеси) и их энергетические уровни, можно определить коэффициенты захвата α_n и α_p для данных центров (и в данном полупроводнике) и сечения захвата S_n и S_p. Такие измерения показывают, что S_n и S_p могут изменяться в чрезвычайно широких пределах, от $\sim 10^{-12}$ см^2 до $\sim 10^{-22}$ см^2. Исследуя зависимость сечений захвата от условий опыта — концентрации носителей заряда, температуры, внешнего электрического поля, — а также анализируя излучение, возникающее при рекомбинации, можно выяснить, какие именно элементарные процессы играют главную роль в явлениях захвата и рекомбинации. К этому вопросу мы вернемся в гл. XVII.

Укажем в заключение на практическую важность рассматриваемых вопросов. Как отмечалось выше, значения сечений захвата для разных примесей можно определить экспериментально. Эти сечения совместно с энергетическими уровнями примеси являются важнейшими характеристиками рассматриваемой примеси в данном полупроводнике и вполне определяют ее влияние на равновесные свойства полупроводника и кинетику электронных процессов. Зная эти характеристики, можно затем рассчитать необходимый состав и количество примесей для получения заданных свойств полупроводника — равновесных концентраций электронов и дырок (по формулам гл. V) и нужных времен их жизни (по формулам настоящей главы), необходимых для изготовления тех или иных полупроводниковых приборов.

Глава X

ПОВЕРХНОСТНЫЕ ЭЛЕКТРОННЫЕ СОСТОЯНИЯ

§ 1. Происхождение поверхностных состояний

Одна из важных особенностей полупроводников заключается в том, что их электрические и оптические свойства могут существенно зависеть от состояния поверхности и изменяться при различной ее обработке (шлифовке, травлении, изменении окружающей среды). Общая причина этих явлений состоит в том, что в ограниченном кристалле возникают не только квантовые состояния электронов, движущихся в объеме кристалла, но еще и дополнительные состояния, в которых электроны локализованы на самой поверхности кристалла. Соответственно, помимо объемных уровней энергии, образующих энергетические зоны безграничного кристалла, появляются локальные уровни энергии, расположенные у самой поверхности.

Наличие локальных поверхностных уровней энергии приводит к тому, что электроны и дырки могут «прилипать» к поверхности, образуя поверхностный электрический заряд. При этом под поверхностью появляется равный по величине и противоположный по знаку индуцированный заряд в объеме, т. е. появляются обогащенные или обедненные приповерхностные слои. Возникновением таких слоев и объясняется влияние поверхности на равновесные свойства полупроводников (электропроводность, работу выхода, контактную разность потенциалов и др.).

Поверхностные уровни энергии могут существенно изменять и кинетику электронных процессов, так как они создают дополнительные центры рекомбинации и генерации носителей заряда. Поэтому все явления, связанные с неравновесными электронами и дырками (фотопроводимость, фотоэдс, процессы в инжекционных полупроводниковых приборах и др.), тоже зависят от состояния поверхности.

Одна из причин возникновения поверхностных состояний заключается в обрыве периодического потенциала кристалла на поверхности. И. Е. Тамм впервые показал в 1932 г., что именно это обстоятельство допускает дополнительные решения уравнения Шредингера для электрона в кристалле (по сравнению с безграничным кристаллом), которые быстро затухают при удалении от поверхности.

Причину возникновения поверхностных состояний можно пояснить на следующем примере. Рассмотрим одномерный кристалл и будем учитывать взаимодействие только между соседними атомами. Как мы видели (гл. III), при этом принципиальные особенности энергетического спектра сохраняются. Тогда для безграничного

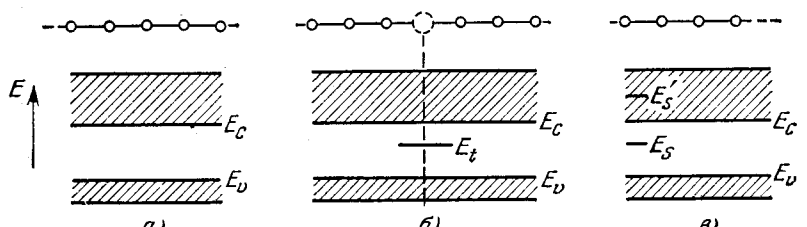


Рис. 10.1. Грубая схема для пояснения возникновения поверхностных уровней энергии.

кристалла мы получаем обычную систему объемных энергетических зон (рис. 10.1, *а*). Положим теперь, что мы удалили один из атомов. Такой структурный дефект, так же как и атом примеси, может создать локальный уровень (или уровни) энергии E_t, лежащий в запрещенной зоне (рис. 10.1, *б*). С другой стороны, в приближении близкого взаимодействия при удалении атома бесконечный кристалл распадается на два невзаимодействующих кристалла, имеющих ограничивающую «поверхность». Тогда для кристалла, имеющего границу, мы приходим к энергетической схеме рис. 10.1, *в*, в которой, кроме объемных зон, имеется еще поверхностный уровень E_s. Если уровень попадает в разрешенную зону (E_s'), то он не отвечает локальному состоянию.

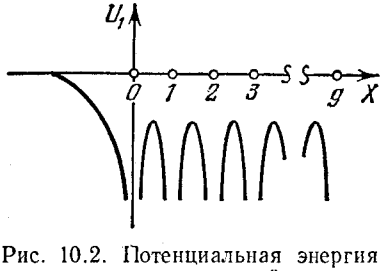


Рис. 10.2. Потенциальная энергия электрона в одномерной ограниченной решетке.

В более точных рассуждениях следует исходить из вида потенциальной энергии электрона в ограниченном кристалле. Она схематически показана на рис. 10.2. В отличие от безграничного кристалла, функция $U_1(x)$ теперь уже не является периодической. Ее можно представить в виде суммы периодической функции $U(x)$, характерной для безграничного кристалла, и добавки $W(x)$, описывающей влияние поверхности. Пользуясь по-прежнему методом сильно связанных электронов (гл. III), можно, однако, и в этом

случае искать решение уравнения Шредингера в виде

$$\psi(x) = \sum_{g=0}^{\infty} c_g \varphi_g, \tag{1.1}$$

где $\varphi_g = \varphi(x - ag)$ — волновая функция электрона в изолированном атоме номер «g», а коэффициенты c_g должны быть подобраны так, чтобы функция ψ удовлетворяла уравнению Шредингера. Теперь, однако, числа g принимают только положительные значения. Подобно тому, что делалось в § III.5, коэффициенты c_g можно искать в виде

$$c_g = Ae^{i\lambda g} + Be^{-i\lambda g}, \tag{1.2}$$

где A и B — некоторые постоянные.

Если оба коэффициента A и $B \neq 0$, то, чтобы $\psi(x)$ оставалась везде конечной, λ должно быть вещественным. Тогда $\psi(x)$ есть осциллирующая функция, что, как мы знаем, соответствует состояниям коллективизированных электронов. Оказывается, однако, что в зависимости от характера взаимодействия электрона с атомами и от вида волновых функций φ_g один из коэффициентов, A или B, может обратиться в нуль. Тогда становятся возможными и комплексные значения λ. Так, если, например, $B = 0$, то $\psi(x)$ остается конечной при

$$\lambda = \lambda_1 + i\lambda_2, \tag{1.3}$$

где λ_1 и λ_2 вещественны и $\lambda_2 > 0$. Тогда, подставляя (1.3) в формулу (1.1), мы имеем

$$\psi(x) = A \sum_{g=0}^{\infty} e^{-\lambda_2 g} e^{i\lambda_1 g} \varphi(x - ag). \tag{1.4}$$

Каждый член этой суммы затухает с увеличением номера атома g. Так как, с другой стороны, атомные функции $\varphi(x - ag)$ заметно отличны от нуля лишь в окрестности своего узла решетки $x \simeq ag$, то волновая функция $\psi(x)$ затухает при увеличении x. В таких квантовых состояниях электрон локализован у границы кристалла, и эти состояния и есть поверхностные электронные состояния Тамма.

Условия, при которых появляются решения вида (1.4), можно установить, подставляя (1.4) в соответствующее уравнение Шредингера. Таким путем можно показать, что решения с чисто вещественными значениями λ существуют всегда, а случай (1.3) реализуется, лишь если изменение потенциальной энергии у поверхности W достаточно велико. При этом значение λ_1 оказывается фиксированным; в зависимости от характера атомных волновых функций оно равно либо нулю, либо π.

Отметим, что существование поверхностных энергетических уровней можно объяснить и анализируя химические связи в кристаллах (подобно тому, как это сделано в гл. II для объяснения других полупроводниковых свойств). С этой точки зрения поверхностные уровни возникают вследствие неспаренных электронов, существующих у поверхности из-за обрыва решетки (Шокли).

В трехмерных кристаллах вместо дискретных поверхностных уровней возникают поверхностные энергетические зоны (рис. 10.3, *а*). Они состоят из близко расположенных уровней энергии, соответствующих различным возможным компонентам квазиимпульса p_y и p_z, параллельным поверхности.

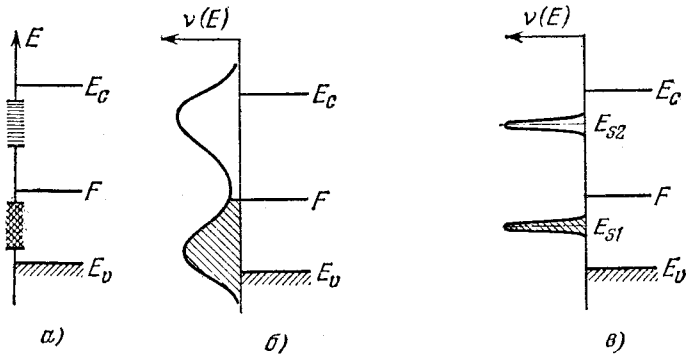


Рис. 10.3. *а*) Поверхностные энергетические зоны. *б*) Непрерывное распределение поверхностных уровней. *в*) Квазидискретные поверхностные уровни.

Если какая-либо из поверхностных зон заполнена электронами лишь частично, то в принципе возможна электропроводность металлического типа вдоль самой поверхности.

В действительности, однако, поверхность кристалла всегда неоднородна и распределение потенциала вдоль нее отнюдь не периодично. По этой причине представление о поверхностных зонах как точном двумерном аналоге объемных энергетических зон кристалла представляет собой слишком большую идеализацию. Однако само понятие поверхностных зон энергии фактом неоднородности реальной поверхности не отменяется.

Поверхностные энергетические зоны могут перекрываться друг с другом, так же как и с объемными зонами. Поэтому часто бывает более правильно говорить о непрерывном распределении поверхностных уровней и описывать его, задавая плотность поверхностных уровней $\nu(E)$ (количество уровней на единицу поверхности и единичный интервал энергии). Пример такого распределения показан на рис. 10.3, *б*. При этом поверхностная концентрация захваченных электронов (с точностью до тепловой размазки функции Ферми)

выражается площадью, ограниченной соответствующим участком кривой $\nu(E)$ (для $E < F$) и осью E.

Иногда при интерпретации экспериментальных данных оказывается возможным считать, что распределение $\nu(E)$ имеет один или несколько резких и узких максимумов (рис. 10.3, *в*). В этих случаях говорят о квазидискретных (или просто дискретных) поверхностных уровнях, в известной мере аналогичных уровням одномерных кристаллов. Такое приближение часто используют при качественном (или полуколичественном) анализе явлений.

Выше мы везде считали, что поверхность кристалла является «атомарно чистой», т. е. лишенной каких-либо чужеродных атомов и дефектов. Поверхности, близкие к атомарно чистым, можно получить с помощью специальных методов (например, образуя поверхность скалыванием в очень глубоком вакууме или очищая поверхность бомбардировкой ионами благородных газов). В обычных же условиях поверхность полупроводника всегда покрыта слоем окислов. Кроме того, на ней имеются адсорбированные атомы, которые тоже создают поверхностные электронные уровни энергии. Эти атомы могут обмениваться электронами с объемом полупроводника и в зависимости от того, что́ является энергетически выгодным, либо отдавать электроны полупроводнику (превращаясь в положительные ионы), либо присоединять к себе дополнительные электроны (заряжаясь отрицательно). Аналогичную роль могут играть и различные «собственные» структурные дефекты поверхности — вакансии и т. д. Электронный обмен между адсорбированными атомами и объемом играет важнейшую роль в современной электронной теории химической адсорбции и гетерогенного катализа [3, 4].

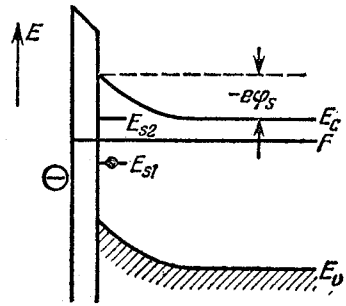


Рис. 10.4. Энергетическая диаграмма реальной поверхности кристалла.

Учитывая сказанное выше, мы приходим к энергетической диаграмме «реальной» поверхности, изображенной на рис. 10.4. Здесь для простоты показаны только два дискретных уровня Тамма (E_{s1} и E_{s2}) и один уровень адсорбированного атома. Вследствие заряда поверхности (созданного уровнями всех типов) электрический потенциал φ в приповерхностном слое полупроводника (толщина которого порядка длины экранирования Дебая) изменяется и, соответственно, энергетические зоны искривляются (рисунок соответствует отрицательному заряду поверхности). Разность потенциалов φ_s между поверхностью полупроводника и его объемом (где зоны уже горизонтальны) мы будем называть *поверхностным потенциалом.*

В дальнейшем мы будем его выражать в безразмерных единицах:

$$Y_s = \frac{e\varphi_s}{kT}, \tag{1.5}$$

так как именно в этом виде он обычно входит в различные формулы.

Приведенная диаграмма является, конечно, тоже очень упрощенной. Так, при достаточно большом числе адсорбированных атомов или иных структурных дефектов поверхность следует рассматривать уже как неупорядоченную систему, применяя для теоретического ее исследования методы, развитые в физике сильно легированных полупроводников (гл. XIX).

§ 2. Влияние поверхностного потенциала на электропроводность

При изменении поверхностного потенциала изменяются концентрации электронов и дырок в приповерхностном слое полупроводника, а следовательно, и его электропроводность. Поэтому, исследуя, как меняется электропроводность при различной обработке поверхности, можно выяснить, как при этом изменяется заряд поверхности, и отсюда получить данные о поверхностных состояниях.

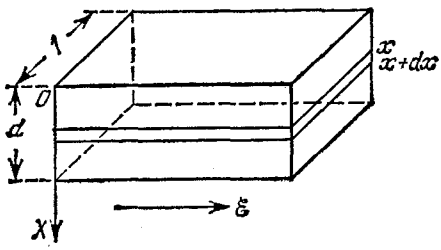


Рис. 10.5. К вычислению влияния поверхности на электропроводность.

Рассмотрим пластинку однородного полупроводника, одна из поверхностей которой лежит в плоскости $x = 0$ (рис. 10.5), а толщина d намного меньше других размеров. Ширину пластинки положим равной единице. Сила тока через бесконечно тонкий слой пластинки $(x,\ x + dx)$ есть

$$e\left[\mu_p p(x) + \mu_n n(x)\right] \mathscr{E}\, dx,$$

а ее изменение по сравнению со случаем неискривленных зон ($Y_s = 0$) равно

$$e\left\{\mu_p\left[p(x) - p_0\right] + \mu_n\left[n(x) - n_0\right]\right\} \mathscr{E}\, dx.$$

Поэтому, обозначая

$$\Gamma_p = \int_0^\infty \left[p(x) - p_0\right] dx, \qquad \Gamma_n = \int_0^\infty \left[n(x) - n_0\right] dx, \tag{2.1}$$

мы находим, что изменение электропроводности пластинки ΔG (рассчитанное на единицу длины и единицу ширины пластинки),

обусловленное влиянием поверхности при $x = 0$, есть

$$\Delta G = \frac{\Delta i}{\mathscr{E}} = e\mu_p (\Gamma_p + b\Gamma_n), \qquad (2.2)$$

где $b = \mu_n/\mu_p$. Величину ΔG часто называют *поверхностной проводимостью*. В интегралах (2.1) в качестве верхнего предела мы положили ∞ (вместо координаты второй поверхности), предполагая, что толщина пластинки d хотя бы в несколько раз превышает длину экранирования.

Мы считали подвижности в приповерхностном слое равными их значениям μ_p и μ_n в объеме. В действительности на поверхности происходит дополнительное рассеяние импульса. Поэтому при точных расчетах вместо μ_n и μ_p нужно пользоваться «поверхностными» подвижностями, которые, при определенных предположениях о

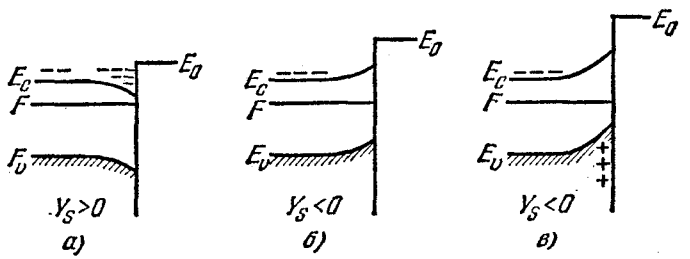


Рис. 10.6. Обогащенный (*а*), обедненный (*б*) и инверсионный (*в*) приповерхностные слои в полупроводнике n-типа.

характере рассеяния на поверхности, можно вычислить (подробнее см., например, в [1]). Однако эта поправка не меняет принципиальные результаты, и мы учитывать ее не будем.

Таким образом, расчет влияния поверхностного потенциала на электропроводность сводится к вычислению распределения $p(x)$ и $n(x)$ в слое объемного заряда. Однако общий характер зависимости ΔG от Y_s можно выяснить без расчета. Положим для определенности, что мы имеем полупроводник n-типа. При положительном потенциале поверхности зоны искривляются вниз и край зоны основных носителей E_c приближается к уровню Ферми (рис. 10.6, *а*). Поэтому у поверхности образуется слой, обогащенный электронами, и ΔG будет непрерывно увеличиваться при возрастании Y_s. При $Y_s < 0$ зоны искривляются вверх (рис. 10.6, *б*) и, соответственно, приповерхностный слой обедняется электронами. Пока $(F - E_v)$ у поверхности остается больше $(E_c - F)$, концентрация дырок $p \ll n$ везде и ΔG уменьшается при увеличении $|Y_s|$. Однако, когда искривление зон становится таким, что F оказывается ближе к E_v, нежели к E_c, концентрация дырок в приповерхностном слое делается

больше концентрации электронов, т. е. образуется инверсионный слой, в данном случае p-типа (рис. 10.6, *в*). Аналогично, в дырочном полупроводнике инверсионный слой имел бы проводимость n-типа. Поэтому при дальнейшем увеличении $|Y_s|$ проводимость ΔG проходит через минимум и снова увеличивается, теперь уже за счет увеличения концентрации дырок в инверсионном слое.

Значение поверхностного потенциала $Y_{s\,\min}$, соответствующее минимуму ΔG, можно найти из следующих простых (хотя и нестрогих) соображений. Очевидно, что при $Y_{s\,\min}$ электроны и дырки дают одинаковый вклад в электропроводность приповерхностного слоя. Поэтому для некоторой средней плоскости в области объемного заряда $x_1 = \text{const}$, где потенциал $Y_s = {}^1/_2\, Y_{s\,\min}$, можно написать

$$\mu_n n(x_1) = \mu_p p(x_1).$$

С другой стороны, ограничиваясь случаем невырожденных полупроводников, мы имеем по закону Больцмана

$$n = n_0 \exp Y, \quad p = p_0 \exp(-Y). \tag{2.3}$$

Подставляя это в предыдущее равенство, мы имеем

$$\mu_n n_0 \exp({}^1/_2 Y_{s\,\min}) = \mu_p p_0 \exp(-{}^1/_2 Y_{s\,\min}). \tag{2.4}$$

Отсюда получаем

$$Y_{s\,\min} = \ln(\xi^2/b), \tag{2.5}$$

где $\xi^2 = p_0/n_0$, а $b = \mu_n/\mu_p$. Для рассматриваемого случая полупроводника n-типа значение $\xi < 1$, а отношение подвижностей b обычно > 1. Поэтому $Y_{s\,\min}$ оказывается отрицательным. Его абсолютное значение тем больше, чем меньше ξ.

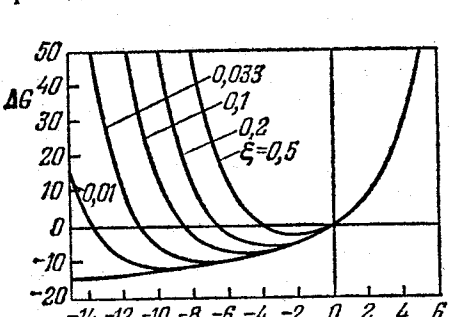


Рис. 10.7. Зависимость изменения проводимости ΔG германия n-типа (произвольные единицы) от поверхностного потенциала $Y_s = e\varphi_s/kT$.

Зависимость ΔG от Y_s, вычисленная для германия n-типа при различных значениях параметра ξ, показана на рис. 10.7. При расчете предполагалось, что доноры и акцепторы полностью ионизованы и что электронный газ не вырожден. Уравнение этих кривых и точный вывод формулы (2.5) даны в Приложении VI.

Рассмотренная зависимость $\Delta G\,(Y_s)$ лежит в основе важного метода экспериментального определения Y_s. Для этого нужно сначала знать отношение p_0/n_0 внутри образца (за пределами слоя объемного заряда). Оно легко может быть найдено из измерений кон-

центрации основных носителей в массивном образце до вырезания тонкой пластинки. Тогда при известном отношении подвижностей b определено, какая из семейства кривых рис. 10.7 относится к данному образцу и чему равно $Y_{s\,\min}$. Затем меняют окружающую среду, чтобы вызвать изменение Y_s. Для этого часто применяют различные газы (азот, кислород, пары воды и др.) при разном давлении и подбирают такую среду, при которой электропроводность проходит через минимум. Это значение $G_{\min}$ измеряют. Тогда при любом другом состоянии поверхности, которому соответствует электропроводность образца G, поверхностный потенциал равен

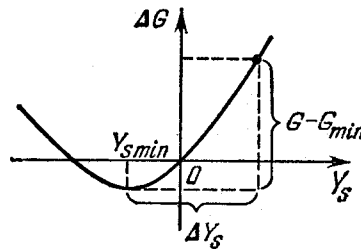


Рис. 10.8. К определению величины поверхностного потенциала.

$$Y_s = Y_{s\min} + \Delta Y_s, \tag{2.6}$$

где ΔY_s определяется величиной $(G - G_{\min})$, как показано на рис. 10.8.

§ 3. Эффект поля

Величину поверхностного потенциала можно изменять не только изменяя окружающую среду, но и создавая у поверхности полупроводника поперечное электрическое поле. Влияние внешнего электрического поля на электропроводность полупроводника получило название *эффекта поля*.

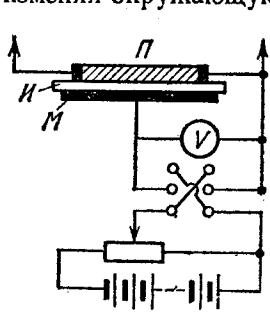


Рис. 10.9. Схема наблюдения стационарного эффекта поля.

Существует большое число разнообразных экспериментальных приемов изучения эффекта поля как в стационарном, так и в нестационарном режимах. Пример стационарного метода показан на рис. 10.9. Пластинка полупроводника $П$ служит одной из обкладок конденсатора, второй обкладкой которого является металлическая пластинка $М$, отделенная от полупроводника тонким слоем изолятора $И$. К конденсатору прикладывается постоянное напряжение, величину и знак которого можно изменять. Концы пластинки полупроводника имеют низкоомные контакты, с помощью которых она включается в мостовую схему (или какую-либо другую) для точного измерения малых изменений проводимости.

Еще удобнее исследовать эффект поля, прикладывая к пластинам конденсатора переменное напряжение низкой частоты (десятки или

сотни герц). Тогда, используя простые схемы и подавая на одну из пар пластин осциллографа сигнал, пропорциональный приложенному напряжению u, а на другую пару пластин — сигнал, пропорциональный ΔG, можно получить на экране всю кривую зависимости ΔG от u. Это позволяет легко определить по формуле (2.6) значение Y_s при любом напряжении на конденсаторе, и в частности, при $u = 0$ (в «естественном» состоянии поверхности). Такой метод имеет большое преимущество перед использованием различных газовых атмосфер, так как эти последние могут не только искривлять энергетические зоны, но и изменять концентрацию поверхностных уровней энергии вследствие адсорбции атомов газов.

Исследование эффекта поля позволяет получить ценную информацию о поверхностных состояниях (энергетических уровнях и их концентрациях). Эта возможность основана на следующем. Электроны и дырки под каждой единицей поверхности создают заряд

$$Q = Q_V + Q_s = e(\Gamma_p - \Gamma_n) + Q_s. \tag{3.1}$$

Здесь Q_V — подвижный заряд в объеме приповерхностного слоя, Q_s — связанный заряд на поверхностных уровнях, а Γ_p и Γ_n — полные количества избыточных дырок и электронов, определяемые формулами (2.1). Кроме того, имеется еще заряд, создаваемый ионами в объеме полупроводника (заряженными донорами и акцепторами), и ионами, адсорбированными на внешней поверхности окисла. В отсутствие внешнего поля заряд ионов равен по величине и противоположен по знаку заряду Q.

Во многих случаях можно считать, что в эффекте поля ионный заряд не изменяется. Это справедливо, если доноры и акцепторы в объеме полупроводника полностью ионизованы. Адсорбированные же ионы обмениваются электронами с полупроводником очень медленно (часто за многие секунды и минуты), и за время измерения их заряд остается постоянным. Поэтому индуцированный заряд равен

$$\delta Q = \delta Q_V + \delta Q_s = e\,(\delta\Gamma_p - \delta\Gamma_n) + \delta Q_s. \tag{3.2}$$

Так как Γ_p и Γ_n — известные функции поверхностного потенциала Y_s (см. Приложение VI), а этот последний, как указывалось выше, тоже определяется из эффекта поля, то величину δQ_V можно легко найти. С другой стороны, суммарный индуцированный заряд (отнесенный к единице площади) равен

$$\delta Q = Cu, \tag{3.3}$$

где C — емкость конденсатора на единицу площади, которая непосредственно определяется на опыте. Поэтому оказывается возможным определить δQ_s и найти, какая доля носителей заряда находится в свободном и связанном состояниях.

Результаты измерения эффекта поля иногда выражают с помощью *эффективной подвижности*. Она, по определению, равна

$$\mu_{\text{эфф}} = \frac{\delta G}{\delta Q}, \tag{3.4}$$

где δG — изменение проводимости в эффекте поля. Эта величина имеет особенно простой физический смысл, если искривление зон невелико и проводимость везде можно считать монополярной. Тогда, скажем для дырочного полупроводника, мы имеем

$$\delta G = e\mu_p \delta \Gamma_p = \mu_p \delta Q_V,$$

а следовательно,

$$\frac{\mu_{\text{эфф}}}{\mu_p} = \frac{\delta Q_V}{\delta Q_V + \delta Q_s}.$$

В этом случае $\mu_{\text{эфф}}$ определяет, какая доля полного заряда поверхности находится в подвижном состоянии.

Исследуя зависимость Q_s от Y_s, можно определить энергетическое положение поверхностных уровней энергии и их концентрацию. Действительно, при изменении Y_s поверхностные уровни, так же как и края зон у поверхности, перемещаются относительно уровня Ферми F. При прохождении какого-либо уровня E_s через F зарядное состояние уровня изменяется. При наличии уровней только одного типа это проявляется на кривых зависимости Q_s от Y_s возникновением ступени.

Чтобы охарактеризовать энергетический спектр поверхностных состояний, будем отсчитывать E_s при неискривленных зонах ($Y_s = 0$) от положения уровня Ферми в собственном полупроводнике F_i, обозначая $\varepsilon_s \equiv E_s - F_i$. Тогда

$$E_s - F = \varepsilon_s - (F - F_i) - kTY_s.$$

Далее, для невырожденных (в объеме) полупроводников из формул (V.5.1) и (V.5.2) имеем

$$F - F_i = \frac{1}{2} kT \ln \frac{n_0}{p_0}.$$

Поэтому, используя функцию распределения (V.9.3), находим, что вероятность заполнения уровня E_s при данном Y_s есть

$$f(\varepsilon_s, Y_s) = \frac{1}{1 + \exp\left(\frac{\varepsilon_s}{kT} - \frac{1}{2} \ln \frac{n_0}{p_0} - Y_s\right)}. \tag{3.5}$$

Здесь для простоты мы опустили множитель, учитывающий вырождение уровня. Если рассматриваемые уровни акцепторные и их поверхностная концентрация равна ν, то заряд поверхности будет равен

$$Q_s = -e\nu f(\varepsilon_s, Y_s). \tag{3.6}$$

Для донорных уровней имеем

$$Q_s = ev\,[1 - f(\varepsilon_s,\ Y_s)]. \tag{3.7}$$

Из формул (3.5) — (3.7) видно, что в обоих случаях на кривых зависимости Q_s от Y_s появляется ступень, а также точка перегиба. Последняя соответствует поверхностному потенциалу

$$Y_{s1} = \frac{\varepsilon_s}{kT} - \frac{1}{2}\ln\frac{n_0}{p_0}. \tag{3.8}$$

Следовательно, определяя по данным эффекта поля Y_{s1}, можно найти ε_s. Величина ступени равна ev.

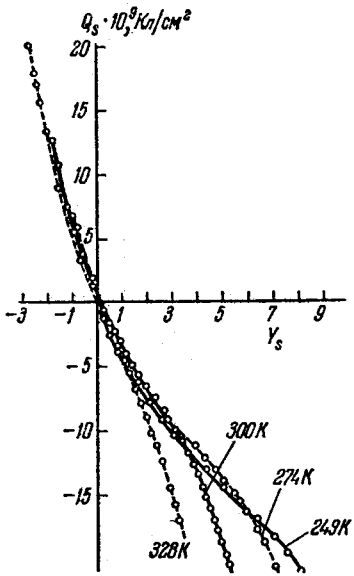


Рис. 10.10. Пример зависимости заряда поверхности от поверхностного потенциала для образца германия.

В случае нескольких дискретных уровней энергии мы имели бы не одну, а несколько точек перегиба.

Если имеется непрерывный энергетический спектр поверхностных состояний, то Q_s получается суммированием выражений типа (3.6) и (3.7). Так, например, для акцепторных уровней мы получаем

$$Q_s = -e\int \frac{\nu(\varepsilon_s)\,d\varepsilon_s}{1+\exp\left(\frac{\varepsilon_s}{kT} + \frac{1}{2}\ln\frac{p_0}{n_0} - Y_s\right)}, \tag{3.9}$$

где интегрирование производится по всей запрещенной зоне энергий. Здесь $\nu(\varepsilon_s)$ есть поверхностная плотность уровней, рассчитанная на единичный интервал энергии. В этом случае задача определения энергетического спектра поверхностных состояний $\nu(\varepsilon_s)$ по экспериментальной зависимости Q_s от Y_s становится гораздо более сложной и требует либо дополнительных данных, либо дополнительных предположений.

Экспериментальное исследование зависимости Q_s от Y_s («кривых захвата») показывает, что кривые захвата, как правило, оказываются плавными и не обнаруживают ясно выраженных ступеней и точек перегиба. Пример таких кривых для германия приведен на рис. 10.10 *). Это показывает, что энергетический спектр поверхностных состояний, ответственных за захват носителей заряда, является квазинепрерывным. Анализ кривых захвата приводит

*) Рисунок из книги [1].

также к заключению, что ν (ε_s) имеет обычно минимальное значение вблизи середины запрещенной зоны и увеличивается по мере приближения к краям разрешенных зон энергии.

Важные данные о свойствах поверхностных уровней можно получить, исследуя изменение ΔG во времени. Для этого разработаны различные методы. Одна из схем показана на рис. 10.11. Здесь на обкладки конденсатора полупроводник — металл подается прямоугольный импульс напряжения с крутым передним фронтом. Об изменении проводимости полупроводника судят по изменению падения напряжения на нем, возникающего от протекания слабого постоянного тока от батареи $Б$. Это напряжение регистрируется осциллографом $О$, развертка которого синхронизована с генератором импульсов $Г$. В подобных импульсных методах, однако, всегда необходимо тщательно исключать наводку в измерительной схеме, возникающую при наложении напряжения на конденсатор (вследствие неполной симметрии схемы), которая может быть намного больше измеряемого сигнала. Для ее исключения образец включают в схему моста $R-C$ и подбирают сопротивления r_1, r_2 и емкости C_1, C_2 таким образом, чтобы при выключенной батарее $Б$ сигнал наводки был минимальным.

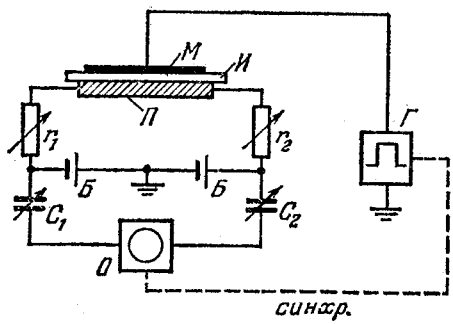


Рис. 10.11. Наблюдение релаксации проводимости в эффекте поля. $П$ — полупроводник, $М$ — металлическая обкладка, $И$ — изолятор, $Г$ — генератор прямоугольных импульсов, $О$ — осциллограф, $Б$ — источники постоянного тока.

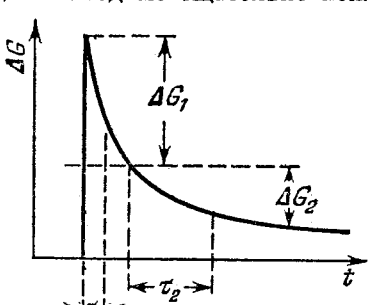


Рис. 10.12. Зависимость изменения проводимости ΔG в эффекте поля от времени (схематически).

Наблюдаемая зависимость ΔG от времени обычно имеет вид, показанный схематически на рис. 10.12. После подачи напряжения, соответствующего обогащению приповерхностного слоя основными носителями, G быстро увеличивается и через короткое время, порядка максвелловского времени релаксации ΔG, достигает наибольшего значения. За эти времена поверхностные уровни еще не успевают заметно изменить свое зарядовое состояние, и весь индуцированный заряд находится в свободном состоянии. Затем начинается перераспределение заряда между объемом и поверхностными уровнями и G

уменьшается. Это изменение имеет сложный, неэкспоненциальный характер. Сначала G изменяется быстро на некоторую величину ΔG_1. Эту стадию процесса можно условно охарактеризовать некоторым временем релаксации τ_1, в течение которого G уменьшается на определенную долю, скажем в e раз. Значения τ_1 зависят от рода полупроводника и состояния его поверхности и обычно лежат в интервале $10^{-2} \div 10^2$ микросекунд. Затем электропроводность изменяется еще на величину ΔG_2, однако гораздо более медленно, и в конце концов достигает стационарного состояния. Характерное время для этой второй стадии процесса может измеряться многими секундами и минутами. Такой характер кинетики показывает, что поверхностные уровни можно разделить на две группы: «быстрые» уровни, которые легко обмениваются электронами с объемом, и «медленные» уровни, обменивающиеся электронами с объемом гораздо хуже.

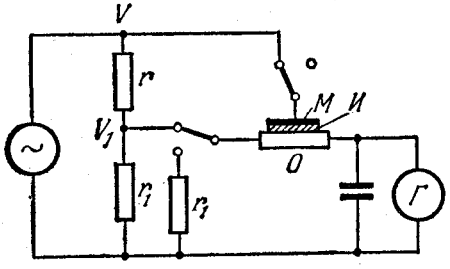


Рис. 10.13. Измерение частотной зависимости эффективной подвижности. O — образец, $И$ — изолятор, M — металлическая обкладка, $Г$ — гальванометр.

Другой прием исследования эффекта поля показан на рис. 10.13. Здесь к конденсатору и к торцам образца прикладываются переменные напряжения одинаковой частоты

$$u = u_0 \cos \omega t, \qquad u_1 = u_{10} \cos \omega t.$$

Эффект поля вызывает модуляцию проводимости образца, которую при малых напряжениях можно считать тоже гармонической:

$$G = G_0 + (\Delta G)_0 \cos(\omega t - \varphi) = G_0 + (\Delta G)_0 (\cos \omega t \cdot \cos \varphi + \sin \omega t \cdot \sin \varphi).$$

Поэтому сила тока через образец есть

$$i = \frac{a}{l} G u_1 = \frac{a}{l} [G_0 + (\Delta G)_0 \cos \varphi \cdot \cos \omega t + (\Delta G)_0 \sin \varphi \cdot \sin \omega t] u_{10} \cos \omega t,$$

где a — ширина пластинки, а l — ее длина. Усредняя это выражение по периоду колебаний, мы находим, что вследствие эффекта поля возникает постоянная составляющая тока

$$\bar{i} = \frac{1}{2} \frac{a}{l} (\Delta G)_0 \cos \varphi \cdot u_{10}.$$

С другой стороны, амплитуда индуцированного заряда на единице поверхности есть $(\Delta Q)_0 = C u_0$. Поэтому для эффективной подвижности получается

$$\mu_{эфф} \cos \varphi = \frac{(\Delta G)_0}{(\Delta Q)_0} \cos \varphi = \frac{2l}{u_0 u_{10} C a} \bar{i}. \qquad (3.10)$$

Пользуясь комплексным описанием гармонических колебаний, можно сказать, что формула (3.10) дает вещественную часть $\mathrm{Re}\ \mu_{эфф}$.

Определение фазового сдвига φ требует дополнительных измерений. Однако уже исследование $\mathrm{Re}\ \mu_{эфф}$ дает много данных. На рис. 10.14 приведен пример зависимости $\mathrm{Re}\ \mu_{эфф}$ от частоты для германия p-типа, поверхность которого была протравлена кислотами (смесь уксусной, азотной и фтористоводородной) в различных газовых атмосферах. Из рисунка видно, что при адсорбции дипольных молекул воды $\mathrm{Re}\ \mu_{эфф}$ при низких частотах оказывается отрицательной, что обозначает существование инверсионного слоя. Однако при увеличении частоты $\mathrm{Re}\ \mu_{эфф}$ становится положительной. В других газах (сухой кислород, озон) инверсионный слой не возникает.

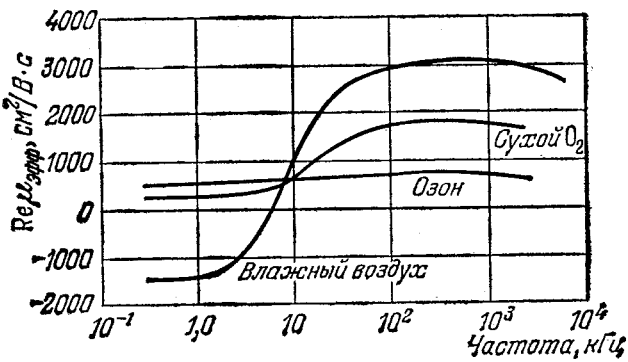


Рис. 10.14. Пример зависимости эффективной подвижности от частоты. Германий p-типа [2].

Исследование зависимости $\mu_{эфф}$ от частоты подтверждает существование быстрых и медленных поверхностных уровней и позволяет оценить их характерные времена релаксации.

Имеющиеся данные показывают, что поверхностные состояния характеризуются сложным и, как правило, непрерывным энергетическим спектром. Часть поверхностных уровней лежит в верхней половине запрещенной зоны, а часть — в нижней половине. Эти уровни могут быть акцепторными и донорными. Их концентрация зависит от обработки поверхности (шлифовка, химическое травление и др.) и, например, в германии может достигать порядка $10^{14} \div 10^{15}$ см$^{-2}$.

Основная часть быстрых поверхностных состояний является по-видимому, состояниями Тамма, уровни которых на реальной поверхности локализованы вблизи раздела полупроводник — окисел. Медленные же состояния обычно связывают с адсорбированными атомами, так как электронный обмен между ними и объемом полупроводника сильно затрудняется наличием слоя окисла. Однако

некоторые из состояний Тамма могут оказаться тоже медленными, если соответствующие им сечения захвата электронов (дырок) (§ 8) достаточно малы.

Поверхностные состояния особого типа возникают при достаточно сильном изгибе зон. Если, например, $Y_s > 0$, то зоны изгибаются вниз и у поверхности возникает потенциальная яма для электронов (рис. 10.15). Форма ее зависит от степени однородности поверхности. Если свойства последней всюду одинаковы, то яма, очевидно, имеет вид желоба с острым дном, вытянутого вдоль поверхности. При достаточной глубине и ширине ямы в ней могут образоваться связанные состояния, находясь в которых электроны локализованы вблизи поверхности. Это явление называется *поверхностным квантованием*. В одномерном случае указанным состояниям отвечали бы дискретные уровни в запрещенной зоне; в трехмерной задаче с однородной поверхностью мы получили бы поверхностные энергетические зоны. Электроны, их заполняющие, могли бы свободно перемещаться вдоль поверхности, соответственно чему их состояния, кроме номера зоны, можно было бы характеризовать двумерным квазиимпульсом с компонентами p_y, p_z. Реальная поверхность, однако, всегда неоднородна, и распределение потенциала вдоль нее не периодично. Поэтому, как уже отмечалось выше, полной аналогии между поверхностными и объемными зонами не имеется. Соответственно, и электропроводность по поверхностным зонам в реальных кристаллах может не проявляться.

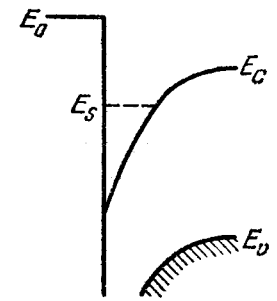


Рис. 10.15. Потенциальная яма для электрона при $Y_s > 0$. Пунктиром обозначен один из уровней электрона в яме.

Очевидно, что сказанное справедливо и для дырок, с той лишь разницей, что для них потенциальная яма возникает при изгибе зон вверх.

§ 4. Некоторые эффекты, связанные с поверхностными состояниями

Так как поверхностные уровни вызывают искривление зон вблизи поверхности, то все явления, связанные с потенциальными барьерами, зависят от поверхностных состояний.

1. Поверхностные уровни изменяют термоэлектронную работу выхода электронов. При положительном потенциале поверхности (рис. 10.16) зоны искривляются вниз на $-e\varphi_s$. Откладывая от дна зоны проводимости E_c у поверхности электронное сродство χ (которое не зависит от изгиба зон), мы получим положение уровня энергии электрона в вакууме E_0. Из рис. 10.16 видно, что термоэлект-

ронная работа выхода $\Phi = E_0 - F$ в этом случае уменьшается. Ее изменение равно $\Delta\Phi = -e\varphi_s$. При отрицательном потенциале поверхности зоны искривляются вверх и Φ увеличивается.

Знаки $\Delta\Phi$ и изменения проводимости ΔG связаны между собой. В полупроводнике n-типа при $\varphi_s > 0$ возникает обогащенный слой и $\Delta G > 0$, а при $\varphi_s < 0$ (при небольшом искривлении зон, пока не образуется инверсионный слой) $\Delta G < 0$. Поэтому $\Delta\Phi$ и ΔG имеют разные знаки. Напротив, в полупроводнике p-типа (как на рис. 10.16) $\Delta\Phi$ и ΔG имеют одинаковые знаки. Это заключение хорошо подтверждается в экспериментах, в которых одновременно измеряются ΔG и изменение контактной разности потенциалов (относительно какого-либо стандартного металла) при изменении газовой атмосферы или в эффекте поля.

2. Поверхностный изгиб зон проявляется и при внешнем фотоэлектрическом эффекте с полупроводниковых фотокатодов. Здесь электроны возбуждаются из валентной зоны в зону проводимости при поглощении света. Если при этом электроны приобретают энергию, достаточную для преодоления потенциального барьера на поверхности, часть из них может выйти в вакуум. Чтобы снизить темновой ток катода, очевидно, выгодно использовать полупроводники p-типа, так как в них концентрация электронов в зоне проводимости без освещения очень мала.

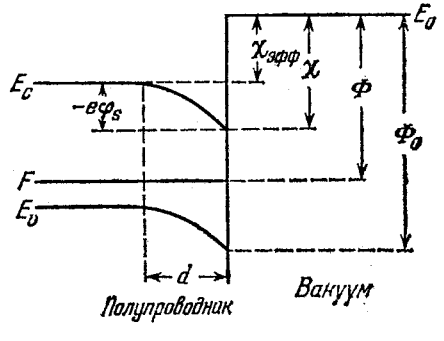


Рис. 10.16. Влияние изгиба зон у поверхности на электронную эмиссию. Φ — термоэлектронная работа выхода, Φ_0 — оптическая работа выхода, χ — электронное сродство, $\chi_{эфф}$ — эффективное электронное сродство.

Квантовый выход фотоэмиссии (число испускаемых электронов на один поглощенный фотон) существенно зависит от соотношения между толщиной слоя объемного заряда d и средней длиной свободного пробега l. Действительно, электроны, возбужденные светом в зону проводимости, диффундируя к поверхности, испытывают соударения с решеткой и на длине, порядка немногих длин свободного пробега, термолизуются. При $d \gg l$ термолизация происходит и в слое объемного заряда, и поэтому в вакуум могут выйти только те электроны, которые возбуждаются светом в тонком слое у поверхности, толщиной порядка l. Эта «глубина выхода» электронов обычно мала по сравнению с глубиной проникновения света и, например, для кремния, германия, соединений $A^{III}B^{V}$ имеет порядок 10^{-6} см. Чтобы в этих условиях электрон мог выйти в вакуум, он должен получить

минимальную энергию $\Phi_0 = E_g + \chi$ (рис. 10.16). Эта величина называется *оптической работой выхода*. Она не зависит от поверхностного потенциала и приповерхностного изгиба зон. Минимальная энергия фотонов (красная граница фотоэмиссии) определяется условием $\hbar\omega_{min} \simeq \Phi_0$ (мы пишем знак «приблизительно равно» вследствие возможного участия фононов в процессах возбуждения электронов, ср. § XVIII.5).

Однако при увеличении степени легирования полупроводника d уменьшается, и в сильно легированных полупроводниках оказывается возможным осуществить условие $d \lesssim l$. В указанных выше полупроводниках это происходит при концентрации основных носителей $\sim 10^{18} \div 10^{19}$ см$^{-3}$, когда уровень Ферми F приближается к E_v. В этом случае электроны практически не испытывают соударений в слое объемного заряда и фотоэмиссия происходит так, как если бы вместо истинного электронного сродства χ мы имели некоторое эффективное сродство $\chi_{эфф} = \chi - e\varphi_s$ (рис. 10.16). При положительном потенциале поверхности $\chi_{эфф} < \chi$. Соответственно красная граница фотоэмиссии теперь определяется условием $\hbar\omega_{min} \approx \approx E_g + \chi - e\varphi_s < \Phi_0$ и сдвигается в область более длинных волн.

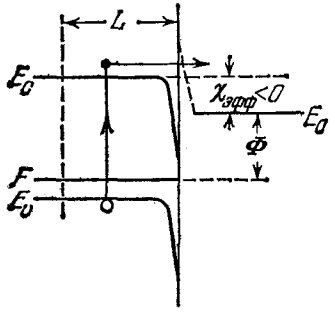


Рис. 10.17. Полупроводник с отрицательным эффективным электронным сродством.

Вторая важная особенность случая $d \lesssim l$ заключается в том, что здесь выйти в вакуум могут все избыточные электроны, которые в процессе диффузии к поверхности не исчезнут вследствие рекомбинации. Поэтому глубина выхода электронов имеет порядок длины диффузии $L \gg l$, вследствие чего квантовый выход таких фотокатодов сильно увеличивается. Таким образом, в сильно легированных полупроводниках ($d \lesssim l$) поверхностные состояния могут сильно сказываться на фотоэмиссии электронов, существенно изменяя красную границу и квантовый выход.

При специальной обработке поверхности можно дополнительно понизить уровень энергии вакуума E_0 относительно краев зон и даже сделать уровень E_0 лежащим ниже уровня E_c в объеме полупроводника (рис. 10.17). Это можно осуществить, например, напыляя в вакууме на поверхность кремния или кристаллов соединений $A^{III}B^{V}$ тонкий слой атомов цезия (один или несколько атомных слоев) и затем окисляя его в атмосфере кислорода. Толщина возникающего при этом потенциального барьера оказывается меньше длины волны де Бройля для электронов, и поэтому значительная часть электронов проходит сквозь этот барьер с помощью туннель-

ного эффекта. В этом случае $\chi_{эфф} = E_0 - E_c$ становится отрицательным. Это значит, что для выхода электрона в вакуум его достаточно только перевести на уровень E_c. Соответственно пороговая энергия фотонов еще уменьшается и становится $\hbar\omega_{min} \simeq E_g$. Используя полупроводники с отрицательным эффективным электронным сродством, оказывается возможным сильно увеличить квантовый выход фотокатодов (примерно на два порядка) и сделать их чувствительными для близкой инфракрасной области спектра. Таким же способом можно существенно увеличить и эффективность эмиттеров вторичных электронов, в которых возбуждение электронов происходит не под действием света, а в результате бомбардировки эмиттера пучком первичных электронов.

3. Поверхностные уровни могут сильно проявляться в электрических свойствах прижимных контактов металл — полупроводник. В §§ VI.11, VI.12 мы видели, что вольтамперная характеристика идеального контакта должна существенно зависеть от контактной разности металл — полупроводник, а следовательно, и от работы выхода металла. Эту «действующую» контактную разность можно определить, исследуя температурную зависимость дифференциального сопротивления контакта при нулевом напряжении (§ VI.13). Однако, как уже упоминалось, для некоторых полупроводников, например для германия и кремния, на опыте не наблюдается никакой корреляции между работой выхода металла и нулевым сопротивлением контакта. Это объясняется тем, что в таких полупроводниках поверхностная концентрация уровней очень велика и поэтому искривление зон в них обусловлено главным образом зарядом поверхности, а не контактным полем.

§ 5. Скорость поверхностной рекомбинации

Поверхностные уровни энергии, так же как и локальные уровни примесных атомов в объеме, могут участвовать в рекомбинации и тепловой генерации электронов и дырок и поэтому могут сильно влиять на их времена жизни. Влияние поверхности на кинетику электронных процессов принято характеризовать *скоростью поверхностной рекомбинации*. К этому понятию мы приходим следующим образом.

Пусть R_{ps} есть суммарный темп рекомбинации дырок на поверхности, т. е. разность темпов захвата и обратного теплового выброса, рассчитанных на единицу поверхности и единицу времени. Обозначим, далее, через p_s концентрацию дырок у поверхности в слое объемного заряда (рис. 10.18). В равновесии, когда $p_s = p_{s0}$, $R_{ps} = 0$. Тогда, аналогично тому как мы поступали, рассматривая темп объемной рекомбинации (§ VII.2), формально можно написать

$$R_{ps} = s'_p (p_s - p_{s0}) = s'_p \delta p_s. \qquad (5.1)$$

Аналогично, для суммарного темпа поверхностной рекомбинации электронов можно положить

$$R_{ns} = s_n' (n_s - n_{s0}) = s_n' \delta n_s. \tag{5.2}$$

Эти соотношения иногда рассматривают как первые члены разложения функций R_{ps} и R_{ns} по степеням переменных δp_s и δn_s соответственно. Однако фактически каждый из коэффициентов s_p' и s_n' может сам зависеть от обеих переменных δp_s и δn_s.

Однако выражения для R_{ps} и R_{ns} принято записывать несколько иначе. Рассмотрим некоторую плоскость $x = 0$, лежащую вблизи границы слоя объемного заряда, но за его пределами (где зоны уже не искривлены), и обозначим через p (0) и n (0) концентрации дырок и электронов в этой плоскости (рис. 10.18). Положим, что толщина слоя объемного заряда (порядка длины экранирования $L_э$) меньше длин диффузии дырок и электронов L_p и L_n. Тогда приближенно рекомбинацией внутри слоя объемного заряда можно пренебречь. Это условие обычно выполняется для технически важных полупроводников — германия и (несколько хуже) кремния (хотя в других полупроводниках оно может и не иметь места). Пусть, далее, темп поверхностной рекомбинации дырок R_{ps} намного меньше каждого из потоков — диффузии и дрейфа дырок у поверхности. То же самое будем предполагать справедливым и для электронов. При этих условиях можно считать, что токи диффузии и токи дрейфа в слое объемного заряда приблизительно уравновешивают друг друга, т. е. что в дырочном и электронном газах имеется диффузионно-дрейфовое равновесие (ср. § VII.3). В этом случае между концентрациями p_s и p (0) существует однозначная связь, такая же, как и в условиях равновесия. В частности, для невырожденного полупроводника мы имеем

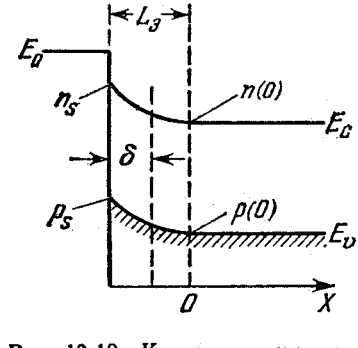


Рис. 10.18. К определению понятия скорости поверхностной рекомбинации.

$$p_s = p(0)\, e^{-Y_s}, \quad n_s = n(0)\, e^{Y_s}. \tag{5.3}$$

Тогда вместо (5.1) и (5.2) можно написать

$$R_{ps} = s_p [p(0) - p_0] = s_p \delta p(0), \qquad R_{ns} = s_n [n(0) - n_0] = s_n \delta n(0), \tag{5.4}$$

где p_0 и n_0 — равновесные концентрации в объеме полупроводника. Коэффициенты s_p и s_n имеют размерность скорости и обычно выражаются в см/с. Они получили название скорости поверхностной рекомбинации для дырок и, соответственно, для электронов. Вели-

чины s_p и s_n в общем случае не равны друг другу. Кроме того, они могут сами зависеть от δp и δn. Таким образом, при выполнении указанных двух условий мы можем под «поверхностью» полупроводника понимать собственно его поверхность и прилегающий к ней слой объемного заряда.

Если полупроводник находится в стационарном (но неравновесном) состоянии и нарушение равновесия мало, то можно ввести некоторую единую скорость поверхностной рекомбинации s, не зависящую от δp и δn. Действительно, в стационарном состоянии мы имеем

$$R_{ps} = R_{ns} = R_s = s_p\, \delta p\,(0) = s_n\, \delta n\,(0). \tag{5.5}$$

С другой стороны, рассматривая рекомбинацию в объеме, мы видели (§ IX.6), что темп объемной рекомбинации $R \sim (pn - p_0 n_0)$. В § 8 мы увидим, что аналогичный результат справедлив и для рекомбинации через поверхностные уровни, а именно $R_s \sim (n\,(0)\, p\,(0) - n_0 p_0)$ (ср. формулу (8.4), в которой при $L_э \ll L$ можно заменить в числителе $n_s p_s$ на $n\,(0)\, p\,(0)$). Поэтому можно написать

$$R_s = s\,\frac{n\,(0)\, p\,(0) - n_0 p_0}{n_0 + p_0}, \tag{5.6}$$

где s есть некоторый коэффициент, имеющий размерность скорости поверхностной рекомбинации. Полагая здесь $n\,(0) = n_0' + \delta n\,(0)$, $p\,(0) = p_0 + \delta p\,(0)$ и линеаризуя это выражение, имеем

$$R_s = s\,\frac{n_0}{n_0 + p_0}\,\delta p\,(0) + s\,\frac{p_0}{n_0 + p_0}\,\delta n\,(0). \tag{5.7}$$

Подставляя сюда из формулы (5.5) $\delta p\,(0) = R_s/s_p$ и $\delta n\,(0) = R_s/s_n$ и разрешая полученное уравнение относительно s, находим

$$s = s_p s_n \left(s_n\,\frac{n_0}{n_0 + p_0} + s_p\,\frac{p_0}{n_0 + p_0} \right)^{-1}. \tag{5.8}$$

В полупроводнике n-типа, когда выполняются условия $n_0 \gg p_0$, $s_n n_0 \gg s_p p_0$, мы имеем $s \simeq s_p$. При этом вторым слагаемым в формуле (5.7) обычно можно пренебречь. Напротив, в полупроводнике p-типа и при условиях $p_0 \gg n_0$, $s_p p_0 \gg s_n n_0$ получается $s \simeq s_n$ и, кроме того, первое слагаемое в формуле (5.7) мало по сравнению со вторым. Следовательно, в обоих случаях s совпадает со скоростью поверхностной рекомбинации неосновных носителей. Этот результат имеет простой физический смысл. Рекомбинация через поверхностные уровни, так же как и через объемные уровни, есть двухступенчатый процесс, состоящий из захвата дырки и захвата электрона. «Узким» местом процесса является захват неосновных носителей, концентрация которых при малом уровне возбуждения остается очень малой по сравнению с концентрацией основных носителей, и поэтому темп рекомбинации определяется темпом захвата неосновных носителей.

Положим теперь, что под влиянием каких-либо внешних воздействий, например при освещении полупроводника сильно поглощаемым светом, происходит поверхностная генерация электронно-дырочных пар. При этом, говоря о «поверхностной генерации», мы имеем в виду, что генерация происходит в тонком слое вблизи поверхности, толщина которого $\delta \lesssim L_э$ (рис. 10.18). Пусть темп поверхностной генерации есть g_s. Для случая оптической генерации

$$\delta \sim \frac{1}{\gamma}, \qquad g_s = I\nu,$$

где γ — коэффициент поглощения света, I — освещенность, а ν — квантовый выход внутреннего фотоэффекта (§ VII.4). Если все фотоны успевают поглотиться в слое объемного заряда (который мы включаем в понятие «поверхность»), то g_s не зависит от γ. Тогда условие баланса для неравновесных носителей заряда, например для дырок, у поверхности дает

$$g_s = \frac{1}{e} j_p(0) + s_p \delta p(0), \tag{5.9}$$

где $j_p(0)$ — плотность дырочного тока у поверхности. Аналогичное соотношение справедливо для неравновесных электронов. При этом поток частиц считается положительным, если он направлен от поверхности вглубь полупроводника. Если $g_s \neq 0$, то $j_p(0) > 0$ и, следовательно, возникает поток частиц от поверхности. При $g_s = 0$ мы имеем $j_p(0) < 0$. В этом случае появляется поток, направленный к поверхности, равный темпу исчезновения частиц вследствие рекомбинации.

В § VII.8 мы видели, что совместные диффузия и дрейф неравновесных дырок и электронов определяются неосновными носителями заряда. Поэтому соотношение (5.9), написанное для неосновных носителей заряда, определяет граничные условия задачи о вычислении распределения неравновесных дырок и электронов в полупроводниках конечных размеров.

Поверхностная рекомбинация проявляется, хотя и в разной степени, во всех неравновесных электронных процессах. Ниже рассмотрены некоторые примеры.

§ 6. Влияние поверхностной рекомбинации на фотопроводимость

а. Стационарная фотопроводимость при объемной однородной генерации. Будем считать, что образец имеет форму прямоугольной пластинки с ребрами $2A$, $2B$ и $2C$ (рис. 10.19), причем $A \ll B$, C. В этом случае концентрации δp и δn будут зависеть лишь от одной координаты z. Положим, что пластинка освещается со стороны широкой грани, а коэффициент поглощения света и толщина пластинки $2A$ таковы, что $2A\gamma \lesssim 1$. Тогда приближенно можно считать генерацию электронно-дырочных пар только объемной и притом однородной. В дальнейшем мы будем также предполагать, что концентрация ловушек в объеме мала, так что $\delta p = \delta n$ и, соответственно, объемное время жизни $\tau_p = \tau_n = \tau$. Скорости поверх-

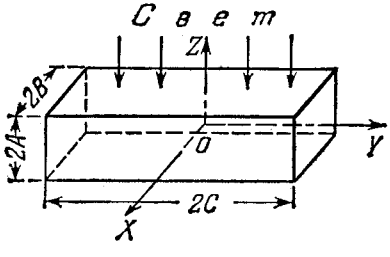


Рис. 10.19. К расчету фотопроводимости в тонких пластинках.

ностной рекомбинации неосновных носителей заряда на обеих гранях будем считать одинаковыми.

Уравнение непрерывности для данного случая имеет вид

$$D\frac{d^2p}{dz^2}-\frac{\delta p}{\tau}+g=0, \tag{6.1}$$

где вместо D нужно подставить коэффициент диффузии неосновных носителей. Граничные условия получаются из формулы (5.9):

$$\mp D\frac{d\delta p}{dz}\Big|_{z=\pm A}=s\delta p\,|_{z=\pm A}. \tag{6.2}$$

Здесь положено $g_s=0$, а s отвечает неосновным носителям. Тогда решение имеет вид

$$\delta p=g\tau\left(1-\frac{S}{S\,\mathrm{ch}\,\frac{A}{L}+\mathrm{sh}\,\frac{A}{L}}\,\mathrm{ch}\,\frac{z}{L}\right). \tag{6.3}$$

Здесь $L=\sqrt{D\tau}$ — длина диффузии, а

$$S=\frac{sL}{D} \tag{6.4}$$

есть безразмерная скорость поверхностной рекомбинации.

Если бы поверхностной рекомбинации не было ($S=0$), то концентрация фотодырок была бы равной $\delta p=g\tau$ и постоянной по объему. При $S\neq 0$ концентрация фотодырок уменьшается в каждой точке, а их распределение становится неоднородным (рис. 10.20).

Найдем еще среднюю концентрацию фотодырок $\overline{\delta p}$. Так как

$$\int_{-A}^{+A}\mathrm{ch}\,\frac{z}{L}\cdot dz=\frac{1}{2}\int_{-A}^{+A}\left(e^{z/L}+e^{-z/L}\right)dz=2L\,\mathrm{sh}\,\frac{A}{L},$$

то

$$\overline{\delta p}=\frac{1}{2A}\int_{-A}^{+A}\delta p\cdot dz=g\tau\left(1-\frac{S\,\frac{L}{A}\,\mathrm{sh}\,\frac{A}{L}}{S\,\mathrm{ch}\,\frac{A}{L}+\mathrm{sh}\,\frac{A}{L}}\right). \tag{6.5}$$

Концентрации фотоэлектронов δn и $\overline{\delta n}$ будут выражаться теми же формулами (6.3) и (6.5). Поэтому, если к торцам пластинки приложено напряжение, то в пластинке возникнет фототок, плотность которого будет

$$\Delta j=e\,(\mu_p+\mu_n)\,\overline{\delta p}\,\mathcal{E}, \tag{6.6}$$

где $\mathcal{E}$ — напряженность электрического поля внутри пластинки.

Из сказанного видно, что поверхностная рекомбинация уменьшает фотоответ, и тем сильнее, чем меньше толщина пластинки.

Для очень тонких пластинок ($A/L\ll 1$) можно положить

$$\mathrm{ch}\,\frac{A}{L}\simeq 1,\qquad \mathrm{sh}\,\frac{A}{L}\simeq\frac{A}{L},$$

и тогда формула (6.5) дает

$$\overline{\delta p} = \overline{\delta n} \simeq g\tau \frac{A/L}{S + \frac{A}{L}}.$$

При уменьшении A/L концентрации $\overline{\delta p}$ и $\overline{\delta n}$, а значит, и фотопроводимость неограниченно уменьшаются. Физический смысл этого результата заключается в том, что, уменьшая толщину пластинки, мы уменьшаем ее объем и, следовательно, полную генерацию электронно-дырочных пар. Рекомбинация же при этом остается прежней, так как поверхность пластинки не изменяется.

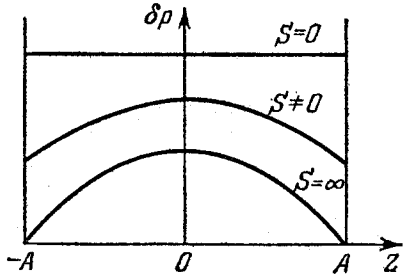


Рис. 10.20. Стационарное распределение фотодырок в тонкой пластинке при разной скорости поверхностной рекомбинации. Однородная генерация в объеме.

б. Стационарная фотопроводимость при поверхностной генерации. Рассмотрим теперь другой крайний случай, когда одна поверхность пластинки освещается сильно поглощаемым светом. В этом случае объемная генерация $g = 0$. Пластинку будем считать «толстой», т. е. $2A \gg L$. Ось Z направим в глубь пластинки и будем теперь отсчитывать z от освещенной поверхности. Тогда граничные условия будут

$$z = 0: \quad g_s = -D\frac{dp}{dz} + s\,\delta p; \qquad z \to \infty: \quad \delta p \to 0. \tag{6.7}$$

Решение уравнения непрерывности при втором граничном условии есть

$$\delta p = \delta n = \delta p\,(0)\, e^{-z/L}. \tag{6.8}$$

Значение граничной концентрации $\delta p\,(0)$ определяется первым граничным условием. Оно дает

$$g_s = \frac{D}{L}\,\delta p\,(0) + s\,\delta p\,(0),$$

а следовательно,

$$\delta p\,(0) = \frac{L}{D}\,\frac{g_s}{1+S}. \tag{6.9}$$

Если к торцам пластинки приложено напряжение и в пластинке имеется электрическое поле $\mathscr{E}$, параллельное освещаемой поверхности (скажем, в направлении оси Y, рис. 10.19), то полная сила фототока равна

$$\Delta i = e\,(\mu_p + \mu_n)\,2B\mathscr{E}\int_0^\infty \delta p \cdot dz = e\,(\mu_p + \mu_n)\,2BL\,\delta p\,(0)\,\mathscr{E},$$

где $2B$ — ширина пластинки. Поэтому, учитывая (6.9), имеем

$$\Delta i = e\,(\mu_p + \mu_n)\,2B\,\frac{g_s\tau}{1+S}\,\mathscr{E}. \tag{6.10}$$

И в этом случае поверхностная рекомбинация уменьшает величину фототока, а ее влияние определяется безразмерной величиной S.

Обобщение формулы (6.10) на случай $\tau_p \neq \tau_n$ (а также на случай наличия магнитного поля) будет дано в Приложении VII.

§ 7. Затухание фотопроводимости в тонких пластинках и нитевидных образцах

Поверхностная рекомбинация изменяет не только стационарную фотопроводимость, но и кинетику ее установления и затухания. Рассмотрим этот вопрос на примере тонкой пластинки с $A \ll B, C$ (рис. 10.19). Положим, что в момент времени $t = 0$ объемная генерация выключается, и найдем закон убывания во времени концентрации избыточных носителей. В этом случае мы должны рассматривать зависящее от времени уравнение непрерывности, которое в отсутствие внешнего электрического поля есть

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial z^2} - \frac{\delta p}{\tau_{об}}. \tag{7.1}$$

Здесь объемное время жизни мы отметили индексом «об», чтобы его не смешивать с временем затухания фотопроводимости. Граничные условия имеют тот же вид, что и в § 6, и выражаются формулой (5.9) при $z = \pm A$ и $g_s = 0$. Из симметрии задачи следует ожидать, что решение (7.1) будет симметрично относительно плоскости $z = 0$.

Нетрудно видеть, что общее решение уравнения (7.1) можно записать в виде

$$\delta p\,(t,\ z) = \sum_{m=1}^{\infty} \alpha_m e^{-t/\tau_m} \cos maz, \tag{7.2}$$

где m — целые числа, а α_m, τ_m и a — постоянные. Действительно, подставляя выражение (7.2) в уравнение (7.1), мы находим, что это уравнение удовлетворяется при условии, что между τ_m и a выполняется соотношение

$$\frac{1}{\tau_m} = \frac{1}{\tau_{об}} + Da^2m^2. \tag{7.3}$$

Полагая в (7.2) $t = 0$, мы также видим, что α_m суть коэффициенты разложения в ряд Фурье начального распределения носителей $\delta p\ (0,\ z)$.

Из формулы (7.2) видно, что затухание δp выражается суммой экспонент и, следовательно, сложным, не экспоненциальным законом. Однако, как показывает формула (7.3), характеристические времена τ_m быстро уменьшаются с увеличением m. Поэтому, если исключить начальный период затухания, приближенно можно пользоваться асимптотическим решением, оставив в формуле (7.2)

только один член с $m = 1$. Тогда

$$\delta p \simeq \alpha e^{-t/\tau} \cos az, \tag{7.2a}$$

где

$$\frac{1}{\tau} = \frac{1}{\tau_{об}} + \frac{1}{\tau_s}, \qquad \frac{1}{\tau_s} = Da^2. \tag{7.3a}$$

Здесь дополнительное слагаемое τ_s^{-1} выражает влияние поверхностной рекомбинации на постоянную времени затухания, а время τ_s часто называют *поверхностным временем жизни* (в отличие от объемного времени жизни $\tau_{об}$).

Постоянную a можно найти из граничных условий. Подставляя решение (7.2a) в соотношение (5.9) (при $g_s = 0$), мы получаем для a трансцендентное уравнение

$$\operatorname{tg}(aA) = \frac{s}{aD}, \tag{7.4}$$

которое требует численного расчета.

Задача упрощается в двух предельных случаях.

а) s очень велика, так что $s/(aD) \gg 1$. Тогда можно приближенно положить $aA \simeq \pi/2$ и формула (7.3a) для τ_s дает

$$\frac{1}{\tau_s} = \left(\frac{\pi}{2}\right)^2 \frac{D}{A^2} \qquad \left(\frac{s}{aD} \gg 1\right). \tag{7.5}$$

В этом случае τ_s вовсе не зависит от скорости поверхностной рекомбинации. Физический смысл этого результата состоит в том, что при очень большой s избыточные носители заряда практически мгновенно рекомбинируют на поверхности, так что «узким местом» процесса является приток частиц к поверхности. А он зависит только от коэффициента диффузии и толщины пластинки.

б) Выполняется условие $s/(aD) \ll 1$. Здесь можно приближенно считать $\operatorname{tg}(aA) \simeq aA$. Тогда уравнение (7.4) дает $a^2 = s/(DA)$ и мы получаем

$$\frac{1}{\tau_s} = \frac{s}{A} \qquad \left(\frac{s}{aD} \ll 1\right). \tag{7.6}$$

В рассматриваемом случае узким местом является поверхностная рекомбинация. Поэтому τ_s^{-1} пропорционально s и не зависит от коэффициента диффузии.

Формулы (7.3a) и (7.6) лежат в основе ряда методов измерения скорости поверхностной рекомбинации. Измеряя сначала объемное время жизни $\tau_{об}$, а затем постоянную затухания в тонкой пластинке τ, можно определить τ_s, а отсюда по формуле (7.6) найти s.

Рассуждая, как и выше, можно легко найти и постоянную затухания в «нитевидном» образце ($A \sim B$, но $C \gg A, B$). Тогда вместо формул (7.5) и (7.6) мы имеем

$$\frac{1}{\tau_s} = \left(\frac{\pi}{2}\right)^2 D\left(\frac{1}{A^2} + \frac{1}{B^2}\right), \tag{7.5a}$$

$$\frac{1}{\tau_s} = s\left(\frac{1}{A} + \frac{1}{B}\right). \tag{7.6a}$$

§ 8. Зависимость поверхностной рекомбинации от поверхностного потенциала

Опыт показывает, что скорость поверхностной рекомбинации s чрезвычайно сильно зависит от обработки поверхности. Так, например, для тонко шлифованной поверхности германия, протравленной в водном растворе H_2O_2 или в специальных смесях кислот (уксусной, азотной, фтористоводородной), при комнатной температуре $s \sim 1 \div 10$ см/с. Но после грубой шлифовки поверхности или помещения образца в вакуум s может возрасти до $10^4 \div 10^5$ см/с. Соответственно время жизни неравновесных носителей в тонких пластинках и в массивных слитках того же самого материала может сильно различаться. Если в очень чистых кристаллах германия τ может достигать $\sim 10^{-1}$ с, то в непротравленных пластинках толщиной ~ 1 мм оно падает до 10^{-6} с и меньше.

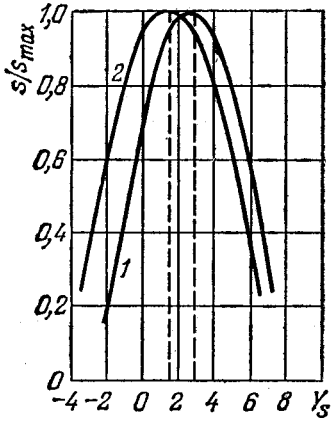


Рис. 10.21. Пример зависимости скорости поверхностной рекомбинации s/s_{max} от поверхностного потенциала в германии. *1* — вакуум, *2* — сухой кислород; 300 К [1].

Величина s сильно зависит от свойств окружающей среды: состава окружающих газов или имеющихся на поверхности различных покрытий, а также от температуры. Она сильно изменяется под действием внешнего поперечного электрического поля. Это показывает, что скорость поверхностной рекомбинации сильно зависит от поверхностного потенциала Y_s.

Зависимость s от Y_s можно получить, исследуя эффект поля и измеряя одновременно с Y_s (ср. § 3) еще и s (например, по скорости затухания фотопроводимости, § 7). Пример такой зависимости для германия показан на рис. 10.21. Величина s имеет максимум при некотором значении Y_{sm} и быстро убывает по обе стороны максимума. Кривые s (Y_s) оказываются различными при разной обработке поверхности.

Для объяснения наблюдаемых экспериментальных фактов рассмотрим подробнее процесс рекомбинации через поверхностные уровни. Для простоты положим, что только один из уровней является эффективным уровнем рекомбинации (в то время как другие уровни участвуют только в искривлении зон). Тогда, как и при рекомбинации в объеме, мы должны учесть четыре типа элементарных актов, показанных на рис. 10.22. Рассуждая так же, как и в § IX.4, мы найдем, что темп захвата электронов из объема на поверхность

можно выразить формулой

$$R_{sn} = \alpha_n \nu [(1 - f_s) n_s - n_1 f_s]. \tag{8.1}$$

Различие по сравнению с формулой (IX.4.7) заключается в том, что слева теперь стоит суммарный темп захвата на поверхность (на 1 см^2 в 1 с), а ν обозначает число поверхностных рекомбинационных уровней на 1 см^2 поверхности; f_s есть (неравновесная) степень заполнения поверхностного уровня электронами, а n_s — полная концентрация свободных электронов возле поверхности, α_n — коэффициент

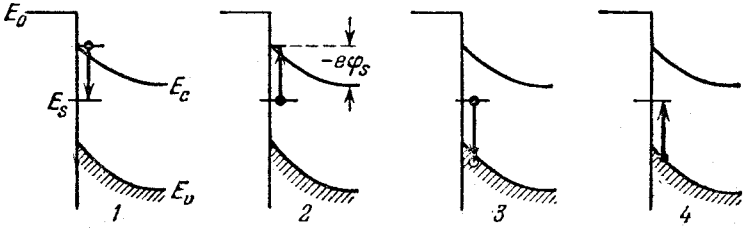


Рис. 10.22. Элементарные акты при рекомбинации через поверхностный уровень энергии. *1* — захват электрона, *2* — эмиссия электрона, *3* — захват дырки, *4* — эмиссия дырки.

захвата поверхностного уровня для электронов, а n_1 по-прежнему выражается формулой (V.9.5), где теперь, однако, $(E_c - E_t)$ есть энергия ионизации поверхностного уровня.

Для захвата дырок справедливо аналогичное выражение:

$$R_{sp} = \alpha_p \nu [f_s p_s - (1 - f_s) p_1], \tag{8.2}$$

где α_p — коэффициент захвата для дырок, а p_1, как и раньше, определяется формулой (V.9.13).

Ограничиваясь стационарными состояниями, мы имеем $R_{sn} = R_{sp} = R_s$. Отсюда определяется f_s:

$$f_s = \frac{\alpha_n n_s + \alpha_p p_1}{\alpha_n (n_s + n_1) + \alpha_p (p_s + p_1)}. \tag{8.3}$$

Подставляя это в одну из формул (8.1) или (8.2), мы находим

$$R_s = \nu \alpha_n \alpha_p \frac{p_s n_s - p_1 n_1}{\alpha_n (n_s + n_1) + \alpha_p (p_s + p_1)}, \tag{8.4}$$

где $n_1 p_1 = n_0 p_0 = n_i^2$, а n_0 и p_0 — концентрации в объеме, за пределами слоя объемного заряда.

В дальнейшем мы будем считать, что а) полупроводник не вырожден и концентрации неравновесных дырок и электронов подчиняются закону Больцмана (5.3); соответственно $p_s n_s = pn$, где p и n — полные концентрации на границе объемного заряда (в плоскости $x = 0$, рис. 10.18); б) $\delta p = \delta n$ (концентрация уровней мала); в) нару-

шение равновесия мало, $\delta n/(n_0+p_0) \ll 1$. Тогда, полагая $n = n_0 + \delta n$, $p = p_0 + \delta p$ и удерживая только малые первого порядка, находим

$$p_s n_s - p_1 n_1 = pn - p_0 n_0 \simeq (n_0 + p_0)\,\delta n,$$

$$R_s = \nu \alpha_n \alpha_p \frac{n_0 + p_0}{\alpha_n (n e^{Y_s} + n_1) + \alpha_p (p e^{-Y_s} + p_1)}\,\delta n. \qquad (8.4a)$$

Если при этом полупроводник не очень близок к собственному, то при малом нарушении равновесия можно еще заменить n и p на n_0 и p_0. Тогда для скорости поверхностной рекомбинации получаем

$$s = \frac{R_s}{\delta n} = \nu \alpha_n \alpha_p \frac{n_0 + p_0}{\alpha_n (n_0 e^{Y_s} + n_1) + \alpha_p (p_0 e^{-Y_s} + p_1)}. \qquad (8.5)$$

Формула (8.5) в ряде случаев удовлетворительно описывает общий характер зависимости s от Y_s. В частности, она показывает, в соответствии с экспериментальными данными, что s имеет максимум при некотором значении $Y_s = Y_{sm}$. Приравнивая нулю производную от знаменателя по Y_s в формуле (8.5), находим

$$Y_{sm} = \ln \left(\frac{p_0}{n_0} \frac{\alpha_p}{\alpha_n} \right)^{1/2}. \qquad (8.6)$$

Определяя на опыте Y_{sm} и зная отношение $p_0/n_0 = n_i^2/n_0^2 = p_0^2/n_i^2$ в объеме (которое легко измеряется), можно найти отношение коэффициентов захвата α_p/α_n для данного поверхностного уровня. А отсюда можно сделать заключение и о том, является ли данный уровень акцепторным или донорным. Действительно, так как захваты при наличии кулоновского притяжения обычно характеризуются большими значениями α, то значения $\alpha_p/\alpha_n \gg 1$, скорее всего указывают на то, что возможные зарядные состояния уровня — отрицательное и нейтральное, а следовательно, уровень является акцепторным. Напротив, значения $\alpha_p/\alpha_n \ll 1$ являются указанием на донорный характер уровня.

Анализ данных по поверхностной рекомбинации показывает, что из всей совокупности поверхностных уровней только часть участвует в процессах рекомбинации. Остальные же уровни могут обмениваться носителями заряда только с одной из зон и являются «уровнями прилипания» для электронов или для дырок.

§ 9. Ток насыщения диодов

Поверхностная рекомбинация может сильно влиять на параметры полупроводниковых приборов с p — n-переходами. Поясним это на простейшем примере диода. Его мы будем представлять (как это часто делают практически) в виде тонкой пластинки полупроводника, для определенности n-типа, в которую вплавлена капля металла

(радиуса a), являющегося акцептором (рис. 10.23). Толщину пластинки $2A$ будем считать намного меньшей эффективной длины диффузии дырок $\lambda_p = \sqrt{D_p\tau}$. Здесь τ — результирующее время жизни, определяемое формулой (7.3а). Как мы знаем (§ VIII.1), обратный ток насыщения диодов образован тепловой генерацией неосновных носителей. Однако в данном случае это будут не только дырки, возникающие в объеме пластинки, но и дырки, генерируемые с поверхностных уровней на обеих поверхностях пластинки (рис. 10.23). Из формулы (5.4) следует, что с каждой единицы поверхности в единицу времени генерируется sp_0 дырок. Поэтому дополнительный ток насыщения, обусловленный поверхностью, есть

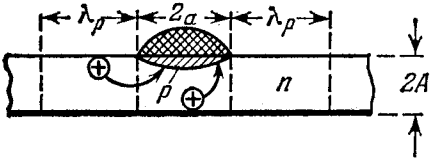


Рис. 10.23. К вычислению тока насыщения диода при учете поверхностной рекомбинации.

$$i'_s = esp_0\,[2\pi\,(\lambda_p + a)^2 - \pi a^2]. \qquad (9.1)$$

В случае тонких пластинок τ_s может быть значительно меньше $\tau_{об}$. Так, например, при $s \sim 10^3$ см/с и $A \sim 10^{-2}$ см мы имеем $\tau_s \sim 10^{-5}$ с, в то время как $\tau_{об}$ часто бывает в десятки раз больше. Поэтому в формуле (9.1) обычно можно считать, что $\lambda_p \simeq \sqrt{D_p\tau_s}$. Тогда весь избыточный обратный ток определяется только свойствами поверхности. Оценки, в согласии с опытом, показывают, что избыточный ток i'_s в диодах с тонкой базой может быть намного больше тока насыщения, обусловленного генерацией носителей в объеме.

Так как s зависит от поверхностного потенциала Y_s, то избыточный ток насыщения изменяется при изменении обработки поверхности и свойств окружающей среды.

Поверхностная рекомбинация существенно влияет на свойства и других полупроводниковых приборов, например на свойства транзисторов с p — n-переходами. А изменение состояния поверхности со временем приводит к нестабильности параметров приборов. Поэтому надежная стабилизация поверхности и ее защита от внешних влияний является одной из важных задач технологии полупроводниковых приборов.

Глава XI

ФОТОЭЛЕКТРОДВИЖУЩИЕ СИЛЫ

§ 1. Роль неосновных носителей

При освещении полупроводника не только изменяется его электропроводность, но возникают также электродвижущие силы. Общая причина возникновения фотоэдс, по крайней мере в наиболее важных известных случаях, одна и та же и заключается в диффузии фотоэлектронов и фотодырок. Диффундируя от места своего возникновения, неравновесные носители заряда создают направленные потоки, что эквивалентно появлению сил некулоновского происхождения, или, иначе, «сторонних» сил.

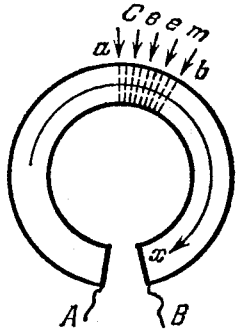


Рис. 11.1. Цепь с частично освещенным полупроводником.

Несмотря на общую причину происхождения, оказывается удобным говорить о разных типах фотоэдс в зависимости от особенностей полупроводниковой структуры и условий опыта. Однако сначала мы остановимся на общем условии, необходимом для возникновения фотоэдс.

Чтобы сделать рассуждения наиболее простыми, рассмотрим полупроводник в виде кольца, часть которого *ab* освещается (рис. 11.1). Кольцо имеет узкий разрез с одинаковыми металлическими электродами A и B для измерения эдс. Полупроводник может быть неоднородным и даже может состоять из нескольких различных веществ. Однако мы будем считать, что разрез сделан в таком месте кольца, где химический состав полупроводника одинаков, а концентрации неравновесных носителей заряда δp и δn равны нулю. При этих условиях контакты не будут давать вклада в эдс. Будем также считать, что подвижности электронов и дырок не изменяются при освещении (условие этого см. в § VII.4). Тогда нетрудно видеть, что фотоэдс в *стационарном* режиме возможна лишь в том случае, когда свет генерирует носители заряда *обоих* знаков.

Пусть концентрации δp и δn зависят только от одной координаты x (толщина кольца мала по сравнению с длинами диффузии дырок и электронов). Тогда в результате диффузии дырок и электронов, созданных светом, в неосвещенные части полупроводника в кольце

появятся токи с плотностями

$$j_p = \sigma_p \mathcal{E} - eD_p \frac{dp}{dx}, \qquad j_n = \sigma_n \mathcal{E} + eD_n \frac{dn}{dx}.$$

Полная плотность тока будет

$$j = j_p + j_n = \sigma \left(\mathcal{E} + e \frac{D_n \frac{dn}{dx} - D_p \frac{dp}{dx}}{\sigma} \right),$$

где $\sigma = e(\mu_n n + \mu_p p)$ — полная электропроводность в данном месте. С другой стороны, согласно закону Ома для проводника с эдс, можно написать

$$j = \sigma (\mathcal{E} + \mathcal{E}^*),$$

где $\mathcal{E}^*$ — напряженность поля сторонних сил. Сравнивая оба выражения для j, находим

$$\mathcal{E}^* = \frac{D_n \frac{dn}{dx} - D_p \frac{dp}{dx}}{\mu_p p + \mu_n n}. \tag{1.1}$$

Отметим, что $\mathcal{E}^*$ совпадает по величине с полем амбиполярной диффузии, но отличается от него по знаку (ср. формулу (VII.8.2)).

Полная фотоэдс в кольце, согласно общему определению эдс, равна

$$V_0 = \oint \mathcal{E}^* \, dx = \oint \frac{D_n \frac{dn}{dx} - D_p \frac{dp}{dx}}{\mu_p p + \mu_n n} \, dx, \tag{1.2}$$

где интегрирование производится вдоль всего кольца. Если в кольце сделать разрез в той его части, где уже не имеется неравновесных носителей заряда, то между концами A и B (рис. 11.1) появится разность потенциалов, выражаемая формулой (1.2). Исследуем теперь подробнее полученное общее выражение для фотоэдс.

Неоднородный полупроводник. Освещения нет. В этом случае $n_0(x)$ и $p_0(x)$ суть равновесные концентрации, которые связаны между собой законами равновесной статистики (гл. V). В частности, для невырожденного полупроводника

$$n_0 p_0 = n_i^2, \qquad \frac{1}{p_0} \frac{dp_0}{dx} = -\frac{1}{n_0} \frac{dn_0}{dx}. \tag{1.3}$$

Кроме того, для μ и D справедливо соотношение Эйнштейна: $D/\mu = kT/e$. Поэтому

$$D_n \frac{dn_0}{dx} - D_p \frac{dp_0}{dx} = \frac{kT}{e} (\mu_p p_0 + \mu_n n_0) \frac{1}{n_0} \frac{dn_0}{dx},$$

и, следовательно, эдс равна

$$V_0 = \frac{kT}{e} \oint \frac{dn_0}{n_0} = 0, \tag{1.4}$$

так как под знаком интеграла стоит полный дифференциал. Отметим, что в окончательной формуле выпали все индивидуальные характеристики полупроводника (n_i, μ_p и μ_n). Поэтому полученный результат справедлив и для любой комбинации различных полупроводников, если только система находится в термодинамическом равновесии.

Однородный полупроводник. Освещение есть. Положим

$$p = p_0 + \delta p, \qquad n = n_0 + a\,\delta p, \qquad \frac{dn_0}{dx} = \frac{dp_0}{dx} = 0.$$

Здесь введением множителя a мы учли возможное прилипание носителей на ловушки. Тогда

$$D_n \frac{dn}{dx} - D_p \frac{dp}{dx} = \frac{kT}{e}(a\mu_n - \mu_p)\frac{d(\delta p)}{dx}$$

и для эдс получается

$$V_0 = \frac{kT}{e}(a\mu_n - \mu_p)\oint \frac{d(\delta p)}{\mu_n n_0 + \mu_p p_0 + (\mu_p + a\mu_n)\,\delta p}. \tag{1.5}$$

Здесь опять подынтегральное выражение есть полный дифференциал, а именно — от функции

$$\frac{\ln[\mu_n n_0 + \mu_p p_0 + (\mu_p + a\mu_n)\,\delta p]}{\mu_p + a\mu_n},$$

и поэтому снова $V_0 = 0$.

Неоднородный полупроводник. Свет создает только основные носители. Будем рассматривать, для определенности, полупроводник n-типа. Тогда, учитывая соотношение (1.3), в формуле (1.2) можно считать

$$D_n \frac{dn}{dx} \gg D_p \frac{dp}{dx} = D_p \frac{dp_0}{dx}, \qquad \mu_n n \gg \mu_p p = \mu_p p_0.$$

Поэтому

$$V_0 = \frac{kT}{e}\oint \frac{dn}{n} = 0, \tag{1.6}$$

так как мы опять приходим к полному дифференциалу.

Таким образом, для возникновения фотоэдс необходимо, чтобы подынтегральное выражение в формуле (1.2) не являлось полным дифференциалом. А для этого, как мы видим, полупроводник, во-первых, должен быть неоднородным и, во-вторых, необходимо, чтобы *свет создавал такие носители заряда, знак которых противоположен знаку темновых носителей заряда.*

Во избежание недоразумения в дальнейшем, отметим, что, говоря о невозможности фотоэдс в однородных полупроводниках, мы имеем в виду однородность в той области, где существуют неравновесные

фотодырки и фотоэлектроны. Это, однако, не относится к случаю однородного кристалла конечных размеров, граница которого освещается (см. § 2). Здесь в области существования фотоносителей имеется «неоднородность», создаваемая самой границей, и поэтому фотоэдс возможна.

§ 2. Фотоэдс в однородных полупроводниках

Рассмотрим полупроводник прямоугольной формы, одна из граней которого освещается сильно поглощаемым светом. Для определенности возьмем полупроводник n-типа. Для простоты расчетов мы будем также предполагать, что а) концентрация ловушек мала ($\delta n = \delta p$) и б) освещение не очень сильное, так что электропроводность σ при освещении в любой точке мало отличается от темновой электропроводности σ_0. Тогда формула (1.2) дает

$$V_0 = \frac{e}{\sigma_0}(D_n - D_p)\int_0^d dp = \frac{e}{\sigma_0}(D_n - D_p)\,[\delta p\,(d) - \delta p\,(0)], \qquad (2.1)$$

где d — толщина образца, а $\delta p\,(0)$ и $\delta p\,(d)$ — концентрации фотоносителей у освещенной и, соответственно, у задней поверхности. Здесь, очевидно, достаточно интегрировать только по толщине пластинки, так как в остальной части цепи $\mathcal{E}^* = 0$.

Если d хотя бы в несколько раз превышает длину диффузии L, то $\delta p\,(d) \ll \delta p\,(0)$. Избыточную концентрацию для «толстой» пластинки мы уже вычислили раньше (§ X.6):

$$\delta p\,(0) = \frac{L}{D}\,\frac{g_s}{1+S}, \qquad (2.2)$$

где D — коэффициент амбиполярной диффузии, а остальные обозначения имеют прежний смысл. Поэтому окончательно получаем

$$V_0 = -\frac{e}{\sigma_0}\,\frac{D_n - D_p}{D_p}\,L\,\frac{g_s}{1+S}. \qquad (2.3)$$

На рис. 11.2 показаны знаки заряда граней для случая $D_n > D_p$, направления поля амбиполярной диффузии $\mathcal{E}$ и сторонней силы $\mathcal{E}^*$. Электроны, диффундирующие быстрее, заряжают нижнюю поверхность отрицательно, а на освещенной грани появляется положительный заряд. Рассмотренная фотоэдс, возникающая вследствие различия коэффициентов диффузии электронов и дырок, получила название эдс *Дембера*.

Величина эдс Дембера невелика. Примем для оценки $(D_n - D_p)/D_p \simeq 1$, $L \sim 0{,}1$ см, $\sigma_0 \sim 1$ Ом$^{-1}$см$^{-1}$, что типично, например, для германия. Безразмерную скорость поверхностной рекомбинации положим $S \ll 1$. Пусть, далее, образец освещается

прямым несфокусированным солнечным излучением. Расчет показывает, что в солнечном излучении число фотонов с энергией $\hbar\omega \geqslant 0{,}65$ эВ, падающих на 1 см^2 в 1 с у поверхности Земли при ясной атмосфере, равно примерно $2 \cdot 10^{17}$ см$^{-2}$ с$^{-1}$. Следовательно, при квантовом выходе $\nu = 1$, $g_s \sim 10^{17}$ см$^{-2}$ с$^{-1}$. Тогда по формуле (2.3) находим, что даже при этой сильной освещенности $V_0 \sim 10^{-3}$ В.

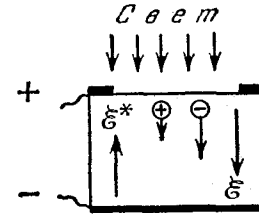


Рис. 11.2. Происхождение эдс Дембера.

Хотя эдс такой величины легко измерить, в действительности экспериментальное определение эдс Дембера в чистом виде сопряжено с большими трудностями. Это происходит по той причине, что при освещении, как правило, возникает еще и вентильная фотоэдс в потенциальных барьерах (§ 5), которая намного больше эдс Дембера. Потенциальные же барьеры всегда существуют как в контактах, так и вблизи свободной поверхности полупроводника вследствие поверхностного искривления зон. Поэтому эдс Дембера обычно играет роль сопутствующего эффекта в различных фотоэлектрических явлениях.

§ 3. Объемная фотоэдс

Объемная, или, иначе, распределенная, фотоэдс возникает в неоднородных полупроводниках, в которых градиент удельного сопротивления отличен от нуля *). Этот тип фотоэдс мы рассмотрим на примере тонкой пластинки (или нитевидного образца), на поверхность которой проецируется световое изображение узкой освещенной щели (рис. 11.3, *а*). Световая щель удалена от торцов образца по крайней мере на несколько диффузионных длин, так что фотоэлектроны и фотодырки рекомбинируют, не достигая торцов, и поэтому не вызывают эдс в контактах. Опыт показывает, что между концами образца появляется напряжение, которое тем больше, чем больше градиент удельного сопротивления в месте положения световой щели.

Возникновение этой эдс поясняет рис. 11.3, *б*. Положим, что сопротивление изменяется вследствие изменения концентрации темновых носителей, вызванного неравномерным распределением доноров и акцепторов вдоль образца. Это значит, что энергетические зоны будут наклонными относительно постоянного уровня Ферми, или, другими словами, что в образце будет существовать внутреннее электрическое поле. Без освещения это поле, конечно, не

*) Объемная фотоэдс была обнаружена и исследована В. Е. Лашкаревым и, независимо, Я. Тауцем.

вызывает тока, так как обусловленный им ток дрейфа в точности компенсируется током диффузии. Однако при освещении это равновесие нарушается и возникают потоки фотоэлектронов и

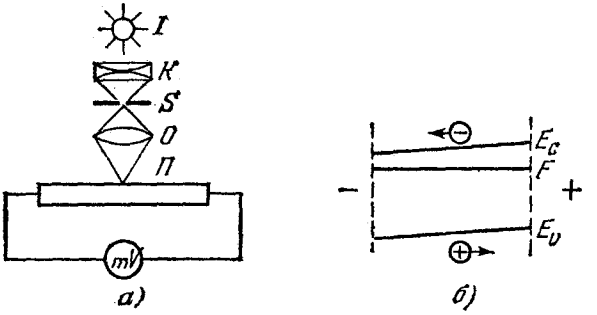


Рис. 11.3. *а*) Наблюдение объемной фотоэдс. *I* — источник света, *K* — конденсор, *S* — щель, *O* — объектив, *П* — полупроводник, *mV* — милливольтметр. *б*) Направления движения фотоэлектронов и фотодырок.

фотодырок, направленные в разные стороны. Если, как показано на рисунке, градиент сопротивления направлен слева направо, то электроны будут двигаться к левому концу образца и заряжать его отрицательно, а дырки — к правому концу, создавая на нем положительный заряд.

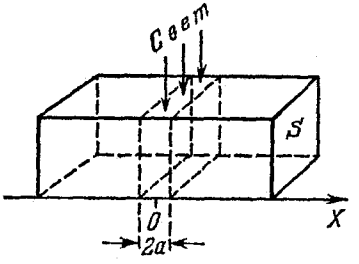


Рис. 11.4. К расчету объемной фотоэдс.

В дальнейшем мы опять будем говорить о полупроводнике *n*-типа. При расчете величины фотоэдс мы будем предполагать, что: 1) δp и δn зависят только от одной координаты x (рис. 11.4); 2) генерация электронно-дырочных пар происходит в тонком слое шириной $2a$, приблизительно равной ширине световой щели, причем $2a \ll$ длины диффузии L (рекомбинация в освещаемом слое не происходит); 3) концентрация избыточных носителей в этом слое постоянна (это приближение существенно упрощает расчет и не вносит большой ошибки); 4) освещение слабое: $\delta p/n_0 \ll 1$. Соответственно мы будем считать, что и при освещении $\mu_n n \gg \mu_p p$. Тогда по формуле (1.2) имеем

$$V_0 = \oint \frac{D_n \frac{dn}{dx} - D_p \frac{dp}{dx}}{\mu_n n} dx = V_{01} + V_{02}.$$

Первое слагаемое

$$V_{01} = \frac{D_n}{\mu_n} \oint \frac{dn}{n} = 0,$$

так как подынтегральная функция есть полный дифференциал. При вычислении второго слагаемого мы положим $n \simeq n_0$ (слабое освещение). Далее, мы учтем, что при темновой концентрации дырок $p = p_0(x)$ эдс $V_{02} = 0$. Поэтому

$$V_0 = -\frac{D_p}{\mu_n} \oint \frac{\frac{d(\delta p)}{dx}}{n_0(x)}\, dx.$$

Введем, наконец, в это выражение темновое удельное сопротивление ρ_0 и плотность тока диффузии избыточных дырок j_p:

$$\rho_0(x) = \frac{1}{e\mu_n n_0(x)}, \qquad j_p(x) = -eD_p \frac{d(\delta p)}{dx}.$$

Тогда для V_0 можно написать

$$V_0 = \int_{-\infty}^{-a} j_p(x)\,\rho_0(x)\,dx + \int_a^{\infty} j_p(x)\,\rho_0(x)\,dx.$$

Здесь мы перешли от интегрирования по замкнутому контуру к бесконечным пределам, так как $j_p(x)$ быстро затухает на расстояниях $\sim L$ от освещенного слоя.

Так как $\delta p(x) \sim \exp(-x/L_p)$ (ср. § VII.9), то по такому же закону изменяется и $j_p(x)$. Поэтому мы имеем:

$$\begin{aligned} x > a: &\quad j_p = j_p(a) \exp\left(-\frac{x-a}{L_p}\right); \\ x < -a: &\quad j_p = -j_p(-a) \exp\frac{x+a}{L_p}, \end{aligned} \tag{3.1}$$

причем из симметрии задачи ясно, что $j_p(-a) = j_p(a)$. Тогда

$$V_0 = -j_p(a) \int_{-\infty}^{-a} \rho_0(x) \exp\left(\frac{x+a}{L_p}\right) dx + j_p(a) \int_a^{\infty} \rho_0(x) \exp\left(-\frac{x-a}{L_p}\right) dx.$$

Заменяя в первом интеграле переменную x на $-x$, написанное выражение можно представить в более удобном виде:

$$V_0 = j_p(a) \int_a^{\infty} \exp\left(-\frac{x-a}{L_p}\right) [\rho_0(x) - \rho_0(-x)]\, dx. \tag{3.2}$$

Распределение $\rho_0(x)$ в узкой области вблизи освещенного слоя можно аппроксимировать линейным законом и приближенно положить

$$\rho_0(x) - \rho_0(-x) = 2\left[\frac{d\rho}{dx}\right]_{x=0} x. \tag{3.3}$$

Тогда получается

$$V_0 = 2j_p(a)\frac{d\rho_0}{dx}\int_a^\infty \exp\left(-\frac{x-a}{L_p}\right)x\,dx = \\ = 2j_p(a)\frac{d\rho_0}{dx}(L_p a + L_p^2) \simeq 2j_p(a)\frac{d\rho_0}{dx}L_p^2, \qquad (3.4)$$

так как $a \ll L_p$.

И, наконец, значение тока на границах $j_p(a)$ получается непосредственно из условия баланса для фотодырок. Если I есть полное число электронно-дырочных пар, генерируемых светом во всем освещенном слое в 1 с (пропорциональное интенсивности света), то

$$2j_p(a)S = eI, \qquad (3.5)$$

где S — сечение образца. Поэтому, учитывая формулы (3.4) и (3.5), находим окончательно

$$V_0 = \frac{e}{S}I \cdot L_p^2\frac{d\rho_0}{dx}. \qquad (3.6)$$

Подчеркнем, что здесь $d\rho_0/dx$ есть градиент темнового сопротивления в месте нахождения световой щели.

Для дырочного полупроводника мы получили бы, очевидно, такую же формулу, с тем отличием, что вместо длины диффузии дырок вошла бы длина диффузии электронов (неосновных носителей).

Оценим порядок величины объемной фотоэдс. Положим, что мы имеем сравнительно чистый германий при комнатной температуре. Тогда в качестве типичного значения можно принять $L_p \sim 0{,}1$ см. Сечение образца примем $S \sim 10^{-2}$ см^2. Пусть, далее, ρ_0 меняется на 10 Ом·см на 1 см длины, т. е. $d\rho_0/dx = 10$ Ом. Тогда, принимая $I \sim 10^{15}$ с$^{-1}$, по формуле (3.6) получаем $V_0 \sim 1$ мВ.

Таким образом, при слабом освещении объемная фотоэдс пропорциональна градиенту темнового сопротивления. Этот вывод хорошо согласуется с опытом. Для его проверки нужно, с одной стороны, независимо определить распределение сопротивления по длине образца. Это можно легко сделать, пропуская через образец слабый постоянный ток и исследуя распределение потенциала вдоль образца с помощью тонкого металлического зонда. С другой стороны, нужно исследовать зависимость фотоэдс от положения светового зонда. Опыт показывает, что распределение величины и знака фотоэдс хорошо повторяет распределение градиента сопротивления. Такое зондирование с одновременным измерением фотонапряжения между концами образца является полезным практическим приемом для определения степени однородности полупроводника.

§ 4. Вентильная фотоэдс

Наиболее важным типом фотоэдс является так называемая *вентильная фотоэдс*. Она возникает в том случае, когда свет генерирует носители заряда вблизи потенциального барьера, т. е. при освещении контактов металл — полупроводник и p—n-переходов. Однако, вследствие сложной микроструктуры контактов полупроводника с металлом, мы ограничимся в дальнейшем наиболее ясным случаем p—n-переходов.

Следует различать две принципиально отличные схемы включения фотоэлемента с p—n-переходом. В первой из них фотоэлемент непосредственно замкнут на внешнюю нагрузку (рис. 11.5, *а*) и

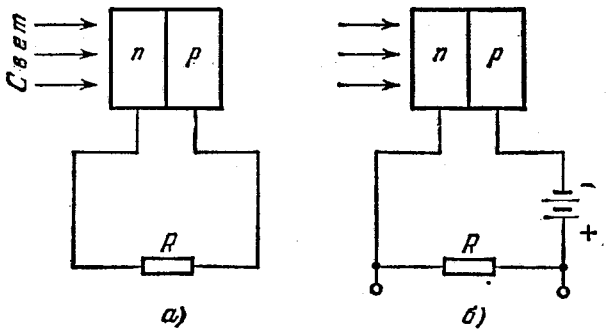


Рис. 11.5. Схемы включения фотоэлемента (*а*) и фотодиода (*б*).

является преобразователем световой энергии в электрическую. Такую схему включения мы будем называть «схемой фотоэлемента». Другая схема (рис. 11.5, *б*) содержит во внешней цепи еще источник тока, включенный так, чтобы на p—n-переходе было напряжение запорного знака. Это напряжение выбирают по возможности большим, но, конечно, меньше пробойного напряжения p—n-перехода. Такая схема получила название «схемы фотодиода». При этой схеме включения существующий в цепи ток (который в темноте есть ток насыщения p—n-перехода и в хороших диодах очень мал) сильно изменяется при освещении и, соответственно, изменяется и падение напряжения на нагрузочном сопротивлении R. При правильном выборе напряжения источника и внешнего сопротивления величина электрического сигнала может быть сделана намного большей, чем в схеме фотоэлемента, и поэтому фотодиоды широко применяются для регистрации и измерения световых сигналов. В дальнейшем мы будем рассматривать фотоэлементы только как источники тока.

Величина вентильной фотоэдс на несколько порядков больше, чем эдс Дембера и объемная фотоэдс. Коэффициент полезного действия вентильных фотоэлементов может быть сделан достаточно большим (см. ниже), и поэтому они находят техническое применение для прямого преобразования световой энергии в электрическую. В частности, вентильные фотоэлементы используются для создания солнечных батарей, широко применяемых на искусственных спутниках Земли и космических кораблях.

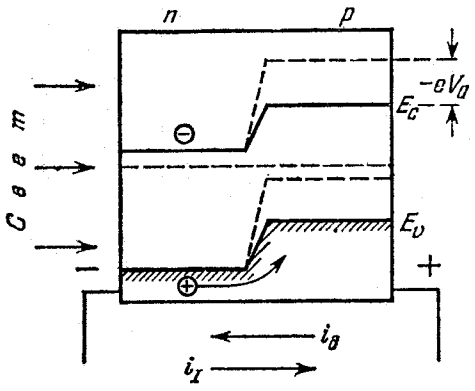


Рис. 11.6. Происхождение вентильной фотоэдс. Пунктиром показаны края энергетических зон в темноте.

Физическая причина возникновения вентильной фотоэдс заключается в том, что потенциальный барьер p—n-перехода разделяет потоки фотодырок и фотоэлектронов. Положим, что сильно поглощаемое излучение падает на одну из граней кристалла, параллельную плоскости p—n-перехода (рис. 11.6). Для определенности будем считать, что освещается n-область, и будем считать сначала, что фотоэлемент разомкнут. Фотоэлектроны и фотодырки будут диффундировать в глубь кристалла, и некоторая их доля, не успевшая рекомбинировать (на поверхности и в объеме), достигнет p—n-перехода. Однако для основных носителей — электронов — в p—n-переходе существует потенциальный барьер, и поэтому практически все они не пройдут в p-область. Напротив, для неосновных носителей — дырок — потенциального барьера нет и все достигшие перехода дырки будут затягиваться полем перехода в глубь кристалла, создавая ток i_I. Если g_s — темп поверхностной генерации пар, а β — доля фотодырок, дошедших до перехода без рекомбинации, то

$$i_I = eg_s\beta S, \tag{4.1}$$

где S — освещаемая площадь. Вследствие появления «светового тока» i_I p-область будет заряжаться положительно, а n-область — отрицательно, и между электродами элемента появится разность потенциалов. Поэтому в фотоэлементе, который представляет собой обычный диод с p—n-переходом, возникнет еще дополнительный ток $i_д$, обусловленный инжекцией дырок в n-область и электронов в p-область и направленный противоположно току i_I. Для «идеального» диода (нет рекомбинации в самом p—n-переходе, токи утечки

малы, падением напряжения в толще кристалла можно пренебречь) этот ток дается формулой (VIII.1.4). В результате между разомкнутыми электродами фотоэлемента установится такое напряжение V_0, при котором полный ток $i = i_I - i_д = 0$.

Если электроды фотоэлемента замкнуты на внешнюю нагрузку, то напряжение между ними u будет меньше V_0 и токи i_I и $i_д$ уже не будут компенсировать друг друга. Поэтому в цепи возникнет ток

$$i = i_I - i_д = i_I - i_s(e^{\alpha u} - 1), \quad (4.2)$$

где $\alpha = e/kT$, а i_s есть ток насыщения. Соотношение (4.2) является основным в теории вентильных фотоэлементов.

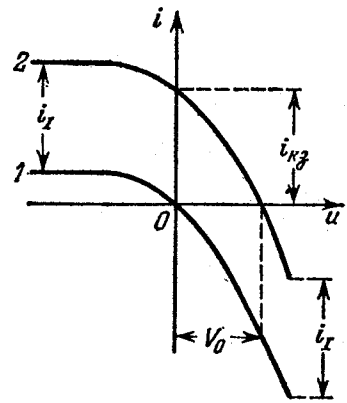


Рис. 11.7. Вольтамперная характеристика идеального фотоэлемента в темноте (*1*) и при освещении (*2*).

Зависимость i от u для идеального вентильного фотоэлемента показана на рис. 11.7. Кривая *1* дает зависимость темнового тока от напряжения и представляет собой обычную вольтамперную характеристику тонкого p—n-перехода. При освещении все точки этой кривой сдвигаются вверх на одну и ту же величину i_I (кривая *2*). Вследствие этого при $i = 0$ (разомкнутая цепь) напряжение на фотоэлементе не равно нулю, а значит, в нем развивается некоторая эдс V_0. Аналогично, при $u = 0$ (короткое замыкание) в цепи существует ток короткого замыкания. Его величина равна световому току i_I.

Из сказанного видно, что свойства идеального вентильного фотоэлемента можно описать с помощью эквивалентной схемы из параллельно соединенных генератора тока и идеального диода (рис. 11.8, *а*). При этом, согласно закону Кирхгофа, $i_I = i_д + i$, что и приводит к формуле (4.2).

Полагая в формуле (4.2) $i = 0$ (разомкнутая цепь), находим, что эдс элемента равна

$$V_0 = \frac{kT}{e} \ln\left(1 + \frac{i_I}{i_s}\right). \quad (4.3)$$

С увеличением освещенности (увеличением i_I) эдс увеличивается, однако не пропорционально i_I, а по логарифмическому закону. Конечно, это увеличение происходит не беспредельно, так как при возрастании освещенности уменьшается высота потенциального барьера в p—n-переходе, и, когда барьер становится порядка kT, увеличение эдс прекращается. Отсюда следует, что эдс не может превышать контактную разность потенциалов между p- и n-обла-

стями. Для наиболее выгодного случая, когда (без освещения) уровень Ферми расположен вблизи краев энергетических зон, $V_{0max} \simeq E_g/e$. Это дает, например, для германия при комнатной температуре $V_{0\,max} \simeq 0{,}6$ В, а для кремния $V_{0max} \simeq 1$ В.

Оценим еще величину тока короткого замыкания. Для этого рассмотрим опять конкретный пример германия и прямого, несфокусированного солнечного излучения ($g_s \sim 10^{17}$ см$^{-2}$с$^{-1}$). Полагая для оценки в формуле (4.1) $\beta = 1$, мы находим $i_I/S \sim 10^{-19} \cdot 10^{17} \sim 10^{-2}$ А/см^2 $= 10$ мА/см^2.

В реальном фотоэлементе необходимо учитывать, во-первых, распределенное сопротивление p- и n-областей и, во-вторых, наличие утечки по поверхности. Поэтому мы приходим к более сложной

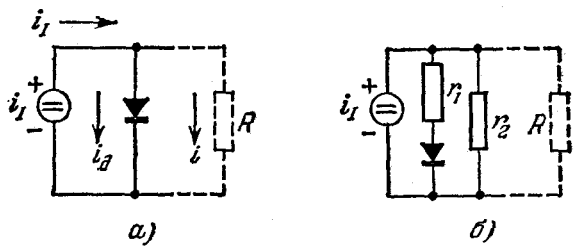


Рис. 11.8. Эквивалентные схемы идеального (*а*) и реального (*б*) вентильных фотоэлементов.

эквивалентной схеме, показанной на рис. 11.8, *б*, где r_1 есть сопротивление кристалла в обеих областях, а r_2 — сопротивление утечки.

Наличие сопротивления r_1 проявляется в зависимости тока короткого замыкания от освещенности. Согласно формуле (4.1) ток короткого замыкания идеального фотоэлемента $i_{кз} = i_I$ должен быть пропорционален g_s, т. е. освещенности. Это действительно и наблюдается в широкой области изменений g_s. Однако при очень больших освещенностях увеличение $i_{кз}$ становится слабее. Это объясняется тем, что при коротком замыкании внешних контактов элемента на самом p—n-переходе напряжение не равно нулю, а равно ir_1, причем потенциал p-области оказывается положительным относительно n-области. Поэтому появляется ток диода $i_д$, противоположный световому току i_I, что и приводит к уменьшению $i_{кз}$.

Приведенные выше простые рассуждения объясняют происхождение вентильной фотоэдс и общие ее особенности. Однако при этом мы, во-первых, не выяснили, от чего зависит коэффициент β, учитывающий потери на рекомбинацию. Во-вторых, мы ничего не говорили о токе насыщения i_s. Хотя в § VIII.1 мы уже вычисляли ток насыщения диода p—n-перехода, эти расчеты относились к «толстому» диоду, у которого толщины p- и n-областей не малы

по сравнению с длинами диффузии. Между тем в вентильных фотоэлементах толщину освещаемой области делают по возможности малой, чтобы увеличить коэффициент β, и поэтому полученное ранее выражение не применимо к реальным фотоэлементам. И, наконец, мы не касались важного вопроса о коэффициенте полезного действия. Поэтому мы остановимся на процессах внутри фотоэлемента более подробно.

§ 5. Вентильные фотоэлементы

1. Рассмотрим фотоэлемент в виде плоскопараллельной пластинки с p—n-переходом, n-область которой освещается сильно поглощаемым светом (рис. 11.9). Толщину освещаемой области обозначим через d. Пластинка имеет два металлических электрода M, M, один из которых сплошной и расположен на тыльной стороне элемента (p-области), а другой, чтобы не препятствовать прохождению света, выполнен в виде узкого кольца или узкой полоски на крае передней стороны (n-области). Толщину самого p—n-перехода будем, как и выше, считать малой (по сравнению с длиной диффузии дырок) и, соответственно, рекомбинацией внутри перехода будем пренебрегать (расчет для «толстых» p—n-переходов см. в [3]).

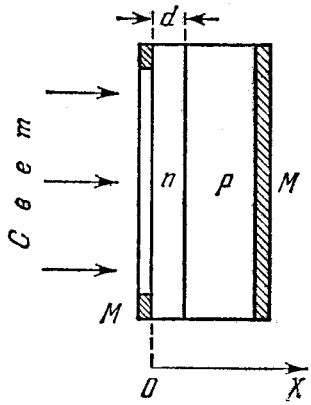


Рис. 11.9. К диффузионной теории вентильного фотоэлемента.

Плотность тока фотодырок в n-области определяется главным образом диффузией и равна

$$j_p = -eD_p \frac{dp}{dx}. \qquad (5.1)$$

Распределение дырок в n-области можно найти из уравнения непрерывности, которое для нашей задачи имеет вид

$$\frac{d^2p}{dx^2} - \frac{\delta p}{L_p^2} = 0. \qquad (5.2)$$

Будем отсчитывать координату x от освещаемой поверхности. Тогда граничные условия будут

$$x = 0: \qquad g_s = \frac{1}{e} j_p(0) + s\,\delta p(0); \qquad (5.3)$$

$$x = d: \quad \delta p(d) = p_n(e^{\alpha u} - 1). \qquad (5.4)$$

Решение уравнения (5.2) при указанных граничных условиях в общем случае имеет довольно громоздкий вид. Однако в реальных

фотоэлементах всегда стремятся сделать $d/L_p \ll 1$. Поэтому мы ограничимся приближенным решением для этого случая и, соответственно, пренебрежем в уравнении (5.2) вторым слагаемым, учитывающим рекомбинацию в объеме. Тогда можно считать

$$\frac{dp}{dx} = \text{const} = \frac{\delta p\,(d) - \delta p\,(0)}{d}, \qquad j_p = \text{const}. \tag{5.5}$$

Определяя здесь $\delta p\,(0)$ из граничного условия (5.3) и подставляя для $\delta p\,(d)$ его значение из граничного условия (5.4), получаем

$$j_p = e\,\frac{g_s - sp_n\,(e^{\alpha u} - 1)}{1 + s\,(d/D_p)}. \tag{5.6}$$

Помимо найденной дырочной составляющей плотности тока, через p—n-переход будет течь еще и электронный ток, обусловленный инжекцией электронов из n-области в p-область. Так как толщина неосвещаемой области обычно бывает не очень малой и охватывает, по крайней мере, несколько диффузионных длин, то эта составляющая тока выражается формулой, уже полученной в § VIII.1:

$$j_n = -\frac{eD_n n_p}{L_n}(e^{\alpha u} - 1). \tag{5.7}$$

Полная сила тока через фотоэлемент равна

$$i = S\,(j_p + j_n) = \frac{Seg_s}{1 + s\,(d/D_p)} - Se\left(\frac{sp_n}{1 + s\,(d/D_p)} + \frac{D_n n_p}{L_n}\right)(e^{\alpha u} - 1). \tag{5.8}$$

Мы видим, что, в соответствии с простыми рассуждениями § 4, полный ток имеет «световую» составляющую i_l (первое слагаемое) и составляющую от инжекции в p—n-переходе (второе слагаемое). При этом коэффициент β оказывается равным

$$\beta = \frac{1}{1 + s\,(d/D_p)}. \tag{5.9}$$

Он тем меньше, чем больше скорость поверхностной рекомбинации s. Ток насыщения диода для данной геометрии выражается формулой

$$i_s = Se\left(\frac{sp_n}{1 + s\,(d/D_p)} + \frac{D_n n_p}{L_n}\right). \tag{5.10}$$

Этот ток определяется не только тепловой генерацией неосновных носителей в объеме (второе слагаемое), но еще и тепловой их генерацией на передней поверхности фотоэлемента (первое слагаемое).

2. Рассмотрим теперь коэффициент полезного действия фотоэлемента. Согласно формуле (4.2) напряжение на электродах фотоэлемента, замкнутого на произвольную нагрузку, есть

$$u = \frac{kT}{e}\ln\left(1 + \frac{i_l}{i_s} - \frac{i}{i_s}\right). \tag{5.11}$$

Поэтому мощность, выделяемая во внешней цепи, равна

$$P = iu = \frac{kT}{e} i \ln\left(1 + \frac{i_l}{i_s} - \frac{i}{i_s}\right). \qquad (5.12)$$

Введем для краткости безразмерные токи

$$y \equiv i/i_s, \qquad z \equiv i_l/i_s \qquad (5.13)$$

и обозначим через G_s полный темп генерации фотоносителей:

$$G_s = g_s S, \qquad i_l = eG_s\beta. \qquad (5.14)$$

Тогда выражение для P можно записать в более удобном виде:

$$P = kT\beta G_s \frac{y \ln(1 + z - y)}{z}. \qquad (5.15)$$

Таким образом, полезная мощность зависит от величины отбираемого тока, т. е. от внешней нагрузки. При разомкнутой цепи ($y = 0$) и при коротком замыкании ($y = z$) эта мощность равна нулю и достигает максимума при некотором значении y_m отбираемого тока.

Для преобразования энергии важно иметь во внешней цепи максимальную мощность. Дифференцируя выражение (5.15) по y и приравнивая производную нулю, мы находим, что это имеет место при токе y_m, который удовлетворяет уравнению

$$(1 + z - y_m)\ln(1 + z - y_m) = y_m \qquad (0 \leqslant y_m \leqslant z). \qquad (5.16)$$

Если освещенность очень мала, так что $z \ll 1$, то можно положить $\ln(1 + z - y_m) \simeq z - y_m$. Тогда, удерживая в уравнении (5.16) только малые 1-го порядка, находим $y_m = {}^1/_2 z$, т. е. $i_m = {}^1/_2 i_l$. Следовательно, в этом случае оптимальная сила тока равна половине тока короткого замыкания. В общем случае уравнение (5.16) требует численных расчетов. Зависимость y_m от z показана на рис. 11.10. При увеличении интенсивности света (увеличении z) y_m монотонно увеличивается и приближается к z. Максимальная мощность во внешней цепи равна

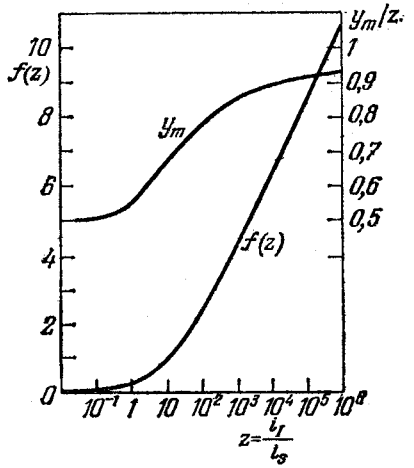


Рис. 11.10. Зависимость оптимального тока фотоэлемента $y_m = i_m/i_s$ от светового тока $z = i_l/i_s$ и функции $f(z)$.

$$P_m = kT\beta G_s f(z), \qquad (5.17)$$

где

$$f(z)=\frac{y_m^2}{z(1+z-y_m)}, \tag{5.18}$$

а y_m есть корень уравнения (5.16).

Положим теперь, что фотоэлемент освещается монохроматическим излучением с энергией фотонов $\hbar\omega > E_g$. Тогда мощность падающего излучения P_0 связана с G_s очевидным соотношением

$$P_0=\frac{G_s\hbar\omega}{(1-R)\nu}, \tag{5.19}$$

где R — коэффициент отражения, а ν — квантовый выход. Поэтому максимальный кпд равен

$$\eta_m=\frac{P_m}{P_0}=\frac{kT}{\hbar\omega}\nu(1-R)\beta f(z). \tag{5.20}$$

Полученное выражение показывает, что η_m тем больше, чем меньше $\hbar\omega$. Так как, с другой стороны, $\hbar\omega$ должно быть $> E_g$ (чтобы была генерация электронно-дырочных пар), то выгодно иметь $\hbar\omega \simeq E_g$. Этот результат очевиден физически, так как избыток энергии фотона $(\hbar\omega - E_g)$ теряется бесполезно (в конечном счете превращается либо в тепло, либо в излучение), что и снижает кпд.

График функции $f(z)$ приведен на рис. 11.10. При увеличении z величина $f(z)$ возрастает, а значит, увеличивается и кпд η_m. Физический смысл этого результата заключается в следующем. Каждая фотодырка, проходя через потенциальный уступ p—n-перехода (рис. 11.6), приобретает определенную дополнительную энергию. Однако эта энергия не выделяется во внешней цепи, а растрачивается внутри фотоэлемента. При увеличении освещенности величина потенциального уступа уменьшается, а значит, уменьшается и бесполезная потеря энергии. Одновременно с этим увеличивается и энергия, выделяемая во внешней цепи, которая пропорциональна напряжению u на зажимах. Оба эти обстоятельства и приводят к увеличению кпд при возрастании освещенности.

Конечно, увеличение η_m происходит только до известного предела. Это связано, во-первых, с тем, что напряжение u не может превышать потенциального барьера в темноте и поэтому при очень большой освещенности u уже не выражается формулой (5.11). И, во-вторых, в реальных фотоэлементах всегда имеется конечное сопротивление p- и n-областей (и контактов), на котором падает часть напряжения. Это падение напряжения увеличивается с ростом тока, что также ограничивает увеличение η_m.

3. Наиболее важное техническое применение вентильные фотоэлементы находят в солнечных батареях для прямого преобразования энергии солнечной радиации в электрическую. Так как солнечное излучение не является монохроматическим, то соотношение (5.19) между P_0 и G_s уже несправедливо и поэтому выражение для кпд

несколько изменяется. Так как активными являются только те фотоны, энергия которых $\hbar\omega \geqslant E_g$, то генерация G_s пропорциональна:

$$G_s \sim \int_{E_g}^{\infty} \rho_0(\hbar\omega)\, d(\hbar\omega),$$

где $\rho_0(\hbar\omega)$ — концентрация фотонов с энергией $\hbar\omega$ (на единичный интервал энергии) в световом потоке. Здесь мы положили квантовый выход $\nu = 1$. Падающая же на фотоэлемент мощность P_0 пропорциональна:

$$P_0 \sim \int_{0}^{\infty} \rho_0(\hbar\omega)\, d(\hbar\omega)\, \hbar\omega,$$

причем коэффициенты пропорциональности, которые мы не выписываем, в обоих выражениях одинаковы. Поэтому вместо формулы (5.19) мы будем иметь

$$\frac{G_s}{P_0} = (1-R)\frac{\int_{E_g}^{\infty} \rho_0(\hbar\omega)\, d(\hbar\omega)}{\int_{0}^{\infty} \rho_0(\hbar\omega)\, d(\hbar\omega)\, \hbar\omega}. \tag{5.21}$$

Так как солнечное излучение по спектральному составу близко к излучению абсолютно черного тела, то для $\rho_0(\hbar\omega)$ можно воспользоваться формулой Планка (IX.3.1), полагая в ней $T = T_1 = 6\cdot 10^3$ K — температуре поверхности Солнца. Тогда, вводя обозначения

$$x \equiv \frac{\hbar\omega}{kT_1}, \qquad x_1 \equiv \frac{E_g}{kT_1}, \tag{5.22}$$

мы имеем

$$\frac{G_s}{P_0} = \frac{1-R}{E_g}\,\psi(x_1), \tag{5.23}$$

где

$$\psi(x_1) = x_1 \frac{\int_{x_1}^{\infty} \frac{x^2\, dx}{e^x - 1}}{\int_{0}^{\infty} \frac{x^3\, dx}{e^x - 1}}. \tag{5.24}$$

Поэтому, учитывая формулы (5.17) и (5.23), мы получаем окончательно для максимального кпд выражение

$$\eta_m = \frac{kT}{E_g}(1-R)\,\beta f(z)\,\psi(x_1). \tag{5.25}$$

Оно отличается от соответствующего выражения (5.20) для монохроматического излучения тем, что вместо $\hbar\omega$ у нас вошло E_g и появился дополнительный множитель $\psi(x_1)$, учитывающий, что только часть фотонов излучения генерирует электронно-дырочные пары.

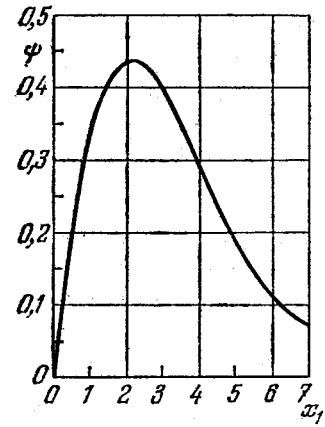


Рис. 11.11. График функции $\psi(x_1)$.

График функции $\psi(x_1)$ показан на рис. 11.11. Эта функция имеет максимум, равный около 44%, при значении $x_{1m} = 2{,}2$. Это значит, что при данной температуре черного излучения имеется оптимальная ширина запрещенной зоны полупроводника E_{gm}. Так как при $T_1 = 6000$ К, $kT_1 = 0{,}50$ эВ, то $E_{gm} = kT_1 \cdot x_{1m} = 2{,}2 \cdot 0{,}50 = 1{,}1$ эВ. Такую ширину запрещенной зоны имеет кремний. Так как, кроме того, технология получения монокристаллического кремния хорошо разработана, то этот материал широко используется для изготовления солнечных батарей.

Таким образом, даже при устранении всех потерь внутри фотоэлементов кпд солнечных батарей не может превышать $\psi_m \simeq 40\%$. Получаемые на практике максимальные кпд значительно меньше и, например, для кремниевых фотоэлементов в настоящее время не превышают 15—16%. Это объясняется тем, что значительная часть энергии растрачивается на поверхностную рекомбинацию, на выделение джоулева тепла внутри элемента, а также теряется при отражении.

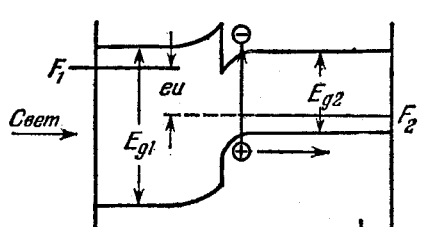


Рис. 11.12. Энергетическая диаграмма фотоэлемента с p — n-гетеропереходом.

Потери на поверхностную рекомбинацию можно уменьшить, используя гетеропереходы (§ VIII.5). На рис. 11.12 показана энергетическая диаграмма фотоэлемента с гетеропереходом, в котором освещаемая сторона состоит из полупроводника n-типа с шириной запрещенной зоны E_{g1}, а тыловая сторона — из полупроводника p-типа с шириной зоны $E_{g2} < E_{g1}$. В этом случае все фотоны с энергией $\hbar\omega < E_{g1}$ поглощаются в n-области очень слабо и поэтому p—n-переход может быть расположен дальше от поверхности фотоэлемента, чем в случае гомоперехода. Возникающий же барьер для дырок препятствует их движению к поверхности. При этом оказывается возможным увеличить толщину освещаемого слоя широко-

зонного полупроводника, а следовательно, и уменьшить внутреннее сопротивление фотоэлемента, что приводит к уменьшению потерь энергии на джоулево тепло.

Так как фотоны с энергией $\hbar\omega > E_{g1}$ сильно поглощаются в n-области и практически не достигают p—n-перехода, а фотоны с $\hbar\omega < E_{g2}$ не генерируют электронно-дырочные пары в области перехода, то такие фотоэлементы чувствительны лишь к фотонам с энергией в интервале $E_{g1} - E_{g2}$. Используя гетеропереходы, возможно получить кпд, превышающие 20%.

§ 6. Поверхностная фотоэдс

Особым случаем вентильной фотоэдс является так называемая поверхностная фотоэдс. Она возникает вследствие существования вблизи поверхности потенциального барьера, вызванного поверхностными уровнями (§ X.1). При генерации светом электронно-дырочных пар вблизи поверхности, электроны и дырки разделяются в поле потенциального барьера, так же как и в поле p—n-перехода, отчего заряд поверхности изменяется. Поэтому при освещении появляется дополнительная разность потенциалов между поверхностью и объемом.

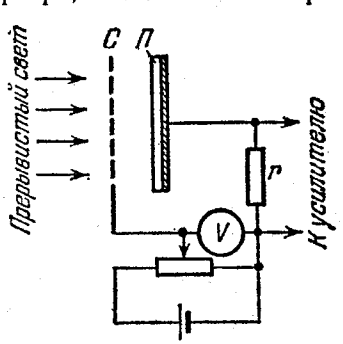


Рис. 11.13. Конденсаторный метод наблюдения фотоэдс.

Знак этой разности потенциалов зависит от знака поверхностного потенциала до освещения. Если $Y_s < 0$ (у поверхности зоны искривлены вверх), то фотодырки будут двигаться к поверхности, а фотоэлектроны — от поверхности и потенциал поверхности будет увеличиваться. Если $Y_s > 0$, то при освещении потенциал поверхности будет уменьшаться. В обоих случаях изгиб энергетических зон у поверхности при освещении уменьшается.

Поверхностную фотоэдс можно наблюдать с помощью «конденсаторного» метода, схема которого показана на рис. 11.13. Здесь исследуемый полупроводник $П$ и прозрачный металлический электрод $С$ (например, металлическая сетка) образуют обкладки конденсатора, который замкнут на сопротивление r. Если бы металлический электрод был вибрирующим, то мы имели бы известную схему Кельвина для измерения контактной разности потенциалов. Однако, что гораздо проще, металлический электрод можно сделать неподвижным, а полупроводник освещать прерывистым светом. Тогда возникающая при освещении поверхностная фотоэдс будет складываться (или вычитаться, в зависимости от ее знака) с темновой контактной разностью, и на сопротивлении r появится переменное

напряжение, по величине которого можно определить величину фотоэдс. Таким образом, поверхностная фотоэдс проявляется как изменение контактной разности потенциалов при освещении.

Необходимо отметить, что изменение контактной разности при освещении может быть вызвано не только поверхностной фотоэдс, но еще и эффектом Дембера (§ 2), и поэтому на опыте наблюдается сумма обоих эффектов. При этом знак эдс Дембера определяется только знаком разности коэффициентов диффузии ($D_n - D_p$) и не зависит от знака Y_s. Опыт, однако, дает, что в германии и кремнии изменение контактной разности при освещении зависит от обработки поверхности и может даже менять знак. Это показывает, что, по крайней мере в этих полупроводниках, конденсаторная фотоэдс обусловлена главным образом поверхностной фотоэдс.

§ 7. Фотоэлектромагнитный эффект

1. В 1934 г. И. К. Кикоин и М. М. Носков открыли новый тип фотоэдс, которая возникает при освещении полупроводника, помещенного в магнитное поле (фотоэлектромагнитный эффект). Наиболее простой случай мы имеем, если одна из граней плоскопараллельной пластинки освещается сильно поглощаемым светом, а магнитная индукция $\mathfrak{B}$ перпендикулярна направлению света (рис. 11.14). Тогда между торцами образца, параллельными направлениям света и $\mathfrak{B}$, появляется разность потенциалов, а при замыкании торцов в цепи возникает ток.

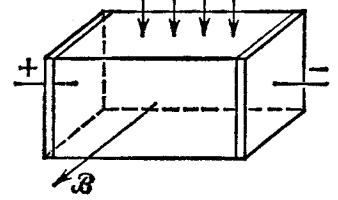


Рис. 11.14. Фотоэлектромагнитный эффект.

Фотоэлектромагнитная эдс может быть намного больше, чем рассмотренные выше фотоэдс. Так, например, в высокоомных кристаллах закиси меди, в которых она была впервые обнаружена, при $\mathfrak{B} \sim 10^3$ Гс ее величина могла достигать многих десятков вольт.

Физическая причина фотоэлектромагнитного эффекта (ФЭМ) заключается в отклонении потоков диффузии фотоэлектронов и фотодырок в магнитном поле (Я. И. Френкель). Полупроводник везде в дальнейшем будем считать изотропным. Направим ось X прямоугольной системы координат вдоль освещаемой поверхности, ось Y — перпендикулярно этой поверхности, а ось Z — по направлению $\mathfrak{B}$ (рис. 11.15). Если бы магнитного поля не было, то потоки фотодырок и фотоэлектронов были бы направлены перпендикулярно к освещаемой поверхности, т. е. вдоль оси Y, оба в одну и ту же сторону, а создаваемые ими токи j_p и j_n были бы противоположны. При включении магнитного поля, вследствие действия силы Лоренца, поток дырок отклоняется на угол Холла φ_p, а поток электронов —

в противоположном направлении на угол Холла φ_n, где

$$\operatorname{tg}\varphi_p = \frac{1}{c}\mu_{pH}\mathscr{B}, \quad \operatorname{tg}\varphi_n = -\frac{1}{c}\mu_{nH}\mathscr{B} \tag{7.1}$$

(ср. § I.3). Поэтому возникают составляющие плотности тока j_{px} и j_{nx}, параллельные освещаемой поверхности и направленные в одну и ту же сторону. Если пластинка не замкнута, то левый ее торец будет заряжаться положительно, а правый — отрицательно, в пластинке появится составляющая электрического поля $\mathscr{E}_x$, а между торцами возникнет разность потенциалов.

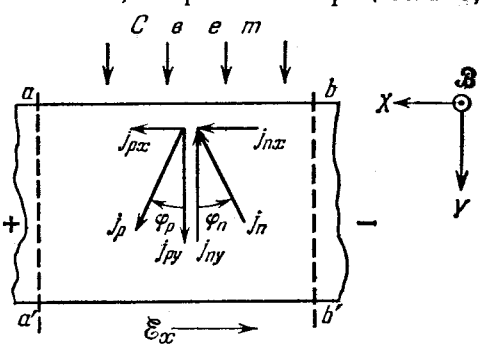


Рис. 11.15. Электронные и дырочные токи при ФЭМ эффекте. Магнитная индукция направлена от чертежа к читателю.

Таким образом, ФЭМ эффект имеет то же происхождение, что и эффект Холла. Однако между ними имеется и различие. В случае эффекта Холла направления токов, создаваемых электронами и дырками, одинаковы, а направления потоков — противоположны. Вследствие этого магнитное поле отклоняет оба эти потока в одну и ту же сторону, и поэтому эффект Холла есть разностный эффект. При этом, как мы видели (§ I.4), результирующие угол Холла и постоянная Холла при смешанной проводимости могут даже обращаться в нуль. В случае же ФЭМ эффекта направления потоков дырок и электронов одинаковы, магнитное поле отклоняет их в разные стороны, и поэтому мы наблюдаем суммарный эффект обоих отклонений.

Рассмотрим теперь распределение электрического поля и плотности тока внутри пластинки. Торцы пластинки будем считать разомкнутыми. Так как магнитное поле постоянно во времени, то согласно уравнениям Максвелла $\operatorname{rot}\mathscr{E} = 0$. Это дает, что

$$\operatorname{rot}_z \mathscr{E} = \frac{\partial \mathscr{E}_y}{\partial x} - \frac{\partial \mathscr{E}_x}{\partial y} = 0.$$

Если длина пластинки (в направлении X) достаточно велика, то все величины вследствие симметрии системы не зависят от координаты x, и поэтому $\partial\mathscr{E}_y/\partial x = 0$. Отсюда следует, что

$$\frac{\partial \mathscr{E}_x}{\partial y} = 0, \quad \mathscr{E}_x = \text{const},$$

т. е. что продольная составляющая электрического поля не зависит от y. Поэтому напряжение ФЭМ эффекта между любой парой

точек a и b (рис. 11.15) на освещенной поверхности такое же, как и между равноудаленными точками a' и b' на темновой поверхности. Этим обстоятельством пользуются при измерениях напряжения ФЭМ эффекта и располагают измерительные зонды на задней стороне пластинки, чтобы исключить появление вентильной фотоэдс в контактах зонд—полупроводник.

Найдем теперь форму эквипотенциальных поверхностей. Электрическое поле $\mathcal{E}$ в каждой точке складывается из постоянной составляющей $\mathcal{E}_x$ и поля $\mathcal{E}_y$, которое есть поле амбиполярной диффузии (поле эффекта Дембера, § 2). Поэтому суммарное поле $\mathcal{E}$, вообще говоря, направлено под углом к освещаемой поверхности (рис. 11.16, $а$). Однако, вследствие рекомбинации электронов и дырок, поле $\mathcal{E}_y$ затухает при увеличении y и поэтому на расстоя-

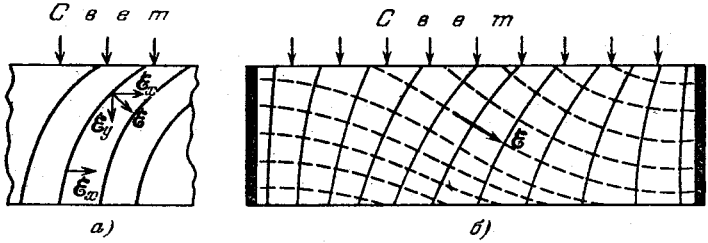


Рис. 11.16. Эквипотенциальные поверхности при ФЭМ эффекте в разомкнутом образце без электродов ($а$) и с электродами ($б$). Пунктир — линии электрического поля.

нии нескольких длин диффузии $\mathcal{E}$ становится параллельным освещенной поверхности. Восстанавливая отсюда форму эквипотенциальных поверхностей, которые ортогональны к $\mathcal{E}$, мы получаем картину рис. 11.16, $а$.

Такая форма эквипотенциальных поверхностей будет, конечно, только в том случае, когда торцы пластинки не имеют металлических электродов. При наличии электродов эквипотенциальные поверхности вблизи торцов искажаются, так как поверхность каждого из электродов должна совпадать с одной из эквипотенциальных поверхностей. Поэтому мы получаем более сложную картину, показанную на рис. 11.16, $б$, где изображены также силовые линии электрического поля. Это искажающее влияние электродов необходимо учитывать при измерении напряжения ФЭМ эффекта, а для этого нужно располагать измерительные зонды на достаточном расстоянии от торцов пластинки.

Обратимся теперь к распределению токов. Для бесконечно длинной пластинки полный ток в направлении Y в любой точке равен нулю:

$$j_y = j_{py} + j_{ny} = 0. \qquad (7.2)$$

Если торцы пластинки замкнуты накоротко, то $\mathscr{E}_x = 0$ и, соответственно, тока дрейфа нет. Поэтому плотность тока определяется только диффузией в магнитном поле и равна

$$j_x^{\text{кз}}(y) = j_{px}(y) + j_{nx}(y). \tag{7.3}$$

При разомкнутых торцах добавляется еще ток дрейфа и плотность тока равна

$$j_x(y) = \sigma \mathscr{E}_x + j_x^{\text{кз}}(y), \tag{7.4}$$

где σ — удельная электропроводность. При этом направления токов дрейфа и диффузии противоположны (ср. рис. 11.15). Так как $\mathscr{E}_x$ не зависит от y, а $j_x^{\text{кз}}$ затухает при увеличении y, то вблизи освещенной поверхности преобладает ток диффузии, а в глубине пластинки остается только один ток дрейфа и направление тока становится противоположным. Вследствие этого в пластинке возникают замкнутые линии тока. Для случая разомкнутой пластинки конечной длины они показаны на рис. 11.17. При этом пластинка приобретает определенный магнитный момент, который можно обнаружить экспериментально.

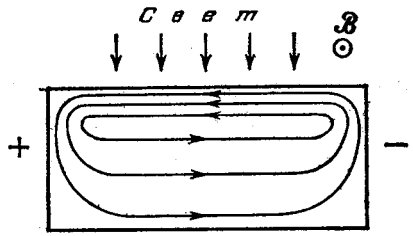


Рис. 11.17. Линии тока при ФЭМ эффекте в разомкнутом образце без электродов.

Найдем теперь величину ФЭМ эффекта. Для этого можно было бы рассмотреть разомкнутую пластинку и вычислить напряжение холостого хода. Однако проще оказывается вычислить ток короткого замыкания $i^{\text{кз}}$, протекающий через все сечение пластинки. Обе эти величины, как и для всякого источника тока, связаны соотношением

$$V_0 = i^{\text{кз}} R, \tag{7.5}$$

где R — полное сопротивление пластинки, и поэтому, зная $i^{\text{кз}}$, можно определить и V_0.

Из сказанного выше следует (ср. рис. 11.15), что

$$j^{\text{кз}} = j_p \sin \varphi_p + j_n \sin \varphi_n,$$

а из формулы (7.2) имеем

$$j_y = j_p \cos \varphi_p - j_n \cos \varphi_n = 0.$$

Поэтому, выражая j_n из второго соотношения и подставляя его в первое, находим

$$j^{\text{кз}} = j_p \cos \varphi_p \cdot \Theta = j_{py} \Theta, \tag{7.6}$$

где для сокращения обозначено

$$\Theta = \Theta_p + \Theta_n = \operatorname{tg}\varphi_p + \operatorname{tg}\varphi_n. \tag{7.7}$$

Сила тока короткого замыкания (рассчитанная на единицу ширины пластинки) есть

$$i^{\text{кз}} = \int_0^d j^{\text{кз}}(y)\,dy = \Theta \int_0^d j_{py}(y)\,dy, \tag{7.8}$$

где d — толщина пластинки. Таким образом, задача сводится к вычислению нормальной составляющей тока диффузии $j_{py}(y)$.

Расчет особенно прост при выполнении следующих условий: 1) нет заметного прилипания на ловушки, так что $\delta p = \delta n$; 2) интенсивность света мала; тогда коэффициент диффузии D и время жизни τ не зависят от y; 3) магнитное поле слабое, $\Theta \ll 1$. В этом случае можно считать, что длина диффузии L заметно не изменяется в магнитном поле. Кроме того, можно пренебречь эффектом магнетосопротивления и (при слабом освещении) считать, что $\sigma(\mathscr{B}) \simeq \sigma_0$, где σ_0 — удельная проводимость без освещения и без магнитного поля. И, наконец, 4) пластинка не очень тонкая ($d/L \gtrsim 2 \div 3$), так что $\delta p(d) \ll \delta p(0)$. Тогда мы имеем

$$j_{py} = -eD\frac{dp}{dy}, \quad i^{\text{кз}} = -eD\Theta \int_0^d \frac{dp}{dy}\,dy = eD\Theta\,\delta p(0).$$

При сделанных предположениях $\delta p(0)$ выражается формулой (X.6.9):

$$\delta p(0) = \frac{L}{D}\frac{g_s}{1+S}.$$

Поэтому для $i^{\text{кз}}$ получается окончательно

$$i^{\text{кз}} = eL\frac{g_s}{1+S}\Theta = eL\frac{g_s}{1+S}\frac{1}{c}(\mu_{pH} + \mu_{nH})\mathscr{B}. \tag{7.9}$$

В слабых магнитных полях ток короткого замыкания пропорционален магнитной индукции $\mathscr{B}$.

Оценим теперь величину эффекта. Рассматривая, как и в прежних примерах, чистый германий при 300 К, мы имеем $\mu_{nH} \simeq \simeq 4\cdot 10^3$ см^2/В·с $= 12\cdot 10^5$ ед. CGSE, $\mu_{pH} \simeq 6\cdot 10^5$ ед. CGSE, что при $\mathscr{B} = 1\cdot 10^3$ Гс дает $\Theta \simeq 6\cdot 10^{-2}$. Следовательно, при выбранном $\mathscr{B}$, $\Theta \ll 1$ и формула (7.9) применима. Положим, далее, $g_s \sim$ $\sim 10^{17}$ см$^{-2}\cdot$с$^{-1}$, $L \sim 0{,}1$ см (что типично для чистого германия при комнатной температуре) и $S \ll 1$. Тогда по формуле (7.9) находим $i^{\text{кз}} \sim 10^{-4}$ А/см. Если длина пластинки ~ 1 см, а толщина $\sim 0{,}1$ см, то ее сопротивление (тоже на единицу ширины) будет порядка $10^2 \div 10^3$ Ом·см. Поэтому напряжение между концами разомкнутой пластинки будет $V_0 = i^{\text{кз}}\cdot R \sim 10$—$100$ мВ.

ФЭМ эффект используют в чувствительных приемниках инфракрасного излучения. Он широко применяется также для измерения времени жизни неравновесных электронов и дырок. Если известен темп поверхностной генерации g_s (т. е. интенсивность активного света) и скорость поверхностной рекомбинации s, то, измеряя $i^{кз}$, можно по формуле (7.9) найти $L\ (=\sqrt{D\tau})$, а следовательно, и τ.

Практически, однако, гораздо удобнее одновременно исследовать ФЭМ эффект и фотопроводимость, что позволяет исключить влияние поверхностной рекомбинации и не требует измерения интенсивности света. Для этого используют метод компенсации ФЭМ эффекта и фотопроводимости, в котором к торцам пластинки прикладывают еще небольшое постоянное напряжение u. Тогда в пластинке, дополнительно к току $i^{кз}_{фэм}$, появляется еще ток фотопроводимости, выражаемый формулой (X.6.10):

$$i_{фп} = e(\mu_p + \mu_n)\frac{g_s\tau}{1+S}\frac{u}{l}. \tag{7.10}$$

Здесь мы положили ширину пластинки равной единице и заменили $\mathscr{E}$ на u/l, где l — длина пластинки. Затем величину и знак напряжения u подбирают так, чтобы результирующий ток обратился в нуль, что соответствует условию $i^{кз}_{фэм} = i_{фп}$. Приравнивая выражения (7.9) и (7.10), мы получаем

$$L = \frac{1}{c} D \frac{\mu_{pH} + \mu_{nH}}{\mu_p + \mu_n} \mathscr{B} \frac{l}{u}. \tag{7.11}$$

При этом темп генерации пар g_s и скорость поверхностной рекомбинации S исключаются.

ФЭМ эффект с успехом используют также для измерения скорости поверхностной рекомбинации. Если известна объемная длина диффузии L, то по формуле (7.9) можно найти S. Конечно, для этого влияние поверхностной рекомбинации должно быть достаточно велико, т. е. нужно иметь $S \gtrsim 1$. Для измерения S, однако, удобнее пользоваться тонкими пластинками ($d/L < 1$). При этом вместо формулы (7.9) получается другое соотношение, на котором мы останавливаться не будем.

2. В приведенном выше расчете мы предполагали, что магнитное поле слабое, и, соответственно, считали, что длина диффузии L и коэффициент диффузии D не изменяются при включении магнитного поля. Если магнитное поле нельзя считать слабым, то для вычисления $i^{кз}$ нужно рассмотреть задачу о диффузии в магнитном поле. В Приложении VII показано, что если $\Theta \gtrsim 1$ и, кроме того, $\tau_p \neq \tau_n$ (но по-прежнему освещение слабое), то формула (7.9) остается справедливой и в этом случае, если только L заменить на длину диффузии в магнитном поле L^*, а D — на коэффициент диффузии в магнитном поле D^*. При этом

$$L^* = \sqrt{D^*\tau_{фэм}}, \tag{7.12}$$

а $\tau_{\text{фэм}}$ есть некоторое комбинированное время, определяемое формулой

$$\tau_{\text{фэм}} = \frac{\tau_p n_0 + \tau_n p_0}{n_0 + p_0}. \tag{7.13}$$

Здесь n_0 и p_0 — равновесные концентрации электронов и, соответственно, дырок. Это время называют «временем жизни фотоэлектромагнитного эффекта». Из формулы (7.13) видно, что для примесного полупроводника ($n_0 \gg p_0$ или $p_0 \gg n_0$) оно практически совпадает с временем жизни неосновных носителей. Величина же D^* равна

$$D^* = D \frac{bn_0 + p_0}{bn_0(1+\Theta_p^2) + p_0(1+\Theta_n^2)}. \tag{7.14}$$

Здесь

$$D = \frac{n_0 + p_0}{\frac{n_0}{D_p} + \frac{p_0}{D_n}}$$

есть коэффициент амбиполярной диффузии без магнитного поля, а $b = \mu_n/\mu_p$. При выводе этих формул считалось: $\mu_{pH} = \mu_p$, $\mu_{nH} = \mu_n$ (что не вносит большой ошибки). Кроме того, считалось, что полупроводник не вырожден, и использовалось соотношение Эйнштейна $\mu_n/\mu_p = D_n/D_p$. В слабых полях ($\Theta_n^2, \Theta_p^2 \ll 1$) D^* переходит в обычный коэффициент амбиполярной диффузии D.

Таким образом, в сильных магнитных полях пропорциональность между $i^{\text{кз}}$ и $\mathscr{B}$ нарушается. При этом зависимость $i^{\text{кз}}(\mathscr{B})$ различна при малой и большой скоростях поверхностной рекомбинации. Из формулы (7.9) видно, что при $S \ll 1$, $i^{\text{кз}} \sim L\mathscr{B}$. С другой стороны, из формул (7.12) и (7.14) следует, что в сильном поле ($\Theta \gg 1$) $D^* \sim 1/\mathscr{B}^2$, а следовательно, $L^* \sim 1/\mathscr{B}$. Поэтому $i^{\text{кз}}$ перестает зависеть от $\mathscr{B}$ и достигает насыщения. В другом предельном случае $S \gg 1$, $i^{\text{кз}} \sim D^*\mathscr{B}$ и в сильном поле уменьшается, как $1/\mathscr{B}$. Вследствие этого при постепенном увеличении магнитного поля $i^{\text{кз}}$ сначала растет линейно, затем проходит через максимум и снова уменьшается. Оба эти типа зависимостей наблюдаются на опыте.

Если пластинка не является толстой (т. е. $d/L^* \lesssim 1$), то использованное нами условие $\delta p\,(d) \ll \delta p\,(0)$ уже несправедливо и полученные соотношения теряют силу. Расчет ФЭМ эффекта для случая пластинок произвольной толщины (для малой и большой интенсивности света) в слабом магнитном поле подробно изложен в работе [5].

3. Выше мы предполагали, что свет поглощается очень сильно (поверхностная генерация) и что магнитная индукция $\mathfrak{B}$ перпендикулярна направлению распространения света. При освещении проникающим излучением диффузионные потоки электронов и дырок изменяются, и поэтому изменяется и величина ФЭМ эффекта.

При этом, как правило, эдс уменьшается и может даже изменить знак. Последнее происходит в том случае, когда скорость поверхностной рекомбинации на передней поверхности больше, чем на задней. Тогда концентрация фотоносителей у задней поверхности становится больше, чем у передней, и потоки диффузии направлены противоположно направлению распространения света.

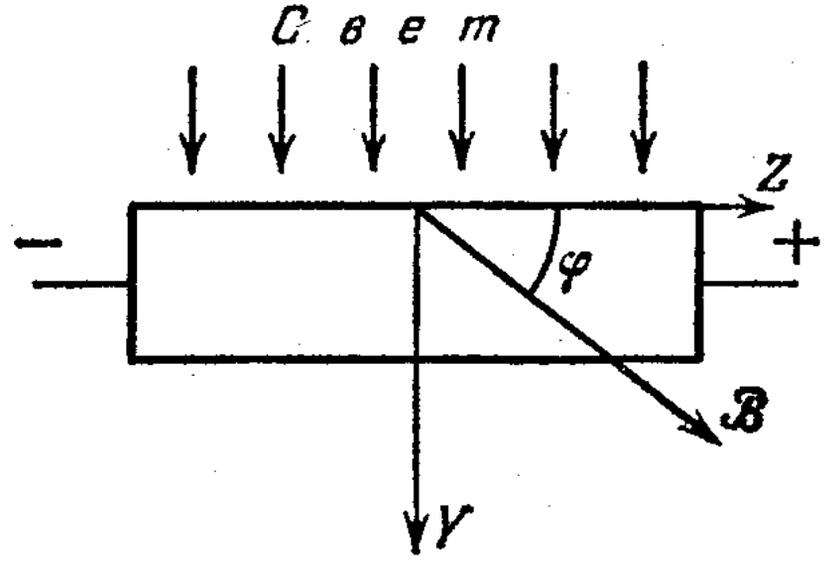


Рис. 11.18. Квадратичный ФЭМ эффект.

Если $\mathfrak{B}$ не перпендикулярно световому потоку, то, помимо рассмотренного линейного ФЭМ эффекта (в слабых магнитных полях пропорционального $\mathscr{B}$), возникает еще квадратичный эффект, пропорциональный $\mathscr{B}^2$. Качественное объяснение квадратичного ФЭМ эффекта заключается в следующем. Положим, что освещаемая поверхность перпендикулярна оси Y, а $\mathfrak{B}$ лежит в плоскости YZ и образует с Z угол φ (рис. 11.18). Тогда z — составляющая индукции $\mathscr{B}\cos\varphi$ вызывает обычный линейный эффект, а ток короткого замыкания $i_x^{\text{кз}}$ (в слабых магнитных полях) пропорционален $\mathscr{B}\cos\varphi$ и направлен вдоль оси X. Однако теперь на этот ток действует еще y — составляющая индукции $\mathscr{B}\sin\varphi$, которая поворачивает его в плоскости XZ. Поэтому вдоль оси Z появляется составляющая тока $i_Z^{\text{кз}}$, пропорциональная $\mathscr{B}\cos\varphi\,\mathscr{B}\sin\varphi \sim \mathscr{B}^2\sin 2\varphi$, а при разомкнутой пластинке — соответствующее напряжение холостого хода. Эта зависимость приближенно наблюдается и на опыте.

Глава XII

КОЛЕБАНИЯ КРИСТАЛЛИЧЕСКОЙ РЕШЕТКИ

§ 1. Малые колебания

Рассмотрим задачу о малых колебаниях атомов кристаллической решетки около положений равновесия. Будем рассматривать кристалл с произвольным (равным r) числом атомов в элементарной ячейке. Равновесное положение любого атома в решетке можно задать с помощью двух величин, одна из которых указывает номер атома в ячейке, а другая — положение данной ячейки в кристалле. Номер атома в ячейке обозначим через h ($1 \leqslant h \leqslant r$). Положение ячейки в кристалле можно задать радиус-вектором ρ какой-либо ее точки, например центра или одного из атомов. Примем одну из ячеек за начальную, направив оси координат вдоль главных осей кристалла. Тогда компоненты радиус-вектора любой ячейки будут целыми кратными соответствующих компонент вектора решетки $\mathbf{a}$:

$$\boldsymbol{\rho} = \{g_x a_x,\ g_y a_y,\ g_z a_z\}. \tag{1.1}$$

Здесь g_x, g_y, g_z — целые числа (или нули); их можно рассматривать как компоненты некоторого вектора $\mathbf{g}$. В качестве основного объема *) кристалла выберем куб, вдоль сторон которого укладывается $G_x = L/a_x$, $G_y = L/a_y$ и $G_z = L/a_z$ элементарных ячеек (таким образом, общее число их есть $G_x G_y G_z = G$). Тогда

$$0 \leqslant g_\alpha \leqslant G_\alpha, \qquad \alpha = x,\ y,\ z. \tag{1.2}$$

(Строго говоря, в правых частях неравенств должно стоять число $G_\alpha - 1$, но ввиду колоссальности числа G_α единицей можно пренебречь.) Для краткости будем обозначать совокупность значков h и g_x, g_y, g_z одним индексом a:

$$a = \{h,\ g_x,\ g_y,\ g_z\}. \tag{1.3}$$

Пусть $\mathbf{R}_a$ — тройка координат a-го атома, а M_a — его масса. Кинетическая энергия решетки будет

$$T = \frac{1}{2} \sum_a M_a \dot{\mathbf{R}}_a^2, \tag{1.4}$$

где $\dot{\mathbf{R}}_a = d\mathbf{R}_a/dt$.

*) Как и в гл. III, основной объем определяется как достаточно большая область кристалла, на границах которой физические условия одинаковы. Возможность ввести основной объем связана с тем, что мы не интересуемся поверхностными свойствами кристалла, считая его достаточно большим.

Потенциальная энергия решетки V представляет собой некоторую функцию всех переменных $\mathbf{R}_a$. Явный вид ее при произвольном расположении атомов может быть очень сложен. Для наших целей, однако, достаточно заметить, что атомы совершают лишь малые колебания около положений равновесия (координаты последних обозначим через $\mathbf{R}_a^{(0)}$). Поэтому функцию V можно представить в виде разложения Тэйлора по степеням отклонений атомов от точек равновесия. Разность $\mathbf{R}_a - \mathbf{R}_a^{(0)}$ мы будем называть вектором смещения a-го атома и обозначать через $\mathbf{Q}_a$:

$$\mathbf{Q}_a = \mathbf{R}_a - \mathbf{R}_a^0. \tag{1.5}$$

Обозначим для краткости всю совокупность переменных $\mathbf{R}_a$ ($\mathbf{R}_a^0$) символом R (R^0) и примем во внимание, что при $R = R^0$ потенциальная энергия V имеет минимум — по определению положения устойчивого равновесия. Получим

$$V(R) = V(R^0) + \frac{1}{2}\sum_{a,\,a'} \Gamma_{aa'}^{\alpha\alpha'} Q_{a,\alpha} Q_{a',\alpha'} + \ldots \tag{1.6}$$

Точками здесь обозначены члены высшего порядка по компонентам вектора смещения. Величина $V(R^0)$ — потенциальная энергия в точке равновесия — в интересующем нас круге задач представляет собой просто аддитивную постоянную. Ее можно вообще исключить из рассмотрения, выбрав соответствующим образом начало отсчета энергии.

Коэффициенты $\Gamma_{aa'}^{\alpha\alpha'}$ суть вторые производные от V по смещениям, взятые в точке минимума:

$$\Gamma_{aa'}^{\alpha\alpha'} = \left.\frac{\partial^2 V}{\partial R_{a,\alpha}\partial R_{a',\alpha'}}\right|_{R=R^0}. \tag{1.7}$$

Очевидно, они симметричны относительно перестановки штрихованных и нештрихованных индексов:

$$\Gamma_{aa'}^{\alpha\alpha'} = \Gamma_{a'a}^{\alpha'\alpha}. \tag{1.7'}$$

Действительно, такая перестановка сводится просто к перестановке порядка дифференцирования, что не влияет на результат. Из условия минимальности потенциальной энергии при $R = R^0$ следует, что величины (1.7) при $\alpha = \alpha'$, $a = a'$ положительны:

$$\Gamma_{aa}^{\alpha\alpha} > 0. \tag{1.7''}$$

Наконец, еще одно, важное для дальнейшего, свойство коэффициентов $\Gamma_{aa'}^{\alpha\alpha'}$ можно установить, замечая, что все элементарные ячейки в основном объеме кристалла физически эквивалентны. Отсюда следует, что энергия взаимодействия между атомами может зависеть лишь от положения их в ячейке и от относительного рас-

положения ячеек, в которых они находятся, но не от номеров этих ячеек по отдельности:

$$\Gamma_{aa'}^{\alpha\alpha'} = \Gamma_{hh'}^{\alpha\alpha'}(\mathbf{g}-\mathbf{g}'). \tag{1.8}$$

В принципе значения $\Gamma_{hh'}^{\alpha\alpha'}$ можно было бы вычислить, зная силы взаимодействия между атомами. Это, однако, требует весьма громоздких расчетов. В ряде задач (в частности, в задаче о рассеянии электронов) достаточно рассматривать $\Gamma_{hh'}^{\alpha\alpha'}$ просто как известные параметры, характеризующие данную решетку. Комбинации этих параметров, фактически входящие в окончательные результаты (например, скорость звука), можно определять непосредственно из опыта.

Полная энергия системы атомов, совершающих малые колебания около периодически расположенных положений равновесия, дается суммой выражений (1.4) и (1.6). Опуская аддитивную постоянную $V(R^0)$, мы имеем

$$H = \frac{1}{2}\sum_{a} M_a \dot{\mathbf{Q}}_a^2 + \frac{1}{2}\sum_{a,\,a'} \Gamma_{hh'}^{\alpha\alpha'}(\mathbf{g}-\mathbf{g}')\,\mathbf{Q}_{a,\,\alpha}\mathbf{Q}_{a',\,\alpha'}. \tag{1.9}$$

При переходе к квантовомеханическому рассмотрению векторы смещения $\mathbf{Q}_a$ и сопряженные им импульсы $M_a\dot{\mathbf{Q}}_a$ следует заменить соответствующими операторами. Мы сделаем это несколько позднее, преобразовав сначала выражение (1.9) к более удобному виду.

§ 2. Нормальные координаты

Выражение (1.9) представляет собой энергию системы связанных гармонических осцилляторов, координаты которых суть $\mathbf{Q}_a$. Эта связь обусловлена наличием в (1.9) «перекрестных» слагаемых $\Gamma_{hh'}^{\alpha\alpha'}\mathbf{Q}_{a,\,\alpha}\mathbf{Q}_{a',\,\alpha'}$ при $a' \neq a$, $\alpha' \neq \alpha$. В результате происходит непрерывный обмен энергией между осцилляторами. Энергия каждого из них в отдельности не является интегралом движения. Соответственно зависимость $\mathbf{Q}_a$ от времени отнюдь не имеет вида простого гармонического колебания, а оказывается довольно сложной.

Естественно попытаться ввести вместо $\mathbf{Q}_a$ новые переменные так, чтобы энергия системы, будучи выражена через них, не содержала перекрестных членов. Эти переменные называются *нормальными координатами*. По определению энергия системы, как функция нормальных координат и соответствующих им скоростей (или импульсов), представляет собой сумму энергий независимых гармонических осцилляторов. Соответственно нормальные координаты меняются со временем по простому гармоническому закону (каждая, вообще говоря, со своей частотой). Колебания, совершаемые нормальными координатами, также называются нормальными.

Определение нормальных координат и частот, с которыми они колеблются, представляет собой, очевидно, не что иное, как задачу о приведении квадратичной формы (1.9) к сумме квадратов. Как известно из линейной алгебры, для этой цели надо произвести линейное преобразование переменных, т. е. в нашем случае представить векторы смещения отдельных атомов $\mathbf{Q}_a$ как линейные комбинации новых обобщенных координат. Вид этой линейной комбинации в интересующей нас задаче можно в известной мере предугадать, принимая во внимание физическую эквивалентность всех элементарных ячеек в основном объеме кристалла. Действительно, на основании этой эквивалентности следует ожидать, что амплитуды смещений соответственных атомов в различных ячейках будут одинаковы и при переходе от одной ячейки к другой будет меняться только фаза колебаний. По этой причине попытаемся представить векторы смещения в виде

$$\mathbf{Q}_a(t) = \frac{1}{G^{1/2}} \sum_{\mathbf{q},\, s} \{\xi_h(\mathbf{q},\, s)\, \eta(\mathbf{q},\, s,\, t)\, e^{i\mathbf{q}\rho}\}. \qquad (2.1)$$

Здесь множитель $G^{-1/2}$ введен для удобства в дальнейшем; $\mathbf{q}$ — некоторый вектор, определяющий изменение фазы при переходе от одной ячейки к другой; ρ — вектор (1.1); комплексные величины $\eta(\mathbf{q}, s, t)$ играют роль новых обобщенных координат (зависящих от времени); $\xi_h(\mathbf{q}, s)$ — векторы, которые надлежит подобрать так, чтобы при подстановке (2.1) в (1.9) «перекрестных» членов не было. Индекс s нумерует различные возможные типы нормальных колебаний, характеризуемых одним и тем же вектором $\mathbf{q}$. Необходимость ввести этот индекс ясна хотя бы из следующих соображений. Поскольку обобщенные координаты η, по определению, должны гармонически зависеть от времени, выражение (2.1) представляет собой сумму волн. Направление распространения каждой такой волны определяется вектором $\mathbf{q}$, а направление колебаний h-го атома — вектором ξ_h. В зависимости от угла между этими векторами мы можем иметь различные типы волн с одной и той же длиной, например продольные и поперечные. Именно это обстоятельство и отражается индексом s *).

Вектор $\mathbf{q}$ называется квазиволновым. Название связано с очевидной аналогией между выражением (2.1) и набором волн в сплошной среде. В следующем параграфе мы увидим, однако, что (как и в случае электронов в решетке, § III.2) эта аналогия неполна, почему и не используется термин «волновой вектор».

Возможные значения компонент $\mathbf{q}$ определяются граничными условиями, накладываемыми на векторы смещения. Как и в гл. III, рассматривая систему в основном объеме V, удобно наложить

*) В решетках, содержащих более одного атома в элементарной ячейке, есть и другие причины, требующие введения индекса s (см. ниже, § 3).

условия периодичности на его границах. Тогда компоненты $\mathbf{q}$ даются формулами вида (III.3.10), в которых следует положить

$$L_\alpha = G_\alpha a_\alpha, \qquad \alpha = x,\ y,\ z.$$

Заметим, что величины $\eta(\mathbf{q},\ s,\ t)$ — комплексные нормальные координаты — не независимы. Действительно, в правой части (2.1) стоят комплексные числа, в то время как компоненты вектора смещения должны быть вещественными. Это возможно, если выполняется равенство

$$\xi_h(\mathbf{q},\ s)\,\eta(\mathbf{q},\ s,\ t) = \xi^*(-\mathbf{q},\ s)\,\eta^*(-\mathbf{q},\ s,\ t). \tag{2.2}$$

Действительно, выражение, комплексно сопряженное правой части (2.1), есть

$$G^{-1/2}\sum_{\mathbf{q},\ s}\xi_h^*(\mathbf{q},\ s)\,\eta^*(\mathbf{q},\ s,\ t)\,e^{-i\mathbf{q}\rho}.$$

Заменяя здесь переменные суммирования q_x, q_y, q_z на $-q_x$, $-q_y$, $-q_z$, получаем

$$G^{-1/2}\sum_{\mathbf{q},\ s}\xi^*(-\mathbf{q},\ s)\,\eta^*(-\mathbf{q},\ s,\ t)\,e^{i\mathbf{q}\rho}.$$

В силу (2.2) это есть не что иное, как сама правая часть (2.1). Таким образом, она совпадает с сопряженным ей выражением, т. е. оказывается вещественной.

Подставим выражение (2.1) в правую часть (1.9). Принимая во внимание (1.8), получим

$$H = T + V, \tag{2.3}$$

где

$$T = \frac{1}{2G}\sum_{\mathbf{q},\ \mathbf{q}'}\sum_{s,\ s'}\sum_{h} M_h\xi_{h\alpha}(\mathbf{q},\ s)\,\xi_{h\alpha}(\mathbf{q}',\ s')\times$$

$$\times\dot{\eta}(\mathbf{q},\ s,\ t)\,\dot{\eta}(\mathbf{q}',\ s',\ t)\sum_{\mathbf{g}} e^{i(\mathbf{q}+\mathbf{q}',\ \rho)}, \tag{2.4}$$

$$V = \frac{1}{2G}\sum_{\mathbf{q},\ \mathbf{q}'}\sum_{s,\ s'}\sum_{h,\ h'}\sum_{\mathbf{g},\ \mathbf{g}'}\Gamma_{hh'}^{\alpha\alpha'}(\mathbf{g}-\mathbf{g}')\times$$

$$\times\xi_{h\alpha}(\mathbf{q},\ s)\,\xi_{h'\alpha'}(\mathbf{q}',\ s')\,\eta(\mathbf{q},\ s,\ t)\,\eta(\mathbf{q}',\ s',\ t)\,e^{i(\mathbf{q}\rho+\mathbf{q}'\rho')}. \tag{2.5}$$

Как показано в Приложении VIII,

$$\sum_{\mathbf{g}} e^{i(\mathbf{q}+\mathbf{q}',\ \rho)} = G\delta_{\mathbf{q}+\mathbf{q}',\ 0}, \tag{2.6}$$

где $\delta_{\mathbf{q}+\mathbf{q}',\ 0}$ есть символ Кронекера, равный единице при $\mathbf{q}+\mathbf{q}'=0$ и нулю при $\mathbf{q}+\mathbf{q}'\neq 0$.

Следовательно, выражение (2.4) с учетом (2.2) принимает вид

$$T = \frac{1}{2}\sum_{\mathbf{q}}\sum_{s,\ s'}\sum_{h} M_h\xi_{h\alpha}(\mathbf{q},\ s)\,\xi_{h\alpha}^*(\mathbf{q},\ s')\,\eta(\mathbf{q},\ s,\ t)\,\eta^*(\mathbf{q},\ s',\ t). \tag{2.4'}$$

Обратимся теперь к вычислению суммы по $\mathbf{g}$ и $\mathbf{g}'$ в правой части (2.5). Заменим переменные суммирования, полагая

$$\mathbf{g}-\mathbf{g}'=\mathbf{g}_1, \quad \mathbf{g}+\mathbf{g}'=2\mathbf{g}_2 \tag{2.7}$$

и, соответственно,

$$\boldsymbol{\rho}-\boldsymbol{\rho}'=\boldsymbol{\rho}_1, \quad \boldsymbol{\rho}+\boldsymbol{\rho}'=2\boldsymbol{\rho}_2.$$

Получим с учетом (2.6)

$$\sum_{\mathbf{g},\,\mathbf{g}'} \Gamma^{\alpha\alpha'}_{hh'}(\mathbf{g}-\mathbf{g}')\, e^{i(\mathbf{q}\boldsymbol{\rho}+\mathbf{q}'\boldsymbol{\rho}')} =$$

$$= \sum_{\mathbf{g}_1} \Gamma^{\alpha\alpha'}_{hh'}(\mathbf{g}_1)\, e^{i/2(\mathbf{q}-\mathbf{q}',\,\boldsymbol{\rho}_1)} \sum_{\mathbf{g}_2} e^{i(\mathbf{q}+\mathbf{q}',\,\boldsymbol{\rho}_2)} =$$

$$= G \sum_{\mathbf{g}_1} \Gamma^{\alpha\alpha'}_{hh'}(\mathbf{g}_1)\, e^{i\mathbf{q}\boldsymbol{\rho}_1}\delta_{\mathbf{q}+\mathbf{q}',\,0}. \tag{2.8}$$

Пределы суммирования здесь легко устанавливаются на основании (2.7) и (1.2): $-G_x \leqslant g_{1x} \leqslant G_x$ и т. д.

Введем обозначение

$$\sum_{\mathbf{g}_1} \Gamma^{\alpha\alpha'}_{hh'}(\mathbf{g}_1)\, e^{i\mathbf{q}\boldsymbol{\rho}_1} = \Gamma^{\alpha\alpha'}_{hh'}(\mathbf{q}). \tag{2.9}$$

Тогда выражение (2.5) с учетом (2.2) и (2.8) принимает вид

$$V = \frac{1}{2}\sum_{s,\,s'}\sum_{h,\,h'}\sum_{\mathbf{q}} \Gamma^{\alpha\alpha'}_{hh'}(\mathbf{q})\,\zeta_{h\alpha}(\mathbf{q},\,s)\,\zeta^*_{h'\alpha'}(\mathbf{q},\,s')\,\eta(\mathbf{q},\,s,\,t)\,\eta^*(\mathbf{q},\,s',\,t). \tag{2.5'}$$

Из формул (2.3), (2.4') и (2.5') видно, что слагаемые с различными $\mathbf{q}$ не «перемешиваются» в выражении для полной энергии: последняя представляется суммой членов, каждый из которых относится только к одному вектору $\mathbf{q}$. Это означает (с учетом (1.1)), что зависимость от $\mathbf{q}$ в линейной комбинации (2.1) выбрана правильно и нормальные колебания действительно можно характеризовать квазиволновым вектором $\mathbf{q}$. Осталось потребовать, чтобы в выражениях (2.4'), (2.5') обратились в нуль слагаемые с $s \neq s'$. Можно убедиться, что это условие выполняется, если величины $\zeta_{h\alpha}$ удовлетворяют системе уравнений

$$\sum_{h} \Gamma^{\alpha\alpha'}_{hh'}(\mathbf{q})\,\zeta_{h\alpha}(\mathbf{q},\,s) = \omega_s^2(\mathbf{q})\, M_{h'}\zeta_{h'\alpha'}(\mathbf{q},\,s), \tag{2.10}$$

где $\omega_s(\mathbf{q})$ — некоторые числа размерности частоты. (Из дальнейшего будет видно, что они представляют собой не что иное, как частоты нормальных колебаний.) Действительно, в Приложении IX показано, что из системы (2.10) вытекает условие ортогональности

$$\sum_{h} M_h \zeta^*_{h\alpha}(\mathbf{q},\,s)\,\zeta_{h\alpha}(\mathbf{q},\,s') = 0, \qquad s' \neq s. \tag{2.11}$$

Из соотношений (2.10) и (2.11) сразу видно, что слагаемые с $s' \neq s$ действительно обращаются в нуль как в (2.4'), так и в (2.5'). Виден также формальный смысл индекса s: он нумерует различные решения системы (2.10) (при заданном векторе $\mathbf{q}$).

Отметим некоторые важные свойства системы (2.10). Прежде всего, из соотношения симметрии (1.7') и определения (2.9) вытекает равенство

$$\Gamma^{\alpha\alpha'}_{hh'}(\mathbf{q}) = [\Gamma^{\alpha'\alpha}_{h'h}(-\mathbf{q})]^*. \tag{2.12}$$

Таким образом, матрица коэффициентов в (2.10) переходит сама в себя, если заменить $\mathbf{q}$ на $-\mathbf{q}$ и одновременно выполнить комплексное сопряжение. То же, очевидно, справедливо и для детерминанта этой матрицы и для его миноров, а потому и для величин $\omega_s^2(q)$ и для векторов $\xi(\mathbf{q}, s)$, которые через них выражаются:

$$\xi(\mathbf{q}, s) = \xi^*(-\mathbf{q}, s), \qquad \omega_s^2(\mathbf{q}) = [\omega_s^2(-\mathbf{q})]^*. \tag{2.13}$$

Еще одно свойство коэффициентов $\Gamma^{\alpha\alpha'}_{hh'}(\mathbf{q})$ можно установить, замечая, что потенциальная энергия кристалла, равно как и ее производные по смещениям, должна быть инвариантной относительно сдвига решетки как целого. Действительно, при таком сдвиге расстояния между атомами не меняются, а потому не могут измениться и силы взаимодействия между ними. С другой стороны, формально мы можем, смещая все атомы на постоянный вектор $\mathbf{Q}_0$, вычислить изменение потенциальной энергии или ее производных при таком сдвиге. Требуя, чтобы это изменение обращалось в нуль, получим некоторое тождественное соотношение между коэффициентами $\Gamma^{\alpha\alpha'}_{hh'}(\mathbf{q})$.

Удобно рассмотреть первую производную от потенциальной энергии по какой-нибудь компоненте вектора смещения $\dfrac{\partial V}{\partial Q_\alpha(\mathbf{\rho}, h)}$ (тот факт, что она равна нулю в условиях равновесия, не играет роли). Очевидно, изменение ее при сдвиге $\mathbf{Q}_0$ есть

$$\frac{\partial V}{\partial Q_\alpha(\mathbf{\rho}, h)} - \frac{\partial V}{\partial Q_\alpha(\mathbf{\rho}, h)}\Big|_0 = \sum_{h'=1}^{r} \sum_{\mathbf{\rho}'} \frac{\partial^2 V}{\partial Q_\alpha(\mathbf{\rho}, h)\, \partial Q_{\alpha'}(\mathbf{\rho}', h')}\Big|_0 Q_{0,\alpha} + \ldots \tag{2.14}$$

Здесь индекс «0» указывает, что соответствующая величина вычисляется при $\mathbf{Q}_0 = 0$, а многоточием обозначены члены, содержащие высшие степени компонент $\mathbf{Q}_0$. Их мы всегда можем отбросить, считая величину $|\mathbf{Q}_0|$ достаточно малой. Принимая во внимание определение величин $\Gamma^{\alpha\alpha'}_{hh'}$ (см. (1.7)), мы получаем из (2.14)

$$\sum_{h'=1}^{r} \sum_{\mathbf{\rho}'} \Gamma^{\alpha\alpha'}_{hh'}(\mathbf{\rho} - \mathbf{\rho}')\, Q_{0,\alpha'} \equiv 0. \tag{2.15}$$

Суммирование по $\mathbf{\rho}'$, очевидно, можно заменить суммированием по разности $\mathbf{\rho} - \mathbf{\rho}' = \mathbf{\rho}''$. Поскольку компоненты $Q_{0,\alpha'}$ независимы,

равенство (2.15) дает

$$\sum_{h'=1}^{r}\sum_{\mathbf{g}'}\Gamma_{hh'}^{\alpha\alpha'}(\mathbf{g}')\equiv 0. \tag{2.16}$$

Ради единства обозначений мы заменили здесь индекс суммирования $\mathbf{g}''$ на $\mathbf{g}'$. Соотношение (2.16) должно выполняться тождественно для любой кристаллической решетки.

В частности, в простой решетке, когда в элементарной ячейке имеется всего один атом ($r = 1$), суммирование по h' отпадает ($h' = h = 1$) и тождество (2.16) принимает вид

$$\sum_{\mathbf{g}'}\Gamma_{11}^{\alpha\alpha'}(\mathbf{g}')\equiv 0. \tag{2.16'}$$

Обращаясь теперь к определению (2.9), видим, что сумма по $\mathbf{g}'$ в левой части (2.16) есть не что иное, как значение $\Gamma_{hh'}^{\alpha\alpha'}(\mathbf{q})$ при $\mathbf{q} = 0$. Таким образом, мы имеем

$$\sum_{h'=1}^{r}\Gamma_{hh'}^{\alpha\alpha'}(\mathbf{q})\,|_{\mathbf{q}=0}=0. \tag{2.17}$$

В простой решетке сумма по h' отпадает ($h = h' = 1$) и

$$\Gamma_{11}^{\alpha\alpha'}(\mathbf{q})\,|_{\mathbf{q}=0}=0. \tag{2.17'}$$

Система уравнений (2.10) представляет собой стандартную математическую задачу на собственные значения: величины $\omega_s^2(\mathbf{q})$ и $\zeta_{h\alpha}(\mathbf{q}, s)$ суть, соответственно, собственные значения и собственные векторы матрицы $\Gamma_{hh'}^{\alpha\alpha'}(\mathbf{q})$. Согласно (2.12) эта матрица эрмитова. В линейной алгебре доказывается, что собственные значения таких матриц $\omega_s^2(\mathbf{q})$ вещественны. Основываясь на свойстве минимальности потенциальной энергии в положении равновесия, можно доказать также [1], что собственные значения $\omega_s^2(\mathbf{q})$ неотрицательны, т. е. сами частоты $\omega_s(\mathbf{q})$ вещественны. В сочетании с (2.13) отсюда следует

$$\omega_s(\mathbf{q}) = \omega_s(-\mathbf{q}), \tag{2.18}$$

т. е. частоты нормальных колебаний суть четные функции квазиволнового вектора.

При $s' = s$ сумма в левой части (2.11) отлична от нуля. Ее значение в этом случае можно назначить по произволу. Действительно, из формулы (2.1) видно, что изменение вектора $\zeta_h(\mathbf{q}, s)$ на постоянный множитель сводится (при заданном векторе смещения $\mathbf{Q}_a$) просто к изменению масштаба обобщенных координат η. Далее, система (2.10) однородна и, следовательно, также определяет $\zeta_h(\mathbf{q}, s)$ лишь с точностью до постоянного множителя. Выбор последнего есть вопрос удобства. Для дальнейшего удобно сохранить за координатами η размерность длины. Тогда функции $\zeta_h(\mathbf{q}, s)$ должны

быть безразмерными и условие нормировки удобно записать в виде

$$\sum_{h=1}^{r} M_h |\xi_h(\mathbf{q}, s)|^2 = M. \tag{2.19}$$

Здесь M есть масса элементарной ячейки:

$$M = \sum_{h=1}^{r} M_h.$$

В частности, если массы всех атомов одинаковы, то вектор ξ_h оказывается единичным:

$$|\xi_h(\mathbf{q}, s)|^2 = 1. \tag{2.19'}$$

Подставляя теперь выражения (2.4') и (2.5') в формулу (2.3) и пользуясь равенствами (2.10), (2.11) и (2.19) мы получаем

$$H = \frac{1}{2} \sum_{\mathbf{q}, s} M \{|\dot{\eta}(\mathbf{q}, s, t)|^2 + \omega_s^2(\mathbf{q}) |\eta(\mathbf{q}, s, t)|^2\}. \tag{2.20}$$

Это выражение формально похоже на энергию системы независимых гармонических осцилляторов, частоты которых равны $\omega_s(\mathbf{q})$. Однако, как уже отмечалось, комплексные нормальные координаты $\eta(\mathbf{q}, s, t)$ фактически не независимы, а связаны условием (2.2). Последнее, согласно (2.13), можно переписать в виде

$$\eta(\mathbf{q}, s, t) = \eta^*(-\mathbf{q}, s, t). \tag{2.2'}$$

Чтобы представить энергию системы в виде суммы энергий независимых гармонических осцилляторов, надо выразить комплексные величины $\eta(\mathbf{q}, s, t)$ через вещественные нормальные координаты так, чтобы условие (2.2') удовлетворялось автоматически. При этом вещественные нормальные координаты, которые мы обозначим через $x(\mathbf{q}, s, t)$, должны гармонически зависеть от времени. Мы обеспечим это, подчинив их уравнению (впредь для краткости мы не будем выписывать аргумент t у функции $x(\mathbf{q}, s, t)$)

$$\ddot{x}(\mathbf{q}, s) + \omega^2(\mathbf{q}, s) x(\mathbf{q}, s) = 0. \tag{2.21}$$

Положим

$$\eta(\mathbf{q}, s, t) = \frac{1}{2} \Big\{x(\mathbf{q}, s) + x(-\mathbf{q}, s) + \frac{i}{\omega(\mathbf{q}, s)} [\dot{x}(\mathbf{q}, s) - \dot{x}(-\mathbf{q}, s)]\Big\}. \tag{2.22}$$

Непосредственно видно, что условие (2.2') при этом удовлетворяется. Заметим, далее, что, согласно (2.21) и (2.22),

$$\dot{\eta}(\mathbf{q}, s, t) = \frac{1}{2} \{\dot{x}(\mathbf{q}, s) + \dot{x}(-\mathbf{q}, s) - i\omega(\mathbf{q}, s) [x(\mathbf{q}, s) - x(-\mathbf{q}, s)]\}. \tag{2.23}$$

Подставляя выражения (2.22) и (2.23) в правую часть (2.20), находим окончательное выражение для энергии кристаллической решетки, совершающей малые колебания около положений равновесия:

$$H=\frac{1}{2}\sum_{\mathbf{q},\,s}\{M\dot{x}^2(\mathbf{q},\,s)+M\omega^2(\mathbf{q},\,s)\,x^2(\mathbf{q},\,s)\}. \tag{2.24}$$

В правой части (2.24) стоит сумма энергий независимых гармонических осцилляторов, частоты которых равны $\omega(\mathbf{q}, s)$, а массы равны M. Координаты этих осцилляторов представляют собой искомые нормальные координаты решетки.

Для перехода к квантовой теории удобно вместо обобщенных скоростей $\dot{x}(\mathbf{q}, s)$ ввести соответствующие импульсы

$$p(\mathbf{q},\,s)=M\dot{x}(\mathbf{q},\,s). \tag{2.25}$$

Тогда

$$H=\frac{1}{2}\sum_{\mathbf{q},\,s}\left\{\frac{p^2(\mathbf{q},\,s)}{M}+M\omega^2(\mathbf{q},\,s)\,x^2(\mathbf{q},\,s)\right\}. \tag{2.24'}$$

Выразим теперь нормальные координаты через векторы смещения. Пользуясь формулами (2.6) и (2.1), мы получаем

$$\eta(\mathbf{q},\,s,\,t)=\frac{1}{MG}\sum_{g,\,h}M_hQ_{g,\,h,\,\alpha}\zeta^*_{h,\,\alpha}(\mathbf{q},\,s)\,e^{-i\mathbf{q}\boldsymbol{\rho}}. \tag{2.26}$$

Отсюда видно, что в идеальной решетке каждое нормальное колебание охватывает, вообще говоря, все атомы и не может быть приписано какому-то одному из них. Иначе говоря, нормальные координаты описывают коллективные движения частиц системы.

Применим уравнение (2.10) к частному случаю линейной цепочки одинаковых атомов, вытянутой вдоль оси x. При этом «элементарная ячейка» состоит из одного атома ($r = 1$), т. е. всюду следует положить $h' = h = 1$ и $M_h = M$; векторные индексы α, α' также принимают лишь по одному значению ($\alpha' = \alpha = x$). Соответственно значки h, h', α, α' можно вообще опустить, и уравнение (2.10) принимает вид

$$\Gamma(\mathbf{q})\,\zeta(\mathbf{q},\,s)=M\omega_s^2(\mathbf{q})\,\zeta(\mathbf{q},\,s). \tag{2.27}$$

Это уравнение имеет нетривиальное решение лишь при

$$\omega_s^2(\mathbf{q})=\Gamma(\mathbf{q})/M. \tag{2.28}$$

Таким образом, здесь возможен лишь один тип нормальных колебаний; по этой причине индекс s также можно опустить. Вектор $\boldsymbol{\rho}$ (см. (1.1)) имеет теперь лишь одну компоненту $\rho = ga$, где g — номер атома в цепочке, а a — расстояние между соседними атомами. Соответственно равенства (1.7), (2.9) и (2.17′) принимают

вид

$$\Gamma(g-g') = \frac{\partial^2 V}{\partial R_g \partial R_{g'}}\Big|_{R=R^0}, \tag{2.29}$$

$$\Gamma(q) = \sum_{g_1=-\infty}^{+\infty} \Gamma(g_1)\, e^{iqag_1} = \Gamma(0) + 2\sum_{g_1=1}^{\infty} \Gamma(g_1)\cos(qag_1) \tag{2.30}$$

и

$$\Gamma(0) + 2\sum_{g_1=1}^{\infty} \Gamma(g_1) = 0. \tag{2.31}$$

При этом, в силу (1.7″), $\Gamma(0) > 0$. Принимая во внимание это обстоятельство и пользуясь равенством (2.31), можем переписать выражение (2.30) в виде

$$\Gamma(q) = -4\sum_{g_1=1}^{\infty} \Gamma(g_1)\sin^2\left(\frac{qag_1}{2}\right), \tag{2.30'}$$

причем

$$-\sum_{g_1=1}^{\infty} \Gamma(g_1) > 0.$$

Наконец, вместо равенства (2.19′) мы получаем

$$\zeta^2 = 1.$$

Таким образом, частоты нормальных колебаний даются выражением

$$\omega(q) = \frac{2}{M}\sqrt{-\sum_{g_1=1}^{\infty} \Gamma(g_1)\sin^2\left(\frac{qag_1}{2}\right)}. \tag{2.32}$$

При малых значениях волнового числа q это дает

$$\omega(q) = \frac{a}{M}\sqrt{\left|\sum_{g_1=1}^{\infty} \Gamma(g_1)\, g_1^2\right|}\, q. \tag{2.32'}$$

§ 3. Частоты нормальных колебаний. Акустические и оптические ветви

Обратимся к более подробному рассмотрению однородной системы уравнений (2.10). Она имеет нетривиальные решения, лишь если детерминант ее обращается в нуль:

$$\begin{vmatrix} \Gamma_{11}^{xx} - M_1\omega_s^2 & \Gamma_{11}^{xy} & \Gamma_{11}^{xz} & \Gamma_{12}^{xx} & \dots & \Gamma_{1r}^{xz} \\ \Gamma_{11}^{yx} & \Gamma_{11}^{yy} - M_1\omega_s^2 & \Gamma_{11}^{yz} & \Gamma_{12}^{yx} & \dots & \Gamma_{1r}^{yz} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \Gamma_{r1}^{zx} & \Gamma_{r1}^{zy} & \Gamma_{r1}^{zz} & \Gamma_{r2}^{zx} & \dots & \Gamma_{rr}^{zz} - M_r\omega_s^2 \end{vmatrix} = 0. \tag{3.1}$$

Поскольку индекс h меняется от 1 до r, а $\alpha = x, y, z$, это есть уравнение степени $3r$ относительно ω_s^2. При заданном векторе $\mathbf{q}$ оно имеет $3r$ корней — собственных частот системы. Соответственно мы получаем $3r$, вообще говоря, различных функций $\boldsymbol{\zeta}(\mathbf{q}, s)$, описывающих различные «ветви» колебаний.

Компоненты вектора $\mathbf{q}$ при этом играют роль параметров. При изменении их собственные частоты и собственные векторы, разумеется, меняются. Зависимость ω_s от $\mathbf{q}$ называется законом дисперсии для нормальных колебаний s-й ветви. Некоторые сведения о законе дисперсии уже были получены в предыдущем параграфе: как мы знаем, частоты представляют собой вещественные и четные функции $\mathbf{q}$. Далее, легко убедиться, что $\Gamma_{hh'}^{\alpha\alpha'}(\mathbf{q})$ периодически зависит от вектора $\mathbf{q}$ с периодом обратной решетки $\mathbf{b}$. Действительно, согласно (1.1) при добавлении к вектору $\mathbf{q}$ слагаемого $\mathbf{b}$ выражение под знаком суммы в формуле (2.9) не изменяется. Отсюда следует, что и величины $\omega_s(q) \equiv \omega(\mathbf{q}, s)$ и $\boldsymbol{\zeta}(\mathbf{q}, s)$ суть периодические функции $\mathbf{q}$ с периодом, равным вектору обратной решетки.

Обратим внимание на аналогию между этими свойствами и соответствующими свойствами энергии и волновой функции электрона в периодической решетке (§§ III.2, III.3). Эта аналогия обусловлена глубокой физической причиной: в обоих случаях мы имеем трансляционно инвариантную систему. Действительно, в гл. III речь шла о движении электрона в периодическом силовом поле; в настоящей же главе рассматриваются малые колебания атомов около периодически распределенных положений равновесия.

Как и в § III.3, периодическая зависимость $\omega(\mathbf{q}, s)$ и $\boldsymbol{\zeta}(\mathbf{q}, s)$ от квазиволнового вектора позволяет ограничить рассматриваемый интервал значений $\mathbf{q}$, введя представление о зонах Бриллюэна. В частности, первая зона Бриллюэна определяется соотношениями (III.3.8). Поскольку последние носят чисто геометрический характер, мы можем говорить просто о зонах Бриллюэна, не уточняя, что́ именно характеризуют соответствующие квазиволновые векторы — движение электрона или нормальные колебания решетки.

Тождества (2.17), (2.17′) позволяют исследовать общий вид закона дисперсии при малых квазиволновых векторах, т. е. для длинных волн («длинными» в данном случае надо называть волны, длины которых велики по сравнению с постоянной решетки).

Рассмотрим сначала простую решетку ($r = 1$). В этом случае уравнение (3.1) — кубическое. Следовательно, имеются три ветви колебаний: $s = 1, 2, 3$. При $\mathbf{q} = 0$ система (2.10) принимает вид

$$\Gamma_{11}^{\alpha\alpha'}(0)\,\zeta_{1\alpha}(0, s) = \omega_s^2(0)\,\zeta_{1\alpha'}(0, s). \tag{3.2}$$

В силу (2.17′) левая часть (3.2) обращается в нуль, и, поскольку вектор $\boldsymbol{\zeta}$ не должен быть тождественно равен нулю:

$$\omega_s(0) = 0.$$

Таким образом, при $\mathbf{q} = 0$ частоты колебаний всех трех ветвей обращаются в нуль. Поскольку $\omega_s(\mathbf{q})$ есть четная функция $\mathbf{q}$, отсюда, в сочетании с определением (2.9), следует, что для длинных волн ω_s^2 есть квадратичная функция $\mathbf{q}$,

$$\omega_s^2(\mathbf{q}) = c_{\alpha\beta}(s)\, q_\alpha q_\beta, \tag{3.3}$$

т. е.

$$\omega_s = c_s q. \tag{3.3'}$$

Величины c_s зависят, вообще говоря, от направления вектора $\mathbf{q}$ (и, конечно, от номера ветви). Аналогичную формулу, но с постоянными величинами c_s мы получили бы, рассматривая распространение малых колебаний плотности в упругом изотропном континууме. При этом, как известно, одна ветвь (припишем ей индекс $s = 1$) описывает продольные волны (это есть обычный звук), а две другие ($s = 2, 3$) — поперечные волны. В анизотропной среде четкое разделение волн на продольные и поперечные возможно лишь для некоторых особых направлений; вообще же говоря, угол между векторами $\mathbf{q}$ и $\boldsymbol{\xi}(\mathbf{q}, s)$ оказывается отличным как от нуля, так и от $\pi/2$. Иными словами, каждая ветвь имеет, вообще говоря, как поперечную, так и продольную компоненты. Тем не менее величины c_s часто называют скоростями звука.

В сложных решетках, когда $r > 1$, также всегда имеются три ветви колебаний, у которых закон дисперсии при малых $\mathbf{q}$ имеет вид (3.3). Этот факт явствует просто из того, что при длинных волнах все атомы в элементарной ячейке могут смещаться практически одинаковым образом, т. е. существуют движения, при которых элементарная ячейка колеблется как целое. Сложная структура ячейки при этом никак не проявляется. Формально в существовании таких ветвей можно убедиться, написав систему (2.10) при $\mathbf{q} = 0$ и предположив заранее, что $\omega_s(0) = 0$:

$$\sum_{h=1}^{r} \Gamma_{hh'}^{\alpha\alpha'}(0)\, \zeta_{h\alpha}(0, s) = 0. \tag{3.4}$$

Из тождества (2.17) вытекает, что эта система имеет нетривиальные решения, не зависящие от h:

$$\zeta_h(0, s) = \xi(s). \tag{3.5}$$

В самом деле, подставляя (3.5) в (3.4), видим, что последнее уравнение удовлетворяется в силу (2.17).

Таким образом, вектор смещения действительно оказывается одним и тем же для всех атомов данной ячейки. Ветви такого типа называются *акустическими*; для некоторых особых направлений их, как и в простой решетке, можно разделить на одну продольную и две поперечные. Мы припишем им значения индекса $s = 1, 2, 3$.

В рассматриваемом случае уравнение (3.1) — степени $3r \geqslant 6$. Следовательно, помимо акустических, должны существовать еще $3r-3$ ветви колебаний, не имеющих себе аналога в динамике сплошной среды. Закон дисперсии для этих ветвей уже не имеет вида (3.3), т. е. при $\mathbf{q} \to 0$ частота колебаний уже не обращается в нуль. Поскольку частота, согласно (2.18), есть четная функция $\mathbf{q}$, для длинных волн величина $\omega(\mathbf{q}, s)$ $(s \geqslant 4)$ почти постоянна *):

$$\omega(\mathbf{q}, s) = \omega_0(s) - \alpha_s \mathbf{q}^2. \tag{3.6}$$

Коэффициент α_s обычно больше нуля. Частота $\omega_0(s)$ называется предельной. Колебания этого типа отличаются тем свойством, что при $\mathbf{q} \to 0$ центр тяжести элементарной ячейки остается в покое. Действительно, при $\mathbf{q} = 0$ и $\omega = \omega_0 \neq 0$ система (2.10) принимает вид

$$\sum_{h=1}^{r} \Gamma_{hh'}^{\alpha\alpha'}(0)\, \zeta_{h\alpha}(0, s) = M_{h'} \omega_0^2(s)\, \zeta_{h'\alpha'}(0, s).$$

Суммируя эти равенства по h' и переставляя в левой части порядок суммирования по h и h', получаем

$$\sum_{h=1}^{r} \zeta_{h\alpha}(0, s) \sum_{h'=1}^{r} \Gamma_{hh'}^{\alpha\alpha'}(0) = \sum_{h'=1}^{r} M_{h'} \omega_0^2(s)\, \zeta_{h'\alpha'}(0, s).$$

Согласно (2.17) левая часть этого равенства тождественно равна нулю. Следовательно, при $\omega_0(s) \neq 0$

$$\sum_{h'=1}^{r} M_{h'} \zeta_{h'\alpha'}(0, s) = 0. \tag{3.7}$$

В левой части (3.7) стоит не что иное, как вектор смещения центра тяжести элементарной ячейки, умноженный на M. Этим и доказывается высказанное выше утверждение. Ветви колебаний с законом дисперсии вида (3.6) называются *оптическими*. Физическую природу различия между акустическими и оптическими колебаниями легко понять, рассматривая кристалл с двумя атомами в элементарной ячейке. Как мы видели, при длинноволновых акустических колебаниях эти атомы движутся почти синфазно, а при длинноволновых оптических, согласно (3.7), — почти в противофазе (рис. 12.1). Очевидно, в ионных кристаллах при колебаниях последнего типа может сильно изменяться дипольный момент элементарной ячейки. Поэтому некоторые оптические колебания в этих кристаллах сравнительно легко возбуждаются переменным электромагнитным полем подходящей частоты. Рассмотренный процесс составляет один из важных механизмов поглощения электромагнит-

*) Выражение (3.6) получается в результате разложения $\omega(\mathbf{q}, s)$ при $s \geqslant 4$ по степеням q^2 и пренебрежения величинами порядка $(q^2)^2$, $(q^2)^3$ и т. д.

ных волн ионными кристаллами. С этим обстоятельством связано и само название «оптические».

В решетке с бо́льшим числом атомов в элементарной ячейке фазовые соотношения между смещениями различных атомов усложняются, однако суть различия между длинноволновыми акустическими и оптическими колебаниями остается прежней: первые



Рис. 12.1. Смещения атомов при длинноволновых акустических и оптических колебаниях решетки в двухатомном кристалле. Кружками изображены атомы, входящие в данную элементарную ячейку.

отвечают смещениям элементарной ячейки как целого, вторые — внутренним деформациям в ней при почти неподвижном центре тяжести ячейки.

При уменьшении длины волны закон дисперсии делается более сложным и его трудно установить в общем виде. Можно лишь утверждать, что при всех возможных значениях q частота колебаний остается ограниченной. В конечном счете это связано с тем, что соответствующая длина волны $2\pi/q$ не может быть меньше расстояния между соседними атомами. Наибольшая частота акустических колебаний называется *дебаевской* (ω_D); с ней можно связать характерную *дебаевскую температуру* T_D, полагая $\hbar\omega_D = kT_D$.

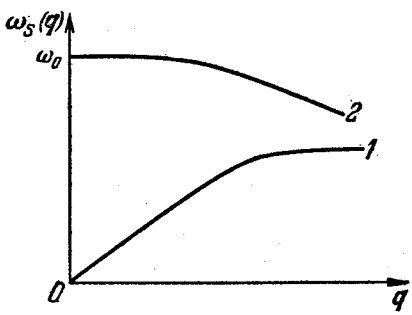


Рис. 12.2. Законы дисперсии для акустических (кривая *1*) и оптических (кривая *2*) фононов в одномерной цепочке атомов (схематически).

Для определения зависимости $\omega_s(\mathbf{q})$ во всей зоне Бриллюэна приходится либо выполнять численные расчеты в рамках каких-либо конкретных моделей, либо использовать экспериментальные данные. В этом отношении весьма полезными оказываются опыты по неупругому рассеянию проникающих излучений, особенно медленных нейтронов, кристаллами. Измеряя угловое распределение рассеянных нейтронов и пользуясь законами сохранения энергии и квазиимпульса, удается с хорошей точностью восстановить вид закона дисперсии $\omega_s(\mathbf{q})$ *). Примерный вид зависимости $\omega_s(\mathbf{q})$ для акустической и оптической ветвей представлен на рис. 12.2 (кривые *1* и *2*).

*) Подробнее об этом можно прочитать в книге [2].

§ 4. Вектор смещения

Выразим теперь вектор смещения a-го атома через вещественные нормальные координаты $x(\mathbf{q}, s)$. Согласно формуле (2.1)

$$\mathbf{Q}_a = G^{-1/2} \sum_{\mathbf{q},\, s} \xi_h(\mathbf{q},\ s)\, \eta(\mathbf{q},\ s)\, e^{i\mathbf{q}\varrho}. \tag{4.1}$$

Подставляя выражение (2.22) в правую часть (4.1), мы получаем

$$\mathbf{Q}_a = \frac{1}{2\sqrt{G}} \sum_{\mathbf{q},\, s} \xi_h(\mathbf{q},\ s)\, e^{i\mathbf{q}\varrho} \times$$

$$\times \left\{ x(\mathbf{q},\ s) + x(-\mathbf{q},\ s) + i\, \frac{p(\mathbf{q},\ s) - p(-\mathbf{q},\ s)}{M\omega(\mathbf{q},\ s)} \right\}. \tag{4.2}$$

Произведем здесь в слагаемых, содержащих $x(-\mathbf{q}, s)$ и $p(-\mathbf{q}, s)$, замену $\mathbf{q} \to -\mathbf{q}$ и примем во внимание равенства (2.13). Получим окончательно

$$\mathbf{Q}_a = \frac{1}{2\sqrt{G}} \sum_{\mathbf{q}} \sum_{s=1}^{r} \left\{ \xi_h(\mathbf{q},\ s) \left[x(\mathbf{q},\ s) + i\, \frac{p(\mathbf{q},\ s)}{M\omega(\mathbf{q},\ s)} \right] e^{i\mathbf{q}\varrho} + \right.$$

$$\left. + \xi_h^*(\mathbf{q},\ s) \left[x(\mathbf{q},\ s) - i\, \frac{p(\mathbf{q},\ s)}{M\omega(\mathbf{q},\ s)} \right] e^{-i\mathbf{q}\varrho} \right\}. \tag{4.3}$$

Поскольку величины $x(\mathbf{q}, s)$ и $p(\mathbf{q}, s)$ гармонически зависят от времени, правая часть формулы (4.3) представляет собой набор бегущих волн; длины волн λ связаны с абсолютными значениями квазиволновых векторов обычным соотношением $\lambda = 2\pi/q$.

Рассмотрим, в частности, одноатомный кристалл ($r = 1$). Тогда компоненты вектора $\boldsymbol{\xi}$ оказываются вещественными. Как видно из системы уравнений (2.10), чтобы убедиться в этом, достаточно доказать вещественность коэффициентов $\Gamma^{\alpha\alpha'}(\mathbf{q})$, определяемых равенством (2.9) (мы опускаем теперь индексы h, h', ибо элементарная ячейка содержит только один атом). В кристаллах рассматриваемого типа каждый атом представляет собой центр симметрии. Следовательно, если в левой части (2.9) имеется слагаемое с некоторым вектором $\mathbf{q}_1$, то там будет и другое слагаемое с равным по величине и противоположно направленным вектором $-\mathbf{q}_1$. Таким образом, равенство (2.9) принимает вид

$$\Gamma^{\alpha\alpha'}(\mathbf{q}) = \Gamma^{\alpha\alpha'}(0) + 2 \sum_{\mathbf{g}_1} \Gamma^{\alpha\alpha'}(\mathbf{g}_1) \cos(\mathbf{q}\varrho_1),$$

т. е. коэффициенты $\Gamma^{\alpha\alpha'}(\mathbf{q})$, а с ними и компоненты вектора $\boldsymbol{\xi}$ вещественны, что и требовалось доказать.

Очевидно, эти соображения справедливы не только для простой решетки, но и вообще для любой решетки, в которой каждый атом является центром симметрии. Видно также, что в любой сколь

угодно сложной решетке коэффициенты $\Gamma_{hh'}^{\alpha\alpha'}(\mathbf{q})$ вещественны при $\mathbf{q} = 0$ и $\mathbf{qa} = \pm\pi$, т. е. для предельно длинных и предельно коротких волн.

Формула (4.3) описывает смещения атомов в дискретной решетке. В ряде явлений, однако, существенны только волны смещения с длиной, значительно превышающей постоянную решетки. При этом факт дискретности решетки не играет роли. Очевидно, в случае длинных волн дискретный аргумент $\boldsymbol{\rho}$ в формуле (4.3) можно заменить непрерывно меняющимся вектором $\mathbf{r}$ — одним и тем же для всех атомов в данной элементарной ячейке. Иначе говоря, в сумме по $\mathbf{q}$ в правой части (4.3) имеет смысл выделить длинноволновую часть $\mathbf{Q}_{дл}$. В нее входят только слагаемые с квазиволновыми векторами, удовлетворяющими условию

$$2\pi/q \gg |\mathbf{a}|. \tag{4.4}$$

Для акустических колебаний переход от дискретного вектора $\boldsymbol{\rho}$ к непрерывному означает не что иное, как переход к макроскопической теории упругости, в которой кристалл рассматривается как непрерывная среда. При этом все атомы в элементарной ячейке колеблются практически в одной и той же фазе и индекс h у вектора $\boldsymbol{\xi}$ можно опустить. Таким образом, длинноволновую часть вектора смещения, связанную с колебаниями акустического типа, можно записать в виде

$$\mathbf{Q}_{дл.ак}(\mathbf{r}) = \frac{1}{2\sqrt{G}} \sum_{\mathbf{q}} \sum_{s=1}^{3} \left\{ \boldsymbol{\xi}(\mathbf{q}, s) \left[x(\mathbf{q}, s) + i \frac{p(\mathbf{q}, s)}{M\omega(\mathbf{q}, s)} \right] e^{i\mathbf{qr}} + \right.$$
$$\left. + \boldsymbol{\xi}^*(\mathbf{q}, s) \left[x(\mathbf{q}, s) - i \frac{p(\mathbf{q}, s)}{M\omega(\mathbf{q}, s)} \right] e^{-i\mathbf{qr}} \right\}. \tag{4.5}$$

С другой стороны, при рассмотрении оптических колебаний, хотя бы и длинноволновых, полностью пренебрегать дискретностью решетки и отбрасывать индекс h нельзя. Так, в двухатомной решетке векторы $\boldsymbol{\xi}_h$ у атомов, входящих в одну ячейку, направлены почти в противоположные стороны. Вместе с тем замена дискретного вектора $\boldsymbol{\rho}$ на непрерывный по-прежнему возможна: разность фаз между колебаниями соответственных атомов в соседних ячейках невелика. Таким образом, длинноволновая часть вектора смещения, связанная с колебаниями оптического типа, также дается формулой вида (4.5):

$$\mathbf{Q}_{дл.опт}(\mathbf{r}) = \frac{1}{2\sqrt{G}} \sum_{\mathbf{q}} \sum_{s \geqslant 4} \left\{ \boldsymbol{\xi}_h(\mathbf{q}, s) \left[x(\mathbf{q}, s) + i \frac{p(\mathbf{q}, s)}{M\omega(\mathbf{q}, s)} \right] e^{i\mathbf{qr}} + \right.$$
$$\left. + \boldsymbol{\xi}_h^*(\mathbf{q}, s) \left[x(\mathbf{q}, s) - i \frac{p(\mathbf{q}, s)}{M\omega(\mathbf{q}, s)} \right] e^{-i\mathbf{qr}} \right\}. \tag{4.6}$$

§ 5. Квантовомеханическое рассмотрение колебаний решетки

Выражение для энергии (2.24′) в равной мере справедливо как в классической, так и в квантовой теории кристаллической решетки. В последнем случае следует лишь считать величины p и x операторами, удовлетворяющими следующему правилу перестановки:

$$x(\mathbf{q}, s)\,p(\mathbf{q}', s') - p(\mathbf{q}', s')\,x(\mathbf{q}, s) = i\hbar\delta_{\mathbf{q}\mathbf{q}'}\delta_{ss'}. \qquad (5.1)$$

Как и раньше, через $\delta_{\mathbf{q}\mathbf{q}'}$ и $\delta_{ss'}$ здесь обозначены символы Кронекера (П.XI.9). В частности, можно рассматривать $x(\mathbf{q}, s)$ как число; тогда $p(\mathbf{q}, s) = -i\hbar\frac{\partial}{\partial x(\mathbf{q}, s)}$. Оператор H теперь представляет собой гамильтониан системы независимых осцилляторов, координаты и импульсы которых суть $x(\mathbf{q}, s)$ и $p(\mathbf{q}, s)$. Соответственно уравнение Шредингера для системы атомов, колеблющихся около периодически расположенных в пространстве положений равновесия, можно записать в виде

$$\sum_{\mathbf{q}, s}\left\{\frac{p^2(\mathbf{q}, s)}{2M} + \frac{M\omega^2(\mathbf{q}, s)}{2}x^2(\mathbf{q}, s)\right\}\Phi = E\Phi. \qquad (5.2)$$

Здесь Φ есть волновая функция, описывающая малые колебания решетки, а E — принадлежащее ей собственное значение энергии. В качестве аргументов Φ можно выбрать, например, совокупность всех нормальных координат $x(\mathbf{q}, s)$. Общее число этих координат, очевидно, равно числу степеней свободы системы $3rG$. Действительно, имеется G независимых компонент вектора $\mathbf{q}$, а индекс s, как мы знаем, принимает все значения от 1 до $3r$. Обозначим для удобства совокупность переменных $\mathbf{q}$ и s одним индексом f:

$$f = \{\mathbf{q}, s\}, \quad f = 1, \ldots, 3rG \equiv N, \quad \omega(\mathbf{q}, s) = \omega_f. \qquad (5.3)$$

Тогда $x(\mathbf{q}, s) = x_f$ и $\Phi = \Phi(x_1, x_2, \ldots, x_N)$, где, например, x_1 есть нормальная координата, описывающая колебания типа $s = 1$ с волновым вектором $\mathbf{q} = \mathbf{q}_1$, и т. д. По правилам квантовой механики величина $|\Phi|^2$ определяет плотность вероятности тех или иных значений $x_1, \ldots, x_N$. Именно, выражение

$$|\Phi(x_1, x_2, \ldots, x_N)|^2\,dx_1\,dx_2\ldots dx_N \qquad (5.4)$$

есть вероятность того, что значение первой нормальной координаты лежит в малом интервале dx_1 около точки x_1, второй — в интервале dx_2 около точки x_2 и т. д. Поскольку, согласно (2.1) и (2.26), нормальные координаты однозначно связаны с компонентами векторов смещения, выражение (5.4) определяет тем самым и распределение вероятностей тех или иных смещений атомов решетки,

Поскольку гамильтониан (5.2) представляет собой сумму независимых однотипных слагаемых, решение уравнения (5.2) легко найти, полагая

$$\Phi = \prod_f \varphi(x_f). \tag{5.5}$$

При этом для всех сомножителей φ получается уравнение одного и того же вида:

$$\left(\frac{p_f^2}{2M} + \frac{M\omega_f^2 x_f^2}{2}\right)\varphi(x_f) = E_f \varphi(x_f). \tag{5.6}$$

Величина E_f есть собственное значение оператора, стоящего в скобках в левой части (5.6). Собственное значение полной энергии системы дается суммой

$$E = \sum_f E_f. \tag{5.7}$$

Уравнение (5.6), дополненное условием квадратичной интегрируемости функции φ в области $-\infty < x_f < +\infty$, представляет собой стандартную задачу о гармоническом осцилляторе в квантовой механике. Решение этой задачи хорошо известно:

$$\varphi \equiv \varphi_{n_f}(x_f) = \left(\frac{M\omega_f}{\pi\hbar}\right)^{1/4} \frac{1}{\sqrt{2^{n_f} n_f!}} \exp\left(-\frac{M\omega_f}{2\hbar} x_f^2\right) H_{n_f}\left(x_f \sqrt{\frac{M\omega_f}{\hbar}}\right), \tag{5.8}$$

$$E_f = \hbar\omega_f (n_f + 1/2). \tag{5.9}$$

Здесь n_f — целое положительное число (вообще говоря, свое для каждого типа нормальных колебаний), а H_{n_f} — n_f-й полином Эрмита. Функция φ, определяемая формулой (5.8), нормирована обычным способом $\int dx \varphi_{n_f}^2(x) = 1$.

Таким образом, состояние отдельного осциллятора, соответствующего нормальному колебанию ветви s с квазиволновым вектором $\mathbf{q}$, задается квантовым числом $n_f \equiv n(\mathbf{q}, s)$. Совокупность чисел $n(\mathbf{q}, s)$ для всех $\mathbf{q}$ и s полностью определяет состояние колеблющейся решетки. Мы будем иногда обозначать эту совокупность одной буквой n, снабжая соответствующим индексом и волновую функцию Φ.

Вектор смещения $\mathbf{Q}_a$ при квантовомеханическом рассмотрении также становится оператором. Для дальнейшего (гл. XIV) нам понадобятся его матричные элементы в системе функций Φ_n, т. е. интегралы вида

$$\mathcal{J}_{nn'} = \int \Phi_{n'}^* \mathbf{Q}_a \Phi_n \, dx_1 \ldots dx_N. \tag{5.10}$$

Пользуясь выражением (4.3) для вектора смещения, можем переписать (5.10) в виде

$$\mathcal{J}_{nn'} = \frac{1}{2} \sum_{\mathbf{q}} \sum_{s=1}^{r} \{\xi_h(\mathbf{q}, s) \mathcal{J}_{nn'}^{+}(f) e^{i\mathbf{q}\mathbf{a}} + \xi_h^*(\mathbf{q}, s) \mathcal{J}_{nn'}^{-}(f) e^{-i\mathbf{q}\mathbf{a}}\}, \tag{5.11}$$

где

$$\mathcal{J}^{\pm}_{nn'}(f)=\int \Phi^{*}_{n'}\left(x_f \pm i\frac{p_f}{M\omega_f}\right)\Phi_n\,dx_1\ldots dx_N. \tag{5.12}$$

Подставляя сюда функции $\Phi^{*}_{n'}$ и Φ_n в виде (5.5), получаем

$$\mathcal{J}^{\pm}_{nn'}(f)=\int dx_f\varphi_{n'_f}(x_f)\left(x_f \pm i\frac{p_f}{M\omega_f}\right)\varphi_{n_f}(x_f)\times \\ \times\prod_{f'\neq f}\int dx_{f'}\varphi_{n'_{f'}}(x_{f'})\,\varphi_{n_{f'}}(x_{f'}). \tag{5.13}$$

Поскольку функции φ_{n_f}, $\varphi_{n'_f}$ образуют ортогональную систему, интегралы по $x_{f'}$ при $f'\neq f$ отличны от нуля только при $n'_{f'}=n_{f'}$, когда они равны единице по условию нормировки. Интегралы по x_f можно вычислить, пользуясь вытекающими из (5.8) рекуррентными соотношениями ([М2], § 33):

$$\left(x_f+\frac{\hbar}{M\omega_f}\frac{\partial}{\partial x_f}\right)\varphi_{n_f}=\sqrt{2\frac{\hbar n_f}{M\omega_f}}\,\varphi_{n_f-1}, \tag{5.14а}$$

$$\left(x_f-\frac{\hbar}{M\omega_f}\frac{\partial}{\partial x_f}\right)\varphi_{n_f}=\sqrt{\frac{2\hbar}{M\omega_f}(n_f+1)}\,\varphi_{n_f+1}. \tag{5.14б}$$

Поскольку $p_f=-i\hbar\dfrac{\partial}{\partial x_f}$, в левых частях (5.14а) и (5.14б) стоят как раз выражения

$$x_f+i\frac{p_f}{M\omega_f} \quad\text{и}\quad x_f-i\frac{p_f}{M\omega_f}.$$

Пользуясь вновь ортогональностью и нормировкой функций φ_{n_f}, мы получаем

$$\int dx_f\varphi_{n'_f}(x_f)\left(x_f+i\frac{p_f}{M\omega_f}\right)\varphi_{n_f}(x_f)= \\ =\left(\frac{2\hbar}{M\omega_f}\right)^{1/2}\cdot\begin{cases}\sqrt{n_f}, & n'_f=n_f-1,\\ 0, & n'_f\neq n_f-1,\end{cases} \tag{5.15а}$$

$$\int dx_f\varphi_{n'_f}(x_f)\left(x_f-i\frac{p_f}{M\omega_f}\right)\varphi_{n_f}(x_f)= \\ =\left(\frac{2\hbar}{M\omega_f}\right)^{1/2}\cdot\begin{cases}\sqrt{n_f+1}, & n'_f=n_f+1,\\ 0 & n'_f\neq n_f+1.\end{cases} \tag{5.15б}$$

Таким образом, интегралы $\mathcal{J}^{\pm}_{nn'}$ отличны от нуля только при совпадении квантовых чисел всех осцилляторов, кроме f-го; значение n_f при этом должно измениться на единицу. Можно сказать, что каждый член суммы (4.3) (или (4.5), (4.6)) при квантовом рассмотрении представляет собой оператор, изменяющий на единицу квантовое число соответствующего осциллятора и не затрагиваю-

щий всех остальных осцилляторов. Удобно ввести новые операторы b_f и b_f^*, полагая

$$\frac{1}{2}\left(x_f + i\,\frac{p_f}{M\omega_f}\right) = \left(\frac{\hbar}{2M\omega_f}\right)^{1/2} b_f, \tag{5.16а}$$

$$\frac{1}{2}\left(x_f - i\,\frac{p_f}{M\omega_f}\right) = \left(\frac{\hbar}{2M\omega_f}\right)^{1/2} b_f^*. \tag{5.16б}$$

Согласно (5.5), (5.15а) и (5.15б)

$$b_f \Phi_n = \sqrt{n_f}\,\Phi_{n'} \delta_{n_f', n_f - 1, 0} \prod_{f' \neq f} \delta_{n_f', n_f}, \tag{5.17а}$$

$$b_f^* \Phi_n = \sqrt{n_f + 1}\,\Phi_{n'} \delta_{n_f', n_f - 1, 0} \prod_{f' \neq f} \delta_{n_f', n_f}. \tag{5.17б}$$

Символы Кронекера здесь выражают полученные ранее условия, которым, согласно (5.15а) и (5.15б), должны удовлетворять состояния n и n'. При выполнении этих условий матричные элементы операторов b_f и b_f^* равны, соответственно, $\sqrt{n_f}$ и $\sqrt{n_f + 1}$.

Подставляя (5.16а, б) в правую часть (4.5), получаем удобную для дальнейшего формулу для длинноволновой части оператора смещения при акустических колебаниях:

$$\mathbf{Q}_{\text{дл. ак}}(\mathbf{r}) = \sum_{\mathbf{q}} \sum_{s=1}^{3} \left[\frac{\hbar}{2M\omega_s(\mathbf{q})\,G}\right]^{1/2} \{\xi(\mathbf{q}, s)\, b_{\mathbf{q}, s} e^{i\mathbf{qr}} + \xi^*(\mathbf{q}, s)\, b_{\mathbf{q}, s}^* e^{-i\mathbf{qr}}\}. \tag{5.18}$$

Аналогично преобразуется и формула (4.6):

$$\mathbf{Q}_{\text{дл. опт}}(\mathbf{r}) = \sum_{\mathbf{q}} \sum_{s \geqslant 4} \left[\frac{\hbar}{2M\omega_s(\mathbf{q})\,G}\right]^{1/2} \{\xi_h(\mathbf{q}, s)\, b_{\mathbf{q}, s} e^{i\mathbf{qr}} + \xi_h^*(\mathbf{q}, s)\, b_{\mathbf{q}, s}^* e^{-i\mathbf{qr}}\}. \tag{5.19}$$

Этими формулами мы воспользуемся дальше, в гл. XIV.

§ 6. Фононы

Согласно (5.9) и (5.7) собственные значения энергии колеблющейся решетки можно записать в виде

$$E = \sum_f \hbar\omega_f (n_f + 1/2) = \tag{6.1}$$

$$= \frac{1}{2}\sum_f \hbar\omega_f + \sum_f n_f \hbar\omega_f. \tag{6.2}$$

Первое слагаемое в (6.2) представляет собой нулевую энергию системы осцилляторов, не исчезающую даже в основном ее состоя-

нии. Коль скоро частоты нормальных колебаний заданы, нулевая энергия есть просто некоторая постоянная. Второе слагаемое в (6.2) описывает энергию возбуждения системы. Различным возбужденным состояниям при этом соответствует различный набор чисел n_f, причем энергия возбуждения увеличивается с возрастанием значений n_f. Формально выражение для энергии возбуждения имеет вид полной энергии идеального газа, состояния «частиц» которого задаются индексом f, т. е. квазиволновым вектором $\mathbf{q}$ и номером ветви s. При этом число «частиц» в состоянии f есть квантовое число соответствующего осциллятора n_f, а энергия «частицы» равна $\hbar\omega_f$. Она, как и следовало ожидать, связана с частотой соответствующего нормального колебания соотношением де Бройля. Представление о таком идеальном газе очень удобно: оно позволяет наиболее простым и наглядным путем выразить то обстоятельство, что энергия колеблющейся решетки может изменяться не «как попало», а только порциями — квантами — величины $\hbar\omega_f$. «Частицы» этого газа — кванты энергии возбуждения — называются *фононами**).

Полное число фононов в кристалле, конечно, не сохраняется: оно дается суммой всех чисел n_f и может быть любым. Число фононов в каждом состоянии n_f также произвольно: как мы видели, число n_f может принимать значения 0, 1, 2, ... Это означает, что газ фононов подчиняется статистике Бозе—Эйнштейна.

Таким образом, мы имеем два эквивалентных языка для описания малых колебаний атомов кристаллической решетки около положений равновесия — язык гармонических осцилляторов и язык фононов. В разных задачах бывает удобно пользоваться либо одним, либо другим из них, и полезно иметь в виду «правила перевода» с одного языка на другой. Составим таблицу (см. табл. 12.1 на стр. 396), указывая в середине то или иное физическое понятие (или процесс), а слева и справа (строкой ниже) — описание его на языках осцилляторов и фононов.

При использовании фононного языка очень ясный и наглядный смысл приобретают операторы b_f^* и b_f (5.16а, б). Первый из них увеличивает, а второй уменьшает на единицу число фононов s-й ветви с квазиволновым вектором $\mathbf{q}$. По этой причине b_f^* и b_f называются, соответственно, операторами порождения и уничтожения фононов.

Рассмотренные только что «правила перевода» в равной мере относятся как к равновесным, так и к неравновесным состояниям решетки. В частности, в условиях термодинамического равновесия представление о малых колебаниях составляет точное содержание понятия «тепловое движение решетки». Последнее можно, таким образом, описывать на языке фононов. Нагревание или охлаждение

---

*) Представление о фононах впервые было введено в 1930 г. И. Е. Таммом [3].

Таблица 12.1

Язык осцилляторов		Язык фононов
	Решетка атомов или ионов, совершающих малые колебания около положений равновесия	
Набор независимых осцилляторов, частоты которых суть частоты нормальных колебаний. Осцилляторы нумеруются вектором $\mathbf{q}$ и числом s		Идеальный газ фононов, энергии которых связаны с частотами нормальных колебаний соотношением де Бройля. Состояние фонона задается вектором $\mathbf{q}$ и числом s
	Величины, задающие состояние решетки	
Квантовые числа всех осцилляторов n_f		Числа фононов во всех состояниях n_f
	Энергия решетки (без нулевой энергии)	
Сумма энергий независимых осцилляторов		Энергия идеального газа фононов
	Изменение состояния решетки	
Изменение квантовых чисел тех или иных осцилляторов		Изменение чисел фононов в тех или иных состояниях
	Возбуждение решетки	
Перевод одного или нескольких осцилляторов в состояние с бо́льшим квантовым числом		Порождение одного или нескольких фононов
	Уменьшение энергии решетки	
Перевод одного или нескольких осцилляторов в состояние с меньшим квантовым числом		Исчезновение одного или нескольких фононов

решетки означает при этом перераспределение фононов по состояниям, а также увеличение или уменьшение общего их числа. В частности, среднее число фононов в состоянии $f=\{\mathbf{q}, s\}$ в условиях термодинамического равновесия дается функцией распределения Бозе—Эйнштейна. При этом следует иметь в виду, что химический потенциал фононного газа равен нулю. Действительно, поскольку общее число n фононов не постоянно, оно определяется (при термодинамическом равновесии) просто из условия минимальности свободной энергии $\mathscr{F}$: $\left(\frac{\partial \mathscr{F}}{\partial n}\right)_{T, V}=0$. Но $\left(\frac{\partial \mathscr{F}}{\partial n}\right)_{T, V}$ есть химический потенциал [М9]. Таким образом, распределение Бозе—Эйнштейна превращается в данном случае в формулу Планка

$$n_f=\left[\exp\left(\frac{\hbar\omega_f}{kT}\right)-1\right]^{-1}. \tag{6.3}$$

Соотношения (6.2) и (6.3) в сочетании с законом дисперсии — зависимостью ω от $\mathbf{q}$ — позволяют вычислять среднюю энергию колеблющейся решетки и, следовательно, все остальные термодинамические характеристики по стандартным формулам теории идеального газа Бозе—Эйнштейна.

В связи с описанием решетки на языке фононного газа следует сделать еще три замечания.

Во-первых, аналогия между фононами и частицами идеального газа основывалась

а) на формуле (6.2) и

б) на соотношении де Бройля между энергией фонона и частотой нормальных колебаний.

Чтобы эта аналогия была последовательной, надо иметь возможность приписать фонону не только энергию, но и квазиимпульс $\mathbf{p}$. При этом квазиимпульс должен быть связан с квазиволновым вектором фонона соотношением де Бройля $\mathbf{p}=\hbar\mathbf{q}$. Рассматривая только идеальный фононный газ, мы не можем решить вопрос о существовании или отсутствии квазиимпульса у фонона. Действительно, квазиимпульс определяется по свойству сохраняться (с точностью до $\hbar\mathbf{b}$) либо в периодическом поле, либо (в сумме по всем частицам) при взаимном их рассеянии. Но в идеальном газе фононов никакие процессы рассеяния не происходят. Поэтому ввести квазиимпульс фонона можно только рассматривая взаимодействие его с электронами или с дырками — объектами, квазиимпульс которых уже определен. Эта задача решается в § XIV.4. Там показано, что фонону действительно следует приписать квазиимпульс $\mathbf{p}=\hbar\mathbf{q}$.

Во-вторых, аналогию между частицами идеального газа и фононами можно провести все же не до конца. Именно, частицы — атомы или молекулы газа — представляют собой в известном смысле

«самостоятельные» образования. Можно, например, поставить вопрос о физическом выделении какого-то их числа из всей системы. Но представление о фононах — это лишь язык для описания нормальных колебаний решетки. Последние представляют собой коллективные движения всех атомов кристалла. Соответственно и фононы характеризуют коллективные свойства системы атомов или ионов, образующих решетку, и существуют лишь постольку, поскольку существует сама эта система. Иначе говоря, фононы ведут себя действительно как частицы — но лишь до тех пор, пока существует физическая система, коллективное движение частиц которой они изображают. По этой причине фононы (как и дырки, § IV.2) называют квазичастицами.

Наконец, в-третьих, представление об идеальном газе фононов возникло у нас при рассмотрении формулы (6.2). Последняя, однако, справедлива лишь в гармоническом приближении: при выводе ее предполагалось, что потенциальную энергию решетки V можно аппроксимировать квадратичной формой по смещениям атомов из положений равновесия $\mathbf{Q}_a$. При учете ангармонизма, т. е. следующих членов разложения V в ряд по $\mathbf{Q}_a$, оператор энергии уже не имеет простого вида (2.20): появляются ангармонические члены с третьей и более высокими степенями величин x_f. Их можно интерпретировать как энергию взаимодействия между фононами. Таким образом, при учете ангармонизма фононный газ становится неидеальным — оказываются возможными, например, процессы рассеяния фононов друг на друге и т. д. Очевидно, роль ангармонических членов тем меньше, чем меньше амплитуда колебаний атомов по сравнению с постоянной решетки. На «фононном» языке это значит, что газ фононов можно считать идеальным лишь пока самих фононов достаточно мало, т. е. пока решетка слабо возбуждена.

Представление о фононах не только полезно и удобно для решения различных конкретных задач. Оно имеет и большое принципиальное значение. Именно, мы начали с рассмотрения системы сильно взаимодействующих друг с другом частиц — атомов, ионов или молекул, образующих кристаллическую решетку. В результате взаимодействия между ними энергия решетки отнюдь не равна сумме энергий отдельных атомов. Выяснилось, однако, что слабо возбужденные состояния рассматриваемой системы можно представить как состояния идеального газа некоторых квазичастиц — фононов, — не имеющих ничего общего с отдельными атомами решетки *). Как мы увидим в дальнейшем (§ XVII, 6), такая ситуация не представляет собой чего-то исключительного, а оказывается типичной для всех систем многих взаимодействующих частиц.

*) Достаточно напомнить, например, что в любом кристалле фононы подчиняются статистике Бозе—Эйнштейна — независимо от статистики, которой подчиняются сами частицы, составляющие кристалл.

Глава XIII

ЭЛЕМЕНТЫ КИНЕТИЧЕСКОЙ ТЕОРИИ ЯВЛЕНИЙ ПЕРЕНОСА

§ 1. Феноменологические соотношения

Задача кинетической теории явлений переноса состоит в вычислении величин, определяющих реакцию системы на те или иные внешние воздействия. К числу их относятся электропроводность, постоянная Холла, термоэдс и т. д.

В настоящей главе мы будем интересоваться поведением вещества в достаточно слабых электрических полях $\mathcal{E}$ или при достаточно малых градиентах температуры ∇T, когда связь их с плотностью тока $\mathbf{j}$ или с потоком тепла оказывается линейной. Точный смысл выражений «достаточно слабое» и «достаточно малый» выяснится в § 7.

Удобно рассмотреть сначала три частных случая, отвечающих поведению носителей заряда в слабых электрическом, температурном и в электрическом и магнитном полях.

а. Носители заряда в постоянном и однородном слабом электрическом поле. Рассмотрим сначала технологически однородный образец, в котором концентрации атомов примеси, электронов и дырок в условиях равновесия не зависят от координат.

Пусть, далее, магнитное поле отсутствует и температура всюду постоянна. Наиболее общее линейное соотношение, связывающее два вектора $\mathbf{j}$ и $\mathcal{E}$, имеет вид обобщенного закона Ома (I.1.5). В отсутствие магнитного поля тензор электропроводности оказывается симметричным: $\sigma_{\alpha\beta} = \sigma_{\beta\alpha}$ (доказательство этого утверждения можно найти в книге [M13]).

Удобно выбрать систему координат, связанную с главными осями тензора $\sigma_{\alpha\beta}$. При этом $\sigma_{\alpha\beta} = 0$ при $\alpha \neq \beta$; σ_{xx}, σ_{yy} и σ_{zz} суть главные значения тензора электропроводности. В кристалле кубической системы $\sigma_{xx} = \sigma_{yy} = \sigma_{zz} = \sigma$, и равенство (I.1.5) принимает вид (I.1.3).

К числу материалов, для которых справедливо равенство (I.1.3) принадлежат многие важные полупроводники: германий, кремний, соединения типа $A^{III}B^{V}$ и т. д. Впредь мы ограничимся рассмотрением только таких систем.

Из сказанного ясен способ вычисления компонент тензора электропроводности. Надо, пользуясь методами статистической физики, явно вычислить плотность тока в рассматриваемых условиях

опыта. В достаточно слабом электрическом поле результат будет иметь вид (I.1.5) — или, в соответствующих случаях, (I.1.3), — но с явно вычисленными коэффициентами пропорциональности между j_α и $\mathcal{E}_\beta$. Эти коэффициенты и следует отождествить с величинами $\sigma_{\alpha\beta}$.

Формула (I.1.3) легко обобщается и на случай неоднородного материала. Надо лишь принять во внимание, что, как указывалось в § VI.3, плотность электрического тока в изотермических условиях определяется градиентом электрохимического потенциала F *). Это означает, что вместо вектора $\mathcal{E} = -\nabla\varphi$ в формуле (I.1.3) и др. теперь будет фигурировать величина

$$\mathcal{E}' = \mathcal{E} + \mathcal{E}^* = \frac{1}{e}\nabla F. \tag{1.1}$$

Здесь $\mathcal{E}$ — напряженность кулоновского поля, $\mathcal{E}^*$ — напряженность поля сторонних сил.

Таким образом, в $\mathcal{E}'$ включаются и поля неэлектростатического происхождения, связанные с разностью давлений электронного газа в разных точках образца (при постоянной температуре). В пространственно однородной среде в изотермических условиях $\mathcal{E}^* = 0$ и $\mathcal{E}'$ переходит в $\mathcal{E}$.

б. Носители заряда в постоянном и однородном слабом температурном поле. Пусть внешние магнитное и электрическое поля отсутствуют, но между торцами образца поддерживается постоянная во времени разность температур (рис. 1.3, стр. 18). При этом в образце установится постоянный градиент температуры ∇T; мы будем считать, что в пространственно однородном образце он не зависит и от координат. Если замкнуть концы образца, то в цепи потечет ток.

При достаточно малом градиенте температуры векторы $\mathbf{j}$, $\mathcal{E}'$ и ∇T линейно связаны друг с другом. В случае кубических кристаллов (или изотропных сред) мы имеем, подобно (I.1.3),

$$\mathbf{j} = a\cdot\nabla T + \sigma\mathcal{E}'. \tag{1.2}$$

Здесь a — скалярный коэффициент, не зависящий ни от ∇T, ни от $\mathcal{E}'$. При разомкнутых концах образца, когда $\mathbf{j} = 0$, мы получаем из (1.2)

$$\mathcal{E}' = -\frac{a}{\sigma}\nabla T. \tag{1.3}$$

Это есть не что иное, как напряженность стороннего поля, вызываемого градиентом температуры. Термоэлектродвижущая сила дается линейным интегралом от $\mathcal{E}'$ по длине образца:

$$V_0 = -\int \frac{a}{\sigma}(\nabla T\cdot d\mathbf{l}). \tag{1.4}$$

*) Это утверждение, справедливое в условиях применимости закона Ома, будет обосновано в § 5.

Здесь $d\mathbf{l}$ — элемент длины, который в данном случае удобно направить вдоль вектора ∇T. Таким образом,

$$V_0 = -\int_{T_1}^{T_2} \frac{a}{\sigma}\, dT. \tag{1.5}$$

Дифференциальную термоэдс определяют как производную от V_0 по температуре горячего спая T_2 (при постоянном значении a/σ это совпадает с элементарным определением величины α как изменения термоэдс при изменении разности температур $(T_2 - T_1)$ на один градус).

Таким образом, равенства (1.1), (1.2) и (1.5) дают

$$\alpha = -a/\sigma = \frac{|\nabla F|}{e\,|\nabla T|}. \tag{1.6}$$

В образцах с биполярной проводимостью дифференциальная термоэдс дается суммой электронного и дырочного вкладов.

В § I.1. мы уже говорили, что полная термоэдс, т. е. разность потенциалов между металлическими электродами на концах полупроводника, складывается из падения напряжения внутри полупроводника и скачков потенциала в контактах полупроводник—металл. Эти скачки возникают потому, что в контактных областях тоже действуют сторонние силы (силы давления электронного газа). Они зависят от температуры, в силу чего при наличии градиента температуры сумма обоих скачков потенциала отлична от нуля. Поскольку ∇F учитывает и эти сторонние силы, формула (1.6) определяет термоэдс, измеряемую на опыте.

Выражая коэффициент a через α и σ, можем переписать соотношение (1.2) в виде

$$\mathbf{j} = \sigma\mathcal{E}' - \alpha\sigma \cdot \nabla T. \tag{1.7}$$

Сопоставляя равенства (1.7) и (1.2), получаем способ вычисления дифференциальной термоэдс: надо, пользуясь методами статистической физики, явно вычислить плотность тока в рассматриваемых условиях опыта. При достаточно малом градиенте температуры результат будет иметь вид (1.2), но с явно вычисленным коэффициентом пропорциональности при векторе ∇T. Этот коэффициент следует отождествить с феноменологически введенной величиной a и воспользоваться формулой (1.6).

Наличие градиента температуры и (или) электрического поля в образце приводит к возникновению не только потока заряда, но и потока энергии. Плотность последнего $\mathbf{I}$, будучи векторной величиной, может выражаться только через независимые векторы, характеризующие рассматриваемую систему. Поскольку три вектора, $\mathbf{j}$, ∇T и $\mathcal{E}'$, связаны соотношением (1.7), независимы из них только два. Выберем в качестве таковых $\mathbf{j}$ и ∇T. Тогда в изотропном

материале (или в кубическом кристалле) плотность потока энергии электронов будет иметь вид

$$\mathbf{I} = -\varkappa \cdot \nabla T + \Pi \mathbf{j} - \frac{F}{e}\mathbf{j}, \tag{1.8}$$

где $\varkappa$ и Π — некоторые скаляры. Формально второе слагаемое в (1.8) можно было бы объединить с третьим. Однако удобно рассматривать их порознь. Дело в том, что, как мы вскоре увидим, второе слагаемое непосредственно связано с выделением тепла Пельтье, а третье вклада в это тепло не дает.

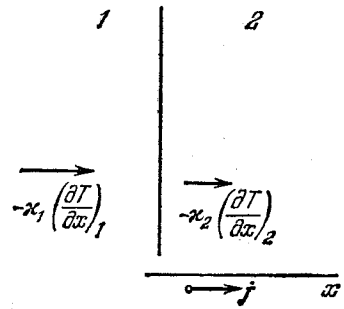


Рис. 13.1. К выводу формулы для тепла Пельтье. Вертикальная линия изображает сечение поверхности раздела плоскостью чертежа. Для определенности градиент температуры принят антипараллельным вектору плотности тока.

Чтобы выяснить смысл величин $\varkappa$ и Π, положим сначала $\mathbf{j} = 0$. При этом возникают стандартные условия для измерения теплопроводности: в электрически изолированном образце создан постоянный градиент температуры и течет поток тепла. Таким образом, в соответствии с § I.1, величина $\varkappa$ есть не что иное, как коэффициент теплопроводности электронного газа.

Далее, рассмотрим контакт двух полупроводников *1* и *2*. Пусть поверхность их раздела есть плоскость YZ, перпендикулярная векторам $\mathbf{j}$ и ∇T (рис. 13.1); вектор $\mathbf{j}$ будем считать направленным вдоль оси X. Вычислим количество тепла Q, выделяющееся или поглощающееся в единицу времени на единице площади данной поверхности. Для этой цели напишем равенство (1.8) дважды — для первого и второго проводников — и воспользуемся условиями непрерывности нормальной (к поверхности раздела) составляющей $\mathbf{j}$, а также условием непрерывности электрохимического потенциала F. Обозначая индексами «1» и «2» значения соответствующих величин на поверхности раздела, получим

$$-\varkappa_1\left(\frac{\partial T}{\partial x}\right)_1 + \varkappa_2\left(\frac{\partial T}{\partial x}\right)_2 = -(\Pi_1 - \Pi_2)\, j. \tag{1.9}$$

В левой части (1.9) стоит суммарная плотность потока тепла, подводимого к поверхности за счет теплопроводности. Поскольку мы рассматриваем стационарные условия, этот поток должен компенсироваться отводом тепла от контакта. Следовательно,

$$Q = \Pi_{12} j = (\Pi_1 - \Pi_2)\, j. \tag{1.10}$$

Сравнивая это выражение с формулой (I.1.16), видим, что соотношение (I.1.17) действительно оправдывается.

Величины Π_1 и Π_2 называются коэффициентами Пельтье (соответственно первого и второго проводника).

Легко связать величины Π и α с коэффициентом Томсона α_T, определяемым равенством (I.1.13). Действительно, по определению дивергенции вектора $\mathbf{I}$ полное количество тепла, выделяющееся в единице объема образца в единицу времени, есть $-\operatorname{div}\mathbf{I}$*). В силу уравнения непрерывности в стационарных условиях $\operatorname{div}\mathbf{j}=0$. Принимая во внимание это обстоятельство, мы получаем из (1.1) и (1.8)

$$-\operatorname{div}\mathbf{I}=\operatorname{div}(\varkappa\cdot\nabla T)-(\mathbf{j},\ \nabla\Pi)+(\mathbf{j},\ \mathcal{E}'). \tag{1.11}$$

Согласно (1.7)

$$\mathcal{E}'=\frac{\mathbf{j}}{\sigma}+\alpha\cdot\nabla T. \tag{1.12}$$

Подставляя (1.12) в формулу (1.11), находим

$$-\operatorname{div}\mathbf{I}=\operatorname{div}(\varkappa\cdot\nabla T)+\frac{j^2}{\sigma}-\left(\frac{d\Pi}{dT}-\alpha\right)(\mathbf{j},\ \nabla T). \tag{1.11'}$$

Первое слагаемое в правой части (1.11') описывает выделение тепла за счет теплопроводности, второе — джоулево тепло, а третье — эффект Томсона. Коэффициент Томсона дается выражением

$$\alpha_T=\frac{d\Pi}{dT}-\alpha. \tag{1.13}$$

Формула (I.1.19) получается отсюда при учете (I.1.18). Последнее соотношение получается более сложным путем (см. § 7).

Из сказанного ясен способ вычисления коэффициента теплопроводности и коэффициента Пельтье: надо, пользуясь методами статистической физики, явно вычислить плотность потока энергии в рассматриваемых нами условиях опыта. При достаточно малых градиентах температуры и электрохимического потенциала результат будет иметь вид (1.8), но с явно вычисленными коэффициентами пропорциональности в первых двух слагаемых. Эти коэффициенты надлежит отождествить с феноменологически введенными величинами $\varkappa$ и Π.

в. Носители заряда в постоянных и однородных электрическом и магнитном полях. Пусть образец, к которому приложено постоянное напряжение, помещен в постоянное и однородное магнитное поле. Температуру образца будем считать всюду постоянной. Электрическое поле в образце по-прежнему будем считать слабым в указанном выше смысле.

Чтобы написать для этого случая феноменологическую формулу типа (I.1.5а), надо принять во внимание, что мы имеем теперь

*) В стационарных условиях эта величина обращается в нуль. Нам, однако, важно лишь узнать, из чего она складывается.

независимый вектор $\mathcal{E}$ и независимый псевдовектор $\mathcal{B}$*). В изотропной среде из них можно образовать только следующие векторные комбинации:

$$\mathcal{E}, \quad [\mathcal{E}\times\mathcal{B}], \quad (\mathcal{B}, \mathcal{E})\mathcal{B}. \tag{1.14}$$

При этом выражение для плотности тока имеет вид

$$\mathbf{j} = a_1\mathcal{E} + a_2[\mathcal{E}\times\mathcal{B}] + a_3(\mathcal{B}, \mathcal{E})\mathcal{B}, \tag{1.15}$$

где a_1, a_2, a_3 — некоторые скаляры. Поскольку магнитное поле не предполагается непременно слабым, они могут зависеть от $\mathcal{B}^2$. Очевидно, формулу (1.15) можно переписать в виде (1.1.5а):

$$j_\alpha = \sigma_{\alpha\beta}(\mathcal{B})\mathcal{E}_\beta. \tag{1.16}$$

Направим ось Z вдоль магнитного поля; тогда, согласно (1.15) и (1.16), компоненты тензора $\sigma_{\alpha\beta}(\mathcal{B})$ будут иметь вид

$$\begin{aligned} &\sigma_{xx} = \sigma_{yy} = a_1, && \sigma_{zz} = a_1 + a_3\mathcal{B}^2, \\ &\sigma_{xy} = -\sigma_{yx} = a_2\mathcal{B}, && \sigma_{xz} = \sigma_{zx} = \sigma_{yz} = \sigma_{zy} = 0. \end{aligned} \tag{1.17}$$

Видим, что при наличии магнитного поля тензор электропроводности имеет как симметричную, так и антисимметричную части $\sigma^s_{\alpha\beta}$ и $\sigma^a_{\alpha\beta}$. Компоненты их суть, соответственно, четные и нечетные функции магнитной индукции. Так,

$$\sigma^a_{xy} = -\sigma^a_{yx} = a_2\mathcal{B}, \quad \sigma^a_{xz} = \sigma^a_{yz} = 0. \tag{1.18}$$

Такая зависимость от $\mathcal{B}$ не случайна. Она вытекает как частный случай из общего принципа симметрии кинетических коэффициентов [М13].

В анизотропных средах — кристаллах — формула (1.16) остается в силе, но набор независимых векторов уже не исчерпывается величинами (1.14). Дело в том, что здесь векторы следует определять по отношению к преобразованиям системы координат, допускаемым симметрией решетки — более низкой, нежели симметрия изотропного пространства. Поясним это на примере слабого магнитного поля, ограничиваясь кристаллами кубической системы. Разлагая тензор $\sigma_{\alpha\beta}(\mathcal{B})$ в ряд по степеням $\mathcal{B}$ и ограничиваясь членами не выше второго порядка, мы получаем

$$j_\alpha = \sigma_{\alpha\beta}\mathcal{E}_\beta + \eta[\mathcal{E}\times\mathcal{B}]_\alpha + \sigma_{\alpha\beta\gamma\delta}\mathcal{E}_\beta\mathcal{B}_\gamma\mathcal{B}_\delta + \ldots \tag{1.16'}$$

*) Напомним, что при поворотах системы координат псевдовектор ведет себя как обычный вектор. Однако, в отличие от компонент вектора, компоненты псевдовектора не меняют знака при замене направлений координатных осей на противоположные (эта операция называется отражением). Скалярное произведение $(\mathcal{B}, \mathcal{E})$ есть псевдоскаляр: в отличие от истинного скаляра, оно меняет знак при отражении. Именно по этой причине оно и введено в последнее из выражений (1.14): произведение псевдоскаляра на псевдовектор есть обычный вектор.

Через $\sigma_{\alpha\beta} = \sigma_{\beta\alpha}$ здесь обозначен тензор $\sigma_{\alpha\beta}(0)$, η — постоянная. Можно считать, что $\sigma_{\alpha\beta\gamma\delta} = \sigma_{\alpha\beta\delta\gamma}$. Действительно, выражение $\mathscr{B}_\gamma \mathscr{B}_\delta$, очевидно, симметрично по индексам γ, δ. Следовательно, антисимметричная по γ и δ часть тензора $\sigma_{\alpha\beta\gamma\delta}$, если бы она и была отлична от нуля, все равно не дала бы вклада в (1.16′).

Число независимых компонент тензора $\sigma_{\alpha\beta\gamma\delta}$ в кубическом кристалле равно трем [М7]. В качестве таковых удобно взять величины

$$\sigma_{\lambda\lambda\lambda\lambda} \equiv \gamma, \quad \sigma_{\lambda\lambda\mu\mu} \equiv \delta, \quad \sigma_{\lambda\mu\lambda\mu} \equiv \beta$$

(без суммирования по λ и μ).

Таким образом, формула (1.16′) принимает вид

$$\mathbf{j} = \sigma\mathscr{E} + \eta[\mathscr{E}\times\mathscr{B}] + \beta\mathscr{B}(\mathscr{B}, \mathscr{E}) + \delta\mathscr{B}^2\mathscr{E} + \gamma\mathscr{K}. \qquad (1.15')$$

Здесь через $\mathscr{K}$ обозначен вектор с компонентами

$$\mathscr{E}_x\mathscr{B}_x^2, \quad \mathscr{E}_y\mathscr{B}_y^2, \quad \mathscr{E}_z\mathscr{B}_z^2.$$

Появление последнего слагаемого в (1.15′) характерно для анизотропной среды. В изотропном случае, когда физически эквивалентны любые три взаимно перпендикулярные оси, коэффициент γ неизбежно обратился бы в нуль. И действительно, в слабом магнитном поле равенство (1.15) принимает вид (1.15′) без последнего слагаемого: надо лишь обозначить предельные (при $\mathscr{B} \to 0$) значения a_1, a_2 и a_3 через σ, η и β соответственно. Подставляя выражения (1.17) в формулы (I.1.9), находим для постоянной Холла

$$R = \frac{a_2}{a_1^2 + a_2^2\mathscr{B}^2}. \qquad (1.19)$$

Далее, пользуясь формулами (I.1.11) и (1.17), получаем для поперечного магнетосопротивления

$$\frac{\rho_\perp - \rho_0}{\rho_0} = \frac{a_1(1-\rho_0 a_1) - \rho_0 a_2^2\mathscr{B}^2}{\rho_0(a_1^2 + a_2^2\mathscr{B}^2)}. \qquad (1.20)$$

Наконец, аналогичным путем можно вычислить и продольное магнетосопротивление. Мы получаем

$$\frac{\rho_\parallel - \rho_0}{\rho_0} = \frac{1 - \rho_0 a_1 - \rho_0 a_3\mathscr{B}^2}{\rho_0(a_1 + a_3\mathscr{B}^2)}. \qquad (1.21)$$

Аналогично, в случае (1.15′) находим

$$R = \eta/\sigma^2, \qquad (1.22)$$

$$\frac{\rho_\perp - \rho_0}{\rho_0} = -\left(\frac{\delta}{\sigma} + \frac{\eta^2}{\sigma^2}\right)\mathscr{B}^2, \qquad (1.23)$$

$$\frac{\rho_\parallel - \rho_0}{\rho_0} = -\frac{\delta+\beta+\gamma}{\sigma}\mathscr{B}^2. \qquad (1.24)$$

Здесь $\sigma = \rho_0^{-1}$ есть электропроводность при $\mathscr{B} = 0$.

Из формул (1.19)—(1.24) ясен способ вычисления постоянной Холла и магнетосопротивления: надо, пользуясь методами статистической физики, вычислить плотность тока в рассматриваемых условиях опыта. В достаточно слабом электрическом поле результат будет иметь вид (1.15) или (1.15′), но с явно вычисленными коэффициентами пропорциональности между векторами $\mathfrak{E}$, $(\mathfrak{B}, \mathfrak{E})\,\mathfrak{B}$, $[\mathfrak{E} \times \mathfrak{B}]$ и $\mathscr{K}$, с одной стороны, и вектором $\mathbf{j}$ — с другой. Эти коэффициенты следует отождествить с феноменологически введенными величинами a_1, a_2, a_3 или $\sigma + \delta\mathscr{B}^2$, η, β, γ.

Подобно (1.16), при наличии как градиента температуры, так и магнитного поля выражения для плотности тока и плотности потока энергии электронов (1.7) и (1.8) следует писать в виде, характерном для анизотропной среды:

$$j_\alpha = \frac{1}{e}\,\sigma_{\alpha\beta}(\mathfrak{B})\,\frac{\partial F}{\partial x_\beta} - \sigma_{\alpha\beta}(\mathfrak{B})\,\alpha_{\beta\gamma}\,\frac{\partial T}{\partial x_\gamma}, \tag{1.25}$$

$$I_\alpha = -\varkappa_{\alpha\beta}(\mathfrak{B})\,\frac{\partial T}{\partial x_\beta} + \Pi_{\alpha\beta}(\mathfrak{B})\,j_\beta - \frac{F}{e}\,j_\alpha. \tag{1.26}$$

При этом аналог соотношения (I.1.18) следует писать в виде

$$\Pi_{\alpha\beta} = T\alpha_{\alpha\beta};$$

тензор $\alpha_{\alpha\beta}$, вообще говоря, не симметричен.

Таким путем вводятся все феноменологические коэффициенты, описывающие явления переноса заряда и энергии при достаточно слабых электрических полях и малых градиентах температуры. Способ явного их вычисления ясен из предыдущего.

§ 2. Кинетические коэффициенты и функция распределения

В теории явлений переноса нас интересует, как правило, поведение газа носителей заряда в условиях, когда электрохимический потенциал его сравнительно медленно изменяется в пространстве: изменение F на расстоянии порядка постоянной решетки очень мало. Согласно результатам § IV.4 носители заряда при этом можно рассматривать как частицы с законом дисперсии $E_l(\mathbf{p})$ и скоростью $\mathbf{v}_l = \nabla_{\mathbf{p}} E_l$. Если, сверх того, электрохимический потенциал медленно меняется и на расстояниях порядка характерной длины волны носителя заряда, то систему свободных электронов (дырок) можно рассматривать просто как газ классически движущихся частиц, «упрятав» все квантовые эффекты в закон дисперсии $E_l(\mathbf{p})$. Главную роль при этом играет представление о функции распределения носителей заряда по квазиимпульсам и координатам $f(\mathbf{p}, \mathbf{r})$. В применении к пространственно однородной системе в условиях равновесия мы уже пользовались этим представлением в гл. V. По определению средняя концентрация электронов проводимости в

элементе объема p-пространства $dp_x dp_y dp_z \equiv d\mathbf{p}$, содержащем точку $\mathbf{p}$, дается выражением

$$\frac{2}{(2\pi\hbar)^3} f(\mathbf{p}, \mathbf{r})\, d\mathbf{p}. \tag{2.1}$$

В силу (2.1) выражения для плотности тока и плотности потока энергии электронов можно записать в виде (опускаем индекс i):

$$\mathbf{j} = -\frac{2e}{(2\pi\hbar)^3} \int \mathbf{v}(\mathbf{p}) f(\mathbf{p}, \mathbf{r})\, d\mathbf{p}, \tag{2.2}$$

$$\mathbf{I} = \frac{2}{(2\pi\hbar)^3} \int \mathbf{v}(\mathbf{p}) [E(\mathbf{p}) - e\varphi] f(\mathbf{p}, \mathbf{r})\, d\mathbf{p}. \tag{2.3}$$

Интегралы здесь берутся, строго говоря, по зоне Бриллюэна. Однако, в силу быстрого убывания функции распределения с увеличением энергии электрона, часто бывает возможно распространить их на все бесконечное p-пространство.

В условиях термодинамического равновесия правые части (2.2) и (2.3) обращаются в нуль. Формально это видно из того, что в указанных условиях $f(\mathbf{p}) = f_0(E(\mathbf{p}))$. Функция f_0 имеет вид (V.3.1) и зависит только от энергии. Поскольку энергия есть четная функция квазиимпульса, а скорость $\mathbf{v}(\mathbf{p})$ — нечетная, интегралы от произведений $f_0\mathbf{v}$ обращаются в нуль.

Факт нарушения термодинамического равновесия проявляется в изменении вида функции распределения. При этом она оказывается зависящей, вообще говоря, не только от энергии, но и от направления вектора $\mathbf{p}$.

Таким образом, вычисление плотности тока и плотности потока энергии сводится к нахождению неравновесной функции распределения $f(\mathbf{p}, \mathbf{r})$ с последующим вычислением интегралов в правых частях (2.2) и (2.3). В зависимости от того, каким именно путем нарушено термодинамическое равновесие в системе носителей заряда, функция распределения их может иметь различный вид. По этой причине естественно искать не универсальное выражение для самой функции $f(\mathbf{p}, \mathbf{r})$, а универсальное уравнение, из которого она определяется и которое отражает как условия опыта, так и свойства материала. В рамках классической кинетической теории газов такая задача была поставлена и решена Л. Больцманом; в дальнейшем его метод был обобщен А. Зоммерфельдом на предмет учета возможного вырождения электронного газа. Уравнение, полученное Больцманом, называется кинетическим. Метод расчета кинетических коэффициентов, основанный на его использовании, принадлежит к числу наиболее популярных и эффективных средств теории явлений переноса (условия применимости его обсуждаются в § XIV.2).

§ 3. Кинетическое уравнение

Кинетическое уравнение представляет собой не что иное, как уравнение непрерывности в фазовом пространстве, координаты которого суть компоненты векторов $\mathbf{p}$ и $\mathbf{r}$. Чтобы написать это уравнение, выберем в фазовом пространстве некоторый фиксированный малый элемент объема, например параллелепипед со сторонами dx, dy, dz, dp_x, dp_y, dp_z, и посмотрим, как изменяется число частиц в нем за бесконечно малый интервал времени dt. Будем при этом опускать слагаемые, содержащие высшие (начиная со второй) степени dt, dx, ..., dp_z. По определению функции f в момент времени t число частиц (с данной проекцией спина) в параллелепипеде есть

$$dn = f(\mathbf{p}, \mathbf{r}, t)\,\frac{d\mathbf{p}\,d\mathbf{r}}{(2\pi\hbar)^3}.$$

Спустя время dt это число будет

$$dn' = f(\mathbf{p}, \mathbf{r}, t+dt)\,\frac{d\mathbf{p}\,d\mathbf{r}}{(2\pi\hbar)^3} = \left\{f(\mathbf{p}, \mathbf{r}, t) + \frac{\partial f(\mathbf{p}, \mathbf{r}, t)}{\partial t}\,dt\right\}\frac{d\mathbf{p}\,d\mathbf{r}}{(2\pi\hbar)^3}.$$

Разность

$$dn' - dn = \frac{\partial f}{\partial t}\,\frac{d\mathbf{p}\,d\mathbf{r}\,dt}{(2\pi\hbar)^3} \tag{3.1}$$

дает результирующий «приход» частиц с данной проекцией спина в данный элемент фазового пространства. Чтобы получить интересующее нас уравнение, надо независимо выразить левую часть (3.1) через функцию распределения f. Для этого надо явно указать процессы, приводящие к изменению числа частиц с данными координатами и компонентами квазиимпульса; именно при этом учитываются конкретные условия опыта.

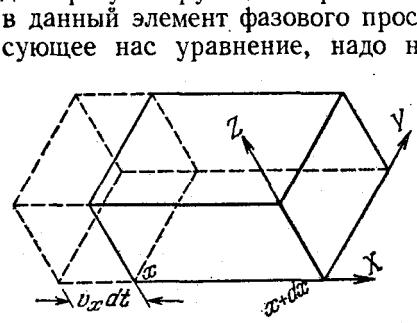


Рис. 13.2. К выводу кинетического уравнения.

Вообще говоря, число частиц в данном элементе фазового объема может измениться за счет.

а) перемещения в пространстве координат (трансляции);

б) действия внешних полей (т. е. ускорения частиц);

в) рассеяния на каких-либо несовершенствах кристаллической решетки или друг на друге («столкновений»);

г) рекомбинации и генерации носителей заряда (или захвата их локальными уровнями).

Для вычисления «трансляционного» вклада заметим, что перемещение частиц в пространстве можно рассматривать как совокуп-

ность трех перемещений вдоль трех произвольно выбранных взаимно перпендикулярных осей координат. Очевидно (рис. 13.2), за время dt через левую грань параллелепипеда в него войдут все частицы (с данной проекцией спина), находившиеся в момент t на расстоянии не более $v_x dt$ от этой грани и движущиеся направо. Число их есть

$$f(\mathbf{p}, x, y, z, t)\frac{v_x\, dt\, dy\, dz\, d\mathbf{p}}{(2\pi\hbar)^3}. \tag{3.2а}$$

Аналогично, число частиц, ушедших за это же время через правую грань, будет

$$f(\mathbf{p}, x+dx, y, z, t)\frac{v_x\, dt\, dy\, dz\, d\mathbf{p}}{(2\pi\hbar)^3}. \tag{3.2б}$$

Разлагая выражение (3.2б) в ряд Тэйлора по dx и вычитая его из (3.2а), получим результирующий «приход», связанный с движением частиц вдоль оси:

$$-\frac{\partial f}{\partial x}v_x\frac{dt\, d\mathbf{r}\, d\mathbf{p}}{(2\pi\hbar)^3}.$$

Аналогично вычисляется и баланс числа частиц за счет перемещений по осям y и z.

Окончательно для изменения числа частиц с данной проекцией спина за время dt за счет перемещения их в пространстве координат находим

$$(dn' - dn)_{\text{трансл}} = -(\mathbf{v}, \nabla f)\frac{dt\, d\mathbf{r}\, d\mathbf{p}}{(2\pi\hbar)^3}. \tag{3.3}$$

Таким же образом можно вычислить и «ускорительный» член — изменение числа частиц в рассматриваемом фазовом параллелепипеде за счет действия сил, обусловленных электрическим и магнитным полями. Действительно, силы вызывают изменение квазиимпульса $\mathbf{p}$, т. е. трансляцию в пространстве квазиимпульсов. Формально этот процесс описывается так же, как и трансляция в обычном пространстве; следует лишь заменить скорость $\mathbf{v} = \frac{d\mathbf{r}}{dt}$ на $\frac{d\mathbf{p}}{dt}$ и вместо градиента в пространстве координат взять градиент в пространстве квазиимпульсов $\nabla_p f$. Итак,

$$(dn' - dn)_{\text{сил}} = -\left(\frac{d\mathbf{p}}{dt}, \nabla_{\mathbf{p}} f\right)\frac{dt\, d\mathbf{r}\, d\mathbf{p}}{(2\pi\hbar)^3}. \tag{3.4}$$

При этом, согласно уравнениям движения (IV.1.9),

$$\frac{d\mathbf{p}}{dt} = \mathbf{F} = -e\mathfrak{E} - \frac{e}{c}[\mathbf{v}\times\mathfrak{B}]. \tag{3.5}$$

Под $\mathfrak{E}$ и $\mathfrak{B}$ здесь следует понимать напряженность электрического и индукцию магнитного полей, мало изменяющихся на протяжении постоянной решетки. Как правило, в число их включаются поля,

созданные внешними источниками (батареей, магнитом, электромагнитной волной и т. д.), а также поля, возникающие в результате перераспределения зарядов в образце при наличии внешних сил. Иногда сюда же включают и поля, созданные структурными дефектами того или иного типа; если эти поля достаточно плавно изменяются в пространстве, то, в зависимости от характера задачи, их можно учитывать либо в уравнении (3.5), либо в выражении, описывающем рассеяние носителей заряда (в «интеграле столкновений» (3.10)).

Обратимся теперь к изменению числа частиц в данном параллелепипеде за счет рассеяния («столкновений»). Рассмотрим сначала рассеяние носителей заряда на каких-либо несовершенствах решетки — «рассеивателях» (но не друг на друге). По смыслу дела столкновения изменяют только распределение носителей по квазиимпульсам (и, может быть, по проекциям спина), но не меняют непосредственно пространственного их распределения. С квантовомеханической точки зрения только такая постановка задачи и имеет смысл: рассеяние есть изменение состояния носителя заряда в результате взаимодействия с рассеивателями; состояние же характеризуется квазиимпульсом и проекцией спина. Таким образом, из аргументов функции распределения $\mathbf{p}$ и $\mathbf{r}$ второй остается неизменным при рассеянии, а меняется только первый. Пусть квазиимпульс носителя заряда до столкновения равен $\mathbf{p}'$, а после столкновения $\mathbf{p}$ *). Очевидно, число таких актов рассеяния в единицу времени пропорционально числу носителей в начальном состоянии и числу свободных мест — в конечном, т. е. оно имеет вид

$$\mathscr{P}_1(\mathbf{p}', \mathbf{p}) f(\mathbf{p}', \mathbf{r})[1-f(\mathbf{p}, \mathbf{r})]\frac{d\mathbf{p}\, d\mathbf{p}'\, d\mathbf{r}}{(2\pi\hbar)^6}. \qquad (3.6)$$

Здесь $\mathscr{P}_1$ — коэффициент пропорциональности, могущий, вообще говоря, зависеть и от $\mathbf{r}$. Строго говоря, величина $\mathscr{P}_1$ могла бы зависеть и от напряженностей внешних электрического и магнитного полей. Очевидно, однако, что для этого последние должны быть сравнимы по величине с напряженностями атомных полей. Столь сильные внешние поля мы рассматривать не будем. Интегрируя (3.6) по $\mathbf{p}'$ и умножая на dt, получим полное число носителей заряда, приходящих за время dt в рассматриваемый элемент фазового объема за счет столкновений:

$$\frac{d\mathbf{r}\, d\mathbf{p}\, dt}{(2\pi\hbar)^3}\int \mathscr{P}_1(\mathbf{p}', \mathbf{p}) f(\mathbf{p}', \mathbf{r})[1-f(\mathbf{p}, \mathbf{r})]\frac{d\mathbf{p}'}{(2\pi\hbar)^3} = A\,\frac{dt\, d\mathbf{r}\, d\mathbf{p}}{(2\pi\hbar)^3}. \qquad (3.7)$$

Интеграл A определяется этим соотношением и описывает «приход» частиц в единичный элемент фазового объема в единицу времени.

*) Мы не указываем явно спиновые переменные, считая в случае необходимости, что они включены в $\mathbf{p}$, $\mathbf{p}'$. Естественно, при этом интегрирование, например, по $\mathbf{p}'$ включает и суммирование по двум соответствующим проекциям спина.

Аналогично, выбирая состояние с квазиимпульсом $\mathbf{p}$ в качестве начального, а состояние с квазиимпульсом $\mathbf{p}'$ — в качестве конечного, находим полное число частиц, уходящих за время dt из рассматриваемого элемента фазового пространства в результате столкновений:

$$\frac{d\mathbf{r}\, d\mathbf{p}\, dt}{(2\pi\hbar)^3} \int \mathscr{P}_2(\mathbf{p}, \mathbf{p}') f(\mathbf{p}, \mathbf{r}) [1 - f(\mathbf{p}', \mathbf{r})] \frac{d\mathbf{p}'}{(2\pi\hbar)^3} = B \frac{dt\, d\mathbf{r}\, d\mathbf{p}}{(2\pi\hbar)^3}. \quad (3.8)$$

Интеграл B определяется этим соотношением и описывает «уход» частиц из единичного элемента фазового объема в единицу времени.

Комбинируя формулы (3.7) и (3.8), получаем полное изменение числа частиц в результате рассеяния:

$$(dn' - dn)_{\text{столкн}} = J[f] \frac{dt\, d\mathbf{r}\, d\mathbf{p}}{(2\pi\hbar)^3}. \quad (3.9)$$

Здесь

$$J[f] \equiv A - B = \frac{1}{(2\pi\hbar)^3} \int d\mathbf{p}' \{\mathscr{P}_1(\mathbf{p}', \mathbf{p}) f(\mathbf{p}', \mathbf{r}) [1 - f(\mathbf{p}, \mathbf{r})] - \mathscr{P}_2(\mathbf{p}, \mathbf{p}') f(\mathbf{p}, \mathbf{r}) [1 - f(\mathbf{p}', \mathbf{r})]\}. \quad (3.10)$$

Выражение в правой части (3.10) называют *интегралом столкновений*. Как мы видим, функция распределения входит в него, вообще говоря, нелинейно.

Коэффициенты $\mathscr{P}_1(\mathbf{p}', \mathbf{p})$ и $\mathscr{P}_2(\mathbf{p}, \mathbf{p}')$, характеризующие темп рассеяния, имеют размерность $\text{см}^3/\text{с}$. Они пропорциональны вероятностям соответствующих переходов. Иногда их представляют как произведения абсолютного значения скорости частицы до столкновения на эффективные сечения рассеяния с данными изменениями квазиимпульса (и, может быть, проекции спина). Для вычисления их надо задать конкретный механизм рассеяния — природу рассеивателей, энергию взаимодействия их с носителями заряда и т. д. — и решить соответствующую задачу динамики.

Заметим, что — в отличие от движения носителей заряда в плавных полях — процессы взаимодействия электронов и дырок с рассеивателями могут и не допускать классического описания. Таким образом, даже в случае плавно меняющихся полей на определенном этапе решения задачи о вычислении кинетических коэффициентов может оказаться необходимым использование методов квантовой механики. Это обстоятельство очень существенно для дальнейшего. Как мы увидим в гл. XIV, необходимость учета квантовых эффектов влечет за собой серьезное ограничение условий применимости метода кинетического уравнения.

В случае рассеяния носителей заряда друг на друге общая формула (3.9) остается в силе, но явный вид интеграла столкновений изменяется, ибо теперь при каждом столкновении меняется состояние не одной, а двух частиц. Пусть квазиимпульсы частиц до столкновения равны $\mathbf{p}_1'$ и $\mathbf{p}_2'$, а после столкновения $\mathbf{p}_1$ и $\mathbf{p}_2$.

Подобно (3.6), число таких столкновений в единицу времени есть

$$\mathscr{P}_1(\mathbf{p}_1', \mathbf{p}_2'; \mathbf{p}_1, \mathbf{p}_2) f(\mathbf{p}_1', \mathbf{r}) \frac{d\mathbf{p}_1'}{(2\pi\hbar)^3} f(\mathbf{p}_2', \mathbf{r}) \frac{d\mathbf{p}_2'}{(2\pi\hbar)^3} \times$$
$$\times [1 - f(\mathbf{p}_1, \mathbf{r})] \frac{d\mathbf{p}_1}{(2\pi\hbar)^3} [1 - f(\mathbf{p}_2, \mathbf{r})] \frac{d\mathbf{p}_2}{(2\pi\hbar)^3},$$

где $\mathscr{P}_1$ — коэффициент пропорциональности.

Полное число столкновений, связанных с приходом носителя заряда в рассматриваемый элемент фазового объема, получится отсюда интегрированием по $\mathbf{p}_1'$, $\mathbf{p}_2$ и $\mathbf{p}_2'$. Принимая во внимание еще обратный процесс («уход»), получаем

$$J[f] = \int \frac{d\mathbf{p}_1' \, d\mathbf{p}_2' \, d\mathbf{p}_2}{(2\pi\hbar)^9} \{\mathscr{P}_1(\mathbf{p}_1', \mathbf{p}_2'; \mathbf{p}_1, \mathbf{p}_2) \times$$
$$\times f(\mathbf{p}_1', \mathbf{r}) f(\mathbf{p}_2', \mathbf{r}) [1 - f(\mathbf{p}_1, \mathbf{r})] [1 - f(\mathbf{p}_2, \mathbf{r})] -$$
$$- \mathscr{P}_2(\mathbf{p}_1, \mathbf{p}_2; \mathbf{p}_1', \mathbf{p}_2') f(\mathbf{p}_1, \mathbf{r}) f(\mathbf{p}_2, \mathbf{r}) [1 - f(\mathbf{p}_1', \mathbf{r})] [1 - f(\mathbf{p}_2', \mathbf{r})]\}. \quad (3.11)$$

Коэффициенты $\mathscr{P}_1$, $\mathscr{P}_2$ пропорциональны вероятностям рассматриваемых актов рассеяния.

Наконец, изменение числа носителей заряда за счет рекомбинации или захвата можно было бы написать по образцу (3.10) — изменился бы только смысл коэффициентов $\mathscr{P}_1$, $\mathscr{P}_2$ и интегрирование по $\mathbf{p}'$ заменилось бы (в случае захвата) суммированием по дискретным переменным, нумерующим различные квантовые состояния электронов на центрах захвата. Чаще всего, однако, рекомбинационные явления можно вообще не принимать во внимание при вычислении кинетических коэффициентов. Дело в том, что скорости процессов рассеяния и рекомбинации носителей заряда, как правило, резко различны. Как мы знаем (гл. IX), характерные рекомбинационные времена — порядка $10^{-8} \div 10^{-6}$ с и более. С другой стороны, типичное среднее время свободного пробега $\tau_p \equiv \langle \tau \rangle$ (см. § I.2), характеризующее темп процессов рассеяния («частоту столкновений»), — порядка $10^{-11} \div 10^{-13}$ с. Таким образом, можно считать, что рассеяние происходит при практически неизменном — безразлично, равновесном или неравновесном — числе свободных носителей заряда. С другой стороны, рекомбинация и захват становятся заметными лишь спустя длительное время после того, как процессы рассеяния (совместно с действием полей и т. д.) «сформируют» функцию распределения носителей по квазиимпульсам. Это разделение процессов на быстрые (в данном случае — рассеяние) и медленные (в данном случае — рекомбинация или захват), протекающие практически независимо друг от друга, весьма характерно для кинетики. По существу, мы уже встречались с ним в гл. VII и др., рассматривая диффузию носителей заряда и связанные с ней явления. Диффузия также представляет собой медленный процесс, характеризуемый коэффициентом диффузии D.

С другой стороны, сама величина D определяется быстрыми процессами рассеяния, формирующими функцию распределения и протекающими зачастую так, как если бы никакой диффузии не было.

Комбинируя теперь формулы (3.1), (3.3), (3.4) и (3.9), получаем уравнение для функции распределения:

$$\frac{\partial f}{\partial t} = -(\mathbf{v}, \nabla f) - (\mathbf{F}, \nabla_p f) + J[f]. \tag{3.12}$$

Это и есть кинетическое уравнение. Интеграл столкновений $J[f]$ дается здесь формулами (3.10) или (3.11) (или их суммой, если существенны оба вида рассеяния), а сила $\mathbf{F}$ — выражением (3.5).

Уравнение (3.12) написано для случая, когда имеются носители заряда только одного типа. При наличии нескольких типов носителей (электронов и дырок, «легких» и «тяжелых» дырок и т. д.) надо ввести свою функцию распределения для каждого типа. Соответственно мы будем иметь столько кинетических уравнений вида (3.12), сколько есть таких типов. При этом в правой части формулы (3.12) появится сумма по всем типам частиц, на которых рассматриваемые носители заряда могут рассеиваться. Таким образом, если существенно взаимное рассеяние или превращение частиц разных типов, то мы приходим к системе связанных кинетических уравнений.

Уравнение (3.12) — интегро-дифференциальное; математическая задача о его решении становится определенной, лишь если задать еще граничные условия. Это, однако, удобно делать не в общем виде, а применительно к тому или иному конкретному случаю, и мы здесь не будем обсуждать этот вопрос.

Вывод кинетического уравнения завершает намеченную в предыдущем параграфе схему вычисления кинетических коэффициентов. Как видно из предыдущего, решение этой задачи делится на два этапа, которые несколько условно можно назвать «механическим» и «статистическим». Первый из них состоит в вычислении коэффициентов пропорциональности $\mathscr{P}_1$, $\mathscr{P}_2$, т. е. вероятностей рассеяния при заданном механизме последнего. Второй этап состоит в решении уравнения (или уравнений) (3.12) при известных функциях $\mathscr{P}_1$, $\mathscr{P}_2$.

§ 4. Термодинамическое равновесие. Принцип детального равновесия

Кинетическое уравнение, выведенное в предыдущем параграфе для произвольных неравновесных условий, сохраняет силу и в условиях термодинамического равновесия. При этом оно определяет равновесную функцию распределения f_0. Последняя, однако, уже известна нам из равновесной статистической физики — это есть функция Ферми (V.3.1). Воспользуемся этим обстоятельством,

чтобы прийти к одному полезному соотношению между коэффициентами $\mathscr{P}_1$ и $\mathscr{P}_2$.

Рассмотрим систему электронов в отсутствие внешнего электрического поля. Поскольку f_0 не зависит от времени, уравнение (3.12) принимает вид

$$-(\mathbf{v},\ \nabla f_0)-(\mathbf{F},\ \nabla_{\mathbf{p}} f_0)+J[f_0]=0. \tag{4.1}$$

При этом согласно (V.3.1)

$$f_0=\left\{\exp\left[\frac{E(\mathbf{p})-e\varphi-F}{kT}\right]+1\right\}^{-1}, \tag{4.2}$$

где φ — электростатический потенциал, отличный от константы, если в условиях равновесия имеется искривление зон.

Обозначим через f_0' производную от f_0 по энергии $E(\mathbf{p})$. Поскольку функция f_0 зависит от координат только через потенциал φ, а последний входит лишь в комбинации $E(\mathbf{p})-e\varphi$, мы имеем

$$\nabla f_0=-ef_0'\cdot\nabla\varphi. \tag{4.3}$$

Далее, в силу (IV.1.3)

$$\nabla_{\mathbf{p}} f_0=(\nabla_{\mathbf{p}} E(\mathbf{p}))\,f_0'=\mathbf{v}f_0'. \tag{4.4}$$

Здесь принято во внимание, что в условиях термодинамического равновесия электрохимический потенциал F и температура T не зависят от координат.

Сила $\mathbf{F}$ дается правой частью уравнения (3.5), где, в рассматриваемом случае,

$$\mathfrak{E}=-\nabla\varphi.$$

Следовательно,

$$(\mathbf{F},\ \nabla_{\mathbf{p}} f_0)=ef_0'\left\{(\mathbf{v},\ \nabla\varphi)-\frac{1}{c}(\mathbf{v},\ [\mathbf{v}\times\mathfrak{B}])\right\}. \tag{4.5}$$

Последнее слагаемое в правой части (4.5) тождественно равно нулю. Это означает, что однородное магнитное поле само по себе не может нарушить равновесное распределение носителей заряда по квазиимпульсам.

Подставляя выражения (4.3) и (4.5) в уравнение (4.1), мы получаем

$$e(\mathbf{v},\ \nabla\varphi)f_0'-e(\mathbf{v},\ \nabla\varphi)f_0'+J[f_0]=0. \tag{4.1'}$$

Видим, что в условиях термодинамического равновесия «трансляционное» и «ускорительное» слагаемые в кинетическом уравнении взаимно уничтожаются. Это означает, что имеет место взаимная компенсация плотностей диффузионного и дрейфового токов (ср. § VI.3). Действительно, диффузионный ток обусловлен как раз неоднородным распределением носителей заряда в пространстве, а дрейфовый — действием силы со стороны электрического поля.

Таким образом, уравнение (4.1) теперь принимает вид

$$J[f_0]=0. \tag{4.1''}$$

Смысл этого равенства ясен: в условиях взаимной компенсации первых двух слагаемых в (4.1) число частиц в любом элементе фазового пространства может измениться только за счет столкновений *). Поскольку в рассматриваемых условиях функция распределения не зависит от времени, «приход» и «уход» частиц должны взаимно компенсироваться.

Уравнение (4.1'') с учетом (4.2) устанавливает некоторую связь между величинами $\mathscr{P}_1(\mathbf{p}', \mathbf{p})$ и $\mathscr{P}_2(\mathbf{p}, \mathbf{p}')$.

Рассмотрим сначала случай, когда рассеяние носителей заряда происходит на каких-либо несовершенствах решетки (но не друг на друге). Тогда, согласно (3.10), равенство (4.1'') принимает вид

$$\int \{\mathscr{P}_1(\mathbf{p}', \mathbf{p}) f_0(\mathbf{p}')[1-f_0(\mathbf{p})] - \mathscr{P}_2(\mathbf{p}, \mathbf{p}') f_0(\mathbf{p})[1-f_0(\mathbf{p}')]\}\, d\mathbf{p}' = 0. \tag{4.6}$$

Допустим, что равенство нулю интеграла в (4.6) влечет за собой и обращение в нуль подынтегрального выражения. Тогда, подставляя в (4.6) явное выражение для функции Ферми, мы получаем

$$\frac{\mathscr{P}_1(\mathbf{p}', \mathbf{p})}{\mathscr{P}_2(\mathbf{p}, \mathbf{p}')} = \exp\frac{E(\mathbf{p}')-E(\mathbf{p})}{kT}. \tag{4.7}$$

Видно, что рассеяние с увеличением энергии менее вероятно, чем с уменьшением ее: при $E(\mathbf{p}') < E(\mathbf{p})$ правая часть (4.7) меньше единицы. При упругом рассеянии, когда энергия частицы не изменяется вовсе, $\mathscr{P}_1(\mathbf{p}', \mathbf{p}) = \mathscr{P}_2(\mathbf{p}, \mathbf{p}')$.

Предположение о равенстве нулю не только интеграла, но и подынтегрального выражения в левой части (4.6) означает, что в условиях равновесия поток частиц из данного элемента фазового пространства в любой другой точно уравновешивается обратным ему потоком. Меньшая вероятность перехода из состояний с малой энергией компенсируется при этом большей их «заселенностью». Иначе говоря, приход и уход частиц уравновешиваются не суммарно, а детально — порознь для каждых двух элементов фазового пространства. Соответственно утверждение, выражаемое равенством (4.7), носит название *принципа детального равновесия.*

Изложенное выше, разумеется, не представляет собой вывода указанного принципа. Это есть лишь способ догадаться, какого соотношения между величинами $\mathscr{P}_1(\mathbf{p}', \mathbf{p})$ и $\mathscr{P}_2(\mathbf{p}, \mathbf{p}')$ можно было бы ожидать. Фактически равенство (4.7) при определенных условиях вытекает из квантовой механики ([М2], § 108), и его можно ис-

*) Ситуация становится особенно ясной, если искривления зон нет вообще, т. е. система носителей заряда пространственно однородна. В этом случае функция распределения не зависит от координат, а $\mathcal{E} = -\nabla\varphi = 0$, т. е. первое и второе слагаемые в (4.1) порознь обращаются в нуль: переноса частиц ни в координатном, ни в импульсном пространстве нет вообще.

пользовать для нахождения явного вида равновесной функции распределения. Так, в частности, обстоит дело при упругом рассеянии бесспиновых частиц (что и будет проиллюстрировано в дальнейшем на конкретных примерах). При этом слово «бесспиновый» не обязательно понимать буквально: речь идет о процессах рассеяния, в которых спин частицы не играет роли. Именно с такими процессами мы чаще всего имеем дело в неферромагнитных полупроводниках и металлах.

Совершенно аналогично можно рассмотреть и рассеяние носителей заряда друг на друге. Интеграл столкновений при этом дается формулой (3.11). Комбинируя ее с соотношением (4.1″), мы получаем, подобно (4.7),

$$\frac{\mathscr{P}_1(\mathbf{p}', \mathbf{p}'; \mathbf{p}_1, \mathbf{p}_2)}{\mathscr{P}_2(\mathbf{p}_1, \mathbf{p}_2; \mathbf{p}_1', \mathbf{p}_2')} = \exp\frac{E(\mathbf{p}')+E(\mathbf{p}')-E(\mathbf{p}_1)-E(\mathbf{p}_2)}{kT}. \quad (4.8)$$

Под знаком экспоненты здесь стоит полное изменение энергии двух частиц при столкновении. В отсутствие обмена энергией с какими-либо третьими телами оно, очевидно, должно равняться нулю. Таким образом, равенство (4.8) принимает более простой вид:

$$\mathscr{P}_1(\mathbf{p}_1', \mathbf{p}_2'; \mathbf{p}_1, \mathbf{p}_2) = \mathscr{P}_2(\mathbf{p}_1, \mathbf{p}_2; \mathbf{p}_1', \mathbf{p}_2'). \quad (4.8')$$

Как и (4.7), равенство (4.8′) справедливо не всегда [М2]. Однако уточнения, которые необходимо внести в общем случае, в интересующих нас задачах ничего не меняют.

§ 5. Малые отклонения от равновесия

Систему электронов можно вывести из состояния термодинамического равновесия, накладывая напряжение на образец, или создавая в нем градиент концентрации носителей заряда или температуры. При этом электрохимический потенциал или температура (или и то и другое) становятся зависящими от координат. Принимая во внимание это обстоятельство, мы можем записать неравновесную функцию распределения в виде

$$f(\mathbf{p}, \mathbf{r}) = f_0(\mathbf{p}, \mathbf{r}) + f_1(\mathbf{p}, \mathbf{r}). \quad (5.1)$$

Здесь

$$f_0 = \left\{\exp\frac{E(\mathbf{p})-e\varphi-F(\mathbf{r})}{kT(\mathbf{r})}+1\right\}^{-1} \quad (5.2)$$

есть функция Ферми с изменяющимися в пространстве температурой и электрохимическим потенциалом, а f_1 — неизвестная пока функция, которую надо определить из кинетического уравнения.

Представление $f(\mathbf{p}, \mathbf{r})$ в виде (5.1) оказывается удобным, если градиенты функций $F(\mathbf{r})$ и $T(\mathbf{r})$ достаточно малы. Как будет

показано в § 7, при этом в существенной области энергий

$$|f_1| \ll f_0. \tag{5.3}$$

Коль скоро условие (5.3) выполняется, говорят о малых отклонениях от термодинамического равновесия; систему носителей заряда при этом называют слабо неравновесной. В применении к таким системам кинетическое уравнение (3.12) несколько упрощается. Действительно, перепишем его, пользуясь явным выражением (3.5) для силы **F**. Получим

$$\frac{\partial f}{\partial t} + (\mathbf{v}, \nabla f) - (e\mathfrak{E}, \nabla_{\mathbf{p}} f) - \frac{e}{c}([\mathbf{v}\times\mathfrak{B}], \nabla_{\mathbf{p}} f) = J[f]. \tag{5.4}$$

Условие (5.3) позволяет заменить f на f_0 во втором и третьем слагаемых в левой части (5.4). Действительно, в условиях, когда $\nabla T = \nabla F = 0$, функция f_1 должна обращаться в нуль. Следовательно, мы можем ожидать, что f_1 будет линейно зависеть от $\mathfrak{E}$ и ∇T (в § 7 это будет доказано). Ограничиваясь, как и в § 1, приближением, линейным по $\mathfrak{E}$ и ∇T, мы должны отбросить члены $(\mathbf{v}, \nabla f_1)$ и $(e\mathfrak{E}, \nabla_{\mathbf{p}} f_1)$, как величины высшего порядка малости. С другой стороны, для вычисления слагаемого с магнитным полем и интеграла столкновений такая аппроксимация недостаточна: как мы видели в § 4, при $f = f_0$ эти выражения обращаются в нуль.

Вычисляя градиенты ∇f_0 и $\nabla_{\mathbf{p}} f_0$, мы получаем, подобно (4.3) и (4.4),

$$\nabla f_0 = \left(-e\nabla\varphi - \nabla F - \frac{E(\mathbf{p}) - e\varphi - F}{kT}\nabla kT\right) f_0', \qquad \nabla_{\mathbf{p}} f_0 = \mathbf{v} f_0',$$

где, как и раньше, $f_0' = \partial f_0/\partial E\,(\mathbf{p})$.

Таким образом, в принятом приближении сумма второго и третьего слагаемых в левой части (5.4) есть

$$\left\{-e(\mathbf{v}, \nabla\varphi) - (\mathbf{v}, \nabla F) - \frac{E(\mathbf{p}) - e\varphi - F}{T}(\mathbf{v}, \nabla T) - e(\mathfrak{E}, \mathbf{v})\right\} f_0'. \tag{5.5}$$

Поскольку $\mathfrak{E} = -\nabla\varphi$, первый и последний члены в (5.5) взаимно уничтожаются. Следовательно, вместо (5.5) в кинетическое уравнение фактически входит выражение

$$\left\{-(\mathbf{v}, \nabla F) - \frac{E(\mathbf{p}) - e\varphi - F}{T}(\mathbf{v}, \nabla T)\right\} f_0'. \tag{5.5'}$$

Видим, что роль напряженности электрического поля действительно играет величина $\mathfrak{E}' = \frac{1}{e}\nabla F$ (1.1).

Итак, при малых отклонениях от равновесия кинетическое уравнение (5.4) принимает вид

$$\frac{\partial f}{\partial t} - (e\mathfrak{E}', \mathbf{v}) f_0' - \frac{E(\mathbf{p}) - e\varphi - F}{T}(\mathbf{v}, \nabla T) f_0' - \frac{e}{c}([\mathbf{v}\times\mathfrak{B}], \nabla_{\mathbf{p}} f_1) = J[f]. \tag{5.4'}$$

В дальнейшем при рассмотрении электронного газа удобно будет отсчитывать энергию $E(\mathbf{p})$ от дна зоны проводимости. При этом $F + e\varphi = \zeta$, где химический потенциал ζ — тот же, что и в гл. V. В стационарных условиях и при $\mathcal{E}' = \nabla T = 0$ отсюда следует, как мы и ожидали, что $f_1 = 0$ и $f = f_0$. Иначе говоря, равновесие нарушается, лишь если отличен от нуля хотя бы один из векторов ∇F и ∇T. Только при этом условии могут возникнуть электрический ток и поток энергии.

§ 6. Интеграл столкновений в случае упругого рассеяния и изотропных изоэнергетических поверхностей. Время релаксации импульса

Кинетическое уравнение (5.4′), как и (3.12), — нелинейное интегро-дифференциальное. В общем случае аналитическое его решение связано с большими математическими трудностями. По этой причине приходится вводить упрощения, основанные на тех или иных физических особенностях задачи.

Прежде всего, заметим, что во многих случаях вероятность изменения проекции спина при рассеянии весьма мала и, кроме того, частицы с «левой» и «правой» проекциями спина рассеиваются практически одинаково. Действительно, проекция спина может измениться только за счет сравнительно слабых магнитных взаимодействий. В пренебрежении последними суммирование по проекциям спина в правой части (3.10) отпадает, и мы можем явно рассматривать только частицы с какой-нибудь одной проекцией спина. Наличие электронов или дырок с другой проекцией спина при этом учитывается просто множителем 2 в формулах для концентрации частиц, плотности тока и плотности потока энергии, как это и сделано в равенствах (2.1), (2.2) и (2.3).

Далее, весьма часто рассеяние носителей заряда носит почти упругий характер. Так, например, обстоит дело при рассеянии их атомами заряженной или нейтральной примеси, дислокациями или иными структурными дефектами решетки. Все эти объекты обладают значительно большей массой, нежели электрон или дырка. Как известно из механики, при столкновении легкой частицы с тяжелой может сильно измениться импульс каждой из них, но обмен энергией между ними весьма затруднен: изменение энергии легкой частицы при столкновении оказывается малым по сравнению с самой этой энергией. В следующей главе мы убедимся, что так же обстоит дело и при рассеянии электронов на акустических колебаниях решетки: изменение энергии при рассеянии пропорционально $(m/M)^{1/2}$, где m — эффективная масса электрона, а M — масса атома решетки.

В соответствии со сказанным предположим, что рассеяние носителей заряда происходит без изменения их энергии: носители

только перераспределяются по данной изоэнергетической поверхности, не сходя с нее. Такой тип рассеяния называется *упругим*.

Разумеется, предположение об упругости рассеяния представляет собой идеализацию, использование которой оправдано не всегда. В самом деле, при полном отсутствии обмена энергией между носителями заряда и их окружением в кристалле не выделялось бы джоулево тепло и, далее, вообще не могло бы установиться термодинамическое равновесие между электронами и дырками, с одной стороны, и кристаллической решеткой — с другой. Утверждение о почти упругом характере рассеяния означает, что процесс обмена энергией между носителями заряда и решеткой протекает гораздо медленнее, чем процесс обмена квазиимпульсом. Принятая нами идеализация означает пренебрежение первым из этих процессов. Это может быть оправдано лишь при достаточно малом нарушении равновесия (например, в слабом электрическом поле), когда можно пренебречь энергией, приобретаемой носителями заряда от внешнего источника.

Математически предположение об упругости рассеяния выражается тем, что коэффициенты $\mathscr{P}_1(\mathbf{p}', \mathbf{p})$ и $\mathscr{P}_2(\mathbf{p}, \mathbf{p}')$, фигурирующие в интеграле столкновений, отличны от нуля лишь при $E(\mathbf{p}') = E(\mathbf{p})$. В то же время интегралы от $\mathscr{P}_1$ и $\mathscr{P}_2$ по $\mathbf{p}'$ должны быть отличны от нуля. Этим требованиям можно удовлетворить, полагая

$$\begin{aligned} \mathscr{P}_1(\mathbf{p}', \mathbf{p}) &= \delta[E(\mathbf{p}') - E(\mathbf{p})]\, S(\mathbf{p}', \mathbf{p}), \\ \mathscr{P}_2(\mathbf{p}, \mathbf{p}') &= \delta[E(\mathbf{p}') - E(\mathbf{p})]\, S(\mathbf{p}, \mathbf{p}'), \end{aligned} \tag{6.1}$$

где $S(\mathbf{p}', \mathbf{p})$, $S(\mathbf{p}, \mathbf{p}')$ — некоторые функции, $\delta[E(\mathbf{p}') - E(\mathbf{p})]$ — дельта-функция Дирака *). Действительно, согласно свойствам δ-функции интегралы от величин (6.1), умноженных на функцию распределения f, во-первых, отличны от нуля и, во-вторых, содержат только значения f и S при $E(\mathbf{p}') = E(\mathbf{p})$. В гл. XIV выражения (6.1) будут получены явно. При этом, в силу принципа детального равновесия (4.7),

$$S(\mathbf{p}', \mathbf{p}) = S(\mathbf{p}, \mathbf{p}').$$

Подставляя выражения (6.1) в правую часть (3.10), мы получаем

$$J[f] = \frac{1}{(2\pi\hbar)^3} \int S(\mathbf{p}, \mathbf{p}')\, \delta[E(\mathbf{p}) - E(\mathbf{p}')] \{f(\mathbf{p}', \mathbf{r}) - f(\mathbf{p}, \mathbf{r})\}\, d\mathbf{p}'. \tag{6.2}$$

Итак, предположение об упругом характере рассеяния привело к тому, что искомая функция распределения вошла под знак интеграла столкновений линейно.

Дальнейшего упрощения интеграла столкновений можно добиться, приняв во внимание свойства симметрии изоэнергетических

*) См. Приложение IV.

поверхностей. Рассмотрим простейший случай, когда названные поверхности не вырождены и изотропны, т. е. представляют собой сферы *)

$$E = E(p^2). \tag{6.3}$$

При этом функцию $E(p^2)$ можно обратить; соответствующая обратная функция есть $p^2(E)$.

Заметим, что по своему физическому смыслу функция $S(\mathbf{p}, \mathbf{p}')$ есть скаляр. Следовательно, она может выражаться лишь через скалярные комбинации векторов $\mathbf{p}$ и $\mathbf{p}'$.

В рассматриваемом случае их пять:

$$p^2, \quad p'^2, \quad (\mathbf{p}, \mathbf{p}') = pp' \cos\theta, \quad E(p^2), \quad E(p'^2). \tag{6.4}$$

Здесь $\theta = \widehat{\mathbf{p}, \mathbf{p}'}$ есть угол рассеяния.

При упругом рассеянии последние две из величин (6.4) оказываются одинаковыми. Далее, в силу (6.3) $p^2 = p'^2$. Таким образом, остаются только два независимых скаляра; в качестве таковых удобно взять энергию $E = E(p^2)$ и $\cos\theta$:

$$S(\mathbf{p}, \mathbf{p}') = S(E, \cos\theta). \tag{6.5}$$

Итак, условия упругости рассеяния и изотропии невырожденных изоэнергетических поверхностей привели к тому, что функция S оказалась зависящей не от пяти, а только от двух аргументов; при этом один из них в процессе рассеяния остается неизменным.

В дальнейшем, при рассмотрении квантовомеханической части задачи (гл. XIV), коэффициенты $\mathscr{P}_1$, $\mathscr{P}_2$ будут явно вычислены для ряда механизмов рассеяния. Мы непосредственно убедимся тогда в справедливости формул (6.1) и (6.5) в разных конкретных случаях.

Оказывается, однако, что для решения статистической части задачи при малых отклонениях от равновесия специализация вида функции $S(E, \cos\theta)$ не требуется. Достаточно представить функцию распределения в виде (5.1).

В соответствии с § 1 ограничимся учетом членов, линейных по $\mathfrak{E}$ и ∇T. Легко сообразить при этом, как должна зависеть функция f_1 от $\mathfrak{E}$, ∇T и $\mathfrak{B}$. Надо лишь принять во внимание, что f_1 есть скаляр, $\mathfrak{E}$ и ∇T — векторы, а $\mathfrak{B}$ — псевдовектор. Рассмотрим сначала те же три частных случая, что и в § 1.

а) Носители заряда в постоянном и однородном слабом электрическом поле. Поскольку функция f_1 должна быть скаляром, линейно содержащим вектор $\mathfrak{E}$, здесь следует положить

$$f_1 = (\mathbf{p}, \mathfrak{E})\psi(E), \tag{6.6а}$$

где $\psi(E)$ — пока неизвестная нам скалярная функция энергии носителя заряда.

*) Предположение о параболичности закона дисперсии не обязательно.

б) Носители заряда в постоянном и однородном слабом температурном поле. Здесь

$$f_1 = (\mathbf{p}, \nabla T)\,\chi_1(E) + (\mathbf{p}, \mathcal{E}')\,\chi_2, \tag{6.6б}$$

где χ_1 и χ_2 — неизвестные пока скалярные функции, а величина $\mathcal{E}'$ связана с нагревом образца и с возможным перераспределением зарядов в нем. Фактически вектор $\mathcal{E}'$ не независим, а выражается через ∇T. Связь между этими векторами, однако, определяется условиями опыта и удобнее здесь ее не специализировать.

в) Носители заряда в постоянных и однородных электрическом и магнитном полях. Здесь имеются три независимых вектора: $\mathcal{E}$, $[\mathcal{E}\times\mathcal{B}]$, $(\mathcal{B}, \mathcal{E})\,\mathcal{B}$. Соответственно мы имеем

$$f_1 = (\mathbf{p}, \mathcal{E})\,\psi_1(E) + (\mathbf{p}, [\mathcal{E}\times\mathcal{B}])\,\psi_2(E) + (\mathbf{p}, \mathcal{B})(\mathcal{B}, \mathcal{E})\,\psi_3(E), \tag{6.6в}$$

где ψ_1, ψ_2 и ψ_3 — пока неизвестные скалярные функции. Они зависят от энергии носителя заряда и, может быть, от величины $\mathcal{B}^2$ *).

Аналогичные выражения можно написать и при наличии сразу и электрического, и магнитного поля, и градиента температуры или только магнитного поля и градиента температуры. В общем виде, следовательно, мы имеем

$$f_1 = (\mathbf{p}, \xi(E)), \tag{6.7}$$

где вектор ξ представляет собой линейную комбинацию векторов, характеризующих внешние воздействия. Скалярные коэффициенты в этой линейной комбинации зависят от E и, может быть, от $\mathcal{B}^2$.

Очевидно, функция (5.1) с учетом (6.7) описывает состояние с отличными от нуля значениями плотности электрического тока и потока энергии. Действительно, f_1, как и $\mathbf{v}(\mathbf{p})$, есть нечетная функция квазиимпульса $\mathbf{p}$ и интегралы в правых частях (2.2) и (2.3) оказываются отличными от нуля. Вместе с тем средняя энергия носителей заряда $\langle E\rangle$ в этом состоянии не отличается от равновесной. В самом деле,

$$\langle E\rangle = \frac{2}{(2\pi\hbar)^3}\int f(\mathbf{p})\,E(\mathbf{p})\,d\mathbf{p}. \tag{6.8}$$

Поскольку $E(\mathbf{p})$ есть четная функция квазиимпульса $\mathbf{p}$, слагаемое f_1 не дает вклада в интеграл (6.8): в принятом приближении носители заряда получают от внешних полей и передают решетке только квазиимпульс, но не энергию **).

Подставим выражение (5.1) в интеграл столкновений (6.2) и примем во внимание равенство (6.5). В силу упругого характера

*) Как мы увидим в следующем параграфе, последняя возможность не противоречит условию (5.3).

**) Этот результат можно было предвидеть заранее, вспомнив, что джоулево тепло выражается через квадрат напряженности электрического поля.

рассеяния $f_0\ (E(p)) = f_0\ (E\ (p'))$ и, следовательно,

$$J[f] = \int \frac{d\mathbf{p}'}{(2\pi\hbar)^3}\,\delta\,(E\,(p) - E\,(p'))\,S\,(E\,(p),\ \cos\theta)\,\{(\mathbf{p}',\ \xi) - (\mathbf{p},\ \xi)\}. \quad (6.9)$$

Интеграл по $\mathbf{p}'$ удобно вычислять в сферических координатах, направив полярную ось по вектору $\mathbf{p}$. Полярные углы векторов $\mathbf{p}'$ и ξ относительно $\mathbf{p}$ обозначим, соответственно, через (θ, φ) и (α, β), а вместо $\mathbf{p}'$ введем переменную $E' = E\ (p')$. Тогда

$$d\mathbf{p}' = \frac{p'^2\,(E')}{v\,(E')}\,dE'\sin\theta\cdot d\theta\,d\varphi,$$

где $v\,(E') = |\,dE'/dp'\,|$.

Введем плотность состояний по формуле (V.7.8а) и выполним интегрирование по E' в (6.9) с помощью δ-функции. Получим

$$J[f] = \frac{N\,(E)\,\xi\,(E)}{8\pi}\int_0^{\pi} d\theta\cdot\sin\theta\cdot S\,(E,\ \cos\theta)\int_0^{2\pi} d\varphi\cdot(p\cos\widehat{(\mathbf{p}',\ \xi)} - p\cos\alpha). \quad (6.10)$$

Как известно из сферической тригонометрии,

$$\cos\widehat{(\mathbf{p}',\ \xi)} = \cos\theta\cdot\cos\alpha + \sin\theta\cdot\sin\alpha\cdot\cos(\varphi - \beta).$$

Подставим это выражение в правую часть (6.10) и проинтегрируем по φ. Получим

$$J[f] = -\frac{\xi p\,(E)\cos\alpha}{\tau\,(E)} \equiv -\frac{f_1}{\tau\,(E)}, \quad (6.11)$$

где

$$\frac{1}{\tau\,(E)} = \frac{N\,(E)}{4}\int_0^{\pi} d\theta\cdot S\,(E,\ \cos\theta)\,(1-\cos\theta)\sin\theta. \quad (6.12)$$

Очевидно, функция $\tau\ (E)$ не отрицательна. В частности, для электронов с квадратичным законом дисперсии мы имеем

$$\frac{1}{\tau\,(E)} = \frac{\sqrt{2m^3\,(E-E_c)}}{(2\pi)^2\hbar^3}\int_0^{\pi} d\theta\cdot S\,(E,\ \cos\theta)\,(1-\cos\theta)\sin\theta. \quad (6.13)$$

Видим, что $\tau^{-1}\ (E)$ представляется в виде суммы двух слагаемых. Первое из них (отвечающее единице в круглых скобках под знаком интеграла) связано с интегралом «ухода» B (3.8), второе (отвечающее $\cos\theta$ в круглых скобках) — с интегралом «прихода» A (3.7). Легко убедиться, что «приходный» член не играет роли, если S — четная функция $\cos\vartheta$ (или не зависит от ϑ).

Функция $\tau\ (E)$, определяемая равенствами (6.12) или (6.13), имеет размерность времени и называется *временем релаксации импульса*. Для краткости мы будем опускать слово «импульса» всюду, где это не может повести к недоразумениям. Часто используется также термин «транспортное время релаксации».

Смысл термина «время релаксации» легко уяснить себе, рассмотрев частный случай, когда внешние поля отсутствуют и образец пространственно однороден ($\nabla f = 0$), но функция распределения все же имеет неравновесный вид (5.1). Так может обстоять дело, например, сразу после выключения (в момент $t = 0$) внешнего электрического поля. Естественно, такое состояние не будет стационарным и функция распределения будет изменяться со временем, постепенно приближаясь к равновесному своему значению. Действительно, кинетическое уравнение (3.12) в данном случае имеет вид

$$\frac{\partial f}{\partial t} + \frac{f - f_0}{\tau(E)} = 0. \tag{6.14}$$

Отсюда

$$f - f_0 = Ce^{-t/\tau(E)}. \tag{6.15}$$

Константа C определяется начальными условиями и нас здесь не интересует. Видно, что величина τ определяет время возвращения системы к состоянию равновесия.

Как будет видно из дальнейшего, время $\tau(E)$ есть не что иное, как время свободного пробега, введенное в § 2 гл. I. Здесь, однако, оно введено не ad hoc, а выведено — в определенных предположениях — из рассмотрения интеграла столкновений. При этом формулы (6.12) и (6.13) позволяют явно вычислять τ как функцию энергии носителя заряда и экспериментально варьируемых параметров, коль скоро известна функция $S(E, \cos\theta)$, т. е. коль скоро решена механическая часть задачи.

Следует подчеркнуть, что предположения, в рамках которых получены простые формулы (6.11) и (6.12), — довольно жесткие. Прежде всего, изоэнергетические поверхности далеко не всегда изотропны (§§ III.8, III.9). Наоборот, энергия электрона чаще зависит не только от абсолютной величины квазиимпульса, но и от ориентации его относительно осей кристалла. В частности, так обстоит дело в германии и кремнии. В таких случаях кинетические свойства системы электронов характеризуются «временем релаксации», зависящим не только от энергии, но и от направления движения частицы относительно кристаллографических осей. Простая формула (6.12) при этом уже несправедлива. Иногда удается ввести три «времени релаксации», каждое из которых зависит только от энергии и соответствует потоку электронов вдоль «своей» кристаллографической оси. Далее, как будет показано в гл. XIV, существуют механизмы рассеяния, для которых предположение об упругом его характере не оправдано.

Задача о решении кинетического уравнения с учетом анизотропии изоэнергетических поверхностей или неупругости рассеяния связана с серьезными математическими осложнениями, и мы здесь не будем на ней останавливаться. Приближенную аналитиче-

скую ее трактовку, справедливую в применении к кристаллам типа германия и кремния, можно найти в монографии [M7]. В более сложных случаях кинетическое уравнение решают численными методами.

Несмотря на известную грубость принятых нами предположений, формулы (6.11) и (6.12) все же оказываются очень полезными. В ряде случаев они позволяют правильно указать зависимость кинетических коэффициентов от экспериментально варьируемых параметров, а также оценить порядок их величины. Дело в том, что анизотропия изоэнергетических поверхностей не всегда играет принципиальную роль. Зачастую учет ее приводит просто к появлению не слишком существенных численных коэффициентов. Более подробно этот вопрос рассматривается в § 7 и в § XIV.6.

§ 7. Элементарные стационарные решения кинетического уравнения в случае малых отклонений от равновесия

Решение кинетического уравнения в условиях, когда справедливы неравенство (5.3) и соотношение (6.11), не представляет труда. В настоящем параграфе мы будем интересоваться только стационарными состояниями, в которых функция распределения f не зависит от времени. Это соответствует постоянным полям (или постоянному градиенту температуры), со времени включения которых прошло время, заметно превышающее как время свободного пробега, так и максвелловское время релаксации.

Вновь рассмотрим по отдельности три случая, указанные в § 6. Ограничимся при этом материалами, пространственно однородными в отсутствие внешних воздействий (в частности, в отсутствие градиента температуры).

а. Статическая электропроводность. В рассматриваемом случае нет причин для возникновения зависимости функции распределения от координат, $\mathfrak{B} = 0$ и $\mathfrak{E}' = \mathfrak{E}$.

Таким образом, кинетическое уравнение (5.4′) с учетом соотношений (6.6а) и (6.11) принимает вид

$$e(\mathfrak{E}, \mathbf{v}) f_0' - (\mathbf{p}, \mathfrak{E}) \frac{\psi}{\tau} = 0. \tag{7.1}$$

В изотропном случае (6.3) скорость $\mathbf{v}$ параллельна $\mathbf{p}$. Положим

$$\mathbf{p} = m(E)\,\mathbf{v}, \tag{7.2}$$

где $m(E)$ — некоторая величина размерности массы. При квадратичном законе дисперсии это есть просто эффективная масса носителя заряда; при учете непараболичности она зависит от энергии.

Уравнение (7.1) теперь принимает вид

$$\left(\frac{e}{m} f_0' - \frac{\psi}{\tau}\right)(\mathbf{p}, \mathfrak{E}) = 0. \tag{7.3}$$

Для краткости мы не выписываем здесь энергетические аргументы у функций m, ψ и τ.

Поскольку угол между векторами $\mathbf{p}$ и $\mathcal{E}$ может быть произвольным, коэффициент при скалярном произведении $(\mathbf{p}, \mathcal{E})$ должен обращаться в нуль. Отсюда

$$\psi = \frac{e}{m}\tau f_0'. \tag{7.4}$$

Поскольку τ и f_0 зависят только от E, видим, что и ψ, как мы и предполагали, действительно зависит только от энергии, но не от направления квазиимпульса. Для функции распределения мы имеем теперь, согласно (7.4), (6.6а) и (5.1),

$$f(\mathbf{p}) = f_0(E) + \frac{e}{m}\tau(\mathbf{p}, \mathcal{E})f_0'. \tag{7.5}$$

Теперь легко раскрыть содержание неравенства (5.3) применительно к рассматриваемой задаче. Для этой цели надо заменить скалярное произведение $(\mathbf{p}, \mathcal{E})$ в (7.5) произведением абсолютных значений и потребовать, чтобы второе слагаемое в (7.5) было по абсолютной величине много меньше первого при всех существенных значениях p и E. Это приводит к следующему неравенству:

$$\frac{e\tau p\mathcal{E}}{mkT} \ll 1. \tag{7.6}$$

В случае невырожденного газа главную роль играют «тепловые» значения

$$p \simeq p_T \equiv \sqrt{mkT}, \qquad v \simeq v_T \equiv \sqrt{kT/m}. \tag{7.7}$$

Таким образом, условие (7.6) принимает вид

$$\frac{e\mathcal{E}}{m}\tau p_T \ll kT, \tag{7.8}$$

причем под τ следует понимать значение функции $\tau(E)$, взятое при $E \simeq kT$. Левая часть (7.8) по порядку величины представляет собой не что иное, как энергию, получаемую электроном (дыркой) от поля за время свободного пробега. Видно, что она должна быть пренебрежимо мала по сравнению со средней энергией носителя заряда, уже имеющейся у него в условиях термодинамического равновесия.

С помощью второй из формул (7.7) условие (7.8) удобно переписать в виде

$$\frac{e\mathcal{E}\tau}{m} \equiv v_d \ll v_T \tag{7.9}$$

или

$$\mathcal{E} \ll \mathcal{E}_{\text{крит}} = \frac{mv_T}{e\tau}. \tag{7.9'}$$

Таким образом, принятое выше приближение оправдано, коль скоро дрейфовая скорость достаточно мала по сравнению с тепловой.

В случае полностью вырожденного газа главную роль играет область энергий $E \simeq \zeta$, где ζ — уровень Ферми, отсчитанный от края зоны, а квазиимпульс и скорость следует заменить «фермиевскими» их значениями $p_\zeta \equiv p(\zeta)$ и $v_\zeta \equiv v(\zeta)$. Вместо неравенства (7.6) при этом получается

$$\frac{e\mathcal{E}}{m}\tau p_\zeta \ll \zeta, \tag{7.6'}$$

причем $\tau = \tau(\zeta)$.

Иначе говоря, мы вновь получаем здесь условие вида (7.9):

$$v_d \ll v_\zeta,$$

или

$$\mathcal{E} \ll \mathcal{E}_{\text{крит}} = \frac{mv_\zeta}{e\tau}. \tag{7.9''}$$

При $T = 77$ К, $m = 0{,}1\, m_0$ и $\tau = 10^{-13}$ с величина $\mathcal{E}_{\text{крит}}$ в невырожденном случае составляет примерно 10^4 В/см.

Неравенства (7.9'), (7.9'') определяют условия применимости линейного (по $\mathcal{E}$) приближения, принятого выше при решении кинетического уравнения. Как будет видно из дальнейшего, они же составляют и условия применимости закона Ома.

Согласно программе, намеченной в § 2, функцию распределения (7.5) надо подставить в формулу для плотности тока (2.2) и, вычислив интеграл, получить явное выражение для $\mathbf{j}$. Найденный таким путем коэффициент пропорциональности между $\mathbf{j}$ и $\mathcal{E}$ следует отождествить с электропроводностью σ. Мы имеем

$$\mathbf{j} = -\frac{2e^2}{(2\pi\hbar)^3}\int \frac{(\mathbf{p}, \mathcal{E})}{m^2(E)}\,\mathbf{p}\tau f_0'\, d\mathbf{p}. \tag{7.10}$$

Как и при выводе формулы (6.11), здесь удобно ввести сферические координаты, направив полярную ось вдоль вектора $\mathcal{E}$. Из соображений симметрии ясно (и легко проверить непосредственно), что компоненты $\mathbf{j}$, перпендикулярные $\mathcal{E}$, равны нулю. Вновь вводя вместо p переменную E, отсчитываемую от E_c, мы получаем с учетом (7.2)

$$\mathbf{j} = -\frac{e^2\mathcal{E}}{3}\int_0^\infty N(E)\,\tau(E)\,v^2(E)\,f_0'\, dE. \tag{7.11}$$

Сравнивая соотношения (7.11) и (I.1.3), находим

$$\sigma = -\frac{e^2}{3}\int_0^\infty N(E)\,\tau(E)\,v^2(E)\,f_0'\, dE. \tag{7.12}$$

Эта формула выражает электропроводность вещества через величину $\tau(E)$, характеризующую решение механической части задачи. Концентрация носителей заряда входит сюда через уровень Ферми, фигурирующий в равновесной функции распределения f_0. Согласно (V.3.1) $f_0' < 0$ и, следовательно, $\sigma > 0$, как и должно быть. Дрейфовая подвижность μ определяется равенством

$$\sigma = en\mu, \tag{7.13}$$

причем концентрация носителей заряда n дается выражением (V.4.1). Комбинируя равенства (7.12), (7.13) и (V.4.1), находим

$$\mu = -\frac{e}{3}\frac{\int_0^\infty \tau v^2 N f_0'\, dE}{\int_0^\infty N f_0\, dE}. \tag{7.14}$$

Это есть не что иное, как формула (I.3.13), с той лишь разницей, что здесь указан точный смысл усреднения, обозначенного там символом $\langle \ldots \rangle$. Именно, определим *эффективную массу электропроводности* m_σ *) соотношением

$$m_\sigma = -\frac{3\int_0^\infty N(E) f_0(E)\, dE}{\int_0^\infty N(E) v^2(E) f_0'\, dE} \equiv m_{\text{opt}}. \tag{7.15}$$

Тогда формулу (7.14) можно переписать в виде

$$\mu = \frac{e\tau_p}{m_\sigma}, \tag{7.14'}$$

где

$$\tau_p \equiv \langle \tau \rangle = \frac{\int_0^\infty f_0 \frac{d}{dE}(\tau v^2 N)\, dE}{\int_0^\infty f_0 \frac{d}{dE}(v^2 N)\, dE} = \frac{\int_0^\infty f_0' \tau v^2 N\, dE}{\int_0^\infty f_0' v^2 N\, dE}. \tag{7.16}$$

Как будет видно из дальнейшего, правило усреднения (7.16) справедливо и в применении к формулам, описывающим термоэдс, постоянную Холла и т. д. Величина m_σ зависит от формы изоэнергетических поверхностей и, вообще говоря, от температуры и от концентрации частиц. В случае изотропного параболического закона

*) Величину m_σ называют также оптической эффективной массой m_{opt}, с чем и связано другое ее обозначение (см. § 8 и гл. XVIII).

дисперсии она превращается в эффективную массу m, фигурирующую в законе дисперсии (III.8.4).

Отметим два предельных случая — полностью вырожденного и невырожденного газа носителей заряда. Первый из них реализуется при $\zeta \gg kT$; как видно из формулы (V.3.1), величина $|f_0'|$ тогда имеет острый максимум при $E = \zeta$ и для вычисления интегралов с гладкими функциями ее можно заменить δ-функцией:

$$f_0' \to -\delta(E-\zeta) \tag{7.17}$$

(см. Приложение XV).

При этом интегралы в (7.16) сразу вычисляются, и мы получаем

$$\tau_p = \tau(\zeta). \tag{7.18}$$

Заметим, что при любом механизме рассеяния подвижность здесь может зависеть от концентрации частиц. Это обусловлено концентрационной зависимостью энергии Ферми ζ, т. е. в конечном счете принципом Паули.

В случае невырожденного газа, когда $f_0 \simeq \exp\left(\frac{\zeta - E}{kT}\right)$, формула (7.14) принимает вид

$$\mu = \frac{e}{3kT} \frac{\int\limits_0^\infty \tau v^2 N \exp\left(-\frac{E}{kT}\right) dE}{\int\limits_0^\infty N \exp\left(-\frac{E}{kT}\right) dE}. \tag{7.19}$$

Поскольку принцип Паули в отсутствие вырождения роли не играет, это выражение может зависеть от концентрации частиц n только через время релаксации $\tau(E)$: функция S, фигурирующая в формуле (6.12) и получающаяся из решения механической части задачи, может зависеть от n, если играет роль взаимодействие электронов друг с другом.

При параболическом законе дисперсии мы имеем согласно (V.2.3)

$$N(E) = \frac{(2m)^{3/2} E^{1/2}}{2\pi^2 \hbar^3}.$$

Следовательно, интеграл в знаменателе (7.19) равен

$$\frac{(2m^3)^{1/2} (kT)^{3/2}}{2\pi^{3/2} \hbar^3}$$

и мы получаем, заменяя в числителе v^2 на $2E/m$,

$$\mu = \frac{4e}{3m (kT)^{5/2} \sqrt{\pi}} \int\limits_0^\infty E^{3/2} \tau \exp\left(-\frac{E}{kT}\right) dE. \tag{7.20}$$

Для дальнейшего нужно знать явную зависимость времени релаксации от энергии. Как будет показано в гл. XIV, весьма часто эта зависимость оказывается степенной:

$$\tau(E) = CE^r, \tag{7.21}$$

где величина C не зависит от энергии, а r — некоторая постоянная (значения ее для ряда механизмов рассеяния указаны в таблице 14.2). Подставляя выражение (7.21) в правую часть (7.20), мы получаем

$$\mu = \frac{4e\Gamma(r+5/2)\,C}{3m\sqrt{\pi}}(kT)^r. \tag{7.20'}$$

Здесь $\Gamma(r+5/2)$ есть Г-функция Эйлера, определяемая равенством

$$\Gamma(x) = \int_0^\infty y^{x-1}e^{-y}\,dy.$$

б. Термоэдс и коэффициент Пельтье. В соответствии с § 1 (п. б) рассмотрим случай, когда внешние электрическое и магнитное поля отсутствуют, но $\nabla T \neq 0$ и имеется поле $\mathcal{E}'$ (§ 1 и § 6 (п. б)). При этом уравнение (5.4′) с учетом (6.11) принимает вид

$$\left\{(e\mathcal{E}', \mathbf{v}) + (\mathbf{v}, \nabla T)\frac{E(\mathbf{p})-\zeta}{T}\right\} f_0' = \frac{f_1}{\tau}. \tag{7.22}$$

Подставляя сюда выражение (6.6б) для f_1 и принимая во внимание соотношение (7.2), мы получаем

$$\chi_1 = \tau\frac{E(\mathbf{p})-\zeta}{mT}f_0', \qquad \chi_2 = \frac{e\tau}{m}f_0'. \tag{7.23}$$

Как мы и предполагали, χ_1 и χ_2 зависят только от энергии электрона. Функция χ_2 имеет, как и следовало ожидать, тот же вид (7.4), что и функция ψ в п. а; соответствующий вклад в плотность тока дается выражением (7.11). Для функции распределения находим согласно (5.1)

$$f = f_0 + \frac{E(\mathbf{p})-\zeta}{mT}\tau f_0'\cdot(\mathbf{p}, \nabla T) + (\mathbf{p}, \mathcal{E}')\frac{e\tau}{m}f_0'. \tag{7.24}$$

Подставляя выражение (7.24) в формулу (2.2) для плотности тока и выполняя интегрирование так же, при выводе формулы (7.11), получим для плотности тока электронов

$$\mathbf{j} = \frac{e}{3}\nabla T\int_0^\infty v^2(E)\,N(E)\frac{\zeta-E}{T}f_0'\tau\,dE + \ldots, \tag{7.25}$$

где многоточием обозначено слагаемое, содержащее $(\mathbf{p}, \mathcal{E}')$; оно имеет вид (7.11) с заменой $\mathcal{E}$ на $\mathcal{E}'$.

Величина ζ в (7.25), как и всегда, отсчитывается от края зоны проводимости. При этом, поскольку перед интегралом уже имеется

множитель ∇T, а градиент электрического потенциала в данном случае отличен от нуля лишь за счет градиента температуры, учитывать искривление зон при вычислении ζ не следует. Сравнивая (7.25) с равенством (1.2), находим фигурирующий там феноменологически введенный коэффициент a и, по формулам (1.6) и (7.12), дифференциальную термоэдс α:

$$\alpha = -\frac{k}{e}\left\{-\frac{\zeta}{kT}+\frac{1}{kT}\frac{\int\limits_0^\infty Ev^2N\tau f_0'\,dE}{\int\limits_0^\infty v^2N\tau f_0'\,dE}\right\}. \tag{7.26}$$

Аналогичная формула справедлива и для тока дырок — с очевидной заменой скорости, плотности состояний, функции распределения и химического потенциала на соответствующие «дырочные» величины. Это означает, в частности, что ζ надо заменить на $-E_g - \zeta$ (ср. § V.4). При параболическом законе дисперсии, когда $v^2 = 2E/m$, выражение (7.26) с учетом (V.2.3) принимает вид

$$\alpha = -\frac{k}{e}\left\{-\frac{\zeta}{kT}+\frac{1}{kT}\frac{\int\limits_0^\infty E^{5/2}\tau f_0'\,dE}{\int\limits_0^\infty E^{3/2}\tau f_0'\,dE}\right\}. \tag{7.27}$$

Это — так называемая формула Писаренко.

Первое слагаемое в фигурных скобках зависит только от равновесных характеристик системы, второе же определяется и механизмом рассеяния.

Рассмотрим формулу (7.27) в частных случаях невырожденного и полностью вырожденного газа носителей заряда.

В первом из них мы получаем

$$\alpha = -\frac{k}{e}\left\{-\frac{\zeta}{kT}+\frac{1}{kT}\frac{\int\limits_0^\infty E^{5/2}\tau\exp(-E/kT)\,dE}{\int\limits_0^\infty E^{3/2}\tau\exp(-E/kT)\,dE}\right\}. \tag{7.28}$$

Второе слагаемое в фигурных скобках есть некоторое безразмерное число, зависящее от вида функции $\tau(E)$. Обычно оно — порядка единицы. Так, в случае (7.21) это отношение составляет

$$\frac{\Gamma(r+7/2)}{\Gamma(r+5/2)} = r+5/2,$$

т. е.

$$\alpha = -\frac{k}{e}\left\{-\frac{\zeta}{kT}+(r+5/2)\right\}. \tag{7.28'}$$

Выражения в фигурных скобках формул (7.28), (7.28′), очевидно, положительны: в отсутствие вырождения $\zeta < 0$.

С другой стороны, обращаясь к дырочному образцу, мы должны были бы, в соответствии с § IV.2, изменить знак перед k/e. Иначе говоря, для дырок мы имеем

$$\alpha = \frac{k}{e}\left\{\frac{\zeta + E_g}{kT} + (r + {}^5/_2)\right\}. \qquad (7.28'')$$

Выражение в фигурных скобках вновь положительно.

Таким образом, знак дифференциальной термоэдс определяется знаком заряда доминирующих носителей.

В случае полного вырождения аппроксимация (7.17) оказывается недостаточной. Действительно, полагая

$$f_0' = -\delta(E - \zeta),$$

мы получили бы из (7.27)

$$\alpha = -\frac{k}{e}\left\{-\frac{\zeta}{kT} + \frac{1}{kT}\frac{\zeta^{3/2}\tau(\zeta)}{\zeta^{3/2}\tau(\zeta)}\right\} \equiv 0.$$

Следовательно, для вычисления интегралов в (7.27) необходимо воспользоваться более точной, нежели (7.17), формулой (П.XV.5). Тогда получается

$$\alpha = -\frac{\pi^2 k^2 T}{3e\zeta}\tau(\zeta)\left[\frac{3}{2} + \frac{d\ln\tau(E)}{d\ln E}\Big|_{E=\zeta}\right]. \qquad (7.28''')$$

В качестве ζ здесь следует взять значение уровня Ферми при $T = 0$ (отсчитанное от дна зоны проводимости).

Выражение в квадратных скобках в формуле (7.28‴) представляет собой число порядка единицы. Так, в условиях (7.21) оно равно

$${}^3/_2 + r.$$

Сравнивая формулы (7.28′) и (7.28‴), видим, что в невырожденных полупроводниках термоэдс значительно больше, нежели в металлах. Действительно, вместо фигурирующего в (7.28′) отношения ζ/kT в правой части (7.28‴) появляется малый множитель kT/ζ (в металлах он составляет величину порядка 10^{-3}).

Вычислим теперь плотность потока энергии. Ограничимся при этом случаем невырожденного газа носителей заряда. Подставляя (7.24) в правую часть (2.3), мы получаем

$$\mathbf{I} = \frac{2}{(2\pi\hbar)^3}\int \mathbf{v}(\mathbf{p})\,\tau\,\frac{E(\mathbf{p}) - \zeta}{mT}(\mathbf{p},\ \nabla T)[E(\mathbf{p}) - e\varphi]f_0'\,d\mathbf{p} + \\ + \frac{2e}{(2\pi\hbar)^3 m}\int \mathbf{v}(\mathbf{p})[E(\mathbf{p}) - e\varphi]\,\tau\,(\mathbf{p}\mathcal{E}')f_0'\,d\mathbf{p}. \qquad (7.29)$$

Интегралы по квазиимпульсам преобразуются здесь так же, как и в п. а. Сравнивая выражения (7.25) и (7.29), видим, что слагаемые

в (7.29), содержащие $e\varphi$, в сумме равны просто $\mathbf{j}\varphi$. Пользуясь для τ формулой (7.21), мы получаем

$$\mathbf{I} = \mathbf{j}\varphi + \mathbf{I}' - \frac{2enC\Gamma(r+{}^7/_2)}{3m\Gamma({}^3/_2)}(kT)^{r+1}\mathcal{E}', \tag{7.29'}$$

где

$$\mathbf{I}' = \frac{4nC(k\tilde{T})^{r+1}\Gamma(r+{}^7/_2)k}{3m\sqrt{\pi}}\left\{-\nabla T - \frac{e}{k}\mathcal{E}' + \left[\frac{\zeta}{kT} - ({}^5/_2+r)\right]\nabla T\right\}. \tag{7.30}$$

Слагаемое, заключенное в (7.30) в квадратные скобки, есть не что инсе, как $e\alpha/k$. Величину $\mathcal{E}'$ можно выразить через $\mathbf{j}$ и ∇T с помощью соотношения (1.7). Обращаясь, далее, к формулам (7.13) и (7.20'), видим, что вклад в $\mathbf{I}$ от второго и третьего слагаемых в фигурных скобках в (7.30) и от последнего слагаемого в (7.29) можно записать в виде

$$-\frac{kT}{e}({}^5/_2+r)\mathbf{j} = -\frac{kT}{e}\left[({}^5/_2+r) - \frac{\zeta}{kT}\right]\mathbf{j} - \frac{\zeta}{e}\mathbf{j} = -\frac{\zeta}{e}\mathbf{j} + T\alpha\mathbf{j}. \tag{7.31}$$

Отсюда явствует, что первое слагаемое в фигурных скобках в (7.30) связано с плотностью потока энергии при $\mathbf{j} = 0$, т. е. с теплопроводностью электронного газа. Сопоставив формулы (7.31), (7.13) и (7.20'), найдем коэффициент теплопроводности газа носителей заряда:

$$\varkappa = \frac{4nkC(kT)^{r+1}\Gamma(r+{}^7/_2)}{3m\sqrt{\pi}} = \frac{k^2T\sigma}{e^2}(r+{}^5/_2). \tag{7.32}$$

Комбинируя теперь выражения (7.29') и (7.30) — (7.32), получаем окончательную формулу для плотности потока энергии электронов:

$$\mathbf{I} = -\frac{\zeta}{e}\mathbf{j} - \varkappa\nabla T + T\alpha\mathbf{j}. \tag{7.33}$$

Сравнивая ее с (1.8) (при выбранном начале отсчета энергии $\zeta = F$), получаем соотношение (I.1.18):

$$\Pi = \alpha T.$$

Заметим, однако, что справедливость последнего равенства не связана непременно с теми упрощающими предположениями, которые мы сделали при выводе формулы (7.33). Фактически оно вытекает из очень общих соотношений термодинамики неравновесных процессов *).

Раскроем теперь условие (5.3) применительно к рассматриваемой задаче. Ограничимся при этом случаем невырожденного полупроводника. Пользуясь формулами (7.2) и (7.23), приходим к

*) Общий вывод термоэлектрических соотношений (I.1.18) и (I.1.19) можно найти в книге [M13].

следующему неравенству:

$$v_T\tau\frac{|\nabla T|}{T}\left|\frac{\zeta}{kT}-1\right|\ll 1. \tag{7.34}$$

Произведение

$$v_T\tau = l \tag{7.35}$$

называют *длиной свободного пробега* (по импульсу). Видим, что для выполнения неравенства (7.34) необходимо (но, вообще говоря, еще недостаточно), чтобы изменение температуры на длине свободного пробега было мало по сравнению с самой температурой. Последнее условие, впрочем, имеет гораздо более общий характер: при нарушении его вообще нельзя ввести представление о температуре, изменяющейся в пространстве. Действительно, понятие температуры вводится в термодинамике равновесных процессов, причем одно из условий равновесия как раз и состоит в постоянстве температуры во всей рассматриваемой системе. Обобщить определение температуры на неравновесный случай можно с помощью так называемой *гипотезы локального равновесия*. Именно, допустим, что рассматриваемое макроскопическое тело можно разделить на ряд физически бесконечно малых объемов — достаточно малых областей, каждая из которых содержит все же макроскопически много частиц (при этом линейные размеры каждой области должны быть велики по сравнению с длиной свободного пробега). Эти области можно в свою очередь рассматривать как макроскопические системы, описывая их состояние с помощью обычных понятий термодинамики. Естественно думать, что равновесие устанавливается прежде всего внутри каждой области (локально); лишь спустя известное время после этого может установиться и равновесие между областями. Мы имеем здесь еще один пример разделения процессов на быстрые и медленные (ср. § 3). Установление равновесия внутри каждой физически бесконечно малой части тела представляет собой быстрый процесс (происходящий за время порядка времени релаксации); установление же равновесия во всем объеме может быть связано с гораздо более медленными макроскопическими процессами теплопроводности, диффузии и т. д. Таким образом, оказывается возможным говорить о локальном равновесии в пределах каждого физически бесконечно малого объема при отсутствии равновесия во всей системе в целом. При этом состояние каждого малого объема можно характеризовать своей (постоянной в его пределах) температурой, своей концентрацией частиц и т. д. Тем самым приобретает ясный смысл и представление об изменяющейся в пространстве температуре. Соображения такого типа постоянно используются в задачах кинетики. В частности, они составляют основу обычной теории теплопроводности, диффузии и т. д. Тем не менее справедливость их заранее не очевидна. Действительно, как уже отмечалось, «физически бесконечно малые объемы» должны быть, с одной стороны,

достаточно велики, а с другой — достаточно малы, и эти два условия должны быть совместны. Именно по этой причине мы употребляем термин «гипотеза локального равновесия», понимая под этим предположение о возможности разбить макроскопическое тело на области указанным выше образом.

в. Постоянная Холла и магнетосопротивление. Как и в случае одного лишь постоянного и однородного электрического поля, здесь нет причин для возникновения координатной зависимости функции распределения и $\mathfrak{E}' = \mathfrak{E}$. Таким образом, уравнение (5.4′), написанное для электронов, принимает вид

$$-e(\mathfrak{E}, \mathbf{v}) f_0' - \frac{e}{c}([\mathbf{v}\times\mathfrak{B}], \nabla_p f_1) = \frac{f_1}{\tau}. \tag{7.36}$$

При этом функция f_1 дается выражением (6.6в).

Преобразуем теперь второе слагаемое в левой части (7.36). Для этой цели воспользуемся известным из векторной алгебры равенством, справедливым для произвольных векторов **a**, **b**, **c** и **d**:

$$([\mathbf{a}\times\mathbf{b}], [\mathbf{c}\times\mathbf{d}]) = (\mathbf{a}, \mathbf{c})(\mathbf{b}, \mathbf{d}) - (\mathbf{a}, \mathbf{d})(\mathbf{b}, \mathbf{c}).$$

Получим, с учетом (6.6в),

$$([\mathbf{v}\times\mathfrak{B}], \nabla_p f_1) = (\mathbf{v}, [\mathfrak{B}\times\mathfrak{E}])\psi_1 + (\mathbf{v}, \mathfrak{E})\mathscr{B}^2\psi_2 - (\mathbf{v}, \mathfrak{B})(\mathfrak{B}, \mathfrak{E})\psi_2. \tag{7.37}$$

Выражение (7.37), вообще говоря, отлично от нуля. Введем по-прежнему массу m (E) равенством (7.2). Тогда, подставляя (7.37) в уравнение (7.36), мы получим

$$\left\{\left(\frac{e}{m}f_0' - \frac{1}{\tau}\right)\psi_1 + \frac{e}{mc}\mathscr{B}^2\psi_2\right\}(\mathbf{p}, \mathfrak{E}) - \left\{\frac{e}{mc}\psi_1 + \frac{\psi_2}{\tau}\right\}(\mathbf{p}, [\mathfrak{E}\times\mathfrak{B}]) - \\ - \left\{\frac{e}{mc}(\mathfrak{B}, \mathfrak{E})\psi_2 + \frac{\psi_3}{\tau}\right\}(\mathbf{p}, \mathfrak{B}) = 0. \tag{7.38}$$

Поскольку углы между векторами **p**, $\mathfrak{E}$ и $\mathfrak{B}$ произвольны, равенство (7.38) влечет за собой условие обращения в нуль порознь всех трех выражений в фигурных скобках — коэффициентов при скалярных произведениях (**p**, $\mathfrak{E}$), (**p**, [$\mathfrak{E}\times\mathfrak{B}$]) и (**p**, $\mathfrak{B}$). Это дает три линейных уравнения для определения трех искомых функций ψ_1, ψ_2 и ψ_3. Решая их, находим

$$\psi_1 = -\frac{e\tau}{m(1+\omega_c^2\tau^2)}f_0', \tag{7.39а}$$

$$\psi_2 = -\frac{1}{c}\frac{e^2\tau}{m^2(1+\omega_c^2\tau^2)}f_0', \tag{7.39б}$$

$$\psi_3 = -(\mathfrak{B}, \mathfrak{E})\frac{1}{c^2}\frac{e^3\tau}{m^3(1+\omega_c^2\tau^2)}f_0'. \tag{7.39в}$$

Здесь

$$\omega_c(E) = \frac{\mathscr{B}e}{m(E)c}. \tag{7.40}$$

При параболическом законе дисперсии, когда функция $m(E)$ превращается в константу, это есть уже известная нам (§ IV.3), циклотронная частота. Произведение $\omega_c\tau$ определяет дугу, описываемую электроном (дыркой) за время свободного пробега τ. Это позволяет наглядно интерпретировать сомножитель $\tau(1 + \omega_c^2\tau^2)^{-1}$, фигурирующий во всех трех формулах (7.39а — в). Как мы видели в гл. I, во время между столкновениями частица в магнитном поле движется не по прямой, а по циклоиде; в то же время вклад в ток определяется лишь проекцией перемещения частицы на направление плотности тока. Таким образом, магнитное поле действует в известном смысле как «рассеиватель», отклоняя частицы от движения вдоль линий тока. Знаменатель $(1 + \omega_c^2\tau^2)$ как раз и учитывает это отклонение, приводя к уменьшению «эффективного времени свободного пробега» $\tau(1 + \omega_c^2\tau^2)^{-1}$ по сравнению с системой без магнитного поля. При исчезновении магнитного поля ($\mathscr{B} \to 0$) решение (6.6в) (7.39а — в), разумеется, переходит в (6.6а), (7.4).

Подставляя теперь выражение (6.6в) в формулу (2.2) для плотности тока и выполняя интегрирование так же, как и в п. *а*, получим выражение вида (1.15), но с явно вычисленными коэффициентами при векторах $(\mathfrak{E}, [\mathfrak{E} \times \mathfrak{B}])$ и $\mathfrak{B}(\mathfrak{B}, \mathfrak{E})$. Эти коэффициенты и следует отождествить с феноменологически введенными там величинами a_1, a_2, a_3.

Ограничимся случаем параболического закона дисперсии. При подстановке (6.6в) с учетом (7.39а — в) в формулу (2.2) появятся интегралы

$$\mathscr{J}_k = -\int_0^\infty E^{3/2} f_0' \frac{\tau^k(E)}{1+\omega_c^2\tau^2}\, dE, \tag{7.41}$$

где k — целые числа.

Получим для коэффициентов a_1, a_2, a_3

$$a_1 = \frac{2e^2\sqrt{2m}}{3\pi^2\hbar^3}\mathscr{J}_1, \quad a_2 = \frac{2e^2\sqrt{2m}}{3\pi^2\hbar^3}\frac{e}{mc}\mathscr{J}_2, \quad a_3 = \frac{2e^2\sqrt{2m}}{3\pi^2\hbar^3}\frac{e^2}{m^2c^2}\mathscr{J}_3. \tag{7.42}$$

В случае полного вырождения

$$\mathscr{J}_k = \frac{\zeta^{3/2}\tau^k(\zeta)}{1+\omega_c^2(\zeta)\tau^2(\zeta)}. \tag{7.41'}$$

В невырожденном случае интегралы $\mathscr{J}_k$ довольно сложным образом зависят от магнитной индукции и от температуры. Явный вид их можно найти, лишь решив механическую часть задачи.

Исключение составляют только случаи слабого и сильного магнитных полей, определяемые, соответственно, неравенствами

$$\omega_c^2\tau^2(E) \ll 1 \quad \text{и} \quad \omega_c^2\tau^2(E) \gg 1. \tag{7.43}$$

В первом из них мы получаем

$$\mathcal{J}_k \simeq -\int_0^\infty E^{3/2}\,\tau^k(E)\,f_0'\,(1-\omega_c^2\tau^2)\,dE, \tag{7.44а}$$

а во втором

$$\mathcal{J}_k \simeq -\frac{1}{\omega_c^2}\int_0^\infty E^{3/2}\,\tau^{k-2}(E)\,f_0'\,dE. \tag{7.44б}$$

Отметим, что при $k=2$ время релаксации выпадает из (7.44б) и мы получаем

$$\mathcal{J}_2 = \frac{3\,(kT)^{3/2}\sqrt{\pi}}{4\omega_c^2}\,\Phi_{1/2}\left(\frac{\zeta}{kT}\right). \tag{7.44б'}$$

Подставляя выражения (7.44а) при $k=1,\ 2,\ 3$ в правые части (7.42), получим для плотности тока в слабом поле выражение вида (1.15′), причем

$$\delta = -\beta = \frac{2e^2\sqrt{2m}}{3\pi^2\hbar^3}\,\frac{e^2}{m^2c^2}\int_0^\infty E^{3/2}\,\tau f_0'\,dE, \qquad \gamma = 0. \tag{7.45}$$

Последнее равенство явствует уже из соображений симметрии изложенных в § 1.

Формулы (7.42) позволяют, в частности, вычислить постоянную Холла и магнетосопротивление для двух специальных ориентаций магнитного поля, рассмотренных в § I.3. Рассмотрим эти два случая по отдельности.

1) Поперечное поле: $\mathfrak{B} \perp \mathbf{j}$. Согласно (1.19) и (7.42) выражение для постоянной Холла имеет вид

$$R = -\frac{3\pi^2\hbar^3}{2ec\sqrt{2m^3}}\,\frac{\mathcal{J}_2}{\mathcal{J}_1^2+\omega_c^2\mathcal{J}_2^2}.$$

Удобно явно ввести сюда концентрацию носителей заряда n, пользуясь формулой (V.4.4).

Тогда для постоянной Холла получается формула вида (I.3.17):

$$R = -\frac{\gamma}{nec}, \tag{7.46}$$

причем безразмерный коэффициент γ дается равенством

$$\gamma = \frac{3\sqrt{\pi}}{4}\,\Phi_{1/2}\left(\frac{\zeta}{kT}\right)(kT)^{3/2}\,\frac{\mathcal{J}_2}{\mathcal{J}_1^2+\omega_c^2\mathcal{J}_2^2}. \tag{7.47}$$

Он зависит от степени вырождения и от вида функции $\tau(E)$, т. е. от механизма рассеяния.

В случае полного вырождения правая часть (7.47) легко вычисляется: согласно § V.6 и (7.41′) в указанных условиях

$$\gamma_{\text{вырожд}} = 1. \tag{7.47′}$$

Этот результат справедлив при любом механизме рассеяния и при любом значении магнитной индукции.

То же значение γ получается и при любой степени вырождения, если магнитное поле сильное. Действительно, согласно (7.44б) и (7.44б′) мы имеем в этом случае

$$\gamma_{\text{сильн. поле}} = 1 + O\left(\frac{1}{\omega_c^2\tau^2}\right). \tag{7.47″}$$

Как можно показать *), этот результат справедлив и при учете квантовых эффектов, если только выполняется второе из неравенств (7.43). Это обстоятельство особенно важно потому, что, как мы знаем (§ IV.5), именно в области сильных магнитных полей квантовые эффекты могут играть заметную роль.

С другой стороны, в слабом поле мы имеем согласно (7.44а)

$$\gamma_{\text{слаб. поле}} = \frac{3\sqrt{\pi}}{4}\Phi_{1/2}\left(\frac{\zeta}{kT}\right)(kT)^{3/2}\frac{\mathcal{J}_2^{(0)}}{(\mathcal{J}_1^{(0)})^2}, \tag{7.48}$$

где

$$\mathcal{J}_k^{(0)} = -\int_0^\infty E^{3/2}\tau^k f_0'\,dE.$$

В частности, для невырожденной системы формула (7.48) принимает вид

$$\gamma_{\text{слаб. поле, невырожд}} = \frac{3\sqrt{\pi}\,(kT)^{5/2}}{4}\frac{\int_0^\infty E^{3/2}\tau^2\exp(-E/kT)\,dE}{\left[\int_0^\infty E^{3/2}\tau\exp(-E/kT)\,dE\right]^2}. \tag{7.48′}$$

Как и в задаче о подвижности, мы получили формулу того же вида (I.3.19), что и в элементарной кинетической теории. В случае степенной зависимости τ от энергии (7.21) формула (7.48′) принимает вид

$$\gamma_{\text{слаб. поле, невырожд}} = \frac{3\sqrt{\pi}\,\Gamma(2r+5/2)}{4\Gamma^2(r+5/2)}. \tag{7.48″}$$

Граница между сильными и слабыми полями определяется характерным значением индукции

$$\mathscr{B}_{\text{крит}} = \frac{mc}{e\tau}. \tag{7.49}$$

Полагая для оценки $\tau = 10^{-13}$ с, мы имеем

$$\mathscr{B}_{\text{крит}} = 6\cdot 10^5\frac{m}{m_0}\ \text{Гс}, \tag{7.49′}$$

*) См. [1] (статья № 10).

где m_0, как всегда, есть масса свободного электрона. Согласно (7.43) магнитное поле называется, соответственно, сильным или слабым, если $\mathscr{B} \gg \mathscr{B}_{\text{крит}}$ или $\mathscr{B} \ll \mathscr{B}_{\text{крит}}$. Видно, что в веществах с малыми эффективными массами сильные в указанном смысле поля могут быть не так уж велики. Так, при $m/m_0 = 0{,}01$ получаем $\mathscr{B}_{\text{крит}} = 6000$ Гс.

Заметим, однако, что, как уже отмечалось в гл. IV, в достаточно сильном магнитном поле может оказаться несостоятельной вся схема расчета, основанная на кинетическом уравнении. Действительно, как мы видели в § IV.5, сильное магнитное поле радикально меняет энергетический спектр носителя заряда. Пренебречь этими квантовыми эффектами можно, лишь если выполняется неравенство (IV.5.19).

Таким образом, рассматривая влияние магнитного поля на кинетические характеристики и энергетический спектр носителей заряда, следует различать три случая:

1) классические слабые поля: $\mathscr{B} \ll \mathscr{B}_{\text{крит}}$, $\dfrac{e\hbar\mathscr{B}}{mc} < \bar{E}$;

2) классические сильные поля: $\mathscr{B} \gg \mathscr{B}_{\text{крит}}$, $\dfrac{e\hbar\mathscr{B}}{mc} < \bar{E}$;

3) квантующие поля: $\dfrac{e\hbar\mathscr{B}}{mc} > \bar{E}$.

Здесь $\bar{E}$ есть характерная энергия носителей заряда. В отсутствие вырождения $\bar{E} \simeq kT$, в условиях полного вырождения $\bar{E} = \zeta$.

Кинетические коэффициенты, описывающие поведение вещества в квантующих полях, вычисляют методами квантовой теории необратимых процессов (см., например, книгу [1]).

Поперечное магнетосопротивление $\dfrac{\rho_\perp - \rho_0}{\rho_0}$, согласно формулам (1.20), (7.12) и (7.42), дается равенством

$$\frac{\rho_\perp - \rho_0}{\rho_0} = \frac{\mathscr{J}_1 \mathscr{J}_1^{(0)}}{\mathscr{J}_1^2 + \omega_c^2 \mathscr{J}_2^2} - 1. \tag{7.50}$$

Подставляя сюда выражения (7.41) и (7.44б), видим, что формула (7.50), как и следовало ожидать, воспроизводит результат элементарной теории (I.3.21).

С помощью формулы (7.41′) легко убедиться, что в случае полного вырождения правая часть (7.50) тождественно обращается в нуль — магнетосопротивление отсутствует. Как можно показать [М7], этот результат связан с использованной нами идеализацией — пренебрежением анизотропией изоэнергетических поверхностей. При учете анизотропии магнетосопротивление вырожденного электронного газа оказывается, вообще говоря, конечным.

Конечное значение магнетосопротивления получается также, если вырождение неполное и время релаксации зависит от энергии. Зависимость магнетосопротивления от величины магнитной

индукции при этом оказывается довольно сложной, определяясь видом функции $\tau(E)$, т. е. механизмом рассеяния. В предельных случаях слабого и сильного поля справедливы формулы (I.3.22) и (I.3.26). В случае, когда зависимость τ от энергии описывается степенной функцией (7.21) и газ носителей заряда не вырожден, мы получаем, пользуясь формулой (7.41),

$$\left.\frac{\rho_\perp-\rho_0}{\rho_0}\right|_{\text{сильн. поле}}=\frac{16}{9\pi}\Gamma(5/2-r)\,\Gamma(5/2+r)-1 \qquad (7.50')$$

и

$$\left.\frac{\rho_\perp-\rho_0}{\rho_0}\right|_{\text{слаб. поле}}=\omega_c^2C^2(kT)^{2r}\,\frac{\Gamma(3r+5/2)\,\Gamma(r+5/2)-\Gamma^2(2r+5/2)}{\Gamma^2(r+5/2)}. \qquad (7.50'')$$

Заметим, однако, что именно магнетосопротивление оказывается очень чувствительным к анизотропии изоэнергетических поверхностей. Поэтому область применимости формул (7.50) — (7.50″) довольно ограничена. В частности, при более сложном законе дисперсии насыщение магнетосопротивления может и не иметь места.

2) Продольное поле: $\mathfrak{B}\parallel\mathbf{j}$. Согласно (1.21) и (7.42) мы имеем

$$\frac{\rho_\parallel-\rho_0}{\rho_0}=\frac{\mathcal{J}_1^{(0)}-\mathcal{J}_1-\omega_c^2\mathcal{J}_3}{\mathcal{J}_1+\omega_c^2\mathcal{J}_3}. \qquad (7.51)$$

Согласно (7.41) эта величина тождественно обращается в нуль: в рамках принятой модели продольное магнетосопротивление отсутствует при любой степени вырождения носителей заряда. Этот результат связан с пренебрежением анизотропией изоэнергетических поверхностей. При учете последней продольное магнетосопротивление становится, вообще говоря, конечным. Действительно, в этом случае равенства (7.45), вообще говоря, уже не имеют места и правая часть равенства (1.24) остается конечной. Таким образом, обнаружив на опыте конечное значение продольного магнетосопротивления в неквантующем магнитном поле, мы можем утверждать, что изоэнергетические поверхности в данном материале анизотропны.

В условиях применимости кинетического уравнения намеченная и проиллюстрированная выше схема вычисления кинетических коэффициентов носит вполне общий характер. Выражения для различных кинетических коэффициентов, в том числе описывающих термомагнитные явления, можно найти, например, в книге [М7].

§ 8. Носители заряда в слабом переменном электрическом поле

Рассмотрим задачу об электропроводности однородного образца n-типа в переменном электромагнитном поле. Пусть напряженности электрического и магнитного полей гармонически меняются со временем с круговой частотой ω. Изменением $\mathfrak{E}$ в пространстве, равно

как и влиянием магнитного поля на поведение носителей заряда, будем пренебрегать. Как видно из (3.5), последняя аппроксимация оправдана, коль скоро средняя скорость носителей заряда мала по сравнению со скоростью света в пустоте. Действительно, в отличие от случая, рассмотренного в § 7 п. в, напряженности электрического и магнитного полей в электромагнитной волне не независимы, а связаны друг с другом. Так, в плоской волне они относятся, как $\sqrt{\mu/\varepsilon}$, где μ и ε — значения магнитной и диэлектрической проницаемости на соответствующей частоте; по порядку величины это отношение обычно не меньше 0,1 (исключение могут составить сегнетоэлектрики при достаточно низких частотах). Условие, при котором допустима первая из указанных аппроксимаций, будет выписано в конце этого параграфа.

В рассматриваемых условиях кинетическое уравнение (3.12) принимает вид

$$\frac{\partial f}{\partial t} - e(\mathcal{E}, \nabla_p f) + \frac{f - f_0}{\tau(E)} = 0. \tag{8.1}$$

Удобно положить

$$\mathcal{E} = \mathcal{E}_m e^{-i\omega t}, \tag{8.2}$$

где $\mathcal{E}_m$ — амплитуда напряженности поля, а произвольная начальная фаза положена равной нулю.

Как и в случае постоянного поля, решение уравнения (8.1) можно записать в виде (5.1) и (6.6а), с той лишь разницей, что теперь напряженность электрического поля дается равенством (8.2). Подставляя выражения (5.1) и (6.6а) в уравнение (8.1) и ограничиваясь по-прежнему членами, линейными по $\mathcal{E}$, мы получаем

$$\psi = \frac{e}{m} \frac{\tau}{1 - i\omega\tau} f_0'. \tag{8.3}$$

Сравнивая это с выражением (7.4), видим, что единственное отличие ψ от случая постоянного поля состоит в замене τ на комплексный множитель $\tau(1 - i\omega\tau)^{-1}$:

$$f = f_0 + \frac{e}{m}(\mathbf{p}, \mathcal{E}_m) \frac{\tau}{1 - i\omega\tau} f_0' e^{-i\omega t}. \tag{8.4}$$

Видим, что между колебаниями неравновесной части функции распределения и напряженности поля имеется сдвиг фаз, равный $\operatorname{arctg} \omega\tau$. Когда период колебаний $2\pi/\omega$ значительно превышает время релаксации ($\omega\tau \ll 1$), этот сдвиг очень мал; в противоположном предельном случае ($\omega\tau \gg 1$) он оказывается близким к $\pi/2$. Происхождение этого сдвига ясно из сказанного в § 6: скорость «приспособления» функции распределения к изменившемуся электрическому полю определяется временем τ; поскольку оно отлично от нуля, колебания f_1 должны отставать по фазе от колебаний $\mathcal{E}(t)$. В этом смысле можно сказать, что газ носителей

заряда (как и любых других частиц) обладает своеобразной «инерцией». Подчеркнем, однако, что этот эффект обусловлен не массой частиц, а статистическими свойствами их системы. Именно, поле ускоряет все электроны (дырки) одинаковым образом, но на виде функции распределения это сказывается лишь в результате столкновений.

В соответствии с формулой (8.4) комплексной оказывается и электропроводность газа носителей заряда в переменном поле. Действительно, подставляя выражение (8.4) в правую часть (2.2) и выполняя те же вычисления, что и в § 7 (п. *а*), мы получаем

$$\sigma = \sigma_1 + i\sigma_2, \tag{8.5}$$

причем вещественная и мнимая части электропроводности σ_1 и σ_2 даются формулами

$$\sigma_1 = -\frac{e^2}{3}\int_0^\infty N(E)\, v^2 \frac{\tau}{1+\omega^2\tau^2} f_0' \, dE \tag{8.6}$$

и

$$\sigma_2 = -\frac{\omega e^2}{3}\int_0^\infty N(E)\, v^2 \frac{\tau^2}{1+\omega^2\tau^2} f_0' \, dE. \tag{8.7}$$

При $\omega \to 0$ величина σ_1 переходит в статическую электропроводность (7.12), а σ_2 обращается в нуль, чего и следовало ожидать. С другой стороны, при возрастании частоты электрического поля σ_1 уменьшается по сравнению со статической проводимостью. Причина этого ясна из сказанного выше о происхождении сдвига фаз $\operatorname{arctg} \omega\tau$: не поспевая полностью за колебаниями напряженности поля, система свободных носителей заряда ведет себя — в какой-то мере — как совокупность связанных зарядов. Последние не дают вклада в ток проводимости, что и отражается в уменьшении σ_1.

При $\omega\tau \to \infty$ формула (8.6) принимает простой вид:

$$\sigma_1 = -\frac{e^2}{3\omega^2}\int_0^\infty N\tau^{-1}v^2 f_0' \, dE. \tag{8.8}$$

Частотная зависимость σ_1 определяется этим выражением явно; она имеет место при любом механизме рассеяния. Заметим, что, в отличие от статической проводимости, σ_1 в рассматриваемых условиях не возрастает, а убывает с увеличением времени свободного пробега: здесь существенно не рассеяние, ограничивающее подвижность носителей заряда, а сдвиг фаз между колебаниями f_1 и $\mathcal{E}$.

Наличие мнимой части у электропроводности, разумеется, не означает, что комплексным будет джоулево тепло Q: последнее

определяется вещественной частью σ. Так, в среднем за период

$$Q = \frac{1}{2}\mathcal{E}_m^2 \sigma_1. \tag{8.9}$$

Физический смысл мнимой части σ ясен из тех же соображений, которые были выше использованы для объяснения частотной зависимости σ_1. В соответствии с ними следует ожидать, что выражение σ_2 будет описывать плотность поляризационного тока, а сама величина σ_2 окажется связанной с вещественной частью диэлектрической проницаемости.

Чтобы найти искомую связь, напишем то уравнение Максвелла, в котором фигурирует плотность тока:

$$\operatorname{rot} \mathcal{H} = \frac{4\pi}{c}\mathbf{j} + \frac{1}{c}\frac{\partial \mathcal{D}}{\partial t}. \tag{8.10}$$

Здесь $\mathbf{j} = (\sigma_1 + i\sigma_2)\,\mathcal{E}$, а $\mathcal{D}$ есть вектор индукции:

$$\mathcal{D} = \varepsilon_0 \mathcal{E}, \tag{8.11}$$

где ε_0 — диэлектрическая проницаемость решетки, вычисленная без учета свободных носителей заряда на частоте ω *). Нас интересует решение с гармонической (вида (8.2)) зависимостью векторов $\mathcal{E}$, $\mathcal{D}$ и $\mathcal{H}$ от времени. При этом уравнение (8.10) принимает вид

$$\operatorname{rot} \mathcal{H} = \frac{4\pi}{c}\left(\omega \frac{\varepsilon_2}{4\pi} + \sigma_1\right)\mathcal{E} - \frac{i\omega}{c}\left(\varepsilon_1 - \frac{4\pi}{\omega}\sigma_2\right)\mathcal{E}. \tag{8.12}$$

Здесь ε_1 и ε_2 — вещественная и мнимая части $\varepsilon_0(\omega)$.

Очевидно, что в смысле создания магнитного поля влияние слагаемого — $4\pi\sigma_2/\omega$ физически неотличимо от влияния ε_1. Следовательно, в соответствии с нашими ожиданиями, мнимая часть электропроводности газа свободных носителей заряда определяет их вклад $\Delta\varepsilon$ в вещественную часть диэлектрической проницаемости **):

$$\Delta\varepsilon = -\frac{4\pi}{\omega}\sigma_2 = +\frac{4\pi e^2}{3}\int_0^\infty N(E)\, v^2(E)\, f_0' \frac{\tau^2}{1+\omega^2\tau^2}\, dE. \tag{8.13}$$

Согласно (8.13) величина $\Delta\varepsilon$ оказывается отрицательной ($f_0' < 0$). При этом абсолютное значение ее может оказаться и больше ε_1.

*) В случае сегнетоэлектрических полупроводников под ε_0 в последующих формулах надо понимать дифференциальную диэлектрическую проницаемость, определяемую для коллинеарных векторов $\mathcal{D}$ и $\mathcal{E}$ как $dD/d\mathcal{E}$.

**) Равенство (8.13) есть частный случай общего соотношения между комплексными электропроводностью и диэлектрической проницаемостью, вытекающего из уравнения (8.12). Это соотношение показывает, что ε и σ представляют собой, в сущности, одну комплексную характеристику системы в переменном поле. Тот факт, что иногда их все же вводят независимо, объясняется лишь историческими причинами и соображениями удобства.

Соответственно вещественная часть полной диэлектрической проницаемости может оказаться отрицательной. Как известно из электродинамики, это означает, что волны соответствующей частоты не могут распространяться в данном веществе, испытывая полное отражение от его поверхности. Такая ситуация иногда реализуется в плазме (в частности, в ионосфере).

Как и (8.6), формулы (8.7) и (8.13) упрощаются при $(\omega\tau)^2 \gg 1$. При этом величины σ_2 и $\Delta\varepsilon$ вообще перестают зависеть от механизма рассеяния. Так, вместо (8.13) мы получаем (ср. (7.15))

$$\Delta\varepsilon = -\frac{4\pi n e^2}{m_{\text{opt}}\omega^2}. \tag{8.14}$$

Согласно (8.14) эффективную массу можно было бы определить, измеряя диэлектрическую проницаемость вещества как функцию частоты (особенно если из других соображений уже известно, что зона — простая параболическая, когда $m_{\text{opt}} = m$). При этом, однако, возникают два осложнения. Во-первых, условие $(\omega\tau)^2 \gg 1$ оказывается довольно жестким: при не слишком низких температурах характерное время свободного пробега может составлять $10^{-13} \div 10^{-12}$ с; соответственно область частот, в которой справедлива формула (8.14), может оказаться неудобной для радиотехнических измерений обычного типа. Во-вторых, учет анизотропии лишает результаты однозначности. Действительно, пусть изоэнергетические поверхности представляют собой эллипсоиды и значения эффективных масс вдоль главных осей суть m_x, m_y, m_z. В пренебрежении рассеянием формулу для $\Delta\varepsilon$ легко обобщить и на этот случай. Для этой цели надо лишь заметить, что массы m_x, m_y и m_z должны входить в формулу для $\Delta\varepsilon$ равноправно, а при совпадении их друг с другом должна получиться формула (8.14). Этим условиям удовлетворяет выражение

$$\Delta\varepsilon = -\frac{4\pi n e^2}{m_{\text{opt}}\omega^2}, \tag{8.14'}$$

где

$$\frac{1}{m_{\text{opt}}} = \frac{1}{3}\left(\frac{1}{m_x} + \frac{1}{m_y} + \frac{1}{m_z}\right). \tag{8.15}$$

Из формулы (8.14′) виден смысл названия *оптическая эффективная масса*: именно эта величина фигурирует при описании ряда оптических явлений, в которых участвуют свободные носители заряда. Оправдан и термин «эффективная масса электропроводности». Действительно, при постоянном и изотропном времени релаксации τ выражение для вещественной части электропроводности кубического кристалла можно было бы записать в виде

$$\sigma_1 = \frac{ne^2}{m_{\text{opt}}}\frac{\tau}{1+\omega^2\tau^2}. \tag{8.14''}$$

Следует, однако, помнить, что фактически время релаксации само зависит от эффективных масс. Более того, как указывалось в § 6, в системе с анизотропным законом дисперсии вообще нельзя ввести единое время релаксации импульса, не зависящее от ориентации последнего относительно осей кристалла.

Из формулы (8.14′) видно, что измерение $\Delta\varepsilon$ может дать не каждую из эффективных масс по отдельности, а только их комбинацию (8.15). В этом отношении метод диамагнитного резонанса (§ IV.3) обладает явным преимуществом.

Чтобы выяснить, в какой мере оправдано принятое выше пренебрежение изменением напряженности поля в пространстве, заменим выражение (8.2) плоской волной с волновым вектором $\mathbf{k}$:

$$\mathcal{E} = \mathcal{E}_m e^{-i\omega t + i\mathbf{kr}}. \tag{8.16}$$

Аналогичными выражениями будут описываться и векторы $\mathfrak{D}$ и $\mathbf{j}$. Неравновесная часть функции распределения при этом также должна содержать множитель $e^{i\mathbf{kr}}$, а кинетическое уравнение надо писать в виде

$$\frac{\partial f}{\partial t} + (\mathbf{v},\ \nabla f) - e\,(\mathcal{E},\ \nabla_p f) + \frac{f - f_0}{\tau\,(E)} = 0. \tag{8.17}$$

Соответственно диэлектрическая проницаемость и электропроводность системы будут зависеть как от ω, так и от $\mathbf{k}$. Последняя зависимость носит название пространственной дисперсии. Очевидно, однако, что второе слагаемое в левой части (8.17) есть $i\,(\mathbf{k},\ \mathbf{v})\,f_1$. Отношение его к последнему слагаемому по абсолютной величине равно

$$kv_\tau\tau = 2\pi\,\frac{l}{\lambda},$$

где $l = v_\tau\tau$ есть длина свободного пробега по импульсу, а $\lambda = 2\pi/k$ — длина электромагнитной волны. Видим, что пространственным изменением напряженности поля в рассматриваемой задаче можно пренебречь, коль скоро длина свободного пробега достаточно мала по сравнению с длиной волны. Типичные значения длины свободного пробега во многих интересных полупроводниках составляют $10^{-6} \div 10^{-4}$ см. Таким образом, принятая нами аппроксимация оправдана в довольно широком интервале длин волн.

§ 9. Плазменные волны

Особый интерес представляет случай, когда полная диэлектрическая проницаемость образца

$$\varepsilon = \varepsilon_0 + \frac{4\pi i}{\omega}\,\sigma \tag{9.1}$$

обращается в нуль. Чтобы выяснить, как влияет последнее обстоятельство на поведение электрического поля в образце, напишем

соответствующие уравнения Максвелла и уравнение непрерывности. Магнитным полем электромагнитной волны при этом будем полностью пренебрегать. Обозначая через ρ объемную плотность заряда, мы имеем

$$\operatorname{rot} \mathfrak{E} = 0, \quad \operatorname{div} \mathfrak{D} = 4\pi\rho, \quad \frac{\partial \rho}{\partial t} + \operatorname{div} \mathbf{j} = 0. \tag{9.2}$$

Будем искать решение этих уравнений в виде (8.16). Тогда (с учетом (8.11))

$$\operatorname{rot} \mathfrak{E} = i[\mathbf{k} \times \mathfrak{E}], \quad \operatorname{div} \mathfrak{D} = i\varepsilon_0 (\mathbf{k}, \mathfrak{E})$$

и для амплитуды поля $\mathfrak{E}_m$ получаются два уравнения:

$$[\mathbf{k} \times \mathfrak{E}_m] = 0, \quad \varepsilon (\mathbf{k}, \mathfrak{E}_m) = 0. \tag{9.3}$$

При $\varepsilon \neq 0$ уравнения (9.3) имеют, очевидно, только тривиальное решение $\mathfrak{E}_m = 0$. Этого и следовало ожидать, ибо мы пренебрегли взаимосвязью электрического и магнитного полей, т. е. как раз тем фактором, который обеспечивает существование электромагнитных волн. Однако при $\varepsilon = 0$ положение меняется. Действительно, второе из уравнений (9.3) при этом превращается в тождество, а из первого следует лишь, что $\mathfrak{E}_m \parallel \mathbf{k}$, т. е. что рассматриваемые волны должны быть продольными. Итак, при $\varepsilon = 0$ оказывается возможным распространение продольных волн напряженности электрического поля и объемной плотности заряда. Эти волны называют *плазменными*. В сущности, они представляют собой один из типов нормальных колебаний поля в среде со свободными зарядами. Физический механизм их распространения состоит в том, что колебания плотности заряда и напряженности поля взаимно поддерживают друг друга.

Рассмотрим условие $\varepsilon = 0$, ограничиваясь для простоты достаточно высокими частотами ($\omega^2\tau^2 \gg 1$), когда $\Delta\varepsilon$ дается выражением (8.14′), а вещественной частью электропроводности σ_1 можно пренебречь. Диэлектрическую проницаемость решетки будем считать вещественной ($\varepsilon_0 = \varepsilon_1$). Это означает, что нас будет интересовать интервал частот, далеких от области поглощения света самой решеткой.

В принятых аппроксимациях $\varepsilon = \varepsilon_0 + \Delta\varepsilon$ и условие $\varepsilon = 0$ принимает вид

$$\omega^2 = \frac{4\pi n e^2}{\varepsilon_0 m_{\text{opt}}} \equiv \omega_{\text{pl}}^2. \tag{9.4}$$

Частота ω_{pl}, определяемая этим соотношением, называется плазменной. В соответствии с аппроксимациями, принятыми при выводе формулы (9.4), последняя имеет смысл, лишь если $(\omega_{\text{pl}}\tau)^2 \gg 1$. В большинстве полупроводников это условие оказывается довольно жестким. Действительно, полагая для оценки $\varepsilon_0 = 16$, $m_{\text{opt}} = 0{,}1m_0$, $n = 10^{16}$ см$^{-3}$, получим из (9.4) $\omega_{\text{pl}} = 4{,}5 \cdot 10^{12}$ с$^{-1}$. Если неравен-

ство $(\omega_{pl}\tau)^2 \gg 1$ не выполняется, то пренебречь величиной σ_1 нельзя. При этом условие $\varepsilon = 0$ состоит из двух уравнений:

$$\operatorname{Im}\varepsilon(\omega) = 0, \qquad \operatorname{Re}\varepsilon(\omega) = 0. \tag{9.5}$$

Удовлетворить им можно, лишь считая саму частоту комплексной величиной. Физически появление мнимой части у частоты означает, что плазменные волны будут затухать со временем, чего и следует ожидать при наличии омических потерь.

Поскольку при не слишком больших частотах формула (8.14) также несправедлива, решение уравнений (9.5) может оказаться довольно громоздким. Как мнимая, так и вещественная части плазменной частоты при этом оказываются зависящими от механизма рассеяния. Существование затухающих плазменных волн проявляется, в частности, в некоторых оптических эффектах (§ XVIII.3).

Формулы (8.13), (8.14) и, следовательно, вытекающие из них выражения для ω_{pl} получены в пренебрежении изменением напряженности поля в пространстве, т. е., формально, при $k \to 0$. Учет этого изменения с помощью уравнения (8.17) привел бы к зависимости ω_{pl} от волнового вектора $\mathbf{k}$. Определяясь, согласно сказанному в § 8, членами порядка $(kl)^2$, эта зависимость не очень сильна. Так, при $\omega_{pl}\tau \gg 1$ она свелась бы просто к появлению в правой части (9.4) малого дополнительного слагаемого *). Однако эта зависимость может оказаться существенной принципиально, ибо: а) благодаря ей мы получаем не одно дискретное значение плазменной частоты, а непрерывный их набор и б) только при учете зависимости ω от k оказывается отличной от нуля групповая скорость плазменных волн, равная $d\omega/dk$.

*) Легко видеть, что при больших волновых числах плазменные волны вообще не могут распространяться. Действительно, согласованное поведение объемной плотности заряда и напряженности электрического поля возможно, лишь если много заряженных частиц движутся почти синфазно. Это сводится к неравенству $kn^{-1/3} < 1$.

Глава XIV

РАССЕЯНИЕ НОСИТЕЛЕЙ ЗАРЯДА В НЕИДЕАЛЬНОЙ КРИСТАЛЛИЧЕСКОЙ РЕШЕТКЕ

§ 1. Постановка задачи. Теория возмущений

Результаты гл. XIII сводят расчет кинетических коэффициентов к вычислению интегралов, содержащих время свободного пробега τ. Последнее в свою очередь выражается через вероятность рассеяния $S(\mathbf{p}, \mathbf{p}')$. Вычисление этой вероятности составляет вторую — механическую — часть задачи, которой посвящена настоящая глава.

В идеальной кристаллической решетке квазиимпульс электрона сохраняется — рассеяние отсутствует. Поэтому в задаче о рассеянии носителей заряда учет неидеальности решетки обязателен. Гамильтониан системы «электрон плюс неидеальная решетка» можно представить в виде

$$H = H_0 + H'. \tag{1.1}$$

Здесь H_0 есть сумма операторов энергии электрона и фононов (XII.6.2) в идеальной решетке. Слагаемое H' описывает изменение энергии электрона при наличии того или иного отклонения силового поля от идеального периодического. Мы будем называть H' гамильтонианом взаимодействия электрона с рассеивателями. Последний термин обозначает любой объект, вызывающий отклонение силового поля решетки от идеально периодического. Во многих случаях энергия взаимодействия носителей заряда с рассеивателями оказывается достаточно малой, так что при определении энергетического спектра ею можно пренебречь (точный смысл слов «достаточно малой» будет раскрыт в дальнейшем, см. §§ 2, 4, 5). Тогда роль рассеивателей состоит лишь в том, что при «столкновениях» с ними носители заряда могут отдавать или получать квазиимпульс.

Мы приходим, таким образом, к стандартной задаче квантовой механики о вычислении вероятности перехода между двумя состояниями невозмущенной системы под влиянием возмущения H'. Роль невозмущенной системы играют в данном случае не взаимодействующие друг с другом электроны и фононы. Соответственно в оператор H_0 входит и энергия колебаний кристаллической решетки (XII.6.2). В частном случае, когда состояние решетки не изменяется, слагаемое (XII.6.2) есть просто аддитивная постоянная, выпадающая из всех последующих формул.

§ 2. Вероятность перехода. Условие применимости кинетического уравнения

Обозначим через λ совокупность квантовых чисел, характеризующих различные состояния невозмущенной системы. Соответствующие волновые функции обозначим через ψ_λ, а принадлежащие им собственные значения энергии — через E_λ:

$$H_0\psi_\lambda = E_\lambda\psi_\lambda. \tag{2.1}$$

Так, в случае одного электрона в идеальной решетке λ есть совокупность номера зоны l и компонент квазиимпульса $\mathbf{p}$ (проекцию спина в случае необходимости будем включать в l). При этом ψ_λ представляет собой функцию Блоха (III.2.15′). Для электронов в колеблющейся решетке λ есть совокупность l, $\mathbf{p}$ и чисел фононов во всех возможных состояниях; при этом ψ_λ представляет собой произведение функции Блоха на волновую функцию решетки (XII.5.5).

Подчиним функции Блоха и нормальные колебания решетки обычным условиям периодичности в кубе объема $V = L^3$. Тогда возможные значения компонент квазиимпульса электрона и квазиволнового вектора фонона даются (III.3.10) и (III.3.10′).

Функции ψ_λ будем считать ортонормированными:

$$\int \psi^*_\lambda\psi_{\lambda'}\,d\tau = \delta_{\lambda\lambda'}, \tag{2.2}$$

где $\delta_{\lambda\lambda'}$ символ Кронекера (многомерный), а символ $\int d\tau \ldots$ означает интегрирование по координатам электрона (по фундаментальному объему) и по вещественным нормальным координатам решетки.

Волновую функцию, вычисленную с учетом энергии взаимодействия H', обозначим через Ψ. Пусть в начальный момент времени ($t = 0$) система находится в состоянии λ:

$$\Psi|_{t=0} = \psi_\lambda. \tag{2.3}$$

В последующие моменты времени волновая функция Ψ будет изменяться в соответствии с уравнением Шредингера

$$i\hbar\frac{\partial\Psi}{\partial t} = H\Psi. \tag{2.4}$$

В отсутствие взаимодействия H' мы получили бы отсюда

$$\Psi = \psi_\lambda \exp\left(-\frac{i}{\hbar}E_\lambda t\right), \tag{2.5}$$

как и должно быть в стационарном состоянии. При учете взаимодействия это уже не так: функции ψ_λ не являются собственными функциями полного гамильтониана $H = H_0 + H'$ и, следовательно, не описывают стационарных состояний. Положим

$$\Psi(t) = \sum_{\lambda''} c_{\lambda''}(t)\psi_{\lambda''}\exp\left(-\frac{i}{\hbar}E_{\lambda''}t\right), \tag{2.6}$$

где $c_{\lambda''}$ — пока неизвестные коэффициенты разложения. Согласно общим правилам квантовой механики величины $|c_{\lambda''}(t)|^2$ представляют собой вероятности обнаружить систему в момент времени t в состоянии λ''. Сумма их, взятая по всем возможным значениям λ'', равна единице при всех t:

$$\sum_{\lambda''} |c_{\lambda''}(t)|^2 = 1. \tag{2.7}$$

Уравнения для коэффициентов $c_{\lambda''}$ легко найти, подставляя функцию (2.6) в уравнение (2.4), умножая результат слева на $\psi^*_{\lambda'}$ и пользуясь равенствами (2.1) и (2.2). Мы получаем

$$i\hbar \frac{\partial c_{\lambda'}(t)}{\partial t} = \sum_{\lambda''} c_{\lambda''}(t) \exp\left[\frac{i}{\hbar}(E_{\lambda'} - E_{\lambda''})\, t\right] \int \psi^*_{\lambda'} H' \psi_{\lambda''}\, d\tau. \tag{2.8}$$

Начальное условие к этой системе уравнений, согласно (2.3), имеет вид

$$c_{\lambda'}|_{t=0} = \delta_{\lambda\lambda'}. \tag{2.9}$$

Интегралы в правой части (2.8) представляют собой матричные элементы оператора H' в системе функций ψ_λ. Введем для них обозначение

$$\int \psi^*_{\lambda'} H' \psi_{\lambda''}\, d\tau = (\lambda' \mid H' \mid \lambda''). \tag{2.10}$$

Таким образом,

$$i\hbar \frac{\partial c_{\lambda'}(t)}{\partial t} = \sum_{\lambda''} (\lambda' \mid H' \mid \lambda'')\, c_{\lambda''}(t) \exp\left[\frac{i}{\hbar}(E_{\lambda'} - E_{\lambda''})\, t\right]. \tag{2.8'}$$

Как видно из уравнений (2.8'), при $t \neq 0$ коэффициенты $c_{\lambda'}$ при $\lambda' \neq \lambda$ становятся, вообще говоря, отличными от нуля, а величина $|c_\lambda|^2$, соответственно, уменьшается в силу (2.7). Это есть математическое выражение нестационарности состояния λ в присутствии возмущения H'.

Точное решение системы (2.8') связано с серьезными математическими трудностями. Приближенное решение можно получить по методу возмущений. Именно, допустим, что энергия взаимодействия, описываемая оператором H', достаточно мала. Тогда систему (2.8') можно решать итерациями, считая матричные элементы $(\lambda'|H'|\lambda'')$ величинами первого порядка малости.

Рассмотрим сначала случай $\lambda' \neq \lambda$. Тогда в первом приближении в правую часть (2.8') можно подставить невозмущенные значения $c_{\lambda''}$. Последние, очевидно, совпадают с начальными значениями (2.9). Действительно, из вида системы (2.8') непосредственно следует, что изменение коэффициентов $c_{\lambda'}$ со временем обусловлено только наличием взаимодействия H'. Таким образом,

при $\lambda' \neq \lambda$

$$i\hbar \frac{\partial c_{\lambda'}}{\partial t} = (\lambda' \mid H' \mid \lambda) \exp\left[\frac{i}{\hbar}(E_{\lambda'} - E_\lambda)\right]. \qquad (2.11)$$

Отсюда, с учетом (2.9), легко находим

$$c_{\lambda'}(t) = \frac{-1 + \exp\left[\frac{i}{\hbar}(E_{\lambda'} - E_\lambda)\, t\right]}{E_\lambda - E_{\lambda'}} (\lambda' \mid H' \mid \lambda). \qquad (2.12)$$

Следовательно, вероятность обнаружить систему в момент времени t в состоянии λ' есть

$$|c_{\lambda'}(t)|^2 = \frac{2\left[1 - \cos\left(\frac{E_{\lambda'} - E_\lambda}{\hbar} t\right)\right]}{(E_\lambda - E_{\lambda'})^2} |(\lambda' \mid H' \mid \lambda)|^2. \qquad (2.13)$$

Подставляя (2.13) в левую часть равенства (2.7), можем найти $|c_\lambda(t)|^2$ — вероятность того, что система останется в состоянии λ.

Дифференцируя соотношение (2.13) по t, получим вероятность перехода, отнесенную к единице времени:

$$\frac{d|c_{\lambda'}(t)|^2}{dt} = \frac{2}{\hbar} \frac{\sin\left(\frac{E_{\lambda'} - E_\lambda}{\hbar} t\right)}{E_{\lambda'} - E_\lambda} |(\lambda' \mid H' \mid \lambda)|^2. \qquad (2.14)$$

Формулу (2.14) еще нельзя непосредственно использовать в кинетическом уравнении, ибо она относится к переходам между состояниями дискретного спектра: мы рассматривали компоненты квазиимпульса, равно как и компоненты квазиволнового вектора фонона, как дискретные величины. В то же время в кинетическом уравнении (XIII.3.12) речь идет о состояниях непрерывного спектра. Переход от дискретного спектра к непрерывному легко выполнить, замечая, что фактически мы всегда имеем дело с системами макроскопически больших размеров. Это позволяет упростить выражение (2.14) при больших t. Именно, согласно (П.IV.6) второй сомножитель в правой части (2.14) асимптотически при $t \to \infty$ превращается в δ-функцию, и мы получаем

$$\frac{d|c_\lambda(t)|^2}{dt} = \frac{2\pi}{\hbar} |(\lambda' \mid H' \mid \lambda)|^2 \delta(E_{\lambda'} - E_\lambda). \qquad (2.15)$$

Таким образом, переходы происходят лишь между состояниями с одинаковой энергией: $E_{\lambda'} = E_\lambda$.

Как уже говорилось, величины $c_{\lambda'}$ не обязательно характеризуют состояния только электрона: в зависимости от конкретных условий они могут относиться и к системе «электрон + фононы» или «электрон + электромагнитное поле световой волны» и т. д. С другой стороны, в интеграле столкновений фигурируют вели-

чины $\mathscr{P}_{1,2}(\mathbf{p}, \mathbf{p}')$, определяющие вероятности данного электронного перехода безотносительно к тому, что делается, например, с фононами. Чтобы получить их, надо просуммировать $\frac{d|c_{\lambda'}(t)|^2}{dt}$ по всем «неэлектронным» квантовым числам, входящим в состав λ' (например, по всем квазиволновым векторам фононов и по всем ветвям фононного спектра).

Обозначим, как и в гл. XII, совокупность чисел фононов во всех состояниях n_f через n. Тогда $\lambda = \{\mathbf{p}, l, n\}$ и отнесенная к единице времени вероятность электронного перехода будет

$$W(l, \mathbf{p}; l', \mathbf{p}') = \sum_{n'} \frac{d|c_{\lambda'}(t)|^2}{dt}. \tag{2.16}$$

В задаче о рассеянии носителей заряда нас интересуют вероятности W при $l' = l$. Величины $\mathscr{P}_1(\mathbf{p}, \mathbf{p}')$ и $\mathscr{P}_2(\mathbf{p}, \mathbf{p}')$ отличаются от них только нормировочным множителем. Вид его легко найти, вычисляя, например, интеграл «ухода» B через коэффициенты c_λ и сравнивая результат с формулой (XIII.3.8) *). Очевидно,

$$B = \sum_{\mathbf{p}'} W(l, \mathbf{p}; l, \mathbf{p}') f(\mathbf{p})[1 - f(\mathbf{p}')] = \\ = \frac{V}{(2\pi\hbar)^3} \int d\mathbf{p}' W(l, \mathbf{p}; l, \mathbf{p}') f(\mathbf{p})[1 - f(\mathbf{p}')]. \tag{2.17}$$

Мы воспользовались здесь правилом (П.Х.3). Функция распределения f относится, разумеется, к зоне номера l — единственной, с которой мы сейчас имеем дело.

Не следует удивляться тому, что в правой части (2.17) и далее явно фигурирует основной объем V. Он входит и в выражение для вероятности перехода W (например, через условие нормировки (2.2)) и выпадает из окончательных выражений для наблюдаемых на опыте величин. Там, как мы увидим, вместо V будут фигурировать только такие величины, как концентрация примеси или объем элементарной ячейки.

Сравнивая формулы (2.17) и (XIII.3.8), получаем

$$\mathscr{P}_2(\mathbf{p}, \mathbf{p}') = VW(l, \mathbf{p}; l, \mathbf{p}'), \tag{2.18}$$

т. е. асимптотически при $t \to \infty$

$$\mathscr{P}_2(\mathbf{p}, \mathbf{p}') = \frac{2\pi}{\hbar} V \sum_{n'} |(l, \mathbf{p}', n'|H'|l, \mathbf{p}, n)|^2 \delta[E(l, \mathbf{p}', n') - E(l, \mathbf{p}, n)]. \tag{2.19}$$

В дальнейшем в этой главе мы будем для краткости опускать индекс l в формулах типа (2.18).

*) Разумеется, тот же самый результат получится, если вместо B вычислить член «прихода» A.

Пользуясь формулами (2.15) и (2.19), следует помнить, что при их выводе было сделано одно важное предположение. Именно, переходя от (2.14) к (2.15), мы совершили предельный переход $t \to \infty$. Однако фактически время t, в течение которого могут быть справедливы равенства (2.11) и (2.14), ограничено. Действительно, в первом приближении теории возмущений не учитываются повторные акты рассеяния. Но таковые с большой вероятностью должны произойти спустя время порядка времени свободного пробега τ. Таким образом, должно выполняться неравенство

$$t \ll \tau. \tag{2.20}$$

Как можно показать *), неравенство типа (2.20) не связано обязательно с использованием теории возмущений в принятой выше форме. Оно всегда появляется, коль скоро мы вообще имеем дело с кинетическим уравнением. Это обстоятельство очень существенно, ибо равенство (2.15) получено из (2.14) формальным предельным переходом $t \to \infty$. Физически это означает, что величина t должна значительно превышать характерное время, определяющее динамику столкновения. Чтобы понять, что́ это за время, запишем аргумент синуса в правой части (2.14) в виде

$$\frac{E_{\lambda'} - E_\lambda}{\bar{E}} \frac{\bar{E}}{\hbar} t.$$

При $\frac{\bar{E}}{\hbar} t \gg 1$ этот аргумент становится, вообще говоря, очень большим и, соответственно, синус, а с ним и вся правая часть (2.14), рассматриваемая как функция $E_{\lambda'}$, быстро осциллирует. Поэтому вклад ее в интеграл по $\mathbf{p}'$, фигурирующий в кинетическом уравнении, будет очень мал. Исключение составляет случай, когда отношение $\frac{|E_{\lambda'} - E_\lambda|}{\bar{E}}$ достаточно мало: $\frac{|E_{\lambda'} - E_\lambda|}{\bar{E}} \ll \frac{\hbar}{\bar{E}t}$. В пределе при $t \to \infty$ это и дает условие сохранения энергии, выражаемое δ-функцией в формуле (2.15).

Таким образом, выражение «$t \to \infty$» фактически означает неравенство

$$t \gg \hbar/\bar{E}. \tag{2.21}$$

Чтобы теория была последовательной, это неравенство не должно противоречить (2.20). Следовательно, время релаксации должно быть достаточно большим:

$$\tau \gg \hbar/kT \tag{2.22а}$$

— в случае невырожденного электронного газа и

$$\tau \gg \hbar/\zeta \tag{2.22б}$$

*) См. сборник [1] (статьи №№ 5, 6), а также конец настоящего параграфа.

— в случае газа, полностью вырожденного (в случае полностью вырожденного дырочного газа знаменатель правой части (2.22б) следует заменить на $-E_{\pi} - \zeta$).

Подчеркнем, что отказаться от условия (2.21) (и, следовательно, от (2.22а, б)) нельзя: без него вообще нельзя было бы ввести представление о не зависящей от времени вероятности перехода (2.15), и стационарная постановка задачи, принятая нами в предыдущей главе (и навязанная опытом), потеряла бы смысл. Иначе говоря, неравенства (2.22а, б) составляют основное условие применимости кинетического уравнения. Вычислив для какого-нибудь механизма рассеяния время релаксации, мы должны затем проверить, удовлетворяется ли это условие. Только если оно удовлетворяется, результат имеет смысл и может быть использован для дальнейшего вычисления кинетических коэффициентов. В противном случае вся постановка задачи нуждается в пересмотре: энергию взаимодействия носителей заряда с соответствующими рассеивателями надо принимать во внимание уже при определении энергетического спектра системы.

Причину неизбежного появления неравенств (2.22а, б) в методе кинетического уравнения можно понять с помощью соотношения неопределенности между энергией и временем. При конечном времени наблюдения τ неопределенность в разности энергий начального и конечного состояний — порядка $\hbar/\tau$. Чтобы можно было, хотя бы приближенно, говорить о сохранении энергии при столкновении, эта неопределенность должна быть мала по сравнению с характерной энергией электрона. Отсюда сразу вытекают неравенства (2.22а, б).

Во избежание недоразумений напомним, что E_{λ} и $E_{\lambda'}$ суть значения энергии невозмущенной системы. Точная энергия, которую надо было бы вычислять с учетом как H_0, так и энергии взаимодействия H', сохраняется, конечно, точно.

Формула (2.19) сводит задачу о расчете времени релаксации к вычислению матричных элементов оператора H'.

§ 3. Энергия взаимодействия носителей заряда с фононами

а. Общие соображения. По определению оператор H' описывает изменение энергии носителя заряда при смещении атомов решетки из положений равновесия. Вычисление этой энергии составляет весьма сложную задачу — хотя бы потому, что атомы обладают конечными размерами и при смещении могут деформироваться. Для определения этой деформации и связанного с ней изменения силового поля надо было бы решить динамическую задачу многих тел.

Положение, однако, значительно упрощается в случае, типичном для задачи о рассеянии. Именно, из дальнейшего (§ 4) будет

видно, что главную роль в рассеянии носителей заряда играют фононы со сравнительно малыми квазиволновыми векторами: длины волн «существенных» фононов оказываются примерно такими же, как у электронов, т. е. значительно больше постоянной решетки. Как всегда в подобных случаях, имеются в виду длины волн большинства электронов (в невырожденной системе — порядка $2\pi\hbar/\sqrt{mkT}$). Приняв это во внимание и пользуясь некоторыми соображениями общего характера, мы можем — не решая указанную выше задачу динамики — написать явное выражение для оператора H', содержащее лишь небольшое число экспериментально определяемых параметров.

б. Взаимодействие носителей заряда с акустическими фононами; метод потенциала деформации. Рассмотрим сначала одноатомный кристалл ($r = 1$).

Заметим прежде всего, что оператор H' должен быть скаляром. Далее, он отличен от нуля, лишь если атомы решетки смещены из положений равновесия. Иначе говоря, оператор H' должен обращаться в нуль вместе с вектором смещения. Рассматривая, как и в гл. XII, случай малых колебаний, мы можем ограничиться линейной зависимостью H' от векторов $\mathbf{Q}$ ($\mathbf{g}$, h) или от их производных *). Наконец, надо принять во внимание, что при сдвиге или повороте всего кристалла как целого энергия электронов не меняется. Такому сдвигу соответствует смещение всех атомов решетки строго в одной фазе. Оператор H' в указанных условиях должен обращаться в нуль. В рассматриваемых сейчас одноатомных кристаллах возможны, как мы знаем, только акустические колебания решетки. При этом синфазному движению атомов соответствует бесконечно большая длина волны, т. е., согласно (XII.4.5), вектор смещения, не зависящий от координат, $\mathbf{r}$. Отсюда следует, что оператор H' должен выражаться не через сам вектор смещения, а через производные от него по координатам — при сдвиге кристалла как целого эти производные обращаются в нуль. Иначе говоря, важно не смещение атомов само по себе, а деформация решетки, при этом возникающая.

Интересуясь только длинноволновыми акустическими фононами, мы можем ограничиться первыми производными по координатам. Действительно, из формулы (XII.4.5) явствует, что каждое дифференцирование $\mathbf{Q}$ по координатам дает множитель $\mathbf{q}$.

Итак, оператор H' представляет собой скаляр, линейно зависящий от первых производных вектора смещения $\mathbf{Q}$ по координатам.

---

*) В гл. XII при разложении потенциальной энергии решетки в ряд по степеням смещений надо было взять квадратичные члены, ибо линейные обращались в нуль по условию минимума. В данном случае оператор H' никакими экстремальными свойствами относительно смещений, вообще говоря, не обладает и линейные члены в нуль обращаться не обязаны.

Простейшее выражение такого типа имеет вид

$$H'_{\text{ак}} = E_1 \operatorname{div} \mathbf{Q}, \tag{3.1}$$

где E_1 — некоторая постоянная. В принципе ее можно было бы вычислить, решая задачу механики об изменении энергии электрона при плавной периодической (с периодом волны) деформации решетки. Проще, однако, рассматривать E_1 как параметр теории, подлежащий определению из опыта (подобно, например, эффективной массе или ширине запрещенной зоны). Значения E_1 для электронов проводимости и для дырок, вообще говоря, различны. Соответственно мы имеем здесь не один, а два параметра, E_{1c} и E_{1v}, относящиеся, соответственно, к зоне проводимости и к валентной зоне.

Выражение (3.1) называется *потенциалом деформации*, а параметр E_1 — *константой потенциала деформации* (иногда слово «константа» опускают, относя термин «потенциал деформации» не только к правой части (3.1), но и к параметру E_1) *). Формула такого же типа описывает и изменение энергии носителя заряда при однородной статической деформации, возникающей, например, при сжатии или растяжении кристалла. Однако параметры, входящие в это последнее выражение, вообще говоря, не совпадают с E_{1c} и E_{1v}. Действительно, длина акустической волны хотя и велика по сравнению с постоянной решетки, но все же конечна, т. е. деформация, создаваемая акустической волной, всегда неоднородна.

Подставим в (3.1) выражение (XII.5.18) для вектора смещения при $\mathbf{q} \to 0$. Согласно (XII.2.13) при этом вектор $\boldsymbol{\zeta}$ оказывается вещественным. Согласно (XII.2.19′) он еще и единичен. Таким образом,

$$H'_{\text{ак}} = \frac{iE_1}{\sqrt{G}} \sum_{\mathbf{q}} \sum_{s=1}^{3} (\mathbf{q}, \boldsymbol{\zeta}(\mathbf{q}, s)) \sqrt{\frac{\hbar}{2M\omega(\mathbf{q}, s)}} \times$$
$$\times \{b(\mathbf{q}, s) e^{i\mathbf{qr}} - b^*(\mathbf{q}, s) e^{-i\mathbf{qr}}\}. \tag{3.2}$$

В рассматриваемом случае длинных акустических волн имеет смысл разделение колебаний на поперечные и продольные. Из формулы (3.2) явствует, что в рамках принятых предположений только продольные волны ($s = 1$) дают вклад в изменение энергии электрона: при $\boldsymbol{\zeta}(\mathbf{q}, s) \perp \mathbf{q}$ соответствующие слагаемые в (3.2) обращаются в нуль. Следовательно,

$$H'_{\text{ак}} = iE_1 \sum_{\mathbf{q}} q \sqrt{\frac{\hbar}{2MG\omega(\mathbf{q}, 1)}} \{b(\mathbf{q}, 1) e^{i\mathbf{qr}} - b^*(\mathbf{q}, 1) e^{-i\mathbf{qr}}\}. \tag{3.3}$$

Здесь принято во внимание, что для продольных волн (снабженных индексом $s = 1$) $(\mathbf{q}, \boldsymbol{\zeta}(\mathbf{q}, 1)) = q$.

*) Выражение (3.1) было использовано в 1935 г. Титейкой в одной задаче теории металлов. В теорию полупроводников оно было введено независимо М. Ф. Дейгеном и С. И. Пекаром, и Дж. Бардином и В. Шокли.

Величина $\operatorname{div}\mathbf{Q}$, фигурирующая в правой части (3.1), имеет простой геометрический смысл. Как показывается в механике сплошных сред, при малых деформациях

$$\operatorname{div}\mathbf{Q}=\frac{\Delta V}{V_0}. \tag{3.4}$$

Здесь V_0 — некоторый объем в недеформированном кристалле (в качестве V_0 можно выбрать, например, объем элементарной ячейки), ΔV — его изменение при деформации, описываемой вектором $\mathbf{Q}$. Таким образом, $\operatorname{div}\mathbf{Q}$ есть относительное изменение объема при малой деформации.

Равенство (3.4) позволяет грубо оценить порядок величины константы E_1. Действительно, изменение энергии электрона или дырки H' можно интерпретировать как смещение границы соответствующей зоны при деформации *). Экстраполируя формулы (3.1), (3.2) на случай не малых деформаций, видим, что E_1 есть, по порядку величины, сдвиг границы зоны при $\Delta V=V_0$, т. е. при изменении объема элементарной ячейки в два раза. Очевидно, это должна быть энергия порядка энергии электрона во внешней атомной оболочке, т. е. $|E_1|\sim 1\div 10$ эВ. При этом знаки E_{1c} и E_{1v} в принципе могут быть любыми: в зависимости от природы кристалла дно зоны проводимости и потолок валентной зоны могут, независимо друг от друга, как подниматься, так и опускаться при сжатии (и, соответственно, опускаться или подниматься при расширении).

Согласно (3.1) и (3.4) колебания решетки, не связанные с изменением объема элементарной ячейки, не меняют энергии носителей заряда и, следовательно, не вызывают рассеяния их. Само выражение (3.1), однако, не вытекает однозначно из поставленных выше условий. Скаляр наиболее общего вида, линейно выражающийся через первые производные вектора смещения по координатам, представляет собой линейную форму

$$H'_{\text{ак}}=\sum_{s=1}^{3} E_{\alpha\beta}(s)\,\frac{\partial Q_{\alpha,s}}{\partial x_\beta}. \tag{3.5}$$

Как показано в Приложении XI, величины $E_{\alpha\beta}(s)$ образуют симметричный тензор 2-го ранга (вообще говоря, свой для каждой ветви колебаний решетки). По этой причине формулу (3.5) можно переписать в виде

$$H'_{\text{ак}}=\sum_{s=1}^{3} E_{\alpha\beta}(s)\,u_{\alpha\beta}(s), \tag{3.5'}$$

*) В рассматриваемом случае неоднородной деформации это смещение зависит от координат, что и приводит к рассеянию.

где $u_{\alpha\beta}(s)=\frac{1}{2}\left(\frac{\partial Q_\alpha}{\partial x_\beta}+\frac{\partial Q_\beta}{\partial x_\alpha}\right)$ — компоненты тензора деформации. Формула (3.1) есть частный случай (3.5), когда тензор $E_{\alpha\beta}(s)$ вырождается в скаляр:

$$E_{\alpha\beta}=E_1\delta_{\alpha\beta}. \tag{3.6}$$

Зависимость от номера ветви при этом становится несущественной, поскольку все равно во взаимодействии с электронами участвуют только продольные фононы ($s=1$). Заметим, однако, что формула (3.6), вообще говоря, несправедлива даже в кубическом кристалле и для точного расчета надо пользоваться общим выражением (3.5′). Благодаря условиям кристаллографической симметрии не все компоненты тензора $E_{\alpha\beta}$ независимы, но все же число параметров, подлежащих определению из опыта, здесь больше, чем в изотропном случае (3.6). Так, в n-Ge и n-Si их два (см. Приложение XI).

Утверждение об исключительной роли продольных фононов в случае (3.5) уже не имеет места. Действительно, вместо (3.3) мы получаем теперь на основании (XII.5.18)

$$H'_{\text{ак}}=i\sum_{s=1}^{3}\sum_{\mathbf{q}}E_{\alpha\beta}q_\alpha\zeta_\beta\sqrt{\frac{\hbar}{2MG\omega(\mathbf{q},\ s)}}\{b(\mathbf{q},\ s)e^{i\mathbf{qr}}-b^*(\mathbf{q},\ s)e^{-i\mathbf{qr}}\}. \tag{3.7}$$

Фигурирующая здесь сумма

$$E_{\alpha\beta}q_\alpha\zeta_\beta,$$

вообще говоря, не обращается в нуль при $\boldsymbol{\zeta}\perp\mathbf{q}$. Соответственно поперечные акустические колебания решетки также дают вклад в рассеяние носителей заряда в анизотропном случае (3.5) (Приложение XI).

Обратимся теперь к кристаллам со сложной структурой элементарной ячейки ($r>1$). Сформулированные выше общие условия, которым должен удовлетворять гамильтониан взаимодействия H', остаются в силе и здесь. Отсюда следует, что энергия взаимодействия электрона с длинноволновыми акустическими фононами и в этом случае дается формулой (3.1) или ее анизотропным обобщением (3.5). В кристаллах с центром симметрии (к числу их принадлежат германий и кремний) остаются в силе и явные формулы (3.3) или (3.7).

в. Взаимодействие носителей заряда с оптическими фононами в гомеополярном кристалле; метод потенциала деформации. Согласно § XII.3 при длинноволновых оптических колебаниях центр тяжести элементарной ячейки почти не смещается. Соответственно здесь нет необходимости требовать, чтобы гамильтониан взаимодействия электронов с оптическими фононами выражался непременно через производные от компонент вектора смещения по координатам. Допустимым может оказаться и выражение, содержащее сам

вектор смещения:

$$H'_{\text{опт}} = \sum_{s \geqslant 4} \mathbf{A}(s)\, \mathbf{Q}_s. \tag{3.8}$$

Здесь $\mathbf{A}(s)$ — некоторый вектор, зависящий, вообще говоря, от номера ветви s; компоненты его должны выражаться через величины, характеризующие зонную структуру носителей заряда в соответствующей идеальной решетке. Нижний предел суммирования $s = 4$ напоминает, что речь идет только об оптических фононах.

Формула (3.8) оправдана, если условия симметрии электронных (дырочных) изоэнергетических поверхностей допускают существование отличного от нуля вектора $\mathbf{A}$. Очевидно, для этого необходимо, чтобы некоторые направления в зоне Бриллюэна были физически выделены. Так обстоит дело, например, если минимум энергии лежит не в центре зоны Бриллюэна (таких точек минимума может быть несколько): в этом случае выделены направления, параллельные радиус-векторам, соединяющим точки минимума с точкой $\mathbf{p} = 0$. Это необходимое условие, однако, еще не является достаточным. Достаточные условия, при которых гамильтониан H' можно записать в виде (3.8), можно получить, лишь исследуя всю совокупность соотношений симметрии, характеризующих поверхности постоянной энергии. Анализ показывает, что для электронов проводимости в германии такая запись возможна, а в кремнии $\mathbf{A}(s) = 0$ *).

При $\mathbf{A}(s) = 0$ в правую часть (3.8) надо включить слагаемые с производными типа (3.5). Разумеется, их можно было бы написать и при $\mathbf{A} \neq 0$; однако в этом случае их роль была бы относительно невелика, поскольку здесь существенны только длинные волны.

Интересуясь электронами вблизи данного минимума, мы можем записать вектор $\mathbf{A}$ в виде

$$\mathbf{A} = \mathbf{n} E_0, \tag{3.9}$$

где $\mathbf{n}$ — единичный вектор вдоль оси, идущей из центра зоны Бриллюэна в рассматриваемый минимум, а E_0 — некоторая постоянная (размерности энергии, деленной на длину). Более общая форма (аналогичная переходу от (3.1) к (3.5)) имеет вид

$$A_\alpha = E^{(0)}_{\alpha\beta} n_\beta, \tag{3.9'}$$

где $E^{(0)}_{\alpha\beta}$ — компоненты некоторого тензора. Очевидно, формула (3.9) получается из (3.9'), если этот тензор вырождается в скаляр.

Подставляя выражение (XII.5.19) в правую часть (3.8), находим окончательный вид гамильтониана взаимодействия электронов с оптическими фононами в гомеополярном кристалле:

$$H'_{\text{опт}} = \sum_{s \geqslant 4} \sum_{\mathbf{q}} E^{(0)}_{\alpha\beta}(s)\, n_\beta \zeta_\alpha \left[\frac{\hbar}{2MG\omega(\mathbf{q}, s)}\right]^{1/2} \{b(\mathbf{q}, s)\, e^{i\mathbf{qr}} + b^*(\mathbf{q}, s)\, e^{-i\mathbf{qr}}\}. \tag{3.10}$$

*) Доказательство этого утверждения можно найти, например, в книге [2].

Для кубических кристаллов величину E_0 (или компоненты $E^{(0)}_{\alpha\beta}$) иногда пишут в виде произведения $\Xi_0 b$ (или, соответственно, $\Xi^{(0)}_{\alpha\beta} b$), где b — абсолютная величина вектора обратной решетки, а Ξ ($\Xi_{\alpha\beta}$) — новая постоянная размерности энергии. Ее называют *константой оптического потенциала деформации.*

г. Взаимодействие носителей заряда с оптическими фононами в гетерополярном кристалле. В гетерополярном полупроводнике атомы решетки обладают отличными от нуля эффективными зарядами. В результате оптические колебания решетки сопровождаются колебаниями дипольного момента, т. е. колебаниями вектора поляризации среды *). Это приводит к изменению энергии электрона (по сравнению с покоящейся решеткой), не связанному непосредственно с потенциалом деформации. Для длинных волн это изменение энергии легко вычислить с помощью законов электростатики.

Пусть изменение дипольного момента элементарной ячейки при смещении атомов из положения равновесия есть $\mathbf{d}$. Соответствующий вектор поляризации $\mathcal{P}$, очевидно, равен

$$\mathcal{P} = \frac{\mathbf{d}}{V_0}, \tag{3.11}$$

где V_0 — объем элементарной ячейки (для краткости мы говорим просто «вектор поляризации» вместо «изменение вектора поляризации»).

Как известно из электростатики, наличие поляризации $\mathcal{P}$ означает, что в решетке имеется связанный заряд с объемной плотностью

$$\rho = -\operatorname{div} \mathcal{P}. \tag{3.12}$$

Этот заряд создает электрическое поле, потенциал которого φ определяется из уравнения Пуассона

$$\nabla^2 \varphi = -4\pi\rho,$$

или, принимая во внимание равенства (3.11) и (3.12), из уравнения

$$\nabla^2 \varphi = \frac{4\pi}{V_0} \operatorname{div} \mathbf{d}. \tag{3.13}$$

Найдя отсюда φ, получим гамильтониан взаимодействия носителя заряда с поляризационными фононами:

$$H'_{\text{пол}} = \pm e\varphi. \tag{3.14}$$

Для вычисления $\mathbf{d}$ надо специализировать природу кристалла. Мы рассмотрим кристалл кубической системы, элементарная ячейка которого содержит два атома различной химической природы. К такому типу относятся, например, щелочно-галоидные кристаллы (NaCl, KCl и др.), кристаллы типа CsCl, а также соедине-

*) По этой причине колебания рассматриваемого типа иногда называют поляризационными.

ния типа $A^{III}B^{V}$ — InSb, GaAs и др. Эффективные заряды атомов той и другой природы обозначим, соответственно, через Ze и $-Ze$. Число Z может быть и не целым, если химическая связь в рассматриваемом кристалле — не чисто ионная (гл. II). Очевидно,

$$\mathbf{d} = Ze\,[\mathbf{Q}\,(\mathbf{r},\,1) - \mathbf{Q}\,(\mathbf{r},\,2)], \tag{3.15}$$

где индексы 1, 2 нумеруют атомы ячейки, а радиус-вектор $\mathbf{r}$ считается непрерывным (мы рассматриваем случай длинных волн). Для вектора смещения $\mathbf{Q}$ надо подставить выражение (XII.5.19). Мы получаем

$$\mathbf{d} = Ze \sum_{s\geqslant 4} \sum_{\mathbf{q}} \left[\frac{\hbar}{2MG\omega\,(\mathbf{q},\,s)}\right]^{1/2} [\xi_1\,(\mathbf{q},\,s) - \xi_2\,(\mathbf{q},\,s)] \times$$
$$\times \{b\,(\mathbf{q},\,s)\,e^{i\mathbf{qr}} + b^*\,(\mathbf{q},\,s)\,e^{-i\mathbf{qr}}\}. \tag{3.16}$$

Уравнение (3.13) принимает вид

$$\nabla^2\varphi = \frac{4\pi i Ze}{V_0} \sum_{s\geqslant 4} \sum_{\mathbf{q}} \left[\frac{\hbar}{2MG\omega\,(\mathbf{q},\,s)}\right]^{1/2} [(\mathbf{q}\xi_1) - (\mathbf{q}\xi_2)] \times$$
$$\times \{b\,(\mathbf{q},\,s)\,e^{i\mathbf{qr}} + b^*\,(\mathbf{q},\,s)\,e^{-i\mathbf{qr}}\}.$$

Нас интересует частное решение этого уравнения, обращающееся в нуль вместе с параметром Z. Представляя φ в виде разложения Фурье:

$$\varphi\,(\mathbf{r}) = \sum_{\mathbf{q}} \{\varphi_{\mathbf{q}} e^{i\mathbf{qr}} + \varphi^*_{\mathbf{q}} e^{-i\mathbf{qr}}\},$$

легко находим

$$\varphi\,(\mathbf{r}) = -\frac{4\pi i Ze}{V_0} \sum_{s\geqslant 4}^{6} \sum_{\mathbf{q}} \frac{1}{q^2} \left[\frac{\hbar}{2MG\omega\,(\mathbf{q},\,s)}\right]^{1/2} [(\mathbf{q},\,\xi_1) - (\mathbf{q},\,\xi_2)] \times$$
$$\times \{b\,(\mathbf{q},\,s)\,e^{i\mathbf{qr}} - b^*\,(\mathbf{q},\,s)\,e^{-i\mathbf{qr}}\}. \tag{3.17}$$

Видно, что поперечные фононы не дают вклада в правую часть (3.17): для них $(\mathbf{q},\,\xi_1) = (\mathbf{q},\,\xi_2) = 0$.

Поскольку в сумме по s в правой части (3.17) остается один член, мы можем вообще опустить индекс s в аргументах функций ω, ξ_1, ξ_2, понимая под ними только величины, относящиеся к продольным оптическим фононам. То же справедливо и для операторов b, b^*.

Задача об определении энергии взаимодействия электрона с длинноволновыми поляризационными фононами свелась теперь к вычислению разности $\xi_1 - \xi_2$ при малых значениях волнового вектора $\mathbf{q}$. Для этой цели воспользуемся соотношениями (XII.2.19) и (XII.3.7). При $r = 2$ и $\mathbf{q} = 0$ они принимают вид

$$M_1\xi_1^2\,(0) + M_2\xi_2^2\,(0) = M_1 + M_2, \qquad M_1\xi_1\,(0) + M_2\xi_2\,(0) = 0.$$

Отсюда находим

$$\zeta_1 = \left(\frac{M_2}{M_1}\right)^{1/2} \zeta, \quad \zeta_2 = -\left(\frac{M_1}{M_2}\right)^{1/2} \zeta, \tag{3.18}$$

где ζ — единичный вектор, определяющий направление смещения первого атома при колебаниях данной ветви. Видно, что атомы действительно смещаются в противоположных направлениях.

Согласно (3.18)

$$\zeta_1(0) - \zeta_2(0) = \frac{M_1 + M_2}{\sqrt{M_1 M_2}} \zeta. \tag{3.19}$$

Интересуясь только длинными волнами, мы можем прямо подставить это выражение в формулу (3.17) — учет зависимости $\zeta_1 - \zeta_2$ от $\mathbf{q}$ дает относительно малые поправки. Подставляя, далее, $\varphi(\mathbf{r})$ в формулу (3.14), находим окончательно

$$H'_{\text{пол}} = \pm \frac{4\pi i e^2 Z}{V_0} \sum_{\mathbf{q}} \left[\frac{\hbar}{2MG\omega(\mathbf{q})}\right]^{1/2} \frac{M_1 + M_2}{\sqrt{M_1 + M_2}} \cdot \frac{1}{q} \{b(\mathbf{q}) e^{i\mathbf{qr}} - b^*(\mathbf{q}) e^{-i\mathbf{qr}}\}. \tag{3.20}$$

Здесь принято во внимание, что для продольных фононов $(\zeta, \mathbf{q}) = q$.

Взаимодействие электронов с колебаниями решетки в гетерополярных кристаллах может быть обусловлено не только переменной поляризацией среды, но и потенциалом деформации. Формулы (3.1), (3.5), (3.10) при этом остаются в силе, ибо при выводе их мы не делали явных предположений о природе кристалла. Однако при не слишком малом эффективном заряде Z «поляризационное» взаимодействие (3.20) оказывается более эффективным, чем «деформационное»: силы, ответственные за возникновение энергии взаимодействия (3.20), дальнодействующие, в результате чего на электрон одновременно оказывают влияние изменения дипольного момента во многих элементарных ячейках.

д. Взаимодействие носителей заряда с пьезоэлектрическими колебаниями решетки. В ряде кристаллов механическая деформация сопровождается их электрической поляризацией и возникновением электрического поля. Такие кристаллы называются пьезоэлектрическими. Пьезоэлектрический эффект объясняется тем, что при механической деформации кристалла, решетка которого содержит разноименно заряженные ионы, отдельные подрешетки из ионов одного и того же знака могут смещаться друг относительно друга, отчего и возникает электрический момент. К числу пьезоэлектриков относятся некоторые кристаллы соединений $A^{II}B^{VI}$ (например, CdS, CdSe), $A^{III}B^{V}$ (GaAs, InSb) и многие другие. Необходимое (но не достаточное) условие существования этого эффекта состоит в отсутствии у данной решетки центра симметрии *). При колебаниях

*) Перечень кристаллических классов, в которых возможен пьезоэлектрический эффект, можно найти в руководствах по кристаллофизике [3, М4].

решетки в пьезоэлектрических кристаллах на носители заряда действуют силы со стороны пьезоэлектрического поля, что приводит к дополнительному (к ранее рассмотренным) механизму рассеяния.

Для количественного описания пьезоэлектрического эффекта мы будем пользоваться двумя основными соотношениями.

Рассмотрим сначала простейший случай, когда эффект определяется только одной компонентой тензора деформации u. Направления электростатической индукции $\mathfrak{D}$ и электрического поля $\mathfrak{E}$ будем считать совпадающими. При малых деформациях возникающий электрический момент можно считать пропорциональным u. Полагая его равным $-\beta u$, получаем первое основное соотношение:

$$\mathfrak{D} = \varepsilon \mathfrak{E} - 4\pi\beta u. \tag{3.21}$$

Здесь β — пьезоэлектрический модуль данного кристалла.

Второе основное соотношение описывает так называемый обратный пьезоэлектрический эффект. Он заключается в том, что в свободном (незакрепленном) пьезоэлектрическом кристалле в электрическом поле появляется дополнительная деформация, пропорциональная первой степени поля. В случае же несвободного кристалла в нем возникают дополнительные механические напряжения. В рассматриваемом простейшем случае механическое напряжение s выражается формулой

$$s = \Lambda u + \beta \mathfrak{E}, \tag{3.22}$$

где Λ — модуль упругости.

В общем случае произвольной деформации в кристалле произвольной симметрии вместо соотношения (3.21) мы получим линейное соотношение между компонентами рассматриваемых величин в виде

$$\mathfrak{D}_\alpha = \varepsilon_{\alpha\beta}\mathfrak{E}_\beta - 4\pi\beta_{\alpha,\mu\nu}u_{\mu\nu}. \tag{3.21'}$$

Здесь, как и всюду выше, индексы α, β, μ, ν принимают значения x, y, z, а по дважды встречающимся индексам подразумевается суммирование. Вместо единственной величины β теперь входит совокупность величин $\beta_{\alpha,\mu\nu}$, образующих тензор пьезоэлектрических модулей 3-го ранга; этот тензор симметричен по двум последним индексам: $\beta_{\alpha,\mu\nu} = \beta_{\alpha,\nu\mu}$.

Аналогично, вместо (3.22) в общем случае справедливо соотношение

$$s_{\mu\nu} = \Lambda_{\alpha\beta;\mu\nu}u_{\alpha\beta} + \beta_{\alpha,\mu\nu}\mathfrak{E}_\alpha, \tag{3.22'}$$

где $\Lambda_{\alpha\beta;\mu\nu}$ — совокупность компонент тензора упругости.

Отметим, что оба соотношения (3.21') и (3.22') взаимно связаны и в них входит один и тот же тензор пьезоэлектрических модулей $\beta_{\alpha,\mu\nu}$. Это становится понятным, если рассмотреть выражение для свободной энергии деформированного пьезоэлектрического кристалла. Если $\mathcal{F}_0$ есть свободная энергия недеформированного кристалла, то при наличии деформации к ней добавляются два слагаемых:

энергия механической деформации и энергия поляризованного диэлектрика в электрическом поле. Записывая компоненты вектора поляризации в виде $-\beta_{\alpha,\,\mu\nu}u_{\mu\nu}$, т. е. исходя из соотношения (3.21′), мы имеем

$$\mathscr{F} = \mathscr{F}_0 + \frac{1}{2}\Lambda_{\alpha\beta;\,\mu\nu}u_{\alpha\beta}u_{\mu\nu} + \beta_{\alpha,\mu\nu}u_{\mu\nu}\mathcal{E}_\alpha.$$

Так как компоненты механического напряжения определяются по формуле

$$s_{\mu\nu} = \left(\frac{\partial\mathscr{F}}{\partial u_{\mu\nu}}\right)_{T,\mathcal{E}},$$

то отсюда и получается соотношение (3.22′).

Число независимых компонент тензоров $\varepsilon_{\alpha\beta}$, $\beta_{\alpha,\,\mu\nu}$ и $\Lambda_{\alpha\beta;\,\mu\nu}$ определяется симметрией кристалла. Рассмотрим кубические кристаллы типа цинковой обманки (к числу их относятся, например, соединения типа $A^{III}B^{V}$). В них имеются три зеркально поворотные оси симметрии четвертого порядка, которые мы и выберем в качестве осей координат. Тогда ([M13], §§ 13, 17) мы имеем случай (3.21): тензор $\varepsilon_{\alpha\beta}$ вырождается в скаляр $\varepsilon\delta_{\alpha\beta}$, а пьезоэлектрический тензор имеет лишь одну независимую компоненту

$$\beta_{x,yz} = \beta_{z,xy} = \beta_{y,zx} = \beta; \qquad \beta_{\lambda,\mu\nu} = 0 \quad \text{при } \mu = \lambda \quad \text{или } \nu = \lambda. \tag{3.23}$$

Вернемся теперь к задаче о рассеянии.

В рассматриваемом случае оператор энергии взаимодействия носителей заряда с фононами имеет вид

$$H'_{\text{пьезо}} = \pm e\varphi. \tag{3.24}$$

Здесь верхний знак относится к дыркам, а нижний — к электронам; через φ обозначен потенциал пьезоэлектрического поля, связанный с напряженностью поля $\mathcal{E}$ обычным равенством

$$\mathcal{E} = -\nabla\varphi. \tag{3.25}$$

Пусть ρ есть плотность объемного заряда. Тогда, пользуясь уравнением Пуассона

$$\operatorname{div}\mathfrak{D} = 4\pi\rho$$

и соотношением (3.21′), можем написать

$$\varepsilon_{\alpha\beta}\frac{\partial^2\varphi}{\partial x_\alpha\,\partial x_\beta} + 4\pi\beta_{\alpha,\mu\nu}\frac{\partial u_{\mu\nu}}{\partial x_\alpha} = -4\pi\rho. \tag{3.26}$$

В рассматриваемом нами технологически однородном кристалле локальная нейтральность может быть нарушена только за счет пространственного перераспределения свободных носителей заряда в поле пьезоэлектрической волны. Последнее есть не что иное, как экранирование пьезоэлектрического поля. При не слишком большой концентрации свободных носителей заряда этот эффект

не играет роли и правую часть уравнения (3.26) можно заменить нулем.

Ограничимся рассмотрением пьезоэлектрических колебаний акустических ветвей. Тогда согласно (XII.5.18) в предельном случае длинных волн мы получим

$$u_{\mu\nu} \equiv \frac{1}{2}\left(\frac{\partial Q_\mu}{\partial x_\nu} + \frac{\partial Q_\nu}{\partial x_\mu}\right) = i \sum_{\mathbf{q}} \sum_{s=1}^{3} \left[\frac{\hbar}{2MG\omega(\mathbf{q}, s)}\right]^{1/2} \times$$

$$\times \{(\zeta_\mu q_\nu + \zeta_\nu q_\mu)\, b(\mathbf{q}, s)\, e^{i\mathbf{qr}} - (\zeta_\mu^* q_\nu + \zeta_\nu^* q_\mu)\, b^*(\mathbf{q}, s)\, e^{-i\mathbf{qr}}\}. \quad (3.27)$$

Представим потенциал φ в виде разложения

$$\varphi(\mathbf{r}) = \sum_{\mathbf{q}} \sum_{s=1}^{3} (\varphi_{\mathbf{q}s} e^{i\mathbf{qr}} + \varphi_{\mathbf{q}s}^* e^{-i\mathbf{qr}}). \quad (3.28)$$

Подставляя выражения (3.27) и (3.28) в уравнение (3.26) (при $\rho = 0$), находим

$$\varphi(\mathbf{q}, s) = -\frac{4\pi\beta \cdot \mu\nu}{\varepsilon_{\alpha\beta} q_\alpha q_\beta} q_\lambda (\zeta_\mu q_\nu + \zeta_\nu q_\mu)\, b(\mathbf{q}, s) \left[\frac{\hbar}{2MG\omega(\mathbf{q}, s)}\right]^{1/2}. \quad (3.29)$$

Специализируем это выражение на случай кристаллов типа $A^{III}B^{V}$, когда справедливы равенства (3.23).

Обозначим через θ и φ полярные углы вектора $\mathbf{q}$ относительно оси Z и введем функцию

$$f(\theta, \varphi) = \zeta_x \sin 2\theta \cdot \sin\varphi + \zeta_y \sin 2\theta \cdot \cos\varphi + \zeta_z \sin^2\theta \cdot \sin 2\varphi. \quad (3.30)$$

Тогда равенство (3.29) (с учетом (3.23)) принимает вид

$$\varphi(\mathbf{q}, s) = -\frac{4\pi\beta}{\varepsilon} f(\theta, \varphi)\, b(\mathbf{q}, s) \left[\frac{\hbar}{2MG\omega(\mathbf{q}, s)}\right]^{1/2}. \quad (3.29')$$

Подставляя это выражение в (3.24), мы получаем

$$H'_{\text{пьезо}} = \mp \frac{4\pi\beta e}{\varepsilon} \sum_{\mathbf{q}} \sum_{s=1}^{3} \left[\frac{\hbar}{2MG\omega(\mathbf{q}, s)}\right]^{1/2} \times$$

$$\times \{f(\theta, \varphi)\, b(\mathbf{q}, s)\, e^{i\mathbf{qr}} + f^*(\theta, \varphi)\, b^*(\mathbf{q}, s)\, e^{-i\mathbf{qr}}\}. \quad (3.31)$$

Легко написать и выражение для гамильтониана взаимодействия носителей заряда с пьезоэлектрическими колебаниями оптических ветвей. Для этой цели надо лишь подставить в общую формулу (XII.5.19) выражения для $\zeta_\nu(\mathbf{q}, s)$, вытекающие из соотношений (XII.2.19) и (XII.3.7) при $\mathbf{q} = 0$.

е. Сводка формул. Выражения (3.7), (3.10), (3.20) и (3.31) можно записать единым образом:

$$H' = \sum_{\mathbf{q}, s} \{H'(\mathbf{q}, s)\, b(\mathbf{q}, s)\, e^{i\mathbf{qr}} + H'^*(\mathbf{q}, s)\, b^*(\mathbf{q}, s)\, e^{-i\mathbf{qr}}\}, \quad (3.32)$$

где

$$H'(\mathbf{q}, s) = \left[\frac{\hbar}{2MG\omega(\mathbf{q}, s)}\right]^{1/2} B(\mathbf{q}, s), \quad (3.32')$$

а через B (q, s) обозначена функция, характеризующая данный тип взаимодействия. Выражения для нее, равно как и значения s, по которым следует производить суммирование, приведены в таблице 14.1.

Таблица 14.1

Коэффициенты B(q, s) для разных механизмов рассеяния

	Акустические фононы (потенциал деформации)	Акустические фононы (пьезоэлектрический потенциал)	Неполярные оптические фононы	Поляризационные (продольные) фононы
B (q, s)	$iE_{\alpha\beta}(s)\, q_\alpha \zeta_\beta$ $(iE_1(\mathbf{q}\xi))$ *)	$\mp \frac{4\pi\beta e}{\varepsilon} f(\theta, \varphi)$	$E^{(0)}_{\alpha\beta} n_\beta \zeta_\alpha$ $(E_0(\mathbf{n}\xi))$ *)	$\frac{4\pi i Z e^2 (M_1+M_2)}{qV_0 \sqrt{M_1 M_2}}$
Значения s	1, 2, 3	1, 2, 3	4, 5, ...	4

*) В скобках во втором и четвертом столбцах указаны упрощенные выражения, соответствующие формулам (3.3) и (3.9).

Выражение, стоящее в четвертом столбце, могло бы описывать и взаимодействие носителя заряда с фононами, ответственными за переходы между различными «долинами» в полупроводнике с несколькими эквивалентными минимумами энергии — в n-Ge, n-Si и др. Следует лишь заменить величину $E^{(0)}_{\alpha\beta} n_\beta \zeta_\alpha$ другой — также определяемой из опыта — константой размерности энергии и разрешить значку s принимать все значения, начиная с единицы. Действительно, точки в зоне Бриллюэна, отвечающие эквивалентным минимумам энергии, обычно отстоят друг от друга на расстояние, сравнимое с размерами самой зоны — близкое к постоянной обратной решетки. Следовательно и квазиволновые векторы фононов, испускаемых или поглощаемых при переходах электронов из одной «долины» в другую, должны быть по модулю порядка этой постоянной. Как мы видели в гл. XII, при этом различие между оптическими и акустическими фононами в значительной мере теряется.

§ 4. Рассеяние носителей заряда фононами

Согласно (2.19) вычисление коэффициентов $\mathscr{P}_1$, $\mathscr{P}_2$ сводится к расчету матричного элемента

$$(\mathbf{p}',\ l,\ n' \mid H \mid \mathbf{p},\ l,\ n) = \int d\mathbf{r} \int \prod_{\mathbf{q},\ s} dx_{\mathbf{q}s} \psi^*_{\mathbf{p}',\ l,\ n'} H' \psi_{\mathbf{p},\ l,\ n}. \qquad (4.1)$$

Как и в § 2, обозначим совокупность чисел $\mathbf{p}$, l, n ($\mathbf{p}'$, l', n') через λ (λ'),

Поскольку функции ψ_λ и $\psi_{\lambda'}$ описывают газ фононов и электрон, не взаимодействующие друг с другом, мы имеем

$$\psi_\lambda = \psi_{\mathbf{p}}(\mathbf{r})\,\Phi_n. \tag{4.2}$$

Функция $\psi_{\mathbf{p}}(\mathbf{r})$ есть не что иное, как функция Блоха, описывающая поведение электрона в идеальной решетке, а Φ_n — волновая функция системы независимых гармонических осцилляторов (XII.5.5).

Подставляя выражения (3.32), (3.32′) и (4.2) в правую часть (4.1), мы получаем

$$(\mathbf{p}',\ l,\ n' \mid H' \mid \mathbf{p},\ l,\ n) = \sum_{\mathbf{q},\ s} [H'(\mathbf{q},\ s)\mathcal{J}_1 + H'^{*}(\mathbf{q},\ s)\mathcal{J}_2], \tag{4.3}$$

где

$$\mathcal{J}_1 = \int \psi^{*}_{\mathbf{p}'} e^{i\mathbf{qr}} \psi_{\mathbf{p}}\, d\mathbf{r} \int \Phi_{n'} b(\mathbf{q},\ s)\,\Phi_n \prod_{\mathbf{q}'',\ s''} dx_{\mathbf{q}''s''}, \tag{4.4а}$$

$$\mathcal{J}_2 = \int \psi^{*}_{\mathbf{p}'} e^{-i\mathbf{qr}} \psi_{\mathbf{p}}\, d\mathbf{r} \int \Phi_{n'} b^{*}(\mathbf{q},\ s)\,\Phi_n \prod_{\mathbf{q}'',\ s''} dx_{\mathbf{q}''s''}. \tag{4.4б}$$

Каждая из величин $\mathcal{J}_1$, $\mathcal{J}_2$ представляет собой произведение интегралов по координатам электрона и по нормальным координатам решетки. Последние легко вычисляются с помощью соотношений (XII.5.16); с точностью до множителя $2\left[\frac{\hbar}{2M\omega}\right]^{1/2}$ это — как раз интегралы (XII.5.12). Для вычисления интегралов по $\mathbf{r}$ заметим, что в интересующем нас случае длинных волн решетки оператор (3.32) представляет собой возмущение, лишь очень медленно изменяющееся на протяжении постоянной решетки a. Следовательно, мы вправе воспользоваться методом эффективной массы, заменяя функцию Блоха $\psi_{\mathbf{p}}$ плоской волной

$$\psi_{\mathbf{p}}(\mathbf{r}) = V^{-1/2} e^{\frac{i}{\hbar}\mathbf{pr}}. \tag{4.5}$$

Множитель $V^{-1/2}$ в (4.5) выбран в соответствии с условием нормировки (2.2).

Подставляя (4.5) в (4.4а, б) и принимая во внимание равенства (XII.5.12) и (XII.5.14), мы получаем

$$\mathcal{J}_1 = \delta_{\mathbf{p}+\hbar\mathbf{k},\,\mathbf{p}'} \sqrt{n(\mathbf{q},\ 1)} \prod_{\mathbf{q}' \neq \mathbf{q}} \prod_{s' \neq 1} \delta_{n'(\mathbf{q}',\ s'),\ n(\mathbf{q}',\ s')} \delta_{n'(\mathbf{q},\ 1),\ n(\mathbf{q},\ 1)-1}, \tag{4.6а}$$

$$\mathcal{J}_2 = \delta_{\mathbf{p}-\hbar\mathbf{k},\,\mathbf{p}'} \sqrt{n(\mathbf{q},\ 1)+1} \prod_{\mathbf{q}' \neq \mathbf{q}} \prod_{s' \neq 1} \delta_{n'(\mathbf{q}',\ s'),\ n(\mathbf{q}',\ s')} \delta_{n'(\mathbf{q},\ 1),\ n(\mathbf{q},\ 1)+1}. \tag{4.6б}$$

Видно, что матричные элементы оператора H' (а с ними и вероятность перехода) отличны от нуля, только если числа фононов $n(\mathbf{q},\ 1)$ в начальном и конечном состояниях различны. Иначе говоря, процесс взаимодействия электронов с фононами состоит в испускании и поглощении последних. При этом в принятом приближении

(первое приближение теории возмущений и линейная по нормальным координатам форма оператора энергии взаимодействия) возможны только однофононные процессы: в каждом акте взаимодействия испускается или поглощается только один фонон. Очевидно, слагаемое с $\mathcal{J}_1$ соответствует рассеянию с поглощением, а с $\mathcal{J}_2$ — рассеянию с испусканием фонона.

Видно, далее, что вероятность перехода отлична от нуля, только если имеют место соотношения

$$\mathbf{p} - \mathbf{p}' = -\hbar\mathbf{q} \tag{4.7а}$$

в случае поглощения фонона и

$$\mathbf{p} - \mathbf{p}' = \hbar\mathbf{q} \tag{4.7б}$$

в случае испускания фонона.

Эти соотношения представляют собой не что иное, как законы сохранения: рассеяние с участием фонона возможно, только если квазиимпульс электрона изменяется на вполне определенную величину $\pm\hbar\mathbf{q}$, связанную формулой де Бройля с квазиволновым вектором фонона $\mathbf{q}$. Это дает основание назвать вектор $\hbar\mathbf{q}$ квазиимпульсом фонона.

В § XII.6 было показано, что фононам следует приписать энергию $\hbar\omega(\mathbf{q}, s)$. Видим теперь, что зависимость $\omega(\mathbf{q}, s)$ можно рассматривать как закон дисперсии для фононов. Тем самым завершается обоснование представления о фононах как о квазичастицах.

В связи с соотношениями (4.7а, б) следует сделать три замечания.

Во-первых, вывод этих законов сохранения не связан непременно с методом эффективной массы. Как можно показать (см. Приложение XIV), тот же результат получится и при использовании точных функций Блоха $\psi_{\mathbf{p}}$ вместо плоских волн (4.5). При этом следует лишь помнить, что векторы $\mathbf{q}$ и $\mathbf{p}/\hbar$ определены с точностью до вектора обратной решетки $\mathbf{b}$. Поэтому и законы сохранения (4.7а, б) надо понимать как равенства по модулю $\hbar\mathbf{b}$. Процессы рассеяния, при которых $\mathbf{p} - \mathbf{p}' = -\hbar\mathbf{q} + \hbar\mathbf{b}$ или $\mathbf{p} - \mathbf{p}' = \hbar\mathbf{q} + \hbar\mathbf{b}$, называются *процессами переброса*. Учет их существен, если вектор $\mathbf{p} - \mathbf{p}'$ может выйти за пределы первой зоны Бриллюэна — тогда добавление слагаемого $\hbar\mathbf{b}$ вновь приводит вектор $\hbar\mathbf{q}$ в первую зону. В дальнейшем мы не будем явно отмечать это обстоятельство. В принятых ранее условиях (минимум энергии носителей заряда расположен в центре зоны Бриллюэна) процессы переброса в интересующих нас задачах оказываются несущественными.

Во-вторых, подчеркнем, что недопустимо смешивать квазиимпульс с импульсом: $\hbar\mathbf{q}$ не есть собственное значение оператора импульса решетки.

Наконец, в-третьих, видно, что длины волн фононов, испускаемых или поглощаемых при рассеянии, — порядка длины волны электрона. Большинство таких фононов (в случае невырожденного электрон-

ного газа) будет иметь длину волны порядка $h/\sqrt{mkT}$ ($\sim 10^{-6}$ см при обычных значениях температуры и эффективной массы). Эта величина значительно превышает постоянную решетки, что и оправдывает сделанное в § 3 предположение о доминирующей роли длинноволновых фононов в процессе взаимодействия с носителями заряда.

Обратимся теперь к вычислению вероятности перехода $\mathscr{P}_2(\mathbf{p}, \mathbf{p}')$. Согласно (4.3) и (4.6а, б)

$$|(\mathbf{p}', l, n'|H'|\mathbf{p}, l, n)|^2 = \\ = |H'(\mathbf{q}, 1)|^2 \prod_{\substack{\mathbf{q}', s' \\ (\neq \mathbf{q}, 1)}} \delta_{n'(\mathbf{q}', s'), n(\mathbf{q}', s')} \{n(\mathbf{q}, 1)\, \delta_{n'(\mathbf{q}, 1), n(\mathbf{q}, 1)-1} + \\ + [n(\mathbf{q}, 1)+1]\, \delta_{n'(\mathbf{q}, 1), n(\mathbf{q}, 1)+1}\}, \quad (4.8)$$

причем под $\mathbf{q}$ следует понимать $-\frac{\mathbf{p}-\mathbf{p}'}{\hbar}$ в первом слагаемом и $\frac{\mathbf{p}-\mathbf{p}'}{\hbar}$ — во втором. Таким образом, вероятность перехода с данным изменением квазиимпульса электрона $\mathbf{p}-\mathbf{p}'$ складывается из вероятностей поглощения и испускания фонона:

$$\mathscr{P}_2(\mathbf{p}, \mathbf{p}') = \mathscr{P}_{\text{погл}} + \mathscr{P}_{\text{исп}}. \quad (4.9)$$

Величины $\mathscr{P}_{\text{погл}}$ и $\mathscr{P}_{\text{исп}}$ легко найти, подставляя (4.8) в формулу (2.19). Для аргументов дельта-функции в (2.19) мы получаем при этом, согласно (XII.6.2),

$$E(\mathbf{p}', l, n') - E(\mathbf{p}, l, n) = \\ = E(\mathbf{p}', l) - E(\mathbf{p}, l) + \sum_{\mathbf{q}', s'} \hbar\omega(\mathbf{q}', s')[n'(\mathbf{q}', s') - n(\mathbf{q}', s')] = \quad (4.10)$$

$$= E(\mathbf{p}', l) - E(\mathbf{p}, l) \pm \hbar\omega\left(\pm\frac{\mathbf{p}-\mathbf{p}'}{\hbar}, s\right). \quad (4.10')$$

Верхний и нижний знаки здесь относятся, соответственно, к испусканию и поглощению фонона. Видно, что испусканию или поглощению фонона с квазиимпульсом $\hbar\mathbf{q} = \pm(\mathbf{p}-\mathbf{p}')$ соответствует уменьшение или увеличение энергии электрона — как раз на величину $\hbar\omega(\mathbf{q}, s)$. Так, конечно, и должно было получиться.

При рассеянии на оптических фононах частота близка к предельному значению ω_0 (§ XII.3). Предположение об упругости рассеяния выполняется при этом, лишь если средняя энергия носителей заряда достаточно велика. Так, в случае невырожденного газа должно выполняться неравенство *)

$$kT \gg \hbar\omega_0. \quad (4.11)$$

*) Условие (4.11) может оказаться очень жестким, ибо значения $\hbar\omega_0$ могут составить несколько сотых эВ.

В противном случае рассеяние становится существенно неупругим и задача о решении кинетического уравнения заметно усложняется (так обычно обстоит дело в случае междолинного рассеяния).

Положение вновь упрощается лишь при достаточно низких температурах, когда $\hbar\omega_0 \gg kT$. При этом практически все электроны могут только поглощать оптические фононы. Поглотив такой фонон, электрон, однако, почти мгновенно вновь испускает его: согласно принципу детального равновесия (XIII.4.4) отношение вероятностей испускания и поглощения фонона составляет $\exp(\hbar\omega_0/kT)$. Этот процесс называется составным рассеянием. В результате энергия электрона изменяется лишь незначительно. В самом деле, если бы частота оптического фонона ω вообще не зависела от его квазиволнового вектора, энергия электрона, испытавшего составное рассеяние, осталась бы неизменной. За счет слабой зависимости ω от $\mathbf{q}$ изменение энергии все же имеет место, но оно оказывается сравнительно небольшим: составное рассеяние — почти упругое. Это позволяет вновь ввести время релаксации [М7]; очевидно, оно пропорционально $\exp(-\hbar\omega_0/kT)$. Другая возможность, также возникающая при низких температурах, рассматривается в конце этого параграфа.

С другой стороны, при рассеянии на акустических фононах рассеяние оказывается почти упругим. Действительно, в этом случае для частоты $\omega(\mathbf{q}, s)$ можно воспользоваться формулой (XII.3.3′). Соответственно выражение (4.10′) принимает вид

$$E(\mathbf{p}') - E(\mathbf{p}) \pm c_s |\mathbf{p} - \mathbf{p}'|, \qquad (4.10'')$$

где c_s — фазовая скорость продольных ($s = 1$) и поперечных ($s = 2, 3$) звуковых волн.

Пусть газ носителей заряда не вырожден. Тогда $|\mathbf{p} - \mathbf{p}'| \sim \sqrt{mkT}$ и, следовательно,

$$\frac{c_s |\mathbf{p} - \mathbf{p}'|}{kT} \sim c_s \sqrt{\frac{m}{kT}} \equiv \frac{c_s}{v_T}, \qquad (4.12)$$

где $v_T = \sqrt{kT/m}$ — величина порядка средней тепловой скорости носителя заряда. Как правило, $v_T \sim 10^7$ см/с, в то время как $c_s \sim 5 \cdot 10^5$ см/с: электроны движутся в среднем гораздо быстрее звука. Причина этого ясна: распространение звуковой волны в кристалле связано со смещениями тяжелых частиц — атомов решетки. Полагая для оценки $m = m_0$, видим, что скорости c_s и v_T сравниваются при температуре $\simeq 3$ К. При более высоких температурах последним слагаемым в (4.10″) можно пренебречь: благодаря малости отношения (4.12) изменение энергии электрона при испускании или поглощении акустического фонона, как правило, мало́ по сравнению с самой этой энергией. При этом аргументы дельта-функций в формулах для $\mathscr{T}_{\text{погл}}$ и $\mathscr{T}_{\text{исп}}$ становятся одинаковыми и равными просто разности энергий электрона в начальном и конечном состояниях.

В условиях почти упругого рассеяния мы получаем, комбинируя формулы (4.8) и (2.19):

$$\mathscr{P}_{\text{погл}} = \frac{V}{G} \cdot \frac{\pi |B(\mathbf{q}, s)|^2}{M\omega(\mathbf{q}, s)} n(\mathbf{q}, s)\, \delta(E(\mathbf{p}) - E(\mathbf{p}')), \tag{4.13а}$$

$$\mathscr{P}_{\text{исп}} = \frac{V}{G} \cdot \frac{\pi |B(\mathbf{q}, s)|^2}{M\omega(\mathbf{q}, s)} [n(\mathbf{q}, s) + 1]\, \delta(E(\mathbf{p}) - E(\mathbf{p}')), \tag{4.13б}$$

причем $\hbar q = |\mathbf{p} - \mathbf{p}'|$.

Отношение $V/G = V_0$ представляет собой, очевидно, не что иное, как объем элементарной ячейки. При $G \to \infty$, $V \to \infty$ эта величина (а с ней и вероятность перехода) остается конечной (и не зависит от V), как это и должно быть. Заметим, что $V_0/M = \rho^{-1}$, где ρ — плотность кристалла.

Как видно из формул (4.13а, б), не только вероятность поглощения, но и вероятность испускания фонона зависит от числа фононов, уже имеющихся в образце. Это есть аналог хорошо известного в оптике явления — вынужденного излучения света (см. ниже § XVIII.4).

При $n(\mathbf{q}, s) \to 0$ вероятность поглощения фонона, естественно обращается в нуль, но вероятность испускания остается конечной благодаря наличию второго слагаемого в квадратных скобках в формуле (4.13б). Это есть аналог спонтанного излучения фотонов (ср. гл. XVIII). Физическая причина его состоит во взаимодействии носителей заряда с нулевыми колебаниями решетки.

Числа заполнения фононных состояний $n(\mathbf{q}, s)$, входящие в формулы (4.13а, б), строго говоря неизвестны: для них надо было бы составить кинетическое уравнение типа (XIII.3.12). Мы получили бы тогда систему кинетических уравнений для функций $f(\mathbf{p})$ и $n(\mathbf{q}, s)$. Заметим, однако, что величины $\mathscr{P}_1$ и $\mathscr{P}_2$ входят в интеграл столкновений, где они умножаются на неравновесную часть функции распределения электронов f_1. Можно думать поэтому, что хорошее приближение получится, если взять для $n(\mathbf{q}, s)$ равновесное выражение (XII.6.3) *). При этом, в силу (4.11) и (4.12), функцию Планка (XII.6.3) можно аппроксимировать классическим выражением, отвечающим равнораспределению энергии по степеням свободы:

$$n[\omega(\mathbf{q}, s)] \simeq \frac{kT}{\hbar\omega(\mathbf{q}, s)}. \tag{4.14}$$

*) Это рассуждение может оказаться несправедливым при наличии в кристалле градиента температуры. Действительно, при этом возникает поток фононов, направленный от «горячего» конца к «холодному». Сталкиваясь с носителями заряда, фононы увлекают их за собой, что может заметно увеличить термоэдс. Этот эффект увлечения был впервые теоретически исследован в 1945 г. Л. Э. Гуревичем. Расчет показывает, что в полупроводниках он может быть важен при низких температурах.

Формулы (4.13а, б) и (4.9) дают теперь

$$\mathscr{P}_2 = \mathscr{P}_1 \equiv \mathscr{P} = \frac{2\pi V_0 kT}{M\omega^2(\mathbf{q}, s)\hbar} |B(\mathbf{q}, s)|^2 \delta[E(\mathbf{p}) - E(\mathbf{p}')]. \quad (4.15)$$

Формула (4.15) становится несправедливой при достаточно низких температурах ($T \ll T_D$ — дебаевской температуры), когда энергия фонона оказывается больше kT и $n(\mathbf{q}, s) \to 0$. В этом случае $\mathscr{P}_{\text{погл}} \to 0$, а для $\mathscr{P}_{\text{исп}}$ мы получаем

$$\mathscr{P}_{\text{исп}} \equiv \mathscr{P} = \frac{2\pi V_0}{M\omega(\mathbf{q}, s)} |B(\mathbf{q}, s)|^2 \delta[E(\mathbf{p}) - E(\mathbf{p}') - \hbar\omega(\mathbf{q}, s)], \quad (4.15')$$

причем $\hbar q = |\mathbf{p} - \mathbf{p}'|$ (разумеется, $E(\mathbf{p}) > \hbar\omega(\mathbf{q}, s)$). Выражение (4.15') характеризует вероятность рассеяния электронов нулевыми колебаниями решетки.

Значения $B(\mathbf{q}, s)$ для разных механизмов рассеяния приведены в таблице 14.1. Из нее явствует, что правая часть (4.15) действительно имеет вид (XIII.6.1), (XIII.6.5), коль скоро речь идет о рассеянии на оптических фононах. Так же обстоит дело и при рассеянии на акустических фононах за счет потенциала деформации, если используется «изотропная» аппроксимация (3.3). С другой стороны, вероятность рассеяния, обусловленного пьезоэлектрическим потенциалом, оказывается существенно анизотропной: согласно (3.30) функция $f(\theta, \varphi)$ существенно зависит от ориентации вектора $\mathbf{q}$ относительно кристаллографических осей. При расчетах ориентировочного характера иногда пользуются упрощенным выражением для $\mathscr{P}_{\text{пьезо}}$, заменяя функцию $f^2(\theta, \varphi)$ средним ее значением

$$\langle f^2(\theta, \varphi)\rangle = \frac{1}{4\pi}\int_0^{2\pi} d\varphi \int_0^{\pi} \sin\theta \cdot f^2(\theta, \varphi)\, d\theta. \quad (4.16)$$

Подставляя сюда выражение (3.30) и выполняя интегрирование, мы получаем

$$\langle f^2(\theta, \varphi)\rangle = 4/15.$$

Пользуясь этим выражением и сравнивая равенства (4.15) и (XIII.6.1), (XIII.6.5), можем вычислить функцию $S(E, \cos\theta)$, фигурирующую в формуле для времени релаксации (XIII.6.12). Ограничимся сначала параболическим законом дисперсии, совмещая по-прежнему начало отсчета энергии электронов с дном зоны проводимости ($E_c = 0$). Получим

$$S(E, \cos\theta) = BE^{-r-1/2}(1 - \cos\theta)^{r'}. \quad (4.17)$$

Здесь, как и в гл. XIII, $\theta = \widehat{\mathbf{p}, \mathbf{p}'}$, есть угол рассеяния; значения постоянных B, r и r' для разных механизмов рассеяния указаны в таблице 14.2.

Таблица 14.2

Значения постоянных B, r, r' и C для разных механизмов рассеяния

	B	r	r'	$C^{-1} = \frac{2B\sqrt{2m^3}}{(2\pi)^2\hbar^3}$
Акустические фононы (потенциал деформации)	$\frac{2\pi V_0 kTE_1^2}{Mc_s^2\hbar}$	$-1/2$	0	$\frac{V_0 kTE_1^2\sqrt{2m^3}}{\pi Mc_s^2\hbar^4}$
Акустические фононы (пьезоэлектрическое рассеяние)	$\frac{32\pi^3\beta^2e^2\hbar V_0 kT}{15Mc_s^2\varepsilon^2 m}$	$+1/2$	-1	$\frac{16\pi\beta^2e^2V_0kT\sqrt{2m}}{15\varepsilon^2Mc_s^2\hbar^2}$
Неполярные оптические фононы *), $kT \gg \hbar\omega_0$	$\frac{2\pi V_0 kTE_0^2}{M\hbar\omega_0^2}$	$-1/2$	0	$\frac{V_0 kTE_0^2\sqrt{2m^3}}{\pi M\hbar^4\omega_0^2}$
Поляризационные фононы, $kT \gg \hbar\omega_0$	$8\pi^3\hbar kT\,(Ze^2)^2 \times \times \frac{(M_1+M_2)}{V_0\omega_0^2 m M_1 M_2}$	$+1/2$	-1	$4\pi kT\,(Ze^2)^2 \times \times \frac{(M_1+M_2)\sqrt{2m}}{V_0 M_1 M_2\hbar^2\omega_0^2}$

*) Принято сокращенное обозначение: $\left| E_{\alpha\beta}^{(0)} n_\beta \xi_\alpha \right|^2 = E_0^2$.

Выражение (4.17) можно обобщить и на случай непараболического (но изотропного) закона дисперсии. Для этой цели надо лишь произвести замену

$$mE \to \frac{1}{2} p^2(E). \tag{4.18}$$

Эффективная масса m при этом выпадает из формул для C (табл. 14.2).

Теперь легко вычислить время релаксации, связанное с рассеянием на фононах какого-либо из рассматриваемых типов. Подставляя выражение (4.17) в правую часть (XIII.6.13) и интегрируя по углу θ, мы получаем формулу (XIII.7.21):

$$\tau(E) = CE^r, \tag{4.19}$$

с известными уже величинами C и r. Значения C указаны в таблице 14.2.

Проверим теперь, выполняется ли основное условие применимости кинетического уравнения, выражаемое в случае невырожденного газа неравенством (2.22а). Очевидно, в качестве τ в это неравенство надо подставить выражение (4.19), взятое для «существенных» электронов, т. е. при $E \sim kT$. Тогда рассматриваемое условие принимает вид

$$\hbar/C\,(kT)^{r+1} \ll 1. \tag{4.20}$$

Подстановка типичных значений параметров ($c_s \simeq 10^5$ см/с, $M \simeq 10^{-22}$ г, $E_1 \simeq 10$ эВ, $V_0 \simeq 10^{-22}$ см^3) показывает, что в случае

акустического рассеяния условие (4.20) обычно выполняется. С другой стороны, при рассеянии на поляризационных оптических колебаниях неравенство (4.20) может и нарушаться. Причина этого состоит в сравнительно большой энергии взаимодействия носителей заряда с поляризационными фононами (§ 3, п. г). Связанные с этим осложнения рассматриваются в гл. XVII.

Как и формула (4.15), данные таблицы 14.2 становятся несправедливыми при низких температурах, когда имеет место только рассеяние на нулевых колебаниях решетки. В этом случае неприменима и формула (XIII.6.13), так как рассеяние носит неупругий характер: согласно (4.12) энергия фонона, испускаемого или поглощаемого в процессе рассеяния, оказывается здесь сравнимой с энергией самого носителя заряда. Все же и здесь можно ввести время релаксации квазиимпульса, определив его как характерное «время испускания» фонона:

$$\frac{1}{\tau(E)} = \frac{1}{(2\pi\hbar)^3}\int d\mathbf{p}'\,\mathscr{P}(\mathbf{p},\,\mathbf{p}'). \tag{4.21}$$

В качестве $\mathscr{P}(\mathbf{p},\,\mathbf{p}')$ сюда следует подставить выражение (4.15′). Так, определенная величина есть не что иное, как «время жизни» электрона в состоянии с квазиимпульсом $\mathbf{p}$. Это время, вообще говоря, не совпадает с определенным ранее временем свободного пробега *).

Вычислим $\tau(E)$, считая, что рассеяние обусловлено оптическими колебаниями решетки (поляризационными или неполярными). Ограничимся при этом изотропным параболическим законом дисперсии и пренебрежем зависимостью ω от квазиволнового вектора оптических фононов.

Для неполярных фононов мы получаем, подставляя (4.15′) в (4.21):

$$\frac{1}{\tau(E)} = \sum_s \frac{\sqrt{2m^3}V_0 B_1}{\pi M\hbar^2\sqrt{\hbar\omega_0(s)}}\left[\frac{E}{\hbar\omega_0(s)} - 1\right]^{1/2}. \tag{4.22}$$

Здесь $B_1 = |\,E^{(0)}_{\alpha\beta}(s)\,n_\beta\xi_\alpha\,|^2$. Заметим, что в этом случае «приход» действительно не играет роли и выражение (4.22) дает обычное время релаксации квазиимпульса. Его можно обобщить и на случай произвольных температур (в предположении о тепловом равновесии фононов). Для этой цели надо принять во внимание рассеяние не только на нулевых, но и на тепловых колебаниях решетки. Иначе говоря, надо принять во внимание зависимость времени релаксации от числа фононов $n\,[\omega(\mathbf{q}, s)]$, уже имеющихся в решетке. Как мы

*) Такое совпадение имело бы место, если бы в кинетическом уравнении можно было опустить член «прихода» (XIII.3.7). Действительно, правая часть (4.21) есть не что иное, как интеграл «ухода» B (XIII.3.8) при $f(\mathbf{p}, r) \simeq 1$, $f(\mathbf{p}', r') \ll 1$. Как видно из формулы (XIII.6.12), член «прихода» не играет роли, если $\mathscr{P}(\mathbf{p}, \mathbf{p}')$ не зависит от угла рассеяния.

знаем, вероятности переходов с испусканием и поглощением фононов пропорциональны, соответственно, $n[\omega(\mathbf{q}, s)]+1$ и $n[\omega(\mathbf{q}, s)]$. Таким образом, искомое обобщение получается, если в правую часть (4.22) ввести множитель

$$\frac{2}{\exp\left[\frac{\hbar\omega_0(s)}{kT}\right]-1}+1$$

(в данном случае $\omega(\mathbf{q}, s)=\omega_0(s)$).

Для поляризационных фононов формула (4.15′) с учетом таблицы 14.1 дает

$$\frac{1}{\tau(E)}=$$

$$=\frac{V_0\sqrt{2m}\,B_2}{4\pi M\sqrt{\hbar\omega_0(s)}}\left[\frac{E}{\hbar\omega_0(s)}-1\right]^{1/2}\frac{1}{\sqrt{E(E-\hbar\omega_0(s))}}\ln\frac{\sqrt{E}+\sqrt{E-\hbar\omega_0(s)}}{\sqrt{E}-\sqrt{E-\hbar\omega_0(s)}}. \quad (4.23)$$

Здесь $B_2=\left[\frac{4\pi Ze^2(M_1+M_2)}{V_0\sqrt{M_1M_2}}\right]^2$; суммирование по s отпадает, так как в данном случае значок s отвечает только продольной ветви.

При $E-\hbar\omega_0(s)\ll E$ равенство (4.23) принимает вид, аналогичный (4.22):

$$\frac{1}{\tau(E)}=\frac{V_0\sqrt{2m}\,B_2}{2\pi M\sqrt{\hbar\omega_0(s)}\,E}\left[\frac{E}{\hbar\omega_0(s)}-1\right]^{1/2}. \quad (4.23')$$

Формулы (4.22) и (4.23′) можно переписать в виде

$$\frac{1}{\tau(E)}=\sum_s\frac{1}{\tau_0(s)}\left[\frac{E}{\hbar\omega_0(s)}-1\right]^{1/2}, \quad (4.24)$$

где $\tau_0(s)$ — характерное время, явное выражение для которого вытекает из сопоставления правых частей (4.24) и (4.22), (4.23′).

При $\hbar\omega_0(s)\gg kT$ и равновесном распределении носителей заряда по энергиям число электронов, способных испускать оптические фононы, пропорционально $\exp[-\hbar\omega_0(s)/kT]$. Следовательно, в этих условиях рассеяние, характеризуемое временами релаксации (4.22) и (4.23), столь же маловероятно, сколь и составное. Положение, однако, может измениться в достаточно сильном электрическом поле, когда электронному газу сообщается достаточно большая энергия (гл. XVI).

§ 5. Рассеяние носителей заряда примесными атомами

При рассеянии электронов неподвижными примесными атомами оператор H' имеет вид

$$H'=\sum_{i=1}^{N}\delta U(\mathbf{r}-\mathbf{R}_i). \quad (5.1)$$

Здесь значок i нумерует атомы примеси, $\mathbf{R}_i$ есть радиус-вектор i-го атома, $\delta U(\mathbf{r}-\mathbf{R}_i)$ — потенциальная энергия взаимодействия электрона с ним, N — полное число атомов примеси в кристалле. Состояние решетки в данном случае не изменяется при рассеянии, поэтому сумма по n' в формуле (2.19) исчезает ($n'=n$), и мы получаем, на основании (2.19) и (5.1),

$$\mathscr{P}_2(\mathbf{p},\ \mathbf{p}') = \frac{2\pi}{\hbar} V \left| \int \psi^*_{\mathbf{p}'}(\mathbf{r}) \sum_{i=1}^{N} \delta U(\mathbf{r}-\mathbf{R}_i)\, \psi_{\mathbf{p}}(\mathbf{r})\, d\mathbf{r} \right|^2 \delta(E(\mathbf{p})-E(\mathbf{p}')). \quad (5.2)$$

В рамках метода эффективной массы, которым мы будем пользоваться, функции Блоха $\psi_{\mathbf{p}}$ и $\psi_{\mathbf{p}'}$ следует заменить просто плоскими волнами (4.5).

Как видно из формулы (5.2), рассеяние в данном случае носит упругий характер. Причина этого ясна: мы рассматриваем атомы примеси просто как фиксированные в пространстве силовые центры, пренебрегая возможным их движением. Такая постановка задачи оправдывается большой (по сравнению с m) массой атома примеси M_t.

Подставляя в (5.2) функции $\psi_{\mathbf{p}}$ и $\psi_{\mathbf{p}'}$ в виде (4.5) и возводя в квадрат сумму в (5.2), получаем

$$\mathscr{P}_1 = \mathscr{P}_2 = \frac{2\pi}{\hbar V}(A_1+A_2)\,\delta(E(\mathbf{p})-E(\mathbf{p}')), \quad (5.3)$$

где

$$A_1 = \sum_{j=1}^{N} \left| \int d\mathbf{r}\, \delta U(\mathbf{r}-\mathbf{R}_j) \exp\left[\frac{i}{\hbar}(\mathbf{p}-\mathbf{p}',\ \mathbf{r})\right] \right|^2, \quad (5.4а)$$

$$A_2 = \sum_{j=1}^{N} \sum_{\substack{j'=1 \\ (j'\neq j)}}^{N} \int d\mathbf{r}_1\, \delta U(\mathbf{r}_1-\mathbf{R}_j) \exp\left[\frac{i}{\hbar}(\mathbf{p}-\mathbf{p}',\ \mathbf{r}_1)\right] \times$$
$$\times \int d\mathbf{r}_2\, \delta U(\mathbf{r}_2-\mathbf{R}_{j'}) \exp\left[-\frac{i}{\hbar}(\mathbf{p}-\mathbf{p}',\ \mathbf{r}_2)\right]. \quad (5.4б)$$

Как известно из квантовой механики ([М2], § 95; [М3], § 29б), выражение

$$\sigma(\mathbf{p},\ \mathbf{p}') = \left(\frac{m}{2\pi\hbar^2}\right)^2 \left| \int d\mathbf{r}\, \delta U(\mathbf{r}-\mathbf{R}_j) \exp\left[\frac{i}{\hbar}(\mathbf{p}-\mathbf{p}',\ \mathbf{r})\right] \right|^2 \quad (5.5)$$

представляет собой отнесенное к единице телесного угла эффективное сечение упругого рассеяния электрона силовым центром, расположенным в точке $\mathbf{R}_j$. Следовательно, правая часть (5.4а) описывает рассеяние электрона всеми N примесными атомами при условии, что рассеяние на каждом из них происходит независимо от всех остальных. Иначе говоря, величина A_1 характеризует рассеяние, которое имело бы место в отсутствие интерференции электронных

волн, рассеянных разными атомами примеси. Такое рассеяние называется некогерентным. Легко убедиться, что все члены, входящие в сумму по j в (5.4а), одинаковы. Действительно, произведем в каждом интеграле замену переменных, полагая

$$\mathbf{r} - \mathbf{R}_j = \mathbf{r}'. \tag{5.6}$$

Получим

$$\int d\mathbf{r}\, \delta U(\mathbf{r} - \mathbf{R}_j) \exp\left[\frac{i}{\hbar}(\mathbf{p} - \mathbf{p}',\ \mathbf{r})\right] =$$
$$= \exp\left[\frac{i}{\hbar}(\mathbf{p} - \mathbf{p}',\ \mathbf{R}_j)\right] \int d\mathbf{r}'\, \delta U(\mathbf{r}') \exp\left[\frac{i}{\hbar}(\mathbf{p} - \mathbf{p}',\ \mathbf{r}')\right].$$

Зависимость от значка j входит сюда только через фазовый множитель, который обращается в единицу при составлении модуля. Таким образом,

$$A_1 = N \left| \int d\mathbf{r}'\, \delta U(\mathbf{r}') \exp\left[\frac{i}{\hbar}(\mathbf{p} - \mathbf{p}',\ \mathbf{r}')\right] \right|^2. \tag{5.7}$$

Заметим, что этот результат не зависит от расположения примесных атомов в решетке. Последнее влияет лишь на слагаемое A_2, которое описывает эффект интерференции электронных волн, рассеянных различными атомами примеси. Действительно, произведем в интегралах, фигурирующих в (5.4б), замену переменных

$$\mathbf{r}_1 - \mathbf{R}_j = \mathbf{r}_1', \qquad \mathbf{r}_2 - \mathbf{R}_{j'} = \mathbf{r}_2'.$$

Тогда

$$A_2 = \sum_{\substack{j=1 \\ (j' \neq j)}}^{N} \sum_{j'=1}^{N} \exp\left[\frac{i}{\hbar}(\mathbf{p} - \mathbf{p}',\ \mathbf{R}_j - \mathbf{R}_{j'})\right] \times$$
$$\times \int \delta U(\mathbf{r}_1') \exp\left[\frac{i}{\hbar}(\mathbf{p} - \mathbf{p}',\ \mathbf{r}_1')\right] d\mathbf{r}_1' \times$$
$$\times \int \delta U(\mathbf{r}_2') \exp\left[-\frac{i}{\hbar}(\mathbf{p} - \mathbf{p}',\ \mathbf{r}_2')\right] d\mathbf{r}_2' =$$
$$= \frac{A_1}{N} \sum_{\substack{j,\ j' \\ (j' \neq j)}} \exp\left[\frac{i}{\hbar}(\mathbf{p} - \mathbf{p}',\ \mathbf{R}_j - \mathbf{R}_{j'})\right]. \tag{5.4б'}$$

Вспомним теперь, что величины $\mathscr{P}_1$, $\mathscr{P}_2$ и, следовательно, A_1 и A_2 должны быть вещественными. Составляя полусумму правой части (5.4б') и сопряженного ей выражения, находим

$$A_2 = \frac{A_1}{N} \sum_{\substack{j,\ j' \\ (j' \neq j)}} \cos\left(\frac{\mathbf{p} - \mathbf{p}'}{\hbar},\ \mathbf{R}_j - \mathbf{R}_{j'}\right). \tag{5.8}$$

Зависимость A_2 от расположения примесных атомов весьма осложняет дальнейший расчет, ибо фактические координаты примес-

ных атомов в решетке нам неизвестны. Однако при хаотическом распределении примеси в кристалле положительные и отрицательные слагаемые в правой части (5.8) будут встречаться одинаково часто. Тогда результат интерференции (5.8) обратится в нуль и атомы примеси будут рассеивать электроны и дырки независимо друг от друга *). В этих условиях, согласно (5.3) и (5.7),

$$\mathscr{P}(\mathbf{p}, \mathbf{p}') = \frac{2\pi N}{\hbar V}\left|\int \delta U(\mathbf{r}') \exp\left[\frac{i}{\hbar}(\mathbf{p}-\mathbf{p}', \mathbf{r}')\right] d\mathbf{r}'\right|^2 \delta(E(\mathbf{p}) - E(\mathbf{p}')). \quad (5.9)$$

Отношение $N/V = N_t$ представляет собой концентрацию примеси в решетке. При $V \to \infty$, $N \to \infty$ эта величина (а с ней и вероятность перехода) остается конечной и не зависит от размеров образца.

Выражение (5.9) имеет вид (XIII.6.1), причем

$$S(\mathbf{p}, \mathbf{p}') = N_t\left|\int \delta U(\mathbf{r}') \exp\left[\frac{i}{\hbar}(\mathbf{p}-\mathbf{p}', \mathbf{r}')\right] d\mathbf{r}'\right|^2. \quad (5.10)$$

Если потенциальная энергия взаимодействия $\delta U(\mathbf{r}')$ изотропна, то функция S зависит только от энергии $E(\mathbf{p})$ и от угла $\theta = \widehat{\mathbf{p}, \mathbf{p}'}$ — в соответствии с (XIII.6.5).

Пользуясь соотношением (5.5), можем переписать формулу (5.10) в виде

$$S(\mathbf{p}, \mathbf{p}') = \left(\frac{2\pi\hbar^2}{m}\right)^2 N_t \sigma(\mathbf{p}, \mathbf{p}'). \quad (5.11)$$

При этом для времени релаксации мы получим, согласно (XIII.6.13),

$$\frac{1}{\tau} = \frac{2\pi N_t \sqrt{2E(\mathbf{p})}}{m} \int_0^\pi (1-\cos\theta)\, \sigma(E(\mathbf{p}), \cos\theta) \sin\theta \cdot d\theta. \quad (5.12)$$

Введение эффективного сечения рассеяния на одном центре имеет не только формальный смысл. Именно, формулы (5.5) и (5.10) получены в первом неисчезающем (борновском) приближении теории возмущений. Это приближение применимо ([М2], § 95), если выполняется хотя бы одно из двух неравенств:

$$|\delta U| \ll \frac{\hbar^2}{md^2}, \quad |\delta U| \ll \frac{\hbar}{d}\sqrt{\frac{E}{m}}. \quad (5.13)$$

Здесь d — радиус действия поля $\delta U(\mathbf{r})$. В задаче о рассеянии электрона атомами заряженной примеси роль d играет радиус экранирования r_0 (см. Приложение XII), а характерное значение $|\delta U| \simeq e^2/\varepsilon r_0$. Таким образом, неравенства (5.13) принимают вид

$$a_B \gg r_0, \quad \sqrt{E_B} \ll \sqrt{E}. \quad (5.14)$$

*) Более точное обоснование возможности пренебречь интерференционным членом (5.8) можно получить, вводя представление об усреднении всех наблюдаемых величин по координатам примесных атомов (см. [1], статьи №№ 5, 6 и Приложение XIII).

Здесь $a_B = \frac{\varepsilon\hbar^2}{me^2}$, $E_B = \frac{me^4}{2\varepsilon^2\hbar^2}$ — соответственно боровский радиус и боровская энергия в кристаллической решетке, статическая диэлектрическая проницаемость которой есть ε. Как мы видели в гл. IV, в Ge $a_B \simeq 40$ Å, а $E_B \simeq 0{,}01$ эВ. Отсюда видно, что первое из неравенств (5.14) может выполняться лишь при очень большой концентрации экранирующих зарядов, достигаемой в предельно сильно легированных полупроводниках (гл. XIX), а второе накладывает ограничение на температуру образца. Действительно, в отсутствие фермиевского вырождения энергия «существенных» электронов или дырок — порядка kT и мы получаем

$$E_B \ll kT. \tag{5.14'}$$

В области низких температур, представляющей, как мы увидим, серьезный интерес именно в случае рассеяния на примеси, это условие может и не выполняться. При этом, благодаря малости концентрации примесных атомов, неравенство (2.22а), может все же оставаться в силе и применение кинетического уравнения, как такового, может быть оправдано. Надо лишь отказаться от борновского приближения в задаче о рассеянии одним отдельным атомом примеси. Соотношениями (5.5) и (5.10) в этом случае пользоваться уже нельзя, однако равенство (5.11), как можно показать ([1], статья № 9), сохраняет силу, если только под $\sigma(\mathbf{p}, \mathbf{p}')$ понимать точное сечение упругого рассеяния. Соответственно остается в силе и формула (5.12).

Все изложенное (кроме явных формул (5.14), (5.14')) относится к рассеянию носителей заряда не только атомами примеси, но и любыми другими «точечными» дефектами решетки — вакансиями, собственными атомами, смещенными в междоузлия, и т. д.

Специализируем теперь общие формулы (5.10) или (5.12) на случай рассеяния электронов и дырок атомами заряженной примеси. В германии и кремнии это могут быть, например, атомы элементов III или V групп (при температуре выше водородной), а также, при определенных условиях, атомы Li, Zn, Mn, Fe, Cu, Ni и т. д. Обозначим заряд примесного атома через Ze (целое число Z может быть как положительным, так и отрицательным). Тогда, казалось бы, потенциальная энергия должна иметь кулоновский вид (IV.7.1). На самом деле, однако, это неверно: формула (IV.7.1) несправедлива как на малых (порядка a), так и на больших расстояниях от примесного центра. Действительно, при $r \sim a$ ионизованный атом примеси уже нельзя рассматривать как точечный заряд, что предполагается в формуле (IV.7.1). Далее, при столь малых расстояниях теряет смысл использование обычной (макроскопической) диэлектрической проницаемости ε. С другой стороны, наличие в кристаллической решетке других свободных зарядов, а также других примесных центров приводит к экранированию всех электрических полей,

в том числе и кулоновского поля данного иона примеси. В результате изменяется сам вид функции δU (r), что существенно на расстояниях, сравнимых с радиусом экранирования.

Первое из указанных только что осложнений не очень существенно, если характерная длина волны носителей заряда велика по сравнению с постоянной решетки. Действительно, при этом взаимодействие на расстояниях порядка a не играет роли: электрон нельзя локализовать в объеме, линейные размеры которого меньше длины волны. В невырожденном газе характерная длина волны есть $h/\sqrt{mkT}$; в газе, полностью вырожденном, она составляет примерно $n^{-1/3}$, где n — концентрация частиц (электронов или дырок). Таким образом, в условиях, когда, в зависимости от степени вырождения,

$$\frac{h}{\sqrt{mkT}} \gg a \quad \text{или} \quad n^{-1/3} \gg a, \tag{5.15}$$

осложнениями, возникающими на расстояниях $r \sim a$, можно пренебречь.

Условия (5.15) обычно хорошо выполняются. Так, например, при $T = 300$ К и $m = m_0$ мы имеем $h/\sqrt{mkT} \simeq 10^{-6}$ см. Поэтому в дальнейшем мы не будем принимать во внимание отклонение δU (r) от вида (IV.7.1) при $r \simeq a$. Соответственно для δU (r) можно воспользоваться выражением, полученным в Приложении XII:

$$\delta U(\mathbf{r}) = \frac{Ze^2}{\varepsilon r} \exp\left(-\frac{r}{r_0}\right). \tag{5.16}$$

Подставляя это в правую часть (5.9), придем к интегралу

$$\mathcal{J} = \int r^{-1} \exp\left[-\frac{r}{r_0} + \frac{i}{\hbar}(\mathbf{p}-\mathbf{p}',\ \mathbf{r})\right] d\mathbf{r}.$$

Главную роль в нем играют значения r, близкие к r_0. Поэтому мы вправе считать здесь объем системы бесконечно большим. Тогда интеграл легко вычисляется (в сферических координатах) и мы получаем

$$S(\mathbf{p},\ \mathbf{p}') = \frac{32\pi^3\hbar^3 N_i Z^2 e^4}{\varepsilon^2 [\hbar^2 r_0^{-2} + (\mathbf{p}-\mathbf{p}')^2]^2}. \tag{5.17}$$

При этом, в силу упругости рассеяния, $E(\mathbf{p}) = E(\mathbf{p}')$. При параболическом законе дисперсии $p^2 = 2mE$ (p) и, следовательно, формулу (5.17) можно переписать в виде

$$S(\mathbf{p},\ \mathbf{p}') = \frac{32\pi^3\hbar^3 N_i Z^2 e^4}{\varepsilon^2[\hbar^2 r_0^{-2} + 4mE(\mathbf{p})\,(1-\cos\theta)]^2}. \tag{5.18}$$

Это есть выражение вида (XIII.6.5).

Комбинируя теперь формулы (5.11), (5.12) и (5.18), мы получаем

$$\frac{1}{\tau(E)} = \frac{\pi N_i Z^2 e^4}{\varepsilon^2 E^{3/2}\sqrt{2m}}\left[\ln(1+x) - \frac{x}{1+x}\right], \tag{5.19}$$

где

$$x = \frac{8mEr_0^2}{\hbar^2} = 16\pi^2 \frac{r_0^2}{\lambda^2}. \tag{5.20}$$

Мы ввели здесь длину волны де Бройля $\lambda = \frac{2\pi\hbar}{\sqrt{2mE}}$. В невырожденном газе при $E \simeq kT$, как и в вырожденном при $E = \zeta$, параметр x обычно оказывается большим ($x \gg 1$). Тогда формула (5.19) принимает вид

$$\frac{1}{\tau(E)} = \frac{\pi N_i Z^2 e^4}{\varepsilon^2 E^{3/2}\sqrt{2m}} \ln\left(\frac{8mEr_0^2}{\hbar^2}\right). \tag{5.19'}$$

Если формально положить здесь $r_0 \to \infty$, то обратное время релаксации τ^{-1} обратится в бесконечность, а время релаксации, подвижность и т. д. — в нуль, что, разумеется, резко противоречит опыту. Случай $r_0 \to \infty$ означает переход от дебаевского закона (5.16) к кулоновскому (IV.7.1), т. е. пренебрежение экранированием. Именно по этой причине учет экранирования в задаче о рассеянии носителей заряда атомами заряженной примеси необходим принципиально. Заметим, однако, что точная форма закона экранирования здесь не очень существенна, ибо радиус экранирования r_0 входит только в аргумент медленно меняющейся функции — логарифма.

Заметим, что, с точностью до медленно меняющегося логарифмического множителя, формулу (5.19') можно переписать в виде (4.19); в рассматриваемом случае

$$r = +3/2. \tag{5.19''}$$

Пределы применимости формулы (5.19') определяются неравенствами (5.14) и (2.22а, б). В случае невырожденного электронного газа условие (2.22а) можно переписать в виде

$$2\pi N_i a_B^3 \ln\left(\frac{8mkT}{\hbar^2} r_0^2\right) \ll \left(\frac{kT}{E_B}\right)^{5/2}. \tag{5.21}$$

При $E_B = 0{,}01$ эВ и $a_B = 4 \cdot 10^{-7}$ см это дает

$$N_i \ll 10^{13} T^{5/2},$$

причем концентрацию N_i следует измерять в см$^{-3}$, а температуру — в градусах Кельвина.

Условие (5.14') может оказаться довольно жестким. Так, при $E_B = 0{,}01$ эВ оно нарушается уже при температуре около 120 К. Соответственно возникает необходимость в расчете, не основывающемся на борновском приближении. Оказывается [4], что соответствующее время релаксации по-прежнему растет с энергией электрона, но несколько слабее, чем это предписывается формулой (5.19').

Обратимся теперь к рассеянию носителей заряда нейтральными атомами примеси. Оно может стать существенным при достаточно низких температурах, когда начинается деионизация примесных

атомов (см. § V.15); при этом в отсутствие компенсации концентрация N_t в формуле (5.19′) стремится к нулю. Общая схема расчета в этом случае остается прежней. Следует лишь принять во внимание то обстоятельство, что силы взаимодействия между электронами и нейтральными атомами примеси сравнительно короткодействующие. По этой причине эффект экранирования здесь не играет заметной роли и энергию взаимодействия можно определить, рассматривая только систему «нейтральный атом примеси + электрон».

Особенно просто обстоит дело в случае примесей, описываемых водородной моделью (§ IV.7). В рамках метода эффективной массы мы получаем здесь, по существу, задачу о рассеянии электронов нейтральными атомами водорода. Последняя хорошо изучена в теории столкновений, и можно прямо воспользоваться известными результатами, заменяя лишь истинную массу электрона на эффективную и деля квадрат заряда e^2 на диэлектрическую проницаемость ε. Точный расчет оказывается возможным провести только численно; однако в интересной области температур для времени релаксации удается получить простую интерполяционную формулу:

$$\frac{1}{\tau} = \frac{20\varepsilon\hbar^3}{m^2e^2} N_0, \tag{5.22}$$

где N_0 — концентрация нейтральной примеси. Это — выражение вида (4.19) при $r = 0$. Его называют формулой Эрджинсоя.

§ 6. Подвижность, холл-фактор и термоэдс при различных механизмах рассеяния

Равенства (4.19), (5.19) и (5.22) в сочетании с таблицей 14.2 позволяют полностью раскрыть содержание формул гл. XIII, описывающих подвижность, термоэдс и холловский фактор газа носителей заряда. В таблице 14.3 приведены основные результаты, относящиеся к невырожденному электронному газу с параболическим законом дисперсии. Поскольку величина C в формуле (XIII.7.20′) может зависеть от температуры, мы представляем дрейфовую подвижность в виде

$$\mu = AT^p, \tag{6.1}$$

где A — множитель, не зависящий или слабо (логарифмически) зависящий от температуры; p — постоянная.

Отметим, в частности, результаты, получающиеся при рассеянии носителей заряда продольными акустическими фононами и атомами заряженной примеси. В первом случае подстановка выражения (4.19) в формулу (XIII.7.20′) дает

$$\mu_{\text{ак. фон}} = \frac{4e\sqrt{\pi}Mc_s^2\hbar^4}{3E_1^2V_0\sqrt{2m^5}(kT)^{3/2}}. \tag{6.2}$$

Таблица 14.3

Значения констант в формуле для подвижности, термоэдс и холловского фактора при различных механизмах рассеяния

Рассеиватель	Коэффициент A в формуле (6.1) для подвижности	Показатель степени p в формуле (6.1) для подвижности	Термоэдс α	Холловский фактор γ
Акустические фононы (потенциал деформации)	$\frac{4e\pi^{1/2}Mc_s^2\hbar^4}{3(2m^5)^{1/2}E_1^2k^{3/2}V_0}$	$-3/2$	$\frac{k}{e}\left(\frac{\zeta}{kT}-2\right)$	$\frac{3\pi}{8}=1{,}18$
Акустические фононы (пьезоэлектрическое рассеяние)	$\frac{5\varepsilon^2Mc_s^2\hbar^2}{2\beta^2eV_0(2k\pi^3m^3)^{1/2}}$	$-1/2$	$\frac{k}{e}\left(\frac{\zeta}{kT}-3\right)$	$\frac{45\pi}{128}=1{,}105$
Неполярные оптические фононы, $kT\gg\hbar\omega_0$	$\frac{4e\pi^{1/2}M\hbar^4\omega_0^2}{3V_0E_0^2(2m^5k^3)^{1/2}}$	$-3/2$	$\frac{k}{e}\left(\frac{\zeta}{kT}-2\right)$	$\frac{3\pi}{8}=1{,}18$
Поляризационные фононы, $kT\gg\hbar\omega_0$	$\frac{eV_0M_1M_2\hbar^2\omega_0^2\sqrt{2}}{3(Ze^2)^2(M_1+M_2)}\times$ $\times(\pi^3km^3)^{-1/2}$	$-1/2$	$\frac{k}{e}\left(\frac{\zeta}{kT}-3\right)$	$\frac{45\pi}{128}=1{,}105$
Заряженная примесь	$\frac{8\varepsilon^2(2k^3)^{1/2}(m\pi^3)^{-1/2}}{e^3N_t\ln\left(\frac{24mkT}{\hbar^2}r_0^2\right)}$	$3/2$	$\frac{k}{e}\left(\frac{\zeta}{kT}-4\right)$	$\frac{315\pi}{512}=1{,}93$
Нейтральная примесь	$\frac{me^3}{20\varepsilon\hbar^3N_0}$	0	$\frac{k}{e}\left(\frac{\zeta}{kT}-\frac{5}{2}\right)$	1

Во втором случае мы получаем, подставляя выражение (5.19′) в (XIII.7.20),

$$\mu_{\text{зар. прим}}=\frac{4\varepsilon^2\sqrt{2}}{3N_te^3\sqrt{m\pi^3}(kT)^{3/2}}\int_0^\infty\frac{E^3\exp\left(-\frac{E}{kT}\right)dE}{\ln\left(\frac{8mE}{\hbar^2}r_0^2\right)}. \qquad (6.3)$$

Функция $E^3\exp\left(-\frac{E}{kT}\right)$ имеет максимум в точке $E=3kT$. Вблизи этой энергии логарифм, стоящий в знаменателе подынтегрального выражения в (6.3), изменяется довольно медленно. По этой причине его можно вынести из-под знака интеграла при $E=3kT$.

Таким путем находим

$$\mu_{\text{зар. прим}} = \frac{8\varepsilon^2\sqrt{2}\,(kT)^{3/2}}{N_i e^3\sqrt{m}\pi^3 \ln\left(\frac{24m\,kT}{\hbar^2}r_0^2\right)}. \qquad (6.3')$$

Это соотношение называют формулой Брукса — Херринга *).

Температурная зависимость подвижности, характеризуемая числом p, имеет ясное происхождение. Действительно, в рассматриваемых условиях среднее число фононов, имеющихся в решетке, пропорционально T; при рассеянии на фононах подвижность, очевидно, обратно пропорциональна этому числу. Еще один множитель, T^r, связан с энергетической зависимостью времени свободного пробега (4.19): в отсутствие фермиевского вырождения каждый множитель E под знаком интеграла в (XIII.7.20) приводит к появлению множителя kT в окончательном результате. Таким образом, при рассеянии на фононах

$$\mu \sim T^{-1+r}, \quad p = r - 1. \qquad (6.4)$$

При рассеянии на примеси первый из указанных выше факторов отпадает (вместо него появляется N_i^{-1}); в результате

$$\mu \sim N_i^{-1} T^r, \quad p = r. \qquad (6.5)$$

Таким образом, температурная зависимость подвижности, выражаемая формулой (6.1) и данными таблицы 14.3, обусловлена двумя факторами: условием равнораспределения энергии по степеням свободы (4.14) (при рассеянии на фононах) и предположением о квадратичном законе дисперсии.

Учет анизотропии изоэнергетических поверхностей (в рамках параболического закона дисперсии) приводит лишь к изменению коэффициента A в формуле (6.1).

Ситуация становится более сложной при наличии магнитного поля: в этом случае анизотропия изоэнергетических поверхностей играет более существенную роль. Действительно, в одном лишь электрическом поле дрейфовая скорость носителей заряда v_d постоянна по направлению, а сила, действующая на электрон, не зависит от v_d, т. е. задана. Поэтому движение электрона в электрическом или температурном поле в любой момент времени определяется одной и той же (хотя, вообще говоря, и зависящей от направления поля) эффективной массой. Результаты для анизотропной и изотропной систем различаются здесь, в конце концов, только численными коэффициентами. С другой стороны, в магнитном поле траектория электрона не прямолинейна, а сила, на него действующая, зависит

*) Аналогичное выражение, отличающееся от (6.3') лишь аргументом логарифма (благодаря менее последовательному учету экранирования), было ранее получено Конвелл и Вайскопфом. По этой причине часто употребляется и название «формула Конвелл—Вайскопфа».

от дрейфовой скорости. Поэтому в анизотропной системе ускорение электрона в различные моменты времени определяется различными эффективными массами. Отличие от изотропной системы здесь оказывается принципиальным.

§ 7. Одновременное действие нескольких механизмов рассеяния

В реальных условиях опыта могут одновременно играть роль несколько механизмов рассеяния. Поэтому возникает вопрос о вычислении подвижности и других кинетических коэффициентов при одновременном действии нескольких таких механизмов.

Будем считать известными вероятности перехода для каждого механизма рассеяния в отдельности. Обозначим их через $\mathscr{P}_i(\mathbf{p}, \mathbf{p}')$, где индекс i нумерует различные механизмы. Допустим, далее, что процессы рассеяния, связанные с различными механизмами, представляют собой независимые события. Тогда интересующая нас полная вероятность рассеяния дается суммой

$$\mathscr{P}(\mathbf{p}, \mathbf{p}') = \sum_i \mathscr{P}_i(\mathbf{p}, \mathbf{p}'). \tag{7.1}$$

Это соотношение справедливо независимо от того, можно ли ввести соответствующее время релаксации для того или иного механизма рассеяния; оно не зависит также от каких-либо предположений относительно изотропии системы. Если можно ввести времена релаксации $\tau_i(E)$, то из равенства (7.1) вытекает правило сложения обратных времен релаксации:

$$\frac{1}{\tau(E)} = \sum_i \frac{1}{\tau_i(E)}. \tag{7.2}$$

Действительно, согласно формуле (XIII.6.12) обратное время релаксации τ_i^{-1} линейно выражается через вероятность соответствующего перехода.

Иногда вместо формулы (7.2) пользуются соответствующим соотношением для обратных подвижностей:

$$\frac{1}{\mu} = \sum_i \frac{1}{\mu_i}. \tag{7.2'}$$

Здесь μ — наблюдаемая на опыте подвижность дрейфа, а μ_i — подвижность, которая наблюдалась бы, если бы имел место только i-й механизм рассеяния. Следует, однако, иметь в виду, что равенство (7.2') эквивалентно (7.2) только в очень специальном случае, когда все времена τ_i не зависят от энергии. Как правило, времена релаксации от энергии зависят и соотношение (7.2'), строго говоря, не выполняется. В зависимости от конкретных условий опыта ошибка при его использовании может достигать 40—50%.

Как видно из формулы (7.2) (и очевидно заранее), главную роль играет тот механизм рассеяния, которому соответствует наименьшее время релаксации (и, следовательно, наименьшее значение подвижности). При изменении температуры относительная роль различных механизмов может меняться. Так, в германии при комнатной температуре и концентрации примеси, не большей примерно 10^{16} см$^{-3}$, главную роль играет рассеяние носителей заряда акустическими фононами. К нему добавляется рассеяние неполярными оптическими фононами. Легко видеть, что последнее приводит к усилению температурной зависимости подвижности. В самом деле, соответствующая вероятность рассеяния обратно пропорциональна числу фононов

$$n = \left[\exp\left(\frac{\hbar\omega_0}{kT}\right) - 1\right]^{-1}. \quad (7.3)$$

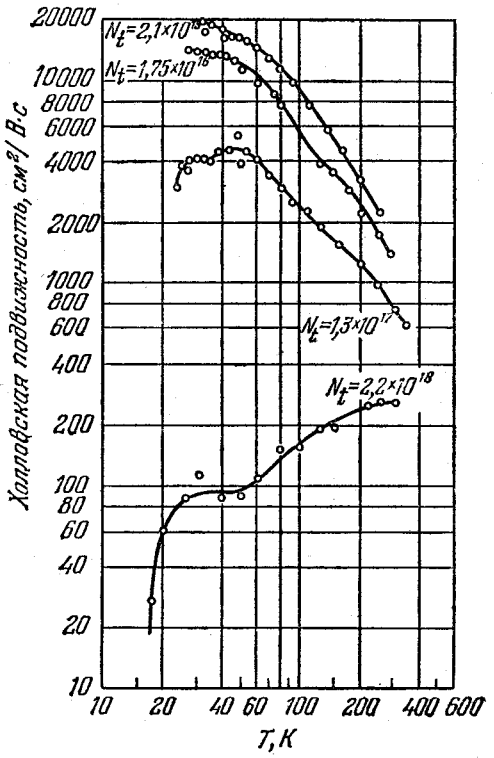


Рис. 14.1. Температурная зависимость холловской подвижности электронов в кремнии, легированном мышьяком. Справа и слева от максимума доминирует, соответственно, рассеяние на акустических фононах и на заряженной примеси.

Как уже отмечалось, для оптических фононов в германии при комнатной температуре аппроксимация (4.14) еще неприменима. По этой причине температурная зависимость подвижности, связанной с рассеянием только на оптических фононах, оказывается более резкой, нежели выражаемая формулой (6.2). Видимо, по этой причине наблюдаемые на опыте подвижности электронов и дырок в этой области температур выражаются эмпирическими формулами

$$\mu_n \sim T^{-1,65}, \quad \mu_p \sim T^{-2,5}.$$

Различие показателей степени для электронов и дырок может быть обусловлено разными значениями соответствующих акустических и оптических потенциалов деформации. Аналогично обстоит дело и в кремнии.

При понижении температуры подвижность носителей заряда увеличивается, пока не начнет превалировать рассеяние на заряженной примеси (в германии при $N_t \simeq 10^{15}$ см$^{-3}$ это происходит в области азотных температур). Далее подвижность начинает уже уменьшаться вместе с температурой, подчиняясь закону (6.3′), пока не начнет заметно изменяться концентрация заряженной примеси. Коль скоро роль последней играют мелкие доноры или акцепторы, зависимость $N_t(T)$ в германии и кремнии становится заметной при температурах, близких к водородной (гл. V). При дальнейшем понижении температуры в отсутствие компенсации концентрация

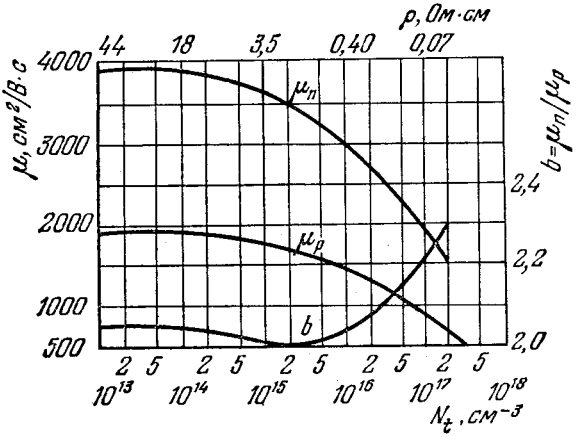


Рис. 14.2. Зависимость подвижностей электронов и дырок в p-германии от концентрации примеси ($T = 300$ °K).

заряженной примеси начинает уменьшаться, а нейтральной — повышаться, в результате чего наиболее существенным может стать рассеяние на нейтральных атомах примеси.

Таким образом, зависимость подвижности от температуры должна изображаться кривой с максимумом, положение которого само зависит от концентрации примеси. На рис. 14.1 представлена температурная зависимость холловской подвижности μ_H электронов в кремнии, легированном мышьяком *) (отличие холловской подвижности от дрейфовой, определяемой формулами (6.1) — (6.3′), здесь несущественно: в условиях, когда доминирует лишь один механизм рассеяния, μ_H отличается от μ только постоянным множителем). При очень большой концентрации заряженной примеси (нижняя кривая) рассеяние на ней доминирует даже при комнатной температуре и максимум может и не наблюдаться.

*) По данным работы *F. T. Morin, I. P. Maita,* Phys. Rev. **96**, 28 (1954).

На рис. 14.2 изображена зависимость подвижности электронов в германии от концентрации заряженной примеси *). Как и следовало ожидать на основании формул (6.2), (6.3′) и (7.2), при достаточно малых значениях N_t подвижность практически не зависит от концентрации примеси, а при повышении N_t начинает уменьшаться.

В соединениях типа $A^{III}B^{V}$ (в достаточно чистых образцах) при комнатных температурах превалирует, по-видимому, рассеяние носителей заряда поляризационными колебаниями решетки.

Помимо рассмотренных выше механизмов рассеяния при определенных условиях может проявиться также рассеяние носителей заряда на других дефектах решетки: дислокациях, границах зерен (в поликристаллических образцах), а также просто на поверхности образца. Последний эффект бывает заметен в образцах, размеры которых (хотя бы в одном направлении) достаточно малы (пленки, проволоки).

Особую роль играет рассеяние носителей заряда друг на друге. Очевидно, в отсутствие перебросов в процессах такого типа полный импульс системы электронов не изменяется; поэтому, в отсутствие каких-либо иных механизмов рассеяния, сопротивление, связанное с одними лишь электрон-электронными столкновениями, без процессов переброса, было бы равно нулю (а подвижность — бесконечна). Тем не менее эти столкновения влияют на подвижность, ибо приводят к перераспределению энергии и квазиимпульса между электронами. Последнее обстоятельство существенно потому, что времена релаксации, связанные, например, с рассеянием на фононах или на ионизованной примеси, зависят от энергии носителей заряда. Таким образом, взаимодействие последних друг с другом, меняя функцию распределения по энергиям, может увеличить или уменьшить интенсивность рассеяния. Так, согласно формуле (5.19′) на заряженной примеси сильнее всего рассеиваются сравнительно медленные электроны. При этом взаимодействие электронов друг с другом, увеличивая процент медленных электронов, приводит к уменьшению подвижности [4].

*) По данным работы *M. Prince*, Phys. Rev. **92**, 681 (1953) (русск. пер.: ПСФ, вып. 2, ИЛ, 1955, стр. 40).

Глава XV

АКУСТО-ЭЛЕКТРОННЫЕ ЯВЛЕНИЯ

§ 1. Предварительные замечания

Взаимодействие колебаний решетки с электронами проводимости проявляется не только в процессах рассеяния квазиимпульса электронов на тепловых колебаниях решетки (гл. XIV), но и в том случае, когда в кристалле распространяются введенные извне упругие волны. В результате этого взаимодействия, с одной стороны, изменяется поведение упругих волн в полупроводниках и, с другой стороны, под действием волн возникают новые электронные процессы.

Ниже мы рассмотрим три наиболее важных акусто-электронных явления: 1) изменение скорости звуковых волн, 2) поглощение звуковых волн и 3) акусто-электрический эффект. Последний заключается в том, что, вследствие увлечения электронов проводимости упругой волной, в полупроводнике возникает электродвижущая сила.

Однако сначала мы должны остановиться на важном общем вопросе о способе описания акусто-электронных явлений. Положим, что звуковая волна, которую мы сейчас будем рассматривать как поток фононов с энергией $\hbar\omega$ и импульсом $\hbar\mathbf{q}$, распространяется вдоль оси X. Рассмотрим, далее, взаимодействие одного электрона с фононом, сопровождающееся поглощением фонона. Как было показано в § XIV.4, при этом должны выполняться законы сохранения квазиимпульса и энергии. При изотропном параболическом законе дисперсии это дает

$$p_{x2} = p_{x1} + \hbar q, \tag{1.1}$$

$$\frac{p_{x2}^2}{2m} = \frac{p_{x1}^2}{2m} + \hbar\omega. \tag{1.2}$$

Здесь p_{x1} и p_{x2} — проекции импульса электрона на направление волны до и после взаимодействия. Подставляя в формулу (1.2) p_{x2} из формулы (1.1) и учитывая, что ω/q есть фазовая скорость волны, обозначаемая в этой главе через v_s, получаем

$$\frac{\hbar q}{m}\left[p_{x1} - \left(mv_s - \frac{1}{2}\hbar q\right)\right] = 0. \tag{1.3}$$

Отсюда следует, что взаимодействовать с решеткой могут только те электроны, импульс которых удовлетворяет условию

$$p_{x1} = mv_s - \frac{1}{2}\hbar q. \tag{1.4}$$

Его поясняет рис. 15.1, где изображено сечение «плоскости взаимодействия» $p_{x1} = \text{const}$ плоскостью рисунка и показано расположение векторов $\mathbf{p}$ и $\mathbf{q}$. Аналогично, при испускании одного фонона концы векторов $\mathbf{p}$ должны лежать на плоскости взаимодействия $p_{x2} = \text{const}$, где

$$p_{x2} = mv_s + \frac{1}{2}\hbar q. \tag{1.4a}$$

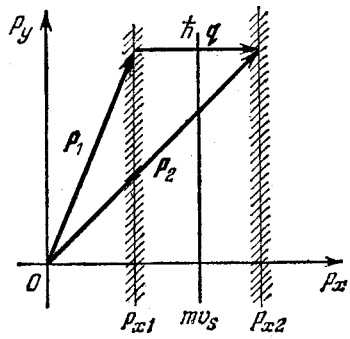


Рис. 15.1. Элементарный акт взаимодействия электрона и фонона.

В этом описании мы не учитывали того, что, помимо взаимодействия со звуковой волной, электроны могут еще рассеиваться на тепловых колебаниях решетки, на ионизованных примесях и вследствие других процессов (гл. XIV). Иными словами, мы рассматривали электронный газ в кристалле как бесстолкновительную плазму. Однако, как мы сейчас увидим, это можно сделать лишь в том случае, когда частота звуковой волны достаточно велика. Для низких же частот звука такое описание невозможно вследствие квантовомеханического соотношения неопределенности. Действительно, если τ есть время свободного пробега электрона, то его энергия определена лишь с точностью

$$\Delta E \sim \hbar/\tau. \tag{1.5}$$

Поэтому вместо формулы (1.3) мы должны писать

$$\frac{\hbar q}{m}\left\{p_x - \left(mv_s - \frac{1}{2}\hbar q\right)\right\} \sim \frac{\hbar}{\tau}. \tag{1.6}$$

Соответственно этому возникает и неопределенность импульса (1.4) взаимодействующих электронов

$$\Delta p_x \sim \frac{\hbar}{\tau}\frac{m}{\hbar q} = \frac{m}{q\tau}, \tag{1.7}$$

а плоскости взаимодействия на рис. 15.1 размываются в полосы шириной $\sim \Delta p_x$. Очевидно, что если

$$\Delta p_x \gtrsim p_x, \tag{1.8}$$

то о взаимодействии индивидуальных электронов говорить уже нельзя и приближение бесстолкновительной плазмы становится

неприменимым. В этом случае процессы имеют коллективный характер и со звуковой волной взаимодействуют сгустки объемного заряда. При этом звуковую волну можно рассматривать как упругую волну в сплошной среде и для описания акусто-электронных явлений пользоваться уравнениями электродинамики и механики сплошных сред (гидродинамическое описание).

Так как в наших рассуждениях составляющие импульса p_y и p_z произвольны, то максимальное значение p_x имеет порядок p. Учитывая, далее, что $p\tau/m = v_T\tau = l$, где l — длина свободного пробега электронов, находим из соотношений (1.7) и (1.8) условие возможности гидродинамического описания:

$$ql \ll 1. \tag{1.9}$$

В случае обратного неравенства, напротив, гидродинамическое описание становится непригодным и мы должны рассматривать взаимодействие отдельных электронов и фононов.

При комнатных температурах обычно $l \sim 10^{-5} \div 10^{-6}$ см и поэтому $ql = 1$ при $q \sim 10^5 \div 10^6$ см$^{-1}$. Так как скорость звука $v_s \sim \sim 10^5$ см/с, то это соответствует частоте $\omega = qv_s \sim 10^{10} \div 10^{11}$ с$^{-1}$, или числу колебаний в секунду $\omega/2\pi \sim 10^9 \div 10^{10}$ Гц. Таким образом, гидродинамическое описание оказывается справедливым вплоть до очень высоких частот, порядка гигагерц.

§ 2. Взаимодействие упругих волн с электронами проводимости

1. В акусто-электронных явлениях обычно приходится учитывать два основных типа взаимодействия электронов со звуковой волной: обусловленное потенциалом деформации (§ XIV, 3) и вызываемое электрическим полем пьезоэлектрического эффекта. Оба эти типа взаимодействия были уже рассмотрены в гл. XIV. Однако сейчас нас будет интересовать гораздо более простой случай, когда в кристалле имеется лишь одна волна, которую для определенности мы будем считать продольной. Далее, мы положим, что волна распространяется вдоль какой-либо оси симметрии кристалла, и, соответственно, будем считать, что все векторные величины направлены вдоль этой оси и зависят только от одной координаты x (и от времени t). Тогда для изменения энергии электрона ΔE при деформации u вместо формулы (XIV.3.5′) мы будем иметь простое соотношение

$$\Delta E = E_1 u, \tag{2.1}$$

в которое входит лишь одна компонента потенциала деформации E_1. Соответственно сила, действующая на электрон, равна

$$F = -\frac{\partial \Delta E}{\partial x} = -E_1 \frac{\partial u}{\partial x} = -E_1 \frac{\partial^2 Q}{\partial x^2}, \tag{2.2}$$

где Q — смещение рассматриваемой точки. Эта сила имеет неэлектростатическое происхождение и должна учитываться как сторонняя (или электродвижущая) сила. Напряженность поля этих сторонних сил $\mathcal{E}^*$ (сила на единицу положительного заряда) есть

$$\mathcal{E}^* = \frac{E_1}{e}\frac{\partial u}{\partial x}. \tag{2.3}$$

Формула для электростатической индукции при учете пьезоэлектрического эффекта имеет вид (XIV.3.21)

$$\mathscr{D} = \varepsilon\mathcal{E} - 4\pi\beta u. \tag{2.4}$$

Здесь β — соответствующая компонента пьезоэлектрического модуля, а $-\beta\mu = \mathscr{P}$ есть электрический момент единицы объема, возникающий при деформации.

Наряду с прямым пьезоэлектрическим эффектом, выражаемым формулой (2.4) (возникновение поляризации при деформации), необходимо также учитывать и обратный эффект (появление деформации или, соответственно, механических напряжений при поляризации). Поэтому, если s — механическое напряжение (в нашем случае сила в направлении X, рассчитанная на единичную площадку с нормалью по оси X), а Λ — модуль упругости, то, согласно (XIV. 3.22),

$$s = \Lambda u + \beta\mathcal{E}, \tag{2.5}$$

где второе слагаемое справа описывает обратный пьезоэлектрический эффект.

Оценим теперь порядок величины обоих взаимодействий. Для этого положим, что в диэлектрическом кристалле распространяется гармоническая волна деформации

$$u = u_1 \exp\left[i\left(qx - \omega t\right)\right]. \tag{2.6}$$

Так как в диэлектрике, согласно уравнению Пуассона,

$$\operatorname{div}\mathfrak{D} = 0,$$

то $\mathfrak{D} = \text{const}$. Тогда из (2.4) следует, что переменное электрическое поле $\tilde{\mathcal{E}}$, вызываемое пьезоэлектрическим эффектом, равно

$$\tilde{\mathcal{E}} = \frac{4\pi\beta}{\varepsilon}u. \tag{2.7}$$

С другой стороны, плотность потока энергии в волне (интенсивность звука) есть

$$I = \frac{1}{2}\Lambda u_1^2 v_s. \tag{2.8}$$

Отсюда можно найти амплитуду поля $\mathcal{E}_1$ при заданной интенсивности звука. Это поле в сильных пьезоэлектриках, каковыми являются, например, кристаллы соединений $A^{II}B^{VI}$ (CdS, ZnS и др.),

может быть весьма значительным. Принимая для оценок типичные значения: $\Lambda \sim 10^{12}$ дин/см^2, $\varepsilon \sim 10$, $v_s \sim 10^5$ см/с и $|\beta| \sim 10^5$ ед. СГСЭ, по формулам (2.7) и (2.8) мы находим, что при $I \sim 1$ Вт/см$^2 \doteq$ $= 10^7$ эрг/см$^2 \cdot$с амплитуда деформации $u_1 \sim 10^{-5} \div 10^{-4}$, а $\tilde{\mathcal{E}}$ может достигать нескольких ед. СГСЭ, или $\sim 10^3$ В/см.

Для взаимодействия через потенциал деформации из (2.3) и (2.6) получается

$$\mathcal{E}^* = i\frac{E_1}{e} qu = i\frac{E_1}{ev_s}\omega u. \tag{2.9}$$

Отношение величин обоих полей равно

$$\frac{|\tilde{\mathcal{E}}|}{|\tilde{\mathcal{E}}^*|} = \frac{4\pi e|\beta| v_s}{\varepsilon |E_1| \omega}. \tag{2.10}$$

Отсюда видно, что относительная роль деформационного взаимодействия тем больше, чем больше частота ω. Так как E_1 обычно имеет порядок $1 \div 10$ эВ ($\sim 10^{-12}$ эрг), то, используя прежние оценочные значения β, ε и v_s, мы находим по формуле (2.10), что $|\tilde{\mathcal{E}}| = |\tilde{\mathcal{E}}^*|$ при $\omega \sim 10^{12}$ с$^{-1}$. Следовательно, при частотах колебаний $\omega/2\pi \ll$ $\ll 10^{11}$ Гц в пьезоэлектрических кристаллах деформационным взаимодействием можно пренебречь по сравнению с пьезоэлектрическим.

2. В дальнейшем мы сначала рассмотрим случай этих «низких» частот звука и пьезоэлектрических кристаллов. В соответствии со сказанным выше мы будем учитывать только пьезоэлектрическое взаимодействие и пользоваться гидродинамическим описанием ($ql \ll 1$). Кристаллы будем считать имеющими электронную проводимость и влиянием дырок будем пренебрегать. Постоянные значения величин в отсутствие волны будем отмечать индексом «0». Тогда для определения электрических и механических величин мы имеем следующую систему уравнений. Выражение для плотности тока

$$j = e\mu(n_0 + \tilde{n})(\mathcal{E}_0 + \tilde{\mathcal{E}}) + eD\frac{\partial \tilde{n}}{\partial x} \tag{2.11}$$

и уравнение непрерывности

$$-e\frac{\partial \tilde{n}_s}{\partial t} + \frac{\partial j}{\partial x} = 0. \tag{2.12}$$

Здесь

$$\tilde{n}_s = \tilde{n} + \tilde{n}_t \tag{2.13}$$

есть полная концентрация избыточных электронов, которая складывается из концентрации в зоне $\tilde{n}$ и концентрации электронов, связанных на ловушках, $\tilde{n}_t$. Отметим, что отношение концентраций

$$\frac{\tilde{n}}{\tilde{n}_t} \equiv f \tag{2.14}$$

в литературе часто называют «фактором прилипания электронов» на ловушки. Далее, мы имеем уравнение Пуассона

$$\frac{\partial \tilde{\mathscr{D}}}{\partial x} = -4\pi e \tilde{n}_s, \tag{2.15}$$

два пьезоэлектрических соотношения

$$\tilde{\mathscr{D}} = e\tilde{\mathscr{E}} - 4\pi\beta u, \tag{2.16}$$

$$\tilde{s} = \Lambda u + \beta\tilde{\mathscr{E}} \tag{2.17}$$

и, наконец, механическое уравнение движения среды

$$\rho \frac{\partial^2 Q}{\partial t^2} = \frac{\partial \tilde{s}}{\partial x} = \Lambda \frac{\partial^2 Q}{\partial x^2} + \beta \frac{\partial \tilde{\mathscr{E}}}{\partial x}, \tag{2.18}$$

где ρ — плотность среды.

Для определения связи между концентрациями $\tilde{n}$ и $\tilde{n}_t$ (или, что то же, фактора прилипания f) необходимы еще уравнения кинетики захвата электронов на примесные центры, рассмотренные в § IX.4. Для простейшего случая центров прилипания одного типа соответствующее уравнение имеет вид

$$\frac{\partial \tilde{n}_t}{\partial t} = \alpha_n (n_0 + \tilde{n})(N_t - n_{t0} - \tilde{n}_t) - \alpha_n n_1 (n_{t0} + \tilde{n}_t), \tag{2.19}$$

где N_t, α_n и n_1 имеют тот же смысл, что и в § IX.4. При этом фактор прилипания в общем случае оказывается комплексным (что указывает на сдвиг фаз колебаний $\tilde{n}$ и $\tilde{n}_s$) и зависящим от частоты волны ω. Однако положение существенно упрощается, если $\omega\tau \ll 1$, где τ — время жизни электронов в зоне. В этом случае ловушки успевают «следить» за изменениями $\tilde{n}$ и f становится постоянным числом, зависящим от состава кристалла и температуры.

Отметим в заключение, что, записывая уравнение движения в форме (2.18), мы не включили в него силы «трения», всегда существующие в реальных кристаллах. Этим мы исключили поглощение волны самой решеткой кристалла, которое, в случае необходимости, должно быть учтено дополнительно.

§ 3. Упругие волны в пьезодиэлектриках

Рассмотрим теперь особенно простой случай, когда упругая волна распространяется в непроводящей пьезоэлектрической среде ($\tilde{n} = \tilde{n}_t = \tilde{n}_s = 0$). Тогда уравнение (2.15) дает $\tilde{\mathscr{D}} = \text{const} = 0$ (так как $\tilde{\mathscr{D}}$ может быть постоянным, только если оно равно нулю), а из (2.16) получается

$$\frac{\partial \tilde{\mathscr{E}}}{\partial x} = \frac{4\pi\beta}{\varepsilon} \frac{\partial u}{\partial x} = \frac{4\pi\beta}{\varepsilon} \frac{\partial^2 Q}{\partial x^2}.$$

Подставляя это в уравнение движения (2.18), мы находим

$$\frac{\partial^2 Q}{\partial t^2} = \frac{\Lambda}{\rho}(1+K^2)\frac{\partial^2 Q}{\partial x^2}, \tag{3.1}$$

где

$$K^2 = \frac{4\pi\beta^2}{\varepsilon\Lambda}. \tag{3.2}$$

Соответственно, дифференцируя обе части уравнения (3.1) по x и заменяя в нем $\partial Q/\partial x$ на u, получаем уравнение, описывающее распространение деформации:

$$\frac{\partial^2 u}{\partial t^2} = \frac{\Lambda}{\rho}(1+K^2)\frac{\partial^2 u}{\partial x^2}. \tag{3.3}$$

Мы получили обычное волновое уравнение. Одно из его решений есть плоская волна, описываемая формулой (2.6). Фазовая скорость волны v_s дается выражением

$$v_s^2 = \frac{\Lambda}{\rho}(1+K^2) = v_{s0}^2(1+K^2), \tag{3.4}$$

где $v_{s0}^2 = \Lambda/\rho$ есть квадрат скорости волны в отсутствие пьезоэлектрического взаимодействия. Мы видим, что в пьезоэлектрической среде скорость волны увеличивается в $\sqrt{1+K^2}$ раз. Можно также сказать, что в этом случае упругие свойства среды описываются измененным, или эффективным, модулем упругости, равным

$$\Lambda' = \Lambda(1+K^2). \tag{3.5}$$

Безразмерная постоянная K носит название константы электромеханической связи. Она определяет влияние пьезоэлектрического эффекта на распространение упругих волн. Величина K у сильных пьезоэлектриков (например, соединений типа $A^{II}-B^{VI}$) имеет порядок 0,1.

§ 4. Упругие волны в пьезоэлектрических полупроводниках

Если электропроводность среды не равна нулю, то необходима вся система уравнений (2.11) — (2.19). Она содержит нелинейные уравнения (2.11) и (2.19), и ее решение в общем случае сложно. Поэтому мы ограничимся случаем малых колебаний и будем пренебрегать теми членами, которые содержат произведения малых переменных величин (при этом интенсивность волны может быть совсем не малой с точки зрения практических применений). Далее, чтобы упростить расчеты, мы сначала 1) пренебрежем диффузией и 2) положим, что прилипания электронов нет ($f=1$, $\tilde{n}_s=\tilde{n}$).

Тогда мы получим систему линеаризованных уравнений

$$j = e\mu(\mathscr{E}_0\tilde{n} + n_0\tilde{\mathscr{E}}), \tag{4.1}$$

$$-e\frac{\partial\tilde{n}}{\partial t} + \frac{\partial\tilde{j}}{\partial x} = 0, \tag{4.2}$$

$$\frac{\partial\tilde{\mathscr{D}}}{\partial x} = -4\pi e\tilde{n}, \tag{4.3}$$

$$\tilde{\mathscr{D}} = \varepsilon\tilde{\mathscr{E}} - 4\pi\beta u, \tag{4.4}$$

$$\tilde{s} = \Lambda u + \beta\tilde{\mathscr{E}}, \tag{4.5}$$

$$\rho\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2\tilde{s}}{\partial x^2}. \tag{4.6}$$

Положим, далее, что в кристалле распространяется плоская волна деформации (2.6). При этом мы будем считать ω вещественным, но q — комплексным, так как вследствие взаимодействия волны с электронами ее амплитуда изменяется. Так как теперь уравнения линейны, то и все другие величины будут изменяться по такому же закону:

$$\tilde{\mathscr{E}} \sim \exp[i(qx - \omega t)] \quad \text{и т. д.} \tag{4.7}$$

Поэтому, подставляя (4.7) в уравнения (4.1) — (4.6), мы получим систему алгебраических уравнений

$$\begin{aligned}
-e\mu\mathscr{E}_0\tilde{n} \quad & -e\mu n_0\tilde{\mathscr{E}} \quad && +\tilde{j} && = 0, && (4.8)\\
\omega e\tilde{n} \quad & && +q\tilde{j} && = 0, && (4.9)\\
4\pi e\tilde{n} \quad & && +iq\tilde{\mathscr{D}} && = 0, && (4.10)\\
& -\varepsilon\tilde{\mathscr{E}} && +\tilde{\mathscr{D}} && +4\pi\beta u = 0, && (4.11)\\
& -\beta\tilde{\mathscr{E}} && && +\tilde{s} \quad -\Lambda u = 0, && (4.12)\\
& && && q^2\tilde{s} \quad -\rho\omega^2 u = 0. && (4.13)
\end{aligned}$$

Дальше задача решается стандартным методом теории колебаний. А именно, так как уравнения (4.8) — (4.13) линейные однородные, то условие существования ненулевых решений, как известно, есть равенство нулю детерминанта коэффициентов этой системы. Это дает соотношение между ω и q:

$$F(\omega, q) = 0, \tag{4.14}$$

или закон дисперсии. Выражая отсюда q, мы получим его комплексным:

$$q = \varkappa(\omega) + i\frac{\gamma(\omega)}{2}. \tag{4.15}$$

Подставляя это в формулу (2.6), мы будем иметь

$$u = u_1 \exp\left\{-\frac{1}{2}\gamma x + i(\varkappa x - \omega t)\right\}, \tag{4.16}$$

что выражает затухающую плоскую волну. Отсюда видно, что коэффициент затухания амплитуды волны $\gamma/2$ и скорость ее распространения v_s получаются непосредственно из дисперсионного уравнения по формулам

$$\frac{1}{2}\gamma = \operatorname{Im} q, \qquad v_s = \frac{\omega}{\varkappa} = \frac{\omega}{\operatorname{Re} q}. \tag{4.17}$$

Однако мы не будем вычислять детерминант системы, а для получения дисперсионного уравнения будем последовательно исключать различные переменные. Это позволит нам получить промежуточные соотношения, описывающие колебания различных электрических величин, полезные для физической интерпретации результатов.

Введем для краткости обозначение

$$\eta = 1 + \frac{q\mu\mathscr{E}_0}{\omega}. \tag{4.18}$$

Здесь $-\mu\mathscr{E}_0$ есть дрейфовая скорость электронов v_d в постоянном внешнем поле. Далее, учтем, что даже в сильных пьезоэлектриках мы имеем $\gamma \ll \varkappa$ (это обозначает, что относительное изменение амплитуды на пути в одну длину волны много меньше единицы). Поэтому в выражении (4.18) можно заменить q на $\varkappa$ и считать $\omega/q = v_s$. Следовательно,

$$\eta = 1 - \frac{v_d}{v_s}, \tag{4.19}$$

где v_d считается положительным, когда электроны движутся в ту же сторону, что и волна. Тогда, исключая $\tilde{j}$ из уравнений (4.8) и (4.9), мы получим

$$\tilde{n} = -\frac{q\sigma_0}{e\omega\eta}\tilde{\mathscr{E}}, \tag{4.20}$$

где $\sigma_0 = e\mu n_0$, а исключая из тех же уравнений $\tilde{n}$, найдем

$$\tilde{f} = \frac{\sigma_0}{\eta}\tilde{\mathscr{E}}. \tag{4.21}$$

Далее, исключим из уравнений (4.10) и (4.11) $\tilde{\mathscr{D}}$. Это дает

$$\tilde{\mathscr{E}} = i\frac{4\pi e}{q\varepsilon}\tilde{n} + \frac{4\pi\beta}{\varepsilon}u. \tag{4.22}$$

По формулам (4.20) и (4.22) можно выразить $\tilde{n}$ и $\tilde{\mathscr{E}}$ как функции деформации. Тогда для $\tilde{\mathscr{E}}$ получается

$$\tilde{\mathscr{E}} = \frac{4\pi\beta}{\varepsilon}\frac{\omega\tau_M\eta}{\omega\tau_M\eta + i}u. \tag{4.23}$$

Здесь $\tau_M = \varepsilon/4\pi\sigma_0$ есть максвелловское время релаксации. Исключим, наконец, s из последней пары уравнений (4.12) и (4.13). Это приводит к соотношению

$$(\rho\omega^2 - q^2\Lambda)u - q^2\beta\tilde{\mathcal{E}} = 0. \tag{4.24}$$

Выражая здесь $\tilde{\mathcal{E}}$ через u по формуле (4.23), мы получаем окончательно

$$u\left\{\rho\omega^2 - q^2\Lambda\left[1 + K^2\frac{\omega\tau_M\eta}{\omega\tau_M\eta + i}\right]\right\} = 0,$$

где K^2 — квадрат константы электромеханической связи (3.2). Отсюда следует, что выражение в фигурных скобках равно нулю, что и дает дисперсионное соотношение

$$q^2 = \frac{\rho\omega^2}{\Lambda}\frac{1}{1 + K^2\dfrac{\omega\tau_M\eta}{\omega\tau_M\eta + i}}. \tag{4.25}$$

Это соотношение можно упростить. Так как $K^2 \ll 1$, то с хорошим приближением можно положить

$$q^2 \simeq \frac{\rho\omega^2}{\Lambda}\left\{1 - K^2\frac{\omega\tau_M\eta}{\omega\tau_M\eta + i}\right\} =$$
$$= \frac{\rho\omega^2}{\Lambda}\left\{1 - K^2\frac{2(\omega\tau_M\eta)^2}{1 + (\omega\tau_M\eta)^2} + iK^2\frac{\omega\tau_M\eta}{1 + (\omega\tau_M\eta)^2}\right\}.$$

Отсюда, с тем же приближением, имеем

$$q \simeq \omega\sqrt{\frac{\rho}{\Lambda}}\left\{1 - \frac{1}{2}K^2\frac{(\omega\tau_M\eta)^2}{1 + (\omega\tau_M\eta)^2} + i\,\frac{1}{2}K^2\frac{\omega\tau_M\eta}{1 + (\omega\tau_M\eta)^2}\right\}. \tag{4.26}$$

Тогда по формулам (4.17) находим для коэффициента поглощения

$$\gamma = K^2\frac{\omega}{v_{s0}}\frac{\omega\tau_M\eta}{1 + (\omega\tau_M\eta)^2}, \tag{4.27}$$

где $v_{s0} = \sqrt{\Lambda/\rho}$. Фазовая скорость волны оказывается равной

$$v_s = v_{s0}\left\{1 + \frac{1}{2}K^2\frac{(\omega\tau_M\eta)^2}{1 + (\omega\tau_M\eta)^2}\right\}. \tag{4.28}$$

Остановимся теперь подробнее на поглощении волны. Удобно исследовать поведение коэффициента поглощения, выраженного в единицах $K^2\omega/v_{s0}$. Из формулы (4.27) видно, что он зависит от частоты ω и τ_M (т. е. от электропроводности кристалла), причем обе эти величины входят только в виде произведения. Если внешнего электрического поля нет ($v_d = 0$, $\eta = 1$), то величина $\gamma v_{s0}/K^2\omega$

имеет максимум при $\omega\tau_M = 1$. При этом

$$\gamma_{\max} = \frac{1}{2} K^2 \frac{\omega}{v_{s0}}. \tag{4.23}$$

Зависимость $\gamma/\gamma_{\max}$ от $\omega\tau_M$ для случая $v_d = 0$ показана на рис. 15.2.

Однако наиболее интересная особенность поведения γ заключается в его зависимости от внешнего поля. Характер этой зависимости показан на рис. 15.3. Когда электроны и волны движутся

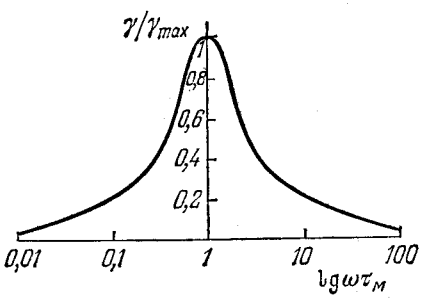


Рис. 15.2. Зависимость электронного коэффициента поглощения звука от $\omega\tau_M$ без внешнего поля.

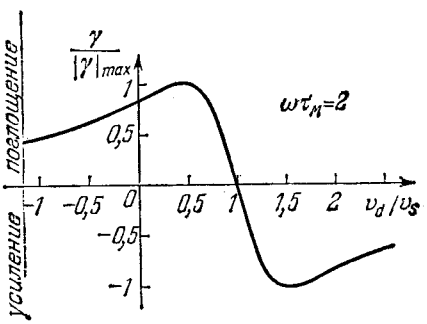


Рис. 15.3. Зависимость электронного коэффициента поглощения звука от дрейфовой скорости электронов.

в одинаковом направлении ($v_d/v_s > 0$), при $v_d > v_s$ в формуле (4.27) η изменяет знак и γ становится отрицательным. Это значит, что волна не затухает, а усиливается сверхзвуковым дрейфом электронов.

Физический смысл этих явлений поясняет рис. 15.4, где показано мгновенное распределение в пространстве различных электрических величин. Упругая волна вызывает колебания электростатического потенциала $\tilde{\varphi}$ и потенциальной энергии $-e\tilde{\varphi}$ электронов. В проводящих средах это приводит к перераспределению электронов, т. е. к экранированию поля, в результате чего образуются электронные сгустки $\tilde{n}$. Если бы волна не двигалась, эти сгустки были бы на дне потенциальных ям. В движущейся волне сгустки расположены на склонах ям. Образование электронных сгустков сопровождается возникновением переменных токов $\tilde{j}$ и электрических полей $\tilde{\mathcal{E}}$. В результате этого в каждой точке кристалла выделяется дополнительная мощность, которая, в расчете на единицу объема, есть

$$(j_0 + \tilde{j})(\mathcal{E}_0 + \tilde{\mathcal{E}}) - j_0\mathcal{E}_0 = \mathcal{E}_0\tilde{j} + j_0\tilde{\mathcal{E}} + \tilde{j}\tilde{\mathcal{E}}.$$

Среднее ее значение равно $\langle\tilde{j}\tilde{\mathcal{E}}\rangle$, так как $\langle\tilde{j}\rangle = \langle\tilde{\mathcal{E}}\rangle = 0$. Если $v_d < v_s$ (т. е. $\eta > 0$), то электроны отстают от звуковой волны и

группируются на заднем фронте потенциальных ям (рис. 15.4, слева). Согласно формуле (4.21) колебания $\tilde{j}$ и $\tilde{\mathcal{E}}$ находятся в фазе и $\langle \tilde{j}\tilde{\mathcal{E}} \rangle$ положительно. Эта мощность заимствуется от волны, что и приводит к ее затуханию. Если же $v_d > v_s$, то электроны, напротив, опережают звуковую волну и электронные сгустки располагаются на переднем фронте потенциальных ям (рис. 15.4, справа). Колебания $\tilde{j}$ и $\tilde{\mathcal{E}}$ находятся в противофазе (формула (4.21) при $\eta < 0$), и поэтому $\langle \tilde{j}\tilde{\mathcal{E}} \rangle$ отрицательно. Это значит, что теперь волна

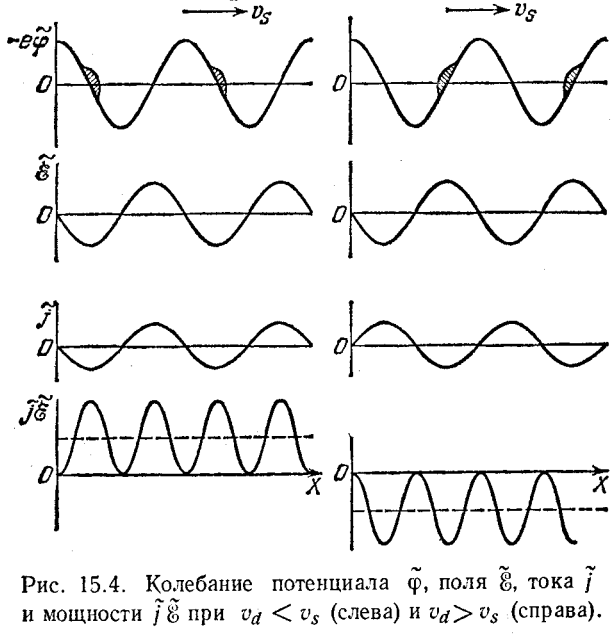


Рис. 15.4. Колебание потенциала $\tilde{\varphi}$, поля $\tilde{\mathcal{E}}$, тока $\tilde{j}$ и мощности $\tilde{j}\tilde{\mathcal{E}}$ при $v_d < v_s$ (слева) и $v_d > v_s$ (справа).

уже не передает свою энергию, а получает ее от дрейфующих электронов. Усиление волны происходит за счет энергии источника тока, поддерживающего дрейф электронов и совершающего дополнительную работу на преодоление тормозящих сил со стороны звуковой волны.

Напомним, что γ есть электронный коэффициент поглощения волны, обусловленный только ее взаимодействием с электронами. Так как имеется еще и поглощение волны самой решеткой, характеризуемое некоторым решеточным коэффициентом поглощения $\gamma_{\text{реш}}$, то суммарный коэффициент поглощения волны равен $(\gamma + \gamma_{\text{реш}})$, а условие ее усиления есть $(\gamma + \gamma_{\text{реш}}) < 0$.

Выше мы везде пренебрегали диффузией электронов. Если бы мы не вычеркнули в формуле (2.11) последнее слагаемое, то,

рассуждая, как и прежде, мы получили бы для коэффициента поглощения более сложное выражение:

$$\gamma = K^2 \frac{\omega}{v_{s0}} \frac{\omega\tau_M \eta}{(1+q_0^2 L_D^2)^2 + (\omega\tau_M \eta)^2}. \qquad (4.30)$$

Здесь L_D — длина экранирования Дебая, выражаемая формулой (VI.7.2), а $q_0 = \omega/v_{s0} = 2\pi/\lambda_з$, где $\lambda_з$ — длина звуковой волны. Сравнивая формулы (4.27) и (4.30), мы видим, что диффузия уменьшает γ. Это и понятно, так как диффузия противодействует группировке электронов. Далее, из этих формул видно, что учет диффузии становится существенным, если только $q_0^2 L_D^2 \gtrsim 1$, или, иначе, когда $\lambda_з < L_D$. Выбирая для n_0 даже не очень большое значение $\sim 10^{12} \div 10^{13}$ см$^{-3}$, мы имеем $L_D \sim 10^{-4} \div 10^{-5}$ см. Учитывая затем, что $v_{s0} \sim 10^5$ см/с, мы находим, что $q_0 L_D \sim 1$ при частоте $\omega \sim 10^9 \div 10^{10}$ с$^{-1}$. Таким образом, вплоть до очень высоких частот влиянием диффузии можно пренебречь.

Если бы мы учли еще и прилипание электронов, считая фактор прилипания f вещественным и постоянным («быстрые» ловушки), то мы получили бы прежние формулы (4.27) или, соответственно, (4.30), однако, в выражении для η (4.19) под v_d следовало бы понимать $-f\mu\mathcal{E}_0$, а длину Дебая L_D нужно было бы заменить на fL_D [3, 4].

§ 5. Электронное поглощение и усиление ультразвуковых волн

Остановимся теперь на экспериментах по электронному поглощению и усилению упругих волн. При звуковых частотах эти эффекты очень малы. Однако так как γ_{max} увеличивается с частотой (ср. (4.29)), то при частотах $\omega/2\pi \gtrsim 1 \div 10$ МГц (ультразвуковые волны) они становятся легко измеримыми.

Чтобы оценить порядок ожидаемых величин, рассмотрим конкретный пример сильного пьезополупроводника — сульфида кадмия (CdS). Его пьезоэлектрические свойства таковы, что если волна распространяется в направлении гексагональной оси C, то только продольные колебания оказываются пьезоэлектрически активными. Напротив, при распространении волны перпендикулярно оси C активной является только поперечная волна, в которой смещение параллельно C. Интересующие нас константы при 300 К имеют следующие значения:

CdS	Продольная волна	Поперечная волна
K^2	0,08	0,036
v_{s0}, см/с	$4{,}3 \cdot 10^5$	$1{,}7 \cdot 10^5$

Положим, далее, $\sigma = 10^{-4}$ Ом$^{-1}\cdot$см$^{-1}$, что для CdS соответствует $\tau_M = 1{,}2\cdot 10^{-8}$ с. Тогда при частоте $\omega/2\pi = 30$ МГц (при которой можно еще пренебречь влиянием диффузии) для продольной волны по формуле (4.29) мы находим

$$|\gamma_{max}| \simeq 18 \text{ см}^{-1}.$$

Укажем для сравнения, что в CdS на этой же частоте $\gamma_{реш} = 0{,}07$ см$^{-1}$, т. е. пренебрежимо мало по сравнению с γ *). Тогда при длине пластинки 0,5 см интенсивность волны должна усиливаться в $\exp(0{,}5\cdot 18) \simeq 10^4$ раз. Это максимальное усиление

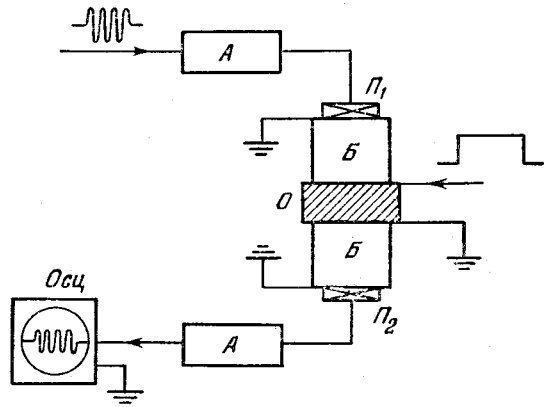


Рис. 15.5. Измерение электронного поглощения и усиления упругих волн в полупроводниках. *А* — аттенюаторы, *Осц* — осциллограф.

будет при $\eta = -1/\omega\tau_M = -0{,}44$, что соответствует $v_d/v_s = 1 - \eta = 1{,}5$. При подвижности электронов $\mu = 200$ см^2/с$\cdot$В для создания такой дрейфовой скорости необходимо электрическое поле $\mathcal{E}_0 = 3{,}2$ кВ/см. При бо́льших частотах ожидаемое усиление (поглощение) будет еще больше.

Схема типичного опыта для исследования электронного поглощения и усиления ультразвука показана на рис. 15.5. Здесь $П_1$ и $П_2$ — электромеханические преобразователи для генерации и регистрации ультразвуковых волн. В простейшем случае это пьезоэлектрические пластинки (например, из кристаллического кварца), ориентированные должным образом, грани которых металлизированы. При подаче на эти грани переменного напряжения с частотой, совпадающей с одной из собственных частот пластинки, в ней возникают механические колебания вследствие обратного

*) Однако следует иметь в виду, что $\gamma_{реш} \sim \omega^2$, и поэтому при более высоких частотах решеточное поглощение может стать заметным.

пьезоэлектрического эффекта, и она становится источником ультразвуковых волн. И обратно, при прохождении волны через преобразователь на обкладках появляется переменное электрическое напряжение в результате прямого пьезоэлектрического эффекта, пропорциональное амплитуде деформации в волне. O — исследуемая пластинка пьезополупроводника, грани которой также металлизированы. К ним прикладывают постоянное электрическое напряжение, создающее тянущее поле $\mathcal{E}_0$. Для уменьшения нагревания кристалла это поле создают обычно в виде коротких импульсов (1 ÷ 10 микросекунд). Между преобразователями и образцом O часто вводят еще диэлектрические звукопроводы — буферы $Б$, вызывающие контролируемое запаздывание звуковой волны. Это позволяет отделить во времени импульс, создаваемый в приемном преобразователе звуковой волной, от возможной радиопомехи, наводимой напряжением на излучающем преобразователе. В подобных опытах особенно удобно исследовать фотопроводящие пьезоэлектрические кристаллы. Тогда, меняя интенсивность освещения, можно изменять в широких пределах электропроводность кристалла и исследовать влияние электропроводности (τ_M) на поглощение и усиление волн.

Такие опыты показывают, что вследствие взаимодействия с электронами интенсивность волн может изменяться во много тысяч и даже десятков тысяч раз. Так, например, высокоомные фоточувствительные кристаллы CdS, которые при частотах порядка 10 ÷ ÷ 100 МГц практически не поглощают волн, при освещении становятся совершенно непрозрачными для ультразвука. При наложении на кристалл постоянного напряжения, большего чем критическое, сильное поглощение сменяется большим усилением. Наблюдаемые значения γ в широком интервале изменения интенсивности хорошо согласуются с формулой (4.30).

Однако это, конечно, не значит, что, увеличивая длину кристалла, можно усилить волну в любой степени. Когда интенсивность волны становится очень большой, возникают различные нелинейные эффекты (не учтенные в приведенной выше линейной теории), которые ограничивают максимальное усиление. Один из таких эффектов заключается в том, что концентрация электронов в сгустках $\tilde{n}$ становится сравнимой с первоначальной концентрацией n_0 и в конце концов все электроны переходят в сгустки. После этого поле $\tilde{\mathcal{E}}$ уже не вызывает токов $\tilde{j}$, и при дальнейшем распространении волны она перестает усиливаться.

§ 6. Акусто-электрический эффект

При поглощении упругих волн на электронах проводимости в кристалле возникает электродвижущая сила и между разомкнутыми гранями кристалла появляется разность потенциалов

(рис. 15.6). Если нанести на грани металлические контакты, то при их замыкании в цепи возникнет электрический ток. При перенесении источника волн на противоположную грань кристалла, т. е. при изменении направления волны на противоположное, эдс изменяет знак и, соответственно, направление тока меняется на обратное.

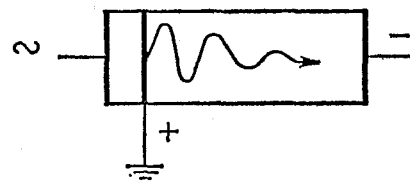


Рис. 15.6. Акусто-электрический эффект. Знак заряда граней показан для полупроводника n-типа.

Общая причина этого эффекта состоит в том, что упругая волна несет с собой определенный механический импульс. При поглощении волны ее импульс передается электронам проводимости, что приводит к появлению некоторой средней силы, действующей на электроны, направленной в сторону распространения волны. Следовательно, акусто-электрический эффект можно описать как увлечение электронов звуковой волной.

Найдем величину акусто-электрической эдс. Наиболее просто это можно сделать, если воспользоваться общим соотношением между интенсивностью волны, т. е. плотностью потока энергии I, и переносимым ею механическим импульсом $\mathbf{P}$ (рассчитанным, как и I, на единицу площади и единицу времени):

$$\mathbf{P} = \frac{I}{v_s}\mathbf{q}_1. \tag{6.1}$$

Здесь $\mathbf{q}_1$ — единичный вектор в направлении волны. Это соотношение получается непосредственно, если рассматривать волну как поток фононов, каждый из которых несет энергию $\hbar\omega$ и импульс $\hbar\mathbf{q}$, и учесть, что $\omega/q = v_s$. Однако формула (6.1) в действительности не связана с квантовыми свойствами волн (что видно хотя бы из того, что в окончательном результате $\hbar$ выпадает). Рассмотрим теперь внутри кристалла бесконечно тонкий слой толщиной dx, перпендикулярный $\mathbf{q}_1$, и с единичной площадью. В этом слое в единицу времени поглощается энергия $\gamma I\, dx$, и он, согласно (6.1), получает импульс $(\gamma I\, dx/v_s)\,\mathbf{q}_1$. Этот импульс равен силе, действующей на $n_0\, dx$ электронов внутри слоя. Поэтому, если $\mathbf{F}_1$ есть средняя сила на один электрон, мы имеем

$$\frac{\gamma I\, dx}{v_s}\mathbf{q}_1 = \mathbf{F}_1 n_0\, dx.$$

Вводя сюда напряженность поля сторонних сил

$$\mathcal{E}^* = -\frac{1}{e}\mathbf{F}_1$$

(силу на единицу положительного заряда), мы получаем окончательно

$$\mathfrak{E}^* = -\frac{\gamma I}{en_0 v_s}\mathbf{q}_1 = -\frac{\mu\gamma I}{\sigma_0 v_s}\mathbf{q}_1. \tag{6.2}$$

В этих рассуждениях мы считали, что прилипания электронов на ловушки нет. Если прилипание есть, то, как уже говорилось в § 4, вместо μ следует подставить $f\mu$, где f — фактор прилипания ($\leqslant 1$). Поэтому в (6.2) войдет дополнительный множитель f.

Полная эдс в пластинке длиной d равна

$$V_0 = \int_0^d \mathfrak{E}^*\, dx = -\frac{f\mu\gamma}{\sigma_0 v_s}\int_0^d I(0)\, e^{-(\gamma+\gamma_{\text{реш}})x}\, dx =$$

$$= -\frac{f\mu}{\sigma_0 v_s}\frac{\gamma}{\gamma+\gamma_{\text{реш}}}[I(0) - I(d)]. \tag{6.3}$$

Такая же разность потенциалов будет между концами пластинки, если внешняя цепь разомкнута.

Найдем еще ток короткого замыкания. Будем считать, что волна поглощается слабо. Тогда можно положить

$$I(0) - I(d) \simeq I(0) - I(0)[1 - (\gamma + \gamma_{\text{реш}})\, d] = (\gamma + \gamma_{\text{реш}})\, Id$$

и

$$V_0 = -f\frac{\mu\gamma I}{\sigma_0 v_s}\, d.$$

Так как сопротивление пластинки есть $R = d/\sigma_0 S$, где S — сечение пластинки, то для плотности тока короткого замыкания

$$j_{ae} = -f\frac{\mu\gamma I}{v_s}. \tag{6.4}$$

Знак минус в формулах (6.2) и (6.4) вошел потому, что мы вели рассуждения для электронов. Он обозначает, что отрицательные электроны, увлекаясь волной, создают ток, текущий противоположно направлению распространения волны. Если бы мы имели полупроводник p-типа, то положительные дырки также увлекались бы в направлении волны. Однако в этом случае направление тока совпадало бы с направлением увлечения и в формулы вошел бы знак плюс.

Так обстоит дело, если скорость дрейфа электронов $v_d = 0$, или $v_d < v_s$. При $v_d > v_s$ вместо поглощения волны происходит ее усиление и γ изменяет знак. При этом направление j_{ae} изменяется на обратное.

Мы получили выражения для акусто-электрических тока и напряжения из общих феноменологических рассуждений. Эти же соотношения можно получить и другим путем, анализируя токи и поля, возникающие в кристалле под действием звуковой волны.

Оценим еще величину эдс акусто-электрического эффекта. Положим в формуле (6.3) $I(d) \ll I(0)$, $\gamma_{\text{реш}} \ll \gamma$, $f = 1$. Примем, далее, $\mu \sim 10^2$ см$^2 \cdot$с$^{-1} \cdot$В$^{-1}$ (что типично, например, для сульфида кадмия при комнатных температурах) и $\sigma \sim 10^{-4}$ Ом$^{-1} \cdot$см$^{-1}$. Тогда, учитывая, что $v_s \sim 10^5$ см$\cdot$с$^{-1}$, мы находим, что при $I(0) \sim 1$ Вт$\cdot$см$^{-2}$ эдс V_0 имеет порядок 10 В.

Отметим в заключение, что эффект увлечения электронов возникает не только при воздействии звуковых волн, но и при поглощении волн других типов, если только эти волны переносят энергию и импульс и взаимодействуют с электронами проводимости. В частности, это имеет место при поглощении электромагнитных волн радио- и светового диапазонов (радиоэлектрический эффект). Однако в этом случае в формулу (6.1) вместо v_s входит скорость света c, которая примерно в 10^5 раз больше, и поэтому возникающая эдс, при прочих равных условиях, оказывается намного меньше.

§ 7. Случай $ql \gg 1$

Выше мы считали, что выполняется неравенство $ql \ll 1$ («низкие» частоты), и, соответственно, пользовались гидродинамическим описанием явлений. Сейчас мы обратимся к противоположному случаю и, соответственно, будем рассматривать взаимодействие одного электрона и одного фонона. Тогда поглощение и усиление звука можно описать следующим образом. Как мы видели в § 1, для возможности поглощения или испускания фонона импульс электрона p_x до взаимодействия должен удовлетворять условию (1.4) или, соответственно, (1.4а). При этом вероятности прямого элементарного акта (поглощения фонона) и ему обратного (испускания фонона) одинаковы. Однако частота таких элементарных актов (в единице объема и в единицу времени) будет различна. Действительно, частота переходов ($p_{x1} \to p_{x2}$) с поглощением фонона будет пропорциональна, во-первых, вероятности f_1 заполнения исходных квантовых состояний электронов с импульсом p_{x1} и, во-вторых, вероятности $(1 - f_2)$ того, что конечные квантовые состояния не заняты электронами. Аналогично, частота обратных переходов ($p_{x2} \to p_{x1}$) с испусканием фонона будет пропорциональна $f_2(1 - f_1)$. Коэффициент поглощения волны γ пропорционален суммарной частоте поглощения фононов, т. е.

$$\gamma \sim f_1(1 - f_2) - f_2(1 - f_1) = f_1 - f_2.$$

Если скорость дрейфа $v_d = 0$, то f_1 и f_2 выражаются функцией Ферми. В случае изотропного параболического закона дисперсии она зависит от аргумента

$$u = \frac{p_x^2 + p_y^2 + p_z^2}{2mkT}, \tag{7.1}$$

а для невырожденного полупроводника переходит в функцию Максвелла — Больцмана

$$f = C\exp\left(-\frac{p_x^2+p_y^2+p_z^2}{2mkT}\right). \tag{7.2}$$

Для дальнейшего мы учтем, что тепловая скорость электронов $v_T \sim 10^7$ см/с, в то время как $v_s \sim 10^5$ см/с, а следовательно,

$$\frac{p_x^2}{2mkT} \sim \frac{v_s^2}{v_T^2} \ll 1. \tag{7.3}$$

Поэтому с достаточной точностью можно положить

$$(f_1 - f_2)_0 = -f'(p_{xs})\left(\frac{\partial u}{\partial p_x}\right)_{p_{xs}}(p_{x2}-p_{x1}) = -f'(p_{xs})\frac{p_{xs}}{mkT}\hbar q. \tag{7.4}$$

Здесь $p_{xs} = mv_s$, а через f' обозначена производная от f по аргументу u.

Положим теперь, что электроны дрейфуют во внешнем электрическом поле в направлении оси X. Будем считать, что время релаксации импульса τ слабо зависит от энергии электронов. Тогда всем электронам можно приписать одинаковую подвижность $\mu = e\tau/m$ и одинаковую скорость дрейфа. Как будет показано в § XVI. 4, при определенных условиях функция распределения по квазиимпульсам при наличии дрейфа зависит от аргумента

$$\frac{(p_x-p_x')^2+p_y^2+p_z^2}{2mkT}, \tag{7.5}$$

где $p_x' = mv_d$. В невырожденном полупроводнике мы получим «смещенное» распределение Максвелла — Больцмана

$$f = C\exp\left(-\frac{(p_x-p_x')^2+p_y^2+p_z^2}{2mkT}\right). \tag{7.6}$$

Поэтому вместо формулы (7.4) мы будем иметь

$$(f_1 - f_2)_{v_d} = -f'(p_{xs}-p_x')\frac{p_{xs}-p_x'}{mkT}\hbar q. \tag{7.7}$$

Для отношения коэффициентов поглощения волны при дрейфе $\gamma(v_d)$ и без дрейфа $\gamma(0)$ получается

$$\frac{\gamma(v_d)}{\gamma(0)} = \frac{(f_1-f_2)_{v_d}}{(f_1-f_2)_0} = \psi\cdot\left(1-\frac{p_x'}{p_{xs}}\right) = \psi\cdot\left(1-\frac{v_d}{v_s}\right), \tag{7.8}$$

где обозначено

$$\psi = \frac{f'(p_{xs}-p_x')}{f'(p_{xs})}.$$

Нетрудно видеть, что функция $\psi \simeq 1$. Действительно, учитывая выражения (7.2) и (7.6), мы имеем

$$\psi = \exp\left(-\frac{(p_{xs}-p'_x)^2-p_{xs}^2}{2mkT}\right) = \exp\left\{\frac{p_{xs}^2}{mkT}\frac{v_d}{v_s}\left(1-\frac{1}{2}\frac{v_d}{v_s}\right)\right\}.$$

В силу неравенства (7.3) показатель экспоненты гораздо меньше единицы, и поэтому приближенно это выражение можно заменить единицей. Таким образом, мы получаем окончательно

$$\gamma(v_d) = \gamma(0)\left(1-\frac{v_d}{v_s}\right). \qquad (7.9)$$

Так же, как и для длинных волн, γ зависит от отношения v_d/v_s. При $v_d > v_s$ поглощение потока фононов сменяется генерацией фононов. Однако кривые зависимости $\gamma(v_d)$ уже не имеют вида, показанного на рис. 15.3, а изображаются прямыми линиями (7.9).

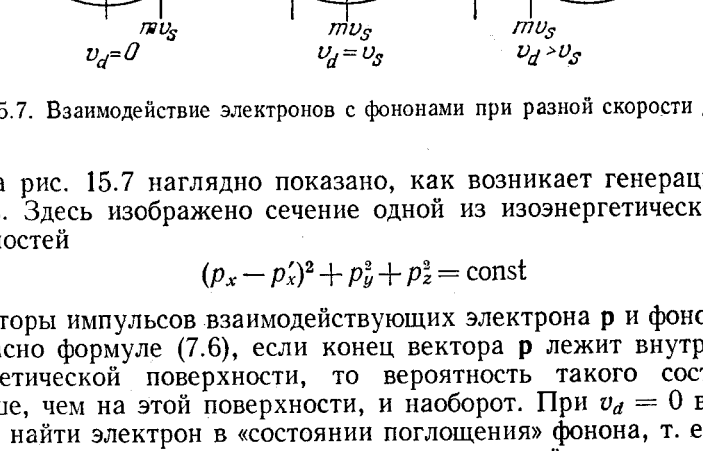


Рис. 15.7. Взаимодействие электронов с фононами при разной скорости дрейфа.

На рис. 15.7 наглядно показано, как возникает генерация фононов. Здесь изображено сечение одной из изоэнергетических поверхностей

$$(p_x-p'_x)^2+p_y^2+p_z^2 = \text{const}$$

и векторы импульсов взаимодействующих электрона $\mathbf{p}$ и фонона $\hbar\mathbf{q}$. Согласно формуле (7.6), если конец вектора $\mathbf{p}$ лежит внутри изоэнергетической поверхности, то вероятность такого состояния больше, чем на этой поверхности, и наоборот. При $v_d = 0$ вероятность найти электрон в «состоянии поглощения» фонона, т. е. с импульсом $\mathbf{p}_1$, больше, чем вероятность «состояний испускания» фонона (импульс $\mathbf{p}_2$), и поэтому преобладают процессы поглощения. При $v_d = v_s$ концы векторов $\mathbf{p}_1$ и $\mathbf{p}_2$ лежат на одной и той же изоэнергетической поверхности (пунктир), а следовательно, частоты актов поглощения и испускания фононов одинаковы. Если же $v_d > v_s$, то концы векторов $\mathbf{p}_2$ располагаются на изоэнергетической поверх-

ности меньшего радиуса, чем концы векторов $\mathbf{p}_1$. В этом случае число электронов в «состоянии испускания» больше, чем в «состоянии поглощения», и суммарный результат есть генерация фононов. Можно сказать, что при сверхзвуковом дрейфе электронов возникает *инверсия заселенности квантовых состояний* (по сравнению со случаем $v_d = 0$), аналогичная инверсии заселенности энергетических уровней в оптических квантовых генераторах (лазерах, см. § XVIII.6).

Рассматриваемые явления имеют глубокую аналогию и с другим важным явлением — *излучением Вавилова — Черенкова*, которое заключается в генерации света под действием быстрых электронов (β-частиц), когда их скорость становится больше фазовой скорости света в среде c/n. Усиление гиперзвука есть, по сути дела, излучение Вавилова — Черенкова, однако не квантов света, а квантов звука — фононов.

§ 8. Усиление тепловых флуктуаций

Вернемся теперь опять к длинным волнам ($ql \ll 1$) и к коллективному взаимодействию. В этом случае поглощение и усиление упругих волн сверхзвуковым дрейфом электронов имеет важную особенность. Если создать скорость дрейфа $v_d > v_s$ в направлении волны и затем изменить направление поля на обратное, то усиление волны в первом случае будет больше, чем ее поглощение при изменении направления дрейфа. Это связано с тем, что кривые $\gamma(v_d)$ симметричны относительно точки $\eta = 1 - v_d/v_s = 0$, но не точки $v_d = 0$. Так, например, если $v_d/v_s = 1{,}5$ (усиление), то в случае рис. 15.3 $|\gamma| = |\gamma|_{\max}$. Изменяя же направление дрейфа (поглощение), мы найдем, что $v_d/v_s = -1{,}5$ соответствует только $\gamma \simeq \simeq {}^1/_3\gamma_{\max}$. Или, иначе, дрейфующая плазма обладает свойством *невзаимности* по отношению к упругим волнам. Поэтому, если волна, полностью отражаясь от граней пластинки, проходит один прямой и один обратный путь (замкнутый цикл), она получает выигрыш мощности.

Подчеркнем, что этот результат справедлив только для коллективного взаимодействия ($ql \ll 1$). Для очень коротких волн ($ql \gg 1$) зависимость γ от η выражается формулой (7.9) и имеет другой характер. В этом случае при замкнутом цикле энергия волны не увеличивается, а, напротив, уменьшается.

Так как при $v_d > v_s$ усиливаются также и тепловые флуктуации решетки, то в режиме усиления наблюдаются акустические шумы. Если импульс ускоряющего поля короче времени прохождения звука через пластинку d/v_s, то флуктуации усиливаются лишь при однократном прохождении в направлении дрейфа (однопролетные шумы). Если же импульс поля длинный, то тепловые флуктуации, многократно отражаясь от граней пластинки, проходят

много замкнутых циклов усиления, и тогда наблюдаются гораздо более сильные шумы (многопролетные шумы).

При определенных условиях в пластинке могут возникнуть не только шумы, но и спонтанные правильные колебания. Причина этого заключается в следующем. Наиболее выгодные условия для усиления мы имеем, когда фаза волны после каждого цикла «вперед — назад» увеличивается на $2\pi n$ ($n = 1, 2, \ldots$), так как в этом случае амплитуды будут складываться вследствие интерференции. Условие этого есть

$$2\pi\left(\frac{d}{\lambda_1}+\frac{d}{\lambda_2}\right)+\varphi_1 = d\left(\frac{\omega_n}{v_{s1}}+\frac{\omega_n}{v_{s2}}\right)+\varphi_1 = 2\pi n. \quad (8.1)$$

Здесь мы учли, что скорость волны вдоль дрейфа v_{s1} слегка отлична от скорости v_{s2} в противоположном направлении, а через φ_1 обозначили изменение фазы при отражениях. Это соотношение определяет дискретный набор частот ω_1, ω_2, ... , для которых усиление будет особенно велико. Однако, конечно, для каждой из этих мод усиление будет различным, так как γ зависит от частоты (ср. формулу (4.27)). Если для какой-либо колебательной моды выигрыш в мощности станет больше потерь (которые обусловлены в первую очередь поглощением в самой решетке и неполным отражением волн от граней), то эта мода будет самовозбуждаться и усиливаться до тех пор, пока возрастающие нелинейные потери не ограничат это усиление. Мы получим генератор ультразвуковых волн. При этом во внешней цепи, вследствие пьезоэлектрического эффекта, появятся колебания тока.

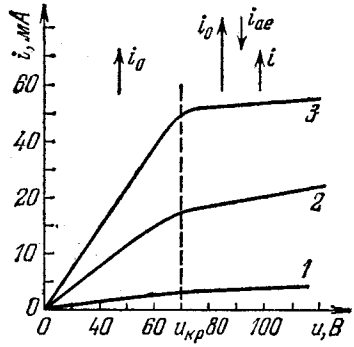


Рис. 15.8. Пример вольтамперных характеристик кристаллов сульфида кадмия. Длина (в направлении поля) 0,05 см, сечение 0,015 мм². Электропроводность увеличивается от кривой *1* к кривой *3*.

При постепенном увеличении электрического поля самовозбуждаться будет все большее число различных мод. В кристалле появятся и их гармоники, так же как и комбинационные частоты ($\omega_n \pm \omega_m$), и форма колебаний станет сложной. Однако из формулы (8.1) видно, что при уменьшении толщины пластинки d разность частот двух соседних мод ($\omega_{n+1} - \omega_n$) увеличивается. Поэтому в тонких пластинках гораздо легче возбудить только одну колебательную моду (или небольшое их число) и получить генерацию приблизительно монохроматических колебаний.

Усиление тепловых колебаний решетки проявляется также в нелинейности вольтамперных характеристик пьезоэлектрических

кристаллов. На рис. 15.8 показана зависимость тока от напряжения в кристаллах сульфида кадмия. Различные кривые соответствуют разным электропроводностям. Пока напряжение остается ниже некоторого критического значения $u_{кр}$, характеристики линейны. Однако при $u > u_{кр}$ ток оказывается меньше, чем следует из закона Ома, и на характеристиках появляется излом, который выражен тем лучше, чем больше электропроводность. При большой электропроводности кристаллов и $u > u_{кр}$ ток почти перестает зависеть от напряжения (насыщение тока). Величина же $u_{кр}$ соответствует условию $v_d \simeq v_s$. Такое поведение кристаллов объясняется тем, что при $v_d > v_s$ возникает черенковское излучение фононов и электроны получают тормозящий импульс. Или, иными словами, в кристалле возникает акустоэлектрический ток i_{ae}, который при $v_d > v_s$ направлен против тока i_0, создаваемого внешним напряжением. Поэтому результирующий ток i оказывается меньше i_0. В предельном случае достаточно большого напряжения, когда практически все электроны захвачены в потенциальные ямы усиленных упругих волн, все электроны движутся со скоростью v_s и плотность тока насыщения равна $j_s = en_0v_s$.



Рис. 15.9. Вольтамперные характеристики кристаллов висмута в сильном магнитном поле. Размеры ~ 1 мм$^2 \times 1$ мм; $T = 2$ К; *1* — $\mathcal{B} = 14$ кГс, *2* — $\mathcal{B} = 20$ кГс.

Аналогичная аномалия в электрическом сопротивлении наблюдается в кристаллах висмута в скрещенных электрическом и магнитном полях. Она очень резко выражена при условии $\omega_c\tau > 1$ (где ω_c — циклотронная частота, τ — время релаксации импульса), что осуществляется при низкой температуре и сильном магнитном поле. В этом случае на вольтамперных характеристиках тоже наблюдается резкий излом (рис. 15.9), однако, в отличие от пьезоэлектрических кристаллов, характеристики изгибаются вверх. Объяснение этого явления, в общих чертах, заключается в следующем. В скрещенных полях носители заряда не только дрейфуют в электрическом поле $\mathcal{E}$, создавая плотность тока $\mathbf{j}_0$, подчиняющуюся закону Ома, но имеют еще скорость переносного движения, перпендикулярную к $\mathcal{E}$ и $\mathcal{B}$ и равную

$$\mathbf{v}_t = c\frac{[\mathcal{E} \times \mathcal{B}]}{\mathcal{B}^2}.$$

В висмуте, который является полуметаллом, имеются и электроны, и дырки, которые движутся со скоростью $\mathbf{v}_t$ в одинаковом

направлении. Дрейфующие дырки и электроны взаимодействуют с фононами — в данном случае висмута — через потенциалы деформации. При $v_t > v_s$ возникает черенковская генерация фононов, и дырки и электроны получают тормозящие импульсы, направленные противоположно нарастающему потоку генерируемых фононов. Это приводит к появлению акусто-электрических токов дырок $\mathbf{j}_p$ и электронов $\mathbf{j}_n$. Эти токи отклоняются в магнитном поле на соответствующие им холловские углы и поэтому не перпендикулярны $\mathbf{v}_t$ и $\mathcal{E}$. Они имеют составляющие вдоль $\mathcal{E}$, которые, как видно из рис. 15.9, прибавляются к току $\mathbf{j}_0$. Поэтому результирующий ток $\mathbf{j}$ оказывается больше $\mathbf{j}_0$.

§ 9. Заключительные замечания

Выше мы везде рассматривали только объемные, продольные или поперечные, упругие волны. Однако все рассмотренные явления — электронное поглощение и усиление волн, акусто-электрический эффект и др. — наблюдаются и в других типах волн. Особенно интересными здесь являются поверхностные волны. В таких

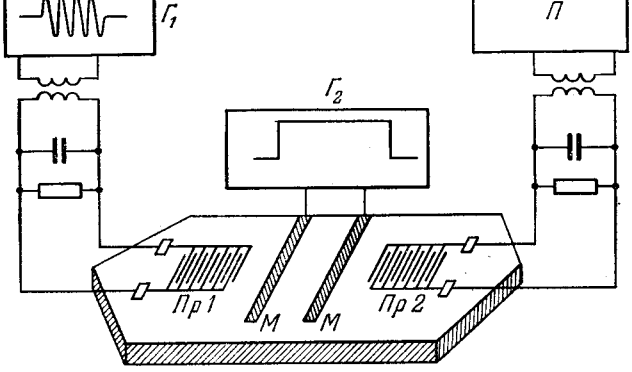


Рис. 15.10. Усиление поверхностных упругих волн дрейфом электронов. $Г_1$ — импульсный генератор высокой частоты, $Г_2$ — генератор прямоугольных импульсов ускоряющего напряжения, *П* — радиоприемник, *Пр1* и *Пр2* — электромеханические преобразователи, *М*, *М* — металлические ускоряющие электроды.

волнах в общем случае имеются нормальная к поверхности составляющая смещения и тангенциальная составляющая, которые, в отличие от объемных волн, затухают при удалении от поверхности.

Схема типичного опыта для наблюдения электронного поглощения и усиления поверхностных волн в пьезоэлектрических

кристаллах показана на рис. 15.10. В качестве электромеханических преобразователей служат две системы гребенчатых металлических электродов, вставленные одна в другую. Переменное электрическое напряжение, подаваемое на преобразователь, должно иметь такую частоту, чтобы длина упругой поверхностной волны была равна периоду гребенки. Разумеется, и здесь кристаллографические оси пластинки должны быть ориентированы так, чтобы смещение в волне было пьезоэлектрически активным.

Исследование акусто-электронных явлений (на объемных и поверхностных волнах) позволяет получить ценную информацию об особенностях взаимодействия колебаний решетки с электронами проводимости. Оно дает новые возможности определения важных характеристик полупроводниковых кристаллов: дрейфовой подвижности носителей заряда, константы электромеханической связи, деформационных потенциалов и др.

Акусто-электронные явления могут также быть использованы для создания новых твердотельных электронных приборов. Так, используя акусто-электрический эффект, можно получить малогабаритные устройства для регистрации и измерения интенсивности ультразвуковых волн. Явление усиления упругих волн с помощью дрейфа электронов можно использовать для создания усилителей и генераторов ультра- и гиперзвуковых колебаний. Используя двойное преобразование: электрический сигнал — звуковая волна и звуковая волна — электрический сигнал, можно получить усилители и генераторы электрических колебаний. Двойное преобразование совместно с усилением звука позволяет осуществить линии задержки электрических сигналов, которые, по сравнению с обычными линиями задержки электромагнитных волн, выгодно отличаются большими временами задержки (вследствие того, что $v_s \ll c$) и компенсацией потерь. Возможно создание и более сложных функциональных схем для различного преобразования и обработки информации, на которых мы, однако, останавливаться не будем.

Глава XVI

ГОРЯЧИЕ ЭЛЕКТРОНЫ

§ 1. Нагрев электронного газа

Ускоряясь электрическим полем, электроны и дырки могут приобретать не только средний импульс, но и энергию. Действительно, средняя энергия, приобретаемая единицей объема газа носителей заряда в единицу времени, равна $(\mathbf{j}, \mathcal{E})$. В стационарном состоянии эта величина должна равняться энергии, в среднем теряемой электронами *) в единицу времени при столкновениях. Обозначим через $\dot{E}$ отнесенное к единице времени изменение энергии носителя заряда за счет столкновений. Эта величина зависит от энергии E, уже имеющейся у электрона. Усредняя ее по всем состояниям электрона, получим интересующую нас величину $\langle \dot{E} \rangle_{\text{ст}}$. Условие баланса энергии в стационарном состоянии поэтому есть

$$(\mathbf{j}, \mathcal{E}) = n \langle \dot{E} \rangle_{\text{ст}}. \tag{1.1}$$

Используемый здесь символ $\langle \ldots \rangle$ не следует смешивать с аналогичным обозначением, использованным в гл. I, XIII. Там усреднение выполнялось с помощью равновесной функции f_0, здесь — с помощью неравновесной функции распределения, описывающей данное неравновесное стационарное состояние газа носителей заряда.

При термодинамическом равновесии, когда $\mathbf{j} = 0$, равенство (1.1) превращается в тождество «0 = 0». Действительно, в этих условиях электроны отдают решетке (в единицу времени) в среднем столько же энергии, сколько и получают от нее, т. е. $\langle \dot{E} \rangle_{\text{ст}} = 0$. В достаточно слабом поле, когда $\mathcal{E} \ll \mathcal{E}_{\text{кр}}$ (см. (XIII.7.9′) или (XIII.7.9″)), величина $(\mathbf{j}, \mathcal{E})$ — второго порядка по напряженности поля. Соответственно очень мала и правая часть (1.1). При возрастании поля положение меняется, ибо энергия, которую электроны получают от поля, заметно возрастает. В то же время правая часть равенства (1.1) определяется вероятностью соответствующих столкновений. Как мы видели в гл. XIV, процессы рассеяния носителей заряда часто оказываются почти упругими. По этой причине величина $\langle \dot{E} \rangle_{\text{ст}}$, вычисленная в условиях, близких к термодинамиче-

*) Мы будем для краткости отождествлять термины «носитель заряда» и «электрон»; исключение составят лишь явно оговариваемые случаи. Будем также рассматривать лишь невырожденный газ.

скому равновесию, может оказаться меньше $(\mathbf{j}, \hat{\mathcal{E}})$. При этом средняя энергия носителя заряда $\langle E\rangle_{\text{ст}}$ начнет возрастать над равновесным своим значением. Так будет продолжаться до тех пор, пока вероятность тех или иных неупругих процессов не увеличится достаточным образом. При не слишком высоких значениях средней энергии электронов роль такого процесса играет обычное рассеяние на фононах, связанное с испусканием их. При дальнейшем повышении напряженности электрического поля и, следовательно, $\langle E\rangle_{\text{ст}}$ могут стать заметными процессы ударной ионизации, при которых энергия электрона расходуется на создание новых носителей заряда (см. ниже § 6, стр. 536).

Замечая, что правая часть (1.1) обращается в нуль при $\langle E\rangle_{\text{ст}} = {}^3/_2\, kT$, можем представить (1.1) в виде

$$(\mathbf{j}, \hat{\mathcal{E}}) = \frac{{}^2/_3 \langle E\rangle_{\text{ст}} - kT}{\tau_e}\, n. \tag{1.2}$$

Величина τ_e, определяемая этим соотношением, имеет размерность времени и называется *средним временем релаксации энергии* (слово «среднее» часто для краткости опускают). Она зависит от механизма рассеяния носителей заряда, от температуры и, может быть, от других величин, характеризующих условия опыта. Оценку τ_e можно получить, измеряя, например, подвижность носителей как функцию напряженности поля. В зависимости от условий опыта τ_e может составлять $10^{-10} \div 10^{-7}$ с (последнее — при достаточно низких температурах и в не слишком сильных полях; при этом рассеяние энергии определяется взаимодействием электронов с фононами, а вероятность испускания их невелика (гл. XIV)). В условиях, когда процессы рассеяния почти упругие, величина τ_e может значительно превышать среднее время релаксации импульса τ_p. По этой причине разность $\frac{2}{3}\langle E\rangle_{\text{ст}} - kT$ может составить заметную долю средней энергии теплового движения электронов. Иначе говоря, благодаря сравнительной медленности процесса обмена энергией между носителями заряда и их окружением в решетке средняя энергия носителей в электрическом поле может оказаться больше своего термодинамически равновесного значения. Это явление называется нагревом электронного газа. Часто его описывают, вводя представление об *электронной температуре* T_e, отличающейся от температуры решетки T. Выбирая систему отсчета, в которой система носителей заряда как целое покоится, мы можем определить T_e равенством

$$\langle E\rangle_{\text{ст}} = {}^3/_2 kT_e. \tag{1.3}$$

Название «горячие электроны» приобретает теперь наглядный смысл: температура T_e может превысить T. Ясна также и причина, по которой система отсчета была выбрана указанным выше образом: в любой другой системе отсчета средняя энергия $\langle E\rangle_{\text{ст}}$ содержала

бы слагаемое, связанное с кинетической энергией движения системы носителей заряда как целого. Эту величину иногда называют энергией дрейфа E_d. Нецелесообразно связывать ее с температурой, ибо последняя величина описывает среднюю энергию беспорядочного движения частиц. Таким образом, при наличии дрейфа носителей заряда вместо (1.3) следует писать

$$\langle E\rangle = \frac{3}{2}kT_e + E_d. \tag{1.3'}$$

С учетом (1.3) правая часть (1.2) принимает такой же вид, как и в обычной теории теплопередачи:

$$\langle \dot{E}\rangle = \frac{k(T_e - T)}{\tau_e}. \tag{1.4}$$

Роль подсистем, обменивающихся теплом, играют здесь газ носителей заряда и кристаллическая решетка.

Следует, однако, иметь в виду, что эта аналогия не всегда точна. Действительно, величина T_e, введенная равенством (1.3), есть, вообще говоря, не более чем обозначение. Она может и не обладать всеми термодинамическими свойствами обычной температуры. Последнее понятие имеет однозначный смысл лишь в условиях термодинамического равновесия; нагрев же электронного газа как раз и означает, что это равновесие нарушено. Как видно из (1.4), «температура» T_e зависит от напряженности электрического поля и от механизмов рассеяния энергии и импульса. Последние определяют времена релаксации энергии и импульса, входящие, соответственно, в формулу (1.4) и в выражение для плотности тока $\mathbf{j}$.

В силу своей наглядности представление об электронной температуре оказывается очень удобным. Так, оно позволяет сразу заметить, что в условиях нагрева электронного газа все кинетические коэффициенты должны зависеть от напряженности электрического поля. Вместе с тем следует помнить, что определение (1.3) — не единственно возможное. Так, можно было бы определить электронную температуру с помощью соотношения Эйнштейна (VI.2.8), полагая

$$\frac{D}{\mu} = \frac{kT_e'}{e}. \tag{1.5}$$

Это определение «температуры» T_e' ничем не хуже и не лучше (1.3); однако значения электронной температуры, определяемые равенствами (1.3) и (1.5), вообще говоря, не совпадают *). Можно было бы ввести и другие определения электронной температуры в соответствии с тем или иным способом ее измерения. Причина такой неоднозначности состоит в том, что, как уже говорилось, понятие температуры имеет однозначный смысл лишь в условиях термоди-

*) Исключение составляет случай, рассматриваемый в § 4.

намического равновесия, когда единую температуру можно ввести обычным статистическим путем — через каноническое распределение Гиббса.

Нагрев электронного газа приводит к ряду следствий, наблюдаемых на опыте и интересных в техническом отношении. Прежде

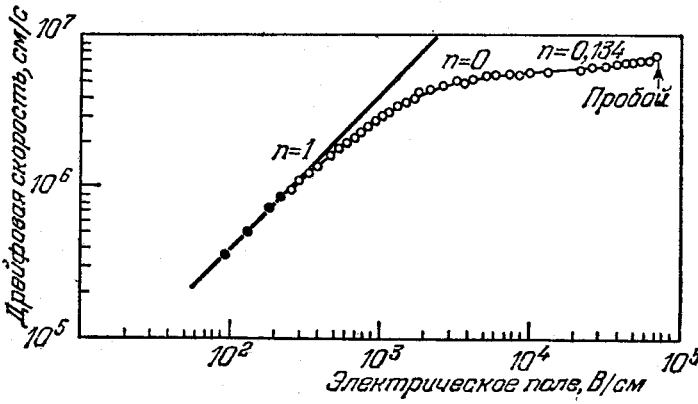


Рис. 16.1. Зависимость дрейфовой скорости от напряженности электрического поля в n-Ge при $T = 300$ К. Число n характеризует зависимость $v_d \sim E^n$.

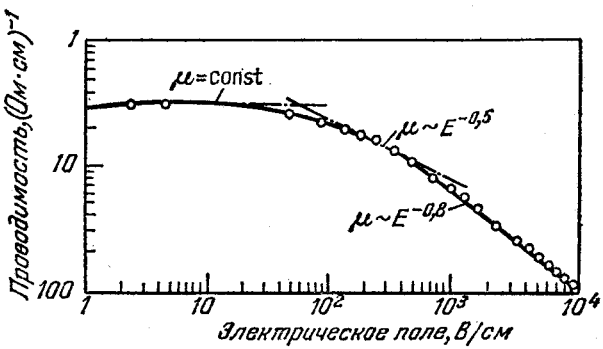


Рис. 16.2. Зависимость проводимости от напряженности электрического поля в p-Ge при $T = 77$ К.

всего, согласно сказанному выше, в условиях нагрева нарушается закон Ома: подвижность и электропроводность начинают зависеть от напряженности поля, а дрейфовая скорость становится нелинейной ее функцией. Примеры такой зависимости приведены на рис. 16.1, 16.2 *); более подробно этот вопрос рассматривается в § 7.

*) По данным работ: *J. B. Gunn,* Journ. Electr. **2**, 87 (1956); *K. S. Mendelson, R. Bray,* Proc. Phys. Soc. **B70**, 899 (1957).

Нагрев электронного газа электрическим полем может привести к зависимости подвижности и электропроводности от направления тока: даже в кубических кристаллах электропроводность оказывается, вообще говоря, тензором второго ранга, а не скаляром.

Физические причины, вызывающие анизотропию электропроводности в сильном поле, могут быть связаны с видом закона дисперсии носителей заряда. Картина особенно ясна в многодолинных материалах типа n-Ge или n-Si. Ускорение электронов в каждой отдельной долине определяется эффективной массой, соответствующей направлению поля (рис. 16.3). Поэтому вклад в электропроводность, обусловленный электронами только данной долины, оказывается анизотропным. Полный ток получается в результате суммирования по всем долинам. В условиях равновесия или в слабом поле все долины энергетически эквивалентны и концентрации электронов в них одинаковы. Соответственно полная электропроводность оказывается одной и той же при всех направлениях поля. С увеличением поля положение меняется. Дело в том, что времена τ_p и τ_ε зависят от эффективных масс. Поэтому в разных долинах электроны будут нагреваться по-разному. Различие электронных температур влечет за собой и различие подвижностей и концентраций электронов в разных долинах. При этом суммирование по всем долинам уже не восстанавливает изотропию электропроводности: она становится зависящей от направления поля. Отсюда следует, что векторы $\mathbf{j}$ и $\mathcal{E}$ оказываются, вообще говоря, не параллельными друг другу: угол между ними зависит от ориентации вектора $\mathbf{j}$ относительно осей кристалла.

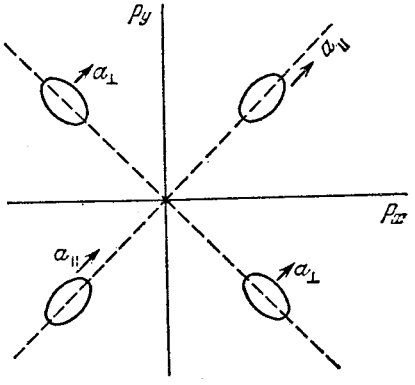


Рис. 16.3. Ускорения электронов $a_\parallel = e\mathcal{E}/m_\parallel$ и $a_\perp = e\mathcal{E}/m_\perp$, принадлежащих различным эллипсоидам энергии, в электрическом поле (n-Ge).

Увеличение средней энергии электронов по сравнению с термодинамически равновесным ее значением приводит к возрастанию тока термоэлектронной эмиссии из полупроводника (при неизменной температуре решетки). В самом деле, здесь существенна не температура решетки сама по себе, а именно энергия электронов. В катодах обычного типа ее повышают, нагревая решетку. Можно, однако, «греть» электроны электрическим полем, уменьшая тем самым бесполезные потери энергии на «раскачивание» атомов решетки. Такие холодные катоды существуют; широкому их применению пока препятствует трудность значительного нагрева электрон-

ного газа при комнатных температурах. Действительно, при заданном поле разность $(T_e - T)$ тем больше, чем больше время τ_ε, а оно становится особенно большим лишь при достаточно низких температурах T.

§ 2. Симметричная и антисимметричная части функции распределения

Функция распределения носителей заряда в слабом электрическом поле была исследована в гл. XIII. Легко убедиться, однако, что, пользуясь выражениями (XIII.5.1) и (XIII.6.7), мы получили бы для средней энергии обычное термодинамически равновесное значение. В самом деле, по определению

$$n\langle E\rangle = \frac{2}{(2\pi\hbar)^3}\int E(\mathbf{p})\,f(\mathbf{p})\,d\mathbf{p}. \tag{2.1}$$

Подставляя (XIII.5.1) в (2.1), видим, что слагаемое f_1, будучи нечетной функцией квазиимпульса, не дает вклада в интеграл, чем и доказывается высказанное утверждение. Этого и следовало ожидать: равенства (XIII.5.1) и (XIII.6.7) получены из кинетического уравнения лишь для достаточно слабого поля, когда квадратом и более высокими степенями его напряженности можно пренебречь.

При учете нагрева функцию распределения $f(\mathbf{p})$ также можно записать в виде суммы слагаемых, четного и нечетного относительно изменения знака $\mathbf{p}$:

$$f(\mathbf{p}) = \frac{f(\mathbf{p})+f(-\mathbf{p})}{2} + \frac{f(\mathbf{p})-f(-\mathbf{p})}{2}. \tag{2.2}$$

Первое слагаемое в правой части (2.2) называется *симметричной частью* функции распределения (f_s), второе — *антисимметричной* (f_a):

$$f(\mathbf{p},\ \mathcal{E}) = f_s + f_a, \tag{2.3}$$

причем

$$f_s(-\mathbf{p}) = f_s(\mathbf{p}), \qquad f_a(-\mathbf{p}) = -f_a(\mathbf{p}).$$

В отсутствие нагрева электронного газа функция f_s превращается в равновесную функцию распределения $f_0(E)$, а f_a — в малую добавку f_1.

Представление функции распределения в виде (2.3) имеет ясный физический смысл. Именно, антисимметричная часть $f_a(\mathbf{p}, \mathcal{E})$ описывает возникновение потоков заряда, энергии и т. д., а симметричная часть $f_s(\mathbf{p}, \mathcal{E})$ связана с функцией распределения носителей заряда по энергии $f(E)$. Так, для средней энергии на один электрон мы имеем

$$\langle E\rangle = \frac{2}{(2\pi\hbar)^3 n}\int E(\mathbf{p})\,f_s(\mathbf{p},\ \mathcal{E})\,d\mathbf{p}. \tag{2.4}$$

Далее, обозначим через $\mathscr{P}_+(E,\ \Delta E)$ и $\mathscr{P}_-(E,\ \Delta E)$ отнесенные к единице времени вероятности того, что электрон с энергией E, соответственно, получит и потеряет в акте рассеяния энергию ΔE. Тогда

$$\langle \dot{E} \rangle = \frac{2}{(2\pi\hbar)^3 n} \int f_s(\mathbf{p},\ \mathfrak{E}) \left[\int d(\Delta E) \left(\mathscr{P}_-(E,\ \Delta E) - \mathscr{P}_+(E,\ \Delta E) \right) \right] d\mathbf{p}. \tag{2.5}$$

Таким путем, зная функцию распределения, можно вычислить время релаксации τ_ε. Соответствующие расчеты можно найти в книгах [1, 2].

Подставляя выражение (2.3) в кинетическое уравнение (XIII.3.12) и разделяя в нем слагаемые, четные и нечетные по $\mathbf{p}$, мы получим систему двух интегро-дифференциальных уравнений для определения двух функций f_s и f_a. Замечая, что производная от четной (нечетной) функции есть нечетная (четная) функция своего аргумента, имеем (в отсутствие магнитного поля, при $\mathbf{F} = -e\mathfrak{E}$)

$$\frac{\partial f_s}{\partial t} + (\mathbf{v},\ \nabla f_a) + (\mathbf{F},\ \nabla_p f_a) = J_s[f], \tag{2.6}$$

$$\frac{\partial f_a}{\partial t} + (\mathbf{v},\ \nabla f_s) + (\mathbf{F},\ \nabla_p f_s) = J_a[f], \tag{2.7}$$

где J_s и J_a — четная и нечетная (относительно изменения $\mathbf{p}$) части интеграла столкновений.

Пусть рассеяние носителей заряда происходит только на несовершенствах решетки (включая фононы). Тогда, поскольку электронный газ не вырожден, интеграл столкновений дается выражением (XIII.3.10), в котором разность $1 - f(\mathbf{p})$ можно заменить единицей. Изменим знак аргумента $\mathbf{p}$ в (XIII.3.10) и одновременно произведем замену переменной интегрирования $\mathbf{p}' = -\mathbf{p}''$. При этом коэффициенты $\mathscr{P}_1$ и $\mathscr{P}_2$, будучи функциями только аргументов (XIII.6.4), останутся неизменными. Следовательно, интеграл $J[f]$ изменит или сохранит свой знак в зависимости от того, содержит ли он функцию f_s или f_a:

$$J_s[f] = J[f_s], \qquad J_a[f] = J[f_a]. \tag{2.8}$$

Такая простая форма связи между J_s, J_a и f_s, f_a обусловлена тем, что в рассматриваемом случае функция распределения входит под знак интеграла столкновений линейно. Если существенно фермиевское вырождение газа носителей заряда или рассеяние последних друг на друге, то соотношения между J_s, J_a и f_s, f_a оказываются более сложными.

Теоретическое исследование нагрева электронного газа в полупроводниках сводится, таким образом, к решению уравнений (2.6) и (2.7) при различных механизмах рассеяния носителей заряда.

§ 3. Уравнения баланса

Исследование системы уравнений (2.6) и (2.7) составляет весьма сложную математическую задачу. Лишь в последние годы были развиты численные методы ее решения с помощью современных электронно-вычислительных машин. По этой причине часто оказываются полезными более простые соотношения, выражающие законы сохранения энергии и квазиимпульса. Они представляют собой не что иное, как уравнения баланса для названных величин и могут быть строго выведены из кинетического уравнения [1, 2, 4]. Особенно простой вид эти соотношения принимают, когда газ носителей заряда пространственно однороден и состояние его стационарно. Действительно, при этом полная энергия и полный квазиимпульс носителей заряда не должны меняться со временем, а потоки, связанные с пространственными градиентами каких-либо величин, отсутствуют. Следовательно, условие постоянства полной энергии носителей имеет вид (1.2). По определению дрейфовой скорости $\mathbf{v}_d$ (гл. I) плотность тока можно записать в виде

$$\mathbf{j} = en\mathbf{v}_d.$$

Соответственно вместо (1.2) мы получим

$$e(\mathbf{v}_d, \mathcal{E}) = \frac{\frac{2}{3}\langle E\rangle - kT}{\tau_e}. \tag{3.1}$$

Как и в предыдущих главах, дрейфовую скорость можно записать в виде

$$\mathbf{v}_d = \frac{e}{m_\sigma}\tau_p\mathcal{E}, \tag{3.2}$$

где m_σ — эффективная масса электропроводности (XIII.7.15). Необходимо лишь иметь в виду, что среднее время релаксации импульса τ_p теперь следует вычислять, принимая во внимание нагрев электронного газа, т. е. с помощью уравнений (2.6) и (2.7), в которых функция f_s отлична от f_0. Так, в рамках изотропной модели τ_p теперь дается выражением (XIII.7.16) с заменой f_0 на f_s.

Соотношения (3.1) и (3.2) называются *уравнениями баланса*. Следует помнить, что входящие в них времена τ_e и τ_p описывают обмен энергией и квазиимпульсом между носителями заряда, с одной стороны, и кристаллической решеткой — с другой. При учете только межэлектронных столкновений без процессов переброса эти времена обращаются в бесконечность: внутренние силы приводят только к перераспределению энергии и квазиимпульса между носителями заряда, но не изменяют полную их энергию и полный квазиимпульс. Вместе с тем, влияя на вид функции распределения, взаимодействие между носителями заряда косвенно влияет и на значения τ_e и τ_p, равно как и на форму их зависимости от температуры, напряженности поля и т. д.

§ 4. Электронная температура

В зависимости от роли столкновений между носителями заряда по сравнению с другими механизмами рассеяния выделяются различные предельные случаи. Рассмотрим сначала материал с носителями заряда одного типа. Обозначим через τ_{ee} среднее время между столкновениями носителей заряда друг с другом. В силу близости эффективных масс носителей заряда при таких столкновениях происходит заметный обмен как квазиимпульсом, так и энергией. Следовательно, время τ_{ee} характеризует скорость обоих этих процессов. Могут иметь место три предельных соотношения между временами τ_{ee}, τ_e и τ_p:

$$\tau_{ee} \ll \tau_p \ll \tau_e, \tag{4.1}$$

$$\tau_p \lesssim \tau_{ee} \ll \tau_e, \tag{4.2}$$

$$\tau_p \lesssim \tau_e \ll \tau_{ee}. \tag{4.3}$$

В случае (4.1) носители заряда успевают многократно обменяться энергией и импульсом до того, как сыграет свою роль рассеяние их на фононах или иных несовершенствах решетки. В случае (4.2) это относится только к обмену энергией. Наконец, в случае (4.3) рассеяние носителей заряда друг на друге вообще не играет роли и им можно пренебречь.

Рассмотрим ситуацию, характеризуемую неравенствами (4.1). Тогда газ носителей заряда можно рассматривать, в первом приближении, как самостоятельную термодинамическую подсистему, лишь слабо взаимодействующую с решеткой. Столкновения между носителями приводят к установлению равновесия по энергии и импульсу внутри данной подсистемы. Это означает, что в данном случае, во-первых, имеет однозначный смысл представление о температуре электронов T_e, отличной, вообще говоря, от температуры решетки T, и, во-вторых, систему носителей заряда можно характеризовать скоростью $\mathbf{v}_d$, с которой все они движутся как целое относительно решетки. В сущности, мы имеем здесь то же, что и при движении жидкости или газа в сосуде. Роль последнего играет неидеальная решетка, а роль «жидкости» — система носителей заряда. По этой причине рассматриваемое приближение часто называют гидродинамическим.

Функция распределения $f(\mathbf{p})$, зависящая от параметров T_e и $\mathbf{v}_d$, должна обращать в нуль интеграл столкновений (XIII.3.11). При $\mathbf{v}_d = 0$ это означало бы, что $f(\mathbf{p})$ есть функция Ферми, в которой надо лишь заменить температуру T на T_e. Так же обстоит дело и при $\mathbf{v}_d \neq 0$, если рассматривать функцию распределения в системе отсчета, в которой электронный газ как целое покоится. Переходя к системе отсчета, связанной с кристаллической решеткой (в которой электронный газ движется), мы должны заменить $\mathbf{v}$ на

$\mathbf{v} - \mathbf{v}_d$. Тогда

$$f(\mathbf{p}) = \left\{\exp \frac{E(\mathbf{v}-\mathbf{v}_d)-F}{kT} + 1\right\}^{-1}. \tag{4.4}$$

Здесь энергия носителя заряда E рассматривается как функция его скорости $\mathbf{v}(\mathbf{p})$.

В отсутствие фермиевского вырождения равенство (4.4) принимает вид

$$f(\mathbf{p}) = \exp \frac{F - E(\mathbf{v}-\mathbf{v}_d)}{kT}. \tag{4.5}$$

Выражение, фигурирующее в правой части (4.5), называется максвелловским распределением с дрейфом.

Пусть скорость дрейфа мала по сравнению с характерной «тепловой» скоростью носителей заряда $v_{T_e} = \sqrt{kT_e/m}$. Тогда правую часть (4.5) можно разложить в ряд по степеням v_{T_e}, и мы получаем выражение вида (2.3), причем теперь

$$f_s = \exp \frac{F - E(\mathbf{v})}{kT}, \tag{4.6}$$

$$f_a = \frac{(\mathbf{v}_d, \nabla_v E)}{kT_e} \exp \frac{F - E(\mathbf{v})}{kT_e}. \tag{4.7}$$

Величины F, T_e и $\mathbf{v}_d$ в формулах (4.4) — (4.7) остаются пока произвольными. Согласно сказанному в § 3, учет одних межэлектронных столкновений (без процессов переброса) в принципе не позволяет найти T_e и $\mathbf{v}_d$. Для определения стационарных значений дрейфовой скорости и электронной температуры в пространственно однородной системе удобно воспользоваться уравнениями баланса (3.1) и (3.2). Действительно, поскольку функция распределения в рассматриваемом случае известна, можно, задавшись теми или иными механизмами рассеяния энергии и импульса (гл. XIV), вычислить времена τ_e и τ_p. Они оказываются теперь функциями $\mathbf{v}_d$ и T_e, а уравнения баланса превращаются в систему уравнений относительно названных переменных:

$$e\mathcal{E} = \frac{m_\sigma \mathbf{v}_d}{\tau_p(\mathbf{v}_d, T_e)}, \tag{4.8}$$

$$e(\mathbf{v}_d, \mathcal{E}) = \frac{k \cdot (T_e - T)}{\tau_e(\mathbf{v}_d, T_e)}. \tag{4.9}$$

Обратимся теперь к ситуации, характеризуемой неравенствами (4.2). Здесь, как и в предыдущем случае, газ носителей заряда достигает равновесия по энергиям за время порядка τ_{ee}, т. е. задолго до того, как станет заметным обмен энергией между электронами и решеткой. Это означает, что в данных условиях представление об электронной температуре по-прежнему имеет однозначный смысл: по отношению к энергии газ носителей заряда можно

рассматривать как почти независимую термодинамическую подсистему. Симметричная часть функции распределения (в системе отсчета, в которой электронный газ как целое покоится) здесь имеет прежний вид (4.6). Однако выражение (4.7) для антисимметричной части в условиях (4.2) оказывается уже несправедливым, сколь бы малой ни была дрейфовая скорость. Действительно, по отношению к импульсу газ носителей заряда в условиях (4.2) отнюдь не образует независимой подсистемы. Рассеяние импульса здесь происходит в основном на несовершенствах решетки, в результате чего сам вид антисимметричной функции распределения оказывается зависящим от механизма рассеяния. Эту функцию здесь следует определять с помощью уравнения (2.7). При этом интеграл столкновений $J_a[f]$, определяющий скорость рассеяния импульса, можно вычислять, пренебрегая межэлектронными столкновениями; рассеяние на несовершенствах решетки здесь, как и в слабом поле, часто можно считать абсолютно упругим *). В частности, при рассмотрении стационарного состояния системы с изотропным законом дисперсии остается в силе выражение (XIII.6.11), в котором надо лишь заменить J на J_a, а f_1 — на f_a:

$$J_a[f] = -\frac{f_a}{\tau(E)}. \tag{4.10}$$

Согласно (4.10) и (2.7) антисимметричная часть функции распределения в постоянном электрическом поле имеет вид

$$f_a(\mathbf{p}, \mathcal{E}) = -\tau(E)\{(\mathbf{v}, \nabla f_s) - e(\mathcal{E}, \nabla_p f_s)\}. \tag{4.11}$$

Поскольку функция f_s уже известна, выражение (4.11) позволяет вычислить плотность тока по формуле (XIII.2.2). В пространственно однородной системе, когда первое слагаемое в (4.11) отсутствует, мы получаем

$$\mathbf{j} = en\mu(T_e, T)\mathcal{E}. \tag{4.12}$$

Условие баланса энергии дает теперь уравнение для определения электронной температуры:

$$e\mu(T_e, T)\mathcal{E}^2 = \frac{k(T_e - T)}{\tau_e(T_e, T)}. \tag{4.13}$$

Приближение, выражаемое формулами (4.6) и (4.11), иногда называют квазигидродинамическим.

При наличии нескольких типов носителей заряда (электронов и дырок, «легких» и «тяжелых» дырок и т. д.) времена релаксации энергии и импульса вводятся для каждого типа в отдельности. Следует различать также времена между столкновениями носителей одного и того же и разных типов. При этом, в зависимости от соот-

*) Исключение, как и в слабом поле, может составить случай рассеяния на оптических фононах (см. гл. XIV и § XVI.5).

ношения между их эффективными массами, столкновения между различными носителями заряда могут как сопровождаться, так и не сопровождаться заметным обменом энергией. Так, при резком различии эффективных масс этот тип рассеяния, по сути дела, не отличается от рассеяния на заряженной примеси. Неравенства типа (4.1) — (4.3) теперь надо писать для разных типов носителей заряда по отдельности, и число различных предельных случаев заметно возрастает. Отметим два из них.

а) Пусть эффективные массы носителей разного типа — одного порядка величины. Тогда взаимное рассеяние разных носителей связано с интенсивным обменом энергией и при достаточно малых временах между соответствующими столкновениями имеет однозначный смысл представление о единой температуре носителей заряда, отличной, вообще говоря, от температуры решетки.

б) Пусть эффективные массы носителей заряда разного типа резко различны, но времена между столкновениями носителей одного и того же типа достаточно малы. Тогда имеет однозначный смысл представление о температурах носителей разного типа (например, электронов и дырок), не совпадающих, вообще говоря, как друг с другом, так и с температурой решетки.

Рассмотрим условия, при которых выполняются соотношения (4.1) и (4.2). Прежде всего заметим, что неравенства (4.2) могут иметь место только при условии $\tau_p \ll \tau_e$, т. е. в случае слабо неупругого рассеяния. Как мы знаем, так обстоит дело, если рассеяние квазиимпульса происходит в основном на атомах примеси (при любом механизме рассеяния энергии) или если рассеяние и энергии, и квазиимпульса обусловлено взаимодействием носителей заряда с акустическими (в том числе и пьезоэлектрическими) колебаниями решетки. Рассеяние на оптических колебаниях оказывается слабо неупругим, лишь если средняя энергия носителя заряда значительно превышает энергию оптического фонона:

$$kT_e \gg \hbar\omega_0.$$

В указанных условиях (4.1) и (4.2) удовлетворяются, если время τ_{ee} достаточно мало́. Чтобы вычислить его (при параболическом законе дисперсии), надо решить задачу о движении двух заряженных частиц с массами m_1 и m_2. Потенциальная энергия их есть

$$U = \frac{e_1 e_2}{\varepsilon r} \exp(-r/r_0). \tag{4.14}$$

Здесь r — расстояние между частицами, e_1 и e_2 — их заряды, r_0 — радиус экранирования.

Как известно из механики, такая задача распадается на две: о свободном движении центра инерции всей системы и о движении одной «частицы» с приведенной массой $m_r = \frac{m_1 m_2}{m_1 + m_2}$ и радиус-

вектором $\mathbf{r}$ в поле $U(\mathbf{r})$. Выражение (4.14) формально не отличается от энергии взаимодействия носителя заряда с ионом примеси. По этой причине мы можем непосредственно воспользоваться результатами § XIV.5, заменяя в них m на m_r и T на T_e. Согласно (XIV.5.19′) для невырожденного газа мы имеем по порядку величины

$$\tau_{ee} \sim \frac{\varepsilon^2 m_r^{1/2}}{ne^4}(kT_e)^{3/2} \ln^{-1}\left(\frac{m_r kT_e}{\hbar^2} r_0^2\right). \tag{4.15}$$

Мы заменили здесь энергию E на $^3/_2 kT_e$ и опустили несущественные численные множители.

Как и следовало ожидать, в невырожденном газе время τ_{ee} обратно пропорционально концентрации носителей заряда n (с точностью до медленно меняющегося логарифмического множителя). Это означает, что неравенства (4.1) и (4.2) выполняются лишь при достаточно больших концентрациях, превышающих некоторые критические значения. Для оценки их надо вычислить отношения τ_e/τ_{ee} и τ_p/τ_{ee}. Приравнивая их единице, получим порядковую оценку критических концентраций n_e и n_p, выше которых рассеяние, соответственно, энергии и квазиимпульса за счет межэлектронных столкновений оказывается доминирующим. При этом вместо того, чтобы вычислять τ_e и τ_p с помощью формул (2.5), (2.7), удобно выразить эти времена через длину свободного пробега по импульсу l_p, определяемую равенством *)

$$l_p = v_{T_e}\tau_p, \tag{4.16}$$

где v_{T_e} — средняя скорость носителей заряда с температурой T_e. При этом

$$\mathbf{v}_d = \frac{e\mathcal{E}l_p}{mv_{T_e}}$$

и, согласно (4.13),

$$\tau_e = \frac{k(T_e - T)}{e^2\mathcal{E}^2 l_p} mv_{T_e}. \tag{4.17}$$

По порядку величины $T_e - T \sim T_e$, и мы получаем

$$\tau_e \simeq \frac{(kT_e)^{3/2}m^{1/2}}{(e\mathcal{E})^2 l_p}. \tag{4.17′}$$

Согласно (4.15) и (4.17′) для критической концентрации n_e находим

$$n_e \simeq \frac{\varepsilon^2\mathcal{E}^2 l_p}{e^2 \ln\left(\frac{8m_r T_e}{\hbar^2} r_0^2\right)}. \tag{4.18}$$

*) Определение (4.16) отличается от используемого в случае слабого поля заменой T на T_e в формуле для средней скорости.

Далее, на основании формул (4.16) и (4.15)

$$n_p \simeq \frac{(\varepsilon k T_e)^2}{e^4 l_p \ln\left(\frac{8 m_r k T_e}{\hbar^2} r_0^2\right)}. \tag{4.19}$$

Составляя отношение правых частей (4.18) и (4.19) и пользуясь оценкой (4.17′), мы получаем.

$$\frac{n_e}{n_p} \simeq \left(\frac{e^2 l_p}{k T_e}\right)^2 \simeq \frac{\tau_p}{\tau_e}. \tag{4.20}$$

Коль скоро рассеяние носит почти упругий характер, правая часть (4.20) значительно меньше единицы.

Таким образом, условия применимости гидродинамического приближения оказываются заметно более жесткими, чем квазигидродинамического. Кроме того, значения n и l_p не всегда независимы: действительно, если свободные носители заряда поставляются примесью, то концентрация их не превышает концентрации заряженной примеси. Это означает, что $\tau_{ee} \simeq \tau_p$, и, следовательно, формула (4.7) для антисимметричной части функции распределения не дает хорошего приближения. С другой стороны, гидродинамическое приближение может оказаться оправданным в условиях сильной инжекции или в применении к очень узкозонным материалам, когда достаточно велика собственная концентрация носителей заряда.

Квазигидродинамическим приближением иногда пользуются и в условиях, когда оно, строго говоря, неприменимо — концентрация носителей заряда невелика по сравнению с n_e. Такой подход имеет смысл, если мы интересуемся лишь оценкой физических величин, получающихся путем усреднения сравнительно плавных функций энергии. Действительно, рассмотрим среднее значение некоторой величины $A(E)$ и будем для простоты считать закон дисперсии изотропным. По определению

$$\langle A \rangle = \frac{2}{(2\pi\hbar)^3 n} \int f_s(E) A(E) d\mathbf{p} = \frac{2}{\pi^2 \hbar^3 n} \int_0^\infty p^2(E) A(E) f_s(E) \frac{dp}{dE} dE. \tag{4.21}$$

Допустим, что функция $A(E)$ не имеет нулей и изменяется значительно медленнее, нежели $f_s(E)$. Тогда с хорошей точностью можно вынести $A(E)$ за знак интеграла при том значении $E = E_m$, которое отвечает максимуму остальных сомножителей в подынтегральном выражении (очевидно, E_m порядка kT_e). Остающийся интеграл дает — независимо от вида $f_s(E)$ — концентрацию частиц n, и мы получаем

$$\langle A \rangle \simeq A(E)\big|_{E=E_m}. \tag{4.22}$$

Неточность связана здесь лишь с отождествлением $E_m = kT_e$. Этот прием иногда называют *методом электронной температуры*.

С другой стороны, если функция $A(E)$ достаточно сильно зависит от E, заметно изменяясь на интервале энергии порядка kT_e, то указанный выше прием может стать непригодным. Действительно, главный вклад в интеграл (4.21) здесь дает лишь сравнительно небольшая область энергий, отвечающая максимуму всего подынтегрального выражения. При этом существенным оказывается явный вид функции распределения $f_s(E)$. Если последняя заметно отличается от (4.6), то метод электронной температуры может оказаться неприменимым.

§ 5. Роль неупругости рассеяния

В условиях, определяемых неравенствами (4.3), носители заряда нельзя рассматривать как независимую подсистему: доминирующую роль играет рассеяние их на несовершенствах решетки. Вид неравновесной функции распределения $f(\mathbf{p})$ здесь заранее не известен: он устанавливается в результате совместного действия внешнего поля и тех же процессов рассеяния, которые определяют и средние значения энергии и импульса носителей заряда. Эти процессы описываются уравнениями (2.6) и (2.7), которые и надо решить для определения функций $f_s(\mathbf{p})$ и $f_a(\mathbf{p})$. При этом результат оказывается не универсальным: вид функции распределения зависит от природы образца и условий опыта. Существенную роль здесь играет степень неупругости рассеяния, определяемая равенством

$$\eta = \frac{\tau_p}{\tau_e}. \tag{5.1}$$

Действительно, рассмотрим уравнения баланса (3.1) и (3.2). Умножая равенство (3.2) скалярно на $\mathbf{v}_d$, получим

$$e(\mathbf{v}_d, \mathcal{E}) = \frac{m_0 v_d^2}{\tau_p}.$$

Сравнивая это соотношение с уравнением (3.1) и принимая во внимание (1.3′), находим

$$kT_e + \frac{2}{3}E_d - kT = \frac{m_0 v_d^2}{\eta}. \tag{5.2}$$

По определению при слабо неупругом рассеянии $\eta \ll 1$, а при сильно неупругом $\eta \simeq 1$. Рассмотрим эти два случая по отдельности.

При $\eta \ll 1$ из соотношения (5.2) вытекает неравенство

$$m_0 v_d^2 \ll k(T_e - T) + {}^2/_3 E_d. \tag{5.3}$$

Поскольку энергия дрейфа E_d обращается в нуль вместе с $\mathbf{v}_d$ (например, при квадратичном законе дисперсии $E_d = {}^1/_2 m v_d^2$), неравенство (5.3) можно переписать в виде

$$m_0 v_d^2 \ll kT_e. \tag{5.3′}$$

Таким образом, при слабо неупругом рассеянии энергия дрейфа мала по сравнению со средней энергией хаотического движения носителей заряда. Иначе говоря,

$$v_d \ll v_{T_e} = \left(\frac{kT_e}{m}\right)^{1/2}. \tag{5.4}$$

Это позволяет думать, что антисимметричная часть функции распределения в данном случае мала по сравнению с симметричной ее частью *). Мы имеем здесь ту же ситуацию, что и в слабом поле, — с той лишь разницей, что вместо равновесной функции распределения f_0 появляется симметричная часть неравновесной f_s. В частности, при исследовании стационарного состояния системы с изотропным законом дисперсии оказываются справедливыми соотношения (4.10) и (4.11). Подстановка выражения (4.11) в уравнение (2.6) (при $\partial f_s/\partial t = \partial f_a/\partial t = 0$) приводит к интегро-дифференциальному уравнению для симметричной части функции распределения.

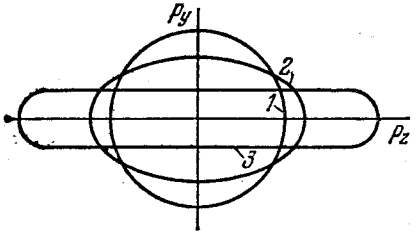


Рис. 16.4. Контуры постоянной функции распределения в пространстве квазиимпульсов. Ось Z направлена вдоль вектора напряженности электрического поля; оси Y и X (последняя не показана) эквивалентны. Кривая *1* изображает контур при $\mathcal{E} = 0$; кривые *2* и *3* отвечают слабо и сильно неупругому рассеянию ($\eta \ll 1$ и $\eta \simeq 1$).

Приближение, основанное на неравенствах (5.3), (5.3′), называется диффузионным. Оно было введено и обосновано Б. И. Давыдовым. Смысл названия состоит в том, что энергия, получаемая носителями заряда от поля, почти равномерно распределяется по всем степеням свободы, функция распределения почти симметрична и вся система электронов лишь очень медленно смещается в направлении действующей силы. Так обстоит дело при рассеянии импульса на ионах заряженной примеси или иных структурных дефектах решетки, на акустических (в том числе и пьезоэлектрических) фононах, а также на оптических фононах, если температура T_e достаточно велика: $kT_e \gg \hbar\omega_0$.

Случай сильно неупругого рассеяния реализуется, если доминирующую роль играет взаимодействие носителей заряда с оптическими фононами, причем $kT \ll \hbar\omega_0$, $kT_e \ll \hbar\omega_0$. При этом из

*) Это можно строго доказать, пользуясь уравнениями (2.6) и (2.7). Высказанное утверждение оказывается справедливым даже для носителей заряда с существенно анизотропной формой изоэнергетических поверхностей, если только вероятность рассеяния на угол θ не слишком сильно зависит от θ [1].

равенства (5.2) вытекает соотношение

$$k(T_e - T) \simeq E_d. \tag{5.5}$$

Энергия дрейфа здесь оказывается сравнимой со средней энергией хаотического движения носителей, а скорость дрейфа — с величиной v_{T_e}. Поэтому при сильно неупругом рассеянии следует ожидать резко анизотропной — вытянутой в направлении силы $\mathbf{F}$ — функции распределения (рис. 16.4). Антисимметричная часть ее здесь отнюдь не мала, и равенства (4.10) и (4.11) неверны. При определенных условиях оказывается справедливым приближение максимальной анизотропии, когда неравновесная функция распределения аппроксимируется выражением

$$f(\mathbf{p}) = \varphi(E)\,\delta(1 - \cos\theta). \tag{5.6}$$

Здесь φ — некоторая функция энергии носителя заряда, подлежащая определению из кинетического уравнения, θ — угол между векторами $\mathbf{p}$ и $\mathbf{F}$. Выражение (5.6) описывает тонкую «иглу», направленную вдоль электрического поля; по этой причине рассматриваемое распределение называется также иглообразным.

§ 6. Зависимость подвижности и концентрации носителей заряда от напряженности поля

Нагрев электронного газа приводит к зависимости подвижности носителей заряда и концентрации их от напряженности электрического поля в образце. Полевая зависимость подвижности может быть обусловлена двумя причинами.

Во-первых, электроны разной энергии рассеиваются по-разному: за редкими исключениями время свободного пробега зависит от энергии носителя заряда. Вызываемое электрическим полем перераспределение носителей по энергиям приводит к тому, что среднее время свободного пробега τ_p, а потому и подвижность μ, оказывается функцией электронной температуры. Вид зависимости $\mu(T_e)$ легко установить, если, как это часто бывает, газ фононов остается в равновесии при температуре T. Тогда мы можем воспользоваться формулами (4.21) и (4.22) при $A(E) = \tau(E)$ *). Пользуясь выражением (XIII.7.20′) и заменяя в нем E на kT_e, получим

$$\mu \sim [T_e(\mathcal{E})]^r. \tag{6.1}$$

Саму функцию $T_e(\mathcal{E})$ можно найти из уравнений баланса. Такой механизм полевой зависимости подвижности называют перегревным: он связан с отклонением электронной температуры от температуры решетки.

*) Речь идет здесь не о точном вычислении подвижности как функции поля, а лишь об ее оценке и об установлении вида зависимости $\mu(\mathcal{E})$. Для точного расчета по формуле (XIII.2.2) следует решить кинетическое уравнение.

Во-вторых, повышение электронной температуры может привести к переходу заметной части носителей заряда в состояния со сравнительно малой подвижностью. Так обстоит дело, если в зоне имеется несколько систем неэквивалентных долин, причем нижним отвечают меньшие значения эффективных масс, чем верхним.

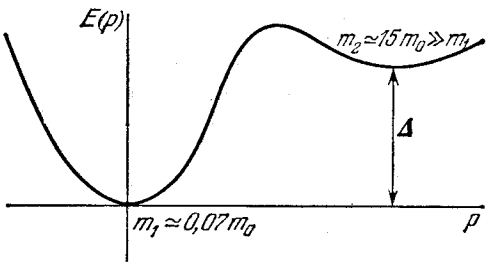


Рис. 16.5. Дисперсионная кривая для электронов проводимости в арсениде галлия в направлении [100] в зоне Бриллюэна (схематически).

Рассмотрим это на примере арсенида галлия. Дисперсионная кривая для электронов проводимости в этом материале изображена на рис. 16.5. Основной минимум расположен в центре зоны Бриллюэна, шесть взаимно эквивалентных побочных минимумов (из которых на рисунке показан один)—в направлении [100] и ему эквивалентных. Энергетическое расстояние между основным и побочными минимумами Δ составляет около 0,34 эВ. Эта величина столь значительна, что при комнатной температуре электронов вблизи побочных минимумов почти нет. При достаточно сильном нагреве электронного газа электрическим полем эти области энергетической зоны начинают заполняться за счет обеднения нижней долины. Наблюдаемую на опыте подвижность электронов μ можно представить в виде

$$\mu = \mu_1 \frac{n_1}{n} + \mu_2 \frac{n_2}{n}. \qquad (6.2)$$

Здесь μ_1 и μ_2, n_1 и n_2 — подвижности и концентрации носителей заряда в нижней и верхней долинах. Заметим, что, приписывая носителям заряда в каждой долине свою подвижность, мы делаем предположение о медленности обмена квазиимпульсами между электронами разных долин по сравнению с электронами одной долины. К этому есть известные основания. Действительно, переход электрона из одной долины в другую связан с большим изменением квазиимпульса, что неизбежно влечет за собой и сравнительно большую энергию испускаемых или поглощаемых фононов. При не слишком высокой электронной температуре этот процесс может оказаться существенно неупругим, что понижает его вероятность. С повышением электронной температуры положение меняется. По этой причине оценку с помощью формулы (6.2) можно рассматривать лишь как ориентировочную.

В слабом поле второе слагаемое в правой части (6.2) близко к нулю; в достаточно сильных полях (порядка 10^3 В/см) оно становится заметным, что и приводит к падению наблюдаемой подвижности.

Аналогичную структуру зоны проводимости имеют и другие материалы: InP, CdTe, ZnSe, GaAsP, $Ga_xIn_{1-x}Sb$, InAs, InSb.

Рассмотренный механизм полевой зависимости подвижности называют механизмом Ридли—Уоткинса—Хилсума. Обычно он становится существенным лишь при достаточно сильных полях. Этот механизм не связан непременно с наличием побочных минимумов в зоне проводимости. Аналогичные эффекты наблюдались и в германии — как электронном, так и дырочном, — подвергнутом одноосному сжатию. Последнее снимало вырождение между подзонами валентной зоны и между различными долинами зоны проводимости; в результате становились возможными индуцированные полем переходы носителей заряда в состоянии с большей эффективной массой.

Явный вид функции $\mu(\mathcal{E})$ зависит от механизма рассеяния. В не слишком сильных полях, однако, можно ограничиться первыми двумя членами разложения подвижности в ряд по степеням $\mathcal{E}^2$:

$$\mu = \mu_0 (1 + \beta \mathcal{E}^2). \tag{6.3}$$

Здесь μ_0 — подвижность в исчезающе слабом поле, β — коэффициент, не зависящий от $\mathcal{E}$ и определяющийся параметрами и температурой образца. При полях, допускающих аппроксимацию (6.3), обычно существен только перегревный механизм. По этой причине знак β определяется знаком показателя степени r в выражении (6.1). Обращаясь к таблице 14.2, видим, что $\beta > 0$, если рассеяние квазиимпульса обусловлено взаимодействием носителей заряда с заряженной примесью, пьезоэлектрическими фононами или поляризационными фононами при температуре решетки выше дебаевской; при рассеянии на акустических фононах $\beta < 0$.

Значения напряженности поля, при которых справедлива аппроксимация (6.3), определяются параметрами и температурой материала. При этом иногда говорят о «теплых» электронах. С дальнейшим повышением напряженности поля вид функции $\mu(\mathcal{E})$ усложняется. В достаточно сильном поле может иметь место насыщение тока, когда дрейфовая скорость перестает зависеть от напряженности поля, и, следовательно, подвижность обратно пропорциональна $\mathcal{E}$. Этот эффект обусловлен сильной неупругостью рассеяния при достаточно сильной связи носителей заряда с оптическими фононами. Действительно, пусть температура решетки $T \ll \hbar\omega_0/k$, а напряженность поля такова, что дрейфовая скорость, ограничиваемая какими-либо другими механизмами рассеяния, становится сравнимой с $v_0 = (2\hbar\omega_0/m)^{1/2}$. Достигнув энергии $\hbar\omega_0$, электрон очень быстро (с характерным временем τ) испустит оптический фонон, резко изменив при этом свою энергию. Поглощением оптических фононов в рассматриваемой нами температурной области можно пренебречь ввиду малого их числа. Таким образом, для среднего значения

величины $\dot{E}$ (см. стр. 513) мы получим

$$\langle \dot{E} \rangle \simeq \hbar\omega_0 \langle \tau^{-1} \rangle.$$

Здесь $\tau(E)$ дается одной из формул (XIV.4.22) или (XIV.4.23). При этом уравнения баланса энергии и квазиимпульса (3.1) и (3.2) принимают вид

$$e(\mathbf{v}_d, \mathcal{E}) \simeq \hbar\omega_0 \langle \tau^{-1} \rangle \tag{6.4}$$

и

$$e\mathcal{E} = m_\sigma \mathbf{v}_d \langle \tau^{-1} \rangle. \tag{6.5}$$

Заметим, что в уравнения (6.4) и (6.5) входит лишь одно характерное время $\langle \tau^{-1} \rangle^{-1}$, поскольку при сильно неупругом рассеянии средние времена релаксации энергии и импульса совпадают. Пользуясь уравнениями (6.4) и (6.5), легко находим

$$\langle \tau^{-1} \rangle^{-1} = \frac{\sqrt{m_\sigma \hbar\omega_0}}{e\mathcal{E}}, \qquad v_d = \sqrt{\frac{\hbar\omega_0}{m_\sigma}}. \tag{6.6}$$

Не следует удивляться тому, что время $\langle \tau^{-1} \rangle^{-1}$ не содержит константы связи носителей заряда с оптическими фононами: это — усредненная величина, для вычисления которой надо было бы решить кинетическое уравнение с учетом всех существенных механизмов рассеяния, а затем усреднить величину $\tau^{-1}(E)$ с помощью соответствующей функции распределения. При этом константа связи с оптическими фононами выпадает именно потому, что она считается достаточно большой. Использование уравнений баланса позволяет избежать этой процедуры.

Видим, что дрейфовая скорость действительно не зависит от $\mathcal{E}$, а подвижность, пропорциональная в данном случае $\langle \tau^{-1} \rangle^{-1}$, обратно пропорциональна $\mathcal{E}$.

Насыщение дрейфовой скорости в достаточно сильном электрическом поле наблюдалось, в частности, в n-Ge (рис. 16.1).

Полевая зависимость концентрации носителей заряда может быть связана со спецификой рекомбинации горячих электронов. Действительно, в условиях равновесия концентрация свободных носителей заряда определяется только положением уровня Ферми и температурой. Это обусловлено тем, что вероятности захвата электронов центрами рекомбинации и вероятности обратного выброса связаны друг с другом принципом детального равновесия (IX.4.6). В результате в указанных условиях остается независимым лишь один из двух коэффициентов, описывающих эти вероятности. Он влияет на кинетику рекомбинации, но не на равновесную концентрацию свободных электронов. При отклонении системы от термодинамического равновесия положение меняется: соотношение (IX.4.6) здесь уже не имеет места, и концентрация свободных носителей заряда определяется не только температурой решетки, но и соотношением между вероятностями захвата и обратного выброса. По-

скольку сечения захвата зависят от энергии захватываемого электрона, коэффициент α_n становится функцией напряженности электрического поля. Соответственно появляется и зависимость времени жизни и концентрации свободных носителей заряда от напряженности поля. Легко понять, в какую сторону должны изменяться названные величины при ее возрастании. Здесь выделяются два случая, отвечающие центрам рекомбинации, заряженным разноименно и одноименно с носителями заряда.

В первом случае, при наличии сил кулоновского притяжения, особенно вероятен захват сравнительно медленных электронов, проводящих много времени вблизи рекомбинационного центра. С другой стороны, в отсутствие вырождения электронного газа вероятность обратного выброса захваченных электронов в зону почти не зависит от напряженности поля (пока последняя не слишком велика) (см. ниже). Таким образом, нагрев электронного газа должен приводить к сдвигу равновесия между процессами захвата и обратного выброса в пользу последних. Иначе говоря, в рассматриваемых условиях время жизни и концентрация свободных носителей заряда должны увеличиваться с ростом напряженности электрического поля. Такие эффекты действительно наблюдались при захвате электронов положительными ионами элементов V группы в кремнии.

Выше мы предполагали, что полевая зависимость коэффициента захвата обусловлена только изменением функции распределения свободных носителей заряда. Однако есть и другой фактор — непосредственное влияние внешнего электрического поля на энергетический спектр связанных электронов. Он оказывается существенным, когда напряженность внешнего поля становится сравнимой с напряженностью поля ловушки. Тогда начинается автоионизация последней и число свободных носителей заряда быстро возрастает.

Во втором случае, при наличии сил кулоновского отталкивания, захватываемый электрон должен преодолевать потенциальный барьер (рис. 16.6) или проникать сквозь него туннельным путем. Фактически в n-Ge, легированном медью или золотом, туннельный эффект доминирует (см. § XVII.9). При этом вероятность захвата электрона быстро возрастает с увеличением его энергии. В результате при нагреве электронного газа равновесие между процессами захвата и обратного выброса сдвигается в пользу первых. Это означает, что время жизни и концентрация свободных носителей заряда должны уменьшаться с увеличением напряженности электрического поля. В сущности, это есть механизм Ридли — Уоткинса — Хилсума: нагрев электронов способствует переходу их в состояние с практически нулевой подвижностью.

Рассматриваемый эффект хорошо наблюдается в кристаллах электронного германия, содержащих в качестве центров рекомбинации отрицательно заряженные ионы золота, меди и других акцеп-

торных атомов. На рис. 16.7 показана полевая зависимость времени жизни электронов в n-Ge с медью, частично компенсированной мелкими донорами (сурьмой) *). Концентрации меди N_{Cu} и компенсирующих доноров N_d удовлетворяли условию $2N_{Cu} < N_d < 3N_{Cu}$. Так как медь в германии создает три акцепторных уровня (см. рис. 2.20), то при пониженных температурах (при которых тепловой

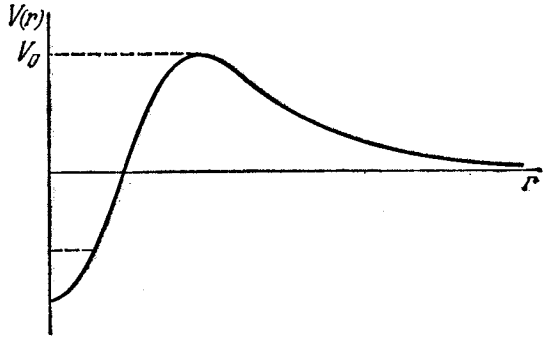


Рис. 16.6. Потенциальная энергия электрона в поле отрицательного иона примеси. На больших расстояниях существует только кулоновское отталкивание. На расстояниях порядка постоянной решетки начинают сказываться короткодействующие силы притяжения, связанные как с формой распределения заряда в самом ионе, так и с влиянием окружающих атомов решетки. Пунктиром показан энергетический уровень, создаваемый потенциальной ямой.

выброс электронов с глубоких уровней меди становится пренебрежимо малым, но мелкие доноры остаются еще ионизованными полностью) все электроны доноров оказываются локализованными на уровнях меди, которая поэтому существует в виде ионов Cu^{2-} и Cu^{3-}. Генерация электронов в зону производилась светом с такой энергией фотонов, чтобы электроны возбуждались только с верхнего уровня меди (с ионов Cu^{3-}). Соответственно захват электронов из зоны происходит на ионы Cu^{2-}. Из рисунка видно, что время жизни τ сильно уменьшается при разогреве электронного газа. Зависимость τ от напряженности электрического поля тем сильнее, чем ниже температура решетки. На рис. 16.8 показана зависимость вероятности захвата электронов на ионы Cu^{2-} (коэффициента захвата α_n^{2-}, ср. § IX.4) от напряженности электрического поля *). Как и следовало ожидать, α_n^{2-} увеличивается при разогреве электронного газа.

*) По данным работы: *В. Г. Алексеева, Н. Г. Жданова, М. С. Каган, С. Г. Калашников, Е. Г, Ландсберг,* ФТП 6, 316 (1972).

На рис. 16.9 *) приведена зависимость времени жизни электронов от напряженности электрического поля в другом хорошо изученном случае — при рекомбинации горячих электронов на ионах золота в n-Ge. Для образца, к которому относится рисунок, также выполнялось соотношение $2N_{Au} < N_d < 3N_{Au}$. Вследствие того, что из четырех энергетических уровней золота в германии (см.

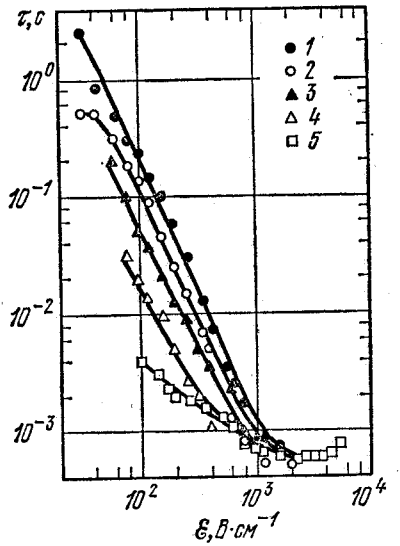


Рис. 16.7. Зависимость времени жизни горячих электронов в n-Ge от напряженности электрического поля при рекомбинации на ионах Cu^{2-}. Температура решетки: *1* — 25; *2* — 35; *3* — 40; *4* — 55; *5* — 90 К.

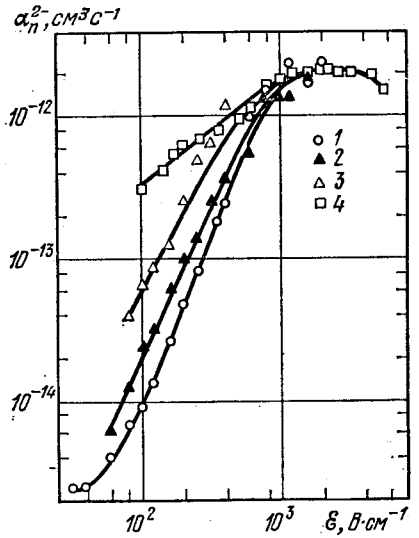


Рис. 16.8. Зависимость коэффициента захвата электронов α_n^{2-} от напряженности электрического поля для меди в германии. Температура решетки: *1* — 30; *2* — 40; *3* — 55; *4* — 90 К.

рис. 2.20) только три являются акцепторными (самый нижний уровень $E_1 - E_v = 0{,}05$ эВ — донорный), при указанном соотношении между концентрациями примесей и при низких температурах в германии существуют лишь ионы Au^{2-} и Au^{3-}. Возбуждение электронов в зону происходило с ионов Au^{3-}, а их захват из зоны — на ионы Au^{2-}.

Так как стационарная концентрация электронов в зоне есть $n = g_n\tau$, где g_n — темп генерации электронов, то уменьшение τ при разогреве электронов сопровождается и соответствующим уменьшением их концентрации.

При достаточно больших значениях напряженности поля концентрация свободных носителей заряда в любом полупроводнике

*) По данным работы: *И. А. Курова, М. Врана, П. Берндт,* ФТП **2**, 1838 (1968).

начинает быстро возрастать. Это обусловлено электрическим пробоем — генерацией свободных носителей заряда, вызванной электрическим полем. Один из механизмов ее, называемый ударной ионизацией, состоит в том, что уже имеющиеся свободные носители, ускоряясь полем, приобретают в нем энергию, достаточную для ионизации примеси (примесный пробой) или — в более сильных полях — для генерации пар электрон — дырка (междузонный пробой). В германии при комнатной температуре междузонный пробой наступает при $\mathscr{E} \simeq$ $\simeq 10^5$ В/см. Примесный пробой возможен лишь при достаточно низких температурах, когда примесные уровни заполнены; при $T \simeq 6{,}5$ К в германии, легированном галлием, это явление наступает при $\mathscr{E} \simeq 10 \div 15$ В/см. Для более глубоких ловушек необходимы и более сильные поля. Так, примесный пробой n-Ge, легированного нейтральными атомами цинка, при 8 К начинается при $\mathscr{E} \simeq 300$ В/см (соответствующая энергия ионизации составляет 0,03 эВ). Другой механизм электрического пробоя обусловлен туннелированием электронов из валентной зоны в зону проводимости в сильном электрическом поле (см. § IV.6).

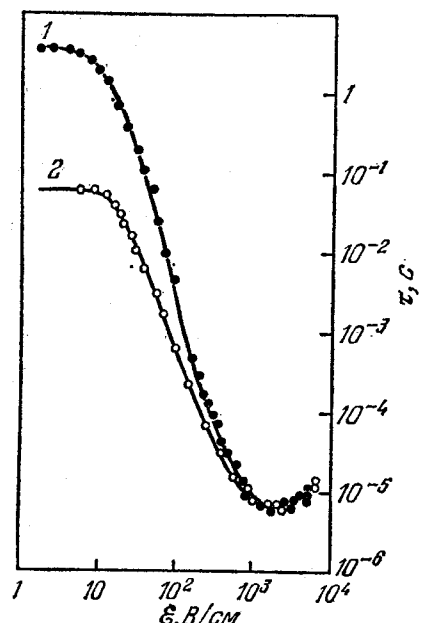


Рис. 16.9. Время жизни электронов в n-Ge с золотом как функция напряженности поля. Кривая *1* соответствует $T = 20$ К, кривая *2* — 30 К. В последнем случае играет роль тепловая генерация носителей заряда.

До сих пор мы рассматривали лишь пространственно однородную систему носителей заряда. В этих условиях электрическое поле в полупроводнике возникает лишь при наложении напряжения и оно неизбежно связано с протеканием тока; энергия, сообщаемая носителям заряда, берется из внешнего источника. При этом совершенно безразлично, рассматривать ли подвижность, время жизни и т. д. как функции напряженности поля или электронной температуры: последние две величины связаны друг с другом уравнениями баланса, и, хотя бы в принципе, можно найти явный вид зависимости $T_e(\mathscr{E})$. С равным успехом можно было бы использовать в качестве аргумента μ и любую другую функцию напряженности поля, например плотность тока. В пространственно неоднородной системе (например, при рассмотрении n—p-переходов) положение значительно усложняется. Действительно, в таких системах существуют внутренние электрические поля. При равновесии они, разумеется, не производят никакой работы над носителями заряда, ибо плотность тока равна нулю: в областях, где зоны искривлены, плотность тока состоит из суммы дрейфовой и диффузионной компонент, которые

взаимно компенсируются. При нарушении равновесия нагрев становится возможным. Тогда уравнения баланса приобретают более сложный вид, нежели (3.1) и (3.2): появляются составляющие плотности тока и потока энергии, связанные с градиентами концентрации электронов и электронной температуры. В результате уравнения баланса становятся дифференциальными. Напряженность поля, концентрация частиц и электронная температура зависят теперь от координат, а соотношения между ними оказываются, вообще говоря, нелокальными: любая из этих величин, взятая в какой-нибудь точке пространства, может зависеть от значений других величин не только в той же точке, но и в других (более подробное рассмотрение этого вопроса можно найти в книге [4]). Некоторые эффекты, связанные с нагревом электронов в таких системах, рассматриваются в книге [3].

§ 7. Дифференциальная проводимость

Плотность тока в пространственно однородной системе дается выражением

$$j_\alpha = \sigma_{\alpha\beta}(T_e)\,\mathcal{E}_\beta, \qquad \sigma_{\alpha\beta} = e\mu_{\alpha\beta}n. \tag{7.1}$$

Уравнение

$$\mathbf{j} = \mathbf{j}(\mathcal{E})$$

определяет вольтамперную характеристику рассматриваемого образца. В условиях применимости закона Ома она линейна. В результате нагрева электронного газа вольтамперная характеристика становится нелинейной.

Для описания таких вольтамперных характеристик удобно ввести представление о дифференциальной проводимости σ_d. В простейшем случае, когда σ есть скаляр, σ_d определяется равенством

$$\sigma_d(\mathcal{E}) = \frac{dj}{d\mathcal{E}} = \sigma + \mathcal{E}\frac{d\sigma}{d\mathcal{E}}. \tag{7.2}$$

Соответствующее тензорное выражение имеет вид

$$\sigma_{d,\,\alpha\beta} = \frac{dj_\alpha}{d\mathcal{E}_\beta} = \sigma_{\alpha\beta} + \mathcal{E}_\gamma\frac{d\sigma_{\alpha\gamma}}{d\mathcal{E}_\beta}. \tag{7.2'}$$

Равенства (7.2), (7.2′) написаны для постоянного тока; в переменном поле они сохраняются в применении к фурье-компонентам $\mathbf{j}(\omega)$ и $\mathcal{E}(\omega)$. Дифференциальная проводимость при этом зависит от частоты поля ω.

Для вычисления σ_d удобно воспользоваться уравнением баланса (3.1), переписав его в виде

$$\sigma\mathcal{E}^2 = nk\frac{T_e - T}{\tau_e}. \tag{7.3}$$

Введем для краткости обозначение

$$nk\frac{T_e - T}{\tau_e} = P. \tag{7.4}$$

Тогда

$$\sigma\mathcal{E}^2 = P. \tag{7.3'}$$

Величина P есть мощность, передаваемая решетке электронами в единице объема.

Рассматривая μ, σ, n и τ_e как функции электронной температуры, мы имеем (согласно (7.2))

$$\sigma_d = \sigma + \mathcal{E}\frac{d\sigma}{dT_e}\frac{dT_e}{d\mathcal{E}}. \qquad (7.2'')$$

Чтобы найти производную $dT_e/d\mathcal{E}$, продифференцируем по $\mathcal{E}$ равенство (7.3'):

$$2\sigma\mathcal{E} + \mathcal{E}^2\frac{d\sigma}{dT_e}\frac{dT_e}{d\mathcal{E}} = \frac{dP}{dT_e}\frac{dT_e}{d\mathcal{E}},$$

откуда

$$\frac{dT_e}{d\mathcal{E}} = \frac{2\sigma\mathcal{E}}{\dfrac{dP}{dT_e} - \mathcal{E}^2\dfrac{d\sigma}{dT_e}}.$$

Подставляя это выражение в (7.2'') и заменяя там $\mathcal{E}^2$ на P/σ, получим с учетом (7.1)

$$\sigma_d = \sigma\frac{\dfrac{dP}{dT_e} + \dfrac{P}{\sigma}\dfrac{d\sigma}{dT_e}}{\dfrac{dP}{dT_e} - \dfrac{P}{\sigma}\dfrac{d\sigma}{dT_e}}. \qquad (7.5)$$

Подчеркнем, что использование здесь представления об электронной температуре отнюдь не означает, что функция распределения непременно имеет больцмановский вид (4.6) (см. § 4).

В зависимости от того, как ведет себя проводимость σ при повышении напряженности поля, кривая $j(\mathcal{E})$ отклоняется вниз или вверх от прямой $j = \sigma_0\mathcal{E}$. В первом случае говорят о сублинейной вольтамперной характеристике ($d\sigma/d\mathcal{E} < 0$), во втором — о суперлинейной ($d\sigma/d\mathcal{E} > 0$).

При $d\sigma/d\mathcal{E} < 0$ проводимость на некотором участке характеристики может столь быстро убывать с ростом $\mathcal{E}$, что это перекроет возрастание тока, связанное с наличием обычного множителя $\mathcal{E}$ в формуле (7.1). Тогда плотность тока будет убывать с возрастанием напряженности поля: сублинейная вольтамперная характеристика превратится в характеристику с падающим участком, чему отвечают отрицательные значения дифференциальной проводимости. При дальнейшем возрастании напряженности поля причины, вызывающие столь быстрое уменьшение проводимости, обычно исчезают: происходит смена механизма рассеяния (в случае перегревного механизма), уравниваются концентрации электронов в нижней и верхних долинах (в случае механизма Ридли — Уоткинса — Хилсума) и т. д. Тогда плотность тока вновь начинает возрастать вместе с полем. В результате получается вольтамперная характеристика, схематически изображенная на рис. 16.10 (ее называют характеристикой N-типа). Как видно из рисунка, в образцах с вольтамперной характеристикой

N-типа, связь напряженности поля с плотностью тока не всегда однозначна: в области

$$j_v < j < j_p$$

каждому значению плотности тока отвечают три возможных значения напряженности поля: $\mathcal{E}_1$, $\mathcal{E}_2$ и $\mathcal{E}_3$. Два из них ($\mathcal{E}_1$ и $\mathcal{E}_3$) находятся на возрастающих ветвях характеристики, одно ($\mathcal{E}_2$) — на падающей.

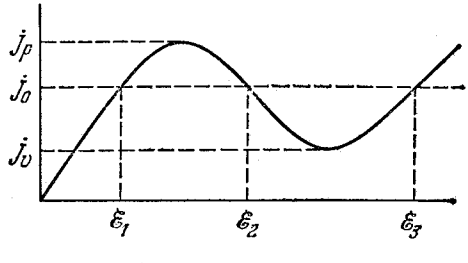


Рис. 16.10. Вольтамперная характеристика N-типа.

При $d\sigma/d\mathcal{E} > 0$ проводимость на некотором участке характеристики может столь быстро возрастать с ростом плотности тока, что напряжение на образце, а следовательно и напряженность поля, будет падать с ростом j. Суперлинейная вольтамперная характеристика превратится в характеристику с областью отрицательной дифференциальной проводимости. При дальнейшем возрастании плотности тока причины, вызывающие столь быстрый рост проводимости, обычно исчезают. Тогда напряженность поля вновь начинает возрастать вместе с плотностью тока. В результате получается вольтамперная характеристика S-типа (рис. 16.11).

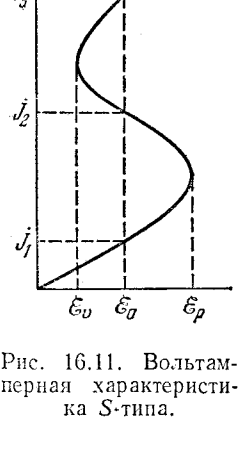


Рис. 16.11. Вольтамперная характеристика S-типа.

Как видно из рис. 16.11, в образцах с вольтамперной характеристикой S-типа связь плотности тока с напряженностью поля не всегда однозначна: в области

$$\mathcal{E}_v < \mathcal{E} < \mathcal{E}_p$$

каждому значению напряженности поля отвечают три возможных значения плотности тока. Два из них (j_1 и j_3) находятся на возрастающих ветвях характеристики, одно (j_2) — на падающей.

Условия образования той или иной характеристики легко найти с помощью формулы (7.5). Действительно, из рис. 16.10 и 16.11 видно, что в случае характеристики N-типа дифференциальная проводимость меняет знак, проходя через нуль, а в случае характеристики S-типа — проходя через бесконечность: в нуль должен обращаться либо числитель, либо знаменатель в правой части (7.5). Таким образом, условия возникновения падающих участков с

образованием характеристик N- и S-типа можно записать в виде

$$\frac{d}{dT_e}\ln(\sigma P)=0$$

и

$$\frac{d}{dT_e}\ln\left(\frac{P}{\sigma}\right)=0$$

или

$$\frac{n^2\mu}{\tau_e}+(T_e-T)\frac{d}{dT_e}\left(\frac{n^2\mu}{\tau_e}\right)=0, \qquad N\text{-тип}, \tag{7.6}$$

и

$$1-\frac{T_e-T}{\mu\tau_e}\frac{d(\mu\tau_e)}{dT_e}=0, \qquad S\text{-тип}. \tag{7.7}$$

Чтобы придать условиям (7.6) и (7.7) явный вид, надо найти подвижность, концентрацию носителей заряда и время релаксации энергии как функции электронной температуры. Расчеты показывают *), что для ряда веществ условия (7.6) и (7.7) действительно удовлетворяются. Так, в однородном n-GaAs уже при комнатной температуре возникает характеристика N-типа. Участок отрицательной дифференциальной проводимости начинается при напряженности поля $\mathcal{E}$ около 2,3 кВ/см и исчезает при $\mathcal{E}\simeq 10$ кВ/см. В n-Ge, легированном золотом или медью, также возникает такая характеристика в области водородных или, соответственно, азотных температур решетки.

§ 8. Флуктуационная неустойчивость

С точки зрения радиотехники образец с вольтамперной характеристикой N- или S-типа представляет собой активный элемент цепи. При определенных условиях в контуре, содержащем такой элемент, могут возбуждаться незатухающие колебания тока. Частота их зависит от параметров контура. Однако активные элементы, рассмотренные в предыдущих параграфах, обладают одной важной особенностью, отличающей их от других известных в радиофизике: мы, по предположению, имели дело с технологически однородными образцами.

Следует, однако, иметь в виду, что представление о пространственной однородности в среднем не исключает малых локальных отклонений — флуктуаций — концентрации электронов, напряженности поля и т. д. от средних их значений. Эти флуктуации обусловлены, с одной стороны, беспорядочным тепловым движением носителей заряда, с другой — случайными неоднородностями в распределении примесных атомов и других структурных дефектов кристаллической решетки. Когда электронный газ находится в состоянии термодинамического равновесия или близком к нему, наличие флуктуаций обычно слабо влияет на явления переноса. Действительно, флук-

*) В применении к n-GaAs подробный расчет можно найти в книге [2].

туации, связанные с тепловым движением, быстро затухают, а вероятность появления больших флуктуаций невелика. Последнее относится и к случайным неоднородностям в распределении примеси в технологически однородном образце. Сверх того, неоднородности в распределении примесных ионов — как и всякий заряд в полупроводнике или металле — экранируются, что еще более локализует их влияние.

Положение может измениться, если электронный газ достаточно сильно нагрет. Действительно, пусть произошла малая локальная флуктуация напряженности поля $\delta\mathcal{E}(x)$ (рис. 16.12); в силу уравнения Пуассона она сопровождается и флуктуациями плотности заряда $\delta\rho = \frac{\varepsilon}{4\pi}\operatorname{div}\delta\mathcal{E}$. Далее, согласно (7.2) должна появиться и малая флуктуация плотности тока

$$\delta\mathbf{j} = \sigma_d\delta\mathcal{E}. \qquad (8.1)$$

Как видно из рис. 16.12, при $\sigma_d > 0$ флуктуация плотности тока такова, что приток заряда в область пониженной концентрации его увеличивается, а в область повышенной концентрации — уменьшается, т. е. флуктуация затухает со временем. С другой стороны, при $\sigma_d < 0$ увеличивается (уменьшается) приток заряда в область, где плотность его и без того уже повышена (понижена). Соответственно флуктуация, раз возникнув, должна, казалось бы, разрастаться — во всяком случае до тех пор, пока ее рост не будет ограничен нелинейными эффектами *). В этом рассуждении содержится, однако, одна неточность. Дело в том, что соотношение (7.2) и, следовательно, (8.1) написаны для пространственно однородной

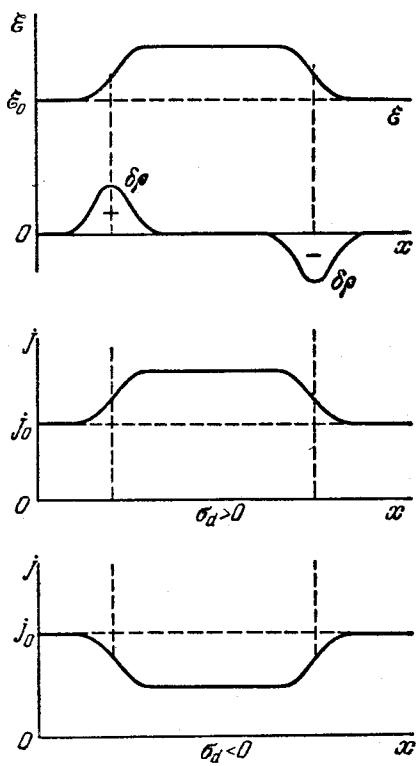


Рис. 16.12. Флуктуации напряженности поля, объемной плотности заряда и плотности тока при положительной и отрицательной дифференциальной проводимости. Через $\mathcal{E}_0$ и j_0 обозначены средние значения напряженности поля и плотности тока.

*) Формально влияние нелинейных эффектов проявится в поправках к формуле (8.1): она вытекает из (7.1) лишь при очень малых (формально — бесконечно малых) значениях $\delta\mathbf{j}$ и $\delta\mathcal{E}$.

системы носителей заряда. При наличии флуктуаций плотности заряда и электронной температуры появляются еще электронные потоки диффузионного и термоэлектрического происхождения, способствующие рассасыванию флуктуации. По этой причине флуктуация начинает расти, лишь если дрейфовая компонента флуктуации плотности тока (8.1) не только антипараллельна $\delta\mathcal{E}$, но и достаточно велика по абсолютной величине: абсолютная величина отрицательной дифференциальной проводимости должна превосходить некоторое критическое значение σ'_d. Для вычисления σ'_d надо исследовать поведение флуктуаций напряженности поля, плотности заряда и т. д. Результат зависит от конкретных особенностей рассматриваемой системы (соответствующие расчеты можно найти в книге [4]).

Коль скоро условие

$$|\sigma_d| > \sigma'_d \tag{8.2}$$

выполняется, газ носителей заряда оказывается неустойчивым относительно развития малых флуктуаций плотности заряда и напряженности поля. Об этом говорят как о флуктуационной неустойчивости системы. В результате распределение плотности заряда, напряженности электрического поля и электронной температуры в технологически однородном образце становится пространственно неоднородным. Иначе говоря, при достаточно большом отклонении от термодинамического равновесия, когда $|\sigma_d|$ достигает значения σ'_d, газ носителей заряда переходит в некоторое новое состояние, в равновесных условиях не наблюдаемое.

§ 9. Электрические домены и токовые шнуры

Исследуем распределение напряженности поля и других физических величин в пространственно неоднородном состоянии, возникающем в результате флуктуационной неустойчивости. С этой целью надо выяснить, какие именно отклонения от постоянных средних значений $\mathbf{j}$, n, $\mathcal{E}$ и T_e «опасны», т. е. не затухают, а нарастают со временем.

Удобно воспользоваться уравнением баланса энергии (7.3'), записав его в виде

$$\Gamma \equiv (\mathbf{j}, \mathcal{E}) - P = 0. \tag{9.1}$$

Величина Γ представляет собой отнесенную к единице объема и единице времени разность между энергией, получаемой электронами от поля, и энергией, отдаваемой ими решетке. В результате флуктуаций электронной температуры, плотности заряда и т. д. значение Γ может стать отличным от нуля. При этом существуют две возможности:

а) Знаки $\delta\Gamma$ и δT_e противоположны. При этом случайное уменьшение (увеличение) электронной температуры компенсируется отно-

сительным ростом (падением) притока энергии от внешнего поля. Такие флуктуации затухают со временем — они безопасны.

б) Знаки $\delta\Gamma$ и δT_e одинаковы. При этом баланс энергии, поступающей в электронный газ, таков, что приводит к дальнейшему росту флуктуации. Такие флуктуации опасны. Коль скоро они имеют место, предположение о сохранении пространственной однородности системы электронов оказывается неправильным.

Примем направление тока за ось X. Пусть электронная температура испытывает флуктуацию $\delta T_e(x)$, постоянную (при данном x) по всему сечению образца; связанная с ней флуктуация напряженности поля есть $\delta\mathcal{E}_x(x)$. Флуктуацию x-компоненты плотности тока будем

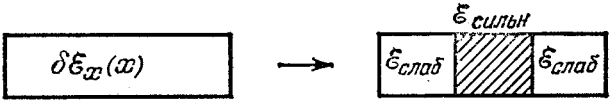


Рис. 16.13. Флуктуации электронной температуры и напряженности поля в образце с характеристикой N-типа приводят к образованию электрического домена. Справа на рисунке заштрихован домен сильного поля.

считать равной нулю: $\delta j_x = 0$ (рис. 16.13). Возможны и флуктуации другого типа, но для наших целей достаточно рассмотреть только такие: задача состоит в том, чтобы найти опасные флуктуации. В рассматриваемом случае левую часть (9.1) удобно переписать в виде

$$\Gamma = \frac{j^2}{\sigma} - P. \tag{9.1'}$$

Флуктуация ее около равновесного состояния, в котором $\Gamma = 0$, есть

$$\delta\Gamma = -\gamma_1 \delta T_e, \tag{9.2}$$

где

$$\gamma_1 = \frac{P}{\sigma}\frac{d\sigma}{dT_e} + \frac{dP}{dT_e}. \tag{9.3}$$

При $\gamma_1 > 0$ реализуется случай а), а при $\gamma_1 < 0$ — случай б). Сравнивая выражения (9.3) и (7.5), видим, что γ_1 меняет знак вместе с дифференциальной проводимостью, которая при этом проходит через нуль. Таким образом, в случае характеристики N-типа опасны флуктуации плотности заряда и компоненты напряженности поля в направлении тока *). Раз возникнув, такие флуктуации будут нарастать, пока нелинейные эффекты не ограничат дальнейшее их развитие. Распределение поля и плотности заряда станет уже

*) Более подробное исследование показывает, что флуктуации другого типа в этом случае безопасны. Доказательство этого и дальнейших утверждений, имеющихся в этом параграфе, выходит за рамки этой книги (см. [4]).

пространственно неоднородным: образец разобьется на области сильного и слабого поля, расположенные так, как показано на рис. 16.13. Область повышенной напряженности электрического поля называют доменом сильного поля или просто электрическим доменом. Распределение напряженности поля внутри самого домена может быть различным в зависимости от природы образца. С другой стороны, в областях слабого поля напряженность его практически постоянна и электронный газ почти не нагрет.

Могло бы возникнуть и несколько доменов сильного поля. Исследование показывает, однако, что такое распределение поля неустойчиво: флуктуации приводят к «слипанию» доменов в один.

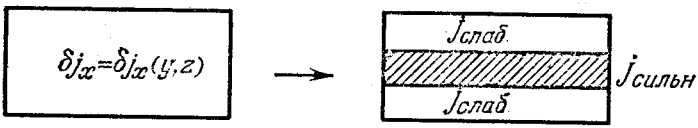


Рис. 16.14. Флуктуации электронной температуры и плотности тока в образце с характеристикой S-типа приводят к образованию токовых шнуров. Справа на рисунке заштрихован шнур сильного тока.

Пусть теперь электронная температура испытывает однородную вдоль оси X флуктуацию $\delta T_e\,(y,\ z)$, сопровождаемую флуктуацией x-компоненты плотности тока $\delta j_x\,(y,\ z)$. Флуктуацию $\delta \mathcal{E}_x$ будем считать равной нулю (рис. 16.14). Выражение (9.1) удобно переписать в виде

$$\Gamma = \sigma \mathcal{E}_x^2 - P. \tag{9.1'}$$

Флуктуация величины Γ равна

$$\delta\Gamma = -\gamma_2 \delta T_e, \tag{9.4}$$

где

$$\gamma_2 = \frac{dP}{dT_e} - \frac{P}{\sigma}\frac{d\sigma}{dT_e}. \tag{9.5}$$

Как и в предыдущем примере, рассматриваемая флуктуация безопасна при $\gamma_2 > 0$ и опасна при $\gamma_2 < 0$. Сравнивая выражения (9.5) и (7.5), видим, что γ_2 меняет знак вместе с дифференциальной проводимостью и последняя в этом случае проходит через бесконечность. Таким образом, при характеристике S-типа опасны флуктуации электронной температуры и x-компоненты плотности тока в направлении, перпендикулярном направлению тока. В результате развития этих флуктуаций образец разбивается на области слабого и сильного тока, расположенные так, как показано на рис. 16.14. Область сильного тока могла бы представлять собой как плоский слой, ориентированный параллельно оси X, так и токовый шнур

цилиндрической формы. Исследование показывает, однако, что устойчивой может быть только последняя конфигурация. Область, заштрихованная на рис. 16.14, изображает сечение шнура плоскостью чертежа.

§ 10. Движущиеся и статические домены

В технологически однородном образце тепловые флуктуации могут привести к образованию домена в любой точке. В силу однородности образца следует ожидать, что, раз возникнув, домен сможет перемещаться по образцу, пока не исчезнет на одном из электродов.

Движение доменов наблюдалось в ряде материалов: в n-GaAs, в n-Ge, легированном золотом или медью, в полуизолирующем арсениде галлия, в сульфиде кадмия n-типа, в p-GaSb и др. Это явление можно обнаружить, измеряя распределение потенциала в образце: напряжение между двумя точками, между которыми находится домен, должно быть повышено по сравнению с тем, что наблюдается в остальной части образца.

Скорость домена оказывается различной — в зависимости от механизма, ответственного за возникновение отрицательной дифференциальной проводимости однородного образца.

Здесь выделяются два предельных случая, о которых говорят, соответственно, как о дрейфовой и рекомбинационной нелинейности. В первом основную роль играет зависимость подвижности от поля — например, в силу механизма Ридли—Уоткинса—Хилсума. При этом захват и генерация носителей заряда не играют существенной роли; при достаточно большой скорости движения домена они вообще не успевают происходить. Скорость домена в этом случае есть дрейфовая скорость основных носителей заряда в слабом поле (соответствующем незаштрихованным областям на рис. 16.13). Она действительно оказывается большой в указанном только что смысле.

При рекомбинационной нелинейности основную роль играют процессы захвата и генерации носителей заряда. В зависимости от напряженности поля соотношение между концентрациями свободных и связанных носителей оказывается различным, и перемещение домена связано с перераспределением электронов между зоной и уровнями захвата. Этот процесс ограничивает скорость домена, которая оказывается значительно меньше дрейфовой. Так, в n-Ge, легированном золотом, скорость домена при водородных температурах может составлять от 10^{-5} см/с до 10^{-2} см/с.

Движение доменов сопровождается колебаниями тока в цепи нагрузки. Действительно, пока домен движется вдоль технологически однородного образца, сила тока не изменяется; но, дойдя до соответствующего электрода, домен исчезает, что ведет к временному повышению силы тока. При последующем возникновении домена на другом электроде сила тока в цепи вновь падает (рис. 16.15).

Время пролета домена по образцу равно $t_0 = L/v_0$, где L и v_0 — длина образца в направлении тока и скорость домена. При не слишком малой длине образца это время значительно превышает времена исчезновения и образования домена. При этом период колебаний тока в цепи нагрузки будет с хорошим приближением равен t_0. Соответственно для частоты этих колебаний ω мы получаем

$$\omega = 2\pi \frac{v_0}{L}. \qquad (10.1)$$

Это соотношение между ω и L хорошо оправдывается на опыте.

В арсениде галлия n-типа частота колебаний (10.1) составляет примерно $10^9 \div 10^{10}$ Гц и выше, в зависимости от длины образца. Возникновение рассматриваемых колебаний в n-GaAs и подобных ему материалах называется *эффектом Ганна*. Он используется в полупроводниковой электронике для создания генераторов и усилителей СВЧ, элементов ЭВМ и ряда других приборов.

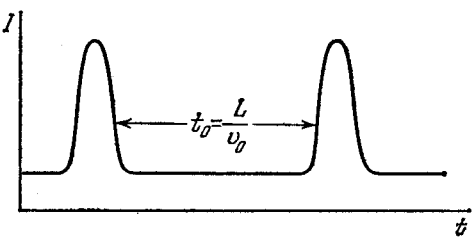


Рис. 16.15. Сила тока I в цепи нагрузки как функция времени t при движении доменов по образцу (схематически).

Электрические домены могут возникать не только в результате тепловых флуктуаций, но и благодаря флуктуациям напряженности поля, связанным со случайными статическими неоднородностями распределения примеси. Последние играют роль затравок, инициирующих возникновение макроскопически неоднородного состояния. При этом домены могут оказаться локализованными вблизи затравок. Такие домены называют статическими. Образование их не влечет за собой колебаний тока в цепи нагрузки. Оно, однако, приводит к насыщению тока: все изменение напряжения на образце падает на домене, размеры которого соответственно изменяются.

С дальнейшим повышением напряжения на образце статический домен может «оторваться» от затравки и превратиться в движущийся.

Помимо случайных неоднородностей указанного выше типа, роль затравки могут играть также электроды. Возникновение домена — статического или движущегося — связано здесь с контактным полем, которое в условиях нагрева действует так же, как и флуктуационное. Домен образуется у катода или анода, в зависимости от типа контакта и от знака носителей заряда.

Глава XVII

ПРОБЛЕМЫ ОБОСНОВАНИЯ ЗОННОЙ ТЕОРИИ И ЗАДАЧИ, ВЫХОДЯЩИЕ ЗА ЕЕ РАМКИ

§ 1. Три вопроса к зонной теории

Как мы видели в предыдущих главах, зонная теория твердого тела позволяет успешно интерпретировать множество экспериментальных данных, касающихся электрических, гальваномагнитных, оптических и других свойств полупроводников. Вместе с тем справедливость основных предположений, на которых базируется зонная теория, отнюдь не очевидна. Поэтому возникает задача об обосновании зонной теории и об установлении границ ее применимости. Здесь можно выделить три основных вопроса, соответствующие трем основным предположениям, принятым в § III.1.

Во-первых, уравнение Шредингера (III.2.4) описывает движение электрона при фиксированном расположении тяжелых частиц, рассматриваемых просто как источники поля. Фактически тяжелые частицы также движутся. Это движение мы изучали в гл. XII. Там, однако, не принималось явно во внимание наличие свободных электронов. Между тем они не только испытывают действие со стороны тяжелых частиц, но и сами действуют на них. Возникает вопрос: можно ли — и если да, то на каком основании — рассматривать по отдельности движение тяжелых частиц — атомных ядер — и электронов?

Во-вторых, уравнение (III.2.4) описывает движение электрона в периодическом поле. Это означает, в частности, что решетка предполагается идеальной: тяжелые частицы считаются покоящимися строго в ее узлах, в результате чего потенциальная энергия электрона оказывается периодической функцией координат. Фактически, однако, такое состояние не может реализоваться: оно отвечало бы точно заданным значениям координат и скоростей частиц, что не допускается принципом неопределенности. Иначе говоря, кристаллическая решетка в принципе не может быть идеальной. Отражением этого обстоятельства являются нулевые колебания решетки. Как мы видели в гл. XIV, наличие колебаний решетки (в том числе и нулевых) приводит к появлению дополнительного слагаемого в операторе энергии электрона — энергии взаимодействия электрона с колебаниями решетки.

Возникает вопрос: можно ли — и если да, то на каком основании — рассматривать задачу об энергетическом спектре электрона

в идеальной решетке, пренебрегая его взаимодействием с ее колебаниями?

Наконец, в-третьих, уравнение (III.2.4) описывает движение одного электрона. На самом деле в любой конденсированной среде, в том числе и в полупроводнике, мы всегда имеем дело со многими электронами, взаимодействующими друг с другом. Такая система описывается волновой функцией, зависящей от координат и спинов всех электронов, а не только одного из них. Для определения этой волновой функции надо было бы написать уравнение Шредингера с гамильтонианом, содержащим не только операторы кинетической энергии электронов и потенциальной энергии их в поле тяжелых частиц, но и операторы энергии взаимодействия электронов друг с другом.

Возникает вопрос: можно ли — и если да, то на каком основании — рассматривать одноэлектронную задачу (III.2.4) вместо многоэлектронной?

Проблема обоснования данной теории твердого тела состоит в том, чтобы выяснить, можно ли дать положительные ответы на три поставленных выше вопроса.

Как мы увидим, такая возможность действительно существует для определенного (достаточно широкого) класса материалов и круга явлений. Обращаясь к изучению других материалов и других явлений, мы тем самым перейдем к задачам, выходящим за рамки зонной теории.

§ 2. Адиабатическое приближение

Обратимся к вопросу о возможности раздельно рассматривать движение электронов и тяжелых частиц. Под последними пока будем понимать ядра атомов кристаллической решетки.

Физическая причина, оправдывающая указанное разделение, состоит в резком различии масс электрона и ядра. В силу этого различия можно ожидать, что скорости и ускорения электронов будут в среднем значительно превышать скорости и ускорения ядер. При этом характер движения электронов будет определяться мгновенным расположением ядер. С другой стороны, ядра, в силу инерционности своего движения, будут «замечать» лишь среднее расположение электронов.

Задача состоит теперь в квантовомеханическом оформлении этих полуинтуитивных соображений классического характера. При этом, в частности, надо придать и точный смысл словам «среднее расположение» и т. п.

Пронумеруем электроны индексами i, j, а тяжелые частицы — индексами a, b. Каждый из этих индексов принимает, независимо от других, столько значений, сколько соответствующих частиц есть в системе. Обозначим радиус-векторы электронов и тяжелых частиц,

соответственно, через $\mathbf{r}_i$ и $\mathbf{R}_a$; совокупности этих переменных обозначим для краткости через $r = \{\mathbf{r}_1, \ldots\}$ и $R = \{\mathbf{R}_1, \ldots\}$. Спиновые координаты электронов явно указывать не будем, подразумевая лишь в случае необходимости, что они включены в r. Массы тяжелых частиц обозначим через M_a. Для дальнейшего существенно лишь, что все они гораздо больше массы электрона m_0. Потенциальные энергии взаимодействия двух электронов друг с другом, электрона с ядром и двух ядер друг с другом обозначим, соответственно, через $U_1(\mathbf{r}_i - \mathbf{r}_j)$, $U_2(\mathbf{r}_i - \mathbf{R}_a)$ и $U_3(\mathbf{R}_a - \mathbf{R}_b)$.

Уравнение Шредингера для рассматриваемой системы частиц имеет вид

$$H\Psi = W\Psi, \tag{2.1}$$

где W — полная энергия системы. Волновая функция Ψ зависит от совокупностей всех координат $r = \{\mathbf{r}_1, \ldots\}$ и $R = \{\mathbf{R}_1, \ldots\}$, а гамильтониан H дается выражением

$$H = -\sum_i \frac{\hbar^2}{2m_0}\nabla_i^2 - \sum_a \frac{\hbar^2}{2M_a}\nabla_a^2 + \frac{1}{2}\sum_{i\neq j} U_1(\mathbf{r}_i - \mathbf{r}_j) + \\ + \sum_{i,a} U_2(\mathbf{r}_i - \mathbf{R}_a) + \frac{1}{2}\sum_{a\neq b} U_3(\mathbf{R}_a - \mathbf{R}_b). \tag{2.2}$$

Здесь символы ∇_i^2 и ∇_a^2 обозначают лапласианы по переменным $\mathbf{r}_i$ и $\mathbf{R}_a$.

Нам предстоит выяснить, нельзя ли свести уравнение (2.1) к системе двух более простых уравнений, описывающих, соответственно, движение электронов при заданном расположении ядер и (более медленное) движение самих ядер. В соответствии с высказанной выше физической идеей представим волновую функцию Ψ в виде

$$\Psi(r, R) = \chi(r; R)\,\Phi(R). \tag{2.3}$$

Здесь χ и Φ — неизвестные пока функции, причем желательно, чтобы χ зависела от координат R только параметрически. Это означает следующее: желательно, чтобы переменные $\mathbf{R}_1, \ldots$ входили в уравнение для χ только как аргументы, от которых зависит потенциальная энергия, но не как координаты, по которым производится дифференцирование.

Подставляя выражение (2.3) в уравнение (2.1), получаем

$$\left\{-\frac{\hbar^2}{2m_0}\sum_i \nabla_i^2\chi + \frac{1}{2}\sum_{i\neq j} U_1(\mathbf{r}_i - \mathbf{r}_j)\chi + \sum_{i,a} U_2(\mathbf{r}_i - \mathbf{R}_a)\chi\right\}\Phi + \\ + \left\{-\frac{\hbar^2}{2}\sum_a \frac{1}{M_a}\nabla_a^2\Phi + \frac{1}{2}\sum_{a\neq b} U_3(\mathbf{R}_a - \mathbf{R}_b)\Phi\right\}\chi + \\ + \sum_a\left\{\left(-\frac{i\hbar}{M_a}\nabla_a\chi,\ -i\hbar\nabla_a\Phi\right) - \frac{\hbar^2}{2M_a}\Phi\nabla_a^2\chi\right\} = W\chi\Phi. \tag{2.4}$$

В третьей строке уравнения (2.4) выписаны слагаемые, содержащие производные от χ по координатам ядер: в первом члене на функцию χ действуют операторы скорости тяжелых частиц $(-i\hbar/M_a)\nabla_a$, во втором — операторы их кинетической энергии $(-\hbar^2/2M_a)\nabla_a^2$. С другой стороны, в первой строке (2.4) на функцию χ действуют операторы кинетической энергии электронов $(-\hbar^2/2m_0)\nabla_i^2$. Поскольку ядра в среднем движутся гораздо медленнее легких частиц, можно ожидать, что при определении энергетического спектра электронов слагаемые в третьей строке (2.4) будут сравнительно несущественны. Пренебрегая ими, мы можем выполнить в (2.4) разделение «электронных» и «ядерных» переменных. В самом деле, приравняем коэффициент при Φ в первой строке (2.4) некоторой функции $E(R)\chi$. Получим

$$-\frac{\hbar^2}{2m_0}\sum_i \nabla_i^2\chi+\frac{1}{2}\sum_{i\neq j} U_1(\mathbf{r}_i-\mathbf{r}_j)\chi+\sum_{i,a} U_2(\mathbf{r}_i-\mathbf{R}_a)\chi=E(R)\chi, \quad (2.5а)$$

$$-\sum_a \frac{\hbar^2}{2M_a}\nabla_a^2\Phi+\frac{1}{2}\sum_{a\neq b} U_3(\mathbf{R}_a-\mathbf{R}_b)\Phi+E(R)\Phi=W\Phi. \quad (2.5б)$$

Равенство (2.5а) есть не что иное, как уравнение Шредингера для системы электронов, взаимодействующих друг с другом и с атомными ядрами, покоящимися в некоторых фиксированных точках пространства. Координаты последних $\mathbf{R}_1$, ... входят в уравнение как параметры. Величина $E(R)$ имеет смысл собственного значения энергии рассматриваемой системы. Естественно, она зависит от параметров $\mathbf{R}_1$, ..., входящих в уравнение, т. е. от расположения ядер в данный момент времени.

Как известно из квантовой механики, $E(R)$ можно рассматривать и как квантовомеханическое среднее значение энергии электронов при данной конфигурации ядер. Действительно, умножим уравнение (2.5а) слева на функцию χ^* (принадлежащую данному собственному значению E) и проинтегрируем по электронным координатам (включая и суммирование по спиновым координатам). Замечая, что имеет место условие нормировки

$$\int dr\,|\chi(r)|^2=1,$$

где dr есть произведение дифференциалов $d\mathbf{r}_1$, ..., получим

$$E(R)=\int \chi^*\left\{-\frac{\hbar^2}{2m_0}\sum_i \nabla_i^2\chi+\right.$$
$$\left.+\frac{1}{2}\sum_{i,j} U_1(\mathbf{r}_i-\mathbf{r}_j)\chi+\sum_{i,a} U_2(\mathbf{r}_i-\mathbf{R}_a)\chi\right\} dr. \quad (2.6)$$

В правой части (2.6) стоит как раз выражение для квантовомеханического среднего значения полной энергии электронов при заданной конфигурации ядер.

Равенство (2.5б) представляет собой не что иное, как уравнение Шредингера для системы ядер. Потенциальная энергия ее есть

$$V(R) = \frac{1}{2}\sum_{a \neq b} U_3(\mathbf{R}_a - \mathbf{R}_b) + E(R). \qquad (2.7)$$

Видим, что в рамках принятого приближения ядра действительно испытывают лишь усредненное воздействие электронов: средняя энергия последних входит как потенциальная энергия в уравнение движения тяжелых частиц.

Уравнения (2.5а) и (2.5б) осуществляют программу, намеченную в начале этого параграфа. Приближение, в рамках которого они получены (пренебрежение слагаемыми в третьей строке уравнения (2.4)), называется *адиабатическим*: функция $\chi(r; R)$ описывает поведение системы электронов при бесконечно медленном («адиабатическом») изменении параметров R.

Слагаемые, содержащиеся в третьей строке уравнения (2.4), называются членами неадиабатичности; они представляют собой результат применения «оператора неадиабатичности»

$$\sum_a \left\{ -\frac{\hbar^2}{M_a}(\nabla_a \Phi, \nabla) - \frac{\hbar^2}{2M_a}\Phi\nabla^2 \right\}$$

к функции χ.

Таким образом, мы получили ответ на первый из вопросов, поставленных в предыдущем параграфе: раздельное рассмотрение движения электронов и ядер возможно в рамках адиабатического приближения.

Для оценки степени точности адиабатического приближения надо, рассматривая оператор неадиабатичности как возмущение, вычислить связанные с ним поправки к энергетическому спектру системы и к другим наблюдаемым величинам. В простейшем случае, когда основное состояние рассматриваемой системы электронов и ядер не вырождено и отделено достаточно большой энергетической щелью от первого возбужденного состояния, результат оказывается довольно простым [1]: поправки к адиабатическому приближению — порядка не ниже $(m_0/M)^{1/4}$, где M — масса ядра. При отказе от указанных только что предположений возникают трудности, связанные с применением теории возмущений в непрерывном спектре. Полная ясность здесь еще не достигнута.

Отметим, что есть явления, при изучении которых учет неадиабатичности необходим принципиально. Так обстоит дело, если мы интересуемся обменом энергией между электронами и решеткой *). Одна из важных задач этого типа есть задача о захвате свободных электронов и дырок ловушками.

*) Рассеяние электронов фононами (гл. XIV) также есть пример неадиабатического процесса. Не случайно константа C в выражении (XIV.4.19) содержит в числителе массу элементарной ячейки M.

§ 3. Приближение малых колебаний

В рамках адиабатического приближения уравнение (2.5а) описывает поведение электронов в любой атомной системе — твердом теле, жидкости, газе, отдельной молекуле и т. д. Специализируя его для идеального кристалла, надо считать, что атомные ядра расположены в узлах кристаллической решетки. Очевидно, соответствующие значения R должны отвечать минимуму потенциальной энергии ядер (2.7): в противном случае решетка была бы неустойчива. Само выражение (2.7), фигурирующее в уравнении (2.5б), естественно разложить в ряд по малым отклонениям от положений равновесия. Это *приближение малых колебаний* оправдано, пока амплитуда колебаний мала по сравнению с постоянной решетки, т. е. пока температура кристалла не слишком близка к точке плавления. Пользуясь таким приближением, мы приходим к постановке задачи о малых колебаниях, принятой в гл. XII.

Согласно (2.7) потенциальная энергия системы ядер зависит от состояния электронов. Следовательно, то же относится и к положениям равновесия, и к частотам нормальных колебаний: последние, как мы знаем, определяются вторыми производными от функции $V(R)$ (2.7), взятыми в точке ее минимума. Таким образом, разным состояниям системы электронов соответствуют, вообще говоря, различные фононные спектры. Это обстоятельство не играло существенной роли в теории подвижности, ибо, как мы видели в § XIV.4, изменение энергии электрона при поглощении или испускании одного акустического фонона сравнительно невелико. Однако при захвате электрона примесным центром или иным структурным дефектом решетки положение может измениться. Как мы увидим в § 9, это обстоятельство существенно для правильной оценки вероятности захвата.

До сих пор мы говорили о тяжелых частицах как об атомных ядрах. Однако весьма часто роль тяжелых частиц могут играть и атомные остовы, содержащие электроны внутренних оболочек. Действительно, энергия связи внутренних электронов с ядром обычно столь велика, что состояния их лишь очень мало меняются как при образовании кристалла, так и при различных электронных процессах в нем *). Далее, внутренние электроны в среднем располагаются гораздо ближе к ядру, нежели валентные. Большей степени локализации отвечает и большая кинетическая энергия: валентные электроны движутся в среднем заметно медленнее внутренних. Это позволяет вновь применить адиабатическое приближение, рассматривая на сей раз внутренние электроны как «быструю подсистему», а валентные — как «медленную»; такой подход оправдан, коль

*) Исключение могут составить лишь атомы самых легких элементов.

скоро энергии ионизации первых гораздо больше, нежели вторых. Таким образом, оказывается возможным рассматривать систему валентных электронов, движущихся в поле атомных ядер и среднем поле внутренних электронов. Иначе говоря, индекс i в уравнении (2.5а) нумерует теперь только валентные электроны, а величина U_2 есть потенциальная энергия электрона в поле атомного остова. Это приближение можно назвать валентным. Им обычно пользуются при изучении большинства полупроводников. Пример, когда валентное приближение может оказаться недостаточным, представляют вещества, содержащие ионы или атомы переходных или редкоземельных элементов: электроны незаполненных внутренних оболочек также надо рассматривать явно.

Подчеркнем, наконец, что разделение всех электронов на «входящие в атомные остовы» и «рассматриваемые явно» не носит абсолютного характера, а зависит от изучаемого явления. Если энергия какого-либо внешнего воздействия сравнима с энергией ионизации внутренних оболочек, то соответствующие электроны надо рассматривать явно. Так, например, обстоит дело в случае поглощения рентгеновских лучей кристаллами или при прохождении быстрых атомных частиц через вещество.

§ 4. Роль колебаний решетки. Полярон

Обратимся ко второму из вопросов, поставленных в § 1, — о роли неизбежной неидеальности решетки, связанной с ее колебаниями. В сущности, ответ на этот вопрос содержится уже в принятой в гл. XIV постановке задачи о рассеянии носителей заряда: в задаче об энергетическом спектре электронов неидеальностью решетки можно пренебречь, если она достаточно мало влияет на полную энергию электрона. Соответствующий количественный критерий не отличается от условий применимости кинетического уравнения (XIV. 2.22а, б).

Как показывают оценки, во всех исследованных до сих пор материалах взаимодействие носителей заряда с акустическими колебаниями оказывается слабым в указанном смысле. То же относится и к взаимодействию с оптическими колебаниями в неполярных кристаллах и в полупроводниках типа $A^{III}B^{V}$. Положение, однако, радикально меняется при переходе к ионным кристаллам — щелочно-галоидным и др. Взаимодействие электронов и дырок с продольными оптическими колебаниями в этих веществах оказывается столь сильным, что знак в неравенстве (XIV. 2.22а) может измениться на обратный.

Причину этого легко понять. В ионных кристаллах продольные оптические колебания решетки связаны с изменениями дипольных моментов элементарных ячеек. При этом меняется и вектор инер-

ционной поляризации среды *). По этой причине колебания указанного типа часто называют поляризационными. Энергия взаимодействия электрона с ними есть не что иное, как изменение энергии взаимодействия заряда с дипольными моментами элементарных ячеек при оптических колебаниях. В силу сравнительно медленного убывания напряженности поля диполя с расстоянием вклад в энергию в каждой точке пространства происходит от большого числа элементарных ячеек. Это приводит к сравнительно большому значению энергии взаимодействия и, соответственно, малому значению времени релаксации, формально вычисленному по обычной схеме кинетического уравнения. Отсюда следует, что в рассматриваемых материалах взаимодействие электронов с поляризационными колебаниями решетки надлежит явно принимать во внимание при определении энергетического спектра электронов и дырок. Представление о движении электрона в идеальной решетке здесь становится неоправданным, и мы приходим к задаче, выходящей за рамки зонной теории.

Характер энергетического спектра электрона, сильно взаимодействующего с поляризационными колебаниями решетки, можно представить себе с помощью наглядных соображений полуклассического характера. Рассмотрим поведение заряженной частицы в инерционно поляризующейся среде. Взаимодействие заряда с ионами среды приводит к появлению поляризации. Связанное с ней электрическое поле в свою очередь действует на заряд, понижая его потенциальную энергию. Возникает созданная самим зарядом поляризационная потенциальная яма, в которой он и находится **). Выйти из нее (т. е. деполяризовать решетку) заряду энергетически невыгодно. Перемещение заряда по решетке сопровождается и перемещением созданной им поляризации. При этом инерционность поляризации замедляет движение заряда, ибо для «рассасывания» потенциальной ямы в одном месте кристалла и возникновения ее в другом требуется конечное время. Таким образом, роль носителей заряда в инерционно поляризующейся среде играют не электроны (или дырки) обычного типа, а более сложные объекты — электроны (дырки) вместе с созданными ими поляризационными ямами. Эти объекты получили название *поляронов* ***).

---

*) Как известно, инерционная поляризация кристалла обусловлена смещениями ионов решетки из их положений равновесия, безынерционная — деформацией электронных оболочек ионов.

**) Об этом иногда говорят как об «автолокализации» электрона. Пользуясь этим термином, следует помнить о несколько условном его смысле: в отсутствие каких-либо структурных дефектов такая яма может быть с равной вероятностью обнаружена в любой точке кристалла.

***) Этот термин был предложен С. И. Пекаром [2], впервые подробно изучившим рассматриваемые состояния. Исходное представление об автолокализации электрона в инерционно поляризующейся среде восходит к работам Л. Д. Ландау и Я. И. Френкеля, выполненным в 30-х годах. Количественная теория

Размер полярона определяется размером поляризационной потенциальной ямы. Здесь выделяются два предельных случая.

1) Радиус ямы значительно превышает постоянную решетки. В этом случае говорят о поляронах большого радиуса. Такие поляроны возникают, видимо, в ряде ионных кристаллов — щелочно-галоидных, закиси меди и др.

2) Радиус ямы близок к постоянной решетки. При этом говорят о поляронах малого радиуса. Подобные объекты, по-видимому, образуются в рутиле.

Движение полярона как целого можно охарактеризовать квазиимпульсом. Как и квазиимпульс свободного электрона в зонной теории, он описывает «почти стационарные» (стационарные в отсутствие рассеяния) состояния системы: энергию движения полярона как целого можно рассматривать как функцию квазиимпульса. Это позволяет ввести представления о поляронных энергетических зонах и об эффективной массе полярона. Очевидно, последняя превышает эффективную массу, которой в том же кристалле обладал бы свободный электрон в пренебрежении его взаимодействием с инерционной поляризацией. Соответственно поляронная зона оказывается более узкой.

Сложные свойства носителя заряда — полярона — особенно ярко проявляются в том, что полярон может иметь внутреннюю структуру: поляризационная потенциальная яма, созданная электроном (дыркой), содержит несколько дискретных уровней, отвечающих разным распределениям плотности заряда в яме и разным значениям ее радиуса *). Переходы между дискретными уровнями в поляроне могут быть вызваны, например, электромагнитным излучением. Связанное с ними поглощение света, по-видимому, наблюдалось в металл-аммиачных растворах; роль носителей заряда при этом играли не электроны, а ионы металла, движущиеся в полярном растворителе.

Другая особенность поляронных состояний состоит в том, что они исчезают в достаточно сильных электрических полях. Действительно, поляризационная яма, сопровождающая электрон, представляет собой не что иное, как набор продольных оптических фононов с разными длинами волн. По этой причине скорость ее перемещения не может превышать групповую скорость поляризационных фононов. Следовательно, при достаточно большой дрейфовой скорости электрон «отрывается» от поляризационной ямы, превращаясь в обычный зонный электрон: поляризационная яма при этом прекращает свое существование.

поляронов большого радиуса была развита в 40—50-х годах в работах Н. Н. Боголюбова, С. И. Пекара и С. В. Тябликова.

*) Термин «дискретный уровень» имеет смысл только при заданном значении квазиимпульса полярона. При учете движения полярона как целого каждый из уровней размывается в узкую зону.

§ 5. Метод самосогласованного поля

Обратимся к вопросу о возможности свести многоэлектронную задачу к одноэлектронной. Для определенности ограничимся валентным приближением, рассматривая явно только электроны валентных оболочек.

В адиабатическом приближении гамильтониан рассматриваемой системы многих электронов есть не что иное, как оператор H_e, действующий на χ в правой части уравнения (2.5а). Перепишем его для краткости в виде

$$H_e = \sum_i \left\{ -\frac{\hbar^2}{2m_0}\nabla_i^2 + U_0(\mathbf{r}_i) \right\} + \frac{1}{2}\sum_{i\neq j} U_1(\mathbf{r}_i - \mathbf{r}_j). \qquad (5.1)$$

Здесь через

$$U_0(\mathbf{r}_i) = \sum_a U_1(\mathbf{r}_i - \mathbf{R}_a)$$

обозначена потенциальная энергия взаимодействия i-го электрона со всеми тяжелыми частицами, расположенными в узлах идеальной решетки данного кристалла.

Последнее слагаемое в правой части (5.1), описывающее взаимодействие между электронами, строго говоря, не позволяет свести уравнение Шредингера с гамильтонианом H_e к одноэлектронному. Можно, однако, попытаться заменить взаимодействие между электронами взаимодействием их с некоторым эффективным внешним полем, подобрав последнее наиболее выгодным образом. Откладывая пока вопрос о подборе эффективного поля, заменим гамильтониан H_e на

$$H'_e = \sum_i H_i, \qquad H_i = -\frac{\hbar^2}{2m_0}\nabla_i^2 + U_0(\mathbf{r}_i) + U'(\mathbf{r}_i), \qquad (5.2)$$

где $U'(\mathbf{r}_i)$ — энергия взаимодействия i-го электрона с эффективным полем. Тогда уравнение Шредингера (2.5а) принимает вид

$$\sum_i H_i(\mathbf{r}_i)\,\chi = E(R)\,\chi. \qquad (5.3)$$

В уравнении (5.3) можно произвести разделение переменных. Положим

$$\chi(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \prod_{j=1}^{N} \psi(\mathbf{r}_j), \qquad (5.4)$$

где N — полное число (валентных) электронов, ψ — пока неизвестные функции, каждая из которых зависит от координат (пространственных и спиновой) только одного из электронов.

Подставляя (5.4) в (5.3) и замечая, что оператор $H_i(\mathbf{r})$ действует только на координаты i-го электрона, получим

$$\sum_i H_i(\mathbf{r}_i)\,\psi(\mathbf{r}_i)\prod_{j\neq i}\psi(\mathbf{r}_j)=E\prod_{j\neq i}\psi(\mathbf{r}_j)\,\psi(\mathbf{r}_i).$$

Это уравнение удовлетворяется, лишь если

$$H_i\psi(\mathbf{r}_i)=E_i\psi(\mathbf{r}_i), \tag{5.5}$$

где E_i — величины, не зависящие от r, причем

$$\sum_i E_i=E. \tag{5.6}$$

Равенство (5.5) есть не что иное, как одноэлектронное уравнение Шредингера. Из дальнейшего будет видно, что в идеальной решетке эффективную потенциальную энергию $U'(\mathbf{r}_i)$ можно определить как периодическую функцию координат. Поскольку функция $U_0(\mathbf{r}_i)$ также периодична, мы получили уравнение (III. 2.4), в котором лишь опущен индекс i. При этом $U=U_0+U'$, а величины $\psi(\mathbf{r}_i)$ имеют смысл волновых функций отдельных электронов. Как и волновые функции ψ_i, величины E_i зависят от квантовых чисел отдельных электронов λ_i. Как мы знаем, в зонной теории роль этих квантовых чисел играют три компоненты квазиимпульса и номер зоны.

Конечно, это еще не есть полное обоснование зонной теории с последовательно многоэлектронной точки зрения. Действительно, возможность перейти к одноэлектронной постановке задачи здесь не доказана, а навязана заранее заменой H_e на H'_e. Смысл метода, излагаемого в настоящем параграфе, состоит лишь в том, чтобы наиболее точно выбрать эффективную потенциальную энергию U', получив тем самым наилучшую аппроксимацию в рамках одноэлектронного приближения.

Обратимся теперь к подбору эффективного поля U'. Равенство

$$\frac{1}{2}\sum_{i\neq j}U_1(\mathbf{r}_i-\mathbf{r}_j)=\sum_i U'(\mathbf{r}_i)$$

в точности не может удовлетворяться. Действительно, в левой его части стоит сумма функций двух переменных, а в правой каждое из слагаемых зависит только от одной переменной. Можно, однако, потребовать, чтобы оно удовлетворялось в среднем. Для этой цели определим $U'(\mathbf{r}_i)$ как потенциальную энергию i-го электрона в поле, создаваемом средней плотностью заряда всех остальных электронов:

$$U'(\mathbf{r}_i)=e^2\sum_{j\neq i}\int\frac{|\psi_{\lambda_j}(\mathbf{r}')|^2}{|\mathbf{r}_i-\mathbf{r}'|}\,d\mathbf{r}'. \tag{5.7}$$

В отличие от внешних полей, потенциальная энергия $U'(\mathbf{r}_i)$ a priori не задана: она зависит от вида волновых функций ψ_{λ_j}, которые в

свою очередь сами определяются через U' из уравнений (5.5). Такое поле называется *самосогласованным* *). Подставляя в (5.7) в качестве $\psi_{\lambda_j}(\mathbf{r}')$ функции Блоха, можно убедиться, что функция $U'(\mathbf{r}_j)$ действительно периодична с периодом решетки.

Волновая функция (5.4) имеет мультипликативный вид. Поэтому такой же оказывается и вероятность обнаружить электроны, соответственно, в элементах объема $d\mathbf{r}_1 \ldots d\mathbf{r}_N$:

$$|\chi(\mathbf{r}_1, \ldots, \mathbf{r}_N)|^2 d\mathbf{r}_1 \ldots d\mathbf{r}_N = \prod_{i=1}^{N} |\psi_{\lambda_i}(\mathbf{r}_i)|^2 d\mathbf{r}_i. \qquad (5.8)$$

По известной теореме теории вероятностей это означает, что в приближении самосогласованного поля попадания электронов в элементы объема $d\mathbf{r}_1$, $d\mathbf{r}_2$, ... суть независимые события. Иначе говоря, волновая функция (5.4) не учитывает корреляции в поведении электронов. Фактически такая корреляция имеет место. Она обусловлена двумя причинами — принципом Паули и взаимодействием между электронами. Согласно принципу Паули волновая функция системы электронов $\psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ должна быть антисимметричной относительно перестановки координат и спинов любых двух электронов (что исключает возможность пребывания двух электронов в одном и том же стационарном состоянии). Волновая функция (5.4) не удовлетворяет этому требованию антисимметрии. Как известно из квантовой механики, для учета принципа Паули надо произведение функций $\psi(\mathbf{r}_j)$ в правой части (5.4) заменить детерминантом

$$\chi = (N!)^{-1/2} \operatorname{Det} |\psi_{\lambda_i}(\mathbf{r}_j)|. \qquad (5.4')$$

Множитель $(N!)^{-1/2}$ здесь введен для нормировки χ на единицу, индексы i и j нумеруют, соответственно, строки и столбцы детерминанта. При этом в правой части (5.7) появляется дополнительное («обменное») слагаемое, также содержащее интегралы от функции ψ_{λ_j} **).

Корреляция, связанная с кулоновским отталкиванием между электронами, должна проявиться в зависимости функции $\chi(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ от расстояний между электронами: вероятность обнаружить электроны на малых расстояниях друг от друга должна убывать с уменьшением этих расстояний. В рамках метода самосогласованного поля этот эффект теряется. Роль его при определении энерге-

*) Метод самосогласованного поля в такой форме был предложен в 20-х годах. Д. Хартри в связи с теорией атома.

**) Этот «метод самосогласованного поля с обменом» был предложен В. А. Фоком в 1930 г. В. А. Фок показал, что решения уравнений самосогласованного поля с обменом — наилучшие из всех одноэлектронных функций вида (5.4'): они дают самое точное значение энергии системы.

тического спектра электронов зависит от их концентрации n и от степени вырождения. В условиях сильного вырождения влияние корреляции оказывается тем меньшим, чем больше концентрация. Действительно, как мы видели в § V.6, для полностью вырожденного идеального газа электронов

$$\zeta = F - E_c \sim \frac{\hbar^2 n^{2/3}}{2m}. \tag{5.9}$$

Средняя энергия, приходящаяся на один электрон, отличается от этого выражения лишь несущественным постоянным множителем и, следовательно, также растет как $n^{2/3}$. Это есть следствие принципа Паули: при увеличении концентрации электронов должны заполняться состояния со все большей и большей энергией. С другой стороны, средняя энергия кулоновского взаимодействия $U_{\text{кул}}$ должна быть порядка $e^2/\varepsilon\tilde{r}$, где $\tilde{r} \sim n^{-1/3}$ — среднее расстояние между электронами (диэлектрическую проницаемость ε здесь следует писать, лишь если $\tilde{r}$ гораздо больше постоянной решетки; в противном случае ε надо заменить единицей):

$$U_{\text{кул}} \sim \frac{e^2}{\varepsilon} n^{1/3}. \tag{5.10}$$

Таким образом, в полностью вырожденном электронном газе

$$\frac{U_{\text{кул}}}{\zeta} \sim \left(\frac{\varepsilon\hbar^2}{me^2} n^{1/3}\right)^{-1}. \tag{5.11}$$

При увеличении n относительная роль взаимодействия между электронами, а потому и роль корреляции уменьшается — сильно вырожденный газ ферми-частиц тем ближе к идеальному, чем больше его концентрация *).

С другой стороны, в невырожденном газе относительная роль взаимодействия между электронами и, следовательно, роль корреляции убывает с уменьшением концентрации электронов.

Как видно из формулы (5.7), система уравнений (5.5) оказывается нелинейной интегро-дифференциальной. Решения ее удается получать лишь численными методами.

Тем не менее метод самосогласованного поля (с обменом) широко применяется в настоящее время для расчета зонного энергетического спектра твердых тел **).

---

*) Современные методы теории многих тел позволяют рассчитать влияние корреляции в пространственном расположении электронов на среднюю энергию и другие термодинамические свойства электронной жидкости. Результаты оказываются в согласии с оценкой (5.11) [3].

**) Обзор работ на эту тему можно найти в книге [4].

§ 6. Электроны и дырки как элементарные возбуждения многоэлектронной системы в полупроводнике

Последовательное обоснование зонной теории твердого тела можно получить с позиций современной теории многих тел. Ведущую роль здесь играет представление об элементарных возбуждениях квантовой системы многих частиц [М9].

Суть дела удобно пояснить, рассматривая уже известный нам пример явного решения задачи многих тел — задачу о малых колебаниях кристаллической решетки. Как мы видели в гл. XII, энергия слабо возбужденных состояний такой системы представляется как энергия газа квазичастиц — элементарных возбуждений системы, — называемых фононами. Идеальность этого газа обусловлена только малостью возбуждения — т. е. малостью амплитуд колебаний атомов по сравнению с постоянной решетки — и никак не связана с какими-либо предположениями о характере взаимодействия между атомами. Ангармоничность колебаний приводит к слабой неидеальности фононного газа.

Этот пример иллюстрирует общую ситуацию: слабо возбужденные состояния любой квантовой системы многих тел можно представить как идеальный или слабо неидеальный газ (или газы) квазичастиц — элементарных возбуждений того или иного типа. Эти квазичастицы могут подчиняться статистике Ферми—Дирака или Бозе—Эйнштейна — независимо от типа статистики, которой подчиняются сами частицы системы *). Соответственно говорят о фермиевской и бозевской ветвях спектра элементарных возбуждений. В зависимости от природы системы квазичастицы могут обладать импульсом или квазиимпульсом (последнее — в кристаллах). Они могут также нести электрический заряд, спин и т. д. Полное число квазичастиц данной ветви бывает как постоянным, так и переменным. Очевидно, что если переменным оказывается число заряженных квазичастиц, то среди них должны быть как положительно, так и отрицательно заряженные. При этом из закона сохранения заряда вытекает, что они могут возникать и исчезать только парами.

Перемещаясь по кристаллической решетке, квазичастицы могут переносить заряд, энергию и т. д. Естественно возникает мысль, что носители заряда в полупроводнике — электроны проводимости и дырки — следует рассматривать как элементарные возбуждения системы многих электронов. Наглядно их можно представлять себе как состояния с избытком или недостатком отрицательного заряда («лишние» электроны и дырки). Чтобы убедиться в справедливости сказанного, надо исследовать структуру энергетического спектра полупроводника, показать, что такие квазичастицы ведут себя во

*) Так, какой бы статистике ни подчинялись атомы кристаллической решетки, статистика фононного газа оказывается бозевской.

внешних электрическом и магнитном полях (в том числе в поле примеси), как этого следует ожидать от электронов проводимости и дырок, и т. д. Эту программу действительно удается реализовать.

Таким образом, обоснование ряда важнейших представлений зонной теории твердого тела оказывается несколько своеобразным: выясняется, что на самом деле эти представления носят строгий многоэлектронный характер. Подчеркнем, однако, что это относится не ко всем зонам. Точное обоснование получают представления о зоне (зонах) проводимости и о дырочной зоне, но отнюдь не о валентной зоне, трактуемой как совокупность одноэлектронных энергий. Иначе говоря, удается обосновать как раз те понятия, которые фактически используются для обработки и интерпретации экспериментальных данных: точный многоэлектронный смысл придается зонной энергетической схеме (с локальными уровнями). Известное обоснование получает также и аппарат зонной теории: уравнения, определяющие энергетический спектр рассматриваемых элементарных возбуждений, приближенно сводятся к уравнениям самосогласованного поля с обменом. Вместе с тем в рамках данного приближения учитывается и ряд многоэлектронных эффектов. С одним из них мы уже встречались в § XIII. 9. Речь идет о коллективных колебаниях плотности газа носителей заряда. Как и в случае колебаний решетки, плазменным волнам можно сопоставить кванты колебаний с энергией $\hbar\omega_{pl}$. Эти кванты иногда называют плазмонами.

Отметим, что кулоновское взаимодействие между носителями заряда (в том числе между электронами и дырками) все же не полностью учитывается с помощью самосогласованного поля. В результате газ носителей заряда оказывается неидеальным. Связанные с этим эффекты называются *корреляционными*. Один из них рассматривается в § 7.

Поляроны также представляют собой элементарные возбуждения, возникающие благодаря кулоновскому взаимодействию свободных электронов и дырок с колеблющимися ионами решетки.

До тех пор, пока мы не рассматриваем явлений, существенно связанных с неидеальностью газа носителей заряда, и не вычисляем параметры энергетического спектра, разница между «невзаимодействующими электронами» зонной теории и «квазичастицами» многоэлектронной теории носит словесный характер. Можно пользоваться и тем и другим языком. Посмотрим, как на этих двух языках описываются одни и те же явления — поглощение и испускание света в беспримесном полупроводнике. Для наглядности изобразим это в виде таблицы. В средней части таблицы (в виде подзаголовков) будем указывать то или иное состояние (или процесс) или наблюдаемую величину. Слева дается их описание на языке одноэлектронной теории, справа — на языке элементарных возбуждений.

Таблица 17.1

Язык одноэлектронной теории		Язык элементарных возбуждений
	Основное состояние полупроводника	
Зона проводимости пуста, валентная зона полностью заполнена		Нет ни электронов проводимости, ни дырок
	Ширина запрещенной зоны	
Наименьшее энергетическое расстояние между нижним краем зоны проводимости и верхним краем валентной зоны		Минимальная работа, необходимая для создания пары «электрон проводимости — дырка»
	Возбужденное состояние полупроводника	
Зона проводимости частично заполнена, в валентной зоне имеются вакантные места		Имеется некоторое количество электронов проводимости и дырок. Энергии их изменяются, соответственно, в пределах зоны проводимости и дырочной
	Поглощение светового кванта	
Переход электрона из валентной зоны в зону проводимости		Образование пары «электрон проводимости — дырка»
	Рекомбинационное излучение	
Переход электрона из зоны проводимости в валентную с испусканием светового кванта		Аннигиляция электрона проводимости и дырки с испусканием светового кванта

Таким же образом можно рассматривать и задачи с участием примеси и т. д. Эквивалентность двух языков имеет место до тех пор, пока мы интересуемся только описанием процесса, но не расчетом его вероятности. Так, например, при одноэлектронном подходе теряется эффект взаимодействия между электроном проводимости и дыркой, образующимися при поглощении светового кванта. Вместе с тем, как мы увидим в следующем параграфе и в гл. XVIII, он может оказаться существенным.

§ 7. Экситон

В ряде веществ (V_2O_5, некоторые молекулярные кристаллы) поглощение света вблизи границы собственной полосы не сопровождается фотопроводимостью (фотоэлектрически неактивно). В других веществах (Cu_2O, CdS) поглощение света с частотой, близкой к $E_g/\hbar$, приводит к фотопроводимости, но величина фототока оказывается пропорциональной концентрации примеси.

Для объяснения этих явлений Я. И. Френкелем и Р. Пайерльсом было введено представление о нейтральной квазичастице — *экситоне* *). Экситон представляет собой связанное состояние электрона и дырки, возникающее в результате кулоновского притяжения их друг к другу. Образовав экситон, электрон и дырка движутся по кристаллу совместно — как одно целое.

Обе указанные выше группы фактов можно объяснить, предположив, что поглощение света в рассматриваемой области частот приводит к образованию экситона. Действительно, будучи в целом нейтральным, экситон не участвует в переносе тока. При столкновении с атомом примеси может произойти ионизация экситона с образованием двух свободных носителей заряда. При больших энергиях световых квантов становятся возможными и переходы с образованием сразу пары свободных носителей заряда.

Рассчитаем энергетический спектр экситона, пользуясь простейшей моделью полупроводника. Именно, будем считать, что минимальные значения энергии электрона и дырки (нижние края соответствующих зон) лежат в центре зоны Бриллюэна, причем законы дисперсии вблизи этих экстремумов имеют вид

$$E_n(\mathbf{p}) = E_g + \frac{p_n^2}{2m_n}, \qquad E_p(\mathbf{p}) = \frac{p_p^2}{2m_p}. \tag{7.1}$$

Здесь E_n и E_p, $\mathbf{p}_n$ и $\mathbf{p}_p$ — энергии и квазиимпульсы электрона и дырки; за нуль энергии принят потолок валентной (т. е. дно дырочной) зоны.

Предположим также (это будет оправдано последующим расчетом), что размер экситона (среднее расстояние между взаимосвязанными электроном и дыркой) велик по сравнению с постоянной решетки. Тогда для описания экситона мы вправе воспользоваться методом эффективной массы. Соответствующее уравнение Шредингера имеет вид

$$-\frac{\hbar^2}{2m_n}\nabla_n^2\psi - \frac{\hbar^2}{2m_p}\nabla_p^2\psi - \frac{e^2}{\varepsilon\,|\mathbf{r}_n - \mathbf{r}_p|}\psi + E_g\psi = E\psi. \tag{7.2}$$

Здесь $\mathbf{r}_n$ и $\mathbf{r}_p$ — радиус-векторы электрона и дырки, ∇_n^2 и ∇_p^2 — лапласианы по координатам электрона и дырки, $\psi(\mathbf{r}_n, \mathbf{r}_p)$ — волновая функция рассматриваемой системы двух частиц, E — соответствующее

*) От слова «excitation» — возбуждение.

собственное значение энергии, отсчитываемое от энергии основного состояния. Через ε обозначена диэлектрическая проницаемость вещества. Строго говоря, надо было бы взять ее значение при частоте $\omega_э = \frac{|E_g - E|}{\hbar}$; во многих полупроводниках, однако, $\omega_э$ лежит далеко от области дисперсии и под ε можно понимать статическую диэлектрическую проницаемость.

Предельный переход к зонной теории получится, если вычеркнуть третье слагаемое в левой части (7.2). При этом функция ψ будет иметь вид

$$\Psi \equiv \psi_0 = A \exp\left[\frac{i}{\hbar}(\mathbf{p}_n, \mathbf{r}_n) + \frac{i}{\hbar}(\mathbf{p}_p, \mathbf{r}_p)\right], \tag{7.3}$$

где A — нормировочный множитель. Соответствующее значение E дается суммой выражений (7.1), при этом $E \geqslant E_g$.

При учете взаимодействия электрона и дырки положение меняется. Решение уравнения (7.2) и в этом случае получается без труда: надо лишь заметить, что (7.2) совершенно идентично уравнению Шредингера для атома водорода, причем m_p и m_n отвечают, соответственно, массе ядра и массе электрона, а разность $E - E_g$ — собственному значению энергии атома.

Введем относительные координаты $\mathbf{r}$ и координаты центра инерции $\mathbf{R}$, полагая

$$\mathbf{r} = \mathbf{r}_n - \mathbf{r}_p, \qquad \mathbf{R} = \frac{m_n \mathbf{r}_n + m_p \mathbf{r}_p}{m_n + m_p}. \tag{7.4}$$

Обозначим также через

$$m = m_n + m_p, \qquad m_r = \frac{m_n m_p}{m_n + m_p}$$

полную и приведенную массы системы. Тогда уравнение (7.2) примет вид

$$-\frac{\hbar^2}{2m}\nabla_R^2\Psi - \frac{\hbar^2}{2m_r}\nabla_r^2\Psi - \frac{e^2}{\varepsilon r}\Psi = (E - E_g)\Psi. \tag{7.2'}$$

С точностью до нормировочного множителя решение уравнения (7.2′) есть

$$\psi(\mathbf{r}, \mathbf{R}) = e^{\frac{i}{\hbar}(\mathbf{pR})}\chi(\mathbf{r}), \tag{7.5}$$

где $\chi(\mathbf{r})$ — новая неизвестная функция, $\mathbf{p}$ — вектор с вещественными компонентами. Для $\chi(\mathbf{r})$ получается уравнение

$$-\frac{\hbar^2}{2m_r}\nabla_r^2\chi - \frac{e^2}{\varepsilon r}\chi = \lambda\chi, \tag{7.6}$$

причем

$$\lambda = E - E_g - \frac{p^2}{2m}. \tag{7.7}$$

Первый сомножитель в правой части (7.5) описывает свободное движение экситона как целого. Вектор $\mathbf{p}$ есть квазиимпульс центра инерции экситона. Функция $\chi(\mathbf{r})$ описывает относительное движение электрона и дырки.

С точностью до замены $m_r \to m$ уравнение (7.6) совпадает с уже известным нам уравнением (IV.7.2), фигурировавшим в «водородной» модели примесного центра. Здесь имеются решения двух типов.

а. Решения, принадлежащие непрерывному спектру. Они существуют при $\lambda \geqslant 0$, т. е., согласно (7.7), при любых значениях E, удовлетворяющих неравенству

$$E \geqslant E_g + \frac{p^2}{2m}. \tag{7.8а}$$

Функция $\chi(\mathbf{r})$ при этом такова, что вероятность найти электрон и дырку на любом — сколь угодно большом — расстоянии друг от друга отлична от нуля. Это — аналог «зонного» решения (7.3): электрон и дырка не связаны друг с другом, а движутся независимо, испытывая лишь взаимное рассеяние. Мы имеем здесь пару свободных носителей заряда.

Энергии световых квантов, способных вызвать образование такой пары, определяются выражением (7.8а). При этом квазиимпульс $\mathbf{p}$ оказывается не произвольным: как будет показано в следующей главе (§ XVIII.5), при поглощении света идеальной решеткой сумма квазиимпульсов электрона и дырки (т. е. квазиимпульс центра их инерции) с точностью до вектора обратной решетки должна равняться импульсу светового кванта. Из-за большого значения скорости света импульс фотона обычно довольно мал (подробнее см. § XVIII.5), и приближенно можно положить $\mathbf{p} = 0$.

б. Решения, принадлежащие дискретному спектру. Они существуют при $\lambda < 0$. Возможные значения λ при этом даются формулой Бора

$$\lambda = -\frac{m_r e^4}{2\varepsilon^2 \hbar^2 n^2}, \tag{7.9}$$

где $n = 1, 2, \ldots$ — главное квантовое число.

Согласно (7.7) это означает, что

$$E = E_g + \frac{p^2}{2m} - \frac{m_r e^4}{2\varepsilon^2 \hbar^2 n^2} < E_g + \frac{p^2}{2m}. \tag{7.8б}$$

В частности, при $\mathbf{p} = 0$ (покоящийся экситон) мы получаем набор дискретных «водородных» уровней, отвечающих энергиям возбуждения, меньшим E_g.

Функции $\chi(\mathbf{r})$, принадлежащие собственным значениям (7.9), быстро (экспоненциально) убывают с расстоянием между электроном

и дыркой r. Например, низшему уровню ($n = 1$) соответствует волновая функция

$$\chi = \text{const} \cdot \exp(-r/a_э). \tag{7.10}$$

Здесь const — нормировочная постоянная, а

$$a_э = \varepsilon\hbar^2/m_r e^2 \tag{7.11}$$

— боровский радиус экситона.

Функция (7.10) описывает связанное состояние электрона и дырки — экситон. Наиболее вероятное расстояние между электроном и дыркой в этом состоянии равно $a_э$. При типичных значениях ε, m_n, m_p оно заметно превышает постоянную решетки, что и оправдывает принятый нами метод расчета.

Энергии световых квантов, при поглощении которых возникают экситоны рассматриваемого типа, определяются формулой (7.8б) (при $\mathbf{p} = 0$). В отличие от случая a они оказываются дискретными.

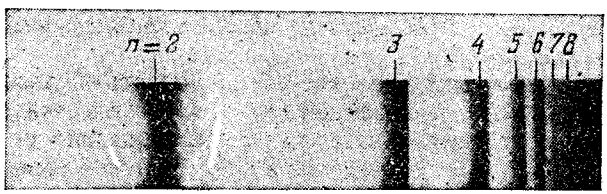


Рис. 17.1. Спектрограмма желтой серии экситона в Cu_2O при $T = 1{,}3$ К.

Таким образом, при частотах выше $E_g/\hbar$ должна возникать сплошная полоса поглощения, а при частотах ниже $E_g/\hbar$ должны быть видны дискретные «водородоподобные линии». Такие линии действительно наблюдались (впервые Е. Ф. Гроссом и его сотрудниками) в закиси меди и в других материалах (рис. 17.1) *).

Для полупроводников с более сложной зонной структурой утверждение о существовании экситонных состояний с энергиями возбуждения, меньшими E_g, остается в силе. Конкретная структура энергетического спектра экситонов, однако, усложняется. Рассмотрим, например, материал типа германия, в котором дно зоны проводимости и потолок валентной зоны отвечают разным точкам зоны Бриллюэна. При этом в той же точке (в данном случае — в центре) зоны Бриллюэна, где расположен потолок валентной зоны, имеется побочный минимум энергии в зоне проводимости (рис. 3.9). Здесь могут существовать два типа экситонов, обычно называемых «прямыми» и «непрямыми». Они образуются из дырки и электрона,

*) *Е. Ф. Гросс*, УФН **63**, 576 (1957).

принадлежащего либо побочному, либо главному минимуму. Соответственно усложняется и спектр поглощения света.

Мы рассмотрели связанные состояния «большого радиуса» ($a_э$ гораздо больше постоянной решетки). Об этих состояниях принято говорить как об экситонах Ваннье—Мотта. Они, по-видимому, реализуются в кристаллах типа германия и т. п. В материалах иной природы (в частности, в молекулярных кристаллах) существуют экситоны другого типа (обычно называемые экситонами Френкеля). Радиус их — порядка постоянной решетки, так что, в сущности, такой экситон есть не что иное, как возбужденное состояние отдельного атома (или молекулы), способное перемещаться по решетке *).

При не слишком малой ширине запрещенной зоны тепловая концентрация экситонов Ваннье—Мотта обычно очень мала. Заметную концентрацию экситонов $N_э$ можно получить, создавая их светом. Пользуясь лазером в качестве источника света, можно добиться значений $N_э$ порядка 10^{18} см$^{-3}$ и более. При этом среднее расстояние между экситонами, по порядку величины близкое к $N_э^{-1/3}$, становится сравнимым с боровским радиусом (7.11). При таких условиях экситоны уже нельзя рассматривать как независимые квазичастицы. Мы имеем здесь существенно неидеальный экситонный газ. При достаточно высокой плотности и достаточно низкой температуре газ экситонов, как и всякий другой газ, должен испытывать конденсацию. Опыт показывает, что при гелиевых температурах в полупроводнике с указанными выше значениями $N_э$ действительно появляются экситонные «капли», обнаруживаемые, например, по вызываемому ими рассеянию света [7]. Линейные размеры капель могут составлять несколько сот микрон и более.

§ 8. Мелкие локальные уровни при учете экранирования примесных центров

Взаимодействие между электронами приводит к экранированию неоднородных электрических полей, создаваемых как носителями заряда, так и атомами примеси, иными несовершенствами решетки или внешними источниками. Есть задачи (см., например, § XIV.5), в которых учет экранирования оказывается принципиально необходимым. В рамках одноэлектронной задачи эта процедура не вполне последовательна. Ее, однако, можно обосновать с позиций многоэлектронной теории твердого тела. При этом дается и способ вычисления потенциальной энергии носителя заряда δU с учетом эффекта экранирования. В частности, при достаточно большом радиусе экранирования r_0 получаются уже известные нам выражения (П.XII.6а и П.XII.6б): они справедливы, если $r_0^3 n \gg 1$, где n —

*) Для дальнейшего знакомства с этим кругом вопросов можно рекомендовать книги [5—7].

концентрация экранирующих носителей заряда. Для определения энергетического спектра электронов и дырок, рассматриваемых как элементарные возбуждения, получается уравнение Шредингера с экранированным потенциалом δU *).

Рассмотрим влияние эффекта экранирования на энергетический спектр локальных уровней, описываемых «водородной» моделью (§ IV.7). Соответствующее уравнение Шредингера имеет вид

$$-\frac{\hbar^2}{2m}\nabla^2\psi+\delta U\cdot\psi=E\psi. \tag{8.1}$$

Здесь, как и в § IV.7, энергия E отсчитывается от края зоны проводимости вверх (если речь идет об электронах) или от края валентной зоны вниз (если рассматриваются дырки). Для $\delta U(r)$ можно воспользоваться, например, выражением (П. XII.7):

$$\delta U=-\frac{e^2}{\varepsilon r}\exp(-r/r_0). \tag{8.2}$$

Для дальнейшего, однако, важен не явный вид функции δU, а лишь то, что она удовлетворяет условию

$$\left|\int \delta U(\mathbf{r})\,d\mathbf{r}\right|<\infty. \tag{8.3}$$

(Тогда говорят о короткодействующих силах.)

При $r_0\to\infty$ (что предполагалось в § IV.7) уравнение (8.1) имеет бесконечное число дискретных собственных значений ($E<0$), отвечающих связанным состояниям электрона на доноре (дырки на акцепторе). При конечном значении радиуса экранирования число дискретных уровней оказывается ограниченным; в частности, они могут и вообще отсутствовать. Дело в том, что в силу принципа неопределенности локализация электрона на малом расстоянии от «ядра» (атомного остова донора) невозможна. Действительно, пусть электрон может с заметной вероятностью отстоять от ядра лишь на расстояние, не превышающее некоторого значения $\bar{r}$. Величину $\bar{r}$ можно рассматривать как неопределенность радиальной координаты. Тогда неопределенность радиального импульса — порядка $\hbar/\bar{r}$. Средняя кинетическая энергия по порядку величины составит $\hbar^2/2m\bar{r}^2$, а полная энергия будет равна

$$E\simeq\frac{\hbar^2}{2m\bar{r}^2}-\frac{e^2}{\varepsilon\bar{r}}\exp\left(-\frac{\bar{r}}{r_0}\right). \tag{8.4}$$

Минимизируя правую часть (8.4) по $\bar{r}$, легко убедиться, что при $r_0\to\infty$ (атом водорода) $\bar{r}=a_B$, а $E=E_B$. Видно, далее, что при достаточно малых значениях $\bar{r}$ энергия E оказывается положительной. Это означает, что столь сильно локализованные состояния существо-

*) Это относится и к теории экситона. Уравнение (7.2) справедливо лишь в условиях слабого экранирования, когда $r_0\gg a_9$.

вать не могут. С другой стороны, если радиус экранирования мал, то притяжение электрона к донору на больших расстояниях от последнего практически исчезает. Отсюда явствует, что уровни, которые в отсутствие экранирования были бы «водородными», должны исчезнуть, когда величина r_0 становится того же порядка, что и радиус соответствующей орбиты. В частности, при $r_0 \sim \varepsilon\hbar^2/me^2$ дискретные уровни должны исчезнуть вообще.

Согласно П.XII.6а и П.XII. 6б радиус экранирования зависит от температуры и от концентрации экранирующих носителей заряда. Последняя может быть и неравновесной: инжектированные носители также принимают участие в экранировании, если их время жизни больше максвелловского времени релаксации. Таким образом, энергетическим спектром примесных уровней можно в известных пределах управлять, варьируя условия опыта. Исчезновение дискретных уровней в условиях сильной инжекции наблюдалось экспериментально: при увеличении уровня инжекции исчезали линии рекомбинационного излучения, связанные с переходами «донорный уровень — валентная зона».

§ 9. Механизмы рекомбинации

Как было показано в гл. IX, для описания рекомбинационных процессов необходимо задать коэффициент прямой междузонной рекомбинации α, значения энергий и факторов вырождения центров рекомбинации, а также по два коэффициента захвата α_n и α_p для каждого из уровней этих центров. Как видно из формул (IX.3.3) и (IX.4.3), коэффициенты захвата определяются вероятностями соответствующих процессов, а также видом функции распределения $f(\mathbf{p})$ в зонах проводимости и дырочной. В отсутствие равновесия по энергиям функция f отличается от фермиевской, чем и описывается возможное влияние нагрева носителей заряда на темп рекомбинации. В этом параграфе мы обсудим, какими физическими факторами определяются значения коэффициентов захвата и вид температурной их зависимости.

Важная роль ловушек в процессе рекомбинации определяется двумя факторами. Во-первых, захватываясь, например, из зоны проводимости на ловушку, электрон должен передать своему окружению в решетке меньшую энергию, нежели при непосредственном переходе в валентную зону. По этой причине вероятность первого процесса может оказаться заметно больше, чем второго.

Во-вторых, при захвате электрона или дырки на ловушку закон сохранения квазиимпульса может и не выполняться. Действительно, носитель заряда взаимодействует при этом не только с идеальной решеткой, но и с атомом примеси или иным структурным дефектом, образующим ловушку. По этой причине потенциальная его энергия уже не обладает свойством пространственной периодичности. Волно-

вая функция электрона на ловушке не имеет вида (III.2.15). Она локализована в пространстве и электрон в таком состоянии не характеризуется определенным квазиимпульсом. Неопределенность квазиимпульса особенно велика для глубоких ловушек, захватившись на которые носители заряда оказываются локализованными в сравнительно небольшой области пространства. С другой стороны, при малой силе взаимодействия с ловушкой, когда энергия ионизации ее невелика, область локализации расширяется, а неопределенность квазиимпульса Δp уменьшается. Какие именно значения Δp следует считать большими, а какие — малыми, зависит от характера задачи. При рассмотрении захвата электрона (или дырки) Δp надо сравнивать с «тепловым» квазиимпульсом p_T (если газ свободных носителей заряда не вырожден).

Как видно из рассуждений предыдущего параграфа, которые привели нас к формуле (8.4), линейный размер области локализации по порядку величины близок к $\hbar/\sqrt{m|E|}$, а неопределенность импульса — к $\sqrt{m|E|}$*); под $|E|$ здесь надо понимать энергию ионизации данного локального уровня (для возбужденных уровней эта величина может быть заметно меньше энергии ионизации ловушки, определяемой как $E_c - E_t$ или — для дырки — как $E_t - E_v$).

Таким образом, при $|E| \ll kT$ неопределенность квазиимпульса мала, а при $|E| \gg kT$ — велика. В первом случае при рассмотрении процесса захвата электрон на ловушке можно рассматривать как «почти свободный» и закон сохранения квазиимпульса должен приближенно выполняться. Во втором случае даже приближенное сохранение квазиимпульса не имеет места. С точки зрения увеличения вероятности захвата это выгодно, так как теперь захват может происходить из любой точки зоны Бриллюэна.

Вероятность захвата на ловушку определяется начальным и конечным состояниями носителей заряда, а также способом отвода энергии, высвобождаемой при захвате. Иначе говоря, чтобы вычислить коэффициенты α_n и α_p, надо задать:

а) волновую функцию системы «решетка + свободный носитель заряда»;

б) волновую функцию системы «решетка + захваченный носитель»;

в) вид оператора энергии взаимодействия носителя заряда с его окружением в решетке.

Для достаточно мелких ловушек волновые функции носителей заряда можно найти с помощью метода эффективной массы (§ IV.4). В задаче о рекомбинации, однако, часто бывают интересны более глубокие центры захвата — с энергиями ионизации, сравнимыми с полушириной запрещенной зоны. Теория таких центров в настоя-

*) В применении к глубоким ловушкам эта оценка имеет лишь ориентировочное значение, ибо методом эффективной массы там пользоваться нельзя.

щее время еще не завершена. По этой причине не всегда удается и полностью рассчитать коэффициенты захвата α_n и α_p. Все же, как мы увидим, температурная их зависимость в ряде важных случаев может быть установлена.

Пункт в) составляет, по существу, часть общей задачи об энергетическом спектре полупроводника. Действительно, в принципе энергия, высвобождаемая при захвате носителя заряда, может быть передана любым элементарным возбуждениям системы «электроны + решетка». В зависимости от конкретной природы этих возбуждений говорят о том или ином механизме рекомбинации (или механизме захвата). Как и при переходах зона—зона (§ IX.1), наиболее интересными здесь оказываются процессы с передачей энергии фотонам, фононам и другим носителям заряда (излучательный, фононный и ударный механизмы).

Механизм захвата существенно зависит от зарядового состояния ловушки до захвата (заряженная или нейтральная) и от знака заряда ловушки относительно захватываемой частицы (имеется ли кулоновское притяжение захватываемой частицы или ее отталкивание). Рассмотрим эти случаи порознь.

При захвате с притяжением сечения захвата очень велики. Этот случай мы имеем, например, при захвате дырок на акцепторные центры Cu^-, Cu^{--}, Au^- и Au^{--} в германии. Сечения захвата для дырок у таких центров при комнатных температурах обычно лежат в пределах $10^{-16} \div 10^{-13}$ см^2 и увеличиваются при понижении температуры. Измерения интенсивности рекомбинационного излучения показывают, что фотоны уносят только очень малую часть освобождаемой энергии, так что переходы оказываются в основном безызлучательными.

Наблюдаемый порядок величины сечений и их температурную зависимость можно хорошо объяснить в предположении, что энергия передается колебаниям решетки, однако с помощью своеобразного процесса. Дело в том, что энергия, освобождаемая при захвате, обычно лежит в пределах $0{,}1 \div 1$ эВ. Это заметно превышает максимальную энергию фонона, благодаря чему для высвобождения всей энергии необходимо участие нескольких фононов. Вероятность одновременного их испускания может оказаться небольшой. Как показывают расчеты, таким путем не всегда удается объяснить наблюдаемые величины сечений. Напротив, ряду экспериментальных данных хорошо соответствует предложенный М. Лэксом каскадный механизм захвата, в котором фононы испускаются последовательно во времени и по одному. Рис. 17.2 поясняет этот процесс. В случае притягивающих сил рассматриваемый примесный атом создает, кроме основного, еще и несколько возбужденных уровней (подобно, например, уровням электрона в атоме водорода). Поэтому электрон из зоны может быть сначала захвачен на один из высоких возбужденных уровней, так что освобождаемая энергия может быть унесена

только одним фононом (переход *1*). Из этих возбужденных состояний часть электронов выбрасывается обратно в зону (переход *2*), но часть из них переходит на более низкие уровни энергии, опять с испусканием только одного фонона (переход *3*). Опускаясь таким образом по цепочке возбужденных уровней, электрон попадает на самый глубокий уровень основного состояния. Отметим, что для наиболее глубоких возбужденных уровней однофононные процессы могут быть и невозможны, так как расстояние между уровнями увеличивается по мере возрастания их глубины (см. рис. 17.2). Однако это уже не играет никакой роли, так как, попав на глубокие возбужденные уровни, электрон уже фактически захвачен и в конце концов (хотя, быть может, и с большим временем жизни в возбужденном состоянии) попадает на основной уровень. Из сказанного также видно, что коэффициенты захвата должны увеличиваться при понижении температуры, так как при этом уменьшается вероятность обратного выброса электронов с возбужденных уровней. Явный вид этой зависимости оказывается степенным. Так, при захвате дырок отрицательно заряженными примесными центрами в германии и кремнии

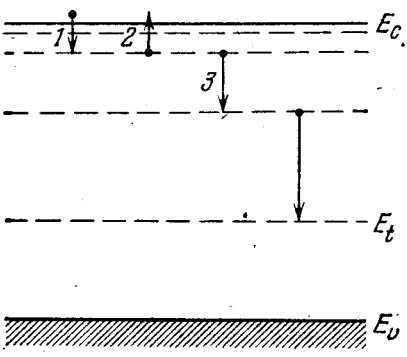


Рис. 17.2. Схема каскадного процесса захвата электрона на притягивающую ловушку.

$$\alpha_p \sim T^{-m}. \tag{9.1}$$

Число m в этой формуле меняется в пределах от $\simeq 1$ до $\simeq 5$ в зависимости от природы материала и примеси.

При фононном механизме захвата коэффициенты захвата α_n и α_p не должны зависеть от концентрации электронов и дырок. Опыт показывает, что во многих полупроводниках это действительно так. Однако, когда концентрация электронов n (или, соответственно, дырок) становится очень большой, α начинает увеличиваться с возрастанием n приблизительно пропорционально n. Так, например, в германии при комнатных температурах α начинает зависеть от n при $n \gtrsim 10^{17}$ см$^{-3}$. Это показывает, что при больших концентрациях электронов и дырок основным механизмом захвата становится уже не фононный, а ударный. Схема ударного механизма захвата неосновного носителя заряда на ловушку показана на рис. 17.3 для полупроводника n-типа (*а*) и p-типа (*б*). Как и раньше, в обоих случаях по вертикали, снизу вверх, отложена полная энергия электрона. В случае *а*) (захват дырки) электрон переходит с уровня

ловушки E_t на один из пустых уровней вблизи края валентной зоны E_v. Выделяемая при этом энергия, равная приблизительно $(E_t - E_v)$, передается одному из основных носителей — электрону (которых много), и последний возбуждается на более высокий уровень энергии E' в зоне проводимости. В случае *б*) электрон переходит

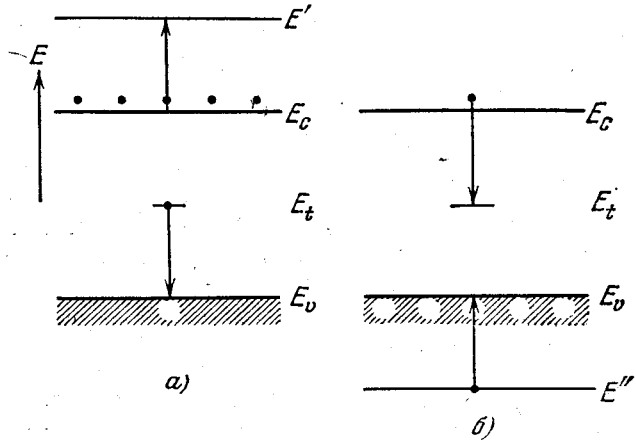


Рис. 17.3. Схема ударного процесса захвата неосновного носителя заряда на ловушку: *а*) захват дырки в полупроводнике *n*-типа, *б*) захват электрона в полупроводнике *p*-типа.

с одного из уровней вблизи E_c на уровень E_t, а выделяемая энергия $(E_c - E_t)$ расходуется на ускорение свободной дырки, переходящей на уровень E''.

Процессы захвата имеют другой характер, когда ловушка несет заряд того же знака, что и захватываемая частица. Здесь на больших расстояниях на частицу действует кулоновская сила отталкивания, а на малых расстояниях преобладают более короткодействующие силы притяжения. Поэтому потенциальная энергия частицы в поле ловушки $V(r)$ имеет вид, показанный на рис. 16.6, и, чтобы частица попала в область сил притяжения и была захвачена, она должна преодолеть потенциальный барьер вокруг центра.

Отсюда можно было бы ожидать, что сечения захвата у отталкивающих ловушек будут на много порядков меньше, чем у незаряженных и притягивающих ловушек, и что они должны быстро уменьшаться при понижении температуры (как $\exp(-V_0/kT)$).

Опыт показывает, что сечения захвата притягивающих центров, как правило, действительно больше сечений отталкивающих центров. Однако при сравнении отталкивающих и нейтральных центров оказывается, что в ряде случаев сечения S_n и S_n^0 имеют одинаковый порядок величины (здесь нижний индекс указывает, какая частица

захватывается, а верхний индекс — зарядовое состояние центра до захвата). Это наблюдается, например, в германии для атомов Au, Cu и Ni при комнатных температурах. При понижении температуры сечения захвата, как и можно было ожидать, уменьшаются, однако гораздо слабее, чем следует из классических представлений.

Такое сравнительно слабое влияние потенциального барьера объясняется квантовомеханическим туннельным эффектом. Так как толщина потенциального барьера сопоставима с длиной волны электронов, то достигают центра не только те электроны, энергия которых $E > V_0$, но и электроны с меньшими энергиями, $E < V_0$, что ослабляет влияние барьера как на величину сечений захвата, так и на их температурную зависимость. Последняя (в случае захвата электронов отрицательно заряженными атомами золота и меди в германии) удовлетворительно описывается формулой

$$\alpha_n \sim \exp\left[-(T_0/T)^{1/3}\right]. \tag{9.2}$$

Здесь

$$T_0 = \frac{27\pi^2 m e^4 Z^2}{2\varepsilon^2 \hbar^2 k}, \tag{9.3}$$

Z — заряд отталкивающего центра в единицах e, а эффективная масса $m \simeq 0{,}2m_0$ (в Ge).

Однако и здесь энергия, освобождаемая при захвате, должна быть отведена от центра, чтобы частица могла остаться на центре.

В случае нейтральных и тем более отталкивающих центров наиболее эффективный каскадный фононный процесс, по-видимому, невозможен, так как нет оснований считать, что такие центры имеют нужный богатый спектр возбужденных уровней. Поэтому остаются в первую очередь многофононный и излучательный процессы. Вероятность многофононного перехода оказывается конечной благодаря отмечавшемуся в § 3 изменению положений равновесия и собственных частот нормальных колебаний при захвате (или обратном выбросе) носителя заряда. Действительно, при этом формулы (XII.5.15а) и (XII.5.15б) уже неприменимы, поскольку функции φ_n, $\varphi_{n'}$ в начальном и конечном состояниях принадлежат разным осцилляторам. Интегралы $\mathcal{J}^{\pm}_{nn'}$ (XII.5.12) оказываются отличными от нуля при любых значениях разности $n - n'$, что и означает возможность испускания (или, при обратном выбросе, поглощения) любого числа фононов. Все же вероятность процесса заметно убывает с увеличением необходимого числа фононов $|n - n'|$ *). По этой причине в зависимости от глубины ловушки и условий опыта главную роль может играть либо фононный, либо излучательный механизм захвата. Если при захвате электрона из зоны проводимости освобождаемая энергия $(E_c - E_t)$ может быть унесена неболь-

*) Обзор теории многофононных переходов можно найти в статье [8].

шим числом фононов, то доминирующим процессом может быть безызлучательный фононный захват. Если, напротив, для этого требуется много фононов, то вероятность фононного механизма уменьшается и тогда основным процессом может стать излучательный захват. Так обстоит дело, например, в кремнии при захвате электронов на нейтральные атомы акцепторов III группы (бор, галлий) при низких температурах. Энергетические уровни этих акцепторов расположены вблизи края валентной зоны, и при захвате на них электрона освобождается большая энергия, равная приблизительно ширине запрещенной зоны кремния $E_g \simeq 1$ эВ, которая намного больше максимальной энергии фононов в кремнии. Поэтому такие захваты оказываются главным образом излучательными.

Глава XVIII

ОПТИКА ПОЛУПРОВОДНИКОВ

§ 1. Поглощение и испускание света полупроводниками. Феноменологические соотношения

В опытах по поглощению света полупроводниками часто используются сравнительно слабые световые потоки. При этом электромагнитная волна не изменяет энергетический спектр носителей заряда (или решетки), а лишь создает новые пары электрон—дырка (или новые фононы) или вызывает перераспределение носителей заряда по состояниям. При этом величины, характеризующие оптические свойства среды, не зависят от интенсивности света. В таком случае говорят о линейном приближении: величина световой энергии, поглощаемой в образце, линейно связана с интенсивностью света. Ограничимся здесь этим приближением. Будем считать также, что длина электромагнитной волны значительно превышает постоянную решетки. Последнее условие обычно хорошо выполняется вплоть до энергий фотонов порядка нескольких сот электронвольт.

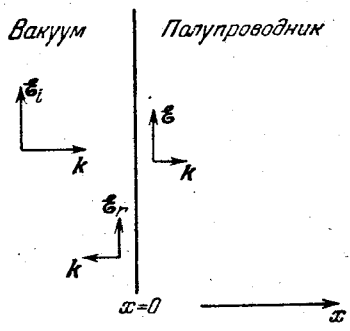


Рис. 18.1. Падающая, прошедшая и отраженная волны.

Опыты, нас здесь интересующие, сводятся в конечном счете к измерению интенсивности света, прошедшего через образец или отраженного от него. Для описания экспериментальных результатов, относящихся к кристаллам кубической симметрии (или к изотропным материалам), вводят две величины: показатели преломления n и поглощения $\varkappa$. Чтобы связать их с микроскопическими характеристиками вещества, рассмотрим задачу о распространении плоской электромагнитной волны, нормально падающей на поверхность образца. Пусть последняя совпадает с плоскостью $x = 0$, причем область $x > 0$ занята полупроводником (рис. 18.1). Размеры образца во всех направлениях будем считать сколь угодно большими.

Обозначим через $\mathfrak{E}$, $\mathfrak{D}$ и $\mathfrak{H}$, $\mathfrak{B}$ векторы напряженности и индукции электрического и магнитного полей электромагнитной

волны. Уравнения Максвелла, описывающие распространение поперечной волны, имеют вид

$$\operatorname{rot} \mathfrak{E} = -\frac{1}{c}\frac{\partial \mathfrak{B}}{\partial t}, \tag{1.1}$$

$$\operatorname{rot} \mathfrak{H} = \frac{4\pi}{c}\mathbf{j} + \frac{1}{c}\frac{\partial \mathfrak{D}}{\partial t}, \tag{1.2}$$

$$\operatorname{div} \mathfrak{D} = 0, \tag{1.3}$$

$$\operatorname{div} \mathfrak{B} = 0. \tag{1.4}$$

Ограничиваясь кристаллами кубической симметрии, можем положить

$$\mathfrak{D} = \varepsilon_0 \mathfrak{E}, \quad \mathfrak{B} = \mu \mathfrak{H}, \quad \mathbf{j} = \sigma \mathfrak{E}, \tag{1.5}$$

причем, в соответствии с § XIII.8, $\sigma = \sigma_1 + i\sigma_2$, $\varepsilon_0 = \varepsilon_1 + i\varepsilon_2$, где σ_1, σ_2, ε_1, ε_2 — вещественные величины, зависящие от частоты падающей волны ω. Магнитную проницаемость μ будем считать вещественной константой, не зависящей от ω; в немагнитных полупроводниках значение μ обычно очень близко к единице.

Возьмем ротор от обеих частей уравнения (1.1). Как известно из векторного анализа,

$$\operatorname{rot}\operatorname{rot} \mathfrak{E} = -\nabla^2 \mathfrak{E} + \operatorname{grad}\operatorname{div} \mathfrak{E}.$$

Пользуясь этим соотношением и равенствами (1.2), (1.3) и (1.5), мы получаем

$$\nabla^2 \mathfrak{E} = \frac{\mu}{c}\frac{\partial}{\partial t}\left(\frac{4\pi}{c}\sigma \mathfrak{E} + \frac{\varepsilon_0}{c}\frac{\partial \mathfrak{E}}{\partial t}\right). \tag{1.6}$$

Такому же уравнению удовлетворяет и вектор $\mathfrak{H}$.

В соответствии с постановкой задачи положим (при $x \geqslant 0$)

$$\mathfrak{E} = \mathfrak{E}_m \xi e^{-i\omega t + ikx}, \tag{1.7}$$

где $\mathfrak{E}_m$ — амплитуда волны, прошедшей в образец, при $x = 0$; ξ — единичный вектор в направлении $\mathfrak{E}$, k — комплексное волновое число. Из физических соображений ясно, что в поглощающей среде мнимая часть его должна быть положительна: она описывает затухание волны по мере углубления ее в среду. Подставляя (1.7) в (1.6), получим

$$k^2 \mathfrak{E} = \left(\frac{4\pi i\omega\mu}{c^2}\sigma + \frac{\varepsilon_0 \mu \omega^2}{c^2}\right)\mathfrak{E}. \tag{1.8}$$

Это уравнение, совместно с условием $\mathfrak{E}_m \neq 0$, дает связь между k и ω, т. е. закон дисперсии электромагнитной волны в рассматриваемой среде.

Удобно ввести комплексную проводимость σ', полагая

$$\sigma' = \sigma_1' + i\sigma_2' = \sigma - \frac{i\omega}{4\pi}\varepsilon_0; \tag{1.9}$$

при этом

$$\sigma_1' = \sigma_1 + \frac{\omega\varepsilon_2}{4\pi}, \quad \sigma_2' = \sigma_2 - \frac{\omega\varepsilon_1}{4\pi}. \tag{1.9'}$$

Положим

$$k = \frac{\omega}{c}(n + i\varkappa), \tag{1.10}$$

где n и $\varkappa$ — безразмерные вещественные положительные величины. Это и есть соответственно показатели преломления и поглощения. Смысл названий становится ясным при подстановке выражения (1.10) в (1.7). Мы получаем при этом

$$\mathfrak{E} = \mathfrak{E}_m \xi \exp\left[-i\omega\left(t - \frac{n}{c}x\right) - \frac{\omega}{c}\varkappa x\right]. \tag{1.7'}$$

При $\varkappa = 0$, т. е. в отсутствие поглощения, выражение (1.7') описывало бы плоскую волну, распространяющуюся с фазовой скоростью c/n и постоянной амплитудой. Если же $\varkappa \neq 0$, то амплитуда волны экспоненциально убывает по мере проникновения ее в образец.

Вместо показателя поглощения часто вводят линейный коэффициент поглощения γ, определяемый равенством

$$\gamma = 2\frac{\omega}{c}\varkappa. \tag{1.11}$$

Смысл величины γ легко понять, вспоминая, что плотность энергии электромагнитной волны пропорциональна квадрату амплитуды последней. Согласно (1.7') это означает, что плотность энергии (и тем самым число фотонов в единице объема) убывает с ростом координаты x, как $\exp\left(-2\frac{\omega}{c}\varkappa x\right) = \exp(-\gamma x)$. Таким образом, γ^{-1} есть длина, на которой плотность энергии волны в результате поглощения убывает в e раз.

Обращая равенство (1.11), получим

$$\varkappa = \frac{\gamma\lambda}{4\pi}, \tag{1.11'}$$

где $\lambda = 2\pi\frac{c}{\omega}$ есть длина световой волны в вакууме.

Подставим выражение (1.10) в равенство (1.8) и сократим последнее на $\mathfrak{E}_m$. Получим

$$n^2 - \varkappa^2 + 2in\varkappa = \frac{4\pi i\mu\sigma'}{\omega}. \tag{1.12}$$

Отделяя здесь вещественную часть от мнимой, находим два алгебраических уравнения для определения n и $\varkappa$. Корни их имеют вид

$$\varkappa=\left[\frac{2\pi\mu}{\omega}\left(\sigma_2'+\sqrt{\sigma_2'^2+\sigma_1'^2}\right)\right]^{1/2}, \tag{1.13а}$$

$$n=\left[\frac{2\pi\mu}{\omega}\left(-\sigma_2'+\sqrt{\sigma_2'^2+\sigma_1'^2}\right)\right]^{1/2}. \tag{1.13б}$$

Формулы (1.13а, б) решают поставленную задачу. Иногда их записывают по-другому, вводя вместо комплексной проводимости σ' комплексную диэлектрическую проницаемость

$$\varepsilon'=\frac{4\pi i}{\omega}\sigma'=\varepsilon_0+\frac{4\pi i}{\omega}\sigma=\varepsilon_1'+i\varepsilon_2'. \tag{1.14}$$

Здесь ε_1' и ε_2' — вещественные величины.

Очевидно,

$$\sigma_1'=\frac{\omega\varepsilon_2'}{4\pi}, \qquad \sigma_2'=-\frac{\omega\varepsilon_1'}{4\pi} \tag{1.15}$$

и, следовательно,

$$\varkappa=\left[\frac{\mu}{2}\left(-\varepsilon_1'+\sqrt{\varepsilon_1'^2+\varepsilon_2'^2}\right)\right]^{1/2}, \tag{1.16а}$$

$$n=\left[\frac{\mu}{2}\left(\varepsilon_1'+\sqrt{\varepsilon_1'^2+\varepsilon_2'^2}\right)\right]^{1/2}. \tag{1.16б}$$

Равенства (1.13а, б) и (1.16а, б) эквивалентны, и выбор тех или других есть дело вкуса.

Согласно (1.16а) показатель поглощения оказывается отличным от нуля, если отлична от нуля мнимая часть диэлектрической проницаемости. Последняя может быть обусловлена как носителями заряда, так и решеткой. В первом случае имеет место поглощение электромагнитной волны, связанное с различными электронными переходами, во втором — поглощение, связанное с передачей энергии непосредственно решетке, т. е. с генерацией только фононов. В обоих случаях, однако, расчет оптических характеристик вещества сводится к вычислению электропроводности $\sigma(\omega)$, определяемой равенством

$$\mathbf{j}(\omega)=\sigma(\omega)\,\mathcal{E}(\omega). \tag{1.17}$$

Здесь $\mathbf{j}(\omega)$ есть плотность электрического тока, возбуждаемого в веществе монохроматической электромагнитной волной вида (1.7).

Отметим два случая.

а. Непоглощающая среда. Пусть $\sigma_1=\varepsilon_2=0$. Тогда, согласно (1.16а, б), $\varkappa=0$ и

$$n=\sqrt{\mu\varepsilon_1'}. \tag{1.18}$$

Это есть формула Максвелла для показателя преломления, дополненная лишь учетом возможной частотной зависимости ε_1'.

б. Слабое поглощение волны достаточно большой частоты. Пусть ω больше плазменной частоты, определяемой равенством (XIII.9.4), т. е. $\varepsilon_1' > 0$. Пусть также

$$\varepsilon_1'^2 \gg \varepsilon_2'^2. \tag{1.19}$$

Тогда показатель преломления по-прежнему дается формулой (1.18), а показатель поглощения есть

$$\varkappa = \frac{\varepsilon_2' \mu^{1/2}}{2(\varepsilon_1')^{1/2}} = \frac{2\pi\sigma_1' \mu^{1/2}}{\omega(\varepsilon_1')^{1/2}}. \tag{1.20}$$

Линейный коэффициент поглощения равен теперь

$$\gamma = \frac{4\pi\sigma_1' \mu^{1/2}}{c(\varepsilon_1')^{1/2}}, \tag{1.21}$$

Смысл условия (1.19) можно выяснить, вспоминая, что длина электромагнитной волны в среде есть $\bar{\lambda} = 2\pi c/n\omega$. С помощью этого соотношения и равенств (1.18), (1.21) легко привести условие (1.19) к виду

$$\gamma^{-1} \gg \frac{\bar{\lambda}}{2\pi}. \tag{1.19'}$$

Иначе говоря, расстояние, на котором волна заметно поглощается, должно быть велико по сравнению с ее длиной.

Условие $\varepsilon_2' > 0$ и неравенство (1.19') выполняется во многих интересных случаях. При этом вклад свободных носителей заряда в диэлектрическую проницаемость образца обычно невелик, т. е. $\varepsilon_1' \simeq \varepsilon_1$. Небольшим оказывается обычно и решеточное поглощение в рассматриваемой области частот: $\varepsilon_2 \ll 1$. При этом формулу (1.21) можно переписать в виде (при $\mu \simeq 1$)

$$\gamma = \frac{4\pi\sigma_1}{c\varepsilon_0^{1/2}}. \tag{1.21'}$$

На опыте часто измеряют еще коэффициент отражения R. Последний определяется равенством

$$R = \frac{|\mathcal{E}_r|^2}{|\mathcal{E}_i|^2}. \tag{1.22}$$

Здесь $\mathcal{E}_i$ — амплитуды волны, падающей на образец, $\mathcal{E}_r$ — амплитуда отраженной волны (рис. 18.1). Пользуясь граничными условиями для компонент вектора $\mathcal{E}$ на поверхности образца, можно выразить R через показатели преломления и поглощения n и $\varkappa$. В случае нормального падения мы имеем [1]

$$R = \frac{(n-1)^2 + \varkappa^2}{(n+1)^2 + \varkappa^2}. \tag{1.23}$$

§ 2. Механизмы поглощения

Процессы поглощения света следует классифицировать по тому, на что непосредственно расходуется энергия поглощенных фотонов. Можно выделить следующие механизмы:

1) Решеточное поглощение: электромагнитная волна непосредственно возбуждает колебания решетки. Этот механизм поглощения особенно важен в ионных кристаллах, в которых генерация оптических фононов приводит к заметному изменению вектора поляризации; однако такое поглощение наблюдается и в гомеополярных материалах. Его испытывают волны, частоты которых близки к предельной частоте оптических фононов ω_0 (§ XII.3). (Обычно это соответствует энергии в несколько сотых долей электронвольта.)

2) Поглощение свободными носителями заряда: энергия расходуется на создание тока высокой (оптической) частоты и в конечном счете переходит в джоулево тепло.

3) Примесное поглощение: энергия поглощается носителями заряда, локализованными на примесных или иных структурных дефектах решетки. Она расходуется либо на перевод носителей с основного уровня примесного центра на возбужденный, либо на ионизацию примеси. В последнем случае электроны (дырки) попадают в зону проводимости (валентную), т. е. имеет место внутренний примесный фотоэффект. Таким путем можно определять энергии ионизации ряда примесей.

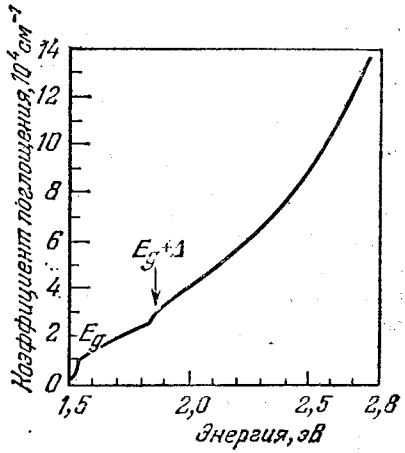


Рис. 18.2. Спектр поглощения пленки арсенида галлия.

4) Междузонное поглощение: энергия фотона расходуется на создание пары «электрон проводимости + дырка». В отсутствие сильного электрического поля и/или большой концентрации примеси этот тип поглощения опознается по наличию граничной частоты ω_m, близкой к $E_g/\hbar$. При $\omega < \omega_m$ поглощение этого типа отсутствует. Следует, однако, заметить, что вид спектра поглощения вблизи частоты $\omega = \omega_m$ в разных материалах оказывается различным. На рис. 18.2 представлен спектр поглощения арсенида галлия*). Видны два края поглощения. Первый из них приближенно

*) По данным работы *M. Sturge*, Phys. Rev. **127**, 768 (1962).

соответствует значению $\hbar\omega_m = E_g$*), второй отвечает энергии $E_g + \Delta$, где Δ — расстояние между потолком валентной зоны и верхним краем валентной подзоны, отщепленной из-за спин-орбитального взаимодействия.

Примерно такие же (в измененном масштабе частот) кривые получаются и при исследовании многих других материалов — антимонида и арсенида индия, антимонида галлия и др. С другой стороны, у ряда интересных полупроводников частотная зависимость

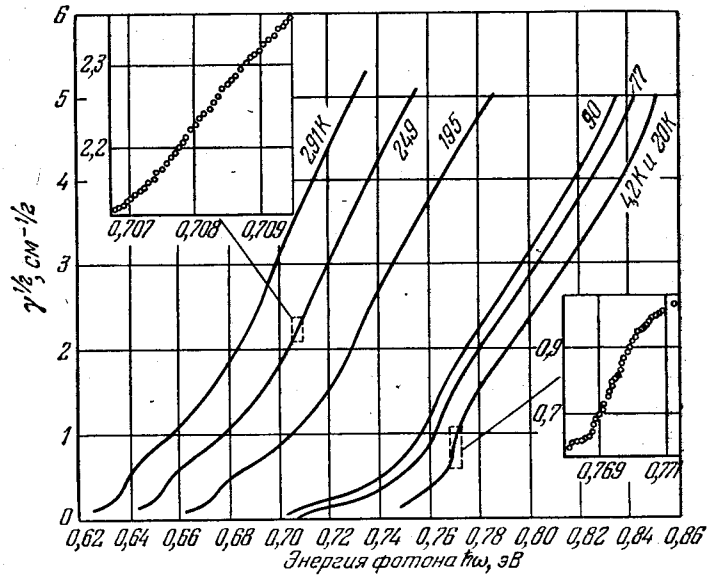


Рис. 18.3. Край поглощения германия при различных температурах.

и величина коэффициента поглощения вблизи красной границы оказываются существенно иными. Так, на рис. 18.3 изображен ход коэффициента поглощения света в германии при различных температурах**). При $\omega = \omega_m$ ($\hbar\omega_m = E_g = 0{,}66$ эВ при комнатной температуре) коэффициент поглощения относительно мал; он становится сравнимым с тем, что наблюдается в арсениде галлия, лишь при $\hbar\omega \simeq E_g + 0{,}1$ эВ. Похожая картина (в другом масштабе частот) наблюдается также в кремнии, фосфиде галлия и других материалах. Как будет видно из дальнейшего (§ 5, 7, 9), это различие

*) Как будет показано в § 4, вблизи красной границы коэффициент поглощения очень мал, поэтому поглощение становится заметным при несколько больших частотах.

) По данным работы *G. G. Macfarlane, T. P. McLean, J. E. Quarrington, V. Roberts,* Phys. Rev. **108, 1377 (1957).

имеет глубокую физическую природу: оно обусловлено тем, что в материалах первого типа экстремумы зон проводимости и валентной лежат в одной точке зоны Бриллюэна, а в материалах второго типа — в разных.

5) Экситонное поглощение: энергия фотона расходуется на образование экситона.

В материалах первого типа экситонному поглощению отвечают узкие пики γ при частотах, несколько меньших ω_m; в материалах второго типа вместо пиков наблюдаются «ступеньки».

§ 3. Поглощение и отражение электромагнитных волн газом свободных носителей заряда

Рассмотрим поглощение света свободными носителями заряда с изотропным законом дисперсии. Комплексная электропроводность такой системы вычислена в § XIII.8 с помощью кинетического уравнения Больцмана. Согласно (XIII.8.6) и (1.21′)

$$\gamma = \frac{4\pi e^2}{3c\varepsilon_0^{1/2}} \int_0^\infty \frac{N(E)\,\tau(E)\,v^2(E)}{1+\omega^2\tau^2(E)} (-f_0')\,dE. \tag{3.1}$$

Начало отсчета энергии здесь совмещено с нижней границей соответствующей зоны.

В отсутствие вырождения $-f_0' = \frac{1}{kT}\exp\frac{F-E}{kT}$ и, как и в статическом случае (§ XIII.7), правая часть (3.1) оказывается пропорциональной концентрации электронов n.

Коэффициент поглощения (3.1) обращается в нуль как при очень сильном рассеянии ($\tau \to 0$), так и в отсутствие его ($\tau \to \infty$). В первом случае носители заряда практически не могут свободно двигаться и образец, в сущности, представляет собой диэлектрик. Во втором случае функция распределения не успевает изменяться вместе с полем (§ XIII.8). В результате обращается в нуль компонента плотности тока, синфазная с полем, и поглощение энергии отсутствует.

Явный вид функции $\gamma(\omega)$ зависит, вообще говоря, от механизма рассеяния. Положение упрощается, если частота достаточно велика, так что в существенном интервале энергий время релаксации удовлетворяет условию

$$(\omega\tau)^2 \equiv \left(\frac{2\pi c}{\lambda}\tau\right)^2 \gg 1. \tag{3.2}$$

При этом равенство (3.1) принимает вид

$$\gamma = \frac{C}{\omega^2} \sim \lambda^2. \tag{3.3}$$

Здесь

$$C=\frac{4\pi e^2}{3c\varepsilon_0^{1/2}}\int_0^\infty \frac{v^2(E)\,N(E)}{\tau(E)}(-f_0')\,dE. \tag{3.4}$$

Как видно из вывода, в рассматриваемых условиях частотная зависимость коэффициента поглощения, определяемая формулой (3.3), должна иметь место при любом механизме рассеяния и при любой температуре. Последние факторы влияют только на величину коэффициента C.

Этот результат, однако, не всегда согласуется с опытом. Так, в n-Ge в области длин волн порядка 100 мкм и меньше коэффициент поглощения света свободными электронами пропорционален величине $\lambda^{3/2}$, если доминирует рассеяние квазиимпульса на акустических колебаниях решетки, и величине $\lambda^{7/2}$, если решающую роль играет рассеяние на заряженной примеси [3].

Причина этого расхождения состоит в невозможности пользоваться классическим кинетическим уравнением, если частота электромагнитной волны недостаточно мала. Действительно, согласно классической механике, на которой основано уравнение Больцмана, энергия, получаемая в единицу времени каждым отдельным носителем заряда от электромагнитного поля волны, есть $e(\mathbf{v}, \mathcal{E})$, где $\mathbf{v}$ — скорость данного носителя. При малой напряженности поля эта энергия может быть сколь угодно мала. На самом деле, однако, передача энергии происходит квантами величины $\hbar\omega$, которая никак не зависит от $\mathcal{E}$. Квадрат напряженности поля определяет лишь среднее число фотонов в единице объема вещества: это число пропорционально среднему по времени (за период волны) от классического значения плотности энергии электромагнитного поля*)

$$W=\frac{\varepsilon_1\mathcal{E}^2+\mu\mathcal{H}^2}{8\pi}. \tag{3.5}$$

Отсюда следует, что расчет коэффициента поглощения с помощью классических соображений может быть оправдан, лишь если энергия светового кванта мала по сравнению с характерной энергией носителя заряда: при этом «столкновение» электрона с фотоном оказывается почти упругим. В отсутствие фермиевского вырождения это означает, что частота волны должна удовлетворять неравенству

$$\hbar\omega \ll kT, \tag{3.6}$$

т. е.

$$\lambda T \gg 1{,}5 \text{ см}\cdot\text{град}. \tag{3.6'}$$

*) Мы считаем здесь $\varepsilon \simeq \varepsilon_1$; в интересующем нас круге вопросов это условие обычно хорошо выполняется, если частота ω не слишком близка к плазменной.

Здесь длина волны должна быть выражена в сантиметрах, а температура — в градусах Кельвина. При $T = 300$ К отсюда следует, что длина волны должна значительно превышать $4 \cdot 10^{-3}$ см $= 40$ мкм.

Коль скоро условие (3.6) не выполняется, возникает необходимость квантовомеханической трактовки поведения электронов. При этом можно по-прежнему пользоваться формулой (1.17); следует лишь вычислять плотность тока методами квантовой механики — как среднее значение соответствующего оператора. Иногда бывает удобнее пользоваться вытекающим из (1.17) выражением для средней энергии Q, получаемой носителями заряда от электромагнитной волны в единицу времени и в единице объема:

$$Q = \sigma \langle \mathcal{E}^2 \rangle = \frac{1}{2} \sigma_1 \mathcal{E}_m^2 . \tag{3.7}$$

Здесь угловые скобки обозначают усреднение по периоду поля.

Легко выразить Q через среднее число фотонов данной энергии $\rho(\hbar\omega)$, рассчитанное на единичный интервал энергии и на единицу объема. Действительно, среднее по времени значение плотности энергии (3.5) равно *)

$$\langle W \rangle = \frac{\varepsilon_1}{8\pi} \mathcal{E}_m^2 .$$

По определению эта же величина должна равняться

$$\hbar\omega \rho(\hbar\omega) \Delta(\hbar\omega),$$

где $\Delta(\hbar\omega)$ есть интервал, в котором заключены возможные значения энергии фотонов.

Таким образом,

$$\mathcal{E}_m^2 = \frac{8\pi}{\varepsilon_1} \hbar\omega \rho(\hbar\omega) \Delta(\hbar\omega) \tag{3.8}$$

и равенство (3.7) можно переписать в виде

$$Q = \frac{4\pi\sigma_1}{\varepsilon_1} \hbar\omega \rho(\hbar\omega) \Delta(\hbar\omega), \tag{3.7'}$$

или, с учетом (1.21'),

$$Q = \frac{\gamma c}{\varepsilon_1^{1/2}} \hbar\omega \rho(\hbar\omega) \Delta(\hbar\omega). \tag{3.9}$$

Очевидно, величину

$$\mathcal{P} = \frac{\gamma c}{\varepsilon_1^{1/2}} \tag{3.10}$$

*) Здесь принято во внимание, что в рассматриваемой нами плоской волне $\varepsilon_1 \mathcal{E}^2 = \mu \mathcal{H}^2$.

можно рассматривать как вероятность поглощения фотона данной частоты, отнесенную к единице времени *).

Расчет, выполненный квантовомеханическим путем **), позволяет успешно объяснить указанные выше экспериментальные данные: при $\hbar\omega \gtrsim kT$ сам вид частотной зависимости γ оказывается различным для разных механизмов рассеяния, причем получаются как раз те соотношения, которые наблюдаются на опыте. С другой стороны, в условиях (3.6) квантовомеханическое выражение, как и следовало ожидать, переходит в классическое (3.3).

Коэффициент отражения электромагнитной волны от вещества со свободными носителями заряда можно найти по формуле (1.23). Здесь особенно интересен случай, когда частота излучения близка к плазменной ω_{pl} (§ XIII.9), а поглощение сравнительно невелико ($\varepsilon_2' \ll 1$). Тогда формулы (1.16а, б) и (1.14) дают

$$\varkappa \ll 1, \qquad n \simeq \left[\mu\left(\varepsilon_1 - \frac{4\pi}{\omega}\sigma_2\right)\right]^{1/2}. \tag{3.11}$$

Для σ_2 в рассматриваемых условиях можно воспользоваться выражением (XIII.8.14), выразив его правую часть через плазменную частоту по формуле (XIII.9.4). Получим

$$n = \left[\mu\varepsilon_1\left(1 - \frac{\omega_{pl}^2}{\omega^2}\right)\right]^{1/2}. \tag{3.12}$$

Видим, что при $\omega = \omega_{pl}$ показатель преломления обращается в нуль, а при

$$\omega = \omega_{pl}\left[\frac{\varepsilon_1\mu}{\varepsilon_1\mu - 1}\right]^{1/2} \equiv \omega_1 \tag{3.13}$$

— в единицу. Соответственно коэффициент отражения оказывается близким к единице ($\omega = \omega_{pl}$) или к нулю ($\omega = \omega_1$).

Во многих полупроводниках величина $\mu\varepsilon_1$ значительно превосходит единицу. При этом частоты ω_1 и ω_{pl} оказываются близкими друг к другу. Следовательно, в интервале длин волн

$$\frac{2\pi c}{\omega_{pl}} > \lambda > \frac{2\pi c}{\omega_1} \tag{3.14}$$

происходит быстрый рост коэффициента отражения с увеличением длины волны. Об этом участке говорят как о *плазменном крае отражения*. Определяя из опыта длину волны, при которой коэффициент отражения минимален, можно найти оптическую эффективную массу носителя заряда.

*) Строго говоря, в формуле для $\mathcal{P}$ должна была бы фигурировать групповая скорость электромагнитной волны, а не фазовая $c/\varepsilon_1^{-1/2}$. Однако сама формула (1.21′) получена уже в пренебрежении дисперсией света (§ 1).

**) Соответствующие вычисления можно найти в книге [2].

§ 4. Коэффициенты поглощения и излучения при оптических переходах зона — зона

Для вычисления коэффициента поглощения, связанного с междузонными переходами, удобно исходить из соотношения (3.7). Среднюю энергию, выделяемую в единицу времени в единице объема образца, следует вычислять квантовомеханическим путем. Для этого надо прежде всего вычислить отнесенную к единице времени вероятность $W(\lambda, \lambda')$ междузонного перехода, связанного с поглощением фотона. В отсутствие внешних электрического и магнитного полей квантовые числа λ, λ' здесь представляют собой совокупности номеров зон l, l' и квазиимпульсов $\mathbf{p}$, $\mathbf{p}'$ электрона в начальном и конечном состояниях:

$$\lambda = \{l, \mathbf{p}\}, \quad \lambda' = \{l', \mathbf{p}'\}.$$

Умножив $W(\lambda, \lambda')$ на энергию фотона $\hbar\omega$ и поделив на объем образца V, мы найдем среднюю энергию, поглощаемую в единицу времени в единице объема образца. Далее ее следует просуммировать по всем конечным состояниям, допускаемым принципом Паули, и усреднить по начальным состояниям с учетом вероятности их заполнения электронами. Таким образом,

$$Q_{\text{погл}} = \frac{1}{V} \sum_{\lambda, \lambda'} W(\lambda, \lambda') f(\lambda) [1 - f(\lambda')] \hbar\omega. \qquad (4.1)$$

Здесь $f(\lambda)$ есть функция распределения электронов.

В первом неисчезающем приближении вероятность $W(\lambda, \lambda')$ дается выражениями (XIV.2.15), (XIV.2.16), (XIV.2.10). При этом функции ψ_λ, $\psi^*_{\lambda'}$ суть обычные функции Блоха, а в качестве H' надо взять оператор энергии взаимодействия электрона с электромагнитным полем. Этот оператор легко найти, записав уравнение Шредингера в присутствии электромагнитного поля. Мы имеем

$$i\hbar \frac{\partial \Psi}{\partial t} = \frac{1}{2m_0} \left(-i\hbar \nabla + \frac{e}{c} \mathcal{A} \right)^2 \Psi - e\varphi\Psi + \beta g (\boldsymbol{\sigma}, \mathfrak{B}) \Psi. \qquad (4.2)$$

Здесь $\mathcal{A}$ и φ — скалярный и векторный потенциалы, связанные с напряженностью электрического поля $\mathfrak{E}$ и магнитной индукцией $\mathfrak{B}$ соотношениями

$$\mathfrak{E} = -\nabla\varphi - \frac{1}{c} \frac{\partial \mathcal{A}}{\partial t}, \quad \mathfrak{B} = \operatorname{rot} \mathcal{A}. \qquad (4.3)$$

Как известно из электродинамики, потенциалы $\mathcal{A}$ и φ можно выбирать любым способом, лишь бы удовлетворялись равенства (4.3). В частности, для поперечной электромагнитной волны скалярный потенциал можно положить равным нулю, подчинив векторный условию

$$\operatorname{div} \mathcal{A} = 0. \qquad (4.4)$$

При этом слагаемое $-e\varphi\Psi$ в (4.2) описывает только взаимодействие заряженной частицы со статическими полями (в том числе с периодическим полем идеальной решетки), а само уравнение можно переписать в виде

$$i\hbar\frac{\partial\Psi}{\partial t}=-\frac{\hbar^2}{2m_0}\nabla^2\Psi-e\varphi\Psi-\frac{ie\hbar}{m_0c}(\mathcal{A}\cdot\nabla\Psi)+\frac{e^2}{2m_0c^2}\mathcal{A}^2\Psi+\beta g(\sigma,\ \mathfrak{B})\Psi. \tag{4.2'}$$

Последние три слагаемых в правой части (4.2′) описывают искомую энергию взаимодействия заряженной частицы с электромагнитным полем. При исследовании структуры зон оптическими методами чаще всего используются сравнительно слабые сигналы: амплитуда напряженности электрического поля волны очень мала по сравнению с напряженностью внутрикристаллического поля. Тогда слагаемым, квадратичным по вектор-потенциалу, можно пренебречь, и оператор энергии взаимодействия электрона с электромагнитным полем принимает вид

$$H'=-\frac{ie\hbar}{m_0c}(\mathcal{A},\ \nabla)+\beta g(\sigma,\ \mathfrak{B}). \tag{4.5}$$

Подчеркнем, что здесь фигурирует масса свободного электрона. Ни метод эффективной массы, ни какие-либо предположения о виде законов дисперсии носителей заряда никак не использовались при выводе формулы (4.5). Второе слагаемое в (4.5) обусловливает возможность поглощения электромагнитных волн с изменением ориентации спина электрона. Однако вероятность таких электронных переходов зачастую невелика по сравнению с вероятностью переходов, при которых спин сохраняется. В дальнейшем мы будем рассматривать только переходы с сохранением ориентации спина; при этом играет роль только первое слагаемое в (4.5), а второе можно опустить.

Рассматривая монохроматическую волну, можем положить, с учетом (4.3),

$$\mathcal{A}=-\frac{ic}{\omega}\xi\mathscr{E}_m e^{ikx-i\omega t}. \tag{4.6}$$

Соответственно

$$H'=\frac{e\hbar}{m_0\omega}e^{ikx-i\omega t}\mathscr{E}_m(\xi,\ \nabla). \tag{4.5'}$$

Амплитуда $\mathscr{E}_m$ рассматривается здесь как классическая величина, а не как оператор.

Таким образом, в первом неисчезающем приближении вероятность перехода $W_{\lambda\lambda'}$, фигурирующая в формуле (4.1), дается выражением

$$W(\lambda,\ \lambda')=\frac{2\pi e^2\hbar}{m_0^2\omega^2}\mathscr{E}_m^2\,|(\xi,\ \mathbf{J}_{\lambda\lambda'})|^2\,\delta(E_{\lambda'}-E_\lambda-\hbar\omega). \tag{4.7}$$

Здесь

$$\mathbf{J}_{\lambda\lambda'} = \int \psi_\lambda^* e^{ikx} \cdot \nabla\psi_\lambda \, d\mathbf{r}. \tag{4.8}$$

Подставляя выражение (4.7) в правую часть (4.1) и сравнивая результат с формулой (3.7), находим вещественную часть электропроводности на частоте ω:

$$\sigma_1 = \frac{4\pi e^2\hbar^2}{V m_0^2 \omega} \sum_{\lambda,\,\lambda'} |(\xi, \mathbf{J}_{\lambda\lambda'})|^2 \, \delta(E_{\lambda'} - E_\lambda - \hbar\omega) f(E_\lambda)[1 - f(E_{\lambda'})]. \tag{4.9}$$

Заметим, что выражение (4.9) имеет меньшую область применимости, нежели (4.1): оно получено в пренебрежении взаимодействием носителя заряда с любыми неидеальностями решетки. В противном случае в качестве H' следовало бы взять сумму выражения (4.5′) и операторов, описывающих названное взаимодействие (гл. XIV).

Одна черта формулы (4.9) имеет, однако, общее значение. Именно, в правой части (4.9) стоит сумма выражений, относящихся к переходам между различными энергетическими зонами. Разность энергий $E(\lambda') - E(\lambda)$ определяется энергией фотона $\hbar\omega$. Последняя чаще всего такова, что пара зон l' и l выделяется однозначно: играют роль только переходы между зонами с данными фиксированными значениями l и l'. Это позволяет ограничиться исследованием лишь одного члена суммы.

Наряду с процессами поглощения фотонов (электронными переходами из более низкого энергетического состояния λ в более высокое λ'), происходят и обратные переходы $\lambda' \to \lambda$, сопровождающиеся излучением фотонов. При этом следует различать два типа таких процессов: самопроизвольные, или спонтанные, переходы и вынужденные переходы, возникающие под воздействием уже имеющегося в образце электромагнитного поля. Необходимость существования вынужденных переходов была впервые указана Эйнштейном в 1917 г. в связи с анализом законов теплового излучения. Энергия, излучаемая в результате вынужденных переходов (в единицу времени и в единице объема), выражается формулой

$$Q_{\text{вын}} = \frac{1}{V} \sum_{\lambda',\,\lambda} W(\lambda,\ \lambda') f(\lambda')(1 - f(\lambda))\,\hbar\omega, \tag{4.10}$$

которая получается из (4.1), если поменять местами индексы λ и λ'. Учитывая выражение (4.7) для $W(\lambda, \lambda')$ и формулу (4.8), можно убедиться, что

$$W(\lambda,\ \lambda') = W(\lambda',\ \lambda). \tag{4.11}$$

Таким образом, вероятности элементарного акта поглощения фотона и обратного ему вынужденного излучения одинаковы.

Из формулы (4.7) также видно, что вероятность вынужденного излучения, так же как и поглощения, пропорциональна $\mathcal{E}_m^2$, т. е., согласно формуле (3.8), пропорциональна концентрации фотонов в образце.

Для спонтанного излучения, аналогично формуле (4.10), можно написать

$$Q_{\text{сп}} = \frac{1}{V} \sum_{\lambda', \lambda} a_{\lambda'\lambda} f(\lambda') (1 - f(\lambda)) \hbar\omega. \tag{4.12}$$

Здесь $a_{\lambda'\lambda}$ — вероятность элементарного акта спонтанного перехода, которая, в отличие от $W(\lambda', \lambda)$, не зависит от концентрации уже имеющихся фотонов.

Рассмотрим, как влияет вынужденное излучение на измеряемый коэффициент поглощения света. Если бы имело место только излучение света, то, продолжая пользоваться формулой (3.7), мы должны были бы считать $\sigma_1 < 0$ и, соответственно, $\gamma < 0$. Фактически наблюдаемый на опыте коэффициент поглощения определяется разностью поглощаемого и испускаемого света. С учетом (1.21′) мы получаем

$$\gamma = \frac{(4\pi e\hbar)^2}{V m_0^2 \omega c \varepsilon_1^{1/2}} \sum_{\lambda, \lambda'} |(\xi, \mathbf{J}_{\lambda\lambda'})|^2 (f(\lambda) - f(\lambda')) \delta(E_{\lambda'} - E_\lambda - \hbar\omega). \tag{4.13}$$

Если равновесие в электронном газе нарушается только за счет поглощения света, то функции $f(\lambda)$ и $f(\lambda')$ можно заменить равновесными функциями Ферми

$$f_0(\lambda) = \frac{1}{\exp \frac{E_\lambda - F}{kT} + 1}, \quad f_0(\lambda') = \frac{1}{\exp \frac{E_{\lambda'} - F}{kT} + 1}. \tag{4.14}$$

Разность $f_0(\lambda) - f_0(\lambda')$ при этом положительна, так как $E_{\lambda'} = E_\lambda + \hbar\omega > E_\lambda$. Следовательно, $\gamma > 0$, т. е. имеет место результирующее поглощение света.

Найдем теперь величину коэффициента спонтанного излучения $a_{\lambda'\lambda}$. Для этого рассмотрим образец в условиях термодинамического равновесия. Так как в этом случае энергия электромагнитного поля не изменяется, то

$$Q_{\text{погл}} - Q_{\text{вын}} - Q_{\text{сп}} = 0. \tag{4.15}$$

Более того, согласно принципу детального равновесия это равенство должно выполняться по отдельности для каждой пары состояний λ, λ'. Рассмотрим две группы состояний с энергиями $E_{\lambda'}$ и, соответственно, E_λ, заключенные в узком интервале энергий ΔE. При этом $E_{\lambda'} - E_\lambda = \hbar\omega$, $\Delta E = \Delta(\hbar\omega)$. Учтем, далее, что для равновесных функций Ферми (4.14)

$$f_0(\lambda)(1 - f_0(\lambda')) = f_0(\lambda')(1 - f_0(\lambda)) \exp(\hbar\omega/kT). \tag{4.16}$$

Тогда, подставляя в условие (4.15) выражения (4.1), (4.10), (4.12) и учитывая (4.11), находим

$$a_{\lambda\lambda'} = W(\lambda, \lambda')\,[\exp(\hbar\omega/kT) - 1]. \tag{4.17}$$

Воспользуемся для $W(\lambda, \lambda')$ формулой (4.7), выражая в ней $\mathcal{E}_m^2$ через концентрацию фотонов $\rho(\hbar\omega)$ по формуле (3.8). Учитывая еще, что в равновесии $\rho_0(\hbar\omega)$ дается формулой Планка

$$\rho_0(\hbar\omega) = \frac{\omega^2 \varepsilon_1^{1/2}}{\pi^2 \hbar c^3} \frac{1}{\exp(\hbar\omega/kT) - 1} \tag{4.18}$$

(в которой мы заменили показатель преломления n на $\varepsilon_1^{1/2}$), получаем

$$a_{\lambda'\lambda} = \frac{16 e^2 \hbar\omega \varepsilon_1^{1/2}}{m_0^2 c^3} |(\xi, \mathbf{J}_{\lambda\lambda'})|^2 \delta(E_{\lambda'} - E_\lambda - \hbar\omega)\,\Delta(\hbar\omega). \tag{4.19}$$

Не завися от вида функций распределения, эта формула остается в силе и в любых неравновесных условиях. Из формулы (4.19) видно, что $a_{\lambda'\lambda}$ не зависит от концентрации фотонов, имеющихся в образце, чего и следовало ожидать для спонтанных переходов.

Полученные результаты позволяют также найти отношение энергий, излучаемых при спонтанных и вынужденных переходах. Из формул (4.10) и (4.12) следует, что при данном переходе

$$\frac{Q_{\text{сп}}}{Q_{\text{вын}}} = \frac{a_{\lambda'\lambda}}{W(\lambda, \lambda')}.$$

Пользуясь теперь формулами (4.17) и (4.18), находим

$$\frac{Q_{\text{сп}}}{Q_{\text{вын}}} = \frac{\omega^2 \varepsilon_1^{1/2}}{\pi^2 c^3 \hbar} \frac{1}{\rho(\hbar\omega)}. \tag{4.20}$$

Таким образом, относительная роль обоих процессов зависит от концентрации фотонов данной частоты, т. е. от интенсивности активного излучения и от ширины его спектра. С помощью количественных оценок можно убедиться, что в обычных лабораторных источниках сплошного спектра (например, лампах накаливания) спонтанное излучение гораздо важнее вынужденного. Напротив, в оптических квантовых генераторах (лазерах, § 6), мощность которых огромна, а спектральный интервал очень узок, $\rho(\hbar\omega)$ очень велика и вынужденное излучение играет основную роль.

§ 5. Прямые и непрямые переходы

Рассмотрим поглощение света в идеальной решетке, связанное с переходами между валентной зоной и зоной проводимости. Для этой цели можно воспользоваться формулой (4.13), взяв в ней слагаемые с $l = c$ и $l' = v$. При этом матричный элемент, фигурирую-

щий в (4.13), содержит функции Блоха (III.2.15′). Таким образом,

$$\mathbf{J}_{\lambda\lambda'} = \int \left(u^*_{\lambda'} \cdot \nabla u_\lambda + \frac{i}{\hbar} \mathbf{p} u^*_{\lambda'} u_\lambda \right) e^{\frac{i}{\hbar}(\mathbf{p}-\mathbf{p}'+\hbar\mathbf{k},\, \mathbf{r})}\, d\mathbf{r}. \qquad (5.1)$$

Для удобства записи мы ввели здесь волновой вектор фотона $\mathbf{k} = \{k, 0, 0\}$.

Ограничимся наиболее интересным случаем сравнительно слабого поглощения, когда справедливо условие (1.19′). Тогда мнимой частью k в формуле (5.1) можно пренебречь. Соответственно подынтегральное выражение в (5.1) представляет собой произведение экспоненциальной функции с мнимым аргументом и функции периодической с периодом решетки. Согласно теореме, доказанной в Приложении XIV, интеграл $J_{\lambda\lambda'}$ может быть отличен от нуля, лишь если выполняется правило отбора

$$\mathbf{p}' = \mathbf{p} + \hbar\mathbf{k} + m\hbar\mathbf{b}, \qquad (5.2)$$

где m — целое число или нуль*). Равенство (5.2) на первый взгляд кажется очевидным. Фактически, однако, выражаемый им закон сохранения не вполне тривиален: вектор $\hbar\mathbf{k}$ есть *импульс* фотона, в то время как $\mathbf{p}$ и $\mathbf{p}'$ суть *квазиимпульсы* электрона в начальном и конечном состояниях.

При не слишком большой разности энергий $E_{\lambda'} - E_\lambda$ величина $\hbar\mathbf{k}$ заметно меньше характерного квазиимпульса электрона в начальном или конечном состояниях. В самом деле, согласно равенству (1.10) и закону сохранения энергии, выражаемому δ-функцией в формуле (4.13), мы имеем $k \sim \frac{n}{\hbar c}(E_{\lambda'} - E_\lambda)$. Следовательно,

$$\frac{\hbar k}{p} \sim \frac{n(E_{\lambda'} - E_\lambda)}{cp} \equiv \frac{\tilde{v}(\mathbf{p})}{c/n}; \qquad (5.3)$$

характерная скорость $\tilde{v}(\mathbf{p})$ определяется этим соотношением. Оптические переходы могут совершать любые электроны, в том числе и обладающие очень малыми квазиимпульсами. Однако число таких электронов невелико, ибо в этой области очень мала плотность состояний (V.2.3). Основную роль играют электроны с конечными значениями p; при этом величина $\tilde{v} = \frac{E_{\lambda'} - E_\lambda}{p}$ обычно оказывается заметно меньше фазовой скорости света в среде c/n. Пусть, например, $E_{\lambda'} - E_\lambda = 1$ эВ и $n = 4$, тогда $k \simeq 10^{+5}$ см$^{-1}$. Пусть, далее, точка с энергией $E_{\lambda'}$ отстоит от края зоны на 10^{-2} эВ (при этом плотность состояний уже не слишком мала). Пусть, наконец, справедлив простой параболический закон дисперсии с эффективной

*) В дальнейшем мы будем для краткости опускать последнее слагаемое в правой части (5.2): поскольку векторы $\mathbf{p}$ и $\mathbf{p}'$ лежат в пределах первой зоны Бриллюэна, число m определяется однозначно.

массой $m = 0{,}1m_0$. Тогда $p/\hbar \sim 10^6$ см$^{-1}$, т. е. правая часть (5.3) составляет около 0,1. Эта ситуация оказывается довольно типичной: с точностью до 10 ÷ 15% импульсом фотона можно пренебречь. Поэтому правило отбора (5.2) принимает более простой вид:

$$\mathbf{p}' = \mathbf{p}. \tag{5.2'}$$

Поскольку вектор —$\mathbf{p}$ есть не что иное, как квазиимпульс дырки, то равенства (5.2), (5.2') имеют очень простой смысл: с точностью до малой величины $\hbar\mathbf{k}$ полный квазиимпульс системы «электрон проводимости + дырка» не изменяется при рассматриваемом оптическом переходе, оставаясь равным нулю (до поглощения или после испускания фотона этот квазиимпульс, разумеется, равен нулю ввиду отсутствия обоих носителей заряда).

При условии (5.2') правая часть равенства (5.1) принимает вид

$$\int \left(u^*_{pl'} \cdot \nabla u_{pl} + \frac{i}{\hbar} \mathbf{p} u^*_{pl'} u_{pl} \right) d\mathbf{r}.$$

Интеграл от второго слагаемого в скобках равен нулю: в силу (III.4.3) при $l' \neq l$ функции $u^*_{pl'}$ и u_{pl} взаимно ортогональны. Таким образом, в рассматриваемом случае

$$\mathbf{J}_{\lambda\lambda'} = \int u^*_{pl'} \cdot \nabla u_{pl}\, d\mathbf{r} \equiv \frac{i}{\hbar} \mathbf{p}_{l'l}(\mathbf{p}), \tag{5.1'}$$

где

$$\mathbf{p}_{l'l}(\mathbf{p}) = -i\hbar \int u^*_{pl'} \cdot \nabla u_{pl}\, d\mathbf{r} \tag{5.4}$$

есть матричный элемент оператора импульса в системе функций $u_{pl'}$, u_{pl}, вычисленный в точке $\mathbf{p}$ зоны Бриллюэна. При этом, пользуясь периодичностью указанных функций, область интегрирования в (5.4) можно ограничить пределами одной элементарной ячейки (нормируя функции $u_{pl'}$, u_{pl} в той же области).

Итак, в идеальной решетке междузонные оптические переходы происходят практически в одной и той же точке зоны Бриллюэна. Такие переходы называются *прямыми* или *вертикальными*. Смысл этого названия становится ясным из рис. 18.4, *а*, где изображены (схематически) законы дисперсии для электронов проводимости и дырок в полупроводнике типа германия. Видно, что равенство (5.2') исключает здесь возможность поглощения фотонов с энергией, близкой к ширине запрещенной зоны. Такое поглощение, однако, становится возможным, если принять во внимание неидеальность решетки. При этом, благодаря взаимодействию носителей заряда с рассеивателями, закон сохранения квазиимпульса в форме (5.2) или (5.2') уже не имеет места: часть квазиимпульса поставляется атомами примеси или фононами, испускаемыми при оптическом переходе. Переходы такого типа называются *непрямыми* или *невертикальными*. В материалах типа германия именно они ответственны за поглощение волн с частотой, близкой к граничной.

Будучи связаны с необходимостью взаимодействия электронов не только со светом, но и с рассеивателями, непрямые переходы менее вероятны, нежели прямые (если последние разрешены). Поэтому для них характерно заметно меньшее значение коэффициента поглощения, что и видно на рис. 18.3: область 0,7 эВ $\lesssim \hbar\omega \lesssim$ 0,8 эВ отвечает там непрямым переходам, а область $\hbar\omega \gtrsim 0{,}8$ эВ — прямым.

Существуют материалы, в которых дно зоны проводимости и потолок валентной зоны располагаются в одной и той же (или почти одной и той же) точке зоны Бриллюэна. К числу таких веществ

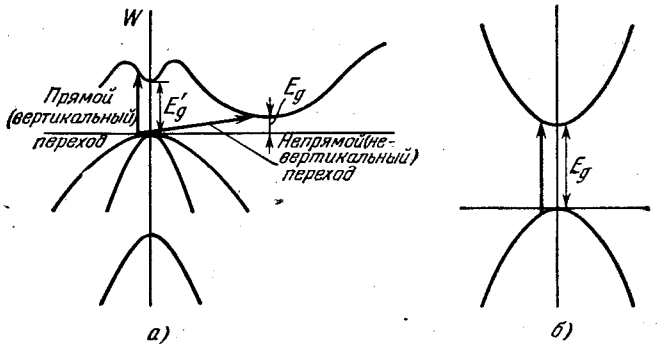


Рис. 18.4. *а*) Вертикальные и невертикальные переходы в «непрямом» полупроводнике. *б*) Вертикальный переход в «прямом» полупроводнике.

относятся, например, арсенид галлия, антимонид индия и др. Прямые переходы в них могут вызываться фотонами с энергией, сколь угодно близкой к красной границе (рис. 18.4, *б*).

О материалах двух указанных типов иногда говорят как о «непрямых» и «прямых» соответственно.

Заметим, что вывод равенства (5.2) никак не был связан с предположением о том, что переходы происходят непременно между различными зонами: последнее условие определяет только область частот ω. В случае внутризонных оптических переходов, рассмотренных в § 3, характерная скорость $\tilde{v}$ — порядка тепловой и правая часть (5.3) также мала. С другой стороны, внутризонные переходы непременно связаны с изменением квазиимпульса, т. е. они по определению невертикальные. Следовательно, они могут происходить лишь при наличии рассеяния носителей заряда, в чем мы уже убедились другим путем в § 3.

§ 6. Полупроводниковые лазеры

В § 4 мы видели, что величина и знак коэффициента поглощения света γ существенно зависят от разности $f(\lambda) - f(\lambda')$, где f — степень заполнения электронами энергетических уровней, между

которыми происходят электронные переходы (формула (4.13)). Там же мы отмечали, что если в отсутствие света электронный газ находится в равновесии, то $\gamma > 0$ и, следовательно, электромагнитная волна ослабляется. Однако если в полупроводник вводить энергию от внешнего источника и этим существенно нарушать равновесие в электронном газе («накачка» энергии), то может оказаться, что $f(\lambda) > f(\lambda')$, т. е. что заселенность расположенных выше уровней станет больше заселенности уровней, лежащих ниже (*инверсия* заселенности уровней). В среде с инверсным распределением электронов γ становится отрицательным, т. е. волна не поглощается, а усиливается. Среды, в которых $\gamma < 0$, называются *активными*.

Остановимся подробнее на условии возникновения инверсного распределения. Рассмотрим сначала однородные полупроводники и прямые электронные переходы зона—зона. Будем считать, что при накачке успевает установиться равновесное энергетическое распределение электронов в зоне проводимости и дырок — в валентной зоне (но, конечно, не имеется равновесия между обеими зонами). Как мы видели в гл. VII, в этом случае можно ввести квазиуровни Ферми для электронов F_n и для дырок F_p. Тогда условие инверсии для какой-либо пары уровней, один из которых, E_λ, лежит в валентной зоне, а другой, $E_{\lambda'}$ — в зоне проводимости, есть

$$f(E_{\lambda'}) = \frac{1}{\exp\dfrac{E_{\lambda'} - F_n}{kT} + 1} > f(E_\lambda) = \frac{1}{\exp\dfrac{E_\lambda - F_p}{kT} + 1}. \qquad (6.1)$$

Отсюда получаем

$$F_n - F_p > E_{\lambda'} - E_\lambda. \qquad (6.2)$$

Так как минимальное значение $(E_{\lambda'} - E_\lambda)$ есть ширина запрещенной зоны E_g, то условие возможности инверсии есть

$$F_n - F_p > E_g. \qquad (6.3)$$

Следовательно, накачка должна быть настолько сильной, чтобы квазиуровни Ферми оказались внутри разрешенных зон энергии; т. е. чтобы электронный и дырочный газы стали вырожденными (рис. 18.5). При этом все уровни в валентной зоне с энергией $E_\lambda > F_p$ будут практически пустыми, а все уровни в зоне проводимости с $E_{\lambda'} < F_n$ — заполнены электронами. Поэтому фотоны, энергия которых лежит в интервале $\omega_{\min} < \omega < \omega_{\max}$, где

$$\hbar\omega_{\max} = F_n - F_p, \qquad \hbar\omega_{\min} = E_g, \qquad (6.4)$$

не могут вызывать электронные переходы валентная зона — зона проводимости и, следовательно, не будут поглощаться. Обратные же переходы зона проводимости—валентная зона возможны, и поэтому

будет происходить излучение фотонов. Соотношения (6.4) определяют частотный интервал, в котором возможно усиление электромагнитных волн. Если инверсное распределение создается освещением от постороннего сильного источника света, то энергия фотонов накачки $\hbar\omega_н$ должна, очевидно, быть (рис. 18.5)

$$\hbar\omega_н \geqslant F_n - F_p = \hbar\omega_{max}. \qquad (6.5)$$

Рассмотрим теперь многодолинные полупроводники и непрямые переходы. Этот случай изображен на рис. 18.6 для конкретного

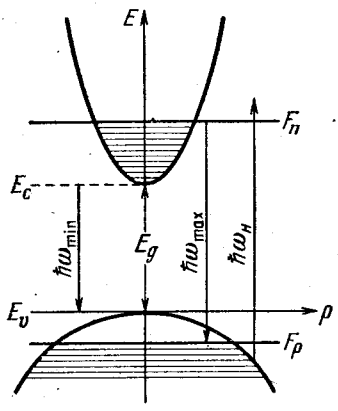


Рис. 18.5. Инверсное распределение электронов в однодолинном полупроводнике.

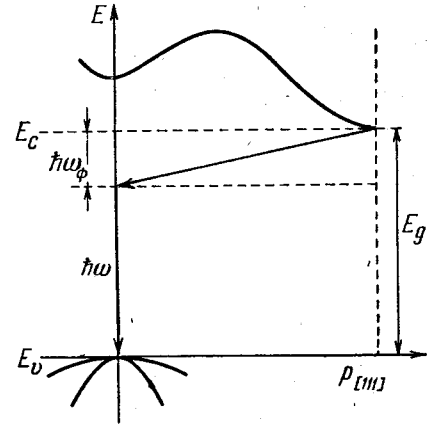


Рис. 18.6. Электронные переходы в многодолинном полупроводнике (германий) с участием фонона.

примера германия. Как показано в § 5, здесь возможны электронные переходы только с участием рассеивателей. Для определенности мы положим, что выполнение законов сохранения энергии и квазиимпульса обеспечивается участием тепловых колебаний решетки, так что при электронном переходе зона проводимости—валентная зона испускается один фонон с энергией $\hbar\omega_ф$. Фононы будем считать равновесными (термолизованными) и находящимися в равновесии с электронным газом. Тогда условие инверсии для пары уровней $E_{\lambda'}$ и E_λ, вместо формулы (6.1), будет иметь вид

$$\left[\exp\frac{E_{\lambda'}-\hbar\omega_ф-F_n}{kT}+1\right]^{-1} > \left[\exp\frac{E_\lambda - F_p}{kT}+1\right]^{-1}. \qquad (6.6)$$

Соответственно условие активности среды, вместо (6.3), теперь будет

$$F_n - F_p > E_g - \hbar\omega_ф.$$

Следовательно, при непрямых переходах для возникновения активности необходима меньшая разность $(F_n - F_p)$, т. е. меньшая накач-

ка. Отметим, однако, что это совсем не значит, что для усиления волн выгоднее использовать непрямые переходы. Действительно, величина усиления зависит от темпа вынужденных излучательных переходов, который, при прочих равных условиях, пропорционален вероятности переходов. А эта последняя, как отмечалось выше (§ 5), для непрямых переходов меньше, чем для прямых.

Существенная особенность усиления электромагнитных волн в активных средах состоит в сужении спектрального интервала при усилении. Это происходит по той причине, что коэффициент усиления γ внутри полосы усиления ($\omega_{max} - \omega_{min}$) зависит от ω и при некоторой частоте имеет максимум. Поэтому преимущественно усиливаются лишь те частоты, которые лежат вблизи нее, и волна в процессе усиления становится более монохроматической.

Использование вынужденного излучения в среде с инверсной заселенностью уровней лежит в основе квантовых генераторов, или, иначе, *лазеров* *), предназначенных для генерации когерентных электромагнитных волн. Принцип их работы заключается в том, что активная среда сопрягается с резонатором колебаний и между ними осуществляется обратная связь. Последняя обычно реализуется тем, что в качестве резонатора используют два плоскопараллельных зеркала (оптический резонатор Фабри и Перро), а активную среду помещают между зеркалами резонатора. Тогда усиленная волна, отражаясь от одного из зеркал, снова вступает в активную среду и вызывает усиленное вынужденное излучение, а, следовательно, и сама усиливается еще больше. Отражаясь от второго зеркала, волна попадает опять в активную среду и снова усиливается и т. д. Если сделать зеркала полупрозрачными, то часть энергии волн будет выходить наружу. Кроме того, часть энергии волн теряется вследствие неизбежных потерь в самом резонаторе. Когда выигрыш в энергии волны становится равным суммарным потерям, в резонаторе устанавливаются стационарные автоколебания и все устройство превращается в источник когерентного электромагнитного излучения высокой степени монохроматичности.

В существующих полупроводниковых лазерах активность среды создается за счет инверсного распределения электронов между валентной зоной и зоной проводимости **). Необходимая для инверсии большая концентрация электронов и дырок может быть создана разными способами: инжекцией в p—n-переходе (инжекционная накачка), освещением сильным (некогерентным) источником света (оптическая накачка), бомбардировкой быстрыми электронами (электронная накачка) и другими приемами.

---

*) Название произошло от первых букв английской фразы «light amplification by stimulated emission of radiation».

**) Возможность использования полупроводников для создания активной среды в лазерах впервые была указана в 1958 г. Н. Г. Басовым и Б. М. Вулом.

На рис. 18.7 показан принцип действия инжекционного лазера. Его основная часть — монокристалл полупроводника, в котором путем надлежащего распределения мелких доноров и акцепторов созданы области p- и n-проводимости. Между ними имеется неизбежная переходная область некоторой толщины d. В слабо легированных полупроводниках создать большую концентрацию электронов и дырок, необходимую для выполнения условия инверсии (6.3), путем одной только накачки трудно. Поэтому p- и n-области делают

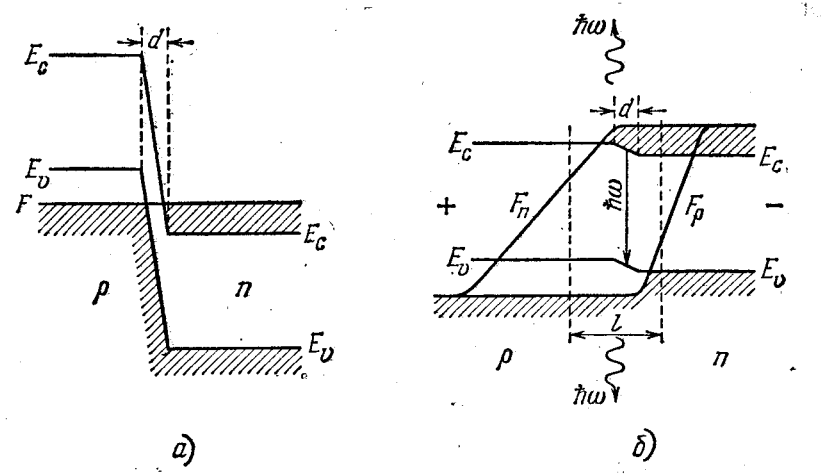


Рис. 18.7. Энергетическая диаграмма инжекционного полупроводникового лазера. *а*) Вырожденный p — n-переход без внешнего напряжения, *б*) Вырожденный p — n-переход при положительном напряжении, d — технологическая ширина перехода, l — ширина активной области ($F_n - F_p \geqslant E_g$).

сильно легированными, чтобы электронный и дырочный газы в n- и, соответственно, в p-области были вырождены. В этом случае уже без инжекции уровень Ферми в p-области лежит внутри валентной зоны, а в n-области — внутри зоны проводимости (рис. 18.7, *а*), отчего выполнение условия (6.3) при накачке облегчается. При приложении к $p-n$-переходу положительного напряжения (плюс источника на p-области, минус на n-области) происходит инжекция дырок (в n-область) и инжекция электронов (в p-область). Неравновесные концентрации носителей постепенно уменьшаются по мере удаления от границ перехода и затухают на расстояниях порядка их длин диффузии. Соответственно первоначальный уровень Ферми расщепляется на квазиуровни F_p и F_n, которые изменяются в пространстве так, как показано схематически на рис. 18.7, *б*. Из рисунка видно, что в окрестности перехода возникает активный слой, в котором $F_n - F_p \geqslant E_c - E_v = E_g$. Толщина активного слоя l может быть значительно больше d.

Многократное отражение излучения происходит от двух параллельных полированных граней кристалла, которые играют роль зеркал резонатора. Излучение выходит через узкие полоски, образованные пересечением активного слоя с частично отражающими гранями кристалла.

Лазерный эффект наблюдается во многих однодолинных полупроводниках. Одним из лучших полупроводниковых лазеров является инжекционный лазер на арсениде галлия. Укажем для ориентировки некоторые его характеристики. Размеры таких лазеров очень малы, что связано с трудностями изготовления больших однородных p—n-переходов. При излучающей поверхности порядка 10^{-4} см^2 излучаемая ими мощность в непрерывном режиме достигает ~ 10 Вт, что в пересчете на 1 см^2 соответствует $\sim 10^2 \div 10^3$ кВт/см^2. В соответствии с шириной запрещенной зоны длина волны излучаемого света равна $\sim$8000—9000 Å (в зависимости от рабочей температуры). Величина пороговой плотности тока через p—n-переход, при которой наступает самовозбуждение, имеет порядок 100 А/см^2 (что соответствует полному току порядка немногих ампер). Коэффициент полезного действия при охлаждении жидким азотом достигает 70—80%.

Характеристики полупроводниковых лазеров можно значительно улучшить, применяя вместо обычных p—n-переходов, созданных в одном и том же полупроводнике, анизотипные гетеропереходы (§ VIII.5). Это позволяет существенно уменьшить пороговый ток и потому легче осуществить непрерывную генерацию в неохлаждаемых переходах при комнатной температуре, что сильно расширяет возможные применения полупроводниковых лазеров. Подробнее о полупроводниковых лазерах см., например, в [4, 5].

§ 7. Коэффициент поглощения при прямых переходах. Комбинированная плотность состояний

Рассмотрим коэффициент поглощения при прямых междузонных переходах. Будем считать, что функции распределения $f(\lambda)$ и $f(\lambda')$ равновесные. Тогда формула (4.9) с учетом (5.1′) и (5.4) примет вид (при $l' = c$, $l = v$)

$$\sigma_1 = \frac{4\pi e^2}{m_0^2 \omega V} \sum_{\mathbf{p}} |(\xi,\ \mathbf{p}_{cv})|^2 f_0(E_v(\mathbf{p}))[1 - f_0(E_c(\mathbf{p}))] \times \\ \times \delta(E_c(\mathbf{p}) - E_v(\mathbf{p}) - \hbar\omega). \quad (7.1)$$

Переходя здесь от суммирования по $\mathbf{p}$ к интегрированию и пользуясь формулой (1.21′), мы получаем

$$\gamma = \frac{2e^2}{\pi c m_0^2 \omega \hbar^3 \varepsilon_1^{1/2}} \int d\mathbf{p}\, |(\xi,\ \mathbf{p}_{cv}(\mathbf{p}))|^2 f_0(E_v(\mathbf{p})) \times \\ \times [1 - f_0(E_c(\mathbf{p}))] \delta(E_c(\mathbf{p}) - E_v(\mathbf{p}) - \hbar\omega). \quad (7.2)$$

Если свет не поляризован, то величину $|(\xi, \mathbf{p}_{cv})|^2$ можно заменить результатом усреднения ее по углам $^1/_3 |\mathbf{p}_{cv}|^2$.

В невырожденных полупроводниках почти все состояния валентной зоны заполнены, а зоны проводимости — свободны. При этом множитель $f_0(E_v(\mathbf{p}))[1 - f_0(E_c(\mathbf{p}))]$ в правой части (7.2) можно заменить единицей. Это означает, что все переходы, дозволенные энергетически, разрешены и принципом Паули. Так, в «прямых» полупроводниках красная граница поглощения ω_m отвечает минимальному значению разности энергий $E_c(\mathbf{p}) - E_v(\mathbf{p})$:

$$\hbar\omega_m = E_g.$$

В условиях сильного вырождения положение меняется. Пусть, например, уровень Ферми лежит в зоне проводимости, причем $F - E_c \gg kT$ (рис. 18.8). Тогда принцип Паули практически

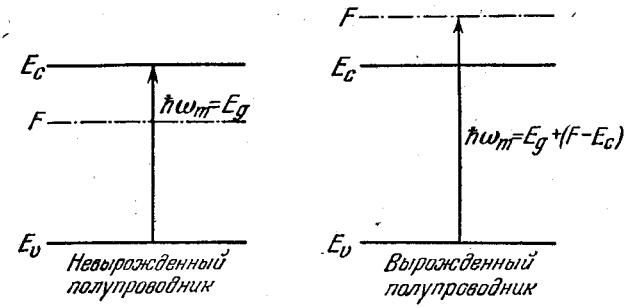


Рис. 18.8. Сдвиг Бурштейна — Мосса.

исключает возможность переходов в состояния зоны проводимости с энергией $E < F$. Красная граница поглощения при этом отвечает энергии $E_g + (F - E_c)$, т. е. смещается вверх (по сравнению с невырожденным образцом) на величину $(F - E_c)/\hbar$. Это явление называется сдвигом Бурштейна—Мосса. Заметим, что в рассматриваемой задаче

$$f_0(E_v(\mathbf{p}))[1 - f_0(E_c(\mathbf{p}))] = f_0(E_v(\mathbf{p})) - f_0(E_c(\mathbf{p}))$$

как для вырожденного, так и для невырожденного полупроводников. Этим оправдано использование формулы (4.9) и, следовательно, (7.2) вместо (4.13).

Для исследования структуры энергетических зон особый интерес представляет частотная зависимость коэффициента поглощения (7.2). Вычисление функции $\gamma(\omega)$ в общем виде связано с большими трудностями, ибо она определяется как законами дисперсии, так и матричным элементом $\mathbf{p}_{cv}(\mathbf{p})$. Заметим, однако, что последняя величина есть гладкая функция $\mathbf{p}$, ибо интеграл (5.4) сходится при любых значениях компонент квазиимпульса. Поэтому мы вправе

выбрать в зоне Бриллюэна некоторую точку $\mathbf{p}_0$ и разложить этот матричный элемент по степеням $\mathbf{p} - \mathbf{p}_0$:

$$p^{\alpha}_{cv}(\mathbf{p}) = p^{\alpha}_{cv}(\mathbf{p}_0) + c_{\alpha\beta}(p_\beta - p_{0,\beta}) + \ldots \tag{7.3}$$

Здесь $c_{\alpha\beta}$ — безразмерный тензор; многоточие означает члены высшего порядка по $(\mathbf{p} - \mathbf{p}_0)$ *).

Разложение (7.3) имеет смысл, если точки $\mathbf{p}$ и $\mathbf{p}_0$ не слишком далеки друг от друга. Иначе говоря, нужно, чтобы были достаточно близки друг к другу разности энергий $E_c(\mathbf{p}) - E_v(\mathbf{p})$ и $E_c(\mathbf{p}_0) - E_v(\mathbf{p}_0)$. Это означает, что при заданной точке $\mathbf{p}_0$ рассматриваемым разложением можно пользоваться лишь для определения функции $\gamma(\omega)$ в не слишком широком интервале частот. При этом величины $\mathbf{p}_{cv}(\mathbf{p}_0)$ и $c_{\alpha\beta}$ можно либо рассматривать как параметры теории, подлежащие определению из опыта, либо вычислять численными методами.

В зависимости от свойств симметрии волновых функций $u^*_{\mathbf{p}_0 c}$ и $u_{\mathbf{p}_0 v}$ матричный элемент

$$\mathbf{p}_{cv}(\mathbf{p}_0) = \int u^*_{\mathbf{p}_0 c}\,(-i\hbar\nabla)\,u_{\mathbf{p}_0 v}\,d\mathbf{r} \tag{7.4}$$

может быть отличным от нуля или обращаться в нуль.

Пусть, например, функции $u_{\mathbf{p}_0 v}$ и $u^*_{\mathbf{p}_0 c}$ характеризуются определенной четностью, т. е. либо меняют знак (нечетные функции), либо остаются неизменными (четные функции) при изменении знаков пространственных координат.

Рассмотрим сначала случай, когда их четность одна и та же, например, обе четные. Поскольку производная от четной функции нечетна, все подынтегральное выражение в (7.4) меняет знак при изменении знаков пространственных координат — переменных интегрирования. То же, очевидно, относится и к самому интегралу. Но при замене переменных интегрирования интеграл измениться не может. Следовательно, он равен нулю. Так, например, обстоит дело при переходах в точке $\mathbf{p}_0 = 0$ между зонами легких и тяжелых дырок в соединениях типа $A^{III}B^{V}$.

С другой стороны, если функции $u_{\mathbf{p}_0 v}$ и $u^*_{\mathbf{p}_0 c}$ обладают разной четностью, то подынтегральное выражение в (7.4) не меняется при изменении знака переменных интегрирования и интеграл может быть отличен от нуля. Это имеет место, например, при переходах между зоной легких дырок и зоной проводимости в арсениде галлия. При более сложных свойствах симметрии правила отбора устанавливаются с помощью теории групп.

---

*) Фактически в дальнейшем значение $\mathbf{p}_0$ будет отвечать либо границе зоны проводимости (или валентной), либо какой-либо другой критической точке (§ 8).

Если интеграл (7.4) отличен от нуля, то говорят о прямых разрешенных переходах, в противном случае — о прямых запрещенных.

Рассмотрим сначала прямые разрешенные переходы в невырожденном полупроводнике. При этом в правой части (7.3) можно оставить только первое слагаемое, т. е. заменить $\mathbf{p}_{cv}(\mathbf{p})$ на $\mathbf{p}_{cv}(\mathbf{p}_0)$ (впредь мы будем для краткости опускать аргумент $\mathbf{p}_0$). Тогда (для неполяризованного света) формула (7.2) принимает вид

$$\gamma = \frac{2e^2 |\mathbf{p}_{cv}|^2}{3\pi c m_0^2 \omega \hbar^3 \varepsilon_1^{1/2}} \int \delta(E_c(\mathbf{p}) - E_v(\mathbf{p}) - \hbar\omega)\, d\mathbf{p}. \tag{7.5}$$

Введем обозначение

$$\rho_{\text{комб}}(\omega) = \frac{2}{(2\pi\hbar)^3} \int \delta(E_c(\mathbf{p}) - E_v(\mathbf{p}) - \hbar\omega)\, d\mathbf{p}. \tag{7.6}$$

Получим

$$\gamma = \frac{2(2\pi)^2 e^2 |\mathbf{p}_{cv}|^2}{3 c m_0^2 \omega \varepsilon_1^{1/2}} \rho_{\text{комб}}(\omega). \tag{7.7}$$

Величина $\rho_{\text{комб}}(\omega)$ называется *комбинированной плотностью состояний*. Смысл этого названия становится ясным, если сравнить правую часть (7.6) с формулами для плотностей состояний в зонах проводимости и валентной (V.7.8б). Разница состоит лишь в том, что, в отличие от последних формул, правая часть (7.6) определяется не видом каждого из законов дисперсии $E_c(\mathbf{p})$ и $E_v(\mathbf{p})$ по отдельности, а разностью $E_c(\mathbf{p}) - E_v(\mathbf{p})$.

Как видно из формулы (7.6), комбинированная плотность состояний отлична от нуля лишь при

$$\omega \geqslant \omega_m = E_g/\hbar. \tag{7.8}$$

Смысл этого неравенства очевиден: в рассматриваемых условиях междузонные переходы могут быть вызваны только фотонами достаточно большой энергии.

Вычислим комбинированную плотность состояний для частот, близких к пороговому значению ω_m. В качестве $\mathbf{p}_0$ здесь удобно выбрать точку в зоне Бриллюэна, отвечающую дну зоны проводимости (по условию ей же соответствует и потолок валентной зоны). Предположим сначала, что $\mathbf{p}_0 = 0$, и аппроксимируем законы дисперсии простыми параболическими выражениями, совмещая начало отсчета энергии с дном зоны проводимости:

$$E_c(\mathbf{p}) = \frac{p^2}{2m_n}, \qquad E_v(\mathbf{p}) = -E_g - \frac{p^2}{2m} \tag{7.9}$$

Подставим выражения (7.9) в правую часть (7.6) и будем вычислять фигурирующий там интеграл в сферических координатах.

Интегрируя по полярным углам, получим

$$\rho_{\text{комб}}(\omega) = \frac{1}{\pi^2\hbar^3}\int_0^\infty p^2\delta\left(\frac{p^2}{2m_r} + E_g - \hbar\omega\right)dp, \tag{7.10}$$

где

$$m_r = \frac{m_n m_p}{m_n + m_p} \tag{7.11}$$

есть приведенная эффективная масса электрона и дырки.

Интеграл, фигурирующий в формуле (7.10), легко вычислить, пользуясь правилами Приложения IV. При $\omega \geqslant \omega_m$

$$\rho_{\text{комб}}(\omega) = \frac{[2m_r^3(\hbar\omega - E_g)]^{1/2}}{2\pi^2\hbar^3}. \tag{7.12}$$

Результат (7.12) легко обобщить и на случай, когда экстремумы зон лежат в точке $\mathbf{p}_0 \neq 0$ и законы дисперсии электронов и дырок анизотропные (но параболические). При этом частотная зависимость $\rho_{\text{комб}}(\omega)$ остается такой же, как и в (7.12); надо лишь заменить там множитель $m_r^{3/2}$ на $(m_{rx}m_{ry}m_{rz})^{1/2}$, где m_{rx}^{-1}, m_{ry}^{-1}, m_{rz}^{-1} суть главные значения тензора $m_{r,\alpha\beta}^{-1} = m_{c,\alpha\beta}^{-1} + m_{v,\alpha\beta}^{-1}$.

Таким образом, при параболических законах дисперсии частотная зависимость коэффициента поглощения при прямых разрешенных переходах имеет вид

$$\gamma(\omega) \sim \frac{1}{\omega}(\omega - \omega_m)^{1/2}. \tag{7.13}$$

Обратимся теперь к прямым запрещенным переходам. При этом в правой части (7.3) надо сохранить и второе слагаемое, в результате чего матричный элемент $p_{cv}(\mathbf{p})$ оказывается обычно отличным от нуля. Полагая вновь $\mathbf{p}_0 = 0$ и пользуясь формулами (7.3), (7.2) (для неполяризованного света) и (7.9), мы получаем (по-прежнему в отсутствие вырождения)

$$\gamma = \sum_{\alpha,\beta,\gamma}\frac{2e^2\,|\,c_{\alpha\beta}c_{\alpha\gamma}\,|^2}{3\pi cm_0^2\omega\hbar^3\varepsilon_1^{1/2}}\int p_\beta p_\gamma\delta\left(\frac{p^2}{2m_r} + E_g - \hbar\omega\right)d\mathbf{p}. \tag{7.14}$$

Как и в случае (7.6), интеграл легко вычисляется в полярных координатах. При $\omega \geqslant \omega_m$

$$\gamma = \sum_{\alpha,\beta}\frac{8e^2\,|\,c_{\alpha\beta}\,|^2\,m_r^{5/2}}{9cm_0^2\hbar^3\varepsilon_1^{1/2}\omega}[2(\hbar\omega - E_g)^3]^{1/2}. \tag{7.15}$$

Этот результат (с точностью до численного множителя) остается в силе и при $\mathbf{p}_0 \neq 0$, и при любой анизотропии тензоров обратных эффективных масс, лишь бы законы дисперсии электронов и дырок

оставались параболическими. Таким образом, частотная зависимость коэффициента поглощения при прямых запрещенных переходах имеет вид

$$\gamma \sim \frac{1}{\omega}(\omega - \omega_m)^{3/2}. \tag{7.16}$$

Как уже отмечалось, формулы (7.13) и (7.16) справедливы при не слишком больших значениях разности $\omega - \omega_m$: энергии электронов и дырок должны быть такими, чтобы законы дисперсии оставались параболическими и чтобы имело смысл разложение (7.3). Количественно эти условия можно сформулировать только применительно к тому или иному конкретному материалу. Следует также иметь в виду, что формулы (7.12) — (7.16) оказываются несправедливыми не только при очень больших частотах, но и при частотах, очень близких к пороговой. Дело в том, что вывод этих формул основан на зонной теории, в рамках которой носители заряда рассматриваются как невзаимодействующие частицы. Фактически, однако, такое взаимодействие имеет место: электрон и дырка испытывают взаимное кулоновское притяжение. Как мы видели в § XVII.7, характерная энергия последнего — порядка боровской энергии в кристалле $\frac{m_r e^4}{2\varepsilon^2\hbar^2}$. Следует ожидать поэтому, что принятый выше упрощенный подход будет оправдан, лишь если

$$\hbar\omega - E_g \gg \frac{m_r e^4}{2\varepsilon^2\hbar^2}. \tag{7.17}$$

Тщательное теоретическое исследование [6] действительно показывает, что в условиях (7.17) формулы (7.12)—(7.16) остаются в силе, в то время как при меньших частотах частотная зависимость коэффициента поглощения оказывается иной.

Во многих кристаллах боровская энергия, фигурирующая в правой части (7.17), сравнительно невелика. Так, в арсениде галлия и фосфиде индия она составляет около 0,004 эВ, в антимониде галлия — около 0,003 эВ. По этой причине область применимости формул (7.12)—(7.16) оказывается довольно широкой. Исключение составляют лишь узкозонные полупроводники, в которых отклонения законов дисперсии от параболических становятся заметными уже при довольно малых энергиях носителей заряда.

§ 8. Критические точки

Обратимся к изучению коэффициента поглощения при прямых разрешенных переходах в области частот, заметно превышающих пороговую. Здесь надо прежде всего выяснить, какие точки в зоне Бриллюэна представляют наибольший интерес. Для этой цели вернемся вновь к формуле (7.6). Фигурирующий в ней тройной

интеграл удобно вычислять, интегрируя сначала по поверхности

$$E_c(\mathbf{p}) - E_v(\mathbf{p}) \equiv E = \text{const}, \tag{8.1}$$

а затем по всем таким поверхностям, т. е. по всем значениям E (сравните с аналогичной выкладкой для плотности состояний в § V.7). Обозначив через dS элемент площади на поверхности (8.1), получим, как и в § V.7,

$$\rho_{\text{комб}}(\omega) = \frac{2}{(2\pi\hbar)^3} \int \frac{dS}{|\nabla_{\mathbf{p}} E|}\Bigg|_{E=\hbar\omega}. \tag{8.2}$$

Вспоминая формулу (IV.1.3), видим, что

$$|\nabla_{\mathbf{p}} E| = |\mathbf{v}_c(\mathbf{p}) - \mathbf{v}_v(\mathbf{p})|, \tag{8.3}$$

где $\mathbf{v}_c(\mathbf{p})$ и $\mathbf{v}_v(\mathbf{p})$ — скорости электрона с квазиимпульсом $\mathbf{p}$ в зонах проводимости и валентной.

Интеграл в (8.2) представляет собой гладкую функцию частоты при всех значениях ω, кроме тех, при которых знаменатель обращается в нуль:

$$|\nabla_{\mathbf{p}} E|_{E=\hbar\omega} = 0. \tag{8.4}$$

Если выполнено равенство (8.4), то интеграл должен иметь особенность: будет расходиться либо сама комбинированная плотность состояний, либо какие-то ее производные по ω (как мы увидим, в трехмерном случае сам интеграл в (8.2) остается конечным). Согласно (8.3) так обстоит дело, если междузонный переход происходит между состояниями с одинаковыми скоростями электронов (в частности, возможен и случай $\mathbf{v}_c(\mathbf{p}) = \mathbf{v}_v(\mathbf{p}) = 0$). Точки в зоне Бриллюэна, для которых

$$|\mathbf{v}_c(\mathbf{p}) - \mathbf{v}_v(\mathbf{p})| = 0, \tag{8.5}$$

называются *критическими*. Соответствующие сингулярности плотности состояний и коэффициента поглощения называются особенностями Ван-Хова.

Можно указать два типа критических точек:

1) критические точки первого рода:

$$\nabla_{\mathbf{p}} E_c(\mathbf{p}) \equiv \mathbf{v}_c(\mathbf{p}) = 0, \quad \nabla_{\mathbf{p}} E_v(\mathbf{p}) \equiv \mathbf{v}_v(\mathbf{p}) = 0; \tag{8.6а}$$

2) критические точки второго рода:

$$\nabla_{\mathbf{p}} E_c(\mathbf{p}) = \nabla_{\mathbf{p}} E_v(\mathbf{p}) \neq 0. \tag{8.6б}$$

Как правило, условия (8.6а) выполняются только в силу тех или иных соотношений симметрии в зоне Бриллюэна. Так, например, в гл. IV было указано, что энергия электрона есть четная функция квазиимпульса и, следовательно, градиенты $\nabla_p E_c(\mathbf{p})$ и $\nabla_p E_v(\mathbf{p})$ суть функции нечетные. Отсюда явствует, что если в центре зоны Бриллюэна энергетические зоны не вырождены, то там

расположена критическая точка. В произвольных точках зоны Бриллюэна критические точки первого рода могут возникнуть только случайно. С другой стороны, равенства (8.6б) в принципе могут удовлетворяться при любых значениях $\mathbf{p}$ (при каких именно — зависит от явного вида законов дисперсии $E_c(\mathbf{p})$ и $E_v(\mathbf{p})$). По этой причине критические точки второго рода встречаются гораздо чаще, нежели первого.

Согласно формуле (V.7.8) и аналогичной ей формуле для $N_v(E)$ в критических точках первого рода особенности имеет не только комбинированная плотность состояний, но и отдельно плотности состояний в валентной зоне и в зоне проводимости. В критических точках второго рода сингулярной оказывается только комбинированная плотность состояний.

Критические точки первого рода представляют особый интерес для исследования структуры зон. Действительно, вблизи них плотности состояний $N_c(E)$ и $N_v(E)$ изменяются наиболее быстро; с другой стороны, в пространстве между ними функции $N_c(E)$ и $N_v(E)$ — гладкие, и их можно аппроксимировать путем интерполяции от одной критической точки до другой. Таким путем, определив из опыта положения критических точек первого рода, можно восстановить вид плотностей состояний — тем точнее, чем больше критических точек известно. Точно так же, определив положения ряда критических точек обоих типов, можно восстановить вид комбинированной плотности состояний. Отсюда явствует, что в качестве $\mathbf{p}_0$ в (7.3) надо выбирать именно критические точки.

Как видно из равенств (8.6а), критические точки первого рода отвечают потолку и дну зоны, а также седловым точкам функций $E_c(\mathbf{p})$ и $E_v(\mathbf{p})$. С одной из таких точек мы уже имели дело в предыдущем параграфе, рассматривая междузонные оптические переходы вблизи порога поглощения. Как видно из формулы (7.12), функция $\rho_{\text{комб}}(\omega)$ действительно имеет особенность при $\hbar\omega = E_g$: производная

$$\frac{d\rho_{\text{комб}}}{d\omega} \sim (\hbar\omega - E_g)^{-1/2} \xrightarrow[\hbar\omega \to E_g]{} \infty. \tag{8.7}$$

Критические точки второго рода отвечают экстремумам и седловым точкам только разности $E(\mathbf{p}) = E_c(\mathbf{p}) - E_v(\mathbf{p})$. Можно доказать, что периодическая функция $E(\mathbf{p})$ всегда имеет такие точки.

Исследуем поведение комбинированной плотности состояний в окрестности критической точки (безразлично, первого или второго рода). Удобно совместить с ней начало координат в $\mathbf{p}$-пространстве (в противном случае величины $\mathbf{p}$ в следующих ниже формулах надо было бы заменить на $\mathbf{p} - \mathbf{p}_0$). В силу (8.4) функцию E вблизи данной критической точки можно представить в виде

$$E(\mathbf{p}) = E(0) + \frac{1}{2} m^{-1}_{\alpha\beta} p_\alpha p_\beta + \ldots \tag{8.8}$$

Здесь многоточие обозначает члены высшего порядка малости, а величины $m^{-1}_{\alpha\beta}$ суть компоненты тензора обратной эффективной массы в точке $\mathbf{p} = 0$. Направим оси координат в $\mathbf{p}$-пространстве вдоль главных осей тензора $m^{-1}_{\alpha\beta}$, обозначив, как всегда, через m^{-1}_{α} главные значения этого тензора. Тогда вместо (8.8) мы получим

$$E(\mathbf{p}) = E(0) + \sum_{\alpha = x,\, y,\, z} \frac{p_\alpha^2}{2m_\alpha}. \tag{8.8'}$$

Эффективные массы m_α могут иметь разные знаки. В зависимости от их сочетания здесь имеются четыре случая:

1) Особая точка типа M_0 (минимум): все три величины m_α положительны.

2) Особая точка типа M_1 (седловая точка): одна из величин m_α (для определенности m_z) отрицательна, две — положительны.

3) Особая точка типа M_2 (седловая точка): две из величин m_α отрицательны, одна (для определенности m_z) — положительна.

4) Особая точка типа M_3 (максимум): все три величины m_α отрицательны.

Для вычисления функции $\rho_{\text{комб}}(\omega)$ в условиях (8.8') удобно воспользоваться выражением (7.6). Следует лишь помнить о необходимости ограничить область интегрирования окрестностью критической точки, где только и имеет смысл разложение (8.8'). Произведем в выражении (7.6) замену переменных интегрирования, полагая

$$p_\alpha = \sqrt{2}\, q_\alpha \,|\, m_\alpha \,|^{1/2}.$$

Получим

$$\rho_{\text{комб}}(\omega) = \frac{2^{5/2} \,|\, m_x m_y m_z \,|^{3/2}}{(2\pi\hbar)^3} I, \tag{8.9}$$

где

$$I = \begin{cases} \int \delta(q^2 + E_{\text{кр}} - \hbar\omega)\, d\mathbf{q}, & \text{точка } M_0, \quad (8.10\text{а}) \\ \int \delta(-q^2 + E_{\text{кр}} - \hbar\omega)\, d\mathbf{q}, & \text{точка } M_3, \quad (8.10\text{б}) \\ \int \delta(q_x^2 + q_y^2 - q_z^2 + E_{\text{кр}} - \hbar\omega)\, d\mathbf{q}, & \text{точка } M_1, \quad (8.10\text{в}) \\ \int \delta(-q_x^2 - q_y^2 + q_z^2 + E_{\text{кр}} - \hbar\omega)\, d\mathbf{q}, & \text{точка } M_2. \quad (8.10\text{г}) \end{cases}$$

Здесь

$$E_{\text{кр}} \equiv E(0) = E_c(0) - E_v(0)$$

есть разность между энергиями электрона в зоне проводимости и в валентной зоне, вычисленная в критической точке. Так, в случае поглощения у порога в прямом полупроводнике, рассмотренном в § 7, $E_{\text{кр}} = E_g$ — ширине запрещенной зоны.

Интеграл I для точки типа M_0 рассматривался в § 7. Опуская несущественные множители, не зависящие от частоты, мы имеем:

$$\text{Точка } M_0\text{:}\quad \rho_{\text{комб}}(\omega) \sim \begin{cases} (\hbar\omega - E_{\text{кр}})^{1/2}, & \hbar\omega \geqslant E_{\text{кр}}, \\ 0, & \hbar\omega < E_{\text{кр}}. \end{cases} \tag{8.11а}$$

Аналогичный результат получается и для точки типа M_3. Действительно, пользуясь соотношением (Приложение IV)

$$\delta(-q^2 + E_{\text{кр}} - \hbar\omega) = \delta(q^2 - E_{\text{кр}} + \hbar\omega),$$

мы получаем:

$$\text{Точка } M_3\text{:}\quad \rho_{\text{комб}}(\omega) \sim \begin{cases} (E_{\text{кр}} - \hbar\omega)^{1/2}, & \hbar\omega \leqslant E_{\text{кр}}, \\ 0, & \hbar\omega > E_{\text{кр}}. \end{cases} \tag{8.11б}$$

Как и в точке M_0, комбинированная плотность состояний имеет здесь корневую особенность.

Обратимся теперь к точкам типа M_1 и M_2. Введем в интеграле (8.10в) цилиндрические координаты, полагая

$$q_x = q_\perp \cos\varphi, \quad q_y = q_\perp \sin\varphi, \quad d\mathbf{q} = q_\perp\, dq_\perp\, d\varphi\, dq_z.$$

Получим

$$I = 2\pi \int\limits_{q_{\min}}^{q_{\max}} q_\perp\, dq_\perp \int dq_z \delta(E_{\text{кр}} - \hbar\omega + q_\perp^2 - q_z^2) =$$

$$= 2\pi \int\limits_{q_{\min}}^{q_{\max}} q_\perp\, dq_\perp\, (E_{\text{кр}} - \hbar\omega + q_\perp^2)^{-1/2}. \tag{8.12}$$

Нижний предел $q_{\min}$ находим из условия положительности выражения в скобках:

$$q_{\min} = \begin{cases} (\hbar\omega - E_{\text{кр}})^{1/2}, & \hbar\omega \geqslant E_{\text{кр}}, \\ 0 & \hbar\omega < E_{\text{кр}}. \end{cases} \tag{8.13}$$

Верхний предел $q_{\max}$ определяется размерами области, в которой справедливо разложение (8.8′); как будет видно из дальнейшего, точное его значение несущественно.

Вычисляя интеграл (8.12), мы получаем:

$$\text{Точка } M_1\text{:}\quad \rho_{\text{комб}}(\omega) \sim \begin{cases} (E_{\text{кр}} - \hbar\omega - q_{\max}^2)^{1/2} - (E_{\text{кр}} - \hbar\omega)^{1/2}, \\ \qquad\qquad \hbar\omega \leqslant E_{\text{кр}}, \\ (E_{\text{кр}} - \hbar\omega - q_{\max}^2)^{1/2} \quad \hbar\omega > E_{\text{кр}}. \end{cases} \tag{8.11в}$$

Таким образом, здесь комбинированная плотность состояний имеет излом в точке $\hbar\omega = E_{\text{кр}}$, точное значение $q_{\max}$ при этом роли не играет.

Аналогично вычисляется и интеграл (8.10г). Мы имеем:

$$\text{Точка } M_2\text{:}\quad \rho_{\text{комб}}(\omega) \sim \begin{cases} (\hbar\omega - E_{\text{кр}} + q^2_{\max})^{1/2} & \hbar\omega < E_{\text{кр}}, \\ (\hbar\omega - E_{\text{кр}} + q^2_{\max})^{1/2} - (\hbar\omega - E_{\text{кр}})^{1/2}, & \hbar\omega \geqslant E_{\text{кр}}. \end{cases} \quad (8.11\text{г})$$

Здесь также имеется излом в точке $\hbar\omega = E_{\text{кр}}$.

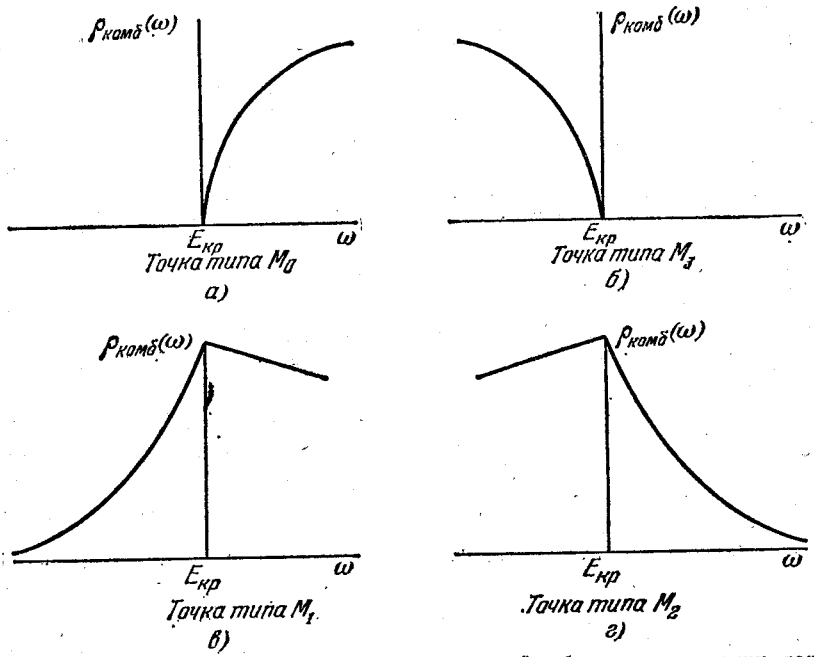


Рис. 18.9. Комбинированная плотность состояний вблизи критических точек (схематически).

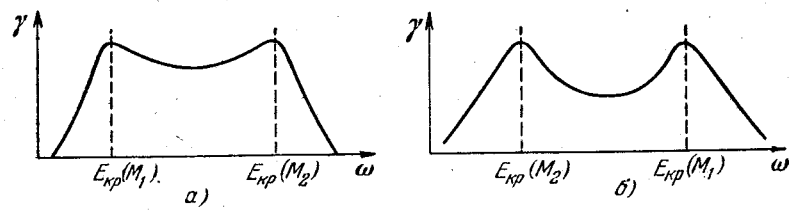


Рис. 18.10. Плато коэффициента поглощения, возникающее при наличии двух критических точек (схематически).

На рис. 18.9 схематически представлен ход комбинированной плотности состояний вблизи критических точек всех четырех

типов. Как видно из рис. 18.9, *в*, *г*, при наличии двух близких по энергии критических точек типа M_1 и M_2 на кривой $\gamma(\omega)$ может возникнуть плато (рис. 18.10). Такие плато действительно иногда наблюдаются.

§ 9. Непрямые переходы

Рассмотрим непрямые междузонные переходы в веществах типа Ge и Si. Ограничимся при этом невырожденными материалами и областью частот вблизи пороговой. Функцию распределения электронов по энергиям будем считать равновесной: $f_0(E_c(\mathbf{p})) \approx 0$, $f_0(E_v(p)) \simeq 1$.

В рассматриваемых веществах точка $\mathbf{p}_0$, в которой расположен главный минимум зоны проводимости, находится недалеко от границы зоны Бриллюэна. Это означает, что непрямой переход, изображенный на рис. 18.4,*а*, связан с изменением квазиимпульса на величину порядка $\hbar/a$, где a — постоянная решетки. Как видно из формулы (XIV.5.18), вероятность передачи квазиимпульса $|\mathbf{p} - \mathbf{p}'|$ при рассеянии на заряженной примеси пропорциональна

$$\frac{N_t Z^2 e^4}{\varepsilon^2 [\hbar^2 r_0^{-2} - (\mathbf{p} - \mathbf{p}')^2]^2}, \tag{9.1}$$

где N_t — концентрация примеси, заряд иона которой (в единицах e) равен Z, а r_0 — радиус экранирования. В не слишком сильно легированных материалах $r_0 \gg a$, и выражение (9.1) оказывается пропорциональным малой величине

$$N_t a^4 a_B^{-1},$$

где, как всегда, через a_B обозначен боровский радиус в кристалле. С другой стороны, согласно (XIV.4.15) вероятность рассеяния на акустических колебаниях решетки слабо зависит от величины передаваемого квазиимпульса. По этим причинам при не слишком большой концентрации заряженной примеси рассеяние носителей заряда на ней не играет заметной роли при непрямых междузонных переходах. Доминирующим — при всех температурах, представляющих экспериментальный интерес, — оказывается рассеяние на акустических колебаниях решетки *). При этом возможны процессы, связанные как с поглощением, так и с испусканием фононов. В большинстве интересных полупроводников энергия взаимодействия электронов с фононами сравнительно невелика. Это позволяет ограничиться рассмотрением только однофононных процессов. Обозначим частоту и квазиволновой вектор фонона через ω_Φ и $\mathbf{q}$. Тогда

*) Положение меняется в сильно легированных полупроводниках, когда из-за взаимодействия с атомами примеси изменяется сама структура энергетического спектра носителей заряда (гл. XIX).

законы сохранения энергии и квазиимпульса при междузонных переходах с поглощением или испусканием фонона будут иметь вид

$$E_v(\mathbf{p}) + \hbar\omega + \hbar\omega_\Phi(\mathbf{q}) = E_c(\mathbf{p}'), \qquad \mathbf{p} + \hbar\mathbf{q} = \mathbf{p}' \tag{9.2а}$$

и

$$E_v(\mathbf{p}) + \hbar\omega - \hbar\omega_\Phi(\mathbf{q}) = E_c(\mathbf{p}'), \qquad \mathbf{p} - \hbar\mathbf{q} = \mathbf{p}'. \tag{9.2б}$$

Рассматриваемые квантовые переходы происходят в результате взаимодействия электронов как с фотонами, так и с фононами. Первое из этих взаимодействий описывается гамильтонианом (4.5), второе — формулой (XIV.3.32). Расчет коэффициента поглощения с учетом обоих этих взаимодействий требует довольно длинных вычислений, которые можно найти, например, в обзоре [6]. Общую структуру выражения для γ можно, однако, установить без всяких вычислений.

Во-первых, ясно, что мы получим здесь сумму двух членов, γ_+ и γ_-, соответствующих поглощению фотона с испусканием и поглощением фонона:

$$\gamma(\omega) = \gamma_+(\omega) + \gamma_-(\omega). \tag{9.3}$$

Во-вторых, в силу законов сохранения (9.2а) и (9.2б) и соображений, связанных с принципом Паули, для γ_+ и γ_- должны получиться выражения типа (7.2). Нужно лишь изменить в (7.2) аргумент δ-функции, включив в него энергию фонона с квазиимпульсом $\pm(\mathbf{p} - \mathbf{p}')$, ввести под знак суммы дополнительный множитель, описывающий вероятность испускания или поглощения фонона, и просуммировать по всем ветвям фононного спектра.

Как и при прямых междузонных переходах здесь следует различать переходы разрешенные и запрещенные — в соответствии с тем, не равно или равно нулю первое слагаемое в правой части формулы (7.3). Рассмотрим сначала разрешенные переходы. Опуская несущественные постоянные множители, мы получим

$$\gamma_\pm(\omega) \sim \sum_s C_s^{-1} \{n(\omega_{\Phi, s})\, I_- + [n(\omega_{\Phi, s}) + 1]\, I_+\}. \tag{9.4}$$

Здесь C_s — константы, описывающие энергию взаимодействия электронов с фононами s-й ветви и указанные в таблице 14.2; $\omega_{\Phi, s}$ есть частота фонона s-й ветви, квазиволновой вектор которого равен $\mathbf{p}_0/\hbar$; $n(\omega_{\Phi, s}) = \left[\exp\frac{\hbar\omega_{\Phi, s}}{kT} - 1\right]^{-1}$ — равновесное число фононов данного типа, а через $I_\pm$ обозначены интегралы

$$I_\pm = \int \delta(E_c(\mathbf{p}') - E_v(\mathbf{p}') - \hbar\omega \pm \hbar\omega_{\Phi, s})\, d\mathbf{p}\, d\mathbf{p}'. \tag{9.5}$$

Правую часть (9.5) легко представить в виде интеграла от произведения плотностей состояний в валентной зоне и в зоне проводимости. Действительно, введем промежуточное интегрирование по

вспомогательной энергетической переменной E, пользуясь тождеством

$$\int \delta(E_c(\mathbf{p}') - E_v(\mathbf{p}) - \hbar\omega \pm \hbar\omega_{\Phi,s})\, d\mathbf{p}\, d\mathbf{p}' =$$
$$= \int_{-\infty}^{+\infty} dE \int \delta(E_c(\mathbf{p}') - E)\, d\mathbf{p}' \int \delta(E - E_v(\mathbf{p}) - \hbar\omega \pm \hbar\omega_{\Phi,s})\, d\mathbf{p}. \quad (9.6)$$

Принимая теперь во внимание равенство (V.7.8 б), находим

$$I_{\pm} \sim \int_{-\infty}^{+\infty} dE\, N_c(E)\, N_v(E - \hbar\omega \pm \hbar\omega_{\Phi,s}). \quad (9.7)$$

Вычислим интегралы (9.7) для параболических (не обязательно изотропных) законов дисперсии (по-прежнему совмещая начало отсчета энергии с дном зоны проводимости). В этом случае (§ V.2)

$$N_c(E) \sim \begin{cases} E^{1/2}, & E \geqslant 0, \\ 0, & E < 0 \end{cases}$$

и

$$N_v(E - \hbar\omega \pm \hbar\omega_{\Phi,s}) \sim$$
$$\sim \begin{cases} (\hbar\omega - E_g - E \mp \hbar\omega_{\Phi,s})^{1/2}, & \hbar\omega - E_g - E \mp \hbar\omega_{\Phi,s} \geqslant 0, \\ 0, & \hbar\omega - E_g - E \pm \hbar\omega_{\Phi,s} < 0. \end{cases}$$

Подставляя эти выражения в правую часть (9.7), получаем

$$I_{\pm} \sim \int_0^{\hbar\omega - E_g \mp \hbar\omega_{\Phi,s}} [E(\hbar\omega - E_g - E \mp \hbar\omega_{\Phi,s})]^{1/2}\, dE =$$
$$= \frac{\pi}{8}(\hbar\omega - E_g \mp \hbar\omega_{\Phi,s})^2. \quad (9.8)$$

Таким образом, для коэффициента поглощения при непрямых разрешенных переходах находим

$$\gamma \sim \sum_s C_s^{-1} \{n(\omega_{\Phi,s})(\hbar\omega - E_g + \hbar\omega_{\Phi,s})^2 +$$
$$+ [n(\omega_{\Phi,s}) + 1](\hbar\omega - E_g - \hbar\omega_{\Phi,s})^2\}. \quad (9.9)$$

Как и следовало ожидать, в отличие от прямых переходов, правая часть (9.9) зависит от температуры.

Формула (9.9) упрощается, если энергия $\hbar\omega_{\Phi,s}$ мала по сравнению с $\hbar\omega - E_g$ *). Тогда

$$\gamma \sim (\hbar\omega - E_g)^2 \sum_s C_s^{-1} \operatorname{cth} \frac{\hbar\omega_{\Phi,s}}{2kT}. \quad (9.9')$$

*) Это условие может и не составлять большого ограничения, так как предельная энергия акустического фонона обычно не превышает нескольких сотых долей электронвольта.

При правильно выбранных значениях C_s выражение (9.9) хорошо описывает температурную и частотную зависимость коэффициента поглощения в германии и кремнии (вблизи порога). Частотная зависимость вида (9.9′) наблюдалась в фосфиде галлия.

§ 10. Электрооптика

При наложении на образец постоянного и однородного электрического поля энергетический спектр носителей заряда претерпевает серьезные изменения (§ IV.6). Соответственно изменяются и спектры поглощения и испускания света полупроводником. Самое характерное их отличие от того, что наблюдается в отсутствие поля, ясно из рис. 4.8: при наложении постоянного и однородного электрического поля становятся возможными междузонные оптические переходы при частоте света, меньшей красной границы *). Это — переходы, невертикальные в плоскости (z, E). Они связаны с туннельным просачиванием электронов и дырок через запрещенную зону. Из рис. 4.8 видно, что энергия поглощаемого фотона $\hbar\omega$ тем меньше, чем на более далекое расстояние туннелирует электрон. Отсюда следует, что вероятность такого перехода, будучи отличной от нуля при любой частоте ω, должна все же быстро убывать с увеличением разности $E_g - \hbar\omega$. Расчет этой вероятности связан с довольно сложными вычислениями. Мы приведем лишь результат для простейшего случая, когда переходы — разрешенные и происходят без участия фононов, причем главные оси тензора эффективной массы — одни и те же для электронов и для дырок. Напряженность электрического поля будем считать направленной вдоль одной из этих осей (оси Z). Введем характерную энергию

$$E_r = \left(\frac{e^2\mathscr{E}^2\hbar^2}{2m_{zr}}\right)^{1/3}, \quad m_{z,r}^{-1} = m_{z,c}^{-1} + m_{z,v}^{-1}. \tag{10.1}$$

Тогда при $\hbar\omega < E_g$ и $E_g - \hbar\omega \gg E_r$ коэффициент поглощения дается выражением [6]

$$\gamma \simeq \frac{AE_r^{3/2}}{\omega(E_g - \hbar\omega)} \exp\left\{-\frac{4}{3}\left(\frac{E_g - \hbar\omega}{E_r}\right)^{3/2}\right\}, \tag{10.2}$$

где A — постоянная.

В соответствии со сказанным выше коэффициент поглощения действительно оказывается отличным от нуля при $\hbar\omega < E_g$, но все же быстро убывает с увеличением параметра $(E_g - \hbar\omega)/E_r$.

С другой стороны, при $\hbar\omega > E_g$ и $\hbar\omega - E_g \gg E_r$ получается

$$\gamma \simeq \frac{A_1}{\omega}(\hbar\omega - E_g)^{1/2}\left\{1 - \frac{E_r^{3/2}}{4(\hbar\omega - E_g)^{3/2}}\cos\left[\frac{4}{3}\left(\frac{\hbar\omega - E_g}{E_r}\right)^{3/2}\right]\right\}, \tag{10.3}$$

*) Это явление называют эффектом Келдыша—Франца.

где A_1 — постоянная. При $\mathcal{E} \to 0$ это выражение (с учетом (10.1)) переходит в формулу (7.13). Поправка, определяемая вторым слагаемым в фигурных скобках в (10.3), оказывается осциллирующей функцией частоты света. Заметим, однако, что условия применимости последнего результата могут оказаться довольно жесткими. Дело в том, что, поднявшись достаточно высоко над нижним краем зоны проводимости, электрон получает возможность участвовать в ряде процессов рассеяния — как упругих, так и неупругих. В результате их рассматриваемое состояние электрона становится нестационарным, что приводит к «размазыванию» осцилляционных пиков коэффициента поглощения.

Как видно из формулы (10.2), коэффициент поглощения света в области $\hbar\omega < E_g$ сильно зависит от разности $E_g - \hbar\omega$. Это позволяет, измеряя $\gamma(\omega, \mathcal{E})$ *), определять ширину запрещенной зоны. Такой способ применим и к «непрямым материалам»; там он дает ширину «прямой зоны», т. е. энергетическое расстояние между потолком валентной зоны и уровнем зоны проводимости, находящимися в той же точке зоны Бриллюэна. Формулы (10.1) — (10.3) позволяют также, зная E_g, определять приведенную эффективную массу.

§ 11. Модуляционная спектроскопия

Чувствительность рассмотренных в предыдущем параграфе методов исследования резко повышается, если модулировать постоянное поле сравнительно слабым переменным. При этом модулированной окажется и интенсивность света, прошедшего через образец или отраженного от него. Выделяя переменную составляющую интенсивности, мы получим сигнал, пропорциональный производной коэффициента поглощения (или отражения) по напряженности поля $\mathcal{E}$.

В сущности, не обязательно модулировать именно напряженность электрического поля. В принципе можно периодически варьировать любую величину, от которой зависит коэффициент поглощения (как в присутствии, так и в отсутствие постоянного поля). Выделяя переменную составляющую интенсивности прошедшего или отраженного света, мы найдем производную γ или R по этой величине. Эта идея лежит в основе метода модуляционной спектроскопии. Кроме электроотражения, используются еще следующие методики:

а) Пьезоотражение. Модулируется (например, акустически) механическое напряжение на образце. В результате периодически изменяются значения всех энергетических зазоров — ширины запре-

*) Непосредственно измеряют чаще коэффициент отражения, вычисляя затем γ по формулам (1.23) и (1.11).

щенной зоны, расстояния между потолком валентной зоны и подзоной, отщепленной благодаря спин-орбитальному взаимодействию, и т. д. Исследование переменной составляющей интенсивности отраженного света дает в этом случае информацию о зависимости ширины запрещенной зоны и других величин от давления.

б) Термоотражение. Модулируется температура образца. При этом будут периодически изменяться и все энергетические зазоры (с разными, вообще говоря, коэффициентами). В результате получается информация о температурной зависимости названных величин.

в) Оптическая модуляция отражения (фотоотражение). Эффект Келдыша—Франца имеет место в электрическом поле любого происхождения — лишь бы оно достаточно плавно изменялось в пространстве (т. е. могло рассматриваться как почти однородное). Это относится и к полю, возникающему вблизи поверхности благодаря изгибу зон. Величину его можно изменять, варьируя концентрацию носителей заряда вблизи поверхности. Последнее достигается оптической инжекцией, при которой электроны и дырки создаются сильно поглощаемым светом (не тем, отражение которого изучается). Исследования этого типа дают ту же информацию, что и электроотражение.

г) Катодоотражение. Идея этого метода та же, что и фотоотражения. Разница состоит в том, что теперь электронно-дырочные пары создаются пучком быстрых электронов («катодных лучей»).

д) Модуляция длины волны. При этом определяется производная от коэффициента отражения (поглощения) по частоте, что полезно при обнаружении критических точек. Действительно, согласно формулам (7.7) и (8.7) в критической точке производная $d\gamma/d\omega$ имеет резкий пик.

Подробное рассмотрение методов модуляционной спектроскопии и изложение результатов исследования, полученных таким путем, можно найти в книге [7].

§ 12. Магнетооптика

Влияние магнитного поля на оптические свойства полупроводника связано как с квантованием Ландау, так и с чисто классическими эффектами. С одним из магнетооптических явлений — диамагнитным резонансом — мы уже встречались в гл. IV. Отметим еще две группы эффектов.

а. Магнетоплазменные эффекты. Газ свободных носителей заряда в полупроводнике представляет собой частный случай плазмы — в целом нейтральной системы заряженных частиц (в монополярном материале нейтральность достигается за счет компенсирующего заряда примесей). Магнетоплазменными называют оптические эффекты, обусловленные влиянием неквантующих магнитных

полей на поведение свободных носителей заряда. Это влияние проявляется через комплексную электропроводность. Согласно § XIII.1в в магнитном поле электропроводность становится тензором, компоненты которого зависят от магнитной индукции $\mathscr{B}$. Особенно интересен случай сравнительно слабого поглощения ($\varkappa \ll 1$), когда при вычислении коэффициентов преломления и отражения поглощением энергии можно пренебречь (о такой системе свободных зарядов говорят как о бесстолкновительной плазме). В этих условиях при параболическом законе дисперсии величины n и R выражаются только через эффективные массы носителей заряда, что и позволяет использовать соответствующие экспериментальные данные для определения компонент тензора $m^{-1}_{\alpha\beta}$. Подробное рассмотрение всех магнетоплазменных эффектов можно найти в статье [8].

б. Междузонные переходы в квантующем магнитном поле. Коэффициент поглощения света при междузонных переходах в квантующем магнитном поле можно вычислить по формулам (4.13) (если ограничиться рассмотрением идеальной решетки). Следует лишь, в соответствии с § IV.5, использовать в качестве λ' и λ квантовые числа k_2, k_3 и n в конечном и начальном состояниях электрона. При этом осцилляции плотности состояний, отмеченные в § V.8, приводят к осциллирующей зависимости коэффициента поглощения от магнитной индукции (при заданной частоте света) или от частоты (при заданной магнитной индукции). В невырожденных зонах пики коэффициента поглощения (без изменения проекции спина) определяются условием, вытекающим из формулы (V.8.7) и такого же выражения для плотности состояний в валентной зоне:

$$\hbar\omega = E_g + \hbar\,(\omega_n - \omega_p)\,(n + {}^1/_2) \pm \beta\,(g_n - g_p)\,\mathscr{B}. \qquad (12.1)$$

Здесь

$$\omega_n = \frac{e\mathscr{B}}{m_n c}, \qquad \omega_p = \frac{e\mathscr{B}}{m_p c},$$

g_n и g_p — гиромагнитные отношения для электронов проводимости и дырок.

Согласно (12.1) экспериментальное исследование междузонных переходов в магнитном поле позволяет определять параметры $\omega_n - \omega_p$ и $g_n - g_p$. Существенно, что при этом мы не обязаны ограничиваться областью энергий, близких к границам запрещенной зоны. Подбирая значение магнитной индукции, можно вести измерения в области частот, заметно превышающих $E_g/\hbar$. При этом получается информация о структуре зоны проводимости (валентной зоны) вдали от ее дна (потолка).

Глава XIX

СИЛЬНО ЛЕГИРОВАННЫЕ ПОЛУПРОВОДНИКИ

§ 1. Примесные уровни и примесные зоны

Определение понятия «сильно легированный полупроводник» можно дать, выяснив особенности материалов, содержащих достаточно много посторонней примеси *). Для этой цели посмотрим прежде всего, как изменяется зонный энергетический спектр при добавлении примеси в кристалл.

Как мы знаем (гл. II и § IV.7), изолированные примесные атомы могут создавать дискретные уровни в запрещенной зоне. Волновые функции электронов, занимающих эти уровни, локализованы вблизи соответствующих атомов примеси. Энергетический спектр такого полупроводника схематически изображен на рис. 19.1, *а* (для определенности мы считаем, что каждому примесному атому принадлежит лишь один дискретный уровень, и рассматриваем материал n-типа).

Электроны и дырки, локализованные на дискретных уровнях, могут перемещаться по кристаллу лишь путем «прыжков» — перескоков с одного уровня на другой. При этом для преодоления потенциального барьера, разделяющего примесные атомы, требуется энергия активации (даже при учете туннельного эффекта). Поэтому, а также из-за больших — в среднем — расстояний между атомами примеси, вероятность такого процесса оказывается небольшой и значения подвижности, соответствующие прыжковой проводимости, также весьма малы (порядка десятых долей $см^2/В \cdot с$ и меньше). Прыжковую проводимость удается наблюдать лишь при достаточно низких температурах, когда концентрация свободных носителей заряда становится совсем небольшой. Вместе с тем температура не должна быть и слишком низкой: при $T = 0$ тепловая активация становится невозможной и прыжковая проводимость (в отсутствие подсветки) обращается в нуль.

Представление об изолированных примесных атомах оправдано, лишь если не перекрываются ни их силовые поля, ни волновые функции электронов, локализованных на дискретных уровнях. Так обстоит дело, если концентрация примеси N_t, боровский радиус

*) Роль примеси могут играть и иные структурные дефекты решетки — междоузельные атомы, вакансии и т. д.

в кристалле a_B и радиус экранирования r_0 удовлетворяют условиям

$$N_t^{-1/3} \gg r_0, \quad N_t^{-1/3} \gg a_B. \tag{1.1}$$

В наиболее интересном для дальнейшего случая «водородоподобных» примесей (§ IV.7) $a_B = \varepsilon\hbar^2/me^2$. Этой формулой мы и

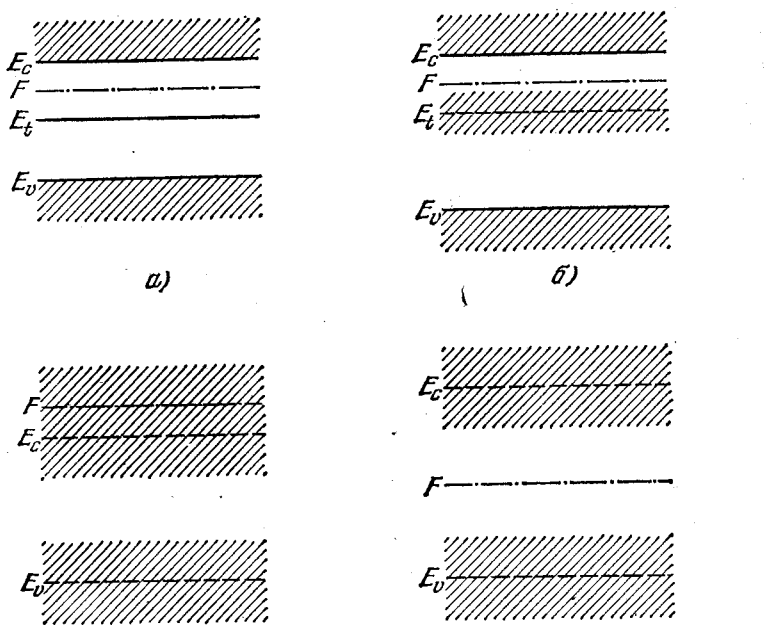


Рис. 19.1. Энергетический спектр полупроводника n-типа в зависимости от уровня легирования (заштрихованы области дозволенных значений энергии). *а*) Слабо легированный полупроводник. Штрихпунктиром указано положение уровня Ферми при низких температурах. *б*) Промежуточно легированный полупроводник. Пунктирная линия указывает положение бывшего уровня примеси, штрихпунктирная — положение уровня Ферми при низких температурах. *в*) Сильно легированный полупроводник. Пунктирными линиями указаны бывшие границы зон проводимости и валентной, штрихпунктирной — уровень Ферми в отсутствие компенсации. *г*) Сильно легированный компенсированный полупроводник.

будем далее пользоваться. При этом второе из неравенств (1.1) есть частный случай (IV.7.8) при квантовом числе $n = 1$. Полупроводники, параметры которых удовлетворяют неравенствам (1.1), называются *слабо легированными*. С ними мы имели дело, в частности, в §§ IV.7 и XIV.5, изучая энергетический спектр мелких доноров и акцепторов и рассеяние на заряженной примеси.

Заметим, что в этом случае радиус экранирования должен превышать «радиус орбиты», иначе отдельный атом примеси вообще не мог бы создать дискретный уровень (§ XVII.8). В справедливости неравенства

$$r_0 > a_B \tag{1.2}$$

в слабо легированных полупроводниках можно убедиться, пользуясь явными выражениями для радиуса экранирования (П.XII.6а, б) (эти выражения справедливы, если экранирование обусловлено только перераспределением свободных зарядов). В отсутствие вырождения и при полном вырождении мы имеем, соответственно,

$$\frac{r_0}{a_B} = \left(\frac{1}{8\pi}\frac{1}{na_B^3}\frac{kT}{E_B}\right)^{1/4} \tag{1.3а}$$

и

$$\frac{r_0}{a_B} = \frac{1}{2}\left(\frac{\pi}{3}\right)^{1/6}(n^{1/3}a_B)^{-1/2}. \tag{1.3б}$$

Здесь n — концентрация электронов, $E_B = me^4/2\varepsilon^2\hbar^2$ — боровская энергия в кристалле.

Полагая теперь $n = N_t$ и пользуясь вторым из неравенств (1.1), видим, что соотношение (1.2) удовлетворяется в широком интервале концентраций примеси.

При повышении концентрации примеси сильные неравенства (1.1) рано или поздно перестанут выполняться. Прежде всего нарушится первое из них: электрон, локализованный вблизи одного из атомов примеси, начнет испытывать воздействие и со стороны других примесных атомов. В результате его энергетический уровень, оставаясь дискретным (пока соблюдается второе из неравенств (1.1)), несколько сдвинется по энергии. Величина этого сдвига зависит от расположения других примесных атомов относительно центра локализации: она тем больше, чем больше атомов примеси отстоит от данного на расстояние, не превышающее примерно r_0. Однако распределение примеси в решетке никогда не бывает строго упорядоченным (см. Приложение XIII). Даже при постоянной по образцу средней концентрации примеси всегда имеют место локальные флуктуации концентрации. Поэтому и сдвиг энергии примесного уровня относительно E_c оказывается случайным и различным в разных местах образца. Иначе говоря, в запрещенной зоне вместо одного дискретного уровня появляется некоторый их набор (рис. 19.1, *б*). Это явление называется *классическим уширением уровней*.

При дальнейшем увеличении концентрации примеси знак строгого неравенства исчезает и во втором из условий (1.1). Тогда становится заметным перекрытие волновых функций электронов,

локализованных на различных (в основном соседних) атомах примеси, и примесный уровень «размывается» в примесную зону (§ IV.7). Об этом эффекте говорят как о *квантовом уширении уровней*.

Полупроводники, в которых образовалась примесная зона, отделенная запрещенными участками как от валентной зоны, так и от зоны проводимости, называют *промежуточно легированными*.

В германии, легированном донорами V группы, эффекты, связанные с образованием примесной зоны, становятся экспериментально заметными при $N_t \simeq 10^{15}$ см$^{-3}$.

При дальнейшем увеличении концентрации примеси примесная зона продолжает расширяться и в конце концов сливается (в образце n-типа) с зоной проводимости (рис. 19.1, *в*). Энергия активации примеси при этом обращается в нуль. В таком материале уже нельзя проводить четкое различие между зоной проводимости и примесной: имеется единая область дозволенных значений энергии, проникающая в глубь запрещенной зоны *). Эту область мы тоже будем называть «примесной».

Полупроводник, в котором примесная зона слилась с ближайшей к ней «собственной» зоной кристалла, называется *сильно легированным*.

Более тщательное исследование (см. ниже § 2) показывает, что даже при легировании примесью только одного типа новая область спектра возникает у обеих собственных зон кристалла. Компенсация приводит лишь к расширению этой области и к очевидному смещению уровня Ферми (рис. 19.1, *г*).

§ 2. Особенности сильно легированных полупроводников

Две особенности сильно легированных полупроводников видны непосредственно из рис. 19.1, *в*, *г*.

Во-первых, в некомпенсированном материале уровень Ферми попадает в дозволенную область энергий: в сильно легированном некомпенсированном полупроводнике имеет место вырождение газа основных носителей заряда. По этой причине такие материалы часто называют вырожденными.

Во-вторых, в сильно (и промежуточно) легированном материале взаимодействие электронов и дырок с примесями играет двоякую роль: формируя определенные участки энергетического спектра системы, оно вместе с тем обусловливает и рассеяние носителей заряда (в том числе и тех, которые движутся в «примесных» областях спектра). С другой стороны, в условиях слабого легирования это взаимодействие играет альтернативную роль: оно либо создает

*) Здесь и в аналогичных случаях в дальнейшем термины «запрещенная зона», «эффективная масса» и т. д. относятся к соответствующему нелегированному полупроводнику.

локальные уровни, находясь на которых электроны и дырки сами не участвуют в явлениях переноса (исключая прыжковую проводимость), либо вызывает рассеяние свободных носителей заряда, находящихся в зонах проводимости и валентной. При этом на структуру указанных зон взаимодействие электронов и дырок с ионами примеси практически не оказывает влияния.

Третья особенность сильно легированных полупроводников связана с обстоятельством, характерным для всех явлений, в которых играет роль взаимодействие носителей заряда с примесными атомами. Как отмечалось в гл. XIV, наблюдаемые на опыте величины представляют собой результат некоторого усреднения по координатам атомов примеси (о совокупности этих координат говорят как о конфигурации примеси). Действительно, из-за флуктуаций в распределении примеси ее конфигурация в разных частях образца оказывается несколько различной; выполняя измерения, мы тем самым производим усреднение по конфигурациям. Именно поэтому экспериментальные данные, относящиеся к электрическим, оптическим и другим характеристикам вещества, оказываются связанными не с координатами каждого из атомов примеси, а лишь с такими величинами, как средняя их концентрация. Чтобы ввести операцию усреднения в аппарат теории, удобно рассматривать координаты примесных атомов как случайные величины и задавать не точные их значения, а лишь вероятность того или иного распределения примеси в пространстве. Этот подход точно соответствует той информации, которой мы фактически располагаем: координаты атомов примеси в опыте не заданы, а о характере их распределения судить можно.

Возможность — и необходимость — такой статистической постановки задачи связана с тем, что в рассматриваемых материалах не имеет места полное термодинамическое равновесие. Действительно, при полном равновесии реализовалось бы упорядоченное распределение примеси, отвечающее условию минимальности свободной энергии системы. Для этого, однако, примесные атомы должны продиффундировать из тех мест, куда они случайно попали при легировании, в положения равновесия. При обычных температурах этот процесс идет очень медленно и распределение примеси остается случайным.

Статистическая постановка задачи о взаимодействии носителей заряда с примесными атомами уже использовалась нами в § XIV.5. Там, однако, операция усреднения носила тривиальный характер, сводясь просто к умножению вероятности рассеяния одним атомом примеси на полное их число. Дело в том, что в условиях слабого легирования (1.1) в каждый данный момент электрон, как правило, взаимодействует только с одним из атомов примеси. При этом энергия взаимодействия остается всегда одной и той же, несмотря на случайный характер распределения примеси в пространстве

(рис. 19.2, *а*). При промежуточном и, особенно, сильном легировании положение меняется: потенциальная энергия носителя заряда зависит уже от координат сразу нескольких атомов примеси и потому становится случайной величиной (рис. 19.2, *б*). Силовые поля, характеристики которых заданы только статистически, называются случайными.

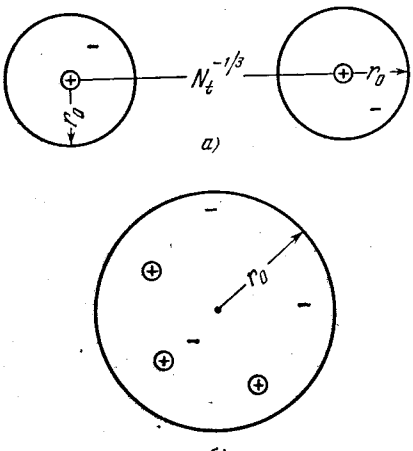


Рис. 19.2. Случайное распределение примеси и случайный характер потенциальной энергии. *а*) Слабое легирование: распределение примеси случайно, но потенциальная энергия электрона в поле примеси не случайна. Символом «+» обозначены атомы примеси, символом «—» — электроны. *б*) Промежуточное и сильное легирование: случайное распределение примеси приводит к случайному характеру потенциальной энергии электрона. Символом «+» обозначены атомы примеси, символом «—» — электроны.

Итак, третья особенность сильно (и промежуточно) легированных полупроводников состоит в том, что носители заряда в них движутся в случайном поле примеси. Это поле представляет собой совокупность случайно расположенных трехмерных потенциальных ям и горбов случайной высоты и формы (см. ниже рис. 19.5). Ямы можно наглядно представлять себе как результат случайных скоплений примесных ионов в областях, линейные размеры которых меньше r_0; горбам соответствуют области обеднения. Такие скопления иногда называют *кластерами* *).

Случайное поле рассматриваемого типа (даже при одном типе примеси) действует на носители заряда обоих знаков. Действительно, потенциальная яма для электрона есть вместе с тем и потенциальный горб для дырки, и наоборот. Именно по этой причине примесные области спектра возникают, вообще говоря, как у верхней, так и у нижней границы запрещенной зоны.

Четвертая особенность сильно (и промежуточно) легированных полупроводников вытекает из второй и третьей. Она состоит в том, что носители заряда в них, вообще говоря, нельзя характеризовать каким бы то ни было законом дисперсии $E(\mathbf{p})$ **). Действительно, квазиимпульс $\mathbf{p}$ сохраняется и, следовательно, характеризует стационарные состояния электрона только в идеально периодиче-

*) Английское слово «cluster» означает «скопление».

**) В этом состоит одно из важных различий между примесной зоной и энергетической зоной в чистом материале.

ском поле. При наличии еще случайного поля происходит рассеяние, в результате которого квазиимпульс изменяется во времени. Иначе говоря, состояния с заданными значениями p оказываются нестационарными. В принципе так обстоит дело при любых процессах рассеяния. Однако, как мы увидим в § 3, в рассматриваемых материалах не при всех существенных значениях энергии электрона выполняется условие малости неопределенности энергии $\hbar/\tau(E)$ по сравнению с самой этой энергией:

$$\frac{\hbar}{\tau(E)} \ll E - E_c \quad \text{(для электронов)}$$

или (2.1)

$$\frac{\hbar}{\tau(E)} \ll E_v - E \quad \text{(для дырок)}.$$

Между тем это условие необходимо для того, чтобы можно было рассматривать состояния с данным квазиимпульсом как почти стационарные (ср. § XIV.2). В условиях, когда неравенство (2.1) не выполняется, энергию взаимодействия, ответственного за столь малые значения времени релаксации, приходится принимать во внимание уже при определении энергетического спектра носителей заряда. В рассматриваемом сейчас случае речь идет о взаимодействии электронов и дырок со случайным полем. Этот фактор особенно важен в области энергий $E_v \leqslant E \leqslant E_c$: в отсутствие примеси там вообще нет дозволенных состояний. С другой стороны, по мере удаления от границы зоны проводимости вверх (или от границы валентной зоны вниз), когда растет полная энергия электрона или дырки, роль случайного поля должна постепенно ослабевать. При достаточно больших значениях разностей $E - E_c$ и $E_v - E$ условия (2.1) будут выполнены. Это означает, что достаточно глубоко в зонах проводимости и валентной представление о законе дисперсии носителей заряда вновь обретает приближенный смысл.

Наконец, пятая особенность сильно легированных полупроводников связана с ролью экранирования поля заряженных атомов примеси. Причины, обусловливающие эффект экранирования, могут быть различными: перераспределение свободных носителей заряда в пространстве (§ VI.7), преимущественное расположение отрицательно заряженных ионов примеси около положительных (при наличии компенсации) и т. д. Для нас сейчас существенны не причины эффекта, а следствия из него. Именно, как мы видели в XVII.8, при $r_0 \lesssim a_B$, экранирование приводит к исчезновению дискретных уровней, создаваемых примесным ионом. Согласно формулам (1.3а, б) именно такая ситуация может возникнуть в сильно легированных полупроводниках. Действительно, в отсутствие компенсации в названных материалах $n \simeq N_t$ и условие $r_0 \lesssim a_B$ достигается (в вырожденном материале) при $N_t \simeq a_B^{-3}$. Аналогичное —

по порядку величины — условие получается и для компенсированных полупроводников. Экранирование там может быть обусловлено, например, второй из указанных выше причин; при этом для оценки отношения r_0/a_B можно воспользоваться формулой (1.3а), заменяя в ней n на N_t, а температуру опыта T — на температуру, при которой примеси вводятся в решетку.

Очевидно, в условиях исчезновения примесных уровней не может возникнуть и примесная область спектра. Следует, однако, иметь в виду, что рассматриваемая сейчас задача отличается от изучавшейся в § XVII.8. Там исследовался энергетический спектр примесного атома при условии, что в образце уже создана (например, подсветкой) определенная концентрация экранирующих свободных носителей заряда. Здесь эти носители поставляются в основном самими атомами примеси. Предположив заранее, что все они ионизованы, мы действительно придем к выводу об исчезновении создаваемых ими локальных уровней. Если, однако, предположить, что электроны в некомпенсированном полупроводнике уже занимают примесные уровни, то эффект экранирования резко ослабнет: атомы примеси окажутся нейтральными, а число свободных зарядов будет невелико. Это означает, что мы не вправе заранее предполагать ни того, ни другого. Дело в том, что роль экранирования определяется концентрациями свободных носителей заряда и заряженных атомов примеси, а эти величины зависят от характера энергетического спектра системы — от того, существуют ли, и в каком количестве, примесные уровни. С другой стороны, сам энергетический спектр надо определять с учетом экранирования. Иначе говоря, энергетический спектр сильно легированного полупроводника следует определять самосогласованным путем.

В этом и состоит пятая особенность названных материалов. Она приводит к необходимости ставить задачу об энергетическом спектре сильно легированного полупроводника в рамках современной теории многих тел. Вообще говоря, это приводит к большим математическим трудностям. Избежать их удается, лишь если какие-либо типы взаимодействия оказываются сравнительно несущественными. При этом важно заранее представить себе характер «самосогласованных» состояний. В связи с этим отметим два обстоятельства. В условиях, обратных (1.1), состояние, в котором все доноры заполнены электронами, не может быть самосогласованным — даже в отсутствие компенсации и при нулевой температуре. Действительно, при $N_t a_B^3 \gtrsim 1$ заведомо возникает примесная зона, т. е. наступает хотя бы частичная делокализация электронов, расположенных на примесных атомах, с неизбежным появлением эффекта экранирования.

Во-вторых, при $r_0 > N_t^{-1/3}$, $a_B > N_t^{-1/3}$ скопление ионов примеси, расстояние между которыми много меньше a_B, можно рассматривать как один «эффективный ион» с зарядом Ze. Энергия

ионизации электрона, захваченного таким ионом, есть Z^2E_B, а характерный «радиус орбиты» равен a_B/Z. Он может оказаться меньше r_0 даже в условиях сильного легирования. Далее, вероятность образования кластера тем меньше, чем больше в нем ионов. Следовательно, при достаточно большом значении Z среднее расстояние между кластерами будет значительно больше $N_t^{-1/3}$; оно может заметно превысить и «радиус орбиты» a_B/Z. Таким образом, «эффективные ионы» с достаточно большими значениями Z могли бы создавать дискретные локальные уровни, даже когда обычные донорные уровни исчезают. Чтобы эта возможность осуществилась, энергия ионизации должна быть меньше ширины запрещенной зоны *):

$$Z^2E_B < E_g. \qquad (2.2)$$

У таких полупроводников, как германий, кремний, арсенид галлия и др., отношение E_g/E_B составляет $70 \div 100$. Поэтому допускаемые неравенством (2.2) значения Z могут удовлетворять поставленным выше условиям; при этом взаимодействие кластеров друг с другом и с отдельными ионами примеси приводит лишь к классическому уширению уровней.

При $r_0 \gg N_t^{-1/3}$ дискретные уровни могут возникать и в достаточно глубоких потенциальных ямах, созданных совокупным действием многих не слишком близких друг к другу ионов примеси. Отличие их от уровней, создаваемых «эффективными ионами», не принципиально: оно состоит лишь в зависимости энергии ионизации от параметров вещества.

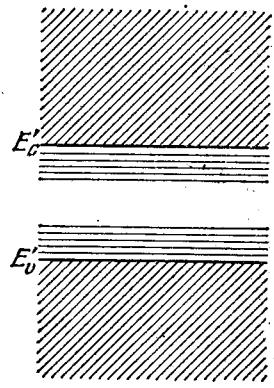


Рис. 19.3. Области непрерывного и дискретного спектров в сильно легированном полупроводнике. Горизонтальная штриховка указывает область дискретного спектра, косая — непрерывного.

Вывод о существовании дискретных локальных уровней в сильно легированном полупроводнике имеет принципиальное значение: оказывается, что при некоторых значениях энергий E_c' и E_v' в примесных областях происходит перестройка энергетического спектра электронов и дырок (рис. 19.3) **). При $E > E_c'$ ($E < E_v'$) состояния носителей заряда принадлежат непрерывному спектру, при

*) Принимая для E_B выражение (IV.7.7), т. е. пользуясь методом эффективной массы, мы должны поставить в (2.2) знак сильного неравенства. Однако дискретные уровни могут возникать и в условиях, когда метод эффективной массы уже неприменим.

**) Этот вывод представляет собой частный случай общего утверждения о перестройке спектра элементарных возбуждений в неупорядоченных системах [1].

$E_v' < E < E_c'$ — дискретному. Соответственно меняется и вид волновых функций: носители заряда, находящиеся в состояниях первого типа, можно с отличной от нуля вероятностью обнаружить в любой точке образца; в состояниях второго типа они локализованы вблизи соответствующих потенциальных ям. При изучении явлений переноса именно величины E_c' и E_v' надо рассматривать как границы запрещенной зоны.

§ 3. Иерархия энергий

В различных явлениях относительная роль носителей заряда различных энергий оказывается различной. Так, при вычислении уровня Ферми важен вклад электронов от всех возможных состояний; в явлениях переноса в вырожденном материале главную роль играют носители заряда с энергией, близкой к фермиевской; при исследовании края оптического поглощения важны электроны и дырки вблизи границ зон и в примесных областях спектра. Особенности сильно легированных полупроводников, рассмотренные в предыдущем параграфе, по-разному сказываются на поведении носителей заряда различных энергий и, следовательно, в разных эффектах проявляются по-разному.

Рассмотрим сначала некомпенсированный и, следовательно, вырожденный материал (для определенности n-типа).

При вычислении уровня Ферми следует принять во внимание три характерные величины: отсчитанную от дна зоны среднюю энергию электронов, которой бы они обладали в идеальном кристалле, E_k (мы будем для краткости называть ее «кинетической»), среднюю потенциальную энергию взаимодействия их друг с другом E_{ee} и с примесью E_{et}. По порядку величины мы имеем (при параболическом законе дисперсии) *)

$$E_k \simeq \frac{\hbar^2 n^{2/3}}{m}, \qquad E_{ee} \simeq \frac{e^2 n^{1/3}}{\varepsilon}, \qquad E_{et} \simeq \frac{e^2}{\varepsilon} N_t^{1/3}. \tag{3.1}$$

Следовательно,

$$\frac{E_k}{E_{ee}} \sim a_B n^{1/3}, \qquad \frac{E_k}{E_{et}} \sim a_B n^{1/3} \left(\frac{n}{N_t}\right)^{1/3}. \tag{3.2}$$

Видим, что относительная роль «кинетической» энергии повышается при увеличении концентрации носителей заряда. Это есть следствие параболичности закона дисперсии и того факта, что электроны подчиняются статистике Ферми (§ V.6). Величины $a_B n^{1/3}$ и $a_B N_t^{1/3}$ называются параметрами сильного легирования. Теорию сильно легированных некомпенсированных полупроводников обычно строят в предположении, что

$$a_B n^{1/3} \gg 1, \qquad a_B N_t^{1/3} \gg 1. \tag{3.3}$$

*) Мы опускаем несущественные для дальнейшего численные коэффициенты.

Условия (3.3) довольно хорошо выполняются в материалах с малыми эффективными массами — сплавах висмута с сурьмой, соединениях PbS, PbTe и т. д. В полупроводниках типа германия левые части неравенств (3.3) даже при наибольших достижимых концентрациях примеси лишь незначительно превышают единицу. Тем не менее расчеты, выполненные в предположениях (3.3), все же дают известную ориентацию в свойствах сильно легированных материалов.

В условиях (3.3) «кинетическая» энергия свободных носителей заряда значительно превышает E_{ee} и E_{et}. Далее, согласно (1.3а, б) при изменении знаков неравенств (1.1) радиус экранирования оказывается значительно меньше a_B, но больше $N_t^{-1/3}$. По этой причине эффект экранирования должен значительно уменьшить концентрацию примесных состояний по сравнению с N_t. Иначе говоря, при любой температуре большинство электронов должно перейти в область непрерывного спектра. И действительно, расчет, выполненный методами многоэлектронной теории твердого тела [2], показывает, что в рассматриваемых условиях

$$\frac{N_t - n}{N_t} \sim (N_t a_B^3)^{-3/8}. \tag{3.4}$$

Таким образом, в сильно легированном некомпенсированном полупроводнике положение уровня Ферми близко к тому, что имело бы место в идеальном вырожденном газе носителей заряда с концентрацией $n \simeq N_t$: с точностью до малых величин оно дается (в электронном образце) формулой (V.6.2). Эти соображения, в частности, позволяют оценивать величину сдвига Бурштейна—Мосса (см. § XVIII.7) в рассматриваемом материале так, как если бы примесь вообще не влияла на поведение свободных носителей заряда.

Обратимся теперь к явлениям переноса в электрическом, магнитном и температурном полях. Согласно (XIV.2.22б) в рассматриваемом случае теория, основанная на кинетическом уравнении Больцмана, оправдана, коль скоро выполняется неравенство (2.1), в котором в качестве $\tau(E)$ надо использовать время релаксации, обусловленное рассеянием носителей заряда атомами заряженной примеси. Согласно (XIV.5.19) с точностью до несущественных численных множителей порядка единицы

$$\frac{1}{\tau(E)} \simeq \frac{e^4 N_t}{\varepsilon^2 E^{3/2}\sqrt{m}} \ln \frac{mEr_0^2}{\hbar^2}. \tag{3.5}$$

При $E = \zeta$ мы получаем, с учетом (V.6.2),

$$\frac{\hbar}{\tau(\zeta)} \simeq \frac{N_t}{n} \frac{me^4}{\varepsilon^2 \hbar^2} \ln (n^{2/3} r_0^2).$$

Как уже отмечалось, в рассматриваемых материалах $n \simeq N_t$. Поскольку $\ln (n^{2/3} r_0^2)$, как правило, не превышает нескольких

единиц, неравенство (2.1) принимает вид

$$n^{2/3}a_B^2 \gg 1. \tag{3.6}$$

В условиях (3.3) это неравенство действительно выполняется.

Видим, что в рассматриваемых материалах роль примеси в явлениях переноса состоит лишь в рассеянии носителей заряда. Представление о законе дисперсии в области $E \sim F - E_c = \zeta$ сохраняет смысл. Вместе с тем относительное изменение подвижности в результате примесного рассеяния может оказаться очень значительным. Так, в n-Ge, содержащем 10^{18} см$^{-3}$ мелких доноров, подвижность электронов при комнатной температуре близка к 250 см^2/В·с (а не 3800 см^2/В·с, что было бы в чистом материале).

Итак, на большинство термодинамических, электрических и оптических характеристик сильно легированных некомпенсированных полупроводников существенное влияние оказывает лишь первая особенность этих материалов — фермиевское вырождение газа свободных носителей заряда. Причина этого очевидна: примесная область спектра не играет роли в рассматриваемых явлениях. Исключение составляют спектральная зависимость коэффициента рекомбинационного излучения (§ 6) и вольтамперная характеристика туннельного диода (§ VIII.3 и § 5).

Положение меняется в сильно компенсированных материалах, содержащих как мелкие доноры, так и мелкие акцепторы, причем

$$\frac{|N_d - N_a|}{N_d} \ll 1. \tag{3.7}$$

Концентрация свободных носителей заряда здесь невелика и уровень Ферми расположен в запрещенной зоне. Следовательно, газ носителей заряда не вырожден вплоть до довольно низких температур, а относительная роль различных видов энергии оказывается совсем не такой, как в отсутствие компенсации. Важную роль здесь играет взаимодействие носителей заряда с примесными атомами и все особенности сильно легированных полупроводников (кроме первой) проявляются в полной мере.

§ 4. Плотность состояний

Особенности сильно легированных полупроводников, связанные с ролью случайного поля примеси и с многоэлектронными эффектами, делают сравнительно малоэффективной привычную нам схему исследования электронных свойств этих материалов. Вместо закона дисперсии здесь целесообразно использовать другие понятия — с одной стороны, более общие, чем представление об этом законе, а с другой — более простые, чем волновая функция всей многоэлектронной системы. Важнейшее из таких понятий есть плотность состояний. Оно возникает уже в обычной зонной

теории: согласно (V.4.1) и (V.4.6) именно через плотности состояний $N_c(E)$ и $N_v(E)$ выражаются концентрации свободных электронов и дырок.

Зная плотности состояний как функции энергии, мы можем найти концентрации носителей заряда как функции температуры, уровня Ферми и, может быть, магнитной индукции. Тем самым определяются все равновесные характеристики электронного газа.

Концентрацию электронов и дырок на локальных уровнях также можно представить в виде (V.4.1). Действительно, вспоминая формулы (V.9.3), видим, что плотность состояний, соответствующую дискретному донорному уровню E_d, можно записать в виде

$$N_d(E) = N_d \delta(E - E_d^*), \tag{4.1}$$

где

$$E_d^* = E_d + kT \ln(g_0/g_1). \tag{4.2}$$

Наконец, в запрещенной зоне (при $E \neq E_d$, $E \neq E_a$) плотность состояний равна нулю.

Таким образом, плотность состояний $N_l(E)$ ($l = c, v$) можно использовать как характеристику энергетического спектра: она отлична от нуля и непрерывна в разрешенных энергетических зонах, равна нулю в запрещенной зоне и имеет дельтообразные особенности при энергиях, отвечающих дискретным уровням. Вид функций $N_c(E)$ и $N_v(E)$ в известной мере характеризует структуру зон: согласно (V.7.8б) эти функции выражаются через законы дисперсии носителей заряда.

Вклад носителей заряда с данной энергией в явления переноса также связан с видом плотности состояний. Действительно, электроны и дырки, локализованные на дискретных уровнях, могут принимать участие в явлениях переноса только путем перескоков и при $T = 0$ их вклад, например, в статическую электропроводность равен нулю. С другой стороны, электроны, находящиеся в зоне проводимости, всегда дают ненулевой вклад в статическую электропроводность. В условиях вырождения так обстоит дело при любой сколь угодно низкой температуре.

Видим, что справедливо следующее утверждение:

1) Статическая электропроводность при $T = 0$ остается отличной от нуля тогда и только тогда, когда уровень Ферми попадает в область, где плотность состояний отлична от нуля и непрерывна.

Видна также и связь плотности состояний с температурной зависимостью статической электропроводности. Именно, если пренебречь малой прыжковой проводимостью, то справедливо следующее утверждение:

2) Если уровень Ферми попадает в область $E_v' < E < E_c'$, где плотность состояний равна нулю или имеет δ-образные особенности, то статическая электропроводность, связанная с электронами

проводимости, пропорциональна выражению

$$\exp\frac{F-E_c'}{kT}. \tag{4.3}$$

Действительно, выражение (4.3) (с точностью до коэффициента пропорциональности) определяет просто число свободных электронов.

Перенос этого утверждения на случай дырок очевиден.

Наконец, плотность состояний коррелирует и с электропроводностью в поле конечной частоты, т. е., согласно гл. XVIII, с коэффициентом поглощения или излучения света. Действительно, эти процессы могут происходить лишь при наличии электронов в начальных состояниях и свободных дозволенных уровней — в конечных. Иначе говоря, справедливо следующее утверждение:

3) Коэффициент поглощения или рекомбинационного излучения электромагнитных волн на частоте ω может быть отличен от нуля, лишь если

$$\hbar\omega = E_2 - E_1, \tag{4.4}$$

где E_2 и E_1 — два значения энергии, при которых плотность состояний отлична от нуля. (Условие непрерывности плотности состояний при этом не обязательно.)

Это утверждение относится и ко взаимодействию электронов со звуком. В случае поглощения или испускания сразу нескольких фотонов (или фононов) левую часть (4.4) следует заменить суммой их энергий.

Утверждения 1)—3) называют теоремами о корреляции. В рамках зонной теории твердого тела они вытекают непосредственно из формулы (XIII.7.12), из закона сохранения энергии при квантовых переходах и из того, что вероятность перехода пропорциональна числу занятых мест в начальном состоянии и числу свободных мест — в конечном. Существенно, однако, что эти теоремы имеют и более общее значение. Действительно, плотность состояний представляет собой не что иное, как число дозволенных энергетических уровней, отнесенное к единице энергии и к единице объема. Это понятие не связано непременно с особенностями поведения одного электрона в периодическом поле. В равной мере им можно характеризовать и энергетический спектр электрона, движущегося в произвольном (постоянном во времени) внешнем поле, в том числе и случайном. В последнем случае плотность состояний удобно вводить как величину, уже усредненную по всем возможным конфигурациям примеси.

Формула (V.4.1) остается в силе и в произвольном внешнем поле, но от определений (V.7.8, а, б) надо, вообще говоря, отка-

заться. Более того, формулу (V.4.1) можно сохранить и при учете взаимодействия электронов друг с другом. В последнем случае, однако, аргумент E уже не имеет смысла одноэлектронной энергии. Его следует понимать как изменение энергии всей системы электронов при изменении числа частиц в ней на единицу. Доказательство этого утверждения дается в общей теории многих тел [2]; там же выводится и общая формула, связывающая функцию $N_l(E)$ с решением соответствующей задачи многих тел.

Таким образом, плотность состояний представляет собой точное понятие, которым можно характеризовать энергетический спектр любой системы частиц. Столь же общими оказываются и теоремы о корреляции. Однако, в общем случае их доказательство требует несколько более тонких рассуждений, выходящих за рамки этой книги.

Пользуясь теоремами о корреляции, следует иметь в виду два обстоятельства.

Во-первых, равенство (4.4) есть необходимое, но, вообще говоря, еще не достаточное условие возможности оптических переходов: могут существовать правила отбора, связанные с другими законами сохранения. Так, в идеальном кристалле закон сохранения квазиимпульса допускает только вертикальные переходы (§ XVIII.5).

Во-вторых, в теоремах говорится о наличии корреляции между плотностью состояний, с одной стороны, и электропроводностью, коэффициентом поглощения и т. д., с другой, но не о возможности непосредственно выразить эти величины через $N_l(E)$. Так, согласно формуле (XIII.7.12) даже в случае почти идеального кристалла электропроводность выражается не только через плотность состояний, но и через вероятность перехода, фигурирующую в формуле для времени релаксации.

В идеальном кристалле плотность состояний не зависит от координат. В образце с примесью плотность состояний, не усредненная по всем конфигурациям примеси, от координат, разумеется, зависит: пики, отвечающие примесным уровням, имеются только там, где расположены атомы примеси. Однако при усреднении по конфигурациям эта зависимость может исчезнуть. Так обстоит дело в «хорошо легированном» материале, когда атомы примеси в среднем равномерно распределены по всему объему образца. Материалы, в которых плотность состояний, усредненная по конфигурациям примеси, не зависит от координат, мы ранее называли технологически однородными. Иногда их называют также макроскопически однородными *). В дальнейшем мы будем рассматривать либо такие материалы, либо контакт между ними.

*) Пример макроскопически неоднородного материала являет собой полупроводник с n—p-переходом.

§ 5. «Хвост» плотности состояний

В полупроводнике с идеальной решеткой плотность состояний имеет вид, схематически изображенный на рис. 19.4, *а*.

При слабом легировании в запрещенной зоне появляются острые (дельтообразные) пики, отвечающие дискретным примесным уровням (рис. 19.4, *б*). В условиях промежуточного легирования эти пики размазываются (рис. 19.4, *в*), причем площадь каждого пика (пиков) остается равной концентрации соответствующей примеси.

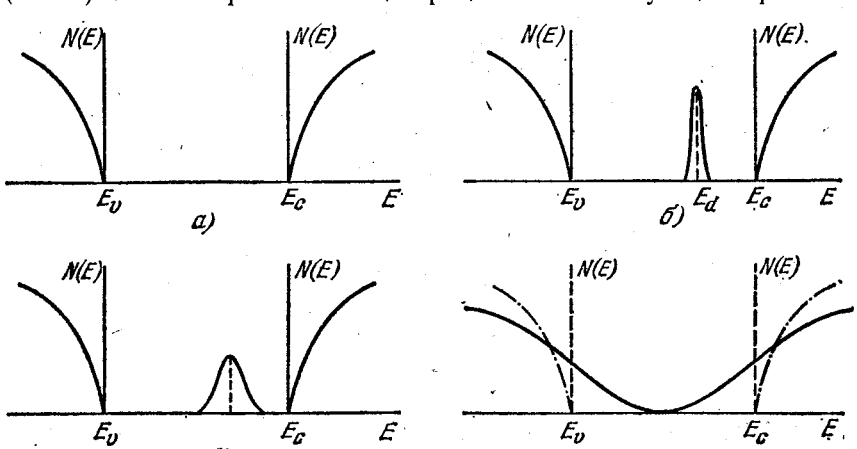


Рис. 19.4. Плотность состояний (схематически). *а*) Полупроводник с идеальной решеткой. *б*) Слабо легированный полупроводник с донорами одного типа. *в*) Промежуточно легированный полупроводник *n*-типа. *г*) Сильно легированный полупроводник. Пунктиром указаны границы зон в чистом материале, штрихпунктиром — ход плотности состояний в нем.

При сильном легировании возникает картина, схематически представленная на рис. 19.4, *г*. Точки $E = E_c$, $E = E_v$, отвечавшие границам запрещенной зоны в чистом образце, равно как и $E = E_d$, здесь ничем не выделены — плотность состояний в них оказывается отличной от нуля и ограниченной. Примесным областям спектра отвечают участки с ненулевой плотностью состояний, постепенно убывающей в глубь запрещенной зоны. Их называют «хвостами» плотности состояний. Своим происхождением они обязаны случайному полю примеси. В тех местах, где флуктуации в распределении примеси привели к образованию потенциальных ям, потенциальная энергия носителя заряда оказывается пониженной по сравнению с тем, что было бы в идеальном кристалле (рис. 19.5). Это влечет за собой и уменьшение полной энергии. Положение оказывается особенно простым, когда радиус экранирования значительно превышает среднее расстояние между атомами примеси, т. е. когда знак

первого из неравенств (1.1) сменяется на обратный. В этих условиях случайное поле в среднем достаточно плавно изменяется в пространстве и поведение электронов и дырок можно рассматривать квазиклассическим путем — с помощью представления об искривленных зонах *). Сплошные линии на рис. 19.5 показывают положения границ таких случайно искривленных зон. При усреднении по конфигурациям примеси локальный характер флуктуаций случайного поля «смазывается», и остается не зависящая от координат

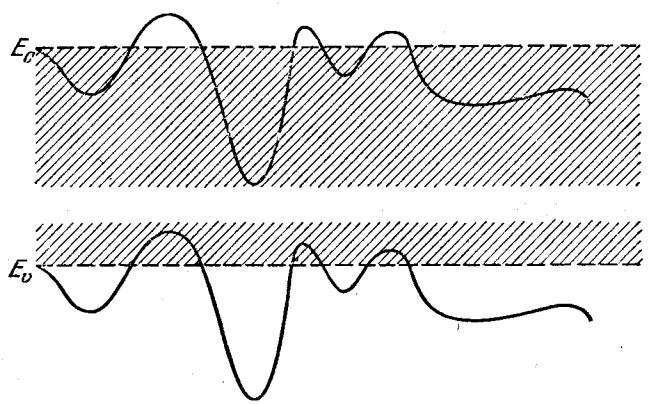


Рис. 19.5. Изменение потенциальной энергии носителей заряда в случайном поле (схематически). По оси абсцисс отложена пространственная координата. Заштрихованы хвосты плотности состояний, получающиеся при усреднении по конфигурациям примеси.

плотность состояний, отличная от нуля во всех точках, до которых доходят границы искривленных зон. Плотность состояний на хвосте при этом пропорциональна вероятности соответствующих флуктуаций потенциальной энергии; по этой причине она довольно быстро уменьшается по мере углубления в запрещенную зону. Это и позволяет говорить по отдельности о хвостах плотности состояний $N_c(E)$ и $N_v(E)$ вблизи дна зоны проводимости и вблизи потолка валентной зоны. По этой же причине полное число состояний

*) Заметим, что в применении к вырожденному полупроводнику слово «квазиклассически» в данном случае имеет ограниченный смысл. Действительно, в наиболее интересном случае мелких доноров или акцепторов $a_B = \varepsilon\hbar^2/me^2$, и большой — по условию — параметр $r_0 N_t^{1/3}$ — порядка $(a_B n^{1/3})^{1/2}$. Это выражение содержит постоянную Планка в числителе. Дело в том, что слово «квазиклассически» относится здесь к поведению носителей заряда в случайном поле примеси при заданном радиусе экранирования. В условиях вырождения последний сам зависит от $\hbar$, что и приводит к появлению этой постоянной в явно раскрытом выражении для $r_0 N_t^{1/3}$.

в области хвоста, отнесенное к единице объема, оказывается небольшим по сравнению с концентрацией примеси.

Роль хвостов плотности состояний в тех или иных явлениях различна в зависимости от степени компенсации материала. В некомпенсированных и, следовательно, вырожденных образцах наличие хвостов заметно в двух эффектах.

Первый из них — избыточный ток в туннельных диодах (§ VIII.3). С учетом хвостов зонная диаграмма туннельного n—p-перехода принимает вид, схематически представленный на рис. 19.6. Видно, что один из возможных механизмов возникновения избыточного

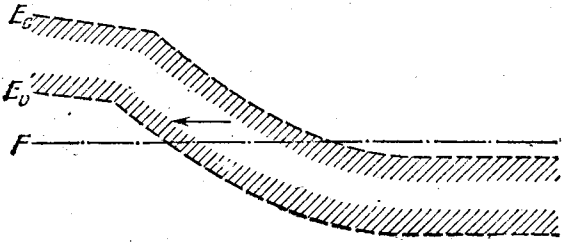


Рис. 19.6. Зонная диаграмма туннельного n — p-перехода с учетом хвоста плотности состояний (схематически). Заштрихованы области энергии, в которых плотность состояний на хвосте достаточно велика. Стрелкой указан один из электронных переходов, дающих вклад в плотность избыточного тока.

тока обусловлен самой природой сильно легированного полупроводника: этот механизм связан с электронными переходами между уровнями, принадлежащими хвостам плотности состояний в n- и p-областях. Исследование избыточного тока в туннельных диодах из германия и арсенида галлия позволило впервые получить прямое доказательство существования хвостов плотности состояний в этих материалах [3].

Второй из указанных эффектов состоит в рекомбинационном излучении фотонов с энергией, меньшей ширины запрещенной зоны. Он рассматривается в следующем параграфе.

Особенно заметна роль хвостов плотности состояний в сильно компенсированных полупроводниках. Один из эффективных способов создания их состоит в облучении образцов нейтронами или быстрыми (с энергией около 1 МэВ) электронами. Отношение $\frac{|N_d - N_a|}{N_d}$ при этом может достигать 10^{-5} (при $N_d \simeq 10^{19}$ см$^{-3}$) В таких материалах уровень Ферми располагается недалеко от середины запрещенной зоны, и, следовательно, большинство носителей заряда находится именно в примесных областях энергетического спектра. При этом важную роль могут играть плавные и достаточно глубокие флуктуации потенциальной энергии носителей

заряда. Действительно, при достаточно сильном искривлении зон границы их могут пересекать уровень Ферми (рис. 19.7). При этом в материале возникают области n- и p-типа, случайно распределенные по объему образца. Вероятность возникновения их может

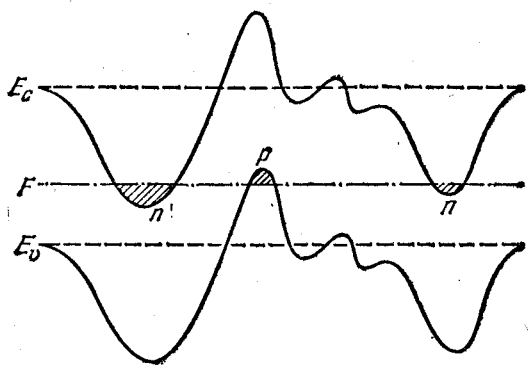


Рис. 19.7. Сильное флуктуационное искривление зон. Буквами n и p отмечены заштрихованные области, в которых при низких температурах образуются «капли» электронов и дырок.

оказаться достаточно большой, так что при низких температурах практически все носители заряда собираются в этих областях. Последние получили название электронных и дырочных «капель».

§ 6. Междузонные оптические переходы в сильно легированных полупроводниках

Особенности сильно легированных полупроводников приводят к некоторым изменениям их оптических характеристик по сравнению с тем, что наблюдается в сравнительно чистых материалах.

а). Переходы в глубину разрешенной зоны. Из-за вырождения газа носителей заряда в некомпенсированных материалах красная граница поглощения в них сдвигается на величину $\Delta\omega_n = \frac{F - E_c}{\hbar}$ или $\Delta\omega_p = \frac{E_v - F}{\hbar}$, соответственно, в образцах n- и p-типа. Это есть сдвиг Бурштейна — Мосса, обсуждавшийся в § XVIII. 7. В соответствии со сказанным в § 3, случайное поле примеси сравнительно слабо влияет на вероятность рассматриваемых переходов. Коэффициенты поглощения и рекомбинационного излучения здесь в основном описываются формулами гл. XVIII. Использование последних для определения ширины запрещенной зоны привело к выводу, что в сильно легированных полупроводниках она несколько меньше, чем в чистом материале. Этот эффект может быть обусловлен

как небольшим изменением постоянной решетки при сильном легировании, так и взаимодействием между электронами.

б). Переходы с участием хвостов плотности состояний. Третья теорема о корреляции указывает на возможность оптических переходов с участием хвостов плотности состояний. В силу быстрого убывания плотности состояний по мере удаления от границ запрещенной зоны эти переходы следует рассматривать как междузонные. Тем не менее энергия соответствующих фотонов $\hbar\omega$ может оказаться меньше ширины запрещенной зоны $E_g = \hbar\omega_m$. Об излучении или поглощении таких фотонов говорят как о хвосте коэффициента рекомбинационного излучения или поглощения (или «оптическом

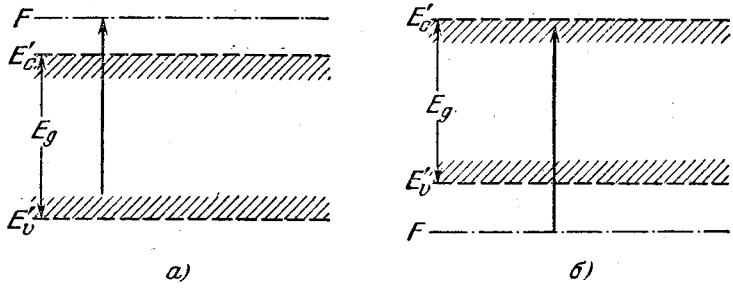


Рис. 19.8. Происхождение хвоста коэффициента поглощения в некомпенсированном полупроводнике. Заштрихованы области энергии, в которых плотность состояний на хвосте достаточно велика. Стрелкой указан один из возможных переходов, ответственных за поглощение фотонов с энергией, меньшей E_g. *а*) *n*-тип; *б*) *p*-тип.

хвосте»). «Величина» этого хвоста, т. е. значение коэффициента рекомбинационного излучения или поглощения света при $\hbar\omega < E_g$, зависит от степени компенсации. При сильном вырождении в отсутствие инжекции переходы в область энергий ниже уровня Ферми практически запрещены принципом Паули. По этой причине в поглощении фотонов с частотой $\omega < \omega_m$ здесь могут участвовать в основном самые глубокие уровни на хвостах плотности состояний (рис. 19.8); при этом в образцах *n*-типа существен хвост плотности состояний вблизи валентной зоны, а в образцах *p*-типа — вблизи зоны проводимости. В результате значения коэффициента поглощения при $\omega < \omega_m$ в некомпенсированных сильно легированных полупроводниках оказываются очень малыми. С другой стороны, хвост коэффициента рекомбинационного излучения в этих материалах может быть вполне заметен. Действительно, при инжекции носителей заряда квазиуровень Ферми может попасть в область соответствующего хвоста плотности состояний.

В компенсированных полупроводниках принцип Паули практически не ограничивает возможность электронных переходов из хво-

ста у валентной зоны в хвост близ зоны проводимости (рис. 19.9). Действительно, уровень Ферми здесь расположен между этими хвостами, в результате чего дозволенные состояния вблизи E_v' почти заполнены, а вблизи E_c' — почти все свободны. Соответственно коэффициент поглощения в области $\omega < \omega_m$ оказывается довольно заметным.

Спектральная зависимость коэффициентов поглощения и рекомбинационного излучения при $\omega < \omega_m$ удовлетворительно описывается формулой

$$\gamma = C \exp\left(-\frac{E_g - \hbar\omega}{E_0}\right). \quad (6.1)$$

Здесь C — сравнительно медленно меняющаяся функция частоты, E_0 — величина размерности энергии.

Соотношение (6.1) называется правилом Урбаха. Оно было найдено эмпирическим путем при исследовании поглощения света в ионных кристаллах. В этих материалах характерная энергия E_0 оказалась пропорциональной температуре:

$$E_0 = cT, \quad (6.2)$$

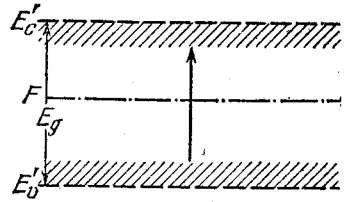


Рис. 19.9. Происхождение хвоста коэффициента поглощения в компенсированном полупроводнике. Заштрихованы области энергии, в которых плотность состояний на хвосте достаточно велика. Стрелкой указан один из возможных переходов, ответственных за поглощение фотонов с энергией, меньшей E_g.

где c — некоторая постоянная. Это указывает на взаимодействие электронов с фононами (а не с примесью) как на возможную причину эффекта. Чтобы совершить междузонный переход, электрон должен получить энергию, не меньшую E_g; часть ее $\hbar\omega$ электрон получает непосредственно от фотона, а дефицит $E_g - \hbar\omega$ покрывается фононами. При этом правая часть равенства (6.1) приближенно воспроизводит просто вероятность соответствующего многофононного перехода. Объяснение правила Урбаха на основе представления об оптических переходах, сопровождаемых поглощением многих фононов, было дано А. С. Давыдовым. Соотношение (6.2) выполняется и в сильно легированных полупроводниках, но лишь при достаточно высокой температуре (в арсениде галлия — при $T \gtrsim 100$ К *). При более низких температурах величина E_0 перестает зависеть от T. Вместо этого она становится зависящей от концентрации примеси, возрастая вместе с ней. Именно тогда наблюдаются переходы с участием хвостов плотности состояний. Теорию этих переходов удается построить в условиях сильного легирования (3.3). При этом приближенно воспроизводится как спектральная зависимость (6.1), так и

*) Цифра взята из работы [4].

зависимость E_0 от концентрации примеси в образце. По существу, мы здесь имеем дело с аналогом эффекта Келдыша — Франца, обусловленного, однако, не внешним электрическим полем, а случайным полем примесей. Последнее, однако, неоднородно в пространстве. Даже в условиях (3.3) оно, будучи плавным в среднем, быстро изменяется вблизи каждого из примесных ионов. По этой причине здесь нельзя воспользоваться формулами § XVIII.10, заменяя в них напряженность поля $\mathcal{E}$ некоторым эффективным ее значением. Подробное изложение расчетов, выполненных в условиях сильного легирования, можно найти в обзоре [5].

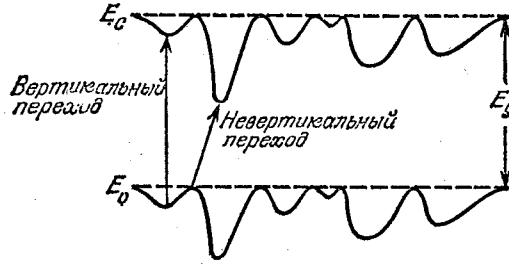


Рис. 19.10. Вертикальные и невертикальные переходы в пространстве «энергия — координата».

Следует заметить, что форма оптического хвоста, т. е. частотная зависимость коэффициента поглощения или рекомбинационного излучения света, вообще говоря, не воспроизводит плотность состояний как функцию энергии. Суть дела легко понять из рис. 19.10, на котором схематически представлена картина искривленных зон; положения их границ зависят от координаты, отложенной вдоль горизонтальной оси. Обе зоны — валентная и проводимости — одинаково искривляются электрическим полем. По этой причине в каждой точке пространства расстояние между их границами E_c и E_v остается неизменным и равным E_g. Электронные переходы с затратой меньшей энергии могут происходить только между точками, разделенными в пространстве: это — «невертикальные» переходы на диаграмме «энергия — координата». Они связаны с туннелированием электрона через запрещенную зону, вероятность чего сильно зависит от энергии поглощаемого или испускаемого фотона.

Возможны и условия, когда вид функции $\gamma(\omega)$ приближенно воспроизводит функцию $N(E)$ вблизи той или иной зоны. Так, например, обстоит дело, если одно из двух состояний, между которыми происходит переход, мало изменяется случайным полем примеси. Однако такая возможность не носит общего характера.

§ 7. Некристаллические полупроводники

Сильно легированные полупроводники составляют простейшую группу неупорядоченных веществ. Под этим термином понимается любая макроскопическая конденсированная система частиц, в расположении атомов которой отсутствует дальний порядок. Последнее условие означает, что пространственное распределение атомов неупорядоченного вещества не периодично. В сильно легированном полупроводнике это относится только к атомам примеси. К числу более сложных объектов этого типа относятся жидкие, аморфные и стеклообразные полупроводники (см. гл. II).

Совокупность экспериментальных данных, относящихся к электрическим и оптическим характеристикам названных материалов, показывает, что введенные в гл. I феноменологические представления о разрешенных и запрещенных энергетических зонах в известной мере сохраняют смысл и здесь *): имеет место резкое изменение коэффициента поглощения света при некоторой частоте ω_m, температурная зависимость статической электропроводности в широком интервале температур описывается обычной экспоненциальной формулой и т. д. Однако для квантовомеханического обоснования представлений о зонах мы уже не можем обратиться к гл. III, IV: в отсутствие дальнего порядка в расположении атомов потенциальную энергию электрона нельзя считать периодической функцией координат, и теорема Блоха не имеет места. Вместе с тем, и в этих материалах сохраняют точный смысл представления о непрерывном и дискретном энергетическом спектре носителей заряда. Области непрерывного спектра электронов (дырок) естественно назвать зоной проводимости (дырочной). Это определение как раз соответствует экспериментальному содержанию представления о разрешенных зонах.

Ряд веществ можно получить как в кристаллическом, так и в аморфном состояниях. Рентгенографические и электронографические исследования показывают, что во многих аморфных веществах расположение атомов, ближайших к данному, почти то же, что и в соответствующем кристалле. Об этом говорят как о сохранении ближнего порядка при переходе в аморфное состояние.

Опыт показывает, что значения ширины запрещенной зоны в том и другом состояниях довольно близки друг к другу. Отсюда следует, что основную роль в определении энергетического спектра этих материалов играет ближний порядок **).

Вместе с тем сходство характеристик кристаллических и аморфных веществ оказывается далеко не полным. Действительно, отсут-

*) Подробную сводку и анализ накопленного к настоящему времени материала можно найти в книгах [6, 7] и в обзоре [8].

**) Эта идея была впервые высказана А. Ф. Иоффе и А. Р. Регелем. Полупроводниковые свойства халькогенидных стекол были обнаружены Б. Т. Коломийцем и Н. А. Горюновой.

ствие дальнего порядка означает, что представление о зонах Бриллюэна в применении к рассматриваемым материалам не имеет смысла, а вторая и четвертая особенности сильно легированных полупроводников (§ 2) характерны для всех неупорядоченных систем. Отсюда следует, что в спектрах поглощения этих веществ не должно быть особенностей Ван-Хова: как мы видели в гл. XVIII, возникновение их связано с существованием закона дисперсии носителей заряда. И действительно, опыт показывает, что при аморфизации эти особенности размываются. Более подробный анализ экспериментальных данных приводит к выводу о существовании двух типов неупорядоченных полупроводников: А и Б.

В полупроводниках типа А, как и в чистых кристаллах (гл. XVIII), наблюдается порог коэффициента поглощения света. При этом пороговая энергия фотона $\hbar\omega_m$ близка к ширине запрещенной зоны, определенной по температурной зависимости статической электропроводности σ. Последняя в широком интервале температур описывается обычным экспоненциальным законом (§ I.5) с энергией активации, близкой к $\hbar\omega_m/2$.

У полупроводников типа Б имеется хвост коэффициента поглощения света. Как и в сильно легированных кристаллических материалах, он удовлетворительно описывается выражением (6.1). Статическая электропроводность σ таких полупроводников при достаточно высоких температурах зависит от температуры по такому же экспоненциальному закону, что и в полупроводниках типа А. Однако при более низких температурах (в аморфном германии — ниже примерно 200 К) температурная зависимость σ лучше описывается соотношением

$$\sigma = A \exp\left[-\left(\frac{T_0}{T}\right)^{1/4}\right], \qquad (7.1)$$

где A — медленно меняющаяся функция T, а T_0 — постоянная (порядка 10^8 К). Соотношение (7.1) называют формулой Мотта.

Поглощение света с частотой меньше пороговой естественно связать с наличием хвостов плотности состояний в запрещенных зонах материалов типа Б. Вывод о существовании хвостов, содержащих и дискретные уровни, подтверждается также результатами исследования кинетики фотопроводимости, токов, ограниченных пространственным зарядом, и др. Полная концентрация уровней на хвостах, меняясь от вещества к веществу и от образца к образцу, часто оказывается довольно большой — до $10^{19} \div 10^{20}$ см$^{-3}$. При этом в ряде материалов ни сам факт наличия рассматриваемых уровней, ни их концентрация не зависит существенным образом от состава и концентрации специально введенной примеси. Зато имеется сильная зависимость от условий приготовления образца: образцы одного и того же вещества, будучи по-разному приготовлены, могут принадлежать как к типу А, так и к типу Б. Первое обычно имеет

место для пленок, подвергнутых достаточно длительному отжигу, второе — при отжиге меньшей длительности. Отсюда следует, что особенности полупроводников типа Б могут быть связаны с наличием всякого рода случайных структурных дефектов в образце. Число их может быть столь большим, что влияние специально вводимой примеси не сказывается. Вместе с тем в хорошо приготовленных образцах, легированных достаточно большим количеством примеси, последняя может играть и свою обычную роль. Так, видимо, обстоит дело в аморфном кремнии и в некоторых халькогенидных стеклах.

Случайное поле, создаваемое структурными дефектами, приводит, в принципе, к тем же следствиям, что и поле примесных ионов в сильно легированных полупроводниках. В частности, в достаточно глубоких потенциальных ямах могут возникать дискретные уровни, отвечающие локализованным состояниям электронов. Эти уровни называют флуктуационными. Как глубина их, так и координаты центров локализации представляют собой случайные величины. Это обстоятельство позволяет понять, почему уровни могут оставаться дискретными даже при таких больших концентрациях, как указанные выше. Дело в том, что размытие дискретных уровней в зону связано с туннельными переходами электрона между центрами локализации (гл. III). При этом указанные центры должны образовывать цепочку макроскопической длины. Туннельные переходы, однако, могут происходить лишь между близко расположенными центрами, причем соответствующие им уровни энергии должны быть одинаковыми или близкими друг к другу. В силу случайного характера рассматриваемых уровней близость их одновременно и в пространстве, и по энергии маловероятна. По этой причине может оказаться отличной от нуля вероятность того, что флуктуационные уровни, возникшие в случайно расположенных потенциальных ямах случайного вида, останутся дискретными, несмотря на большое их число *). Значения энергий ионизации уровней, будучи случайными, изменяются практически непрерывно в пределах целой энергетической полосы; последняя может заполнять и всю запрещенную зону.

В этих условиях термин «запрещенная зона» становится неудобным. Вместо этого говорят о щели для подвижности. Действительно, электроны, занимающие дискретные уровни, могут участвовать в явлениях переноса только путем прыжков с одного центра на другой; при $T = 0$ соответствующая прыжковая проводимость также обращается в нуль.

При понижении температуры в рассматриваемых полупроводниках происходит смена механизма проводимости: по мере перехода электронов на локальные уровни она становится прыжковой.

*) Об образовании таких уровней говорят как об андерсоновской локализации.

При этом случайный характер расположения уровней в пространстве координат и энергии приводит к специфической температурной зависимости σ (7.1). В самом деле, пусть энергия активации, необходимая для одного отдельного прыжка, есть E_a. В рассматриваемой системе уровней эта величина — случайная. Как мы уже видели, малые значения E_a чаще встречаются при сравнительно больших расстояниях R между центрами. При этом в той области пространства, где волновые функции электронов, локализованных на первом и втором центрах, перекрываются друг с другом, мы можем воспользоваться их асимптотическими выражениями. Как известно из квантовой механики, последние (с точностью, может быть, до сравнительно медленно изменяющегося предэкспоненциального множителя) имеют вид $\psi \sim \exp(-\gamma r)$. Здесь γ^{-1} — параметр, зависящий от энергии ионизации данного уровня и называемый радиусом локализации электрона, r — расстояние между данной точкой и центром локализации. Вероятность прыжка определяется произведением вероятности тепловой ионизации $\exp(-E_a/kT)$ (в случае $E_a > kT$), на интеграл, содержащий волновые функции электрона на первом и втором центрах. Таким образом, с точностью до предэкспоненциальных множителей вероятность прыжка пропорциональна

$$\exp\{-(\gamma_1+\gamma_2)R\}\exp(-E_a/kT), \qquad (7.2)$$

где γ_1 и γ_2 — радиусы локализации электронов на первом и втором уровнях. Первый сомножитель в (7.2) убывает с ростом R, а второй — в среднем возрастает. Разумеется, могут найтись и такие пары центров локализации, которые, будучи близкими по энергии, окажутся близко расположенными и пространственно. Однако они с большой вероятностью будут расположены далеко от других центров локализации. Вместе с тем для переноса заряда через весь образец требуется образование макроскопической цепочки из центров локализации, между которыми электроны могут совершать прыжки. Таким образом, величина σ определяется в данном случае вероятностью возникновения цепочки, в которой оптимальным образом сочетаются не слишком большие расстояния между центрами локализации и не слишком большие значения энергий E_a.

Расчет, основанный на формуле (7.2) (его можно найти в книге [6]), показывает, что при не слишком высоких температурах эти соображения приводят в формуле (7.1).

В ряде аморфных, стеклообразных и жидких полупроводников наблюдается эффект переключения. Он состоит в резком и обратимом изменении электропроводности образца под влиянием электрического поля. Время переключения может оказаться небольшим: достигнуты значения до 10^{-10} с. Это позволяет использовать такие вещества в переключающих элементах для ЭВМ. Природа эффекта пока не вполне ясна.

ПРИЛОЖЕНИЯ

Приложение I. К доказательству теоремы Блоха

Пусть индексы i, j, k нумеруют различные собственные функции ψ_i, ψ_j, ψ_k, принадлежащие одному и тому же собственному значению E. Производя замену аргумента $\mathbf{r} \to \mathbf{r} + \mathbf{a}_n$, мы получим, вообще говоря,

$$\psi_i(\mathbf{r}+\mathbf{a}_n) = \sum_j c_{ij}\psi_j(\mathbf{r}), \tag{П.I.1}$$

где c_{ij} — некоторые коэффициенты.

Введем вместо ψ_i, ψ_j линейные комбинации их

$$\psi'_k = \sum_i b_{ki}\psi_i. \tag{П.I.2}$$

Коэффициенты b_{ki} (пока неизвестные) образуют некоторую матрицу. Элементы обратной матрицы обозначим через b^{-1}_{ki}. По определению

$$\sum_k b^{-1}_{jk} b_{ki} = \delta_{ij}, \tag{П.I.3}$$

где δ_{ij} — символ Кронекера (П.II.9) и, следовательно,

$$\psi_i = \sum_l b^{-1}_{il}\psi'_l. \tag{П.I.2'}$$

Согласно (П.I.1)—(П.I.2')

$$\psi'_k(\mathbf{r}+\mathbf{a}_n) = \sum_i b_{ki}\psi_i(\mathbf{r}+\mathbf{a}_n) = \sum_{i,j} b_{ki}c_{ij}\psi_j(\mathbf{r}) = \sum_{i,j,l} b_{ki}c_{ij}b^{-1}_{jl}\,\psi'_l(\mathbf{r}). \tag{П.I.4}$$

Наложим на коэффициенты b_{ki} условия

$$\sum_{i,j} b_{ki}c_{ij}b^{-1}_{jl} = C\delta_{kl}, \tag{П.I.5}$$

где C — постоянная (зависящая, может быть, от k и l). Тогда равенство (П.I.4) принимает вид

$$\psi'_k(\mathbf{r}+\mathbf{a}_n) = C_{kn}\psi'_k(\mathbf{r}), \tag{П.I.6}$$

что совпадает с (III.2.9).

Равенства (П.I.5) можно привести к более удобному виду, умножая их на $b_{lk'}$ и суммируя по l. Принимая во внимание определение (П.I.3), мы получаем при этом

$$\sum_{i,j} b_{ki}c_{ij}\delta_{jk'} = Cb_{kk'}, \tag{П.I.7}$$

т. е.

$$\sum_i b_{ki}c_{ik'} = Cb_{kk'}.$$

Это есть система линейных однородных уравнений относительно неизвестных $b_{kk'}$. Условие разрешимости ее дает нам значения C, фигурирующие в (П.I.6), а решения — коэффициенты преобразования b_{ki}.

Таким образом, взяв произвольный набор собственных функций, принадлежащих данному собственному значению энергии, мы можем составить из них новые собственные функции — линейные комбинации (П.I.2), удовлетворяющие условию (III.2.9).

Приложение II. Интегралы с функциями Блоха

Напишем уравнение (III.4.1), а также сопряженное ему уравнение с другим собственным значением энергии E' и другим квазиимпульсом $\mathbf{p}'$:

$$-\frac{\hbar^2}{2m_0}\nabla^2 u_{\mathbf{p}l} - \frac{i\hbar}{m_0}(\mathbf{p},\ \nabla u_{\mathbf{p}l}) + U u_{\mathbf{p}l} = \left(E - \frac{p^2}{2m_0}\right) u_{\mathbf{p}l}, \qquad (\text{П.II.1})$$

$$-\frac{\hbar^2}{2m_0}\nabla^2 u^*_{\mathbf{p}'l'} + \frac{i\hbar}{m_0}(\mathbf{p}',\ \nabla u^*_{\mathbf{p}'l'}) + U u^*_{\mathbf{p}'l'} = \left(E' - \frac{p'^2}{2m_0}\right) u_{\mathbf{p}'l'}. \qquad (\text{П.II.2})$$

При этом $E = E(\mathbf{p},\ l)$, $E' = E'(\mathbf{p}',\ l')$.

Умножим уравнение (П.II.1) на $u^*_{\mathbf{p}'l'}$, уравнение (П.II.2) — на $u_{\mathbf{p}l}$, затем вычтем почленно второе уравнение из первого и проинтегрируем результат по основному объему. Получим

$$-\frac{\hbar^2}{2m_0}\int \{u^*_{\mathbf{p}'l'}\nabla^2 u_{\mathbf{p}l} - u_{\mathbf{p}l}\nabla^2 u^*_{\mathbf{p}'l'}\}\, d\mathbf{r} -$$

$$-\frac{i\hbar}{m_0}\int \{u^*_{\mathbf{p}'l'}(\mathbf{p},\ \nabla u_{\mathbf{p}l}) + u_{\mathbf{p}l}(\mathbf{p}',\ \nabla u^*_{\mathbf{p}'l'})\}\, d\mathbf{r} =$$

$$= \left(E - E' - \frac{p^2 - p'^2}{2m_0}\right)\int u^*_{\mathbf{p}'l'} u_{\mathbf{p}l}\, d\mathbf{r}. \qquad (\text{П.II.3})$$

Подынтегральное выражение в первом интеграле в левой части (П.II.3) можно переписать в виде

$$\nabla(u^*_{\mathbf{p}'l'}\nabla u_{\mathbf{p}l}) - (\nabla u^*_{\mathbf{p}'l'},\ \nabla u_{\mathbf{p}l}) - \nabla(u_{\mathbf{p}l}\nabla u^*_{\mathbf{p}'l'}) + (\nabla u_{\mathbf{p}l},\ \nabla u^*_{\mathbf{p}'l'}) =$$

$$= \operatorname{div}[u^*_{\mathbf{p}'l'}\nabla u_{\mathbf{p}l} - u_{\mathbf{p}l}\nabla u^*_{\mathbf{p}'l'}].$$

Получающийся интеграл можно преобразовать по теореме Гаусса:

$$\int \operatorname{div}[u^*_{\mathbf{p}'l'}\nabla u_{\mathbf{p}l} - u_{\mathbf{p}l}\nabla u^*_{\mathbf{p}'l'}]\, d\mathbf{r} = \int ([u^*_{\mathbf{p}'l'}\nabla u_{\mathbf{p}l} - u_{\mathbf{p}l}\nabla u^*_{\mathbf{p}'l'}],\ d\mathbf{S}), \qquad (\text{П.II.4})$$

где $d\mathbf{S}$ — элемент поверхности, охватывающей основной объем. В силу условий (III.3.9) абсолютная величина подынтегрального выражения на противоположных гранях куба периодичности одна и та же. Следовательно, рассматриваемый интеграл равен нулю.

Второй интеграл в левой части (П.II.3) можно переписать в виде

$$\int \{u^*_{\mathbf{p}'l'}(\mathbf{p},\ \nabla u_{\mathbf{p}l}) + \nabla(u_{\mathbf{p}l}\mathbf{p}' u^*_{\mathbf{p}'l'}) - u^*_{\mathbf{p}'l'}(\mathbf{p}',\ \nabla u_{\mathbf{p}l})\}\, d\mathbf{r}.$$

Подобно (П.II.4) второе слагаемое здесь обращается в нуль. Таким образом, уравнение (П.II.3) принимает вид

$$-\frac{i\hbar}{m_0}\left(\mathbf{p}-\mathbf{p}',\ \int u^*_{\mathbf{p}'l'}\nabla u_{\mathbf{p}l}\, d\mathbf{r}\right) = \left(E - E' - \frac{p^2 - p'^2}{2m_0}\right)\int u^*_{\mathbf{p}'l'}u_{\mathbf{p}l}\, d\mathbf{r}. \qquad (\text{П.II.5})$$

При $\mathbf{p}'=\mathbf{p}$ мы получаем отсюда

$$(E-E')\int u^*_{\mathbf{p}l'}u_{\mathbf{p}l}\, dr = 0. \qquad (\text{П.II.6})$$

При $l'=l$ равенство (П.II.6) удовлетворяется тождественно. Значение интеграла в левой части (П.II.6) при этом определяется условием нормировки (III.2.8). С другой стороны, при $l'\neq l$, $E(\mathbf{p},\ l)\neq E(\mathbf{p},\ l')$ из него вытекает условие ортогональности:

$$\int u^*_{\mathbf{p}l'}u_{\mathbf{p}l}\, d\mathbf{r} = 0. \qquad (\text{П.II.7})$$

При $l'\neq l$, но $E(\mathbf{p},\ l)=E(\mathbf{p},\ l')$ имеет место вырождение: одному и тому же значению энергии отвечают две (или больше) волновые функции. Известно (см. ссылку [1] к гл. III), что эти функции всегда можно выбрать так, чтобы условие (П.II.7) удовлетворялось.

Комбинируя равенства (П.II.7) и (III.2.8), можем написать

$$\int u^*_{\mathbf{p}l'}u_{\mathbf{p}l}\, d\mathbf{r} = \delta_{ll'}. \qquad (\text{П.II.8})$$

Здесь $\delta_{ll'}$ есть символ Кронекера:

$$\delta_{ll'} = \begin{cases} 1, & l'=l, \\ 0, & l'\neq l. \end{cases} \qquad (\text{П.II.9})$$

Положим теперь в (П.II.5) $l'=l$ и будем рассматривать близкие значения $\mathbf{p}$ и $\mathbf{p}'$, ограничиваясь членами, линейными по $\mathbf{p}-\mathbf{p}'$. Тогда

$$E-E' \equiv E(\mathbf{p},\ l) - E(\mathbf{p}',\ l) \simeq (\mathbf{p}-\mathbf{p}',\ \nabla_{\mathbf{p}}E(\mathbf{p},\ l)),\quad p^2-p'^2 \simeq 2(\mathbf{p},\ \mathbf{p}-\mathbf{p}'),$$

а функции $u^*_{\mathbf{p}'l}$ под знаками интегралов в левой и правой частях (П.II.5) можно заменить на $u^*_{\mathbf{p}l}$. В результате равенство (П.II.5) с учетом (III.2.8) примет вид

$$-\frac{i\hbar}{m_0}\left(\mathbf{p}-\mathbf{p}',\ \int u^*_{\mathbf{p}l}\nabla u_{\mathbf{p}l}\, d\mathbf{r}\right) - (\mathbf{p}-\mathbf{p}',\ \nabla_{\mathbf{p}}E(\mathbf{p},\ l)) + \left(\mathbf{p}-\mathbf{p}',\ \frac{\mathbf{p}}{m_0}\right) = 0. \qquad (\text{П.II.5}')$$

Поскольку ориентация вектора $\mathbf{p}-\mathbf{p}'$ произвольна, это равенство может выполняться, только если обращается в нуль коэффициент при $\mathbf{p}-\mathbf{p}'$:

$$-\frac{i\hbar}{m_0}\int u^*_{\mathbf{p}l}\nabla u_{\mathbf{p}l}\, d\mathbf{r} = \nabla_{\mathbf{p}}E(\mathbf{p},\ l) - \frac{\mathbf{p}}{m_0}. \qquad (\text{П.II.10})$$

Приложение III. Таблица значений интеграла $\Phi_{1/2}$

(см. V.4.5) [1]; приведенное в таблице значение нужно умножить на 10 в степени, указанной в скобках.

ζ^*	$\Phi_{1/2}(\zeta^*)$	ζ^*	$\Phi_{1/2}(\zeta^*)$	ζ^*	$\Phi_{1/2}(\zeta^*)$
−4,0	1,8199 (−2)	−0,3	6,0022 (−1)	3,5	5,4580 (0)
−3,9	2,0099 (−2)	−0,2	6,5161 (−1)	3,6	5,6623 (0)
−3,8	2,2195 (−2)	−0,1	7,0654 (−1)	3,7	5,8699 (0)
−3,7	2,4510 (−2)			3,8	6,0806 (0)
−3,6	2,7063 (−2)	0,0	7,6515 (−1)	3,9	6,2945 (0)
		0,1	8,2756 (−1)		
−3,5	2,9880 (−2)	0,2	8,9388 (−1)	4,0	6,5115 (0)
−3,4	3,2986 (−2)	0,3	9,6422 (−1)	4,2	6,9548 (0)
−3,3	3,6412 (−2)	0,4	1,0387 (0)	4,4	7,4100 (0)
−3,2	4,0187 (−2)			4,6	7,8769 (0)
−3,1	4,4349 (−2)	0,5	1,1173 (0)	4,8	8,3550 (0)
		0,6	1,2003 (0)		
−3,0	4,8933 (−2)	0,7	1,2875 (0)	5,0	8,8442 (0)
−2,9	5,3984 (−2)	0,8	1,3791 (0)	5,2	9,3441 (0)
−2,8	5,9545 (−2)	0,9	1,4752 (0)	5,4	9,8546 (0)
−2,7	6,5665 (−2)			5,6	1,0375 (+1)
−2,6	7,2398 (−2)	1,0	1,5756 (0)	5,8	1,0906 (+1)
		1,1	1,6806 (0)		
−2,5	7,9804 (−2)	1,2	1,7900 (0)	6,0	1,1447 (+1)
−2,4	8,7944 (−2)	1,3	1,9038 (0)	6,2	1,1997 (+1)
−2,3	9,6887 (−2)	1,4	2,0221 (0)	6,4	1,2556 (+1)
−2,2	1,0671 (−1)			6,6	1,3125 (+1)
−2,1	1,1748 (−1)	1,5	2,1449 (0)	6,8	1,3703 (+1)
		1,6	2,2720 (0)		
−2,0	1,2930 (−1)	1,7	2,4035 (0)	7,0	1,4290 (+1)
−1,9	1,4225 (−1)	1,8	2,5393 (0)	7,2	1,4886 (+1)
−1,8	1,5642 (−1)	1,9	2,6794 (0)	7,4	1,5491 (+1)
−1,7	1,7193 (−1)			7,6	1,6104 (+1)
−1,6	1,8889 (−1)	2,0	2,8237 (0)	7,8	1,6725 (+1)
		2,1	2,9722 (0)		
−1,5	2,0740 (−1)	2,2	3,1249 (0)	8,0	1,7355 (+1)
−1,4	2,2759 (−1)	2,3	3,2816 (0)	8,2	1,7993 (+1)
−1,3	2,4959 (−1)	2,4	3,4423 (0)	8,4	1,8639 (+1)
−1,2	2,7353 (−1)			8,6	1,9293 (+1)
−1,1	2,9955 (−1)	2,5	3,6070 (0)	8,8	1,9954 (+1)
		2,6	3,7755 (0)		
−1,0	3,2780 (−1)	2,7	3,9480 (0)	9,0	2,0624 (+1)
−0,9	3,5841 (−1)	2,8	4,1241 (0)	9,2	2,1301 (+1)
−0,8	3,9154 (−1)	2,9	4,3040 (0)	9,4	2,1986 (+1)
−0,7	4,2733 (−1)			9,6	2,2678 (+1)
−0,6	4,6595 (−1)	3,0	4,4876 (0)	9,8	2,3378 (+1)
		3,1	4,6747 (0)		
−0,5	5,0754 (−1)	3,2	4,8653 (0)	10,0	2,4085 (+1)
−0,4	5,5224 (−1)	3,3	5,0595 (0)		
		3,4	5,2571 (0)		

ЛИТЕРАТУРА

1. *Дж. Блекмор*, Статистика электронов в полупроводниках, «Мир», М., 1964.

Приложение IV. Дельта-функция

Одномерная дельта-функция Дирака $\delta(x)$ определяется соотношениями

$$\int_a^b f(x)\,\delta(x)\,dx = \begin{cases} f(0), & a<0<b, \\ {}^1/_2 f(0), & a=0 \text{ или } b=0, \\ 0, & 0<a<b \text{ или } a<b<0, \end{cases} \tag{П.IV.1}$$

$$\delta(x)=0, \qquad x\neq 0. \tag{П.IV.2}$$

Здесь $f(x)$ — любая регулярная функция; значения пределов a и b произвольны (в частности, возможен и предельный переход $a\to-\infty$ и/или $b\to+\infty$).

Определения (П.IV.1), (П.IV.2) легко обобщаются на многомерный случай: многомерная δ-функция есть произведение одномерных. Так, если $\mathbf{r}$ есть вектор с компонентами x, y, z, то

$$\delta(\mathbf{r})=\delta(x)\,\delta(y)\,\delta(z), \tag{П.IV.3}$$

$$\int dx\,dy\,dz\,f(x,\,y,\,z)\,\delta(\mathbf{r})=f(0,\,0,\,0) \tag{П.IV.4}$$

при условии, что точка $\mathbf{r}=0$ лежит внутри области интегрирования.

Согласно (П.IV.1) и (П.IV.2) $\delta(x)$ не относится к числу функций, рассматриваемых в классическом анализе. Действительно, интеграл

$$\int \delta(x)\,dx,$$

взятый по любому отрезку, содержащему точку $x=0$, конечен, в то время как «функция» $\delta(x)$ отлична от нуля лишь на множестве меры нуль. Все же определениям (П.IV.1), (П.IV.2) можно придать точный смысл, рассматривая $\delta(x)$ как символ некоторого предельного перехода, выполняемого над интегралом от регулярных функций. Так, по теореме Фурье мы имеем при любой достаточно регулярной функции $f(x)$

$$\lim_{a\to\infty}\int_{-\infty}^{+\infty} f(x)\,\frac{\sin ax}{x}\,dx=\pi f(0). \tag{П.IV.5}$$

Сравнивая это соотношение с (П.IV.1), получаем следующее символическое представление δ-функции:

$$\delta(x)=\frac{1}{\pi}\lim_{a\to\infty}\frac{\sin ax}{x}. \tag{П.IV.6}$$

Существует много соотношений такого типа. Все они, как и (П.IV.6), приобретают точный смысл, лишь будучи проинтегрированы с какой-либо регулярной функцией *).

Из определений (П.IV.1), (П.IV.2) вытекают следующие правила преобразования δ-функций:

---

*) Понимая соотношение (П.IV.6) буквально, мы получили бы

$$\delta(0)=\frac{1}{\pi}\lim_{a\to\infty} a=\infty.$$

Это «соотношение» часто используется как дополнение к (П.IV.2).

1) Пусть C — постоянная. Тогда

$$\delta(Cx) = |C|^{-1}\delta(x). \qquad (\text{П.IV.7})$$

2) Пусть $\varphi(x)$ — непрерывная функция, все вещественные корни которой простые. Обозначим их через x_n ($n = 1, 2, \ldots$). Тогда

$$\delta[\varphi(x)] = \sum_n \frac{\delta(x - x_n)}{|\varphi'(x_n)|}. \qquad (\text{П.IV.8})$$

Можно определить и производные любого порядка от δ-функции. Действительно, пользуясь формальным правилом интегрирования по частям и соотношением (П.IV.1), мы имеем

$$\int_a^b f(x)\,\delta^{(n)}(x)\,dx = (-1)^n f^{(n)}(0) \qquad a < 0 < b, \qquad (\text{П.IV.9})$$

где n — любое целое число.

Приложение V. **Рекомбинация через многозарядные ловушки**

Пусть ловушка может захватывать 0, 1, 2, ..., M электронов, создавая в запрещенной зоне энергий M локальных уровней. Тогда суммарный темп захвата электронов из зоны проводимости на уровень j (при этом захваченный электрон становится j-м) можно выразить формулой, аналогичной (IX. 4.7):

$$R_{nj} = \alpha_{nj} N_t (f_{j-1} n - f_j n_j). \qquad (\text{П.V.1})$$

Здесь α_{nj} — коэффициент захвата электронов на уровень j (до захвата ловушка имела $(j-1)$ электронов); $N_t f_j$ и $N_t f_{j-1}$ — неравновесные концентрации ловушек в j-м и, соответственно, в $(j-1)$-м зарядных состояниях; n_j — величина, аналогичная n_1 для однозарядных ловушек, равная (с точностью до отношения статистических весов) равновесной концентрации электронов в зоне при совпадении уровня Ферми с уровнем E_j. Совершенно так же суммарный темп захвата дырок на уровень j (до захвата дырки ловушка имела j электронов) выражается формулой, аналогичной (IX. 4.7а):

$$R_{pj} = \alpha_{pj} N_t (f_j p - f_{j-1} p_j), \qquad (\text{П.V.2})$$

где α_{pj} — коэффициент захвата дырки на уровень E_j, а p_j — равновесная концентрация дырок при совпадении уровня Ферми с уровнем E_j. Полные темпы захвата электронов и дырок на ловушки в различных зарядных состояниях равны

$$R_n = \sum_{j=1}^{M} R_{nj}, \qquad R_p = \sum_{j=1}^{M} R_{pj}. \qquad (\text{П.V.3})$$

В стационарном состоянии концентрация ловушек в каждом зарядном состоянии не зависит от времени. Это дает

$$R_{n1} = R_{p1}, \qquad R_{nM} = R_{pM},$$
$$R_{nj} + R_{p(j+1)} = R_{n(j+1)} + R_{pj} \qquad (j \neq 1, M).$$

Отсюда следует, что

$$R_{nj} = R_{pj} \qquad (j = 1, 2, \ldots, M). \qquad (\text{П.V.4})$$

Кроме того, мы имеем еще условие постоянства полной концентрации ловушек

$$\sum_{j=0}^{M} f_j = 1. \qquad (\text{П.V.5})$$

Написанные уравнения вполне определяют задачу. Условия (П.V.4) и (П.V.5) дают $(M+1)$ уравнений для определения всех неравновесных чисел заполнения f_j. В частности, уравнения (П.V.4) дают

$$f_j = f_{j-1}\frac{\alpha_{nj}n + \alpha_{pj}p_j}{\alpha_{nj}n_j + \alpha_{pj}p}. \qquad (\text{П.V.6})$$

Подставляя это в выражения (П.V.1) и (П.V.2), после несложных преобразований получаем

$$R_{nj} = R_{pj} = R_j = (f_j + f_{j-1})\frac{np - n_i^2}{(\alpha_{pj}N_t)^{-1}(n+n_j) + (\alpha_{nj}N_t)^{-1}(p+p_j)}. \qquad (\text{П.V.7})$$

При выводе этой формулы мы считали полупроводник невырожденным и, соответственно, полагали $n_j p_j = n_i^2$, где n_i^2 есть квадрат собственной концентрации электронов.

Формула (П.V.7) выражает темп рекомбинации, обусловленный одним энергетическим уровнем ловушек. Если ловушки могут захватывать только один электрон, то $M=1$, $f_j + f_{j-1} = f_1 + f_0 = 1$ и формула (П.V.7) переходит в формулу (IX.6.6) для простых ловушек.

Полученные формулы упрощаются для малой концентрации ловушек (когда $\delta p \simeq \delta n$) и малого нарушения равновесия. Тогда, полагая в формуле (П.V.7)

$$n = n_0 + \delta n, \qquad p = p_0 + \delta p, \qquad f_j = f_j^0 + \delta f_j$$

и удерживая только члены первого порядка малости, получаем

$$R_j = (f_j^0 + f_{j-1}^0)\frac{n_0 + p_0}{(\alpha_{pj}N_t)^{-1}(n_0+n_j) + (\alpha_{nj}N_t)^{-1}(p_0+p_j)}\delta n, \qquad (\text{П.V.8})$$

Поэтому общее время жизни электронов и дырок τ выражается формулой

$$\frac{1}{\tau} = \frac{1}{\delta n}\sum_{j=1}^{M} R_j = (n_0+p_0)\sum_{j=1}^{M}\frac{f_j^0 + f_{j-1}^0}{(\alpha_{pj}N_t)^{-1}(n_0+n_j) + (\alpha_{nj}N_t)^{-1}(p_0+p_j)}, \qquad (\text{П.V.9})$$

в которую входят только равновесные концентрации носителей заряда n_0 и p_0 и равновесные числа заполнения уровней f_j^0.

Отметим в заключение два важных предельных случая. Положим, что мы имеем сильно легированный полупроводник p-типа, так что равновесный уровень Ферми расположен ниже наинизшего уровня ловушек E_1 (как на рис. 9.8, $F = F_1$). Тогда $f_0^0 \simeq 1$, а все другие $f_j^0 \ll 1$, и поэтому в формуле (П.V.9) остается только одно слагаемое с $j=1$. При этом $p_0 \gg n_0$, $n_0 \ll n_1$, $p_0 \gg p_1$. Если равновесная концентрация дырок достаточно велика, так что $(\alpha_{n1}N_t)^{-1}p_0 \gg (\alpha_{p1}N_t)^{-1}n_1$, то из формулы (П.V.9) следует, что время жизни, так же как и для простых ловушек, достигает предельного значения, не зависящего от концентрации основных носителей p_0 и равного $\tau_{n0} = (\alpha_{n1}N_t)^{-1}$.

Аналогично, для сильно легированного полупроводника n-типа, когда равновесный уровень Ферми становится выше наивысшего (M-го) уровня ловушек, мы имеем $f_M \simeq 1$, $f_j \ll 1$ $(j \neq M)$. Кроме того, $n_0 \gg p_0$, $n_M \ll n_0$, $p_0 \ll p_M$. Поэтому при больших концентрациях электронов, когда выполняется неравенство $(\alpha_{pM}N_t)^{-1}n_0 \gg (\alpha_{nM}N_t)^{-1}p_M$, τ становится равным другому предельному значению $\tau_{p0} = (\alpha_{pM}N_t)^{-1}$, которое тоже не зависит от концентрации основных носителей n_0.

Приложение VI. Интеграл поверхностной проводимости

Как и в гл. X, мы ограничимся случаем невырожденных полупроводников для которых справедливы соотношения (X.2.3). Кроме того, будем считать, что доноры и акцепторы полностью ионизованы (что имеет место, например, в германии и кремнии с примесями элементов III и V групп при температурах выше температуры жидкого азота). Тогда, следуя работе [1], изменение электропроводности ΔG можно вычислить следующим образом. В рассматриваемом случае объемная плотность заряда равна

$$\rho = e\,[(p - p_0) - (n - n_0)] = e\,[p_0\,(e^{-Y} - 1) - n_0\,(e^{Y} - 1)].$$

Поэтому уравнение Пуассона для безразмерного потенциала Y

$$\frac{d^2Y}{dx^2} = -\frac{4\pi}{\varepsilon}\frac{e}{kT}\rho$$

принимает вид

$$\frac{d^2Y}{dx^2} = -\frac{1}{L_i^2}\left[\xi\,(e^{-Y} - 1) - \frac{1}{\xi}\,(e^{Y} - 1)\right]. \qquad (\text{П.VI.1})$$

Здесь $\xi = (p_0/n_0)^{1/2}$, а L_i — длина Дебая в собственном полупроводнике:

$$L_i^2 = \frac{\varepsilon kT}{4\pi e^2 n_i}. \qquad (\text{П.VI.2})$$

Граничные условия задачи:

$$x = 0: \quad Y = Y_s; \qquad x = \infty: \quad \frac{dY}{dx} = Y = 0. \qquad (\text{П.VI.3})$$

Умножая обе части уравнения (П.VI.1) на dY/dx, получаем

$$d\left(\frac{dY}{dx}\right)^2 = -\frac{2}{L_i^2}\left[\xi\,(e^{-Y} - 1) - \frac{1}{\xi}\,(e^{Y} - 1)\right]dY.$$

Интегрируя обе части этого уравнения от $Y = 0$ до произвольного значения Y, находим

$$\frac{dY}{dx} = \frac{\sqrt{2}}{L_i}F\,(Y,\ \xi), \qquad (\text{П.VI.4})$$

где

$$F\,(Y,\ \xi) = \mp\left[\xi\,(e^{-Y} - 1) + \frac{1}{\xi}\,(e^{Y} - 1) + \left(\xi - \frac{1}{\xi}\right)Y\right]^{1/2}. \qquad (\text{П.VI.5})$$

Так как при $Y_s > 0$ потенциал Y убывает с увеличением x, то в формуле (П.VI.5) для этого случая нужно брать знак минус. При $Y_s < 0$ следует выбирать знак плюс. Формулы (П.VI.4) и (П.VI.5) выражают первый интеграл уравнения Пуассона и дают распределение градиента потенциала (электрического поля) в функции самого потенциала Y.

С другой стороны, в интегралах Γ_p и Γ_n (§ X.2) интегрирование по x можно заменить интегрированием по Y:

$$\Gamma_p = \int_0^\infty [p\,(x) - p_0]\,dx = \int_{Y_s}^0 [p\,(Y) - p_0]\frac{dx}{dY}\,dY,$$
$$\Gamma_n = \int_{Y_s}^0 [n\,(Y) - n_0]\,\frac{dx}{dY}\,dY. \qquad (\text{П.VI.6})$$

Подставляя сюда dx/dY из формулы (П.VI.4) и выражая $p(Y)$ и $n(Y)$ по формулам (X.2.3), мы получаем для изменения проводимости ΔG выражение

$$\Delta G = c\mu_p(\Gamma_p + b\Gamma_n) = \frac{1}{\sqrt{2}} e\mu_p L_i n_i \int_{Y_s}^{0} \frac{\xi(e^{-Y}-1) + \frac{b}{\xi}(e^{Y}-1)}{F(Y, \xi)} dY. \quad (\text{П.VI.7})$$

Входящий сюда «интеграл поверхностной проводимости» в общем случае не выражается аналитически и требует численных расчетов. Кривые рис. 10.7 дают значения этого интеграла (умноженного на $\xi^{1/2}$) как функцию Y_s.

Из формулы (П.VI.7) непосредственно получается величина Y_{sm}, соответствующая минимуму ΔG. Дифференцируя эту формулу по Y_s и приравнивая производную нулю, мы имеем

$$\frac{\xi(e^{-Y_{sm}}-1) + \frac{b}{\xi}(e^{Y_{sm}}-1)}{F(Y_{sm}, \xi)} = 0. \quad (\text{П.VI.8})$$

Условие обращения в нуль числителя приводит к квадратному уравнению относительно $\exp Y_{sm}$, которое дает два корня:

$$\exp Y_{sm} = 1, \qquad \exp Y_{sm} = \xi^2/b.$$

Однако первый корень не есть решение уравнения (П.VI.8), так как при этом и $F = 0$. Поэтому остается только один, второй корень, который и дает формулу (X.2.5).

Найдем еще величину подвижного объемного заряда в приповерхностном слое, находящегося под единицей площади поверхности. Она равна

$$Q_V = e(\Gamma_p - \Gamma_n) = \frac{en_i L_i}{\sqrt{2}} \int_{Y_s}^{0} \frac{\xi(e^{-Y}-1) - \frac{1}{\xi}(e^{Y}-1)}{F(Y, \xi)} dY =$$

$$= -\sqrt{2}\, en_i L_i \int_{Y_s}^{0} \frac{dF(Y, \xi)}{dY} dY.$$

Учитывая, что $F(0, \xi) = 0$, получаем

$$Q_V = \sqrt{2}\, en_i L_i F(Y_s, \xi). \quad (\text{П.VI.9})$$

В более сложных случаях, когда доноры и акцепторы не полностью ионизованы или электронный (дырочный) газ вырожден, задача решается аналогично. Примеры таких расчетов можно найти в работах [2, 3].

ЛИТЕРАТУРА

1. *C. Garrett, W. Brattain*, Phys. Rev. **99**, 376 (1955).
2. *R. Seiwatz, M. Green*, Journ. Appl. Phys. **29**, 1034 (1958).
3. *Ю. И. Горкун*, ФТТ. **3**, 1061 (1961).

Приложение VII. Диффузия неравновесных носителей заряда в магнитном поле

Строгая теория фотоэлектромагнитного эффекта в произвольном магнитном поле требует решения кинетических уравнений Больцмана для дырок и электронов (см. например, [1]). Однако интересующие нас соотношения (XI.7.12) — (XI.7.14) можно получить из более простых феноменологических рассуждений.

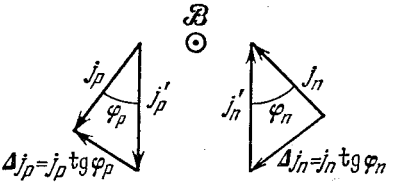


Рис. П.VII.

Будем, как и раньше, считать полупроводник изотропным. Запишем плотности токов дырок и электронов в магнитном поле в виде

$$\mathbf{j}_p = \mathbf{j}_p' + \Delta\mathbf{j}_p, \qquad \mathbf{j}_n = \mathbf{j}_n' + \Delta\mathbf{j}_n,$$

где

$$\mathbf{j}_p' = \sigma_p \mathcal{E} - eD_p \cdot \nabla p, \qquad \mathbf{j}_n' = \sigma_n \mathcal{E} + eD_n \cdot \nabla n$$

— плотности токов, обусловленные дрейфом в электрическом поле и диффузией, а $\Delta\mathbf{j}_p$ и $\Delta\mathbf{j}_n$ — добавки, возникающие под действием силы Лоренца. Эффектом магнетосопротивления будем пренебрегать и, соответственно, считать σ и D не зависящими от магнитного поля. Так как сила в магнитном поле перпендикулярна направлению тока, то $\Delta\mathbf{j}_p \perp \mathbf{j}_p$ и $\Delta\mathbf{j}_n \perp \mathbf{j}_n$. Тогда (рис. П.VII)

$$\Delta j_p = j_p \operatorname{tg}\varphi_p = j_p \frac{1}{c}\mu_{pH}\mathcal{B}, \qquad \Delta j_n = j_n \operatorname{tg}\varphi_n = j_n \frac{1}{c}\mu_{nH}\mathcal{B},$$

где φ_p и φ_n — абсолютные значения холловских углов. Поэтому для плотностей токов в магнитном поле мы получаем

$$\mathbf{j}_p = \mathbf{j}_p' + \frac{1}{c}\mu_{pH}[\mathbf{j}_p \times \mathcal{B}], \tag{П.VII.1}$$

$$\mathbf{j}_n = \mathbf{j}_n' - \frac{1}{c}\mu_{nH}[\mathbf{j}_n \times \mathcal{B}]. \tag{П.VII.2}$$

Обобщение этих соотношений для анизотропных полупроводников дано в [2]. Кроме того, мы имеем уравнение непрерывности

$$-\frac{1}{e}\operatorname{div}\mathbf{j}_p - \frac{\delta p}{\tau_p} = 0, \tag{П.VII.3}$$

$$\frac{1}{e}\operatorname{div}\mathbf{j}_n - \frac{\delta n}{\tau_n} = 0. \tag{П.VII.4}$$

Применим теперь эти уравнения к задаче о пластинке в магнитном поле (§ XI.7). В дальнейшем для простоты мы положим $\mu_{nH} = \mu_n$, $\mu_{pH} = \mu_p$. Полупроводник будем считать невырожденным и, соответственно, $\mu_n/\mu_p = D_n/D_p$. Далее, для сокращения введем обозначения:

$$\frac{1}{c}\mu_n\mathcal{B} \equiv \Theta_n, \qquad \frac{1}{c}\mu_p\mathcal{B} \equiv \Theta_p, \qquad b \equiv \frac{\mu_n}{\mu_p}. \tag{П.VII.5}$$

Тогда в координатах (направленных, как на рис. 11.15) уравнения (П.VII.1) и (П.VII.2) будут

$$bj_{px} = e\mu(p_0 + \delta p)\mathcal{E}_x + \Theta j_{py}, \tag{П.VII.6}$$

$$bj_{py} = e\mu(p_0 + \delta p)\mathcal{E}_y - eD\frac{dp}{dy} - \Theta j_{px}, \tag{П.VII.7}$$

$$j_{nx} = e\mu(n_0 + \delta n)\mathcal{E}_x - \Theta j_{ny}, \tag{П.VII.8}$$

$$j_{ny} = e\mu(n_0 + \delta n)\mathcal{E}_y + eD\frac{dn}{dy} + \Theta j_{nx}. \tag{П.VII.9}$$

Здесь величины μ, D и Θ относятся к электронам, а индекс n мы не выписываем. Уравнения непрерывности:

$$\frac{d}{dy} j_{py} = -e\frac{\delta p}{\tau_p}, \tag{П.VII.10}$$

$$\frac{d}{dy} j_{ny} = e\frac{\delta n}{\tau_n}. \tag{П.VII.11}$$

Кроме того, для нашей задачи

$$j_{py} + j_{ny} = 0. \tag{П.VII.12}$$

Исключим из уравнений (П.VII.7) и (П.VII.9) $\mathscr{E}_y$ и учтем соотношение (П.VII.12). Это дает

$$j_{py}\,[b\,(n_0+\delta n)+(p_0+\delta p)] =$$
$$= -eD\left[(n_0+\delta n)\frac{dp}{dy}+(p_0+\delta p)\frac{dn}{dy}\right]-\Theta\,[(n_0+\delta n)\,j_{px}+(p_0+\delta p)\,j_{nx}]. \tag{П.VII.13}$$

Здесь, согласно уравнениям (П.VII.13) и (П.VII.11),

$$\frac{dp}{dy}=\frac{d\delta p}{dy}=-\frac{\tau_p}{e}\frac{d^2 j_{py}}{dy^2},$$
$$\frac{dn}{dy}=\frac{\tau_n}{e}\frac{d^2 j_{ny}}{dy^2}=-\frac{\tau_n}{e}\frac{d^2 j_{py}}{dy^2}.$$

Далее, используя уравнения (П.VII.6), (П.VII.8) и (П.VII.12), мы имеем

$$(n_0+\delta n)\,j_{px}+(p_0+\delta p)\,j_{nx} =$$
$$= e\mu\,(p_0+\delta p)\,(n_0+\delta n)\left(1+\frac{1}{b}\right)\mathscr{E}_x+\Theta\left[\frac{1}{b}\,(n_0+\delta n)+(p_0+\delta p)\right]j_{py}.$$

Подставляя эти выражения в уравнение (П.VII.13), мы получаем для тока j_{py} дифференциальное уравнение 2-го порядка

$$D\,[\tau_p\,(n_0+\delta n)+\tau_n\,(p_0+\delta p)]\,\frac{d^2 j_{py}}{dy^2}-$$
$$-\left\{b\,(n_0+\delta n)+(p_0+\delta p)+\Theta^2\left[\frac{1}{b}\,(n_0+\delta n)+(p_0+\delta p)\right]\right\}j_{py}-$$
$$-e\mu\Theta\,(p_0+\delta p)\,(n_0+\delta n)\left(1+\frac{1}{b}\right)\mathscr{E}_x=0. \tag{П.VII.14}$$

Это уравнение нелинейно, и поэтому для произвольной интенсивности света точное вычисление $j_{py}(y)$ сложно.

Для интересующего нас случая короткого замыкания $\mathscr{E}_x=0$. Если, кроме того, освещенность не слишком велика, так что выполняются условия

$$\delta n,\ \delta p \ll n_0+p_0, \qquad \tau_p\delta n+\tau_n\delta p \ll \tau_p n_0+\tau_n p_0, \tag{П.VII.15}$$

то уравнение (П.VII.14) становится линейным и принимает вид

$$\frac{d^2 j_{py}}{dy^2}-\frac{1}{L^{*2}}\,j_{py}=0, \tag{П.VII.16}$$

где

$$L^{*2}=D_n\frac{\tau_p n_0+\tau_n p_0}{b n_0+p_0+\Theta_n^2\left(\frac{n_0}{b}+p_0\right)}. \tag{П.VII.17}$$

Поэтому для толстого образца ($j_{py}(d) \ll j_{py}(0)$) уравнение (П.VII.16) дает

$$j_{py} = j_{py}(0)\, e^{-y/L^*}. \qquad (\text{П.VII.18})$$

Отсюда видно, что L^* есть длина диффузии в магнитном поле, на которой ток j_{py} затухает в e раз. Выражение (П.VII.17) соответствует формулам (XI.7.12)—(XI.7.14), в которых D_n выражено через коэффициент амбиполярной диффузии D и отношение подвижностей b и учтено, что $\Theta_n = b\Theta_p$.

Вычисление $j_{p0}(0)$ у освещенной поверхности производится так же, как и без магнитного поля. Однако при $\delta p \neq \delta n$ нужно учесть, что скорости поверхностной рекомбинации для дырок s_p и электронов s_n неодинаковы (§ X.5). Поэтому граничное условие будет

$$g_s = \frac{1}{e} j_{py}(0) + R_s, \qquad (\text{П.VII.19})$$

где

$$R_s = s\frac{n_0}{n_0+p_0}\delta p(0) + s\frac{p_0}{n_0+p_0}\delta n(0), \qquad (\text{П.VII.20})$$

а s есть результирующая скорость поверхностной рекомбинации на освещаемой поверхности, выражаемая формулой (X.5.8).

В формуле (П.VII.20) можно выразить $\delta p(0)$ и $\delta n(0)$ через $j_{py}(0)$, пользуясь уравнениями непрерывности (П.VII.10) и (П.VII.11) и законом затухания (П.VII.18). Это дает

$$\delta p = -\frac{\tau_p}{e}\frac{dj_{py}}{dy} = j_{py}(0)\frac{\tau_p}{eL^*}e^{-y/L^*}, \qquad (\text{П.VII.21})$$

$$\delta n = -\frac{\tau_n}{e}\frac{dj_{py}}{dy} = j_{py}(0)\frac{\tau_n}{eL^*}e^{-y/L^*}. \qquad (\text{П.VII.22})$$

Отсюда

$$\delta p(0) = \frac{\tau_p}{eL^*} j_{py}(0), \qquad \delta n(0) = \frac{\tau_n}{eL^*} j_{py}(0). \qquad (\text{П.VII.23})$$

Тогда из граничного условия (П.VII.19) получается

$$j_{py}(0) = \frac{eg_s}{1+s\tau_{\text{фэм}}/L^*} = \frac{eg_s}{1+sL^*/D^*}, \qquad (\text{П.VII.24})$$

где $\tau_{\text{фэм}}$ определяется формулой (XI.7.13).

Подставляя в формулу (XI.7.8) j_{py} из (П.VII.18) и учитывая (П.VII.24), мы опять приходим к формуле (XI.7.9) для $i^{\text{кз}}$. Однако в данном случае время жизни пар τ везде заменяется на комбинированное время $\tau_{\text{фэм}}$, а вместо L и D входят L^* и D^*.

Остановимся еще на фотопроводимости в магнитном поле.

Сила тока фотопроводимости есть

$$i_{\text{фп}} = e\mathscr{E}_x \int_0^d (\mu_n \delta n + \mu_p \delta p)\, dy. \qquad (\text{П.VII.25})$$

Подставляя сюда для δp и δn их выражения (П.VII.21) и (П.VII.22) и выполняя интегрирование, находим для толстого образца

$$i_{\text{фп}} = \mathscr{E}_x \tau_{\text{фп}} (\mu_n + \mu_p)\, j_{py}(0), \qquad (\text{П.VII.26})$$

где $\tau_{\text{фп}}$ есть введенное ранее (§ VII.4) время жизни фотопроводимости.

В методе компенсации ФЭМ эффекта и фотопроводимости (§ XI.7) мы имеем

$$\Theta L^* = \mathscr{E}_x \tau_{\text{фп}} (\mu_n + \mu_p). \qquad (\text{П.VII.27})$$

В эту формулу удобно ввести новое характерное время

$$\tau_к = \frac{\tau_{фп}^2}{\tau_{фэм}} \qquad (П.VII.28)$$

и новую характерную длину

$$l^* = \sqrt{D^* \tau_к}. \qquad (П.VII.29)$$

Тогда равенство (П.VII.27) можно записать в виде

$$l^* = \frac{\Theta D^*}{\mathscr{E}_x (\mu_n + \mu_p)} = \frac{1}{c} D^* \frac{\mathscr{B}}{\mathscr{E}_x}. \qquad (П.VII.30)$$

Оно по форме совпадает (при $\mu_{pH} = \mu_p$ и $\mu_{nH} = \mu_n$) с формулой (XI.7.11) и является обобщением последней на случай не слабого магнитного поля.

Из формул (П.VII.28) — (П.VII.30) видно, что измеряемое методом компенсации время $\tau_к$ в общем случае не совпадает ни с $\tau_{фэм}$, ни с $\tau_{фп}$. Однако при $\tau_p = \tau_n = \tau$ мы имеем $\tau_{фэм} = \tau_{фп} = \tau$, и в этом случае $\tau_к$ дает время жизни электронно-дырочных пар τ.

ЛИТЕРАТУРА

1. *А. И. Ансельм*, ЖТФ 24, 2064 (1954).
2. *W. van Roosbroeck*, *Phys. Rev.* **101**, 1713 (1956).

Приложение VIII. **Вычисление сумм (XII.2.6)**

Согласно (XII.1.1) сумму (XII.2.6) можно переписать в виде

$$S \equiv \sum_{\mathbf{g}} e^{i(\mathbf{k}+\mathbf{k}', \rho)} = \prod_{\alpha = x, y, z} \sum_{g_\alpha = 0}^{G_\alpha} e^{i(k_\alpha + k'_\alpha) g_\alpha a_\alpha} \equiv \prod_{\alpha = x, y, z} S_\alpha \qquad (П.VIII.1)$$

(без суммирования по α в экспоненте).

Поскольку

$$L_\alpha = G_\alpha a_\alpha, \qquad k_\alpha = \frac{2\pi}{L_\alpha} n_\alpha \qquad (\alpha = x, y, z),$$

мы имеем по формуле геометрической прогрессии

$$S_\alpha = \frac{1 - \exp 2\pi i (n_\alpha + n'_\alpha)}{1 - \exp \frac{2\pi i}{L_\alpha} (n_\alpha + n'_\alpha) a_\alpha} = \begin{cases} G_\alpha, & n_\alpha + n'_\alpha = 0, \\ 0, & n_\alpha + n'_\alpha \neq 0. \end{cases} \qquad (П.VIII.2)$$

Подставляя это выражение в формулу (П.VIII.1), получаем результат (XII.2.6).

Приложение IX. **Вывод условия ортогональности (XII.2.11)**

Опуская для краткости аргумент q, перепишем еще раз уравнение (XII.2.10):

$$\sum_{h, \alpha} \Gamma_{hh'}^{\alpha\alpha'} \zeta_{h\alpha}(s) = \omega_s^2 M_{h'} \zeta_{h'\alpha'}(s). \qquad (П.IX.1)$$

С учетом вещественности ω_s^2, вытекающей из (XII.2.10), можем написать сопряженное уравнение (при том же $\mathbf{q}$ и другом, вообще говоря, значении s) в виде

$$\sum_{h, \alpha} \left(\Gamma_{hh'}^{\alpha\alpha'}\right)^* \zeta_{h\alpha}^*(s') = \omega_{s'}^2 M_{h'} \zeta_{h'\alpha'}^*(s'). \qquad (П.IX.2)$$

Умножим уравнения (П.IX.1) и (П.IX.2) соответственно на $\zeta^*_{h'\alpha'}(s')$, $\zeta_{h'\alpha'}(s)$ и просуммируем по h', α'. Получим

$$\sum_{h',\,\alpha'}\sum_{h,\,\alpha}\Gamma^{\alpha\alpha'}_{hh'}\zeta^*_{h'\alpha'}(s')\,\zeta_{h\alpha}(s)=\omega_s^2\sum_{h',\,\alpha'}M_{h'}\zeta^*_{h'\alpha'}(s')\,\zeta_{h'\alpha'}(s), \quad (\text{П.IX.1}')$$

$$\sum_{h',\,\alpha'}\sum_{h,\,\alpha}\left(\Gamma^{\alpha\alpha'}_{hh'}\right)^*\zeta_{h'\alpha'}(s)\,\zeta^*_{h\alpha}(s')=\omega_{s'}^2\sum_{h',\,\alpha'}M_{h'}\zeta^*_{h'\alpha'}(s')\,\zeta_{h'\alpha'}(s). \quad (\text{П.IX.2}')$$

Примем теперь во внимание равенство (XII.2.12) и вычтем почленно уравнение (П.IX.2′) из (П.IX.1′):

$$\sum_{h,\,h',\,\alpha,\,\alpha'}\left\{\Gamma^{\alpha\alpha'}_{hh'}\zeta^*_{h'\alpha'}(s')\,\zeta_{h\alpha}(s)-\Gamma^{\alpha'\alpha}_{h'h}\zeta^*_{h\alpha}(s')\,\zeta_{h'\alpha'}(s)\right\}=$$

$$=(\omega_s^2-\omega_{s'}^2)\sum_{h',\,\alpha'}M_{h'}\zeta^*_{h'\alpha'}(s')\,\zeta_{h'\alpha'}(s). \quad (\text{П.IX.3})$$

Выполняя во втором слагаемом в левой части (П.IX.3) замену индексов суммирования

$$h\rightleftarrows h', \qquad \alpha\rightleftarrows\alpha',$$

видим, что оно совпадает с первым. Таким образом, равенство (П.IX.3) принимает вид

$$(\omega_s^2-\omega_{s'}^2)\sum_{h',\,\alpha'}M_{h'}\zeta^*_{h'\alpha'}(s')\,\zeta_{h'\alpha'}(s)=0. \quad (\text{П.IX.4})$$

Следовательно, при $\omega_s^2\neq\omega_{s'}^2$

$$\sum_{h',\,\alpha'}M_{h'}\zeta^*_{h'\alpha'}(s')\,\zeta_{h'\alpha'}(s)=0. \quad (\text{П.IX.5})$$

При однозначной зависимости s от ω_s^2 отсюда вытекает условие (XII.2.11). В противном случае частоты могут оказаться одинаковыми и при $s'\neq s$. Это, однако, означает вырождение: одному и тому же значению ω_s^2 будет отвечать несколько решений $\zeta_{h\alpha}(s)$ уравнения (XII.2.10). Всегда можно выбрать такую их линейную комбинацию, чтобы равенство (П.IX.5), т. е. условие (XII.2.11), все же удовлетворялось при $s'\neq s$.

Приложение X. **Переход от суммирования по дискретным компонентам квазиимпульса к интегрированию**

Рассмотрим сумму

$$S=\sum_{n_x,\,n_y,\,n_z}f(\mathbf{p}), \quad (\text{П.X.1})$$

где $f(\mathbf{p})$ — любая регулярная функция, $p_x=\dfrac{2\pi\hbar}{L}n_x$, $p_y=\dfrac{2\pi\hbar}{L}n_y$, $p_z=\dfrac{2\pi\hbar}{L}n_z$, а n_x, n_y, n_z — положительные или отрицательные целые числа или нули. Разность между соседними значениями p_α ($\alpha=x,\ y,\ z$) составляет

$$\Delta p_\alpha=\frac{2\pi\hbar}{L}.$$

Таким образом, выражение (П.X.1) можно переписать в виде

$$S=\left(\frac{L}{2\pi\hbar}\right)^3\sum_{n_x,\,n_y,\,n_z}f(\mathbf{p})\,\Delta p_x\,\Delta p_y\,\Delta p_z. \quad (\text{П.X.2})$$

Пусть при $L\to\infty$ функция $f(\mathbf{p})$ не зависит от L. В правой части (П.X.2) стоит интегральная сумма, и при $L\to\infty$ мы получаем

$$S=\frac{V}{(2\pi\hbar)^3}\int f(\mathbf{p})\,d\mathbf{p},\qquad V=L^3. \tag{П.X.3}$$

Величина S/V при этом остается конечной и не зависит от V.

Соотношение (П.X.3) составляет правило перехода от суммирования по n_x, n_y, n_z к интегрированию по трехмерному $\mathbf{p}$-пространству. Такое же правило применимо и в случае одного или двух измерений: надо лишь заменить в (П.X.3) $V=L^3$ на L или L^2, а тройной интеграл — на однократный или двойной.

Приложение XI. Гамильтониан взаимодействия электронов с акустическими фононами

Согласно (XIV.3.5) напишем гамильтониан взаимодействия электронов с акустическими фононами в виде (опускаем временно сумму по s):

$$H'=E_{\alpha\beta}\frac{\partial Q_\alpha}{\partial x_\beta}. \tag{П.XI.1}$$

Не предполагая пока симметричности тензора $E_{\alpha\beta}$, представим его в виде суммы симметричной и антисимметричной частей:

$$E_{\alpha\beta}=E^s_{\alpha\beta}+E^a_{\alpha\beta}, \tag{П.XI.2}$$

где

$$E^s_{\alpha\beta}=\frac{1}{2}(E_{\alpha\beta}+E_{\beta\alpha}),\qquad E^a_{\alpha\beta}=\frac{1}{2}(E_{\alpha\beta}-E_{\beta\alpha}). \tag{П.XI.3}$$

Соответственно

$$H'=\frac{1}{2}E^s_{\alpha\beta}\left(\frac{\partial Q_\alpha}{\partial x_\beta}+\frac{\partial Q_\beta}{\partial x_\alpha}\right)+\frac{1}{2}E^a_{\alpha\beta}\left(\frac{\partial Q_\alpha}{\partial x_\beta}-\frac{\partial Q_\beta}{\partial x_\alpha}\right). \tag{П.XI.4}$$

Компоненты антисимметричного тензора можно представить в виде [1]

$$E^a_{\alpha\beta}=\varepsilon_{\alpha\beta\gamma}A_\gamma, \tag{П.XI.5}$$

где A_γ — компоненты некоторого псевдовектора, а $\varepsilon_{\alpha\beta\gamma}$ — совершенно антисимметричный единичный тензор ($\varepsilon_{\alpha\beta\gamma}=-\varepsilon_{\beta\alpha\gamma}=-\varepsilon_{\alpha\gamma\beta}$; $\varepsilon_{123}=1$).

Замечая, что, по определению операции взятия вихря,

$$\varepsilon_{\alpha\beta\gamma}\left(\frac{\partial Q_\alpha}{\partial x_\beta}-\frac{\partial Q_\beta}{\partial x_\alpha}\right)=(\mathrm{rot}\,\mathbf{Q})_\gamma, \tag{П.XI.6}$$

можем переписать второе слагаемое в правой части (П.XI.4) в виде

$$\frac{1}{2}(\mathbf{A},\,\mathrm{rot}\,\mathbf{Q}). \tag{П.XI.7}$$

Как известно, если компоненты $\mathrm{rot}\,\mathbf{Q}$ не зависят от координат, то выражение $\mathrm{rot}\,\mathbf{Q}$ описывает не деформацию, а поворот тела как целого. Соответственно слагаемое вида (П.XI.7) не может входить в выражение для энергии взаимодействия электронов с фононами *), и антисимметричная часть тензора $E_{\alpha\beta}$ тождественно равна нулю:

$$\mathbf{A}=0,\qquad E_{\alpha\beta}=E^s_{\alpha\beta}. \tag{П.XI.8}$$

*) Этот вывод может быть несправедлив, если вектор $\mathbf{Q}$ описывает заведомо неоднородную деформацию, созданную, например, дислокацией.

Гамильтониан H' сводится к произведению некоторого постоянного симметричного тензора на тензор деформации:

$$H'=\frac{1}{2}E^s_{\alpha\beta}\left(\frac{\partial Q_\alpha}{\partial x_\beta}+\frac{\partial Q_\beta}{\partial x_\alpha}\right)\equiv E_{\alpha\beta}\frac{\partial Q_\alpha}{\partial x_\beta}. \qquad (\text{П.XI.9})$$

Тензор $E_{\alpha\beta}$ называется иногда тензором потенциалов деформации; компоненты его суть потенциалы деформации.

Приведем тензор $E_{\alpha\beta}$ к главным осям, обозначив главные его значения через E_α. Тогда соотношение (П.XI.9) примет вид

$$H'=\sum_{\alpha=x,\,y,\,z}E_\alpha\frac{\partial Q_\alpha}{\partial x_\alpha}. \qquad (\text{П.XI.10})$$

Формула (П.XI.10) превращается в (XIV.3.1), если тензор $E_{\alpha\beta}$ вырождается в скаляр, т. е. если $E_x=E_y=E_z=E_1$. Так обстоит дело в кубическом кристалле, если дно зоны проводимости лежит в центре зоны Бриллюэна. Главные оси в данном случае суть оси [100], [010] и [001] в обратной решетке.

Оговорка о расположении минимума зоны проводимости в центре зоны Бриллюэна весьма существенна. Действительно, в других точках зоны Бриллюэна симметрия системы, вообще говоря, ниже кубической: имеется физически выделенное направление вдоль радиус-вектора, связывающего данную точку с центром зоны. Главные оси тензора $E_{\alpha\beta}$ здесь уже не обязаны совпадать с осями куба и могут оказаться неэквивалентными. Соответственно тензор $E_{\alpha\beta}$ может и не выродиться в скаляр даже в кубическом кристалле. Так, в частности, обстоит дело для электронов в германии и кремнии. Формулу (П.XI.9) для такой «многоэллипсоидной» системы следует писать в виде

$$H'=\sum_j E_{\alpha\beta}(j)\frac{\partial Q_\alpha}{\partial x_\beta}, \qquad (\text{П.XI.9}')$$

где индекс j нумерует эллипсоиды и симметричный тензор $E_{\alpha\beta}(j)$ описывает взаимодействие фононов с электронами, находящимися вблизи j-го минимума энергии. Очевидно, все эти тензоры получаются один из другого преобразованием симметрии, переводящим друг в друга соответствующие минимумы; поэтому достаточно рассмотреть структуру лишь одного из них.

Естественно ожидать, что главные оси $E_{\alpha\beta}(j)$ будут совпадать с осями j-го эллипсоида. Соответственно, выбирая ось вдоль большой оси эллипсоида, можем написать (опуская для краткости индекс j)

$$E_{xx}=E_{yy}=E', \qquad E_{zz}=E''\neq E'. \qquad (\text{П.XI.11})$$

Подставляя в (П.XI.9′) вектор смещения $\mathbf{Q}$ в виде (XII.5.18) и приводя тензор $E_{\alpha\beta}(j)$ к главным осям, получаем под знаком суммы по j, $\mathbf{q}$ выражения

$$E'(q_x\zeta_x+q_y\zeta_y)+E''q_z\zeta_z. \qquad (\text{П.XI.12})$$

Рассмотрим отдельно случаи взаимодействия электронов с продольными и поперечными фононами.

Для продольных фононов по определению векторы $\mathbf{q}$ и $\boldsymbol{\zeta}$ параллельны, т. е.

$$\mathbf{q}=\frac{\boldsymbol{\zeta}}{|\boldsymbol{\zeta}|}|\mathbf{q}|. \qquad (\text{П.XI.13})$$

Соответственно выражение (П.XI.12) примет вид

$$q|\zeta|\left\{E'-(E'-E'')\frac{\zeta_z^2}{|\boldsymbol{\zeta}|^2}\right\}\equiv q|\zeta|\{E'-(E'-E'')\cos^2\vartheta\}, \qquad (\text{П.XI.14а})$$

где ϑ — полярный угол вектора $\boldsymbol{\zeta}$ в системе координат с полярной осью, параллельной главной оси рассматриваемого эллипсоида.

Для поперечных фононов мы имеем $(\mathbf{q}, \zeta)=0$ и выражение (П.XI.12) принимает вид

$$(E'-E'')(q_x\zeta_x+q_y\zeta_y)=(E'-E'')\,q\zeta\sin\theta\cdot\sin\alpha\cdot\cos(\varphi-\beta). \quad (\text{П.XI.14б})$$

Здесь θ и φ — полярные углы вектора ζ, α и β — полярные углы вектора $\mathbf{q}$ в прежней системе координат.

Видим, что в данном случае тензор потенциала деформации определяется двумя независимыми константами. В качестве таковых можно выбрать, например, $E'-E''=\Xi_t$ и $E'=\Xi_l$.

Подробную теорию потенциала деформации в применении к электронам проводимости в германии и кремнии можно найти в работе [2].

Особенно сложной становится теория потенциала деформации в случае вырожденных зон. Соответствующие результаты можно найти в книге [3].

ЛИТЕРАТУРА

1. *Л. Д. Ландау, Е. М. Лифшиц*, Механика, Физматгиз, М., 1958.
2. *C. Herring, E. Vogt*, Phys. Rev. **101**, 944 (1956) (см. русский перевод в сб. «Проблемы физики полупроводников», ИЛ, М., 1957).
3. *Г. Л. Бир, Г. Е. Пикус*, Симметрия и деформационные эффекты в полупроводниках, гл. V, «Наука», 1972.

Приложение XII. Потенциал заряженного центра при учете экранирования свободными носителями заряда

Как показано в § VI.6, для вычисления потенциала φ с учетом экранирования надо решить уравнение Пуассона, в правой части которого фигурирует заряд экранирующих носителей, сам зависящий от φ. В гл. VI соответствующий расчет был выполнен для невырожденного газа при условии, что концентрация носителей заряда и потенциал зависят только от одной координаты. Здесь мы рассмотрим трехмерный случай при произвольной степени вырождения. Пусть, для определенности, образец будет n-типа со средней концентрацией электронов, равной n_0. Концентрация электронов проводимости вблизи примесного иона дается формулой (V.4.4а). Допустим, что в области, нас интересующей, $e\varphi/kT \ll 1$ (это неравенство может и не иметь места; при этом расчет несколько усложняется). Тогда функцию $\Phi_{1/2}\left(\dfrac{\zeta+e\varphi}{kT}\right)$ в (V.4.4а) можно заменить первыми двумя членами разложения в ряд Тэйлора:

$$\Phi_{1/2}\left(\frac{\zeta+e\varphi}{kT}\right)\simeq\Phi_{1/2}\left(\frac{\zeta}{kT}\right)+\frac{e\varphi}{kT}\left.\frac{d\Phi_{1/2}(z)}{dz}\right|_{z=\zeta/T}.$$

Соответственно вместо (V.4.4а) мы получим

$$n(r)=n_0+N_c\frac{e\varphi}{kT}\Phi'_{1/2}\left(\frac{\zeta}{kT}\right), \quad (\text{П.XII.1})$$

где $\Phi'_{1/2}(z)=d\Phi(z)/dz$. Уравнение Пуассона примет вид

$$\nabla^2\varphi=\frac{4\pi e^2N_c}{\varepsilon kT}\Phi'_{1/2}\left(\frac{\zeta}{kT}\right)\varphi. \quad (\text{П.XII.2})$$

Введем обозначение

$$r_0^{-2}=\frac{4\pi e^2N_c}{\varepsilon kT}\Phi'_{1/2}\left(\frac{\zeta}{kT}\right) \quad (\text{П.XII.3})$$

и перепишем уравнение (П.XII.2) для потенциальной энергии электрона $\delta U = -e\varphi$:

$$\nabla^2 \delta U = r_0^{-2} \delta U. \qquad (П.XII.4)$$

Пусть примесный ион (с зарядом Ze) помещен в начале координат ($r = 0$). Тогда при $\mathbf{r} \to 0$ решение уравнения (П.XII.4) должно иметь кулоновский вид (IV.7.1). С другой стороны, при $\mathbf{r} \to \infty$ потенциальная энергия $\delta U(\mathbf{r})$ должна убывать по абсолютной величине. Естественно считать также, что функция $\delta U(\mathbf{r})$ будет сферически симметричной. Анизотропия изоэнергетических поверхностей, если она есть, в данном случае «смазывается» благодаря суммированию по всем минимумам, неявно выполненному в правой части (V.4.4а). Решение (П.XII.4), удовлетворяющее поставленным условиям, имеет вид

$$\delta U(\mathbf{r}) = \frac{Ze^2}{\varepsilon r} \exp\left(-\frac{r}{r_0}\right). \qquad (П.XII.5)$$

Величина r_0, определяемая равенством (П.XII.3), имеет размерность длины и называется *радиусом экранирования*. Как видно из формул (П.XII.1) и (П.XII.5), она определяет размеры области, в которой в основном находится экранирующий заряд. Видно, что при $r \gg r_0$ потенциальная энергия $\delta U(r)$ быстро убывает по абсолютной величине и система «заряженный центр + экранирующие носители» ведет себя как нейтральная.

Выражение (П.XII.3) заметно упрощается, если рассматривать невырожденный или, наоборот, полностью вырожденный газ. Как показано в §§ V.5, V.6, в первом случае мы имеем

$$\Phi_{1/2}\left(\frac{\zeta}{kT}\right) = \exp\left(\frac{\zeta}{kT}\right) \quad \text{и} \quad \exp\frac{\zeta}{kT} = \frac{n_0}{N_c},$$

а во втором случае

$$\Phi_{1/2}\left(\frac{\zeta}{kT}\right) = \frac{4}{3\sqrt{\pi}}\left(\frac{\zeta}{kT}\right)^{3/2} \quad \text{и} \quad \zeta = \frac{(3\pi^2 n_0)^{2/3}\hbar^2}{2m_d}.$$

Соответственно в невырожденном полупроводнике

$$r_0 = \left(\frac{\varepsilon kT}{4\pi n_0 e^2}\right)^{1/2}, \qquad (П.XII.6а)$$

а в условиях полного вырождения

$$r_0 = \frac{1}{2}\left(\frac{\pi}{3}\right)^{1/6}\left(\frac{\varepsilon\hbar^2}{m_d e^2} n^{-1/3}\right)^{1/2}. \qquad (П.XII.6б)$$

При наличии экранирующих носителей заряда обоих знаков формулы (П.XII.3) и (П.XII.6а, б) несколько изменяются. Выражение (П.XII.5), однако, остается в силе, коль скоро $\frac{|\delta U|}{kT} \ll 1$ или $\frac{|\delta U|}{\zeta} \ll 1$. Так же обстоит дело и при других механизмах экранирования (последнее может быть обусловлено, например, корреляцией в пространственном распределении примесных атомов, возникающей при введении их в решетку) *).

ЛИТЕРАТУРА

1. *Дж. Займан*, Принципы теории твердого тела, изд. 2. «Мир», М., 1974.

*) В металлах, где мы можем иметь сильно вырожденный электронный газ при сравнительно малой концентрации примеси, вид функции $\delta U(\mathbf{r})$ отличается от (П.XII.5) [1].

Приложение XIII. Усреднение по координатам примесных атомов

Вероятность перехода (XIV.5.3) зависят, очевидно, от $3N$ координат примесных атомов:

$$\mathscr{P}_1 = \mathscr{P}_2 = \mathscr{P} = \mathscr{P}(\mathbf{p},\ \mathbf{p}';\ \mathbf{R}_1,\ \ldots,\ \mathbf{R}_N). \qquad (\text{П.XIII.1})$$

Точный вид этой зависимости характеризует данный конкретный образец и обычно не представляет физического интереса. Используя совокупность большого числа кристаллов, мы получим среднее значение

$$\langle\mathscr{P}\rangle = \int \mathscr{P}(\mathbf{p},\ \mathbf{p}';\ \mathbf{R}_1,\ \ldots,\ \mathbf{R}_N)\, F(\mathbf{R}_1,\ \ldots,\ \mathbf{R}_N)\, d\mathbf{R}_1 \ldots d\mathbf{R}_N. \qquad (\text{П.XIII.2})$$

Здесь интегрирование по каждой из координат $\mathbf{R}_1$, ... производится в пределах основного объема (куба) V; $F(\mathbf{R}_1,\ \ldots,\ \mathbf{R}_N)\, d\mathbf{R}_1 \ldots d\mathbf{R}_N$ есть вероятность обнаружить 1-й, 2-й, ... атомы примеси, соответственно, в элементах объема $d\mathbf{R}_1$, $d\mathbf{R}_2$, ..., $d\mathbf{R}_N$ около точек $\mathbf{R}_1$, $\mathbf{R}_2$, ..., $\mathbf{R}_N$. Она удовлетворяет очевидному условию

$$\int F(\mathbf{R}_1,\ \mathbf{R}_2\ \ldots,\ \mathbf{R}_N)\, d\mathbf{R}_1\, d\mathbf{R}_2 \ldots d\mathbf{R}_N = 1. \qquad (\text{П.XIII.3})$$

Заметим, что роль «совокупности большого числа кристаллов» может играть и совокупность многих малых объемов, на которые можно разбить данный (достаточно большой) образец. Действительно, каждый из таких объемов можно охарактеризовать, вообще говоря, своей конфигурацией примеси; макроскопический же опыт относится ко всему кристаллу, и, следовательно, в нем автоматически производится усреднение, описываемое формулой (П.XIII.2) *).

Пока концентрация атомов примеси не слишком велика, вероятности попадания их в те или иные элементы объема можно считать независимыми:

$$F(\mathbf{R}_1,\ \ldots,\ \mathbf{R}_N) = f(\mathbf{R}_1) \ldots f(\mathbf{R}_N), \qquad (\text{П.XIII.4})$$

причем

$$\int f(\mathbf{R}_i)\, d\mathbf{R}_i = 1, \qquad i = 1, 2, \ldots, N. \qquad (\text{П.XIII.3}')$$

Строго говоря, представление вероятности $F(\mathbf{R}_1,\ \ldots,\ \mathbf{R}_N)$ в виде (П.XIII.4) носит приближенный характер. Так, здесь не исключена возможность одновременного попадания многих атомов примеси в один и тот же узел решетки, что физически невозможно. Если, однако, среднее расстояние между атомами примеси $N_t^{-1/3}$ заметно превышает постоянную решетки a, то такие события будут сравнительно редкими. Поэтому учет корреляции в расположении примесных атомов, исключающий названную «нефизическую возможность», практически несуществен и аппроксимация (П.XIII.4) достаточна для вычисления кинетических коэффициентов. Более того, в макроскопически однородной системе примесный атом с одинаковой вероятностью может попасть в любую точку кристалла, т. е. функция f фактически есть константа. Согласно условию нормировки (П.XIII.3')

$$f = V^{-1}. \qquad (\text{П.XIII.5})$$

Итак, при не слишком большой концентрации примеси $\left(N_t^{-1/3} \gg a\right)$ в макроскопически однородной системе

$$\langle\mathscr{P}\rangle = \frac{1}{V^N}\int \mathscr{P}(\mathbf{p},\ \mathbf{p}';\ \mathbf{R}_1,\ \ldots,\ \mathbf{R}_N)\, d\mathbf{R}_1 \ldots d\mathbf{R}_N. \qquad (\text{П.XIII.6})$$

---

*) Речь идет о кристалле, макроскопически однородном, и о макроскопически однородных условиях опыта. В макроскопически неоднородной системе (например, в кристалле с n—p-переходом) интегрирование в (П.XIII.2) удобно производить особо для каждой однородной области.

Подставляя сюда выражение (XIV.5.3), получаем

$$\langle \mathscr{P} \rangle = \frac{2\pi}{\hbar V} \delta (E(\mathbf{p}) - E(\mathbf{p}')) (\langle A_1 \rangle + \langle A_2 \rangle). \qquad (\text{П.XIII.6}')$$

Поскольку A_1 не зависит от координат примесных атомов,

$$\langle A_1 \rangle = \frac{1}{V^N} \int d\mathbf{R}_1 \ldots d\mathbf{R}_N A_1 = A_1. \qquad (\text{П.XIII.7})$$

Далее, согласно (XIV.5.8)

$$\langle A_2 \rangle =$$

$$= \left| \int \delta U(\mathbf{r}') e^{\frac{i}{\hbar}(\mathbf{p}-\mathbf{p}', \mathbf{r}')} d\mathbf{r}' \right|^2 \sum_{\substack{i, j \\ (i \neq j)}} \frac{1}{V^N} \int \cos \left(\frac{\mathbf{p}-\mathbf{p}'}{\hbar}, \mathbf{R}_i - \mathbf{R}_j \right) d\mathbf{R}_1 \ldots d\mathbf{R}_N.$$

Очевидно, все члены суммы по i, j здесь одинаковы, ибо отличаются друг от друга только заменой переменных интегрирования. Пусть $i=1$, $j=2$. Тогда интегралы по $\mathbf{R}_3, \ldots, \mathbf{R}_N$ сразу берутся, и мы получаем, принимая во внимание, что полное число членов в сумме равно $N(N-1)$,

$$\langle A_2 \rangle = \frac{N(N-1)}{V^2} \left| \int \delta U(\mathbf{r}') \exp \left[\frac{i}{\hbar} (\mathbf{p}-\mathbf{p}', \mathbf{r}') \right] d\mathbf{r}' \right|^2 \times$$

$$\times \int d\mathbf{R}_1 \, d\mathbf{R}_2 \cos \left(\frac{\mathbf{p}-\mathbf{p}'}{\hbar}, \mathbf{R}_1 - \mathbf{R}_2 \right). \qquad (\text{П.XIII.8})$$

Введем вместо $\mathbf{R}_1$ и $\mathbf{R}_2$ новые переменные интегрирования

$$\mathbf{R} = \frac{1}{2}(\mathbf{R}_1 + \mathbf{R}_2), \qquad \mathbf{R}' = \mathbf{R}_1 - \mathbf{R}_2.$$

Тогда фигурирующей в (П.XIII.8) интеграл по $\mathbf{R}_1$ и $\mathbf{R}_2$ примет вид

$$\int \cos \left(\frac{\mathbf{p}-\mathbf{p}'}{\hbar}, \mathbf{R}' \right) d\mathbf{R} \, d\mathbf{R}' = V \prod_{\alpha = x, y, z} \frac{2\hbar}{(p_\alpha - p'_\alpha)} \sin \left(\frac{p_\alpha - p'_\alpha}{2\hbar} L \right), \qquad (\text{П.XIII.9})$$

где L есть ребро куба.

При неограниченном возрастании размеров системы правая часть (П.XIII.9) стремится к произведению δ-функций

$$V \prod_{\alpha} \pi \delta \left(\frac{p_\alpha - p'_\alpha}{2\hbar} \right).$$

Таким образом, при $N \to \infty$ величина $\langle A_2 \rangle$ отлична от нуля лишь при $\mathbf{p}' = \mathbf{p}$, т. е. когда рассеяние отсутствует. Иначе говоря, вклад в рассеяние от этого слагаемого — порядка $1/V$, и для макроскопической системы им можно пренебречь. Окончательно:

$$\langle \mathscr{P} \rangle = \frac{2\pi}{\hbar V} A_1 \delta (E(\mathbf{p}) - E(\mathbf{p}')). \qquad (\text{П.XIII.10})$$

В основном тексте книги мы опускаем угловые скобки, понимая под всеми величинами типа (П.XIII.1) только усредненные их значения.

Вычислим среднее значение квадрата флуктуации потенциальной энергии электрона в поле одинаковых хаотически расположенных ионов примеси. Сог-

ласно (XIV.5.1) эта энергия дается выражением

$$H' = \sum_{i=1}^{N} \delta U(\mathbf{r} - \mathbf{R}_i). \qquad (\text{П.XIII.11})$$

Представим правую часть (П.XIII.11) в виде

$$\langle H' \rangle + \delta H',$$

где $\delta H'$ — интересующая нас флуктуация энергии, а $\langle H' \rangle$ — среднее (по координатам примесных ионов) значение H':

$$\langle H' \rangle = \frac{1}{V^N} \int H'\, d\mathbf{R}_1 \ldots d\mathbf{R}_N. \qquad (\text{П.XIII.12})$$

В силу (П.XIII.11) в правой части (П.XIII.12) фигурирует сумма N одинаковых слагаемых. Следовательно, это соотношение можно переписать в виде

$$\langle H' \rangle = N_t \int \delta U(\mathbf{r} - \mathbf{R})\, d\mathbf{R}. \qquad (\text{П.XIII.13})$$

Интеграл в правой части (П.XIII.13) есть некоторая постоянная, зависящая от явного вида экранированного потенциала δU. Обозначив ее через u_0, имеем *)

$$\delta H' = \sum_{i=1}^{N} \delta U(\mathbf{r} - \mathbf{R}_i) - N_t u_0. \qquad (\text{П.XIII.14})$$

Таким образом,

$$\langle (\delta H')^2 \rangle \equiv \psi_1 = V^{-N} \int \left\{ \sum_{i=1}^{N} \sum_{j=1}^{N} \delta U(\mathbf{r} - \mathbf{R}_i)\, \delta U(\mathbf{r} - \mathbf{R}_j) - \right.$$
$$\left. - 2N_t u_0 \sum_{i=1}^{N} \delta U(\mathbf{r} - \mathbf{R}_i) + N_t^2 u_0^2 \right\} d\mathbf{R}_1 \ldots d\mathbf{R}_N. \qquad (\text{П.XIII.15})$$

Очевидно, второе и третье слагаемые в правой части (П.XIII.15) в сумме составляют

$$-2N_t^2 u_0^2 + N_t^2 u_0^2 = -N_t^2 u_0^2. \qquad (\text{П.XIII.16})$$

В двойной сумме по i, j удобно выделить слагаемые с $i = j$. Тогда первый член в правой части (П.XIII.15) можно переписать в виде

$$V^{-N} \int \left\{ \sum_{i=1}^{N} [\delta U(\mathbf{r} - \mathbf{R}_i)]^2 + \sum_{\substack{i,\, j=1 \\ i \neq j}}^{N} \delta U(\mathbf{r} - \mathbf{R}_i)\, \delta U(\mathbf{r} - \mathbf{R}_j) \right\} d\mathbf{R}_1 \ldots d\mathbf{R}_N =$$
$$= N_t \int [\delta U(\mathbf{r} - \mathbf{R})]^2\, d\mathbf{R} + \frac{N(N-1)}{V^2} u_0^2. \qquad (\text{П.XIII.17})$$

*) Фактически в макроскопически однородном образце второе слагаемое в (П.XIII.14) следует отбросить в силу условия полной нейтральности материала. Действительно, заряд примесных ионов точно компенсируется свободными электронами и дырками (либо — в условиях компенсации — ионами другого знака). Следовательно, к выражению (П.XIII.11) надо добавить еще постоянное слагаемое, описывающее энергию взаимодействия электрона с размазанным в пространстве зарядом всех других электронов и дырок. Можно доказать, что это слагаемое равно $-N_t u_0$. При вычислении интересующей нас величины $\langle (\delta H')^2 \rangle$ это, однако, не играет роли.

Асимптотически при $N \to \infty$, $V \to \infty$, $N_t = \frac{N}{V} < \infty$ правая часть (П.XIII.17) принимает вид

$$N_t \int [\delta U(\mathbf{r} - \mathbf{R})]^2 \, d\mathbf{R} + N_t^2 u_0^2. \qquad (\text{П.XIII.17}')$$

Очевидно, интеграл по $\mathbf{R}$ здесь не зависит от $\mathbf{r}$: можно ввести новую переменную $\mathbf{r}' = \mathbf{r} - \mathbf{R}$. Складывая выражения (П.XIII.16) и (П.XIII.17'), мы получаем

$$\psi_1 = N_t \int [\delta U(\mathbf{r}')]^2 \, d\mathbf{r}'. \qquad (\text{П.XIII.18})$$

Пусть, например, потенциальная энергия электрона в поле одного заряженного центра описывается формулой (П.XII.5)

$$\delta U(\mathbf{r}') = \frac{Ze^2}{\varepsilon r'} \exp\left(-\frac{r'}{r_0}\right). \qquad (\text{П.XIII.19})$$

Подставляя это выражение в (П.XIII.18), находим

$$\psi_1 = \frac{2\pi Z^2 e^4 N_t r_0}{\varepsilon^2}. \qquad (\text{П.XIII.20})$$

Приложение XIV. Теорема об интеграле от периодической функции

Рассмотрим выражение

$$\mathcal{J} = \int e^{i\mathbf{f}\mathbf{r}} \Phi(\mathbf{r}) \, d\mathbf{r}, \qquad (\text{П.XIV.1})$$

где $\Phi(\mathbf{r})$ — функция, периодическая с периодом решетки $\mathbf{a}$, $\mathbf{f}$ — некоторый вектор, а интеграл берется по фундаментальному объему. Компоненты $\mathbf{f}$ подчиняются условиям периодичности вида (III. 3.10).

Произведем в (П.XIV.1) замену переменных, полагая

$$\mathbf{r} = \mathbf{r}' + \mathbf{a}.$$

Получим, с учетом свойства периодичности функции Φ,

$$\mathcal{J} = \exp(i\mathbf{f}\mathbf{a}) \int e^{i\mathbf{f}\mathbf{r}'} \Phi(\mathbf{r}' + \mathbf{a}) \, d\mathbf{r}' = \exp(i\mathbf{f}\mathbf{a}) \int e^{i\mathbf{f}\mathbf{r}'} \Phi(\mathbf{r}') \, d\mathbf{r}'. \qquad (\text{П.XIV.2})$$

В правой части (П.XIV.2) стоит прежний интеграл $\mathcal{J}$; таким образом,

$$\mathcal{J} = \mathcal{J} \exp(i\mathbf{f}\mathbf{a}).$$

Отсюда явствует, что интеграл $\mathcal{J}$ может быть отличен от нуля, лишь если $\exp(i\mathbf{f}\mathbf{a}) = 1$, т. е.

$$\mathbf{f} = \mathbf{b}m, \qquad (\text{П.XIV.3})$$

где $\mathbf{b}$ — вектор обратной решетки, а m — целое число или нуль.

Приложение XV. Интегралы с функцией Ферми в условиях сильного вырождения

Рассмотрим интеграл

$$\mathcal{J} = \int_0^\infty \varphi(E) f_0'(E) \, dE. \qquad (\text{П.XV.1})$$

где $\varphi(E)$ — некоторая гладкая функция энергии, а $f_0' = \frac{df_0}{dE}$. В условиях сильного вырождения функция $f_0'(E) \neq 0$ лишь при $E \simeq \zeta$ и в интеграле

(П.XV.1) существенны только значения E, близкие к ζ. По этой причине удобно ввести новую переменную интегрирования

$$\eta=\frac{E-\zeta}{kT}$$

и разложить выражение $\varphi(E)\equiv\varphi(\zeta+\eta kT)$ по степеням η.

Замечая, что

$$f_0'=\frac{1}{kT}\frac{df_0}{d\eta}=-\frac{1}{kT}(e^{\eta}+1)^{-1}(e^{-\eta}+1)^{-1}, \qquad (\text{П.XV.2})$$

мы получаем

$$\mathcal{J}=\mathcal{J}_1\varphi(\zeta)+kT\varphi'(\zeta)\,\mathcal{J}_2+\frac{1}{2}(kT)^2\varphi''(\zeta)\,\mathcal{J}_3+\ldots \qquad (\text{П.XV.3})$$

Здесь

$$\mathcal{J}_n=\int\limits_{-\zeta/kT}^{\infty}\eta^{n-1}f_0'(\eta)\,d\eta, \qquad n=1,2,3. \qquad (\text{П.XV.4})$$

По условию сильного вырождения нижний предел в интегралах $\mathcal{J}_n$ отрицателен и велик по абсолютной величине. Как видно из формулы (П.XV.2), при больших значениях $|\eta|$ функция $f_0'(\eta)$ экспоненциально мала. Пренебрегая величинами порядка $\exp(-\zeta/kT)$, мы можем распространить область интегрирования до $-\infty$. Тогда

$$\mathcal{J}_1\simeq\int\limits_{-\infty}^{+\infty}f'(\eta)\,d\eta)=-1, \qquad \mathcal{J}_2\simeq\int\limits_{-\infty}^{+\infty}\eta f'(\eta)\,d\eta=0,$$

так как f_0' есть четная функция η, и, наконец,

$$\mathcal{J}_3\simeq-\int\limits_{-\infty}^{+\infty}\frac{\eta^2\,d\eta}{(1+e^{-\eta})(1+e^{\eta})}=-2\int\limits_{0}^{\infty}\frac{\eta^2e^{-\eta}\,d\eta}{(1+e^{-\eta})^2}=$$

$$=-2\sum_{k=0}^{\infty}(-1)^k(k+1)\int\limits_{0}^{\infty}\eta^2e^{-(k+1)\eta}\,d\eta=-4\sum_{k=0}^{\infty}\frac{(-1)^k}{(k+1)^2}.$$

Сумма последнего ряда составляет $\pi^2/12$ [1]. Таким образом,

$$\mathcal{J}_3=-\pi^2/3$$

и

$$\mathcal{J}=-\varphi(\zeta)-\frac{\pi^2}{6}(kT^2)\varphi''(\zeta)+\ldots \qquad (\text{П.XV.5})$$

В пренебрежении величиной порядка $(kT/F)^2$ формула (П.XV.5) дает

$$\mathcal{J}=-\varphi(\zeta),$$

что и оправдывает выражение (XIII.7.17).

ЛИТЕРАТУРА

1. *И. М. Рыжик. И. С. Градштейн,* Таблицы интегралов, сумм, рядов и произведений, изд. 3, переработанное, стр. 21, Гостехиздат, М., 1951.

ЛИТЕРАТУРА

ОСНОВНАЯ ЛИТЕРАТУРА

М1. *Д. И. Блохинцев,* Основы квантовой механики, «Наука», М., 1976.
М2. *А. С. Давыдов,* Квантовая механика, изд. 2, «Наука», М., 1973.
М3. *А. А. Соколов, Ю. М. Лоскутов, И. М. Тернов,* Квантовая механика, изд. 2, «Просвещение», М., 1965.
М4. *Г. С. Жданов,* Физика твердого тела, изд-во МГУ, М., 1961.
М5. *Н. Е. Кочин,* Векторное исчисление и начала тензорного исчисления, изд. 9, «Наука», М., 1965.
М6. *Г. Джонс,* Теория зон Бриллюэна и электронные состояния в кристаллах, «Мир», М., 1968.
М7. *А. И. Ансельм,* Введение в теорию полупроводников, Физматгиз, М.—Л., 1962.
М8. *И. М. Цидильковский,* Электроны и дырки в полупроводниках, «Наука», М., 1972.
М9. *Л. Д. Ландау, Е. М. Лифшиц,* Статистическая физика (классическая и квантовая), изд. 3, «Наука», М., 1976.
М10. *А. И. Ансельм,* Основы статистической физики и термодинамики, «Наука», М., 1973.
М11. *Л. Д. Ландау, Е. М. Лифшиц,* Механика сплошных сред, изд. 2, Физматгиз, М., 1963.
М12. *И. Е. Тамм,* Основы теории электричества, изд. 9, «Наука», М., 1976.
М13. *Л. Д. Ландау, Е. М. Лифшиц,* Электродинамика сплошных сред, Физматгиз, М., 1959.

К главе II

1. *Ч. Коулсон,* Валентность, «Мир», М., 1965.
2. *Н. Мотт, Р. Герни,* Электронные процессы в ионных кристаллах, ИЛ, М., 1950.
3. *А. Ф. Иоффе, А. Р. Регель,* Некристаллические, аморфные и жидкие электронные полупроводники, Избранные труды А. Ф. Иоффе, т. 2, «Наука», М., 1975.
4. *Н. Мотт, Э. Дэвис,* Электронные процессы в некристаллических веществах, «Мир», М., 1974.
5. *В. С. Вавилов,* Действие излучений на полупроводники, Физматгиз, М., 1963.
6. *В. Т. Рид,* Дислокации в кристаллах, ИЛ, М., 1957.
7. *У. Бардсли,* Влияние дислокаций на электрические свойства полупроводников, УФН **73**, 121 (1961).

К главе III

1. *В. И. Смирнов,* Курс высшей математики, изд. 18, т. II, Физматгиз, М., 1961.
2. *Д. Гильберт, Р. Курант,* Методы математической физики, т. I, ГТТИ, М. 1933.

3. *Дж. Займан,* Вычисление блоховских функций, «Мир», М., 1973.
4. *О. Маделунг,* Физика полупроводниковых соединений элементов III и V групп, «Мир», М., 1967.

К главе IV

1. *Дж. Займан,* Принципы теории твердого тела, изд. 2, «Мир», М., 1974.
2. *С. И. Пекар,* Исследования по электронной теории ионных кристаллов, Гостехиздат, М., 1951.
3. «Туннельные явления в твердых телах» (сб. под ред. Э. Бурштейна и С. Лундквиста. Пер. с англ. под ред. В. И. Переля), гл. 6, «Мир», М., 1973.
4. *К. Киттель, К. Митчелл* (в сб. «Проблемы физики полупроводников»), ИЛ, М., 1957.
5. *Дж. М. Латтинджер, В. Кон,* (в сб. «Проблемы физики полупроводников»), ИЛ, М., 1957.
6. *А. И. Ансельм, Л. И. Коровин,* ЖТФ **24**, 8044 (1955).

К главе V

1. *Дж. Блекмор,* Статистика электронов в полупроводниках, «Мир», М., 1964.

К главе VI

1. *Н. Мотт, Р. Герни,* Электронные процессы в ионных кристаллах, ИЛ, М., 1950.
2. *Г. Е. Пикус,* Основы теории полупроводниковых приборов, «Наука», М.—Л., 1965.
3. *А. Роуз,* Основы теории фотопроводимости, «Мир», М., 1966.
4. *М. Ламперт, П. Марк,* Инжекционные токи в твердых телах, «Мир», М., 1973.
5. *А. Милнс, Д. Фойхт,* Гетеропереходы и переходы металл — полупроводник, «Мир», М., 1975.

К главе VII

1. *С. М. Рывкин,* Фотоэлектрические явления в полупроводниках, Физматгиз, М., 1963.
2. *В. С. Вавилов,* Действие излучений на полупроводники, Физматгиз, М., 1963.
3. *W. Van Roosbroeck,* Phys. Rev. **91**, № 2, 282 (1953).
4. Статьи №№ 21—26 в сб. «Проблемы физики полупроводников», ИЛ, М., 1957.

К главе VIII

1. *Г. Е. Пикус,* Основы теории полупроводниковых приборов, «Наука», М.—Л., 1965.
2. *Я. А. Федотов,* Основы физики полупроводниковых приборов, «Советское радио», М., 1969.
3. *А. Милнс, Д. Фойхт,* Гетеропереходы и переходы металл — полупроводник, «Мир», М., 1975.

К главе IX

1. *С. М. Рывкин,* Рекомбинация в полупроводниках. Полупроводники в науке и технике, т. II, Изд-во АН СССР, М., 1958.
2. *Дж. С. Блэкмор,* Статистика электронов в полупроводниках, «Мир», М., 1964.
3. *W. van Roosbroeck, W. Shockley,* Phys Rev. **94**, 1558 (1954).

К главе X

1. *А. В. Ржанов,* Электронные процессы на поверхности полупроводников, «Наука», М., 1971.
2. Электронные явления на поверхности полупроводников, Сб. под ред. В. И. Ляшенко, «Наукова думка», Киев, 1968.
3. *Ф. Ф. Волькенштейн,* Физико-химия поверхности полупроводников, «Наука», М., 1973.
4. *В. Ф. Киселев,* Поверхностные явления в полупроводниках и диэлектриках, «Наука», М., 1970.
5. *С. Дэвисон, Дж. Левин,* Поверхностные (таммовские) состояния, «Мир», М., 1973.

К главе XI

1. *С. М. Рывкин,* Фотоэлектрические явления в полупроводниках, Физматгиз, М., 1963.
2. *Я. Тауц,* Фото- и термоэлектрические явления в полупроводниках, ИЛ, М., 1962.
3. *Г. Е. Пикус,* Основы теории полупроводниковых приборов, «Наука», М.—Л., 1965.
4. *А. М. Васильев, А. П. Ландсман,* Полупроводниковые фотопреобразователи, «Советское радио», М., 1971.
5. *W. van Roosbroeck,* Phys. Rev. **101**, № 6, 1713 (1956).

К главе XII

1. *Макс Борн, Хуан Кунь,* Динамическая теория кристаллических решеток, ИЛ, М., 1958.
2. *Ч. Киттель,* Квантовая теория твердых тел, «Наука», М., 1967.
3. *И. Е. Тамм,* Zs. f. Phys. **60**, 345 (1930).

К главе XIII

1. «Вопросы квантовой теории необратимых процессов» (сб. статей, перевод с англ.), ИЛ, М., 1961.
2. *Б. М. Аскеров,* Кинетические эффекты в полупроводниках, «Наука», Л., 1970.

К главе XIV

1. «Вопросы квантовой теории необратимых процессов» (сб. статей, перевод с англ.), ИЛ, М, 1961.
2. *Г. А. Бир, Г. Е. Пикус,* Симметрия и деформационные эффекты в полупроводниках, «Наука», М., 1972.
3. *Ю. И. Сиротин, М. П. Шаскольская,* Основы кристаллофизики, «Наука», М., 1975.
4. *Ф. Блатт,* Теория подвижности электронов в твердых телах, ИЛ., М., 1963, гл. V.

К главе XV

1. *В. Л. Гуревич,* Физика и техника полупроводников **2**, № 11, 1557 (1968).
2. *В. И. Пустовойт,* УФН **97**, № 2, 257 (1969).
3. *A. R. Hutson, J. H. McFee, D. L. White,* Phys. Rev. Lett. **7**, 237 (1961).
4. *D. L. White,* Journ. Appl. Phys. **33**, 2547 (1962).
5. *A. B. Pippard,* Phil. Mag. **8**, № 85, 161 (1963).

К главе XVI

1. *Ф. Г. Басс, Ю. Я. Гуревич,* Горячие электроны и сильные электромагнитные волны в плазме полупроводников и газового разряда, «Наука», М., 1975.

2. *Э. Конуэлл,* Кинетические свойства полупроводников в сильных электрических полях, «Мир», М., 1970.
3. *В. Денис, Ю. Пожела,* Горячие электроны, «Минтис», Вильнюс, 1971.
4. *В. Л. Бонч-Бруевич, И. П. Звягин, А. Г. Миронов,* Доменная электрическая неустойчивость в полупроводниках, «Наука», М., 1972.

К главе XVII

1. *Макс Борн, Хуан Кунь,* Динамическая теория кристаллических решеток, ИЛ, М., 1958, § 14.
2. *С. И. Пекар,* Исследования по электронной теории ионных кристаллов, Гостехиздат, М., 1951.
3. *Д. Пайнс,* Элементарные возбуждения в твердых телах, «Мир», М., 1965.
4. *Дж. Займан,* Вычисление функций Блоха, «Мир», М., 1973.
5. *Р. Нокс,* Теория экситонов, «Мир», М., 1966.
6. *А. С. Давыдов,* Теория молекулярных экситонов, «Наука», М., 1968.
7. «Экситоны в полупроводниках» (сб.), «Наука», М., 1972.
8. *Ю. Е. Перлин,* УФН **80**, вып. 4, 553 (1963).

К главе XVIII

1. *Г. С. Ландсберг,* Оптика, «Наука», М., 1976.
2. *Г. Фэн,* Фотон-электронное взаимодействие в кристаллах, «Мир», М., 1969.
3. *В. С. Вавилов,* Действие излучений на полупроводники, Физматгиз, М., 1963.
4. *О. В. Богданкевич, С. А. Даржнек, П. Г. Елисеев,* Полупроводниковые лазеры, «Наука», М., 1976.
5. Квантовая электроника (отв. редактор М. Е. Жаботинский), «Советская энциклопедия», М., 1969.
6. *Е. Джонсон,* Поглощение вблизи края фундаментальной полосы (в кн. «Оптические свойства полупроводников (полупроводниковые соединения типа $A^{III}B^{V}$)», под ред. Р. Уиллардсона и А. Бира), гл. 6, «Мир», М., 1970.
7. *М. Кардона,* Модуляционная спектроскопия, «Мир», М., 1972.
8. *Е. Патлик, Дж. Райт,* Магнетоплазменные эффекты (в кн. «Оптические свойства полупроводников (полупроводниковые соединения типа $A^{III}B^{V}$)», под ред. Р. Уиллардсона и А. Бира), гл. 10, «Мир», М., 1970.

К главе XIX

1. *И. М. Лифшиц,* УФН **83**, 617 (1964); ЖЭТФ **53**, 743 (1967).
2. *В. Л. Бонч-Бруевич,* Вопросы электронной теории сильно легированных полупроводников (в кн. «Итоги науки. Физика твердого тела» под ред. С. В. Тябликова, ВИНИТИ, М., 1965).
3. *В. И. Фистуль,* Сильно легированные полупроводники, «Наука», М., 1967.
4. *М. А. Афромович, Д. Редфилд,* Труды IX Межд. конф. по физике полупроводников, «Наука», Л., 1969, т. I, стр. 103.
5. *В. Л. Бонч-Бруевич,* Квазиклассическая теория движения частиц в случайном поле (в кн. «Статистическая физика и квантовая теория поля» под ред. Н. Н. Боголюбова, «Наука», М., 1973).
6. *Н. Мотт, Э. Дэвис,* Электронные процессы в некристаллических веществах, «Мир», М., 1974.
7. *Н. Мотт,* Электроны в неупорядоченных структурах, «Мир», М., 1969.
8. *В. А. Алексеев, А. А. Андреев, В. Я. Прохоренко,* Электрические свойства жидких металлов и полупроводников, УФН **106**, вып. 3, 393 (1972).

ОСНОВНЫЕ ОБОЗНАЧЕНИЯ

a — постоянная решетки (в кубическом кристалле); совокупность номера атома в элементарной ячейке и координат центра тяжести ячейки.

a_B — боровский радиус в кристалле.

$\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ — основные векторы решетки.

$\mathfrak{B}$ — вектор магнитной индукции.

$\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ — основные векторы обратной решетки.

c — скорость света в вакууме.

c_s — фазовая скорость звуковых волн.

$\mathfrak{D}$ — вектор электрической индукции.

D — коэффициент амбиполярной диффузии.

$E(\mathbf{p}, l)$ — энергия носителя заряда, зависящая от квазиимпульса $\mathbf{p}$ и номера зоны l.

E_c — энергия дна зоны проводимости.

E_v — энергия потолка валентной зоны.

E_g — ширина запрещенной зоны.

E — собственные значения энергии электрона.

E_B — боровская энергия в кристалле.

$\mathfrak{E}$ — вектор напряженности электрического поля.

e — абсолютная величина заряда электрона.

$F = \zeta - e\varphi$ — уровень Ферми (электрохимический потенциал).

$F_{n\,(p)}$ — квазиуровень Ферми для электронов (дырок).

f — функция распределения электронов; совокупность номера ветви и квазиволнового вектора фонона.

f_0 — функция Ферми.

$g_{n(p)}$ — темп внешней генерации свободных электронов (дырок).

H — оператор энергии (гамильтониан).

$\hbar$ — постоянная Планка, деленная на 2π.

$\mathbf{I}$ — плотность потока энергии.

I — интенсивность света.

i — сила тока.

$\mathbf{j}$ — вектор плотности конвекционного тока.

$\mathfrak{H}$ — вектор напряженности магнитного поля.

$\mathbf{k}$ — квазиволновой вектор электрона.

k — постоянная Больцмана; координационное число.

l — длина свободного пробега; номер энергетической зоны.

L_D — длина экранирования Дебая.

$L_э$ — длина экранирования в общем случае; толщина слоя объемного заряда в контакте или у поверхности полупроводника.

$L = \sqrt{D\tau}$ — длина диффузии.

$l(\mathfrak{E})$ — длина затягивания неравновесных носителей заряда в электрическом поле.

m_0 — масса изолированного электрона.

$m_{n\,(p)}$ — скалярная эффективная масса электрона (дырки).

$m^{-1}_{\alpha\beta}$ — компоненты тензора обратной эффективной массы.

$m_{||}$ — эффективная масса для движения вдоль оси вращения эллипсоида энергии.

$m_\perp$ — эффективная масса для движения перпендикулярно оси вращения эллипсоида энергии.

m_{nd}, m_{pd} — эффективная масса плотности состояний для электронов, дырок.

$m_\sigma = m_{\text{opt}}$ — эффективная масса электропроводности, оптическая эффективная масса.

M_h — масса h-го атома решетки.

M — масса элементарной ячейки.

n — концентрация свободных электронов; совокупность чисел n_f со всеми значениями f.

N_f — концентрация примеси (любой).

N_d — концентрация доноров.

n_t — концентрация связанных электронов на акцепторных уровнях.

n_1 — характерная концентрация электронов.

n_f — число фононов типа f (квантовое число гармонического осциллятора, соответствующего нормальному колебанию типа f).

$N_{c\,(v)}$ — эффективная плотность состояний в зоне проводимости (в валентной зоне).

$N_{c\,(v)}(E)$ — плотность состояний в зоне проводимости (в валентной зоне).

$\mathcal{P}$ — вектор поляризации.

$\mathbf{p}$ — квазиимпульс электрона.

p — концентрация свободных дырок.

p_t — концентрация связанных дырок на донорных уровнях.

p_1 — характерная концентрация дырок.

Q — количество тепла.

$\mathbf{Q}$ — вектор смещения.

$\mathbf{q}$ — волновой вектор фонона.

R — постоянная Холла.

$\mathbf{R}_\alpha$ — тройка координат α-го атома решетки.

$\mathbf{r}$ — тройка координат электрона.

r_n — темп захвата электронов из зоны на локальные уровни.

r_p — темп захвата дырок из зоны на локальные уровни.

$R_n = r_n - g_{nT}$ — суммарный темп рекомбинации электронов.

$R_p = r_p - g_{pT}$ — суммарный темп рекомбинации дырок.

r — число атомов в элементарной ячейке; показатель степени в зависимости времени релаксации от энергии.

r_0 — радиус экранирования.

S — энтропия; поверхность; безразмерная скорость поверхностной рекомбинации.

$s_{\alpha\beta}$ — компоненты тензора механического напряжения.

S_n, S_p — эффективное сечение захвата электронов, дырок на локальные уровни.

s_n, s_p — скорость поверхностной рекомбинации электронов, дырок.

T_D — температура Дебая.

T_e — электронная температура.

u — электрическое напряжение; деформация.

$u_{\mu\nu}$ — компоненты тензора деформации.

$U(\mathbf{r})$ — потенциальная энергия электрона в идеальной кристаллической решетке двух атомов.

V — потенциальная энергия колеблющейся решетки; фундаментальный (основной) объем.

V_0 — объем элементарной ячейки; электродвижущая сила.

$\mathbf{v}_d$ — дрейфовая скорость носителя заряда.

$\mathbf{v}(\mathbf{p}, l)$ — скорость носителя заряда с квазиимпульсом $\mathbf{p}$ в l-той зоне.

v_T — тепловая скорость.

v_s — фазовая скорость звуковых волн.

$Y_s = e\varphi_s/kT$ — безразмерный поверхностный потенциал.

α — дифференциальная термоэдс; коэффициент рекомбинации зона — зона; коэффициент усиления тока в биполярном диффузионном транзисторе в схеме с общей базой.

α_n, α_p — коэффициент захвата электрона, дырки на локальные уровни.

β — магнетон Бора; коэффициент передачи неосновных носителей в транзисторах и фотоэлементах; коэффициент в формуле зависимости подвижности от напряженности поля; пьезоэлектрический модуль.

$\beta_{\alpha,\mu\nu}$ — компоненты тензора пьезоэлектрических модулей.

γ — коэффициент поглощения света или звука; холл-фактор; магнитная длина.

Δ — энергия спин-орбитального расщепления валентной зоны в центре зоны Бриллюэна.

ΔG — поверхностная проводимость.

$\varepsilon = \varepsilon_1 + i\varepsilon_2$ — диэлектрическая проницаемость решетки (ε_1, ε_2 — вещественные числа).

$\zeta = F - E_c$ (r) — химический потенциал для электронов.

$\zeta^* = \zeta/kT$ — безразмерный химический потенциал для электронов.

$\boldsymbol{\xi}_h$ (q, s) — вектор, указывающий направление смещения h-го атома решетки при нормальном колебании ветви s с квазиволновым вектором $\mathbf{q}$.

$\eta = E_v$ (r) $- F$ — химический потенциал для дырок; кпд; степень неупругости рассеяния.

$\eta^* = \eta/kT$ — безразмерный химический потенциал для дырок.

λ — длина волны де Бройля; совокупность квантовых чисел, характеризующих стационарные состояния носителей заряда.

$\Lambda_{\alpha\beta,\mu\nu}$ — компоненты тензора модулей упругости.

$\mu_{n\,(p)}$ — дрейфовая подвижность электронов (дырок).

ν (E) — плотность поверхностных уровней на единичный интервал энергии.

ν (ω) — квантовый выход внутреннего фотоэффекта.

ρ — удельное сопротивление; объемная плотность электрического заряда; плотность кристалла.

ρ ($\hbar\omega$) — среднее число фононов в единице объема и в единичном интервале энергии, содержащем энергию $\hbar\omega$.

σ — удельная электропроводность.

σ (ω) $= \sigma_1 + i\sigma_2$ — комплексная электропроводность на частоте ω (σ_1, σ_2 — вещественные числа).

$\boldsymbol{\sigma}$ — спиновый вектор Паули; компоненты его суть матрицы Паули σ_x, σ_y, σ_z.

τ (E) — время релаксации импульса.

$\tau_M = \varepsilon/4\pi\sigma$ — максвелловское время релаксации.

$\tau_{\mathbf{p}} = \langle\tau\,(E)\rangle$ — среднее время релаксации импульса (время свободного пробега).

τ_e — среднее время релаксации энергии.

τ — время жизни электронно-дырочных пар.

$\tau_{n\,(p)}$ — среднее время жизни избыточных (неравновесных) электронов (дырок).

τ_s — время жизни неравновесных носителей заряда, обусловленное поверхностной рекомбинацией.

φ — электрический потенциал; угол Холла.

Φ_n — интеграл Ферми — Дирака с индексом n.

Φ — термоэлектронная работа выхода; волновая функция колеблющейся решетки.

ψ (r) — одноэлектронная волновая функция.

χ (r) — сглаженная волновая функция в методе эффективной массы.

χ — электронное сродство полупроводника.

$\Xi_{l\,(t)}$ — константа акустического потенциала деформации для продольных (поперечных) фононов.

$\Xi_{\text{опт}}$ — константа оптического потенциала деформации.

ω — круговая частота электромагнитной или звуковой волны.

ω_c — циклотронная частота.

ω_0 — предельная частота оптического фонона.