621.38
К84

Ю. А. Кругляк

НАНОЭЛЕКТРОНИКА «СНИЗУ – ВВЕРХ»

$$I=\frac{2q}{h}\int T(E)M(E)(f_1(E;E_{F1},T_1)-f_2(E;E_{F2},T_2))dE$$

$$f_1(E)=\frac{1}{e^{(E-E_{F1})/kT_1}+1} \qquad f_2(E)=\frac{1}{e^{(E-E_{F2})/kT_2}+1}$$

$$T(E)=\frac{\lambda(E)}{\lambda(E)+L}\Rightarrow\begin{Bmatrix} L\gg\lambda\ ;\ T\ll 1\\ L\approx\lambda\ ;\ T<1\\ L\ll\lambda\ ;\ T\to 1\end{Bmatrix}$$

$$M(E)=\begin{Bmatrix} M_{1D}(E) = \frac{h}{4}\langle v_x^+(E)\rangle D_{1D}(E)\\ WM_{2D}(E) = W\frac{h}{4}\langle v_x^+(E)\rangle D_{2D}(E)\\ AM_{3D}(E) = A\frac{h}{4}\langle v_x^+(E)\rangle D_{3D}(E)\end{Bmatrix}$$

$$G\equiv\frac{I}{V}=\int_{-\infty}^{+\infty}dE\left(-\frac{\partial f_0}{\partial E}\right)G(E) \qquad R=\rho(L+\lambda)\left\{1,\ \frac{1}{W},\ \frac{1}{A}\right\}$$

$$D(E)v(E)p(E)=N(E)\cdot d$$

Одесса
ТЭС
2015

621.38
К84

Ю. А. Кругляк

НАНОЭЛЕКТРОНИКА «СНИЗУ – ВВЕРХ»

Дар от автора
8.03.2016

Одесса
ТЭС
2015

ББК 22.332
УДК 537.32
PACS: 71.15.Mb, 71.20.-b, 73.22.Pr , 73.23.Ad, 84.32.Ff, 85.35.-p

К 840 Кругляк, Юрий Алексеевич

Наноэлектроника «снизу – вверх» /Ю.А.Кругляк. - Одесса: ТЭС, 2015 – 536 стр.; 276 рис.; 3 табл.; 505 лит.

ISBN 978-617-7337-15-6

Книга посвящена современной наноэлектронике в концепции «снизу – вверх», основы которой были заложены еще Рольфом Ландауэром, а ныне активно развиваемой Суприе Даттой, Марком Лундстромом и другими исследователями. Подробно рассмотрена обобщенная модель транспорта электронов и переноса тепла в микро- и наноэлектронике, одинаково пригодная для 0D, 1D, 2D и 3D резисторов произвольного масштаба, характеризуемых любой наперед заданной дисперсией и работающих в баллистическом, диффузионном и промежуточных режимах. В первой части рассматривается классический транспорт, а во второй – квантовый, включая начала спинтроники, магнетроники и молекулярной электроники, на основе метода неравновесных функций Грина в матричном представлении вплоть до квантовой природы классики.

Книга предназначена прежде всего для студентов, магистрантов, аспирантов и преподавателей университетов естественно-научного профиля, как физиков, химиков, так и будущих электроинженеров. Чтобы сделать книгу доступной даже студентам младших курсов, пожалуй, достаточно знать лишь основы линейной алгебры и дифференциальных уравнений, а в области физики понадобятся лишь самые начальные сведения из квантовой механики. Вместе с тем в книге используются новые подходы в преподавании весьма сложных разделов общей физики, что позволило включить в книгу новейшие результаты в области нанофизики, в частности, спинтроники, мало известные русскоязычной университетской аудитории.

ББК 22.332
УДК 537.32
PACS: 71.15.Mb, 71.20.-b, 73.22.Pr , 73.23.Ad, 84.32.Ff, 85.35.-p

ISBN 978-617-7337-15-6

474392
НТБ ВНТУ
м.Вінниця

Жене Наталье,
дочери Наташе, сыновьям Алеше и Жене,
внукам Варе, Паше, Ксюше, Антону и Денису,
правнуку Родиону

Предисловие

У каждого из нас есть мобильный телефон или смартфон или планшет, а многие, особенно студенты и офисные служащие пользуются ноутбуками и стационарными персональными компьютерами. Каждое из этих удивительных устройств имеет от одного до четырех и более миллиардов нанотранзисторов (нанорезисторов), токопроводящая область которых составляет всего несколько сот атомов. В лабораториях уже созданы нанотранзисторы, работающие на одном атоме или на одной молекуле (от молекулы водорода до практически любой другой). Это – наше будущее. Предстоящие годы ознаменуются постепенным переходом от зарядовой архитектуры компьютеров к спиновой архитектуре, появятся компьютеры со спиновой структурой хранения и передачи информации, а для этого нужны магниты для создания спиновых конденсаторов и устройства для передачи информации от магнита к магниту. Надежды связывают с молекулярными магнитами.

Бурное развитие наноэлектроники в последние 15 – 20 лет привело не только к созданию и широкому использованию нанотранзисторов и других разнообразных наноразмерных устройств электроники, но и к более глубокому пониманию причин возникновения тока, обмена и рассеяния энергии и принципов работы электронных устройств в целом как наноразмерных, так и привычных электронных приборов. В наши дни фактически происходят революционные события в электронике, что влечет за собой необходимость пересмотра содержания университетского физического образования. Похожая революционная ситуация наблюдалась 65 лет назад после открытия транзистора, что привело не только к повсеместному использованию устройств микроэлектроники, но и к коренному пересмотру университетских и инженерных курсов общей физики, не говоря уже о специальных курсах в области электроники и смежных дисциплин.

С времен становления физики твердого тела используемые в электронике материалы характеризовались интегральными свойствами такими, например, как подвижность носителей тока или коэффициент оптического поглощения, с дальнейшим их использованием для объяснения наблюдаемых физических явлений и моделирования различных электронных устройств (концепция «сверху – вниз»). С переходом в наши дни к нанотехнологиям нано- и молекулярные транзисторы требуют с самого начала для своего описания и моделирования применения законов квантовой механики и неравновесной статистической термодинамики, что неизбежно приведет к пересмотру физического образования уже на начальных университетских курсах.

Обращает на себя внимание то обстоятельство, что впечатляющие и взрывные успехи экспериментальной наноэлектроники практически не повлияли на то как мы думаем, обучаем и объясняем понятия сопротивления, электрической проводимости, переноса тепла и работу электронных устройств в целом. И поныне по историческим причинам доминирует привычная концепция «сверху – вниз», от массивных проводников до атомов и молекул. Такой подход был вполне приемлем до тех пор, пока не было достаточного массива экспериментальных данных по измерению электрической проводимости нанопроводников. В последнее десятилетие ситуация изменилась. Накоплены обширные экспериментальные данные для предельно малых проводников. Началась разработка концепции электрической проводимости «снизу – вверх», которая не только оказалась совместимой с концепцией «сверху – вниз», но и привела к переосмысливанию принципов работы обычных электронных устройств.

Есть еще один круг задач, для решения которых концепция «снизу – вверх» особенно привлекательна. Это – транспортные задачи, задачи переноса электронов и передачи тепла. В обычной электронике перенос электронов описывается законами механики – классической или квантовой. Перенос электронов по массивному проводнику сопровождается выделением тепла, что описывается законами термодинамики. Процессы в механике обратимы, а в термодинамике необратимы. Разделить эти два процесса – движения электронов и выделения тепла – строго говоря, невозможно. Совсем другая ситуация в наноэлектронике. Здесь процессы движения электронов и выделения тепла пространственно разделены: электроны движутся упруго (без потери энергии), иначе говоря, баллистически, а выделение тепла происходит лишь на границах проводника с его контактами. Развитие концепции «снизу – вверх» Р. Ландауэром, С. Даттой, М. Лундстромом и другими исследователями привело в конечном итоге к созданию объединенной модели переноса электронов и тепла в электронных устройствах как наноскопических, так и микро- и макроскопических, произвольной размерности 1D, 2D и 3D и работающих в баллистическом, диффузионном и промежуточных режимах, изложению которой и посвящена настоящая книга.

Книга предназначена прежде всего для студентов, аспирантов и преподавателей университетов естественно-научного профиля, как физиков, химиков, так и инженеров. Чтобы сделать книгу доступной даже студентам младших курсов, пожалуй, достаточно знать лишь основы линейной алгебры и дифференциальных уравнений, а в области физики понадобятся лишь самые начальные сведения из квантовой механики. Вместе с тем в книге используются новые подходы в преподавании весьма сложных разделов общей физики, что позволило включить в книгу новейшие результаты в области нанофизики, в частности, спинтроники, мало известные русскоязычной университетской аудитории.

Содержание

Предисловие

I. Обобщенная модель транспорта электронов и переноса тепла в микро- и наноэлектронике

Глава 1. Обобщенная модель электронного транспорта

1.1. Концепция «снизу – вверх» в современной наноэлектронике 15

1.2. Причины возникновения электрического тока 19
1.2.1. Роль фермиевских функций 22
1.2.2. Квазиравновесный транспорт электронов 24

1.3. Модель электронного транспорта Ландауэра – Датты – Лундстрома 27
1.3.1. Число электронов и ток в проводнике 28
1.3.2. Моды проводимости 31
1.3.3. Коэффициент прохождения 36
1.3.4. Режим линейного отклика 39
1.3.5. Транспорт в массивных проводниках 40
1.3.6. Баллистический и диффузионный режимы проводимости 43
1.3.6.1. 2D Баллистические резисторы 44
1.3.6.2. Широкие 2D баллистические резисторы при $T = 0° K$ 46
1.3.6.3. Широкие 2D баллистические резисторы при $T > 0° K$ 46
1.3.6.4. Интегралы Ферми – Дирака 48
1.3.6.5. 2D диффузионные резисторы 49
1.3.6.6. Широкие 2D диффузионные резисторы при $T = 0° K$ 49
1.3.6.7. Широкие 2D диффузионные резисторы при $T > 0° K$ 50
1.3.7. О понятии подвижности электронов 51
1.3.8. Способы записи 2D проводимости 54
1.3.9. Диссипация мощности в баллистических резисторах 56
1.3.10. Падение напряжения в баллистических резисторах 57
1.3.11. 1D и 3D резисторы 58
1.3.12. Модель упругого резистора в концепции Ландауэра – Датты 61
1.3.12.1. Диссипация тепла упругим резистором 65
1.3.12.2. Модель упругого резистора и макропроводники 66
1.3.12.3. Баллистический и диффузионный режимы 67
1.3.12.4. О модах проводимости в общем случае 71
1.3.12.5. Фундаментальное соотношение 72

1.3.12.6. Дисперсии $E(p)$ для кристаллических твердых тел 75
1.3.12.7. Еще раз о формуле Друде 77
1.3.12.8. Пропорциональна ли электропроводимость электронной плотности? 80
1.3.12.9. Квантование проводимости и целочисленность мод проводимости 82

Литература 84

Глава 2. Обобщенная модель переноса тепла

2.1. Термоэлектрические явления и устройства 89
2.1.1. Эффект Зеебека 90
2.1.2. Эффект Пельтье 94
2.1.3. Термоэлектрические устройства 98

2.2. Термоэлектрические коэффициенты 104
2.2.1. «Движущие силы» электрического тока 105
2.2.2. Ток в массивных проводниках 109
2.2.3. Перенос тепла электронами 110
2.2.4. Обращенная форма транспортных уравнений 111
2.2.5. Транспортные коэффициенты для 1D проводников 112
2.2.6. Закон Видемана – Франца и числа Лоренца 117
2.2.7. Полупроводники p-типа и биполярная проводимость 118
2.2.8. Транспортные коэффициенты для 3D проводников в диффузионном режиме 120
2.2.9. Коэффициент Зеебека в вырожденном пределе: формула Мотта 127
2.2.10. Анализ экспериментальных данных для Ge: коэффициент Зеебека и термоэлектрические коэффициенты 123
2.2.11. Термоэлектрические коэффициенты для 1D, 2D и 3D проводников с параболической дисперсией 126
2.2.11.1. Термоэлектрические коэффициенты для 1D баллистических резисторов 127
2.2.11.2. Термоэлектрические коэффициенты для 1D диффузионных резисторов 127
2.2.11.3. Термоэлектрические коэффициенты для 2D баллистических резисторов 128
2.2.11.4. Термоэлектрические коэффициенты для 2D диффузионных резисторов 128

2.2.11.5. Термоэлектрические коэффициенты для 3D баллистических резисторов 129
2.2.11.6. Термоэлектрические коэффициенты для 3D диффузионных резисторов 129

2.3. Перенос тепла фононами 130
2.3.1. Элементарная физика фононов 130
2.3.2. Обобщенная модель переноса тепла фононами 133
2.3.2.1. Теплопроводность массивных проводников 137
2.3.2.2. Удельная теплоемкость 138
2.3.2.3. Средняя длина свободного пробега и коэффициент прохождения 139
2.3.2.4. Плотность фононных состояний и фононные моды 142
2.3.2.5. Дебаевская модель 146
2.3.2.6. Рассеяние фононов 148
2.3.2.7. Температурная зависимость теплопроводности 149
2.3.2.8. Различие между решеточной теплопроводностью и электрической проводимостью 150
2.3.2.9. Квантование теплопроводности 150
2.3.3. Термоэлектричество «снизу – вверх» в концепции Ландауэра – Датты
2.3.3.1. Еще раз о коэффициенте Зеебека 154
2.3.3.2. О показателях качества ТЭ устройств и материалов 156
2.3.3.3. Перенос тепла электронами в концепции «снизу – вверх» 157
2.3.3.4. Квазиравновесный режим 159
2.3.3.5. Дельта-термоэлектрик 160
2.3.3.6. Оптимизация силового коэффициента 163
2.3.3.7. О транспорте тепла фононами в концепции «снизу – вверх» 165
2.3.3.8. Баллистический транспорт фононов 166
2.3.3.9. Фононная теплопроводность 167

Литература 170

Глава 3. Графен

3.1. Элементарная теория зонной структуры графена 176
3.1.1. Обратная решетка и зона Бриллюэна графена 181

3.2. Плотность состояний и плотность носителей тока 183

3.3. Число мод и проводимость 184

3.4. Рассеяние в графене 185

3.5. Подвижность и формула Друде 187

3.6. Циклотронная частота в графене 187

3.7. Эффективная масса в графене 190

3.8. Плотность фононных состояний в графене 192

3.9. Вклад электронов и фононов в теплопроводность графена 193

3.10. Максимальная электрическая проводимость графена 195

3.11. Термоэлектрические коэффициенты графена 196
3.11.1. Баллистический режим 196
3.11.2. Диффузионный режим 197

Литература 198

Глава 4. Учет рассеяния в обобщенной транспортной модели

4.1. Качественные аспекты физики рассеяния электронов 201

4.2. Коэффициент прохождения и средняя длина свободного пробега 206

4.3. Средняя длина свободного пробега и время рассеяния 208

4.4. Оценка усредненного значения средней длины свободного пробега из экспериментальных измерений 209

4.5. Оценка длины свободного пробега из коэффициента диффузии 210

4.6. Связь длины свободного пробега с подвижностью 212

4.7. Усредненная длина среднего свободного пробега для степенного закона рассеяния 213

4.8. Подвижность при постоянном значении времени рассеяния 213

4.9. Анализ экспериментальных данных для Si MOSFET 214

Литература 218

Глава 5. Обобщенная модель переноса и транспортное уравнение Больцмана

5.1. Транспортное уравнение Больцмана 219

5.2. Решение уравнения Больцмана в режиме динамического равновесия 223

5.3. Вычисление транспортных коэффициентов 224

5.4. Учет магнитного поля 230

Литература 237

Глава 6. Измерение электрофизических свойств

6.1. Методы измерения сопротивления (проводимости) 241

6.2. Измерение эффекта Холла 243

6.2.1. Пример анализа результатов холловского эксперимента 245

6.3. Метод измерений ван дер Пау 246

6.4. Температурные измерения 251

6.4.1. Температурный артефакт при измерении эффекта Холла (эффект Нернста) 253

6.5. Измерения в сильных магнитных полях 255

6.5.1. Осцилляции Шубникова – де Гааза 259

Литература 261

Глава 7. Роль электростатики и контактов

7.1. Диффузия и дрейф 264

7.2. Уравнение диффузии для баллистического транспорта 265

7.3. Электрохимические потенциалы вдали от равновесия 271

7.4. Токи в режиме неравновесных потенциалов 274

7.5. Сопротивление R_{int} на границе контакта двух проводников с разным числом мод 275

7.6. Уравнение диффузии как следствие уравнения Больцмана 277

7.7. Эффекты при распространении тока по квантовой проволоке 282

7.8. Роль электростатики 287
7.8.1. Полевой транзистор 288
7.8.2. Ток стремится к насыщению 290
7.8.3. Роль заряжания проводника 292
7.8.3.1. Электростатический «выпрямитель» 295
7.8.4. Расширенная модель проводника 297

7.9. Контакты 300
7.9.1. $p - n$ Переходы 301
7.9.2. Вольт-амперные характеристики 303
Литература 307

Глава 8. Термодинамика проводника с током

8.1. Термодинамика проводника с током с учетом окружающей среды 309

8.2. Асимметрия процессов поглощения и излучения энергии 311

8.3. Энтропия 314
8.3.1. Энтропия всегда растет 316
8.3.2. Свободная энергия всегда уменьшается 317

8.4. Универсальный закон равновесия Больцмана 318

8.5. Состояния в пространстве Фока 320
8.5.1. Фермионы 320
8.5.2. Бозоны 321
8.5.3. Взаимодействующие электроны 322

8.6. Альтернативное выражение для энтропии 323
8.6.1. Равновесное распределение из минимума свободной энергии 325

8.7. Глобальная ценность информации 326

8.8. Информационно-управляемый аккумулятор 329
8.8.1. Важно знать детальную информацию 332

8.9. Принцип Ландауэра о минимуме энергии, необходимом для стирания одного бита информации 333

8.10. Демон Максвелла 334

Литература 336

II. Квантовый транспорт, спинтроника, магнетроника, молекулярная электроника и смежные вопросы

Глава 9. Спинтроника и магнетроника

9.1. Введение 339

9.2. Граничное сопротивление и несовпадение мод проводимости 343

9.3. Спиновые потенциалы 346
9.3.1. Разность нелокальных спин-потенциалов 349

9.4. Спиновый момент 352

9.5. Уравнение Ландау – Лифшица – Гильберта 353
9.5.1. Выделенная ось магнита 355

9.6. Обращение намагниченности спиновым током 357

9.7. Поляризаторы и анализаторы спинового тока 358

Литература 360

Глава 10. Метод неравновесных функций Грина

10.1. Квантовый транспорт 363

10.2. Уравнения метода НРФГ в матричном представлении 366

10.3. Одноуровневый резистор: полуклассический подход 369

10.4. Одноуровневый резистор: квантовый подход 371

10.5. Квантовое уширение 374

10.6. Интерференция источников поступления электронов в резистор 375

10.7. Квантовый транспорт по многоуровневому проводнику 376

10.8. Функция проводимости для когерентного транспорта 381

10.9. Модельные транспортные задачи 383
10.9.1. Моделирование 1D проводника 383
10.9.1.1. Баллистический 1D проводник 387
10.9.1.2. Плотность состояний 1D проводника 388
10.9.1.3. 1D проводник с одним рассеивающим центром 389

10.9.2. Моделирование однородного 2D проводника 392
10.9.2.1. Контактные собственные энергии для 2D проводника 398
10.9.3. Графен как пример неоднородного 2D проводника 400
10.9.4. Общий подход к построению контактных матриц 404
10.9.5. Баллистическая проводимость графена 408

10.10. Дефазировка 409
10.10.1. Упругая дефазировка 411
10.10.2. 1D проводник с двумя и более рассеивающими центрами 413
10.10.3. Скачок потенциала на дефектах 417
Литература 421

Глава 11. Эффекты Холла, измерение электрохимических потенциалов, транспорт спинов в модели НРФГ и квантовая природа классики

11.1. Эффекты Холла 425

11.2. Почему отличаются между собой *n*- и *p*-проводники? 429

11.3. Пространственный профиль электрохимического потенциала 430

11.4. Измерение электрохимических потенциалов 434
11.4.1. Формулы Ландауэра 438
11.4.2. Формула Бюттекера 441
11.4.2.1. Не тривиально ли уравнение Бюттекера? 445

11.5. Измерение потенциала Холла 446

11.6. Квантовый эффект Холла и краевые состояния в графене 449

11.7. Формула Кубо 457
11.7.1. Формула Кубо для упругого резистора 460

11.8. Невзаимные электрические цепи 463
11.8.1. Соотношения Онзагера 464

11.9. Спинорное представление 466
11.9.1. Одноуровневый спиновый вентиль 470
11.9.2. Вращение магнитных контактов 474

11.10. Спиновые гамильтонианы 476
11.10.1. Гамильтониан с зеемановским расщеплением 476
11.10.2. Гамильтониан Рашбы 477

11.11. Спиноры и векторы 478

11.12. Прецессия спина 482

11.13. Квантовый спиновый эффект Холла 485

11.14. Метод НРФГ и диффузионный транспорт 486
11.14.1. Электронная плотность в матричном представлении 487
11.14.2. Измерение спинового потенциала 488
11.14.3. Транспорт спинов в четырехкомпонентном формате 489

11.15. Квантовая природа классики 492
11.15.1. Спиновая когерентность 492
11.15.2. Псевдоспин 494
11.15.3. Квантовая энтропия 496
11.15.4. Сколько информации переносит один электрон? 497
11.15.5. Увеличивает ли энтропию взаимодействие частиц? 498
11.15.6. Спины и магниты 499

Литература 500

Глава 12. Кулоновская блокада и одноэлектронный нанотранзистор на молекуле бензола

12.1. За пределами метода НРФГ 505

12.2. Кулоновская блокада в одноэлектронике 506
12.2.1. Вольт-амперные характеристики 509
12.2.2. Моделирование заряжания проводника в пространстве Фока 510
12.2.3. Вычисление тока в фоковском пространстве 513

12.3. Одноэлектронный транзистор на молекуле бензола 515
12.3.1. Элементарная теория SET 517
12.3.2. Условия переноса электрона с истока на сток 520
12.3.3. Формализм теории функционала плотности 521
12.3.3.1. Учет внешнего электростатического поля 524
12.3.4. Результаты расчетов 526

Литература 532

I. Обобщенная модель транспорта электронов и тепла в микро- и наноэлектронике

Глава 1. Обобщенная модель электронного транспорта

1.1. Концепция «снизу – вверх» в современной наноэлектронике

Бурное развитие наноэлектроники в последние 15 – 20 лет привело не только к созданию и широкому использованию нанотранзисторов и других разнообразных наноразмерных устройств электроники, но и к более глубокому пониманию причин возникновения тока, обмена и диссипации энергии и принципов работы устройств в целом как наноразмерных, так и привычных электронных приборов [1 – 4]. В наши дни фактически происходят революционные изменения в электронике, что влечет за собой необходимость пересмотра содержания университетского физического образования. Похожая революционная ситуация наблюдалась 70 лет назад после открытия транзистора, что привело не только к повсеместному использованию устройств микроэлектроники, но и к коренному пересмотру университетских и инженерных курсов общей физики, не говоря уже о специальных курсах в области электроники и смежных дисциплин. С времен становления физики твердого тела используемые в электронике материалы и вещества характеризовались интегральными свойствами такими, например, как подвижность носителей тока или коэффициент оптического поглощения, с дальнейшим их использованием для объяснения наблюдаемых физических явлений и моделирования различных электронных устройств. С переходом в наши дни к мезо- и наноскопике нано- и молекулярные транзисторы требуют с самого начала для своего описания и моделирования применения законов квантовой механики и неравновесной статистической термодинамики, что неизбежно приведет к пересмотру физического образования уже на начальных университетских курсах.

Феноменологическая модель электронной проводимости Друде [5, 6] и поныне занимает центральное место и в преподавании и в экспериментальных измерениях электрических характеристик различных материалов [7]. В качестве причины, вызывающей электрический ток, принимается внешнее электрическое поле

$$\mathrm{E}_x = -\frac{dV}{dx} = \frac{V}{L}, \qquad (1)$$

возникающее как градиент разности потенциалов V, подаваемой на проводник (рис. 1).

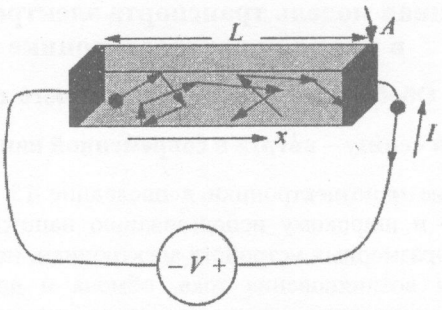


Рис. 1. Электронный транспорт в режиме диффузии через проводник длины L и площади поперечного сечения A. Используется традиционное соглашение о знаках: ток I и направление движения электронов противоположны.

Электрон испытывает действие силы

$$F_e = -q\mathrm{E}_x \tag{2}$$

со стороны электрического поля и движется с дрейфовой скоростью

$$v_d = -\mu \mathrm{E}_x, \tag{3}$$

характеризуемой подвижностью μ.

Ток равен суммарному заряду

$$Q = nqA \cdot L, \tag{4}$$

где n – объемная плотность электронов, перенесенная за время τ,

$$I = Q/\tau, \tag{5}$$

которое в свою очередь определяется скоростью дрейфа

$$\tau = L/v_d. \tag{6}$$

Для тока окончательно имеем

$$I = nqv_d A = nq\mu \mathrm{E}_x A \tag{7}$$

и

$$I = nq\mu \frac{A}{L} V, \tag{8}$$

а для проводимости G и ее удельного значения σ имеем:

$$G = nq\mu \frac{A}{L} = \sigma \frac{A}{L} \tag{9}$$

и

$$\sigma = nq\mu. \tag{10}$$

Согласно закону Ома, сопротивление

$$R \equiv V/I = \rho L/A, \qquad (11)$$

где удельное сопротивление ρ, как и обратная ему удельная проводимость σ в выражении для проводимости (9) не зависят от геометрии проводника и являются свойствами материала, из которого изготовлен проводник. Закон Ома утверждает, что если уменьшить длину проводника в несколько раз, то в это же число раз уменьшится его сопротивление. А если уменьшить длину канала проводимости до очень маленьких размеров, то сопротивление практически «занулится»?

При обычном «диффузионном» движении электронов по проводнику среднее значение длины свободного пробега в полупроводниках меньше 1 *мкм* и изменяется в широких пределах в ту и в другую стороны в зависимости от температуры и природы материала, из которого сделан проводник. Длина канала проводимости в современных полевых транзисторах $\sim 30-40$ *нм*, что соответствует нескольким сотням атомов. Вполне уместно задаться вопросом: а если длина проводника меньше диффузионной длины свободного пробега, то движение электрона станет баллистическим? Будет ли сопротивление подчиняться закону Ома? А если уменьшить длину канала проводимости до нескольких атомов? Имеет ли смысл в этом случае говорить о сопротивлении как таковом? Все эти вопросы были предметом жарких дискуссий еще лет 15 – 20 назад. Сегодня ответы на эти вопросы даны и надежно подкреплены многочисленными экспериментальными данными. Измерено даже сопротивление молекулы водорода [8] и создан полевой транзистор на одиночном атоме фосфора [9].

Сейчас надежно экспериментально установлено, что сопротивление баллистического резистора

$$R^{ball} = \frac{h}{q^2}\frac{1}{M},$$

где фундаментальная константа $h/q^2 = 25.8128$ *кОм* и носит название постоянной Клитцинга в честь ученого, открывшего целочисленный квантовый эффект Холла и впервые доказавшего возможность квантования сопротивления, а целочисленная величина M есть число мод (каналов) проводимости резистора. Этот результат хорошо известен, однако, обычно считают, что он имеет отношение только к ультракоротким резисторам и является предметом обсуждения в спецкурсах по мезофизике и наноэлектронике. Что не очень широко известно, так это то, что сопротивление и для длинных и для коротких резисторов может быть записано в виде

$$R = \frac{h}{q^2}\frac{1}{M}\left(1+\frac{L}{\lambda}\right),$$

где λ есть среднее значение длины свободного пробега в проводнике.

НТБ ВНТУ

Баллистические и диффузионные проводники – это не два разных мира, а один и тот же объект, плавно превращающийся из баллистического резистора в диффузионный проводник по мере того как его длина увеличивается. Баллистический предел очевиден при $L \ll \lambda$, тогда как при $L \gg \lambda$ получаем стандартную запись закона Ома (11). Действительно, сопротивление можно записать в виде

$$R = \frac{\rho}{A}(L + \lambda), \tag{12}$$

тогда для удельного сопротивления получим выражение

$$\rho = \frac{h}{q^2}\left(\frac{1}{M/A}\right)\frac{1}{\lambda},$$

позволяющее по новому интерпретировать сопротивление через удельное число мод проводимости и среднюю длину свободного пробега. Выражение (12) справедливо и для длинных массивных проводников и для ультракоротких нанорезисторов и известно как обобщенный закон Ома, а вытекающие из него следствия будут еще предметом обсуждения в этой книге.

Обращает на себя внимание то обстоятельство, что впечатляющие успехи экспериментальной наноэлектроники практически не повлияли на то как мы думаем, обучаем и объясняем понятия сопротивления, проводимости и работу электронных устройств в целом. И поныне, по-видимому, по историческим причинам доминирует привычная концепция «сверху – вниз», от массивных проводников до атомов и молекул. Такой подход был вполне приемлем до тех пор, пока не было достаточного массива экспериментальных данных по измерению проводимости наноразмерных проводников. В последнее десятилетие ситуация изменилась. Накоплены обширные экспериментальные данные и для больших и для предельно малых проводников. Началась разработка концепции проводимости «снизу – вверх» [10 – 13], которая не только оказалась комплементарной концепции «сверху – вниз», но и привела к переосмысливанию принципов работы обычных электронных устройств. Вспомним, что в квантовой механике с самого начала доминировала концепция «снизу – вверх»: от атома водорода в направлении твердого тела.

Есть еще один круг задач в наноэлектронике, для решения которых концепция «снизу – вверх» особенно привлекательна. Это – транспортные задачи. В обычной электронике транспорт частиц описывается законами механики – классической или квантовой. Транспорт по массивному проводнику сопровождается выделением тепла, что описывается законами термодинамики – обычной или статистической. Процессы в механике обратимы, а в термодинамике необратимы. Разделить эти два процесса – движения и выделения тепла – строго говоря, невозможно. Совсем другая ситуация в наноэлектронике. Здесь процессы движения электронов и выделения тепла

пространственно разделены: электроны движутся упруго, баллистически («упругий резистор»), а выделение тепла происходит лишь на границах контакта проводника с электродами. Концепция «упругого резистора» была предложена Ландауэром в 1957 году [14–16] задолго до ее экспериментального подтверждения в нанотранзисторах. Концепция «упругого резистора», строго говоря, является идеализацией, но она надежно подтверждена многочисленными экспериментальными данными для ультрамалых нанотранзисторов. Развитие концепции «снизу – вверх» Рольфом Ландауэром, Суприе Даттой и Марком Лундстромом (ЛДЛ) привело в конечном итоге к созданию обобщенной модели транспортных явлений в электронных устройствах как наноскопических, так и микро- и макроскопических, произвольной размерности 1D, 2D и 3D и работающих в баллистическом, квазибаллистическом или диффузионном режимах, изложению которой и посвящено настоящая глава, как и книга в целом. В своем изложении мы будем ближе всего следовать лекциям Марка Лундстрома «Near-Equilibrium Transport: Fundamentals and Applications» [11], прочитанных им в 2011 году, и его лекциям «Nanotechnology: Fundamentals of Nanotransistors», прочитанных в 2016 году, а также лекциям Суприе Датты «Fundamentals of Nanoelectronics, Part 1: Basic Concepts» и «Fundamentals of Nanoelectronics, Part 2: Quantum Transport», прочитанных им в 2012 и в 2015 годах в рамках инициативы Purdue University / nanoHUB-U (www.nanohub.org/u) [11,12, 19].

Начнем мы с казалось бы очевидного вопроса о причинах возникновения электрического тока.

1.2. Причины возникновения электрического тока

На вопрос о причине возникновения тока при приложении разности потенциалов к концам проводника обычно ссылаются на связь плотности тока $\vec{J}$ с внешним приложенным электрическим полем $\vec{E}$

$$\vec{J} = \sigma \vec{E},$$

что следует из (7); другими словами, причиной возникновения тока обычно считают электрическое поле. Ответ, в лучшем случае, не полный. Еще до подключения проводника к клеммам источника напряжения на электроны проводника действуют сильные электрические поля, создаваемые ядрами атомов, а ток тем не менее не возникает. Почему сильные внутренние электрические поля не вызывают движение электронов, а намного более слабое внешнее электрическое поле батареи вызывает движение электронов? Обычно говорят, что внутренние микроскопические поля якобы не могут вызвать

движение электронов, необходимо приложить внешнее макроскопическое поле. Трудно признать такое объяснение удовлетворительным. В современных экспериментах по измерению проводимости отдельных молекул невозможно с определенностью вычленить отдельно внутренние и внешние электрические поля. Приходится считаться с этим уроком, преподнесенным нам современной экспериментальной наноэлектроникой, и заново задаться вопросом почему же движутся электроны при подключении батареи к концам проводника.

Для ответа на вопрос о причине возникновения тока нам с самого начала потребуются два понятия -- плотности состояний проводника на единицу энергии $D(E)$, занятых электронами и свободных и электрохимического потенциала E_{F0} (рис. 2).

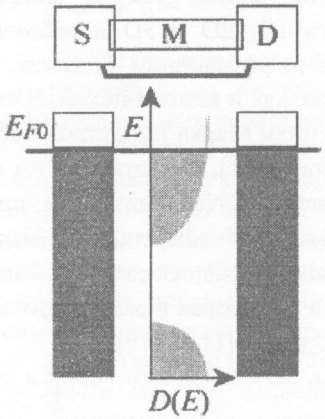


Рис. 2. Первым шагом в объяснении работы электронного устройства должно быть задание плотности состояний $D(E)$ внутри проводника M и определение равновесного значения электрохимического потенциала E_{F0}, отделяющего заполненные электронами состояния от пустых состояний.

Простоты ради, что никак не скажется на окончательных выводах, будем пользоваться так называемой точечной моделью проводника, которая предполагает неизменность плотности состояний $D(E)$ при движении вдоль проводника. Если система, включающая истоковый электрод (S/Source), проводник M и стоковый электрод (D/Drain), находится в равновесии (закорочена), то электрохимический потенциал E_{F0} везде одинаков и все состояния с $E < E_{F0}$ заполнены электронами, а состояния с $E > E_{F0}$ пустые (рис. 2).

При включении в цепь источника напряжения (рис. 3) разность потенциалов V понижает все энергии на положительном электроде D на

величину qV, где q – заряд электрона, в результате чего на электродах создается разность электрохимических потенциалов

$$E_{F1} - E_{F2} = qV . \qquad (13)$$

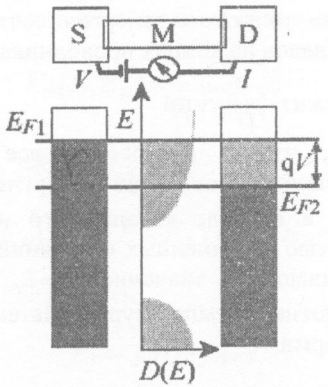


Рис. 3. При подаче напряжения V на клеммы проводника потенциал анода D понижается на величину qV, создавая разность электрохимических потенциалов на концах проводника.

Точно так же как разность температур приводит к потоку тепла, а различие в уровнях жидкости ведет к ее перетоку, так и разность электрохимических потенциалов является причиной возникновения тока. Только состояния проводника в окне $E_{F1} - E_{F2}$ и находящиеся достаточно близко к значениям E_{F1} и E_{F2} дают вклад в поток электронов, тогда как все состояния, лежащие значительно выше E_{F1} и ниже E_{F2}, не играют никакой роли. Причина этого кроется в следующем.

Каждый контакт стремится привести токовый канал в равновесие с собой путем заполнения электронами всех состояний канала с энергией, меньшей электрохимического потенциала E_{F1}, и опорожнения состояний канала с энергией, большей потенциала E_{F2}. Рассмотрим токовый канал с состояниями, энергия которых меньше E_{F1}, но больше E_{F2}. Контакт 1 стремится заполнить эти состояния, поскольку их энергия меньше E_{F1}, а контакт 2 стремится опорожнить эти состояния, поскольку их энергия больше E_{F2}, что и приводит к непрерывному движению электронов от контакта 1 к контакту 2.

Рассмотрим теперь состояния канала с энергией, большей E_{F1} и E_{F2}. Оба контакта стремятся опорожнить эти состояния, но они и так пустые и не дают вклада в электрический ток. Аналогична ситуация с состояниями, энергия которых одновременно меньше обоих потенциалов E_{F1} и E_{F2}. Каждый из контактов стремится заполнить их электронами, но они уже заполнены, и

вклада в ток дать не могут, а точнее не могут в пределах нескольких kT от ширины окна, в чем мы убедимся позже.

Подобная картина выглядит почти очевидной, если бы не привычное утверждение о том, что электроны движутся под действием электрического поля внутри проводника. Если бы это было так, то вклад в ток должны были бы дать все электроны, а не только те, энергия состояний которых лежит в пределах разности потенциалов на концах проводника.

1.2.1. Роль фермиевских функций

Итак, утверждалось, что в равновесии все состояния с энергией $E < E_{F0}$ заполнены электронами, а состояния с энергией $E > E_{F0}$ пустые (рис. 2). Это справедливо только в пределе абсолютного нуля температуры. Более точно, переход от полностью заполненных состояний к пустым совершается в зазоре $\sim \pm 2\,kT$, охватывающем значение $E = E_{F0}$, где k – постоянная Больцмана, T – абсолютная температура. Математически этот переход описывается функцией Ферми

$$f(E) = \frac{1}{e^{(E-E_F)/kT} + 1}. \qquad (14)$$

График функции Ферми показан на рис. 4 слева, возможно, в несколько непривычном виде с энергией в безразмерных единицах по вертикальной оси, что позволит нам позже при объяснении причины токообразования совместить функцию Ферми с графиком плотности состояний $D(E)$.

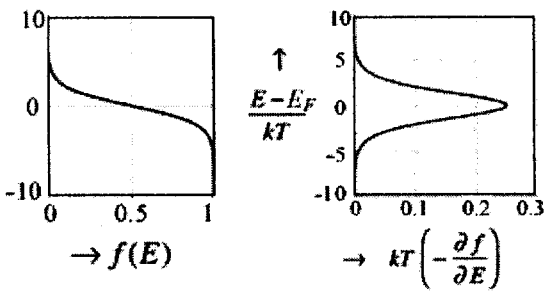


Рис. 4. Графики функции Ферми и нормированной функции теплового уширения.

Функция Ферми играет ключевую роль в статистической механике, однако же для наших целей достаточно понимать, что состояния с низкой энергией всегда заняты ($f = 1$), тогда как состояния с высокой энергией всегда пустые ($f = 0$), а переход от $f = 1$ к $f = 0$ происходит в узком интервале энергии $\sim \pm 2\,kT$, охватывающем значение $E = E_{F0}$.

Действительно, на рис. 4 показана производная от функции Ферми, помноженная на kT с тем, чтобы сделать ее безразмерной:

$$F_T\left(E,E_F\right)=kT\left(-\frac{\partial f}{\partial E}\right). \tag{15}$$

Подставляя выражение для фермиевской функции (14), видно, что

$$F_T\left(E,E_F\right)=\frac{e^x}{\left(e^x+1\right)^2}, \tag{16}$$

где $x\equiv (E-E_F)/kT$. Из (16) видно, что

$$F_T\left(E,E_F\right)=F_T\left(E-E_F\right)=F_T\left(E_F-E\right), \tag{17}$$

а из уравнений (16) и (14) следует, что

$$F_T=f\left(1-f\right). \tag{18}$$

Интегрирование функции (18) во всем интервале изменения энергии дает площадь, равную kT:

$$\int_{-\infty}^{+\infty}dEF_T\left(E,E_F\right)=kT\int_{-\infty}^{+\infty}dE\left(-\frac{\partial f}{\partial E}\right)=kT\left[-f\right]_{-\infty}^{+\infty}=kT\left(1-0\right)=kT. \tag{19}$$

Так что функцию F_T можно приблизительно представить себе в виде прямоугольного «импульса», центрированного относительно значения $E=E_F$ с высотой, равной ¼ и шириной ~ $4\,kT$. Функция $(-{}^{\partial f}\!/_{\partial E})$ получила название фермиевского окна проводимости.

Когда рассматриваемая система находится в равновесии (рис. 2), электроны распределяются по имеющимся состояниям в соответствии с функцией Ферми. При выходе из равновесия нет простых правил для вычисления функции распределения электронов. Все зависит от конкретной задачи, решать которую нужно методами неравновесной статистической механики.

В нашем специальном случае выхода из равновесия (рис. 3) можно надежно утверждать, что оба контакта S и D настолько велики по сравнению с каналом переноса электронов, что они не могут выйти из равновесия. Каждый из контактов локально находится в равновесии со своим собственным электрохимическим потенциалом, порождая две фермиевских функции (рис. 5):

$$f_1\left(E\right)=\frac{1}{e^{(E-E_{F1})/kT}+1} \tag{20}$$

и

$$f_2(E)=\frac{1}{e^{(E-E_{F2})/kT}+1}. \tag{21}$$

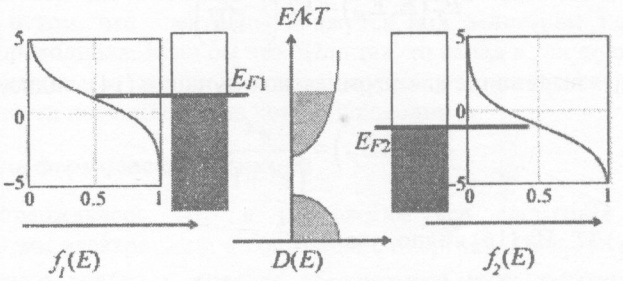


Рис. 5. При выходе из равновесия электроны в контактах занимают доступные им состояния в соответствии с фермиевскими распределениями и значениями электрохимических потенциалов.

Подводя итоги, утверждается, что *причиной появления тока является различие в подготовке равновесных состояний контактов, отображаемое их соответствующими фермиевскими функциями* $f_1(E)$ *и* $f_2(E)$. Качественно это справедливо для любых проводников – и наноразмерных и макроскопических. Однако, для наноразмерных проводников ток при любых значениях энергии электронных состояний в проводнике, как будет показано позже, пропорционален разности фермиевских распределений в обоих контактах: $I(E)\sim f_1(E)-f_2(E)$. Эта разность зануляется, если энергия E больше E_{F1} и E_{F2}, поскольку в этом случае обе функции Ферми равны нулю. Эта разность также зануляется, если энергия E меньше и E_{F1} и E_{F2},поскольку в этом случае обе фермиевские функции равны единице. Ток возникает лишь в окне $E_{F1}-E_{F2}$, если в этом окне есть хотя бы одно электронное состояние проводника.

1.2.2. Квазиравновесный транспорт электронов

Вольт-амперная характеристика обычно нелинейная, но из нее можно вычленить участок «линейного отклика», под которым подразумевается проводимость dI/dV при $V\to 0$.

Построим функцию разности двух фермиевских функций, нормированную к приложенному напряжению,

$$F(E)=\frac{f_1(E)-f_2(E)}{qV/kT}, \tag{22}$$

где

$$\begin{aligned} E_{F1}&=E_{F0}+(qV/2),\\ E_{F2}&=E_{F0}-(qV/2). \end{aligned} \tag{23}$$

Функция разности $F(E)$ сужается по мере того, как напряжение V, помноженное на заряд электрона, становится все меньше величины kT (рис. 6).

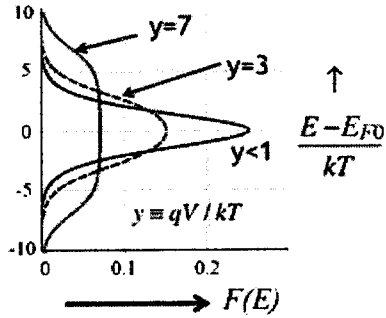


Рис. 6. График функции разности $F(E)$ в зависимости от значения $(E-E_{F0})/kT$ для различных значений $qV/kT \equiv y$.

Отметим также, что по мере того, как kT начинает превосходить энергию qV, функция $F(E)$ все больше приближается к функции теплового уширения:

$$F(E) \to F_T(E) \text{ при } qV/kT \to 0,$$

так что из ур-я (22) следует, что

$$f_1(E) - f_2(E) = \frac{qV}{kT} F_T(E, E_{F0}) = \left(-\frac{\partial f_0}{\partial E} \right) qV, \tag{24}$$

если приложенное напряжение, помноженное на заряд электрона, $qV = E_{F1} - E_{F2}$ становится намного меньшим kT.

Нам потребуется также следующее выражение

$$f(E) - f_0(E) = -\frac{\partial f_0}{\partial E}(E_F - E_{F0}), \tag{25}$$

которое, как и уравнение (24), можно получить следующим образом.

Для функции Ферми

$$f(x) \equiv \frac{1}{e^x + 1}, \quad x \equiv \frac{E - E_F}{kT} \tag{26}$$

имеем

$$\begin{aligned} \frac{\partial f}{\partial E} &= \frac{df}{dx}\frac{\partial x}{\partial E} = \frac{df}{dx}\frac{1}{kT}, \\ \frac{\partial f}{\partial E_F} &= \frac{df}{dx}\frac{\partial x}{\partial E_F} = -\frac{df}{dx}\frac{1}{kT}, \\ \frac{\partial f}{\partial T} &= \frac{df}{dx}\frac{\partial x}{\partial T} = -\frac{df}{dx}\frac{E - E_F}{kT^2}, \end{aligned} \tag{27}$$

откуда

$$\frac{\partial f}{\partial E_F} = -\frac{\partial f}{\partial E},$$
$$\frac{\partial f}{\partial T} = -\frac{E - E_F}{T}\frac{\partial f}{\partial E}. \qquad (28)$$

Уравнениее (25) получается из разложения функции Ферми вблизи точки равновесия в ряд Тейлора

$$f\left(E, E_F\right) \cong f\left(E, E_{F0}\right) + \left(\frac{\partial f}{\partial E_F}\right)_{E_F = E_{F0}} \left(E_F - E_{F0}\right). \qquad (29)$$

Из уравнения (28) следует

$$\left(\frac{\partial f}{\partial E_F}\right)_{E_F = E_{F0}} = \left(-\frac{\partial f}{\partial E}\right)_{E_F = E_{F0}}. \qquad (30)$$

Пусть $f(E)$ соответствует $f(E, E_F)$, а $f_0(E)$ соответствует $f(E, E_{F0})$, тогда

$$f(E) \approx f_0(E) + (E_F - E_{F0})\left(-\frac{\partial f}{\partial E}\right), \qquad (31)$$

что после перегруппировки дает искомое уравнение (25), которое справедливо при $E_F - E_{F0} << kT$.

Подведем итоги. Проводимость веществ может меняться более чем в 10^{20} раз, переходя, например, от серебра до стекла – веществ, весьма удаленных друг от друга в шкале проводимости. Стандартное объяснение различия в их проводимости якобы состоит в том, что плотность «свободных электронов» в этих веществах существенно отличается. Подобное объяснение немедленно требует пояснения, какие электроны свободные, а какие нет. Это различие становится все более абсурдным по мере перехода к наноразмерным проводникам.

Концепция «снизу – вверх» предлагает следующий простой ответ. Проводимость зависит от плотности состояний в окне шириной несколько kT, охватывающих равновесный электрохимический потенциал E_{F0}, определяемый функцией F_T (ф-ла (15), рис. 4), которая отлична от нуля в небольшом промежутке шириной несколько kT вокруг равновесного значения электрохимического потенциала.

Дело не в суммарном числе электронов, которое одного порядка как в серебре, так и в стекле. Ключевым моментом является наличие электронных состояний в области значений электрохимического потенциала E_{F0} (фермиевское окно проводимости), что в корне отличает одно вещество от другого.

Настоящий ответ нельзя назвать новым, и он хорошо известен специалистам в области микро- и наноэлектроники. Тем не менее обсуждение проводимости и поныне обычно начинается с теории Друде [7], которая сыграла важнейшую историческую роль в понимании природы электрического тока. К сожалению, подход Друде породил два недоразумения, которые следовало бы преодолеть, а именно:

(1) ток порождается электрическим полем и

(2) ток зависит от числа электронов.

Оба недоразумения связаны друг с другом, поскольку если бы ток действительно порождался бы электрическим полем, то все электроны были бы подвержены влиянию поля.

Уроки, преподнесенные нам экспериментальной наноэлектроникой, показывают, что ток порождается «подготовкой» двух контактов $f_1(E) - f_2(E)$, и эта разница не нулевая только в фермиевском окне проводимости вокруг равновесного электрохимического потенциала E_{F0}. Проводимость канала высокая или низкая зависит только от наличия электронных состояний в этом окне. К этому выводу обычно приходят на основе транспортного уравнения Больцмана [22] или формализма Кубо [23], тогда как предлагаемая концепция «снизу – вверх» сразу дает корректную картину возникновения тока.

1.3. Модель электронного транспорта Ландауэра – Датты – Лундстрома

Мы рассмотрим модель электронного транспорта довольно простую и вместе с тем удивительно полезную в прикладном отношении и нашедшую широкое применение при анализе как электронного, так и фононного транспорта не только в режиме линейного отклика, но и в высоковольтном режиме горячих электронов, нелокального и квантового транспорта, транспорта в неупорядоченных и наноструктурированных материалах как наноразмерных, так и протяженных, в которых проводимость определяется только свойствами проводника. Свое начало она берет в работах Ландауэра [14 – 16], переосмысленных Даттой [12, 17, 18] и далее развитых Лундстромом применительно к самым различным материалам [13, 19].

Центральное место в любом электронном устройстве играет канал проводимости, характеризуемый плотностью состояний $D(E-U)$, где E – энергия состояний проводника, а U – самосогласованный электростатический потенциал затвора, позволяющий смещать состояния резистора вверх или вниз по шкале энергии (рис. 7).

Далее будем рассматривать двухтерминальное устройство $(U=0)$. Канал проводимости связан с двумя «идеальными» контактами, способными к быстрому восстановлению равновесия в процессе электронного транспорта и характеризуемыми фермиевскими функциями (20) и (21) с соответствующими электрохимическими потенциалами E_{F1} и E_{F2}. При подаче напряжения V электрохимический потенциал E_{F2} понижается относительно потенциала E_{F1} на величину qV.

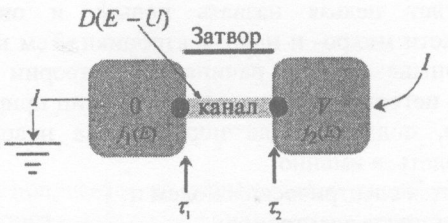


Рис. 7. Типичное наноразмерное электронное устройство с контактами, характеризуемыми временами пролета τ.

Связь контактов с каналом проводимости характеризуется временами пролета τ, дающими представление о том, как быстро электроны могут покинуть контакт или проводник. Для наноразмерных резисторов, например, молекул, времена τ контролируются контактами. Для протяженных резисторов с хорошими контактами, как мы убедимся позже, времена τ становятся сопоставимыми с временами пролета всей протяженности проводника. В общем случае качество обоих контактов может быть различным, что ведет к различным значениям τ_1 и τ_2. Иногда времена τ удобнее выражать в единицах энергии $\gamma = \hbar/\tau$. Если в роли канала проводимости выступает одиночная молекула, величина γ приобретает простой физический смысл уширения энергетических уровней молекулы за счет конечности времени жизни электронов на молекулярных уровнях.

Предполагается, что канал имеет зонную структуру $E(k)$. Это требование, однако, не является обязательным [17]. Обсуждение ограничений и применимости модели ЛДЛ можно найти в [12, 13, 17, 18].

Перейдем к построению математической модели транспорта ЛДЛ в концепции «снизу – вверх». Начать надо с нахождения зависимости плотности электронов и тока в проводнике в зависимости от фермиевских распределений и соответствующих электрохимических потенциалов, плотности состояний проводника и характеристических времен τ пролета.

1.3.1. Число электронов и ток в проводнике

Пусть пока только левый контакт 1 связан с каналом проводимости. Он стремится пополнить проводник электронами в соответствии со своим электрохимическим потенциалом E_{F1}. В конце концов равновесие между контактом и каналом проводимости наступит тогда, когда число электронов с энергиями между E и $E+dE$ станет равным

$$N'_{01}(E)dE = D(E)dE\,f_1(E), \tag{32}$$

где $D(E)$ – плотность состояний с энергией E (пусть каждое состояние заполнено двумя электронами с противоположными спинами), а $f_1(E)$ есть равновесная функция Ферми контакта 1.

Процесс установления равновесия между контактом 1 и каналом описывается простым кинетическим уравнением

$$F_1 = \left.\frac{dN'(E)}{dt}\right|_1 = \frac{N'_{01}(E) - N'(E)}{\tau_1(E)}, \tag{33}$$

где скорость подачи электронов в канал dN'/dt положительна, если число электронов в канале N' меньше его равновесного значения N'_{01}, и отрицательна в противном случае. Если канал первоначально пустой, то он заполняется электронами до их равновесного числа; если же канал переполнен электронами, то он опустошается до тех пор, пока в канале не будет достигнуто равновесное число электронов.

Если канал проводимости связан только с контактом 2, аналогичные рассуждения приводят к следующей паре уравнений:

$$N'_{02}(E)dE = D(E)dEf_2(E), \tag{34}$$

$$F_2 = \left.\frac{dN'(E)}{dt}\right|_2 = \frac{N'_{02}(E) - N'(E)}{\tau_2(E)}. \tag{35}$$

Когда канал проводимости связан одновременно с двумя контактами и оба контакта одновременно пополняют электронами канал и опорожняют его, имеем

$$\left.\frac{dN'(E)}{dt}\right|_{1+2} = F_1 + F_2 = \left.\frac{dN'(E)}{dt}\right|_1 + \left.\frac{dN'(E)}{dt}\right|_2. \tag{36}$$

В состоянии динамического равновесия левые части кинетического уравнения (36) равны нулю. Приравнивая нулю правую часть равенства (36) и подставляя выражения для скоростей на обоих контактах 1 и 2 по уравнениям (33) и (35), для числа электронов в канале проводимости получим:

$$N'(E) = \frac{(1/\tau_1)}{(1/\tau_1)+(1/\tau_2)}N'_{01}(E) + \frac{(1/\tau_2)}{(1/\tau_1)+(1/\tau_2)}N'_{02}(E). \tag{37}$$

Простоты ради, положим $\tau_1 = \tau_2$ и подставим равновесные значения числа электронов N'_{01} и N'_{02} по уравнениям (32) и (34). Тогда для числа электронов в канале в состоянии динамического равновесия в дифференциальной форме получим:

$$N'(E)dE = \frac{D(E)dE}{2}f_1(E) + \frac{D(E)dE}{2}f_2(E). \tag{38}$$

Остается проинтегрировать по всему спектру энергий и получим число электронов в канале проводимости в состоянии динамического равновесия двух

контактов с электрохимическими потенциалами E_{F1} и E_{F2}, разница между которыми пропорциональна напряжению, поданному на концы проводника, а именно:

$$N = \int N'(E)dE = \int \left[\frac{D(E)}{2} f_1(E) + \frac{D(E)}{2} f_2(E) \right] dE . \qquad (39)$$

Получен ответ на первый из поставленных нами двух вопросов: найдено число электронов, или иначе плотности электронов, если известны геометрические параметры проводника, в зависимости от плотности состояний проводника и фермиевских распределений на контактах. Когда электронное устройство (рис. 7) находится в равновесии ($f_1 = f_2 \equiv f_0$), получаем стандартное выражение для равновесного числа электронов в проводнике [20]

$$N_0 = \int D(E) f_0(E) dE , \qquad (40)$$

в отличие от которого уравнение (39) пригодно как для равновесного состояния устройства (рис. 7), так и для случая, когда электронное устройство выведено далеко за пределы равновесия.

Напомним, что для 1D-, 2D- и 3D-проводников плотность состояния пропорциональна, соответственно, длине L проводника, площади его поперечного сечения A и объему проводника Ω:

$$D(E) \approx \{L, A, \Omega\}, \qquad (41)$$

а плотность электронов определяется, соответственно, как

$$n_L = N/L,\ n_S = N/A,\ n = N/\Omega . \qquad (42)$$

Перейдем к вычислению тока в состоянии динамического равновесия. В этом состоянии один из контактов пополняет канал проводимости электронами, а другой контакт опустошает его. Если $E_{F1} > E_{F2}$, контакт 1 инжектирует электроны, а контакт 2 поглощает их, и наоборот, если $E_{F1} < E_{F2}$.

Скорости, с которыми электроны покидают или поглощаются контактами 1 и 2, даются уравнениями (33) и (35). В состоянии динамического равновесия

$$F_1 + F_2 = 0 . \qquad (43)$$

По договоренности, ток считается положительным, если он входит в контакт 2 (рис. 7), так что

$$I' = qF_1 = -qF_2 . \qquad (44)$$

Выразив ток один раз через F_1, а второй раз через F_2 согласно (44), и сложив эти одинаковые токи, сразу получим, что

$$I' = \frac{q}{2}(F_1 - F_2) . \qquad (45)$$

Остается подставить значения для скоростей (33) и (35) в предположении одинаковости контактов $(\tau_1 = \tau_2 \equiv \tau)$, затем воспользоваться (32) и (34), и мы имеем:

$$I'(E) = \frac{q}{2\tau(E)}\left(N'_{01} - N'_{02}\right) = \frac{2q}{h}\frac{\gamma(E)}{2}\pi D(E)(f_1 - f_2), \qquad (46)$$

где характеристическое время пролета τ выражено в единицах энергии, а именно:

$$\gamma \equiv \frac{\hbar}{\tau(E)}. \qquad (47)$$

Интегрируя (46) по всему спектру энергий, окончательно для тока в состоянии динамического равновесия получаем

$$I = \int I'(E)dE = \frac{2q}{h}\int \gamma(E)\pi\frac{D(E)}{2}\left(f_1 - f_2\right)dE. \qquad (48)$$

Согласно (48), ток возникает лишь тогда, когда фермиевские распределения на контактах различны, то ли по причине различия электрохимических потенциалов, то ли по причине различия в температуре контактов, то ли сразу по этим обеим причинам. При различии в электрохимических потенциалах контактов один из контактов стремится передать электроны проводнику, а другой стремится принять их на себя. Ток при различии в температуре контактов рассматривается в главе 2.

Подведем итоги. Получены два основных уравнения транспортной модели ЛДЛ:

$$N = \int\frac{D(E)}{2}\left(f_1 + f_2\right)dE, \qquad (49)$$

$$I = \frac{2q}{h}\int\gamma(E)\pi\frac{D(E)}{2}\left(f_1 - f_2\right)dE. \qquad (50)$$

Первое из них выражает число электронов в проводнике в состоянии динамического равновесия через плотность состояний проводника и фермиевские функции контактов, а второе уравнение дает ток через эти же характеристики электронного устройства и через характеристическое время пролета. Дальнейшее построение модели ЛДЛ фактически сводится к использованию и развитию этих двух основных уравнений.

1.3.2. Моды проводимости

Как и ожидалось, ток пропорционален разности фермиевских функций контактов. Позже мы убедимся, что комбинация фундаментальных констант $2q/h$ играет важную роль в выражении для тока (50). А каков смысл произведения $\gamma\pi D/2$?

Параметр γ (47) имеет размерность энергии, а размерность плотности состояний D есть (энергия)$^{-1}$, так что интересующее нас произведение $\gamma\pi D/2$

не имеет размерности. В отношении размерности плотности состояний обратим внимание на то, что мы оперируем с числом электронов, а не с плотностями электронов (42), как это обычно принято в физике полупроводников, где плотность состояний для 3D-образца имеет размерность (энергия·объем)$^{-1}$. Мы сейчас увидим, что произведение $\gamma(E)\pi\dfrac{D(E)}{2}=M(E)$ имеет физический смысл числа мод (иначе, каналов) проводимости резистора.

Рассмотрим 2D баллистический резистор длины L, меньшей средней длины свободного пробега λ, вытянутый вдоль оси x и шириной W вдоль оси y. Полная плотность состояний

$$D(E)=D_{2D}(E)L\cdot W, \tag{51}$$

$$D_{2D}(E)=g_v\frac{m^*}{\pi\hbar^2}, \tag{52}$$

где D_{2D} есть удельная плотность состояний (на единицу площади), выписанная для параболической зонной структуры с эффективной массой m^* и долинным вырождением g_v [20].

Определим характеристическое время τ. Из (38) и (46) имеем:

$$\frac{qN'(E)dE}{I'(E)dE}=\frac{\hbar}{\gamma}\frac{\left(f_1+f_2\right)}{\left(f_1-f_2\right)}. \tag{53}$$

Приложим к концам проводника достаточно большое напряжение такое, чтобы имело место $E_{F2}\ll E_{F1}$, так что $f_2\ll f_1$. Тогда из (53) имеем

$$\frac{qN'(E)dE}{I'(E)dE}=\frac{\text{Накопленный заряд}}{\text{Ток}}=\frac{\hbar}{\gamma}=\tau(E). \tag{54}$$

Число электронов в канале

$$N'(E)=n'_S(E)L\cdot W, \tag{55}$$

где n'_S – удельная электронная плотность (на единицу площади), а ток

$$I'(E)=qWn'_S(E)\left\langle v_x^+(E)\right\rangle, \tag{56}$$

так что из (54) имеем:

$$\tau(E)=\frac{L}{\left\langle v_x^+(E)\right\rangle}, \tag{57}$$

что есть просто среднее время пролета электрона через всю длину упругого резистора.

Для оценки τ нужно вычислить среднюю скорость пролета от контакта 1 до контакта 2 в направлении $+x$ (рис. 1). Эта скорость в случае баллистического транспорта, т.е. без изменения направления движения

$$\left\langle v_x^+(E)\right\rangle = v(E)\left\langle\cos\theta\right\rangle, \tag{58}$$

где угол θ отсчитывается от положительного направления оси x. Поскольку

$$\left\langle\cos\theta\right\rangle = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\cos\theta d\theta = \frac{2}{\pi}, \tag{59}$$

для средней скорости в случае параболического дисперсионного соотношения и изотропии скорости имеем:

$$\left\langle v_x^+(E)\right\rangle = \frac{2}{\pi}v = \frac{2}{\pi}\sqrt{\frac{2\left(E-E_C\right)}{m^*}}. \tag{60}$$

Определяя число мод проводимости как

$$M(E) = \gamma(E)\pi\frac{D(E)}{2} \tag{61}$$

и используя определение γ по (47) и плотности состояний для 2D проводника по (51) для числа мод 2D проводника окончательно получаем

$$M(E) = WM_{2D}(E) = W\frac{h}{4}\left\langle v_x^+(E)\right\rangle D_{2D}(E). \tag{62}$$

Аналогичные соображения для 1D и 3D проводников дают

$$\begin{aligned} M(E) &= M_{1D}(E) = \frac{h}{4}\left\langle v_x^+(E)\right\rangle D_{1D}(E), \\ M(E) &= WM_{2D}(E) = W\frac{h}{4}\left\langle v_x^+(E)\right\rangle D_{2D}(E), \\ M(E) &= AM_{3D}(E) = A\frac{h}{4}\left\langle v_x^+(E)\right\rangle D_{3D}(E). \end{aligned} \tag{63}$$

Обратим внимание, что в случае 2D проводников число мод проводимости пропорционально ширине проводника W, а для 3D проводников – площади их поперечного сечения A.

Заслуживает внимания физический смысл полученных результатов (63). Так, для 2D проводника с учетом (60) и (52) число мод проводимости

$$WM_{2D}(E) = g_v W\frac{\sqrt{2m^*(E-E_C)}}{\pi\hbar}. \tag{64}$$

Для частного случая параболической дисперсии

$$E(k) = E_C + \frac{\hbar^2 k^2}{2m^*} \tag{65}$$

и выражая волновой вектор через де бройлевскую длину волны

$$\lambda_B = 2\pi/k, \tag{66}$$

вместо общего выражения (64) для числа мод проводимости 2D проводника имеем:

$$WM_{2D}(E) = g_v \frac{Wk}{\pi} = g_v \frac{W}{\lambda_B(E)/2}, \tag{67}$$

что открывает возможность их физической интерпретации, а именно: число мод проводимости 2D проводника (с учетом долинного вырождения) показывает сколько де бройлевских полуволн энергии E укладывается по ширине 2D проводника. Целочисленность мод обеспечивается граничными условиями, согласно которым волновые функции электронов на обеих сторонах 2D проводника должны обращаться в нуль. Сам термин «моды» заимствован из теории волноводов.

Теперь мы можем базовые уравнения модели ЛДЛ (49) и (50) переписать в виде

$$N = \int \frac{D(E)}{2}(f_1 + f_2)dE,$$

$$I = \frac{2q}{\hbar}\int \gamma(E)\pi \frac{D(E)}{2}(f_1 - f_2)dE. \tag{68}$$

Таким образом, для вычисления числа электронов и тока в проводнике кроме фермиевских функций контактов нужно знать плотность состояний $D(E)$ и число мод $M(E)$ проводника.

Для параболической дисперсии (65) плотности состояний хорошо известны [20]:

$$D_{1D}(E)L = \frac{L}{\pi\hbar}\sqrt{\frac{2m^*}{(E-E_C)}}H(E-E_C), \tag{69}$$

$$D_{2D}(E)A = A\frac{m^*}{\pi\hbar^2}H(E-E_C), \tag{70}$$

$$D_{3D}(E)\Omega = \Omega\frac{m^*\sqrt{2m^*(E-E_C)}}{\pi^2\hbar^3}H(E-E_C), \tag{71}$$

где $H(E - E_c)$ – ступенчатая функция Хевисайда, а соответствующие моды проводимости таковы:

$$M_{1D}(E) = H\left(E - E_C\right), \tag{72}$$

$$M_{2D}(E) = g_v \frac{\sqrt{2m^*(E - E_C)}}{\pi\hbar} H\left(E - E_C\right), \tag{73}$$

$$M_{3D}(E) = g_v \frac{m^*(E - E_C)}{2\pi\hbar^2} H\left(E - E_C\right). \tag{74}$$

Поведение плотности состояний $D(E)$ и числа мод $M(E)$ для 1D-, 2D- и 3D-проводников с параболической дисперсией качественно показано на рис. 8.

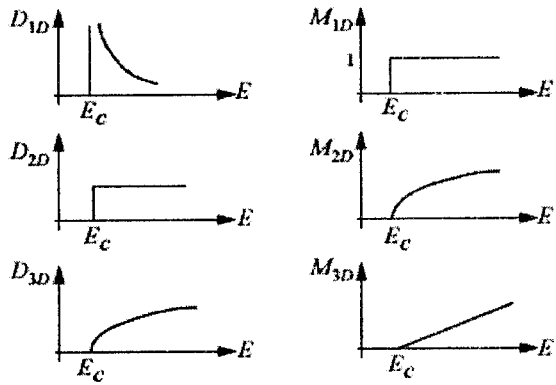


Рис. 8. Сравнительное поведение плотности состояний $D(E)$ и числа мод $M(E)$ для 1D-, 2D- и 3D-проводников с параболической дисперсией (65).

Плотность состояний для 1D проводника обратно пропорциональна $\sqrt{E}$, для 2D проводника постоянна, для 3D – прямо пропорциональна $\sqrt{E}$. Что касается мод проводимости, то 1D проводник характеризуется единственной модой проводимости, равной функции Хевисайда, что позволяет определить ее калибровку для использования ее в формулах (69) – (74). Число же мод 2D проводника прямо пропорциональна $\sqrt{E}$, а 3D проводника растет с энергией линейно.

Заключая этот раздел, подведем итоги:

1. Плотность состояний $D(E)$ нужна для вычисления плотности носителей тока.
2. Число мод $M(E)$ используется для вычисления тока.
3. Число мод $M(E)$ пропорционально средней скорости носителей тока в направлении их движения, помноженной на плотность состояний $D(E)$.
4. Число мод $M(E)$ зависит от зонной структуры и размерности проводника.

Хотя выше в качестве примера и рассмотрены проводники с параболической дисперсией, формулы (63) носят общий характер. Мы воспользуемся ими в главе 3, посвященной графену, дисперсия у которого линейная [21]. Для произвольной дисперсионной зависимости $E(k)$ разработаны численные методы расчета числа мод проводимости [22].

1.3.3. Коэффициент прохождения

До сих пор мы рассматривали баллистический транспорт. Перейдем к рассмотрению диффузионного транспорта, когда $L \gg \lambda$. Электроны, инжектируемые контактами 1 и 2 , подвержены случайным блужданиям. Некоторые из них заканчиваются на инжектирующем контакте, а иные – на другом контакте. Если на контакт 2 подан положительный потенциал, то большая часть блужданий закончится на этом контакте.

Ключевым параметром в модели ЛДЛ есть величина $\gamma \pi D(E)/2$, которая в случае баллистического транспорта оказывается равной числу мод проводимости $M(E)$. Уширение γ и время пролета τ связаны между собой: $\gamma = \hbar/\tau$. В режиме диффузионного транспорта следует ожидать увеличения времени τ, что влечет за собой уменьшения величины $\gamma \pi D(E)/2$. Сейчас покажем, что в случае диффузионного транспорта $\gamma \pi D(E)/2 = M(E) \cdot T(E)$, где $T(E) \le 1$ получил название коэффициента прохождения (transmission coefficient).

В режиме баллистического транспорта электроны инжектируются в проводник под разными углами, в результате чего время пролета описывается некой функцией распределения. Величину γ мы оценивали через среднее время пролета

$$\gamma(E) = \frac{\hbar}{\langle \tau(E) \rangle}, \tag{75}$$

где

$$\langle \tau(E) \rangle = \frac{L}{\langle v_x^+(E) \rangle} = \frac{L}{v(E)\langle \cos\vartheta \rangle} = \frac{L}{v(E)(2/\pi)}. \tag{76}$$

Наша задача сейчас оценить $\langle \tau(E) \rangle$ в режиме диффузионного транспорта.

Воспользуемся первым законом диффузии Фика применительно к потоку электронов в 2D проводнике:

$$J = q\bar{D}\frac{dn_S}{dx}, \tag{77}$$

где плотность тока J пропорциональна градиенту плотности электронов с коэффициентом диффузии $\bar{D}$ в роли коэффициента пропорциональности.

Пусть электроны в проводник с подавляющим преимуществом поставляет только контакт 1, т.е. $f_1 \approx 1$, а $f_2 \approx 0$. На левом конце проводника $(x = 0)$ есть конечная плотность электронов $\Delta n_S(0)$ (рис. 9).

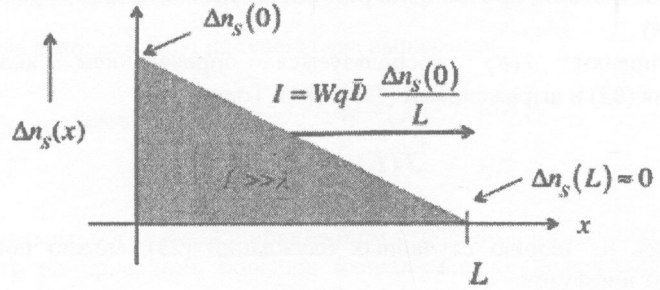


Рис. 9. К выводу времени пролета τ_D в диффузионном режиме.

Поскольку процессы рекомбинации/рождения носителей тока не учитываются, то профиль плотности электронов на всем протяжении проводника будет линейным с практически нулевым значением плотности на правом конце проводника: $\Delta_S(L) \to 0$ за счет $f_2 \approx 0$. Время пролета, согласно (54), есть отношение накопленного в проводнике заряда к току, а именно:

$$\tau_D = \frac{qN}{I} = \frac{Wq\Delta n_S(0)L/2}{Wq\bar{D}\Delta n_S(0)/L} = \frac{L^2}{2\bar{D}}, \tag{78}$$

где накопленный в проводнике заряд qN есть половина площади прямоугольника (рис. 9), построенного для 2D проводника на его длине L и ширине W, а ток $I = J \cdot W$ и, согласно (77), $dn_S/dx = \Delta n_S(0)/L$. Таким образом, время пролета в диффузионном режиме

$$\tau_D = \frac{L^2}{2\bar{D}}, \tag{79}$$

тогда как в баллистическом режиме

$$\tau_B = \frac{L}{\langle v_x^+ \rangle}. \tag{80}$$

Собирая все вместе, имеем:

$$\gamma(E)\pi\frac{D(E)}{2} = \frac{\hbar}{\tau_D}\pi\frac{D(E)}{2} = \frac{\hbar}{\tau_B}\pi\frac{D(E)}{2}\frac{\tau_B}{\tau_D} \equiv M(E)\cdot T(E), \tag{81}$$

где коэффициент прохождения есть отношение двух времен пролета – в баллистическом режиме и в диффузионном:

$$T(E) = \frac{\tau_B}{\tau_D}. \tag{82}$$

Другими словами, при наличии рассеяния число мод $M(E)$ нужно заменить на $M(E)\cdot T(E)$.

Для оценки $T(E)$ воспользуемся определением коэффициента прохождения (82) и выражениями (79) и (80). Тогда

$$T(E)=\frac{2\bar{D}}{L\left\langle v_x^+\right\rangle}. \tag{83}$$

Опираясь на теорию случайных блужданий [23], можно показать, что коэффициент диффузии

$$\bar{D}=\frac{\left\langle v_x^+\right\rangle\lambda}{2}, \tag{84}$$

что вместе с (83) дает простое выражение для коэффициента прохождения

$$T(E)=\frac{\lambda}{L}\ll 1. \tag{85}$$

Как и ожидалось, число мод проводимости в диффузионном режиме $M(E)\cdot T(E)$ резко уменьшается.

Выражение для коэффициента прохождения (85) верно для диффузионного режима. Можно показать [24], что в общем случае коэффициент прохождения

$$T(E)=\frac{\lambda(E)}{\lambda(E)+L}. \tag{86}$$

Это выражение справедливо как для диффузионного режима $(L>>\lambda)$, так и для баллистического $(L<<\lambda)$, приближаясь к единице, так и для промежуточных случаев.

Подводя итоги, в общем случае имеет место соотношение

$$\gamma(E)\pi\frac{D(E)}{2}=M(E)\cdot T(E), \tag{87}$$

в котором число мод определяется общими выражениями (63), а в отношении коэффициента прохождения в реальных экспериментах обычно рассматривают три режима:

$$\begin{gathered}\text{Диффузионный: } L>>\lambda;T=\lambda/L<<1,\\ \text{Баллистический: } L<<\lambda;T\to 1,\\ \text{Квази-баллистический: } L\approx\lambda;T<1.\end{gathered} \tag{88}$$

Примечательно, что сравнительно простая транспортная модель ЛДЛ применима ко всем трем режимам.

1.3.4. Режим линейного отклика

Для тока в модели ЛДЛ получены два выражения:

$$I=\frac{2q}{h}\int\gamma(E)\pi\frac{D(E)}{2}\left(f_1-f_2\right)dE,$$
$$I=\frac{2q}{h}\int T(E)M(E)\left(f_1-f_2\right)dE. \qquad (89)$$

В них не учитывается важный в прикладном отношении низковольтный режим. Если подать на проводник большое напряжение, то вследствие появления множества неупругих столкновений в режиме горячих электронов нарушатся принятые нами допущения, которые привели нас к модели независимых мод проводимости. Мы далее продолжим строить модель ЛДЛ для режима линейного отклика, что позволит упростить уравнение (89).

В режиме линейного отклика, иначе линейного транспорта или почти равновесного транспорта (near-equilibrium transport), фермиевские функции контактов (20) и (21) и равновесная функция Ферми

$$f_0\left(E\right)=\frac{1}{e^{(E-E_{F0})/kT}+1} \qquad (90)$$

находятся в соотношении

$$f_0(E)\approx f_1(E)>f_2(E)\approx f_0(E), \qquad (91)$$

при этом подаваемое на концы проводника напряжение

$$V=\Delta E_F/q=(E_{F1}-E_{F2})/q \qquad (92)$$

незначительно. Пользуясь малостью ΔE_F, фермиевскую функцию контакта 2 можно разложить в ряд Тейлора и ограничиться первой степенью, а именно:

$$f_2=f_1+\frac{\partial f_1}{\partial E_F}\Delta E_F=f_1+\frac{\partial f_1}{\partial E_F}qV, \qquad (93)$$

а учитывая (30) и (91), окончательно для разности фермиевских функций контактов в режиме линейного отклика имеем:

$$f_1-f_2=(-\frac{\partial f_0}{\partial E})qV. \qquad (94)$$

Окончательно для тока (89) и проводимости в режиме линейного отклика модели ЛДЛ имеем:

$$I=\left[\frac{2q^2}{h}\int T(E)M(E)\left(-\frac{\partial f_0}{\partial E}\right)dE\right]V=GV, \qquad (95)$$

$$G=\frac{2q^2}{h}\int T(E)M(E)\left(-\frac{\partial f_0}{\partial E}\right)dE, \qquad (96)$$

где последнее уравнение известно как уравнение Ландауэра для проводимости.

Теперь мы имеем выражение для проводимости, в котором проводимость связана со свойствами материала проводника. Эти выражения справедливы для 1D-, 2D- и 3D-проводников как в баллистическом режиме, так и в диффузионном, если пользоваться соответствующими выражениями для числа мод (63).

1.3.5. Транспорт в массивных проводниках

Модель ЛДЛ пригодна как для коротких, наноразмерных, так и длинных, массивных проводников. Когда канал проводимости достаточно длинный, роль контактов пренебрежимо мала и проводимость определяется свойствами материала проводника. Выражения для тока в массивных проводниках в режиме линейного отклика можно получить из любого из общих уравнений (89). Воспользуемся первым их них.

Перепишем разность фермиевских функций (94) в виде

$$f_1 - f_2 \approx -\frac{\partial f_0}{\partial E_{F0}} \Delta E_F . \tag{97}$$

Тогда для тока имеем:

$$I = \frac{2q}{h} \int \left[\gamma(E) \pi \frac{D(E)}{2} \left(-\frac{\partial f_0}{\partial E_{F0}} \right) \Delta E_F \right] dE . \tag{98}$$

В массивных проводниках всегда реализуется диффузионный режим, так что

$$\gamma(E) = \frac{\hbar}{\tau(E)} = \frac{\hbar}{L^2 / 2D(E)} . \tag{99}$$

Пусть мы рассматриваем 2D проводник. Для него, согласно (70),

$$D(E) = WLD_{2D}(E) . \tag{100}$$

Подставляя (99) и (100) в (98), для плотности тока получаем

$$J_x = I / W = \left[\int qD(E) D_{2D}(E) \left(-\frac{\partial f_0}{\partial E_{F0}} \right) dE \right] \frac{\Delta E_F}{L} . \tag{101}$$

Поскольку массивный проводник находится в режиме линейного отклика и почти равновесного транспорта по всей длине проводника, мы вправе предположить линейное падение электрохимического потенциала от контакта 1 до контакта 2 так, что $\Delta E_F / L$ есть градиент фермиевской функции dE_F / dx, и (101) можно переписать в виде

$$J_x = \frac{\sigma}{q} \frac{dE_F}{dx}, \tag{102}$$

где удельная проводимость

$$\sigma = \int q^2 D(E) D_{2D}(E) \left(-\frac{\partial f_0}{\partial E} \right) dE . \qquad (103)$$

Оба уравнения (102) и (103) представляют собой стандартный результат, получаемый обычно в термодинамике необратимых процессов или же путем решения транспортного уравнения Больцмана [7]. Мы же получили эти стандартные результаты для массивных проводников, предположив в модели ЛДЛ лишь то, что длина проводника намного превышает длину свободного пробега.

Уравнение (102) можно переписать иначе. Из теории полупроводников для невырожденных *n*-проводников известно [20], что

$$n_S = N_{2D} \exp\left[\left(E_F - E_C \right) kT \right],$$

$$N_{2D} = \frac{m^* kT}{\pi \hbar^2},$$

$$E_F = E_C + kT \ln \frac{N_{2D}}{n_S},$$

$$\sigma = nq\mu .$$

Для этого случая уравнение для плотности тока (102) можно переписать в виде:

$$J_x = n_S q \mu \mathrm{E}_x + q \bar{D} \frac{dn_S}{dx}, \qquad (104)$$

где

$$\frac{\bar{D}}{\mu} = \frac{kT}{q} \qquad (105)$$

есть соотношение Эйнштейна между коэффициентом диффузии электронов и их подвижностью.

Уравнение (104) дрейфа и диффузии хорошо известно в физике полупроводников и обычно служит начальным приближением при анализе полупроводниковых устройств [20]. Оно справедливо для невырожденных проводников в предположении постоянства температуры вдоль проводника и, как мы видели, предполагает почти равновесный транспорт.

До сих пор речь шла только о транспорте электронов. Если же речь зайдет о транспорте «дырок», то в модели ЛДЛ нужно лишь помнить, что «дырки» описываются своим электрохимическим потенциалом $E_F^{(p)}$, отличным от

электрохимического потенциала электронов E_F, в обозначении которого мы просто опускали верхний индекс (n) за очевидностью его. Это связано с тем, что в валентной зоне и в зоне проводимости носители тока отделены друг от друга и в равновесии находятся лишь носители тока в каждой из зон. Процессы рекомбинации/рождения носителей тока, которые связывают заселенности в этих зонах, идут намного медленнее, чем процессы рассеяния, устанавливающие равновесие в каждой из зон.

Для электронов в зоне проводимости имеем уравнения (102), (103) и (91). Уравнения (102) и (103) выводились без учета природы носителей тока. Поэтому для «дырок» в валентной зоне эти уравнения нужно просто переписать с «дырочным» электрохимическим потенциалом, а именно:

$$J_x^{(p)} = \frac{\sigma}{q}\frac{dE_F^{(p)}}{dx},$$

$$\sigma_p = \int q^2 D_p(E) D_{2D}(E) \left(-\frac{\partial f_0^{(p)}}{\partial E} \right) dE,$$

$$f_0^p(E) = \frac{1}{e^{(E-E_{F0}^p)/kT}+1}.$$

Полный ток есть сумма токов в каждой из зон.

Подведем итоги. Уравнение Ландауэра для проводимости (96) описывает электронный транспорт в проводнике с общих позиций. Проводимость пропорциональна фундаментальным константам ($G \approx 2q^2/h$), которые определяют квант проводимости, ассоциируемый с контактами, в чем мы убедимся далее. Проводимость зависит от числа мод проводимости $M(E)$ и от коэффициента прохождения $T(E)$, представляющего собой вероятность того, что электрон с энергией E, инжектированный контактом 1, достигнет контакта 2. Полную проводимость находим путем интегрирования вкладов от всех мод проводимости. Уравнение справедливо для 1D-, 2D- и 3D-проводников, нужно лишь просто корректно вычислить число мод проводимости по (63). Справедливо оно как для баллистических нанорезисторов, так и для массивных проводников и корректно описывает промежуточные ситуации.

Далее мы займемся вычислением проводимости для 1D-, 2D- и 3D-проводников, работающих в режиме баллистическом или диффузионном, как, впрочем, и в любом промежуточном. Нужно также помнить, что электронный транспорт рассматривается нами в изотермических условиях. Термоэлектрические явления и термоэлектрики с позиций транспортной модели ЛДЛ будут рассмотрены в главе 2.

1.3.6. Баллистический и диффузионный режимы проводимости

Рассмотрим 1D-, 2D- и 3D-проводники, начиная с очень коротких баллистических резисторов, затем перейдем к обычным резисторам, работающим в диффузионном режиме, и рассмотрим также промежуточные режимы.

В 1D проводнике («нанопроволока») длины L электроны могут перемещаться лишь в одном измерении, в 2D проводнике («наноляма») ширины W – в двух измерениях, в 3D проводнике с площадью поперечного сечения A все три измерения доступны электронам. Согласно общепринятой договоренности [20], сопротивление R и его удельное значение ρ для проводников всех трех размерностей запишем следующим образом:

$$R_{1D} = \rho_{1D} L; \quad \rho_{1D} = \frac{1}{n_L q \mu}, \tag{106}$$

$$R_{2D} = \rho_{2D} \frac{L}{W}; \quad \rho_{2D} = \frac{1}{n_S q \mu}, \tag{107}$$

$$R_{3D} = \rho_{3D} \frac{L}{A}; \quad \rho_{3D} = \frac{1}{n q \mu}, \tag{108}$$

где μ – подвижность электронов, а n_L, n_S и n есть плотность электронов для проводников размерности 1D, 2D и 3D, соответственно, удельное сопротивление которых измеряется в разных единицах в зависимости от размерности проводника. Подобная запись сопротивления для проводников разной размерности физически вполне корректна: во всех случаях сопротивление пропорционально длине проводника, а для 2D и 3D проводников еще и обратно пропорционально, соответственно, ширине W и площади A поперечного сечения проводника, поскольку увеличение W и A физически эквивалентно параллельному соединению мод проводников. И вместе с тем мы увидим, что уравнения (106) – (108) не всегда корректны: удивительное встречается даже в таких, казалось бы, простых ситуациях, какими нам представляются проводники разной размерности.

Уравнение Ландауэра для проводимости

$$G = \frac{2q^2}{h} \int T(E) M(E) \left(-\frac{\partial f_0}{\partial E} \right) dE, \tag{109}$$

послужит нам отправным пунктом; оно справедливо для проводников любой размерности, если только используются корректные выражения для числа мод проводимости $M(E)$.

Рассмотрим 2D проводник, проводимость которого запишем в виде

$$G = \frac{1}{\rho_{2D}} \frac{W}{L} = \sigma_S \frac{W}{L}. \tag{110}$$

Мы позже убедимся в том, что у массивных и длинных 2D проводников, работающих в диффузионном режиме, удельная поверхностная проводимость σ_S не зависит от ширины W и длины L проводника, для коротких же проводников σ_S становится зависящей от длины проводника, а для узких проводников проводимость увеличивается с ростом их ширины ступенчато.

Основное внимание мы уделим 2D проводникам, хотя похожие рассуждения и выкладки могут быть проведены и для 1D и 3D проводников, для которых мы приведем только окончательные результаты. В отношении формулы Ландауэра (109) напомним, что множитель $(-\partial f_0/\partial E)$ происходит из разложения в ряд Тейлора разности электрохимических потенциалов на контактах проводника $f_1 - f_2$ в предположении, что температура обоих контактов одинакова. Термоэлектрические явления в модели ЛДЛ рассматриваются в главе 2.

1.3.6.1. 2D Баллистические резисторы

У баллистических 2D резисторов коэффициент прохождения $T(E)=1$, а число мод проводимости в (109) $M(E)=WM_{2D}(E)$, а общие выражения для мод проводимости для любых дисперсионных соотношений даются формулами (63), а для параболических дисперсий – формулами (72) – (74). Рассмотрим еще раз важный множитель $(-\partial f_0/\partial E)$ в уравнении Ландауэра (109), который получил название фермиевского окна проводимости.

На рис. 10 качественно показан график функции Ферми $f_0(E)$ и функции $(-\partial f_0/\partial E)$ в зависимости от энергии E. Более подробное и количественное обсуждение фермиевского окна дано выше в § 1.2.1.

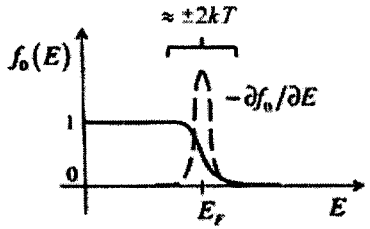


Рис. 10. Фермиевское окно проводимости.

Мы видим, что функция $(-\partial f_0/\partial E)$ существенна лишь в окрестности $\sim \pm 2kT$ вокруг энергии Ферми E_F. В § 1.2.1 показано, что площадь под этой кривой равна единице, так что для низких температур фермиевское окно проводимости можно записать через дельта-функцию Кронекера:

$$-\frac{\partial f_0}{\partial E} \approx \delta(E - E_F). \qquad (111)$$

Используя (111) и $T(E)=1$, из уравнения Ландауэра (109) получаем выражение для баллистической проводимости

$$G^{ball}=\frac{2q^2}{h}M(E_F), \tag{112}$$

которое в обращенной форме выше лишь констатировалось и которое является общим и справедливым для резисторов любой размерности. Если число мод невелико, его можно вычислить из экспериментальных измерений, тогда проводимость и соответственно сопротивление не может принимать произвольные значения, а квантуется согласно

$$R^{ball}=\frac{h}{2q^2}\frac{1}{M(E_F)}=\frac{12.9}{M(E_F)}[кОм]. \tag{113}$$

Обратим внимание, что сопротивление в баллистическом режиме не зависит от длины проводника, как и ожидалось для режима баллистического транспорта.

Тот факт, что сопротивление квантуется, надежно установлено экспериментально [25, 26]. В экспериментах ширина резистора W контролировалась электростатически напряжением на затворе V_G. Балистический режим транспорта обеспечивался низкой температурой. По мере роста ширины резистора проводимость растет ступенчато (рис. 11) в соответствии с уравнением (112).

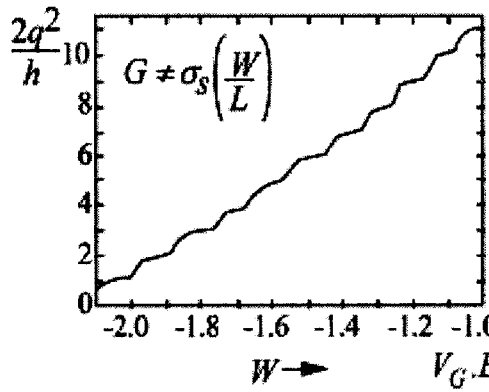


Рис. 11. Квантование проводимости электронного газа в интерфейсе $AlAs/Al_xGa_{1-x}As$ [25, 26].

Отметим, что ступенчатый характер проводимости в современных экспериментах наблюдается даже при комнатных температурах благодаря проведению измерений на нанорезисторах.

1.3.6.2. Широкие 2D баллистические резисторы при $T = 0°\ K$

Ширина W резистора может стать столь значительной, что на ней укладывается большое число де бройлевских полуволн и число мод измерить экспериментально становится затруднительно. В этом случае, например, для параболической дисперсии из уравнений (64) и (73) имеем:

$$M(E_F) = WM_{2D}(E_F) = g_v W \frac{\sqrt{2m^*(E_F - E_C)}}{\pi\hbar}. \qquad (114)$$

Число мод M_{2D} удобно связать с поверхностной плотностью электронов n_S, которая обычно известна из экспериментов. Все состояния с волновым вектором $k < k_F$ заняты при $T = 0° K$. Для поверхностной плотности имеем:

$$n_S = g_v \frac{\pi k_F^2}{(2\pi)^2} \times 2 = g_v \frac{k_F^2}{2\pi}, \qquad (115)$$

где двойка учитывает спиновое вырождение, а g_v – долинное вырождение. Отсюда находим k_F через поверхностную плотность и подставляем в (67), так что окончательно

$$M_{2D}(E_F) = \sqrt{2g_v \frac{n_S}{\pi}}. \qquad (116)$$

Уравнение (116) связывает число мод при энергии, равной энергии Ферми, с поверхностной плотностью электронов 2D баллистического проводника. Обратим внимание на то, что этот результат не предполагает задание какой-либо конкретной зонной структуры, важно лишь, чтобы она была изотропной. А вот увязка k_F и E_F уже требует задания конкретной дисперсии. Например, для параболической дисперсии волновой вектор k_F находим из

$$\frac{\hbar^2 k_F^2}{2m^*} = E_F - E_C. \qquad (117)$$

1.3.6.3. Широкие 2D баллистические резисторы при T > 0° K

Приближение (111) хорошо работает при низких температурах. При комнатных температурах и выше интеграл Ландауэра (109) нужно добросовестно вычислять. Пользуясь записью фермиевской функции (14), для баллистической 2D проводимости ($T = 1$) имеем:

$$G_{2D}^{ball} = \frac{2q^2}{h} \int W M_{2D}(E) \left(-\frac{\partial}{\partial E}\right) \frac{1}{\exp\left(E - E_F/kT\right) + 1} dE. \qquad (118)$$

Подобные интегралы часто встречаются в физике полупроводников. Разберемся в качестве примера с этим интегралом.

Фермиевская функция обладает полезным свойством (30), а именно:

$$\left(-\frac{\partial}{\partial E}\right)=\left(+\frac{\partial}{\partial E_F}\right), \tag{119}$$

которое позволяет производную в уравнении (118) вынести из под знака интеграла. Тогда вместо (118) с учетом (119) получим:

$$G_{2D}^{ball}=\frac{2q^2}{h}\frac{g_v W\sqrt{2m^*}}{\pi\hbar}\left(\frac{\partial}{\partial E_F}\right)\int_0^\infty\frac{\sqrt{E-E_C}}{\exp\left(E-E_F/kT\right)+1}dE. \tag{120}$$

Далее введем новые переменные:

$$\eta\equiv(E-E_C)/kT;\ \ \eta_F\equiv(E_F-E_C)/kT, \tag{121}$$

что позволяет интеграл (120) преобразовать к виду

$$G_{2D}^{ball}=\frac{2q^2}{h}\frac{g_v W\sqrt{2m^*kT}}{\pi\hbar}\left(\frac{\partial}{\partial\eta_F}\right)\int_0^\infty\frac{\sqrt{\eta}}{\exp\left(\eta-\eta_F\right)+1}d\eta. \tag{122}$$

Интегралы этого типа не берутся аналитически. Они часто встречаются в физике полупроводников и получили название интегралов Ферми – Дирака. В нашем случае интеграл (122) пропорционален интегралу

$$\Im_{1/2}(\eta_F)=\frac{2}{\sqrt{\pi}}\int_0^\infty\frac{\eta^{1/2}}{\exp\left(\eta-\eta_F\right)+1}d\eta, \tag{123}$$

который называется интегралом Ферми – Дирака порядка одна вторая.

При дифференцировании интеграла Ферми – Дирака по его параметру порядок интеграла понижается на единицу:

$$\frac{d\Im_j(\eta_F)}{d\eta_F}=\Im_{j-1}(\eta_F). \tag{124}$$

Воспользуемся этим свойством и перепишем выражение (122) следующим образом:

$$G_{2D}^{ball}=\frac{2q^2}{h}\frac{g_v W\sqrt{2m^*kT}}{\pi\hbar}\frac{\sqrt{\pi}}{2}\Im_{-1/2}(\eta_F)=\frac{2q^2}{h}\left\langle WM_{2D}\right\rangle, \tag{125}$$

где

$$\left\langle M\right\rangle=\left\langle WM_{2D}\right\rangle=\frac{\sqrt{\pi}}{2}WM_{2D}(kT)\Im_{-1/2}(\eta_F), \tag{126}$$

а $WM_{2D}(kT)$ есть $WM_{2D}(E-E_C)$, вычисленное при энергии $E-E_C=kT$. Сравнивая (125) с (112), мы видим, что проводимость 2D баллистических резисторов при конечных температурах имеет тот же вид, что и при $T=0°K$: нужно лишь заменить число мод $M(E_F)$ на $\langle M\rangle$ по (126). Величина $\langle M\rangle$ есть число мод проводимости 2D баллистического резистора при конечных температурах в фермиевском окне проводимости $-\partial f_0/\partial E$.

При обработке экспериментальных данных для 2D проводников обычно легче определить поверхностную плотность электронов n_S, чем энергию уровня Ферми E_F, однако, они связаны между собой, и зная n_S, можно найти E_F. Для параболической дисперсии эта связь дается следующим выражением:

$$n_S=\int_0^\infty D_{2D}(E)f_0(E)dE=g_v\frac{m^*kT}{\pi\hbar^2}\Im_0(\eta_F)=N_{2D}\Im_0(\eta_F). \qquad (127)$$

Мы рассмотрели только один пример 2D резистора с параболической дисперсией. Аналогичные интегралы Ферми – Дирака разного порядка часто встречаются в задачах о проводниках разной размерности и различной зонной структуры. Полезно рассмотреть интегралы Ферми – Дирака более подробно.

1.3.6.4. Интегралы Ферми – Дирака

Интеграл Ферми – Дирака порядка j

$$\Im_j(\eta_F)=\frac{1}{\Gamma(j+1)}\int_0^\infty\frac{\eta^j}{\exp(\eta-\eta_F)+1}d\eta, \qquad (128)$$

где гамма-функция при n целом положительном

$$\Gamma(n)=(n-1)!; \qquad (129)$$

при этом

$$\Gamma(p+1)=p\Gamma(p) \qquad (130)$$

и

$$\Gamma(1/2)=\sqrt{\pi}. \qquad (131)$$

Важным свойством интеграла Ферми – Дирака является понижение его порядка при дифференцировании интеграла по его параметру (124). Например, известно аналитическое выражение для интеграла нулевого порядка [27]:

$$\Im_0(\eta_F)=\log(1+e^{\eta_F}). \qquad (132)$$

Тогда для интеграла минус первого порядка имеем:

$$\Im_{-1}(\eta_F)=\frac{d\Im_0(\eta_F)}{d\eta_F}=\frac{1}{e^{-\eta_F}+1}. \qquad (133)$$

Продолжая таким же образом далее, можно получить аналитические полиномиальные выражения для любых целых отрицательных порядков [27 – 29].

Для невырожденных полупроводников

$$\eta_F = (E_F - E_C)/kT \ll 0, \tag{134}$$

в связи с чем для них интегралы Ферми – Дирака любого порядка сводятся к экспонентам:

$$\Im_j(\eta_F) \to e^{\eta_F}. \quad (\eta_F \ll 0) \tag{135}$$

Роль интегралов Ферми – Дирака в физике полупроводников, производные интегралов, асимптотические и приближенные выражения для них, методы и алгоритмы высокоточных и приближенных методов их расчета можно найти в [29].

1.3.6.5. 2D диффузионные резисторы

Перейдем от рассмотрения баллистического транспорта с коэффициентом прохождения всегда равным единице, $T(E) = 1$, к диффузионному транспорту с $T(E) = \lambda(E)/L$. Модель ЛДЛ успешно применяется к проводникам любой размерности, однако, мы по-прежнему сосредоточимся на 2D резисторах, а позже приведем окончательные результаты для 1D и 3D резисторов.

Уравнение Ландауэра (109) перепишем с учетом (63) для 2D резистора в диффузионном режиме, а именно:

$$G_{2D}^{diff} = \frac{2q^2}{h}\frac{W}{L}\int \lambda(E) M_{2D}(E)\left(-\frac{\partial f_0}{\partial E}\right)dE. \tag{136}$$

Появление множителя W/L согласуется с традиционным определением (107) сопротивления 2D проводника. В модели же ЛДЛ пропорциональность проводимости ширине резистора W объясняется тем, что число мод проводимости 2D проводника должно быть пропорционально его ширине, а обратная пропорциональность длине проводника L вызвана диффузионным режимом.

1.3.6.6. Широкие 2D диффузионные резисторы при $T = 0^o$ K

Выражение (136) для низких температур существенно упрощается за счет сужения фермиевского окна проводимости (111) и сразу можно записать, что

$$G_{2D}^{diff} = \frac{2q^2}{h}\frac{W}{L} M_{2D}(E_F)\lambda(E_F) = \frac{\lambda(E_F)}{L} G_{2D}^{ball}, \tag{137}$$

где баллистическая проводимость дается выражением (112). Если определена длина свободного пробега λ, то проводимость в диффузионном режиме можно вычислить через ее значение в баллистическом режиме.

Мы уже обсудили проводимость раздельно в баллистическом и в диффузионном режимах при низких температурах. Однако, можно рассмотреть весь спектр режимов при переходе от баллистического режима к диффузионному. Воспользуемся общим выражением для коэффициента прохождения (86), а именно:

$$T(E)=\frac{\lambda(E)}{\lambda(E)+L}, \tag{138}$$

тогда

$$G_{2D}=\frac{2q^2}{h}W M_{2D}(E_F)\frac{\lambda(E_F)}{\lambda(E_F)+L}=\frac{\lambda(E_F)}{\lambda(E_F)+L}G_{2D}^{ball}, \tag{139}$$

а для сопротивления имеем:

$$R_{2D}=\left(1+\frac{L}{\lambda(E_F)}\right)R_{2D}^{ball}. \tag{140}$$

Полученный результат свидетельствует, что при низких температурах сопротивление 2D резистора в диффузионном режиме пропорционально длине проводника, а в баллистическом режиме от длины проводника не зависит.

1.3.6.7. Широкие 2D диффузионные резисторы при $T > 0°$ K

При конечных температурах интеграл (136) нужно вычислять. Поступим следующим образом. Умножая и деля его на

$$\langle M_{2D}\rangle \equiv \int M_{2D}(E)\left(-\frac{\partial f_0}{\partial E}\right)dE, \tag{141}$$

переписываем (136) следующим образом:

$$G_{2D}^{diff}=\frac{2q^2}{h}\langle WM_{2D}\rangle\frac{\langle\langle\lambda\rangle\rangle}{L}=\frac{\langle\langle\lambda\rangle\rangle}{L}G_{2D}^{ball}, \tag{142}$$

где $\langle WM_{2D}\rangle$ дается выражением (126), а усредненное значение средней длины свободного пробега определяется как

$$\langle\langle\lambda\rangle\rangle=\frac{\int\lambda(E)M_{2D}(E)\left(-\frac{\partial f_0}{\partial E}\right)dE}{\int M_{2D}(E)\left(-\frac{\partial f_0}{\partial E}\right)dE}=\frac{\langle M\lambda\rangle}{\langle M\rangle}. \tag{143}$$

Уравнение (142) по форме такое же как и уравнение (137). При низких температурах мы $\langle M \rangle$ заменяем на $M(E_F)$, а $\langle\langle \lambda \rangle\rangle$ на $\lambda(E_F)$. Одиночные и двойные угловые скобки указывают лишь на то, как производится усреднение.

Чтобы реально оценить усредненное значение средней длины свободного пробега, нужно задаться дисперсионным соотношением и выражением для $\lambda(E)$. В случае обычных и наиболее распространенных механизмов рассеяния, которые достаточно подробно для наших целей обсуждаются в главе 4, простейший путь записать выражение для $\lambda(E)$ это воспользоваться степенным законом в форме [30]

$$\lambda(E) = \lambda_0 \left(\frac{E - E_C}{kT} \right)^r , \tag{144}$$

где значение показателя степени r определяется выбором того или иного механизма рассеяния, а константа λ_0 в типичных случаях еще и температурозависимая. Например, при рассеянии электронов на акустических фононах в 3D проводнике $r = 0$, а на заряженных примесях $r = 2$.

Оба интеграла в (143) выражаются через интегралы Ферми – Дирака следующим образом:

$$\langle\langle \lambda \rangle\rangle = \lambda_0 \times \left(\frac{\Gamma(r+3/2)}{\Gamma(3/2)} \right) \times \left(\frac{\Im_{r-1/2}(\eta_F)}{\Im_{-1/2}(\eta_F)} \right). \tag{145}$$

Если $r = 0$, средняя длина свободного пробега не зависит от энергии и $\langle\langle \lambda \rangle\rangle = \lambda_0$.

Мы не будем рассматривать весь спектр перехода от баллистического режима к диффузионному; это выведет нас далеко за пределы главы 1. Заметим лишь, что в случае не зависящей от энергии средней длине свободного пробега сопротивление резистора

$$R = \left(1 + \frac{L}{\lambda_0} \right) R^{ball} , \tag{146}$$

что очень похоже на уже известный результат (140).

Далее мы обсудим с позиций транспортной модели ЛДЛ некоторые общие вопросы электронного транспорта такие, как понятие подвижности носителей тока, способы записи выражений для поверхностной 2D проводимости, диссипацию мощности и падение напряжения в баллистическом резисторе, а также приведем, наконец, сводку основных результатов для 1D и 3D проводников.

1.3.7. О понятии подвижности электронов

Традиционно обсуждение сопротивления проводников начинается с определений (106) – (108), которые, однако, не применимы к проводникам, работающим в баллистическом или квазибалистическом режимах. В этих

режимах не совсем ясно, как оценить подвижность носителей тока. Транспортная модель ЛДЛ начинается с уравнения Ландауэра (109) и пригодна для любых транспортных режимов, от баллистического до диффузионного и во всем спектре режимов между ними. В рамках модели ЛДЛ вообще нет нужды обращаться к понятию подвижности. Вместе с тем понятие подвижности и поныне широко используется несмотря на то, что в отдельных случаях оно приводит к недоразумениям. Так уравнения (106) – (108) утверждают, что проводимость пропорциональна произведению плотности неких электронов и подвижности, тогда как согласно уравнению Ландауэра (109) проводимость обеспечивается только теми электронами, энергия которых попадает в фермиевское окно проводимости $(-\partial f_0/\partial E)$. Для полупроводников *n*-типа это могут быть все электроны в зоне проводимости (невырожденные полупроводники), а может быть только малая их доля (вырожденные полупроводники). Поскольку понятие подвижности и поныне широко используется, полезно обсудить его в рамках концепции ЛДЛ.

Лучше всего опять начать с уравнения Ландауэра (109) и приравнять его проводимости 2D проводника по уравнению (107), а именно:

$$G_{2D} = \frac{2q^2}{h}\int T(E)M(E)\left(-\frac{\partial f_0}{\partial E}\right)dE \equiv n_S q\mu\frac{W}{L}, \tag{147}$$

откуда кажущаяся подвижность для 2D резистора

$$\mu_{2D}^{app} \equiv \frac{2q}{h}\frac{1}{n_S}\int T(E)LM_{2D}(E)\left(-\frac{\partial f_0}{\partial E}\right)dE \tag{148}$$

берется нами за определение подвижности в отличие от определения подвижности по модели Друде [31]

$$\mu = \frac{q\tau}{m^*}, \tag{149}$$

где τ – среднее время рассеяния, иначе говоря, время импульсной релаксации. Мы называем подвижность по (148) кажущейся, поскольку определение (148) пригодно как для баллистической подвижности, так и для диффузионной. Например, положив время прохождения $T(E)=1$, для баллистической подвижности имеем:

$$\mu_{2D}^{ball} \equiv \frac{2q}{h}\frac{1}{n_S}\int LM_{2D}(E)\left(-\frac{\partial f_0}{\partial E}\right)dE; \tag{150}$$

аналогично, выбрав $T(E)=\lambda(E)/L$, получим выражение для диффузионной подвижности

$$\mu_{2D}^{diff} \equiv \frac{2q}{h}\frac{1}{n_S}\int \lambda(E)M_{2D}(E)\left(-\frac{\partial f_0}{\partial E}\right)dE. \tag{151}$$

Концепция баллистической подвижности была введена в обращение Шуром [32] и оказалась полезной при анализе различных устройств [33]. Сравнивая выражения (150) и (151), мы видим, что диффузионная подвижность отличается от баллистической только заменой средней длины свободного пробега просто на длину баллистического резистора. Подобная замена выглядит вполне физически приемлемой. Вспомним, что в контактах равновесие быстро достигается за счет интенсивных процессов рассеяния. Электрон, инжектированный в баллистический канал истоком, участвовал в рассеянии в контакте с истоком, а затем следующий раз участвовал в рассеянии лишь в контакте со стоком. Расстояние между этими двумя событиями есть как раз длина баллистического резистора, которая и играет роль средней длины свободного пробега в диффузионном канале.

Проблемы в наноэлектронике возникают в области, промежуточной между баллистическим и диффузионным предельными режимами. В этой области коэффициент прохождения определяется выражением (138) и можно показать [33], что для кажущейся подвижности имеет место следующее соотношение

$$\frac{1}{\mu^{app}} = \frac{1}{\mu^{diff}} + \frac{1}{\mu^{ball}}, \tag{152}$$

которое выглядит как традиционное правило Матиссена [30, 7]. По другому в этой переходной области можно поступить, если в (151) среднюю длину свободного пробега заменить на ее кажущееся значение, а именно:

$$\frac{1}{\lambda^{app}} = \frac{1}{\lambda} + \frac{1}{L}. \tag{153}$$

Кажущуюся длину свободного пробега можно интерпретировать либо как реальную среднюю длину свободного пробега в условиях диффузионного режима, либо как длину резистора в зависимости от того, какая из этих длин меньше.

Предполагая параболическую зонную структуру, подвижность даже в самом общем случае можно выразить через интегралы Ферми – Дирака. Для примера рассмотрим более простой случай, когда $T = 0°K$. Плотность электронов в зоне проводимости

$$n_S = g_v \frac{m^*}{\pi\hbar^2}(E_F - E_C) = D_{2D}(E_F - E_C), \tag{154}$$

а также из формул (63) следует, что

$$M_{2D} = \frac{h}{4}\left\langle v_x^+ \right\rangle D_{2D}, \tag{155}$$

где скорость

$$\left\langle v_x^+ \right\rangle = \frac{2}{\pi} v_F \tag{156}$$

выражается через фермиевскую скорость v_F.

Используя (150) и (151) вместе с (154) и (155), находим, что

$$\mu_{2D}^{ball} = \frac{\bar{D}_{2D}^{ball}}{(E_F - E_C)/q}, \tag{157}$$

$$\mu_{2D}^{diff} = \frac{\bar{D}_{2D}^{diff}}{(E_F - E_C)/q}, \tag{158}$$

где коэффициенты диффузии даются выражениями:

$$\bar{D}_{2D}^{ball} = \left\langle v_x^+ \right\rangle \frac{L}{2}, \tag{159}$$

$$\bar{D}_{2D}^{diff} = \left\langle v_x^+ \right\rangle \frac{\lambda(E_F)}{2}. \tag{160}$$

Уравнения (157) и (158) напоминают соотношение Эйнштейна между подвижностью и коэффициентом диффузии (105) с разностью $E_F - E_C$, играющей роль kT, поскольку с самого начала предполагалось, что $T = 0°K$.

Уравнение (160) дает привычное определение коэффициента диффузии, а вот уравнение (159) вводит новое понятие «баллистического коэффициента диффузии».

1.3.8. Способы записи 2D проводимости

Согласно уравнению (137), удельная диффузионная проводимость, иначе поверхностная проводимость, при $T = 0°K$ дается выражением

$$\sigma_S = \frac{2q^2}{h} M_{2D}(E_F)\lambda(E_F). \tag{161}$$

Выражения для поверхностной проводимости встречаются в различных формах записи. Полезно рассмотреть наиболее часто встречающиеся.

Нам уже известно выражение $M_{2D}(E)$ из (63), а именно:

$$M_{2D}(E) = \frac{h}{4}\left\langle v_x^+(E) \right\rangle D_{2D}(E), \tag{162}$$

где для средней скорости в случае параболической дисперсии и изотропии скорости имеем уравнение (60), а именно:

$$\left\langle v_x^+ \right\rangle = \frac{2}{\pi} v. \tag{163}$$

Позже при рассмотрении общих вопросов рассеяния при электронном транспорте в главе 4 для средней длины свободного пробега в (161) мы получим

$$\lambda(E) = \frac{\pi}{2} v(E) \tau(E), \tag{164}$$

где τ есть время импульсной релаксации, иначе говоря, время между двумя соседними актами рассеяния. Используя (163) и (164), для поверхностной проводимости (161) получаем типичное выражение:

$$\sigma_S = q^2 D_{2D}(E_F)\frac{v^2(E_F)\tau(E_F)}{2}. \qquad (165)$$

Определив электронный коэффициент диффузии

$$\bar{D}(E_F) = \frac{v^2(E_F)\tau(E_F)}{2}, \qquad (166)$$

переписываем (165) в более привычном виде:

$$\sigma_S = q^2 D_{2D}(E_F)\bar{D}(E_F). \qquad (167)$$

Приведем еще один способ записи поверхностной проводимости. Для параболической дисперсии

$$\frac{1}{2}m^* v^2(E_F) = E_F - E_C. \qquad (168)$$

Используя это выражение и (154), переписываем (165) в виде

$$\sigma_S = n_S q\mu, \qquad (169)$$

где подвижность

$$\mu = \frac{q\tau(E_F)}{m^*}. \qquad (170)$$

Уравнение (170) хорошо известно, но его нельзя назвать удачным в качестве базового уравнения для анализа работы электронного устройства. Типичным примером может служить ситуация когда не работает предположение о параболической зонной структуре, порождающее понятие эффективной массы носителя тока.

Подведем итоги. Получены четыре разных способа записи удельной 2D проводимости при $T = 0°K$. Это уравнения (161), (165), (167) и пара уравнений (169) – (170). Все уравнения эквивалентны, но обладают они разными возможностями. Уравнение (161) увязывает удельную проводимость с числом мод проводимости. По уравнениям (165) и (167) нужно знать плотность состояний резистора и скорость или коэффициент диффузии при энергии, равной фермиевскому значению. Уравнение (169) связывает удельную проводимость с поверхностной плотностью электронов. Конечно, роль фермиевского окна проводимости $(-\partial f_0/\partial E)$ остается прежней: проводимость обеспечивают лишь те электроны, энергия которых попадает в окно $\sim \pm 2kT$, охватывающее фермиевскую энергию E_F.

При конечных температурах вместо уравнения (161) имеем:

$$\sigma_S = \frac{2q^2}{h}\int M_{2D}(E)\lambda(E)\left(-\frac{\partial f_0}{\partial E}\right)dE. \qquad (171)$$

Удобно ввести удельную поверхностную проводимость в дифференциальной форме $\sigma'_S(E)$, а именно:

$$\sigma_S = \int \sigma'_S(E)dE, \qquad (172)$$

где

$$\sigma'_S(E) = \frac{2q^2}{h} M_{2D}(E)\lambda(E)\left(-\frac{\partial f_0}{\partial E}\right). \qquad (173)$$

Полную проводимость (172) получаем суммированием проводимостей по всем модам резистора. Уравнение (173) при конечных температурах можно записать по разному так же, как это было показано выше для $T = 0°K$.

1.3.9. Диссипация мощности в баллистических резисторах

Любой проводник, на который подано напряжение V, диссипирует мощность $P = VI = V^2/R$. Мощность обычно диссипирует в результате электрон-фононного взаимодействия, энергия которого передается решетке проводника и проводник греется. В баллистических резисторах перенос электронов происходит без рассеяния, однако, диссипация мощности по прежнему равна V^2/R. Где же эта мощность диссипирует? Поскольку диссипация мощности не может происходить в канале проводимости, то есть единственный ответ – на контактах, где происходят интенсивные процессы рассеяния, направленные на быстрое восстановление равновесных значений электрохимических потенциалов.

Диссипация мощности в баллистическом резисторе иллюстрируется на рис. 12.

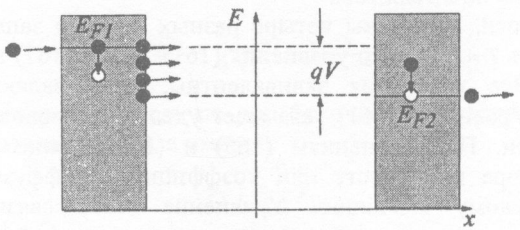


Рис. 12. Диссипация мощности в баллистическом канале происходит на контактах.

Ток возникает в фермиевском окне, созданном разностью фермиевских функций $f_1 - f_2$, отличной от нуля, и следовательно разностью электрохимических потенциалов $E_{F1} - E_{F2} = qV$. Как схематически показано на рис. 12, когда электрон покидает контакт 1, образуется незаполненное состояние («дырка») в спектре контакта 1. Достигнув баллистически контакта 2 с избытком энергии по сравнению с химпотенциалом E_{F2} («горячий электрон»), он теряет этот избыток в результате неупругого рассеяния на контакте 2. При равноправии контактов потеря энергии на контакте 2 составит $qV/2$, так что половина мощности диссипирует на контакте 2. Зарядовая нейтральность контакта 2 восстанавливается после того как электрон покинет контакт 2 и по внешней цепи достигнет контакта 1. Электрон входит в контакт 1 с энергией E_{F1}, теряет энергию неупруго, заполняя собой «дырку», и диссипируя мощность $qV/2$ на контакте 1.

Итак, в баллистическом канале в режиме линейного отклика на каждом контакте диссипирует примерно половина мощности VI.

1.3.10. Падение напряжения в баллистических резисторах

В диффузионном режиме работы однородного резистора напряжение падает линейно вдоль его длины. А как падает напряжение в баллистическом резисторе? Ответ, наверное, ожидаемый – на контактах (рис. 13).

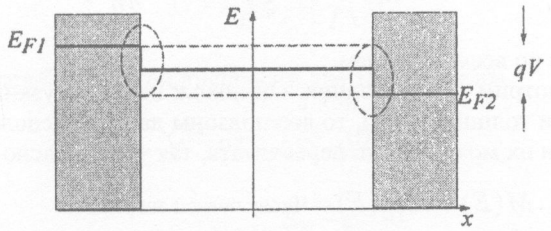


Рис. 13. В баллистическом резисторе при равноправии контактов напряжение $V = (E_{F1} - E_{F2})/q$ падает на контактах поровну.

Внутри контакта 1 есть один хорошо определенный уровень Ферми E_{F1}, аналогично и в контакте 2 есть один уровень E_{F2}. Внутри же всего устройства уровней Ферми два. Часть электронных состояний резистора заполняется истоком. Поскольку они находятся в равновесии с ним, то заполнены они в соответствии с фермиевской функцией контакта 1. Аналогичная ситуация имеет место на стоке. Состояния заполняются в соответствии с фермиевской функцией контакта 2. Расчет среднего значения электрохимического потенциала внутри баллистического резистора [17] приводит к результату, показанному на рис. 13. При равноправии контактов половина падения электрохимического потенциала происходит на истоке, а вторая половина –

на стоке. По этой причине баллистическое сопротивление в 12.9 *кОм* называют квантом контактного сопротивления.

1.3.11. 1D и 3D резисторы

Уравнение Ландауэра для проводимости (109) справедливо для резисторов любой размерности, если только корректно учтено число мод проводимости $M(E)$. Вернемся к 2D проводнику. Он удлинен в направлении движения электронов и имеет конечную ширину W. Движение электронов ограничено потенциальной ямой, их энергия квантуется, а именно:

$$\varepsilon_n = \frac{\hbar^2\pi^2}{2m^*a^2}n^2, \tag{174}$$

где a – ширина потенциальной ямы, n – целое квантовое число, перечисляющее уровни энергии («подзоны»), которые заселяются в соответствии с положением уровня Ферми, и каждая из подзон есть мода проводимости для тока. При достаточно большой ширине резистора W энергии нижних подзон близки друг к другу и многие из этих подзон заселены электронами. В этом случае число подзон пропорционально ширине резистора W и для числа мод проводимости имеем:

$$M(E) = WM_{2D}(E) = W\sum_{n=1}^{N} g_v \frac{\sqrt{2m^*(E-\varepsilon_n)}}{\pi\hbar}, \tag{175}$$

где сумма берется по всем подзонам.

Теперь рассмотрим 1D проводник. Он похож на очень узкий 2D резистор. Если его ширина и толщина малы, то все подзоны далеко расположены друг от друга по энергии и их можно легко пересчитать, так что согласно (63)

$$M(E) = M_{1D}(E) = \text{Числу подзон энергии } E. \tag{176}$$

Наконец, если ширина и толщина резистора велики, то мы имеем фактически 3D проводник, все подзоны близко расположены друг по отношению к другу по энергии и число мод проводимости согласно (63)

$$M(E) = AM_{3D}(E) = g_v \frac{m^*(E-E_C)}{2\pi\hbar^2}, \tag{177}$$

где A – площадь поперечного сечения 3D проводника.

Для 1D резистора имеет место строгое квантовое ограничение в двух измерениях и число мод проводимости дается простым выражением (176). Для 2D резистора строгое квантовое ограничение имеет место в одном измерении и число мод дается выражением (175). Для 3D резистора квантовых ограничений нет и число мод дается выражением (177). Из (175) и (177) видно, что эти формулы выписаны для параболической дисперсии. С таким же успехом их

можно выписать для любой аналитически заданной дисперсии, например, для линейной как в графене, которому мы намерены посвятить в книге отдельную главу 3. Если дисперсия задается эмпирически, то остается прибегнуть только к численным методам. Как только число мод выявлено, интеграл Ландауэра для проводимости (109) не трудно оценить для проводников любой размерности.

Уравнение Ландауэра (109) бывает удобнее переписать в другом виде, а именно:

$$G = \frac{2q^2}{h}\left\langle\left\langle T\right\rangle\right\rangle\left\langle M\right\rangle,$$

$$\left\langle M\right\rangle = \int M(E)(-\frac{\partial f_0}{\partial E})dE, \tag{178}$$

$$\left\langle\left\langle T\right\rangle\right\rangle = \frac{\int T(E)M(E)(-\frac{\partial f_0}{\partial E})dE}{\int M(E)(-\frac{\partial f_0}{\partial E})dE}.$$

Воспользуемся общим выражением для коэффициента прохождения (138) и простоты ради предположим, что средняя длина свободного пробега постоянна и равна λ_0. Тогда для проводимости имеем:

$$G = \frac{2q^2}{h}\frac{\lambda_0}{\lambda_0 + L}\left\langle M\right\rangle. \tag{179}$$

В случае параболической дисперсии для 1D резистора

$$\left\langle M_{1D}\right\rangle = \sum_i \Im_{-1}(\eta_{Fi}), \tag{180}$$

где

$$\eta_{Fi} = \frac{E_F - \varepsilon_i}{kT} \tag{181}$$

и суммирование в (180) ведется по подзонам.

При $T = 0°\ K$ проводимость 1D резистора упрощается до

$$G_{1D} = \frac{2q^2}{h}\frac{\lambda_0}{\lambda_0 + L} \times \text{Число подзон энергии } E_F. \tag{182}$$

В случае статистики Максвелла – Больцмана

$$G_{1D} = q n_L \mu^{app}\frac{1}{L}, \tag{183}$$

где

$$\mu^{app} = \frac{\bar{D}}{(kT/q)}, \tag{184}$$

$$\bar{D} = \frac{v_T \lambda^{app}}{2}, \tag{185}$$

$$v_T = \sqrt{2kT/\pi m^*}, \tag{186}$$

$$\frac{1}{\lambda^{app}} = \frac{1}{\lambda_0} + \frac{1}{L}. \tag{187}$$

Для 2D резистора имеем:

$$\langle M \rangle = W \langle M_{2D} \rangle = \frac{\sqrt{\pi}}{2} W M_{2D}(kT) \sum_i \Im_{-1/2}(\eta_{Fi}), \tag{188}$$

где для параболической дисперсии

$$M_{2D}(kT) = g_v \frac{\sqrt{2m^* kT}}{\pi\hbar}. \tag{189}$$

При $T = 0°\ K$ выражение (179) для 2D резистора упрощается до

$$G_{2D} = \frac{2q^2}{h} \frac{\lambda_0}{\lambda_0 + L} W M_{2D}(E_F). \tag{190}$$

В случае статистики Максвелла – Больцмана

$$G_{2D} = q n_S \mu^{app} \frac{W}{L}. \tag{191}$$

Наконец, для 3D резистора

$$\langle M \rangle = A \langle M_{3D} \rangle = A M_{3D}(kT) \Im_0(\eta_F), \tag{192}$$

где

$$M_{3D}(kT) = g_v \frac{m^* kT}{2\pi\hbar^2} \tag{193}$$

и

$$\eta_F = \frac{E_F - E_C}{kT}. \tag{194}$$

При $T = 0°\ K$ проводимость для 3D резистора упрощается до

$$G_{3D} = \frac{2q^2}{h} \frac{\lambda_0}{\lambda_0 + L} A M_{3D}(E_F). \tag{195}$$

В случае статистики Максвелла – Больцмана

$$G_{3D} = qn\mu^{app}\frac{A}{L}. \qquad (196)$$

В качестве примера приведем результаты анализа экспериментальной вольт-амперной характеристики (ВАХ) углеродных нанотрубок. Это почти идеальные 1D проводники. Экспериментальные ВАХ одностеночных углеродных нанотрубок в металлическом режиме длиной 1 *мкм* при $T = 4$, 100, 200° K в широком диапазоне напряжений воспроизведены из [34] на рис. 14.

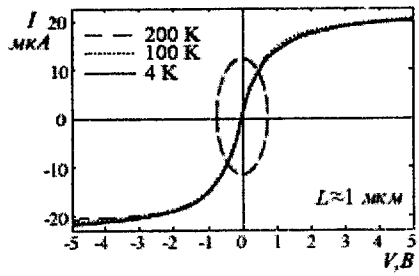


Рис. 14. Вольт-амперная характеристика одностеночных углеродных нанотрубок в металлическом режиме при трех разных температурах [34]. Обведен участок линейного отклика.

Из линейного участка ВАХ для проводимости получается 22 *мкСм*. Для таких углеродных нанотрубок приближение при $T = 0° K$ хорошо работает даже при весьма высоких температурах. Так что можно пользоваться выражением для проводимости (182). Долинное вырождение для углеродных нанотрубок равно двойке. В предположении, что заполнена одна подзона для баллистической проводимости получается $G^{ball} = 154\,мкСм$. Оценка среднего пути свободного пробега (182) дает $\lambda_0 = 167\,нм$, что намного меньше физической длины использованных в эксперименте нанотрубок в 1 *мкм*, так что транспорт в таких нанотрубках диффузионный.

1.3.12. Модель упругого резистора в концепции Ландауэра – Датты

Итак, ток порождается «подготовкой» двух контактов 1 и 2 с фермиевскими функциями $f_1(E)$ и $f_2(E)$. Отрицательному контакту 1 соответствует большее значение электрохимического потенциала, а положительному – меньшее значение. Отрицательный контакт стремится передать электроны в канал проводимости, а положительный контакт стремится извлечь электроны из канала проводимости. Это справедливо для любых проводников – и нано- и макроразмерных.

До сих пор мы пользовались моделью упругого резистора Ландауэра [13–15], особенно не вникая в детали этой важной модели. Ниже мы

рассмотрим эту концепцию более подробно с учетом ее дальнейшего развития Даттой [12].

В модели упругого резистора электроны проскакивают канал проводимости от истока S до стока D упруго, без потери или приобретения энергии (рис. 15).

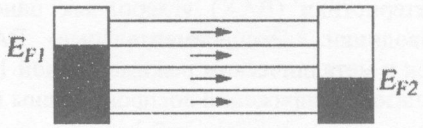


Рис. 15. В упругом резисторе электроны движутся баллистически по каналам с постоянной энергией.

В упругом резисторе ток в промежутке с энергией от E до $E+dE$ отделен от каналов с другими значениями энергии, что позволяет нам записать для тока в дифференциальной форме

$$dI = dE\,G(E)\Big(f_1(E)-f_2(E)\Big), \tag{197}$$

а после интегрирования получить выражение для полного тока. Используя далее (24), получим выражение для низковольтовой проводимости (линейный отклик)

$$G \equiv \frac{I}{V} = \int_{-\infty}^{+\infty} dE\left(-\frac{\partial f_0}{\partial E}\right)G(E), \tag{198}$$

в котором отрицательную производную $(-\partial f_0/\partial E)$ представляют себе в виде прямоугольного импульса, площадь которого равна единице при ширине $\sim \pm 2kT$ (рис. 4, 6, 10). Согласно (198), функция проводимости $G(E)$ для упругого резистора, будучи усредненной по интервалу $\sim \pm 2kT$, включающему значение электрохимического потенциала E_{F0}, дает экспериментально измеряемую проводимость G. При низких температурах можно просто использовать значение $G(E)$ при $E = E_{F0}$.

Такой энергетический подход к проводимости в модели упругого резистора дает существенное упрощение в понимании причин возникновения тока, хотя и звучит парадоксально, поскольку мы традиционно связываем ток I, текущий по проводнику с сопротивлением R, с джоулевым теплом I^2R. Как можно говорить о сопротивлении, если движущиеся по проводнику электроны не теряют энергию?

Ответ на этот вопрос следующий: поскольку электроны не теряют энергию при движении по упругому резистору, потеря энергии происходит на границах проводника с истоком и стоком, на которых и происходит диссипация джоулева тепла. Иначе говоря, упругий резистор, характеризуемый сопротивлением R

канала проводимости, диссипирует джоулево тепло I^2R за пределами канала проводимости. На это указывают различные многочисленные экспериментальные измерения, косвенные и прямые, на наноразмерных проводниках [3, 4], не говоря уже о том, что диссипация тепла, будь она шла бы на одиночной молекуле или на нанопроводнике привела бы к их сгоранию, и общее мнение сейчас таково, что сгорания в реальных экспериментах не происходит, поскольку подавляюще бóльшая часть тепла генерируется на контактах, которые достаточно массивны и тем самым с достаточной легкостью диссипируют тепло.

В понятие упругого резистора не вкладывается обязательного следования электроном по прямолинейной траектории от истока к стоку, допускается и «диффузионное» движение с переменным вектором импульса, но без изменения энергии.

Модель упругого резистора вводится не только в качестве полезного понятия для объяснения работы наноразмерных устройств, но также и потому, что эта модель позволяет объяснить такие транспортные свойства, как проводимость в макроразмерных проводниках. Именно эта модель делает концепцию «снизу – верх» столь эффективной в объяснении транспортных явлений в целом. Еще вернемся к этому утверждению позже, а сейчас получим выражение для проводимости упругого резистора.

Следуя духу концепции «снизу – вверх», рассмотрим простейший упругий резистор с одним каналом энергии ε, с которой электрон проскакивает от истока к стоку (рис. 16).

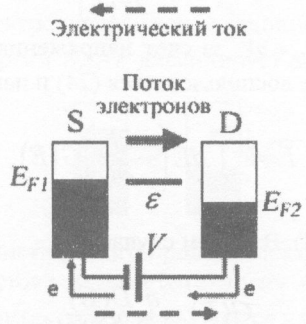


Рис. 16. Одноуровневая модель упругого резистора с энергией в канале ε.

Напомним, что в результате приписывания в конце XIX века электрону отрицательного заряда, что уже изменить невозможно, контакт с бóльшим напряжением D имеет меньший электрохимический потенциал, а движение электрона по каналу происходит от большего значения электрохимического потенциала к меньшему, так что направление тока противоположно реальному движению электронов от истока S к стоку D. Часто удобнее вести речь о токе реальных электронов, а не о токе в исторически сложившемся привычном понимании (рис. 16).

Результирующий одноканальный ток будет

$$I=\frac{q}{\tau}\left(f_1(\varepsilon)-f_2(\varepsilon)\right), \tag{199}$$

где τ есть время, затрачиваемое электроном на проскок от истока к стоку. Теперь можно выражение (199) обобщить на произвольный упругий резистор (рис. 15) с произвольной плотностью состояний $D(E)$, имея в виду, что все энергетические каналы в упругом резисторе проводят независимо друг от друга и в параллельном режиме. Выпишем вначале выражение для тока в канале с энергией от E до $E+dE$:

$$dI=dE\frac{D(E)}{2}\frac{q}{\tau}\left(f_1(E)-f_2(E)\right), \tag{200}$$

в котором учтено, что в таком канале имеется $D(E)dE$ состояний, из которых только половина дает вклад в ток от истока к стоку. Интегрируя, получаем выражение для тока по упругому резистору

$$I=\frac{1}{q}\int_{-\infty}^{+\infty}dEG(E)\left(f_1(E)-f_2(E)\right), \tag{201}$$

где

$$G(E)=\frac{q^2D(E)}{2\tau(E)}. \tag{202}$$

Если разность $E_{F1}-E_{F2}=qV$ за счет напряжения на контактах V намного меньше, чем kT, мы вправе воспользоваться (24) и написать

$$I=V\int_{-\infty}^{+\infty}dE\left(-\frac{\partial f_0}{\partial E}\right)G(E), \tag{203}$$

что и приводит к ур-ю (198). В общем случае

$$G(E)=\frac{q^2D(E)}{2\tau(E)}, \tag{204}$$

однако, нужно помнить, что плотность состояний $D(E)$ и время пролета $\tau(E)$ в общем случае зависят от энергии и должны быть усреднены в пределах $\sim\pm 2kT$, включающих электрохимический потенциал E_{F0}. Это выражение представляется правильным и на интуитивном уровне. Оно утверждает, что проводимость определяется двумя факторами, а именно: наличием энергетических состояний (D) и легкостью, с которой электрон преодолевает расстояние от истока к стоку ($1/\tau$). Это ключевой результат для дальнейшего

обсуждения. Сейчас же мы перейдем к более подробному рассмотрению диссипации тепла упругим резистором.

1.3.12.1. Диссипация тепла упругим резистором

Сопротивление R упругого резистора определяется каналом, а соответствующее тепло I^2R выделяется вне канала проводимости. Рассмотрим подробнее эту ситуацию на примере одноуровневой модели упругого резистора с энергией в канале ε (рис. 16). Каждый раз когда электрон из истока баллистически проскакивает канал, он оказывается в состоянии «горячего электрона» на стоке с энергией ε, превышающей электрохимический потенциал стока E_{F2} (рис. 17).

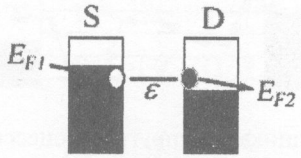


Рис. 17. Мгновенная картина после проскока электрона с истока на сток с энергией в канале ε, превышающей электрохимический потенциал стока E_{F2}.

Диссипационные процессы в стоковом контакте быстро рассеивают избыток энергии $\varepsilon - E_{F2}$. Подобным же образом на истоковом контакте «дырка» с энергией ε, меньшей электрохимического потенциала истока E_{F1} заполняется электроном, а избыток энергии $E_{F1} - \varepsilon$ диссипирует на истоке (рис. 18).

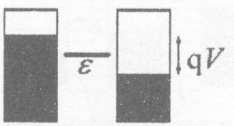


Рис. 18. После баллистического пролета электрона с истока на сток по каналу с энергией ε, на истоке выделяется энергия $E_{F1} - \varepsilon$, а на стоке – энергия $\varepsilon - E_{F2}$, и контакты снова находятся в равновесии.

Суммарная энергия, диссипирующая на контактах, равна $E_{F1} - E_{F2} = qV$. Если N электронов проскакивают с истока на сток за время τ, то на контактах диссипирует мощность

$$P = V \cdot I = qV \cdot N/\tau . \qquad (205)$$

Таким образом, тепло, выделяемое при прохождении тока по упругому резистору, рассеивается на контактах, чему уже имеется достаточно экспериментальных подтверждений на нано- и мезоразмерных проводниках

[35]. Привлекательность модели упругого резистора состоит в том, что механические и термодинамические процессы пространственно разделены.

1.3.12.2. Модель упругого резистора и макропроводники

Естественно задаться вопросом о применимости и пользе модели упругого резистора для объяснения и понимания физики явлений протекания тока в обычных макропроводниках, движение электронов в которых очевидно неупругое. В макропроводниках неупругие процессы совмещены с упругими и беспорядочно распределены по всему проводнику (рис. 19).

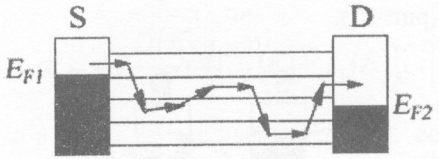


Рис. 19. В макропроводниках неупругие процессы совмещены с упругими.

Разобъем такой макропроводник с беспорядочно распределенными неупругими столкновениями на последовательность упругих резисторов (рис. 20) длины l, намного меньшей длины L реального макропроводника и с перепадом напряжения между соседними упругими резисторами, составляющим лишь незначительные доли от разности потенциалов на концах реального проводника $E_{F1} - E_{F2} = qV$.

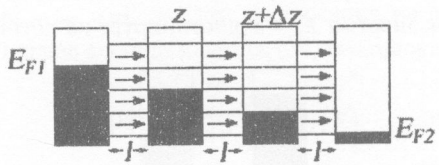


Рис. 20. Гипотетическое разбиение реального макропроводника на последовательность упругих резисторов.

В разбиении длины l должны быть меньше длины λ, представляющей собой среднюю длину пробега электрона до очередного неупругого столкновения. Кроме условия $l < \lambda$ при разбиении должно соблюдаться также требование падения напряжения между соседними упругими резисторами $\Delta V < kT/q$.

Разбиение крупного проводника на упругие резисторы требует некоторой осторожности. Как уже было упомянуто выше, стандартное выражение для закона Ома следует заменить на

$$R = \rho\left(L + \lambda\right)/A, \tag{206}$$

в котором дополнительное сопротивление $\rho\lambda/A$ не зависит от длины проводника и интерпретируется как пограничное сопротивление, возникающее на границе канал/контакт. В выражении (206) λ есть средняя длина свободного пробега, так что такая модификация закона Ома существенна лишь для баллистических проводников ($L \sim \lambda$) и не существенна для крупных проводников ($L \gg \lambda$). Однако, концептуально это дополнительное интерфейсное сопротивление оказывается исключительно важным, если пользоваться гипотетической структурой на рис. 20 для объяснения реальной ситуации на рис. 19. Структура на рис. 20 имеет много пограничных поверхностей раздела, отсутствующих в реальной ситуации (рис. 19), так что следует избавиться от этих виртуальных границ. Например, если каждая секция длины l на рис. 20 характеризуется сопротивлением (206), то правильным выражением для проводника на рис. 19, например, длины $3l$ будет

$$R = \rho\left(3l + \lambda\right)/A, \tag{207}$$

а не

$$R = \rho\left(3l + 3\lambda\right)/A. \tag{208}$$

Таким образом, для получения корректного выражения для проводимости длинного проводника в рамках модели упругого резистора следует аккуратно отделить пограничное сопротивление от сопротивления проводника, зависящего от его длины.

1.3.12.3. Баллистический и диффузионный режимы

Как мы видели выше, проводимость упругого резистора дается выражением (204). Покажем, что время пролета τ через резистор длины L при диффузионном режиме с длиной свободного пробега λ связано с временем пролета при баллистическом режиме τ_B соотношением

$$\tau = \tau_B\left(1 + \frac{L}{\lambda}\right). \tag{209}$$

Подставим (209) в (204), учтем, что

$$G^{ball} \equiv q^2 D/2\tau_B \tag{210}$$

и окончательно для проводимости в диффузионном режиме получим

$$G = \frac{G^{ball}\lambda}{L + \lambda}. \tag{211}$$

Обращая проводимость (211), для закона Ома в обобщенной формулировке, получаем

$$R = \frac{\rho}{A}\left(L + \lambda\right), \tag{212}$$

где

$$\frac{\rho}{A}=\frac{1}{\sigma A}=\frac{1}{G^{ball}\lambda}. \tag{213}$$

До сих пор речь шла о трехмерном резисторе с площадью поперечного сечения A. Разнообразные эксперименты выполняются на двухмерных проводниках шириной W с одномерным поперечным сечением. Для таких 2D-резисторов соответствующие выражения для обобщенного закона Ома, очевидно, имеют вид

$$R=\frac{\rho}{W}\left(L+\lambda\right), \tag{214}$$

где

$$\frac{\rho}{W}=\frac{1}{\sigma W}=\frac{1}{G^{ball}\lambda}. \tag{215}$$

Наконец, для одномерных проводников имеем

$$R=\rho\left(L+\lambda\right), \tag{216}$$

где

$$\rho=\frac{1}{\sigma}=\frac{1}{G^{ball}\lambda}. \tag{217}$$

Запишем обобщенный закон Ома компактно для проводников всех трех размерностей

$$R=\rho\left(L+\lambda\right)\left\{1,\ \frac{1}{W},\ \frac{1}{A}\right\}, \tag{218}$$

где

$$\frac{1}{\rho}=\sigma=G^{ball}\lambda\left\{1,\ \frac{1}{W},\ \frac{1}{A}\right\}. \tag{219}$$

Выражение в фигурных скобках соответствует 1D-, 2D- и 3D-проводникам. Заметим, что удельные сопротивление и проводимость имеют разную размерность в зависимости от размерности проводника, а проводимость и длина по-прежнему измеряются в сименсах и метрах.

Стандартный закон Ома утверждает, что сопротивление стремится к нулю при уменьшении длины проводника до нуля. Никто, конечно, не ожидает, что сопротивление обратится в нуль, но обычное общее мнение таково, что сопротивление стремится к некому пограничному сопротивлению, которое можно сделать произвольно малым по мере совершенствования измерительной техники. Что сейчас надежно установлено экспериментально, так это то, что при наиболее тщательно приготовленных контактах наименьшее наблюдаемое сопротивление связано с каналом проводимости и не зависит от контактов. Обобщенный закон Ома отражает это обстоятельство: даже при приближении

длины проводника к нулю остается остаточное сопротивление, связанное с эффективной длиной λ. Уместно, однако, спросить себя, какой смысл имеет говорить о ненулевой длине λ при нулевой длине проводника. Ответ состоит в том, что для наноразмерных проводников ни удельное сопротивление ρ, ни длина λ не имеют смысла порознь, существенно только их произведение.

Рассмотрим как плотность состояний D и время пролета τ в выражении для проводимости (204) соотносятся с размерами канала в массивных проводниках. Что касается плотности состояний, то это аддитивное свойство. В два раза больший канал имеет в два раза больше электронных состояний, так что плотность состояний для крупных проводников должна быть пропорциональна объему проводника $A{\cdot}L$.

Что касается времени пролета τ, обычно рассматривают два транспортных режима: баллистический с $\tau \sim L$ и диффузионный с $\tau \sim L^2$. Баллистическая проводимость пропорциональна площади поперечного сечения A проводника и, согласно (204), не зависит от длины проводника. Такое «неомическое» поведение действительно наблюдается в наноразмерных проводниках [36].Что же касается проводников с диффузионным транспортным режимом, то они демонстрируют нормальное омическое поведение проводимости $G \sim A/L$.

Различие в двух транспортных режимах можно пояснить следующим образом. В баллистическом режиме время пролета от истока к стоку

$$\tau^{Ball} = \frac{L}{\bar{u}}, \tag{220}$$

где

$$\bar{u} = \left\langle \left| v_z \right| \right\rangle \tag{221}$$

есть среднее значение скорости электронов вдоль оси z – направления движения электронов от истока к стоку.

В случае диффузионного режима время τ зависит от длины проводника квадратично

$$\tau = \frac{L}{\bar{u}} + \frac{L^2}{2\bar{D}}, \tag{222}$$

где величина $\bar{D}$ в рамках теории случайных блужданий [23] есть коэффициент диффузии

$$\bar{D} = \left\langle v_z^2 \tau \right\rangle, \tag{223}$$

где τ есть время свободного пробега.

Используя определение (220), перепишем (222) в виде

$$\tau = \tau^{Ball}\left(1 + \frac{L\bar{u}}{2\bar{D}}\right), \tag{224}$$

которое вместе с (209) для длины λ дает значение

$$\lambda = \frac{2\bar{D}}{\bar{u}}. \tag{225}$$

Для вычисления констант $\bar{u}$ по (221) и $\bar{D}$ по (223) требуется усреднить скорость электронов в направлении их движения (ось z) по всем угловым переменным в зависимости от размерности проводника d = {1, 2, 3}. Простые вычисления дают для

$$\text{1D проводника: } \left\langle |v_z| \right\rangle = v \quad \text{и} \quad \left\langle |v_z^2| \right\rangle = v^2; \tag{226}$$

$$\text{2D проводника: } \left\langle |v_z| \right\rangle = 2v/\pi \quad \text{и} \quad \left\langle |v_z^2| \right\rangle = v^2/2; \tag{227}$$

$$\text{3D проводника: } \left\langle |v_z| \right\rangle = v/2 \quad \text{и} \quad \left\langle |v_z^2| \right\rangle = v^2/3, \tag{228}$$

так что в итоге

$$\bar{u} = \left\langle |v_z| \right\rangle = v(E)\left\{1, \frac{2}{\pi}, \frac{1}{2}\right\}, \tag{229}$$

$$\bar{D} = \left\langle v_z^2 \tau \right\rangle = v^2 \tau(E)\left\{1, \frac{1}{2}, \frac{1}{3}\right\}, \tag{230}$$

или окончательно для средней длины свободного пробега λ имеем

$$\lambda = \frac{2\bar{D}}{\bar{u}} = v\tau\left\{2, \frac{\pi}{2}, \frac{4}{3}\right\}. \tag{231}$$

Подчеркнем, что длина λ включает численный множитель, зависящий от размерности проводника, сравнительно с ее стандартным значением $\lambda = v\cdot\tau$. Можно ли просто использовать это стандартное значение λ? Можно, но тогда в обобщенной формулировке закона Ома (218) нужно не просто заменять L на $L+\lambda$, а на L прибавить λ, помноженное на численный фактор, зависящий от размерности проводника, либо пользоваться определением длины λ по (231). Любопытно, что этот фактор даже для одномерного проводника не равен единице, а равен двум. Величина τ есть среднее время свободного пробега до очередного столкновения. Поскольку рассеяние предполагается изотропным, только половина актов столкновения ведет электрон от истока к стоку. С учетом обратного рассеяния значение длины λ для 1D проводника оказывается

равным $2v\tau$, в чем мы убедимся при обсуждении вопросов рассеяния в главе 4 [37].

Теперь получим уравнение для баллистической проводимости. Из уравнений (231) и (220) имеем

$$G^{Ball} \equiv \frac{q^2 D\bar{u}}{2L}, \tag{232}$$

а после подстановки (229) получим

$$G^{Ball} \equiv \frac{q^2 Dv}{2L}\left\{1, \frac{2}{\pi}, \frac{1}{2}\right\}. \tag{233}$$

Окончательно, подставляя в (219) уравнения (231) и (233) и учитывая (230) для удельной проводимости получаем

$$\sigma = q^2 \bar{D} \frac{D}{L}\left\{1, \frac{1}{W}, \frac{1}{A}\right\}. \tag{234}$$

Итак, на основе общего выражения для проводимости (225) и времен пробега (220) и (222) получены выражения для проводимости в баллистическом режиме (234) и в диффузионном режиме (232).

1.3.12.4. О модах проводимости в общем случае

Из уравнения (233) видно, что баллистическая проводимость пропорциональна плотности состояний на единицу длины проводника, D/L. Поскольку плотность состояний пропорциональна объему, то можно ожидать, что баллистическая проводимость пропорциональна площади поперечного сечения A 3D проводника или ширине W 2D проводника.

Многочисленные эксперименты показывают [36], что для наноразмерных проводников баллистическая проводимость не изменяется линейно с их площадью поперечного сечения, а кратна кванту проводимости

$$G^{Ball} \equiv \frac{q^2}{h} M. \tag{235}$$

Другими словами, реальный проводник можно рассматривать как M независимых мод проводимости, дающих в сумме баллистическую проводимость. Учитывая (233), для числа мод проводимости в общем случае получаем

$$M \equiv \frac{hDv}{2L}\left\{1, \frac{2}{\pi}, \frac{1}{2}\right\}, \tag{236}$$

а из уравнений (219) и (235) удельная проводимость выражается через число мод проводимости M и среднюю длину свободного пробега λ

$$\sigma = \frac{q^2}{h} M \lambda \left\{ 1, \frac{1}{W}, \frac{1}{A} \right\}. \tag{237}$$

Более подробно концепцию мод проводимости рассмотрим ниже.

1.3.12.5. Фундаментальное соотношение

Стандартное выражение для удельной проводимости дается формулой Друде [7], связывающей проводимость σ с электронной плотностью n, эффективной массой m и временем свободного пробега τ:

$$\sigma \equiv \frac{1}{\rho} = \frac{q^2 n \tau}{m}, \tag{238}$$

или используя понятие подвижности

$$\mu = \frac{q\tau}{m}, \tag{239}$$

получаем общеизвестное уравнение (10), а именно:

$$\sigma = qn\mu. \tag{240}$$

С другой стороны, для удельной проводимости в рамках концепции «снизу – вверх» получаются два эквивалентных выражения, одно из которых выражает проводимость через произведение плотности состояний D и коэффициента диффузии $\overline{D}$ (234), а другое – через произведение числа мод M в канале проводимости и средней длины пробега λ (237).

Как и проводимость

$$G \equiv \frac{I}{V} = \int_{-\infty}^{+\infty} dE \left(-\frac{\partial f_0}{\partial E} \right) G(E),$$

удельную проводимость по уравнениям (234) и (237) нужно усреднить по энергии в области нескольких kT, включающей $E = E_{F0}$, с помощью функции теплового уширения, а именно:

$$\sigma = \int_{-\infty}^{+\infty} dE \left(-\frac{\partial f_0}{\partial E} \right) \sigma(E). \tag{241}$$

Уравнение (234) хорошо известно, выводится в стандартных учебниках по физике твердого тела [7], чего нельзя сказать об эквивалентном ему уравнении (237), вывод которого обычно требует использования статистической термодинамики необратимых процессов, например, формализма Кубо [38, 39].

По поводу модели Друде хотелось бы подчеркнуть следующее. Применимость модели Друде весьма ограничена, в то время как уравнения для проводимости (234) и (237) имеют самое общее значение. Например, эти уравнения применимы к графену с непараболическим поведением зон и «безмассовыми» электронами – свойствами, описание которых невозможно в модели Друде. Один из уроков, преподнесенных нам наноэлектроникой, – широкая применимость уравнений для проводимости (234) и (237).

Фундаментальное различие между уравнениями (234) и (237) и теорией Друде состоит в том, что усреднение (241) делает проводимость свойством фермиевской поверхности: проводимость определяется уровнями энергии, близкими к $E = E_{F0}$. А согласно уравнениям (238) – (240) теории Друде проводимость зависит от общей электронной плотности, суммированной по всему спектру энергий, что и приводит к ограниченной применимости этой модели. Проводимость веществ меняется в очень широких пределах несмотря на то, что число электронов в различных веществах приблизительно одинаково. Низкая проводимость стекла объясняется не тем, что в стекле мало так называемых «свободных» электронов, а потому, что для стекла характерна очень низкая плотность состояний и числа мод вблизи $E = E_{F0}$. Понятие же «свободных» электронов относится к интуитивным понятиям.

Для произвольных проводников, как с кристаллической структурой, так и аморфных, так и для молекулярных проводников, следуя [12], покажем, что независимо от дисперсии $E(p)$ плотность состояний $D(E)$, скорость $v(E)$ и импульс $p(E)$ связаны с числом электронных состояний $N(E)$ с энергией, меньшей значения E, соотношением

$$D(E)v(E)p(E) = N(E)\cdot d, \tag{242}$$

где d – размерность проводника. Используя (242) для вычисления проводимости (234) с учетом коэффициента диффузии (230)

$$\bar{D} = \left\langle v_z^2 \tau \right\rangle,$$

для 3D проводника получаем

$$\sigma(E) = q^2 \frac{N(E)}{A\cdot L}\frac{\tau(E)}{m(E)}, \tag{243}$$

где масса определена как

$$m(E) = \frac{p(E)}{v(E)}. \tag{244}$$

Легко убедиться, что фундаментальное соотношение (242) справедливо и для параболической зависимости $E(p)$ и для линейной, как в графене. Для параболической зависимости масса носителя тока от энергии не зависит, что в общем случае не так.

Формула (243) выглядит как выражение (238) теории Друде, если $N/(A\cdot L)$ считать электронной плотностью n. При низких температурах это действительно так, поскольку усреднение (241) при $E = E_{F0}$ дает

$$\sigma = \left(q^2 \frac{N}{A\cdot L} \cdot \frac{\tau}{m} \right)_{E=E_{F0}} = q^2 n\tau/m, \qquad (245)$$

так как $N(E)$ при $E = E_{F0}$ есть полное число электронов (рис. 21).

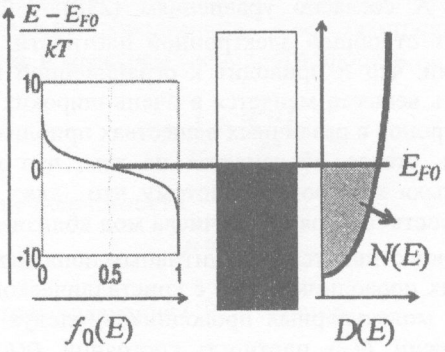


Рис. 21. Равновесная функция Ферми $f_0(E)$, плотность состояний $D(E)$ и полное число электронов $N(E)$.

При ненулевой температуре ситуация тем более сложнее, чем плотность состояний все более непараболическая. Отметим, что ключевым моментом в сведении общего выражения для проводимости (234) к выражению (243), похожему на формулу Друде (238), есть фундаментальное соотношение (242), связывающее плотность состояний $D(E)$, скорость $v(E)$ и импульс $p(E)$ для данного значения энергии с полным числом состояний $N(E)$, полученных интегрирование плотности:

$$N(E) = \int_{-\infty}^{E} dE\, D(E). \qquad (246)$$

Каким образом полное число состояний $N(E)$ по (246) может быть однозначно связано с плотностью состояний $D(E)$, скоростью $v(E)$ и импульсом $p(E)$ для конкретного значения энергии? Ответ состоит в том, что уравнение (242) удовлетворяется только в том случае, когда уровни энергии вычисляются однозначно из выражения для дисперсии $E(p)$. Оно может не выполняться в области энергий перекрывающихся зон или, например, для

аморфных тел, когда не известна зависимость $E(p)$. В этих случаях уравнения (234) и (237) не эквивалентны ур-ю (243) и пользоваться можно только первыми.

Теперь посмотрим, как одиночные зоны, описываемые различными дисперсионными соотношениями $E(p)$ приводят к фундаментальному уравнению (242) и таким образом откроется возможность установления связи между выражениями для проводимости (234) и (237) и формулами Друде (238) – (240). Это приведет также к новой интерпретации мод $M(E)$ и к объяснению их целочисленности.

1.3.12.6. Дисперсии $E(p)$ для кристаллических твердых тел

Пусть стандартно связь энергии с импульсом параболическая (рис. 22)

$$E(p)=E_C+\frac{p^2}{2m}, \tag{247}$$

где под m подразумевается эффективная масса. Будем пользоваться соотношениями $E(p)$ вместо $E(k)$, хотя всегда можно перейти к волновому вектору $k = p/\hbar$. Дисперсия (247) широко используется для различных веществ – и для металлов и для полупроводников.

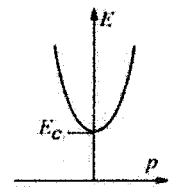


Рис. 22. Параболическая дисперсия.

Но это не единственная возможность. Для графена имеет место линейная зависимость энергии от импульса (рис. 23):

$$E=E_C+v_0 p, \tag{248}$$

где v_0 – константа, равная приблизительно 1/300 от скорости света.

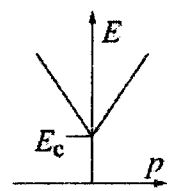


Рис. 23. Линейная дисперсия.

Здесь, как и ранее, под импульсом p подразумевается его абсолютное значение. Другими словами, подразумевается, что зависимость $E(p)$ изотропная.

Для изотропных $E(p)$ скорость сонаправлена с импульсом, а величина ее равна

$$v \equiv \frac{dE}{dp}. \tag{249}$$

Подсчитаем число состояний для проводников разной размерности. На длине резистора L должно укладываться целое число де-бройлевских волн с длиной $\lambda = h/p$:

$$\frac{L}{h/p} = \text{целому числу или } p = \text{целое число} \times \left(h/L\right).$$

Это означает, что разрешенные состояния однородно распределены для данного значения p и каждое из состояний занимает промежуток

$$\Delta p = \frac{h}{L}. \tag{250}$$

Определим функцию $N(p)$ как полное число состояний со значениями импульса меньшими, чем заданное значение p. Для одномерных проводников 1D (рис. 24) эта функция равна отношению доступной длины $2p$ (от $-p$ до $+p$) к промежутку Δp:

$$N(p) = \frac{2p}{h/L} = 2L\left(\frac{p}{h}\right). \tag{251}$$

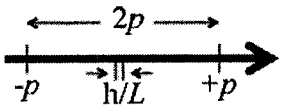


Рис. 24. К подсчету числа состояний для 1D проводника.

Для 2D проводников (рис. 25) нужно разделить площадь поперечного сечения πp^2 на промежутки по длине h/L и по площади поперечного сечения h/W, так что окончательно

$$N(p) = \frac{\pi p^2}{(h/L)(h/W)} = \pi W \cdot L\left(\frac{p}{h}\right)^2. \tag{252}$$

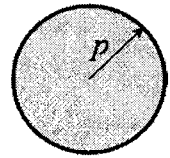


Рис. 25. К подсчету числа состояний для 2D проводника.

Для 3D проводников объем сферы радиуса p делится на произведение промежутков $(h/L)\cdot(h/W_1)\cdot(h/W_2)$, где площадь поперечного сечения $A=W_1\cdot W_2$, так что окончательно

$$N(p)=\frac{(4\pi/3)p^3}{(h/L)(h^2/A)}=\frac{4\pi}{3}A\cdot L\left(\frac{p}{h}\right)^3, \tag{253}$$

или собирая все вместе для $d=\{1,\ 2,\ 3\}$ имеем

$$N(p)=\left\{2\frac{L}{h/p},\ \ \pi\frac{L\cdot W}{\left(h/p\right)^2},\ \ \frac{4\pi}{3}\frac{L\cdot A}{\left(h/p\right)^3}\right\}. \tag{254}$$

Задавшись законом дисперсии $E(p)$, теперь можно вычислить зависимость числа состояний $N(E)$ с энергией, меньшей заданного значения E.

Полученная таким образом функция числа состояний $N(E)$ должна быть равна плотности состояний $D(E)$, проинтегрированной вплоть до энергии состояний E согласно (246), так что плотность состояний

$$D(E)=\frac{dN}{dE}, \tag{255}$$

а с использованием (254) получается

$$D(E)=\frac{dN}{dp}\frac{dp}{dE}=\frac{dp}{dE}\frac{p^{d-1}d}{h^d}\left\{2L,\ \ \pi LW,\ \ \frac{4}{3}\pi LA\right\}. \tag{256}$$

Воспользовавшись (249) и (254), окончательно получим искомое фундаментальное уравнение (242), независимое от закона дисперсии.

1.3.12.7. Еще раз о формуле Друде

Как уже было показано, используя (242) для вычисления проводимости (234) для 3D проводника получаем выражение (243), в котором зависящая от энергии масса определена уравнением (244). Также было показано, что ур-е (243) сводится к формуле Друде (238) при температурах, близких к нулю.

Теперь отдельно рассмотрим проводники *n*–типа и *p*-типа при температурах, отличных от нуля, а также рассмотрим графен.

Проводники n–типа. Используя уравнение (243) и предполагая в нем независимость от энергии массы m и времени τ, получаем

$$\sigma = \frac{q^2\tau}{m}\frac{1}{A\cdot L}\int_{-\infty}^{+\infty} dE\left(-\frac{\partial f_0}{\partial E}\right)N(E). \tag{257}$$

Интегрируя по частям, имеем

$$\int_{-\infty}^{+\infty} dE\left(-\frac{\partial f_0}{\partial E}\right)N(E) = \left[-N(E)f_0(E)\right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} dE\frac{dN}{dE}f_0(E) = \tag{258}$$

$$= \left[0-0\right] + \int_{-\infty}^{+\infty} dE\,D(E)f_0(E) = \text{ полному числу электронов,}$$

поскольку произведение $dE\cdot D(E)\cdot f_0(E)$ есть число электронов в области энергий от E до $E+dE$. Таким образом, уравнение (257) сводится к формуле Друде

$$\sigma = \frac{q^2\tau}{m}\frac{N}{A\cdot L}, \tag{259}$$

имея в виду, что $N/(A\cdot L) = n$.

Проводники p–типа. Любопытная ситуация имеет место для *p*-проводников с ниспадающей дисперсией, например,

$$E(p) = E_c - \frac{p^2}{2m}. \tag{260}$$

Вместо числа состояний по (246) мы теперь имеем (рис. 26)

$$N(E) = \int_{E}^{+\infty} dE\,D(E), \tag{261}$$

что дает

$$D(E) = -\frac{dN}{dE}. \tag{262}$$

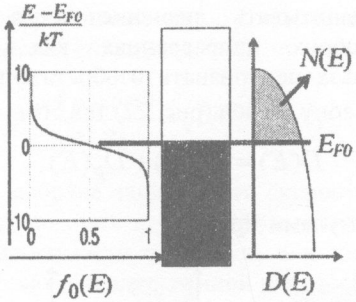


Рис. 26. Равновесная функция Ферми $f_0(E)$, плотность состояний $D(E)$ и число состояний $N(E)$ для p-проводника с дисперсией (260).

Так получается потому, что функция $N(E)$ определяется через функцию $N(p)$, которая дает полное число состояний с импульсом, меньшим заданного значения p, что соответствует энергиям, большим чем заданное значение E согласно дисперсионному соотношению (260).

Если, как и прежде, проинтегрировать по частям

$$\int_{-\infty}^{+\infty} dE\left(-\frac{\partial f_0}{\partial E}\right)N(E)=\left[-N(E)f_0(E)\right]_{-\infty}^{+\infty}+\int_{-\infty}^{+\infty} dE\frac{dN}{dE}f_0(E), \qquad (263)$$

теперь первое слагаемое не зануляется, поскольку и $N(E)$ и $f_0(E)$ в нижнем пределе не нули.

Обойти эту ситуацию можно следующим образом: вместо производной от f_0 брать производную от $(1-f_0)$:

$$\int_{-\infty}^{+\infty} dE\left(-\frac{\partial(1-f_0)}{\partial E}\right)N(E)=\left[-N(E)\left(1-f_0(E)\right)\right]_{-\infty}^{+\infty}+\int_{-\infty}^{+\infty} dE\frac{dN}{dE}\left(1-f_0(E)\right)= \qquad (264)$$

$$=\left[0-0\right]+\int_{-\infty}^{+\infty} dE\,D(E)\left(1-f_0(E)\right)= \text{ полному числу «дырок».}$$

Другими словами, для проводников p-типа можно пользоваться формулой Друде

$$\sigma=q^2n\tau/m, \qquad (265)$$

если под значением n понимать число «дырок»: большему значению n соответствует меньшее число электронов.

Графен. Как подсчитывать значение n в случае, когда зоны распространяются в обоих направлениях как в графене с дисперсией $E = \pm v_0 p$ (рис. 15). Нельзя не признать изобретательным разбивать зоны в графене на зону n-типа и зону p-типа (рис. 27) так, что

$$D(E) = D_n(E) + D_p(E), \tag{266}$$

а затем пользоваться формулами Друде.

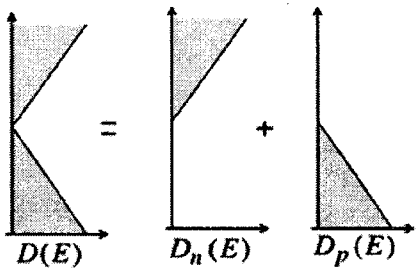


Рис. 27. Искусственное разбиение зонной структуры графена на зоны n- и p-типа.

Хотелось бы подчеркнуть, что нет нужды для подобной изобретательности, поскольку уравнения (234) и (237) применимы во всех случаях и корректно отражают физику проводимости.

1.3.12.8. Пропорциональна ли электропроводимость электронной плотности?

Экспериментальные измерения проводимости часто ведутся в зависимости от электронной плотности, которые, согласно теории Друде, связаны линейно, так что отклонения от линейности трактуют как проявления зависимости времени свободного пробега от энергии. При этом не учитывают, что для непараболических дисперсий масса носителя тока, сама определяемая как p/v, может зависеть от энергии и тем самым приводить к нелинейности проводимости от плотности электронов.

Сначала определим электронную плотность из уравнения (254):

$$n(p) = \left\{ 2\frac{p}{h}, \pi\frac{p^2}{h^2}, \frac{4\pi}{3}\frac{p^3}{h^3} \right\}, \tag{267}$$

где плотность n равна N/L, $N/W \cdot L$ и $N/A \cdot L$ для d = 1, 2 и 3. Перепишем (267) в виде

$$n(p) = K p^d, \tag{268}$$

где коэффициент пропорциональности $K = \{2/h, \pi/h^2, 4\pi/3h^3\}$. Теперь для проводимости (266) с учетом (265) имеем

$$\sigma = q^2 \frac{n(p)\tau(p)}{m(p)} = q^2 K p^{d-1} v(p)\tau(p). \qquad (269)$$

Если известна или выбрана зависимость скорости и времени свободного пробега от энергии, а стало быть и от импульса, в уравнениях (268) и (269) можно избавиться от зависимости от импульса и таким образом установить связь между проводимостью σ и электронной плотностью n.

Например, в случае графена $E = \pm v_0 p$, скорость dE/dp постоянна и равна v_0 и не зависит от импульса. Полагая время свободного пробега не зависящим от энергии, проводимость от плотности электронов из уравнений (268) и (269) с учетом (252) для длины свободного пробега

$$\lambda = \frac{2\bar{D}}{\bar{u}} = v\tau\left\{2, \frac{\pi}{2}, \frac{4}{3}\right\}$$

получается следующей:

$$\sigma = \frac{q^2}{h}\lambda\sqrt{\frac{4n}{\pi}} \qquad (270)$$

или с учетом g-фактора (для графена $g = 4$)

$$\sigma = \frac{q^2}{h}\lambda\sqrt{\frac{4gn}{\pi}}. \qquad (271)$$

Таким образом, проводимость в графене получается пропорциональной $\sqrt{n}$, а не как обычно предполагается ~ n, и с временем свободного пробега, не зависящем от энергии. Вычисления по формуле (271) с $\lambda = 2\mu м$ и $\lambda = 300 нм$ (рис. 28) согласуются с экспериментальными данными [40].

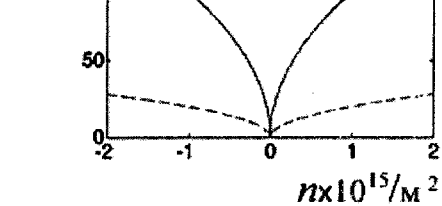


Рис. 28. Проводимость графена по (271) как функция электронной плотности для значений $\lambda = 2\mu м$ (сплошная) и $\lambda = 300 нм$ (пунктир) согласуется с экспериментальными данными (рис. 1 в [40]).

1.3.12.9. Квантование проводимости и целочисленность мод проводимости

Баллистическая проводимость квантуется, $G^{Ball} = q^2 M/h$, где для низкоразмерных проводников при низких температурах число M – целое. Выше было получено выражение для числа мод

$$M \equiv \frac{hD\nu}{2L}\left\{1, \frac{2}{\pi}, \frac{1}{2}\right\} \tag{272}$$

через произведение плотности состояний D и скорости электронов ν, и совершенно не очевидна целочисленность выражения (272). Пользуясь выражениями для дисперсии $E(p)$ оказывается возможным дать иную интерпретацию $M(p)$, свидетельствующую о целочисленной природе числа мод проводимости M.

Используя (242), перепишем (272) в виде

$$M = \frac{hN(p)}{2Lp}\left\{1, \frac{4}{\pi}, \frac{3}{2}\right\}, \tag{273}$$

где $N(p)$ есть полное число состояний с импульсом, меньшим заданного значения p. С использованием (254) преобразуем (273) к виду:

$$M(p) = \left\{1, 2\frac{W}{h/p}, \pi\frac{A}{(h/p)^2}\right\}. \tag{274}$$

Так же как число состояний $N(p)$ дает нам число де-бройлевских длин волн; укладывающихся в проводнике, так и $M(p)$ дает нам число мод, укладывающихся в поперечном сечении проводника, причем это число не зависит от закона дисперсии, поскольку при выводе (274) нигде не использовался какой-либо конкретный закон дисперсии.

При оценке чисел $N(p)$ и $M(p)$ в конкретной задаче на практике получаются, конечно, дробные числа. Однако, по физическому смыслу эти числа должны быть целыми. В больших проводниках при ненизких температурах квантование $M(p)$ смазывается, однако, в мезо- и наноразмерных проводниках наблюдается целочисленная природа числа мод $M(p)$ и квантование проводимости. Поэтому (274) правильнее переписать в виде

$$M(p) = Int\left\{1, 2\frac{W}{h/p}, \pi\frac{A}{(h/p)^2}\right\}, \tag{275}$$

где под $Int\{x\}$ подразумевается наибольшее целое число, меньшее значения x.

В одномерных проводниках число мод совпадает с g-фактором, равным числу долин, умноженным на вырождение по спину, равное 2. Сопротивление баллистических проводников $\sim M \cdot h/q^2$, так что сопротивление баллистического 1D проводника приблизительно равно $25K\Omega$, деленных на g, что и наблюдается экспериментально: большинство металлов и полупроводников типа GaAs имеют $g = 2$ и баллистическое сопротивление 1D образцов $\sim 12.5K\Omega$, а углеродные нанотрубки двухдолинные с $g = 4$ и их баллистическое сопротивление $\sim 6.25K\Omega$.

Мы ставили перед собой задачу в этой главе, написанной по нашим журнальным публикациям, показать как пользоваться уравнением проводимости Ландауэра (109) при отсутствии перепада температур на концах проводника. Основные результаты можно сформулировать коротко следующим образом:

Все проводники обладают постоянным сопротивлением даже при отсутствии рассеяния электронов. Баллистическое сопротивление есть нижний предел сопротивления независимо от того, насколько мал проводник. Этот баллистический предел сопротивления становится важным даже при работе электронных устройств при комнатной температуре.

Баллистическое сопротивление квантуется и квантом сопротивления является величина $h/2q^2$.

Вся область перехода от баллистического транспорта к диффузионному стандартно трактуется в обобщенной модели ЛДЛ с помощью коэффициента прохождения.

Резисторы всех размерностей 1D, 2D и 3D трактуются в формализме ЛДЛ единообразно, а сама трактовка допускает любой вид дисперсии.

При изучении электрических свойств любого нового материала, включая наносистемы, начинать следует не с договоренностей (106) – (108), а с уравнения Ландауэра [41 – 60].

Литература

1. V. V. Mitin, M. A. Kochelap, M. A. Stroscio. *Introduction to Nanoelectronics: Science, Nanotechnology, Engineering, and Applications*. Cambridge: Cambridge University Press, 2012.
2. C. Durkan. *Current at the Nanoscale. An Introduction to Nanoelectronics.* London: World Scientific, 2014.
3. Дж. М. Мартинес-Дуарт, Р. Дж. Мартин-Пальма, Ф. Агулло-Руеда. *Нанотехнологии для микро- и оптоэлектроники*. Москва: Техносфера, 2007.
4. В. П. Драгунов, И. Г. Неизвестный, В.А. Гридчин. *Основы наноэлектроники*. Москва: Логос, 2006.
5. P. Drude. Zur Elektronentheorie der metalle, *Ann. Physik*, v. 306, n 3, 566 (1900).
6. P. Drude. Zur Elektronentheorie der Metalle: Galvanomagnetische und thermomagnetische Effecte, *Ann. Physik*, v. 308, n 11, 369 (1900).
7. Н. Ашкрофт, Н. Мермин. *Физика твердого тела.* Москва: Мир, 1979.
8. R. H. M. Smit, Y. Noat, C. Untiedt, N. D. Lang, M. C. van Hemert, J. M. van Ruitenbeek. Measurement of the conductance of a hydrogen molecule, *Nature*, v. 419, n 3, 906 – 909 (2002).
9. M. Fuechsle, J. A. Miwa, S. Mahapatra, H. Ryu, S. Lee, O. Warschkow, L. C. L. Hollenberg, G. Klimeck, M. Y. Simmons. A single-atom transistor, *Nature Nanotech.*, v. 7, 242 – 246 (2012).
10. M. Lundstrom, S. Datta. *Electronics from the Bottom Up: A New Approach to Nanoelectronic Devices and Materials*: www.nanohub.org/topics/ElectronicsFromTheBottomUp.
11. M. Lundstrom. *Near-Equilibrium Transport: Fundamentals and Applications*: www.nanohub.org/resources/11763.
12. S. Datta. *Lessons from Nanoelectronics: A New Perspective on Transport.* Hackensack, New Jersey: World Scientific Publishing Company, 2013: www.nanohub.org/courses/FoN1.
13. M. Lundstrom, C. Jeong. *Near-Equilibrium Transport: Fundamentals and Applications*. Hackensack, New Jersey: World Scientific Publishing Company, 2013.
14. R. Landauer. Spatial variation of currents and fields due to localized scatterers in metallic conduction, *IBM J. Res. Dev.*, v. 1, n.3, 223 – 231 (1957).
15. R. Landauer. Electrical resistance of disordered one dimensional lattices, *Philos. Mag.*, v. 21, 863 – 867 (1970).
16. R. Landauer. Electrical resistance of disordered one dimensional lattices, *J. Math. Phys.*, v. 37, n. 10, 5259 (1996).
17. S. Datta. *Electronic Transport in Mesoscopic Systems.* Cambridge: Cambridge University Press, 2001.
18. S. Datta. *Quantum Transport: Atom to Transistor.* Cambridge: Cambridge University Press, 2005.
19. M. Lundstrom. *Nanoscales Transistors*: www.nanohub.org/courses/NT.

20. R. F. Pierret. *Semiconductor Device Fundamentals*. Reading, MA: Addison–Wesley, 1996.
21. Ю. А. Кругляк, Н. Е. Кругляк. Методические аспекты расчета зонной структуры графена с учетом σ-остова. Теоретические основы, *Вестник Одес. гос. эколог. ун-та,* № 13, 207 – 218 (2012).
22. C. Jeong, R. Kim, M. Luisier, S. Datta, M. Lundstrom. On Landauer versus Boltzmann and full band versus effective mass evaluation of thermoelectric transport coefficients, *Appl. J. Phys.*, v. 107, n. 2, 023707 (2010).
23. H. C. Berg. *Random walks in biology*. Princeton: Princeton University Press, 1993.
24. R. Waser. *Nanoelectronics and Information Technology*, New York: John Wiley & Sons, 2012.
25. B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel, C. T. Foxon. Quantized conductance of point contacts in a two-dimensional electron gas, *Phys. Rev. Lett.*, v. 60, 848 (1988).
26. D. F. Holcomb. Quantum Electrical Transport in samples of limited dimensions, *Amer. J. Phys.*, v. 67, n. 4, 278 (1999).
27. D. Cvijovic. Fermi – Dirac and Bose – Einstein functions of negative integer order, *Theoret. Math. Phys.*, v. 161, n. 3, 163 (2009).
28. R. Dingle. The Fermi – Dirac Integrals, *Appl. Scientific Res.*, v. 6, n. 1,225 (1957).
29. R. Kim, M. S. Lundstrom. Notes on Fermi - Dirac Integrals: www.nanohub.org/resources/5475.
30. M. Lundstrom. *Fundamentals of Carrier Transport*. Cambridge: Cambridge Univ. Press: 2000.
31. P. Yu, M. Cardona. *Fundamentals of Semiconductors. Physics and Materials Properties.* Berlin: Springer, 2010.
32. M.S.Shur. Low Ballistic Mobility in GaAs HEMTs, *IEEE Electron Dev. Lett.*, v. 23, n. 9, 511 (2002).
33. J. Wang, M. Lundstrom. Ballistic Transport in High Electron Mobility Transistors, *IEEE Trans. Electron Dev.*, v. 50, n. 7, 1604 (2003).
34. Z. Yao, C. L. Kane, C. Dekker. High-Field Electrical Transport in Single-Wall Carbon Nanotubes, *Phys. Rev. Lett.* v. 84, n. 13, 2941 (2000).
35. M. Lundstrom, J. Guo. *Nanoscale Transistors: Physics, Modeling, and Simulation.* Berlin: Springer, 2006.
36. Yu. V. Nazarov, Ya. M. Blanter. *Quantum Transport. Introduction to nanoscience.* Cambridge: Cambridge University Press, 2009.
37. Ю. А. Кругляк. Учет рассеяния в транспортной модели Ландауэра – Датты – Лундстрома, *ScienceRise*, т. 3, № 2(8), 99 – 107 (2015).
38. F. W. Sears, G. L. Salinger. *Thermodynamics, Kinetic Theory, and Statistical Thermodynamics.* Boston: Addison-Wesley, 1975.
39. R. Kubo. Statistical-Mechanical Theory of Irreversible Processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems,

J. Phys. Soc. Japan, v. 12, 570 – 586 (1957).
40. K. I. Bolotin, K. J. Sikes, J. Hone, P. Kim, H. L. Stormer. Temperature-Dependent Transport in Suspended Graphene, *Phys. Rev. Lett.* v. 101, 096802 (2008).
41. Ю. О. Кругляк, Н. Ю. Кругляк, М. В. Стріха. Уроки наноелектроніки: Виникнення струму, формулювання закону Ома і моди провідності в концепції «знизу – вгору», *Сенсор. електр. мікросист. техн.*, т. 9, № 4, 5 – 29 (2012).
42. Ю. А. Кругляк, Н. Е. Кругляк. Уроки наноэлектроники. 1. Причины возникновения тока в концепции "снизу – вверх", *Вестник Одес. гос. эколог. ун-та,* № 14, 197 – 206 (2012).
43. Ю. А. Кругляк, Н. Е. Кругляк. Уроки наноэлектроники. 2. Модель упругого резистора и новая формулировка закона Ома в концепции "снизу – вверх", *Вестник Одес. гос. эколог. ун-та,* № 14, 207 – 216 (2012).
44. Ю. А. Кругляк, Н. Е. Кругляк. Уроки наноэлектроники. 3. Электронная проводимость и моды проводимости в концепции "снизу – вверх", *Вестник Одес. гос. эколог. ун-та,* № 15, 213 – 222 (2013).
45. Ю. А. Кругляк, Н. Е. Кругляк. Уроки наноэлектроники. 1. Причины возникновения тока в концепции "снизу – вверх", *Физич. образов. в вузах,* т. 19, № 1, 50 – 61 (2013).
46. Ю. А. Кругляк, Н. Е. Кругляк. Уроки наноэлектроники. 2. Модель упругого резистора и новая формулировка закона Ома в концепции "снизу – вверх", *Физич. образов. в вузах*, т. 19, № 2, 161 – 173 (2013).
47. Ю. А. Кругляк, Н. Е. Кругляк. Уроки наноэлектроники. 3. Электронная проводимость и моды проводимости в концепции "снизу – вверх", *Физич. образов. в вузах*, т. 19, № 3, 99 – 110 (2013).
48. Ю. А. Кругляк. Обобщённая модель электронного транспорта Ландауэра – Датты – Лундстрома, *Наносистемы, наноматериалы, нанотехнологии,* т. 11, № 3, 519 – 549 (2013); Erratum: *ibid*, т. 12, № 2, 415 (2014).
49. Ю. А. Кругляк. От баллистической проводимости к диффузионной в транспортной модели Ландауэра – Датты – Лундстрома, *Наносистемы, наноматериалы, нанотехнологии*, т. 11, № 4, 655 – 677 (2013).
50. Ю. А. Кругляк, П. А. Кондратенко, Ю. М. Лопаткин. Наноэлектроника: возникновение тока, новая формулировка закона Ома и моды проводимости в концепции «снизу – вверх», *Ж. нано- электр. физ.*, т. 5, № 1, 01023(15) (2013).
51. Yu. A. Kruglyak, P. A. Kondratenko, Yu. M. Lopatkin. Nanoelectronics: Current Generation, New Formulation of Ohm's Law and Conduction Modes by "Bottom – Up" Approach, *J. Nano- Electron. Phys.*, v. 5, n. 1, 01023(15) (2013).
52. Yu. A. Kruglyak. Landauer – Datta – Lundstrom Generalized Transport Model for Nanoelectronics, *Proc. Intern. Conf: Nanomaterials: Applications and Properties*, v. 3, N 1, 01001(6pp) (2014).
53. Yu. A. Kruglyak. Landauer – Datta – Lundstrom Generalized Transport Model

for Nanoelectronics, *Proc. Intern. Conf: Nanomaterials: Applications and Properties*, v. 3, N 2, 02NAESF01 (2014).
54. Yu. Kruglyak. Landauer – Datta – Lundstrom Generalized Transport Model for Nanoelectronics, *J. Nanoscience*, v. 2014, Art. ID 725420, 15 pp. (2014); DOI: 10.1155/2014/725420.
55. Ю. А. Кругляк. Обобщенная модель транспорта электронов и тепла Ландауэра – Датты – Лундстрома в микро- и наноэлектронике, *ScienceRise*, т. 5, № 3(5), 21 – 38 (2014).
56. Ю. А. Кругляк. Обобщенная модель электронного транспорта Ландауэра – Датты – Лундстрома, *Ж. физ. хим.*, т. 88, № 11, 1648 – 1659 (2014).
57. Yu. A. Kruglyak. Generalized Landauer – Datta – Lundstrom electron transport model, *Russian J. Phys. Chem.*, v. 88, n. 11 1826 – 1836 (1914).
58. Ю. А. Кругляк. Наноэлектроника «снизу – вверх»: Возникновение тока, обобщенный закон Ома, упругий резистор, моды проводимости, термоэлектричество, *ScienceRise*, т. 7, № 2(12), 76 – 100 (2015).
59. Ю. А. Кругляк, Л. С. Кострицкая. Обобщенная модель электронного транспорта и переноса тепла в микро- и наноэлектронике, *Вестник Одес. гос. эколог. ун-та*, № 19, 155 – 169 (2015).
60. Yu. A. Kruglyak, M. V. Strikha. Landauer – Datta – Lundstrom generalized electron transport model for micro- and nanoelectronics, *Proc. 2015 IEEE 35th Intern. Conf. Electronics Nanotech.* (ELNANO), April 21 – 24, 2015, Kyiv, Ukraine, pp. 70 – 74; DOI: 10.1109/ELNANO.2015.7146837.

Глава 2. Обобщенная модель переноса тепла

2.1. Термоэлектрические явления и устройства

Обобщенная транспортная модель для задач наноэлектроники, в равной мере пригодная также и для моделирования устройств микро- и макроэлектроники, изложена в главе 1 для случая, когда температура контактов резистора одинакова [1, 2]. В главе 2 рассмотрим термоэлектрические (ТЭ) эффекты и обобщенную модель переноса тепла в концепции Ландауэра – Датты – Лундстрома (ЛДЛ). Напомним, что термоэлектрические устройства конвертируют тепло в электрический ток или же ток используется для нагревания/охлаждения [3 – 8].

При рассмотрении транспорта в массивных проводниках (§1.3.5) для плотности тока вдоль длинной оси проводника x получено выражение через градиент электрохимического потенциала E_F с удельной проводимостью σ в роли коэффициента пропорциональности:

$$J_x = \sigma \frac{d\left(E_F/q\right)}{dx}. \; [A/м^2] \tag{1}$$

Перепишем это выражение через удельное сопротивление ρ:

$$\frac{d\left(E_F/q\right)}{dx} = \rho J_x. \tag{2}$$

а для однородных проводников

$$\frac{d\left(E_F/q\right)}{dx} = \mathrm{E}_x, \tag{3}$$

где E_x – электрическое поле.

Сначала рассмотрим диффузионный транспорт в массивных 3D проводниках. Аналогичным образом можно рассматривать 1D и 2D проводники. Позже рассмотрим транспорт в баллистическом и квази-баллистическом режимах.

Как нужно изменить выписанные выше уравнения с учетом температурного градиента? Мы увидим ниже, что ответ на этот вопрос можно записать следующим образом:

$$J_x = \sigma \frac{d\left(E_F/q\right)}{dx} - S\sigma \frac{dT}{dx}, \tag{4}$$

$$\frac{d\left(E_F/q\right)}{dx} = \rho J_x + S \frac{dT}{dx}, \tag{5}$$

где S – коэффициент Зеебека в $[B/K]$.

Термоэлектричество сочетает потоки электронов и тепла. Таким образом, в дополнение к уравнению для потока электронов нужно иметь уравнение для потока тепла. Поскольку тепло распространяется против градиента температуры, то для потока тепла можно ожидать пропорциональности вида

$$J_{Qx} = -\kappa \frac{dT}{dx}. \; [Bm/м^2] \tag{6}$$

Как изменится это уравнение в присутствии электрического тока? Ответ таков:

$$J_{Qx} = TS\sigma \frac{d\left(E_F/q\right)}{dx} - \kappa_0 \frac{dT}{dx}, \tag{7}$$

$$J_{Qx} = \pi J_x - \kappa \frac{dT}{dx}, \tag{8}$$

где коэффициент Пельтье

$$\pi = TS, \; [Bm/A] \tag{9}$$

и

$$\kappa = \kappa_0 - S^2 \sigma T \; [Bm/м \cdot K] \tag{10}$$

есть электронная теплопроводность в условиях незамкнутой электрической цепи, а κ_0 – в условиях короткозамкнутой цепи. И в том и в другом случае речь идет о переносе тепла только электронами. Перенос тепла фононами рассмотрим в этой же главе позже.

Рассмотрим эффекты Зеебека и Пельтье подробнее.

2.1.1. Эффект Зеебека

Если температура контактов проводника разная, возникает поток электронов от более нагретого контакта к менее нагретому (эффект Зеебека) и генерируется напряжение Зеебека V_{oc}, измерить которое можно прилагая к концам незамкнутой ос (open circuit) цепи напряжение противоположной полярности до прекращения тока (рис. 1).

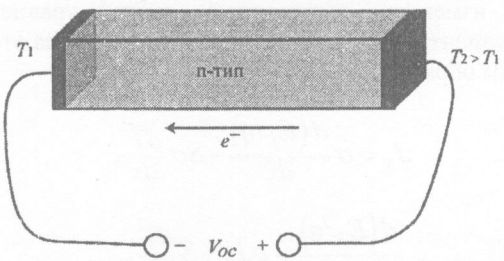


Рис. 1. При разной температуре контактов электроны движутся в направлении от более нагретого контакта к менее нагретому.

Эффект Зеебека физически корректно рассматривать с учетом поведения фермиевских функций при разных температурах (рис. 2).

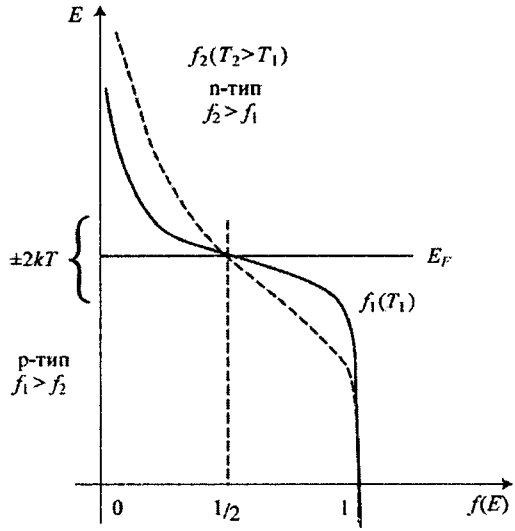


Рис. 2. Качественный ход двух фермиевских функций с разными температурами.

Фермиевское окно проводимости составляет $\sim \pm 2kT$ и охватывает значение $E = E_F$. Чем выше температура, тем доступнее окно проводимости. Ток зависит от разности фермиевских функций $f_1 - f_2$. Если рассматривается проводник *n*-типа, то электронные состояния, обеспечивающие ток, находятся в зоне проводимости выше уровня Ферми, где $f_2 > f_1$, а в проводнике *p*-типа ток обеспечивают «дырочные» состояния, лежащие ниже уровня Ферми в валентной зоне, где $f_1 > f_2$ (рис. 2). В первом случае для прекращения тока ($f_1 = f_2$) на контакт 2 нужно подать положительный потенциал, а во втором случае – отрицательный. В соответствии с этим напряжение Зеебека положительное для проводников *n*-типа и отрицательное для проводников *p*-типа. Этим пользуются для определения типа проводимости резистора.

Теперь посмотрим что же контролирует саму величину напряжения Зеебека. Рассмотрим *n*-проводник в состоянии равновесия, когда $E_{F1} = E_{F2}$ и $T_1 = T_2$. Для любого состояния проводника вероятность f_1 быть заселенным со стороны контакта 1 такая же как и вероятность f_2 быть заселенным со стороны контакта 2: $f_1 = f_2$. Тока нет.

На рис. 3 показан проводник с подачей положительного потенциала на его контакт 2, что приводит к понижению электрохимического потенциала $E_{F2} = E_{F1} - qV$.

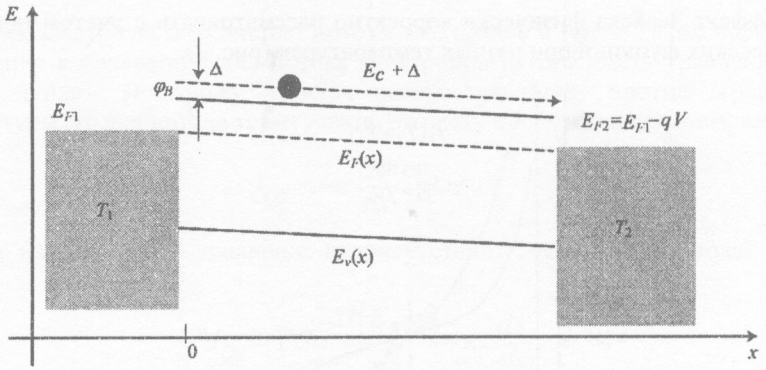


Рис. 3. Энергетическая диаграмма проводника *n*-типа при подаче разности потенциалов $V_2 > V_1$; разность температур контактов может или иметь место или быть равной нулю.

Пусть вначале температура контактов одинакова ($T_1 = T_2$). Теперь для любого состояния (E, x) проводника $f_1 > f_2$ и электроны потекут слева направо, что соответствует току в направлении отрицательной полуоси x. Рассмотрим ситуацию когда и потенциал и температура контакта 2 больше чем для контакта 1: $V_2 > V_1$, $T_2 > T_1$. Тот факт, что $T_2 > T_1$ вынуждает электроны перетекать с контакта 2 на контакт 1, а больший потенциал на контакте 2 по сравнению с контактом 1 вынуждает электроны двигаться в противоположном направлении. В условиях незамкнутой цепи обе тенденции уравновешивают друг друга и ток равен нулю.

Пусть $T_2 > T_1$, а к контакту 2 приложен такой потенциал, что ток равен нулю (рис. 3). Электроны движутся с энергией, лежащей чуть выше (на величину Δ) дна зоны проводимости. В некоторой промежуточной точке проводника (на бесконечно малом отрезке dx) энергия есть $E_C(0) + \Delta$. Численное значение Δ зависит от зонной структуры проводника и физики его рассеивающих центров, для невырожденных же проводников $\Delta \approx 2kT$. Ток через эту точку не идет, поскольку состояние в этой промежуточной точке характеризуется равенством $f_1 = f_2$. В условиях незамкнутой цепи

$$f_1[E_C(0) + \Delta] = f_2[E_C(0) + \Delta] \tag{11}$$

или в развернутом виде

$$\frac{1}{\exp\left[\left(E_C(0)+\Delta-E_{F1}\right)/kT\right]+1} = \frac{1}{\exp\left[\left(E_C(0)+\Delta-E_{F1}+q\delta V\right)/kT\right]+1}, \tag{12}$$

где учтено то обстоятельство, что на бесконечно малом отрезке dx $E_{F2} = E_{F1} - q\delta V$ (рис. 3). Из равенства показателей экспонент непосредственно следует, что

$$\delta V = -S\delta T \tag{13}$$

с $\delta T = T_1 - T_2$ и

$$S = -\frac{E_C(0) + \Delta - E_{F1}}{qT_1} \equiv -\frac{E_{av} - E_{F1}}{qT_1}, \tag{14}$$

где $E_C(0) + \Delta \equiv E_{av}$ есть среднее значение энергии, при которой движутся электроны. Коэффициент Зеебека пропорционален разности между этой средней энергией и фермиевской энергией на левом контакте E_{F1}. По договоренности, для *n*-проводников коэффициент Зеебека отрицательный. Перепишем ур-е (14) следующим образом:

$$S(T) = -\frac{k}{q}\left(\frac{E_C - E_F}{kT} + \delta\right), \tag{15}$$

где $\delta = \Delta / kT$, а фундаментальная константа $k/q = 86\,мкВ/К$. Для невырожденных 3D проводников безразмерный параметр $\delta \approx 2$, а для сильно вырожденных

$$\delta \to \frac{E_F - E_C}{kT}, \tag{16}$$

что в пределе по мере удаления фермиевской энергии от дна зоны проводимости ведет к занулению коэффициента Зеебека (рис. 4).

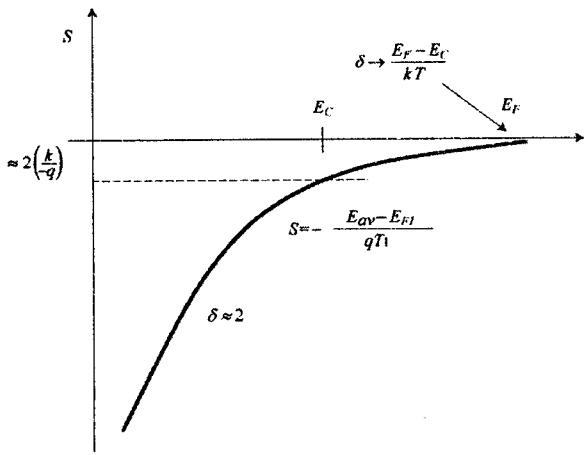


Рис. 4. Поведение коэффициента Зеебека в зависимости от положения уровня Ферми E_F относительно дна зоны проводимости E_C.

Коэффициент Зеебека (с учетом знака) тем больше, чем глубже находится уровень Ферми относительно зоны проводимости (слабо легированные *n*-полупроводники). При нахождении уровня Ферми на дне зоны проводимости коэффициент Зеебека равен $\approx 2 \times 86\, мкВ/К$. Чем выше поднимается уровень Ферми относительно дна зоны проводимости, тем меньше коэффициент Зеебека (низкие значения коэффициента Зеебека у металлов).

Подставим (15) в (13) и просуммируем по всей длине проводника, в результате чего получим перепад напряжения на концах проводника

$$\Delta V = -\int_{T_1}^{T_2} S(T)\, dT\,. \; (J_x = 0) \qquad (17)$$

Приведенные рассуждения показывают происхождение второго слагаемого в уравнении (5).

2.1.2. Эффект Пельтье

Охлаждение и нагревание за счет эффекта Пельтье при прохождении тока по проводнику, поддерживаемому в изотермических условиях ($T_1 = T_2$) иллюстрируется на рис. 5.

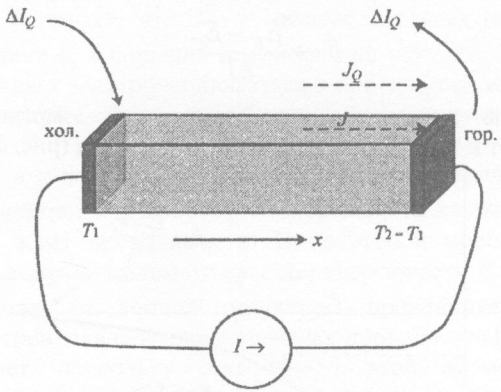


Рис. 5. Прохождение тока по проводнику в изотермических условиях сопровождается охлаждением одного контакта и нагреванием другого (эффект Пельтье).

В соответствии с выбранным направлением тока электроны начинают движение с небольшой дрейфовой скоростью слева направо. По ходу движения электроны рассеиваются на фононах, приобретая при этом намного большую тепловую скорость. В результате поток электронов сопровождается потоком тепла (эффект Пельтье): правый контакт отдает тепло, а левый контакт поглощает тепло. Если направление тока изменить на противоположное, то миссии контактов поменяются местами.

Для оценки теплового потока обратимся к рис. 6.

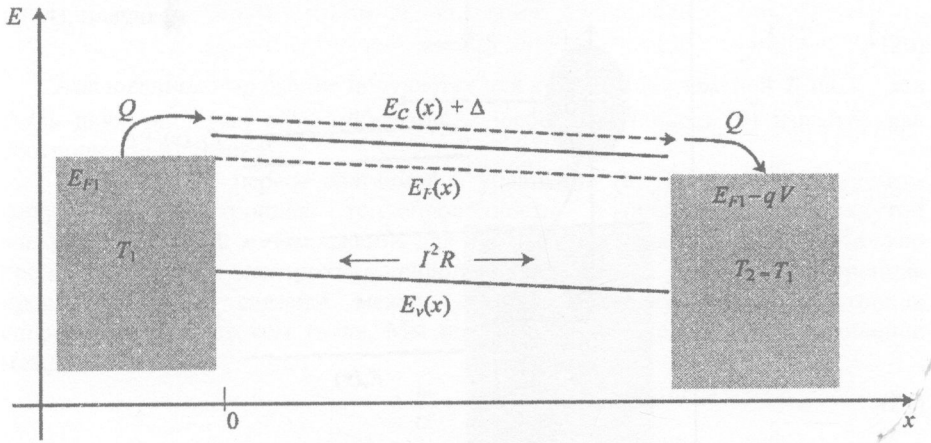


Рис. 6. Энергетическая диаграмма слабо допированного полупроводника *n*-типа.

Металлические контакты сильно вырождены, так что фермиевское окно проводимости $-(\partial f_0/\partial E)$ представляет собой фактически δ-функцию при $E=E_F$. В металлических контактах ток течет очень близко к фермиевской энергии. В слабо допированном полупроводнике ток течет чуть выше дна зоны проводимости при $E_{av}\equiv E_C(0)+\Delta$. На границе металл/полупроводник энергия с фермиевского уровня возрастает до этого среднего значения E_{av} с поглощением тепла $Q=E_C(0)+\Delta-E_{F1}$. Эта тепловая энергия берется из решетки металлического контакта 1 (фононы). На противоположном контакте происходит обратное: тепло $Q=E_C(L)+\Delta-E_{F2}$ диссипируется металлическим контактом 2, где L – длина проводника. Мы все это время рассматриваем массивный 3D проводник в диффузионном режиме. При прохождении тока по резистору выделяется также джоулево тепло I^2R, пропорциональное квадрату тока. Тепло Пельтье пропорционально первой степени тока. В случае сильно допированного полупроводника можно ожидать намного меньшего охлаждения/нагревания контактов за счет эффекта Пельтье, поскольку средняя энергия E_{av} за счет меньшего потенциала барьера φ_B (рис. 3) лежит намного ближе к фермиевской энергии по сравнению со слабо допированным полупроводником.

Рассмотрим подробнее область вблизи левого контакта 1 (рис. 7).

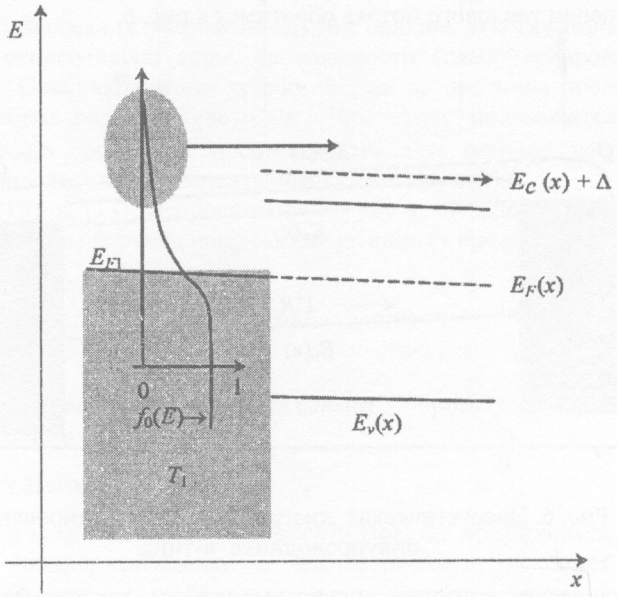


Рис. 7. К механизму поглощения тепла на левом контакте 1.

Электроны с энергией большей, чем высота барьера φ_B (рис. 3) термически эмитируются из металла в полупроводник. Металл покидают высокоэнергетические электроны (серый овал на рис. 7), нарушая при этом равновесное фермиевское распределение. Процесс перехода сопровождается электрон-фононным рассеянием с поглощением тепла проводником и охлаждением контакта. Энергия поглощается из решетки металлического контакта, поднимая энергию низкоэнергетических электронов выше и восстанавливая равновесное фермиевское распределение. Этот процесс напоминает испарение жидкости, в котором роль жидкости играют электроны в металле.

Вычислим поток тепла, сопровождающий поток электронов. Поток электронов в направлении положительной полуоси x есть $J_x/(-q)>0$, поскольку $J_x<0$. Каждый электрон, переходящий из контакта 1 в проводник, переносит тепловую энергию в количестве $Q=E_C(0)+\Delta-E_{F1}$. Таким образом, поток тепла будет

$$J_{Q1}=[E_C(0)+\Delta-E_{F1}]\times J_x/(-q)=\pi J_x, \tag{18}$$

где коэффициент Пельтье

$$\pi=-[E_C(0)+\Delta-E_{F1}]/q. \tag{19}$$

Коэффициент Пельтье отрицательный для *n*-проводников. Сравнивая (19) и (14), получаем

$$\pi = T_1 S. \tag{20}$$

Аналогичное выражение получается для контакта 2 с заменой T_1 на T_2. Эта связь двух тепловых коэффициентов – Зеебека и Пельтье (9) известна как соотношение Кельвина.

Мы получили первое слагаемое в уравнении (8). Во втором слагаемом фигурирует электронная теплопроводность в условиях незамкнутой электрической цепи κ. Выражение для нее получим позже. Сейчас же уместно предположить, что электронная теплопроводность κ и удельная электронная проводимость σ связаны между собой, поскольку поток электронов сопровождается потоком тепла. Мы позже получим следующее соотношение между ними:

$$\frac{\kappa}{\sigma} = LT, \tag{21}$$

где L называют числом Лоренца, а соотношение (21) – законом Видемана – Франца [9]. Они не столь фундаментальны как соотношение Кельвина (9), поскольку зависят от особенностей зонной структуры и характера рассеяния в конкретном проводнике [10]. Для типичного полупроводника с параболической дисперсией и с постоянным средним значением длины свободного пробега

$$L \approx 2 \times \left(\frac{k}{q}\right)^2, \text{ (невырожденный проводник)}$$
$$L \approx \frac{\pi^2}{3} \times \left(\frac{k}{q}\right)^2. \text{ (вырожденный проводник)} \tag{22}$$

Термин «закон Видемана – Франца» обычно подразумевает формулу (21) с числами Лоренца (22). У проводников меньшей размерности числа Лоренца совсем не похожи на (22). Общее выражение для чисел Лоренца L получим позже.

Итак, основными уравнениями ТЭ служат уравнения (5), (8) и (9) с четырьмя транспортными коэффициентами: 1) удельное сопротивление $\rho = 1/\sigma$; 2) коэффициент Зеебека S; 3) коэффициент Пельтье π; 4) электронная теплопроводность κ.

Электрическая проводимость σ подробно обсуждалась в главе 1. Для массивного 3D проводника в диффузионном режиме для удельной и дифференциальной проводимости σ' имеем:

$$\sigma = \int \sigma'(E) dE = \frac{2q^2}{h} \langle M_{3D} \rangle \langle\langle \lambda \rangle\rangle, \tag{23}$$

$$\sigma'(E) = \frac{2q^2}{h} M_{3D}(E) \lambda(E) \left(-\frac{\partial f_0}{\partial E}\right), \left[1/Ом \cdot м \cdot Дж\right] \tag{24}$$

где $M_{3D}(E)$ – число мод проводимости при энергии E на единицу площади поперечного сечения проводника.

Коэффициент Зеебека дается уравнением (15) с безразмерным параметром

$$\delta = \left(E_{av} - E_C\right)/kT, \tag{25}$$

определяющим среднюю энергию, с которой электрон движется по проводнику, относительно дна зоны проводимости. Поскольку дифференциальная проводимость определяет распределение тока по энергии, то

$$\delta = \frac{1}{kT}\left(\frac{\int (E - E_C)\sigma'(E)dE}{\int \sigma'(E)dE}\right). \tag{26}$$

Мы также знаем, что коэффициент Пельтье связан с коэффициентом Зеебека соотношением Кельвина

$$\pi(T) = T \cdot S(T). \tag{27}$$

Уравнения (5) и (8) являются частным случаем более общих уравнений связанных потоков [11]. В нашем случае температурный градиент порождает электрический ток, а он в свою очередь порождает поток тепла. Перекрестные коэффициенты S и π фундаментально связаны друг с другом соотношением Кельвина, которое является частным случаем более общих соотношений Онзагера [12].

Мы также имеем уравнение для электронной теплопроводности

$$\kappa = L\sigma T, \tag{28}$$

которое учитывает перенос тепла только электронами. В металлах большую часть тепла переносят электроны, а в полупроводниках – фононы, к рассмотрению которых в транспортной модели ЛДЛ мы вернемся позже в этой главе.

2.1.3. Термоэлектрические устройства

Пионерские теоретические и прикладные исследования А. Ф. Иоффе в 50 – 60 годы в Ленинграде [3] сыграли решающую роль в развитии физики ТЭ явлений. С конца 60-х годов начала формироваться научная школа в области термоэлектричества в Черновицком университете, возглавляемая Л.И.Анатычуком [4 – 8, 13]. Последние 30 лет характеризуются неуклонным прогрессом в области физики ТЭ явлений и ее многочисленных приложений – от переносных холодильников для пикника до генераторов тока для дальних космических аппаратов. В последние годы надежды на улучшение показателей эффективности работы ТЭ устройств и показателей качества термиков

возлагают на наноструктурированные материалы [14 – 16]. Так это или не так покажут только экспериментальные исследования.

Принципиальная схема ТЭ охладителя показана на рис. 8.

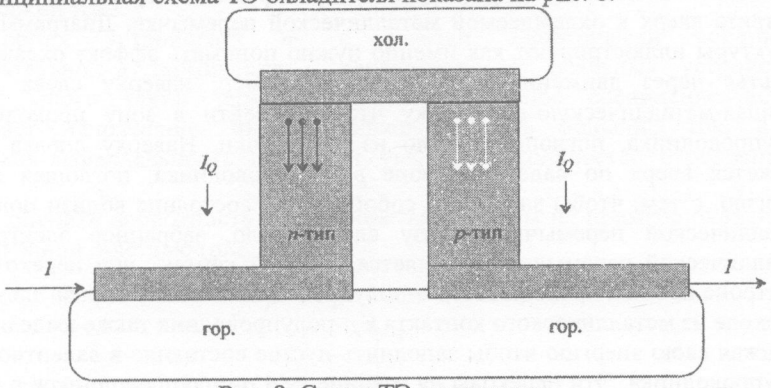


Рис. 8. Схема ТЭ охладителя.

Ток подается на *n*- и *p*-ветви, которые соединены последовательно металлической перемычкой. Электроны и дырки движутся сверху вниз, унося с собой тепло из перемычки через переходы металл/полупроводник.

Каким образом следует обсуждать работу ТЭ устройства в терминах только электронных потоков, а не потоков электронов и «дырок», как это проводилось до сих пор. «Дырки» являются, в лучшем случае, концептуальным понятием, а измеряемые на практике эффекты вызваны движением электронов и не могут зависеть от субъективных, хотя и очень удобных, договоренностей. Не говоря уже о том, что вся развитая в [1, 2] транспортная модель ЛДЛ выписана для электронов, хотя и было показано как трансформировать ее для «дырок».

Электронные потоки в ТЭ охладителе показаны на рис. 9.

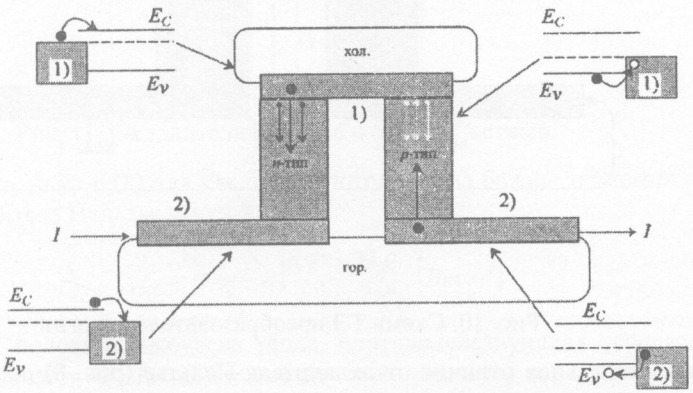


Рис. 9. Движение электронов в ТЭ охладителе (рис. 8) в *n*- и *p*-ветвях показано темными стрелками.

По левой *n*-ветви электроны движутся сверху от охлаждаемой металлической перемычки вниз к нагреваемому контакту, через который ток подается в охладитель. В правой *p*-ветви электроны движутся снизу от правого контакта вверх к охлаждаемой металлической перемычке. Диаграммы зонной структуры иллюстрируют как именно нужно понимать эффект охлаждения по Пельтье через движение электронов. Например, наверху слева электрон, покидая металлическую перемычку чтобы перейти в зону проводимости *n*-полупроводника, поглощает тепло из перемычки. Наверху справа электрон движется вверх по валентной зоне *p*-полупроводника, поглощая при этом энергию, с тем, чтобы заполнить собой пустое состояние вблизи поверхности металлической перемычки. Внизу слева тепло, забранное электроном из металлической перемычки, выделяется в левый контакт при переходе в него электрона из зоны проводимости *n*-полупроводника. Внизу справа электрон при переходе из металлического контакта в *p*-полупроводник также выделяет тепло, понижая свою энергию чтобы заполнить пустое состояние в валентной зоне *p*-полупроводника. Эти переходы на границе *n*- и *p*-полупроводников с металлом можно обсуждать через потоки электронов, либо через потоки электронов и дырок, как это кому удобно.

Для конкретного устройства охлаждения надлежит ответить на следующие очевидные вопросы: 1) Какая максимальная разница температур может быть достигнута? 2) Какое количество тепла можно забрать? 3) Что является показателем эффективности охладителя Пельтье?

На рис. 10 показана схема ТЭ преобразователя энергии.

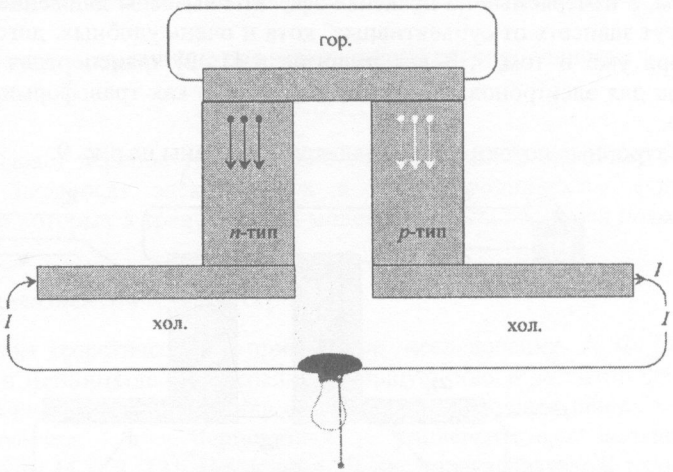


Рис. 10. Схема ТЭ преобразователя энергии.

Принципиальное отличие от охладителя Пельтье (рис. 8) состоит лишь в том, что теперь тепло подается извне, а подложка охлаждается, в результате чего во внешней цепи (на рисунке с нагрузкой) возникает ток Зеебека.

Направление движения носителей тока точно такое же, как и в охладителе Пельтье. Основной вопрос: что определяет эффективность конвертации тепла в электрический ток?

В обоих устройствах *n*- и *p*-ветви электрически соединены последовательно, а термически – параллельно. В реальных устройствах создается много таких термопар. Последовательное соединение ветвей увеличивает напряжение, что позволяет передать больший ток, а их термическая параллельность позволяет увеличить количество переносимого тепла.

Как оценить добротность ТЭ устройства покажем на примере охладителя Пельтье с одной *n*-ветвью (рис. 11).

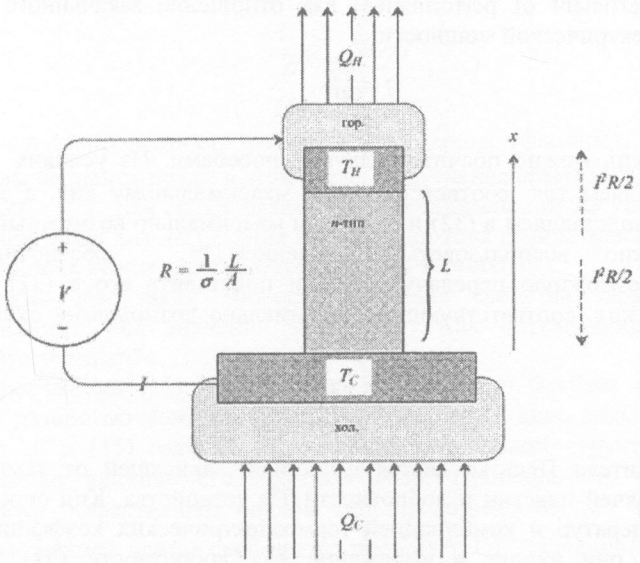


Рис. 11. Охладитель Пельтье с одной *n*-ветвью.

Опираясь на ур-е (8) для удельного потока тепла баланс тепловых потоков такого охладителя Пельтье имеет вид:

$$Q_C = \pi\frac{I}{A} - \kappa\frac{\Delta T}{L} - \frac{I^2R}{2A},\ [Bm/м^2] \qquad (29)$$

где учтена половина джоулева тепла, противодействующая передаче тепла снизу вверх, а $\Delta T = T_H - T_C$. Максимальное количество передаваемого тепла Q_C^{max} находится из условия $dQ_C / dI = 0$, учитывая которое сначала находим I^{max}, а после подстановки его в (29) находим Q_C^{max}. Для определения

максимально достижимой разности температур ΔT^{max} положим $Q_C^{max} = 0$ и найдем, что

$$\Delta T^{max} = \frac{1}{2} Z T_C^2, \tag{30}$$

где

$$Z = \frac{S^2 \sigma}{\kappa} [K^{-1}] \tag{31}$$

есть добротность ТЭ устройства (thermoelectric figure of merit) – важнейший показатель эффективности работы ТЭ устройства. Максимальная разность температур достигается при токе I^{max} и $Q_C^{max} = 0$.

Следующий вопрос заключается в определении эффективности охлаждения, определяемой коэффициентом полезного действия (кпд) устройства (coefficient of performance) как отношение закачанного тепла к затраченной электрической мощности:

$$\eta = \frac{Q_C}{P_{in}}. \tag{32}$$

Значение кпд можно посчитать двумя способами. Из условия $d\eta / dI = 0$ сначала определяем ток, соответствующий максимальному кпд, а затем это значение тока подставляем в (32) и получаем максимально возможный кпд. По другому, можно воспользоваться значением I^{max}, обеспечивающему максимально возможную передачу тепла, и подставить его в (32) и таким образом найти кпд, соответствующий максимально возможному охлаждению [17 – 19]:

$$\eta = \frac{Q_C}{P_{in}} = f_P(T_H, T_C, Z). \tag{33}$$

Кпд охладителя Пельтье дается функцией, зависящей от температуры холодной и горячей пластин и добротности ТЭ устройства. Кпд определяется разностью температур и комбинацией термоэлектрических коэффициентов в том виде, как они входят в выражение для добротности (31): бóльшая проводимость понижает потери джоулева тепла, бóльший коэффициент Зеебека увеличивает количество отобранного тепла, меньшая теплопроводность уменьшает возврат тепла от горячей пластины к охлаждаемой.

Аналогичные рассуждения можно провести для ТЭ преобразователя энергии (рис. 9). Опять, записывается тепловой баланс для горячей пластины: поток тепла, который мы хотим превратить в электрический ток, равен передаваемому теплу Пельтье плюс тепло, диффундирующее от горячей пластины к холодной, минус половина противодействующего джоулева тепла. Значение кпд ТЭ генератора энергии будет просто отношение полученной мощности к затраченной:

$$\eta = \frac{P_{out}}{P_{in}} = \frac{I^2 R_L}{A Q_{in}}, \tag{34}$$

где R_L – сопротивление нагрузки. Ток определяется перепадом температур пластин. Максимально возможное значение кпд сводится к решению уравнения $d\eta / dR_L = 0$, которое дает оптимальное значение сопротивления нагрузки, после подстановки которого в (34) получаем искомый кпд. Как и в случае с охладителем Пельтье, кпд генератора энергии определяется добротностью ТЭ устройства Z.

И в случае охладителя и в случае генератора энергии кпд определяется добротностью ТЭ устройства, которая существенно определяется свойствами термиков. На практике контакты и интерфейсные соединения, как электрические, так и термические, так или иначе понижают кпд ТЭ устройства, и решающую роль играют свойства используемых термиков.

Осталось обсудить еще роль добротности ТЭ устройства в обеспечении эффективности его работы. Добротность обычно записывают в безразмерном виде:

$$ZT = \frac{S^2 \sigma T}{\kappa + \kappa_L}, \qquad (35)$$

где κ_L – теплопроводность решетки проводника (фононы), к обсуждению которой мы вернемся позже в этой главе. Длительное время не удается выйти за пределы $ZT \sim 1-3$. Для самых разнообразных применений ТЭ устройств как для охлаждения, так и для нагревания и генерации тока крайне желательно достичь $ZT > 3$. Это пока несбыточная мечта в физике термиков.

Обсудим два таких вопроса: 1) какие свойства термика являются определяющими для значения ZT? 2) Как оптимизировать ZT для выбранного термика?

Из выражения (15) следует, что коэффициент Зеебека определяется в основном разностью между фермиевской энергией и дном зоны проводимости. Параметр δ в (15) зависит от особенностей зонной структуры и физики рассеяния, но влияние его все-таки мало. Проводимость, согласно (23), определяется эффективным числом мод проводимости $\langle M \rangle$ и усредненным значением средней длины свободного пробега $\langle\langle \lambda \rangle\rangle$. Значение $\langle M \rangle$ тем больше, чем выше энергия Ферми в зоне проводимости. Для больших $\langle\langle \lambda \rangle\rangle$ рассеяние должно быть слабым (бóльшая подвижность). В знаменателе добротности (35) обычно $\kappa_L \gg \kappa$; теплопроводность решетки в обобщенной транспортной модели учтем ниже.

Характер зависимости коэффициента Зеебека и электронной проводимости от положения уровня Ферми показано на рис. 12.

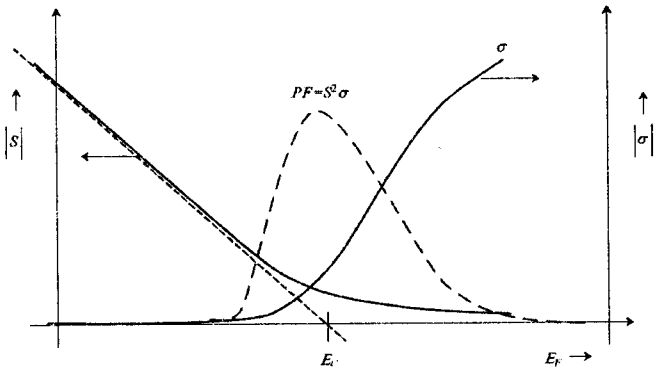


Рис. 12. Качественная зависимость коэффициента Зеебека (слева) и электронной проводимости (справа) от положения уровня Ферми относительно дна зоны проводимости. Показана также максимизация фактора мощности PF (power factor) вблизи дна зоны проводимости.

По мере того как уровень Ферми приближается к дну зоны проводимости снизу, а затем движется вверх по зоне проводимости, коэффициент Зеебека уменьшается. В то же время электронная проводимость растет за счет появления все большего числа мод проводимости. Их произведение называют фактором мощности PF, который максимален в районе дна зоны проводимости. Положение максимума для конкретного термика зависит от особенностей зонной структуры проводника и физики его рассеивающих центров. На практике стараются путем допирования полупроводника сместить уровень Ферми поближе к дну зоны проводимости.

Итак, на примере 3D резистора в диффузионном режиме мы обсудили физику эффектов Зеебека и Пельтье, качественно показали происхождение основных уравнений термоэлектричества, рассмотрели каким образом четыре ТЭ параметра зависят от свойств термиков и как работают ТЭ охладитель и генератор, ввели понятия добротности и кпд ТЭ устройств.

Для количественного описания ТЭ эффектов в резисторах любой размерности и любого масштаба, работающих в баллистическом, диффузионном или квазибаллистическом режиме, потребуется строгий вывод соответствующих транспортных уравнений, чем мы сейчас и займемся.

2.2. Термоэлектрические коэффициенты

Обобщенная транспортная модель применительно к нано- и микроэлектронике, изложенная в главе 1, на многочисленных примерах обработки экспериментальных данных показала свою эффективность при вычислении проводимости и других электрофизических свойств резисторов любой размерности, любого масштаба и произвольной дисперсии, работающих в баллистическом, квазибаллистическом или диффузионном режиме как вблизи $0°K$, так и при более высоких температурах. С позиций концепции «снизу –

вверх» выше рассмотрена физика термоэлектрических явлений Зеебека и Пельтье и качественно обсуждены закон Видемана – Франца, числа Лоренца и основные уравнения термоэлектричества с соответствующими транспортными коэффициентами.

Теперь получим строгие результаты, сформулированные выше без доказательства, для 1D, 2D и 3D проводников с произвольной дисперсией и любого масштаба в режимах от баллистического до диффузионного, включая переходные режимы. С подробными выкладками покажем как получаются термоэлектрические коэффициенты для 1D баллистического транспорта и для 3D диффузионного. Основным объектом обсуждения будут *n*-проводники, вместе с тем покажем каким образом распространить результаты на *p*-проводники. В рамках обобщенной транспортной модели рассмотрим биполярную проводимость, закон Видемана – Франца и числа Лоренца, соотношение Мотта.

Затем в справочных целях приведем сводку термоэлектрических коэффициентов для 1D, 2D и 3D полупроводников с параболической дисперсией в баллистическом режиме, а также в диффузионном со степенным законом рассеяния, выраженных через стандартные интегралы Ферми – Дирака.

2.2.1. «Движущие силы» электрического тока

Общее выражение для электрического тока, полученное в главе 1,

$$I = -I_x = \frac{2q}{h}\int T(E)M(E)\left(f_1 - f_2\right)dE \qquad (36)$$

утверждает, что любые причины, вызывающие различие между фермиевскими функциями контактов проводника, порождают ток, где $T(E)$ – коэффициент прохождения, а $M(E)$ – число мод проводимости. Знак «–» в этом выражении указывает, что ток считается положительным, если он входит в контакт 2 извне (рис. 1/гл. 1) в направлении отрицательной полуоси x, электроны при этом движутся по проводнику слева – направо от контакта 1 к контакту 2 в направлении положительной полуоси x.

Различают две «движущие силы», порождающих ток: разность потенциалов и разность температур контактов. Электроны переносят как заряд, так и тепло. Зарядовый ток дается уравнением (36). Чтобы получить общее выражение для термотока, рассуждаем следующим образом. Электроны внутри контакта движутся при энергии $E \approx E_F$. Чтобы зайти в моду проводимости $M(E)$ резистора со стороны истока S электрон должен выделить $(E_F > E)$, а со стороны стока D – поглотить $(E_F < E)$ тепловую энергию в количестве $|E - E_F|$. Выражение для термотока получается просто внесением $(E - E_F)/q$ под знак интеграла в (36), так что в итоге термоток

$$I_Q = \frac{2}{h}\int (E - E_F)T(E)M(E)\left(f_1 - f_2\right)dE. \qquad (37)$$

Уравнения (36) и (37) есть общие выражения для зарядового тока и для термотока. Далее сразу перейдем к рассмотрению тока в режиме линейного отклика $f_0(E) \approx f_1(E) > f_2(E) \approx f_0(E)$, который мы рассмотрели подробно в § 1.3.4.

На рис. 13 показан ход фермиевских функций $f_1 > f_2$ при одинаковой температуре контактов и $\Delta V = V_2 - V_1 > 0$, так что ток положительный.

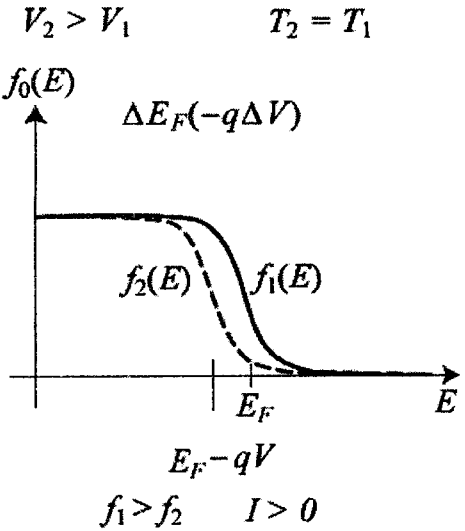


Рис. 13. Качественный ход фермиевских функций при приложении разности потенциалов к контактам, находящимся при одной и той же температуре.

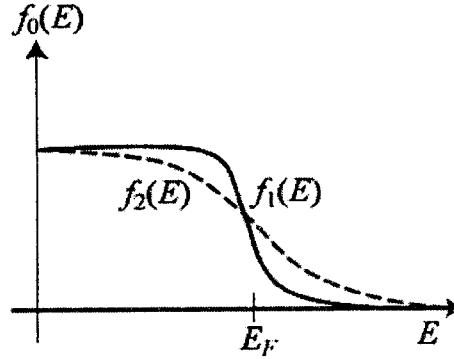
В этом случае знак тока не зависит от того, речь идет о *n*-проводнике или о *p*-проводнике. В режиме линейного отклика имеем:

$$f_1 - f_2 \approx \left(-\frac{\partial f_0}{\partial E}\right) q\Delta V, \qquad (38)$$

где фермиевское окно проводимости $(-\partial f_0/\partial E) \approx \pm\, 2kT$, охватывающее равновесное значение электрохимического потенциала $E = E_{F0}$, «фильтрует» электроны и пропускает в проводник только те из них, энергия которых лежит в пределах окна Ферми.

Рис. 14 показывает ситуацию, когда электрохимические потенциалы контактов одинаковы $(E_{F1} = E_{F2}: V_2 = V_1)$, а температура контактов разная с $\Delta T = T_2 - T_1 > 0$.

$$V_2 = V_1 \qquad T_2 > T_1$$



n-тип: $E > E_F$: $f_2 > f_1$

p-тип: $E < E_F$: $f_1 > f_2$

Рис. 14. Качественный ход фермиевских функций без приложения разности потенциалов к контактам, находящимся при разных температурах.

В этом случае $f_1 > f_2$ для энергий электронов ниже уровня Ферми $(E < E_F)$ и $f_1 < f_2$ для энергий электронов выше уровня Ферми $(E > E_F)$. Направление тока зависит от того, как расположены моды проводимости проводника: если выше уровня Ферми (*n*-проводник), ток положительный, если же ниже уровня Ферми (*p*-проводник), ток отрицательный (электроны движутся по проводнику справа – налево от контакта 2 к контакту 1). В режиме линейного отклика разность фермиевских функций разложим в ряд Тейлора и ограничимся членом первого порядка:

$$f_1 - f_2 \approx f_1 - \left(f_1 + \frac{\partial f_1}{\partial T}\Delta T \right) = \left(-\frac{\partial f_0}{\partial T} \right) \Delta T . \qquad (39)$$

Дифференцирование функции Ферми по температуре дает:

$$f_1 - f_2 \approx -\left(-\frac{\partial f_0}{\partial E} \right) \frac{E - E_F}{T} \Delta T , \qquad (40)$$

где $T = (T_1 + T_2)/2$.

Действие обеих движущих сил – и разности электрохимических потенциалов и разности температур контактов иллюстрируется рис. 15.

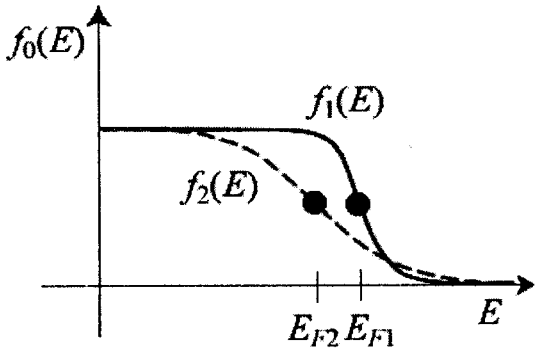


Рис. 15. Качественный ход фермиевских функций при приложении разности потенциалов к контактам, находящимся при разных температурах.

В режиме линейного отклика эффекты от обеих «сил» просто складываются:

$$f_1 - f_2 \approx \left(-\frac{\partial f_0}{\partial E}\right) q\Delta V - \left(-\frac{\partial f_0}{\partial E}\right) \frac{E - E_F}{T} \Delta T . \qquad (41)$$

Окончательное выражение для суммарного тока есть сумма вкладов от всех мод проводимости:

$$I = \int I'(E) dE , \qquad (42)$$

где ток в дифференциальной форме

$$I'(E) = \frac{2q}{h} T(E) M(E) \left(f_1 - f_2\right). \qquad (43)$$

Подставляя разность фермиевских функций через (41), получаем

$$I'(E) = G'(E)\Delta V + S'_T(E)\Delta T , \qquad (44)$$

где

$$G'(E) = \frac{2q^2}{h} T(E) M(E) \left(-\frac{\partial f_0}{\partial E}\right) \qquad (45)$$

есть дифференциальная проводимость, а

$$S'_T(E) = -\frac{2q^2}{h} T(E) M(E) \left(\frac{E - E_F}{qT}\right) \left(-\frac{\partial f_0}{\partial E}\right) = -\frac{k}{q} \left(\frac{E - E_F}{kT}\right) G'(E) \qquad (46)$$

есть коэффициент Соре термодиффузии в дифференциальной форме. Видно, что он отрицательный для энергий электронов выше энергии Ферми

(*n*-проводники) и положительный для энергий электронов ниже энергии Ферми (*p*-проводники).

Остается проинтегрировать дифференциальный ток (7). Окончательно получаем:

$$I = G\Delta V + S_T \Delta T, \tag{47}$$

где проводимость

$$G = \int G'(E)dE, \tag{48}$$

коэффициент Соре термодиффузии

$$S_T = \int S_T'(E)dE = -\frac{k}{q}\int\left(\frac{E-E_F}{kT}\right)G'(E)dE\,.\,[A/K] \tag{49}$$

Оба эти уравнения справедливы для проводников любой размерности и любого масштаба в режимах от баллистического через квазибаллистический до диффузионного.

2.2.2. Ток в массивных проводниках

Ток в массивных проводниках в режиме диффузионного транспорта описывается уравнениями (1) – (3). Из них уравнение (2) для плотности тока, выраженной через электрическое поле (3), справедливо для однородных проводников без учета коэффициента диффузии. Как переписать эти уравнения с учетом перепада температур на контактах?

Воспользуемся уравнением (47) и напомним, что в этом уравнении ток считается положительным, если он направлен в сторону отрицательной полуоси x, от правого контакта 2 к левому контакту 1. Делим ток на площадь поперечного сечения проводника со знаком «минус» и получаем плотность тока в направлении положительной полуоси x:

$$J_x = -\frac{G}{A}\Delta V - \frac{S_T}{A}\Delta T\,. \tag{50}$$

Разделим и умножим это уравнение на длину проводника L. Тогда

$$J_x = -G\frac{L}{A}\frac{\Delta V}{L} - S_T\frac{L}{A}\frac{\Delta T}{L}. \tag{51}$$

В диффузионном режиме

$$G = \sigma\frac{A}{L} \tag{52}$$

и по аналогии

$$S_T = s_T\frac{A}{L}, \tag{53}$$

где s_T есть удельное значение термодиффузионного коэффициента Соре. С учетом также того, что $-q\Delta V/L \approx d(E_F)/dx$ и $\Delta T/L \approx dT/dx$, получается искомое уравнение для плотности тока в массивном проводнике с учетом градиента не только электрохимического потенциала, но и температуры:

$$J_x = \sigma \frac{d\left(E_F/q\right)}{dx} - s_T \frac{dT}{dx}. \; [A/м^2] \tag{54}$$

Для *n*-проводников проводимость положительна, а коэффициент Соре отрицательный. Для *p*-проводников и электронная проводимость и коэффициент Соре положительны. К подробному обсуждению проводимости в *n*- и *p*-проводниках еще вернемся позже.

2.2.3. Перенос тепла электронами

Потоки тепла на контактах определяются уравнением (37). Перепишем его в дифференциальной форме

$$I'_Q(E) = \frac{2}{h}(E - E_F)T(E)M(E)\left(f_1 - f_2\right), \tag{55}$$

так что

$$I_Q = \int I'_Q(E)dE. \tag{56}$$

В режиме линейного отклика для разности фермиевских функций воспользуемся выражением (41), а также выражениями (45) и (46) для дифференциальных проводимости $G'(E)$ и коэффициента Соре $S'_T(E)$. Тогда

$$I'_Q(E) = -T S'_T(E)\Delta V - K'_0(E)\Delta T, \tag{57}$$

где

$$K'_0(E) = \frac{(E - E_F)^2}{q^2 T} G'(E). \tag{58}$$

Интегрирование по всему спектру энергий дает:

$$I_Q = -T S_T \Delta V - K_0 \Delta T, \; [Bm] \tag{59}$$

где

$$K_0 = T\left(\frac{k}{q}\right)^2 \int \left(\frac{E - E_F}{kT}\right)^2 G'(E)dE \; [Bm/K] \tag{60}$$

есть электронная теплопроводность в условиях короткозамкнутой цепи $(\Delta V = 0)$.

Итак, из общих уравнений для электрического тока (36) и термотока (37) получены соответствующие выражения (47) и (59) для режима линейного отклика. Эти выражения показывают, что различие в электрохимических потенциалах и температуре контактов ведет к электрическому току и переносу тепла электронами. Этому режиму соответствуют три транспортных параметра: проводимость G (48), коэффициент термодиффузии Соре S_T (49) и электронная теплопроводность K_0 (60). Они справедливы для проводников любой размерности и для режимов от баллистического до диффузионного и всех промежуточных.

Для 3D-проводника в диффузионном режиме электрический ток описывается транспортным уравнением (54). Плотность теплового потока дается уравнением

$$J_{Qx} = T s_T \frac{d(E_F/q)}{dx} - \kappa_0 \frac{dT}{dx}, \ [Вт/м^2] \tag{61}$$

а соответствующие этим двум уравнениям (54) и (61) транспортные параметры включают удельную проводимость σ и удельные значения термодиффузионного коэффициента Соре и удельную теплопроводность, выраженные через дифференциальную проводимость $\sigma'(E)$:

$$\sigma = \int \sigma'(E) dE = \frac{2q^2}{h} \left\langle M_{3D} \right\rangle \left\langle\left\langle \lambda \right\rangle\right\rangle,$$

$$\sigma'(E) = \frac{2q^2}{h} M_{3D}(E) \lambda(E) \left(-\frac{\partial f_0}{\partial E} \right), \ [1/Ом \cdot м \cdot Дж]$$

$$s_T = -\frac{k}{q} \int \left(\frac{E - E_F}{kT} \right) \sigma'(E) dE, \ [А/м \cdot К] \tag{62}$$

$$\kappa_0 = T \left(\frac{k}{q} \right)^2 \int \left(\frac{E - E_F}{kT} \right)^2 \sigma'(E) dE. \ [Вт/м \cdot К] \tag{63}$$

Отметим, что уравнения (54) и (61) справедливы и для 1D и 2D проводников, но единицы измерения будут другими.

2.2.4. Обращенная форма транспортных уравнений

Пары уравнений для токов (47) и (59) и для плотностей (54) и (61) получились естественным образом из общих уравнений (36) и (37) применительно к режиму линейного отклика. Они соответствуют экспериментальной ситуации, когда для заданной разности потенциалов и температур измеряются токи и потоки тепла. Если первые могут быть заданы независимо друг от друга, то потоки электронов и тепла взаимозависимы. В таком представлении вклады от всех мод проводимости складываются друг с

другом. С точки зрения проведения экспериментов часто бывает удобным переписать эти уравнения таким образом, чтобы электрический ток и разность температур стали независимыми величинами. Обращенные таким образом уравнения (47) и (59) таковы:

$$\Delta V = RI - S\Delta T, \tag{64}$$

$$I_Q = -\Pi I - K\Delta T, \tag{65}$$

где

$$S = S_T / G, \tag{66}$$

$$\Pi = TS, \tag{67}$$

$$K = K_0 - \Pi S G. \tag{68}$$

В этом представлении вклады от разных мод проводимости не суммируются, например, сопротивление $R \neq \int R(E)dE$.

Для 3D проводника в диффузионном режиме обращение уравнений (54) и (61) дает следующую пару уравнений:

$$\frac{d(E_F/q)}{dx} = \rho J_x + S\frac{dT}{dx}, \tag{69}$$

$$J_{Qx} = TSJ_x - \kappa\frac{dT}{dx} \tag{70}$$

с транспортными коэффициентами

$$\rho = 1/\sigma, \tag{71}$$

$$S = s_T / \sigma, \tag{72}$$

$$\kappa = \kappa_0 - S^2\sigma T, \tag{73}$$

Часто можно встретить уравнение (69) с левой частью, записанной через электрическое поле.

2.2.5. Транспортные коэффициенты для 1D проводников

Рассмотрим 1D проводник в баллистическом режиме с одной занятой подзоной. Вычислим его транспортные коэффициенты из общих уравнений (47) и (59) или (64) и (65). В рассматриваемом случае коэффициент прохождения $T(E) = 1$ и число мод проводимости $M(E) = g_v$, где g_v есть долинное вырождение. Хотя мы и предполагаем, что занята только одна подзона, но в зависимости от зонной структуры g_v может отличаться от единицы. Например, в углеродных нанотрубках, часто рассматриваемых в качестве идеальных 1D

проводников, $g_v = 2$. В случае 1D проводников все, что нам нужно знать об их зонной структуре, это только значение долинного вырождения.

Для вычисления проводимости сначала обратимся к дифференциальной проводимости (45). В нашем случае

$$G'(E) = \frac{2q^2}{h} g_v \left(-\frac{\partial f_0}{\partial E} \right), \tag{74}$$

так что проводимость

$$G = \frac{1}{R} = \frac{2q^2}{h} g_v \int_{E_C}^{\infty} \left(-\frac{\partial f_0}{\partial E} \right) dE = \frac{2q^2}{h} \langle M \rangle. \tag{75}$$

Интеграл в этом выражении есть число мод проводимости, которое в случае полного вырождения ($T = 0°\ K$) $\langle M \rangle = g_v$. В общем же случае

$$\langle M \rangle = g_v \int_{E_C}^{\infty} \left(-\frac{\partial f_0}{\partial E} \right) dE = g_v \frac{\partial}{\partial E_F} \int_{E_C}^{\infty} f_0 \, dE, \tag{76}$$

где использовано свойство фермиевской функции

$$-\frac{\partial f_0}{\partial E} = +\frac{\partial f_0}{\partial E_F}. \tag{77}$$

Для вычисления правого интеграла в (76)

$$\langle M \rangle = g_v \frac{\partial}{\partial E_F} \int_{E_C}^{\infty} \frac{dE}{e^{(E-E_F)/kT} + 1} \tag{78}$$

сделаем замену переменных

$$\eta \equiv (E - E_C)/kT; \quad \eta_F \equiv (E_F - E_C)/kT, \tag{79}$$

так что

$$\langle M \rangle = g_v \frac{\partial}{\partial \eta_F} \int_0^{\infty} \frac{d\eta}{e^{\eta - \eta_F} + 1} \tag{80}$$

выражается через интеграл Ферми – Дирака (§ 1.3.6.4) 0-го порядка $\Im_0(\eta_F)$. Дифференцирование его по параметру дает

$$\langle M \rangle = g_v \Im_{-1}(\eta_F), \tag{81}$$

так что искомая проводимость

$$G = \frac{2q^2}{h} \langle M \rangle = \frac{2q^2}{h} g_v \Im_{-1}(\eta_F). \tag{82}$$

Для невырожденных проводников $\eta_F \ll 0$ и $\Im_{-1}(\eta_F) \to e^{\eta_F}$. Для сильно вырожденных проводников $\eta_F \gg 0$. Интеграл Ферми – Дирака $\Im_{-1}(\eta_F)$ как аналитическая функция

$$\Im_{-1}(\eta_F) = \frac{\partial \Im_0(\eta_F)}{\partial \eta_F} = \frac{\partial \left[\ln\left(e^{\eta_F}+1\right)\right]}{\partial \eta_F} = \frac{e^{\eta_F}}{e^{\eta_F}+1}. \qquad (83)$$

Для сильно вырожденного проводника ($\eta_F \gg 0$) $\Im_{-1}(\eta_F) \to 1$ и

$$G = \frac{2q^2}{h} g_v, \qquad (84)$$

как и ожидалось.

Зная G и $R = 1/G$, вычислим теперь коэффициенты Соре и Зеебека. Из (49) и (74) следует, что

$$S_T = -\frac{k}{q} \int_{E_C}^{\infty} \left(\frac{E - E_F}{kT}\right) \left[\frac{2q^2}{h} g_v \left(-\frac{\partial f_0}{\partial E}\right)\right] dE. \qquad (85)$$

Разделим и умножим это выражение на проводимость согласно (75). Тогда

$$S_T = -\frac{k}{q} G \frac{\int_{E_C}^{\infty} \left[\left(E - E_F\right)/kT\right]\left(-\partial f_0/\partial E\right) dE}{\int_{E_C}^{\infty} \left(-\partial f_0/\partial E\right) dE}. \qquad (86)$$

Коэффициент Соре пропорционален проводимости. Остается вычислить два интеграла. Интеграл в знаменателе равен $\Im_{-1}(\eta_F)$. Это очевидно из сравнения (81) с (76). Интеграл в числителе преобразуем следующим образом:

$$\int_{E_C}^{\infty} \left[\left(E - E_F\right)/kT\right]\left(-\partial f_0/\partial E\right) dE = \int_{E_C}^{\infty} \left[\left(E - E_C + E_C - E_F\right)/kT\right]\left(-\partial f_0/\partial E\right) dE =$$

$$= \int_{E_C}^{\infty} \left(E - E_C/kT\right)\left(+\partial f_0/\partial E_F\right) dE - \eta_F \int_{E_C}^{\infty} \left(+\partial f_0/\partial E_F\right) dE, \qquad (87)$$

где использовалось равенство (77). Выносим $\partial/\partial E_F$ из под знака интеграла и делаем замену переменных (79). Тогда числитель в (86) оказывается равным

$$\int_{E_C}^{\infty} \left[\left(E - E_F\right)/kT\right]\left(-\partial f_0/\partial E\right) dE =$$

$$= \frac{\partial}{\partial \eta_F} \int_0^{\infty} \frac{\eta\, d\eta}{e^{\eta - \eta_F} + 1} - \eta_F \frac{\partial}{\partial \eta_F} \int_0^{\infty} \frac{d\eta}{e^{\eta - \eta_F} + 1} = \Im_0(\eta_F) - \eta_F \Im_{-1}(\eta_F). \qquad (88)$$

Окончательно коэффициент Соре (86)

$$S_T = -\frac{k}{q}\left(-\eta_F + \frac{\Im_0(\eta_F)}{\Im_{-1}(\eta_F)}\right)G. \tag{89}$$

Поскольку коэффициенты Соре и Зеебека связаны друг с другом через проводимость соотношением (66), то коэффициент Зеебека

$$S = -\frac{k}{q}\left(-\eta_F + \frac{\Im_0(\eta_F)}{\Im_{-1}(\eta_F)}\right) \tag{90}$$

или иначе

$$S = -\frac{k}{q}\left(\frac{E_C - E_F}{kT} + \delta\right), \tag{91}$$

где

$$\delta = \frac{\Im_0(\eta_F)}{\Im_{-1}(\eta_F)}. \tag{92}$$

Коэффициент Зеебека получился таким же, как и ожидалось, с

$$\delta = \Delta / kT. \tag{93}$$

В невырожденном пределе ($\eta_F \ll 0$) интегралы Ферми – Дирака становятся просто экспонентами с $\delta \to 1$ и $\Delta \to kT$.

В невырожденном 1D баллистическом резисторе ток идет по единственной моде с энергией приблизительно на kT выше дна зоны проводимости. Для сильно вырожденных резисторов ($\eta_F \gg 0$) $\delta \to \eta_F = (E_F - E_C)/kT$, так что согласно (91) $S \to 0$, а $\Delta \to (E_F - E_C)$, что намного больше кванта тепловой энергии kT, характерного для невырожденных резисторов.

Итак, для 1D баллистического резистора вычислены G, R, S_T, S, Π. Перейдем к вычислению электронной теплопроводности K_0 и K. Следуем той же процедуре, что и при вычислении коэффициента Соре. После умножения и деления теплопроводности K_0 (60) на проводимость (75) имеем:

$$K_0 = T\left(\frac{k}{q}\right)^2 G \frac{\int_{E_C}^{\infty} [(E - E_F)/kT]^2 \left(-\partial f_0 / \partial E\right) dE}{\int_{E_C}^{\infty} \left(-\partial f_0 / \partial E\right) dE}. \tag{94}$$

Сразу видим, что теплопроводность пропорциональна проводимости (закон Видемана – Франца). Как и в коэффициенте Соре (86), интеграл в знаменателе равен $\Im_{-1}(\eta_F)$. Остается разобраться с интегралом в числителе. Он равен

$$\int_{E_C}^{\infty}[(E-E_F)/kT]^2\left(-\partial f_0/\partial E\right)dE=$$

$$=\int_{E_C}^{\infty}[(E-E_C)+(E_C-E_F)/kT]^2\left(-\partial f_0/\partial E\right)dE=$$

$$=\int_{E_C}^{\infty}[(E-E_C)/kT]^2\left(+\partial f_0/\partial E_F\right)dE- \qquad (95)$$

$$-2\eta_F\int_{E_C}^{\infty}[(E-E_C)/kT]\left(+\partial f_0/\partial E_F\right)dE+$$

$$+(\eta_F)^2\int_{E_C}^{\infty}\left(+\partial f_0/\partial E_F\right)dE.$$

Последнее слагаемое есть $(\eta_F)^2\Im_{-1}(\eta_F)$, второе слагаемое равно $-2\eta_F\Im_0(\eta_F)$, а первое есть $2\Im_1(\eta_F)$, так что электронная теплопроводность в условиях короткозамкнутой цепи

$$K_0=T\left(\frac{k}{q}\right)^2 G\left[2\frac{\Im_1(\eta_F)}{\Im_{-1}(\eta_F)}-2\eta_F\frac{\Im_0(\eta_F)}{\Im_{-1}(\eta_F)}+(\eta_F)^2\right]. \qquad (96)$$

Электронная теплопроводность в условиях незамкнутой цепи K дается выражением (68) и с учетом (67) и (90) окончательно имеем:

$$K=T\left(\frac{k}{q}\right)^2 G\left[2\frac{\Im_1(\eta_F)}{\Im_{-1}(\eta_F)}-\left(\frac{\Im_0(\eta_F)}{\Im_{-1}(\eta_F)}\right)^2\right]. \qquad (97)$$

Аналогичным образом можно получить аналитические выражения для термоэлектрических параметров для резисторов любой размерности, любого масштаба и любых режимов работы. Если подобные выкладки кажутся слишком скучными, вполне можно прибегнуть к численному интегрированию, что обычно достаточно для сравнения с экспериментальными данными.

Ниже в § 2.2.11 мы приведем окончательные аналитические выражения для термоэлектрических коэффициентов резисторов с параболической дисперсией всех размерностей, в том числе и для случая диффузионного режима со степенным законом рассеяния

$$\lambda(E)=\lambda_0\left(\frac{E-E_C}{kT}\right)^r. \qquad (98)$$

Аналогичные аналитические выражения приведем в § 3.11 и для линейной дисперсии в графене, которому будет посвящена глава 3.

2.2.6. Закон Видемана – Франца и числа Лоренца

Электроны переносят как заряд, так и тепло. Выпишем связь между электронной проводимостью и теплопроводностью в явном виде. Для этого нам потребуется процедура усреднения по удельной проводимости

$$\sigma = \int \sigma'(E) dE, \tag{99}$$

а именно:

$$\langle (*) \rangle \equiv \frac{\int (*) \sigma'(E) dE}{\int \sigma'(E) dE}. \tag{100}$$

Умножаем и делим удельное значение коэффициента Соре (62) на удельную проводимость. Тогда

$$s_T = -\left(\frac{k}{q}\right) \sigma \langle \frac{E - E_F}{kT} \rangle. \tag{101}$$

Аналогично поступаем с коэффициентом Зеебека (72). Тогда

$$S = \frac{s_T}{\sigma} = -\left(\frac{k}{q}\right) \langle \frac{E - E_F}{kT} \rangle = -\left(\frac{k}{q}\right) \frac{E_{av} - E_F}{kT} \tag{102}$$

и мы видим, что коэффициент Зеебека пропорционален среднему значению энергии выше уровня Ферми, при которой течет ток.

Для удельной теплопроводности (63) и (73) имеем:

$$\kappa_0 = T \left(\frac{k}{q}\right)^2 \sigma \langle (\frac{E - E_F}{kT})^2 \rangle, \tag{103}$$

$$\kappa = T \left(\frac{k}{q}\right)^2 \sigma \left[\langle (\frac{E - E_F}{kT})^2 \rangle - \langle \frac{E - E_F}{kT} \rangle^2 \right]. \tag{104}$$

Обе теплопроводности пропорциональны электронной проводимости. Отсюда

$$\frac{\kappa}{T\sigma} \equiv L = \left(\frac{k}{q}\right)^2 \left[\langle (\frac{E - E_F}{kT})^2 \rangle - \langle \frac{E - E_F}{kT} \rangle^2 \right], \tag{105}$$

есть закон Видемана – Франца, а L – число Лоренца. Множитель в квадратных скобках зависит от профиля зон, вырожденности и характера рассеяния. Для типичного полупроводника с параболической дисперсией и с постоянным средним значением длины свободного пробега этот множитель приблизительно равен 2 для невырожденного проводника или $\pi^2/3$ для вырожденного (§ 2.1.2).

2.2.7. Полупроводники *p*-типа и биполярная проводимость

Рассмотрим 3D проводник с параболической дисперсией. Для зоны проводимости

$$M_{3D}(E) = g_v \frac{m^*}{2\pi\hbar^2}(E - E_C), \quad (E \geq E_C) \tag{106}$$

а для валентной зоны

$$M_{3D}^{(v)}(E) = g_v \frac{m_p^*}{2\pi\hbar^2}(E_V - E). \quad (E \leq E_V). \tag{107}$$

Проводимость обеспечивается двумя вкладами. Со стороны зоны проводимости

$$\sigma = \frac{q^2}{h} \int_{E_C}^{\infty} M_{3D}(E)\lambda(E)\left(-\frac{\partial f_0}{\partial E}\right)dE \tag{108}$$

и со стороны валентной зоны

$$\sigma_p = \frac{q^2}{h} \int_{-\infty}^{E_V} M_{3D}^v(E)\lambda_p(E)\left(-\frac{\partial f_0}{\partial E}\right)dE. \tag{109}$$

Коэффициент Зеебека для электронов зоны проводимости следует из (62):

$$\begin{aligned} \sigma &= \int_{E_C}^{\infty} \sigma'(E)dE, \\ \sigma'(E) &= \frac{2q^2}{h} M_{3D}(E - E_C)\lambda(E)\left(-\frac{\partial f_0}{\partial E}\right), \\ s_T &= -\frac{k}{q} \int_{E_C}^{\infty} \left(\frac{E - E_F}{kT}\right)\sigma'(E)dE, \\ S &= s_T / \sigma, \end{aligned} \tag{110}$$

и по аналогии для электронов валентной зоны

$$\begin{aligned} \sigma_p &= \int_{-\infty}^{E_V} \sigma_p'(E)dE, \\ \sigma_p'(E) &= \frac{2q^2}{h} M_{3D}^v(E_V - E)\lambda_p(E)\left(-\frac{\partial f_0}{\partial E}\right), \\ s_T^{(v)} &= -\frac{k}{q} \int_{-\infty}^{E_V} \left(\frac{E - E_F^{(p)}}{kT}\right)\sigma_p'(E)dE, \\ S_p &= s_T^{(v)} / \sigma_p, \end{aligned} \tag{111}$$

причем знак S_p положительный.

В чем проявляются особенности когда обе зоны дают вклад в проводимость? Это встречается в узкозонных полупроводниках или при высоких температурах. В этом случае мы просто интегрируем по всем модам проводимости:

$$\sigma^{tot} \equiv \sigma + \sigma_p = \frac{q^2}{h} \int_{E_1}^{E_2} M_{3D}^{tot}(E)\,\lambda(E)\left(-\frac{\partial f_0}{\partial E}\right) dE, \qquad (112)$$
$$M_{3D}^{tot}(E) = M_{3D}(E) + M_{3D}^{v}(E);$$

при этом нет нужды заботиться об интегрировании до верхнего предела зоны проводимости или от самого дна валентной зоны, поскольку фермиевские функции гарантируют экспоненциальное падение до нуля на границах зон. Биполярная проводимость дается суммой вкладов от обеих мод проводимости с соответствующими значениями $\lambda(E)$ в нужных нам пределах энергий (рис. 16).

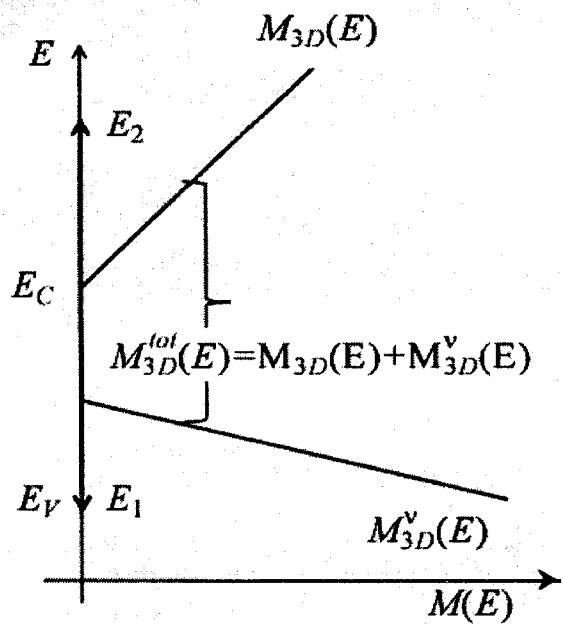


Рис. 16. К вычислению биполярной проводимости.

Биполярная проводимость σ^{tot} физически обеспечивается электронами и описывается общим уравнением (112): нет нужды менять знаки для валентной зоны или заменять f_0 на $1-f_0$.

Каким будет коэффициент Зеебека в режиме биполярной проводимости?

Вспомним, что при вычислении транспортных коэффициентов вклады от всех мод проводимости складываются. Для коэффициента Соре в случае биполярной проводимости имеем:

$$s_T^{tot} = -\frac{k}{q}\int_{-\infty}^{+\infty}\left(\frac{E-E_F}{kT}\right)\sigma'(E)dE = S\sigma + S_p\sigma_p, \tag{113}$$

тогда коэффициент Зеебека в режиме биполярной проводимости

$$S^{tot} = \frac{S\sigma + S_p\sigma_p}{\sigma + \sigma_p}; \tag{114}$$

коэффициенты Зеебека для электронов валентной зоны и зоны проводимости имеют противоположные знаки, так что в режиме биполярной проводимости суммарный коэффициент Зеебека близок к нулю и добротность ТЭ устройства резко падает.

2.2.8. Транспортные коэффициенты для 3D проводников в диффузионном режиме

Получим, наконец, транспортные коэффициенты (71) – (73) для массивных 3D диффузионных резисторов для транспортных уравнений (69) и (70). Начнем с удельного сопротивления (71). Удельную проводимость

$$\sigma = \frac{2q^2}{h}\int_{E_C}^{\infty} M_{3D}(E)\lambda(E)\left(-\frac{\partial f_0}{\partial E}\right)dE \tag{115}$$

перепишем следующим образом:

$$\sigma = \frac{2q^2}{h}\left[\frac{\int_{E_C}^{\infty} M_{3D}(E)\lambda(E)\left(-\frac{\partial f_0}{\partial E}\right)dE}{\int_{E_C}^{\infty} M_{3D}(E)\left(-\frac{\partial f_0}{\partial E}\right)dE}\right]\int_{E_C}^{\infty} M_{3D}(E)\left(-\frac{\partial f_0}{\partial E}\right)dE. \tag{116}$$

Знаменатель есть среднее число мод в фермиевском окне проводимости вблизи уровня Ферми

$$\int_{E_C}^{\infty} M_{3D}(E)\left(-\frac{\partial f_0}{\partial E}\right)dE \equiv \langle M_{3D}\rangle. \tag{117}$$

Выражение в квадратных скобках

$$\frac{\int_{E_C}^{\infty} M_{3D}(E)\lambda(E)\left(-\frac{\partial f_0}{\partial E}\right)dE}{\int_{E_C}^{\infty} M_{3D}(E)\left(-\frac{\partial f_0}{\partial E}\right)dE} = \frac{\langle M_{3D}\lambda(E)\rangle}{\langle M_{3D}\rangle} \equiv \langle\langle\lambda\rangle\rangle \tag{118}$$

есть средняя длина свободного пробега, усредненная по всем модам в окне проводимости. Таким образом, вычисление удельной проводимости 3D диффузионного резистора сводится к вычислению среднего числа мод в окне проводимости и усредненного по этим модам средней длины свободного пробега:

$$\sigma = \frac{2q^2}{h}\langle M_{3D}\rangle\langle\langle\lambda\rangle\rangle. \tag{119}$$

Для параболической дисперсии (106) из (117) имеем:

$$\begin{aligned}\langle M_{3D}\rangle &= g_v \frac{m^*}{2\pi\hbar^2} kT \int_{E_C}^{\infty} \left(\frac{E-E_C}{kT}\right)\left(-\frac{\partial f_0}{\partial E}\right) dE = \\ &= M_{3D}(kT)\int_{E_C}^{\infty} \eta\left(+\frac{\partial f_0}{\partial \eta_F}\right) d\eta = M_{3D}(kT)\Im_0(\eta_F),\end{aligned} \tag{120}$$

где $M_{3D}(kT)$ это выражение (106), вычисленное при $E - E_C = kT$, а новые переменные определены в (79).

Усредненное значение $\langle\lambda\rangle$ по (118) вычислим для степенного закона рассеяния (98), в котором показатель степени r принимает разные значения для разных механизмов рассеяния. Числитель в (118) преобразуется следующим образом:

$$\begin{aligned}\int_{E_C}^{\infty} M_{3D}(E)\lambda(E)\left(-\frac{\partial f_0}{\partial E}\right) dE &= \int_{E_C}^{\infty} g_v \frac{m^*}{2\pi\hbar^2}(E-E_C)\lambda_0\left(\frac{E-E_C}{kT}\right)^r\left(-\frac{\partial f_0}{\partial E}\right) dE \\ &= M_{3D}(kT)\lambda_0\Gamma(r+2)\Im_r(\eta_F).\end{aligned} \tag{121}$$

Поскольку знаменатель в (118) есть $\langle M_{3D}\rangle$ по (120), то

$$\langle\langle\lambda\rangle\rangle = \lambda_0\Gamma(r+2)\frac{\Im_r(\eta_F)}{\Im_0(\eta_F)}, \tag{122}$$

а проводимость (119) с учетом (120)

$$\sigma = \frac{2q^2}{h} M_{3D}(kT)\Im_0(\eta_F)\left[\lambda_0\Gamma(r+2)\frac{\Im_r(\eta_F)}{\Im_0(\eta_F)}\right]. \tag{123}$$

Итак, вычислили удельную проводимость и далее удельное сопротивление ρ по (71). Теперь вычислим коэффициент Зеебека по (72) и (62). С учетом (98) коэффициент Зеебека дается уравнением (91) с

$$\delta = \frac{\int\left(\frac{E-E_C}{kT}\right)\sigma'(E)dE}{\int\sigma'(E)dE} = (r+2)\frac{\Im_{r+1}(\eta_F)}{\Im_r(\eta_F)}. \tag{124}$$

Более общее определение этого параметра есть (§ 2.1.2)

$$\delta = \left(E_{av} - E_C\right)/kT\,. \tag{125}$$

Рассмотрим невырожденный предел ($\eta_F \ll 0$). Оба интеграла Ферми – Дирака в (124) становятся просто e^{η_F} и $\delta = r+2$. При $r=0$ параметр $\delta = 2$ и согласно (125) это означает, что ток течет при средней энергии на $2kT$ выше дна зоны проводимости E_C. При $r>0$ средняя длина свободного пробега (98) растет с ростом энергии, в результате чего ток течет при более высокой энергии и соответственно увеличивается коэффициент Зеебека. Для $r=2$, что характерно для рассеяния на ионизированных примесях, $\delta = 4$ и ток течет при средней энергии на $4kT$ выше E_C.

Остается вычислить удельную теплопроводность (73) или иначе число Лоренца (105):

$$L = \frac{\kappa}{T\sigma} = \left(\frac{k}{q}\right)^2 \left\{ \left(r+2\right)\left(r+3\right)\frac{\Im_{r+2}(\eta_F)}{\Im_r(\eta_F)} - \left[\left(r+2\right)\frac{\Im_{r+1}(\eta_F)}{\Im_r(\eta_F)}\right]^2 \right\}. \tag{126}$$

Этот частный результат полезно сравнить с общим выражением (105), которое не предполагает параболическую дисперсию и степенной закон рассеяния. В невырожденном пределе $r=0$ и выражение в фигурных скобках есть просто 2 как и ожидалось. В вырожденном пределе ($\eta_F \gg 0$) можно воспользоваться разложением интегралов Ферми – Дирака в ряд и показать, что выражение в фигурных скобках равно $\pi^2/3$.

2.2.9. Коэффициент Зеебека в вырожденном пределе: формула Мотта

Коэффициент Зеебека (91) с параметром δ по (124) справедлив для параболической дисперсии и степенного закона рассеяния. В невырожденном пределе интегралы Ферми – Дирака становятся просто экспонентой, в результате чего выражение для коэффициента Зеебека упрощается до

$$S = -\left(\frac{k}{q}\right)\left[\left(r+2\right) - \eta_F\right]. \tag{127}$$

Упростить коэффициент Зеебека можно и в сильно вырожденном пределе ($\eta_F \gg 0$) следующим образом. В этом пределе интеграл Ферми – Дирака порядка r аппроксимируется рядом [21]

$$\Im_r(\eta_F) \approx \frac{(\eta_F)^{r+1}}{\Gamma(r+2)} + \frac{(\eta_F)^{r-1}}{\Gamma(r)}\varsigma(2) + \ldots, \tag{128}$$

где $\varsigma(2) = \pi^2/6$ есть дзета-функция Римана. Воспользуемся этим разложением для коэффициента Зеебека (91) с параметром δ по (124)

$$S = -\frac{k}{q}\left[\left(r+2\right)\frac{\Im_{r+1}(\eta_F)}{\Im_r(\eta_F)} - \eta_F\right] \tag{129}$$

и найдем

$$S \approx -\frac{k}{q}\left[\frac{2(r+1)\eta_F \varsigma(2)}{(\eta_F)^2 + r(r+1)\varsigma(2)}\right], \tag{130}$$

что в сильно выраженном пределе ($\eta_F \gg 0$) дает

$$S \approx -\frac{k}{q}\left[\frac{2(r+1)\varsigma(2)}{\eta_F}\right]. \tag{131}$$

Мы видим из (127), что в невырожденном пределе

$$|S| \propto -\eta_F, \ (\eta_F \ll 0) \tag{132}$$

а из (131) в вырожденном пределе

$$|S| \propto 1/\eta_F. \ (\eta_F \ll 0) \tag{133}$$

Наконец, вспомним, что

$$\sigma'(E) \propto M_{3D}(E)\lambda(E) \propto (E - E_C)^{r+1}, \tag{134}$$

откуда

$$\frac{1}{\sigma'(E)}\frac{d\sigma'(E)}{dE}\Big|_{E=E_F} = \frac{r+1}{E_F - E_C}. \tag{135}$$

Подставляя в (131) $\varsigma(2)$ и $r+1$ из (135) получаем, наконец, известное соотношение Мотта для коэффициента Зеебека

$$S \approx -\frac{\pi^2 k^2 T}{3q}\left[\frac{\partial \ln \sigma'(E)}{\partial E}\right]_{E=E_F}, \tag{136}$$

которое широко используется для вырожденных полупроводников. Для получения этого соотношения мы воспользовались уравнением (129), для которого предполагается параболическая зонная структура и степенной закон рассеяния. Этот же результат можно получить и из более общих соображений, если воспользоваться известным разложением Зоммерфельда [9, 22].

2.2.10. Анализ экспериментальных данных для Ge: коэффициент Зеебека и термоэлектрические коэффициенты

На рис. 17 показана зависимость коэффициентов Зеебека для Ge *n*- и *p*-типа при комнатной температуре в зависимости от положения уровня Ферми E_F относительно дна зоны проводимости E_C.

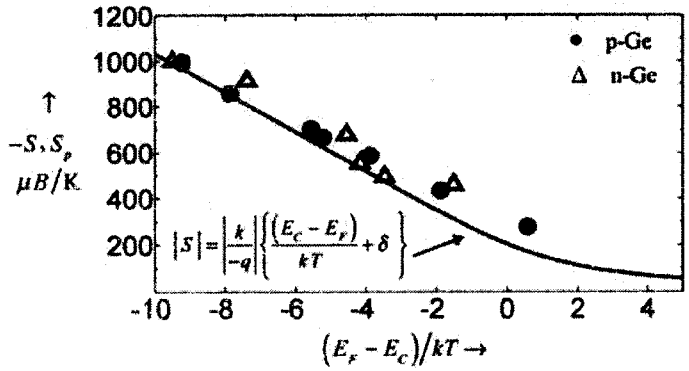


Рис. 17. Коэффициент Зеебека для Ge *n*- и *p*-типа при $T = 300K$: экспериментальные данные по [23], расчетная кривая согласно [20].

Покажем, что экспериментальные результаты согласуются с уравнением (91) и указывают на то, что поправка δ (93) мала и постоянна у невырожденных проводников. Покажем также, что поправка δ растет по мере удаления уровня Ферми от дна зоны проводимости. Рассмотрим также ситуацию с *p*-проводниками.

Согласно (91) коэффициент Зеебека $|S|$ должен падать линейно с ростом E_F, если только поправка δ (93) постоянна. Этот вывод подтверждается экспериментальными данными (рис. 17). Так

$$|S| = 86 \times (-\eta_F + \delta), \quad [\mu B/K] \tag{137}$$

где η_F определяется по (79). При $\eta_F = -10$ величина $|S| \approx 1000\ \mu B/K$, откуда следует, что $\delta \approx 2$ в соответствии с приведенными выше простыми рассуждениями для невырожденных проводников.

Если бы поправка δ была всегда постоянна, то коэффициент Зеебека менял бы знак при $E_F > E_c + 2kT$. На самом деле этого не происходит (рис. 17), что указывает на то, что поправка δ должна увеличиваться по мере того как уровень Ферми поднимается все выше по зоне проводимости. При $\eta_F = 4$ коэффициент Зеебека $|S| \approx 50\ \mu B/K$ (рис. 17), что соответствует $\delta \approx 4.6$, а это в два раза превышает значение этой поправки для невырожденных условий. Как было показано выше, поправка

$$\delta = (E_{av} - E_c)/kT, \tag{138}$$

где E_{av} есть среднее значение энергии, при которой движутся носители тока, показывает нам на каком удалении от дна зоны проводимости движутся электроны. У сильно вырожденных проводников $E_{av} \rightarrow E_F$, так что $\delta \rightarrow \eta_F$ и $S \rightarrow 0$ при $E_F \gg E_c$.

Для n-проводников коэффициент Зеебека (по модулю) тем больше, чем ниже уровень Ферми находится по отношению к дну зоны проводимости. Знак коэффициента Зеебека отрицательный. В случае p-проводников коэффициент Зеебека тем больше, чем выше уровень Ферми находится по отношению к потолку валентной зоны и знак его положительный. Токовые уравнения (4) и (5) для n-проводников при переходе к p-проводникам не изменяются, а коэффициент Зеебека запишется следующим образом:

$$\frac{d\left(E_F^{(p)}/q\right)}{dx} = \rho_p J_{px} + S_p \frac{dT}{dx},$$
$$S_p(T) = \left(\frac{k}{+q}\right)\left(\frac{E_F^{(p)} - E_v}{kT} + \delta_p\right), \tag{139}$$

где $\delta_p = (E_v - E_{av})/kT$ является положительным числом, которое показывает нам насколько ниже потолка валентной зоны движутся «дырки». Как мы уже видели выше, в условиях биполярной проводимости и «дырки» и электроны дают вклад в коэффициент Зеебека. Учет этого обстоятельства становится важным при высоких температурах и для узкощелевых полупроводников.

Обратимся к термоэлектрическим коэффициентам для Ge. Оценим их для слаболегированного n-Ge при комнатной температуре. Для него известно, что при $T = 300\,K$ подвижность $\mu = 3200\,см^2/В\cdot с$, а равновесная плотность носителей тока $n_0 \approx N_D = 10^{-15} см^{-3}$ [10]. Вспомним также, что для невырожденных полупроводников справедливо простое соотношение между плотностью носителей тока и положением уровня Ферми, а именно [24]:

$$n_0 = N_c e^{(E_F - E_c)/kT}, \tag{140}$$

где «эффективная плотность состояний» для Ge при комнатной температуре $N_c = 1.09 \times 10^{19} см^{-3}$ [24].

Основные уравнения термоэлектричества для массивного образца даются (69) и (70) с четырьмя транспортными коэффициентами: удельное сопротивление ρ (71), коэффициент Зеебека S (72), электронная теплопроводность κ (73) и коэффициент Пельтье $\pi = TS$ (соотношение Кельвина).

Имеющиеся в нашем распоряжении данные позволяют сразу оценить удельную проводимость $\sigma = n_0 q \mu$, откуда удельное сопротивление $\rho \approx 2\,\Omega\cdot см$.

Коэффициент Зеебека дается соотношением (56) и он зависит от положения уровня Ферми. Из (140) имеем

$$\frac{E_c - E_F}{kT} = \ln\frac{N_c}{n_0} \approx 9.3. \tag{141}$$

В предположении, что $\delta = 2$, для коэффициента Зеебека получаем $S \approx -970\,мкВ/К$.

Коэффициент Пельтье находим из соотношения Кельвина: $\pi = -0.3\,Вт/А$.

Электронную теплопроводность найдем через число Лоренца:

$$\kappa = T\sigma L. \qquad (142)$$

Поскольку в нашем случае речь идет о слаболегированном 3D полупроводнике с почти параболической зонной структурой, то мы вправе предположить, что число Лоренца как для невырожденного проводника $L \approx 2(k/q)^2$, откуда $\kappa \approx 2.2 \times 10^{-4}\,Вт/м \cdot К$.

Рассмотрен довольно простой пример. Полезно продумать что изменится в этой сводке термоэлектрических коэффициентов Ge, если температура понизится до 77 $К$ или температура останется комнатной, а легирование возрастет до $10^{20}\,см^3$. Нужно помнить, что решетка также проводит тепло. Решеточная теплопроводность Ge при комнатной температуре достигает $58\,Вт/м \cdot К$, что на пять порядков больше оцененной выше электронной теплопроводности. В сильнолегированных полупроводниках с низкой решеточной теплопроводностью вклад электронов в теплопроводность может быть существенным.

Подведем итоги. В нашу задачу входило получить строгие результаты, сформулированные вначале без доказательства, для 1D, 2D и 3D проводников с произвольной дисперсией и любого масштаба в режимах от баллистического до диффузионного. Показано с подробными выкладками как получаются термоэлектрические коэффициенты для 1D баллистического транспорта и для 3D диффузионного. Основным объектом обсуждения были n-проводники, вместе с тем мы показали каким образом распространить результаты на p-проводники. Рассмотрели биполярную проводимость, закон Видемана – Франца и числа Лоренца, соотношение Мотта. Процессы рассеяния учитывали феноменологически. В коэффициенте прохождения $T(E) = \lambda(E)/[\lambda(E) + L]$ среднюю длину свободного пробега описывали степенным законом.

Далее в справочных целях собраны воедино термоэлектрические коэффициенты для 1D, 2D и 3D полупроводников с параболической дисперсией [20, 25]. А затем перейдем к рассмотрению переноса тепла фотонами.

2.2.11. Термоэлектрические коэффициенты для 1D, 2D и 3D проводников с параболической дисперсией

В приведенных ниже формулах предполагается параболическая зонная структура

$$E(k) = \frac{\hbar^2 k^2}{2m^*} \qquad (143)$$

и степенной закон рассеяния

$$\lambda(E) = \lambda_0 \left(\frac{E}{kT} \right)^r. \qquad (144)$$

Положение уровня Ферми E_F относительно дна зоны проводимости E_C определяется параметром

$$\eta_F = \frac{E_F - E_C}{kT}. \qquad (145)$$

2.2.11.1. Термоэлектрические коэффициенты для 1D баллистических резисторов

$$G = \frac{2q^2}{h}\Im_{-1}(\eta_F); \quad S = -\frac{k}{q}\left[\frac{\Im_0(\eta_F)}{\Im_{-1}(\eta_F)} - \eta_F\right];$$

$$S_T = -\frac{k}{q}\frac{2q^2}{h}\left[\Im_0(\eta_F) - \eta_F\Im_{-1}(\eta_F)\right];$$

$$K_0 = T\left(\frac{k}{q}\right)^2\frac{2q^2}{h}\left[2\Im_1(\eta_F) - 2\eta_F\Im_0(\eta_F) + \eta_F^2\Im_{-1}(\eta_F)\right];$$

$$K = T\left(\frac{k}{q}\right)^2\frac{2q^2}{h}\left[2\Im_1(\eta_F) - \frac{\Im_0^2(\eta_F)}{\Im_{-1}(\eta_F)}\right];$$

2.2.11.2. Термоэлектрические коэффициенты для 1D диффузионных резисторов

$$G = \frac{2q^2}{h}\left(\frac{\lambda_0}{L}\right)\Gamma(r+1)\Im_{r-1}(\eta_F); \; S = -\frac{k}{q}\left[\frac{(r+1)\Im_r(\eta_F)}{\Im_{r-1}(\eta_F)} - \eta_F\right];$$

$$S_T = -\frac{k}{q}\frac{2q^2}{h}\left(\frac{\lambda_0}{L}\right)\Gamma(r+1)\left[(r+1)\Im_r(\eta_F) - \eta_F\Im_{r-1}(\eta_F)\right];$$

$$K_0 = T\left(\frac{k}{q}\right)^2\frac{2q^2}{h}\left(\frac{\lambda_0}{L}\right)\left[\begin{matrix}\Gamma(r+3)\Im_{r+1}(\eta_F) - 2\eta_F\Gamma(r+2)\Im_r(\eta_F) + \\ +\eta_F^2\Gamma(r+1)\Im_{r-1}(\eta_F)\end{matrix}\right];$$

$$K = T\left(\frac{k}{q}\right)^2\frac{2q^2}{h}\left(\frac{\lambda_0}{L}\right)\Gamma(r+2)\left[(r+2)\Im_{r+1}(\eta_F) - \frac{(r+1)\Im_r^2(\eta_F)}{\Im_{r-1}(\eta_F)}\right];$$

Проводимость $G = \sigma_{1D}/L$ приведена в *Сименсах*: $[\sigma_{1D}] = 1 См \cdot м$. Аналогично для остальных коэффициентов: $s_T = S_T L; \;\; \kappa_0 = K_0 L; \;\; \kappa = KL.$

2.2.11.3. Термоэлектрические коэффициенты для 2D баллистических резисторов

$$G = W\frac{2q^2}{h}\frac{\sqrt{2\pi m^* kT}}{h}\Im_{-1/2}(\eta_F);\; S = -\frac{k}{q}\left[\frac{3\Im_{1/2}(\eta_F)}{2\Im_{-1/2}(\eta_F)} - \eta_F\right];$$

$$S_T = -W\frac{k}{q}\frac{2q^2}{h}\frac{\sqrt{2\pi m^* kT}}{h}\left[\frac{3}{2}\Im_{1/2}(\eta_F) - \eta_F\Im_{-1/2}(\eta_F)\right];$$

$$K_0 = WT\left(\frac{k}{q}\right)^2\frac{2q^2}{h}\frac{\sqrt{2\pi m^* kT}}{h}\left[\frac{15}{4}\Im_{3/2}(\eta_F) - 3\eta_F\Im_{1/2}(\eta_F) + \eta_F^2\Im_{-1/2}(\eta_F)\right];$$

$$K = WT\left(\frac{k}{q}\right)^2\frac{2q^2}{h}\frac{\sqrt{2\pi m^* kT}}{h}\left[\frac{15}{4}\Im_{3/2}(\eta_F) - \frac{9\Im_{1/2}^2(\eta_F)}{4\Im_{-1/2}(\eta_F)}\right];$$

2.2.11.4. Термоэлектрические коэффициенты для 2D диффузионных резисторов

$$G = W\frac{2q^2}{h}\left(\frac{\lambda_0}{L}\right)\frac{\sqrt{2m^* kT}}{\pi h}\Gamma(r+\frac{3}{2})\Im_{r-1/2}(\eta_F);$$

$$S = -\frac{k}{q}\left[\frac{(r+3/2)\Im_{r+1/2}(\eta_F)}{\Im_{r-1/2}(\eta_F)} - \eta_F\right];$$

$$S_T = -W\frac{k}{q}\frac{2q^2}{h}\left(\frac{\lambda_0}{L}\right)\frac{\sqrt{2m^* kT}}{\pi h}\left[\Gamma(r+\frac{5}{2})\Im_{r+1/2}(\eta_F) - \eta_F\Gamma(r+\frac{3}{2})\Im_{r-1/2}(\eta_F)\right];$$

$$K_0 = WT\left(\frac{k}{q}\right)^2\frac{2q^2}{h}\left(\frac{\lambda_0}{L}\right)\frac{\sqrt{2m^* kT}}{\pi h}\times$$

$$\times\left[\Gamma(r+\frac{7}{2})\Im_{r+3/2}(\eta_F) - 2\eta_F\Gamma(r+\frac{5}{2})\Im_{r+1/2}(\eta_F) + \eta_F^2\Gamma(r+\frac{3}{2})\Im_{r-1/2}(\eta_F)\right];$$

$$K = WT\left(\frac{k}{q}\right)^2\frac{2q^2}{h}\left(\frac{\lambda_0}{L}\right)\frac{\sqrt{2m^* kT}}{\pi h}\Gamma(r+\frac{5}{2})\left[(r+\frac{5}{2})\Im_{r+3/2}(\eta_F) - \frac{(r+\frac{3}{2})\Im_{r+1/2}^2(\eta_F)}{\Im_{r-1/2}(\eta_F)}\right];$$

Проводимость $G = \sigma_{2D}W/L$ приведена в *Сименсах*: $[\sigma_{2D}] = 1См$. Аналогично для остальных коэффициентов: $s_T = S_T L/W$; $\kappa_0 = K_0 L/W$; $\kappa = KL/W$.

2.2.11.5. Термоэлектрические коэффициенты для 3D баллистических резисторов

$$G = A\frac{2q^2}{h}\frac{m^* kT}{2\pi\hbar^2}\Im_0(\eta_F);\quad S = -\frac{k}{q}\left[\frac{2\Im_1(\eta_F)}{\Im_0(\eta_F)} - \eta_F\right];$$

$$S_T = -A\frac{k}{q}\frac{2q^2}{h}\frac{m^* kT}{2\pi\hbar^2}\left[2\Im_1(\eta_F) - \eta_F\Im_0(\eta_F)\right];$$

$$K_0 = AT\left(\frac{k}{q}\right)^2\frac{2q^2}{h}\frac{m^* kT}{2\pi\hbar^2}\left[6\Im_2(\eta_F) - 4\eta_F\Im_1(\eta_F) + \eta_F^2\Im_0(\eta_F)\right];$$

$$K = AT\left(\frac{k}{q}\right)^2\frac{2q^2}{h}\frac{m^* kT}{2\pi\hbar^2}\left[6\Im_2(\eta_F) - \frac{4\Im_1^2(\eta_F)}{\Im_0(\eta_F)}\right];$$

2.2.11.6. Термоэлектрические коэффициенты для 3D диффузионных резисторов

$$G = A\frac{2q^2}{h}\left(\frac{\lambda_0}{L}\right)\frac{m^* kT}{2\pi\hbar^2}\Gamma(r+2)\Im_r(\eta_F);\quad S = -\frac{k}{q}\left[\frac{(r+2)\Im_{r+1}(\eta_F)}{\Im_r(\eta_F)} - \eta_F\right];$$

$$S_T = -A\frac{k}{q}\frac{2q^2}{h}\left(\frac{\lambda_0}{L}\right)\frac{m^* kT}{2\pi\hbar^2}\left[\Gamma(r+3)\Im_{r+1}(\eta_F) - \eta_F\Gamma(r+2)\Im_r(\eta_F)\right];$$

$$K_0 = AT\left(\frac{k}{q}\right)^2\frac{2q^2}{h}\left(\frac{\lambda_0}{L}\right)\frac{m^* kT}{2\pi\hbar^2}\left[\begin{array}{l}\Gamma(r+4)\Im_{r+2}(\eta_F) - 2\eta_F\Gamma(r+3)\Im_{r+1}(\eta_F) + \\ +\eta_F^2\Gamma(r+2)\Im_r(\eta_F)\end{array}\right];$$

$$K = AT\left(\frac{k}{q}\right)^2\frac{2q^2}{h}\left(\frac{\lambda_0}{L}\right)\frac{m^* kT}{2\pi\hbar^2}\Gamma(r+3)\left[(r+3)\Im_{r+2}(\eta_F) - \frac{(r+2)\Im_{r+1}^2(\eta_F)}{\Im_r(\eta_F)}\right].$$

Проводимость $G = \sigma_{3D}A/L$ приведена в *Сименсах*: $[\sigma_{3D}] = 1См/м$. Аналогично для остальных коэффициентов: $s_T = S_T L/A;\quad \kappa_0 = K_0 L/A;\quad \kappa = KL/A.$

2.3. Перенос тепла фононами

При транспорте электронов переносится как заряд, так и тепло. В металлах почти все тепло переносится электронами, а в полупроводниках только меньшая часть тепла переносится электронами, и почти все тепло переносится фононами.

Поток тепла, вызванный транспортом фононов вдоль проводника (ось x)

$$J_{Qx}^{ph} = -\kappa_L \frac{dT}{dx}, \; [Вт/м^2] \qquad (146)$$

где коэффициентом пропорциональности потока тепла градиенту температуры является удельная решеточная теплопроводность κ_L. У алмаза, одного из лучших проводников тепла, $\kappa_L \approx 2000$ *Вт/м·К*, у такого плохого проводника тепла как стекло $\kappa_L \approx 1$ *Вт/м·К*, а у воздуха $\kappa_L \approx 0.025$ *Вт/м·К*. Если электрическая проводимость твердых материалов меняется на 20 и больше порядков, то их решеточная теплопроводность изменяется всего на 3 – 4 порядка. Тем не менее, есть много общего в описании переноса тепла электронами и фононами в рамках обобщенной транспортной модели. Далее упомянем лишь необходимые нам сведения о фононах.

2.3.1. Элементарная физика фононов

Физика фононов подробно изложена в [9, 26 – 28]. Характерный ход дисперсии $E(k)$ для электронов качественно показан на рис. 18, а дисперсии $\hbar\omega(q)$ для фононов – на рис. 19.

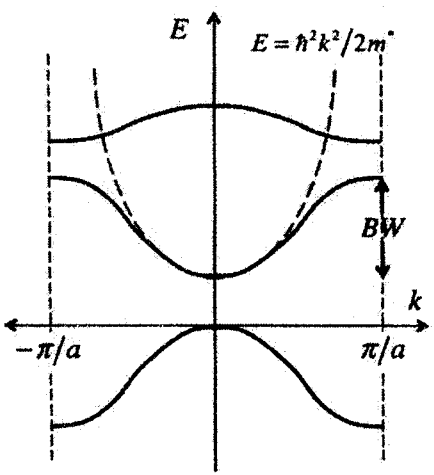


Рис. 18. Упрощенная зонная структура. Ширина зоны проводимости помечена BW/Band Width. Часто используемая параболическая дисперсия показана пунктиром.

Кристаллическая решетка периодична в реальном пространстве, а дисперсия – в k-пространстве. Аналогично для фононов: только в q-пространстве. В обоих случаях решения реализуются в зоне Бриллюэна и ширины обоих зон конечны.

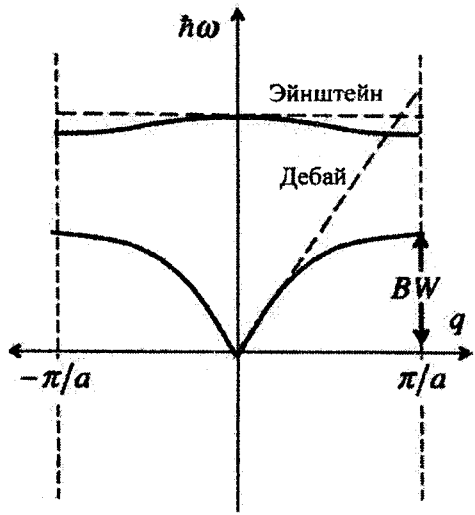


Рис. 19. Упрощенный ход дисперсии для фононов. Пунктиром показаны простейшие модели Эйнштейна и Дебая.

Подробнее ход дисперсии для фононов показан на рис. 20.

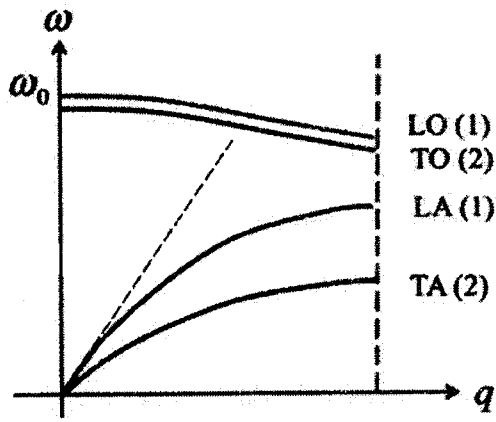


Рис. 20. Дисперсионные кривые для фононов.

Для кристаллических 3D проводников при распространении колебаний по решетке возможны три состояния поляризации: одно продольное, в направлении распространения колебаний (longitudinal/L) и два поперечных (transverse/T). В области низких энергий фононные моды называют акустическими/A. Вблизи нулевых волновых чисел ($q \to 0$) у акустических мод дисперсия линейная. Для продольной акустической моды LA

$$\omega(q) = v_s q, \qquad v_s = \sqrt{k_{elas} / m_{eff}}, \tag{147}$$

где скорость звука v_s определяется упругой постоянной k_{elas} и массой m_{eff}. Характерные скорости звука порядка 5×10^3 *м/с*, что примерно в 20 раз меньше скорости электронов. Из уравнений (147) следует, что у веществ с тяжелыми атомами скорость звука меньше и соответственно меньше ширина фононной зоны дисперсии по сравнению с веществами из легких атомов.

На рис. 20 показаны также три оптические моды. Дисперсия этих мод незначительна: частота оптических мод L и T сравнительно мало зависит от q. Различие между акустическими и оптическими модами состоит в том, что в области малых $q \approx 0$ соседние атомы смещаются в одном и том же направлении для акустических мод и в противоположных направлениях для оптических мод. В полярных веществах свет взаимодействует с оптическими модами; отсюда и их название. Скорость фононов определяется наклоном зависимости частоты от смещения $\omega(q)$: акустические моды характеризуются относительной бóльшей скоростью фононов по сравнению с оптическими модами. Поэтому можно ожидать, что именно акустические моды переносят большую часть тепла.

В качестве реалистического примера приводим фононный спектр графена (рис. 21), построенный на основе экспериментальных данных, хорошо согласующихся с результатами расчетов методами теории функционала плотности [29 – 31].

В элементарной ячейке графена два атома углерода, что обусловливает наличие в спектре дисперсии графена шести фононных ветвей – трех оптических (LO, TO, ZO) и трех акустических (LA, TA, ZA). Последние связаны со смещением всей ячейки из положения равновесия, а оптическим модам отвечают сдвиги атомов в самой элементарной ячейке с сохранением положения центра масс. Фононные моды LA и TA соответствуют продольным и поперечным колебаниям атомов углерода в плоскости графена, а мода ZA (flexural mode) – колебаниям атомов углерода в направлении, перпендикулярном направлению колебаний мод LA и TA, с выходом атомов из плоскости графена. Моды LA и TA характеризуются линейной дисперсией; соответствующие этим модам скорости звука составляют 21300 и 13600 *м/с*, соответственно. Все теплофизические свойства графена (теплопроводность, теплоемкость и др.) определяются фононами.

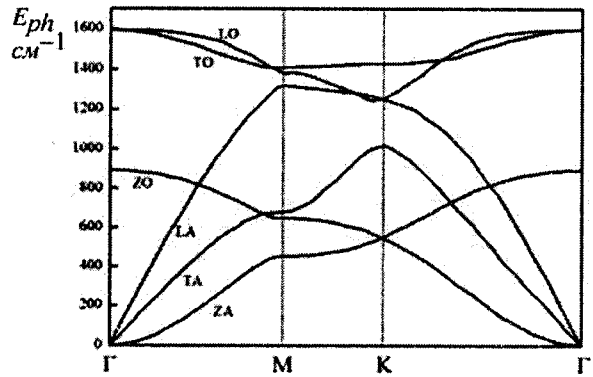


Рис. 21. Фононный спектр графена, опирающийся на совокупные экспериментальные и расчетные данные.

2.3.2. Обобщенная модель переноса тепла фононами

Электрический ток в обобщенной транспортной модели (глава 1)

$$I = \frac{2q}{h}\int T_{el}(E)M_{el}(E)\left(f_1 - f_2\right)dE \ [A] \tag{148}$$

пропорционален произведению коэффициента прохождения $T_{el}(E)$ или иначе вероятности прохождения проводника по каналу проводимости энергии E, числу мод проводимости $M_{el}(E)$ при этой энергии и разности фермиевских функций на контактах проводника.

Для описания транспорта фононов нам необходимо уравнение, аналогичное (148). Как и в случае электронов, контакты на концах проводника достаточно массивны настолько, чтобы поддерживать тепловое равновесие фононов при разных температурах контактов. Если в случае электронов состояния контактов заселяются в соответствии со статистикой Ферми – Дирака, то в случае фононов – в соответствии со статистикой Бозе – Эйнштейна

$$n(\hbar\omega) = \frac{1}{e^{\hbar\omega/kT} - 1}. \tag{149}$$

Приведем важное свойство бозевской функции, которое понадобятся нам вскоре:

$$n(x) \equiv \frac{1}{e^x - 1}, \quad x \equiv \frac{\hbar\omega}{kT}; \quad \frac{\partial n}{\partial T} = \frac{dn}{dx}\frac{\partial x}{\partial T} = -\frac{\hbar\omega}{kT^2}\frac{dn}{dx}.$$

Пусть температура на левом и правом контактах будет соответственно T_1 и T_2. Как и в случае электронов, предполагается, что контакты идеальные, другими словами, отражение фононов на контактах не происходит, и коэффициент прохождения действительно соответствует прохождению фононов через весь канал проводимости.

Теперь легко переписать выражение (148) применительно к фононам. Для потока фононов энергию электронов E заменим на энергию фононов $\hbar\omega$. В случае электронов переносятся заряды q, а в случае фононов – кванты энергии $\hbar\omega$: заменим q на $\hbar\omega$ и внесем $\hbar\omega$ под знак интеграла. Двойка в уравнении (148) для тока электронов соответствует вырождению по спину. Исключим ее из уравнения для фононов, а число состояний поляризации фононов включим в число фононных мод $M_{ph}(\hbar\omega)$. Окончательно для тепла, переносимого фононами, имеем:

$$Q=\frac{1}{h}\int(\hbar\omega)T_{ph}(\hbar\omega)M_{ph}(\hbar\omega)\left(n_1-n_2\right)d(\hbar\omega).\ [Bm] \qquad (150)$$

Как и в случае электронов, нас интересует режим линейного отклика, или иначе квази-равновесного транспорта. В этом режиме $T_1\approx T_2$ и $n_0\approx n_1\approx n_2\approx n_0$, что позволяет разложить n_2 относительно n_1 в ряд Тейлора и ограничиться первым порядком:

$$n_2\approx n_1+\frac{\partial n_0}{\partial T}\Delta T, \qquad (151)$$

так что нужная нам разность

$$n_1-n_2\approx-\frac{\partial n_0}{\partial T}\Delta T. \qquad (152)$$

Производную $\partial n_0/\partial T$ находим дифференцированием бозевского распределения (149):

$$\frac{\partial n_0}{\partial T}=\frac{\hbar\omega}{T}\left(-\frac{\partial n_0}{\partial(\hbar\omega)}\right), \qquad (153)$$

где

$$\frac{\partial n_0}{\partial(\hbar\omega)}=\left(-\frac{1}{kT}\right)\frac{e^{\hbar\omega/kT}}{(e^{\hbar\omega/kT}-1)^2}. \qquad (154)$$

Теперь уравнение (150) для малой разности температур на контактах можно переписать следующим образом:

$$Q=-K_L\Delta T, \qquad (155)$$

где решеточная теплопроводность

$$K_L=\frac{k^2T}{h}\int T_{ph}(\hbar\omega)M_{ph}(\hbar\omega)\left[\left(\frac{\hbar\omega}{kT}\right)^2\left(-\frac{\partial n_0}{\partial(\hbar\omega)}\right)\right]d(\hbar\omega).\ [Bm/K] \qquad (156)$$

Уравнение (155) есть закон теплопроводности Фурье, согласно которому поток тепла антинаправлен градиенту температуры. Полезно обратить внимание на следующую схожесть теплопроводности (156) и электронной проводимости

$$G = \frac{2q^2}{h}\int T_{el}(E)M_{el}(E)\left(-\frac{\partial f_0}{\partial E}\right)dE. \qquad (157)$$

Производная

$$W_{el}(E) = \left(-\frac{\partial f_0}{\partial E}\right) \qquad (158)$$

известна как фермиевское окно проводимости, выделяющее те моды проводника, которые только и могут внести вклад в электронную проводимость. Окно проводимости нормировано на единицу:

$$\int_{-\infty}^{+\infty}\left(-\frac{\partial f_0}{\partial E}\right)dE = 1. \qquad (159)$$

Аналогичную роль окна проводимости для фононов играет следующее выражение:

$$W_{ph}(\hbar\omega) = \frac{3}{\pi^2}\left(\frac{\hbar\omega}{kT}\right)\left(\frac{\partial n_0}{\partial(\hbar\omega)}\right), \qquad (160)$$

в котором численный коэффициент необходим для нормировки на единицу выражения в квадратных скобках уравнения (156):

$$\int_0^{+\infty}\left(\frac{\hbar\omega}{kT}\right)^2\left(-\frac{\partial n_0}{\partial(\hbar\omega)}\right)d(\hbar\omega) = \frac{\pi^2}{3}, \qquad (161)$$

так что окончательно решеточная теплопроводность

$$K_L = \frac{\pi^2k^2T}{3h}\int T_{ph}(\hbar\omega)M_{ph}(\hbar\omega)W_{ph}(\hbar\omega)d(\hbar\omega). \qquad (162)$$

В этом выражении численный коэффициент

$$g_0 \equiv \pi^2k^2T/3h \approx (9.456\times10^{-13}Bm/K^2)T, \qquad (163)$$

который мы обсудим позже, есть квант теплопроводности, впервые подтвержденный экспериментально в 2000 году [32].

Сравнивая выражения для решеточной теплопроводности (162) и электронной проводимости (157), опять убеждается в их схожести: и та и другая проводимости пропорциональны соответствующему кванту проводимости, умноженному на интеграл от произведения трех величин – коэффициента прохождения, числа мод и окна проводимости.

Профили окон проводимости для электронов и фононов показаны на рисунках 22 и 23.

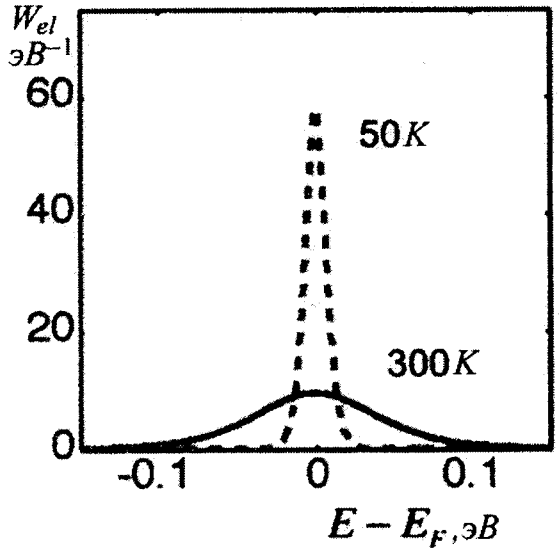


Рис. 22. Окно проводимости для электронов (158) при двух разных температурах. Энергия E может быть как больше, так и меньше фермиевской.

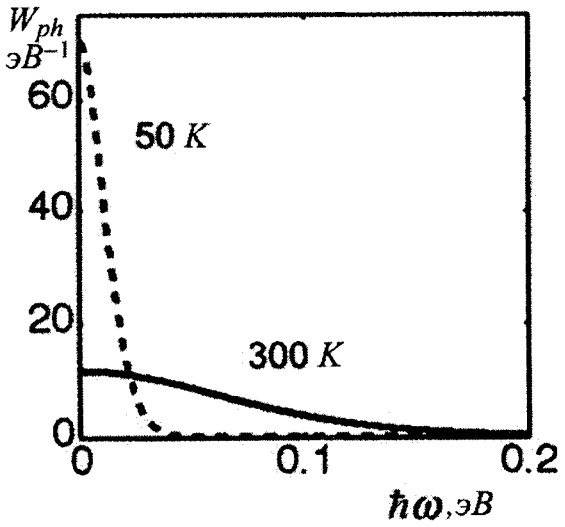


Рис. 23. Окно проводимости для фононов (160) при двух разных температурах. Энергия фононов может быть только положительной.

Окна проводимости для электронов и фононов не только схожи, но и имеют приблизительно одинаковую ширину в $\approx \pm 2\,kT$.

2.3.2.1. Теплопроводность массивных проводников

Теплопроводность диффузионных проводников является их ключевым свойством, определяющим работоспособность разнообразных электронных устройств. По аналогии с транспортом электронов (глава 1) коэффициент прохождения для фононов

$$T_{ph}(\hbar\omega) = \frac{\lambda_{ph}(\hbar\omega)}{\lambda_{ph}(\hbar\omega)+L}\Big|_{L\gg\lambda_{ph}} \to \frac{\lambda_{ph}(\hbar\omega)}{L}, \tag{164}$$

где L – длина 3D проводника, а $\lambda_{ph}(\hbar\omega)$ есть длина свободного пробега фонона частоты ω. Кроме того, число фононных мод, очевидно, пропорционально площади поперечного сечения 3D проводника:

$$M_{ph}(\hbar\omega) \propto A, \tag{165}$$

Умножая и деля уравнение (155) на A/L, получаем

$$Q = -\left(K_L \frac{L}{A}\right) A \frac{\Delta T}{L}, \tag{166}$$

откуда следует постулированное в самом начале уравнение (146) для потока тепла:

$$\frac{Q}{A} \equiv J_{Qx}^{ph} = -\kappa_L \frac{dT}{dx} \tag{167}$$

с удельной решеточной теплопроводностью

$$\kappa_L = K_L \frac{L}{A} \tag{168}$$

или после подстановки (164) в (162) в развернутом виде окончательно получаем

$$\kappa_L = \frac{\pi^2 k^2 T}{3h} \int \frac{M_{ph}(\hbar\omega)}{A} \lambda_{ph}(\hbar\omega) W_{ph}(\hbar\omega)\, d(\hbar\omega). \tag{169}$$

Поскольку именно окно проводимости фононов $W_{ph}(\hbar\omega)$ определяет какие именно фононные моды могут принимать участие в проводимости тепла, естественно ввести среднее число активных фононных мод на единицу площади поперечного сечения проводника

$$\left\langle M_{ph}/A \right\rangle \equiv \int \frac{M_{ph}(\hbar\omega)}{A} W_{ph}(\hbar\omega)\, d(\hbar\omega), \tag{170}$$

так что в компактном виде

$$\kappa_L = \frac{\pi^2 k^2 T}{3h}\left\langle M_{ph}/A\right\rangle\left\langle\left\langle \lambda_{ph}\right\rangle\right\rangle, \tag{171}$$

где усредненное значение длины свободного пробега фононов

$$\left\langle\left\langle \lambda_{ph}\right\rangle\right\rangle = \frac{\int \frac{M_{ph}(\hbar\omega)}{A}\lambda_{ph}(\hbar\omega)W_{ph}(\hbar\omega)d(\hbar\omega)}{\int \frac{M_{ph}(\hbar\omega)}{A}W_{ph}(\hbar\omega)d(\hbar\omega)}. \tag{172}$$

Итак, паре фононных транспортных уравнений (167) и (171) соответствует схожая по форме пара электронных транспортных уравнений (глава 1):

$$J_x = \frac{\sigma}{q}\frac{d\left(E_F\right)}{dx}, \tag{173}$$

$$\sigma = \frac{2q^2}{h}\left\langle M_{el}/A\right\rangle\left\langle\left\langle \lambda_{el}\right\rangle\right\rangle. \tag{174}$$

Градиент электрохимического потенциала вызывает поток электронов, а градиент температуры – поток тепла. Термоэлектрические коэффициенты (171) и (174) имеют одинаковую структуру. Это всегда произведение соответствующего кванта проводимости на число мод проводимости, участвующих в транспорте, и на усредненное значение средней длины свободного пробега соответствующего носителя (тока или тепла). К оценке этих трех величин для фононов мы вернемся позже.

2.3.2.2. Удельная теплоемкость

Хорошо известна пропорциональность между удельной теплопроводностью и удельной теплоемкостью при постоянном объеме C_V [9, 26 – 28]. Установим эту связь и покажем, что коэффициент пропорциональности между ними определяется усредненной должным образом средней длиной свободного пробега фононов $\left\langle\left\langle \Lambda_{ph}\right\rangle\right\rangle$ и средней скоростью фононов $\left\langle v_{ph}\right\rangle$, а именно:

$$\kappa_L = \frac{1}{3}\left\langle\left\langle \Lambda_{ph}\right\rangle\right\rangle\left\langle v_{ph}\right\rangle C_V. \tag{175}$$

Энергия фононов на единицу объема, очевидно, дается выражением:

$$E_{ph} = \int_0^\infty (\hbar\omega)D_{ph}(\hbar\omega)n_0(\hbar\omega)d(\hbar\omega), \tag{176}$$

где $D_{ph}(\hbar\omega)$ есть плотность фононных состояний. По определению,

$$C_V \equiv \frac{\partial E_{ph}}{\partial T} = \frac{\partial}{\partial T}\int_0^\infty (\hbar\omega) D_{ph}(\hbar\omega) n_0(\hbar\omega) d(\hbar\omega) = \\ = \int_0^\infty (\hbar\omega) D_{ph}(\hbar\omega)\left(\frac{\partial n_0(\hbar\omega)}{\partial T}\right) d(\hbar\omega) = \frac{\pi^2 k^2 T}{3}\int_0^\infty D_{ph}(\hbar\omega) W_{ph}(\hbar\omega) d(\hbar\omega), \quad (177)$$

где использовались уравнения (153) и (160).

Умножение и деление коэффициента теплопроводности (169) на теплоемкость (177) дает искомую пропорциональность:

$$\kappa_L = \left[\frac{\frac{1}{h}\int_0^\infty \frac{M_{ph}(\hbar\omega)}{A}\lambda_{ph}(\hbar\omega) W_{ph}(\hbar\omega) d(\hbar\omega)}{\int_0^\infty D_{ph}(\hbar\omega) W_{ph}(\hbar\omega) d(\hbar\omega)}\right] C_V. \quad (178)$$

Для получения окончательного результата в виде уравнения (175) и корректной интерпретации коэффициента пропорциональности между κ_L и C_V нужно вернуться к связи между средней длиной свободного пробега и коэффициентом прохождения (164), имеющей одинаковый вид как для электронов, так и для фононов.

2.3.2.3. Средняя длина свободного пробега и коэффициент прохождения

Связь этих двух важнейших характеристик диффузионного транспорта проще всего проиллюстрировать на примере электронного транспорта по 1D проводнику в диффузионном режиме (рис. 24).

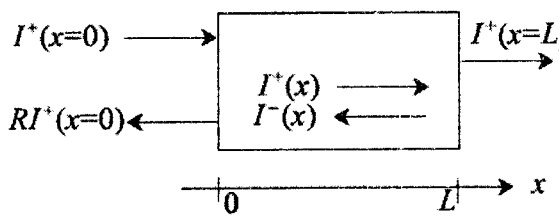


Рис. 24. К выводу связи между коэффициентом прохождения и средней длиной свободного пробега на примере однородного 1D проводника.

Рассмотрим однородный 1D проводник в диффузионном режиме. Левый контакт впрыскивает электроны с током $I^+(x=0)$. Доля электронов T войдет в правый контакт с током $I^+(x=L) = T\,I^+(x=0)$. Оставшаяся доля вернется в левый контакт (рассеяние назад) с током $I^-(x=0) = R\,I^+(x=0)$. В отсутствии процессов

рекомбинации $T+R=1$. Предполагается также, что правый контакт идеальный, поглощает все входящие в него электроны. Результирующий ток, очевидно, будет $I=(1-R)I^{+}(0)=T\,I^{+}(0)$. Так или иначе, в проводнике имеют место как прямые, так и обратные потоки электронов, и нам необходимо описать их пространственное распределение.

Определим обратное значение средней длины свободного пробега $1/\lambda$ как вероятность (на единицу длины) обращения положительного потока электронов в отрицательный и наоборот. Именно исходя из этого определения, величину λ правильно называть средней длиной свободного обратного рассеяния (mean-free-pass for backscattering) или проще средней длиной свободного пробега. В диффузионном проводнике некоторая доля положительного потока в результате рассеяния назад обращается в отрицательный. В проводнике формируется отрицательный поток и его некоторая доля в результате рассеяния назад обращается, усиливая положительный поток. В результате градиент положительного тока складывается из двух величин:

$$\frac{dI^{+}(x)}{dx}=-\frac{I^{+}(x)}{\lambda}+\frac{I^{-}(x)}{\lambda}. \tag{179}$$

Пренебрегая процессами рекомбинации, результирующий ток

$$I=I^{+}(x)-I^{-}(x) \tag{180}$$

является постоянной величиной, так что градиент тока

$$\frac{dI^{+}(x)}{dx}=-\frac{I}{\lambda} \tag{181}$$

является константой. Другими словами, ток спадает линейно вдоль проводника:

$$I^{+}(x)=I^{+}(0)-\frac{I}{\lambda}x. \tag{182}$$

Воспользуемся полученным уравнением для вычисления тока, входящего в правый контакт:

$$I^{+}(L)=I^{+}(0)-\frac{I}{\lambda}L=I^{+}(0)-\left[I^{+}(L)-I^{-}(L)\right]\frac{L}{\lambda}=I^{+}(0)-I^{+}(L)\frac{L}{\lambda}, \tag{183}$$

где мы воспользовались уравнением баланса (180) и тем, что правый контакт идеальный ($I^{-}(L)=0$). Из последнего равенства находим

$$I^{+}(L)=\frac{\lambda}{\lambda+L}I^{+}(0)=T\,I^{+}(0), \tag{184}$$

Если провести аналогичные рассуждения для впрыскивания электронов правым контактом, то получим аналогичное уравнение для электронов, входящих в левый контакт, а именно: $I^{-}(0)=T'I^{-}(L)$. Для однородного

проводника $T' = T$. Проводник под напряжением не является однородным, но нас интересует режим линейного отклика, так что вполне приемлемо положить $T' \approx T$. Окончательно, в предположении независимости друг от друга мод проводимости получаем искомое уравнение (164), связывающее коэффициент прохождения со средней длиной свободного пробега:

$$T(E) = \frac{\lambda(E)}{\lambda(E) + L}. \tag{185}$$

Вывод уравнения (185) сделан в рамках простой модели, что никак не мешает успешному и широкому использованию его на практике. Важным моментом в проведенных рассуждениях является интерпретация $1/\lambda$ как вероятности (на единицу длины) обращения потока частиц в обратном направлении в результате рассеяния. Именно поэтому саму длину λ, как уже упоминалось, часто называют средней длиной свободного рассеяния назад.

Установим связь между λ и временем рассеяния τ. Пусть электрон совершает акт рассеяния в изотропном 1D проводнике. У него есть две возможности: рассеяться вперед и рассеяться назад. Только рассеяние назад существенно для определения средней длины свободного пробега. Отсюда следует, что средняя длина рассеяния назад равна удвоенному значению средней длины рассеяния:

$$\lambda_{1D}(E) = 2\Lambda(E) = 2v(E)\tau(E). \tag{186}$$

Для проводника произвольной размерности средняя длина рассеяния назад дается выражением [33]

$$\lambda(E) = 2\frac{\left\langle v_x^2 \tau \right\rangle}{\left\langle |v_x| \right\rangle}, \tag{187}$$

где усреднение для 2D и 3D проводников ведется по углам. Для изотропных проводников

$$\lambda_{2D}(E) = \frac{\pi}{2} v(E)\tau(E), \tag{188}$$

$$\lambda_{3D}(E) = \frac{4}{3} v(E)\tau(E). \tag{189}$$

Для времени рассеяния τ часто используется степенной закон

$$\tau(E) = \tau_0 \left(\frac{E - E_C}{kT} \right)^s, \tag{190}$$

где значение показателя степени отражает тот или иной механизм рассеяния. Для акустических фононов в 3D проводниках с параболической дисперсией $s = -1/2$, для рассеяния на ионизированных примесях $s = +3/2$ [34].

Аналогичный степенной закон часто используется для средней длины рассеяния:

$$\lambda(E)=\left(\frac{E-E_C}{kT}\right)^r\lambda_0. \tag{191}$$

Для параболической зонной структуры $v(E)\propto E^{1/2}$, так что $r=s+1/2$. Для акустических фононов, соответственно, $r=0$, а для рассеяния на ионизированных примесях $r=2$.

Возвращаясь к нашей первоначальной задаче – выводу уравнения (175) из (178), для 3D проводника, согласно (189), имеем:

$$\lambda_{ph}(\hbar\omega)=\frac{4}{3}v_{ph}(\hbar\omega)\tau_{ph}(\hbar\omega), \tag{192}$$

где, согласно (186),

$$v_{ph}(\hbar\omega)\tau_{ph}(\hbar\omega)=\Lambda_{ph}(\hbar\omega), \tag{193}$$

так что окончательно

$$\lambda_{ph}(\hbar\omega)=\frac{4}{3}\Lambda_{ph}(\hbar\omega). \tag{194}$$

2.3.2.4. Плотность фононных состояний и фононные моды

Ранее было показано, что для 3D проводника число мод и плотность электронных состояний связаны соотношением (глава 1):

$$M_{el}(E)=AM_{3D}(E)=A\frac{h}{4}\left\langle v_x^+(E)\right\rangle D_{3D}(E). \tag{195}$$

Мы хотим эту формулу переписать применительно к фононам. Сначала вспомним, что спиновое вырождение $g_s=2$ входит в плотность состояний, так что в пересчете на один спин

$$D_{3D}(E)=2D'_{3D}(E). \tag{196}$$

Далее, для сферических зон в 3D

$$\left\langle v_x^+(E)\right\rangle=\frac{v_{el}(E)}{2}. \tag{197}$$

Сочетая (196) и (197) с (195) и переходя к фононам, получаем

$$M_{ph}(\hbar\omega)=A\frac{h}{2}\left(\frac{v_{ph}(\hbar\omega)}{2}\right)2D_{ph}(\hbar\omega)=A\frac{h}{4}v_{ph}(\hbar\omega)D_{ph}(\hbar\omega). \tag{198}$$

Подставляя (194) и (198) в (178), получаем очередной промежуточный результат:

$$\kappa_L = \left[\frac{\frac{1}{3}\int_0^\infty \Lambda_{ph}(\hbar\omega) v_{ph}(\hbar\omega) D_{ph}(\hbar\omega) W_{ph}(\hbar\omega) d(\hbar\omega)}{\int_0^\infty D_{ph}(\hbar\omega) W_{ph}(\hbar\omega) d(\hbar\omega)} \right] C_V . \qquad (199)$$

Умножая и деля (199) на

$$\int_0^\infty v_{ph}(\hbar\omega) D_{ph}(\hbar\omega) W_{ph}(\hbar\omega) d(\hbar\omega), \qquad (200)$$

окончательно получаем искомое уравнение (175) с коэффициентом пропорциональности между κ_L и C_V в виде произведения средней длины свободного пробега фононов

$$\left\langle\left\langle \Lambda_{ph} \right\rangle\right\rangle \equiv \frac{\int_0^\infty \Lambda_{ph}(\hbar\omega) v_{ph}(\hbar\omega) D_{ph}(\hbar\omega) W_{ph}(\hbar\omega) d(\hbar\omega)}{\int_0^\infty v_{ph}(\hbar\omega) D_{ph}(\hbar\omega) W_{ph}(\hbar\omega) d(\hbar\omega)} \qquad (201)$$

и средней скорости фононов

$$\left\langle v_{ph} \right\rangle \equiv \frac{\int_0^\infty v_{ph}(\hbar\omega) D_{ph}(\hbar\omega) W_{ph}(\hbar\omega) d(\hbar\omega)}{\int_0^\infty D_{ph}(\hbar\omega) W_{ph}(\hbar\omega) d(\hbar\omega)}, \qquad (202)$$

усредненных должным образом.

Уравнение (175) часто используется для вычисления средней длины сводного пробега фононов из результатов измерений теплопроводности и теплоемкости. Для этого нужно знать среднюю скорость фононов, которая нередко предполагается равной продольной составляющей скорости звука. Приведенные выше выкладки точно определяют смысл усреднений при вычислении средней длины свободного пробега фононов и их средней скорости. Например, задавшись дисперсией фононов всегда можно вычислить их среднюю скорость по (202). Как правило, вычисленная таким образом скорость сильно отличается от продольной составляющей скорости звука, что далее ведет к ошибочным результатам по средней длине свободного пробега фононов [35].

Рассмотрим плотность фононных состояний и число фононных мод с более общих позиций. Плотность состояний фононов определяется аналогично

плотности состояний электронов, как число разрешенных фононных состояний на единицу энергии и единицу длины L для 1D проводников, площади поперечного сечения A для 2D проводников и на единицу объема $\Omega = L \cdot A$ для 3D проводников. В случае электронов, как показано в главе 1, число разрешенных состояний со значением импульса меньшим, чем заданное значение p, для проводников размерности $d = \{1, 2, 3\}$ [36]

$$N(p) = \left\{ 2\frac{L}{h/p},\ \pi\frac{L \cdot W}{(h/p)^2},\ \frac{4\pi}{3}\frac{L \cdot A}{(h/p)^3} \right\}. \tag{203}$$

Аналогично, число разрешенных фононных состояний со значением волнового вектора $k = p/\hbar$ меньшим, чем заданное значение k,

$$\begin{aligned} N_{1D}(k) &= \frac{2k}{2\pi / L}, \\ N_{2D}(k) &= \frac{\pi k^2}{4\pi^2 / L^2}, \\ N_{3D}(k) &= \frac{4\pi k^3 / 3}{8\pi^3 / L^3}, \end{aligned} \tag{204}$$

где единообразия ради далее удобно площадь 2D проводника обозначить как $L^2 \equiv L \cdot W$, а объем для 3D проводника как $L^3 \equiv \Omega = L \cdot A$. Величину L^d назовем «объемом» проводника размерности d.

Плотность фононных состояний, согласно определению выше, очевидно, такова:

$$\begin{aligned} D_{1D}(k) &= \frac{1}{L}\frac{dN_{1D}(k)}{dk} = \frac{1}{\pi}, \\ D_{2D}(k) &= \frac{1}{L^2}\frac{dN_{2D}(k)}{dk} = \frac{k}{2\pi}, \\ D_{3D}(k) &= \frac{1}{L^3}\frac{dN_{3D}(k)}{dk} = \frac{k^2}{2\pi^2}. \end{aligned} \tag{205}$$

Плотность фононных состояний часто выражают через групповую скорость фононов

$$v_g(\omega) = d\omega / dk, \tag{206}$$

а именно [37]:

$$\begin{aligned} D_{1D}(\omega) &= \frac{1}{L}\frac{dN_{1D}(k)}{d\omega} = \frac{1}{L}\frac{dN_{1D}(k)}{dk}\frac{dk}{d\omega} = \frac{1}{\pi v_g(\omega)}, \\ D_{2D}(\omega) &= \frac{1}{L^2}\frac{dN_{2D}(k)}{d\omega} = \frac{1}{L^2}\frac{dN_{2D}(k)}{dk}\frac{dk}{d\omega} = \frac{k(\omega)}{2\pi v_g(\omega)}, \\ D_{3D}(\omega) &= \frac{1}{L^3}\frac{dN_{3D}(k)}{d\omega} = \frac{1}{L^3}\frac{dN_{3D}(k)}{dk}\frac{dk}{d\omega} = \frac{k^2(\omega)}{2\pi^2 v_g(\omega)}. \end{aligned} \tag{207}$$

Количество тепла, переносимого фононами, определяется выражением (150), в котором число состояний поляризации фононов учтено в числе фононных мод $M_{ph}(\hbar\omega)$. Перепишем это выражение через интеграл по частоте:

$$Q=\frac{1}{2\pi}\int_0^\infty(\hbar\omega)T_{ph}(\omega)M_{ph}(\omega)\left(n_1-n_2\right)d\omega. \tag{208}$$

При переносе тепла по 1D проводнику $M_{ph}(\omega)$ есть просто единица, по 2D проводнику $M_{ph}(\omega)$ пропорционально ширине проводника, а по 3D проводнику – площади его поперечного сечения:

$$M_{ph}(\omega)=\left\{M_{1D}(\omega),\ WM_{2D}(\omega),\ AM_{3D}(\omega)\right\}. \tag{209}$$

Для вычисления и интерпретации $M_{2D}(\omega)$ и $M_{3D}(\omega)$ нам потребуются соответствующие выражения для потоков тепла

$$J_{Qx}^{ph}=\left\{J_Q^{ph}(1D),\ J_Q^{ph}(2D),\ J_Q^{ph}(3D)\right\}=\left\{Q,\ Q/W,\ Q/A\right\}, \tag{210}$$

связанных непосредственно с количеством переносимого тепла (178), а именно [37]:

$$\begin{aligned}
J_Q^{ph}(1D)&=\int_0^\infty(\hbar\omega)T_{ph}(\omega)\left[\frac{1}{2}v_g(\omega)D_{1D}(\omega)\right]\left(n_1-n_2\right)d\omega,\\
J_Q^{ph}(2D)&=\int_0^\infty(\hbar\omega)T_{ph}(\omega)\left[\frac{1}{2}\frac{2v_g(\omega)}{\pi}D_{2D}(\omega)\right]\left(n_1-n_2\right)d\omega,\\
J_Q^{ph}(3D)&=\int_0^\infty(\hbar\omega)T_{ph}(\omega)\left[\frac{1}{2}\frac{v_g(\omega)}{2}D_{3D}(\omega)\right]\left(n_1-n_2\right)d\omega,
\end{aligned} \tag{211}$$

где плотности фононных состояний определяются уравнениями (207). Уравнивая потоки тепла (211) с количеством перенесенного тепла (208) согласно (210) и учитывая выражения для плотности фононных состояний (207), окончательно получаем:

$$\begin{aligned}
M_{1D}(\omega)&=\pi\left[v_g(\omega)\right]D_{1D}(\omega)=1,\\
M_{2D}(\omega)&=\pi\left[\frac{2v_g(\omega)}{\pi}\right]D_{2D}(\omega)=\frac{k(\omega)}{\pi},\\
M_{3D}(\omega)&=\pi\left[\frac{v_g(\omega)}{2}\right]D_{3D}(\omega)=\frac{k^2(\omega)}{4\pi},
\end{aligned} \tag{212}$$

или иначе число фононных мод (209), определяющих количество перенесенного тепла (208),

$$M_{ph}(\omega)=\begin{cases} M_{1D}(\omega)=\pi\left[v_g(\omega)\right]\dfrac{1}{\pi v_g(\omega)}=1, \\ W\,M_{2D}(\omega)=W\pi\left[\dfrac{2v_g(\omega)}{\pi}\right]\dfrac{k(\omega)}{2\pi v_g(\omega)}, \\ A M_{3D}(\omega)=A\pi\left[\dfrac{v_g(\omega)}{2}\right]\dfrac{k^2(\omega)}{2\pi^2 v_g(\omega)}, \end{cases} \tag{213}$$

что позволяет интерпретировать число мод $M_{ph}(\omega)$ как максимальное число де-бройлевских полуволн, укладывающихся в «сечении» проводника, поперечном относительно направления движения фононов. Действительно, предполагая линейную зависимость $\omega(k)$, что вполне обосновано для LA и TA акустических мод, имеем $k=2\pi/\lambda$, так что уравнения (213) сводятся к

$$M_{ph}(\omega)=\begin{cases} M_{1D}(\omega)=1, \\ W\,M_{2D}(\omega)=W/(\lambda/2), \\ A M_{3D}(\omega)=A/[(4/\pi)(\lambda/2)^2], \end{cases} \tag{214}$$

где множитель $4/\pi$ появляется в связи с тем, что поперечное сечение 3D проводника предполагается круговым с площадью A, равной квадрату диаметра окружности, деленному на $4/\pi$. Обратим также внимание на то, что выражения в квадратных скобках уравнений (213) есть ничто иное как усредненные по всем угловым переменным значения скорости фононов в направлении их движения (вдоль оси $+x$ в нашем рассмотрении) [38].

2.3.2.5. Дебаевская модель

В случае электронов ширина зоны проводимости $BW \gg kT$, так что заняты только состояния вблизи дна зоны проводимости и поэтому так неплохо работает модель эффективной массы с параболической дисперсией (рис. 18). Совсем другая ситуация в случае фононов: $BW \approx kT$ (рис. 21). Заняты состояния по всей зоне Бриллюэна. Акустические моды тем лучше аппроксимируются дебаевским линейным приближением (19), чем ближе волновые числа к центру зоны Бриллюэна (рис. 20 и 21).

Перепишем (19) в виде

$$\hbar\omega=\hbar v_D q, \tag{215}$$

где под дебаевской скоростью v_D понимают среднюю скорость продольных и поперечных акустических мод. Тогда, по аналогии с электронами, плотность фононных состояний

$$D_{ph}(\hbar\omega)=\frac{3(\hbar\omega)^2}{2\pi^2(\hbar v_D)^3},\ [Дж^{-1}\cdot м^{-3}] \tag{216}$$

где множитель 3 учитывает три поляризации акустических мод. Обратим внимание, что в учебниках плотность фононных состояний чаще выражается в $Гц^{-1} \cdot м^{-3}$:

$$D_{ph}(\hbar\omega) = \frac{D_{ph}(\omega)}{\hbar}.$$

Число мод на единицу площади поперечного сечения проводника вычисляем по (198):

$$M_{ph}(\hbar\omega) = \frac{3(\hbar\omega)^2}{4\pi(\hbar v_D)^2}. \qquad (217)$$

Поскольку все состояния в зоне Бриллюэна при умеренных температурах стремятся быть заполненными, нужно проконтролировать, что действительно учтены все состояния. В кристалле с N/Ω атомами на единицу объема имеем $3N/\Omega$ состояний на единицу объема. Полное число состояний дается интегралом

$$\int_0^{\hbar\omega_D} D_{ph}(\hbar\omega)\, d(\hbar\omega), \qquad (218)$$

в котором верхний предел, определяемый дебаевской частотой, выбирается таким образом, чтобы получить правильное число состояний $3N/\Omega$, а именно:

$$\hbar\omega_D = \hbar v_D \left(\frac{6\pi^2 N}{\Omega}\right)^{1/3} \equiv kT_D. \qquad (219)$$

Дебаевская частота является предельной, выше которой фононные состояния уже не существуют или не учитываются. Это ограничение можно ввести и через дебаевский волновой вектор q_D или через дебаевскую температуру T_D через (219). Для $T \ll T_D$ заняты только состояния с $q \to 0$, для которых хорошо выполняется дебаевское приближение.

Теперь интегрирование в (169) можно вести только до дебаевской энергии:

$$\kappa_L = \frac{\pi^2 k^2 T}{3h} \int_0^{\hbar\omega_D} \frac{M_{ph}(\hbar\omega)}{A} \lambda_{ph}(\hbar\omega) W_{ph}(\hbar\omega)\, d(\hbar\omega) \qquad (220)$$

и брать число фононных мод по (217). Интеграл можно взять численно или аналитически с приемлемым выражением для средней длины свободного пробега фононов. Именно таким образом была впервые вычислена решеточная теплопроводность [39, 40]. Дальнейшее развитие теория и методы вычисления термоэлектрических транспортных коэффициентов получили в работах [33, 35, 41].

2.3.2.6. Рассеяние фононов

Фононы рассеиваются на незаряженных дефектах, примесных атомах, изотопах, на других фононах, на поверхностях и границах, на электронах. Фонон-фононное рассеивание обязано тому, что потенциальная энергия колеблющегося атома в кристалле не совсем гармоническая. Все более высокие порядки учитываются в виде потенциала рассеяния. Рассматривают две возможности фонон-фононного взаимодействия. В обычном процессе два фонона взаимодействуя порождают третий фонон с сохранением импульса и энергии:

$$\begin{aligned} \hbar\vec{q}_1 + \hbar\vec{q}_2 &= \hbar\vec{q}_3 \\ \hbar\vec{w}_1 + \hbar\vec{w}_2 &= \hbar\vec{w}_3 \end{aligned} \quad (221)$$

Поскольку суммарный импульс сохраняется, то такой процесс не может дать существенный вклад в перенос тепла.

В umklapp/U-процессах рассеяния суммарный импульс не сохраняется, генерируемый фонон характеризуется импульсом за пределами первой зоны Бриллюэна. Это может быть как ангармоничное фонон-фононное взаимодействие, так и электрон-фононное. U-рассеяние является основным процессом в переносе тепла, особенно при высоких температурах. Ощутимый вклад дают также рассеяние на дефектах/D и на границах/B. Скорости всех трех процессов складываются, так что

$$\frac{1}{\tau_{ph}(\hbar\omega)} = \frac{1}{\tau_U(\hbar\omega)} + \frac{1}{\tau_D(\hbar\omega)} + \frac{1}{\tau_B(\hbar\omega)} \quad (222)$$

или через соответствующие длины свободного пробега

$$\frac{1}{\lambda_{ph}(\hbar\omega)} = \frac{1}{\lambda_U(\hbar\omega)} + \frac{1}{\lambda_D(\hbar\omega)} + \frac{1}{\lambda_B(\hbar\omega)}, \quad (223)$$

поскольку для 3D сферических дисперсий λ и τ связаны уравнением (192).

Для всех механизмов рассеяния предложены выражения для скоростей [42]. Например, при рассеянии на дефектах

$$\frac{1}{\tau_D(\hbar\omega)} \propto \omega^4, \quad (224)$$

что напоминает рэлеевское рассеяние света на частицах пыли в атмосфере; высокочастотные фононы особенно чувствительны к рассеянию на малых дефектах. При рассеянии на границах и поверхностях

$$\frac{1}{\tau_B(\hbar\omega)} \propto \frac{v_{ph}(\hbar\omega)}{L^{\min}}, \quad (225)$$

где L^{min} – наименьшая протяженность образца. Сложнее с U-рассеянием. Обычно используется выражение

$$\frac{1}{\tau_U(\hbar\omega)} \propto T^3\omega^2 e^{-T_D/bT}. \qquad (226)$$

2.3.2.7. Температурная зависимость теплопроводности

Температурная зависимость решеточной теплопроводности κ_L кремния представлена на рис. 25.

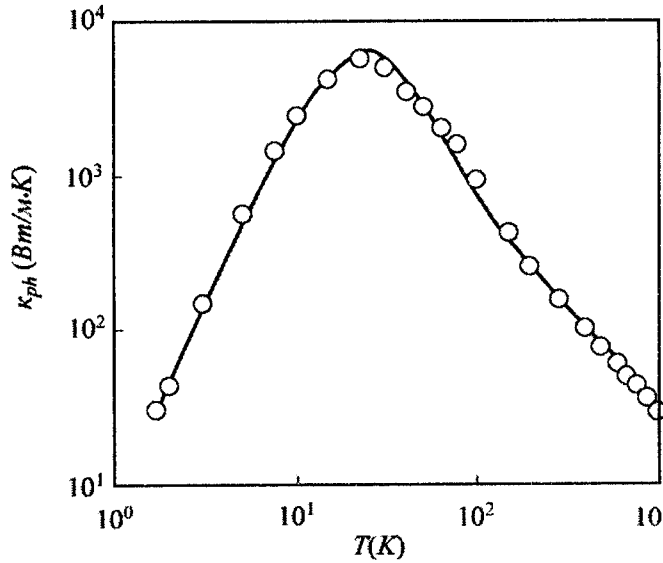


Рис. 25. Температурная зависимость решеточной теплопроводности кремния: экспериментальные данные по [43], расчетная кривая по [35].

Согласно уравнению (171), κ_L пропорциональна числу фононных мод $\langle M_{ph}/A \rangle$ и усредненному значению средней длины свободного пробега фононов $\langle\langle \lambda_{ph} \rangle\rangle$. Изменение теплопроводности с температурой качественно можно объяснить следующим образом. Из определения числа мод (170) можно показать, что при низких температурах

$$\langle M_{ph} \rangle \propto T^3; \ (T \to 0) \qquad (227)$$

быстрое увеличение числа заселенных мод с температурой вызывает рост теплопроводности. Важную роль играет рассеяние на границах. По мере роста температуры все больше продуцируется коротковолновых фононов (бóльшие

значения q). Возрастает роль рассеяния таких фононов на дефектах. По мере приближения температуры к дебаевской заселяются все фононные моды и дальнейшее увеличение температуры уже не увеличивает числа мод. Более того, высокие температуры порождают U-процессы рассеяния и теплопроводность начинает падать.

2.3.2.8. Различие между решеточной теплопроводностью и электрической проводимостью

Уже подчеркивалось близкое подобие пары фононных транспортных уравнений (167) и (171) и пары электронных уравнений (173) и (174). Длины свободных пробегов электронов и фононов одного порядка. Почему же электрическая проводимость веществ меняется на много порядков, тогда как теплопроводность всего разнообразия веществ изменяется лишь на несколько порядков? Ответ кроется в различии между окнами проводимости для электронов (158) и фононов (160). Для тех и других рост температуры расширяет окна проводимости и тем самым увеличивает заселенность состояний. Для электронов, однако, все зависит от положения уровня Ферми, сдвигая который вверх или вниз можно менять проводимость на много порядков. Для фононов же ширина окна проводимости определяется только температурой.

Подчеркнем еще одно практически важное различие между описанием электронов и фононов, связанное с особенностями заселения состояний. В термоэлектрических устройствах $E_F \approx E_C$ и окна проводимости электронов и фононов схожи. У электронов, однако, ширина зон очень большая и лишь небольшое число состояний вблизи дна зоны проводимости заселено. Хорошо работает параболическая дисперсия и легко получить аналитические решения. В сопоставимой ситуации у фононов ширина зон незначительная, при повышенных температурах практически все состояния в зоне Бриллюэна заселены и простые аналитические модели для вычисления теплопроводности не работают.

2.3.2.9. Квантование теплопроводности

По аналогии с квантованием электронной проводимости

$$G^{ball} = \frac{2q^2}{h} M(E_F) \tag{228}$$

еще 30 лет назад Пендри [44] показал, что должны существовать квантовые ограничения на перенос тепла. Действительно, в уравнении (162) при $T \to 0$ окно фононной проводимости $W_{ph}(\hbar\omega)$ стремится к δ-функции, так что в пределе

$$K_L = \frac{\pi^2 k^2 T}{3h} T_{ph}(0) M_{ph}(0). \tag{229}$$

Для массивных проводников $M_{ph}(\hbar\omega) \to 0$ при $\hbar\omega \to 0$, однако, для наноразмерных проводников всегда есть конечное число фононных мод. В случае баллистического транспорта $T_{ph} = 1$ и в этом случае можно ожидать что

$$K_L = \frac{\pi^2 k^2 T}{3h} M_{ph}. \qquad (230)$$

Именно этот результат был подтвержден экспериментально на 4-хмодовом нанорезисторе при $T < 0.8K$ [32]; результаты измерений теплопроводности полностью согласуется с предсказаниями для одномерных баллистических резисторов [45 – 47].

Квант теплопроводности (163)

$$g_0 \equiv \pi^2 k^2 T / 3h$$

есть максимальная энергия, переносимая в пересчете на одну фононную моду. Удивительно, что квант теплопроводности не зависит от статистики частиц, переносящих тепло, он универсален для фермионов, бозонов и любых аньенов/anyons [48 – 50].

Подведем итоги. В этой главе подчеркивается, что концепции, используемые для описания электронного транспорта, с успехом могут быть перенесены на фононный транспорт. И в том и в другом случае подход Ландауэра, обобщенный впоследствии Даттой и Лундстромом, позволяет количественно описать транспортные процессы в проводниках любой размерности и при любых режимах транспорта – от диффузионного до баллистического. Решеточная теплопроводность и электронная проводимость описываются очень схожими по форме уравнениями. Эти оба свойства всегда пропорциональны соответствующему кванту проводимости, умноженному на интеграл от произведения трех величин – коэффициента прохождения, числа мод и окна проводимости. Есть, однако, два существенных отличия, связанных с физикой электронных и фононных процессов.

Первое отличие касается ширины зон. В случае электронов ширина зон намного превышает kT и при комнатной температуре заселены лишь уровни вблизи дна зоны проводимости. Ширина фононных дисперсий всего лишь порядка kT и при комнатной температуре все акустические моды в пределах всей зоны Бриллюэна практически заселены. В результате дебаевская модель для акустических мод далеко не столь успешно работает по сравнению с широкой применимостью параболической дисперсии для электронов.

Второе существенное различие связано с тем, что в случае электронов заселенность мод проводимости, контролируемая фермиевским окном проводимости, зависит не только от температуры, как это имеет место в случае фононов, но и от положения уровня Ферми. Именно по этой причине электронная проводимость различных материалов может изменяться в очень широких пределах, тогда как теплопроводность изменяется в очень узких пределах всего на несколько порядков.

Эту главу мы закончим рассмотрением термоэлектрических явлений в концепции «снизу – вверх» Ландауэра – Датты.

2.3.3. Термоэлектричество «снизу – вверх» в концепции Ландауэра – Датты

Измерения проводимости сами по себе не несут никакой информации о физической природе проводимости в том или ином проводнике. В замкнутой цепи электроны всегда движутся от отрицательно заряженного контакта через проводник к положительно заряженному контакту какая бы ни была физико-химическая природа проводника. Это верно для любых проводников и поэтому измерения проводимости ничего не говорят нам о самом проводнике.

Термоэлектрические же явления предоставляют возможность судить о физической природе проводимости в изучаемом проводнике. Простейшим примером могут служить проводники *n*- и *p*-типа. В проводниках *n*-типа электроны движутся от горячего контакта через проводник к холодному контакту, а в проводниках *p*-типа направление движения электронов обратное. Почему? Часто объясняют таким образом, мол, в проводниках *p*-типа носители тока (так называемые «дырки») заряжены противоположно заряду электрона и поэтому направление движения тока обратное. В главе 1 уже обращалось внимание на то, что в проводниках *p*-типа плотность состояний $D(E)$ нисходящая с ростом энергии, а в *n*-типа – восходящая (рис. 26).

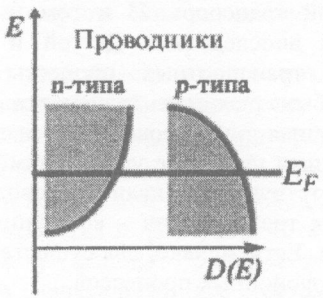


Рис. 26. В проводниках *n*-типа электрохимический потенциал E_F расположен ближе к дну зоны проводимости, а в *p*-типа – ближе к верхушке валентной зоны.

Для проводников *p*-типа электроны движутся с энергиями вблизи верхушки валентной зоны, отслеживать удобно пустые состояния выше значения электрохимического потенциала E_F, а не занятые состояния ниже значения E_F. Эти пустые состояния называют, «дырками», поскольку в этих состояниях нет электронов и ведут они себя как положительно заряженные «сущности». Подобное объяснение нельзя признать удовлетворительным, поскольку в реальности движутся только электроны. «Дырки» же являются, в лучшем случае, концептуальным понятием, а измеряемые в экспериментах эффекты вызваны только движением электронов и не могут зависеть от субъективных, хотя и очень удобных, договоренностей. Мы ниже еще раз убедимся в том, что объяснение различия между *n*- и *p*-проводниками не

требует никаких новых концепций кроме очевидного утверждения, что ток порождается различием между фермиевскими функциями на контактах проводника. Существенно лишь то, чем отличаются *n*- и *p*-проводники друг от друга: для первых плотность энергии растет с ростом энергии, а для вторых – падает. Для молекул в роли проводников вообще нет понятия дисперсии $E(p)$, а в этой роли молекулы обнаруживают весьма заметные термоэлектрические эффекты [51].

Начнем с выражения для тока:

$$I=\frac{1}{q}\int_{-\infty}^{+\infty} dEG(E)\Big(f_1(E)-f_2(E)\Big). \qquad (231)$$

Различие между фермиевскими функциями порождаться также различием в температуре контактов:

$$f_1(E)=\frac{1}{e^{(E-E_{F1})/kT_1}+1},\quad f_2(E)=\frac{1}{e^{(E-E_{F2})/kT_2}+1}. \qquad (232)$$

Как уже обсуждалось выше причина кроется в поведении разности фермиевских функций горячего и холодного контактов (рис. 27).

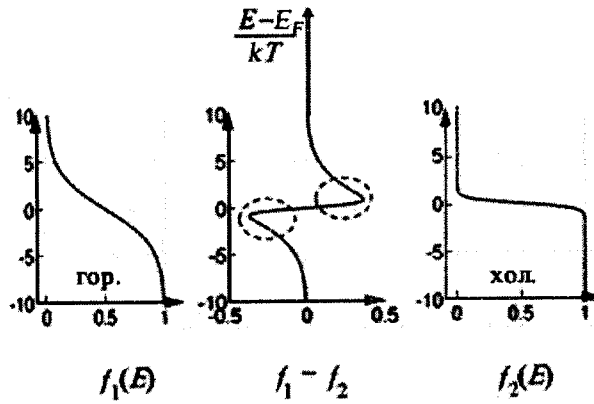


Рис. 27. Два контакта, горячий и холодный, с одним и тем же значением E_F: разность фермиевских функций положительна для $E > E_F$ и отрицательна для $E < E_F$.

В канале *n*-типа проводимость $G(E)$ – возрастающая функция энергии, так что в суммарном токе доминируют состояния с $E > E_F$, электроны движутся от левого контакта к правому, от горячего к холодному (рис. 28). В *p*-каналах ситуация противоположная. Проводимость – убывающая функция

энергии, так что в суммарном токе доминируют состояния с $E < E_F$, электроны движутся от правого контакта к левому, от холодного к горячему (рис. 28).

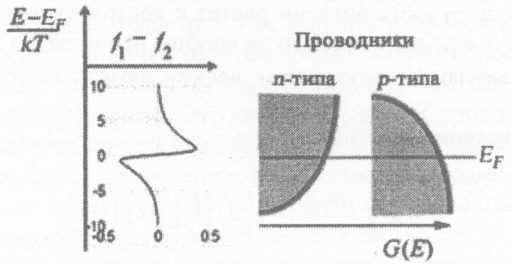


Рис. 28. Для n-проводников ток для $E > E_F$ доминирует над током для $E < E_F$, а для p-проводников наоборот – ток для $E < E_F$ доминирует над током для $E > E_F$.

2.3.3.1. Еще раз о коэффициенте Зеебека

Для вычисления тока за счет различия в температуре контактов можно пользоваться формулой (231) напрямую, не вводя никаких упрощений. Однако, в случае небольших различий в температуре контактов, как это уже было продемонстрировано выше для случая небольшой разницы между электрохимическими потенциалами, что соответствует участку линейного отклика вольт-амперной характеристики, можно воспользоваться разложением разности фермиевских функций в (231) в ряд Тейлора и ограничиться линейным членом. Тогда получим

$$I = G(V_1 - V_2) + G_S(T_1 - T_2). \tag{233}$$

Отсюда видно, что электроны движутся от горячего контакта через n-проводник к холодному контакту, а в p-проводниках наоборот. Перепишем (233) иначе:

$$\Delta V = I/G + (-G_S/G)\Delta T, \tag{234}$$

где V_1 и V_2 определены как E_{F1}/q и E_{F2}/q, а проводимость G дается формулой Ландауэра

$$G = \int_{-\infty}^{+\infty} dE\, G(E)\left(\frac{\partial f_0}{\partial E_F}\right) = \int_{-\infty}^{+\infty} dE\left(-\frac{\partial f_0}{\partial E}\right) G(E), \tag{235}$$

тогда как зеебековская проводимость

$$G_S = \frac{1}{q}\int_{-\infty}^{+\infty} dE\, G(E)\left(\frac{\partial f_0}{\partial T}\right) = \int_{-\infty}^{+\infty} dE\left(-\frac{\partial f_0}{\partial E}\right)\frac{E - E_{F0}}{qT} G(E). \tag{236}$$

В уравнении (234) коэффициентом Зеебека S называют отношение

$$S = -G_S/G. \tag{237}$$

Уравнение (236) математически выражает тот факт, что состояния с энергией, большей и меньшей электрохимического потенциала, дают вклад с разными знаками в термоэлектрический коэффициент Зеебека S. Из этого же уравнения ясно, что при подборе материала с наилучшим коэффициентом Зеебека следует выбирать такой материал, в котором вся плотность состояний сосредоточена по одну сторону от электрохимического потенциала E_{F0}, поскольку даже небольшая часть состояний по другую сторону от E_{F0} лишь ухудшит коэффициент Зеебека.

Соотношения (233) и (234) можно проиллюстрировать двумя электрическими цепями (рис. 29): короткозамкнутой с током (слева на рисунке)

$$I_{sc} = G_S(T_1 - T_2) \tag{238}$$

и незамкнутой (справа на рис. 29), которая обычно и используется в термоэлектрических измерениях и дающая напряжение

$$V_{oc} = -\frac{I_{sc}}{G} = -\frac{G_S}{G}(T_1 - T_2). \tag{239}$$

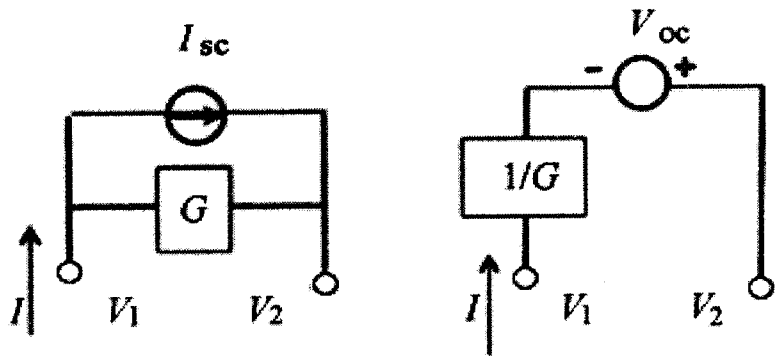


Рис. 29. Короткозамкнутая sc (short circuit) и незамкнутая oc (open circuit) цепи, иллюстрирующие уравнения (233) и (234). Для направления тока (потока электронов) и знака напряжения используются естественные представления, которые противоположны обычно условно принятым (рис. 16/гл. 1).

К примеру, для n-проводников зеебековская проводимость положительна, так что напряжение V_{oc} отрицательное, если $T_1 > T_2$. Это означает, что контакт с более высокой температурой заряжен отрицательно (положительно в общепринятом смысле). По договоренности, коэффициент Зеебека

$$S \equiv \frac{V_{oc}}{T_1 - T_2} = -\frac{G_S}{G}. \tag{240}$$

2.3.3.2. О показателях качества ТЭ устройств и материалов

Максимальная мощность P^{max} при конвертации избыточного тепла в электрический ток генерируется тогда, когда сопротивление нагрузки $R_L = 1/G$ (рис. 30):

$$P^{max} = V_{oc}^2 G/4 = S^2 G (T_1 - T_2)^2/4 . \qquad (241)$$

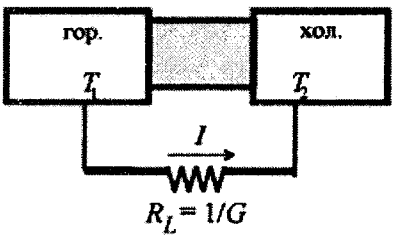


Рис. 30. Термоэлектрический генератор создает ток за счет разности температур.

К.п.д. ТЭ устройства как отношение максимальной генерируемой мощности к мощности внешнего источника

$$\eta \equiv \frac{P^{max}}{G_K (T_1 - T_2)} = \frac{S^2 G T}{G_K} \frac{T_1 - T_2}{4T}, \qquad (242)$$

где G_K – теплопроводность проводника, а T есть средняя температура $(T_1 + T_2)/2$. Стандартный показатель качества термоэлектрического материала ZT пропорционален отношению $S^2 G$ к G_K:

$$ZT \equiv \frac{S^2 G T}{G_K} = \frac{S^2 \sigma T}{\kappa}, \qquad (243)$$

где удельная теплопроводность κ связана с теплопроводностью G_K материала проводника:

$$G_K = \kappa \cdot A/L \qquad (244)$$

так же, как проводимость связана с удельной проводимостью. Закон Фурье для теплопроводности, аналог закона Ома для электропроводности, в случае применения к нанопроводникам нуждается в такой же коррекции $(L \Rightarrow L + \lambda)$ как и закон Ома (глава 1).

Если электрическая проводимость определяется исключительно переносом электронов, то теплопроводность включает также существенную фононную составляющую. Сейчас же мы рассматриваем перенос тепла только электронами.

2.3.3.3. Перенос тепла электронами в концепции «снизу – вверх»

До сих пор при обсуждении ТЭ явлений проводимость материала предполагадась произвольной $G(E)$. Модель упругого резистора (§ 1.3.12) привлекательна тем, что моды проводимости при разных энергиях проводят электроны одновременно и в параллельном режиме, так что можно рассмотреть проводимость при некотором одном значении энергии, а затем отсуммировать по всему спектру энергий. Рассмотрим небольшой промежуток энергии между значениями энергии E и $E+dE$, расположенный как выше значения электрохимического потенциала E_{F0} (моды n-типа), так и ниже его (моды p-типа) (рис. 31). Как уже упоминалось выше, вклады этих двух разных типов мод в зеебековскую проводимость будут противоположны по знаку.

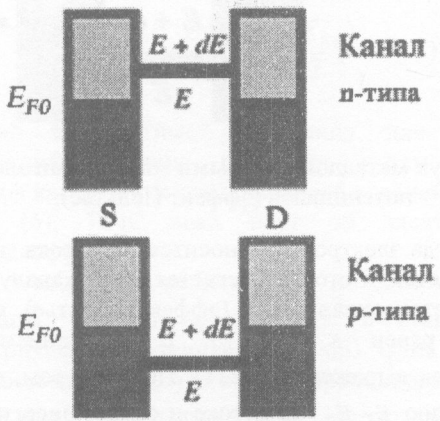


Рис. 31. Простейшая одномодовая проводимость $G(E)$ с каналом проводимости между энергиями E и $E+dE$ ($dE \ll kT$) выше (канал n-типа) и ниже (канал p-типа) значения электрохимического потенциала.

В упругом резисторе при прохождении тока джоулево тепло I^2R диссипирует на контактах проводника с истоком S и стоком D. В системах n- и p-типа на рис. 31 тепло на контактах не выделяется, поскольку тока в каналах нет (одинаковые электрохимические потенциалы на истоке и стоке). Рассмотрим, однако, ситуацию на рис. 32 с разными электрохимическими потенциалами на истоке и на стоке, что соответствует, например, контакту двух разных металлов.

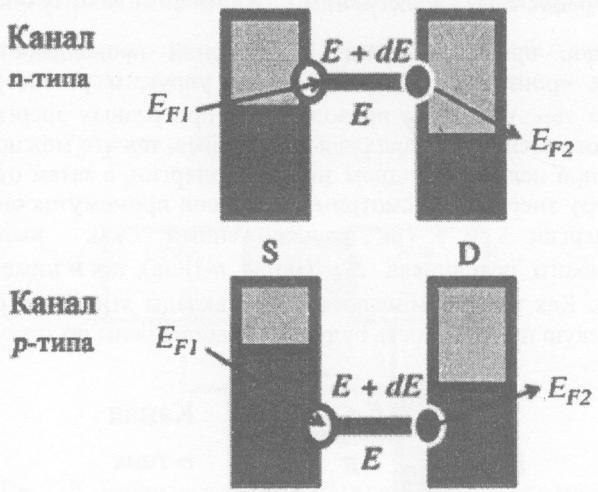


Рис. 32. Контакт двух металлов с разными значениями электрохимических потенциалов (эффект Пельтье).

Каждый раз, когда электрон переносится с истока на сток по каналу *n*-типа исток охлаждается, а сток нагревается, а по каналу *p*-типа наоборот – исток нагревается, а сток охлаждается (эффект Пельтье), однако, каждый раз суммарный эффект равен $E_{F1}-E_{F2}=qV$ с выделением джоулева тепла. Зарядовый ток I дается выражением (231). Помня о том, что в окне $E_{F1}-E_{F2}$ электрон теряет энергию $E-E_{F1}$ на истоке и отдает энергию $E-E_{F2}$ на стоке, выпишем энергетические вклады термотоков отдельно для истока и стока

$$I_{Q1}=\frac{1}{q}\int_{-\infty}^{+\infty}dE\frac{E-E_{F1}}{q}G(E)(f_1(E)-f_2(E)) \tag{245}$$

и

$$I_{Q2}=\frac{1}{q}\int_{-\infty}^{+\infty}dE\frac{E_{F2}-E}{q}G(E)(f_1(E)-f_2(E)). \tag{246}$$

Энергия, извлекаемая из внешнего источника напряжения,

$$I_E=\frac{E_{F1}-E_{F2}}{q}I=V\cdot I, \tag{247}$$

так что закон сохранения энергии соблюдается

$$I_{Q1}+I_{Q2}+I_E=0. \tag{248}$$

2.3.3.4. Квазиравновесный режим

Линеаризуем уравнения (245) и (246) точно так же, как было получено ур-е (233) из ур-я (231), и получим

$$I_Q = G_P\left(V_1 - V_2\right) + G_Q\left(T_1 - T_2\right), \tag{249}$$

где

$$G_P = \int_{-\infty}^{+\infty} dE\left(-\frac{\partial f_0}{\partial E}\right)\frac{E - E_{F0}}{q}G(E) \tag{250}$$

и

$$G_Q = \int_{-\infty}^{+\infty} dE\left(-\frac{\partial f_0}{\partial E}\right)\frac{(E - E_{F0})^2}{q^2 T}G(E) \tag{251}$$

представляют собой стандартные выражения для термоэлектрических коэффициентов, которые обычно получают путем применения таких изощренных подходов как техника транспортного уравнения Больцмана [5] или формализма Кубо [6]. Речь пока идет об электронных вкладах в термоэлектрические коэффициенты.

Следует иметь в виду, что величина G_Q не есть теплопроводность G_K материала проводника, обычно используемая в стандартном показателе качества термоэлектрического материала ZT (243). Одна из причин этого, как уже упоминалось, состоит в том, что в теплопроводности G_K есть фононный вклад, который пока что не учитывался нами. Но есть еще одна, совершенно другая причина.

Величина G_K определяется как теплопроводность проводника в условиях незамкнутой электрической цепи ($I = 0$):

$$G_K = \left(\frac{\partial I_Q}{\partial(T_1 - T_2)}\right)_{I=0}, \tag{252}$$

тогда как величина G_Q есть теплопроводность проводника в условиях короткозамкнутой цепи ($V = 0 : V_1 = V_2$), а именно:

$$G_Q = \left(\frac{\partial I_Q}{\partial(T_1 - T_2)}\right)_{V_1=V_2}. \tag{253}$$

Можно, однако, переписать уравнения (233) и (249) таким образом, чтобы в явном виде выделялись коэффициенты, обычно экспериментально измеряемые в условиях незамкнутой электрической цепи:

$$(V_1 - V_2) = \frac{1}{G} I - \frac{G_S}{G}(T_1 - T_2), \tag{254}$$

$$I_Q = \frac{G_P}{G} I + \left(G_Q - \frac{G_P G_S}{G} \right)(T_1 - T_2), \tag{255}$$

имея, конечно, в виду, что V и I есть напряжение E_F/q и ток, отличающиеся по знаку от общепринятых значений, где $-G_S/G$ есть коэффициент Зеебека (240), коэффициент Пельтье

$$\Pi = -G_P / G, \tag{256}$$

а теплопроводность

$$G_K = G_Q - G_P G_S / G. \tag{257}$$

Если обратиться к уравнениям (236) и (250), коэффициенты Зеебека и Пельтье в уравнениях (254) и (255) оказываются связанными соотношением Кельвина

$$\Pi = T S, \tag{258}$$

которое является частным случаем фундаментальных соотношений Онзагера [7].

2.3.3.5. Дельта-термоэлектрик

Иногда дельта-термоэлектрик называют одноуровневым или одномодовым термоэлектриком. Это гипотетическое вещество с узким каналом проводимости при энергии ε шириной $\Delta\varepsilon << kT$ (рис. 33). Полезно рассмотреть термоэлектрические коэффициенты для такого гипотетического вещества.

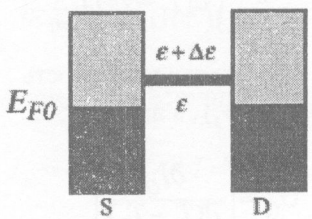


Рис. 33. Модель дельта-термоэлектрика.

Выпишем сводку ранее уже полученных термоэлектрических коэффициентов:

$$G=\int_{-\infty}^{+\infty}dE\left(-\frac{\partial f_0}{\partial E}\right)G(E),$$

$$G_S=\int_{-\infty}^{+\infty}dE\left(-\frac{\partial f_0}{\partial E}\right)\frac{E-E_{F0}}{qT}G(E),$$

$$G_P=\int_{-\infty}^{+\infty}dE\left(-\frac{\partial f_0}{\partial E}\right)\frac{E-E_{F0}}{q}G(E),$$

$$G_Q=\int_{-\infty}^{+\infty}dE\left(-\frac{\partial f_0}{\partial E}\right)\frac{(E-E_{F0})^2}{q^2T}G(E).$$

Проводимость $G(E)$ в подынтегральном выражении отлична от нуля в очень узкой области энергии, а множители $(E-E_{F0})$ практически постоянны и могут быть вынесены из под знака интеграла, что дает

$$G=G(\varepsilon)\Delta\varepsilon\left(-\frac{\partial f_0}{\partial E}\right)_{E=\varepsilon},\qquad(259)$$

$$G_S=\frac{\varepsilon-E_{F0}}{qT}G,\qquad(260)$$

$$G_P=\frac{\varepsilon-E_{F0}}{q}G,\qquad(261)$$

$$G_Q=\frac{(\varepsilon-E_{F0})^2}{q^2T}G,\qquad(262)$$

так что из уравнений (254) и (255) теперь легко получаются термоэлектрические коэффициенты для дельта-термоэлектрика:

$$S=-\frac{G_S}{G}=-\frac{\varepsilon-E_{F0}}{qT},\qquad(263)$$

$$\Pi=-\frac{G_P}{G}=-\frac{\varepsilon-E_{F0}}{q},\qquad(264)$$

$$G_K=G_0-\frac{G_PG_S}{G}=0.\qquad(265)$$

Нулевая теплопроводность G_K для дельта-термоэлектрика заслуживает обсуждения. Однако, сначала рассмотри физический смысл коэффициентов

Зеебека (263) и Пельтье (264) для дельта-термоэлектрика. Коэффициент Зеебека (263) есть напряжение в незамкнутой цепи, обеспечивающее нулевой ток. Дельта-термоэлектрик проводит только при энергии $E = \varepsilon$. Чтобы тока не было, фермиевские функции при этой энергии должны быть одинаковы: $f_1(\varepsilon) = f_2(\varepsilon)$, что в свою очередь дает

$$\frac{\varepsilon - E_{F1}}{kT_1} = \frac{\varepsilon - E_{F2}}{kT_2} = \frac{(\varepsilon - E_{F1}) - (\varepsilon - E_{F2})}{k(T_1 - T_2)} = -\frac{E_{F1} - E_{F2}}{k(T_1 - T_2)}. \qquad (266)$$

Поскольку же коэффициент Зеебека при нулевом токе

$$S = \frac{(E_{F1} - E_{F2})/q}{T_1 - T_2}\Big|_{I=0}, \qquad (267)$$

окончательно получаем

$$S = -\frac{\varepsilon - E_{F1}}{qT_1} = -\frac{\varepsilon - E_{F2}}{qT_2} = -\frac{\varepsilon - E_{F0}}{qT}, \qquad (268)$$

что согласуется с ранее полученным выражением (263).

Выражение (264) для коэффициента Пельтье дельта-термоэлектрика также легко понять, если вспомнить, что каждый электрон в дельта-термоэлектрике переносит тепло $(\varepsilon - E_{F0})$, так что отношение перенесенного тепла к заряду электрона $(-q)$ должно быть $(\varepsilon - E_{F0}) / (-q)$.

В итоге имеет место нулевая теплопроводность (265), что свидетельствует о том, что в дельта-термоэлектрике в условиях незамкнутой цепи (нет транспорта электронов) перенос тепла не происходит. Вроде бы так. Однако, все же если нет электрического тока, может ли иметь место перенос тепла? Если бы модель дельта-термоэлектрика была достаточной для описания реального термоэлектрика, то не должен был бы наблюдаться перенос тепла в любом термоэлектрике.

Чтобы разобраться в этой ситуации, усложним модель и рассмотрим двухуровневый термоэлектрик с перепадом температур (рис. 34).

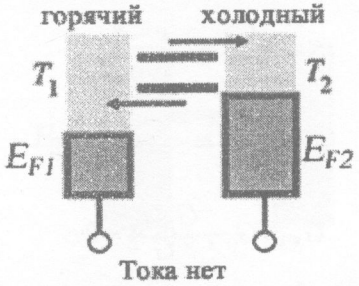


Рис. 34. Двухуровневый термоэлектрик с перепадом температур в режиме незамкнутой цепи.

В этом случае между двумя контактами с $E_{F1} < E_{F2}$ возникает разность потенциалов. Хотя суммарный ток нулевой, но индивидуальные токи отличны от нуля на разных уровнях. Они одинаковы по величине, но противоположны по направлению, так что суммарный ток равен нулю. Однако, потоки тепла на обоих уровнях не компенсируют друг друга, поскольку по уровню с большей энергией переносится больше тепла, чем по уровню с меньшей энергией. Таким образом, нулевой электрический ток не означает нулевой перенос тепла в реальных термоэлектриках, разве что это имеет место лишь в дельта-термоэлектрике.

Поскольку дельта-термоэлектрик имеет нулевую теплопроводность, то стандартный показатель качества термоэлектрического материала ZT (243) для дельта-термоэлектрика должен быть очень большим с претензией на идеальный термоэлектрик. Как уже упоминалось выше, даже при нулевом переносе тепла электронами имеет место перенос тепла фононами, что не позволяет показателю качества термоэлектрика ZT принимать слишком большие значения. Роль фононов мы подробно рассмотрим ниже.

2.3.3.6. Оптимизация силового коэффициента

Обсудим факторы, которые могли бы максимизировать силовой коэффициент S^2G в выражении (241) для максимальной мощности $P^{\max}$. Если стремиться к максимально возможному коэффициенту Зеебека, то энергию ε в (263) нужно выбирать как можно дальше от электрохимического потенциала E_{F0}. Это, однако, приводит к неприемлемо низкому значению проводимости G (259), поскольку множитель $(-\partial f_0/\partial E)$ затухает очень быстро при удалении E от E_{F0}. Из уравнений (263) и (259) имеем

$$S^2G = G(\varepsilon)\Delta\varepsilon\left(\frac{\varepsilon - E_{F0}}{qT}\right)^2\left(-\frac{\partial f_0}{\partial E}\right)_{E=\varepsilon} = G(\varepsilon)\frac{\Delta\varepsilon}{kT}\left(\frac{k}{q}\right)^2 x^2\frac{e^x}{(e^x+1)^2}, \quad (269)$$

где $x \equiv (\varepsilon - E_{F0})/kT$, а зависимость силового коэффициента от x дается функцией $F(x) = x^2e^x/(e^x + 1)^2$, график которой показан на рис. 35.

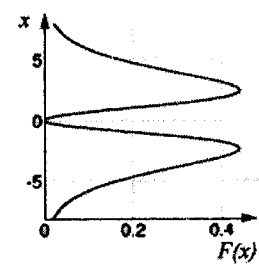


Рис. 35. График функции $F(x) = x^2e^x/(e^x + 1)^2$.

Максимум функции $F(x)$ лежит при $x \sim \pm 2$, т.е. в идеальном случае энергия в канале проводника должна лежать выше или ниже значения электрохимического потенциала на величину, равную приблизительно $2kT$. Соответствующие значения коэффициента Зеебека и силового коэффициента приблизительно равны:

$$S \approx 2\frac{k}{q}, \tag{270}$$

$$S^2 G \approx 0.5\left(\frac{k}{q}\right)^2 G(\varepsilon)\frac{\Delta\varepsilon}{kT}. \tag{271}$$

Наилучшие термоэлектрики характеризуются коэффициентами Зеебека, не сильно отличающимися от значения $2(k/q) = 170\,\mu B/K$, ожидаемого согласно (270). Они обычно проектируются таким образом, чтобы электрохимический потенциал E_{F0} находился чуть ниже дна зоны проводимости, так что произведение $G(E)$ и $(-\partial f_0/\partial E)$ выглядит как дельта-функция, центрированная при $E = \varepsilon$ и лежащая чуть выше дна зоны проводимости, как показано на рис. 36.

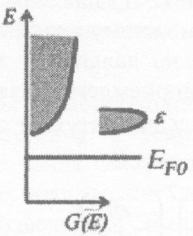


Рис. 36. К проектированию наилучших термоэлектриков.

Проблема, однако, в том, что при таком проектировании проводимость недостаточно велика, каковой она могла бы быть, если бы электрохимический потенциал E_{F0} располагался бы повыше (рис. 37), что характерно для металлов.

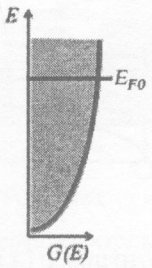


Рис. 37. К увеличению проводимости термоэлектрика.

Однако, металлы являются плохими термоэлектриками, потому что их коэффициенты Зеебека намного меньше k/q, поскольку их электрохимические потенциалы лежат ближе к середине зоны, так что выше E_{F0} находится приблизительно так же много состояний, как и ниже. По этой причине среди хороших термоэлектриков доминируют полупроводники, которые характеризуются достаточно высокими значениями силовых коэффициентов. Вместе с тем, силовой коэффициент определяет только числитель показателя качества термоэлектрика ZT (243). Как уже упоминалось, теплопроводность в знаменателе этого уравнения существенно зависит от транспорта тепла фононами, физика которого весьма отличается от электронного транспорта (§ 2.3.1).

2.3.3.7. О транспорте тепла фононами в концепции «снизу – вверх»

Покажем по-иному как легко модель упругого транспорта электронов (§ 1.3.12) может быть переписана применительно к фононам.

Равновесное распределение фононов описывается функцией Бозе

$$n(\omega)=\frac{1}{e^{\hbar\omega/kT}+1}. \tag{272}$$

Полученное выше выражение для электрического тока

$$I=\frac{q}{h}\int_{-\infty}^{+\infty}dE\left(\frac{M\lambda}{L+\lambda}\right)_{el}(f_1(E)-f_2(E)), \tag{273}$$

где фермиевские функции (232) записаны для разности потенциалов на контактах $V=(E_{F1}-E_{F2})/q$ и для разных температур, переписывается в фононный поток тепла

$$I_Q=\frac{1}{h}\int_{0}^{+\infty}d(\hbar\omega)\left(\frac{M\lambda}{L+\lambda}\right)_{ph}(\hbar\omega)(n_1(\omega)-n_2(\omega)) \tag{274}$$

просто путем замены $q\to\hbar\omega$, фермиевских функций (232) на бозевские функции

$$n_1(\omega)=\frac{1}{\exp(\hbar\omega/kT_1)-1},\quad n_2(\omega)=\frac{1}{\exp(\hbar\omega/kT_2)-1} \tag{275}$$

и нижнего предела интегрирования в (273) на нулевой.

Стандартная линеаризация выражения (274) дает

$$I_Q\approx G_K(T_1-T_2), \tag{276}$$

где теплопроводность, обязанная фононам,

$$G_K = \frac{k^2 T}{h}\int_0^\infty dx\left(\frac{M\lambda}{L+\lambda}\right)_{ph}\frac{x^2 e^x}{\left(e^x-1\right)^2},\quad x\equiv\frac{\hbar\omega}{kT}. \tag{277}$$

Обратим внимание на то, что как и модель упругого резистора для электронов не учитывает эффекты неупругого рассеяния мод проводимости, так и изложенная выше фононная модель не учитывает ангармоничность взаимодействия фононов, которая вынуждает фононы менять свою частоту. В отличие от достаточно хорошо изученных устройств на баллистическом транспорте электронов, и поныне немного известно об устройствах на баллистическом транспорте фононов [52 – 54].

2.3.3.8. Баллистический транспорт фононов

Баллистическая фононная проводимость записывается обычно в форме, напоминающей закон Стефана – Больцмана для фотонов.

Из уравнения (277) для баллистической теплопроводности ($L << \lambda$) имеем

$$G_K^{ball} = \frac{k^2 T}{h}\int_0^\infty dx M_{ph}\frac{x^2 e^x}{(e^x-1)^2}. \tag{278}$$

Как и для случая электронов, число мод для фононов

$$M_{ph} = \frac{3\pi A}{\lambda_{ph}^2}, \tag{279}$$

где учтены три разрешенных направления поляризации фононов, а

$$\lambda_{ph} = \frac{c_s}{\omega/2\pi}, \tag{280}$$

где c_s – скорость фононов, так что

$$M_{ph} = \frac{3\omega^2 A}{4\pi c_s^2} = \frac{3k^2T^2A}{4\pi\hbar^2 c_s^2}x^2, \tag{281}$$

и окончательно из (278) имеем

$$G_K^{ball} = \frac{3k^4T^3}{8\pi^2\hbar^3 c_s^2}\int_0^\infty dx\frac{x^4 e^x}{\left(e^x-1\right)^2} = \frac{\pi^2 k^4 T^3}{10\hbar^3 c_s^2}, \tag{282}$$

где интеграл слева равен $4\pi^4/15$. Теперь из (276) для баллистического потока тепла имеем

$$I_Q^{Ball} = \frac{\pi^2 k^4 T^3}{10\hbar^3 c_s^2}\Delta T, \tag{283}$$

где $\Delta T = T_1 - T_2$, а поскольку $T^3 \Delta T = \Delta(T^4/4)$, окончательно

$$I_Q^{Ball} = \frac{\pi^2 k^4}{40\hbar^3 c_s^2}(T_1^4 - T_2^4). \tag{284}$$

Аналогичный результат для фотонов известен как закон Стефана – Больцмана, отличающийся от баллистического потока фононов (284) коэффициентом 2/3, поскольку в случае фотонов число разрешенных поляризаций равно двум.

Теперь рассмотрим диффузионный транспорт фононов.

2.3.3.9. Фононная теплопроводность

Фононная теплопроводность определяется уравнением (277) и связана обычным образом со своим удельным значением

$$G_K = \frac{\kappa A}{L + \lambda}, \tag{285}$$

так что окончательно удельная фононная теплопроводность

$$\kappa = \frac{k^2 T}{h} \int_0^{+\infty} dx \left(\frac{M\lambda}{A} \right)_{ph} \frac{x^2 e^x}{(e^x - 1)^2} \tag{286}$$

подобна удельной электронной проводимости

$$\sigma = \frac{q^2}{h} \int_{-\infty}^{+\infty} dx \left(\frac{M\lambda}{A} \right)_{el} \frac{e^x}{(e^x + 1)^2}. \tag{287}$$

Функция уширения в электронных транспортных коэффициентах (§ 1.2.1)

$$F_T(x) \equiv \frac{e^x}{(e^x + 1)^2} \tag{288}$$

отличается от аналогичной функции в фононных транспортных коэффициентах

$$F_T^{ph}(x) \equiv \frac{3}{\pi^2} \frac{x^2 e^x}{(e^x - 1)^2}, \tag{289}$$

однако, схожесть их налицо (рис. 38).

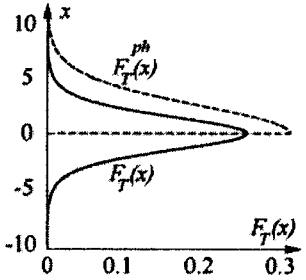


Рис. 38. Сравнение функций уширения в электронных и фононных транспортных коэффициентах. Множитель $3/\pi^2$ в (289) нужен для того, чтобы площади под кривыми были одинаковы и равны единице как в электронной функции $F_T(x)$.

Таким образом, по крайней мере качественно, электропроводность и теплопроводность можно описывать схожим образом. Так, если удельная электрическая проводимость для 3D-проводника дается выражением (глава 1):

$$\sigma = \frac{q^2}{h}\left(\frac{M\lambda}{A}\right), \tag{290}$$

где коэффициент $q^2/h \approx 38\,мкС$ есть обратное значение фундаментальной константы Клитцинга, то удельная теплопроводность дается выражением

$$\kappa_{ph} = \frac{\pi^2}{3}\frac{k^2T}{h}\left(\frac{M_{ph}\lambda_{ph}}{A}\right)_{ph}, \tag{291}$$

где коэффициент $(\pi^2/3)(k^2T/h) \approx 325\,nBm/K$ с множителем $\pi^2/3$, обратным тому значению, которое необходимо для нормировки функции уширения для фононов, а число мод на единицу поперечного сечения проводника M_{ph}/A и средняя величина свободного пробега λ_{ph} оцениваются в зазоре частот $\hbar\omega$, равном приблизительно нескольким kT. Поляризационный g-фактор $g = 3$ предполагается учтенным в значении числа мод M_{ph}.

Для термоэлектрика с коэффициентом Зеебека $S \sim 2(k/q)$ и доминирующей фононной составляющей удельной теплопроводности (291) для показателя качества термоэлектрика ZT можно написать

$$ZT \approx 4\frac{k^2T}{q^2}\frac{\sigma}{\kappa+\kappa_{ph}} \approx 4\frac{k^2T}{q^2}\frac{\sigma}{\kappa_{ph}} \tag{292}$$

или иначе, используя (290) и (291),

$$ZT \approx \frac{M\lambda/A}{(M\lambda/A)_{ph}}, \qquad (293)$$

где опущен множитель $12/\pi^2 \sim 1$. Это полезное выражение свидетельствует о том, что как только термоэлектрик оптимизирован по коэффициенту Зеебека S соответствующее ему значение ZT приближенно показывает отношение $M\lambda/A$ для электронов и фононов.

В стремлении повысить коэффициент Зеебека S мы тем самым понижаем удельное значение числа мод M/A для электронов. С другой стороны, отношение M_{ph}/A для фононов часто намного больше, а именно, $\sim 1 нм^{-2}$, так что их отношение в уравнении (293) всего ~ 0.1 или даже меньше. Однако, у электронов средняя длина свободного пробега больше, благодаря чему в наилучших термоэлектриках $ZT \sim 1$. В настоящее время представляется, что наиболее многообещающий путь улучшить значение ZT заключается в подавлении средней длины пробега для фононов, не затрагивая при этом поведение электронов (т. наз. PGEC/Phonon-Glass-Electron Crystals [55, 56]).

Мы ставили перед собой задачу в этой главе показать как пользоваться уравнением проводимости Ландауэра (109) при наличии перепада температур на концах проводника. Основные результаты можно сформулировать следующим образом:

Концепции, используемые для описания электронного транспорта, с успехом могут быть перенесены на фононный транспорт. И в том и в другом случае подход Ландауэра – Датты – Лундстрома, позволяет количественно описать транспортные процессы в проводниках любой размерности и при любых режимах транспорта – от диффузионного до баллистического.

Решеточная теплопроводность и электронная проводимость описываются схожими по форме уравнениями. Оба свойства всегда пропорциональны соответствующему кванту проводимости, умноженному на интеграл от произведения трех величин – коэффициента прохождения, числа мод и окна проводимости.

Имеют место два существенных отличия в физике транспорта электронов и фононов. Первое отличие касается ширины зон. В случае электронов ширина зон намного превышает kT и при комнатной температуре заселены лишь уровни вблизи дна зоны проводимости. Ширина фононных дисперсий всего лишь порядка kT и при комнатной температуре все акустические моды в пределах всей зоны Бриллюэна практически заселены. В результате дебаевская модель для акустических мод далеко не столь успешно работает по сравнению с широкой применимостью параболической дисперсии для электронов.

Второе существенное различие связано с тем, что в случае электронов заселенность мод проводимости, контролируемая фермиевским окном проводимости, зависит не только от температуры, как это имеет место в случае фононов, но и от положения уровня Ферми. Именно по этой причине

электронная проводимость различных материалов может изменяться в очень широких пределах, тогда как теплопроводность изменяется в очень узких пределах всего на несколько порядков [57 – 74].

Литература

1. Ю. О. Кругляк, Н. Ю. Кругляк, М. В. Стріха, Уроки наноелектроніки: Виникнення струму, формулювання закону Ома і моди провідності в концепції «знизу – вгору», *Сенсор. електр. мікросист. техн.*, т. 9, № 4, 5 – 29 (2012).
2. Ю. О. Кругляк, М. В. Стріха, Узагальнена модель електронного транспорту в мікро- і наноелектроніці, *Сенсор. електр. мікросист. техн.*, т. 12, № 3, 4 – 27 (2015).
3. A. F. Ioffe. *Semiconductor Thermoelements and Thermoelectric Cooling.* London: Infosearch, 1957.
4. Л. И. Анатычук, *Термоэлементы и термоэлектрические устройства.* Киев: Наукова думка, 1979.
5. Л. И. Анатычук, В. А. Семенюк, *Оптимальное управление свойствами термоэлектрических материалов и приборов.* Черновцы: Прут, 1992.
6. Л. И. Анатычук, Л. П. Булат, *Полупроводники в экстремальных температурных условиях.* Ленинград: Наука, 2001.
7. Л. И. Анатычук, *Термоэлектричество. Т. 2. Термоэлектрические преобразователи энергии.* Киев – Черновцы: Букрек, 2003.
8. Л. И. Анатычук, *Термоэлектричество. Т. 1. Физика термоэлектричества.* Киев – Черновцы: Букрек, 2009.
9. Н. Ашкрофт, Н. Мермин, *Физика твердого тела.* Москва: Мир, 1979.
10. G. D. Mahan, M. Bartkowiak. *Appl. Phys. Lett.*, v. 74, n 7, 953 – 954 (1999).
11. A. C. Smith, J. Janak, R. Adler. *Electronic Conduction in Solids.* New York: McGraw-Hill, 1965.
12. L. Onsager, *Phys. Rev.*, v. 37, n 4, 405 – 426 (1931).
13. Институт термоэлектричества НАНУ/МОН Украины: www.inst.cv.ua.
14. A. Majumdar. *Science*, v. 303, 778 – 779 (2004).
15. M. Dresselhaus, G. Chen, M. Tang, R. Yang, H. Lee, D. Wang, Z. Ren, J.-P. Fleureal, P. Gogna. *Adv. Materials*, v. 19, n 8, 1043 – 1053 (2007).
16. A. J. Minnich, M. S. Dresselhaus, Z. F. Ren, G. Chen. *Energy and Environmental Science*, v. 2, 466 – 479 (2009).
17. M. Hode. *IEEE Trans. Compon. Packag. Techn.*, v. 28, 218 – 229 (2005).
18. M. Hode. *IEEE Trans. Compon. Packag. Techn.*, v. 30, 50 – 58 (2007).
19. M. Hode. *IEEE Trans. Compon. Packag. Techn.*, v. 33, 307 – 318 (2010).
20. M. Lundstrom, C. Jeong. *Near-Equilibrium Transport: Fundamentals and Applications.* Hackensack, New Jersey: World Scientific Publishing Company, 2013; www.nanohub.org/resources/11763.
21. R. Kim, M. S. Lundstrom. Notes on Fermi – Dirac Integrals, *arXiv*:0811.0116; www.nanohub.org/resources/5475.

22. A. Sommerfeld. An electronic theory of the metals based on Fermi's statistics. *Z. Phys.*, v. 47, n 1, 1 (1928).
23. T. N. Geballe, G. W. Hull. Seebeck Effect in Germanium. *Phys. Rev.*, v. 94, 1134 (1954).
24. R. F. Pierret. *Semiconductor Device Fundamentals*. Reading, MA: Addison–Wesley, 1996.
25. Ra Seong Kim. *Physics and Simulation of Nanoscale Electronic and Thermoelectric Devices*. West Lafayette: Purdue University, 2011.
26. Дж. Займан. Электроны и фононы. Теория явлений переноса в твердых телах. Москва: ИИЛ, 1962.
27. Дж. Займан. Принципы теории твердого тела. Москва: Высшая школа, 1974.
28. Ч. Киттель. Введение в физику твердого тела. Москва: Наука, 1978.
29. M. Mohr, J. Maultzsch, E. Dobardžić, S. Reich, I. Milošević, M. Damnjanović, A. Bosak, M. Krisch, C. Thomsen. Phonon dispersion of graphite by inelastic x-ray scattering, *Phys. Rev. B*, v. 76, n 3, 035439/7 (2007); DOI: 10.1103/PhysRevB.76.035439.
30. А.В. Елецкий, И.М. Искандарова, А.А. Книжник, Д.Н. Красиков, Графен: методы получения и теплофизические свойства, *УФН*, т. 181, 227 (2011); DOI: 10.3367/UFNr.0181.201103a.0233.
31. M. I. Katsnelson. *Graphene: Carbon in Two Dimensions*. New York: Cambridge University Press, 2012.
32. K. Schwab, E. A. Henriksen, J. M. Worlock, M. L. Roukes. Measurement of the quantum of thermal conductance, *Nature*, v. 404, 974 (2000); DOI:10.1038/35010065.
33. C. Jeong, R. Kim, M. Luisier, S. Datta, M. Lundstrom, On Landauer vs. Boltzmann and Full Band vs. Effective Mass Evaluation of Thermoelectric Transport Coefficients, *J. Appl. Phys.*, v. 107, 023707 (2010).
34. M. Lundstrom. *Fundamentals of Carrier Transport*. Cambridge: Cambridge University Press, 2012.
35. C. Jeong, S. Datta, M. Lundstrom. Full Dispersion vs. Debye Model Evaluation of Lattice Thermal Conductivity with a Landauer Approach, *J. Appl. Phys.*, v. 109, 073718/8 (2011).
36. Ю. А. Кругляк, Н. Е. Кругляк, Уроки наноэлектроники. 3. Электронная проводимость и моды проводимости в концепции «снизу – вверх», *Физич. образов. в вузах*, т. 19, № 3, 99 – 110 (2013).
37. T. S. Fisher. *Thermal Energy at the Nanoscale*. Hackensack, New Jersey: World Scientific Publishing Company, 2013; www.nanohub.org/courses/2.
38. Ю. А. Кругляк, Н. Е. Кругляк, Уроки наноэлектроники. 2. Модель упругого резистора и новая формулировка закона Ома в концепции «снизу – вверх», *Физич. образов. в вузах*, т. 19, № 2. 161 – 173 (2013).
39. J. Callaway. Model for lattice thermal conductivity at low temperatures, *Phys. Rev.*, v. 113, n 4, 1046 (1959).

40. M. G. Holland. Analysis of lattice thermal conductivity, *Phys. Rev.*, v. 132, n 6, 2461 (1963).
41. C. Jeong, S. Datta, M. Lundstrom. Thermal conductivity of bulk and thin-film silicon: a Landauer approach, *J. Appl. Phys.*, v. 111, 093708 (2012).
42. Gang Chen. *Nanoscale Energy Transport and Conversion: A Parallel Treatment of Electrons, Molecules, Phonons, and Photons*. New York: Oxford University Press, 2005.
43. C. J. Glassbrenner, G. A. Slack. Thermal Conductivity of Silicon and Germanium from 3°K to the Melting Point, *Phys. Rev. A*, v. 134, n 4, 1058 (1964); DOI: 10.1103/PhysRev.134.A1058.
44. J. B. Pendry. Quantum limits to the flow of information and entropy, *J. Phys. A.*, 16, 2161 – 2171 (1983).
45. D. E. Angelescu, M. C. Cross, M. L. Roukes. Heat transport in mesoscopic systems, *Superlatt. Microstruct.*, v. 23, 673 (1998).
46. L. G. C. Rego, G. Kirczenow. Quantized Thermal Conductance of Dielectric Quantum Wires, *Phys. Rev. Lett.*, v. 81, 232 (1998).
47. M. P. Blencowe. Quantum energy flow in mesoscopic dielectric structures, *Phys. Rev. B.*, v. 59, 4992 (1999).
48. L. G. C. Rego, G. Kirczenow. Fractional exclusion statistics and the universal quantum of thermal conductance: A unifying approach, *Phys. Rev. B.*, v. 59, 13080 (1999).
49. I. V. Krive, E. R. Mucciolo. Transport properties of quasiparticles with fractional exclusion statistics, *Phys. Rev. B.*, v. 60, 1429 (1999).
50. C. M. Caves, P. D. Drummond. Quantum limits on bosonic communication rates, *Rev. Mod. Phys.*, v. 66, 481 (1994).
51. K. Baheti, J. A. Malen, P. Doak, P. Reddy, Sung-Yeon Jang, T. D. Tilley, A. Majumdar, R. A. Segalman. Probing the Chemistry of Molecular Heterojunctions Using Thermoelectricity, *Nano Letters*, v. 8, n 2, 715 – 719 (2008).
52. G. Chen. Thermal conductivity and ballistic-phonon transport in the cross-plane direction of superlattices, *Phys. Rev. B*, v. 57, 14958 – 14973 (1998).
53. H.-Y. Chiu, V. V. Deshpande, H. W. Ch. Postma, C. N. Lau, C. Mikó, L. Forró, M. Bockrath. Ballistic Phonon Thermal Transport in Multi-Walled Carbon Nanotubes, *Phys. Rev. Letts.*, v. 95, 226101 (2005).
54. N. Zuckerman, J. R. Lukes. Atomistic Visualization of ballistic phonon transport, *Proc. ASME-JSME Thermal Engineering Summer Heat Transfer Conf., Vancouver, Canada*, N 32674, 825 – 833 (2007).
55. G. S. Nolas, D. T. Morelli, T. M. Tritt. SKUTTERUDITES: A Phonon-Glass-Electron Crystal Approach to Advanced Thermoelectric Energy Conversion Applications, *Ann. Rev. Mater. Sci.*, v. 29, 89 – 116 (1999).
56. G. Min, D. M. Rowe. A serious limitation to the phonon glass electron crystal (PGEC) approach to improved thermoelectric materials, *J. Mater. Sci. Lett.*, v. 18, n 16, 1305 – 1306 (1999).

57. Ю. О. Кругляк, Н. Ю.Кругляк, М. В.Стріха, Уроки наноелектроніки. Термоелектричні явища в концепції «знизу – вгору», *Сенсор. електр. мікросист. техн.*, т. 10, № 1, 6 – 21 (2013).
58. Ю. А. Кругляк. Уроки наноэлектроники. 4. Термоэлектрические явления в концепции «снизу – вверх», *Физич. образов. в вузах*, т. 19, № 4, 70 – 85 (2013).
59. Ю. А. Кругляк. Уроки наноэлектроники. 5. Перенос тепла фононами в концепции «снизу – вверх» // *Физич. образов. в вузах*, т. 20, № 1, 39 – 43 (2014).
60. Ю. А. Кругляк, П. А. Кондратенко, Ю. М. Лопаткин. Наноэлектроника: Термоэлектрические явления в концепции «снизу – вверх», *Ж. нано- электр. физ.*, т. 6, № 1, 01013(9) (2014).
61. Yu. A. Kruglyak. Heat Transfer by Phonons in Landauer-Datta-Lundstrom Approach, *Proc. Intern. Conf: Nanomaterials: Applications and Properties*, v. 3, n 2, 02NAESF02 (2014).
62. Ю. А. Кругляк. Термоэлектрические явления и устройства в концепции Ландауэра – Датты – Лундстрома, *Наносистемы, наноматериалы, нанотехнологии*, т. 12, № 3, 601 – 621 (2014).
63. Ю. А. Кругляк. Термоэлектрические коэффициенты в транспортной модели Ландауэра – Датты – Лундстрома, *Наносистемы, наноматериалы, нанотехнологии*, т. 12, № 4, 751 – 776 (2014).
64. Ю. А. Кругляк. Електричний струм, термострум і тепловий потік в нано- і мікроелектроніці: модель переносу, *Термоелектрика*, № 6, 7 – 28 (2014).
65. Ю. А. Кругляк. Електричний струм, термострум і тепловий потік в нано- і мікроелектроніці: спеціальні питання, *Термоелектрика*, № 6, 29 – 47 (2014).
66. Yu. A. Kruglyak. Electric current, thermocurrent, and heat flux in nano- and microelectronics: transport model, *J. Thermoelectricity*, n 6, 7 – 25 (1914).
67. Yu. A. Kruglyak. Electric current, thermocurrent, and heat flux in nano- and microelectronics: selected topics, *J. Thermoelectricity*, n 6, 26 – 41 (1914).
68. Ю. А. Кругляк. Электрический ток, термоток и тепловой поток в нано- и микроэлектронике: модель переноса, *Термоэлектричество*, № 6, 7 – 28 (2014).
69. Ю. А. Кругляк. Электрический ток, термоток и тепловой поток в нано- и микроэлектронике: специальные вопросы, *Термоэлектричество*, № 6, 29 – 47 (2014).
70. Ю. А. Кругляк. Термоэлектрические явления и устройства в концепции Ландауэра – Датты – Лундстрома, *ScienceRise*, т. 1, № 2(6), 69 – 77 (2015).
71. Ю. А. Кругляк. Термоэлектрические коэффициенты в транспортной модели Ландауэра – Датты – Лундстрома, *ScienceRise*, т. 1, № 2(6),78 – 89 (2015).
72. Ю. А. Кругляк. Перенос тепла фононами в транспортной модели Ландауэра – Датты – Лундстрома, *ScienceRise*, т. 2, № 2(7), 81 – 93 (2015).

73. Yu. A. Kruglyak, M. V. Strikha. Heat flow by phonons in Landauer – Datta – Lundstrom transport model for micro- and nanoelectronics, *Proc. 2015 IEEE 35th Intern. Conf. Electronics Nanotech. (ELNANO), April 21 – 24, 2015, Kyiv, Ukraine*, pp. 75 – 80; DOI: 10.1109/ELNANO.2015.7146838.
74. Ю. А. Кругляк. Транспорт тепла фононами в модели Ландауэра – Датты – Лундстрома, *Наносистемы, наноматериалы, нанотехнологии*, т. 13, № 3, 549 – 576 (2015).

Глава 3. Графен

Обобщенная модель транспорта электронов и тепла в режиме линейного отклика, изложенная в главах 1 и 2, сводится к паре транспортных уравнений [1]:

$$J_x = \sigma \frac{d\left(E_F / q\right)}{dx} - S\sigma \frac{dT}{dx}, \ [A/м^2] \tag{1}$$

$$J_{Qx} = \pi\sigma \frac{d(E_F / q)}{dx} - \kappa_0 \frac{dT}{dx} \ [Вт/м^2] \tag{2}$$

или в обращенной форме

$$\frac{d(E_F / q)}{dx} = \rho J_x + S\frac{dT}{dx}, \tag{3}$$

$$J_{Qx} = \pi J_x - \kappa \frac{dT}{dx} \tag{4}$$

с транспортными коэффициентами для 2D проводников

$$\sigma = 1/\rho = \int \sigma'(E)dE, \tag{5}$$

$$\sigma'(E) = \frac{2q^2}{h}\frac{M(E)}{W}\lambda(E)\left(-\frac{\partial f_0}{\partial E}\right), \tag{6}$$

$$S = -\frac{k}{q}\frac{1}{\sigma}\int\left(\frac{E-E_F}{kT}\right)\sigma'(E)dE = \frac{s_T}{\sigma} = \frac{\pi}{T}, \tag{7}$$

$$\kappa_0 = T\left(\frac{k}{q}\right)^2 \int\left(\frac{E-E_F}{kT}\right)^2 \sigma'(E)dE = \kappa + \pi S\sigma. \tag{8}$$

Уравнения (1) – (8) справедливы для диффузионных резисторов любой размерности и любого масштаба, любого типа проводимости, включая биполярную проводимость. Уравнение (6) выписано для 2D проводника; это видно по множителю $M(E)/W$, где W – ширина 2D проводника. При рассмотрении 3D проводников этот множитель нужно заменить на $M(E)/A$, где A-- площадь поперечного сечения 3D проводника, а для 1D проводника заменить просто на число мод проводимости $M(E)$. При рассмотрении транспорта во всех режимах от диффузионного до баллистического среднюю длину свободного пробега $\lambda(E)$ нужно заменить на кажущуюся длину по уравнению $1/\lambda^{app} = 1/\lambda + 1/L$, где L – длина проводника. Фактически в этих уравнениях нужно задаться только двумя параметрами, $M(E)$ и $\lambda(E)$. Число

мод для проводников любой размерности с параболической дисперсией определены в [2], но сами уравнения (1) – (8) справедливы для любой зонной структуры, нужно лишь корректно задать число мод проводимости. Зависимость $\lambda(E)$ нужно определить в соответствии с характером рассеяния в проводнике; часто используется степенной закон рассеяния [1].

В настоящем главе рассматривается применение обобщенной транспортной модели проводимости к графену.

3.1. Элементарная теория зонной структуры графена

Пожалуй, сегодня нет другого объекта, который вызывал бы столь пристальное и широкое внимание физиков, химиков и технологов, как графен. Опубликованы подробные обзоры по физике графена [3 – 9]. В приближении эффективной массы поведение электрона в графене дается уравнением Дирака – Вейля для безмассовых нейтрино [10 – 12]. В такой экзотической электронной системе особый интерес вызывают ее транспортные свойства [13 – 17]. После демонстрации получения образцов графена микромеханическим методом расщепления графита [18] качество образцов графена оказалось настолько высоким, что удалось реализовать баллистический транспорт на этих образцах [18, 19] и наблюдать квантовый эффект Холла, что в свою очередь явилось фактическим подтверждением справедливости «нейтринной» трактовки электронных состояний графена [20, 21]. Это обстоятельство послужило «спусковым крючком» взрывного интереса к всесторонним исследованиям графена вплоть до включения расчетов зонной структуры графена в университетские курсы, например, в Стенфордском университете [22] и в Университете Пердью [23 – 25].

Вначале мы дадим сжатое изложение простейшей π-модели графена в рамках теории сильной связи [26] и приведем необходимые сведения о его зонной структуре, что послужит нам основой для расчета транспортных и не только свойств графена. Дадим также сводку всех термоэлектрических свойств графена как в баллистическом режиме, так и в диффузионном.

Кристаллическая решетка графена представляет собой совокупность двух взаимнопроникающих решеток Бравэ A и B с элементарной ячейкой в виде правильного ромба (рис. 1).

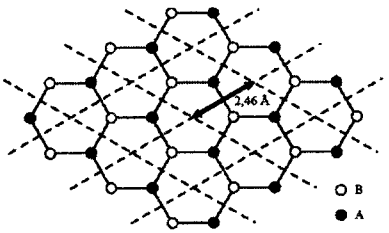


Рис. 1. Элементарная ячейка графена и вектор трансляции решеток Бравэ A и B.

Период этих решеток равен $a = a_0\sqrt{3} \approx 2.46$ Å, где длина связи СС $a_0 = 1.42$ Å. Зона Бриллюэна представляет собой правильный шестиугольник со стороной $4\pi / 3a$ (рис. 2).

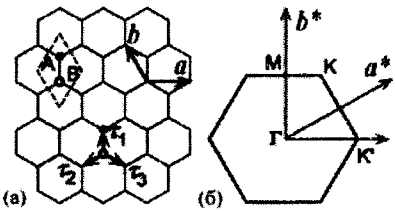


Рис. 2. Решетка графена (а) и его первая зона Бриллюэна (б).

Примитивные векторы трансляции (рис. 2) выбраны следующим образом: $\boldsymbol{a} = a(1,0)$ и $\boldsymbol{b} = a(-1/2, \sqrt{3}/2)$. Векторы, связывающие соседние атомы, равны $\boldsymbol{\tau}_1 = a(0, 1/\sqrt{3})$, $\boldsymbol{\tau}_2 = a(-1/2, -1/2\sqrt{3})$, $\boldsymbol{\tau}_3 = a(1/2, -1/2\sqrt{3})$. Вектора обратной решетки равны $\boldsymbol{a}^* = (2\pi/a)(1, 1/\sqrt{3})$ и $\boldsymbol{b}^* = (2\pi/a)(0, 2/\sqrt{3})$. Первая зона Бриллюэна имеет две узловые точки К и К′. Соответствующие волновые векторы даются выражениями $\boldsymbol{K} = (2\pi/a)(-2/3, 0)$ и $\boldsymbol{K}' = (2\pi/a)(2/3, 0)$. Далее

$$\begin{aligned} &\exp\left(i\boldsymbol{K}\cdot\boldsymbol{\tau}_1\right) = \omega, \quad \exp\left(i\boldsymbol{K}\cdot\boldsymbol{\tau}_2\right) = \omega^{-1}, \quad \exp\left(i\boldsymbol{K}\cdot\boldsymbol{\tau}_3\right) = 1, \\ &\exp\left(i\boldsymbol{K}'\cdot\boldsymbol{\tau}_1\right) = 1, \quad \exp\left(i\boldsymbol{K}'\cdot\boldsymbol{\tau}_2\right) = \omega^{-1}, \quad \exp\left(i\boldsymbol{K}'\cdot\boldsymbol{\tau}_3\right) = \omega, \end{aligned} \tag{9}$$

где $\omega = \exp(2\pi i / 3)$ и удовлетворяет условию $1 + \omega + \omega^{-1} = 0$.

Пусть $\phi(\boldsymbol{r})$ обозначает $2p_z$-АО атома углерода С. Периодический потенциал решетки создается двумя смещенными друг относительно друга подрешетками А и В. Волновая функция электрона в таком потенциале дается линейной комбинацией двух блоховских волн, построенных на этих подрешетках

$$\psi(\boldsymbol{r}) = \sum_{\boldsymbol{R}_A} \psi_A\left(\boldsymbol{R}_A\right)\varphi\left(\boldsymbol{r} - \boldsymbol{R}_A\right) + \sum_{\boldsymbol{R}_B} \psi_B\left(\boldsymbol{R}_B\right)\varphi\left(\boldsymbol{r} - \boldsymbol{R}_B\right), \tag{10}$$

где амплитуды $\psi_A(\boldsymbol{R}_A)$ и $\psi_B(\boldsymbol{R}_B)$ берутся в точках $\boldsymbol{R}_A = n_a\boldsymbol{a} + n_b\boldsymbol{b} + \boldsymbol{\tau}_1$ и $\boldsymbol{R}_B = n_a\boldsymbol{a} + n_b\boldsymbol{b}$ при целочисленных значениях n_a и n_b. Пренебрежем интегралом перекрывания двух соседних АО $(S[2p_z(A), 2p_z(B)] = 0)$ и учтем резонансный интеграл только между соседними атомами ($\gamma_0 \approx 3$ эВ), что позволит получить все решения в аналитическом виде. Тогда

$$E\psi_A\left(\boldsymbol{R}_A\right) = -\gamma_0 \sum_{l=1}^{3} \psi_B\left(\boldsymbol{R}_A - \boldsymbol{\tau}_l\right), \quad E\psi_B\left(\boldsymbol{R}_B\right) = -\gamma_0 \sum_{l=1}^{3} \psi_A\left(\boldsymbol{R}_B + \boldsymbol{\tau}_l\right), \tag{11}$$

где за ноль отсчета энергии принята энергия $2p_z$-АО атома С.

Положив $\psi_A(\boldsymbol{R}_A) = f_A(\boldsymbol{k})\exp(i\boldsymbol{k}\cdot\boldsymbol{R}_A)$ и $\psi_B(\boldsymbol{R}_B) = f_B(\boldsymbol{k})\exp(i\boldsymbol{k}\cdot\boldsymbol{R}_B)$, получим

$$\begin{pmatrix} 0 & h_{AB}(\boldsymbol{k}) \\ h_{AB}(\boldsymbol{k})^* & 0 \end{pmatrix}\begin{pmatrix} f_A(\boldsymbol{k}) \\ f_B(\boldsymbol{k}) \end{pmatrix} = E\begin{pmatrix} f_A(\boldsymbol{k}) \\ f_B(\boldsymbol{k}) \end{pmatrix},\quad h_{AB}(\boldsymbol{k}) = -\gamma_0\sum_{l=1}^{3}\exp(-i\boldsymbol{k}\cdot\boldsymbol{\tau}_l). \qquad (12)$$

Тогда для энергии зон имеем

$$E_{\pm}(\boldsymbol{k}) = \pm\gamma_0\left|\sum_{l=1}^{3}\exp(-i\boldsymbol{k}\cdot\boldsymbol{\tau}_l)\right|. \qquad (13)$$

В соответствии с уравнением (10) имеем $E_{\pm}(\boldsymbol{K}) = E_{\pm}(\boldsymbol{K}') = 0$, т.е. в так называемых дираковских точках K и K' отсутствует щель между валентной зоной π и зоной проводимости π^* (рис. 3 и 4).

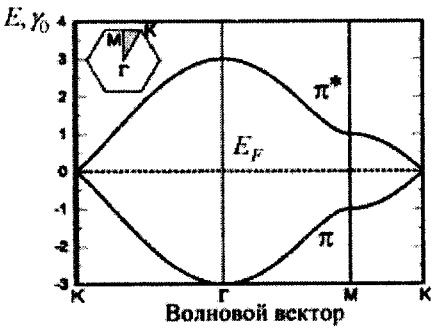


Рис. 3. Симметричная зонная структура графена в приближении Хюккеля: (S = 0; $\gamma_0 \neq 0$).

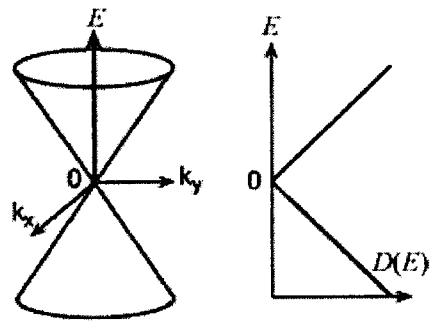


Рис. 4. Дисперсия энергии и плотность состояний $D(E)$ в окрестности точек K и K'.

Посмотрим на поведение энергии в окрестности точки K. Перепишем $\boldsymbol{k}$ как $\boldsymbol{K}+\boldsymbol{k}$ и разложим энергию $E_{\pm}(\boldsymbol{k})$ по степеням $|\boldsymbol{k}|a$. В наинизшем, первом порядке получим

$$-\gamma_0\sum_{l=1}^{3}\exp\left[-i\left(\boldsymbol{K}+\boldsymbol{k}\right)\cdot\boldsymbol{\tau}_l\right]=-\omega^{-1}\gamma\left(k_x-ik_y\right) \tag{14}$$

$$\gamma=\frac{\sqrt{3}}{2}a\gamma_0. \tag{15}$$

Поэтому когда мы переопределяем $f_A(\boldsymbol{K}+\boldsymbol{k})=\tilde{f}_A(\boldsymbol{k})$ и $f_B(\boldsymbol{K}+\boldsymbol{k})=-\omega\tilde{f}_B(\boldsymbol{k})$, имеем

$$\gamma\begin{pmatrix}0 & k_x-ik_y\\ k_x+ik_y & 0\end{pmatrix}\begin{pmatrix}\tilde{f}_A(\boldsymbol{k})\\ \tilde{f}_B(\boldsymbol{k})\end{pmatrix}=E\begin{pmatrix}\tilde{f}_A(\boldsymbol{k})\\ \tilde{f}_B(\boldsymbol{k})\end{pmatrix}. \tag{16}$$

Используя матрицы Паули

$$\sigma_x=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\ \sigma_y=\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix},\ \sigma_z=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \tag{17}$$

предыдущее уравнение (16) можно переписать в виде

$$\gamma\left(\boldsymbol{\sigma}\cdot\boldsymbol{k}\right)\tilde{\boldsymbol{f}}\left(\boldsymbol{k}\right)=E\,\tilde{\boldsymbol{f}}\left(\boldsymbol{k}\right),\ \ \tilde{\boldsymbol{f}}\left(\boldsymbol{k}\right)=\begin{pmatrix}\tilde{f}_A(\boldsymbol{k})\\ \tilde{f}_B(\boldsymbol{k})\end{pmatrix}, \tag{18}$$

где $\boldsymbol{\sigma}=(\sigma_x,\sigma_y)$. Аналогичным образом получаем уравнение для точки K'. Фактически, достаточно лишь $\boldsymbol{\sigma}$ заменить на $\boldsymbol{\sigma}^*$.

Таким образом, в окрестностях точек K и K' дисперсия энергии дается простым выражением

$$E_s\left(\boldsymbol{k}\right)=s\gamma\left|\boldsymbol{k}\right|,\ \ s=\pm1, \tag{19}$$

а плотность состояний равна

$$D\left(E\right)=\frac{g_v g_s}{L^2}\sum_{s,k}\delta\left(E-s\gamma\left|\boldsymbol{k}\right|\right)=\frac{g_v g_s\left|E\right|}{2\pi\gamma^2}, \tag{20}$$

где L_xL_y- площадь графена, вырождение по спину $g_s=2$, а наличие двух долин K и K' приводит к пространственному (долинному) вырождению $g_v=2$. Полученные результаты иллюстрируются рис. 4. Плотность состояний исчезает

в точке $E=0$. Поэтому графен часто рассматривают как двумерный полупроводник с нулевой запрещенной зоной или же как полуметал.

В рассматриваемом приближении эффективной массы, иначе, схемы $\boldsymbol{k}\cdot\boldsymbol{p}$, движение электрона вблизи точек K и K' описывается уравнением Шредингера, в котором вектор $\boldsymbol{k}$ заменяется оператором $\hat{\boldsymbol{k}}=-i\bar{\nabla}$. Имеем

$$\gamma\left(\boldsymbol{\sigma}\cdot\hat{\boldsymbol{k}}\right)\boldsymbol{F}^{K}(\boldsymbol{r})=E\,\boldsymbol{F}^{K}(\boldsymbol{r}),\ \ \gamma\left(\boldsymbol{\sigma}^{*}\cdot\hat{\boldsymbol{k}}\right)\boldsymbol{F}^{K'}(\boldsymbol{r})=E\,\boldsymbol{F}^{K'}(\boldsymbol{r}),\tag{21}$$

где $\boldsymbol{F}^{K}(\boldsymbol{r})$ и $\boldsymbol{F}^{K'}(\boldsymbol{r})$ являются двухкомпонентной волновой функцией

$$\boldsymbol{F}^{K}(\boldsymbol{r})=\begin{pmatrix}F_{A}^{K}(\boldsymbol{r})\\F_{B}^{K}(\boldsymbol{r})\end{pmatrix},\ \ \boldsymbol{F}^{K'}(\boldsymbol{r})=\begin{pmatrix}F_{A}^{K'}(\boldsymbol{r})\\F_{B}^{K'}(\boldsymbol{r})\end{pmatrix}.\tag{22}$$

Уравнения (21) – это ничто иное как уравнения Дирака – Вейля для нейтрино, исключая разве что изотропную скорость

$$v_{F}=\frac{1}{\hbar}\frac{\partial E}{\partial k}=\frac{\gamma}{\hbar},\tag{23}$$

меньшую скорости света приблизительно в 300 раз. Столь высокая скорость электронов в графене привлекла внимание специалистов по наноэлектронике.

Итак, вблизи точки Дирака

$$E(k)=\pm\hbar v_{F}k=\pm\hbar v_{F}\sqrt{k_{x}^{2}+k_{y}^{2}}\,.\tag{24}$$

Графен с $E_{F}>0$ назовем n-проводником, а с $E_{F}<0$ – p-проводником. Щели между зоной проводимости и валентной зоной нет. Обе зоны всегда спарены (рис. 5).

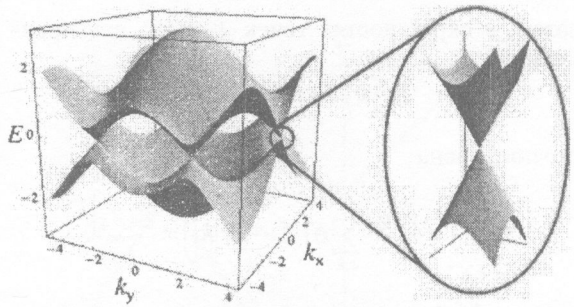


Рис. 5. Зонная структура графена. Увеличенным масштабом показан линейный спектр носителей тока в точке Дирака.

Электроны в графене описываются двухкомпонентной волновой функцией

$$\begin{pmatrix}\psi_A\\ \psi_B\end{pmatrix}=\frac{1}{\sqrt{2}}\frac{1}{\sqrt{L_xL_y}}\begin{pmatrix}1\\ se^{i\theta}\end{pmatrix}e^{i\left(k_xx+k_yy\right)}, \tag{25}$$

где $\theta = arctg(k_y / k_x)$, что влечет за собой важные следствия; например, поскольку электроны вдоль $+k_x$ и $-k_x$ описываются ортогональными волновыми функциями, то вероятность обратного рассеяния на 180° в графене нулевая.

3.1.1. Обратная решетка и зона Бриллюэна графена

Построение обратной решетки графена и его зоны Бриллюэна почти очевидно. Примитивные векторы трансляции можно выбрать любым приемлемым способом. В отличие от выбора на рис. 2, пусть сейчас векторы трансляции выберем иначе (рис. 6).

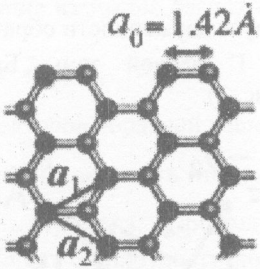


Рис. 6. Выбор примитивных векторов трансляции в решетке графена.

Эти базисные векторы, очевидно, таковы:

$$\boldsymbol{a}_1=\frac{3}{2}a_0\hat{x}+\frac{\sqrt{3}}{2}a_0\hat{y},$$
$$\boldsymbol{a}_2=\frac{3}{2}a_0\hat{x}-\frac{\sqrt{3}}{2}a_0\hat{y},$$
$$\boldsymbol{a}_3=c\hat{z},$$

где вектор $\boldsymbol{a}_3$ играет вспомогательную роль: периодичность перпендикулярно плоскости графена отсутствует, величина этого вектора, как будет видно ниже, дается большим числом.

Произвольные векторы прямой решетки $\boldsymbol{R}=n_1\boldsymbol{a}_1+n_2\boldsymbol{a}_2+n_3\boldsymbol{a}_3$ и обратной $\boldsymbol{G}=k_1\boldsymbol{b}_1+k_2\boldsymbol{b}_2+k_3\boldsymbol{b}_3$ связаны соотношением:

$$\boldsymbol{R}\cdot\boldsymbol{G}=2\pi\times Int,$$

где *Int* есть произвольное целое число, а трансляционные векторы обратной решетки

$$\boldsymbol{b}_i = 2\pi \frac{\boldsymbol{a}_j \times \boldsymbol{a}_k}{\boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)},$$

где знаменатель есть объем элементарной ячейки. Вычисляем векторы обратной решетки:

$$\boldsymbol{b}_1 = 2\pi \frac{\boldsymbol{a}_2 \times \boldsymbol{a}_3}{\boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)} = \frac{2\pi}{a_0}\left(\frac{1}{3}\hat{x} + \frac{1}{\sqrt{3}}\hat{y}\right),$$

$$\boldsymbol{b}_2 = 2\pi \frac{\boldsymbol{a}_3 \times \boldsymbol{a}_1}{\boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)} = \frac{2\pi}{a_0}\left(\frac{1}{3}\hat{x} - \frac{1}{\sqrt{3}}\hat{y}\right),$$

$$\boldsymbol{b}_3 = 2\pi \frac{\boldsymbol{a}_1 \times \boldsymbol{a}_2}{\boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)} = \frac{2\pi}{c}\hat{z},$$

где $\boldsymbol{b}_3 \big|_{c\to\infty} \to 0$, что соответствует двумерности обратной решетки.

Относительно центра Г первой зоны Бриллюэна строим векторы $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_1 + \boldsymbol{b}_2$ и им обратные (рис. 7).

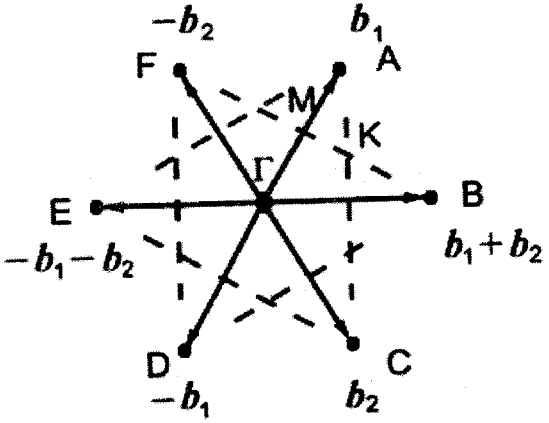


Рис. 7. Построение первой зоны Бриллюэна.

Концы векторов A, C, E и B, D, F образуют два симметрично пересекающихся треугольника, образующих первую зону Бриллюэна в виде правильного шестиугольника с особыми точками M и K. Из геометрических построений для стороны зоны Бриллюэна получаем $2MK = 4\pi / 3a_0\sqrt{3}$.

Сопоставление прямой и обратной решеток показано на рис. 8.

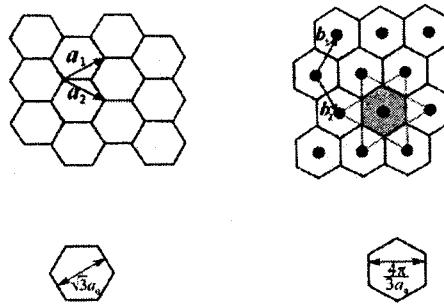


Рис. 8. Прямая (слева) и обратная (справа) решетки графена; первая зона Бриллюэна затенена.

Обращает на себя внимание взаимное расположение векторов трансляции обеих решеток:

$$\boldsymbol{b}_1 \perp \boldsymbol{a}_2, \quad \boldsymbol{b}_2 \perp \boldsymbol{a}_1,$$

что следует из определения векторов обратной решетки.

3.2. Плотность состояний и плотность носителей тока

Вычислим проводимость графена. Проводимость зависит от E_F, а положение уровня Ферми зависит от плотности носителей тока, определяемой плотностью состояний $D(E)$.

Опуская подробности вычисления плотности состояний [27, 28], для 2D проводника число состояний между k и $k+dk$

$$N(k)dk = g_s g_v \frac{2\pi k dk}{(2\pi / L_x)(2\pi / L_y)}, \tag{26}$$

где $L_x L_y = A$ есть площадь графена. В знаменателе стоит площадь k-пространства, занятая k-состояниями. Для дисперсии (24)

$$N(k)dk = g_v A \frac{EdE}{\pi(\hbar v_F)^2}. \tag{27}$$

Определяя $D(E)dE$ как плотность состояний на единицу поверхности графена в интервале энергий между E и $E+dE$ и учитывая, что энергия может быть как положительной, так и отрицательной, для плотности состояний графена окончательно получаем:

$$D(E) = \frac{2|E|}{\pi \hbar^2 v_F^2}, \tag{28}$$

а для образца графена с площадью поверхности $W \cdot L$ это удельное выражение нужно умножить на $W \cdot L$.

Число электронов в зоне проводимости

$$n_s(E_F) = \int_0^\infty D(E) f_0(E) dE, \tag{29}$$

где индексом s пометим свойство, отнесенное к поверхности. Поскольку графен вырожден, то T = 0° K будет неплохим приближением даже при комнатной температуре. Уравнение (29) становится следующим:

$$n_S(E_F) = \int_0^{E_F} D(E) dE = \frac{2}{\pi \hbar^2 v_F^2} \int_0^{E_F} dE, \tag{30}$$

так что окончательно для плотности носителей тока имеем:

$$n_S(E_F) = \frac{E_F^2}{\pi \hbar^2 v_F^2}. \tag{31}$$

3.3. Число мод и проводимость

Для 2D проводника число мод (глава 1)

$$M(E) = W M_{2D}(E) = W \frac{h}{4} \left\langle v_x^+(E) \right\rangle D_{2D}(E), \tag{32}$$

где W – ширина проводника. Для графена

$$\left\langle v_x^+(E) \right\rangle = \frac{2}{\pi} v_F, \tag{33}$$

плотность состояний дается формулой (28), так что для числа мод окончательно имеем:

$$M(E) = W \frac{2|E|}{\pi \hbar v_F}. \tag{34}$$

В отличии от параболической дисперсии, для которой поведение числа мод и плотности состояний с энергией различное (глава 1), в графене оба эти свойства пропорциональны энергии:

$$D(E), M(E) \propto E. \tag{35}$$

Теперь можно сразу вычислить проводимость. Согласно (6) с учетом вырожденности (T = 0° K),

$$\sigma'(E) = \frac{2q^2}{h} \frac{M(E)}{W} \lambda(E) \left(-\frac{\partial f_0}{\partial E} \right) = \frac{2q^2}{h} \frac{M(E)}{W} \lambda(E) \delta(E - E_F), \tag{36}$$

так что окончательно поверхностная проводимость (5)

$$\sigma_S = \int \sigma'(E)dE = \frac{2q^2}{h}\left(\frac{2E_F}{\pi\hbar v_F}\right)\lambda(E_F). \qquad (37)$$

Для не зависящей от энергии средней длины свободного пробега проводимость графена пропорциональна энергии Ферми, а по (31) стало быть линейно зависит от $\sqrt{n_S}$.

3.4. Рассеяние в графене

Характер зависимости проводимости от энергии Ферми или от поверхностной плотности носителей тока диктуется зависимостью среднего пути свободного пробега от энергии $\lambda(E)$, который в свою очередь определяется временем релаксации импульса согласно [29]

$$\lambda(E) = \frac{\pi}{2} v_F \tau_p. \qquad (38)$$

Там же показано, что для короткодействующих потенциалов рассеяния и рассеяния на акустических фононах скорость рассеяния пропорциональна плотности состояний:

$$\frac{1}{\tau_p} \propto D(E) \propto E, \qquad (39)$$

так что в итоге

$$\tau_p \propto \lambda(E) \propto E^{-1} \qquad (40)$$

и не зависит от температуры. Согласно же (37), теперь нужно признать, что для короткодействующих потенциалов рассеяния и рассеяния на акустических фононах поверхностная проводимость σ_S должна быть константой, не зависящей от E_F или n_S. Поскольку

$$\sigma_S = q n_S \mu, \qquad (41)$$

то подвижность μ должна быть обратно пропорциональна n_S. Это необычная ситуация. Ведь обычно чем больше плотность носителей тока, тем больше проводимость. Можно заключить, что если доминируют короткодействующие потенциалы рассеяния или рассеяние на акустических фононах, то можно ожидать независимости проводимости от n_S [13] (рис. 9).

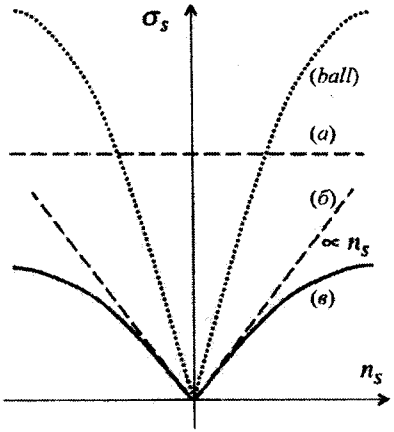


Рис. 9. Качественная картина зависимости проводимости от плотности носителей тока для графена: (а) для короткодействующих потенциалов рассеяния и рассеяния на акустических фононах зависимость отсутствует, (б) рассеяние на заряженных примесях дает линейную зависимость, (в) суммарная картина; показан также баллистический режим,

В случае рассеяния на экранированных и неэкранированных ионизированных примесях для $\lambda(E)$ можно ожидать линейную зависимость от энергии [15 – 17]

$$\lambda(E) = \lambda_{ii} E, \tag{42}$$

что после подстановки в (37) дает квадратичную зависимость от E_F. Плотность носителей тока (31) также квадратична по E_F. Таким образом, рассеяние на заряженных примесях ведет к проводимости, линейно зависящей от n_S и, согласно (41), к постоянному значению подвижности [14].

На рис. 9 качественно иллюстрируются два рассмотренных выше механизма рассеяния: случай короткодействующих потенциалов рассеяния (sr/short range) и рассеяние на ионизированных примесях (ii/ionized impurities). Скорости рассеяния складываются, так что для суммарного значения средней длины свободного пробега имеем

$$\frac{1}{\lambda_{tot}} = \frac{1}{\lambda_{ii}} + \frac{1}{\lambda_{sr}} \tag{43}$$

и

$$\frac{1}{\sigma_{tot}} = \frac{1}{\sigma_{ii}} + \frac{1}{\sigma_{sr}}. \tag{44}$$

Для выбранного значения n_S наименьший из двух вкладов лимитирует суммарное значение σ_S, из чего следует нелинейная суммарная зависимость $\sigma_S(n_S)$.

Еще несколько механизмов рассеяния могут оказаться важными в графене. Это, например, короткодействующее рассеяние на дефектах сотовой структуры графена и рассеяние на полярных оптических фононах в SiO_2, часто служащем подложкой для графена [15 – 17].

3.5. Подвижность и формула Друде

Подвижность в графене находим уравниванием выражений (37) и (41):

$$\mu = \frac{2q}{h}\frac{1}{n_S}\left(\frac{2E_F}{\pi\hbar v_F}\right)\lambda(E_F). \tag{45}$$

После подстановки выражения для n_S (31) получаем

$$\mu = \frac{2qv_F}{\pi E_F}\lambda(E_F). \tag{46}$$

Часто пишут, что электроны в графене «безмассовые». Это утверждение не лишено смысла, поскольку линейная дисперсия в графене (24) похожа на частотную дисперсию для фотонов $\omega(k)$. Соответственно этому для нахождения подвижности не стоит пользоваться формулой Друде с некой эффективной массой

$$\mu = \frac{q\tau_p}{m^*}, \tag{47}$$

а пользоваться напрямую формулой (46). Вместе с тем, связь между двумя подходами можно установить. Для этого выразим длину свободного пробега в (46) через время релаксации по (38). Тогда

$$\mu = \frac{q}{E_F / v_F^2}\tau_p(E_F), \tag{48}$$

что напоминает формулу Друде для подвижности (47), если под так называемой эффективной массой в графене подразумевать

$$m^* = \frac{E_F}{v_F^2}, \tag{49}$$

что напоминает релятивистское выражение для энергии $E = mc^2$.

3.6. Циклотронная частота в графене

При помещении графена в магнитное поле, направленное перпендикулярно его поверхности, циклотронная частота

$$\omega_c = 2\pi / T, \tag{50}$$

где T – период обращения электрона по круговой орбите вокруг вектора магнитного поля (рис. 10)

$$\vec{B} = B_z \hat{z}, \tag{51}$$

для слабых полей циклотронная частота связана со средним временем свободного пробега τ_p соотношением

$$\omega_c \tau_p \ll 1. \tag{52}$$

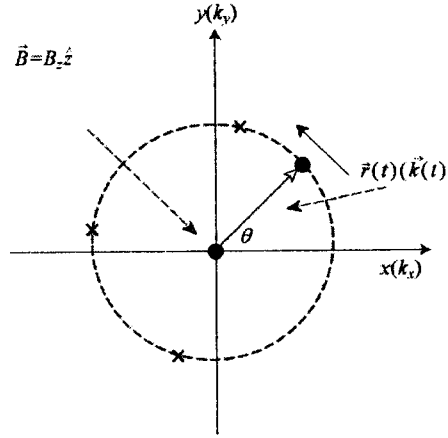


Рис. 10. К вычислению циклотронной частоты.
Крестиком на траектории отмечены акты рассеяния.

Для параболической дисперсии, как известно [30],

$$\omega_c = \frac{qB_z}{m^*}. \tag{53}$$

Тогда условие для слабых полей (52) с учетом (47) можно переписать иначе:

$$\mu B_z \ll 1. \tag{54}$$

Из (50) и (52) следует, что для слабых полей

$$T \gg \tau_p, \tag{55}$$

другими словами, период обращения электрона намного превышает время между столкновениями, в результате чего лишь редкие орбиты оказываются замкнутыми, что и приводит к наблюдению классического эффекта Холла. В противном случае, когда

$$\omega_c \tau_p \gg 1, \tag{56}$$

что соответствует сильным магнитным полям, наблюдается квантовый эффект Холла.

Получим общее выражение для циклотронной частоты, пригодное для произвольной дисперсии. Для силы Лоренца, действующей на электрон в магнитном поле, имеем

$$\frac{d\vec{p}}{dt} = -q\vec{v} \times \vec{B}. \tag{57}$$

Учтем, что импульс

$$\vec{p} = \hbar\vec{k}, \tag{58}$$

и распишем (57) покомпонентно:

$$\hbar\frac{dk_x}{dt} = -qv_yB_z,$$

$$\hbar\frac{dk_y}{dt} = +qv_xB_z,$$

что дает (рис. 10):

$$\hbar k\frac{d(\cos\theta)}{dt} = -q(v\sin\theta)B_z, \tag{59}$$

$$\hbar k\frac{d(\sin\theta)}{dt} = +q(v\cos\theta)B_z. \tag{60}$$

Дифференцируя (59) по времени и используя (60), получаем

$$\frac{d^2(\cos\theta)}{dt^2} = -\omega_c^2\cos\theta,$$

где искомая циклотронная частота для изотропной зонной структуры

$$\omega_c = \frac{qv}{\hbar k}B_z. \tag{61}$$

Для параболической дисперсии

$$v = \frac{\hbar k}{m^*}, \tag{62}$$

что и дает для этой дисперсии циклотронную частоту (53).

Для интересующего нас случая графена $v = v_F$ и $E = \hbar v_F k$ для циклотронной частоты в графене из (61) получаем

$$\omega_c = \frac{qB_z}{E_F / v_F^2}. \tag{63}$$

Таким образом, и здесь справедливо полученное ранее выражение (49) для эффективной массы.

3.7. Эффективная масса в графене

Выражение для эффективной массы в графене из подвижности (49) и циклотронной частоты (63) указывает на то, что должен быть более общий подход к понятию эффективной массы, чем традиционное определение ее через кривизну дисперсионного соотношения

$$m^* = \hbar^2 / \left(\frac{d^2 E}{dk^2} \right)_{k=0}. \tag{64}$$

Разные способы записи проводимости для 2D проводников приведены в [31]. В общем случае имеем два эквивалентных выражения для проводимости (глава 1), одно из которых выражает проводимость через произведение плотности состояний $D(E)$ и коэффициента диффузии $\overline{D}$

$$\sigma(E) = q^2 \overline{D} \frac{D(E)}{L} \left\{ 1, \ \frac{1}{W}, \ \frac{1}{A} \right\}, \tag{65}$$

а другое – через произведение числа мод $M(E)$ в канале проводимости и средней длины пробега $\lambda(E)$

$$\sigma(E) = \frac{q^2}{h} M(E) \lambda(E) \left\{ 1, \ \frac{1}{W}, \ \frac{1}{A} \right\}, \tag{66}$$

где выражение в фигурных скобках показывает множители для проводников разной размерности.

Применимость модели Друде ограничена, в то время как уравнения для проводимости (65) и (66) имеют самое общее значение. Фундаментальное различие между этими уравнениями и теорией Друде состоит в том, что усреднение (5), (6)

$$\sigma = \int_{-\infty}^{+\infty} \sigma(E) \left(-\frac{\partial f_0}{\partial E} \right) dE \tag{67}$$

делает проводимость свойством поверхности Ферми: проводимость определяется уровнями энергии, близкими к $E = E_{F0}$. А согласно теории Друде проводимость зависит от общей электронной плотности, суммированной по всему спектру энергий, что и приводит к ограниченной применимости модели Друде. Проводимость веществ меняется в очень широких пределах несмотря на то, что число электронов в разных веществах приблизительно одинаково. Низкая проводимость стекла объясняется не тем, что в стекле мало так

называемых «свободных» электронов, а потому, что для стекла характерна очень низкая плотность состояний и числа мод вблизи $E = E_{F0}$. Понятие же «свободных» электронов относится к интуитивным понятиям.

Для произвольных проводников, как с кристаллической структурой, так и аморфных, так и для молекулярных проводников в главе 1 показано [32], что независимо от функциональной зависимости дисперсионного соотношения $E(p)$ плотность состояний $D(E)$, скорость $v(E)$ и импульс $p(E)$ связаны с числом электронных состояний $N(E)$ с энергией, меньшей значения E, фундаментальным соотношением

$$D(E)v(E)p(E) = N(E)\cdot d, \tag{68}$$

где d – размерность проводника. Используя (68) для вычисления проводимости (65) с учетом коэффициента диффузии [33, 34]

$$\bar{D} = \left\langle v_x^2 \tau_p \right\rangle \tag{69}$$

для 3d-проводника получим

$$\sigma(E) = q^2 \frac{N(E)}{A\cdot L}\frac{\tau_p(E)}{m(E)}. \tag{70}$$

где масса определена как

$$m(E) = \frac{p(E)}{v(E)}. \tag{71}$$

Легко убедиться, что фундаментальное соотношение (68) справедливо и для параболической дисперсии и для линейной. Для параболической зависимости масса носителей тока от энергии не зависит, что в общем случае не так.

Формула (70) выглядит как в теории Друде, если $N/A\cdot L$ считать электронной плотностью n. При низких температурах это действительно так, поскольку усреднение (67) при $E = E_{F0}$ дает

$$\sigma = \left(q^2 \frac{N(E)}{A\cdot L}\frac{\tau_p}{m} \right)_{E=E_{F0}} = q^2 n \tau_p / m, \tag{72}$$

так как $N(E)$ при $E = E_{F0}$ есть полное число электронов. При ненулевой температуре ситуация тем более сложнее, поскольку плотность состояний непараболическая. Отметим, что ключевым моментом в сведении общего выражения для проводимости (65) к выражению (70), похожему на формулу Друде, есть фундаментальное соотношение (68), связывающее плотность состояний $D(E)$, скорость $v(E)$ и импульс $p(E)$ для данного значения энергии с полным числом состояний $N(E)$, полученных интегрирование плотности состояний

$$N(E) = \int_{-\infty}^{E} D(E)\,dE. \tag{73}$$

3.8. Плотность фононных состояний в графене

Плотность фононных состояний и фононные моды для проводников разной размерности подробно рассмотрены в главе 2. В качестве примера обратимся к вычислению плотности состояний для фононных акустических мод в графене, ход которых качественно показан на рис. 11.

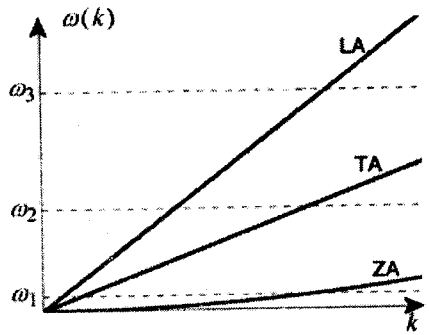


Рис. 11. К вычислению плотности фононных состояний на разных частотах.

Вычислим плотность фононных состояний для конкретных частот ω_1, ω_2 и ω_3. Если продольные LA и поперечные TA акустические моды можно аппроксимировать линейной зависимостью $\omega(k)$, а именно:

$$\omega = v_g^{\mathrm{LA}} k, \qquad (74)$$

$$\omega = v_g^{\mathrm{TA}} k,$$

то изгибные моды (flexural/ZA) ближе всего аппроксимируются квадратичной зависимостью частоты

$$\omega \propto Const^{ZA} \cdot k^2, \qquad (75)$$

где для графена [35, 36]

$$v_g^{\mathrm{LA}} \approx 2\times 10^4,\ v_g^{\mathrm{TA}} \approx 1.5\times 10^4 м/с,\ \ Const^{ZA} \approx 5\times 10^{-7} м^2/с.$$

Подстановка (74) в (207/гл.2) окончательно для плотности фононных LA мод дает

$$D_{2D}^{LA}(\omega) = \frac{\omega}{2\pi (v_g^{\mathrm{LA}})^2}$$

и аналогично для TA мод

$$D_{2D}^{TA}(\omega) = \frac{\omega}{2\pi (v_g^{\mathrm{TA}})^2}.$$

Для ZA мод с учетом (204/гл.2) и (74) окончательно имеем

$$D_{2D}^{ZA}(\omega)=\frac{1}{L^2}\frac{dN_{2D}(k)}{dk}\frac{dk}{d\omega}=\frac{1}{4\pi\cdot Const^{ZA}}. \tag{76}$$

На частоте ω_1 вклад в плотность фононных состояний дают все три моды (рис. 11):

$$\begin{aligned} D_{2D}^{tot}(\omega_1)&=D_{2D}^{LA}(\omega_1)+D_{2D}^{TA}(\omega_1)+D_{2D}^{ZA}(\omega_1)= \\ &=\frac{\omega_1}{2\pi(v_g^{\mathrm{LA}})^2}+\frac{\omega_1}{2\pi(v_g^{\mathrm{TA}})^2}+\frac{1}{4\pi\cdot Const^{ZA}}, \end{aligned} \tag{77}$$

на частоте ω_2 – только моды LA и TA:

$$D_{2D}^{tot}(\omega_2)=D_{2D}^{LA}(\omega_2)+D_{2D}^{TA}(\omega_2)=\frac{\omega_2}{2\pi(v_g^{\mathrm{LA}})^2}+\frac{\omega_2}{2\pi(v_g^{\mathrm{TA}})^2},$$

а на частоте ω_3 – только мода LA:

$$D_{2D}^{tot}(\omega_3)=D_{2D}^{LA}(\omega_3)=\frac{\omega_3}{2\pi(v_g^{\mathrm{LA}})^2}.$$

3.9. Вклад электронов и фононов в теплопроводность графена

Уникальная зонная структура графена в точках Дирака порождает исключительно высокую электронную проводимость графена, что в свою очередь привело к множеству разнообразных предложений и идей по применению графена в различных и порой необычных электронных устройствах [37]. В отличие от большинства известных материалов линейность зонной структуры графена напоминает поведение длинноволновых акустических фононов. В отличие от исключительного внимания к электрическим свойствам графена его термические свойства менее изучены. Вместе с тем именно фононы ответственны за его теплофизические свойства. Приведем качественное объяснение причин такой ситуации. Для этого нужно обратиться к фононным модам.

Согласно (34), в случае электронов для плотности числа мод имеем

$$M_{2D}(E)=\frac{2|E|}{\pi\hbar v_F}. \tag{78}$$

Для плотности числа мод фононов, согласно (212/гл.2), имеем

$$M_{2D}(\omega)=2v_g(\omega)D_{2D}(\omega)=\frac{k(\omega)}{\pi}. \tag{79}$$

Для LA и TA акустических фононов воспользуемся, согласно (74), линейным дебаевским приближением, а для ZA фононов – квадратичной аппроксимацией (75). Тогда для плотности фононных состояний имеем:

$$M_{2D}^{LA}(\omega)=\frac{\omega}{\pi v_g^{LA}},$$
$$M_{2D}^{TA}(\omega)=\frac{\omega}{\pi v_g^{TA}}, \qquad (80)$$
$$M_{2D}^{ZA}(\omega)=\frac{1}{\pi}\sqrt{\frac{\omega}{Const^{ZA}}}$$

с параметрами v_g^{LA}, v_g^{TA} и $Const^{ZA}$ по (75).

Сравнивая плотность числа электронных мод (78) с плотностью LA и TA фононных состояний (80), мы фактически имеем одно и то же выражение ($E=\hbar\omega$) за исключением двойки, учитывающей вырождение по спину. Отличаются эти выражения лишь скоростью: скорость электронов v_F на два порядка больше групповой скорости фононов v_g. В результате плотность электронных мод намного меньше плотности фононных мод. Это наглядно видно из рис. 12, где зависимость плотности всех акустических фононных мод отложена вместе с плотностью электронных мод в интервале $0<E\le 50\,мэВ$ с максимальным значением, соответствующим $\approx 2kT$ при комнатной температуре.

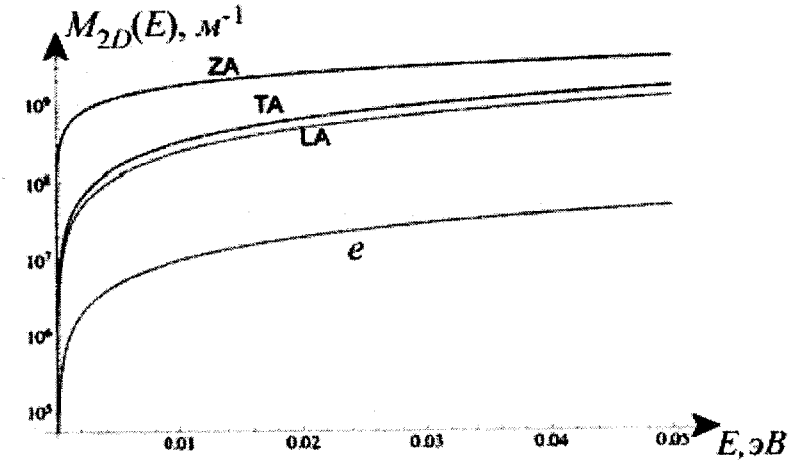


Рис. 12. Плотность электронных и фононных состояний в графене вплоть до 50 *мэВ* [36].

Как следствие, нужно ожидать намного более существенный вклад фононов в передачу тепла по сравнению с электронами. А среди фононных

акустических мод наиболее важную роль в передаче тепла играют флексюральные моды ZA по крайней мере вплоть до комнатных температур.

3.10. Максимальная проводимость графена

Рассмотрим идеально чистый графен без дефектов и без заряженных примесей, рассеивающих электроны. Какую проводимость можно ожидать? При высоких температурах имеет место рассеяние на фононах, вблизи же комнатной температуры доминирует рассеяние на акустических фононах. Скорость рассеяния пропорциональна плотности состояний, а для времени рассеяния получено следующее выражение [28]:

$$\tau_p(E) = \frac{4\hbar^3 \rho_m v_F^2 v_S^2}{D_A^2 kT}\left(\frac{1}{E}\right), \tag{81}$$

где ρ_m – массовая плотность, v_S – скорость звука, D_A – акустический деформационный потенциал, служащий мерой электрон-фононной связи. Тогда для среднего пути свободного пробега, согласно (38), получаем:

$$\lambda(E) = \frac{2\pi\hbar^3 \rho_m v_F^3 v_S^2}{D_A^2 kT}\left(\frac{1}{E}\right) \tag{82}$$

и далее из уравнения (37) для поверхностной проводимости имеем:

$$\sigma_S = \frac{4q^2\hbar \rho_m v_F^2 v_S^2}{\pi D_A^2 kT}. \tag{83}$$

Подставляя следующие значения свойств для графена [29]

$$v_S \approx 2.1\times10^4\, м/с, \quad \rho_m \approx 7.6\times10^{-7}\, кг/м^2, \quad D_A \approx 18\, эВ,$$

получаем

$$\rho_S = \frac{1}{\sigma_S} \approx 30\, Ом/\square. \tag{84}$$

Оценка (84) дает ожидаемый нижний предел для удельного поверхностного сопротивления (верхний предел для удельной поверхностной проводимости). В реальности другие механизмы рассеяния увеличат поверхностное сопротивление. Как соотносится полученная оценка с поверхностным сопротивлением для других материалов? Рассмотрим, например, Si MOSFET с $n_S \approx 10^{13} см^{-3}$ и подвижностью в инверсионном слое $\mu_{eff} \approx 250\, см^2/В\cdot с$. Для такого 2D проводника $\rho_S \approx 2500\, Ом/\square$. Для высокопроводящего InGaAs с $n_S \approx 2\times10^{12} см^{-3}$ и подвижностью

$\mu_{eff} \approx 1\times10^5\ см^2/В\cdot с$ находим, что $\rho_S \approx 300\ Ом/\square$. Действительно, для графена можно ожидать исключительно высокую проводимость.

Далее в справочных целях приводим термоэлектрические коэффициенты для графена в баллистическом и диффузионном режимах [29, 38].

3.11. Термоэлектрические коэффициенты графена

В приведенных ниже формулах предполагается линейная дисперсия (24). Число мод дается выражением (34). Рассеяние описывается степенным законом

$$\lambda(E) = \lambda_0\left(\frac{E}{kT}\right)^r.$$

Положение уровня Ферми E_F относительно дна зоны проводимости E_C определяется параметром

$$\eta_F = \frac{E_F - E_C}{kT}.$$

Интегралы Ферми – Дирака порядка j определяются следующим образом (§ 1.3.6.4):

$$\Im_j(\eta_F) = \frac{1}{\Gamma(j+1)}\int_0^\infty \frac{\eta^j}{\exp(\eta-\eta_F)+1}d\eta.$$

Проводимость $G = \sigma W/L$ приведена в *Сименсах*: $[\sigma] = 1См$. Аналогично для остальных удельных термоэлектрических коэффициентов:

$$s_T = S_T L/W;\ \kappa_0 = K_0 L/W;\ \kappa = KL/W.$$

3.11.1. Баллистический режим

$$G^{ball} = W\frac{2q^2}{h}\left(\frac{2kT}{\pi\hbar v_F}\right)\left[\Im_0(\eta_F)+\Im_0(-\eta_F)\right];$$

$$S^{ball} = -\frac{k}{q}\left\{\frac{2\left[\Im_1(\eta_F)-\Im_1(-\eta_F)\right]}{\Im_0(\eta_F)+\Im_0(-\eta_F)}-\eta_F\right\};$$

$$S_T^{ball} = -W\frac{2q^2}{h}\left(\frac{2kT}{\pi\hbar v_F}\right)\left(\frac{k}{q}\right)\left\{2\left[\Im_1(\eta_F)-\Im_1(-\eta_F)\right]-\eta_F\left[\Im_0(\eta_F)+\Im_0(-\eta_F)\right]\right\};$$

$$K^{ball}=WT\frac{2q^2}{h}\left(\frac{2kT}{\pi\hbar v_F}\right)\left(\frac{k}{q}\right)^2\left\{6\left[\Im_2(\eta_F)+\Im_2(-\eta_F)\right]-\frac{4\left[\Im_1(\eta_F)-\Im_1(-\eta_F)\right]^2}{\Im_0(\eta_F)+\Im_0(-\eta_F)}\right\};$$

$$K_0^{ball}=WT\frac{2q^2}{h}\left(\frac{2kT}{\pi\hbar v_F}\right)\left(\frac{k}{q}\right)^2\times$$

$$\times\left\{6\left[\Im_2(\eta_F)+\Im_2(-\eta_F)\right]-4\eta_F\left[\Im_1(\eta_F)-\Im_1(-\eta_F)\right]+\eta_F^2\left[\Im_0(\eta_F)+\Im_0(-\eta_F)\right]\right\}$$

3. 11.2. Диффузионный режим

$$G^{diff}=W\frac{2q^2}{h}\left(\frac{2kT}{\pi\hbar v_F}\right)\left(\frac{\lambda_0}{L}\right)\Gamma(r+2)\left[\Im_r(\eta_F)+\Im_r(-\eta_F)\right];$$

$$S^{diff}=-\frac{k}{q}\left\{\frac{(r+2)\left[\Im_{r+1}(\eta_F)-\Im_{r+1}(-\eta_F)\right]}{\Im_r(\eta_F)+\Im_r(-\eta_F)}-\eta_F\right\};$$

$$S_T^{diff}=-W\frac{2q^2}{h}\frac{k}{q}\left(\frac{2kT}{\pi\hbar v_F}\right)\left(\frac{\lambda_0}{L}\right)\left\{\begin{array}{l}\Gamma(r+3)\left[\Im_{r+1}(\eta_F)-\Im_{r+1}(-\eta_F)\right]-\\-\eta_F\Gamma(r+2)\left[\Im_r(\eta_F)+\Im_r(-\eta_F)\right];\end{array}\right\}$$

$$K=WT\frac{2q^2}{h}\left(\frac{k}{q}\right)^2\left(\frac{2kT}{\pi\hbar v_F}\right)\left(\frac{\lambda_0}{L}\right)\times$$

$$\times\Gamma(r+3)\left\{(r+3)\left[\Im_{r+2}(\eta_F)+\Im_{r+2}(-\eta_F)\right]-\frac{(r+2)\left[\Im_{r+1}(\eta_F)-\Im_{r+1}(-\eta_F)\right]^2}{\Im_r(\eta_F)+\Im_r(-\eta_F)}\right\};$$

$$K_0=WT\frac{2q^2}{h}\left(\frac{k}{q}\right)^2\left(\frac{2kT}{\pi\hbar v_F}\right)\left(\frac{\lambda_0}{L}\right)\times$$

$$\times\left\{\begin{array}{l}\Gamma(r+4)\left[\Im_{r+2}(\eta_F)+\Im_{r+2}(-\eta_F)\right]-2\eta_F\Gamma(r+3)\left[\Im_{r+1}(\eta_F)+\Im_{r+1}(-\eta_F)\right]+\\+\eta_F^2\Gamma(r+2)\left[\Im_r(\eta_F)+\Im_r(-\eta_F)\right].\end{array}\right\}$$

Итак, в этой главе приведена элементарная теория зонной структуры графена и рассмотрены такие свойства графена как плотность электронных состояний и носителей тока, число мод и максимальная проводимость, рассеяние и подвижность электронов в графене, циклотронная частота и эффективная масса, плотность фононных состояний, относительный вклад электронов и фононов в теплопроводность графена [41, 42].

Литература

1. Yu. Kruglyak. Landauer – Datta – Lundstrom Generalized Transport Model for Nanoelectronics, *J. Nanoscience*, v. 2014, Art. ID 725420, 15 pp. (2014); DOI: 10.1155/2014/725420.
2. Ю. А. Кругляк. Обобщённая модель электронного транспорта Ландауэра – Датты – Лундстрома, *Наносистемы, наноматериалы, нанотехнологии*, т. 11, № 3, 519 – 549 (2013); Erratum: *ibid*, т. 12, № 2, 415 (2014).
3. М. В. Стріха. Фізика графену: стан і перспективи, *Сенсор. електр. мікросист. техн.*, т. 7, № 3, 5 – 13 (2010).
4. A. K. Geim. Graphene: Status and Prospects, *Science*, v. 324, 1530 – 1534 (2009).
5. K. S. Novoselov. Beyond the wonder material, Physics World, v. 22, n 8, 27 – 30 (2009).
6. Ю. Е. Лозовик, С. П. Меркулова, А. А. Соколик. Коллективные электронные явления в графене, Усп. физ. наук, т. 178, № 7, 757 – 776 (2008).
7. С. В. Морозов, К. С. Новоселов, А. К. Гейм. Электронный транспорт в графене, Усп. физ. наук, т. 178, № 7, 776 – 780 (2008).
8. T. Ando. Physics of Graphene. Zero-Mode Anomalies and Roles of Symmetry, *Prog. Theor. Phys. Suppl.*, n 176, 203 – 226 (2008).
9. A. K. Geim, K. S. Novoselov. The Rise of Graphene, *Nature Mat.*, v. 6, 183 – 191 (2007).
10. J. W. McClure. Diamagnetism of Graphite, *Phys. Rev.*, v. 104, 666 – 671 (1956).
11. J. C. Slonczewski, P. R. Weiss. Band Structure of Graphite, *Phys. Rev.*, v. 109, n 2, 272 – 279 (1958).
12. T. Ando. Theory of electronic states and transport in carbon nanotubes, *J. Phys. Soc. Japan*, v. 74, 777 – 817 (2005).
13. N. H. Shon, T. Ando. Quantum transport in two-dimensional graphite system, *J. Phys. Soc. Japan*, v. 67, 2421 – 2429 (1998).
14. N. M. R. Peres, J. M. B. Lopes dos Santos, T. Stauber. Phenomenological study of the electronic transport coefficients of graphene, *Phys.Rev.B*, v. 76, n 7, 073412 (2007).

15. W. Zhu, V. Perebeinos, M. Freitag, P. Avouris. Carrier scattering, mobilities, and electrostatic potential in monolayer, bilayer, and trilayer graphene, *Phys. Rev. B*, v. 80, 235 – 402 (2009).
16. V. Perebeinos, P. Avouris. Inelastic scattering and current saturation graphene, *Phys. Rev.B*, v. 81, 195442 (2010).
17. S. Das Sarma, S. Adam, E. H. Hwang, E. Rossi. Electronic transport in two-dimensional graphene, *Rev. Mod. Phys.*, v. 83, 407 – 470 (2011).
18. R. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, A. A. Firsov. Electric Field Effect in Atomically Thin Carbon Films, *Science,* v. 306, 666 – 669 (2004).
19. S. V. Morozov, K. S. Novoselov, F. Schedin, D. Jiang, A.A. Firsov, A. K. Geim. Two-dimensional electron and hole gases at the surface of graphite, *Phys. Rev. B*, v. 72, n 20, 201401 (2005).
20. R. S. Novoselov, A. K. Geim, S. V. Morozov, Jiang D., Katsnelson M.I., Grigorieva I.V., Dubonos S.V., Firsov A.A. Two-Dimensional Gas of Massless Dirac Fermions in Graphene, Nature, v. 438, 197 – 200 (2005).
21. Y. Zhang, Y.-W. Tan, H. L. Stormer, P. Kim. Experimental observation of the quantum Hall effect and Berry's phase in grapheme, *Nature*, v. 438, 201 – 204 (2005).
22. Robert B. Laughlin. Condensed Matter Theory (II): Graphene Band Structure / Graphene Density of States: http://large.stanford.edu/courses/.
23. Supriyo Datta. Graphene Bandstructures (Purdue University, 2008: www.nanohub.org/resources/5710).
24. Supriyo Datta. Graphene Density of States I (Purdue University, 2008: www.nanohub.org/resources/5721).
25. Supriyo Datta. Graphene Density of States II (Purdue University, 2008: www.nanohub.org/resources/5722).
26. Ю. А. Кругляк, Н. Е. Кругляк. Методические аспекты расчета зонной структуры графена с учетом σ-остова. Теоретические основы, *Вестник Одес. гос. эколог. ун-та*, № 13, 207 – 218 (2012).
27. Mark Lundstrom. Sums in *k*-space / Integrals in Energy Space (Purdue University, 2009: www.nanohub.org/resources/7296).
28. Dionisis Berdebes, Tony Low, Mark Lundstrom. Lecture Notes on Low Bias Transport in Graphene: An Introduction (Purdue University,2009: www.nanohub.org/resources/7435).
29. M. Lundstrom, C. Jeong, *Near-Equilibrium Transport: Fundamentals and Applications* (Hackensack, New Jersey: World Scientific Publishing Company: 2013; www.nanohub.org/resources/11763).
30. Ю. О. Кругляк, М. В. Стріха. Уроки наноелектроніки: ефект Холла і вимірювання електрохімічних потенціалів у концепції «знизу – вгору», *Сенсор. електр. мікросист. техн.*, т. 11, № 1, 5 – 27 (2014).

31. Ю. А. Кругляк. От баллистической проводимости к диффузионной в транспортной модели Ландауэра – Датты – Лундстрома, *Наносистемы, наноматериалы, нанотехнологии*, т. 11, № 4, 655 – 677 (2013).
32. Ю. А. Кругляк, Н. Е. Кругляк. Уроки наноэлектроники. 3. Электронная проводимость и моды проводимости в концепции "снизу – вверх", *Физич. образов. в вузах*, т. 19, № 3, 99 – 110 (2013).
33. Ю. А. Кругляк, Н. Е. Кругляк. Уроки наноэлектроники. 2. Модель упругого резистора и новая формулировка закона Ома в концепции «снизу – вверх», *Физич. образов. в вузах*, т. 19, № 2, 161 – 173 (2013).
34. Yu. A.Kruglyak. Heat transfer by phonons in Landauer – Datta – Lundstrom approach, *Proc. Intern. Conf. "Nanomaterials: Applications and Properties"*, v. 3, n 2, 02NAESF02 (2014).
35. D. Singh, J. Y. Murthy, T. S. Fisher. Spectral phonon conduction and dominant scattering pathways in graphene, *J. Appl. Phys.*, v. 110, n 9, 094312 (2011).
36. Timothy S. Fisher. *Thermal Energy at the Nanoscale.* Hackensack, New Jersey: World Scientific Publishing Company, 2013; www.nanohub.org/courses/2.
37. S. Das Sarma, S. Adam, E. H. Hwang, E. Rossi. Electronic transport in two-dimensional graphene, *Rev. Mod. Phys.*, v. 83, n *2*, 407 (2011).
38. Ra Seong Kim. *Physics and Simulation of Nanoscale Electronic and Thermoelectric Devices.* West Lafayette: Purdue University, 2011.
39. Supriyo Datta. *Lessons from Nanoelectronics: A New Perspective on Transport.* Hackensack, New Jersey: World Scientific Publishing Company, 2012; www.nanohub.org/courses/FoN1.
40. Ю. О. Кругляк, М. В. Стріха. Уроки наноелектроніки: Метод нерівноважних функцій Гріна у матричному зображенні. II. Модельні транспортні задачі, *Сенсор. електр. мікросист. техн.*, т. 10, № 4, С. 5 – 22 (2013).
41. Ю. А. Кругляк. Графен в транспортной модели Ландауэра – Датты – Лундстрома, *ScienceRise*, т. 2, № 2(7), 93 – 106 (2015).
42. Ю. А. Кругляк. Свойства графена в транспортной модели Ландауэра – Датты – Лундстрома, *Наносистемы, наноматериалы, нанотехнологии*, т. 13, № 2, 215 – 242 (2015).

Глава 4. Учет рассеяния в обобщенной транспортной модели

4.1. Качественные аспекты физики рассеяния электронов

При описании переноса электронов в диффузионном режиме обобщенной транспортной модели (глава 1) [1 – 4] важную роль играет средняя длина свободного обратного рассеяния (mean-free-pass for backscattering) λ , которая определяет коэффициент прохождения

$$T(E)=\frac{\lambda(E)}{\lambda(E)+L}, \tag{1}$$

где L – длина проводника. Откуда берется это выражение?

Обычно под λ подразумевается среднее расстояние между двумя актами рассеяния. В подходе Ландауэра [5, 6] величина λ имеет специальное значение: это длина, на которой коэффициент прохождения (1) уменьшается вдвое.

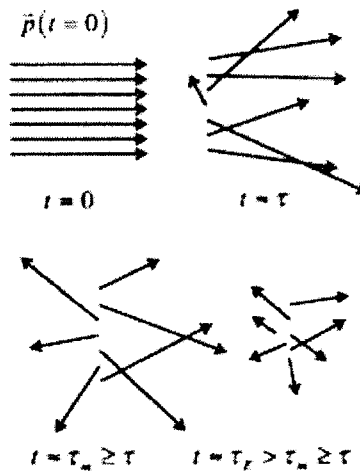


Рис. 1. Качественная визуализация различных времен рассеяния. В момент времени $t = 0$ электроны с одинаковыми импульсами впрыскиваются в проводник в одном направлении. Электроны испытывают в среднем одно столкновение через время $t\approx\tau(E)$. Суммарный импульс первоначального ансамбля электронов зануляется через время $t\approx\tau_m(E)\geq\tau(E)$, а энергия принимает свое равновесное значение через время $t\approx\tau_E(E)>\tau_m(E)\geq\tau(E)$ [2, 7].

Мы позже увидим, что средняя длина свободного рассеяния назад пропорциональна средней длине свободного пробега, Λ,

$$\lambda(E) \propto \Lambda(E) = v(E)\tau(E), \tag{2}$$

где средняя скорость определяется зонной структурой резистора, а среднее время рассеяния определяется также и физикой процессов рассеяния. Хотелось бы глубже понять как упомянутые скорость и время определяют величину λ. В этой главе обсудим также каким образом среднюю длину свободного пробега определить из измерений проводимости или подвижности. Начнем с того как контролируется время рассеяния $\tau(E)$.

Пусть ансамбль электронов с энергией E впрыскивается в проводник в момент времени t = 0 (рис. 1). Первоначальные импульсы одинаковы и однонаправлены. Через время $t \approx \tau(E)$ электроны испытают в среднем одно столкновение. В зависимости от физики процессов рассеяния импульсы электронов (направления стрелок на рис. 1) могут измениться, а их энергия увеличится или уменьшится (длины стрелок на рис. 1). Если рассеяние анизотропно и стремится отклонить электроны на небольшой угол, то одного акта рассеяния недостаточно для зануления суммарного импульса ансамбля электронов. По истечении еще некоторого времени $t \approx \tau_m(E) \geq \tau(E)$ импульсы электронов рандомизируются. Если, однако, доминирующий механизм рассеяния упругий, то первоначальная энергия ансамбля электронов не примет еще своего равновесного значения. По происшествии еще некоторого времени $t \approx \tau_E(E) \gg \tau_m(E), \tau(E)$ первоначальный избыток энергии занулится и энергия примет свое равновесное значение. Рис. 1 наглядно иллюстрирует три характеристических времени рассеяния: 1) среднее время между двумя последовательными актами рассеяния $\tau(E)$; 2) время релаксации импульса $\tau_m(E)$; 3) время релаксации энергии $\tau_E(E)$. В общем случае $\tau_m(E) \geq \tau(E)$ и $\tau_E(E) \gg \tau_m(E), \tau(E)$.

Поскольку мы рассматриваем транспорт зарядов и перенос тепла, то наибольший интерес для нас представляет время импульсной релаксации, а также то каким образом это время зависит от физики рассеяния.

Фундаментальным понятием в теории рассеяния является матрица рассеяния $S(\vec{p} \to \vec{p}')$, переводящая систему частиц из первоначального состояния $\vec{p}$ в некое конечное состояние $\vec{p}'$.

Скорость рассеяния, иначе вероятность рассеяния за единицу времени, есть просто единица, деленная на среднее время между соседними

столкновениями, и получается путем суммирования по всем возможным конечным состояниям, а именно:

$$\frac{1}{\tau(\vec{p})}=\sum_{\vec{p}'}S(\vec{p}\to\vec{p}'). \tag{3}$$

В предположении, что электроны впрыскиваются в проводник с начальным импульсом, направленным вдоль оси z, для скорости импульсной релаксации, по аналогии с (3), имеем [7]:

$$\frac{1}{\tau_m(\vec{p})}=\sum_{\vec{p}'}S(\vec{p}\to\vec{p}')\frac{\Delta p_z}{p_z}. \tag{4}$$

Анизотропное рассеяние стремится отклонить электроны на небольшие углы, в результате чего время импульсной релаксации растет, а соответствующая скорость падает. Аналогичным образом записывается скорость релаксации энергии через время соответствующей релаксации.

Время импульсной релаксации можно рассчитать из скорости перехода. Техника вычислений скорости перехода подробно изложена в [7]. Здесь мы обрисуем лишь ее основные моменты. Рассмотрим акт рассеяния подробнее (рис. 2).

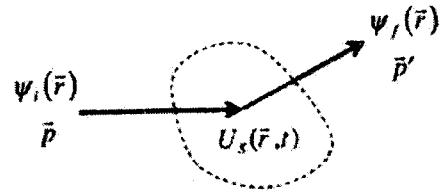


Рис. 2. Переход электрона из начального состояния в конечное в результате акта рассеяния на потенциале $U_S(\vec{r},t)$.

Пусть электрон с импульсом $\vec{p}=\hbar k$ и описываемый волновой функцией $\psi_i(\vec{r})$ оказался в области действия потенциала рассеяния $U_S(\vec{r},t)$, который может быть как статическим (рассеяние на заряженных примесях), так и динамическим (рассеяние на фононах). В результате акта рассеяния первоначальный импульс электрона становится иным $\vec{p}'$, изменяется и его волновая функция на $\psi_f(\vec{r})$. Нужно рассчитать вероятность (за единицу времени) $S(\vec{p}\to\vec{p}')$ перехода электрона из первоначального состояния $\vec{p}$ в конечное $\vec{p}'$.

В первом порядке теории возмущений

$$S(\vec{p} \to \vec{p}') = \frac{2\pi}{\hbar}\left|H_{p',p}\right|^2 \delta(E' - E - \Delta E), \qquad (5)$$

где матричный элемент перехода

$$\left|H_{p',p}\right|^2 = \int_{-\infty}^{+\infty} \psi_f^*(\vec{r}) U_S(\vec{r}) \psi_i(\vec{r}) d\vec{r}\,.$$

Выражение (5) известно в квантовой теории как золотое правило Ферми – Дирака [8, 9]. Фигурирующая в (5) δ- функция ответственна за сохранение энергии. В случае статического потенциала рассеяния (например, на заряженных примесях) энергия не релаксирует ($\Delta E = 0$). Для периодического потенциала с характерной частотой ω (колебания решетки) $\Delta E = \pm\hbar\omega$, что соответствует излучению или поглощению фонона.

В любом случае сначала нужно задаться потенциалом рассеяния, затем воспользоваться (5) и далее вычислить характерное время (3) или (4). Позже будет показано как из времени импульсной релаксации вычислить среднюю длину свободного рассеяния назад.

Некоторые простые потенциалы рассеяния, например, короткодействующие δ-потенциалы, а также потенциалы рассеяния на акустических и оптических фононах в неполярных материалах просто равновероятно отклоняют носители тока с сохранением энергии. В таких случаях можно ожидать, что скорость рассеяния будет пропорциональна плотности конечных состояний. При упругом рассеянии $1/\tau(E) \propto D(E)$, при рассеянии с поглощением фонона $1/\tau(E) \propto D(E + \hbar\omega)$, а с эмиссией фонона $1/\tau(E) \propto D(E - \hbar\omega)$. Поскольку обычно плотность состояний растет с энергией (n-проводники), можно ожидать уменьшения времени рассеяния с ростом энергии носителей тока или тепла.

Ситуация с рассеянием на заряженных примесях или на фононах в полярных средах иная. Случайно расположенные заряды создают флуктуации у дна зоны проводимости $E_c(\vec{r})$, что ведет к рассеянию электронов и фононов (рис. 3).

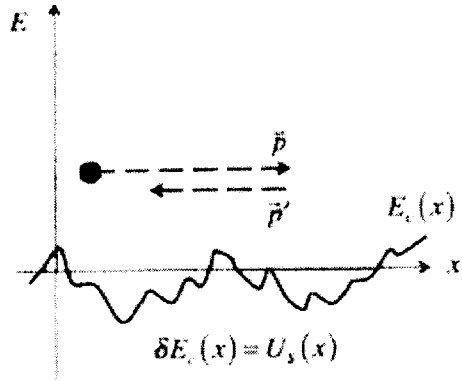


Рис. 3. Флуктуирующий потенциал, создаваемый случайно расположенными зарядами.

Высокоэнергетические электроны менее чувствительны к флуктуирующему потенциалу рассеяния по сравнению с низкоэнергетическими электронами, так что в случае рассеяния на заряженных примесях (и полярных фононах) с ростом энергии носителей время рассеяния $\tau(E)$ будет расти, а скорость рассеяния $1/\tau(E)$ будет падать.

Для некоторых механизмов рассеяния время рассеяния можно записать в виде степенного закона

$$\tau(E) = \tau_0 \left(\frac{E - E_c}{kT} \right)^s, \tag{6}$$

в котором показатель степени разный для различных механизмов рассеяния. Так, для рассеяния на акустических фононах в 3D проводниках с параболической дисперсией $s = -1/2$, а для рассеяния на ионизированных примесях $s = +3/2$ [7].

Наша задача в этом главе лучше понять каким образом рассеяние влияет на среднюю длину свободного пробега и на коэффициент прохождения. Мы ожидаем, что длина свободного пробега будет пропорциональна произведению скорости носителя и времени рассеяния. Мы стали лучше понимать какие факторы влияют на время рассеяния. Прежде чем увязать между собой длину свободного пробега и время рассеяния, сначала рассмотрим как связаны друг с другом длина свободного пробега и коэффициент прохождения.

4.2. Коэффициент прохождения и средняя длина свободного пробега

Связь этих двух важнейших характеристик диффузионного транспорта проще всего проиллюстрировать на примере электронного транспорта по 1D проводнику в диффузионном режиме (рис. 4).

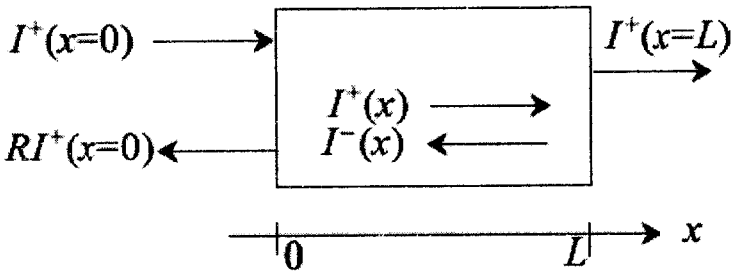


Рис. 4. К выводу связи между коэффициентом прохождения и средней длиной свободного пробега на примере однородного 1D проводника.

Рассмотрим однородный 1D проводник в диффузионном режиме. Левый контакт впрыскивает электроны с током $I^+(x=0)$. Доля электронов T войдет в правый контакт с током $I^+(x=L)=TI^+(x=0)$. Оставшаяся доля вернется в левый контакт (рассеяние назад) с током $I^-(x=0)=RI^+(x=0)$. В отсутствии процессов рекомбинации $T+R=1$. Предполагается также, что правый контакт идеальный, поглощает все входящие в него электроны. Результирующий ток, очевидно, будет $I=(1-R)I^+(0)=TI^+(0)$. Так или иначе, в проводнике имеют место как прямые, так и обратные потоки электронов, и нам необходимо описать их пространственное распределение.

Определим обратное значение средней длины свободного пробега $1/\lambda$ как вероятность (на единицу длины) обращения положительного потока электронов в отрицательный и наоборот. Именно исходя из этого определения, величину λ и называют средней длиной свободного обратного рассеяния (mean-free-pass for backscattering) или проще средней длиной свободного пробега. В диффузионном проводнике некоторая доля положительного потока в результате рассеяния назад обращается в отрицательный. В проводнике формируется отрицательный поток и его некоторая доля в результате рассеяния назад обращается, усиливая положительный поток. В результате градиент положительного тока складывается из двух величин:

$$\frac{dI^+(x)}{dx}=-\frac{I^+(x)}{\lambda}+\frac{I^-(x)}{\lambda}. \qquad (7)$$

Пренебрегая процессами рекомбинации, результирующий ток

$$I = I^{+}(x) - I^{-}(x) \tag{8}$$

является постоянной величиной, так что градиент тока

$$\frac{dI^{+}(x)}{dx} = -\frac{I}{\lambda} \tag{9}$$

является константой. Другими словами, ток спадает линейно вдоль проводника:

$$I^{+}(x) = I^{+}(0) - \frac{I}{\lambda}x. \tag{10}$$

Воспользуемся полученным уравнением для вычисления тока, входящего в правый контакт:

$$I^{+}(L) = I^{+}(0) - \frac{I}{\lambda}L = I^{+}(0) - \left[I^{+}(L) - I^{-}(L)\right]\frac{L}{\lambda} = I^{+}(0) - I^{+}(L)\frac{L}{\lambda}, \tag{11}$$

где мы воспользовались уравнением баланса (8) и тем, что правый контакт идеальный ($I^{-}(L) = 0$). Из последнего равенства находим

$$I^{+}(L) = \frac{\lambda}{\lambda + L} I^{+}(0) = T\, I^{+}(0). \tag{12}$$

Если провести аналогичные рассуждения для впрыскивания электронов правым контактом, то получим аналогичное уравнение для электронов, входящих в левый контакт, а именно: $I^{-}(0) = T' I^{-}(L)$. Для однородного проводника $T' = T$. Проводник под напряжением не является однородным, но нас интересует режим линейного отклика, так что вполне приемлемо положить $T' \approx T$. Окончательно, в предположении независимости друг от друга мод проводимости получаем искомое уравнение (1), связывающее коэффициент прохождения со средней длиной свободного пробега:

$$T(E) = \frac{\lambda(E)}{\lambda(E) + L}.$$

Вывод этого уравнения сделан в рамках простой модели, что никак не мешает успешному и широкому использованию его на практике. Важным моментом в проведенных рассуждениях является интерпретация $1/\lambda$ как вероятности (на единицу длины) обращения потока частиц в обратном

направлении в результате рассеяния. Именно поэтому саму длину λ, как уже упоминалось, часто называют средней длиной свободного рассеяния назад.

4.3. Средняя длина свободного пробега и время рассеяния

Установим связь между λ и временем рассеяния τ_m. Пусть электрон совершает акт рассеяния в изотропном 1D проводнике. У него есть две возможности: рассеяться вперед и рассеяться назад. Только рассеяние назад существенно для определения средней длины свободного пробега. Отсюда следует, что средняя длина рассеяния назад равна удвоенному значению средней длины рассеяния:

$$\lambda_{1D}(E) = 2\Lambda(E) = 2v(E)\tau_m(E).$$

Для проводника произвольной размерности средняя длина рассеяния назад дается выражением [10]

$$\lambda(E) = 2\frac{\left\langle v_x^2 \tau_m \right\rangle}{\left\langle |v_x| \right\rangle},$$

где усреднение для 2D и 3D проводников ведется по углам. Для изотропных проводников

$$\lambda_{2D}(E) = \frac{\pi}{2} v(E)\tau_m(E), \tag{13}$$

$$\lambda_{3D}(E) = \frac{4}{3} v(E)\tau_m(E).$$

Аналогичный (6) степенной закон часто используется для средней длины рассеяния:

$$\lambda(E) = \lambda_0 \left(\frac{E - E_C}{kT} \right)^r. \tag{14}$$

Для параболической зонной структуры $v(E) \propto E^{1/2}$, так что $r = s + 1/2$, где s – показатель степени в степенном законе (6) для времени рассеяния. Для акустических фононов $r = 0$, а для рассеяния на ионизированных примесях $r = 2$.

4.4. Оценка усредненного значения средней длины свободного пробега из экспериментальных измерений

Для 2D проводника в диффузионном режиме (глава 1) [11]

$$G_{2D}^{diff} = \frac{1}{\rho_{2D}} \frac{W}{L} = \sigma_S \frac{W}{L}, \tag{15}$$

где поверхностная проводимость

$$\sigma_S = \frac{2q^2}{h} \int M_{2D}(E) \lambda(E) \left(-\frac{\partial f_0}{\partial E} \right) dE \tag{16}$$

или иначе

$$\sigma_S = \frac{2q^2}{h} \langle M_{2D} \rangle \langle\langle \lambda \rangle\rangle, \tag{17}$$

где

$$\langle M_{2D} \rangle = \frac{\sqrt{\pi}}{2} M_{2D}(kT) \Im_{-1/2}(\eta_F) \tag{18}$$

выражается через интеграл Ферми – Дирака с $\eta_F = (E_F - E_C)/kT$, а усредненное значение длины свободного пробега [11]

$$\langle\langle \lambda \rangle\rangle = \frac{\int \lambda(E) M_{2D}(E) \left(-\frac{\partial f_0}{\partial E} \right) dE}{\int M_{2D}(E) \left(-\frac{\partial f_0}{\partial E} \right) dE} = \frac{\langle M\lambda \rangle}{\langle M \rangle}. \tag{19}$$

Измерив экспериментально поверхностную проводимость (17), мы хотим вычислить $\langle\langle \lambda \rangle\rangle$ из (17), а именно:

$$\langle\langle \lambda \rangle\rangle = \frac{\sigma_S}{\left(2q^2/h\right)} \frac{1}{\langle M_{2D} \rangle}. \tag{20}$$

Для вычисления $\langle M_{2D} \rangle$ согласно (18) нужно знать η_F или иначе положение уровня Ферми относительно дна зоны проводимости. Измерения одной проводимости недостаточно, нужно еще измерить поверхностную плотность электронов [11]

$$n_S = g_v \frac{m^* kT}{\pi \hbar^2} \Im_0(\eta_F) = N_{2D} \Im_0(\eta_F), \tag{21}$$

откуда вычислить η_F, далее $\langle M_{2D} \rangle$ по (18) и, наконец, $\langle\langle \lambda \rangle\rangle$ по (20).

Для невырожденных проводников ситуация упрощается, поскольку в этом случае интегралы Ферми – Дирака сводятся к экспонентам и усредненное значение длины свободного пробега можно сразу записать в явном виде через измеряемые поверхностные проводимость и плотность

$$\langle\langle \lambda \rangle\rangle = \frac{2(kT/q)}{qv_T}\left(\frac{\sigma_S}{n_S}\right),$$

где v_T есть однонаправленная термическая скорость электронов [11]

$$v_T = \sqrt{2kT/\pi m^*}. \qquad (22)$$

Часто, однако, измеряются коэффициент диффузии и подвижность, так что нам нужно увязать $\langle\langle \lambda \rangle\rangle$ с этими измеряемыми величинами.

4.5. Оценка длины свободного пробега из коэффициента диффузии

Вернемся к 1D проводнику на рис. 4. Предположим, что речь идет о транспорте электронов в диффузионном режиме и вычислим ток. На левом конце ($x = 0$) число электронов, движущихся в направлении $+x$, есть $n^+(0) = I^+(0)/\langle v_x^+ \rangle$, где $\langle v_x^+ \rangle$ есть средняя скорость в направлении $+x$. В режиме квазиравновесного транспорта $\langle v_x^- \rangle \approx \langle v_x^+ \rangle$, так что $n^-(0) = I^-(0)/\langle v_x^+ \rangle$. Суммарная плотность электронов

$$n(0) = \frac{(1+R)I^+(0)}{\langle v_x^+ \rangle} = \frac{(2-T)I^+(0)}{\langle v_x^+ \rangle}. \qquad (23)$$

На правом конце проводника $n^+(L) = I^+(L)/\langle v_x^+ \rangle$, а $n^-(L) = 0$, поскольку электроны не впрыскиваются с правого контакта. Суммарная плотность электронов на правом контакте

$$n(L) = \frac{I^+(L)}{\langle v_x^+ \rangle} = \frac{T\,I^+(0)}{\langle v_x^+ \rangle}, \qquad (24)$$

и она меньше, чем на левом контакте:

$$n(0) - n(L) = 2(1-T)\frac{I^+(0)}{\langle v_x^+ \rangle}. \qquad (25)$$

Воспользуемся выражением для суммарного тока $I = T I^{+}(0)$ и найдем его, опираясь на ур-е (25). В результате получим

$$I = \frac{\langle v_x^+ \rangle}{2} \frac{TL}{1-T} \times \left[\frac{n(0)-n(L)}{L} \right] = -\frac{\langle v_x^+ \rangle \lambda}{2} \times \frac{dn(x)}{dx}. \qquad (26)$$

Поскольку коэффициент диффузии в нашем случае (глава 1)

$$\bar{D} = \frac{\langle v_x^+ \rangle \lambda}{2}, \qquad (27)$$

то в итоге мы получаем хорошо известный закон диффузии Фика

$$I = -\bar{D} \times \frac{dn}{dx}. \qquad (28)$$

Этот результат может показаться удивительным, поскольку принято считать, что диффузионный закон Фика выполняется на расстояниях, намного превышающих среднюю длину свободного пробега. Однако, подобного допущения не делалось при выводе ур-я (28). В этой связи еще Шокли [12] заметил, что закон Фика не ограничен большими расстояниями, он применим и для описания баллистического и квазибаллистического транспорта, нужно лишь аккуратно учитывать граничные условия.

Поскольку при выводе использовался 1D проводник, то ток в (28) соответствует одной моде с энергией *E*. При наличии многих каналов проводимости полный ток получается интегрированием по всем модам, так что полученный таким образом коэффициент диффузии будет соответствовать усредненной по энергии длине свободного пробега.

Простая ситуация имеет место в невырожденных проводниках со средней длиной свободного пробега λ_0, не зависящей от энергии. В этом случае коэффициент диффузии зависит от термической скорости (22)

$$\bar{D} = \frac{v_T \lambda_0}{2}, \qquad (29)$$

что открывает простую возможность оценить длину свободного пробега из коэффициента диффузии.

4.6. Связь длины свободного пробега с подвижностью

И поныне часто измеряют подвижность, зная которую можно оценить среднюю длину свободного пробега. Удельная поверхностная проводимость (16) может быть записана как (глава 1)

$$\sigma_S = qn_S\mu, \tag{30}$$

уравнивая которую с (16), для подвижности получаем:

$$\mu \equiv \frac{\frac{2q}{h}\int M_{2D}(E)\lambda(E)\left(-\frac{\partial f_0}{\partial E}\right)dE}{n_S}, \tag{31}$$

и возьмем эту формулу, известную также как формула Кубо – Гринвуда, за определение подвижности.

Уравнение (31), используя (20) и (30), можно переписать в виде

$$\mu = \frac{1}{n_S}\frac{2q}{h}\left\langle\left\langle\lambda\right\rangle\right\rangle\left\langle M_{2D}\right\rangle. \tag{32}$$

Отсюда можно вычислить $\langle\langle\lambda\rangle\rangle$, задавшись измеренными значениями подвижности и поверхностной плотности электронов, как и прежде для вычисления η_F и далее $\langle M_{2D}\rangle$ по (18), а именно:

$$\left\langle\left\langle\lambda\right\rangle\right\rangle = \frac{2kT\mu}{qv_T}\frac{\Im_0(\eta_F)}{\Im_{-1/2}(\eta_F)}. \tag{33}$$

Для невырожденных проводников последняя дробь равна единице, так что $\langle\langle\lambda\rangle\rangle$ легко вычисляется:

$$\left\langle\left\langle\lambda\right\rangle\right\rangle = \frac{2kT\mu}{qv_T}, \tag{34}$$

откуда для подвижности в невырожденных проводниках

$$\mu = \frac{v_T\left\langle\left\langle\lambda\right\rangle\right\rangle}{2}\times\frac{1}{kT/q}, \tag{35}$$

где первый сомножитель можно определить как коэффициент диффузии

$$\bar{D} = \frac{v_T\left\langle\left\langle\lambda\right\rangle\right\rangle}{2}, \tag{36}$$

откуда получается известное выражение Эйнштейна для невырожденных проводников

$$\frac{\bar{D}}{\mu}=\frac{kT}{q}. \tag{37}$$

4.7. Усредненная длина среднего свободного пробега для степенного закона рассеяния

Из определения $\langle\langle\lambda\rangle\rangle$ по (19) с использованием степенного закона рассеяния (14) имеем:

$$\langle\langle\lambda\rangle\rangle=\lambda_0\frac{\int\left(\frac{E-E_C}{kT}\right)^r M_{2D}(E)\left(-\frac{\partial f_0}{\partial E}\right)dE}{\int M_{2D}(E)\left(-\frac{\partial f_0}{\partial E}\right)dE}, \tag{38}$$

что сводится к интегралам Ферми – Дирака, а именно:

$$\langle\langle\lambda\rangle\rangle=\lambda_0\frac{\Gamma(r+3/2)}{\Gamma(3/2)}\times\frac{\Im_{r-1/2}(\eta_F)}{\Im_{1/2}(\eta_F)}. \tag{39}$$

Для невырожденных проводников последний множитель равен единице.

4.8. Подвижность при постоянном значении времени рассеяния

Воспользуемся предыдущим результатом и найдем подвижность для невырожденного проводника, характеризуемого постоянным значением времени рассеяния τ_0. Из (35) с подстановкой (39) имеем:

$$\mu=\frac{v_T\lambda_0}{2}\frac{1}{kT/q}\frac{\Gamma(r+3/2)}{\Gamma(3/2)}. \tag{40}$$

Для 2D проводника из (13)

$$\lambda(E)=\frac{\pi}{2}\nu(E)\tau_0=\left(\frac{\pi}{2}\sqrt{\frac{2kT}{m^*}}\right)\tau_0\left(\frac{E-E_c}{kT}\right)^{1/2}, \tag{41}$$

откуда следует, что показатель степени в законе рассеяния $r=1/2$ и

$$\lambda_0=\left(\frac{\pi}{2}\sqrt{\frac{2kT}{m^*}}\right)\tau_0, \tag{42}$$

а после подстановки в (40) получаем ожидаемый результат:

$$\mu = \frac{q\tau_0}{m^*}. \tag{43}$$

4.9. Анализ экспериментальных данных для Si MOSFET

В качестве примера проведем анализ экспериментальных данных для полевого транзистора Si MOSFET в разных приближениях в рамках обобщенной транспортной модели (глава 1). Вольт-амперная характеристика (ВАХ) Si MOSFET с длиной канала проводимости 60 *нм* приведена на рис. 5.

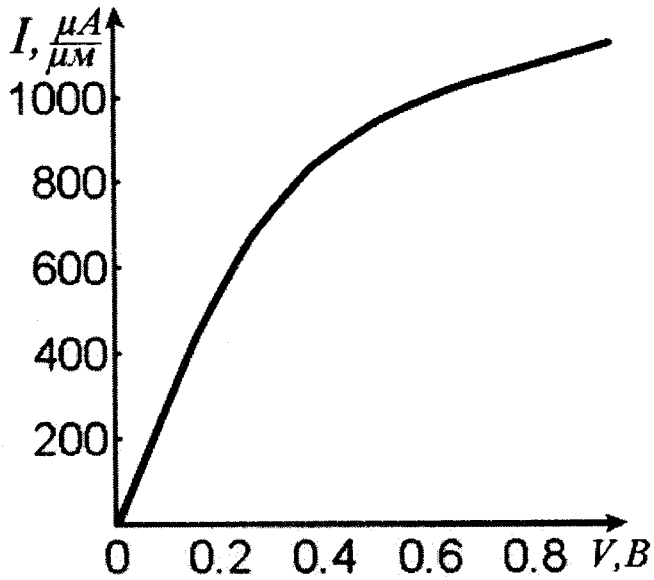


Рис. 5. Вольт-амперная характеристика Si MOSFET при $V_G = 1.2\ B$ [13].

Для интересующего нас линейного участка ВАХ измеренные значения при комнатной температуре

$$n_S \approx 6.7 \times 10^{12} см^{-2},$$

$$R \approx 215\ Ом \cdot \mu м, \tag{44}$$

$$\mu \approx 260\ см^2/B \cdot c.$$

Зададим себе два вопроса:

1) Сколько мод проводимости обеспечивают ток?

2) Насколько измеренное сопротивление близко к баллистическому пределу?

Сначала вспомним, что у кристаллического Si долинное вырождение минимумов эллипсоидальной формы равно шести [14]. В нашем образце квантовые ограничения снимают вырождение до $g_v = 2$ с $m^* = m_t = 0.19 m_0$ [15, 16]. Ответить на интересующие нас вопросы можно с разной степенью достоверности. Для упрощения вычислений можно воспользоваться простой моделью при $T = 0^\circ K$ (§1.3.6.2), что, конечно, может быть недостаточно удовлетворительным, особенно для комнатной температуры. Далее, можно предположить максвелл-больцмановскую статистику для носителей тока (невырожденные проводники), выкладки в этом случае также не вызовут затруднений, однако, выше порогового напряжения допущение невырожденности также неудолетворительно. Наконец, отказаться от каких-либо допущений и добросовестно просчитать интегралы Ферми – Дирака.

Пойдем самым простым путем, что, по крайней мере, даст нам возможность почувствовать числа. В модели $T = 0^\circ K$ плотность числа мод $M_{2D}(E_F) \approx 150\,\mu м^{-1}$. Для наименьшего по размерам из изученных транзисторов $W / L = 2$ [13], а с учетом $L = 60\,нм$ для числа мод, обеспечивающих ток, получаем $M_{2D}(E_F) \approx 18$. Это весьма небольшое число мод.

Баллистический предел находим из выражения (113/гл.1): $R_{2D}^{ball} \approx 90\,Ом \cdot \mu м$, что примерно в два раза меньше измеренного сопротивления (44). Теперь можем получить оценку длины свободного пробега из формулы (140/гл.1): $\lambda(E_F) \approx 40\,нм$. Более тщательный анализ предполагает использование максвелл-больцмановской модели, а еще точнее – вычисление интегралов Ферми – Дирака с учетом заселенности подзон. Однако, так или иначе уже ясно, что современные Si MOSFET работают в квазибаллистическом режиме, а не в баллистическом или чисто диффузионном режиме.

Далее обратимся к оценке баллистической подвижности нашего образца Si MOSFET. Измеренная подвижность (44) относится к достаточно протяженному проводнику и традиционно является диффузионной. Чтобы оценить баллистическую подвижность опять обратимся к простейшей модели $T = 0^\circ K$, в рамках которой уравнение (157/гл.1) можно переписать следующим образом

$$\mu_{2D}^{ball} = \frac{2q}{h} L \sqrt{2 g_v / \pi n_s}\,. \qquad (45)$$

Подставляя все известные величины, находим $\mu \approx 1200\,см^2/В\cdot с$, что в несколько раз превышает диффузионную подвижность. Кажущаяся подвижность, согласно уравнению (152/гл.1), будет несколько меньше подвижности нашего образца за счет баллистической подвижности. К приведенным оценкам нельзя относиться слишком строго из-за явной недостаточности модели $T = 0^\circ K$.

Подвижность (44) была измерена при комнатной температуре. Уточним модель $T = 0^\circ K$, перейдя к максвелл-больцмановской для невырожденных проводников. Из соотношения Эйнштейна (37) найдем коэффициент диффузии

$$\bar{D} = \frac{kT}{q}\mu = 6.7\,см^2/с. \tag{46}$$

Теперь можно вычислить усредненную длину свободного пробега $\langle\langle\lambda\rangle\rangle$ по (36). Для этого нужно оценить термическую скорость v_T (22), что в свою очередь требует знания эффективной массы. Для электронов в инверсионном слое (100) Si, когда заселена только лишь одна подзона зоны проводимости, $m^* = m_t = 0.19 m_0$ [16], откуда $v_T = 1.2\times10^7\,см/с$ и далее

$$\left\langle\left\langle\lambda\right\rangle\right\rangle_{MB} \approx 11\,нм. \tag{47}$$

Еще более точный результат можно получить перейдя к статистике Ферми – Дирака. Из (33) следует

$$\left\langle\left\langle\lambda\right\rangle\right\rangle = \left\langle\left\langle\lambda\right\rangle\right\rangle_{MB} \times \frac{\Im_0(\eta_F)}{\Im_{-1/2}(\eta_F)}. \tag{48}$$

Значение η_F находим через известную измеренную поверхностную плотность

$$n_S = N_{2D}\Im_0(\eta_F) = \left(g_v \frac{m^* kT}{\pi\hbar^2}\right)\Im_0(\eta_F). \tag{49}$$

Для электронов в первой подзоне инверсионного слоя (100) Si $g_v = 2$, так что $N_{2D} = 4.1\times10^{11}\,см^{-2}$. Интеграл Ферми – Дирака вычисляется аналитически

$$\Im_0(\eta_F) = \ln(1+e^{\eta_F}), \tag{50}$$

так что

$$\eta_F = \ln\left(e^{n_S/N_{2D}} - 1\right) = 1.42. \tag{51}$$

Окончательно

$$\left\langle\left\langle \lambda \right\rangle\right\rangle_{FD} = 11 \times \frac{\Im_0(1.42)}{\Im_{-1/2}(1.42)} \approx 15 нм, \tag{52}$$

что и является наилучшей из возможных оценок для рассматриваемого транзистора длиной 60 *нм*. Длина этого транзистора не может считаться слишком большой по сравнению с длиной свободного пробега, так что физически корректно считать, что этот транзистор работает в квазибаллистическом режиме.

Подведем итоги. Основное внимание в этой главе уделено понятию длины свободного пробега λ как длины рассеяния назад и ее связи с коэффициентом прохождения. Показана связь между λ и временем импульсной релаксации для проводников разной размерности. Показано также как вычислить усредненное значение λ из экспериментальных измерений: через коэффициент диффузии или через подвижность. В качестве примера проанализированы экспериментальные данные для Si MOSFET с привлечением моделей различной достоверности [17].

Литература

1. Datta Supriyo. *Lessons from Nanoelectronics: A New Perspective on Transport*. Hackensack, New Jersey: World Scientific Publishing Company, 2012; www.nanohub.org/courses/FoN1.
2. M. Lundstrom, Changwook Jeong. *Near-Equilibrium Transport: Fundamentals and Applications*. Hackensack, New Jersey: World Scientific Publishing Company, 2013; www.nanohub.org/resources/11763.
3. Ю. А. Кругляк. Обобщённая модель электронного транспорта Ландауэра – Датты – Лундстрома, *Наносистемы, наноматериалы, нанотехнологии*, т. 11, № 3, 519 – 549 (2013); Erratum: *ibid*, т. 12, № 2, 415 (2014).
4. Ю. А. Кругляк. Перенос тепла фононами в транспортной модели Ландауэра – Датты – Лундстрома, *ScienceRise*, т. 2, №2(7), 81 – 93 (2015).
5. Rolf Landauer. Spatial variation of currents and fields due to localized scatterers in metallic conduction, *IBM J. Res. Dev.*, v. 1, n.3, 223 – 231 (1957).
6. Rolf Landauer. Electrical resistance of disordered one dimensional lattices, *Philos. Mag.*, v. 21, 863 – 867 (1970).
7. M. Lundstrom. *Fundamentals of Carrier Transport*. Cambridge: Cambridge Univ. Press, 2000.
8. P. A. M. Dirac. The Quantum Theory of Emission and Absorption of Radiation, *Proc. Royal Soc. A*, v. 114, n 767, 243 – 265 (1927).
9. E. Fermi. *Nuclear Physics*. University of Chicago Press, 1950.
10. C. Jeong, R. Kim, M. Luisier, S. Datta, M. Lundstrom. On Landauer versus Boltzmann and full band versus effective mass evaluation of thermoelectric transport coefficients, *J. Appl. Phys.*, v. 107, 023707 (2010).
11. Ю.А.Кругляк. От баллистической проводимости к диффузионной в транспортной модели Ландауэра – Датты – Лундстрома, *Наносистемы, наноматериалы, нанотехнологии*, т. 11, № 4, 655 – 677 (2013).
12. W. Shockley. Diffusion and Drift of Minority Carriers in Semiconductors for Comparable Capture and Scattering Mean Free Paths, *Phys. Rev.*, v. 125, 1570 –1576 (1962).
13. Changwook Jeong, D. A. Antoniadis, M. S. Lundstrom. On backscattering and mobility in nanoscale silicon MOSFETs, *IEEE Trans. Electron Dev.*, v. 56, n 11, 2762 – 2769 (2009).
14. R. F.Pierret. *Semiconductor Device Fundamentals*. Reading, MA: Addison–Wesley, 1996.
15. M. Lundstrom, *ECE 612: Nanoscale Transistors. Lecture 4. Polysilicon Gates/QM Effects*, 2008: www.nanohub.org/resourses/5364.
16. Yuan Taur, Tak Ning. *Fundamentals of Modern VLSI Devices*. Cambridge univ. Press: Cambridge, UK, 2009.
17. Ю. А. Кругляк. Учет рассеяния в транспортной модели Ландауэра – Датты – Лундстрома, *ScienceRise*, т. 3, №2(8), 99 – 107 (2015).

Глава 5. Обобщенная модель переноса и транспортное уравнение Больцмана

Транспортное уравнение Больцмана (ТУБ) в наноэлектронике [1] решает те же самые задачи, что и обобщенная транспортная модель Ландауэра – Датты – Лундстрома (ЛДЛ) [2]. Для одних задач лучше пользоваться ТУБ, а для других – моделью ЛДЛ. При правильной постановке вычислений оба подхода приводят, естественно, к одним и тем же результатам. Настоящая глава посвящена в первую очередь ТУБ и его роли в концепции ЛДЛ. В качестве приложения ТУБ рассмотрим поведение тока во внешнем магнитном поле.

Основной задачей теории транспортных явлений Больцмана [3] является вычисление функции распределения $f(\vec{r},\vec{k},t)$ как вероятности того, что состояние в точке $\vec{r}$ с волновым вектором $\vec{k}$ (или импульсом $\vec{p}=\hbar\vec{k}$) занято в момент времени t. В состоянии равновесия ответ известен: это – функция Ферми $f_0(E)$. Нас же в этом обзоре будут интересовать следующие вопросы: как составить уравнение для $f(\vec{r},\vec{k},t)$ за пределами равновесия, как решить это уравнение в режиме линейного отклика, другими словами, вблизи равновесия, как увязать полученные таким образом результаты с теми, которые получаются в модели ЛДЛ в диффузионном режиме транспорта, как учесть внешнее магнитное поле и его влияние на электронный транспорт. Более подробное рассмотрение этих вопросов можно найти в [4 – 8].

5.1. Транспортное уравнение Больцмана

Электроны в кристалле с медленно меняющимся потенциалом могут рассматриваются как квазичастицы. Ньютоновское уравнение движения

$$\frac{d\vec{p}}{dt}=\frac{d(\hbar\vec{k})}{dt}=\vec{F}_e, \tag{1}$$

где сила, действующая на электрон,

$$\vec{F}_e=-\nabla E_c-q\vec{v}\times\vec{B}=-q\vec{E}-q\vec{v}\times\vec{B}. \tag{2}$$

Для нахождения $\vec{k}(t)$ нужно вычислить

$$\hbar\vec{k}(t)=\hbar\vec{k}(0)+\int_0^t\vec{F}_e\left(\vec{r}(t'),t'\right)dt'. \tag{3}$$

Зная $\vec{k}(t)$, по зонной структуре определяем скорость электронов

$$\vec{v}_g(t)=\frac{1}{\hbar}\nabla_k E\left(\vec{k}(t)\right) \tag{4}$$

и далее их траекторию

$$\vec{r}(t) = \vec{r}(0) + \int_0^t \vec{v}_g(t')dt'. \qquad (5)$$

Уравнения (3) и (5) есть полуклассические уравнения движения электронов в кристаллическом твердом теле. В полуклассическом подходе предполагается, что потенциал $E_c(\vec{r})$ меняется медленно, так что отсутствуют квантовая интерференция и туннелирование и электроны можно рассматривать как частицы. Положение электрона и его импульс одновременно определены, так что квантовомеханическая неопределенность $\Delta r \Delta p \geq \hbar/2$ мала.

Уравнения движения описывают положение электрона в фазовом пространстве. На рис. 1 показана произвольная траектория $T[x(t), p_x(t)]$ движения электрона в 2D-фазовом пространстве (x, p_x).

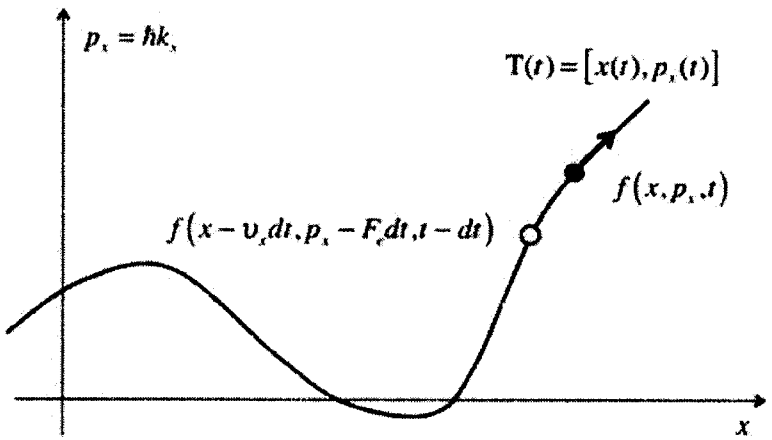


Рис.1. Траектория движения электрона $T(t) = [x(t), p_x(t)]$ в фазовом пространстве (x, p_x). Определить нужно вероятность $f(x, p_x, t)$ того, что состояние, обозначенное черным кружком, занято в момент времени t. Эта вероятность должна совпадать с вероятностью того, что предыдущее состояние, обозначенное светлым кружком, также занято в момент времени $t - dt$.

Черным кружком на траектории показано состояние, для которого мы хотим определить вероятность $f(x, p_x, t)$ того, что оно занято. Из непрерывности движения электрона вдоль траектории следует, что эта вероятность должна совпадать с такой же вероятностью для предшествующего состояния, показанного светлым кружком, а именно:

$$f(x, p_x, t) = f(x - v_x dt, p_x - F_e dt, t - dt), \qquad (6)$$

другими словами, полная производная вдоль траектории равна нулю

$$\frac{df}{dt}=0. \tag{7}$$

Взяв производную от функции нескольких переменных (цепное правило), получаем

$$\frac{df}{dt}=\frac{\partial f}{\partial t}+\frac{\partial f}{\partial x}\frac{\partial x}{\partial t}+\frac{\partial f}{\partial p_x}\frac{\partial p_x}{\partial t}=0, \tag{8}$$

или иначе

$$\frac{\partial f}{\partial t}+\frac{\partial f}{\partial x}v_x+\frac{\partial f}{\partial p_x}F_{e,x}=0. \tag{9}$$

В 3D-фазовом пространстве, очевидно, имеем

$$\frac{\partial f}{\partial t}+\vec{v}\cdot\nabla_r f+\vec{F}_e\cdot\nabla_p f=0, \tag{10}$$

где

$$\begin{aligned}\nabla_r f&=\frac{\partial f}{\partial x}\hat{x}+\frac{\partial f}{\partial y}\hat{y}+\frac{\partial f}{\partial z}\hat{z},\\ \nabla_p f&=\frac{\partial f}{\partial p_x}\hat{p}_x+\frac{\partial f}{\partial p_y}\hat{p}_y+\frac{\partial f}{\partial p_z}\hat{p}_z.\end{aligned} \tag{11}$$

Уравнение (10) иногда называют бесстолкновительным уравнением Больцмана. Это уравнение описывает баллистический транспорт или же состояние равновесия, в котором каждый акт столкновения компенсируется ему противоположным актом в соответствии с принципом детального равновесия. Поясним последнее утверждение.

Процессы рассеяния увеличивают или уменьшают плотность электронов, приходящуюся на инфинитезимальный объем $d\vec{k}$. Вводят величины

$$\left(\frac{\partial f(\vec{r},\vec{k},t)}{\partial t}\right)_{in} \text{ и } \left(\frac{\partial f(\vec{r},\vec{k},t)}{\partial t}\right)_{out}$$

такие, что выражения

$$\left(\frac{\partial f(\vec{r},\vec{k},t)}{\partial t}\right)_{in}\frac{d\vec{k}}{(2\pi)^3}dt \text{ и } \left(\frac{\partial f(\vec{r},\vec{k},t)}{\partial t}\right)_{out}\frac{d\vec{k}}{(2\pi)^3}dt$$

дают прирост (*in*) или отток (*out*) числа электронов на единицу объема в элементе объема $d\vec{k}$, охватывающего значение $\vec{k}$ в инфинитезимальный временной интервал dt. В состоянии динамического равновесия прирост и отток уравниваются. При полуклассическом подходе к транспорту электронов предполагается, что рассеяние происходит под действием короткодействующих сил настолько быстро, что положение электрона не изменяется в процессе столкновения, изменяется лишь импульс электрона.

При учете рассеяния ТУБ принимает вид

$$\frac{\partial f}{\partial t}+\vec{v}\cdot\nabla_r f+\vec{F}_e\cdot\nabla_p f=\hat{S}f, \tag{12}$$

в котором выбор оператора рассеяния $\hat{S}$ определяется постановкой задачи и может быть весьма сложным. При полуклассическом рассмотрении транспорта электронов обычно используется приближение времени релаксации (ВР) [6]

$$\hat{S}f=-\left(\frac{f(\vec{p})-f_0(\vec{p})}{\tau}\right)=-\frac{\delta f(\vec{p})}{\tau}, \tag{13}$$

где δf есть отклонение функции распределения от ее равновесного значения, а τ есть характеристическое время релаксации импульса [4].

Почувствовать приближение ВР можно на таком простом примере. Пусть мы рассматриваем пространственно однородный материал в отсутствии электрических или магнитных полей. Тогда второе и третье слагаемые в ТУБ (12) зануляются и остается уравнение

$$\frac{\partial f}{\partial t}=-\frac{\delta f}{\tau}, \tag{14}$$

решением которого является

$$\delta f(t)=\delta f(0)e^{-t/\tau}, \tag{15}$$

согласно которому нарушения в системе, вызванные отклонением от равновесного состояния, распадаются экспоненциально с характеристическим временем τ. Рассеяние восстанавливает равновесие.

Приближение ВР вполне приемлемо для описания квазиравновесного транспорта в режиме линейного отклика, однако, подобное приближение для интеграла столкновений может быть обосновано лишь вблизи равновесия и только для упругих и/или изотропных процессов рассеяния [4]. Этим приближением широко пользуются для получения результатов в аналитическом виде. Будем далее пользоваться им и мы. Нужно, однако, помнить, что есть немало ситуаций, когда приближение ВР не работает (см., например, [9]).

5.2. Решение уравнения Больцмана в режиме динамического равновесия

В качестве примера решения ТУБ рассмотрим случай динамического равновесия. Внешнее магнитное поле пока не будем учитывать. В приближении ВР для системы в состоянии динамического равновесия ($\partial f/\partial t = 0$) имеем:

$$\vec{v}\cdot\nabla_r f - q\vec{E}\cdot\nabla_p f = -\frac{\delta f}{\tau}. \tag{16}$$

Решение сводится к равновесному значению $f_0(\vec{r},\vec{p},t)$ с учетом малой поправки $\delta f(\vec{r},\vec{p},t)$. Естественно предположить, что

$$\begin{aligned} \nabla_r f &\approx \nabla_r f_0, \\ \nabla_p f &\approx \nabla_p f_0. \end{aligned} \tag{17}$$

Тогда для поправки δf из уравнения (16) имеем:

$$\delta f = -\tau\left(\vec{v}\cdot\nabla_r f_0 - q\vec{E}\cdot\nabla_p f_0\right). \tag{18}$$

Преобразуем правую часть решения (18) следующим образом. Равновесную функцию Ферми запишем в виде

$$\begin{aligned} f_0(\vec{p}) &= \frac{1}{e^{\Theta}+1}, \\ \Theta &= \frac{E(\vec{r},\vec{p}) - E_F(\vec{r})}{kT} = \frac{E_c(\vec{r}) + E(\vec{p}) - E_F(\vec{r})}{kT}, \end{aligned} \tag{19}$$

где $E_F(\vec{r})$ – электрохимический потенциал. Для градиентов равновесной функции Ферми имеем:

$$\begin{aligned} \nabla_r f_0 &= \frac{\partial f_0}{\partial\Theta}\nabla_r\Theta, \\ \nabla_p f_0 &= \frac{\partial f_0}{\partial\Theta}\nabla_p\Theta; \end{aligned} \tag{20}$$

$$\frac{\partial f_0}{\partial\Theta} = kT\frac{\partial f_0}{\partial E}. \tag{21}$$

Подставляя (20) и (21) в (18), получаем:

$$\delta f = \tau kT\left(-\frac{\partial f_0}{\partial E}\right)\left(\vec{v}\cdot\nabla_r\Theta - q\vec{E}\cdot\nabla_p\Theta\right). \tag{22}$$

С учетом

$$\nabla_r\Theta=\frac{1}{kT}\nabla_r\left(E_c(\vec{r})-E_F(\vec{r})\right),\quad -\nabla_r E_c(\vec{r})=-q\vec{E}(\vec{r}),\quad \nabla_p\Theta=\frac{\vec{v}}{kT},$$

квазиравновесное решение ТУБ (22) окончательно можно записать следующим образом:

$$\delta f=\tau\left(-\frac{\partial f_0}{\partial E}\right)\left(\vec{v}\cdot\vec{F}_g\right),$$
$$\vec{F}_g=-\nabla_r E_F(\vec{r})+T\left[E_c(\vec{r})+E(\vec{p})-E_F(\vec{r})\right]\nabla_r\left(\frac{1}{T}\right). \tag{23}$$

Силу $\vec{F}_g$ называют обобщенной силой; именно она выводит систему из состояния равновесия. Обобщенная сила содержит две компоненты: градиент электрохимического потенциала и градиент (обратной) температуры. Именно эти две причины в транспортной модели ЛДЛ выводят систему (проводник) из состояния равновесия и обеспечивают возникновение электрического тока [2, 10]. В модели ЛДЛ это очевидный физический результат, подтвержденный теперь математическим решением ТУБ в квазиравновесном режиме в приближении ВР.

5.3. Вычисление транспортных коэффициентов

Располагая решением (23) ТУБ в квазиравновесном режиме, можно рассчитать различные свойства проводника. К примеру, для 2D проводника поверхностная плотность электронов

$$n_S(\vec{r})=\frac{1}{A}\sum_{\vec{k}}\left[f_0(\vec{r},\vec{k})+\delta f(\vec{r},\vec{k})\right]=\frac{1}{A}\sum_{\vec{k}}f_0(\vec{r},\vec{k}), \tag{24}$$

где окончательный результат есть следствие того, что функция $\delta f(\vec{r},\vec{k})$ является нечетной по импульсу и при интегрировании в k-пространстве зануляется. Суммирование в k-пространстве подробно рассматривается во второй лекции курса М. Лундстрома «Электронный транспорт в полупроводниках» [11].

В случае плотности тока в 2D проводнике нужно вычислить только

$$\vec{J}(\vec{r})=\frac{1}{A}\sum_{\vec{k}}(-q)\vec{v}(\vec{k})\delta f(\vec{k}), \tag{25}$$

поскольку существенна только поправка δf, а произведение $\vec{v}f_0$ нечетное в k-пространстве и при интегрировании зануляется, произведение же $\vec{v}\delta f$ в (25) четное и дает вклад в плотность тока.

Плотность энергии электрического тока дается выражением

$$\vec{J}_W(\vec{r}) = \frac{1}{A}\sum_{\vec{k}} E(\vec{k})\vec{v}(\vec{k})\delta f(\vec{k}), \tag{26}$$

а термоток

$$\vec{J}_Q(\vec{r}) = \frac{1}{A}\sum_{\vec{k}} (E - E_F)\vec{v}(\vec{k})\delta f(\vec{k}), \tag{27}$$

где суммарная энергия электрона $E = E_c(\vec{r}) + E(\vec{k})$. Разность $E - E_F$ есть энергия, которую должен поглотить электрон, находящийся в контакте при энергии $E_F < E$, чтобы войти в моду проводника при энергии E [12].

Если вычислять зарядовый ток и термоток опираясь на решение ТУБ (23), для каждого из токов получим двухчленное выражение, поскольку обобщенная сила в (23) дает два вклада – за счет разности электрохимических потенциалов и за счет разности температур контактов. В итоге получаются четыре термоэлектрических коэффициента. Вычислим один из них – поверхностную проводимость σ_S.

Итак, вычисляем зарядовый ток (25) с учетом (23):

$$\vec{J}(\vec{r}) = \frac{(-q)}{A}\sum_{\vec{k}} \tau\left(-\frac{\partial f_0}{\partial E}\right)\vec{v}(\vec{v}\cdot\vec{F}_g) = \frac{(-q)}{A}\sum_{\vec{k}} \tau\left(-\frac{\partial f_0}{\partial E}\right)(\vec{v}\vec{v})\cdot\vec{F}_g, \tag{28}$$

где произведение $(\vec{v}\vec{v})$ в общем случае есть тензор, так что ТУБ позволяет относительно просто рассматривать анизотропный транспорт, для описания которого потребуются тензорные выражения для транспортных коэффициентов. Упрощая ситуацию, далее рассмотрим ток лишь в направлении x в предположении, что электрохимический потенциал изменяется только в этом направлении и температура вдоль проводника постоянна, а именно:

$$\vec{J}_x(\vec{r}) = \left[\frac{1}{A}\sum_{\vec{k}} q v_x^2 \tau\left(-\frac{\partial f_0}{\partial E}\right)\right]\frac{dE_F}{dx}, \tag{29}$$

что можно переписать в стандартной форме

$$J_x = \sigma_S \frac{d(E_F/q)}{dx}, \tag{30}$$

где

$$\sigma_S = \frac{1}{A}\sum_{\vec{k}} q^2 v_x^2 \tau\left(-\frac{\partial f_0}{\partial E}\right). \tag{31}$$

Для вычисления удельной поверхностной проводимости нужно отсуммировать в k-пространстве.

При конечных размерах проводника число электронных состояний в проводнике счетно. Если речь идет о нанопроводнике, число состояний невелико и суммирование в (31) можно выполнить непосредственно. В достаточно объемных проводниках состояния располагаются плотно и суммирование в (31) можно заменить интегрированием. Рецепт простой:

$$\sum_{\vec{k}}(\bullet)\to\int(\bullet)N_k d\vec{k}\,, \tag{32}$$

где N_k есть плотность состояний в k-пространстве.

Для проводников разной размерности окончательно имеем [4, 11]:

$$\begin{aligned} 1D:&\ N_k d\vec{k}=2\times\left(\frac{L}{2\pi}\right)dk=\frac{L}{\pi}dk,\\ 2D:&\ N_k d\vec{k}=2\times\left(\frac{A}{4\pi^2}\right)dk_x dk_y=\frac{A}{2\pi^2}dk_x dk_y,\\ 3D:&\ N_k d\vec{k}=2\times\left(\frac{\Omega}{8\pi^3}\right)dk_x dk_y dk_z=\frac{\Omega}{4\pi^3}dk_x dk_y dk_z, \end{aligned} \tag{33}$$

где общая двойка учитывает вырождение по спину. Плотность состояний N_k не зависит от зонной структуры.

Вместо (31) с учетом долинного вырождения g_v имеем:

$$\sigma_S=\frac{1}{A}g_v\frac{A}{2\pi^2}\int_0^{\infty}\int_0^{2\pi}q^2v_x^2\tau\left(-\frac{\partial f_0}{\partial E}\right)d\theta\,kdk\,. \tag{34}$$

После взятия интеграла по углу с $v_x=v\cos\theta$ имеем:

$$\sigma_S=g_v\frac{q^2}{2\pi}\int_0^{\infty}v^2\tau(k)\left(-\frac{\partial f_0}{\partial E}\right)kdk\,. \tag{35}$$

Для параболической дисперсии

$$kdk=\frac{m^*}{\hbar^2}dE\ \text{ и }\ v^2=\frac{2(E-E_c)}{m^*}.$$

Пусть еще, упрощения ради, характеристическое время рассеяния постоянно: $\tau(E)=\tau_0$. Тогда вместо (35) имеем:

$$\sigma_S=g_v\frac{q^2\tau_0}{\pi\hbar^2}\int_0^{\infty}(E-E_c)\left(-\frac{\partial f_0}{\partial E}\right)dE\,, \tag{36}$$

что после взятия интеграла дает

$$\sigma_S = g_v \frac{q^2 \tau_0 kT}{\pi \hbar^2} \Im_0(\eta_F), \tag{37}$$

где интеграл Ферми – Дирака порядка j (§ 1.3.6.4)

$$\Im_j(\eta_F) = \frac{1}{\Gamma(j+1)} \int_0^\infty \frac{\eta^j}{\exp(\eta - \eta_F) + 1} d\eta,$$

с $\eta_F = (E_F - E_C)/kT$.

Если вспомнить ф-лу (127/гл.1) [13], что поверхностная плотность электронов

$$n_S = \int_0^\infty D_{2D}(E) f_0(E) dE = g_v \frac{m^* kT}{\pi \hbar^2} \Im_0(\eta_F) = N_{2D} \Im_0(\eta_F),$$

то удельная поверхностная проводимость (37) выразится привычным образом:

$$\sigma_S = n_S q \left(\frac{q\tau_0}{m^*} \right) = n_S q \mu. \tag{39}$$

Однако, как общеизвестный результат соотносится с моделью ЛДЛ? Для ответа на этот вопрос нужно вернуться к уравнению (35) и перейти в нем к интегрированию по энергии:

$$\sigma_S = \frac{q^2}{2\pi} \int_0^\infty v^2 \tau(E) \left(-\frac{\partial f_0}{\partial E} \right) \left(\frac{g_v m^*}{\hbar^2} \right) dE. \tag{40}$$

Поскольку плотность 2D-состояний

$$D_{2D}(E) = \frac{g_v m^*}{\pi \hbar^2},$$

то вместо (40) имеем:

$$\sigma_S = \frac{q^2}{2} \int_0^\infty (v\tau) v D_{2D}(E) \left(-\frac{\partial f_0}{\partial E} \right) dE. \tag{41}$$

Введем множители $\pi/2$ и $h/4$ и им обратные и перекомпонуем (41) следующим образом:

$$\sigma_S = \frac{q^2}{2} \frac{4}{h} \int_0^\infty \left(\frac{\pi}{2} v\tau \right) \left[\left(\frac{2}{\pi} v \right) \frac{h}{4} D_{2D}(E) \right] \left(-\frac{\partial f_0}{\partial E} \right) dE. \tag{42}$$

В подынтегральном выражении первый сомножитель в круглых скобках, согласно (13/гл.4) [14], есть средняя длина свободного обратного рассеяния $\lambda(E)$, сомножитель $2v/\pi$, согласно (60/гл.1), есть средняя скорость $\langle v_x^+(E)\rangle$ при энергии E в направлении $+x$, а все выражение в квадратных скобках, согласно (63/гл.1), есть ни что иное как число мод проводимости $M_{2D}(E)$. Таким образом, полученное выражение (42) есть хорошо известное в модели ЛДЛ выражение для поверхностной проводимости

$$\sigma_S = \frac{2q^2}{h}\int \lambda(E) M_{2D}(E)\left(-\frac{\partial f_0}{\partial E}\right)dE. \qquad (43)$$

Итак, результатом решения ТУБ в квазиравновесном режиме в приближении ВР, или иначе в диффузионном приближении, для поверхностной проводимости получается привычное для модели ЛДЛ выражение. Показано также [15], что ТУБ в приближении ВР дает точно такие же выражения для коэффициента Зеебека и для электронной теплопроводности, что и транспортная модель ЛДЛ. Преимущество модели ЛДЛ в ее физической прозрачности, а также в том, что она позволяет рассматривать квазибаллистический и баллистический режимы транспорта так же просто, как и диффузионный режим. С другой стороны, без ТУБ не обойтись при изучении анизотропного транспорта. Кроме того, легко учесть магнитное поле в ТУБ, как это будет показано ниже. Не говоря уже о том, что ТУБ позволяет рассматривать транспорт вдали от равновесия [4].

Прежде чем перейти к учету магнитного поля в ТУБ, получим выражение для усредненного времени рассеяния $\langle\langle\tau\rangle\rangle$.

Если при переходе от (35) к (36) сохранить зависимость времени рассеяния от энергии, то

$$\sigma_S = g_v \frac{q^2}{\pi\hbar^2}\int_0^\infty \tau(E)(E-E_c)\left(-\frac{\partial f_0}{\partial E}\right)dE, \qquad (44)$$

а самый простой путь учесть зависимость времени рассеяния от энергии это выбрать для этой зависимости степенной закон

$$\tau(E) = \tau_0\left(\frac{E-E_c}{kT}\right)^s. \qquad (45)$$

Поверхностная проводимость пропорциональна поверхностной плотности электронов

$$n_S = \int_0^\infty D_{2D}(E) f_0(E) dE = \int_0^\infty \left(g_v \frac{m^*}{\pi\hbar^2}\right) f_0(E) dE.$$

Умножая и деля σ_S по (44) на n_S, получаем

$$\sigma_S = q^2 \frac{\left(\frac{g_v}{\pi\hbar^2}\right)\int_0^\infty \tau(E)(E-E_c)\left(-\frac{\partial f_0}{\partial E}\right)dE}{m^*\int_0^\infty \left(\frac{g_v}{\pi\hbar^2}\right) f_0(E)dE} n_S, \tag{46}$$

что позволяет записать поверхностную проводимость в виде

$$\sigma_S = n_S q \frac{q\langle\langle\tau\rangle\rangle}{m^*}, \tag{47}$$

где должным образом усредненное время рассеяния

$$\langle\langle\tau\rangle\rangle \equiv \frac{\int_0^\infty \tau(E)(E-E_c)\left(-\frac{\partial f_0}{\partial E}\right)dE}{\int_0^\infty f_0(E)dE}. \tag{48}$$

В предположении степенного закона рассеяния (45)

$$\langle\langle\tau\rangle\rangle = \tau_0 \frac{\int_0^\infty \left(\frac{E-E_c}{kT}\right)^s (E-E_c)\left(-\frac{\partial f_0}{\partial E}\right)dE}{\int_0^\infty f_0(E)dE}. \tag{49}$$

Интегрирование знаменателя дает $kT\Im_0(\eta_F)$, а числителя – $kT\Gamma(s+2)\Im_s(\eta_F)$, так что окончательно

$$\langle\langle\tau\rangle\rangle = \tau_0 \frac{\Gamma(s+2)\Im_s(\eta_F)}{\Im_0(\eta_F)}. \tag{50}$$

Точно так же выражения для $\langle\langle\tau\rangle\rangle$ можно получить для 1D и 3D проводников. Для невырожденных полупроводников интегралы Ферми – Дирака обращаются в экспоненты, так что для таких проводников

$$\langle\langle\tau\rangle\rangle = \tau_0\Gamma(s+2). \tag{51}$$

При рассеянии на акустических фононах в 2D проводнике $s=0$, а на заряженных примесях $s\approx 1$. Таким образом, для 2D проводников с такими режимами рассеяния $\langle\langle\tau\rangle\rangle$ превышает τ_0 в $1-2$ раза.

5.4. Учет магнитного поля

Под влиянием магнитного поля транспортные коэффициенты изменяются. Измерение эффекта Холла позволяет полнее охарактеризовать свойства материала проводника [16]. Решение ТУБ с учетом магнитного поля в общем случае вполне достижимо, хотя и требует подчас значительных вычислительных ресурсов. Мы же рассмотрим частный случай измерения проводимости в 2D проводнике, ориентированном в плоскости xy с магнитным полем $\vec{B}$, направленным перпендикулярно вдоль оси z.

Начнем с приближения ВР для системы в состоянии динамического равновесия ($\partial f/\partial t = 0$) и, простоты ради, предположим пространственную однородность проводника ($\nabla_r f = 0$). Тогда ТУБ (16) с учетом магнитного поля будет:

$$-q\vec{E}\cdot\nabla_p f - q\left[\vec{v}\times\vec{B}\right]\cdot\nabla_p f = -\frac{\delta f}{\tau}. \tag{52}$$

Сразу хочется допустить, что

$$\nabla_p f \approx \nabla_p f_0, \tag{53}$$

как это предполагалось ранее в (17), но теперь это невозможно. Приближение (53) вполне пригодно, как мы видели выше, для первого слагаемого с электрическим полем в левой части (52), но не приемлемо для второго слагаемого с магнитным полем. Причина кроется в том, что

$$\nabla_p f_0 = \frac{\partial f_0}{\partial E}\nabla_p E = \frac{\partial f_0}{\partial E}\vec{v}, \tag{54}$$

и будучи подставленным во второе слагаемое с магнитным полем, получаем $\left[\vec{v}\times\vec{B}\right]\cdot\vec{v} = 0$. Чтобы получить решение ТУБ с учетом магнитного поля, уравнение (52) нужно переписати в виде

$$-q\vec{E}\cdot\nabla_p f_0 - q\left[\vec{v}\times\vec{B}\right]\cdot\nabla_p(\delta f) = -\frac{\delta f}{\tau}, \tag{55}$$

решить которое, правда, труднее, поскольку δf входит в обе части уравнения. Решить его можно параметрически. Пусть решение имеет вид (23), но обощенную силу $\vec{F}_g$ в нем заменим на неизвестный вектор $\vec{G}$, а именно:

$$\delta f = \tau\left(-\frac{\partial f_0}{\partial E}\right)\left(\vec{v}\cdot\vec{G}\right), \tag{56}$$

причем искомый вектор $\vec{G}$ не зависит от импульса $\vec{p}$. После подстановки этого пробного решения в (55) нужно вычислять $\nabla_p(\delta f)$. Учитывая (54) и рассматривая параболическую зонную структуру ($\nabla_p \vec{v} = 1/m^*$), находим, что

$$\nabla_p(\delta f) = \tau\left(-\frac{\partial f_0}{\partial E}\right)\frac{\vec{G}}{m^*}. \tag{57}$$

Остается подставить (54), (56) и (57) в (55) и получим:

$$q\vec{E}\cdot\vec{v} - \frac{q\tau}{m^*}\left(\vec{v}\times\vec{B}\cdot\vec{G}\right) + \vec{v}\cdot\vec{G} = 0. \tag{58}$$

Пользуясь свойством тройного векторного произведения $\left(\vec{v}\times\vec{B}\cdot\vec{G}\right) = \left(\vec{v}\cdot\vec{B}\times\vec{G}\right)$, переписываем (58) в виде

$$\vec{v}\cdot\left(-q\vec{E} + \frac{q\tau}{m^*}\left[\vec{B}\times\vec{G}\right] - \vec{G}\right) = 0. \tag{59}$$

Поскольку (59) должно удовлетворяться для любого $\vec{v}$, то выражение в круглых скобках должно быть равно нулю, а именно:

$$\vec{G} = -q\vec{E} + \frac{q\tau}{m^*}\left[\vec{B}\times\vec{G}\right]. \tag{60}$$

Для векторного уравнения $\vec{c} = \vec{a} + \vec{b}\times\vec{c}$ решением является

$$\vec{c} = \frac{\vec{a} + [\vec{b}\times\vec{a}] + (\vec{a}\cdot\vec{b})\vec{b}}{1+b^2}. \tag{61}$$

Тогда решением уравнения (60) будет

$$\vec{G} = \frac{-q\vec{E} - (q^2\tau/m^*)\left[\vec{B}\times\vec{E}\right] - q(q\tau/m^*)^2\left(\vec{E}\cdot\vec{B}\right)\vec{B}}{1+(\omega_c\tau)^2}, \tag{62}$$

где циклотронная частота

$$\omega_c = \frac{qB}{m^*}, \tag{63}$$

а плоскость круговой орбиты электрона перпендикулярна вектору магнитного поля.

Первое слагаемое в числителе дает вклад в проводимость, второе слагаемое ответственно за эффект Холла, третье слагаемое как и знаменатель пропорциональны B^2 и определяют магнитосопротивление.

Рассмотрим стандартную ситуацию при измерении эффекта Холла, когда $\vec{B} \perp \vec{E}$. В этой конфигурации полей третье слагаемое зануляется. Будем предполагать также, что магнитное поле слабое, чему соответствует

$$\omega_c \tau \ll 1. \tag{64}$$

В слабых магнитных полях электрон испытывает много столкновений прежде чем замкнет свою круговую орбиту. Для слабых магнитных полей имеем

$$\vec{G} = -q\vec{E} - (q^2\tau/m^*)\left[\vec{B} \times \vec{E}\right]. \tag{65}$$

Вектор $\vec{G} \perp \vec{B}, \vec{E}$, лежит в плоскости xy с проекциями на оси

$$\begin{aligned} G_x &= -qE_x + \left(\frac{q^2\tau}{m^*}\right) E_y B_z, \\ G_y &= -qE_y - \left(\frac{q^2\tau}{m^*}\right) B_z E_x. \end{aligned} \tag{66}$$

Итак, решение ТУБ при наличии магнитного поля в приближении ВР получено. Решением является уравнение (56) с вектором $\vec{G}$ по (62) или же уравнение (65) при $\vec{G} \perp \vec{B}, \vec{E}$. В отсутствии магнитного поля решение дается уравнением (23). Предполагалось, что градиенты плотности электронов и температуры отсутствуют. Если нужно учесть изменение концентрации электронов по длине проводника или перепад температуры, следует $-q\vec{E}$ заменить обобщенной силой $\vec{F}_g$ (23).

Теперь рассмотрим как будут меняться транспортные коэффициенты в присутствии магнитного поля. Для этого полученные решения нужно подставить в (25) и (27) и посмотреть как изменится зарядовый ток и термоток. Проще всего рассмотреть эту ситуацию на примере проводимости.

Заменим обобщенную силу в уравнении (28) на некий вектор $\vec{G}$:

$$\vec{J}(\vec{r}) = \frac{(-q)}{A} \sum_{\vec{k}} \tau \left(-\frac{\partial f_0}{\partial E}\right) \vec{v}(\vec{v} \cdot \vec{G}). \tag{67}$$

Проекции вектора плотности тока получаются следующими:

$$\begin{aligned} J_x &= \frac{(-q)}{A} \sum_{\vec{k}} \tau \left(-\frac{\partial f_0}{\partial E}\right) v_x (v_x G_x + v_y G_y), \\ J_y &= \frac{(-q)}{A} \sum_{\vec{k}} \tau \left(-\frac{\partial f_0}{\partial E}\right) v_y (v_x G_x + v_y G_y). \end{aligned} \tag{68}$$

Слагаемые $v_x v_y = v^2 \cos\theta \sin\theta$ после интегрирования от 0 до 2π дадут ноль, так что вместо (68) запишем

$$J_x = \frac{(-q)}{A}\sum_{\vec{k}} \tau\left(-\frac{\partial f_0}{\partial E}\right) v_x^2 G_x,$$
$$J_y = \frac{(-q)}{A}\sum_{\vec{k}} \tau\left(-\frac{\partial f_0}{\partial E}\right) v_y^2 G_y. \qquad (69)$$

В состоянии равновесия $v_x^2 + v_y^2 = v^2$, а после усреднения по углам $v_x^2 = v_y^2 = v^2/2$. Для параболической дисперсии $m^* v^2/2 = E - E_c$, так что v_x^2 и v_y^2 в (69) можно заменить на $(E - E_c)/m^*$:

$$J_x = \frac{(-q)}{A}\sum_{\vec{k}} \frac{(E-E_c)\tau(E)}{m^*}\left(-\frac{\partial f_0}{\partial E}\right) G_x,$$
$$J_y = \frac{(-q)}{A}\sum_{\vec{k}} \frac{(E-E_c)\tau(E)}{m^*}\left(-\frac{\partial f_0}{\partial E}\right) G_y. \qquad (70)$$

Используя выражения для проекций вектора $\vec{G}$ (66), имеем:

$$J_x = \frac{1}{A}\sum_{\vec{k}} \frac{(E-E_c)\tau(E)}{m^*}\left(-\frac{\partial f_0}{\partial E}\right)[q^2 E_x - (q^3\tau/m^*)E_y B_z],$$
$$J_y = \frac{1}{A}\sum_{\vec{k}} \frac{(E-E_c)\tau(E)}{m^*}\left(-\frac{\partial f_0}{\partial E}\right)[q^2 E_y + (q^3\tau/m^*)B_z E_x]. \qquad (71)$$

Окончательные выражения для компонент плотности тока получаем делением правых частей (71) на

$$n_S = \frac{1}{A}\sum_{\vec{k}} f_0(E) \qquad (72)$$

с одновременным умножением их на n_S, так что в конечном итоге базовые токовые уравнения таковы:

$$J_x = \sigma_S E_x - \sigma_S \mu_H E_y B_z,$$
$$J_y = \sigma_S E_y + \sigma_S \mu_H B_z E_x, \qquad (73а)$$

или в матричной форме

$$\begin{pmatrix} J_x \\ J_y \end{pmatrix} = \begin{bmatrix} \sigma_S & -\sigma_S \mu_H B_z \\ +\sigma_S \mu_H B_z & \sigma_S \end{bmatrix} \begin{pmatrix} E_x \\ E_y \end{pmatrix} \qquad (73б)$$

и в векторном обозначении

$$\vec{J} = \sigma_S \vec{E} - \sigma_S \mu_H \left[\vec{E} \times \vec{B} \right]. \qquad (73в)$$

Позже нам понадобится запись уравнений (73) в индексных обозначениях:

$$J_i = \sum_j \sigma_{ij}(B_z) E_j, \qquad (73г)$$

где индексы i и j пробегают все координатные оси x, y, z или 1, 2, 3. Тензор проводимости σ_{ij} представим матрицей 2×2 (73б). Запись (73г) часто пишут без значка суммирования

$$J_i = \sigma_{ij}(B_z) E_j, \qquad (73д)$$

имея в виду, что суммирование ведется по повторяющемуся индексу.

Уравнение (73в) записано в векторной форме. В индексных же обозначениях

$$J_i = \sigma_S E_i - \sigma_S \mu_H \epsilon_{ijk} B_k E_j, \qquad (73е)$$

где ϵ_{ijk} есть символ Леви-Чивита, иначе перестановочный единичный тензор

$$\epsilon_{ijk} = \begin{cases} 1, & \text{если } i, j, k \text{ все различны и переставляются в циклическом порядке } x, y, z, \\ -1, & \text{если } i, j, k \text{ все различны и переставляются в ациклическом порядке } x, z, y, \\ 0 & \text{в противном случае.} \end{cases}$$

В уравнениях (73)

$$\sigma_S = n_S q \mu \qquad (74)$$

и

$$\mu = \frac{q \left\langle \left\langle \tau \right\rangle \right\rangle}{m^*} \qquad (75)$$

с

$$\left\langle \left\langle \tau \right\rangle \right\rangle \equiv \frac{\sum\limits_{\vec{k}} (E - E_c) \tau(E) \left(-\frac{\partial f_0}{\partial E} \right)}{\sum\limits_{\vec{k}} f_0(E)}, \qquad (76)$$

что то же самое, что и уравнение (48).

Холловская подвижность в (73) определяется в виде

$$\mu_H = \mu r_H \qquad (77)$$

через холловский фактор, играющий важную роль в холловских измерениях,

$$r_H \equiv \frac{\left\langle\left\langle \tau^2 \right\rangle\right\rangle}{\left\langle\left\langle \tau \right\rangle\right\rangle^2} \tag{78}$$

с

$$\left\langle\left\langle \tau^2 \right\rangle\right\rangle \equiv \frac{\sum_k (E - E_c)\tau^2(E)\left(-\frac{\partial f_0}{\partial E}\right)}{\sum_k f_0(E)}. \tag{79}$$

Уравнения (73) и есть искомые уравнения для тока в плоском образце в присутствии слабого магнитного поля, направленного перпендикулярно к поверхности образца. Из этих уравнений следует, что в отсутствии магнитного поля электрическое поле в направлении x порождает ток в этом же направлении, а в присутствии магнитного поля, перпендикулярного поверхности образца, ток порождается также и в направлении y. В присутствии магнитного поля электрическое же поле в направлении y порождает ток и в направлении x и в направлении y, проводимость приобретает тензорный характер.

Аналогичным образом можно получить выражения для коэффициента Зеебека и для электронной теплопроводности, и эти оба термические коэффициенты оказывается также имеют тензорную природу [17].

Сделаем одно замечание относительно холловского фактора применительно к 2D проводникам со степенным законом рассеяния.

Для холловского фактора можно получить простое выражение, если рассеяние описывается степенным законом (45). Нам нужно вычислить (79). Поскольку

$$\tau^2(E) = \tau_0^2 \left(\frac{E - E_c}{kT}\right)^{2s}, \tag{80}$$

то при вычислении (79) результат окажется таким же, как и (51) только лишь с заменой показателя степени s на $2s$, а именно:

$$\left\langle\left\langle \tau^2 \right\rangle\right\rangle = \tau_0^2 \Gamma(2s+2), \tag{81}$$

предполагая, что речь идет о невырожденных полупроводниках.

Теперь холловский фактор (78)

$$r_H \equiv \frac{\left\langle\left\langle \tau^2 \right\rangle\right\rangle}{\left\langle\left\langle \tau \right\rangle\right\rangle^2} = \frac{\Gamma(2s+2)}{\left[\Gamma(s+2)\right]^2}. \tag{82}$$

Поскольку при рассеянии в 2D проводнике на акустических фононах и на заряженных примесях показатель степени s в законе рассеяния (45) изменяется в пределах от 0 до 1, то холловский фактор изменяется от 1 до 1.5. При выполнении холловских измерений механизм рассеяния в образце как правило не известен и холловский фактор обычно полагают $r_H = 1$. Нужно, однако, помнить, что такой подход вносит неопределенность в результаты измерений.

В заключение еще раз отметим, что в отсутствие магнитного поля ТУБ и обобщенная транспортная модель приводят к одинаковым результатам для проводников, работающих в квазиравновесном диффузионном режиме. Модель ЛДЛ привлекательна тем, что она физически прозрачна, а также тем, что транспорт в баллистическом и квазибаллистическом режимах описывается так же просто как и транспорт в диффузионном режиме. Кроме того, обобщенная транспортная модель не предполагает наличия периодической кристаллической структуры материала проводника, и она успешно применяется к аморфным, поликристаллическим и нанокомпозитным материалам. Преимуществом ТУБ является возможность описания анизотропного транспорта и транспорта в магнитном поле. Кроме того транспортное уравнение Больцмана позволяет описывать транспорт вдали от квазиравновесных условий [18, 19].

Литература

1. Ю. О. Кругляк, М. В. Стріха, Уроки наноелектроніки: Роль електростатики й контактів у концепції "знизу – вгору", *Сенсор. електр. мікросист. техн.*, т. 11, № 4, 27 – 42 (2014).
2. Yuriy Kruglyak. Landauer – Datta – Lundstrom Generalized Transport Model for Nanoelectronics, *J. Nanoscience*, v. **2014**, Art. ID 725420; DOI:10.1155/2014/725420.
3. Людвиг Больцман. *Избранные труды*. Москва: Мир, 1984.
4. Mark Lundstrom. *Fundamentals of Carrier Transport*. Cambridge UK: Cambridge University Press, 2000.
5. Mark Lundstrom, Jeong Changwook. *Near-Equilibrium Transport: Fundamentals and Applications*. Hackensack, New Jersey: World Scientific Publishing Company, 2013; www.nanohub.org/resources/11763.
6. F. W. Sears, G. L. Salinger. *Thermodynamics, Kinetic Theory, and Statistical Thermodynamics*. Boston: Addison-Wesley, 1975.
7. Дж. Займан. *Принципы теории твердого тела*. Москва: Высшая школа, 1974.
8. Н. Ашкрофт, Н. Мермин. *Физика твердого тела*. Москва: Мир, 1979.
9. D. I. Pikulin, C.-Y. Hou, C. W. J. Beenakker. Nernst effect beyong the relaxation-time approximation, *Phys. Rev. B*, v. 84, 035133 (2011).
10. Ю. О. Кругляк, Н. Ю. Кругляк, М. В. Стріха. Уроки наноелектроніки: Термоелектричні явища в концепції "знизу – вгору", *Сенсор. електр. мікросист. техн.*, т. 10, № 1, 6 – 21 (2013).
11. Mark Lundstrom. *Electronic Transport in Semiconductors*, 2011; www.nanohub.org/resources/11872.
12. Ю. А. Кругляк. Термоэлектрические явления и устройства в концепции Ландауэра – Датты – Лундстрома, *ScienceRise*, т. 1, № 2(6), 69 – 77 (2015).
13. Ю. О. Кругляк, М. В. Стріха. Узагальнена модель електронного транспорту Ландауера – Датты – Лундстрома, *Сенсор. електр. мікросист. техн.*, т. 12, № 3, 4 – 27 (2015).
14. Ю. А. Кругляк, Учет рассеяния в транспортной модели Ландауэра – Датты – Лундстрома, *ScienceRise*, т. 3, № 2(8), 99 – 107 (2015).
15. C. Jeong, R. Kim, M. Luisier, S. Datta, M. Lundstrom. On Landauer versus Boltzman and full band versus effective mass evaluation of thermoelectric transport coefficients, *J. Appl. Phys.*, v. 107, 023707 (2010).
16. Ю. О. Кругляк, М. В. Стріха. Уроки наноелектроніки: Ефект Холла і вимірювання електрохімічних потенціалів у концепції "знизу – вгору", *Сенсор. електр. мікросист. техн.*, т. 11, № 1, 5 – 27 (2014).
17. C. M. Wolfe, N. Holonyak, G. E. Stillman. *Physical Properties of Semiconductors*. Englewood Cliffs, N. J.: Prentice Hall, 1989.

18. Ю. А. Кругляк. Модель проводимости Ландауэра – Датты – Лундстрома в микро- и наноэлектронике и транспортное уравнение Больцмана, *ScienceRise*, т. 3, № 2(8), 108 – 116 (2015).
19. Ю. А. Кругляк. Кинетическое уравнение Больцмана в наноэлектронике: эффект Холла в графене, *Сб. науч. тр. конф. «Функциональная база наноэлектроники», Харьков – Одесса, 28 сент. – 2 окт.*, 71 – 74 (2015).

Глава 6. Измерение электрофизических свойств

Заключая рассмотрение обобщенной транспортной модели (главы 1 и 2) [1 – 4], остановимся в этой главе на методах измерений проводимости и эффекта Холла в слабых и сильных магнитных полях, в том числе при перепаде температур на контактах проводника, а также на том, как избавиться при экспериментальном измерении от артефактов таких как сопротивление на контактах и от неконтролируемых термоэлектрических эффектов, продемонстрируем также обработку результатов при конкретных измерениях.

Для характеристики нового материала проводника в принципе нужно измерять все четыре термоэлектрических коэффициента (глава 1) – проводимость (сопротивление), коэффициент Зеебека (или Соре), коэффициент Пельтье и электронную теплопроводность. Соотношение Кельвина позволяет измерять либо коэффициент Зеебека, либо коэффициент Пельтье. Техника измерения коэффициента Зеебека требует немалого искусства, обсуждать ее здесь не будем, рекомендуем обратиться к отчету [5]. Не легче измерять и электронную теплопроводность, обычно ее оценивают через закон Видемана – Франца. В этой главе мы сконцентрируемся на измерении проводимости.

Будем рассматривать токи в диффузионном режиме с уравнением для плотности тока

$$J_x = \sigma \frac{d(E_F/q)}{dx}. \tag{1}$$

При однородном распределении плотности электронов $d(E_F/q)/dx = E_x$, так что

$$\begin{aligned} J_x &= \sigma E_x, \\ E_x &= \rho J_x. \end{aligned} \tag{2}$$

Обычно меряют проводимость или сопротивление какого-либо материала в диффузионном режиме, поскольку в этом режиме проводимость зависит от материала проводника, а не от его размерности. Стоит также упомянуть, что уравнения (1) и (2) одинаково пригодны как для *n*-проводников, так и для *p*-проводников.

Уравнения (1) и (2) применимы к 1D, 2D и 3D проводникам, предполагая для простоты пространственную неоднородность только в направлении x. Единицы измерения зависят от размерности образца, а именно:

$$\begin{array}{lllllll} 1D: & E_x & В/м & J_x & А & \rho & Ом/м \\ 2D: & E_x & В/м & J_x & А/м & \rho & Ом \\ 3D: & E_x & В/м & J_x & А/м^2 & \rho & Ом\cdot м \end{array} \tag{3}$$

Для 2D проводника удельное сопротивление ρ обычно называют поверхностным сопротивлением ρ_S и в качестве единицы используют *Ом/□* (Ω/sq: Ohms per square) с единицей измерения в Омах.

Рассмотрим плоский резистор, показанный на рис. 1.

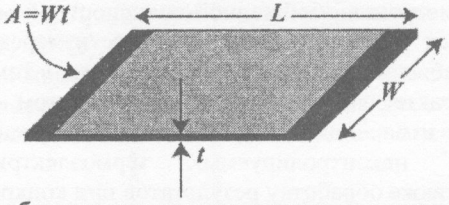


Рис. 1. Принятые обозначения для плоского проводника толщины t. Если толщина проводника больше половины дебройлевской длины волны для электрона, то электроны в проводнике трехмерны, если же толщина t мала по сравнению с половиной длины дебройлевской волны, то электроны в таком проводнике двухмерны.

Для тока имеем:

$$I = GV = \left(\sigma \frac{A}{L}\right)V = \left(\sigma \frac{Wt}{L}\right)V, \tag{4}$$

при этом $\sigma_S = (nt)q\mu = n_S q\mu$ и проводимость

$$G = \sigma_S \left(\frac{W}{L}\right), \tag{5}$$

где $\sigma_S = 1/\rho_S$ есть удельная поверхностная проводимость, а n_S есть поверхностная плотность электронов. Для 3D образца $G = \sigma A/L$, а для 1D проводника $G = \sigma/L$ [4].

Вычисление поверхностной проводимости плоского проводника толщины t зависит от того, ведут ли себя электроны в образце как 3D объекты или как 2D объекты. Если t достаточно велико, так что квантовое удержание является слабым, то электроны нужно рассматривать как 3D частицы и тогда поверхностная проводимость определяется как

$$\sigma_S = \frac{2q^2}{h}\int tM_{3D}(E)\lambda(E)\left(-\frac{\partial f_0}{\partial E}\right)dE. \tag{6}$$

Именно это допущение $(n_S = nt)$ было принято при переходе от (4) к (5), где n есть плотность электронов в 3D образце.

Если же толщина образца t достаточно мала, так что квантовое удержание является сильным, то электроны нужно рассматривать как 2D частицы и тогда поверхностная проводимость [4]

$$\sigma_S = \frac{2q^2}{h}\int M_{2D}(E)\lambda(E)\left(-\frac{\partial f_0}{\partial E}\right)dE. \tag{7}$$

Напомним, что $M_{3D} = M/A$ есть удельное число мод проводимости, равное числу мод проводимости, приходящемуся на единицу площади поперечного сечения 3D проводника, а $M_{2D} = M/W$ есть удельное число мод проводимости, равное числу мод проводимости, приходящемуся на единицу длины поперечного (перпендикулярного направлению тока) сечения 2D проводника. Если режим транспорта в проводнике оказался квазибаллистическим, то среднюю длину свободного пробега $\lambda(E)$ нужно заменить ее кажущимся значением [4].

Как видно из (6) и (7), проводимость зависит от положения уровня Ферми, которое в свою очередь зависит от плотности электронов. Таким образом, в конечном итоге проводимость зависит от плотности, $\sigma_S(n_S)$. Измерив экспериментально зависимость удельной проводимости от плотности электронов, найдем подвижность электронов из $\sigma_S = n_S q \mu$. Конечно, надо иметь в виду, что уравнения (6) и (7) одинаково пригодны как для *n*-проводников, так и для *p*-проводников.

6.1. Методы измерения сопротивления (проводимости)

Рассмотрим простейшую задачу измерения сопротивления R_{ch} материала, из которого сделан резистор длины L (рис. 2): $R_{ch} = \rho_S L / W$.

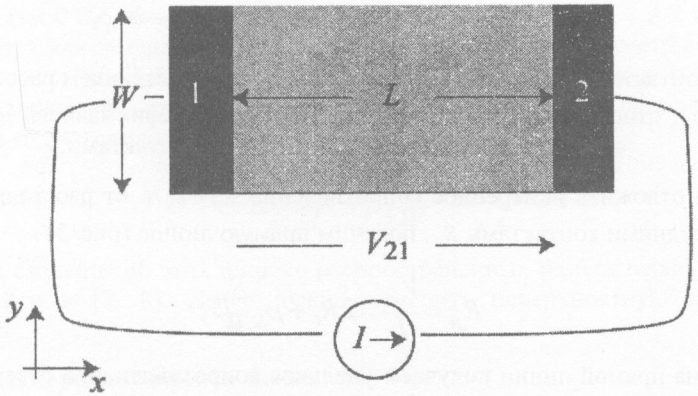


Рис. 2. Плоский резистор с двумя металлическими контактами 1 и 2.

При подаче тока на контакт 2 и выведения его с контакта 1 измеряется разность потенциалов V_{21} и

$$R = \frac{V_{21}}{I} = R_{ch} + 2R_c, \qquad (8)$$

где R_c есть сопротивление контакта метал/полупроводник. Измеренное сопротивление включает в себя сопротивление контакта. Нужно либо отдельно измерить сопротивление контакта, либо выполнить измерения по такой схеме, которая исключала бы сопротивление контактов.

Сначала рассмотрим первый вариант измерения. Он известен как метод переменной длины проводника (Transmission Line Method) [6]. Используется плоский резистор с контактами 1, 2, 3,..., расположенными на разном расстоянии друг относительно друга (рис. 3).

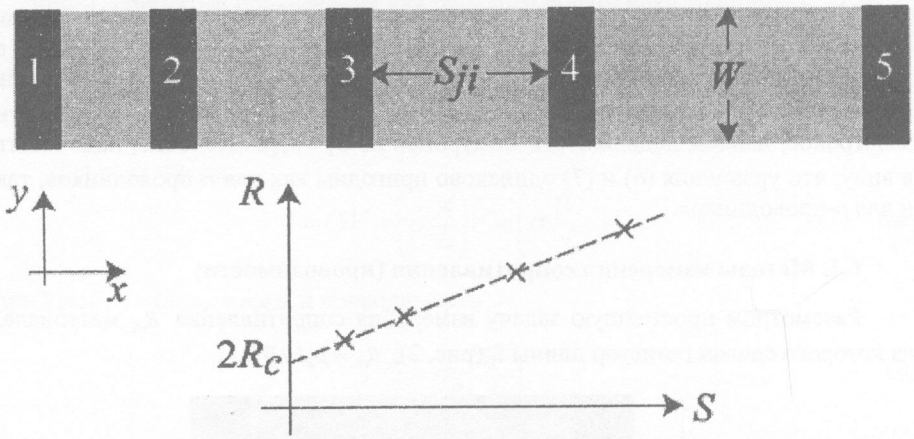


Рис. 3. Контакты расположены, например, на все возрастающем расстоянии S_{ji} друг относительно друга. Показан типичный график зависимости сопротивления от расстояния между контактами.

Если отложить измеренное сопротивление $R_{ji} = V_{ji}/I$ от расстояния между двумя соседними контактами S_{ji}, получим прямую линию (рис. 3)

$$R_{ji} = \frac{V_{ji}}{I} = 2R_c + \rho_S \frac{S_{ji}}{W}, \qquad (9)$$

Из наклона прямой линии получаем удельное сопротивление, а отсекаемый на оси сопротивления отрезок дает сопротивление контакта. Можно также извлечь удельное сопротивление контакта в $Ом \cdot м^2$ и расстояние, на которое ток проникает в метал контакта [6].

Теперь рассмотрим второй вариант измерения сопротивления с исключением сопротивления контактов путем использования четырехточечной схемы измерений (рис. 4).

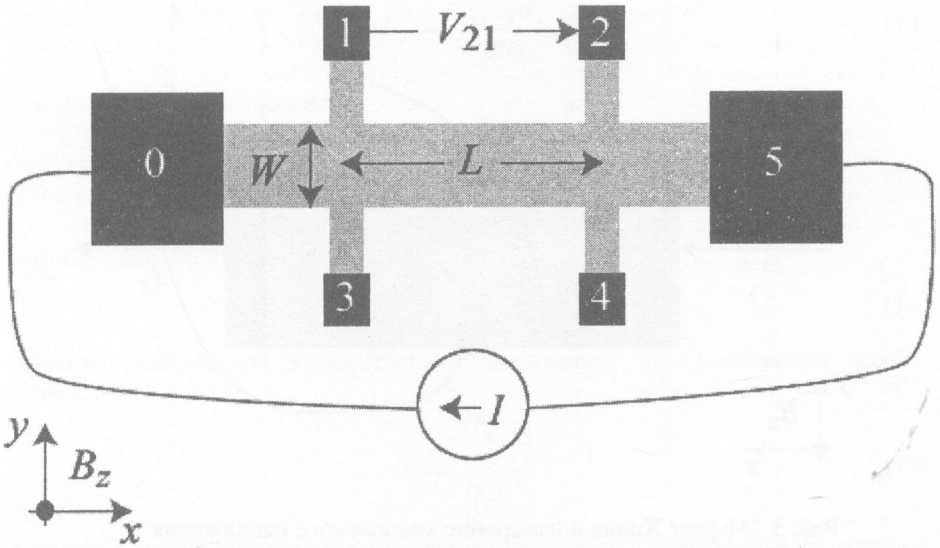


Рис. 4. Схема четырехточечной схемы измерений сопротивления, используемой также для измерения эффекта Холла (холловская линейка).

Контакты 0 и 5 используют для подвода тока, а контакты 1 и 2 или 3 и 4 используются для замера напряжения. Если используется вольтметр с высоким импедансом, ток между измерительными контактами отсутствует, так что нет падения напряжения на границе контактов с измеряемым проводником и замеряется только сопротивление материала проводника:

$$R = \frac{V_{21}}{I} = \rho_S \frac{L}{W}. \tag{10}$$

Детальные сведения об этих широко распространенных измерительных схемах можно найти в [7, 8]. Далее нужно измерить поверхностную плотность электронов.

6.2. Измерение эффекта Холла

Если в четырехточечной измерительной схеме (рис. 4) ток подать в направлении $+x$, магнитное поле приложить в направлении $+z$, а напряжение мерять в направлении $+y$ между контактами 1 и 3 или 2 и 4, то результирующее холловское напряжение будет положительным для n-проводников и отрицательным для p-проводников. Эта схема измерений позволяет рассчитать поверхностную плотность электронов.

Поскольку ток идет в направлении $+x$, электроны движутся в направлении $-x$ со средней скоростью, определяемой из $I_x = Wn_S q\langle v_x \rangle$ (рис. 5).

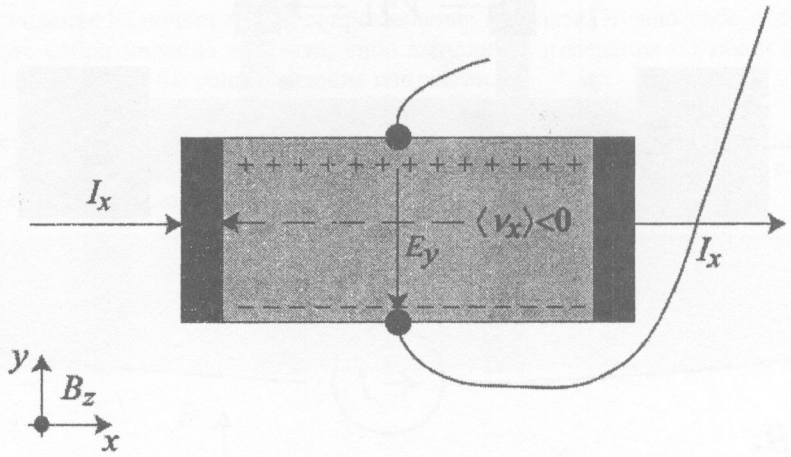


Рис. 5. Эффект Холла и измерение холловского напряжения V_H.

Сила Лоренца $\vec{F}_e = -q\vec{v} \times \vec{B}$ действует на электроны в направлении $-y$, что приводит к накоплению электронов на нижней стороне 2D проводника, а на верхней стороне создается дефицит электронов, что в свою очередь порождает электрическое поле в направлении $-y$, компенсирующее действие магнитного поля. Соответствующее напряжение $V_H = -WE_y$ называют холловским.

Анализ эффекта Холла начинается с уравнения (глава 5) [9]

$$\vec{J} = \sigma_S \vec{E} - \sigma_S \mu r_H \left[\vec{E} \times \vec{B} \right], \tag{11}$$

где r_H – холловский фактор, которое далее удобнее выписать покомпонентно, а именно:

$$\begin{aligned} J_x &= \sigma_S E_x - \sigma_S \mu r_H E_y B_z, \\ J_y &= \sigma_S E_y + \sigma_S \mu r_H E_x B_z, \end{aligned} \tag{12}$$

и переписать относительно электрического поля:

$$\begin{aligned} E_x &= \rho_S J_x + (\rho_S \mu r_H B_z) J_y, \\ E_y &= \rho_S J_y - (\rho_S \mu r_H B_z) J_x. \end{aligned} \tag{13}$$

Поскольку в описанных выше экспериментальных условиях $J_y = 0$, то из второго уравнения имеем:

$$E_y = -\frac{r_H B_z J_x}{q n_S}, \tag{14}$$

что можно переписать в виде

$$R_H \equiv \frac{E_y}{B_z J_x} = \frac{r_H}{(-q) n_S} = \frac{1}{(-q) n_H}, \tag{15}$$

где R_H называют коэффициентом Холла, а

$$n_H \equiv \frac{n_S}{r_H} \tag{16}$$

называют холловской концентрацией. Напомним, что холловский фактор (глава 5)

$$r_H \equiv \frac{\left\langle\left\langle \tau^2 \right\rangle\right\rangle}{\left\langle\left\langle \tau \right\rangle\right\rangle^2} \tag{17}$$

близок к единице (сверху) и его значение определяется механизмом рассеяния в проводнике.

Наконец, выразим коэффициент Холла через экспериментально измеряемое холловское напряжение $V_H = -WE_y$ и ток $I = WJ_x$, а именно:

$$R_H = \frac{-V_H}{IB_z} = \frac{1}{(-q) n_H}. \tag{18}$$

Итак, экспериментально измеряемый коэффициент Холла напрямую связан с холловской концентрацией (для p-проводников $(-q)$ заменяется на $(+q)$). Холловская концентрация обычно приводится в качестве поверхностной плотности электронов. Нужно, однако, помнить, что холловская концентрация и поверхностная плотность электронов связаны друг с другом соотношением (16).

6.2.1. Пример анализа результатов холловского эксперимента

Проанализируем следующие результаты измерения эффекта Холла на холловской линейке (рис. 4):

$$L = 100 мкм, \quad W = 50 мкм, \quad I = 1 мкА, \quad B_z = 0.2 Тл,$$
$$V_{21}|_{B_z=0} = 0.54 мВ, \quad V_{24}|_{B_z=0.2Тл} = 0.13 мВ.$$

Из этих результатов измерений можно получить сопротивление измеренного образца, выяснить это проводник n- или p-типа, найти поверхностную

плотность электронов (точнее, холловскую концентрацию) и подвижность (точнее, холловскую подвижность).

В отсутствии магнитного поля

$$R_{xx} = V_{21} / I = 540 Ом. \quad (19)$$

Поскольку измеренное сопротивление $R_x = \rho_S L / W$ пропорционально удельному поверхностному сопротивлению, то

$$\rho_S = 270 Ом/\square. \quad (20)$$

Измеренное холловское напряжение V_{24} положительно, так что речь идет о проводнике *n*-типа. Из уравнения (18) находим коэффициент Холла

$$R_H = \frac{-V_{24}}{IB_z} = -650 м^2/Кл \quad (21)$$

и холловскую концентрацию

$$n_H = r_H n_S = 9.6 \times 10^{15} м^{-2} = 9.6 \times 10^{11} см^{-2}. \quad (22)$$

Далее, для удельной поверхностной проводимости имеем:

$$\sigma_S = \frac{1}{\rho_S} \equiv n_S q \mu = n_H q \mu_H, \quad (23)$$

откуда для холловской подвижности получаем

$$\mu_H = r_H \mu = 24100 \, см^2/В \cdot с. \quad (24)$$

Можно предположить тот или иной механизм рассеяния, найти холловский фактор r_H и определить истинные значения поверхностной плотности электронов и их подвижности. Полученные таким образом значения n_S и μ предполагают некоторую неопределенность, связанную с выбором модели рассеяния. Поэтому обычно просто приводят холловские концентрации и подвижности.

6.3. Метод измерений ван дер Пау

Метод ван дер Пау позволяет измерять сопротивление и холловские эффекты плоского образца произвольной формы [7, 8, 10], не прибегая к холловский линейке (рис. 6).

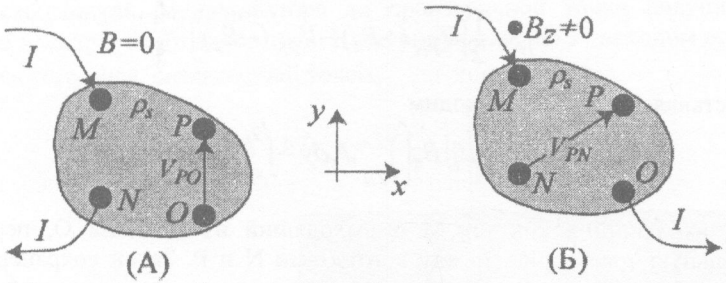


Рис. 6. Метод измерений ван дер Пау: (А) схема измерения сопротивления; (Б) схема измерения эффекта Холла.

Измеряемый образец находится в плоскости *xy*. Он может иметь произвольную форму, но проводник должен быть однородным с изотропной проводимостью и не иметь отверстий. Контакты должны быть достаточно точечными и располагаться по периметру образца. При измерении сопротивления ток подается на два соседних контакта, например, M и N, а напряжение меряется между двумя другими контактами P и O (рис. 6А). Измеренное таким образом сопротивление

$$R_{MN,OP} = \frac{V_{PO}}{I}$$

позволяет вычислить удельное поверхностное сопротивление.

При измерении эффекта Холла магнитное поле направляется перпендикулярно образцу, ток подается на два противоположных контакта, например, M и O, а напряжение меряется между двумя другими контактами N и P. Измеренное напряжение

$$R_{MO,NP} = \frac{V_{PN}}{I} \tag{25}$$

позволяет вычислить холловское напряжение.

Анализ измерительной схемы ван дер Пау начинаем с уравнений (13)

$$E_x = \rho_S J_x + (\rho_S \mu r_H B_z) J_y,$$
$$E_y = \rho_S J_y - (\rho_S \mu r_H B_z) J_x.$$

Разность потенциалов между контактами N и P есть интеграл по траектории

$$V_{PN}(B_z) = -\int_N^P \vec{E} \cdot d\vec{l} = -\int_N^P (E_x dx + E_y dy). \tag{26}$$

Холловское напряжение определяется как

$$V_H \equiv \frac{1}{2}\left[V_{PN}(+B_z) - V_{PN}(-B_z)\right]. \tag{27}$$

Подставляя (13) в (26), находим

$$V_H = \rho_S \mu_H B_z \left(\int_{y_N}^{y_P} J_x dy - \int_{x_N}^{x_P} J_y dx \right). \tag{28}$$

Ток, входящий в контакт M и выходящий из контакта O, пересекает умозрительную траекторию между контактами N и P. Закон сохранения тока требует, чтобы

$$I = \int_N^P \vec{J} \cdot \hat{n}\, dl, \tag{29}$$

где $\hat{n}$ есть единичный вектор, нормальный к траектории, связывающей точки N и P, а именно:

$$\hat{n}\, dl = d\vec{l} \times \hat{z} = dy\,\hat{x} - dx\,\hat{y}. \tag{30}$$

Подставляя (30) в (29), находим

$$I = \int_{y_N}^{y_P} J_x dy - \int_{x_N}^{x_P} J_y dx, \tag{31}$$

и согласно (28) окончательно

$$V_H = \rho_S \mu_H B_z I. \tag{32}$$

Таким образом, геометрия схемы ван дер Пау действительно позволяет измерить эффект Холла.

Перейдем к измерению сопротивления по ван дер Пау. На рис. 7 схеме измерения сопротивления по ван дер Пау (рис. 6А) сопоставлена более простая для анализа геометрия контактов, расположенных в ряд по краю бесконечной полуплоскости.

Ван дер Пау, используя технику конформных преобразований, показал, что результаты измерений в обеих схемах (А) и (Б) рис. 7 идентичны при условии, что измеряемый материал проводника однородный, изотропный и не содержит дыр [10].

Рассмотрим, что происходит при входе тока в точечный контакт M бесконечной полуплоскости (рис. 7Б). Ток расходится радиально относительно точки входа с плотностью и полем

$$J_r = \frac{I}{\pi r}, \quad E_r = \frac{I\rho_S}{\pi r}, \tag{33}$$

где r есть модуль радиус-вектора до произвольной точки полуплоскости, берущего начало в точке входа тока. Интегрируя от r_0 до r, находим потенциал электрического поля, создаваемый током,

$$V(r)-V(r_0)=-\frac{I\rho_S}{\pi}\ln\left(\frac{r}{r_0}\right). \tag{34}$$

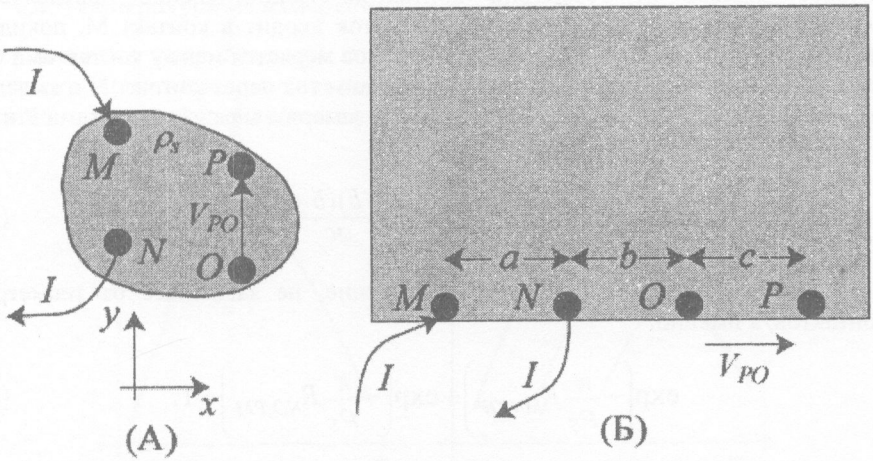


Рис. 7. Сравнение геометрии схемы (А) измерения сопротивления по ван дер Пау с умозрительной схемой (Б) расположения контактов по краю бесконечной полуплоскости и дающей те же результаты измерений, что и схема (А) при вполне приемлемых ограничениях, накладываемых на измеряемый объект [10].

Теперь можно написать (рис. 7), что

$$V(P)=-\frac{I\rho_S}{\pi}\ln\left(\frac{a+b+c}{r_0}\right),$$

$$V(O)=-\frac{I\rho_S}{\pi}\ln\left(\frac{a+b}{r_0}\right), \tag{35}$$

$$V_{PO}=V(P)-V(O)=-\frac{I\rho_S}{\pi}\ln\left(\frac{a+b+c}{a+b}\right).$$

Еще есть потеря в потенциале, вызванная током, уходящим из контакта N, а именно:

$$V_{PO}^{(N)}=+\frac{I\rho_S}{\pi}\ln\left(\frac{b+c}{b}\right). \tag{36}$$

Разность потенциалов между контактами O и P есть сумма обоих вкладов. После деления ее на ток получим сопротивление

$$R_{MN,OP}=\frac{V_{PO}+V_{PO}^{(N)}}{I}=\frac{\rho_S}{\pi}\ln\left[\frac{(a+b)(b+c)}{b(a+b+c)}\right], \tag{37}$$

которое зависит от геометрии контактов. Сопротивление определяется последовательностью его индексов в (37): ток входит в контакт M, покидает образец через контакт N, разность потенциалов меряется между контактами O и P с P как положительным терминалом. Подадим ток через контакт N и выведем его через контакт O, а разность потенциалов замерим между контактами P и M. Тогда для сопротивления имеем:

$$R_{NO,PM}=\frac{\rho_S}{\pi}\ln\left[\frac{(a+b)(b+c)}{ac}\right]. \tag{38}$$

Комбинируя (37) и (38), получим выражение, не зависящее от геометрии контактов, а именно:

$$\exp\left(-\frac{\pi}{\rho_S}R_{MN,OP}\right)+\exp\left(-\frac{\pi}{\rho_S}R_{NO,PM}\right)=1. \tag{39}$$

Аналогичный результат получен ван дер Пау для геометрии рис. 7A.

Итак, метод измерения сопротивления по ван дер Пау выглядит следующим образом. Измеряете сопротивление (рис. 6A) направляя ток через соседние контакты M и N, а разность потенциалов меряете между двумя другими контактами O и P: $R_{MN,OP}=V_{PO}/I$. Затем делаете аналогичное измерение пропуская ток через контакты N и O, разность потенциалов меряете между P и M: $R_{NO,PM}=V_{PM}/I$. Используя два полученных результата измерений, решаете уравнение (39) для определения значения удельного поверхностного сопротивления ρ_S. Образец, используемый в эксперименте, может не иметь какой-либо специальной формы и контакты могут располагаться более или менее произвольно и на разных расстояниях. Для квадратного образца $a=b=c$ и $R_{MN,OP}=R_{NO,PM}=V/I$, так что уравнение (39) упрощается до

$$\rho_S=\frac{\pi}{\ln 2}\frac{V}{I} \tag{40}$$

и нет нужды решать трансцендентное уравнение (39) методом итераций.

Для измерения эффекта Холла (рис. 6Б) направьте магнитное поле в направлении $+z$, ток проведите между двумя несоседними контактами M и O, замерьте разность потенциалов между двумя другими контактами N и P. Поменяйте направление магнитного поля и измерьте ту же самую разность

потенциалов. Согласно (32) среднее значение двух этих измерений есть холловское напряжение, которое позволяет вычислить далее холловские коэффициент, концентрацию и подвижность.

Холловская линейка и геометрия ван дер Пау предоставляют два разных способа выполнить одно и то же измерение. Метод ван дер Пау с квадратным расположением контактов особенно удобен.

6.4. Температурные измерения

Измерение сопротивления и эффекта Холла при разных температурах образца позволяет пролить свет на физику процессов рассеяния в проводнике. Обычно строят график зависимости подвижности от температуры (рис. 8А).

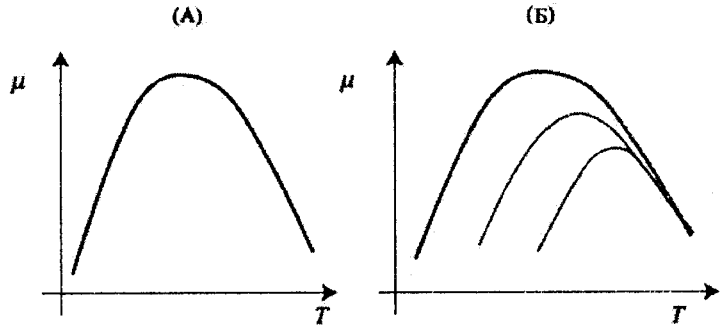


Рис. 8. Типичная зависимость подвижности от температуры:
(А) Характерный вид $\mu(T)$;
(Б) Сдвиг $\mu(T)$ для трех образцов со все большей концентрацией заряженных примесей.

Если подвижность была определена из холловских измерений, то на оси ординат отложена фактически холловская подвижность. С ростом температуры подвижность вначале растет, выходит на плато, а затем падает. Рост подвижности с температурой обычно связывают с рассеянием электронов на заряженных или ионизированных примесях, а падение подвижности с ростом температуры – с рассеянием на фононах. Если работают одновременно два механизма рассеяния, то наблюдаемая подвижность качественно описывается правилом Матиссена [11, 12]:

$$\frac{1}{\mu_{obs}} = \frac{1}{\mu_{ii}} + \frac{1}{\mu_{ph}}, \tag{41}$$

которое утверждает, что наименьшая подвижность контролирует наблюдаемую подвижность.

Чтобы понять увеличение подвижности с ростом температуры, обратимся к рис. 3/гл.4 [13]. Заряженные примеси порождают случайные флуктуации на дне зоны проводимости, которые проявляются в виде случайно расположенных центров рассеяния. Чем выше кинетическая энергия электронов, а кинетическая энергия носителей тока характеризуется расстоянием от дна зоны проводимости, тем меньше электроны чувствуют флуктуирующий потенциал примесей. Время рассеяния должно расти с ростом энергии. Поскольку средняя кинетическая энергия для невырожденных проводников пропорциональна kT, то время рассеяния τ, а поэтому и подвижность $\mu = q\tau/m^*$, должна расти с увеличением температуры. Количественное описание зависимости подвижности от температуры предполагает усреднение должным образом зависимости времени рассеяния от энергии $\tau(E-E_c)$ согласно уравнению (48/гл.5). В 3D-проводниках рассеяние на заряженных примесях ведет к температурной зависимости $\sim T^{3/2}$ [11]. Важно следующее. Как только эксперимент показывает рост подвижности с увеличением температуры, это означает, что участие в рассеянии принимают заряженные примеси.

Уменьшение подвижности при высоких температурах, как правило, вызывается рассеянием на фононах. Рассматривая фононы как частицы, скорость рассеяния оказывается пропорциональной числу заселенности фононов

$$\frac{1}{\tau(E)} \propto n_{ph}, \tag{42}$$

которое описывается статистикой Бозе – Эйнштейна:

$$n_{ph}(\hbar\omega) = \frac{1}{e^{\hbar\omega/kT} - 1}. \tag{43}$$

В ковалентных полупроводниках доминирует рассеяние на акустических фононах. Опираясь на законы сохранения энергии и импульса, можно показать, что основной вклад дают низкоэнергетические фононы. С учетом этого обстоятельства

$$e^{\hbar\omega/kT} \approx 1 + \hbar\omega / kT,$$

так что

$$n_{ph}(\hbar\omega) \approx \frac{kT}{\hbar\omega}.$$

Число фононов пропорционально температуре. Тогда из (42) следует, что

$$\tau \propto \frac{1}{kT},$$

так что подвижность должна уменьшаться с ростом температуры. Уточнение показателя степени при температуре требует корректного усреднения времени рассеяния по энергии. В [11] показано, что в 3D проводниках

$$\mu \propto T^{-3/2}.$$

В полярных полупроводниках в рассеянии доминируют оптические фононы с энергией $\hbar\omega_0 \approx kT$ при комнатной температуре. В этом случае экспоненту в бозевском распределении (43) нельзя разложить по малому показателю степени, и нужно использовать (43) напрямую. Таким образом, в полярных полупроводниках подвижность уменьшается экспоненциально с ростом температуры [11].

На рис. 8Б показана тенденция уменьшения максимально достижимой подвижности в трех образцах, характеризуемых все большей концентрацией заряженных примесей. По мере роста подвижности с увеличением температуры преобладающая роль рассеяния на заряженных примесях постепенно нивелируется в конкуренции с рассеянием на фононах. Подвижность достигает своего максимума при равнодействии обоих механизмов рассеяния. Дальнейшее увеличение температуры приводит к уменьшению подвижности по мере все большего преобладания рассеяния на фононах над рассеянием на заряженных примесях. Достижение максимума подвижности контролируется рассеянием на заряженных примесях. Чем больше концентрация заряженных примесей в образце, тем требуется все более высокая температура для создания все большей плотности фононов, нивелирующих рассеяние на заряженных примесях и уменьшающих максимально достижимую подвижность, тем меньшего максимума достигает подвижность и тем дальше максимум сдвинут в сторону увеличения температуры. Чем меньше концентрация заряженных примесей в образце, тем большего максимума достигает подвижность, а сам максимум достигается при меньшей температуре. Измерение подвижности при низких температурах позволяет оценить общую концентрацию заряженных примесей в проводнике. Более детальное обсуждение практики измерений электрофизических свойств проводников в условиях квазиравновесного режима транспорта можно найти в [7, 8].

Далее речь пойдет об одном артефакте при холловских измерениях (эффект Нернста) и затем об измерениях в сильных магнитных полях.

6.4.1. Температурный артефакт при измерении эффекта Холла (эффект Нернста)

Вернемся к холловский линейке на рис. 4. Ток подается на контакт 0 и покидает проводник через контакт 5. Нужно ожидать охлаждения контакта 5 и нагревания контакта 0 в результате эффекта Пельтье (§ 2.1.2) [14]. Ранее всегда предполагалось, что измерения проводятся в изотермических условиях. Если же подложка холловской линейки не обладает достаточными теплоотводящими

свойствами, то возможен градиент температуры в направлении оси x. Как это скажется на измерении эффекта Холла?

Вернемся к транспортным уравнениям в обращенной форме для 3D проводника в диффузионном режиме (§ 2.2.4) [15], а именно:

$$\frac{d(E_F/q)}{dx} = \rho J_x + S\frac{dT}{dx},$$

$$J_{Qx} = TSJ_x - \kappa\frac{dT}{dx}$$

и перепишем их в индексных обозначениях (§ 5.4):

$$\begin{aligned} E_i &= \rho_{ij}(\vec{B})J_j + S_{ij}(\vec{B})\partial_j T, \\ J_{Qi} &= \pi_{ij}(\vec{B})J_j - \kappa_{ij}(\vec{B})\partial_j T. \end{aligned} \tag{44}$$

У изотропного полупроводника все четыре транспортных тензора диагональны, однако, в присутствии магнитного поля появляются недиагональные компоненты тензора. В слабых магнитных полях диагональные компоненты тензора не изменяются, а недиагональные компоненты появляются из векторных произведений силы Лоренца. Для материалов с кубической симметрией транспортные коэффициенты в присутствии слабого магнитного поля можно записать следующим образом:

$$\begin{aligned} \rho_{ij}(\vec{B}) &= \rho_0 + \rho_0\mu_H\epsilon_{ijk}B_k + \ldots, \\ S_{ij}(\vec{B}) &= S_0 + S_1\epsilon_{ijk}B_k + \ldots, \\ \pi_{ij}(\vec{B}) &= \pi_0 + \pi_1\epsilon_{ijk}B_k + \ldots, \\ \kappa_{ij}(\vec{B}) &= \kappa_0 + \kappa_1\epsilon_{ijk}B_k + \ldots. \end{aligned} \tag{45}$$

Каждый из транспортных коэффициентов в слабом магнитном поле имеет одну и ту же структуру в виде суммы диагональной компоненты, помеченной индексом 0, и недиагональных компонент, происходящих из векторного произведения силы Лоренца и контролируемых по знаку и по отличию от нуля символом Леви-Чивита. Исключая удельное сопротивление, для которого $\rho_1 = \rho_0\mu_H$, для остальных термоэлектрических коэффициентов недиагональные компоненты не раскрыты, им просто присвоен индекс 1.

Вернемся, наконец, к вопросу как влияет температурный градиент в направлении x на измеряемое холловское напряжение. Из уравнения (44) имеем:

$$E_y = \rho_0 J_y + \rho_0\mu_H\epsilon_{yjk}B_k J_j + S_0\partial_y T + S_1\epsilon_{yjk}B_k\partial_y T. \tag{46}$$

Условия проведения эксперимента таковы, что ток направлен только по оси x, а магнитное поле направлено по оси z. Предположим также, что градиент температуры вдоль оси y отсутствует. В этих условиях вместо (46) имеем:

$$E_y = +\rho_0 \mu_H \epsilon_{yxz} B_z J_x + S_1 \epsilon_{yxz} B_z \partial_x T . \tag{47}$$

Наконец, используя значение символа Леви-Чивита $\epsilon_{yxz} = -1$ (§ 5.4), получаем

$$E_y = -\rho_0 \mu_H B_z J_x - S_1 B_z \partial_x T . \tag{48}$$

Первое слагаемое – это чистый эффект Холла. Второе же слагаемое – артефакт, вызванный градиентом температуры в направлении x, который был создан током в том же направлении, и порождающий дополнительную компоненту электрического поля в направлении y (эффект Нернста). Из (48) следует, что при одновременном изменении направления тока и магнитного поля на противоположное, слагаемое, ответственное за эффект Холла, знак не меняет, а второе слагаемое – меняет знак на противоположный. Таким образом, измерив холловское напряжение, а затем переключив направления электрического и магнитного полей на противоположные, измерив еще раз холловское напряжение и взяв среднее из двух измерений, можно избавиться от этого артефакта. Это переключение нужно сделать достаточно быстро настолько, чтобы градиент температуры $\partial T/\partial x$ не успел изменить свой знак. Подобные термомагнитные эффекты могут исказить интерпретацию результатов холловских измерений [11, 16]. Даже в отсутствие магнитного поля продольный температурный градиент, вызванный эффектом Пельтье, может исказить результаты измерения сопротивления.

6.5. Измерения в сильных магнитных полях

При решении транспортного уравнения Больцмана с учетом магнитного поля (§ 5.4) предполагалось, что магнитное поле слабое, так что

$$\omega_c \tau \ll 1, \tag{49}$$

где для параболической дисперсии циклотронная частота

$$\omega_c = \frac{qB}{m^*} . \tag{50}$$

Из этих двух уравнений следует условие слабости магнитного поля

$$\mu B \ll 1. \tag{51}$$

Ниже обсуждаются три вопроса: 1) физический смысл условия слабости магнитного поля (49); 2) циклотронная частота для произвольной дисперсии $E(k)$; 3) влияние величины магнитного поля на результаты измерений.

На рис. 9 показана траектория движения электрона в плоскости xy под действием магнитного поля, направленного в сторону $+z$.

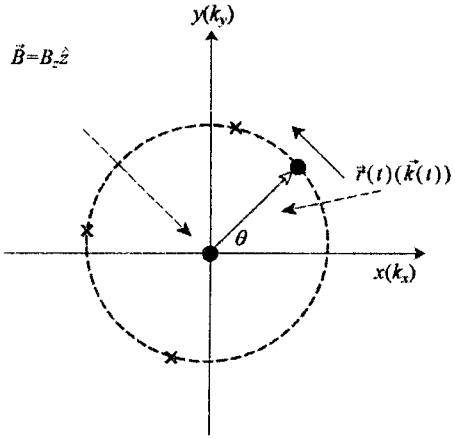


Рис. 9. К вычислению циклотронной частоты.
Крестиком на траектории отмечены акты рассеяния.

Период обращения по орбите равен $T = 2\pi/\omega_c$, так что условие слабости магнитного поля (49) эквивалентно $T \gg \tau$. В слабых магнитных полях период обращения электрона намного превышает время между двумя последовательными столкновениями, так что электроны редко завершают круговые орбиты. В сильных магнитных полях $T \ll \tau$ и электроны, прежде чем столкнуться, успевают совершить одно или несколько круговых движений.

Уравнение движения электрона (рис. 9) есть

$$\frac{d(\hbar\vec{k})}{dt} = -q\vec{v}\times\vec{B} \tag{52}$$

или покомпонентно

$$\hbar\frac{dk_x}{dt} = -qv_yB_z,$$

$$\hbar\frac{dk_y}{dt} = +qv_xB_z,$$

что дает (рис. 9):

$$\hbar k \frac{d(\cos\theta)}{dt} = -q(v\sin\theta)B_z,$$
$$\hbar k \frac{d(\sin\theta)}{dt} = +q(v\cos\theta)B_z. \tag{53}$$

Дифференцируя по времени первое из уравнений (53) и используя второе уравнение, получим

$$\frac{d^2(\cos\theta)}{dt^2} = -\omega_c^2 \cos\theta, \tag{54}$$

где искомая циклотронная частота для произвольной, но изотропной зонной структуры

$$\omega_c = \frac{qv}{\hbar k} B_z. \tag{55}$$

Для параболической дисперсии

$$v = \frac{\hbar k}{m^*},$$

что и дает для этой дисперсии циклотронную частоту (50). Выражение (55) используется для нахождения циклотронной частоты в графене (§ 3.6) [17].

Рассмотрим несколько типичных примеров. У слабодопированного Si подвижность составляет около $1000\,см^2/В\cdot с$. У широко распространенного лабораторного магнита B = 0.2 *Тл*, так что $\mu B \approx 0.02 \ll 1$, и мы гарантировано находимся в режиме слабого магнитного поля. Не трудно приобрести магнит с полем, бóльшим раз в 10, но мы все еще будем в режиме слабых полей. Для работы в режиме уже сильного магнитного поля нужны магниты с полем, бóльшим раз в 100. Подобные магниты доступны, но редко используются в рутинных экспериментах.

Ситуация совсем другая для проводников с высокой подвижностью, например, для полупроводников группы III-V или для модуляционно-легированных гетероструктур. Например, рассмотрим InGaAs при $300\,K$. Подвижность возможна порядка $10^4\ см^2/В\cdot с$. Тогда $\mu B \approx 0.2 < 1$ и мы все еще в режиме слабого поля. Если понизить температуру образца до $77\,K$ и желательно ниже, то вполне достижима подвижность в 10^5 и даже выше. Теперь $\mu B \approx 2 > 1$ и режим сильного магнитного поля уже достижим для распространенных лабораторных магнитов. В сильных полях наблюдаются интересные эффекты.

Обратимся к уравнению (54). Оно описывает гармонические колебания. Уровни энергии гармонического осциллятора квантуются согласно

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega_c. \tag{56}$$

Применительно к рассматриваемой нами ситуации эти уровни энергии называют уровнями Ландау.

Рассмотрим 2D проводник в магнитном поле B_z. В отсутствие магнитного поля плотность состояний (рис. 10А)

$$D_{2D}(E) = g_v \frac{m^*}{\pi\hbar^2}, \quad (E > \varepsilon_1) \tag{57}$$

где ε_1 – дно первой подзоны, и предполагается, что следующие подзоны не заполнены.

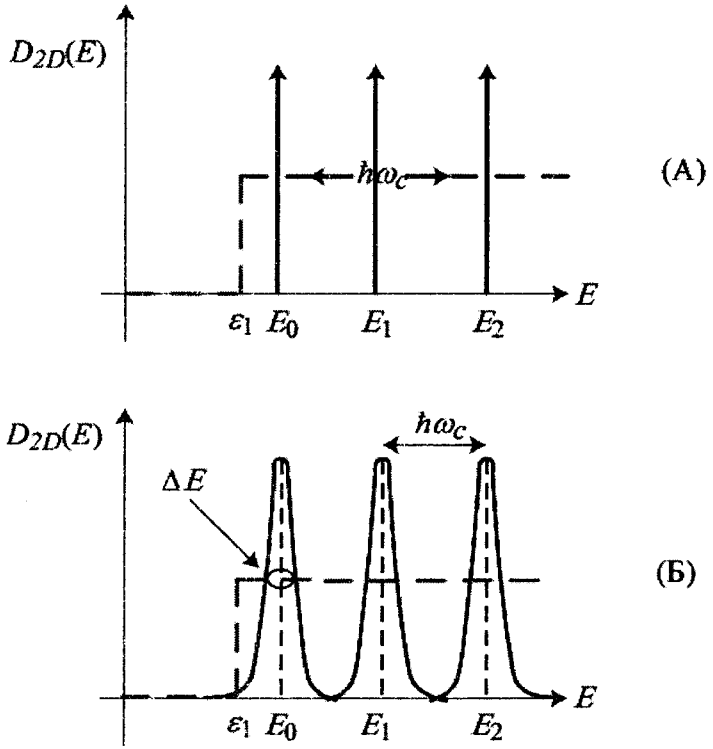


Рис. 10. Магнитное поле изменяет плотность состояний 2D проводника: (А) Рассеяние отсутствует; (Б) Рассеяние уширяет уровни Ландау.

В присутствии сильного магнитного поля плотность состояний дается суммой δ-функций:

$$D_{2D}(E, B_z) = D_0 \sum_{n=0}^{\infty} \delta\left[E - \varepsilon_1 - \left(n + \frac{1}{2}\right)\hbar\omega_c\right], \tag{58}$$

где D_0 – вырождение уровней Ландау. Расстояние между ними $\hbar\omega_c$ и вырождение

$$D_0 = \hbar\omega_c\left(g_v\frac{m^*}{\pi\hbar^2}\right) = g_v\frac{2qB_z}{h}. \qquad (59)$$

Рассеяние уширяет уровни Ландау (рис. 10Б). Ширина каждого уровня определяется принципом неопределенности $\Delta E\Delta t \approx \hbar$. Приняв Δt близким к времени рассеяния, $\Delta t \approx \tau$, получаем $\Delta E \approx \hbar/\tau$. Сильное магнитное поле определяется как поле, достаточное для различения отдельных уровней Ландау. Для этого расстояние между уровнями Ландау $\hbar\omega_c$ должно превышать уширение ΔE отдельной линии, что ведет к известному неравенству $\omega_c\tau \gg 1$ и тем самым демонстрирует еще одну физическую интерпретацию неравенства (49).

6.5.1. Осцилляции Шубникова – де Гааза

Интересные эффекты возникают при измерении сопротивления и эффекта Холла в изменяемом сильном магнитном поле. Пусть у нас есть модуляционно-легированная пленка с $n_S = 5\times10^{11}см^{-2}$ и подвижностью $\mu = 1\times10^5 см^2/В\cdot с$. Зададимся вопросом, сколько уровней Ландау заселено, если $B_z = 1Тл$.

Сначала нужно выяснить, расщепляются ли уровни Ландау. Для подвижности $\mu = 1\times10^5 см^2/В\cdot с = 10\ м^2/В\cdot с$ имеем

$$\mu B_z = 10 \gg 1 \qquad (60)$$

и мы явно находимся в режиме сильного магнитного поля. Далее вычисляем вырождение уровней Ландау. Предполагая $g_v = 1$, находим

$$D_0 = \frac{2qB_z}{h} = 4.8\times10^{10}см^{-2},$$

так что

$$\frac{n_S}{D_0} = 10.4,$$

т.е. первые 10 уровней Ландау заселены, а уровень Ферми лежит в пределах 11-го уровня Ландау.

Рис. 11 показывает результаты измерения подобной структуры [18].

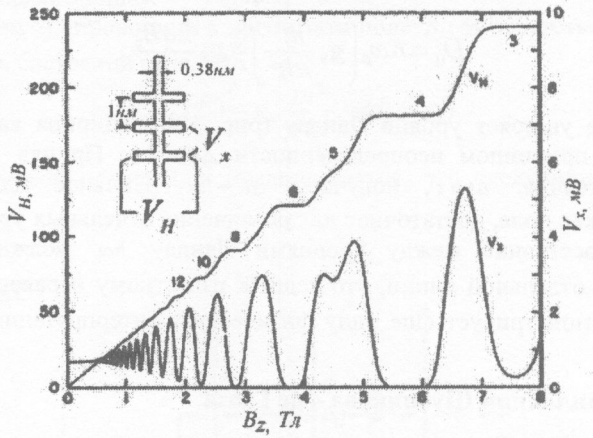


Рис. 11. Холловские измерения модуляционно-легированной структуры AlGaAs/GaAs в сильном магнитном поле при 1.2 *K*. Инжектированный ток – 25.5 *мкА* и $n_S = 5.6 \times 10^{11} см^{-2}$ [18].

Рассмотрим сначала продольное сопротивление $R_{xx} = V_x / I$ (правая ось на рис. 11). При слабых полях сопротивление не зависит от магнитного поля и является сопротивлением образца. По мере увеличения магнитного поля R_{xx} начинает осциллировать. Осцилляции становятся все более выраженными по мере того как все больше проявляются уровни Ландау. Эти осцилляции R_{xx} известны как осцилляции Шубникова – де Гааза [19, 20]. Период осцилляций позволяет измерить поверхностную плотность электронов [21]. Когда магнитное поле становится очень сильным, R_{xx} фактически зануляется. Это – квантовый эффект Холла [18, 21 – 24].

Поведение холловского напряжения также представляет интерес (левая ось на рис. 11). В слабых магнитных полях холловское напряжение пропорционально полю, согласно (32). По мере усиления магнитного поля начинается квантование холловского напряжения, сопровождаемое ростом высоты и ширины ступенек квантования. Сами ступеньки совпадают с участками зануления R_{xx}. Квантование холловского напряжения настолько прецизионное, что может служить стандартом сопротивления [18].

Подведем итоги. Холловская линейка и измерительные геометрии ван дер Пау позволяют измерить как сопротивление, так и холловскую концентрацию и холловскую подвижность. При желании можно оценить и холловский фактор, что позволит более точно определить поверхностную плотность электронов и их истинную подвижность. Однако, на практике обычно приводят холловскую концентрацию и холловскую подвижность. Температурные измерения

помогают идентифицировать доминирующий механизм рассеяния. Холловские измерения требуют особой тщательности с целью избежать проявления гальвано- и термомагнитных эффектов, таких как эффект Нернста, поперечный термомагнитный эффект Риги – Ледюка, поперечный гальванотермомагнитный эффект Эттингсхаузена. Существенно новую информацию предоставляют измерения в сильных магнитных полях, однако для подобных экспериментов нужны материалы с высокой подвижностью [25, 26].

Литература

1. S. Datta. *Lessons from Nanoelectronics: A New Perspective on Transport.* Hackensack, New Jersey: World Scientific Publishing Company, 2012; www.nanohub.org/courses/FoN1.
2. M. Lundstrom, Jeong Changwook. *Near-Equilibrium Transport: Fundamentals and Applications.* Hackensack, New Jersey: World Scientific Publishing Company, 2013; www.nanohub.org/resources/11763.
3. Yuriy Kruglyak. Landauer – Datta – Lundstrom Generalized Transport Model for Nanoelectronics, *J. Nanoscience*, v. **2014**, Art. ID 725420; www.dx.doi.org/10.1155/2014/725420.
4. Ю. А. Кругляк. Обобщенная модель транспорта электронов и тепла Ландауэра – Датты – Лундстрома в микро- и наноэлектронике, *ScienceRise*, т. 5, № 3(5), 21 – 38 (2014).
5. L. Danielson. *Measurement of the thermoelectric properties of bulk and thin film materials*, 1996: www.osti.gov/scitech/biblio/663573.
6. H. H. Berger. Models for contacts to planar devices, *Solid-State Electron.*, v. 15, 145 – 158 (1972).
7. D. K. Schroder. *Semiconductor Material and Device Characterization.* New York: Wiley Interscience, 2006.
8. D. C. Look. *Electrical Characterization of GaAs Materials and Devices.* New York: John Wiley and Sons, 1989.
9. Ю. А. Кругляк, Модель проводимости Ландауэра – Датты – Лундстрома в микро- и наноэлектронике и транспортное уравнение Больцмана, *ScienceRise*, т. 3, № 2(8), 108 – 116 (2015).
10. L. J. van der Pauw. A method for measuring specific resistivity and Hall effect of discs of arbitrary shape, *Phillips Res. Rep.*, v. 13, 1 – 9 (1958).
11. Mark Lundstrom. *Fundamentals of Carrier Transport.* Cambridge UK: Cambridge University Press, 2000.
12. Н. Ашкрофт, Н. Мермин. *Физика твердого тела.* Москва: Мир, 1979.
13. Ю. А. Круглякю. Учет рассеяния в транспортной модели Ландауэра – Датты – Лундстрома, *ScienceRise*, т. 3, № 2(8), 99 – 107 (2015).
14. Ю. А. Кругляк. Термоэлектрические явления и устройства в концепции Ландауэра – Датты – Лундстрома, *ScienceRise*, т. 1, № 2(6), 69 – 77 (2015).

15. Ю. А. Кругляк. Термоэлектрические коэффициенты в транспортной модели Ландауэра – Датты – Лундстрома, *ScienceRise*, т. 1, №2(6), 78 – 89 (2015).
16. C. M. Wolfe, N. Holonyak, G. E. Stillman. Physical Properties of Semiconductors. Englewood Cliffs, N. Jersey: Prentice Hall, 1989.
17. Ю. А. Кругляк Графен в транспортной модели Ландауэра – Датты – Лундстрома, *ScienceRise*, т. 2, №2(7), 93 – 106 (2015).
18. M. E. Cage, R. F. Dziuba, B. F. Field. A test of the quantum Hall effect as a resistance standard, *IEEE Trans. Instr. Meas.*, v. 34, 301 – 303 (1985).
19. L. W. Schubnikov, W. J. de Haas. *Proc. Royal Netherlands Acad. Arts Science*, v. 33, 130, 163 (1930).
20. Л. В. Шубников. *Избранные труды. Воспоминания*. Киев: Наукова думка, 1990.
21. D. F. Holcomb. Quantum electrical transport in samples of limited dimensions, *Amer. J. Phys.*, v. 67, 278 – 297 (1999).
22. J. H. Davies. *The Physics of Low-Dimensional Semiconductors*. Cambridge: Cambridge Univ. Press, 1998.
23. S. Datta. *Electronic Transport in Mesoscopic Systems*. Cambridge: Cambridge University Press, 2001.
24. Ю. О. Кругляк, М. В. Стріха, Уроки наноелектроніки: ефект Холла і вимірювання електрохімічних потенціалів у концепції «знизу – вгору», *Сенсор. електр. мікросист. техн.*, т. 11, № 1, 5 – 27 (2014).
25. Ю. А. Кругляк. Измерение электрофизических свойств в концепции транспортной модели Ландауэра – Датты – Лундстрома, *ScienceRise*, т. 4, № 2(9), 77 – 88 (2015).
26. Ю. А. Кругляк. Измерение сопротивления и холловских эффектов в микро- и наноэлектронике методом ван дер Пау, *Сб. науч. тр. конф. «Функциональная база наноэлектроники», Харьков – Одесса, 28 сент. – 2 окт.*, 102 – 105 (2015).

Глава 7. Роль электростатики и контактов

До сих пор при изложении причин возникновения электрического тока (глава 1) [1, 2], ни при обсуждении далее уравнения диффузии для описания баллистического транспорта электронов не привлекалось и не привлекается внешнее электрическое поле в качестве причины, создающей разность потенциалов на концах проводника, хотя именно оно и в поныне практикуемом стандартном изложении объявляется причиной, порождающей электрический ток. Может показаться парадоксальным, что фактически вся теория электрической проводимости строится [1 – 3] без привлечения понятия внешнего электрического поля.

Сами электрические поля удобно представлять в виде градиента электростатического потенциала U/q. Мы же при объяснении причины возникновения электрического тока пользовались электрохимическим потенциалом E_F. Различие между этими двумя потенциалами следующее:

$$E_F = (E_F - U) + U, \qquad (1)$$

где первое слагаемое обычно называют электрохимическим потенциалом, а второе – электростатическим потенциалом. Соотношение между потенциалами и их связь с плотностью состояний $D(E)$ показана на рис. 1.

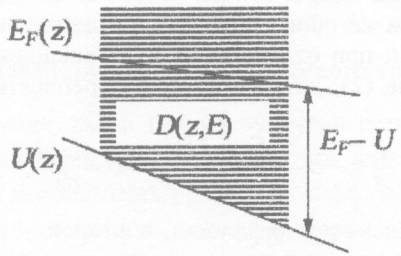


Рис. 1. Качественная картина изменения электрохимического E_F/q и электростатического U/q потенциалов вдоль длины проводника z. Разность между ними есть изменяющаяся вдоль длины проводника плотность состояний.

Электрохимический потенциал E_F фиксирует энергию, вплоть до которой включительно заполнены электронные состояния проводника, а электростатический потенциал U определяет сдвиг всех доступных электронных состояний проводника, так что их разность $E_F - U$ есть мера заполненности уровней энергии электронами и просто равна числу электронов.

Уравнение диффузии, связывающее электрический ток с градиентом электрохимического потенциала,

$$\frac{I}{A} = -\sigma \frac{d(E_F / q)}{dz} \tag{2}$$

где σ – удельная проводимость проводника с площадью поперечного сечения A, мы получим и подробно обсудим позже. Однако, фактически (2) это уравнение

$$\frac{I}{A} = -\sigma \frac{d(E_F - U)/q}{dz} \tag{3}$$

в предположении, что внешнее электрическое поле $d(U/q)/dz$ равно нулю. Каким же должно быть уравнение диффузии, если мы хотим учесть отличное от нуля внешнее электрическое поле?

7.1. Диффузия и дрейф

На первый взгляд кажется, что для вывода такого уравнения потребуется с самого начала учесть влияние электрического поля $d(U/q)/dz$ на все электроны проводника. Однако, необходимости решать такую задачу «с нуля» нет, поскольку согласно постулатам равновесной статистической механики электрический ток отсутствует, если электрохимический потенциал E_F на обоих концах проводника один и тот же, точно так же как и отсутствует поток тепла, если температура на обоих концах проводника одна и та же. Поэтому уравнением диффузии и при отличном от нуля внешнем электрическом поле должно быть уравнение (2), которое удобно переписать в виде суммы двух вкладов, а именно:

$$\frac{I}{A} = -\sigma \frac{d(E_F - U)/q}{dz} - \sigma \frac{dU/q}{dz}, \tag{4}$$

первый из которых описывает диффузию, а второй – дрейф. При этом оба вклада пропорциональны одному и тому же значению удельной проводимости – требование, ведущее к соотношению Эйнштейна между диффузией и дрейфом заряженных частиц [4 – 6]. Именно поэтому мы можем найти электрическую проводимость σ, рассматривая лишь диффузию электронов в отсутствии внешнего электрического поля ($U = 0$), заменив ур-е (3) уравнением (2), которое, как оказывается, уже все учитывает. Нет нужды отдельно рассматривать задачу о дрейфе. Другими словами, диффузионное уравнение (2) является на самом деле уравнением, описывающем и диффузию и дрейф, хотя сам дрейф в явном виде не рассматривается.

А не могли бы мы вместо изложенного выше подхода пренебречь в уравнении (4) диффузионным вкладом и учесть лишь дрейфовую составляющую? Ведь в этой модели ток действительно порождается внешним электрическим полем и нет нужды обращаться к электрохимическим

потенциалам. Проблема с таким подходом, однако, состоит в том, что теперь надо привлекать некие мистические силы, чтобы объяснить почему внешнее электрическое поле действует лишь на некоторые электроны проводника, а не на все электроны. Энергетическое окно для транспорта электронов определяется функцией теплового уширения F_T (§ 1.2.1), возникающей естественным образом как результат «подготовки» двух контактов:

$$f_1(E) - f_2(E) = \left(-\frac{\partial f_0}{\partial E} \right) (E_{F1} - E_{F2}) . \qquad (5)$$

Позже мы обоснуем пренебрежение дрейфовой составляющей в (4) опираясь на транспортное уравнение Больцмана (ТУБ), более известное в русскоязычной литературе как кинетическое уравнение Больцмана, которое обычно является отправной точкой при обсуждении любых задач транспорта частиц. Мы уже обращались к ТУБ в связи с эффектом Холла (глава 5) [7]. Как и прежде, не вдаваясь в детали, покажем лишь, что в режиме линейного отклика ток действительно зависит лишь от градиента электрохимического потенциала dE_F/dz, а не от градиента электростатического потенциала dU/dz. Покажем также, что в случае динамического, переменного во времени транспорта пренебрежение учетом электрических полей может привести к ошибочным результатам. Даже в условиях динамического равновесия электрические поля играют важную роль в описании всей вольт-амперной характеристики как только мы выходим за пределы режима линейного отклика.

7.2. Уравнение диффузии для баллистического транспорта

Звучит противоречиво как и термин «упругий резистор» (глава 1). Разве диффузионное уравнение не должно было бы описывать диффузионный транспорт? Можно ли использовать уравнение диффузии для баллистического транспорта? С позиций концепции «снизу – вверх» оба режима переноса – диффузионный и баллистический – существенно близки.

Уравнение диффузии связывает электрический ток с градиентом электрохимического потенциала $E_F(z)$, а именно:

$$\frac{I}{A} = -\frac{\sigma}{q} \frac{dE_F}{dz}, \qquad (6)$$

где удельная проводимость σ дается уравнениями (234/гл.1) и (237/гл.1). Это уравнение можно получить рассматривая проводник как последовательность упругих резисторов (рис. 2).

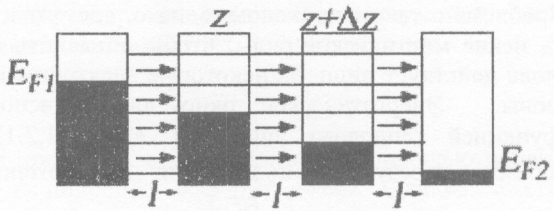


Рис. 2. Условное разбиение реального макропроводника на последовательность упругих резисторов (глава 1).

Используя ур-е (201/гл.1), для тока $I(z)$ в отдельной секции проводника можно написать:

$$I(z)=\frac{1}{q}\int_{-\infty}^{+\infty} dE\,G(E)\Big(f(z,E)-f(z+\Delta z,E)\Big). \tag{7}$$

Из уравнений (211/гл.1) и (219/гл.1) для проводимости в диффузионном режиме имеем

$$G=\frac{\sigma}{L+\lambda}\{1,W,A\}, \tag{8}$$

откуда следует, что

$$\frac{1}{G(E)}=\rho\frac{\Delta z+\lambda}{A}; \tag{9}$$

однако, при этом нужно отметить, что сопротивление (9) включает в себя граничные сопротивления, которые на самом деле не существуют, разве что на физически реальных концах проводника. Опуская их, для проводимости имеем

$$G(E)=\frac{\sigma A}{\Delta z}, \tag{10}$$

Комбинируя (10) с уже привычным линейным разложением для малой разности электрохимических потенциалов

$$f(z,E)-f(z+\Delta z,E)=\left(-\frac{\partial f_0}{\partial E}\right)\Big(E_F(z)-E_F(z+\Delta z)\Big), \tag{11}$$

которое следует из (5), и определяя удельную проводимость σ как термически среднее $\bar{\sigma}$ от $\sigma(E)$, получаем

$$I(z)=\frac{1}{q}\frac{\sigma A}{\Delta z}\Big(E_F(z)-E_F(z+\Delta z)\Big). \tag{12}$$

Обратим внимание на то, что удельные проводимости (234/гл.1) и (237/гл.1), как и проводимости выше в уравнениях (10) и (12), зависят от энергии. Они должны быть усреднены в промежутке нескольких kT, включая $E = E_{F0}$, с использованием функции термического уширения

$$\bar{\sigma} = \int_{-\infty}^{+\infty} dE \left(-\frac{\partial f_0}{\partial E} \right) \sigma(E). \tag{13}$$

Именно такая термически усредненная проводимость $\bar{\sigma}$ должна сравниваться с удельной проводимостью в классических формулах теории Друде (формулы (238/гл.1) и (240/гл.1)). В вырожденных проводниках усредненная проводимость $\bar{\sigma}$ приблизительно равна проводимости при $E = E_{F0}$:

$$\bar{\sigma} \approx \sigma\big|_{E=E_{F0}}. \tag{14}$$

Вернемся к уравнению (12). Устремляя $\Delta z \to 0$, получим искомое уравнение диффузии (6).

Уравнение диффузии обычно идет в паре с уравнением непрерывности. В одномерных проводниках, как на рис. 3, в условиях равновесия ток постоянен на всем протяжении проводника

$$\frac{dI}{dz} = 0. \tag{15}$$

Решение системы уравнений (6) и (15) ищется при граничных условиях

$$\begin{aligned} E_F(z=0) &= E_{F1}, \\ E_F(z=L) &= E_{F2}. \end{aligned} \tag{16}$$

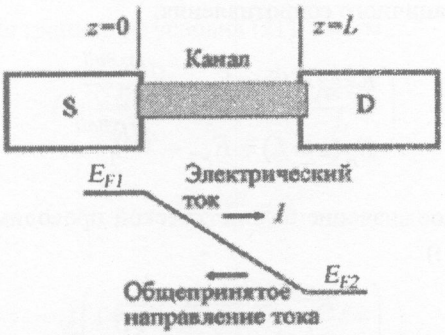


Рис. 3. К решению системы уравнений (6) и (15) с граничными условиями (16). Как и на рис. 16/гл.1, здесь и далее (§§ 7.2 – 7.4) используется направление тока S → D в отличие от общепринятого направления.

Линейное решение, графически показанное на рис. 3, удовлетворяет систему уравнений (6) и (15) с граничными условиями (16), поскольку линейная зависимость $E_F(z)$ имеет постоянный наклон

$$\frac{dE_F}{dz} = -\frac{E_{F1} - E_{F2}}{L}, \tag{17}$$

так что из уравнения (6) имеем постоянный ток с $dI/dz = 0$:

$$I = \frac{\sigma A}{q}\frac{E_{F1} - E_{F2}}{L}. \tag{18}$$

Разность электрохимических потенциалов $E_{F1} - E_{F2} = qV$. Имеем стандартный закон Ома

$$I = \frac{\sigma A}{L}V, \tag{19}$$

а не обобщенный закон Ома, пригодный также для учета баллистического транспорта (206/гл.1),

$$I = \frac{\sigma A}{L + \lambda}V. \tag{20}$$

Можно ли получить обобщенный закон Ома (20) из уравнений диффузии и непрерывности (6) и (15)? На первый взгляд нет, поскольку традиционная проводимость и коэффициент диффузии не имеют смысла для баллистического транспорта. И все же можно пользоваться уравнениями (6) и (15) для баллистического транспорта, если модифицировать граничные условия (16) путем учета в них граничного сопротивления:

$$\begin{aligned} E_F(z=0) &= E_{F1} - \frac{qIR^{Ball}}{2}, \\ E_F(z=L) &= E_{F2} - \frac{qIR^{Ball}}{2}, \end{aligned} \tag{21}$$

где R^{Ball} есть обратное значение баллистической проводимости G^{Ball} (формулы (219/гл.1) и (235/гл.1))

$$R^{Ball} = \frac{\lambda}{\sigma A} = \frac{h}{q^2 M}. \tag{22}$$

Новые граничные условия (21) можно реализовать в виде граничных сопротивлений $R^{Ball}/2$, что ведет к скачкам химпотенциалов, как показано на рис. 4.

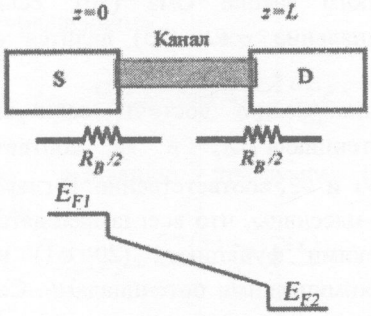


Рис. 4. Уравнения (6) и (15) можно использовать не только для описания диффузионного транспорта, но и для баллистического транспорта, если граничные условия (16) модифицировать путем введения граничных сопротивлений $R^{Ball}/2$.

Теперь легко убедиться, что новые граничные условия (21) в применении к однородному проводнику ведут к обобщенному закону Ома (20). Поскольку $E_F(z)$ меняется линейно от $z=0$ до $z=L$, ток по уравнению (6)

$$I=\frac{\sigma A}{q}\frac{E_F(0)-E_F(L)}{L}. \tag{23}$$

Используя новые граничные условия (21), имеем

$$I=\frac{\sigma A}{q}\left(\frac{E_{F1}-E_{F2}}{L}-\frac{qIR^{Ball}}{L}\right). \tag{24}$$

Поскольку

$$\sigma AR^{Ball}=\lambda, \tag{25}$$

то

$$I\left(1+\frac{\lambda}{L}\right)=\frac{\sigma A}{q}\left(\frac{E_{F1}-E_{F2}}{L}\right). \tag{26}$$

Учитывая, что $E_{F1} - E_{F2} = qV$, окончательно получаем обобщенный закон Ома (20).

Можно ли обосновать новые граничные условия (21)? Да, поскольку они следуют из обобщенного закона Ома (20), если предположить, что дополнительное сопротивление $\sigma A/\lambda$ (25) делится поровну между двумя границами проводника.

Лучшее обоснование можно достичь, если ввести два различных электрохимических потенциала E_F^+ и E_F^-, соответствующих движению электронов вдоль осей $+z$ и $-z$, соответственно. В главе 1 предполагалось, что оба контакта настолько массивны, что всегда находятся вблизи равновесия и описываются фермиевскими функциями (20/гл.1) и (21/гл.1) с хорошо определенными электрохимическими потенциалами. Сейчас же мы говорим о $E_F(z)$ в канале, не находящемся в равновесии, когда электронные состояния, переносящие электроны, заселены различно для электронов, движущихся вдоль направлений $+z$ и $-z$, в противном же случае тока не будет. Это различие в заселенности находит свое отражение в различии E_F^+ и E_F^-, и мы позже покажем, что ток пропорционален этой разности:

$$I = \frac{q}{h} M \left(E_F^+(z) - E_F^-(z) \right), \tag{27}$$

что можно переписать используя (235/гл.1) в виде

$$I = \frac{1}{qR^{Ball}} \left(E_F^+(z) - E_F^-(z) \right) = \frac{\sigma A}{q\lambda} \left(E_F^+(z) - E_F^-(z) \right). \tag{28}$$

Корректные граничные условия для E_F^+ и E_F^- следующие:

$$\begin{aligned} E_F^+(z=0) &= E_{F1}, \\ E_F^-(z=L) &= E_{F2}, \end{aligned} \tag{29}$$

которые можно понять из следующих соображений (рис. 5).

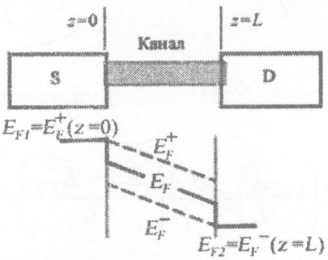


Рис. 5. Профиль электрохимических потенциалов E_F^+ и E_F^- в канале проводимости.

Электроны, генерируемые на границе $z=0$ в направлении $+z$, подчиняются фермиевскому распределению с потенциалом E_{F1}. Аналогично, электроны, генерируемые на границе $z=L$ в направлении $-z$, подчиняются фермиевскому распределению E_{F2} на правом контакте. Ток связан с потенциалами E_F^+ и E_F^- уравнениями

$$I=-\frac{\sigma A}{q}\frac{dE_F^+}{dz}=-\frac{\sigma A}{q}\frac{dE_F^-}{dz}, \tag{30}$$

которые эквивалентны уравнению диффузии (6), примененному к усредненному потенциалу

$$E_F(z)=\frac{E_F^+(z)+E_F^-(z)}{2}. \tag{31}$$

Уравнения (30) решаются с граничными условиями (29) и дают графики для E_F^+ и E_F^-, показанные на рис. 5, и их среднее значение действительно выглядит как на рис. 4 с соответствующими скачками потенциала на концах. И все же, нет нужды отказываться от традиционного уравнения диффузии (6) в пользу нового уравнения (30). Те же результаты можно просто получить модифицируя граничные условия для $E_F(z)$ с использованием уравнений (27) – (30) следующим образом для левого конца проводника

$$E_F(z=0)=\left(\frac{E_F^++E_F^-}{2}\right)_{z=0}=\left(E_F^+-\frac{E_F^+-E_F^-}{2}\right)_{z=0}=E_{F1}-\frac{qIR^{Ball}}{2} \tag{32}$$

и для правого конца

$$E_F(z=L)=\left(E_F^-+\frac{E_F^+-E_F^-}{2}\right)_{z=L}=E_{F2}+\frac{qIR^{Ball}}{2}. \tag{33}$$

Это в точности те же самые граничные условия для стандартного уравнения диффузии, что и выписанные выше (21).

7.3. Электрохимические потенциалы вдали от равновесия

Как уже упоминалось выше в отношении электрохимических потенциалов внутри контактов, оба контакта настолько массивны, что всегда находятся вблизи равновесия и описываются фермиевскими функциями с хорошо определенными электрохимическими потенциалами. Канал проводимости,

однако, не находится в равновесии, так что распределение электронов по доступным состояниям может и не описываться фермиевскими функциями.

В общем случае приходиться решать транспортное уравнение Больцмана (глава 5) [14, 15], а в квантовом случае использовать формализм неравновесных функций Грина (глава 10) [16 – 18] для получения соответствующих функций распределения $f(z,E)$. Можно ли представить эти распределения с использованием электрохимических потенциалов $E_F^+(z)$ и $E_F^-(z)$?

В канале с идеальной баллистической проводимостью использование $E_F^+(z)$ и $E_F^-(z)$ является строгим решением, а не приближенным. Все электроны, движущиеся от истока S в направлении $+z$ (рис. 6), подчиняются фермиевскому распределению на этом контакте с $E_F^+ = E_{F1}$:

$$f^+(z,E) = f_1(E) \equiv \frac{1}{e^{(E-E_{F1})/kT}+1}, \tag{34}$$

а все электроны стока D, движущиеся в направлении $-z$, подчиняются распределению на стоке с $E_F^- = E_{F2}$

$$f^-(z,E) = f_2(E) \equiv \frac{1}{e^{(E-E_{F2})/kT}+1}. \tag{35}$$

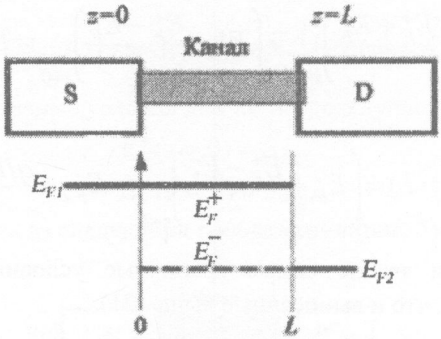


Рис. 6. Профили электрохимических потенциалов $E_F^+(z)$ и $E_F^-(z)$ в канале с идеальной баллистической проводимостью.

В дополнение к сказанному заметим, что связанные со стоком D моды, берущие свое начало на истоке S, заполнены только электронами, идущими из истока, так что эти моды остаются в равновесии с истоком с функцией распределения $f_1(E)$. Аналогично, связанные с истоком моды и берущие свое

начало на стоке находятся в равновесии со стоком с функцией распределения $f_2(E)$.

Пусть при некоторой энергии $f_1(E) = 1$ и $f_2(E) = 0$, так что множество электронов на истоке S готовы к транспорту на сток D, но ни один электрон на стоке D не готов к транспорту на исток S (рис. 7).

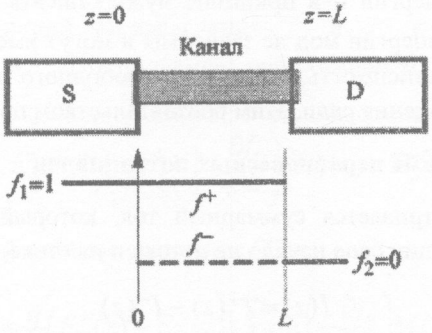


Рис. 7. Профили заселенности f^+ и f^- в канале с идеальной баллистической проводимостью.

Можно ожидать, что связанные со стоком моды, берущие свое начало на истоке, будут вплотную заполнены электронами (трафик «бампер-к-бамперу» на скоростном шоссе), тогда как связанные с истоком моды и берущие свое начало на стоке будут пустыми (трафик в обратном направлении отсутствует). Конечно, такая идеализированная модель баллистического канала предполагает, что в процессе транспорта электроны не возвращаются назад ни по ходу своей траектории, ни в ее конце. Именно это имеется в виду под баллистическим каналом с хорошими контактами, когда в канале есть достаточное число мод чтобы электроны легко покинули исток с практически нулевой вероятностью вернуться назад. Если же имеют место плохие контакты или транспорт в канале проводимости носит диффузионный характер, ожидать решение с функциями распределения (34) и (35) не приходится. В главе 9 при обсуждении спиновых вентилей будет показано к каким последствиям ведут плохие контакты. Сейчас же мы сосредоточимся на диффузионных каналах с хорошими контактами.

Функции распределения (34) и (35) представляются нам достаточно хорошими для диффузионного канала. Предполагается, что распределения подобны фермиевским, но учитывают пространственную зависимость электрохимических потенциалов:

$$f^+(z,E)=\frac{1}{e^{(E-E_F^+(z)/kT}+1},$$
$$f^-(z,E)=\frac{1}{e^{(E-E_F^-(z))/kT}+1}. \qquad (36)$$

Полноты ради заметим, что электрохимические потенциалы в общем случае зависят от энергии и в принципе нужно писать $E_F^+(z,E)$ и $E_F^-(z,E)$. В упругих резисторах энергии мод не зависимы и могут иметь свою характерную пространственную зависимость, если длина свободного пробега от энергии не зависит. В (36), упрощения ради, этим обстоятельством пренебрегается.

7.4. Токи в режиме неравновесных потенциалов

Обычно рассматривается суммарный ток, который представляет собой разность токов, берущих свое начало на истоке и на стоке,

$$I(z)=I^+(z)-I^-(z). \qquad (37)$$

Ток I^+ равен заряду, переносимому направо за единицу времени (рис. 8).

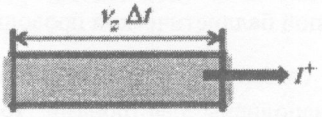


Рис. 8. К подсчету тока, берущему свое начало на истоке.

За временной интервал Δt заряд находится на длине $v_z\cdot\Delta t$, так что

$$I^+(z)=q\cdot(\text{число электронов на единице длины})\cdot v_z. \qquad (38)$$

Число электронов на единице длины равно половине плотности состояний на единице длины $D(E)/2L$, умноженной на долю f^+ занятых состояний, так что

$$I^+(z,E)=q\frac{D(E)}{2L}\bar{u}(E)f^+(z,E), \qquad (39)$$

где $\bar{u}$ есть среднее значение скорости v_z согласно уравнениям (220/гл.1) – (221/гл.1), а произведение $D(E)/2L$ на скорость есть $M(E)/h$ согласно (236/гл.1), так что

$$I^+(z,E)=\frac{qM(E)}{h}f^+(z,E) \qquad (40)$$

и аналогично

$$I^-(z,E)=\frac{qM(E)}{h}f^-(z,E). \qquad (41)$$

В итоге суммарный ток (37)

$$I(z)=\int_{-\infty}^{+\infty}dE\left(I^{+}(z,E)-I^{-}(z,E)\right)=\frac{q}{h}\int_{-\infty}^{+\infty}dE\left(f^{+}(z,E)-f^{-}(z,E)\right)M(E). \quad (42)$$

Для перехода от функций распределения f^{+} и f^{-} к электрохимическим потенциалам E_F^{+} и E_F^{-} воспользуемся приближением линейного отклика (25/гл.1):

$$f^{+}(z,E)-f^{-}(z,E)=\left(-\frac{\partial f_0}{\partial E}\right)\left(E_F^{+}(z)-E_F^{-}(z)\right), \quad (43)$$

так что из (42) получим искомое уравнение (27):

$$I(z)=\frac{q}{h}\left(E_F^{+}(z)-E_F^{-}(z)\right)\int_{-\infty}^{+\infty}dE\left(-\frac{\partial f_0}{\partial E}\right)M(E), \quad (44)$$

имея в виду, что стоящий справа интеграл есть термически усредненное число мод M.

7.5. Сопротивление R_{int} на границе контакта двух проводников с разным числом мод

Рассмотрим границу раздела между двумя проводниками с разным числом мод проводимости $M_1 > M_2$, граничащих с двумя массивными контактами на обоих концах, число мод в которых эффективно бесконечно велико (рис. 9), а также электрохимические потенциалы E_F^{+} и E_F^{-}, соответствующие движению электронов направо и налево, соответственно. Как показано выше, граничные условия имеют вид

$$E_F^{+}(L)=E_{F1} \text{ и } E_F^{-}(R)=E_{F2}. \quad (45)$$

Ток направо и налево одинаков и равен:

$$I=\frac{q}{h}M_1(E_F^{+}-E_F^{-})_L=\frac{q}{h}M_2(E_F^{+}-E_F^{-})_R. \quad (46)$$

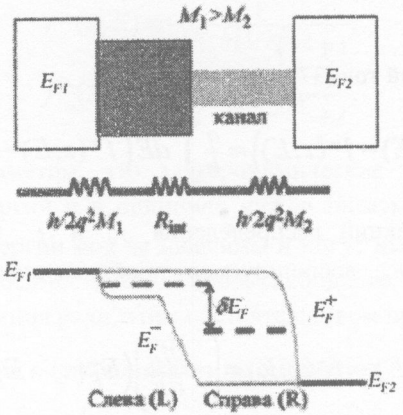


Рис. 9. Граница раздела между двумя каналами проводимости (широким и узким) с модами $M_1 > M_2$, граничащих с массивными контактами на обоих концах, число мод в которых эффективно бесконечно велико.

Электроны движутся свободно через границу раздела так, что движущиеся направо потоки в узком канале находятся в равновесии с движущимися направо потоками в широком канале:

$$E_F^+(R) = E_{F1}. \tag{47}$$

Движущиеся налево потоки в широком канале не могут быть адекватно заполнены узким каналом и соответствующий потенциал *a priori* не известен. Для его определения из (46) имеем

$$E_F^+(L) - E_F^-(L) = \frac{M_2}{M_1}\left(E_F^+(R) - E_F^-(R)\right). \tag{48}$$

Подставляя далее (45) и (47), получаем

$$E_F^-(L) = E_{F1} - \frac{M_2}{M_1}(E_{F1} - E_{F2}). \tag{49}$$

Для вычисления граничного сопротивления R_{int} нужно вычислить скачок потенциала на границе контакта двух проводников

$$\delta E_F = \left(\frac{E_F^+ + E_F^-}{2}\right)_L - \left(\frac{E_F^+ + E_F^-}{2}\right)_R. \tag{50}$$

Используя (45), (47) и (49), получаем

$$\delta E_F = \left(1 - \frac{M_2}{M_1}\right)(E_{F1} - E_{F2}), \tag{51}$$

$$I = \frac{q}{h} M_2 (E_{F1} - E_{F2}), \tag{52}$$

и окончательно получаем искомую формулу для граничного сопротивления

$$R_{int} \equiv \frac{\delta E_F / q}{I} = \frac{h}{2q^2}\left(\frac{1}{M_2} - \frac{1}{M_1}\right). \tag{53}$$

7.6. Уравнение диффузии как следствие уравнения Больцмана

При рассмотрении токов в режиме неравновесных потенциалов (§7.4) были введены функции распределения f^+ и f^- и соответствующие электрохимические потенциалы E_F^+ и E_F^-, которые описывают токи I^+ и I^-, берущие свое начало на истоке и стоке, соответственно (рис.10).

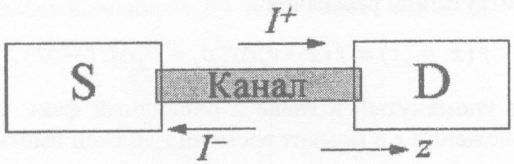


Рис. 10. Токи I^+ и I^-, порождаемые, соответственно, истоком S и стоком D.

Токи I^+ и I^-, а рассматриваем мы, как и прежде, упругие резисторы, формируются электронами, движущимися под разными углами по отношению к оси проводника и таким образом имеющих различные значения z-проекции импульса p_z, хотя полный импульс у них сохраняется и энергия электронов постоянна. Для учета влияния электрического поля на такой проводник нужны функции распределения $f^+(z, p_z, t)$ и $f^-(z, p_z, t)$ с учетом зависимости от импульса в явном виде. Такие функции даются решениями ТУБ

$$\frac{\partial f}{\partial t} + v_z \frac{\partial f}{\partial z} + F_z \frac{\partial f}{\partial p_z} = \hat{S} f \tag{54}$$

и показывают они заселенность состояний с заданными значениями скорости v_z и импульса p_z в точке проводника z в момент времени t, где F_z есть сила, действующая на электрон в тот же момент времени t, а под символом $\hat{S}$ подразумевается соответствующий оператор рассеяния, распределяющий электроны по состояниям со всеми возможными значениями скорости.

Если в уравнении (54) правую часть занулить (рассеяние отсутствует), а именно:

$$\frac{\partial f}{\partial t}+v_z\frac{\partial f}{\partial z}+F_z\frac{\partial f}{\partial p_z}=0, \tag{55}$$

то такое уравнение полностью эквивалентно описанию набора частиц с импульсами $p_z(t)$ в точках проводника $z(t)$, динамика которых описывается законами классической механики

$$v_z\equiv\frac{dz}{dt}=\frac{\partial E}{\partial p_z}, \tag{56}$$

$$F_z\equiv\frac{dp_z}{dt}=-\frac{\partial E}{\partial z}, \tag{57}$$

где $E(z,p_z,t)$ – полная энергия, с одним лишь отличием: если уравнения (56) и (57) описывают динамику одиночной частицы, положение и импульс которой меняются со временем, то в (55) все три переменных z, p_z, t равноправны и независимы друг от друга, в результате чего достигается описание коллективного транспортного процесса.

Для вывода ур-я (55) из уравнений (56) и (57) обратим внимание прежде всего на то, что в отсутствии рассеяния

$$f(z,p_z,t)=f(z-v_z\Delta t,p_z-F_z\Delta t,t-\Delta t), \tag{58}$$

что отражает уже упомянутый в главе 5 очевидный факт: любой электрон с импульсом p_z в положении z в момент времени t должен иметь импульс $p_z-F_z\Delta t$ в положении $z-v_z\Delta t$ в более ранний момент времени $t-\Delta t$.

Далее поступим просто: разложим (58) в ряд Тейлора и ограничимся лишь первым порядком:

$$f(z,p_z,t)=f(z,p_z,t)-\frac{\partial f}{\partial z}v_z\Delta t-\frac{\partial f}{\partial p_z}F_z\Delta t-\frac{\partial f}{\partial t}\Delta t, \tag{59}$$

сократим подобные и получим ТУБ в отсутствии рассеяния (55).

Далее для наших целей достаточно рассеяние $\hat{S}f$ в (54) учесть лишь в приближении времени релаксации (ВР) (глава 5) [7, 8]

$$\hat{S}f\cong-\frac{f-f_0}{\tau}, \tag{60}$$

которое предполагает, что эффект от влияния рассеяния просто пропорционален отклонению искомой функции распределения f от ее равновесного значения f_0. Такой подход надо считать квазиклассическим. И вот почему. Само ур-е (54) можно считать классическим в том смысле, что электроны рассматриваются в нем как частицы. Но оно не совсем классическое применительно к электронам, поскольку оператор рассеяния $\hat{S}$ в принципе

имеет квантовую природу, а кроме того в левой части уравнения фактически учитывается дисперсионное соотношение. Для графена, например, оно имеет вид

$$\vec{E} = \pm v_0 \vec{p},$$

следующий из его зонной структуры [9, 10], также имеющей квантовую природу. Если согласиться с квазиклассической природой транспортного уравнения Больцмана применительно к электронам, то многие транспортные свойства графена получают свое правдоподобное объяснение [11 – 16], чего сам Людвиг Больцман [17] в 1872 году предвидеть, конечно, не мог.

Из полного ТУБ (54) в приближении ВР (60) в условиях динамического равновесия ($\partial/\partial t = 0$)

$$v_z \frac{\partial f}{\partial z} + F_z \frac{\partial f}{\partial p_z} = -\frac{f - f_0}{\tau}. \tag{61}$$

Полная энергия в присутствии внешнего электрического поля

$$E(z, p_z) = \varepsilon(p_z) + U(z), \tag{62}$$

где дисперсионное соотношение $\varepsilon(p_z)$ испытывает сдвиг со стороны внешнего поля $U(z)$, как это показано на рис. 11.

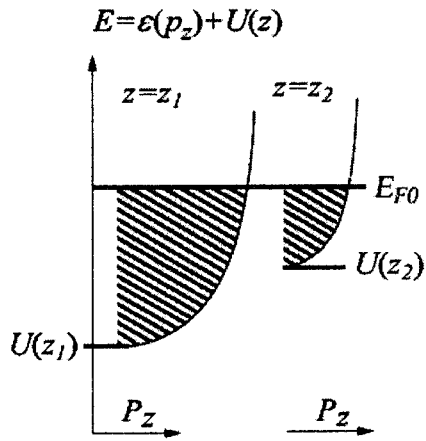


Рис. 11. Дисперсия при $U = 0$ ($z = z_1$) испытывает сдвиг при включении поля $U = U(z_2) - U(z_1)$. В состоянии равновесия электрохимический потенциал $E_{F0} = \text{Const}$.

Сначала обратим внимание на то, что равновесное распределение

$$f_0(z,p_z)=\frac{1}{e^{[E(z,p_z)-E_{F0}]/kT}+1} \tag{63}$$

при постоянном значении электрохимического потенциала E_{F0} удовлетворяет ТУБ (61). Правая часть (61) очевидно равна нулю. Убедимся, что и левая часть ур-я (13) также равна нулю. Для этого введем новую переменную

$$X_0 \equiv E(z,p_z)-E_{F0}=\varepsilon(p_z)+U(z)-E_{F0} \tag{64}$$

и учтем соотношения (56) и (57), а именно:

$$\begin{aligned} v_z\frac{\partial f_0}{\partial z}+F_z\frac{\partial f_0}{\partial p_z}&=\left(\frac{\partial f_0}{\partial X_0}\right)\left(v_z\frac{\partial X_0}{\partial z}+F_z\frac{\partial X_0}{\partial p_z}\right)= \\ &=\left(\frac{\partial f_0}{\partial X_0}\right)\left(v_z\frac{\partial E}{\partial z}+F_z\frac{\partial E}{\partial p_z}\right)=0. \end{aligned} \tag{65}$$

Ситуация в отсутствие равновесия качественно показана на рис. 12.

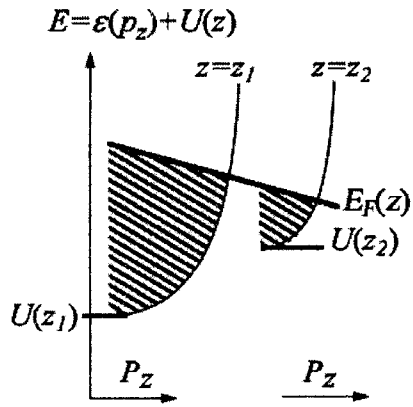


Рис. 12. В отсутствие равновесия электрохимический потенциал изменяется вдоль проводника: $E_F = E_F(z)$.

Пусть функция распределения в отсутствие равновесия имеет тот же вид (63), но с пространственно переменным электрохимическим потенциалом:

$$f(z,p_z)=\frac{1}{e^{[E(z,p_z)-E_F(z)]/kT}+1}. \tag{66}$$

Введем с учетом (24) новую переменную

$$X \equiv E(z,p_z) - E_F(z) = X_0(z,p_z) + E_{F0} - E_F(z), \tag{67}$$

в которой фигурирует отличие текущего значения электрохимического потенциала $E_F(z)$ от его равновесного значения E_{F0}. С новой переменной левая часть ур-я (61) оказывается равной

$$\left(\frac{\partial f}{\partial X}\right)\left(v_z\frac{\partial X}{\partial z} + F_z\frac{\partial X}{\partial p_z}\right) = \left(\frac{\partial f}{\partial X}\right)\left(-v_z\frac{\partial E_F}{\partial z}\right), \tag{68}$$

в чем убедиться можно следующим образом: в левую часть (68) подставим вместо скорости и силы уравнения (56) и (57) и возьмем производные от новой переменной X (67); в результате сократится слагаемое $(\partial E/\partial z)(\partial E/\partial p_z)$ и останется лишь $-v_z dE_F/dz$.

В условиях малости отклонения электрохимического потенциала $E_F(z)$ от его равновесного значения E_{F0} переменные (67) и (64) близки друг к другу, так что правую часть ур-я (28) можно записать в виде

$$\left(\frac{\partial f}{\partial X}\right)_{X=X_0}\left(-v_z\frac{\partial E_F}{\partial z}\right), \tag{69}$$

чему равна левая часть исходного ур-я (61). В его правой части стандартно перейдем от функций распределения к электрохимическим потенциалам:

$$-\frac{f-f_0}{\tau} \cong \left(\frac{\partial f}{\partial X}\right)_{X=X_0}\frac{E_F(z)-E_{F0}}{\tau}. \tag{70}$$

Сравнивая левую (29) и правую (30) части исходного ТУБ (61), получаем искомое уравнение диффузии

$$v_z\frac{dE_F}{dz} = -\frac{E_F(z)-E_{F0}}{\tau}. \tag{71}$$

Введем раздельные электрохимические потенциалы E_F^+ и E_F^- для электронов, движущихся слева направо ($v_z > 0$) и справа налево ($v_z < 0$) (рис. 10):

$$\frac{dE_F^+}{dz} = -\frac{E_F^+ - E_{F0}}{v_z\tau}, \quad \frac{dE_F^-}{dz} = \frac{E_F^- - E_{F0}}{v_z\tau}. \tag{72}$$

Выбрав равновесное значение химического потенциала как полусумму $E_{F0} = \left(E_F^+ + E_F^-\right)/2$, имеем:

$$\frac{dE_F^+}{dz} = -\frac{E_F^+ - E_F^-}{\lambda} = \frac{dE_F^-}{dz}, \tag{73}$$

где $\lambda = 2v_x\tau$. Комбинируя (73) с уравнениями для тока (28) окончательно получаем диффузионные уравнения для тока (30), приведенные выше без доказательства.

Обратим внимание на методически важное обстоятельство: при выводе уравнений (73) мы в явном виде учитывали внешнее электрическое поле и показали, что его влияние оказывается нулевым: внешнее приложенное электрическое поле не влияет на низковольтную проводимость (режим линейного отклика). Вместе с тем, любые поля, встроенные в проводник, находящийся в условиях равновесия электрохимического потенциала, влияют на низковольтную проводимость. Интересное обсуждение этого важного вопроса и роли гетероструктур можно найти в [3].

7.7. Эффекты при распространении тока по квантовой проволоке

До сих пор мы рассматривали равновесный транспорт электронов от истока к стоку и показали, что для понимания низковольтной проводимости достаточно электрохимического потенциала E_F и нет нужды учитывать электростатический потенциал U.

Межэлементные соединения в интегральных схемах обычно металлические, часто медные или алюминиевые. В связи с продолжающейся миниатюризацией рассматривают и другие варианты, среди которых нужно в первую очередь упомянуть углеродные нанотрубки (УНТ) с металлической проводимостью. Именно для них была построена токовая модель линии передачи, основанная на жидкостной теории Люттинжера [18]. Сравнение медных и металлических УНТ межэлементных соединений выполнено в [19] на основе потенциальных оценок Бюрке [20, 21]. Именно эти пионерские работы Бюрке поставили следующий важный вопрос [22]. Традиционные металлические межэлементные соединения обычно моделируются линией передачи с магнитной индуктивностью L_M и электростатической емкостью C_E (рис. 13).

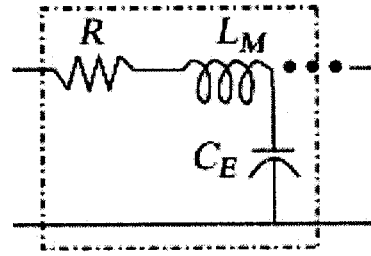


Рис. 13. Традиционная линия передачи: сопротивление и последовательно с ним индуктивность шунтируются емкостью между линией передачи и землей.

В жидкостной модели Люттинжера линии передачи дополнительно появляются кинетическая индуктивность L_K и квантовая емкость C_Q (рис. 14).

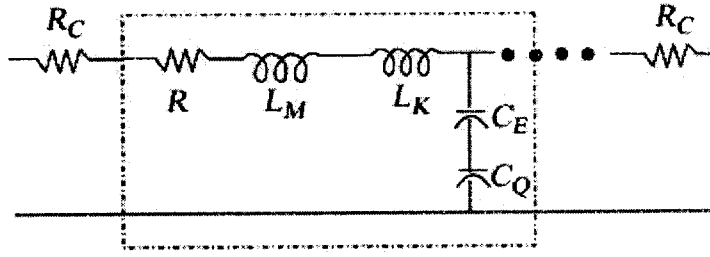


Рис. 14. Эквивалентная схема линии передачи в модели Люттинжера: вклад в суммарную индуктивность дают магнитная и кинетическая индуктивности, а в суммарную емкость – электростатическая и квантовая емкости.

Как связаны между собой эти две предельные модели? Если поперечное сечение медного межэлементного соединения существенно уменьшится, окажется ли существенным вклад L_K и C_Q? Насколько надо уменьшить поперечное сечение чтобы их вклад оказался существенным? Ответ на эти вопросы дан в модели [22], вывод основных уравнений которой мы рассмотрим достаточно подробно.

Речь пойдет о транспортных эффектах распространения переменного тока по нанопроволоке вдали от контактов, где мы зададимся локальными напряжением $V(z,t)$ и током $I(z,t)$ (рис. 15).

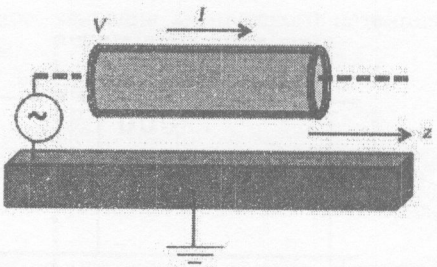


Рис. 15. Динамический транспорт по нанопроволоке задается локальными значениями $V(z,t)$ и $I(z,t)$.

Эта задача для нас интересна тем, что потребуется воспользоваться различием между раздельными электрохимическими потенциалами E_F^+ и E_F^- (72) и (73) с тем, чтобы получить корректные выражения для индуктивности и емкости.

Начнем опять с ТУБ с приближении ВР (60), но теперь уже откажемся от динамического равновесия ($\partial/\partial t = 0$):

$$\frac{\partial f}{\partial t} + v_z \frac{\partial f}{\partial z} + F_z \frac{\partial f}{\partial p_z} = -\frac{f - f_0}{\tau},$$

линеаризуем его имея в виду что теперь функция распределения, по сравнению с (66), имеет вид

$$f\left(z, p_z, t\right) = \frac{1}{e^{[E(z,p_z,t) - E_F(z,t)/kT]} + 1}. \tag{74}$$

По сравнению со стационарным уравнением диффузии (71) теперь мы имеем слева два дополнительных слагаемых:

$$\frac{\partial E_F(z,t)}{\partial t} + v_z \frac{dE_F(z,t)}{dz} - \frac{\partial E(z,t)}{\partial t} = -\frac{E_F(z,t) - E_{F0}}{\tau}, \tag{75}$$

определяющих динамику E и E_F.

Как и в случае стационарного уравнения диффузии (71), расщепим это уравнение на два уравнения для потенциалов E_F^+ и E_F^-, сумма и разность которых дают напряжение и ток, так что окончательно получаем следующую систему уравнений:

$$\frac{d(E_F/q)}{dz} = -(L_M + L_K)\frac{\partial I}{\partial t} - \frac{I}{\sigma A}, \tag{76}$$

$$\frac{d(E_F/q)}{dt} = -\left(\frac{1}{C_E} + \frac{1}{C_Q}\right)\frac{\partial I}{\partial z}, \tag{77}$$

которая выглядит как система уравнений линии передачи с последовательным соединением магнитной и кинетической индуктивностей и электростатической и квантовой емкостей, шунтирующих линию передачи (рис. 16).

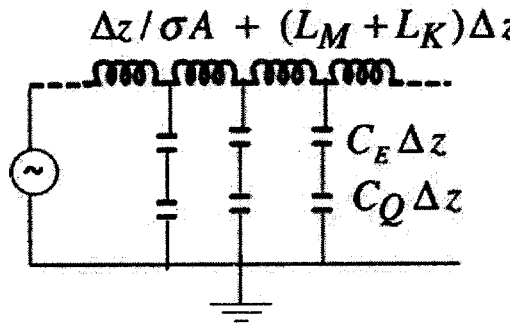


Рис. 16. Эквивалентная схема линии передачи (рис. 7) с учетом L_K и C_Q.

Приведем более подробные выкладки при переходе от (75) к (76) и (77). Итак, расщепляем уравнение (75) на два уравнения:

$$\frac{\partial E_F^+}{\partial t} + v_z \frac{dE_F^+}{dz} - \frac{\partial E^+}{\partial t} = -\frac{E_F^+ - E_{F0}}{\tau}, \tag{78}$$

$$\frac{\partial E_F^-}{\partial t} - v_z \frac{dE_F^-}{dz} - \frac{\partial E^-}{\partial t} = -\frac{E_F^- - E_{F0}}{\tau}. \tag{79}$$

Складывая и вычитая (78) и (79) и используя

$$I = \frac{q}{h} M (E_F^+ - E_F^-),$$

$$E_F = \frac{1}{2}(E_F^+ + E_F^-),$$

где M – число мод проводника, получаем

$$2\frac{\partial E_F}{\partial t} + \frac{v_z}{qM/h}\frac{\partial I}{\partial z} - \frac{\partial}{\partial t}(E^+ + E^-) = 0,$$

$$\frac{1}{qM/h}\frac{\partial I}{\partial t} + 2v_z \frac{\partial E_F}{\partial z} - \frac{\partial}{\partial t}(E^+ - E^-) = -\frac{I}{(qM/h)\tau}.$$

После перегруппировки имеем

$$\frac{\partial(E_F/q)}{\partial t} = -\frac{1}{C_Q}\frac{\partial I}{\partial z} + \frac{1}{2q}\frac{\partial}{\partial t}(E^+ + E^-), \tag{80}$$

$$\frac{\partial(E_F/q)}{\partial z} = -L_K\frac{\partial I}{\partial t} + \frac{1}{2qv_z}\frac{\partial}{\partial t}(E^+ - E^-) - RI, \tag{81}$$

где квантовая емкость C_Q и кинетическая индуктивность L_K (на единицу длины проводника) возникают как транспортные эффекты:

$$C_Q = \frac{q^2}{h}\left\langle\frac{2M}{v_z}\right\rangle, \tag{82}$$

$$L_K = \frac{h}{q^2}\left\langle\frac{1}{2Mv_z}\right\rangle, \tag{83}$$

тогда как C_E и L_M являются обычными электростатической емкостью и магнитной индуктивностью, следующими из уравнений электростатики и магнитостатики.

Обратим внимание на следующее обстоятельство. Если не учитывать поля, входящие в выражение для $E(z,t)$, мы лишимся производной $\partial E/\partial t$ в (75), а именно:

$$\frac{\partial E_F(z,t)}{\partial t} + v_z\frac{dE_F(z,t)}{dz} = -\frac{E_F(z,t) - E_{F0}}{\tau},$$

и после преобразований, аналогичным проведенным выше, получим

$$\frac{d(E_F/q)}{dz} = -L_K\frac{\partial I}{\partial t} - \frac{I}{\sigma A}, \tag{84}$$

$$\frac{d(E_F/q)}{dt} = -\frac{1}{C_Q}\frac{\partial I}{\partial z}. \tag{85}$$

Для каких проводников эти уравнения будут справедливы? Только для нанопроводников с очень малыми значениями поперечного сечения, следствием чего у них будут небольшие значения числа мод проводимости M, что в свою очередь, согласно (82) и (83), ведет к $L_K \gg L_M$ и $C_Q \ll C_E$, т.е. к уравнениям (84) и (85) вместо (76) и (77).

Если изучать проводимость только в условиях равновесия ($\partial/\partial t = 0$), то обе пары уравнений (84) – (85) и (76) – (77) сводятся к обычным уравнениям диффузии (6) и (15). Как уже неоднократно указывалось, при изучении проводимости в режиме линейного отклика нет нужды учитывать внешнее приложенное электрическое поле.

Вернемся к уравнениям (24) и (25) и рассмотрим слагаемые, содержащие энергии $E^{\pm}$. Энергия $E(z,p_z,t)$, ассоциируемая с поперечной электромагнитной

волной, может быть выражена через $U(z,t)$ и векторный потенциал $A_z(t)$, ориентированный вдоль оси распространения волны z, а именно:

$$E=\frac{\left(p_z-qA_z(t)\right)^2}{2m}+U(z,t).$$

Учитывая, что

$$v=\frac{\partial E}{\partial p_z}=\frac{p_z-qA_z}{m},$$

можно записать

$$\frac{\partial E}{\partial t}=v_z\left(-q\frac{\partial A_z}{\partial t}\right)+\frac{\partial U}{\partial t},$$

так что

$$\frac{\partial}{\partial t}(E^+ + E^-)=2\frac{\partial U}{\partial t},$$

$$\frac{\partial}{\partial t}(E^+ - E^-)=2v_z\left(-q\frac{\partial A_z}{\partial t}\right).$$

Теперь уравнения (80) и (81) могут быть записаны в виде:

$$\frac{\partial(E_F/q)}{\partial t}=-\frac{1}{C_Q}\frac{\partial I}{\partial z}+\frac{\partial(U/q)}{\partial t},$$

$$\frac{\partial(E_F/q)}{\partial z}=-L_K\frac{\partial I}{\partial t}-\frac{\partial A_z}{\partial t}-RI,$$

что сводится к (76) и (77) используя уравнение непрерывности

$$\frac{\partial Q}{\partial t}+\frac{\partial I}{\partial z}=0,$$

и имея в виду, что

$$A_z=L_M I,\quad U/q=C_E Q.$$

7.8. Роль электростатики

Внешнее электрическое поле играет важную роль как только мы выходим за пределы режима линейного отклика. Роль внешнего электростатического потенциала и соответствующего электрического поля мы продемонстрируем на примере простейшего полевого нанотранзистора, вольт-амперная характеристика которого за пределами режима линейного отклика контролируется приложенным электростатическим потенциалом.

Ток в модели упругого резистора дается формулой (глава 1):

$$I = \frac{1}{q}\int_{-\infty}^{+\infty} dE\, G(E)\left(f_1(E) - f_2(E)\right).$$

При учете внешнего электростатического потенциала U

$$I = \frac{1}{q}\int_{-\infty}^{+\infty} dE\, G(E-U)\left(f_1(E) - f_2(E)\right), \qquad (86)$$

при этом на первых порах будем пользоваться точечной моделью канала переноса электронов, которая предполагает неизменность плотности состояний $D(E)$ при движении вдоль проводника. Позже перейдем к стандартной расширенной модели проводника, учитывающей градиент $D(E)$ при переходе вдоль проводника от истока к стоку.

7.8.1. Полевой транзистор

В простейшем варианте это трехтерминальное устройство (рис. 17) (исток/source/S – сток/drain/D – затвор/gate/G), при этом затвор отделен от канала проводимости диэлектрической прокладкой, так что ток через затвор не идет: потенциал затвора V_G позволяет контролировать ток $I(V_D)$ в канале проводимости.

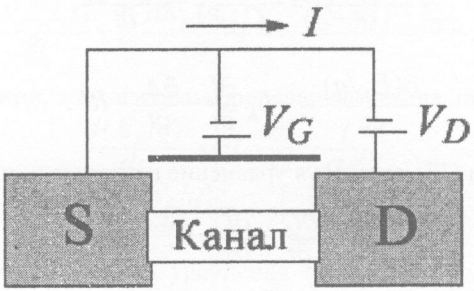


Рис. 17. Простейший трехтерминальный полевой транзистор.

Характерная вольт-амперная характеристика полевого транзистора показана на рис. 18.

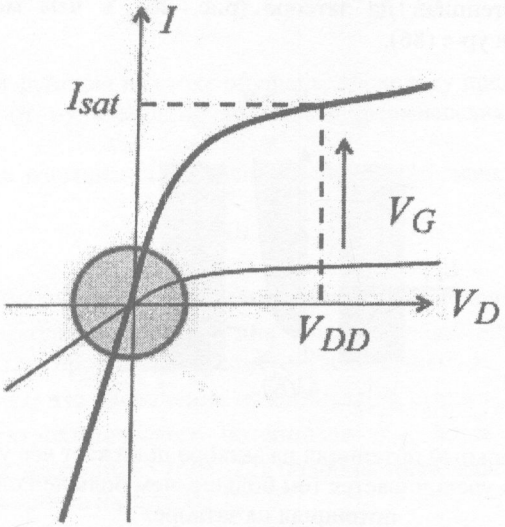


Рис. 18. Типичная вольт-амперная характеристика $I(V_D)$ полевого транзистора и ее изменение с ростом потенциала затвора V_G. Кружком показана область режима линейного отклика.

Все ясно с низковольтной проводимостью. Однако, с ростом V_D происходит нарушение закона Ома и появляются эффекты, требующие пересмотра и дополнения уже привычных представлений.

Пусть потенциал на затворе нулевой, а в щели $E_{F1} - E_{F2}$ нет электронных состояний проводника (рис. 19): тока нет.

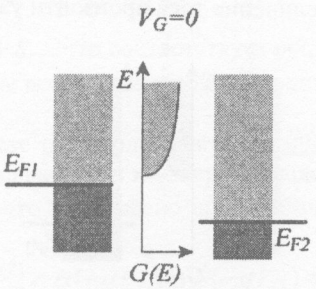


Рис. 19. В проводнике нет тока.

При подаче положительного потенциала на затвор $V_G > 0$ все уровни энергии проводника понижаются, ток увеличивается тем больше, чем больше

положительный потенциал на затворе (рис. 20), в чем можно убедиться подставляя $U = qV_G$ в ур-е (86).

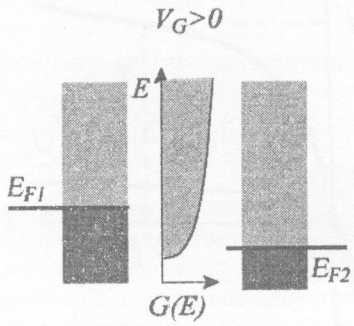


Рис. 20. Положительный потенциал на затворе понижает все уровни энергии проводника. Ток увеличивается тем больше, чем больше положительный потенциал на затворе.

Такова ситуация для *n*-проводников. Для проводников *p*-типа плотность состояний нисходящая с ростом энергии (глава 1) [23] и возрастание положительного потенциала на затворе ведет к противоположному эффекту по току. Далее будем рассматривать только проводники *n*-типа.

7.8.2. Ток стремится к насыщению

По мере роста напряжения V_D ток не возрастает линейно согласно закону Ома, а стремится к насыщению (рис. 18). Причина очевидна. Как только электрохимический потенциал стока E_{F2} опустился ниже дна зоны проводимости (рис. 21), увеличение тока произойти уже не может.

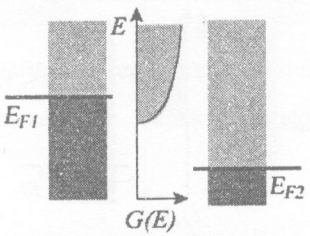


Рис. 21. Ток испытывает насыщение как только электрохимический потенциал стока E_{F2} оказывается ниже дна зоны проводимости.

Величина тока насыщения определяется из ур-я (86), а именно:

$$I_{sat} = \frac{1}{q}\int_{-\infty}^{+\infty} dE\, G(E-U) f_1(E), \tag{87}$$

где фермиевская функция на стоке опущена, поскольку предполагается, что для всех отличных от нуля мод проводимости фермиевская функция на стоке нулевая.

Если бы ток насыщения определялся бы только значением потенциала на затворе,

$$U^{(1)} = -qV_G, \tag{88}$$

то насыщение было бы полным (горизонтальная полочка) как только E_{F2} окажется на несколько kT ниже дна зоны проводимости. В реальных же экспериментах ток продолжает расти с увеличением V_D (рис. 22). Причина заключается в том, что увеличение потенциала на стоке V_D не только ведет к понижению электрохимического потенциала E_{F2}, но и к понижению всех уровней энергии проводника (рис. 22) подобно тому, как это происходит при увеличении потенциала на затворе V_G.

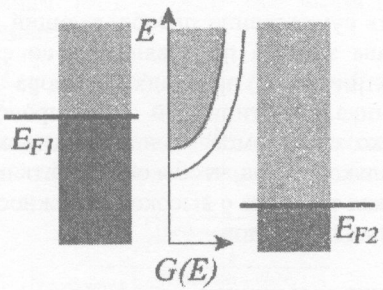


Рис. 22. При понижении E_{F2}, что соответствует росту потенциала на стоке V_D, понижаются также все уровни в зоне проводимости проводника.

В конечном результате ток продолжает несколько возрастать по мере того, как проводимость $G(E)$ ползет вниз по энергии на некоторую долю α ($\alpha < 1$) от потенциала на стоке V_D, что и отражено ниже в более точной модели для тока насыщения:

$$U^{(2)} = \alpha(-qV_D) + \beta(-qV_G) \equiv U_L. \tag{89}$$

Задача создания хорошего полевого транзистора сводится к тому, чтобы величину α сделать как можно меньшей и тем самым влияние потенциала на стоке V_D свести к минимуму. Если величина α окажется нулевой, ток выйдет

на полное насыщение (рис. 23), что и соответствует идеальному полевому транзистору: в таком транзисторе ток насыщения полностью определяется потенциалом на затворе V_G и никак не зависит от потенциала на стоке V_D.

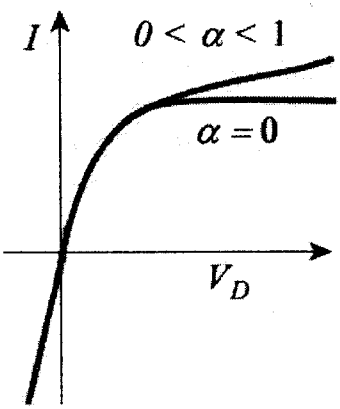


Рис. 23. Ток в полевом транзисторе испытывает полное насыщение, если он в режиме насыщения не подвержен влиянию потенциала на стоке V_D.

Чтобы гарантировать существенно преобладающий контроль над каналом проводимости со стороны затвора по сравнению со стоковым терминалом, нужно толщину диэлектрической прокладки затвора сделать существенно меньшей длины проводника. При типичной длине проводника в современных транзисторах в несколько сот атомов, толщина диэлектрической прокладки должна быть лишь несколько атомов, чтобы обеспечить нужное малое значение α. Поразительно, как такая толщина с высокой надежностью контролируется в производстве современных ноутбуков.

7.8.3. Роль заряжания проводника

Есть еще один эффект, который ведет к увеличению тока насыщения сверх того значения, которое получается при использовании (89) в (86). В проводнике под напряжением заселенность состояний проводника меньше по сравнению с заселенностью проводника в состоянии динамического равновесия, когда оба контакта стремятся максимально заполнить доступные состояния проводника. Тогда как под напряжением исток стремится пополнить электронами состояния проводника, а сток стремится электроны удалить из проводника. Уменьшение максимально возможной заселенности электронами состояний проводника под напряжением ведет к заряжанию проводника положительным зарядом, что в свою очередь ведет к понижению всех состояний проводника (рис. 22) и даже в случае идеального транзистора ($\alpha = 0$) увеличивает ток.

Учесть эффект заряжания в рамках точечной модели проводника можно путем добавления к потенциалу проводника U_L (89) еще одного слагаемого

$$U = U_L + U_0(N - N_0), \tag{90}$$

которое учитывает отличие равновесного числа электронов в проводнике N_0 от его неравновесного значения N в условиях нахождения проводника под напряжением, где U_0 есть вклад в потенциал проводника в пересчете на один электрон.

Равновесное число электронов в проводнике N_0 вычисляется через равновесное значение фермиевской функции $f_0(E)$:

$$N_0 = \int_{-\infty}^{+\infty} dE\, D(E-U) f_0(E), \tag{91}$$

а соответствующее неравновесное значение

$$N = \int_{-\infty}^{+\infty} dE\, D(E-U) \frac{f_1(E) + f_2(E)}{2} \tag{92}$$

в предположении, что проводник с обоими контактами взаимодействует одинаково хорошо.

Как только найден потенциал проводника U, удовлетворяющий одновременно уравнениям (90) – (92), ток вычисляется по (86) или по эквивалентному ему уравнению

$$I = \frac{1}{q} \int_{-\infty}^{+\infty} dE\, G(E)(f_1(E+U) - f_2(E+U)). \tag{93}$$

Вычисление потенциала проводника U, сводится к итерационному процессу, принципиальная блок-схема которого показана на рис. 24.

Рис. 24. Самосогласованная процедура вычисления потенциала проводника U и тока в точечной модели проводника.

Простая точечная модель проводника оказывается вполне приемлемой по сравнению с более строгими подходами [24].

Результаты численного расчета вольт-амперных характеристик (ВАХ) 2D проводника с параболической дисперсией и плотностью состояний [25 – 27]

$$D(E) = g\frac{mLW}{2\pi\hbar^2}H(E-E_c),$$

где $H(E-E_c)$ – ступенчатая функция Хэвисайда, показаны на рис. 25.

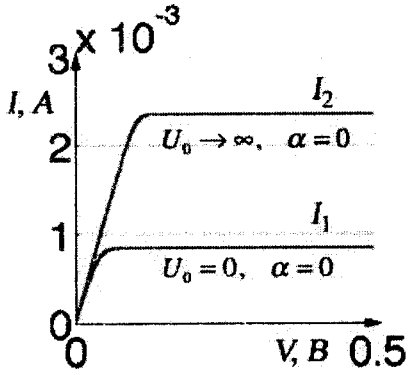


Рис. 25. ВАХ 2D проводника с параболической дисперсией в точечной модели проводника в режиме баллистического транспорта [3].

Вычисления проводились для $g=2$, $m=0.2\times9.1\times10^{-31}Kz$, $kT=5\,мэВ$, $\beta=1$, $\alpha=0$ и $U_0=0$ или $U_0=\infty$, $L=W=1\,\mu м$ в модели баллистического транспорта, так что

$$G(E)=\frac{q^2}{h}M(E),$$

а число мод проводимости (глава 1) [26, 27]

$$M(E)=g\frac{2W}{h}\sqrt{2m(E-E_c)}H(E-E_c).$$

Вычисленные ВАХ (рис. 25) имеют две характерных области: линейный рост и ток насыщения. Хотя эти графики и были получены численно, однако, и наклон и ток насыщения могут быть вычислены аналитически, особенно просто если принять низкотемпературное приближение, в котором фермиевские функции изменяются скачкообразно от 1 до 0 в узком промежутке, включающем $E=E_F$: действительно, в расчетах использовалось значение теплового кванта энергии $kT=5\,мэВ$ вместо его значения при комнатной

температуре $kT = 25\ мэВ$, так что приведенные численные ВАХ ближе сопоставимы с низкотемпературными оценками.

Этими графиками мы хотим проиллюстрировать два важных результата. Начальный наклон ВАХ не зависит от энергии заряжания U_0 (78). Он определяется низковольтной проводимостью. Именно по этой причине теорию проводимости в режиме линейного отклика мы имели право строить без учета электростатики (глава 1).

Во-вторых, ток насыщения сильно зависит от электростатики и увеличивается в ~ 2.8 раз при переходе от модели с нулевой энергией заряжания U_0 к модели с очень большим значением U_0. Причина этого фактически уже упоминалась выше. При нулевой энергии заряжания спектр состояний канала проводимости остается неизменным и число электронов N в канале проводимости равно $N_0/2$, поскольку $f_1 = 1$ и $f_2 = 0$. При очень большом значении U_0 во избежание доминирования $U_0(N - N_0)$ значение N должно быть практически равным его равновесному значению N_0 даже при наполовину заполненных состояниях. Для этого состояния в проводнике должны опуститься ниже при увеличении тока, как это показано на рис. 22.

7.8.3.1. Электростатический «выпрямитель»

Рассмотрим высококачественный полевой транзистор (ПТ) с $\alpha = 0$ как в стандартной конфигурации (рис. 26а), так и с плавающим затвором (рис. 26б).

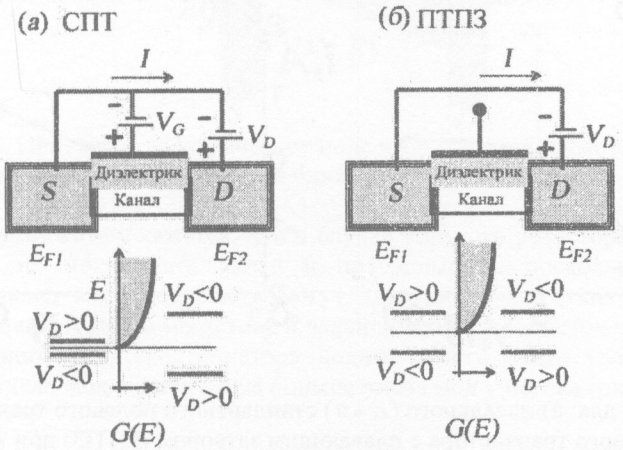


Рис. 26. Полевой транзистор: а) стандартный (СПТ);
б) с плавающим затвором (ПТПЗ).

Пусть контакты S и D в ПТ одинакового качества, иначе говоря, симметричны. В реальной ситуации это не так, однако, для нас это сейчас не важно, поскольку мы хотим качественно продемонстрировать два полезных

результата. В стандартной конфигурации ПТ (рис. 26а) имеет место электростатическая асимметрия, поскольку затвор G поддерживается при фиксированном потенциале V_G относительно истока S, но не относительно стока D, что и продемонстрировано на зависимости $G(E)$. В транзисторе же с плавающим затвором (рис. 26б) имеет место электростатическая симметрия, поскольку потенциал на затворе устанавливается посередине потенциалов на S и D. Мы хотим проиллюстрировать один существенный результат, а именно: контактно-симметричный ПТ может иметь асимметричную ВАХ (условно называемый «выпрямитель»), иначе говоря,

$$I(+V_D) \neq I(-V_D)$$

просто в результате электростатической асимметрии.

На рис. 27 показаны ВАХ, рассчитанные для приведенной выше точечной модели проводника (рис. 25) за исключением того что электрохимический потенциал расположен точно на дне зоны проводимости, как это показано на рис. 26: $E_{F0} = E_C$.

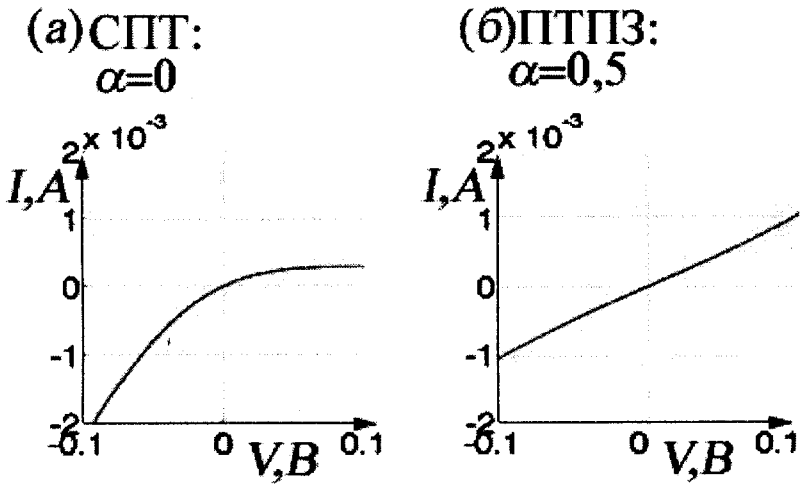


Рис. 27. ВАХ для: а) идеального ($\alpha = 0$) стандартного полевого транзистора (СПТ); б) полевого транзистора с плавающим затвором (ПТПЗ) при $\alpha = 0.5$ [3].

В случае стандартного ПТ полагалась $\alpha = 0$, что соответствует высококачественному ПТ, тогда как расчет ПТПЗ проводился с $\alpha = 0.5$. В первом случае имеет место «выпрямляющая» ВАХ (рис. 27а), а во втором случае – линейная, т.е. «омическая» ВАХ (рис. 27б). Вывод можно сделать

следующий: для получения асимметричной ВАХ нет нужды конструировать ПТ с асимметричными контактами. Даже симметричный по контактам ПТ может иметь несимметричную «выпрямляющую» ВАХ, если имеет место электростатическая асимметрия.

Обратим также внимание на то обстоятельство, что наклон ВАХ $dI/dV|_{V\to 0}$ не зависит от α, что видно из сравнения наклона ВАХ при $V=0$ на рис. 27а и рис. 27б, и может быть предсказан без учета энергии заряжания, т.е. без учета электростатики.

7.8.4. Расширенная модель проводника

Точечная модель проводника не учитывает градиент электрического поля в канале проводимости и предполагает, что плотность состояний $D(E)$ остается постоянной на всем протяжении проводника от истока до стока. В реальности электрическое поле понижает уровни энергии проводника вблизи стока в большей степени по сравнению с истоком, в результате чего наблюдается градиент плотности состояний $D(z, E)$ (рис. 28). Как это влияет на ток?

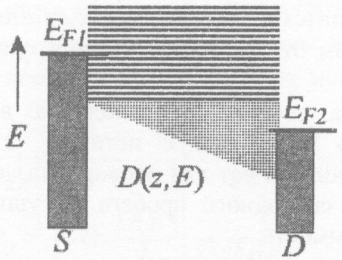


Рис. 28. Ненулевое электрическое поле в канале проводимости больше понижает состояния проводника ближе к стоку по сравнению с истоком.

На первый взгляд кажется, что в случае упругого резистора «косоидущие состояния» не могут дать вклад в проводимость, поскольку в упругом резисторе вклад в проводимость могут дать лишь состояния с постоянной энергией, начинающиеся на истоке и заканчивающиеся на стоке. Однако, даже для идеального упругого резистора должно иметь место увеличение тока, поскольку число мод вблизи стока больше числа мод у истока (рис. 29).

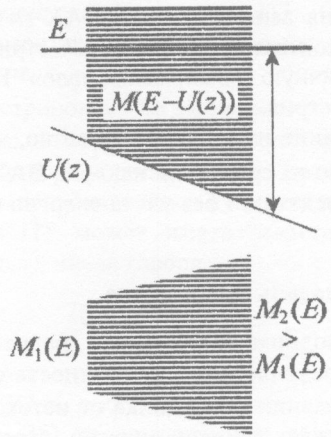


Рис. 29. Градиент электрического поля в канале проводимости приводит к большему числу мод вблизи стока по сравнению с истоком.

Вызвано это тем, что число мод с энергией E зависит от того, насколько далеко эта энергия находится от дна зоны проводимости, положение которого определяется потенциалом $U(z)$, значение которого ниже у стока по сравнению с истоком (рис. 29).

Структура спектра состояний проводника выглядит как бы более растянутой у стока по сравнению с истоком. Это не существенно для баллистического проводника, а вот для проводника, длина которого L намного больше средней длины свободного пробега λ, уширение спектра на стоке может увеличить проводимость.

В общем случае можно написать, что

$$\frac{q^2}{h}\frac{M_1\lambda}{L+\lambda} \le G(E) \le \frac{q^2}{h}M_1. \qquad (94)$$

Этот эффект не существенен для квазибаллистических упругих каналов, поскольку для таких каналов минимальные и максимальные значения проводимости по (81) фактически одинаковы. На самом деле увеличение проводимости можно приписать зависимости средней длины свободного пробега от электрического поля; этим можно пренебречь в режиме линейного отклика, чем мы фактически и пользовались.

Как учесть градиент электрического поля количественно? Для этого в диффузионном уравнении (2) учтем пространственное изменение удельной проводимости вдоль длины проводника:

$$\frac{I}{A} = -\frac{\sigma(z)}{q}\frac{dE_F}{dz}. \qquad (95)$$

Если ранее удельная проводимость усреднялась по энергии в пределах фермиевского окна проводимости шириной в несколько kT в окрестности $E = E_{F0}$:

$$\sigma = \int_{-\infty}^{+\infty} dE\,\sigma(E)\left(-\frac{\partial f}{\partial E}\right)_{E=E_{F0}}, \tag{96}$$

то теперь переменный вдоль длины проводника электростатический потенциал $U(z)$ сдвигает состояния проводника по энергии, в результате чего усреднение нужно вести в окрестности $E = E_F(z) - U(z)$:

$$\sigma = \int_{-\infty}^{+\infty} dE\,\sigma(E)\left(-\frac{\partial f}{\partial E}\right)_{E=E_F(z)-U(z)}. \tag{97}$$

Обратим внимание на то, что использование уравнений (95) и (97) требует вычисления $E_F(z) - U(z)$ из итерационной процедуры решения уравнения Пуассона

$$\frac{d}{dz}\left(\varepsilon\frac{dU}{dz}\right) = q^2(n - n_0), \tag{98}$$

где ε – относительная диэлектрическая проницаемость прокладки затвора, а n_0 и $n(z)$ – электронная плотность на единицу длины в канале проводимости в состоянии равновесия и вне равновесия, которые можно вычислить согласно определению, а именно:

$$n(z) \equiv \int_{-\infty}^{+\infty} dE\,\frac{D(E - U(z))}{L}\,\frac{1}{e^{(E-E_F(z))/kT}+1}, \tag{99}$$

$$n_0 = \int_{-\infty}^{+\infty} dE\,\frac{D(E)}{L}\,\frac{1}{e^{(E-E_{F0})/kT}+1}. \tag{100}$$

Принципиальная блок-схема вычисления тока в расширенной модели проводника показана на рис. 30.

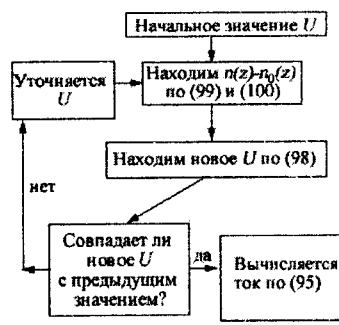


Рис. 30. Самосогласованная процедура вычисления потенциала проводника U и тока в расширенной модели проводника.

Нужно иметь в виду, что изложенный выше подход основан на модели локальных электрохимических потенциалах $E_F^{\pm}(z)$, описывающих транспорт электронов направо и налево, среднее значение которых фигурирует в уравнении диффузии (71). В общем же случае распределение электронов может настолько сильно отличаться от фермиевского, что концепция электрохимических потенциалов окажется неадекватной и потребуется решать транспортное уравнение Больцмана (глава 5), что является несравненно более сложной задачей. Несмотря на значительные успехи в этом направлении, уравнение диффузии, основанное на концепции локальных электрохимических потенциалов, по-прежнему остается главным инструментом в моделировании электронных устройств.

Концепция «снизу – вверх» расширяет область применимости уравнений диффузии (71) и (95). Ими можно пользоваться и для моделирования баллистических каналов, если только должным образом модифицировать граничные условия с целью учесть граничные сопротивления.

7.9. Контакты

Обобщенный закон Ома, применимой и для нанотранзисторов и для макропроводников,

$$R = \frac{\rho(L+\lambda)}{A} \tag{101}$$

содержит дополнительное сопротивление $\rho\lambda/A$, которое определяется только свойствами канала проводимости и не может быть устранено лучшим подбором терминальных контактов.

Как показано выше в этой главе, ключевым моментом в идентификации этого дополнительного сопротивления как граничного было различие между электрохимическими потенциалами E_F^+ и E_F^-, соответствующих модам проводимости слева – направо и справа – налево (рис. 31).

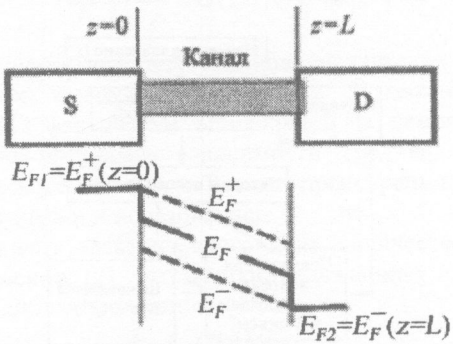


Рис. 31. Профиль электрохимических потенциалов E_F^+ и E_F^- в канале проводимости.

Из уравнений (20) и (28) и помня о том, что $E_{F1} - E_{F2} = qV$, для разности между потенциалами E_F^+ и E_F^- имеем:

$$\delta E_F \equiv E_F^+ - E_F^- = \frac{E_{F1} - E_{F2}}{1 + L/\lambda}. \qquad (102)$$

Контакты, поддерживаемые при различных электрохимических потенциалах E_{F1} и E_{F2}, выводят из равновесия моды проводимости, связанные с истоком и со стоком, тогда как процессы обратного рассеяния, контролируемые величиной λ, стремятся восстановить равновесие. Результат конкуренции этих двух тенденцій описывается уравнением (102).

Подчеркнем, что соотнесение различных электрохимических потенциалов к разным модам состояний проводника повсеместно используется при анализе и конструировании электронных устройств. Примером могут служить $p-n$ переходы.

7.9.1. $p-n$ переходы

На рис. 32 показано характерное изменение плотности состояний $D(z,E)$ в $p-n$ переходе. Положительный потенциал приложен справа, так что $E_{F2} < E_{F1}$.

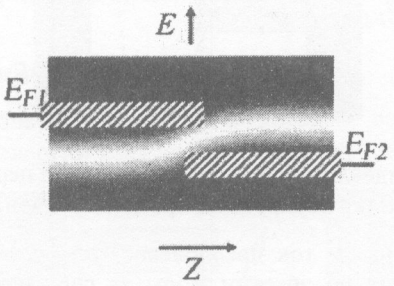


Рис. 32. Типичная картина пространственного изменения плотности состояний в $p-n$ переходе.

Левый контакт взаимодействует с модами проводимости шириной в несколько kT, показанными штриховкой слева в районе значения E_{F1}, а правый контакт – с модами проводимости, показанными штриховкой справа в районе значения E_{F2}. Можно представить себе идеализированную диаграмму этих двух групп мод, идущих от одного контакта, но не приходящих к другому контакту (рис. 33).

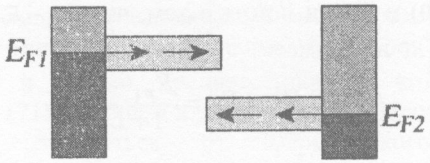


Рис. 33. Идеализированная модель $p-n$ перехода на рис. 32.

На самом деле ни левая, ни правая группы мод не изолированы полностью от противоположного контакта. Разработчикам электронных устройств для достижения максимально возможной изоляции часто приходится существенно удлинять проводник.

Возможен ли ток в подобном устройстве? Безусловно нет, если речь идет об упругом резисторе: нет канала проводимости, который позволил бы электрону пролететь баллистически от одного контакта к другому. Но ток действительно имеет место и обеспечивается он неупругими процесами в зоне перехода, которые позволяют электрону изменить свою энергию и достичь противоположного контакта (рис. 34).

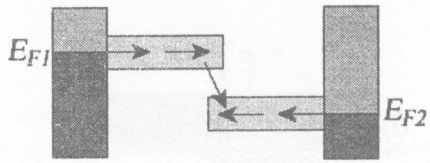


Рис. 34. Ток в идеализированной модели $p-n$ перехода (рис. 33) обеспечивается неупругими процесами в зоне перехода.

Именно таким образом ток преодолевает $p-n$ переходы: с верхних по энергии мод на нижние по энергии моды за счет неупругих RG-процессов (Recombination / Generation), название которых берет свое начало со все еще широко распространенных представлений об электронно-дырочных взаимодействиях в $p-n$ переходах.

Физически корректно обе группы мод рассматривать порознь путем введения разных для них электрохимических потенциалов, перечисляемых специальным индексом:

$$I_n = -\frac{\sigma_n}{q}\frac{dE_{Fn}}{dz}. \qquad (103)$$

Все токи увязаны друг с другом неупругими RG-процессами в области $p-n$ перехода

$$\frac{dI_n}{dz} = \sum_m \left\{ \left[RG\right]_{m\to n} - \left[RG\right]_{n\to m} \right\}, \qquad (104)$$

которые электроны из одной группы мод m перебрасывают в другую группу мод n и наоборот. Именно так моделируют $p-n$ переходы.

Ток в $p-n$ переходе дается выражением вида [25]

$$I = I_0(e^{qV/\nu kT} - 1), \tag{105}$$

где значения ν и I_0 определяются природой неупругих RG-процессов. Проводимости σ_n в (103) обеих групп мод практически не играют никакой роли в токе $p-n$ перехода (105). Физическая причина этого очевидна. Лимитирующую стадию тока $p-n$ перехода определяют неупругие процессы, перебрасывающие электроны из одной группы мод в другую группу, а транспорт электронов в пределах каждой из групп добавляет лишь дополнительное сопротивление в электронную схему соответствующего устройства. Ниже мы дадим вывод вольт-амперной характеристики $p-n$ перехода (105) для частного случая $\nu = 1$.

До сих пор мы рассматривали только лишь проводимость однородных p- и n-проводников. И этого может быть достаточно для анализа и моделирования многих электронных устройств, например, полевых транзисторов. Другая ситуация, однако, с устройствами, в которых используются $p-n$ переходы. Здесь иные вольт-амперные характеристики и фактически другая физика лежит в их основе.

7.9.2. Вольт-амперные характеристики

Рассмотрим устройство (рис. 35), в котором верхняя группа мод A кластеризуется вокруг энергии ε_A, а нижняя B – вокруг энергии ε_B.

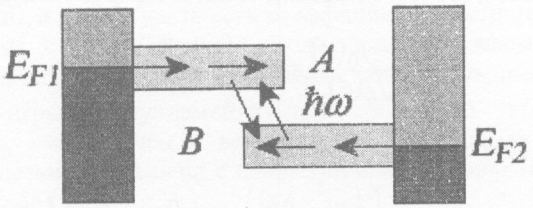


Рис. 35. Та же самая ситуация, что и на рис. 34, с двумя группами мод, обозначенными буквами A и B и увязанными неупругими взаимодействиями.

Физика подобных устройств описывается не уравнениями типа (103), а уравнениями (104), которые для двухуровневой ситуации (рис. 35) можно переписать в виде

$$I \sim D_{B\leftarrow A} f_A(\varepsilon_A)[1 - f_B(\varepsilon_B)] - D_{A\leftarrow B} f_B(\varepsilon_B)[1 - f_A(\varepsilon_A)], \tag{106}$$

где коэффициенты $D_{B\leftarrow A}$ и $D_{A\leftarrow B}$ характеризуют интенсивность неупругих процессов, индуцирующих переходы из A в B и наоборот из B в A.

Любопытно, что эти две скорости $D_{B\leftarrow A}$ и $D_{A\leftarrow B}$ в общем случае не одинаковы: $D_{A\leftarrow B}$ соответствует поглощению энергии в количестве

$$\hbar\omega = \varepsilon_A - \varepsilon_B, \tag{107}$$

тогда как $D_{B\leftarrow A}$ соответствует выделению такого же количества энергии. А фундаментальный принцип равновесия в статистической механике гласит: если система с неупругими взаимодействиями находится в равновесии при температуре T_0, то всегда труднее отнять энергию от такой системы, чем передать ей энергию, и отношение этих двух противоположных процессов дается выражением (помним о том, что направление перехода определяется от второго индекса к первому):

$$\frac{D_{A\leftarrow B}}{D_{B\leftarrow A}} = \exp\left(-\frac{\hbar\omega}{kT_0}\right). \tag{108}$$

Теперь для тока из (106) можно написать

$$I \sim D_{A\leftarrow B} f_B(\varepsilon_B)(1 - f_A(\varepsilon_A))(X - 1), \tag{109}$$

где

$$X \equiv \frac{D_{B\leftarrow A}}{D_{A\leftarrow B}} \frac{f_A(\varepsilon_A)}{1 - f_A(\varepsilon_A)} \frac{1 - f_B(\varepsilon_B)}{f_B(\varepsilon_B)}. \tag{110}$$

Используя (109), (110) и очевидное свойство фермиевской функции

$$\frac{1 - f_0(\varepsilon)}{f_0(\varepsilon)} = \exp\left(\frac{\varepsilon - E_{F0}}{kT}\right), \tag{111}$$

переписываем (110) следующим образом:

$$X = \exp\left(\frac{\hbar\omega}{kT_0} - \frac{\hbar\omega}{kT}\right) \exp\left(\frac{E_{FA} - E_{FB}}{kT}\right). \tag{112}$$

Поскольку группа мод А берет свое начало на контакте 1, а группа мод В – на контакте 2, и если работают слабые неупругие взаимодействия, перебрасывающие электроны с А на В, то можно считать, что моды А практически находятся в равновесии с контактом 1, а моды В – с контактом 2, так что

$$E_{FA} - E_{FB} \cong E_{F1} - E_{F2} = qV. \tag{113}$$

Если $T_0 = T$, то согласно (109), ток

$$I \sim (X-1) \sim e^{qV/kT} - 1. \qquad (114)$$

Таким образом, мы получили стандартное выражение для вольт-амперной характеристики $p-n$ перехода (105) для частного случая $\nu = 1$. Для значений ν, отличных от 1, требуется учитывать разнообразные *RG*-процессы в явном виде.

Обратим внимание на то обстоятельство, что устройства на $p-n$ переходах когут быть использованы для генерирования энергии. Если температура T_0 в зоне $p-n$ перехода отличается от температуры T, при которой поддерживается само устройство $(T_0 \neq T)$, то можно получить ток даже в отсутствие приложенного напряжения, например, в короткозамкнутой цепи (short circuit/sc)

$$I_{SC} \equiv I\Big|_{V=0} \sim \exp\frac{\hbar\omega}{k}\left(\frac{1}{T_0} - \frac{1}{T}\right) - 1. \qquad (115)$$

Таким образом, устройства на $p-n$ переходах принципиально пригодны для конвертации разности температур в электрический ток. Ток в короткозамкнутых цепях меняет направление на обратное при переходе от $T_0 > T$ к $T_0 < T$ [23].

Из уравнения (109) также следует, что в условиях незамкнутой ($I = 0$) цепи (open circuit/oc) должно быть $X = 1$, так что ур-е (112) дает

$$\frac{qV_{OC}}{\hbar\omega} = 1 - \frac{T}{T_0}. \qquad (116)$$

Левая часть равенства (116) есть доля энергии в пересчете на один фотон, сгенерированной в условиях пренебрежимо малого тока (практически незамкнутой цепи), а правая часть есть коэффициент полезного действия Карно тепловой машины в виде, скажем, солнечного фотоелемента. Поскольку обычно $T_0 \gg T$, то кпд такой солнечной ячейки может быть очень большим.

Контакты играют фундаментальную роль во всех устройствах наноэлектроники, как впрочем и во многих устройствах микроэлектроники. Получить ток в незамкнутой цепи с симметричными одинаковыми контактами невозможно (рис. 36).

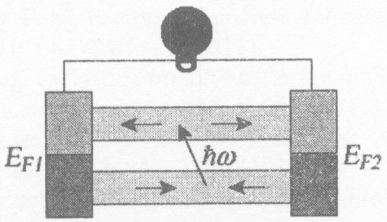


Рис. 36. Незамкнутая цепь с симметричными контактами генерировать ток не может.

Если в такой же цепи контакты асимметричны и дискриминированы (рис. 37), внешнее воздействие, например, солнечный свет, может электрон с мод B перебросить на одну из мод A и направить далее во внешнюю цепь через левый контакт, тогда как вакансия, образовавшаяся в одной из мод B заполнится электроном, прибывшим через нагрузку на правый контакт.

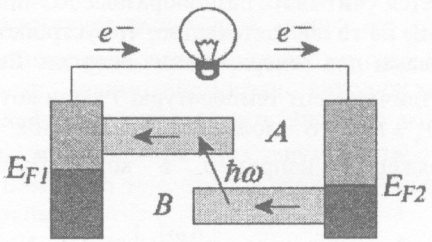


Рис. 37. Возможность получить ток во внешней цепи, если контакты асимметричны и дискриминированы как в $p-n$ переходах.

Именно асимметрия контактов в $p-n$ переходах принципиально отличает соответствующие устройства от, например, полевых транзисторов, что проявляется не только в вольт-амперных характеристиках, но и в самой физике $p-n$ переходов.

Подведем итоги. В рамках концепции «снизу – вверх» наноэлектроники рассмотрены диффузионно-дрейфовая модель тока на основе транспортного уравнения Больцмана, транспортные эффекты при распространении тока по квантовой проволоке роль внешнего электрического поля при выходе за пределы режима линейного отклика, полевой транзистор и ток насыщения, роль заряжания проводника, точечная и расширенная модели проводника, роль контактов, модели $p-n$ переходов и генерация тока в проводнике с асимметричными контактами [28, 29].

Литература

1. Ю. О. Кругляк, Н. Ю. Кругляк, М. В. Стріха. Уроки наноелектроніки: Виникнення струму, формулювання закону Ома і моди провідності в концепції «знизу – вгору», *Сенсор. електр. мікросист. техн.*, т. 9, № 4, 5 – 29 (2012).
2. Ю. А. Кругляк, Н. Е. Кругляк. Уроки наноэлектроники. 1. Причины возникновения тока в концепции "снизу – вверх", *Вестник Одес. гос. эколог. ун-та*, № 14, 197 – 206 (2012).
3. Suprio Datta. *Lessons from Nanoelectronics: A New Perspective on Transport*. Hackensack, New Jersey: World Scientific Publishing Company, 2012; www.nanohub.org/courses/FoN1.
4. A. Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen, *Ann. Physik*, v. 322, n 8, 549 – 560 (1905).
5. S. Lindsay. *Introduction to Nanoscience*. Oxford, England: Oxford University Press, 2009.
6. Н. Ашкрофт, Н. Мермин. Физика твердого тела. Москва: Мир, 1979.
7. Ю. А. Кругляк, Эффекты Холла, измерение электрохимических потенциалов и транспорт спинов в модели НРФГ, *ScienceRise*, т. 10, № 2(15), 35 – 67 (2015).
8. F. W. Sears, G. L. Salinger. *Thermodynamics, Kinetic Theory, and Statistical Thermodynamics*. Boston: Addison-Wesley, 1975.
9. Ю. А. Кругляк. Графен в транспортной модели Ландауэра – Датты – Лундстрома, *ScienceRise*, т. 2, № 2(7), 93 – 106 (2015).
10. Ю. А. Кругляк, Н. Е. Кругляк. Методические аспекты расчета зонной структуры графена с учетом σ-остова. Теоретические основы, *Вестник Одес. гос. эколог. ун-та*, № 13, 207 – 218 (2012).
11. M. Rabiu, S. Y. Mensah, S. S. Abukari. General Scattering Mechanism and Transport in Graphene, *Graphene*, v. 2, n 1, 49 – 54 (2013).
12. N. Bode, E. Mariani, F. von Oppen. Transport properties of graphene functionalized with molecular switches, *J. Phys.: Cond. Matter*, v. 24, 394017/1 – 10 (2012).
13. H. M. Dong, W. Xu, F. M. Peeters. High-field transport properties of graphene, *J. Appl. Phys.*, v. 110, 063704/1 – 6 (2011).
14. J. Chauhan, Guo Jing. Inelastic Phonon Scattering in Graphene FETs, *IEEE Trans. Electron Dev.*, v. 58, n 11, 3997 – 4003 (2011).
15. N. M. R. Peres. The transport properties of graphene. An introduction, *Rev. Mod. Phys.*, v. 82, n 3, 2673 – 2700 (2010).
16. A. Barreiro, M. Lazzeri, J. Moser, F. Mauri, A. Bachtold. Transport properties of graphene in the high-current limit, Phys. Rev. Lett., v. 103, 076601/1 – 4 (2009).
17. Людвиг Больцман. *Избранные труды*. Москва: Мир, 1984.

18. R. Tarkiainen R, M. Ahlskog, J. Penttila, P. Roschier, P. Hakonen, V. Paalanen, E. Sonin. Multiwalled carbon nanotube: Luttinger versus Fermi liquid, *Phys. Rev. B*, v. 64, n 19, 195412/1 – 4 (2001).
19. A. Naeemi, R. Savari, D. Meindl. Perfomance comparison between carbon nanotube and copper interconnects for GSI, *IEEE Intern. Elect. Dev. Meet. Tech. Dig.*, 699 – 702 (2004).
20. P. J. Burke. Lüttinger liquid theory as a model of the gigahertz electrical properties of carbon nanotubes, *IEEE Trans. Nanotech.*, v. 1, n 3, 129 – 144 (2002).
21. P. J. Burke. An RF circuit model for carbon nanotubes, *IEEE Trans. Nanotech.*, v. 2, n 1, 55 – 58 (2003).
22. S. Salahuddin, M. Lundstrom, S. Datta. Transport Effects on Signal Propagation in Quantum Wires, *IEEE Trans. Elect. Dev.*, v. 52, n 8, 1734 – 1742 (2005).
23. Ю. О. Кругляк, Н. Ю.Кругляк, М. В.Стріха, Уроки наноелектроніки. Термоелектричні явища в концепції «знизу – вгору», *Сенсор. електр. мікросист. техн.*, т. 10, № 1, 6 – 21 (2013).
24. A. Rahman, Guo Jing, S. Datta, M. Lundstrom. Theory of Ballistic Nanotransistors, *IEEE Trans. Elect. Dev.*, v. 50, n 9, 1853 – 1864 (2003).
25. R. F. Pierret. *Semiconductor Device Fundamentals*. Reading, MA: Addison–Wesley, 1996.
26. Ю. А. Кругляк. Обобщенная модель транспорта электронов и тепла Ландауэра – Датты – Лундстрома в микро- и наноэлектронике, *ScienceRise*, т. 5, № 3(5), 21 – 38 (2014).
27. Yu. A. Kruglyak, M. V. Strikha. Landauer – Datta – Lundstrom generalized electron transport model for micro- and nanoelectronics, *Proc. 2015 IEEE 35th Intern. Conf. Electronics Nanotech.* (ELNANO), April 21 – 24, 2015, Kyiv, Ukraine, pp. 70 – 74; DOI: 10.1109/ELNANO.2015.7146837.
28. Ю. О. Кругляк, М. В. Стріха. Уроки наноелектроніки: Роль електростатики й контактів у концепції «знизу – вгору», *Сенсор. електр. мікросист. техн.*, т. 11 № 4, 27 – 42 (2014).
29. Ю. А. Кругляк, Л. В. Ременяк. Роль электростатики и контактов в концепции наноэлектроники «снизу – вверх», *Вестник Одес. гос. эколог. ун-та*, № 19, 182 – 193 (2015).

Глава 8. Термодинамика проводника с током

8.1. Термодинамика проводника с током с учетом окружающей среды

Рассмотрим потоки электронов и тепла в простейшем электронном устройстве, состоящем из проводника с контактами – истоком S с электрохимическим потенциалом E_{F1} при температуре T_1 и стоком D с электрохимическим потенциалом E_{F2} при температуре T_2. Среда, окружающая проводник, находится при температуре T_0 и обменивается с проводником только теплом, например, путем эмиссии или поглощения фононов или фотонов, но не электронами (рис. 1).

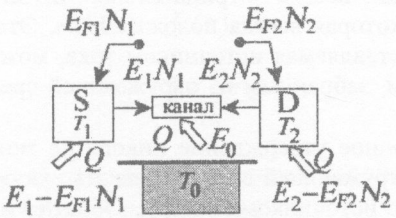


Рис. 1. Потоки электронов и тепла, показанные стрелками, в простейшем электронном устройстве.

Пусть N_1 электронов впрыскивается в проводник контактом 1 и N_2 – контактом 2. Числа электронов N_1 и N_2 не могут быть одного знака: инжектирование N_1 электронов в канал проводимости приводит к оттоку того же количества электронов через контакт 2, так что

$$N_1+N_2=0. \tag{1}$$

Полноты ради на рис. 1 показан также обмен теплом с окружающей средой. Если проводником будет упругий резистор, то его обмен теплом с окружающей средой отсутствует (§ 1.3.12) [1].

Закон сохранения энергии предполагает, что

$$E_1+E_2+E_0=0. \tag{2}$$

Это ограничение – первый закон термодинамики.

Есть, однако, еще одно ограничение, известное как второй закон термодинамики, а именно:

$$\frac{E_1-E_{F1}N_1}{T_1}+\frac{E_2-E_{F2}N_2}{T_2}+\frac{E_0}{T_0}\le 0. \tag{3}$$

Первый закон термодинамики описывается равенством, а второй закон допускает неравенство. Если температура обоих контактов совпадает с температурой окружающей среды $(T_1 = T_2 = T_0)$, то суммарное тепло, поглощенное проводником,

$$(E_1 - E_{F1}N_1) + (E_2 - E_{F2}N_2) + E_0 \le 0. \quad (4)$$

С использованием первого закона термодинаміки (2) это означает, что

$$E_{F1}N_1 + E_{F2}N_2 \ge 0. \quad (5)$$

Другими словами, полная энергия (2), которая сохраняется в этом процессе, состоит из двух частей: тепловой энергии со стороны окружающей среды (4), которая всегда отрицательна, и энергии (5), поставляемой источником тока, которая всегда положительна. Это же самое можно сказать иначе: энергия, поставляемая источником тока, может быть конвертирована в тепло, а вот теплом, забранным из окружающей среды, подзарядить источник тока невозможно.

Сформулированное утверждение никого не может удивить. Если бы мы могли теплом из окружающей среды заряжать аккумулятор, у нас не было бы никаких проблем с источником энергии. Обратим внимание на то, что это не противоречило бы первому закону термодинамики: суммарная энергия сохранилась бы и была бы прежней. Это второй закон термодинамики делает различие между энергией, накопленной в аккумуляторе, и тепловой энергией окружающей среды. Первую легко конвертировать во вторую, диссипировать в виде тепла, но не наоборот, поскольку тепловая энергия распределена по многим степеням свободы. Мы можем взять энергию, запасенную в состоянии с одной степенью свободы, и распределить ее по многим степеням свободы, но обратное действие невозможно – энергию, распределенную по многим степеням свободы, невозможно сконцентрировать в состоянии, имеющем одну степень свободы. Подобное интуитивное ощущение нашло свое отражение во втором законе термодинамики (3), имеющем неопровержимые экспериментальные подтверждения.

Упругий резистор не обменивается теплом с окружающей средой ($E_0 = 0$) и второй закон термодинамики (3) для него имеет вид:

$$\frac{\varepsilon - E_{F1}}{T_1}N_1 + \frac{\varepsilon - E_{F2}}{T_2}N_2 \le 0, \quad (6)$$

где, имея в виду одноуровневый резистор, энергии $E_1 = \varepsilon N_1$ и $E_2 = \varepsilon N_2$, так что электрон, входящий и выходящий из резистора, имеет энергию ε. С учетом материального баланса (1) это означает, что

$$\left(\frac{\varepsilon - E_{F1}}{T_1} - \frac{\varepsilon - E_{F2}}{T_2}\right)N_1 \le 0. \quad (7)$$

Физика упругого резистора всегда удовлетворяет этому условию, поскольку ток определяется разностью фермиевских функций контактов $f_1 - f_2$, где

$$f_1(\varepsilon) = \frac{1}{e^{(\varepsilon - E_{F1})/kT} + 1}, \tag{8}$$

$$f_2(\varepsilon) = \frac{1}{e^{(\varepsilon - E_{F2})/kT} + 1}. \tag{9}$$

Число электронов N_1 принимает положительное значение, указывая на поток электронов от истока в сторону стока, если

$$f_1(\varepsilon) > f_2(\varepsilon), \tag{10}$$

$$\frac{\varepsilon - E_{F1}}{T_1} < \frac{\varepsilon - E_{F2}}{T_2}. \tag{11}$$

Аналогичным образом, N_1 принимает отрицательное значение, указывая на поток электронов от стока в сторону истока, если

$$\frac{\varepsilon - E_{F1}}{T_1} > \frac{\varepsilon - E_{F2}}{T_2}. \tag{12}$$

В любом случае соблюдается второй закон термодинамики для упругого резистора (7). Если же речь пойдет о диффузионном резисторе, предполагающем обмен энергией с окружающей средой, то естественно задаться вопросом каким образом гарантировать выполнение второго закона термодинамики?

8.2. Асимметрия процессов поглощения и излучения энергии

Ответ на поставленный вопрос состоит в следующем. Модель резистора должна быть такой, чтобы гарантировать, что для всех процессов, включающих обмен энергией и электронами резистора с контактом, поддерживаемым при электрохимическом потенциале E_F и температуре T, вероятность инжектирования электронов в канал проводимости $P(+E,+N)$ должна относиться к вероятности приема электронов из канала проводимости $P(-E,-N)$ (рис. 2) как

$$\frac{P(+E,+N)}{P(-E,-N)} = e^{-(E - E_F N)/kT}. \tag{13}$$

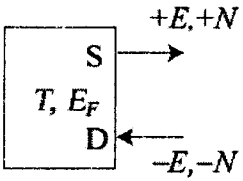


Рис. 2. Процессы поглощения проводником (инжектирование) и «излучения» им (прием контактом) энергии и электронов.

Если же речь идет только об обмене энергией, но не электронами, то

$$\frac{P(+E)}{P(-E)} = e^{-E/kT}. \tag{14}$$

Убедимся в том, что уравнения (13) и (14) гарантируют выполнимость второго закона термодинамики (3). Для этого рассмотрим ситуацию на рис. 1, где обмен электронами и/или энергией происходит в режиме трех «терминалов». Вероятность инжектирования электронов в канал проводимости пропорциональна произведению трех вероятностей

$$P(+E_1,+N_1)P(+E_2,+N_2)P(+E_0), \tag{15}$$

а вероятность обратного процесса пропорциональна

$$P(-E_1,-N_1)P(-E_2,-N_2)P(-E_0). \tag{16}$$

Доминирование инжектирования в канал над обратным процессом (ток идет от истока 1 к стоку 2) обеспечивается отношением этих двух вероятностей большим единицы:

$$\frac{P(+E_1,+N_1)P(+E_2,+N_2)P(+E_0)}{P(-E_1,-N_1)P(-E_2,-N_2)P(-E_0)} \geq 1. \tag{17}$$

Если все эти процессы подчиняются уравнениям (13) и (14), то

$$\exp\left(-\frac{E_1 - E_{F1}N_1}{kT_1}\right)\exp\left(-\frac{E_2 - E_{F2}N_2}{kT_2}\right)\exp\left(-\frac{E_0}{kT_0}\right) \geq 1, \tag{18}$$

откуда немедленно следует второй закон термодинамики (3), если вспомнить, что $\exp(-x)$ больше единицы, если x меньше нуля.

Обратим внимание на то, что равенство во втором законе термодинамики (3) соответствует бесконечно малому превалированию инжектирования в

резистор по сравнению с обратным процессом, так что суммарный ток инфинитезимально мал. Увеличение тока будет расти по мере усиления неравенства в (3).

Итак, почему мы уверены, что наша модель резистора удовлетворяет требованиям (13), (14)? Рассмотрим резистор с одним рассеивающим центром, слева от которого энергия электрона равна E_1, а справа равна $E_2 < E_1$ (рис. 3).

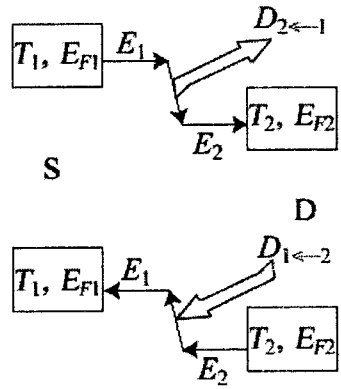


Рис. 3. Процессы выделения и поглощения энергии при переносе электронов с одного контакта на другой по проводнику с одним рассеивающим центром.

Вероятность процесса переноса электронов от контакта 1 до контакта 2 с выделением энергии есть

$$D_{2\leftarrow 1} f_1(E_1)(1-f_2(E_2)), \tag{19}$$

тогда как вероятность обратного процесса с поглощением энергии есть

$$D_{1\leftarrow 2} f_2(E_2)(1-f_1(E_1)), \tag{20}$$

где $D_{2\leftarrow 1}$ и $D_{1\leftarrow 2}$ – скорости соответствующих процессов. Поскольку в одном процессе речь идет о выделении энергии, а в другом – о поглощении энергии, то соответствующие скорости должны подчиняться закону (14), а именно:

$$\frac{D_{2\leftarrow 1}}{D_{1\leftarrow 2}} = e^{(E_1-E_2)/kT_0}, \tag{21}$$

каковым мы также уже пользовались ранее (108/гл.7), правда, для другой задачи [2]. Температура T_0 – это температура окружающей среды, с которой электроны обмениваются энергией.

Ток в таком неупругом резисторе

$$I \sim D_{2\leftarrow 1} f_1(E_1)(1-f_2(E_2)) - D_{1\leftarrow 2} f_2(E_2)(1-f_1(E_1)), \qquad (22)$$

и сводится к знакомому выражению для упругого резистора

$$I \sim (f_1 - f_2), \qquad (23)$$

только если

$$D_{2\leftarrow 1} = D_{1\leftarrow 2}, \qquad (24)$$

что соответствует упругому резистору с $E_2 = E_1$. На практике проводники представляют собой сложную смесь перекрывающихся упругих и неупругих «доменов» и не существует простых выражений, связывающих ток с фермиевскими функциями.

Итак, любая модель канала проводимости с учетом обмена энергией должна гарантировать соблюдение отношения скоростей выделения и поглощения энергии согласно закону (21), если окружающая среда находится в равновесии с проводником при температуре T_0. Любая транспортная теория, полуклассическая или квантовая, чтобы не нарушать второй закон термодинамики, должна обеспечивать выполнимость закона (21), должна учитывать энтропийные процессы наряду с механическими.

8.3. Энтропия

Асимметрия выделения и поглощения энергии привычна настолько, что мы и не замечаем ее. Если речь зайдет, скажем, об атоме водорода, и его единственный электрон пусть находится в возбужденном состоянии 2p, он немедленно вернется в основное состояние 1s с выделением энергии. Если же электрон находится в состоянии 1s, он в этом состоянии будет находиться бесконечно долго. Привычное объяснение этой ситуации заключается в том, что электрон, мол, стремится занять состояние с наименьшей энергией.

На самом деле в этом объяснении нет ничего естественного. Любое взаимодействие, классическое или квантовое, переводя электрон в основное состояние, с таким же успехом должно перевести его из основного в исходное возбужденное состояние. Кажущееся нам естественным стремление электрона занять состояние с наименьшей энергией имеет немеханическую природу, ни классическую, ни квантовую, а энтропийную, связанную с окружающей средой и отображаемую законами (13) и (14), согласно которым получить энергию из окружающей среды намного сложнее, чем излучить энергию в окружающую среду. При нулевой температуре система частиц может только излучать энергию, но не поглощать ее. Точно так же и электрон в возбужденном состоянии атома водорода может только излучать энергию, переходя в свое

основное состояние, характеризуемое наименьшей возможной энергией, но не поглощать ее, переходя в следующее, еще более возбужденное состояние.

Итак, каково происхождение предпочтительной тенденции терять энергию, а не приобретать ее из «среды»? Уравнение (14) можно понять, рассуждая следующим образом. Когда электрон теряет энергию в проводнике, контакт («среда») приобретает энергию, так что отношение скорости потери энергии к скорости приобретения энергии равно отношению плотности состояний при энергии $E_0+\varepsilon$ к плотности состояний при E_0 (рис. 4):

$$\frac{P(-\varepsilon)}{P(+\varepsilon)}=\frac{W(E_0+\varepsilon)}{W(E_0)}, \tag{25}$$

где $W(E)$ есть число состояний, доступных в области энергии E в контакте, которое связано с энтропией уравнением Больцмана

$$S=k\ln W, \tag{26}$$

так что

$$\frac{P(-\varepsilon)}{P(+\varepsilon)}=\exp\frac{S(E_0+\varepsilon)-S(E_0)}{k}. \tag{27}$$

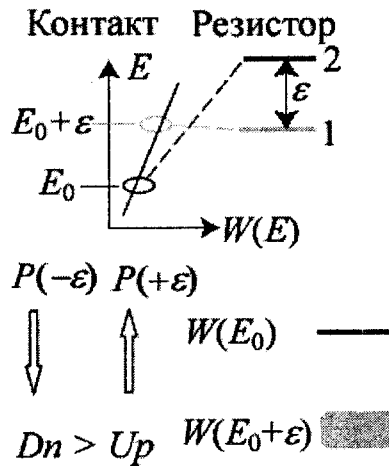


Рис. 4. Электроны предпочитают понижать свою энергию, поскольку выделяемая при этом энергия переходит в среду с большей плотностью состояний.

Считая обмениваемую энергию ε величиной малой по сравнению с энергией среды E_0, можно написать

$$S(E_0+\varepsilon)-S(E_0)=\varepsilon\left(\frac{dS}{dE}\right)_{E=E_0}=\frac{\varepsilon}{T} \tag{28}$$

с температурой, определяемой выражением

$$\frac{1}{T}=\left(\frac{dS}{dE}\right)_{E=E_0}. \tag{29}$$

Это, на самом деле, принципиально важный результат. Он показывает, что независимо от того как устроена окружающая среда, как только она находится в равновесии с изучаемой системой частиц, производная от энергии системы по энтропии есть температура системы. Если мы принимаем такое определение температуры, то уравнение Больцмана (26) становится основой, позволяющей рассматривать обмен энергией между системой и средой, находящихся в равновесии при температуре T:

$$\frac{P(-\varepsilon)}{P(+\varepsilon)}=e^{\varepsilon/kT}, \tag{30}$$

что уже утверждалось ранее в уравнении (14).

Если эмиссия энергии контактом сопровождается также эмиссией им электрона (рис. 2), который покидает контакт с энергией E_F, тогда ε нужно заменить на $\varepsilon-E_F$ в соответствии с уравнением (13).

8.3.1. Энтропия всегда растет

Энтропия определена. Теперь можно второй закон термодинамики (3) переписать иначе. Из рис. 1 следует, что

$E_1-E_{F1}N_1$ есть энергия, обмениваемая со средой при температуре T_1,

$E_2-E_{F2}N_2$ есть энергия, обмениваемая со средой при температуре T_2,

E_0 есть энергия, обмениваемая со средой при температуре T_0.

Опираясь на определение температуры по (28) и (29), эти энергии можно переписать через изменения энтропии:

$$(\Delta S)_1=\frac{-(E_1-E_{F1}N_1)}{T_1},\quad (\Delta S)_0=\frac{-E_0}{T_0},\quad (\Delta S)_2=\frac{-(E_2-E_{F2}N_2)}{T_2}. \tag{31}$$

Получились те же самые выражения, за исключением знака «минус», что и слагаемые в записи второго закона термодинамики в виде неравенства (3). Подставив эти выражения в (3), получим

$$(\Delta S)_1 + (\Delta S)_2 + (\Delta S)_0 \geq 0. \tag{32}$$

Окончательно, второй закон термодинамики в применении к рассматриваемому нами простейшему электронному устройству (рис. 1) утверждает, что при прохождении тока по проводнику суммарное изменение энтропии при транспорте электронов и обмене энергией с окружающей средой всегда положительно.

8.3.2. Свободная энергия всегда уменьшается

При нулевой температуре любая система частиц, достигая равновесия с окружающей средой, приходит в состояние с наименьшей возможной энергией. Как уже упоминалось, это связано с тем, что при нулевой температуре система частиц может только излучать энергию, но не поглощать ее. При ненулевой температуре оказывается возможным определить так называемую свободную энергию

$$F = E - TS \tag{33}$$

такую, что система частиц в состоянии равновесия с окружающей средой находится в состоянии с наименьшей свободной энергией. При нулевой температуре свободная энергия F совпадает с полной энергией системы E.

Убедиться в этом можно путем следующих рассуждений. Рассмотрим систему S, которая обменивается энергией с окружающей средой R, при этом полная энергия сохраняется. Выпишем первый и второй законы термодинамики для составной системы $S + R$:

$$(\Delta E)_S + (\Delta E)_R = 0, \tag{34}$$

$$(\Delta S)_S + (\Delta S)_R \geq 0. \tag{35}$$

Поскольку

$$(\Delta S)_R = \frac{(\Delta E)_R}{T}, \tag{36}$$

то подставляя (36) в (35) и комбинируя с (34), получаем

$$(\Delta E)_S - T(\Delta S)_S \equiv (\Delta F)_S \leq 0. \tag{37}$$

Таким образом, все процессы с обменом энергии, подчиняющиеся первому и второму законам термодинамики, вынуждены понижать свободную энергию системы, а прийдя в равновесие с окружающей средой, система оказывается в состоянии с наименьшей возможной свободной энергией.

8.4. Универсальный закон равновесия Больцмана

Тенденция предпочтительнее терять энергию, чем приобретать ее из окружающей среды, выраженная уравнениями (13) и (14), ведет к универсальному закону, который утверждает следующее: любая равновесная система, имеющая состояния i с энергиями E_i и числом частиц N_i, находится в этих состояниях с вероятностями

$$p_i = \frac{1}{Z} e^{-(E_i - E_F N_i)/kT}, \tag{38}$$

где Z есть константа, выбранная таким образом, чтобы все вероятности, будучи суммированы, дали единицу.

Чтобы убедиться в этом, обратимся к уравнению (13). Пусть мы имеем систему с двумя состояниями, в которой происходит обмен электронами и энергией с окружающей средой (рис. 5).

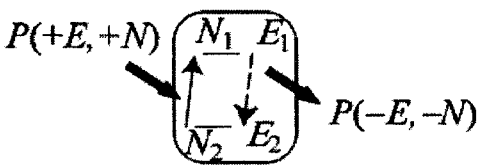


Рис. 5. Переходы в системе с двумя состояниями.

В состоянии равновесия требуется, чтобы переходы между состояниями уравновешивали друг друга, а именно:

$$p_2 \cdot P(+E,+N) = p_1 \cdot P(-E,-N). \tag{39}$$

Обращаясь к уравнению (13), получаем

$$\frac{p_1}{p_2} = \frac{P(+E,+N)}{P(-E,-N)} = e^{-((E_1 - E_F N_1) - (E_2 - E_F N_2))/kT} \tag{40}$$

и убеждаемся, что вероятности (38) удовлетворяют этому требованию, и таким образом правильно описывают состояние равновесия.

Чем вызвана универсальность закона равновесия (38)? Почему он распространяется на любые системы независимо от их внутреннего устройства? Потому что, как уже упоминалось, он связан со свойствами среды, окружающей систему, а не с самой системой.

Уравнение (38) есть основной постулат равновесной статистической механики, известный как уравнение Больцмана. Ричард Фейнман назвал его «вершиной»/«summit» [3]. Мы до сих пор пользовались распределениями Ферми (транспорт электронов) и Бозе (перенос тепла фононами (§ 2.3) [4]). Все три распределения описываются схожим образом:

$$f_i = \frac{1}{e^{(E_i - E_F)/kT} + \gamma}, \qquad (41).$$

где имеем статистику

Ферми – Дирака при $\gamma = +1$,
Бозе – Эйнштейна при $\gamma = -1$,
Максвелла – Больцмана при $\gamma = 0$.

Распределение Максвелла – Больцмана является предельным и для фермиевского и для бозевского распределений при $E - E_F(\hbar\omega) \gg kT$. В области энергий, сопоставимых с kT, эти оба распределения ведут себя существенно различным образом (рис. 6).

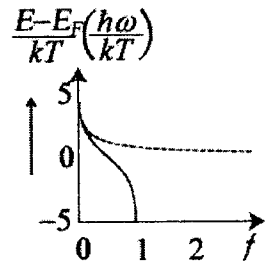


Рис. 6. Поведение функций Ферми (сплошная) и Бозе (пунктиром) в области энергий, сопоставимых с kT.

Фермиевская заселенность ограничена 0 и 1, что является следствием принципа запрета Паули, а бозевская заселенность ничем не ограничена кроме как числом частиц.

Важно отметить, что оба распределения – и фермиевское и бозевское – являются специальными случаями более общего закона Больцмана (38). Чтобы продемонстрировать это утверждение, нужно обратиться к концепции пространства Фока [5]. Дело в том, что энергии состояний E_i, фигурирующие в уравнении Больцмана (38), не являются энергиями одно-электронных состояний, как это обычно нами до сих пор предполагалось. Эти состояния являются состояниями в пространстве Фока, которые мы сейчас и рассмотрим, а затем получим фермиевское и бозевское распределения. Заметим лишь, что фоковское пространство лежит в основе техники вторичного квантования, используемой в современных методах кванговой химии [6 – 9], а операторы

вторичного квантования рождения и уничтожения исключительно удобны для описания электронной структуры молекул, например, полиенов [10, 11] и кумуленов [12].

8.5. Состояния в пространстве Фока

8.5.1. Фермионы

Мы уже фактически использовали фоковские состояния на рис. 5. Рассмотрим простую систему с одним уровнем энергии ε (рис. 7).

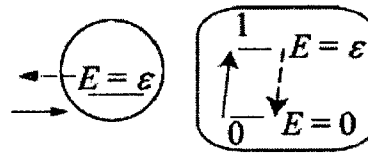


Рис. 7. Одноэлектронное состояние (слева) и фоковское состояние для электрона (справа).

В концепции одноэлектронных состояний электроны занимают или покидают это состояние. В концепции пространства Фока рассматриваются два возможных состояния этой системы: одно соответствует незанятому электроном состоянию с энергией $E=0$, а другое – занятому электроном состоянию с энергией $E=\varepsilon$. Нахождение системы в том или ином состоянии зависит от того, имеется ли электрон в системе или он покинул ее.

Применив закон Больцмана (38) к рассматриваемой системе в пространстве Фока, для вероятностей получаем:

$$p_0 = 1/Z; \quad p_1 = e^{-x}/Z, \tag{42}$$

где $x \equiv (\varepsilon - E_F)/kT$. Поскольку сумма вероятностей нормируется к единице, то

$$Z = 1 + e^{-x}, \tag{43}$$

$$p_0 = \frac{1}{e^{-x}+1} = 1 - f_0(\varepsilon), \tag{44}$$

$$p_1 = \frac{e^{-x}}{e^{-x}+1} = \frac{1}{e^{x}+1} = f_0(\varepsilon), \tag{45}$$

где правое равенство в (45) имеет место согласно (41). Вероятность p_1 того, что система занята электроном, есть просто функция Ферми, а вероятность p_0 того, что электрон покинул систему равна единице минус функция Ферми.

8.5.2. Бозоны

Распределение Бозе также следует из закона Больцмана (38). Теперь его нужно применить к системе, в которой число частиц (например, фотонов или фононов) может быть любым (рис. 8).

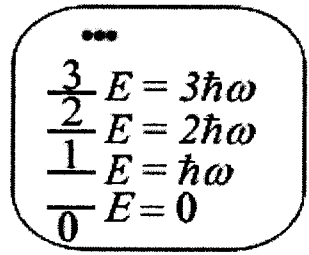


Рис. 8. Фоковские состояния для фононов и фотонов.

Согласно закону Больцмана, вероятность того, что система заполнена N бозонами

$$p_N = \frac{e^{-Nx}}{Z}, \tag{46}$$

где $x \equiv \hbar\omega / kT$. Из нормировки суммы всех вероятностей к единице имеем:

$$Z = \sum_{N=0}^{\infty} e^{-Nx} = \frac{1}{1-e^{-x}}, \tag{47}$$

так что среднее число бозонов

$$n = \sum_{N=0}^{\infty} N p_N = \frac{1}{Z} \sum_{N=0}^{\infty} N e^{-Nx}, \tag{48}$$

а с учетом того, что сумма справа

$$\sum_{N=0}^{\infty} N e^{-Nx} = -\frac{d}{dx} \sum_{N=0}^{\infty} e^{-Nx} = -\frac{d}{dx} Z, \tag{49}$$

окончательно получаем функцию Бозе (41)

$$n = \frac{1}{e^x - 1}. \tag{50}$$

Причина, по которой энергия $E - E_F$ фигурирует в фермиевской функции для электронов, а в случае фононов и фотонов фигурирует лишь $\hbar\omega$, а не $\hbar\omega - E_F$, связано со следующим. Когда электрон покидает проводник и входит в

контакт с энергией E, он релаксирует до энергии ε_F, характерной для этого контакта, а разница $E-E_F$ диссипирует. В случае же фононов и фотонов с энергией $\hbar\omega$, излученной или поглощенной, диссипирует вся их энергия $\hbar\omega$.

8.5.3. Взаимодействующие электроны

Закон равновесия Больцмана (38), в принципе, позволяет также описывать равновесное состояние сложной системы взаимодействующих частиц, если, конечно, удастся вычислить энергии соответствующих фоковских состояний. Рассмотрим систему из двух взаимодействующих одноэлектронных состояний (рис. 9).

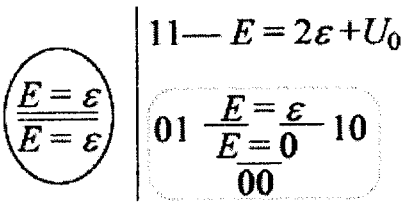


Рис. 9. Два взаимодействующих одноэлектронных состояния и соответствующие им четыре фоковских состояний.

Им соответствует четыре фоковских состояния 00, 01, 10 и 11. В последнем состоянии оба электрона взаимодействуют друг с другом с энергией U_0. Какова будет заселенность электронами системы, если она находится в равновесии с электрохимическим потенциалом E_F?

Пусть опять $x \equiv (\varepsilon - E_F)/kT$. Согласно закону Больцмана (38) имеем:

$$p_{00}=\frac{1}{Z};\quad p_{01}=p_{10}=\frac{e^{-x}}{Z};\quad p_{11}=\frac{e^{-2x}}{Z}e^{-U_0/kT}. \tag{51}$$

Среднее число электронов в системе

$$n=0\cdot p_{00}+1\cdot p_{01}+1\cdot p_{10}+2\cdot p_{11}=$$

$$=\frac{2}{Z}(e^{-x}+e^{-2x}e^{-U_0/kT}). \tag{52}$$

Эту задачу полезно рассмотреть для заданного значения энергии взаимодействия электронов, и мы здесь ограничимся лишь предельными случаями. Прежде всего для случая невзаимодействующих электронов с $U_0 \to 0$ имеем:

$$Z=1+2e^{-x}+e^{-2x}=(1+e^{-x})^2, \tag{53}$$

так что

$$n = 2\frac{1}{e^{(\varepsilon - E_F)/kT} + 1} \quad (U_0 \to 0) \tag{54}$$

есть просто удвоенное значение фермиевской функции, что и следовало ожидать, поскольку речь идет о двух невзаимодействующих состояниях.

Другой предельный случай – это сильно взаимодействующие электроны с $U_0 \to \infty$. В этом случае, согласно (51), $p_{11} = 0$, так что

$$Z = 1 + 2e^{-x}, \tag{55}$$

$$n = \frac{1}{\frac{1}{2}e^{(\varepsilon - E_F)/kT} + 1} \quad (U_0 \to \infty), \tag{56}$$

результат, который не представляется очевидным и следующим из распределения Ферми. Вместе с тем, можно показать, что в случае g одноэлектронных состояний

$$n = \frac{1}{g^{-1}e^{(\varepsilon - E_F)/kT} + 1} \quad (U_0 \to \infty), \tag{57}$$

результат, известный в физике полупроводников для подсчета числа локализованных состояний.

Закон равновесия Больцмана (38) и энтропия (26) формулируются в пространстве Фока. Мы показали, как закон Больцмана можно переформулировать через одноэлектронные состояния для невзаимодействующих электронов. Сейчас мы покажем как сделать то же самое для энтропии.

8.6. Альтернативное выражение для энтропии

Рассмотрим систему невзаимодействующих локализованных спинов, например, магнитных примесей в канале проводимости. В состоянии равновесия половина спинов случайным образом ориентирована вверх, а другая половина – вниз (рис. 10).

$$S = Nk\,\ell n 2$$

Рис. 10. Энтропия равновесной системы N локализованных спинов.

Выражение (26) определяет энтропию как $S = k \ln W$, где W есть полное число фоковских состояний, доступных рассматриваемой системе. В нашем случае каждому спину доступны два состояния – «вверх» и «вниз», так что N спинам доступны 2^N состояний: $W = 2^N$ и энтропия

$$S = k \ln W = Nk \ln 2 . \tag{58}$$

Все правильно, но есть еще альтернативное выражение для энтропии, которым можно пользоваться всегда, когда мы рассматриваем систему, состоящую из большого числа одинаковых невзаимодействующих подсистем, например, тех. же N локализованных спинов, а именно:

$$S = -Nk \sum_i \tilde{p}_i \ln \tilde{p}_i , \tag{59}$$

где $\tilde{p}_i$ есть вероятность найти отдельную подсистему в i-ом состоянии. В нашем примере отдельный спин характеризуется вероятностью ½ быть обнаруженным в состоянии «вверх» и «вниз», так что уравнение (59) дает

$$S = -Nk \left(\frac{1}{2} \ln \frac{1}{2} + \frac{1}{2} \ln \frac{1}{2} \right) = Nk \ln 2 , \tag{60}$$

в точности тот же результат, что и ранее уравнение (58).

Формула (59) более гибкая по сравнению с (26) в том смысле, что вероятности $\tilde{p}_i$ могут принимать произвольные значения для каждой из подсистем в i-ом состоянии. Не говоря уже о том, что она поразительно похожа на выражение для шенноновской энтропии, ассоциируемой с количеством информации в строке из N символов, каждый из которых может принимать различные значения i с вероятностью $\tilde{p}_i$. Мы позже вернемся к шенноновской энтропии и рассмотрим ее подробно применительно к электронике.

А сейчас покажем как формула (59) следует из формулы (26) [13]. Рассмотрим очень большое число N идентичных подсистем, каждая с набором состояний $\{E_i\}$, заселенных с вероятностями $\{\tilde{p}_i\}$ такими, что число этих подсистем в состоянии i дается выражением

$$N_i = N \tilde{p}_i . \tag{61}$$

Полное число возможностей получить набор подсистем $\{N_i\}$ будет

$$W = \frac{N!}{N_1! N_2! N_3! \cdots} . \tag{62}$$

Используя формулу Стирлинга

$$\ln n! \cong n \ln n - n , \tag{63}$$

получаем

$$\ln W = \ln N! - \ln N_1! - \ln N_2! - \ln N_3! - \cdots \cong$$

$$\cong N \ln N - N\tilde{p}_1 \ln N\tilde{p}_1 - N\tilde{p}_2 \ln N\tilde{p}_2 - N\tilde{p}_3 \ln N\tilde{p}_3 - \cdots . \qquad (64)$$

Учитывая нормировку всех вероятностей $\{\tilde{p}_i\}$ к единице, получаем

$$\ln W = -N\left(\tilde{p}_1 \ln \tilde{p}_1 + \tilde{p}_2 \ln \tilde{p}_2 + \tilde{p}_3 \ln \tilde{p}_3 + \cdots\right) = -N\sum_i \tilde{p}_i \ln \tilde{p}_i . \qquad (65)$$

что и устанавливает соответствие между формулами для энтропии (59) и (26).

8.6.1. Равновесное распределение из минимума свободной энергии

В общем случае система может находиться в произвольном состоянии (не обязательно равновесном), в котором каждый уровень энергии E_i заселен с вероятностью $\tilde{p}_i$. Для равновесного же состояния, согласно уравнению Больцмана (38),

$$\left[\tilde{p}_i\right]_{eq} = \frac{1}{Z} e^{-E_i/kT} \equiv p_i . \qquad (66)$$

Мы также уже знаем, что равновесное состояние характеризуется минимумом свободной энергии (33). Покажем, что из всех возможных выборов вероятностей $\{\tilde{p}_i\}$ только равновесные распределения $\{p_i\}$ минимизируют свободную энергию.

Энергия системы

$$E = \sum_i E_i \tilde{p}_i . \qquad (67)$$

Воспользуемся выражением для удельной энтропии S/N из определения энтропии по (59) и выпишем свободную энергию в виде

$$F = \sum_i \tilde{p}_i \left(E_i + kT \ln \tilde{p}_i\right), \qquad (68)$$

позволяющем минимизировать ее по изменениям в $\{\tilde{p}_i\}$, а именно:

$$dF = \sum_i d\tilde{p}_i \left(E_i + kT \ln \tilde{p}_i\right) = 0, \qquad (69)$$

если учесть, что

$$\sum_i d\tilde{p}_i = 0, \qquad (70)$$

поскольку сумма всех вероятностей фиксирована (нормирована на единицу).

Гарантировать же равенство dF нулю при произвольном выборе $d\tilde{p}_i$ можно лишь положив

$$E_i + kT \ln \tilde{p}_i = 0, \tag{71}$$

что и ведет нас к равновесным вероятностям (66).

Если же система не находится в равновесии, то можно воспользоваться выражением (68) для расчета свободной энергии в неравновесном состоянии, если нам только известны вероятности $\tilde{p}_i$, и получим большее значение свободной энергии по сравнению с ее равновесным значением.

В работе [14] утверждается, что можно построить такое устройство, которое избыток свободной энергии в неравновесном состоянии системы

$$\Delta F = F - F_{eq} \tag{72}$$

преобразует в полезную работу. Этот избыток состоит из двух частей:

$$\Delta F = \Delta E - T\Delta S, \tag{73}$$

где ΔE есть избыток реальной энергии, а вот второе слагаемое содержит изменение энтропии, которое можно связать с информацией, позволяющей конвертировать энергию из окружающей среды в полезную работу. Далее мы изложим эту «взрывную» концепцию болем подробно.

8.7. Глобальная ценность информации

Система в равновесном состоянии не содержит информации, поскольку состояние равновесия не зависит от предистории системы, оказавшейся в своем равновесном состоянии. Обычно информация содержится в системе, которая «застряла» в каком-то своем неравновесном состоянии. Утверждается, что имея такую неравновесную систему, можно построить такое устройство, которое может экстрагировать энергию, меньшую или равную

$$E_{avail} = F - F_{eq}, \tag{74}$$

где F есть свободная энергия неравновесной системы, F_{eq} есть свободная энергия системы, пришедшей в состояние равновесия. Приведем соображения в пользу такого утверждения.

Вернемся к рис. 1. Пусть температура обоих контактов одинакова и равна T, а среда находится в метастабильном состоянии, так что электроны в проводнике взаимодействуют с неравновесной средой. В этом случае о температуре среды говорить не приходиться.

Например, в системе невзаимодействующих спинов в равновесном состоянии при любой температуре половина спинов находится в состоянии «вверх», а другая половина – в состоянии «вниз» (рис. 11).

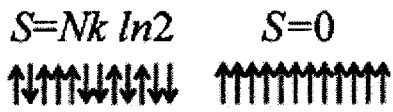


Рис. 11. Равновесная (слева) и неравновесная (справа) системы локализованных спинов.

Если же систему перевести в неравновесное состояние со всеми спинами «вверх», то невозможно говорить о температуре такой системы, а энтропию по (59) вычислить можно и она окажется в этом случае равной нулю.

Вернемся к нашему примеру со средой в неравновесном состоянии. Заменяя E_0/T_0 на $-\Delta S$, второй закон термодинамики (3) перепишем в виде

$$\frac{E_1-E_{F1}N_1}{T_1}+\frac{E_2-E_{F2}N_2}{T_2}-\Delta S\le 0, \qquad (75)$$

а закон сохранения энергии (2) требует

$$E_1+E_2=-E_0\equiv\Delta E, \qquad (76)$$

где ΔE есть изменение энергии окружающей среды.

При равенстве температур контактов ($T_1=T_2\equiv T$) и очевидном равновесии токов (1) с учетом (76) из второго закона термодинамики для среды в неравновесном состоянии (75) получаем

$$\left(E_{F1}-E_{F2}\right)N_1\ge\Delta E-T\Delta S=\Delta F. \qquad (77)$$

Обычно изменение свободной энергии ΔF может быть только положительным, поскольку система в равновесном состоянии при выходе из равновесия может только увеличить свободную энергию. Тогда (77) требует, чтобы число электронов N_1, инжектируемое левым контактом 1 в проводник (рис. 1) было одного знака с разностью электрохимических потенциалов $E_{F1}-E_{F2}$, подаваемых на проводник: электроны движутся от контакта с большим электрохимическим потенциалом к контакту с меньшим, как это и должно быть.

В нашем же случае ситуация необычная: система находится в неравновесном состоянии и стремится перейти в равновесное состояние, так что $\Delta F<0$ и N_1 может иметь знак, противоположный разности $E_{F1}-E_{F2}$, не нарушая при этом условие (77). Электроны могут переходить от контакта с меньшим электрохимическим потенциалом к контакту с большим потенциалом, как это происходит при зарядке аккумулятора.

Основной вывод следующий: метастабильное неравновесное состояние системы, в принципе, может быть использовано для конструирования

аккумулятора. Как, например, это имеет место в реальных аккумуляторах. Возьмем, например, литий-ионный аккумулятор. Заряженный аккумулятор находится в метастабильном состоянии с избытком ионов Li^+, интеркалированных в графитовый анод. Когда аккумулятор разряжается, поставляя электроны во внешнюю цепь, ионы Li^+ диффундируют на катод, содержащий, например, $LiCoO_2$. При зарядке аккумулятора идут обратные процессы. В качестве примера приведем окислительно-восстановительные реакции на положительном электроде

$$LiCoO_2 \rightleftarrows Li_{1-n}CoO_2 + nLi^+ + ne^-$$

и на отрицательном электроде

$$nLi^+ + ne^- + C \rightleftarrows Li_nC.$$

Разрядившись, аккумулятор переходит в равновесное состояние с наименьшим значением свободной энергии. Максимальная энергия, которая может быть получена при работе аккумулятора, есть изменение его свободной энергии.

Обычно изменение свободной энергии (77) обязано изменению в реальной энергии E. Это как бы естественно. Если неравновесная система характеризуется энергией E, а по мере перехода в равновесное состояние ее энергия уменьшается до E_{eq}, то при удачном инженерном решении устройства разницу $E - E_{eq}$ можно превратить в полезную работу.

Вместе с тем не видно причин почему бы изменение в энтропии при переходе из неравновесного состояния в равновесное нельзя использовать для совершения полезной работы. Возьмите для примера систему локализованных невзаимодействующих спинов в неравновесном состоянии, когда все спины ориентированы «вверх» (рис. 11). В этом состоянии энтропия системы равна нулю. Перейдя в равновесное состояние, в котором спины поровну «вверх» и «вниз», энергия остается прежней, а энтропия увеличивается:

$$\Delta E = 0, \quad \Delta S = Nk\ln 2, \quad \Delta F = \Delta E - T\Delta S = -NkT\ln 2. \qquad (78)$$

В согласии с (77) должна быть принципиальная возможность построить такое устройство, которое выработает энергию, равную, например, $NkT\ln 2$. В работе [14] предлагается такое устройство реализовать в виде спинового вентиля (глава 9) с антипараллельной ориентацией намагниченности контактов [15]. Мы далее подробно рассмотрим такое возможное устройство.

С практической точки зрения $NkT\ln 2$ дает $\sim 2.5\,кДж/моль$, что на 2 – 3 порядка меньше теплотворной способности угля и нефти, которая обеспечивается прежде всего величиной ΔE.

С концептуальной точки зрения ситуация в рассматриваемом примере не обычная. Полученный выйгрыш в энергии связан не с системой спинов, энергия которых не изменилась, а с окружающей средой. Энергия получена из окружающей среды. Стандартно второй закон термодинамики запрещает получение энергии из окружающей среды для выполнения полезной работы. Однако, информация, содержащаяся в неравновесном состоянии системы в виде «отрицательной энтропии» позволяет получать энергию из окружающей среды, не нарушая при этом второй закон термодинамики.

Воспользуемся определением свободной энергии $F = E - TS$ и расщепим правую часть уравнения (74), а именно:

$$E_{avail} = \left(E - E_{eq}\right) + T\left(S_{eq} - S\right) \tag{79}$$

на действительную энергию $\left(E - E_{eq}\right)$ и на информационную энергию $T\left(S_{eq} - S\right)$, которая может быть извлечена из окружающей среды в результате использования информации, доступной нам в форме дефицита энтропии S относительно ее равновесного значения S_{eq}. В приведенном примере с системой локализованных спинов доступная для превращения в полезную работу энергия вся полностью является информационной энергией, поскольку изменение реальной энергии равно нулю.

8.8. Информационно-управляемый аккумулятор

Рассмотрим подробнее устройство для извлечения информационной энергии из системы локализованных спинов.

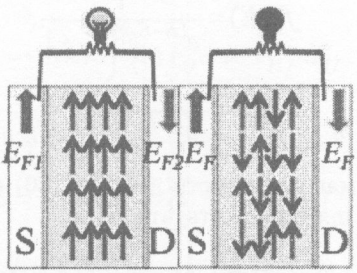


Рис. 12. Информационно-управляемый аккумулятор для получения избытка свободной энергии из системы многих локализованных спинов: в заряженном неравновесном состоянии все спины «вверх» (слева); аккумулятор «садится» после рандомизации всех спинов и перехода системы в равновесное состояние (справа).

Оно может представлять из себя спиновый вентиль (глава 9) с антипараллельной ориентацией намагниченности контактов, в котором

ферромагнитный исток может инжектировать в проводник и экстрагировать из проводника только электроны со спином «вверх», а ферромагнитный сток – только электроны со спином «вниз» (рис. 12).

Канальные электроны в проводнике обменно взаимодействуют с локализованными спинами:

$$u + D \Leftrightarrow U + d, \tag{80}$$

где u и d – странствующие в проводнике электроны, а U и D – локализованные спины.

Обычно такая обменная реакция должна идти в обе стороны с одинаковой скоростью. Пусть мы начинаем с заряженного неравновесного состояния с $U \gg D$ и током, идущим по внешней цепи. Превалировать будет обменная реакция (80), идущая справа налево: избыток канальных электронов u извлекается из проводника левым контактом, а недостаток электронов d компенсируется правым контактом. По истечении некоторого времени устанавливается равновесие с одинаковым числом локализованных спинов $U = D$, обменная реакция пойдет с одинаковой скоростью в обоих направлениях, выделение энергии прекратится и аккумулятор разрядится.

Какое количество энергии может быть получено по мере перехода $(U \gg D) \Rightarrow (U = D)$? Ответ уже был дан: $NkT \ln 2$. Получим ли мы тот же результат в ходе стандартного анализа такого аккумулятора? В предположении малости взаимодействия в обменной реакции (80) можно ожидать, что электроны u находятся в равновесии с контактом 1, а электроны d – в равновесии с контактом 2 так что

$$f_u(E) = \frac{1}{e^{(E-E_{F1})/kT} + 1}, \tag{81}$$

$$f_d(E) = \frac{1}{e^{(E-E_{F2})/kT} + 1}. \tag{82}$$

В предположении малой скорости реакции (80) настолько, что она почти сбалансирована, можно предположить, что

$$P_D f_u \left(1 - f_d\right) = P_U f_d \left(1 - f_u\right), \tag{83}$$

или для вероятностей имеем:

$$\frac{P_U}{P_D} = \frac{f_u}{1 - f_u} \frac{1 - f_d}{f_d} = e^{\Delta E_F / kT}, \tag{84}$$

где $\Delta E_F \equiv E_{F1} - E_{F2}$. Задавшись электрохимическими потенциалами на контактах, получили отношение распределений локализованных спинов U и D. Иначе

говоря, разность электрохимических потенциалов порождается различием в распределении спинов U и D:

$$\Delta E_F = kT \ln \frac{P_U}{P_D}. \tag{85}$$

Первоначально различие в электрохимических потенциалах велико в результате существенного превалирования спинов U над спинами D, однако, со временем распределения спинов уравниваются, что ведет к уравниванию электрохимических потенциалов на контактах и к прекращению тока.

Обменная реакция (80) сводится к тому, что каждый раз когда спин D обращается в спин U, электрон u обращается в d, который выводится из проводника через сток. И наоборот, когда U обращается в D, электрон d обращается в u, который выводится из проводника через исток. Полное число электронов, переносимых от истока S к стоку D, равно недостатку спинов U по сравнению со спинами D:

$$n_{S\to D} = -\Delta N_U. \tag{86}$$

Извлекаемая энергия как произведение разности электрохимических потенциалов на число перенесенных электронов из исходного неравновесного состояния в конечное равновесное

$$E = -\int_{Ini}^{Fin} \Delta E_F \, dN_U. \tag{87}$$

Подстановка (85) дает

$$E = -NkT \int_{Ini}^{Fin} \left(\ln P_U - \ln P_D\right) dP_U. \tag{88}$$

Поскольку $dP_U + dP_D = 0$ и учитывая, согласно (59), что

$$S = -Nk\left(P_U \ln P_U + P_D \ln P_D\right), \tag{89}$$

подынтегральное выражение в (88) можно переписать как

$$\left(\ln P_U - \ln P_D\right) dP_U = d\left(P_U \ln P_U + P_D \ln P_D\right) = -dS/Nk, \tag{90}$$

так что

$$E = T \int_{Ini}^{Fin} dS = T\Delta S, \tag{91}$$

что и является основным результатом, который и ожидался, а именно: метастабильное состояние со всеми спинами «вверх» может быть в принципе

использовано для конструирования аккумулятора с выходом по энергии во внешнюю цепь до величины

$$T\Delta S = NkT\ln 2. \qquad (92)$$

8.8.1. Важно знать детальную информацию

До сих пор утверждалось, что система со всеми спинами «вверх» имеет бóльшую свободную энергию по сравнению с рандомизированной системой спинов в равновесном состоянии и эту разницу в свободной энергии в принципе можно извлечь и превратить в полезную работу с помощью подходящего устройства. А что делает рандомизированную систему так сильно отличающейся от упорядоченной системы? Как заметил Ричард Фейнман, кажется необычным встретить на дороге автомашину с номером 9999 по сравнению с неким безликим номером, например, 7048. Аналогично, если спины расположены пусть случайным, но точно известным нам образом, то энтропия такой системы может быть принятой равной нулю как и в случае, когда точно известно, что все спины «вверх». Возможность получить энергию проистекает не от того, что в исходном состоянии все спины «вверх», а от точного знания в каком именно состоянии находится каждый спин. Рассмотрите, например, систему на рис. 13.

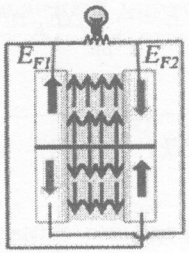


Рис. 13. Информационно-управляемый двухсекционный аккумулятор с переключением намагниченностей контактов.

Система спинов не совсем рандомизирована, но все же … половина спинов «вверх», а другая половина «вниз». При стандартном включении односекционного вентиля (рис. 12) получить ток в незамкнутой цепи невозможно: одна половина спинов компенсирует другую половину спинов. Но можно в двух разных секциях аккумулятора намагниченность контактов переключить таким образом, чтобы ток пошел. Этим примером иллюстрируется то обстоятельство, что априорное знание исходной конфигурации спинов может открыть возможность такого подключения контактов, которое позволит получить полезную работу. Детальное владение информацией о неравновесном

состоянии системы может в принципе быть основой для построения источника полезной энергии.

Шеннон [16] ввел понятие информационной энтропии

$$H = -\sum_i p_i \ln p_i , \qquad (93)$$

как меры информации, содержащейся в строке символов i, выпадающей с вероятностью p_i. Шенноновская энтропия выглядит как термодинамическая энтропия (59), исключая константу Больцмана k. Обе энтропии тесно связаны друг с другом [17]. Утверждается, что если мы имеем систему с состояниями i и равновесными вероятностями p_i, тогда $k \cdot H$ есть энтропия равновесной системы, не содержащей никакой информации. Если система зафиксирована в одном из своих возможных равновесных состояниях, то энтропия понижается на величину $Nk \cdot H$, увеличивая свободную энергию системы на $NkT \cdot H$. Этот избыток свободной энергии в принципе может быть превращен в полезную работу.

8.9. Принцип Ландауэра о минимуме энергии, необходимом для стирания одного бита информации

Идея о том, что обладание полными сведениями о метастабильном состоянии системы позволяет создать аккумулятор на ее основе, перекликается с принципом Ландауэра о том, какой минимум энергии необходим, чтобы стереть один бит информации [18, 19].

В рассматриваемом нами контексте стирание эквивалентно переводу системы спинов из равновесного состояния с $F_{eq} = -NkT \ln 2$ в стандартное неравновесное состояние с известной структурой спинов с $F = 0$ (рис. 14).

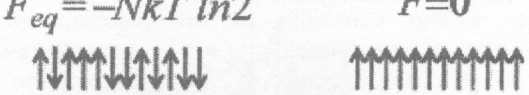


Рис. 14. Стирание информации путем перевода системы спинов из равновесного состояния с $S = NkT \ln 2$ в неравновесное с $S = 0$.

Какой минимум энергии нужен для достижения сформулированной цели? Выше было показано, что как только структура спинов стандартна (все спины «вверх»), оказывается возможным построить аккумулятор и извлечь энергию в количестве $F - F_{eq}$. В циклическом процессе на пути от равновесного состояния системы со свободной энергией F_{eq} к неравновесному с энергией F тратится энергия на стирание E_{err}, затем в результате работы аккумулятора извлекается энергия $F - F_{eq}$, так что полная энергия, затрачиваемая на стирание, есть

$E_{err} - (F - F_{eq})$ и она должна быть больше нуля, ибо в противном случае получится вечный двигатель,

$$E_{err} \geq (F - F_{eq}), \tag{94}$$

что и дает нам принцип Ландауэра

$$E_{err} \geq NkT \ln 2. \tag{95}$$

Принцип Ландауэра, устанавливающий связь термодинамики с информацией, недавно был, наконец, успешно экспериментально подтвержден [20].

Представляется правдоподобным, что стирание не обязательно должно заканчиваться приведением всех спинов в положение «вверх», как в рассматриваемой нами задаче, а может заключаться в приведении системы в состояние с известной структурой спинов. Кроме того, минимальная энергия стирания не обязательно должна диссипировать. Стирание заканчивается именно диссипацией только лишь потому, что сегодня еще нет технической возможности запустить информационно-управляемый аккумулятор.

50 лет тому назад Рольф Ландауэр задумался над глубокими вопросами, время ответа на которые тогда еще не пришло. Сегодня в результате удивительно быстрого развития наноэлектроники некоторые вопросы уже сняты, а постановка других вопросов стала намного яснее. Изучение ряда проблем перенесено уже на квантовый уровень и сейчас активно обсуждается.

8.10. Демон Максвелла

Информационно-управляемый аккумулятор имеет прямое отношение к демону Максвелла [21, 22], который во-время открывая и закрывая щель между двумя резервуарами «горячие» частицы пропускает только в правый резервуар, а «холодные» – только в левый (рис. 15).

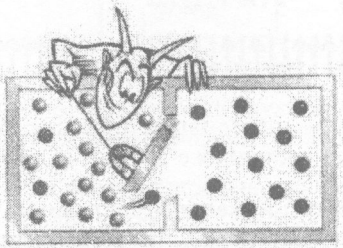


Рис. 15. Демон Максвелла создает разность температур, сортируя частицы на горячие (черные кружки) и холодные (белые кружки).

Аналогия между информационно-управляемым аккумулятором и демоном Максвелла видна из следующего сопоставления:

Горячие частицы $\Leftrightarrow$ Электроны со спином «вверх»
Холодные частицы $\Leftrightarrow$ Электроны со спином «вниз»
Демон Максвелла $\Leftrightarrow$ Система локализованных спинов с двумя состояниями
Резервуары $\Leftrightarrow$ Контакты

Аккумулятор работает за счет обменной реакции (80), в которой роль демона играют локализованные спины. Демон использует информацию, а не энергию создавая разность температур так же, как и аккумулятор использует низкоэнтропийное состояние локализованных спинов для создания разности электрохимических потенциалов. Как и локализованные спины, демон также дрейфует к высокоэнтропийному состоянию, которое не позволяет ему дискриминировать горячие и холодные частицы. Как и акумулятор, демон в конце концов прекращает функционировать, когда энтропия достигает своего максимального значения и теряется вся исходная информация.

Транспортные процессы всегда сочетают в себе механические и энтропийные процессы, потому нередко говорят о механических и энтропийных силах и факторах, которые и управляют большинством окружающих нас событий.

Полностью поляризованная система спинов с $S=0$ самопроизвольно переходит в неполяризованное состояние с $S=Nk\ln 2$. Такой направленный процесс возникает потому, что полностью поляризованное состояние является по сути одним состоянием, а неполяризованное состояние допускает множество различных состояний локализованных спинов емкостью 2^N. Именно такая громадная емкость управляет примесями при их релаксации из состояния с низкой энтропией в состояние с высокой энтропией, но никак не наоборот. Многие окружающие нас явления управляются энтропийными силами, которые принципиально отличаются от механических сил, способных лишь перевести систему из одного состояние в какое-то другое, но опять одно. Транспортные явления потому так сложны, что их описание требует одновременного привлечения как механических представлений, так и термодинамических. Модель упругого резистора (глава 1) хороша тем, что в ней действие энтропийных сил ограничено лишь контактами, а в самом резисторе действуют законы механики.

Подведем итоги. Построена равновесная термодинамика проводника с током, подчеркнута роль фоковских состояний, обращено внимание на накопление информации в неравновесном состоянии и подробно проанализирована модель информационно-управляемого аккумулятора и связь ее с принципом Ландауэра о минимальной энергии, необходимой для стирания одного бита информации [23, 24].

Литература

1. Ю. А. Кругляк. Наноэлектроника «снизу – вверх»: Возникновение тока, обобщенный закон Ома, упругий резистор, моды проводимости, термоэлектричество, *ScienceRise*, т. 7, № 2(12), 76 – 100 (2015).
2. Ю. О. Кругляк, М. В. Стріха. Уроки наноелектроніки: Роль електростатики й контактів у концепції «знизу – вгору», *Сенсор. електр. мікросист. техн.*, т. 11 № 4, 27 – 42 (2014).
3. R. P. Feynman. *Statistical Mechanics*. San Francisco: Benjamin/Cummings, 1972.
4. Ю. А. Кругляк. Перенос тепла фононами в транспортной модели Ландауэра – Датты – Лундстрома, *ScienceRise*, т. 2, № 2(7), 81 – 93 (2015).
5. V. Fock. Konfigurationsraum und zweite Quantelung, *Z. Phys.*, v. 75, 622 – 647 (1932).
6. Yu. A. Kruglyak. Configuration interaction in the second quantization representation: basics with application up to full CI, *ScienceRise*, v. 4, n 2(4), 98 – 115 (2014).
7. V. A. Kuprievich, Yuri A. Kruglyak, E. V. Mozdor. Full configuration interaction for the benzyl radical, *Intern. J. Quantum Chem.*, v. 4, n 1, 73 – 87 (1970).
8. V. A. Kuprievich, Yu. A. Kruglyak, E. V. Mozdor. The Configuration Interaction Method in the Second Quantization Representation, *Croat. Chem. Acta,* v. 43, 1 – 13 (1971).
9. Yu. A. Kruglyak, E. V. Mozdor, V.A. Kuprievich. Study of the Electronic Structure of Radicals by the CI Method. 1. Matrix Elements of the Physical Value Operators, *Croat. Chem. Acta*, v. 43, 15 – 23 (1971).
10. Yuri A. Kruglyak, I. I. Ukrainsky. Study of the electronic structure of alternant radicals by the DODS method, *Intern. J. Quantum Chem.*, v. 4, n 1, 57 – 72 (1970).
11. Yu. A. Kruglyak. Quantum-chemical studies of quasi-one-dimensional electron systems. 1. Polyenes, *ScienceRise*, v. 5, n 2 (10), 69 – 105 (2015).
12. Yu. A. Kruglyak. Quantum-chemical studies of quasi-one-dimensional electron systems. 2. Polyenes, *ScienceRise*, v. 6, n 2 (11), 122 – 148 (2015).
13. K. A. Dill, S. Bromberg. *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience*. New York: Garland Science, 2010.
14. S. Salahuddin, S. Datta. An All Electrical Spin Detector, *Sixth IEEE Conf. Nanotech.*, v. 2, 834 – 837 (2006).
15. Ю. А. Кругляк. Наноэлектроника «снизу – вверх»: Начало спинтроники и магнетроники, *ScienceRise*, т. 8, № 2(13), 51 – 68 (2015).

16. К. Шеннон. *Работы по теории информации и кибернетике*. Москва: ИЛ, 2002.
17. М. В. Волькенштейн. *Энтропия и информация*. Москва: Наука, 2006.
18. R. Landauer. Irreversibility and heat generation in the computing process, *IBM J. Res. Dev.*, v. 5, 183 – 191 (1961).
19. C. H. Bennett. Notes on Landauer's principle. Reversible Computation and Maxwell's Demon, *History Phil. Mod. Phys.*, v. 34, 501 – 510 (2003).
20. A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, E. Lutz. Experimental verification of Landauer's principle linking information and thermodynamics, *Nature*, v. 483, 187 – 189 (2012).
21. H. S. Leff, A. F. Rex (eds.). *Maxwell's Demon 2: Entropy, Classical and Quantum Information Computing*. Bristol: Institute of Physics Publishing, 2003.
22. S. Datta S. Nanodevices and Maxwell's Demon. *Lecture Notes in Nanoscale Science and Technology, Vol. 2, Nanoscale Phenomena: Basic Science to Device Applications*, Eds. Z. K. Tang and P. Sheng. Berlin: Springer, 2008.
23. Ю. О. Кругляк, М. В. Стріха. Уроки наноелектроніки: Електричний струм і другий закон термодинаміки в концепції «знизу – вгору», *Сенсор. електр. мікросист. техн.*, т. 12 № 2, 5 – 26 (2015).
24. Ю. А. Кругляк. Наноэлектроника «снизу – вверх»: Термодинамика проводника с током, информационно-управляемый аккамулятор и квантовая энтропия, *ScienceRise*, т. 11, № 2(16), 55 – 71 (2015).

II. Квантовый транспорт, спинтроника, магнетроника, молекулярная электроника и смежные вопросы

Глава 9. Спинтроника и магнетроника

9.1. Введение

В этой главе в рамках концепции «снизу – вверх» наноэлектроники [1 – 3] рассмотрим такие ключевые вопросы спинтроники как спиновый вентиль, граничное сопротивление при несовпадении мод проводимости, спиновые потенциалы и разность нелокальных спин-потенциалов, спиновый момент и его транспорт, уравнение Ландау – Лифшица – Гильберта применительно к выделенной оси магнита, рассмотрим обращение намагниченности спиновым током, поляризаторы и анализаторы спинового тока.

Электроника второй половины XX века основывалась на транспорте заряда электронов и управления им электрическими и магнитными полями (зарядовая электроника). В конце прошлого века началось бурное развитие нового направления, основанного на том, что электроны имеют не только электрический заряд, но и спин и связанный с ним магнитный момент. Это направление получило название спиновой электроники или спинтроники (*spin-tr*ansport electr*onics*).

Среди работ, предвосхитивших развитие спинтроники, отметим пионерские исследования М.И.Дьяконова и В.И.Переля, показавших возможность ориентации спинов при протекании тока [4], М.Жюльера по туннельному магнитосопротивлению [5], А.Г.Аронова и Г.Е.Пикуса по спиновой инжекции в полупроводниках [6]. И поныне, 40 лет спустя, исследования в области спинтроники ведутся в области этих трех открытых эффектов – инжекции в магнитных переходах носителей с определенным направлением спина, переключения таких переходов спин-поляризованным током и гигантского магнитосопротивления.

Началом современного этапа исследований в области спинтроники принято считать работы [7, 8], в которых было экспериментально показано, что электронный ток в ферромагнитном металле поляризован по спину и было открыто явление гигантского магнитосопротивления. Поляризация тока открыла возможность управления транспортом спинов в ферромагнитных структурах с помощью магнитных полей. В 2007 году Альберт Ферт и Петер Грюнберг были удостоены Нобелевской премии по физике за открытие гигантского магнитосопротивления.

Основным объектом исследований в спинтронике и поныне остается спиновый вентиль (spin valve). В простейшем случае он состоит из двух токонесущих ферромагнитных (ФМ) контактов, разделенных достаточно тонким каналом транспорта электронов (спейсер / spacer). Спейсер может быть металлическим, но не магнитным, может быть диэлектриком, его роль могут играть отдельные молекулы, кластеры и любые наноразмерные структуры.

Перенос электронов по спейсеру обычно баллистический или туннельный. Один из ферромагнитных контактов (он именуется свободным / free) характеризуется малой энергией анизотропии и легко меняет направление своей намагниченности под действием внешнего магнитного поля соответствующей ориентации. Другой ферромагнитный контакт (его называют закрепленным / pinned) характеризуется существенно большей энергией анизотропии и требует существенно более сильных полей для изменения своей намагниченности. Сильная анизотропия закрепленного контакта может быть природно присущей ему или же наведенной в процессе изготовления.

Для спинового вентиля характерна сильная зависимость электрического сопротивления спейсера при протекании тока между магнитными контактами от взаимной ориентации намагниченности контактов: при параллельной ориентации (*P*) сопротивление значительно меньше, чем при антипаралельной ориентации (*AP*)

$$R_P < R_{AP}. \qquad (1)$$

Поскольку ориентация намагниченности свободного ферромагнитного контакта может меняться под действием внешнего магнитного поля, то это приводит к сильной зависимости сопротивления проводника между контактами от приложенного магнитного поля.

Понять экспериментально наблюдаемое неравенство сопротивлений (1) качественно можно на основе двухканальной модели Мотта [9, 10], в которой перенос мажоритарных электронов (направление спина параллельно намагниченности) и миноритарных электронов (направление спина антипараллельно намагниченности) условно осуществляется по двум независимым спиновым подзонам (рис. 1) в условиях отсутствия спин-флип рассеяния (to flip – переворачивать), к рассмотрению которого вернемся позже.

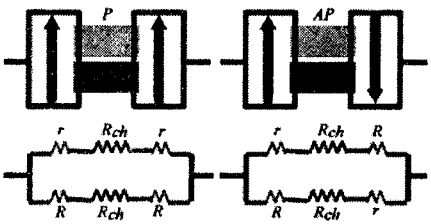


Рис. 1. Параллельная и антипараллельная ориентации намагниченности контактов спинового вентиля и соответствующие эквивалентные схемы сопротивления для мажоритарных (слева) и миноритарных (справа) носителей заряда. Канал транспорта электронов условно разбит на две спиновых подзоны: для электронов со спином «вверх» (*up*) подзона закрашена светлосерым, а со спином «вниз» (*dn*) – темносерым цветом.

Электрон из определенной подзоны одного контакта может туннелировать только в такую же подзону другого контакта. Если намагниченность контактов параллельна, то вероятность такого туннелирования будет намного больше, а электрическое сопротивление будет соответственно меньше, чем в случае антипараллельной намагниченности контактов [11]. Рассмотрим ситуацию подробнее.

Количественную оценку неравенства (1) можно получить в модели, согласно которой спиновая подзона имеет различное граничное сопротивление с контактом в зависимости от того, речь идет о переносе спинов параллельных (мажоритарных спинов) или антипараллельных (миноритарных спинов) намагниченности контактов. Граничное сопротивление для мажоритарных спинов меньше, чем для миноритарных $(r < R)$. Соответствующие эквивалентные схемы сопротивления показаны на рис. 1. Полноты ради, учтено также сопротивление каналов подзон R_{ch}.

Из элементарной теории электрических цепей следует, что для параллельной ориентации намагниченности контактов

$$R_P = \left(\frac{1}{2r + R_{ch}} + \frac{1}{2R + R_{ch}} \right)^{-1} = \frac{(2r + R_{ch})(2R + R_{ch})}{2(R + r + R_{ch})}, \qquad (2)$$

а для антипараллельной ориентации

$$R_{AP} = \frac{r + R + R_{ch}}{2}. \qquad (3)$$

Качество спинового вентиля определяется различием между R_P и R_{AP}. Можно ожидать, что качество вентиля будет выше, если сопротивлением канала можно пренебречь $(R_{ch} << r, R)$, так что качество вентиля определяется лишь граничными сопротивлениями. Тогда

$$R_P = \frac{2rR}{r + R} \qquad (4)$$

и

$$R_{AP} = \frac{r + R}{2}, \qquad (5)$$

откуда сразу следует неравенство (1), стоит лишь в (4) и (5) бóльшее сопротивление R устремить к бесконечности.

В пределе $R_{ch} \to 0$ получим максимально возможное значение магнитосопротивления (МС)

$$MR \equiv \frac{R_{AP} - R_P}{R_P} = \frac{R_{AP}}{R_P} - 1 = \frac{(r + R)^2}{4rR} = \frac{(R - r)^2 / (R + r)^2}{1 - (R - r)^2 / (R + r)^2}, \qquad (6)$$

если $R_{ch} = 0$.

Поляризация ФМ контакта определяется как

$$P \equiv \frac{R-r}{R+r} \tag{7}$$

и является мерой его эффективности, так что магнитосопротивление

$$MR = \frac{P^2}{1-P^2}, \tag{8}$$

если $R_{ch} = 0$.

Зависимость МС от сопротивления канала R_{ch} показана на рис. 2.

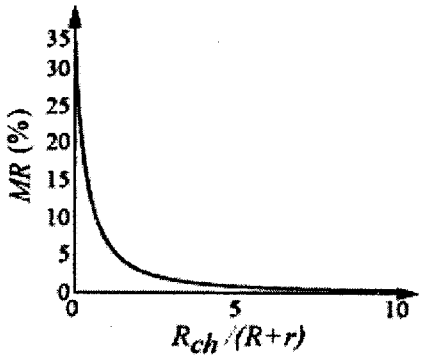


Рис. 2. Падение МС с ростом нормированного сопротивления канала при поляризации $P = 0.5$.

Обращает на себя внимание быстрое зануление МС с ростом нормированного сопротивления канала, начиная, скажем, со значения, равного пяти. Выражение (8) для МС справедливо для металлических немагнитных проводников. В этом случае сопротивление двух последовательно соединенных сопротивлений R_1 и R_2 равно сумме этих сопротивлений $R_1 + R_2$. Если же проводником является диэлектрик, то имеет место магнитный туннельный переход (МТП), а сопротивление двух последовательно соединенных сопротивлений R_1 и R_2 пропорционально произведению этих сопротивлений $KR_1 \cdot R_2$, что следует из физики туннельных проводников, так что для параллельной P ориентации намагниченностей контактов имеем

$$R_P = \frac{Kr^2R^2}{r^2+R^2}, \tag{9}$$

а для антипараллельной AP

$$R_{AP} = \frac{KrR}{2}, \tag{10}$$

так что

$$\frac{R_{AP}}{R_P}=\frac{r^2+R^2}{2rR}=\frac{(R+r)^2+(R-r)^2}{(R+r)^2-(R-r)^2}=\frac{1+P^2}{1-P^2}, \tag{11}$$

а магнитосопротивление МТП

$$MR=\frac{2P^2}{1-P^2} \tag{12}$$

отличается двойкой от МС металлического проводника (8).

9.2. Граничное сопротивление и несовпадение мод проводимости

Поначалу в спиновых вентилях использовались металлические спейсеры, в частности, медные. Оказалось, однако, что во многих приложениях лучше себя показывают непроводящие оксиды в режиме МТП, обеспечивая более высокие значения МС. Попытки использовать полупроводниковые спейсеры были неудачными приблизительно до 2000 года, когда стало ясно, что причина неудач кроется в высоких значениях R_{ch} сравнительно с суммой $(r + R)$, приводящих к низким значениям МС [12, 13]. Выход был найден в увеличении граничных сопротивлений за счет дополнительных барьерных слоев на границах с контактами (рис. 3). Сейчас это стандартная процедура при работе с полупроводниковыми каналами.

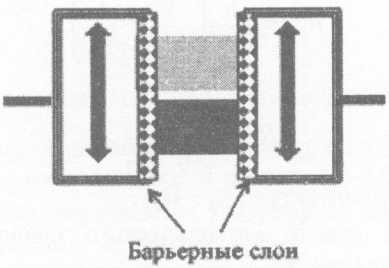


Рис. 3. Дополнительные барьерные слои с целью увеличить граничные сопротивления при инжекции спинов в полупроводниковый канал проводимости.

Как же это работает? Стандартное объяснение очевидно. Барьерные слои увеличивают граничные сопротивления r и R, уменьшая тем самым отношение $R_{ch}/(r+R)$ и увеличивая МС (рис. 2). Однако, если бы дело было только в этом, то можно было бы уменьшить толщину спейсера настолько, чтобы перейти в баллистический режим транспорта $(L \ll \lambda)$. Эта идея не нашла, однако, экспериментального подтверждения.

Число мод $M(E)$ или же плотность состояний $D(E)$ в обычном канале проводимости и в спейсере спинового клапана схематически показаны на рис. 4. В обычном канале обе спиновые подзоны одинаковы. В спиновом же вентиле полоса миноритарных спинов обычно сдвинута вверх по энергии, в результате чего число мод в районе $E = E_{F0}$ меньше для миноритарных спинов (M_{dn}), чем для мажоритарных (M_{up}). Чему будут равны граничные сопротивления?

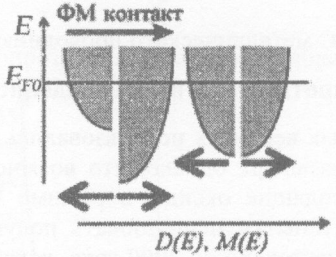


Рис. 4. Спиновые подзоны в обычном канале проводимости (справа) и в спиновом вентиле (слева).

В § 7.5 было показано, что сопротивление R_{int} на границе контакта двух проводников с разным числом мод $(M_1 > M_2)$

$$R_{int} = \frac{h}{2q^2}\left(\frac{1}{M_2} - \frac{1}{M_1}\right). \tag{13}$$

Если $M_1 \gg M_2$, то

$$R_{int} = \frac{h}{2q^2 M_2}, \tag{14}$$

что отвечает «хорошему контакту» $(M_1 > M_2)$.

Число мод M в канале металлического проводника обычно имеет промежуточное значение (рис. 5).

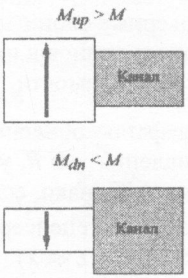


Рис. 5. Металлический спейсер между двумя ФМ контактами.

В идеале же

$$M_{up} \gg M \gg M_{dn}, \tag{15}$$

так что ФМ контакт «хорош» для мажоритарных спинов, но не для миноритарных:

$$r = \frac{h}{2q^2}\left(\frac{1}{M} - \frac{1}{M_{up}}\right) \approx \frac{h}{2q^2 M},$$
$$R = \frac{h}{2q^2}\left(\frac{1}{M_{dn}} - \frac{1}{M}\right) \approx \frac{h}{2q^2 M_{dn}}. \tag{16}$$

Число же мод M в полупроводящем канале (рис. 6) обычно меньше числа мод в обоих ФМ контактах:

$$M_{up} > M_{dn} \gg M, \tag{17}$$

так что

$$r = R = \frac{h}{q^2 M} \tag{18}$$

и поляризация P нулевая.

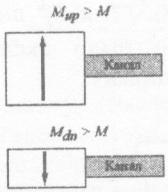


Рис. 6. Полупроводящий спейсер между двумя ФМ контактами.

Другими словами, проблема инжекции спинов в полупроводящий канал не только вызвана высоким сопротивлением канала R_{ch}, которое можно было бы уменьшить в режиме баллистического транспорта, но и тем, что теряется различие между граничными сопротивлениями R и r для обоих спинов. Скажем, если канал имеет 10 мод проводимости, то ему безразлично имеет ли ФМ контакт 100 мод (миноритарные спины) или 1000 мод (мажоритарные спины). И в том и в другом случае недостатка электронов в канале не будет.

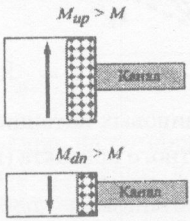


Рис. 7. Барьеры на границе ФМ контактов и канала проводимости.

При наличии барьеров на границе ФМ контакта и полупроводящего канала (рис. 7) граничное сопротивление уже не дается формулой (13), а полагают, что оно пропорционально произведению плотности состояний, а стало быть и числа мод на обеих сторонах туннельного барьера так что

$$r = K \cdot M_{up} M, R = K \cdot M_{dn} M \tag{19}$$

с константой пропорциональности K. Теперь поляризация P не зависит от числа мод в канале:

$$P = \frac{M_{up} - M_{dn}}{M_{up} + M_{dn}} \tag{20}$$

и ей можно придать нужное численное значение. Граничные сопротивления теперь, конечно, больше по сравнению с омическим сопротивлением (13).

9.3. Спиновые потенциалы

Различие в граничном сопротивлении между магнитным контактом и спиновыми подзвонами для спинов вверх (*up*) и вниз (*dn*) позволяет ввести понятие о спиновых потенциалах E_F^{up} и E_F^{dn} внутри немагнитного проводника. Различие между ними вначале было экспериментально обнаружено на металлах, а затем и на полупроводниках.

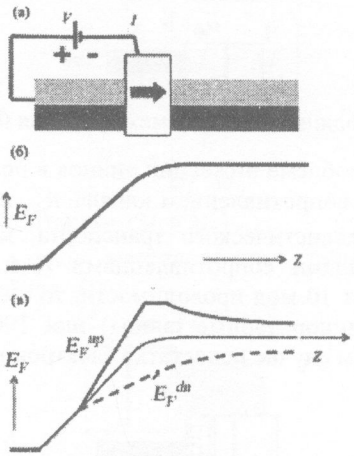


Рис. 8. Сепарирование спиновых потенциалов E_F^{up} и E_F^{dn} в канале с использованием магнитного контакта (качественная картина).

Концепцию спинового потенциала продемонстрируем на простой структуре с одним магнитным контактом (рис. 8а). Если не учитывать спины,

профиль электрохимического потенциала качественно выглядел бы как на рис. 8б. Количественное решение дают уравнения диффузии и непрерывности с соответствующими граничными условиями для $E_F(z)$ на контактах, что подробно рассмотрено в главе 7. Поскольку граничные сопротивления между магнитным контактом и спиновыми подзонами проводника различны, следует ожидать различное падение электрохимических потенциалов на границе между контактом и спиновыми подзонами, и при решении соответствующих уравнений диффузии профили электрохимических потенциалов для спинов *up* и *dn* будут различны, как это качественно показано на рис. 8в.

Электрохимические потенциалы для двух спинов сепарируются на магнитном контакте, однако, затем стремятся вернуться к исходному значению в результате спин-флип-релаксации, которая непрерывно стремится восстановить локальное равновесие путем уравнивания E_F^{up} и E_F^{dn}. Количественно поведение спиновых потенциалов дается уравнениями диффузии для спинов *up* и *dn*:

$$I_{up} = -\frac{\sigma A}{2q}\frac{dE_F^{up}}{dz},$$
$$I_{dn} = -\frac{\sigma A}{2q}\frac{dE_F^{dn}}{dz}, \qquad (21)$$

в которых для каждого из спинов учитывается половина проводимости по сравнению с уравнением для суммарного тока.

Спин-флип-релаксация обращает ток I_{up} в ток I_{dn} и наоборот, так что

$$\frac{dI_{up}}{dz} = -\frac{dI_{dn}}{dz} = -K(E_F^{up} - E_F^{dn}), \qquad (22)$$

где константа пропорциональности K есть мера эффективности спин-флип-релаксации, стремящейся уравнять спиновые потенциалы E_F^{up} и E_F^{dn}.

Комбинируя (22) и (21), получаем

$$\frac{d^2E_F^{up}}{dz^2} = \frac{E_F^{up} - E_F^{dn}}{2\lambda_{sf}^2} = -\frac{d^2E_F^{dn}}{dz^2}, \qquad (23)$$

где длина

$$\lambda_{sf} = \frac{1}{2}\sqrt{\sigma A / qK} \qquad (24)$$

есть характеристическое расстояние, на котором электрон меняет свой спин на противоположный. Характерные значения длины спин-флипа меняются в широких пределах от нескольких десятков нанометров до сотен микрометров в зависимости от среды и температуры.

Уравнение (23) известно как уравнение Вале – Ферта [13]. Изначально оно было получено как следствие транспортного уравнения Больцмана [14, 15] и ныне широко используется при обсуждении диффузионных задач с учетом спина электронов.

Введем понятия зарядового и спинового потенциалов

$$E_F \equiv (E_F^{up} + E_F^{dn})/2, \tag{25}$$

$$E_F^s \equiv E_F^{up} - E_F^{dn} \tag{26}$$

и аналогично – зарядового и спинового токов

$$I = I_{up} + I_{dn}, \tag{27}$$

$$I_s = I_{up} - I_{dn}. \tag{28}$$

Зарядовые потенциалы и токи удовлетворяют обычным диффузионным уравнениям (глава 7), а спиновый потенциал определяется длиной спин-флипа:

$$\frac{d^2 E_F^s}{dz^2} = \frac{E_F^s}{\lambda_{sf}^2}. \tag{29}$$

Можно ли измерить спиновую разность потенциалов внутри канала проводимости? Можно, и не только в пределах канала проводимости, но и за его пределами, как это показано на рис. 9. Подобные измерения известны как измерения разности нелокальных спин-потенциалов и сейчас являются рутинными при исследовании спин-транспортных задач [19].

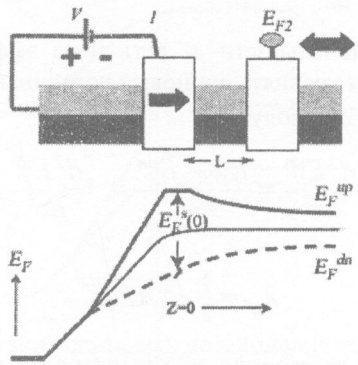


Рис. 9. К измерению спиновой разности потенциалов за пределами проводника тока.

Спиновая разность потенциалов измеряется при изменении намагниченности пробного электрода за пределами проводника (рис. 9) с параллельного режима P на антипараллельный AP и как будет далее показано равна

$$V_S \equiv \frac{E_F^P - E_F^{AP}}{q} = P_1 P_2 I R_S e^{-L/\lambda_{sf}}, \tag{30}$$

где P_1 и P_2 – поляризации инжектирующего и детектирующего ФМ контактов (рис. 9), а спиновое сопротивление

$$R_S = \lambda_{sf} / \sigma A. \tag{31}$$

9.3.1. Разность нелокальных спин-потенциалов

Уравнение (30) можно получить в два шага. Сначала покажем, что спиновый потенциал инжектирующего контакта

$$E_F^s(0) = P_1 q I R_s. \tag{32}$$

Затем покажем, что разность

$$E_F^P - E_F^{AP} = P_2 E_F^s(0) e^{-L/\lambda_{sf}}, \tag{33}$$

откуда сразу получается уравнение (30).

Поведение спиновых потенциалов описывается уравнением (29), согласно которому спиновый потенциал уменьшается экспоненциально в обе стороны от инжектирующего контакта

$$E_F^s = E_F^s(0) e^{-|z|/\lambda_{sf}}, \tag{34}$$

как это показано на рис. 10.

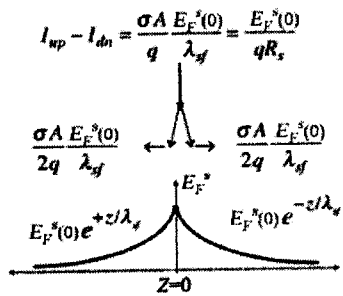


Рис. 10. К расчету суммарного спинового тока, порождаемого инжектирующим контактом на рис. 9.

Теперь вычислим спиновый ток в обоих направлениях от инжектирующего контакта, а именно:

$$I_S = -\frac{\sigma A}{2q} \frac{dE_F^s}{dz}, \tag{35}$$

выражение для которого следует из (21) и (25) – (28). Сами токи в обоих направлениях показаны на рис. 10. Их сумма с учетом спинового сопротивления (31) дает суммарный спиновый ток

$$I_{up} - I_{dn} = \frac{E_F^s(0)}{qR_S}. \tag{36}$$

Теперь рассмотрим ток от инжектирующего контакта через его граничные проводимости g_{up} и g_{dn} для спинов *up* и *dn* (рис. 11).

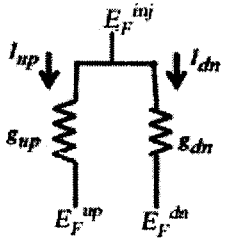


Рис. 11. К вычислению токов от инжектирующего контакта через его граничные проводимости g_{up} и g_{dn} для спинов *up* и *dn*.

Из теории электрических цепей имеем

$$\frac{E_F^s(0)}{q} \equiv \frac{E_F^{up} - E_F^{dn}}{q} = \frac{I_{dn}}{g_{dn}} - \frac{I_{up}}{g_{up}}, \tag{37}$$

что можно переписать в виде

$$\frac{E_F^s(0)}{q} = \frac{g_{up} + g_{dn}}{2g_{up}g_{dn}}\left(P_1 I - (I_{up} - I_{dn})\right) \tag{38}$$

через поляризацию инжектирующего контакта

$$P_1 \equiv \frac{g_{up} - g_{dn}}{g_{up} + g_{dn}}, \tag{39}$$

а с использованием (36) имеем

$$(I_{up} - I_{dn})R_s = \frac{g_{up} + g_{dn}}{2g_{up}g_{dn}}\left(P_1 I - (I_{up} - I_{dn})\right) \tag{40}$$

или иначе

$$\frac{P_1 I}{I_{up} - I_{dn}} - 1 = \frac{2R_s}{\frac{1}{g_{up}} + \frac{1}{g_{dn}}}. \tag{41}$$

Спиновое сопротивление R_s (31) есть сопротивление той части канала проводимости, длина которой соответствует спин-флип-длине λ_{sf}, и оно намного меньше чем граничные сопротивления $1/g_{up}$ и $1/g_{dn}$ (рис. 11), которые особенно велики при использовании барьеров для усиления поляризации контакта. В этих условиях правая часть равенства (41) зануляется, так что окончательно

$$I_{up} - I_{dn} = P_1 I, \tag{42}$$

что вместе с (40) окончательно дает искомое выражение (32).

Для получения на втором шаге выражения (33) начинаем со спинового потенциала на детектирующем контакте (рис. 9)

$$E_F^s(L) = E_F^s(0) e^{-L/\lambda_{sf}}. \tag{43}$$

Для нахождения потенциала, регистрируемого детектирующим контактом, воспользуемся цепью на рис. 12, аналогичной использованной для инжектирующего контакта на рис. 11.

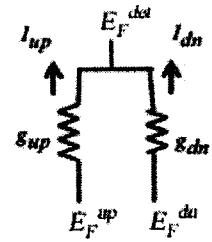


Рис. 12. К вычислению токов на детектирующем контакте через его граничные проводимости g_{up} и g_{dn} для спинов *up* и *dn*.

Поскольку суммарный ток на детектирующем контакте равен нулю, то для цепи на рис. 12 при параллельной ориентации намагниченности контакта имеем

$$I = 0 = g_{up}(E_F^{up} - E_F^P) + g_{dn}(E_F^{dn} - E_F^P), \tag{44}$$

так что

$$E_F^P = \frac{g_{up}E_F^{up} + g_{dn}E_F^{dn}}{g_{up} + g_{dn}}. \tag{45}$$

В случае же антипараллельной ориентации в числителе появляются перекрестные произведения

$$E_F^{AP} = \frac{g_{dn}E_F^{up} + g_{up}E_F^{dn}}{g_{up} + g_{dn}}. \tag{46}$$

Итак,

$$E_F^P - E_F^{AP} = \frac{(g_{up} - g_{dn})(E_F^{up} - E_F^{dn})}{g_{up} + g_{dn}} = P_2 E_F^s(L), \tag{47}$$

где поляризация детектирующего контакта P_2 определяется через граничные проводимости точно так же, как и поляризация инжектирующего контакта (39). Из (47) и (43) получаем искомое уравнение (33). Это же уравнение можно получить несколько иначе.

Перепишем (44) в общем виде:

$$I = 0 = g_{up}(E_F^{up} - E_F^{\det}) + g_{dn}(E_F^{dn} - E_F^{\det}) \tag{48}$$

так что

$$E_F^{\det} = \frac{g_{up}E_F^{up} + g_{dn}E_F^{dn}}{g_{up} + g_{dn}}. \tag{49}$$

Используя уравнения (25) и (26), перепишем E_F^{up} и E_F^{dn} через E_F и E_F^s:

$$\begin{aligned} E_F^{up} &= E_F + \frac{E_F^s}{2}, \\ E_F^{dn} &= E_F - \frac{E_F^s}{2}, \end{aligned} \tag{50}$$

так что для параллельной ориентации намагниченности детектирующего контакта

$$E_F^P = E_F + \frac{P_2 E_F^s}{2}, \tag{51}$$

а для антипараллельной ориентации

$$E_F^{AP} = E_F - \frac{P_2 E_F^s}{2}, \tag{52}$$

где поляризация P_2 определена выше. Таким образом, мы снова пришли к уравнению (47):

$$E_F^P - E_F^{AP} = P_2 E_F^s(L). \tag{53}$$

9.4. Спиновый момент

Спиновый вентиль и многочисленные различные устройства электроники на его основе явились наиболее значительным достижением спинтроники [11, 20 – 22]. Другим удивительным достижением явилось экспериментальное обнаружение транспорта спинового момента [23 – 25], предложенного в [26, 27] и которое заключается в том, что спиновые токи могут менять намагниченность

наноконтакта [28 – 30]. Схема эксперимента по транспорту спинового момента показана на рис. 13.

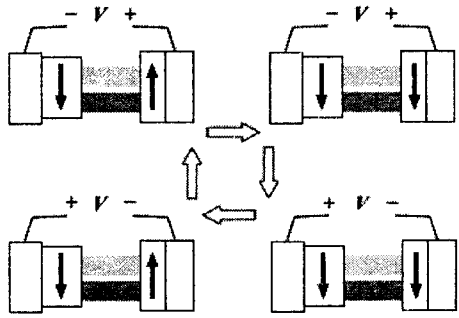


Рис. 13. Демонстрация эксперимента по транспорту спинового момента.

Намагниченность закрепленного левого контакта спинового вентиля фиксирована и направлена вниз. Правый наноконтакт свободен и его намагниченность может изменять свое направление. Подача отрицательного потенциала на закрепленный контакт порождает отрицательный спиновый потенциал

$$E_F^s \equiv E_F^{up} - E_F^{dn} < 0, \tag{54}$$

который вызывает перенос спинового момента на наноконтакт и, если спиновый потенциал достаточно большой, то намагниченность свободного контакта меняется с направления «вверх» на направление «вниз». Если теперь поменять полярность разности потенциалов, подаваемой на вентиль, то появление положительного потенциала на закрепленном контакте вытягивает из канала электроны со спиной «вниз» и таким образом меняет знак спинового потенциала на обратный

$$E_F^s \equiv E_F^{up} - E_F^{dn} > 0 \tag{55}$$

Опять же, если положительный спиновый потенциал достаточно большой, то он вернет намагниченность наноконтакта в исходное состояние. Этот эффект надежно экспериментально подтвержден, и представляется весьма вероятным, что он будет вскоре использоваться для записи информации на ФМ наноноситель так же, как явление магнитосопротивления сейчас широко используется для считывания информации, например, с жесткого диска.

9.5. Уравнение Ландау – Лифшица – Гильберта

Эти два экспериментальных достижения – магнитное генерирование избытка спинов одного сорта и обращение намагниченности образца за счет этого избытка фактически объединили спинтронику с магнетроникой (рис. 14)

в единую область исследований, в которой намагничивание и спиновый транспорт играют равновеликие роли.

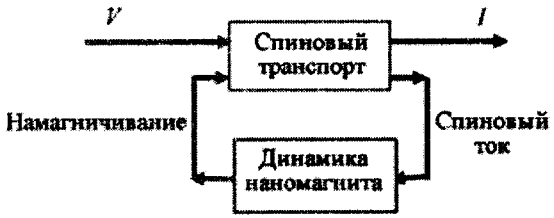


Рис. 14. Спиновый транспорт и динамика перемагничивания наномагнитов тесно связаны.

Модель, описывающая динамику перемагничивания наномагнитных структур под действием спинового тока, основана на уравнении Ландау – Лифшица – Гильберта (ЛЛГ) [31 – 34].

Магнитный момент электрона пропорционален магнетону Бора

$$\mu_{el} = \frac{g_s}{2}\mu_B, \tag{56}$$

$$\mu_B \equiv \frac{q\hbar}{2m} = 9.274 \cdot 10^{-24}\, A \cdot м^2, \tag{57}$$

где g-фактор g_s для спина электрона в вакууме очень близок к 2 (точнее равен 2.002329), но может существенно отличаться от 2 для электронов в твердых телах, что для нас сейчас не существенно, так что будем считать, что $g_s = 2$, а $\mu_{el} = \mu_B$. Из (57) видно, что магнитный момент в один магнетон Бора создается током в приблизительно 10 μA, циркулирующему по квадратному контуру со стороной в 1 *нм*.

В немагнитных телах все спины скомпенсированы. В магнитных телах величина намагниченности пропорциональна числу нескомпенсированных спинов N_s в объеме Ω:

$$M_s = \mu_B \frac{N_s}{\Omega}, \tag{58}$$

а направление вектора намагниченности, задаваемое его единичным вектором $\hat{m}$, меняется с магнитным полем $\vec{H}$ согласно уравнению ЛЛГ

$$(1+\alpha^2)\frac{d\hat{m}}{dt} = -\gamma\mu_0(\hat{m}\times\vec{H}) - \alpha\gamma\mu_0(\hat{m}\times\hat{m}\times\vec{H}), \tag{59}$$

где гиромагнитное отношение, как отношение заряда электрона к его массе,

$$\gamma \equiv \frac{q}{m} = \frac{2\mu_B}{\hbar}, \tag{60}$$

а магнитная постоянная $\mu_0 = 1/(\varepsilon_0 \cdot c^2)$ связана с электрической постоянной ε_0 через скорость света c.

В уравнении ЛЛГ (59) первое слагаемое описывает динамику намагниченности [32], а второе слагаемое – диссипацию динамического процесса с параметром затухания Гильберта α [33], характерное значение которого обычно ~ 0.01.

9.5.1. Выделенная ось магнита

Воспользуемся уравнением ЛЛГ для описания фундаментального экспериментального факта о наличии у магнита выделенной оси (пусть это будет ось z). Внешнее магнитное поле H_{ext}, если оно превышает некоторое критическое значение H_K, может быть использовано для изменения намагниченности между значениями $m_z = -1$ и $m_z = +1$ (рис. 15).

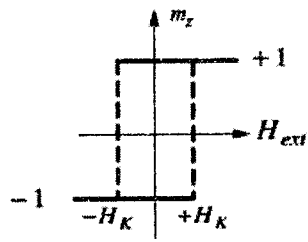


Рис. 15. Магнит имеет выделенную ось (пусть ось z). Внешнее магнитное поле H_{ext}, если оно превышает некоторое критическое значение H_K, меняет намагниченность между значениями $m_z = -1$ и $m_z = +1$.

С магнитным полем, направленным вдоль оси z,

$$\vec{H} = H\hat{z} \tag{61}$$

и пренебрегая $\alpha^2 \ll 1$, уравнение ЛЛГ упрощается до

$$\frac{d\hat{m}}{dt} = -\gamma\mu_0 H(\hat{m}\times\hat{z}) - \alpha\gamma\mu_0 H(\hat{m}\times\hat{m}\times\hat{z}). \tag{62}$$

Выполнив векторные произведения

$$\hat{m}\times\hat{z} = m_z, \quad (\hat{m}\times\hat{z})\cdot\hat{z} = 0, \quad -\hat{z}\cdot(\hat{m}\times\hat{m}\times\hat{z}) = 1 - m_z^2, \tag{63}$$

получим

$$\frac{dm_z}{dt} = (1 - m_z^2)\alpha\gamma\mu_0 H . \tag{64}$$

Равновесное состояние требует

$$\frac{dm_z}{dt} = 0, \tag{65}$$

так что единичный вектор намагниченности может принимать только два значения

$$m_z = -1 \text{ и } m_z = +1, \tag{66}$$

что и соответствует наличию у магнита выделенной оси.

Остается вопрос о стабильности решения уравнения (65). Пусть

$$m_z = +1 - \delta . \tag{67}$$

Тогда вместо (64) имеем

$$-\frac{d\delta}{dt} \approx (2\alpha\gamma\mu_0 H)\delta , \tag{68}$$

что означает невозможность отклонения m_z от +1 при положительном значении магнитного поля H. Аналогично, при

$$m_z = -1 - \delta \tag{69}$$

равенство

$$\frac{d\delta}{dt} \approx (2\alpha\gamma\mu_0 H)\delta \tag{70}$$

свидетельствует о невозможности отклонения m_z от −1 при отрицательном значении магнитного поля. Иначе говоря,

$$m_z = +1 \text{ устойчиво при } H > 0, \tag{71}$$

$$m_z = -1 \text{ устойчиво при } H < 0. \tag{72}$$

Теперь вернемся к рис. 15. Мы до сих пор не конкретизировали магнитное поле H. Оно включает в себя внешнее магнитное поле H_{ext} и внутреннее магнитное поле, которое каждый электрон чувствует со стороны всех остальных электронов со знаком, определяемым значением m_z,

$$H = H_{ext} + H_K m_z . \tag{73}$$

Теперь из условий устойчивости (71) и (72) следует

$$m_z = +1 \text{ устойчиво при } H_{ext} > -H_K, \tag{74}$$

$$m_z = -1 \text{ устойчиво при } H_{ext} < +H_K. \tag{75}$$

что и показано графически на рис. 15.

9.6. Обращение намагниченности спиновым током

Для обсуждения динамики намагничивания в уравнение ЛЛГ (59) добавим еще одно слагаемое ($\alpha^2 << 1$), а именно:

$$\frac{d\hat{m}}{dt} = -\gamma\mu_0(\hat{m}\times\vec{H}) - \alpha\gamma\mu_0(\hat{m}\times\hat{m}\times\vec{H}) - \left(\hat{m}\times\hat{m}\times\frac{\vec{I}_s}{qN_s}\right), \tag{76}$$

пропорциональное спиновому току $\vec{I}_s$ в пересчете на один спин, где N_s есть число спинов, обеспечивающих намагниченность. Почему дополнительный член берется в виде двойного векторного произведения

$$\hat{m}\times\hat{m}\times\frac{\vec{I}_s}{qN_s}, \tag{77}$$

а не просто

$$\frac{\vec{I}_s}{qN_s}? \tag{78}$$

Двойное векторное произведение $\hat{m}\times\hat{m}\times\frac{\vec{I}_s}{qN_s}$ с произвольным вектором $\vec{V}$ (рис. 16) сводится к вычитанию из вектора $\vec{V}$ компоненты этого вектора вдоль единичного вектора $\hat{m}$:

$$-\hat{m}\times\hat{m}\times\vec{V} = \vec{V} - (\hat{m}\cdot\vec{V})\hat{m}. \tag{79}$$

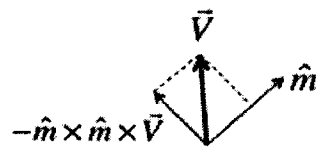


Рис. 16. К вычислению двойного векторного произведения.

Поэтому член $\hat{m}\times\hat{m}\times\frac{\vec{I}_s}{qN_s}$ равен компоненте вектора спинового тока $\frac{\vec{I}_s}{qN_s}$, перпендикулярной намагниченности, величина же намагниченности не изменяется, обращается только ее направление. Это гарантируется тем, что вся правая часть уравнения ЛЛГ должна быть перпендикулярна намагниченности.

Есть еще один дополнительный член в правой части уравнения ЛЛГ, также перпендикулярный намагниченности, а именно:

$$\alpha \hat{m} \times \frac{\vec{I}_s}{qN_s}, \tag{80}$$

но мы им пренебрегли, поскольку параметр затухання Гильберта α обычно очень мал.

Проектируя уравнение ЛЛГ (76) на выделенную ось, получаем

$$\frac{dm_z}{dt} = (1-m_z^2)\left(\alpha\gamma\mu_0 H_K m_z + \frac{I_s}{qN_s}\right). \tag{81}$$

Как и в случае с уравнением (64), критическое значение спинового тока, необходимое для обращения намагниченности, дается уравнением

$$\left(\frac{I_s}{qN_s}\right)_{crit} = \alpha\gamma\mu_0 H_K, \tag{82}$$

а с использованием (58) для критического значения спинового тока имеем

$$(I_s)_{crit} = \frac{4q\alpha}{\hbar}\left(\frac{1}{2}\mu_0 H_K M_s \Omega\right). \tag{83}$$

Величина в круглих скобках для критического тока есть энергия барьера, разделяющего два состояния магнита. Для устойчивого состояния магнита с той или иной намагниченностью (вверх или вниз) барьер должен быть не меньше нескольких десятков kT. В противном случае намагниченность магнита будет обращаться циклически практически бесконечно долго. При барьере $\sim 40kT$ и $\alpha = 0.01$ уравнение (83) для критического значения спинового тока дает $\sim 10\mu A$. Экспериментально наблюдались токи $\sim 50-100\mu A$.

Наглядные апплеты по динамике намагничивания с переносом спинового момента выставлены на сайте [35].

9.7. Поляризаторы и анализаторы спинового тока

Пусть регистрирующий ФМ контакт 2 расположен под углом по отношению к инжектирующему контакту (рис. 17). Какая разность потенциалов будет измерена?

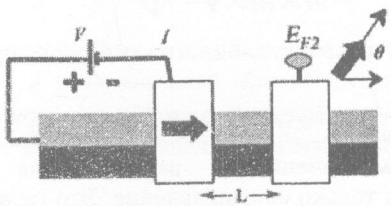


Рис. 17. Регистрирующий контакт 2 в роли анализатора спинового тока.

Ответ представляется довольно простым:

$$E_{F2} = E_F + \frac{\vec{P}_2 \cdot \vec{E}_F^s}{2}, \tag{84}$$

где вектор поляризации совпадает с направлением регистрирующего контакта, а вектор спинового потенциала совпадает с направлением спиновой поляризации канала проводимости, которое по договоренности есть направление намагниченности инжектирующего контакта. Ранее мы уже рассматривали два частных случая взаимной ориентации намагниченности контактов: параллельно P и антипараллельно AP (53). Как интерпретировать более общий результат (84) ?

Проведем аналогию с поляризацией потока фотонов. Интенсивность света, прошедшего через анализатор, пропорциональна квадрату косинуса угла между плоскостями пропускания поляризатора и анализатора $I/I_0 = \cos^2\theta$ (закон Малюса). Интенсивность прошедшего света максимальна при совпадении плоскостей пропускания поляризатора и анализатора ($\theta = 0°$) и минимальна, когда плоскости перпендикулярны ($\theta = 90°$). Иная ситуация с потоком электронов.

Пусть все электроны в потоке имеют спин «вверх». Тогда по определению (25) и (26)

$$E_F^s = E_F^{up} = 2E_F, \tag{85}$$

если же повернуть намагниченность на регистрирующем контакте на угол θ, измеряемая разность потенциалов, как следует из (84), станет

$$\frac{E_{F2}}{E_F} = 1 + P_2 \cos\theta. \tag{86}$$

Как и в случае потока фотонов, разность потенциалов максимальна, когда регистрирующий и инжектирующий контакты параллельны ($\theta = 0°$). Если же в случае потока фотонов интенсивность прошедшего через анализатор света минимальна при $\theta = 90°$, то в случае потока электронов минимум разности потенциалов достигается при антипараллельной ($\theta = 180°$) ориентации намагниченности контактов (рис. 18).

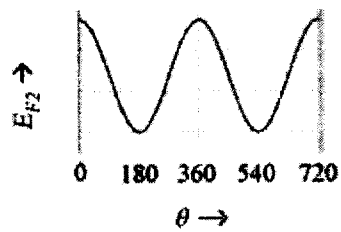


Рис. 18. Колебания нелокального спинового потенциала в зависимости от угла между инжектирующим и регистрирующим ФМ контактами.

В предположении идеального регистрирующего контакта ($P_2 = 1$) из (86) следует

$$\frac{E_{F2}}{E_F} = 1 + \cos\theta = 2\cos^2\frac{\theta}{2}, \qquad (87)$$

так что если анализатор фотонов пропускает через себя количество фотонов, пропорциональное $cos^2\theta$, то спиновый анализатор электронов пропускает через себя количество электронов, пропорциональное $cos^2(\theta/2)$. Есть надежда, что уже в недалеком будущем спиновый анализатор электронов будет ключевым измерительным устройством в спиновом квантовом компьютере так же, как закон Малюса уже сейчас используется в фотонных квантовых компьютерах.

Подведем итоги. В рамках концепции «снизу – вверх» рассмотрены ключевые вопросы спинтроники – спиновый вентиль, граничное сопротивление при несовпадении мод проводимости, спиновые потенциалы и разность нелокальных спин-потенциалов, спиновый момент и его транспорт, уравнение Ландау – Лифшица – Гильберта и на его основе обсуждается выделенная ось магнита, рассматриваются обращение намагниченности спиновым током, поляризаторы и анализаторы спинового тока [36 – 41].

Литература

1. Ю. О. Кругляк, Н. Ю. Кругляк, М. В. Стріха. Уроки наноелектроніки: Виникнення струму, формулювання закону Ома і моди провідності в концепції «знизу – вгору», *Сенсор. електр. мікросист. техн.*, т. 9, № 4, 5 – 29 (2012).
2. Ю. А. Кругляк. Наноэлектроника «снизу – вверх»: Возникновение тока, обобщенный закон Ома, упругий резистор, моды проводимости, термоэлектричество, *ScienceRise*, т. 7, № 2(12), 76 – 100 (2015).
3. Datta Supriyo. *Lessons from Nanoelectronics: A New Perspective on Transport*. Hackensack, New Jersey: World Scientific Publishing Co., 2012; www.nanohub.org/courses/FoN1.
4. M. I. Dyakonov, V. I. Perel. Current-induced spin orientation of electrons in semiconductors, *Phys. Lett. A*, v. 35, 459 – 460 (1971).
5. M. Julliere. Tunneling between ferromagnetic films, *Phys. Lett. A*, v. 54, n 3, 225 – 226 (1975).
6. А. Г. Аронов, Г. Е. Пикус. Спиновая инжекция в полупроводниках, Физика техн. полупров., № 10, 1177 – 1180 (1976).
7. M. N. Baibich, J. M. Broto, A. Fert, F. Nguyen Van Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friederich, J. Chazelas. Giant Magnetoresistance of (001)Fe/(001)Cr Magnetic Superlattices, *Phys. Rev. Lett.*, v. 61, n 21, 2472 – 2475 (1988).

8. G. Binasch, P. Grünberg, F. Saurenbach, W. Zinn. Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange, *Phys. Rev. B*, v. 39, 4828 – 4830 (1989).
9. N. F. Mott. The Electrical Conductivity of Transition Metals, *Proc. Roy. Soc.*, v. 153, 699 – 717 (1936).
10. N. F. Mott. Electrons in Transition Metals, *Adv. Phys.*, v. 13., 325 (1964).
11. А. М. Погорілий, С. М. Рябченко, О. І. Товстолиткін. Спінтроніка. Основні явища. Тенденції розвитку, *Укр. фіз. журн. Огляди*, т. 6, № 1, 37 – 97 (2010).
12. G. Schmidt. Concepts for Spin Injection into Semiconductors – a Review, *J. Phys. D: Appl. Phys.*, v. 38, R107 – R122 (2005).
13. T. Valet, A. Fert. Theory of the perpendicular magnetoresistance in magnetic multilayers, *Phys. Rev. B.*, v. 48, 7099 (1993).
14. F. W. Sears, G. L. Salinger. *Thermodynamics, Kinetic Theory, and Statistical Thermodynamics*. Boston: Addison-Wesley, 1975.
15. R. Kubo. Statistical-Mechanical Theory of Irreversible Processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems, *J. Phys. Soc. Japan*, v. 12, 570 – 586 (1957).
16. P. C. Martin, J. Schwinger. Theory of many-particle systems. I, *Phys. Rev.*, v. 115, n 6, 1342 – 1373 (1959).
17. L. P. Kadanoff, G. Baym. *Quantum Statistical Mechanics*. New York: W. A. Benjamin, 1962.
18. L. V. Keldysh. Diagram Technique for Non-Equilibrium Processes, *ЖЭТФ*, т. 47, 1515 – 1527 (1964); *Sov. Phys. JETP*, v. 20, 1018 (1965).
19. S. Takahashi, S. Maekawa. Spin Injection and Detection in Magnetic Nanostructures, *Phys. Rev. B*, v. 67, 052409 (2003).
20. О. В. Третяк, В. А. Львов, О. В. Барабанов. *Фізичні основи спінової електроніки*. Київ: Вид-во Київського університету, 2002.
21. Ю. А. Данилов, Е. С. Демидов, А. А. Ежевский. *Основы спинтроники*. Нижний Новгород: Нижегородский гос. ун-т, 2009.
22. С. С. Аплеснин. *Основы спинтроники*. Санкт-Петербург: Лань, 2010.
23. M. Tsoi, A. G. M. Jansen, J. Bass, W.-C. Chiang, M. Seck, V. Tsoi, P. Wyder. Excitation of a Magnetic Multilayer by Electric Current, *Phys. Rev. Lett.*, v. 80, 4281 (1998).
24. E. B. Myers, D. C. Ralph, J. A. Katine, R. N. Louie, R. A. Buhrman. Current-Induced Switching of Domains in Magnetic Multilayer Devices, *Science*, v. 285, 867 – 870 (1999).
25. J. A. Katine, F. J. Albert, R. A. Buhrman, E. B. Myers, D. C. Ralph. Current-Driven Magnetization Reversal and Spin-Wave Excitations in Co/Cu/Co Pillars, *Phys. Rev. Lett.*, v. 84, 3149 – 3152 (2000).
26. L. Berger. Emission of spin waves by a magnetic multilayer traversed by a current, *Phys. Rev. B.*, v. 54, n 13, 9353 – 9358 (1996).

27. J. C. Slonczewski. Current-driven excitation of magnetic multilayers, *J. Magn. Magn. Mater.*, v. 159, L1 (1996).
28. Y. B. Bazaliy, B. A. Jones, S.-C. Zhang. Modification of the Landau-Lifshitz equation in the presense of a spin-polarized current and colossal- and giant-magnetoresistive materials, *Phys. Rev. B*, v. 57, R3213 – R3216 (1998).
29. J. Z. Sun. Spin-current interaction with a monpdomain magnetic body: A model study, *Phys. Rev. B*, v. 62, 570 (2000).
30. D. C. Ralph, M. D. Stiles. Spin transfer torques, *J. Magn. Magn. Mater.*, v. 320, 1190 – 1216 (2008).
31. Л. Д. Ландау, Е. М. Лифшиц. К теории дисперсии магнитной проницаемости ферромагнитных тел, *Phys. Z. Sowjetunion*, т. 8, 153 – 169 (1935).
32. Л. Д. Ландау, Е. М. Лифшиц. К теории дисперсии магнитной проницаемости ферромагнитных тел, В кн. *Л. Д. Ландау. Собрание трудов*, т. 1. Москва: Наука, 1969.
33. T. Gilbert. A phenomenological theory of damping in ferromagnetic materials, *IEEE Trans. Magn.*, v. 40, n 6, 3443 – 3449 (2004).
34. А. К. Звездин, К. А. Звездин, А. В. Хвальковский. Обобщенное уравнение Ландау — Лифшица и процессы переноса спинового момента в магнитных наноструктурах, *Усп. физич. н.*, т. 178, 436 – 442 (2008).
35. T. Mewes. Magnetization dynamics including spin-torque et al: www.bama.ua.edu/~tmewes/.
36. Ю. О. Кругляк, Н. Ю. Кругляк, М. В. Стріха. Уроки наноелектроніки: Спінтроніка в концепції «знизу – вгору», *Сенсор. електр. мікросист. техн.*, т. 10, № 4, 5 – 37 (2013).
37. Ю. А. Кругляк. Начала спинтроники в концепции "снизу – вверх", *Вестник Одес. гос. эколог. ун-та*, № 16, 168 – 195 (2013).
38. Ю. А. Кругляк, Н. Е. Кругляк. Спинтроника: граничное сопротивление и несовпадение мод проводимости, *Proc. Intern. Conf: Nanomaterials: Applications and Properties*, v. 2, N 1, 01001(4) (2013).
39. Ю. А. Кругляк. Уроки наноэлектроники. 6. Начала спинтроники в концепции "снизу – вверх", *Физич. образов. в вузах*, т. 20, № 2, 88 – 108 (2014).
40. Ю. А. Кругляк, П. А. Кондратенко, Ю. М. Лопаткин. Спинтроника в концепции «снизу – вверх», *Ж. нано- электр. физ.*, т. 6, № 4, 04042(6) (2014).
41. Ю. А. Кругляк. Наноелектроника «снизу – вверх»: начала спинтроники и магнетроники, *ScienceRise*, т. 8, № 2(13), 51 – 68 (2015).

Глава 10. Метод неравновесных функций Грина

10.1. Квантовый транспорт

Продолжая развитие концепции «снизу – вверх» применительно к наноэлектронике [1 – 5], начнем рассмотрение квантового транспорта с формулировки метода неравновесных функций Грина (НРФГ) в матричном представлении и применим этот метод для анализа и решения модельных транспортных задач переноса электронов на примере однородных и неоднородных 1D и 2D проводников, в том числе и графена, а также сформулируем общий метод учета электрических контактов в уравнении Шредингера. Затем рассмотрим модели упругой дефазировки и спиновой дефазировки, учет некогерентных процессов с использованием зонда Бюттекера, 1D проводник с двумя и более рассеивающими центрами, явление квантовой интерференции, режимы сильной и слабой локализации, скачок потенциала на дефектах, квантовые осцилляции в методе НРФГ без учета дефазировки и с ее учетом, эффекты деструктивной и конструктивной интерференции, четырехкомпонентное описание спинового транспорта с учетом дефазировки. Закончим эту главу обсуждением квантовой природы классики на таких примерах как спиновая когерентность, псевдоспин, квантовая энтропия, спины и магниты, ответим на вопросы увеличивает ли энтропию взаимодействие частиц и сколько информации переносит один электрон.

Современный ноутбук содержит более 4 миллиардов полевых транзисторов и резисторов с длиной канала проводимости $\sim 30\,нм$ и даже меньше, что соответствует сотне – другой атомов.

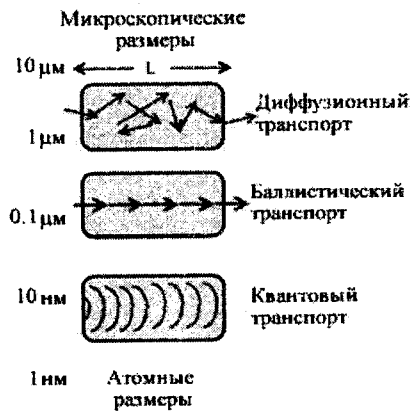


Рис. 1. С уменьшением длины канала проводимости L физическая природа транспорта электронов изменяется качественно -- от диффузионного к баллистическому и далее к квантовому.

По мере уменьшения длины канала проводимости физическая природа электронного транспорта изменяется качественно (рис. 1). Для достаточно длинных проводников транспорт является диффузионным с траекторией, напоминающей случайные блуждания. Если длина канала проводимости становится меньше средней длины свободного пробега, транспорт электронов переходит в режим баллистического переноса. При еще более коротких длинах канала проводимости начинает проявляться волновая природа электронов в виде таких квантовых эффектов как интерференция и туннелирование.

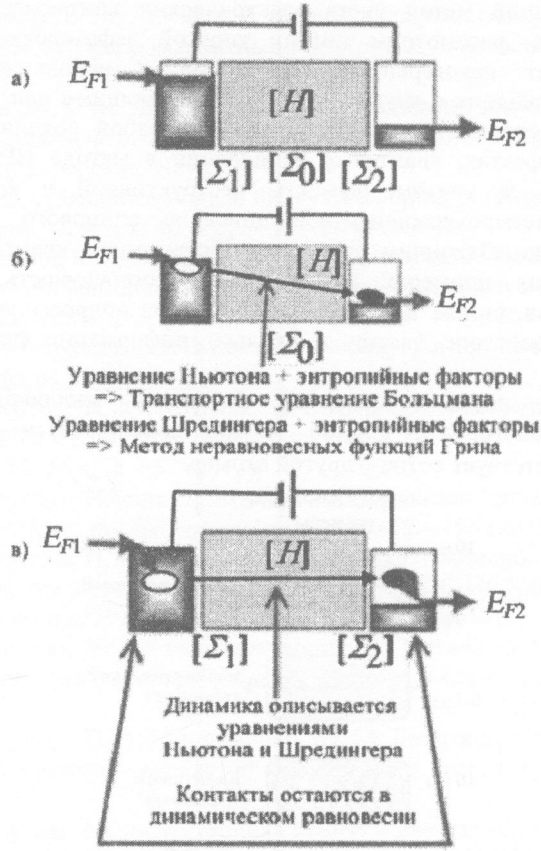


Рис. 2. Схематическое описание (а) активного канала проводимости, взаимодействующего с истоком и стоком, и двух предельных случаев – (б) диффузионного транспорта в макроскопических проводниках и (в) баллистического транспорта в нанотранзисторах.

Исторически все более глубокое понимание физической природы электрической проводимости приходило ***сверху – вниз***: от массивных макроскопических проводников до транзисторов молекулярных и даже атомных размеров. Еще лет 20 – 25 тому назад были обычными рассуждения о том, как понимать концепцию об электрическом сопротивлении, если размер проводника приближается к атомным размерам. Несмотря на исключительные достижения в области экспериментальной мезо- и нанофизики, и поныне при обсуждении проводимости доминирует концепция сверху – вниз вместо более естественной концепции ***снизу – вверх***, что делает анализ и обсуждение устройств наноэлектроники подчас нелепо сложными [6, 7]. Концепцию снизу – вверх проиллюстрируем на рис. 2.

Любое устройство наноэлектроники имеет активный канал проводимости, который описывается гамильтонианом $[H]$, включающем потенциальную энергию U, обязанную всем другим зарядам как внешним (на контактах), так и внутренним (в самом канале). Канал проводимости взаимодействует с истоком и стоком и с любыми другими контактами в конкретном устройстве, которые находятся в локальном равновесии, определяемом соответствующими электрохимическими потенциалами (рис. 2а). Взаимодействие между каналом и контактами описывается матрицами собственной энергии (self-energy) $[\Sigma_1]$ и $[\Sigma_2]$ [8]. Взаимодействие электрона в канале с его окружением описывается матрицей собственной энергии $[\Sigma_0]$, которая в отличие от матриц $[\Sigma_1]$ и $[\Sigma_2]$, должна быть вычислена самосогласовано. Размерность этих квадратных матриц определяется числом N базисных функций, используемых для квантовомеханического описания канала проводимости и контактов. Конкретный вид матриц определяется используемым методом решения уравнения Шредингера – полуэмпирическим, на основе теории функционала плотности или из первых принципов, а также выбором базисных функций. Как только эти матрицы составлены, дальнейшая процедура вычисления проводимости, тока и других электрофизических свойств является стандартной, чему в основном и посвящена настоящая глава на примере модельных транспортных задач, представляющих не только научный, но и педагогический и даже познавательный интерес.

На рис. 2 показаны два предельных случая электронного транспорта – диффузионный (рис. 2б) и баллистический (рис. 2в). В баллистическом пределе транспорт электронов контролируется контактными матрицами $[\Sigma_1]$ и $[\Sigma_2]$, тогда как взаимодействия внутри канала пренебрежимо малы. В противовес этому, в диффузионном пределе транспорт электронов контролируется взаимодействиями внутри канала, описываемыми матрицей $[\Sigma_0]$, а роль контактных матриц $[\Sigma_1]$ и $[\Sigma_2]$ пренебрежимо мала. Не удивительно, что до примерно 1990 года контакты даже не изображались на схемах. Между матрицей гамильтониана $[H]$ и матрицами $[\Sigma_{0,1,2}]$ есть существенное различие:

матрица гамильтониана представляет консервативные динамические силы и является эрмитовой, тогда как матрицы собственной энергии учитывают энтропийные факторы и не являются эрмитовыми.

Уравнение Шредингера само по себе не пригодно для объяснения таких, кажущихся очевидными процессов, как например, самопроизвольный переход электрона из возбужденного состояния в основное и невозможность самопроизвольного обратного процесса. Эта тенденция для систем любой сложности релаксировать однонаправлено в сторону понижения энергии кажется очевидной и не имеет объяснения в рамках квантовой механики. Подобные процессы, как и разнообразные однонаправленные явления в окружающем нас мире, имеют энтропийную природу. При конструировании и анализе работы любого электронного устройства, квантового или классического, уравнения динамики должны быть дополнены «энтропийными силами». Так возникла статистическая механика неравновесных процессов, центральное место в которой вот уже 140 лет занимает транспортное уравнение Больцмана (глава 5) [9, 10]. Квантовым аналогом уравнения Больцмана является метод неравновесных функций Грина, основы которого были заложены в работах Мартина и Швингера [11], Каданова и Бейма [12] и Келдыша [13].

Оба подхода – классический Больцмана и квантовый формализм НРФГ объединяет то обстоятельство, что в них одновременно учитываются и динамические и энтропийные процессы. В баллистическом пределе, однако, динамические и энтропийные процессы пространственно разделены (рис. 3в). Электроны проскакивают от одного контакта к другому под действием только динамических сил. Внутри контактов электроны оказываются не в равновесии, но быстро приходят в равновесное состояние под действием энтропийных сил. Подобную модель развития событий в баллистическом пределе называют моделью упругого резистора Ландауэра, предложенную Рольфом Ландауэром еще в 1957 году [14 – 16] задолго до ее триумфального экспериментального подтверждения в нанотранзисторах. Сегодня действительно надежно установлено, что баллистические резисторы выдерживают довольно сильные токи благодаря тому, что выделение джоулева тепла пренебрежимо мало. Тепло выделяется на контактах, которые благодаря своей относительной массивности быстро диссипируют его. Пространственное разнесение динамики и термодинамики в баллистических устройствах служит весомым аргументом в пользу концепции «снизу – вверх», которая представляется нам привлекательной не только научно, но и педагогически.

10.2. Уравнения метода НРФГ в матричном представлении

Задачу в этой главе мы видим в том, чтобы дать компактное изложение формализма НРФГ с учетом модели Ландауэра применительно к наноэлектронным устройствам. Если опираться на пионерские работы от Швингера до Келдыша [11 – 13], основанные на квантовой многочастичной

теории возмущений (МЧТВ) и диаграммной технике, то не хватит и нескольких семестров для овладения методом НРФГ. В своем изложении мы будем следовать работам Датты, Мейра и Вингрина [4 – 7, 17 – 20], как наиболее адекватным стоящей перед нами задачи.

Начнем с рассмотрения упругого резистора, энтропийные процессы диссипации тепла в котором происходят лишь на контактах, а саму задачу о сопротивлении резистора рассмотрим в простейшем одночастичном приближении квантовой химии методом Хюккеля, который в физике твердого тела более известен под названием метод сильной связи в ортогональном базисе с учетом соседних взаимодействий. В методе Хюккеля решение уравнения Шредингера

$$\hat{H}\Psi = E\Psi$$

сводится к задаче на собственные значения

$$\left[H\right]\{\psi\} = E\{\psi\}, \tag{1}$$

где $[H]$ – матрица гамильтониана, а $\{\psi\}$ – столбцовые матрицы решений задачи (1). Для учета контактов проводника с источником тока к матрице гамильтониана добавляются еще два члена, описывающих отток электронов из проводника в контакты (outflow),

$$\left[\Sigma\right] = \left[\Sigma_1\right] + \left[\Sigma_2\right] \tag{2}$$

и приток электронов в проводник с контактов (inflow)

$$\{s\} = \{s_1\} + \{s_2\}, \tag{3}$$

а именно:

$$E\{\psi\} = \left[H\right]\{\psi\} + \left[\Sigma\right]\{\psi\} + \{s\}. \tag{4}$$

Теперь решение уравнения Шредингера для проводника с учетом контактов можно сразу записать через обратную матрицу:

$$\{\psi\} = [EI - H - \Sigma]^{-1}\{s\}, \tag{5}$$

где I – единичная матрица.

Матрица

$$G^R = [EI - H - \Sigma]^{-1} \tag{6}$$

получила название запаздывающей (Retarded) функции Грина, а эрмитово сопряженная ей матрица

$$G^A = [G^R]^+ \tag{7}$$

называется опережающей (Advanced) функцией Грина. Происхождение этих и других терминов, общепринятых в формализме НРФГ, нам для дальнейшего не

существенно, подробнее познакомиться с терминологией рекомендуем работы [6, 11 – 13]. Отметим лишь, что формализм НРФГ в применении к задачам в наноэлектронике сводится к четырем уравнениям, **первое** из которых есть выражение (6) для запаздывающей функции Грина. Теперь уравнение Шредингера (5) перепишем в виде

$$\{\psi\} = [G^R]\{s\}. \tag{8}$$

Произведение столбца $\{\psi\}$ на эрмитово сопряженную ему строку $\{\psi\}^+$ дает

$$\{\psi\}\{\psi\}^+ = [G^R]\{s\}\{s\}^+[G^A], \tag{9}$$

где учтено, что транспонирование произведения матриц меняет порядок сомножителей на обратный.

Неравновесная функция Грина определяется как

$$G^n = 2\pi\{\psi\}\{\psi\}^+, \tag{10}$$

так что число электронов дается выражением

$$N = Tr[G^n]/2\pi. \tag{11}$$

Аналогично описывается приток электронов, а именно:

$$\Sigma^{in} = 2\pi\{s\}\{s\}^+, \tag{12}$$

так что теперь уравнение (9) имеет вид

$$G^n = G^R\Sigma^{in}G^A \tag{13}$$

и будет **вторым** уравнением в формализме НРФГ.

Полноты ради отметим, что обозначения по сравнению с [6] упрощены, а именно: Σ вместо Σ^R, G^n вместо $-iG^<$, Σ^{in} вместо $-i\Sigma^<$, однако, первое и второе уравнения (6) и (13) существенно те же, что и уравнения (75) – (77) в основополагающей работе Келдыша [13], полученные диаграммной техникой МЧТВ. Мы полагаем, что вычленение этих двух уравнений, а далее еще двух из МЧТВ, первоначально использованной для их вывода, сделает метод НРФГ более прозрачным и доступным, и он будет читаться студентам физикам и инженерам-электронщикам для решения задач, когда возникает необходимость учесть электрические контакты в уравнении Шредингера.

Приведем оставшиеся два уравнения формализма НРФГ. **Третье** уравнение есть матричная форма плотности состояний $D(E)$, умноженной на 2π, и называется спектральной функцией A:

$$2\pi \cdot D(E) = A = G^R\Gamma G^A = G^A\Gamma G^R = i[G^R - G^A], \tag{14}$$

где матрицы G^R и G^A даются уравнениями (6) и (7), а матрица [Γ] есть антиэрмитовая часть соответствующей контактной матрицы (2):

$$\Gamma = i[\Sigma - \Sigma^+] \tag{15}$$

и описывает взаимодействие электронов в канале с контактами. Вывод формулы (14) для спектральной функции и эквивалентность всех трех ее выражений сделаем позже в этой главе (§ 10.7).

Все матрицы [Σ], [Γ] и [Σ^{in}] содержат слагаемые, относящиеся к конкретным терминалам (контактам), учитываемым в конкретной задаче. В уравнениях (6), (13) и (14) суммирование по терминалам уже выполнено.

Четвертым уравнением в формализме НРФГ служит уравнение для тока через терминал с номером m

$$\tilde{I}_m = \frac{q}{h} Tr\left[\Sigma_m^{in} A - \Gamma_m G^n\right], \tag{16}$$

куда входят только те компоненты матриц, которые относятся к данному терминалу m. Это удельный ток (на единицу энергии), его нужно еще проинтегрировать по всему спектру энергий, чтобы получить полный ток через терминал m.

Далее рассмотрим простые модельные транспортные задачи, затем вернемся к обоснованию и выводу уравнений метода НРФГ и обсудим более сложные транспортные задачи.

10.3. Одноуровневый резистор: полуклассический подход

Почувствовать физику метода НРФГ можно уже на простейшей транспортной задаче об одноуровневом проводнике, описываемом матрицами 1×1, т.е. числами с $[H] = \varepsilon$. На этом примере увидим как учитываются контакты в уравнении Шредингера, а затем перейдем к проводнику с произвольным числом мод, описываемых матрицами $N \times N$. Сначала, однако, рассмотрим одноуровневую задачу в полуклассическом приближении (рис. 3).

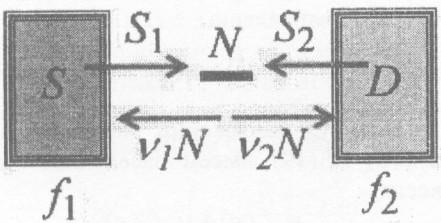


Рис. 3. Модель одноуровневого проводника в полуклассическом приближении.

Одноуровневый проводник контактирует с двумя контактами с фермиевскими заселенностями $f_1(\varepsilon)$ и $f_2(\varepsilon)$. Поначалу предположим, что на истоке S фермиевская функция $f_1 = 1$, а на стоке D функция $f_2 = 0$. Это означает, что исток стремится заполнить уровень с энергией ε, а сток стремится только извлечь электроны с этого уровня. Окончательно при вычислении суммарного тока помножим его на $f_1(\varepsilon) - f_2(\varepsilon)$, имея в виду, что впрыскивание электронов происходит с обоих контактов, и суммарный ток есть разностный эффект.

При $f_1 = 1$ на истоке и $f_2 = 0$ на стоке среднее число электронов N удовлетворяет уравнению

$$\frac{dN}{dt} = -(\nu_1 + \nu_2)N + S_1 + S_2, \tag{17}$$

где ν_1 и ν_2 есть скорости, с которыми электроны покидают проводник в направлении контактов, а S_1 и S_2 есть скорости, с которыми электроны впрыскиваются контактами в проводник. В условиях динамического равновесия $dN/dt = 0$, так что заселенность одиночного уровня электронами

$$N = \frac{S_1 + S_2}{\nu_1 + \nu_2}. \tag{18}$$

Если мысленно отключить сток, то число электронов будет равно фермиевской функции на истоке, если же отключить исток, то число электронов будет равно фермиевской функции на стоке:

$$\frac{S_1}{\nu_1} = f_1(\varepsilon) \quad \text{и} \quad \frac{S_2}{\nu_2} = f_2(\varepsilon). \tag{19}$$

Перегруппируем уравнение (17) в

$$\frac{dN}{dt} = (S_1 - \nu_1 N) + (S_2 - \nu_2 N), \tag{20}$$

тогда первое слагаемое есть поток электронов, создаваемый истоком, а второе слагаемое – стоком. В условиях динамического равновесия оба потока одинаковы и противоположно направлены, а сам ток равен

$$I = q(S_1 - \nu_1 N) = q(\nu_2 N - S_2), \tag{21}$$

при этом любое из этих двух выражений может быть использовано для вычисления тока. Из (18) и (19) имеем заселенность уровня в условиях динамического равновесия

$$N = \frac{\nu_1 f_1(\varepsilon) + \nu_2 f_2(\varepsilon)}{\nu_1 + \nu_2}, \tag{22}$$

подставляя которую в выражение для тока (21) и учитывая (19), для тока I через фермиевские функции окончательно имеем

$$I = q\frac{\nu_1\nu_2}{\nu_1+\nu_2}(f_1(\varepsilon)-f_2(\varepsilon)). \tag{23}$$

10.4. Одноуровневый резистор: квантовый подход

Стационарное уравнение Шредингера

$$[H]\{\psi\} = E\{\psi\}$$

следует из зависящего от времени уравнения Шредингера

$$i\hbar\frac{\partial}{\partial t}\{\tilde{\psi}(t)\} = [H]\{\tilde{\psi}(t)\} \tag{24}$$

в результате подстановки

$$\{\tilde{\psi}(t)\} = \{\psi\}e^{-iEt/\hbar}. \tag{25}$$

Для описания динамического равновесия обычно достаточно стационарного уравнения Шредингера, но в некоторых случаях не обойтись без нестационарного уравнения (24), например, как мы увидим далее, при интерпретации некоторых зависимостей от матриц собственных энергий.

Для одноуровневой задачи $[H]=\varepsilon$ эволюция волновой функции описывается уравнением

$$ih\frac{d}{dt}\tilde{\psi} = \varepsilon\tilde{\psi}, \tag{26}$$

с использованием которого и ему комплексно сопряженного имеем

$$\frac{d}{dt}(\tilde{\psi}\tilde{\psi}^*) = 0; \tag{27}$$

другими словами заселенность изолированного уровня энергии не изменяется со временем.

Нас же интересует не изолированная система, а проводник, соединенный с двумя контактами. Стандартные учебники по квантовой механике, к сожалению, не показывают как описать интересующую нас ситуацию. Модифицируем уравнение Шредингера следующим образом

$$ih\frac{d}{dt}\tilde{\psi} = \left(\varepsilon - i\frac{\gamma_1+\gamma_2}{2}\right)\tilde{\psi} \tag{28}$$

с тем, чтобы эволюция электронной плотности

$$\frac{d}{dt}\tilde{\psi}\tilde{\psi}^* = -\left(\frac{\gamma_1+\gamma_2}{\hbar}\right)\tilde{\psi}\tilde{\psi}^* \tag{29}$$

походила на динамическое уравнение полуклассики (17), разве что за исключением скоростей S_1 и S_2, с которыми электроны впрыскиваются контактами в проводник, но мы к ним вернемся позже. Уравнения (29) и (17) согласуются друг с другом, если положить

$$\gamma_1 = \hbar\nu_1, \tag{30}$$

$$\gamma_2 = \hbar\nu_2. \tag{31}$$

Стационарный аналог уравнения (28)

$$E\psi = \left(\varepsilon - i\frac{\gamma_1+\gamma_2}{2}\right)\psi \tag{32}$$

получается подстановкой в него решения для одиночного значения энергии $E=\varepsilon$:

$$\tilde{\psi}(t) = \psi(E)e^{-iEt/\hbar}. \tag{33}$$

Уравнение (32) имеет очевидное решение $\psi = 0$, что указывает на то, что в динамическом равновесии электроны не могут заселять уровень с энергией $E=\varepsilon$. Пока не включен источник S_1 поступления электронов в проводник электроны могут лишь покидать проводник, уходя в контакты (рис. 4).

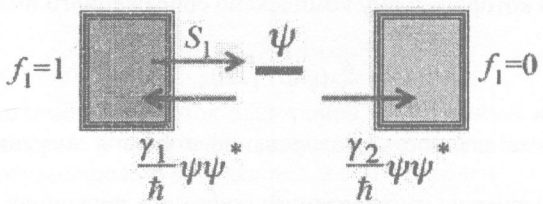


Рис. 4. Опорожнение и поступление электронов в квантовой модели одноуровнего проводника.

Введем в стационарное уравнение Шредингера (32) контактный член s_1 в качестве источника электронов:

$$E\psi = \left(\varepsilon - i\frac{\gamma}{2}\right)\psi + s_1, \tag{34}$$

где $\gamma = \gamma_1 + \gamma_2$.

В отличие от полуклассической модели (17) в квантовой модели вводится один источник электронов, а не два, причина чего станет ясной ниже.

Уравнение (34) позволяет связать волновую функцию с источником электронов

$$\psi = \frac{s_1}{E - \varepsilon + i(\gamma/2)}. \tag{35}$$

Обратим внимание на то, что волновая функция приобретает свое максимальное значение, когда энергия электрона E становится равной энергии уровня ε. Особенность квантовой модели в том, что волновая функция не существенно исчезает при отклонении E от ε на величину, меньшую γ. Это пример «уширения» или неопределенности энергии, отсутствующих в полуклассике.

Оценим качество источника электронов s_1 следующим образом. Проинтегрируем полное число электронов по всему спектру энергий и приравняем его полуклассическому выражению (22) с учетом того, что $f_1 = 1$ и $f_2 = 0$ (рис. 4):

$$\int_{-\infty}^{+\infty} dE\,\psi\psi^* = \frac{\nu_1}{\nu_1 + \nu_2} = \frac{\gamma_1}{\gamma_1 + \gamma_2}, \tag{36}$$

где второе равенство получено с учетом (30) и (31).

Вычислим левую часть уравнения (36), воспользовавшись выражением для волновой функции (35),

$$\int_{-\infty}^{+\infty} dE\,\psi\psi^* = \int_{-\infty}^{+\infty} dE \frac{s_1 s_1^*}{(E-\varepsilon)^2 + \left(\frac{\gamma}{2}\right)^2} = \frac{2\pi s_1 s_1^*}{\gamma}, \tag{37}$$

где

$$\int_{-\infty}^{+\infty} dE \frac{\gamma}{(E-\varepsilon)^2 + \left(\frac{\gamma}{2}\right)^2} = 2\pi. \tag{38}$$

Приравнивая (36) и (37), получим

$$2\pi s_1 s_1^* = \gamma_1. \tag{39}$$

Другими словами, качество источника электронов пропорционально скорости удаления электронов из проводника, что представляется правдоподобным: если контакт хорошо соединен с проводником, то электроны так же хорошо покидают контакт, как и возвращаются назад из проводника.

Как и в случае классического выражения для тока (21), ток в квантовой модели получим из скорости изменения электронной плотности (29):

$$\frac{d}{dt}\tilde{\psi}\tilde{\psi}^* = (\text{Приток из контакта 1}) - \frac{\gamma_1}{\hbar}\tilde{\psi}\tilde{\psi}^* - \frac{\gamma_2}{\hbar}\tilde{\psi}\tilde{\psi}^* \qquad (40)$$

добавив лишь приток из инжектирующего контакта s_1, не учтенного во временном уравнении Шредингера (28).

Левая и правая части уравнения (40) равны нулю, поскольку речь идет о токе в состоянии динамического равновесия. Как и в классической модели, ток можно вычислить либо как сумму первых двух слагаемых, либо как третье слагаемое в уравнении (40):

$$\frac{I}{q} = (\text{Приток из контакта 1}) - \frac{\gamma_1}{\hbar}\tilde{\psi}\tilde{\psi}^* = \frac{\gamma_2}{\hbar}\tilde{\psi}\tilde{\psi}^*. \qquad (41)$$

Интегрируя по всему спектру энергий, для тока получаем

$$I = q\int_{-\infty}^{+\infty} dE\frac{\gamma_2}{\hbar}\psi\psi^* \qquad (42)$$

и с подстановкой (35) и (39) для тока в квантовой модели окончательно имеем

$$I = \frac{q}{\hbar}\frac{\gamma_1\gamma_2}{2\pi}\int_{-\infty}^{+\infty} dE\frac{1}{(E-\varepsilon)^2+(\gamma/2)^2}, \qquad (43)$$

что можно сравнить с выражением для тока в полуклассической модели (23), имея в виду, что $f_1 = 1$ и $f_2 = 0$, а $\gamma = \gamma_1 + \gamma_2$:

$$I = \frac{q}{\hbar}\frac{\gamma_1\gamma_2}{\gamma_1-\gamma_2}. \qquad (44)$$

10.5. Квантовое уширение

Вычисление тока в квантовой модели по (43) предполагает интегрирование по всему спектру энергий, поскольку при квантовом рассмотрении одиночный локализованный уровень размывается в непрерывное распределение по энергии (рис. 5) согласно плотности состояний $D(E)$:

$$D = \frac{\gamma/2\pi}{(E-\varepsilon)^2+(\gamma/2)^2}. \qquad (45)$$

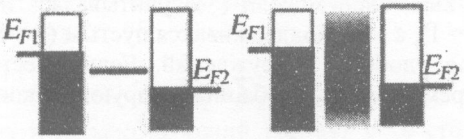


Рис. 5. В отличие от классических представлений (слева) одиночный локализованный уровень энергии в квантовой модели (справа) размывается вследствие принципа неопределенности Гейзенберга.

Прямое экспериментальное измерение проводимости одиночной молекулы водорода [21], спектр которой фактически соответствует одноуровневой модели резистора, может служить прямым доказательством справедливости квантового уширения уровней энергии.

Сравнивая (43) с выражением для тока упругого резистора (201/гл.1), получаем проводимость одноуровневой модели с учетом квантового уширения

$$G(E)=\frac{q^2}{h}\frac{\gamma_1\gamma_2}{(E-\varepsilon)^2+(\gamma/2)^2}. \tag{46}$$

Предполагая равноценную связь проводника с обоими контактами

$$\gamma_1=\gamma_2=\frac{\gamma}{2} \tag{47}$$

и достаточно низкую температуру для того, чтобы измеряемая проводимость была равной $G(E=E_{F0})$, имеем

$$G\approx G(E=E_{F0})=\frac{q^2}{h}\frac{(\gamma/2)^2}{(E_{F0}-\varepsilon)^2+(\gamma/2)^2}. \tag{48}$$

Таким образом, квантовая модель одноуровнего резистора показывает, что измеряемая проводимость будет максимальной и равной кванту проводимости q^2/h, если электрохимический потенциал E_{F0} будет достаточно близок к энергии уровня ε. Экспериментально измеряемый квант проводимости равен $2q^2/h$, где двойка связана с вырождением по спину, поскольку в действительности все уровни энергии идут парами со спинами α и β, так что одноуровневая модель резистора фактически есть двухуровневой с учетом вырождения по спину.

10.6. Интерференция источников поступления электронов в резистор

В отличие от классической модели резистора, в которой одновременно учитывались оба источника поступления электронов с двух контактов в

проводник (17), в квантовой модели (34) учитывалась инжекция электронов только с истока ($f_1 = 1$), а сток поддерживался пустым ($f_2 = 0$).

Это не вопрос удобства рассуждений. Если вместо (34) в уравнении Шредингера одновременно учесть оба инжектирующих контакта

$$E\psi = \left(\varepsilon - i\frac{\gamma}{2}\right)\psi + s_1 + s_2, \tag{49}$$

для волновой функции получим

$$\psi = \frac{s_1 + s_2}{E - \varepsilon + i\frac{\gamma}{2}}, \tag{50}$$

так что в электронной плотности появятся два перекрестных интерференционных слагаемых:

$$\psi\psi^* = \frac{1}{(E-\varepsilon)^2 + (\gamma/2)^2}\left(s_1 s_1^* + s_2 s_2^* + s_1 s_2^* + s_2 s_1^*\right), \tag{51}$$

которые никогда не наблюдались экспериментально, поскольку при инжекции от двух отдельных контактов электроны поступают в проводник с нескоррелированными фазами, меняющимися во времени произвольным образом и в среднем дающими нулевой вклад. Первые же два попарных произведения дают положительные вклады и наблюдаются экспериментально.

Контакты в электронных устройствах обычно не когерентны и это обстоятельство нужно учитывать при их моделировании. В уравнении Шредингера нельзя одновременно учитывать несколько электрических контактов. Учитывать контакты надо по одному, вычислять свойства, зависящие от произведения волновых функций, таких как электронная плотность, ток и другие, а затем суммировать эти вклады от разных контактов.

10.7. Квантовый транспорт по многоуровневому проводнику

Обобщим одноуровневую модель (34) на многоуровневую (рис. 6) с матрицей гамильтониана $N \times N$, имеющей N собственных значений.

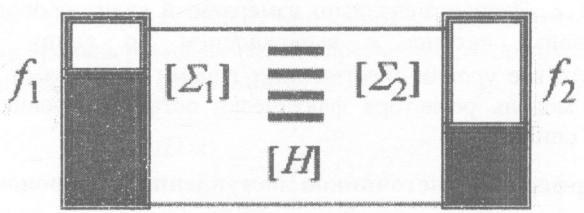


Рис. 6. Транспортная модель для многоуровневого проводника.

Уравнение Шредингера с двумя контактами, один из которых инжектирующий, имеет вид:

$$E\{\psi\}=[H+\Sigma_1+\Sigma_2]\{\psi\}+\{s_1\}, \tag{52}$$

где волновая функция и инжектирующий контакт есть столбцовые матрицы $N\times 1$, а контактные $N\times N$ матрицы Σ_1 и Σ_2 неэрмитовы с антиэрмитовыми компонентами

$$\begin{aligned}\Gamma_1 &= i\left[\Sigma_1-\Sigma_1^+\right],\\ \Gamma_2 &= i\left[\Sigma_2-\Sigma_2^+\right],\end{aligned} \tag{53}$$

выполняющими роль γ_1 и γ_2 в одноуровневой задаче.

Сначала покажем как четыре основных уравнения метода НРФГ (6), (13), (14) и (16) следуют из уравнения Шредингера с учтенными контактами (52).

Из (52) непосредственно имеем

$$\{\psi\}=\left[G^R\right]\{s_1\}, \tag{54}$$

где запаздывающая функция Грина G^R дается выражением (6) с

$$\Sigma=\Sigma_1+\Sigma_2. \tag{55}$$

Неравновесная функция Грина

$$G^n=2\pi\{\psi\}\{\psi\}^+=2\pi[G^R]\{s_1\}\{s_1\}^+[G^A], \tag{56}$$

где опережающая функция Грина G^A есть эрмитово сопряженная запаздывающая функция (7).

Для одноуровневой задачи $2\pi s_1 s_1^* = \gamma_1$ (39), а для многоуровневой – это матрицы

$$2\pi\{s_1\}\{s_1\}^+=[\Gamma_1], \tag{57}$$

так что

$$G^n=[G^R][\Gamma_1][G^A]. \tag{58}$$

Это неравновеская гриновская функция для одного инжектирующего источника. Для нескольких источников матрицы электронной плотности, вопреки волновым функциям, все складываются, взвешенные соответствующими фермиевскими функциями, для получения (13) с матрицей Σ^{in}, представляющей собой некогерентную сумму всех независимых источников, так что для нашего случая двух источников

$$[\Sigma^{in}]=[\Gamma_1]f_1(E)+[\Gamma_2]f_2(E). \tag{59}$$

Уравнение (13) с (59) дает матрицу электронной плотности G^n через фермиевские функции для двух контактов. Если обе фермиевские функции равны единице, то все состояния заняты электронами, так что матрица электронной плотности становится равной матрице плотности состояний, называемой в методе НРФГ матрицей спектральной функции $[A]$. Полагая в (13) с (59) $f_1 = 1$ и $f_2 = 1$, спектральная функция

$$[A]=[G^R][\Gamma][G^A] \tag{60}$$

с $\Gamma = \Gamma_1 + \Gamma_2$.

Получим еще два других выражения для спектральной функции, приведенных в (14). Из (54) с учетом (52) и (55) имеем

$$G^R = [EI - H - \Sigma]^{-1}, \tag{61}$$

так что для обратной матрицы запаздывающей функции Грина имеем

$$[G^R]^{-1} = EI - H - \Sigma. \tag{62}$$

Выполнив эрмитовое сопряжение равенства (62), получим

$$\left[\left[G^R\right]^{-1}\right]^{+} = \left[\left[G^R\right]^{+}\right]^{-1} = EI - H - \Sigma^{+} \tag{63}$$

или иначе, учитывая (7),

$$\left[G^A\right]^{-1} = EI - H - \Sigma^{+}. \tag{64}$$

Вычитая (62) из (64) и учитывая

$$\Gamma = i\left[\Sigma - \Sigma^{+}\right], \tag{65}$$

следующее из (53), получим

$$\left[G^R\right]^{-1} - \left[G^A\right]^{-1} = i[\Gamma]. \tag{66}$$

Умножая (66) слева на $[G^R]$, а справа на $[G^A]$, получим еще одно выражение для спектральной функции A (14):

$$i\left[\left[G^R\right] - \left[G^A\right]\right] = G^R \Gamma G^A. \tag{67}$$

Умножая (66) слева на $[G^A]$, а справа на $[G^R]$, получим третье выражение для спектральной функции A (14)

$$i\left[\left[G^R\right]-\left[G^A\right]\right]=G^A\Gamma G^R. \tag{68}$$

Осталось получить выражение для тока. Как в случае и классического и квантового рассмотрения одноуровневой модели, выражение для тока получим как изменение во времени числа электронов. Начинаем с временного уравнения Шредингера

$$i\hbar\frac{d}{dt}\{\psi\}=\left[H+\Sigma\right]\{\psi\}+\{s\} \tag{69}$$

и его эрмитово сопряженного

$$-i\hbar\frac{d}{dt}\{\psi\}^+=\{\psi\}^+\left[H+\Sigma^+\right]+\{s\}^+. \tag{70}$$

Имеем

$$\begin{aligned} i\hbar\frac{d}{dt}\{\psi\}\{\psi\}^+&=\left(i\hbar\frac{d}{dt}\{\psi\}\right)\{\psi\}^+ +\{\psi\}\left(i\hbar\frac{d}{dt}\{\psi\}^+\right)= \\ &=\left(\left[H+\Sigma\right]\{\psi\}+\{s\}\right)\{\psi\}^+ -\{\psi\}\left(\{\psi\}^+\left[H+\Sigma^+\right]+\{s\}^+\right)= \\ &=\left[\left(H+\Sigma\right)\psi\psi^+ -\psi\psi^+\left(H+\Sigma^+\right)\right]+\left[ss^+G^A-G^Rss^+\right], \end{aligned} \tag{71}$$

где использовались уже известное соотношение (54) и его эрмитово сопряженное

$$\{\psi\}=\left[G^R\right]\{s\} \quad \text{и} \quad \{\psi\}^+=\{s\}^+\left[G^A\right]. \tag{72}$$

Поскольку след матрицы $[\psi\psi^+]$ дает число электронов, определив производную этой матрицы по времени, найдем матрицу оператора тока, след которой даст ток. Учитывая (10) и (12), для матрицы оператора тока из (71) имеем

$$\left[\hat{I}\right]=\frac{\left[HG^n-G^nH\right]+\left[\Sigma G^n-G^n\Sigma^+\right]+\left[\Sigma^{in}G^A-G^R\Sigma^{in}\right]}{i2\pi\hbar}. \tag{73}$$

Учитывая, что след произведения матриц не зависит от порядка сомножителей, для скорости изменения числа электронов в канале имеем

$$\frac{dN}{dt}=\frac{-i}{h}Tr\left(\left[\Sigma G^n-G^n\Sigma^+\right]+\left[\Sigma^{in}G^A-G^R\Sigma^{in}\right]\right), \tag{74}$$

а с учетом еще (14) и (15) окончательно имеем

$$\frac{dN}{dt}=\frac{1}{h}Tr\left[\Sigma^{in}A-\Gamma G^n\right]. \tag{75}$$

Далее нужно учесть следующее. В уравнении (73) обе части равенства равны нулю, поскольку речь идет о токе в системе (рис. 6) в состоянии динамического равновесия. Обе части уравнения (73) разбиваются на слагаемые, относящиеся к контактам 1 и 2. Сумма их равна нулю в согласии с законом Кирхгофа для электрических цепей в состоянии динамического равновесия. Обобщая на произвольный номер контакта m, получаем уже упомянутое уравнение (16) для тока

$$\tilde{I}_m=\frac{q}{h}Tr\left[\Sigma_m^{in}A-\Gamma_m G^n\right]. \tag{76}$$

Классическая и квантовая модели сравниваются на рис. 7, где множителем D в классической модели учитывается многоуровневость классической задачи.

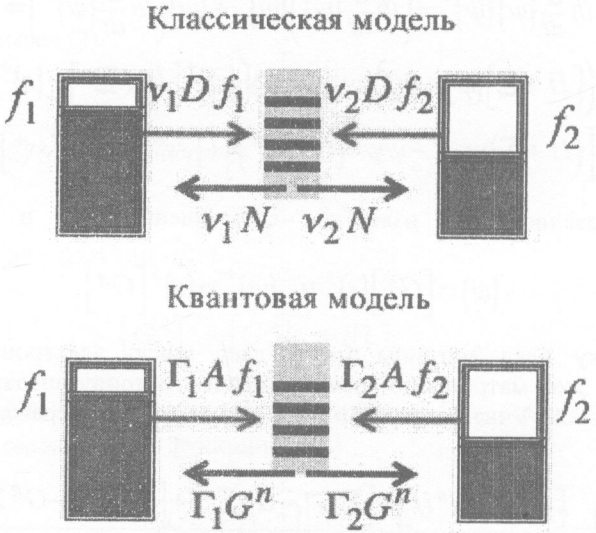


Рис. 7. Пополнение и опорожнение канала проводимости в классической и квантовой моделях транспорта электронов.

10.8. Функция проводимости для когерентного транспорта

Преобразуем уравнение (76) следующим образом. Учтем (13) и (14), а также

$$\Gamma=\sum_n\Gamma_n,\quad \Sigma^{in}=\sum_n\Sigma_n^{in},\quad \Sigma_n^{in}=\Gamma_n f_n\left(E\right). \tag{77}$$

Тогда имеем

$$\tilde{I}_m=\frac{q}{h}\sum_n\bar{T}_{mn}\left(f_m\left(E\right)-f_n\left(E\right)\right), \tag{78}$$

где коэффициент прохождения между контактами m и n

$$\bar{T}_{mn}\equiv Tr\left[\Gamma_m G^R\Gamma_n G^A\right]. \tag{79}$$

Пользуясь перестановочностью матриц под знаком следа, легко доказывается полезное свойство коэффициента прохождения, а именно:

$$\sum_n\bar{T}_{mn}=\sum_n\bar{T}_{nm}=Tr\left[\Gamma_m A\right]. \tag{80}$$

Перепишем выражение для тока (78) с использованием (79) применительно к двухтерминальному устройству

$$\tilde{I}\left(E\right)=\frac{q}{h}Tr\left[\Gamma_1 G^R\Gamma_2 G^A\right]\left(f_1(E)-f_2(E)\right) \tag{81}$$

и сравнивая это выражение для тока с аналогичным выражением для тока в упругом резисторе (201/гл.1)

$$I=\frac{1}{q}\int_{-\infty}^{+\infty}dEG\left(E\right)\left(f_1\left(E\right)-f_2\left(E\right)\right),$$

получаем квантовый аналог функции проводимости

$$G\left(E\right)=\frac{q^2}{h}Tr\left[\Gamma_1 G^R\Gamma_2 G^A\right]=\frac{q^2}{h}\bar{T}_{12}. \tag{82}$$

Для интерпретации экспериментальных данных, полученных на многотерминальных устройствах, Бюттекер [22] предложил элегантную формулу, связывающую ток I_m на контакте m с электрохимическими потенциалами на остальных контактах

$$I_m=\left(1/q\right)\sum_n G_{m,n}\left(\mu_m-\mu_n\right), \tag{83}$$

где $G_{m,n}$ есть проводимость, определяемая коэффициентом прохождения между контактами m и n.

В области линейного отклика воспользуемся нашим обычным разложением разности фермиевских функций в ряд Тейлора (глава 1) для получения соответствующей разности электрохимических потенциалов, тогда уравнение (78) окажется фактически уравнением Бюттекера (83) с проводимостью

$$G_{m,n}(E) \equiv \frac{q^2}{h} Tr\left[\Gamma_m G^R \Gamma_n G^A\right],$$

которую нужно еще усреднить для упругого резистора обычным образом

$$G_{m,n} = \int_{-\infty}^{+\infty} dE \left(-\frac{\partial f_0}{\partial E}\right) G_{m,n}(E).$$

До сих пор рассматривались только физические контакты $[\Sigma_{1,2}]$ в квантовой модели когерентного транспорта, в которой электроны движутся когерентно от истока к стоку по каналу, описываемому статическим гамильтонианом $[H]$ в отсутствии взаимодействия электрона с окружением $[\Sigma_0]$ при его движении по каналу (рис. 2а). Учет взаимодействия $[\Sigma_0]$ с формальной точки зрения проблемы не представляет. Все уравнения метода НРФГ остаются прежними, в матрицах же Σ, Γ и Σ^{in} появятся лишь дополнительные члены:

$$\begin{aligned} &\Sigma = \Sigma_1 + \Sigma_2 + \Sigma_0, \\ &\Gamma = \Gamma_1 + \Gamma_2 + \Gamma_0, \\ &\left[\Sigma^{in}\right] = \left[\Gamma_1\right] f_1(E) + \left[\Gamma_2\right] f_2(E) + \left[\Sigma_0^{in}\right]. \end{aligned} \qquad (84)$$

Однако, что из себя физически представляет взаимодействие Σ_0? С точки зрения электрона, движущегося в твердом теле, среда не представляется электрону статической, описываемой статическим гамильтонианом $[H]$, а весьма турбулентной средой со случайно меняющимся потенциалом U_R, который флуктуирует в пикосекундной шкале времени. Даже при достаточно низких температурах с замороженными фононными модами электрон движется в поле флуктуирующего потенциала, создаваемого всеми другими электронами (приближение самосогласованного поля). Даже в этом случае имеют место фазовые флуктуации (дефазировка), приводящие к флуктуации тока. Типичные измерения тока дают нам его среднее значение в интервале нескольких наносекунд, микросекунд или даже миллисекунд. Этот эффект усреднения нужно адекватно моделировать, если мы хотим правильно интерпретировать экспериментальные данные.

Метод НРФГ был первоначально разработан именно для учета неупругих процессов квантового транспорта в массивных проводниках. Мы же изложили его применительно к упругим резисторам. Вопросы дефазировки и обобщение метода НРФГ на неупругие процессы транспорта рассмотрим позже в этой главе. Сейчас же мы сформулируем и решим модельные задачи квантового транспорта, представляющие не только педагогический, но и научный интерес. Начнем мы с квантового транспорта в 1D проводниках не только в режиме баллистического транспорта, но и с учетом рассеивающих центров. Обосновано ли пренебрегать эффектами интерференции на дефектах и предполагать, что электроны диффундируют как классические частицы? Такой вопрос поставил Андерсон еще в 1958 году [23] и пришел к выводу, что диффузия может быть существенно подавлена или даже ею можно полностью пренебречь в результате квантовой интерференции между рассеивающими центрами. Мы не намерены углубляться в теорию андерсоновской локализации [24], покажем лишь как даже упрощенная модель НРФГ помогает глубже понять физику квантового транспорта.

Для любой транспортной задачи нужно написать гамильтониан $[H]$ и матрицы собственной энергии Σ. Как только это сделано, дальнейшие вычисления методом НРФГ выполняются стандартно.

10.9. Модельные транспортные задачи

Мы рассмотрим однородные 1D и 2D проводники, а также графен в качестве примера неоднородного проводника и сформулируем общий метод построения матриц собственной энергии.

10.9.1. Моделирование 1D проводника

Для наших целей достаточно одномерной модели неограниченного однородного проводника в приближении сильной связи с учетом взаимодействия лишь соседних атомов (рис. 8) в ортогональной базисе. Это приближение известно в квантовой химии с 1931 г. как метод молекулярных орбиталей Хюккеля [25]. Даже такая простая модель корректно описывает не все, но многие свойства протяженных полиенов $-(CH{=}CH{-})_nCH{=}$ [26 – 28], графена [29, 30], полиацетиленов и кумуленов $={(C{=})_n}C{=}$ [31 – 33], в последних, правда, каждый атом поставляет по два взаимно ортогональных π-электрона, что требует лишь незначительной модификации модели.

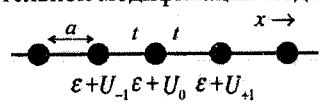


Рис. 8. К учету изменяющегося вдоль канала проводимости потенциала $U(x)$ в трансляционно-инвариантной цепочке атомов с периодом a, поставляющих каждый по одному электрону, и характеризуемой в приближении Хюккеля двумя параметрами – кулоновским интегралом ε и резонансным интегралом t.

Однородный 1D проводник описывается стандартной теорией одномерной модели кристаллов Кронига – Пенни и подчиняется параболическому дисперсионному соотношению с эффективной массой:

$$E = E_C + \frac{\hbar^2 k^2}{2m}. \tag{85}$$

Нам нужно переписать дисперсионное соотношение (85) применительно к используемому нами приближению сильной связи с учетом взаимодействия лишь соседних атомов (рис. 8) в ортогональной базисе. Другими словами, как выбрать параметры ε и t так, чтобы аппроксимировать дисперсионное соотношение (85).

Из уравнения Шредингера

$$E\psi_n = \sum_m H_{nm}\psi_m \tag{86}$$

энергия

$$E = \sum_m H_{nm}\frac{\psi_m}{\psi_n}. \tag{87}$$

Решения уравнения Шредингера (86) есть плоские волны (теорема Блоха)

$$\psi_n = \exp\left(ikna\right), \tag{88}$$

так что из (87) следует

$$E\left(k\right) = \sum_m H_{nm}\exp\left(ik\left(m-n\right)a\right). \tag{89}$$

Матрица $[H]$ имеет следующую структуру: по главной диагонали стоят кулоновские интегралы ε, на соседних верхней и нижней диагоналях стоят резонансные интегралы t, остальные элементы равны нулю,

$$H = \begin{bmatrix} \ddots & \cdots & & \iddots \\ & \varepsilon & t & 0 & \\ \vdots & t & \varepsilon & t & \vdots \\ & 0 & t & e & \\ \iddots & \cdots & & & \ddots \end{bmatrix},$$

так что для любой строки n матрицы $[H]$ согласно уравнению Шредингера (86) имеем

$$E\psi_n = t\psi_{n-1} + \varepsilon\psi_n + t\psi_{n+1} \tag{90}$$

или иначе

$$E = t\frac{\psi_{n-1}}{\psi_n} + \varepsilon + t\frac{\psi_{n+1}}{\psi_n}, \tag{91}$$

а с учетом (88) и после перегруппировки

$$E(k) = \varepsilon + t\exp(+ika) + t\exp(-ika) = \varepsilon + 2t\cos ka. \tag{92}$$

В области малых k

$$\cos ka \approx 1 - \frac{(ka)^2}{2}. \tag{93}$$

Сравнивая дисперсионное соотношение (92) для однородной цепочки в хюккелевском приближении в области малых k с правильным параболическим дисперсионным приближением (85), находим резонансный и кулоновский интегралы

$$E_c = \varepsilon + 2t, \tag{94}$$

$$-t \equiv t_0 \equiv \frac{\hbar^2}{2ma^2}. \tag{95}$$

Если при моделировании электронного устройства потребуется учесть изменяющийся вдоль канала проводимости потенциал $U(x)$, это делается путем добавления локального значения потенциала U к кулоновскому интегралу (рис. 8). Теперь, конечно, трансляционная инвариантность нарушается, дисперсионное соотношение будет другим, но новый гамильтониан вполне пригоден для численных расчетов и удовлетворительно описывает физику транспорта для потенциалов $U(x)$, не слишком быстро меняющихся в масштабе межатомных расстояний.

С найденными значениями интегралов ε (94) и t (95) легко выписывается матрица гамильтониана $[H]$. Теперь обсудим контактные матрицы собственных энергий. Основная идея заключается в том, чтобы бесконечно протяженный проводник, описываемый гамильтонианом $[H]$, заменить на проводник конечной длины, описываемый матрицей $[H + \Sigma_1 + \Sigma_2]$, с открытыми граничными условиями на концах, подразумевая под этим «хорошие» контакты, не создающие на своих концах отраженных потоков (рис. 9).

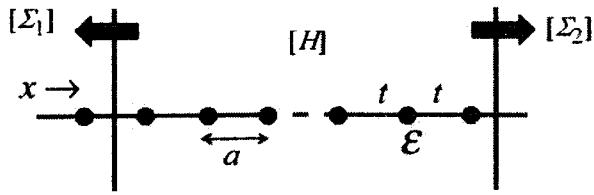


Рис. 9. Открытым граничным условиям соответствуют «хорошие» контакты, не создающие на своих концах отраженных потоков.

Проиллюстрируем эту идею на одномерной решетке. Пусть проводник ограниченной длины имеет n атомов, перенумерованных от 1 до n. Тогда левый контакт 1 начинается перед атомом цепи с номером 1, а правый контакт 2 – после атома цепи с номером n (рис. 10).

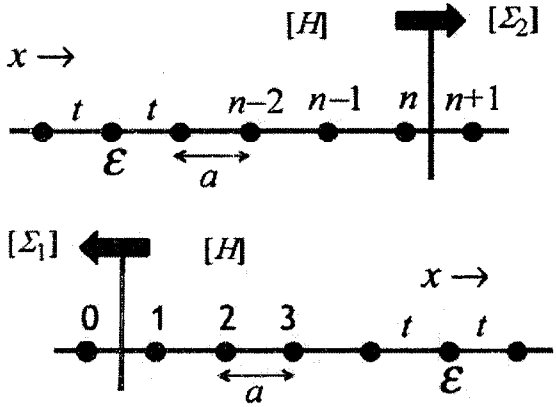


Рис. 10. К составлению контактных матриц 1D проводника из n атомов.

Контакты не имеют входящих потоков, только выходящие. В n-ой строке (90) уравнения Шредингера (86) слагаемое $t\psi_{n+1}$ уже принадлежит контакту 2, который, согласно уравнению (91), дает вклад в энергию, равный $t\psi_{n+1}/\psi_n$. Эта энергия и есть собственная энергия контакта 2. С учетом

$$\psi_{n+1} = \psi_n e^{ika} \tag{96}$$

вместо (90) имеем

$$E\psi_n = t\psi_{n-1} + (\varepsilon + te^{ika})\psi_n, \tag{97}$$

где добавка к кулоновскому интегралу в n-ой строке (97) есть собственная энергия контакта 2 и помещается она как элемент (n, n) соответствующей контактной матрицы

$$\Sigma_2 = \begin{bmatrix} \ddots & & \cdots & \\ & 0 & 0 & 0 \\ \cdots & 0 & 0 & 0 \\ & 0 & 0 & te^{ika} \end{bmatrix}. \tag{98}$$

Такую же собственную энергию имеет контакт 1, а в соответствующей контактной матрице помещается она как элемент (1, 1)

$$\Sigma_1 = \begin{bmatrix} te^{ika} & 0 & 0 & \cdots \\ 0 & 0 & 0 & \\ 0 & 0 & 0 & \\ \cdots & & & \ddots \end{bmatrix}. \qquad (99)$$

Остальные элементы матриц Σ_1 и Σ_2 нулевые.

10.9.1.1. Баллистический 1D проводник

Энергетические матрицы H, Σ_1 и Σ_2 составлены, вычисляем запаздывающую гриновскую функцию G^R (6), опережающую функцию G^A (7), матрицы Γ_1 и Γ_2 (53) и, наконец, коэффициент прохождения T_{12} и проводимость $G(E)$ (82). Вычисление проводимости баллистического 1D проводника является хорошей проверкой того, что контактные матрицы определены правильно. Проводимость должна быть равна кванту проводимости q^2/h, помноженному на число мод $M(E)$ 1D проводника, равное единице (спины не учитываются). Согласно (82) это означает, что в рассматриваемом примере коэффициент прохождения должен быть равен единице в области энергий $0 < E - E_C < 4t_0$, покрываемых дисперсионным соотношением

$$E = \varepsilon + 2t\cos ka = E_C + 2t_0\left(1 - \cos ka\right), \qquad (100)$$

и нулю за пределами этой области (рис. 11, $U = 0$).

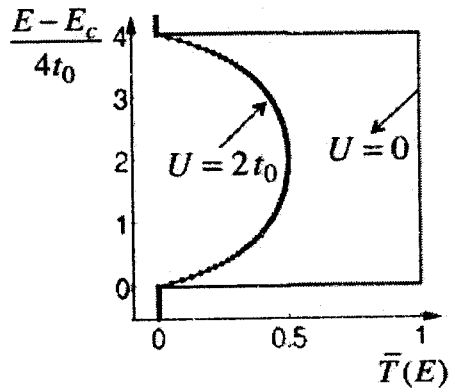


Рис. 11. Коэффициент прохождения в баллистическом 1D проводнике без рассеяния ($U = 0$) и с одним точечным рассеивающим центром ($U = 2t_0$).

10.9.1.2. Плотность состояний 1D проводника

Вычислим сначала $D(E)$ 1D проводника из элементарных соображений. Согласно (251/гл.1), число состояний 1D проводника длины L со значениями импульса меньшими, чем заданное значение p,

$$N(p)=\frac{2L}{h/p}. \tag{101}$$

Плотность состояний

$$D(E)=\frac{dN}{dE}=\frac{2L}{h}\frac{dp}{dE}=\frac{L}{\pi\hbar v}, \tag{102}$$

где учтено то обстоятельство, что для изотропного закона дисперсии $E(p)$ скорость $v = dE/dp$ [1].

Получим это же выражение для плотности состояний методом НРФГ. Для изотропного 1D проводника достаточно учесть только один атом в цепи (рис. 12).

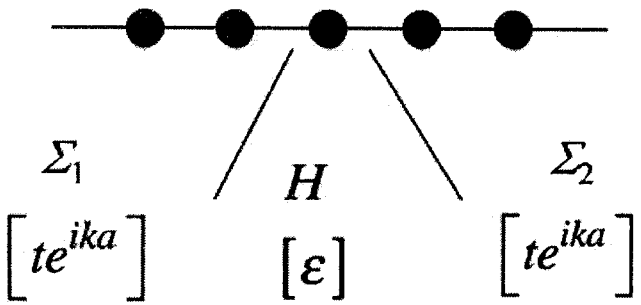


Рис. 12. К вычислению плотности состояний $D(E) = A/2\pi$ через спектральную функцию A (14).

В этом случае роль длины проводника L играет период решетки a. Для запаздывающей гриновской функции имеем

$$G^R=\left[E-\varepsilon-2te^{ika}\right]^{-1}. \tag{103}$$

Представляя экспоненту через синус и косинус и учитывая (100), получим

$$G^R=i/2t\sin ka. \tag{104}$$

Имеем также

$$\hbar v = \frac{dE}{dk} = -2at\sin ka, \tag{105}$$

где первое равенство следует из изотропности 1D проводника, а второе – из (100). С учетом (105) для G^R окончательно имеем

$$G_R = \frac{i}{2t\sin ka} = \frac{-i}{\hbar v / a}, \tag{106}$$

а опережающая функция Грина

$$G^A = \frac{ia}{\hbar v}. \tag{107}$$

Спектральная функция

$$A = i\left[G^R - G^A\right] = \frac{2a}{\hbar v}, \tag{108}$$

а плотность состояний

$$D(E) = \frac{A}{2\pi} = \frac{a}{\pi\hbar v} \tag{109}$$

совпадает с выражением (102), полученным ранее из элементарных соображений.

10.9.1.3. 1D проводник с одним рассеивающим центром

В гамильтониане рассеивающий центр моделируем путем добавления к кулоновскому интегралу одного из атомов цепочки постоянного потенциала U:

$$H = \begin{bmatrix} \ddots & \cdots & & \\ & \varepsilon & t & 0 \\ \vdots & t & \varepsilon+U & t \\ & 0 & t & \varepsilon \\ & & \cdots & \ddots \end{bmatrix}. \tag{110}$$

При такой постановке задачи коэффициент прохождения может быть вычислен аналитически (рис. 13).

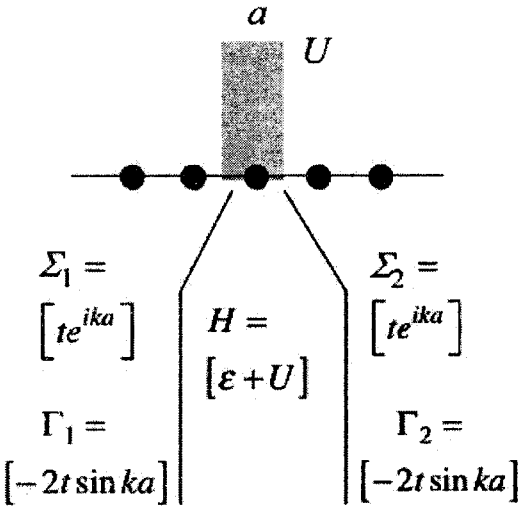


Рис. 13. К расчету коэффициента прохождения в 1D проводнике с одним рассеивающим центром методом НРФГ.

Контактные матрицы 1×1 заданы, по ним вычислили Γ_1 и Γ_2 по (53), гриновская функция с учетом (100)

$$G^R(E) = \frac{1}{E - (\varepsilon + U) - 2te^{ika}} = \frac{1}{-U - i2t\sin ka}, \tag{111}$$

от нее комплексным сопряжением берем опережающую функцию G^A и сразу получаем коэффициент прохождения

$$\Gamma_1 G^R \Gamma_2 G^A = \frac{(2t\sin ka)^2}{U^2 + (2t\sin ka)^2} \tag{112}$$

или с учетом (105) окончательно

$$\bar{T}(E) = \frac{(2t\sin ka)^2}{U^2 + (2t\sin ka)^2} = \frac{(\hbar v/a)^2}{U^2 + (\hbar v/a)^2}. \tag{113}$$

Результаты вычислений в отсутствии рассеивающего центра ($U = 0$) и с его учетом $U = 2t_0$ показаны на рис. 11.

Выражение для коэффициента прохождения (113) полезно вывести также из элементарных соображений (рис. 14).

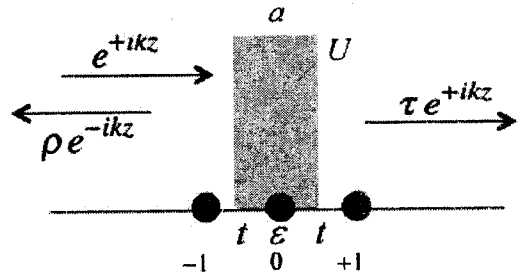


Рис. 14. К вычислению коэффициента прохождения из условия непрерывности волновой функции.

Удобства ради, перенумеруем атомы проводника так, чтобы дефект находился на атоме с номером 0. Падающая на рассеивающий центр волна $\exp(+ikz)$ отражается $\rho\cdot\exp(-ikz)$ с коэффициентом отражения ρ и проходит далее $\tau\cdot\exp(+ikz)$ с коэффициентом прохождения τ. Из уравнения Шредингера в хюккелевском приближении имеем

$$E\psi_0 = (\varepsilon + U)\psi_0 + t\psi_{-1} + t\psi_{+1}. \qquad (114)$$

Из закона сохранения и условия непрерывности волновой функции на дефекте ($z = 0$) имеем

$$1+\rho = \tau = \psi_0. \qquad (115)$$

Подстановка в (114) волновых функций на соседних с дефектом атомах и с учетом (115) дает

$$(E-\varepsilon-U)\tau = t\left(e^{-ika} + (\tau-1)e^{ika}\right) + t\tau e^{ika}, \qquad (116)$$

а после простых преобразований и перегруппировок с учетом (100) и (105) для коэффициента прохождения τ получаем

$$\tau = \frac{i\hbar v/a}{-U + i\hbar v/a}, \qquad (117)$$

что после умножения на его комплексно сопряженное получаем уже выведенное ранее методом НРФГ выражение для коэффициента прохождения (113).

10.9.2. Моделирование однородного 2D проводника

Среди фундаментальных экспериментов 80-х годов, приведших к рождению мезофизики, были наблюдения проводимости баллистических 2D проводников, оказавшейся целочисленно пропорциональной кванту проводимости $2q^2/h$. Для понимания физики и интерпретации таких экспериментов нужна как можно более простая модель 2D проводника.

Для наших целей, по аналогии с 1D проводниками, достаточно двумерной модели неограниченного однородного проводника в хюккелевской модели (рис. 15).

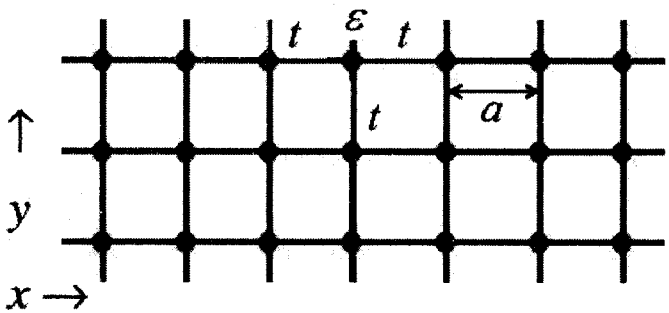


Рис. 15. Трансляционно-инвариантная решетка атомов с периодом a, поставляющих каждый по одному электрону, и характеризуемая в приближении Хюккеля двумя параметрами – кулоновским интегралом ε и резонансным интегралом t.

Параметры модели ε и t выберем таким образом, чтобы выполнялось стандартное дисперсионное соотношение с эффективной массой

$$E(k_x,k_y)=E_C+\frac{\hbar^2(k_x^2+k_y^2)}{2m}. \qquad (118)$$

Решение уравнения Шредингера (86) выберем в виде

$$\psi_n=\exp(i\vec{k}\cdot\vec{r}), \qquad (119)$$

где волновой вектор определяется его проекциями k_x и k_y в выбранной системе координат, а радиус-вектор указывает положение n-го атома в решетке. Подставляя это решение в уравнение (87), получаем дисперсионное соотношение

$$E(\vec{k})=\sum_m H_{nm}\exp\left(i\vec{k}\cdot(\vec{r}_m-\vec{r}_n)\right), \qquad (120)$$

которое для хюккелевской модели решетки дает

$$E(\vec{k}) = \varepsilon + t\exp(+ik_x a) + t\exp(-ik_x a) + t\exp(+ik_y a) + t\exp(-ik_y a) = \\ = \varepsilon + 2t\cos(k_x a) + 2t\cos(k_y a). \quad (121)$$

Пользуясь теми же соображениями, что и для одномерной цепочки, для резонансного и кулоновского интегралов как параметров модели получаем

$$t = -\hbar^2 / 2ma^2, \quad (122)$$

$$\varepsilon = E_C - 4t \quad (123)$$

или несколько иначе

$$E_C = \varepsilon + 4t, \quad (124)$$

$$-t \equiv t_0 \equiv \frac{\hbar^2}{2ma^2}. \quad (125)$$

Составление хюккелевского гамильтониана H (рис. 16) не вызывает затруднений, обсудить лишь надо построение контактных матриц для 2D проводника.

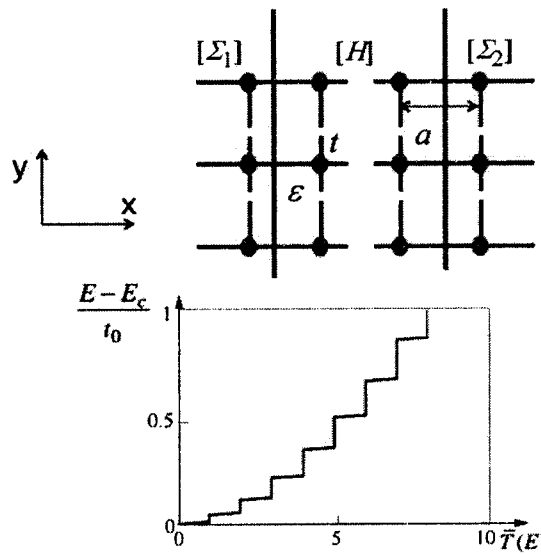


Рис. 16. Вычленение из двумерной решетки области собственно проводника и контактов для метода НРФГ и результаты расчета коэффициента прохождения методом НРФГ при числе атомов по ширине проводника, равном 25.

Однако, сначала попробуем понять причины появления ступенек в зависимости коэффициента прохождения от энергии. Простое объяснение этого экспериментально наблюдаемого факта может быть следующим. Для баллистического проводника коэффициент прохождения есть просто число мод $M(E)$, равное половине числа де-бройлевских длин волн, укладывающихся в поперечном сечении проводника W (275/гл.1),

$$M = Int\left(\frac{2W}{h/p}\right) = Int\left(\frac{2W}{h}\sqrt{2mE'}\right), \qquad (126)$$

где под значком $Int(x)$ подразумевается наибольшее целое число, меньшее значения x, а правая часть (126) выписана для параболической дисперсии $E' = E - E_C = p^2/2m$. Экспериментально при низких температурах [34, 35] измеряется число мод $M(E = E_{F0})$, которое будучи по своей физической природе целочисленным свойством показывает скачки при изменении энергии $E = E_{F0}$ и поперечного сечения 2D проводника.

Правильное поведение расчетного коэффициента прохождения с ростом энергии (рис. 16), о чем свидетельствует не только само наличие ступенек, но и их равноширинность, подтверждаемая экспериментальными данными [34, 35], свидетельствует о корректном задании энергетических матриц в расчете методом НРФГ. Покажем как появляются ступеньки (рис. 16) и построим гамильтониан и контактные матрицы для 2D решетки.

Пусть проводник имеет p атомов по ширине проводника и q атомов по длине проводника, т.е. матрица проводника имеет форму $(p \times q)$. Такой 2D проводник условно можно представить себе как p соединенных параллельно 1D проводников, каждый длиной q. Матрица $(p \times q)$ есть строчная матрица длины q, элементами которой являются столбцовые матрицы длины p. На рис. 17 показан проводник с одним столбцом формы $(p \times 1)$.

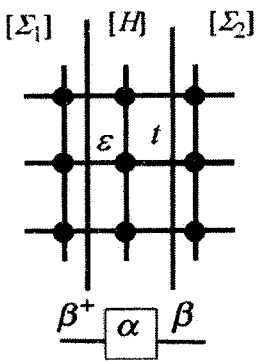


Рис. 17. Проводник формы $(p \times 1)$.

Каждый из q столбцов описывается своей хюккелевской матрицей α порядка p. Например, при $p = 3$ она имеет вид

$$\alpha = \begin{bmatrix} \varepsilon & t & 0 \\ t & \varepsilon & t \\ 0 & t & \varepsilon \end{bmatrix}. \tag{127}$$

Столбцы связаны друг с другом резонансными интегралами t следующим образом. Рассмотрим связь между столбцами с соседними номерами n и $n + 1$. Эта связь описывается скалярной матрицей $\beta = t \cdot I$ порядка p, где I – единичная матрица. В нашей модели, естественно, $\beta = \beta^+$.

Например, при $p = 3$

$$\beta = \begin{bmatrix} t & 0 & 0 \\ 0 & t & 0 \\ 0 & 0 & t \end{bmatrix}. \tag{128}$$

Гамильтониан H имеет блочную структуру. На его главной диагонали стоят одинаковые матрицы α порядка p, а ближайшие к ней диагонали сверху и снизу заполнены матрицами β тоже порядка p, остальные элементы нулевые. Если длина проводника составляет, скажем, $q = 10$ атомам, а по ширине проводника располагается $p = 5$ атомов, то порядок матрицы H равен $p \times q = 50$.

Решение задачи на собственные значения гамильтониана H сводится к диагонализации матрицы $[\alpha]$

$$[\tilde{\alpha}] = [V]^+ [\alpha] [V], \tag{129}$$

где столбцы матрицы $[V]$ есть собственные векторы матрицы $[\alpha]$, так что

$$\tilde{\alpha} = \begin{bmatrix} \varepsilon_1 & 0 & 0 \\ 0 & \varepsilon_2 & 0 \\ 0 & 0 & \varepsilon_3 \end{bmatrix}. \tag{130}$$

Всегда возможно обратное преобразование из собственного, иначе модового базиса в исходный решеточный базис:

$$[\alpha] = [V] [\tilde{\alpha}] [V]^+. \tag{131}$$

Матрица β не затрагивается преобразованием базиса, поскольку она уже диагональная. Диагонализация матрицы α формы $(p \times p)$ приводит к занулению

резонансных интегралов t, связывающих строки исходной матрицы гамильтониана, т. е. к превращению 2D проводника в p параллельно соединенных с контактами 1D проводников, каждый длиной q атомов (рис. 18) с энергиями $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_p$, равными собственным значениям матрицы α:

$$\varepsilon_n = \varepsilon - 2t_0 \cos k_n a, \tag{132}$$

с

$$k_n a = \frac{n\pi}{p+1}. \tag{133}$$

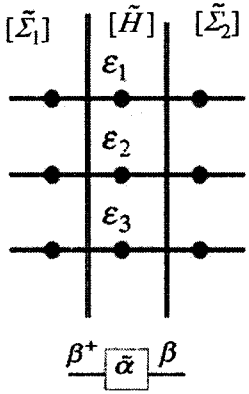


Рис. 18. 2D проводник в хюккелевском приближении после диагонализации гамильтониана H.

Для каждого из p параллельных 1D проводников коэффициент прохождения равен единице в области энергий ($t_0 \equiv |t|$) $\varepsilon_n - 2t_0 < E < \varepsilon_n + 2t_0$, как показано на рис. 19.

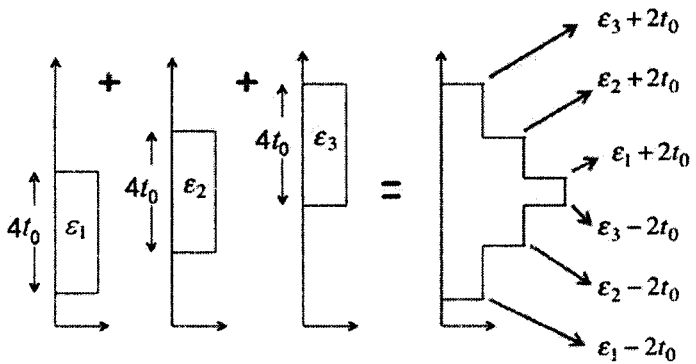


Рис. 19. К образованию ступенек (рис. 16) в зависимости коэффициента прохождения от энергии.

Складывая коэффициенты прохождения для всех p мод проводника, получаем зависимость коэффициента прохождения от энергии в виде поднимающихся ступенек в нижней части графика (рис. 16) и опускающихся ступенек в верхней части графика. При моделировании проводников n-типа обычно вычисляют нижнюю часть зоны (рис. 16) и мы видим лишь идущие вверх ступеньки с энергиями $\varepsilon_n - 2t_0$. Используя (132), (133) и (124), находим положение ступенек

$$\varepsilon_n - 2t_0 = E_c + 2t_0\left(1 - \cos\frac{n\pi}{p+1}\right). \tag{134}$$

На рис. 20 результаты расчета коэффициента прохождения методом НРФГ при числе атомов по ширине проводника p = 25 (рис. 16) показаны вместе с огибающей ступенек, вычисленной по (134) при той же ширине проводника $p = 25$.

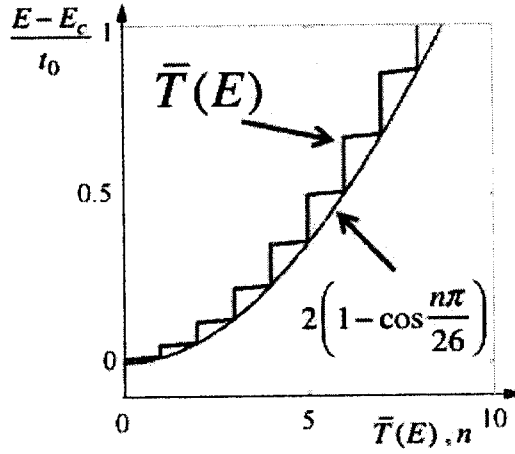


Рис. 20. Сопоставление численных результатов расчета коэффициента прохождения методом НРФГ с аналитической аппроксимацией по (134) при $p = 25$.

Представление 2D/3D проводника в виде параллельных 1D проводников представляется нам не только физически корректным, но и исключительно полезным подходом при интерпретации экспериментальных данных. Каждый их этих 1D проводников называют модой или подзоной с дисперсионным соотношением

$$E_n(k_x) = \varepsilon_n - 2t_0 \cos k_x a, \tag{135}$$

как показано на рис. 21.

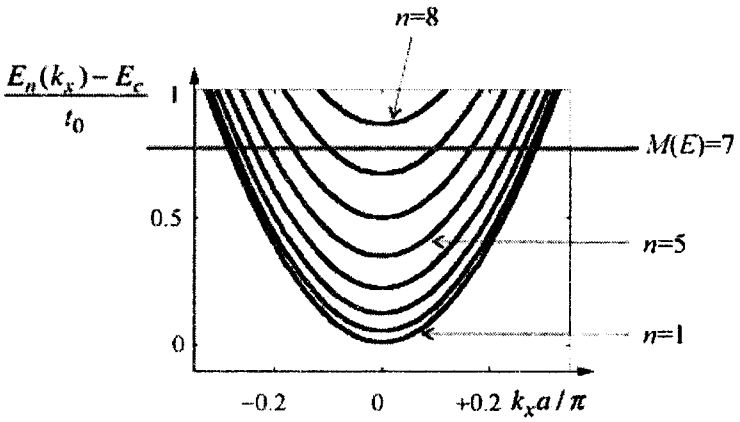


Рис. 21. Нижние восемь подзон хюккелевской модели 2D проводника.

Дисперсионные соотношения для подзон получаются из общего выражения (121) при требовании, чтобы k_y принимал квантованные значения

$$k_y a = \frac{n\pi}{p+1}, \tag{136}$$

где каждое значение n порождает одну соответствующую подзону (рис. 21).

Горизонтальная линия, проведенная при определенной энергии E, пересекает число подзон, равное удвоенному значению мод при этой энергии, поскольку каждая мода порождает два пересечения, одно для состояния с положительной скоростью, а другое – с отрицательной.

10.9.2.1. Контактные собственные энергии для 2D проводника

В идеальном случае контакты должны быть такими, чтобы электроны беспрепятственно покидали проводник и не возвращались в результате отражения от границ контактов. Мы рассмотрим простейшую модель таких контактов, которая сводится к тому, что контакт является естественным продолжением самого проводника.

2D проводник ширины p эквивалентен p параллельным 1D проводникам в результате преобразования базиса

$$\left[\tilde{X}\right] = \left[V\right]^{+}\left[X\right]\left[V\right], \tag{137}$$

где матрица X в решеточном базисе преобразуется в матрицу $\tilde{X}$ в модовом базисе или обратно

$$[X]=[V][\tilde{X}][V]^{+}, \tag{138}$$

как это было продемонстрировано выше для гамильтониана H. В результате для каждого из p независимых 1D проводников легко выписывается своя контактная матрица, а затем они все собираются в полную матрицу Σ.

В хюккелевской модели 2D проводника каждый из p 1D проводников характеризуется собственной энергией контакта $t\cdot\exp(ika)$ с соответствующим значением ka для конкретного 1D проводника при заданной энергии E. Для моды с номером n

$$E=\varepsilon_n-2t_0\cos k_n a, \tag{139}$$

так что контактная матрица в модовом базисе

$$[\tilde{\Sigma}_1]=\begin{bmatrix} te^{ik_1a} & 0 & 0 & \cdots \\ 0 & te^{ik_2a} & 0 & \\ 0 & 0 & te^{ik_3a} & \\ \cdots & & & \ddots \end{bmatrix}, \tag{140}$$

а после преобразования в решеточный базис (рис. 22)

$$[\Sigma_1]=[V][\tilde{\Sigma}_1][V]^{+}. \tag{141}$$

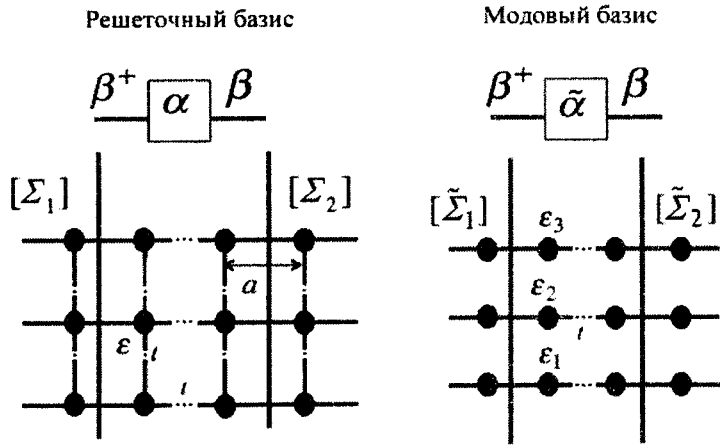


Рис. 22. К построению контактных матриц в модовом базисе с обратным преобразованием в решеточный базис.

Изложенный выше метод построения контактных матриц подходит для однородных решеток, но не пригоден в общем случае, поскольку далеко не всегда удается одним и тем же преобразованием базиса диагонализовать одновременно и матрицу α и матрицу β. Примером подобного рода может служить решетка графена. А в рассмотренном нами случае матрица β вообще скалярная, т.е. уже диагонализована.

10.9.3. Графен как пример неоднородного 2D проводника

В рассмотренной однородной решетке (рис. 15) каждый атом находится в одном и том же окружении. В правильной гексагональной решетке графена (рис. 23) есть два сорта атомов.

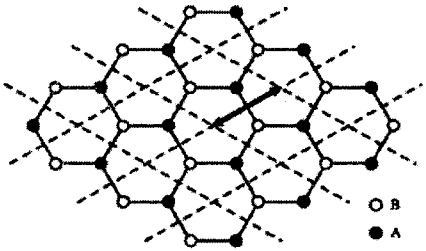


Рис. 23. Элементарная ячейка графена и вектор трансляции решеток Бравэ А и В.

Атомы сортов А и В находятся в разном окружении. У атома сорта А слева один атом, а справа два, а у атома сорта В наоборот. Трансляционную инвариантность графеновой решетки обеспечивают два атома, образующие элементарную ячейку в виде правильного ромба. Учитывая наличие двух неэквивалентных атомов в элементарной ячейке, уравнение Шредингера (86) перепишем в виде

$$E\{\psi\}_n = \sum_m [H]_{nm} \{\psi\}_m, \tag{142}$$

где $\{\psi\}_n$ есть вектор-столбец формы (2×1), компоненты которого соответствуют двум атомам сорта А и В, образующим элементарную ячейку с номером n. Гамильтониан формы (2×2) устанавливает связь между атомами сорта А и В ячейки с номером n и атомами сорта А и В ячейки с номером m.

Решение запишем в виде

$$\{\psi\}_n = \{\psi\}_0 \exp(i\vec{k}\cdot\vec{r}_n), \tag{143}$$

после подстановки которого в (142), имеем

$$E\{\psi\}_0 = \left[h(\vec{k})\right]\{\psi\}_0, \tag{144}$$

где гамильтониан с учетом фазового множителя

$$\left[h(\vec{k})\right]=\sum_{m}\left[H\right]_{nm}\exp\left(i\vec{k}\cdot(\vec{r}_m-\vec{r}_n)\right) \tag{145}$$

также имеет форму (2×2). Вычислим его, воспользовавшись вырезкой из графена, содержащей ячейку в окружении ее четырех ближайших ячеек (рис. 24).

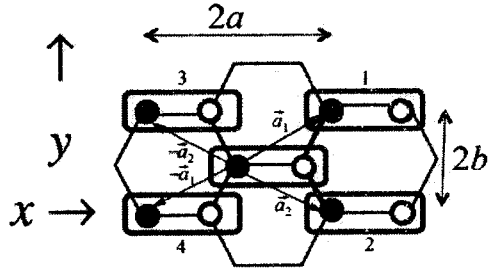


Рис. 24. Элементарная ячейка графена с номером 0 в окружении ближайших четырех ячеек с номерами 1, 2, 3, 4.

Потребуются также два вектора

$$\begin{aligned}\vec{a}_1&=a\hat{x}+b\hat{y}\\ \vec{a}_2&=a\hat{x}-b\hat{y}\end{aligned}\,, \tag{146}$$

связанные с геометрическими параметрами графеновой сетки

$$a=3a_0 \text{ и } b=\sqrt{3}a_0/2, \tag{147}$$

где a_0 – длина связи CC в графене, обычно принимаемая равной 1.42 $\mathring{A}$.

Вклады в гамильтониан (145) от пяти ячеек с учетом векторов (146)

$$\begin{matrix} e^{-i\vec{k}\cdot\vec{a}_2}\begin{bmatrix}0 & t\\ 0 & 0\end{bmatrix} & & \begin{bmatrix}0 & 0\\ t & 0\end{bmatrix}e^{+i\vec{k}\cdot\vec{a}_1}\\ + & & +\\ & \begin{bmatrix}\varepsilon & t\\ t & \varepsilon\end{bmatrix} & \\ + & & +\\ e^{-i\vec{k}\cdot\vec{a}_1}\begin{bmatrix}0 & t\\ 0 & 0\end{bmatrix} & & \begin{bmatrix}0 & 0\\ t & 0\end{bmatrix}e^{+i\vec{k}\cdot\vec{a}_2}\end{matrix} \tag{148}$$

нужно суммировать вместе с соответствующими фазовыми множителями. У центральной ячейки фазовый множитель отсутствует, поскольку это диагональный вклад в гамильтониан ($n = m = 0$). Взаимодействие ячеек с $n = 0$ и $m = 1$ в хюккелевской модели сводится лишь к одному резонансному интегралу t между атомом В центральной ячейки и атомом А ячейки $m = 1$. Аналогично для взаимодействия центральной ячейки с ячейкой $m = 2$. Отличаются они только фазовыми множителями: сдвиг первой ячейки по отношению к центральной определяется вектором $\vec{a}_1$, а второй ячейки – вектором $\vec{a}_2$ (рис. 24). Аналогично получаются вклады от ячеек 3 и 4.

Суммирование пяти матриц (148) дает для гамильтониана (145)

$$\left[h(\vec{k})\right]=\begin{bmatrix}\varepsilon & h_0^* \\ h_0 & \varepsilon\end{bmatrix}, \tag{149}$$

где

$$h_0 \equiv t + te^{+i\vec{k}\cdot\vec{a}_1} + te^{+i\vec{k}\cdot\vec{a}_2} \tag{150}$$

или с учетом (146)

$$h_0 \equiv t\left(1+2\cos(k_y b)\exp(+ik_x a)\right). \tag{151}$$

Диагонализация (144) с учетом (149) дает дисперсионное соотношение

$$E(\vec{k}) = \varepsilon \pm \left|h_0(\vec{k})\right|, \tag{152}$$

привести которое к привычному для графена виду

$$E = \pm v_0 p\ , \tag{153}$$

где v_0 – фермиевская скорость, составляющая примерно 1/300 от скорости света, можно путем линеаризации h_0 вблизи дираковских точек, в которых

$$h_0(\vec{k}) = 0, \tag{154}$$

так что

$$E(\vec{k}) = \varepsilon\ . \tag{155}$$

В этих точках равновесный электрохимический потенциал расположен при ε для нейтральных систем, у которых ровно половина уровней энергии заполнена (валентная зона), а другая половина – пустая (зона проводимости). Дираковских точек – три пары (рис. 25).

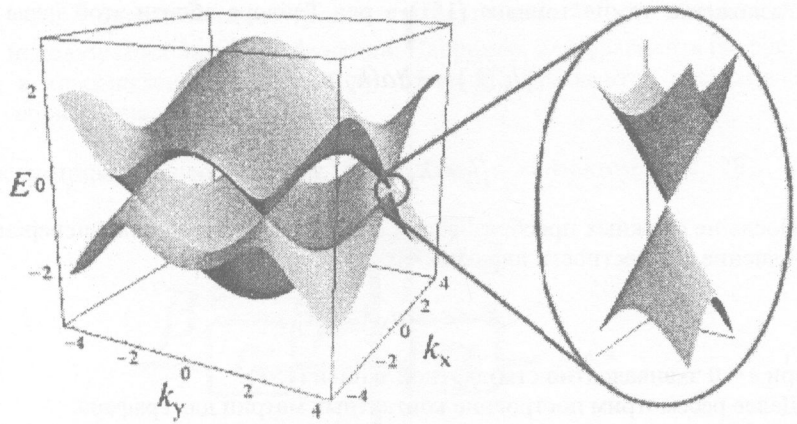


Рис. 25. Зонная структура графена. Увеличенным масштабом показан линейный спектр электронов в окрестности точки Дирака $k_x a = 0$ и $k_y b = +2\pi/3$ (рис. 26).

Выберем пару, соответствующую $k_x = 0$. Тогда гамильтониан (151) обращается в ноль при $k_y b = \pm 2\pi/3$.

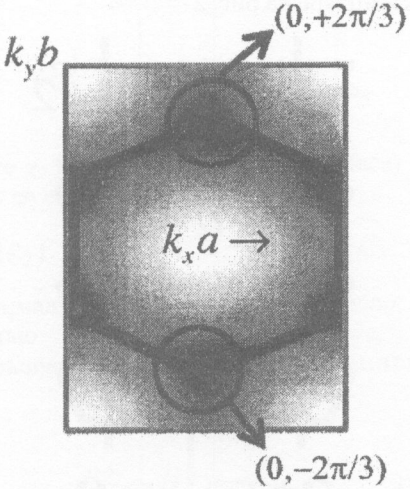


Рис. 26. Сечение зонной структуры графена по дираковским точкам, область вблизи которых зачернена.

Разложение гамильтониана (151) в ряд Тейлора вблизи этой пары точек дает

$$h_0(\vec{k}) \approx \pm ita(k_x \mp i\beta_y), \tag{156}$$

где

$$\beta_y \equiv k_y \mp 2\pi/3b. \tag{157}$$

После не сложных преобразований окончательно получаем дисперсионное соотношение в окрестности дираковских точек

$$E = \varepsilon \pm at\sqrt{k_x^2 + k_y^2}, \tag{158}$$

что при $\varepsilon = 0$ эквивалентно стандартной записи (153).

Далее рассмотрим построение контактных матриц для графена.

10.9.4. Общий подход к построению контактных матриц

Любой 2D проводник с одинаковым по всей длине проводника сечением по его ширине может быть разбит на фрагменты, повторяющие друг друга по всей длине проводника. Например, в случае графена таким фрагментом может быть вырезка, повторяющаяся по всей длине графенового проводника и показанная в одном экземпляре на рис. 27.

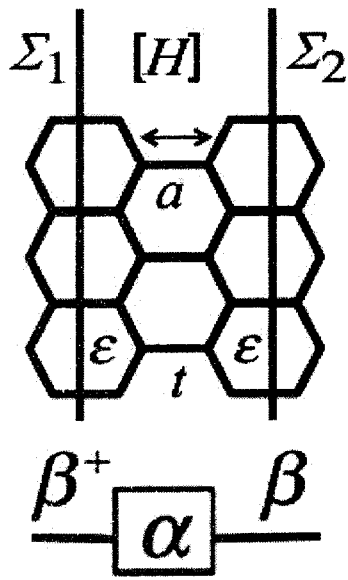


Рис. 27. К построению контактных матриц для графена.

Фрагмент описывается матрицей гамильтониана $[H] \equiv \alpha$ порядка n по числу используемых базисных функций. Например, для фрагмента на рис. 27 n = 12 в хюккелевском приближении. Матрицы фрагментов α связываются между собой матрицами β.

Рассмотрим правую границу проводника с контактом (рис. 28).

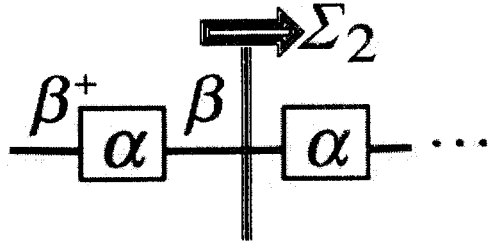


Рис. 28. Правая граница проводника с контактом.

Контактная блочная матрица везде нулевая кроме последнего диагонального элемента с номером n

$$\Sigma_2(E) = \begin{matrix} & & n-2 & n-1 & n \\ & & \cdots & & \\ \end{matrix} \begin{bmatrix} \cdots & 0 & 0 \\ \cdots & 0 & \beta g_2 \beta^+ \end{bmatrix}. \qquad (159)$$

В этом ненулевом блоке g_2 называют граничной (surface) функцией Грина для контакта 2 и вычисляют ее итерационно из уравнения

$$[g_2]^{-1} = (E + iO^+)I - \alpha - \beta^+ g_2 \beta, \qquad (160)$$

где O^+ – инфинитезимальное положительное число, выбором которого управляют сходимостью итерационного процесса, а слагаемое iO^+I обеспечивает отрицательность мнимых частей и неэрмитовость матриц g и Σ.

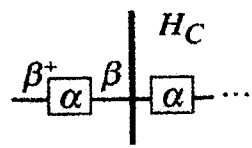


Рис. 29. Правая граница проводника с контактом в общем случае.

Рассмотрим вывод уравнения (160) в более общем случае (рис. 29), когда последний фрагмент проводника связан с неограниченным контактом, описываемым гамильтонианом H_c

$$\begin{bmatrix} \alpha & B \\ B^+ & H_c \end{bmatrix}, \tag{161}$$

где

$$[B] \equiv [\beta \quad 0 \quad 0 \quad \cdots]. \tag{162}$$

Результирующая запаздывающая функция Грина вычисляется из равенства

$$\begin{bmatrix} A & -B \\ -B^+ & A_c \end{bmatrix}^{-1} = \begin{bmatrix} G^R & \cdots \\ \cdots & \cdots \end{bmatrix}, \tag{163}$$

где

$$A \equiv (E + i0^+)I - \alpha, \tag{164}$$

$$A_c \equiv (E + i0^+)I_c - H_c. \tag{165}$$

Для вычисления G^R из (163) обратимся к очевидному равенству для блочных матриц

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} \times \begin{bmatrix} p & q \\ r & s \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}, \tag{166}$$

перемножая которые после несложных преобразований получим

$$p = (A - BD^{-1}C)^{-1}. \tag{167}$$

Воспользовавшись матричным равенством (167), для G^R из (163) имеем

$$G^R = \left[A - BA_c^{-1}B^+\right]^{-1}, \tag{168}$$

так что контактная матрица

$$\Sigma = BA_c^{-1}B^+. \tag{169}$$

Поскольку матрица B (162) имеет единственный ненулевой элемент β, имеем

$$\Sigma = \beta g \beta^+, \tag{170}$$

где g представляет собой верхний блок матрицы $[A_c]^{-1}$:

$$\begin{bmatrix} A & -\beta & 0 & 0 & \cdots \\ -\beta^{+} & A & -\beta & 0 & \cdots \\ 0 & -\beta^{+} & A & -\beta & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}^{-1} = \begin{bmatrix} g & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix}. \tag{171}$$

Искомое уравнение (160) получается, если уравнение (168) применить к матрице $N \times N$ (171), рассматривая первый блок A как проводник, а остальное $(N-1)\times(N-1)$ как контакт. Тогда

$$g_N = \left[A - \beta g_{N-1} \beta^{+}\right]^{-1}, \tag{172}$$

где g_N соответствует g в правой части уравнения (171), если матрица слева имеет размеры $N \times N$. Уравнение (172) решается итеративно, начиная с известного g_1, далее g_2 и так далее, пока g_N не совпадет с g_{N-1} с наперед заданной точностью, давая решение уравнения (160)

$$g = \left[A - \beta g \beta^{+}\right]^{-1}. \tag{173}$$

Полезно получить уже известное нам значение Σ для однородного 1D проводника, воспользовавшись общим методом построения контактных матриц Σ по уравнению (170). Сначала найдем g по уравнению (173), в котором $A = EI - \alpha$. Для однородного проводника $\alpha = \varepsilon$, $\beta = t$, $I = 1$, так что

$$g^{-1} = E - \varepsilon - t^2 g$$

или иначе

$$t^2 g^2 - (E - \varepsilon) g + 1 = 0,$$

решением которого является

$$g = \frac{(E-\varepsilon) \pm \sqrt{(E-\varepsilon)^2 - 4t^2}}{2t^2} = \frac{1}{t}\left(\cos ka \pm \sqrt{\cos^2 ka - 1}\right) = \frac{e^{\pm ika}}{t}, \tag{174}$$

где мы воспользовались дисперсионным соотношением для однородного 1D проводника (92). Подставляя (174) в (170), получаем уже известное нам выражение для однородного проводника $t \cdot \exp(ika)$, где в показателе экспоненты знак «+» выбран по физическим соображениям.

10.9.5. Баллистическая проводимость графена

В качестве иллюстративного примера приведем результаты расчета коэффициента прохождения и плотности состояний хюккелевской модели графеновых лент для его двух конфигураций границ (рис. 30) – зигзагообразной (Zigzag Graphene Nano Ribon / ZGNR) и креслоподобной (Armchair Graphene Nano Ribon / AGNR) [36]. Расчеты проводились для лент шириной $W = 53$ *нм* с резонансным интегралом $t_0 = 2.7$ *эв*.

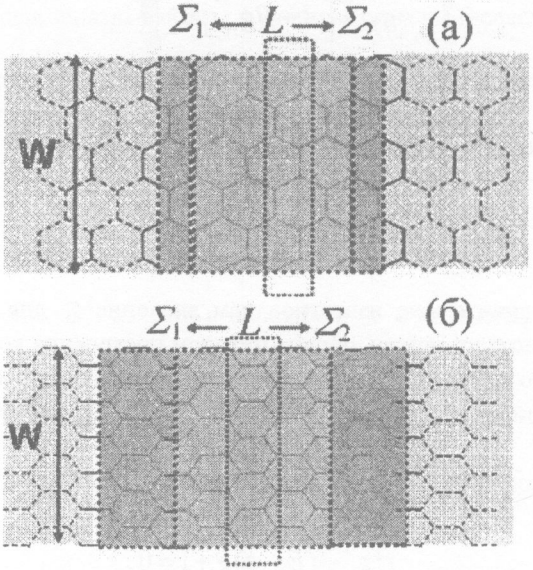


Рис. 30. Выбор повторяющихся фрагментов (показаны прямоугольниками, выступающими за пределы ширины лент W) двух конфигураций границ – зигзагообразной (а) и креслоподобной (б).

Поскольку рассматриваемые идеальные графеновые ленты являются баллистическими проводниками, то коэффициенты прохождения равны числу мод

$$M = Int\left(\frac{2W}{h/p}\right) = Int\left(\frac{2W}{h}\frac{E}{v_0}\right), \qquad (175)$$

где использовано дисперсионное соотношение для графена (153). Фермиевская скорость v_0 полагалась равной 10^6 *м/с*, обычно используемой в литературе [37].

Результаты расчета коэффициента прохождения и плотности состояний методом НРФГ представлены на рис. 31.

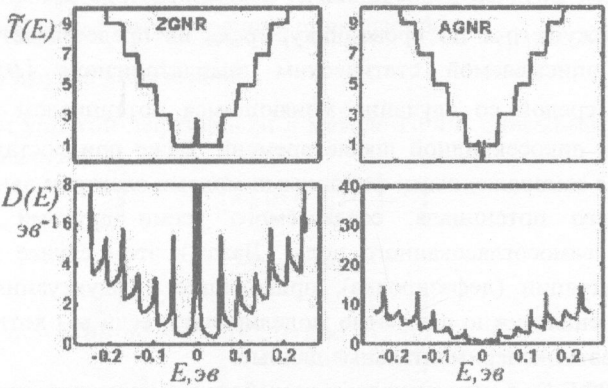


Рис. 31. Коэффициент прохождения и плотность состояний для двух конфигураций границ – зигзагообразной ZGNR (слева) и креслоподобной AGNR (справа).

Обращает на себя внимание высокая плотность состояний при $E = 0$ для зигзагообразной конфигурации графеновой ленты. Это так называемые краевые локальные состояния вблизи уровня Ферми, отсутствующие у креслоподобной конфигурации [38, 39]. В нашу задачу не входит анализ результатов расчета, тем более, что роль конфигурации границ и ширины графеновых лент детально изучены [38 – 42]. Отметим лишь высокую правдоподобность результатов, получаемых даже в рамках такой простейшей хюккелевской модели графена.

10.10. Дефазировка

До сих пор рассматривались только физические контакты $[\Sigma_{1,2}]$ в квантовой модели когерентного транспорта, в которой электроны движутся когерентно от истока к стоку по каналу, описываемому статическим гамильтонианом $[H]$ в отсутствии взаимодействия электрона с окружением при его движении по каналу, описываемому собственной энергией $[\Sigma_0]$ (рис. 32).

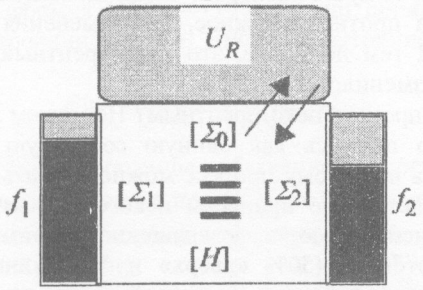


Рис. 32. Модель квантового транспорта с простой упругой дефазировкой.

Что из себя физически представляет взаимодействие Σ_0? С точки зрения электрона, движущегося по проводнику, среда не представляется электрону статической, описываемой статическим гамильтонианом $[H]$, а весьма турбулентной средой со случайно меняющимся потенциалом U_R, который флуктуирует в пикосекундной шкале времени. Даже при достаточно низких температурах с замороженными фононными модами электрон движется в поле флуктуирующего потенциала, создаваемого всеми другими электронами (приближение самосогласованного поля). Даже в этом случае имеют место фазовые флуктуации (дефазировка), приводящие к флуктуации тока. Этот эффект усреднения нужно адекватно моделировать, если мы хотим правильно интерпретировать экспериментальные данные.

Метод НРФГ был первоначально разработан именно для учета неупругих процессов квантового транспорта в массивных проводниках. Мы же пока что изложили его выше применительно к упругим резисторам. Сейчас мы рассмотрим сравнительно простую модель упругой дефазировки.

Что представляет из себя процесс дефазировки? Или другими словами, чем когерентный процесс отличается от некогерентного? Когерентный транспорт это такой процесс движения электрона по проводнику, при котором состояние остальных частиц в канале проводимости не изменяется. Простой отскок электрона от упругого дефекта обратно в канал проводимости есть процесс когерентный. Если же электрон передал часть своей энергии, скажем, атому решетки и атом начал совершать колебания с бóльшей амплитудой, то такой процесс некогерентный. Это вроде бы означает, что некогерентный процесс также должен быть неупругим, происходящим с обменом энергии. Однако же, это не обязательно так, и классическим примером может служить спиновая дефазировка.

Пусть имеем магнитную примесь с двумя состояниями спина одинаковой энергии (вырождение). И пусть электрон, взаимодействуя с примесью, меняет направление спина на противоположное, без изменения его энергии. Такой процесс – упругий. И тем не менее, это некогерентный процесс, поскольку состояние электрона изменилось.

Что делает такой процесс некогерентным? Не можем ли мы рассматривать электрон и спиновую примесь как единую составную систему, состояние которой не изменилось и поэтому процесс можно считать когерентным? Что в действительности делает такие процессы некогерентными, так это внешние силы, которые инспирируют возвращение примесных спинов в неполяризованное состояние (50% «вверх» и 50% «вниз»). Такой процесс «стирания информации» и есть суть спиновой дефазировки. В общем случае

под дефазировкой понимается такой процесс, при котором квантовомеханическая интерференция разрушается.

10.10.1. Упругая дефазировка

Процессы упругой дефазировки в методе НРФГ описываются матрицами собственной энергии

$$\left[\Sigma_0\right] = D \times \left[G^R\right], \tag{176}$$

$$\left[\Sigma_0^{in}\right] = D \times \left[G^n\right], \tag{177}$$

$$\left[\Gamma_0\right] = D \times \left[A\right], \tag{178}$$

где «крест» × означает поэлементное умножение матриц, а элементы матрицы D описывают корреляцию между случайными значениями потенциала в узлах i и j проводника:

$$D_{ij} = \left\langle U_{Ri} U_{Rj} \right\rangle. \tag{179}$$

Две предельные модели корреляционных связей представляют интерес. В первой из них случайный потенциал полностью скоррелирован во всех узлах по всей длине канала проводимости, имеет одно и то же значение для любого узла i, так что все элементы матрицы D одинаковы и равны D_0:

(*Модель A*) $$D_{ij} = D_0. \tag{180}$$

Во второй модели корреляция отсутствует между узлами проводника, матрица D скалярная:

(*Модель B*) $$D_{ij} = D_0 \cdot \delta_{ij}, \tag{181}$$

где δ_{ij} – дельта-функция Кронекера. Реальные процессы описываются обычно промежуточными случаями.

Происхождение формул (176) – (178) можно понять, если обратиться к исходному при построении метода НРФГ в матричной форме уравнению Шредингера

$$E\left\{\psi\right\} = \left[H + \Sigma_1 + \Sigma_2\right]\left\{\psi\right\} + \left\{s_1\right\}, \tag{182}$$

где $[\Sigma_1]$ и $[\Sigma_2]$ – контактные матрицы собственной энергии, а матрица-столбец $\{s_1\}$ описывает приток электронов в проводник с левого контакта 1 (рис. 32).

Случайный потенциал вводится в уравнение Шредингера по аналогии с матрицами собственной энергии

$$E\left\{\psi\right\} = \left[H + \Sigma_1 + \Sigma_2\right]\left\{\psi\right\} + U_R\left\{\psi\right\} + \left\{s_1\right\} \tag{183}$$

с соответствующим притоковым членом

$$\Sigma_0^{in} A = 2\pi U_R U_R^* \{\psi\}\{\psi\}^+ = D_0 G^n, \tag{184}$$

выписанным для модели А. Несколько более длинные выкладки ведут к аналогичному выражению для модели В. В итоге имеем формулу (177).

Для обоснования формулы (178) заметим, что эта формула вместе с (177) должны обеспечивать нулевой ток на «терминале 0», что следует из четвертого уравнения (16) в формализме НРФГ, а именно:

$$I_0 = \frac{q}{h} Tr\left[\Sigma_0^{in} A - \Gamma_0 G^n\right] = D_0 \frac{q}{h} Tr\left[G^n \Gamma_0 - \Gamma_0 G^n\right] = 0 \tag{185}$$

и является обязательным условием, поскольку «терминал 0» не является физическим контактом, где электроны могут переходить в проводник или покидать его.

Действительно, в свое время Бюттекер предложил для учета некогерентных процессов вводить фиктивный контакт (зонд Бюттекера), электрохимический потенциал которого подбирается таким образом, чтобы гарантировать на этом контакте нулевой ток [22, 43]. В формализме НРФГ это ведет к утверждению, что

$$\Sigma_0^{in} = \Gamma_0 f_P, \tag{186}$$

где число f_P подбирается таким, чтобы обеспечить нулевой ток на фиктивном терминале. Это эквивалентно рассуждениям выше, если связь Γ_0 «терминала» с проводником выбрать пропорциональным спектральной функции $[A]$ как это требуется согласно (178).

Обратим внимание на то обстоятельство, что уравнения (176) – (178) предполагают самосогласованную процедуру вычислений, поскольку Σ и Σ^{in} зависят от G^R и G^n, которые в свою очередь, согласно уравнениям (6) и (13), зависят от Σ и Σ^{in}; а также на то, что модель А (180) предполагает вычисление полной гриновской функции, что для достаточно больших устройств ведет к появлению в расчетах матриц весьма большого порядка, в то время как модель В (181) не требует столь значительных вычислительных ресурсов, поскольку вычислению подлежат только диагональные элементы гриновских матриц.

Существенное концептуальное различие между моделями А и В состоит в том, что модель А разрушает фазировку, но не затрагивает импульсы, в то время как модель В разрушает также и импульсы [44]. Процесс дефазировки можно представить себе как «извлечение» электрона из проводника в состоянии $[G^n]$ и последующей «инжекции» его обратно в проводник в состоянии $D \times G^n$. Модель А предполагает умножение гриновской матрицы $[G^n]$ на константу, так что электрон «реинжектируется» точно в то же самое состояние, в котором он был «извлечен», без потери импульса, тогда как в модели В отбрасываются недиагональные элементы матрицы D и при

«реинжектировании» электрон меняет траекторию движения на некую произвольную по сравнению с той траекторией, с которой он был «извлечен». Вся эта ситуация станет яснее ниже при рассмотрении конкретных примеров.

Вместо учета упругой дефазировки через матрицу собственной энергии $[\Sigma_0]$ можно было бы включить случайный потенциал U_R в гамильтониан с дальнейшим усреднением его по возможным случайным реализациям U_R. Оба подхода хотя и не совсем эквивалентны, однако, в ряде случаев приводят к одинаковым результатам, в чем мы также убедимся ниже на конкретных примерах.

Полноты ради, отметим, что в самом общем случае D есть тензор 4-го ранга, связывающий друг с другом пары матриц, а именно:

$$\left[\Sigma_0\right]_{ij} = \sum_{k,l} D_{ijkl}\left[G^R\right]_{kl}, \tag{187}$$

$$\left[\Sigma_0^{in}\right]_{ij} = \sum_{k,l} D_{ijkl}\left[G^n\right]_{kl}, \tag{188}$$

а упругая дефазировка (176) – (178) реализуется через ненулевые D_{ijkl} при $i=k$ и $j=l$.

10.10.2. 1D проводник с двумя и более рассеивающими центрами

Выше подробно рассматривался 1D проводник с одним рассеивающим центром (§ 10.9.1.3), характеризуемым коэффициентом прохождения T. В следующей главе будет показано (§ 11.4), что сопротивление такого проводника R_1 может быть разбито на две части – сопротивление рассеивающего центра и граничное сопротивление (см. также (34) и (35) в [45]):

$$R_1 = \frac{h}{q^2 M}\left(\frac{1-T}{T}+1\right). \tag{189}$$

Зададимся вопросом, каково будет сопротивление R_2 проводника с двумя одинаковыми рассеивающими центрами, каждый с коэффициентом прохождения T. Такой проводник можно рассматривать как последовательное соединение двух проводников, каждый с одним и тем же рассеивающим центром (рис. 33).

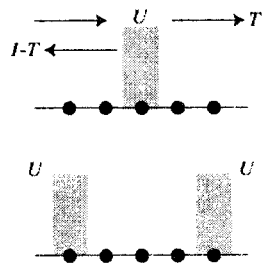


Рис. 33. Проводник с одним и с двумя одинаковыми рассеивающими центрами.

Можно ожидать, что в проводнике с двумя центрами вклад от рассеивающих центров удвоится:

$$R_2 = \frac{h}{q^2 M}\left(2\frac{1-T}{T}+1\right) = \frac{h}{q^2 M}\frac{2-T}{T}, \tag{190}$$

так что

$$R_2 = R_1\left(2-T\right). \tag{191}$$

Если коэффициент прохождения рассеивающих центров T близок к единице, то имеет место баллистический предел проводимости с $R_2 \approx R_1$: два одинаковых проводника, каждый с одним и тем же рассеивающим центром и соединенных последовательно, проводят так же, как один из них, поскольку все сопротивление является граничным.

Если $T \ll 1$, имеет место омический предел с $R_2 \approx 2R_1$: два таких проводника, соединенных последовательно, удваивают сопротивление одного проводника, поскольку все сопротивление связано с каналом проводимости.

Но может ли R_2 быть меньше, чем R_1? Если бы электроны подчинялись законам классической механики, то, конечно, нет. Увеличение количества препятствий на автомобильном шоссе не может увеличить предельный трафик на этом шоссе. А вот на «квантовом шоссе» такое вполне возможно благодаря квантовой (волновой) интерференции.

Для решения задач такого типа воспользуемся рассмотренной выше моделью однородного 1D проводника в хюккелевском приближении (§ 10.9.1.3). Вспомним поведение коэффициента прохождения в проводнике с одним рассеивающим центром $U = 2t_0$ (рис. 11) и сравним его с таким же проводником, но с двумя одинаковыми такими же рассеивающими центрами (рис. 34).

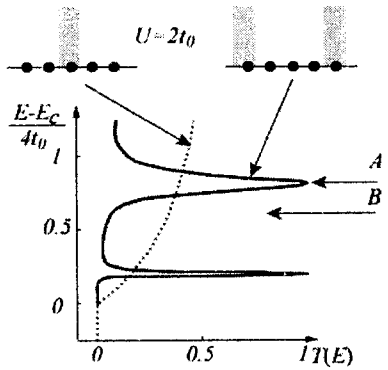


Рис. 34. Коэффициент прохождения в одномодовом 1D проводнике с одним и с двумя одинаковыми рассеивающими центрами с $U = 2t_0$.

Если электрохимический потенциал в устройстве с проводником с двумя рассеивающими центрами окажется на уровне B (рис. 34), то коэффициент прохождения окажется меньше, чем в проводнике с одним центром; другими словами, сопротивление R_2 окажется больше, чем R_1. Если же случится так, что электрохимический потенциал окажется на уровне A (рис. 34), то R_2 окажется даже меньше, чем R_1.

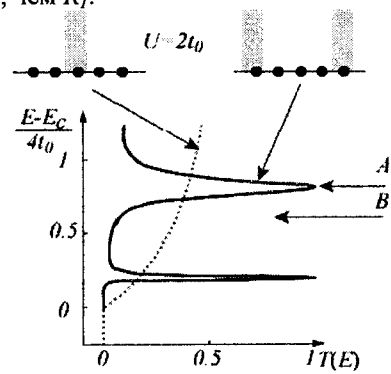


Рис. 34. Коэффициент прохождения в одномодовом 1D проводнике с одним и с двумя одинаковыми рассеивающими центрами с $U = 2t_0$.

Рассмотрим проводник со случайно распределенными рассеивающими центрами. Квантовый расчет методом НРФГ для шести таких центров показывает, что проводимость весьма и весьма низкая (практически омическое поведение), исключая многочисленные всплески проводимости, редко вплоть до значений, близких к единице (рис. 35).

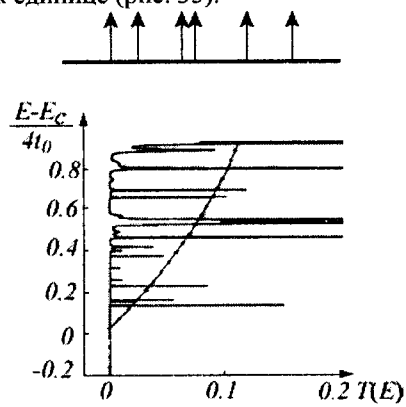


Рис. 35. Коэффициент прохождения в одномодовом 1D проводнике с шестью одинаковыми случайно разбросанными рассеивающими центрами с $U = 2t_0$ [4]. Точками показан результат полуклассического расчета (рис. 11) в пересчете на шесть центров.

На рисунке точками показан также полуклассический результат, полученный вначале для одиночного рассеивающего центра (рис. 11), а затем вклад от него увеличен в шесть раз. Для сопротивления такого проводника имеем

$$R_6 = \frac{h}{q^2 M}\left(6\frac{1-T}{T}+1\right) = \frac{h}{q^2 M}\frac{6-5T}{T}, \qquad (192)$$

где первое слагаемое в скобках обусловлено шестью рассеивающими центрами, а второе слагаемое – граничным сопротивлением.

Квантовые расчеты показывают аналогичное поведение и в мультимодовых проводниках как только классический коэффициент прохождения $M\lambda/L$ оказывается меньше единицы. О таких проводниках говорят, что они находятся в режиме сильной локализации. Любопытно, что даже если $M\lambda/L$ превосходит единицу, то квант проводимости лишь не намного меньше классического значения; о таких проводниках говорят, что они находятся в режиме слабой локализации.

Подобные эффекты локализации обычно экспериментально наблюдаются только при низких температурах. При комнатной температуре крайне редко можно встретиться с отклонением от закона Ома. Рассмотрим, например, медную проволоку с поперечным сечением $10нм \times 10нм$, которое содержит приблизительно 1000 атомов, и стало быть число мод $M \approx 1000$. Пусть средний свободный пробег $\lambda \approx 40нм$. Тогда такая проволока длиной не более чем $M\lambda \approx 40\mu м$ должна обнаруживать неомическое поведение, что не подтверждается экспериментальными наблюдениями. Причина этого кроется в том, что эффекты локализации вызываются квантовой интерференцией и наблюдаются только в том случае, когда имеет место фазовая когерентность по всей длине проводника. Медная проволока длиной $L \approx 40\mu м$ фазово некогерентна, тем более при комнатной температуре. Концептуально реальный проводник можно рассматривать как последовательность отдельных когерентных проводников, каждый длиной, равной длине фазовой когерентности L_p. Видим мы или не видим эффекты локализации зависит не от $M\lambda/L$, а от $M\lambda/L_p$.

Основной вывод состоит в том, что для интерпретации реальных экспериментов при комнатной температуре в формализме НРФГ, как правило, необходимо в той или иной степени учитывать процессы дефазировки, как это описано выше. Пока мы так или иначе не учтем дефазировку, в квантовых моделях электронного транспорта проявятся эффекты интерференции, ведущие к сильной локализации или к резонансному туннелированию. В подтверждение этого рассмотрим изменение потенциала вдоль канала проводника с дефектами методом НРФГ.

10.10.3. Скачок потенциала на дефектах

Профиль потенциала одномодового проводника с одним дефектом прозрачности T рассматривался выше (§ 10.9.1.3). Проводимость такого проводника вычислим позже (§ 11.4) [45/(34)], а его сопротивление, нормированное на квант сопротивления,

$$R_{Norm} = (1-T)/T\,. \tag{193}$$

Полуклассическое выражение для коэффициента прохождения T дается выше формулой (113). Тогда нормированное сопротивление на дефекте равно

$$R_{Norm} = (Ua/\hbar v)^2\,. \tag{194}$$

Полуклассический профиль потенциала со скачками на граничных сопротивлениях $\hbar/q^2$ и на сопротивлении на дефекте (194) показан на рис. 36 вместе с результатом, полученным методом НРФГ без учета дефазировки.

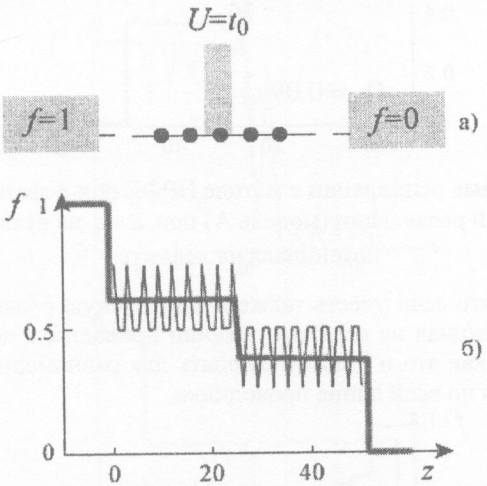


Рис. 36. (а) Схема одномодового проводника с дефектом $U = t_0$; (б) Квантовые осцилляции в методе НРФГ без дефазировки ($D_0 = 0$) в режиме когерентности при $E = t_0$ на фоне полуклассических скачков потенциала.

Техника вычисления заселенности f в методе НРФГ через функцию Грина и спектральную функцию и далее электрохимического потенциала будет описана в следующей главе (§ 11.6) [45/(87),(88)].

Заселенность в расчете методом НРФГ в результате квантовой интерференции осциллирует настолько сильно, что становится затруднительным увидеть скачок потенциала на дефекте (рис. 36).

Экспериментально профиль потенциала обычно измеряется сканирующим туннельным микроскопом и при комнатной температуре квантовые осцилляции обычно не видны из-за процессов дефазировки, неизбежно имеющих место в таких условиях. Приведенный выше пример еще раз показывает, что при моделировании и интерпретации реальных транспортных измерений при комнатной температуре необходимо в той или иной степени учитывать дефазировку эффектов квантовой интерференции.

Действительно, если для той же модели проводника (рис. 36а) при расчете методом НРФГ включить дефазировку с учетом только фазовой релаксации (модель А, ур-е (180)), то получим профиль потенциала с ясно выраженным скачком на дефекте в соответствии с полуклассической моделью (рис. 37).

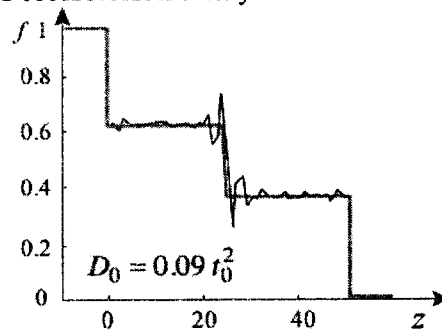


Рис. 37. Квантовые осцилляции в методе НРФГ при дефазировке с учетом только фазовой релаксации (модель А) при $E = t_0$ не вуалируют скачок потенциала на дефекте.

Любопытно, что если учесть также и импульсную релаксацию (модель В, ур-е (181)), то потенциал на всем протяжении проводника подает практически линейно (рис. 38), как это и должно ожидать для равномерно распределенных упругих резисторов по всей длине проводника.

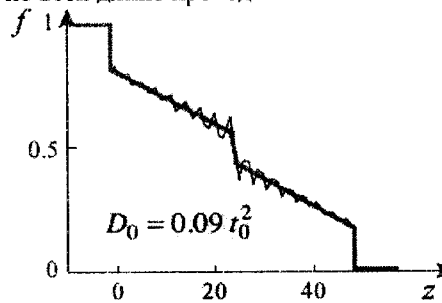


Рис. 38. Учет в методе НРФГ как фазовой, так и импульсной релаксации (модель Б) практически нивелирует скачок потенциала на дефекте, превращая проводник с дефектом в последовательность упругих резисторов.

Сопротивление в пересчете на один узел такого однородного 1D проводника можно получить из (194) путем замены U^2 на D_0:

$$R_{Norm} = D_0 \left(a/v\right)^2 \left(L/a\right), \tag{195}$$

где L/a дает число узлов в проводнике.

Еще одним полезным примером может служить уже рассмотренный выше проводник с двумя одинаковыми рассеивающими центрами (рис. 34). Значения электрохимического потенциала на уровнях A и B соответствуют конструктивной ($R_2 < R_1$) и деструктивной ($R_2 > R_1$) квантовой интерференции. Различие между ними ярко проявляется при расчете такого проводника методом НРФГ в режиме когерентного транспорта без дефазировки с $D_0 = 0$ (рис. 39).

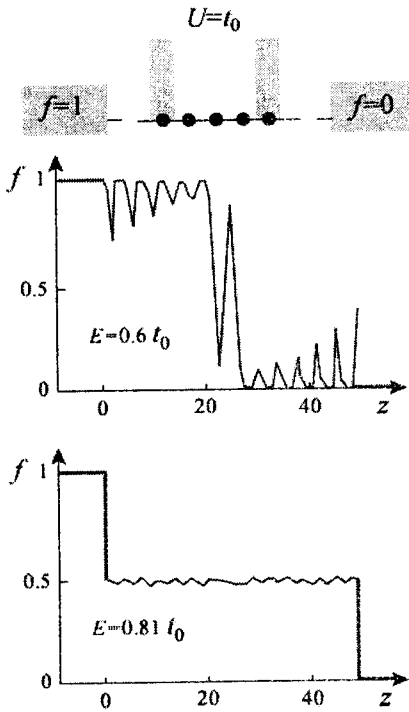


Рис. 39. Квантовые осцилляции в расчете проводника с двумя одинаковыми рассеивающими центрами с $U = t_0$ методом НРФГ без дефазировки ($D_0 = 0$) в режиме когерентности при $E = 0.6t_0$ свидетельствуют о деструктивной интерференции (режим B с $R_2 > R_1$), а при $E = 0.81t_0$ – о конструктивной интерференции (режим A с $R_2 < R_1$).

На уровне B профиль потенциала выглядит так, как будто мы имеем дело с большим скачком потенциала на дефекте, да еще наложенными на него двумя всплесками (деструктивная интерференция). Небольшое увеличение электрохимического потенциала до $E = 0.81t_0$ (уровень A) кардинально меняет профиль потенциала. Теперь он выглядит так (рис. 39), как можно было бы ожидать для баллистического проводника со скачками потенциала лишь на границах с контактами (конструктивная интерференция).

На поставленный выше вопрос «может ли R_2 быть меньше, чем R_1?» ответ однозначный – «Да, может»: два дефекта могут создать меньшее сопротивление, чем один такой дефект. И этот «странный» результат обязан квантовой интерференции.

Стоит лишь в обоих случаях A и B (рис. 39) учесть дефазировку в виде фазовой релаксации с достаточным отличным от нуля значением D_0, как профиль потенциала сразу становится похожим на полуклассический профиль резистора (рис. 40).

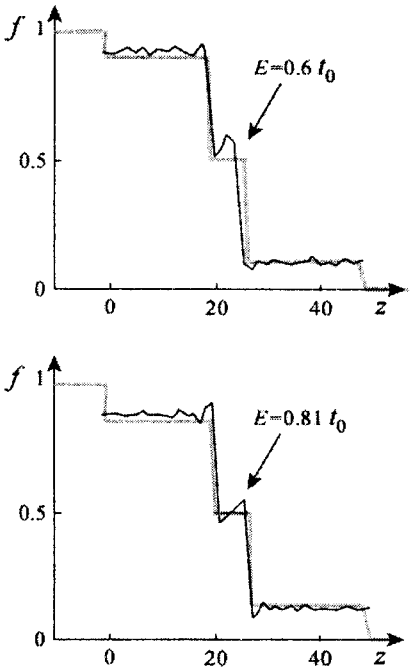


Рис. 40. Квантовые осцилляции в методе НРФГ при дефазировке с учетом только фазовой релаксации ($D_0 = 0.09t_0^2$) при обоих значениях электрохимического потенциала $E = 0.6t_0$ и $E = 0.81t_0$ на фоне полуклассических скачков потенциала.

В вычислительном отношении обратим внимание на то, что учет дефазировки в методе НРФГ алгоритмически не вызывает затруднений, однако, если в когерентном приближении ($D_0 = 0$) или с учетом лишь импульсной релаксации (модель В, ур-е (181)) достаточно вычислять лишь диагональные элементы гриновской функции, то в случае учета лишь фазовой релаксации необходимо вычислять полную функцию Грина, что существенно увеличивает время вычислений и требует намного больших вычислительных ресурсов.

Резонно также поставить вопрос: а что если вместо учета дефазировки через матрицу собственной энергии Σ_0 потенциал U_R включить в гамильтониан H, а затем усреднить его по всем возможным реализациям, приведет ли это к тем же самым результатам?

Для коротких резисторов таких, как на рис. 34, ответ будет скорее всего положительным, а вот для длинных резисторов, как на рис. 35, ответ будет отрицательным. В случае проводника в режиме сильной локализации (рис. 35) трудно представить себе, каким образом усреднение когерентного квантового состояния по многим возможным конфигурациям приведет к полуклассическому результату.

Метод НРФГ с дефазировкой, не вдаваясь в тонкости, не сводится просто к усреднению по многим конфигурациям, а включает в себя также усреднение по фрагментам отдельных конфигураций, в результате чего и достигается полуклассическое описание проводимости, столь хорошо зарекомендовавшее себя при интерпретации реальных измерений при комнатной температуре.

Подведем итоги. В рамках концепции «снизу – вверх» наноэлектроники рассмотрен метод неравновесных функций Грина в матричной формулировке и применение его к модельным транспортным задачам 1D и 2D проводников в хюккелевской модели. Сформулирован общий метод учета электрических контактов в уравнении Шредингера для решения задач квантового транспорта электронов. Рассмотрены модели упругой дефазировки и спиновой дефазировки, учет некогерентных процессов с использованием зонда Бюттекера, 1D проводник с двумя и более рассеивающими центрами, явление квантовой интерференции, режимы сильной и слабой локализации, скачок потенциала на дефектах, квантовые осцилляции в методе НРФГ без учета дефазировки и с ее учетом, эффекты деструктивной и конструктивной интерференции [46 – 50].

Литература

1. Ю. О. Кругляк, Н. Ю. Кругляк, М. В. Стріха. Уроки наноелектроніки: виникнення струму, формулювання закону Ома і моди провідності в концепції «знизу – вгору», *Сенсор. електр. мікросист. техн.*, т. 9, № 4, 5 – 29 (2012).

2. Ю. А. Кругляк. Наноэлектроника «снизу – вверх»: Возникновение тока, обобщенный закон Ома, упругий резистор, моды проводимости, термоэлектричество, *ScienceRise*, т. 7, № 2(12), 76 – 100 (2015).
3. Ю. А. Кругляк. Наноэлектроника «снизу – вверх»: Начало спинтроники и магнетроники, *ScienceRise*, . – 2015. – т. 8, № 2(13), 51 – 68 (2015).
4. S. Datta. *Lessons from Nanoelectronics: A New Perspective on Transport.* Hackensack, New Jersey: World Scientific Publishing Company, 2012; www.nanohub.org/courses/FoN1.
5. S. Datta. Nanoelectronic devices: A unified view, In *The Oxford Handbook on Nanoscience and Nanotechnology: Frontiers and Advances, v. 1*, Eds. A.V. Narlikar and Y.Y.Fu. Oxford: Oxford University Press, 2012.
6. S. Datta. *Quantum Transport: Atom to Transistor*. Cambridge: Cambridge University Press, 2005.
7. S. Datta. Nanodevices and Maxwell's demon, In *Lecture Notes in Nanoscale Science and Technology, v. 2, Nanoscale Phenomena: Basic Science to Device Applications*, Eds. Z.K. Tang and P. Sheng. Berlin: Springer, 2008.
8. C. Caroli, R. Combescot, P. Nozieres, D. Saint-James. A direct calculation of the tunneling current: IV. Electron phonon interaction effects, *J. Phys. C,* v. 5, 21 (1972).
9. R. Kubo. Statistical-Mechanical Theory of Irreversible Processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems, *J. Phys. Soc. Japan,* v. 12, 570 – 586 (1957).
10. F. W. Sears, G. L. Salinger. *Thermodynamics, Kinetic Theory, and Statistical Thermodynamics*. Boston: Addison-Wesley, 1975.
11. P. C. Martin, J. Schwinger. Theory of many-particle systems. I, *Phys. Rev.*, v. 115, n 6, 1342 – 1373 (1959).
12. L. P. Kadanoff, G. Baym. *Quantum Statistical Mechanics*. New York: W. A. Benjamin, 1962.
13. Л. В. Келдыш. Диаграммная техника для неравновесных процессов, *ЖЭТФ*, т. 47, 1515 – 1527 (1964); L. V. Keldysh. Diagram Technique for Non-Equilibrium Processes, *Sov. Phys. JETP*, v. 20, 1018 (1965).
14. R. Landauer. Spatial variation of currents and fields due to localized scatterers in metallic conduction, *IBM J. Res. Dev.*, v. 1, n.3, 223 – 231 (1957).
15. R. Landauer. Electrical resistance of disordered one dimensional lattices, *Philos. Mag.*, v. 21, 863 – 867 (1970).
16. R. Landauer. Electrical resistance of disordered one dimensional lattices, *J. Math. Phys.*, v. 37, n. 10, 5259 (1996).
17. S. Datta. Steady-state quantum kinetic equation, *Phys. Rev. B,* v. 40, 5830 (1989).
18. S. Datta. A simple kinetic equation for steady-state quantum transport, *J. Phys. Cond. Matt.*, v. 2, 8023 – 8052 (1990).
19. Y. Meir, N. S. Wingreen. Landauer formula for the current through an interacting electron region, *Phys. Rev. Lett.*, v. 68, 2512 – 2515 (1992).

20. S. Datta. *Electronic Transport in Mesoscopic Systems*. Cambridge: Cambridge University Press, 2001.
21. R. H. M. Smit, Y. Noat, C. Untiedt, N. D. Lang, M. C. van Hemert, J. M. van Ruitenbeek. Measurment of the conductance of a hydrogen molecule, *Nature*, v. 419, n 3, 906 – 909 (2002).
22. M. Buttiker. Symmetry of Electrical Conduction, *IBM J. Res. Dev.*, v. 32, n 3, 317 – 334 (1988).
23. P. W. Anderson. Absence of Diffusion in Certain Random Lattices, *Phys. Rev.*, v. 109, n 5, 1492 – 1505 (1958).
24. P. W. Anderson. New method for scaling theory of localization. II. Multichannel theory of a "wire" and possible extension to higher dimensionality, *Phys. Rev. B*, v. 23, n 10, 4828 – 4836 (1981).
25. Ю. А. Кругляк *et al. Методы вычислений в квантовой химии. Расчет π-электронной структуры молекул простыми методами молекулярных орбиталей*. Киев: Наукова думка, 1967.
26. Yu. A. Kruglyak, I. I. Ukrainsky. Study of the electronic structure of alternant radicals by the DODS method, *Intern. J. Quantum Chem.*, v. 4, n 1, 57 – 72 (1970).
27. Yu. A. Kruglyak. Quantum-chemical studies of quasi-one-dimensional electron systems. 1. Polyenes, *ScienceRise*, т. 5, № 2 (10), 69 – 105 (2015).
28. G. F. Kventsel, Yu. A. Kruglyak, Local Electronic States in Long Polyene Chains, Theor. Chim. Acta, v. 12, 1 – 17 (1968).
29. М. В. Стріха. Фізика графену: стан і перспективи, *Сенсор. електр. мікросист. техн.*, т. 7, № 3, 5 – 13 (2010).
30. Ю. А. Кругляк. Графен в транспортной модели Ландауэра – Датты – Лундстрома, *ScienceRise*, т. 2, № 2 (7), 93 – 106 (2015).
31. Yu. A. Kruglyak, G. G. Dyadyusha. Torsion Barriers of End-Groups in Cumulenes. I. General Consideration, *Theor. Chim.Acta*, v. 10, 23 – 32 (1968).
32. Yu. A. Kruglyak, G. G. Dyadyusha. Torsion Barriers of End-Groups in Cumulenes. II. Results of Calculations and Discussion, *Theor. Chim.Acta*, v. 12, 18–28 (1968).
33. Yu. A. Kruglyak, Quantum-chemical studies of quasi-one-dimensional electron systems. 2. Cumulenes and origin of the forbidden zone, *ScienceRise*, т. 6, № 2 (11), 122 – 148 (2015).
34. B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel, C. T. Foxon. Quantized Conductance of Point Contacts in a Two-Dimensional Electron Gas, Phys. Rev. Lett., v. 60, n 9, 848 – 850 (1988).
35. D. A. Wharam, T. J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J. E. F. Frost, D. G. Hasko, D. C. Peacock, D. A. Ritchie, G. A. C. Jones. One-Dimensional Transport and the Quantization of the Ballistic Resistance, *J. Phys. C*, v. 21, n 8, L209 – L214 (1988).

36. R. G. Mojarad, A. N. M. Zainuddin, G. Klimeck, S. Datta. Atomistic non-equilibrium Green's function simulations of graphene nano-ribbons in the quantum hall regime, *J. Comput. Electron.*, v. 7, 407 – 410 (2008).
37. C. Berger, S. Zhimin, Xuebin Li, Xiaosong Wu, N. Brown, C. Naud, D. Mayou, Tianbo Li, J. Hass, A. N. Marchenkov, E. H. Conrad, P. N. First, W. A. de Heer. Electronic Confinement and Coherence in Patterned Epitaxial Graphene, Science, v. 312, 1191 – 1196 (2006).
38. M. Fujita, K. Wakabayashi, K. Nakada, K. Kusakabe. Peculiar Localized State at Zigzag Graphite Edge, *J. Phys. Soc. Japan*, v. 65, n 7, 1920 (1996).
39. K. Nakada, M. Fujita, G. Dresselhaus, M. S. Dresselhaus. Edge state in graphene ribbons: Nanometer size effect and edge shape dependence, *Phys. Rev. B*, v. 54, n 24, 17954 (1996).
40. L. Brey, H. A. Fertig. Electronic states of graphene nanoribbons studied with the Dirac equation, *Phys. Rev. B*, v. 73, n 23, 235411 (2006).
41. K. Wakabayashi, Y. Takane, M. Yamamoto, M. Sigrist. Electronic transport properties of graphene nanoribbons, *New J. Phys.*, v. 11, 095016 (2009).
42. M. Koch, F. Ample, C. Joachim, L. Grill. Voltage-dependent conductance of a single graphene nanoribbon, *Nature Nanotech.*, v. 7, 713 – 717 (2012).
43. M. Buttiker. Four-terminal phase-coherent conductance, *Phys. Rev. Lett.*, v. 57, 1761 (1986).
44. R. Golizadeh-Mojarad, S. Datta. Non-equilibrium Green's function based model for dephasing in quantum transport, *Phys. Rev. B*, v. 75, n 8, 081301 (2007).
45. Ю. О. Кругляк, М. В. Стріха. Уроки наноелектроніки: Ефект Хола і вимірювання електрохімічних потенціалів в концепції «знизу – вгору», *Сенсор. електр. мікросист. техн.*, т. 11, № 1, 5 – 27 (2014).
46. Yu. A. Kruglyak. Non-equilibrium Green's function method in matrix representation and model transport problems of nanoelectronics, *Proc. Intern. Conf: Nanomaterials: Applications and Properties*, v. 2, N 4, 04NAESP01(6pp) (2013).
47. Ю. А. Кругляк, Т. В. Крыжановская. Метод неравновесных функций Грина в матричном представлении. 1. Теоретические основы, *Вестник Одес. гос. эколог. ун-та*, № 18, 175 – 192 (2014).
48. Ю. А. Кругляк, Т. В. Крыжановская. Метод неравновесных функций Грина в матричном представлении. 2. Модельные транспортные задачи, *Вестник Одес. гос. эколог. ун-та*, № 19, 202 – 209 (2015).
49. Ю. А. Кругляк. Наноэлектроника «снизу – вверх»: метод неравновесных функций Грина, модельные транспортные задачи и квантовая интерференция, *ScienceRise*, т. 9, № 2(14), 41 – 72 (2015).
50. Ю. А. Кругляк. Метод неравновесных функций Грина в матричном представлении: модельные транспортные задачи наноэлектроники в концепции «снизу – вверх», *Наносистемы, наноматериалы, нанотехнологии*, т. 13, № 2, 243 – 292 (2015).

Глава 11. Эффекты Холла, измерение электрохимических потенциалов, транспорт спинов в модели НРФГ и квантовая природа классики

11.1. Эффекты Холла

Эдвин Холл в 1879 году обнаружил возникновение поперечной разности потенциалов при помещении тонкой золотой пластинки с постоянным током в магнитное поле [1]. В простейшем варианте эффект Холла проявляется следующим образом. Поместим 2D проводник с электрическим током в слабое магнитное поле $\vec{B}$, направленное перпендикулярно к поверхности проводника – вдоль оси y на рис. 1.

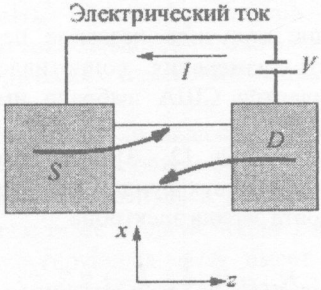


Рис. 1. Магнитное поле, направленное перпендикулярно к проводнику с током, отклоняет электроны к краям проводника.

Магнитное поле под действием силы Лоренца

$$\vec{F} = \frac{d\vec{p}}{dt} = -q\vec{v} \times \vec{B} \qquad (1)$$

отклоняет электроны, движущиеся со скоростью $\vec{v}$ от истока S к стоку D, от их траектории вдоль или против вектора скорости к боковым краям проводника (рис. 1). Критерием малости магнитного поля обычно служит условие, чтобы электроны не начали двигаться по циклоиде.

Поскольку больше электронов движется от истока к стоку, создается холловская разность потенциалов V_H в поперечном направлении x (рис. 2).

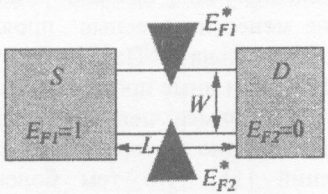


Рис. 2. Холловская разность потенциалов создается в поперечном направлении.

Эффект Холла всегда вызывал значительный интерес. Он особенно возрос в 1980 году, когда К. фон Клитцинг обнаружил, что холловское сопротивление

$$R_H = \frac{V_H}{I} \tag{2}$$

принимает дискретные значения [2, 3], которые могут быть вычислены с помощью следующей простой формулы

$$R_H = \frac{25812.80}{k}, k = 1,2,3,... \tag{3}$$

Обнаруженное явление получило название целочисленного квантового эффекта Холла. Точность измерения сопротивления столь велика, что Национальное бюро стандартов США выбрало именно эффект Холла для калибровки сопротивлений [4].

К. фон Клитцинг показал [2, 3], что холловское сопротивление фактически измеряет отношение двух фундаментальных физических констант – постоянной Планка и квадрата заряда электрона

$$R_H = \frac{h}{q^2} \cdot \frac{1}{k}, \quad k = 1,2,3,..., \tag{4}$$

а через скорость света c связано с постоянной тонкой структуры

$$\alpha = \frac{q^2}{\hbar c} = \frac{1}{137.0359895}, \tag{5}$$

что имеет исключительно важное значение для метрологии фундаментальных констант физики.

Уже в 1982 году был экспериментально обнаружен дробный квантовый эффект Холла [5, 6]. В ферромагнетиках в холловском сопротивлении обнаружен вклад, зависящий непосредственно от намагниченности материала, причем этот вклад может быть намного больше нормального холловского сопротивления. Это явление получило название аномального эффекта Холла [7]. Известны и другие, не менее интересные проявления эффекта Холла – спиновый эффект Холла Дьяконова – Переля [8, 9], квантовый спиновый эффект Холла в графене [10], как и иные проявления действия магнитного поля на электрические цепи, как, например, невзаимные электрические цепи (non-reciprocity circuits), которые трудно интерпретировать в терминах обычных эффективных сопротивлений [11], что тем более интересно, поскольку появился новый класс материалов – топологические диэлектрики (topological insulators), которые как будто бы обладают невзаимностью даже в отсутствии магнитных полей [12]. Эти и другие смежные вопросы вплоть до квантовой

природы классики мы рассмотрим в этой главе с разной степенью полноты с позиций концепции «снизу – вверх» современной наноэлектроники [13].

Рассмотрим подробнее обычный классический эффект Холла при низких магнитных полях. Одной из причин, вызвавших повышенный интерес к эффекту Холла, был тот факт, что холловское сопротивление имеет разный знак для *n*- и *p*-проводников, что открывает возможность экспериментально определять различие между ними. Этой же цели можно достичь, измеряя направление тока по проводнику между контактами, находящимися при разных температурах [14]. И в том и в другом случае обычно пользуются понятием «дырок» как носителей положительного заряда в *p*-проводниках. Подобную трактовку нельзя признать удовлетворительной, поскольку и в *n*- и в *p*-проводниках носителями тока являются только электроны.

В случае термоэлектрических измерений направление тока определяется наклоном плотности состояний $D(E)$, то ли эта функция является растущей с ростом энергии (*n*-проводник), то ли падающей (*p*-проводник) [14]. Знак же холловского эффекта определяется знаком эффективной массы, определяемой по (244/гл.1) [15] как отношение импульса p к скорости dE/dp. В результате, хотя сила Лоренца (1) одна и та же и для *n*- и для *p*-проводников, давая одно и то же значение dp/dt, результирующая dv/dt имеет разные знаки для *n*- и *p*-проводников, закручивая электроны в последних в противоположном направлении, что и приводит к изменению знака холловской разности потенциалов.

Обратимся к элементарной теории эффекта Холла, как она обычно излагается в учебниках по физике твердого тела, например [16]. Ток дается выражением

$$I = q\left(N/L\right)v_d, \tag{6}$$

где N/L есть линейная плотность электронов, а их дрейфовая скорость равна произведению подвижности электронов на электрическое поле вдоль длины проводника

$$v_d = \mu\left(V/L\right). \tag{7}$$

В состоянии равновесия поперечное поле компенсирует силу Лоренца

$$V_H/W = v_d B, \tag{8}$$

так что используя (6) и (7) для холловского сопротивления имеем

$$R_H = \frac{V_H}{I} = \frac{B}{q\left(N/LW\right)}. \tag{9}$$

Это уравнение широко используется для определения поверхностной плотности электронов N/LW по наклону зависимости холловского сопротивления от магнитного поля.

Эта элементарная теория эффекта Холла излагается с теми или иными незначительными нюансами во всех учебниках. И везде фигурирует электрическое поле как в продольном, так и в поперечном направлении в качестве причины возникновения электрического тока. Согласиться с этим никак нельзя [15]. Согласно (9), холловское сопротивление зависит от электронной плотности по всему спектру энергий, что, конечно, не так. Как и любые другие транспортные коэффициенты [14], холловское сопротивление – это свойство фермиевской поверхности и зависит оно только от электронов, энергия которых лежит в интервале приблизительно нескольких kT вокруг значения $E = E_{F0}$ [15].

Мы покажем, что холловское сопротивление для одномодового упругого резистора дается выражением

$$R_H(E) = \frac{2BLW}{qD(E)v(E)p(E)}, \tag{10}$$

которое нужно еще усреднить по энергии в интервале нескольких kT вокруг значения $E = E_{F0}$, используя стандартную функцию теплового уширения F_T (глава 1), а именно:

$$\frac{1}{R_H} = \int_{-\infty}^{+\infty} dE \left(-\frac{\partial f_0}{\partial E}\right) \frac{1}{R_H(E)}. \tag{11}$$

Обратим внимание на то, что в общем случае нужно усреднять проводимость $1/R_H(E)$, а не сопротивление $R_H(E)$, поскольку все модовые каналы проводимости работают параллельно и находятся под одной и той же разностью потенциалов.

Уравнения (10) – (11) сводятся к стандартному выражению (9), используя фундаментальное соотношение [15]

$$D(E)v(E)p(E) = N(E)\cdot d \tag{12}$$

для 2D проводника (d = 2) и усредняя $N(E)$, как это было показано в главе 1. Если же дисперсионное соотношение для изучаемого проводника не известно, то остается только пользоваться уравнениями (10) – (11).

В любом случае уравнение (11) свидетельствует о том, что эффект Холла не затрагивает все электроны при всех возможных энергиях. Есть, правда, одно явление, которое затрагивает все электроны при любых энергиях, – это существование равновесных токов в проводнике, помещенном в магнитное поле (рис. 3).

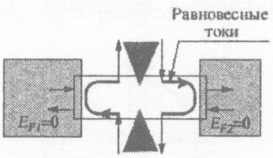


Рис. 3. Равновесные токи существуют в любом проводнике, находящемся в магнитном поле.

Однако же, неисчезающие токи, которые существуют даже в атоме водорода, помещенном в магнитное поле, не имеют никакого отношения к транспортным коэффициентам. Транспортные модели надлежит строить таким образом, чтобы нефермиевские токи элиминировались с самого начала. Подобная проблема по отношению к спиновым токам возникает даже без учета внешнего магнитного поля [17].

Далее мы рассмотрим расчет неравновесных электрохимических потенциалов внутри проводника, а также моделирование проводника в четырехтерминальном окружении с использованием уравнения Бюттикера [18, 19]. Однако, сначала конспективно рассмотрим общую динамику электронов в магнитном поле.

11.2. Почему отличаются между собой *n*- и *p*-проводники?

Почему холловское сопротивление имеет противоположные знаки для *n*- и *p*-проводников? Основная причина состоит в том, что в *n*-проводниках скорость и импульс сонаправлены, а в *p*-проводниках – антинаправлены; и поскольку $v = dE/dp$, а в *p*-проводниках энергия уменьшается с ростом импульса (рис. 4).

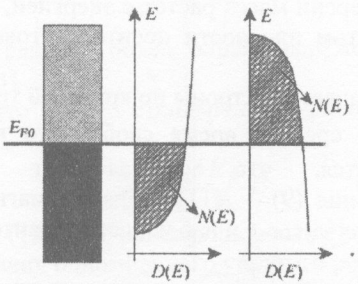


Рис. 4. Холловское сопротивление имеет разные знаки для *n*-проводников (слева) и *p*-проводников (справа) и обратно пропорционально $N(E)$ при $E = E_{F0}$.

Чтобы увидеть какую роль играет различие в знаках скорости и импульса, обратимся к силе Лоренца (1). Для любого изотропного дисперсионного соотношения скорость и импульс коллинеарны (параллельны или

антипараллельны) и пусть находятся, например, под углом θ к продольной оси z проводника (рис. 5).

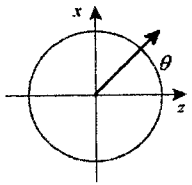


Рис. 5. К разложению импульса и скорости электрона на проекции вдоль и поперек 2D проводника (рис. 1).

Тогда

$$\begin{aligned} \vec{p} &= p\cos\theta\hat{z} + p\sin\theta\hat{x} \\ \vec{v} &= v\cos\theta\hat{z} + v\sin\theta\hat{x} \end{aligned}, \tag{13}$$

а после подстановки в (1) получим

$$\frac{d\theta}{dt} = \frac{qvB}{p}. \tag{14}$$

Другими словами, угол θ меняется линейно со временем, а векторы скорости и импульса вращаются с постоянной угловой скоростью ω_C. Однако, для n- и p-проводников вращение происходит в противоположных направлениях, поскольку отношение p/v у них имеет разные знаки. Это отношение определяется как масса (244/гл.1) и является константой для параболической дисперсии, так что циклотронная частота

$$\omega_C = \left|\frac{qvB}{p}\right|_{E=E_{F0}} = \left|\frac{qB}{m}\right|_{E=E_{F0}}. \tag{15}$$

Для линейной дисперсии масса растет с энергией, и циклотронная частота (15) уменьшается с ростом плотности носителей тока, что и наблюдается в графене [20].

Магнитное поле вращает электроны по круговой траектории с частотой ω_C. Если $\omega_C\tau \ll 1$, где τ – среднее время свободного вращения, то вращение никогда не заканчивается, что соответствует режиму наблюдения холловского сопротивления (9) – (11) в слабом магнитном поле. В режиме сильного магнитного поля $\omega_C\tau \gg 1$ и наблюдается квантовый эффект Холла.

11.3.Пространственный профиль электрохимического потенциала

Приведем соображения в пользу формулы (10) – (11) для холловского сопротивления, опираясь на общую теорию баллистического транспорта (глава 1) [21].

В модели упругого резистора скорость дрейфа определяется разностью потенциалов

$$\delta E_F = E_F^+ - E_F^- \tag{16}$$

между состояниями, связанными со стоком, и состояниями, связанными с истоком, так что вместо уравнения для тока (6) запишем уравнение (глава 1) [21]

$$I(E) = \frac{q}{h} M(E) \left(-\frac{\partial f_0}{\partial E} \right) \delta E_F, \tag{17}$$

где для числа мод используется известная формула (глава 1)

$$\frac{M(E)}{h} = \frac{D(E)\nu(E)}{\pi L}. \tag{18}$$

Из уравнений (20/гл.7) и (29/гл.7) вместо уравнения (7) имеем разность потенциалов, связанную с напряжением, приложенным к концам проводника,

$$\delta E_F = \frac{qV\lambda}{L+\lambda} \cong q\lambda \frac{V}{L}. \tag{19}$$

Так же как из (6) и (7) следует формула Друде для проводимости

$$G = \frac{I}{V} = \sigma \frac{W}{L}, \tag{20}$$

где удельная проводимость

$$\sigma = \frac{q\mu N}{WL}, \tag{21}$$

так и из (17) и (18) следует более общее выражение для проводимости, полученное в главе 1 [14].

Сейчас покажем, что вместо стандартного уравнения эффекта Холла (8) имеет место уравнение

$$\frac{V_H}{W} = \frac{2}{\pi} \frac{\delta E_F}{p} B \tag{22}$$

которое вместе с (17) и (18) дает приведенное выше выражение (10) – (11) для холловского сопротивления одномодового упругого резистора.

Ранее в главе 7 было получено уравнение (17) [21], а именно:

$$v_z \frac{\partial E_F}{\partial z} = -\frac{E_F - E_{F0}}{\tau}, \tag{23}$$

решения которого определяют электрохимические потенциалы E_F^+ и E_F^- (рис. 6).

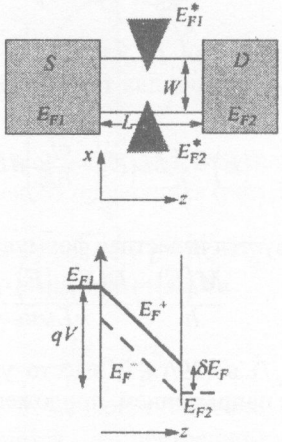


Рис. 6. Пространственный профиль потенциалов E_F^+ и E_F^- вдоль длины проводника.

Покажем, что эти решения могут быть записаны в виде

$$E_F(z,\theta) \equiv \bar{E}_F(z) + \frac{2}{\pi}\delta E_F \cos\theta . \qquad (24)$$

Поскольку речь идет об упругом резисторе, электроны в котором движутся при фиксированной энергии и стало быть имеют постоянный импульс, удобно использовать цилиндрические координаты p, θ (рис. 7) вместо декартовых p_x, p_y.

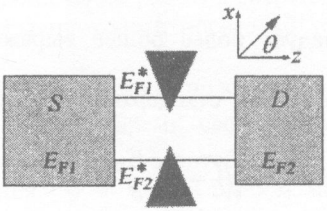


Рис. 7. К ориентации цилиндрических координат p, θ относительно канала проводимости.

Пусть мы рассматриваем электроны, движущиеся под фиксированным углом θ. Тогда с учетом (19) получим

$$E_F(z) = \bar{E}_F(z) + \frac{qV}{L}\nu\tau\cos\theta , \qquad (25)$$

имея в виду, что в этом простом случае длина свободного пробега

$$2v_z\tau = 2v\tau\cos\theta. \tag{26}$$

Сравнивая (25) и (24) и выполнив усреднение по углам для 2D проводника (глава 1 [15]), для разности потенциалов получаем

$$\delta E_F \approx \frac{qV}{L}\frac{\pi}{2}v\tau, \tag{27}$$

и, наконец, комбинируя (25) и (27), получаем искомое выражение (24).

Теперь вопрос в том, как изменится решение (24) уравнения (23) при учете силы Лоренца, создаваемой магнитным полем. Для этого нам потребуется транспортное уравнение Больцмана в приближении времени релаксации (глава 5) [22]:

$$v_x\frac{\partial E_F}{\partial x} + v_z\frac{\partial E_F}{\partial z} + F_x\frac{\partial E_F}{\partial p_x} + F_z\frac{\partial E_F}{\partial p_z} = -\frac{E_F - E_{F0}}{\tau}, \tag{28}$$

где слева сохранены все x- и z-компоненты, поскольку речь идет о 2D проводнике в плоскости xz (рис. 1). Это уравнение по сравнению с его частным случаем (23) содержит три новых слагаемых. Два из них связаны с силой Лоренца и могут быть переписаны в виде

$$F_x\frac{\partial E_F}{\partial x} + F_z\frac{\partial E_F}{\partial z} = \vec{F}\cdot\vec{\nabla}_\mu = \frac{F_\theta}{p}\frac{\partial E_F}{\partial\theta} + F_r\frac{\partial E_F}{\partial r}, \tag{29}$$

где радиальная составляющая силы Лоренца (1) отсутствует, а угловая составляющая

$$F_r = 0, \quad F_\theta = -qvB. \tag{30}$$

Теперь уравнение (28) запишется в виде

$$v_x\frac{\partial E_F}{\partial x} + v_z\frac{\partial E_F}{\partial z} - \frac{qvB}{p}\frac{\partial E_F}{\partial\theta} = -\frac{E_F - E_{F0}}{\tau}. \tag{31}$$

С учетом того, что решением уравнения (23) служит уравнение (24), не трудно убедиться, что решением уравнения (31) будет

$$E_F(z,\theta,x) = \bar{E}_F(z) + \frac{2}{\pi}\delta E_F\cos\theta - \frac{2}{\pi}\frac{\delta E_F}{p}qBx, \tag{32}$$

откуда следует

$$-qV_H = E_F(x=W) - E_F(x=0) = -\frac{2}{\pi}\frac{\delta E_F}{p}qBW \tag{33}$$

и соответственно искомое уравнение (22).

Прежде чем перейти к интересующему нас измерению потенциала Холла в четырехтерминальной модели проводника рассмотрим более подробно общие вопросы измерения электрохимических потенциалов.

11.4. Измерение электрохимических потенциалов

Классическое распределение Ферми с высокой точностью пригодно для массивных контактов, которые всегда остаются практически в равновесном состоянии, но это не обязательно справедливо для небольших проводников даже при малой приложенной разности потенциалов. В главе 7 [21] показано, как важно ввести два отдельных электрохимических потенциала E_F^+ и E_F^- для понимания физического смысла граничного сопротивления как основы обобщенной формулировки закона Ома. Неравновесные электрохимические потенциалы этого типа оказываются весьма полезными и широко используются в практической наноэлектронной инженерии, хотя и поныне все еще высказываются сомнения в их концептуальной ценности и даже измеримости.

Далее на простом примере мы рассмотрим особенности неравновесных потенциалов и вопросы их измеримости, что позволит увязать их с представлениями и формулами Ландауэра и Бюттекера [18, 23 – 27], которые занимают центральное место в теории транспортных процессов современной мезофизики.

Следуя Ландауэру [23 – 27], рассмотрим баллистический канал с одним локализованным дефектом, пропускающим долю T падающих на дефект электронов и отражающим долю $1 - T$ назад (рис. 8). Мы могли бы следовать логике рассуждений в главе 7 [21] и получили бы пространственные профили потенциалов E_F^+ и E_F^- в канале с дефектом, как это качественно показано на рис. 8, и таким образом оценили бы сопротивление, оказываемое дефектом, как это проделал Ландауэр еще в 1957 году.

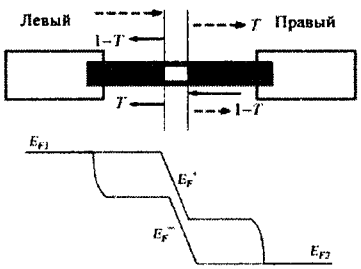


Рис. 8. Профиль неравновесных потенциалов в канале с одним дефектом.

Преодолеть высказываемые сомнения в реалистичности неравновесных потенциалов могли бы четырехтерминальные эксперименты (рис. 9) с

использованием двух дополнительных контактов, проводящих незначительный ток, с тем, чтобы измерить падение напряжения на дефекте.

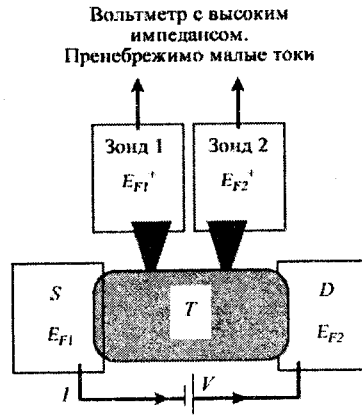


Рис. 9. Четырехтерминальный эксперимент по измерению проводимости 1D канала с одним дефектом с коэффициентом пропускания T.

Мы покажем, что в случае одинаковых и слабосвязанных пробных контактов четырехтерминальная проводимость

$$G_{4t} = \frac{I}{\left(E_{F1}^{*} - E_{F2}^{*}\right)/q} = M\frac{q^2}{h}\frac{T}{1-T}, \tag{34}$$

где M – число мод проводника, а двухтерминальная проводимость имеет меньшее значение

$$G_{2t} = \frac{I}{\left(E_{F1} - E_{F2}\right)/q} = M\frac{q^2}{h}T, \tag{35}$$

так что различие между ними и есть граничное сопротивление (рис. 10)

$$\frac{1}{G_{2t}} - \frac{1}{G_{4t}} = \frac{h}{q^2 M}. \tag{36}$$

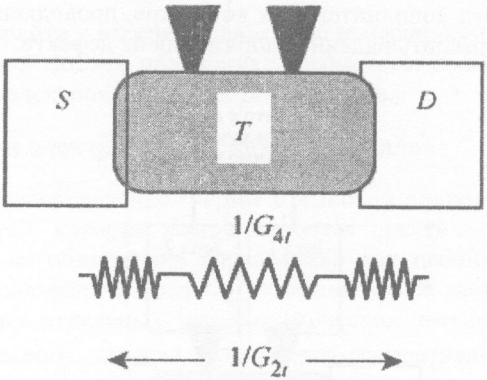


Рис. 10. Двухтерминальное сопротивление можно рассматривать как сопротивление на дефекте, соединенное последовательно с граничными сопротивлениями.

Хотя граничное сопротивление было предсказано и обнаружено Ю.В.Шарвиным у металлических проводников еще в 1965 году [28, 29], его вездесущая роль признана все еще недостаточно широко и сегодня. В первой половине 80-х годов шла дискуссия о смысле и различии двух формул проводимости (34) и (35), когда, наконец, в 1986 г. Имри не идентифицировал различие между ними (36) как граничное сопротивление [30, 31]. При становлении мезофизики уравнение (35) получило широкую известность и было известно как формула Ландауэра, тогда как действительно оригинальная формула Ландауэра (34) была почти забыта. Падение напряжения на дефекте, контактное сопротивление и обе формулы Ландауэра подробно обсуждаются в обзоре Лесовика и Садовского [32].

Одна из проблем в годы становления мезофизики состояла в том, что контакты были сильно связаны с основным проводником и вели себя как «дополнительные дефекты», исключить которые было не так просто. Бюттекер [18, 27] с целью интерпретации четырехтерминальных измерений предложил элегантную формулу

$$I_m = \left(1/q\right)\sum_n G_{m,n}\left(E_{Fm} - E_{Fn}\right), \tag{37}$$

связывающую ток на контакте с номером m с электрохимическими потенциалами на других контактах, где $G_{m,n}$ есть проводимость, определяемая через коэффициент прохождения $T_{m,n}$ между контактами m и n. Для двухтерминальной схемы измерений формула Бюттекера упрощается до

$$I_1 = \left(1/q\right)G_{1,2}\left(E_{F1} - E_{F2}\right) = -I_2 \tag{38}$$

и совпадает с двухтерминальной формулой Ландауэра (35), если проводимость $G_{1,2}$ идентифицировать как $(q^2/h)M$.

Если бы мы знали все электрохимические потенциалы E_{Fm}, могли бы воспользоваться формулой Бюттекера (37) и вычислили бы все токи I_m. Но в четырехтерминальных измерениях (рис. 9) нам не известны электрические потенциалы на контактах 1* и 2*, так что не известны и электрохимические потенциалы E^*_{F1} и E^*_{F2}. Однако, мы знаем токи I^*_1 и I^*_2 и они должны быть равны нулю (рис. 9). Если же нам известны на каждом контакте либо E_{Fm}, либо I_m, формула Бюттекера позволяет посчитать все недостающие величины.

Рассмотрим ландауэровское падение напряжения на дефекте (рис. 8) и покажем, что для слабосвязанных контактов формула Бюттекера дает те же самые результаты, что и неравновесные потенциалы E^+_F и E^-_F внутри проводника.

Потенциал, измеряемый пробным контактом (рис. 9), есть некое среднее значение потенциалов E^+_F и E^-_F (рис. 11), при этом корректное усреднение определяется конструкцией контактов. Промоделируем такое измерение путем сопоставления проводимостей g^+ и g^- двум потокам электронов I^+ и I^- (рис. 11).

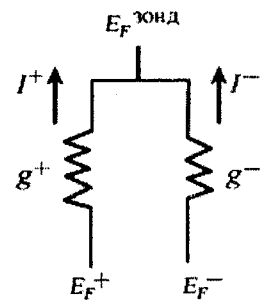


Рис. 11. К подсчету электрохимического потенциала на пробном контакте.

Полагая ток на пробном контакте нулевым, имеем

$$g^+\left(E_F^+ - E_F^{зонд}\right) + g^-\left(E_F^- - E_F^{зонд}\right) = 0, \tag{39}$$

так что

$$E_F^{зонд} = \frac{g^+}{g^+ + g^-}E_F^+ + \frac{g^-}{g^+ + g^-}E_F^-, \tag{40}$$

где весовые коэффициенты

$$\frac{g^+}{g^+ + g^-} = \alpha, \quad \frac{g^-}{g^+ + g^-} = 1 - \alpha. \tag{41}$$

Предполагая, что в измерениях используются неинвазивные пробные контакты атомных размеров, можно ожидать, что проводимости будут

практически одинаковыми, так что весовой коэффициент $\alpha \sim 50\%$ и пробный контакт меряет среднее арифметическое

$$E_F^{зонд} = \left(E_F^+ + E_F^-\right)/2. \tag{42}$$

Если контакт недостаточно мал и по разному возмущает потоки Γ^+ и Γ^-, весовой коэффициент α отличен от 50%. Если к тому же оба пробных контакта 1* и 2* в измерительной схеме (рис. 9) отличаются весовыми коэффициентами, то измеренное сопротивление дефекта окажется иным по сравнению с ландауэровским значением (34).

11.4.1. Формулы Ландауэра

Вернемся теперь к задаче нахождения профиля неравновесных потенциалов в канале с дефектом (рис. 8). Сначала сбалансируем выходящие и входящие токи (рис. 12):

$$\begin{aligned} I_R^+ &= TI_L^+ + \left(1-T\right)I_R^-, \\ I_L^- &= \left(1-T\right)I_L^+ + TI_R^-. \end{aligned} \tag{43}$$

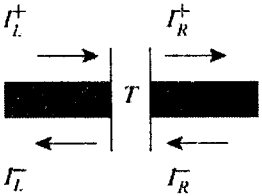


Рис. 12. Распределение токов слева и справа от дефекта.

Затем, следуя формулам (40/гл.7) и (41/гл.7) [21], от токов перейдем к функциям распределения

$$\begin{aligned} f_R^+ &= Tf_L^+ + \left(1-T\right)f_R^-, \\ f_L^- &= \left(1-T\right)f_L^+ + Tf_R^-, \end{aligned} \tag{44}$$

а затем к электрохимическим потенциалам

$$\begin{aligned} E_{FR}^+ &= TE_{FL}^+ + \left(1-T\right)E_{FR}^- = TE_{F1} + \left(1-T\right)E_{F2}, \\ E_{FL}^- &= \left(1-T\right)E_{FL}^+ + TE_{FR}^- = \left(1-T\right)E_{F1} + TE_{F2}. \end{aligned} \tag{45}$$

Упростим выкладки, выбрав потенциал на одном контакте, равным нулю, а на другом, равным единице (рис. 13).

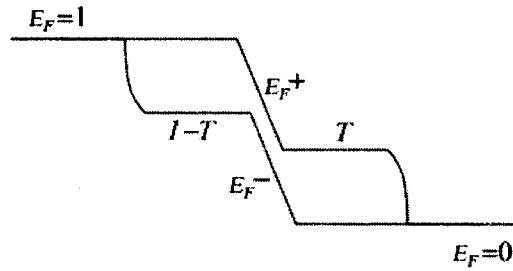


Рис. 13. Профили потенциалов E_F^+ и E_F^- на дефекте, нормированные на суммарную разность потенциалов, равную единице.

Нужную разность потенциалов всегда можно учесть путем умножения на $E_{F1} - E_{F2} = qV$. Тогда уравнения (45) упрощаются до

$$E_F^+ - E_F^- = T\left(E_{F1} - E_{F2}\right) \tag{46}$$

и вместе с выражением для тока

$$I = \frac{q}{h} M\left(E_F^+ - E_F^-\right) \tag{47}$$

дают стандартную формулу Ландауэра (35).

Для получения первой формулы Ландауэра (34) находим скачок на дефекте либо потенциала E_F^+, либо E_F^- :

$$\begin{aligned} E_{FL}^+ - E_{FR}^+ &= \left(1-T\right)\left(E_{F1} - E_{F2}\right), \\ E_{FL}^- - E_{FR}^- &= \left(1-T\right)\left(E_{F1} - E_{F2}\right), \end{aligned} \tag{48}$$

а затем делим ток на эту разность и получаем формулу (34).

Не забудем, что нас интересует измерение электрохимических потенциалов внутри проводника. Как это соотносится с измерением напряжения неинвазивным зондом сканирующего туннельного микроскопа (СТМ)? Если предположить, что такой зонд СТМ измеряет среднее значение электрохимических потенциалов E_F^+ и E_F^-, то получим график на рис. 14, сравнение которого с рис. 13 и позволяет утверждать, что неинвазивный зонд СТМ действительно измеряет среднее значение электрохимического потенциала $(E_F^+ + E_F^-)/2$.

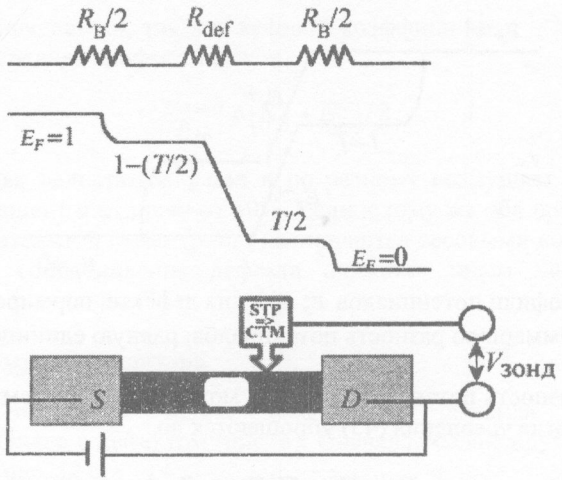


Рис. 14. Сканирующий туннельный микроскоп измеряет среднее значение электрохимического потенциала $(E_F^+ + E_F^-)/2$.

Что если зонд СТМ измеряет средневзвешенное значение E_F^+ и E_F^- с весовым коэффициентом α (40), отличным от 50% ? До тех пор пока весовой коэффициент α остается одинаковым для обоих потенциалов E_F^+ и E_F^-, скачок на дефекте (48) остается прежним

$$E_{FL}^{зонд} - E_{FR}^{зонд} = \left(1-T\right)\left(E_{F1} - E_{F2}\right), \tag{49}$$

что ведет к той же самой первой формуле Ландауэра (34). Если же весовые коэффициенты различны для E_F^+ и E_F^-, то формула (34) окажется неверной. В предельном случае, когда весовой коэффициент α равен нулю слева и единице справа (рис. 11),

$$E_{FL}^{зонд} - E_{FR}^{зонд} = \left(1-2T\right)\left(E_{F1} - E_{F2}\right), \tag{50}$$

что ведет к отрицательному сопротивлению при T > 0.5.

Концепция неравновесных потенциалов E_F^+ и E_F^- ведет к интуитивно правильным результатам и сейчас уже широко используется, однако, ее применение требует осторожности. Именно дефект вызывает скачок потенциала, а не область баллистического транспорта (рис. 14), подразумевая, что сопротивление по (34) определяется дефектом. Нужно помнить, что мы все еще рассматриваем упругие резисторы. Да, скачок напряжения IR остается, а диссипация джоулева тепла I^2R в проводнике отсутствует и происходит лишь на контактах.

11.4.2. Формула Бюттекера

Формула Бюттекера (37) содержит непосредственно измеряемые величины, оставляя за ее пределами любые вопросы, касающиеся немых внутренних переменных. Хотелось бы подчеркнуть общую применимость формулы Бюттекера, независимо от того проводник является упругим или не является таковым. Действительно, как мы сейчас увидим, для вывода формулы (37) потребуется лишь совсем немного сверх обычной линейной теории электрических цепей.

Определим для начала мультитерминальную проводимость

$$G_{m,n} \equiv -\frac{\partial I_m}{\partial\left(E_{Fn}/q\right)}, \quad m \neq n$$
$$G_{m,m} \equiv +\frac{\partial I_m}{\partial\left(E_{Fm}/q\right)}, \tag{51}$$

где естественно задаться вопросом, почему при $m \neq n$ знак выбран отрицательным в отличие от случая, когда $m = n$. Выбор знака можно проиллюстрировать на примере четырехтерминального измерения канала с одним дефектом (рис. 15).

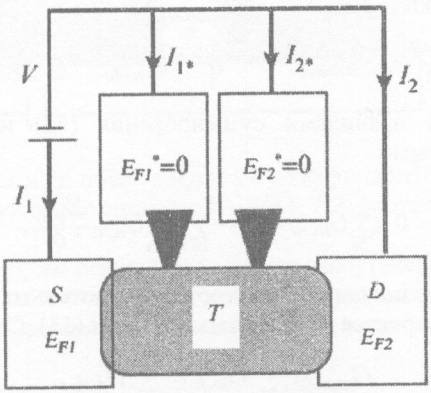


Рис. 15. Мысленный эксперимент с четырехтерминальным измерением канала с одним дефектом как на рис. 8.

Увеличение электрохимического потенциала E_{F1} ведет к увеличению тока, входящего в контакт 1 (положительный ток), и к оттоку от всех других контактов (отрицательные токи). Знаки в (51) выбраны таким образом, чтобы

коэффициенты проводимости $G_{m,n}$ всегда были положительными, как того интуитивно требует смысл понятия проводимости.

Теперь ток (37) перепишем через коэффициенты проводимости

$$I_m = G_{m,m}\frac{E_{Fm}}{q} - \sum_{n\neq m} G_{m,n}, \tag{52}$$

при этом коэффициенты должны удовлетворять следующим двум правилам суммируемости.

Во-первых, если все электрохимические потенциалы одинаковы (состояние равновесия), все токи по (52) должны быть равны нулю, что влечет за собой правило сумм

$$G_{m,m} = \sum_{n\neq m} G_{m,n}. \tag{53}$$

Во-вторых, при произвольных значениях электрохимических потенциалов сумма всех токов должна быть равна нулю, что приводит ко второму правилу сумм

$$G_{m,m} = \sum_{n\neq m} G_{n,m},$$

доказательство которого не столь очевидно, как это имело место с первым правилом сумм, и получить его можно опираясь на уравнение (52). Сначала просуммируем все токи (52)

$$\sum_m I_m = 0 = \sum_m G_{m,m}\frac{E_{Fm}}{q} - \sum_m \sum_{n\neq m} G_{m,n}\frac{E_{Fn}}{q}, \tag{54}$$

затем воспользуемся правилами суммирования (53) и поменяем местами индексы в двойной сумме

$$0 = \sum_m G_{m,m}\frac{E_{Fm}}{q} - \sum_m \sum_{n\neq m} G_{n,m}\frac{E_{Fm}}{q}, \tag{55}$$

что при произвольных значениях электрохимических потенциалов справедливо только, если удовлетворяется второе правило сумм (53). Окончательно,

$$G_{m,m} = \sum_{n\neq m} G_{m,n} = \sum_{n\neq m} G_{n,m}. \tag{56}$$

Воспользовавшись правилами сумм (56), перепишем первое слагаемое в (52) в виде суммы и в результате получим знаменитую формулу Бюттикера (37). Обратим внимание на то, что в формуле Бюттикера суммирование нет нужды ограничивать по $n \neq m$, поскольку слагаемое при $n = m$ всегда равно нулю.

Иногда формулу Бюттекера (37) удобнее записать через так называемые коэффициенты отклика [33]

$$I_m = \sum_n g_{m,n} \frac{E_{Fn}}{q}, \tag{57}$$

которые определяются следующим образом:

$$\begin{aligned} g_{m,n} &\equiv -G_{m,n}, \quad m \neq n \\ g_{m,m} &\equiv G_{m,m} \end{aligned} \tag{58}$$

с правилами суммирования

$$\sum_n g_{m,n} = \sum_n g_{n,m} = 0. \tag{59}$$

Применим подход Бюттекера к ситуации на рис. 14, где утверждается, что два неинвазивных зонда, расположенные до и после дефекта с коэффициентом пропускания T, генерируют потенциалы соответственно $1 - T/2$ и $T/2$ (рис. 16).

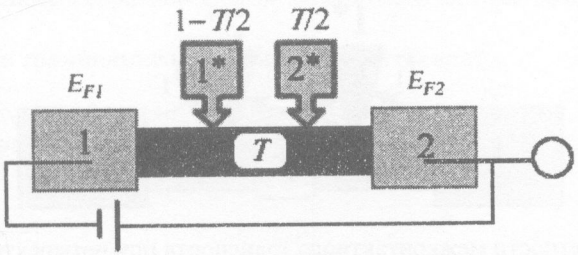


Рис. 16. На зондах до и после дефекта с коэффициентом пропускания T генерируются потенциалы $1 - T/2$ и $T/2$, соответственно.

Получим ли мы те же результаты методом Бюттекера? Имея в виду, что в этом измерении речь идет о четырех токах и четырех потенциалах на контактах 1, 2, 1* и 2*, запишем уравнение (57) в матричном виде

$$\begin{Bmatrix} I_1 \\ I_2 \\ I_{1^*} \\ I_{2^*} \end{Bmatrix} = \frac{Mq}{h} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{Bmatrix} E_{F1} \\ E_{F2} \\ E_{F1^*} \\ E_{F2^*} \end{Bmatrix}, \tag{60}$$

где матрицы A, B, C и D имеют форму (2×2).

Поскольку

$$\begin{Bmatrix} I_{1*} \\ I_{2*} \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}, \tag{61}$$

то можно записать, что

$$\begin{Bmatrix} E_{F1*} \\ E_{F2*} \end{Bmatrix} = -D^{-1}C \begin{Bmatrix} E_{F1} \\ E_{F2} \end{Bmatrix} \tag{62}$$

с матрицами C и D в виде

$$\begin{bmatrix} C & D \end{bmatrix} = \begin{bmatrix} -t_1 & -t_2 & r & 0 \\ -t_2' & -t_1' & 0 & r' \end{bmatrix}, \tag{63}$$

где элементы матрицы C интерпретируются как вероятности межконтактного транспорта 1 на 1*, 2 на 1*, 2 на 2* и 1 на 2* (рис. 17), причем предполагается, что контакты 1* и 2* слабо связаны и транспортом между ними можно пренебречь.

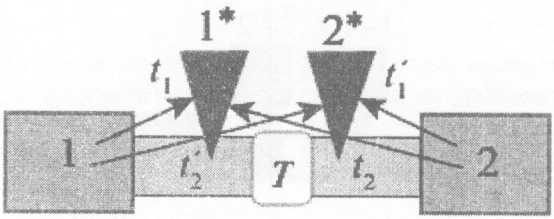


Рис. 17. Вероятности межконтактного транспорта при четырехтерминальном измерении проводимости канала с одним дефектом с коэффициентом пропускания T.

Правила сумм (59) дают

$$\begin{aligned} r &= t_1 + t_2, \\ r' &= t_1' + t_2'. \end{aligned} \tag{64}$$

Из уравнений (62) – (64) имеем

$$\begin{aligned} E_{F1*} &= \frac{t_1}{t_1 + t_2} E_{F1} + \frac{t_2}{t_1 + t_2} E_{F2}, \\ E_{F2*} &= \frac{t_2'}{t_1' + t_2'} E_{F1} + \frac{t_1'}{t_1' + t_2'} E_{F2}. \end{aligned} \tag{65}$$

Учтем также, что вероятность t_1 (рис. 17) может быть представлена как сумма вероятности τ прямого переноса 1 → 1* и вероятности $(1 - T)\cdot\tau$

отразиться от дефекта, а затем достичь контакта 1*, и аналогичные соображения для вероятности t_2:

$$t_1 = \tau + (1-T)\tau, \qquad t_2 = T\tau, \tag{66}$$

поскольку на пути электрона с контакта 2 сначала нужно преодолеть дефект (вероятность T), а затем попасть на контакт 1* (вероятность τ). По тем же соображениям $t_1 = t_1'$ и $t_2 = t_2'$ (рис. 17).

Пользуясь уравнениями (66) и положив $E_{F1} = 1$ и $E_{F2} = 0$, из уравнений (65) получаем ожидаемые значения для потенциалов на зондах 1* и 2*:

$$E_{F1^*} = 1 - (T/2), \qquad E_{F2^*} = T/2, \tag{67}$$

в чем можно было бы и не сомневаться, поскольку в формуле Бюттекера фигурируют только характеристики контактов и фактически свернуты все тонкости, связанные с неравновесными электрохимическими потенциалами.

11.4.2.1. Не тривиально ли уравнение Бюттекера?

В конце концов, любую сложную электрическую схему всегда можно представить эквивалентной схемой эффективных резисторов, как например на рис. 18 для трех терминалов.

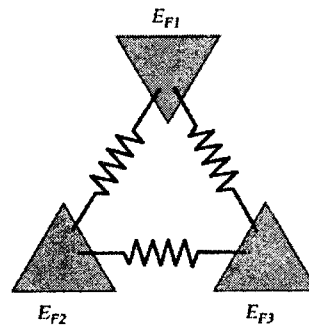


Рис. 18. Уравнение Бюттекера эквивалентно сети эффективных резисторов только в том случае, если все проводимости обратимы ($G_{m,n} = G_{n,m}$).

Даст ли нам стандартная теория электрических цепей такой же результат как и уравнение Бюттекера (37)? Ответ на этот вопрос несомненно положительный, однако, только для нормальных взаимных электрических

цепей, у которых проводимости между всеми терминалами полностью обратимы

$$G_{m\leftarrow n} = G_{n\leftarrow m}, \tag{68}$$

другими словами, электроны одинаково легко транспортируются в обе стороны для каждой пары терминалов (m, n).

Область применимости же уравнения Бюттекера (37) намного более широкая, чем нормальные взаимные электрические цепи, например, она применима для проводников в магнитном поле, когда не имеет места взаимная проводимость:

$$G_{m\leftarrow n} \neq G_{n\leftarrow m}. \tag{69}$$

Для таких ситуаций обосновать применимость уравнения Бюттекера (37) не так то просто. Если в уравнении Бюттекера (37) проводимость $G_{m,n}$ поменять на $G_{n,m}$, то мы просто получим ошибочную формулу, применяя которую к многотерминальным невзаимным цепям получим ошибочные результаты, не соответствующие формуле Бюттекера (37).

11.5. Измерение потенциала Холла

Обратимся, наконец, к вычислению холловского напряжения с использованием формулы Бюттекера (37) при четырехтерминальном эксперименте, отличающемся от рассмотренного выше (рис. 17) размещением пробных зондов 1* и 2* поперек проводника (рис. 19).

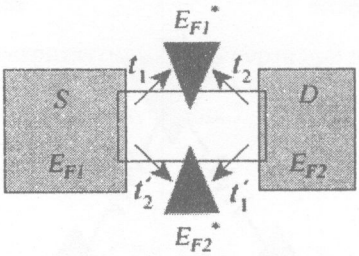


Рис. 19. Вероятности межконтактного транспорта при четырехтерминальном измерении холловского напряжения.

В отсутствии магнитного поля разность зондовых потенциалов нулевая. При включении магнитного поля перпендикулярно к поверхности проводника (плоскости рисунка) электроны, движущиеся слева направо, отклоняются вдоль или против направления движения к боковым краям проводника (рис. 1), создавая различие между зондовыми потенциалами E_{F1^*} и E_{F2^*}, которое мы и хотим вычислить в зависимости от напряжения, создающего ток в проводнике.

Уравнения для зондовых потенциалов остаются прежними (65), если вероятностям межконтактного транспорта придать смысл, показанный на рис. 19. Вычисление этих вероятностей либо численно, либо в полуклассическом приближении приводит к холловскому напряжению V_H

$$-qV_H = E_{F1^*} - E_{F2^*} \tag{70}$$

в согласии с уже полученным ранее уравнением (22).

Уравнение (22) было получено опираясь на транспортное уравнение Больцмана, но его можно получить и из уравнений (65), если переписать их применительно к тонкому проводнику ширины где-нибудь посередине длинного проводника меньшей, чем средняя длина свободного пробега, с электронами, движущимися слева, где потенциал E_F^+, и справа, где потенциал E_F^-, для потенциалов на середине проводника на его боковых сторонах в точках, имитирующих зонды 1* и 2* (рис. 20), а именно:

$$\begin{aligned} E_F(x+dx) &= \frac{t_1}{t_1+t_2}E_F^+ + \frac{t_2}{t_1+t_2}E_F^-, \\ E_F(x) &= \frac{t_2'}{t_1'+t_2'}E_F^+ + \frac{t_1'}{t_1'+t_2'}E_F^-. \end{aligned} \tag{71}$$

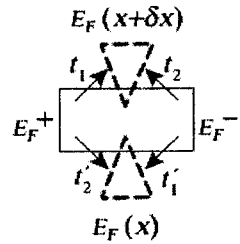


Рис. 20. К выводу уравнения (22) из вероятностей межконтактного транспорта при четырехтерминальном измерении холловского напряжения.

Для вычисления вероятностей в уравнениях (71) заметим, что в отсутствии магнитного поля электроны со скоростями, лежащими в пределах угла θ (рис. 5)

$$0 < \theta < +\pi/2,$$

движутся вверх, а лежащими в пределах

$$-\pi/2 < \theta < 0,$$

движутся вниз (рис. 21).

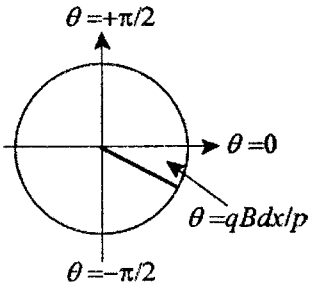


Рис. 21. К вычислению угла заворота электрона под действием силы Лоренца.

Магнитное поле закручивает траекторию электронов наверх на угол, равный произведению угловой скорости (14) на время dx/v, затрачиваемое на преодоление расстояния dx, а именно:

$$\frac{qvB}{p}\frac{dx}{v}=\frac{qBdx}{p}. \tag{72}$$

Тогда вероятности межконтактного транспорта будут пропорциональны углам

$$\begin{aligned} t_1 &= t_1' \sim \frac{\pi}{2}+\frac{qB}{p}dx, \\ t_2 &= t_2' \sim \frac{\pi}{2}-\frac{qB}{p}dx, \end{aligned} \tag{73}$$

и подстановка их в уравнения (71) дает

$$E_F\left(x+dx\right)-E_F\left(x\right)=\frac{2qBdx}{\pi p}\left(E_F^+-E_F^-\right), \tag{74}$$

так что

$$\frac{dE_F}{dx}=\frac{2qB}{\pi p}\delta E_F \tag{75}$$

в согласии с уравнением (32), а стало быть и с (22).

11.6. Квантовый эффект Холла и краевые состояния в графене

Одним из наиболее эффектных приложений подхода Бюттекера служит интерпретация режима квантового эффекта Холла, когда магнитное поле настолько велико, что электроны, движущиеся от истока, не достигают зонда 2*, а электроны, движущиеся от стока, не достигают зонда 1* (рис. 19). В результате

$$\frac{V_H}{\left(E_{F1}-E_{F2}\right)/q}=\frac{t_1-t_2'}{t_1+t_2'}=1, \tag{76}$$

поскольку $t_2 = t_2' = 0$, так что холловское напряжение оказывается равным разности потенциалов на концах проводника, а холловское сопротивление равным обычному двухтерминальному сопротивлению.

Интересно, что в этом режиме сопротивление дается выражением (4) с высочайшей степенью точности [4]. Это можно представить себе так, как будто бы мы имеем идеальный баллистический проводник с одним только граничным сопротивлением. Длина проводников в экспериментах нередко достигает нескольких сот микрометров, подобная баллистичность просто удивительна, что и было отмечено Нобелевской премией в 1985 году (Клаус фон Клитцинг).

Обратимся к графену. В главе 10 сформулирован метод НРФГ в матричной формулировке применительно к решению модельных транспортных задач наноэлектроники, в частности, для графеновых лент были приведены результаты расчета методом НРФГ коэффициента прохождения и плотности состояний для двух конфигураций ленточных границ – зигзагообразной и креслоподобной в приближении сильной связи в ортогональном базисе с параметрическим учетом взаимодействия лишь соседних атомов через кулоновский ε и резонансный t_0 интегралы (хюккелевское приближение) в полном согласии с результатами вычислений с более точными модельными гамильтонианами. Сейчас мы учтем внешнее магнитное поле для графеновых лент в этом же приближении и приведем результаты расчета коэффициента прохождения и плотности состояний в присутствии магнитного поля, а также квантового эффекта Холла и проводимости графеновых лент в магнитном поле [34].

Изучаемая нами система представляет собой длинный 2D проводник с двумя пробными зондами по обеим сторонам проводника (рис. 22) для измерения поперечного холловского напряжения, возникающего в присутствии магнитного поля, направленного перпендикулярно к поверхности 2D проводника.

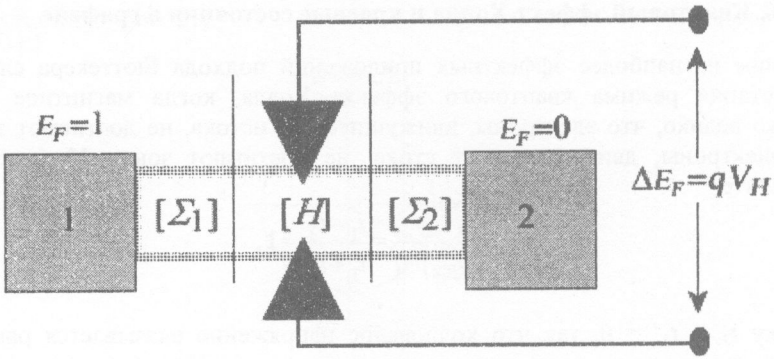


Рис. 22. К измерению холловского напряжения через разность химпотенциалов, создаваемую между поперечными зондами при включении перпендикулярного к проводнику магнитного поля.

Вначале учтем магнитное поле в гамильтониане проводника H и контактных матрицах Σ_1 и Σ_2 метода НРФГ.

Энергия частицы с зарядом $-q$ в поле векторного потенциала $\vec{A}$ дается выражением

$$E(\vec{r},\vec{p}) = \frac{\left(\vec{p}+q\vec{A}\right)\cdot\left(\vec{p}+q\vec{A}\right)}{2m} + U(\vec{r}). \tag{77}$$

Для нахождения параметров хюккелевского гамильтониана с учетом магнитного поля рассмотрим однородную решетку атомов с постоянным значением E_C и постоянным векторным потенциалом. Сначала рассмотрим 1D проводник, вытянутый вдоль оси x

$$E(p_x) = E_C + \frac{\left(p_x+qA_x\right)\left(p_x+qA_x\right)}{2m}, \tag{78}$$

так что дисперсионное соотношение имеет вид

$$E(k_x) = E_C + \frac{\left(\hbar k_x+qA_x\right)\left(\hbar k_x+qA_x\right)}{2m} \tag{79}$$

и может быть аппроксимировано функцией косинуса

$$E(k_x) = \varepsilon + 2t\cos\left(k_x a + \frac{qA_x a}{\hbar}\right) \tag{80}$$

с хюккелевскими параметрами ε и t [19]

$$\varepsilon = E_C - 2t$$
$$t = -\hbar^2 / 2ma^2 = -t_0 \quad , \tag{81}$$

где a – постоянная решетки. Отличие от однородной решетки без учета поля векторного потенциала (§ 10.9.2) [19] состоит лишь в появлении фазового множителя $\exp(\pm iqA_x a/\hbar)$ у резонансных интегралов (рис. 23).

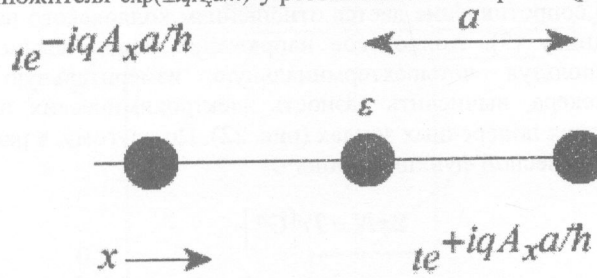


Рис. 23. Однородный 1D проводник в поле векторного потенциала.

Для однородной решетки 2D проводника фазовые множители с фазами

$$\varphi_x = \frac{qA_x a}{\hbar}, \quad \varphi_y = \frac{qA_y a}{\hbar} \tag{82}$$

появляются у резонансных интегралов в обоих направлениях плоскости проводника (рис. 24).

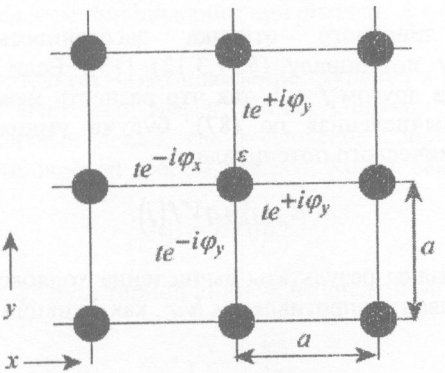


Рис. 24. Учет магнитного поля, направленного вдоль оси z, сводится к учету фазовых множителей у резонансных интегралов хюккелевской решетки.

Магнитное поле

$$\vec{B} = \vec{\nabla} \times \vec{A}, \tag{83}$$

так что интересующее нас магнитное поле, перпендикулярное плоскости 2D проводника,

$$B_z = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}. \tag{84}$$

Холловское сопротивление дается отношением холловского напряжения к току по проводнику (9). Холловское напряжение можно вычислить двумя способами. Используя четырехтерминальную измерительную схему и уравнение Бюттекера, вычислить разность электрохимических потенциалов, создаваемых на двух поперечных зондах (рис. 22). По другому, в рамках метода НРФГ след неравновесной функции Грина G^n

$$2\pi N = Tr\left[G^n\right] \tag{85}$$

пропорционален числу электронов, а спектральная функция A

$$2\pi D(E) = A \tag{86}$$

пропорциональна плотности состояний, так что заселенность состояния j есть отношение соответствующих диагональный элементов функции Грина и спектральной функции

$$f(j) = \frac{G^n(j,j)}{A(j,j)}. \tag{87}$$

В области линейного отклика заселенность пропорциональна электрохимическому потенциалу (§ 1.3.12) [15]. Если на одном контакте положить $f = 0$, а на другом $f = 1$, так что разность между ними есть qV, то заселенность $f(j)$, вычисленная по (87), будучи умноженной на qV, даст значение электрохимического потенциала

$$E_F(j) = qVf(j). \tag{88}$$

На рис. 25 показаны результаты вычисления холловского сопротивления, нормированного на квант сопротивления h/q^2, как функции магнитного поля.

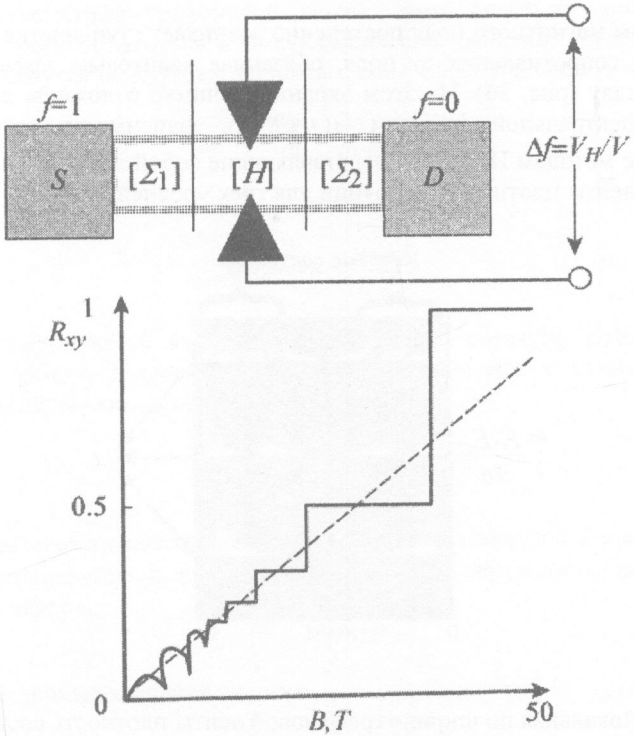


Рис. 25. Нормированное холловское сопротивление в зависимости от магнитного поля для графеновой ленты шириной $W = 26a = 65нм$ в хюккелевском приближении при энергии $E = t_0$ [19].

Замечательный результат – это появление ступенек в сильных магнитных полях, известный как квантовый эффект Холла. Однако, сначала обратимся к области слабых магнитных полей, где холловское сопротивление, согласно стандартной полуэмпирической теории эффекта Холла, зависит от магнитного поля линейно (9):

$$R_H = B/qn, \tag{89}$$

что и показано на рис. 25 пунктирной линией, вычисленной для

$$n = \frac{N}{LW} = \frac{k^2}{4\pi} \tag{90}$$

при $E = t_0 = 2.7эВ$ и $ka = \pi/3$ с $a = 2.5нм$ [19].

С ростом магнитного поля постепенно возникает ступенчатая зависимость холловского сопротивления от поля, обязанная квантовым эффектам в виде уровней Ландау (рис. 26). На этом экранном снимке отложены диагональные элементы спектральной функции $A(j,j;E)$ в зависимости от энергии E, вычисленные методом НРФГ, и представляющие собой локальную по ширине j графеновой ленты плотность состояний для трех уровней Ландау.

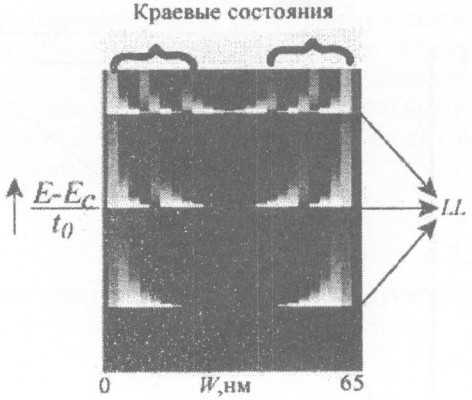


Рис. 26. Локальная по ширине графеновой ленты плотность состояний в зависимости от энергии уровней Ландау при $B = 20$ T.

Энергия уровней Ландау

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega_c, \tag{91}$$

где n – целое число, ω_c – циклотронная частота (15). Эквидистантные уровни Ландау располагаются друг относительно друга с инкрементом

$$\hbar\omega_c = \frac{\hbar qB}{m} = \frac{2qBa^2}{\hbar}t_0, \tag{92}$$

выражение для которого получено с использованием (15) и (81). При использованных в расчете значениях B = 20 T и a = 2.5 *нм* инкремент получается равным $\hbar\omega_c \approx 0.37t_0$ в неплохом согласии с результатами расчета методом НРФГ (рис. 26).

Выражение для энергии уровней Ландау (91) есть чисто квантовый результат аналитического решения уравнения Шредингера с учетом векторного потенциала, лежащего также в основе численного решения методом НРФГ для

конкретной геометрии графеновой ленты. Этот результат можно понять качественно, если обратиться к полуклассической модели, согласно которой электроны движутся по круговым орбитам, совершая один оборот за время

$$t_c = \frac{2\pi}{\omega_c} = \frac{2\pi p}{qvB}, \tag{93}$$

так что длина окружности орбиты с радиусом r_c

$$2\pi r_c = vt_c = \frac{2\pi p}{qB}. \tag{94}$$

Если теперь учесть квантовое ограничение, согласно которому длина окружности орбиты должна быть такой, чтобы на ней укладывалось целое число K де-бройлевских длин волн h/p, то получим

$$\frac{2\pi p}{qB} = K \cdot h/p. \tag{95}$$

Полуклассически электрон может иметь любую энергию $E = p^2/2m$, однако, квантовое ограничение в виде равенства длины окружности целому числу K длин волн ведет к

$$p^2 = K \cdot hqB, \tag{96}$$

так что разрешенные значения энергии

$$E = K\frac{\hbar qB}{2m} = K\frac{\hbar\omega_c}{2}, \tag{97}$$

что не совсем то же самое, что и правильный результат (91), но качественно вполне приемлемо.

Если число краевых состояний обозначить K_{es}, то результирующий ток

$$I_H = \frac{q^2}{h} V \cdot K_{es}, \tag{98}$$

поскольку холловское напряжение V_H просто равно напряжению V, приложенному к концам проводника, так как один край проводника находится в равновесии с истоком, а другой – со стоком.

Это ведет к квантованию холловского сопротивления

$$R_H = \frac{h}{q^2}\frac{1}{K_{es}}, \tag{99}$$

порождая у нормированного холловского сопротивления плато при ¼, ⅓, ½ и 1 (рис. 25) по мере того как магнитное поле генерирует уровни Ландау, изменяя число краевых состояний K_a при энергии $E = t_0$ с 4-х до 3-х, до 2-х и до 1-го (рис. 26).

Что касается распределения электрического тока в 2D проводнике в режиме наблюдения квантового эффекта Холла, то часто используют [35, 36] полуклассическую модель движения электронов по так называемым [37] открытым орбитам (skipping orbits) (рис. 27). В результате образуются краевые токи, обеспечивающие исключительно высокую баллистическую проводимость.

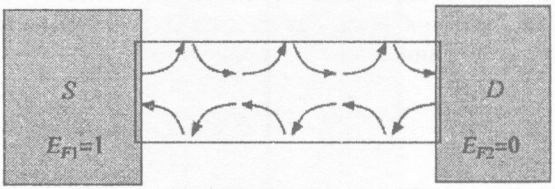


Рис. 27. Предполагают, что электроны в 2D проводнике в сильном магнитном поле движутся по открытым орбитам, создавая компактные трассы краевых токов, как бы локализованных в квантовом волноводе [37].

Ранее уже приводились результаты расчета методом НРФГ коэффициента прохождения и плотности состояний графеновых лент для двух конфигураций ленточных границ – зигзагообразной ZGNR / Zigzag Graphene NanoRibbon и креслоподобной AGNR / Armchair GNR в отсутствии магнитного поля [19]. Влияние сильного магнитного поля показано на рис. 28.

Уровни Ландау LL/Landau Levels в сильных магнитных полях образуются в таких 2D проводниках, ширина которых достаточна для образования циклотронных орбит. Чем больше ширина проводника, тем требуется меньшее магнитное поле для генерации первого уровня Ландау. В графене энергии уровней Ландау

$$E_n^{LL} = \pm v_0 \sqrt{2q\hbar nB}, \qquad (100)$$

где фермиевская скорость принималась равной $v_0 = 10^6$ *м/с* [38]. Согласно (100), первый и второй уровни Ландау при $B = 20$ *Т* возникают при $E_1^{LL} = 0.15$ *эв* и $E_2^{LL} = 0.23$ *эв*. Есть также уровень Ландау при $E = 0$, это нулевой уровень Ландау E_0^{LL} [39 – 41]. Результаты расчета методом НРФГ (рис. 28) полностью согласуются с приведенными оценками.

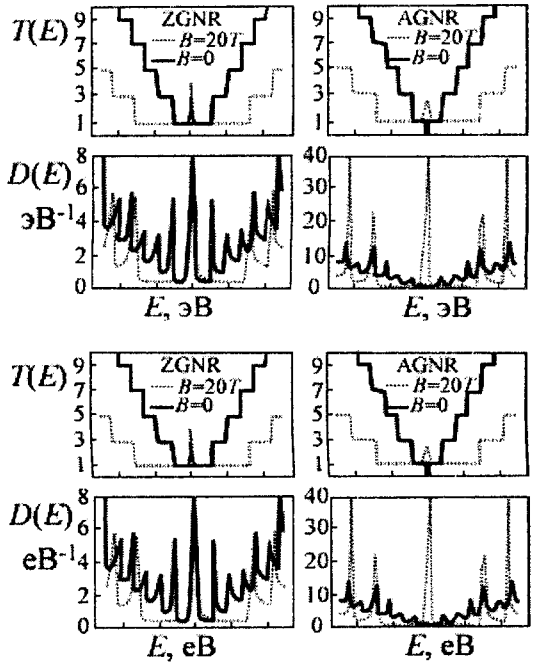


Рис. 28. Влияние сильного магнитного поля $B = 20\ T$ на коэффициент прохождения $T(E)$ и плотность состояний $D(E)$ графеновых лент шириной $W = 53$ *нм* для двух конфигураций границ – ZGNR и AGNR [34].

11.7. Формула Кубо

Точно так же как тепло передается от более нагретого конца проводника к менее нагретому, а различие в уровнях жидкости ведет к её перетоку, так и электрический ток течет от контакта с большим значением электрохимического потенциала к контакту с меньшим его значением (глава 1). Под направлением тока, вопреки исторически сложившейся договоренности, мы далее понимаем направление движения электронов. Неоднократно подчеркивалась неравновесная природа образования электрического тока. Общее выражение для тока (глава 1)

$$I = q\int_{-\infty}^{+\infty} dE \frac{D(E)}{2t(E)}\left[f_1(E) - f_2(E)\right] \qquad (101)$$

справедливо для любой по величине разности потенциалов, подаваемой на концы проводника. В этом выражении $D(E)$ – плотность состояний в

проводнике, $t(E)$ – время пролета электрона от истока S до стока D, $f_i(E)$– фермиевские функции контактов $(i=1,2)$. В случае интересующего нас низковольтного предела проводимость

$$G = q^2 \int_{-\infty}^{+\infty} dE \left(-\frac{\partial f_0}{\partial E} \right) \frac{D(E)}{2t(E)}. \qquad (102)$$

Хотя это выражение и было получено из общего неравновесного выражения (101), полезно отметить, что проводимость в режиме линейного отклика является в действительности равновесным свойством. Имеет место фундаментальная флуктуационно-диссипационная теорема, связывающая проводимость в квазиравновесном режиме с токовыми флуктуациями в состоянии равновесия, когда разность потенциалов на концах проводника нулевая.

При ненулевой температуре тепловые флуктуации являются одной из причин зашумленности проводника с током, что приводит, в частности, к флуктуации заселенности состояний проводника. Заселенность состояний проводника g обычно характеризуется числом, которое может принимать значение 1 или 0. Термодинамически среднее значение заселенности $\langle g \rangle$ определяется функцией Ферми

$$f(E) = 1/[e^{(E-E_F)/kT} + 1] \equiv \langle g \rangle.$$

В равновесном состоянии вероятность того, что состояние не заселено, дается в среднем значением $1-f$, а вероятность того, что состояние заселено в среднем равно f. Флуктуации заселенности от их средних значений характеризуются выражением

$$\left(g - \langle g \rangle\right)^2 = g^2 - 2g\langle g \rangle + \langle g \rangle^2.$$

Для фермиевских систем $g^2 = g$, так что флуктуации заселенности (в состоянии равновесия) от их термодинамически средних значений даются выражением

$$\left\langle \left(g - \langle g \rangle\right)^2 \right\rangle = f(1-f).$$

Эти среднеквадратичные флуктуации исчезают в нулевом температурном пределе. При высоких температурах и достаточно больших энергиях фермиевская функция намного меньше единицы, так что множитель $1-f$ можно заменить на 1: флуктуации подчиняются статистике Максвелла – Больцмана.

Флуктуации заселенности порождают токовые флуктуации (в состоянии равновесия) во внешней цепи, которые через флуктуационно-диссипационную теорему, иначе формулу Кубо, связаны с проводимостью проводника. Таким образом, исследование токових флуктуаций в состоянии равновесия дает нам ту же самую информацию, что и исследование проводимости. Обсудим этот вопрос подробнее.

Рассмотрим ситуацию, когда проводник не находится под напряжением, электрохимический потенциал на контактах S и D один и тот же E_{F0} (рис. 29).

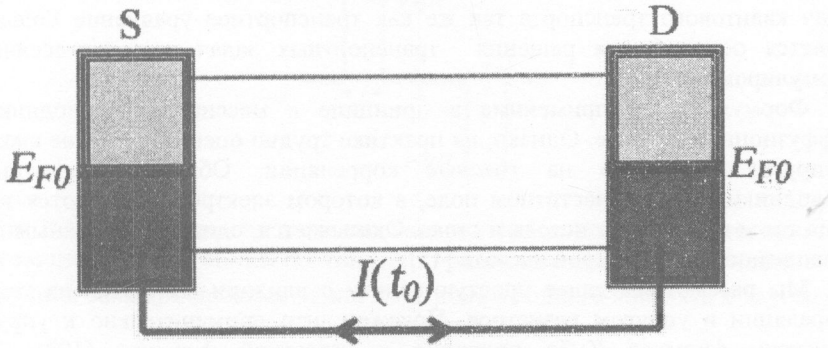


Рис. 29. В состоянии равновесия электрохимический потенциал один и тот же E_{F0} на обоих контактах. Суммарный ток нулевой, однако, имеют место взаимно компенсирующиеся токи между контактами.

Время от времени некоторое число электронов проходит по проводнику от S к D и такое же число электронов возвращается от D к S так, что

$$\left\langle I(t_0)\right\rangle_{eq} = 0, \tag{103}$$

где скобки означают либо усреднение по ансамблю многих одинаковых проводников, либо усреднение по времени в течение времени пролета t_0.

Однако, вычисления показывают, что токовый коррелятор

$$C_I = \int_{-\infty}^{+\infty} d\tau \left\langle I(t_0+\tau)I(t_0)\right\rangle_{eq} \tag{104}$$

отличен от нуля даже в состоянии равновесия, и формула Кубо увязывает этот коррелятор с проводимостью в режиме линейного отклика [42, 43]:

$$G = \frac{C_I}{2kT} = \frac{1}{2kT}\int_{-\infty}^{+\infty} d\tau \left\langle I(t_0+\tau)I(t_0)\right\rangle_{eq}. \tag{105}$$

Формула Кубо позволяет найти проводимость путем вычисления токовых корреляций методами равновесной статистической механики, которые разработаны намного лучше по сравнению с методами неравновесной статистической механики. Это действительно сильный результат. Еще до триумфального появления мезофизики в конце 80-х годов формула Кубо была фактически единственным подходом в моделировании квантового транспорта в электронике. Однако, ее применение ограничено транспортными задачами в режиме линейного отклика. Позже был окончательно сформулирован метод НРФГ, который и стал основным инструментом рассмотрения неравновесных задач квантового транспорта так же как транспортное уравнение Больцмана является основой для решения транспортных задач в полуклассической формулировке [44].

Формула Кубо применима в принципе к массивным проводникам в диффузионном режиме. Однако, на практике трудно оценить влияние сложных неупругих процессов на токовые корреляции. Обычно рассматривают проводники в высокочастотном поле, в котором электроны дергаются туда и сюда так и не достигая истока и стока. Оказывается, однако, возможным найти проводимость при нулевой частоте [45].

Мы рассмотрим далее простую задачу о влиянии контактов на токовые корреляции в упругом резисторе. Покажем, что применительно к упругому резистору формула Кубо приводит к известной формуле (102). Затем конспективно обсудим соотношения Онзагера.

11.7.1. Формула Кубо для упругого резистора

Как и ранее в концепции «снизу – вверх» (глава 1), рассмотрим одноуровневый резистор с одним и тем же электрохимическим потенциалом E_{F0}, а стало быть и с одной и той же фермиевской функцией $f_0(E)$ на контактах (рис. 30).

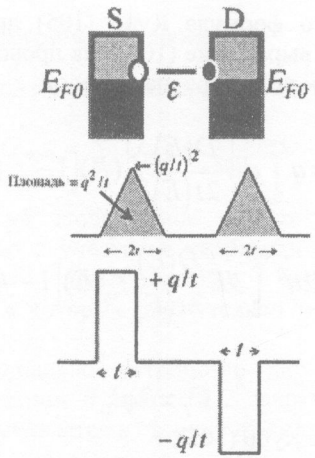


Рис. 30. При равновесии с одинаковым значением электрохимического потенциала E_{F0} на контактах тока нет. Имеют место, однако, случайные импульсы тока при проскакивании электрона от одного контакта к другому и наоборот.

Происходят случайные положительные и отрицательные импульсы тока, вызванные проскакиванием электрона от S к D и наоборот, соответственно. Положительный ток в среднем одинаков с отрицательным током, и этот равновесный ток I_0 можно выразить через время пролета (199/гл.1), а именно:

$$I_0 = \frac{q}{t} f_0(\varepsilon)\left[1 - f_0(\varepsilon)\right], \tag{106}$$

где произведение $f_0(\varepsilon)[1 - f_0(\varepsilon)]$ есть вероятность того, что на контакте S есть электрон, готовый к переходу на D, а на контакте D нет электрона, готового к переходу на S. Корреляция электронных переходов имеет место, поскольку предполагается, что переходы носят стохастический характер.

Подынтегральное выражение коррелятора (104) выглядит как последовательность треугольных импульсов (рис. 30) каждый площадью q^2/t, так что

$$C_I = \frac{2q^2}{t} f_0(\varepsilon)\left[1 - f_0(\varepsilon)\right], \tag{107}$$

где появление двойки вызвано тем, что ток (106) учитывает только положительные импульсы в то время как вклад в коррелятор дают также и отрицательные импульсы.

Сейчас покажем, что формула Кубо (105) применительно к упругому резистору дает известное выражение (102) для проводимости. Обобщая модель одноуровневого упругого резистора на таковой с произвольной плотностью состояний, имеем:

$$I_0 = q\int_{-\infty}^{+\infty} dE \frac{D(E)}{2t(E)} f_0(E)\left[1-f_0(E)\right], \tag{108}$$

$$C_I = 2q^2\int_{-\infty}^{+\infty} dE \frac{D(E)}{2t(E)} f_0(E)\left[1-f_0(E)\right], \tag{109}$$

так что

$$C_I = 2qI_0. \tag{110}$$

Согласно же формуле Кубо (105),

$$G = \frac{C_I}{2kT} = q^2\int_{-\infty}^{+\infty} dE \frac{D(E)}{2t(E)} \frac{f_0(E)\left[1-f_0(E)\right]}{kT}, \tag{111}$$

что сразу дает искомое выражение для проводимости (102), поскольку (глава 1)

$$\frac{f_0(E)\left[1-f_0(E)\right]}{kT} = \left(-\frac{\partial f_0}{\partial E}\right). \tag{112}$$

Равенство (112) играет ключевую роль в выводе выражения для проводимости (102). Ведь элементарные соображения показывают, что проводимость пропорциональна

$$\frac{f_1 - f_2}{E_{F1} - E_{F2}} \cong -\frac{\partial f_0}{\partial E}, \tag{113}$$

тогда как учет токовых корреляций по формуле Кубо приводит к проводимости, пропорциональной $f_0(E)\left[1-f_0(E)\right]/kT$.

Итак, изучение токовых флуктуаций (в состоянии равновесия) эквивалентно исследованию проводимости в режиме линейного отклика. Это не так в случае дробового шума в электрических цепях, обусловленного дискретностью заряда электронов как носителей электрического тока. Подобные тонкие корреляционные эффекты в упругих резисторах в связи с развитием мезоэлектроники интенсивно изучаются [46].

11.8. Невзаимные электрические цепи

В электрических цепях с взаимной проводимостью имеет место правило сумм (56):

$$G_{m,m} = \sum_{n \neq m} G_{m,n} = \sum_{n \neq m} G_{n,m} \,. \tag{114}$$

Обратим внимание на любопытное и полезное свойство цепей с невзаимной проводимостью с участием, например, проводников с краевыми токами, обеспечивающих исключительно высокую баллистическую проводимость, электроны в которых движутся по так называемым открытым орбитам [37].

Рассмотрим двухтерминальное устройство (рис. 31) с такой ориентацией магнитного поля, что электрон с контакта 2 закручивается на контакт 1, а электрон с контакта 1 закручивается в сторону от контакта 2.

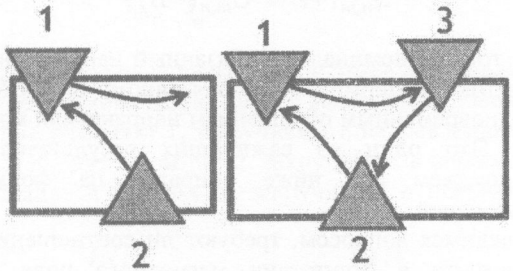


Рис. 31. Магнитное поле ориентировано таким образом, что электрон с контакта 2 закручивается на контакт 1, а электрон с контакта 1 закручивается от контакта 2. Возможна ли невзаимность проводимостей $G_{1,2} \neq G_{2,1}$? Да, возможна, если число терминалов больше двух, но не для двухтерминального устройства.

Для двухтерминального устройства невзаимность проводимостей $G_{1,2} \neq G_{2,1}$ невозможна, поскольку согласно правилу сумм (114)

$$G_{1,1} = G_{1,2} = G_{2,1} \,. \tag{115}$$

Однако, для трехтерминального устройства уже имеем

$$G_{1,1} = G_{1,2} + G_{1,3} = G_{2,1} + G_{3,1} \,, \tag{116}$$

и таким образом снимается требование взаимности проводимости. Эффекты такой невзаимной проводимости с очевидностью обнаружены на примере краевых состояний в режиме квантового эффекта Холла, к чему мы еще вернемся ниже.

Концепция краевых состояний, обеспечивающих однонаправленный баллистический транспорт на макроскопические расстояния, представляется исключительно интересной несмотря на то, что наблюдение таких состояний ограничено низкими температурами и сильными магнитными полями, что малоприемлемо с прикладной точки зрения. Ситуация, однако, может измениться с открытием новых материалов, например, таких как топологические диэлектрики, которые демонстрируют краевые состояния даже без приложения магнитного поля.

11.8.1. Соотношения Онзагера

Зададимся вопросом, возможна ли невзаимная проводимость в отсутствии магнитного поля. В общем случае матрица проводимости подчиняется соотношениям взаимности Онзагера

$$G_{n,m}(+B) = G_{m,n}(-B), \tag{117}$$

согласно которым ток на терминале n, обязанный напряжению, поданному на терминал m, равен току на терминале m, обязанному напряжению, поданному на терминал n, с одновременным обращением направления магнитного поля на противоположное. Это один из важнейших результатов статистической механики. Мы докажем его ниже опираясь на формулу Кубо для многотерминальных цепей.

Сейчас же зададимся вопросом, требуют ли соотношения Онзагера (117) взаимной проводимости в отсутствии магнитного поля ($B = 0$). Ответ положительный, но при условии, что проводники изготовлены из немагнитных материалов. В противном случае, взаимность обеспечивается не только обращением магнитного поля, но и обращением знака намагниченности материала:

$$G_{n,m}(+B,+M) = G_{m,n}(-B,-M). \tag{118}$$

Например, пусть один из контактов изготовлен из магнитного материала (рис. 32).

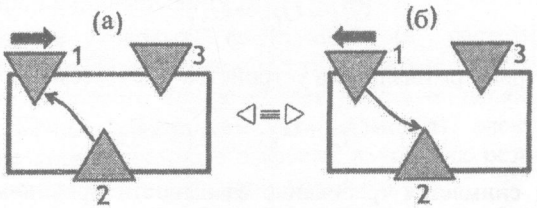


Рис. 32. Контакт 1 имеет отличную от нуля намагниченность.

Тогда, согласно соотношениям Онзагера (118), проводимость $G_{1,2}$ цепи (а) должна совпадать с проводимостью $G_{2,1}$ цепи (б). Однако, для одной и той же цепи, (а) или (б), проводимости $G_{1,2}$ и $G_{2,1}$ не совпадают.

При нынешнем понимании физики топологических диэлектриков, являющихся немагнитными материалами, топологические диэлектрики не обнаруживают невзаимную проводимость с немагнитными контактами в отсутствии магнитного поля, однако могут проявить невзаимность с магнитными контактами. Эти и ряд других аспектов физики топологических диэлектриков еще ждут своего более глубокого понимания [47].

Важнейшим приложением формулы Кубо (105) является доказательство справедливости соотношений взаимности Онзагера (117). Для этого используется многотерминальная версия формулы Кубо

$$G_{m,n}=\frac{1}{2kT}\int_{-\infty}^{+\infty}d\tau\left\langle I_m(t_0+\tau)I_n(t_0)\right\rangle_{eq}, \tag{119}$$

учитывающая корреляцию токов на двух различных терминалах.

Рассмотрим трехтерминальную цепь с магнитным полем ($B>0$), которое электроны, входящие в контакт 1, закручивает в направлении контакта 2, те же электроны, которые входят в контакт 2, закручивает в направлении контакта 3, и электроны, входящие в контакт 3, закручивает в направлении контакта 1 (рис. 33).

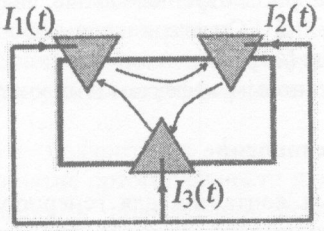


Рис. 33. Трехтерминальное устройство с магнитным полем, направляющим электроны с одного терминала на другой.

Можно ожидать, что коррелятор $\left\langle I_2(t_0+\tau)I_1(t_0)\right\rangle_{eq}$ должен выглядеть приблизительно как на рис. 34 с корреляцией, спадающей медленнее с ростом $\tau>0$.

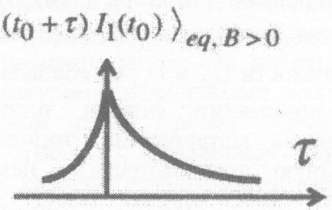


Рис. 34. Токовый коррелятор в окрестности момента времени t_0 при движении электронов от терминала 1 к терминалу 2 (качественно).

Объясняется это тем, что электроны движутся от первого терминала ко второму (рис. 33), и таким образом ток I_1 в момент времени t_0 сильнее скоррелирован с током I_2 в последующие моменты времени ($\tau > 0$), чем с током в предыдущие моменты ($\tau < 0$). Обращение направления магнитного поля на противоположное ($B < 0$) изменяет траектории движения электронов на обратные, так что

$$\left\langle I_1(t_0+\tau) I_2(t_0)\right\rangle_{eq,B<0} = \left\langle I_2(t_0+\tau) I_1(t_0)\right\rangle_{eq,B>0}. \qquad (120)$$

Считая это утверждение справедливым, соотношения Онзагера (117) немедленно следуют из формулы Кубо (119).

Соотношения Онзагера, рассмотренные выше, являют собой лишь частный случай общих соотношений Онзагера для обобщенных транспортных коэффициентов разного вида [43].

Вернемся, однако, к спиновым эффектам и их учету в методе НРФГ.

11.9. Спинорное представление

Применение магнитных контактов для генерирования спинового тока в немагнитном канале проводимости даже за пределами протекания тока с использованием пробного магнитного контакта 2 с поляризацией $\vec{P}_2$ (рис. 35) рассматривалось в главе 9 [21] на основе концепции о зарядовом электрохимическом потенциале E_F и спиновом потенциале $\vec{E}_F^*$.

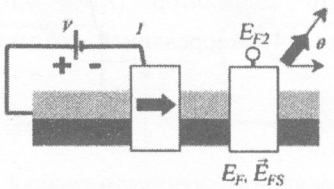


Рис. 35. Регистрирующий контакт 2 в роли анализатора спинового тока.

Связь между зарядовым и спиновым потенциалами дается простым выражением

$$E_{F2} = E_F + \frac{\vec{P}_2 \cdot \vec{E}_F^s}{2}, \qquad (121)$$

в котором вектор поляризации совпадает с направлением регистрирующего контакта 2 относительно направления тока, а вектор спинового потенциала совпадает с направлением спиновой поляризации канала проводимости, которое по договоренности есть направление намагниченности инжектирующего контакта. Обычно рассматриваемые параллельная и антипараллельная ориентации намагниченности контактов являются частными случаями общего выражения (121). У нас пока нет надежных сведений, выполнялись ли на самом деле подобные эксперименты, т.е. спины инжектировались контактом с фиксированным направлением намагниченности, а разность потенциалов замерялась вращающимся магнитным контактом. Неоднократно, однако, экспериментально измерялась фактически аналогичная ситуация, когда оба магнитных контакта были фиксированы друг относительно друга, а спин электронов в проводнике вращался внешним магнитным полем.

Все подобные эффекты исключительно интересны и важны сами по себе. Мы же, однако, далее займемся тем, что покажем как подобные спиновые эффекты учесть в методе НРФГ применительно к модельным транспортным задачам наноэлектроники, в частности, к рассмотрению квантового спинового эффекта Холла в графене. Покажем также как метод НРФГ приводит к результатам, соответствующим уравнению (121). Учет спина в методе НРФГ потребует введения матричных потенциалов формы (2×2), которые, как окажется, могут быть выражены через зарядовые и спиновые потенциалы, фигурирующие в (121).

Спин электронов и поляризация фотонов во многом похожи, за исключением того, что понятие ортогональности для них имеет разный смысл. Для фотонов ортогональность поляризации реализуется при 90°, а для электронов – при 180°, что проявляется в колебании нелокального спинового потенциала в зависимости от угла между инжектирующим и регистрирующим магнитными контактами (рис. 36) [21].

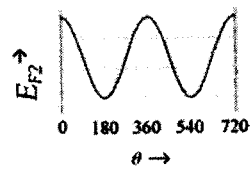


Рис. 36. Колебания нелокального спинового потенциала в зависимости от угла между инжектирующим и регистрирующим магнитными контактами (рис. 35).

Математически адекватное описание ортогональности электронных спинов требует спинорного исчисления, а не векторного как в случае поляризации фотонов. Вектор описывается тремя действительными компонентами вдоль осей *x, y, z* декартовой системы координат, а спинор описывается двумя комплексными компонентами вдоль направлений вверх (*up*) и вниз (*dn*):

$$\begin{Bmatrix} n_x \\ n_y \\ n_z \end{Bmatrix}, \quad \begin{Bmatrix} \psi_{up} \\ \psi_{dn} \end{Bmatrix},$$

где здесь и далее, как и прежде, векторы-столбцы будем брать в фигурные скобки, а квадратные матрицы -- в прямоугольные.

И тем не менее, спинор нужно воспринимать как объект, имеющий определенное направление как и вектор. Как же согласуется такое визуальное восприятие спинора с его двухкомпонентной структурой?

Спинор, ориентированный вдоль направления единичного вектора $\hat{n}$ с проекциями на оси *x, y, z*

$$\hat{n} \equiv \begin{Bmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{Bmatrix} \tag{122}$$

описывается двумя комплексными компонентами *c* и *s*:

$$\begin{Bmatrix} \cos\frac{\theta}{2}e^{-i\phi/2} \equiv c \\ \sin\frac{\theta}{2}e^{+i\phi/2} \equiv s \end{Bmatrix}. \tag{123}$$

Это совсем не очевидно, но позже мы покажем, что векторное (122) и спинорное (123) представления одного и того же абстрактного вращающегося объекта идентичны, точнее изоморфны. Пока что воспримем эквивалентность (122) и (123) как нечто заданное.

При учете спина в формализме НРФГ соответствующие уравнения, естественно, остаются прежними, а вот порядок всех матриц – гамильтониана [*H*], контактных [Σ], неравновесной функции Грина [G^n], спектральной функции [*A*] (глава 10) удваивается за счет того, что электроны, находящиеся в одном и том же квантовом координатном состоянии, но с разными спинами, теперь описываются разными базисными функциями. Если раньше, когда не учитывалось различие между спинами двух электронов, находящихся в одном и том же квантовом состоянии, число базисных функций было равно *N*, то теперь число базисных функций равно 2*N*. В квантовой химии этот подход известен как неограниченный (по спину) метод Хартри – Фока (Unrestricted Hartee –

Fock) [48 – 49] или иначе метод разных орбиталей для разных спинов (Different Orbitals for Different Spins) [50] или расщепленных орбиталей [51].

В простейшем варианте описания канала проводимости без потери «физики», к чему мы всегда стремимся, в хюккелевском приближении, более известном в физике твердого тела как методе сильной связи в ортогональном базисе с учетом взаимодействия лишь соседних атомов, канал проводимости считается не чувствительным к спину, другими словами, одинаково взаимодействует с электронами, находящимися в разных спиновых состояниях. Гамильтониан $[H]$ проводника в хюккелевском приближении теперь запишется точно так же, как и в [19], только кулоновский ε и резонансный t интегралы пишутся теперь в виде матриц 2×2 $[\alpha]$ и $[\beta]$ (рис. 37), а именно:

$$\begin{aligned}[\alpha]&=\varepsilon[I],\\ [\beta_x]&=[\beta_y]=t[I],\end{aligned} \tag{124}$$

где $[I]$ – единичная матрица формы 2×2. Фактически налицо два расщепленных гамильтониана, что не несет какой-либо новой «физики».

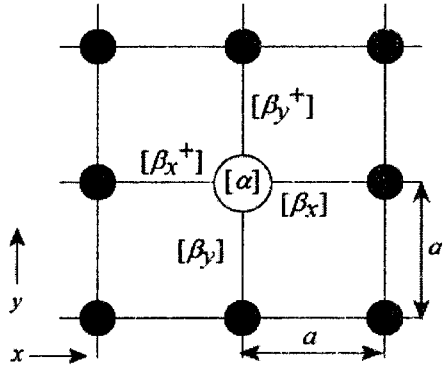


Рис. 37. Однородная 2D решетка с параметром решетки a, в которой каждый атом описывается кулоновским $[\alpha]$ и резонансными $[\beta]$ интегралами в виде матриц 2×2.

Аналогично записываются матрицы собственной энергии $[\Sigma]$ контактов, которые в этом приближении одинаково реагируют на разные спины, перемножая их обычные выражения на единичную матрицу $[I_2]$. В этом приближении результат ожидается тривиальным в том смысле, что мы как бы учитываем лишь спины одного сорта, и лишь в конце вычислений все результаты просто удваиваем.

Все спиновые эффекты, рассмотренные, например, в главе 9, возникают либо как результат учета нетривиальных контактов, чувствительных к спину и описываемых соответствующими контактными матрицами $[\Sigma]$, либо канал проводимости чувствителен к спину и описывается соответствующим гамильтонианом $[H]$, либо имеют место оба обстоятельства.

Сейчас же пока рассмотрим несколько модельных транспортных задач методом НРФГ с учетом спина и начнем с простой одноуровневой модели спинового вентиля. С вычислительной точки зрения вопрос лишь в том, как записать гамильтониан и контактные матрицы, а далее ход вычислений стандартный (§ 10.2).

11.9.1. Одноуровневый спиновый вентиль

Как известно (§ 9.1), спиновый вентиль обнаруживает различную проводимость G_P или G_{AP} в зависимости от того, намагниченность обоих контактов параллельна P или антипараллельна AP. Магнитосопротивление

$$MR = \left(G_p - G_{AP}\right)/G_{AP} \tag{125}$$

выражается через поляризацию P магнитных контактов

$$MR = \frac{P^2}{1-P^2}, \tag{126}$$

которая в свою очередь определяется граничними сопротивлениями вентиля

$$P \equiv \frac{R-r}{R+r} \tag{127}$$

при условии равенства нулю сопротивления канала проводимости.

В режиме магнитного туннельного перехода (МТП) в выражении для магнитосопротивления диэлектрического проводника появляется дополнительная двойка:

$$MR = \frac{2P^2}{1-P^2}, \tag{128}$$

появление которой можно понять из физики туннельных проводников, согласно которой результирующее сопротивление двух последовательно соединенных проводников пропорционально их произведению, а не сумме.

Мы сейчас получим этот же результат, включая множитель двойку, применив метод НРФГ к одноуровневому вентилю в предположении, что равновесный электрохимический потенциал E_{F0} лежит на много kT ниже энергии ε диэлектрического проводника (рис. 38).

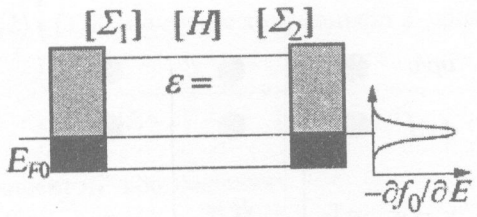


Рис. 38. Модель равновесного диэлектрического одноуровневого проводника.

Согласно квантовой модели одноуровневого однородного 1D резистора без учета спина (§ 10.4) гамильтониан и контактные матрицы есть просто числа (рис. 39).

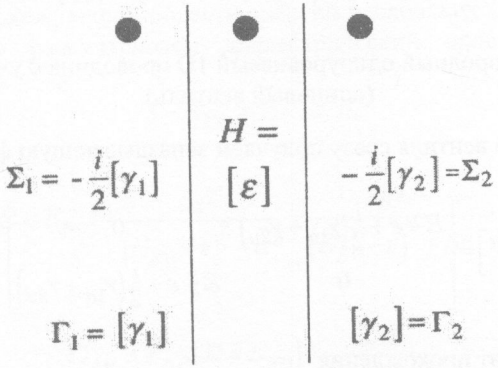


Рис. 39. Однородный одноуровневый 1D проводник без учета спина.

Собственные энергии контактов

$$\gamma_1 = \hbar\nu_1, \qquad \gamma_2 = \hbar\nu_2 \tag{129}$$

пропорциональны скоростям, с которыми электроны покидают проводник в направлении контактов.

При учете спина число базисных функций удваивается, вместо чисел появляются матрицы 2-го порядка: гамильтониан, однако, есть просто диагональная матрица, а вот контактные матрицы различные для спинов «вверх» *u* и спинов «вниз» *d* (рис. 40).

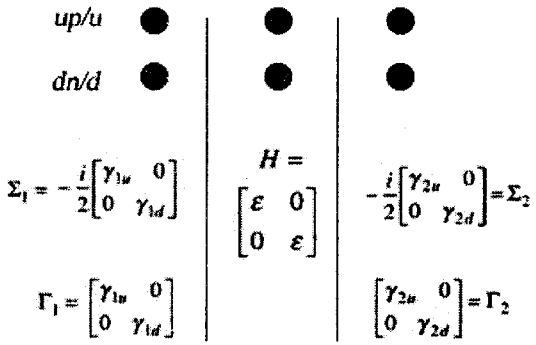


Рис. 40. Однородный одноуровневый 1D проводник с учетом спина (спиновый вентиль).

Для спинового вентиля сразу получаем запаздывающую функцию Грина

$$\left[G^R\right] = \begin{bmatrix} E-\varepsilon+\frac{i}{2}\left(\gamma_{1u}+\gamma_{2u}\right) & 0 \\ 0 & E-\varepsilon+\frac{i}{2}\left(\gamma_{1d}+\gamma_{2d}\right) \end{bmatrix}^{-1} \tag{130}$$

и далее коэффициент прохождения

$$\bar{T} = Tr\left[\Gamma_1 G^R \Gamma_2 G^A\right] = \frac{\gamma_{1u}\gamma_{2u}}{\left(E-\varepsilon\right)^2+\left(\frac{\gamma_{1u}+\gamma_{2u}}{2}\right)^2} + \frac{\gamma_{1d}\gamma_{2d}}{\left(E-\varepsilon\right)^2+\left(\frac{\gamma_{1d}+\gamma_{2d}}{2}\right)^2}. \tag{131}$$

Нужно помнить, что сейчас мы рассматриваем «не чувствительный к спину» канал проводимости.

Для параллельной конфигурации намагниченности контактов P оба контакта можно считать одинаковыми, так что ($\alpha > \beta$)

$$\begin{aligned} \gamma_{1u} &= \gamma_{2u} = \alpha, \\ \gamma_{1d} &= \gamma_{2d} \equiv \beta, \end{aligned} \tag{132}$$

тогда как для антипараллельной конфигурации AP второй контакт опрокидывает спин по сравнению с параллельной конфигурацией P, а именно:

$$\begin{aligned} \gamma_{1u} &= \gamma_{2d} = \alpha, \\ \gamma_{1d} &= \gamma_{2u} \equiv \beta. \end{aligned} \tag{133}$$

Подставляя (132) в (128), получаем коэффициент P-прохождения

$$\bar{T}_P = \frac{\alpha^2}{\left(E-\varepsilon\right)^2+\alpha^2} + \frac{\beta^2}{\left(E-\varepsilon\right)^2+\beta^2}, \tag{134}$$

а (133) дает коэффициент AP-прохождения

$$\bar{T}_{AP} = \frac{2\alpha\beta}{\left(E-\varepsilon\right)^2+\left(\frac{\alpha+\beta}{2}\right)^2}.$$

Измеряемая проводимость определяется усредненным значением коэффициента прохождения в области энергий нескольких kT вокруг значения E_{F0}. Мы сейчас рассматриваем диэлектрический проводник (рис. 38). Предполагая

$$\varepsilon - E_{F0} \gg kT, \alpha, \beta, \tag{135}$$

получаем

$$G_p \sim \bar{T}_p\left(E=E_{F0}\right) = \frac{\alpha^2}{\left(E_{F0}-\varepsilon\right)^2+\alpha^2} + \frac{\beta^2}{\left(E_{F0}-\varepsilon\right)^2+\beta^2} \approx \frac{\alpha^2+\beta^2}{\left(E_{F0}-\varepsilon\right)^2} \tag{136}$$

и

$$G_{AP} \sim \bar{T}_{AP}\left(E=E_{F0}\right) \approx \frac{2\alpha\beta}{\left(E_{F0}-\varepsilon\right)^2},$$

а для магнитосопротивления получаем искомую формулу (128)

$$MR \equiv \frac{G_p}{G_{AP}} - 1 = \frac{\alpha^2+\beta^2}{2\alpha\beta} - 1 = \frac{2P^2}{1-P^2} \tag{137}$$

с поляризацией

$$P \equiv \frac{\alpha-\beta}{\alpha+\beta}. \tag{138}$$

Точно так же можно получить формулу для магнитосопротивления (126) металлического немагнитного проводника, если предположить, что электрохимический потенциал E_{F0} лежит вблизи уровня с энергией ε, а $kT \gg \alpha, \beta$.

11.9.2. Вращение магнитных контактов

В спиновом вентиле в режиме *AP* второй контакт опрокидывает спины по сравнению с вентилем в режиме *P*, так что

$$\Gamma_1 = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}, \quad \Gamma_2 = \begin{bmatrix} \beta & 0 \\ 0 & \alpha \end{bmatrix}. \tag{139}$$

Как записать матрицу Γ для контакта, ориентированного в произвольном направлении единичного вектора $\hat{n}$ через его проекции (122) вдоль осей *x, y, z* ? Окончательная формула имеет вид

$$\Gamma = \frac{\alpha+\beta}{2}[I] + \frac{\alpha-\beta}{2}\begin{bmatrix} n_z & n_x - in_y \\ n_x + in_y & -n_z \end{bmatrix}, \tag{140}$$

где *I* – единичная матрица 2-го порядка. Формула сразу не очевидна, но мы скоро ее получим. А пока что убедитесь, что эта формула содержит частные случаи параллельной и антипараллельной ориентации намагниченностей контактов, соответственно:

$$\begin{aligned} n_z = +1, \quad n_x = n_y = 0: \quad \Gamma &= \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}, \\ n_z = -1, \quad n_x = n_y = 0: \quad \Gamma &= \begin{bmatrix} \beta & 0 \\ 0 & \alpha \end{bmatrix}. \end{aligned} \tag{141}$$

Убедиться в справедливости общей формулы (140) можно путем следующего наблюдения. Матрица

$$\tilde{\Gamma} = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix} \tag{142}$$

описывает магнитный контакт, ориентированный вдоль произвольно направленного единичного вектора $\hat{n}$, если в качестве оси отсчета выбрать направления $+\hat{n}$ и $-\hat{n}$ вместо стандартных декартовых полуосей $+\hat{z}$ и $-\hat{z}$. Вернуться назад к стандартному базису $\pm\hat{z}$ можно путем унитарного преобразования $\Gamma = V\,\tilde{\Gamma}\,V^{+}$ матрицы (142):

$$\begin{matrix} & \hat{n} & -\hat{n} \\ \hat{z} \\ -\hat{z} \end{matrix}\!\!\begin{bmatrix} c & -s^* \\ s & c^* \end{bmatrix}\!\!\begin{matrix} \\ \hat{n} \\ -\hat{n} \end{matrix}\!\!\begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}\!\!\begin{matrix} \\ \hat{n} \\ -\hat{n} \end{matrix}\!\!\begin{bmatrix} c^* & s^* \\ -s & c \end{bmatrix}\!\!\begin{matrix} & \hat{z} & -\hat{z} \\ \\ \\ \end{matrix}, \tag{143}$$

$$V \qquad\qquad\qquad\qquad V^{+}$$

где колонки матрицы унитарного преобразования $[V]$ есть компоненты спинора *c* и *s* (123) вдоль направлений $\pm\hat{n}$. Первая колонка матрицы V соответствует направлению $+\hat{n}$, а вторая колонка соответствует направлению $-\hat{n}$ с

компонентами спинора $-s^*$ и c^*, которые получаются из (123) путем преобразования углов

$$\theta \to \pi - \theta \quad \text{и} \quad \phi \to \pi + \phi \tag{144}$$

и удаления из обеих компонент общего фазового множителя.

Перемножение трех матриц в (143) дает

$$\Gamma = \begin{bmatrix} c & -s^* \\ s & c^* \end{bmatrix} \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix} \begin{bmatrix} c^* & s^* \\ -s & c \end{bmatrix} = \begin{bmatrix} c & -s^* \\ s & c^* \end{bmatrix} \begin{bmatrix} \alpha c^* & \alpha s^* \\ -\beta s & \beta c \end{bmatrix} = \begin{bmatrix} \alpha cc^* + \beta ss^* & (\alpha - \beta) cs^* \\ (\alpha - \beta) sc^* & \alpha ss^* + \beta cc^* \end{bmatrix}. \tag{145}$$

Используя определение компонент спинора c и s (123) и тригонометрических равенств

$$2\cos^2\frac{\theta}{2} = 1 + \cos\theta, \quad 2\sin^2\frac{\theta}{2} = 1 - \cos\theta \quad \text{и} \quad 2\sin\frac{\theta}{2}\cos^2\frac{\theta}{2} = \sin\theta, \tag{146}$$

перепишем (145) в виде

$$\Gamma = \frac{1}{2}\begin{bmatrix} (\alpha+\beta) + (\alpha-\beta)\cos\theta & (\alpha-\beta)\sin\theta e^{-i\phi} \\ (\alpha-\beta)\sin\theta e^{+i\phi} & (\alpha+\beta) - (\alpha-\beta)\cos\theta \end{bmatrix}, \tag{147}$$

что приводит к искомой формуле (140), если воспользоваться выражениями (122) для компонент единичного вектора $\hat{n}$ вдоль осей x, y, z.

Если вектор поляризации

$$\vec{P} \equiv P\hat{n} = \frac{\alpha - \beta}{\alpha + \beta}\hat{n} \tag{148}$$

определить его величиной (138) и единичным вектором $\hat{n}$, то уравнение (140) можно переписать в виде

$$\Gamma = \frac{\alpha + \beta}{2}\left([I] + \begin{bmatrix} P_z & P_x - iP_y \\ P_x + iP_y & -P_z \end{bmatrix} \right) \tag{149}$$

или же еще иначе

$$\frac{[\Gamma]}{(\alpha+\beta)/2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + P_x \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + P_y \begin{bmatrix} 0 & -i \\ +i & 0 \end{bmatrix} + P_z \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{150}$$

через широко используемые в квантовой теории спиновые матрицы Паули

$$\sigma_x \equiv \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y \equiv \begin{bmatrix} 0 & -i \\ +i & 0 \end{bmatrix}, \quad \sigma_z \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{151}$$

В конечном итоге уравнение (149) можно переписать компактно в виде

$$\Gamma = \frac{\alpha+\beta}{2}\left([I]+[\sigma_x]P_x+[\sigma_y]P_y+[\sigma_z]P_z\right) = \frac{\alpha+\beta}{2}\left([I]+[\vec{\sigma}]\cdot\vec{P}\right). \qquad (152)$$

Подобным же образом можно компактно переписать контактные матрицы собственной энергии. Пусть, например, в базисе $\pm\hat{n}$

$$\tilde{\Sigma} = -\frac{i}{2}\begin{bmatrix}\alpha & 0\\ 0 & \beta\end{bmatrix}, \qquad (153)$$

тогда в базисе $\pm\hat{z}$ имеем:

$$\Sigma = -i\frac{\alpha+\beta}{4}[I] - i\frac{\alpha-\beta}{4}[\vec{\sigma}]\cdot\hat{n} = -i\frac{\alpha+\beta}{4}\left([I]+[\vec{\sigma}]\cdot\vec{P}\right). \qquad (154)$$

11.10. Спиновые гамильтонианы

Перейдем к обсуждению спин-зависимых гамильтонианов и рассмотрим учет классического спин-орбитального зеемановского расщепления и гамильтониан Рашбы [52, 53].

11.10.1. Гамильтониан с зеемановским расщеплением

Пусть направление магнитного поля $\vec{B}$ определяется единичным вектором $\hat{n}$. Тогда гамильтониан в базисе $\pm\hat{n}$ имеет вид

$$\mu_{el}\begin{bmatrix}+B & 0\\ 0 & -B\end{bmatrix}, \qquad (155)$$

где μ_{el} – магнитный момент электрона. Уровень с положительным спином смещается вверх на величину $\mu_{el}B$, а уровень с отрицательным спином смещается вниз на эту же величину (зеемановское расщепление).

В базисе $\pm\hat{z}$ имеем:

$$H_B = \mu_{el}\vec{\sigma}\cdot\vec{B}. \qquad (156)$$

Добавить еще нужно спин-независимую часть гамильтониана. Для случая параболической дисперсии имеем

$$H = \frac{\hbar^2}{2m}\left(k_x^2+k_y^2\right)[I] + \mu_{el}\vec{\sigma}\cdot\vec{B}, \qquad (157)$$

а для однородной 2D решетки в хюккелевском приближении с резонансным и кулоновским интегралами (§ 10.9.2)

$$t = -\hbar^2/2ma^2, \qquad (158)$$

$$\varepsilon = E_C - 4t \qquad (159)$$

гамильтониан имеет вид:

$$H=\left(\varepsilon+2t\cos k_x a+2t\cos k_y a\right)\left[I\right]+\mu_{el}\vec{\sigma}\cdot\vec{B}, \tag{160}$$

где a – параметр решетки. Для такой решетки (рис. 37) кулоновская и резонансная матрицы имеют вид:

$$\begin{aligned}\alpha&=\varepsilon\left[I\right]+\mu_{el}\vec{\sigma}\cdot\vec{B}\\ \beta_x&=t\left[I\right],\quad \beta_y=t\left[I\right]\end{aligned} \tag{161}$$

По сравнению со спин-независимым гамильтонианом (124) изменяется только кулоновская матрица.

11.10.2. Гамильтониан Рашбы

Более полный учет спин-орбитального взаимодействия дает гамильтониан

$$H_R=\eta\hat{z}\cdot\left(\vec{\sigma}\times\vec{k}\right)=\eta\left(\sigma_x k_y-\sigma_y k_x\right), \tag{162}$$

где η – параметр Рашбы [52 – 54]. Этот гамильтониан имеет релятивистское происхождение, обусловленные им эффекты широко наблюдаются экспериментально и надежно интерпретированы [54]. Нас же сейчас интересует только одно – как учесть взаимодействие Рашбы для нашей однородной 2D решетки в хюккелевском приближении. С этой целью аппроксимируем (162) синусами

$$H_R=\frac{\eta}{a}\left(\sigma_x\sin k_y a-\sigma_y\sin k_x a\right) \tag{163}$$

и запишем их удобства ради через экспоненты

$$H_R=\frac{\eta}{2ia}\sigma_x\left(e^{+ik_y a}-e^{-ik_y a}\right)-\frac{\eta}{2ia}\sigma_y\left(e^{+ik_x a}-e^{-ik_x a}\right), \tag{164}$$

сомножители перед которыми

$$\begin{aligned}\beta_x&=\frac{i\eta}{2a}\sigma_y,\quad \beta_x^{+}=-\frac{i\eta}{2a}\sigma_y,\\ \beta_y&=-\frac{i\eta}{2a}\sigma_x,\quad \beta_y^{+}=\frac{i\eta}{2a}\sigma_x\end{aligned} \tag{165}$$

обеспечивают правильное дисперсионное соотношение (157).

Учтем спин-независимую часть гамильтониана 2D решетки (4), а также постоянное магнитное поле (41) через зеемановское расщепление. Окончательно кулоновская и резонансная матрицы однородной 2D решетки с

учетом зеемановского спин-орбитального взаимодействия и взаимодействия Рашбы имеют вид:

$$\alpha = \varepsilon I + \mu_{el}\vec{\sigma}\cdot\vec{B},$$
$$\beta_x = tI + \frac{i\eta}{2a}\sigma_y, \quad \beta_x^+ = tI - \frac{i\eta}{2a}\sigma_y, \tag{166}$$
$$\beta_y = tI - \frac{i\eta}{2a}\sigma_x, \quad \beta_y^+ = tI + \frac{i\eta}{2a}\sigma_x.$$

Обратим внимание на то, что взаимодействие Рашбы учитывается через параметр Рашбы и матрицы Паули в резонансных матрицах 2D решетки, а зеемановское взаимодействие – в кулоновских матрицах.

11.11. Спиноры и векторы

Мы обычно визуально представляем себе спин как вращающийся объект, ориентированный в определенном направлении. Он описывается двумя комплексными компонентами спинора (123), а вектор в этом же направлении задается тремя действительными проекциями (122) на оси декартовой системы координат. Чтобы почувствовать связь между двумя представлениями одного и того же направленного объекта – векторным и спинорным полезно рассмотреть прецессию спина в магнитном поле в обоих представлениях.

Рассмотрим одноуровневый резистор с $\varepsilon = 0$ в магнитном поле, направленном вдоль оси z. Двухкомпонентное уравнение Шредингера

$$\frac{d}{dt}\begin{Bmatrix}\psi_u\\ \psi_d\end{Bmatrix} = \frac{\mu_{el}B_z}{i\hbar}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{Bmatrix}\psi_u\\ \psi_d\end{Bmatrix}, \tag{167}$$

где справа стоит спиновая матрица σ_z, фактически есть два дифференциальных уравнения с очевидными решениями

$$\psi_u(t) = \psi_u(0)e^{-i\omega t/2},$$
$$\psi_d(t) = \psi_d(0)e^{+i\omega t/2}. \tag{168}$$

где частота прецессии

$$\omega \equiv \frac{2\mu_{el}B_z}{\hbar}. \tag{169}$$

Пусть спин электрона в начальный момент времени $t = 0$ направлен вдоль единичного вектора $\hat{n}$ под углами θ и ϕ относительно выбранной системы отсчета (рис. 41).

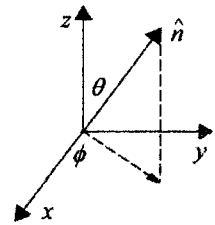


Рис. 41. К прецессии по углу ϕ спина, направленного вдоль орта $\hat{n}$.

В этот начальный момент времени его волновые функции

$$\left\{\begin{aligned}\psi_u(0)&=\cos\frac{\theta}{2}e^{-i\phi/2}\\ \psi_d(0)&=\sin\frac{\theta}{2}e^{+i\phi/2}\end{aligned}\right\}, \tag{170}$$

а в последующие моменты времени

$$\left\{\begin{aligned}\psi_u(t)&=\cos\frac{\theta}{2}e^{-i\phi/2}e^{-i\omega t/2}\\ \psi_d(t)&=\sin\frac{\theta}{2}e^{+i\phi/2}e^{+i\omega t/2}\end{aligned}\right\}, \tag{171}$$

что соответствует вращению спина вокруг оси z под фиксированным к ней углом θ, тогда как угол ϕ растет линейно со временем:

$$\phi(t)=\phi(0)+\omega t. \tag{172}$$

Согласно (122), для проекций единичного вектора имеем:

$$\begin{aligned}n_x&=\sin\theta\cos\phi(t),\\ n_y&=\sin\theta\sin\phi(t),\\ n_z&=\cos\theta,\end{aligned} \tag{173}$$

откуда с использованием (172) получаем

$$\frac{dn_x}{dt}=-\omega n_y,\quad \frac{dn_y}{dt}=+\omega n_x \tag{174}$$

или в матричном виде для всех трех компонент единичного вектора

$$\frac{d}{dt}\begin{Bmatrix}n_x\\ n_y\\ n_z\end{Bmatrix}=\omega\begin{bmatrix}0&-1&0\\ +1&0&0\\ 0&0&0\end{bmatrix}\begin{Bmatrix}n_x\\ n_y\\ n_z\end{Bmatrix}, \tag{175}$$

где квадратная матрица, обозначаемая обычно R_z, есть матрица вращения единичного вектора вокруг оси z.

Сравните полученное уравнение (125) с уравнением Шредингера (167) в спинорном представлении, переписанном через частоту прецессии (169):

$$\frac{d}{dt}\begin{Bmatrix}\psi_u\\ \psi_d\end{Bmatrix} = \frac{\omega}{2i}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{Bmatrix}\psi_u\\ \psi_d\end{Bmatrix}, \tag{176}$$

где фигурирует матрица Паули σ_z.

Пусть мы хотим описать прецессию спина электрона в магнитном поле, направленном вдоль оси x. Для этого в уравнении (175) достаточно лишь выполнить циклическую перестановку переменных: $x \to y$, $y \to z$, $z \to x$ и мы получим уравнение с матрицей вращения R_x:

$$\frac{d}{dt}\begin{Bmatrix}n_x\\ n_y\\ n_z\end{Bmatrix} = \omega\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & -1\\ 0 & +1 & 0\end{bmatrix}\begin{Bmatrix}n_x\\ n_y\\ n_z\end{Bmatrix}; \tag{177}$$

если в уравнении (175) -1 стояла на пересечении строки x и колонки y, то в (177) она окажется на пересечении строки y и колонки z.

Сразу не очевидно как именно с той же целью модифицировать уравнение (176). Правильный ответ таков:

$$\frac{d}{dt}\begin{Bmatrix}\psi_u\\ \psi_d\end{Bmatrix} = \frac{\omega}{2i}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{Bmatrix}\psi_u\\ \psi_d\end{Bmatrix}, \tag{178}$$

где фигурирует матрица Паули σ_x.

Уравнения (175) и (176), как и уравнения (177) и (178), описывают один и тот же физический процесс, а именно, вращение спина электрона в магнитном поле, направленном вдоль оси z, соответственно вдоль оси x, в первом случае – через три действительных компоненты, а во втором – через две комплексных. Циклическая перестановка в противоположном направлении даст нам матрицу вращений R_y, а соответствующая ей спиновая матрица σ_y приведена в сводке (151).

Что делает матрицы вращения и спиновые матрицы изоморфными, позволяя им описывать один и тот же физический процесс? Ответ состоит в том, что у этих матриц одинаковые коммутационные соотношения.

Легко убедиться, что матрицы вращения

$$R_x = \begin{bmatrix}0 & 0 & 0\\ 0 & 0 & -1\\ 0 & +1 & 0\end{bmatrix},\quad R_y = \begin{bmatrix}0 & 0 & +1\\ 0 & 0 & 0\\ -1 & 0 & 0\end{bmatrix},\quad R_z = \begin{bmatrix}0 & -1 & 0\\ +1 & 0 & 0\\ 0 & 0 & 0\end{bmatrix} \tag{179}$$

подчиняются следующим коммутационным соотношениям:

$$\begin{aligned}\left[R_x,R_y\right]&\equiv R_xR_y-R_yR_x=R_z,\\ \left[R_y,R_z\right]&\equiv R_yR_z-R_zR_y=R_x,\\ \left[R_z,R_x\right]&\equiv R_zR_x-R_xR_z=R_y,\end{aligned}\qquad(180)$$

где слева стоят стандартные обозначения соответствующих коммутаторов.

Спиновые матрицы Паули подчиняются таким же коммутационным соотношениям, если R формально заменить на $\sigma/2i$:

$$\begin{aligned}\left[\sigma_x,\sigma_y\right]&\equiv \sigma_x\sigma_y-\sigma_y\sigma_x=2i\sigma_z,\\ \left[\sigma_y,\sigma_z\right]&\equiv \sigma_y\sigma_z-\sigma_z\sigma_y=2i\sigma_x,\\ \left[\sigma_z,\sigma_x\right]&\equiv \sigma_z\sigma_x-\sigma_x\sigma_z=2i\sigma_y.\end{aligned}\qquad(181)$$

Заметим, что в стандартных учебниках квантовой механики спиновые матрицы Паули обычно вводятся постулативно.

Какие должны быть компоненты у спинора, направленного вдоль оси z? Поскольку спинор не изменяется при вращении вокруг оси z, то его компонентами должны быть собственные векторы спиновой матрицы σ_z в сводке (151):

$$\begin{Bmatrix}1\\0\end{Bmatrix}\quad\text{или}\quad\begin{Bmatrix}0\\1\end{Bmatrix},\qquad(182)$$

представляющие действительно спин «вверх» или спин «вниз» вдоль оси z.

Если же мы хотим узнать компоненты спинора, направленного вдоль оси x, то нужно найти собственные векторы спиновой матрицы σ_x, каковыми являются векторы

$$\begin{Bmatrix}+1\\+1\end{Bmatrix}\Big/\sqrt{2}\quad\text{или}\quad\begin{Bmatrix}+1\\-1\end{Bmatrix}\Big/\sqrt{2},\qquad(183)$$

представляющие направление спина «вверх» вдоль положительной полуоси или спина «вниз» вдоль отрицательной полуоси оси x.

В общем случае, если мы хотим найти компоненты спинора, направленного вдоль произвольно ориентированного единичного вектора $\hat{n}$ (122), нужно искать собственные векторы матрицы

$$\vec{\sigma}\cdot\hat{n}=\sigma_x\sin\theta\cos\phi+\sigma_y\sin\theta\sin\phi+\sigma_z\cos\theta=\begin{bmatrix}\cos\theta & \sin\theta e^{-i\phi}\\ \sin\theta e^{+i\phi} & -\cos\theta\end{bmatrix},\qquad(184)$$

обозначенные ранее как

$$\begin{Bmatrix}c\\s\end{Bmatrix}\quad\text{и}\quad\begin{Bmatrix}-s^*\\c^*\end{Bmatrix},$$

первый из которых соответствует ориентации спина вдоль положительного направления вектора $\hat{n}$ и в явном виде выписан в (123).

11.12. Прецессия спина

Мы уже знаем, как составить гамильтониан H и контактные матрицы Σ, в том числе и с учетом нетривиальных спиновых эффектов, и готовы просчитать численные модели вычисляя электронную плотность через неравновесную функцию Грина G^n, плотность состояний через спектральную функцию A и ток (глава 10).

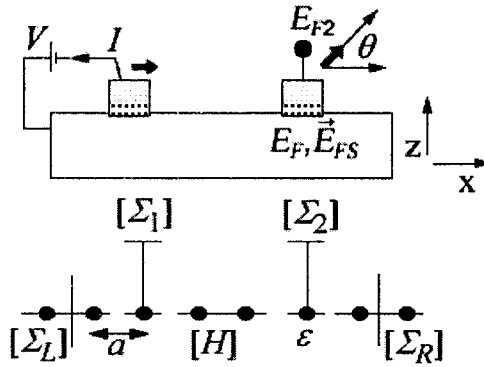


Рис. 42. К измерению и расчету методом НРФГ спинового потенциала в рамках четырехтерминальной модели однородного 1D проводника с двумя магнитными зондами, описываемыми контактными матрицами Σ_1 и Σ_2.

Сначала рассчитаем потенциал на 2-ом зонде, описываемом контактной матрицей Σ_2, однородного 1D проводника (рис. 42) методом НРФГ с простым гамильтонианом без учета спиновых эффектов в зависимости от угла наклона θ магнитного зонда 2 по отношению к линейному проводнику (плавающий контакт). Результаты расчета для различных значений поляризации (127) P_2 2-го зонда относительно 1-го представлены на рис. 43.

Полностью подтверждается расчетом методом НРФГ формула (121): измеряемый потенциал осциллирует в зависимости от угла наклона одного магнитного контакта относительно другого. При нулевой поляризации контактов измеряемый потенциал остается постоянным.

Как уже упоминалось вначале, нет сведений, что подобные эксперименты с плавающим магнитным контактом действительно проводились. Проводились, однако, эксперименты с вращением спина электронов в канале проводимости при фиксированной геометрии взаимного расположения магнитных контактов.

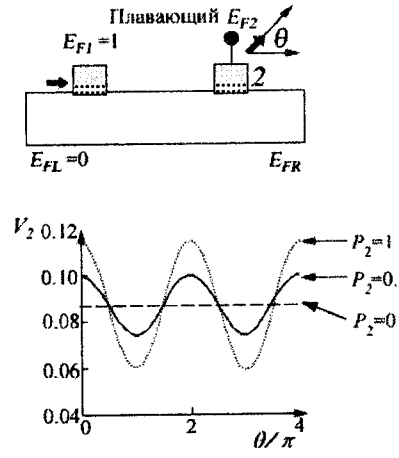


Рис. 43. Изменение потенциала 2-го зонда относительно 1-го в зависимости от угла наклона, нормированного на 180°, при разных значениях взаимной поляризации зондов [13].

Один из широко распространенных методов заключается в приложении внешнего магнитного поля, что приводит к прецессии спина относительно направления магнитного поля с циклической частотой ω (169). Это означает, что спиновый потенциал E_{F2} в точке, где находится пробный зонд 2, повернется на угол $\omega\tau$, где τ есть время прохождения электроном пути от инжектирующего контакта до пробного зонда 2. Записав $\tau = L/v$, из уравнений (121) и (169) имеем:

$$E_{F2} = E_F + P_2 E_F^s \cos\frac{2\mu_{el}L}{\hbar v} B_z . \qquad (185)$$

Можно ожидать появления осциллирующего потенциала на зонде 2, зависящего от величины приложенного магнитного поля. Однако, обычно наблюдают не баллистический транспорт, и имеет место большой разброс в значениях времен τ, в результате чего потенциал, усредненный по всем временам τ зануляется. Типичные эксперименты вместо осциллирующего сигнала, зависящего от магнитного поля, показывают просто уменьшение потенциала от значения $P_2 E_F^s$ вплоть до нуля (эффект Хэнли [55, 56]). И тем не менее, осцилляции в эффекте Хэнли наблюдать удалось за счет уменьшения разброса времен τ настолько, что разброс оказался намного меньше среднего значения времени τ [57].

Другим подходом к вращению спина электрона может служить эффект Рашбы в проводниках с сильным спин-орбитальным взаимодействием. Сейчас

хорошо установлено на примере многих полупроводников, что электрическое поле вдоль оси z (рис. 42) ведет к появлению эффективного магнитного поля, зависящего от импульса электрона. В этом можно убедиться, сравнивая гамильтониан с учетом магнитного поля (156) с гамильтонианом Рашбы (162), который, удобства ради, перепишем в виде

$$H_R = \eta\hat{z}\cdot\left(\vec{\sigma}\times\vec{k}\right) = \eta\vec{\sigma}\cdot\left(\vec{k}\times\hat{z}\right), \tag{186}$$

откуда для эффективного магнитного поля, обязанного взаимодействию Рашбы, получаем

$$\mu_{el}\vec{B}_{eff} = \eta\hat{z}\times\vec{k}, \tag{187}$$

так что согласно (185) можно ожидать осциллирующий потенциал вида

$$E_{F2} = E_F + P_2 E_F^s \cos\frac{2\eta kL}{\hbar v} \tag{188}$$

с периодом

$$\frac{2kL}{\hbar v}\Delta\eta = 2\pi \quad \text{или иначе} \quad \Delta\eta = \frac{2\pi a t_0}{kL}\sin ka. \tag{189}$$

Этот вывод находится в неплохом согласии с численными результатами, полученными методом НРФГ для энергии, соответствующей $ka = \pi/3$ при расстоянии между инжектором и детектором $L = 40a$ (рис. 44).

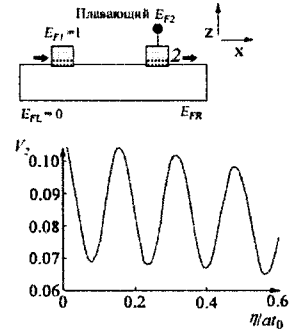


Рис. 44. В проводниках с большим значением коэффициента Рашбы η потенциал зонда 2 осциллирует, если намагниченности обоих зондов направлены вдоль проводника (ось x); осцилляции исчезают, если намагниченности направлены перпендикулярно к проводнику (ось z) [13].

В этой модели электроны, движущиеся вдоль оси x в ее положительном направлении, испытывают влияние эффективного магнитного поля вдоль оси y.

Поскольку инжектируемые электроны имеют спиновый потенциал $\vec{E}_F'$, направленный вдоль намагниченности зондов 1 и 2 (ось x), спиновый потенциал должен вращаться. Осцилляции потенциала исчезают, если намагниченности обоих зондов направлены вдоль оси y. Этот эффект подтверждается экспериментально [58, 59].

11.13. Квантовый спиновый эффект Холла

Транспорт поляризованных спинов в немагнитных материалах, таких как медь, изучен достаточно полно и физические основы в целом ясны. Этого нельзя сказать о диффузии спинов в материалах с сильным спин-орбитальным взаимодействием.

В последнее десятилетие пристальное внимание уделялось квантовому спиновому эффекту Холла (КСХ). В обычном классическом эффекте Холла электроны, движущиеся от истока к стоку вдоль оси x, под действием магнитного поля, направленного перпендикулярно поверхности 2D проводника, вдоль оси z, закручиваются таким образом, что создается поперечная разность потенциалов вдоль оси y (рис. 45).

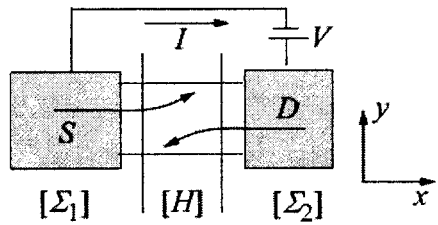


Рис. 45. К выбору системы координат при обсуждении эффекта КСХ.

Эффект КСХ можно понять в рамках модели, согласно которой эффективное магнитное поле закручивает спины, поляризованные вдоль полуосей $+z$ и $-z$, в противоположных направлениях, тогда как в классическом эффекте Холла оба сорта спинов закручиваются в одном и том же направлении.

Поскольку электроны в проводниках с сильным спин-орбитальным взаимодействием испытывают действие эффективного магнитного поля, естественно задать вопрос, можно ли наблюдать холловский эффект в отсутствии внешнего магнитного поля, а лишь под влиянием внутреннего эффективного магнитного поля.

Ответ оказывается положительным, но подобное явление с закручиванием спинов, поляризованных вдоль $+z$ и $-z$, в противоположных направлениях представляется весьма деликатным явлением, поскольку заряды не накапливаются в поперечном направлении и холловское напряжение зануляется. Однако, спиновый потенциал может быть измерен магнитными контактами.

В первых экспериментах накопление спинов противоположной поляризации в поперечном направлении (КСХ) обнаружили в массивных полупроводниках оптическими зондами [60], а недавно КСХ этого типа измерили магнитными зондами в нанопроводниках [59].

Транспортная модель на основе метода НРФГ адекватно описывает КСХ, о чем свидетельствуют результаты расчета однородного 2D проводника (рис. 44) с коэффициентом Рашбы $\eta = 1 \cdot 10^{-11}$ *эВ·м* с энергией $E = 0.05 t_0$ и переменной ширине W (рис. 46). Для вычленения z-компоненты спиновой плотности из гриновской функции G^n использовалось уравнение (185).

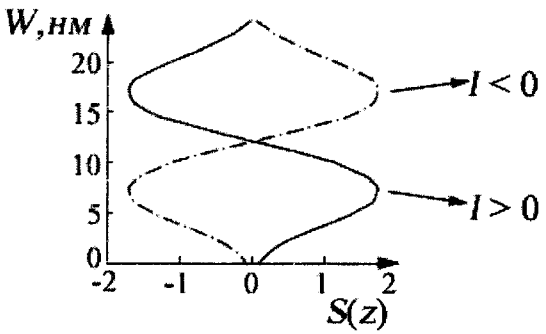


Рис. 46. Квантовый спиновый эффект Холла. Численные результаты, полученные методом НРФГ, показывают накапливание $+z$-спинов и $-z$-спинов $S(z)$ (в произвольных единицах) на противоположных сторонах 2D проводника, при этом изменение направления тока вдоль проводника на противоположное приводит к обращению накапливаемых спинов [13].

В последние несколько лет значительно возрос интерес к транспорту спинов в материалах с сильным спин-орбитальным взаимодействием, особенно в топологических диэлектриках [21], которые обнаруживают совсем удивительные проявления КСХ [60].

11.14. Метод НРФГ и диффузионный транспорт

В заключение обсудим связь метода НРФГ с диффузионным подходом для баллистического транспорта, развитым в главе 9. Мы уже знаем, что численные результаты, полученные выше методом НРФГ, качественно верно описываются уравнением (121), сформулированным в § 9.7 из эвристических соображений. Однако, причина этого согласия не очевидна, тем более что в методе НРФГ для моделирования транспорта спинов первично используются двухкомпонентные комплексные спиноры, а спиновый потенциал $\vec{E}_F^s$ описывается трехкомпонентным действительным вектором.

Выше уже подчеркивалась связь между единичным вектором $\hat{n}$, вдоль которого направлен спин, и спиновыми волновыми функциями ψ_{up} и ψ_{dn}. Чтобы установить связь между уравнением (121) и методом НРФГ для транспорта спинов, нужно рассматривать такие величины как гриновская функция $G^n \sim \psi\psi^+$, а не просто волновая функция ψ, поскольку метод НРФГ исходно формулируется через G^n. Не говоря уже о том, что именно гриновская функция G^n, а не волновая функция, является наблюдаемой величиной, пригодной для установления связи с экспериментальными результатами.

11.14.1. Электронная плотность в матричном представлении

Мы уже не раз использовали электронную плотность в матричной записи $[G^n]$, диагональные элементы которой дают нам число электронов в данной точке проводника (глава 10). С учетом спина матрица $[G^n]$ в данной точке есть матрица (2×2) и она дает нам число электронов или суммарный спин $\vec{S}$. Убедиться в этом можно рассматривая электрон со спином, направленным вдоль произвольного направления $\hat{n}$ (рис. 41) и описываемым волновой функцией в спинорном представлении (170), так что соответствующая матрица электронной плотности $[G^n]$ формы (2×2) дается выражением

$$\psi\psi^+ = \begin{Bmatrix} c \\ s \end{Bmatrix} \begin{Bmatrix} c^* & s^* \end{Bmatrix} = \begin{bmatrix} cc^* & cs^* \\ sc^* & ss^* \end{bmatrix} \tag{190}$$

или иначе, используя (122) и (146),

$$\psi\psi^+ = \frac{1}{2}\begin{bmatrix} 1+n_z & n_x - in_y \\ n_x + in_y & 1-n_z \end{bmatrix} = \frac{1}{2}\left[I + \vec{\sigma}\cdot\hat{n}\right]. \tag{191}$$

Суммируя вклады в $\psi\psi^+$ от всех N электронов, для матрицы гриновской функции окончательно получим

$$\frac{G^n}{2\pi} = \frac{1}{2}\begin{bmatrix} N+S_z & S_x - iS_y \\ S_x + iS_y & N-S_z \end{bmatrix} = \frac{1}{2}\left(N\left[I\right] + \vec{\sigma}\cdot\vec{S}\right). \tag{192}$$

Вычислив матрицу $[G^n]$, полное число электронов и суммарный спин можно найти из

$$N = \frac{1}{2\pi}Tr\left[G^n\right], \quad \vec{S} = \frac{1}{2\pi}Tr\left[\vec{\sigma}\,G^n\right], \tag{193}$$

что следует из равенства нулю следа спиновых матриц Паули (151) и из следующих свойств спиновых матриц, вытекающих из (151) и (181), а именно:

$$\sigma_x^2=\sigma_y^2=\sigma_z^2=I,$$
$$\sigma_x\sigma_y=-\sigma_y\sigma_x=i\sigma_z,$$
$$\sigma_y\sigma_z=-\sigma_z\sigma_y=i\sigma_x,$$
$$\sigma_z\sigma_x=-\sigma_x\sigma_z=i\sigma_y. \qquad (194)$$

В итоге, вся информация, содержащаяся в эрмитовой матрице $[G^n]$, может быть представлена четырьмя действительными величинами в данной точке проводника – полным числом электронов N и тремя действительными компонентами вектора полного спина $\vec{S}$, которые можно извлечь из (193).

11.14.2. Измерение спинового потенциала

Мы подошли к установлению связи между уравнением (121) и методом НРФГ для транспорта спинов. Скалярную версию (53/§9.3) уравнения (121) мы получили в рамках полуклассической модели вычисления токов на детектирующем зонде через его граничные проводимости для спинов противоположной направленности (рис. 12/§9.3), полагая зондовый ток нулевым. Теперь мы получим полное уравнение (121) в векторной форме, опираясь на метод НРФГ (рис. 47).

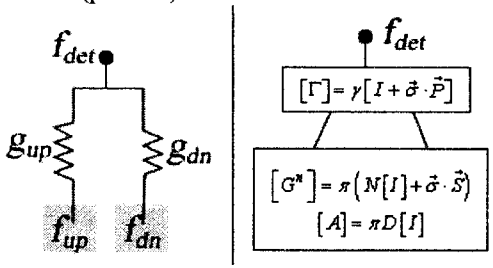


Рис. 47. Моделирование детектирующего зонда в полуклассической модели и в методе НРФГ.

Для тока имеем четвертое уравнение (16/гл.10) в формализме НРФГ

$$I\sim Tr\left[\Gamma\right]\left[f_{\det}\left[A\right]-\left[G^n\right]\right], \qquad (195)$$

так что при нулевом токе на детектирующем зонде (рис. 47)

$$f_{\det}=\frac{Tr\left[\Gamma\right]\left[G^n\right]}{Tr\left[\Gamma\right]\left[A\right]}. \qquad (196)$$

Подставляя гриновскую функцию G^n (192), спектральную функцию A выражая через спин-независимую плотность состояний D

$$\frac{A}{2\pi}=\frac{D}{2}[I], \tag{197}$$

описывая связь детектирующего зонда с каналом проводимости согласно (152)

$$\Gamma=\gamma\left[I+\vec{\sigma}\cdot\vec{P}\right], \tag{198}$$

и учитывая нулевой след у всех матриц Паули, окончательно получим:

$$f_{\det}=Tr\left[I+\vec{\sigma}\cdot\vec{P}\right]\left[\frac{N}{D}I+\vec{\sigma}\cdot\frac{\vec{S}}{D}\right]. \tag{199}$$

Учитывая тождество для двух произвольных векторов $\vec{P}$ и $\vec{B}\equiv\vec{S}/D$

$$\left[\vec{\sigma}\cdot\vec{P}\right]\left[\vec{\sigma}\cdot\vec{B}\right]=\left(\vec{P}\cdot\vec{B}\right)[I]+i\vec{\sigma}\cdot\left[\vec{P}\times\vec{B}\right], \tag{200}$$

имеем

$$\left[I+\vec{\sigma}\cdot\vec{P}\right]\left[bI+\vec{\sigma}\cdot\vec{B}\right]=\left(b+\vec{P}\cdot\vec{B}\right)[I]+\vec{\sigma}\cdot\left[\vec{P}+\vec{B}+i\vec{P}\times\vec{B}\right], \tag{201}$$

и окончательно, используя равенство нулю следа матриц Паули, вместо (199) имеем:

$$f_{\det}=\frac{N}{D}+\vec{P}\cdot\frac{\vec{S}}{D}\equiv f+\vec{P}\cdot\frac{\vec{f}_s}{2}, \tag{202}$$

где зарядовая заселенность f и спиновая f_s определены таким образом, чтобы их максимально возможные значения не превышали единицы. Чтобы перейти от заселенностей к электрохимическим потенциалам и получить искомое уравнение (121), воспользуемся линейной зависимостью между ними в режиме линейного отклика (глава 1).

11.14.3. Транспорт спинов в четырехкомпонентном формате

Подчеркнем еще раз, что квантовый подход естественным образом ведет к комплексной матрице $[G^n]$ формы (2×2) в каждой точке проводника, и вместе с тем гриновскую матрицу всегда можно переписать через четыре действительных и физически понятных числа таких как $(N,\vec{S})$ или $(E_F,\vec{E}_F^s)$.

Мы уже видели в главе 9 как много различных явлений спинового транспорта можно описать с помощью уравнения Вале – Ферта вместе с моделью спин-зависимых пограничных сопротивлений. Однако, этот подход был ограничен применением к задачам транспорта спинов только в одном направлении (ось z). Сейчас же, имея полную версию метода НРФГ, можно рассматривать транспорт спинов, ориентированных в любом произвольном направлении, хотя подобные расчеты и требуют бóльших вычислительных ресурсов.

Транслируя матрицу $[G^n]$ в четырехкомпонентный формат $(N,\ \vec{S})$ спин-зависимые граничные проводимости (глава 9) могут быть замещены на матрицу проводимости G формы (4×4), которая свяжет четыре компоненты потенциалов с четырьмя компонентами тока:

$$\begin{Bmatrix} I \\ I_{sx} \\ I_{sy} \\ I_{sz} \end{Bmatrix} = \begin{bmatrix} G \\ 4\times4 \end{bmatrix} \begin{Bmatrix} \Delta E_F \\ \Delta E_F^{sx} \\ \Delta E_F^{sy} \\ \Delta E_F^{sz} \end{Bmatrix}. \tag{203}$$

Аналогичным образом, двухкомпонентное уравнение Вале – Ферта для $(E_F, \vec{E}_F^{\,s})$ может быть конвертировано в четырехкомпонентное уравнение диффузии для $(E_F, \vec{E}_F^{\,s})$.

Как уже упоминалось выше, учет различия между двумя состояниями спина удваивает порядок всех матриц в методе НРФГ. Например, если в 1D проводнике учитываются три атома, то блочная матрица $[G^n]$ будет 6-го порядка:

$$\begin{array}{c|cccccc} & 1up & 1dn & 2up & 2dn & 3up & 3dn \\ 1up & \multicolumn{2}{c}{\left[N_1,\vec{S}_1\right]} & & & & \\ 1dn & & & & & & \\ 2up & & & \multicolumn{2}{c}{\left[N_2,\vec{S}_2\right]} & & \\ 2dn & & & & & & \\ 3up & & & & & \multicolumn{2}{c}{\left[N_3,\vec{S}_3\right]} \\ 3dn & & & & & & \end{array}, \tag{204}$$

в которой диагональные блоки (2×2) соответствуют четырехкомпонентному формату описания транспорта.

Чего не хватает в гриновской матрице (204), так это недиагональных элементов, связывающих соседние и более удаленные друг от друга атомы проводника. Эти элементы гриновской матрицы порождают эффекты квантовой интерференции. Для некоторых транспортных задач это не существенно, поскольку эти эффекты часто нивелируются процессами дефазировки (§ 10.10). В этой связи заметим, что времена спиновой релаксации имеют порядок наносекунд, а времена фазовой релаксации меньше на три порядка и даже более. Раз так, то важно сохранить информацию, хранящуюся в диагональных блоках, даже если не учитываются сведения из недиагональных блоков.

В формализме НРФГ нужно подобрать подходящую матрицу D в уравнении (177/гл.10):

$$\left[\Sigma^{in}\right] = D \times \left[G^{n}\right],$$

поскольку процесс дефазировки можно визуально представить себе как «извлечение» электрона из проводника в состоянии $[G^n]$ и последующей «инжекции» его обратно в проводник в состоянии $D \times G^n$.

В главе 10 были уже рассмотрены две модели для матрицы $[D]$ – модели A и B, описываемые уравнениями, соответственно, (180/гл.10) и (181/гл.10). Модель A эквивалентна умножению матрицы $[D]$ на константу: другими словами, электрон «инжектируется» в то же самое состояние, из которого он был «извлечен», так что импульс электрона сохраняется. В модели B зануляются недиагональные элементы, что ведет к потере информации об импульсе электрона, что хорошо иллюстрируется рис. 38/гл.10.

Определим модель C такую, в которой матрица $[D]$ сохраняет всю информацию о спинах и зануляет импульсную релаксацию, а именно:

$$\frac{\left[D\right]}{D_0} = \begin{array}{c|cccccc} & 1up & 1dn & 2up & 2dn & 3up & 3dn \\ 1up & 1 & 1 & 0 & 0 & 0 & 0 \\ 1dn & 1 & 1 & 0 & 0 & 0 & 0 \\ 2up & 0 & 0 & 1 & 1 & 0 & 0 \\ 2dn & 0 & 0 & 1 & 1 & 0 & 0 \\ 3up & 0 & 0 & 0 & 0 & 1 & 1 \\ 3dn & 0 & 0 & 0 & 0 & 1 & 1 \end{array}, \qquad (205)$$

другими словами, модель C напоминает модель B в отношении решетки проводника, а модель A в отношении спиновой информации.

Перепишем гриновскую матрицу

$$G^{n} = G^{R}\Sigma^{in}G^{A} \qquad (206)$$

как

$$\left[G^{n}\right]_{ii} = \sum \left[G^{R}\right]_{ij}\left[\Sigma^{in}\right]_{jj}\left[G^{A}\right]_{ji} = D_0 \sum \left[G^{R}\right]_{ij}\left[G^{n}\right]_{jj}\left[G^{A}\right]_{ji}, \qquad (207)$$

где индексы i,j нумеруют атомы решетки проводника и учитывается то обстоятельство, что в модели C матрица Σ^{in} в решеточном базисе диагональна.

Мы уже знаем, что для любого узла решетки проводника его диагональную матрицу (2×2) гриновской матрицы (196) можно записать через

числа $(N,\vec{S})$, так что уравнение (199) путем несложных преобразований можно записать в виде

$$\begin{Bmatrix} N \\ S_x \\ S_y \\ S_z \end{Bmatrix}_i = \sum_j \begin{bmatrix} \text{Резонансная} \\ \text{матрица} \\ (4\times 4) \end{bmatrix}_{ij} \begin{Bmatrix} N \\ S_x \\ S_y \\ S_z \end{Bmatrix}_j , \qquad (208)$$

где «резонансная» матрица (4×4) может быть интерпретирована как взаимодействие $(N,\vec{S})$ в узле i с $(N,\vec{S})$ в узлах j, находящихся от i в одном шаге по решетке. Аналогичное одномерное уравнение напоминает стандартное описание броуновского движения на решетке и ведет к диффузионному уравнению дрейфа.

11.15. Квантовая природа классики

Нет ничего удивительного и неожиданного в названии этого раздела. Из разобранных выше примеров мы видели, что для понимания физики и для интерпретации реальных экспериментальных измерений, как правило, достаточно полуклассического рассмотрения. Почти все, что мы до сих пор обсуждали, является следствием транспортного уравнения Больцмана. Более того, все миллиарды транзисторов в современных ноутбуках, состоящих всего из нескольких сотен атомов, по существу являются классическими устройствами электроники, контролируемыми, в основном, электростатикой, а не законами квантовой механики. Недавно выполненные тщательные и всесторонние измерения и расчеты проводимости металлического 2D проводника с сечением (1×4) атомов [61] продемонстрировали его нормальное омическое поведение. Резонно задаться вопросом удастся ли, опираясь на кажущиеся нам фантастическими успехи нанофизики и наноэлектроники, создать принципиально иные квантовые устройства, которые выведут нас за пределы нынешней «зарядовой» парадигмы обработки информации. Коснемся лишь некоторых соображений в этом направлении [62].

11.15.1. Спиновая когерентность

Различие между квантовой физикой и классической лучше всего прослеживается на таком фундаментальном свойстве как спин электрона.

Рассмотрим, например, эксперимент с инжекцией спина в канал проводимости с дальнейшим измерением потенциала вторым зондом (§ 11.14.2)

$$E_{F2} = E_F + \frac{\vec{P}_2 \cdot \vec{E}_F^s}{2}, \qquad (209)$$

зависящим от косинуса угла между инжектирующим и регистрирующим магнитными контактами. Гриновская матрица имела вид (204) с диагональными блоками (2×2)

$$\begin{bmatrix} N+S_z & S_x-iS_y \\ S_x+iS_y & N-S_z \end{bmatrix}, \tag{210}$$

элементы которых содержат физически понятные величины электронной и спиновой плотности $(N,\vec{S})$. Формализм НРФГ приводит далее к уравнению (209) с $(E_F,\vec{E}_F^s)$, транслируемыми в $(N,\vec{S})$ согласно ур-ю (203).

Этот простой пример позволяет проиллюстрировать связь между квантовым и классическим описанием. Если перекрестные компоненты матрицы (210) пренебрежимо малы, то лишь двух чисел N и S_z достаточно для описания всей физики транспорта. Оставшиеся ненулевыми диагональные компоненты могут быть интерпретированы $N+S_z$ как число электронов со спином «вверх», $N-S_z$ как число электронов со спином «вниз» и далее выписаны полуклассические уравнения для двух сортов электронов. Фактически такой подход реализован в главе 9 [63].

Когда работает такая модель? Одна из возможностей предоставляется тогда, когда намагниченности контактов коллинеарны и спин-орбитальное взаимодействие в канале проводимости отсутствует, так что угол θ является целым кратным 180°. Другая возможность реализуется тогда, когда спиновая дефазировка настолько значительна, что перекрестными спиновыми компонентами в диагональных блочных матрицах (2×2) гриновской матрицы (210) можно пренебречь. Наконец, если можно пренебречь и z-компонентами спина, тогда об учете спина речь вообще не идет.

Что если коллинеарные магнитные контакты направлены не вдоль оси z, а вдоль оси x? В этом случае блоки (2×2) гриновской матрицы не диагональны:

$$\frac{[G^n]}{2\pi} \Rightarrow \begin{bmatrix} N & S_x \\ S_x & N \end{bmatrix},$$

и полуклассическое описание скорее всего окажется невозможным. В этом случае нужно иначе выбрать систему координат, или в общем случае иначе выбрать разложение по базисным функциям. Выбрать базис нужно таким образом, чтобы объекты, например, спин, имеющие направления «вверх» и «вниз», были ориентированы, соответственно, вдоль $+x$ и $-x$; тогда блоки гриновской матрицы будут диагональны:

$$\frac{[G^n]}{2\pi} \Rightarrow \begin{bmatrix} N+S_{up} & 0 \\ 0 & N+S_{dn} \end{bmatrix}.$$

Рекомендация иначе выбрать базис может показаться тривиальной, но это не совсем так, в чем мы сейчас убедимся.

11.15.2. Псевдоспин

Формализм спиновых матриц Паули замечателен тем, что он не исчерпывается только спином электрона, а распространяется на любые объекты, имеющих двухкомпонентную структуру в комплексной плоскости. Вспомним графен. Атомы углерода в графене топологически не эквивалентны и могут быть либо сорта A, либо сорта B (гл. 3). Волновая функция элементарной ячейки графена описывается двухкомпонентным вектором в комплексной плоскости

$$\{\psi\}=\begin{Bmatrix}\psi_A\\ \psi_B\end{Bmatrix},$$

так что соответствующая ей функция Грина (10) может быть формально переписана следующим образом:

$$\frac{[G^n]}{2\pi}=\begin{bmatrix}\psi_A\psi_A^* & \psi_A\psi_B^*\\ \psi_B\psi_A^* & \psi_B\psi_B^*\end{bmatrix}\Rightarrow\begin{bmatrix}N+S_z & S_x-iS_y\\ S_x+iS_y & N-S_z\end{bmatrix},$$

что не имеет никакого отношения к реальным спинам, а лишь формально математически соответствует (210).

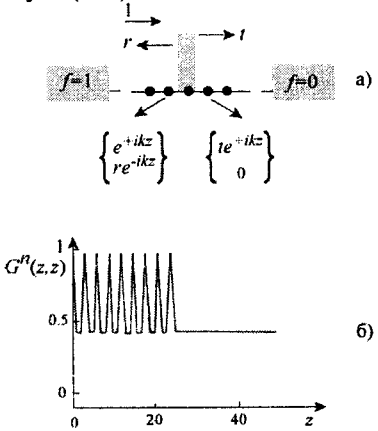


Рис. 48. (а) Схема одномодового проводника с одним дефектом, характеризуемым коэффициентом прохождения t и коэффициентом отражения r; стрелками указаны псевдоспинорные волновые функции электрона до и после прохождения дефекта. (б) Поведение диагональных элементов гриновской матрицы слева и справа от дефекта.

Рассмотрим еще один менее знакомый пример использования концепции псевдоспина, преследуя прежде всего педагогические цели. Вернемся к модели на рис. 36а/гл.10. Посмотрим на поведение диагональных элементов гриновской матрицы $[G^n]$ слева и справа от дефекта (рис. 48б). Слева от дефекта «наблюдаются» осцилляции электронной плотности, тогда как после прохождения дефекта плотность постоянна. Заметим, что в такой же модели на рис. 36 – 40/гл.10 осцилляции имеют место и справа от дефекта. Но там речь шла об осцилляциях заселенности, вычисляемой из G^n/A, а спектральная функция A, пропорциональная плотности состояний D, осциллирует как до, так и после прохождения дефекта, так что к рассматриваемой нами ситуации эти осцилляции справа на рис. 36 – 40/гл.10 отношения не имеют.

Посмотрите как можно используя концепцию псевдоспина понять поведение диагональных элементов гриновской матрицы на рис. 48б. Ниже показаны псевдоспинорные волновые функции электрона и электронные плотности слева/L и справа/R от дефекта и представление последних с использованием концепции псевдоспина:

$$\{\psi\} \to \begin{Bmatrix} e^{+ikz} \\ re^{-ikz} \end{Bmatrix}_L, \quad \begin{Bmatrix} re^{+ikz} \\ 0 \end{Bmatrix}_R;$$

$$\{\psi\}\{\psi\}^+ \to \begin{bmatrix} 1 & r^* e^{+i2kz} \\ re^{-i2kz} & rr^* \end{bmatrix}_L, \quad \begin{bmatrix} tt^* & 0 \\ 0 & 0 \end{bmatrix}_R \to \begin{bmatrix} N+S_z & S_x - iS_y \\ S_x + iS_y & N-S_z \end{bmatrix}.$$

Считая коэффициенты t и r действительными величинами, для электронной псевдоплотности и псевдоспиновой плотности слева и справа от барьера, соответственно, получим:

$$\begin{aligned} N &= (1+r^2)/2 \quad N = r^2/2 \\ S_x &= +r\cos 2kz \quad S_x = 0 \\ S_y &= -r\sin 2kz \quad S_e = 0 \\ S_z &= (1-r^2)/2 \quad S_z = r^2/2 \end{aligned} \qquad (211)$$

Другими словами, слева от барьера псевдоспин вращается в плоскости xy. При построении графика $G^n(z,z)$ на рис. 48б находилась сумма двух компонент псевдоспина, а затем брался ее квадрат, что эквивалентно равенству

$$Tr\{1 \quad 1\}\{\psi\}\{\psi\}^+ \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} = Tr \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \{\psi\}\{\psi\}^+. \qquad (212)$$

Фактически мы имеем дело с псевдомагнитом с

$$\Gamma = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \tag{213}$$

который поляризован на 100% вдоль оси x. Таким образом, согласно (209) замеряемый потенциал должен быть пропорционален $N + \hat{x} \cdot \vec{S}$, что ведет слева от барьера к осцилляциям по косинусу, а справа – к отсутствию осцилляций

$$N + \hat{x} \cdot \vec{S} \to \frac{1+r^2}{2} + r\cos 2kz \Big|_L \; ; \; \frac{r^2}{2}\Big|_R ,$$

что хорошо согласуется с результатами расчета методом НРФГ (рис. 48б).

11.15.3. Квантовая энтропия

Вернемся к полностью поляризованной системе спинов (глава 8) и к ее равновесному состоянию (рис. 49)

Рис. 49. Неравновесная система спинов (А) и ее равновесное состояние (Б).

и к выражению для энтропии (59/гл.8)

$$\frac{S}{k} = -\sum_i p_i \ln p_i . \tag{214}$$

Волновая функция отдельного спина в системе А

$$\psi = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} \text{ дает } \psi\psi^+ = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \tag{215}$$

а диагональные элементы плотности можно интерпретировать как значения p_i в (214). Плотность же $\psi\psi^+$ для спина в системе Б требует суммирования двух равноправных возможностей:

$$\psi\psi^+ = 0.5\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + 0.5\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}, \tag{216}$$

что дает с такой плотностью уже известный ответ (60/гл.8).

Для системы же В (рис. 50), в которой спины ориентированы вдоль оси x, а не z, как в системе А, волновая функция

$$\psi = \frac{1}{\sqrt{2}}\begin{Bmatrix}1\\1\end{Bmatrix} \text{ дает } \psi\psi^{+} = \begin{bmatrix}0.5 & 0.5\\0.5 & 0.5\end{bmatrix}. \qquad (217)$$

Рис. 50. Система В это система А, переориентированная вдоль оси x.

Если мы опять посчитаем энтропию по диагональным элементам (217), то получим, что энтропия системы В такая же, как и системы Б, что явно неверно. Она должна быть такая же как и системы А.

Правильный ответ получится, если мы сначала диагонализуем плотность $\psi\psi^{+}$, а потом полученные диагональные элементы как собственные значения плотности используем как вероятности p_i в (214). Это достигается путем обобщения (214) следующим образом:

$$\frac{S}{k} = -Tr\left[\rho \ln \rho\right], \qquad (218)$$

где матрица плотности $\rho = \psi\psi^{+}$.

11.15.4. Сколько информации переносит один электрон?

Воспользуемся для передачи информации устройством (рис. 35), в котором намагниченность регистрирующего контакта может меняться в пределах $\pi/2$ относительно намагниченности инжектирующего контакта с мелким шагом, скажем 10°.

На первый взгляд кажется, что инжектируя один электрон таким образом можно передать большой объем информации, поскольку есть много возможностей зарегистрировать ориентацию спина электрона детектором. При шаге в 10° объем переданной информации вроде бы был бы равен $S/k = \ln 10$ вместо $\ln 2$. Ведь ранее мы всегда считали, что одному электрону всегда соответствует $S/k = \ln 2$. Эти два ответа на самом деле не противоречат друг другу. Ведь для измерения спинового тока, зависящего от угла θ (глава 11), к детектору нужно подвести очень много электронов с тем, чтобы должным образом усреднить результаты измерений. При установке регистрирующего детектора на определенный угол θ подошедший к детектору электрон может либо передать информацию, если его поляризация соответствует поляризации детектора, либо нет. Передать действительно можно информацию объемом $\ln 10$, но для этого нужно пропустить через детектор множество должным

образом подготовленных электронов с тем, чтобы детектор выдал правдоподобный результат.

А можно ли клонировать электрон в определенном состоянии в виде множества идентичных электронов с одинаковыми волновыми функциями? Теоремы в квантовой механике, запрещающей клонирование, нет. Просто нужно инжектировать много электронов, приготовленных одинаковым образом, а затем грамотно произвести обработку результатов измерений.

Все эти вопросы являются предметом рассмотрения быстро развивающейся квантовой информатики. Но уже сейчас стало очевидным, что спин электрона по сравнению с его зарядом открывает качественно новые перспективы в обработке информации.

11.15.5. Увеличивает ли энтропию взаимодействие частиц?

Вернемся к информационно-управляемому аккумулятору (§ 8.8), по ходу работы которого энтропия растет. А как именно это происходит?

Взаимодействие электронов и локализованных спинов описывалось обменной реакцией (80/гл.8). С точки зрения квантовой механики обменное взаимодействие порождает суперпозицию волновых функций (рис. 51).

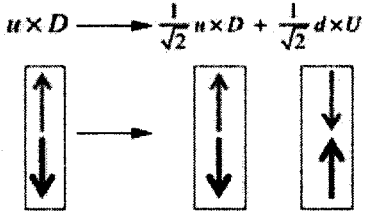


Рис. 51. Обменное взаимодействие в реакции (80/гл.8) порождает суперпозиционное состояние.

Суперпозиционное состояние может быть не таким простым, как на рис. 51, но в любом случае сумма квадратов комплексних коэффициентов должна быть равна единице.

Важно то, что суперпозиционное состояние

$$\frac{1}{\sqrt{2}}u\times D+\frac{1}{\sqrt{2}}d\times U$$

имеет нулевую энтропию как и исходное состояние $u\times D$. Действительно, можно прибегнуть к технике псевдоспина (§ 11.15.2), выбрав его компонентами «вверх» и «вниз» состояния $u\times D$ и $d\times U$. Обменное взаимодействие фактически только вращает псевдо-спин с направления вдоль оси z на

направление вдоль оси x, а выше уже было показано, что вращение системы спинов (псевдо-спинов) не производит энтропию.

Как же все-таки происходит увеличение энтропии? Канальные электроны в проводнике рано или поздно покидают проводник. В момент ухода суперпозиционное состояние разрушается («коллапс волновой функции») и становится либо состоянием $u \times D$, либо $d \times U$, в зависимости от того, канальный электрон уходит через исток со спином «вверх», либо через сток со спином «вниз». Локализованный спин остается с равной вероятностью 50% либо в состоянии «вниз», либо «вверх». Это тот момент, когда энтропия увеличивается на $k \ln 2$.

11.15.6. Спины и магниты

Спины с момента своего экспериментального обнаружения сразу стали нетривиальным свойством частиц. Уже научились с помощью твердотельных устройств управлять и манипулировать не только отдельными спинами, но и такими странными квантовыми объектами как перепутанные спины [64]. Многие квантовые объекты допускают псевдо-спиновую трактовку как двухкомпонентных спинов. Однако, совсем не обязательно ограничиваться двухкомпонентными объектами. При рассмотрении транспорта спинов в четырехкомпонентном формате (§ 11.14.3) диагональные блоки (2×2) гриновской матрицы G^n содержали четыре действительных числа $(N, \bar{S})$. Можно пойти дальше и рассматривать блоки, скажем, (12×12), соответствующие многоатомным молекулярным орбиталям, сохраняющим фазовую когерентность в течение относительно длительного времени, как гигантские квази-спиновые объекты. Нам потребуется 144 действительных компонент, чтобы отразить всю информацию, содержащуюся в эрмитовой матрице 12×12, а затем воспользоваться методом НРФГ и изучить диффузию таких гигантских квази-спиновых объектов так, как это уже проводилось выше для спинов.

Можем ли мы использовать эти гигантские квази-спины для кодирования и передачи информации? Вероятно, да, но есть одна до сих пор нерешенная проблема – интеграция спинтроники и магнетроники. Осознали эту проблему всего лет 10 назад. Спинтроника занималась манипулированием и управлением индивидуальными спинами и мотивировалась физикой сверхнизких температур. Магнетроника носила более прикладной характер и занималась классическими магнитами при комнатных температурах, магнетизм которых был результатом взаимодействия и упорядочения громадного числа элементарных магнитных моментов.

Первым шагом на пути интеграции этих двух фундаментальных направлений исследований явилось использование магнитных контактов для инжектирования спинов, что привело к спиновым вентилям и магнитным туннельным переходам, лежащим в основе современных технологий

считывания информации с магнитных носителей. Вторым шагом стала демонстрация транспорта магнитного момента, что позволило обращать намагниченность образца (§ 9.6), если только его толщина не превышала нескольких атомных слоев. Этот эффект можно было бы использовать для записи информации на наномагниты. Устройства памяти, использующие спин как для чтения, так и для записи информации сейчас интенсивно разрабатываются.

Для передачи и управления информацией нужны не только спиновые устройства, но и магнитные. В нынешней зарядовой архитектуре компьютеров для хранения информации используются конденсаторы. В предстоящей спиновой архитектуре нужны магниты для создания спиновых конденсаторов и устройства для передачи информации от магнита к магниту. Последние 10 лет научных и технологических исследований вселяют надежду на создание компьютеров со спиновой структурой хранения и передачи информации. Надежды связывают с молекулярными магнитами [65, 66].

Напрашивается также идея выйти за пределы обычных спинов и использовать разнообразные экзотические псевдо-спины и даже многокомпонентные квази-спины, сохраняющие когерентность в течение достаточно длительного времени. Ключевым здесь будет вопрос о наличии квази-магнитов для генерирования и детектирования квази-спинов.

Подведем итоги. В рамках концепции «снизу – вверх» современной наноэлектроники рассмотрены классический и квантовые эффекты Холла, методы измерения электрохимических потенциалов, формулы Ландауэра и Бюттекера, измерение холловского потенциала, учет магнитного поля в методе НРФГ, уровни Ландау и краевые состояния в графене, спиновый транспорт в формализме НРФГ в спинорном представлении, в частности, подробно обсуждены спиновый вентиль, вращение магнитных контактов, прецессия спина и вращение спинов, роль спиновых гамильтонианов Зеемана и Рашбы, квантовый спиновый эффект Холла, вычисление спинового потенциала, а также четырехкомпонентный формат описания спинового транспорта.В заключение обсуждаются некоторые аспекты квантовой природы классики [67 – 69].

Литература

1. E. H. Hall. On a New Action of the Magnet on Electric Currents, *Amer. J. Math.*, v. 2, n 3, 287 – 292 (1879).
2. K. von Klitzing, G. Dorda, M. Pepper. New Method for High-Accuracy Determination of the Fine-Structure Constant Based on Quantized Hall Resistance, Phys. Rev. Lett., v. 45, 494 – 497 (1980).

3. К. фон Клитцинг. Квантовый эффект Холла: Нобелевские лекции по физике, *УФН,* т. 150, № 1, 107 – 126 (1985).
4. G. M. Reedtz, M. E. Cage. An Automated Potentiometric System For Precision Measurement Of the Quantized Hall Resistance, *J. Res. Nat. Bureau Stand.*, v. 92, n 5, 303 – 310 (1987).
5. D. S. Tsui, H. L. Stormer, A. C. Gossard. Two-dimensional magnetotransport in the extreme quantum limit, *Phys. Rev. Lett.*, v. 48, n 22, 1559 – 1962 (1982).
6. Ю. П. Степановский. Дробный квантовый эффект Холла, Электромагн. явления, т. 1, № 3, 427 – 442 (1998).
7. N. Nagaosa, J. Sinova, S. Onoda, A. H. MacDonald, N. P. Ong. Anomalous Hall effect, *Rev. Mod. Phys.*, v. 82, n 2, 1539 – 1592 (2010).
8. M. I. Dyakonov, V. I. Perel. Possibility of orientating electron spins with current, *Sov. Phys. JETP Lett.*, v. 13, 467 (1971).
9. M. I. Dyakonov, V. I. Perel. Current-induced spin orientation of electrons in semiconductors, *Phys. Lett. A*, v. 35, n 6, 459 – 460 (1971).
10. C. L. Kane, E. J. Mele. Quantum Spin Hall Effect in Graphene, *Phys. Rev. Lett.*, v. 95, 226801/1 – 4 (2005).
11. S. Srinivasan, A. Sarkar, B. Behin-Aein, S. Datta. All-Spin Logic Device with inbuilt Non-Reciprocity, *IEEE Trans. Magnetics*, v. 47, n 10, 4026 – 4032 (2011).
12. C. L. Kane, J. E. Moore. Topological Insulators, Physics World, v. 24, 32 – 36 (2011).
13. S. Datta. *Lessons from Nanoelectronics: A New Perspective on Transport.* Hackensack, New Jersey: World Scientific Publishing Company, 2012.
14. Ю. А. Кругляк. Наноэлектроника «снизу – вверх»: Возникновение тока, обобщенный закон Ома, упругий резистор, моды проводимости, термоэлектричество, *ScienceRise*, т. 7, № 2(12), 76 – 100 (2015).
15. Ю. О. Кругляк, Н. Ю. Кругляк, М. В. Стріха. Уроки наноелектроніки: Виникнення струму, формулювання закону Ома і моди провідності в концепції «знизу – вгору», *Сенсор. електр. мікросист. техн.*, т. 9, № 4, 5 – 29 (2012)..
16. Н. Ашкрофт, Н. Мермин. *Физика твердого тела.* Москва: Мир, 1979.
17. E. I. Rashba. Spin Currents in Thermodynamic Equilibrium: The Challenge of Discerning Transport Currents, *Phys. Rev. B*, v. 68, n 24, 241315R (2003).
18. M. Buttiker. Symmetry of Electrical Conduction, *IBM J. Res. Dev.*, v. 32, n 3, 317 – 334 (1988).
19. Ю. А. Кругляк. Наноэлектроника «снизу – вверх»: метод неравновесных функций Грина, модельные транспортные задачи и квантовая интерференция, *ScienceRise*, т. 9, № 2(14), 41 – 72 (2015).
20. Ю. А. Кругляк. Графен в транспортной модели Ландауэра – Датты – Лундстрома, *ScienceRise*, т. 2, № 2(7), 93 – 106 (2015).
21. Ю. А. Кругляк. Наноэлектроника «снизу – вверх»: Начала спинтроники и магнетроники, *ScienceRise*, т. 8, № 2(13), 51 – 68 (2015).

22. Sears F.W., Salinger G.L. Thermodynamics, Kinetic Theory, and Statistical Thermodynamics. – Boston: Addison-Wesley. – 1975.
23. R. Landauer. Spatial variation of currents and fields due to localized scatterers in metallic conduction, *IBM J. Res. Dev.*, v. 1, n.3, 223 – 231 (1957).
24. R. Landauer. Electrical resistance of disordered one dimensional lattices, *Philos. Mag.*, v. 21, 863 – 867 (1970).
25. R. Laundauer. Spatial variation of currents and fields due to localized scatterers in metallic conduction, *IBM J. Res. Dev.*, v. 32, 306 (1988).
26. R. Landauer. Electrical resistance of disordered one dimensional lattices, *J. Math. Phys.*, v. 37, n. 10, 5259 (1996).
27. M. Buttiker. Four-terminal phase-coherent conductance, *Phys.Rev. Lett.*, v. 57, 1761 (1986).
28. Ю. В. Шарвин. Об одном возможном методе исследования поверхности Ферми, *ЖЭТФ*, т. 48, № 3, 984 – 985 (1965); *Sov. Phys. JETP*, v. 21, 655 – 656 (1965).
29. Yu. V. Sharvin, N. I. Bogatina. Investigation of Focusing of Electron Beams in a Metal by a Longitudinal Magnetic Field, *Sov. Phys. JETP*, v. 29, n 3, 419 – 423 (1969); *ЖЭТФ*, т. 56, № 3, 772 – 779 (1969).
30. Y. Imry in *Directions in Condensed Matter Physics*, ed. G. Grinstein, G. Mazenko, p. 101. Singapore: World Scientific, 1986.
31. Y. Imry, R. Laundauer. Conductance viewed as transmission, *Rev.Mod.Phys.*, v. 71, n 2, S306 – S312 (1999).
32. Г. Б. Лесовик, И. А. Садовский. Описание квантового электронного транспорта с помощью матриц рассеяния, *УФН*, т. 181, № 10, 1041 – 1096 (2011).
33. A. D. Stone, A. Szafer. What is measured when you measure a resistance? The Landauer formula revisited, *IBM J. Res. Dev.*, v. 32, n 3, 384 – 413 (1988).
34. R. G. Mojarad, A. N. M. Zainuddin, G. Klimeck, S. Datta. Atomistic non-equilibrium Green's function simulations of graphene nano-ribbons in the quantum hall regime, *J. Comput. Electron.*, v. 7, 407 – 410 (2008).
35. R. J. Haug. Edge-state transport and its experimental consequences in high magnetic fielgs, *Semicond. Sci. Technol.*, v. 8, 131 – 153 (1993).
36. M. E. Cage. Current Distributions in Quantum Hall Effect Devices, *J. Res. Natl. Inst. Stand. Technol.*, v. 102, 677 (1997).
37. Дж. М. Мартинес-Дуарт, Р. Дж. Мартин-Пальма, Ф. Агулло-Руеда. *Нанотехнологии для микро- и оптоэлектроники*. Москва: Техносфера, 2007.
38. C. Berger, S. Zhimin, Xuebin Li, Xiaosong Wu, N. Brown, C. Naud, D. Mayou, Tianbo Li, J. Hass, A. N. Marchenkov, E. H. Conrad, P.N. First, W. A. de Heer. Electronic Confinement and Coherence in Patterned Epitaxial Graphene, *Science*, v. 312, 1191 – 1196 (2006).

39. L. Brey, H. A. Fertig. Edge states and quantized Hall effect in Graphene, *Phys. Rev. B*, v. 73, 195408 (2006).
40. N. M. Peres, A. H. Castro Neto, F. Guinea, *Phys. Rev. B*, v. 73, 195411 (2006).
41. D. A. Abanin, P. A. Lee, L. S. Levitov. Spin-filtered edge states and quantum Hall effect in Graphene, *Phys. Rev. Lett.*, v. 96, 176803 (2006).
42. R. Kubo. Statistical-Mechanical Theory of Irreversible Processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems, *J. Phys. Soc. Japan*, v. 12, 570 – 586 (1957).
43. W. Yourgrau, A. van der Merwe, G. Raw. *Treatise on Irreversible and Statistical Thermophysics*. New York: Dover Publication, 1982.
44. Ю. А. Кругляк. Модель проводимости Ландауэра – Датты – Лундстрома в микро- и наноэлектронике и транспортное уравнение Больцмана, *ScienceRise*, т. 3, № 2(8), 108 – 116 (2015).
45. E. Doniach, E.H. Sondheimer. *Green's Functions for Solid State Physicists*. London: Imperial College Press, 1999.
46. Ya. Blanter, M. Buttiker. Noise in Mesoscopic Conductors, *Phys. Reports*, v. 336, 1 – 166 (2000).
47. M. Z. Hasan, C. L. Kane. Topological insulators, *Rev. Mod. Phys.*, v. 82, 3045 (2010).
48. Ю. А. Кругляк, Х. Прейсс, Р. Яношек. Расчет электронных оболочек бензильного радикала неограниченным методом Хартри – Фока на гауссовом базисе, *Ж. структ. хим.*, т. 12, № 4, 689 – 696 (1971).
49. Ю. А. Кругляк, Обобщенный метод Хартри – Фока и его версии: от атомов и молекул до полимеров, *ScienceRise*, т. 5, № 3 (5), 6 – 21 (2014).
50. Yu. A. Kruglyak, I. I. Ukrainsky. Study of the electronic structure of alternant radicals by the DODS method, *Intern. J. Quantum Chem.*, v. 4, n 1, 57 – 72 (1970).
51. Yu. A. Kruglyak. Quantum-chemical studies of quasi-one-dimensional electron systems. 1. Polyenes, *ScienceRise*, т. 5, № 2 (10), 69 – 105 (2015).
52. Э. И. Рашба. Свойства полупроводников с петлей экстремумов. I. Циклотронный и комбинированный резонанс в магнитном поле, перпендикулярном плоскости петли, *ФТТ*, т. 2, № 6, 1224 – 1238 (1960); E. I. Rashba E.I. Properties of semiconductors with an extremum loop. 1. Cyclotron and combinational resonance in a magnetic field perpendicular to the plane of the loop, *Sov. Phys. Solid. State*, v. 2, 1109 (1960).
53. Yu. A. Bychkov, E. I. Rashba. Oscillatory effects and the magnetic susceptibility of carriers in inversion layers, *J. Phys. C*, v. 17, n 33, 6039 – 6045 (1984).
54. R. Winkler. *Spin-Orbit Coupling Effects in Two-Dimensional Electron and Hole Systems*. Berlin: Springer, 2003.
55. W. Hanle. Uber magnetische Beeinflussung der Polarisation der Resonanzfluorescenz, *Z. Physik*, v. 30, n 1, 93 – 105 (1924).

56. R. Van Dyck, J. Stoltenberg, D. Pengra. *The Hanle Effect*. Washington: University of Washington, 2006.
57. B. Huang, Jang Hyuk-Jae, Ian Appelbaum. Geometric dephasing-limited Hanle effect in long distance lateral silicon spin transport devices, *Appl. Phys. Lett.*, v. 93, n 16, 162508/1 – 3 (2008).
58. H. C. Koo, J. H. Kwon, J. Eom, J. Chang, S. H. Han, M. Johnson. Control of Spin Precession in a Spin-Injected Field Effect Transistor, *Science*, v. 325, . 1515 (2009).
59. J. Wunderlich, Park Byong-Guk, A. C. Irvine, L. P. Zârbo, E. Rozkotová, P. Nemec, V. Novák, J. Sinova, T. Jungwirth. Spin Hall Effect Transistor, *Science*, v. 330, n 6012, 1801 – 1804 (2010).
60. V. Sih, W. H. Lau, R. C. Myers, V. R. Horowitz, A. C. Gossard, D. D. Awschalom. Generating Spin Currents in Semiconductors with the Spin Hall Effect, *Phys. Rev. Lett.*, v. 97, n 9, 096605/1 – 4 (2006).
61. B. Weber, S. Mahapatra, H. Ryu, S. Lee, A. Fuhrer, T. C. G. Reusch, D. L. Thompson, W. C. T. Lee, G. Klimeck, L. C. L. Hollenberg, M. Y. Simmons. Ohm's Law Survives to the Atomic Scale, *Science*, v. 335, 64 – 67 (2012).
62. W. H. Zurek. Decoherence, Einselection and the Quantum Origins of the Classical, *Rev. Mod. Phys.*, v. 75, 715 – 775 (2003).
63. Ю. О. Кругляк, Н. Ю. Кругляк, М. В. Стриха. Уроки наноелектроніки: спінтроніка в концепції «знизу - вгору», *Сенсор. електр. мікросист. техн.*, т. 10, № 4, 5 – 37 (2013).
64. R. Horodecki, P. Horodecki, M. Horodecki, K. Horodecki. Quantum entanglement, *Rev. Mod. Phys.*, v. 81, n 2, 865 – 942 (2007).
65. J. R. Friedman, M. P. Sarachik. Single-Molecule Nanomagnets, *Ann. Rev. Cond. Matter Phys.*, v.1, 109 – 128 (2010).
66. Molecular Magnetism Web. *A gate to molecular magnetism*: www.molmag.de.
67. Ю. О. Кругляк, М. В. Стриха. Уроки наноелектроніки: ефект Холла і вимірювання електрохімічних потенціалів у концепції «знизу - вгору», *Сенсор. електр. мікросист. техн.*, т. 11, № 1, 5 – 27 (2014).
68. Ю. А. Кругляк, П. А. Кондратенко, Ю. М. Лопаткин. Наноэлектроника: эффект Холла и измерение электрохимических потенциалов в концепции «снизу – вверх», *Ж. нано- электр. физ.*, т. 7, № 2, 02031(15pp) (2013).
69. Ю. А. Кругляк. Наноэлектроника «снизу – вверх»: Эффекты Холла, измерение электрохимических потенциалов и транспорт спинов в модели НРФГ, *ScienceRise*, т. 10, № 2 (15), 35 – 67 (2015).

Глава 12. Кулоновская блокада и одноэлектронный нанотранзистор на молекуле бензола

12.1. За пределами метода НРФГ

До сих пор мы рассматривали квантовый транспорт электронов в упругом проводнике методом неравновесных функций Грина (НРФГ) в матричном представлении [1 – 3]. Проводник описывался матрицей гамильтониана $[H]$, взаимодействие между каналом проводимости и контактами 1 и 2 задается матрицами собственной энергии $[\Sigma_1]$ и $[\Sigma_2]$, а взаимодействие электрона в канале с его окружением описывается матрицей собственной энергии $[\Sigma_0]$, которое может рассматриваться как дополнительный виртуальный контакт (рис.1).

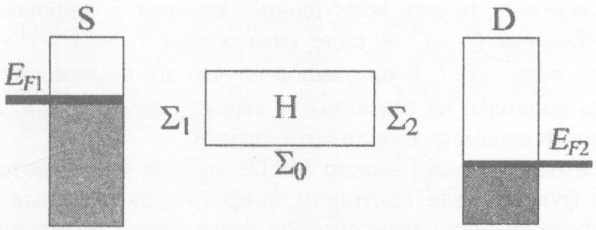


Рис. 1. Проводник взаимодействует с контактами ($[\Sigma_1]$ и $[\Sigma_2]$) и окружающей средой ($[\Sigma_0]$). Поданное на проводник напряжение определяется разностью электрохимических потенциалов $E_{F1} - E_{F2} = qV$.

Если перечисленные входные матрицы заданы, задача сводится к четырем уравнениям (6), (13), (14) и (16) в [1, 3], решение которых дает всю электрофизику рассматриваемого устройства. Однако, что касается взаимодействия внутри канала $[\Sigma_0]$, мы либо его вообще не учитывали (когерентный транспорт), либо рассматривали лишь процессы упругой дефазировки, описываемые уравнениями (1) в [2] и (176) – (178) в [3]. Подобный выбор $[\Sigma_0]$ не позволяет учесть обмен энергией с окружающей средой, но влияние на транспорт электронов оказывает в результате учета обмена импульсами и «фазами» [2, 3]. Фактически в этих работах речь шла об упругих резисторах с учетом квантовомеханических эффектов. Общий метод НРФГ применялся к упругим резисторам по аналогии с рассмотрением упругих проводников с использованием общего транспортного уравнения Больцмана (ТУБ) (глава 5) [4 – 6].

Как выйти за пределы упругой модели проводника? Что касается квазиклассических моделей электронного транспорта, то в принципе известно как учесть самые различные взаимодействия в ТУБ, и в этом направлении достигнут значительный прогресс. Аналогичная ситуация имеет место и при рассмотрении квантовых моделей транспорта электронов в рамках метода

НРФГ. Рассмотрение упругой дефазировки в [2, 3] демонстрирует учет лишь некоторых взаимодействий, а достаточно полный учет взаимодействий можно найти в классических работах, например, [7 – 9].

Моделирование проводников с симметричными контактами (рис. 28 в [10]) не может тестировать физику неупругого транспорта и служить критерием различия между хорошей теоретической моделью и неудовлетворительной. Хорошей проверкой моделей неупругого рассеяния может служить канал проводимости как на рис. 29 в [10], в котором весь ток на обоих контактах обусловлен неупругими процессами.

А вообще, в принципе, может ли метод НРФГ учесть всё и вся?

Метод НРФГ был разработан в начале 60-х годов на основе квантовой многочастичной теории возмущений (МЧТВ) [11 – 13]. Казалось, что МЧТВ может, в принципе, учесть любые взаимодействия. Но это не совсем так, поскольку это все-таки теория возмущений, которая в широком понимании напоминает вычисление $(1-x)^{-1}$ в виде разложения $1+x+x^2+x^3+...$, которое хорошо работает, если $|x|<1$. В противном случае это разложение не работает, и нужны другие подходы, не связанные с теорией возмущений, либо удается найти другой малый параметр теории возмущений.

С равновесными задачами вместо МЧТВ хорошо справляются, например, методы теории функционала плотности, широко используемые в квантовой химии [14]. Моделирование неравновесных задач намного сложнее, поскольку в отсутствии равновесия возникает целый спектр многочастичных состояний, взаимодействующих с контактами сложным образом. Обычно же в той или иной специальной задаче физику транспортного процесса удается так или иначе успешно моделировать.

Есть, однако, экспериментально наблюдаемые факты, например, явление кулоновской блокады в одноэлектронике [15 – 17], когда традиционная хартри – фоковская модель учета электронного взаимодействия в рамках теории самосогласованного поля не работает. В будущем, возможно, и удастся так изощренно построить матрицы собственной энергии $[\Sigma]$ и $[\Sigma^{in}]$ в методе НРФГ, чтобы распространить его и на подобные явления.

12.2. Кулоновская блокада в одноэлектронике

Следуя духу концепции «снизу – вверх» по аналогии с простейшим одноуровневым резистором с одной модой проводимости при энергии ε (§3.1 в [18]), рассмотрим двухмодовый резистор, каждый из двух уровней которого занят одним электроном – один со спином «вверх», второй – со спином «вниз», находящихся при одной и той же энергии ε (рис. 2).

Также предположим, что электрохимический потенциал E_F совпадает точно с энергией уровней ε, так что каждый из уровней заполнен наполовину, поскольку фермиевская функция $f_0|_{\varepsilon=\mu}=0.5$. Следует ожидать высокой проводимости поскольку электрохимический потенциал лежит строго посередине каждого из двух одинаково размытых уровней энергии ε [18].

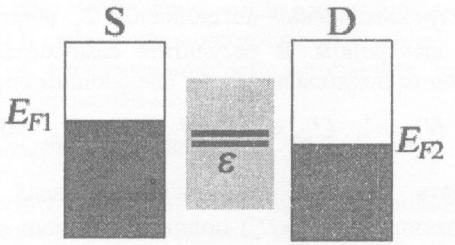


Рис. 2. Исходное состояние наполовину заполненных уровней в явлении кулоновской блокады.

Однако, если удельная (в пересчете на один электрон) энергия заряжания проводника U_0 [10] достаточно велика, то спектр энергий в проводнике изменяется таким образом, что один из уровней вместе с электроном поднимается выше на величину U_0 (рис. 3).

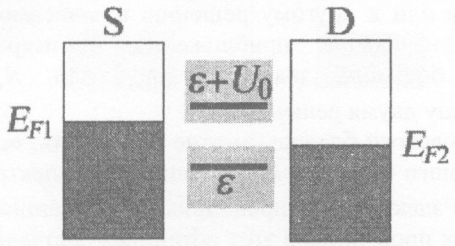


Рис. 3. Кулоновская блокада с позиции концепции «снизу – вверх»: двухмодовый канал проводимости (рис. 2) обнаруживает значительно меньшую плотность состояний в области $E = E_F$ и стало быть большее сопротивление, если удельная энергия заряжания проводника U_0 достаточно велика.

Почему вверх не поднимается и второй уровень вместе с электроном? Потому что уровень энергии с электроном не может чувствовать потенциал, создаваемый своим же электроном. Этот эффект самовзаимодействия не учитывается в ограниченном хартри – фоковском методе самосогласованного поля, когда мы раньше писали [10], что $U = U_0 N$. Нужен неограниченный подход в методе самосогласованного поля [19 – 21], когда каждый уровень i испытывает действие потенциала U_i, который зависит от изменения числа электронов, занимающих все уровни кроме i-го:

$$U_i = U_0 (N - N_i). \qquad (1)$$

Если использовать (1) вместо $U = U_0 N$, мы получим картину заселенности как на рис. 3 с электрохимическим потенциалом E_F вблизи моды проводника, заселенной одним электроном. В результате самосогласованное решение в рамках неограниченного подхода выглядит следующим образом:

$$N_{dn} = 1,\ U_{up} = U_0,\ N_{up} = 0,\ U_{dn} = 0,$$

что нужно понимать как то, что нижний уровень заселен ($N_{dn} = 1$) и результирующий потенциал ($U_{up} = U_0$) поднимет уровень электрона со спином «вверх» и этот уровень будет пустым ($N_{up} = 0$). Поскольку поднятый уровень не заселен, потенциал, испытываемый нижним уровнем, равен нулю ($U_{dn} = 0$), и он остается на прежнем месте.

Естественно, эта картина симметрична относительно замены «вверх» на «вниз»:

$$N_{up} = 1,\ U_{dn} = U_0,\ N_{dn} = 0,\ U_{up} = 0$$

также является решением. Численно процедура самосогласования будет сходиться к одному или к другому решению в зависимости, в частности, от того, будет ли начальное приближение в итерационном процессе характеризоваться бóльшим значением N_{up} или N_{dn}, или же будет флюктуировать между двумя решениями.

Явление кулоновскогй блокады можно наблюдать, если удельная энергия заряжания U_0 намного больше kT и уширения спектра проводника. При добавлении одного электрона в проводник его потенциал увеличивается на U_0/q. Для массивных проводников этот потенциал составляет микровольты или даже меньше и не наблюдаем даже при очень низких температурах. В конце концов, любое энергетическое свойство уширяется на kT, составляющее $\sim 25\ мэВ$ при комнатной температуре и $\sim 200\ мкэВ$ при температуре $\sim 1K$. Эффект заряжания проводника, вызывающий явление кулоновской блокады, становится наблюдаемым по крайней мере при низких температурах, если проводник достаточно мал, чтобы энергия U_0 стала порядка, скажем, 1 *мэВ*. Для проводников молекулярного размера энергия U_0 может быть сотни мэВ, что делает кулоновскую блокаду наблюдаемой даже при комнатной температуре.

Есть и второй фактор, ограничивающий наблюдаемость эффекта кулоновской блокады. В дополнение к температурному уширению $\sim kT$ есть еще более фундаментальное уширение $\gamma \sim h/t$, связанное с временем пролета. Эффект заряжания проводника наблюдаем только если кулоновская щель U_0 существенно превышает это уширение: $U_0 \gg h/t$. Наблюдение эффекта кулоновской блокады даже для самых малых проводников предполагает наличие хороших контактов в измерительной схеме.

Позже условия наблюдаемости кулоновской блокады мы обсудим в традиционном контексте через контактные сопротивления и параметризацию всех взаимодействий в канале проводимости суммарной емкостью [22, 23]. А сейчас перейдем к рассмотрению вольт-амперных характеристик (ВАХ) рассматриваемой двухуровневой модели.

12.2.1. Вольт-амперные характеристики

Упрощая рассмотрение двухмодовой модели проводника, пусть его уровни энергии фиксированы относительно истока S и не подвержены электростатическому влиянию [10] со стороны стока D. Пренебрегая электрон-электронным взаимодействием ($U = 0$), имеем ВАХ качественно как на рис. 4.

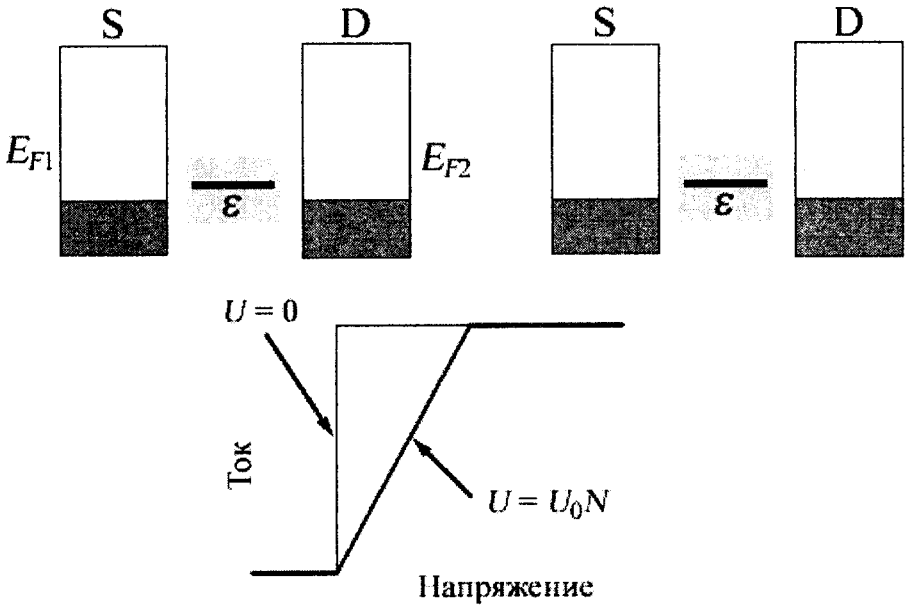


Рис. 4. Качественный ход ВАХ двухмодового резистора при $U = 0$ и $U = U_0 N$.

Как только электрохимический потенциал E_{F2} пересекает энергию уровней ε, ток тотчас достигает своего максимального значения.

Если теперь учесть эффект заряжания проводника через самосогласованный потенциал $U = U_0$, всплеск тока протянется по шкале напряжения на $\sim U_0/q$, поскольку заряжание поднимает уровни вверх и требуется большего напряжения полностью покрыть уровни энергии проводника.

Если же учесть эффект самовзаимодействия в неограниченном подходе (1) мы получим ВАХ с промежуточным плато, происхождение которого очевидно из диаграмм на рис. 5.

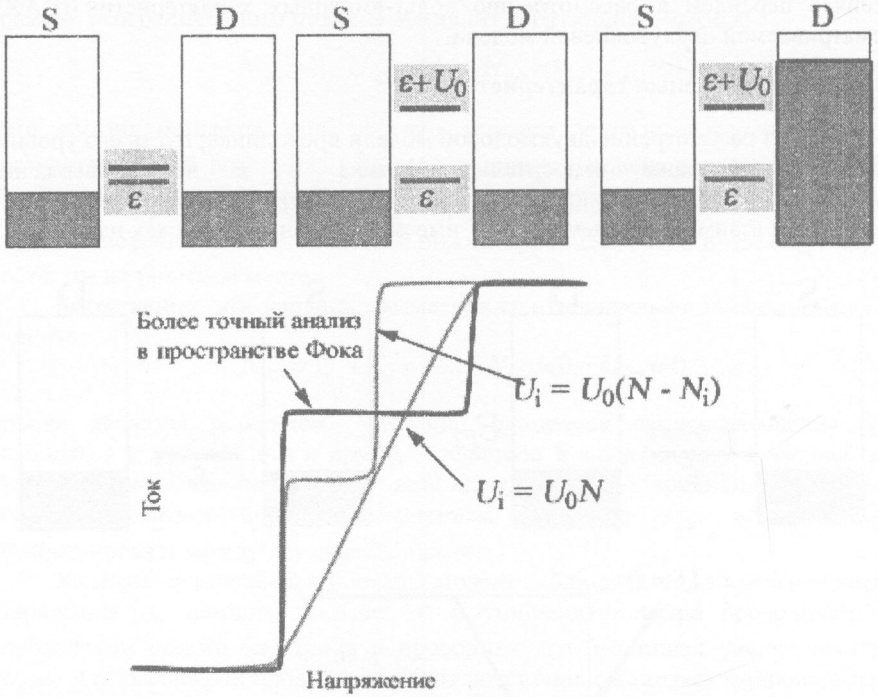


Рис. 5. Более точное положение плато на ВАХ согласно анализу в пространстве Фока.

Как показано на диаграммах, поначалу только нижний уровень дает половинный вклад в полную проводимость (срединная диаграмма), а полный ток достигается лишь тогда (диаграмма справа), когда подан достаточный потенциал μ_2, чтобы пересечь уровень с энергией $\varepsilon + U_0$ и получить полный ток.

Это промежуточное плато действительно неоднократно наблюдалось экспериментально, но количественно его высота не половинная, а составляет, по-видимому, две трети от полного тока $2q/t$. Эту разницу невозможно объяснить и учесть в рамках одночастичной модели проводника. Однако, она получается естественным образом при переходе к пространству Фока.

12.2.2. Моделирование заряжания проводника в пространстве Фока

Мы уже пользовались описанием квантового состояния проводника в пространстве Фока [24, 25]. Если в одночастичной модели мы говорим о заполнении или опустошении уровней энергии проводника электронами, поставляемыми контактами, то фоковское представление дает возможность

говорить о переходе всей системы электронов проводника из одного состояния в другое.

Так, рис. 6 демонстрирует одноуровневый резистор в одночастичном представлении и в пространстве Фока: состояние 0 – пустое, а состояние 1 – заселенное.

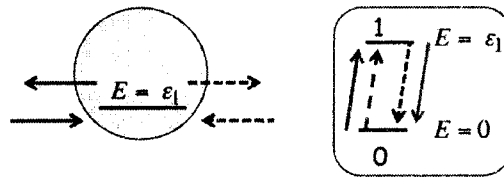


Рис. 6. Одномодовый резистор в одночастичном представлении (слева) и в фоковском пространстве (справа).

На рис. 7 одночастичная модель двухуровневого резистора сравнивается с его же фоковским представлением, демонстрирующим четыре возможных состояния проводника, отличающихся заселенностью этих состояний электронами.

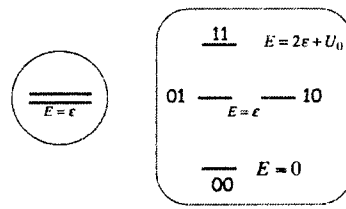


Рис. 7. Двухмодовый резистор в одночастичном представлении (слева) и в фоковском пространстве (справа).

В общем случае N-уровневый резистор имеет 2^N состояний в пространстве Фока. Мы еще вернемся к рассмотрению общего случая.

А сейчас обратимся к описанию равновесия в пространстве Фока. Процедура вычисления вероятности найти систему в состоянии i в условиях равновесия хорошо известна [24, 25]:

$$p_i = \frac{1}{Z} e^{-(E_i - E_F N_i)/kT}, \qquad (2)$$

где Z – нормировочная постоянная.

Пусть мы хотим найти равновесное число электронов n в двухмодовом проводнике (рис. 7) в зависимости от электрохимического потенциала E_F. Для n в фоковском пространстве имеем:

$$n = \sum_i N_i p_i = p_{01} + p_{10} + 2p_{11},$$

где вероятности p_i вычисляются по (2). Зададимся значениями $\varepsilon = 10kT$ и $U_0 = 20kT$. Вычисления дают графики, показанные на рис. 8.

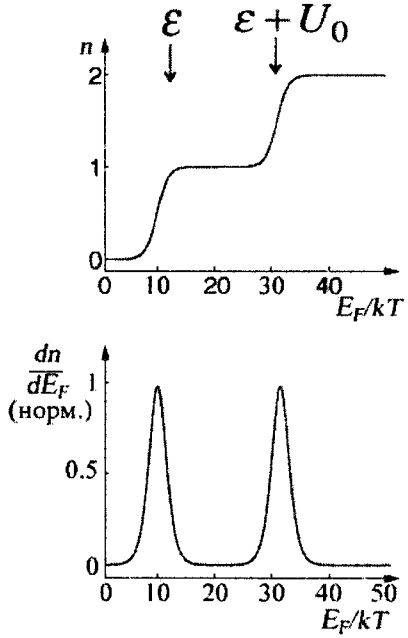


Рис. 8. Равновесное число электронов n в двухмодовом проводнике (рис. 7) как функция E_F/kT при $\varepsilon = 10kT$ и $U_0 = 20kT$ [26].

Проводимость пропорциональна $dn/d\mu$ и принимает пиковые значения при $E_F = \varepsilon$ и $E_F = \varepsilon + U_0$. Обратим внимание, что число электронов в канале меняется на единицу как только E_F принимает значение ε и далее значение $\varepsilon + U_0$, как и следовало ожидать из рис. 3. Заметим, однако, что при этом мы не подразумевали как на рис. 3 два одноэлектронных состояния при различных энергиях. Мы подразумевали два одноэлектронных состояния с одной и той же энергией, но учли энергию взаимодействия в пространстве Фока (рис. 7).

Именно пиковые значения проводимости наблюдаются экспериментально в низковольтном режиме кулоновской блокады для нанорезисторов при условии, что U_0 превышает как квант тепловой энергии kT, так и фундаментальное уширение ($U_0 \gg h/t$).

Низковольтная проводимость является равновесным свойством [27] и для ее описания достаточно равновесной статистической механики. Ток за пределами режима линейного отклика при более высоком напряжении требует уже методов неравновесной статистической механики. Далее мы покажем как

получить плато на уровне две трети от полного тока (рис. 5) путем вычисления тока при высоком напряжении в фоковском представлении.

12.2.3. Вычисление тока в фоковском пространстве

Для вычисления тока нужно выписать уравнения для вероятностей найти систему в одном из доступных состояний, при этом сумма вероятностей должна быть равна единице. Так, для тока в одноуровневом резисторе в полуклассической модели (§10.3) имели:

$$\nu_1 p_0 = \nu_2 p_1 \Rightarrow p_1/p_0 = \nu_1/\nu_2 \Rightarrow p_1 = \nu_1/(\nu_1+\nu_2),$$

где левый контакт посылает систему из состояния 0 в состояние 1 со скоростью ν_1, а правый контакт реализует возврат системы со скоростью ν_2, и оба процесса в состоянии равновесия должны уравновешивать друг друга, причем $p_0 + p_1 = 1$. В итоге ток

$$I = q\nu_2 p_1 = q\frac{\nu_1\nu_2}{\nu_1+\nu_2}$$

в согласии с результатом, полученным ранее в одночастичной модели (§ 10.3).

Такой подход особенно эффективен при рассмотрении систем с несколькими взаимодействующими состояниями. Рассмотрим, например, двухуровневый резистор, на который подано такое напряжение, что электроны, входящие в проводник из левого контакта, переводят систему из состояния 00 в состояние 01 или 10, но не 11 из-за большого значения удельной энергии заряжания U_0 (рис. 9) .

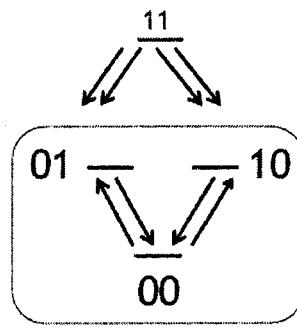


Рис. 9. Условия появления плато на уровне две трети от полного тока (рис. 5).

Сейчас мы покажем, что именно в этих условиях появляется плато на уровне две трети от полного тока (рис. 5), когда система не может перейти в состояние 11, а может лишь покинуть его.

Из кинетики процессов с участием только трех состояний 00, 01 и 10 в условиях динамического равновесия имеем:

$$2\nu_1 p_{00} = \nu_2 (p_{01} + p_{10}) \Rightarrow \frac{p_{01} + p_{10}}{p_{00}} = \frac{2\nu_1}{\nu_2},$$

откуда с учетом нормировки $p_{00} + p_{01} + p_{10} = 1$:

$$p_{01} + p_{10} = \frac{2\nu_1}{2\nu_1 + \nu_2},$$

так что в конечном итоге ток

$$I = q\nu_2 (p_{01} + p_{10}) = q\frac{2\nu_1\nu_2}{2\nu_1 + \nu_2}.$$

В условиях равенства прямых и обратных потоков электронов в проводнике ($\nu_1 = \nu_2$) ток составляет именно 2/3 от его максимального значения:

$$I = \frac{2}{3} q\nu_1.$$

Этот результат невозможно получить в одночастичной модели. Наш двухмодовый резистор в оговоренных условиях может находиться в состояниях 00, 01 и 10, но не в состоянии 11. В одночастичной модели электрон может находиться в любом из этих состояний с вероятностью 1/3. Если электроны независимы (не взаимодействуют), вероятность заселить одно и то же состояние (со спинами «вверх» и «вниз») будет 1/9. На самом деле эта вероятность равна нулю, что указывает на сильно коррелированное движение электронов (рис. 10).

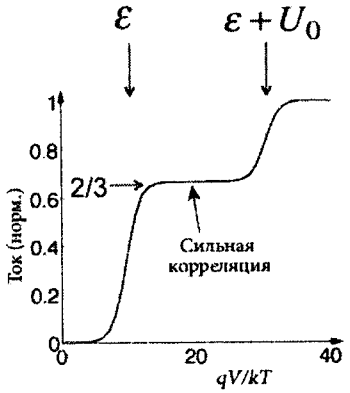


Рис. 10. Промежуточное плато в ВАХ (рис. 5) соответствует сильно коррелированному состоянию электронов.

Следуя духу концепции «снизу – вверх» [26], мы рассмотрели явление кулоновской блокады в молекулярной электронике, точнее в одноэлектронике. Традиционно явление кулоновской блокады рассматривается с позиций концепции «сверху – вниз», привлекая такое понятие как электроемкость нанопроводника. Рассмотрим физику кулоновской блокады и с таких позиций и обсудим результаты расчета одноэлектронного транзистора, в котором роль проводника выполняет одиночная молекула бензола [22, 23].

12.3. Одноэлектронный транзистор на молекуле бензола

Успехи последнего десятилетия в области молекулярной электроники, в первую очередь, экспериментальные, завершились, в частности, созданием одноэлектронного полевого транзистора (Single-Electron Transistor/SET) [28 – 36], схематически показанного на рис. 11.

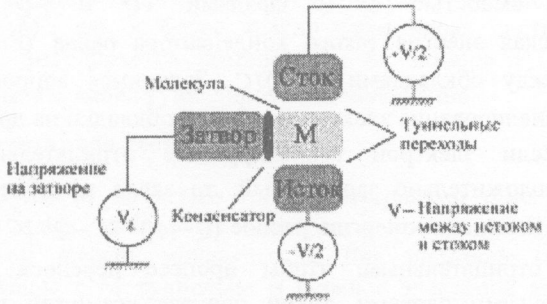


Рис. 11. Принципиальная схема SET. Энергия электронных состояний молекулы М контролируется электростатическим полем затвора.

Рассматривают два механизма переноса электронов в SET: когерентное туннелирование и последовательное туннелирование. Когерентное туннелирование реализуется в случае сильной связи молекулы М с металлической поверхностью электродов, например, через сульфидные мостики. Время жизни электронов на М короткое, локализоваться электроны не успевают и движутся когерентно к стоковому электроду. При когерентном туннелировании наличие электронных состояний молекулы М в окне туннелирования не обязательно, электроны могут транспортироваться через «хвосты» короткоживущих уширенных состояний молекулы, например, нижнего незаполненного состояния (Lowest Unoccupied Molecular Orbital/LUMO) (рис. 12).

Последовательное туннелирование, известное также как механизм кулоновской блокады [15], реализуется в случае слабой связи молекулы с поверхностью электродов. Волновые свойства электрона позволяют ему преодолеть туннельный барьер, а его корпускулярная природа приводит к дискретности переноса заряда, в результате чего при определенных условиях в

туннельных наноконтактах возникает подавление электронного транспорта («кулоновская блокада»). Мы далее выполним и обсудим расчеты именно этой модели на примере молекулы бензола.

Молекулы – наноразмерные проводники. С уменьшением размеров проводника уменьшается его электрическая емкость. Сегодня, при вполне достижимых экспериментальных условиях емкость C может стать настолько малой, что даже кулоновская энергия q^2/C одного дополнительного электрона на молекуле может оказаться существенной. Приходится учитывать влияние кулоновских эффектов на перенос заряда через молекулу, что требует запаса энергии, и может реализоваться явление кулоновской блокады, когда к электродам приложено конечное напряжение, а ток тем не менее не идет.

Приведем простой и очевидный пример кулоновской блокады. Рассмотрим конденсатор с емкостью C и зарядами $+Q$ и $-Q$ на обкладках. Электростатическая энергия такого конденсатора равна $Q^2/2C = CV^2/2$, где напряжение между обкладками $V = Q/C$. Зададимся вопросом: при каком напряжении туннелирование электрона с одной обкладки на другую становится возможным? Если электрон туннелирует с отрицательно заряженного электрода на положительно заряженный, то заряд на обкладке становится равным $Q-|q|$, а изменение энергии, равное $(Q-|q|)^2/2C - Q^2/2C = (q^2 - 2|q|Q)/2C$, должно быть отрицательным, чтобы процесс переноса электрона стал возможным. Другими словами, такой процесс возможен при $V > |q|/2C$ и электронный транспорт невозможен при напряжениях на обкладках $|V| < |q|/2C$. Итак, имеем нулевой туннельный ток при конечном напряжении на обкладках, что может служить простым примером кулоновской блокады. Туннелировав с истока на молекулу М, электрон живет достаточно долго на М, успевает локализоваться и, таким образом, теряет всю информацию о своей предистории. Дальнейшее туннелирование этого электрона на сток уже никак не связано с его туннелированием с истока на молекулу. Это механизм последовательного туннелирования. Экспериментально показано [30], что реализоваться он может только при наличии дискретного электронного состояния молекулы М в окне туннелирования, например, состояния EA (Electron Affinity), соответствующего захвату молекулой дополнительного электрона M^-, энергия которого определяется сродством молекулы к электрону EA или состояния IP (Ionization Potential), соответствующего потере одного электрона молекулой M^+, энергия которого определяется потенциалом ионизации молекулы IP (рис.12).

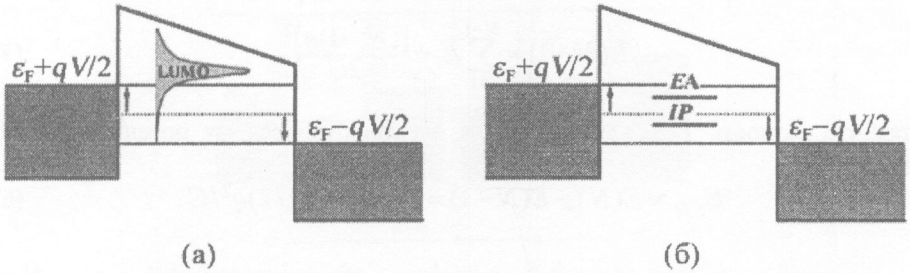


Рис. 12. Когерентное (а) и последовательное (б) туннелирование электрона в SET.

Положение уровней энергии *EA* и *IP* однократно заряженных состояний молекулы M^- и M^+, как и более кратно заряженных состояний в канале туннелирования регулируется потенциалом на затворе V_g, что позволяет закрывать и открывать канал туннелирования электронов.

12.3.1. Элементарная теория SET

Связь молекулы M со всеми тремя электродами (рис. 11) емкостная: изменение потенциала любого из электродов влечет за собой изменение электростатической энергии молекулы M. Два электрода (истоковый S и стоковый D) связаны с M туннельно и перенос электрона возможен только между этими электродами. Туннельная связь означает, что переносимый электрон находится либо на M, либо на одном из этих двух электродов. Упростим ситуацию. Предположим, что все взаимодействия между электроном, переносимым на молекулу M, как и всех остальных электронов M и обоих электродов, можно параметризовать суммарной емкостью C. Предположим также, что значение C не зависит от возможных заряженных состояний молекулы M. Тогда электростатическая энергия молекулы M с N электронами равна

$$Q^2/2C = (Nq)^2/2C.$$

Полная энергия молекулы M с N электронами равна

$$E(N)=\sum_{i=1}^{N}E_i+(Nq)^2/2C, \qquad (3)$$

где E_i – энергия i-го электрона в самосогласованном поле остальных электронов. При появлении дополнительного электрона у молекулы M полная энергия становится равной

$$E(N+1)=\sum_{i=1}^{N+1}E_i+\frac{[(N+1)q]^2}{2C}, \qquad (4)$$

а уход электрона с молекулы М дает

$$E(N-1)=\sum_{i=1}^{N-1}E_i+\frac{[(N-1)q]^2}{2C}, \qquad (5)$$

так что разность $E(N)-E(N-1)$ равна электрохимическому потенциалу N-го электрона

$$E_{FN}\equiv E(N)-E(N-1)=E_N+(N-1/2)q^2/C, \qquad (6)$$

определяемого как минимальная энергия, необходимая для добавления N-го электрона. Как только E_{FN} окажется меньше E_{FS} и E_{FD}, N-ый электрон перенесется на М. Чтобы добавить еще один электрон на молекулу М с N электронами электрохимический потенциал

$$E_{F,N+1}=E_{FN}+\frac{q^2}{C}+\Delta E_N, \qquad (7)$$

где $\Delta E_N=E_{N+1}-E_N$ должна быть меньше обоих E_{FS} и E_{FD}. Для упрощения реальной ситуации предположим, что ΔE_N слабо зависит от величины заряда на молекуле М, так что далее опустим индекс N в ΔE_N. Таким образом, энергия N+1-ого электрона должна быть больше энергии N-ого электрона на величину энергии заряжания $q^2/C+\Delta E$. Первое слагаемое в энергии заряжания молекулы М есть энергия $q^2/C\equiv E_C$, необходимая для преодоления кулоновского отталкивания между электронами. Второе слагаемое ΔE есть результат дискретности спектра молекулы.

Проиллюстрируем полученные выводы на рис. 13 для двух ситуаций: (а) перенос электрона невозможен (состояние транзистора «off») и (б) перенос разрешен (состояние транзистора «on»). Пусть в первом случае энергетическая диаграмма SET такова: $E_{F,N+1}>E_{FD}>E_{FS}>E_{FN}$, а во втором случае $E_{FD}>E_{F,N+1}>E_{FS}>E_{FN}$. В первом случае каналом туннелирования является молекула М с N электронами. Ближайший уровень с энергией $E_{F,N+1}$ не заполнен электронами и лежит выше энергии Ферми истокового и стокового электродов, в канале туннелирования нет ни одного свободного уровня, перенос электрона «блокирован», состояние транзистора – «off».

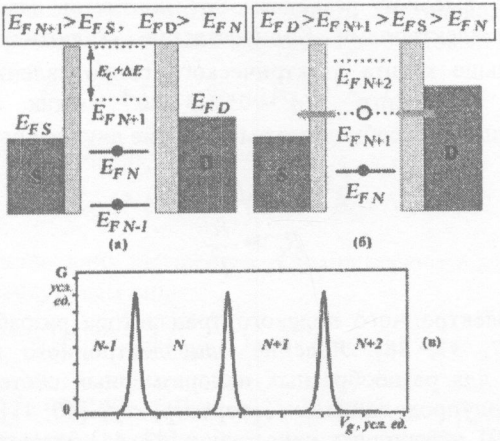


Рис. 13. Перенос электрона в SET. Диаграммы энергии показаны для двух различных ситуаций. В случае (а) число электронов молекулы М фиксировано значением N, перенос электрона на М «блокирован», состояние транзистора – «off»; в случае (б) число электронов на молекуле М осциллирует между значениями N и $N+1$, состояние – «on»; (в) Кулоновские осцилляции.

Противоположная ситуация показана на рис. 13б. В канале туннелирования есть состояние с энергией $E_{F,N+1}$, электрон может перейти со стокового электрода на истоковый. Электрический ток можно обеспечить путем периодического изменения зарядового состояния молекулы в канале туннелирования с N на $N+1$. Изменить число электронов на молекуле М можно в результате изменения потенциала затвора V_g, поскольку от него зависит энергия заряжания молекулы.

Наблюдаемая характерная зависимость проводимости G от потенциала затвора V_g в SET при небольших значениях разности потенциалов на электродах в виде резких пиков и долин показана на рис. 13в. В долинах число электронов молекулы М фиксировано значениями $N-1$, N, $N+1$, $N+2$ и т.д., и ток блокируется энергией заряжания $q^2/C + \Delta E$, что соответствует ситуации на рис. 13а. Пики проводимости соответствуют ситуации на рис. 13б, когда молекула в канале туннелирования осциллирует между двумя своим состояниями. Например, пик проводимости между долинами с N и $N+1$ электронами соответствует осцилляции молекулы М между ее состояниями с N и $N+1$ электронами. Это – кулоновские осцилляции.

Для наблюдения кулоновских осцилляций энергия заряжания молекулы $q^2/C + \Delta E$ должна быть намного больше кванта тепловой энергии kT. В противном случае тепловые флуктуации станут доминирующими и перекроют кулоновские осцилляции. Также необходимо, чтобы число электронов молекулы было хорошо определенной наблюдаемой величиной, что влечет за

собой требование высокой резистивности контактов между молекулой и токообразующими электродами. Количественно, сопротивление контакта R_c должно быть больше кванта электрического сопротивления, определяемого константой фон Клитцинга $h/q^2 \sim 25.813\ кОм^{-1}$. Итак, для наблюдения кулоновских осцилляций необходимо выполнение двух условий:

$$q^2/C + \Delta E \gg kT, \tag{8}$$

$$R_c \gg \frac{h}{q^2}. \tag{9}$$

Теория одноэлектронного полевого транзистора разработана достаточно глубоко [15 – 17, 37, 38]. Явление одноэлектронного переноса изучено экспериментально для разнообразных наноразмерных систем: металлических наночастиц [39], полупроводниковых гетероструктур [40, 41], полупроводящих нанокристаллов [42], углеродных нанотрубок [43, 44] и отдельных молекул [28 – 36].

12.3.2. Условия переноса электрона с истока на сток

Для нахождения этих условий выпишем очевидные соотношения между энергиями $E_M(N)$ молекулы с числом электронов N в начальном незаряженном состоянии и энергиями $E_S(N')$ и $E_D(N'')$ электронов в истоке и стоке, учитывая то обстоятельство, что при переходе одного электрона из истокового электрода на молекулу энергия системы должна по крайней мере понизиться [35]:

$$E_S(N') + E_M(N) \geq E_S(N'-1) + E_M(N+1). \tag{10}$$

Аналогичное неравенство имеет место при переходе одного электрона с молекулы на стоковый электрод:

$$E_M(N+1) + E_D(N'') \geq E_M(N) + E_D(N''+1). \tag{11}$$

Если работу выхода электрона из металла обозначить W, то максимальная энергия электрона в истоковом электроде будет $-W+qV/2$, где V – прилагаемая к молекуле М разность потенциалов. Если предположить, что с истока на молекулу туннелирует электрон с максимальной энергией, то, очевидно, имеем

$$E_S(N') - E_S(N'-1) = -W + qV/2. \tag{12}$$

Тогда условие туннелирования одного электрона с истока на молекулу примет следующий вид:

$$-W + qV/2 + E_M(N) \geq E_M(N+1). \tag{13}$$

Поскольку минимальная энергия электрона в стоковом электроде есть $-W-qV/2$, то при переходе электрона с молекулы на стоковый электрод имеем:

$$E_{\text{M}}(N+1) \geq -W-qV/2+E_{\text{M}}(N). \quad (14)$$

Введем энергию заряжания молекулы

$$\Delta E_{\text{M}}(N) = E_{\text{M}}(N+1)-E_{\text{M}}(N). \quad (15)$$

Тогда из двух последних неравенств условия переноса электрона с истока на сток получим в следующем виде:

$$q|V|/2 \geq \Delta E_{\text{M}}(N)+W \geq -q|V|/2. \quad (16)$$

Не учтено лишь влияние потенциала затвора V_g на спектр молекулы, используемой в качестве канала передачи электрона. Пусть в первом приближении имеет место линейная зависимость энергии заряжания от потенциала затвора:

$$\Delta E_{\text{M}}(N,V_g) = \Delta E_{\text{M}}(N)+\alpha V_g, \quad (17)$$

где константа связи затвора α пока что не известна. Положим $\alpha = 1$. Окончательно, условия переноса электрона с истока на сток теперь запишутся так:

$$q|V|/2 \geq \Delta E_{\text{M}}(N,V_g)+W \geq -q|V|/2. \quad (18)$$

Выполненные нами самосогласованные расчеты с учетом поляризации молекулы электростатическим полем затвора покажут (§ 12.3.4) [22, 23], что значение константы связи затвора в случае молекулы бензола действительно близко к 1, а зависимость энергии заряжания молекулы от потенциала затвора действительно близка к линейной. Мы воспользуемся этими соотношениями для вычисления диаграмм зарядовой стабильности одноэлектронного полевого транзистора, которые показывают зависимость числа заряженных состояний молекулы в канале туннелирования от напряжения, подаваемого на электроды, и от потенциала затвора.

12.3.3. Формализм теории функционала плотности

Спектр молекулы, реализующей канал туннелирования, в нашем случае молекулы бензола, предпочтительнее вычислять методами теории функционала плотности (Density Functional Theory, DFT) [14, 45]. В отличии от решения уравнения Шредингера для многоэлектронных систем тем или иным вариационным методом, когда время вычислений пропорционально в лучшем случае $\mathbb{N}^6$, где $\mathbb{N}$ – число базисных функций, или даже $\mathbb{N}!$, если речь идет о полном конфигурационном взаимодействии [19, 45, 46], в методах DFT, к

которым относится и использованное нами в расчетах приближение локальной плотности (Local Density Approximations, LDA), время вычислений пропорционально $\sim N^3$, что позволяет рассчитывать молекулярные системы с большим числом электронов.

Практическая реализация методов DFT сводится к решению одноэлектронного уравнения Кона – Шема [14]

$$\hat{\boldsymbol{H}}_{el} = -\frac{1}{2}\nabla^2 + V^{eff}[n](\boldsymbol{r}), \tag{19}$$

где второе слагаемое описывает эффективную потенциальную энергию электрона в усредненном поле всех других электронов, электронная плотность которых обозначена n.

Одноэлектронные состояния гамильтониана Кона – Шема находятся как решения одночастичного уравнения Шредингера

$$\hat{\boldsymbol{H}}_{el}\psi_\alpha(\boldsymbol{r}) = \varepsilon_\alpha\psi_\alpha(\boldsymbol{r}) \tag{20}$$

путем разложения решений по базисным функциям

$$\psi_\alpha(\boldsymbol{r}) = \sum_i a_{\alpha i}\phi_i(\boldsymbol{r}). \tag{21}$$

В результате решение сводится к задаче на собственные значения и собственные функции $c_{\alpha j}$:

$$\sum_j H_{ij}c_{\alpha j} = \varepsilon_\alpha \sum_j S_{ij}c_{\alpha j}. \tag{22}$$

Здесь матричные элементы гамильтониана $H_{ij} = \langle\phi_i|\hat{\boldsymbol{H}}_{el}|\phi_j\rangle$ и матрицы перекрывания $S_{ij} = \langle\phi_i|\phi_j\rangle$ берутся по базисным функциям. Занятые состояния определяют электронную плотность

$$[n](\boldsymbol{r}) = \sum_\alpha |\psi_\alpha(\boldsymbol{r})|^2 f\left(\frac{\varepsilon_\alpha - \varepsilon_F}{kT}\right), \tag{23}$$

где фермиевская функция $f(x) = 1/(e^x + 1)$, ε_F – энергия Ферми и T – электронная температура. Электронная плотность традиционно выражается через матрицу плотности

$$[n](\boldsymbol{r}) = \sum_{ij} D_{ij}\phi_i(\boldsymbol{r})\phi_j(\boldsymbol{r}), \tag{24}$$

которая в свою очередь выражается через коэффициенты разложения $c_{\alpha i}$:

$$D_{ij} = \sum_\alpha c^*_{\alpha i}c_{\alpha j} f\left(\frac{\varepsilon_\alpha - \varepsilon_F}{kT}\right). \tag{25}$$

Эффективный потенциал в уравнении Кона – Шема (19) в общем случае имеет три слагаемых:

$$V^{eff}[n]=V^{H}[n]+V^{xc}[n]+V^{ext}[n], \tag{26}$$

где два первых члена зависят от электронной плотности: хартриевский потенциал $V^{H}[n]$ описывает взаимодействие электрона с усредненным полем остальных электронов, а второе слагаемое $V^{xc}[n]$ есть обменно-корреляционный потенциал, учитывающий квантовую природу электрона. Последний член $V^{ext}[n]$ учитывает, если необходимо как в нашем случае, другие электростатические взаимодействия, например, с внешним электростатическим полем, или же ионные потенциалы.

Хартриевский потенциал вычисляется из уравнения Пуассона

$$\nabla^2 V^{H}[n](\boldsymbol{r})=-4\pi[n](\boldsymbol{r}) , \tag{27}$$

для решения которого задаются граничные условия периодичности кристаллической решетки.

Обменно-корреляционный потенциал $V^{xc}[n]$ есть усредненный потенциал квантового взаимодействия между электронами и определяется он как функциональная производная от обменно-корреляционной энергии

$$V^{xc}[n](\boldsymbol{r})=\frac{\delta E^{xc}}{\delta n}(\boldsymbol{r}). \tag{28}$$

В DFT полная энергия является функционалом электронной плотности n и дается уравнением

$$E[n]=T[n]+E^{xc}[n]+E^{H}[n]+E^{ext}[n], \tag{29}$$

где $T[n]$ есть кинетическая энергия кон-шемовских орбиталей

$$T[n]=\sum_{\alpha}\langle\psi_\alpha|-\frac{1}{2}\nabla^2|\psi_\alpha\rangle f\left(\frac{\varepsilon_\alpha-\varepsilon_F}{kT}\right), \tag{30}$$

$E^{xc}[n]$ – обменно-корреляционная энергия, $E^{H}[n]$ – хартриевская энергия и $E^{ext}[n]$ есть энергия взаимодействия с внешними электростатическими полями.

В используемом нами приближении LDA обменно-корреляционный функционал зависит от локальной плотности

$$E^{LDA}[n]=\int n(\boldsymbol{r})\varepsilon^{LDA}\big(n(\boldsymbol{r})\big)d\boldsymbol{r}, \tag{31}$$

где $\varepsilon^{LDA}(n(\boldsymbol{r}))$ есть плотность обменно-корреляционной энергии однородного электронного газа плотности $n(\boldsymbol{r})$.

12.3.3.1. Учет внешнего электростатического поля

Используемая далее в расчетах модель DFT основана на псевдопотенциалах с численными локализованными базисными функциями, как это предложено в основополагающей работе Солера и др. [49]. Введем для каждого атома молекулы компенсационный заряд $\rho_i^{comp}(\boldsymbol{r})$, равный заряду Z_i псевдопотенциала и экранирующий электростатические взаимодействия.

Тогда разностная электронная плотность

$$\delta n(\boldsymbol{r}) = n(\boldsymbol{r}) - \sum_i \rho_i^{comp}(\boldsymbol{r}), \tag{32}$$

где $n(\boldsymbol{r})$ – определенная выше суммарная электронная плотность молекулы в точке $\boldsymbol{r}$.

Введем также экранированный локальный псевдопотенциал («нейтральный атом»)

$$V_i^{NA}(\boldsymbol{r}) = V_i^{loc}(\boldsymbol{r}) - \int d\boldsymbol{r}' \frac{\rho_i^{comp}(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \tag{33}$$

по отношению к локальному псевдопотенциалу $V_i^{loc}(\boldsymbol{r})$.

Тогда функционал полной энергии можно представить в виде

$$E[n] = T[n] + E^{xc}[n] + \frac{1}{2}\int \delta V^H(\boldsymbol{r})\delta n(\boldsymbol{r})d\boldsymbol{r} + \int \sum_i V_i^{NA}(\boldsymbol{r})\delta n(\boldsymbol{r})d\boldsymbol{r} + \frac{1}{2}\sum_{ij} U_{ij}, \tag{34}$$

где первые два слагаемых определены выше, а следующие три слагаемых представляют собой перегруппированные электростатические члены. Первое из них есть разностная хартриевская энергия, вычисляемая через разностный хартриевский потенциал $\delta V^H(\boldsymbol{r})$, который в свою очередь получается в результате решения уравнения Пуассона для разностной электронной плотности $\delta n(\boldsymbol{r})$. Следующее слагаемое учитывает взаимодействие между электронами и экранированными ионами, а последнее слагаемое включает в себя все электростатические взаимодействия, которые не зависят от электронной плотности. Это последнее слагаемое вычисляется из $V_i^{loc}(\boldsymbol{r})$, $\rho_i^{comp}(\boldsymbol{r})$ и Z_i, а поскольку уравнение Пуассона линейное, то это слагаемое можно переписать в виде суммы парных потенциалов.

Далее в функционале полной энергии нужно учесть влияние на молекулу трех металлических электродов и диэлектрической подложки со стороны затвора (рис. 11). Потенциал внутри металлических электродов определяется напряжением, заданным для каждой пары электродов, а на граничных поверхностях электродов электрическое поле равно нулю (граничные условия Неймана).

Решив уравнение Пуассона в отсутствии молекулы, получим распределение в пространстве внешнего потенциала, создаваемого электростатическим окружением:

$$-\nabla\cdot[\varepsilon(\boldsymbol{r})\nabla V^{ext}(\boldsymbol{r})]=0, \tag{35}$$

где $\varepsilon(\boldsymbol{r})$ – относительная диэлектрическая проницаемость. Добавляем молекулу и снова решаем уравнение Пуассона методом самосогласования, в результате чего получаем разностную электронную плотность и полный разностный хартриевский потенциал:

$$-\nabla\cdot[\varepsilon(\boldsymbol{r})\nabla\delta V^{H+ext}(\boldsymbol{r})]=\delta n(\boldsymbol{r}). \tag{36}$$

Наконец, определяем молекулярную часть полного разностного хартриевского потенциала

$$\delta V^{H}(\boldsymbol{r})=\delta V^{H+ext}(\boldsymbol{r})-V^{ext}(\boldsymbol{r}). \tag{37}$$

Именно этот разностный хартриевский потенциал входит в функционал полной энергии, приведенный выше. Следуя [47], вклад в полную энергию системы за счет внешнего электростатического поля дается выражением

$$\Delta E=\int V^{ext}(\boldsymbol{r})n(\boldsymbol{r})d\boldsymbol{r}-\sum_i V^{ext}(\boldsymbol{R}_i)Z_i, \tag{38}$$

где $\boldsymbol{R}_i$ и Z_i есть, соответственно, радиус-вектор атома i и валентность его псевдопотенциала.

Окончательно, полная энергия системы (молекула в электростатическом поле полевого транзистора) дается суммой вкладов от уравнений (34) и (38). Отметим, что это верно только если последнее слагаемое в уравнении (34) в виде суммы парных потенциалов не зависит от электростатического окружения, что соответствует тому, что компенсационный заряд $\rho_i^{comp}(\boldsymbol{r})$ и экранированный локальный псевдопотенциал $V_i^{NA}(\boldsymbol{r})$ не перекрываются с металлическими электродами и диэлектрической прослойкой затвора [35]. Отметим также, что поскольку разностный хартриевский потенциал $\delta V^{H}(\boldsymbol{r})$ в уравнении (34) вычисляется с учетом электростатического окружения, то разностная хартриевская энергия [третье слагаемое в уравнении (34)] учитывает энергию электростатического взаимодействия между молекулой и индуцированными поляризационными зарядами окружающих ее трех металлических электродов [34].

12.3.4. Результаты расчетов

Вычисления проводились по программе ATK 10.8.2 [48 – 50].

Сначала рассмотрим изолированную молекулу бензола (B) в ее экспериментальной геометрии D_{6h} с длинами связей CC = 1.40 и CH = 1.10 Å [51] и вычислим ее спектр в модели LDA-PZ теории функционала плотности [14] с базисными функциями DoubleZetaPolarized программы ATK без учета остовных электронов. Для одночастичных верхнего заполненного состояния HOMO и нижнего пустого – LUMO получаем, соответственно, $\varepsilon_{14} = -2.5890$ и $\varepsilon_{15} = +2.5890$ при энергии Ферми $\varepsilon_F = -3.3182$ *эВ*, что дает для вертикальных значений потенциала ионизации и сродства к электрону изолированной молекулы бензола значения, соответственно, $I = 5.90_7$ и $A = 0.72_9$ *эВ* в противоречии с экспериментальными значениями $I_{эксп} = 9.25$ и $A_{эксп} \sim -1.10$ *эВ* [52]. Нужны расчетные значения, более точно отражающие экспериментальные данные.

С этой целью в рамках той же самой модели DFT/LDA-PZ мы выполнили самосогласованные расчеты не только нейтральной молекулы бензола B^0, но и ее заряженных форм B^+, B^-, B^{2+} и B^{2-}. Результаты самосогласованного расчета полной энергии этих систем приведены в табл. 1.

Таблица 1

Самосогласованные значения полной энергии свободной молекулы бензола и ее заряженных форм в модели LDA-PZ и энергии заряжания молекулы бензола $\Delta E_B(N)$, *эВ*

Состояние	B^{2+}	B^+	B^0	B^-	B^{2-}
Полная энергия	–1014.60	–1030.33	–1039.45	–1037.14	–1028.75
Энергия заряжания	–15.73	– 9.12	2.31	8.39	–

Самосогласованные значения потенциала ионизации и сродства к электрону молекулы бензола получаются равными, соответственно, I = 9.12 и A = –2.31 эВ, что намного лучше согласуется с приведенными выше экспериментальными данными и вполне приемлемо для дальнейшего обсуждения. В табл. 1 приведены также значения определенной ранее энергии заряжания (15) $\Delta E_M(N)$ молекулы бензола в ее разных заряженных состояниях. Для построения диаграммы зарядовой стабильности полевого транзистора на молекуле бензола осталось задать работу выхода W электрона из металлического электрода. Пусть $W = 5.28$ *эВ*, что соответствует выбору электродов истока и стока из золота [29].

Задавая различное напряжение V, подаваемое на канал туннелирования электрона при разных значениях потенциала затвора V_g, из полученных выше условий переноса электрона с истока на сток (18) вычислим число заряженных состояний в канале туннелирования. Соответствующая вычисленная нами

диаграмма зарядовой стабильности полевого транзистора на молекуле бензола без учета среды SET показана на рис. 14.

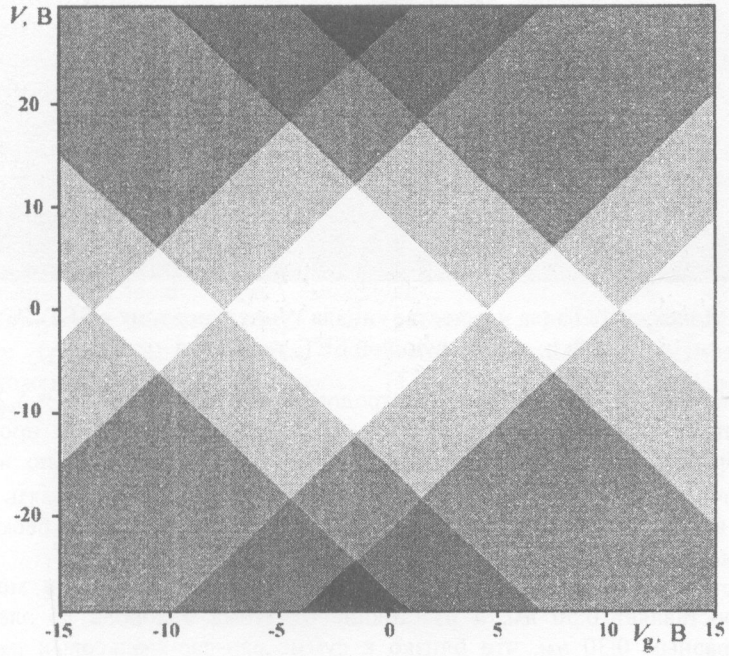


Рис. 14. Диаграмма зарядовой стабильности полевого транзистора на молекуле бензола без самосогласованного учета поляризации молекулы бензола электростатическим полем затвора. Диаграмма показывает число заряженных состояний молекулы в канале туннелирования в зависимости от значений V и V_g: в белых полях заряженных состояний нет, в остальных – число заряженных состояний изменяется от 1 до 4, возрастая с плотностью закраски полей. Из-за некоторого сходства многоугольников с гранями алмаза подобные графики в англоязычной литературе принято называть кулоновскими алмазами (Coulomb diamonds).

Перейдем к рассмотрению молекулы бензола в качестве канала туннелирования в реалистической модели SET в виде прямоугольной ячейки $1.90 \times 1.20 \times 1.20\ нм^3$ с молекулой бензола, помещенной посередине между двумя электродами (рис. 15).

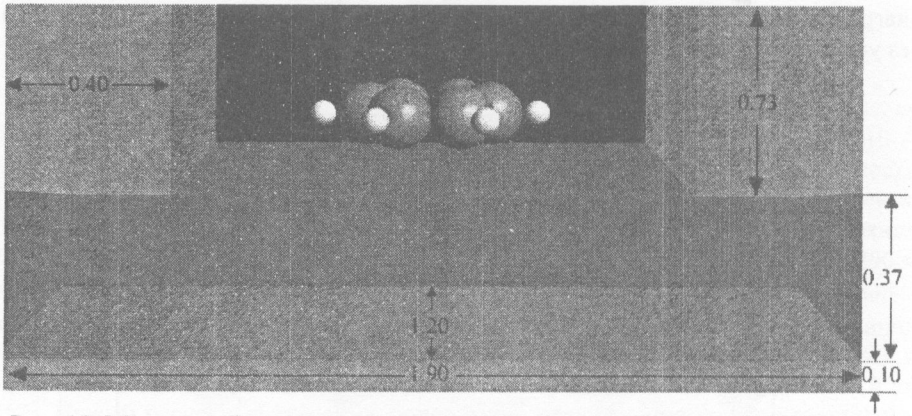


Рис. 15. Молекула бензола в качестве канала туннелирования в SET. Размеры ячейки, моделирующей SET, указаны в *нм*.

Работа выхода для всех трех электродов полагалась равной $W = 5.28$ *эВ*, а относительная диэлектрическая постоянная для диэлектрической прослойки затвора бралась равной $\varepsilon_r = 10$ (high-k Al_2O_3). Такая сравнительно высокая диэлектрическая постоянная обеспечит достаточную емкостную связь между затвором и молекулой. Электроды истока и стока моделировались бесконечно длинными металлическими блоками.

Расстояние между противоположными атомами водорода в молекуле бензола составляет 0.50 *нм*, а расстояние от атома водорода до электрода выбрано равным 0.30 *нм*, что близко к сумме ван-дер-ваальсовых радиусов атомов H и Au, которая равна $0.27_5 = 0.10_9 + 0.16_6$, в *нм*, соответственно [51]. Расстояние между поверхностью диэлектрика и плоскостью молекулы бензола выбрано равным 0.13 *нм*. Остальные геометрические параметры ячейки, моделирующей SET, показаны на рис. 15.

В расчетах не учитывалось изменение геометрии заряженных форм молекулы бензола относительно приведенной выше геометрии нейтральной молекулы бензола. Анализ фотоэлектронного спектра B^+ с полностью разрешенной вращательной структурой показал, что симметрия иона B^+ остается D_{6h}, а изменения длин связей CC при ионизации молекулы бензола не превышает 0.1% [54], что подтверждается и нашими квантовомеханическими расчетами, в том числе и для дважды заряженных состояний молекулы бензола, в модели B3LYP/6-311G** [45]. В табл. 2 собраны результаты расчета полной энергии и энергии заряжания молекулы B и ее заряженных форм в описанной выше реалистической модели SET при нулевом потенциале на затворе.

Таблица 2

Самосогласованные значения полной энергии молекулы бензола и ее заряженных форм в модели LDA-PZ и энергии заряжания молекулы бензола $\Delta E_B(N)$, эВ в реалистической модели SET при $V_g = 0$.

Состояние	B^{2+}	B^{+}	B^{0}	B^{-}	B^{2-}
Полная энергия	–1021.84	–1032.01	–1039.51	–1039.41	–1037.06
Энергия заряжания	–10.17	–7.50	0.10	2.35	–

Сравнение с аналогичными результатами для свободной молекулы бензола (табл. 1) показывает, как и ожидалось, уменьшение энергии заряжания при помещении молекулы в среду SET. Это уменьшение вызвано, в основном, стабилизацией заряда на молекуле диэлектрической подложкой. Соответствующая диаграмма зарядовой стабильности показана на рис. 16. Как и ранее (рис. 14), диаграмма зарядовой стабильности вычислялась в предположении линейной зависимости (17) энергии заряжания молекулы от потенциала затвора.

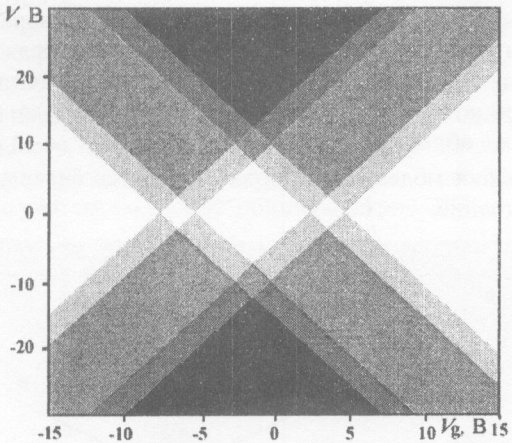


Рис. 16. Диаграмма зарядовой стабильности полевого транзистора на молекуле бензола в его реалистической модели при $V_g = 0$. Диаграмма показывает число заряженных состояний молекулы в канале туннелирования в зависимости от значений V и V_g: в белых полях заряженных состояний нет, в остальных – число заряженных состояний изменяется от 1 до 4, возрастая с плотностью закраски полей.

Можно визуализовать электростатический потенциал в области расположения молекулы, индуцированный при подаче напряжения на затвор. Разность между потенциалами, индуцированным при $V_g = 0$ и $V_g = 2$ B, показана графически на рис. 17.

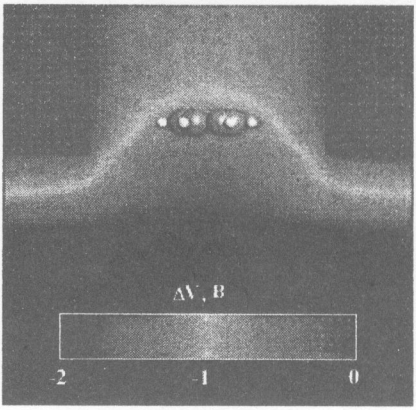


Рис. 17. Электростатический потенциал в SET на молекуле бензола, индуцируемый при $V_g = 2\ B$.

На рис. 18 для разных заряженных состояний молекулы бензола показана зависимость полной энергии SET от потенциала затвора. Отрицательный потенциал на затворе стабилизирует положительно заряженные состояния молекулы бензола и наоборот, положительный потенциал на затворе стабилизирует отрицательно заряженные состояния бензола. Обращает на себя внимание, что в области потенциала затвора $V_g < -4\ B$ положительно заряженные состояния молекулы бензола становятся более устойчивыми, чем незаряженное состояние.

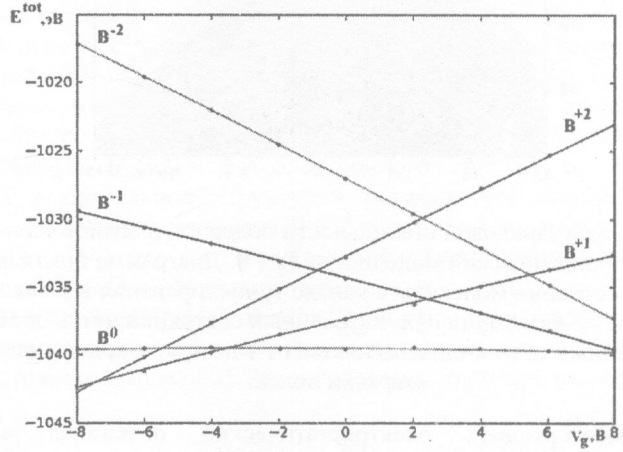


Рис. 18. Зависимость полной энергии SET на молекуле бензола от напряжения на затворе для разных заряженных состояний молекулы.

Зависимость полной энергии SET $E^{tot} = E[n] + \Delta E$ (см. (34) и (38)) от потенциала затвора в случае молекулы бензола близка к линейной, а наклон определяется знаком и величиной заряда молекулы Q. Линеаризация полученных данных (рис. 18) в предположении линейной зависимости полной энергии от заряда молекулы

$$E(Q,V_g) = \alpha Q V_g \tag{39}$$

позволяет количественно оценить константу связи α в ур-ях (17) и (39) для молекулы бензола (табл. 3).

Таблица 3

Линеаризация зависимости полной энергии SET на молекуле бензола от потенциала затвора (рис. 18)

Заряд Q	+2	+1	0	– 1	– 2
Величина αQ	1.2323	0.6104	– 0.0143	– 0.6384	– 1.2660
Константа связи α	0.616	0.610	–	0.638	0.633

Среднее значение константы связи $<\alpha>$ для молекулы бензола в предположении линейной зависимости полной энергии SET от заряда молекулы получается равным $<\alpha> = 0.62$ по сравнению со значением $\alpha = 1$, которое было использовано выше при построении диаграмм зарядовой стабильности полевого транзистора на молекуле бензола (рис. 14 и 16). Окончательная диаграмма зарядовой стабильности полевого транзистора на молекуле бензола с учетом $<\alpha> = 0.62$ показана на рис. 19.

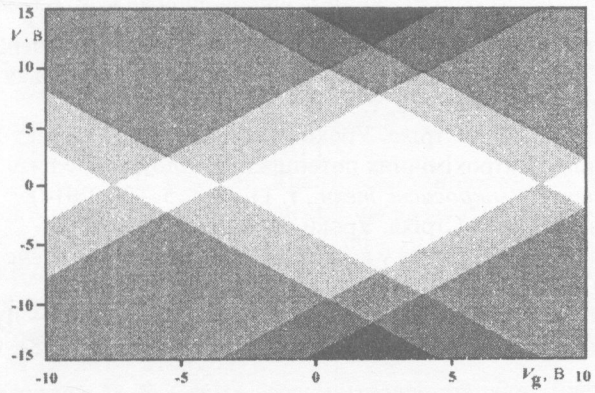


Рис. 19. Диаграмма зарядовой стабильности полевого транзистора на молекуле бензола. Диаграмма показывает число заряженных состояний молекулы в канале туннелирования в зависимости от значений V и V_g: в белых полях заряженных состояний нет, в остальных – число заряженных состояний изменяется от 1 до 4, возрастая с плотностью закраски полей.

Настоящие расчеты объясняют главные причины уменьшения энергии заряжания молекулы в условиях полевого транзистора, но, конечно, не все возможные эффекты были учтены. В частности, не учтено возможное изменение геометрии молекулы вблизи токообразующих электродов, как и образование поляронов при заряжании молекулы.

Подведем итоги. Явление кулоновской блокады в одноэлектронике рассмотрено с позиций двух казалось бы разных подходов: в концепции «снизу – вверх» наноэлектроники, в которой важную роль играет неограниченный по спину подход в теории самосогласованного поля в фоковском пространстве, и в традиционной концепции «сверху – вниз» с привлечением макроскопических свойств и понятий. Далее диаграмма зарядовой стабильности одноэлектронного полевого нанотранзистора на молекуле бензола в качестве проводящего канала в режиме кулоновской блокады рассчитана из первых принципов. Энергии заряжания молекулы вычислены квантовомеханически в приближении теории функционала плотности, а взаимодействие молекулы с окружающей ее средой в реалистической модели нанотранзистора учитывалось самосогласовано. Найдены условия функционирования такого одномолекулярного полевого нанотранзистора [22, 23].

Литература

1. Ю. О. Кругляк, М. В. Стріха. Уроки наноелектроніки: Метод нерівноважних функцій Гріна у матричному зображенні. 1. Теорія, *Сенсор. електр. мікросист. техн.*, т. 10, № 3, 22 – 35 (2013).
2. Ю. О. Кругляк, М. В. Стріха. Уроки наноелектроніки: Квантова інтерференція і дефазування в методі нерівноважних функцій Гріна, *Сенсор. електр. мікросист. техн.*, т. 11, № 3, 5 – 18 (2014).
3. Ю. А. Кругляк. Наноэлектроника «снизу – вверх»: Метод неравновесных функций Грина, модельные транспортные задачи и квантовая интерференция, *ScienceRise,* т. 9, № 2(14), 41 – 72 (2015).
4. Ю. О. Кругляк, М. В. Стріха. Уроки наноелектроніки: Ефект Холла і вимірювання електрохімічних потенціалів у концепції «знизу – вгору», *Сенсор. електр. мікросист. техн.*, т. 11, № 1, 5 – 27 (2014).
5. Ю. О. Кругляк, М. В. Стріха. Уроки наноелектроніки: Роль електростатики й контактів у концепції «знизу – вгору», *Сенсор. електр. мікросист. техн.*, т. 11, № 4, 27 – 42 (2014).
6. Ю. А. Кругляк. Модель проводимости Ландауэра – Датты – Лундстрома и транспортное уравнение Больцмана, *ScienceRise*, т. 3, № 2(8), 108 – 116 (2015).
7. P. Danielewicz. Quantum Theory of Non-Equillibrium Processes, *Ann. Phys.*, v.152, 239 (1984).
8. G. D. Mahan. Quantum Transport Equation for Electric and Magnetic Fields, *Phys. Rep.*, v. 145, 251 (1987).
9. S. Datta. *Quantum Transport: Atom to Transistor*. Cambridge: Cambridge University Press, 2005.

10. Ю. А. Кругляк. Наноэлектроника «снизу – вверх»: Роль электростатики и контактов, *ScienceRise*, т. 12, № 2(17) (2015)
11. P. C. Martin, J. Schwinger. Theory of many-particle systems. I, Phys. Rev., v. 115, n 6, 1342 – 1373 (1959).
12. L. P. Kadanoff, G. Baym. Quantum Statistical Mechanics. New York: W. A. Benjamin, 1962.
13. Л. В. Келдыш. Диаграммная техника для неравновесных процессов, ЖЭТФ, т. 47, 1515 – 1527 (1964); L. V. Keldysh. Diagram Technique for Non-Equilibrium Processes, Sov. Phys. JETP, v. 20, 1018 (1965)..
14. E. S. Kryachko, E. V. Ludena. Density functional theory: Foundations reviewed, *Physics Rep.*, v. 544, 123 – 239 (2014); DOI: 10.1016/j.physrep.2014.06.002.
15. D. V. Averin, K. K. Likharev. Single electronics: A correlated transfer of single electrons and Cooper pairs in systems of small tunnel junctions, In B. L. Altshuler, P. A. Lee, R. A. Webb (Eds.). *Mesoscopic Phenomena in Solids*. New York: Elsevier, 1991.
16. C. W. J. Beenakker. Theory of Coulomb-Blockade Oscillations in the Conductance of a Quantum Dot, *Phys. Rev. B*, v. 44, n 4, 1646 – 1656 (1991).
17. H. Grabert, M. H. Devoret (Eds.). *Single Charge Tunneling. Coulomb Blockade Phenomena In Nanostructures*. New York: Plenum Press, 1992.
18. Ю. А. Кругляк. Наноэлектроника «снизу – вверх»: Возникновение тока, обобщенный закон Ома, упругий резистор, моды проводимости, термоэлектричество, *ScienceRise*, т. 7, № 2(12), 76 – 100 (2015).
19. Ю. А. Кругляк. Обобщенный метод Хартри – Фока и его версии: от атомов и молекул до полимеров, *ScienceRise*, т. 5, № 5(3), 6 – 21 (2014).
20. Kruglyak Yu. A., Quantum-chemical studies of quasi-one-dimensional electron systems. 1. Polyenes // ScienceRise. – 2015. – т. 5, N 2(10). – С. 69 – 105.
21. Yu. A. Kruglyak. Quantum-mechanical studies of quasi-one-dimensional electron systems, Chapter 2 in Kruglyak Yu.A., Glushkov A.V. et al, *Calculational Methods in Quantum Geometry and Chaos Theory, Part 4.* Odessa: TES Publishing House, 2015.
22. Ю. А. Кругляк, Н. Е. Кругляк. Квантовомеханический расчет одноэлектронного полевого транзистора на молекуле бензола, *Сенсор. електр. мікросист. техн.*, т. 8, № 3, 60 – 70 (2011).
23. Ю. А. Кругляк, Н. Е. Кругляк. Одноэлектронный одномолекулярных полевой транзистор: квантовомеханическое и электродинамическое рассмотрение на примере молекулы бензола, *Вестник Одес. гос. эколог. ун-та*, № 12, 201 – 214 (2011).
24. Ю. О. Кругляк, М. В. Стріха. Уроки наноелектроніки: Електричний струм і другий закон термодинаміки у концепції «знизу – вгору», *Сенсор. електр. мікросист. техн.*, т. 12, № 2, 5 – 26 (2015).
25. Ю. А. Кругляк. Наноэлектроника «снизу – вверх»: Термодинамика проводника с током, информационно-управляемый акумулятор и квантовая энтропия, *ScienceRise*, т. 11, № 2(16), 55 – 71 (2015)

26. S. Datta. *Lessons from Nanoelectronics: A New Perspective on Transport.* Hackensack, New Jersey: World Scientific Publishing Company, 2012; www.nanohub.org/courses/FoN1.
27. Ю. А. Кругляк. Эффекты Холла, измерение электрохимических потенциалов и транспорт спинов в модели НРФГ, *ScienceRise*, т. 10, № 2(15), 35 – 67 (2015).
28. J. Park, A. N. Pasupathy, J. I. Goldsmith, C. Chang, Y. Yaish, J. R. Petta, M. Rinkoski, J. P. Sethna, H. D. Abruna, P. L. McEuen, D. C. Ralph. Coulomb blockade and the Kondo effect in single-atom transistors, *Nature*, v. 417, 722 – 725 (2002).
29. W. Liang, M. P. Shores, M. Bockrath, J. R. Long, H. Park. Kondo resonance in a single-molecule transistor, Nature, v. 417, n 13, 725 – 729 (2002).
30. S. Kubatkin, A. Danilov, M. Hjort, J. Cornil, J.-L. Bredas, N. Stuhr-Hansen, P. Hedegard, T. Bjornholm. Single Electron Transport Through a Single Conjugated Molecule Visiting Several Redox States, *Nature*, v. 425, 698 – 701 (2003).
31. E. A. Osorio, K. O'Neill, N. Stuhr-Hansen, O. F. Nielsen, T. Bjornholm, H. S. van der Zant. Addition energies and vibrational fine structure measured in electromigrated single-molecule junctions based on an oligophenylenevinylene derivative, *Adv. Mater.*, v. 19, 281 – 285 (2007).
32. A. Danilov, S. Kubatkin, S. Kafanov, P. Hedegard, N. Stuhr-Hansen, K. Moth-Poulsen, T. Bjornholm. Electronic Transport in Single Molecule Junctions: Control of the Molecule-Electrode Coupling through Intramolecular Tunneling Barriers, *Nano Lett.*, v. 8, n 1, 1 – 5 (2008).
33. J. M. Thijssen, H. S. van der Zant. Charge transport and single-electron effects in nanoscale systems, *Phys. Status Solidi B*, v. 245, n 8, 1455 – 1470 (2008).
34. K. Kaasbjerg, K. Flensberg. Strong Polarization-Induced Reduction of Addition Energies in Single-Molecule Nanojunctions, *Nano Lett.*, v. 8, n 11, 3809 – 3814 (2008).
35. K. Stokbro. First-Principles Modeling of Molecular Single-Electron Transistors, *J. Phys. Chem. C*, v. 114, n 48, 20461 – 20465 (2010).
36. I. O. Kulik, R. I. Shekhter. Kinetic Phenomena and ChargeDiscreteness Effects in Granulated Media, *Sov. Phys. JETP* , v. 4 , n 2, 308 (1975).
37. L. P. Kouwenhoven, C. M. Marcus, P. L. McEuen, S. Tarucha, R. M. Westervelt, N. S. Wingreen. Electron transport in quantum dots, In *Mesoscopic Electron Transport*, L. L. Sohn, L. P. Kouwenhoven, G. Schön (Eds.). New York: Kluwer, 1997.
38. U. Meirav, E. B. Foxman. Single-electron phenomena in semiconductors, *Semicond. Sci. Tech.*, v. 11, n 3, 255 – 284 (1996).
39. T. A. Fulton, G. J. Dolan. Observation of Single-Electron Charging Effects in Small Tunnel-Junctions, *Phys. Rev. Lett.*, v. 59, n 1, 109 – 112 (1987).
40. J. H. F. Scott-Thomas, S. B. Field, M. A. Kastner, H. I. Smith, D. A. Antoniadis. Conductance Oscillations Periodic in the Density of a One-Dimensional Electron Gas, *Phys. Rev. Lett.*, v. 62, n 5, 583 – 586 (1989).

41. M. A. Reed, J. N. Randall, R. J. Aggarwal, R. J. Matyi, T. M. Moore. Observation of Discrete Electronic States in a Zero-Dimensional Semiconductor Nanostructure, *Phys. Rev. Lett.*, v. 60, n 6, 535 – 537 (1988).
42. D. L. Klein, R. Roth, A. K. L. Lim, A. P. Alivisatos, P. L. McEuen. A single-electron transistor made from a cadmium selenide nanocrystal, *Nature*, v. 389, n 6652. 699 – 701 (1997).
43. S. J. Tans, M. H. Devoret, H. Dai, A. Thess, R. E. Smalley, L. G. Geerligs, C. Dekker. Individual single-wall carbon nanotubes as quantum wires, *Nature*, v. 386, n 6624, 474 – 477 (1997).
44. M. Bockrath, D. H. Cobden, P. L. McEuen, N. G. Chopra, A. Zettl, A. Thess, R. E. Smalley. Single-electron transport in ropes of carbon nanotubes, *Science*, v. 275, n 5308, 1922 – 1925 (1997).
45. Yu. A. Kruglyak. Configuration interaction in the second quantization representation: basics with applications up tp full CI, *ScienceRise*, т. 4, № 2(4), 98 – 115 (2014).
46. Yu. A. Kruglyak. Configuration interaction in the second quantization representation: basics with applications up tp full CI, Chapter 2 in A. V. Glushkov, O. Yu. Khetselius, Yu. A. Kruglyak *et al. Calculational Methods in Quantum Geometry and Chaos Theory, Part 3*. Odessa: TES Publishing House, 2015.
47. J. Neugebauer, M. Scheffler. Adsorbate-substrate and adsorbate-adsorbate interactions of Na and K adlayers on Al(111), *Phys. Rev. B.*, v. 46, n 24, 16067 – 16080 (1992).
48. J. Taylor, H. Guo, J. Wang. Ab initio modeling of quantum transport properties of molecular electronic devices, *Phys. Rev. B.*, v. 63, 245407 (2001).
49. J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejon, D. Sanchez-Portal. The SIESTA method for ab initio order-N materials simulation, *J. Phys. Condens. Matter.*, v. 14, 2745 – 2779 (2002).
50. M. Brandbyge, J.-L. Mozos, P. Ordejón, J. Taylor, K. Stokbro. Density-functional method for nonequilibrium electron transport, *Phys. Rev. B*, v. 65, n 16, 165401 (2002).
51. *Handbook of Chemistry and Physics*, 90th Ed., Editor-in-Chief David R. Lide. Cleveland: CRC Publishing, 2010.
52. Л. В. Гурвич, Г. В. Карачевцев, В. Н. Кондратьев, Ю. А. Лебедев, В. А. Медведев, В. К. Потапов, Ю. С. Ходеев. *Энергии разрыва химических связей. Потенциалы ионизации и сродство к электрону.* Москва: Наука, 1974.
53. J. C. Rivière. The Work Function of Gold, *Appl. Phys. Lett.*, v. 8, n 7, 172 (1966).
54. R. Lindner, H. Sekiya, B. Beyl, K. Muller-Dethlefs. Structure and Symmetry of the Benzene Cation, *Angew. Chemie Intern.*, v. 32, n 4, 603 – 606 (1993).

Ю. А. Кругляк
Профессор, д.ф.-м.н.
Действительный член
Академии наук
высшей школы Украины
по отделению физики

Научные интересы:
Квантовая химия
Физика молекул
Наноэлектроника

Основные результаты в области:
Методов самосогласованного поля
Учета корреляции электронов
Расчета молекул ***ab initio***
Молекулярной электроники
Прогнозирования свойств веществ

9 786177 337156