

НОБЕЛЕВСКИЕ ЛЕКЦИИ ПО ФИЗИКЕ — 2000

Квазиэлектрическое поле и разрывы зон. Обучение электронов новым фокусам

Г. Крёмер

(Нобелевская лекция. Стокгольм, 8 декабря 2000 г.)

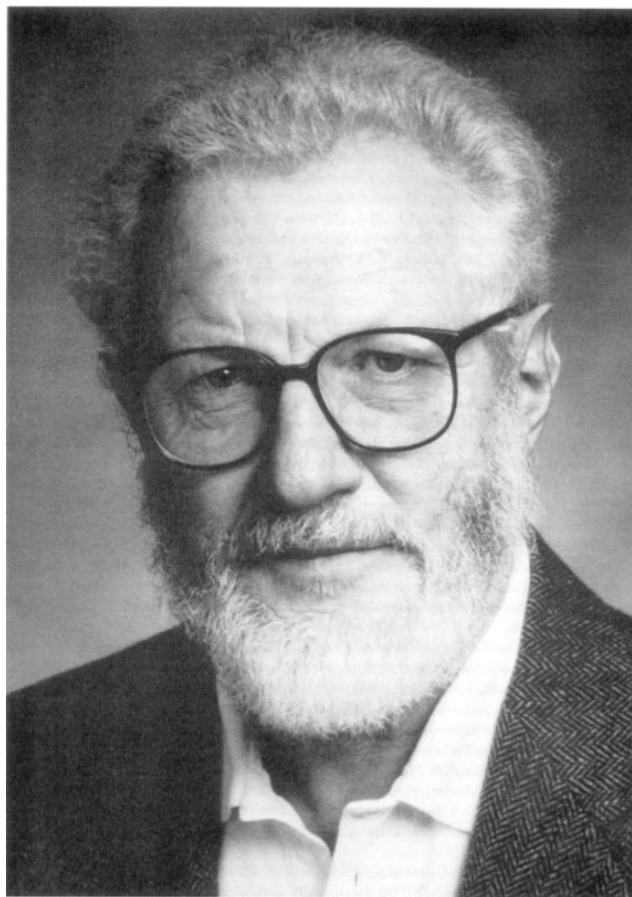
PACS number: 01.60. + q

Автобиография

Я родился 25 августа 1928 г. в Веймаре в Германии. Мой отец был чиновником городской администрации, мать — классической немецкой домохозяйкой — "Hausfrau". Оба происходили из семей квалифицированных ремесленников. Оба не имели среднего образования, но, несомненно, хотели дать детям наилучшее доступное для них образование. В частности, моя мать прилагала все усилия, чтобы мы были лучшими в школе — учиться просто хорошо было недостаточно. К счастью, я прошел через все 12 лет школы без особых усилий, ни разу не попросив помощи у своих родителей при выполнении домашних заданий.

Несмотря на стремление к совершенству, родители никогда не направляли меня жестко в какую-либо область знаний — я имел возможность свободно следовать своим наклонностям, занимаясь математикой, физикой и химией. Когда в конце концов я сказал родителям, что намерен посвятить себя именно физике, отец только поинтересовался, что это такое и удастся ли мне на этом поприще обеспечить себе средства к существованию. Конечно, можно было стать учителем физики в средней школе (называвшейся "Gymnasium"), что было вполне уважаемой профессией.

Впрочем, в школе у меня была одна большая проблема — поведение! Я часто нарушал дисциплину, скучая на уроках и развлекая себя по-разному. Обычным наказанием были замечания, которые записывались в классный журнал, "Klassenbuch". Подобные замечания рассматривались как серьезные прегрешения, и не будь я отличником, меня могли бы просто исключить из школы. Однажды, когда я мешал на занятиях всему классу, классная воспитательница — д-р Эдит Рихтер (Edit Richter), которую я обожал, — спросила меня в отчаянии: "Ну почему ты так себя ведешь?". Я объяснил, что мне скучно. В ответ она взорвалась: "Г-н Крёмер, одна из целей школьного образования состоит в том,



G. Kroemer

чтобы научиться скучать в благопристойной форме". Слова эти я никогда не забывал, но, к сожалению, благопристойно скучать так и не научился.

У другого учителя, Вилибальда Виммера (Willibald Wimmer), был свой собственный остроумный способ общения со мной. Еще до окончания войны он был

Г. Крёмер (H. Kroeмер). ECE Department, University of California, Santa Barbara, CA 93106, USA

© The Nobel Foundation 2000

© Перевод на русский язык — Российская академия наук, "Успехи физических наук" 2002

преподавателем местного технического училища, а затем стал преподавать математику и физику в нашей школе. Он привык иметь дело со студентами и обращался с нами как со взрослыми. Я шел по математике впереди учебной программы и постоянно демонстрировал это. Более того, я обучил некоторых своих одноклассников математическим приемам, которые не входили в нашу учебную программу. В результате г-н Виммер заключил со мной "договор". Он не мог освободить меня от посещения занятий, но гарантировал мне высшую оценку без выполнения домашних заданий. На уроке мне разрешалось делать все что угодно *при условии*, что я буду молчать, за исключением тех случаев, когда меня вызывали к доске. Мы оба соблюдали этот договор.

Виммер был у нас также и учителем физики, о которой он знал немногим больше того, что имелось в учебнике. Поняв, что я уже глубоко вник в физику, он попросту поручил мне и еще одному ученику помогать ему в подготовке лекций, например, выяснять, какие приборы сохранились со времени войны. Однажды он даже попросил меня, чтобы я сам прочел лекцию, при этом он сидел в первом ряду и наслаждался зрелищем. Это было прекрасно!

После окончания гимназии в 1947 г. меня приняли студентом по специальности "физика" в Университет в Йене, где я попал под обаяние великого Фридриха Хунда (Friedrich Hund), самого замечательного лектора, которого мне довелось когда-либо слушать. Но счастье было недолгим. В начале 1948 г. политическое давление в Восточной Германии стало слишком ощутимым, особенно в таком непокорном заведении как Йенский университет. Каждую неделю кто-то из моих товарищей неожиданно исчезал, и нельзя было узнать, отправился ли он на Запад или попал в немецкое отделение сталинского ГУЛАГа, вроде урановых шахт на границе с Чехословакией. Во времена "Берлинского воздушного моста"¹ я был в Берлине на летней практике в компании Сименс. Тогда я принял решение перебраться на Запад на одном из самолетов, выполняющих порожняком обратный рейс.

Еще будучи в Берлине, я обратился в некоторые западногерманские университеты, в том числе и в Гёттингенский, с просьбой о зачислении, но до отъезда из Берлина не получил ни одного ответа (фактически они мне отказали). Один из моих преподавателей в Йене сказал мне перед моим отъездом из Берлина: "Почему бы Вам не передать привет от меня профессору Кёнигу (König) в Гёттингене?" Я последовал этому совету. Кёниг ответил, что зачисление на специальность "физика" уже закончено и направил меня, как он сказал, "для дружеского разговора" к профессору Рихарду Беккеру (Richard Becker) и его *alter ego*, ассистенту д-ру Гюнтеру Лейбфриду (Günther Leibfried). Они, в свою очередь, рекомендовали меня Вольфгангу Паулю (Wolfgang Paul, Нобелевская премия 1989 г.) и, я думаю, также Роберту Поллю (Robert Pohl). Скоро мне стало ясно, что это были не просто обычные дружеские беседы с людьми, которым было нечего делать, но, фактически, строгие экзамены. Я помню один из вопросов, заданных мне Паулем: "Вы знаете, что зеркало превращает левое в правое? Тогда

почему оно не переставляет верх и низ?". В конце концов, меня отправили обратно к Беккеру, который сообщил, что двое из уже зачисленных студентов не придут и что на завтра назначено совещание, на котором решат, кого взять на две появившиеся вакансии. Двумя днями позже я получил почтовую открытку с сообщением, что я принят.

Послевоенный Гёттинген в интеллектуальном отношении был удивительно стимулирующим творчеством местом. Мне понравился один из самых молодых преподавателей — "Privatdozent", д-р Хелльвеге (Hellwege), который предложил мне посещать так называемый "предсеминар"², на котором студенты делали доклады на рекомендованные темы. Я принимал участие в этом семинаре в течение нескольких семестров. Однажды знаменитый Фриц Хоутерманс (Fritz Houtermans) посетил Хелльвеге и прослушал несколько докладов, в том числе и мой. Я говорил о чьих-то данных, хорошо укладывающихся на прямую линию в двойной логарифмической шкале, и гордо утверждал, что этим данным соответствует степенная зависимость. На Хоутерманса это не произвело никакого впечатления: "В двойной логарифмической шкале моя бабушка тоже соответствует прямой линии". Я теперь всегда цитирую "бабушку Хоутерманса" моим студентам. В конечном счете я записался к Хелльвеге для выполнения дипломной работы, предполагавшей экспериментальное исследование оптических спектров некоторых солей редкоземельных элементов. Но у Хелльвеге был длинный список претендентов на работу, а тем временем профессор Фриц Заутер (Fritz Sauter) — беженец, который нашел временный приют в виде должности приглашенного исследователя в Институте теоретической физики Беккера, — предложил мне теоретическую тему для дипломной работы, основанную на содержании моего доклада, который я сделал на одном из его семинаров. Хелльвеге посоветовал мне принять предложение Заутера: "Вы с ним уже закончили работу до того, как сможете начать ее со мной". Так я стал теоретиком.

Тема диплома была продолжением работы Шокли (Shockley) 1939 г. о природе поверхностных состояний в одномерном потенциале. Работая над этой темой, я решил рассмотреть границу двух разных периодических потенциалов и в первый раз столкнулся с тем, что теперь принято называть разрывом зон на гетерогранице.

Когда я работал у Заутера, была еще одна неожиданная встреча с гетероструктурами. Мы отправились в исследовательскую лабораторию электротехнического концерна Allgemeine Elektrizitäts-Gesellschaft в Белеке (Belecke), маленьком городке в Вестфалии, и там д-р Поганьски (Poganski) убедительно продемонстрировал, что селеновый выпрямитель является не барьером Шоттки, а р-п-переходом между селеном р-типа и CdSe п-типа, т.е. истинным гетеропереходом (хотя в то время этого термина еще не существовало). По-видимому, это оказало на меня, по меньшей мере, подсознательное влияние: когда я позже начал всерьез размышлять о гетеропереходах, вопрос, может ли такой объект существовать как реальный прибор, имел очевидный ответ: конечно!

¹ Берлинский воздушный мост был организован западными союзниками (США, Великобритания и Франция) во время советской блокады Западного Берлина в 1948–1949 гг. (Примеч. ред.)

² Proseminar — занятия для студентов и аспирантов, проводящиеся в форме научного семинара. (Примеч. переводчика.)

Во время работы над дипломом я выступил на коллоквиуме Заутера с докладом по известной работе Бардина (Bardeen) и Браттейна (Brattain) "Физические принципы работы транзистора" (или с каким-то похожим названием). В конце я высказал свои соображения по поводу некоторых нерешенных проблем, сформулированных авторами. Заутер заинтересовался и предложил взять эти проблемы в качестве возможной темы докторской диссертации. Несколько позже он пришел ко мне в комнату и предложил приостановить работу над дипломом и просто написать о том, что уже сделано к настоящему моменту. Несмотря на мои протесты он настоял на том, что пришло время заниматься реальными проблемами, т.е. диссертацией.

Таким образом, я столкнулся с твердым убеждением Заутера, по-видимому, восходящим к традициям 20-х годов, что ученые степени присуждаются не "за выслугу лет", а за способность соискателя выполнить теоретическую работу самостоятельно, без руководителя. Фактически, он явно предпочитал, чтобы диссертации делались быстро. В результате я получил ученую степень Ph. D. до того как мне исполнилось 24 года, т.е. довольно рано даже для теоретика. Это было замечательно!

Диссертация была посвящена явлению, которое мы теперь назвали бы *влиянием горячих электронов* на слой пространственного заряда в коллекторе транзистора — нового для того времени устройства. Идея была простой. О зонной структуре Ge было почти ничего не известно, но из выполненных кем-то теоретических оценок следовал неверный вывод об очень узких зонах, в частности, очень узкой валентной зоне. В этом случае при достаточно сильном поле дырки в валентной зоне должны были демонстрировать явление, которое мы теперь называем блоховскими осцилляциями. С помощью нескольких алгебраических формул можно было показать, что при заданной плотности тока концентрация свободных дырок должна расти с увеличением поля — так называемый "Stau-effect" (*нем.* эффект накопления), приводя к сильным изменениям пространственного заряда. Влияние этого пространственного заряда на вольт-амперные характеристики диодов с точечными контактами и транзисторов было основным содержанием диссертации.

Мои алгебраические расчеты предсказывали также уменьшение дрейфовой скорости электронов при увеличении поля, что означало отрицательную дифференциальную проводимость. Не зная практически ничего о теории электрических цепей, я не подозревал, насколько полезным может быть такое явление, пока Шокли не указал мне на это в частном разговоре двумя годами позже.

Однако скоро стало ясно, что моя диссертация не имела отношения к действительности. Мои предположения о зонной структуре и о не зависящей от энергии длине свободного пробега были неверны, и после открытия лавинного пробоя стало очевидно, что такие огромные поля, которые необходимы для блоховских осцилляций в *объемных* полупроводниках, вообще невозможно получить. Спустя 20 лет, после пионерской работы Есаки (Esaki) и Цу (Tsu) об отрицательной дифференциальной проводимости в сверхрешетках, я понял, что, фактически, предвидел их результаты, хотя и в более примитивной форме — то, что было невозможно в объемных материалах, оказалось возможным в сверхрешетках, период которых намного больше.

Вернемся к Заутеру. Его не интересовало скрупулезное руководство работой студентов, он просто наблюдал, что они делают самостоятельно. Тем не менее, он оказал на меня огромное влияние в вопросах методологии. Каждый раз, когда я приходил к нему с чисто физической идеей, он непременно говорил с легким сарказмом: "Но, г-н Крёмер, Вы должны сформулировать это математически". Если же я приходил к нему с математической формулировкой, то слышал произнесенное тем же тоном: "Но, г-н Крёмер, это всего лишь математика, где же физика?". После нескольких таких встреч суть его требований становилась понятной: необходимо было уметь с легкостью переходить от одного описания к другому. Однако при окончательном анализе физической концепции отдавался приоритет перед формализмом, последний был просто средством (необходимым) для завершения работы.

Приоритеты, таким образом, были четко расставлены, и такой подход сильно повлиял на меня. Ученик Зоммерфельда, Заутер сам был превосходным математиком, но он терпеть не мог, когда люди для демонстрации своих математических способностей использовали более сложный формализм, чем было необходимо для решаемой задачи. Поскольку он был признанным специалистом по функциям Бесселя, я однажды почувствовал непреодолимое желание включить в рукопись моей диссертации некую специально придуманную задачу, решение которой требовало использования функций Бесселя. Он был не в восторге: "Здесь это не нужно. Вы включили этот материал, чтобы произвести на меня впечатление. Уберите его!"

У Рихарда Беккера было такое же отношение к делу (они были близкими друзьями), и позже я встретил подобное же отношение у Шокли. Под таким влиянием я не превратился в "закоренелого Теоретика с заглавной буквы Т", а стал скорее "концептуалистом" — генератором идей, который остро осознает ограниченность своих возможностей как "формалиста" и для которого Нильс Бор, более чем кто-либо другой из Великих Физиков, является примером для подражания.

В Германии 1952 г. рынка труда для физиков-теоретиков практически не существовало. Не было новых университетских должностей, но было много специалистов старшего поколения, ожидающих любого освободившегося места. Поэтому я даже не рассматривал университетскую карьеру в качестве возможной. Ситуация в промышленности была не лучше. Однако по счастливой случайности небольшой исследовательской группе, занимающейся полупроводниками в Центральной телекоммуникационной лаборатории (Fernmelde-technisches Zentralamt, FTZ) Германского почтового ведомства, понадобился "домашний теоретик", знающий физику полупроводников, и мне удалось получить эту работу. Мои обязанности были простыми. Я должен был быть всегда доступен для обсуждения любых теоретических вопросов, возникающих у членов группы; моя активная роль заключалась в том, чтобы совать свой нос в дела моих коллег — экспериментаторов и технологов, а также находить самому, чем я мог бы быть полезен, при условии, что не буду притрагиваться к экспериментальному оборудованию. Каждые 1-2 недели я должен был делать одно- или двухчасовой доклад, самостоятельно выбирая такую тему, которую, по моему мнению, полезно было бы освоить членам

группы. Помимо этого я был полностью свободен в выборе тех проблем, которыми, как я считал, стоило бы заняться. Таким образом, я стал чем-то вроде "профессора", обучающего небольшую, но целеустремленную группу "студентов". С первого же дня мне пришлось учиться общению не с другими теоретиками, а с экспериментаторами и технологами. Мне была предоставлена замечательная возможность испытать себя, занимаясь широким кругом вопросов, выходящих далеко за пределы того, что я изучил в Гёттингене, а часто и вообще за рамки физики — в металлургию, химию или электронную технику.

Конечно, я перестал быть "настоящим" физиком-теоретиком, если я когда-либо им был. Если угодно, меня можно называть "прикладным теоретиком". Однако уверенность, что я действительно делаю что-то полезное, помогла мне преодолеть неприятное ощущение завершения карьеры теоретика в самом ее начале. В конце концов, оглядываясь назад, можно сказать, что это было не таким уж плохим поворотом судьбы!

В качестве темы моих исследований в FTZ я решил заняться поисками путей решения проблемы жесткого частотного ограничения в новых транзисторах. Именно эта проблема привела меня прямо к идее гетероперехода. В моей публикации 1954 г. имеется пара разделов, в которых в примитивной форме впервые излагается идея устройства, позже названного биполярным транзистором на гетероструктурах (heterostructure bipolar transistor, HBT). Я предложил транзистор с базой, имеющей переменную ширину запрещенной зоны, и более простую разновидность с широкозонным эмиттером. Остальное — это история, изложенная достаточно подробно в моей Нобелевской лекции, и потому я остановлюсь здесь только на ее основных моментах.

В 1954 г., вскоре после моего перехода в лабораторию RCA в Принстоне, штат Нью-Джерси, я вернулся к проблеме гетероперехода. Фактически я пытался (неудачно) создать HBT-транзистор на основе эмиттера из сплава Ge/Si и базы из Ge. Тем не менее, я внес свой вклад в эту область — наиболее важными здесь являются две теоретические статьи. Одна из них, опубликованная в *RCA Review*, малоизвестна и сегодня, хотя в ней ясно сформулирована концепция *квазиэлектрического поля*, которую я считаю основным принципом конструирования для всех гетероструктур.

Последний шаг был сделан в 1963 г., когда я работал в Varian Associates в Пало-Альто, Калифорния. Мой коллега, д-р Сол Миллер (Sol Miller) докладывал на семинаре о новых полупроводниковых лазерных диодах. В частности, он сказал, что, по мнению специалистов, в принципе нельзя получить равновесную инверсную заселенность при комнатной температуре, поскольку инжектированные носители слишком быстро диффундируют на другую сторону гетерограницы. Я сразу возразил: "Это ерунда, нужно просто сделать внешнюю область из широкозонного материала". Я сформулировал свою идею и направил статью в *Applied Physics Letters*, но она была отклонена. Мне посоветовали не бороться с этим решением, а представить статью в *Proceedings of the IEEE*, где она и была опубликована, но на нее никто не обратил внимания. Я также оформил патент, который написан, по-видимому, лучше, чем статья в *Proc. IEEE*.

И вот, наконец, последняя насмешка судьбы: мне было отказано в финансировании работы по лазерам нового типа на том основании, что она не будет иметь никакого практического выхода. По случайному стечению обстоятельств, в этот момент был открыт эффект Ганна. Давно испытывая интерес к явлению отрицательного сопротивления на горячих электронах, я следующие 10 лет работал над эффектом Ганна и не принимал участия в окончательной технологической реализации этих лазеров.

Я ушел из Varian в 1966 г., а в 1968 г. устроился в Колорадский университет. Там я в конце концов вернулся к гетероструктурам. В начале 70-х я занялся теорией разрывов зон совместно с моим аспирантом Биллом Френсли (Bill Frensley), который работает теперь в Техасском университете в Далласе. Он и разработал первую *ab initio* теорию разрывов зон. Вскоре после этого, уже в Калифорнийском университете в Санта-Барбаре, я предложил эффективный экспериментальный метод определения разрывов зон путем измерения зависимости емкости от напряжения через границу гетероперехода.

В конце 70-х я вернулся к прибору, с которого все и начиналось, а именно к НВТ. Технологические достижения, которые сделали возможным создание ДН-лазера (double heterostructure laser — лазер на двойной гетероструктуре), открывали перспективы также и для создания НВТ, и я стал горячим сторонником разработки этого устройства.

Помимо гетероструктур, я занимался многими другими проблемами полупроводников, независимо от того, относились они к физике, материаловедению, приборостроению или технологии. Вторым моим неизменным пристрастием после гетероструктур было отрицательное сопротивление на горячих электронах — этим эффектом я начал интересоваться еще во время работы над диссертацией на степень Ph. D. Я уже упоминал одну мою работу по эффекту Ганна, но их было несколько. В годы работы в RCA я увлекся сумасшедшим проектом — получить отрицательное сопротивление в направлении, *перпендикулярном* сильному полю смещения, используя тот факт, что некоторые тяжелые дырки в Ge имеют отрицательную поперечную эффективную массу, т.е. получить отрицательное сопротивление в направлении, фактически перпендикулярном скорости этих дырок. С точки зрения эксперимента это была очередная неудача, но сама идея казалась мне чрезвычайно вдохновляющей. Другим тоже так казалось, и тогда это принесло мне определенную известность. Сегодня я вернулся к одному из грехов моей юности — блоховскому осциллятору на сверхрешетке, интересной комбинации физики гетероструктур и физики горячих электронов.

Совсем в другой области по отношению к горячим электронам находится моя недавняя работа по слабой связи в сверхпроводящей цепи, в которой квантовая яма InAs/AlSb при близком к вырождению модуляционном легировании осуществляет баллистическую связь между сверхпроводящими Nb электродами. При этом обнаружено восхитительно большое расхождение экспериментальных данных и существующей теории.

Помимо вышеперечисленных, в моей профессиональной жизни было множество других проблем. В сущности, я всегда старался не упустить свой шанс и ничуть этого не стыжусь.

PACS numbers: 01.30.Bb, 42.55.Px, 78.67.-n

Нобелевская лекция

1. Введение

Гетероструктуры, о которых я здесь буду говорить, можно определить как неоднородные полупроводниковые структуры, изготовленные из двух или более различных материалов таким образом, что переходный слой, или граница раздела двух материалов играет важную роль в любом протекающем в приборе процессе. Можно даже сказать, что *техническим устройством является сама граница раздела*.

Все материалы, из которых делаются гетероструктуры, относятся к центральной части периодической системы элементов (табл. 1). В середине находится кремний — основа современной электроники. Под кремнием находится Ge. Хотя сам Ge используют крайне редко, сплавы Ge–Si разного состава играют все возрастающую роль в современной технологии гетероструктур. Исторически это было первое технически эффективное устройство с гетероструктурой, хотя его практическая разработка потребовала больших усилий из-за 4%-ного различия постоянных решетки Si и Ge.

Таблица 1. Центральная часть Периодической таблицы элементов. Показаны элементы II–VI групп, которые широко используются в современной технологии гетероструктур

II	III	IV	V	VI
Zn	Al	Si	P	S
Cd	Ga	Ge	As	Se
Hg	In		Sb	Te

Кремний играет такую же важную роль в технологии электронных материалов, что и сталь в металлургии материалов конструкционных. Но, подобно тому, как современная металлургия занимается кроме стали и другими металлами, электроника, помимо кремния, имеет дело, например, с полупроводниковыми твердыми растворами. Каждый элемент III группы может вступать в соединение с любым элементом V группы. При этом возникают соединения элементов III и V групп, так называемые $A^{III}B^V$. Из приведенных в табл. 1 элементов можно создать 12 различных соединений. Наиболее часто используемое в технике соединение — арсенид галлия GaAs, однако любое из этих соединений может применяться в гетероструктурах — это зависит от конкретной цели. Фактически сегодня соединения $A^{III}B^V$ используются, как правило, не в чистом виде, а в составе гетероструктур.

Для получения твердых растворов могут использоваться два или большее число отдельных соединений. Наглядный пример — соединение алюминий–галлий–мышьяк, $Al_xGa_{1-x}As$, где x — доля узлов элементов III группы, занятых атомами Al, а $(1-x)$ — доля узлов, занятых атомами Ga. Таким способом можно получить не только 12 соединений, а непрерывный ряд материалов. Это открывает возможность создания гетероструктур с переменным составом, меняющимся вдоль структуры непрерывно, а не скачком.

Подобно соединениям $A^{III}B^V$, каждый элемент II группы может соединяться с каждым элементом

VI группы, при этом возникают соединения $A^{II}B^{VI}$, причём снова можно получить непрерывный ряд сплавов.

2. Зонная структура и квазиэлектрические силы

Всякий раз, когда я читаю лекцию по физике полупроводниковых приборов, одна из основных идей, которую я стараюсь донести до аудитории, — это важность построения схемы энергетической зонной структуры. Я часто формулирую это в виде "Леммы Крёмера о Непонимании":

Если при обсуждении проблем физики полупроводников вы не можете нарисовать *Энергетическую Зонную Структуру*, значит, вы не понимаете, о чем вы говорите.

Отсюда следует, что если вы можете ее нарисовать, но не делаете этого, значит ваши слушатели не понимают, о чем вы говорите.

Это тем более верно при обсуждении гетероструктур, понимание сути которых во многом зависит от умения изобразить схему зонной структуры и осознать, что она означает.

Чтобы пояснить эти утверждения, рассмотрим однородный кусочек полупроводника, например, однородно легированного кремния, к которому приложено постоянное электрическое поле. Зонная структура тогда выглядит так, как показано на рис. 1а. На рисунке имеются две параллельные линии, изображающие края зоны проводимости и валентной зоны. Зазор между линиями — это запрещенная зона полупроводника. Угол наклона зон определяется произведением элементарного заряда электрона e на электрическое поле E . Если электрон или дырка помещены в такую систему, то на электрон действует сила $-eE$, а на дырку — сила $+eE$. Эти силы

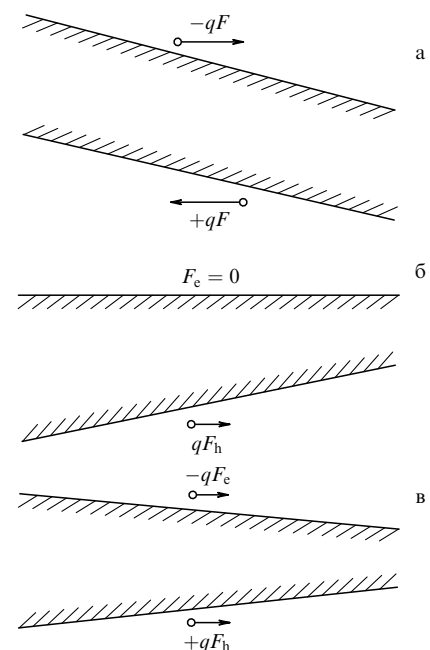


Рис. 1. Квазиэлектрическое поле. (а) Истинное электрическое поле просто наклоняет зоны; (б) квазиэлектрическое поле, в котором нет силы, действующей на электроны, но есть сила, действующая на дырки; (в) квазиэлектрическое поле, перемещающее электроны и дырки в одном направлении. (Из работы [1].)

равны по величине и противоположны по направлению, причем их величина дает углы наклона зон, отличающиеся только знаками.

В гетероструктуре ширина запрещенной зоны зависит от координаты, и углы наклона зон оказываются неодинаковыми. Разными являются при этом и величины сил. Например, может существовать сила, действующая только на один тип носителей (рис. 1б), или силы, которые действуют в одном и том же направлении на оба типа носителей (рис. 1в). Сила чисто электрической природы в однородном материале не может оказывать такое действие. Поэтому я называю эти силы "квазиэлектрическими". Они дают разработчику приборов дополнительную степень свободы, позволяющую получить эффекты, которые в принципе невозможно получить с помощью "настоящих" электрических полей.

Это и есть основной принцип конструирования приборов на основе гетеропереходов, впервые высказанный мною в статье 1957 г. [1]. Фактически, предшествующий абзац является несколько улучшенной версией основного утверждения этой статьи.

Когда я писал эти строки, я еще не знал о знаменитом патенте Шокли 1951 г. [2], где прямо говорится, что можно изготовить биполярный транзистор с эмиттером из широкозонного материала. Однако, по-видимому, идея широкозонного эмиттера упоминалась там главным образом для того, чтобы перекрыть дорогу другим возможным предложениям, как часто бывает в патентной практике. В патенте не было указано, почему широкозонный вариант эмиттера имеет заметные преимущества перед приборами на гомопереходах, и в нем еще меньше говорилось об общем принципе конструирования, который мог бы быть использован и для устройств другого типа. Моя собственная формулировка может рассматриваться как широкое обобщение идеи, высказанной в патенте Шокли. Но моя отправная точка была совсем иной: вместо резкого изменения ширины запрещенной зоны, сопровождающегося разрывом зон, непрерывное ее изменение на любую наперед заданную величину, для которого резкий скачок является просто частным случаем.

Возвращаясь к рис. 1б, следует подчеркнуть, что показанный там нулевой наклон зоны проводимости не означает отсутствия электрического поля. Истинное электрическое поле, конечно, присутствует, и его можно, в принципе, найти из уравнения Пуассона при условии, что локальное пространственное распределение плотностей зарядов известно, хотя и это само по себе часто представляет сложную задачу. Но это истинное поле не является частью схемы энергетических зон. Оно не действует на электроны: важным является наклон зон, а не истинное электрическое поле. Различие между этими двумя полями становится еще более заметным на рис. 1в, из которого невозможно определить даже направление внешнего поля, не говоря уже о его величине.

3. Биполярные транзисторы на гетероструктурах

3.1. Транзисторы с переменной шириной запрещенной зоны

К идее 1957 г. меня привел вполне практический вопрос, возникший еще в 1953–1954 гг., когда я работал в

Центральной телекоммуникационной лаборатории (Fernmeldetechnisches Zentralamt, FTZ) Почтового ведомства ФРГ. Дело в том, что первые биполярные транзисторы на р–п-переходах не годились для практических применений в телекоммуникациях, так как работали слишком медленно. Я поставил себе задачу объяснить теоретически причину такого частотного ограничения и выяснить, что с этим можно сделать. Одна из возможностей (не единственная) состояла в увеличении скорости потока неосновных носителей из эмиттера в коллектор путем встраивания электрического поля в область базы. Это можно сделать с помощью неоднородного легирования базы, спадающего экспоненциально от эмиттера к коллектору, — так называемый дрейфовый транзистор [3]. Разрабатывая этот транзистор, я понял, что

"... дрейфовое поле можно также создавать, изменяя ширину самой запрещенной зоны. Это можно сделать, изготавливая область базы из нестехиометрического твердого раствора полупроводников с различными ширинами запрещенной зоны (например, Ge–Si), причем химический состав должен непрерывно изменяться вдоль базы." (Цитата из моей статьи 1954 г. [4].)

Это еще не было полной формулировкой принципа конструирования, но здесь уже содержалась оригинальная идея устройства, которое впоследствии стало называться "биполярным транзистором на гетеропереходе" (heterostructure bipolar transistor, НВТ) и, более того, общая идея эффективного поля в устройстве на гетеропереходе.

Соответствующая схема энергетических зон (рис. 2) была изображена в указанной выше статье 1957 г., где я привел эту идею 1954 г. в качестве примера использования общего принципа конструирования. Заметим, что на рис. 2 изображена плоская зона проводимости, как и должно быть в случае достаточно сильного однородного легирования. Зонная структура на рис. 1б представляет, в основном, область базы в этом старом варианте. Рисунок 1в иллюстрирует общую идею принципа конструирования.

Следует отметить, что в моих первоначальных работах в качестве примера была явно указана система Ge–Si, а не система на основе соединений $A^{III}B^V$. Прошло почти четыре десятилетия, пока НВТ-транзисторы на основе Ge–Si поступили, наконец, в продажу. Это

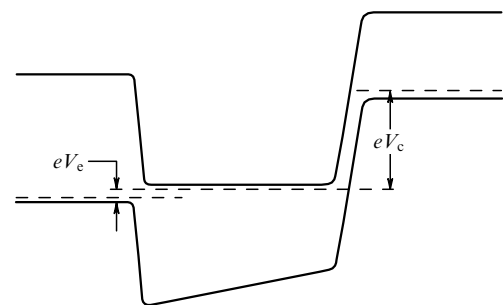


Рис. 2. р–п–р-транзистор с базовой областью, имеющей переменную ширину запрещенной зоны для ускорения потока неосновных носителей от эмиттера к коллектору (из работы [1]). р–п–р-транзисторы были предпочтительной конструкцией для транзисторов на основе Ge в середине 1950-х.

случилось намного позже, чем для устройств на основе $A^{III}B^V$.

3.2. Широкозонный эмиттер

Предложенная мною структура базы с переменной шириной запрещенной зоны не соответствовала технологическим возможностям того времени, и такая ситуация сохранялась в течение десятилетий. Единственный возможный вариант, предугаданный одним из моих коллег, Альфонсом Хенляйном (Alfons Hähnlein), — это конструкция, в которой эмиттер сделан из более широкозонного полупроводника, чем материал базы, с квазирезким переходом на границе раздела этих двух материалов, что приводило к зонной структуре, схематически изображенной на рис. 3. В сущности, Хенляйн воспроизвел (неосознанно) замысел Шокли.

Конечно, было очевидно, что встроить дрейфовое поле в базу транзистора таким путем невозможно. Однако обдумав возможные свойства такой структуры, я понял, что широкозонный эмиттер имеет свои преимущества [5, 6]. Одна из задач, общая для всех биполярных транзисторов, состоит в минимизации крайне нежелательной обратной инжекции основных носителей из базы (электронов в $p-n-p$ -транзисторах) в эмиттер. В транзисторе на основе гомоперехода это требование резко ограничивает уровень легирования базы, что, в свою очередь, может приводить к другим нежелательным последствиям, таким, как большое входное сопротивление базы. Широкозонный эмиттер в значительной мере подавляет этот ток обратной инжекции. Если не использовать термины квазиэлектрических сил и говорить о соответствующих потенциалах, любые уходящие из базы в эмиттер электроны должны преодолевать более высокий потенциальный барьер, чем дырки, поступающие в базу из эмиттера. В результате электронная плотность тока утечки уменьшается примерно в $\exp[-\Delta E_g/kT]$ раз, где ΔE_g — разность ширины запрещенных зон. Это весьма эффективный метод: легко достижимая разница ширины зон порядка 0,2 эВ ($\approx 8kT$) приводит к уменьшению плотности тока в $\exp 8 \approx 3000$ раз.

В таких условиях можно гораздо сильнее легировать базу, понижая тем самым ее сопротивление. Но при наличии неустраняемых емкостей на границах раздела

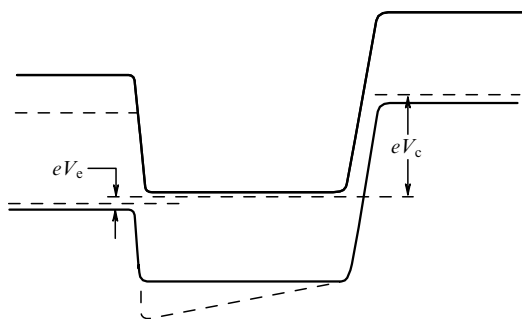


Рис. 3. Эмиттер с широкой запрещенной зоной. Область изменения ширины запрещенной зоны сжата в квазирезкий переход на границе эмиттер–база. Область базы все еще имеет однородную ширину запрещенной зоны без способствующего переносу квазиполя, но уже появился потенциальный барьер, препятствующий уходу электронов из базы в эмиттер, который больше барьера для дырок, входящих в базу из эмиттера.

уменьшение сопротивления базы уменьшает RC-временную постоянную прибора и, следовательно, увеличивает его быстродействие.

Благодаря большой технологической простоте изготовления широкозонного эмиттера по сравнению с конструкцией с переменной шириной запрещенной зоны, именно широкозонный эмиттер чаще всего используется в НВТ-технологии, но в наилучших НВТ-транзисторах теперь применяют оба метода [7].

3.3. Дальнейшие шаги

Из-за отсутствия какой-либо надежной технологии я не добивался полной реализации поставленных в 1954 г. целей до того времени, когда тремя годами позже перешел работать в RCA Laboratories в Принстоне, Нью-Джерси. Я отдавал себе отчет в общности сформулированного выше основного принципа конструирования и опубликовал статью в *RCA Review* [1] со ссылкой на мою более раннюю публикацию [4]. Эта статья была почти сразу же забыта, но не только потому, что *RCA Review* был малоизвестным журналом, но, возможно, в большей степени из-за того, что сам я никогда в своих последующих работах не ссылался явно ни на эту статью, ни на предшествующую ей работу 1954 г. вплоть до статьи 1996 г., опубликованной почти через сорок лет [8]. Сам общий принцип конструирования подробно обсуждался в обзоре по НВТ 1982 г. [6], но ссылка на статью 1954 г. и статью в *RCA Review* 1957 г. в нем не было.

Моя действительно широко цитируемая статья 1957 г. [5] была второй публикацией в том году. В ней был дан подробный анализ варианта широкозонного эмиттера для НВТ. Она была опубликована в более известном журнале, а потому привлекла большое внимание и стимулировала несколько попыток других ученых реализовать широкозонный эмиттер в НВТ еще в 1960-е гг. К сожалению, технология тогда еще не была достаточно развита, и ни одна из этих ранних попыток не была успешной. К 1970 году большинство ученых, по-видимому, перестали заниматься этой проблемой.

Когда я работал в RCA, я также предпринял неудачную попытку изготовить Ge транзистор с эмиттером из твердого раствора Ge–Si, которая представляется достаточно любопытной, чтобы упомянуть здесь о ней, и характерной для примитивного состояния технологии в 1957 г. [9]. Я решил использовать тот факт, что на фазовой диаграмме Au–Si имелась эвтектика с низкой температурой плавления (370 °C). Я приготовил такую эвтектику, измельчил молотком весьма хрупкий материал до состояния крупного порошка, поместил частицы порошка на кристалл Ge и расплавил все это при температуре где-то между 500 и 600 °C. Сплав Au–Si, расплавившись, должен был проникнуть в кристалл Ge и растворить некоторое количество Ge. После охлаждения должен был кристаллизоваться Ge–Si эмиттер (рис. 4). Я действительно получил таким способом один-два работающих транзистора, но в большинстве случаев термические деформации, возникающие при затвердевании эвтектики, приводили к растрескиванию кристалла Ge. Мои эксперименты были настолько неудачными, что я даже не опубликовал эту работу. Попытку повторили Дидрих и Йоттен в 1961 г. [10], которые знали о моей работе, но методика явно была неперспективной, и

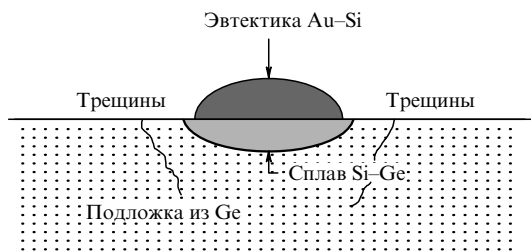


Рис. 4. Попытка создания Ge транзистора с эмиттером из сплава Ge–Si. Кусочек эвтектики Au–Si плавляется в Ge базу. Затем при охлаждении образуется Si–Ge сплавной эмиттер. (Из работы [9].)

прошло еще несколько десятилетий, прежде чем были созданы работающие Ge–Si НВТ-транзисторы.

4. Лазеры на двойных гетероструктурах

Ни НВТ с переменной шириной запрещенной зоны, ни широкозонный эмиттер в НВТ не выражают в полной мере общий принцип конструирования, состоящий в том, что "в квазиэлектрическом поле можно получать эффекты, которые в принципе невозможны в "реальных" электрических полях". Да, в них были существенные улучшения, но было ли в них что-то такое, что в принципе нельзя получить иным способом?

Пример эффекта, который действительно невозможно было получить другим путем, появился неожиданно в марте 1963 г. Я работал тогда в Varian Associates в Пало-Альто, и мой коллега, д-р Сол Миллер (Sol Miller), очень заинтересовался новыми полупроводниковыми лазерами на р–п-переходах, которые появились в 1962 г. (меня этот вопрос тогда не интересовал). На коллоквиуме по этой теме он сделал блестящий обзор достижений в данной области, не преминув отметить, что такой лазер работает хорошо или при низких температурах, или в импульсном режиме с малой частотой повторения, а чаще, когда выполняется и то, и другое. Когда его спросили, каков шанс получить лазер непрерывного действия, Миллер ответил, что, согласно мнению некоторых экспертов, это в принципе невозможно.

Полезно обсудить здесь эту точку зрения. Рассмотрим свёрхупрощенную схему энергетических зон для

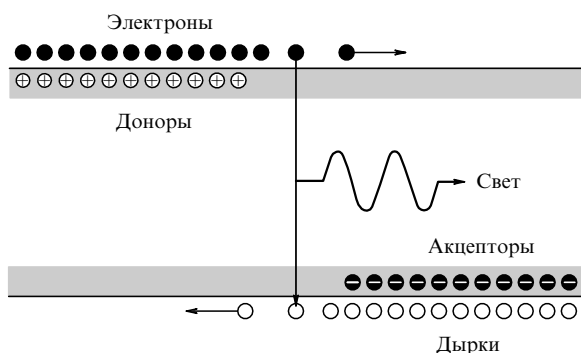


Рис. 5. Схема энергетических зон р–п-гомоперехода при прямом смещении, создающем плоские зоны. При этом вблизи плоскости перехода создается высокая концентрация электронно-дырочных пар, и в результате возникает рекомбинационное излучение.

GaAs р–п-перехода, сильно легированного с обеих сторон и с таким приложенным положительным смещением, что возникают плоские зоны (рис. 5). Тогда электроны диффундируют со стороны полупроводника n-типа в сторону р-типа, а дырки перемещаются в противоположном направлении, создавая некоторую концентрацию электронно-дырочных пар в области границы раздела; их рекомбинация должна давать электромагнитное излучение. Но чтобы возник *лазерный эффект*, нужно получить *инверсную заселенность*, т.е. добиться, чтобы в активной области заселенность нижних состояний зоны проводимости превышала заселенность верхних состояний валентной зоны. *Необходимое* условие для такой инверсии состоит в том, чтобы прямое смещение стало больше ширины запрещенной зоны. Но даже при выполнении этого условия получить инверсную заселенность в обычном р–п-переходе достаточно сложно. Прежде всего, концентрация электронов в активной области всегда будет ниже, чем в легированной n-области. Аналогичное ограничение имеется и для дырок. Таким образом, инверсия требует сильного легирования с обеих сторон, соответствующего вырождению носителей. Но даже при вырождении как электроны, так и дырки диффундируют из активной области непосредственно в соседнюю противоположно легированную область, разрушая тем самым инверсную заселенность. Увеличение прямого смещения не слишком помогает, так как оно увеличивает скорость ухода носителей в той же степени, как и скорость инжекции.

Я немедленно возразил против этого утверждения приблизительно такими словами: "но ведь это чушь, все что нужно сделать — это использовать широкозонный инжектор". Как показано на рис. 6, такое усовершенствование создаст отталкивающее электроны квазиэлектрическое поле с р⁺-стороны и такой же отталкивающий дырки барьер с n⁺-стороны. Таким образом будет достигнуто пространственное ограничение носителей.

При дальнейшем увеличении прямого смещения образуются потенциальные ямы как для электронов, так и для дырок (рис. 7) с квазиэлектрическими силами с *обеих* сторон, втягивающими *и электроны, и дырки* в активную область. В результате концентрации электронов и дырок могут стать намного выше тех, которые устанавливаются уровнями легирования в контактных областях. Возникает инверсная заселенность, необходи-

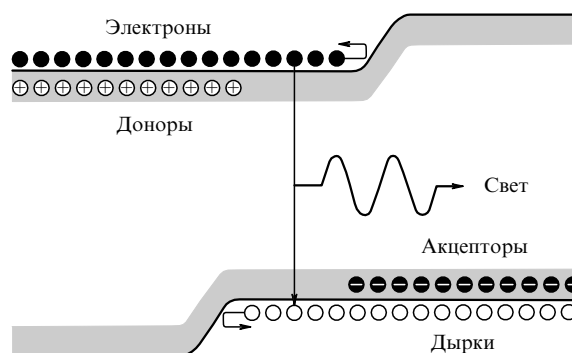


Рис. 6. Пространственное ограничение носителей в двойной гетероструктуре, создаваемое квазиэлектрическими потенциальными барьерами на границах излучающей активной области, которые предотвращают утечку инжектированных электронов и дырок, не влияя на потоки основных носителей из инжекторов.

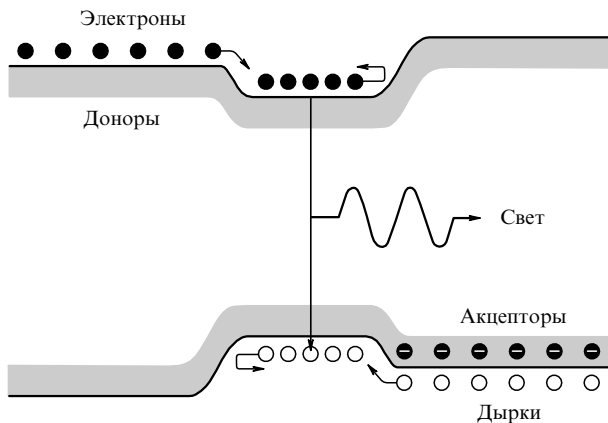


Рис. 7. При дальнейшем увеличении смещения в прямом направлении и для электронов, и для дырок образуются потенциальные ямы, в которых происходит накопление инжектированных носителей вплоть до концентраций вырождения, более высоких, чем концентрации в областях инжекторов.

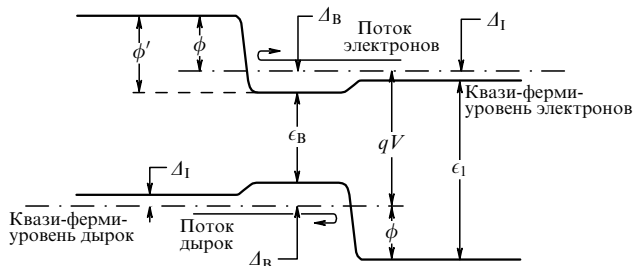


Рис. 8. Схема зон для лазера на двойной гетероструктуре, первоначально опубликованная в работе [11].

мая для лазерного действия. Такой лазер с двойной гетероструктурой (ДН-лазер) является устройством, которое абсолютно невозможно создать в однородных структурах с реальными электрическими полями. Отмечу, что эта идея возникла фактически в тот же самый момент, когда я понял, в чем состоит проблема.

Я написал статью с изложением идеи ДН-лазера и одновременно подал заявку на патент. Статья была представлена в *Applied Physics Letters*, однако была отклонена. Мне посоветовали не бороться с этим решением, а послать статью в *Proceedings of the IEEE*, где она и была опубликована [11], но осталась практически незамеченной. На рисунке 8 приведена та же схема энергетических зон, которая была тогда опубликована.

Патент был зарегистрирован в 1967 г. [12]. Он, пожалуй, написан лучше, чем заметка в *Proc. IEEE*. Срок патента истек в 1985 г.

Повторюсь, в нем содержалась идея, реализовать которую тогда было технически невозможно. Непрерывный режим работы ДН-лазера при комнатной температуре был продемонстрирован впервые Алферовым и др. в 1970 г. [13], а вскоре после этого, также в 1970 г., Хаяши и др. [14]. Исторический обзор экспериментальных работ см. в Нобелевской лекции Алферова [15], а также в [16, 17].

По причинам, которые описаны ниже, сам я не мог быть участником технологической реализации идеи. В

последующие 10 лет я занимался исследованиями эффекта Ганна и вернулся к гетероструктурам только в середине 1970-х.

5. Как не надо оценивать новые технологии

Когда я представил проект с предложением разработать технологию для создания ДН-лазера, мне отказали в его финансировании на том основании, что "этот прибор не может, по-видимому, иметь никакого практического применения", или что-то в этом роде. Сейчас, разумеется, очевидно, насколько ошибочной была эта оценка.

Это был классический случай оценки принципиально новой технологии не с той позиции, какие новые области применения она может породить, а только лишь с точки зрения ее полезности для уже существующих приложений. Это в высшей степени близорукая тактика, однако она стара как мир, стара, как сама техника. ДН-лазер — это просто еще один пример в длинной цепи подобных ошибок. Не будет он и последним. Поэтому я думаю, что стоит сказать здесь несколько слов об оценках такого рода.

Любой внимательный взгляд на историю дает поразительные доказательства моей *Леммы Новых Технологий*, как я ее назвал (по другому поводу) [18]:

Основные области применения любой, достаточно новой инновационной технологии всегда были — и будут — порождены самой этой технологией. Как правило, такие области появляются всегда, хотя обычно не сразу. ДН-лазер — просто хороший современный пример.

Это означает, что следует быть дальновидным при оценке потенциальных возможностей новой технологии. Ее *нельзя* оценивать только по тому, как она может встроиться в уже существующие области приложений, где она едва ли выиграет соревнование с уже внедренными технологиями. Отказ от нее на том основании, что она не имеет известных применений, только затормозит развитие тех прикладных областей, которые могут из нее *вырасти*.

Я не думаю, что можно реалистически предсказать, какие новые приборы и применения возникнут, но полагаю, что можно создать способствующую прогрессу атмосферу, не всегда немедленно спрашивая, для чего эта новая наука может быть полезна (и отказывая в финансировании, если ответ не содержит нереальных, фантастических обещаний). В частности, мы должны объяснить этот исторический факт нашим финансовым ведомствам. Это трудно, но необходимо. Мы должны найти приемлемое решение вопроса о практических приложениях и руководствоваться этим решением. По меньшей мере, поиск области применения должен рассматриваться как часть самого исследования, а не как заранее обещанный результат. Лучше всех по этому поводу высказался Дэвид Мермин в своем недавнем так называемом "стратегическом исследовании" [19]:

"Я жду того дня, когда люди поймут, что открытие совершается не потому, что кто-то захотел его сделать — и сделал."

Что *абсолютно* неприемлемо и от чего мы должны воздерживаться — это попытки мотивировать научное исследование вероятными многообещающими мифическими усовершенствованиями в уже *существующих* при-

кладных областях. Большинство таких обещаний, скорее всего, нереалистично и легко опровергается. Они только вызывают критику, дискредитируя этим всю работу.

В конечном счете, прогресс в прикладной науке не *предопределен*, он имеет *случайный* характер, используя для новых применений все достижения, которые появились в науке и технологии.

6. Важные ограничения

6.1. Согласование решеток

Перейдем теперь к проблеме изготовления гетероструктур. Если два материала с сильно различающимися постоянными решетками выращиваются один на другом, то, независимо от того, является ли переход между ними плавным или нет, при увеличении толщины быстро развиваются большие деформации и возникают дислокации несоответствия — дефекты, не имеющие полезных свойств. Поэтому необходимость согласования решеток становится почти очевидной. Эта проблема оказывается несколько менее острой для изготавливаемых сегодня структур, состоящих из очень тонких слоев (см. ниже), но даже здесь поиск согласованных решеток всегда служит отправной точкой.

Исторически важность согласования параметров решеток была осознана почти с самого начала, особенно для таких биполярных устройств, как лазеры. В моем патенте 1967 г. для ДН-лазера [12] я привел таблицу, в которой были перечислены многие полупроводники в порядке возрастания их параметра решетки (табл. 2). В сопровождающем таблицу тексте пары полупроводников с рассогласованием менее 0,01 Å ($\approx 0,2\%$) были названы самыми многообещающими; это показывает, насколько жесткие требования предъявляются к согласованию решеток. Правда, в тексте патента было ясно сказано, что можно получать согласованные решетки путем сплавления. По иронии судьбы, в 1963 г. литературные данные именно для постоянной решетки AlAs были неверны. Поэтому тогда казалось, что пара GaAs — AlAs не удовлетворяет указанным жестким требованиям, и кроме того, склонность бинарного соединения AlAs к окислению ухудшала дело. Понадобилось некоторое время, чтобы установить перспективность AlAs не

Таблица 2. Фрагмент таблицы 1963 г. [12]. Полупроводники расположены в порядке возрастания параметра решетки (второй столбец). В третьем столбце приведено увеличение параметра решетки по отношению к предшествующему материалу. Следует отметить, что не проводилось никакого различия между элементами IV группы, соединениями III — V и II — VI. Данные 1963 г. для параметра решетки AlAs ошибочны. Правильное значение при комнатной температуре (5,661 Å) в действительности на 0,02 Å больше, чем постоянная решетки GaAs, а при типичных температурах роста кристаллов разница еще меньше. (Здесь приведены полупроводники только до ZnSe; полную таблицу 1963 г. можно найти в [8].)

Полупроводник	a , Å	Δa , Å
ZnS	5,406	
Si	5,428	0,022
GaP	5,450	0,022
AlP	5,46	0,01
AlAs	5,63	0,17
GaAs	5,653	0,02
Ge	5,658	0,005
ZnSe	5,667	0,009
...

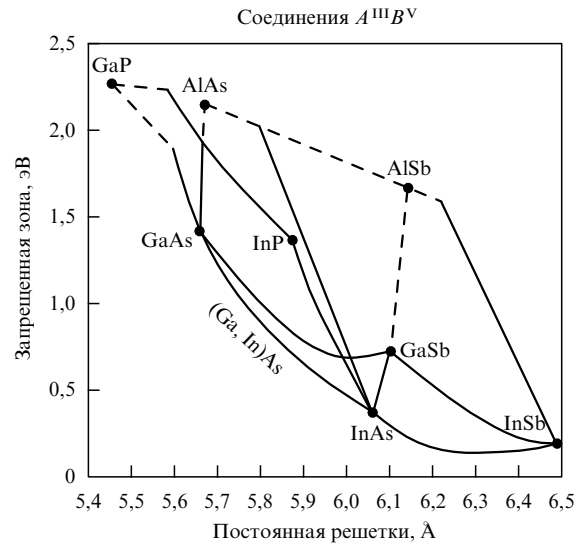


Рис. 9. Фрагмент "Карты Мира", показывающий зависимость ширины запрещенной зоны различных полупроводниковых соединений III — V от параметра решетки. На карте не обозначены "Континенты Старого Света" из IV группы, полупроводники II — VI и "Новый Свет" — нитриды.

как бинарного материала, а в сплаве с GaAs, для которого процессы окисления менее существенны, а рассогласование решеток сводится к пренебрежимо малой величине.

Более наглядным способом представить информацию табл. 2, включающим и ширину запрещенной зоны, является так называемая "Карта Мира", на которой изображены зависимости ширин запрещенных зон перспективных полупроводников от их постоянных решетки (рис. 9) с соединительными линиями, указывающими возможные бинарные твердые растворы.

Основная причина неизменного доминирования сплавов (Al, Ga)As в гетероструктурных исследованиях — это Великая Кристаллографическая Случайность, тот факт, что AlAs и GaAs имеют почти одинаковые параметры решетки. Такое естественное согласование решеток означает, в частности, что имеется идеальная готовая подложка для роста этих гетероструктур, а именно объемный GaAs, существующий в виде высококачественных монокристаллов, в том числе и полуизоляторов, с малой плотностью дислокаций. Если и остался *один* неудобный аспект системы (Al, Ga)As, то это вредное химическое средство алюминия к кислороду, являющееся причиной многих остаточных дефектов. В соответствии с моим предложением 1983 г. [7], твердый раствор (Ga, In)P, изопериодический с GaAs, недавно привлек внимание как альтернатива (Al, Ga)As, в частности, для НВТ, для которых расположение зон на границе раздела (Ga, In)P — GaAs предпочтительнее, чем на границе (Al, Ga)As — GaAs.

Второй естественной подложкой является InP, широко используемый и в оптоэлектронике, и в быстродействующих устройствах, которые менее требовательны к ширине запрещенной зоны, чем приборы на GaAs. Двойных соединений $A^{III}B^V$, изопериодических с InP, не существует, но InP применяется в приборах в комбинации с широким набором твердых растворов от (Ga, In)As до Al(As, Sb).

С появлением квантовых ям, сверхрешеток и других структур, в которых имеются очень тонкие слои, проблема деформаций, индуцированных рассогласованием решеток, несколько утратила свою тираническую сущность. В достаточно тонкослойных структурах очень большие деформации могут сохраняться без появления дислокаций. Более того, модификация зонной структуры гетероструктур путем *целенаправленного* введения деформаций стала сама по себе важным способом конструирования приборов. Недавнее появление успешно работающих Ge–Si HBT-транзисторов стало впечатляющим триумфом этой идеи (см. [20, 21]), да и другие примеры не за горами, такие, как полевые транзисторы (field-effect transistor, FET) или оптические устройства. Новая, недавно разработанная технология получения самоорганизующихся квантовых точек также основана на использовании деформаций непосредственно в процессе роста кристалла.

6.2. Согласование валентностей

Если бы согласование постоянных решетки было единственной проблемой, система Ge–GaAs была бы идеальной гетеросистемой, как фактически считали некоторые, включая и меня, в начале 1960-х. В то время самыми удачными были гетероструктуры Ge на GaAs, исследованные Андерсоном [22]. Им предсказывали прекрасное будущее (термин гетеропереход, по-видимому, появился впервые в статье Андерсона), что нашло свое отражение в табл. 2. В ней объединены данные для соединений III–V, соединений II–VI и полупроводников IV группы, и среди них система GaAs–Ge кажется наиболее привлекательным кандидатом. Понадобилось несколько лет для понимания того, почему это был полный тупик.

Неудача не была следствием химической несовместимости или перекрестного легирования. Ковалентные химические связи между германием, с одной стороны, и Ga и As, с другой, образуются с легкостью, но мне хотелось бы их назвать *валентно-несогласованными* в том смысле, что число поставляемых всеми атомами электронов не равно в точности каноническому "два электрона на ковалентную связь". Как было впервые отмечено в "обязательной для прочтения" статье Харрисона и др. [23], сами связи не являются электронейтральными.

Рассмотрим гипотетическую идеальную (001)-поверхность раздела между Ge и GaAs, относительно которой Ge расположен слева, а GaAs — справа (рис. 10). В GaAs один атом мышьяка дает 5 электронов (= 5/4 электронов на одну связь) и должен быть окружен 4-мя атомами Ga, каждый из которых дает 3 электрона (3/4 на связь). Эти заряды дают в сумме правильное число электронов на одну ковалентную связь в GaAs: $8/4 = 2$. Но если на (001)-границе раздела атомы As имеют в качестве партнеров атомы Ge, то каждый атом Ge дает по 1 электрону на связь, что на 1/2 электрона больше, чем требуется. Иначе говоря, атом As "не знает", является ли он основным атомом в GaAs или донором в Ge.

В результате каждая связь Ge–As ведет себя как донор с дробным зарядом, а каждая связь Ge–Ga — как акцептор с дробным зарядом противоположного знака. Для обеспечения электронейтральности граница раздела Ge–GaAs должна иметь равное число зарядов

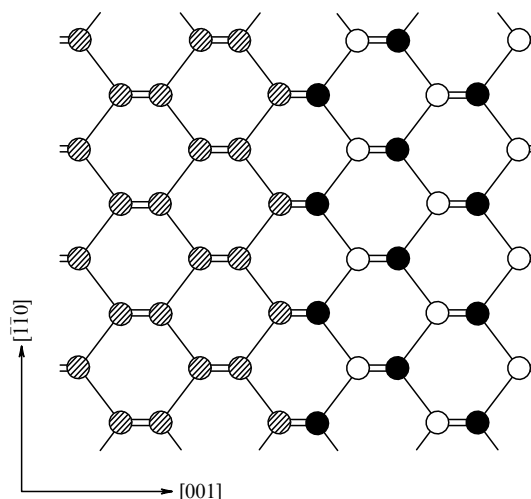


Рис. 10. Отклонение от электронейтральности на "математически плоской" (001)-границе раздела Ge–GaAs. Различные атомы — Ga, As (белые и черные кружки) и Ge (серые кружки) — не дают правильного числа электронов для образования электронейтральных ковалентных связей Ga–Ge или As–Ge, соответствующих 2 электронам на атом. (Из работы [23].)

обоего знака — локально, а не в среднем по большой площади. Однако из-за образования химических связей в процессе эпитаксиального роста такого упорядочения, конечно, не происходит. Если имеется только один тип химических связей, как на рис. 10, заряд границы раздела создает электрическое поле, равное 4×10^7 В см⁻¹. Такое огромное поле должно перестраивать расположение атомов в процессе роста, стремясь уравнять число связей Ge–As и Ge–Ga. Однако эта перестройка никогда не завершается, остаются плохо определяемые флуктуирующие остаточные локальные заряды, приводящие к вредным последствиям для любого приборного применения. Границы с идеальной компенсацией зарядов на связях для любого прибора легко нарисовать на бумаге, но на практике всегда имеется некоторое локальное отклонение от идеальной компенсации зарядов, приводящее к ухудшению рабочих характеристик прибора из-за хаотических флуктуаций потенциала на границе раздела.

Хотя Харрисон и др. [22] обсуждают только границу раздела GaAs–Ge, их аргументация применима и к границам между другими полупроводниковыми материалами из разных групп Периодической системы. В специальном случае сложного полупроводникового соединения, выращенного на границе одноэлементного полупроводника IV группы, на стороне этого соединения возникает проблема антифазных доменов (см., например, [24]).

Вышеприведенная дискуссия относится к наиболее часто используемой границе раздела (001). Заряд на валентно-несогласованной границе зависит, в действительности, от кристаллографической ориентации. Райтом и др. было показано [25, 26], что идеальная граница раздела (112) не имеет ни заряда, ни антифазных доменов. Оказалось, что границы GaP–Si, имеющие достаточно низкую плотность дефектов, можно использовать в качестве эмиттеров в GaP–Si HBT-транзисторах. Однако такие приборы работали не слишком хорошо, и исследования в этом направлении не продолжались.

7. Молекулярно-лучевая эпитаксия и резкие гетерограницы

Лазеры на двойных гетероструктурах, полученные в 1970-е гг., были изготовлены с помощью жидкофазной эпитаксии (liquid-phase epitaxy, LPE), красивой и простой технологии, имеющей, однако, определенные ограничения. Настоящий прорыв в создании гетероструктур произошел только с появлением молекулярно-лучевой эпитаксии (molecular beam epitaxy, MBE) — практической технологии роста, в значительной мере созданной Элом Чо (Al Cho) (позднее появилась газофазная эпитаксия из паров металлоорганических соединений). В отличие от LPE, MBE допускает комбинирование широкого круга полупроводниковых материалов, даже таких гетероэпитаксиальных композиций, как GaP или GaAs на Si. Более того, в MBE оказалось возможным обеспечить высокую степень локального контроля состава соединений практически на уровне одного атомного слоя. Неожиданно выяснилось, что можно экспериментально реализовать практически любую зонную структуру, которую можно нарисовать, по крайней мере, в направлении роста (латеральный контроль с тем же пространственным масштабом остается до сих пор недостижимым). К 1980 г. прогресс в области гетероструктур был настолько велик, что я решился сделать приглашенный доклад с вызывающим названием "Принцип конструирования 1980-х: гетероструктуры на любой вкус?" ("Heterostructures for Everything: Device Principle of the 1980's?") (см. [27]). Это оказалось правильным предвидением.

В частности, появилась возможность растить почти автоматически гетероструктуры с очень резкой границей. Это означало также, что две гетерограницы можно расположить настолько близко друг к другу, что в промежутке между ними квантовые эффекты станут определяющими. Это открыло возможности для создания приборов нового типа. Наиболее очевидный прогресс состоял в использовании квантовых ям (quantum well, QW), особенно для создания лазеров, и в результате лазеры на квантовых ямах вскоре стали самыми распространенными. Возросло также использование гетероструктур и в монополярных приборах, где общий принцип конструирования на основе квазиэлектрических полей нашел применение уже вне первоначальных границ.

Одним из примеров является использование двух туннельных барьеров в резонансных туннельных диодах, которые применяются как источники высокочастотного излучения, вплоть до субтерагерцового частотного диапазона. Еще один пример — это предложение Есаки и Цу использовать периодические гетероструктурные сверхрешетки в качестве объемной среды с отрицательным сопротивлением для генерации в диапазоне еще более высоких частот [28]. Задача до сих пор остается нерешенной, но она является областью очень активных научных исследований, в том числе и моих.

Я бы хотел выделить здесь менее очевидную новую идею модуляционного легирования, выдвинутую Динглом и др. [29]. Рассмотрим гетероструктуру, в которой легирована только одна сторона, с более высоко расположенной зоной проводимости (рис. 11). Идущая вниз ступенька квазиэлектрического потенциала заставляет электроны перетекать в расположенную ниже по энергии зону проводимости с другой стороны от границы раздела

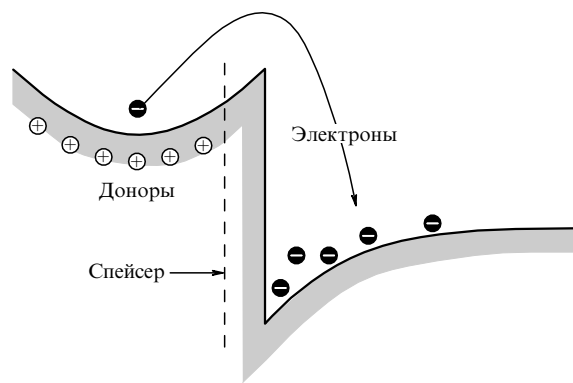


Рис. 11. Модуляционное легирование. При резкой гетерогранице электроны, созданные донорами на высокоэнергетической стороне, перетекают на низкоэнергетическую сторону, где образуется квазидвумерный электронный газ. Поскольку электроны оказываются при этом пространственно отделенными от доноров, рассеяние на примесях уменьшается, особенно если на высокоэнергетической стороне имеется спейсер. Показанное на рисунке искривление зон обусловлено пространственными зарядами, находящимися по обе стороны от границы раздела.

двух материалов. Когда электроны находятся в области квазиэлектрической потенциальной ступеньки, созданной этой резкой гетерограницей, они все еще чувствуют обычное электрическое поле кулоновского притяжения доноров, оставленных позади, на другой стороне перехода. Оно подтягивает электроны к границе раздела, создавая при этом в треугольной квантовой яме двумерный электронный газ (2DEG). Более того, и это самое главное, поскольку электроны пространственно отделились от "своих" доноров, рассеяние на примесях ослабевает, а подвижность электронов растет. Чтобы максимально использовать эти преимущества, вблизи от границы оставляют нелегированную область — спейсер.

Эта идея имеет необычайно далеко идущие последствия как для приборных применений, так и для фундаментальной физики твердого тела. В приборостроении она составляет основу полевых транзисторов (FET) нового типа, обычно называемых транзисторами с высокой подвижностью электронов (high-electron-mobility transistor, HEMT) [30, 31]. По качеству эти транзисторы лучше FET-транзисторов более раннего поколения. Из-за низких шумов их используют теперь в качестве чувствительных входных каскадов сотовых телефонов. Таким образом, они внесли вклад в быстрый рост этого направления современных информационных технологий.

В фундаментальной физике подавление примесного рассеяния модуляционным легированием с оптимально выбранными спейсерами позволило получить огромные подвижности при низких температурах. Это прямой путь от идеи модуляционного легирования к открытию квантового эффекта Холла, сделанному Цуи, Штёрмером и Госсардом [32, 33] в образцах 2DEG высокого структурного совершенства, выращенных Госсардом. Последующая теоретическая интерпретация этого эффекта Лафлином [34] показала, что это действительно прорыв в физике твердого тела. За эти работы Цуи, Штёрмер и Лафлин получили Нобелевскую премию по физике 1998 г. К сожалению, правила присуждения Нобелевской премии запрещают давать ее более, чем трем людям. Поэтому Госсард был исключен из этого коллектива.

8. Разрывы зон

После достижений в МВЕ-технологии в начале 1970-х я вернулся к моим исследованиям гетероструктур, в частности, к проблеме разрывов зон в резких гетеропереходах. В пределе такого резкого перехода в энергетической зонной структуре возникает разрыв, и главным, как с экспериментальной, так и с теоретической точки зрения становится вопрос о взаимном расположении зон по обе стороны от границы. Одна из причин, по которой на всех моих прежних схемах зонных структур изображен плавный переход, состоит в том, чтобы проследить шаг за шагом, как выстраиваются зоны; при этом суть проблемы была для меня вполне ясна.

8.1. Типы разрывов зон

В случае двух полупроводников имеются три возможных варианта взаимного расположения зон (рис. 12).

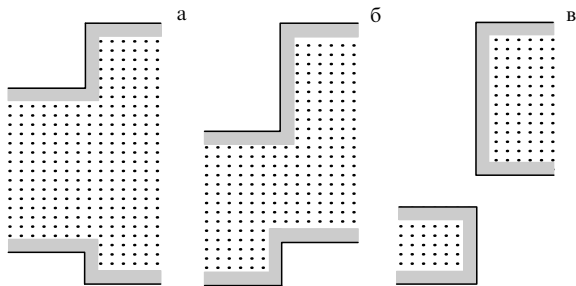


Рис. 12. Раздвижка зон (а), ступенчатое расположение зон (б) и разрыв энергетической щели (в).

8.1.1. Раздвинутые зоны. Это наиболее частый случай взаимного расположения зон, при котором разрывы зон проводимости и валентной зоны имеют разный знак. По сути дела, это предельный случай перехода от структуры с плавно меняющейся запрещенной зоной (рис. 1в) к структуре с резкой гетерограницей. В квантовых ямах и сверхрешетках, изготовленных из таких пар полупроводников, нижние электронные состояния зоны проводимости находятся с той же стороны от границы, что и верхние состояния валентной зоны. Именно это делает пару полупроводников с таким взаимным расположением зон наиболее интересной для оптоэлектронных приборов, например, лазеров, которые являются биполярными устройствами, поскольку в их работу вовлечены носители обоих типов — как электроны, так и дырки. В этом случае оба типа носителей находятся в *одном и том же* слое, поэтому такие структуры называют иногда *пространственно прямыми*. Подобное расположение зон типично для множества современных оптоэлектронных приборов, таких, например, как лазеры на квантовых ямах. К этому типу структур относятся наиболее широко исследованные системы с гетероструктурами GaAs – (Al, Ga)As. Имеется еще целый ряд систем, обладающих подобными свойствами, например, (Ga, In)As, согласованный по параметру решетки с InP, и (Ga, In)P, согласованный с GaAs.

8.1.2. Ступенчатое расположение зон. Для некоторых пар материалов и зона проводимости, и валентная зона сдвигаются в одном направлении. При этом возникает

такая зонная структура, в которой дно зоны проводимости расположено с одной стороны от перехода, а потолок валентной зоны — с другой. Энергетический зазор между ними при этом меньше, чем наименьшая из двух запрещенных зон в исходных материалах. Такими парами являются, например, AlAs – Al_xGa_{1-x}As при $x > 0,3$ или (Al, In)As, согласованный по параметру решетки с InP; имеются и другие примеры. В биполярных структурах с таким расположением зон электроны и дырки локализованы в *различных* слоях, поэтому такие структуры называются *пространственно непрямыми*. Несмотря на это, волновые функции перекрываются на границе раздела, и становится возможной излучательная рекомбинация, при которой энергия испущенных фотонов меньше, чем ширина более узкой из двух запрещенных зон [35, 36].

Ступенчатое расположение означает, что больший разрыв зон имеется либо в зоне проводимости, либо в валентной зоне. Для некоторых приложений это свойство даже важнее, чем пространственное разделение носителей. Например, сдвиг вверх зоны проводимости на границе раздела InAs – AlSb равен 1,35 эВ [37]; это самый большой сдвиг для соединений III – V. Некоторые приложения основаны на использовании этого свойства и малой величины эффективной массы в InAs. Самый быстродействующий из известных резонансный туннельный диод с частотой 712 ГГц создан на основе этой системы [38].

Высокие потенциальные барьеры обеспечивают сильное пространственное ограничение электронов в FET, что позволяет достигать чрезвычайно высоких концентраций электронов (вплоть до 10^{13} см⁻²) с помощью модуляционного легирования (т.е. путем введения доноров не в потенциальную яму, а в барьер) при сохранении высокой подвижности. Такие свойства делают гетеропереход InAs – AlAs идеальной системой для исследования свойств квантовых ям в пределе металлической проводимости, что интересно, например, для нового класса сверхпроводящих приборов, основанных на слабой связи [39].

8.1.3. Разрыв запрещенной зоны. Если ступенчатое расположение зон довести до логического конца, то возникнет разрыв запрещенной зоны, когда дно зоны проводимости полупроводника с одной стороны от перехода окажется ниже потолка валентной зоны полупроводника с другой стороны. Имеется по крайней мере одна почти изопериодическая пара такого типа — InAs – GaSb с разрывом запрещенной зоны на границе раздела порядка 150 мэВ [40].

Расположение зон с разрывом энергетической щели в InAs – GaSb само по себе является весьма экзотическим и представляет для физиков особый интерес. С точки зрения теоретика, занимающегося объяснением разрывов зон, возможность предсказывать величину разрыва, хотя бы приблизительно, является лакмусовой бумажкой для проверки любой модели, описывающей взаимное расположение зон. Современные теории проходят эту проверку, как говорится, "на ура".

8.2. Теория

Из сказанного должно быть ясно, что вопрос о точных значениях разрывов зон для различных, представляющих интерес пар полупроводников является централь-

ным как с теоретической, так и с экспериментальной точки зрения. Я старался внести вклад и в теорию, и в эксперимент.

В конце 1960-х единственная возможность для теоретической оценки разрывов зон состояла в использовании *правила электронного сродства* [22]. Согласно этому правилу, разрыв зоны проводимости должен равняться разности электронного сродства для двух свободных поверхностей полупроводников. В статье 1975 г. [41] я показал, что это очень ненадежное правило. Даже если бы имелись надежные данные по электронному сродству, применимость этого правила все равно зависела бы от неявных приближений, сделанных при описании границы раздела двух полупроводников и гораздо более резких границ раздела полупроводник–вакуум. Такие приближения почти наверняка были бы неприменимы. Харрисон прекрасно охарактеризовал это правило, сказав, что "оно заменяет одну простую задачу двумя очень трудными проблемами" [42].

Мне нужна была такая теория, которая определяла бы разрывы зон из *объемных* свойств граничащих полупроводников. Я предложил эту задачу в качестве темы диссертации Биллу Френсли (теперь он работает в Техасском университете в Далласе). В частности, я попросил Билла разобраться в том, могут ли в действительности возникать разрывы запрещенных зон. Разработанная нами теория [43, 44] была основана на методе псевдопотенциала. Она не только впервые дала возможность, исходя из объемных свойств, сделать полуколичественные оценки уже известных разрывов зон, например, в GaAs–AlAs, но и обладала большой предсказательной силой. В частности, с помощью этой теории было предсказано, что гетероструктура InAs–GaSb имеет либо разрыв запрещенной зоны, либо очень близкое к этому расположение зон.

После теории Френсли–Крёмера появились другие работы, в которых предлагались иные теоретические подходы (см. [42, 45]).

8.3. Определение разрывов зон по зависимости емкости от напряжения

Однажды, в 1979 г. Джим Харрис (Jim Harris, тогда он работал в Rockwell Science Center, а теперь — в Стэнфордском университете) показал мне данные по зависимости емкости от напряжения (C – V -данные) для гетероструктуры (Al, Ga)As–GaAs, выращенной методом LPE. C – V -измерения — это обычная техника определения концентрации электронов по измерению емкости барьера Шоттки, изготовленного на поверхности полупроводника, при обратном смещении. Изменяя напряжение, можно исследовать распределение электронов по глубине. Данные Харриса, относящиеся к гетерогранице, ясно указывали, что имеется обогащенный электронами слой со стороны GaAs и обедненный слой со стороны (Al, Ga)As, как и можно было ожидать из соответствующей зонной структуры. Однако наблюдаемая концентрация электронов оказалась сильно размытой из-за усреднения по дебаевской длине. Когда я попытался описать процесс усреднения количественно, мне стало ясно, что дипольный момент, связанный с комбинацией обогащенного и обедненного слоев, при усреднении должен сохраняться, и измеряя его, можно определить разрыв зоны проводимости [46–48]. С помощью этого подхода были получены разрывы зон,

составляющие приблизительно 66 % от разности ширины запрещенных зон, что близко к принятому сегодня значению 62 %.

Описанная C – V -методика затем использовалась многими другими авторами. Она дала наиболее надежные данные по разрывам зон для многих гетероструктурных пар полупроводников.

9. Эпилог

В этой лекции я сделал обзор своих работ по гетероструктурам, в основном ранних — до 1963 г., в которых были сформулированы основные принципы биполярных устройств. Однако, чтобы достичь современного уровня знаний о гетероструктурах, потребовалась дальнейшая работа многих других исследователей — как в области технологии, так и в теории — главным образом, по небиполярным структурам. Только благодаря их достижениям в областях, выходящих за рамки моей деятельности, со временем выяснилось значение сделанного мной. За это я выражаю им всем мою благодарность.

Перевела с английского *И.П. Ипатова*
Научная редакция перевода *Р.А. Сурица*

Список литературы

1. Kroemer H *RCA Rev.* **18** 332 (1957); reprinted from the *Proc. of the Symp. "The Role of Solid State Phenomena in Electric Circuits"*, Polytechnic Institute of Brooklyn, April 1957, p. 143
2. Shockley W, US Patent 2, 569, 347 (filed 26 June 1948) (September 25, 1951)
3. Krömer H *Naturwissenschaften* **40** 578 (1953)
4. Krömer H *Arch. Elekt. Übertrag.* **8** 499 (1954)
5. Kroemer H *Proc. IRE* **45** 1535 (1957)
6. Kroemer H *Proc. IEEE* **70** 13 (1982)
7. Kroemer H *J. Vac. Sci. Technol. B* **1** 126 (1983)
8. Kroemer H *Phys. Scripta* **T68** 10 (1996)
9. Kroemer H, unpublished (1957)
10. Diedrich H, Jötten K, in *Communications Présentées Colloque Intern. sur les Dispositifs à Semiconducteurs, Paris, 20–25 Février 1961* (Paris: Éditions Chiron, 1961) p. 330
11. Kroemer H *Proc. IEEE* **51** 1782 (1963)
12. Kroemer H, US Patent 3, 309, 553 (filed Aug. 16, 1963) (March 14, 1967)
13. Алферов Ж И, Андреев В М, Гарбузов Д З, Жиляев Ю В, Морозов Е П, Портной Е Л, Трофим В Г *ФТП* **4** 1826 (1970) [Alferov Zh I, Andreev V M, Garbuzov D Z, Zhilyaev Yu V, Morozov E P, Portnoi E L, Trofim V G *Sov. Phys. Semicond.* **4** 1573 (1971)]
14. Hayashi I, Panish M B, Foyt P W, Sumski S *Appl. Phys. Lett.* **17** 109 (1970)
15. Alferov Z I, in *Les Prix Nobel: The Nobel Prizes 2000* (Ed. T Frängsmyr) (Stockholm, Sweden: Almqvist & Wiksell Intern., 2001) p. 65 [Алферов Ж И *УФН* **172** 1068 (2002)]
16. Alferov Z I *Phys. Scripta* **T68** 32 (1996)
17. Casey H C (Jr), Panish M B *Heterostructure Lasers — Part A: Fundamental Principles* (New York: Academic Press, 1978) Sec. 1.2
18. Kroemer H, in *Proc. of NATO Adv. Res. Workshop on Future Trends in Microelectronics, Bendor, France, 1995* (Eds S Luryi, J Xu, A Zaslavsky, NATO ASI Series, Ser. E, Vol. 323) (Dordrecht: Kluwer Acad. Publ., 1996) p. 1
19. Mermin D *Phys. Today* **52** (8) 11 (1999)
20. Abstreiter G *Phys. Scripta* **T68** 68 (1996)
21. König U *Phys. Scripta* **T68** 90 (1996)
22. Anderson R L *IBM J. Res. Dev.* **4** 283 (1960)
23. Harrison W A, Kraut E A, Waldrop J R, Grant R W *Phys. Rev. B* **18** 4402 (1978)
24. Kroemer H *J. Cryst. Growth* **81** 193 (1987)
25. Wright S L, Inada M, Kroemer H *J. Vac. Sci. Technol.* **21** 534 (1982)

26. Wright S L, Kroemer H, Inada M *J. Appl. Phys.* **55** 2916 (1984)
27. Kroemer H *Jpn. J. Appl. Phys. Suppl.* **20-1** 9 (1981)
28. Esaki L, Tsu R *IBM J. Res. Dev.* **14** 61 (1970)
29. Dingle R, Störmer H L, Gossard A C, Wiegmann W *Appl. Phys. Lett.* **33** 665 (1978)
30. Mimura T, Hiyamizu S, Fujii T, Nanbu K A *Jpn. J. Appl. Phys.* **19** L225 (1980)
31. Delagebeaudeuf D, Delescluse P, Etienne P, Laviron M, Chaplart J, Linh N T *Electron. Lett.* **16** 667 (1980)
32. Tsui D C, Stormer H L, Gossard A C *Phys. Rev. Lett.* **48** 1559 (1982)
33. Stormer H L *Rev. Mod. Phys.* **71** 875 (1999)
34. Laughlin R B *Rev. Mod. Phys.* **71** 863 (1999)
35. Kroemer H, Griffiths G *IEEE Electron. Dev. Lett.* **EDL-4** 20 (1983)
36. Caine E J, Subbanna S, Kroemer H, Merz J L, Cho A Y *Appl. Phys. Lett.* **45** 1123 (1984)
37. Nakagawa A, Kroemer H, English J H *Appl. Phys. Lett.* **54** 1893 (1989)
38. Brown E R, Söderström J R, Parker C D, Mahoney L J, Molvar K M, McGill T C *Appl. Phys. Lett.* **58** 2291 (1991)
39. Kroemer H, Nguyen C, Hu E L, Yuh E L, Thomas M, Wong K C *Physica B* **203** 298 (1994)
40. Sakaki H, Chang L L, Ludeke R, Chang C-A, Sai-Halasz G A, Esaki L *Appl. Phys. Lett.* **31** 211 (1977)
41. Kroemer H *Crit. Rev. Solid State Sci.* **5** 555 (1975)
42. Harrison W A *J. Vac. Sci. Technol.* **14** 1016 (1977)
43. Frensley W R, Kroemer H *J. Vac. Sci. Technol.* **13** 810 (1976)
44. Frensley W R, Kroemer H *Phys. Rev. B* **16** 2642 (1977)
45. Christensen N E *Phys. Rev. B* **38** 12687 (1988)
46. Kroemer H, Chien W-Y, Harris J S (Jr), Edwall D D *Appl. Phys. Lett.* **36** 295 (1980)
47. Kroemer H, Chien W-Y *Solid State Electron.* **24** 655 (1981)
48. Kroemer H *Appl. Phys. Lett.* **46** 504 (1985)