

PACS: 84.32.Ff, 85.35.-p

Ю.А.Кругляк, д.х.н.

Одесский государственный экологический университет

Н.Е.Кругляк, к.х.н.

Одесский национальный университет им. И.И.Мечникова

УРОКИ НАНОЭЛЕКТРОНИКИ. 1. ПРИЧИНЫ ВОЗНИКНОВЕНИЯ ТОКА В КОНЦЕПЦИИ «СНИЗУ – ВВЕРХ»

В рамках концепции «снизу – вверх» современной теоретической и прикладной наноэлектроники рассматриваются причины возникновения тока и роль электрохимических потенциалов и фермиевских функций в этом процессе.

Ключевые слова: наноэлектроника, молекулярная электроника, снизу-вверх, электрический ток, электрохимический потенциал, функция Ферми.

Введение. Бурное развитие наноэлектроники в последние 10 – 15 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 50 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 законов квантовой механики и неравновесной статистической термодинамики, что неизбежно приведет к пересмотру физического образования уже на начальных университетских курсах.

Согласно закону Ома, сопротивление R и проводимость G проводника длины L с площадью поперечного сечения A даются выражениями

$$R \equiv V/I = \rho L/A \text{ и } 1/R \equiv G = \sigma A/L, \quad (1)$$

где удельное сопротивление ρ и обратная ему удельная проводимость σ 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 $1 \mu\text{m}$ 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 ~ 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 [5].

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 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» [6 – 8], 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 [9 – 11] 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» [12] привело к созданию единой картины транспортных явлений в электронных устройствах как наноразмерных, так и макроразмерных.

В настоящей работе в рамках концепции «снизу – вверх» рассматриваются причины возникновения тока и роль электрохимических потенциалов и фермиевских функций в этом процессе. В следующем сообщении рассматривается модель «упругого резистора» и дается новая формулировка закона Ома. В рамках концепции «снизу – вверх» будут также рассмотрены общие вопросы электронной проводимости, в том числе на примере графена.

Причина возникновения тока. На вопрос о причине возникновения тока при приложении разности потенциалов к концам проводника обычно ссылаются на связь плотности тока \mathbf{j} с внешним приложенным электрическим полем \mathbf{E}

$$\mathbf{j} = \sigma \mathbf{E}, \quad (2)$$

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 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 на единицу энергии $D(E)$, занятых электронами и свободных и электрохимического потенциала μ_0 (рис. 1). Простоты ради, что никак не скажется на окончательных выводах, будем пользоваться точечной моделью

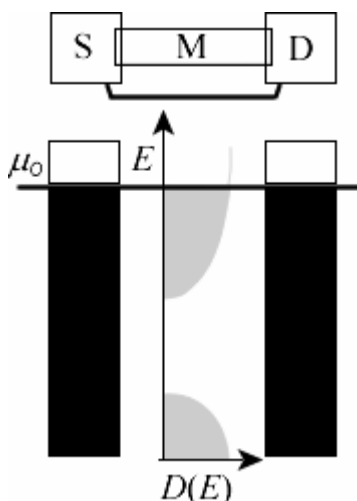


Рис. 1 – Первым шагом в объяснении работы любого электронного устройства должно быть задание плотности состояний $D(E)$ как функции энергии E внутри проводника M и определение равновесного значения электрохимического потенциала μ_0 , отделяющего заполненные электронами состояния от пустых состояний.

проводника (канала переноса электронов), 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 μ_0 везде одинаков и все состояния с $E < \mu_0$ заполнены электронами, а состояния с $E > \mu_0$ пустые (рис. 1).

При включении в цепь источника напряжения (рис. 2) 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

$$\mu_1 - \mu_2 = qV. \quad (3)$$

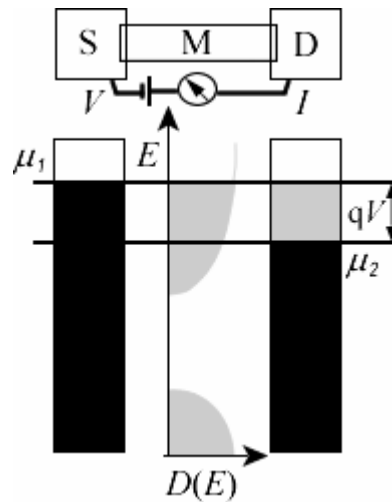


Рис. 2 – При подаче напряжения V на клеммы проводника потенциал анода D понижается на величину qV , создавая разность электрохимических потенциалов на концах проводника $\mu_1 - \mu_2 = qV$.

Точно так же как разность температур приводит к потоку тепла, а различие в уровнях жидкости ведет к ее перетoku, так и разность электрохимических потенциалов является причиной возникновения тока. Только состояния проводника в окне $\mu_1 - \mu_2$ и находящиеся достаточно близко к значениям μ_1 и μ_2 дают вклад в поток электронов, тогда как все состояния значительно выше μ_1 и ниже μ_2 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 μ_1 , и опорожнения состояний канала с энергией, большей потенциала μ_2 . Рассмотрим токовый канал с состояниями, энергия которых меньше μ_1 , но больше μ_2 . Контакт 1 стремится заполнить эти состояния, поскольку их энергия меньше μ_1 , а контакт 2 стремится опорожнить эти состояния, поскольку их энергия больше μ_2 , что и приводит к непрерывному движению электронов от контакта 1 к контакту 2.

Рассмотрим теперь состояния канала с энергией, большей μ_1 и μ_2 . 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 μ_1 и μ_2 . 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.

Роль фермиевских функций. Итак, утверждалось, что в равновесии все состояния с энергией $E < \mu_0$ заполнены электронами, а состояния с энергией $E > \mu_0$ пустые. 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 $E = \mu_0$, где k – постоянная Больцмана, T – абсолютная температура. Математически этот переход описывается функцией Ферми

$$f(E) = \frac{1}{\exp\left(\frac{E - \mu}{kT}\right) + 1}. \quad (4)$$

График функции Ферми показан на рис. 3 слева, возможно, в несколько

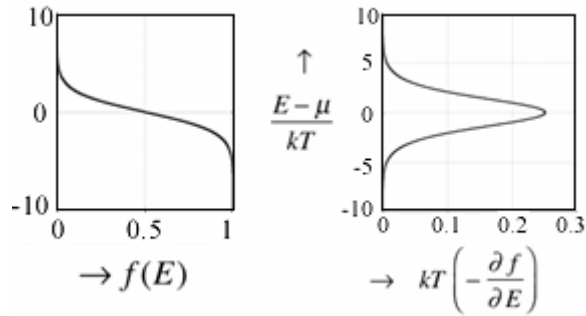


Рис. 3 – Графики функции Ферми и нормированной функции теплового уширения.

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 $D(E)$. 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 пусты ($f = 0$), а переход от $f = 1$ к $f = 0$ происходит в узком интервале энергии $\sim \pm 2 kT$, охватывающем значение $E = \mu_0$.

Действительно, на рис. 3 показана производная от функции Ферми, помноженная на kT с тем, чтобы сделать ее безразмерной

$$F_T(E, \mu) = kT \left(-\frac{\partial f}{\partial E} \right). \quad (5)$$

Подставляя выражение для функции (4), сразу видно, что

$$F_T(E, \mu) = \frac{e^x}{(e^x + 1)^2}, \quad (6)$$

где $x \equiv (E - \mu) / kT$. Из (6) сразу видно, что

$$F_T(E, \mu) = F_T(E - \mu) = F_T(\mu - E), \quad (7)$$

а из ур-й (6) и (4) видно, что

$$F_T = f(1 - f). \quad (8)$$

Интегрирование функции (8) во всем интервале изменения энергии дает площадь, равную kT

$$\int_{-\infty}^{+\infty} dE F_T(E, \mu) = kT \int_{-\infty}^{+\infty} dE \left(-\frac{\partial f}{\partial E} \right) = kT [-f]_{-\infty}^{+\infty} = kT(1 - 0) = kT. \quad (9)$$

Так что функцию F_T можно приблизительно представить себе в виде прямоугольного «импульса», центрированного относительно значения $E = \mu$ с высотой, равной $1/4$ и шириной $\sim 4 kT$.

Выход из равновесия. Когда рассматриваемая система (рис. 1) находится в равновесии, 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. 2) 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. 4)

$$f_1(E) = \frac{1}{\exp\left(\frac{E - \mu_1}{kT}\right) + 1} \quad (10)$$

и

$$f_2(E) = \frac{1}{\exp\left(\frac{E - \mu_2}{kT}\right) + 1}. \quad (11)$$

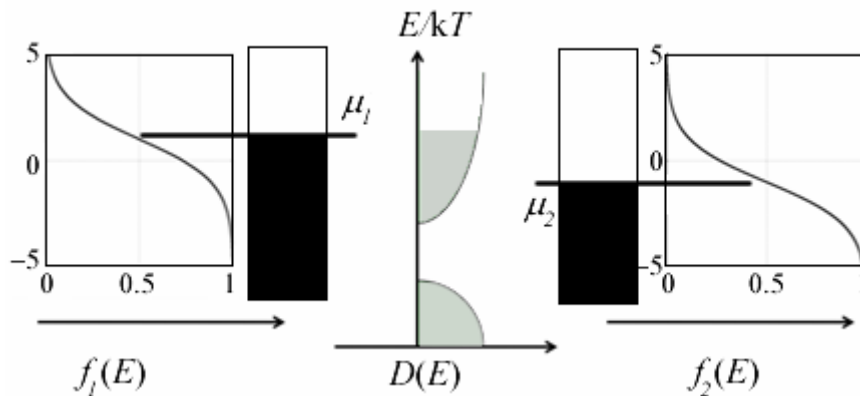


Рис. 4 – При выходе из равновесия электроны в контактах занимают доступные им состояния в соответствии с фермиевскими распределениями и значениями электрохимических потенциалов.

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 макроразмерных. Однако, для наноразмерных проводников ток при любых значениях энергии электронных состояний в проводнике пропорционален разности $I(E) \sim f_1(E) - f_2(E)$ фермиевских распределений в обоих контактах. Эта разность зануляется, если энергия E больше μ_1 и μ_2 , поскольку в этом случае обе функции Ферми равны нулю. Эта разность также зануляется, если энергия E меньше μ_1 и μ_2 , поскольку в этом случае обе фермиевские функции равны единице. Ток возникает лишь в окне $\mu_1 - \mu_2$, если в этом окне есть хотя бы одно электронное состояние проводника.

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

Построим функцию разности двух фермиевских функций, нормированную к приложенному напряжению

$$F(E) = \frac{f_1(E) - f_2(E)}{qV/kT}, \quad (12)$$

где

$$\begin{aligned} \mu_1 &= \mu_0 + (qV/2) \\ \mu_2 &= \mu_0 - (qV/2) \end{aligned} \quad (13)$$

Функция разности $F(E)$ сужается по мере того, как напряжение V , помноженное на заряд электрона, становится все меньше величины kT (рис. 5). Отметим также, что по мере

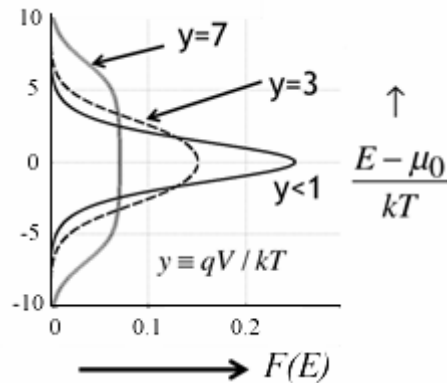


Рис. 5 – График функции разности $F(E)$ в зависимости от значения $(E - \mu_0)/kT$ для различных значений $qV/kT \equiv y$.

того, как kT начинает превосходить энергию qV , функция $F(E)$ все больше приближается к функции теплового уширения (5)

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

так что из ур-я (12) следует, что

$$f_1(E) - f_2(E) = \frac{qV}{kT} F_T(E, \mu_0) = \left(-\frac{\partial f_0}{\partial E} \right) qV, \quad (14)$$

если приложенное напряжение, помноженное на заряд электрона, $\mu_1 - \mu_2 = qV$ становится намного меньшим kT .

Нам потребуется также следующее выражение

$$f(E) - f_0(E) = -\frac{\partial f_0}{\partial E} (\mu - \mu_0), \quad (15)$$

которое, как и ур-е (14), можно получить следующим образом.

Для функции Ферми

$$f(x) \equiv \frac{1}{e^x + 1}, \quad x \equiv \frac{E - \mu}{kT} \quad (16)$$

имеем

$$\frac{\partial f}{\partial E} = \frac{df}{dx} \frac{\partial x}{\partial E} = \frac{df}{dx} \frac{1}{kT}$$

$$\frac{\partial f}{\partial \mu} = \frac{df}{dx} \frac{\partial x}{\partial \mu} = -\frac{df}{dx} \frac{1}{kT} \quad , \quad (17)$$

$$\frac{\partial f}{\partial T} = \frac{df}{dx} \frac{\partial x}{\partial T} = -\frac{df}{dx} \frac{E - \mu}{kT^2}$$

откуда

$$\frac{\partial f}{\partial \mu} = -\frac{\partial f}{\partial E} \quad . \quad (18)$$

$$\frac{\partial f}{\partial T} = -\frac{E - \mu}{T} \frac{\partial f}{\partial E}$$

Уравнение (15) получается из разложения функции Ферми вблизи точки равновесия в ряд Тейлора

$$f(E, \mu) \cong f(E, \mu_0) + \left(\frac{\partial f}{\partial \mu} \right)_{\mu=\mu_0} (\mu - \mu_0). \quad (19)$$

Из уравнения (18) следует

$$\left(\frac{\partial f}{\partial \mu} \right)_{\mu=\mu_0} = \left(-\frac{\partial f}{\partial E} \right)_{\mu=\mu_0} \quad . \quad (20)$$

Пусть $f(E)$ соответствует $f(E, \mu)$, а $f_0(E)$ соответствует $f(E, \mu_0)$, тогда

$$f(E) \approx f_0(E) + (\mu - \mu_0) \left(-\frac{\partial f}{\partial E} \right), \quad (21)$$

что после перегруппировки дает искомое ур-е (15), которое справедливо при $\mu - \mu_0 \ll kT$.

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 нескольких kT , охватывающих равновесный электрохимический потенциал μ_0 , определяемый функцией F_T (ф-ла 5,

рис. 3), 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 μ_0 , что в корне отличает одно вещество от другого.

Настоящий ответ нельзя назвать новым, и он хорошо известен специалистам в области нанoeлектроники. Тем не менее обсуждение проводимости и поньше обычно начинается с теории Друде [13], которая сыграла важнейшую историческую роль в понимании природы 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 нанoeлектроникой, показывают, что ток порождается «подготовкой» двух контактов

$$f_1(E) - f_2(E),$$

и эта разница не нулевая только в окне вокруг равновесного электрохимического потенциала μ_0 . Проводимость канала высокая или низкая зависит только от наличия электронных состояний в этом окне. К этому выводу обычно приходят на основе транспортного уравнения Больцмана [14] или формализма Кубо [15], тогда как предлагаемая концепция «снизу – вверх» сразу дает корректную картину возникновения тока.

Настоящая работа явилась результатом посещения одним из нас (ЮАК) курса лекций «Fundamentals of Nanoelectronics, Part I: Basic Concepts», прочитанного он-лайн в январе – феврале 2012 года проф. С.Датта (Supriyo Datta) в рамках инициативы Purdue University / nanoHUB-U [www.nanohub.org/u].

Список литературы

1. *Mitin Vladimir V., Kochelap Viatcheslav A., Stroscio Michael A.* Introduction to Nanoelectronics: Science, Nanotechnology, Engineering, and Applications. – Cambridge: Cambridge University Press. – 2012. – pp. 346.
2. *Hoefflinger Bernd (Editor).* Chips 2020: A Guide to the Future of Nanoelectronics (Frontiers Collection). – Berlin: Springer-Verlag. – 2012. – pp. 505.
3. *Мартинес-Дуарт Дж.М., Мартин-Палма Р.Дж., Агулло-Рueda Ф.* Нанотехнологии для микро- и оптоэлектроники. – Москва: Техносфера. – 2007. – 368 с.
4. *Драгунов В. П., Неизвестный И. Г., Гридчин В. А.* Основы нанoeлектроники. – Москва: Логос. – 2006. – 496 с.
5. *Smit R.H.M., Noat Y., Untiedt C., Lang N.D., van Hemert M.C., van Ruitenbeek J.M.* Measurment of the conductance of a hydrogen molecule // Nature. – 2002. – V. 419, N 3. – P. 906 – 909.

6. *Datta Supriyo*. Electronic Transport in Mesoscopic Systems.- Cambridge: Cambridge University Press. – 2001. – pp. 377.
7. *Datta Supriyo*. Quantum Transport: Atom to Transistor. – Cambridge: Cambridge University Press. – 2005. – pp. 404.
8. www.nanohub.org/topics/ElectronicsFromTheBottomUp
9. *Landauer Rolf*. Spatial variation of currents and fields due to localized scatterers in metallic conduction // IBM J. Res. Dev. – 1957. – V. 1, N 3. – P. 223 – 231.
10. *Landauer Rolf*. Electrical resistance of disordered onedimensional lattices // Philos. Mag. – 1970. – V. 21. – P. 863 – 867.
11. *Landauer Rolf*. Spatial variation of currents and fields due to localized scatterers in metallic conduction // J. Math. Phys. – 1996. – V. 37, N 10. – P. 5259.
12. *Datta Supriyo*. Lessons from Nanoelectronics: A New Perspective on Transport. – Hackensack, New Jersey: World Scientific Publishing Company. – 2012. – pp. 340.
13. *Ашкрофт Н., Мермин Н.* Физика твердого тела, том 1. – М: Мир. – 1979. – 400 с.
14. *Sears F.W., Salinger G.L.* Thermodynamics, Kinetic Theory, and Statistical Thermodynamics. – Boston: Addison-Wesley. – 1975. – pp. 331 – 336, 355 – 361.
15. *Kubo R.* Statistical-Mechanical Theory of Irreversible Processes.I. General Theory and Simple Applications to Magnetic and Conduction Problems // J.Phys.Soc. Japan. – 1957. – V. 12. – P. 570 – 586.

Уроки наноелектроніки. 1. Причини виникнення струму в концепції «знизу – нагору»

Кругляк Ю.О., Кругляк Н.Ю.

В рамках концепції «знизу – нагору» сучасної теоретичної і прикладної наноелектроніки розглядаються причини виникнення току та роль електрохімічних потенціалів і фермієвських функцій в цьому процесі.

Ключові слова: *наноелектроніка, молекулярна електроніка, знизу – нагору, електричний струм, електрохімічний потенціал, функція Фермі.*

Lessons of nanoelectronics. 1. Current generation by bottom – up approach

Kruglyak Yu.A., Kruglyak N.E.

Current generation with the use of electrochemical potentials and Fermi functions is discussed in the frame of the bottom – up approach of modern theoretical and applied nanoelectronics.

Keywords: *nanoelectronics, molecular electronics, bottom – up, electric current, electrochemical potential, Fermi function.*

*Надійшла до редакції 27.02.2012
Прийнята до публікації 27.06.2012*