

6. МОДЕЛИРОВАНИЕ ТЕХНОЛОГИИ И ТРАНЗИСТОРОВ ИС

6.1. Общие положения

Транзистор – основной элемент современных цифровых ИС. Характерным примером схемных решений, применяемых в таких микросхемах, является логический элемент 2И-НЕ, рассмотренный в главе 1. Отметим, что эта схема содержит только транзисторы. Поэтому такие важнейшие характеристики современных ИС, как быстродействие (тактовая частота), потребляемая мощность и некоторые другие в сильной степени зависят от электрических параметров транзистора. Этим определяется важность этапа проектирования транзистора ИС.

Опытный специалист, занимающийся разработкой конструкции современного транзистора (технолог-маршрутчик) может предложить разные технические решения, в том числе и весьма экзотические. Но такой прибор нужно еще и уметь изготовить, причем в огромных количествах (несколько миллиардов штук на одной микросхеме). Поэтому проектирование прибора всегда тесно связано с разработкой технологии его изготовления. А технология прибора в значительной степени определяет технологию изготовления ИС, построенных на таком приборе.

Уже в конце 80-х годов XX века развитие технологии ИС достигло такого уровня, при котором общепринятые (экспериментальные) методы оптимизации конструкции транзисторов микросхем, являющиеся по своей сути методами технологических проб и ошибок, оказываются неприемлемыми. В этих условиях математическое моделирование, подразумевающее строгое, без упрощающих предположений, решение уравнений физики полупроводников является эффективным средством улучшения технико-эксплуатационных характеристик, снижения затрат материальных и трудовых ресурсов и сроков разработки новых изделий.

Упомянутые уравнения позволяют промоделировать разные транзисторы. Для анализа процессов в конкретном приборе необходимо описать его конструкцию: расположение и размеры областей разных материалов (кремний, окись кремния, нитрид кремния, поликремний, металл), образующих прибор, а в полупроводнике (в кремнии) – задать распределение примеси. На ранних этапах проектирования прибора нет необходимости детально описывать эти величины. В этом случае области прибора задаются в виде прямоугольного параллелепипеда, дополненного пирамидальными и/или призматическими участками, а примесный профиль – простыми аналитическими функциями. Но когда процесс проектирования доходит до момента оптимизации конструкции прибора и его технологии возникает необходимость детального описания этой конструкции. И здесь приходится использовать модели технологии ИС. В этом случае моделирование прибора выглядит следующим образом. По заданному описанию техпроцесса изготовления транзистора и его топологии выполняется моделирование этого процесса. Рассчитывается рельеф границ раздела разных материалов, присутствующих в приборе, и распределение примеси в них. Эта информация передается в программу моделирования прибора, которая рассчитывает его характеристики и параметры его схемотехнической модели. Последняя используется в программе схемотехнического моделирования для расчета электрических параметров основных логических элементов, применяемых в будущей цифровой ИС. Именно такова структура современных САПР приборно-технологического уровня (ТСАПР), называемых иногда виртуальными фабриками.

6.2. Моделирование технологии интегральных схем

Технология производства монокристаллических ИС представляет собой совокупность механических, физико-химических, электрофизических и других видов обработки полупроводниковых, диэлектрических и прочих материалов и полуфабрикатов, а также методов их контроля. Своеобразие технологического процесса изготовления микросхем заключается в том, что вся последовательность многочисленных (до двухсот и более) технологи-

ческих операций относится ко всем микросхемам, изготавливаемым на одной кремниевой пластине (подложке).

Технологические операции условно можно разделить на две группы: основные и вспомогательные. Первые используются для формирования компонентов микросхемы и металлизации. Вторые, как правило, применяются только для того, чтобы устранить влияние одной основной операции на другую. Например, после вскрытия окна в окисле необходимо удалить фоторезист. Если для этого используется жидкий селективный травитель, удаляющий только фоторезист, подложку помещают в сосуд с травителем. Если эта операция выполнена качественно, весь фоторезист удален, моделировать эту операцию не нужно. Важно знать только размеры и расположение окна в окисле. Если же на пластине остались следы (остатки) фоторезиста, т.е. пластина грязная, это резко отрицательно скажется на выходе годных изделий, будет много брака. В этом случае специалист, занимающийся только процессом фотолитографии (технолог-операционщик) может, используя специальные средства, проводить моделирование процессов травления фоторезиста чтобы обеспечить требуемое качество этой операции.

После удаления фоторезиста нужно очистить пластину от остатков травителя (промыть деионизованной водой) а затем просушить. Это тоже вспомогательные операции.

Основными являются следующие операции:

1. ионная имплантация,
2. диффузионное введение (загонка) примеси (например, из поликремния),
3. отжиг,
4. эпитаксия,
5. окисление,
6. осаждение (нанесение) слоев разных материалов на поверхность пластины,
7. травление этих слоев.

Именно их и нужно моделировать.

Одна из первых удачных программ одномерного (в глубь пластины) моделирования технологии кремниевых интегральных схем SUPREM II была разработана в Стэнфордском университете в США (1978 г.). В 1993 году появилась более совершенная версия этой программы (SUPREM III) используемая многими производителями современных ТСАПР.

Структура этой программы такова, что каждая технологическая операция может быть промоделирована как отдельно от других, так и в общей последовательности в соответствии с технологическим маршрутом изготовления ИС. Выходными параметрами программы в конце каждой операции являются одномерные распределения всех примесей, используемых в техпроцессе. Очевидно, что при последовательном постадийном расчете выходные параметры очередной операции являются начальными условиями для следующей.

6.2.1. Модель ионной имплантации

Ионная имплантация представляет собой обработку пластины пучком высокоэнергетических ионов примеси. Энергия ионов в диапазоне 3–500 кэВ достаточна для их проникновения в приповерхностную область кремниевой подложки на глубину 10–1000 нм. При этом естественный окисел толщиной порядка 3 нм практически не влияет на результаты проведения этой операции.

Основным преимуществом технологии ионной имплантации является возможность точного управления количеством внедренных атомов примеси, их распределением по глубине, возможность введение примеси в кремний через слои других материалов, например, легирование канала МОП транзистора сквозь подзатворный окисел.

В ранних работах, посвященных воздействию высокоэнергетических частиц на материалы, основной теорией, позволяющей рассчитать распределение частиц по глубине,

являлась теория ЛШШ (по первым буквам фамилий авторов, Линхард, Шиотт, Шафт). Концентрация внедренных ионов в зависимости от положения иона в мишени определяется выражением

$$C(x) = \frac{\Phi}{\sqrt{2\pi}\Delta R_p} \exp\left[-\frac{(x - R_p)^2}{2\Delta R_p^2}\right] \quad (1)$$

где x – расстояние от поверхности в глубину образца в направлении падения пучка; Φ – доза имплантации; R_p – проективный пробег; ΔR_p – среднеквадратичное отклонение. Эксперименты, однако, показали, что реальный профиль распределения примеси не симметричный. Чтобы это учесть в современных программах используются распределения типа двойной Гаусс, т.е. в (1) при $x \leq R_p$ $\Delta R_p = \Delta R_{p1}$, а при $x > R_p$ $\Delta R_p = \Delta R_{p2}$. Для некоторых видов примеси (например, для бора) лучшую точность дает распределение PEARSON IV:

$$C(x) = A \left| b_0 + b_1 z + b_2 z^2 \right|^{1/2b_2} \exp\left[-\frac{b_1/b_2 + 2a}{D} \operatorname{arctg} \left| \frac{2b_2 z + b_1}{D} \right| \right] \quad (2)$$

где $z = x - R_p$, $D = \sqrt{4b_0 b_2 - b_1^2}$ и A, b_0, b_1, b_2 – параметры распределения. Параметры обоих распределений получены путем подгонки по экспериментальным примесным профилям.

В современных условиях, когда не редко легирование проводится через многослойные структуры, последующие отжиги проводятся при относительно низких температурах за очень короткое время, упомянутые распределения не обеспечивают требуемой точности. Поэтому уже в первой версии программы SUPREM III для этого случая использовался метод кинетического моделирования ионной имплантации. При этом детально рассматривается движение иона в подложке, учитываются вероятности его взаимодействия с атомами полупроводника или диэлектрика, а также примесными атомами, введенными в моделируемую структуру ранее. Такое моделирование обеспечивает высокую точность расчетов, однако и время таких расчетов достаточно велико.

6.2.2. Модель высокотемпературной операции

Высокотемпературной называется операция, при которой разогрев пластины настолько высок, что приходится учитывать диффузионное перераспределение примеси в пластине. Это операции 2-5. Причем для окисления и эпитаксии диффузия – не основной процесс, но учесть ее необходимо.

Основой для построения моделей всех высокотемпературных операций является уравнение непрерывности, которое в интегральной форме имеет следующий вид

$$\frac{d}{dt} \int_{V(t)} C dV = - \oint_{S(t)} (\bar{F} \cdot \bar{\eta}) dS + \int_{V(t)} (G - R) dV \quad (3)$$

где $V(t), S(t)$ – объем и поверхность интегрирования, которые в общем случае являются функциями времени (например, при моделировании процесса окисления), \bar{F} – диффузионный поток, $\bar{\eta}$ – внешняя нормаль к поверхности интегрирования, G, R – скорости “появления” и “исчезновения” примеси (например, при переходе из активного состояния в неактивное). В простейшем случае диффузионный поток определяется первым законом Фика

$$\bar{F} = -D_k \nabla C \quad (4)$$

где D_k – коэффициент диффузии, $\nabla = \frac{\partial}{\partial x} \bar{i} + \frac{\partial}{\partial y} \bar{j} + \frac{\partial}{\partial z} \bar{k}$ – оператор градиента.

В таком виде поток можно рассчитывать только при очень низкой концентрации примеси. В общем случае это уравнение имеет вид:

$$\bar{F} = -D\nabla C + Z\mu N\bar{E} \quad (5)$$

где D – условный коэффициент диффузии, Z – зарядовое состояние примеси (+1 для доноров, -1 для акцепторов), μ – подвижность примеси, N – концентрация электрически активной примеси, \bar{E} – напряженность электрического поля.

Следуя общепринятой практике, параметр D будем называть просто коэффициентом диффузии. Строго говоря, это уже не коэффициент (не константа), а сложная функция условий, существующих в кристалле полупроводника. Он зависит от концентрации моделируемой примеси, от вида и концентрации других примесей, используемых в техпроцессе, от кристаллографической ориентации подложки, от условий обработки (например, при окислении коэффициент диффузии значительно изменяется). Если в кристалле присутствуют электрически активные виды примеси, возникает электрическое поле, появляется еще один, дрейфовый, механизм переноса примеси (второе слагаемое в (5)). Это приводит к необходимости решать еще и уравнение Пуассона, чтобы рассчитать величину \bar{E} . Далее будем считать, что электрическое поле достаточно мало, что позволяет пренебречь вторым слагаемым в (5).

Для решения уравнений (3) и (5) необходимо определить область моделирования (участок пространства, в котором решается задача), задать начальные и граничные условия. Строго говоря, задача имеет три пространственных измерения. Однако ее сложность чрезвычайно велика. Поэтому на ранних этапах разработки технологии ИС используют одно- и двумерные приближения.

Для одномерного потока вдоль оси x , перпендикулярной поверхности подложки и направленной вглубь, уравнение (3) принимает вид

$$\frac{d}{dt} \int_{x_1}^{x_2} C(x) dx = -(F(x_2) - F(x_1)). \quad (6)$$

При записи этого уравнения предполагается, что C – полная концентрация примеси, поэтому $G = 0$ и $R = 0$. Уравнение (5) в одномерном случае имеет вид

$$F = -D \frac{\partial C}{\partial x}. \quad (7)$$

Областью моделирования в этом случае должен быть весь кристалл и граничные условия нужно ставить на его границах (на его верхней и нижней поверхности). Но расчет распределения примеси по всей толщине кристалла будет крайне неэффективным, т.к. элементы ИС расположены в тонком слое у поверхности пластины. Поэтому отрезок, на котором решается (6), выбирается следующим образом: начало координат (начало отрезка) помещается на поверхность подложки, ось координат (ось x) направлена вниз, вглубь кристалла, длина отрезка L (глубина моделирования) выбирается из условия, что фронт диффузии не проникнет на эту глубину при проведении всего техпроцесса. Это позволяет в точке L поставить следующее граничное условие

$$C(t, x = L) = C_{\text{фон}}, \quad (8)$$

где $C_{\text{фон}}$ – фоновая концентрация примеси.

Как уже отмечалось ранее, при последовательном, пооперационном моделировании техпроцесса, распределение примеси, рассчитанное при моделировании текущей операции, является начальным условием для моделирования следующей операции.

Граничное условие на поверхности подложки определяется типом высокотемпературной операции. Рассмотрим эти операции.

1. Диффузионная загонка примеси. Такой процесс будет происходить, например, при нанесении сильно легированного поликремния на кремний. В приповерхностный слой кремния будет проникать примесь и диффундировать вглубь. Концентрация примеси на поверхности равна предельной растворимости:

$$C(t, x = 0) = C_{\text{ПРЕД}}, \quad (9)$$

2. Отжиг. При этом кристалл кремния нагревается до высокой температуры и выдерживается в нагретом состоянии некоторое время. Эта операция проводится после ионной имплантации для того, чтобы устранить дефекты решетки, возникшие из-за воздействия высокоэнергетических ионов. Даже небольшой слой естественного окисла устраняет испарение примеси в окружающую среду, поэтому здесь используется условие отсутствия потока примеси:

$$\frac{\partial C}{\partial x} \Big|_{t, x=0} = 0, \quad (10)$$

3. Эпитаксия. Это выращивание на поверхности подложки монокристалла, повторяющего кристаллическую структуру исходной пластины кремния, но имеющего другой тип и/или концентрацию примеси. Достаточно простой подход, позволяющий, однако, с хорошей точностью рассчитать распределение примеси в структуре эпитаксиальной пленки – подложка состоит в следующем. Скорость роста пленки значительно больше, чем скорость диффузии примеси из подложки в растущую пленку и из пленки в подложку. В этом случае можно считать, что пленка и присутствующая в ней примесь появляются мгновенно (отрезок НО на рис.1), информация об этом заносится в соответствующие массивы и далее моделируется отжиг структуры пленка–подложка как единого целого (начало координат переносится в точку Н, длина области моделирования равна Н+L). Время моделирования равно времени эпитаксии.

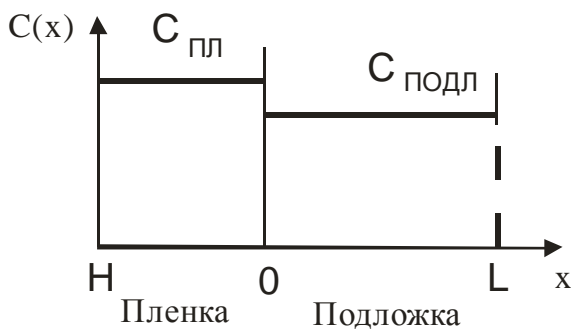


Рис. 1

Более точный подход применен в программе SUPREM III. Здесь рассматривается динамика движения границы растущей пленки. На рис.2 показана дискретизация пространства, используемая в программе. Сплошная вертикальная линия разделяет газовую и твердую (подложка + часть выросшей пленки) фазы. Область монокристалла кремния разделена на дискретные элементы штриховыми линиями, которые являются границами этих элементов. Концентрация легирующей примеси считается постоянной внутри каждого элемента ($C_i, C_{i+1}, C_{i+2}, \dots$). Учет поверхностного граничного условия проводится в два этапа.

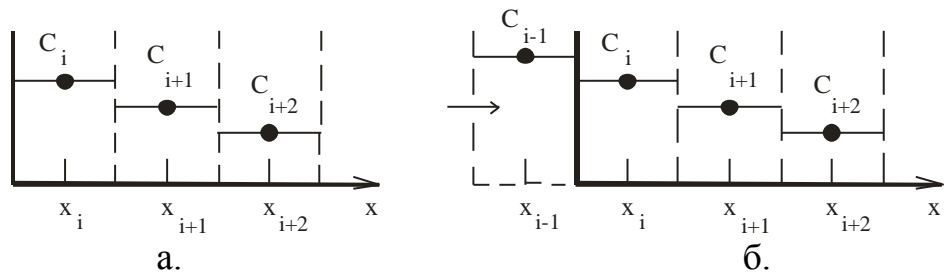


Рис.2

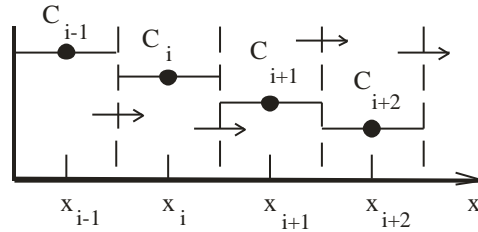


Рис. 2.в

Этап 1. В момент времени $t = t_0$ (рис.2.а) известно распределение примеси в кремнии. Оно является либо начальным условием, либо результатом расчетов, проведенных до момента времени t_0 . Очередной цикл расчетов начинается с добавления нового элемента x_{i-1} (рис.2.б). Концентрация примеси C_{i-1} внутри нового элемента определяется условиями его выращивания. При этом учтены только процессы проникновения примеси в добавляемый элемент (горизонтальная стрелка на его левой границе) и не учтен процесс одновременного перераспределения примеси в кремнии.

Этап 2. Теперь рассчитывается только перераспределение примеси, происходящее в процессе роста элемента x_{i-1} . Это фактически диффузионное перераспределение примеси за один временной интервал Δt (горизонтальные стрелки между границами элементов).

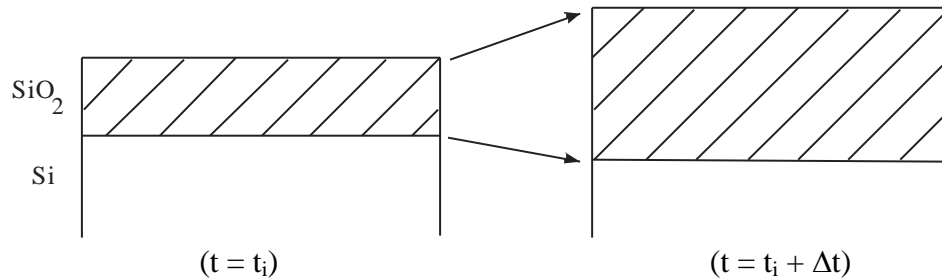


Рис.3

4. Окисление. При моделировании этой операции приходится учитывать, во-первых, движение оси координат вверх. Это объясняется тем, что на образование единицы толщины слоя окисла (h_{OX}) затрачивается 0,44 единицы толщины слоя кремния (h_{Si}), т.е. $0,44 \cdot h_{OX} = h_{Si}$ (рис.3). Во-вторых, граница раздела Si-SiO₂ по этой же причине движется вниз. В-третьих, на границе раздела присутствует сегрегационный поток примеси F_S

$$F_S = h(C_1 - C_2/m_{1-2}), \quad (11)$$

где h – коэффициент массопереноса, C_1 – концентрация примеси на границе раздела в области 1 (в окисле), C_2 – концентрация примеси на границе раздела в области 2 (в кремнии), m_{1-2} – коэффициент сегрегации ($m_{1-2} = C_2/C_1$). В-четвертых, как в кремнии, так и в окисле происходит диффузионное перераспределение примеси. На поверхности растущего окисла обычно используется условие типа (10).

6.2.3. Численная реализация модели высокотемпературной операции

В качестве примера рассмотрим реализацию модели диффузионной загонки. При этом, используя уравнения (6), (7) и граничные условия (8) и (9), необходимо рассчитать распределение примеси на интервале (0, L). Будем решать эту задачу численно, проводя ее дискретизацию. Для этого введем разностные сетки по координате и по времени. Обе сетки имеют, в общем случае, неравномерный шаг. Пронумеруем узлы и шаги сетки. Индекс i (нижний индекс) используем для обозначения номера узла пространственной сетки и шага сетки (h_i), расположенного справа от этого узла (рис.4).

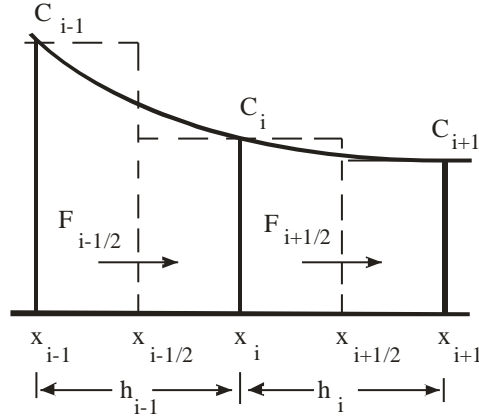


Рис.4

Индекс k (верхний индекс) определяет текущий момент времени и предыдущий шаг по времени (τ^k). Для аппроксимации интеграла в (6), а также потока и коэффициента диффузии в (7), введем вспомогательную сетку по координате (рис.4), сдвинутую на пол шага относительно основной. Для обозначения узлов этой сетки используем полуцелые нижние индексы. Учитывая эти замечания, получим:

$$\int_{x_{i-1/2}}^{x_{i+1/2}} C(x,t) dx = C_i(t) \frac{h_{i-1} + h_i}{2}, \quad (12)$$

$$F_{i+1/2}(t) = -D_{i+1/2}(t) \frac{C_{i+1}(t) - C_i(t)}{h_i}. \quad (13)$$

Используя эти соотношения, получим дискретный по пространству аналог уравнения (6):

$$\frac{d}{dt} C_i(t) \frac{h_{i-1} + h_i}{2} = D_{i+1/2}(t) \frac{C_{i+1}(t) - C_i(t)}{h_i} - D_{i-1/2}(t) \frac{C_i(t) - C_{i-1}(t)}{h_{i-1}} \quad (14)$$

Для дискретизации уравнения (15) по времени используем полуявную разностную схему

$$\frac{C_i^k - C_i^{k-1}}{\tau^k} = \frac{2}{h_i + h_{i-1}} \left[D_{i+1/2}^{k-1} \frac{C_{i+1}^k - C_i^k}{h_i} - D_{i-1/2}^{k-1} \frac{C_i^k - C_{i-1}^k}{h_{i-1}} \right]. \quad (15)$$

Эта формула применима ко всем узлам разностной сетки, кроме первого и последнего. Для первого узла используем аналог уравнения (9)

$$C_1^k = C_{\text{ПРЕД}}, \quad (16)$$

а для последнего – аналог уравнения (8)

$$C_N^k = C_{\text{ФОН}}. \quad (17)$$

Соотношения (15)-(17) образуют систему линейных алгебраических уравнений (СЛАУ), которую необходимо решать на каждом временном слое. Использование (15) вместе с (16) для второго узла сетки дает линейное уравнение с двумя неизвестными (C_2^k, C_3^k). Для предпоследнего узла используются (15) и (17), что тоже приводит к линейному уравнению с двумя неизвестными (C_{N-2}^k, C_{N-1}^k). Уравнение (15), используемое во всех остальных узлах, содержит три неизвестных. Таким образом, эта СЛАУ имеет трехдиагональную матрицу, что позволяет применять для ее решения метод прогонки. Размерность этой системы (m) равна $N-2$.

Запишем покомпонентно СЛАУ такого вида

$$\begin{aligned} b_1 x_1 + c_1 x_2 &= d_1 \\ a_2 x_1 + b_2 x_2 + c_2 x_3 &= d_2 \\ \dots & \\ a_j x_{j-1} + b_j x_j + c_j x_{j+1} &= d_j \\ \dots & \\ a_{m-1} x_{m-2} + b_{m-1} x_{m-1} + c_{m-1} x_m &= d_m \\ a_m x_{m-1} + b_m x_m &= d_m. \end{aligned} \quad (18)$$

Найдем x_1 из первого уравнения системы

$$x_1 = \alpha_1 x_2 + \beta_1, \quad (19)$$

где $\alpha_1 = -c_1/b_1$, $\beta_1 = d_1/b_1$, и подставим его во второе уравнение системы

$$a_2(\alpha_1 x_2 + \beta_1) + b_2 x_2 + c_2 x_3 = d_2. \quad (20)$$

Перепишем его в виде, подобном (19),

$$x_2 = \alpha_2 x_3 + \beta_2 \quad (21)$$

где $\alpha_2 = -c_2/(b_2 + a_2\alpha_1)$, $\beta_2 = (d_2 - a_2\beta_1)/(b_2 + a_2\alpha_1)$. Опять подставим его в третье уравнение системы и т.д.

На i -м шаге этого процесса ($1 < i < m$) имеем

$$x_i = \alpha_i x_{i+1} + \beta_i, \quad (22)$$

где $\alpha_i = -c_i/(b_i + a_i\alpha_{i-1})$, $\beta_i = (d_i - a_i\beta_{i-1})/(b_i + a_i\alpha_{i-1})$.

На последнем, m -м шаге, подстановка в последнее уравнение выражения $x_{m-1} = \alpha_{m-1}x_m + \beta_{m-1}$ дает

$$a_m(\alpha_{m-1}x_m + \beta_{m-1}) + b_m x_m = d_m, \quad (23)$$

откуда

$$x_m = (d_m - a_m\beta_{m-1})/(b_m + a_m\alpha_{m-1}). \quad (24)$$

Значения остальных неизвестных x_i для $i = m - 1, m - 2, \dots, 1$ теперь легко вычисляются по формуле (22).

6.2.4 Особенности реализации

При записи уравнения (15) нами была использована полуявная разностная схема. Это было сделано для того, чтобы это уравнение было линейным, т.к. если использовать чисто неявную разностную схему, т.е. коэффициент диффузии (D) брать в момент времени k , это уравнение стало бы нелинейным из-за зависимости коэффициента диффузии от концентрации примеси. Такой подход является достаточно точным, т.к. изменение D при проведении технологической операции невелико. Сделать его еще меньшим можно выбором достаточно малой величины шага по времени τ^k . Кроме того, в тех случаях, когда D все же изменяется достаточно сильно, при переходе от момента времени $k-1$ к моменту k используется итерационный процесс (метод простой итерации), когда такой переход повторяется несколько раз.

6.2.5. Модель операций травления и осаждения

В одномерном случае модели этих операций очень просты. Обе операции низкотемпературные, диффузионного перераспределения примеси нет. Поэтому в моделирующей программе просто добавляются или удаляются элементы массивов, содержащих распределения примеси, соответствующие толщине добавляемого или удаляемого слоёв.

6.3. Моделирование транзисторов интегральных схем

Моделирование полупроводникового прибора представляет собой анализ процессов переноса зарядов в нем, позволяющей рассчитать все электрические характеристики прибора. Для этого используется математическая модель, описывающая с требуемой точностью эти процессы. Чаще всего в качестве такой модели используется система уравнений в частных производных, которую называют гидродинамической моделью в диффузионно-дрейфовом приближении или просто диффузионно-дрейфовой моделью (ДДМ). Эта модель, в стационарном случае, включает три основных уравнения, которые мы запишем в интегральной форме. Это уравнение Пуассона

$$\oint_S (\bar{D} \cdot \bar{\eta}) ds = \int_V \rho dv, \quad (25)$$

уравнения непрерывности для электронов

$$-\frac{1}{q} \oint_S (\bar{j}_n \cdot \bar{\eta}) ds = \int_V (g - r) dv \quad (26)$$

и дырок

$$\frac{1}{q} \oint_S (\bar{j}_p \cdot \bar{\eta}) ds = \int_V (g - r) dv. \quad (27)$$

Уравнения (26) и (27) дополняются уравнениями переноса, которые при отсутствии внешних магнитных полей имеют вид

$$\bar{j}_n = q\mu_n n \bar{E} + qD_n \nabla n, \quad (28)$$

$$\bar{j}_p = q\mu_p p \bar{E} - qD_p \nabla p. \quad (29)$$

Для умеренных электрических полей

$$\bar{D} = \varepsilon \varepsilon_0 \bar{E} = -\varepsilon \varepsilon_0 \nabla \varphi \quad (30)$$

и суммарный заряд единицы объёма равен

$$\rho = q(p - n + C_D - C_A). \quad (31)$$

Здесь использованы следующие обозначения:

\vec{D} – вектор электрической индукции;

\vec{j}_n, \vec{j}_p – плотности тока электронов и дырок;

g, r – скорости генерации и рекомбинации электронов и дырок;

q – заряд электрона;

n, p – концентрация электронов и дырок;

\bar{E} – напряженность электрического поля;

μ_n, μ_p – подвижности электронов и дырок;

D_n, D_p – коэффициенты диффузии электронов и дырок, связанные с их подвижностями соотношением Эйнштейна:

$$\frac{D_n}{\mu_n} = \frac{kT}{q}, \quad \frac{D_p}{\mu_p} = \frac{kT}{q}; \quad (32)$$

φ – электростатический потенциал;

ε_0 – электрическая постоянная вакуума;

ε – относительная диэлектрическая проницаемость среды;

C_D, C_A – концентрация донорной и акцепторной примеси;

$\nabla = \frac{\partial}{\partial x} \vec{i} + \frac{\partial}{\partial y} \vec{j} + \frac{\partial}{\partial z} \vec{k}$ – оператор дифференцирования.

Решение этой системы уравнений позволяет рассчитать трехмерные распределения концентраций дырок, электронов и потенциала. Знание этих величин позволяет найти все электрические параметры прибора. Однако расчет трехмерных величин – очень трудоемкая задача. Поэтому, например, при моделировании МОП-транзистора достаточно рассмотреть его поперечное сечение (двумерная задача). Для двумерной задачи интегрирование в правой части (25)-(27) будет вестись по площади S , а в левой части – по периметру S , который обозначим L .

Стационарное приближение – еще один прием, позволяющий сократить затраты на проектирование прибора. Знание упомянутых распределений позволяет рассчитать не только статические характеристики прибора (ВАХ, ток насыщения, напряжение пробоя и т.п.), но и его динамические параметры (заряды и емкости).

Для завершения постановки задачи необходимо описать конструкцию прибора, задать граничные условия на внешних и внутренних границах (границах раздела разных материалов, образующих прибор).

С целью некоторого упрощения записи формул выполним нормировку рассматриваемой задачи. Для этого умножим выражения (25)-(32) на специально подобранные константы. В результате они приобретут вид:

$$\oint_L (\vec{D} \cdot \vec{n}) dl = \int_S (p - n + N) ds, \quad (33)$$

$$\oint_L (\vec{j}_n \cdot \vec{n}) dl = - \int_S R ds, \quad (34)$$

$$\oint_L (\vec{j}_p \cdot \vec{n}) dl = \int_S R ds, \quad (35)$$

$$\bar{j}_n = -\mu_n (n \nabla \varphi - \nabla n), \quad (36)$$

$$\bar{j}_p = -\mu_p (p \nabla \varphi + \nabla p), \quad (37)$$

$$\bar{D} = -\varepsilon \nabla \varphi, \quad (38)$$

где $N = C_D - C_A$ и $R = g - r$.

6.3.1 Разностная аппроксимация уравнений сохранения

Аналитическое решение этой задачи получить не удастся, поэтому ее решают численно. Численное решение на ЭВМ краевых задач для уравнений в частных производных предполагает замену непрерывной задачи ее разностным аналогом – системой алгебраических уравнений относительно конечного числа неизвестных. Для этого область моделирования покрывается сеткой и уравнения модели аппроксимируются на этой сетке.

Выполним дискретизацию задачи, для чего в области прибора построим две сетки – основную

$$x = x_i, i = 1, 2, \dots, n; \quad y = y_j, j = 1, 2, \dots, m$$

с шагом h_i по координате x и шагом r_j по координате y , т.е.

$$x_1 = 0, x_{i+1} = x_i + h_i, i = 1, 2, \dots, n - 2, x_n = L_x$$

$$y_1 = 0, y_{j+1} = y_j + r_j, j = 1, 2, \dots, m - 2, y_m = L_y$$

где L_x, L_y – размеры области моделирования, и вспомогательную. Основная сетка строится таким образом, что линии основной сетки проходят по границам области моделирования. Линии вспомогательной сетки смещаются относительно линий основной сетки на величину, равную половине соответствующего шага дискретизации.

Точки пересечения взаимно перпендикулярных линий основной сетки, их называют узлами сетки, образуют область определения функции дискретного изменения аргумента, которую в свою очередь называют сеточной функцией. В этих точках будем искать функции $\varphi(x, y)$, $n(x, y)$, $p(x, y)$.

На рис.5 приведен участок сетки. Здесь непрерывные отрезки – линии основной сетки, штриховые – линии вспомогательной сетки. Как и при моделировании технологии ИС ось X направлена вглубь подложки, ось Y – вправо. Ячейкой сетки (ЯС) является прямоугольник, ограниченный смежными линиями основной сетки (прямоугольник АКВО), а ячейкой алгебраизации (ЯА) – прямоугольник, ограниченный смежными линиями вспомогательной сетки (прямоугольник EFGH). ЯА линиями основной сетки разделена на четыре квадранта (римские цифры).

Основные уравнения модели (33)-(35) имеют одинаковый вид, поэтому и дискретизация их выполняется одинаково. В качестве примера рассмотрим уравнение непрерывности для дырок. В каждом узле сетки, в том числе и в граничных узлах, заменим интегралы, имеющиеся в уравнениях, их приближенными значениями. Для этого рассмотрим ЯА, приведенную на рис.5. Пронумеруем линии сетки и используем целые индексы i, j для обозначения линий основной сетки, и полуцелые индексы — для обозначения линий вспомогательной сетки так, как это показано на рис.5. Двойной индекс будем использовать для обозначения точки пересечения соответствующих линий сетки.

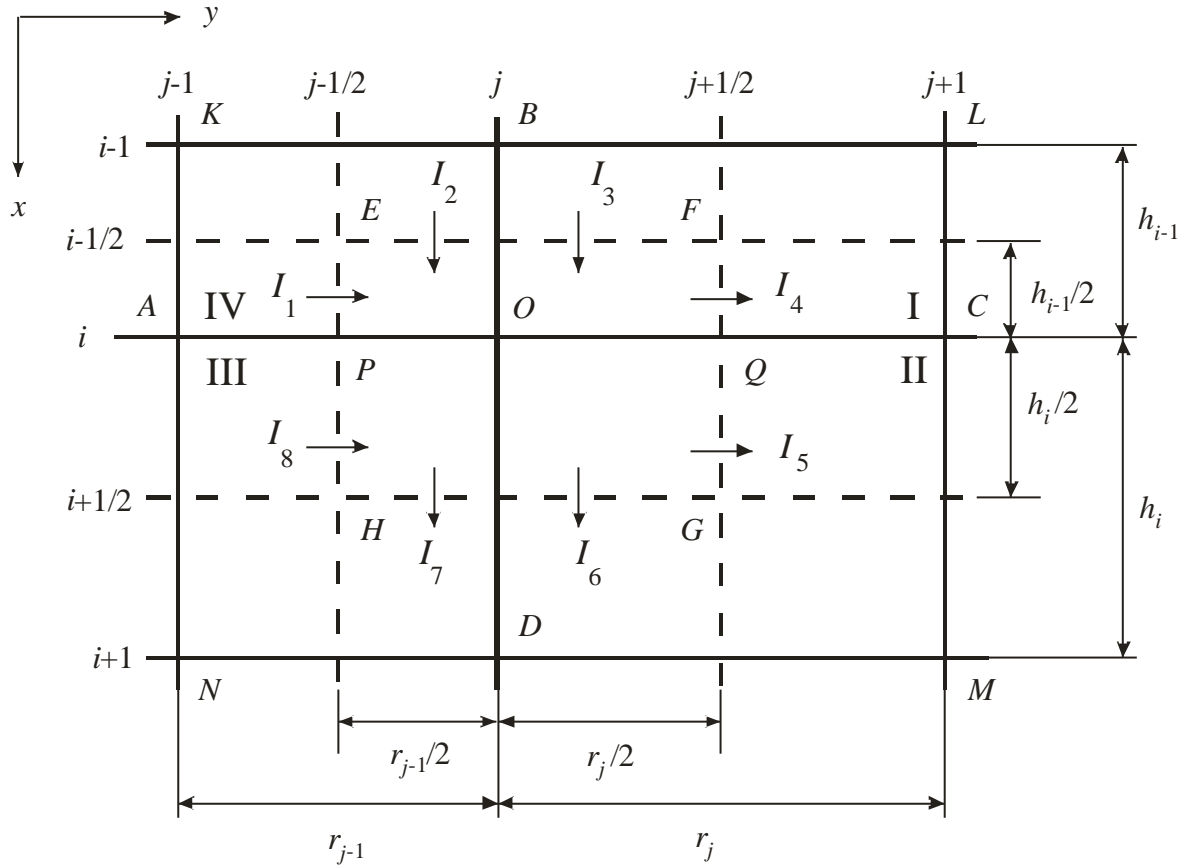


Рис. 5 Ячейка сетки и ячейка алгебраизации.

Уравнение (35) справедливо для любой замкнутой области, в том числе и для ЯА. Ее внешней границей является прямоугольник EFGH, стороны которого равны $\frac{h_{i-1} + h_i}{2}$ и $\frac{r_{j-1} + r_j}{2}$. По построению ток дырок, текущий из A в O, перпендикулярен отрезку EH, поэтому полный ток дырок, втекающий в ЯА через ее левую границу равен (здесь и далее индекс p опускаем)

$$I_{AO} = j_{i,j-1/2} \cdot \frac{h_{i-1} + h_i}{2}.$$

Определив аналогичным образом токи, текущие через остальные границы ЯА, получим аппроксимацию интеграла по контуру EFGH в уравнении (35). При этом учтем, что втекающий ток нужно брать со знаком минус:

$$\oint_L (\vec{j} \cdot \vec{\eta}) ds \approx (j_{i,j+1/2} - j_{i,j-1/2}) \frac{h_{i-1} + h_i}{2} + (j_{i+1/2,j} - j_{i-1/2,j}) \frac{r_{j-1} + r_j}{2}. \quad (39)$$

Интеграл, присутствующий в правой части уравнения, аппроксимируем следующим образом:

$$\int_S R ds = R_{i,j} \cdot \frac{r_{j-1} + r_j}{2} \cdot \frac{h_{i-1} + h_i}{2}. \quad (40)$$

Следовательно, аппроксимация (35) имеет вид

$$(j_{i,j+1/2} - j_{i,j-1/2}) \frac{h_{i-1} + h_i}{2} + (j_{i+1/2,j} - j_{i-1/2,j}) \frac{r_{j-1} + r_j}{2} = R_{i,j} \cdot \frac{r_{j-1} + r_j}{2} \cdot \frac{h_{i-1} + h_i}{2}. \quad (41)$$

На рис.5 каждый из токов, текущих через границы ЯА, разбит на две составляющих. Так $I_{AO} = I_1 + I_8$. Это объясняется тем, что если узел сетки лежит на верхней границе области моделирования, т.е. эта граница проходит по отрезку AC, ячейка алгебраизации содержит только II и III квадранты. Через границы такой ЯА текут токи I_5, I_6+I_7, I_8 . Ток через отрезок PQ равен нулю (это следует из граничных условий).

6.3.2 Разностная аппроксимация уравнений переноса

Выполним теперь разностную аппроксимацию уравнения (37) на отрезке OD. Для этого используем следующие приближенные выражения (здесь и далее индекс j опускаем):

$$\frac{dp}{dx} \approx \frac{p_{i+1} - p_i}{h_i}, \quad \frac{d\varphi}{dx} \approx \frac{\varphi_{i+1} - \varphi_i}{h_i}, \quad p_{i+1/2,j} \approx \frac{p_{i+1} + p_i}{2}. \quad (42)$$

Тогда из (37) имеем

$$j_{i+1/2} = -\mu_{i+1/2} \left[\frac{p_{i+1} + p_i}{2} \cdot \frac{\varphi_{i+1} - \varphi_i}{h_i} + \frac{p_{i+1} - p_i}{h_i} \right]. \quad (43)$$

Эта разностная аппроксимация использовалась в первых программах численного моделирования полупроводниковых приборов. Однако очень скоро выяснилось, что такой подход приводит к грубым ошибкам. Нередко при расчетах величина концентрации дырок и/или электронов была отрицательной (концентрация по определению величина положительная). Покажем причину этого. После несложных преобразований имеем

$$p_{i+1} = -\frac{2h_i j_{i+1/2}}{\mu_{i+1/2}(\varphi_{i+1} - \varphi_i + 2)} - \frac{(\varphi_{i+1} - \varphi_i - 2)}{(\varphi_{i+1} - \varphi_i + 2)} p_i \quad (44)$$

Если предположить, что в некоторой области прибора плотность тока имеет очень малую величину ($j_{i+1/2} \approx 0$), перепад потенциала $\varphi_{i+1} - \varphi_i > 2$ и $p_i > 0$, тогда оказывается, что $p_{i+1} < 0$. Это вынуждает, при использовании разностной аппроксимации (42), строить очень густую разностную сетку, чтобы обеспечить малость величины $\varphi_{i+1} - \varphi_i$.

Причина упомянутой проблемы проста. Аппроксимация производных в виде (42) будет достаточно точной только при условии, что зависимость $p = f(\varphi)$ на интервале $[x_i, x_{i+1}]$ будет близка к линейной. Но теоретические исследования статистики электронов и дырок в полупроводниках показали, что эта зависимость является экспоненциальной.

Решение этой задачи было предложено уже в конце 60-х годов XX века. Его суть состоит в том, что соотношение (37) предлагается рассматривать как обыкновенное дифференциальное уравнение, решением которого будет зависимость $p = p(x)$. Перепишем его в виде

$$\frac{dp}{dx} + p \frac{d\varphi}{dx} = -\frac{j}{\mu} \quad (45)$$

Предположим, что $\frac{d\varphi}{dx} = const$ на отрезке (x_i, x_{i+1}) . Проинтегрируем однородное уравнение ($j = 0$) на интервале (x_i, x) , где $x_i \leq x \leq x_{i+1}$,

$$\int_{p(x_i)}^{p(x)} \frac{dp}{p} = -\frac{d\varphi}{dx} \int_{x_i}^x dx',$$

откуда

$$\ln[p(x)] - \ln[p(x_i)] = \ln \frac{p(x)}{p(x_i)} = -\frac{d\varphi}{dx} (x - x_i)$$

или

$$p(x) = p_i \exp(\alpha \Delta x_i), \quad (46)$$

где

$$\alpha = -\frac{d\varphi}{dx}, \quad \Delta x_i = x - x_i.$$

Из (46) имеем

$$p(x) - p_i = p_i [\exp(\alpha \Delta x_i) - 1]$$

и

$$p_{i+1} - p_i = p_i [\exp(\alpha h_i) - 1].$$

Разделив первое равенство на второе, получим

$$p(x) = p_i + \Delta p_i \frac{\exp(\alpha \Delta x_i) - 1}{\exp(\alpha h_i) - 1}, \quad (47)$$

где

$$\Delta p_i = p_{i+1} - p_i.$$

Из формулы (47) имеем

$$\frac{dp}{dx} = \Delta p_i \alpha \frac{\exp(\alpha \Delta x_i)}{\exp(\alpha h_i) - 1}. \quad (48)$$

Подставив (47) и (48) в (45) и выполнив несложные преобразования, получаем искомое выражение

$$j_{i+1/2} = \mu_{i+1/2} \frac{\Delta \varphi_i}{h_i} \left[\frac{p_{i+1}}{\exp(-\Delta \varphi_i) - 1} + \frac{p_i}{\exp(\Delta \varphi_i) - 1} \right], \quad (49)$$

где $\frac{d\varphi}{dx} = \frac{\Delta \varphi_i}{h_i}$ и $\Delta \varphi_i = \varphi_{i+1} - \varphi_i$.

6.3.3 Решение алгебраизованной задачи

Разностная аппроксимация уравнений сохранения (33)-(35) вместе с уравнениями переноса (36)-(38) позволяет рассчитать три сеточные функции $\varphi_{i,j}$, $n_{i,j}$ и $p_{i,j}$. Для этого нужно решать систему нелинейных алгебраических уравнений (СНАУ), имеющую размерность $3N$, где N – число узлов разностной сетки. Представим эту систему в виде трех подсистем

$$F_\varphi(\bar{\varphi}, \bar{p}, \bar{n}) = 0, \quad (50)$$

$$F_n(\bar{\varphi}, \bar{p}, \bar{n}) = 0, \quad (51)$$

$$F_p(\bar{\varphi}, \bar{p}, \bar{n}) = 0. \quad (52)$$

N -го порядка. Первая подсистема соответствует алгебраизованному уравнению Пуассона, вторая и третья - алгебраизованным уравнениям непрерывности для электронов и дырок.

Для решения таких СНАУ в практике численного моделирования находят применение алгоритмы последовательного и параллельного типов. Алгоритмы первого типа основываются на последовательном решении в итерационном цикле уравнений непрерывности и Пуассона, в то время как в алгоритмах второго типа подсистемы (50)-(52) решаются одновременно.

Последовательный алгоритм представляет собой несколько итерационных процессов, включенных один в другой. Внешний цикл содержит три этапа, на каждом из которых решается одно из уравнений системы (50) – (52). На каждом этапе уточняется распределение потенциала, концентраций электронов и дырок, рассчитываются токи, текущие через выводы прибора. Критерием окончания внешних итераций является величина максимальной абсолютной и относительной поправки потенциала, относительной поправки токов и величина консервативности. Последнее значение определяется следующим образом

$$\Delta = \frac{\sum(I_K)}{\sum(|I_K|)},$$

где I_K – ток, текущий через K -й контакт. Суммирование ведется по всем контактам.

Уравнение Пуассона сильно не линейно по правой части (концентрации дырок и электронов экспоненциально зависят от потенциала). Поэтому для его решения используются итерации Ньютона (второй по уровню вложенности итерационный цикл). Здесь контролируются абсолютное и относительное изменение потенциала и полное число итераций. Следует подчеркнуть, что поправки потенциала в ньютоновом процессе и на внешних итерациях это разные величины и они задаются разными параметрами.

Линеаризованные системы (50)-(52) решаются одним из вариантов метода сопряженных градиентов с неполной факторизацией (НФСГ). Это итерационный метод и он образует третий по уровню вложенности итерационный цикл.

Параллельные алгоритмы приходится использовать в том случае, когда, например, необходимо моделировать работу полупроводниковых приборов в режимах высокого уровня инжекции или предпробойных режимах. В этом случае из-за сильной связи между уравнениями системы (50)–(52) сходимость последовательных алгоритмов становится чрезвычайно медленной, либо отсутствует вовсе. Суть параллельного алгоритма состоит в том, что для решения задачи применяют метод Ньютона сразу ко всем трем группам уравнений. При этом выполняется следующий итерационный процесс:

$$\begin{bmatrix} \varphi^{I+1} \\ n^{I+1} \\ p^{I+1} \end{bmatrix} = \begin{bmatrix} \varphi^I \\ n^I \\ p^I \end{bmatrix} + \begin{bmatrix} \delta\varphi^{I+1} \\ \delta n^{I+1} \\ \delta p^{I+1} \end{bmatrix}, \quad I = 0, 1, 2, \dots \quad (53)$$

$$\begin{bmatrix} \left(\frac{\partial F_\varphi}{\partial \varphi}\right)^I & \left(\frac{\partial F_\varphi}{\partial n}\right)^I & \left(\frac{\partial F_\varphi}{\partial p}\right)^I \\ \left(\frac{\partial F_n}{\partial \varphi}\right)^I & \left(\frac{\partial F_n}{\partial n}\right)^I & \left(\frac{\partial F_n}{\partial p}\right)^I \\ \left(\frac{\partial F_p}{\partial \varphi}\right)^I & \left(\frac{\partial F_p}{\partial n}\right)^I & \left(\frac{\partial F_p}{\partial p}\right)^I \end{bmatrix} \times \begin{bmatrix} \delta\varphi^{I+1} \\ \delta n^{I+1} \\ \delta p^{I+1} \end{bmatrix} = - \begin{bmatrix} F_\varphi(\varphi^I, n^I, p^I) \\ F_n(\varphi^I, n^I, p^I) \\ F_p(\varphi^I, n^I, p^I) \end{bmatrix}. \quad (54)$$

Размерность СЛАУ (54), которую приходится решать на каждом шаге параллельного процесса, будет в три раз больше, чем в случае последовательного алгоритма. Ее матрица будет сильно разреженной, т.к. при использовании пятиточечного шаблона каж-

дое из разностных уравнений системы (54) будет включать значения трех переменных в пяти узлах, т.е. не более 15 ненулевых коэффициентов.

Свойства матрицы не позволяют использовать современные высокоэффективные итерационные методы решения СЛАУ. Приходится применять прямые методы, Гаусса или LU-разложения.

6.4. Кинетические коэффициенты

Моделирование технологии и транзисторов ИС подразумевает на разных этапах решение одного или системы дифференциальных уравнений в частных производных. Современный уровень развития математики и вычислительной техники позволяет, используя численные методы, решать практически любые из упомянутых задач с достаточно высокой точностью. Поэтому результирующая точность расчетов практически полностью определяется набором коэффициентов (их называют кинетическими). Это такие параметры, как коэффициент диффузии, коэффициент сегрегации, параметры, определяющие скорость роста окисла и т. п., применяемые при моделировании технологии. При моделировании транзистора это подвижности носителей, их скорости генерации и рекомбинации и некоторые другие величины.

Многие из этих параметров имеют сложную, еще до конца не понятую природу. Поэтому их численные значения и зависимость от условий, существующих в соответствующем объекте моделирования, являются обобщением огромного экспериментального материала. Это позволило создать достаточно полный набор инженерных моделей (формул) этих параметров. Поскольку коэффициенты этих моделей получены настройкой на экспериментальные данные, расчеты по этим моделям, как правило, хорошо согласуются с экспериментом в своем ограниченном диапазоне входных параметров. Так как многие из этих моделей невозможно обосновать строго физически, а некоторые из них – просто противоречивы, их следует рассматривать просто как набор формул, как макромодели, тех или иных кинетических коэффициентов.

Эта особенность кинетических коэффициентов учтена во многих современных коммерческих программах. Так, например, вариант программы SUPREM III, продаваемый фирмой Silvaco, дает возможность пользователю изменять значения коэффициентов упомянутых моделей, настраивая программу на особенности технологии конкретного производства.

6.5. Уровни моделей технологических САПР

Все рассмотренные выше модели (за исключением модели операции ионной имплантации) используют так называемое гидродинамическое приближение. Суть его состоит в том, чтобы рассматривать систему огромного числа частиц как непрерывную среду. Такой подход применяется при моделировании жидкости (гидродинамика), газа (газовая динамика) и других систем многих частиц. Для этого вводится понятие концентрации как количества частиц в единице объема. Считается, что она является непрерывной функцией времени и координат. Далее, используя законы сохранения (количества вещества, заряда, импульса, энергии) выводятся дифференциальные уравнения в частных производных. Такой подход применим только в том случае, когда система содержит очень много частиц.

Оценим количество частиц (электронов), осуществляющих перенос заряда в современном nМОП транзисторе. В середине 2010 Intel выпускала микропроцессоры по проектным нормам 32 нм, в середине 2011 года было объявлено о переходе на 22 нм техпроцесс. В это же время (2011 г.) в США начата постройка нового завода фирмы Intel, который в 2013 г. должен выпускать процессоры по 14 нм технологии.

32 нм это разрешающая способность технологии фотолитографии и соответствующего оборудования, которые определяют минимально возможные размеры элементов то-

пологии ИС (проектные нормы). Будем считать, что длина и ширина истокового контакта равна 32 нм. Зазор между контактами затвора и истока – 32 нм. Размеры прибора в глубину кристалла значительно меньше его размеров в плане. Положим, что глубина истока – 16 нм. Таким образом, исток это прямоугольный параллелепипед размерами 64 нм (длина) × 32 нм (ширина) × 16 нм (глубина) и его объем $32768 \text{ нм}^3 = 32,768 \cdot 10^{-18} \text{ см}^3$. Если концентрация примеси в области истока имеет очень большую величину – 10^{20} см^{-3} , общее количество электронов в истоке менее 3500. Если в процессе работы полевого транзистора сразу все электроны из истока перейдут в канал (что мало вероятно), их слишком мало, чтобы можно было вводить понятие концентрации и использовать гидродинамическое приближение. Кроме того, постоянная решетки кремния равна $5,43 \text{ \AA}$ (0,543 нм), следовательно, глубина области истока меньше 60 периодов решетки. Предположим, что глубина (толщина) канала равна половине глубины истока (реально она еще меньше), тогда оказывается, ток канала течет в области толщиной всего несколько десятков атомных слоев. Следовательно, в области канала и полупроводник нельзя рассматривать как непрерывную среду. Все это приводит к необходимости использовать кинетический подход.

В этом случае электронно-дырочный газ (плазма) рассматривается как система ограниченного числа взаимодействующих частиц. Тем или иным способом (метод молекулярной динамика, метод Монте-Карло, метод кинетических уравнений) рассчитываются координаты и скорости всех частиц системы, что далее позволяет рассчитать любые параметры анализируемого прибора.