

## ЗУ.

Основные материалы по разделу – в тетрадях. Тут некоторые определения, общие слова и рисунки...

### 1.Классификация запоминающих устройств.

ЗУ – устройство, предназначенное для записи, хранения и чтения данных. В основе работы запоминающего устройства может лежать любой физический эффект, обеспечивающий приведение системы к двум или более устойчивым состояниям.

ЗУ можно классифицировать по следующим некоторым параметрам:

*По форме записанной информации:*

- аналоговые;
- цифровые.

*По возможности перезаписи:*

- С однократной записью без возможности перезаписи (ПЗУ) (например, CD-R, масочные микросхемы ПЗУ).
- Перепрограммируемые ЗУ (ППЗУ) - запоминающие устройства с возможностью многократной перезаписи на специализированном/дополнительном оборудовании (например, CD-RW, микросхемы EPROM). Как следствие, интегральное время перезаписи велико (минуты).
- Устройства с многократной перезаписью по месту функционирования (например, жёсткие магнитные диски, микросхемы оперативной памяти). Нет чёткой границы с предыдущим.

*По назначению:*

- Оперативная память (ОЗУ) – память, в которой размещаются данные, над которыми непосредственно производятся операции процессора. Оперативная память может иметь несколько иерархических уровней. Примеры: SRAM, DRAM.
- Внутренние устройства для долговременного хранения информации (например, CMOS-память, жёсткие диски, SSD).
- Внешние носители, предназначенные для резервного хранения либо переноса информации от одного устройства к другому (например, дискеты, флешки).
- Запоминающие устройства для идентификации и платежей (например, магнитные карты, метки RFID).

*По энергозависимости:*

- энергонезависимые, записи в которых не стираются при снятии электропитания;
- энергозависимые, записи в которых стираются при снятии электропитания;
- статические, которым для хранения информации достаточно сохранения питающего напряжения;
- динамические, в которых информация со временем разрушается (деградирует), и, кроме подачи электропитания, необходимо производить её периодическое восстановление (регенерацию).

*По типу доступа:*

- С последовательным доступом (например, магнитные ленты).
- С произвольным доступом (RAM; например, оперативная память).

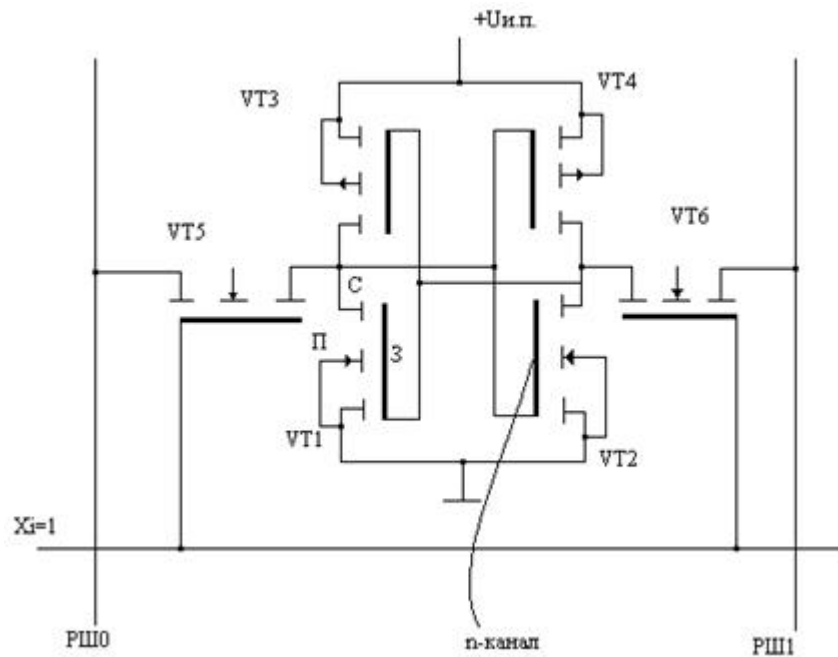
- С прямым доступом (например, жёсткие диски).
- С ассоциативным доступом (специальные устройства, для повышения производительности баз данных).

## 2. Архитектура полупроводниковых ЗУ

Ячейка памяти (блок схема), адресная шина, шина данных, запись, чтение, хранение (регенерация).

N-координатная выборка.

## 3. Статическая 6-транзисторная ячейка памяти

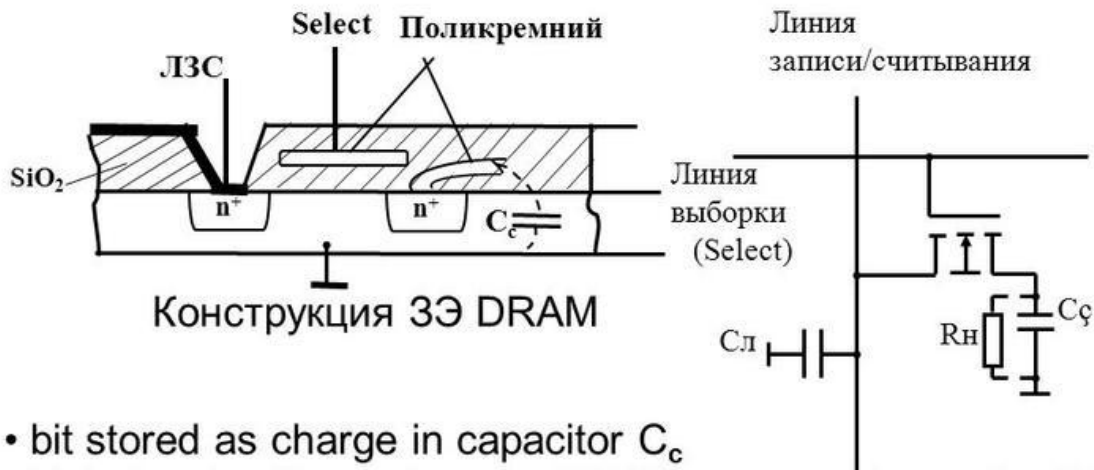


Вариант, когда шины данных подтянуты к «0». На лекции – шины подтянуты к «1».

## 4. Динамическая 4-транзисторная ячейка

## 5. Однотранзисторная ячейка динамической памяти

### Динамическая память



- bit stored as charge in capacitor  $C_c$
- high density (1 transistor for DRAM vs 6 transistors for SRAM)
  - destructive read (capacitor discharge on a read)
  - read is automatically followed by a write (to restore)
  - charge leaks away over time (need to refresh)

MyShared

Так лучше:

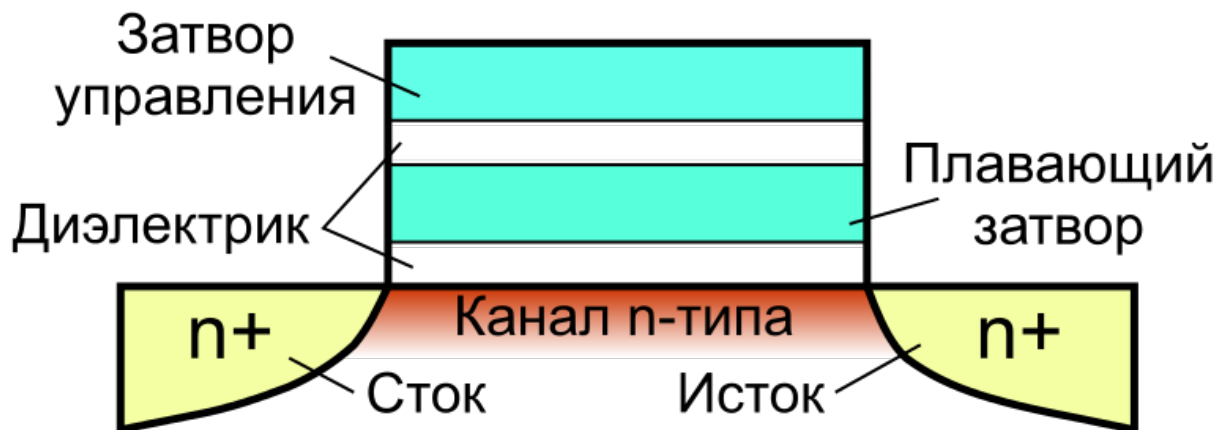


## 6. Флеш-память

**Флеш-память** (*flash memory*) — разновидность полупроводниковой технологии электрически перепрограммируемой памяти (EEPROM). Это же слово используется в электронной схемотехнике для обозначения технологически законченных решений постоянных запоминающих устройств в виде микросхем на базе этой полупроводниковой технологии. В быту это словосочетание закрепилось за широким классом твердотельных устройств хранения информации.

Основным компонентом в флеш памяти является транзистор с плавающим затвором, который является разновидностью МОП-транзисторов. Его отличие в том, что у него есть дополнительный затвор (плавающий), расположенный между управляющим затвором и р-слоем.

В таких транзисторах изменение электрического заряда внутреннего, изолированного слоями диэлектрика затвора производится чисто электрическим - за счет туннелирования электронов и обратимого лавинного пробоя тончайших (порядка нескольких нм) слоёв диэлектрика, обусловленные высокой напряжённостью электрического поля в диэлектрике. При изменении электрического заряда на плавающем затворе изменяется вид вольтамперной характеристики структуры, в частности, изменяется пороговое напряжение. При изменении напряжения на управляющем затворе, считывается состояние транзистора (есть/нет заряд), что позволяет в этой структуре хранить 1 бит информации. Так как заряд плавающего, изолированного от всех электрических цепей затвора сохраняется (при не очень сильных электрических полях в слоях диэлектрика), то микросхемы, построенные на таких структурах, хранят информацию при отключённом напряжении питания.



Другой вариант данной технологии – МНОП:

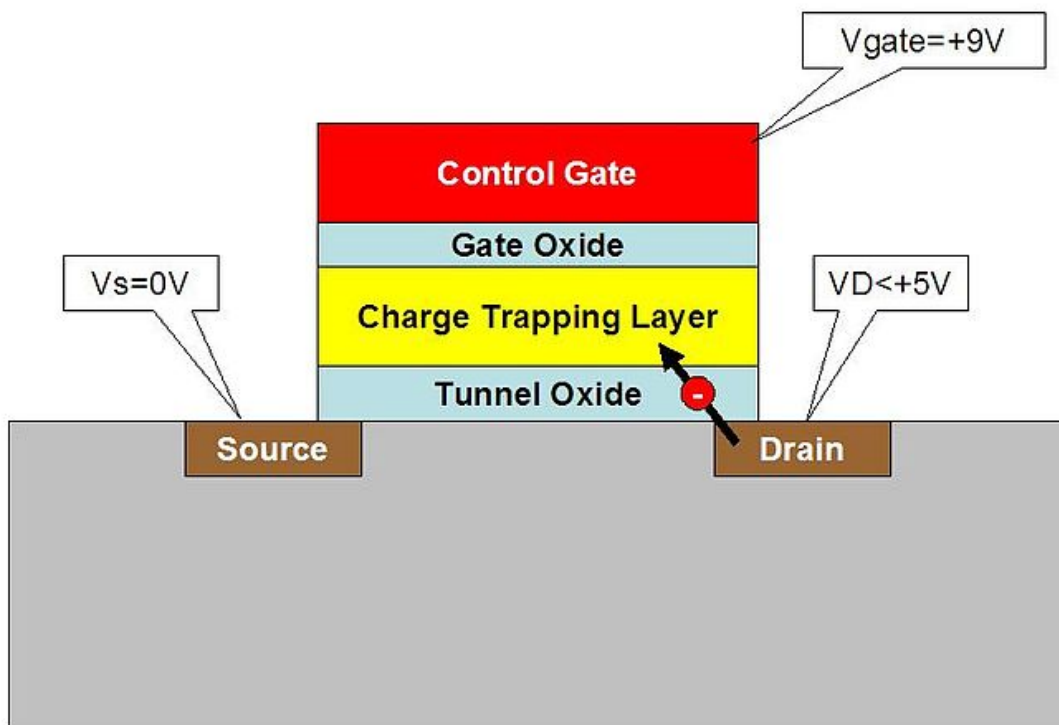
Charge Trap Flash (CTF, память с ловушкой заряда) — технология компьютерной флеш-памяти известная с 1967 года и используемая для создания NOR и NAND накопителей с 2002 и 2008 годов соответственно. Она отличается от широко использовавшейся до 2010 года технологии флеш-памяти на MOSFET-транзисторах с плавающим затвором тем, что использует для хранения электронов пленку нитрида кремния ( $\text{Si}_3\text{N}_4$ ), а не поликремний с допирующими элементами. За счет перехода на CTF производители памяти смогли снизить стоимость производства за счет:

- меньшего количества технологических этапов для формирования ячейки
- возможности использования более тонких технологических процессов (30, 20 нм и немного меньше)
- упрощения хранения нескольких бит в одной ячейке (например, MLC — хранение 2 бит в виде 4 возможных уровней заряда)
- повышения надежности
- более высокого выхода годных, поскольку технология меньше подвержена точечным дефектам в туннельном слое оксида.

Производство флеш-памяти на основе CTF было освоено AMD в партнерстве с Fujitsu ещё в 2002 году. В 2008 году CTF память составляла около 30 % от рынка NOR памяти, общим объёмом в 2,5 млрд долларов США.

Многие производители NAND флеш-памяти перешли с плавающих затворов на CTF в 2008—2010 годах, когда техпроцесс стал приближаться к 20 нм.

Во всех вариантах трехмерной компоновки ячеек флеш-памяти (3D NAND), включая V-NAND (Samsung), применяется CTF.



## SLC- и MLC-приборы

Различают приборы, в которых элементарная ячейка хранит один бит информации и несколько бит. В однобитовых ячейках различают только два уровня заряда на плавающем затворе или слое нитрида кремния. Такие ячейки называют одноуровневыми (single-level cell, SLC). В многобитовых ячейках различают больше уровней заряда; их называют многоуровневыми (multi-level cell, MLC). MLC-приборы дешевле и более ёмки, чем SLC-приборы, однако имеют более высокое время доступа и примерно на порядок меньшее максимальное количество перезаписей.

Обычно под MLC понимают память с 4 уровнями заряда (2 бита) на каждую ячейку. Более дешёвую в пересчёте на объём память с 8 уровнями (3 бита) чаще называют TLC (Triple Level Cell) или 3bit MLC (MLC-3).

Многоуровневая память доминирует на рынке. Тем не менее SLC изделия, несмотря на многократно меньшую ёмкость, продолжают разрабатываться и выпускаться для особо ответственных применений.

### Чтение

Для чтения подаётся положительное напряжение на управляющий затвор. Если в плавающем затворе отсутствует заряд, то транзистор начнёт проводить ток. В противном случае ток между истоком и стоком не возникает. Для MLC ячеек необходимо произвести несколько измерений.

### Запись

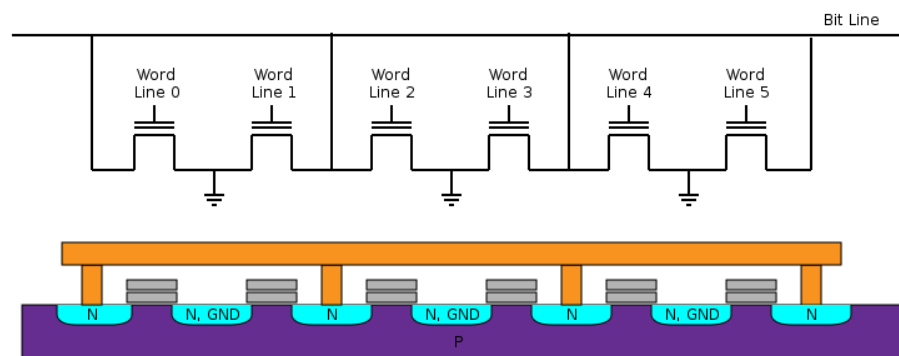
Для записи заряды должны попасть в плавающий затвор, однако он изолирован слоем оксида. Для переноса зарядов может использоваться эффект туннелирования. Для разряда ранее записанного отрицательного заряда необходимо подать большое положительное напряжение на управляющий затвор: отрицательный заряд с помощью туннельного эффекта покинет плавающий затвор. И, наоборот, для заряда плавающего затвора необходимо подать большое отрицательное напряжение (заряд) на затвор.

Также запись может быть реализована с помощью инжекции горячих носителей. При протекании тока между истоком и стоком, при повышенном напряжении, электроны могут преодолевать слой оксида и оставаться в плавающем затворе. При этом необходимо, чтобы на управляющем затворе присутствовал положительный заряд, который создавал бы потенциал для инжекции.

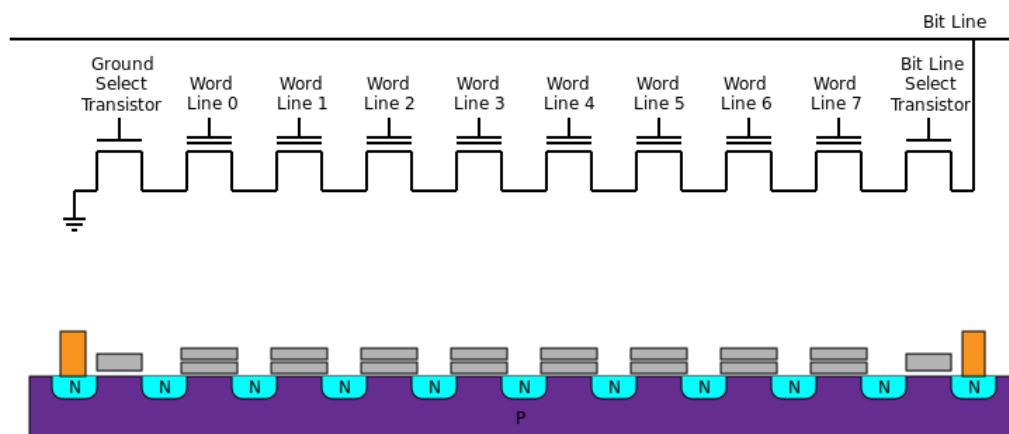
В MLC для записи разных значений используются разные напряжения и время подачи.

Каждая запись наносит небольшой ущерб оксидному слою, поэтому число записей ограничено.

Запись в NOR и NAND компоновке состоит из двух стадий: вначале все транзисторы в линии устанавливаются в 1 (отсутствие заряда), затем нужные ячейки устанавливаются в 0.



NOR flash memory wiring and structure on silicon



NAND flash memory wiring and structure on silicon



Недостатки – ограниченное количество циклов записи, не очень большое время хранения информации (5-15 лет в зависимости от типа).

### Синхронное двухпортовое ОЗУ (Synchronous Dual-Port RAMs)

К отличительным особенностям семейства Synchronous Dual-Port RAMs относятся: синхронный интерфейс с отдельными сигналами синхронизации CLK\_R и CLK\_L и внутренние счетчики (internal counters) для организации пакетного режима передачи данных. Поскольку обязательным условием доступа активных устройств к пространству такой памяти является их взаимная синхронизация от одного системного таймера, никакой дополнительной логики (арбитраж, семафоры или прерывания) для разрешения конфликтных ситуаций не требуется. На рис. 1 показано отличие синхронного двухпортового ОЗУ от асинхронного. Операции обращения к ячейкам асинхронной памяти могут выполняться в произвольные моменты времени при условии соблюдения необходимых временных соотношений между сигналами установки адреса, управления, чтения/записи данных. Операции обращения к ячейкам синхронного двухпортового ОЗУ осуществляются строго под управлением внешнего сигнала синхронизации (CLK\_R для порта R и CLK\_L для порта L).

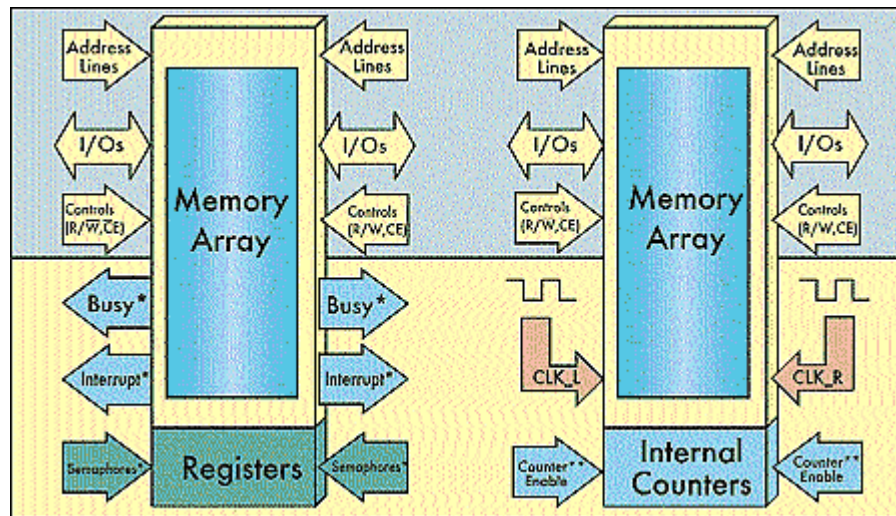


Рис. 1. Отличие синхронного двухпортового ОЗУ от асинхронного

Архитектура синхронного двухпортового ОЗУ оптимизирована для применения в вычислительных сетях (ATM и Ethernet коммутаторы/маршрутизаторы) и системах беспроводной телефонии (базовые станции), обеспечивая следующие синхронные режимы работы памяти: Pipelined (конвейерный), Flow-through (сквозной) и Burst (пакетный). Режимы Pipelined и Flow-through отличаются структурой выходного устройства (рис. 2). В первом случае дополнительный буферный регистр-защелка по выходу (Output register) позволяет организовать конвейерный доступ к данным (одновременно с чтением по предыдущему адресу осуществляется запрос по следующему). Поэтому включение микросхемы двухпортового ОЗУ в режиме Pipelined позволяет сократить общее время обращения к памяти. Единственным недостатком этого режима является задержка на один период сигнала синхронизации при считывании первого слова. Во втором случае (режим Flow-through) считываемые данные непосредственно поступают на выходную шину микросхемы памяти (Output). Это позволяет обеспечить минимальную задержку при считывании первого слова. Однако все последующие обращения к памяти в этом режиме будут проходить за более длительное время, чем в режиме Pipelined. Режимы Flow-through и Pipelined задаются пользователем аппаратно.

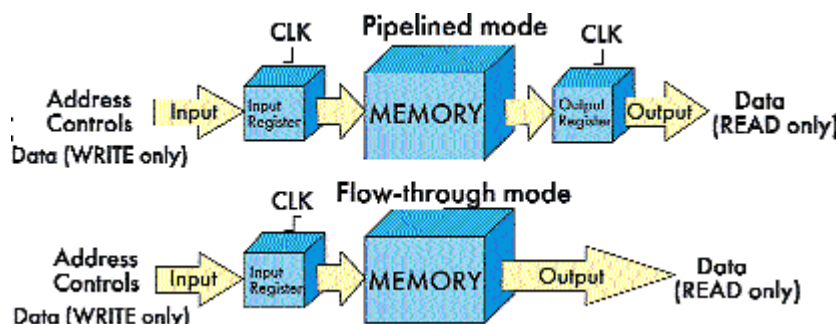


Рис. 2. Режимы Pipelined и Flow-through

Режим Burst предназначен для выполнения операций над последовательными потоками параллельных данных (например, потоками речевых сообщений) и имеет некоторое сходство с



## **Технология. Факты последних лет.**

### **Литография.**

Базовый вариант - аргонно-фторидный лазер. Длина волны в этом случае составляла всего 193 нм. Для того чтобы использовать еще более короткие волны (например, глубокий ультрафиолет) весь литографический процесс необходимо проводить в вакууме. Причина этого заключается в том, что воздух поглощает все излучение с длиной волны меньше 186 нм. На практике даже излучение с длиной волны 193 нм, которое вырабатывается аргонно-фторидным лазером, отчасти поглощается воздухом.

### **Иммерсионная литография**

Еще одним способом увеличения разрешающей способности является иммерсионная литография. Разные производители начали использовать эту технологию в разное время. Это произошло в период между выпуском первых 65-нм и 32-нм чипов. Иммерсионная литография является относительно простым процессом. Его смысл заключается в том, что воздушное пространство между последней линзой и пленкой фоторезиста заполняется жидкостью с показателем преломления больше единицы. Благодаря этому повышается числовая апертура оптической системы, потому что свет от источника лучше рассеивается. В качестве иммерсионной жидкости, как правило, используется высокоочищенная (деионизированная) вода.

С появлением фотолитографии в глубоком ультрафиолете (EUV, Extreme ultraviolet lithography) использовать метод иммерсионной литографии стало невозможным, поскольку используемые жидкости поглощают всю выделяемую энергию.

### **Мультипаттернинг**

А вот технологию мультипаттернинга (Multiple Patterning) использовать вместе с фотолитографией в глубоком ультрафиолете уже возможно. При этом существует далеко не один способ ее реализации. Суть мультипаттернинга заключается в следующем. К примеру, если определенная система умеет печатать линии на расстоянии 64 нм друг от друга за один проход, то во время второго прохода она сможет нанести еще две такие же линии с таким же интервалом, причем одна из них будет располагаться между линиями первого прохода. Как итог, всего будет нанесено 4 линии на расстоянии 32 нм друг от друга.

Одним из основных способов мультипаттернинга является двойное формирование рисунка LELE (Litho-Etch, Litho-Etch). В процессе LELE используется два отдельных фоторезистивных слоя. Во время первого прохода литография производится на твердой маске, после чего следует проявление. Очень важно использовать именно твердую маску, так как во время второго прохода используется еще один слой фоторезиста, который впоследствии также экспонируется и проявляется. Если бы не твердая маска, то после применения второго слоя фоторезиста от первого рисунка ничего бы не осталось. Что касается технологии LFLE (Litho-Freeze, Litho-Etch), то она является своего рода модификацией LELE. В LFLE не используется твердая маска. После начального литографического процесса фоторезист замораживают, покрывая его определенным химическим веществом, а затем проявляют. Тем самым первый резистивный слой отделяется от второго. После этого применяется второй литографический проход, и процесс завершается.

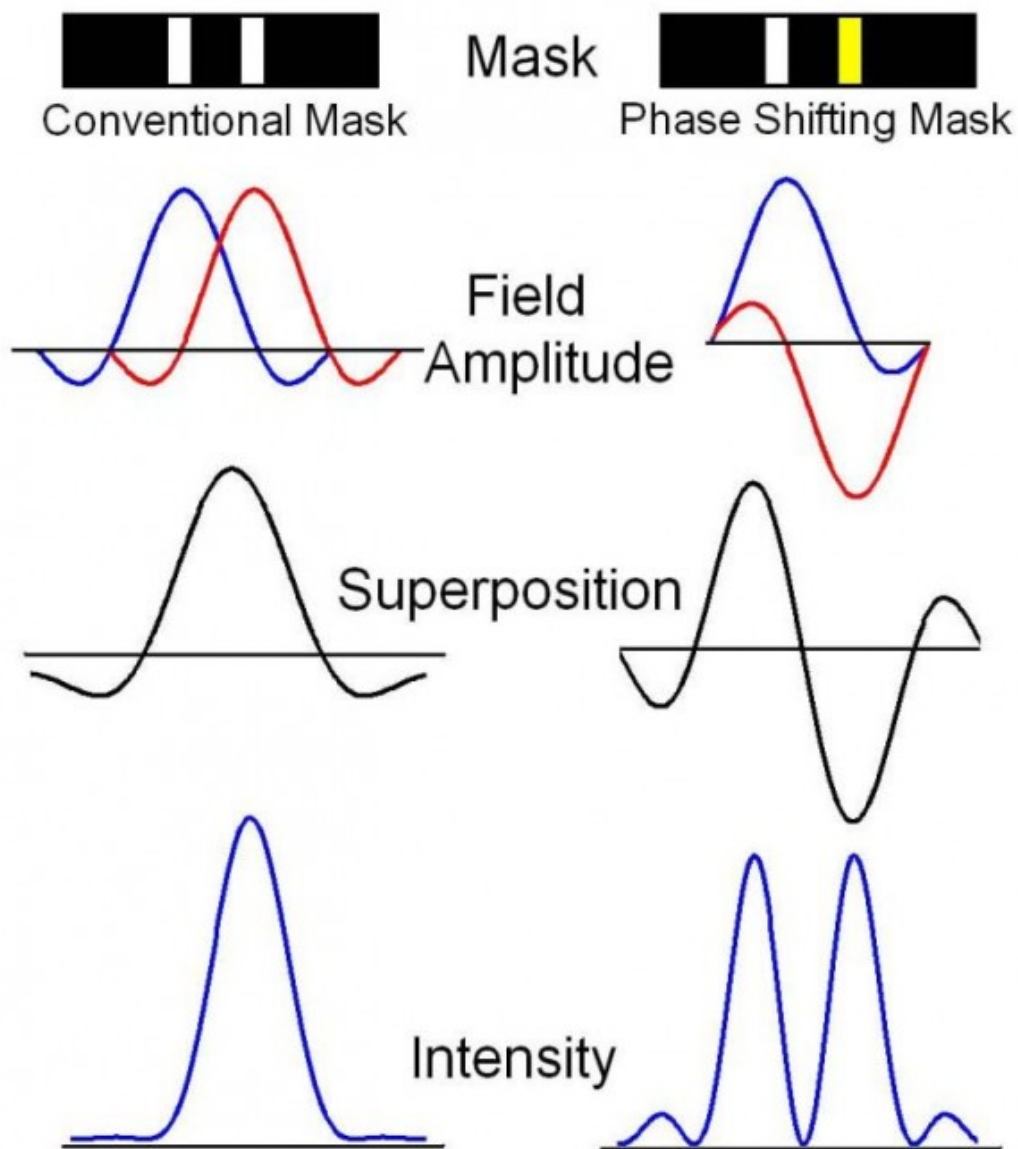
Помимо LELE и LFLE, применяется способ под названием SADP (self-aligned double patterning) — двойное формирование рисунка с самовыравниванием. Этот способ также состоит из двух этапов. На первом шаге происходит травление временных шаблонов, которые впоследствии становятся линиями травления в финальном шаблоне. После этого временные шаблоны заполняются твердой маской, к которой несколько позже применяется процесс травления для экспонирования боковых стенок линий временного шаблона. Затем временный шаблон удаляется, а к экспонированному оксиду также применяется процесс травления. В результате получаются линии, расположенные вдвое ближе друг к другу, причем результат достигается только за один литографический проход.

В названии технологии SADP не зря указано слово «самовыравнивание». Благодаря тому, что после наложения временного шаблона нет повторного процесса экспонирования, беспокоиться о выравнивании линий не приходится. Вдобавок ко всему, из-за того, что боковые стенки создаются с помощью твердой маски, результирующие линии получаются максимально ровными и точными.

Мультипаттернинг может показаться идеальным способом дальнейшего увеличения разрешающей способности, однако это не совсем так. Главный его недостаток (и весьма существенный) — его дороговизна. Даже из-за самого минимального несоответствия выравниваний между двумя шаблонами пластина приходит в негодность.

### **Вычислительная литография**

Вычислительная литография представлена двумя основными методами: технологией фазосдвигающих масок (phase shifting masks, PSM) и оптической коррекцией близости (optical proximity correction, OPC). Суть первой заключается в коррекции толщины отдельных пикселей маски для изменения их прозрачности, что изменяет фазу проходящего сквозь них света. В результате достигается больший контраст между экспонированными и неэкспонированными зонами, что увеличивает разрешающую способность.



Сравнение фазосдвигающих и обычных масок

Процесс оптической коррекции близости кардинально отличается от технологии фазосдвигающих масок. По сути, эти два метода вообще не связаны между собой. Одним из самых больших недостатков литографии является то, что рисунок с фотомаски переносится на фоторезист с искажениями. Например, ширина полученных линий во многом зависит от плотности шаблона около этих самых линий. В итоге результирующие полосы, как правило, получаются более узкими и располагаются не совсем так, как на маске. Технология OPC как раз отвечает за коррекцию искажений. С ее помощью создается такая фотомаска, чтобы в итоге все элементы располагались на нужных позициях.

## Архитектура транзистора.

### От SOI к FINFET

По мере того, как уменьшаются размеры транзисторов, факторы, которые не имели особого значения в прошлом, становятся все более значимыми. Одной из самых больших проблем является ток утечки.

Прежде всего, нужно отметить, что поведение электронов больше похоже на поведение системы из квантовой механики, нежели из классической механики. Это приводит, например, к туннельному эффекту (quantum tunneling), когда электроны проходят через изолирующие слои. И если раньше туннельный эффект не оказывал столь значительного влияния, то с переходом на более тонкие техпроцессы проблема постоянно возрастала.

Всего существует пять эффектов короткого канала. Наибольший интерес для полупроводниковой индустрии представляет DIBL-эффект (drain-induced barrier lowering). Суть этого эффекта заключается в том, что из-за небольшой длины канала напряжение, которое прикладывается к стоку, может оказывать влияние и на исток.

Это означает, что эффективность затвора в плане контроля канала и протекающего тока значительно снижается. При этом сдвигается и уровень порогового напряжения.

Решить эту проблему пытались путем легирования определенных областей между истоком и стоком, однако нужного эффекта это не принесло. Почему? Во-первых, в легированных областях был довольно большой разброс электронов, и сами частицы становились менее подвижными. Во-вторых, легирование привело к тому, что пороговое напряжение у каждого транзистора довольно сильно различалось.

Так или иначе, в результате DIBL-эффекта резко падает производительность. Для предотвращения действия эффектов короткого канала существует несколько технологий: растянутый кремний (straining silicon), кремний на изоляторе SOI (silicon on insulator), металлический затвор с высоким значением диэлектрической константы (high-k metal gate) и FINFET.

### Кремний на изоляторе (SOI)

Суть метода SOI (кремний на изоляторе) заключается в использовании трехслойной подложки со структурой кремний-диэлектрик-кремний вместо обычной монокристаллической кремниевой пластины. Этот метод имеет свои преимущества и недостатки.

Прежде всего, стоит сказать, что при использовании технологии SOI удается значительно сократить паразитную емкость. Проблема паразитной емкости состоит в том, что на нее затрачивается достаточно большая часть потребляемой мощности в момент переключения транзистора из одного состояния в другое. А время, затраченное на этот переход, как раз и определяет общее быстродействие. Добавление слоя оксида кремния под транзистор позволяет сократить паразитную емкость, а значит уменьшить и время, которое затрачивается на переключение транзистора.

Главным недостатком SOI, как и в случае с мультипаттернингом, является высокая стоимость. Кроме этого, из-за того, что транзисторы располагаются на изолирующем слое, рассеивание выделяемого им тепла не столь эффективно, как в случае с монокристаллическим кремнием. Кстати, это стало причиной тому, что компания AMD вернулась к использованию монокристаллического кремния.

### Растянутый кремний



Если метод SOI утратил былые позиции, то технология растянутого кремния, наоборот, становится популярнее. Путем помещения германий-кремния (SiGe) или карбида кремния (SiC) в исток и сток кремний в канале растягивается, эффекты электростатических сил в нем сокращаются. Носители заряда становятся более подвижными, что увеличивает производительность транзистора. Существует далеко не один способ реализации технологий, но смысл остается тем же.

Впервые технология использовалась компанией Intel при производстве 90-нм процессоров в 2003 году.

### **Металлические затворы и материалы с высоким показателем диэлектрической константы**

Применение металлических затворов и материалов с высоким показателем диэлектрической константы началось с 45-нм процессоров Intel Penryn, которые увидели свет в 2007 году. Почему же эти технологии так важны? Тут опять необходимо вернуться к квантовой механике. Уменьшая расстояние между пластинами конденсатора, увеличивается его емкость. При этом эффект поля усиливается, и транзистор переключается быстрее. Однако при малой толщине случается так, что электроны начинают туннелировать из затвора в кремниевый канал, минуя на своем пути изолятор. Это приводит к тому, что токи утечки находятся на достаточно высоком уровне, даже если транзистор не переходит из одного состояния в другое. Поэтому уменьшать толщину еще больше становится невозможно.

Для того чтобы это все-таки было возможно, и применяются high-k диэлектрики. Обратной стороной медали является то, что использование таких материалов приводит к определенным сложностям в производстве. Возникают проблемы с формированием каналов, поэтому ради увеличения производительности транзистора необходимо использовать металлические затворы.

### **FINFET**

С переходом на 22-нм техпроцесс влияние эффектов короткого канала еще больше увеличилось. И хотя с помощью металлических затворов и high-k диэлектриков удастся уменьшить токи утечки и улучшить производительность транзистора, остается проблема контроля за каналом со стороны затвора, поскольку его размеры относительно подложки постоянно уменьшаются. Исправить это могут Tri-Gate транзисторы, в которых структура канала немного видоизменена и похожа на плавник (fin). Увидеть отличия в дизайне Tri-Gate от обычного транзистора можно на картинке ниже.

# 22 nm Tri-Gate Transistor

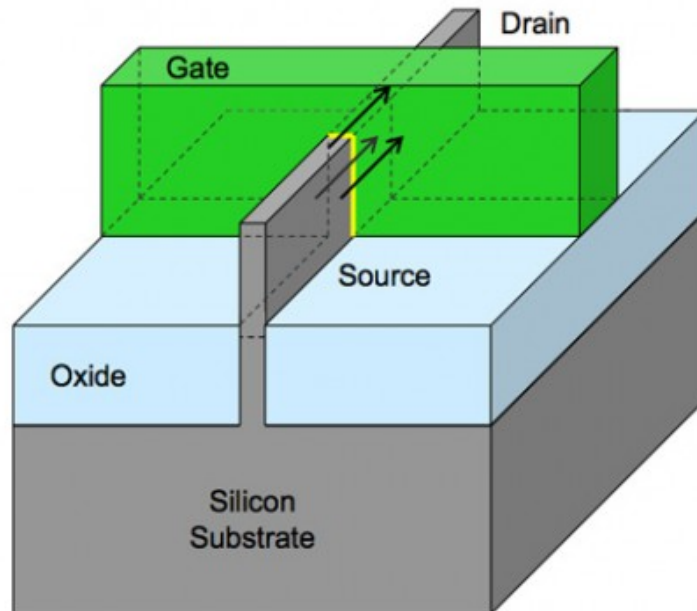


Схема Tri-Gate транзистора

Основным плюсом Tri-Gate транзистора является то, что влияние DIBL-эффекта значительно снижается. Это также означает, что больше нет необходимости легировать канал. Следовательно, пороговые напряжения у каждого транзистора будут одинаковыми. Кроме этого, увеличивается и производительность транзистора за счет более быстрого перехода между состояниями.

## Перспективы.

Что ждет полупроводниковую индустрию в ближайшем будущем. Скорее всего, производство 10-нм процессоров станет лебединой песней для 193-нм иммерсионной литографии. Огорчаться не стоит, поскольку технология и так используется намного дольше, чем предрекали специалисты. Что может прийти ей на смену? Это литография в глубоком ультрафиолете, нанопечатная литография (nanoimprint lithography) и электронная литография (electron-beam lithography).

## Литография в глубоком ультрафиолете

В отличие от иммерсионной литографии, этот метод использует излучение с длиной волны около 13,5 нм. Столь короткие волны и порождают массу проблем. Во-первых, источники глубокого ультрафиолета не являются лазерами. Это значит, что энергоэффективность такого источника света во много раз ниже, чем у современных лазеров. Поэтому для генерации необходимой энергии потребуются очень мощные источники. Во-вторых, воздух поглощает глубокий ультрафиолет, поэтому литография должна осуществляться в вакууме. Кроме того, процесс производства в вакууме не столь быстрый, как в воздухе. За один час в вакууме



производится значительно меньше пластин, нежели в воздушном пространстве. Вдобавок ко всему в вакууме невозможно использовать линзы — их заменяют зеркалами.

Наконец, глубокий ультрафиолет — своего рода ионизирующая радиация. Это значит, что когда молекула поглощает EUV-свет, то ее электроны могут высвободиться. К сожалению, это снижает разрешающую способность, поскольку фотоэлектроны двигаются через фоторезист случайным образом. В целом все недостатки литографии в глубоком ультрафиолете снижают ее разрешение примерно до 15-19 нм. А для того, чтобы достичь разрешающей способности хотя бы 10 нм, требуется двойной паттернинг, о котором шла речь выше.

### **Нанопечатная литография**

Нанопечатная литография является достаточно простым методом. Здесь изображение образуется за счет механической деформации полимерного покрытия штампом, после чего к деформированному покрытию применяется травление. Казалось бы, ничего сложного, но и здесь есть свои нюансы. Во-первых, нанопечатная литография, как и технология EUV, тоже должна выполняться в вакууме. Во-вторых, после многократного использования штампа его разрешающая способность снижается. Ну и в-третьих, для производства самого штампа требуется очень точная литография.

### **Электронная литография**

Электронная литография — это метод нанолитографии с использованием электронного пучка. Главным ее преимуществом является то, что она позволяет достичь невероятно высоких разрешений (менее 10 нм). С помощью электронной литографии даже можно переносить шаблон напрямую на кремниевую пластину, минуя этап с фоторезистом.

Увы, но основным недостатком технологии является ее низкая производительность. Электронная литография позволяет получать всего около десяти кремниевых пластин в час. Для сравнения: большинство других современных методов литографии имеют производительность более 100 пластин в час. Для того чтобы понять, насколько медленным является электронная литография, приведем пример. Используя один-единственный электронный пучок, литография одной 300-миллиметровой пластины заняла бы порядка 22 лет! Конечно, для ускорения процесса можно одновременно использовать тысячи пучков. Однако в данном случае одни электроны будут влиять на траектории других. Поэтому здесь необходимы сложные предварительные расчеты.