Флаш памети – NOR и NAND флаш. Сравнителен анализ.

I. Въведение

Флаш паметта е тип EEPROM (Electrically Erasable Programmable Read-Only Memory) – енергонезависима памет, която се изтрива и програмира по електрически път. Тази памет намира широко приложение в редица устройства: мемори карти, USB флаш памети, цифрови камери, мобилни телефони, цифрови плейъри, лаптопи и др. Флаш паметта е измислена от д-р Fujio Masuoka докато работи за Toshiba през 1980 г. Според Toshiba, името "флаш" е предложено от колегата на д-р Masuoka, г-н Шоджи Ариууми, защото процесът на изтриване на съдържанието на паметта му напомня за светкавицата на фотоапарат. Д-р Masuoka представя изобретението си през 1984 г. на IEEE International Electron Devices Meeting (IEDM), проведено в Сан Франциско, Калифорния. Първите NOR флаш памети са изобретени през 1984 година от сътрудници на фирма Toshiba. Toshiba анонсира NAND flash паметите на IEEE IEDM през 1987г.

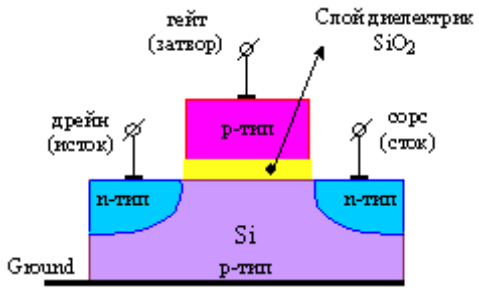
Всяка запомняща клетка от флаш паметта може да съхранява информация за един или няколко бита. В зависимост от това те се делят на:

- Single-Level Cell (SLC) – всяка запомняща клетка съхранява 1 бит информация (лог. 0 или лог. 1).
- Multi-Level Cell (MLC) – всяка запомняща клетка съхранява повече от 1 бит информация, най-често 2 (00, 01, 10, 11). Тези памети памет имат по-голям обем при една и съща площ на кристала, но имат по-бавен достъп и по-малко полезни работни цикли на изтриване (около 10,000);

Всяка запомняща клетка на флаш паметта е MOS транзистор с плаващ гейт (FGMOS).

II. MOS транзистори

При интегралните схеми, основния компонент на базата на който се изгражда всяка логика, е MOS транзистора. Той има три извода - гейт, сорс и гейт (виж Фиг. 1)

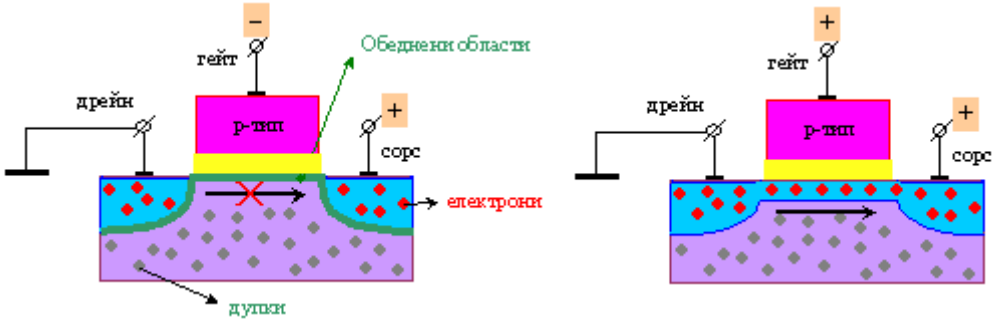


Фиг. 1. *Структура на MOS транзистор*

MOS транзисторът най-често се използва като ключов елемент. Транзисторът има две възможни състояния – отпушено и запушено (отворено, затворено). Става дума за канала между дрейна и сорса, по който протича (или не протича) ток. Състоянието на канала се управлява чрез подаване на подходящо напрежение на гейта. Транзисторът е запушен, когато съпротивлението между дрейна и сорса е много голямо (десетки MΩ) – няма проводящ канал между дрейна и сорса. Транзисторът е отпушен когато се изгради

проводящ канал между дрейна и сорса. В този случай съпротивлението на канала е малко (части от ома). Съществуват два вида MOS транзистори в зависимост от това каква е проводимостта на канала: транзистори с *n* канал и транзистори с *p* канал. При транзисторите с *n* канал проводимостта се формира от електрони (отрицателен заряд), а три транзисторите с *p* канал – от дупки (положителен заряд). Самият транзистор е технологично изграден върху полупроводникова подложка с *p* проводимост. Гейтът се намира между сорса и дрейна и е отделен от тях с много тънък слой диелектрик от SiO_2. Дебелината на този диелектрик е не повече от 90nm (10^{-9} m).

На Фиг. 2 е показано какви трябва да бъдат захранващите напрежения, за да се запуши или отпуши един MOS транзистор от *n* тип.



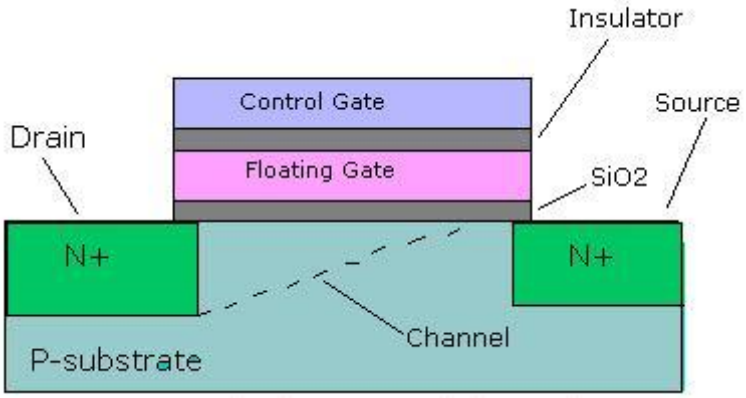
Фиг. 2. *Работа на MOS транзистора като ключ: а) запушен; б) отпушен*

При свързването, показано на Фиг 2а, на сорса се подава положително напрежение (спрямо маса), маса на дрейна и отрицателно напрежение на гейта. При това захранване не се получава проводящ канал между дрейна и сорса. Причина за това е дифузията на дупки от *p* гейта и обратната дифузия на електрони в *p* гейта. Това не позволява протичане на ток между дрейна и сорса и следователно MOS транзисторът е запушен.

При свързването, показано на Фиг 2б се подава положително отпушващо напрежение на гейта. Дупките се изтласкват в дълбочина на силициевата подложка и на тяхно място се натрупват електрони от дрейна. Така се получава проводящ канал от електрони между дрейан и сорса и транзисторът се отпушва.

III. MOS транзистори с плаващ гейт

Ще разгледаме SLC транзистор с плаващ гейт, който запомня един бит информация (логическа единица или логическа нула). Плаващият гейт е изграден в *p* проводимата (наситена с дупки) област на управляващият гейт (виж Фиг. 3). Самият плаващ гейт е нанесен върху много тънък (около 10nm) слой диелектрик (SiO_2). Този допълнителен гейт се използва за съхранение на електрони. Електроните в плаващият гейт могат да се съхраняват ограничен, но достатъчно дълъг период от време. Ако в него няма съхранени електрони, то при подаване на отпушващо напрежение към управляващият гейт, транзисторът ще се отпуши (това състояние съответства на лог. 0). Ако в него има съхранени електрони, то транзисторът ще остане запушен (това състояние съответства на лог. 1). Лесно може да се разбере дали има електрони в плаващия гейт или няма – подава се отпушващо напрежение на управляващия гейт и се проверява дали е протекъл ток между дрейна и сорса.



Фиг. 3. *MOS транзистор с плаващ гейт*

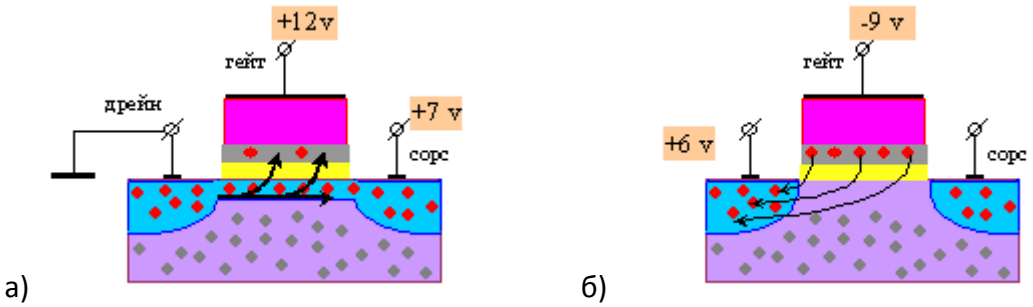
Записът на информация в клетка от флаш се състои в натрупване на отрицателен заряд върху плаващия гейт. То се реализира или чрез метода на инжекция на горещи електрони или чрез тунелиране на електрони (базира се на тунелния ефект на Fowler-Nordheim).

При метода на инжекция на горещи електрони, на сорса и на управляващия гейт се подава високо напрежение, което е необходимо за ускоряване на електроните до такава степен, при която те преодоляват потенциалната бариера на тънкия изолационен слой и достигнат до плаващия гейт. Потенциалната разлика в напреженията трябва да е достатъчно висока, така че между дрейна и сорса да протече ток не по-малък от 1mA.

При тунелния ефект се използват вълновите свойства на електрона. При него не се налага използването на висока потенциална разлика, за да се получи натрупване на електрони в плаващия гейт в резултат на което запомнящите клетки имат по-малки размери. Програмирането при този метод обаче е по-бавно.

Изтриването на информация от клетка на флаш памет се свежда до премахване на натрупания в плаващия гейт заряд. За целта на управляващия гейт се подава високо отрицателно напрежение, а на дрейна - положително. С това се създава силно обратно електростатично поле и електроните мигрират от плаващия гейт към дрейна.

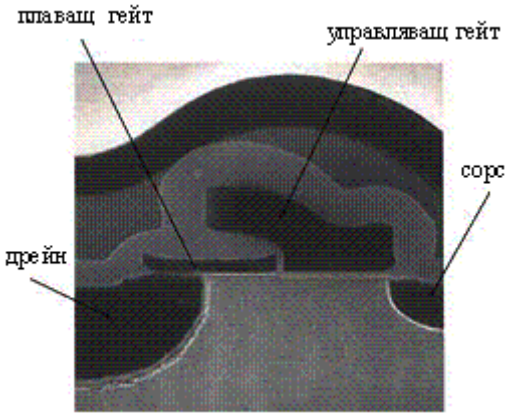
Следователно, при запис във флаш запомняща клетка се запомня лог. 1, а при изтриване – лог. 0 (виж. Фиг. 4).



Фиг. 4. *Програмиране на флаш памет: а) запис (лог. 1); б) изтриване (лог. 0)*

Основният проблем при флаш е свързан с мащабируемостта на паметта. Поради сравнително високите програмиращи напрежения се налага по-голямо отстояние между

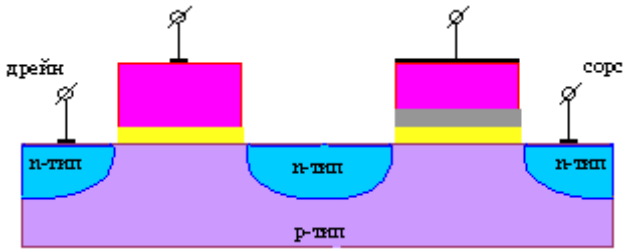
свързващите интегрални проводници. Това води до необходимост от голяма площ на една запомняща клетка. Този недостатък се отстранява частично технологично - гейтовете на флаш транзистора имат специална форма и разположение, както е показано на Фиг. 5.



Фиг. 5. *Рентгенова снимка на реална флаш клетка*

Извитата форма на гейта и припокриването на плаващия гейт с управляващия гейт позволява намаляване на програмиращите напрежения. Мигрирането на електроните от плаващия гейт се реализира чрез подаване на положително напрежение на управляващия гейт. Електроните тунелират не към дрейна, а към управляващия гейт, за което спомага и извитата форма на двата гейта. За натрупване на заряд върху плаващия гейт към дрейна и към управляващия гейт се подава положително напрежение, а към сорса – маса. В управляващия гейт се формира проводящ канал, а напрежението между дрейна и сорса ускорява електроните така, че те преодоляват потенциалната бариера и правят пробив към плаващия гейт.

Друг начин за подобряване на мащабируемостта е показан на Фиг. 6.



Фиг. 6. *Двутранзисторна флаш клетка*

При него за всяка клетка от флаш се използват два транзистора – един обикновен MOS транзистор и един MOS транзистор с плаващ гейт. Обикновенния MOS трянзистор се използва за изолация на транзистова с плаващ гейт от линията за данни. Тъй като се използва тунелния ефект при запис се работи със значително по-ниски напрежения. Двутранзисторният флаш елемент е в основата на NAND флаш паметта.

Срокът за съхранение на информацията трябва да се има предвид, тъй като той не е безкраен. Той е до 20 години. Причина за това е много тънката изолация на плаващия гейт. Външни фактори като температура и най-вече радиация оказват влияние на срока на съхранение на информация.

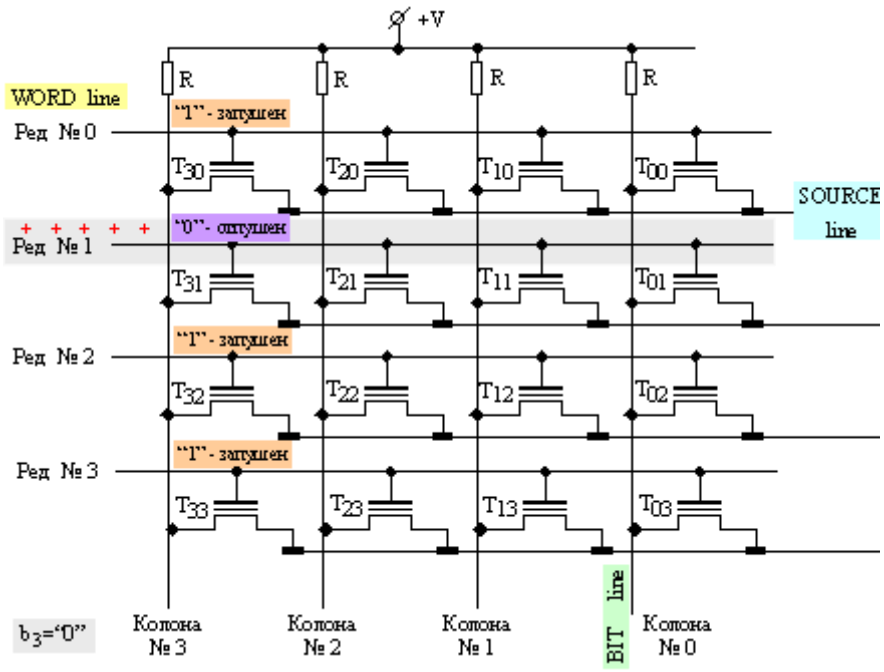
Броят на записите също е ограничен като брой. Тази стойност най-често е до 100,000 за SLC флаш и до 10,000 за MLC флаш. Основната причина за това е свързан с това че записът и изтриването се изпълнява върху множество клетки едновременно (блоков режим). С времето зарядите в отделните клетки се разсъгласуват и в определен момент е възможно някой от зарядите да излезе от допустимите граници. Друга причина, влияеща на броя на записите, е взаимната дифузия на атомите от изолираните и проводящите области в общата полупроводникова структура. Дифузията се повишава с времето поради множеството пробиви на изолатора при запис и изтриване.

IV. Флаш архитектурни решения

Двата основни типа флаш памет са NOR flash и NAND flash. Intel е първата компания, която въведе търговски NOR флаш чип през 1988 г., а Toshiba пуска на пазара първата в света NAND флаш памет през 1989 г. Имената NOR и NAND следват от структурата, използвана за формиране на връзките между клетките на паметта.

4.1. NOR флаш памет

NOR флаш паметите са изградени от еднотранзисторни елементи подредени в матрица. Те образуват клетки и блокове. На Фиг. 7 е показана структурата на NOR флаш памет с организация на матрицата 4x4.



Фиг. 7. *Структура на NOR флаш памет*

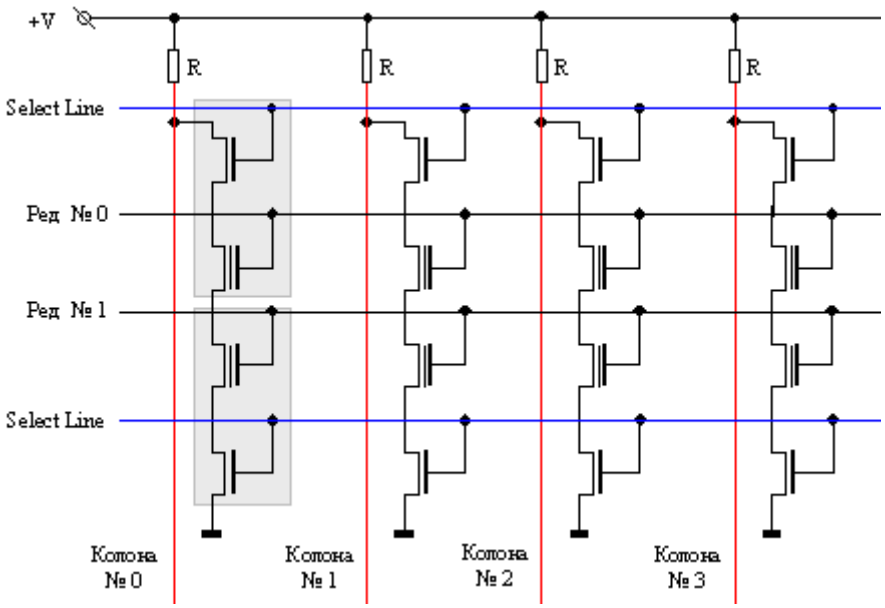
Запомнящите елементи се управляват по три електрода – word линия, source линия и bit линия. Всички елементи от една клетка се управляват едновременно. Техните гейтове са свързани към общ проводник. Това означава, че за достъп до дадена клетка от паметта е нужно да се подаде положително напрежение на word линията на целия ред в който тя се намира (виж ред № 1). Едновременно се четат 8, 16 или 32 бита информация (в случая 4). Информацията в дадения бит се прочита по съответната bit линия. Сорсовете на транзисторите са свързани към маса.

Схемата реализира логическа операция ИЛИ-НЕ (NOR). При тази операция се получава лог. 1 само в случай, че на всички входове има лог. 0. Във всички останали случаи резултатът е лог. 1. Логическа операция ИЛИ се формира чрез свързване на дрейновете на всички транзистори от колоните към захранващо напрежение през общо съпротивление R. Инвертирането се реализира от самия MOS транзистор. При NOR паметта word линиите са адресни, а по bit линиите се четат данните. Една bit линия ще върне лог. 1, само ако всички транзистори в колонката са запушени. При избор на ред № 1, транзисторите от реда стават достъпни. Тези транзистори, които нямат натрупани електрони в плаващия гейт ще върнат лог. 0, тъй като ще се отпушат. Останалите ще върнат лог. 1, тъй като ще останат запушени. Тези данни ще се прочетат чрез bit линиите (колонки № 0, 1, 2 и 3).

При NOR паметта операция четене се реализира бързо, тъй като всички битове се прочитат едновременно. Операции запис и изтриване са по-бавни. Запис на лог. 0 в даден запомнящ елемент се осъществява чрез едновременно подаване на положително напрежение на линия word и високо положително напрежение на данновата линия през съпротивлението R, с което се осигурява отрицателна потенциална разлика между гейта и сорса. В резултат на това електроните, натрупани в плаващият гейт, мигрират към дрейна. При следващо четене този транзистор ще се отпуши и ще върне лог. 0. При изтриване, всички клетки в един блок на паметта се изтриват едновременно. Основното предимство на NOR флаш паметите е, че можем да прочетем коя да е клетка от паметта. Това е така, защото памаетта позволява адресиране на произволна клетка. Следователно, тази памет ще позволи изпълнение на програмен код директно от флаш, без да е необходимо преместването му в RAM.

4.2. NAND флаш памет

При NAND флаш паметите транзисторите са свързани последователно в колони, както е показано на Фиг. 8. Това определя основното предимство на NAND паметите – те имат много по-добра мащабируемост при сравнение с NOR паметите, тъй като при един и същ капацитет на паметите запомнящите клетки заемат по-малка площ върху подложката.



Фиг. 8. *Структура на NAND флаш памет*

С NAND паметта се работи в блоков режим. Всеки блок се състои от множество страници. Страницата обединява няколко групи запомнящи клетки (най-често 16). Следователно адртесирането на произволна клетка на паметта е невъзможно. Операция запис е за цяла страница, а операция изтриване – за цял блок от паметта (или група от блокове).

Схемата реализира логическата операция И-НЕ (NAND). При тази операция се получава лог. 0, само ако всички входни данни са лог. 1 . При всички останали случаи резултатът е лог. 1. Този ефект се дължи на последователното свързване на транзисторите.

Гейтовете на транзисторите от даден ред са свързани към съответната word линия. Включването на последователно свързваните транзистори от всяка колона към захранващо напрежение се реализира чрез линии select. Те управляват обикновени MOS транзистори, които „включват" колоните към маса и захранващото напрежение +V. За да се осъществи достъп до избран бит, първо се подава положително напрежение на съответната word линия – активират се всички транзистори от дадения ред. След това е нужно транзисторите от другите редове да бъдат в такова състояние че да пропускат електричество по начин независим от техните стойности. Така при четене на данни безпроблемно протича електричество по цялата серия последователно свързани транзистори, до достигане на избрания бит за четене. По този начин информацията от избраните битове за четене се появява на bit линията (колона № 0, 1, 2 или 3). Ако избрания транзистор е запушен се получава лог. 1, а когато е отпушен – лог .0.

Страниците най-често са с размер 512B, 2KiB или 4KiB. За всяка страница няколко байта (в зависимост от размера на страницата) са служебни и се използват с цел откриване и отстраняване на грешки - Error Correcting Code (ECC). Ако в страницата има грешни данни, те могат да бъдат коригирани чрез ECC данните. Размерът на блоковете най-често е 16KiB, 128 KiB, 256KiB или 512KiB.

Основното предимство на NAND флаш паметите е много добрата мащабируемост, която гарантират. Недостатъците на тези памети са свързани с блоковия режим на работа. Той не позволява изпълнение на програмен код директно от флаш паметта.

Основното приложение на NAND паметите е свързано с така наречените интегрални дискове - Solid State Drive (SSD) като алтернатива на трърдите дискове. Благодарение на специфични архитектурни решения е възможно получаване на дискове с капацитет от няколко стотин GiB до няколко TiB.

V. Въпроси и задачи за самостоятелна работа

1. Каква е разликата между технологии „горещо инжектиране на електрони" и „тунелен ефект", използвани при флаш паметите?

2. Анализирайте колко точно байта се заделят за ECC данни при конкретна флаш памет.

3. Къде в момента намират приложение NOR и къде NAND флаш паметите?

4. Използвайте специализирана литература и анализирайте какво представляват така наречените V NAND флаш памети.