INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato inovace předmětu *Hardware a komunikační technologie* je spolufinancována Evropským sociálním fondem a Státním rozpočtem ČR, projekt č. CZ.1.07/2.2.00/28.0014, „Interdisciplinární vzdělávání v ICT s jazykovou kompetencí“.

Flash paměti a SSD (Solid State Drive)

Šárka Vavrečková

Poslední aktualizace: 16. dubna 2015

Obsah

1 Vlastnosti a struktura NAND flash pamětí

SSD (Solid State Drive) je vnější paměť typu NAND flash (tj. elektronická paměť) vybavená vhodným řadičem. Co to znamená?

- NAND flash paměti jsou permanentní (tj. po odpojení od el. proudu jejich obsah zůstává),
- rychlost čtení je typicky vyšší než rychlost zápisu, nicméně obojí je vyšší než u klasického HDD disku,
- spotřeba energie se v průběhu činnosti moc nemění (narozdíl od HDD, který spotřebovává energii jen tehdy, když se plotny otáčejí), obvykle bývá nižší než u HDD,
- životnost paměťových buněk není neomezená, závisí na několika faktorech (viz dále),
- na kvalitě řadiče závisí jak životnost, tak i přístupové doby pro čtení a zejména zápis.

Pozor, nejde o „disk“, nic kruhového na tomto zařízení není.

1.1 Paměťové buňky

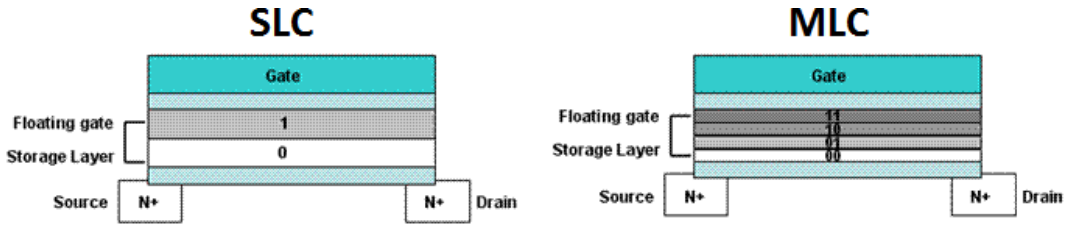
V paměťových buňkách flash pamětí (tj. nejen SSD, ale i třeba USB flash disků) se data ukládají ve formě elektrického náboje.

Pokud do takové paměťové buňky chceme uložit údaj, je možné to provést tak, že v buňce bude rozpoznáván určitý počet *úrovní napětí*. Například pokud má být buňka schopna uložit hodnotu jednoho bitu (tedy máme dvě možnosti: buď tam bude uložena hodnota 0, a nebo hodnota 1), pak buňka musí umožnit rozpoznání dvou úrovní napětí - jednu úroveň pro hodnotu 0, druhou pro hodnotu 1. Podle toho, jaká úroveň je v buňce indikována (nízká nebo vysoká), poznáme při čtení buňky, zde je v ní uložena hodnota 1 nebo 0.

V současné době se u SSD používají tři *typy paměťových buněk*:

- SLC (Single-Level Cell)
  - rozpoznávají se dvě úrovně napětí (0 % nebo 100 %), úroveň 0 % znamená uloženou hodnotu 1, úroveň 100 % znamená uloženou hodnotu 0,
  - do jedné buňky lze uložit 1 bit (uloženo 1 nebo 0),
  - počet zápisových operací pro jednu buňku je omezen na cca 100 000.
- MLC (Multi-Level Cell)
  - rozpoznávají se čtyři úrovně napětí (0, 33, 66 a 100 %), úroveň 0 % znamená uloženou hodnotu 3, úroveň 100 % znamená uloženou hodnotu 0,
  - do jedné buňky lze uložit 2 bity (uloženo binárně 11, 10, 01 nebo 00, tj. čísla 3, 2, 1 nebo 0),
  - počet zápisových operací pro jednu buňku je omezen na cca 3 000.
- TLC (Triple-Level Cell)
  - rozpoznává se osm ($2^3 = 8$) úrovní napětí (0, 14, 28, 42, 56, 70, 84 a 100 %), úroveň 0 % znamená uloženou hodnotu 7, úroveň 100 % znamená uloženou hodnotu 0,
  - do jedné buňky lze uložit 3 bity (uloženo binárně 111, 110, 101, 100, 011, 010, 001 nebo 000, tj. čísla 7, 6, . . . , 0),
  - počet zápisových operací pro jednu buňku je omezen na cca 1 000.

Na jedné SSD kartě (ano, není to disk, ale karta, která může/nemusí být „zaobalena“ do podobného pouzdra jako HDD) mohou být MLC a TLC flash čipy kombinovány.



Obrázek 1: Srovnání paměťových buněk typu SLC a MLC[1]

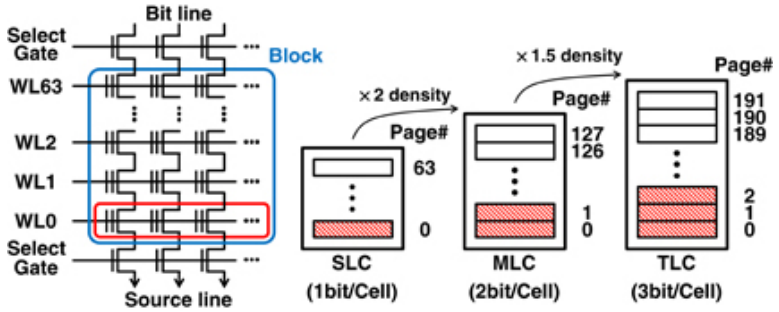
[1] Zdroj: http://en.expreview.com/2012/05/15/supersspeed-hyper-slc-ssd-review/23066.html

Jak vidíme, typ paměťové buňky má velmi podstatný vliv na životnost této buňky. Uvědomme si, že se stoupajícím množstvím dat, která lze do paměťové buňky uložit, stoupá počet potenciálních zápisů *exponenciálně*, proto se exponenciálně urychluje příchod konce životnosti buňky.

1.2 Organizace buněk

Paměťové buňky nejsou používány každá zvlášť, flash paměti obecně se vyznačují blokovou strukturou s využitím stránek.

Paměťové buňky jsou sdruženy do *stránek*. Z důvodu kompatibility řadičů SSD a HDD je velikost stránky typicky 4 KiB (tj. 4096 B). Tyto stránky jsou dále organizovány v *blocích*. Obvykle platí, že při použití SLC buněk je 64 stránek v jednom bloku (256 KiB na blok), u MLC buněk dvojnásobek – 128 stránek v jednom bloku (512 KiB) a u TLC buněk trojnásobek – 192 stránek v bloku (768 KiB).



Obrázek 2: Struktura flash paměti – stránky a bloky[2]

V hierarchii se ve skutečnosti jde ještě dále, ale vyšší členění už nemá přímo vliv na průběh zápisu. Bloky jsou organizovány do rovin (*plane*). V jedné plane je obvykle 1024 bloků a na jednom vyrobeném plátu (*die*, viz procesory v Technickém vybavení osobních počítačů) je vždy určitý počet planes. Výrobce může do jednoho flash čipu umístit určitý počet takovýchto dies, tedy jejich význam je spíše „fyzicky organizační“.

1.3 Řadič

Řadiče flash pamětí (flash memory controller) musejí být o něco inteligentnější a propracovanější než řadiče HDD disků. Na jejich kvalitě hodně závisí přenosové rychlosti (čtení i zápisu) a také životnost paměťových buněk. Veškeré postupy, které jsou dále popisovány, zajišťuje právě řadič.

Uvědomme si, že se to netýká pouze SSD a USB flash disků, ale také například SD karet a dalších vnějších flash pamětí.

V běžných NAND flash pamětech včetně SSD se většinou setkáváme s řadiči od firmy SandForce, nicméně někteří výrobci flash čipů používají řadiče od jiných výrobců (Marvell, JMicron, vlastní Samsung, atd.).

Pokud je nutné využít služeb firmy pro záchranu dat z SSD, s většinou řadičů obvykle nebývá problém, bohužel kromě toho nejběžnějšího – SandForce. Je dobré se v dané firmě

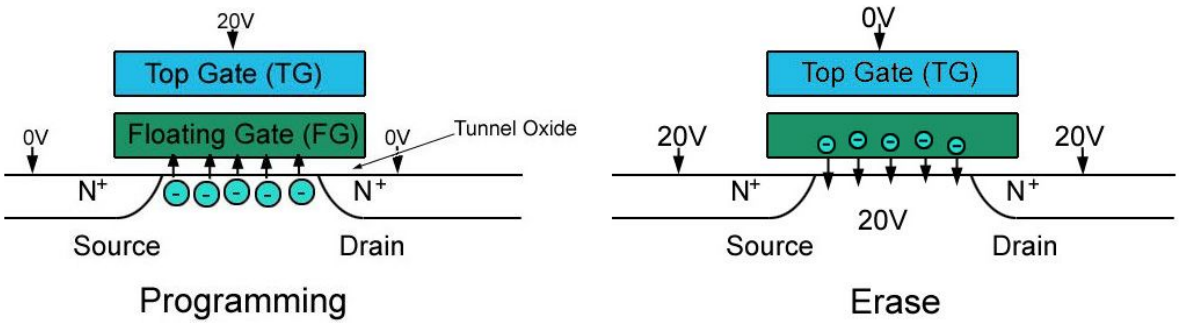
[2] Zdroj: http://m.iopscience.iop.org/1347-4065/53/4S/04EE04/article

předem informovat, zda jsou vybaveni k záchraně dat z SSD s tímto řadičem, mnohé firmy nejsou. V každém případě je lepší, když k takové situaci vůbec nedojde, protože záchrana dat z SSD bývá technicky i finančně náročnější než z klasického disku, takže zálohujte, zálohujte, zálohujte.

2 Průběh zápisu a čtení

2.1 Zápis a čtení na úrovni buněk

SLC bunky se mohou nacházet v jednom ze dvou stavů – buď jsou *vybité* (úroveň napětí 0, buňka nese údaj „1“), a nebo *nabité* (úroveň napětí 1, buňka nese údaj „0“). Výchozí stav buňky je „vybitá“ (obsahuje 1), a pokud chceme do buňky zapsat hodnotu 0, je třeba ji naprogramovat (zvednout napětí). Naopak, kdy do buňky, která je „nabitá“ (obsahuje 0), chceme zapsat hodnotu 1, musíme provést mazání (vybití).



Obrázek 3: Paměťová buňka typu SLC – vlevo průběh programování, vpravo průběh mazání[4]

Čtecí a zápisový proces řídí řídicí brána (top gate). Nabití a vybití se vztahuje ke stavu plovoucí brány, na obrázcích výše je označena jako *floating gate*.

Během nabíjení se do plovoucí brány dostávají přes izolační vrstvu (která je pod ní) elektrony, které po nabití zůstávají v bráně i po přerušení proudu (proto se jedná o permanentní paměť). Nabitá plovoucí brána (obsahující elektrony, tj. záporně nabitá) pak tvoří elektrické pole, které brání průchodu elektrického proudu během čtení (proud neprojde ⇒ přečteme 0, proud projde ⇒ přečteme 1).

Pokud chceme do buňky zapsat hodnotu 1, je třeba plovoucí bránu vybít přivedením opačného napětí, které z brány *vytlačí* elektrony, jedná se o „mazání“ buňky.

Čtení tedy probíhá tak, že se přes řídicí bránu zavede na plovoucí bránu (coby tranzistor) nízké napětí (které přímo elektrony neovlivní) a změří se průchod napětí.

2.2 Čtení a zápis stránky

Operační systém pracuje se stránkami, což se plně respektuje při čtení z flash čipu, čte se vždy jedna stránka najednou. Ovšem zápis je složitější. Jakékoliv flash čipy (NOR jako vnitřní paměti i NAND jako vnější paměti) jsou paměti s blokovým zápisem, tedy se vždy maže a zapisuje celý blok. Na tento proces se podíváme podrobněji.

[4] Zdroj: http://www.lostcircuits.com/mambo//index.php?option=com_content&task=view&id=69&Itemid=1

Pokud má být změněn obsah některého bloku (třeba i jen jediné stránky v tomto bloku, nicméně zapisuje se vždy celý blok najednou), postupuje se takto:

- Řadič používá rychlejší pomocnou paměť (buffer), která není permanentní.
- Pokud do dané stránky nebylo ještě zapisováno či byla smazána (všechny paměťové buňky jsou vybité, mají hodnotu binárně 1, 11 nebo 111 podle typu buňky), je situace jednoduchá – do příslušné stránky je možné data zapsat přímo.
- Pokud však příslušná stránka není označena jako smazaná, je třeba provést smazání (celého bloku), a pak teprve zapisovat. Postup:
  - Do *bufferu* se zkopíruje obsah celého bloku, ve kterém se stránka nachází.
  - Blok je podroben operaci smazání, teď všechny paměťové buňky obsahují binárně maximální možnou hodnotu.
  - V bufferu se provede kompletace finálního stavu bloku, tj. do příslušné stránky provedeme zápis, v této chvíli máme v bufferu blok zkompletován v tom stavu, v jakém ho budeme chtít přímo ve flash paměti.
  - Obsah bufferu (jeden blok) je zapsán na příslušné místo do flash paměti.

Nejpomalejší z výše popsaných operací je operace mazání, a to až řádově.

Kdyby bylo nutné provádět mazání bloku při každém zápisu do paměti, pak by tento zápis byl vždy velmi pomalý. Nové flash čipy jsou smazány už z výroby, tedy zápis do „nových“ stránek je velmi rychlý, pomalý je až zápis do stránek, kde „už něco bylo“. Řadič se (pokud je to možné) pokouší pro zápis nových dat vybírat spíše takové stránky, které už jsou předem smazány, čímž kladně ovlivňuje rychlost zápisu. Pokud mají být přepsána existující data, zápis nemusí nutně být proveden do téže stránky, ve které data byla původně – stránka je označena jako nepoužívaná a určená ke smazání (ale není smazána), a je vybrána jiná (už předem smazaná) stránka.

U používanějších flash pamětí proto dochází k situaci, kdy žádné volné nesmazané stránky nejsou k dispozici, a tedy se řadič mazání bloků nevyhne. Proto u takových disků může docházet k postupnému zvyšování doby zápisu. Kvalitní řadič si však s tím dokáže poradit a bloky, ve kterých je takto označeno dostatečné množství stránek, operativně maže a tím připravuje k rychejšímu zápisu.

2.3 ATA TRIM

Jeden z ATA příkazů (tj. fungující přes rozhraní SATA, operační systém tento příkaz může poslat řadiči) je TRIM. Aby bylo možné tento příkaz používat, musí mu „rozumět“ jak řadič, tak i operační systém, což bychom si u řadiče měli ověřit při koupi SSD, u operačního systému bychom tuto informaci měli znát z výuky Technického vybavení.

Příkaz ATA TRIM má řešit jisté „nedorozumění“ mezi fungováním flash pamětí a běžným chováním souborových systémů. Pokud v souborovém systému (FAT, NTFS, ext3fs apod.) smažeme soubor, ve skutečnosti není smazána celá oblast, která dotyčný soubor obsahovala, ale pouze je odstraněn odkaz na tento soubor v příslušných datových strukturách, které si souborový systém vede: v adresáři (složce), kde byl soubor umístěn, je smazána příslušná evidenční položka s přístupovými a bezpečnostními údaji o souboru, a dále v tabulce volných clusterů/bloků je daný cluster/blok označen jako volný. Ovšem řadič flash paměti není

informován o tom, že oblast, kterou soubor fyzicky zabíral, už není využívána a může být smazána, proto ji nemůže přiřadit jinému souboru.

V případě klasického pevného disku by to problém nebyl, protože umístění souboru do konkrétních clusterů určuje přímo operační (resp. souborový) systém, ale u SSD je to čistě věc řadiče.

Pokud operační systém podporuje příkaz ATA TRIM, dokáže řadiči sdělit, které clustery (mapované na stránky ve flash paměti) jsou ve skutečnosti volné, a řadič následně může k nim přidružené stránky označit jako určené ke smazání (smažou se až tehdy, když řadič stanoví ke smazání celý blok). Využitím ATA TRIM tedy uvolňujeme kapacitu flash paměti, osvobozujeme ji od „virtuálně zabraných“ stránek, důsledkem je zvýšení rychlosti zápisu (protože více bloků je včas smazáno, řadič má více na výběr, do kterých stránek zapisovat).

3 Životnost paměťových buněk

Než dojde k naprostému ukončení životnosti paměťové buňky, probíhají procesy, kterými se řadič pokouší životnost prodloužit.

3.1 Vliv operací mazání na životnost flash paměti

Kdykoliv dochází k zápisu do paměťové buňky, opotřebovává se *izolační vrstva*, přes kterou procházejí elektrony při mazání a zápisu do buňky – když už je příliš poškozená, řadič ji označí za vadnou a buňka není dále používána (vlastně celý blok). Likvidační jsou především operace mazání, protože ty se provádějí vždy pro celý blok najednou (navíc zápis do stránky je obvykle spojen se smazáním celého bloku, ve kterém se stránka nachází, má tedy smysl počítat spíše operace mazání než zápisu).

Věc se komplikuje tím, že postupem času v izolační vrstvě uvíznou sem tam nějaké elektrony, čímž se (kromě samotného ničení izolační vrstvy) také zhoršuje možnost správného přečtení obsahu buňky, zachycené elektrony tento výsledek ovlivňují. Řadič se to pokouší kompenzovat zvýšením čtecího proudu a jeho aplikováním po delší dobu, což však může ještě zvýšit rychlost opotřebení izolační vrstvy a prodloužit dobu zápisu. Pokud už délka zápisu přesáhne stanovenou mez, blok taktéž přestane být používán.

Řadič u každého bloku eviduje počet provedených mazacích cyklů. Tato hodnota je pak důležitým ukazatelem při algoritmech vyvažování zátěže (určujeme, které bloky budou pro zápis používány přednostně – spíše ty, které byly dosud méněkrát mazány).

3.2 Zvyšování životnosti buněk

Výše naznačené limity životnosti (například 3000 cyklů pro MLC buňky) vypadají hrozivě. Ale uvědomme si, že ne každý den přepisujeme celý obsah flash paměti.

Řadič to samozřejmě nemůže nechat *jen tak*. Některé možnosti zvýšení životnosti byly popsány výše, především efektivnější stanovování stránek pro zápis. Další možnosti:

- Kromě deklarovaného množství paměti disponuje řadič ještě (často nezveřejněným množstvím) paměti navíc (obvykle asi tak o třetinu). Tato paměť je používána pro nahrazování (přemosťování) oblastí paměti, jejichž životnost již vypršela.

- Frekventované soubory (log soubory apod.) bývají čas od času přesouvány na stránky, na kterých nejsou paměťové buňky příliš opotřebované. Toto vyrovnávání opotřebení se nazývá *statický Wear Leveling*.
- Řadič má implementovány algoritmy *dynamického Wear Levelingu*, které se používají při každém zápisu. Obecně to znamená, že přednostně jsou pro zápis používány stránky s nižší mírou opotřebení, aby zápisové a následně čtecí doby byly co nejdelší dobu nízké.
- V operačním systému bychom měli vypnout určité funkce, které nejsou potřeba a zbytečně snižují životnost flash paměti, ideálně by to měl provést samotný operační systém (pokud detekuje flash paměť) – vypnout defragmentaci disku, omezit používání log souborů, dočasných souborů, případně cookies, ve Windows funkce Prefetch a Superfetch, Ready Boost, Volume Shadow Copy, atd.

Vzhledem k tomu, jakým způsobem dnes řadiče flash pamětí fungují, výrobci nabízejí garanci životnosti v rozsahu 3–5 let, u kvalitnějších i déle. Pokud na SSD nezapisujeme každodenně kvanta dat, životnost naprosto dostačuje a odpovídá běžné životnosti celého stroje.

Pokud máme pocit (umocněný použitím některého z následně uváděných nástrojů), že nás SSD se blíží konci svého života, můžeme ho i dále bez problémů používat pro data, která jen velmi zřídka měníme a spíše je čteme (například zásobárna filmů, ty moc nemažeme). Čtením se totiž paměťové buňky nijak neopotřebovávají.

4 Nástroje

Co se SSD týče, může se hodit nástroj, který sleduje S.M.A.R.T. hodnoty. Narozdíl od klasických pevných disků je určení životnosti SSD (tj. dopočítání ukazatele z hodnot poskytnutých řadičem) poněkud náročnější, ne všechny nástroje ji zvládají dostatečně kvalitně.

Stejně jako u klasických disků, i zde můžeme využít známý *CrystalDiskInfo*, ovšem jeho odhady životnosti z načítaných hodnot nejsou moc přesné. O něco přesnější výsledky podává *SSDLife*, který od chvíle své instalace neustále sleduje jednotlivé parametry a své výsledky průběžně zpřesňuje (takže jím uváděné hodnoty jsou postupem času čím dál kvalitnější).

Vyšší nároky na softwarovou výbavu máme v případě, že dojde k poruše disku a je třeba z něj zachránit data. Není řečeno, že nutně musíme hledat specializovanou firmu (záleží na typu poruchy), někdy můžeme sami vytvořit obraz disku, ze kterého již data dostaneme.

Vytvořit image disku lze pomocí nástroje *Paragon Backup & Recovery Free* nainstalovaného ve Windows, kdy poškozený disk (pokud je vyjmutý) připojíme přes tutéž redukci jako klasický disk (záleží na datovém rozhraní, které na SSD máme, SATA či jiné). Pokud se to nepovede, můžeme vyzkoušet získání dat pomocí některé Linux Live distribuce. Pro záchranu dat se často používá distribuce *Parted Magic*.

Odkazy na další zdroje

- ČERNÝ, Jan. Solidní budoucnost pevných disků – úvod k velkému testu SSD disků [online]. *PCTuning Týden.cz*, 2010. Dostupné na: http://pctuning.tyden.cz/hardware/disky-cd-dvd-br/18914-solidni-budoucnost-pevnych-disku-uvod-k-velkemu-testu-ssd-disku
- HACHIYA, Shogo, et al. Hybrid triple-level-cell/multi-level-cell NAND flash storage array with chip exchangeable method [online]. *Japanese Journal of Applied Physics* 53(4S).

IOP Publishing, 2011. Dostupné na:
http://m.iopscience.iop.org/1347-4065/53/4S/04EE04/article

- HANDY, Jim. *How Controllers Maximize SSD Life – Better Wear Leveling* [online]. The SSD Guy, 2012. Dostupné na: http://thessdguy.com/how-controllers-maximize-ssd-life-better-wear-leveling/
- HORT, Tomáš. Technologie a zajímavosti z oblasti SSD disků [online]. *PCTuning Týden.cz*, 2011. Dostupné na: http://pctuning.tyden.cz/hardware/disky-cd-dvd-br/22588-technologie-a-zajimavosti-z-oblasti-ssd-disku
- Increasing Flash SSD Reliability [online]. *Storage Search*. Dostupné na: http://www.storagesearch.com/siliconsys-art1.html
- SCHUETTE, Michael. The Brave New World of SSDs [online]. *Lost Circuits*, 2009. Dostupné na: http://www.lostcircuits.com/mambo//index.php?option=com_content&task=view&id=69&Itemid=1