Uticaj disk keš bafera na performanse SSD diskova

Valentina Timčenko, Borislav Đorđević
Institut Mihajlo Pupin
Beograd, Srbija
valentina@kondor.imp.bg.ac.rs , bora@impcomputers.com

Slobodan Obradović, Nikola Čorni
VIPOS
Valjevo, Srbija
sobradovic@vipos.edu.rs, nedzadin@gmail.com

Sadržaj—**U ovom radu su predstavljeni rezultati dobijeni ispitivanjem performansi SSD diskova pod MS Windows operativnim sistemom. Detaljno je analiziran uticaj kompresije i disk keš bafera na samom disk drajvu na performanse SSD diskova. Performanse su merene primenom ATTO i AS SSD benchmark aplikacija.**

Ključne reči – SSD, disk bafer, MS Windows, sistemi datoteka, performanse

I. UVOD

Sve veći i brži razvoj računarske tehnike za različite potrebe tržišta, uslovio je potražnju za sve većim performansama računarskog hardvera. Pored snažnih mikroprocesora koji obrađuju veliku količinu podataka u jedinici vremena, sve je veća količina podataka koje treba uskladištiti. Štaviše, pored velikog skladišnog kapaciteta za podatke, potreban je i brži pristup istima. Tradicionalni hard disk nije mogao da obezbedi zahtevane performanse novim potrebama, pa je zato napravljen novi elektronski medijum za masovno skladištenje podataka - Solid State Drive, SSD.

Elektronski način skladištenja podataka poznat je već nekoliko decenija, ali je tadašnji spor razvoj elektronike u ovom pravcu ograničavao pristupačnu komercijalnu proizvodnju ovakve vrste memorije. Međutim, u poslednjih nekoliko godina, razvoj je napredovao u velikoj meri, pa je usled toga tema SSD diskova aktuelna danas.

Proizvodnja SSD diskova je komercijalno počela pre oko pet godina, kada se na tržištu pojavilo svega nekoliko modela najistaknutijih svetskih proizvođača memorija. To je bila prva generacija SSD diskova koja je sa sobom nosila prilično visoku cenu, uz takođe veliki broj falinki.

Pre oko tri godine izdata je druga generacija SSD-ova sa većim brzinama čitanja i upisa, otklonjenim nedostacima i nešto nižim cenama. U to vreme se pojavio i SATA 3 standard.

Pre nešto manje od 2 godine, proizvođači su počeli da izdaju SSD jedinice koje nude brzine iz opsega SATA 3 standarda i ponovo je snižena cena na sada već pristupačnu vrednost. Sad već i jedan malo napredniji korisnik računara može sebi priuštiti SSD jedinicu i poboljšati ličnu ugodnost u radu ili zabavi.

Obzirom da je do sada na tržište izdato mnoštvo SATA 3 SSD diskova, postavlja se pitanje: Koji model izabrati ?

Iz razloga da odgovori na ovo aktuelno pitanje, napisan je i ovaj rad koji za glavni cilj ima da testiranjem pokaže koji proizvođač nudi najbolje performanse.

II. CILJ I MOTIV RADA

Cilj rada je da se na sistematičan načim predstave rezultati dobijeni ispitivanjem performansi SSD diskova i tumače rezultati na bazi internih faktora SSD diskova. Motivaciju za ovaj rad smo pronašli u činjenici da novi SSD diskovi uključuje brojna poboljšanja u odnosu na svoje prethodnike, a da više internih faktora SSD diskovao, kao što je interni keš bafer mogu imati dosta uticaja na performanse.

III. PRINCIP RADA SSD DISKOVA

SSD (Solid State Drive) disk trajno skladišti podatke kao i hard disk, ali ih umesto na magnetnim diskovima, čuva u flash memorijskim čipovima. Zbog toga što se podaci čuvaju u elektronskom umesto u magnetnom obliku, SSD-ovi su brži od hard diskova iz dva razloga. Prvo, nije potrebna konverzija električnog oblika informacije u magnetni oblik, a drugo, SSD ne koristi pokretne mehaničke delove, pa bilo kojim uskladištenim podacima može da pristupa gotovo trenutno, tj. sa mnogo kraćim vremenom pristupa. SSD jedinice se takođe proizvode sa nekoliko različitih interfejsa za prenos podataka sa i na SSD, ali se SATA interfejs najčešće koristi.

Unutrašnjost SSD jedinice čine tri osnovne komponente:

- Flash memorija
- Kontroler
- Bafer

Flash memorija sadrži same podatke. Ovo je ista vrsta memorije koja se koristi u prenosivim USB Flash memorijskim stikovima i u na memorijskim karticama za digitalne kamere. Flash memorija može biti proizvedena u dve različite tehnologije - NAND i NOR, a sve SSD jedinice koriste memoriju proizvedenu tehnologijom NAND. Flash NAND memorija je dostupna u dve različite vrste memorijskog pakovanja:

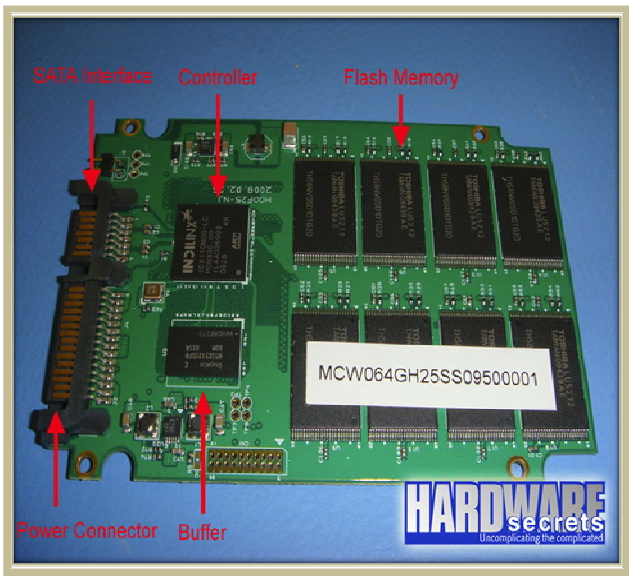
- SLC (Single Level Cell)
- MLC (Multi Level Cell)

Kod SLC čipova, u jednom ciklusu upisa, u čip se upisuje samo jedan bit podataka, dok se kod MLC čipova, u jednom ciklusu upisuje više bitova podataka. Pored toga, MLC čipovi su jeftiniji, jer zbog gušćeg memorijskog pakovanja u jedan svoj čip smeste više podataka nego jedan SLC čip. Dakle, jeftinije SSD jedinice koriste MLC čipove, dok skuplje jedinice koriste SLC čipove. Mana MLC čipova je što oni imaju kraći životni vek u odnosu na SLC čipove iz razloga veće ograničenosti broja upisnih ciklusa. Prosečan SLC čip dozvoljava do 100.000 upisnih ciklusa pre nego što postane neupotrebljiv, dok MLC čipovi imaju ograničenje na oko 10.000 ciklusa. Najnoviji MLC čipovi proizvedeni 25-nanometarskim proizvodnim procesom imaju ograničenje na samo 5.000 ciklusa.

Kontroler predstavlja glavni i najbrži deo SSD jedinice. Nekoliko kompanija proizvode ovu vrstu čipa, a najpoznatije su Indilinx, Intel, Samsung, Sand Force i JMicron. Kontroleri organizuju nizove flash memorije u kanale kojima se teoretski može pristupiti nezavisno. Broj ovih kanala trenutno se kreće od 4 do 10. Tako će npr. kontroler koji flash memoriju organizuje u 10 kanala biti teoretski brži od kontrolera koji memoriju deli na 8 ili manje kanala. Kontroler može biti povezan sa magistralom preko bafera ili direktno, u zavisnosti od toga da li ima bafer ili ne. Kontroler čip se naziva još i SOC čip (System On a Chip) i u njemu se nalazi firmware uređaja.

Jedan čip memorije tipa SDRAM male potrošnje, koristi se da ubrza i uravnoteži brzinu komunikacije između kontrolera i SATA interfejsa. Veličina memorije ovog čipa u SSD jedinicama je tipično 64 MB ili 128 MB. Međutim, SATA 3 SSD jedinice bazirane na Sand Force kontroleru ne sadrže ovaj čip iz razloga što je brzina kontrolera bliska brzini SATA 3 magistrale, pa mu nije potrebna amortizacija brzine pri upisu ili čitanju.

Sl. 1 prikazuje unutrašnjost jedne SSD jedinice, sa obeleženim osnovnim komponentama.



Slika 1. Unutrašnjost SSD jedinice

IV. ČITANJE, UPIS, BRISANJE, WEAR LEVELING, OVER-PROVISIONING, WRITE AMPLIFICATION

Da bismo objasnili osnovne operacije koje vrši SSD jedinica, moramo dublje ući u suštinu fizičke organizacije memorijskog prostora unutar jedinice.

Memorijska ćelija predstavlja osnovnu memorijsku komponentu koja čuva jedan bit podataka u slučaju SLC čipa, a dva bita podataka kada je u pitanju MLC čip.

Stranica (page) je veličine 4 KB i nju čine organizovane grupe ćelija. Stranice predstavljaju najmanju strukturu koja se može očitavati ili na koju se može upisivati.

Blok (block) se sastoji od 128 stranica, pa je njegova veličina 128 x 4 KB = 512 KB. Blok predstavlja najmanju memorijsku strukturu koja se može brisati.

Ravan (plane) čine 1024 bloka, a njena veličina je 1024 x 512 KB = 512 MB.

Fizički čip sadrži više ravni, a ima kapacitet od 8, 16 ili 32 GB.

SSD jedinica ima 8 ili 16 fizičkih čipova koji svojim brojem i kapacitetom određuju ukupan kapacitet jedinice u rangu od 64 do 512 GB.

Proces čitanja je najprostiji, a vrši se adresiranjem određene stranice unutar nekog bloka u jednoj od ravni u čipu. Tom prilikom se očitava 4 KB podataka.

Proces upisa se odvija isto kao i proces čitanja, samo što se tom prilikom upisuje u stranicu minimum 4 KB podataka. Ono što je važno napomenuti je da svakim novim upisom u neku od stranica, ukupan broj upisnih ciklusa te stranice smanjujemo za jedan. Kada se broj upisnih ciklusa "potroši", tj. dostigne brojku od oko 10.000 za MLC ili oko 100.000 za SLC čip, tu stranicu više nije moguće brisati i ponovo u nju upisivati.

Proces brisanja je malo složeniji proces i sastoji se od četiri koraka. Prvi korak zahteva operaciju čitanja celog bloka (512 KB) u internu memoriju, zatim se u drugom koraku u memoriji modifikuju neke od stranica (4 KB), u trećem koraku sledi operacija samog brisanja celog bloka (512 KB) flash-a i u četvrtom koraku operacija ponovnog upisa stranica (4 KB) čiji sadržaj nije menjan i novih stranica umesto starih ne validnih.

Ovako komplikovan proces brisanja uslovljen je zbog činjenice da kod SSD-a nije moguće prosto prepisati određenu stranicu, kao što je to moguće kod hard diska. Razlog za to je problem fizičke mikroelektronike koji se javlja zbog veoma male površine i istovremeno veoma kompleksne električne šeme unutar svakog čipa.

Sledeće tri funkcije su usko povezane i najviše zavise od samog kontrolera, a imaju za cilj da što više produže životni vek SSD jedinice.

Obzirom da je broj upisnih ciklusa svih stranica ograničen, kontroler mora voditi računa da jednako koristi svaku stranicu. Za to je zadužena funkcija **wear leveling** koja ima zadatak da nove upise koordinira tako da se podjednako "troše" sve stranice.

Iako se sve trenutne SSD jedinice prave u stvarnim fizičkim kapacitetima od 64 - 512 GB, nijedna od njih ne dozvoljava korisniku da iskoristi ovaj pun kapacitet za svoje potrebe. Umesto toga, zavisno od proizvođača, ali najčešće 7,5% od ukupnog kapaciteta jedinice se odvaja za funkciju **over-provisioning**. Taj prostor je skriven od korisnika, a jedinica balansira nove upisne cikluse delom u slobodne stranice u korisničkom prostoru, a delom u over-provisioning zonu. Time smanjuje broj ciklusa brisanja i ponovnog pisanja, čime ostvaruje dva benefita. Prvo, štedi broj upisnih ciklusa, a drugo, ne gubi na performansama prilikom datog upisa.

WA (write amplification) je funkcija koja direktno smanjuje broj upisnih ciklusa. WA predstavlja odnos količine podataka zadate kontroleru jedinice na upis i količine podataka stvarno upisane u memorijske ćelije. Ova razlika postoji zbog mogućnosti koju poseduju najnoviji kontroleri, a to je kompresija podataka. Sand Force kontroler, koji je trenutno na vrhu po svim performansama, vrši kompresiju podataka "na živo". On se ne bavi tipom sadržaja, niti njegovom stvarnom ukupnom količinom, već on u trenutku kada je zadat upis podataka preuzima svaki novopristigli blok za upis i pokušava da kompresuje. Koliko će u ovom poslu biti uspešan direktno zavisi od tipa sadržaja nad kojim vrši kompresiju. Ukoliko je podatke moguće više kompresovati, sačuvaće se i memorijski prostor jedinice, pa će samim tim i broj stranica na koje se fizički upisuje/prepisuje biti manji, čime se sporije troši ograničen broj upisnih ciklusa, a na kraju se dobijaju i bolje upisne performanse.

Zbog svih navedenih funkcija koje obavlja, jasno je da kontroler SSD jedinice predstavlja vrlo komplikovan fizički čip, a ono što je još komplikovanije je napisati što bolji firmware pod kojim kontroler radi. Usled složenosti i jednog i drugog problema, proizvođači izdaju po više modela kontrolera jedne iste serije, a često se rade i revizije prethodnog firmware koda za već izdate jedinice.

V. MEĐUSOBNO POREĐENJE – SPECIFIKACIJA

Za potrebe ovog rada korišćeno je 5 vrsta diskova sa karakteristikama koje su date u tabelama I, II, III, IV i V:

TABELA I. SPECIFIKACIJA ZA MODEL A-DATA S511

Kapacitet	120 GB
Tip flash memorije	Intel 25nm NAND MLC
Kontroler	Sand Force SF-2281
Bafer	Nema
Read	550 MB/s max
Write	510 MB/s max

TABELA II. SPECIFIKACIJA ZA MODEL CORSAIR FORCE 3

Kapacitet	120 GB
Tip flash memorije	Micron 25nm NAND MLC
Kontroler	Sand Force SF-2281
Bafer	Nema
Read	550 MB/s max
Write	510 MB/s max

TABELA III. SPECIFIKACIJA ZA MODEL KINGSTON HYPER X

Kapacitet	120 GB
Tip flash memorije	Intel 25nm NAND MLC
Kontroler	Sand Force SF-2281
Bafer	Nema
Read	555 MB/s max
Write	510 MB/s max

TABELA IV. SPECIFIKACIJA ZA MODEL CRUCIAL M4

Kapacitet	128 GB
Tip flash memorije	Micron 25nm NAND MLC
Kontroler	Marvell 88SS9174-BLD2
Bafer	Micron 1GD22-D9LGQ (256 MB)
Read	500 MB/s max
Write	175 MB/s max

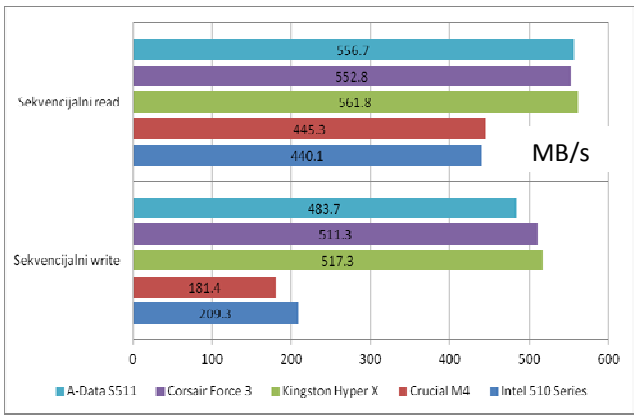
TABELA V. SPECIFIKACIJA ZA MODEL INTEL 510 SERIES

Kapacitet	120 GB
Tip flash memorije	Intel 34nm NAND MLC
Kontroler	Marvell 88SS9174-BKK2
Bafer	Hynix H5TQ1G63BFR (128 MB)
Read	450 MB/s max
Write	210 MB/s max

VI. DETALJNA ANALIZA

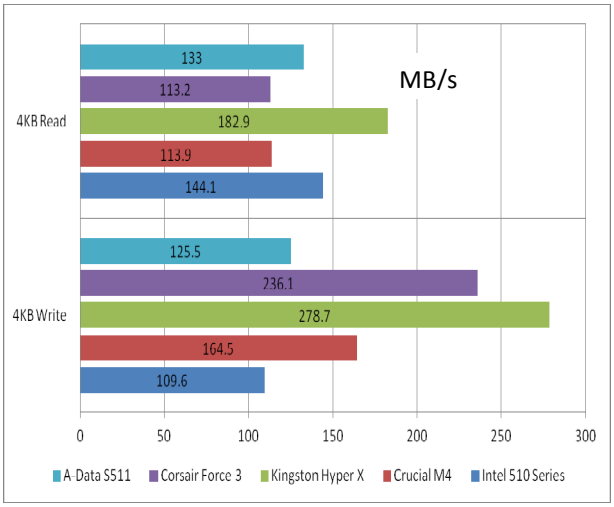
Za potrebe ovog rada korišćeni su ATTO Disk Benchmark i AS SSD Benchmark. ATTO Disk Benchmark formira upisne šeme podataka nad kojima Sand Force kontroler može da prikaže prednosti kompresije. AS SSD Benchmark formira upisne šeme gde je mali stepen moguće kompresije, pa njegovi write testovi prikazuju direktne performanse samih MLC čipova. Na Sl. 2 prikazan je grafik za sekvencijalni read/write a na Sl. 3 grafik za 4KB read/write.

Testiranjem između pet SSD jedinica najpoznatijih svetskih proizvođača, istakao se apsolutni pobednik - Kingston Hyper X. Poredeći vrednosti na sumarnim graficima vidimo sledeće stanje. Na sekvencijalnom read testu Kingston je za vrlo malo nadmašio svoja dva najsličnija konkurenta - model A-Data S511 i Corsair Force 3. Kod sekvencijalnog write testa Kingston je za 6,9% ispred Corsaira. Naime, sva tri modela u osnovi imaju isti model Sand Force kontrolera, pa su zato slične performanse bile i očekivane.



Slika 2. Grafik za sekvencijalni read/write

U read/write performansama na 4KB, Kingston je pokazao mnogo veću razliku, nadmašivši model A-Data za 37,5% u read, a Corsair za 18% u write testu.



Slika 3. Grafik za 4KB read/write

Sand Force kontroler vrši kompresiju podataka, pa je zato u slučaju prve tri jedinice dat i grafik kompresije. Upoređujući podatke sa grafika, vidimo da A-Data ima dosta dobru krivu za opis read kompresije, sličnu Kingston-ovoj, ali joj write kriva startuje sa nesto niže pozicije. Corsair ima nešto bolju krivu za write, ali takođe i lošu krivu za read. Kingston objedinjuje dobar read i write, a takođe doseže i nešto više vrednosti na kraju grafika.

Crucial M4 i Intel 510 Series jedinicama upravlja Marvelov kontroler 9174. Ova dva modela se razlikuju u MLC memoriji koja je u slučaju Intela proizvedena u 34nm procesu za razliku od 25nm MLC čipovima koju koriste Crucial i prethodna tri modela.

Obzirom da Marvel kontroler ne vrši kompresiju sadržaja, performanse za write kod Crucial i Intel jedinica prestaju da rastu već na upisnim sekvencama od oko 16KB i ostaju pri sličnim vrednostima za veće sekvence. Iz ovog razloga grafik kompresije ne postoji za modele Crucial i Intel na testiranju performansi.

VII. ZAKLJUČAK

"Na sekvencijalnom testu razlika u brzinama read/write praktično je nebitna i ne zavisi od toga da li bafer postoji ili ne. Isto tako na read/write 4BK testu postojanje bafera nije dalo bolje performanse, naprotiv i u ovom testu su bolje karakteristike pokazali diskovi bez bafera. Zbog kompresije, sve tri jedinice bez bafera postižu veće sekvencijalne brzine upisa u odnosu na preostala dva modela - Crucial M4 i Intel 510 Series.

Drugo što se može zaključiti jeste da razlika je u veličine bafera - Intel koristi Hynih bafer kapaciteta 128 MB, a Crucial Micronov bafer veličine 256 MB, što za posledicu ima nešto veći sekvencijalni read, kao i čak 50,1% veći 4KB write u odnosu na Intela. Ova brzina je čak veća i od brzine Sand Force baziranih jedinica u slučaju kada one upisuju sadržaj koji se ne može kompresovati.

ZAHVALNICA

Rad je finansiran od strane Ministarstva prosvete i nauke Republike Srbije (Projekat TR 032025).

REFERENCE

[1] M. Seltzer, G. Ganger, M. McKusick, K. Smith, C. Soules, C. Stein, “Journaling versus Soft Updates: Asynchronous Meta-data Protection in File Systems”, USENIX Conf. Proc., San Diego, June 2000. pp. 71-84

[2] Tweedie S., “EXT3, Journaling Filesystem”, July 2000.

[3] A. Mathur, M. Cao, S. Bhattacharya, A. Dilger, A. Tomas, L. Vivier, "The new ext4 filesystem: current status and future plans" in Proceedings of the Linux Symposium, Ottawa, Canada, June 2007.

[4] Roderick W. Smith, "Migrating to Ext4". DeveloperWorks. IBM, http://www.ibm.com/developerworks/linux/library/l-ext4/, April 2008.

[5] "Ext4 Howto", [Online]. https://ext4.wiki.kernel.org/index.php/Ext4_Howto#Bigger_File_System_and_File_Sizes, January 2011.

[6] „First benchmarks of ext4“, [Online]. http://www.linuxinsight.com/first_benchmarks_of_the_ext4_file_system.html, October 2006.

[7] B. Baude, “RAID on Linux on POWER”, IBM eServer Solutions Enablement, November 2005.

[8] A. Thomasian, J. Xu, „Reliability and Performance of Mirrored Disk Organizations“ Computer Journal, January 2008.

[9] A. Lebrecht, N. Dingle, W. Knottenbel, „Analytical and Simulation Modelling of Zoned RAID Systems“ in Computer Journal, June 2010.

[10] V. Danen, “Set up Logical Volume Manager in Linux”, March 2007.

[11] J. Katcher, “PostMark: A New File System Benchmark”, Technical Report TR3022. Network Appliance Inc, October 1997.

ABSTRACT

This paper presents the results obtained for performance solid-state drives under MS Windows operating system. We analized in detail the impact of the compression and disk buffer cache to the performance of solid-state drives. Performance are measured using the ATTO and AS SSD benchmark applications.

IMPACT OF DISK CACHE BUFFERS ON THE PERFORMANCE OF SOLID-STATE DRIVES

Valentina Timčenko, Borislav Đorđrvić, Slobodan Obradović, Nikola Čorni