Uchovávání dat v SSD

Co se děje s mými daty?

Aleš Padrta

Karel Nykles

Obsah

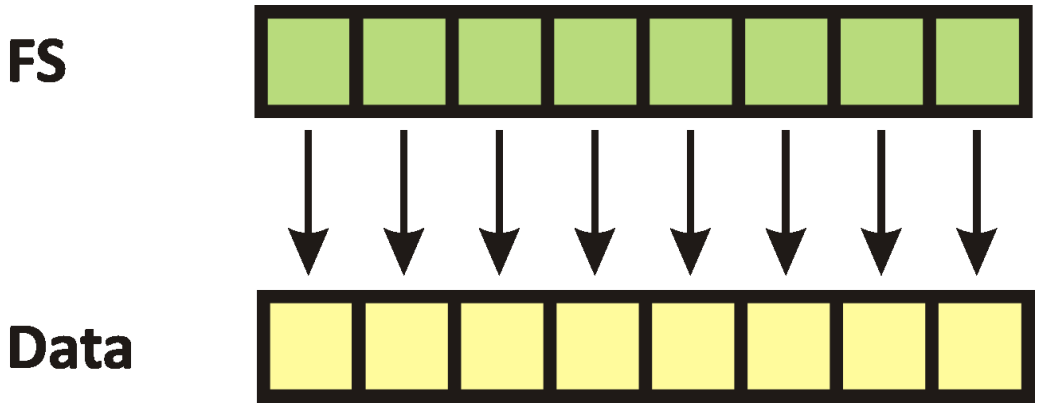
- Úvodní slovo
- Technické pozadí
  - Technologie flash
  - Servisní procedury
  - Kontroler
- Experimenty
  - Testovací prostředí
  - Výsledky
- Shrnutí

Úvodní slovo

- SSD = Solid State Drive
- Rozšířen
  - Rychlost
  - Hardwarová velikost
  - Žádné pohyblivé části
- Jiná technologie (než klasický HDD)
  - Dopad na uložení
  - Dopad na zacházení
  - Ovlivní získávání el. důkazů
- Analýza chování SSD

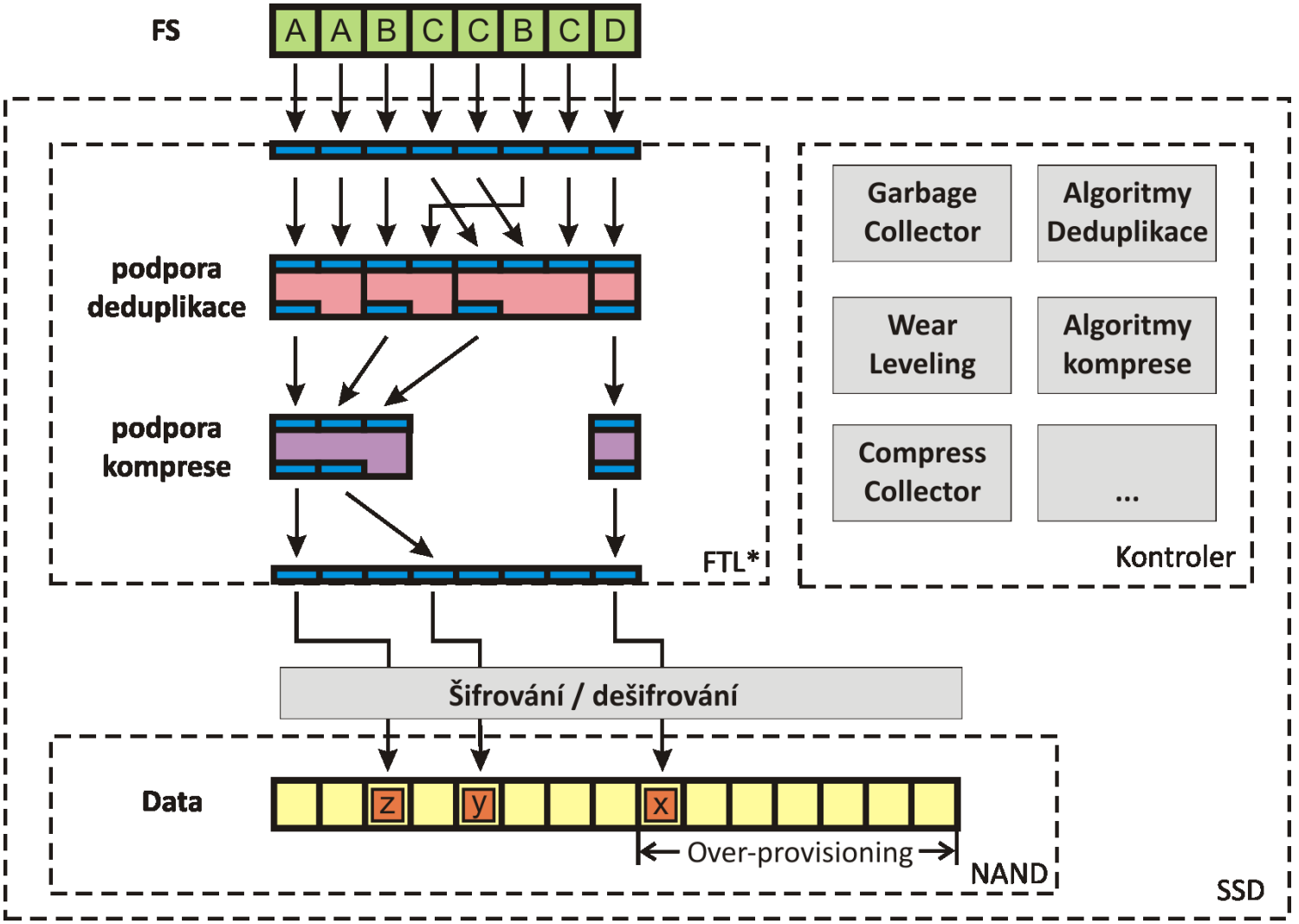
První pohled do útrob SSD

- Naivní = stejné jako HDD
  - Adresa ve FS = umístění na médiu



Druhý pohled do útrob SSD

- Realita = mnohem složitější

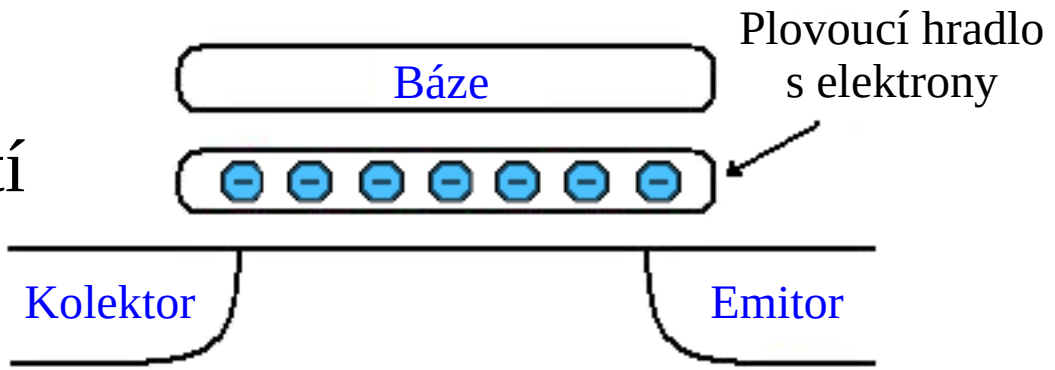


Technické pozadí

Paměťová buňka

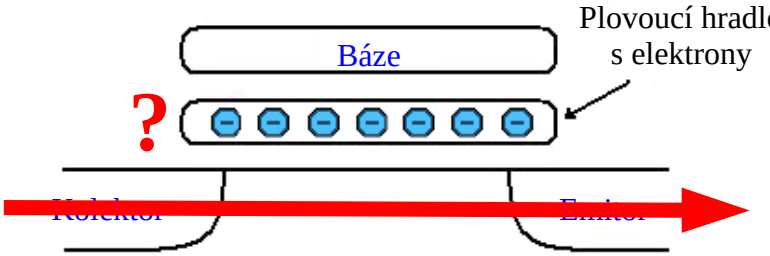
- Základní jednotka
  - Tranzistor s plovoucím hradlem
  - Schopnost udržet náboj
- Nabíjení
  - Emitor – uzemněn
  - Báze – nízké napětí
- Vybíjení
  - Opačná polarita
  - **Vysoké napětí**
  - Část náboje nelze vybít (kumuluje se)

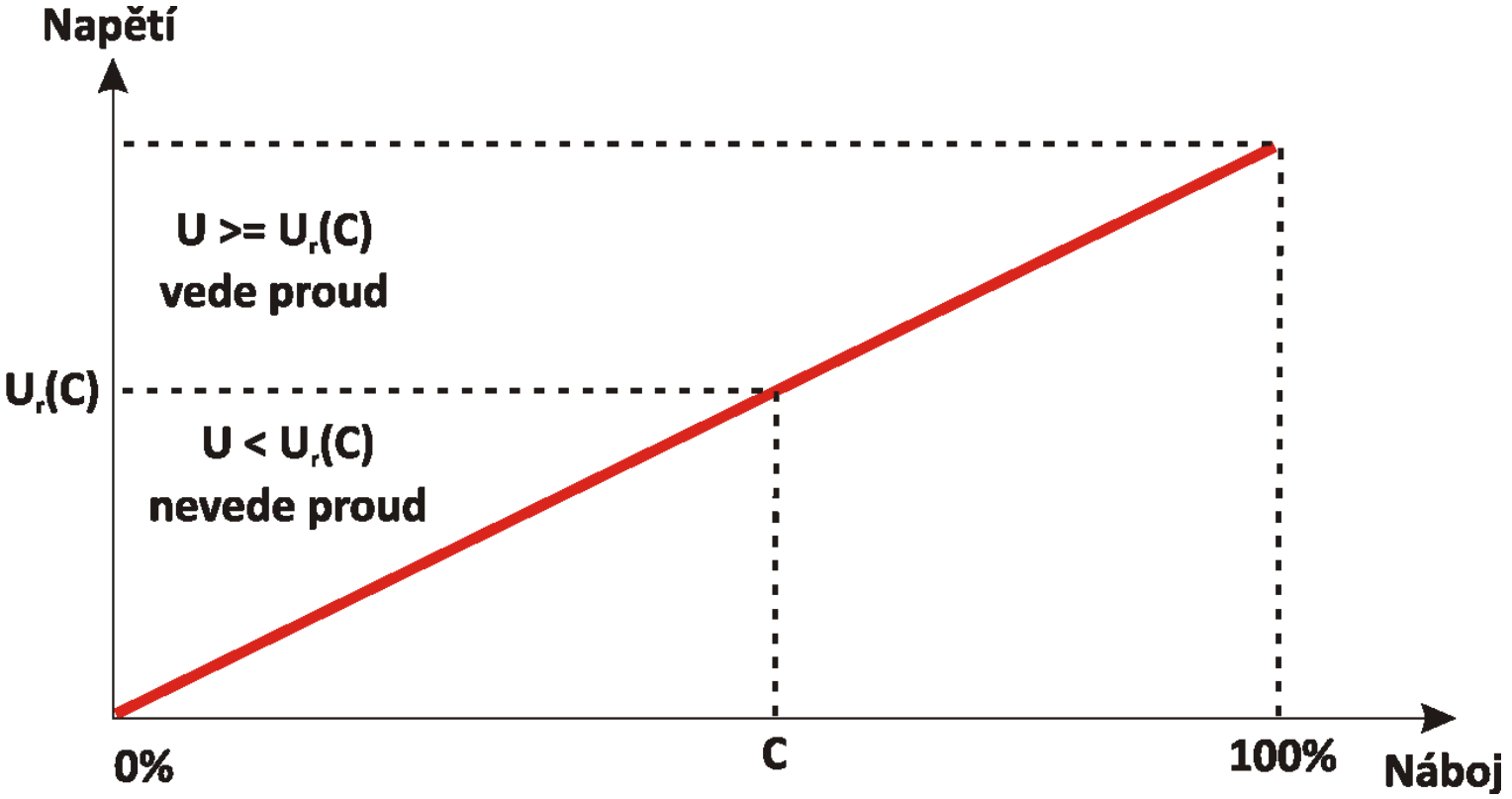
    ⇨ degradace v čase



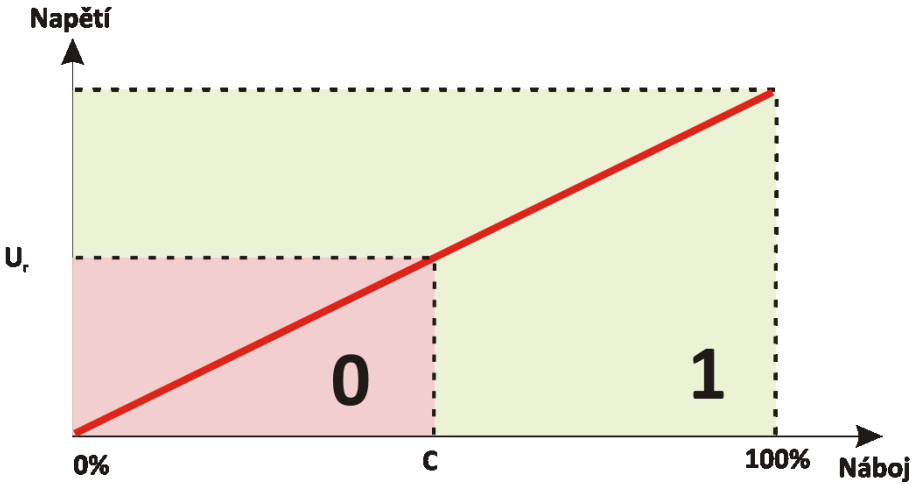
Paměťová buňka

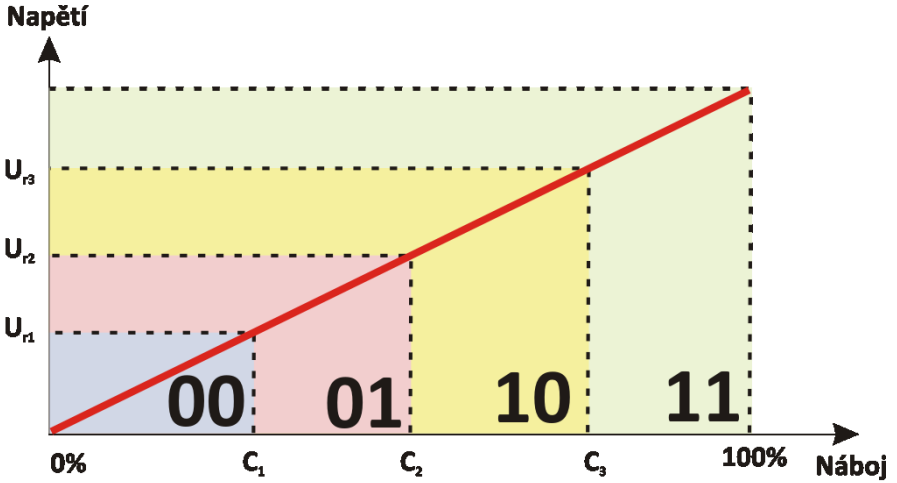
- Vodivost kolektor → emitor
  - Ovlivněno nábojem C
  - Napětí **U >= U$_r$(C)** → vede proud





Paměťová buňka





- SLC (single-layer cell)
  - 1 bit = 2 úrovně (vede/nevede proud při daném U)
  - 100 000 vybíjecích cyklů
- MLC (multi-layer cell), TLC (triple-layer cell)
  - N bitů = 2^N úrovní (více testovaných U_i, i=1, ..., N)
  - Více úrovní = vyšší citlivost na “zbytkový náboj”
    ⇨ pouze 10 000 vybíjecích cyklů

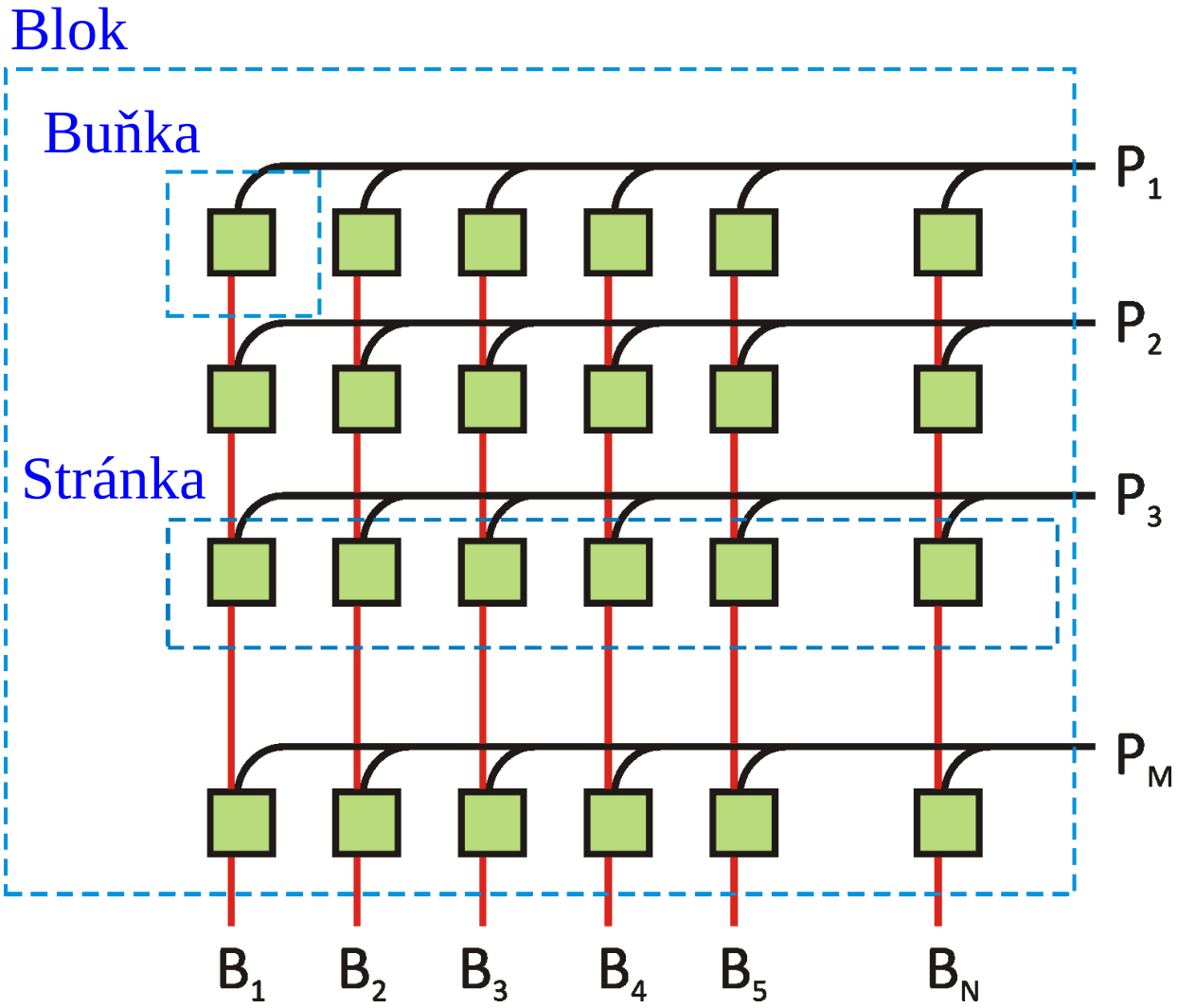
Paměťové buňky ve struktuře

- NAND uspořádání
  - Méně drátů
- Hierarchie
  - Buňka
  - Stránka
  - Blok
- Typické velikosti

  1 stránka = 8kB

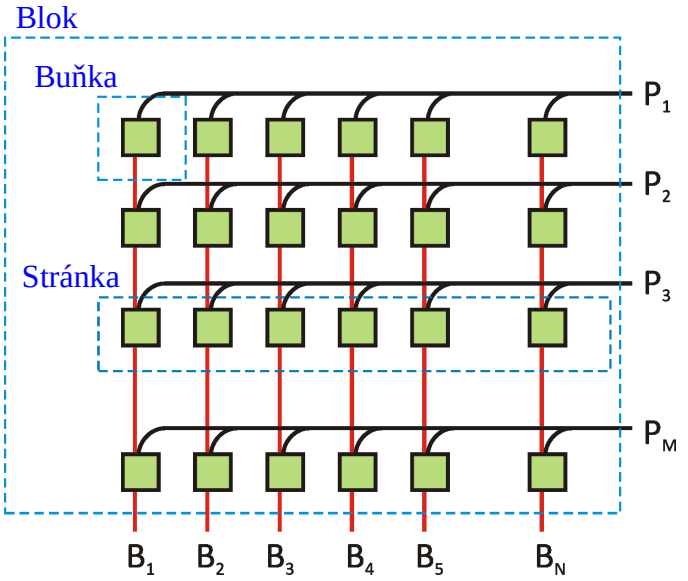
  1 blok = 256 stránek

  tj. blok = 2MB



Základní operace

- Čtení – celá stránka (rychlé)
  - P_m uzemněno, ostatní U_r
  - Vodivost B_n - P_m → 1 or 0
- Zápis – celá stránka (rychlé)
  - P_m nízké napětí U_w
  - B_n uzemněna → náboj je akumulován
  - Zápis může **pouze zvýšit náboj**
- Reset (vybití) – celý **blok** (pomalé)
  - Opačná polarita, vysoké napětí
    ⇨ může ovlivnit sousední buňky

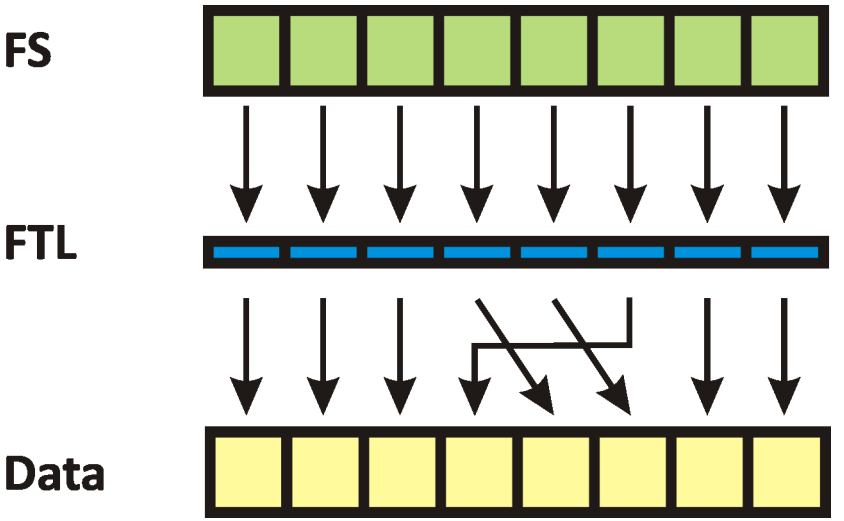


Technické pozadí

Pokročilé operace

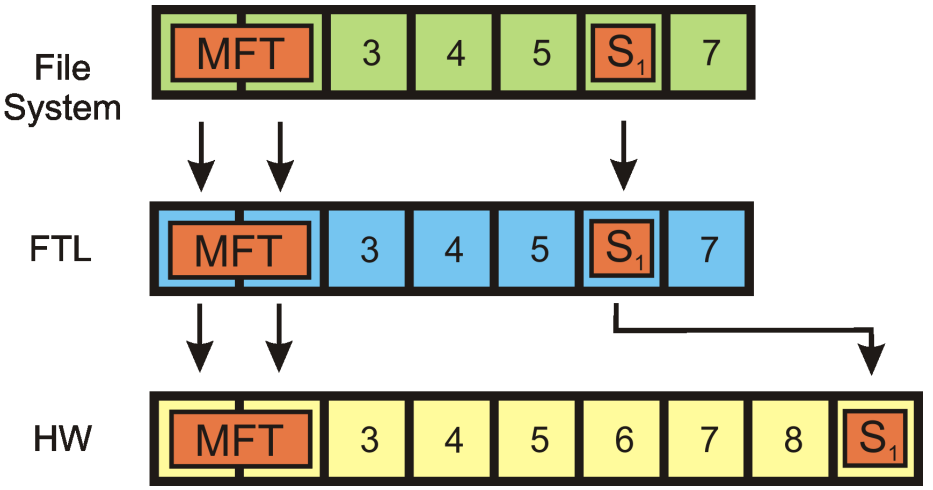
Vyrovnávání opotřebení

- Úroveň opotřebení = počet vybíjecích cyklů
  - Nerovnoměrné (MFT vs. AV soubory, ...)
  - ⇨ různá životnost bloků (nejnižší = životnost SSD)
- Flash Translation Layer (FTL)
  - Indexová tabulka ⇨ změna fyzického umístění
- Vyrovnávací procedura
  - Stejné opotřebení
  - Výběr při zápisu
  - Optimalizace (prohození bloků)



Over provisioning

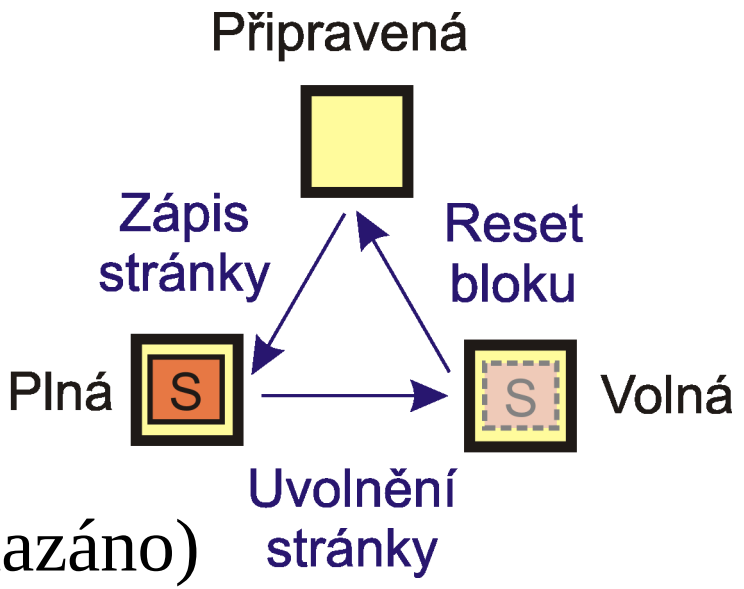
- FTL
  - Ukazatele na data
  - |FS| = |FTL|
  - |HW| >= |FTL|



- Over-provisioning
  - Vyšší interní kapacita
  - Typicky cca 15% (80GB SSD = 96 GB)
- Přínosy
  - Delší životnost (|HW| / |FTL|)
  - “Odkládací prostor” pro vyrovnávání a recyklaci

Recyklace paměťových buněk

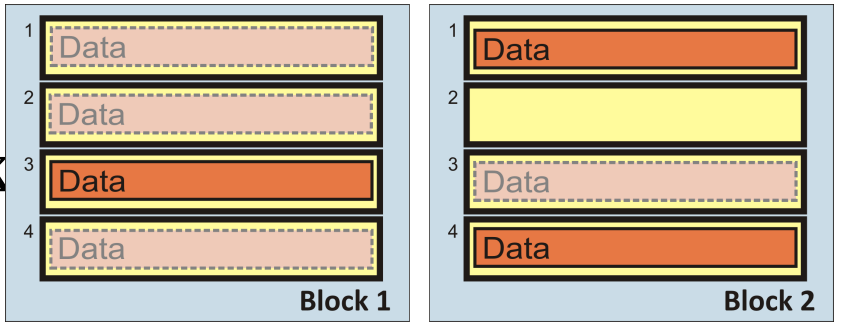
- Tři stavy
  - Připravená/Plná/Volná
- Přechody
  - Reset volných bloků
  - Zápis do stránek
  - Nastavení jako volné (smazáno)



- Smazání dat
  - Na úrovni OS/FS MFT … SSD nemůže zachytit
    → TRIM (protokol – OS posílá info SSD)
  - “Přeuložení” na jiné místo
  - Vlastní analýza MFT

Recyklace paměťových buněk

- Problém částečně volných
  - Zápis, uvolnění = stránk
  - Reset = blok
  - Čekat na celý blok? Deadlock!
  - Přesun stránek jinam = více zápisů

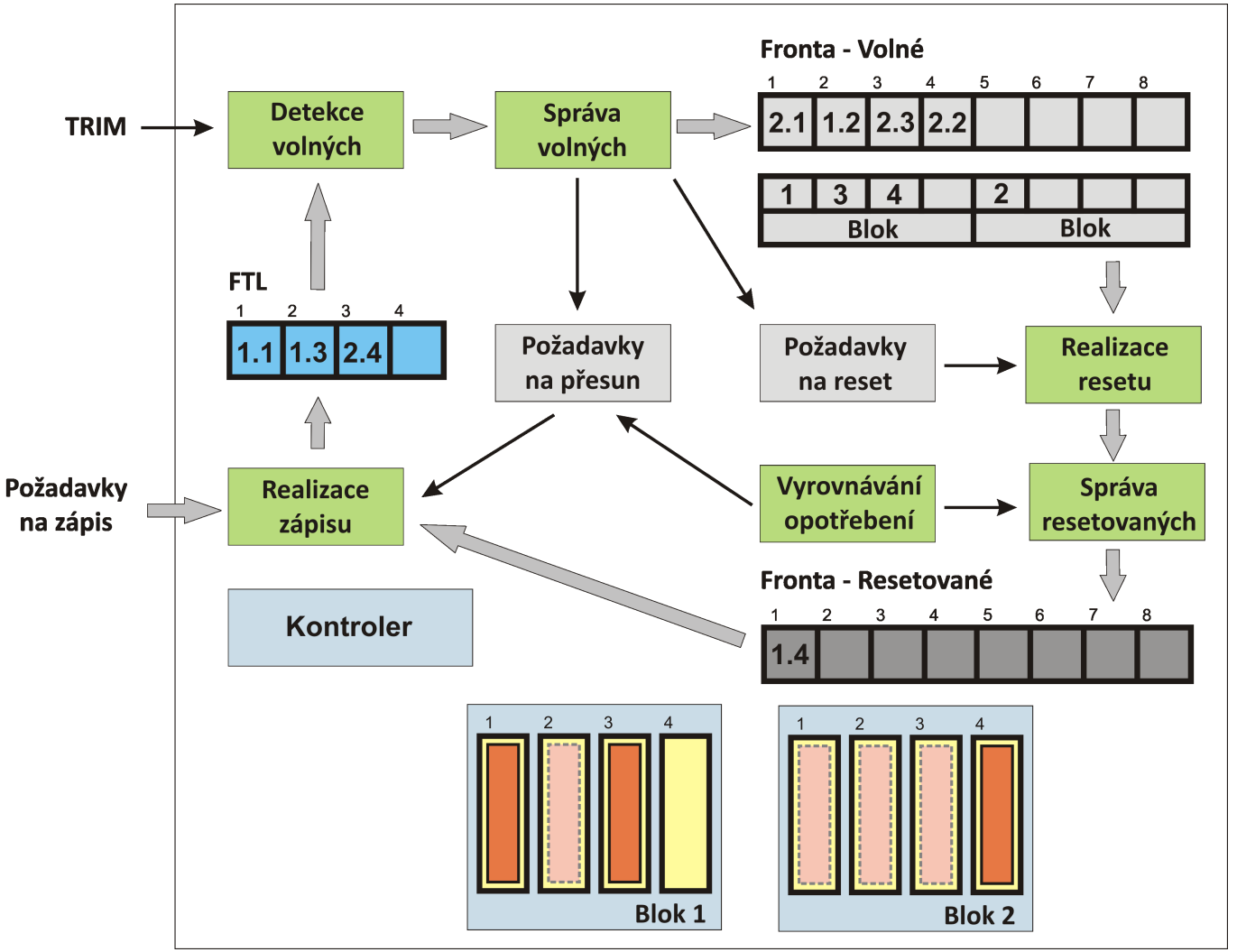


- Koeficient zesílení zápisu (Write Amplification Factor)
  - $WAF = V_{SSD} / V_{FS}$
  - Měl by být co nejmenší
  - Přesun stránek kvůli resetování → WAF > 1,

    … ne nutně (viz později)

Recyklace paměťových buněk

- Garbage kolektor – nezávislá rutina

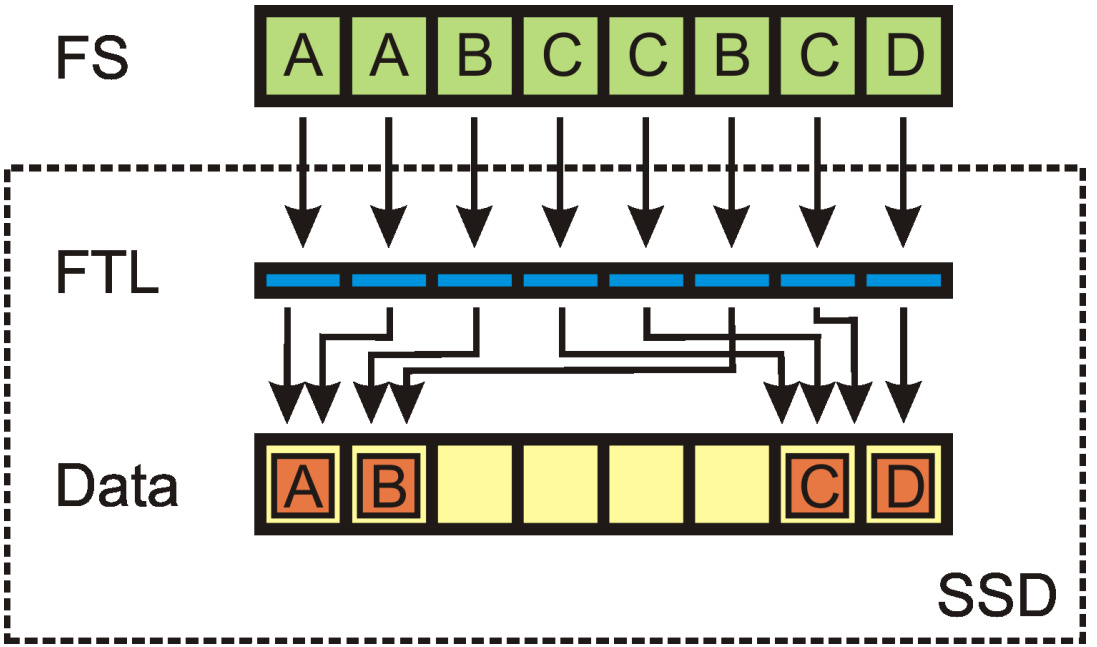


Technické pozadí

Ještě pokročilejší operace

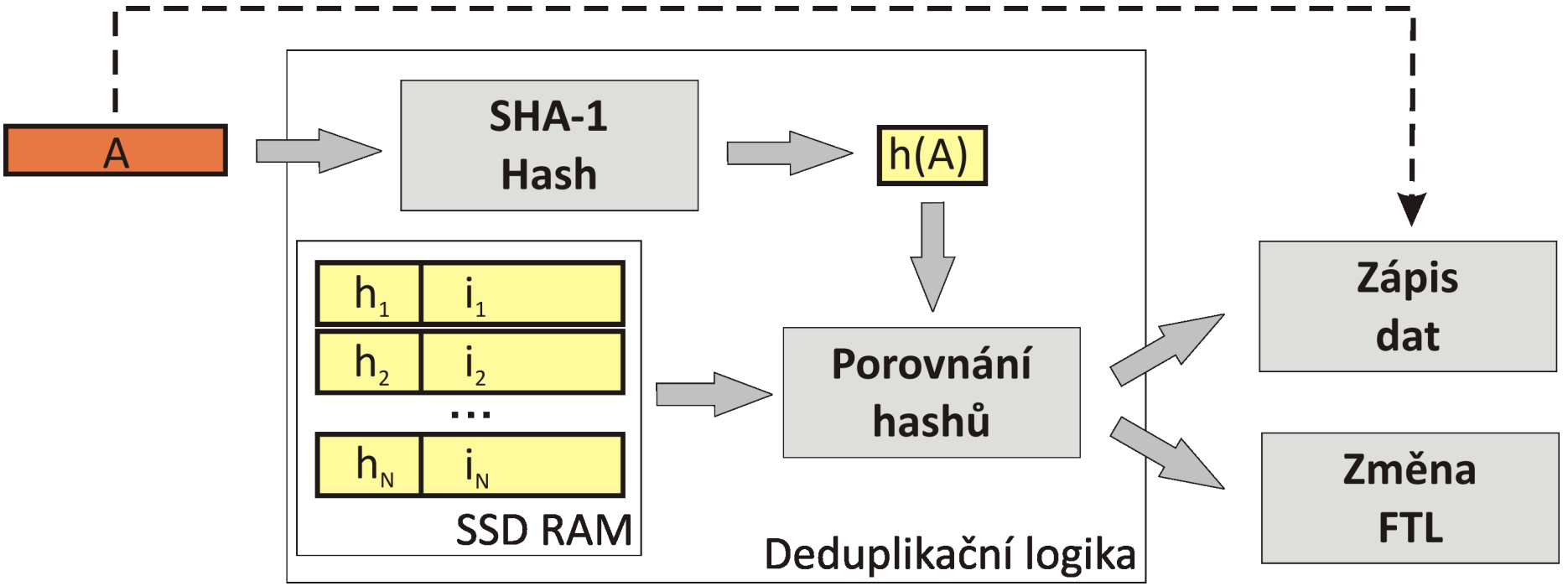
Deduplikace

- FTL = zprostředkovaný (!) přístup k datům
  - Uvnitř SSD = může se stát cokoli
  - Předzpracování dat
- Deduplikace
  - Úroveň stránek
  - Ukazatele N:1
  - Nižší WAF
  - Delší životnost
  - “Rychlejší” zápis



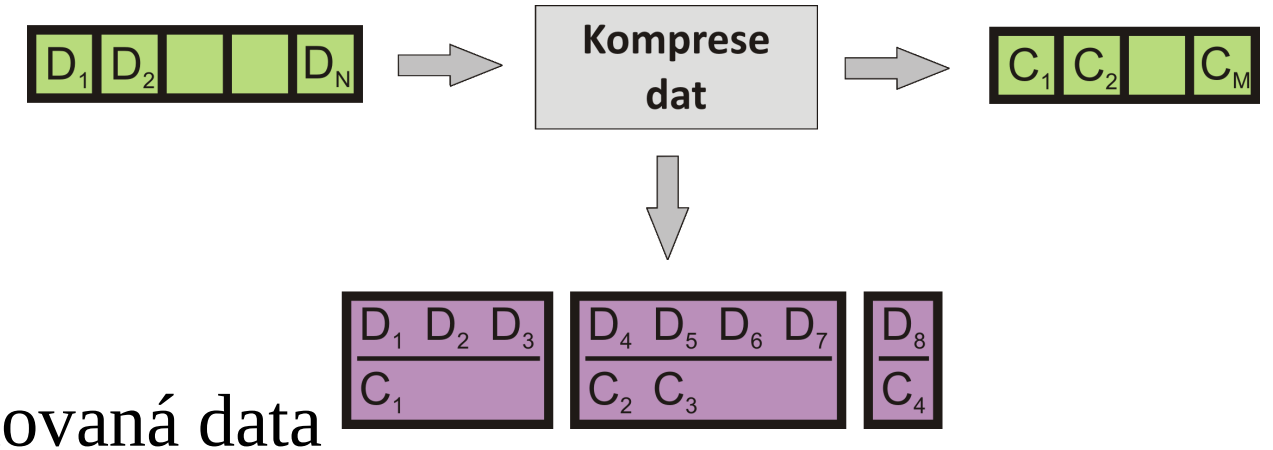
Deduplikace

- Prováděna uvnitř SSD
  - Hash (malá pravděpodobnost kolize)
  - Omezená SSD RAM → částečná deduplikace



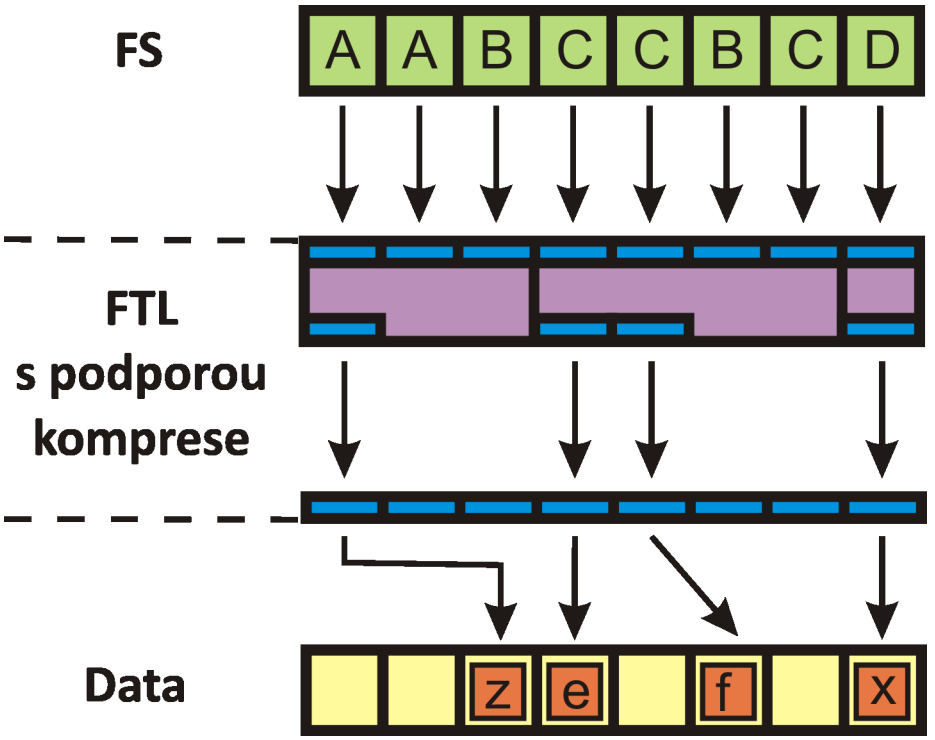
Komprese dat

- Komprese = menší objem dat
- Zápis po celých stránkách
  - 8kB data = 8kB stránka
  - 1kB data = 8kB stránka (žádná úspora!)
    - → Komprese N stránek na M stránek (!)
- Cache = vstupní proud
- Omezené N, M
- Výstup
  - Metadata
  - Komprimovaná data



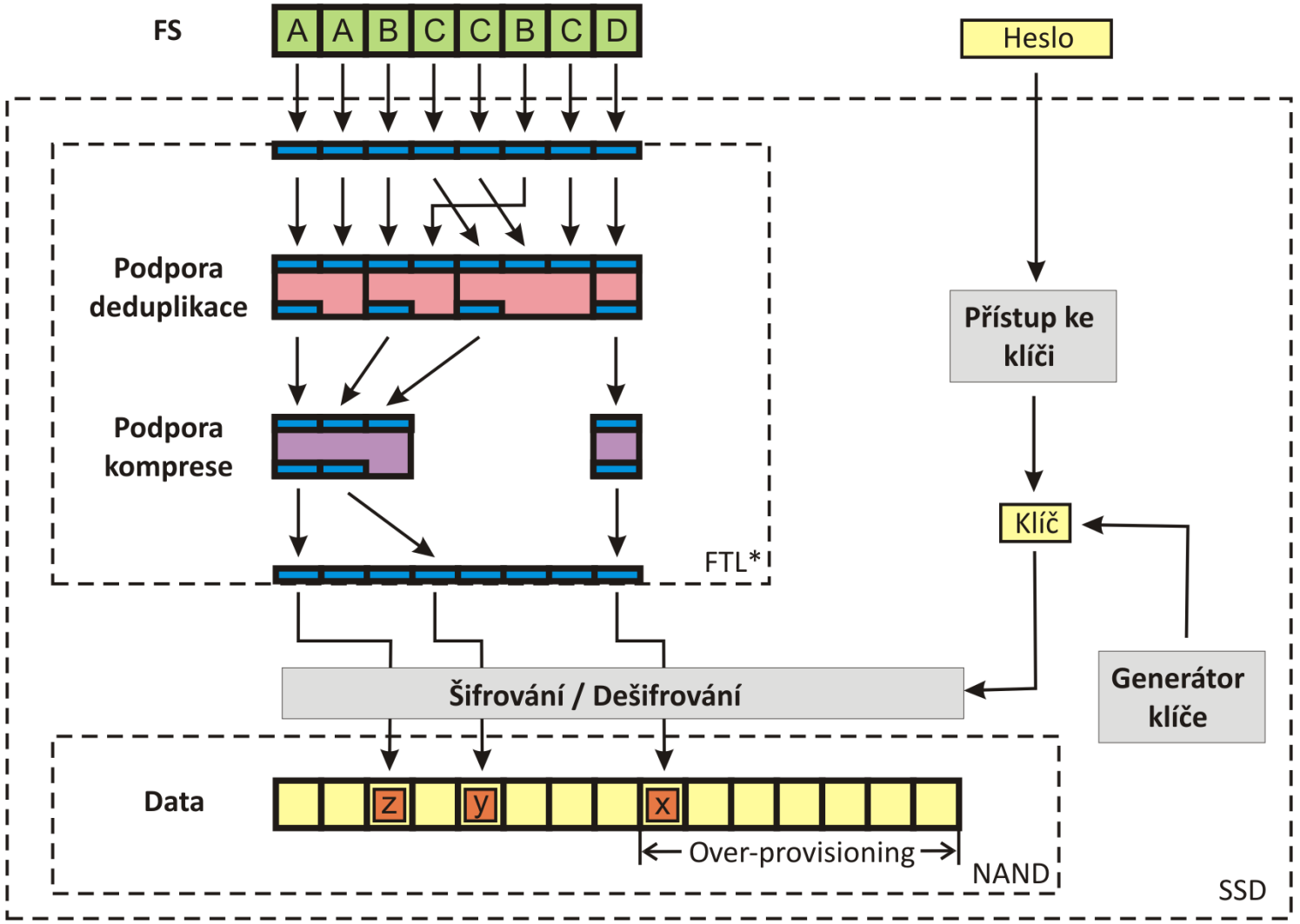
Komprese dat

- Modifikovaná FTL



Šifrování

- Šifrování/dešifrování dat na úrovni stránky



Technické pozadí

Kontroler (UI)

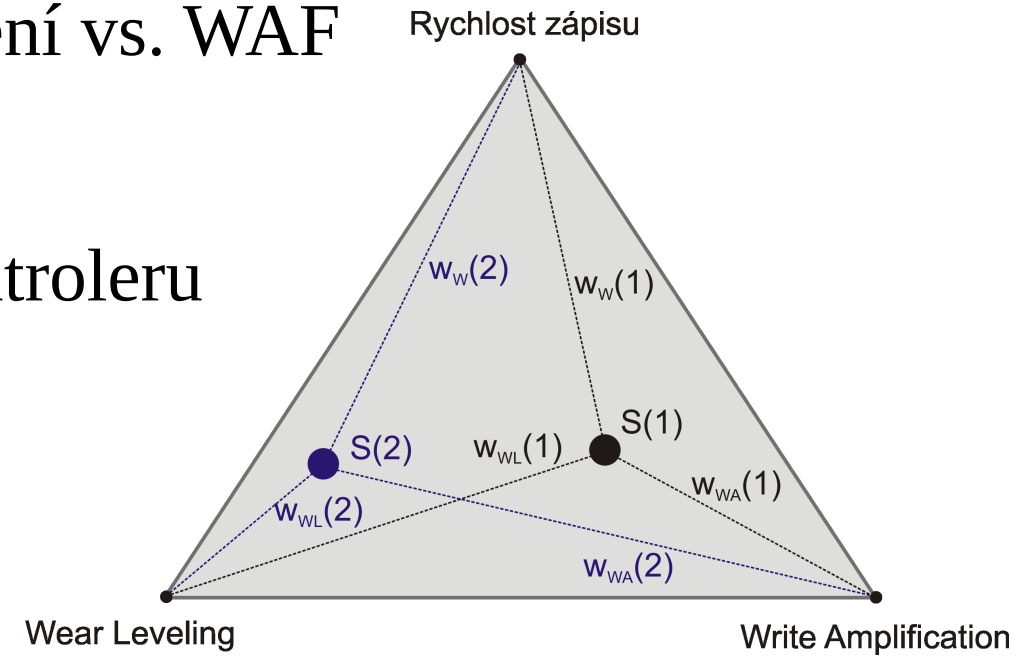
Kontroler

- SSD obsahuje
  - RAM, CPU, I/O, úložiště → vypadá jako počítač
- Kontroler
  - Firmware / Mozek SSD … “umělá inteligence”
- Zajišťuje
  - Maximální rychlost čtení/zápisu
  - Minimální WAF
  - Stejnou úroveň opotřebení
  - ...

→ Využívá (více) pokročilé operace

Kontroler

- Protichůdné požadavky
  - Úroveň opotřebení vs. rychlost zápisu
  - Úroveň opotřebení vs. WAF
- Rozhodnutí
  - Konfigurace kontroleru
  - Aktuální stav (!)
- Celkové kritérium
  - Minimalizace



$$J = \frac{1}{w_W} J_W(A_i) + \frac{1}{w_{WA}} J_{WA}(A_i) + \frac{1}{w_{WL}} J_{WL}(A_i)$$

Technické pozadí

Shrnutí

Kde jsou moje data?

- Data ve flash paměti
  - Fragmentována na stránky
  - Stránky umístěny “náhodně” (FTL)
  - Sdílení některých stránek (Deduplikace)
  - Vybrané množiny stránek = více stránek (Komprese)
  - Stránky mohou být šifrovány (Šifrování)
- Kritická metadata
  - FTL, Deduplikační a kompresní tabulky
  - Heslo/klíč k dešifrování

Jak moje data cestují?

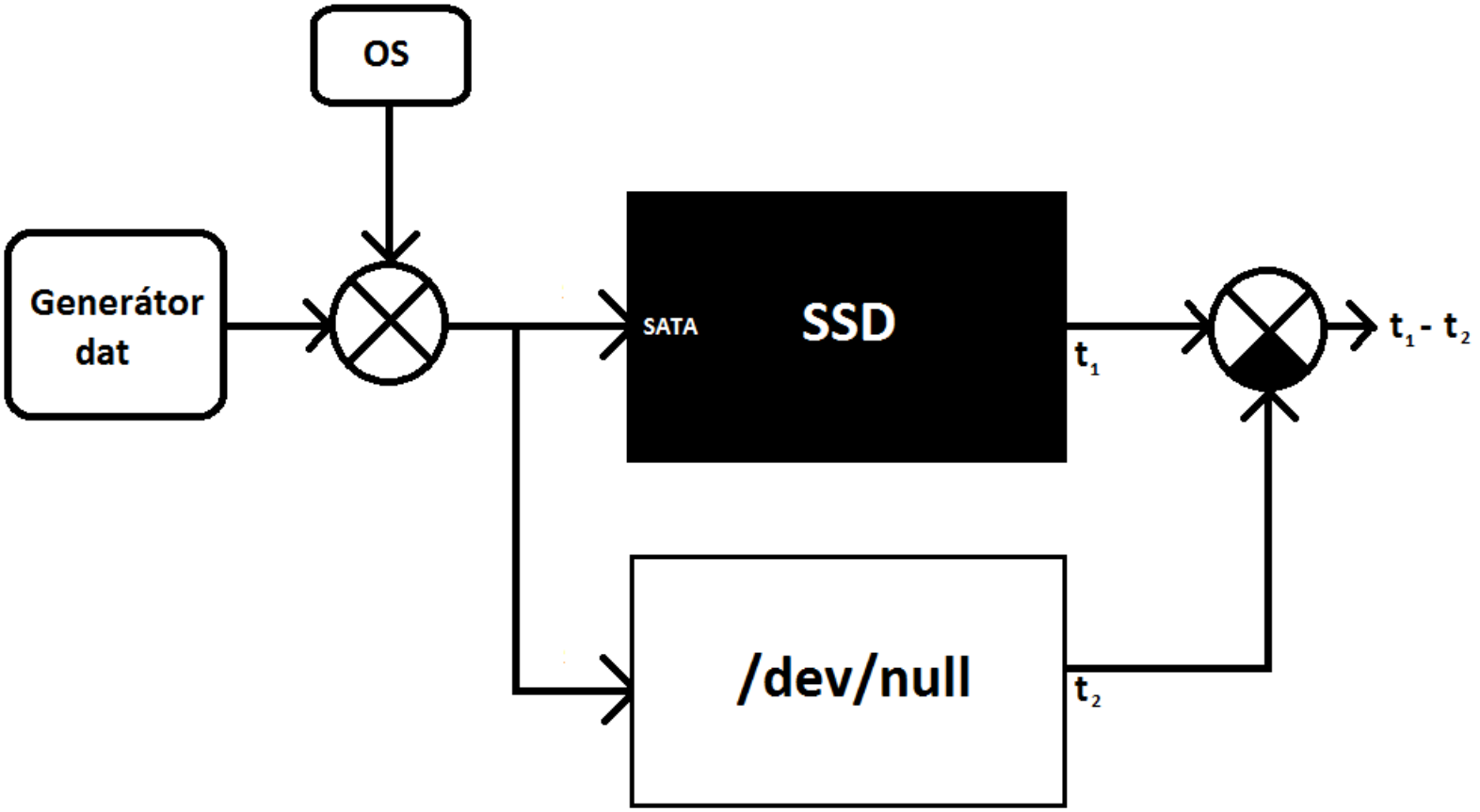
- SSD provádí operace na pozadí
  - Když uzná za vhodné, zapnutý
- Smazaná data mizí (viditelné – uživatel/interface)
  - TRIM
  - Garbage kolektor
  - Problematické pro forenzní obrazy
- Přesuny dat (na fyzické vrstvě)
  - Garbage kolektor
  - Vyrovnávání opotřebení
  - Deduplikační optimalizátor
  - Kompresní optimalizátor
- Závisí na rozhodování kontroleru

Experimenty

Cíle

- Chování SSD ke smazaným/formátovaným datům
  - Obnova smazaných/formátovaných dat
- Identifikace vnitřních rutin SSD
  - TRIM, Garbage Kolektor, Komprese, ...
- Zjištění vnitřních parametrů SSD
  - Velikost stránky, rychlost zápisu, Over-provisioning
- Nezničit hardware
  - Prozatím :-)

Jak zkoumat



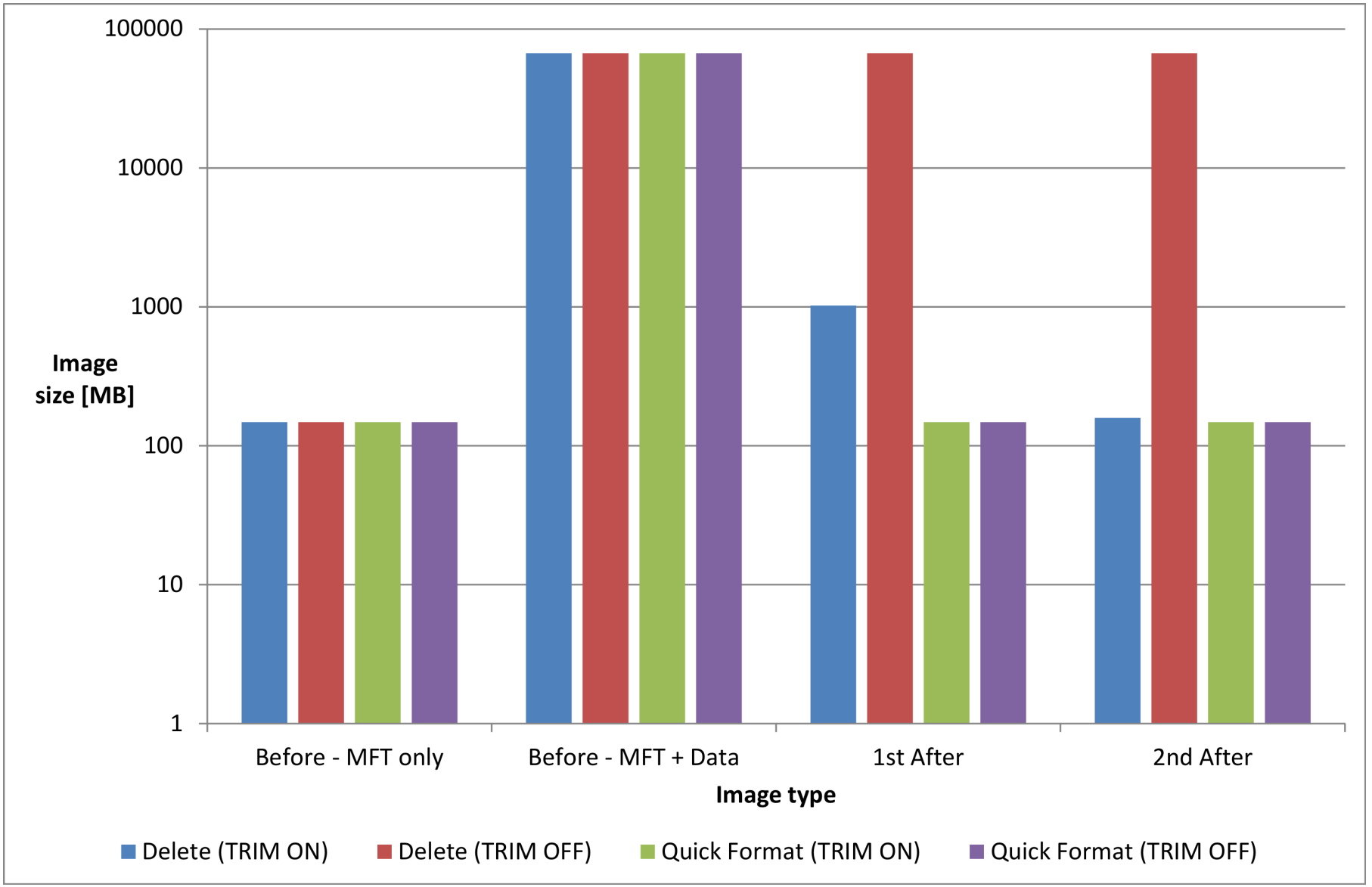
Slepé uličky

- Použití Windows exploreru pro výkonové testy
  - Nereálná rychlost zápisu, ~600MB/sec = vliv cache
- I/O peklo (paralelní zápis na FS)
  - Neúspěšný pokus zahltit SSD Kontroler
  - Identifikována velikost systémové cache v RAM
- Cyklické přepisování (náhodně, nulami)
  - Pokus o zjištění velikosti over-provisioningu
  - Konstantní rychlost

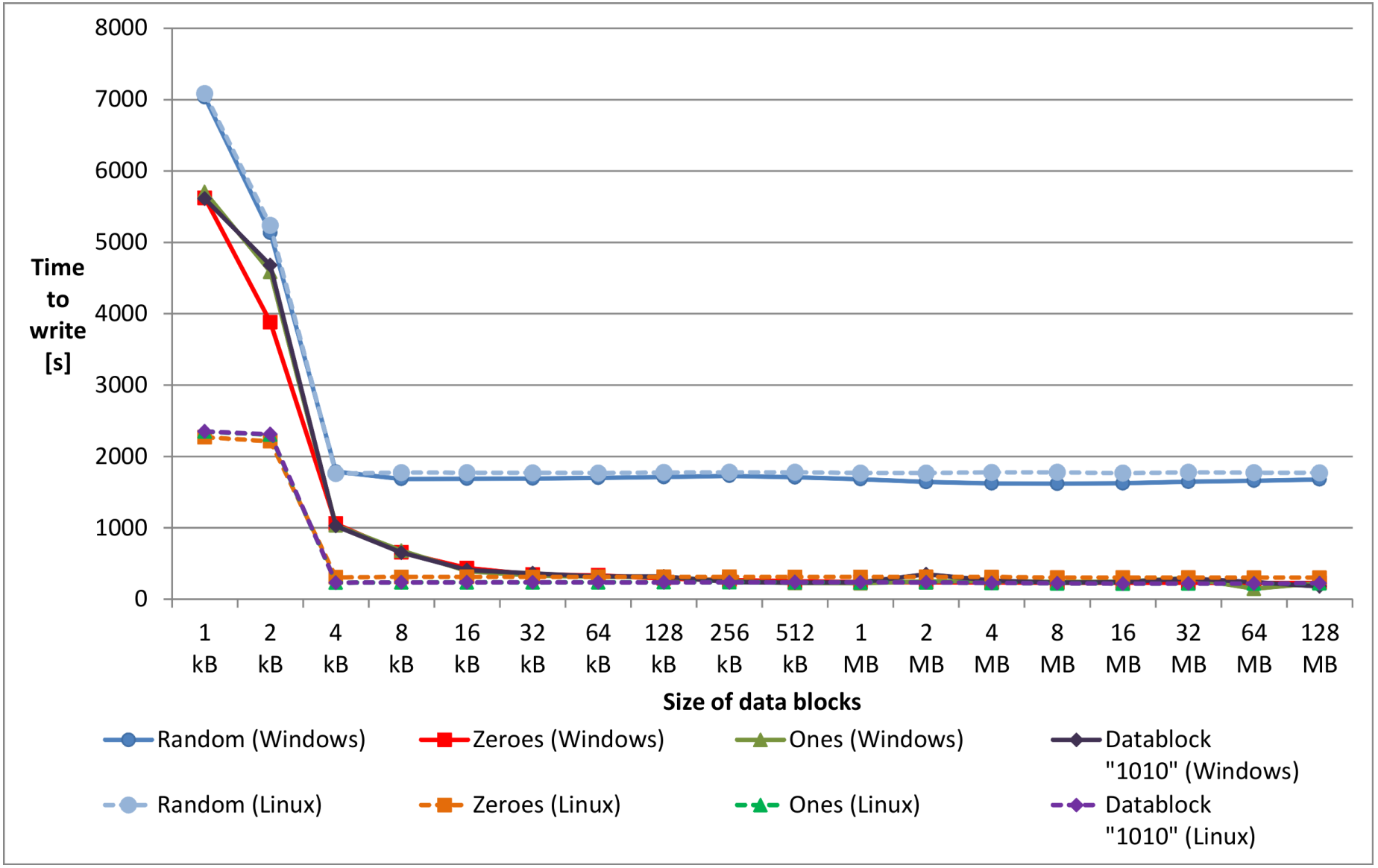
Popis experimentů

- TRIM
  - Jak SSD zachází s daty?
  - TRIM – zapnuto, vypnuto?
  - Windows/Linux (Knoppix)?
- Rychlost zápisu
  - Závislost rychlosti zápisu na velikost stránky
- Deduplikace a komprese
  - Ověřit (ne)přítomnost těchto rutin

TRIM


100000
10000
1000
100
10
1
Image size [MB]
Before - MFT only
Before - MFT + Data
1st After
2nd After
Image type
Delete (TRIM ON)
Delete (TRIM OFF)
Quick Format (TRIM ON)
Quick Format (TRIM OFF)

Velikost stránky



Výsledky

- TRIM ničí data *RYCHLE!*
  - 10 min smazaná, 59 sec rychlý formát
  - TRIM, Linux, Windows – různé chování
- Rychlost zápisu závisí na velikosti zapisovaného bloku
  - Min. alokační jednotka = 4kB = ½ stránky
  - Zvýšení rychlosti pro bloky >=4kB
  - Nejrychlejší zápis pro násobky 4kB
- SSD provádí deduplikaci nebo kompresi
  - Velikost deduplikační/kompresní jednotky ~4-8kb
  - Potřeba dalších experimentů

Shrnutí

Shrnutí experimentů

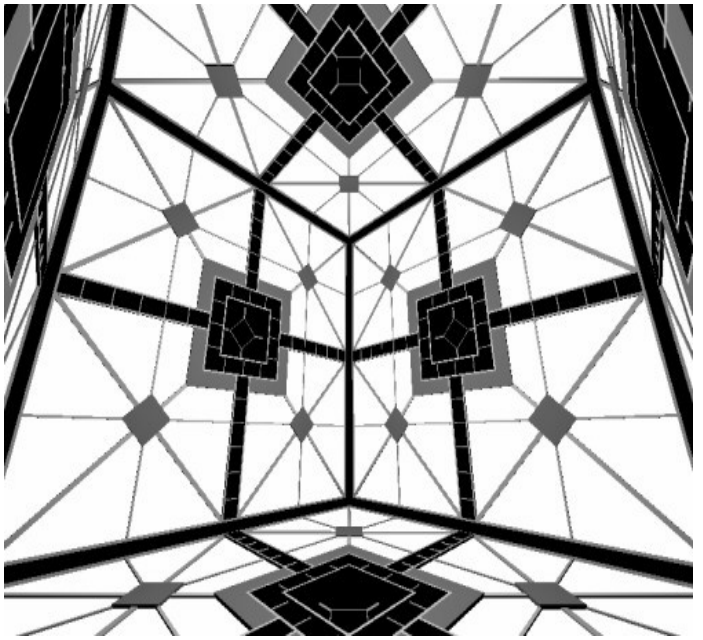
- Typicky: smazáno = ztraceno
  - Windows (7+) – TRIM defaultně zapnutý
  - Smazání + 10 min = 0b (přes SATA)
  - Velká výzva pro obnovu dat a forenzní analýzu
  - „Rychlejší“ analýza (netřeba hledat smazaná data)
- FTL, deduplikace, komprese
  - Na fyzické úrovni = max 4kB fragmenty ...
  - … pravděpodobně komprimovaných dat

Budoucí experimenty

- Rozlišit od sebe deduplikaci a kompresi
  - Vyšší rychlost pro uniformní data, proč?
  - Příprava testovacího prostředí – generátor dat
- TRIM vs šifrování (TrueCrypt)
  - Může TRIM fungovat přes šifrovací vrstvu?
  - Jak?
- FDE vs výkon
  - Jak FDE ovlivní rychlost zápisu a čtení?

Shrnutí

- Co se děje s mými daty?
  - Fragmentována
  - Deduplikována
  - Komprimována
  - (Šifrována)
- A co smazaná data?
  - Zničená reference v FTL
  - Opravdu rychle (max 10 min na celý disk)
  - Obnova nemožná (přes SATA)
  - `fsutil behavior query disabledeletenotify`



^^Hyperkrychle^^

Dotazy

???