淘宝网
Taobao.com

# SSD Internals

伯瑜

2012-11-23

http://weibo.com/pagefault

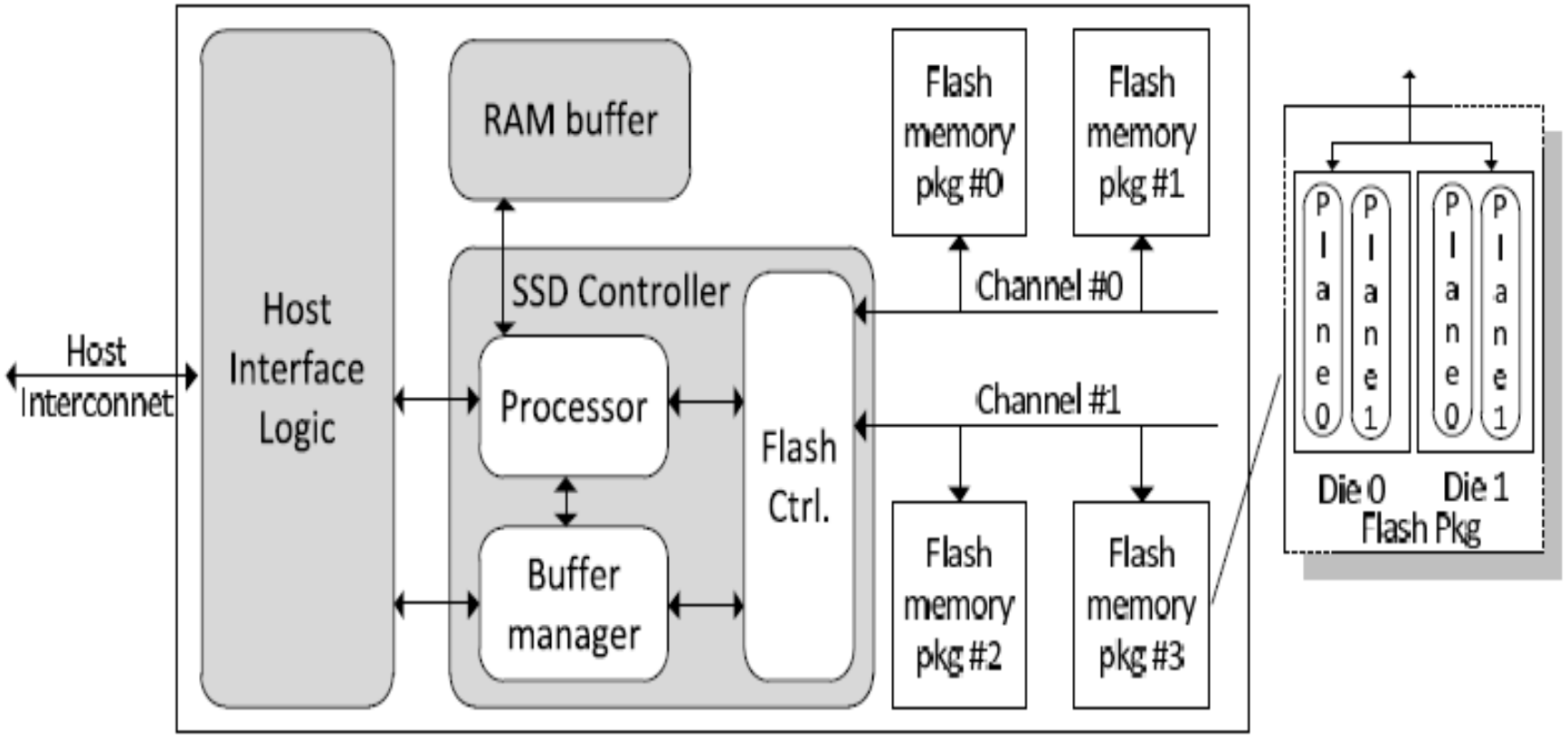
淘宝网 Taobao.com

# 议题

- SSD 的构造
- SSD 的特性
- 对 FTL 行为的分析
- SSD 内部并行的分析

淘宝网
Taobao.com

# Inside SSD



# NAND 芯片性能

| | |
|---|---|
| Page Read to Register | 25$\mu s$ |
| Page Program (Write) from Register | 200$\mu s$ |
| Block Erase | 1.5ms |
| Serial Access to Register (Data bus) | 100$\mu s$ |
| Die Size | 2 GB |
| Block Size | 256 KB |
| Page Size | 4 KB |
| Data Register | 4 KB |
| Planes per die | 4 |
| Dies per package (2GB/4GB/8GB) | 1,2 or 4 |
| Program/Erase Cycles | 100 K |

淘宝网 Taobao.com

# NAND 芯片性能分析

- 读性能很好 ,1 个页需要 200us(100us 读取，传输 +25ns/byte*4096 大约 100us）
- 写速度很慢 300us(200us 写 +100us 传输）
- 如果写有 erase 的参与则更慢（ erase 需要 1.5ms)
- 即使能够 interleave, 单个芯片也只能达到 40MB/s 的读

淘宝网 Taobao.com

# 对 FTL 的要求

- NAND 芯片都一样，关键在 FTL
- 需要有效的屏蔽掉 erase 过程
- 单芯片需要很好的 interleave
- 需要多芯片很好的并行

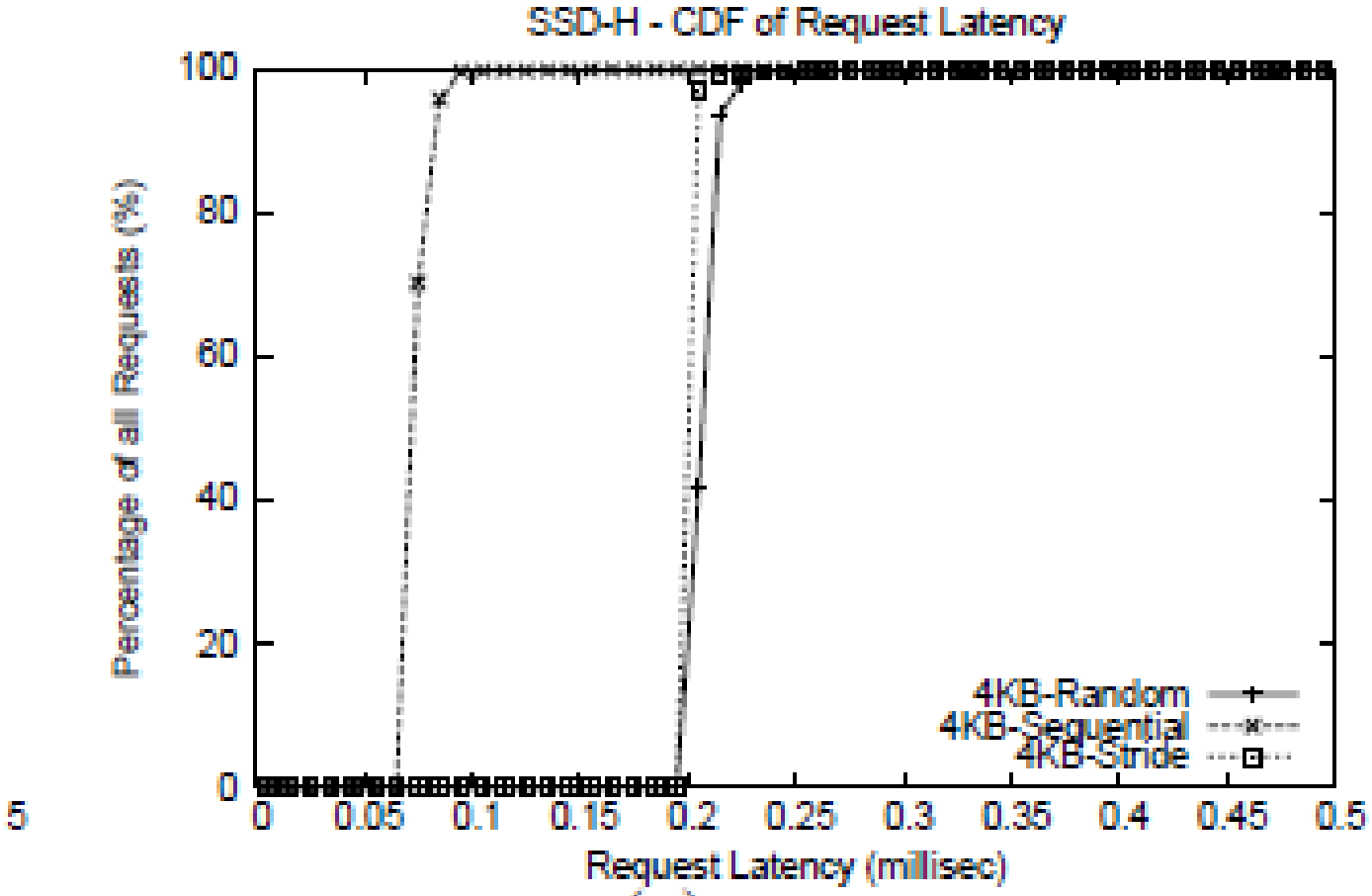
淘宝网 Taobao.com

# FTL 的设计要点

- Logical Block Map
- Cleaning, Garbage Collection
- Parallelism and Interconnect Density
- Persistence, wear leveling

# 对 FTL 行为的分析

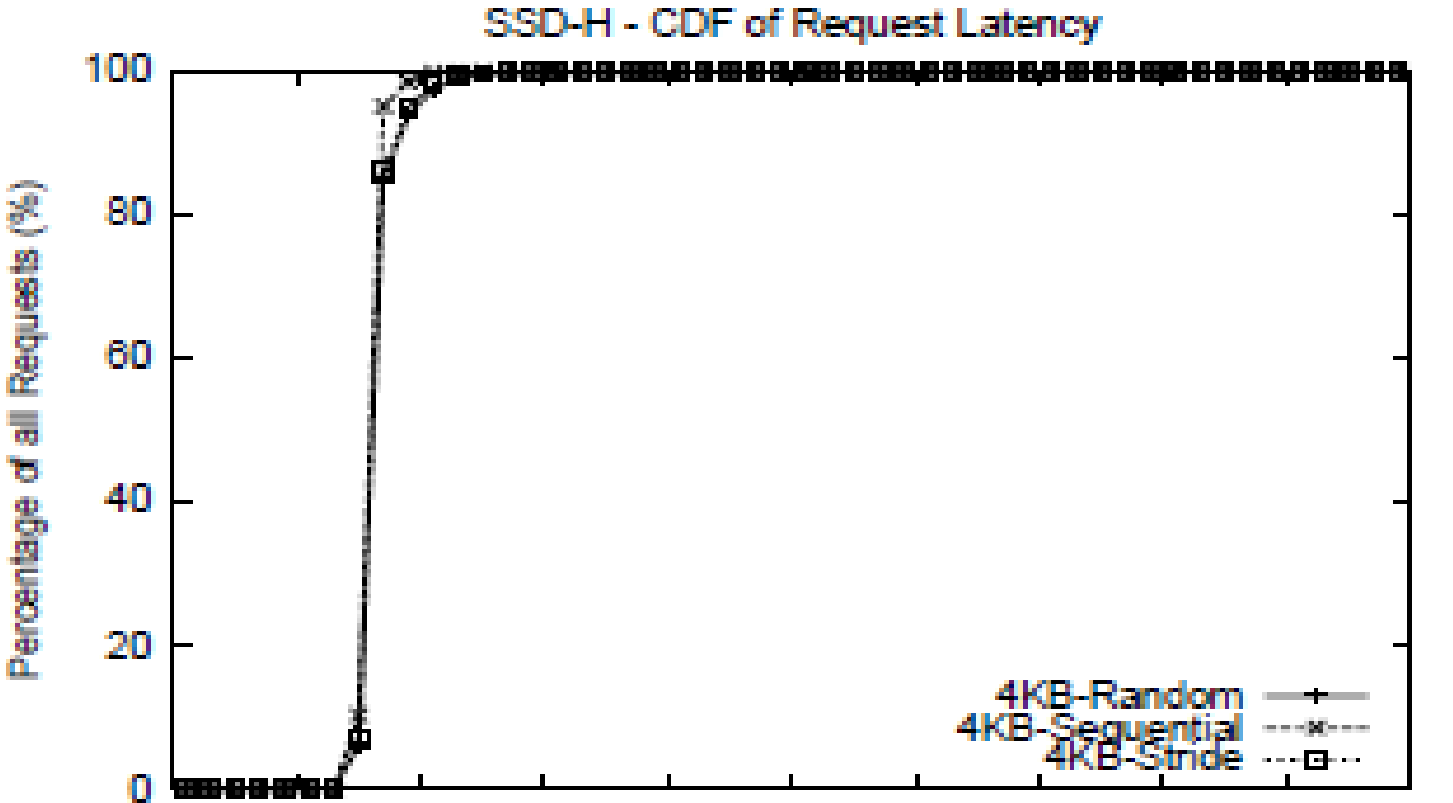
- RANDOM READ vs SEQUENTIAL READ




(c) SSD-H

# FTL 的行为的分析

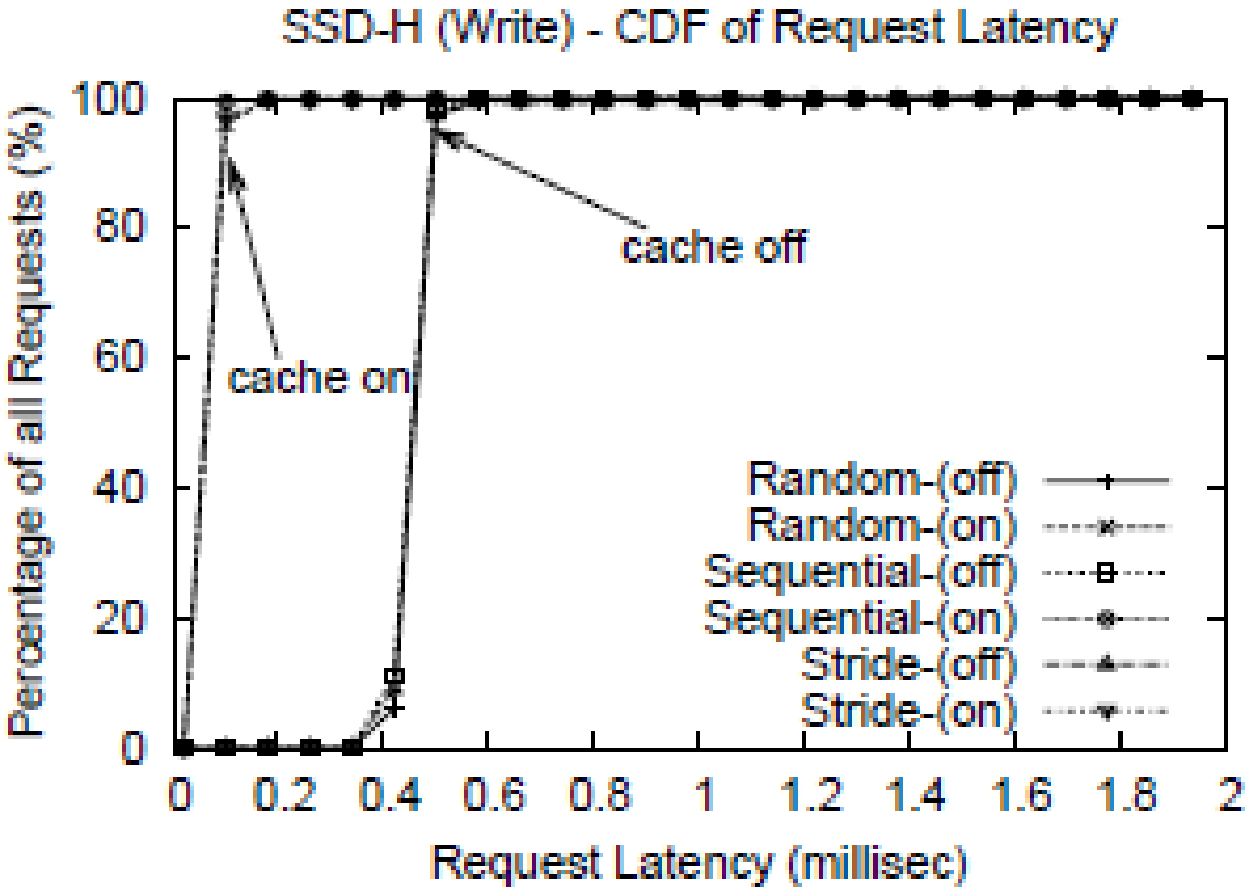
- Sequential write vs. random write




(c) SSD-H

# FTL 的行为的分析

- Disk Cache




(b) SSD-H

淘宝网 Taobao.com

# FTL 的行为的分析

- Fragmentation affects performance

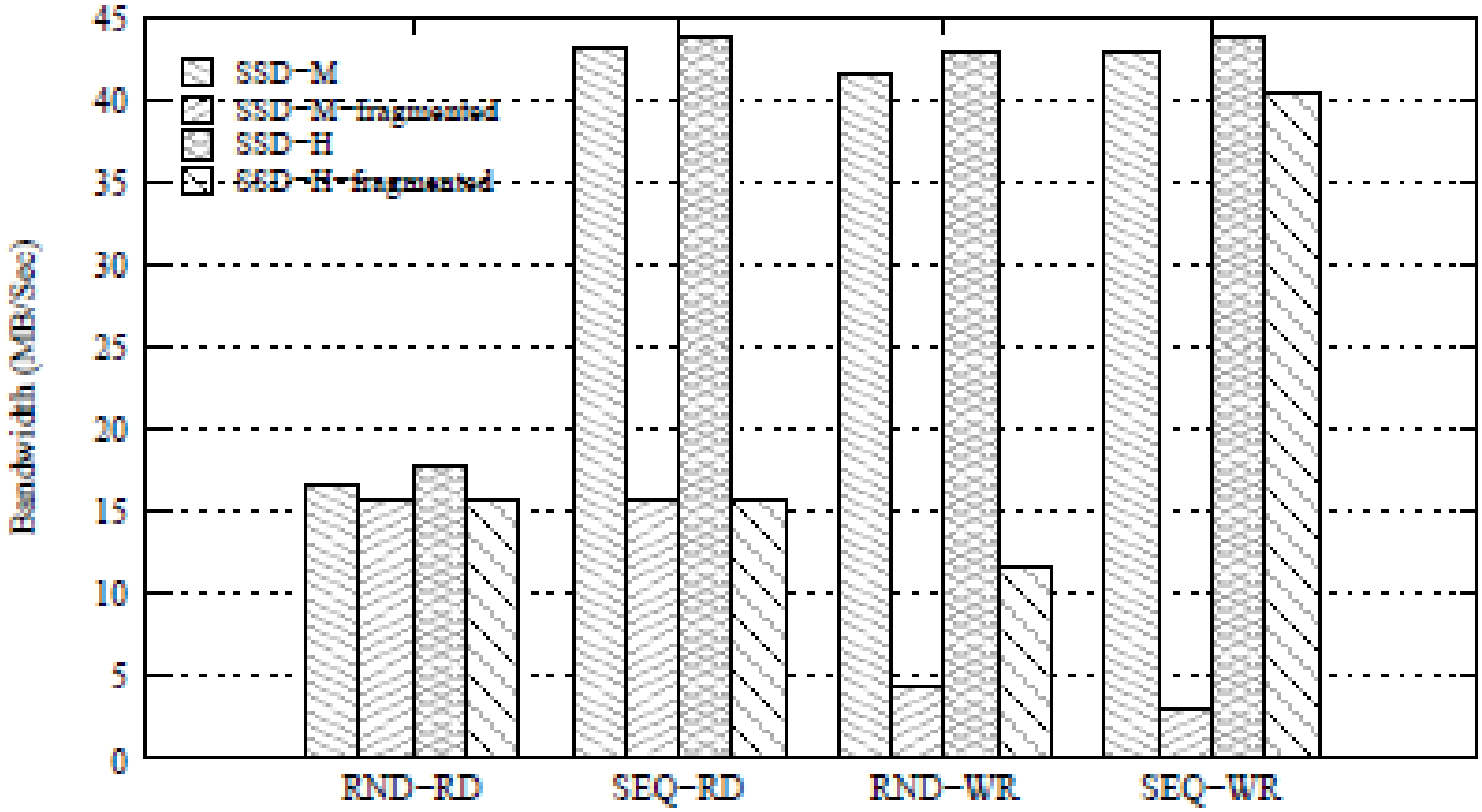
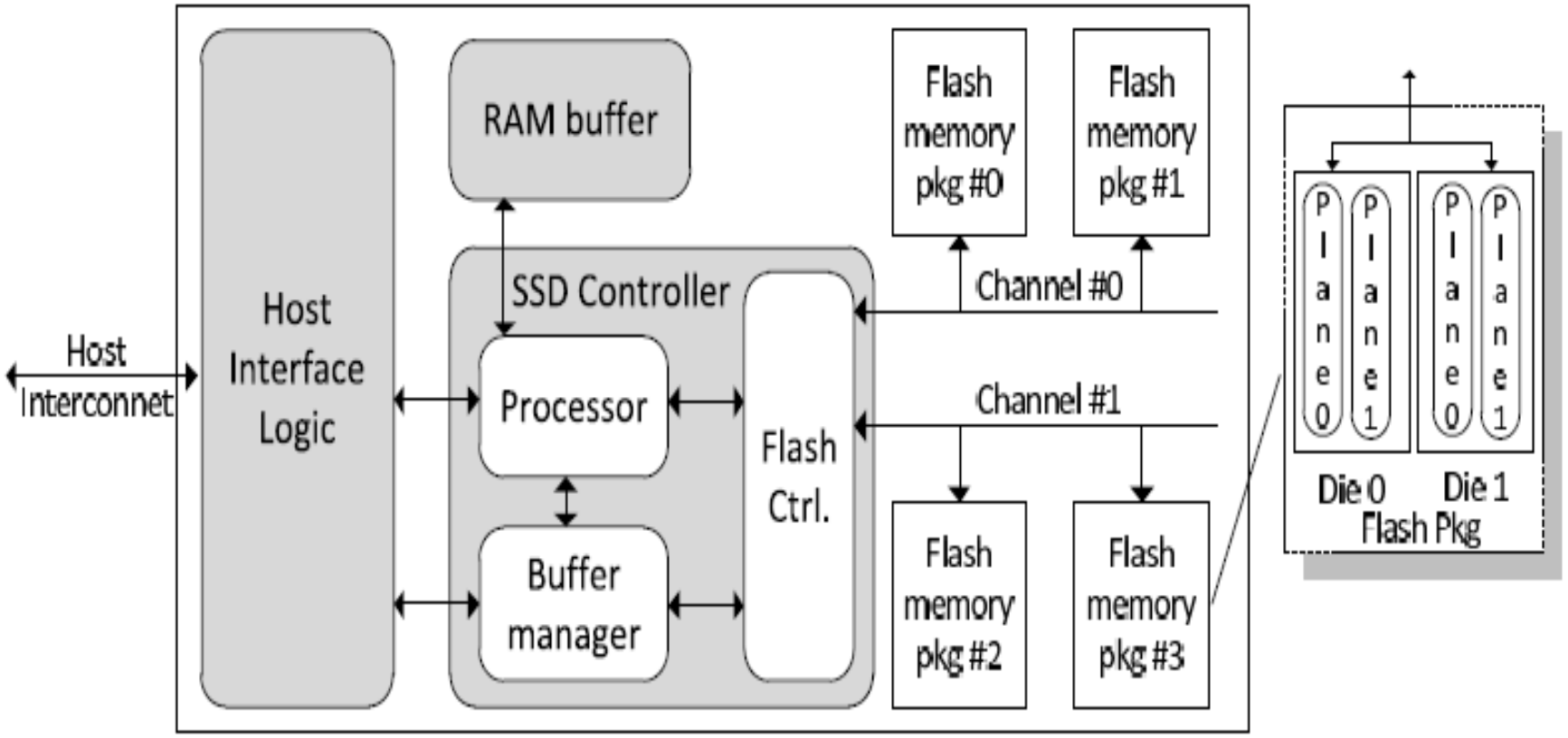



Figure 9: Bandwidth of fragmented SSDs. Four

# SSD 内可以并行的地方

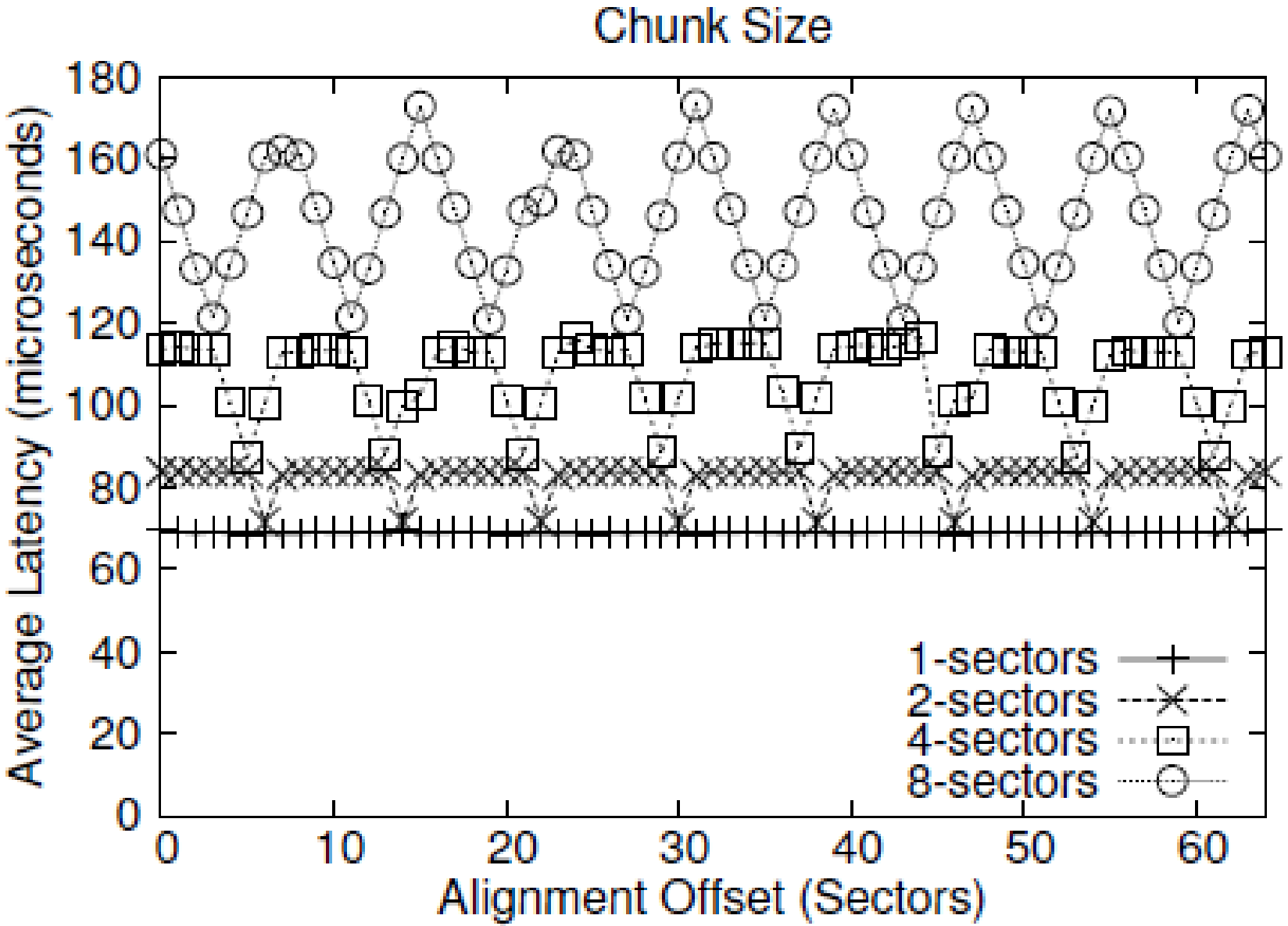


淘宝网 Taobao.com

# FTL 的并行分析

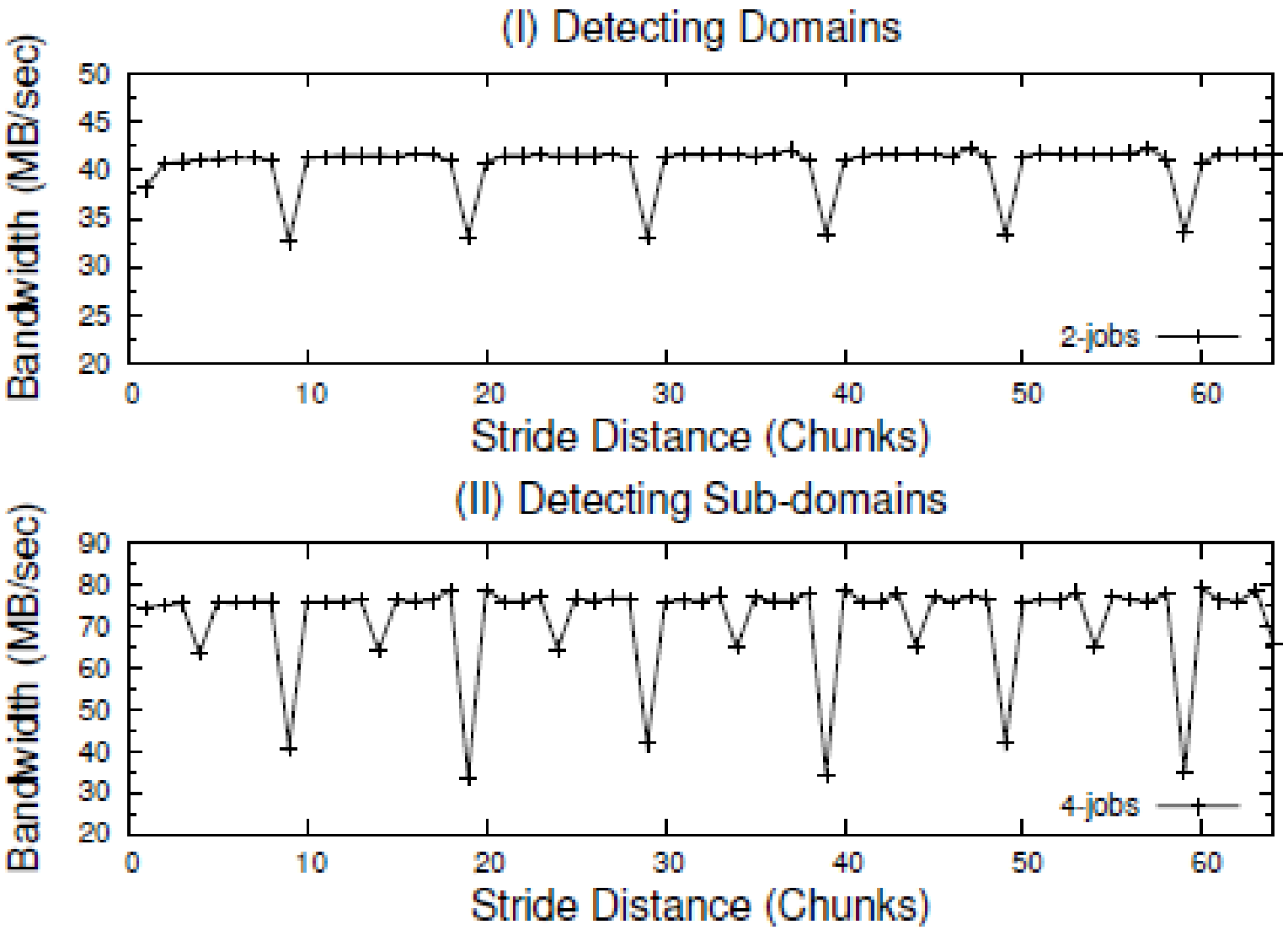
- **Chunk size** – the size of the largest unit of data that is continuously mapped within an individual domain.
- **Interleaving degree** – the number of domains at the same level. The interleaving degree is essentially determined by the redundancy of the resources (e.g. channels).
- **Mapping policy** – the method that determines the domain to which a chunk of logical data is mapped. This policy determines the physical data layout.

# Chunk size = 4KB

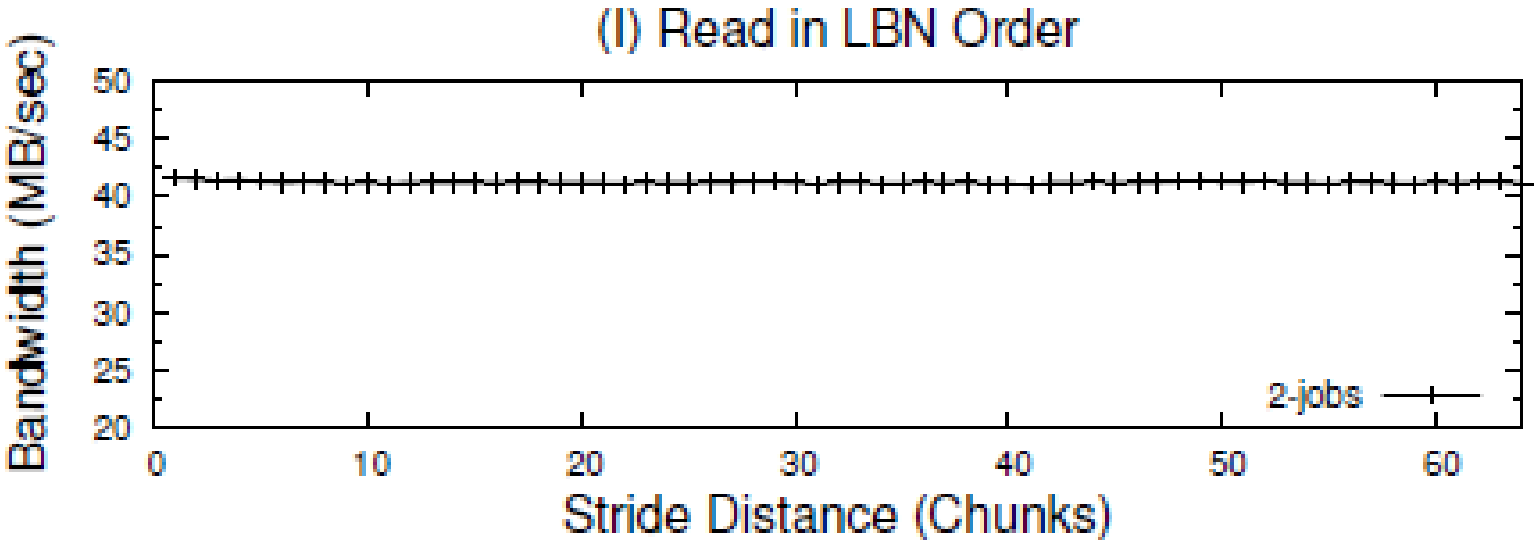


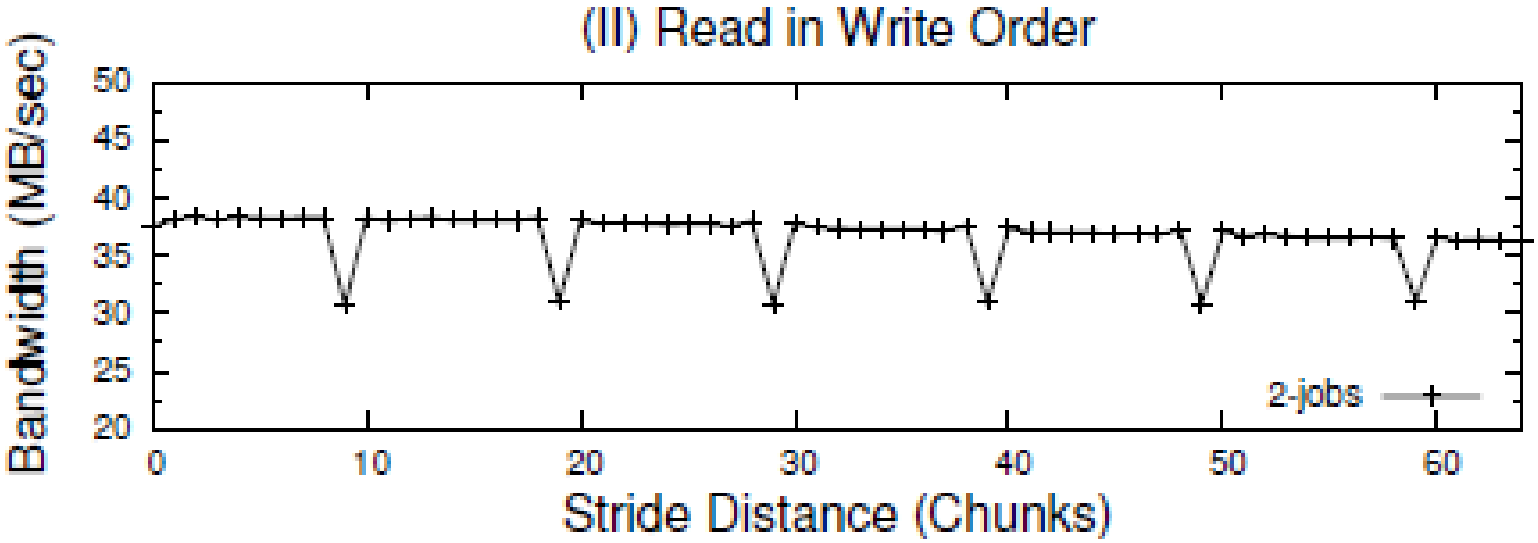
淘宝网
Taobao.com

# Interleave degree

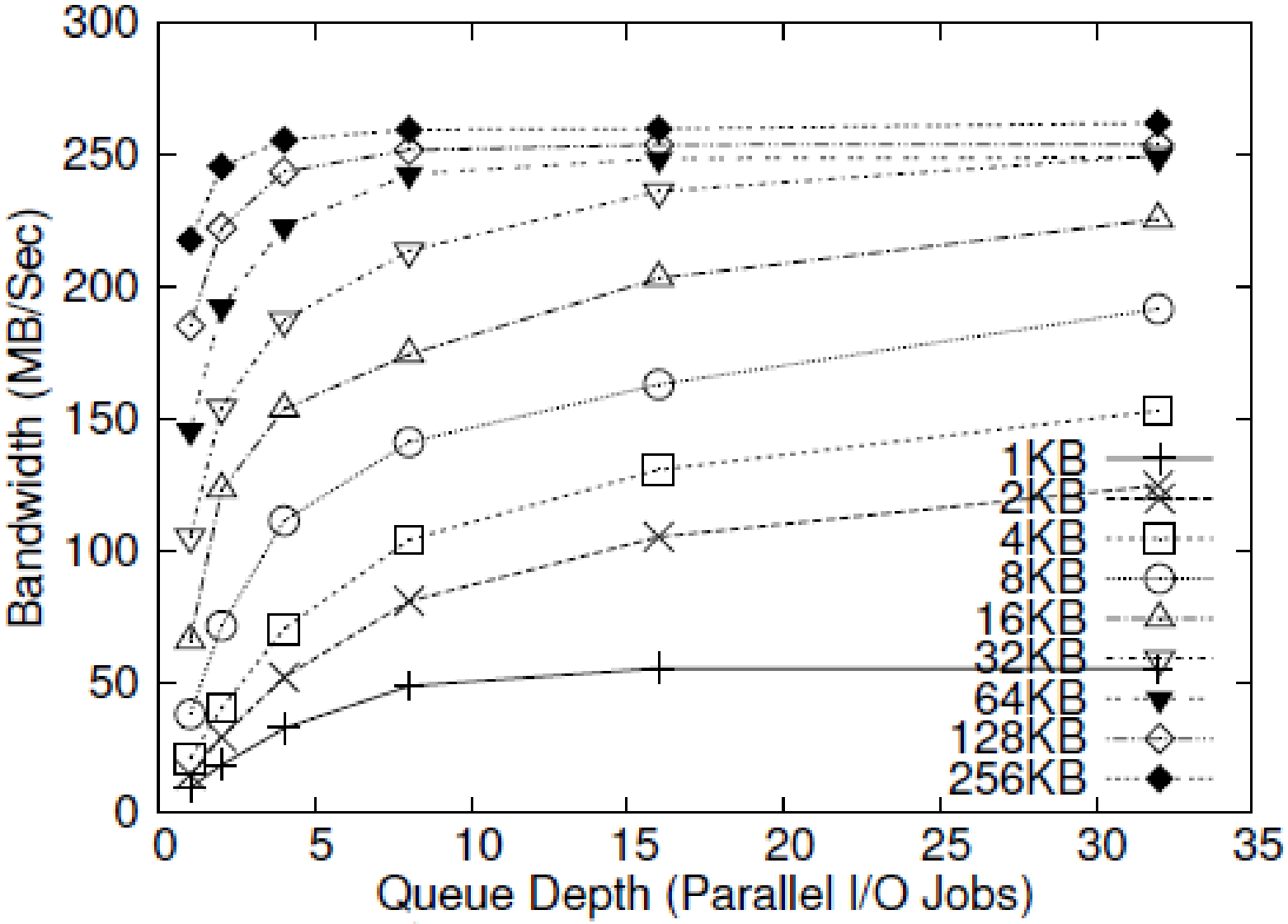


# Mapping Policy



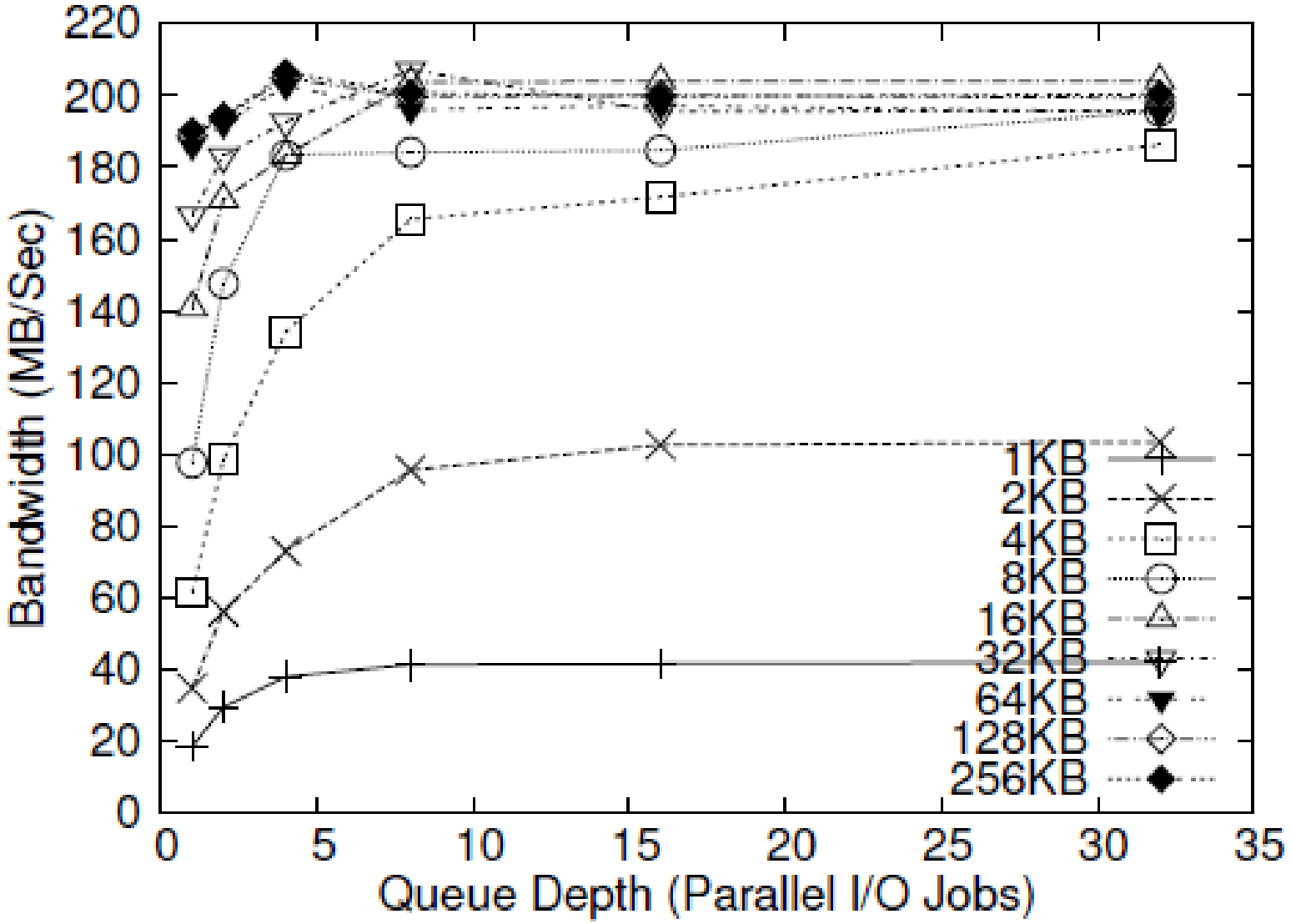



# 并行的好处




(b) SSD-S (Random Read)

# 并行的好处




(c) SSD-S (Sequential Write)

# 并行读写

| | Seq. Write | Rnd Write | None |
|---|---|---|---|
| Seq. Read | 109.2 | 103.5 | 72.6 |
| Rnd. read | 32.8 | 33.2 | 21.3 |
| None | 61.4 | 59.4 | |

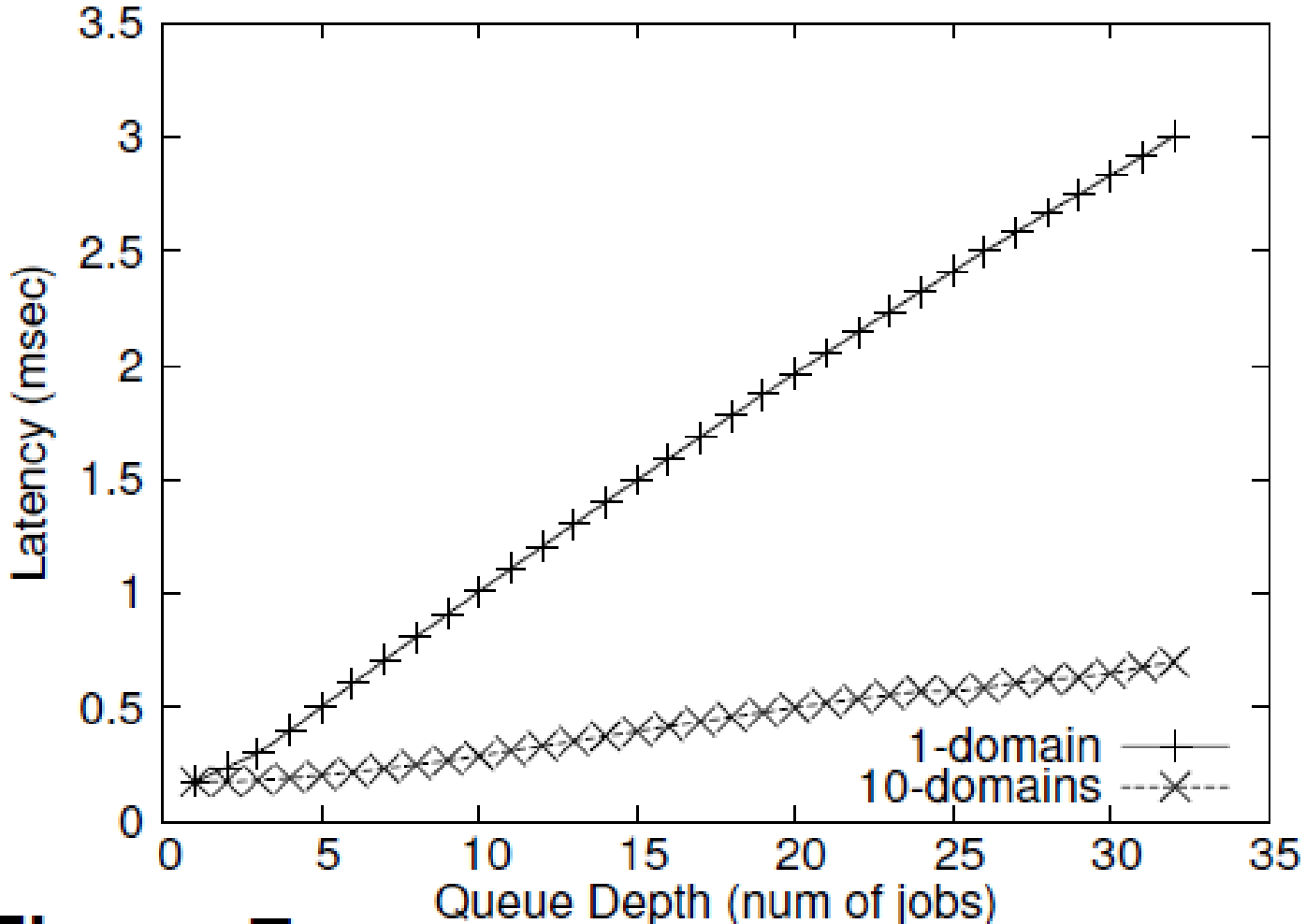
读写干扰严重， randread+seqwrite 条件下：

write 61.4->13.4

read 21.3->19.4

# 并行分析

- ill-mapped data layout impact read



淘宝网 Taobao.com

# 参考论文

- Design Tradeoffs for SSD Performance
- Understanding Intrinsic Characteristics and System Implications of Flash Memory based Solid State Drives
- Essential Roles of Exploiting Internal Parallelism of Flash Memory based Solid State Drives in High-Speed Data Processing