
# A Case for Flash Memory SSD in OLTP Applications

## One FlashSSD can beat Ten 15K RPM HDDs
### - Performance, Price, Power -

Sang-Won Lee

http://icc.skku.ac.kr/~swlee

(Joint Work: Bongki Moon@Univ. of Arizona, Tony Park@Indilinx)


# Flash Memory Markets

- Mobile devices
- PC, Laptop
- Enterprise server storages

# A Gloomy Marketing Strategy

- One SSD vs. One Harddisk
  - e.g One SSD's Capacity: 32 GB, 80 GB
- Someone says that "SSD can penetrate the market only when it matches HDD price"
  - Partially true in PC / Laptop market
- Under this strategy, the market would be invulnerable to SSD

# Three Truths / Myths on SSD

- SSD is expensive!
- SSD's power consumption is non-trivial!
- Write performance is problematic!

# Motivations

- "FlashSSD's message is still unclear in the market"
  - [Personal Communication] Ken Salem, University of Waterloo
- It is urgent to develop "the case for flash memory SSD" (or killer applications) and "the right message"
- Debunk those myths on SSDs
  - From OLTP Perspectives

# IOPS Crisis in OLTP(1): with HDDs

IBM for TPC-C (2008 Dec.)

- 800 TB Storage
  - 11K 73.4GB disks (each 15k rpm)
- Total cost: 35M $
  - Server HW: 12M $
  - Server SW: 2M $
  - Storage: 20M $
  - Client HW/SW: 1M $
  - To improve IOPS
- They buy IOPS, not capacity!

IBM | IBM Power 595 Model 9119-FHA | TPC-C Rev. 5.9 | Report Date: June 10, 2008

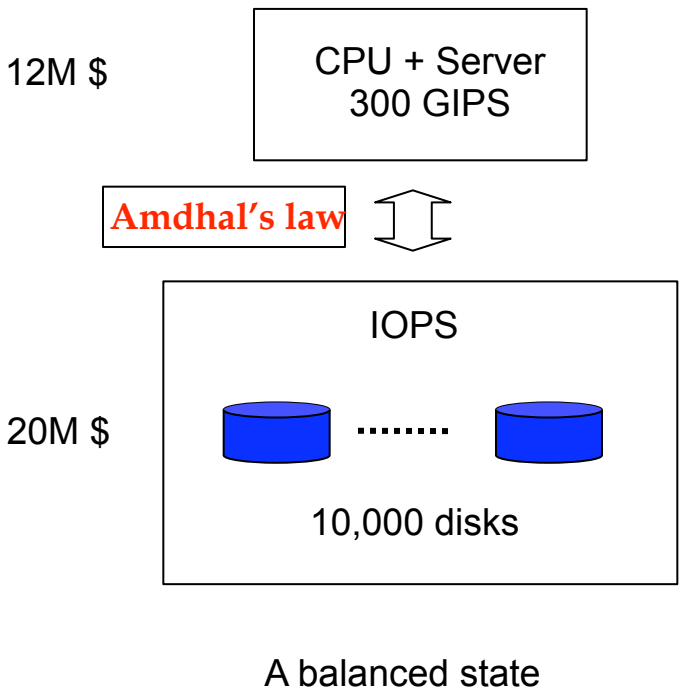
| Total System Cost | TPC-C Throughput | Price/Performance | Availability Date |
|---|---|---|---|
| $17,111,788 USD | 6,085,166 | $2.81 USD | December 10, 2008 |

| Database Server Processor Chip/Core/Thread | Database Manager | Operating System | Other Software | No. Users |
|---|---|---|---|---|
| 32/64/128 POWER6 5.0GHz | DB2 9.5 | AIX 5L V5.3 | Microsoft Visual C++ Microsoft COM+ | 5,184,000 |

| System Components | Server | | Each of the 128 Clients | |
|---|---|---|---|---|
| | Quantity | Description | Quantity | Description |
| Processors Chips /Cores/Threads | 32/64/128 | 5.0GHz POWER6 | 1/2/2 | 2.0GHz 4MB L2 Xeon Processor |
| Memory | 64 | 64GB | 2 | 512 MB |
| Disk Controllers | 1<br>68<br>68 | SAS Controller<br>4Gb FC Adapters<br>IBM System Storage DS4800 | 1 | SAS Controller |
| Disk Drives | 8<br>10,992 | 146.8GB 15K RPM SCSI<br>73.4GB 15K RPM 4Gb FC | 1 | 73.4GB 15K RPM SAS |
| Total Storage | | 746,467GB | | 67.86GB |
| Terminals | 1 | System Console | 1 | System Console |

# IOPS Crisis in OLTP(2): with HDDs

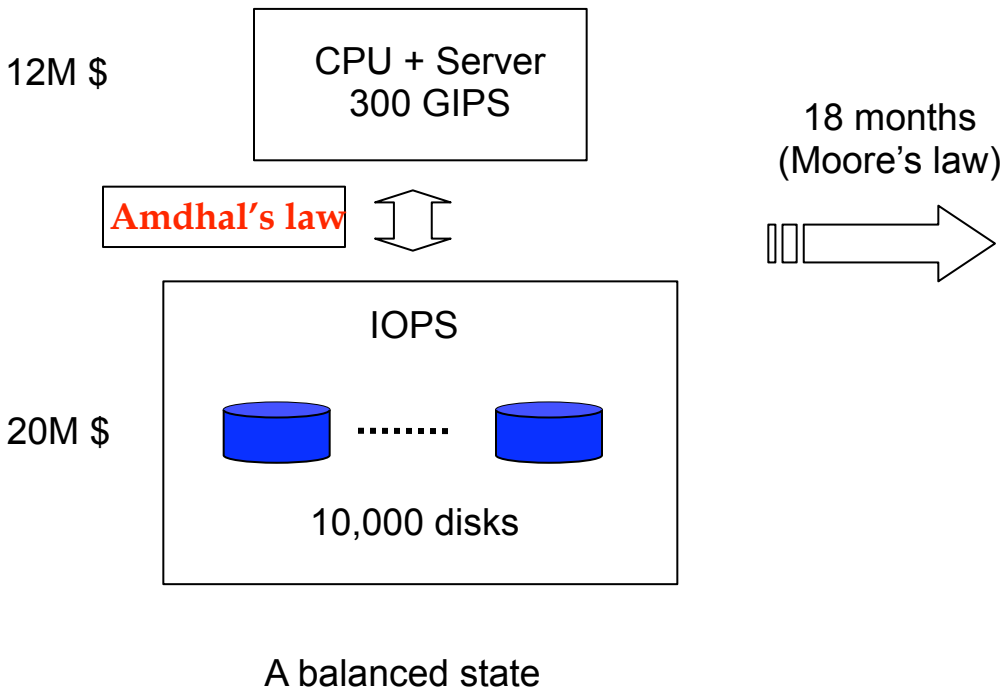
- For balanced systems, OLTP systems pay huge $$$ on disks for high IOPS; IOPS crisis would be worse and worse




A balanced state

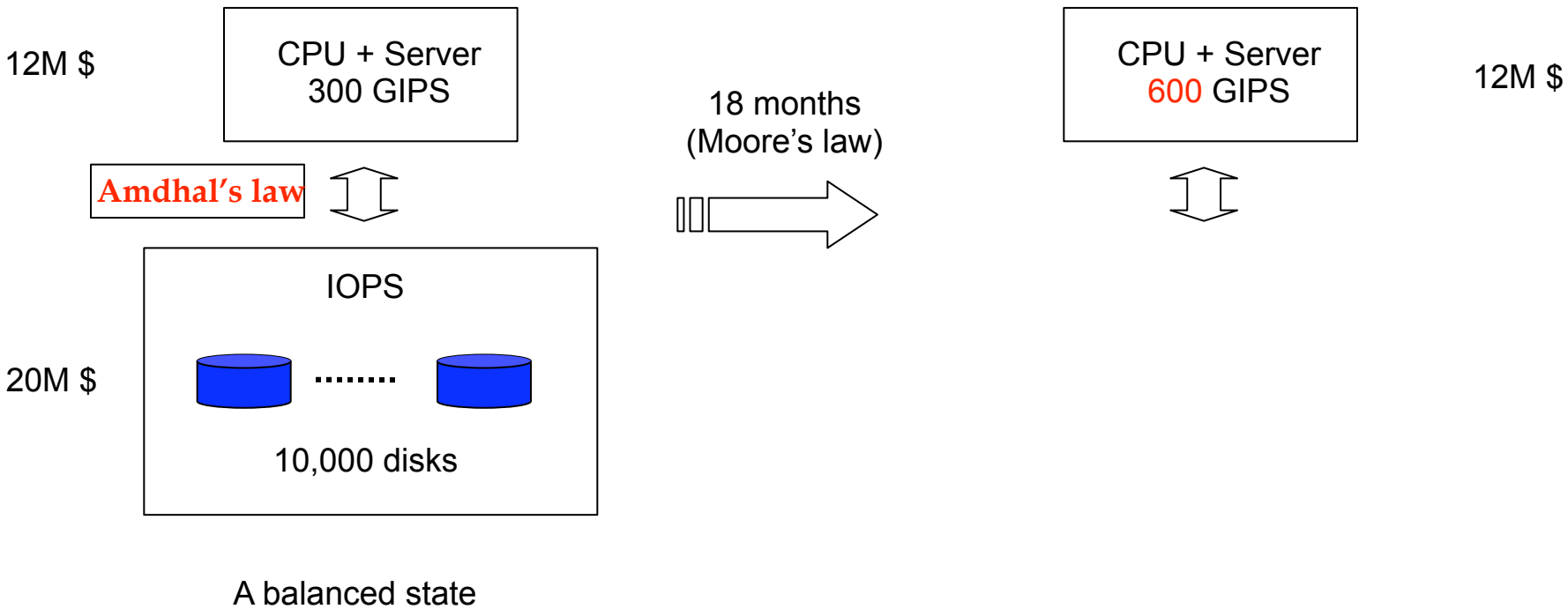
# IOPS Crisis in OLTP(2): with HDDs

- For balanced systems, OLTP systems pay huge $$$ on disks for high IOPS; IOPS crisis would be worse and worse



# IOPS Crisis in OLTP(2): with HDDs

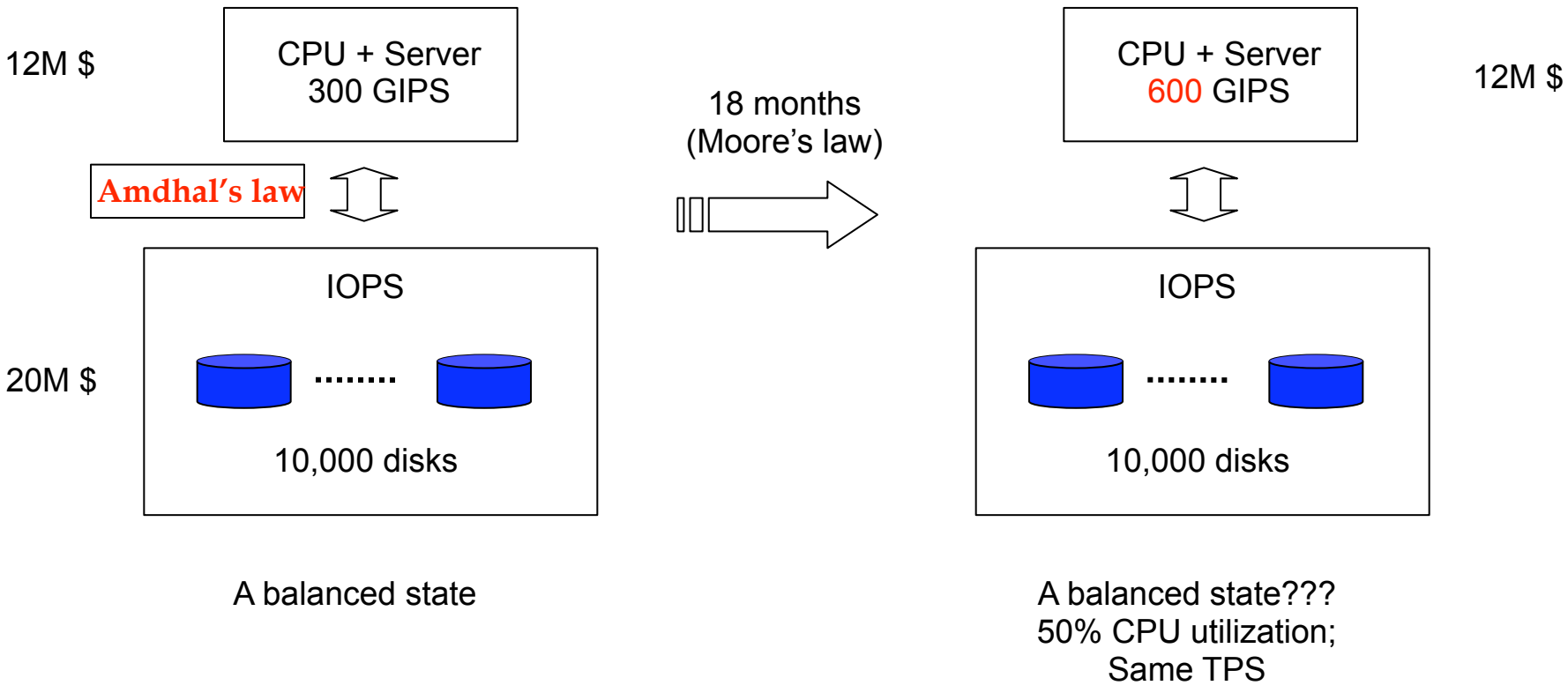
- For balanced systems, OLTP systems pay huge $$$ on disks for high IOPS; IOPS crisis would be worse and worse




A balanced state

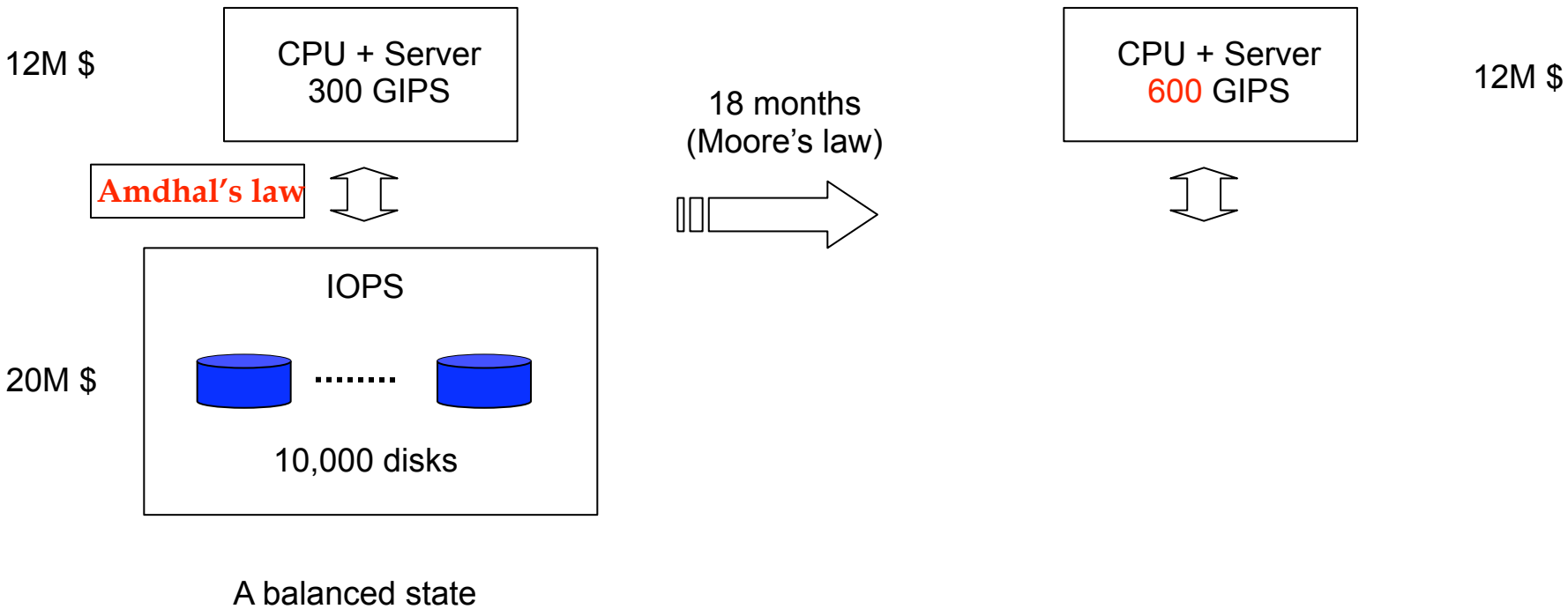
# IOPS Crisis in OLTP(2): with HDDs

- For balanced systems, OLTP systems pay huge $$$ on disks for high IOPS; IOPS crisis would be worse and worse



# IOPS Crisis in OLTP(2): with HDDs

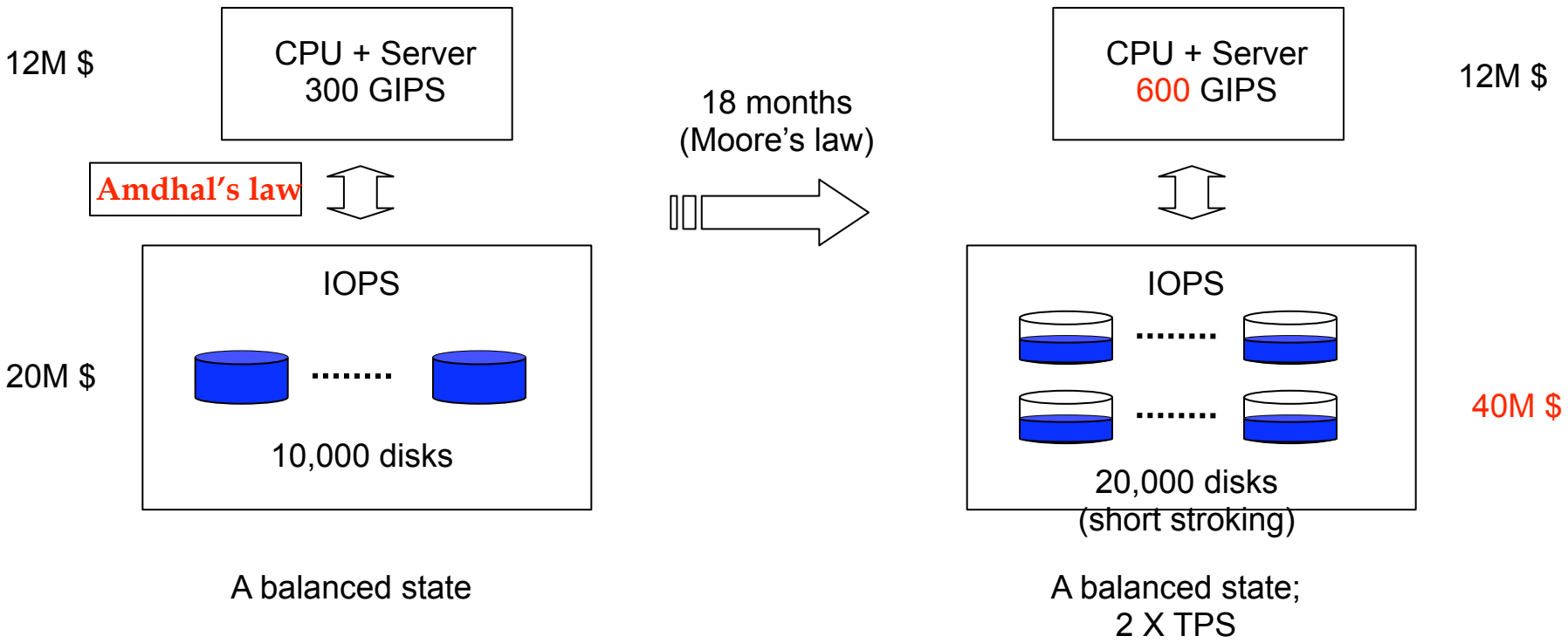
- For balanced systems, OLTP systems pay huge $$$ on disks for high IOPS; IOPS crisis would be worse and worse




A balanced state

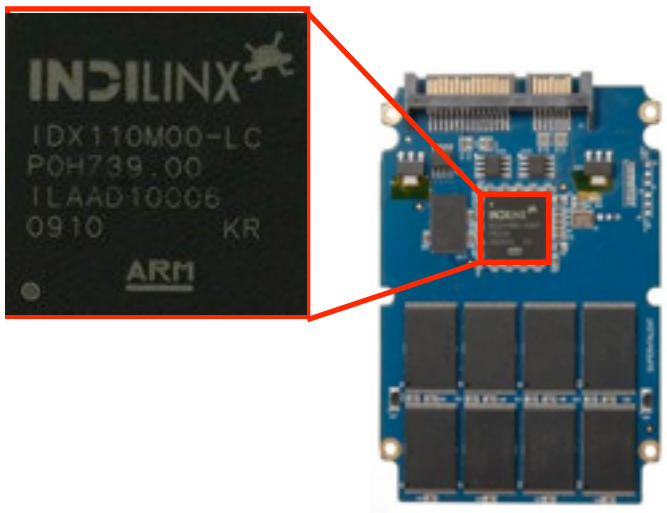
# IOPS Crisis in OLTP(2): with HDDs

- For balanced systems, OLTP systems pay huge $$$ on disks for high IOPS; IOPS crisis would be worse and worse



# Indilinx SSD vs. HDD

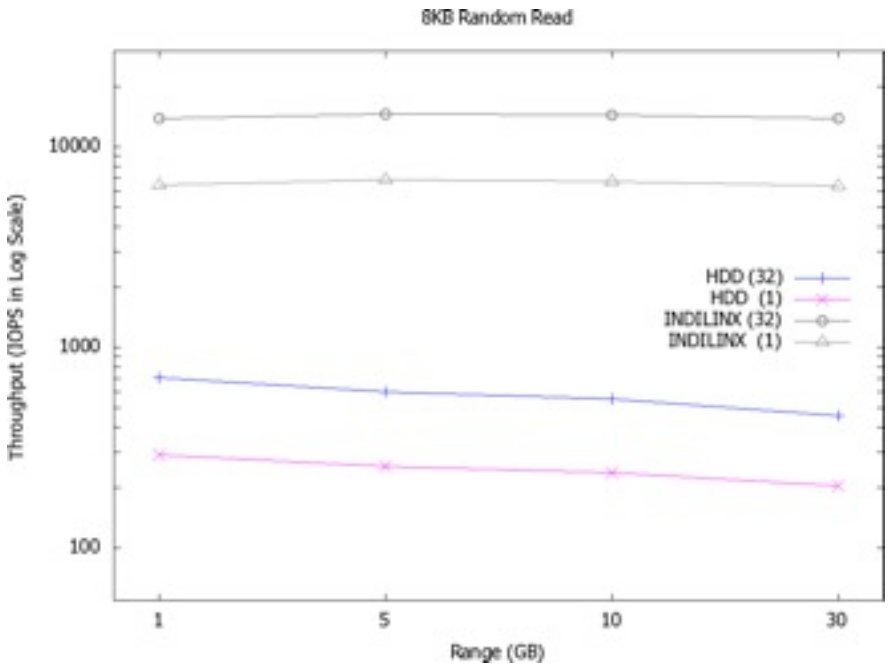
- **INDILINX Barefoot Controller:** 2.5” 32GB SLC SSD with

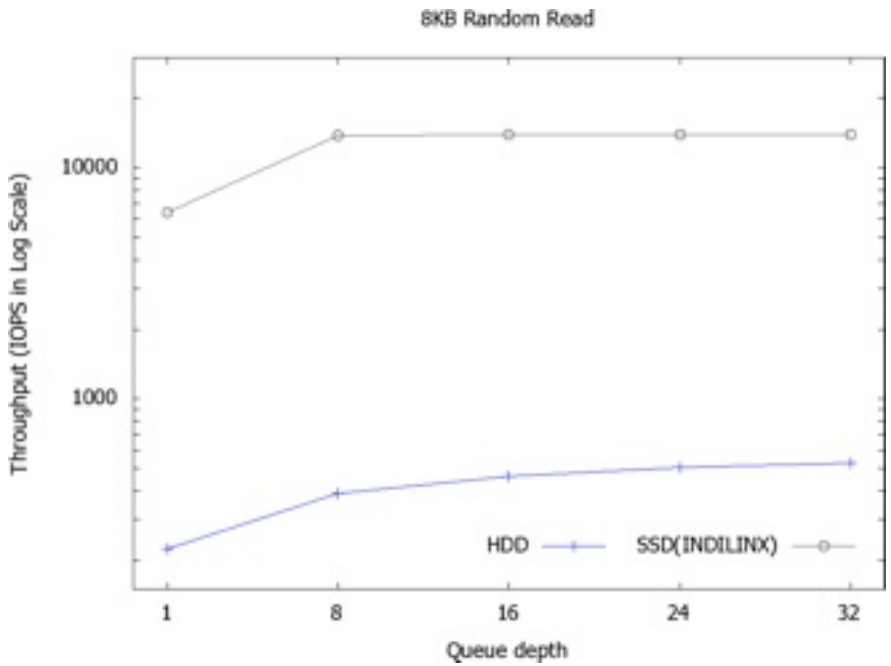



- HDD: 15K rpm Seagate 73.4GB SAS Cheetah 15K.5 model (model no.: ST373455SS)

# Simple IOPS(1)

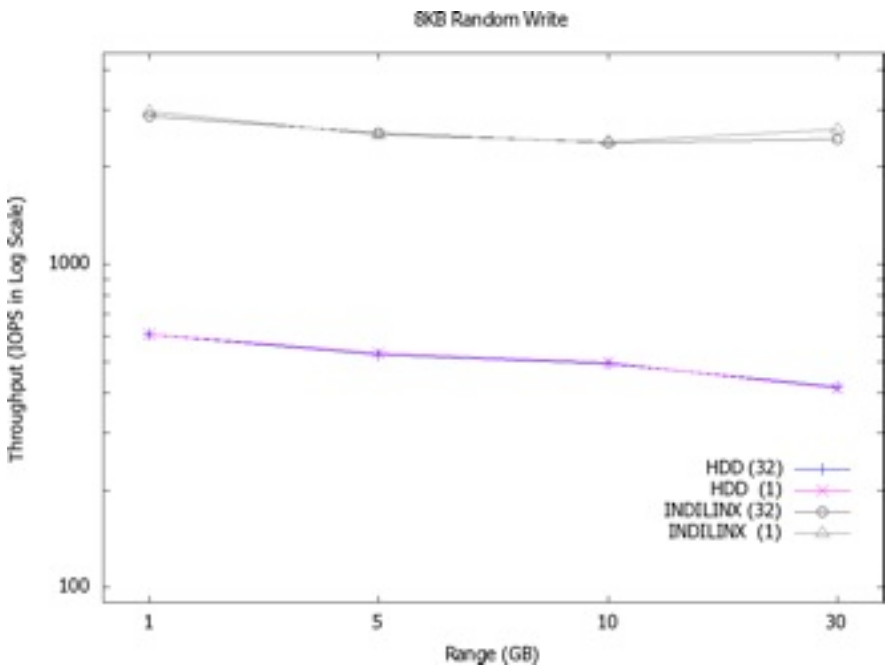
- Random read vs. data range vs. NCQ queue depth
  - 20 ~ 30 times faster (8KB)

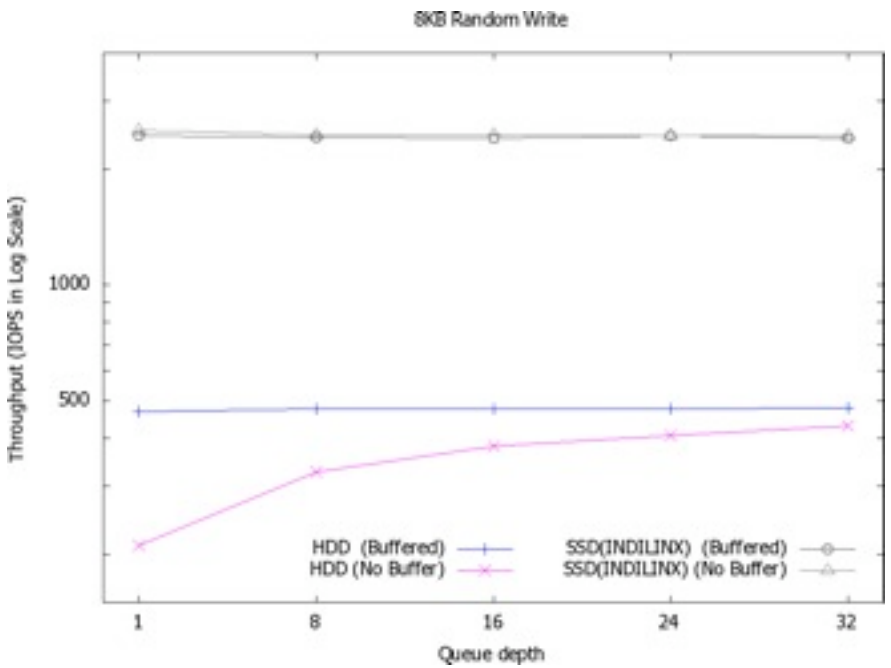


# Simple IOPS(2)

- Random write vs. data range vs. NCQ queue depth
  - 5 ~ 6 times faster (8KB)



# TPC-C Bechmark

- TPC = Transaction Processing Performance Council
  - TPC-C: De facto industry standard benchmark for OLTP performance
  - 5 types of transactions:
    - ✓ Read only: Order-status(4%), Stock-level(4%)
    - ✓ Read/Write mixed: New-order(45%, heavy write), Payment(43%, light write), Delivery(4%, medium write)

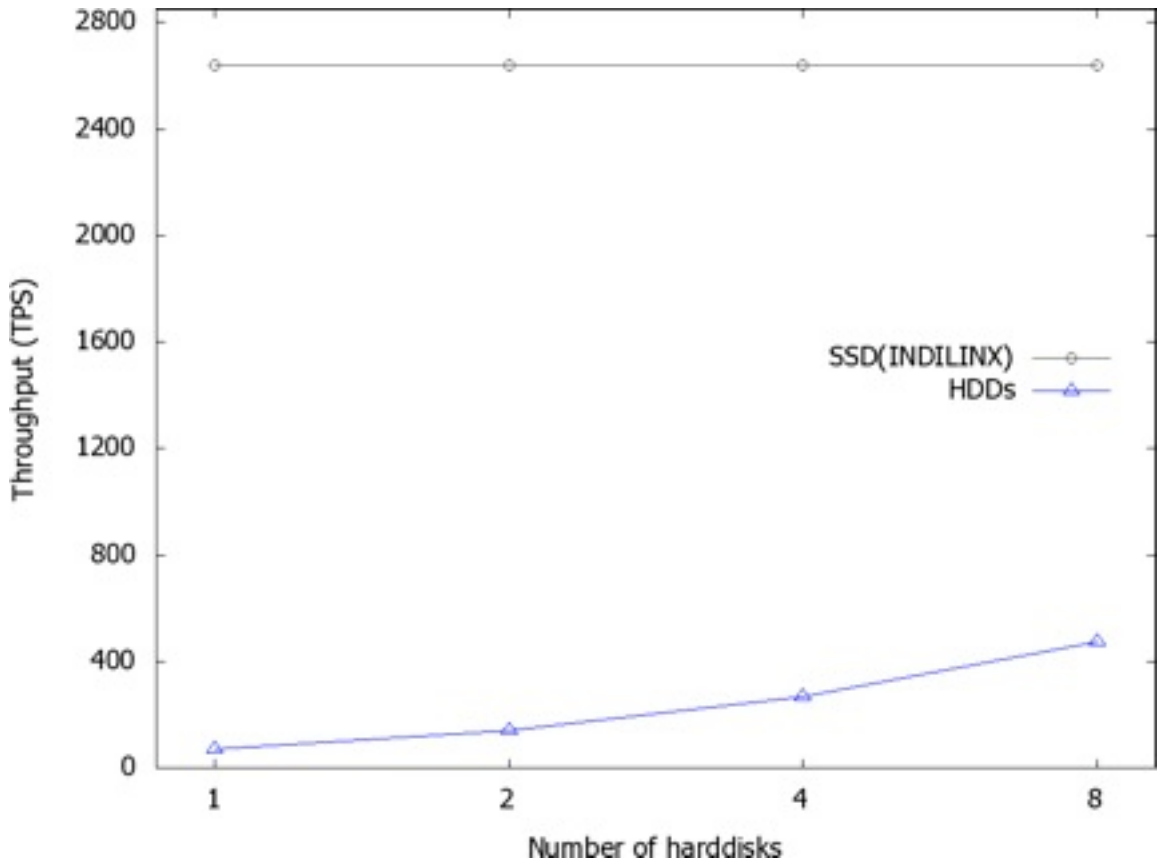
- IO Characteristics
  - Unit of IO: 2 ~ 8K page
  - Ratio of read and write ~~ 1:1

# TPC-C Benchmark: Experimental Setups

- CPU: Intel Core 2 Quad Q6600
- Mother Board: ASUS P5K-E
- RAM: Samsung DDR2 1GB × 2 (2GB)
- OS: Oracle Enterprise Linux 5.1
- DBMS: Oracle 10g R2 (10.2.0.1.0) for Linux x86
- RAID Controller: Intel RAID Controller SRCSASRB
- TPC-C benchmark software: BMFactory
  - 10GB database
  - 100MB buffer
- 8-HDDs vs. 1-SSD

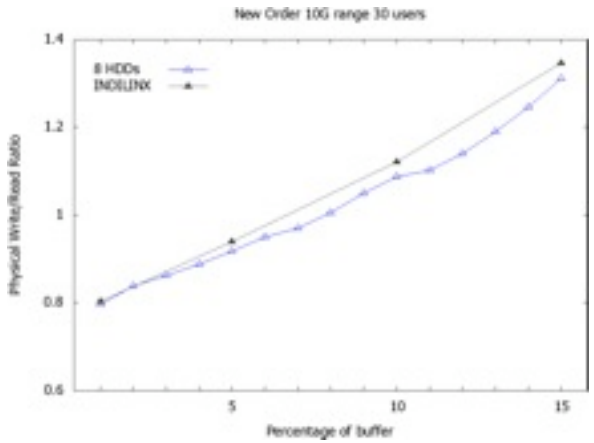
# TPC-C Performance: Read Only TPS

- Order_Status: One SSD vs.8 HDDs = 5:1

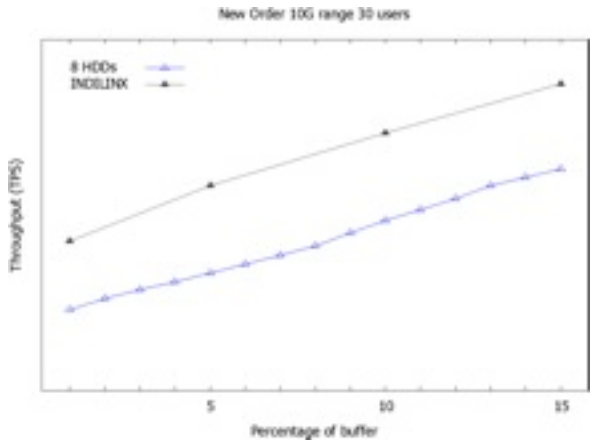


# TPC-C Performance: Read/Write TPS

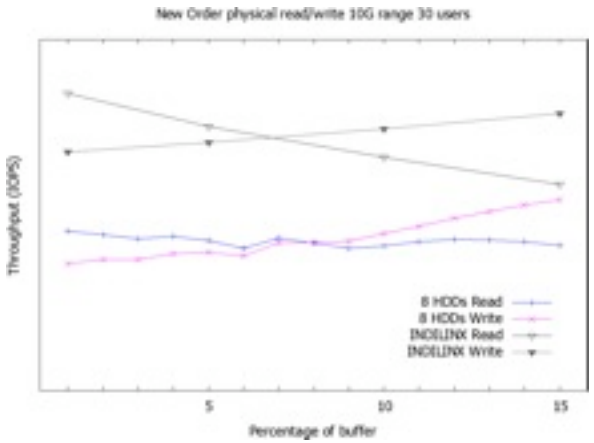
- New_Order: One SSD vs. 8 HDDs = 1.5 ~ 2 : 1
  - Figure A: large buffer means higher physical W/R ratio
  - Figure B: TPS increases
  - Figure C: But, the performance improvement ratio in SSD lags that of HDD because of random write bottleneck




< Figure A >



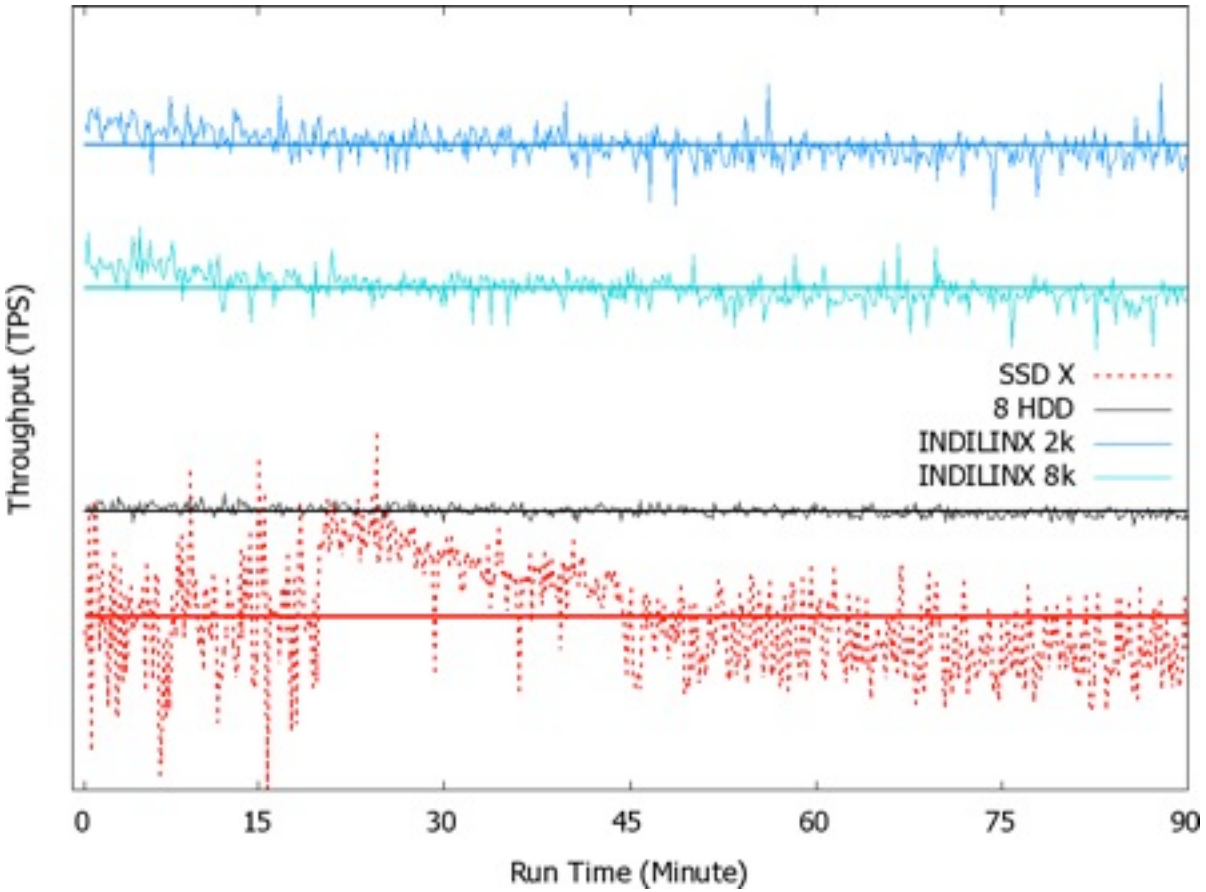

< Figure B >




< Figure C >

# TPC-C Performance: Read/Write TPS

- TPS change over time



# Power Consumption in OLTP

- Meikel Poess et al., Energy Cost: The Key Challenge of Today's Data Centers: A Power Consumption Analysis of TPC-C Results, VLDB 2008
  - In OLTP, storage component consumes 80% of the whole OLTP system
  - Energy metrics will be added in future TPC-C benchmark
- One SSD vs. 8 HDDs
  - Performance: SSD >> 8 HDDs
  - Power: SSD(5W) << HDD(104W)
- In other words, SSD shows

  very low (Watt or $) / IO / sec

# Conclusion

- OLTP: The Case for Flash Memory SSD
- One FlashSSD can beat 10 15K RPM harddisks
  - Performance, prices, and, power consumption
- The key metric in OLTP storage
  - IOPS / GB, rather than sequential bandwidth and capacity
- We should be more enthusiastic in developing the various cases, where SSD is definitely winner both in terms of price and performances