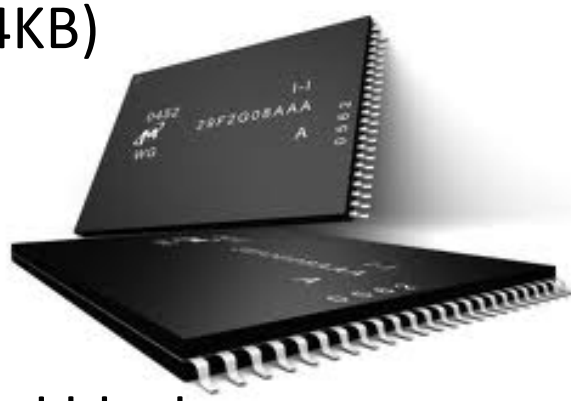
Performance Modeling and Analysis of Flash-based Storage Devices

H. Howie Huang, Shan Li
George Washington University
Alex Szalay, Andreas Terzis
Johns Hopkins University

MSST ‘11
May 26, 2011

NAND Flash

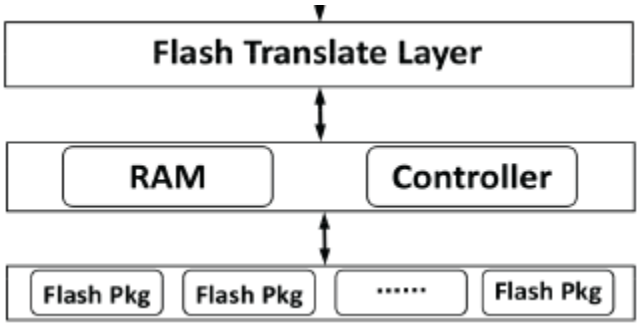
- Each NAND flash package contains a small number of dies where digital logic gates (memory cells) are grouped into blocks (e.g., 256KB) and pages (e.g., 2KB and 4KB)
- Support three kinds of operations
  - Page-level read (fast)
  - Page-level write (slow)
  - Block erase (much slower)



- Page writes can only be performed to an erased block
  - A page becomes available for writes after the entire block is erased
- Wear-leveling is used for improving the lifetime
  - Limited erase cycles per cell

Solid-State Drives (SSDs)

- Resembles the form factor (2.5 or 3.5 in)
- Emulate block-level interfaces (SCSI and SATA)
- Internal organization
  - Flash packages, RAM (cache buffer), host interface logic
  - FTL (flash translation layer) mimics a hard disk and manages the mappings from logical block addresses (LBA) to physical flash locations



Motivation

- Flash-based SSDs appear in a wide spectrum of systems, e.g.,
  - Mobile computers where SSDs provide low power consumption and resist rough handling
  - Enterprise class server and storage where SSDs promise high data transfer rate and low access latency



- For SSDs, time-sensitive and I/O-intensive applications are often considered as good candidates

- Good performance model can help
  - Understand the state-of-the-art of SSDs
  - Provide the tools for exploring design space of flash-based storage systems

Overview

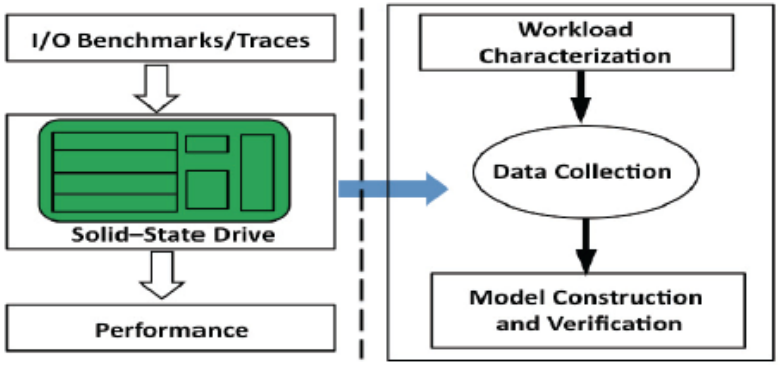
- Utilize the black-box modeling approach to analyze and evaluate SSD performance
- This approach is attractive because it requires limited information about a storage device
- Beneficial for SSDs, as the vendors are reluctant to reveal the design details

Contributions

- Analyze a number of different workload characteristics for SSDs modeling
  - Need further improvements on traditional performance models that were designed for hard drives
- Propose an extended model to properly correlate the SSD performance and I/O requests
  - Investigate the models for each specific data access pattern
- Evaluate this approach on a variety of SSDs
  - The model can produce accurate predictions under a collection of different workloads

Approach

- Build a black-box model to predict the performance of a given SSD through different workload characteristics
- The performance tends to be correlated with workload characteristics
  - E.g., SSD latency and throughput fluctuate when the percentage of write requests varies
- Construct the model
  - Benchmark an SSD and collect the training data
  - Utilize statistical methods to quantify the correlations



Performance Model

- The model takes the workload characteristics (wc) as input parameters and outputs the predicted performance metric (p)

$$p = F(wc).$$

- Focus on three performance metrics: latency (lat), bandwidth (bw), and throughput in IOs per second (iops)

$$p = lat|bw|iops.$$

Basic Model

- Characterize a stream of IOs in four fundamental parameters
  - Read and write ratio (rw_ratio): the percentage of writes in the request
  - Request size (req_size): the number of bytes transferred to/from the storage device
  - Queue depth (q_dep): number of outstanding I/Os
  - Request randomness (rand): the percentage of random accesses in the I/O request stream

$$wc =< wr\_ratio, q\_dep, req\_size, rand > .$$

Extended Model

- The four parameters in the basic model somewhat capture the correlation between the workloads and SSD performance

- Consider additional parameters to improve model accuracy
  - Read and write stride for the effect of request alignments
  - Read and write size because of SSD asymmetric read/write performance
  - Read and write randomness that can also have varied impacts on the SSD performance

$$wc =< wr\_ratio, q\_dep, wr\_size, rd\_size, wr\_rand, rd\_rand, wr\_stride, rd\_stride > .$$

Regression Tree

- Apply statistical machine learning algorithms
- Use the least-squares approach to fit the linear model
- Construct a regression tree from the function
  - Recursively split the input variables into leaf nodes to minimize mean square errors
  - Leaf nodes provide the predicted values for dependent variables as a constant function of independent variables

Experiment Setup

- Run experiments on the machines with Intel Core 2 Duo 2.93 GHz, 4GB memory, and Linux kernel 2.6
- Test on three SSDs, one hard drive, as well as an SSD RAID

	HDD_S [9]	SSD_I [10]	SSD_A [11]	SSD_S [12]
Capacity	500GB	80GB	120GB	32GB
Buffer Size	8MB	Unknown	64MB	Unknown
Read Bandwidth	-	250MB/s (seq)	250MB/s	100MB/s (seq)
Write Bandwidth	-	70MB/s (seq)	100MB/s (sustained)	80MB/s (seq)
Latency	5.6ms (avg)	85μs (Read) 115μs (Write)	$< 100\mu$s	-

- The training data is generated by a synthetic I/O workload Generator
  - Each I/O request is run for one minute
  - For each device, we run 12,000 one-minute workloads
    - In total take about 200 hours (about 8 days) to complete
- Evaluate with synthetic I/O requests, and four real-world traces from OLTP applications and a web search engine

Evaluation Metrics

- Mean Absolute Error (MAE) is defined as the difference between the observed and predicated performance
- Mean Relative Error (MRE) is defined the ratio between the absolute error and the observed performance
- $R^2 = 1 – SSE/SST$ is used to determine how well the performance is likely to be predicted by the model

- A better model has
  - Smaller MAE and MRE
  - R^2 close to 1

Microbenchmarks

TABLE II
PREDICTION ACCURACY OF BASIC MODELS

(a) Latency

Device	R^2	MAE(Mean)	MRE
HDD_S	0.808	28.94(94.61)	90%
SSD_I	0.627	6.90 (15.97)	63%
SSD_A	0.926	5.61 (36.31)	23%
SSD_S	0.693	14.21 (34.90)	55%

(b) Bandwidth

Device	R^2	MAE(Mean)	MRE
HDD_S	0.281	7.29(14.63)	110%
SSD_I	0.515	21.87(68.61)	40%
SSD_A	0.570	15.72(38.17)	86%
SSD_S	0.548	13.66(36.33)	63%

(c) Throughput

Device	R^2	MAE(Mean)	MRE
HDD_S	0.080	467(664)	50%
SSD_I	0.500	1,547(3,967)	53%
SSD_A	0.765	246(1,054)	19%
SSD_S	0.459	749(1,702)	48%

TABLE III
PREDICTION ACCURACY OF EXTENDED MODELS

(a) Latency

Device	R^2	MAE(Mean)	MRE
HDD_S	0.866	17.96(94.61)	26%
SSD_I	0.986	1.42 (15.97)	12%
SSD_A	0.976	3.16(36.31)	9%
SSD_S	0.911	6.22(34.90)	20%

(b) Bandwidth

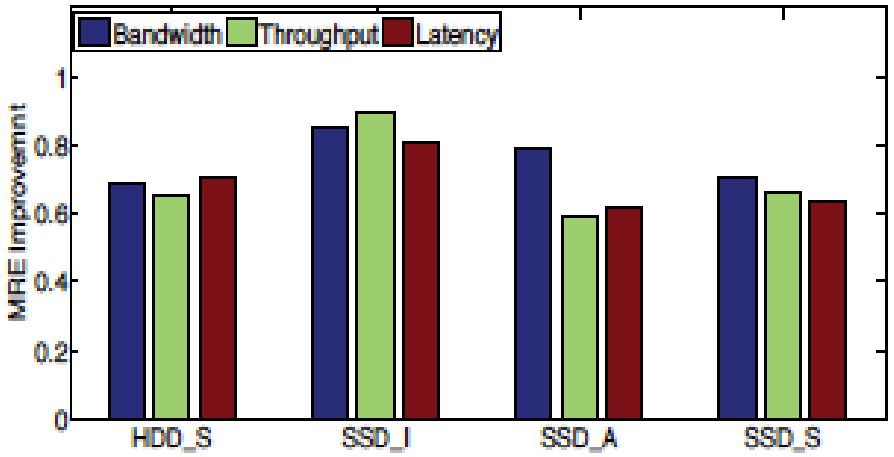
Device	R^2	MAE(Mean)	MRE
HDD_S	0.768	3.67(14.63)	35%
SSD_I	0.981	3.91(68.61)	6%
SSD_A	0.882	6.29(38.17)	18%
SSD_S	0.917	5.21(36.33)	19%

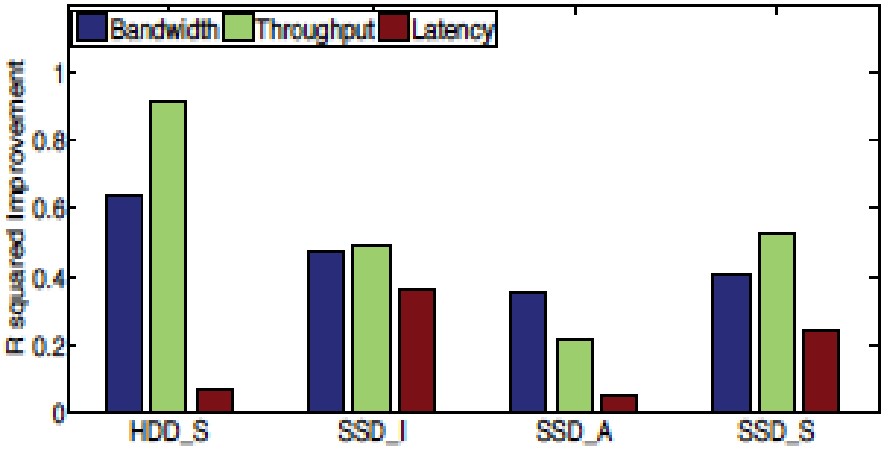
(c) Throughput

Device	R^2	MAE(Mean)	MRE
HDD_S	0.870	152(664)	18%
SSD_I	0.970	254(3,967)	6%
SSD_A	0.971	74(1,054)	8%
SSD_S	0.951	212(1,702)	15%

- Extended models significantly improve model accuracy for SSDs
- Latency remains most difficult to predict
  - Consistent with the results from prior research

Model Improvements





- For MRE, all the devices have close to or higher than 60% improvements for three performance models
- For SSD_I there is 80% improvement
- Large increases in R^2 values for SSD_I and SSD_S

Work-specific Models

- Explore the SSD performance models for eight special cases
  - Each workload reflects only one type of access pattern each time
- SSD models work well in this case

TABLE IV
PREDICTION ACCURACY OF WORKLOAD-SPECIFIC MODELS

(a) Latency

	R^2			MRE		
Workloads	HDD_S	SSD_I	SSD_A	HDD_S	SSD_I	SSD_A
rd_only	0.953	0.999	0.999	18%	4%	1%
wr_only	0.942	0.996	0.990	17%	4%	7%
rand_only	0.910	0.975	0.991	29%	15%	5%
seq_only	0.933	0.954	0.989	16%	12%	6%
rand_rd	0.965	0.999	0.968	46%	2%	8%
rand_wr	0.984	0.993	0.993	9%	9%	5%
seq_read	0.964	0.999	0.999	8%	2%	6%
seq_wr	0.961	0.999	0.998	7%	7%	3%

SSD Array

- Extended models perform better again
- The model accuracy decreases when compared to the single device
- Further investigations will look into this issue

	R^2	MAE(Mean)	MRE
Latency	0.939	1.43(8.15)	23%
Bandwidth	0.885	18.45(101.70)	25%
Throughput	0.860	934(4,875)	17%

IO Traces

- Able to achieve high accuracy for four different traces
- Two search engine traces produce much better accuracy for all devices
  - Web search engine traces are read intensive with a very high read to write ratio
  - Financial traces are write heavy

(a) Latency (s)

Device	*Financial1*	*Financial2*	*WebSearch1*	*WebSearch2*
HDD_S	16%	12%	1%	3%
SSD_I	7%	19%	1%	1%
SSD_S	18%	15%	2%	1%
Array_I	19%	11%	1%	1%

(b) Bandwidth (MB/s)

Device	*Financial1*	*Financial2*	*WebSearch1*	*WebSearch2*
HDD_S	18%	30%	5%	5%
SSD_I	11%	38%	1%	1%
SSD_S	17%	14%	2%	1%
Array_I	26%	25%	2%	2%

(c) Throughput (IO/s)

Device	*Financial1*	*Financial2*	*WebSearch1*	*WebSearch2*
HDD_S	15%	26%	5%	5%
SSD_I	9%	37%	1%	1%
SSD_S	15%	14%	2%	1%
Array_I	25%	12%	1%	1%

Related Work

- Extensive research in performance modeling studies on hard drives
  - Analytical modeling [33]–[35]
  - Simulation [25],[36]
  - Benchmarking [37], [38]
  - Black-box approach [4], [5], [39], [40]
- Inspire our work on SSD models

Conclusion

- Flash-based solid-state drives play an important role in today's storage systems
- An accurate performance model will help
- A good black-box model can be constructed for SSDs

- Future research directions:
  - Evaluate our models against existing simulators, e.g., SSDSim and FlashSim
  - Apply our black-box models, preferably in an autonomic manner, to help design and configure heterogeneous storage systems

Thank You

Howie@gwu.edu

http://www.seas.gwu.edu/~howie