How Suspending Benefits Enterprise Apps Performance

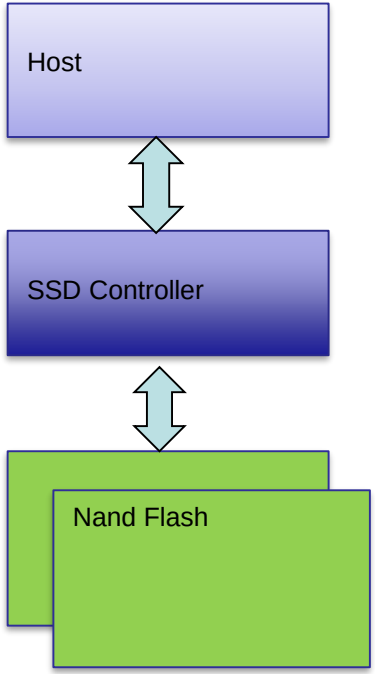
debin.wu@starblaze-tech.com

Principal Engineer, SSD design

Beijing Starblaze Technology

Nand Flash SSD I/O operation models

- Host View
  - Host logical write
  - Host logical read
- Nand Flash View
  - Nand physical Erase
  - Nand physical Program
  - Nand physical Read



SSD Performance

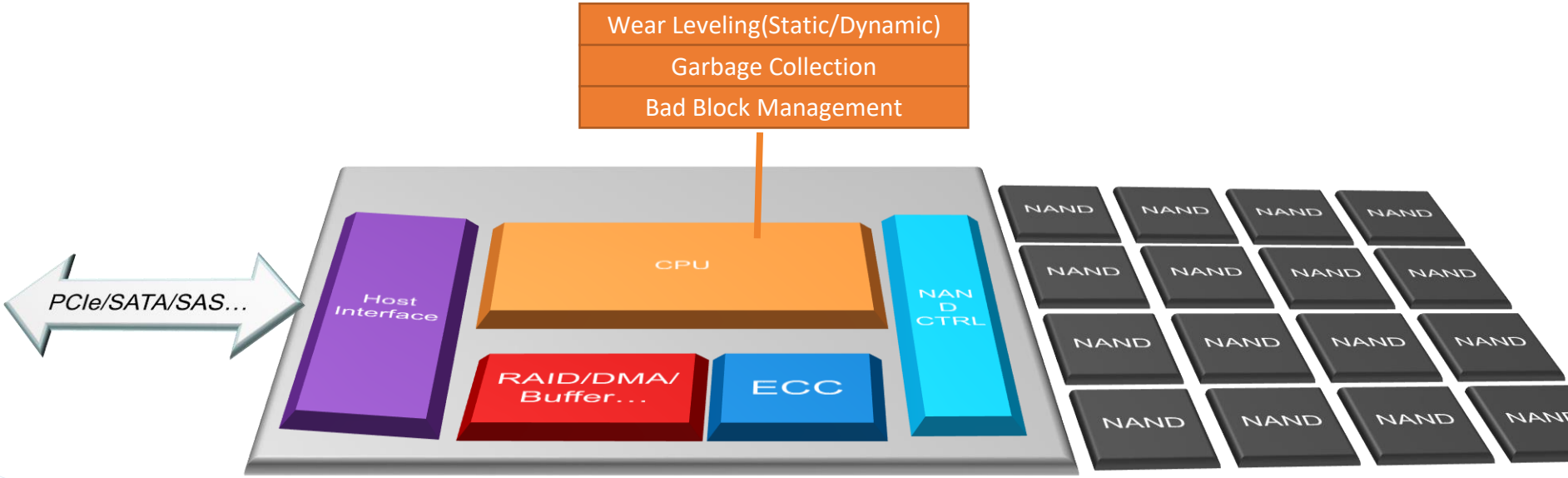
- Bandwidth
- IOPS
- Latency
- Consistency

e.g.

Performance			
Sequential Read (up to)	2800 MB/s	Sequential Write (up to)	1900 MB/s
Random Read (100% Span)	450000 IOPS	Random Write (100% Span)	150000 IOPS
Latency - Read	20 µs	Latency - Write	20 µs

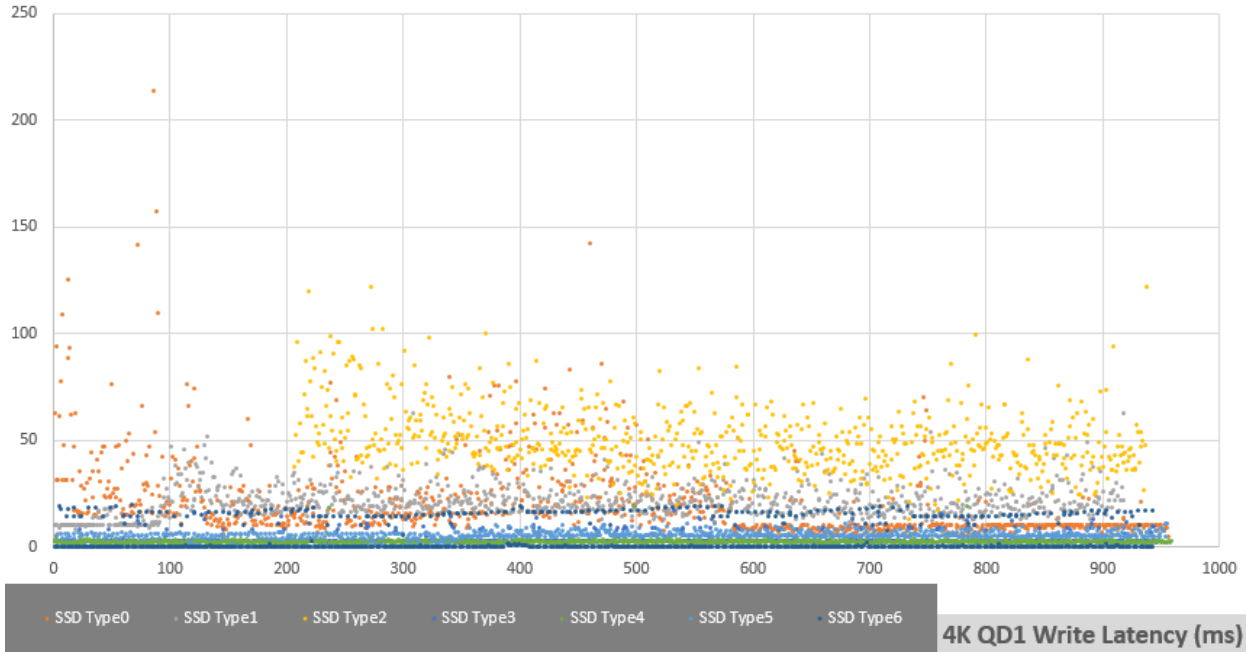
Flash Memory Summit

SSD Controller Architecture



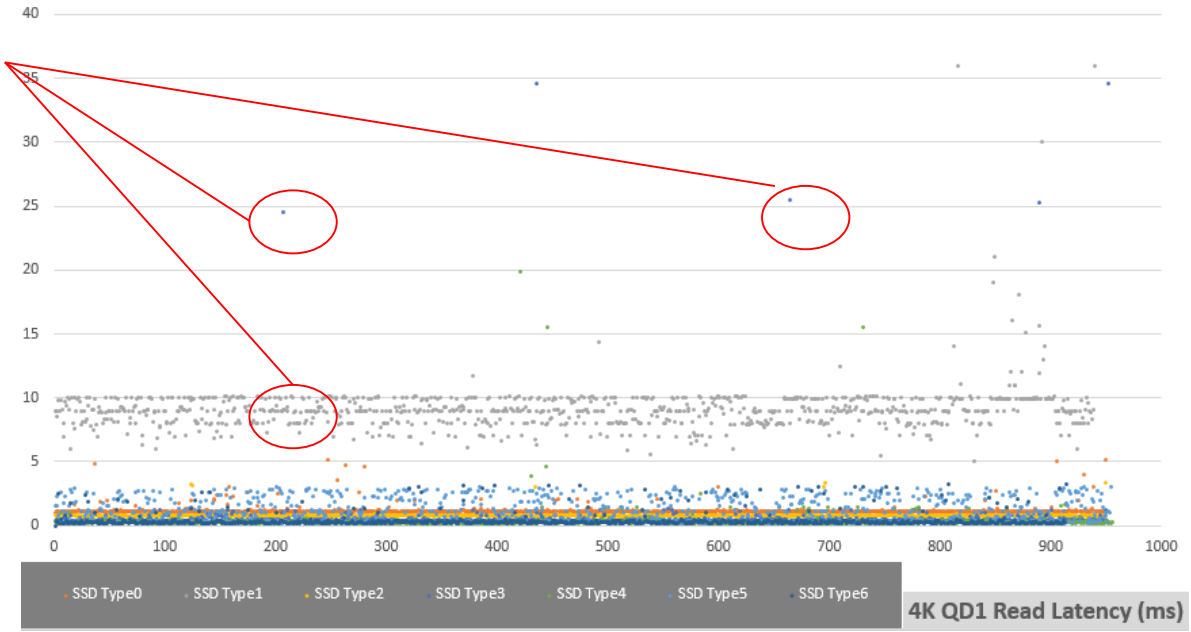
Write Latency Influence from Erase/Program

- GC erase/program/read
- Write latency is influenced by GC indirectly – write buffer (SRAM/DRAM)



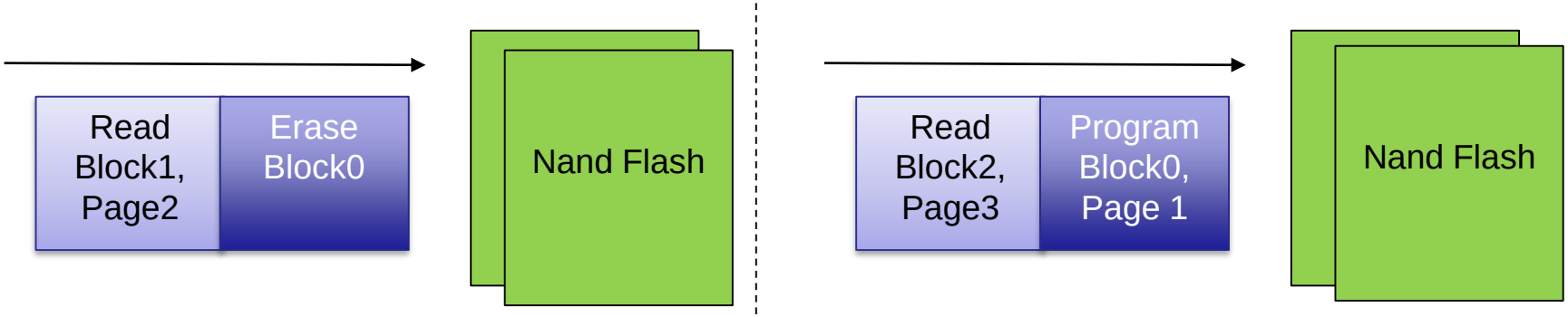
Read Latency Influence from Erase/Program

- Read latency is influenced by GC directly – nand flash



Read and Program/Erase Conflict

- Read after Erase/Program
- Latency ~= tErase + tRead or tProgram + tR
- tR << tErase, tR << tProgram



Flash Memory Summit

Nand Flash Timing Parameters

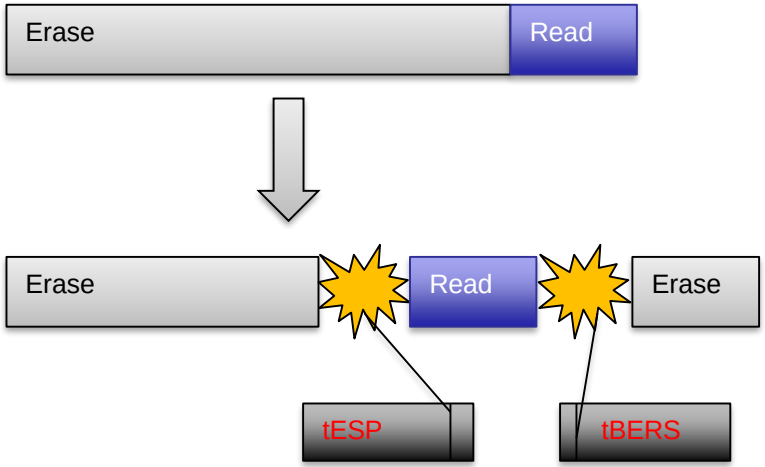
Nand Flash	2D		3D	
Parameter	SLC	MLC	SLC*	TLC
tErase (ms, typ/max)	5/10	5/10	10/25	10/25
tProg (ms, typ/max)	0.3/3	1.4/3	0.25/0.6	2.3/11
tRead (us, typ/max)	35/50	50/85	37/65	67/100

Enterprise Database Application Test

Test Cases	Description	RW	Bandwidth	IOPS	Avg Latency(usec)	Max Latency(usec)
1.1	2rr1rw_1thread_db_fio	read	23215KB/s	5803	126.01	9970
1.2	2rr1rw_1thread_db_fio	write	12302KB/s	3076	17.05	409
2.1	5rr4rw_1thread_db_fio	Read	15130KB/s	3782	147.39	9814
2.2	5rr4rw_1thread_db_fio	Write	12950KB/s	3238	16.8	133
3.1	2rr1rw_2thread_db_fio	Read	28708KB/s	7177	120.03	10582
3.2	2rr1rw_2thread_db_fio	Write	14770KB/s	3693	18.05	270
4.1	5rr4rw_2thread_db_fio	Read	26057KB/s	6514	141.9	12598
4.2	5rr4rw_2thread_db_fio	Write	21100KB/s	5276	17.58	405

Read Conflict with Program/Erase

- Nand Operations are in serial for single LUN



Read After Erase

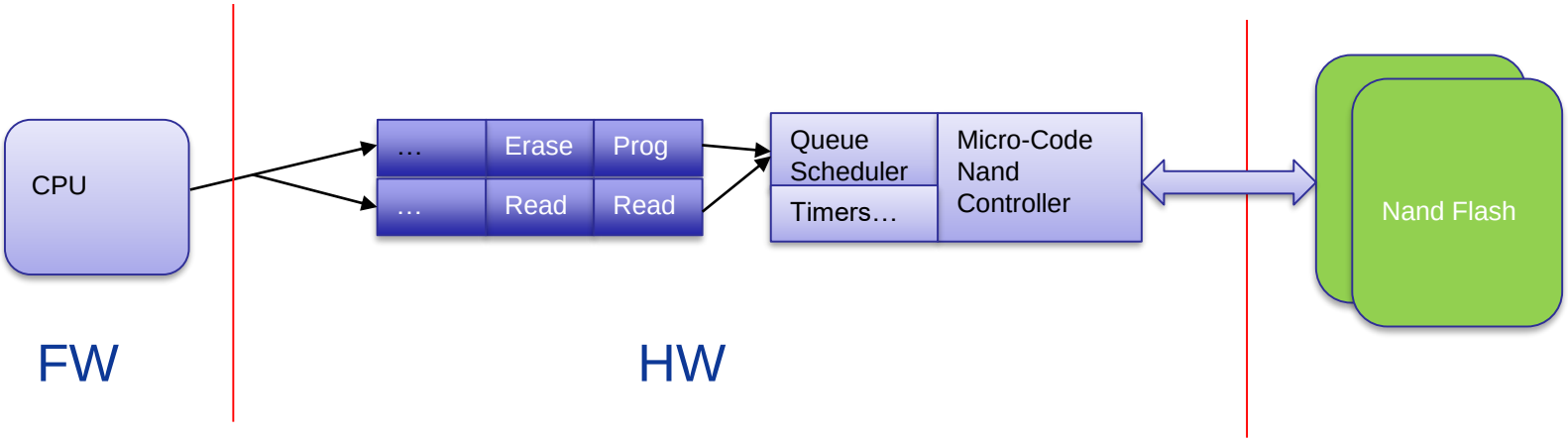
Read With Erase/Program Suspension

Challenges in implementation

- **WHICH** lun/block/page address is suitable
- **WHEN** will suspension benefits most and when resume pended operation
- **HOW** FW/HW scheduler is able to handle
- **WHAT** will be handled if cache/multi-plane operations are involved
- Frequently suspension will cause cell worn out soon. Suspension is taken into P/E cycle
- GC is interrupted and may cause some negative effects

Starblaze's solution

- Micro-Code Nand Controller with HW queue scheduler
- SMART processing mechanism between GC read and host read

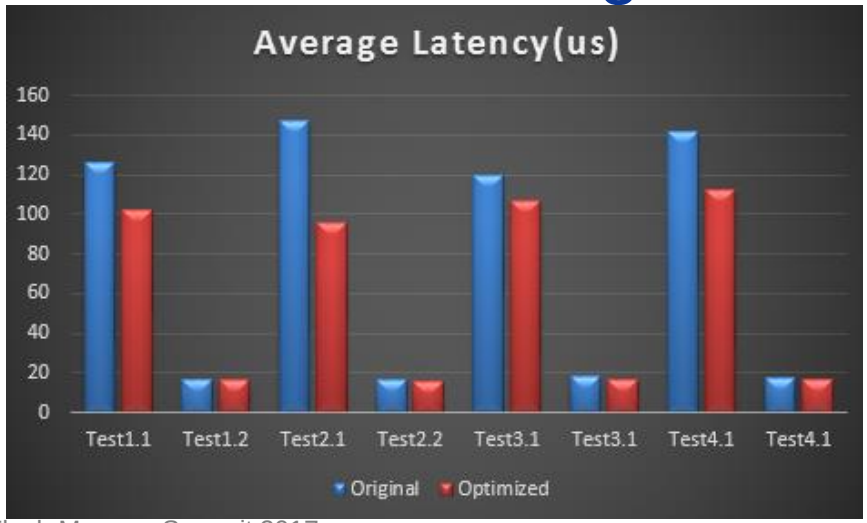


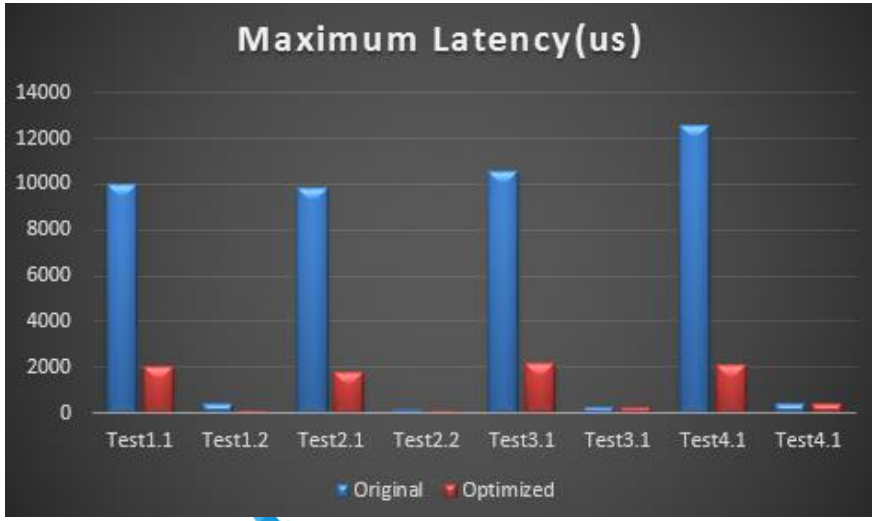
Enterprise Database Application Test -- Optimized

Test Cases	Description	RW	Bandwidth	IOPS	Avg Latency(usec)	Max Latency(usec)
1.1	2rr1rw_1thread_db_fio	read	23215KB/s	5803	102.53	2047
1.2	2rr1rw_1thread_db_fio	write	12302KB/s	3076	16.28	117
2.1	5rr4rw_1thread_db_fio	Read	15130KB/s	3782	95.34	1809
2.2	5rr4rw_1thread_db_fio	Write	12950KB/s	3238	16.14	106
3.1	2rr1rw_2thread_db_fio	Read	28708KB/s	7177	107.1	2145
3.2	2rr1rw_2thread_db_fio	Write	14770KB/s	3693	16.82	221
4.1	5rr4rw_2thread_db_fio	Read	26057KB/s	6514	112.61	2098
4.2	5rr4rw_2thread_db_fio	Write	21100KB/s	5276	16.88	393

Enterprise Database Application Test -- Optimized

- Both average and maximum latency are better than original





Conclusion

- Benefits
  - Avg/Maximum read latency value
  - Consistency/QoS benefits
  - Power Loss Protection with erase suspension
  - Real database application test shows the results
  - Starblaze’s controller demonstrates a feasible solution


Thank You!!!

Q & A

- Starblaze Booth: #647
-

Invitation	Thursday, August 10th , 1:30-4:15pm Forum C-32: SSD Concepts (SSDs Track)
Topic	Creating SSD Designs That Can Readily Be Migrated Between Form Factors
Presenter	Feng Tang, Senior Manager, SSD FW