

Marktüberblick

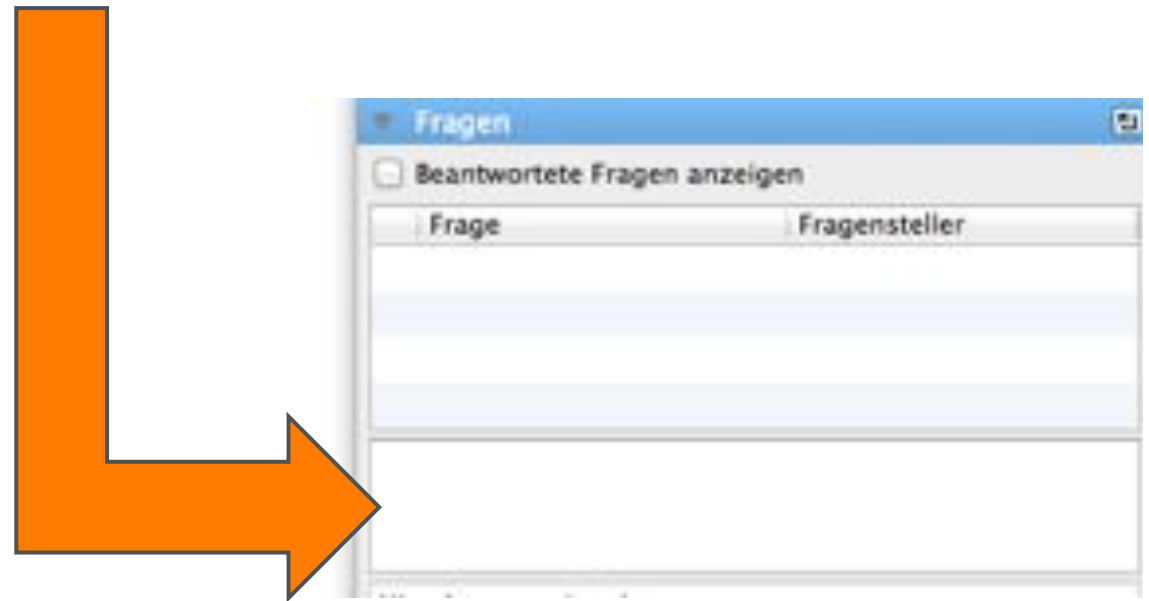
SSD und Flash Technologie

Florian Hettenbach
26.03.2014



Organisatorisches

- _ Dauer ca. 1 Stunde
- _ Webcast wird aufgezeichnet
- _ Fragen während des Webcasts?



Florian Hettenbach



- Seit 7 Jahren bei Thomas Krenn
- Produktmanagement
Storage und Virtualisierung
- Business Development:
Neue Geschäftsbereiche
und Technologien

Agenda

- _ Was man vorab wissen sollte
- _ Marktüberblick
- _ Unsere Erfahrungen
- _ Entscheidungshilfe
- _ Ausblick
- _ Fazit



Aufgepasst!



Flash ist schnell!



80 IOPS

375x



30.000 IOPS

375x in Relation



4 km/h

1500 km/h

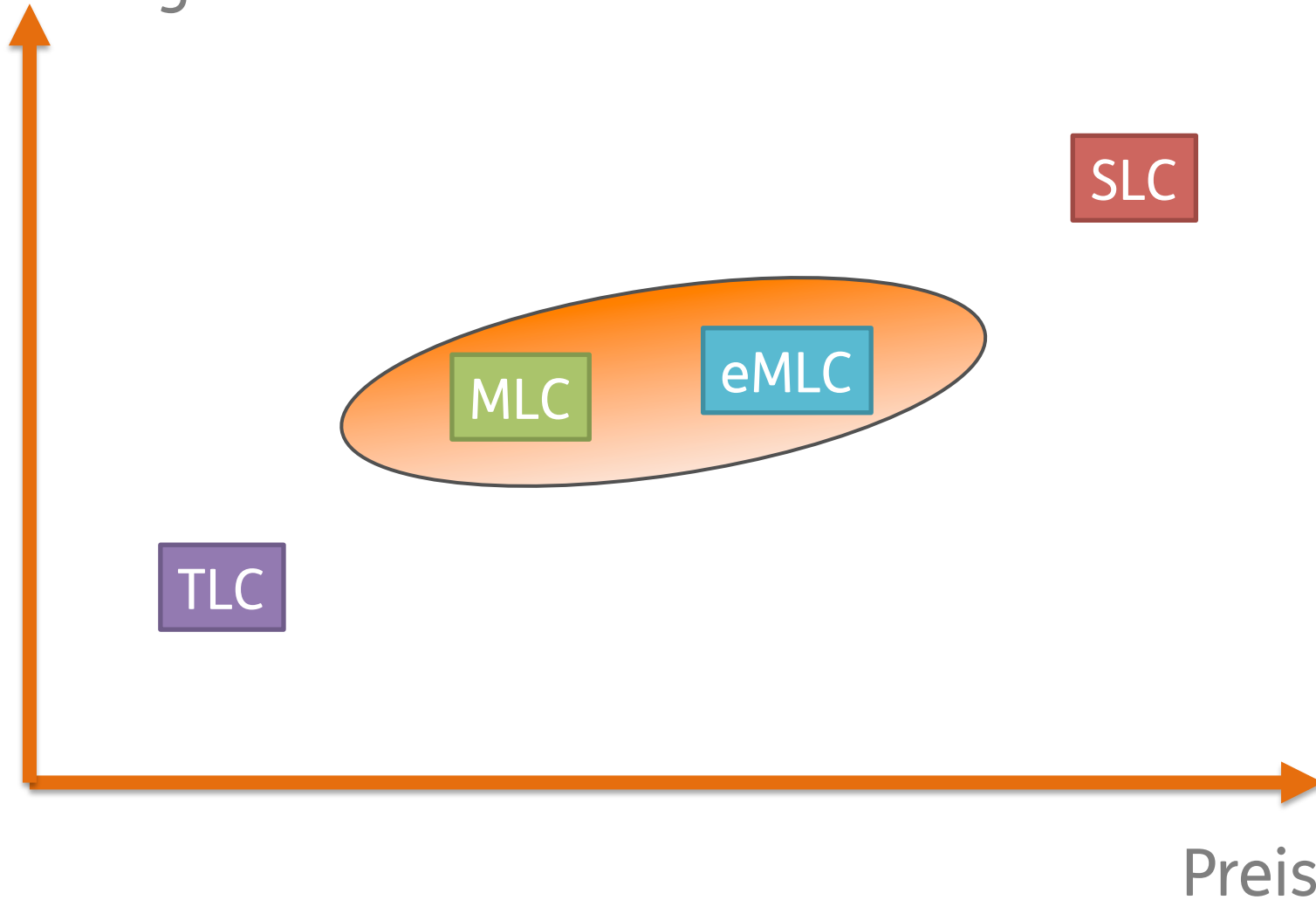
4 wichtige Punkte



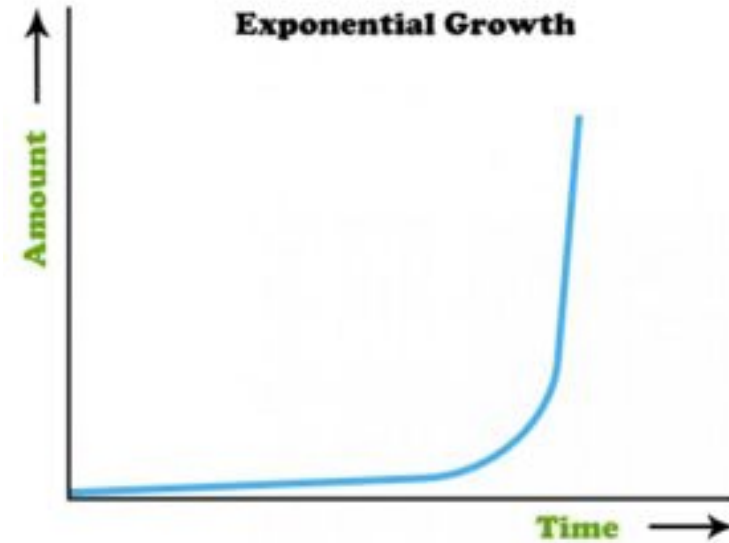
1. Flash nutzt sich ab


2. Flash Varianten

Haltbarkeit
Geschwindigkeit




3. Flash fällt anders aus





**4. Lesen = umsonst,
Schreiben = teuer**





Marktüberblick



Hersteller

San



TOSHIBA



mushkin



23 Hersteller x Modelle x Kapazität = ???

Marktkonsolidierung

- _ Toshiba { OCZ
- _ Seagate + Samsung
- _ HGST { sTec + Virident
- _ SanDisk { Smart Storage
- _ Intel → Micron

Nand- und Controller



■ Nand

■ Controller

Grundlagen - Geschwindigkeit

— Transferrate / Datendurchsatz

— MByte / s

— Vergleich: Personen/h pro Strecke



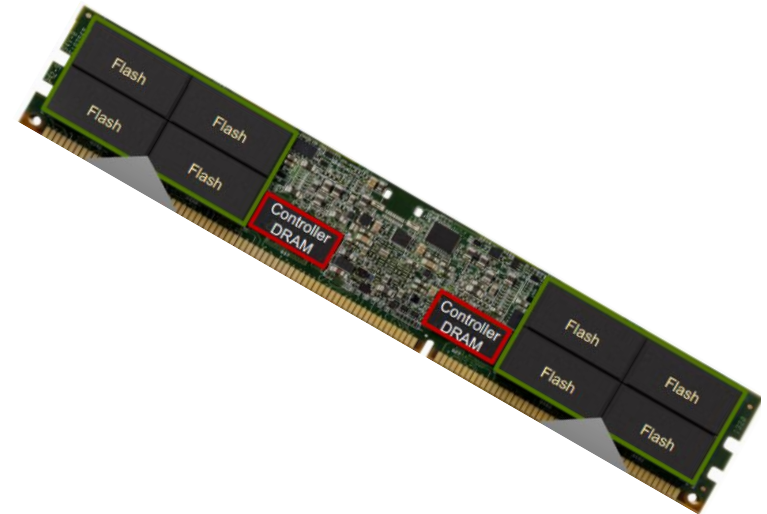
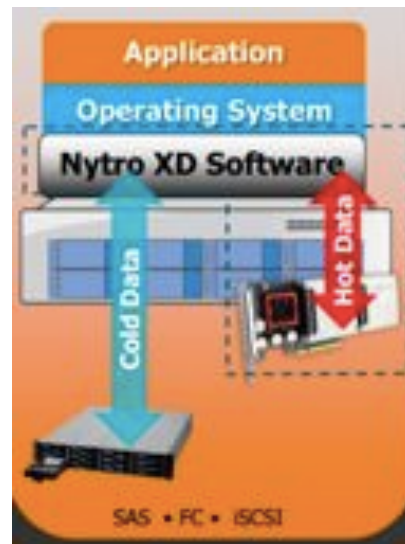
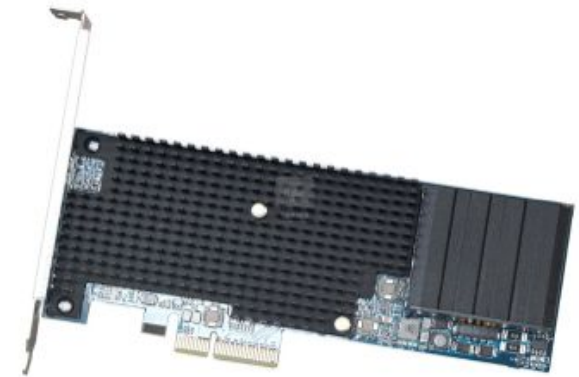
— I/O Operationen pro Sekunde

— IOPS

— Vergleich: Anzahl möglicher indiv. Fahrten



Einsatzmöglichkeiten



Beispiel: LSI NytroMegaRaid



Consumer? Enterprise?

— TeraByte Writen (TBW)
— 90/10 Read / Write

— Drive Writes per Day
(DWPD)
— 50/50 RW - 10/90

*Positionierung durch Hersteller:
Bspw. „Read intensive“*

Konnektivität

- SATA 3 (6 Gbit)
- M.2 SATA (SATA über Mini-PCIe Konnektor)
- SAS 3 (12 Gbit)
- PCIe (x8 - bis zu 62 Gbit)
- NVMe (PCIe)





Unsere Erfahrungen





1 Setup Information
Tested device

- Model Number: 800G2M40003
- Serial Number: 8120110000000000
- Firmware Revision: 00100006
- Media Serial Num:
- Media Manufacturer:
- Device size with M = 1000000 200000 MB (200 GB)

Used benchmark tool

- Configuration: fio test mode: fio_01 (scale: x1.2) out: 16

Performance System:

- FioTest version: 1.1
- Fio version: fio_2.10
- Date of test run: 2016-12-08
- Number of jobs: 2
- Number of benchmark file instances: 16

Operating System:

- Kernel Version: 3.2.0-30-generic
- Description: Ubuntu 12.04.1 LTS

2 General Information

- Condition: The percentage of read operations in this random mixed workload, in the ratio the term "100000 Reads, 100% Read and 0% Write, 100% RW, Read and 0% Write, and so on"
- Test mode: The block size of I/O is the default 4096 bytes.
- Measurement window: Those hours, when the dependence variable became stable.
- Measurement window: A specific type of test variable to determine the steady state.

2.1 Steady State

The Steady State is to determine if a test has reached a steady performance level. Each test has a defined dependence variable to check if the test has already been reached. To check for the steady state the performance values of a test measurement window are taken (the last 5 hours). The steady state is reached if:

- The maximum value deviation is less than 20% of the average in the measurement window.
- The speed of the final test I/O is less than 10% of the average in the measurement window.

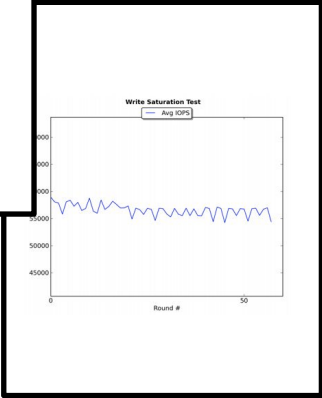
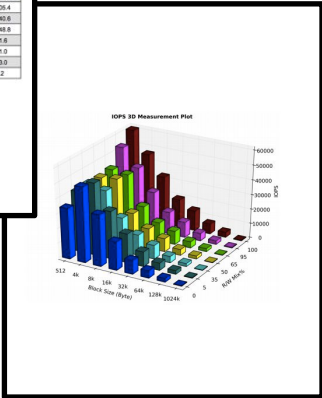
If these two conditions are met the steady state has been reached for the specific dependence variable. Therefore, the test can be stopped and the performance values of the measurement window can be taken for the final test results. The numbers for the test variables are:

- Measurement window: 6

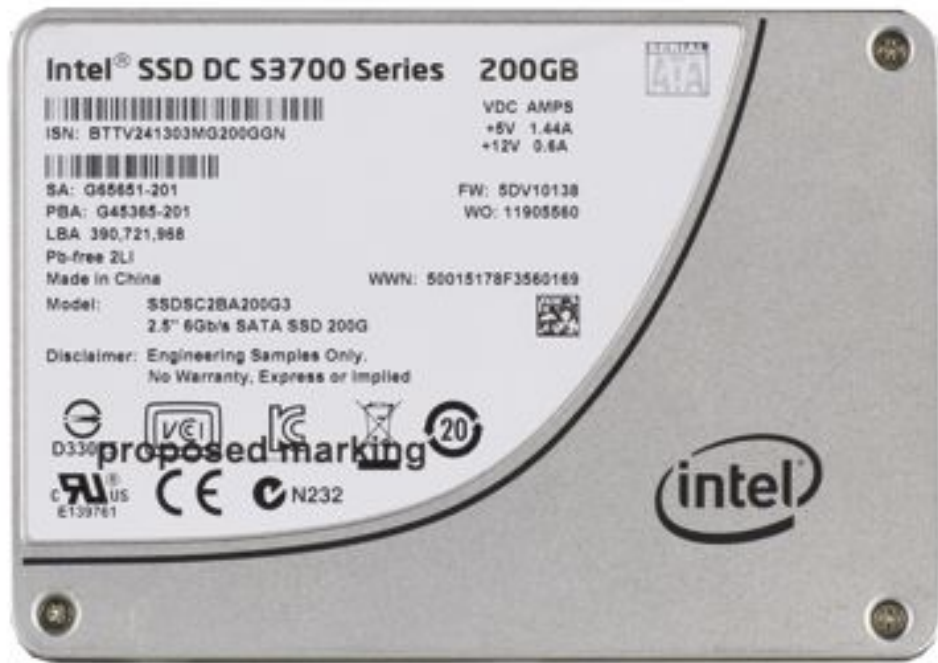
<http://www.thorsten-sonnen.com/> Page 2 of 12

Average IOPS vs. Block Size and RW Mix %

Block Size	100%	90%	80%	70%	60%	50%	40%	30%	20%	10%
512	60205.4	63360.2	42985.2	41482.8	41864.6	41196.0	34573.0			
1024	46482.0	41588.2	41586.2	42856.4	39593.2	49120.4	61005.4			
2048	33073.2	26893.2	21620.4	21693.0	23203.6	32827.8	38140.8			
4096	23420.4	18216.8	12733.0	13637.8	14114.0	19861.2	18448.8			
8192	13854.8	10568.0	8020.0	9108.2	9484.4	10114.8	8611.6			
16384	8997.0	6169.0	4377.2	4065.8	4376.4	5086.2	6141.0			
32768	3831.6	2651.8	2146.0	2101.8	2074.8	2502.0	2463.0			
65536	432.8	373.2	275.2	270.8	266.8	321.2	308.2			



FIO



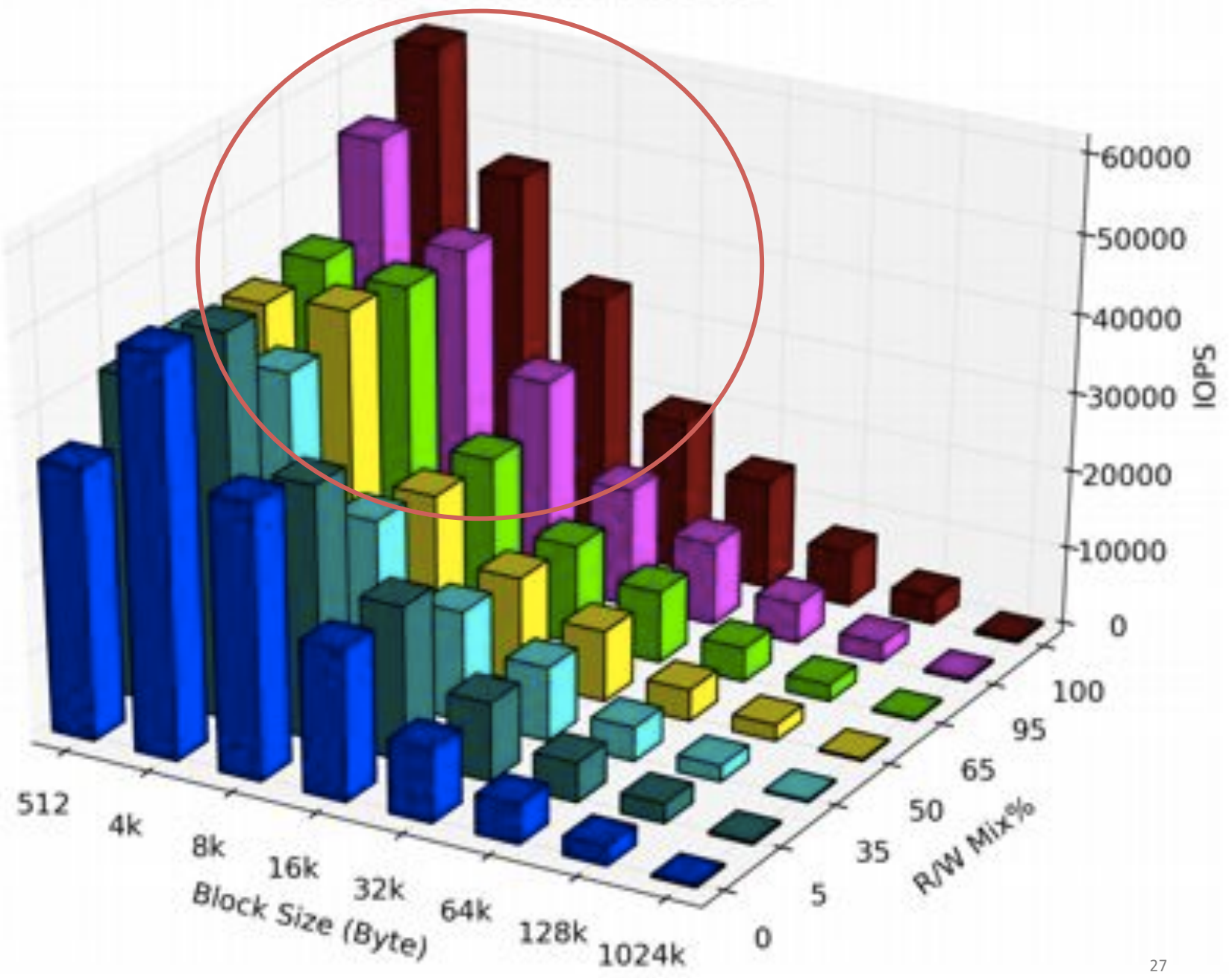

```
$ sudo tkperf ssd intels3700 /dev/sda -nj 2 -iod 16 -rfb
```

Intel S3700 200GB FW 5DV10206

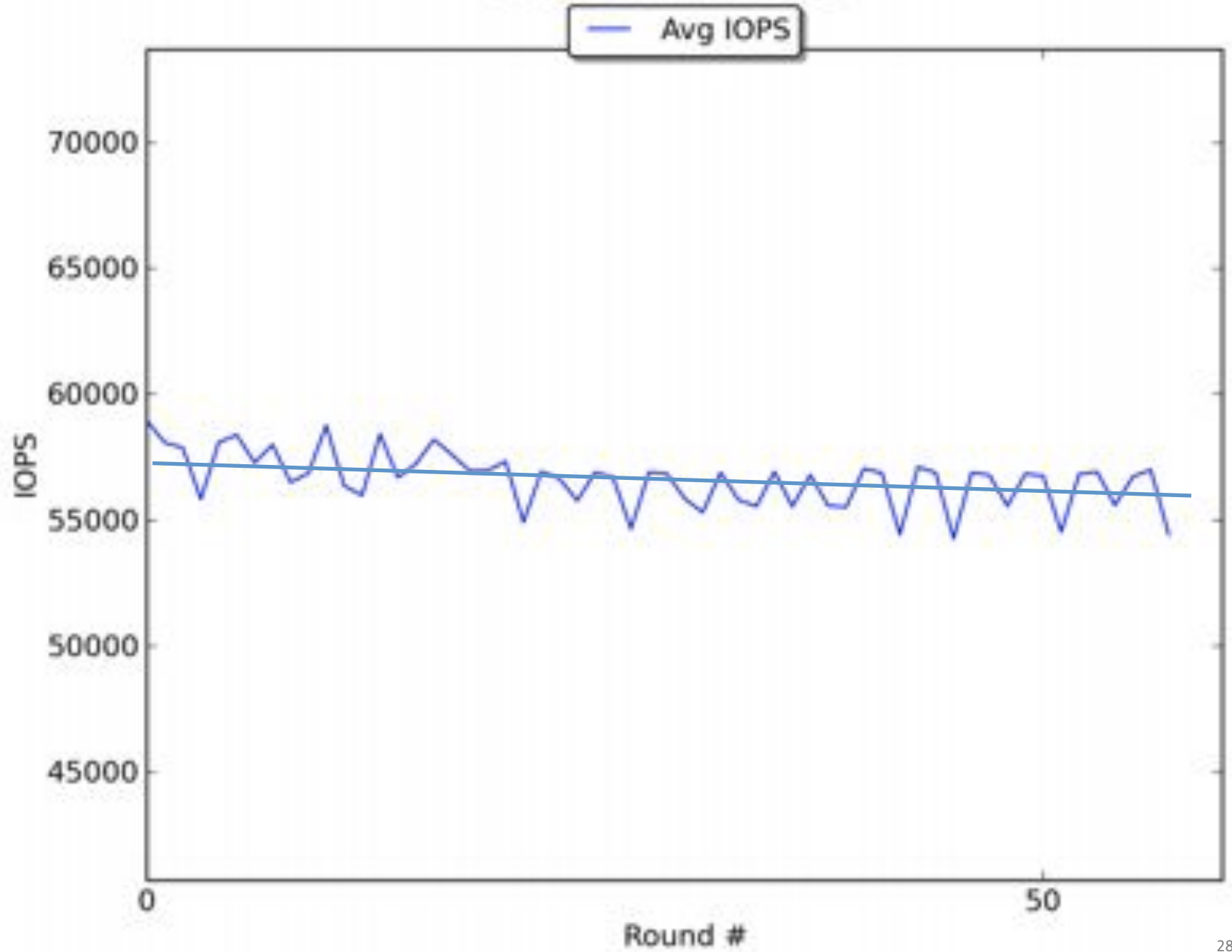
Average IOPS vs. Block Size and R/W Mix %

Block Size ↓	Wtd. → 100/0	95/5	65/35	50/50	35/65	5/95	0/100
512	60925.4	53380.2	42585.2	41462.8	41884.6	41156.0	34573.0
4k	46492.0	41586.2	41356.2	42856.4	39593.2	40120.4	51005.4
8k	33073.2	26893.2	21630.4	21685.0	23293.8	32627.8	35140.6
16k	19430.4	15215.6	12733.0	13487.8	14114.0	19861.2	19448.8
32k	13834.8	10569.0	9200.0	9108.2	9494.4	10114.6	9511.6
64k	6997.0	5169.0	4377.2	4055.8	4376.4	5066.2	5141.0
128k	3531.8	2531.8	2136.0	2101.8	2074.8	2525.2	2493.0
1024k	432.8	373.2	275.2	270.8	266.8	321.2	306.2

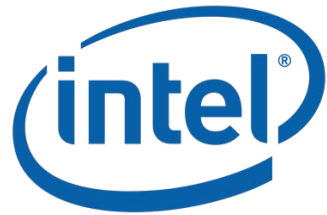
IOPS 3D Measurement Plot



Write Saturation Test



Unsere Flash Performance Tests



Bisherige Erfahrungen

- _ Boot-Datenträger (<300GB)
- _ Performance variiert extrem
- _ Zeitersparnis bei Administration
- _ Erwartung klären/definieren
- _ Selbst testen!
- _ *SSDs sind kein Allheilmittel*

Raidcontroller und SSDs

- _ Funktioniert das ohne Probleme?
 - _ Grundsätzlich JA
- _ Tuning möglich/notwendig

*Sie haben konkrete
Performanceanforderungen?
Kontaktieren Sie uns!*



Entscheidungshilfe





€/GB

80 IOPS

€/IO

30.000 IOPS





Kosten pro GB

0,08€

1,10€

Kosten pro IO

0,98€

0,01€

*Performance einer SSD:
98 HDDs → 7.742€!*

x15!

Ohne Strom- und Kühlungersparnis einzurechnen!

Arbeitslast einordnen



XX ms, variabel ok

Benötigte Latenz

<1-3ms, Konstante Rate



1k-10ks IOPS

I/O Last

10k-100ks IOPS



Schreiblastig

Zugriffsmuster

Leselastig



4k-8k

Blockgröße

Unterschiedliche Blockgrößen



Sequentiell,
Vorhersehbar,
Cachebar

Art der Zugriffe

Zufällig, nicht vorhersehbar

Wichtige Eigenschaften

- _ Stromschutz
- _ Haltbarkeit / Garantie
- _ Geschwindigkeit

Der feine Unterschied

Intel DC S3500

_ Write 11.000

_ Read 75.000

_ 275 TBW

_ 0,15 DWPD

Intel DC S3700

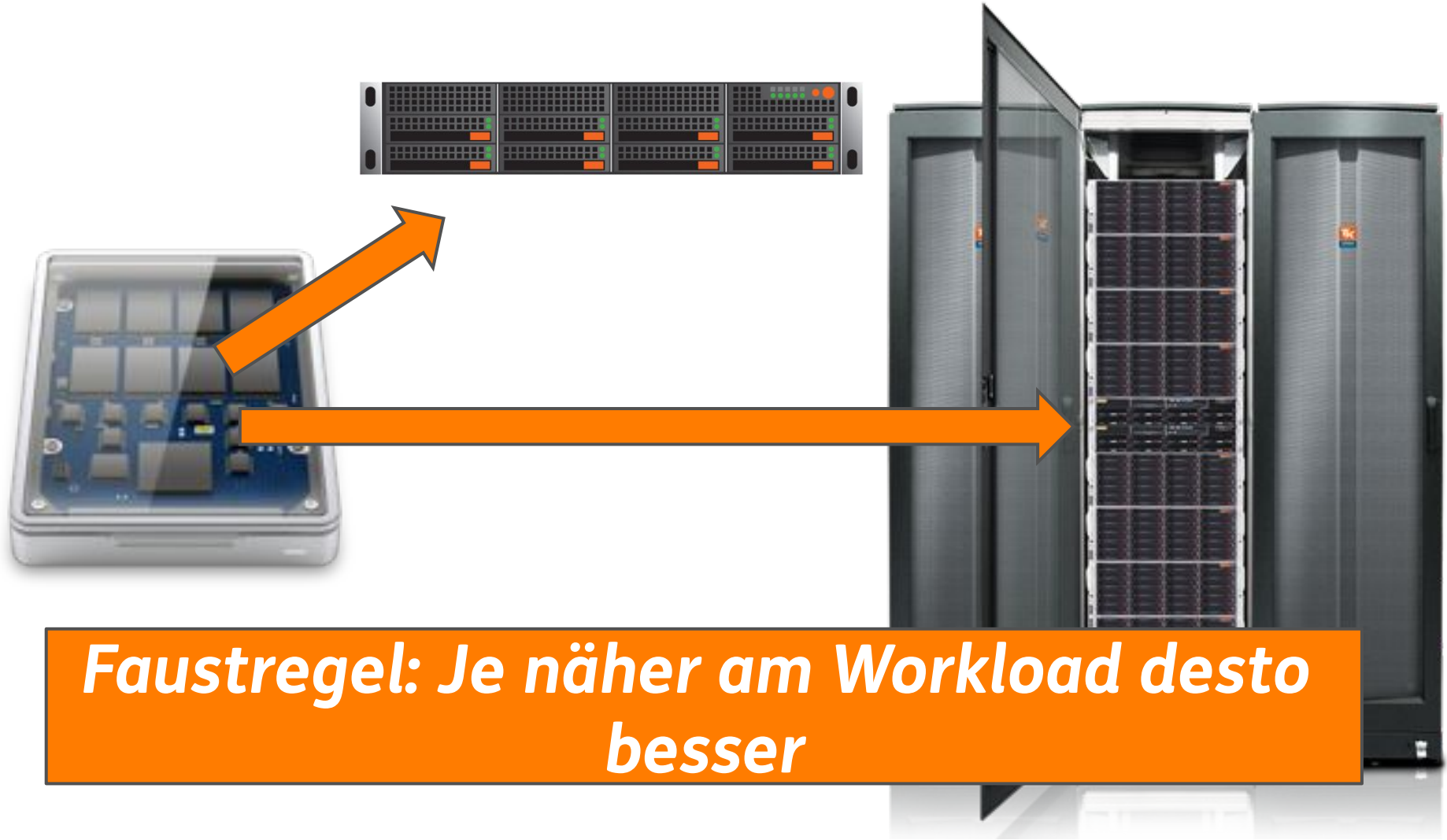
_ Write 36.000

_ Read 75.000

_ 7,3 PBW

_ 10 DWPD

Wohin mit dem Flash?



Faustregel: Je näher am Workload desto besser



Ausblick



Ausblick

Allgemein

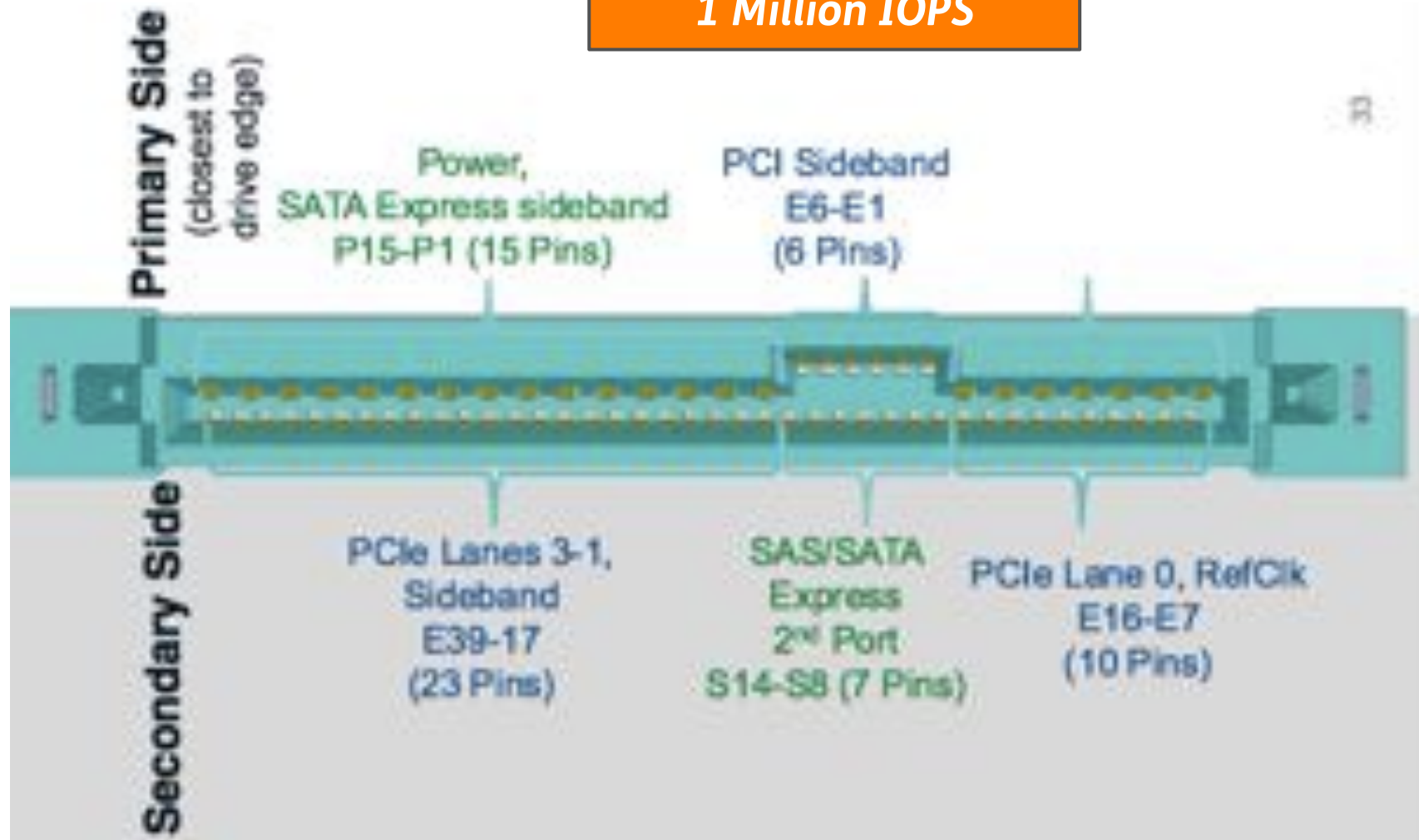
- Prognose unter 1TB werden wir nur noch SSDs sehen
- MLC wird sich durchsetzen

Technologisch

- 3D Nand (Samsung: VNAND)
- 12 Gbit SAS
- NVMe Interface

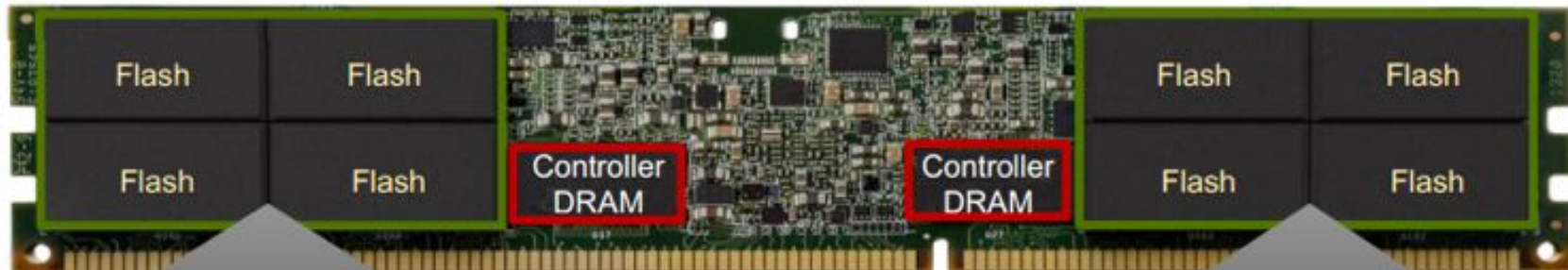
NVMe Interface SFF-8639

1 Million IOPS



Flash DIMM

- _ Geringe Latenz (5-10 Microsekunden)
- _ 200/400GB Speicher
- _ Leichter Aufrüsten
- _ Aber: Treiber notwendig



Fazit

- Produkte sind stabil geworden
 - ABER: Firmware Updates beachten
- Entwicklung noch am Anfang
- Genaue Kenntnis über Workload ist entscheidend!

€ pro IO

Weitere Informationsquellen

[http://www.thomas-krenn.com/de/wiki/
Intel DC S3700 Series SSDs](http://www.thomas-krenn.com/de/wiki/Intel_DC_S3700_Series_SSDs)

[http://www.thomas-krenn.com/de/wiki/Solid-
State Drive](http://www.thomas-krenn.com/de/wiki/Solid-State_Drive)

<http://www.thomas-krenn.com/de/wiki/TKperf>

Tkmag:

[https://www.thomas-krenn.com/de/tkmag/
expertentipps/unsere-tkperf-performance-
ergebnisse-der-uebersicht/](https://www.thomas-krenn.com/de/tkmag/expertentipps/unsere-tkperf-performance-ergebnisse-der-uebersicht/)

**Vielen Dank für Ihre
Aufmerksamkeit!**



