

Dr. Axel Koester

Technologist – IBM Systems & Technology Group



Ersetzt „Solid State Memory“ die Festplatte?



Wird Speicher weiterhin billiger?

Ersetzt Flash Memory die Festplatte?

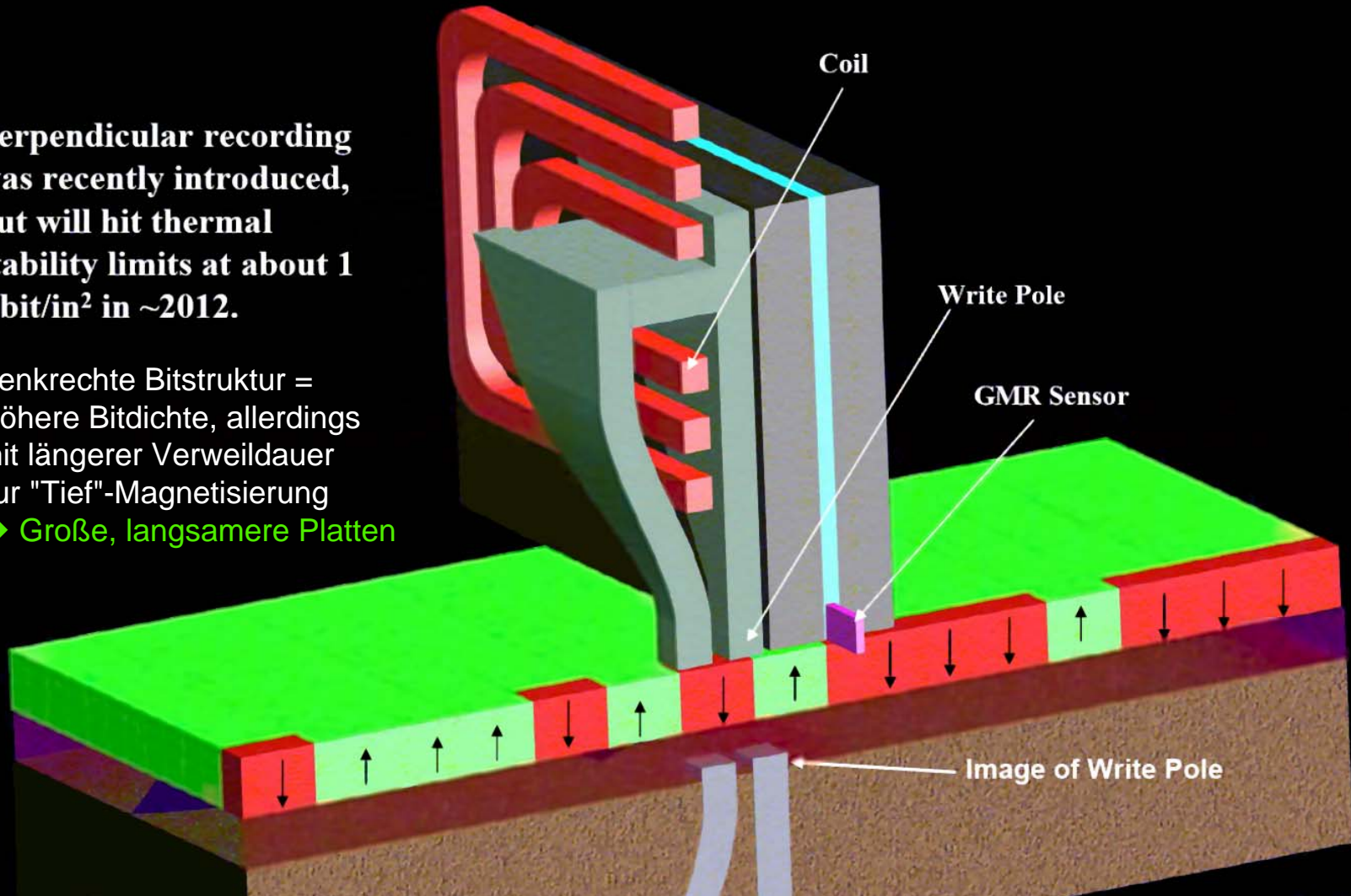
Quo vadis, Virtualisierung?

Wie kann Speicher weiterhin billiger werden?

Moderner Schreib/Lesekopf für senkrechte Bits

Perpendicular recording was recently introduced, but will hit thermal stability limits at about 1 Tbit/in² in ~2012.

Senkrechte Bitstruktur =
Höhere Bitdichte, allerdings
mit längerer Verweildauer
zur "Tief"-Magnetisierung
→ Große, langsamere Platten

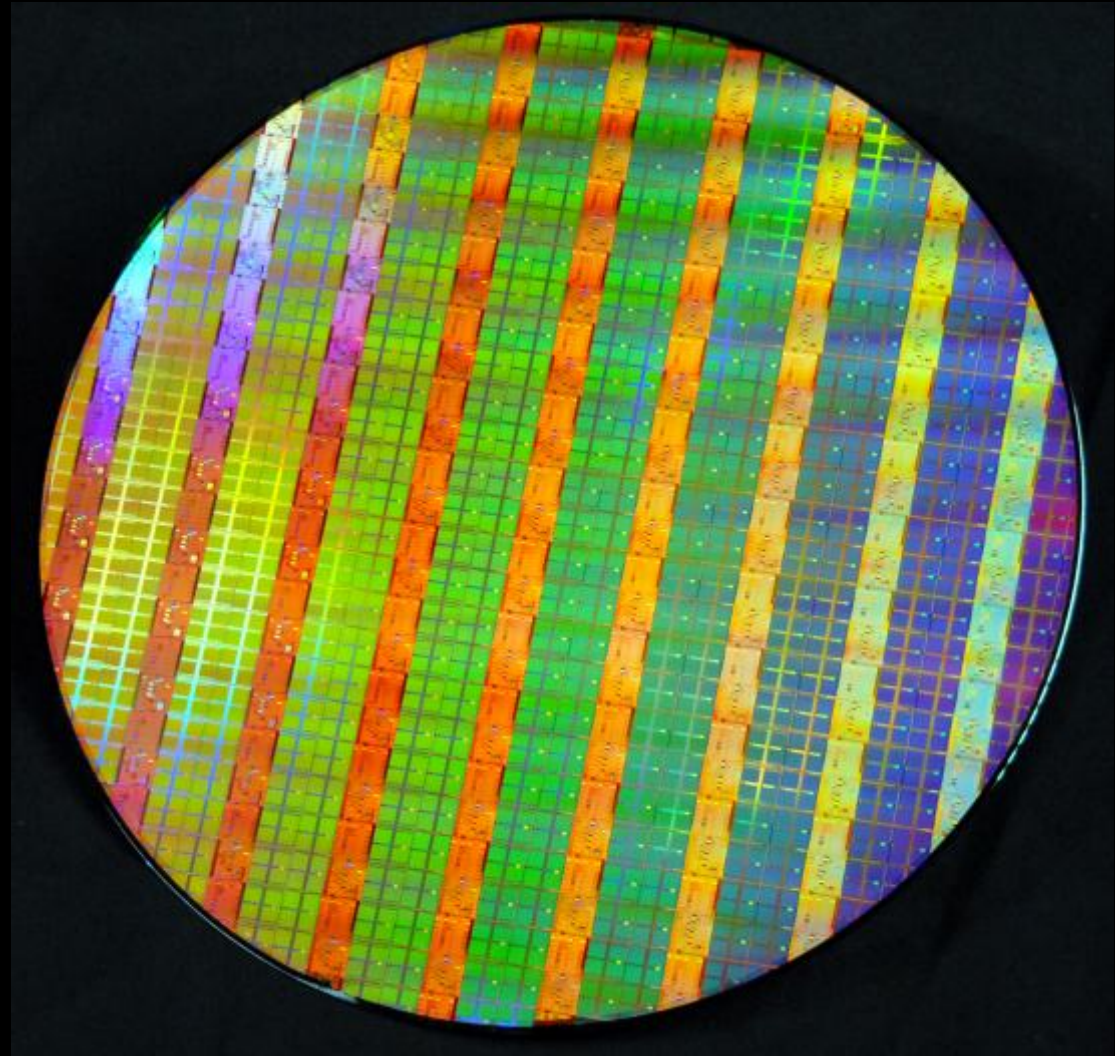


(!) Drehzahl & Thermische Bitstabilität

Moderner Chipwafer (30cm Ø, 45 nm Struktur)

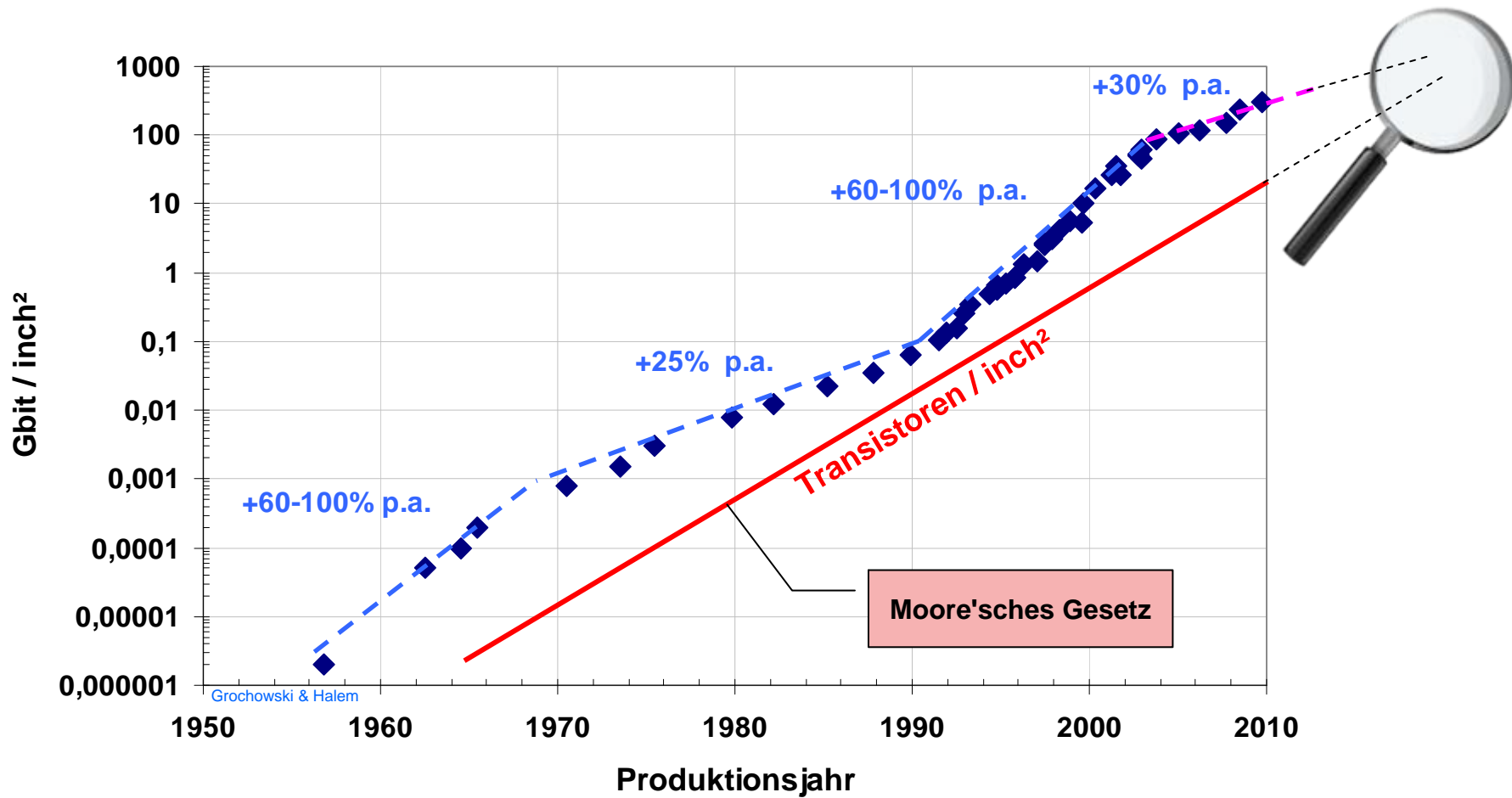
Freiheitsgrade zur Vergünstigung
des Preises von Chip-Speicher:

1. Größere Waferdurchmesser
2. Feinere Lithografie
3. Mehr Bits/Transistor



- (!) Durchmesser
- (!) Ausschuss
- (!) Lithografiefortschritt < 40nm

Speicherdichten-Trends : Disk ---- Chip —



Gordon Moore, April 1965 : Verdopplung der Transistordichte alle 1,5...2 Jahre

Werden Festplatten von Flash Memory verdrängt?



Flash Vorteile überwiegen bei mobilen Geräten



Disk

*Verzicht auf
Mehrkapazität* →

Flash



Online-Bandbreite fördert Flash in PCs



mit Festplatte:
Kapazität

**Fotos, Video,
Audio, Präsentationen,
Dokumente, Software**

*Verzicht auf
tragbare Kapazität* →

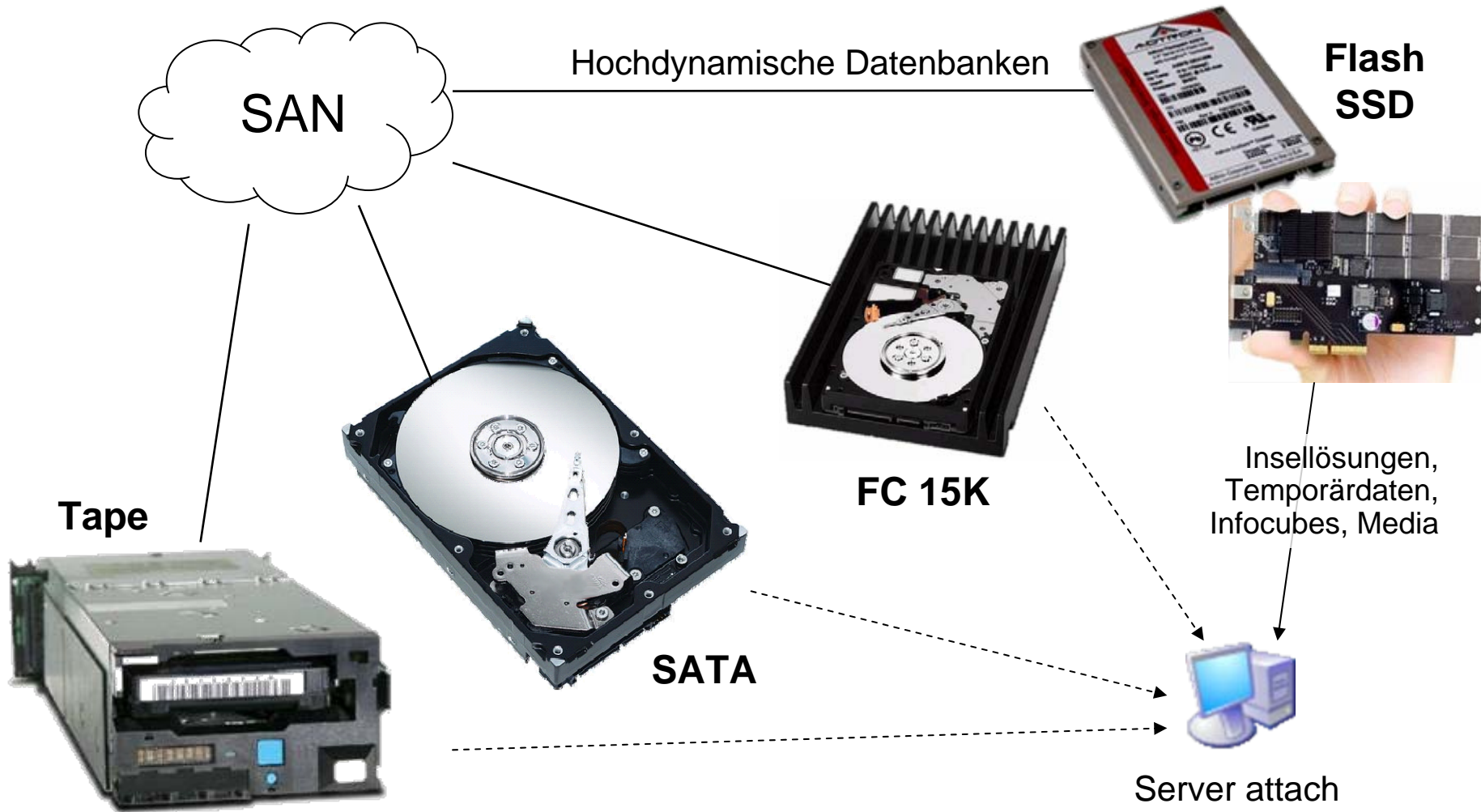


mit Flashdrive:
Low Power

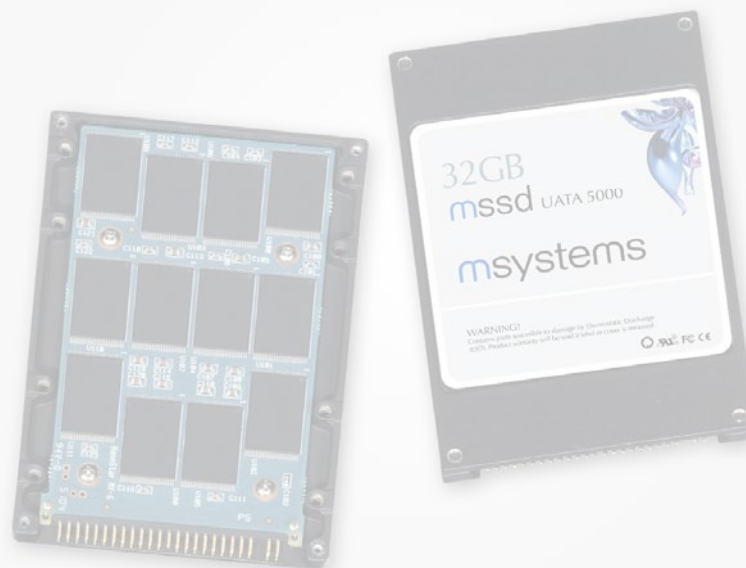
**Verzicht auf Kapazität →
Auslagerung auf Homeserver**



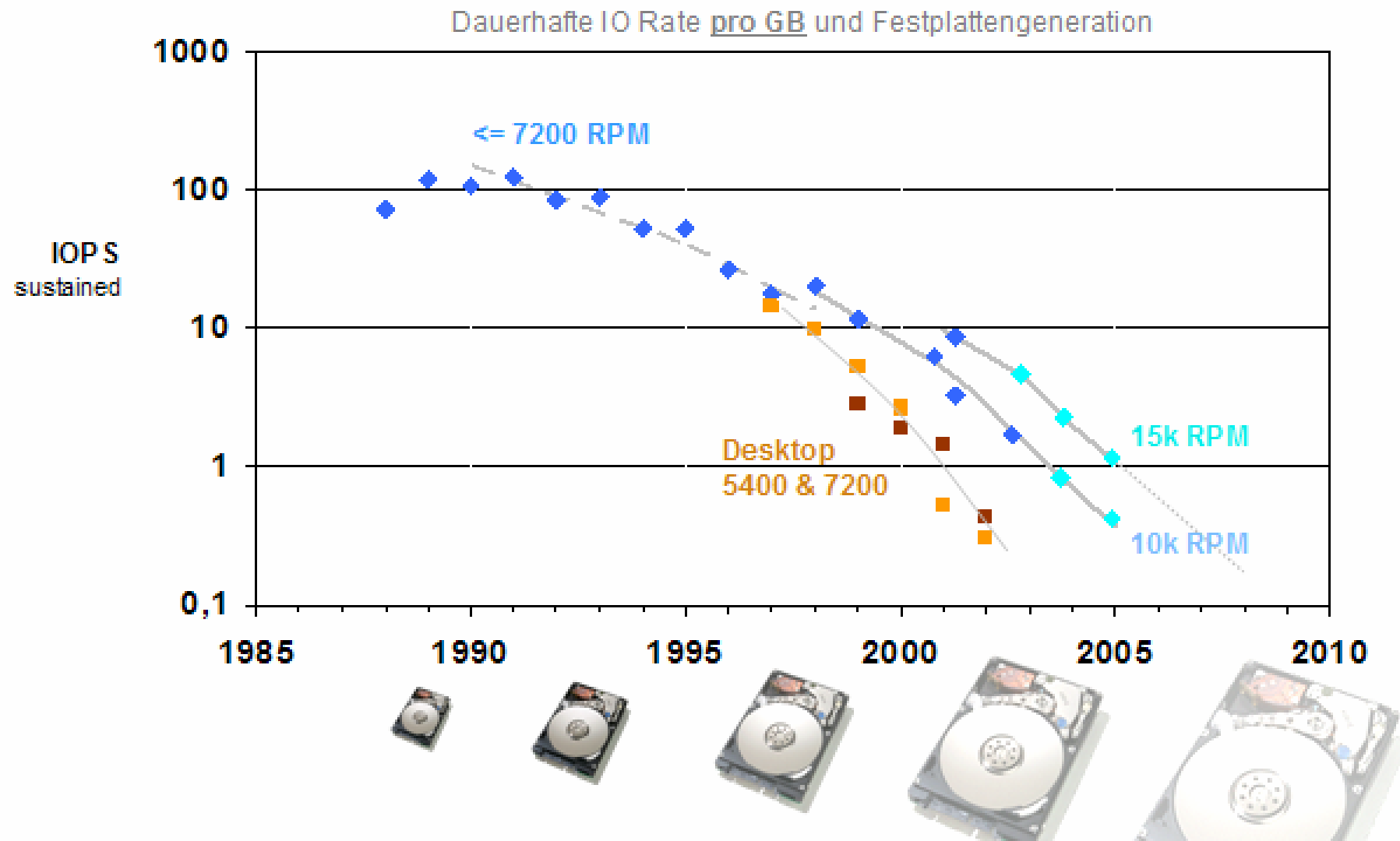
Speicherklasse "Flash" im Rechenzentrum



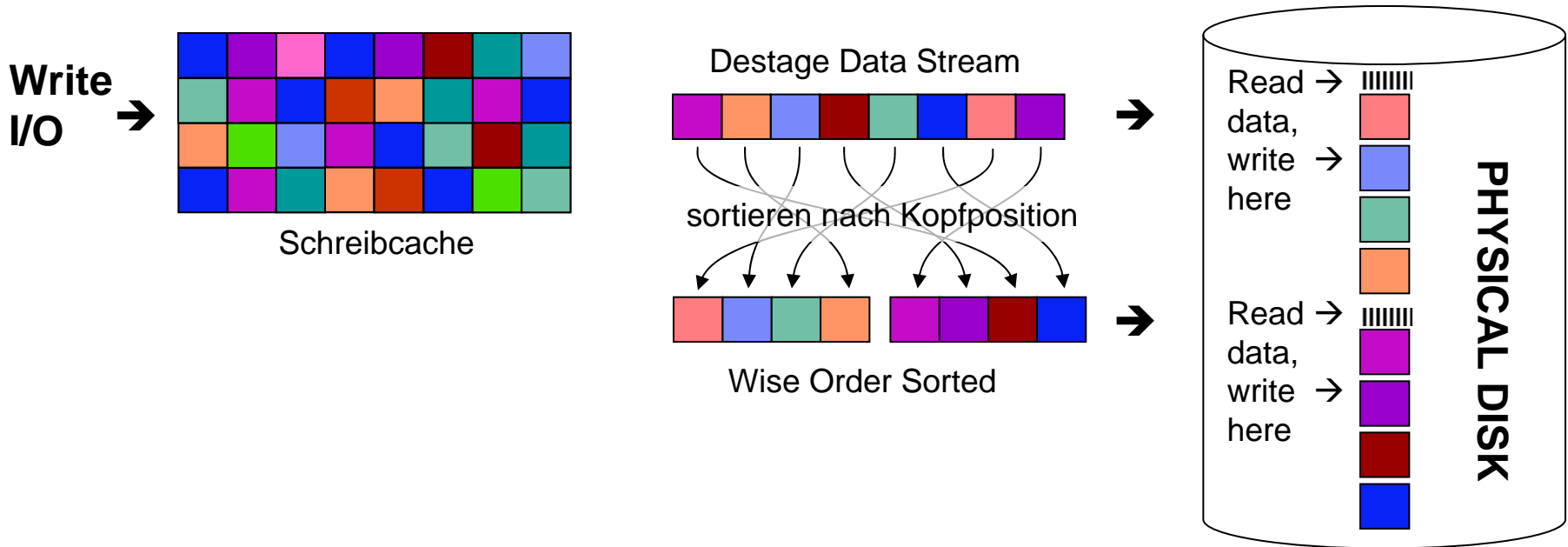
Disk Technologiegrenzen



Festplatten-Datenrate *pro Gigabyte* fällt dramatisch

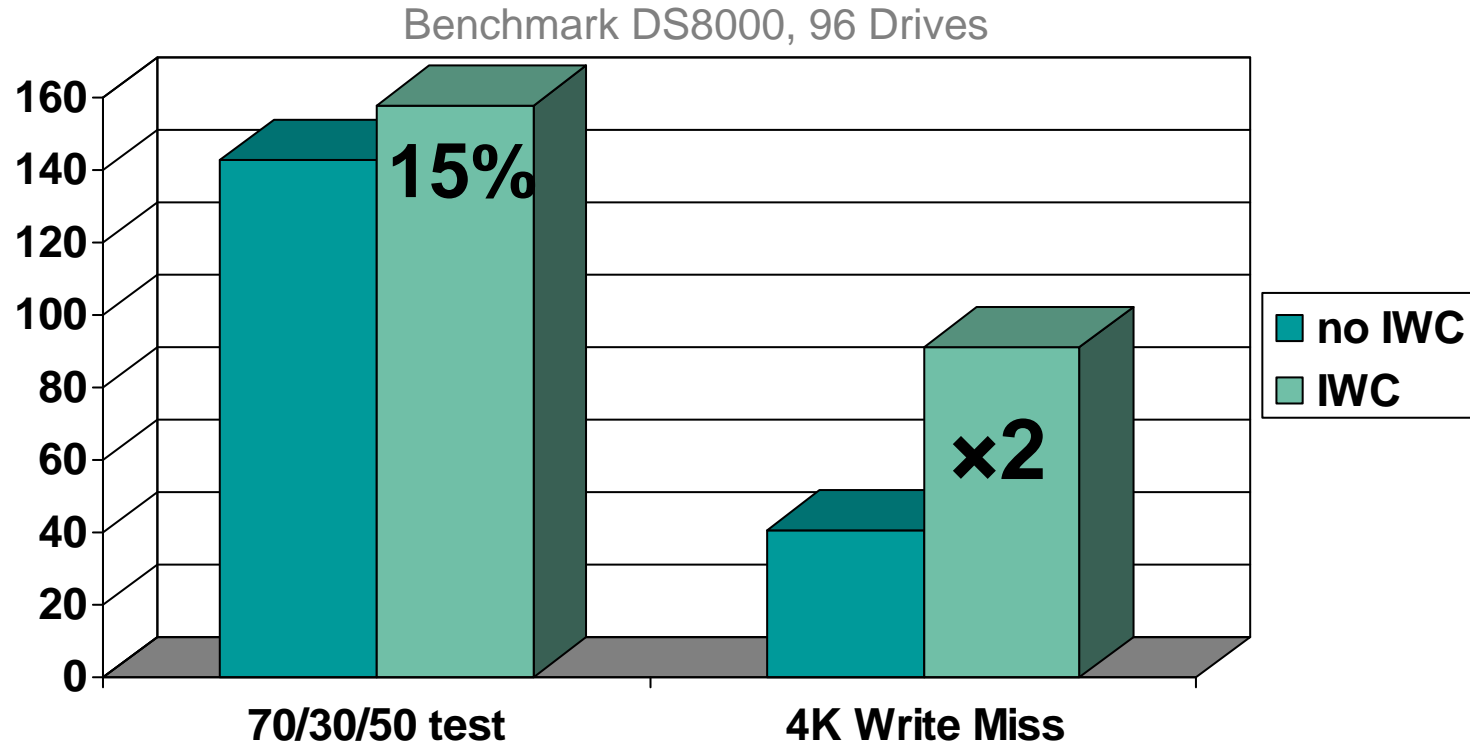


Neu: Intelligent Write Caching (seit IBM DS8000 R4.2)



- Optimiert für **minimale Kopfbewegungen**
- Verringerter Schreib-Overhead durch Ausnutzung der aktuellen Kopfposition
- "Wise Order Writes"

Intelligent Write Caching = Turbo für Datenbanken

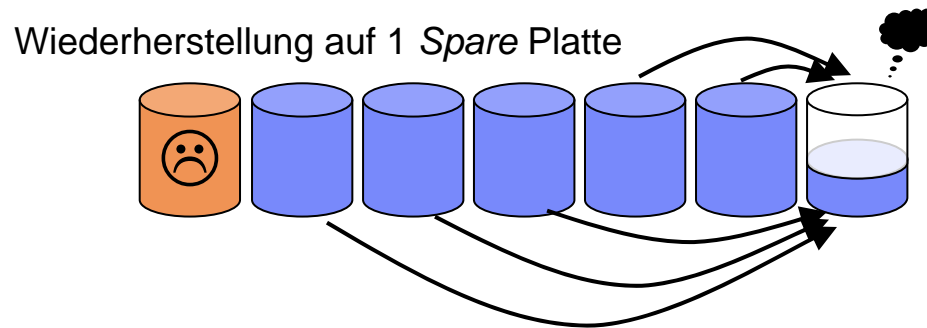


RAID 10, 15K RPM, mix of 146, 300 and 450 GB
(64) (16) (16)

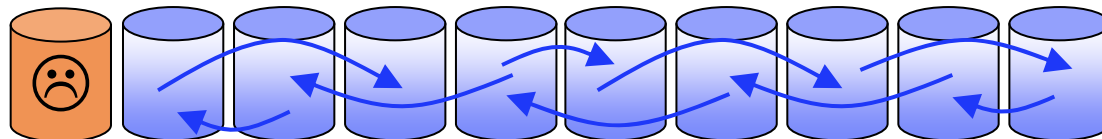
Bessere Verfügbarkeit trotz größerer Kapazität



Klassischer RAID Schutz bei Terabyte-Platten

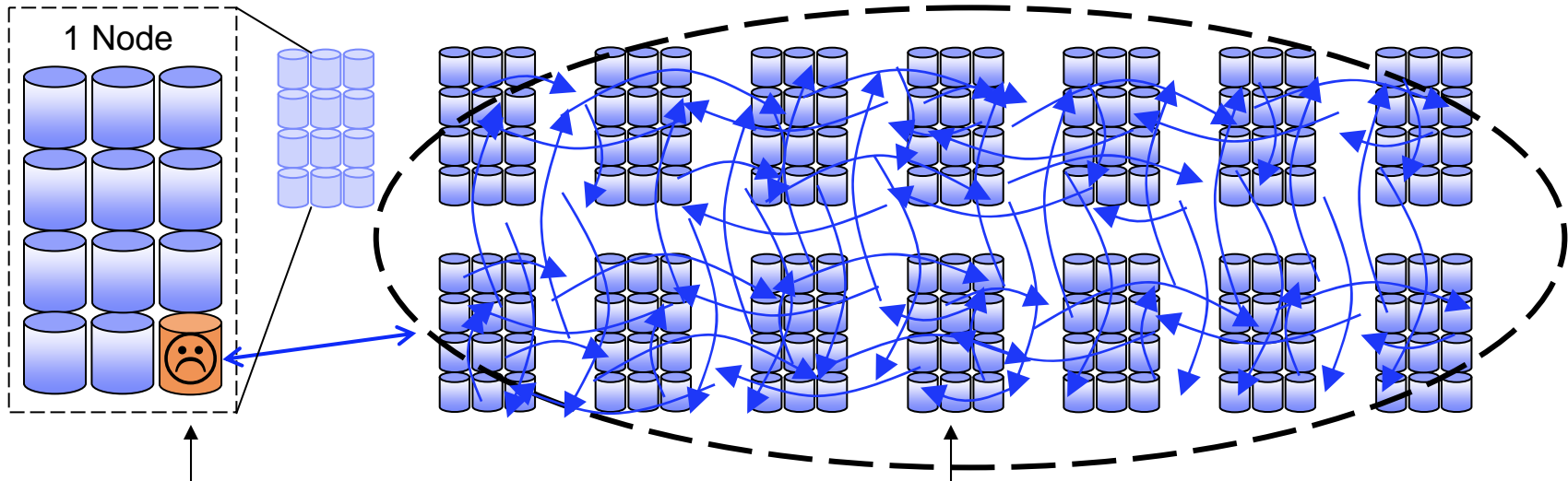


- **Declustered RAID:** parallele Wiederherstellung der Redundanz



Jede Platte handhabt 1/10 der Wiederherstellungs-Arbeit in 1/10 der Zeit
Schneller, umso mehr Platten involviert sind

Das Redundanzprinzip des IBM XIV Speichersystems



Redundanzdaten jeder Platte sind auf allen übrigen Platten in anderen Nodes verteilt.

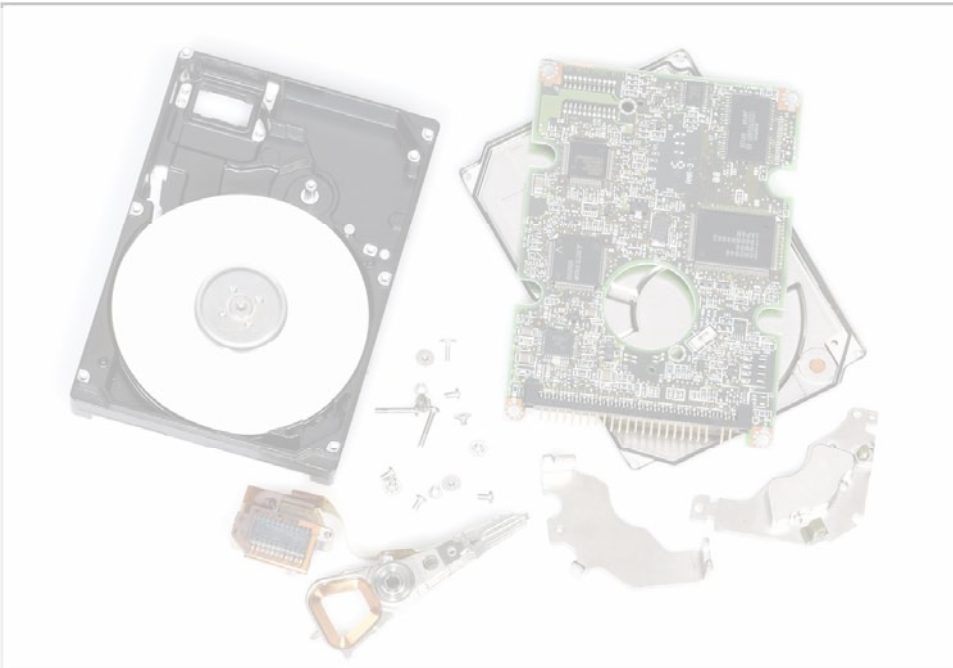
Paralleler Rebuild-Prozess auf allen Platten zugleich
(180 pro Rack, abzgl. 12)

≤ 30 min

Redundanz kann mehrfach hintereinander wiederhergestellt werden.
Platten- und Node-Redundanz nutzen denselben Mechanismus.
"Worst case" Rebuild Zeit für 1TB ≤ 30min.

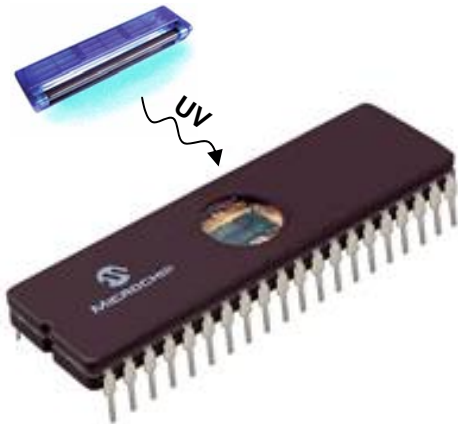


Chip Technologiegrenzen



Flash-Memory = Viele EEPROM-Blöcke

EPROM
(1970)



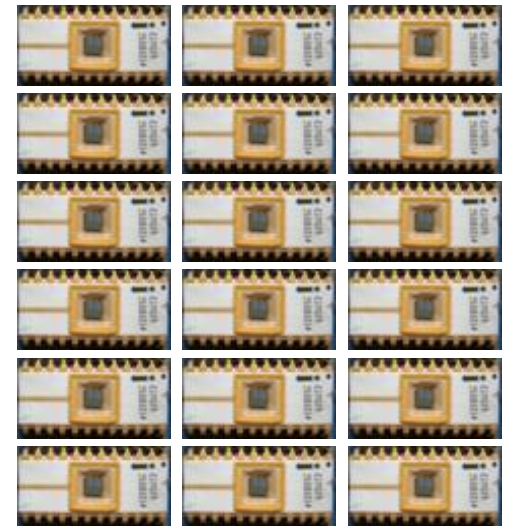
Löschen mit UV,
programmieren mit 12 V

EEPROM



Löschen mit 12V,
programmieren mit 5 V

Flash Memory



Blockweise
Löschen

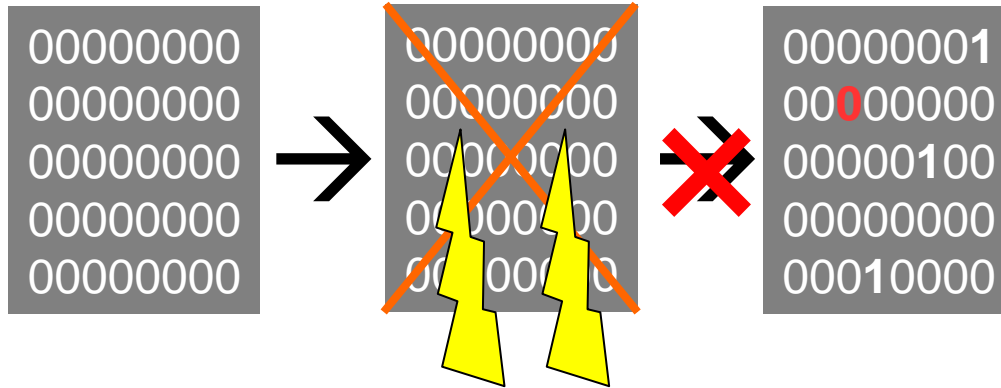
(max. 100.000x)



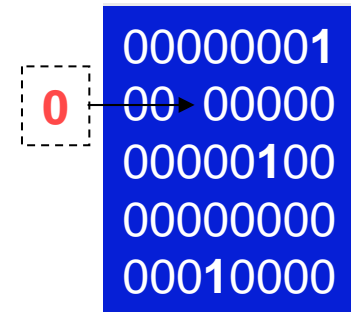
Kein Schreiben ohne Löschen

Löschimpuls = Alterung

Flash-Datenblock

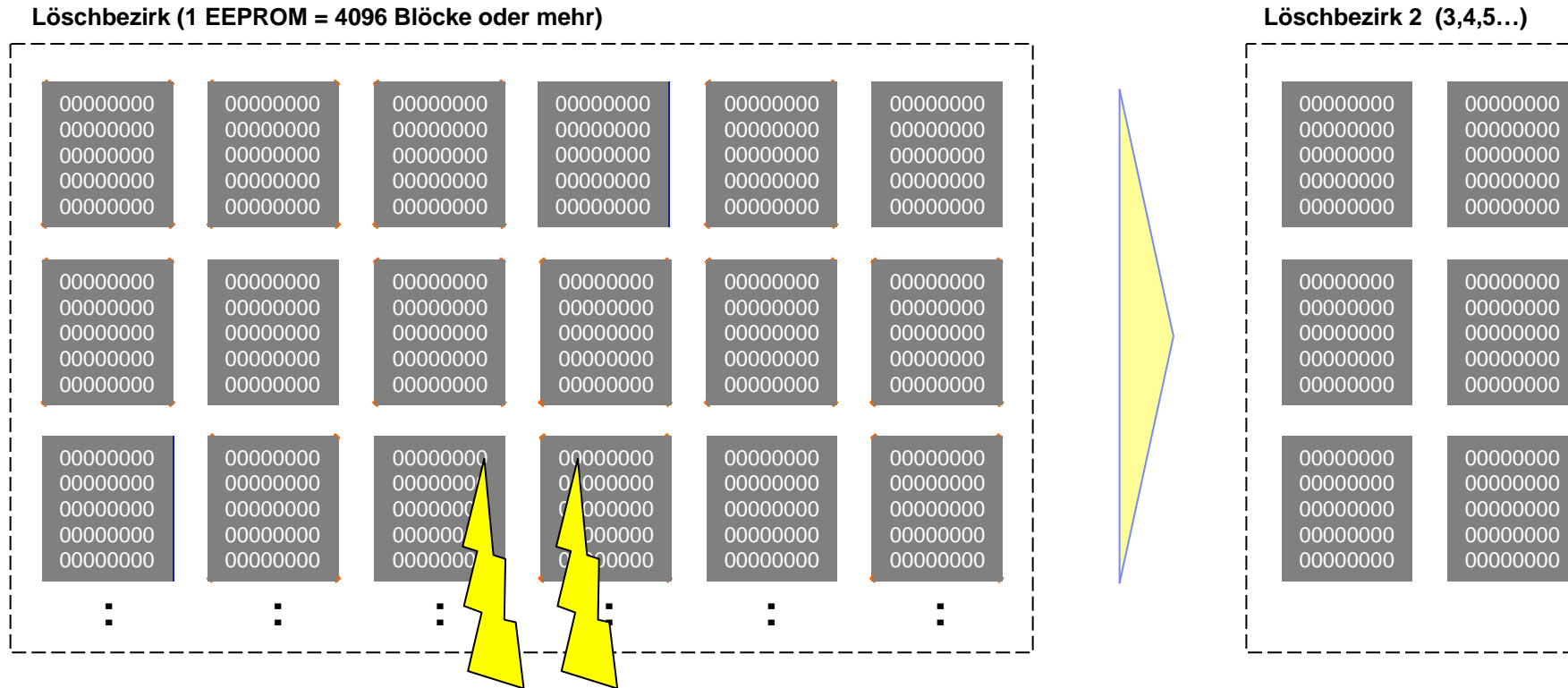


Schreiboperationen
nur auf gelöschte Blöcke



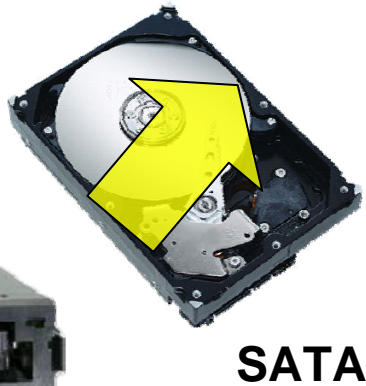
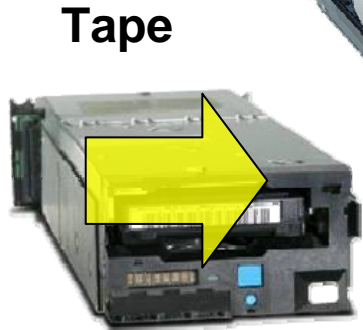
- Alte Blöcke bleiben bis zum Löschimpuls **gesperrt**
- 100.000 Löschimpulse je Block möglich, daher optimal streuen
- **Völlig andere optimale Cache Algorithmen (Lebensdauer!)**

Garbage Collection: verlängert die Lebensdauer

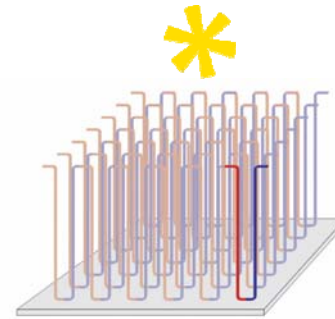


- Seltener löschen = längere Haltbarkeit, bessere Performance
- Im Vergleich zu DRAM aufwändiges *wear levelling*
- **Viel Overprovisioning = höherer Preis**

Speicherklassen im Rechenzentrum – was ist zu erwarten?



FC 15K
SAS 15K

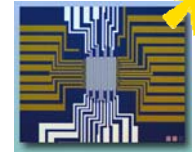


3D Chip-
Memory



Flash

Non-volatile
RAM *



Speicherklassen im Rechenzentrum

Heute:

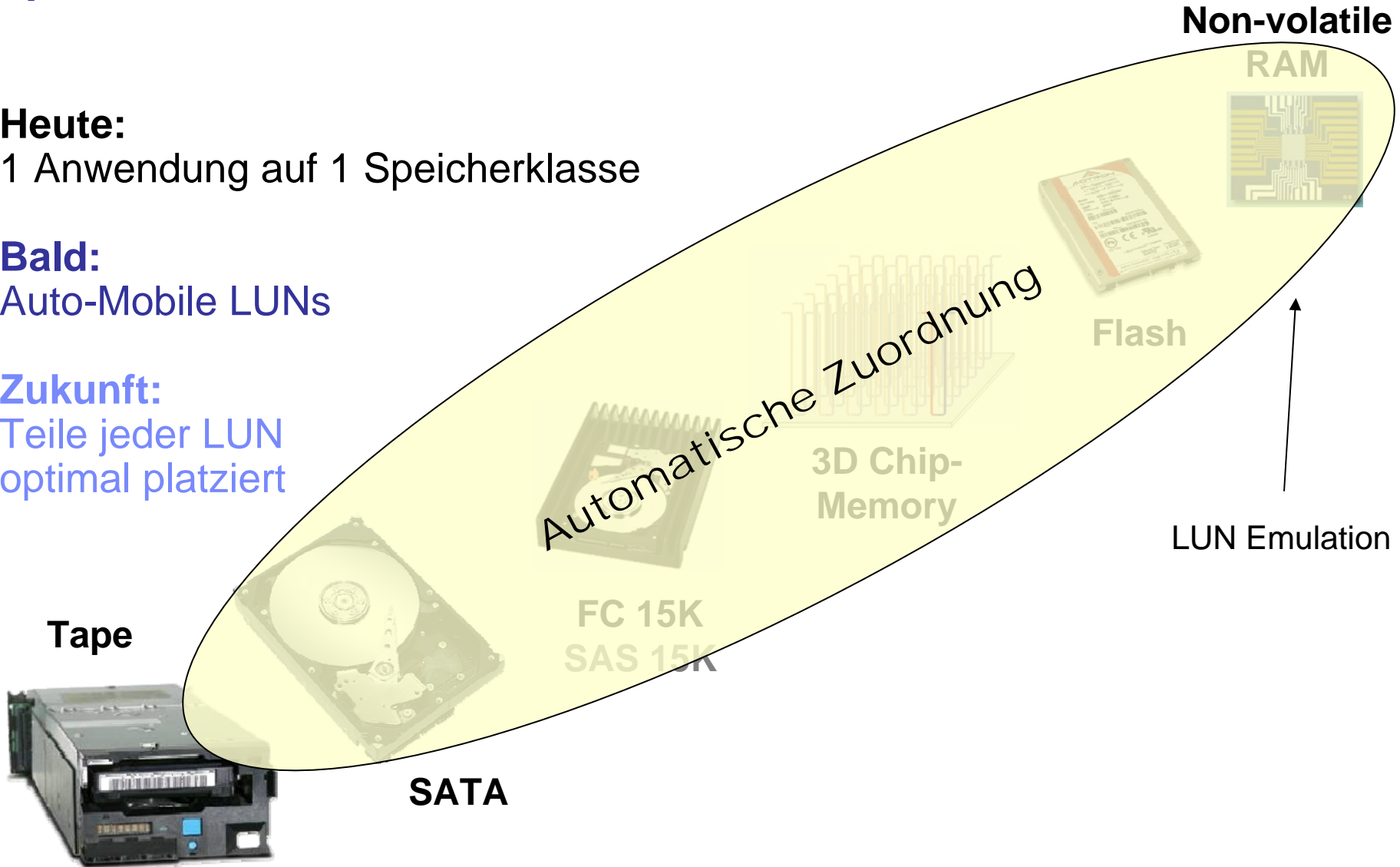
1 Anwendung auf 1 Speicherklasse

Bald:

Auto-Mobile LUNs

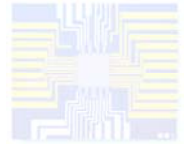
Zukunft:

Teile jeder LUN optimal platziert

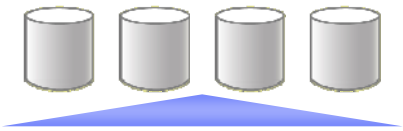


SAN Volume Controller R5: QoS Virtualisierung

Preview



Applikationen

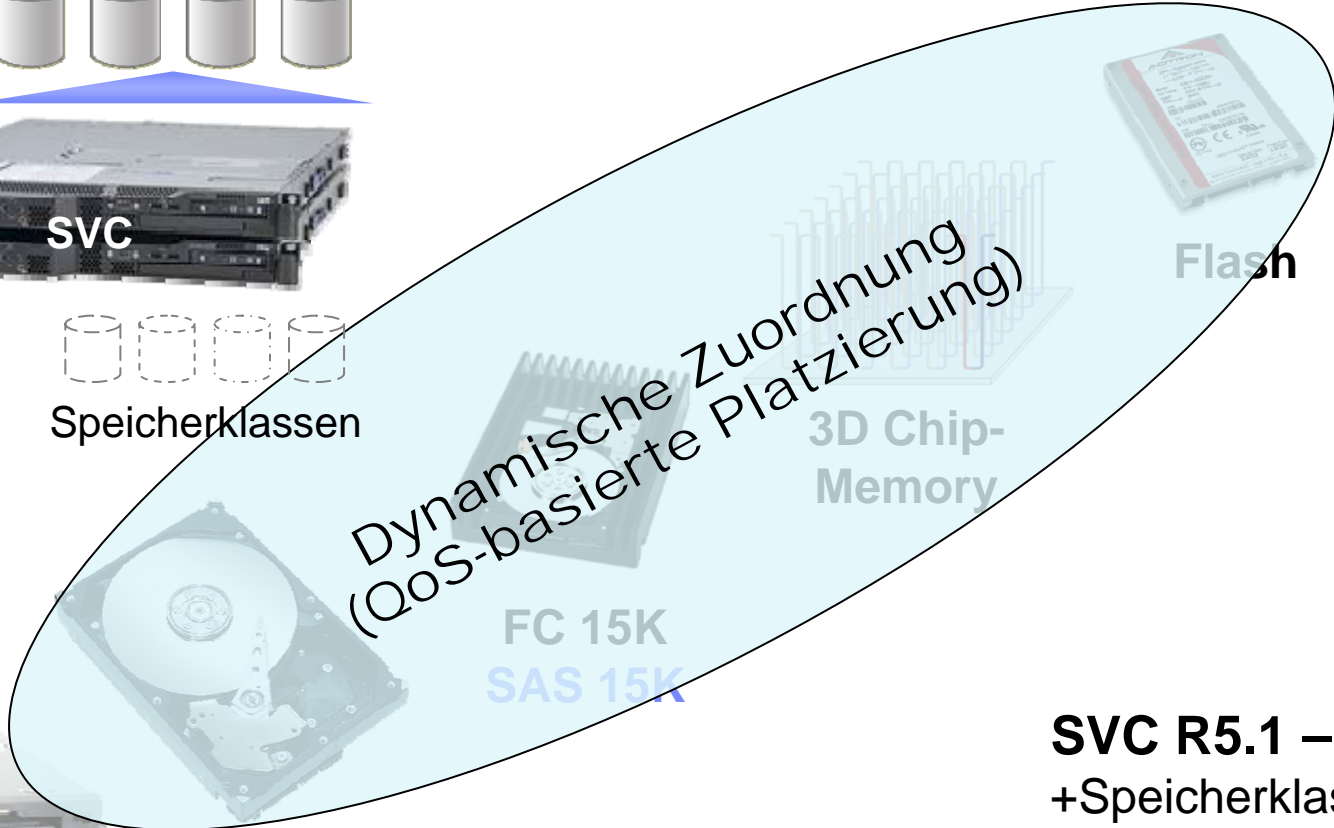


SVC



Speicherklassen

Speicherklassen



Dynamische Zuordnung
(QoS-basierte Platzierung)

FC 15K
SAS 15K

3D Chip-Memory

Flash

Tape



SATA

SVC R5.1 – Q4 2009
+Speicherklasse Flash

SVC R5.x – 2010
+QoS Virtualisierung

Screenshot: SVC auto-tiering Software

Aspect: Total capacity Object color: Object type

BVQ ▶ Cluster ▶ Storage class ▶ MDisk group ▶ VDisk ▶

Search...

Cluster **svcarz** 28.46 TiB

Storage class **Gold** 41.04 TiB

MDg **ds8300arz_300** 9.78 TiB

VD **esx_data_v1**
1,000 GiB

VD **fsc11_dat**
800 GiB

VD **fsc21_de**
650 GiB

VD **esx_**

VD **prac**

VD **prac_**

VD **srac_**

VD **srac_**

VD **srac_**

VD **fs**

VD **aj**

VD **e**

VD **srac_**

MDg **ds8300arz_450** 9.27 TiB

VD **backupi46_v00**
1.95 TiB

VD **esx_data**
1,000 GiB

VD **esx_dat**
600 GiB

VD **c**

VD **c**

MDg **ds8300arz_146** 3.04 TiB

VD **es**

VD **es**

VD **ori**

VD **ts**

Storage class **Bronze** 12.72 TiB

MDg **ds4700b1** 6.36 TiB

Cluster **svcrz** 27.7 TiB

Storage class **Gold** 41.04 TiB

MDg **ds8300rz_300** 9.78 TiB

VD **fsc11_dat**
800 GiB

VD **fsc21_de**
650 GiB

VD **esx_**

VD **prac**

VD **sra**

VD **sra**

VD **cd10**

VD **fs**

VD **e**

MDg **ess800rz** 6.12 TiB

VD **esx_data_v1**
1,000 GiB

VD **esx_**

VD **es_**

VD **apj**

VD **z**

VD **sv2**

MDg **ds8300rz_146** 3.04 TiB

VD **or**

VD **or**

VD **ts**

VD **ts**

Storage class **Bronze** 12.72 TiB

MDg **ds4700a1** 6.36 TiB

VD **es**

Storage class **Bronze**

MDg

2

B

Alle großen, Gold-Klasse VDisks mit geringer Last.

Per drag'n'drop in die optimale Klasse ziehen.

SVC erledigt den Rest unterbrechungsfrei.

"Auto-tiering"



**Valet Parking,
gesehen in
SAN Francisco**



axel.koester@de.ibm.com