

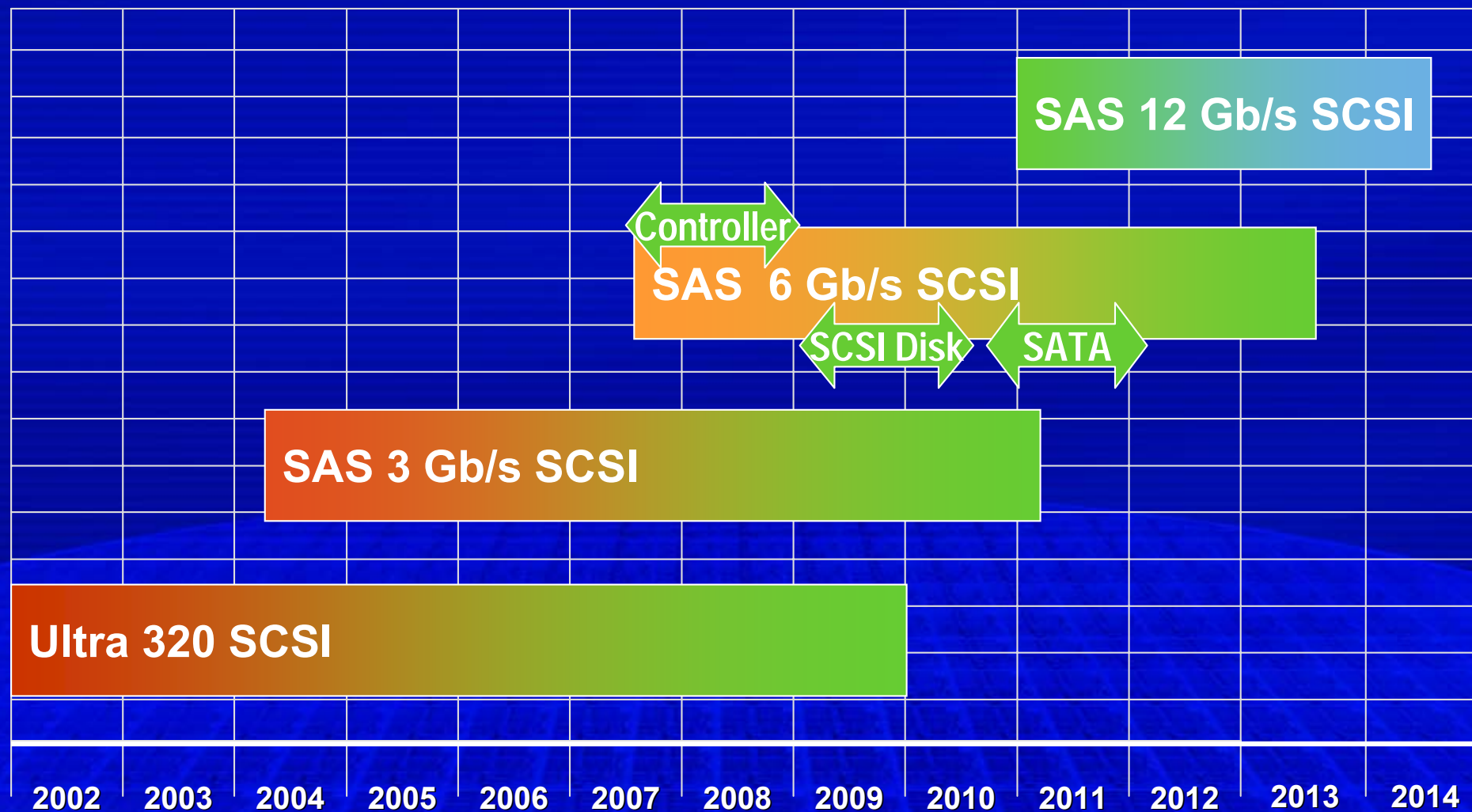
SAS 6Gb/3Gb Link Rate Matching (without reinventing the link layer)

Bob Sheffield

Intel Corporation

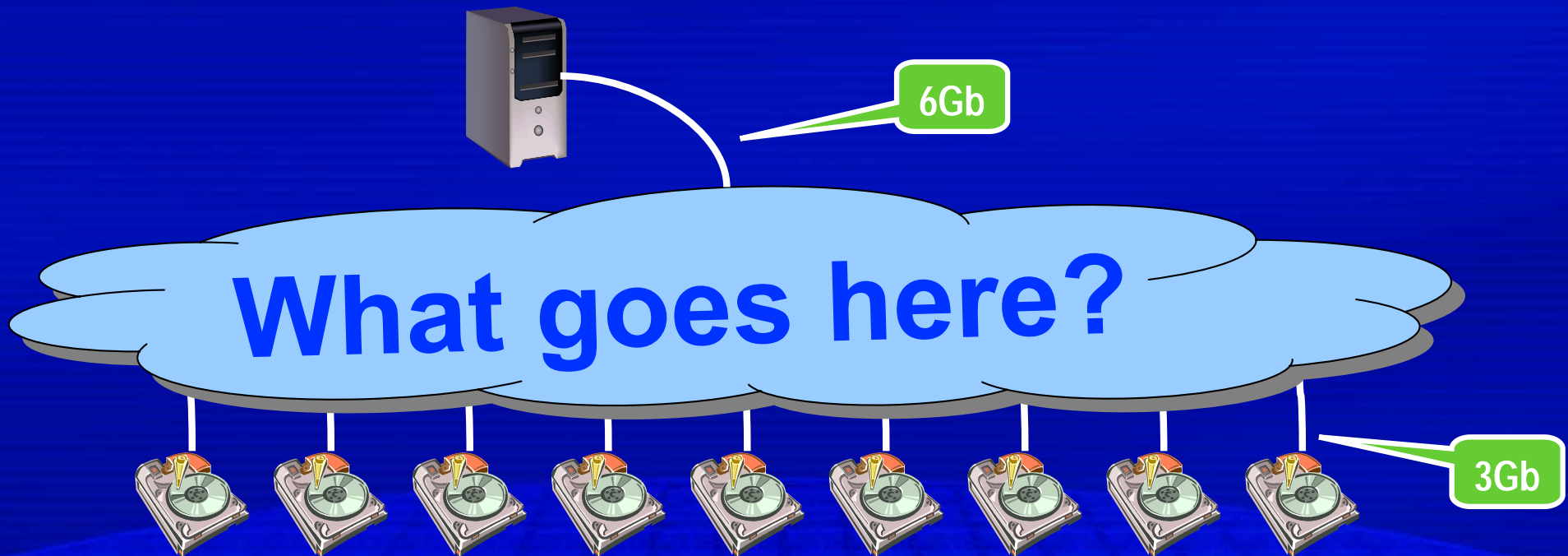
7 November 2005

SAS Roadmap (STA)



Note: Beginning of bars denote first plugfest utilizing the technology

Problem Statement

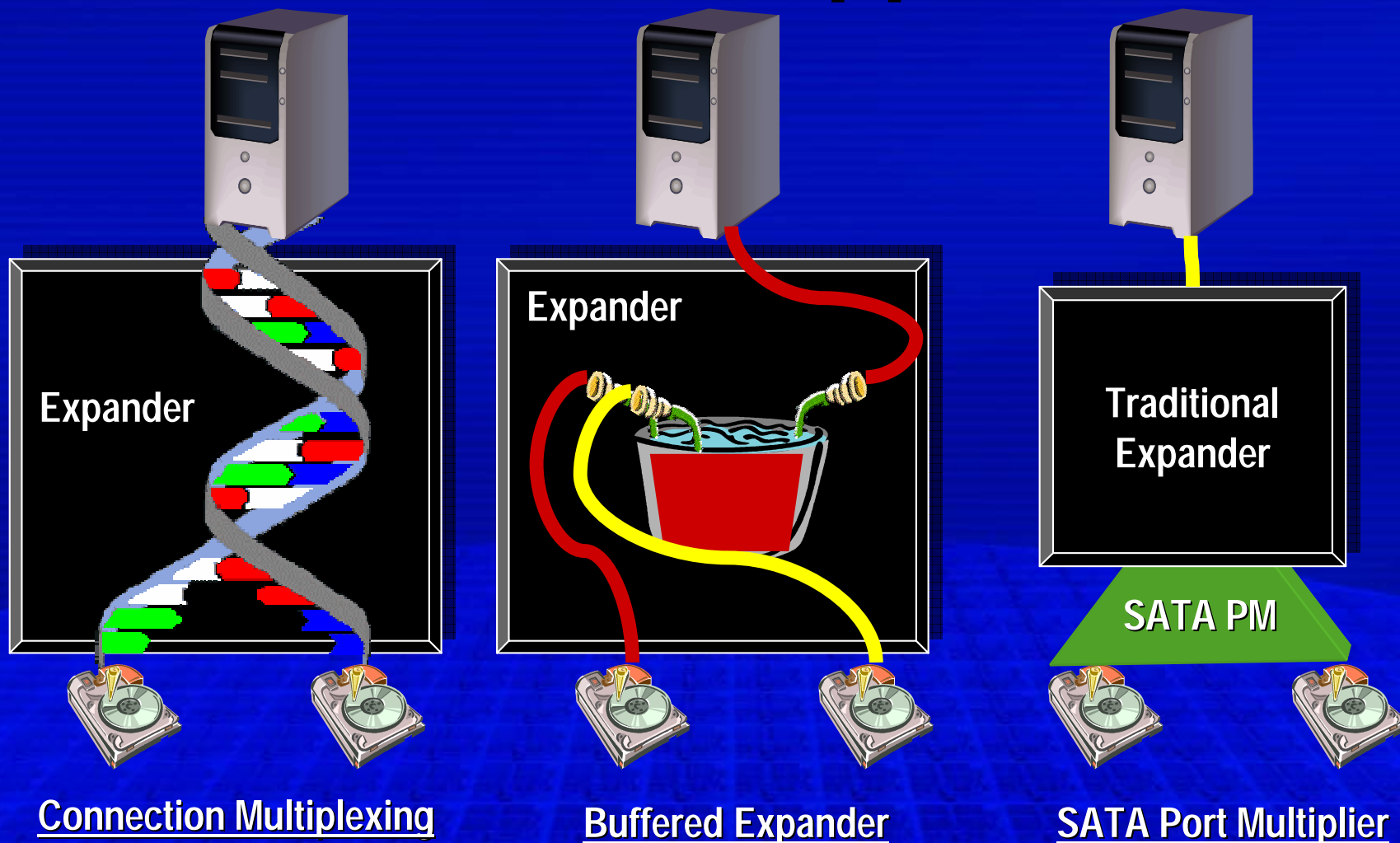


Using a connection-based transport, how do we get 6Gbps effective throughput per phy at the initiator when target phys operate at 3Gbps?

Problem goes away when targets support 6Gbps.

Solution should have *minimal* impact.

Three Basic Approaches



Need to Consider Alternate Approaches

What's the Real Benefit of Muxing?

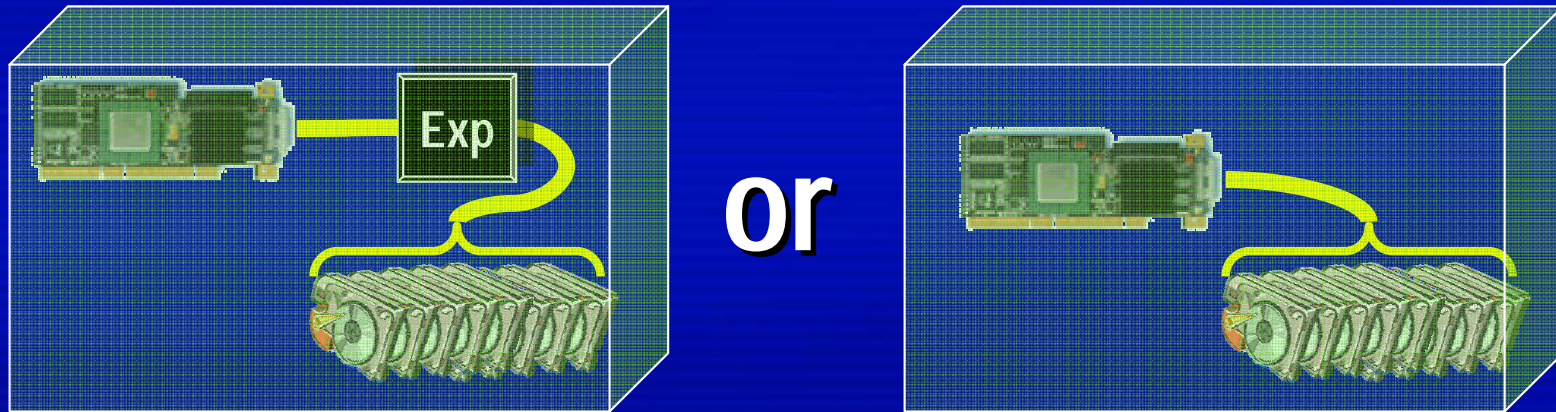


- SAS 6Gbps x 8 = 4.8 GBytes/S
- SAS 3Gbps x 8 = 2.4 GBytes/S
- PCI-e @ 5Gbps x 8 x 67% = 2.7 GBytes/S
- $2.7 / 2.4 = 12.5\%$ more throughput (optimistic)

Can't do much better than 3Gb SAS.

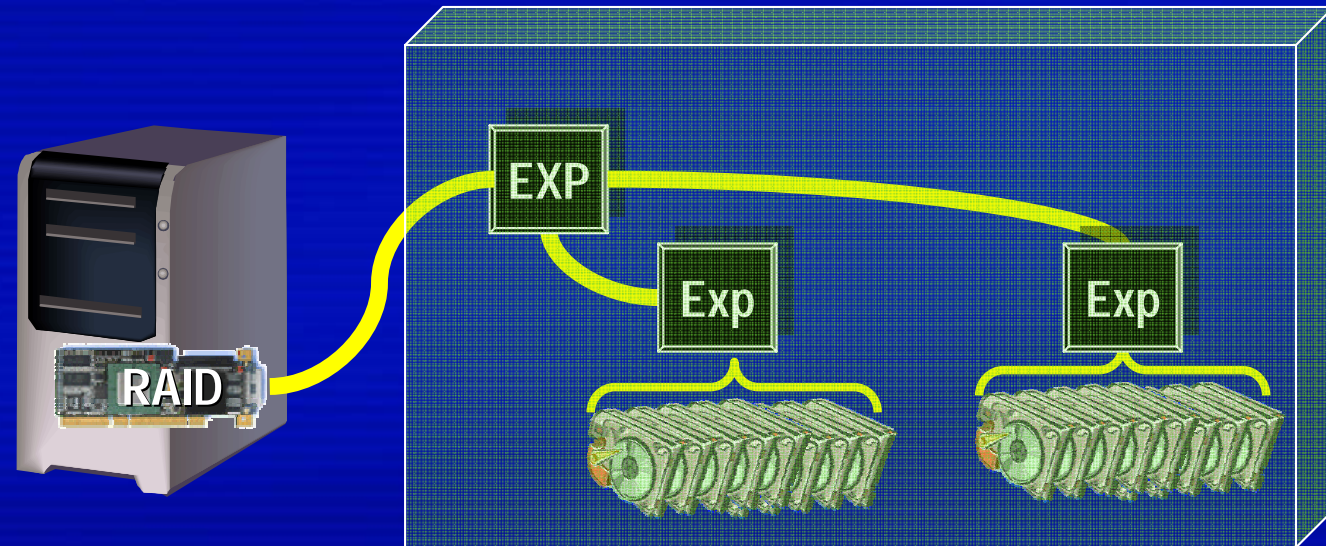
Does the benefit justify the complexity?

Internal Direct-Attach Disk



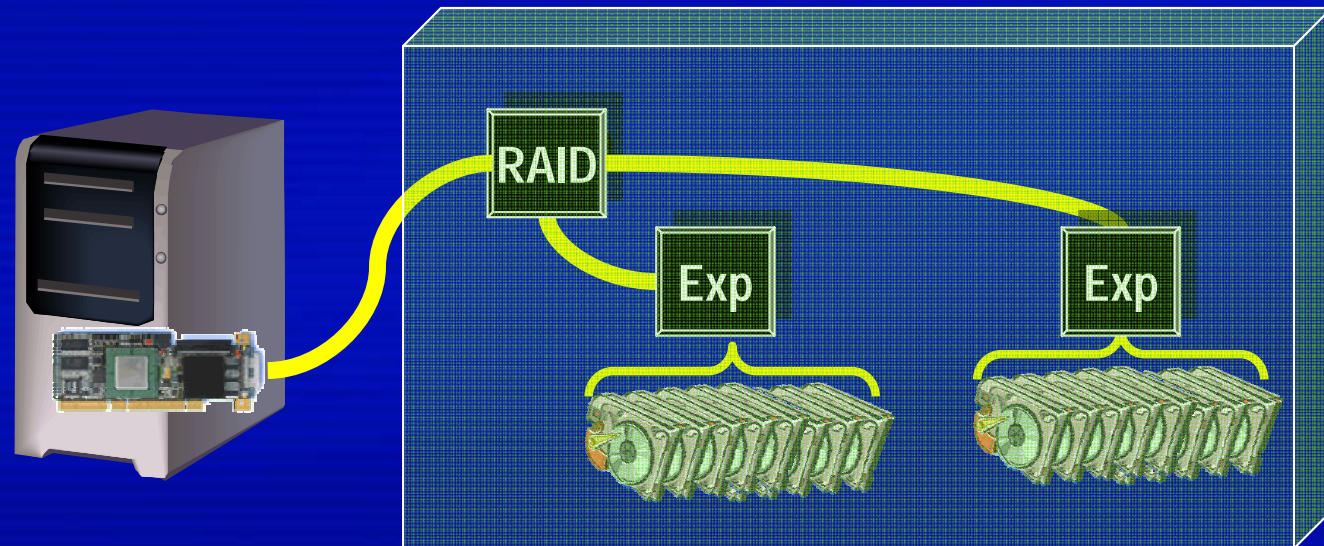
- Not enough disks to saturate 3Gb x 8 SAS
- No benefit from 6Gbps SAS
- Will HBAs support Muxing anyway?

External JBOD



- **With MUX approach**
 - Expanders and HBA need to implement
- **With Buffer Approach**
 - Only top-level expander impacted

External RBOD

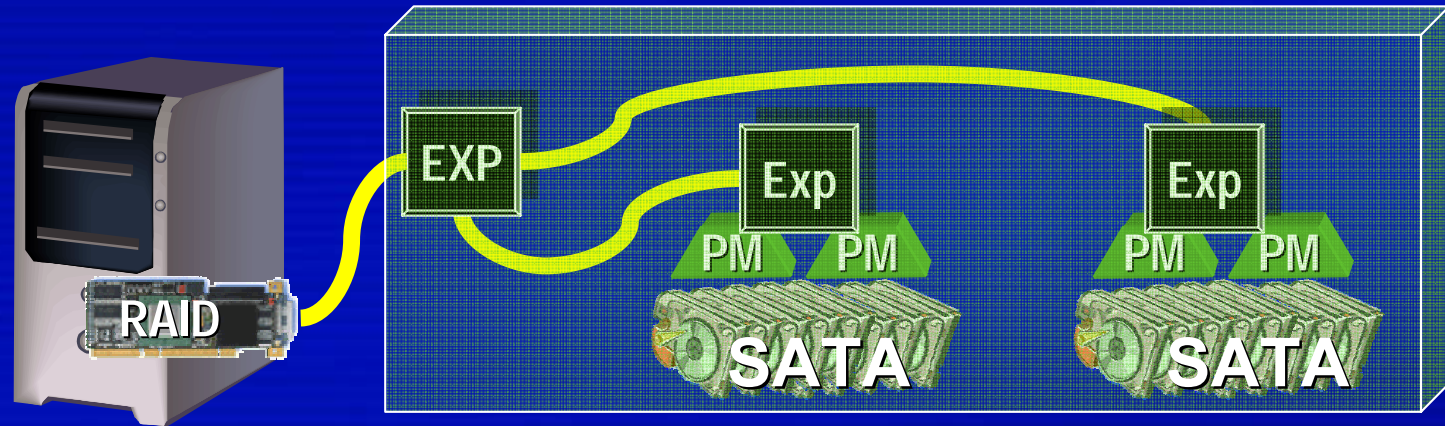


- **RAID ↔ HBA is 6G at both ends**
 - No benefit from multiplexing
- **Disk ↔ RAID I/F *could* benefit from Mux**

Where is complexity applied?

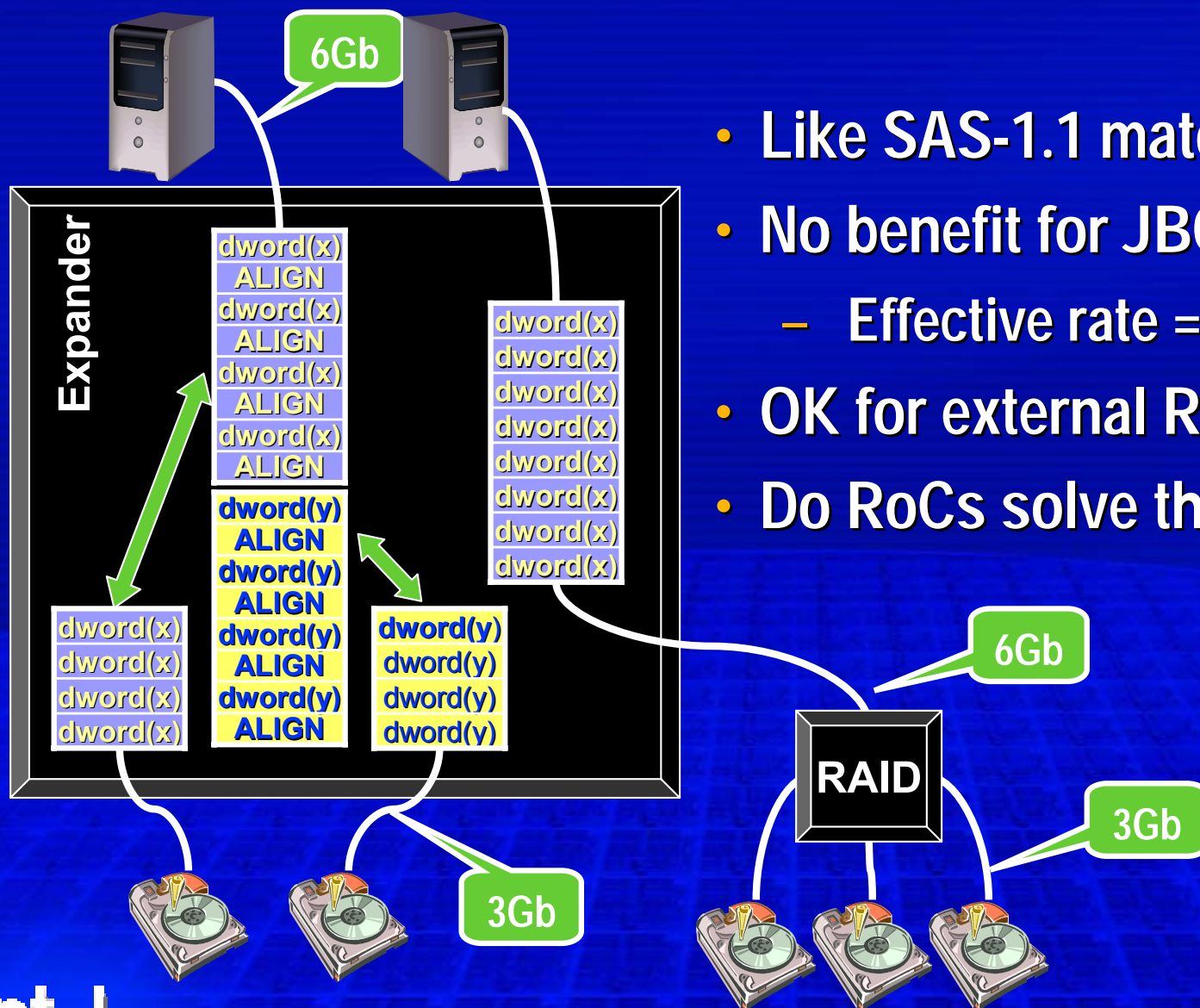
Usage		Mux	Bfr	PM
Internal DAS	Helps?	No	No	No
	Impacts?	I+E	No	No
External RBOD	Helps?	No	No	Yes
	Impacts?	I+E	No	E/M
External JBOD or Ext RAID Disk I/F	Helps?	Yes	Yes	Yes
	Impacts?	I+E	E	E/M
Initiator-to-Initiator	Helps?	No	No	No
	Impacts?	I+E	No	No

SATA Considerations



- **SATA link utilization is low @ 1.5 Gbps**
 - Maybe 50% best case?
 - 4x mux yields < 50% utilization of 6Gbps vs 12%
- **Today Port Multipliers aggregate BW**
 - FIS switching: 1.5 Gbps \leftrightarrow 3Gbps
 - Expanders work with Port Multipliers
 - Buffered PM approaches 100% utilization
- **Need PM to work with 6Gbps Expander**
 - Work to define SATA 6G for Port Multiplier; or...
 - Integrate PM function in the STP/SATA bridge

Pad 6Gb link w/ ALIGN



- Like SAS-1.1 matches 1.5G \leftrightarrow 3G
- No benefit for JBODs
 - Effective rate = 3Gbps anyway
- OK for external RAID
- Do RoCs solve this problem?

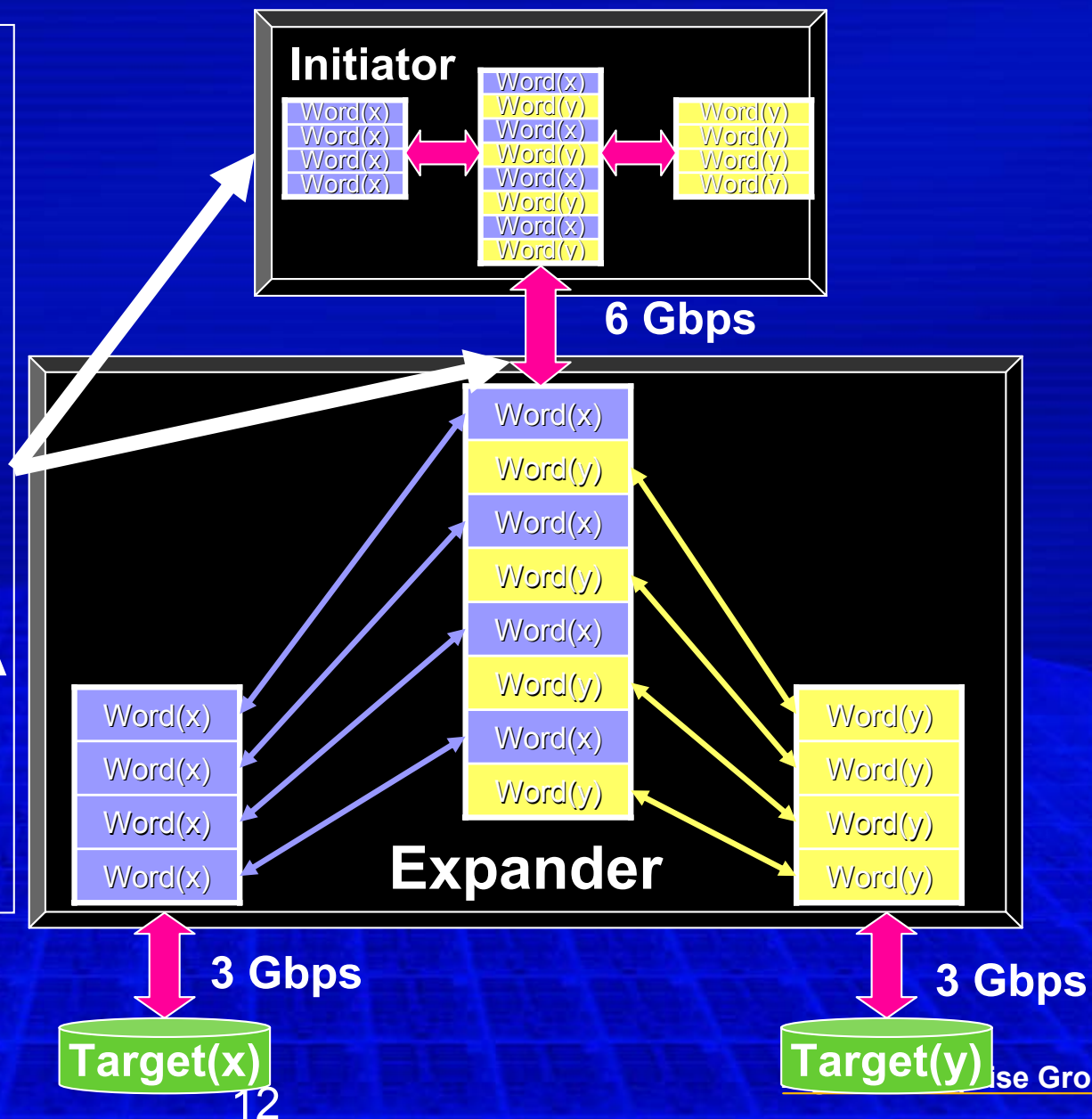
SAS Connection Multiplexing

T10 proposal 05-381

Initiator & Expander coordinate dword interleave for two simultaneous I_T connections.

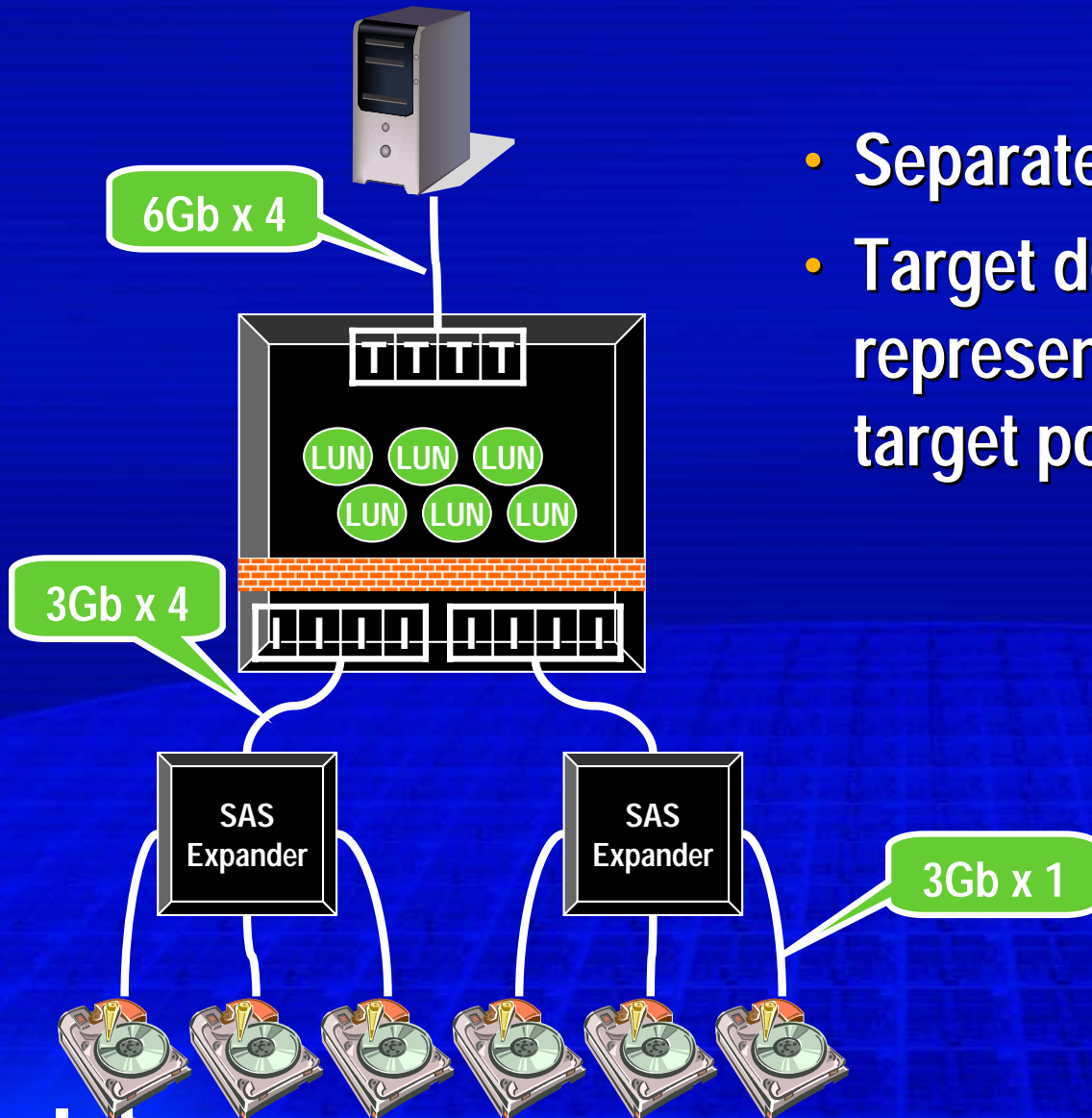
Heavy burden for **Initiators & Expanders**.
Complicates protocol.
Heaviest burden on HBA
Incurs cost on every HBA, needed or not.

May delay TTM.

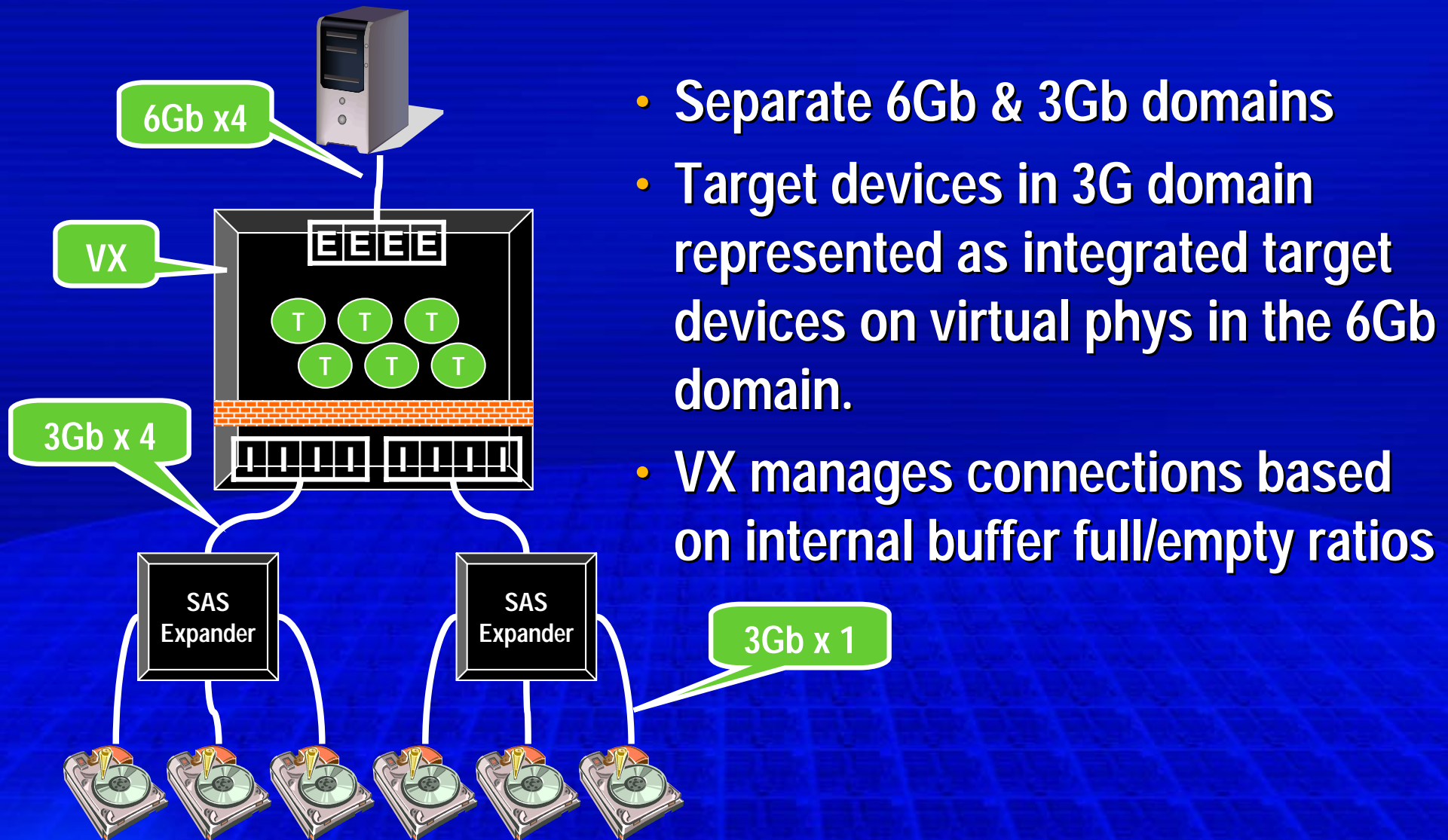


External RBOD Model

- Separate 6Gb & 3Gb domains
- Target devices in 3G domain represented by LUNs behind 6Gb target ports in the 6Gb domain.

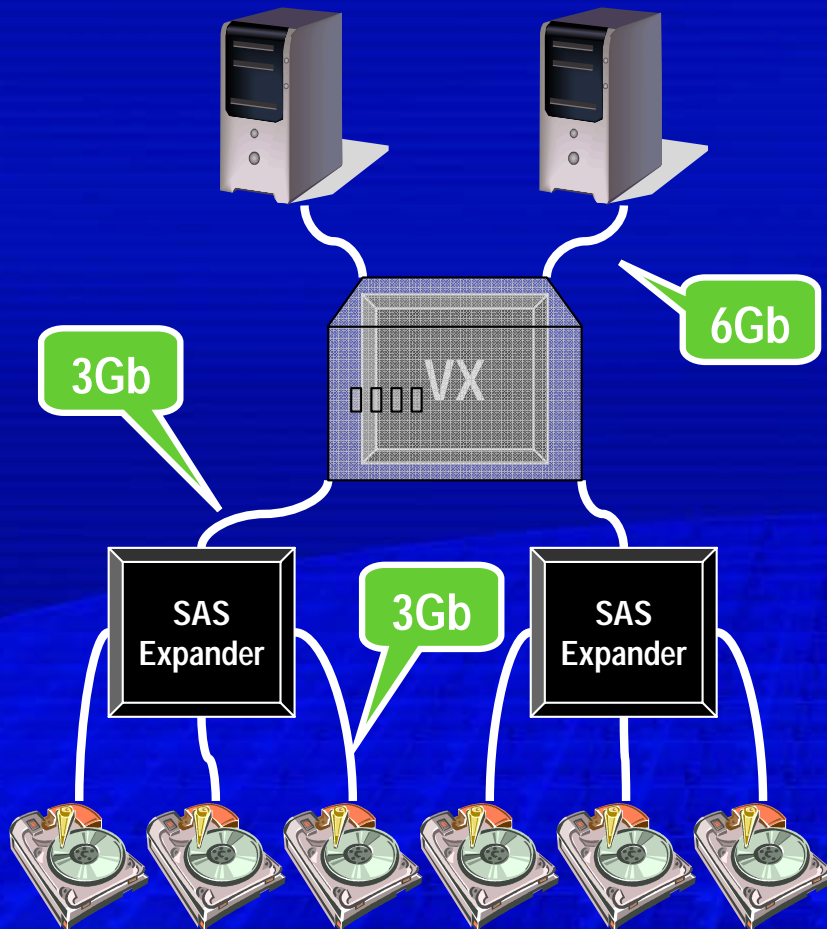


Virtualizing Expander Model



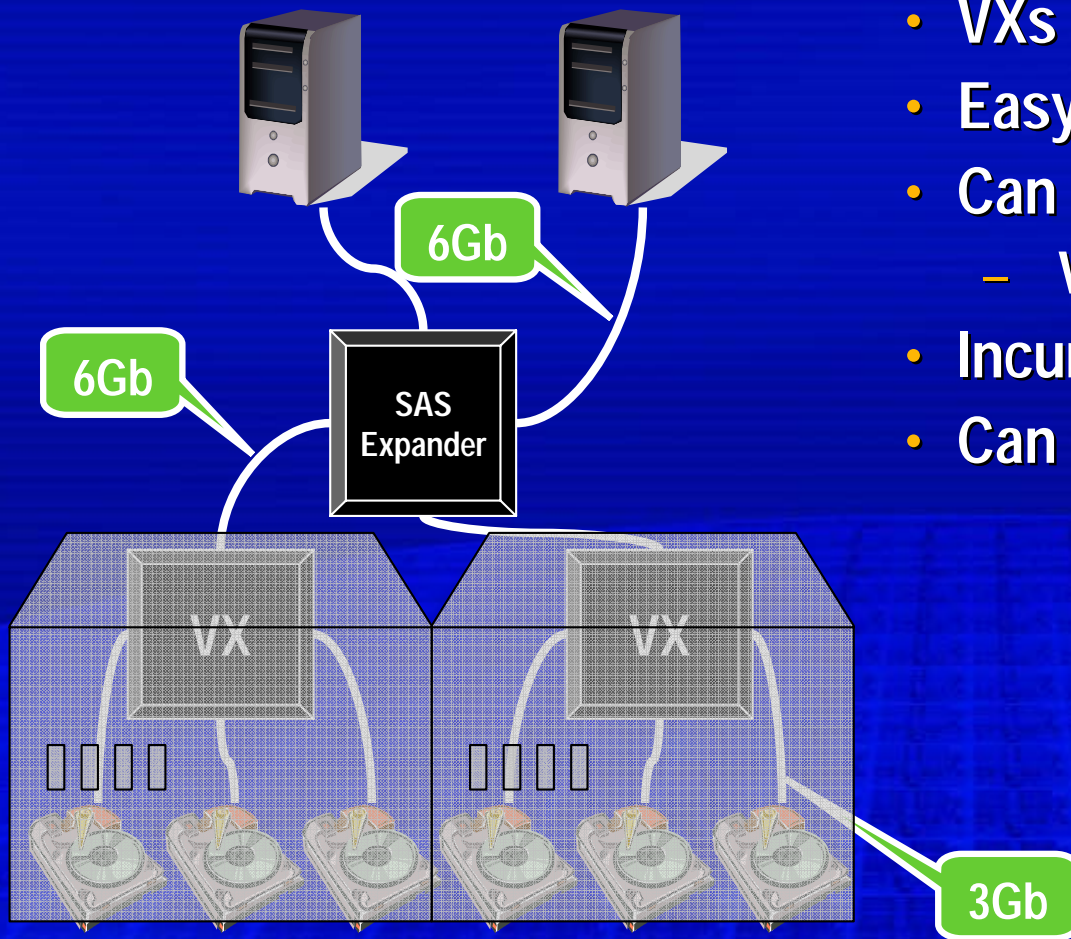
- Separate 6Gb & 3Gb domains
- Target devices in 3G domain represented as integrated target devices on virtual phys in the 6Gb domain.
- VX manages connections based on internal buffer full/empty ratios

Virtualizing Expander @ Host Port



- Use 3Gb Expanders at Drives
- Consolidate cost at common point
- Central point for management
- Max reuse of deployed 3G infrastructure
- Incur cost only where needed
- Can co-exist with BW Mux solutions

Virtualizing Expanders w/ Drives



- VXs appear as multi-LUN targets
- Easy to put on disk backplane
- Can be a modular upgrade
 - VX replaces 3Gb Exp module
- Incur cost only where needed
- Can co-exist with BW Mux solutions

Rollup

- **Connection Mux Aggregates BW, but...**
 - Helps only in External JBOD
 - Impacts Initiators whether used or not
 - Less effective in offsetting SATA utilization
 - Limited to 2.7 GBytes/Sec by PCI-e
 - May complicate achieving interoperability
- **Buffered Expander requires memory**
 - Only used where needed in domain
 - Minimum burden to SAS initiators
 - Minimizes interoperability problems (VX model)
- **Port Multiplier option**
 - PM aggregation already defined in SATA
 - Best option for maximum SATA link utilization
 - Doesn't change SAS link-layer protocol

Conclusions

- **SAS Multiplexing is just one alternative**
 - Has benefits as well as challenges
- **Buffered Expanders and SATA PM are reasonable alternatives**
 - With benefits as compared to Mux
- **SAS-2 should consider all alternatives**
- **All 3 solutions could be included**
 - And still be interoperable