



STORAGE DEVELOPER CONFERENCE

SNIA ■ SANTA CLARA, 2014

Next Generation iSCSI Enterprise Grade Data Integrity and Performance

**Wael Nouredine
Chelsio Communications**

Outline

- ❑ iSCSI Overview
- ❑ iSCSI HBA Update
- ❑ Benchmarks and roadmap
 - ❑ Performance
 - ❑ Virtualization
- ❑ Data integrity protection

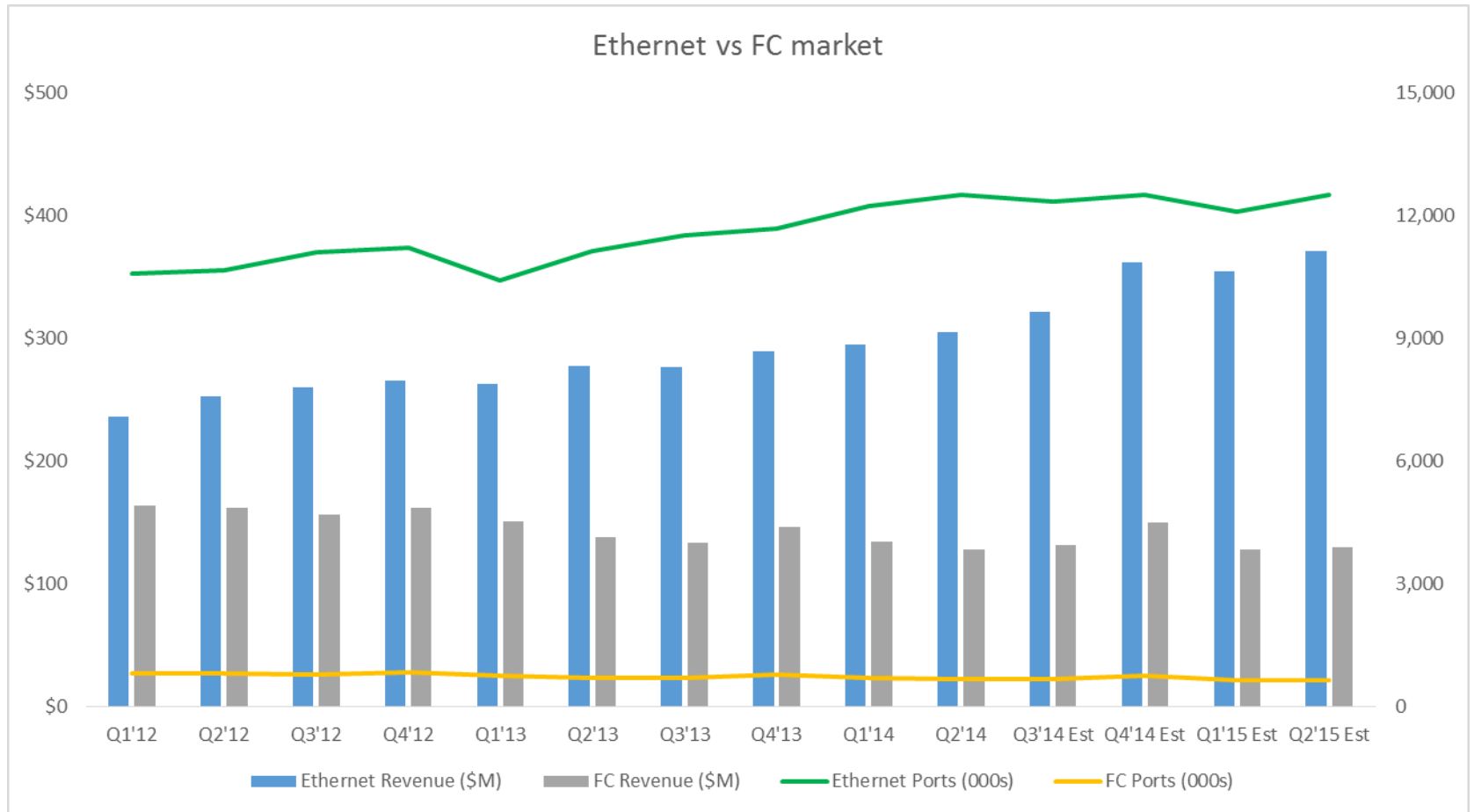
iSCSI Timeline

- ❑ RFC 3720 in 2004
 - ❑ Latest RFC 7143 in April 2014
- ❑ Designed for Ethernet-based storage area networks
 - ❑ Data protection
 - ❑ Performance
 - ❑ Latency
 - ❑ Flow control
- ❑ Leading Ethernet based SAN technology
 - ❑ In-boxed initiators
 - ❑ Plug-and-play
- ❑ Closely tracks Ethernet speeds
 - ❑ Increasingly high bandwidth
- ❑ 10GbE, IEEE 802.3ae 2002
 - ❑ First 10Gbps hardware iSCSI in 2004 (Chelsio)
- ❑ 40/100GbE, IEEE 802.3ba 2010
 - ❑ First 40Gbps hardware iSCSI in 2013 (Chelsio)
 - ❑ First 100Gbps hardware iSCSI expected in 2016
- ❑ 400GbE, IEEE P802.3bs
 - ❑ Task Force formed March 2014

iSCSI Trends

- ❑ iSCSI growth
 - ❑ FC in secular decline
 - ❑ FCoE struggles with limitations
- ❑ Ethernet flexibility
 - ❑ iSCSI for both front and back end networks
- ❑ Convergence
 - ❑ Block-level and file-level access in one device using a single Ethernet controller
 - ❑ Converged adapters with RDMA over Ethernet and iSCSI consolidate front and back end storage fabrics
- ❑ Hardware offloaded 40Gbps iSCSI aligns with migration from spindles to NVRAM
 - ❑ Unlocks potential of new low latency, high speed SSDs
- ❑ Virtualization
 - ❑ Native iSCSI initiator support in all major OS/hypervisors
 - ❑ Simplifies storage virtualization

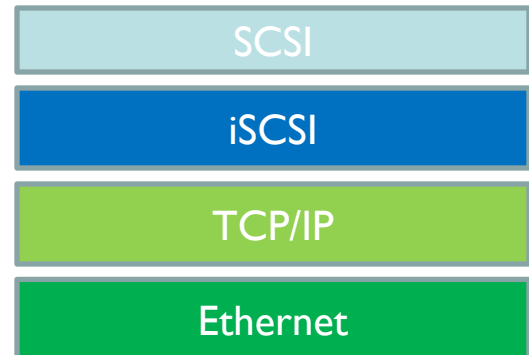
iSCSI Trends



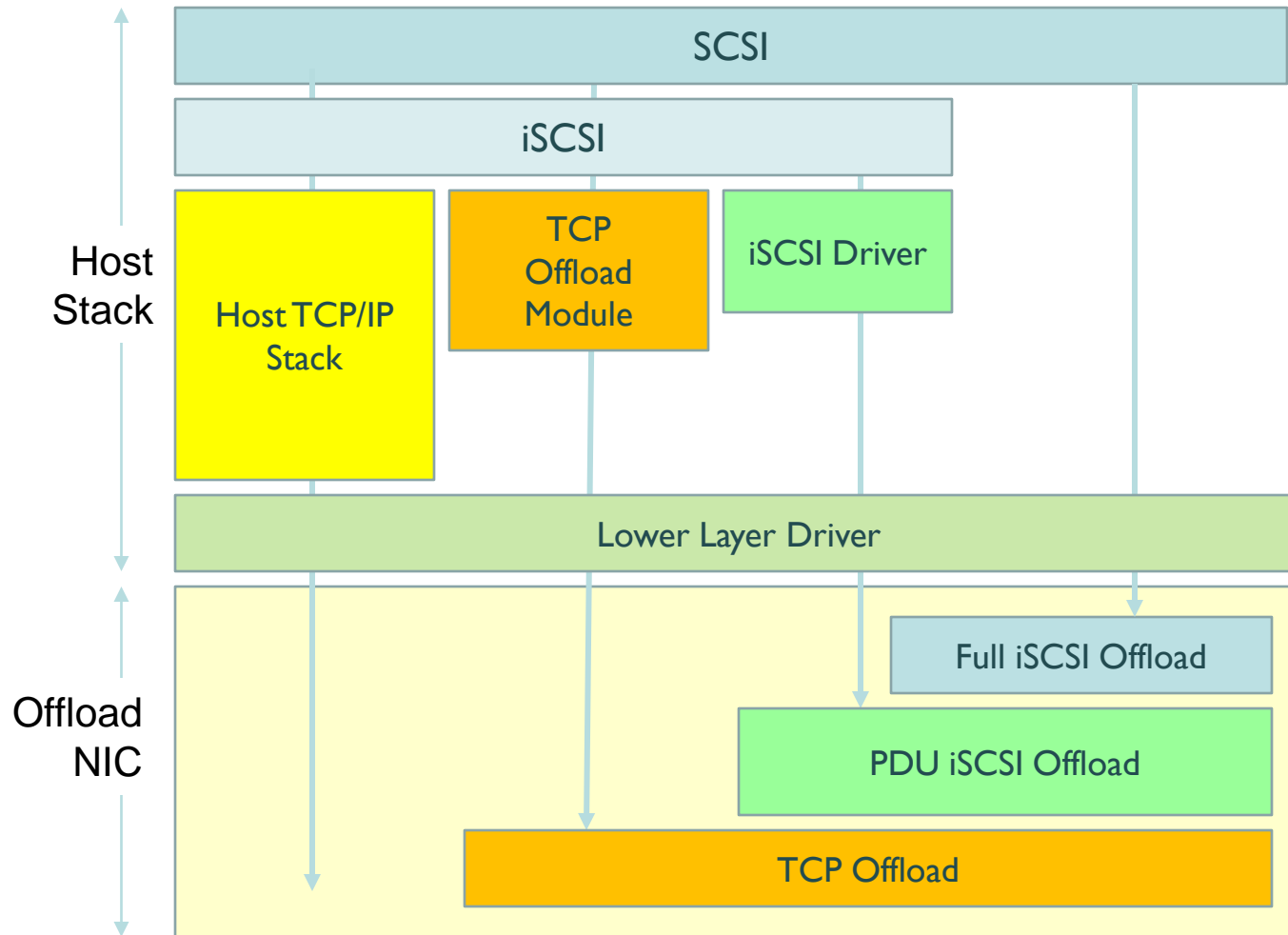
Source: Crehan Research - 2Q14 CREHAN Quarterly Market Share Tables

iSCSI Overview

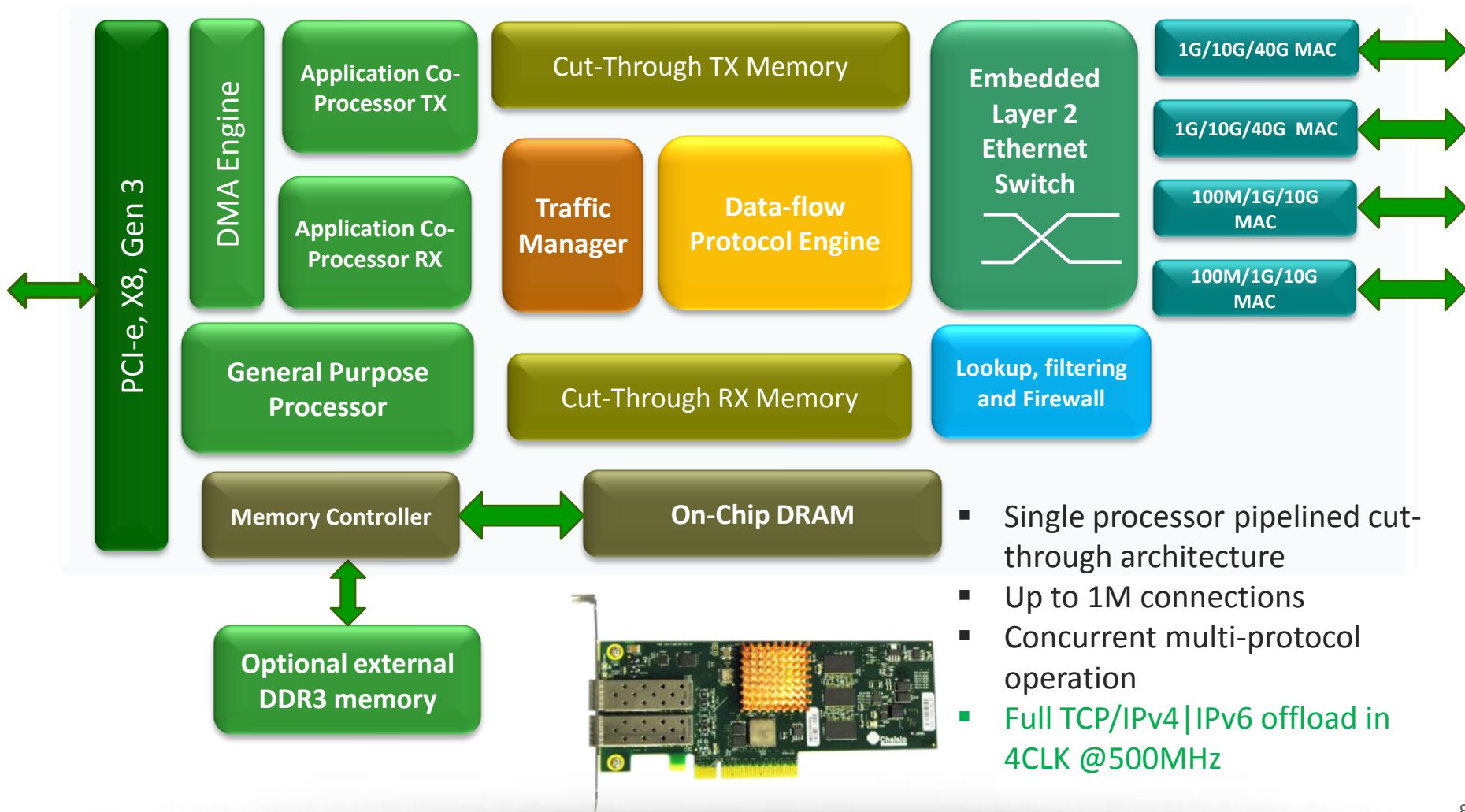
- ❑ High performance
 - ❑ Zero copy DMA on both ends
 - ❑ Hardware TCP/IP offload
 - ❑ Hardware iSCSI processing
- ❑ Data protection
 - ❑ CRC-32 for header
 - ❑ CRC-32 for payload
 - ❑ No overhead with hardware offload
- ❑ Scalable TCP/IP foundation
 - ❑ IP routability to datacenter, WAN and Cloud scales
 - ❑ Reliability/robustness even over wireless links
 - ❑ Congestion and flow control
 - Leverages all infrastructure



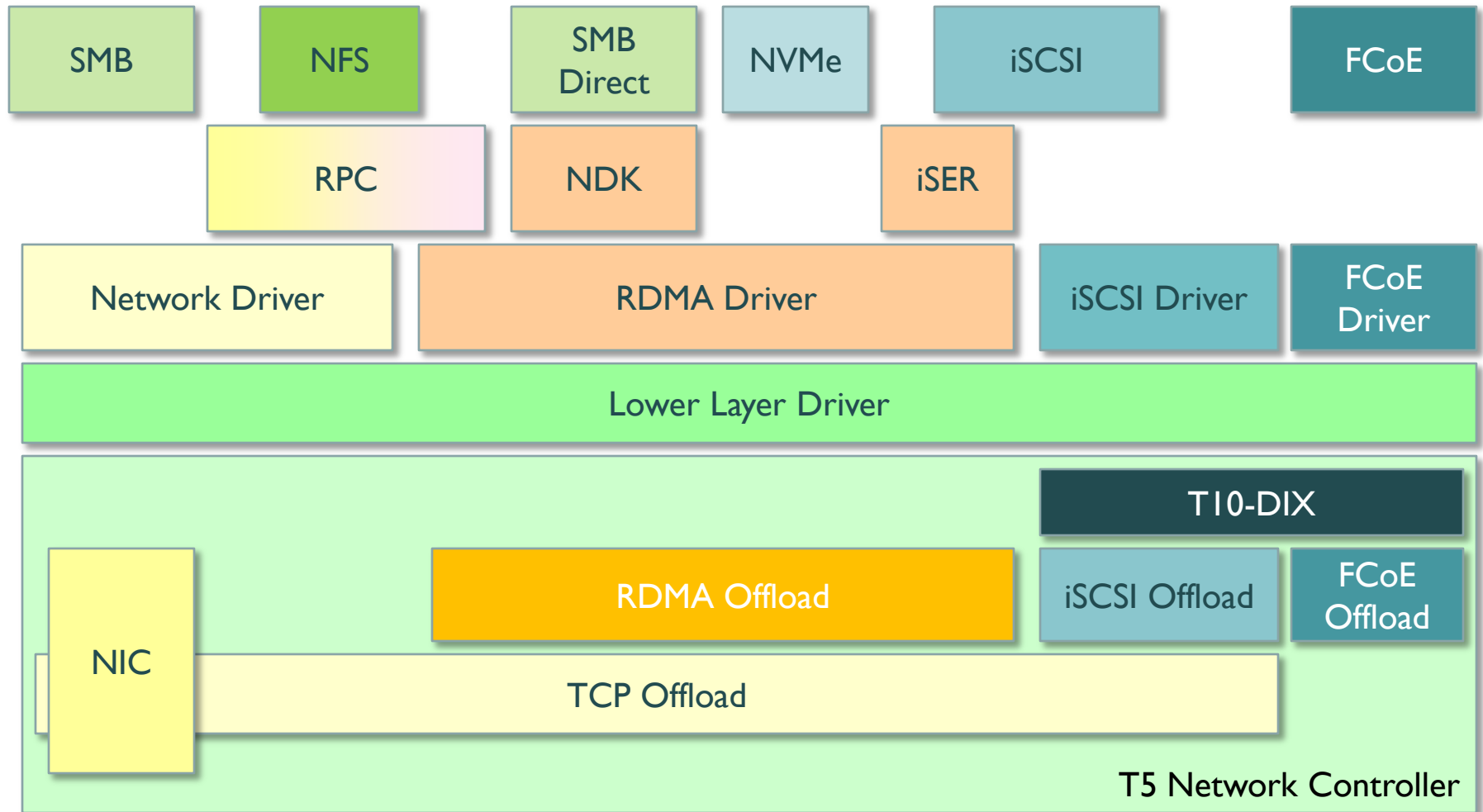
iSCSI Layering



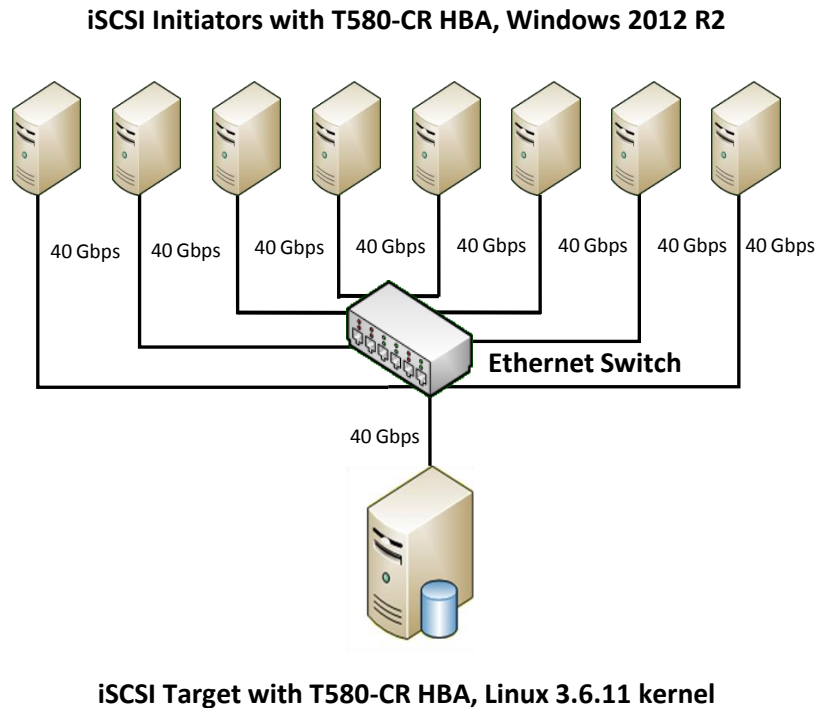
Chelsio T5 Ethernet Controller ASIC



T5 Storage Protocol Support

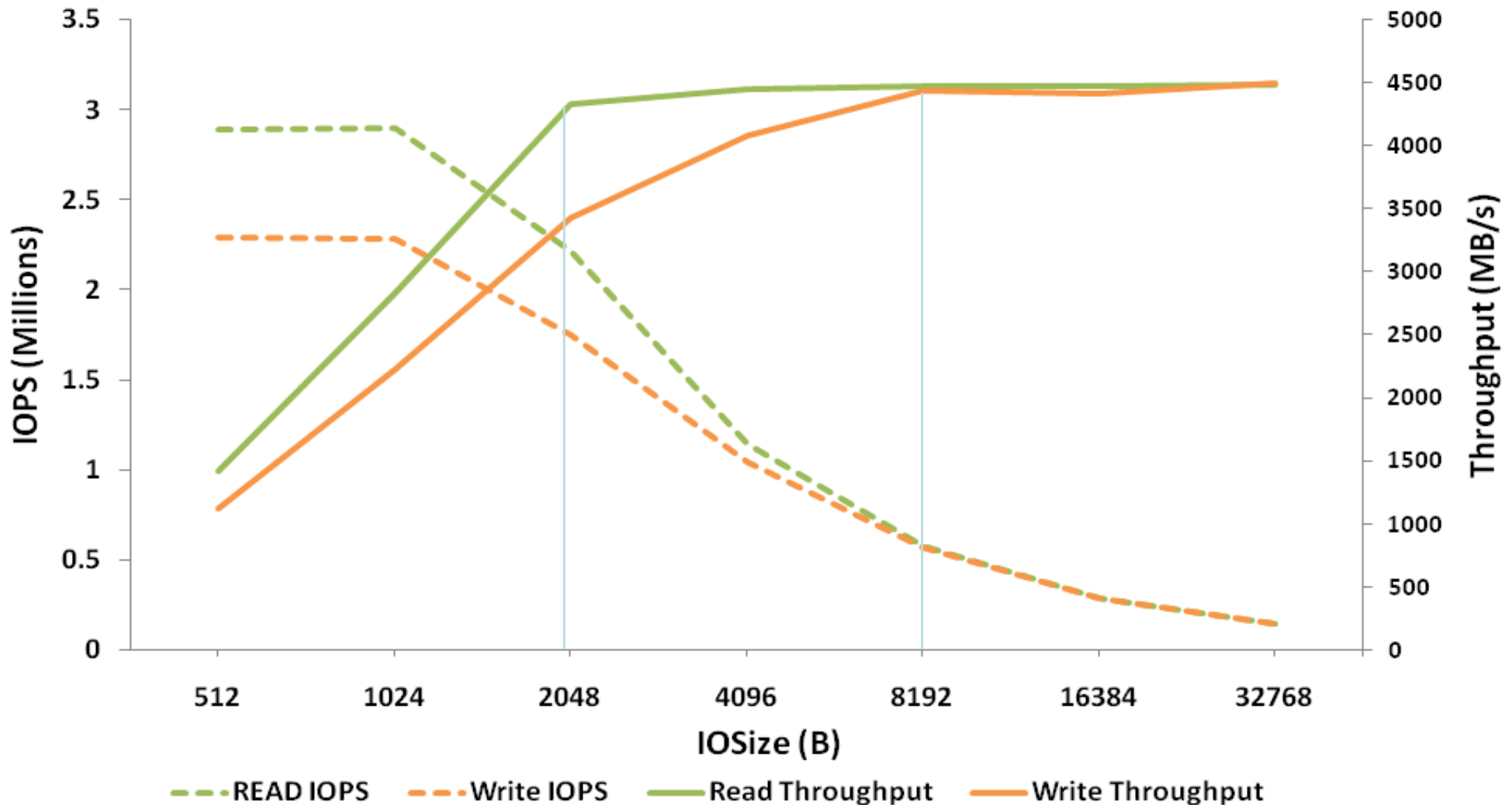


iSCSI Performance at 40Gbps

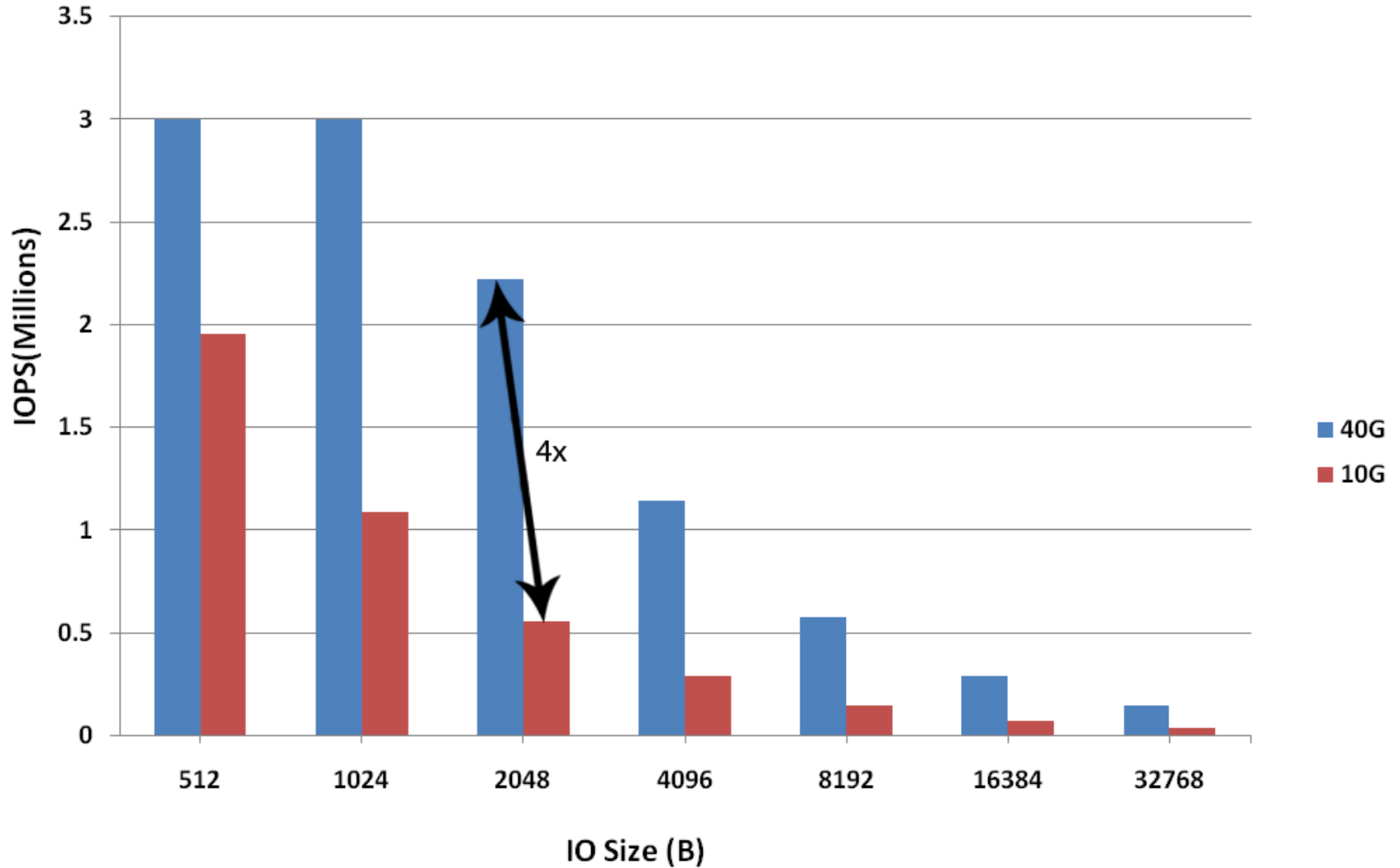


- ❑ Storage array with 64 targets connected to 8 initiator machines through 40Gbps switch
 - ❑ Targets are *ramdisk null-rw*
 - ❑ Each initiator connects to 8 targets
- ❑ Iometer configuration on initiators
 - ❑ Random access pattern
 - ❑ 50 outstanding IO per target
 - ❑ 8 worker threads, one per target
 - ❑ IO size ranges from 512B to 32KB

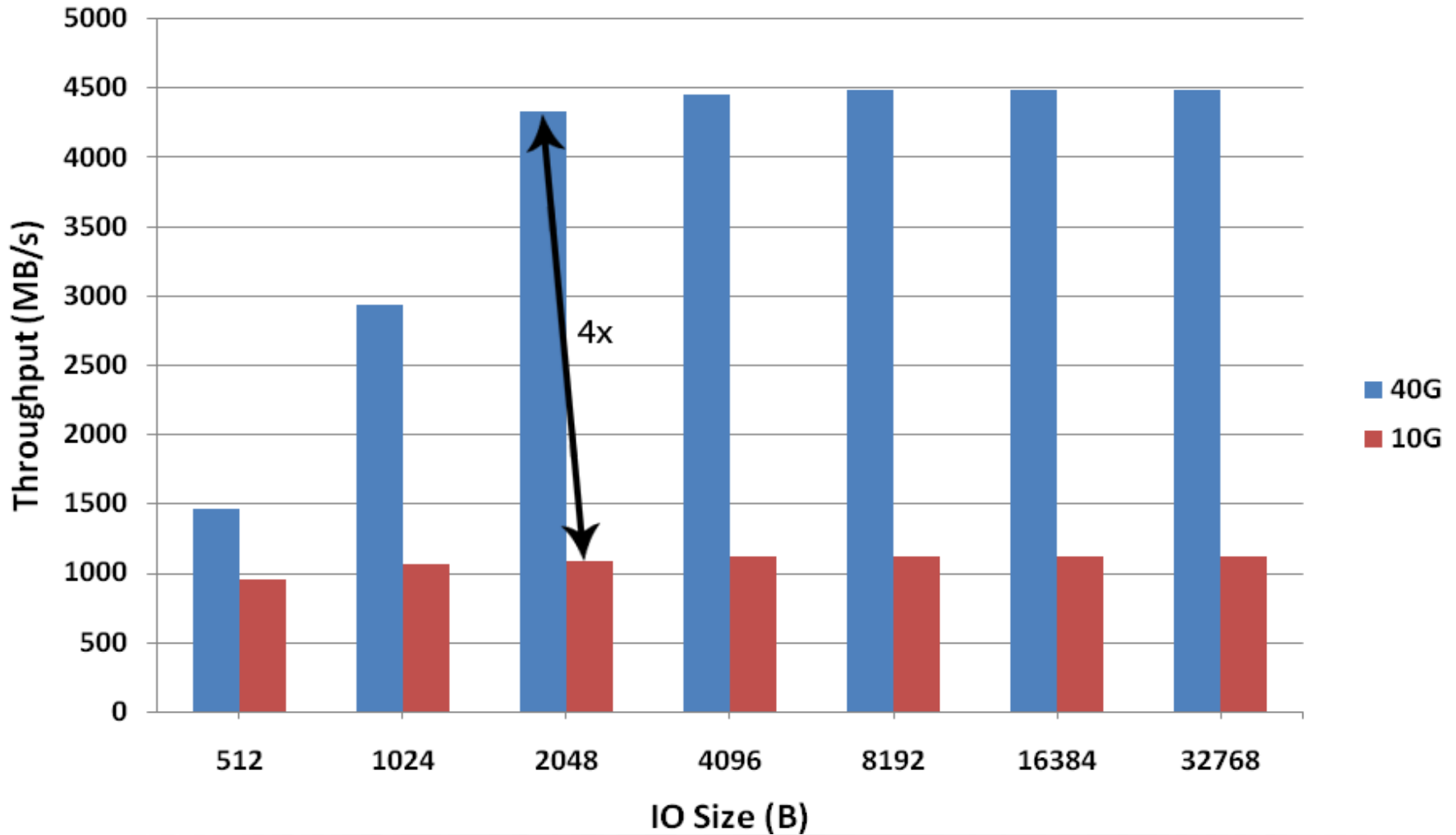
iSCSI Performance at 40Gbps



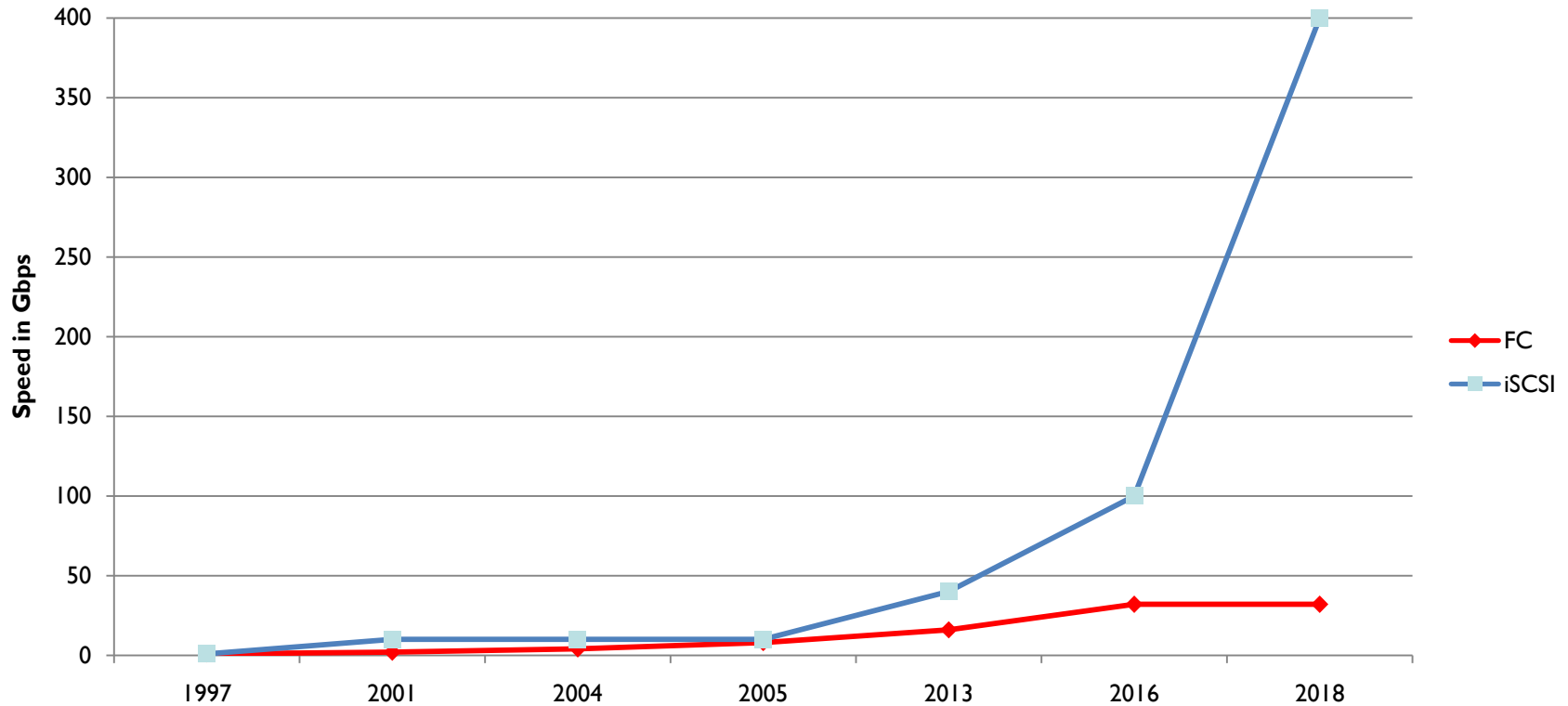
iSCSI READ IOPS – 10Gbps vs. 40Gbps



iSCSI READ BW – 10Gbps vs. 40Gbps

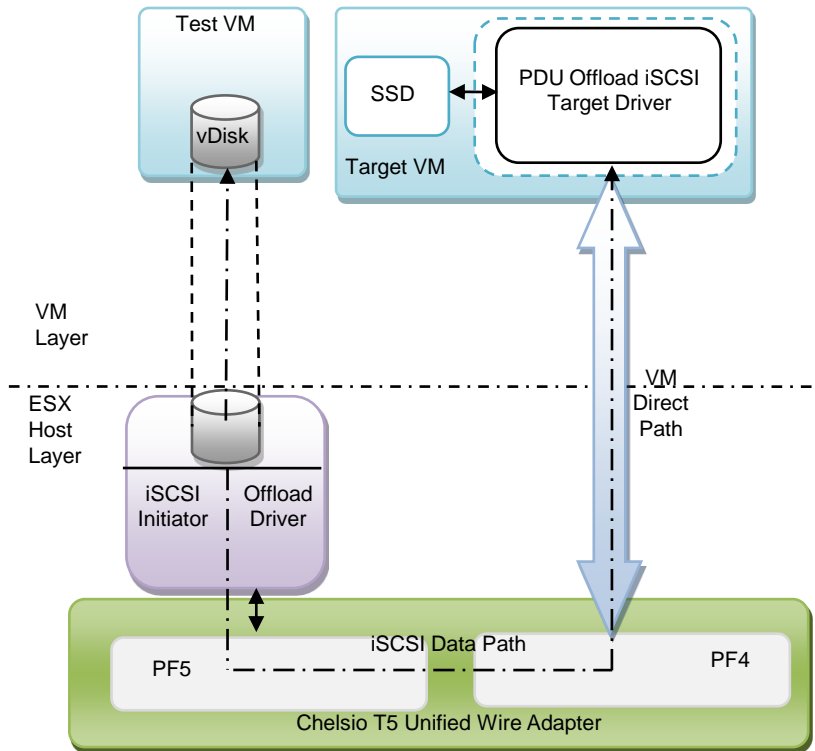


iSCSI Bandwidth Roadmap



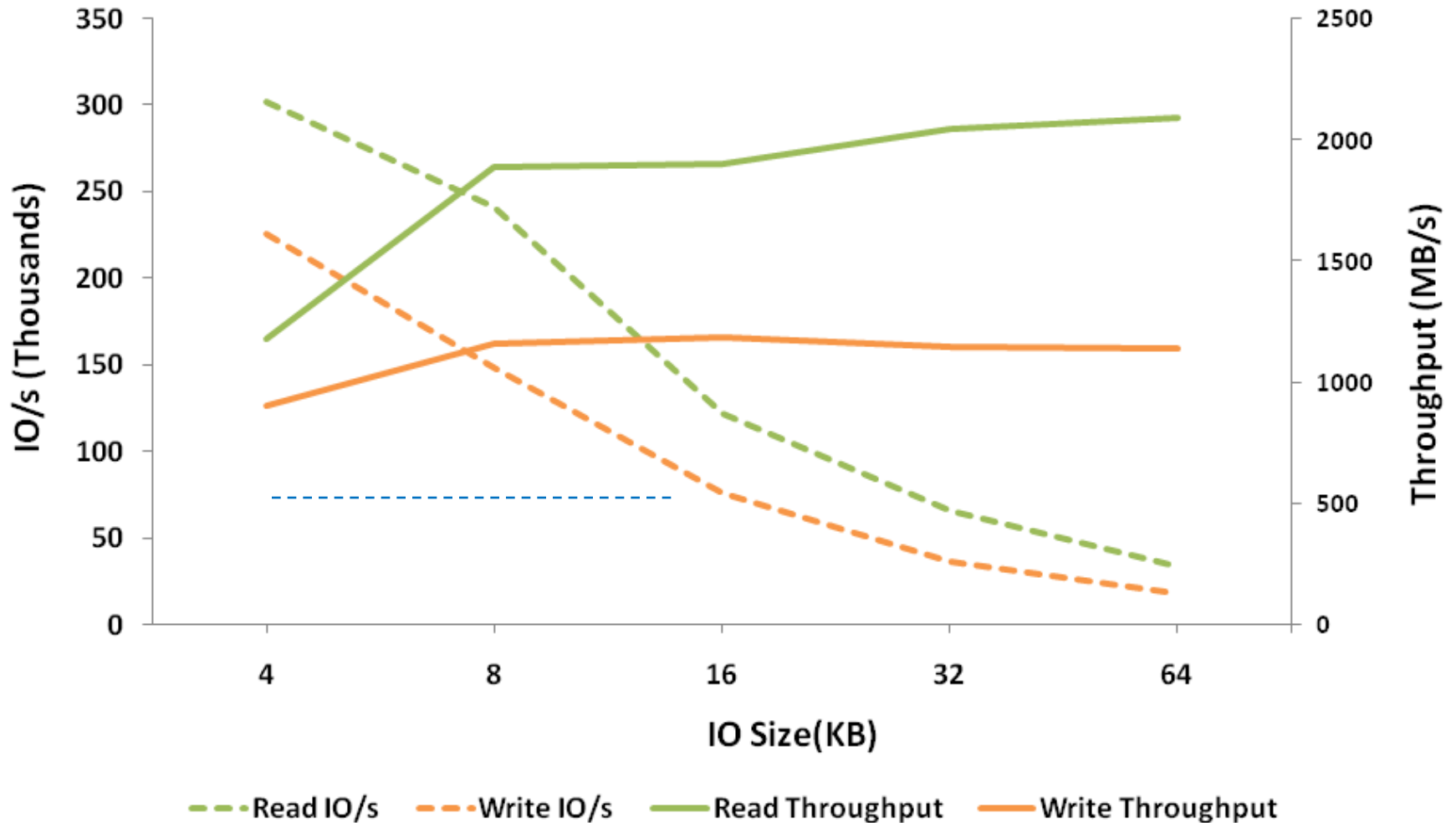
As of 2013, T5 offload engine iSCSI PDU processing capacity sufficient for standard frames at 400Gbps rate.

Virtualized iSCSI



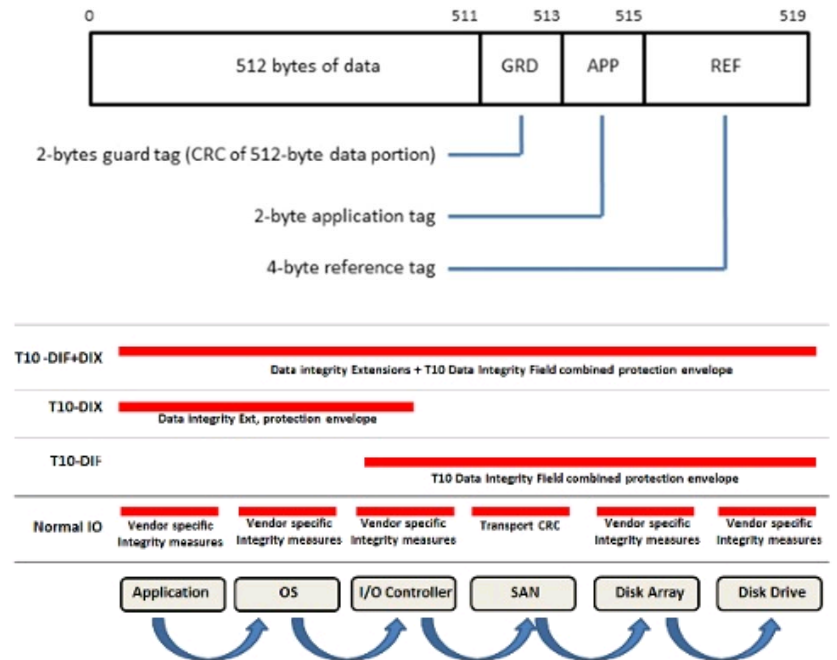
- ❑ Initiator VM and target VM running on the same system
- ❑ Communication through T5 on-chip embedded switch
- ❑ Target VM communicates through VM Direct Path to the T5 adapter
- ❑ Initiator VM runs a paravirtualized driver to utilize the fully offloaded T5 initiator

Virtualized iSCSI IOPS and Throughput



Advanced Data Integrity Protection

- Above and beyond iSCSI CRC-32
- Data Integrity Field (DIF) protects against silent data corruption with 16b CRC
 - Adds 8-bytes of Protection Information (PI) per block
- Data Integrity Extension (DIX) allows this check to be done between application and HBA
 - iSCSI CRC-32 handoff possible
- T10-DIF+DIX provide a full end-to-end data integrity check
- T5 supports hardware offloaded T10-DIF+DIX for iSCSI (and FCoE)



Martin Petersen, Oracle, <https://oss.oracle.com/~mkp/docs/dix.pdf>

iSCSI Summary

- ❑ Mature protocol with wide industry support
- ❑ Native initiator in-boxed in all major operating systems/hypervisors
 - ❑ Back-end and front-end applicability, virtualization
- ❑ Hardware offloaded iSCSI shipping at 40Gbps
 - ❑ High IOPS and throughput
 - ❑ Low latency
- ❑ Robust TCP/IP foundation allows operation over Wireless, LAN and WAN networks
 - ❑ Hardware offload eliminates overhead
 - ❑ No specialized cables, equipment, switches, or forwarders
 - ❑ True network convergence
- ❑ Roadmap to 100Gbps, 400Gbps and beyond
- ❑ Hardware based end-to-end data integrity protection



STORAGE DEVELOPER CONFERENCE

SNIA ■ SANTA CLARA, 2014

Thank You

Ask about Chelsio's 40Gbps iSCSI
evaluation program at: sales@chelsio.com

Visit www.chelsio.com for more info