

An abstract graphic consisting of several interlocking rings of varying colors (silver, orange, blue) and sizes, arranged in a circular pattern. The rings have a metallic, reflective appearance. The largest ring is silver and orange, positioned in the upper left. Other rings in orange, silver, and blue are scattered in the lower right and bottom center.

Brookhaven Laboratories Technology Overview:

Neterion's X3 series
10GbE Adapters

VIRTUAL I/O
SOLUTIONS

AGENDA

- Introduction
- X3100 series hardware
- Features & functions
- Using with virtualized servers
- Traditional server use case
- Future





VIRTUAL I/O SOLUTIONS

Focus exclusively on innovative technology

- Emerging I/O virtualization technologies
- High function, commodity-based clustered storage

Portfolio includes products from companies such as:

- Aprius
- Neterion
- Scale Computing
- Voltaire
- Xsigo

Design & implementation services

Offices in Alpharetta, GA and Tucson, AZ



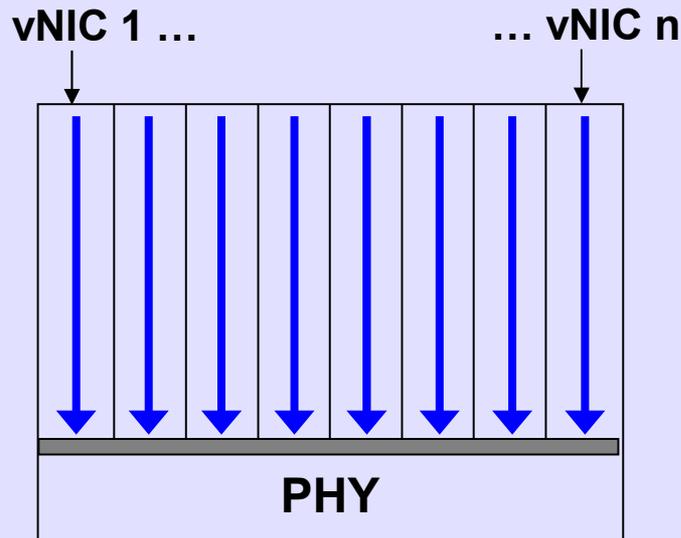
About Glenn

Glenn Hechler currently meets with customers throughout the world to discuss IT optimization, strategy, and solutions, with a particular focus on I/O and storage virtualization. He started using and building computers back in 1978. His career took him from being an engineer creating robotics, databases and other software, to systems integration, sales, and finally shifting to a consulting role. Before joining Virtual I/O Solutions, he worked with storage virtualization solutions at IBM, including Scale-Out File Services, SAN Volume Controller, SAN File System, and Virtual Tape Systems.

VIRTUAL I/O
SOLUTIONS

NETERION X3100

Neterion's Independent Hardware IOV



- Neterion is the 1st fully virtualized 10GbE adapter
 - Provides multiple queues
 - Each queue associated with MAC addresses
 - MAC addresses can be changed dynamically without resetting the entire physical NIC
 - Each queue can be quiesced independently

Layer 2 switch in the silicon!



HARDWARE T10GBITS

- PCIe x8 adapter
 - Up-pluggable in x16 slot
 - Will negotiate x8 slot wired x4
- 32MB memory
- 6 models
 - Single-port and Dual-port adapters
 - Cable options
 - Copper – 5m (7m certified)
 - 62.5 μ m, 50 μ m optical fibre - 300m
 - 10 μ m optical fibre - 10km



PLATFORM SUPPORT

- Linux 2.4 and 2.6 kernels
 - In kernel 2.6.30
 - RHEL 4, 5 – shipping with 5.4 and on (vxge)
- Xen (RH, SLES, Citrix) 3, 4, 5
 - Control via xe/xm and Linux commands
- Also ESX/ESXi, FreeBSD, KVM, Windows
 - Via OEM drivers when available: AIX, HP-UX, Irix, Mac OS X, Solaris

NOTE: s2io only supports Xframe I/II adapters. vxge is required for X3 adapters.



HARDWARE MODES

- Single function mode (up to 17 queues)
- Multi-function mode
- SR-IOV 1.0
- MR-IOV 1.0
- Configurable in the BIOS (will need to reboot to switch between modes)



MULTI-FUNCTION MODE

- Multiple instances of the adapter in OS
 - Similar to SR-IOV without requiring SR-IOV hardware or OS support (of SR-IOV)
- OS limits # of instances (functions)
 - Xen 5.4 allows for up to 8
 - Linux allows for up to 17 (adapter maximum)
 - 2 functions, 8 vpaths each
 - 4 functions, 4 vpaths each
 - 8 functions, 2 vpaths each



- Single Root I/O Virtualization
- Requires hardware that supports SR-IOV
- Intended usage: hypervisors
 - Enables direct access of the device by the guest, reducing virtualization overhead
- Modes supported:
 - 17 VF, 1 vpath each
 - 1 PF, 7 VF, 2 vpaths each
 - 1 PF, 3 VF, 4 vpaths each

PF – Physical Function
VF – Virtual Function



- Multi-Root I/O Virtualization
 - Requires hardware that supports MR-IOV
- Intended usage: Blades
 - I/O module plugs directly into PCIe bus
 - Eliminates mezzanine card in each blade
- Modes supported:
 - 17 virtual hierarchies, 1 vpath/function/hierarchy
 - 8 virtual hierarchies, 2 vpaths/function/hierarchy
 - 4 virtual hierarchies, 4 vpaths/function/hierarchy



NIFTY FEATURES

- Multiple hardware queues
- IOQoS™
 - Each queue can have a throughput limit (or be left unlimited)
 - Prioritize iSCSI workloads
 - Guarantee bandwidth for services
- iSCSI
- FCoE



MORE NIFTY FEATURES

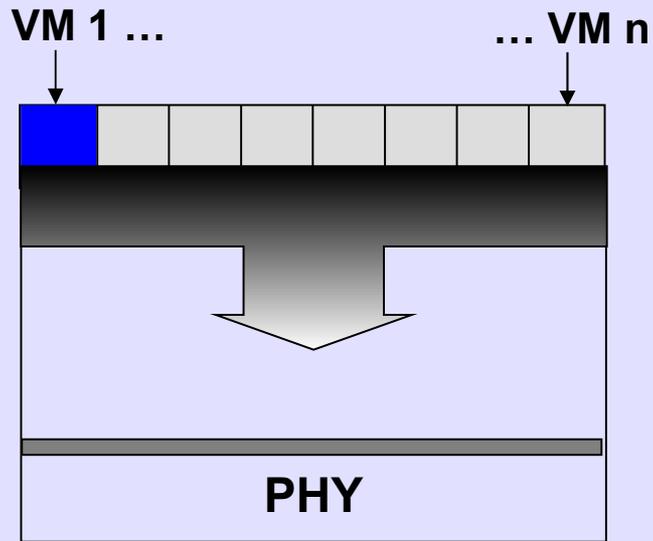
- Jumbo frames
 - Up to 9600B, configurable via ipconfig
- Linux offloading support
 - Checksum offload (TCP/UDP/IP) on transmit and receive paths
 - TSO on transmit path
 - GRO on receive path

TSO – TCP Segmentation Offload
GRO – Generic Receive Offload

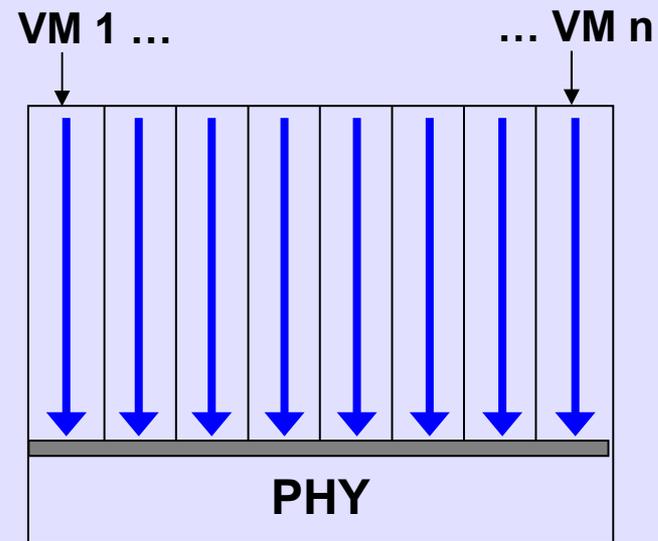


IOV ARCHITECTURE

Other Superficial Front-End IOV



Neterion's Independent Hardware IOV

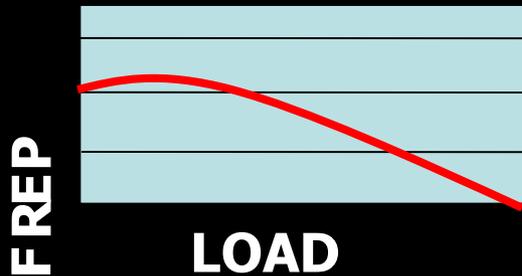
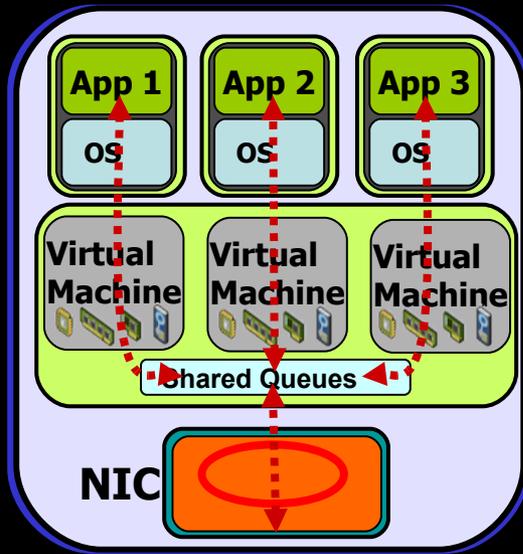


Some architectures are providing features via firmware without sufficient hardware to support the function very well.

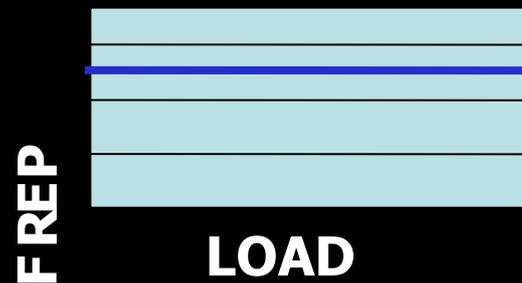
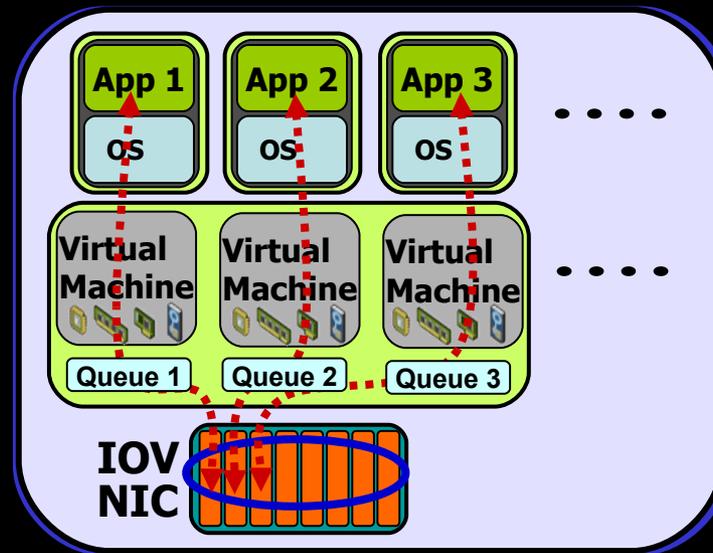


IOV MAINTAINS HIGHER PERFORMANCE

No IOV



With IOV

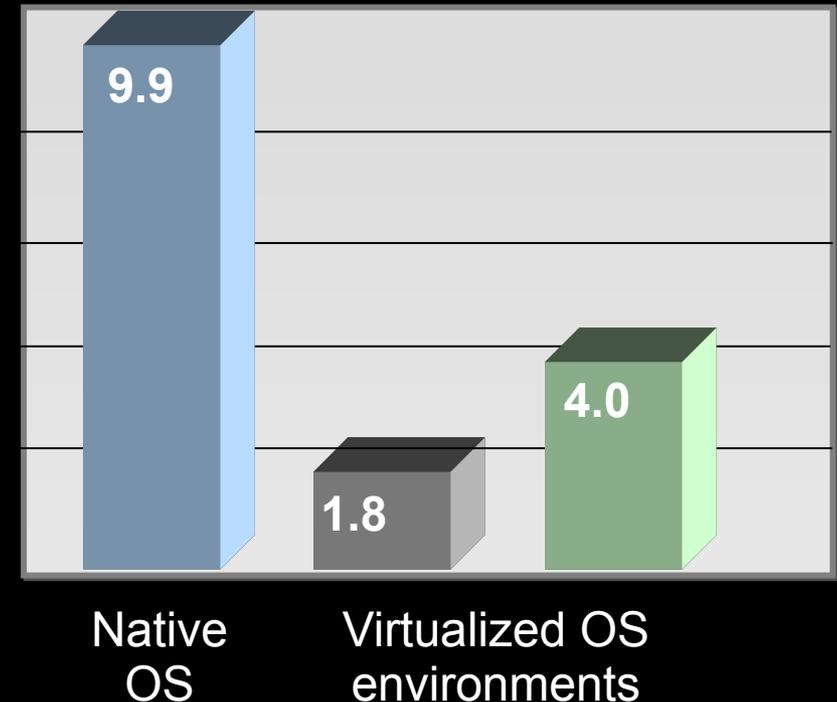


Independent hardware paths for each instance that can be configured and quiesced without impacting the entire adapter



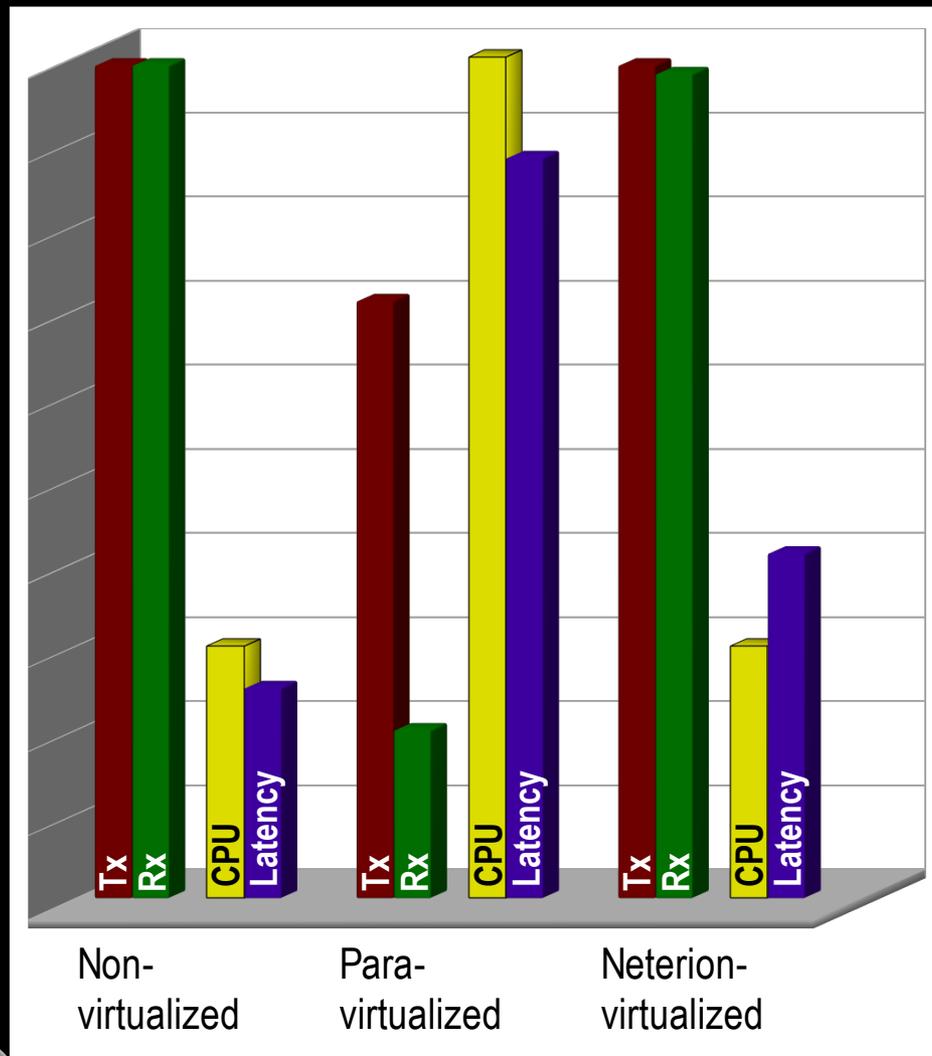
PROBLEM: LOADING 10Ge

- Upwards of 1M+ packets/sec
- Hypervisor bogs down
 - Too much context switching processing inbound packets
 - CPU utilization goes up
 - Latency increases
 - Performance of virtual machines much lower versus native OS



FULLY VIRTUALIZED 10Ge

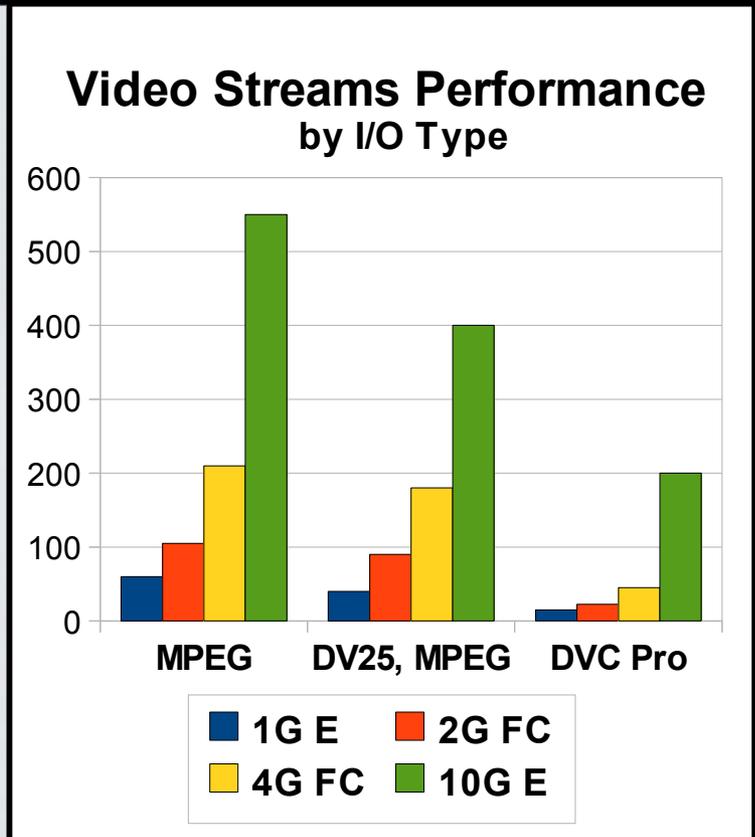
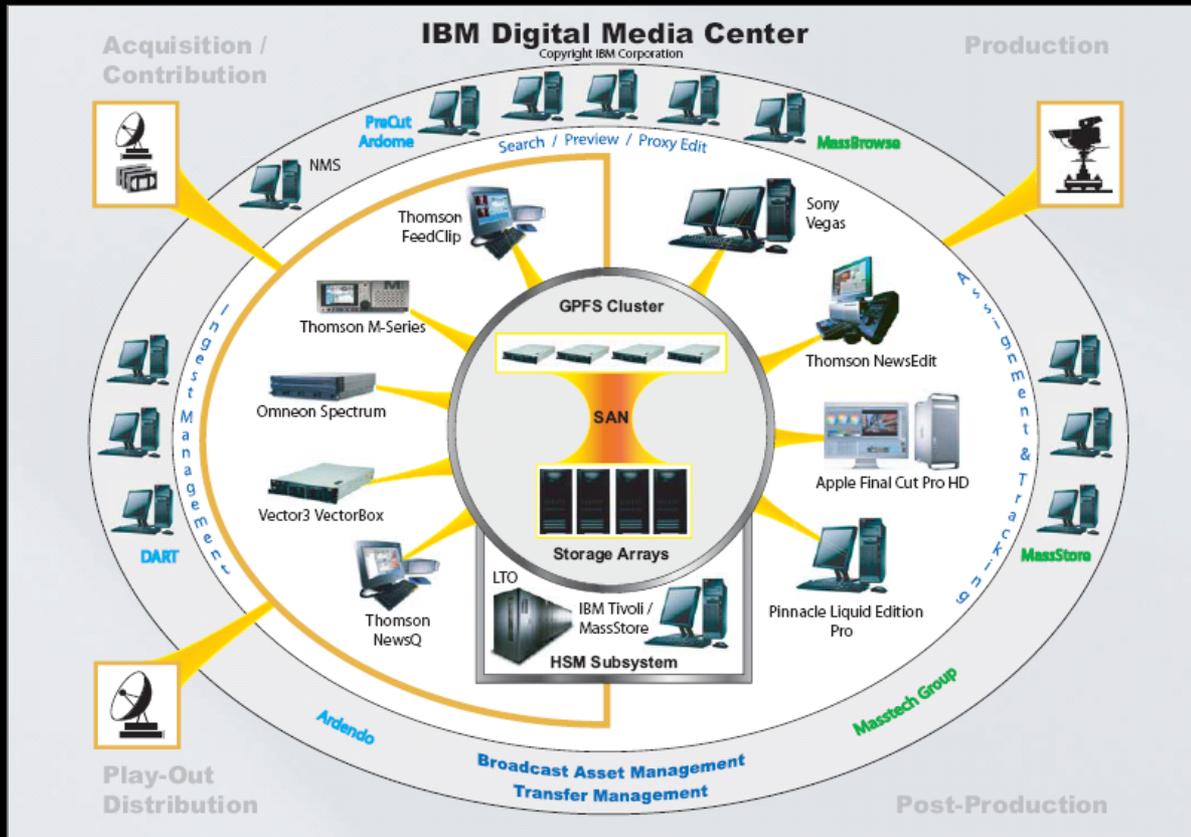
Network Performance



- Direct Assignment – Bypass Hypervisor Overhead
- Offload CPU Intensive Tasks to adapter H/W
- Smash Virtualization’s “Glass Ceiling”
- iSCSI Compatible

Tx,Rx – taller is better
CPU, Latency – smaller is better

USE CASE



State-of-the-art post-production digital solution



10Ge VALUE PROPOSITION

- 1) 10 times the performance of 1Ge
- 2) Simplify management complexity
- 3) Reduced power and cooling
- 4) Enables server and drawer consolidation
- 5) Future-proof the data center – ready for Unified Fabric

“A fat pipe requires no management!”

*Radia Perlman,
commonly referred to as the ‘Mother of the Internet’,
inventor of the Ethernet spanning-tree protocol.*



WHAT'S NEXT?

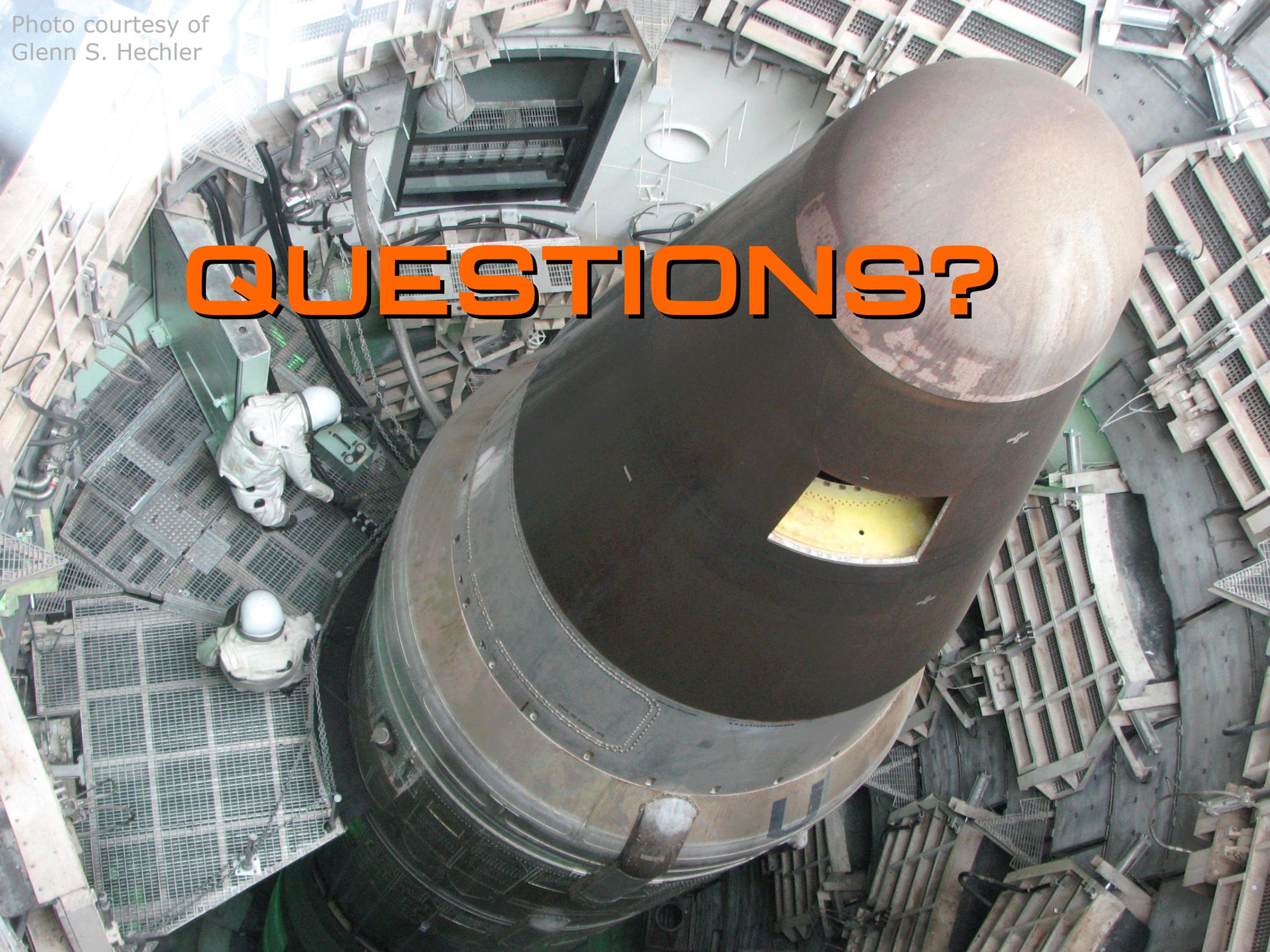
New silicon that will provide:

- **HW accelerations**
 - **iSCSI**
 - **FCoE**
- **More memory**
- **More instances**



Photo courtesy of
Glenn S. Hechler

QUESTIONS?



Trademarks & Disclaimers

The following are trademarks or registered trademarks of other companies:

Microsoft, Hyper-V, Windows, Windows NT and the Windows logo are trademarks or registered trademarks of Microsoft Corporation.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation.

UNIX is a registered trademark of The Open Group.

Linux is a trademark of Linus Torvalds.

InfiniBand is a registered trademark and servicemark of InfiniBand Trade Association.

Intelligent Clustered Storage is a trademark of Scaling Computing.

Other company, product, or service names may be trademarks or service marks of others.

NOTES:

Any performance data in this document was determined in a controlled environment. Actual results may vary significantly and are dependent on many factors including system hardware configuration and software design and configuration. There is no guarantee these measurements will be the same on commercially-available systems. Users of this document should verify the applicable data for their specific environment.

Information is provided “AS IS” without warranty of any kind.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. Virtual I/O Solutions may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice.

