Achieving Near-Native GPU Performance in the Cloud

John Paul Walters

Project Leader, USC Information Sciences Institute

jwalters@isi.edu

USC Viterbi School of Engineering

Information Sciences Institute

Outline

- Motivation
- ISI's HPC Cloud Effort
- Background: PCI Passthrough, SR-IOV
- Results
- Conclusion

Motivation

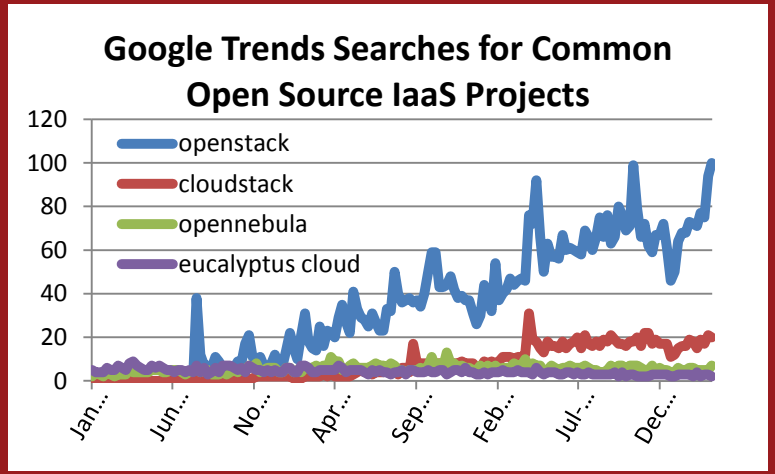
- Scientific workloads demand increasing performance with greater power efficiency
  - Architectures have been driven towards specialization, heterogeneity
- Infrastructure-as-a-Service (IaaS) clouds can democratize access to the latest, most powerful accelerators
  - If performance goals are met
- Can we provide HPC-class performance in the cloud?

ISI's HPC Cloud Work

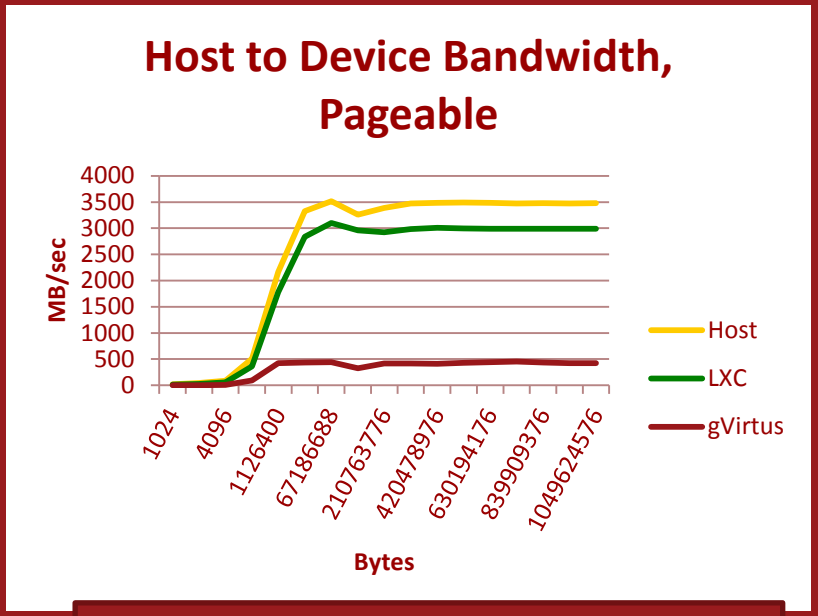
- Cloud computing is traditionally seen as a resource for IT
  - Web servers, databases
- More recently researchers have begun to leverage the public cloud as an HPC resource
  - AWS virtual cluster is 101 on Top500 list
- Major difference between HPC and IT in the cloud:
  - *Types of resources, heterogeneity*
- Our contribution: we're developing the heterogeneous HPC extensions for the OpenStack cloud computing platform

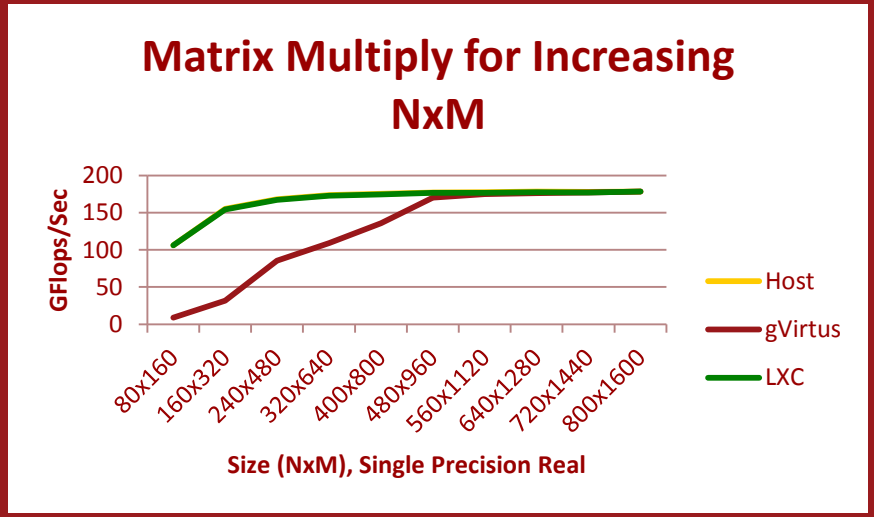
OpenStack Background

- OpenStack founded by Rackspace and NASA
- In use by Rackspace, HP, and others for their public clouds
- Open source with hundreds of participating companies
- In use for both public and private clouds
- Current stable release: OpenStack Juno
  - OpenStack Kilo to be released in April



Accessing GPUs from Virtual Hosts Using API Remoting

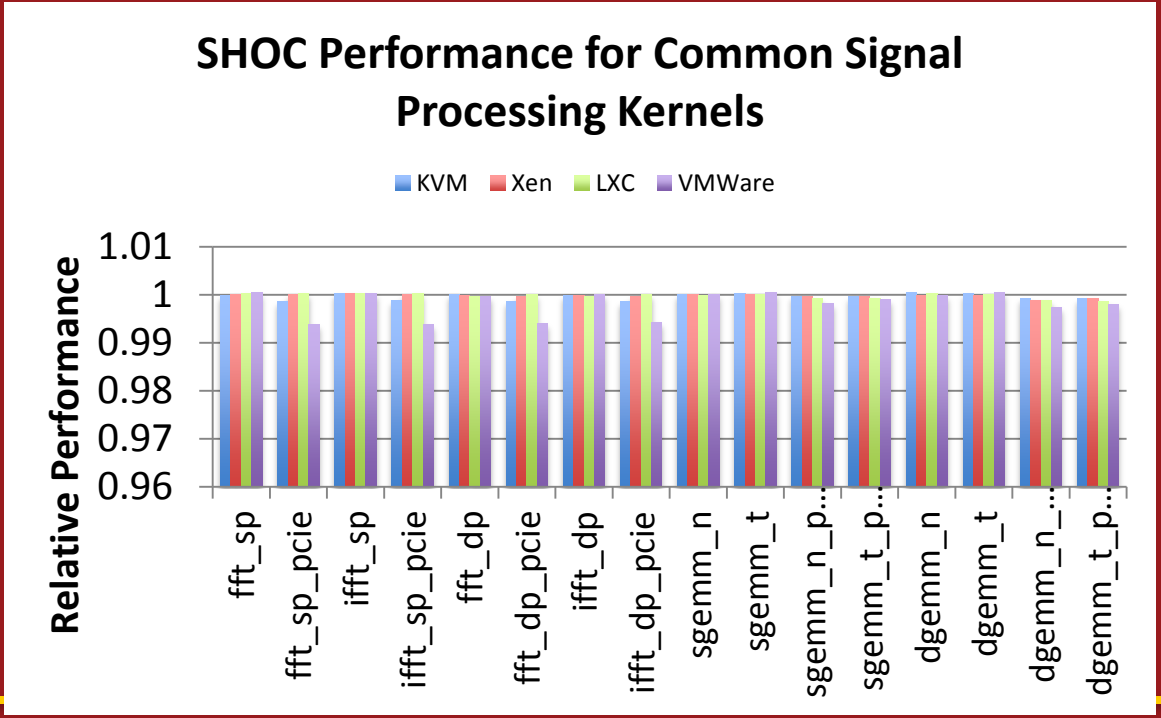




I/O performance low for gVirtus/KVM, LXC much closer to native performance.

Larger matrix multiply amortizes I/O transfer cost, LXC and native performance indistinguishable.

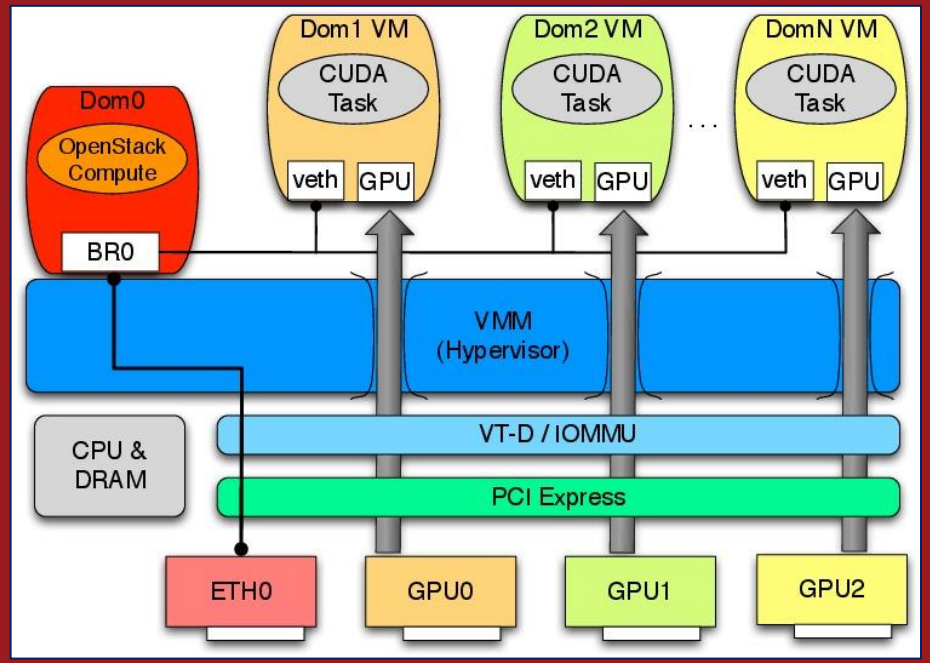
Accelerators and Virtualization



- Combine non-virtualized accelerators with virtual hosts
- Results in > 99% efficiency

PCI Passthrough Background

- 1:1 mapping of physical device to virtual machine
- Device remains non-virtualized



SR-IOV Background

- SR-IOV partitions a single physical device into multiple virtual functions
- Virtual functions almost indistinguishable from physical functions.
- Virtual functions passed to virtual machines using PCI passthrough

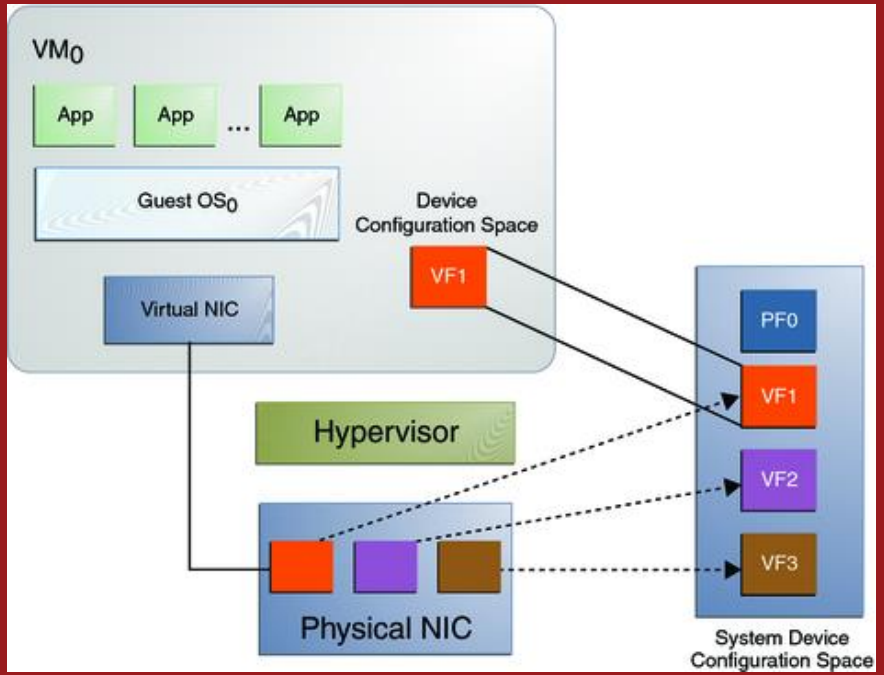
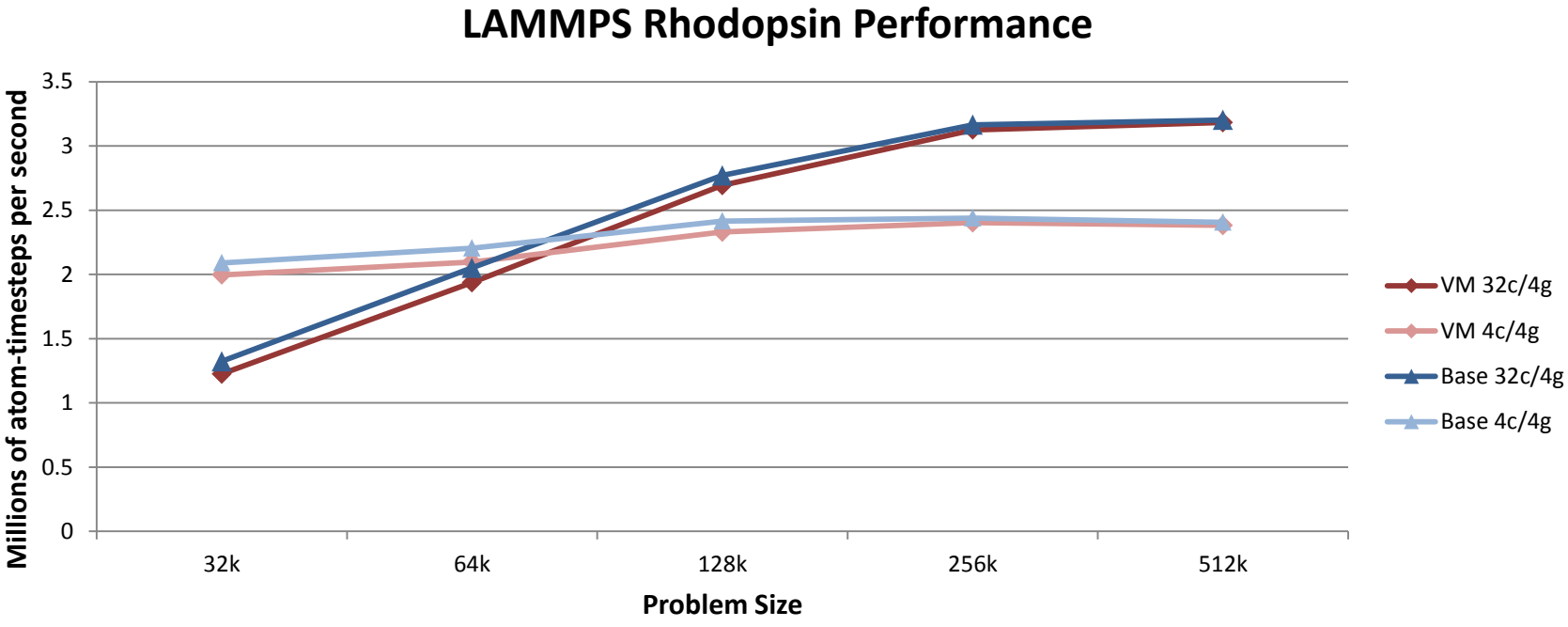


Image from: http://docs.oracle.com/cd/E23824_01/html/819-3196/figures/sriov-intro.png

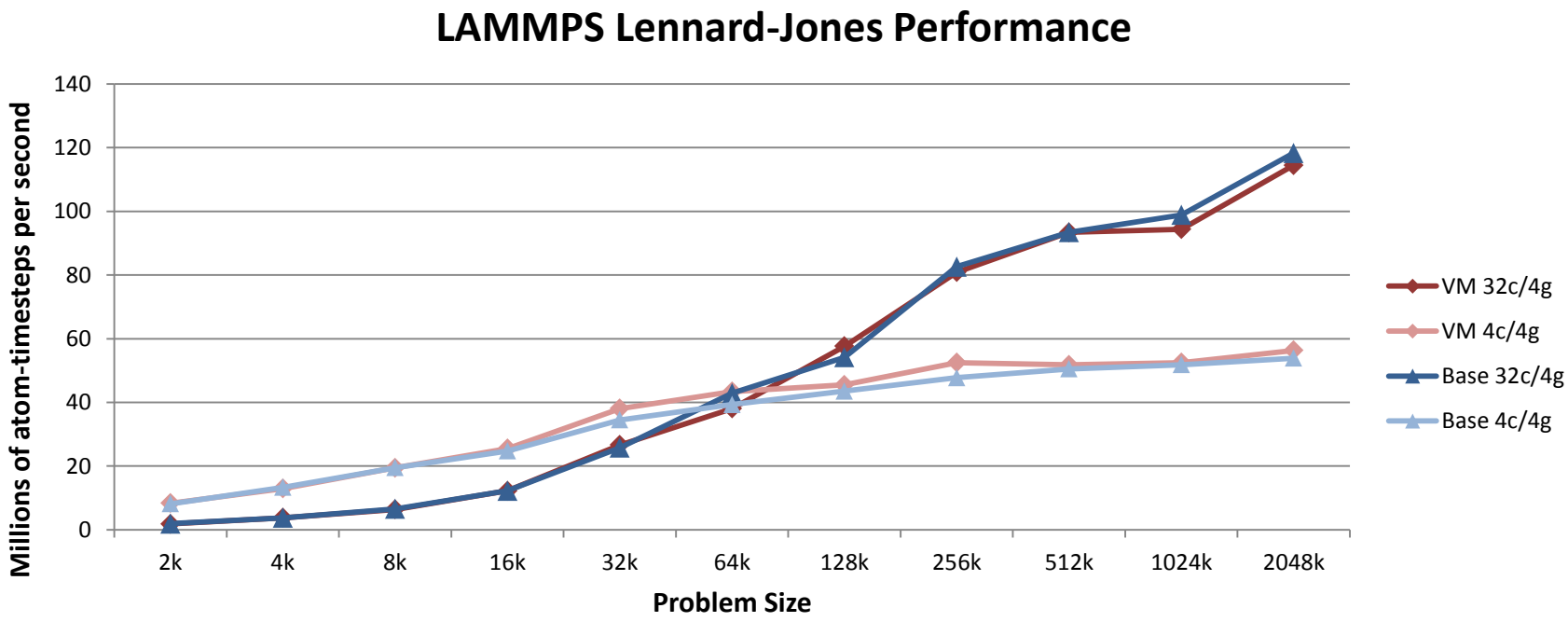
Multi-GPU with SR-IOV and GPUDirect

- Many real applications extend beyond a single node's capabilities
- Test multi-node performance with Infiniband SR-IOV and GPUDirect
- 4 Sandy Bridge nodes equipped with K20/K40 GPUs
  - ConnectX-3 IB with SR-IOV enabled
  - Ported Mellanox OFED 2.1-1 to 3.13 kernel
  - KVM hypervisor
- Test with LAMMPS, OSU Microbenchmarks, and HOOMD

LAMMPS Rhodopsin with SR-IOV Performance



LAMMPS Lennard-Jones with SR-IOV Performance



LAMMPS Virtualized Performance

- Achieve 96% - 99% efficiency
  - Performance gap decreases with increasing problem size
- Future work needed to validate results across much larger systems
  - This work is in the early stages

GPUDirect Advantage

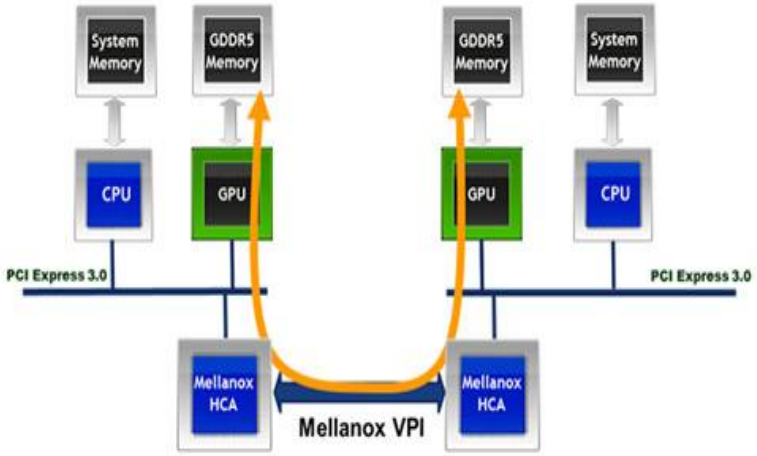
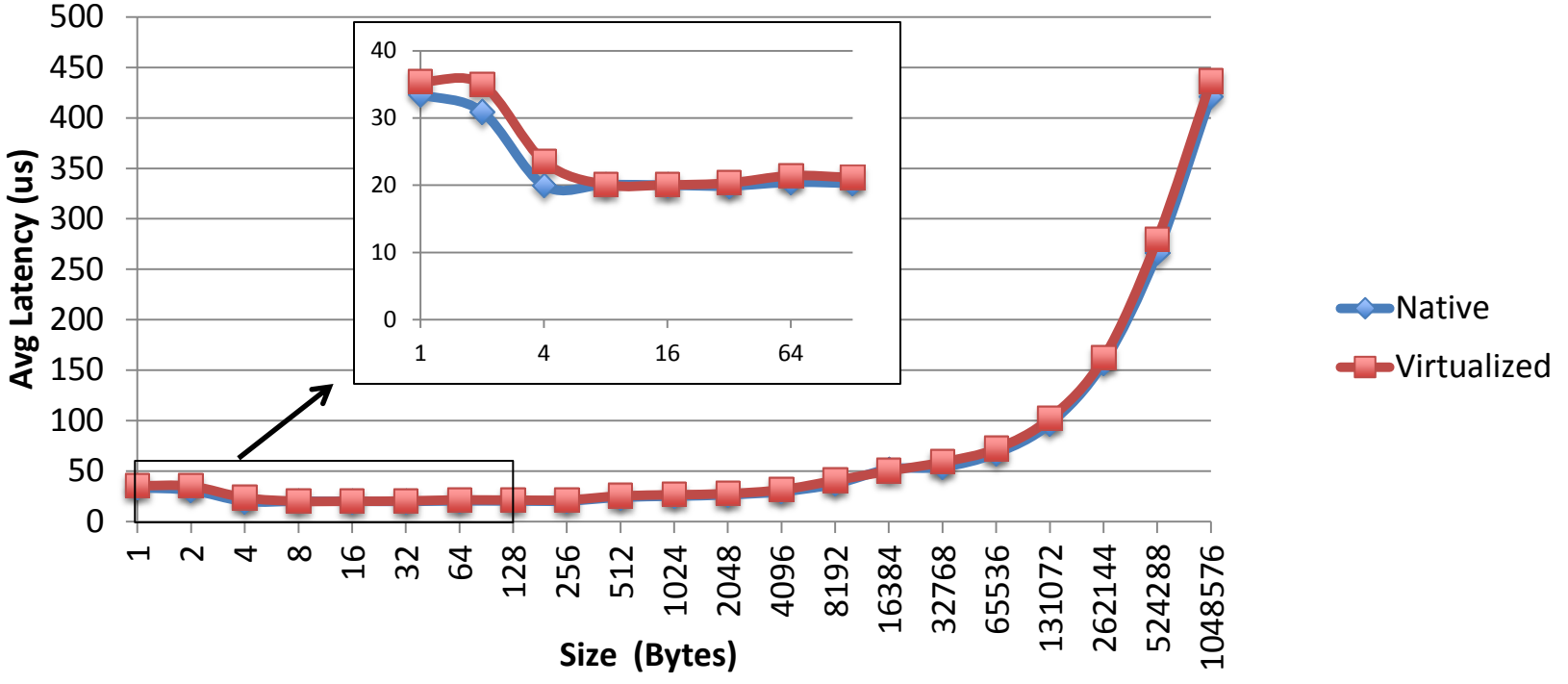


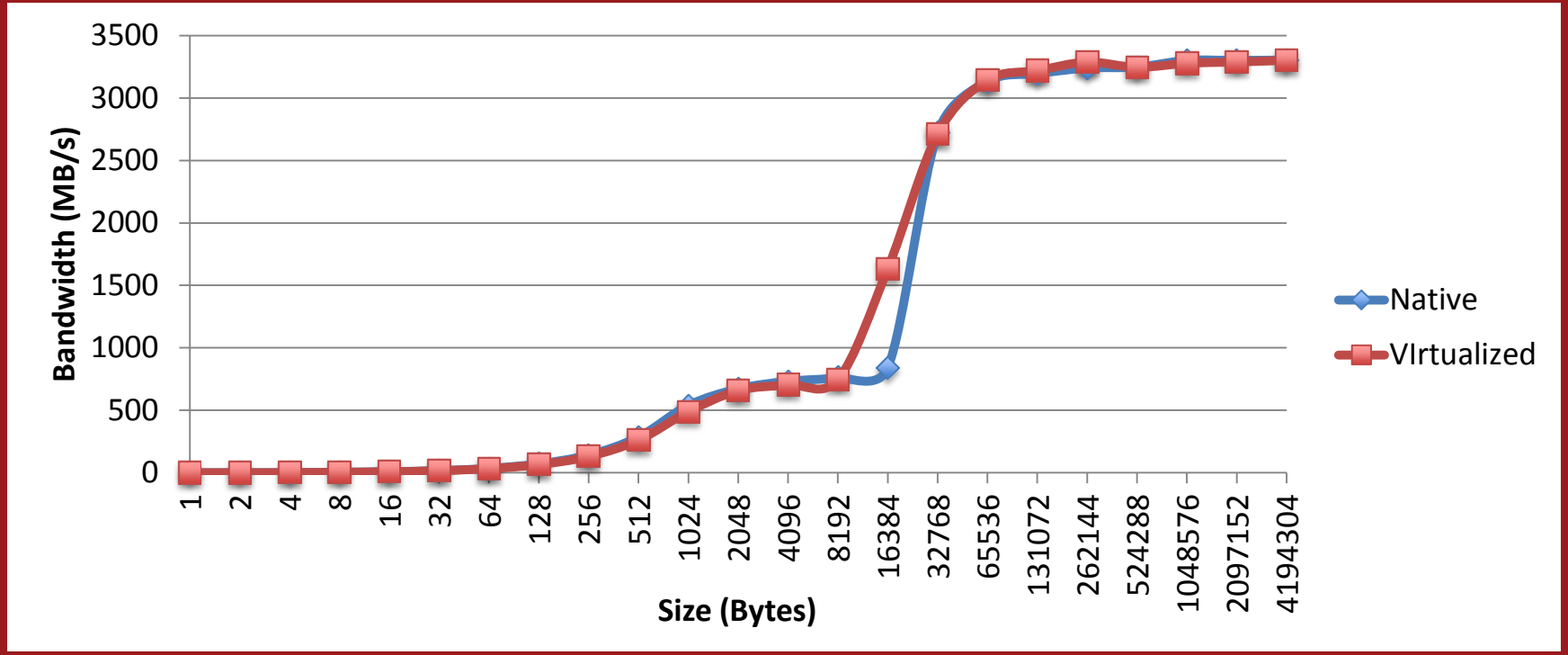
Image source: http://old.mellanox.com/content/pages.php?pg=products_dyn&product_family=116

- Validate GPUDirect over SR-IOV
  - Uses nvidia_peer_memory-1.0-0 kernel module
- OSU GDR Microbenchmarks
- HOOMD MD

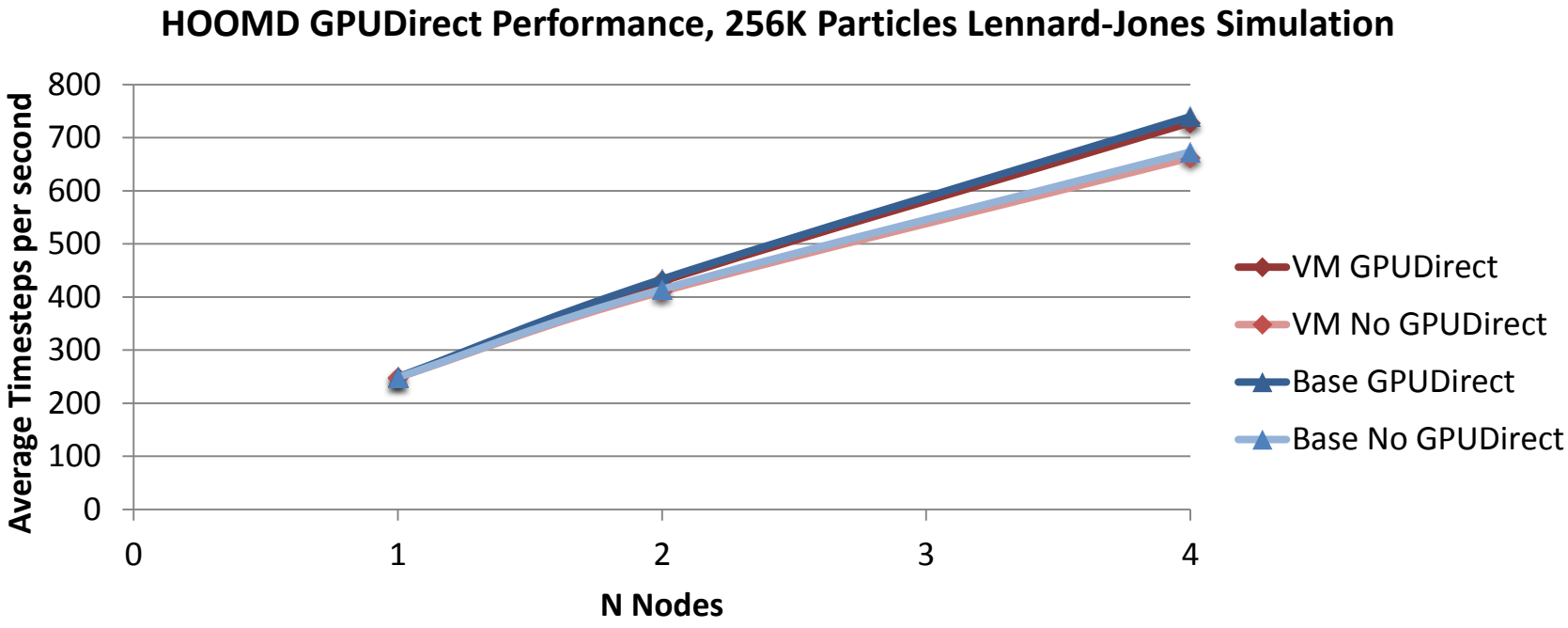
OSU GDR Microbenchmarks: Latency



OSU GDR Microbenchmarks: Bandwidth



GPUDirect-enabled VM Performance



Discussion

- Take-away: GDR provides nearly 10% improvement
- SR-IOV interconnect results in < 2% overhead
- Further work needed to validate these results in larger systems
  - Small-scale results are promising

Future Work

- **For full results see:**
  - J.P. Walters, et al. *GPU Passthrough Performance: A Comparison of KVM, Xen, VMWare ESXi, and LXC for CUDA and OpenCL Applications*, IEEE Cloud 2014
  - A.J. Younge, et al. *Supporting High Performance Molecular Dynamics in Virtualized Clusters using IOMMU, SR-IOV, and GPUDirect,* to appear in VEE 2015.
- **Next steps:**
  - Extend scalability results
  - OpenStack integration
- **Code: https://github.com/usc-isi/nova**

Questions and Comments

- **Contact me:**
  - jwalters@isi.edu
  - www.isi.edu/people/jwalters/