Performance Evaluation of the XDEM framework on the OpenStack Cloud Computing Middleware

X. Besseron[1] V. Plugaru[2] A. H. Mahmoudi[1]
S. Varrette[2] B. Peters[1] P. Bouvry[2]

[1]Luxembourg XDEM Research Centre (LuXDEM)
[2]Parallel Computing and Optimisation Group (PCOG)
Faculty of Science, Technology and Communication
University of Luxembourg, Luxembourg

SC-Camp 2015
(initially presented at PARENG 2015)

Outline

Cloud Computing for Discrete Element Method?

Processing of granular materials

- Snow, sand, gravel, coke, iron oxide, biomass, food, tablets, ...
- Widely used in industry
- Discrete Element Method (DEM)

DEM, HPC and Cloud Computing

- Huge computation time $\Rightarrow$ Parallel execution required
- Traditionally addressed using High Performance Computing (HPC) platforms
  - $\hookrightarrow$ Cloud Computing (CC) appears as a promising alternative

$\Longrightarrow$ How suitable is the Cloud Computing approach for an HPC workflow such as a DEM application?

Cloud Computing for Discrete Element Method?

Processing of granular materials

- Snow, sand, gravel, coke, iron oxide, biomass, food, tablets, ...
- Widely used in industry
- Discrete Element Method (DEM)

DEM, HPC and Cloud Computing

- Huge computation time $\Rightarrow$ Parallel execution required
- Traditionally addressed using High Performance Computing (HPC) platforms
  - $\hookrightarrow$ Cloud Computing (CC) appears as a promising alternative

$\Longrightarrow$ How suitable is the Cloud Computing approach for an HPC workflow such as a DEM application?

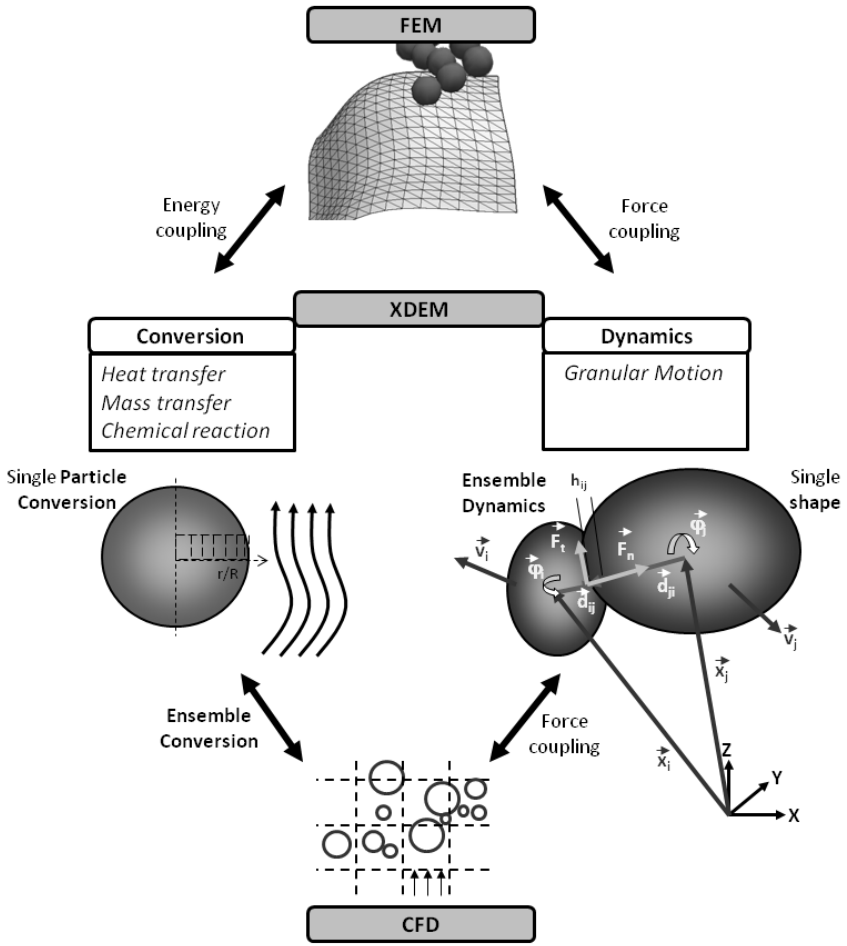
eXtended Discrete Element Method (XDEM)

XDEM software

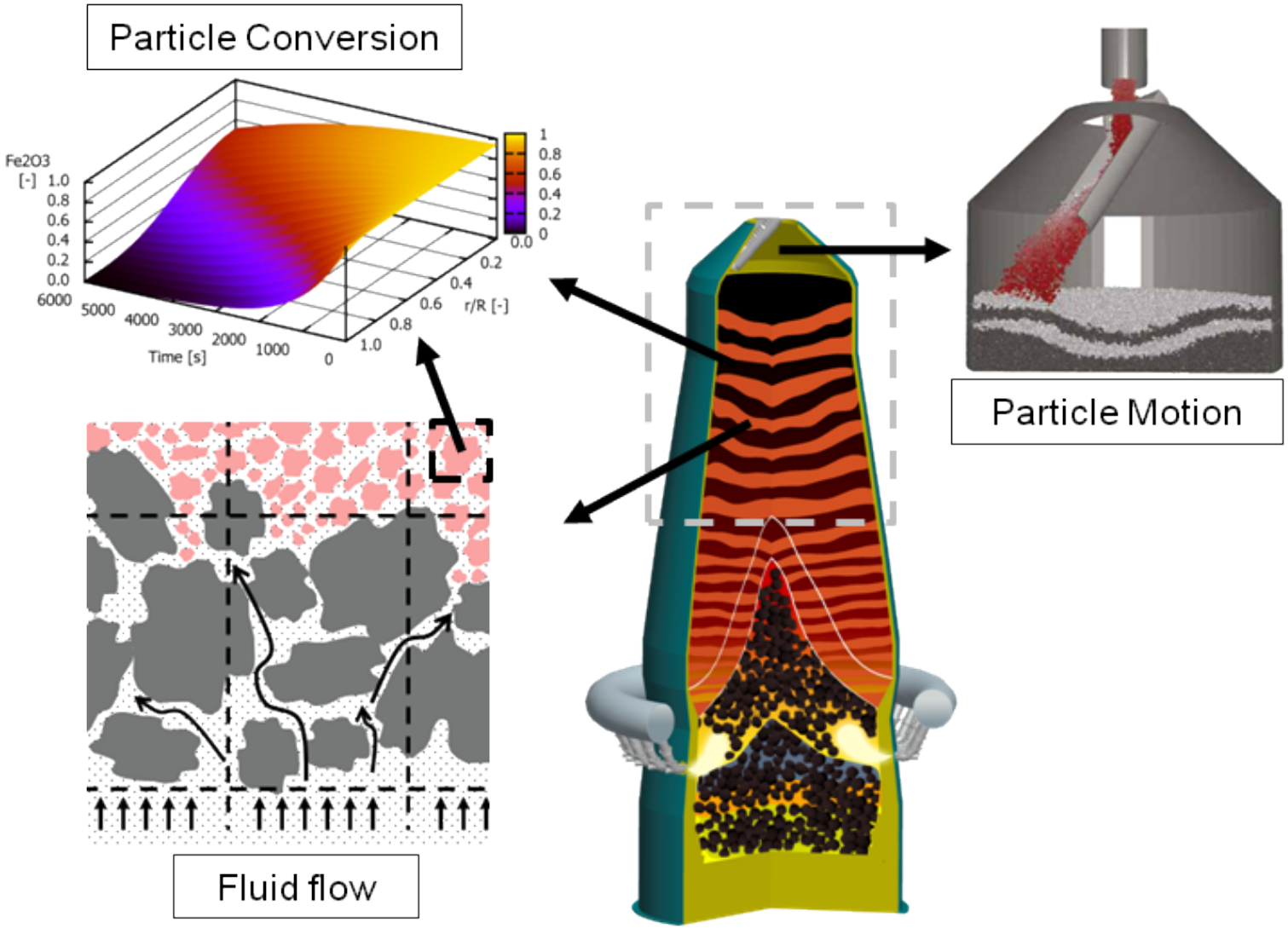
- numerical simulation framework
- extends the classic DEM approach
- parallel execution with MPI

Multi-physics simulation

- Particle motion
- Chemical conversion
- Finite Element Method (FEM) coupling (with Diffpack)
- Computational Fluid Dynamics (CFD) coupling (with OpenFoam)



XDEM example: Blast Furnace



Used in metallurgy to produce hard metals

Cloud Computing

On-demand, online access to computing resources and services

What kind of services?

- Software, Platform as a Service (SaaS, PaaS)
- Infrastructure as a Service (IaaS)
  - ↪ *i.e.* deploy your own OS, software layer and applications

IaaS relies on a virtualization layer

- Provisions user virtual machines on-demand
- Provides flexibility yet adds overhead to operations
- Hypervisors: Xen, KVM, VMWare ESXi, Microsoft Hyper-V, ...
- Cloud middleware: OpenStack, Eucalyptus, Nimbus, OpenNebula, VMWare vCloud, ...

Current study

Objectives

- Study the overhead of Cloud Computing middleware on an HPC workflow
- Extend previous work [1] to a real application

Systematic Performance Evaluation

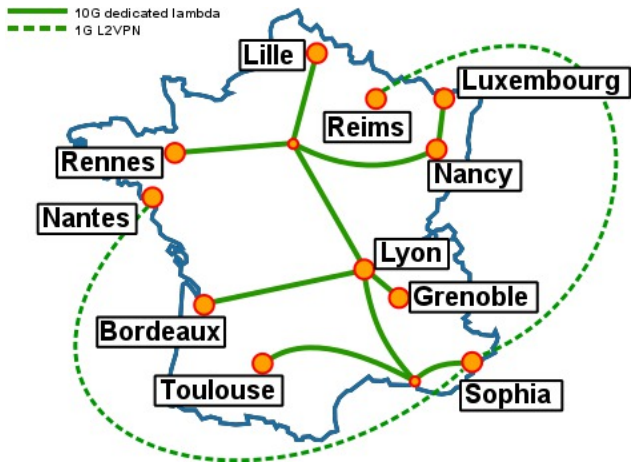
- Fair comparison using the exact same hardware
- Large scale distributed execution totalling hundreds of cores
- Real application with XDEM and a real-life test case
- Automated experimental framework and reproducible measurements

[1] S. Varrette, V. Plugaru, M. Guzek, X. Besseron, P. Bouvry
HPC Performance and Energy-Efficiency of the OpenStack Cloud Middleware
Heterogeneous and Unconventional Cluster Architectures and Applications Workshop (HUCAA'14)
Proc. of the 43rd Intl. Conf. on Parallel Processing (ICPP-2014)

Grid5000 & Kadeploy

Grid'5000

- Large scale nation wide infrastructure
  - ↪ 8 sites in France, 1 in Luxembourg
- 23 clusters, 941 nodes, 7494 cores
- Designed for large scale parallel and distributed computing research



Kadeploy

- Scalable, efficient and reliable deployment system
- Used as bare-metal provisioning solution
- Many OS environments pre-defined, easily customizable
- Integrates with KaVLAN to deploy inside isolated, routed or grid-global VLANs

Experimental setup

Three configurations

1. Native (no virtualization)
2. OpenStack with Xen hypervisor
3. OpenStack with KVM hypervisor

Two clusters

- *PetitPrince*: Intel-based, 10-Gigabit Ethernet
- *StRemi*: AMD-based, 1-Gigabit Ethernet

Computing and networking performance

- Performance metric: XDEM iteration time
  - ↪ Reported value: average of at least 20 measurements
- Only **one virtual machine per physical node**
- In-memory file reading/writing operations

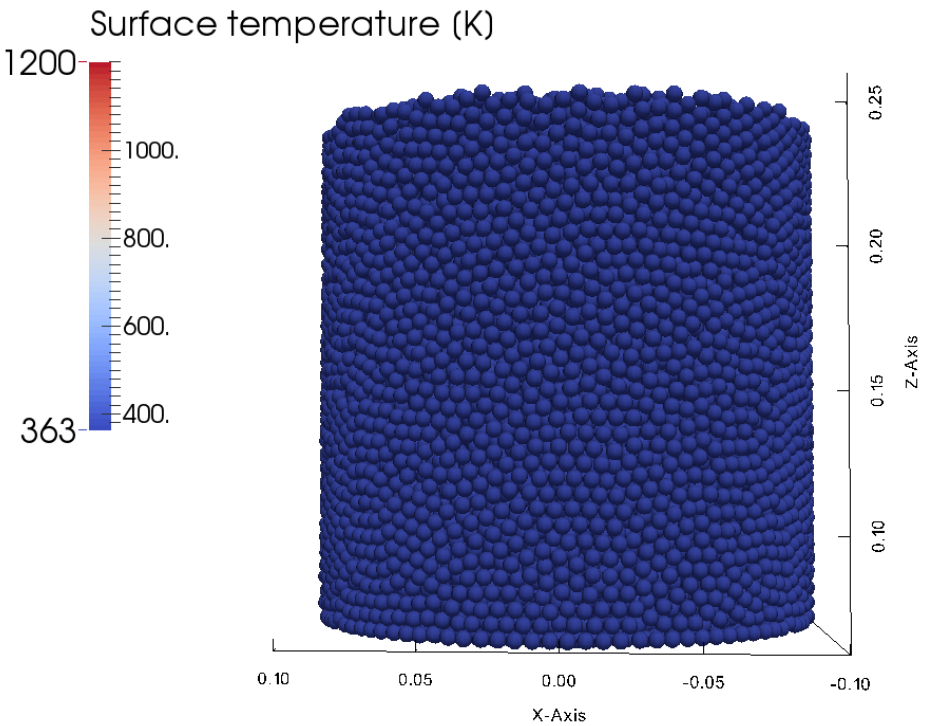
Test case: Biomass pyrolysis with XDEM

Wood decomposition reactions

$$
\begin{aligned}
Wood &\rightarrow Char \\
Wood &\rightarrow Tar \\
Wood &\rightarrow \nu_{CO}\ CO + \nu_{CO_2}\ CO_2 \\
&+ \nu_{H_2O}\ H_2O + \nu_{H_2}\ H_2 + \nu_{CH_4}\ CH_4
\end{aligned}
$$

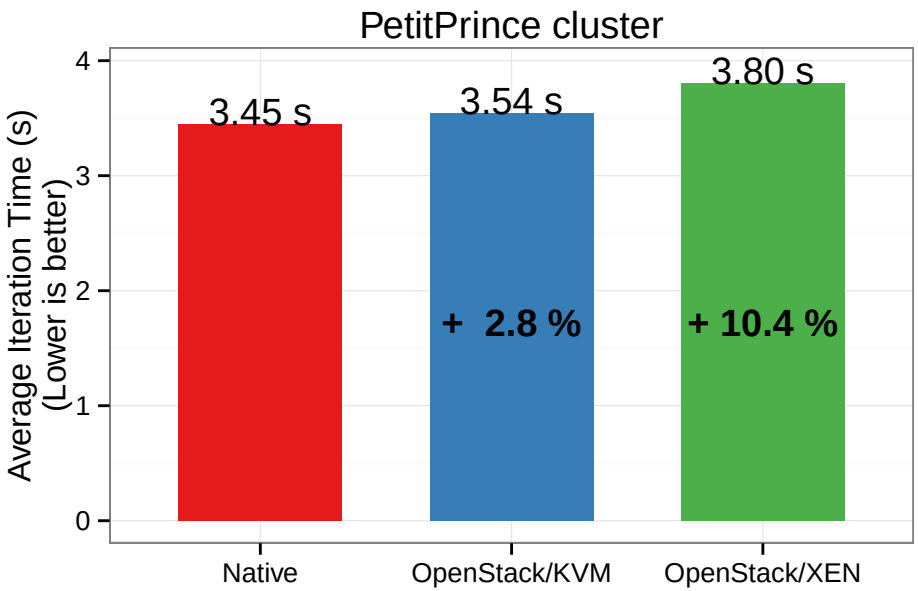
Initial conditions

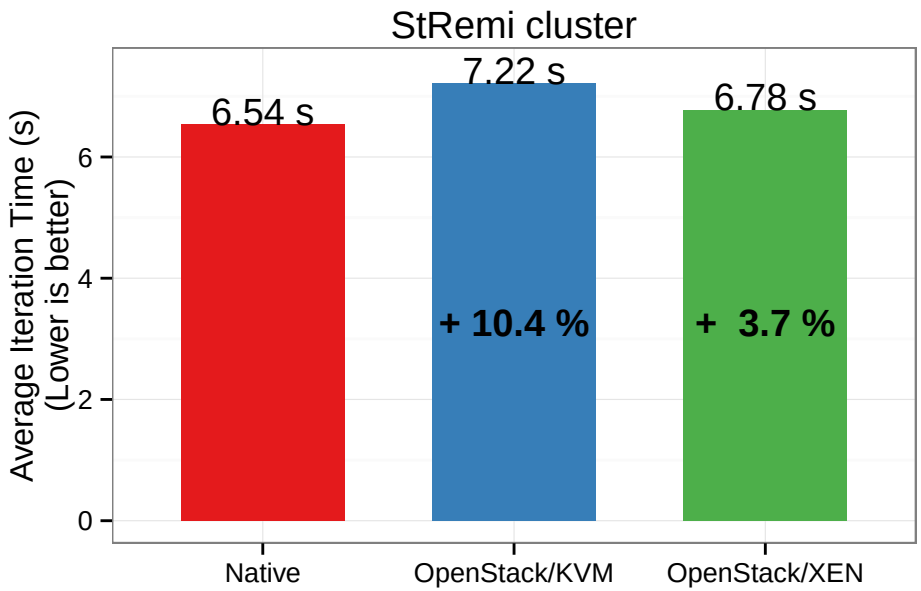
- Wood packed bed of 19 *cm*
- 32.000 spherical particles with diameter of 6.2 *mm*
- Wood particles initially at 363 *K* with 8% *wb* moisture
- Surrounding gas temperature 1200 *K* and pressure of 1 *bar*



Single-core performance

- Sequential execution, only one process with one thread
- Overhead of virtualization on computing performance

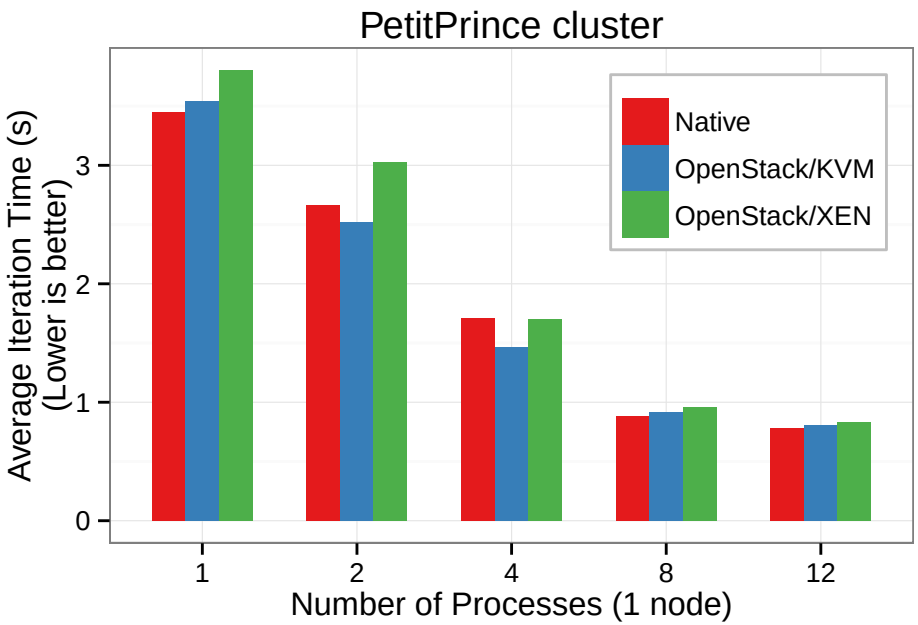


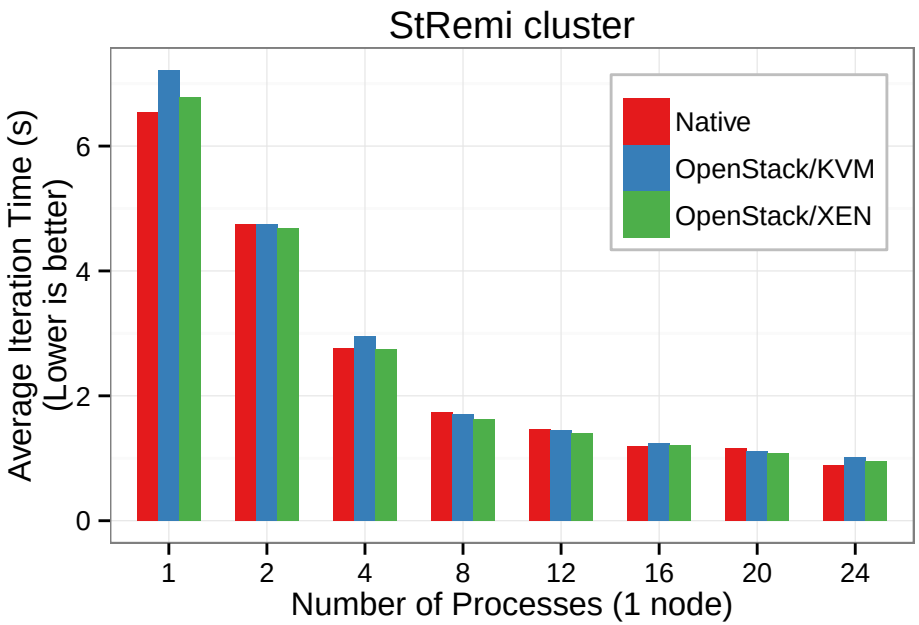


$\Longrightarrow$ Overhead between 2.8 % and 10.4 %

Single-node performance

- Execution on one full node, one process per core
- MPI communications using shared memory

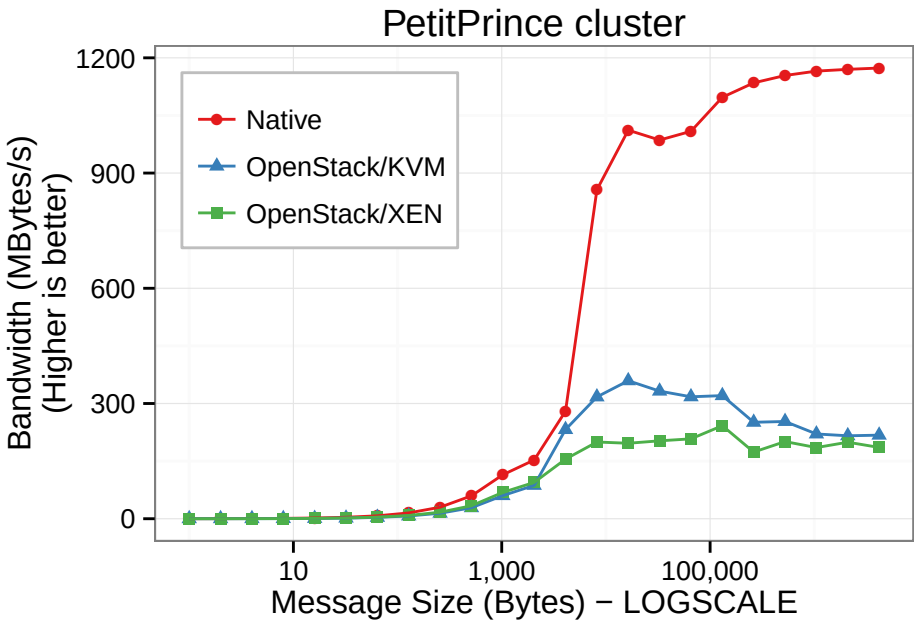


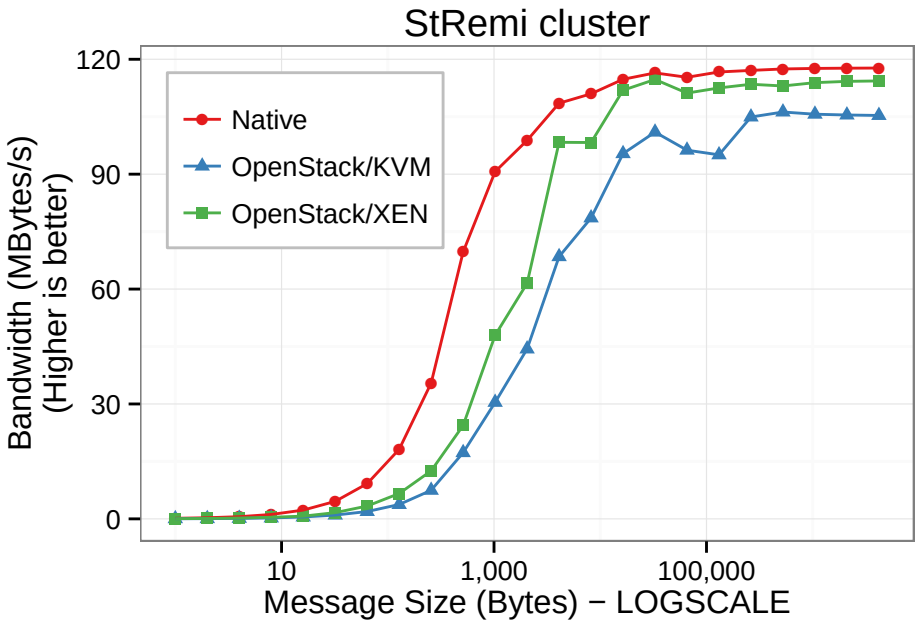


	KVM overhead	XEN overhead
PetitPrince (12 processes)	3.5%	6.4%
StRemi (24 processes)	14.0%	8.1%

Internode communication

- OSU Micro-Benchmarks
- MPI internode bandwidth
- Two processes on two different nodes



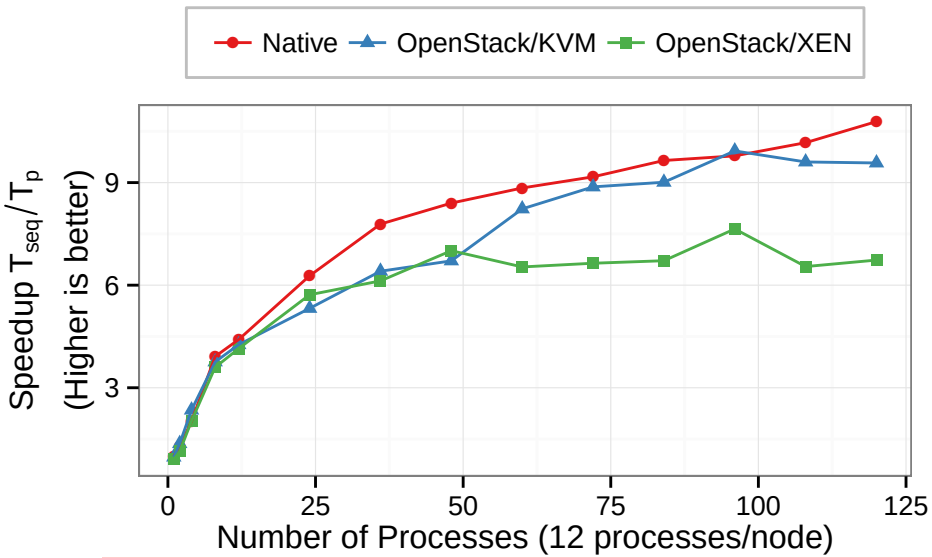


- Virtualized environments cannot sustain more than 25 % of the available bandwidth

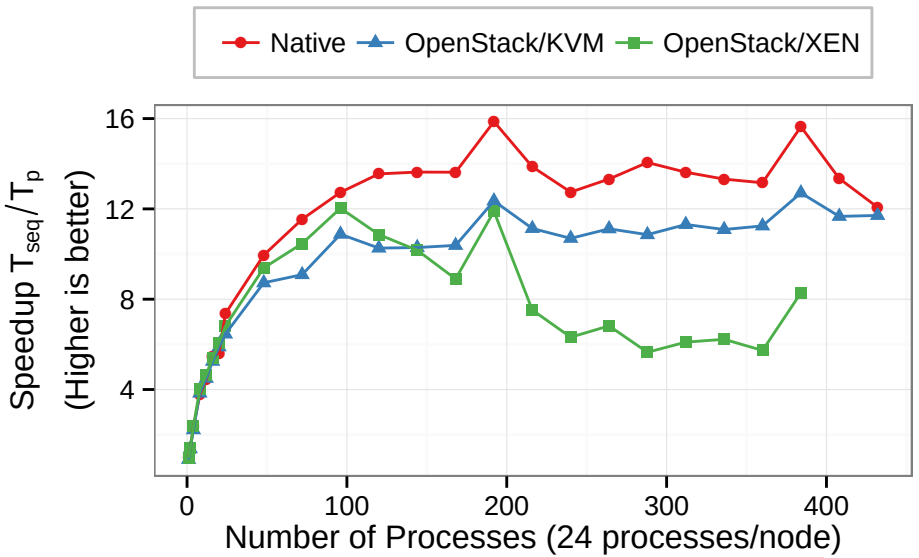
Multi-node performance

- Distributed execution, one process per core
- MPI communication using Ethernet network between nodes
- Speedup = T_{seq}^{Native} / T_p

PetitPrince cluster



StRemi cluster



$\Longrightarrow$ OpenStack/KVM performs better than OpenStack/Xen

Multi-node performance

Summary for the best configuration in the scalability range

	PetitPrince 10 nodes, 120 processes		*StRemi* 8 nodes, 192 processes	
	Iteration Time	**Speedup**	**Iteration Time**	**Speedup**
Native	0.32 s	10.8	0.41 s	15.9
OpenStack/KVM	0.36 s + 12.6%	9.6	0.53 s + 28.5%	12.4
OpenStack/XEN	0.51 s + 60.1%	6.7	0.55 s + 33.4%	11.9

Conclusions & Future Works

Conclusions – HPC vs CC

- **Sequential** executions: moderate overhead < 11 %
- **Distributed** executions: moderate (10 %) to significant (60 %) overhead
- Communication appears to be the bottleneck
  - ↪ study limited to Ethernet (no InfiniBand)
- Cloud Computing offers other advantages
  - ↪ on-demand provisioning, cost reduction, ...

Future Works

- Other test cases, other applications
- Newer versions of OpenStack, Xen, KVM
- Single Root I/O Virtualization (SR-IOV) technology (promising for InfiniBand)
- LinuX Containers (LXC)

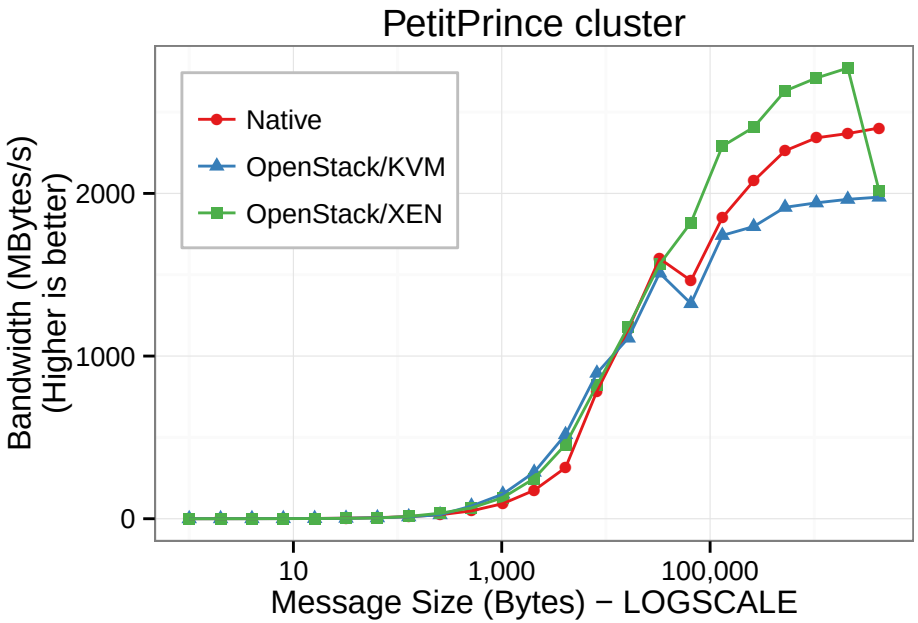
Thank you for your attention!

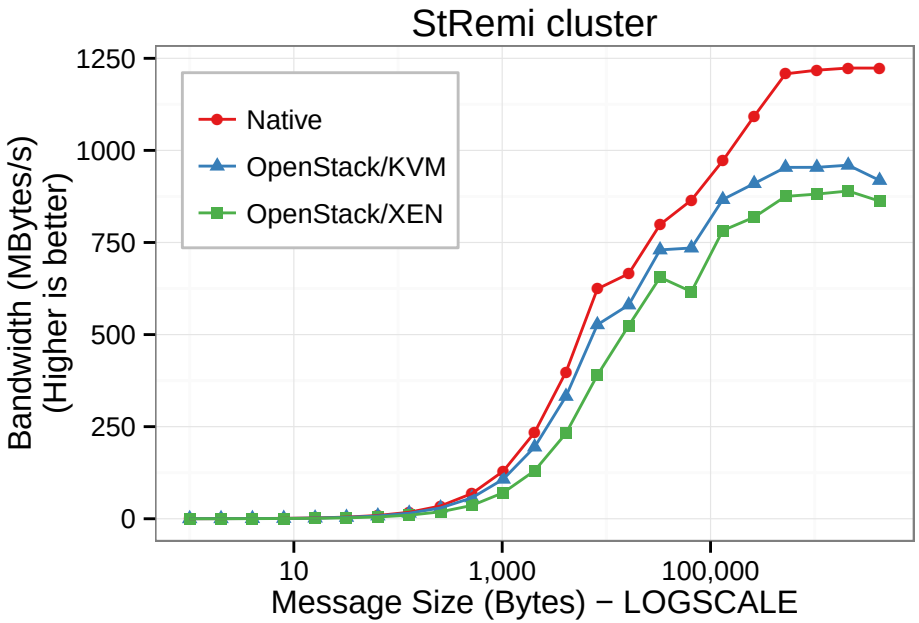
Questions?

Luxembourg XDEM Research Centre (LuXDEM) `http://luxdem.uni.lu`
Parallel Computing and Optimisation Group (PCOG) `http://pcog.uni.lu`

Intranode communication

- OSU Micro-Benchmarks
- MPI intranode bandwidth
- Two processes on the same node





Hardware and software configuration

Cluster	*PetitPrince*	*StRemi*
Site	Luxembourg	Reims
Processor type	Intel Xeon	AMD Opteron
Processor model	E5-2630L @ 2GHz	6164 HE @ 1.7GHz
#nodes	16	44
#CPUs per node	2	2
#cores per node	6	12
Memory per node	32 GBytes	48 GBytes
Network	10-Gigabit Ethernet	1-Gigabit Ethernet
Operating System (Hyp.)	Ubuntu 12.04 LTS, Linux 3.2	
Operating System (VM)	Debian 7.1, Linux 3.2	
Cloud Middleware	OpenStack Essex	
OpenMPI	1.4.3	
OSU Micro Benchmark	4.4.1	
XDEM software	Internal v2015.01.05	