
CERN

GPUs in OpenStack

Experiences in validating performance

Konstantinos Samaras-Tsakiris

IT-CM-RPS

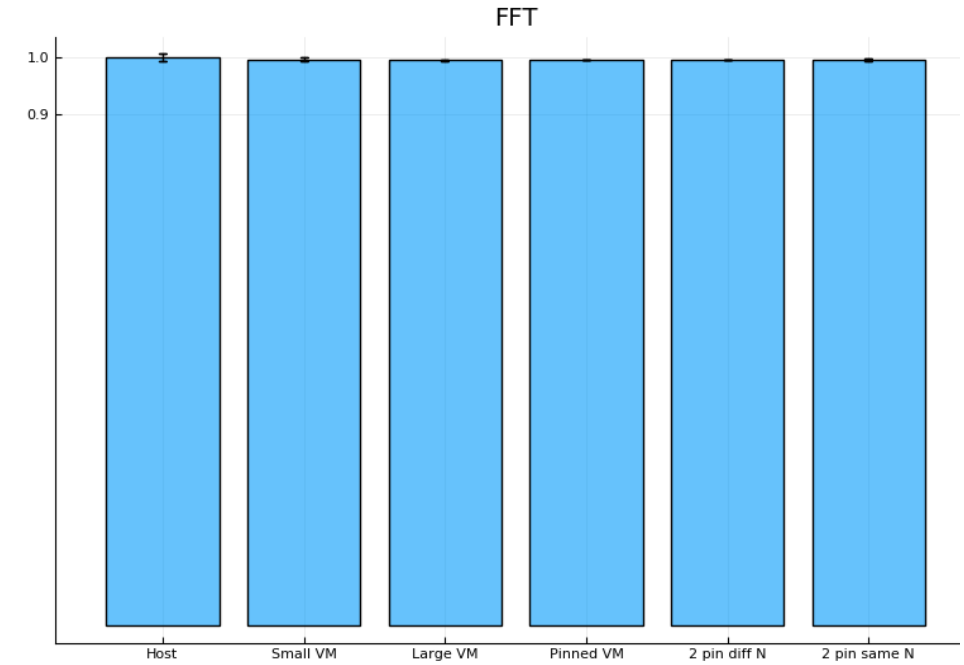
- **Performance penalty for GPU computations in VMs?**

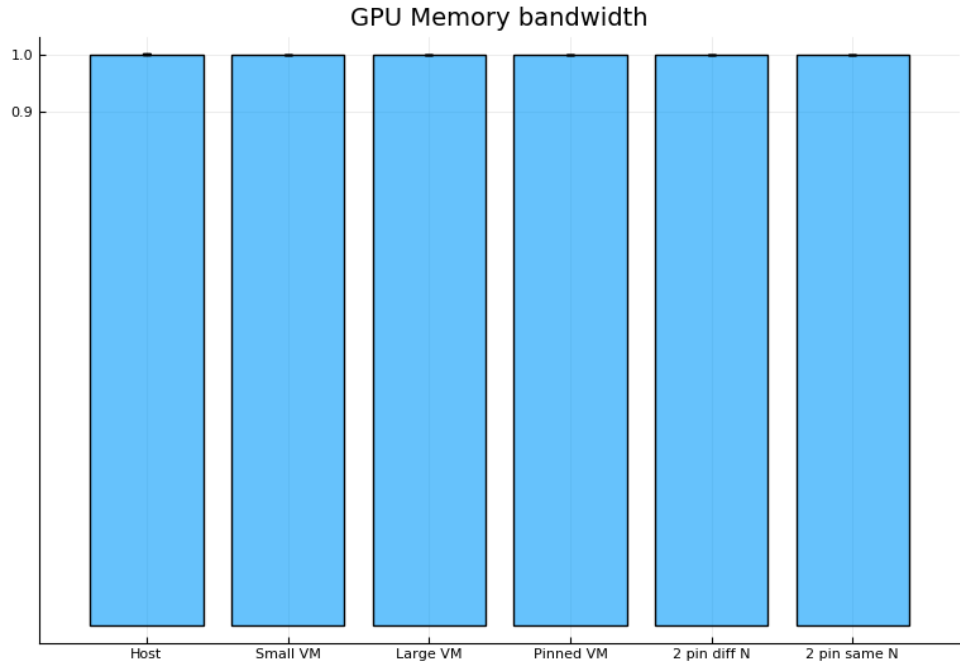
- **Performance penalty for GPU computations in VMs?**
  - ⇒ **Benchmarking**

GPU performance: VM vs baremetal

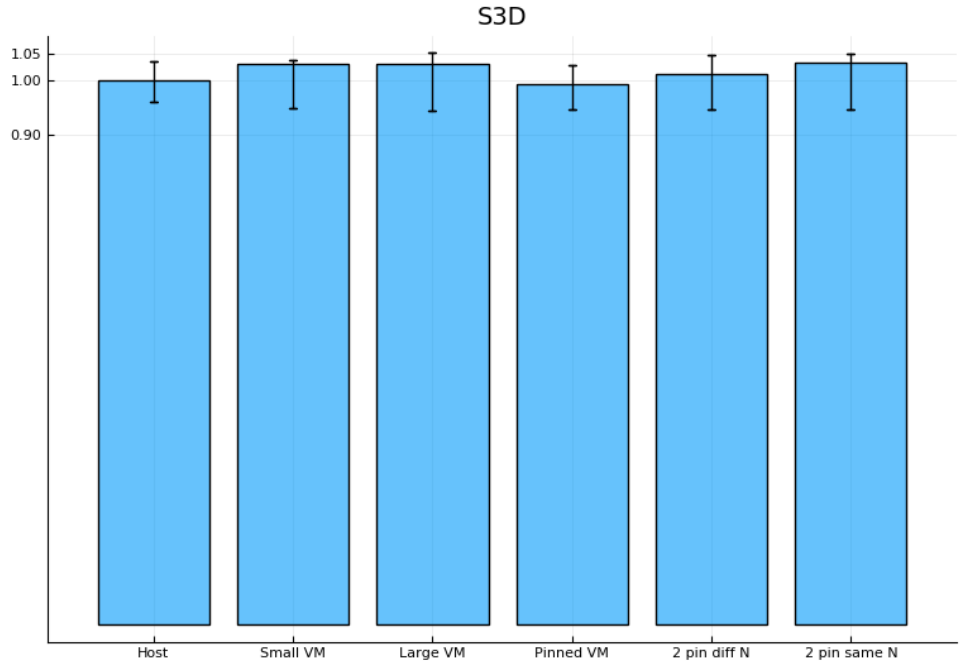
- PCI passthrough GPU
  - ⇒ no penalty expected
- Benchmarks
  - CUDA samples
  - SHOC
    - Low-level characteristics like bus bandwidth
    - Base algorithms like FFT
    - Applications like S3D
    - ➢ but *short run times*

GPU performance: VM vs baremetal



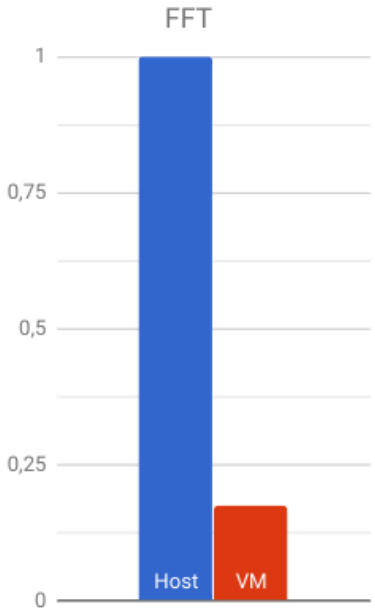


GPU performance: VM vs baremetal



No significant performance impact, as expected

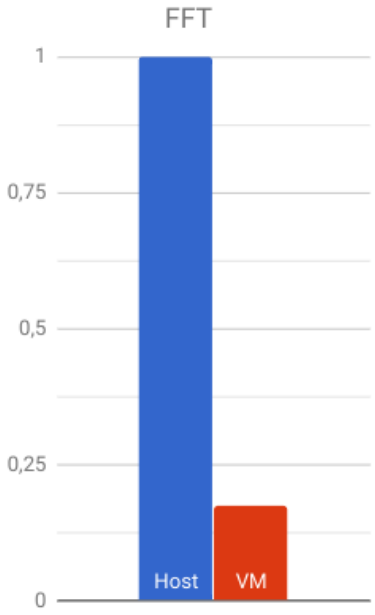
But…!



This result also came up occasionally!

- Difference: `nvidia-smi` (GPU monitoring) wasn't running
- Benchmark run time: ~1sec

But…!



This result also came up occasionally!

- Difference: `nvidia-smi` (GPU monitoring) wasn't running
- Benchmark run time: ~1sec

Hypothesis

- some GPU management operation has extra latency (eg changing power state)
- *Never appears on longer-running benchmarks*

Real world testing

Real world testing

- ➔ Tensorflow: deep learning
- ➔ BioDynaMo: simulation of biological tissue dynamics

Very low performance on BioDynaMo!

- Large slowdown on CPU-only parts (also on CPU-GPU parts)
- Hypervisor
  - `acpi_pad` takes up all CPU
  - CPU frequency: 800MHz (nominal: 2100MHz)

Very low performance on BioDynaMo!

- Large slowdown on CPU-only parts (also on CPU-GPU parts)
- Hypervisor
  - `acpi_pad` takes up all CPU
  - CPU frequency: 800MHz (nominal: 2100MHz)

Fix: kernel parameters `intel_idle.max_cstate=0 processor.max_cstate=0`

- No `acpi_pad`
- Frequency reported as 2100MHz
- BioDynaMo performs much better (*still less than reference*)

Not a VM problem

After the “fix”

- VM performance similar to hypervisor
- ⅓ of a similar machine!

Not a VM problem

After the “fix”

- VM performance similar to hypervisor
- ⅓ of a similar machine!

➢ HEPSPEC on hypervisor
  - Score now: 90.82
  - At procurement: **330**
  - Same result on another node from the same delivery

Not a VM problem

After the “fix”

- VM performance similar to hypervisor
- ⅓ of a similar machine!

➢ HEPSPEC on hypervisor
  - Score now: 90.82
  - At procurement: **330**
  - Same result on another node from the same delivery

⇒ Hardware / firmware / BIOS configuration issue

… But a firmware issue

BIOS update fixed it!
(at least on the sibling machine, haven't applied it yet to hypervisor)

- HEPSPEC: **331** ≈ procurement score

Future work

What about a standardized GPU benchmark?

GPU use cases

- Particle simulations
- Machine learning

Different requirements, different hardware optimizations
(Tesla V100 tensor cores)

Proposition for machine learning:

➢ MLPerf
https://www.hpcwire.com/2018/05/02/mlperf-will-new-machine-learning-benchmark-help-propel-ai-forward/

Virtualized GPU

Alternative way to provision GPUs, vendor-specific. For Nvidia:

- Only device management actions are virtualized
- Divide GPU memory among tenants
- Share compute resources

Benchmark

- Virtualization penalty (again, none expected)
- QoS

Summary

- GPUs can be cloud provisioned
  - Device passthrough OR virtual GPU
- No performance (throughput) penalty with passthrough GPUs
  - Possibility of extra latency for device management operations
- Underlying hypervisor problem (CPU firmware) discovered inadvertedly
- Need for a standardized GPU benchmark: MLPerf?