Live Migration of vGPU

Aug 2016

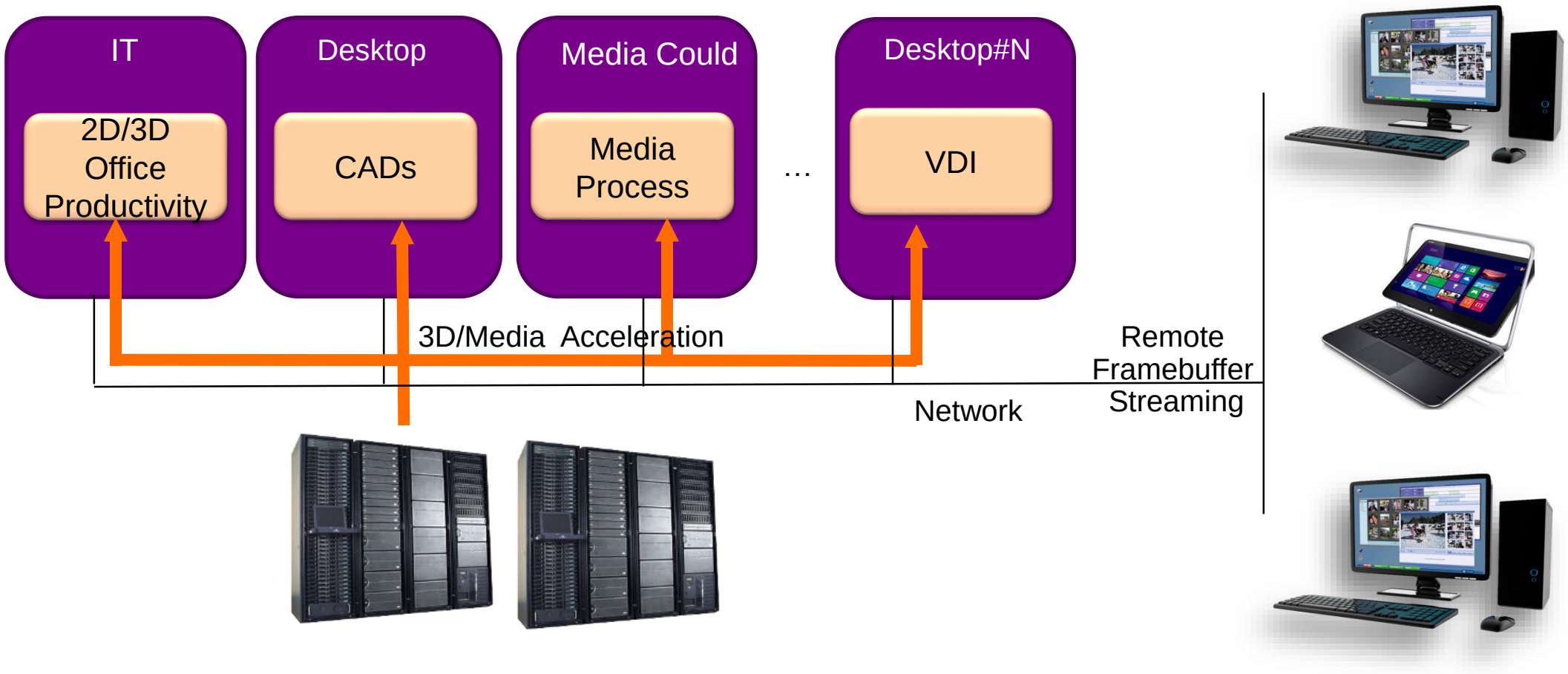
Xiao Zheng xiao.zheng@Intel.com

Kevin Tian kevin.tian@Intel.com

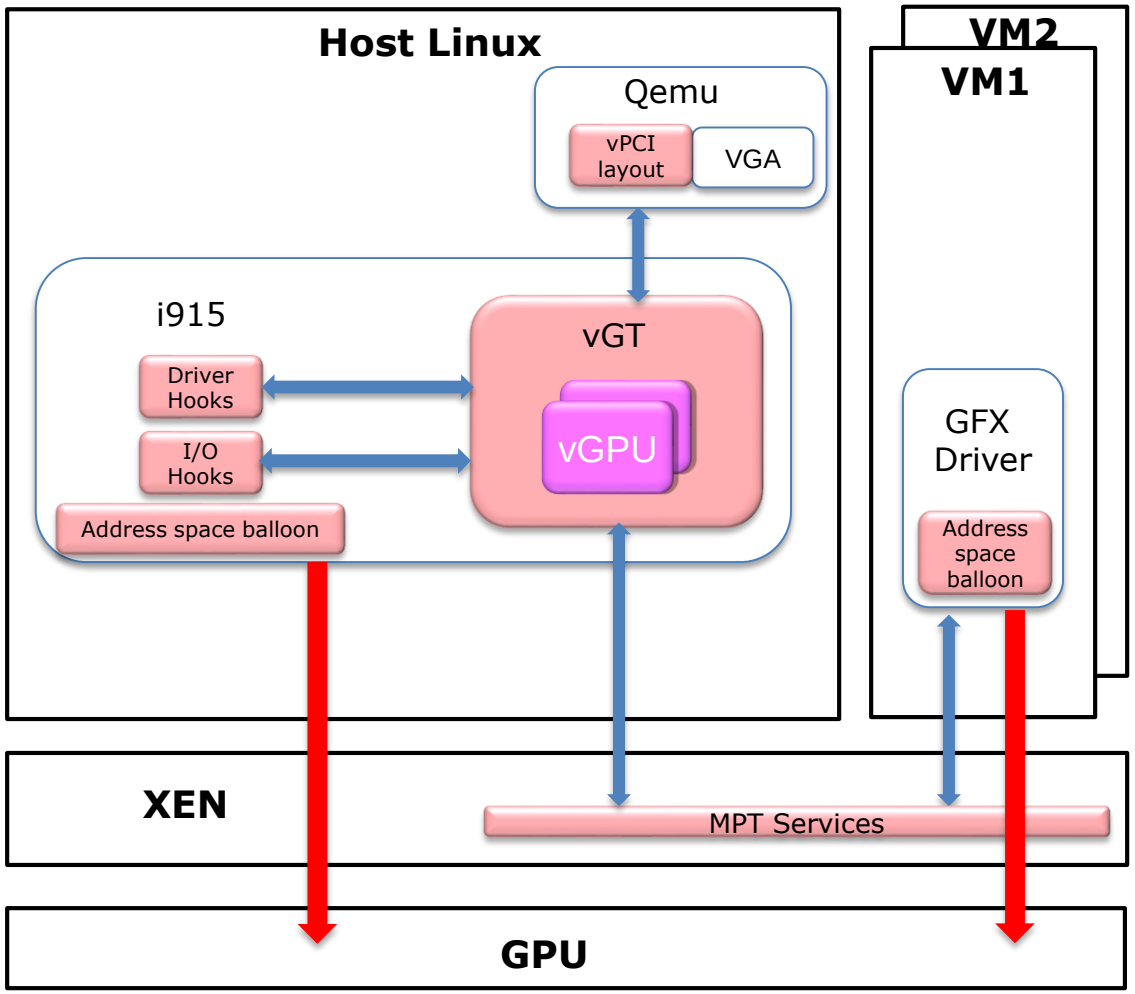
Agenda

- GPU Virtualization and vGPU Live Migration
- vGPU Resources
- Design and Solution
- Current Status
- Summary

GPU Virtualization Usage Cases



XENGT Architecture – Mediated Pass-through

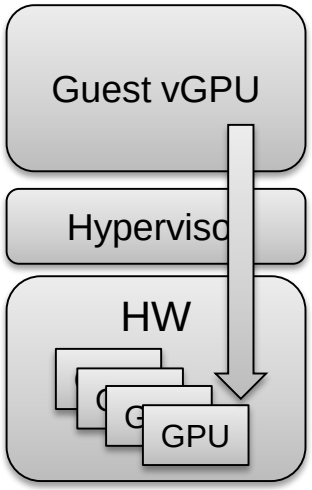


- pass-through for performance critical resource
- Trap and emulate for privileged resource
- Time-shared among VMs

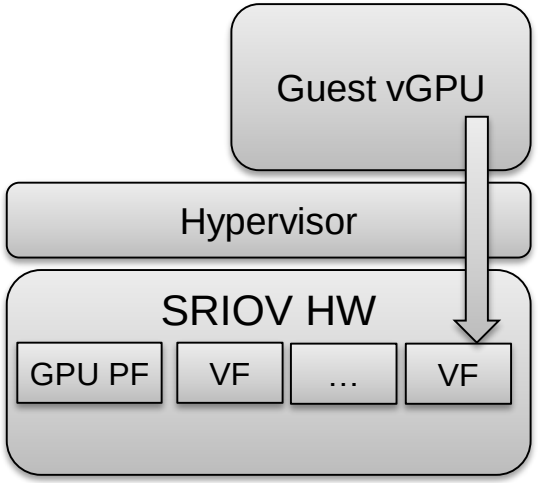
vGPU Live Migration

Live Migration: Load balance, Maintenance, Fault recovery
Unfortunately most of vGPU solutions do not support migration except Intel® GVT-g

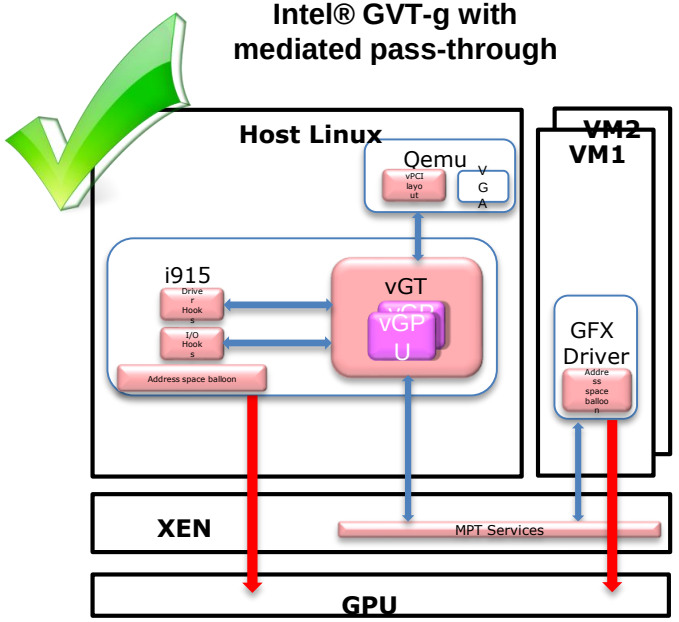
Typically a GPU pass-through solution



Typically a GPU SRIOV solution



Intel® GVT-g with mediated pass-through

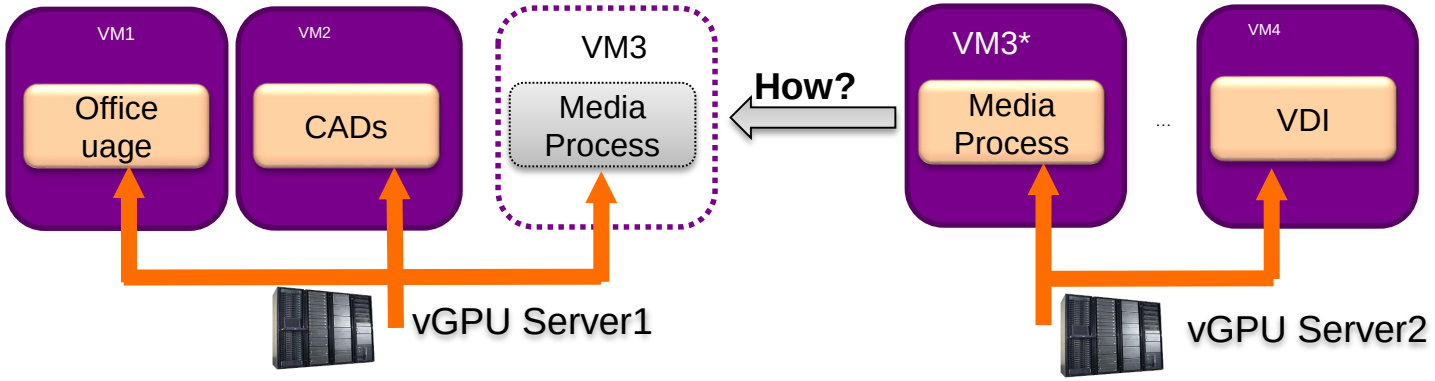


Intel® GVT-g architecture (Mediation) make it possible for seamless live migration

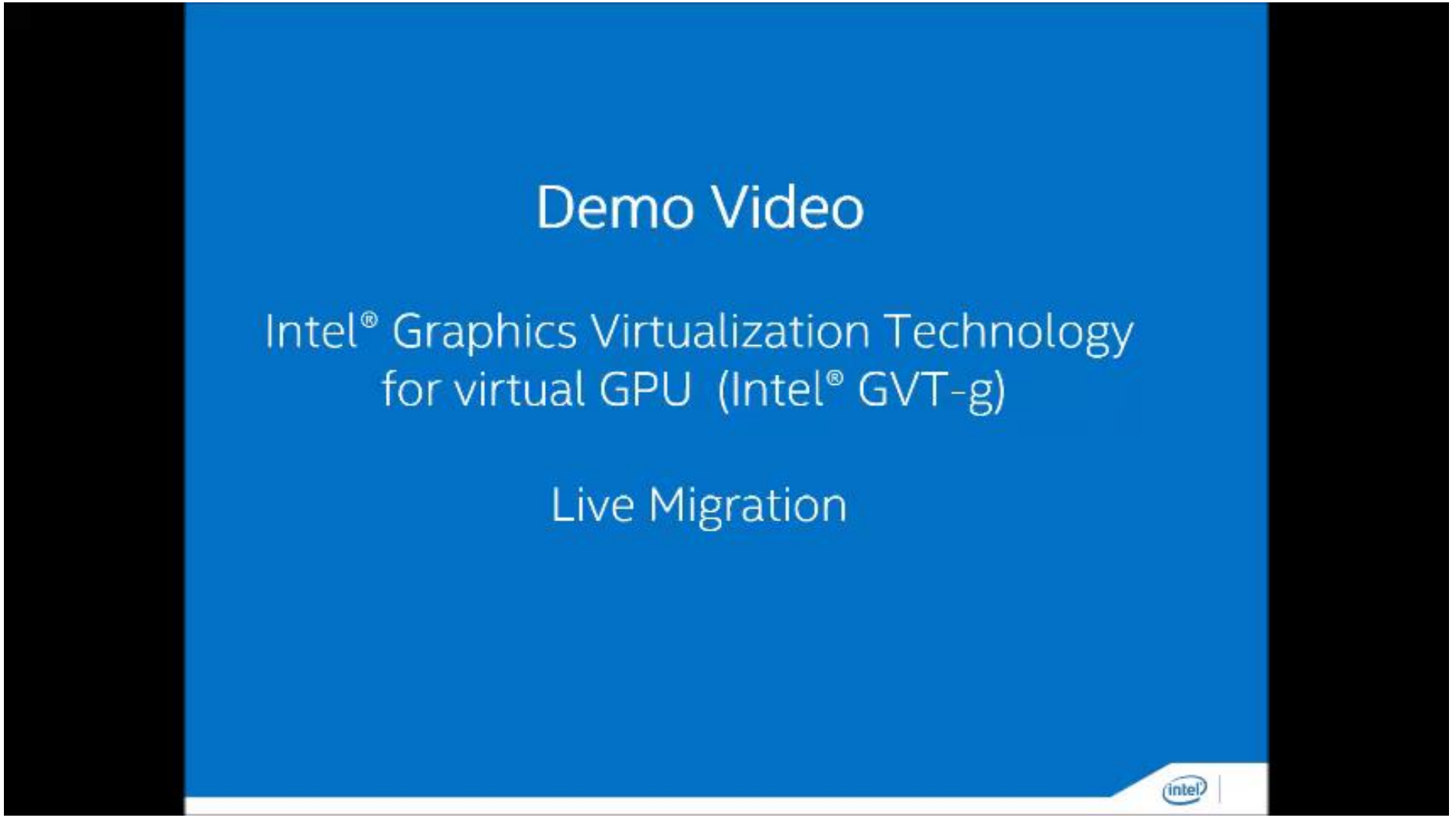
Live Migration of vGPU in Intel® GVT-g

Highlight feature:

- Intel® GVT-g is Open Source project, upstream ongoing
- vGPU Live Migration follows existing hypervisor migration flow
- 3D/2D/Media graphics workload seamless migrated between Servers or Local machine
- Support Linux/Windows Guest
- Live Migration Service downtime latency < 0.3 sec (Guest RAM 2GB, assigned 512MB vGPU memory, 10Gpbs adapter)



Demo: vGPU Live Migration with 3D workload



https://www.youtube.com/watch?v=y2SkU5JODIY

Agenda

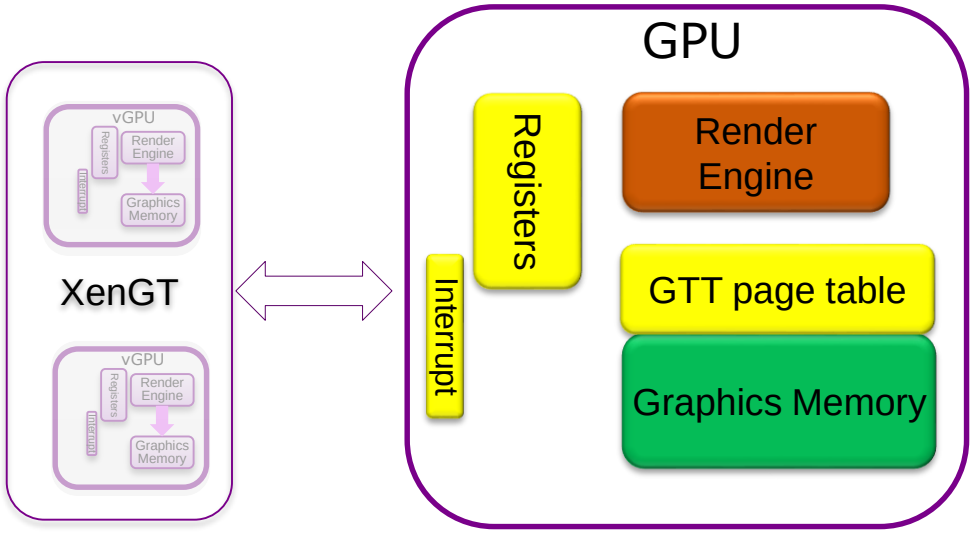
- GPU Virtualization and vGPU Live Migration
- vGPU Resources
- Design and Solution
- Current Status
- Summary

Inside of vGPU instance

pass-through for performance critical resource

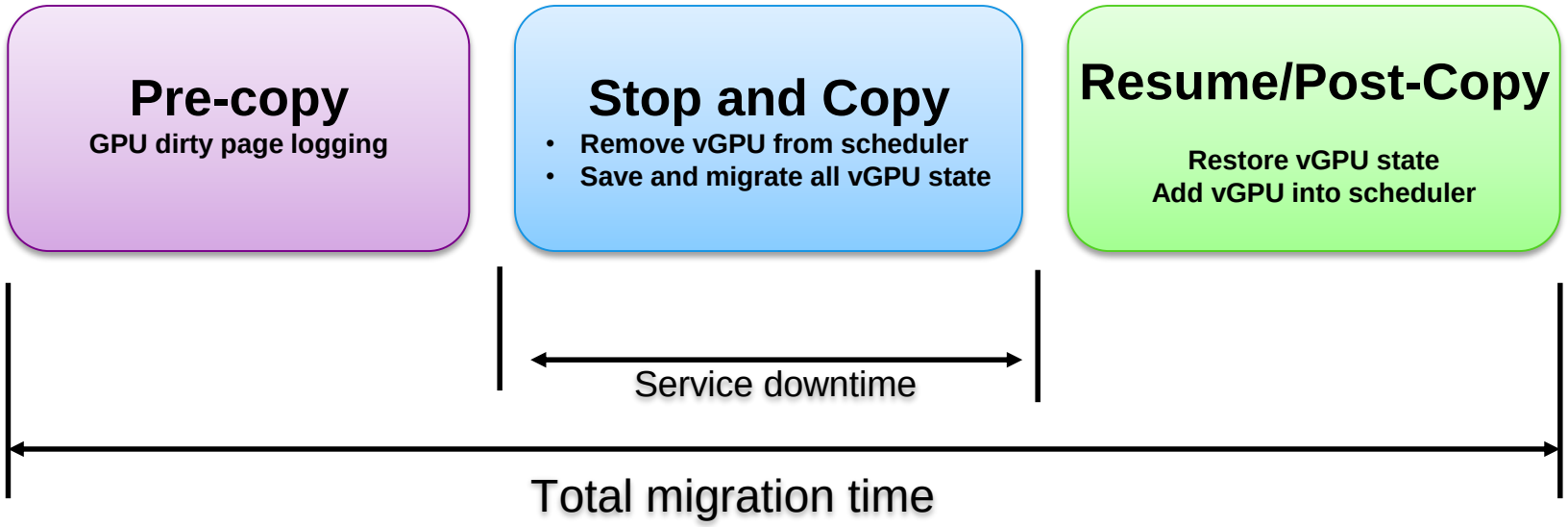
Trap and emulate for privileged resource

Time-shared among VMs



Challenge of Migrating vGPU Instance

- When and how to migrate Graphics Memory
- When and how to migrate Guest Graphics Page Table
- When and how to migrate Render Engine State



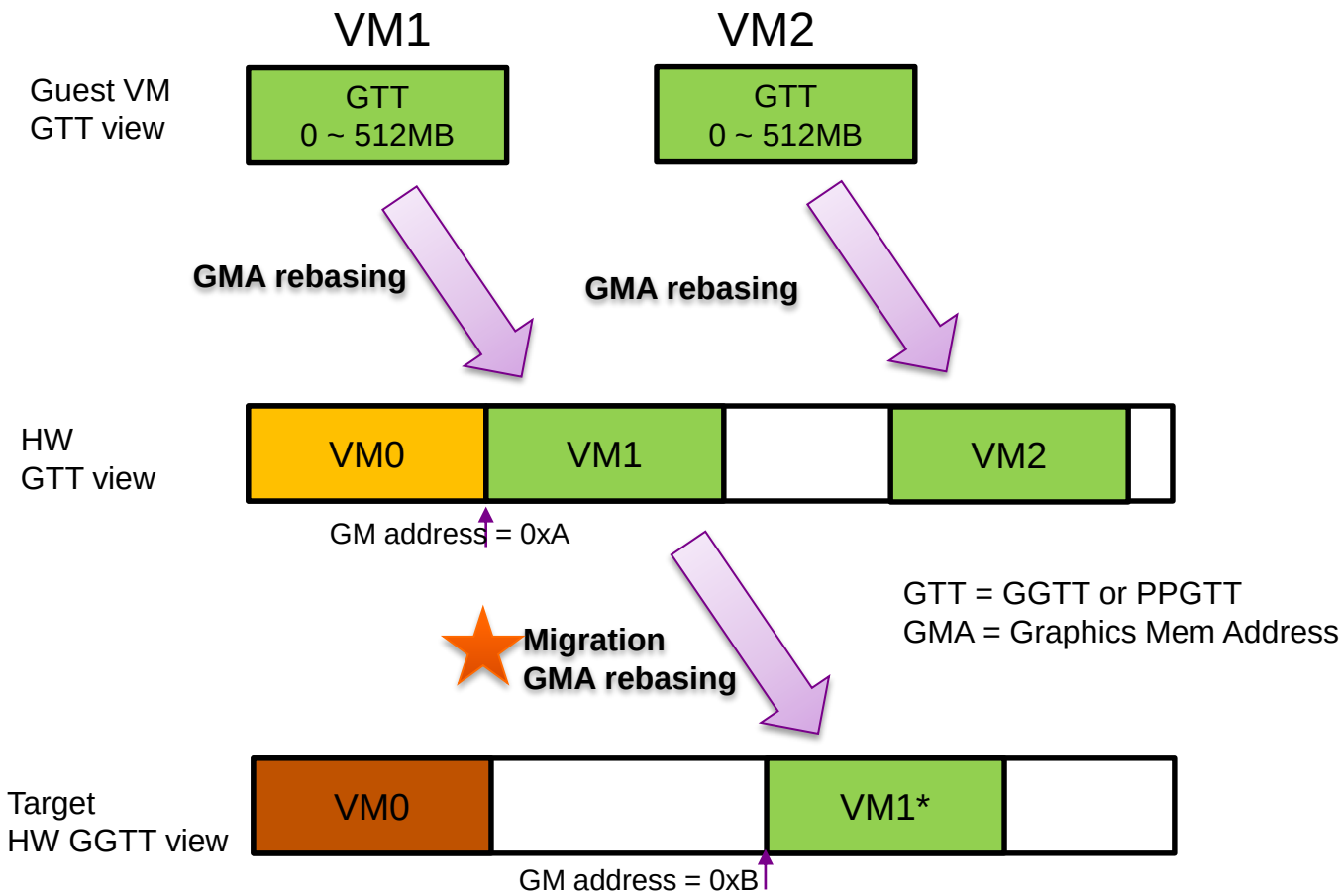
Migration Policies for Different vGPU Resources

Registers	Copy and Restore
GTT page table	Recreate Shadowing
Graphics Memory	Track Dirty and Copy
Context: Render Engine	Recreate Shadowing

Agenda

- GPU Virtualization and vGPU Live Migration
- vGPU Resources
- Design and Solution
- Current Status
- Summary

Guest GTT Page Table Migration



- Both GGTT and PPGTT are shadowed for Guest
- GGTT required rebasing due to GGTT partition among VMs
- Migration process actually:
  - A. Copy entire Guest GTT page table
  - B. Re-create the shadow page table for Guest on Target side
  - C. Rebasing GGTT for GPU commands

Graphics Memory Address rebasing:

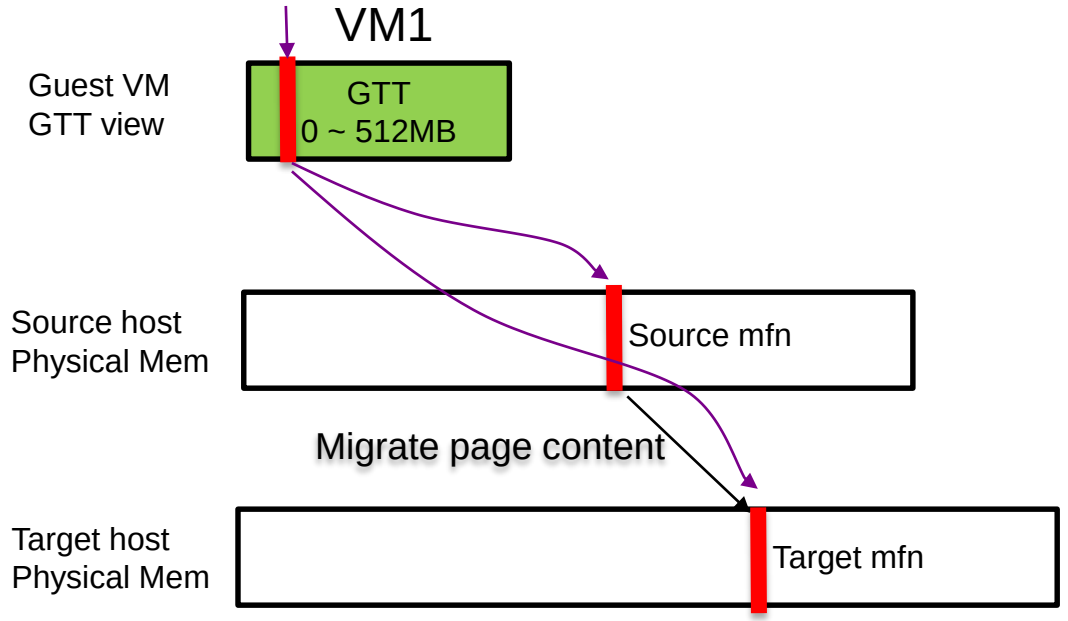
All vGPU cmds from Guest need to be rebased on new address in GVT-g before send to real GPU HW

Guest Graphics Memory Migration

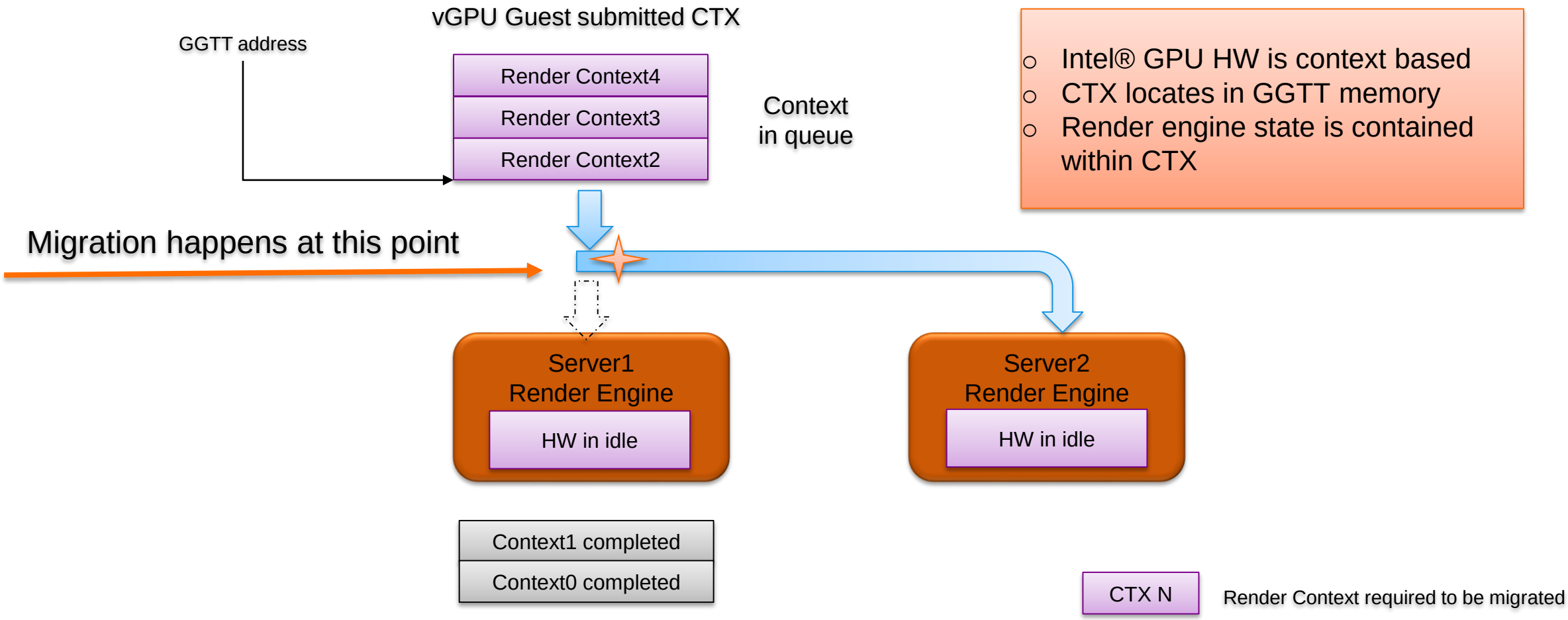
- **Pre-copy:** Logging dirty graphics memory pages
- **Stop-and-Copy:** Migrate contents to target
- **Resume/Post-copy:** Recreate GTT page table based on target mfn

Problem:
Intel® GPU page table entities has no Dirty or Accessed flags to track dirty pages

Solution:
Copy all used graphics memory to target.



Render Engine State Migration



- Intel® GPU HW is context based
- CTX locates in GGTT memory
- Render engine state is contained within CTX

Agenda

- GPU Virtualization and vGPU Live Migration
- vGPU Resources
- Design and Solution
- Current Status
- Summary

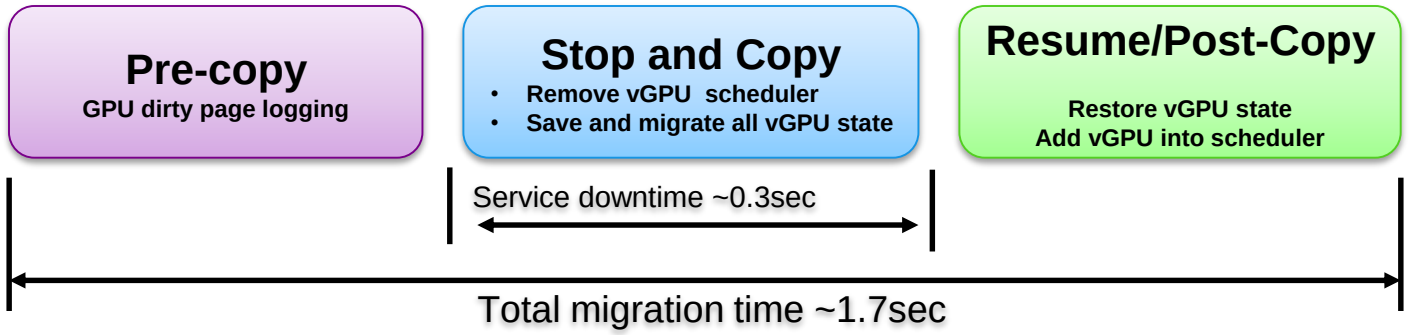
Current Status

- Experimental support both KVMGT and XENGT
- Platforms: Intel® 5th /6th Generation Intel® Core™ Processors
- Benchmarks covered:

  Windows guest: Heaven, 3Dmark06, Trophic, Media encoding/decoding, Linux guest: lightsmark, 2D
- Quality: 12hours overnight testing, migrating every 30sec
- Timing: (Guest RAM 2GB including 512MB Graphics memory, 10Gbps adapter)

  Service downtime ~0.3sec

  Total migration time: ~1.7sec



Summary

- Need 3D/2D/Media workload in virtualization?

  GVT-g is the choice

- Need GPU virtualization with migration support?

  GVT-g is the choice ☺

Resource Links

- Project webpage and release: https://01.org/igvt-g
- Project public papers and document: https://01.org/group/2230/documentation-list
- Intel® IDF: GVT-g in Media Cloud: https://01.org/sites/default/files/documentation/sz15_sfts002_100_engf.pdf
- XenGT introduction in summit in 2015: http://events.linuxfoundation.org/sites/events/files/slides/XenGT-Xen%20Summit-REWRITE%203RD%20v4.pdf
- XenGT introduction in summit in 2014: http://events.linuxfoundation.org/sites/events/files/slides/XenGT-LinuxCollaborationSummit-final_1.pdf

Legal Notices and Disclaimers

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer.

No computer system can be absolutely secure.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit http://www.intel.com/performance.

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

Statements in this document that refer to Intel's plans and expectations for the quarter, the year, and the future, are forward-looking statements that involve a number of risks and uncertainties. A detailed discussion of the factors that could affect Intel's results and plans is included in Intel's SEC filings, including the annual report on Form 10-K.

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

Intel, Core and the Intel logo are trademarks of Intel Corporation in the United States and other countries.

OpenCL and the OpenCL logo are trademarks of Apple Inc. used by permission by Khronos.

*Other names and brands may be claimed as the property of others.

© 2016 Intel Corporation.