gScale : Improve vGPU Scalability
Using Dynamic Resource Sharing

Aug 2016

Pei Zhang Pei.zhang@intel.com

Kevin Tian Kevin.tian@intel.com

Xiao Zheng Xiao.zheng@intel.com

Agenda

- vGPU Scalability of GVT-g
- gScale Design for Doubled vGPU Density
- Evaluation
- Summary

vGPU Scalability in GVT-g

GVT-g Introduction

- Full GPU virtualization solution with mediated pass-through approach
- Open source implementation for Xen/KVM (aka XenGT/KVMGT)
  - Support a rich span of Intel® Processor Graphics
  - Available in https://01.org/igvt-g

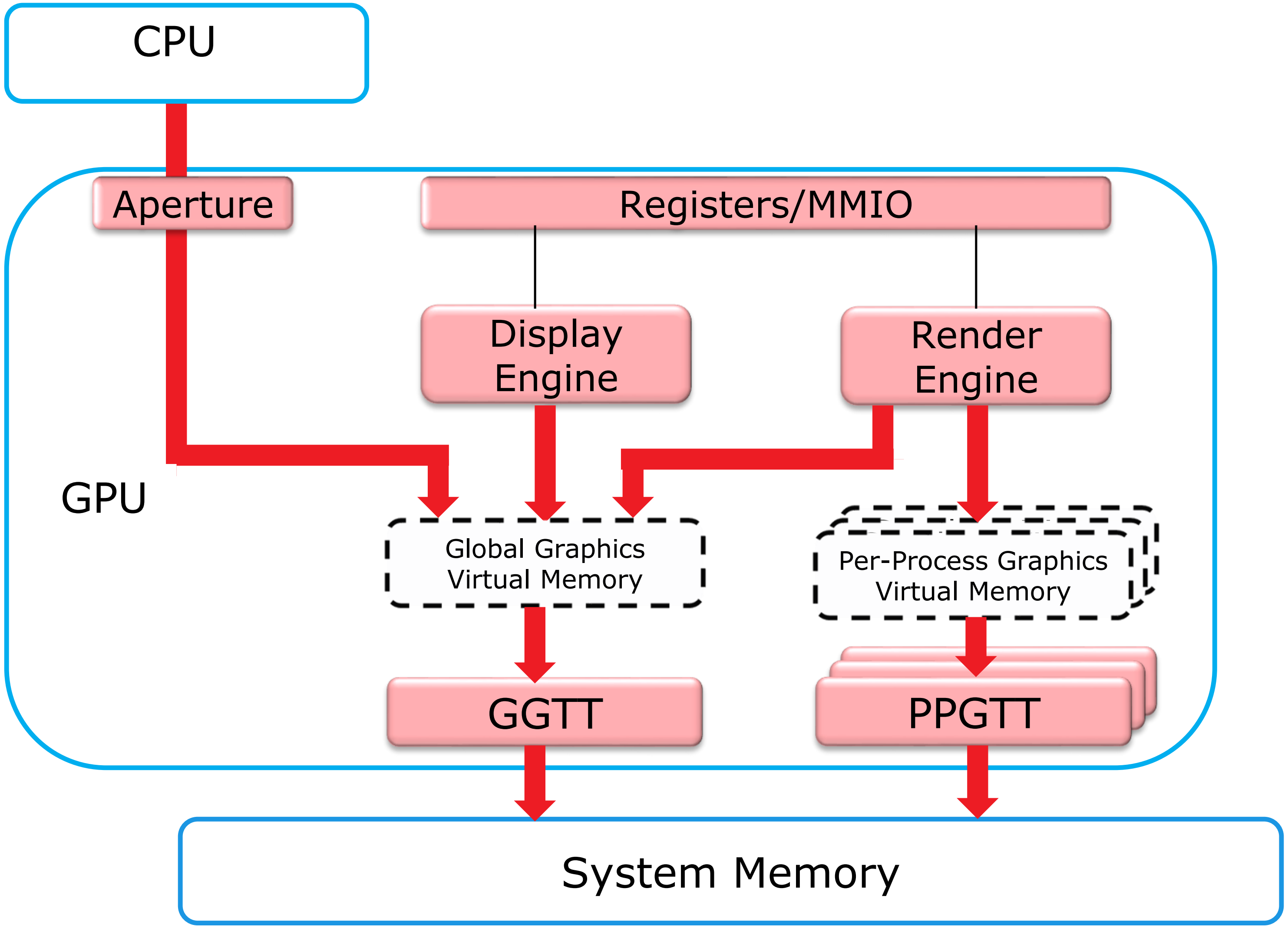
Near to native GPU performance + **Full GPU capabilities in VM** + **Flexible VMs sharing (Up to 7VMs)**

gScale

>= 2X higher density!

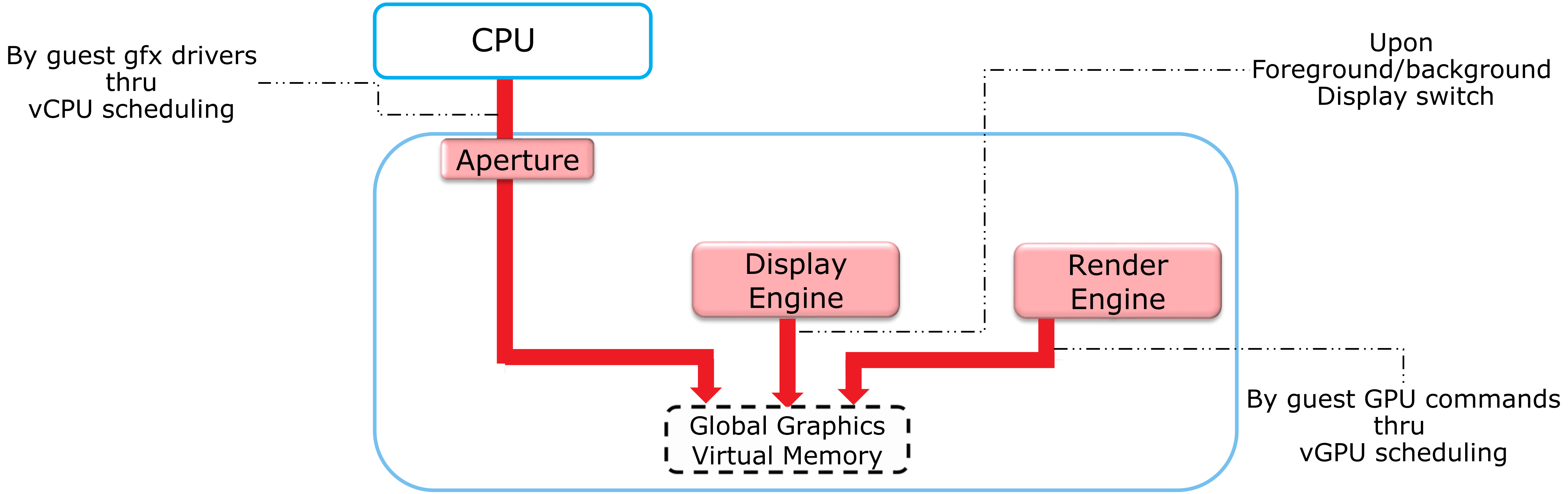
- *Intel and the Intel logo are trademarks of Intel Corporation or its subsidiaries in the U.S. and/or other countries.*
- *For more complete information about performance and benchmark results, visit www.intel.com/benchmarks.*

Processor Graphics: Components

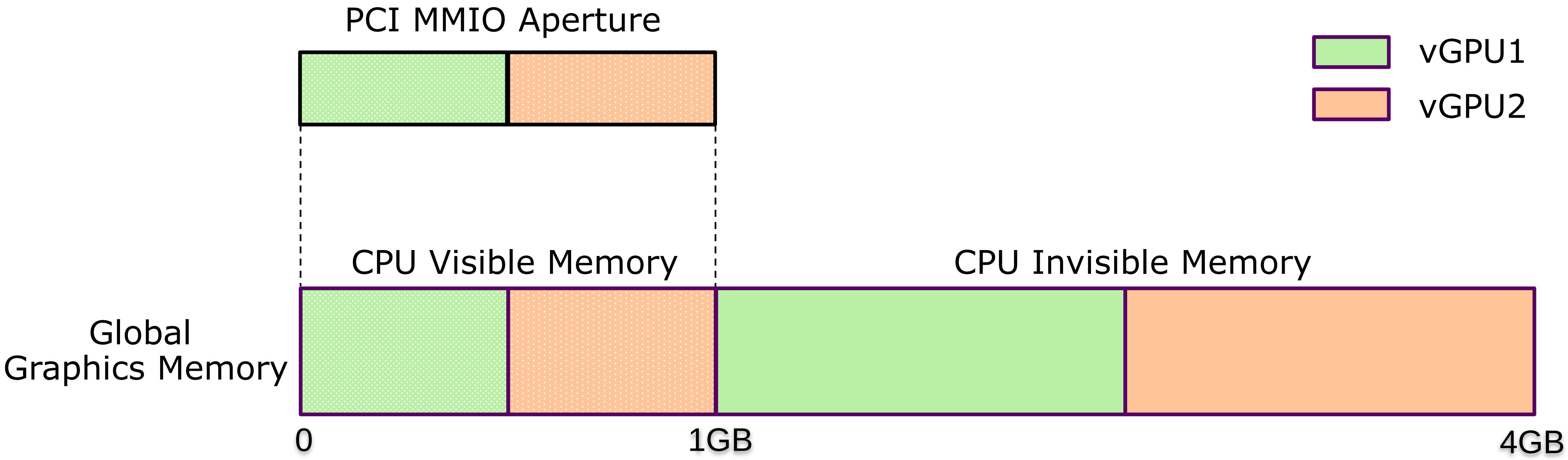


Shared Global Graphics Memory in GVT-g

- Parallel access from multiple engines due to split vCPU/vGPU scheduling



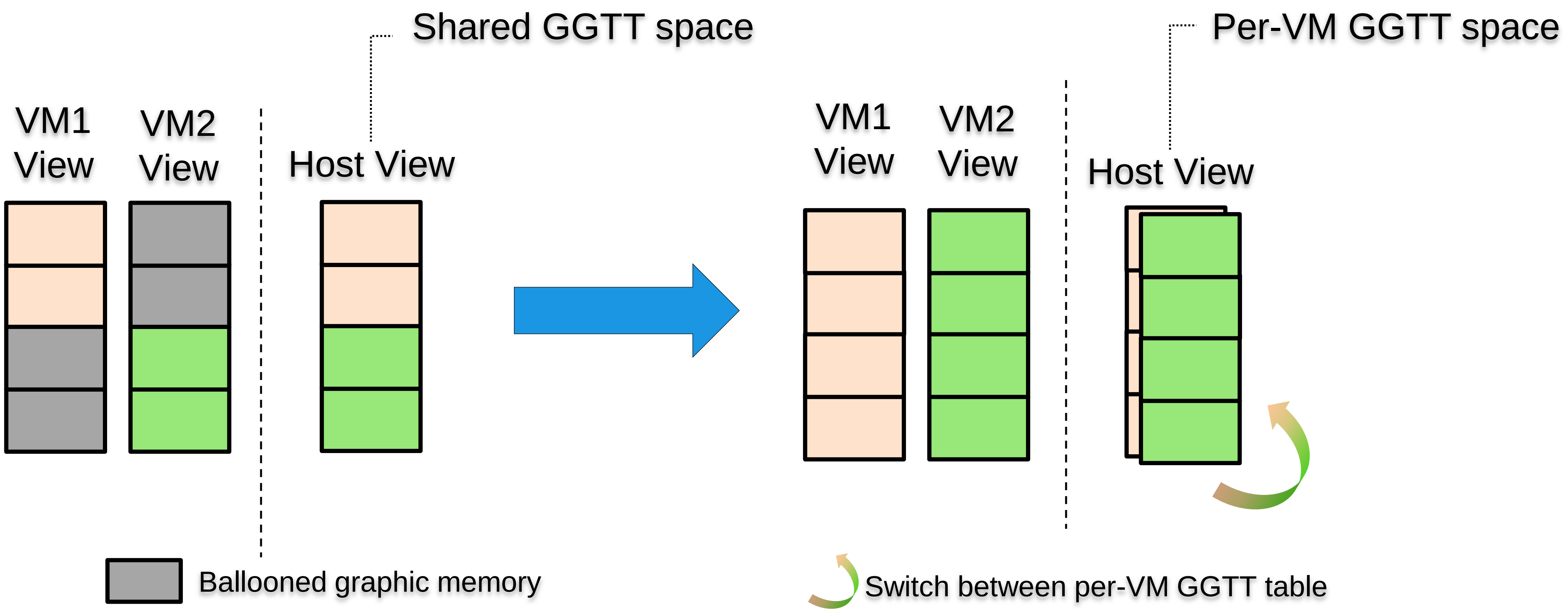
Static Partitioning of Global Graphics Memory



- Minimal 128MB visible/384MB invisible per vGPU
- Means up to 7vGPUs possible (besides reserved for Dom0)
- Some BIOS may support a smaller aperture which means more limitation

gScale Design for Doubled vGPU Density

gScale: per-VM Global Graphics Memory



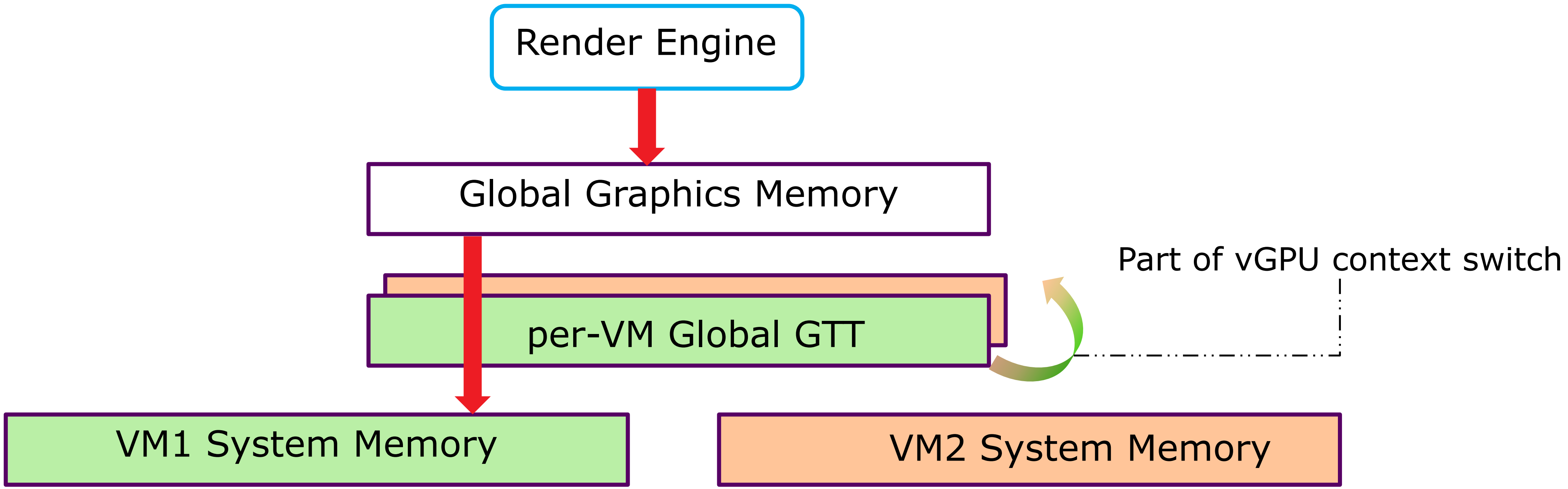
Key challenge is to remove parallel accesses!

3 Components in Parallel Access to GGTT

- Render engine access
- Display engine access
- CPU tiled/non-tiled memory access

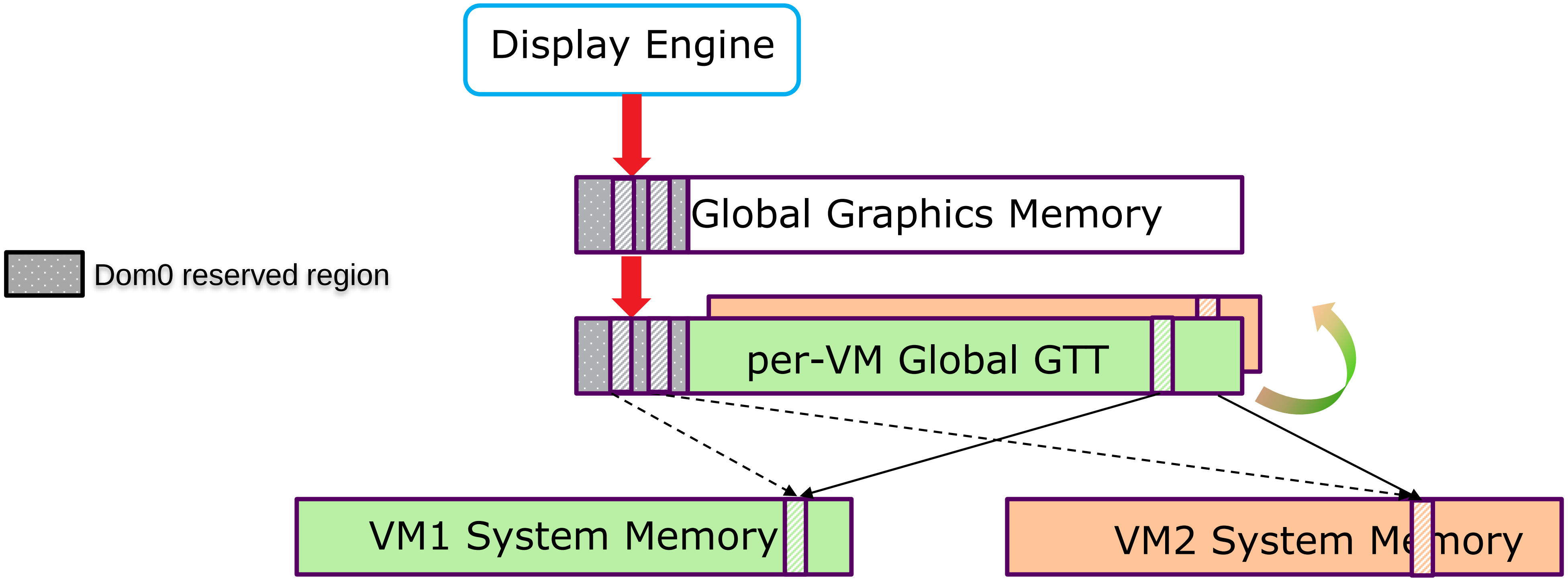
Render Engine Accesses

- At anytime, only one vGPU's graphics memory is accessible by render engine
  Controlled by vGPU scheduler
- Dynamically switch per-VM GGTT table at vGPU context switch



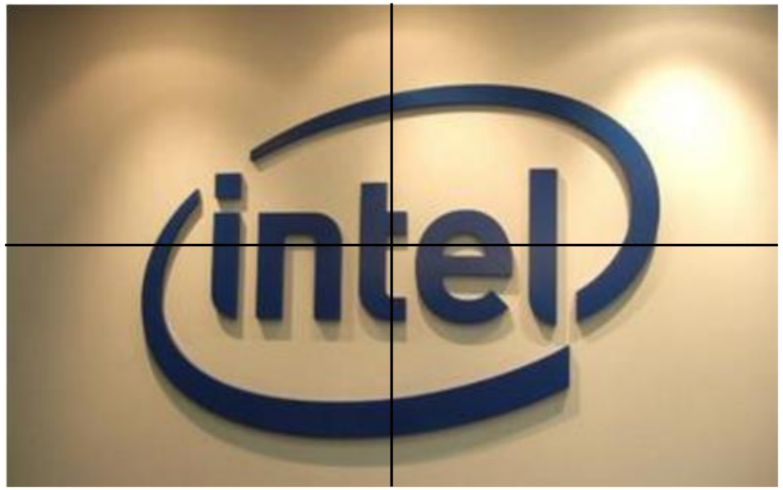
Display Engine Accesses

- Display engine accesses is restricted to the Dom0 reserved region
  Alias mapping of guest framebuffer into reserved graphics memory for Dom0



Tiled/non-Tiled and Aperture Memory

- What is tiled memory
- Tiled memory – aperture access



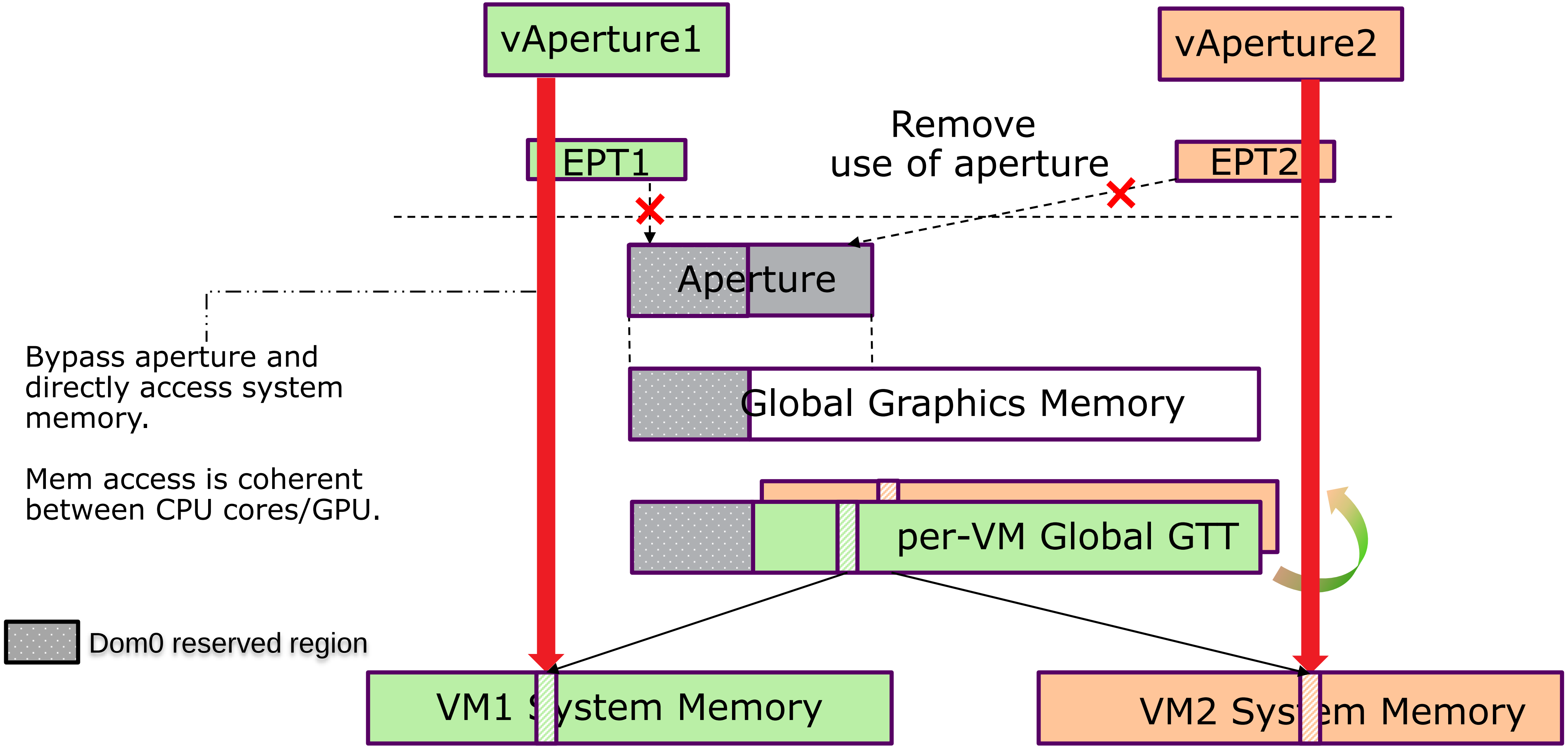
Linear surface

Tiled surface

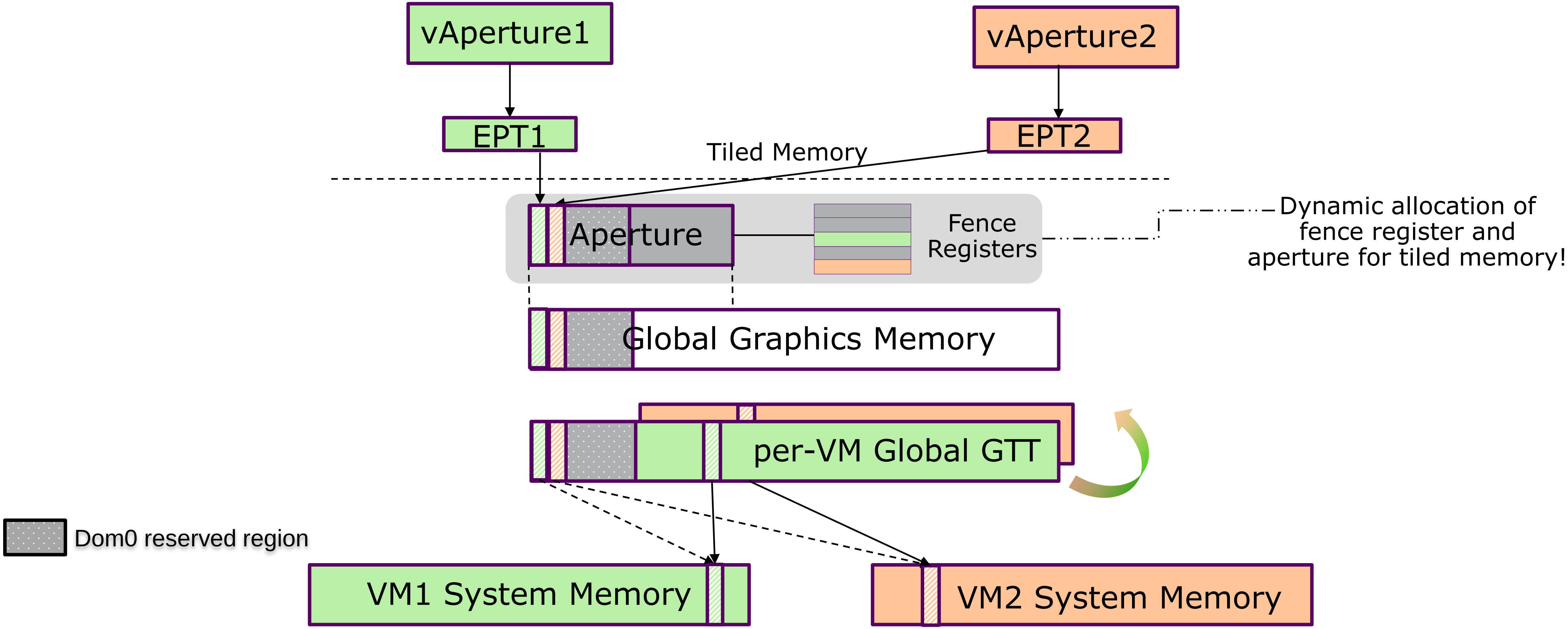
CPU Accesses to Non-Tiled Memory

Bypass aperture and directly access system memory.

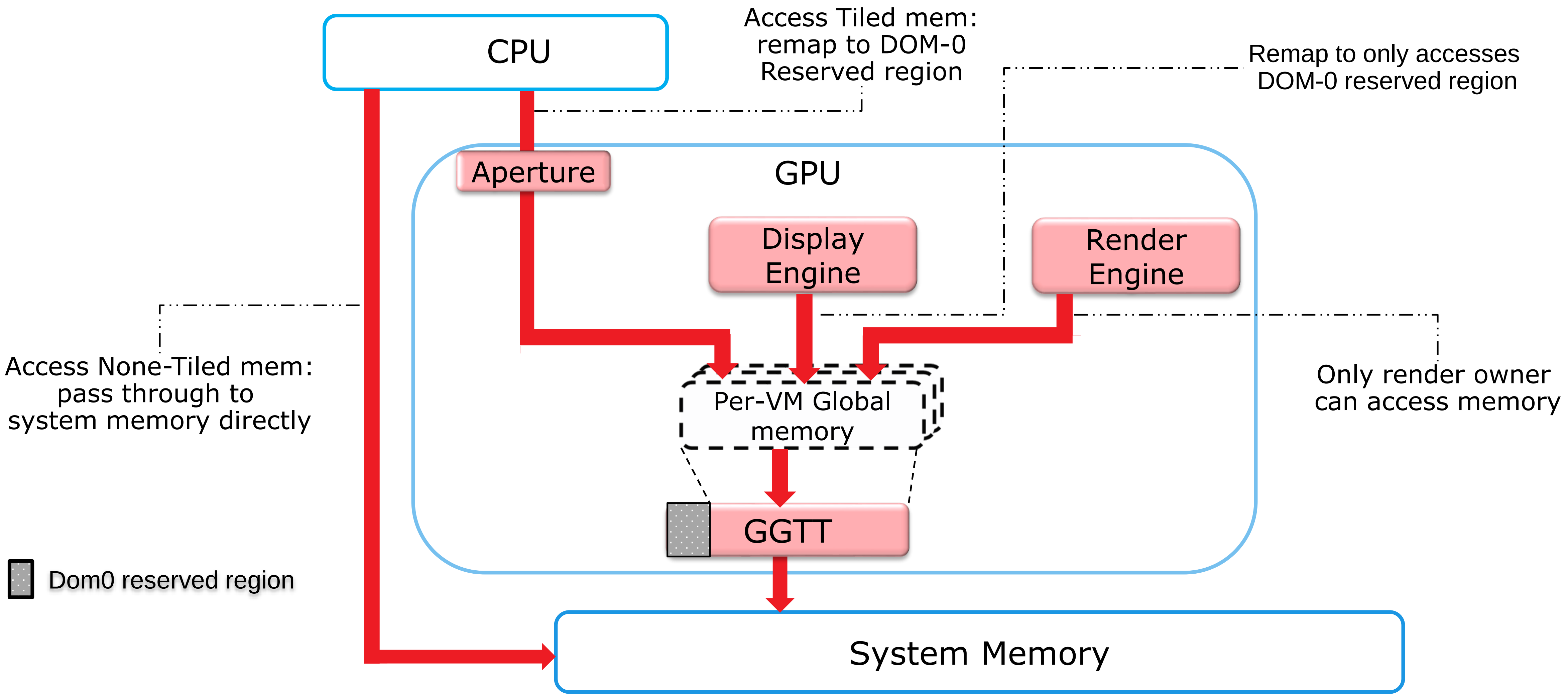
Mem access is coherent between CPU cores/GPU.



CPU Access for Tiled Memory



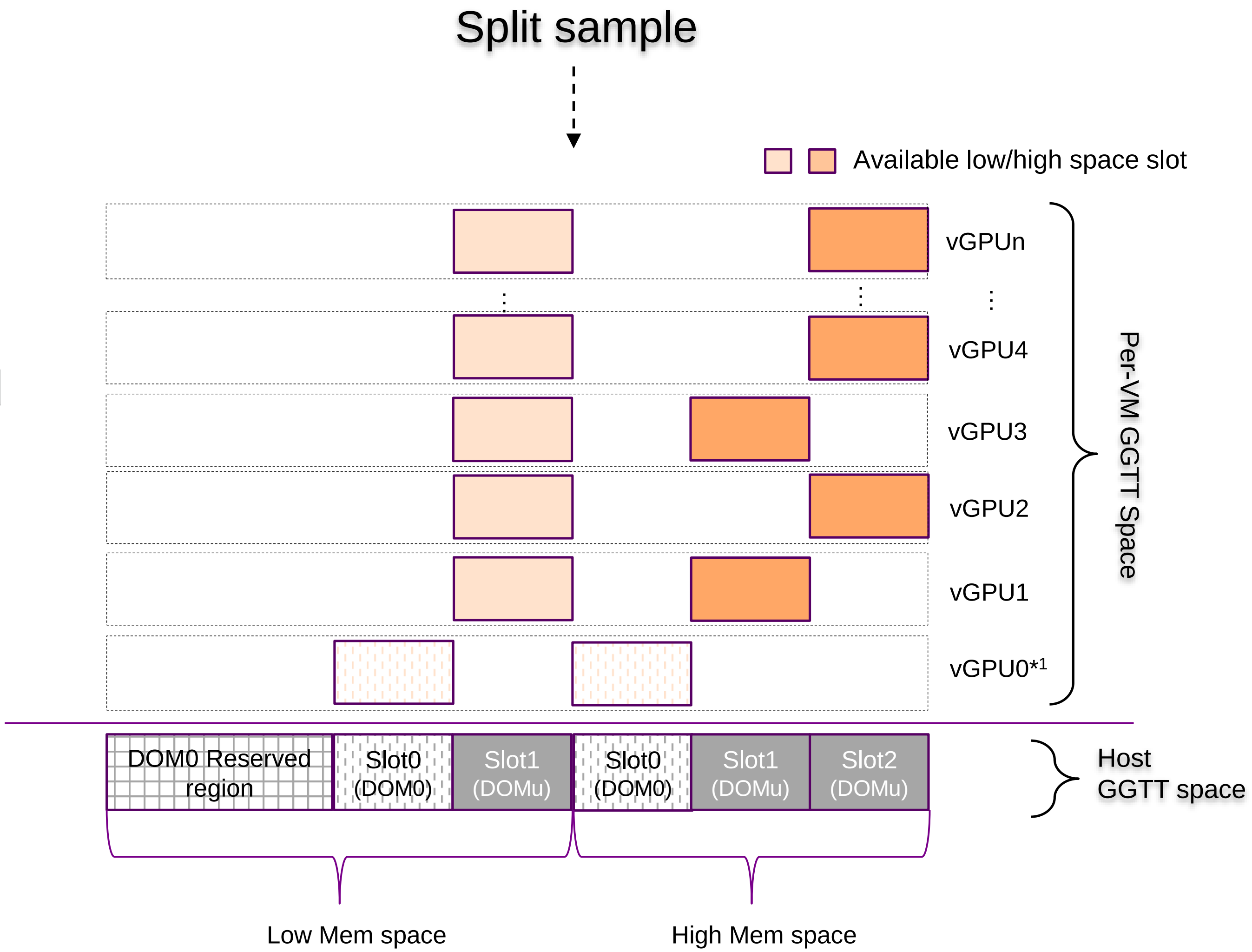
Summary: Global Graphic memory access



Optimization: Split per-VM GGTT into slots

Whole per-VM GTT table size is about 2MB
Switching entire GTT table is time consumed.

- per-VM GTT switching is based on slots
- Switching for those only slots that shared between other VMs

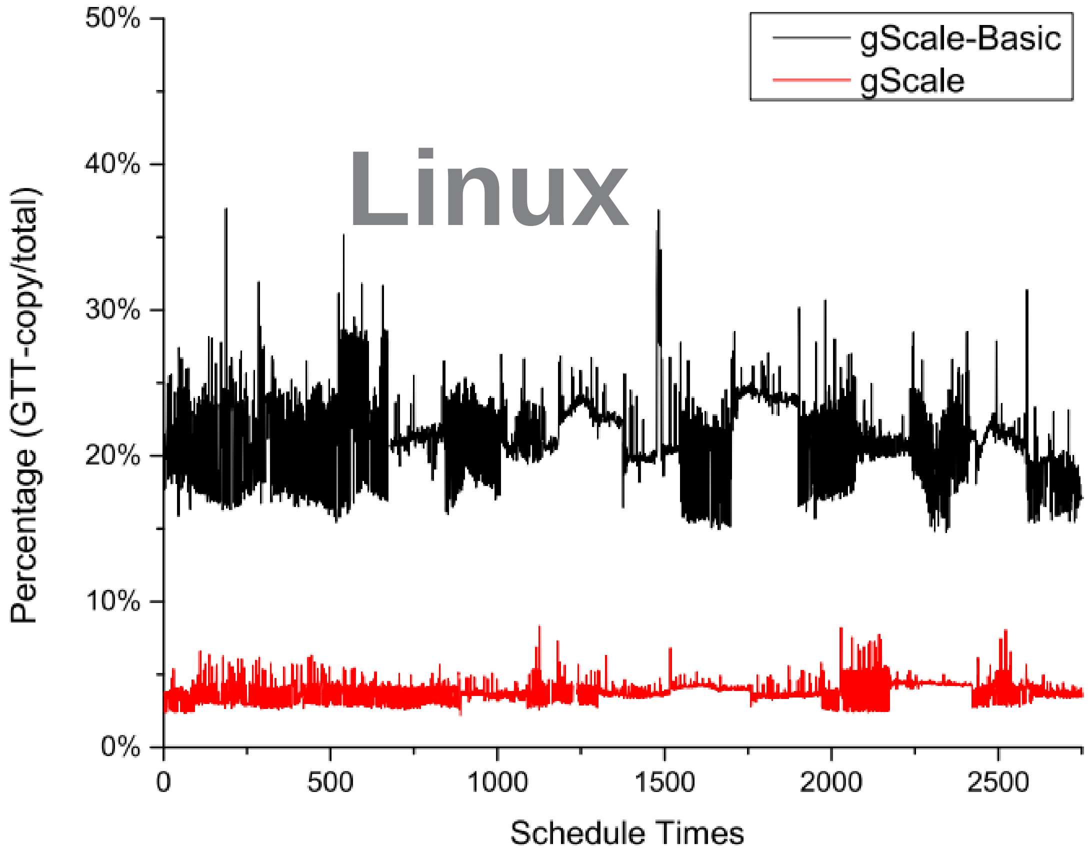


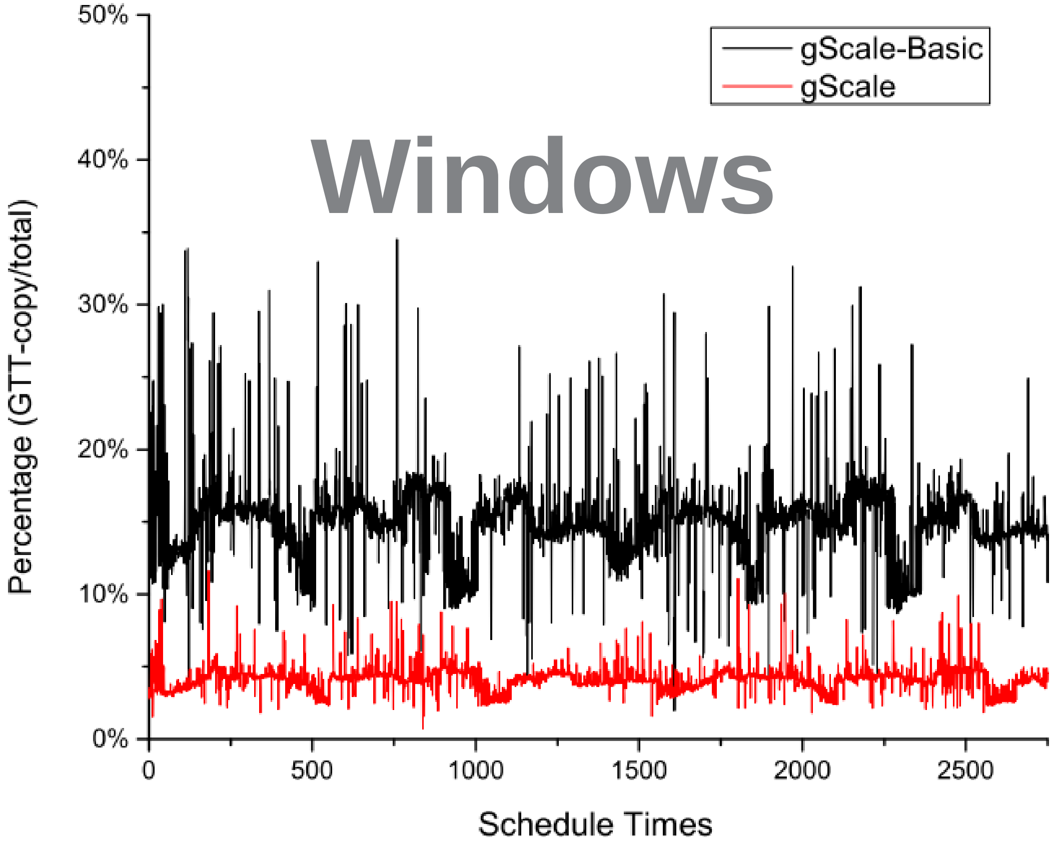
*** Notes:**

1. vGPU0 (for DOM0) has dedicated mem space slots which won't be shared by other vGPUs.

Context switch performance
- Private GTT table copy

1. *gScale-Basic* is the data without memory slot split.
2. Performance data is sampled with 12 VMs.

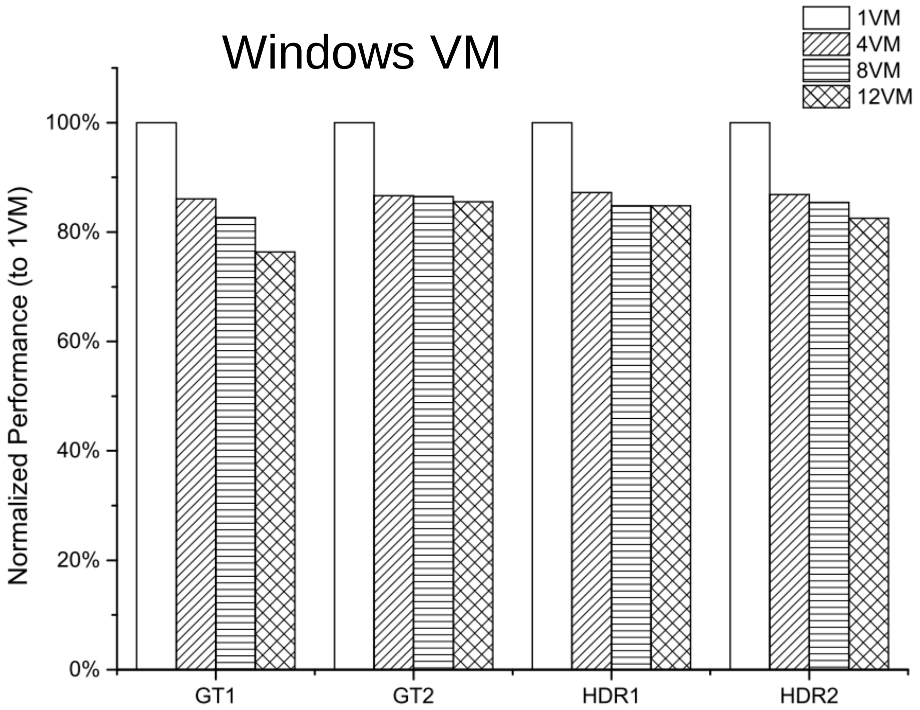


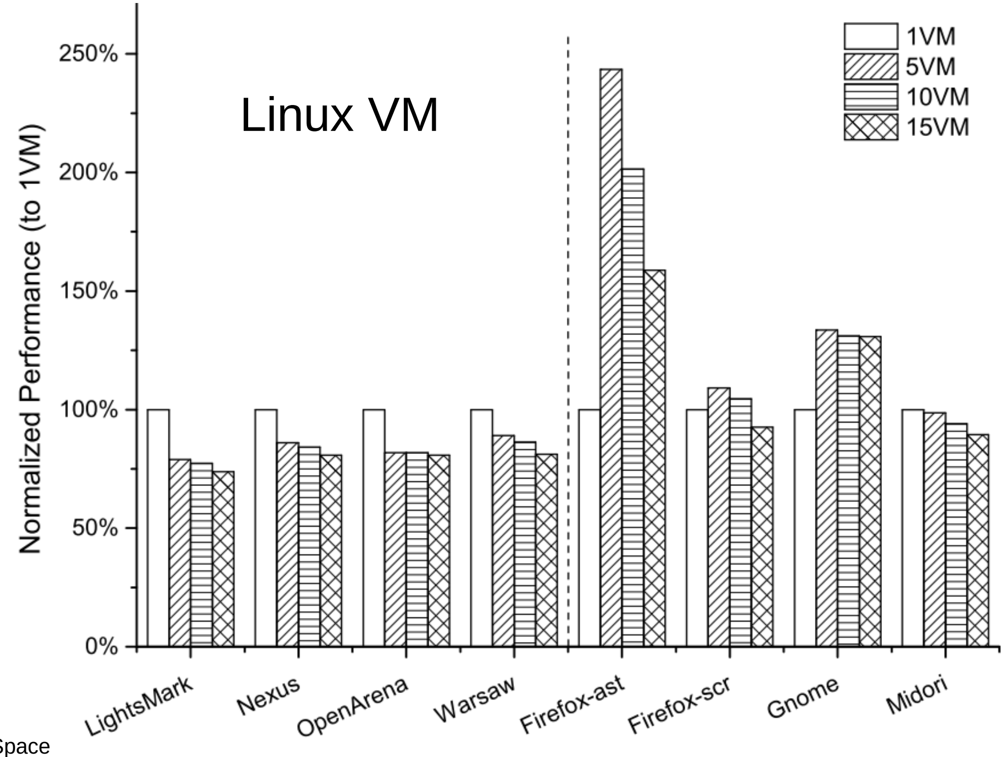


Source: USENIX ATC (2015), gScale: Scaling up GPU Virtualization with Dynamic Sharing of Graphics Memory Space

Evaluation

Scalability Evaluation





Source: *USENIX ATC* (2015), gScale: Scaling up GPU Virtualization with Dynamic Sharing of Graphics Memory Space

Windows Guest: Scalability 3D performance with 12 VMs could achieves up to 80% of 1 VM.

Linux Guest: Scalability of 3D performance with 15VMs could achieves up to 80% of 1 VM.
2D performance is better by burn out GPU power.

Summary

Summary

- gScale breaks graphics memory resource limitation by introducing per-VM GGTT design
- gScale doubles vGPU density in Intel®@ GVT-g with good scalability
- Converged performance of 15 vGPUs reaches 96% of native performance

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

- Slide4 is measured by computing the VM creation number in HSW platform.
- Slide18 is measured by computing the context switch time.
- Slide20 is measured by running different 3D workload and compute the total FPS.
- Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.
- All test in this document are with same hardware configuration: CPU Intel E3-1285 v3(4Cores, 3.6GHz), GPU Intel HD Graphics P4700, Memory 32GB, Storage SAMSUNG 850Pro 256GB*3. Test is made by Xue, Mochi.
- For more information go to http://www.intel.com/performance.