LINUTRONIX

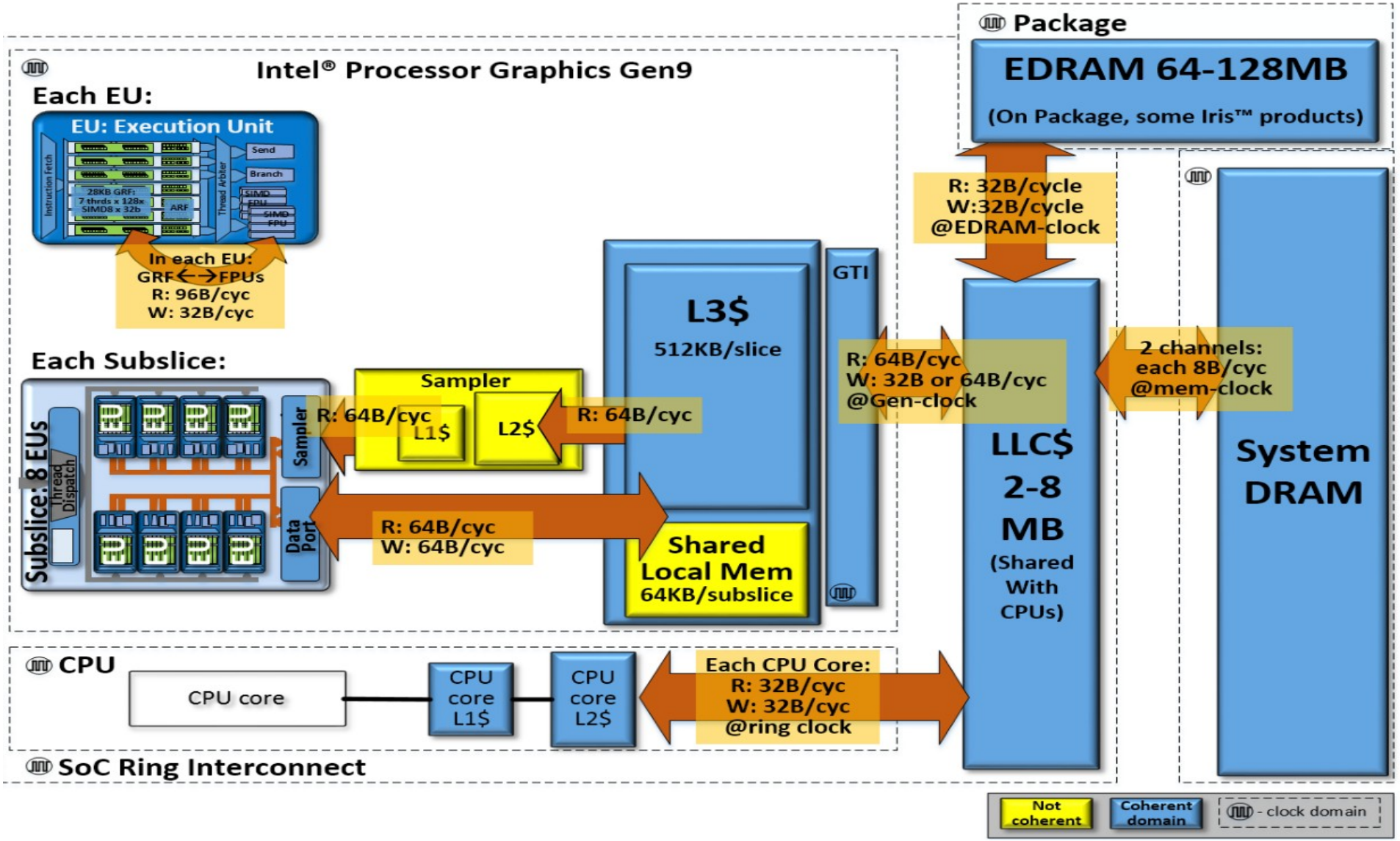
Realtime and Graphics – a Contradiction in Terms?

electronic displays Conference 2021
01. - 05. March 2021

Poster Session

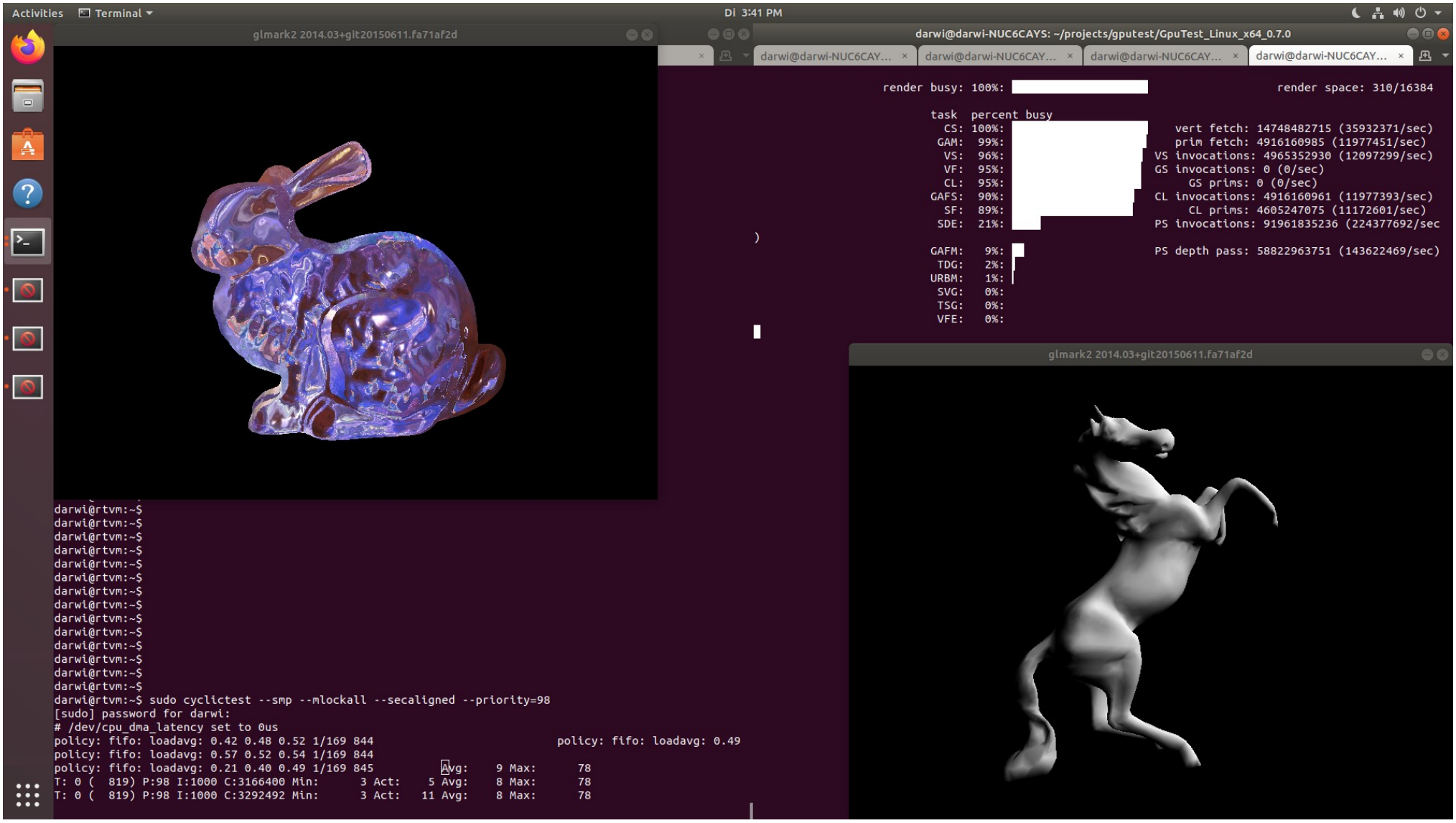
Ahmed S. Darwish <a.darwish@linutronix.de>

Problem: GPUs share resources with CPUs



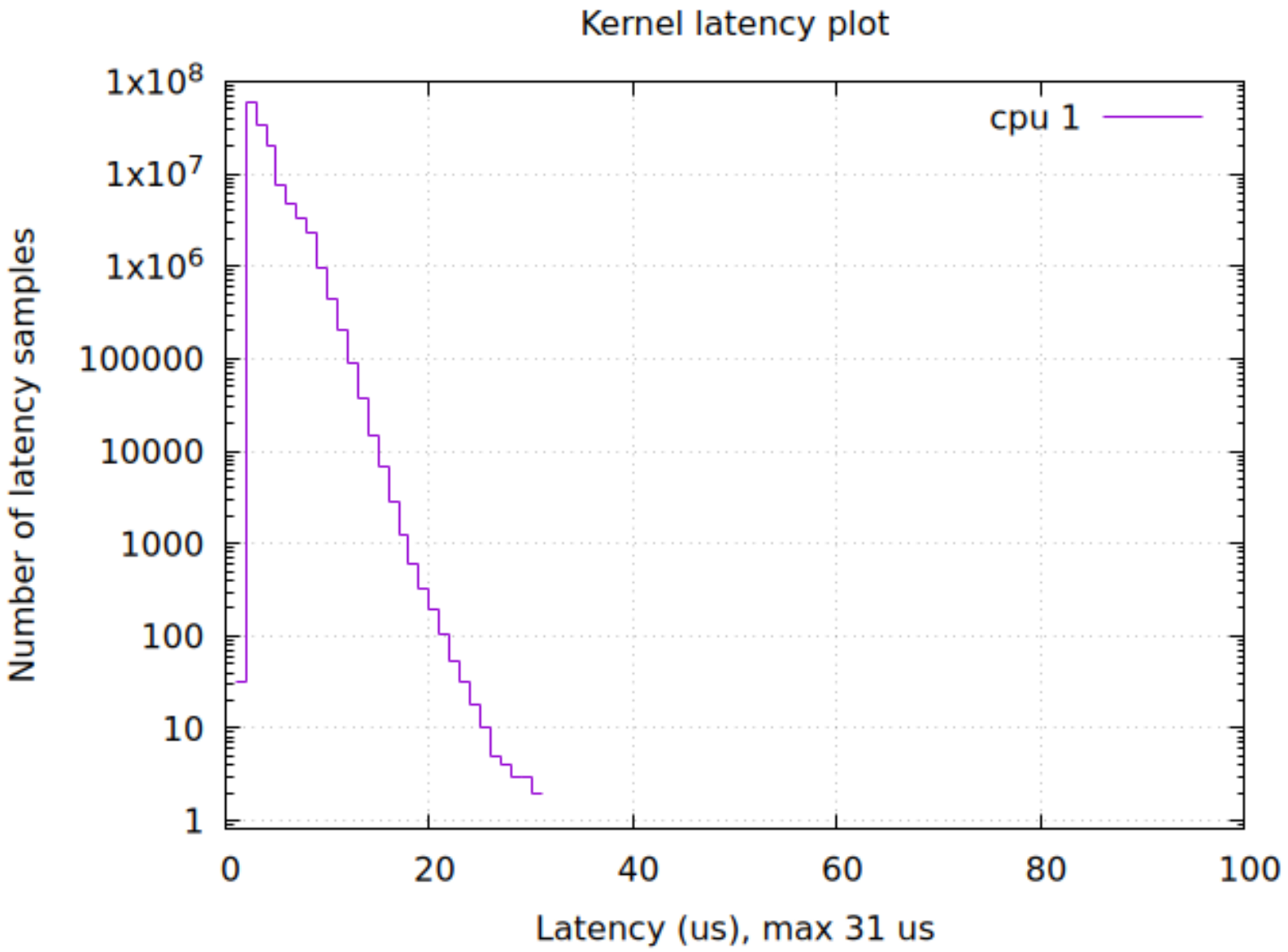
Source: "The Compute Architecture of Intel Processor Graphics Gen9", Intel

Noisy Neighbours: even with VM Isolation

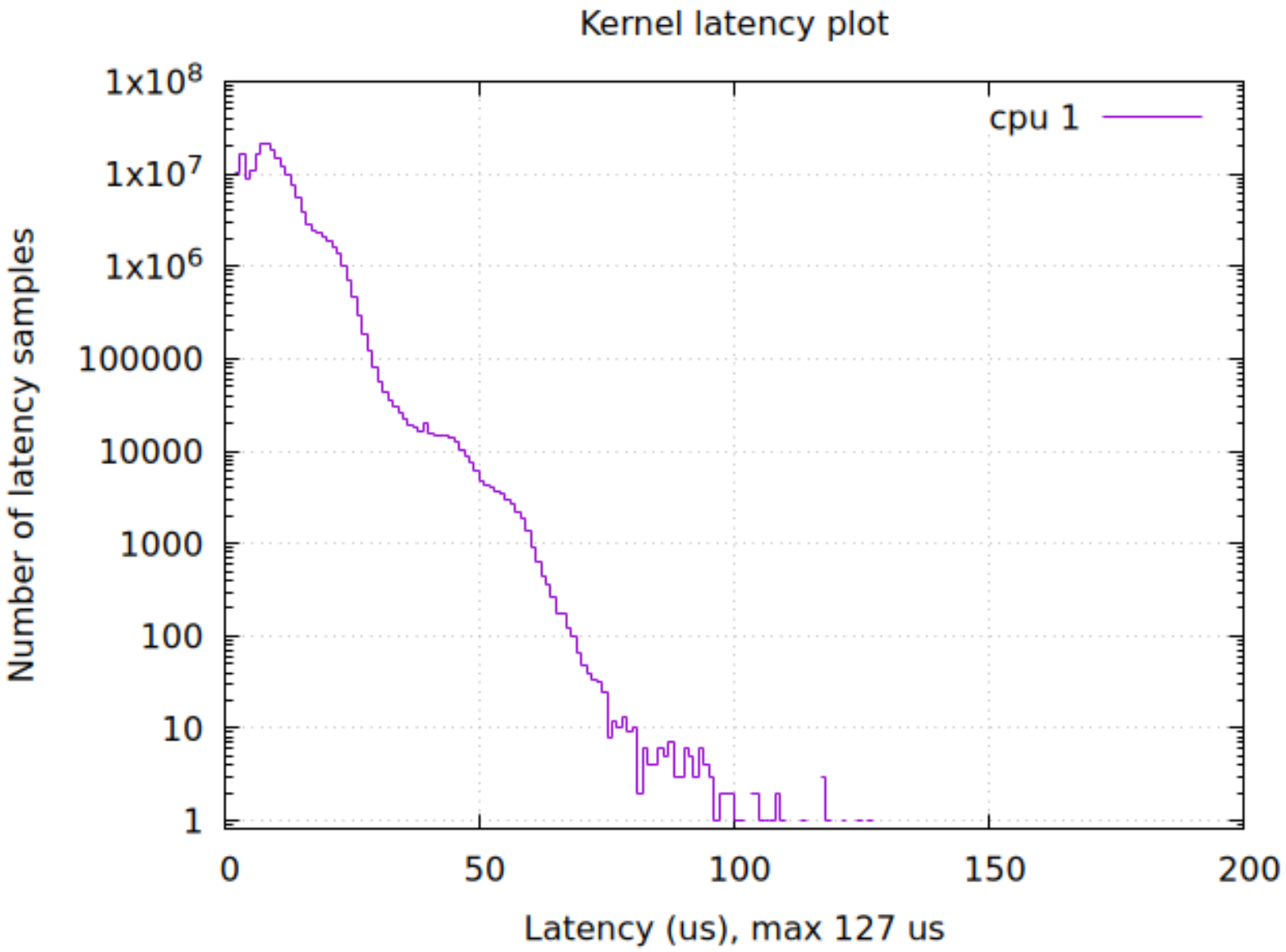


Bottom Left: RT VM running over the ACRN Hypervisor + cyclictest latency measurement. Top right: GPU usage. Note: Graphical applications are running outside the RT virtual machine.

RT VM Latency: No External GPU Workload



RT VM Latency: with External GPU Workload



Segmenting Caches: Intel Cache Allocation Tech

	M7	M6	M5	M4	M3	M2	M1	M0
COS0	A	A	A	A	A	A	A	A
COS1	A	A	A	A	A	A	A	A
COS2	A	A	A	A	A	A	A	A
COS3	A	A	A	A	A	A	A	A

Default Bitmask

	M7	M6	M5	M4	M3	M2	M1	M0
COS0	A	A	A	A	A	A	A	A
COS1					A	A	A	A
COS2							A	A
COS3								A

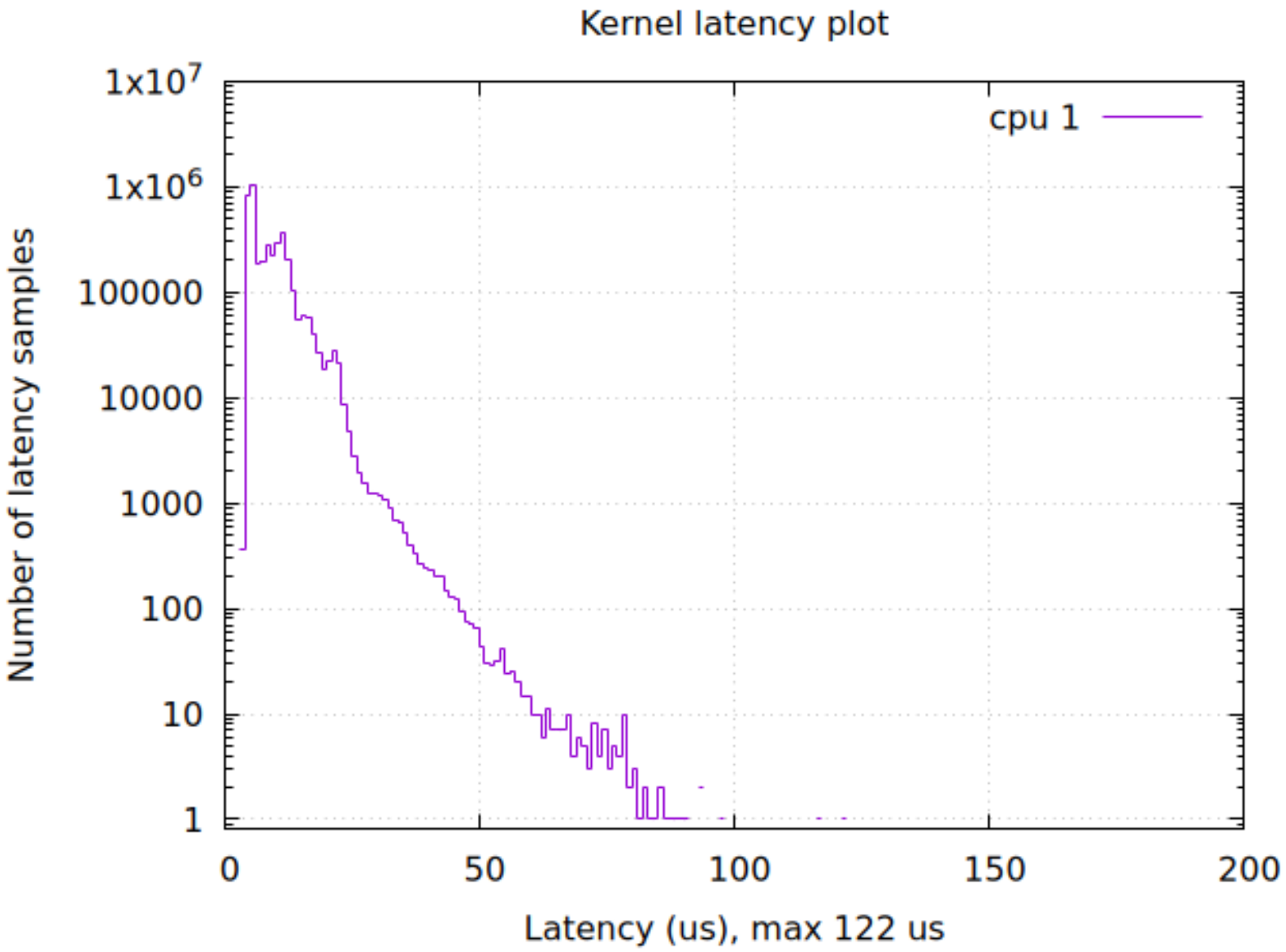
Overlapped Bitmask

	M7	M6	M5	M4	M3	M2	M1	M0
COS0	A	A	A	A				
COS1					A	A		
COS2							A	
COS3								A

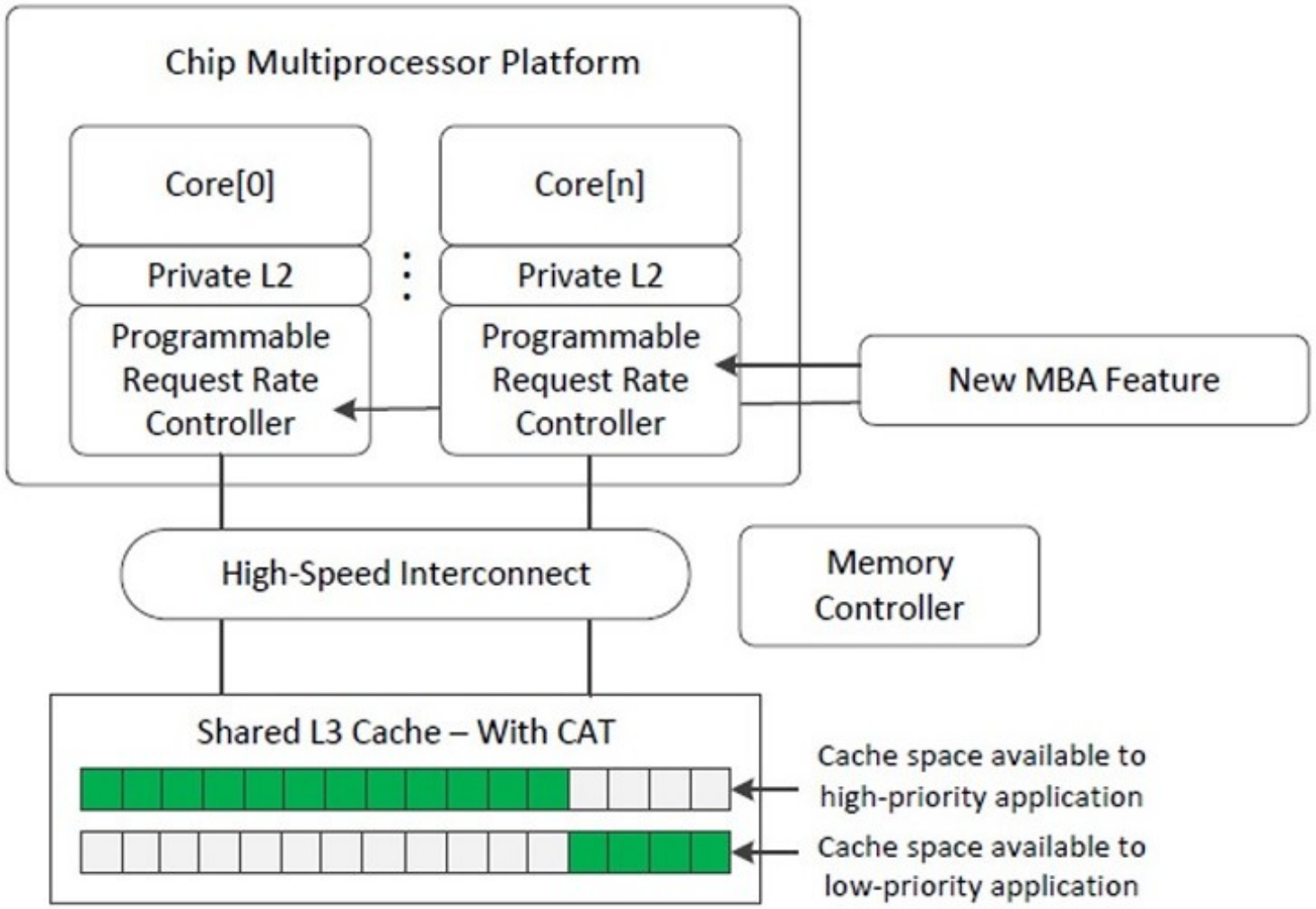
Isolated Bitmask

Source: Intel® 64 and IA-32 Architectures Software Developer's Manual. Vol. 3A. Section 17.19.2 Cache Allocation Technology Architecture

RT VM Latency: External GPU Workload + CAT



Segmenting Memory Bandwidth: Intel MBA



Source: Intel® 64 and IA-32 Architectures Software Developer's Manual. Vol. 3A. Section 17.19.7 Introduction to Memory Bandwidth Allocation

Questions / Comments

Thank you for your attention.

a.darwish@linutronix.de
info@linutronix.de