LOW-LATENCY GPGPU

A 5-minute intro and investigation

Matheus Vitti Santos
@ Meeting C++ 2019

Disclaimer

- These findings reflect the point of view of someone who's been courting only CUDA in a hobbyist setting since 2010, and in a (lightweight) professional setting since 2017

- I'd love to hear the viewpoints of AMD, Intel, Direct3D, Metal, and Vulkan folks on this - hit me up afterwards!
  - *And maybe give me some hardware to play with...?*

So, GPGPUs have latency issues...

- Calling GPU functions **takes time**
- Moving memory around **takes a lot of time**
- *The GPU Driver* **takes its sweet, sweet time**

- ...right?

Matheus Vitti Santos
@ Meeting C++ 2019

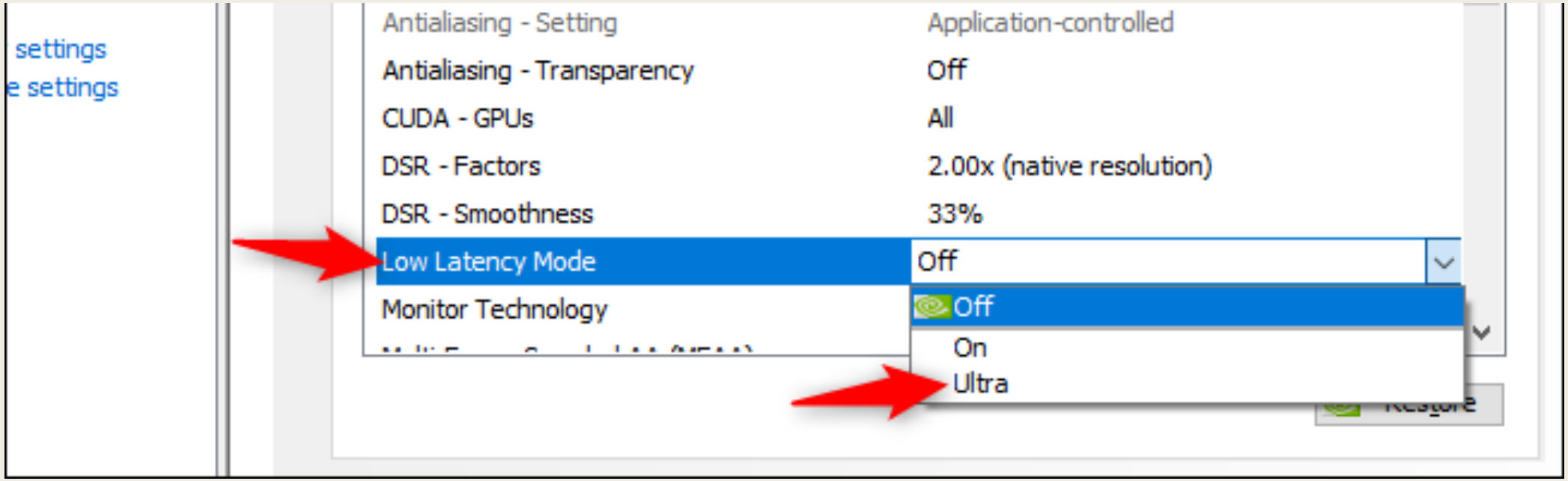
Matheus Vitti Santos
@ Meeting C++ 2019

10+ years of GPGPU

- APIs are slimmer!
  - *(D3D12 / Vulkan / Metal vs. OpenGL/D3D9)*
- GPUs are faster!
- PCI Express is faster!

- Most of all, **Drivers are faster**!

Literally 2 weeks ago: NVIDIA introduces “Ultra Low Latency Mode”



Matheus Vitti Santos
@ Meeting C++ 2019

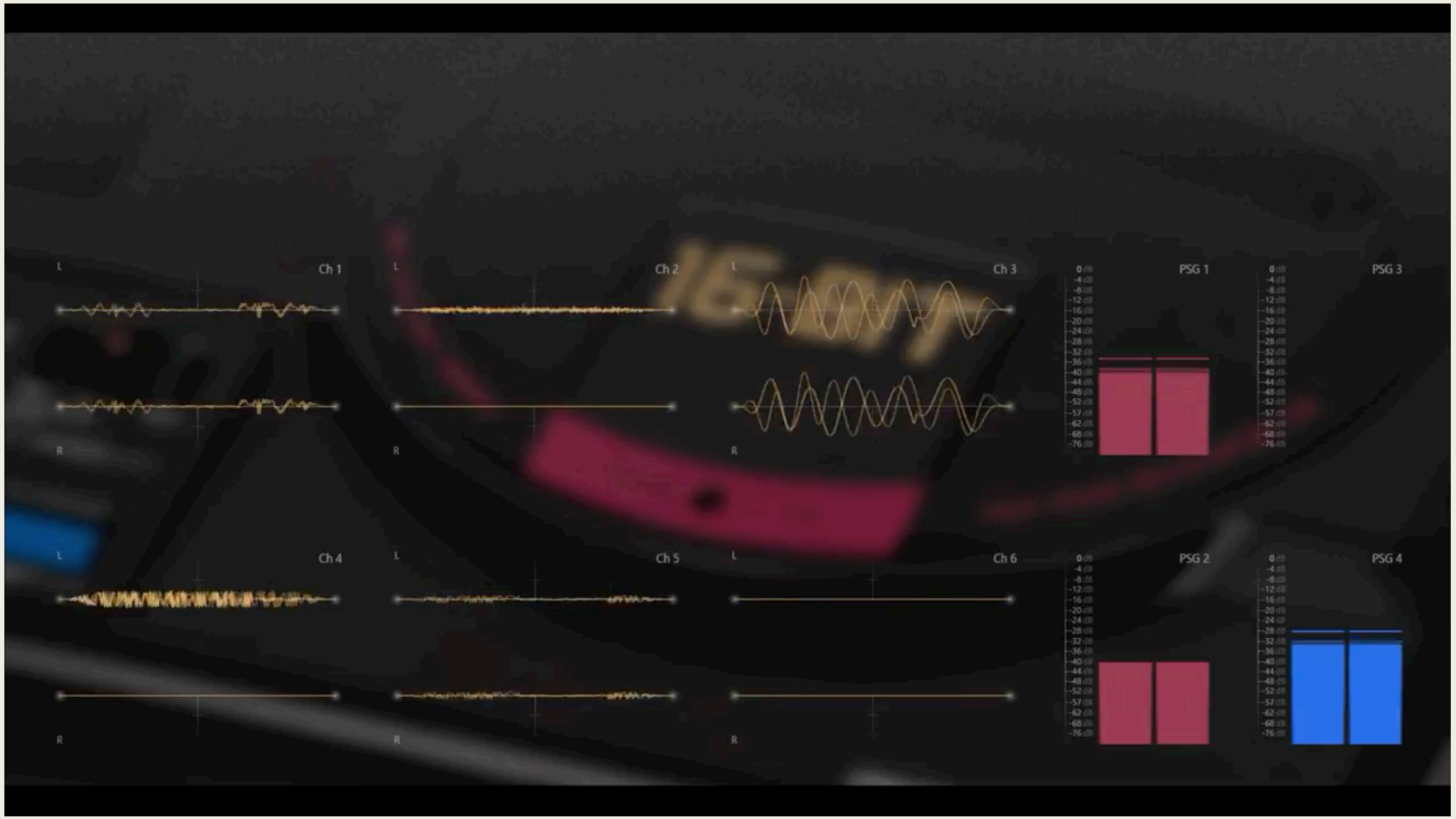
“Latency” is relative

Domain	Acceptable Latency
Protein Folding Simulation	Days
Weather Simulation	Hours
Radar Signal Convolution	200~300ms
Videogame	10~30ms
Audio Processing	5~10ms
High Frequency Stock Trading	<1ms

“Latency” is relative

Domain	Acceptable Latency
Protein Folding Simulation	Days
Weather Simulation	Hours
Radar Signal Convolution	200~300ms
Videogame	10~30ms
Audio Processing	5~10ms
High Frequency Stock Trading	<1ms

Experiment: Real-time FM Synth



[Solar Modulation - Savaged Regime]

Matheus Vitti Santos
@ Meeting C++ 2019

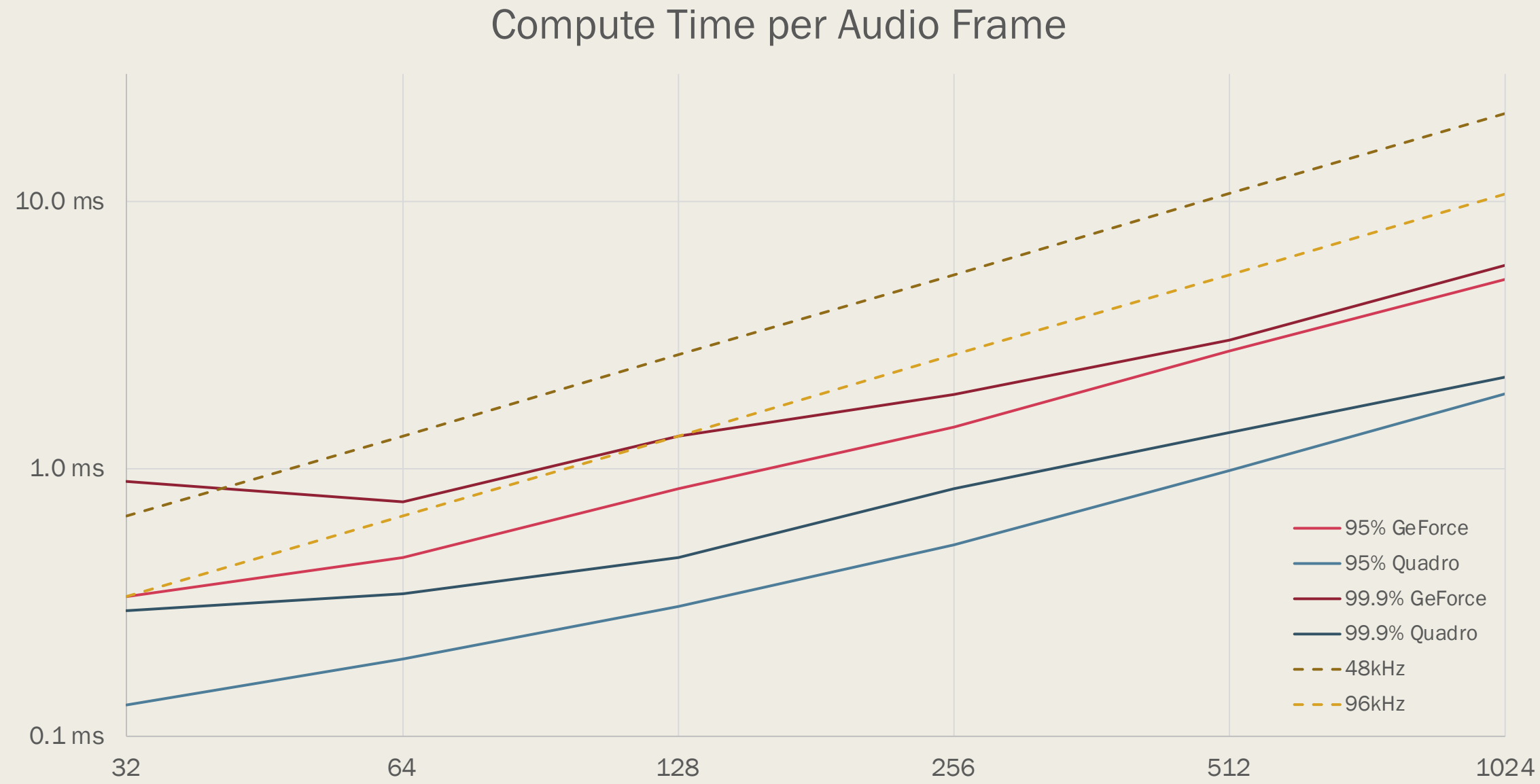
Test subjects

GeForce 640M
(this computer)

- Kepler Architecture, SM 3.0, 2012
- 2GB GDDR5 / 128bit / 900MHz
- ~390 Gflops
  - *2x PlayStation 3*
  - *Intel UHD 620*
- ~25 Giops

Quadro P400
(office workstation)

- Pascal Architecture, SM 6.1, 2017
- 2GB GDDR5 / 64bit / 2GHz
- ~630 Gflops
  - *½ Xbox One*
  - *2x Intel Iris 5100*
- ~200 Giops



Matheus Vitti Santos
@ Meeting C++ 2019

DEMO TIME!

THANK YOU!

And go do something awesome
with that GPU of yours!

Image Sources

- Screenshot: Marble Madness, c. Atari 1984
- Nvidia Control Panel: https://www.howtogeek.com/437761/how-to-enable-ultra-low-latency-mode-for-nvidia-graphics/
- FM Music Video: Savaged Regime