

NVIDIA TESLA

A UNIFIED GRAPHICS AND COMPUTING ARCHITECTURE

Cyrus Rashtchian

March 11, 2013

Tesla: motivation

- Unified graphs processor, combining previously separate vertex and pixel graphics (sub)processors.
- The vertex processor is the part responsible for doing the math geometry lines/points/shapes stuff
- The pixel process is for filling/texturing the interior of objects
- The goal was to combine these to gain through dynamic load balancing and more stages in the architecture

Tesla: some details

- it's pretty darn complicated... lots of clusters and local/shared memory and local controllers and a multilevel threading/processing architecture
- But easily(?) programmed using CUDA in C/C++
- Subcomponents execute 100s of threads and Tesla can execute over 12,000 concurrent threads
- The ISA supports everything and more... floating point, integer, bit ops... memory, flow control.... and also special functions like $\log_2 x$, 2^x , $\frac{1}{x}$, $\sin x$, specifically for small $x \approx 1 \pm \epsilon$.

Questions

- 1) Tesla sounds awesome... but what are its problems?
- 2) This came out a couple years ago... did it start a wave of unified graphics processors? or are there even better solutions?