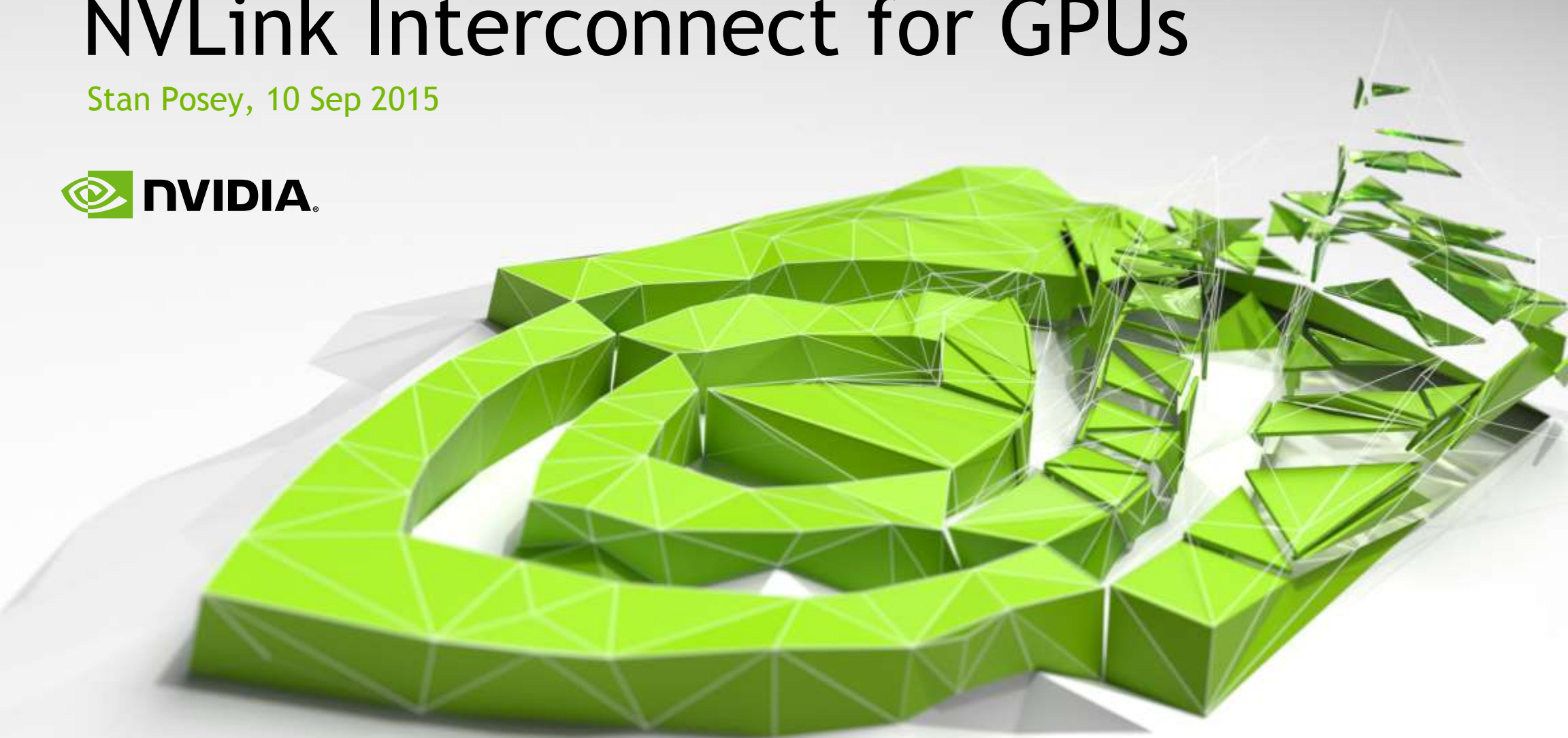


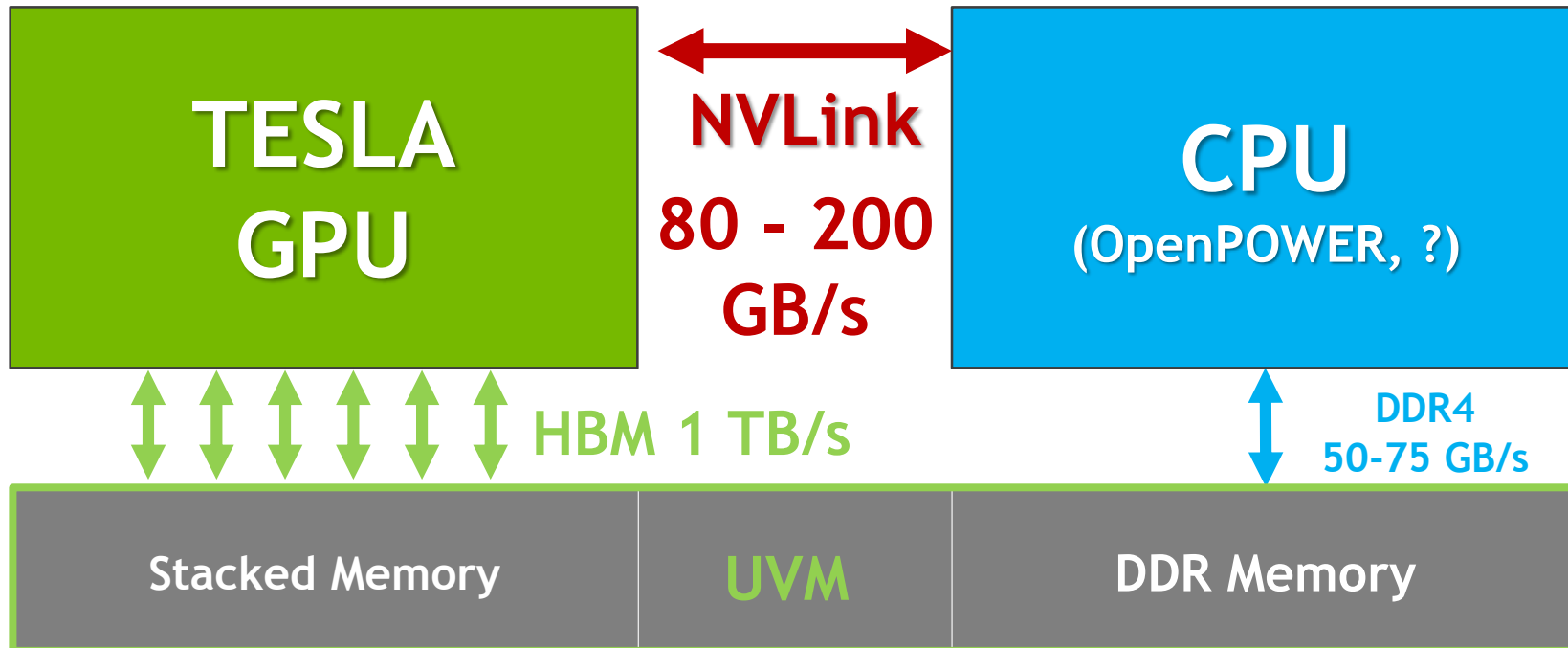
NVLink Interconnect for GPUs

Stan Posey, 10 Sep 2015



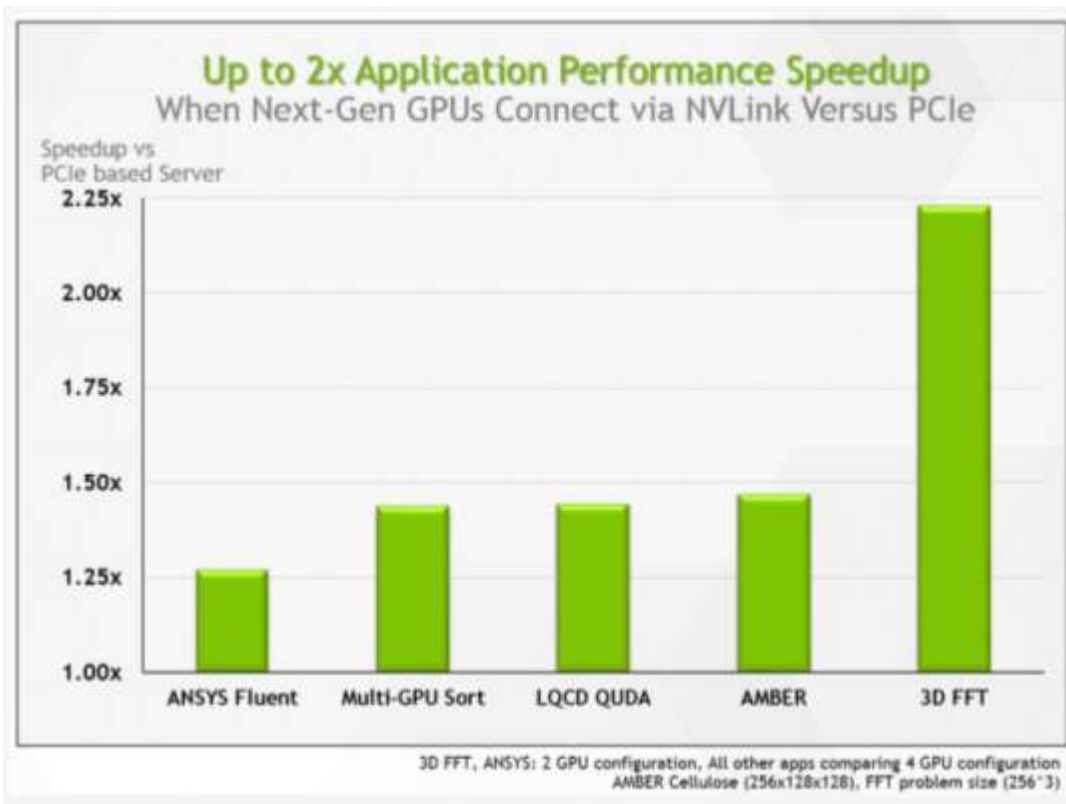
What is the Disruptive Technology?

- **NVLink Interconnect: GPU-to-GPU and CPU-to-GPU**
 - Interconnect alternative to PCIe Gen-3 (5x to 12x faster)
 - GPUs and CPUs share data structures at CPU memory speeds
 - Simplified programming model: NVLink with Unified Memory



Why is it Disruptive - Benefits to the End User?

- NVLink Application Performance
- Up to 2x vs. PCIe for GPU-to-GPU
- NVLink in Action: CORAL
- NVLink between CPU-to-GPU



CORAL Summit System
150 PF: 5-10x Faster than Titan
1/5th the Nodes: 3.4K vs. 18K)
Same Energy Use as Titan: 10MW

3

What are NVIDIA's Market Expectations

- GPU Density Increasing



CRAY
THE SUPERCOMPUTER COMPANY
Cray CS-Storm: 8 x K80



Dell C4130: 4 x K80



HP SL270: 8 x K80

- HPC Vendors & NVLink—2016

- All will support GPU-to-GPU
- Several to support CPU-to-GPU

