# CRADA Report: Scientific Computing with Video-Gaming Technologies

Gaurav Khanna & Glenn Volkema
UMass Dartmouth, Physics Dept.

# Talk Outline

- Recent exploration of video-gaming technologies with OpenCL: **Fusion +Radeon**, **Liquid-Cooled GPUs**, Apple's **Mac Pro**
- CRADA: Computational Science Productivity (**Black Hole Physics & Cryptography**)

(thanks for support from NSF & AFRL!)

# Why video-gaming hardware for scientific computation?

- Commodity-consumer hardware like the game consoles (PlayStation), video-gaming graphics cards (AMD Radeon, Nvidia GeForce), mobile-devices (ARM, "fused" processors: APU, Tegra, etc)
- High-performance (driven by strong consumer demand – 4K gaming!)
- High-availability (easy to obtain in large volumes – gaming market is huge)
- Very low-cost (intense competition!)
- Excellent growth path (rapid innovation)

# Design #1: liquid-cooled discrete GPUs node

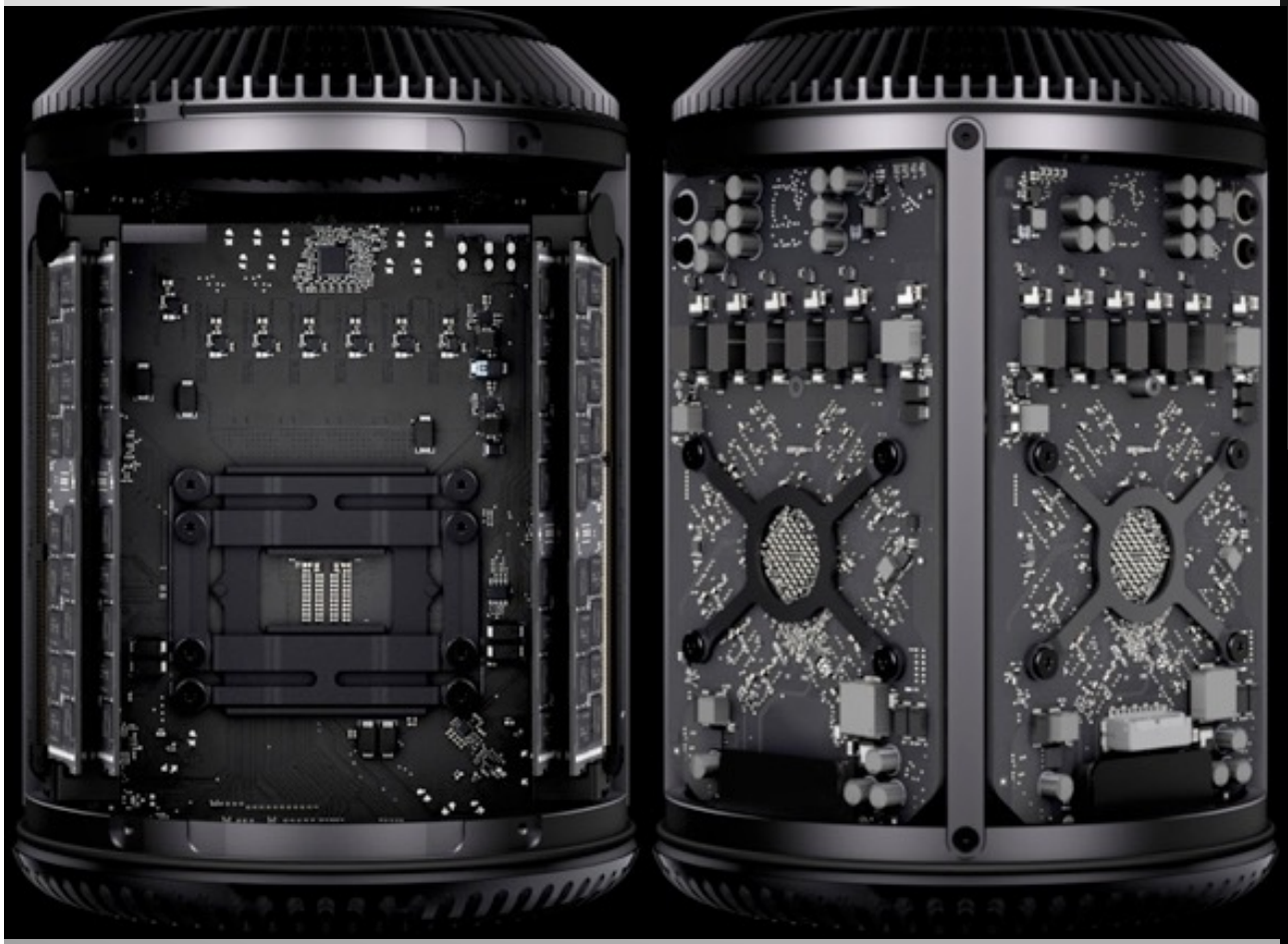
- AMD multi-core CPU processor accelerated by multiple dual Radeon HD
- Programming framework: OpenCL -- provides access to ALL resources (CPU-cores & discrete-GPU)
- NEED liquid-cooling for full performance
- Tested node: 8-core AMD FX 5GHz + 3 dual Radeon HD R9 295X2; E-ATX form
- GFLOPS: 300 (CPU) + 3x11,500 (295X2); Cost: $5K; Watts: ~2kW; load test 6 mos
- Performance-per-Watt: ~17 GFLOPS/Watt
- Performance-per-$: ~7 GFLOPS/$


Liquid-cooled
(closed loop)
Radeon HD R9
295x2 dual GPUs

36 TFLOPS
@2kWatts

# Design #2: Apple Mac Pro

- Intel multi-core Xeon CPU accelerated by two AMD FirePro GPUs
- Programming framework: OpenCL -- provides access to ALL resources (CPU-cores & discrete-GPU)
- Special thermal design via metal core
- Tested node: 8-core CPU 3GHz + two FirePro D700; very small form factor
- GFLOPS: 400 (CPU)+ 2x3,500 (D700x2); Cost: $6K; Watts: ~400W; load test 1 yr
- Performance-per-Watt: ~18 GFLOPS/Watt
- Performance-per-$: ~1.2 GFLOPS/$



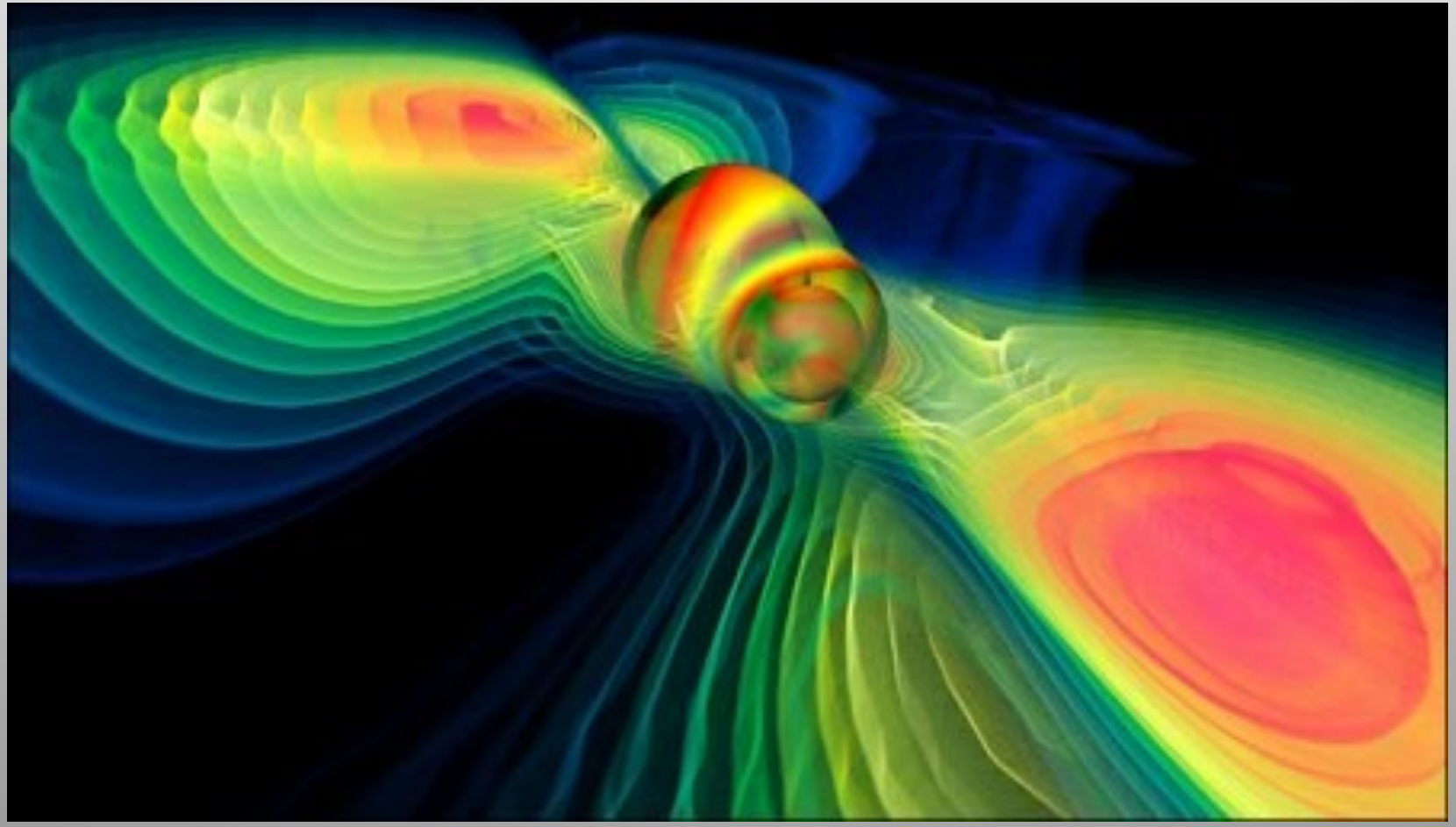
# Design #3: APU+GPU node

- AMD Fusion APU (4-core) accelerated by discrete PCIe Radeon HD GPU
- Programming framework: OpenCL -- provides access to ALL resources (CPU-cores, integrated-GPU & discrete-GPU) to a single application
- Integrated-GPU acts as accelerator with no PCIe bottleneck! Contributes ~20%
- Tested node: Kaveri APU 4GHz + Radeon HD 7970/7990/R9-290X; micro-ATX form
- GFLOPS: 128 (CPU)+922 (iGPU) + 5,600 (R9-290X); Cost: $1 – 1.5K; Watts: ~600W
- Performance-per-Watt: ~10 GFLOPS/Watt
- Performance-per-$: ~6 GFLOPS/$

# Future Exploration

- Build cluster from “best” node design …
- Perform various scaling tests / benchmarks …
- Other technologies to explore: OpenPOWER (includes many features IBM POWER8, CAPI, NV-link, etc.); Tegra; OpenCL 2.0
- Others: Altera FPGAs (now support OpenCL!) and supposedly offer excellent performance-per-Watt (but, developmental systems in the $8K range!)
- Other??

# CRADA Report 2010 - 2014



# PS3: 150 GFLOPS; 100W

# AFRL CONDOR
# 1716 PS3s!

# CRADA Transfer

- AFRL granted UMass Dartmouth 4 full CONDOR racks = 4x44 = 176 PS3s with network, cables, PDUs etc.
- Each rack 5kW power + 16k BTU/hr cooling -> 20kW power + over 60k BTU/hr
- Some infrastructure challenge
- Need to keep costs low; install racks and make them operational quickly
- Built-in flexibility / scalability
- Consider installing racks in a “reefer”!

# PS3 “reefer”

- Milk/Meat freezer containers very common
- Used by stores to ship foods across the country
- Very high cooling capacity
- Easy to find; many options
- Easy drop-off / installation
- Ideal “portable data-center”
  – Cost $30K

Media attention (Forbes, NYTimes, Space.com, local TV, Radio ..)

# PS3 reefer