Commodity-based Scalable Visualization: Graphics Cluster Components

Randall Frank

Lawrence Livermore National Laboratory

UCRL-VG-143528

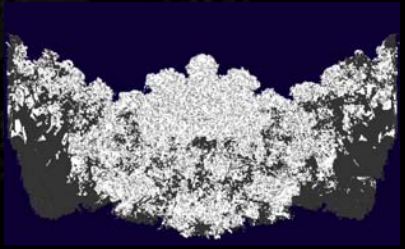
SIGGRAPH 2001 EXPLORE INTERACTION AND DIGITAL IMAGES

Scalable Rendering Clusters

What makes a scalable rendering cluster unique?

- **Generation of graphical primitives**
  - Graphics computation: primitive extraction/computation
  - Multiple rendering engines
- **Video displays**
  - Routing of video tiles
  - Aggregation of multiple rendering engines
- **Interactivity (not a render-farm!)**
  - Real-time imagery
  - Interaction devices, human in the loop
- **Unique I/O requirements**
  - Access patterns/performance

Argonne Chiba City

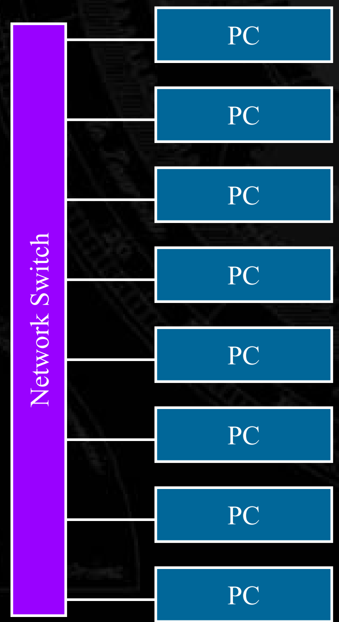


469M Triangle Isosurface

Graphics Cluster Anatomy: The Cluster

Start with a basic computational cluster

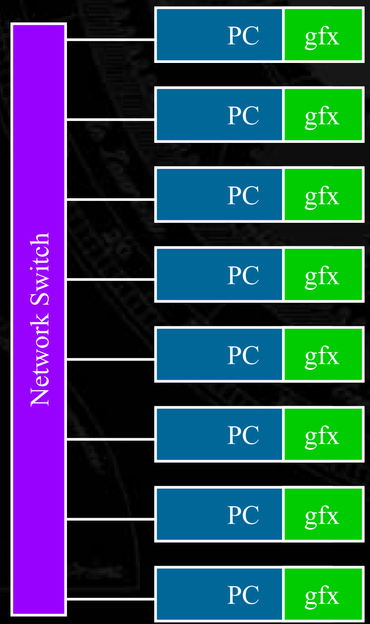
- COTS computational nodes
- High-speed interconnect
  - GigE, Myrinet, ServerNet II, Quadrics, InfiniBand...



Graphics Cluster Anatomy: Rendering

Add multiple rendering resources

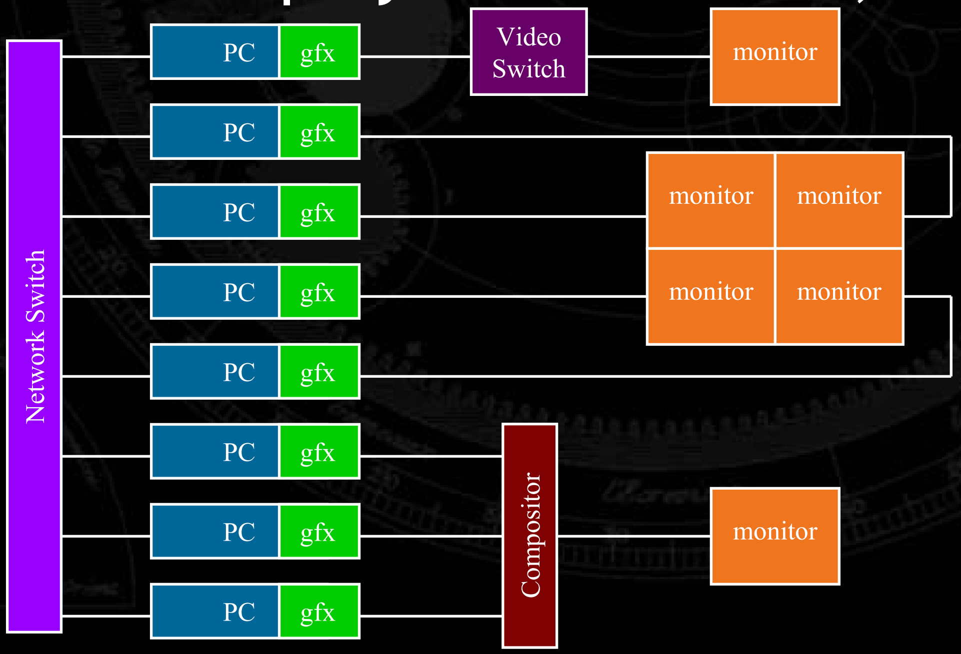
- Software rendering (Mesa, custom, …)
- Hardware rendering cards
  - nVidia, ATI, 3dfx, intense3d, …



Graphics Cluster Anatomy: Displays

Attach one or more displays

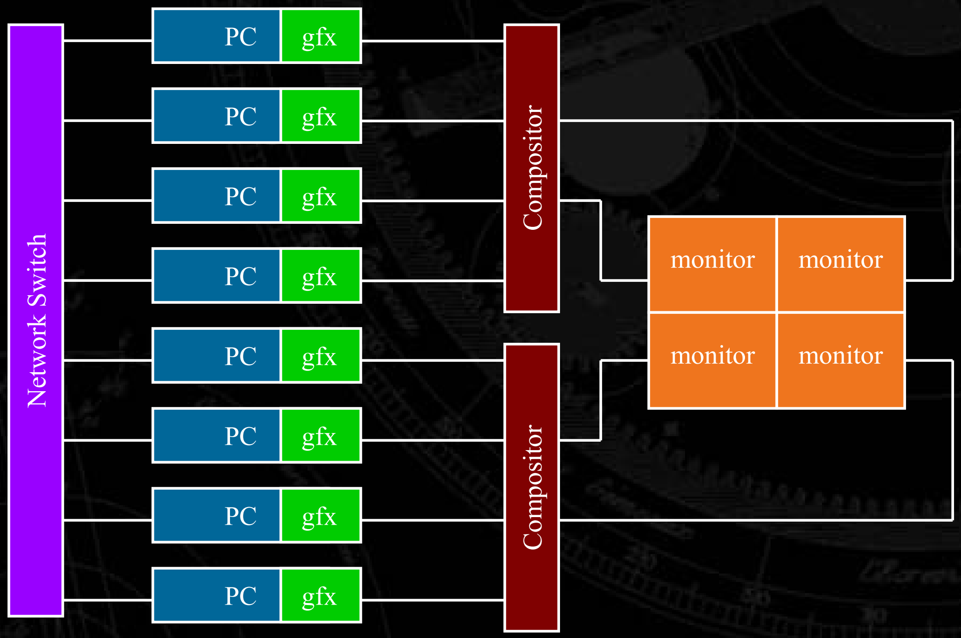
- Direct display monitors
- Tiled displays (PowerWalls)
- Composite displays: M renderers, N displays

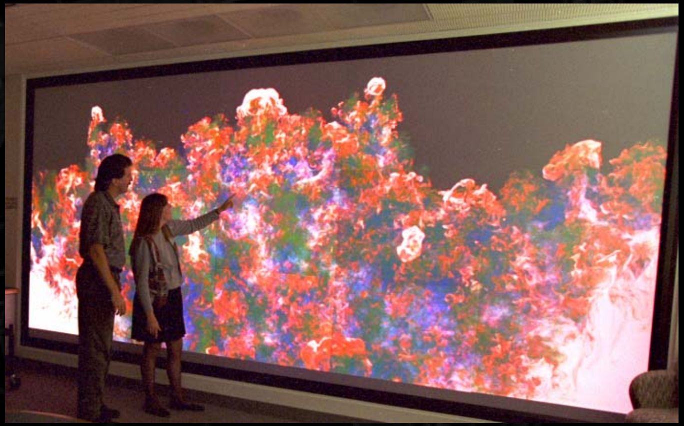


Graphics Cluster Anatomy: Displays

Advanced layouts

- Combinations of tiling and compositing





LLNL PowerWall (6400x3072)

IBM T220 “Bertha” (3840x2400)

PC Graphics Cards:What are they?

PCI and AGP commodity graphics cards

- Cluster-capable PC architectures
  - Intel CPUs + AGP + independent PCI 64/66 (e.g. i840 chipset)
- Common 3D Graphics APIs: OpenGL/DirectX

Why are we interested?

- Large numbers of cards - low cost
- Games + fast PC hardware - speed
- Graphics "innovation" leadership

Broad categories

- Consumer - Games, Media playback
- Professional - CAD, Media generation

PC Cards: Consumer

Consumer: nVidia, ATI, 3Dfx, Matrox

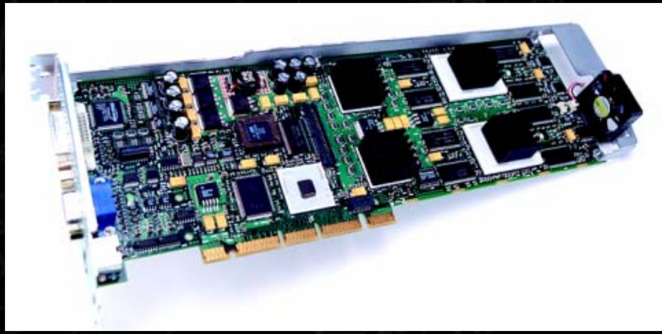
- Pros
  - High fill rates (600-2000Mpixels)
  - Hardware T&L (8-25Mtris) in most recent versions
  - Innovations: cube maps, texture combiners, vertex programs
  - Cheap (<$400), price sensitive/competitive market
- Cons
  - Driven by games
    - OpenGL can be a secondary consideration
    - Poor line drawing rates/quality
    - Windowing issues
    - Readback and buffer access issues
  - Difficult to achieve "ultimate" performance
  - Bit depth issues - good enough quality
    - Screen and pipeline (e.g. Texture compression)

nVidia GeForce 2

PC Cards: Professional

Professional: HP, IBM, 3DLabs/Intense3D, nVidia?

- Pros
  - Full accelerated OpenGL 1.2: 3D texture support
  - Finer attention to OpenGL detail
  - Deeper intermediate computations
  - Non-game features
    - Higher line drawing performance/quality
    - Larger memory
    - Concurrent multi-bit depth/screen support
    - Enhanced video output options (e.g. genlock)
- Cons
  - Lower fill rates (100-400Mpixels, application market bias)
  - Fewer "innovative" extensions: Cube mapping
  - (More) Expensive



3DLabs Wildcat II 5110

PC Cards: What should you expect?

- Are they really Infinite Reality™ pipes?
  - Basic rendering and raw speed: for most measures, yes
  - Image quality/integrity: no, improving
  - Flexible output options: no + DVI, improving, but no DG5-8s
  - System bandwidths: maybe
- Easily rival present desktop workstation graphics
  - Vendors are shipping them as options
- System stability issues (Read the game torture test reviews)
- High fill rates (Not high enough, thank the BSP tree)
- Future feature sets
  - Exceed the IR in many ways, can be raw and complex
  - Extensions: increase the difficulty in writing portable code

Graphics Cluster Anatomy: Issues

- System bus contention
  - Simultaneous graphics AGP bandwidth and interconnect PCI bandwidth
  - Careful selection of motherboards (e.g. i840)
- CPU options (number/speed)
  - System overhead (e.g. TCP/IP stacks)
- Core system interconnect
  - Bandwidth/latency
- Operating system selection
  - Drivers/cluster management software

Aggregation: Tiling Vs Compositing

Goal: aggregate multiple rendering engines, combining their outputs on a single display to scale rendering "performance"

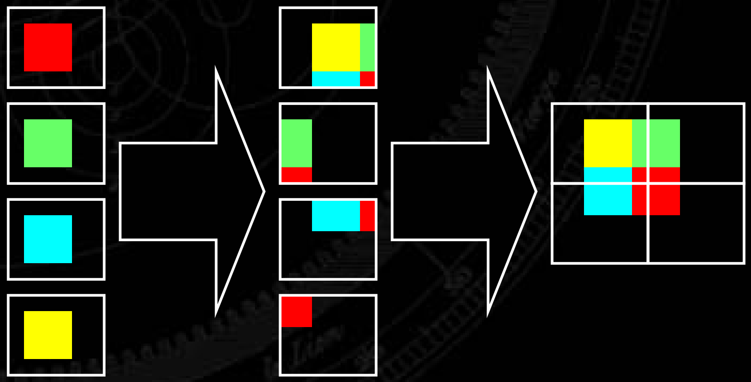
- 2D - "screen space"
  - "Sort-first" rendering model
  - Targets display scalability, higher frame rates
- 3D - "data space"
  - "Sort-last" rendering model
  - Targets large data scalability, higher polygon counts

Aggregation: Tiling

Tiling (2D decomposition in screen space)

Route portions of a final aggregate display to their final destination with no overlap

- Order independent
- Destination determines bandwidth
- Graphics primitives may be moved, replicated or sorted for load balancing
- RGB data

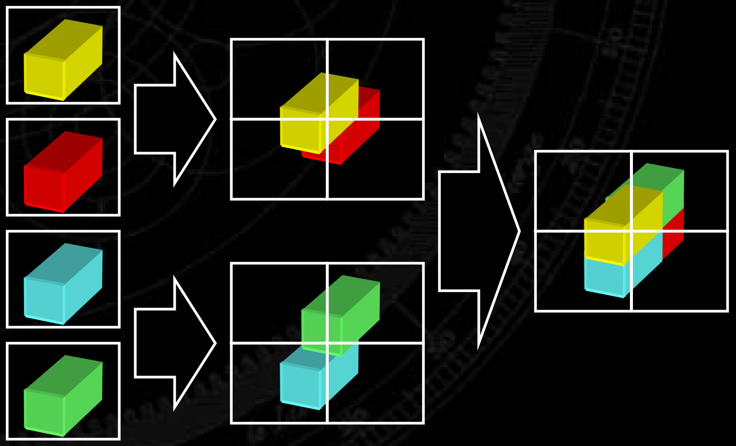


Aggregation:Compositing

Compositing (3D decomposition in data space)

3D blocks that are combined using classic graphics operators (e.g. Z-buffering, alpha blending, etc)

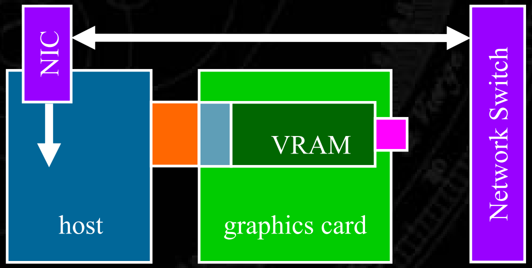
- Z, α, stencil enhanced pixels
- Fixed 3D data decompositions (data need not move)
- Bandwidth exceeds that of output display (3D vs 2D)
- Hierarchy trades bandwidth for latency
- Ordering may be critical



Implementing Aggregation

Composition datapaths are targets for specialized parallel and asynchronous interconnects

- Basic operation
  - Access the rendered imagery in digital form
  - Route image fragments to composition mechanism
  - Composite the fragments
  - Display the results



- Approaches
  - Reuse the cluster interconnect
  - Utilize digital video interface (DVI) output
  - Use a dedicated interconnect

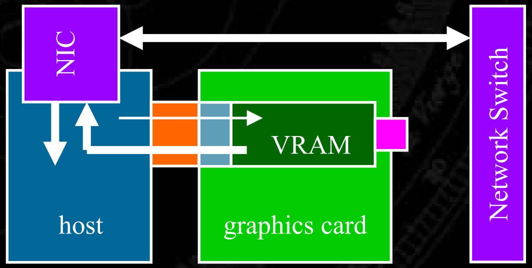
Reuse Core Cluster Interconnect

Compositing/tiling directly on the nodes

- Image or primitive exchange over the interconnect
- Readback of graphics card buffers (RGB,z,α,stencil)
- Flexible computation of aggregate imagery by host CPU

Current solutions

- Quadrics, Myrinet, ServerNet, GigE
- MPI, VIA, TCP/IP,GM



Issues

- Processor overhead (second CPU?)
- Available bandwidth and latency
- Framebuffer readback performance

Myricom Myrinet 2000

Digital Video Interface Interconnect

Video based solutions

- Ideally suited to tiling, DVI inputs/outputs
- Asynchronous operation, Avoids readback

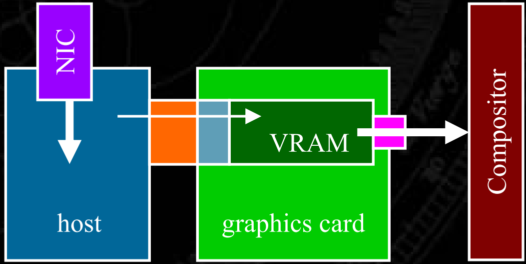
Examples

- Stanford: Lightning-2, U. Texas: MetaBuffer

Lightning-2

Issues

- Synchronization issues
  - Tagged imagery
  - Auxiliary signals
- DVI signal and pixel format limits
- Limited compositing functions/ordering options
- Scalability of mesh architectures



Dedicated Compositing Interconnect

Secondary interconnect dedicated to compositing

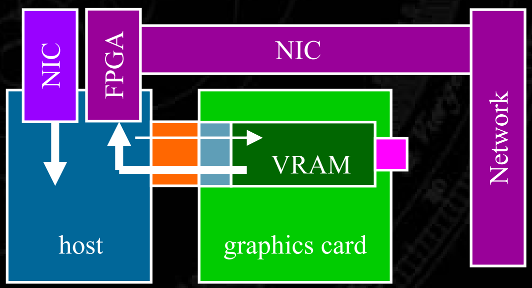
- Need not be fully connected (data decomposition)
- Offload operation from host onto custom chips (FPGA)
- General pixel formats, programmable composition functions
- Interconnect switch for ordering

Examples

- Compaq: Sepia, IBM: SGE

Issues

- Framebuffer readback
- Additional host bus demand
- Bandwidth-pixel count/format



Sepia-2

Composition and Interconnects: Issues

- Multi-pass rendering algorithms
- Framebuffer readback
  - Performance and availability of graphics APIs
  - Limitations of DVI: distance, pixel formats, bandwidth
- Graphics card bit depth limitations (e.g. global Z)
- Latency and ultimate framerate issues
- Protocol/API inefficiencies
  - TCP/IP: High overhead, Jumbo frames (M-VIA over gigE?)
  - MPI: Design issues for streaming transport
- Flexible/scalable software interfaces
  - Data partitioning: The "zoom" problem
  - Anisotropic rendering environments