

Overview and Introduction to Scientific Visualization

Paul Navrátil, Ph.D.

Texas Advanced Computing Center

The University of Texas at Austin

Visualization Training

September 8, 2011

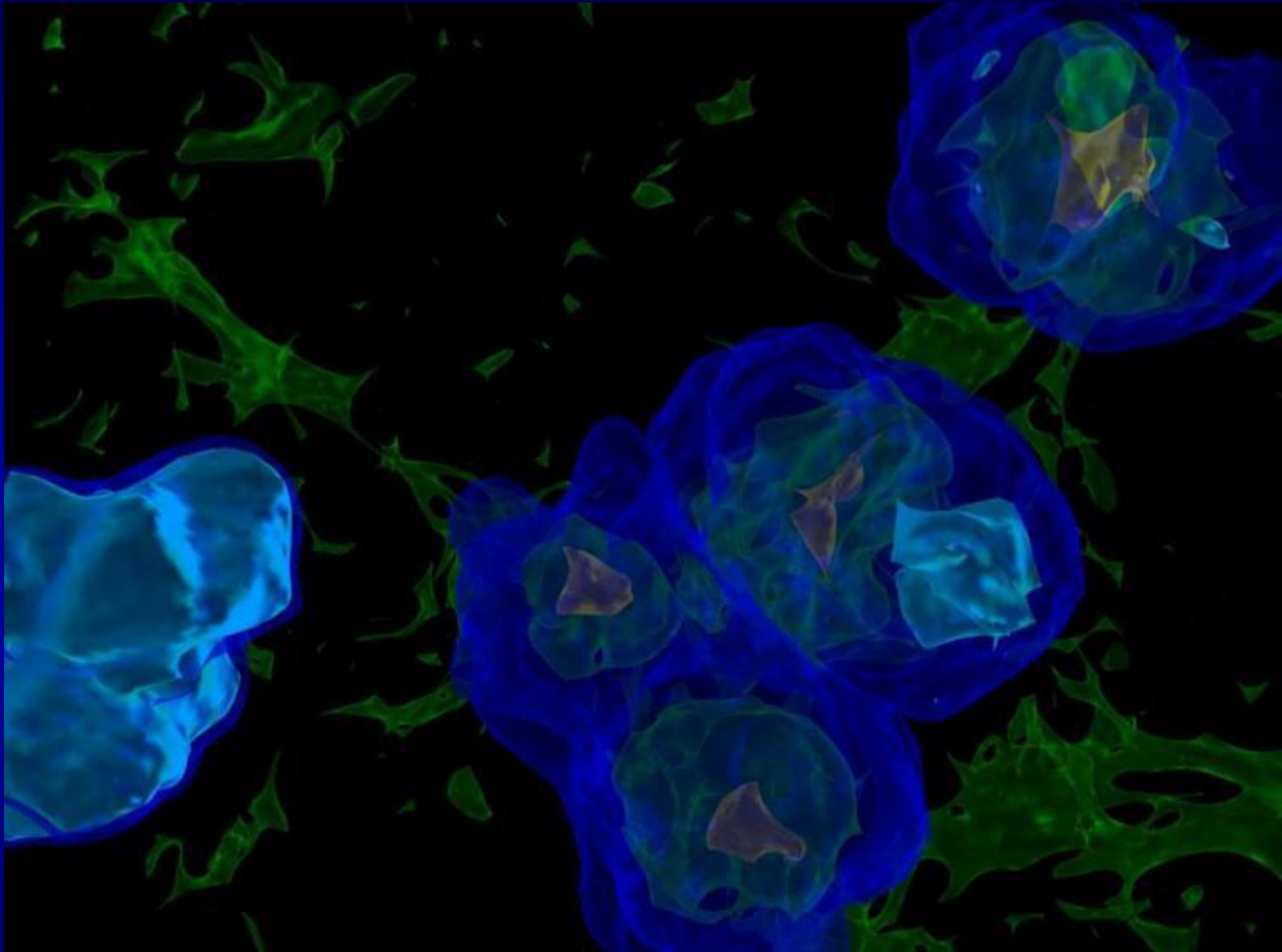
Scientific Visualization

“The purpose of computing is insight not numbers.”

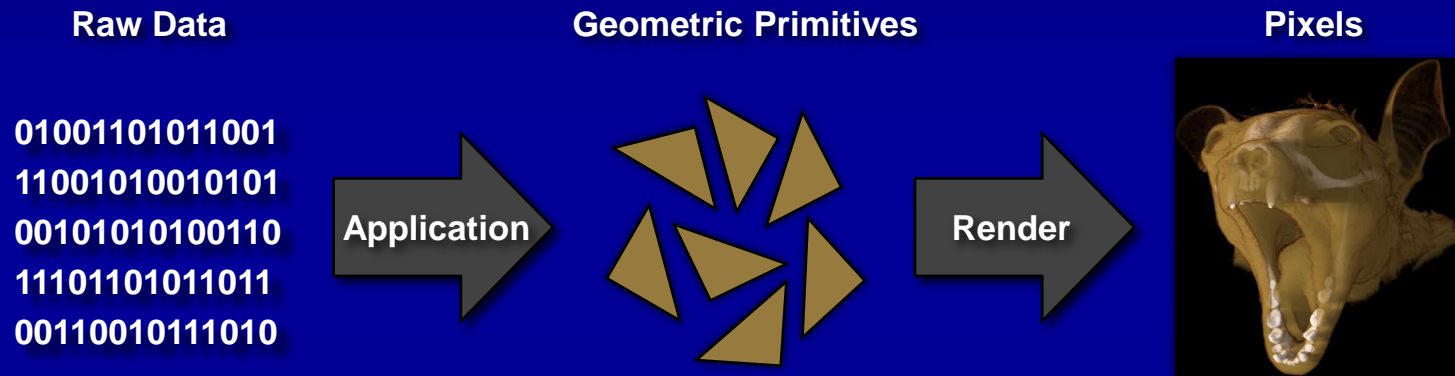
-- R. W. Hamming (1961)

x/w+bfGoZnyKDqfxH5BN0Mpt9Lx2TD08LCMwGVI89TJU72KMfDiS0J+UfbiAkLhEXf9XgzrP0GBCZjdvxWUCuFyS1XVFZuB02p8kY88
+KCIHMKdPrfPFgavrxjayRjxbDh0Rovb/VF5N0iLJuTjeEIBIQJLUJrcx/J+MDnrovmyGEq3cS9BFpvmEmFkgp58Fxt/k9kEfrUQoHAQWEmwLZRAsY+yDDdxm9EVAzXyGtniHH71Q5/gqH
OKQvH4PHfjnyXSMV69K9kevnyGnTYbvivS9FN80TUKakwoReySsGOZIVTglnGWR40ag5pzuuFC/Klp09MXrYwR8R36+OmROW8Bkq4r fht80Fb fh309VRXaa28/ fYtY8TKVv0swYc6a0e/
+twg6Zab2TduYwvBpDwf3wUVs2YgKUvpuH31N0K0xRtes+aN/CwaQtYy16otGREPz12Qh27VUwvym2xRYTZsc6w2dcNm5snI3BjEpbZJaD0qljccPNA5gHt0sKlMpn+nLus+
55uWd09xG337nm9SafNPr f+foi9XnhXfLedIMRGWBFKmtbRQxRkLh4zE/Zmzyb7d1pLVj67E5RXnkb/1NzwsnYZ/7eIwJjWAl0wva8fiXEYOetgHo+d2b8Z5t2m3p0i f4IY1x7tDEdsXg9
dmwVI4DGuJ42F92xX+rVT4U0ygbkvz fIAP/ubUI3P42J4da2U+/vSza+hC+prCupWfljtjyH6BRYMokd7yuCfx6LxjnstagY04tUcLlDnN0uzb5xrMMmXasPvmyeTEUBclUjGZr0v4/j7fF
rWk+aBARjtCTAz/y74uSo189JQKGOEESRorISwoc+XcMNXabGdnHluBD558RQ4xDluBgFC+U1VPFle9ppVyx86rWnlie17Vd0bgX9iUvflhS0PvSzGnHWKje076v2W/qlA8zQG5CVYYdV
20HQs98PRJNCTHsrDjnlY0s2P3DlPnLwm+w3fjNuJwHPVYxv71sH2YRiel fEli f/sAhcEVApI/nd+UP7rr6d394V2mHPbDU+neqHJsAgswWLMqr00K1x34Ua9XxH2dAiuZNRgn9X6/d20
VdrWm6Ccvz/WS3Am44Cnd5gvAb4qTjyx2TP81ePBVg/WwxhzbEGzqxc6zovEuDsK+Y08mkiVFMqMD2QbRL/MMW+VVBtpi87
+N1T9N24/0aW/YhUQhKx18cnf0c6YNxJa+xztIbPIjuMyRqBu49G4uWSdLQ+PEFeb/7AAcEyDD4UC5eTIIHNdiyid6bjD7ueJvGv89oXwvSAsHMYGth0
+tcjwsdAVubVscbluAdqMmYK/rBZhlK87YXtvFwJ1W1LHQYj5ShL56Mp09tbrP0SPAoSxh0TX4QY7jXG4To80YIsYUQ/woLiNuaxs2c/Ln4lwzK7w+rXkvhaitF+nARZowWY2t7GRoftEXuDe
a5MQJCvGwix/MfLfnlWwntDEVLR3R45oBlyzXgPdgFKZcosX8zE5tx6HY0
+me6MGc44PI2auw5boWv/7KmMFXB FUN/nJOTj0XxXfG47Vj8XrAGjcmHlzWCUEh4TgmiUCEdyQePvD7q0PubALJyXZG5TxdagmvLA2+
3xtKjUP74vK/k9E0rgCHZ26aiTmEmWkX3vPWS6dh72RFA0GxHUukdNI7
+gg98no8rNwYwKlL2XTR6D1L/u3Jc2yVA3gMBDlpq6c02SptsitViz47/TcqmAtRYym/cZKo+e1PIb0iuvc4ZT7/7JERZTwlMf8ixPFpC9WkixjyucYQoshLh+iQJfi0V8P/KC2m/u3Hu3
djl05bYnv+2QYX4z3RvUOC2jknYPGMAM0fNDjgpcA/bhIkWsmQXk f1UecTXTwvufSErtyHQcn3oVhhYY+FWx0x0dj5I88Yd3+
0aLoQ0zd8Q06k95SvJiJuhucs2vj fRr7XgUR2NUH/q00i1YYzZfUyNjYUY76Ps1IXF+IsMtj8Sy7D5pSheHxxmcdPDgQDev0wcdLm/D3zFKsKnmNKecL0wDMKGWjLkhaHYffjyPhHu6Eh
hYS2PxELK8j0vme2y/8WH4+dW9kEYE/N7Pbagz1I20fWHL2/LzYnxoowqB3WvYlp1dJ5MmV+
3D0BtPxsW1OKjUg+qjx fV0ZUaw7eK0EUS/bUnGNC66cd6PiqyqDvwAryNVjHWVT9tUSY0D14M/RjwfgoLoXazqlYeSjvKJMxK80i1FYQMbhD90o0q2K4IW+
2DfUD+E3/DDh9uBEJEI fSMP8Yvsc23p2Z6Zcd0bEb+QIwiPwR6kJGXFSPysJve87zweLz9KwOBZf02+wFnuhplLLlau5RzZY4fatVIRNjwRI3Y441m6irKFE71f5T6NwYcGK6DM+BaNQpe
ibP0kCFRZuLEjEicIG3ajxXwyzqzbBDpyAg/reulH38EoERVmz4v/ueQEBNEW70U7LWb008Ku7vDRbA52wYLe6Ly/HnwuPDNTXyeikx+77
+e7Rd9lpj3YV0GFPjAZ632QRB0S20QILS3kcaycMX+wpKPURAcM4XxG1IoNgYNS/pi27NEKM6Pc00L3TE6v3
+SHUJdW7R3t12C5LPHv1r fRf9q9y4fqB6yMfG9X9Rxy30u5JHn5JvbK8xg31ycihl1SbCx8/4oS502YZQ/euNprq0e/vl0iak644FoKqZf2laKmq8skbjHFS0F9ByiJ9V/dUNC7Hh0H5
eDy6vo/YSJK0kTD+/vuuBoaaCut2VhezdtXq0ytMS3vY7YISvH/9/xSuS12ZiNP03fbBBGjDv7u2gV/JuLReVx+PyMQg7n0d7orPk0fu2PoK2iGLG1j ffn1f0uESGv10rTlvd3A2pu29tS
8bwS1/4Q1r8DvnGyFFb f16jYyBHLCPwReEJ/Dpb+PJ0qxj9TrHB3skpPlro fcsj4111L8viSj/eznhCgqeJ3TrEnFd0E4N5t2aUvhv1T8XxHeG13x0B59HxCPXz/tBehXzWtb/bfr0C350
64/LULLuRoMh7WD5uJSLid7gJ3Ihp7JwWgtbcflk5Sf5S3AcnSmzI3wak2g2ZMCD/zE0DbKAAxvA/Hj0C+
8V1sztvM/slm2/r92KFNzc0Gpm2KaTktP0uDR8vHA0/6YIcs8WPZu70tEN1AUB9nKiYUFza7rpH/hpicvfk+2H9kwUlnkC8zjAUD8bHyvqKJypmEo7nztY5ULExojJd261yJk2Dq4z1
fCm9I0TJRCSp2/Hn51kiPyqspzzIkBBxQYv/HxwLtoSfoFiPR+ooN3rVDy0yB1DhLkHxT8a17ni54MteLlRdU8HztUeujJgppMSKU196/y0TH1S8wOzj f+mQGlP80wmd0oJwdbAr703
FrPnxw28mPvgb51Pxo992QfXFfvA/HYH0lh1h+6cPXuY74+e7G1S/745pUb6M8tHYM/LRuf9b7+8
+TgAvS+P2iikN/d512Sk1FW9pQHfK29QCQWFP09iv0chziwYhF8tKsfg9Du043pQ/eEm47UEmu5aeuLZtSjMvRuKjEwg1sIFTq21GBru79G1f3cpIejgidsn3BHCUomGL1JgsBh9Bogx8
zB7u3tWu5a49DvqJkSmgSwaKzlgS1snzH6h9f0nU6u0PxcV7RCU7+tkalN2lpKiepg9LGR9EBgfVd/Rh6QGJmiD0uhlCmmhjUg4mH2DijRzG7Y+
6dJdRwqV5FD E+m0uZwG1S2T+Pm7I92a1l1sytlM09JkBBGrWEWArNT5u0KYMnrlts4p9rLr10es65MBU8kBCXput8Ueaiw8c7eIM50puV154G9ItwzWJsc3ccH0Wv7Y9Ht3yMIHY3tad0Y
55ocKPr53Sr8jX9SVWnxxrLT84Jnxwtw820Ek672oGFJpZ/XFVfir3umVwUxrZvyJV0b9YkU1p05vhQ1VwKftr16k5dHPFlir2f3p7GnPYra2552LC7NU4CubU1lkZbUWwdfg/S+feA
1RVaj7WEZcP+s0NjMFVdSH+0B6CN/2Z9w7pput0UHW7b5v5JcX0VpJ7tw2d/9X0c33Y7J8Lx2SIOb6dNQ5/nZXimF5evm7Kp7n3VeiF8xjDXLa6ybIrt8N0gX2r8EohAL1 fcdpRiHCvq4p9
Et0H2K20Iiz2+n59Ln2

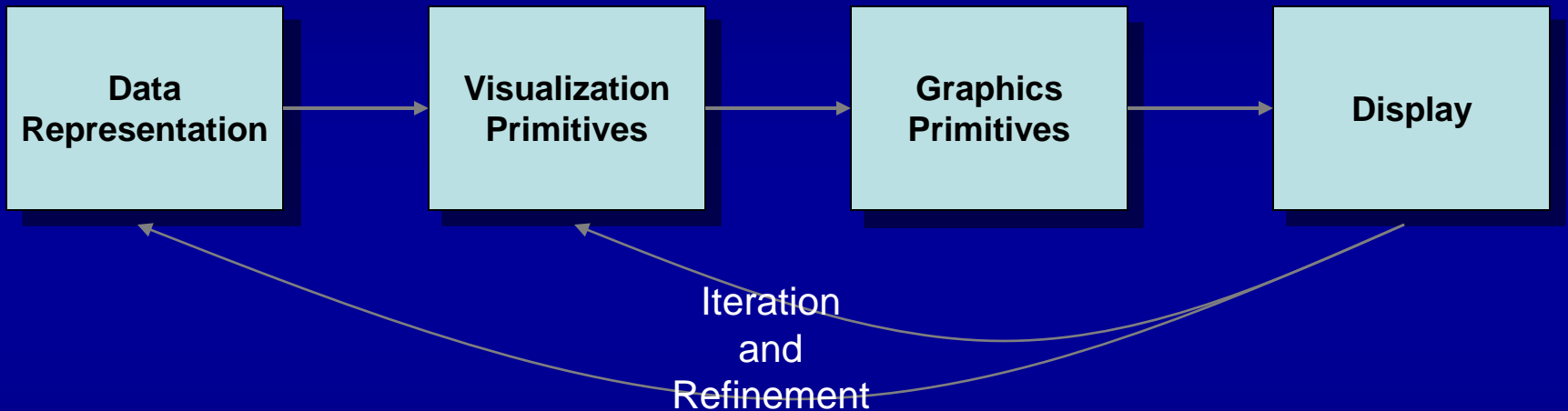




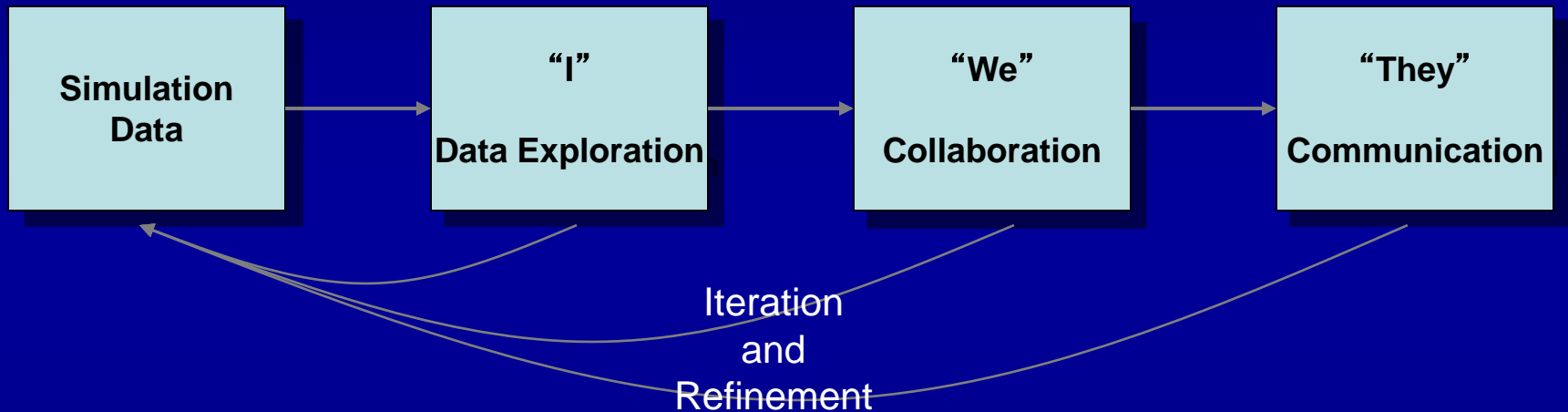
Visualization Allows Us to “See” the Science



Getting from Data to Insight



“I, We, They” Development Path



Visualization Process Summary

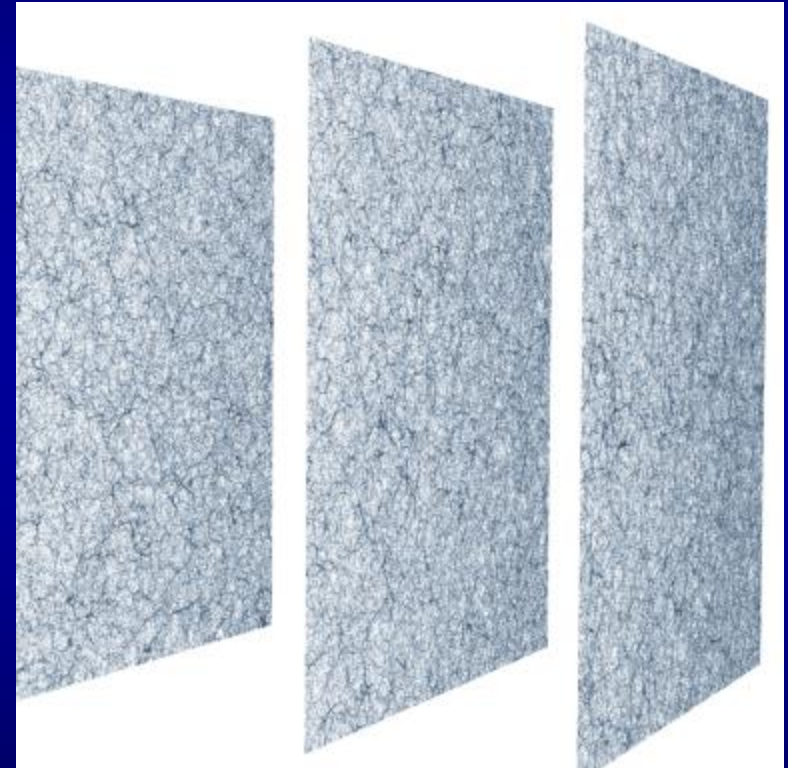
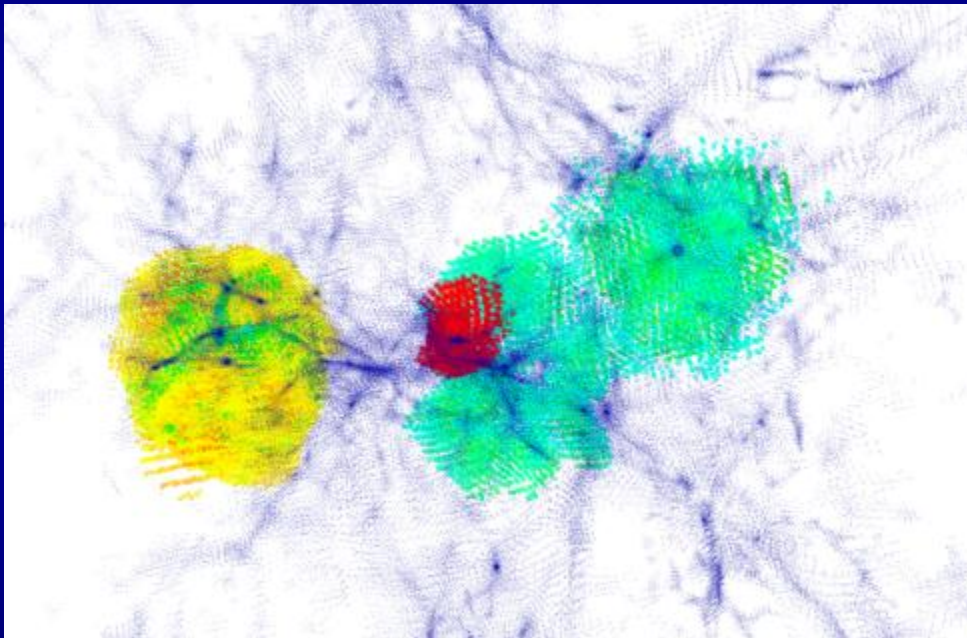
- The primary goal of visualization is *insight*
- A picture is worth not just 1000 words, but potentially tera- or peta-bytes of data
- Larger datasets demand not just visualization, but advanced visualization resources and techniques
- Visualization system technology improves with advances in GPUs and LCD technology
- Visualization software slower to adapt

Types of Input Data

- Point / Particle
 - N-body simulation
- Regular grid
 - Medical scan
- Curvilinear grid
 - Engineering model
- Unstructured grid
 - Extracted surfaces

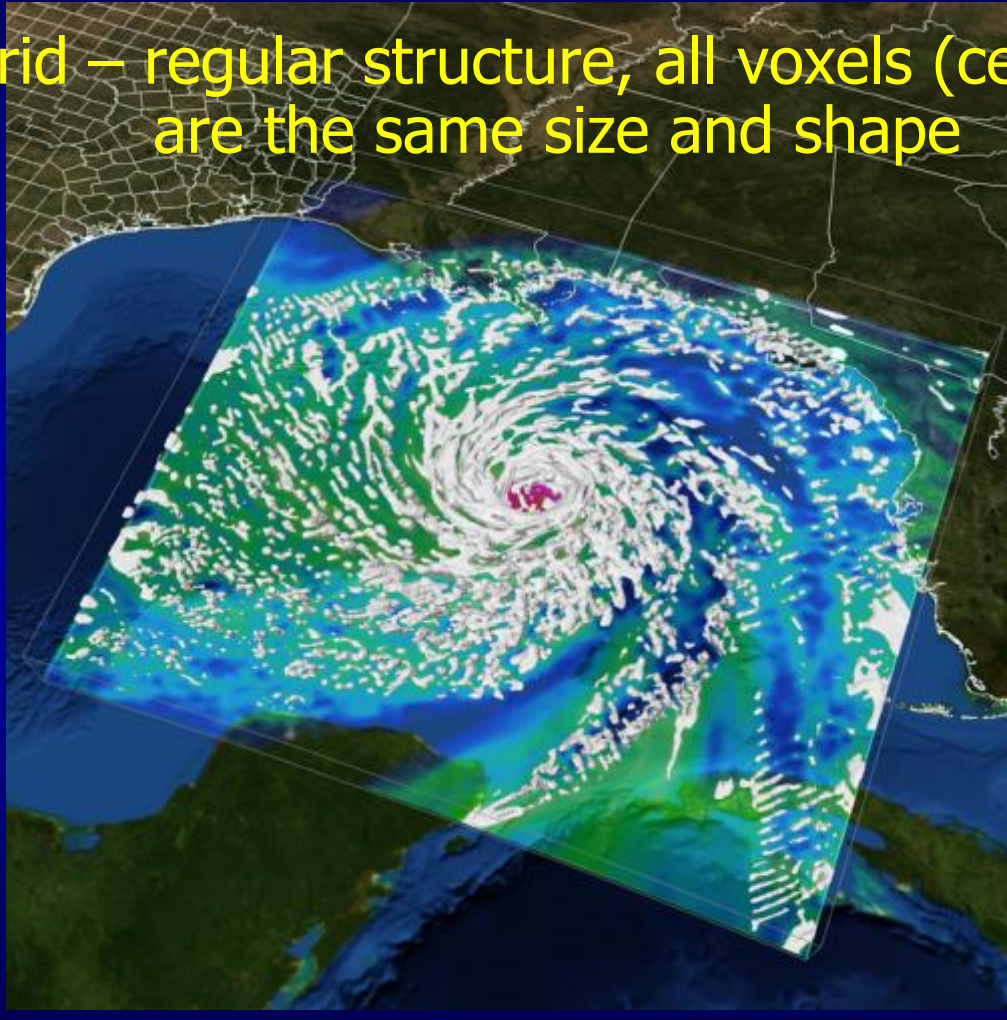
Types of Input Data

Point – scattered values with no defined structure



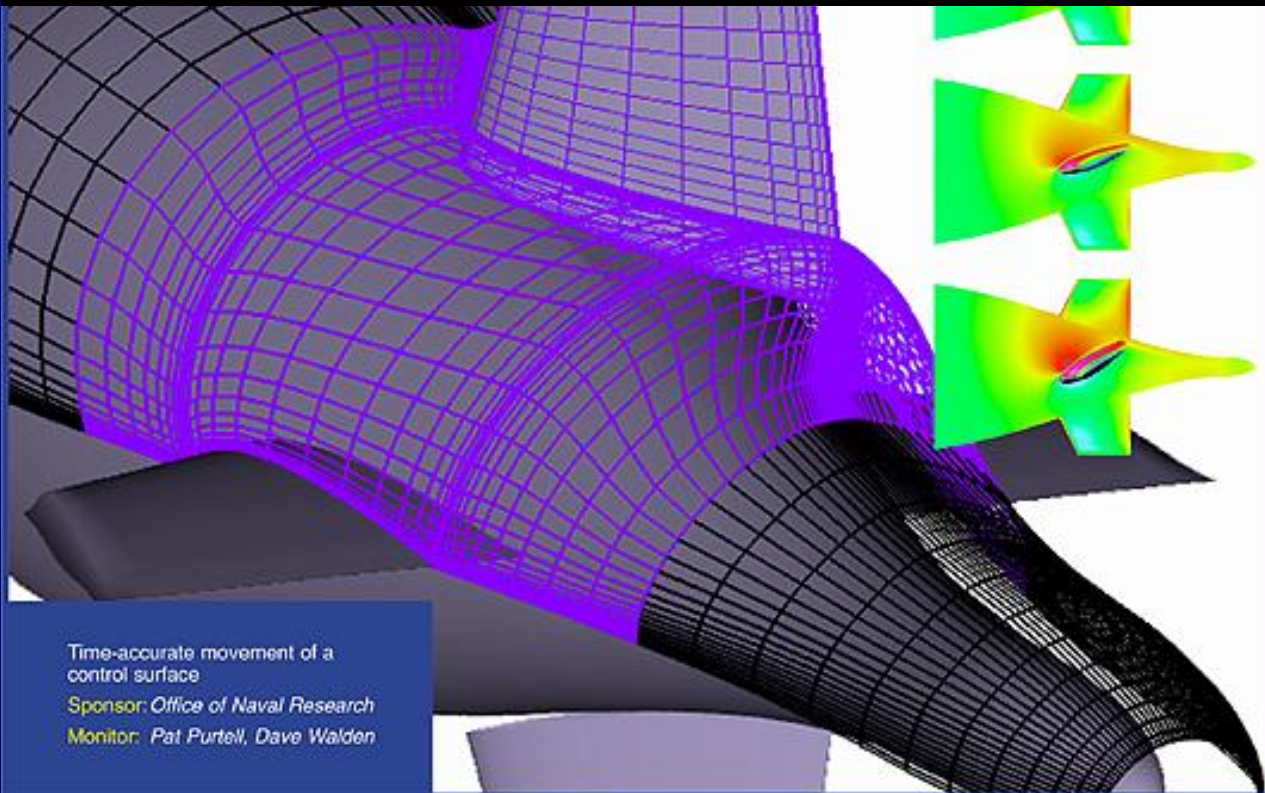
Types of Input Data

Grid – regular structure, all voxels (cells) are the same size and shape



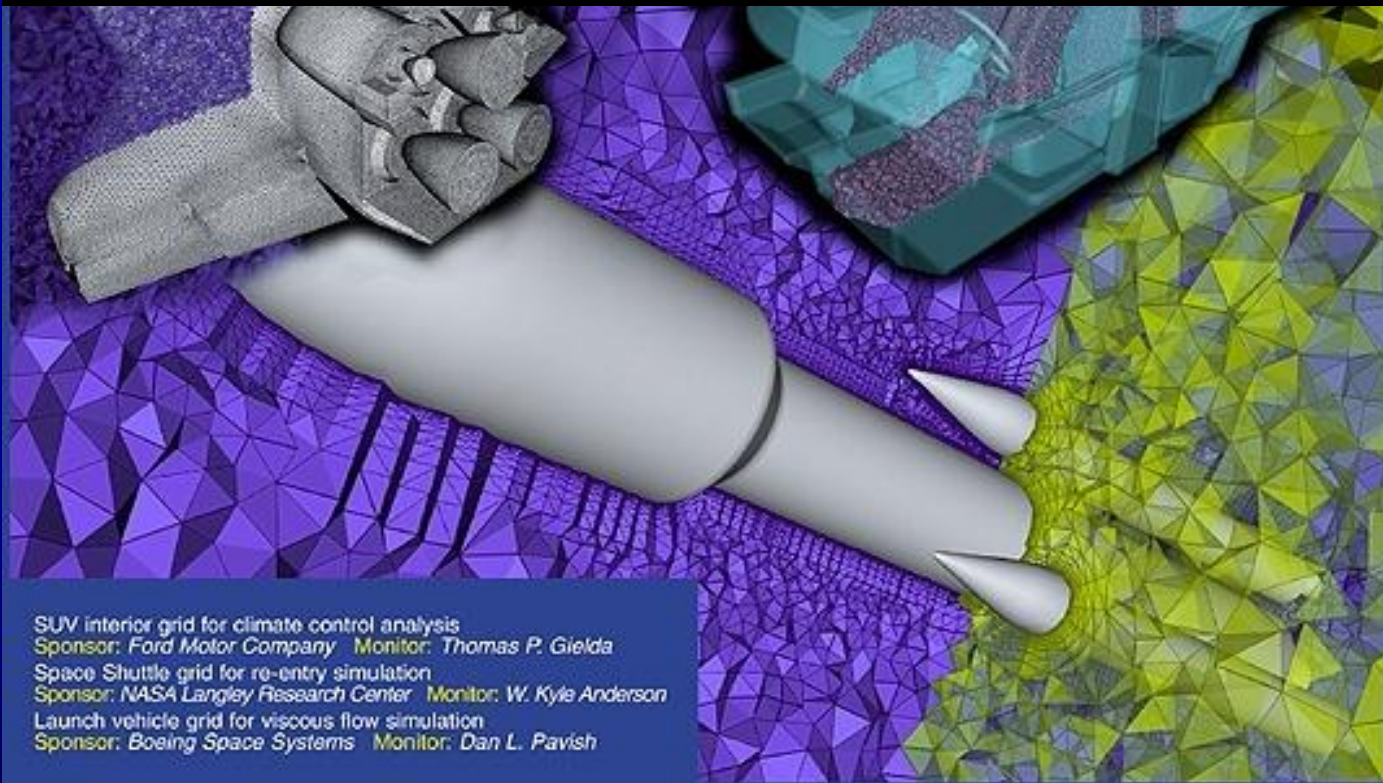
Types of Input Data

Curvilinear – regularly grided mesh
shaping function applied



Types of Input Data

Unstructured grid – irregular mesh typically composed of tetrahedra, prisms, pyramids, or hexahedra.



SUV interior grid for climate control analysis
Sponsor: Ford Motor Company Monitor: Thomas P. Gieda
Space Shuttle grid for re-entry simulation
Sponsor: NASA Langley Research Center Monitor: W. Kyle Anderson
Launch vehicle grid for viscous flow simulation
Sponsor: Boeing Space Systems Monitor: Dan L. Pavish

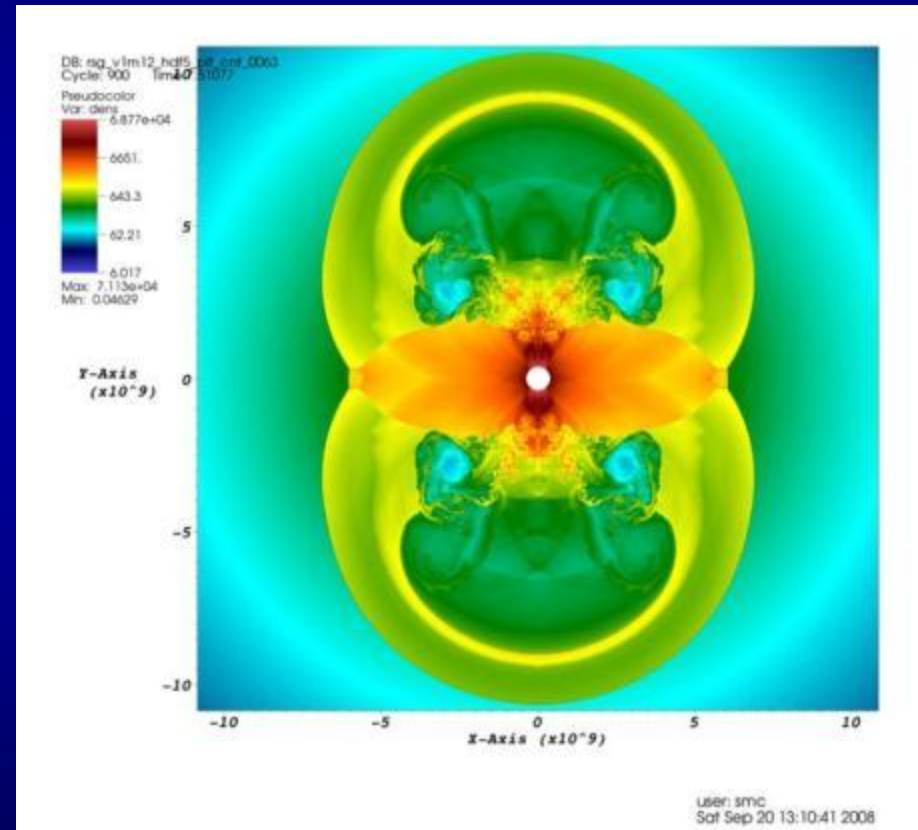
Visualization Operations

- Surface Shading (Pseudocolor)
- Isosurfacing (Contours)
- Volume Rendering
- Clipping Planes
- Streamlines

Surface Shading (Pseudocolor)

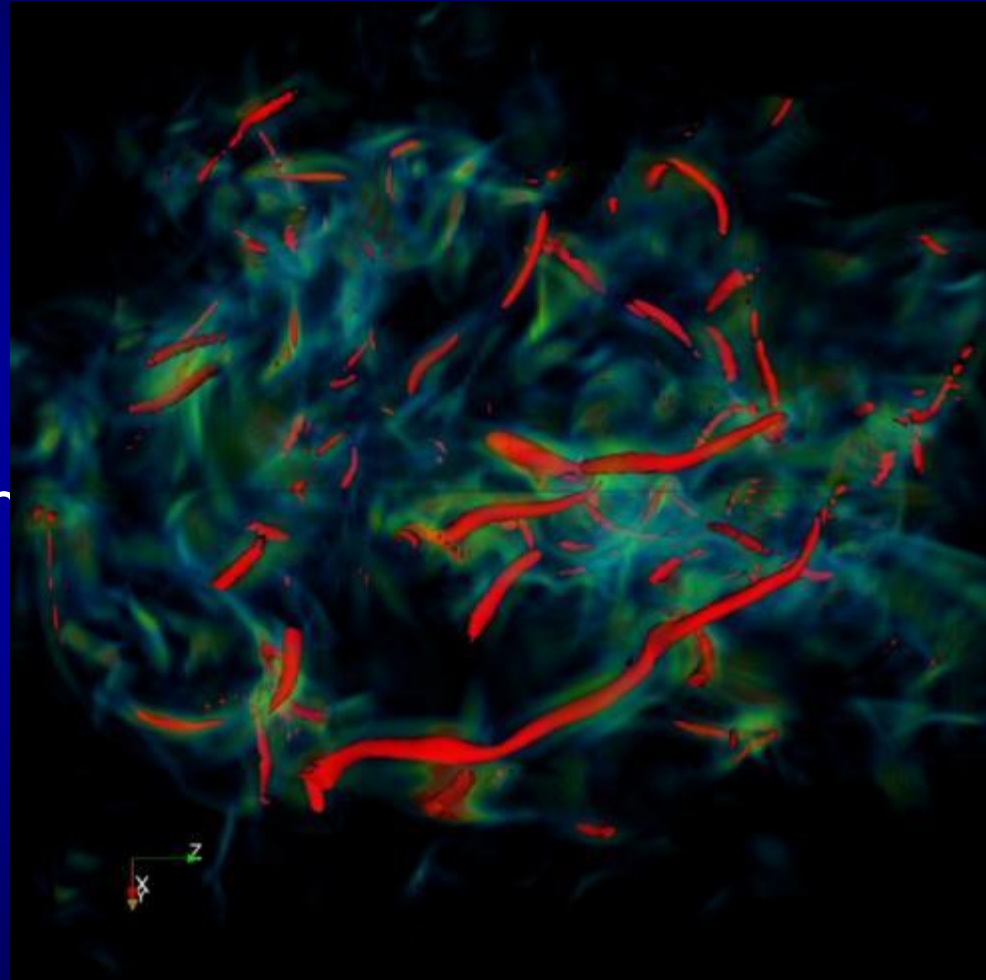
Given a scalar value at a point on the surface and a color map, find the corresponding color (and opacity) and apply it to the surface point.

Most common operation, often combined with other ops



Isosurfaces (Contours)

- Surface that represents points of constant value with a volume
- Plot the surface for a given scalar value.
- Good for showing known values of interest
- Good for sampling through a data range

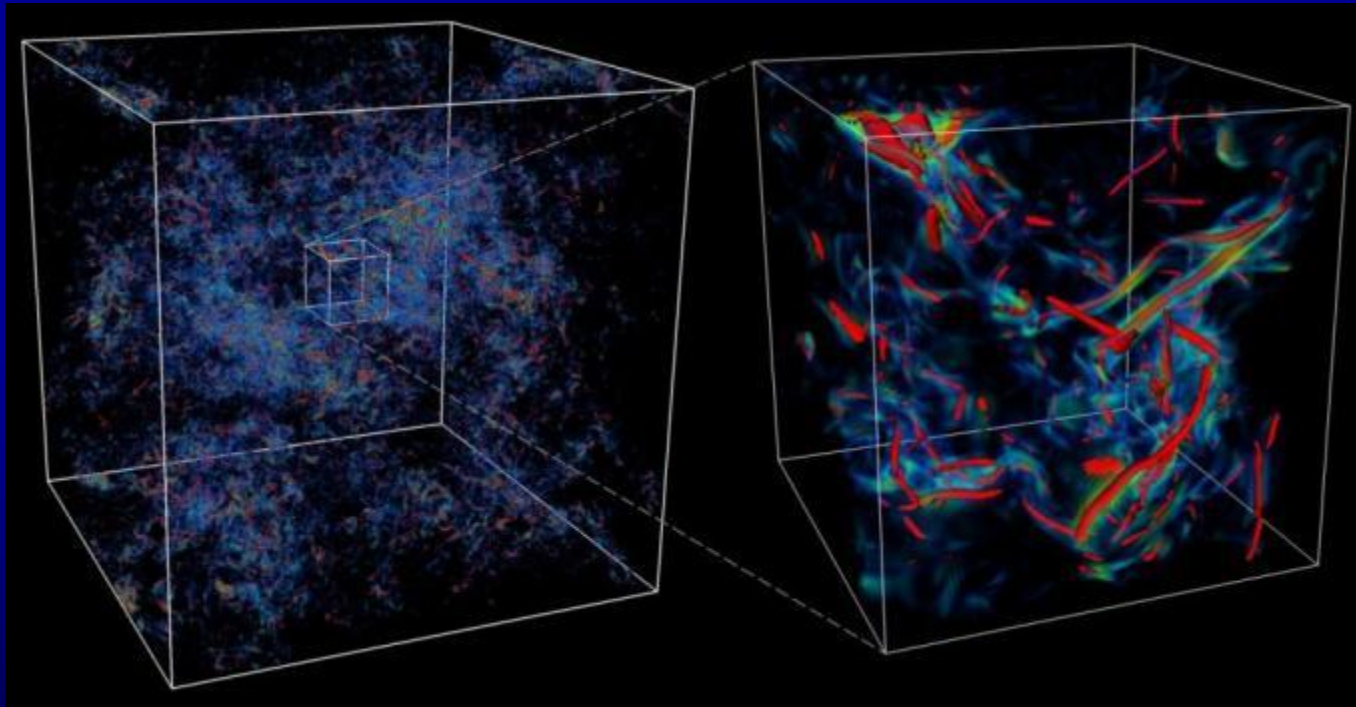


Volume Rendering

Expresses how light travels through a volume

Color and opacity controlled by transfer function

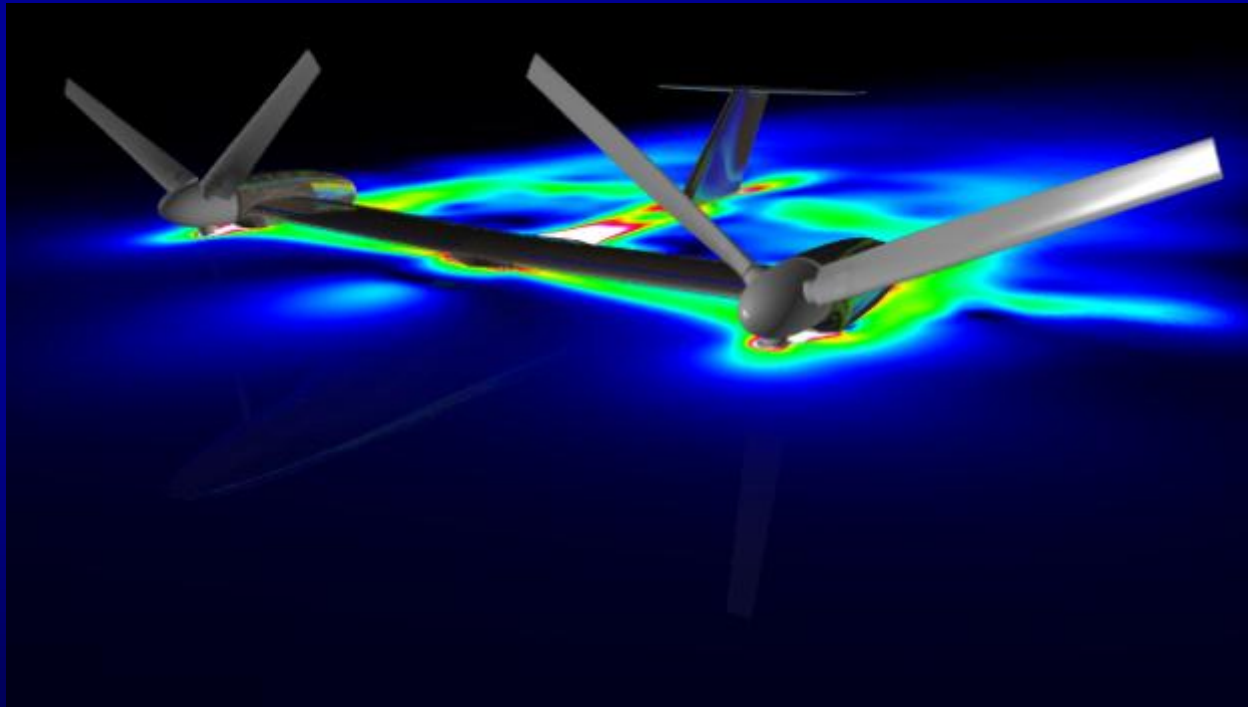
Smother transitions than isosurfaces



Clipping / Slicing Planes

Extract a plane from the data to show features

Hide part of dataset to expose features

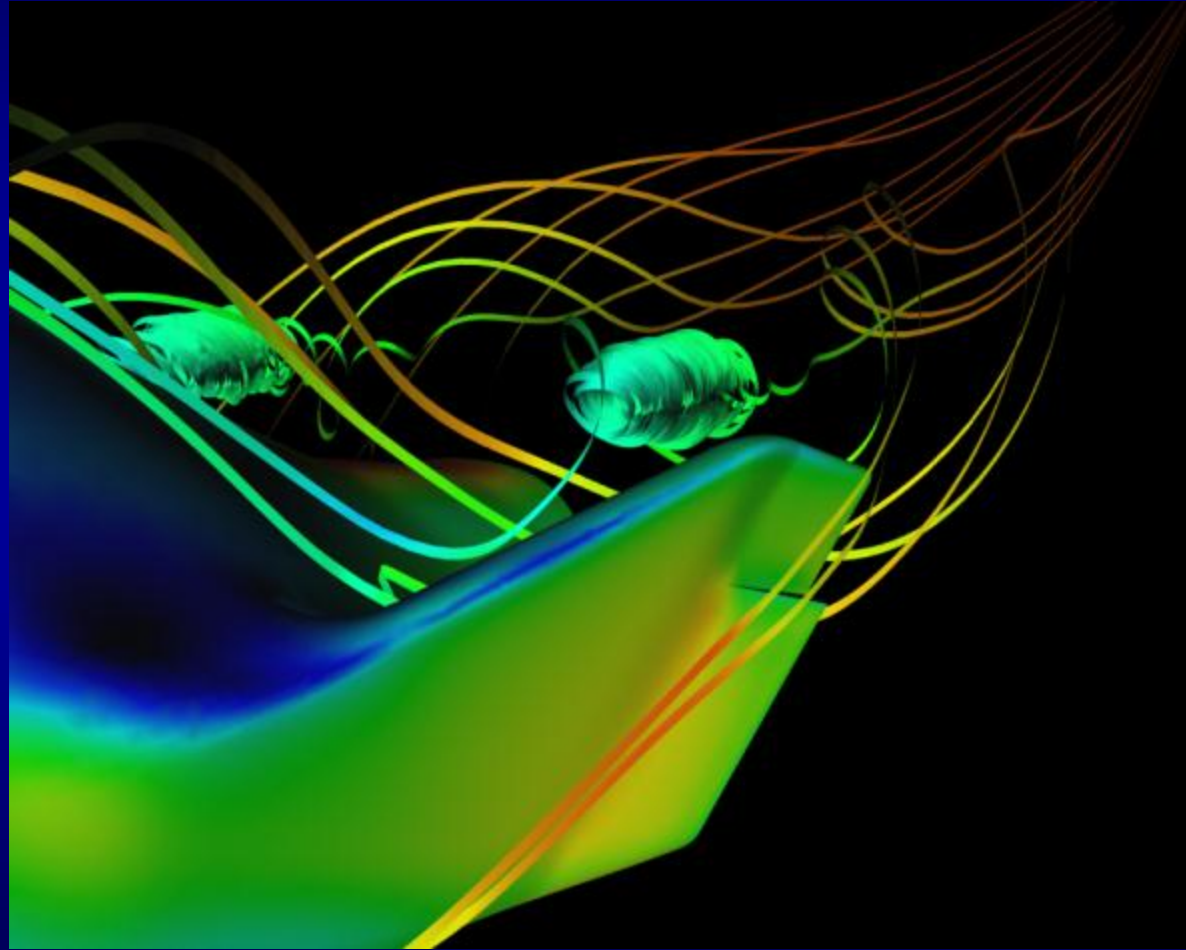


Particle Traces (Streamlines)

Given a vector field, extract a trace that follows that trajectory defined by the vector.

$$P_{\text{new}} = P_{\text{current}} + V_P \Delta t$$

Streamlines – trace in space
Pathlines – trace in time



Visualization Resources

- Personal machines
 - Most accessible, least powerful
- Projection systems
 - Seamless image, high purchase and maintenance costs
- Tiled-LCD displays
 - Lowest per-pixel costs, bezels divide image
- Remote visualization
 - Access to high-performance system, latency can affect user experience

XSEDE Visualization Resources

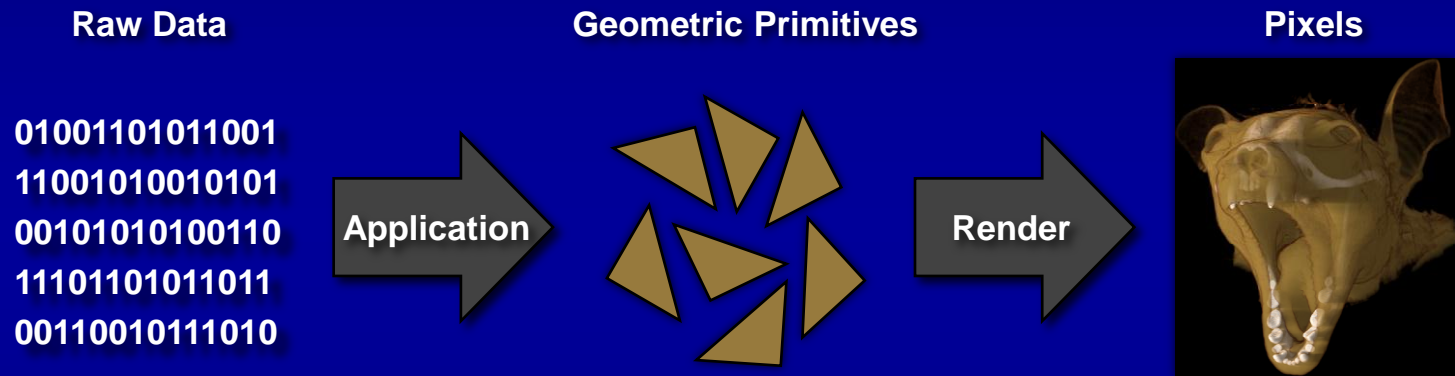
- Longhorn (TACC)
 - 256 nodes, 2048 total cores, 512 total GPUs
13.5 TB aggregate memory, QDR InfiniBand interconnect
 - Ranger file system mounted read-only across system
 - Longhorn Visualization Portal
 - <https://portal.longhorn.tacc.utexas.edu/>
 - Visualization job submission and monitoring
 - Remote, interactive, web-based visualization
 - Guided visualization using EnVision
- Spur (TACC)
 - 8 nodes, 128 total cores, 32 total GPUs
1 TB aggregate memory, SDR InfiniBand interconnect
 - Shares interconnect and file system with Ranger
- Lonestar (TACC)
 - 8 nodes, 96 total cores, 32 total GPUs (expand to 72 nodes soon) 192 GB aggregate memory, QDR InfiniBand interconnect
 - Shares interconnect and file system with Lonestar HPC nodes

TeraGrid Visualization Resources

- Nautilus (NICS)
 - SMP, 1024 Total Cores, 8 GPUs
 - 4 TB Global Shared Memory, SGI NUMalink 5 interconnect
- TeraDRE Condor Pool (Purdue)
 - 1750 Nodes, 14000 Total Cores, 48 Nodes with GPUs
 - 28 TB Aggregate Memory, no interconnect

Visualization Challenges

Visualization Allows Us to “See” the Science



But what about large, distributed data?

01001101011001 01001101011001
11001010010101 11001010010101
00101010100110 00101010100110
11101101011011 11101101011011
00110010111010 00110010111010

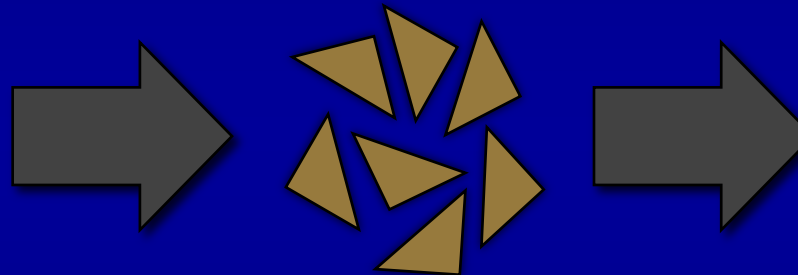
01001101011001 01001101011001
11001010010101 11001010010101
00101010100110 00101010100110
11101101011011 11101101011011
00110010111010 00110010111010

01001101011001 01001101011001
11001010010101 11001010010101
00101010100110 00101010100110
11101101011011 11101101011011
00110010111010 00110010111010

01001101011001 01001101011001
11001010010101 11001010010101
00101010100110 00101010100110
11101101011011 11101101011011
00110010111010 00110010111010

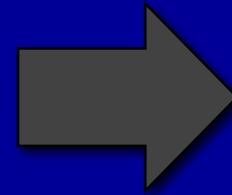
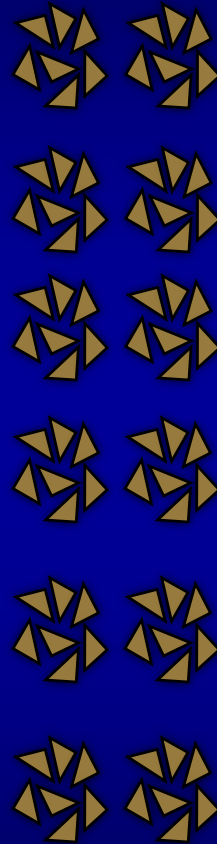
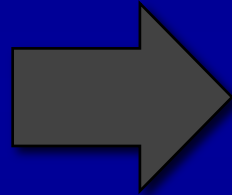
01001101011001 01001101011001
11001010010101 11001010010101
00101010100110 00101010100110
11101101011011 11101101011011
00110010111010 00110010111010

01001101011001 01001101011001
11001010010101 11001010010101
00101010100110 00101010100110
11101101011011 11101101011011
00110010111010 00110010111010



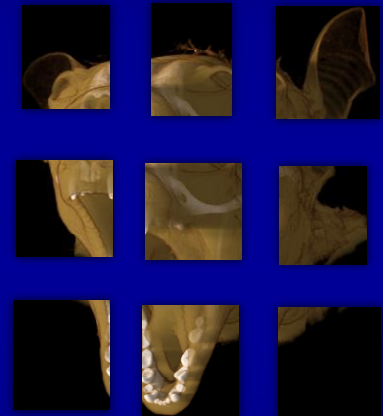
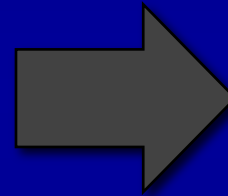
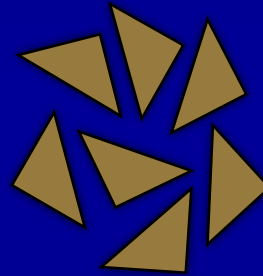
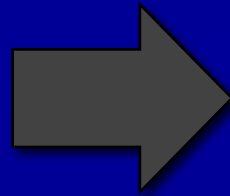
Or distributed rendering?

01001101011001
11001010010101
00101010100110
11101101011011
00110010111010



Or distributed displays?

01001101011001
11001010010101
00101010100110
11101101011011
00110010111010



Or all three?

01001101011001 01001101011001
11001010010101 11001010010101
00101010100110 00101010100110
11101101011011 11101101011011
00110010111010 00110010111010

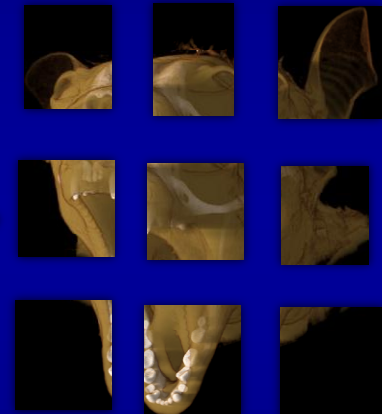
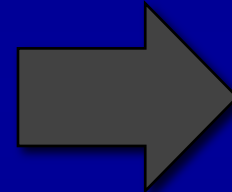
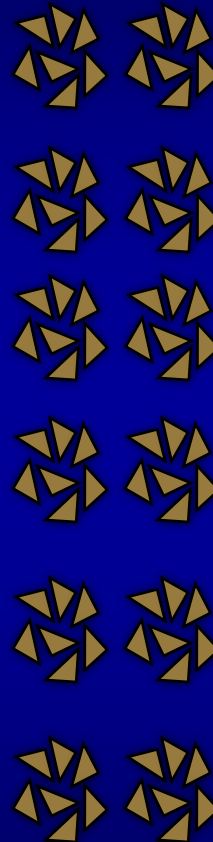
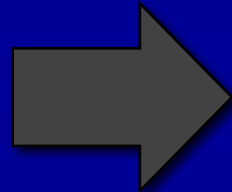
01001101011001 01001101011001
11001010010101 11001010010101
00101010100110 00101010100110
11101101011011 11101101011011
00110010111010 00110010111010

01001101011001 01001101011001
11001010010101 11001010010101
00101010100110 00101010100110
11101101011011 11101101011011
00110010111010 00110010111010

01001101011001 01001101011001
11001010010101 11001010010101
00101010100110 00101010100110
11101101011011 11101101011011
00110010111010 00110010111010

01001101011001 01001101011001
11001010010101 11001010010101
00101010100110 00101010100110
11101101011011 11101101011011
00110010111010 00110010111010

01001101011001 01001101011001
11001010010101 11001010010101
00101010100110 00101010100110
11101101011011 11101101011011
00110010111010 00110010111010

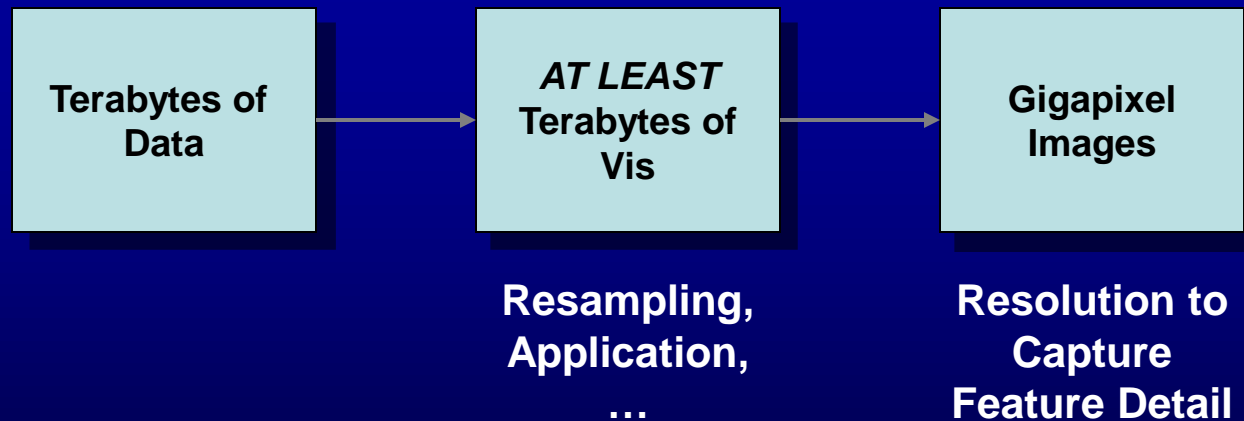


Visualization Scaling Challenges

- Moving data to the visualization machine
- Most applications built for shared memory machines, not distributed clusters
- Image resolution limits in some software cannot capture feature details
- Displays cannot show **entire** high-resolution images **at their native resolution**

Visualization scales with HPC

Large data produced by large simulations require large visualization machines and produce large visualization results



Moving Data

- How long can you wait?

File Size	10 Gbps	54 Mbps
1 GB	1 sec	2.5 min
1 TB	~17 min	~43 hours
1 PB	~12 days	~5 years

Analyzing Data

- Visualization programs **only beginning** to **efficiently** handle ultrascale data
 - 650 GB dataset -> 3 TB memory footprint
 - Allocate HPC nodes for RAM not cores
 - N-1 idle processors per node!
- Stability across many distributed nodes
 - Rendering clusters typically number $N \leq 64$
 - Data must be dividable onto N cores

Remember this when resampling!

Imaging Data



Image: NASA Blue Marble Project

Hypothetical fly-around movie

4096 x 2160 PNG ~ 10 MB
x 360 degrees ~ 3.6 GB

x 30 days ~ 108 GB
x 12 months ~ 1.3 TB

@ 10 fps 3.6 hours
@ 60 fps 36 min

Displaying Data



Dell 30" flat-panel LCD

4 Megapixel display

2560 x 1600 resolution

Displaying Data



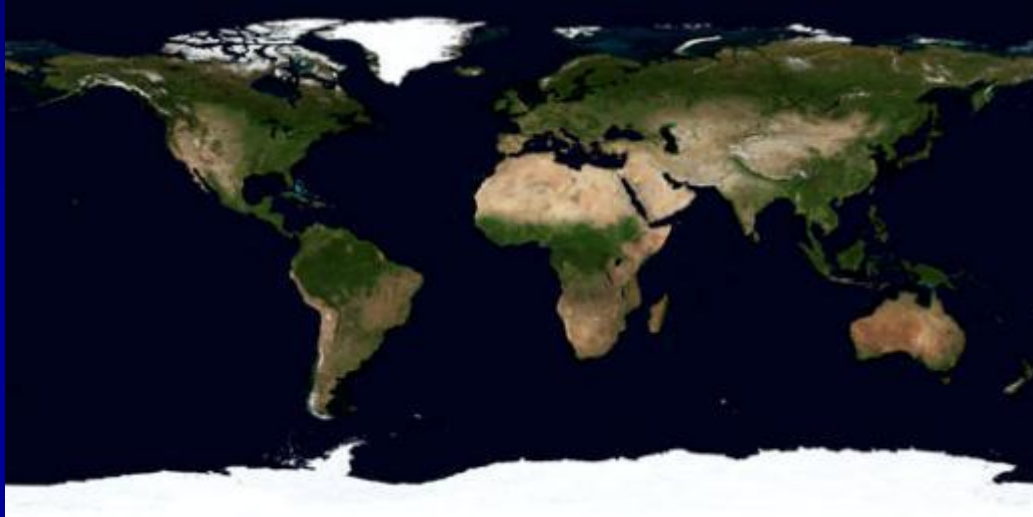
Stallion – currently world's highest-resolution tiled display

307 Megapixels
38400 x 8000 pixel resolution



Dell 30" LCD

Displaying Data



NASA Blue Marble
0.5 km² per pixel

3732 Mpixel
(86400 x 43200)



Stallion – 307 Mpixel (38400 x 8000)

- Dell 30" LCD – 4 Mpixel (2560 x 1600)

Solution by Partial Sums

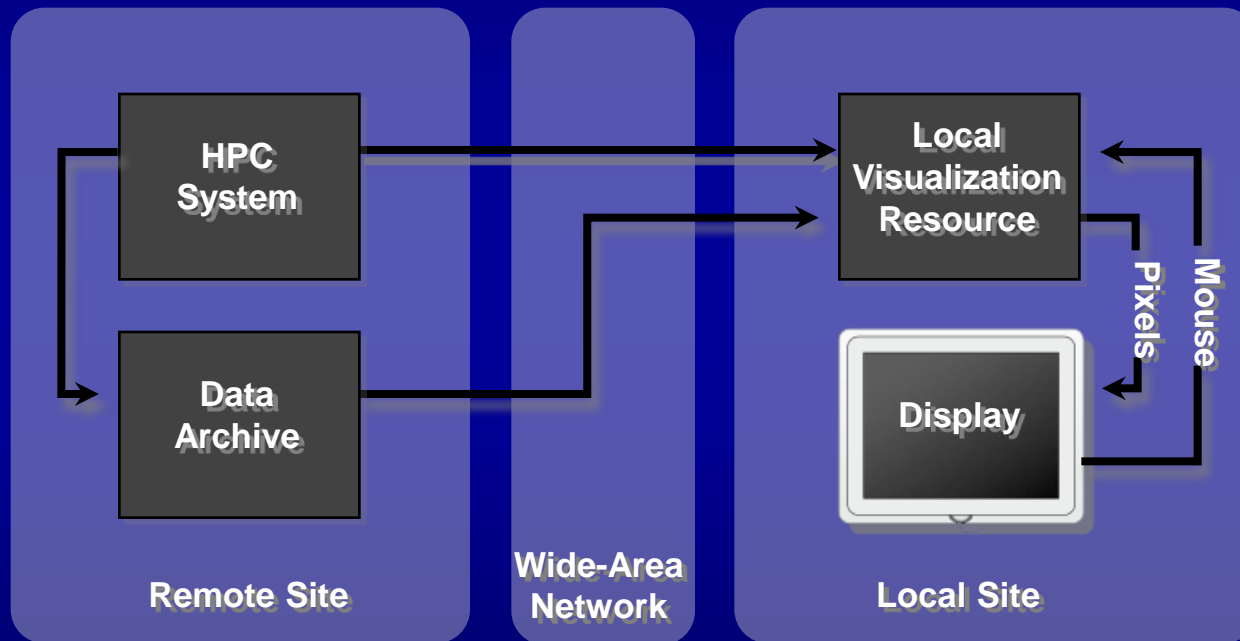
- Moving data – integrate vis machine into simulation machine. **Move the machine** to data!
 - Ranger + Spur: shared file system and interconnect
- Analyzing data – create larger vis machines and develop more efficient vis apps
 - Smaller memory footprint
 - More stable across many distributed nodes

Until then, **the simulation machine is the vis machine!**

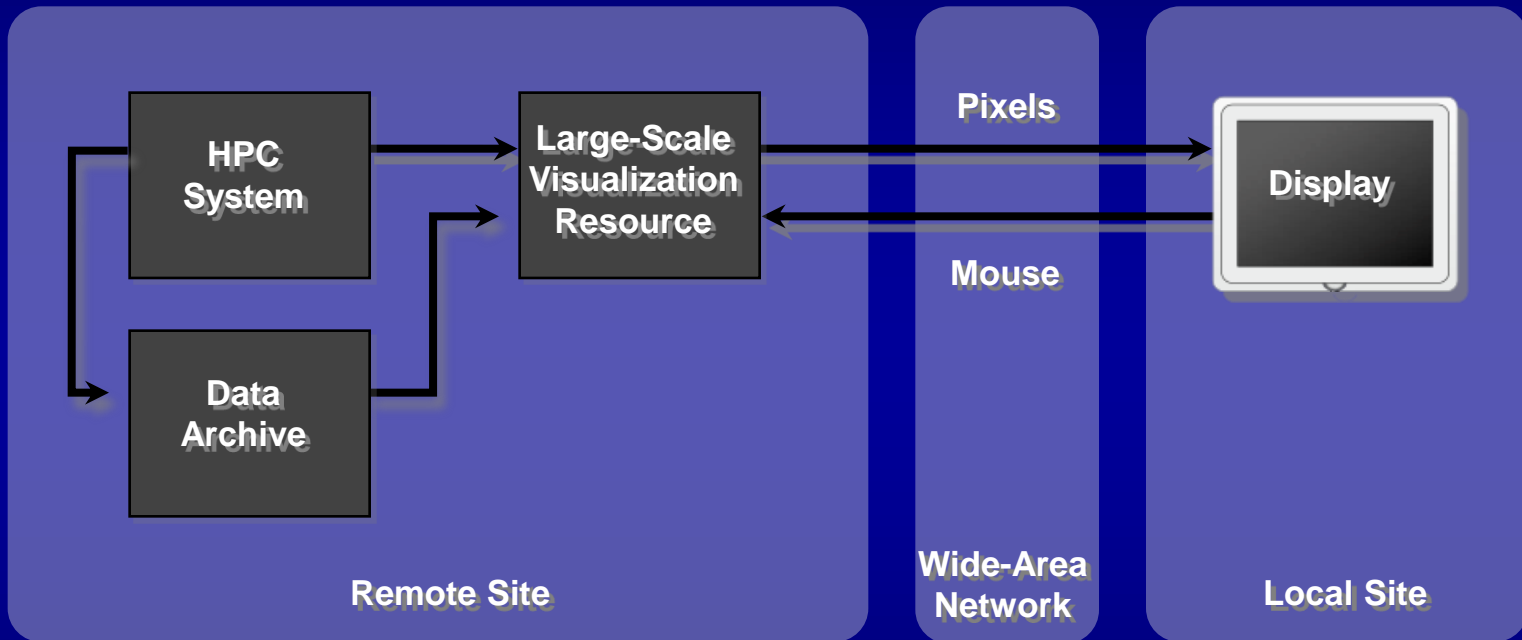
Solution by Partial Sums

- Imaging data – focus vis effort on interesting features
parallelize image creation
 - Feature detection to determine visualization targets
but can miss “unknown unknowns”
 - Distribute image rendering across cluster
- Displaying data – high resolution displays
multi-resolution image navigation
 - Large displays need large spaces
 - Physical navigation of display provides better insights

Old Model (No Remote Capability)



New Model Remote Capability



Using GPUs

- More than for just rendering!
 - HPC applications and Visualization algorithms
- Parallelism – kernel should be highly SIMD/SIMT
 - Switching kernels is expensive!
 - Fermi *hardware* supports multiple kernel execution
- Control Flow – avoid conditionals in kernels
 - Implemented with predication, harms utilization
- Job size – high workload per thread + many threads
 - amortize thread initialization and memory transfer costs
 - GPU is a throughput machine, must keep it busy!
- Memory footprint – task must decompose well
 - local store per GPU core is low (16 KB on Longhorn)
 - card-local RAM is limited (4GB on Longhorn)
 - access to system RAM is slow (treat like disk access)
- More on this Thursday during CUDA for HPC

Summary

- Challenges at every stage of visualization when operating on large data
- Partial solutions exist, though not integrated
- Problem sizes continue to grow at every stage
- Vis software community must keep pace with hardware innovations



Thank you!

pnav@tacc.utexas.edu