I/O Efficient Analysis and Visualization of Overset Grid Results

A presentation for the 11th Symposium on Overset Composite Grids and Solution Quality

Scott T. Imlay, Durrell K. Rittenberg, Craig A. Mackey, and Christopher C. Nelson

tecplot Master the View

Outline

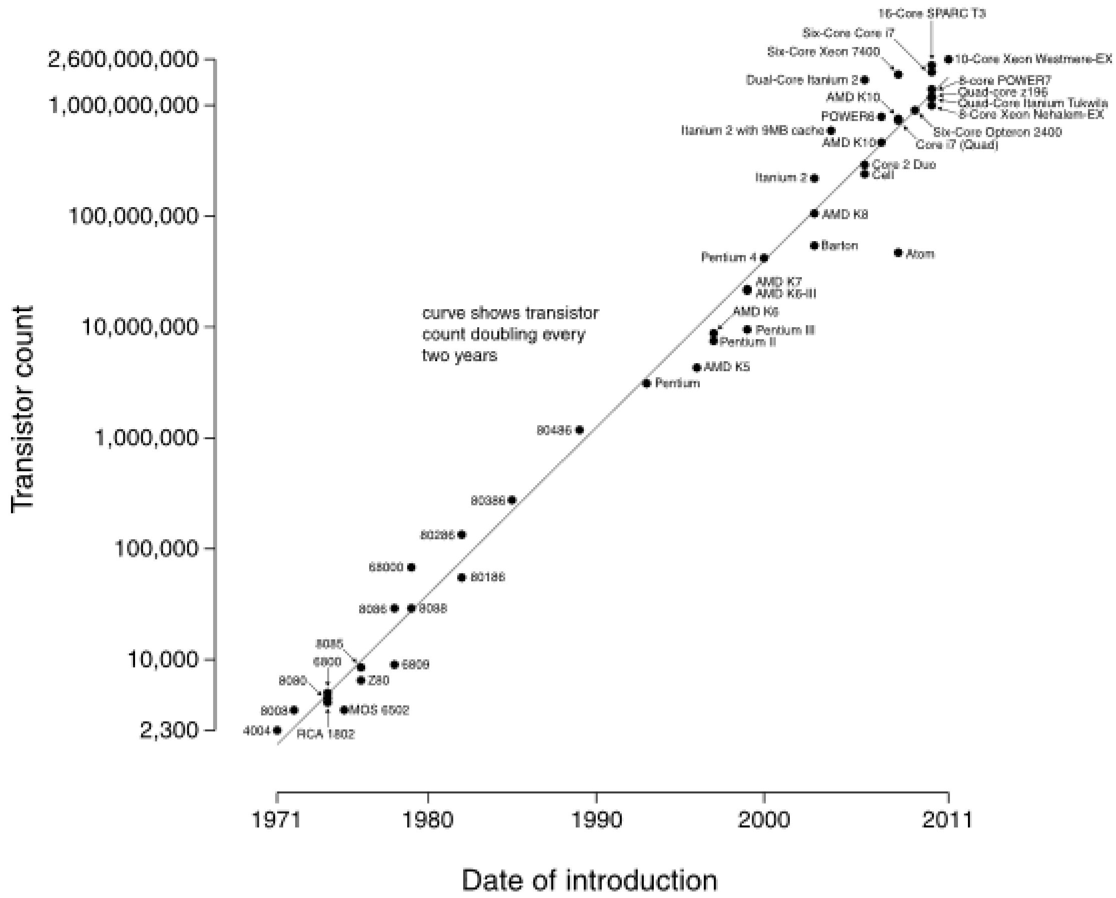
- Trends (hardware and data)
- Proposed solution
- Results

- Goal: Improve performance for large remote data.

tecplot Master the View

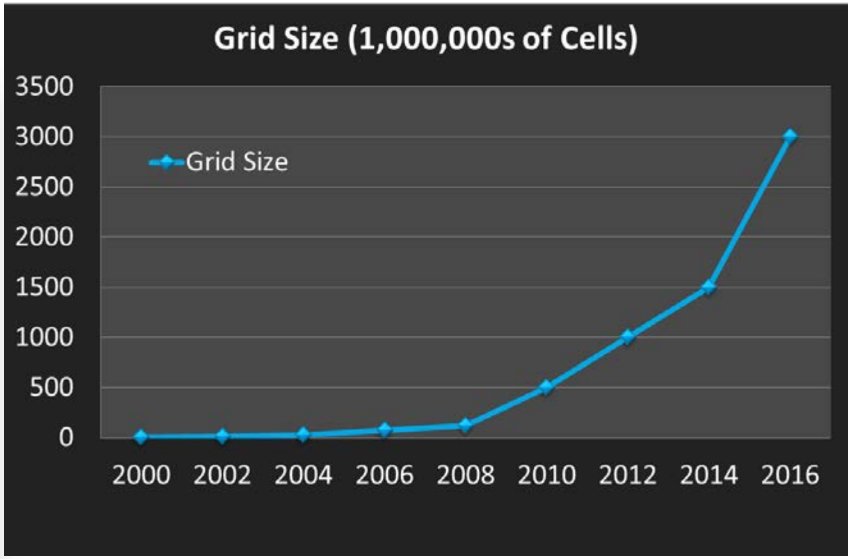
Driving Force is Moore's Law

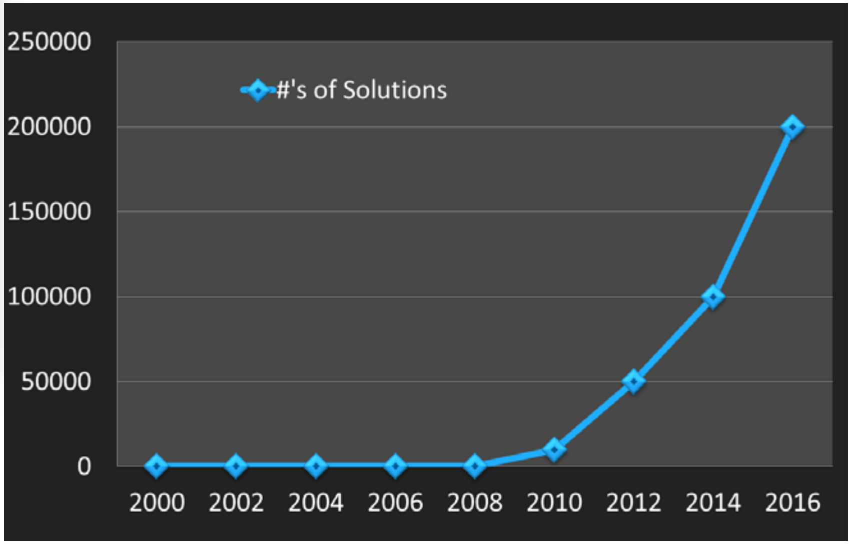
Microprocessor Transistor Counts 1971-2011 & Moore's Law



tecplot. Master the View

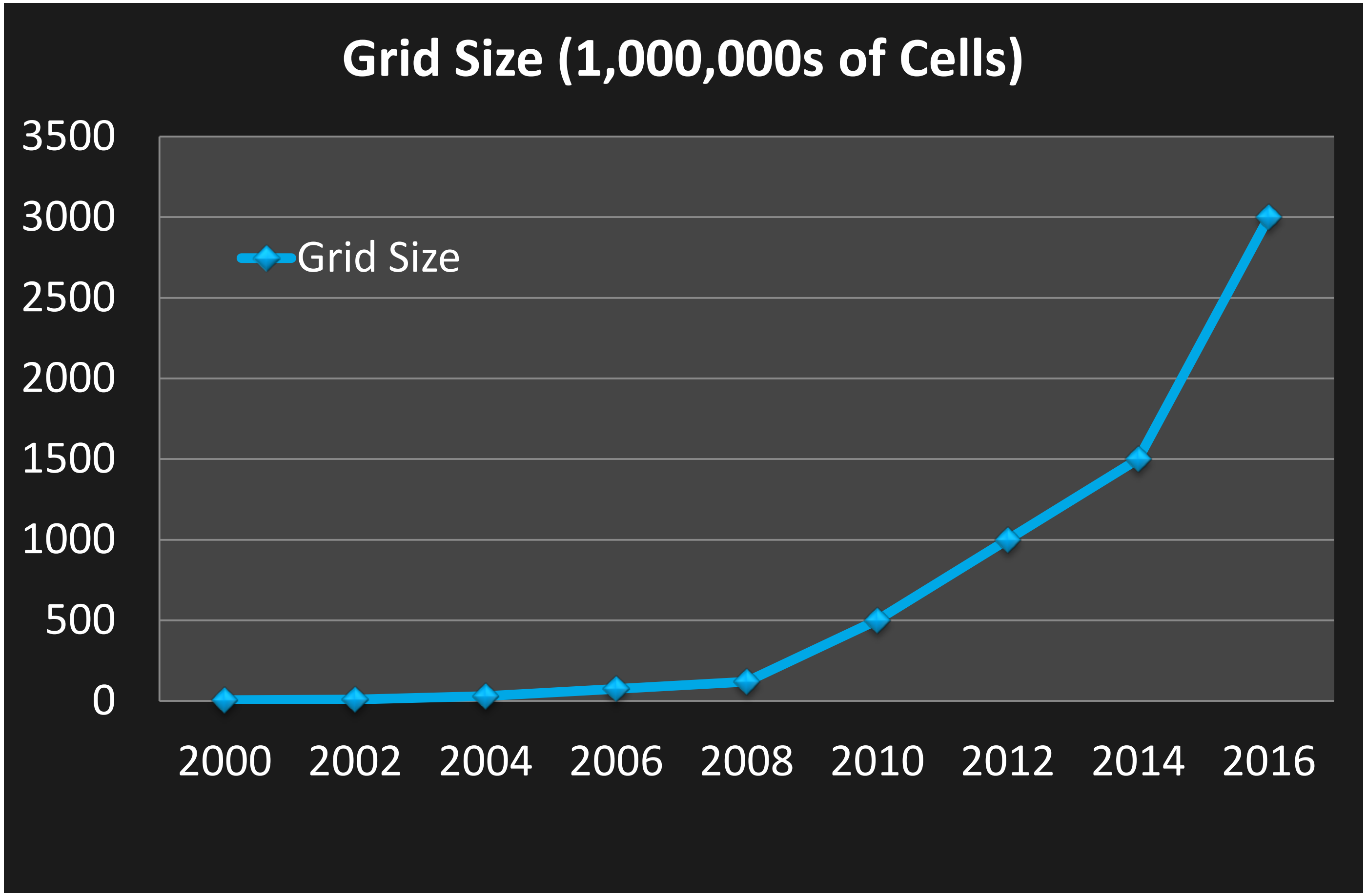
Impact of Low Cost Computational Resources





tecplot. Master the View

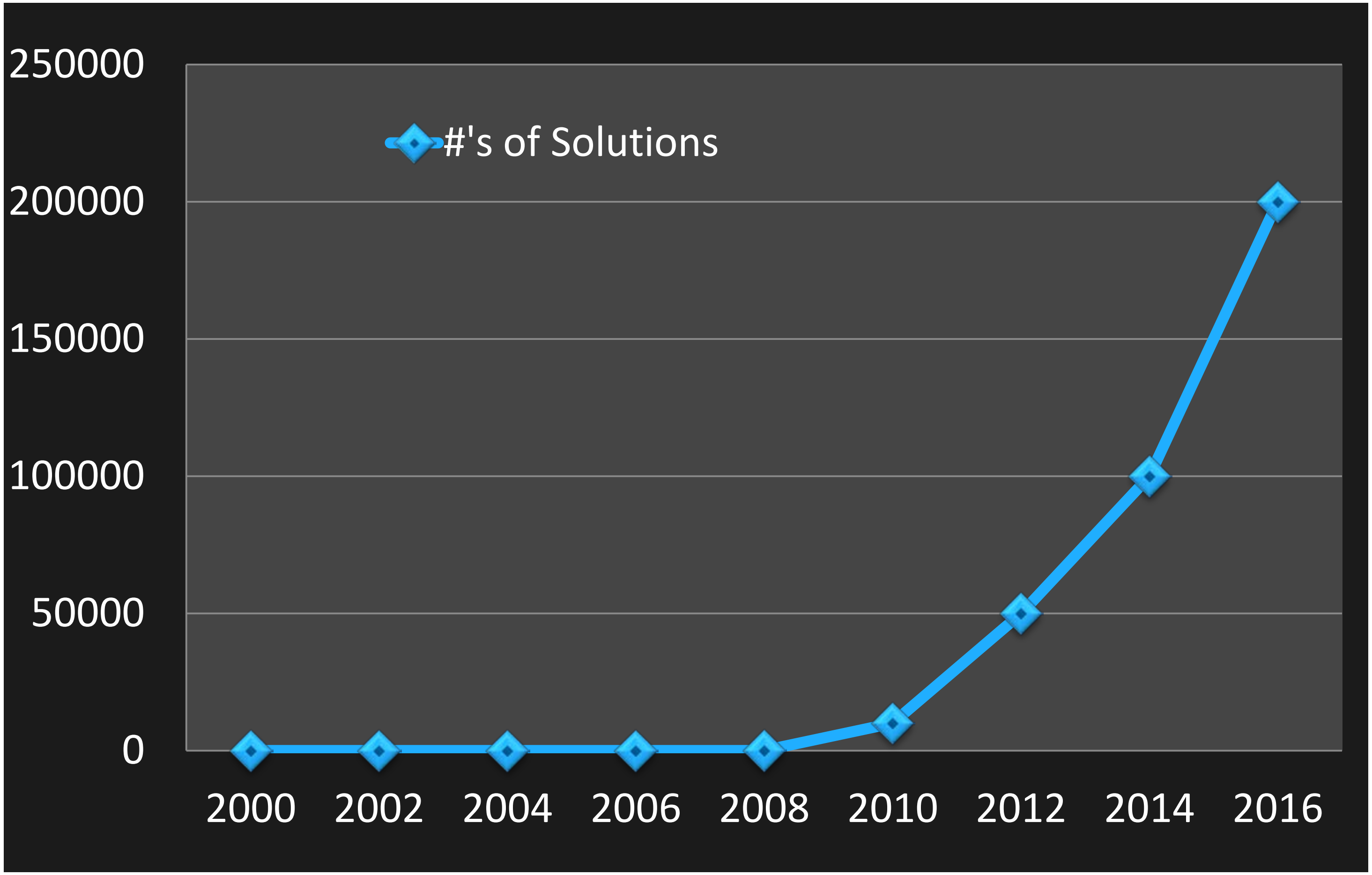
Impact of Low Cost Computational Resources



Engineers are tackling more interesting problems that require better resolution of physical phenomena

tecplot. Master the View

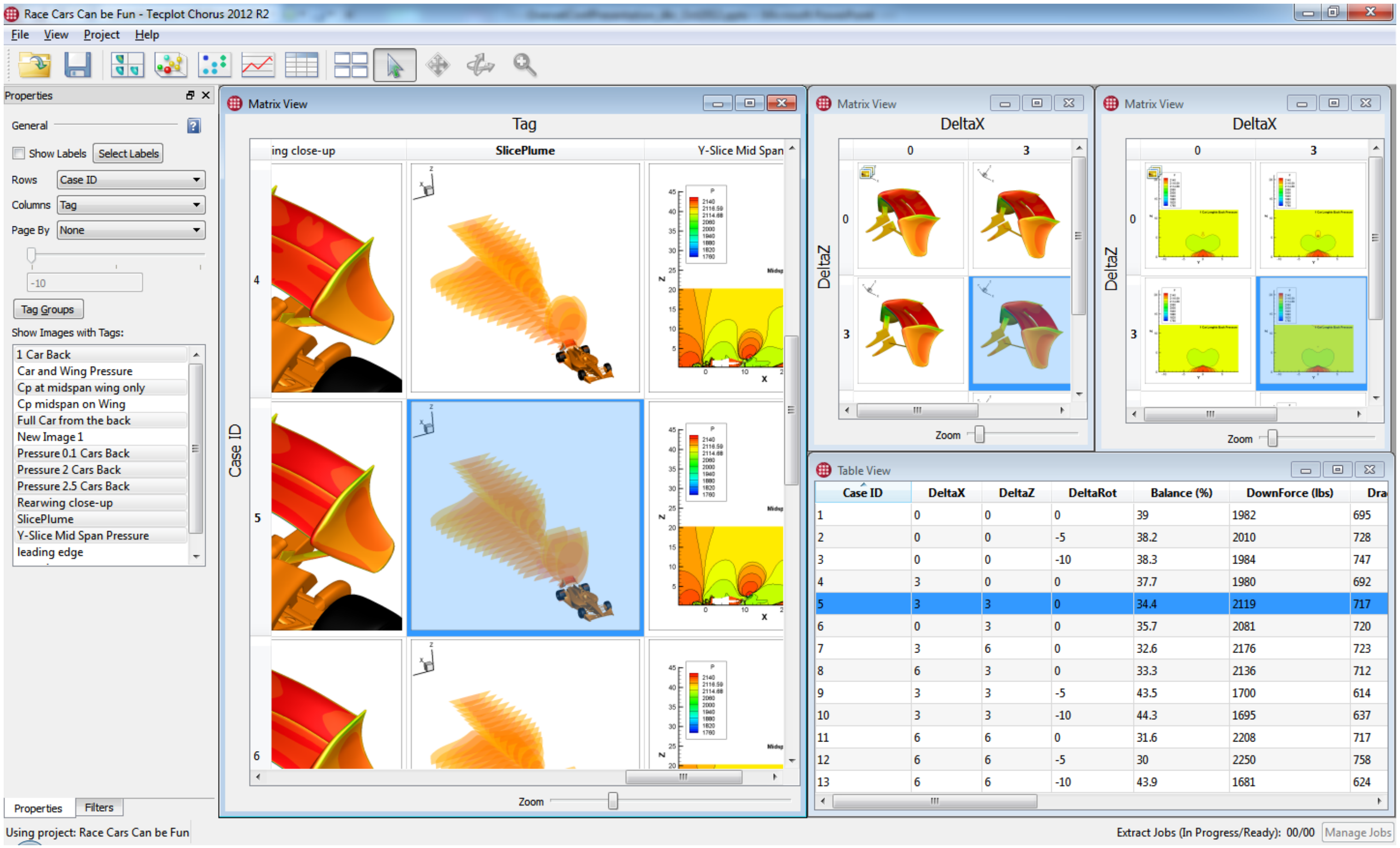
Impact of Low Cost Computational Resources



Engineers using High Fidelity simulation to predict performance of vehicles and optimizations

tecplot® Master the View

Tecplot Chorus For Collections of CFD Results



Evaluating overall system performance and allowing engineers to compare results quickly

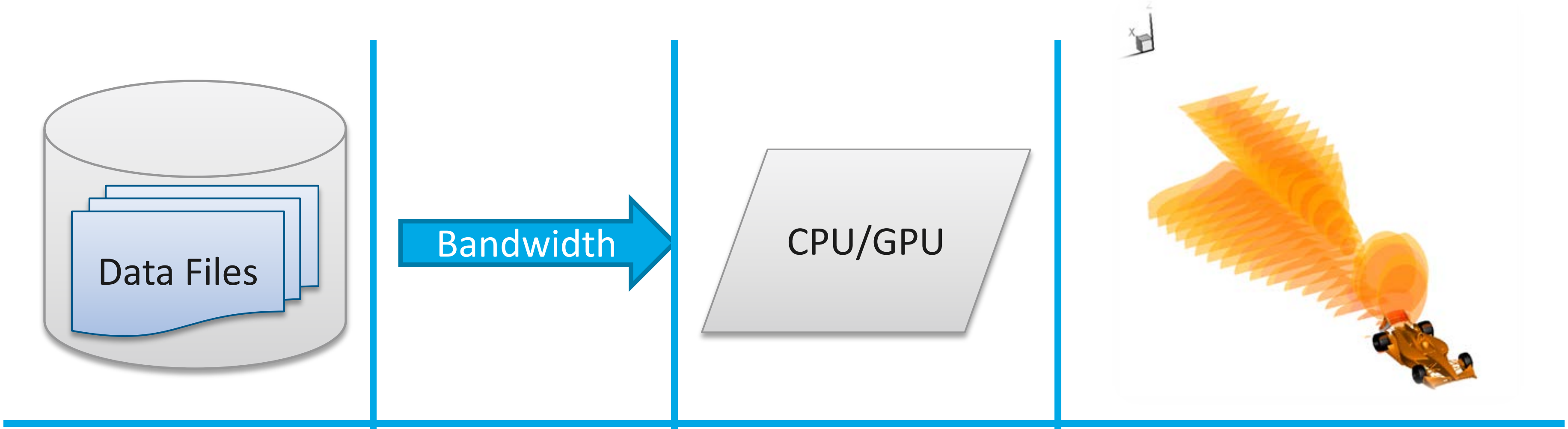
tecplot
Master the View

Field-Data Operations for Parametric Data

- Common Field-Data Operations
  - Comparative analysis (what has changed)
  - Integrations (forces and moments)
  - Extractions (snapshots)
  - Aggregations (statistics, box-plots)
- Operations over enormous amounts of data
  - Example: Aero database development
    - Thousands of 100M Cell CFD solutions
  - Some aggregations require data from all solutions to be loaded simultaneously
    - If not clever, work through equivalent of 100B Cells
- Exaggerates the performance issues resulting from the Red Shift

tecplot Master the View

Data Processing Pipeline

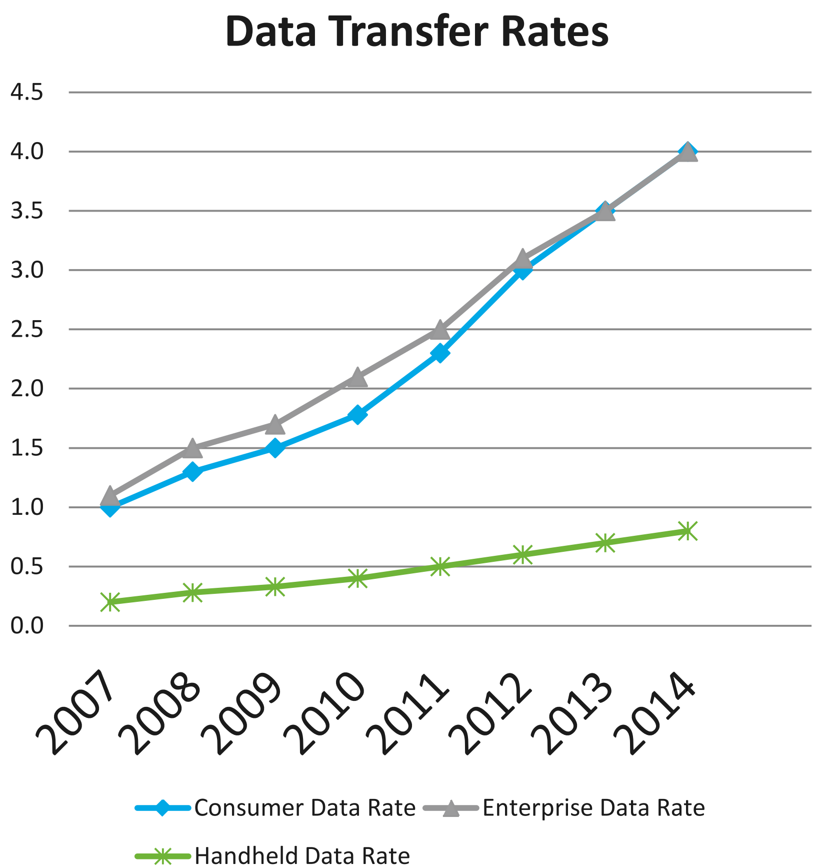


Double every 36 months	Double every 16 months	Double every 18 months	

Data IO is the current rate determining step in the visualization pipeline.

tecplot Master the View

Hard-drive Load Times Dominate

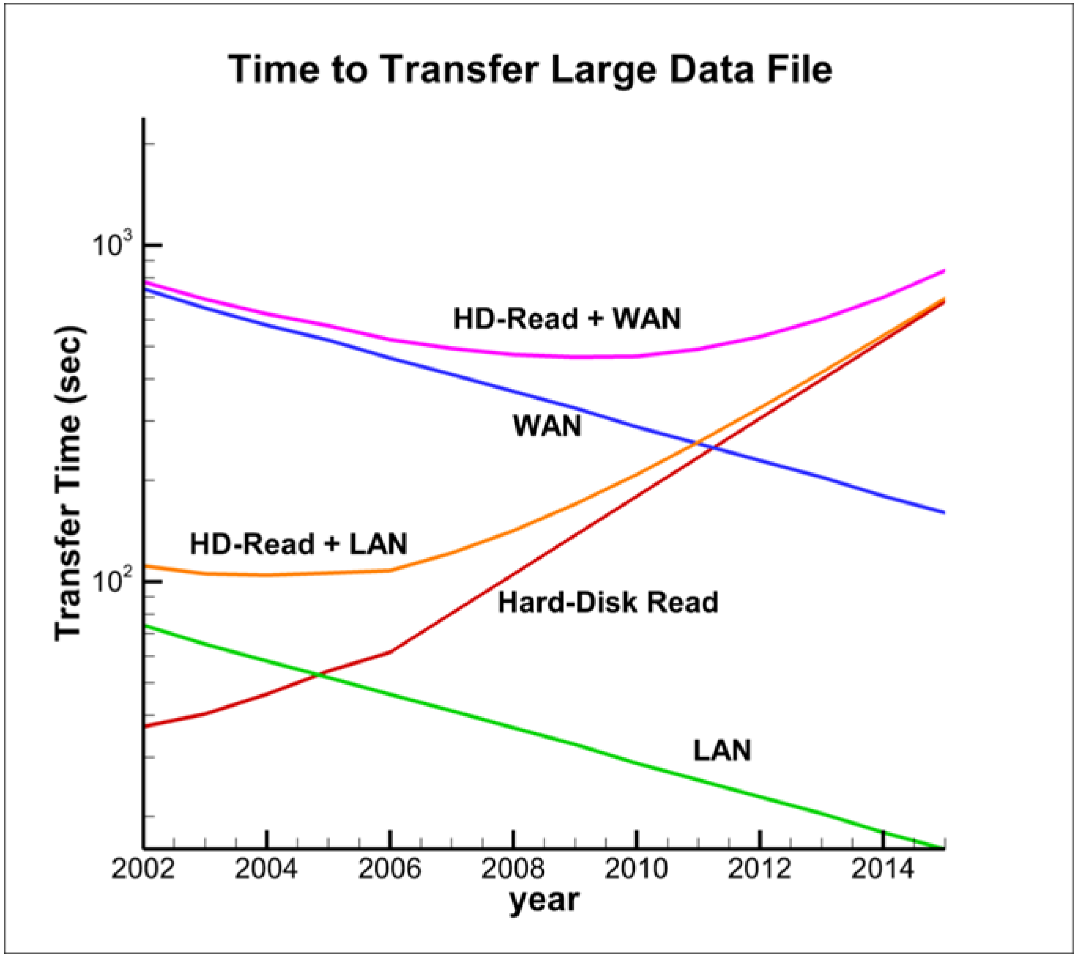


- Disk Capacity is doubling ever 12 months
- Disk I/O doubling ever 36 months

tecplot® Master the View

Consequence of Red Shift

- Current visualization architectures will perform worse as time goes on!



tecplot Master the View

Overcoming Data Transfer Bottleneck
Popular Approaches in Industry

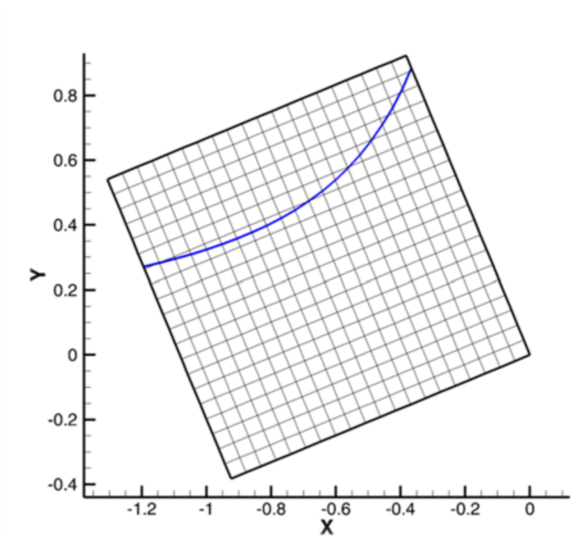
- Hardware/System Improvements
  - Parallel file systems (delay problem, but can't outgrow Moore's law by adding spindles)
  - New types of memory
    - SSD (maybe, but probably expensive for many of our customers)
    - Holographic memory, etc. (wouldn't that be lovely)
- In Situ visualization
  - Link libraries into CFD code to extract desired data or images (Don't save volume data)
  - Circumvents the disk transfer rate bottleneck
  - What about aggregations and data mining?
- Parallel visualization
  - Doesn't entirely solve disk transfer rate problem
  - May help some if it uses efficient parallel data reads
  - Red Shift doesn't need more compute power!

Overcoming Data Transfer Bottleneck
Our Solution

- Reduce the amount of data you read!
  - Must scale sub-linearly with the size of the grid
- Subzone Load-on-Demand (SZLoD)
  - Save indexed volume data file
  - Load only the data you need (Lazy Loading)
  - Related work
    - Out-of-Core algorithms of the 1990's
    - Field Encapsulation library of Patrick Moran at NASA Ames
      - Patrick Moran, et. al. "Field Encapsulation Library: The FEL 2.2 User Guide", NAS Technical Report NAS-00-002. NASA Ames Research Center, January, 2000

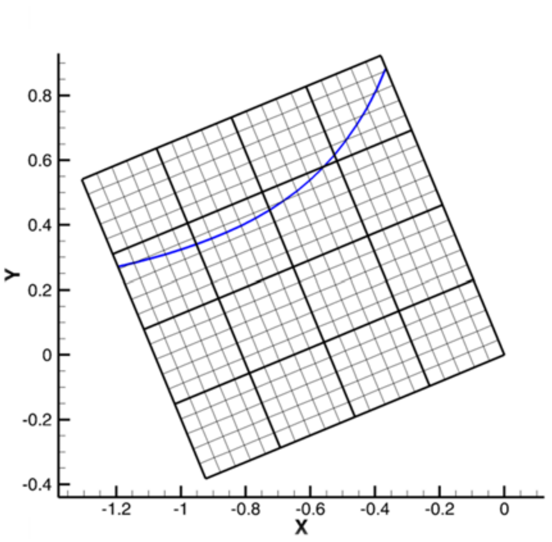
tecplot® Master the View

How Does SZLoD Work?



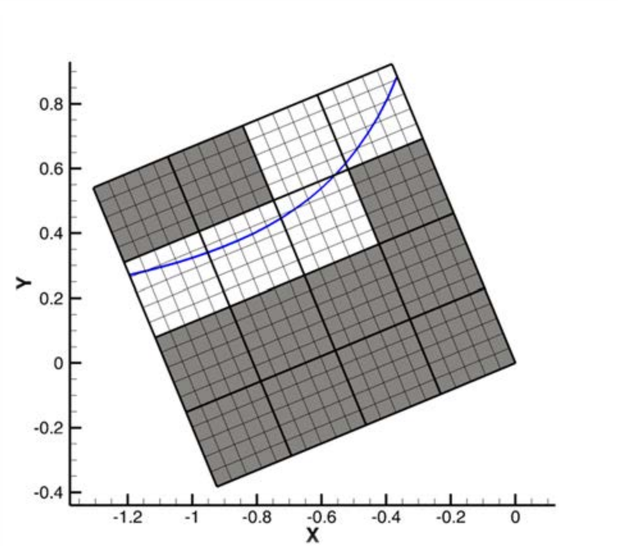
Example 2D Contour Line

- Current Methodologies require loading data for zone
- For Large data loading can be time intensive



Domain can be indexed

- Decomposition of domain into smaller subdomains
- These subdomains can be indexed

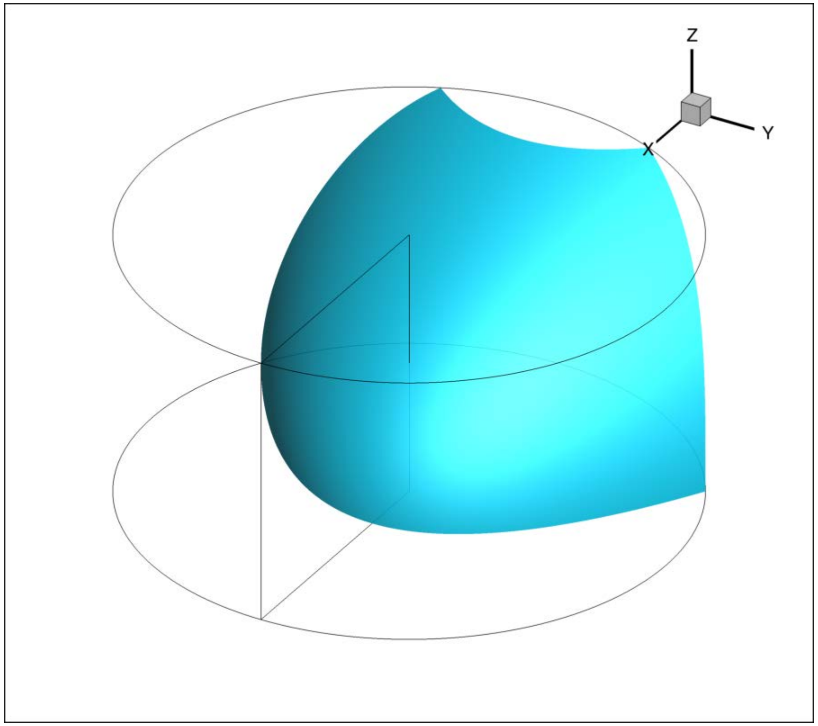


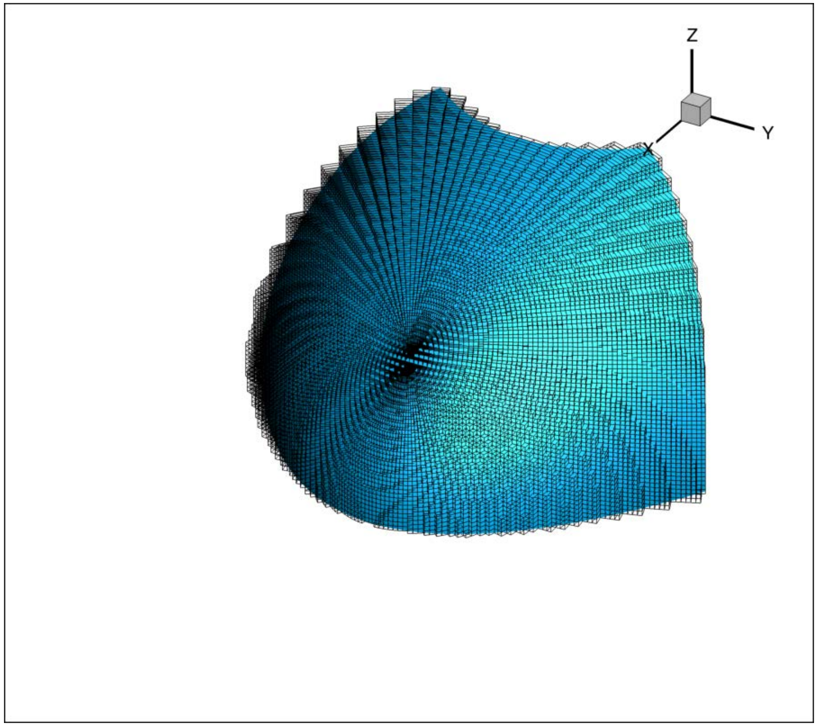
Data Required for Line 5/16 of total data

- Loading time reduced
- Memory requirements reduced

tecplot® Master the View

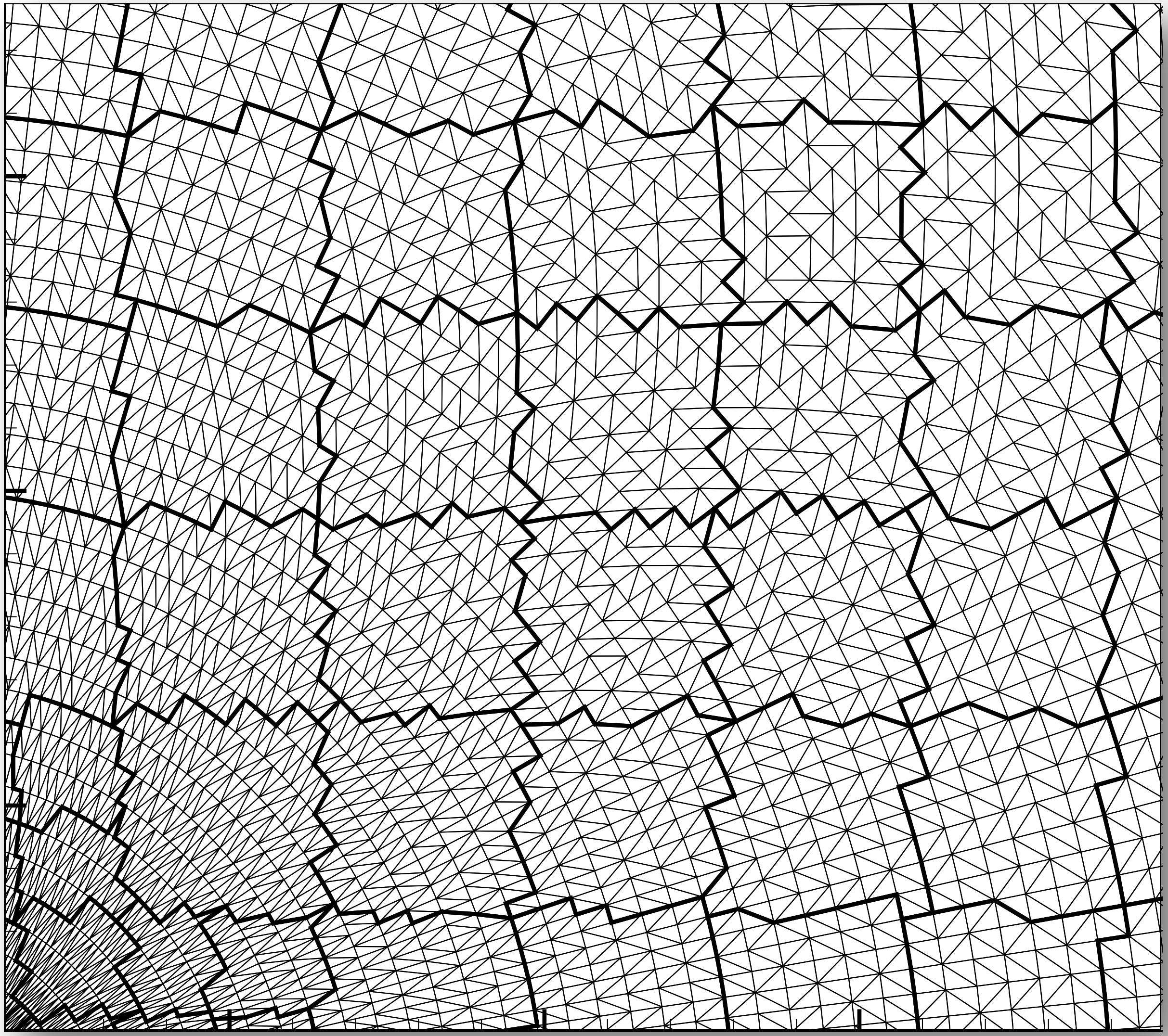
SZLoD Similar in 3D





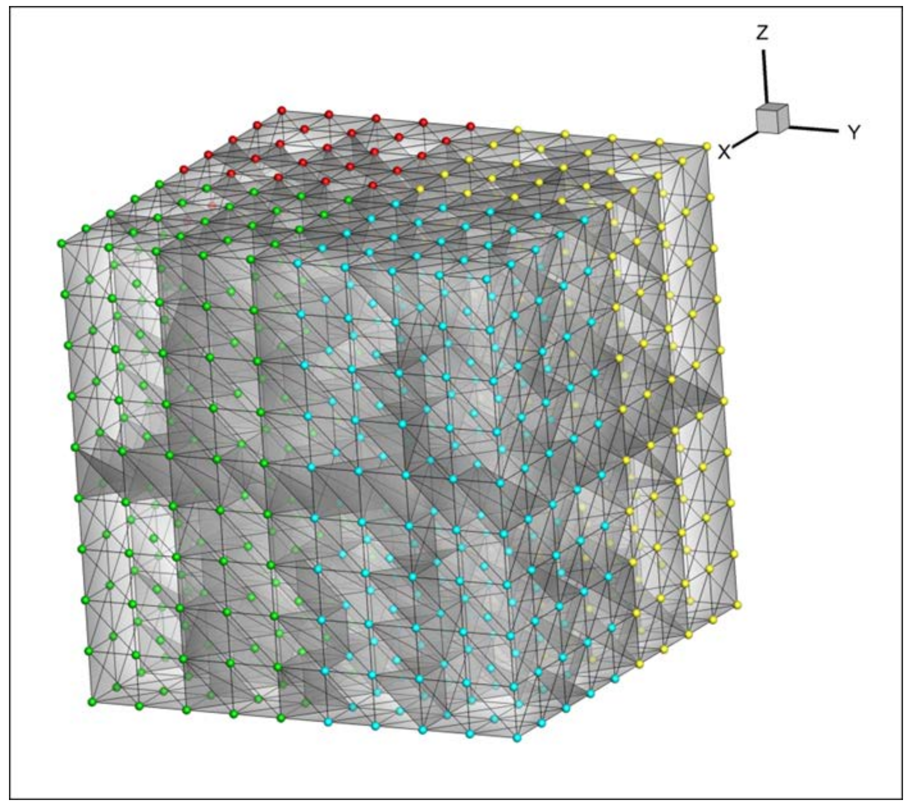
- The indexed decomposition can be extended to 3D for iso-surfaces, slices and streamtraces

SZLoD Extended to Unstructured Data



3D Unstructured-Grid Domain Decomposition

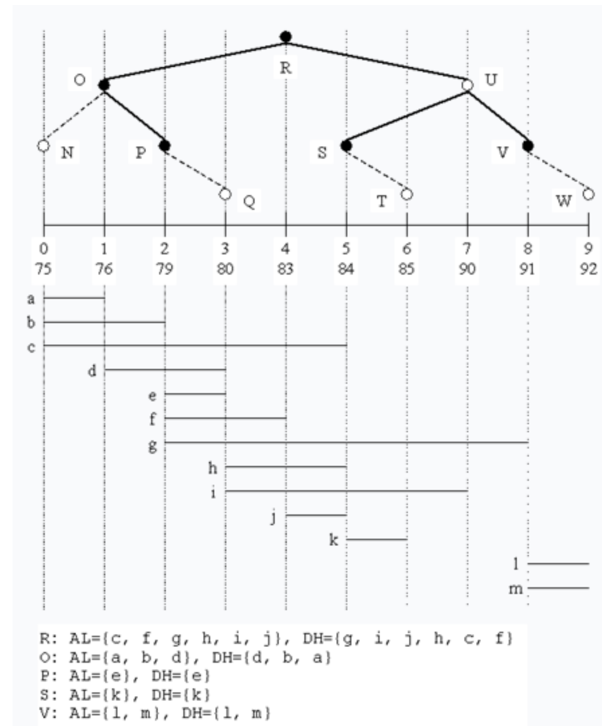
Subdivision using Recursive Orthogonal Bisection



Indexing for Subzone Selection - Interval Tree

Binary tree of intervals (value ranges)

- Return all intervals that contain a specified value of the variable
- 255 cells per subzone
- Query is O(log(N))

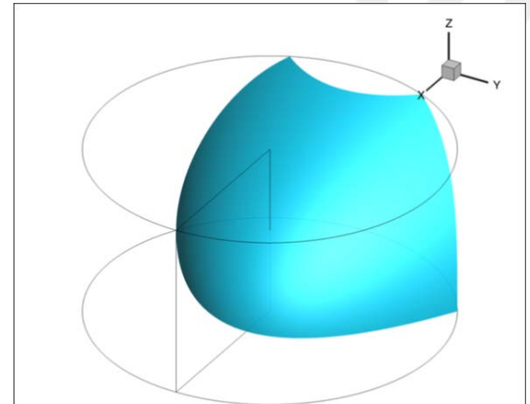


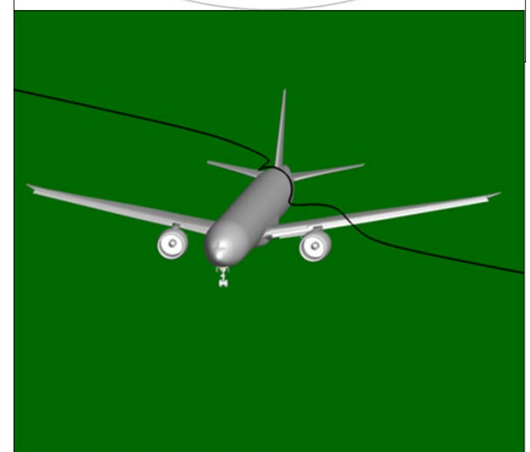
Grid Size (Cells)	Size (subzones)	Query (no tree)	Query (tree)	Tree file size
1B	4M	17ms	0.12ms	62.8MB
10B	40M	160ms	1.4ms	620MB

Master the View

Test Cases

- Synthetic test dataset
  - Scaling up to a billion cells
- Transport aircraft
  - 187 Million cell finite-element grid
- Unsteady wind-turbine analysis
  - Overflow results

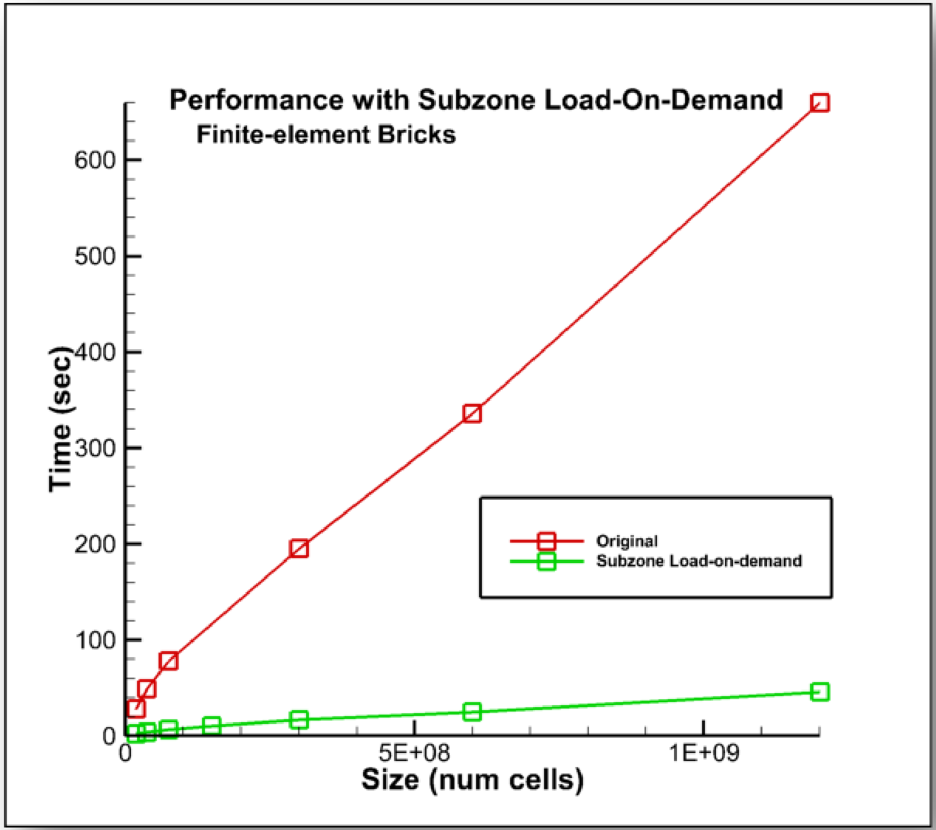




Master the View

Preliminary Results of **1 Billion** Cell Test Case

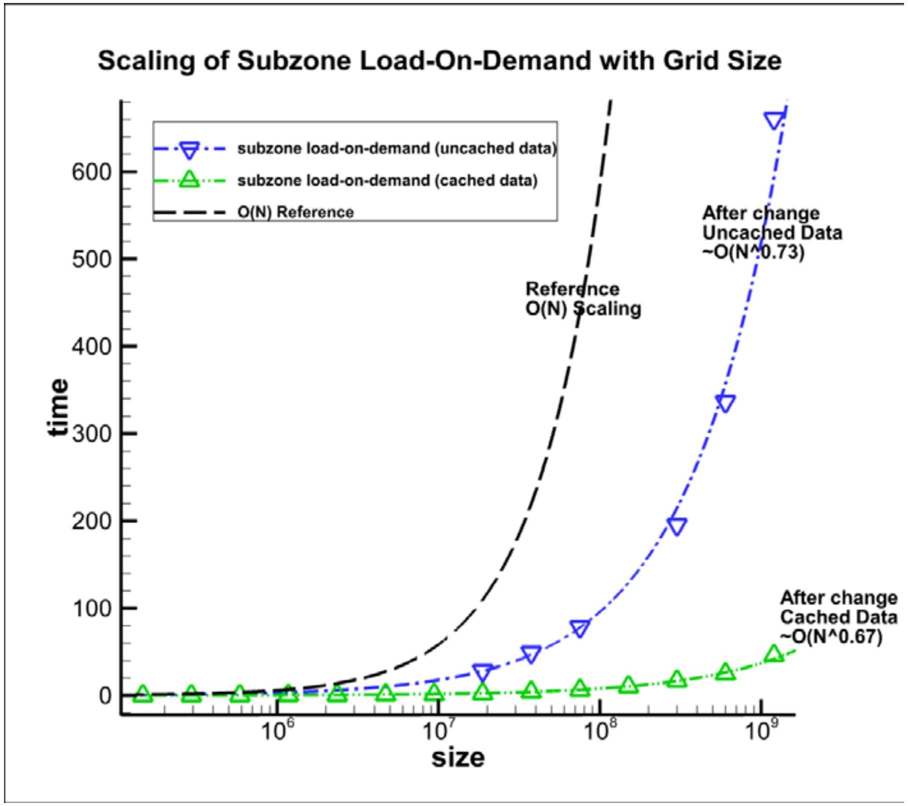
- SZLOD >**15x** faster to created iso-surface
  - The improvement is more pronounced with Larger datasets
- Uses >**15x** less RAM
  - Load a 75GB file with less than 8GB of RAM



tecplot Master the View

Scaling of Subzone LOD

- Overcoming Red Shift
  - Need sub-linear scaling with number of cells
  - Blue and green use binary search over subzones (O(log(n)))
  - Close to O(N^2/3)



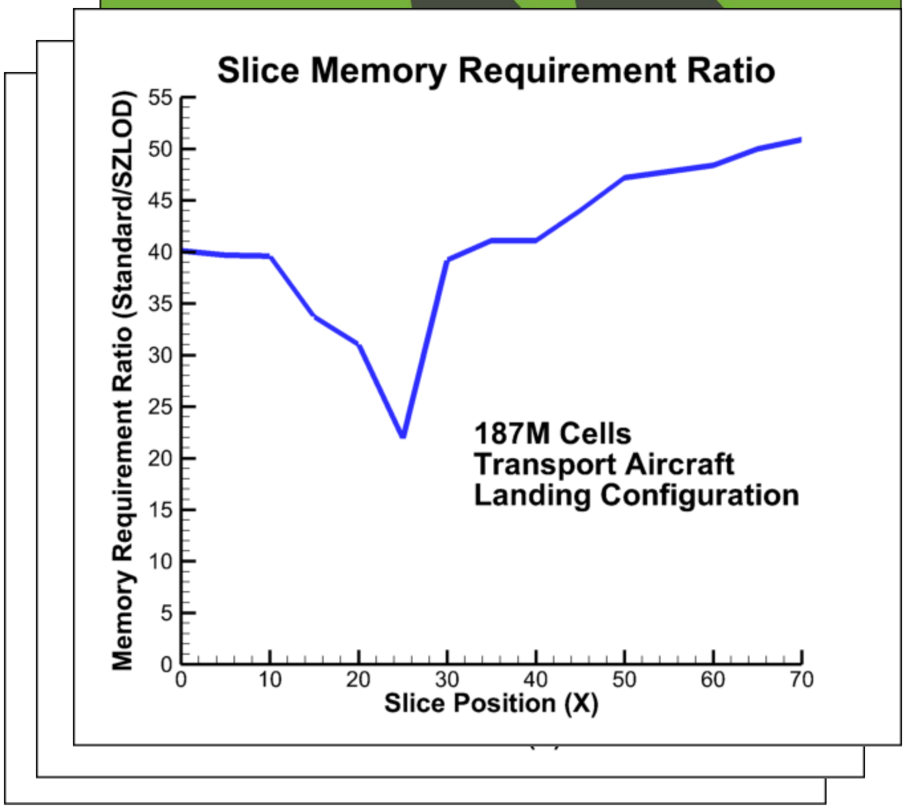
tecplot® Master the View

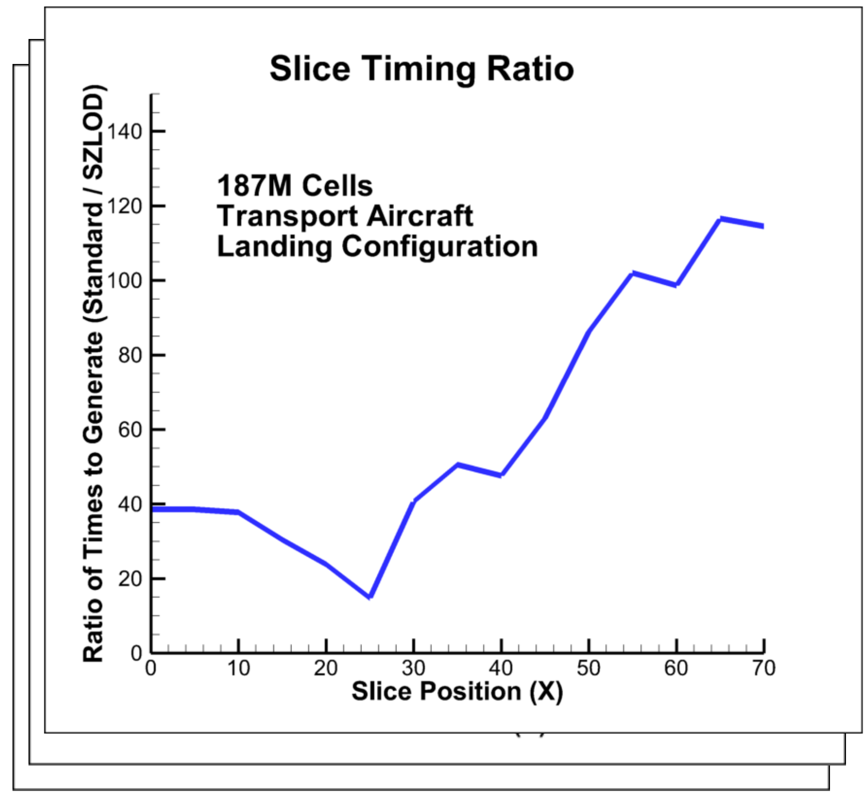
FE Transport Example

- Geometry
  - Generic Transport in Landing Configuration
  - Strong crosswind
- Grid
  - Finite-element (Tets, Prisms)
  - 187 Million Cells
- Solution
  - Velocity Data in the Volume
  - File - 6.9GB
- Compare
  - Generation of slice and Isosurface

tecplot. Master the View

FE Transport Aircraft - Slice





FE Transport Aircraft - Streamtrace

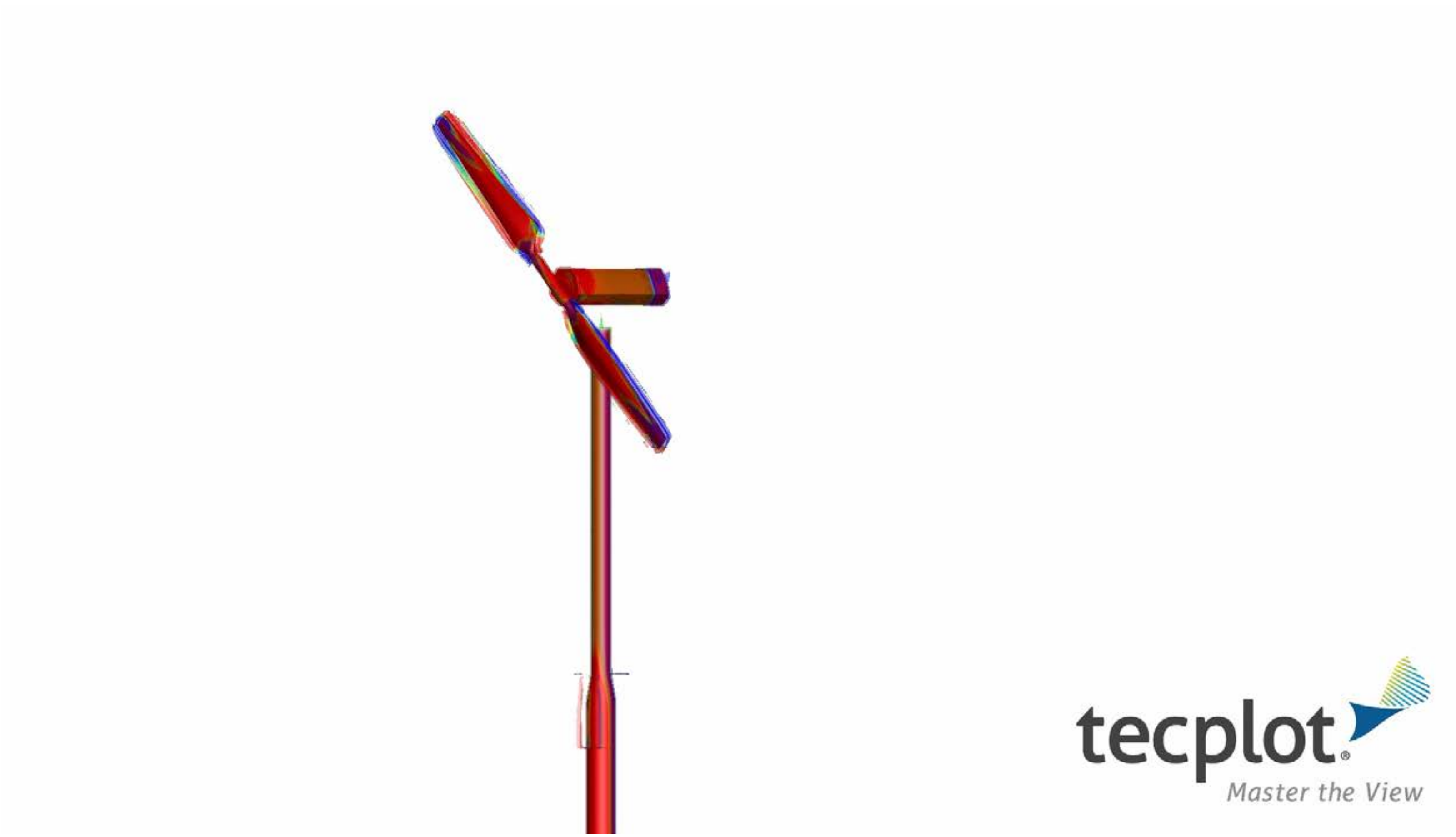
- Tecplot
  - 170 sec
  - 16 GB
- SZLOD
  - 2.2 sec
  - 1.3GB max
  - 0.7GB resting

tecplot® Master the View

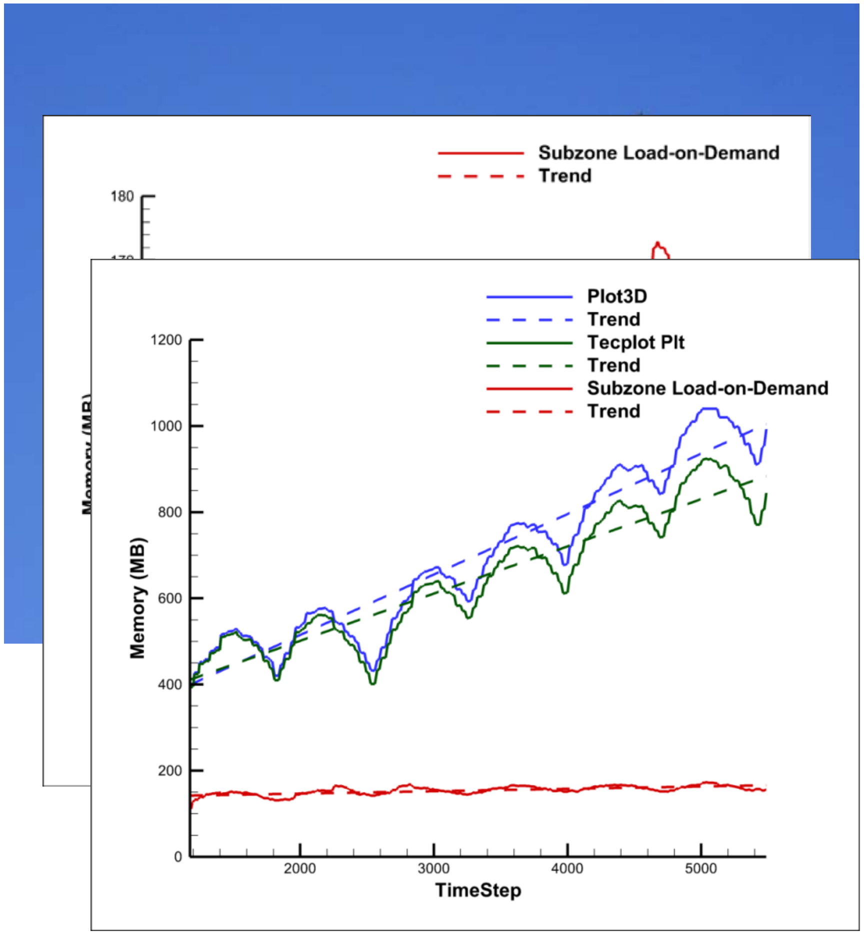
NREL Wind-Turbine Analysis

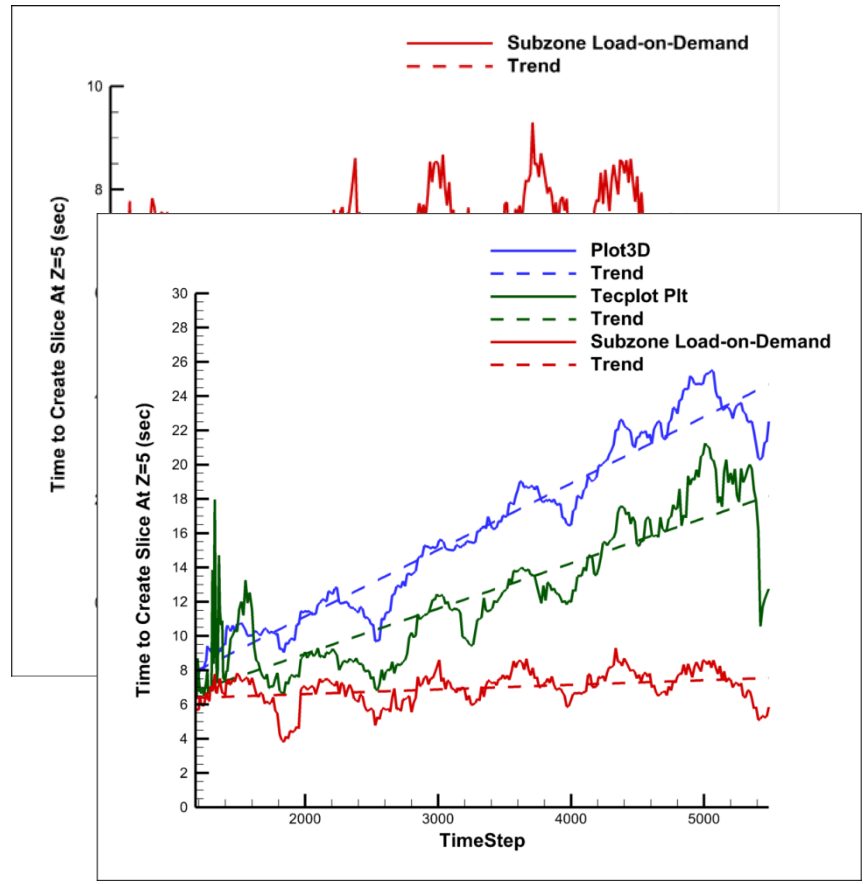
- 10m research turbine
- Acoustics (coupled to PSU WOPWOP)
- OVERFLOW 2.2
  - OVERFLOW-D mode
    - Domain Connectivity Function (DCF)
    - Geometry Manipulation Protocol (GMP)
  - 16 near-body blocks (2.6M points)
  - Adaptive Mesh Refinement
  - 2nd-order differencing near-body
  - 4th-order differencing off-body
- Results
  - 10 m/s aligned with turbine
  - 7152 time step (saved every 16 time steps)
  - Start: 30.5M nodes and 68 blocks
  - End: 260M nodes and 5600 blocks

Animation of Wind Turbine Vorticity Magnitude



SZLoD Performance for Overset Grid





Conclusions

- Dramatic reduction in memory requirements
  - Factor of 4 to 50 less memory used
  - Scaling for isosurface and slices is $O(N^{2/3})$ - critical for maintaining performance into the future
  - Scaling for a streamtrace is $O(N^{1/3})$
- Significant improvements in speed for most cases
  - 15 to 120 times faster for synthetic data and transport aircraft
  - 3 times faster for overset data with large number of zones
- Similar benefits when network bandwidth is bottleneck
- Downside
  - Speedups depend on using new file format (but you can still get memory reductions with native files)

Future Work

- Optimize for large number of zones (Overset)
  - Eliminate O(M) operations (M=number of zones)
  - Improve handling of blanking
  - Should achieve speedups comparable to single-zone structured grid (15x)
- Further speed improvements (factor of 3-10 faster)
  - Threading (extraction is now the bottleneck)
  - Eliminating remaining O(N) operations
- Integrate into Tecplot products
  - Currently implemented in add-on (work-in-progress)

Questions?

If you are interested in testing this technology, please talk with Scott (s.imlay@tecplot.com) or Durrell (d.rittenberg.@tecplot.com).