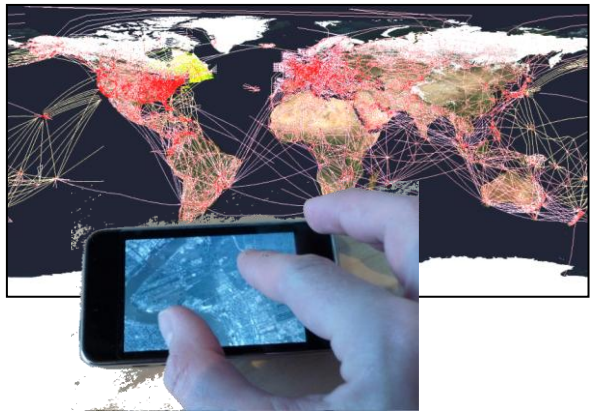
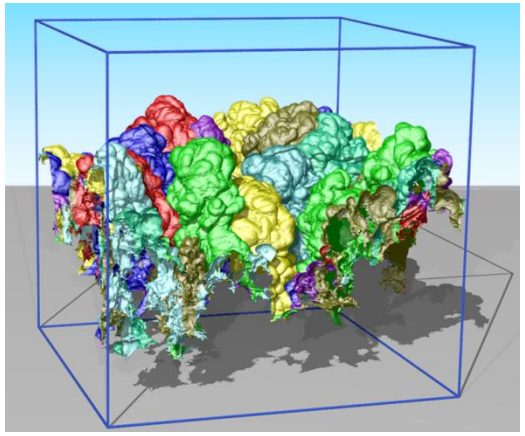
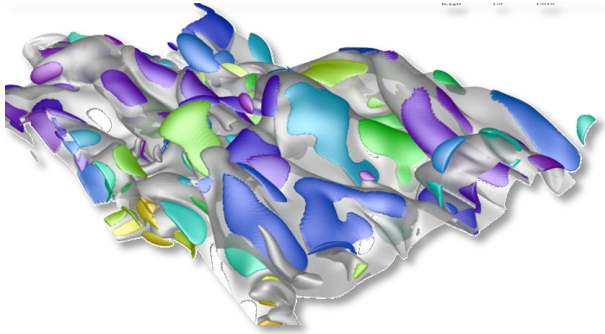
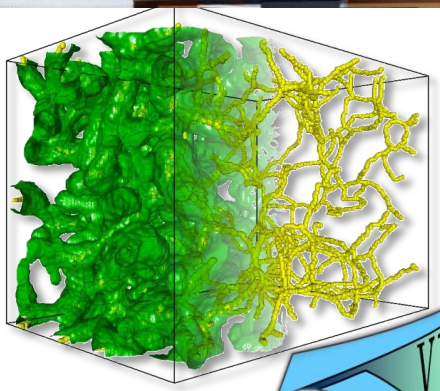
ViSUS: Massive Data Management, Analysis, and Visualization with Scaling From Handheld Devices to Supercomputers







ViSUS Framework



Valerio Pascucci

Director, CEDMAV

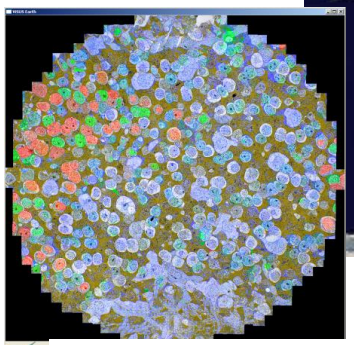
Professor, SCI Institute & School of Computing

Laboratory Fellow, PNNL

Massive Simulation and Sensing Devices Generate Great Challenges and Opportunities


BlueGene/L


Satellite

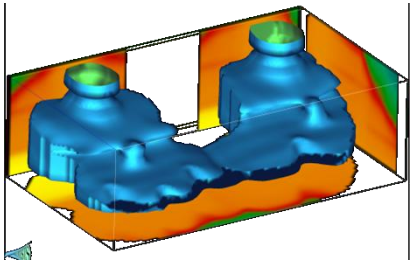

Retinal Connectome

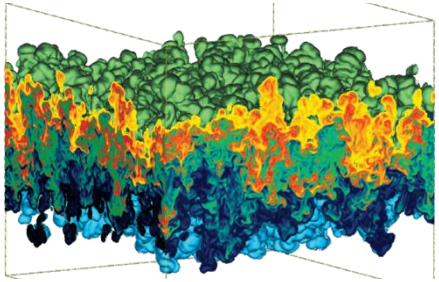

Earth Images

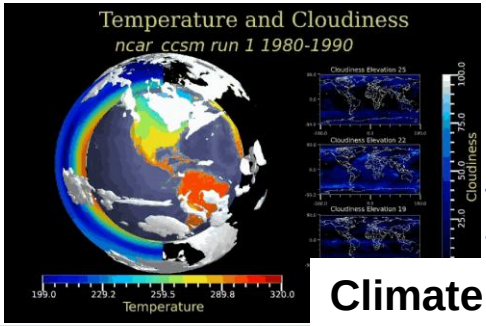

Cameras


EM

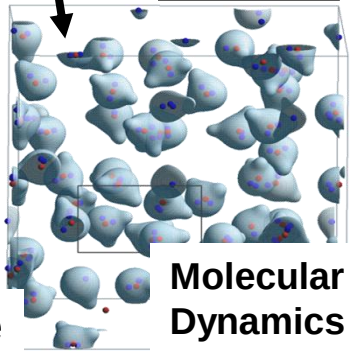
Jaguar

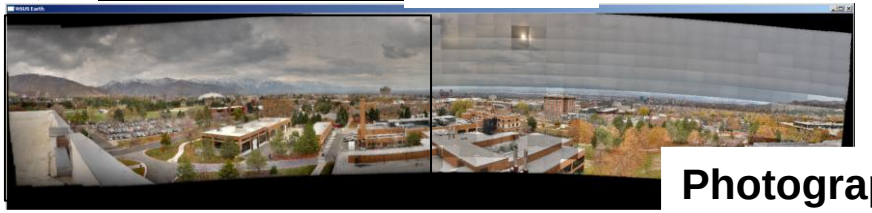

Carbon Seq. (Subsurface)

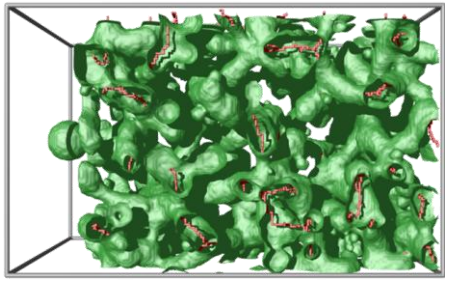

Hydrodynamic Inst.

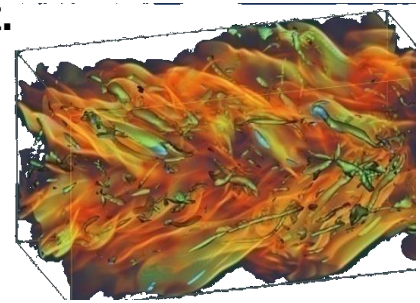


Climate


Molecular Dynamics

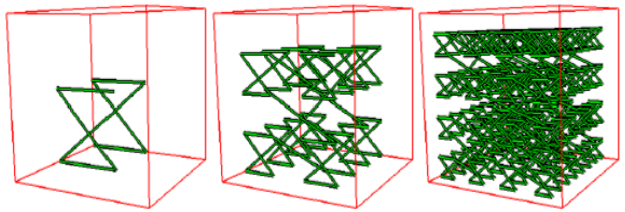

Photography


Porous Materials

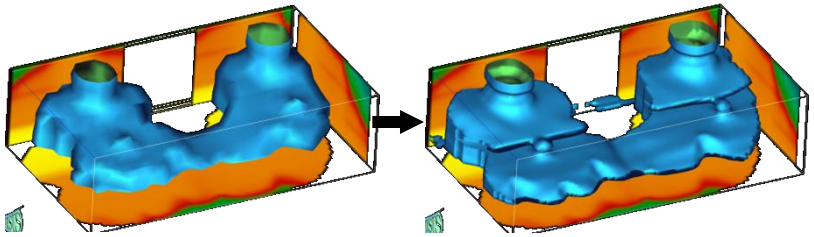

Turbulent Combustion

A Cyberinfrastructure Requires Efficient Data Management and Processing

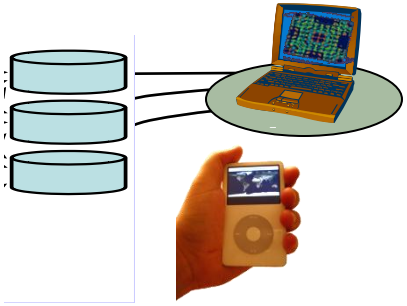
- **Advanced data storage techniques:**
  - **Data re-organization.**
  - **Compression.**



- **Advanced algorithmic techniques:**
  - **Streaming.**
  - **Progressive multi-resolution.**
  - **Out of core computations.**

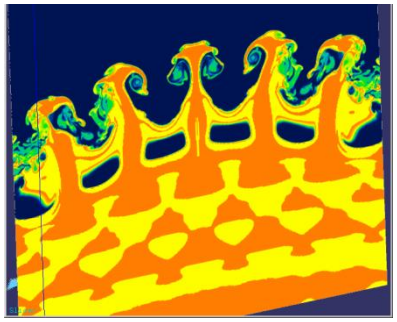


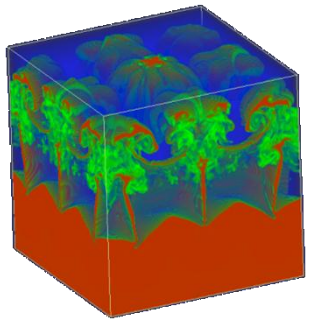
- **Scalability across a wide range of running conditions:**
  - **From laptop, to office desktop, to cluster of PC, to BG/L.**
  - **Memory, to disk, to remote data access.**

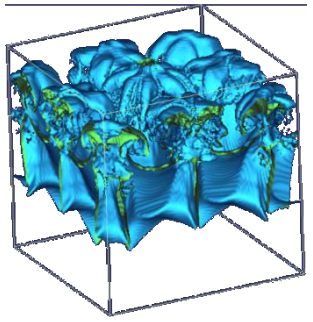


We Redesigned the Data Management and Visualization Pipeline with New Principles

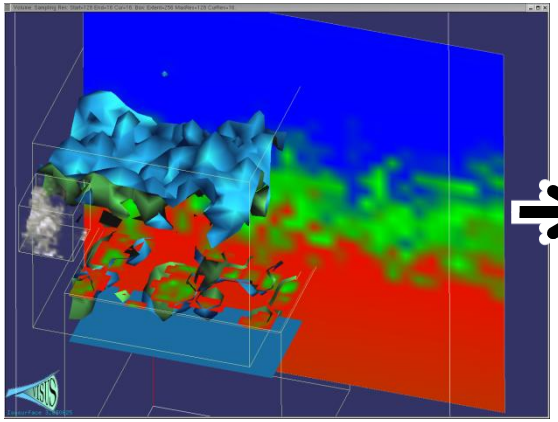
- **Basic core techniques:**
  - **Slicing**
  - **Volume rendering**
  - **Iso-surfaces**

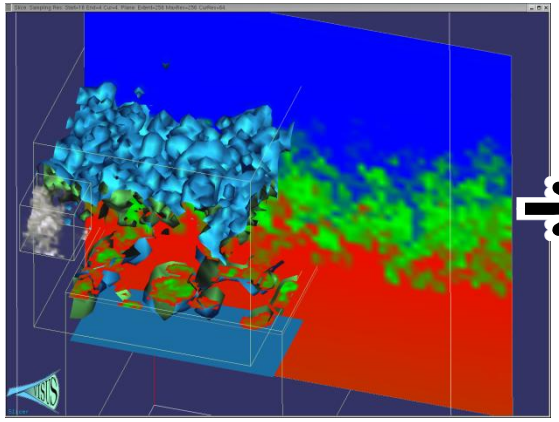


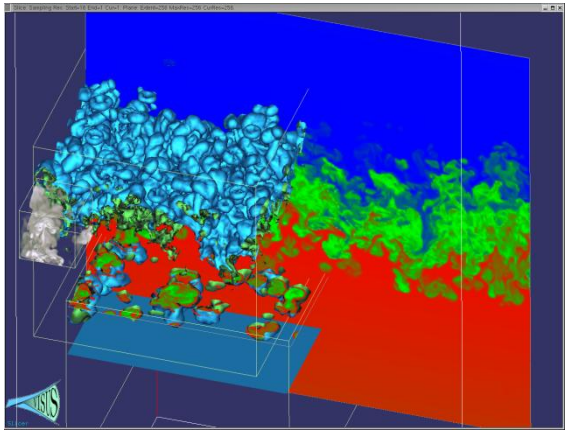




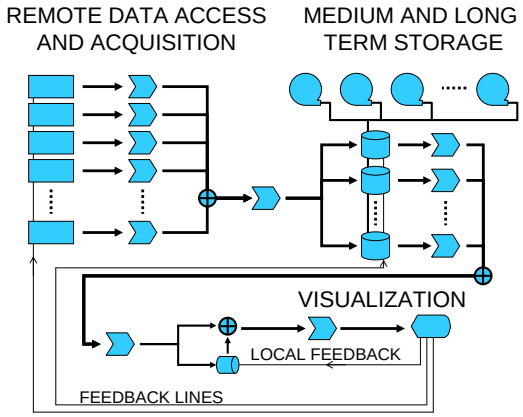
- **Cache-oblivious out-of-core processing optimizing access locality for any size of data blocks**
- **Pipelines of progressive algorithms**
- **Coarse-to-fine construction of multi-resolution models**
- **Remote data streaming**

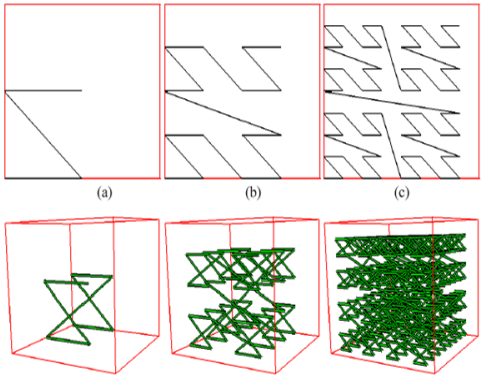






We Consider the Three Main Components Defining a Computing Infrastructure

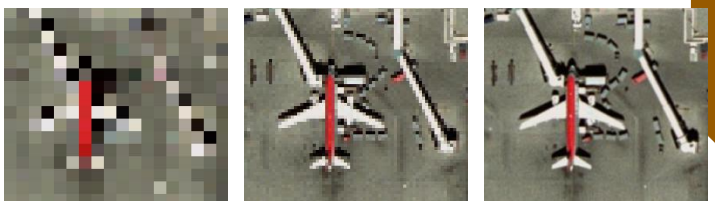




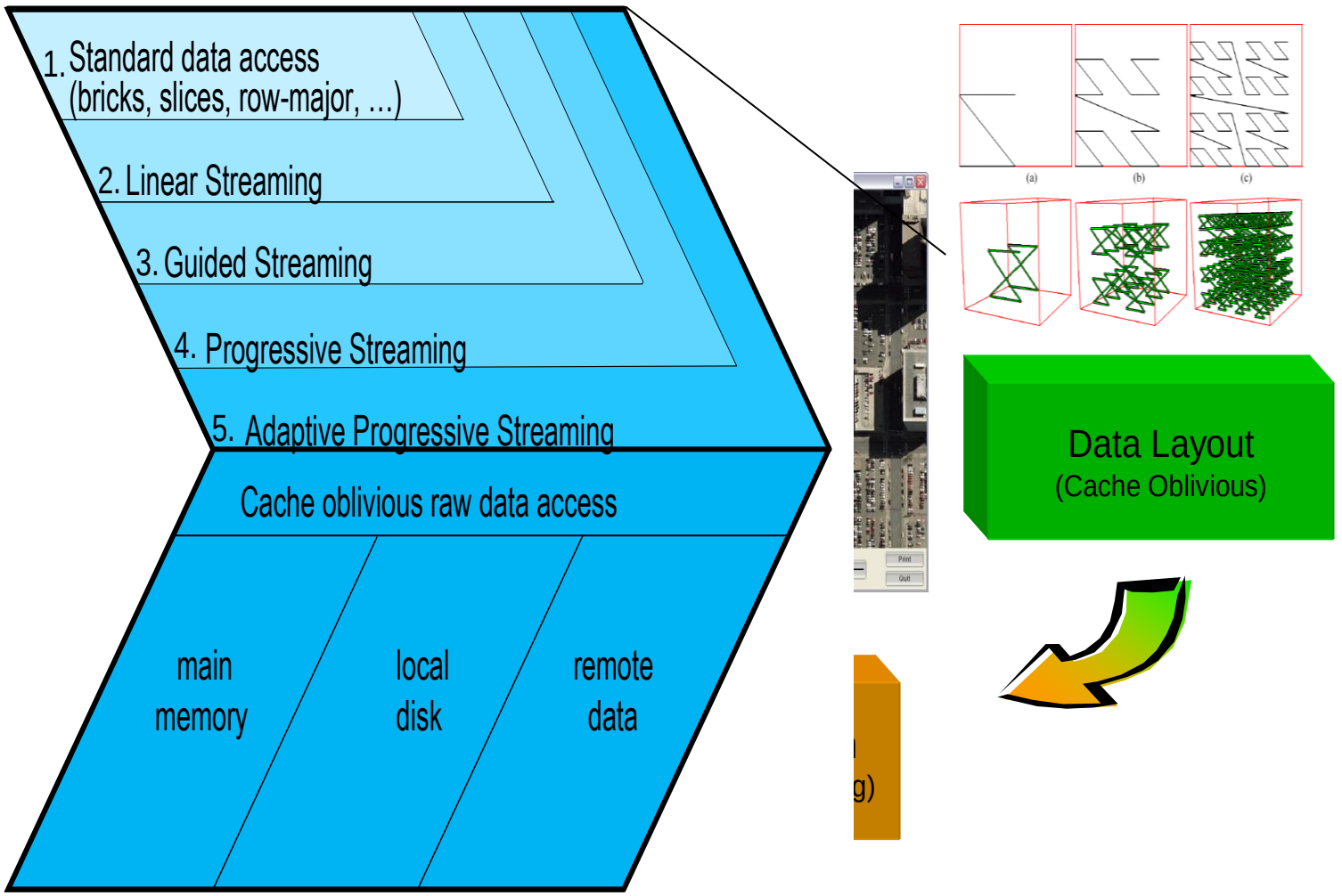
Processing Network
(Data Access Path)

Data Layout
(Cache Oblivious)

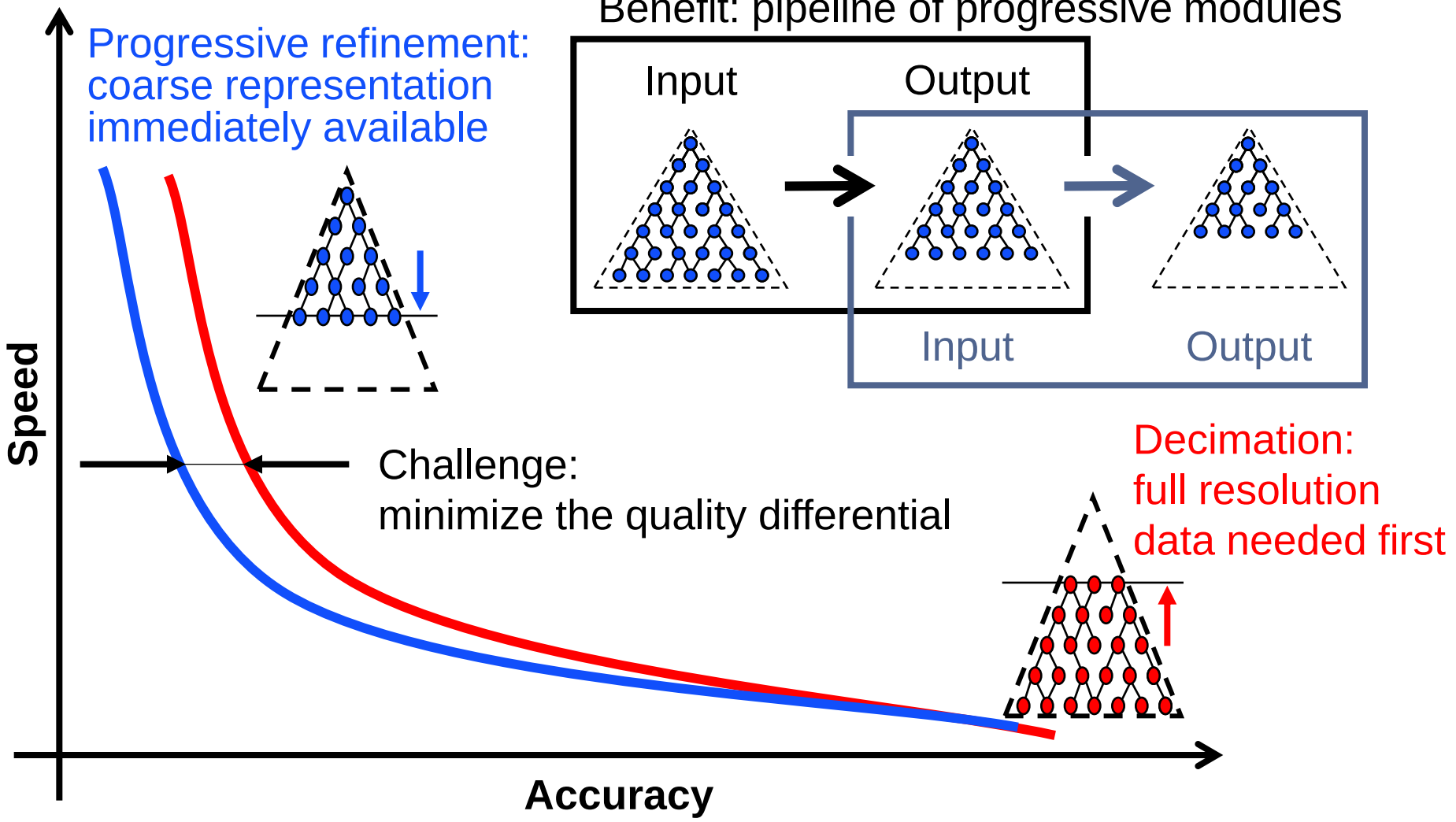
Algorithm Design
(Progressive Processing)



We Characterize Algorithmic Classes Based on Effect in a Processing Network

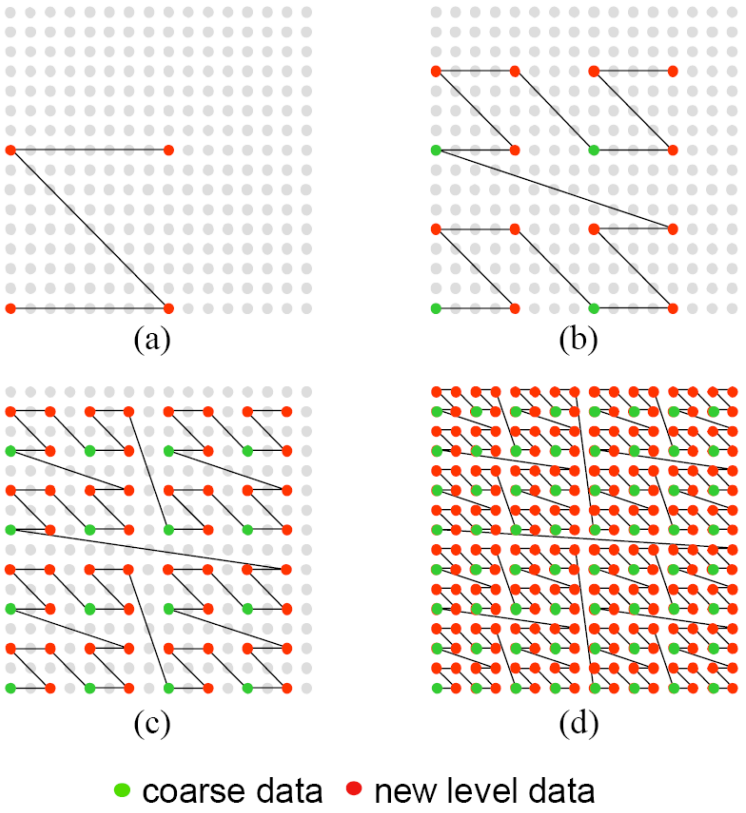


The use of top-down and bottom-up processes have a strong impact on the data stream

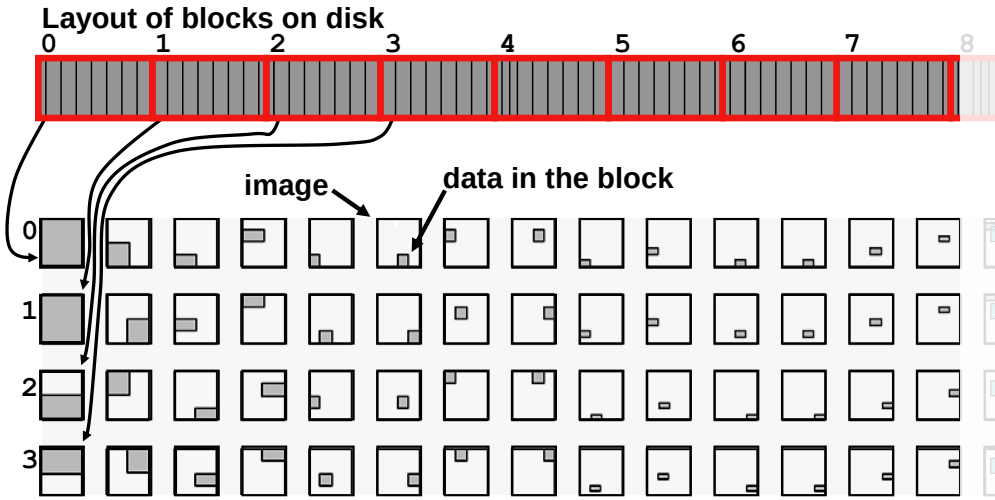


We Introduced Multi-resolution Cache Oblivious Layouts for Image Data

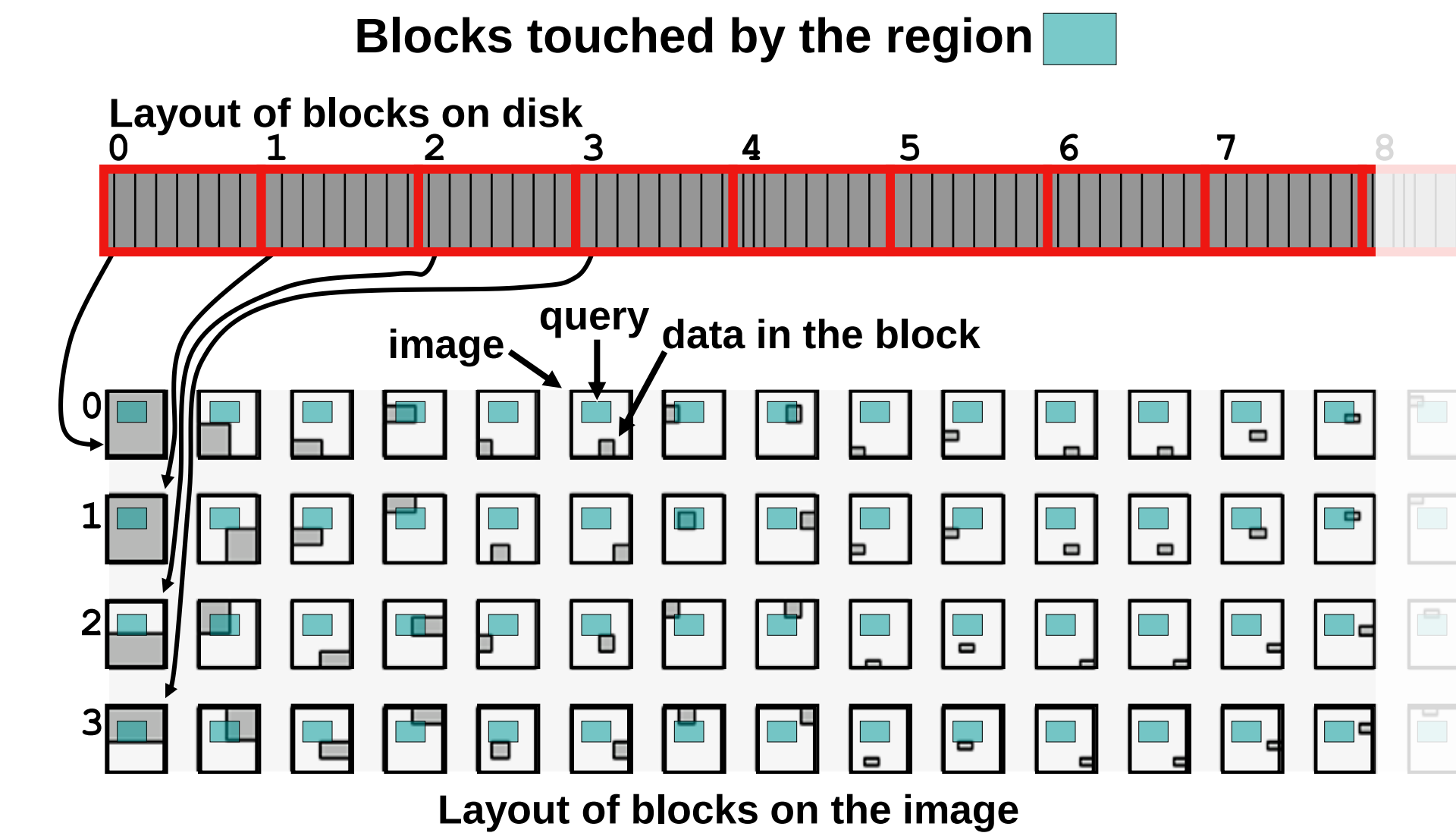
- Z-order curve used to define a hierarchical sub-sampling over a grid



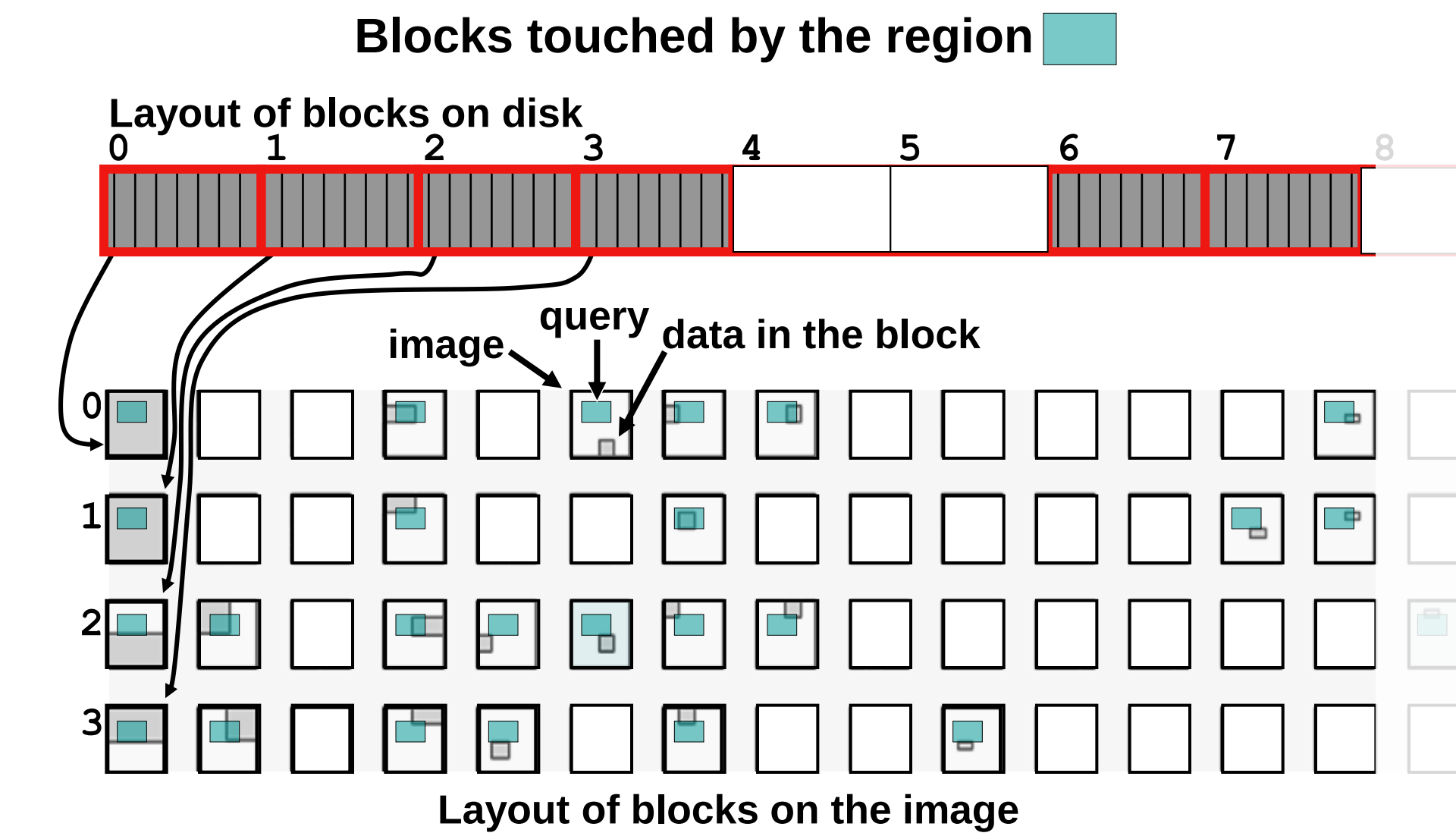
- Improve access locality:
  - **Interleaving hierarchical levels**
  - **Maintaining geometric proximity**
- Data layout is independent of the traversal of the data



We Introduced a Progressive Range Query Avoiding Unnecessary Data Access

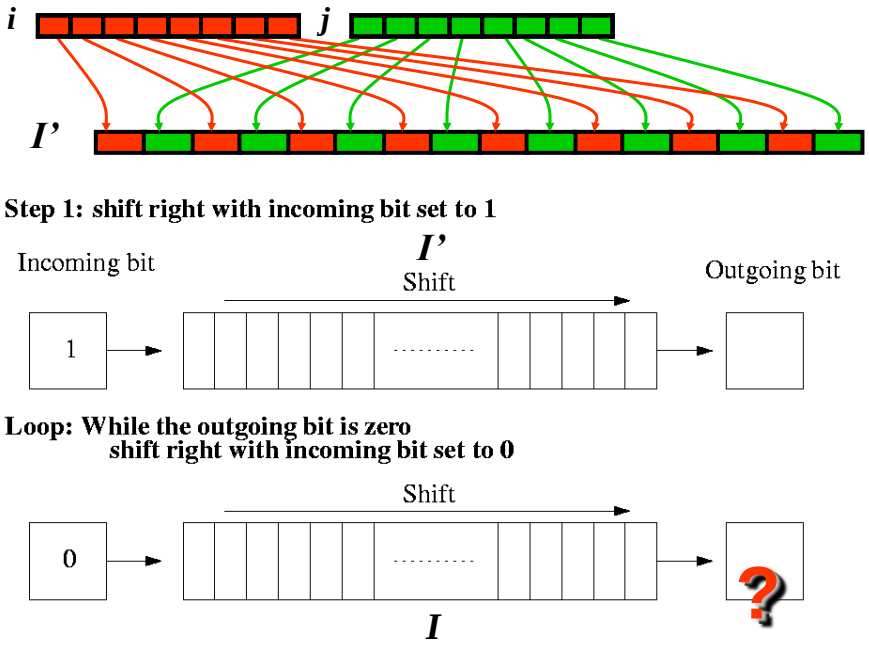


We Introduced a Progressive Range Query Avoiding Unnecessary Data Access

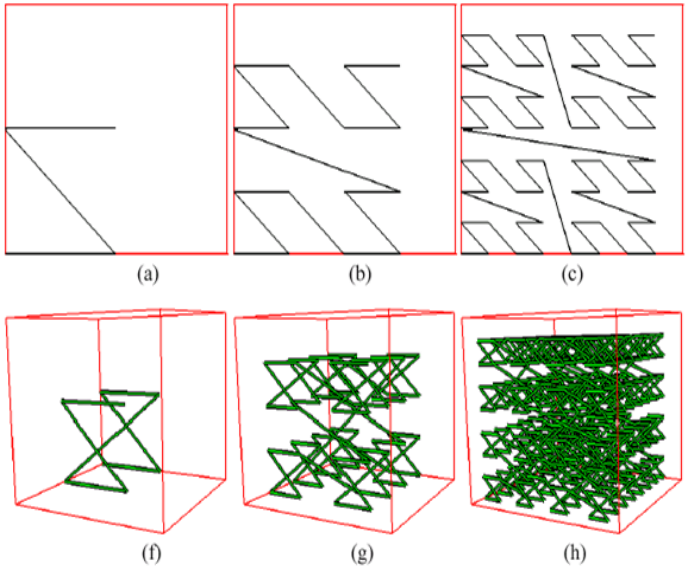


We Provided a Fast Address Computation Based on Simple Bit Manipulation

- **Simple bit manipulations to convert row major to hierarchical Z-order**



- **3D version (also nD): basic Z shape replaced by a connected pair of Z shapes**

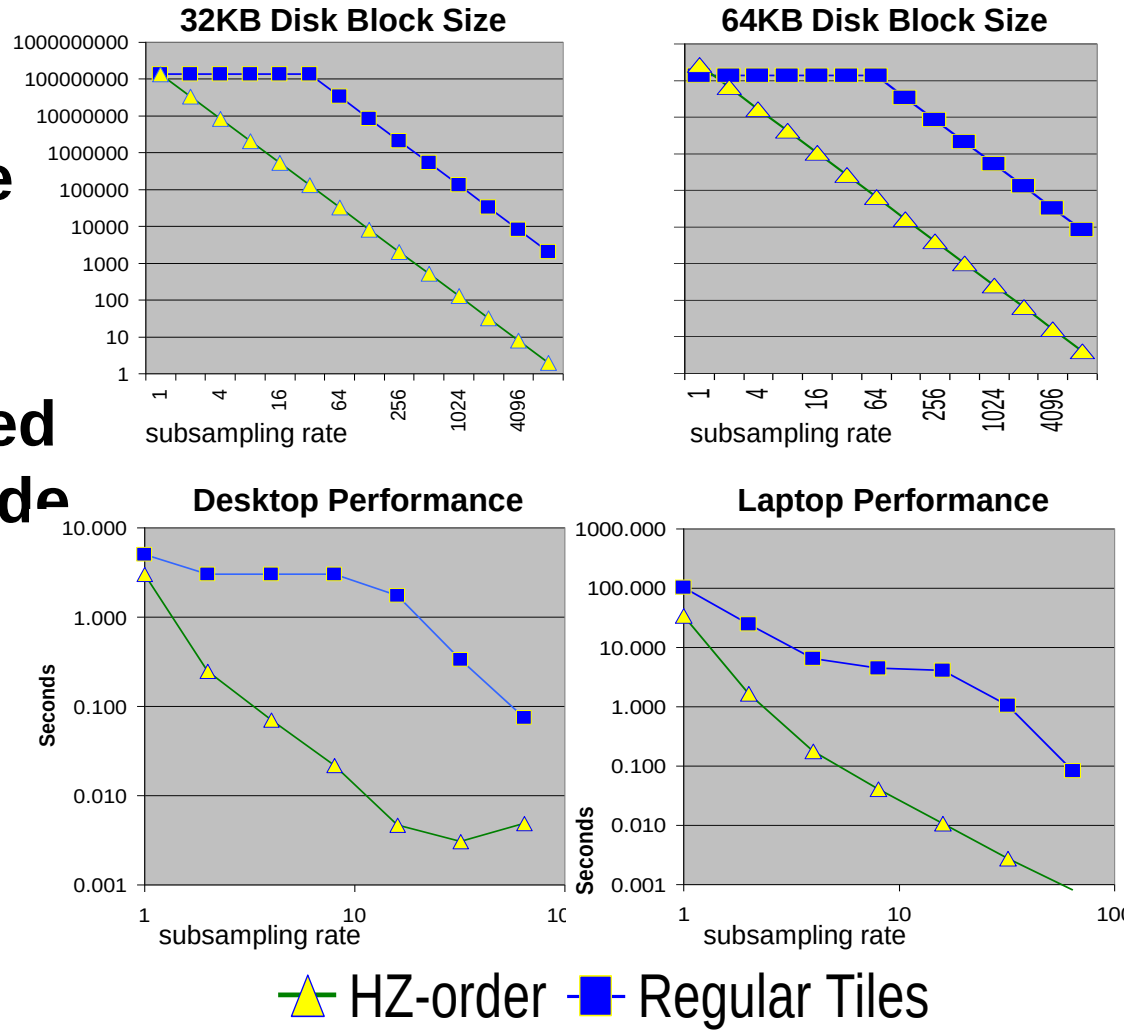


Cache-Oblivious Data Layouts Scale Well Across Different Storage Blocking Factors

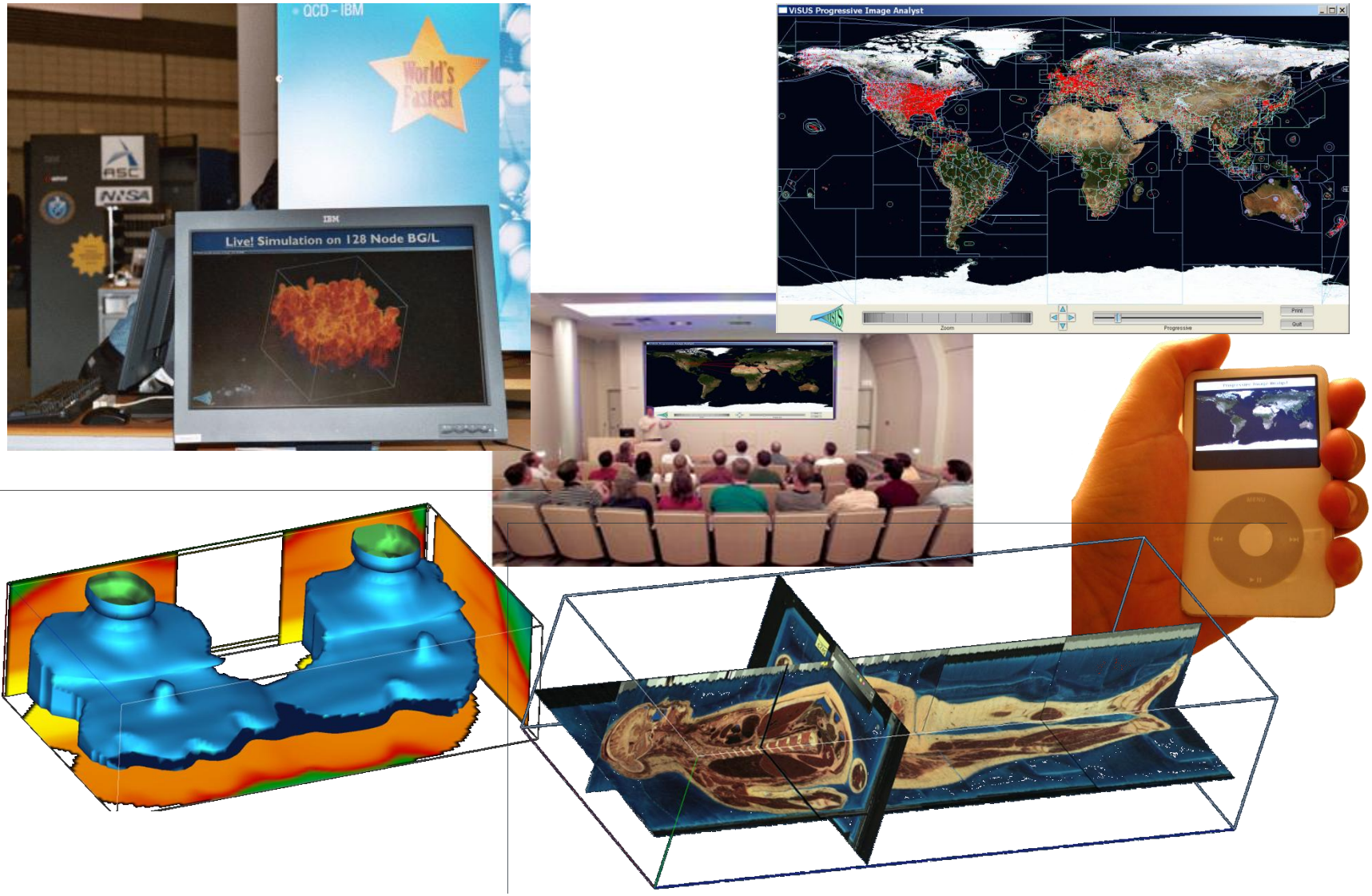
Formal analysis predicts performance and scalability

Performance improved by orders of magnitude

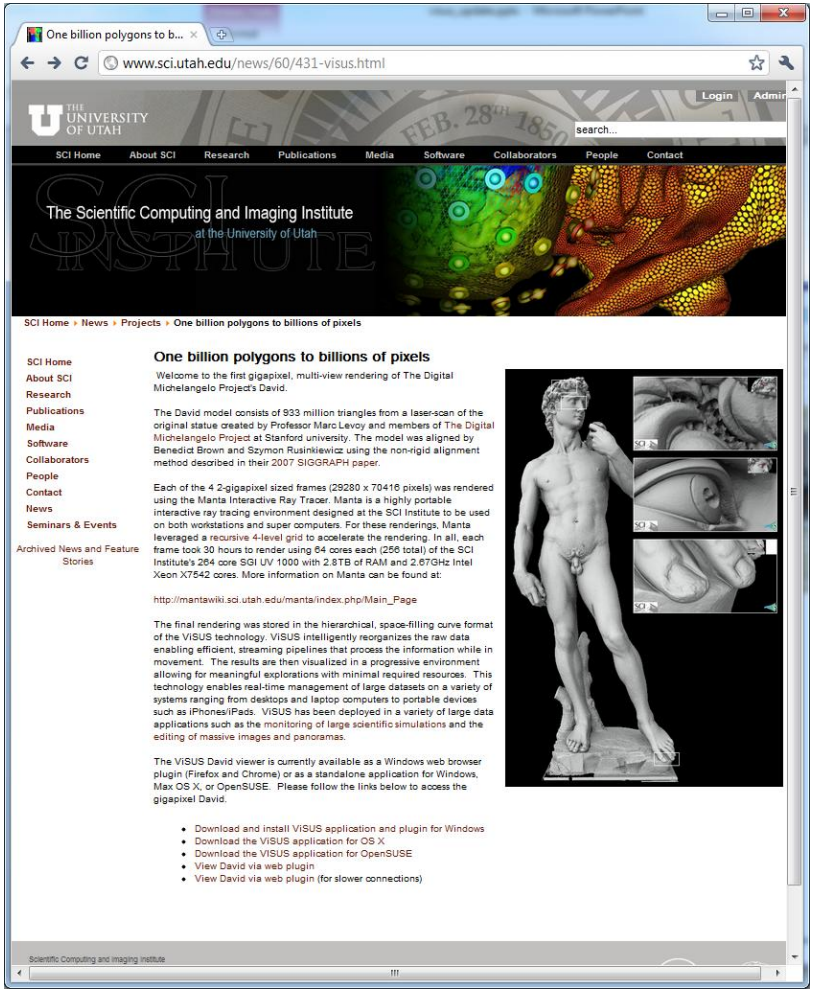
Independence of architecture and storage characteristics

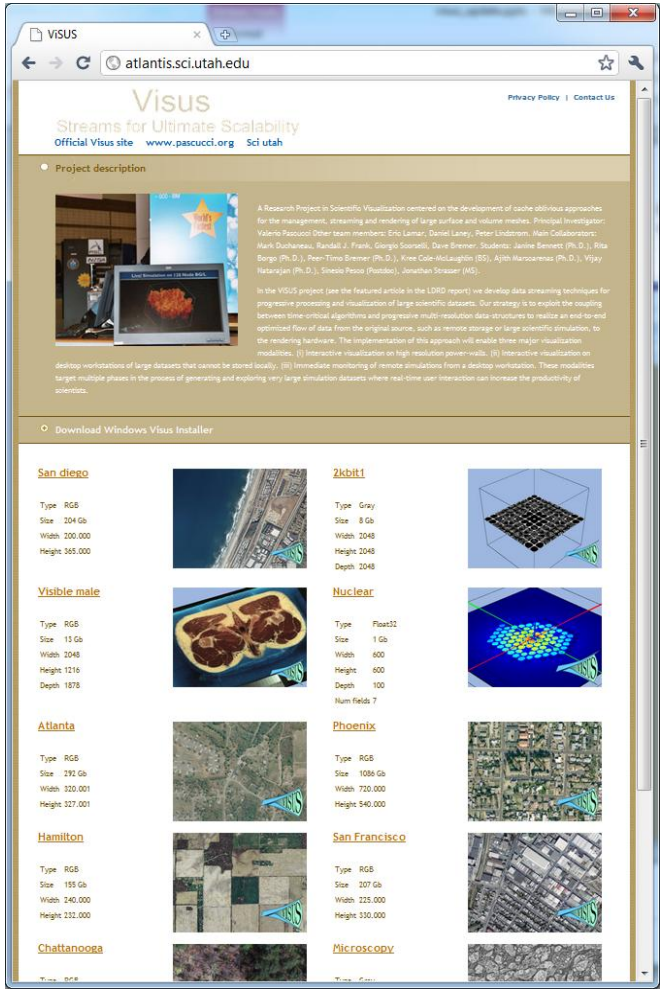


We Demonstrated Performance and Scalability in a Variety of Applications



Server can be wrapped in Apache plug-in
Client can be run in a web browser





Geospatial Data Rendering on iPad

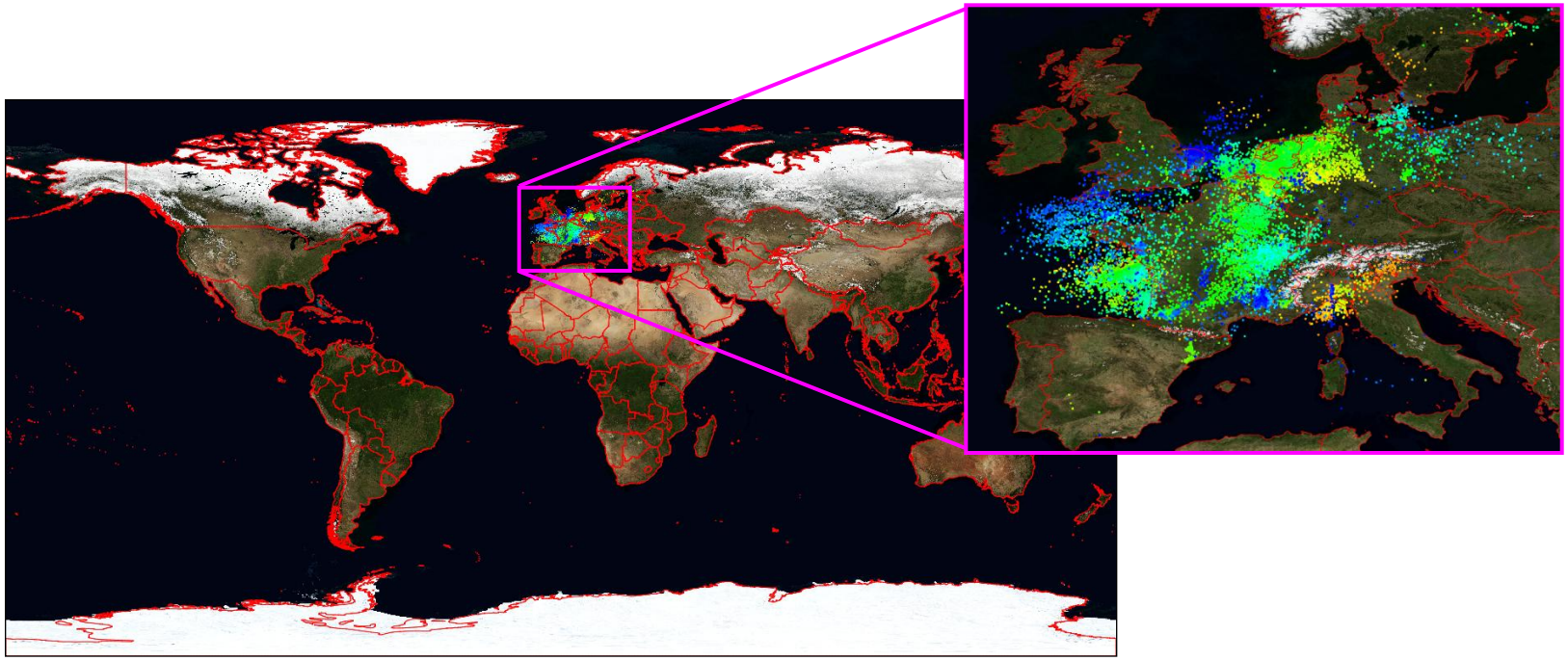
Both client and SERVER run of handheld devices, e.g. multiple iPhones can be clients and servers for each other to share information on the field

We Address the Need for Scalable Algorithms and Infrastructures

- Data formats that minimize I/O and memory transfer for most frequent operations
- New algorithms and data structures for management of large collections of time dependent information
- New theoretical models that predict the behavior of modern architectures
- New algorithms that are “intrinsically scalable” with respect to:
  - Processing capabilities
  - Diversity of hardware available
  - Locality of data
- Can benefit a variety of tools
- Scalable system infrastructures

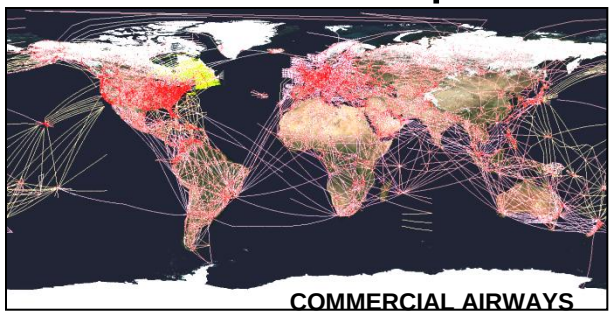
We provide real-time access to large scale time dependent data and sensory data

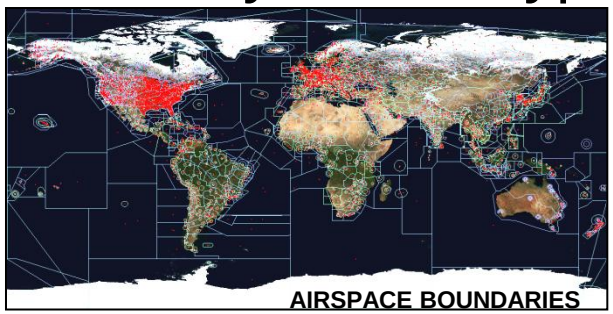
- **Blue Marble Earth (next generation) provided by NASA:**
  - **Twelve months in 2004 (11GB per month)**
  - **Resolution: 86400 x 43200 pixels (500 meters per pixel)**
- **Lightning events from distributed sensing devices over Central Europe (Blitzortung.org)**

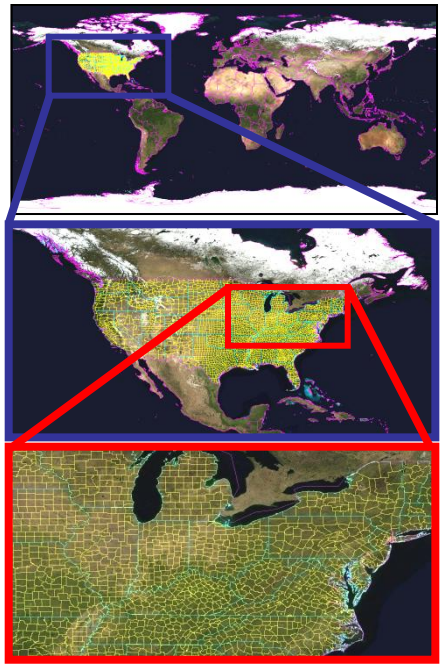


We Are Moving Towards Support of a Combination of Different Data Types

- Concurrent access to several information sources requires similar techniques for a variety of data types.


COMMERCIAL AIRWAYS

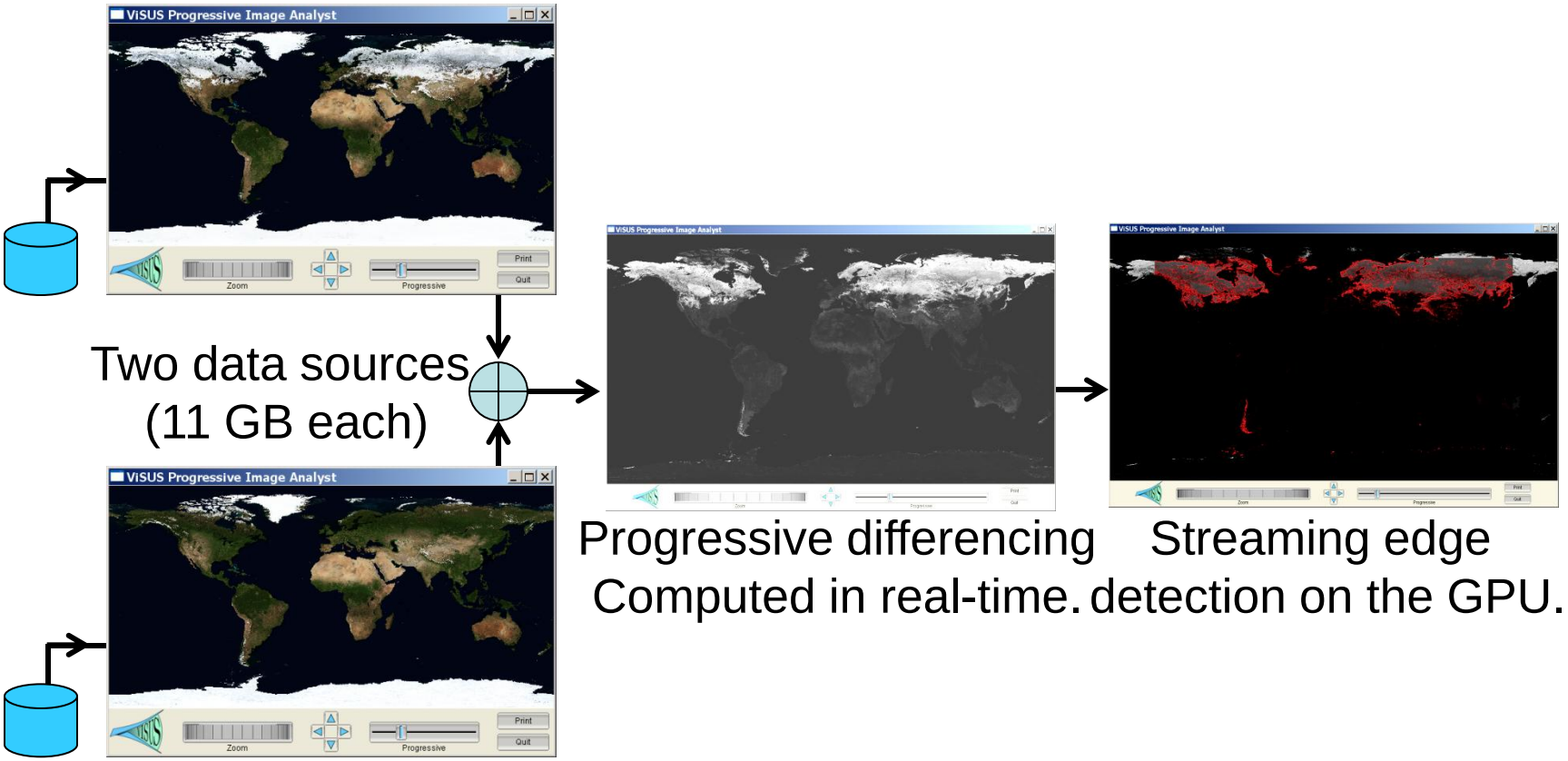

AIRSPACE BOUNDARIES


BOUNDARIES OF US COUNTIES

- **Algorithmic directions:**
  - **integration with other frameworks**
  - **progressive streaming infrastructure for "vector data" (points, lines, polygons);**
  - **uncertainty/incompleteness in the data**
  - **progressive resolution of queries with quantification and visualization of error/confidence;**
  - **on-line update of internal data structures to render new data immediately available.**

We Allow Distributed Computations at Different Stages of the Data Stream

- Progressive Image Differencing + Editable GPU filter.

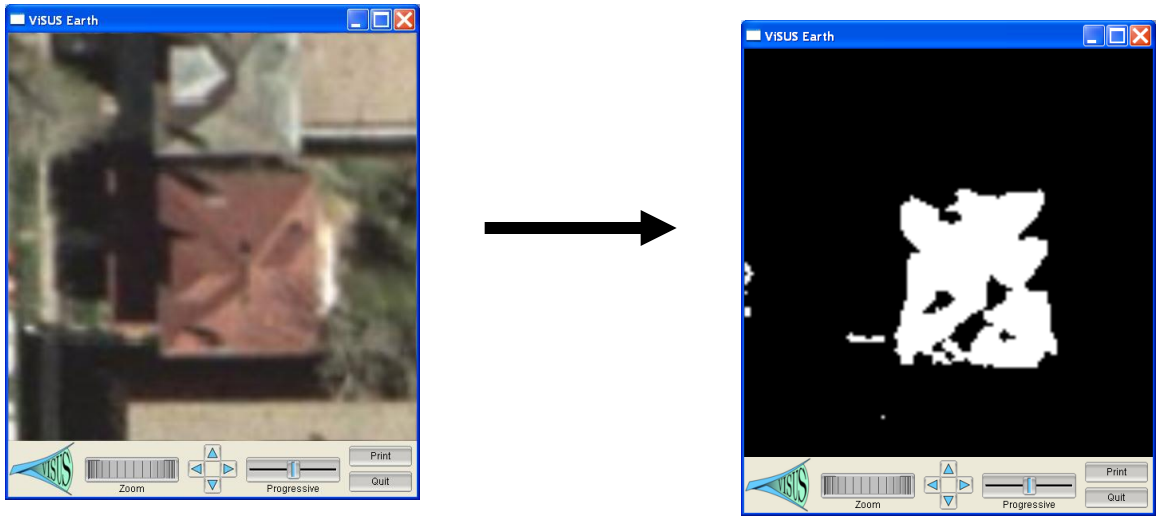


Progressive differencing Computed in real-time.

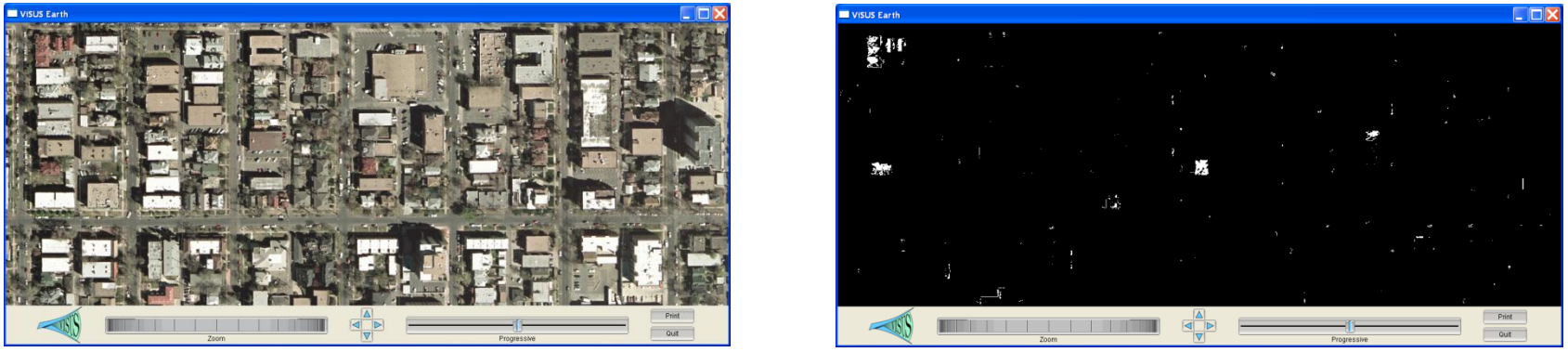
Streaming edge detection on the GPU.

We are Developing Progressive Scheme for Content Based Image Processing

- **Hypothesis:**



- **Progressive Analysis:**

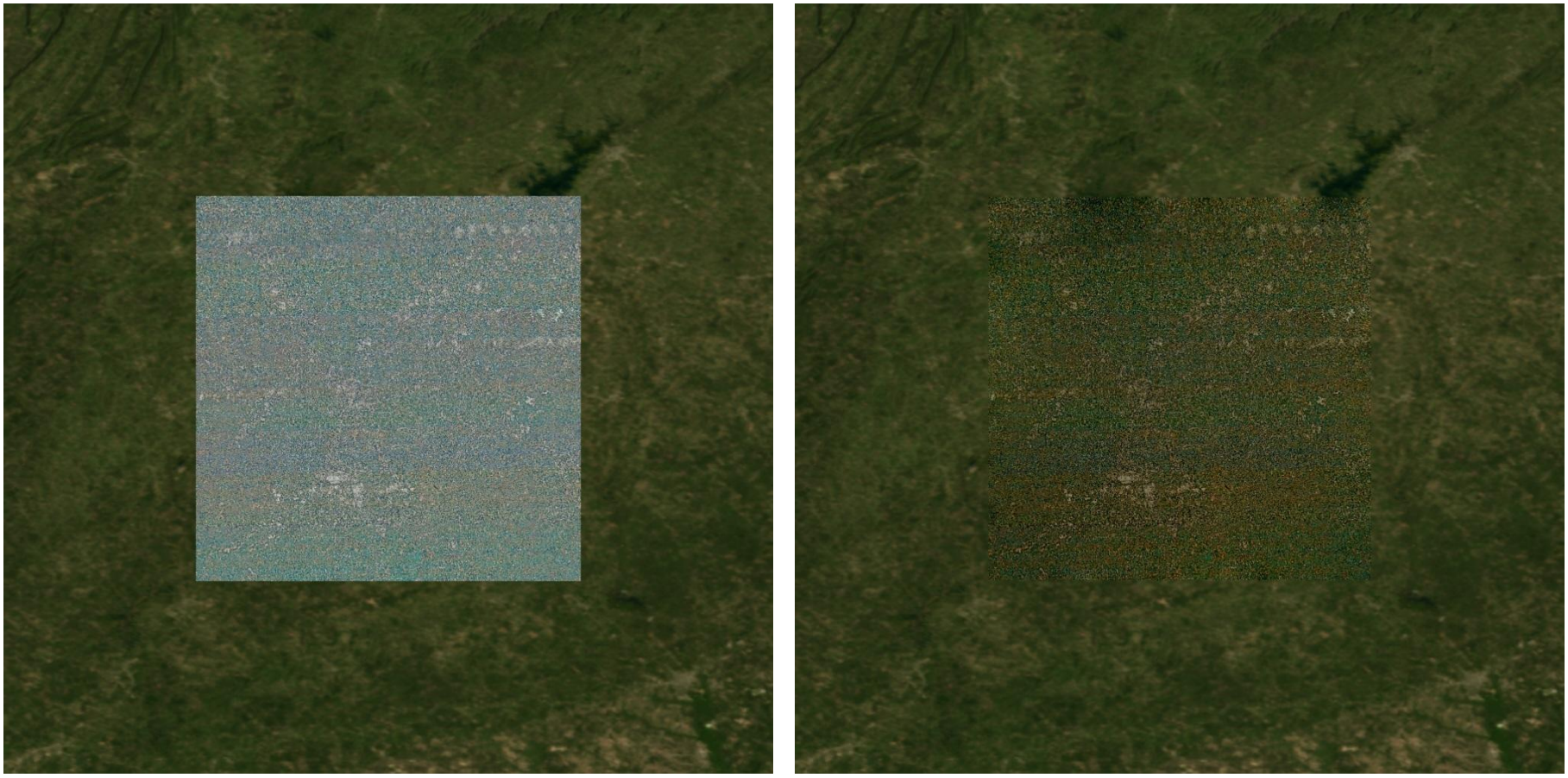


Poisson Solver for Image Cloning in Massive Image Collections

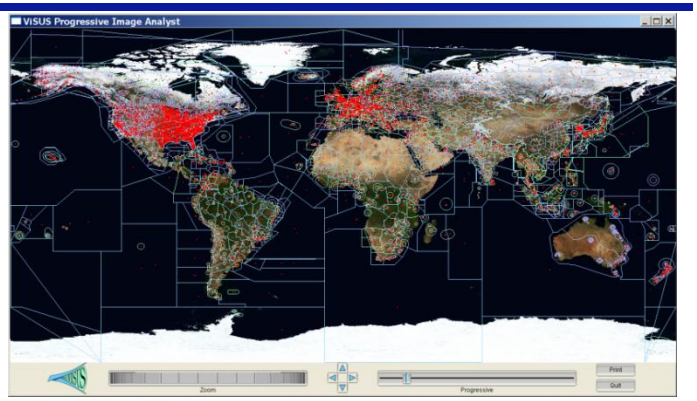
- **Color correction of 600+ images in real time**

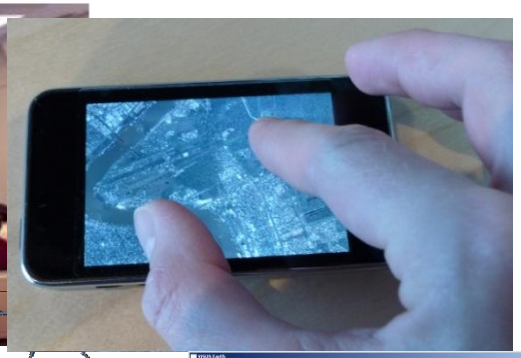
Poisson Solver for Image Cloning in Massive Image Collections

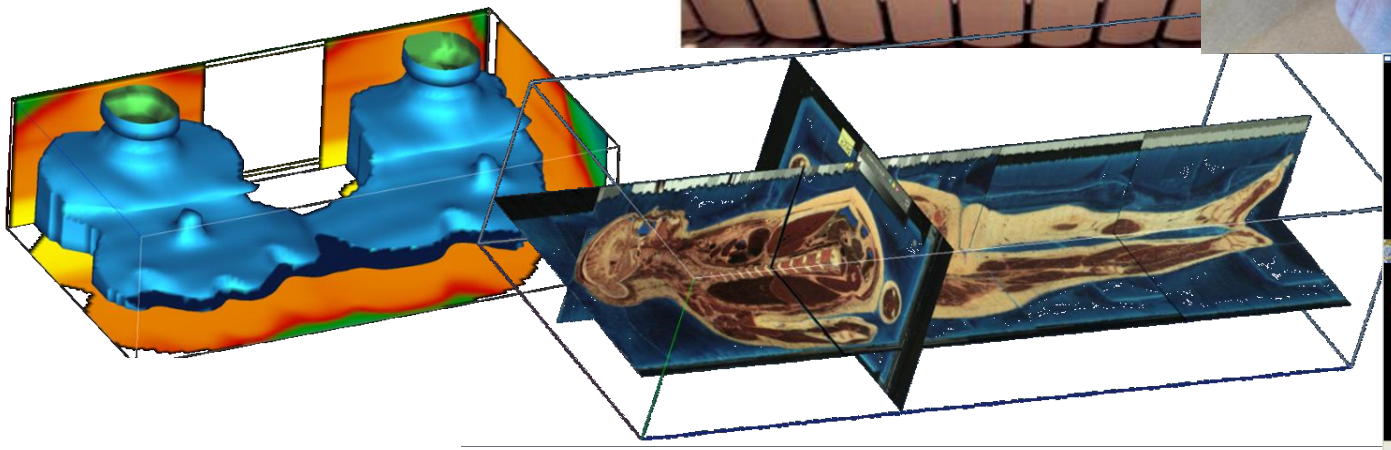
- **Pasting a 300GB satellite image of a city in background world map merged in real time**

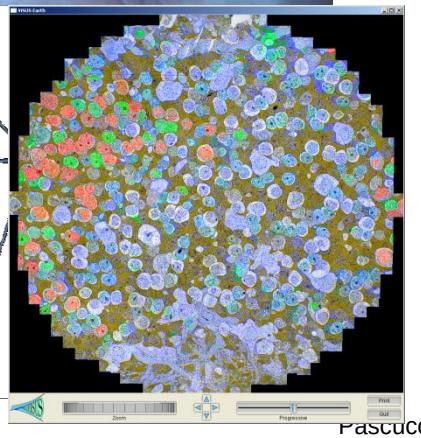


We Demonstrated Performance and Scalability in a Variety of Applications

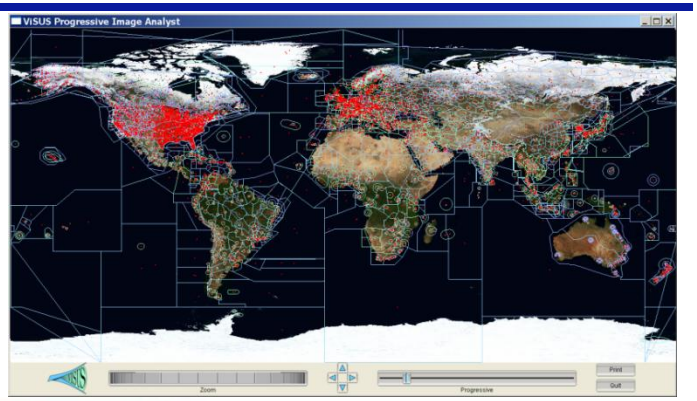


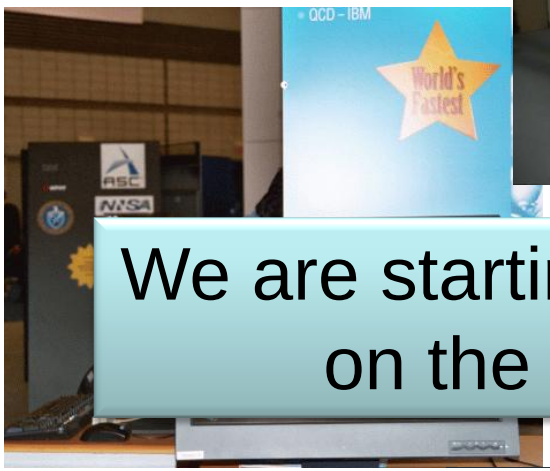




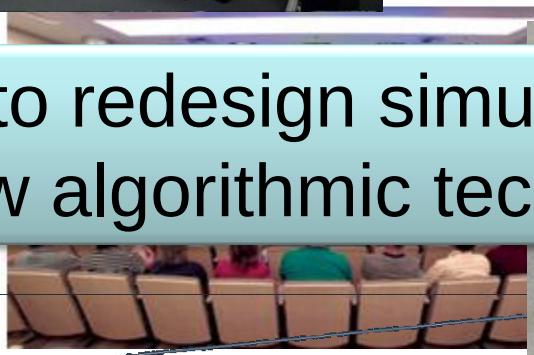


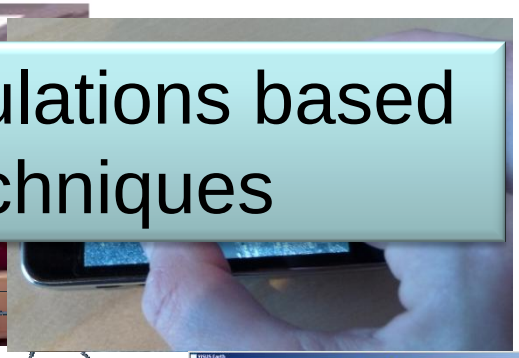
We Demonstrated Performance and Scalability in a Variety of Applications

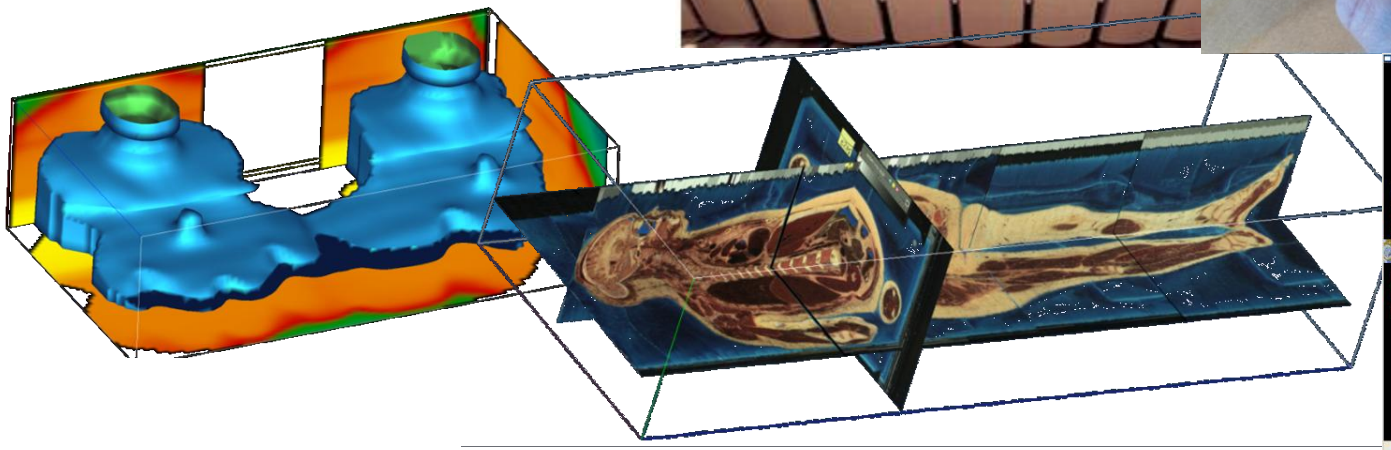


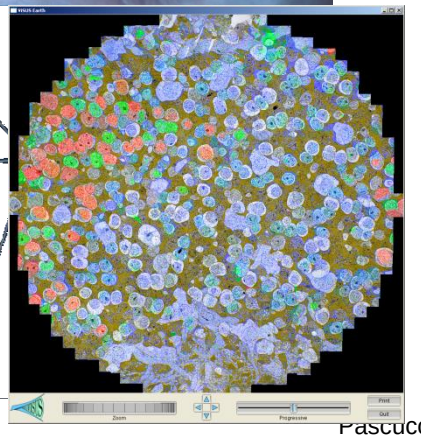


We are starting to redesign simulations based on the new algorithmic techniques









We Are Moving Towards a Distributed Storage and Processing Environment

- **Distributed storage**
- **Data redundancy**
- **Security**
- **Heterogeneous collaborative infrastructure**
- **Multi-scale collaborative interfaces accessing shared data sources:**
  - data collection and validation
  - interactive analytics
  - decision making