GPU accelerated heterogeneous computing for Particle/FMM Approaches and for Acoustic Imaging

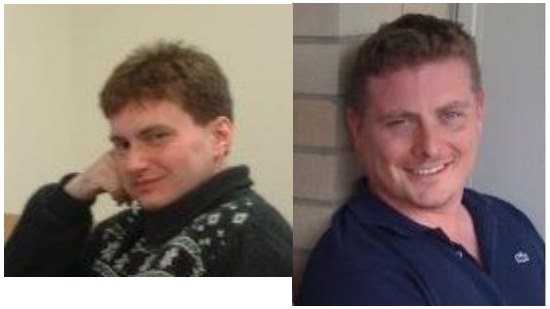
Ramani Duraiswami

University of Maryland, College Park

http://www.umiacs.umd.edu/~ramani

With Nail A. Gumerov, Qi Hu, Dmitry Zotkin, Adam O'Donovan

Work supported by NSF, AFOSR, ONR, DARPA and UMIACS



Fast Multipole Methods – Particle Methods

- Fast summation of radial functions
  - Green's functions $\Phi(x,y)$ for classical operators L
  - Solution y to forcing at points x
- Applications
  - Electrostatics, Molecular/Stellar dynamics, Vortex
  - Boundary Integral Methods
  - Particle discretization of field equations
  - Machine Learning

$$s(x_j) = \sum_{i=1}^{N} \alpha_i \phi(x_j - x_i), \quad \{s_j\} = [\Phi_{ji}]\{\alpha_i\}.$$

What is the FMM?

N-body problem

- Source points: $\{\mathbf{x}_i\}, i = 1, \ldots, N$
- Receiver points: $\{\mathbf{y}_j\}, j = 1, \ldots, M$
- The matrix-vector product:

$$\phi(\mathbf{y}_j) = \sum_{i=1}^{N} q_i \Phi(\mathbf{y}_j - \mathbf{x}_i), \quad j = 1, 2, \ldots, M$$

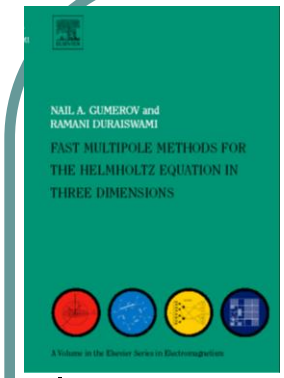
$$\mathbf{x}_i, \mathbf{y}_j \in \mathbb{R}^d.$$

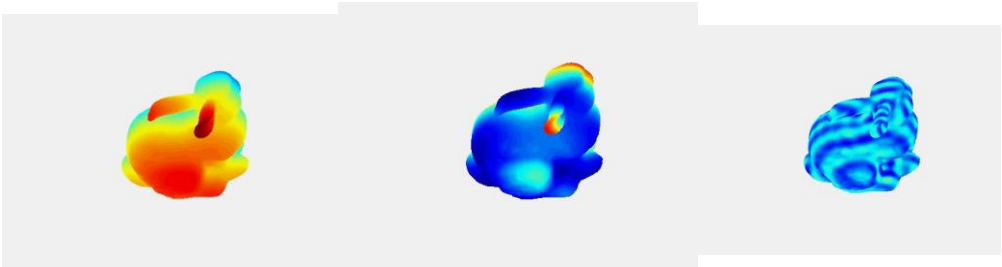
- Complexity of direct method: $O(NM)$

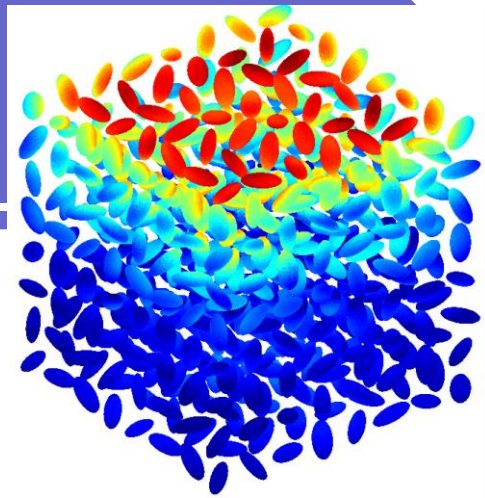
- Decompose singular sum in to
  - "local" or sparse part +
  - "far-field" and dense part

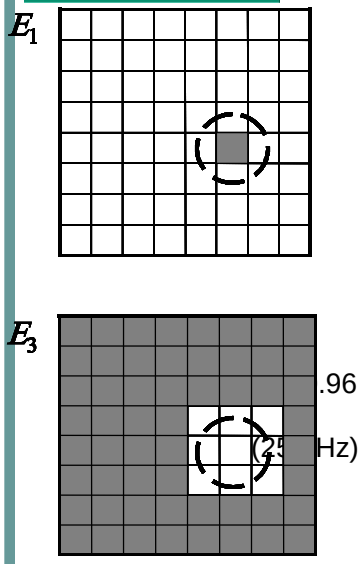
$$\phi(\mathbf{y}_j) = \sum_{\mathbf{x}_i \notin \Omega(\mathbf{y}_j)} q_i \Phi(\mathbf{y}_j - \mathbf{x}_i) + \sum_{\mathbf{x}_i \in \Omega(\mathbf{y}_j)} q_i \Phi(\mathbf{y}_j - \mathbf{x}_i)$$

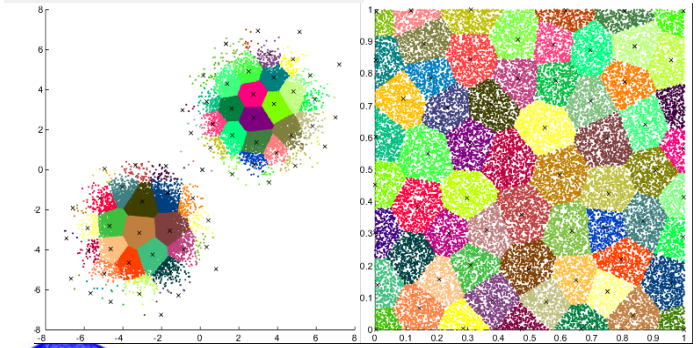
A few pretty pictures from papers

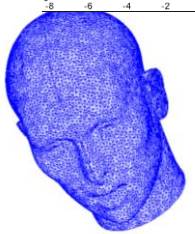


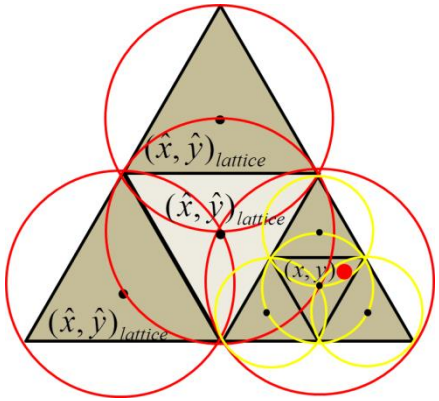


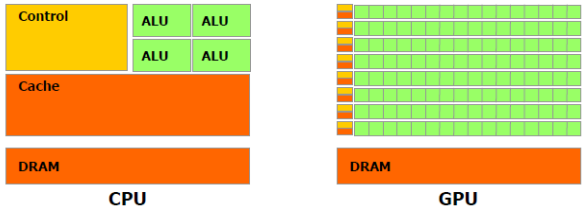


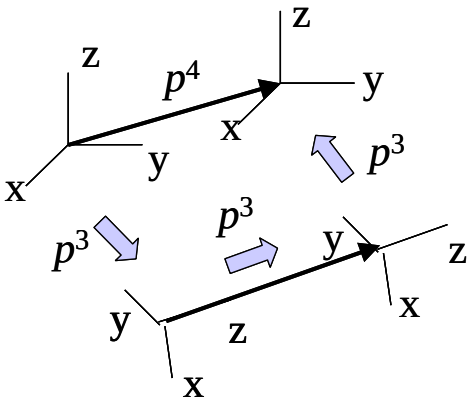


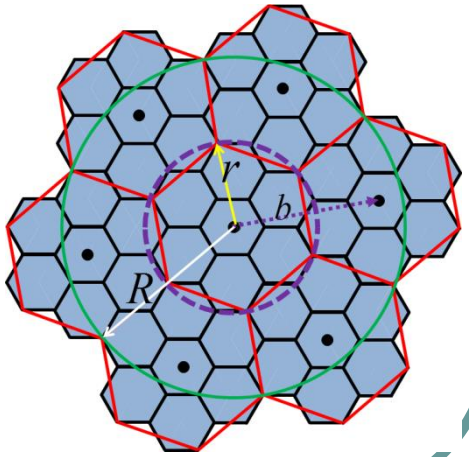












FMM for stellar/molecular dynamics

$$G(\mathbf{r}) = \frac{1}{r}, \quad r = |\mathbf{r}| = \sqrt{x^2 + y^2 + z^2},$$

then

$$G(\mathbf{r} - \mathbf{r}_0) = \frac{1}{|\mathbf{r} - \mathbf{r}_0|} = \frac{1}{\sqrt{(\mathbf{r} - \mathbf{r}_0) \cdot (\mathbf{r} - \mathbf{r}_0)}} = \frac{1}{\sqrt{r^2 - 2\mathbf{r} \cdot \mathbf{r}_0 + r_0^2}}$$

$$= \frac{1}{\sqrt{r^2 - 2rr_0 \cos\theta + r_0^2}} = \frac{1}{\sqrt{r^2 - 2\mu r r_0 + r_0^2}}$$

$$= \begin{cases} r_0^{-1} \sum_{n=0}^{\infty} P_n(\mu)(r/r_0)^n, & r < r_0, \\ r^{-1} \sum_{n=0}^{\infty} P_n(\mu)(r_0/r)^n = r_0^{-1} \sum_{n=0}^{\infty} P_n(\mu)(r/r_0)^{-n-1}, & r > r_0. \end{cases}$$

At $r = r_0$ the series also converges, if $\cos\theta \neq 1$ ($\mathbf{r} \neq \mathbf{r}_0$).

Addition Theorem for Spherical Harmonics

Spherical Harmonics

$$P_n(\cos\theta) = \frac{4\pi}{2n+1}\sum_{m=-n}^{n} Y_n^{-m}(\theta',\varphi')Y_n^m\left(\hat{\theta},\hat{\varphi}\right),$$

order

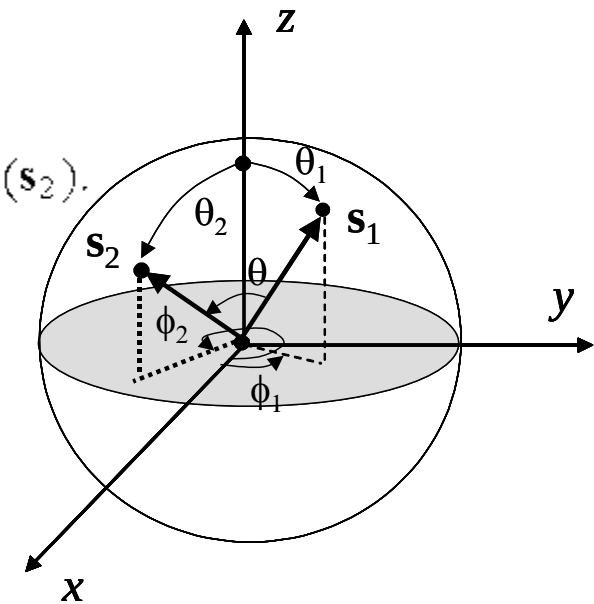
$$Y_n^m(\theta,\varphi) = (-1)^m\sqrt{\frac{2n+1}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}}\,P_n^{|m|}(\mu)e^{im\varphi}, \quad \mu = \cos\theta.$$

where θ is the angle between two points on a sphere with spherical angles (θ',φ') and $\left(\hat{\theta},\hat{\varphi}\right)$.

degree

$$P_n(\mathbf{s}_1\cdot\mathbf{s}_2) = \frac{4\pi}{2n+1}\sum_{m=-n}^{n} Y_n^{-m}(\mathbf{s}_1)Y_n^m(\mathbf{s}_2) = \frac{4\pi}{2n+1}\sum_{m=-n}^{n} Y_n^m(\mathbf{s}_1)Y_n^{-m}(\mathbf{s}_2).$$

Vector form of the addition theorem



Data structures for the FMM

- ***Separation:*** Cells satisfy local separation and WSPD properties across all levels.
- ***Indexing:*** Cell indices satisfy some order relation in memory.
- ***Spatial addressing:*** Cell centers are computable from addresses and each particle can find its bounding cell.
- ***Hierarchical addressing:*** Cell children, parent, and neighbor indices are computable from addresses.
- All this must be done in O(N log N) and must be parallelizable

1. Nail A. Gumerov, Ramani Duraiswami, and Y.A. Borovikov. Data structures, optimal choice of parameters, and complexity results for generalized fast multipole methods in d dimensions. CS/UMIACS UMIACS-TR-2003-28, CS-TR-4458.
2. Qi Hu, Nail Gumerov and Ramani Duraiswami. Parallel Algorithms for FMM Data Structures. 2011
3. Yuancheng Luo and Ramani Duraiswami Alternative Tilings for the Fast Multipole Method on the Plane submitted. Available on arXiv. 2012

GPU/heterogeneous FMM

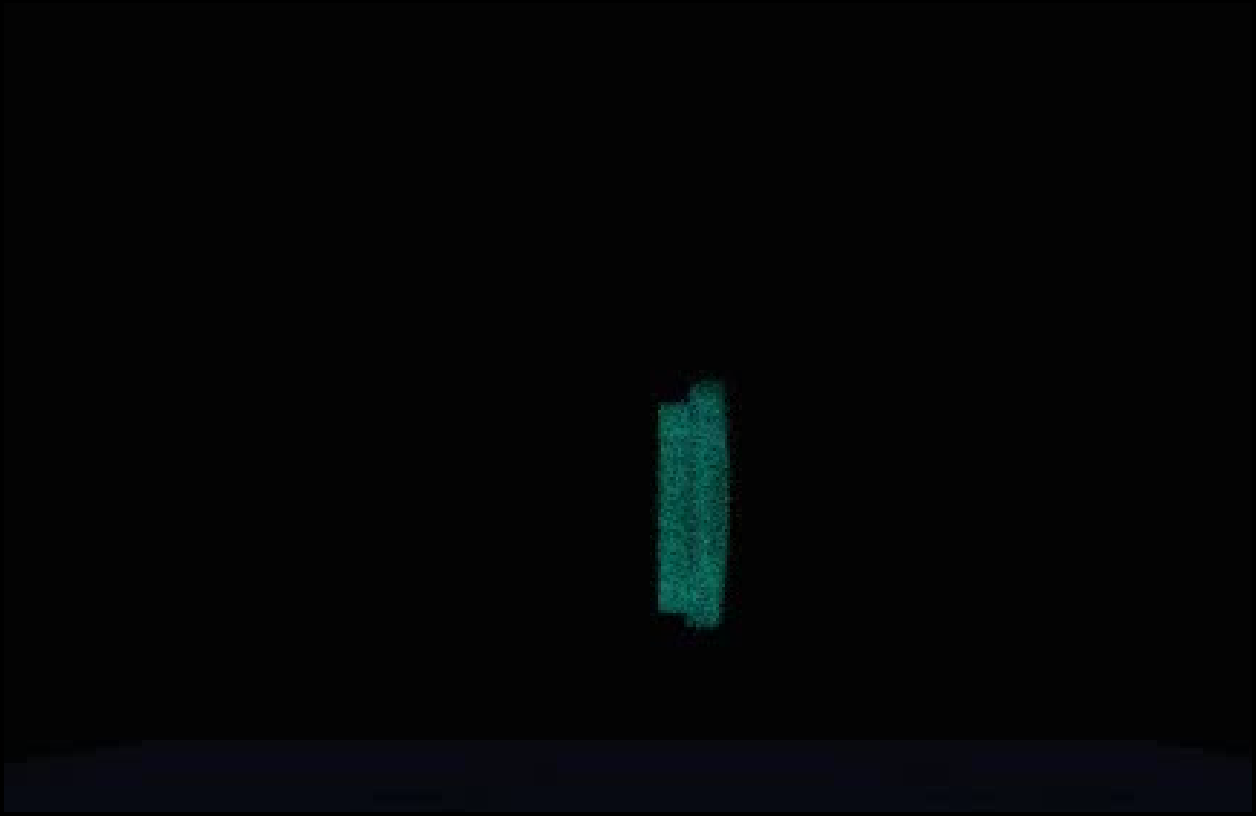
- Moved to implement FMM on GPU almost with the introduction of CUDA (2007)
- FMM translation based on rotation translation, and showing that it is competitive or superior to asymptotically faster diagonal translations
- Real valued formulations
- Algorithm splitting

Nail A. Gumerov, Ramani Duraiswami. Fast Multipole Methods on Graphics Processors, Journal of Computational Physics, 2008.

Two vortex rings interaction demo

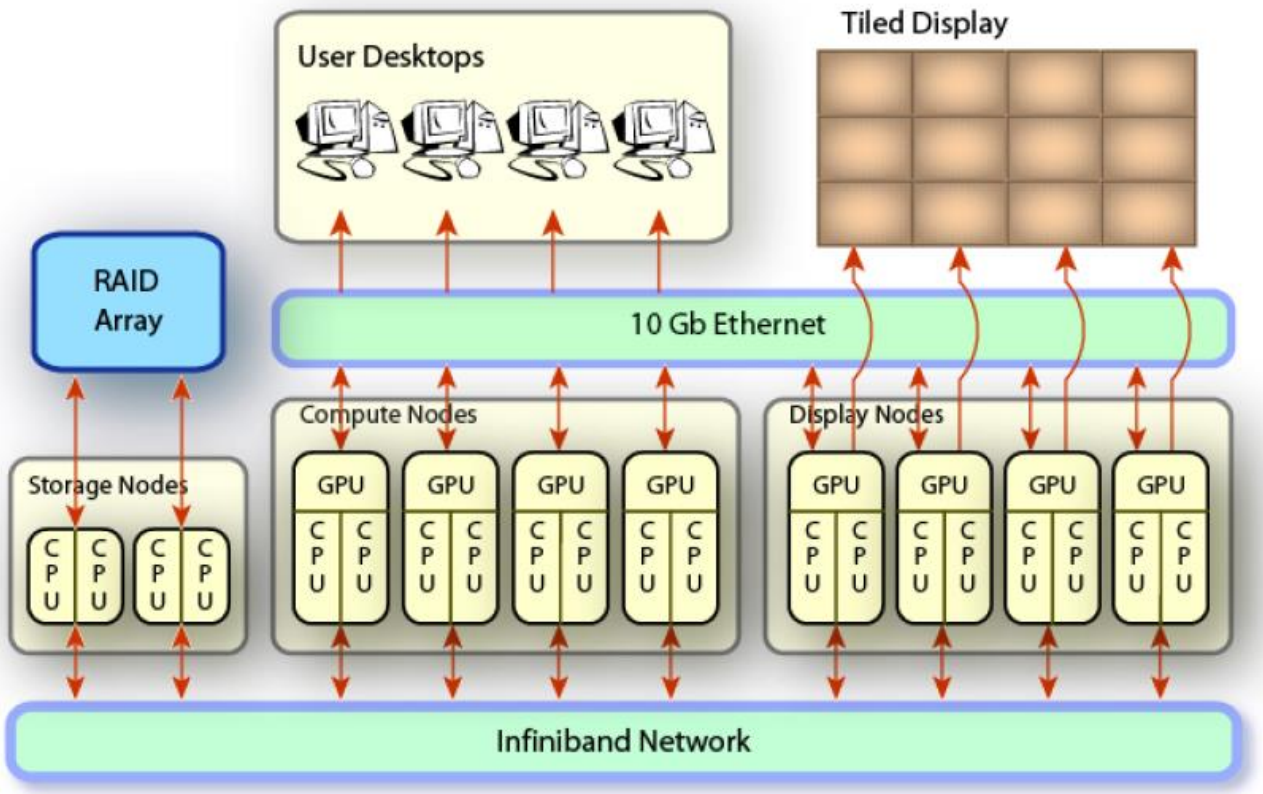
- **Two vortex rings move at the same direction**
- **Two vortex rings collision**

- **Visualization as computation proceeds!**



Heterogeneous architectures

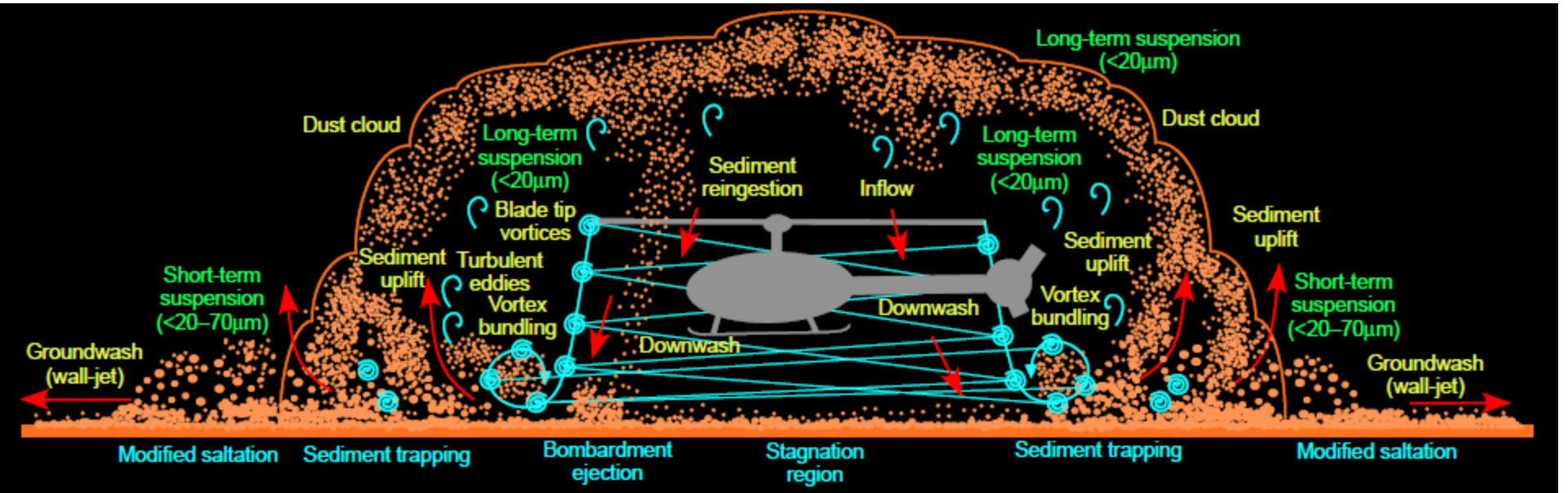
- GPUs and MIC live on a bus which sits inside a CPU box
- Insight --- why do gymnastics to fit tree algorithms with irregular access on the GPU
- CPU is multicore
- Let the GPU do stuff it is good at and let CPU do what it is good at



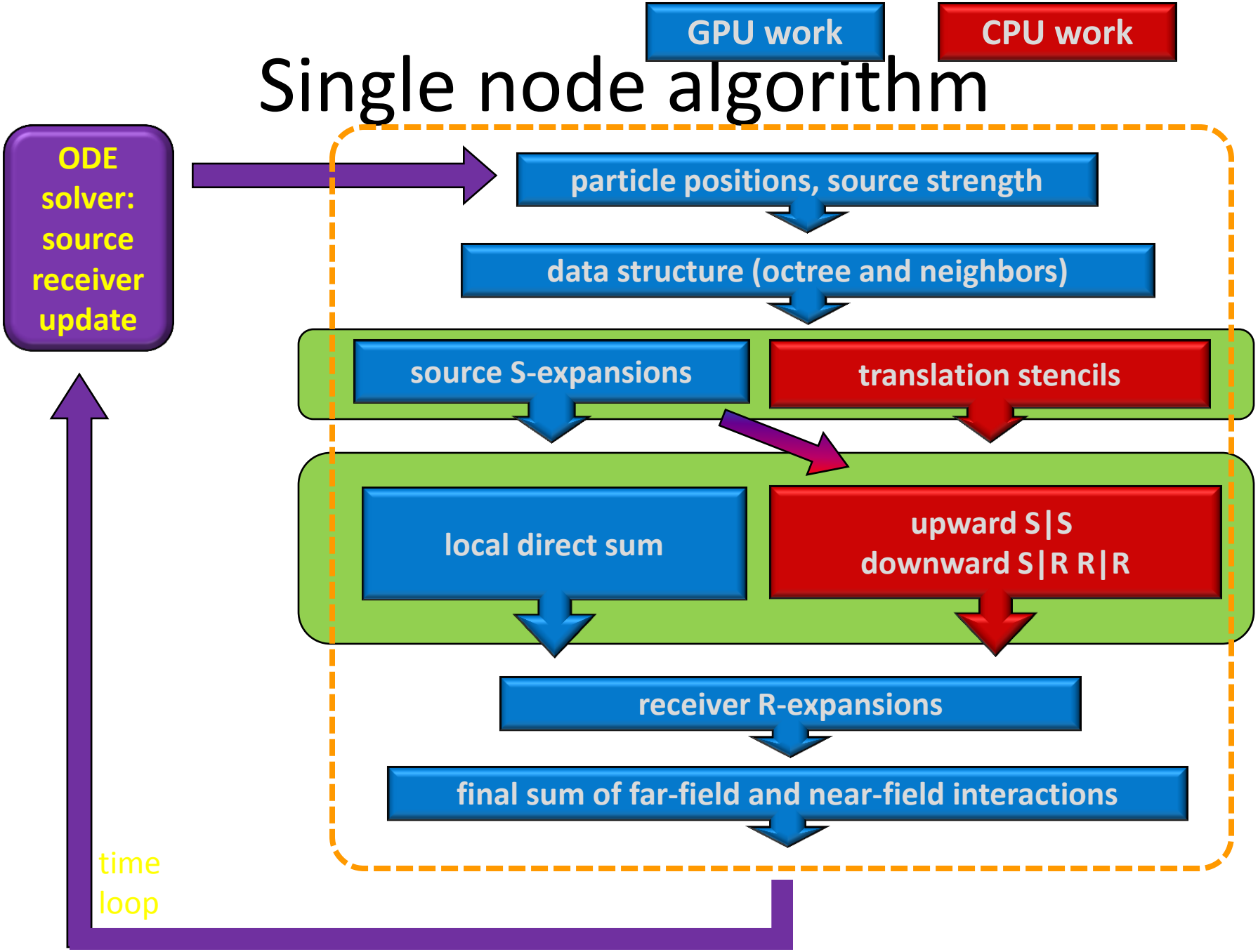
Motivation -- Helicopter Brownout



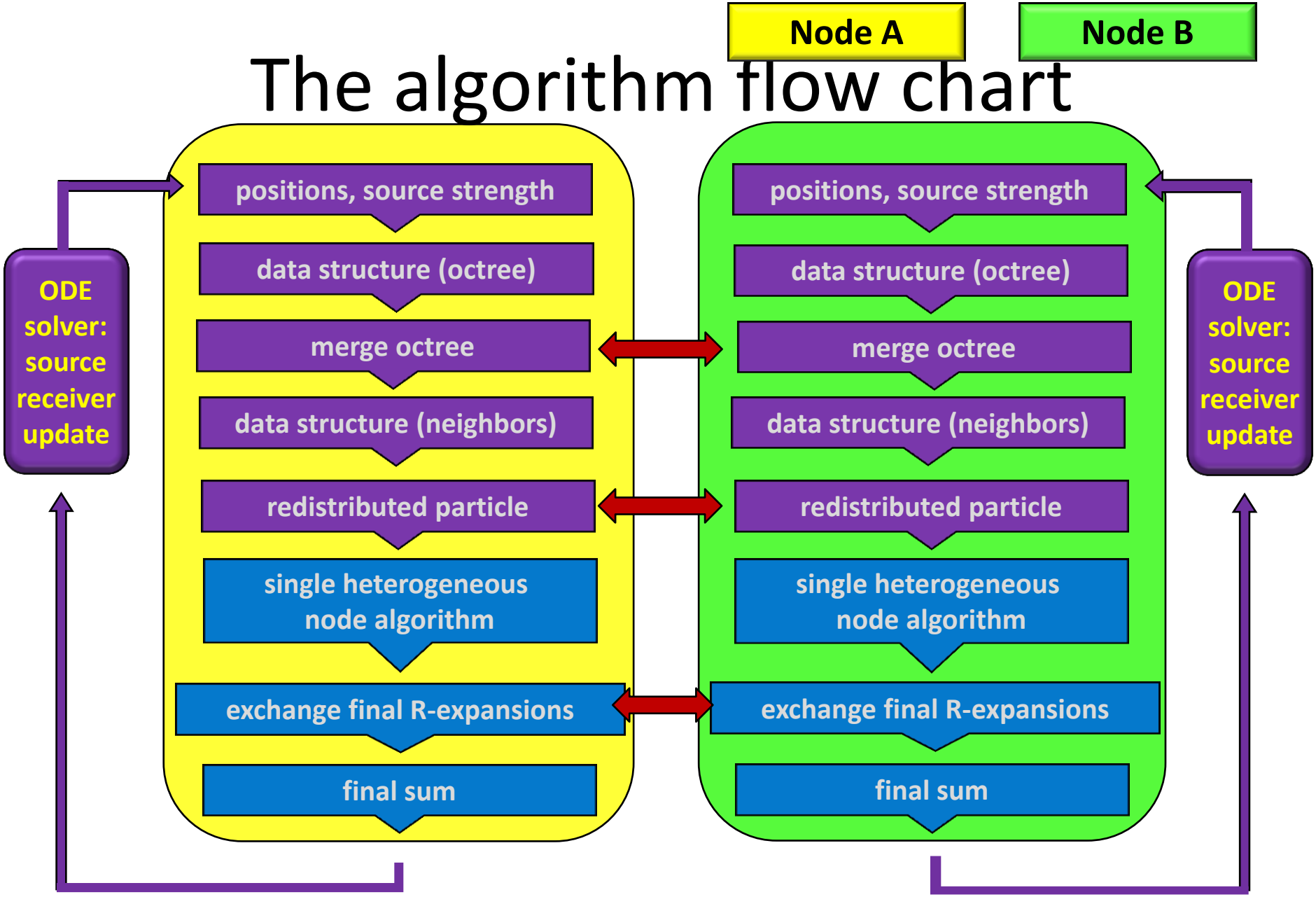
- Complicated phenomena involving interaction between rotorcraft wake, ground, and dust particles
- Causes accidents to poor visibility and damage to helicopters
- Understanding can lead to mitigation strategies



Single node algorithm

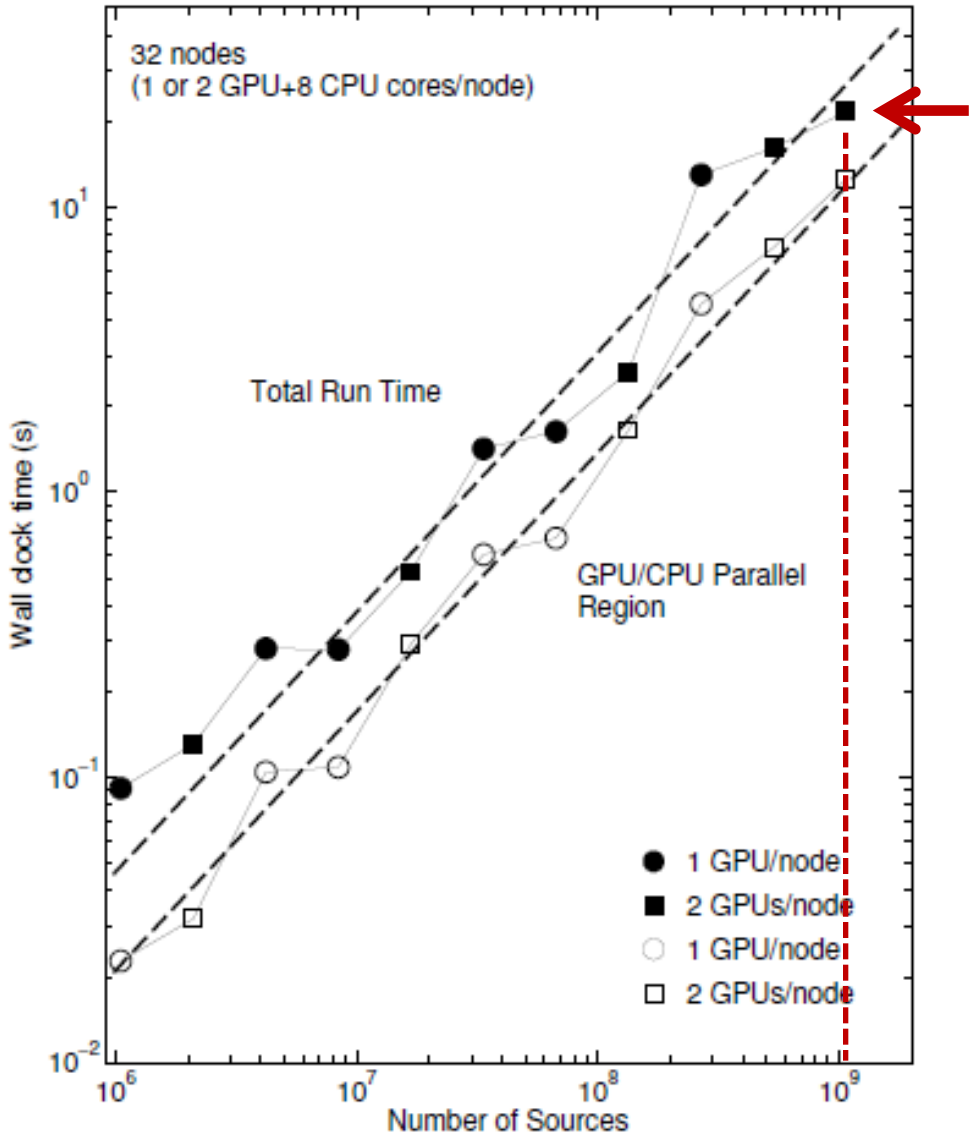


The algorithm flow chart



The billion size test case

- Using all 32 Chimera nodes
- Timing excludes the partition
- 1 billion runs potential kernel in 21.6 s



Scalar potential formulation of Equations of Mathematical Physics

- Vector valued functions are subject to constraints
  - E.g. Stokes flow, Maxwell
- Represent solutions via “potentials”, and remove degrees of freedom via gauge selection
- Approach not usable with FMM directly
- With Nail Gumerov developed this idea for FMM and showed speedup over best other algorithm proposed
- Biharmonic (2/5); Stokes (3/4); Vortex (2/5); Maxwell (2/6); Elasticity (in progress)
- Boundary element formulations for all these

Interpolation

$$\nabla^{2n}\phi = 0$$ **Polyharmonic**

$$\nabla^4 \mathbf{u} = \mathbf{0}$$
$$\nabla \cdot \mathbf{u} = \mathbf{0}$$
Elasticity

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial}{\partial t}\mathbf{B}$$
$$\nabla \times \mathbf{B} = \frac{4\pi}{c}\mathbf{J} + \frac{1}{c}\frac{\partial}{\partial t}\mathbf{E}$$
$$\nabla \cdot \mathbf{E} = 4\pi\rho$$
$$\nabla \cdot \mathbf{B} = \mathbf{0}$$
Maxwell

Electrodynamics
Plasma physics

Waves, Diffusion,
Quantum mechanics
(spectral domain,
k is complex)

Electrostatics,
Gravity,
Potential flows

$$\nabla^2\phi = 0$$ **Laplace**
$$\nabla^2\phi = f$$ **Poisson**

$$\nabla^2\phi + k^2\phi = 0$$ **Uniform Helmholtz**
$$\nabla^2\phi + k^2\phi = f$$ **Non-uniform Helmholtz**

$$\nabla^2 \mathbf{u} = \mathbf{f}$$
$$\nabla \cdot \mathbf{u} = q$$
Constrained Vector Poisson

$$\nabla^2 \mathbf{u} - \nabla p = \mathbf{0}$$
$$\nabla \cdot \mathbf{u} = \mathbf{0}$$
Stokes

$$\mu\nabla^2 \mathbf{u} - \nabla p = \rho\frac{\partial}{\partial t}\mathbf{u}$$
$$\nabla \cdot \mathbf{u} = \mathbf{0}$$
Unsteady Stokes

Vortical flows,
Reconstruction of
vector fields

Creeping flows,
Micro- and
nanofluidics

High Strouhal
number flows

Fluid dynamics

Teaching the FMM

- Course notes since 2004 online
- www.umiacs.umd.edu/~ramani/cmsc878R
- Java Applet that demos the FMM
- 10 papers directly from the course
- enabled solidifying broad themes

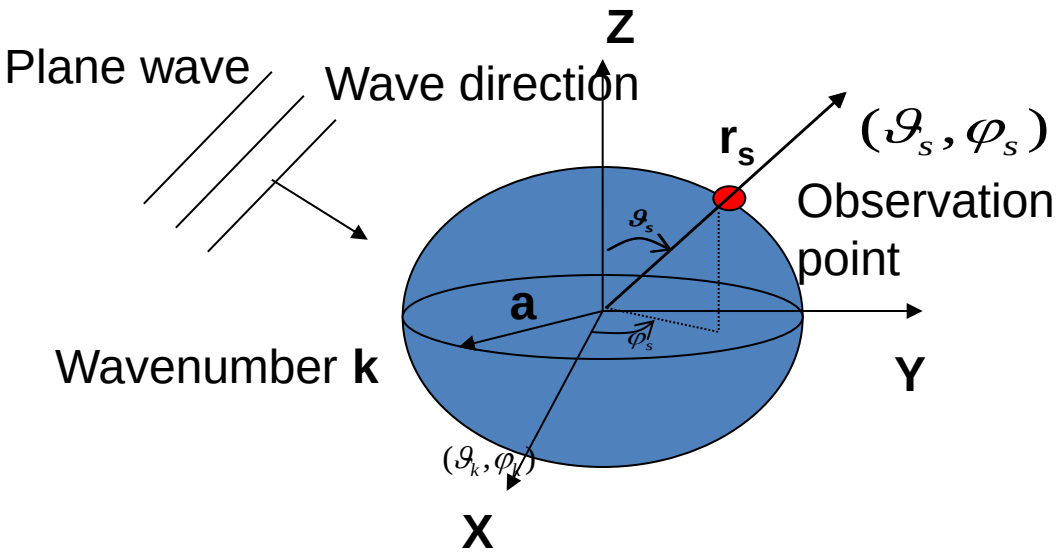
A computational camera for spatial sound

VisiSonics Corporation

www.visisonics.com

Spherical Beamforming

$$(\nabla^2 + k^2)\psi(\mathbf{r}) = 0$$



wave scattering from a rigid (sound hard) surface
find solution to Helmholtz equation which satisfies:
the rigid surface,
$\partial \psi / \partial n = 0$
radiation condition on ψ_{scat}

$$\psi(\boldsymbol{\theta}_s, \boldsymbol{\theta}_k, ka)$$
$$= [\psi_{in}(\mathbf{r}_s, \mathbf{k}) + \psi_{scat}(\mathbf{r}_s, \mathbf{k})]|_{r_s=a}$$
$$= 4\pi \sum_{n=0}^{\infty} i^n b_n(ka) \sum_{m=-n}^{n} Y_n^m(\boldsymbol{\theta}_k) Y_n^{m*}(\boldsymbol{\theta}_s),$$

$$b_n(ka) = j_n(ka) - \frac{j_n'(ka)}{h_n'(ka)} h_n(ka),$$

Allows the creation of arbitrary beampatterns using spherical harmonics

VisiSonics 5/64 Audio Visual Camera

64 microphones: 20 Hz – 20 kHz, 48 kHz sampling, 24 bit digitization

Collects video with 5 HD cameras

Produces a frame-rate video panorama using proprietary GPU based stitching algorithms

Performs beamforming in over 10,000 directions at frame-rate, to produce “audio images”

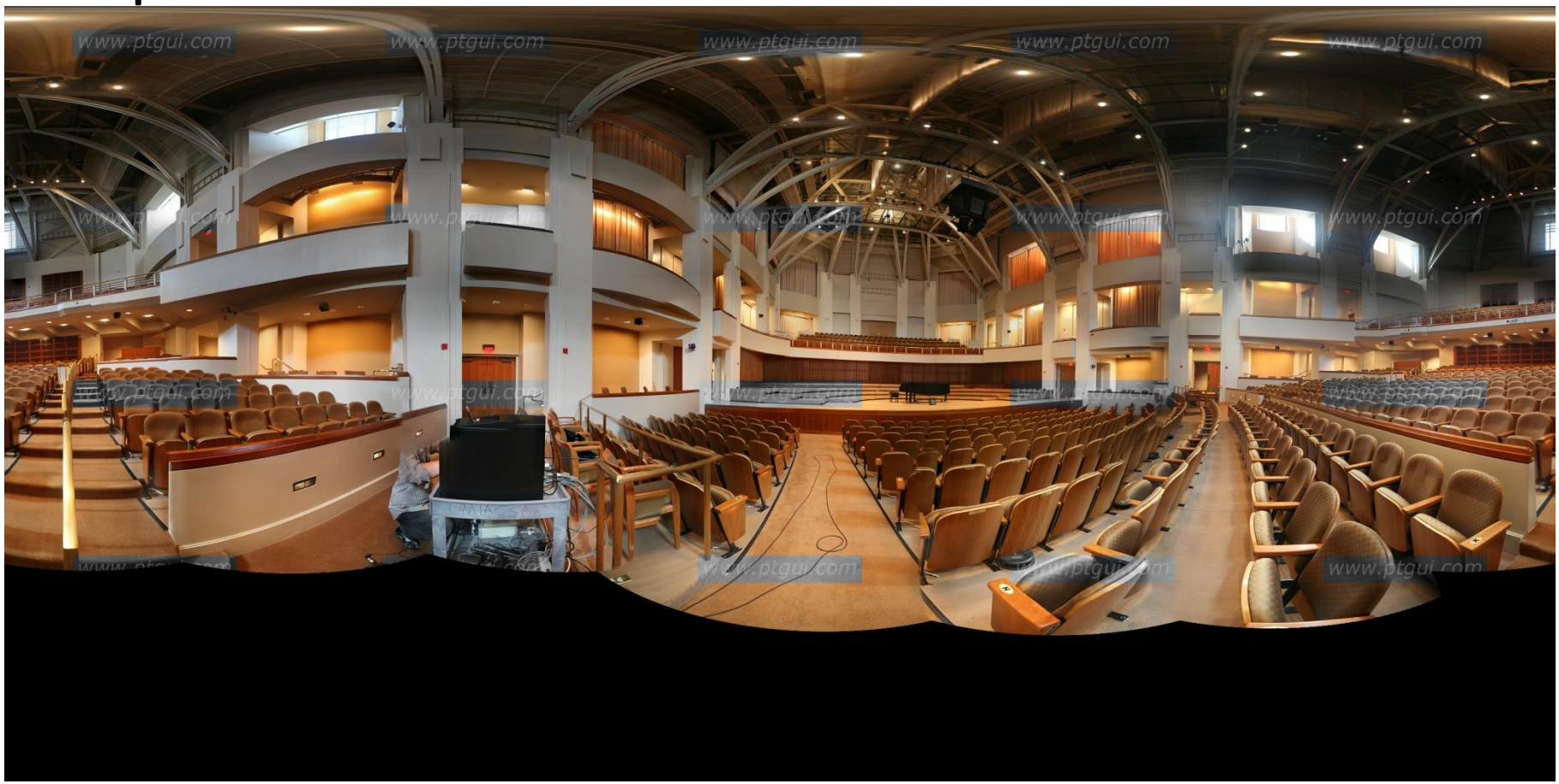
Portable – the compact microphone and camera layout is achieved via a novel microphone architecture

Connects to Laptop with GPU, no other expensive hardware

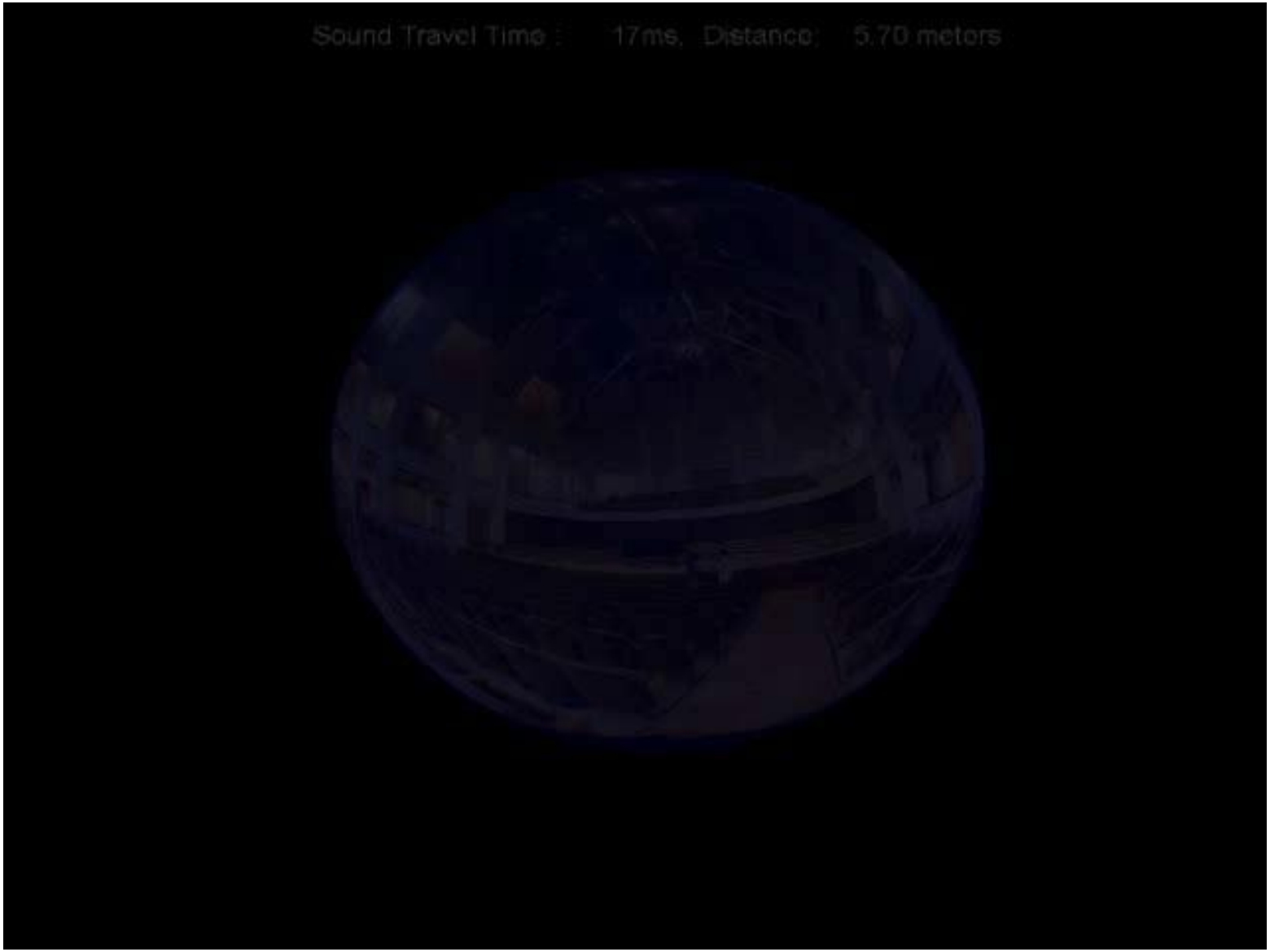
Offline processing of listening spaces enabled by a spectrotemporal analysis tool.

Dekelbaum theater at Clarice Smith Performing arts Center at UMD

- Mercator projection created from 24 snapshots

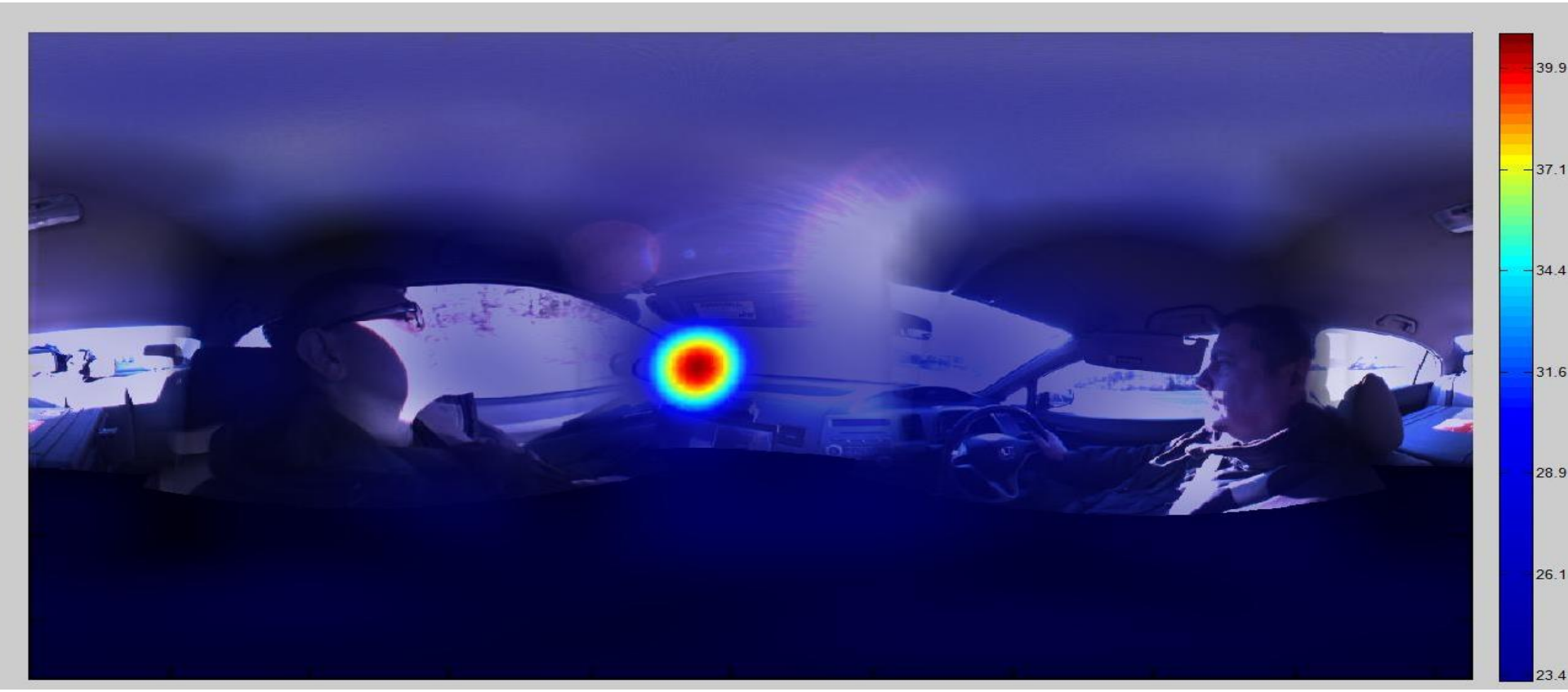


Studying Reverberation



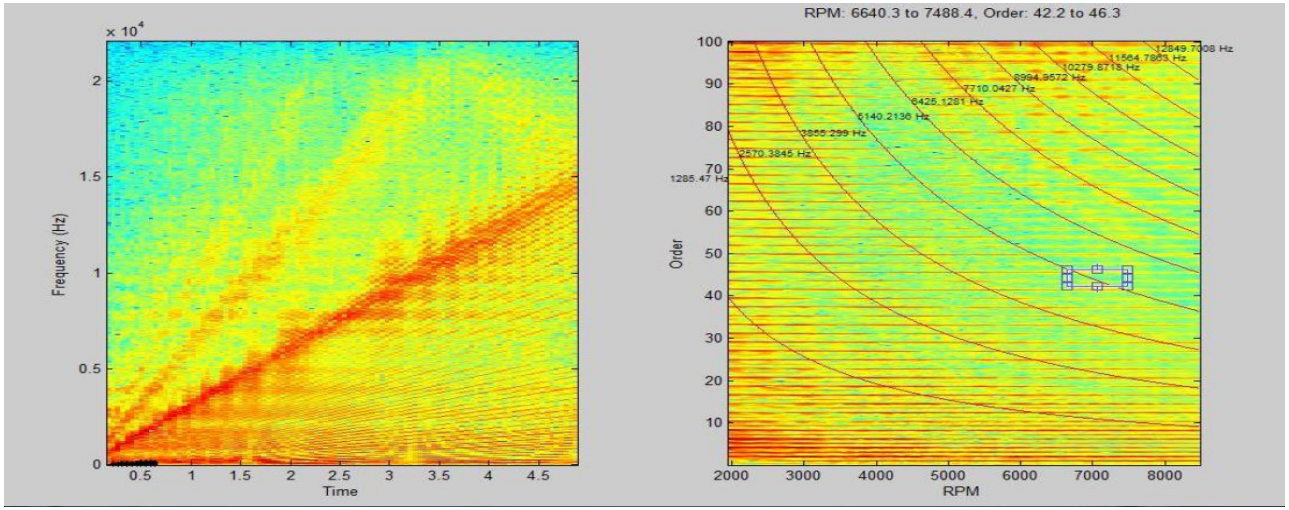
Surveillance

In Car Example Image



Direct Audio Imaging of Harmonic Orders

- Order Tracking module



Conclusions

- Two successful applications shown where GPUs were used to accelerate computing
- In each case accounting for the fact that GPUs are a component in a heterogeneous system lead to success
- GPUs are important, but the algorithm/system has to be considered in its entirety
- Fast multipole summation in heterogeneous environment has lead to the best scaled code capable of exascale performance
- Combining signal acquisition with cameras and microphones, PC based control and display, GPU based beamforming and video stitching, has lead to a unique tool for audio scene analysis