IEEE HPEC 1-1: General Purpose GPU Computing Session

GPU-Accelerated Discontinuous Galerkin Methods: 30x Speedup on 345 Billion Unknowns

September 21, 2020

Andrew C. Kirby[1]

Dimitri Mavriplis[2]

[1]Research Scientist, Scientific Simulations, LLC.

[2]Max Castagne Professor, M.E. University of Wyoming

COLLEGE OF ENGINEERING & APPLIED SCIENCE

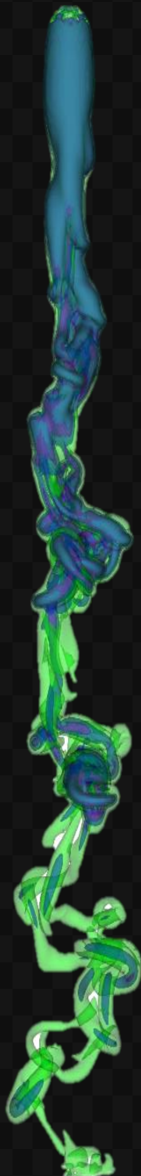
UNIVERSITY of WYOMING

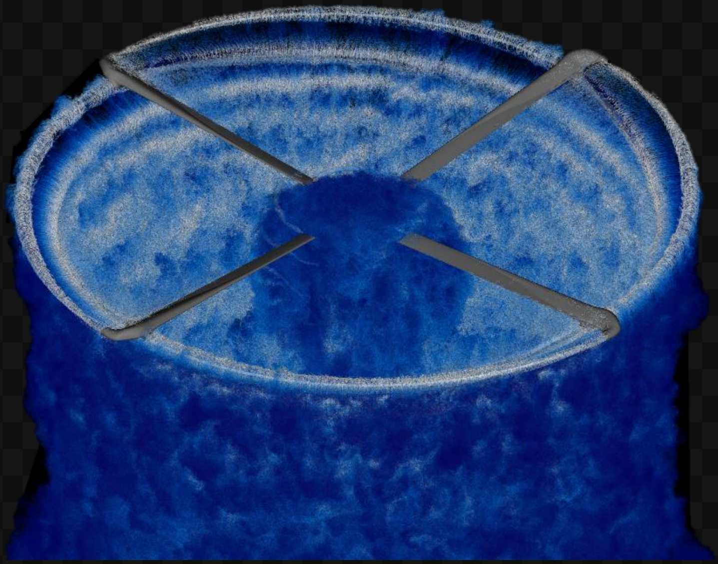
COLLEGE OF ENGINEERING & APPLIED SCIENCE

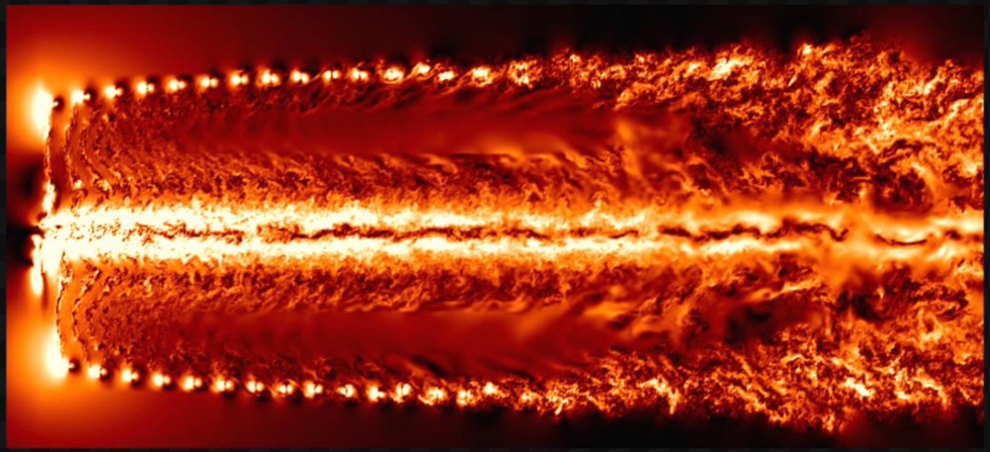
UNIVERSITY of WYOMING

COLLEGE OF ENGINEERING & APPLIED SCIENCE
UNIVERSITY of WYOMING

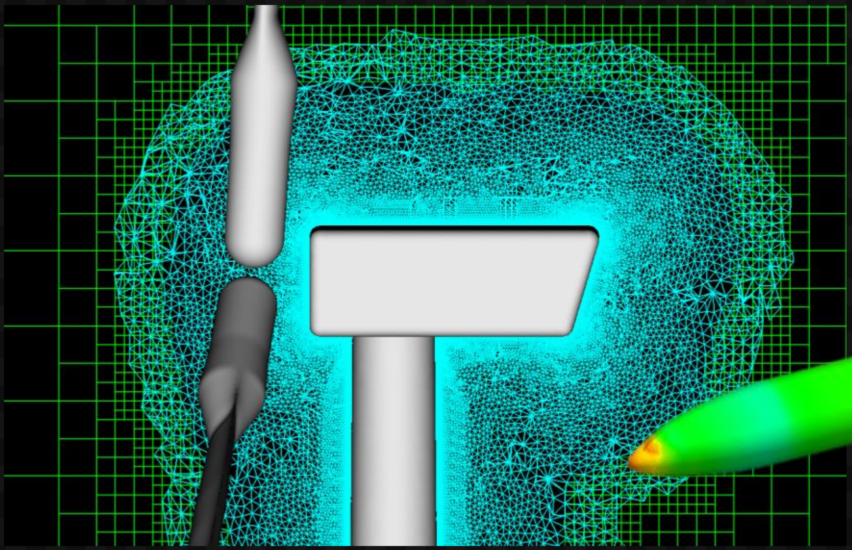
Computational Fluid Dynamics



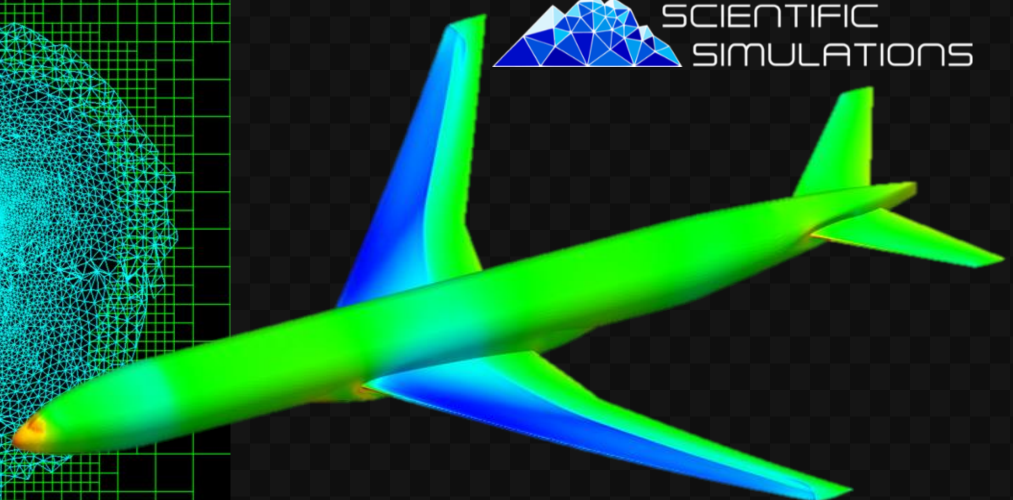




Wyoming Wind and Aerospace Applications Komputation Environment

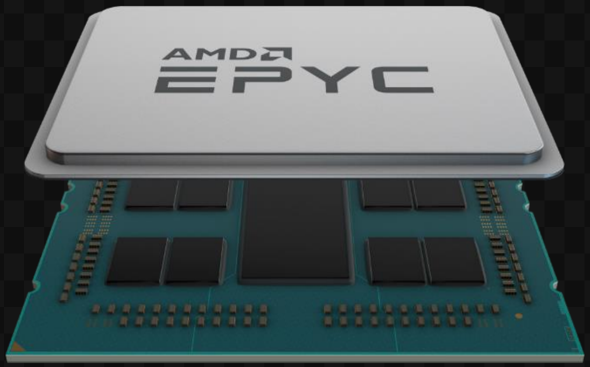


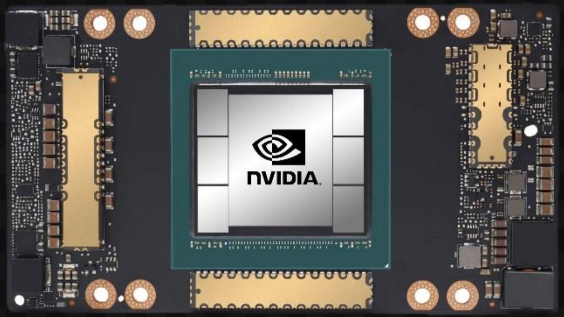
SCIENTIFIC SIMULATIONS



Motivation

- Leadership-Class Supercomputing Facilities: Heterogeneous Computing





NVIDIA A100
9.7 TFLOPS Double Precision

Governing Equations

Compressible Euler Equations

Cartesian Meshes

$$\frac{\partial \mathbf{Q}(\boldsymbol{x},t)}{\partial t} + \vec{\nabla}\cdot\mathbf{F}(\mathbf{Q}(\boldsymbol{x},t)) = 0$$

$$\mathbf{Q} = \begin{Bmatrix} \rho \\ \rho u \\ \rho v \\ \rho w \\ \rho E \end{Bmatrix}, \mathbf{F} = \begin{Bmatrix} \underline{\mathbf{F}^1} & \underline{\mathbf{F}^2} & \underline{\mathbf{F}^3} \\ \rho u & \rho v & \rho w \\ \rho u^2 + p - \tau_{11} & \rho u v - \tau_{12} & \rho u w - \tau_{13} \\ \rho u v - \tau_{21} & \rho v^2 + p - \tau_{22} & \rho v w - \tau_{23} \\ \rho u w - \tau_{31} & \rho v w - \tau_{32} & \rho w^2 + p - \tau_{33} \\ \rho u H + q_1 - \tau_{1j} u_j & \rho v H + q_2 - \tau_{2j} u_j & \rho w H + q_3 - \tau_{3j} u_j \end{Bmatrix}$$

Discretization

$$\frac{\partial \mathbf{Q}\left(\boldsymbol{x},t\right)}{\partial t} + \vec{\nabla} \cdot \mathbf{F}\left(\mathbf{Q}\left(\boldsymbol{x},t\right)\right) = 0$$

Discontinuous Galerkin Method

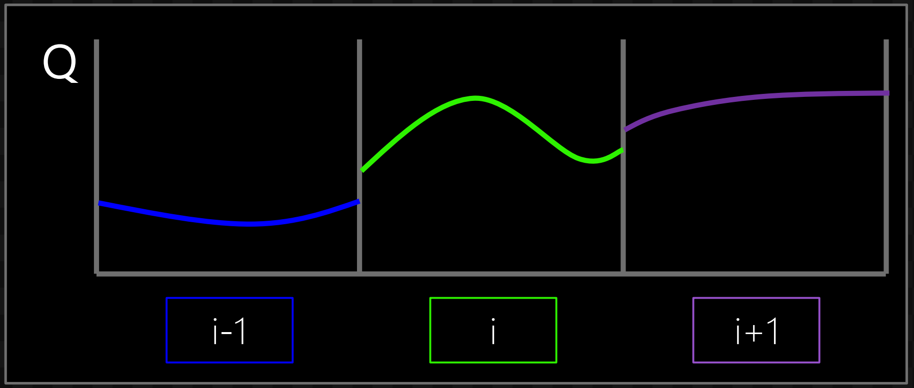
$$\int_{\Omega_k} \left(\frac{\partial \mathbf{Q}}{\partial t} + \vec{\nabla} \cdot \mathbf{F} \right) \psi(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = 0$$

$$\mathbf{R}^{\mathrm{Weak}} = \int_{\Omega_k} \frac{\partial \mathbf{Q}}{\partial t} \psi(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} - \int_{\Omega_k} \left(\mathbf{F} \cdot \vec{\nabla}\right) \psi(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} + \int_{\Gamma_k} \left(\mathbf{F}^* \cdot \vec{\mathrm{n}}\right) \psi(\boldsymbol{x}|_{\Gamma_k}) \mathrm{d}\Gamma_k = 0$$

Nodal Basis Functions

$$\ell_s\left(x\right) = \prod_{i=1, i\neq s}^{N} \frac{\left(x - \xi_i\right)}{\left(\xi_s - \xi_i\right)}$$

Lagrange Interpolating Polynomial
One-Dimensional

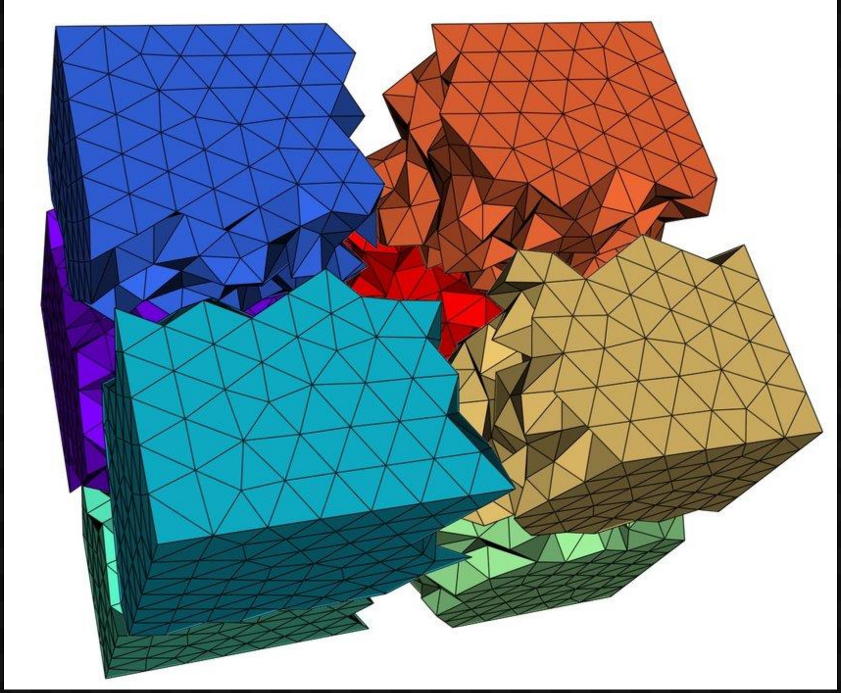


HPC

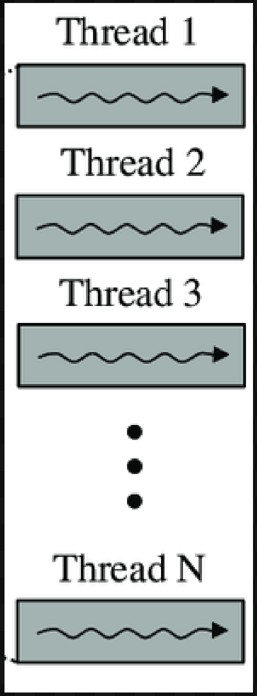
MPI + X Programming Model

$$\mathbb{M}\frac{\partial \mathbf{Q}_{ijk}(t)}{\partial t} + \mathbf{R}_{ijk}\left(\mathbf{Q}\right) = 0$$

MPI: coarse-grained parallelism

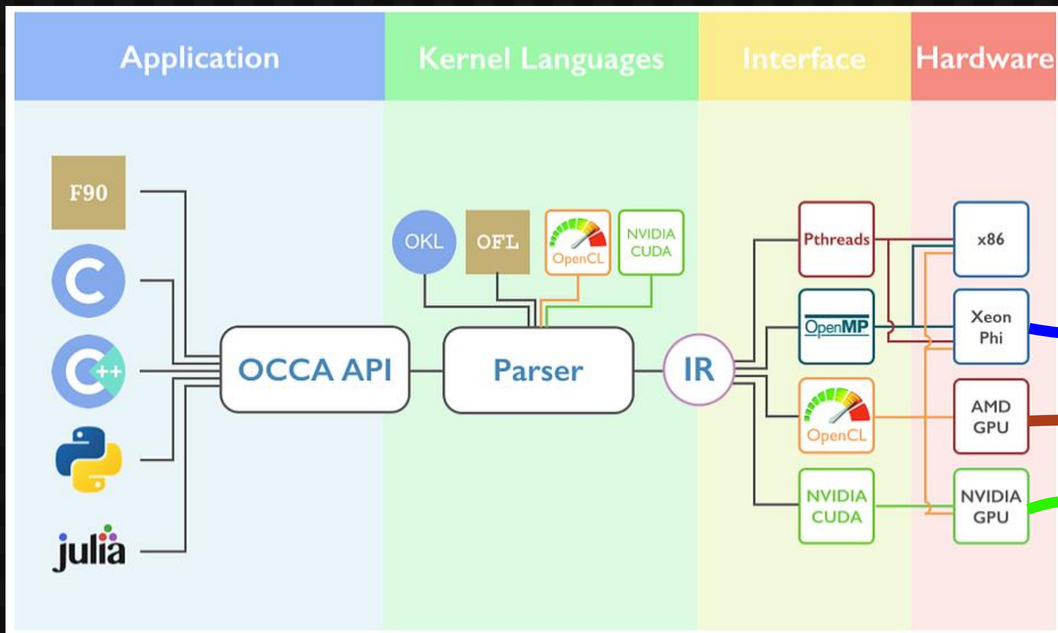


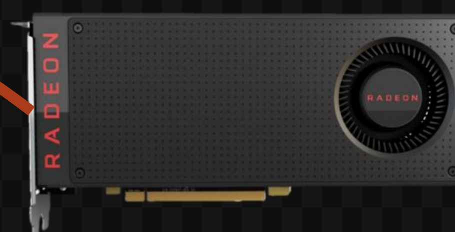
X: fine-grained parallelism

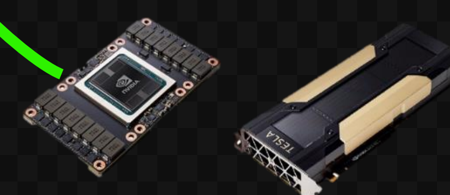


OCCA

A Unified API for programming devices







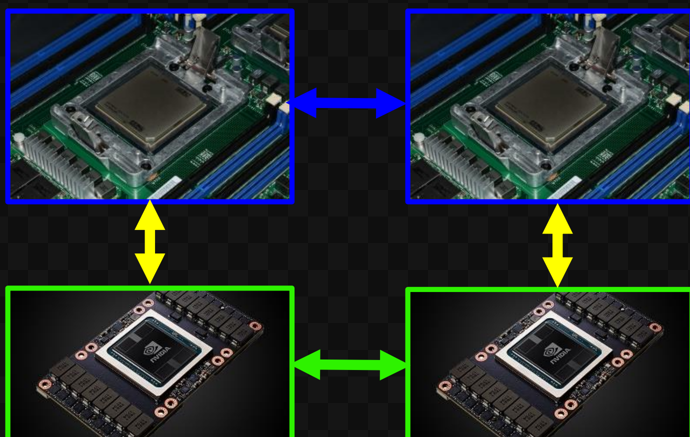
MIT Satori Results

<u>64 Nodes</u>

- 2 IBM Power9 CPUs
- 4 NVIDIA Volta V100
- EDR Infiniband

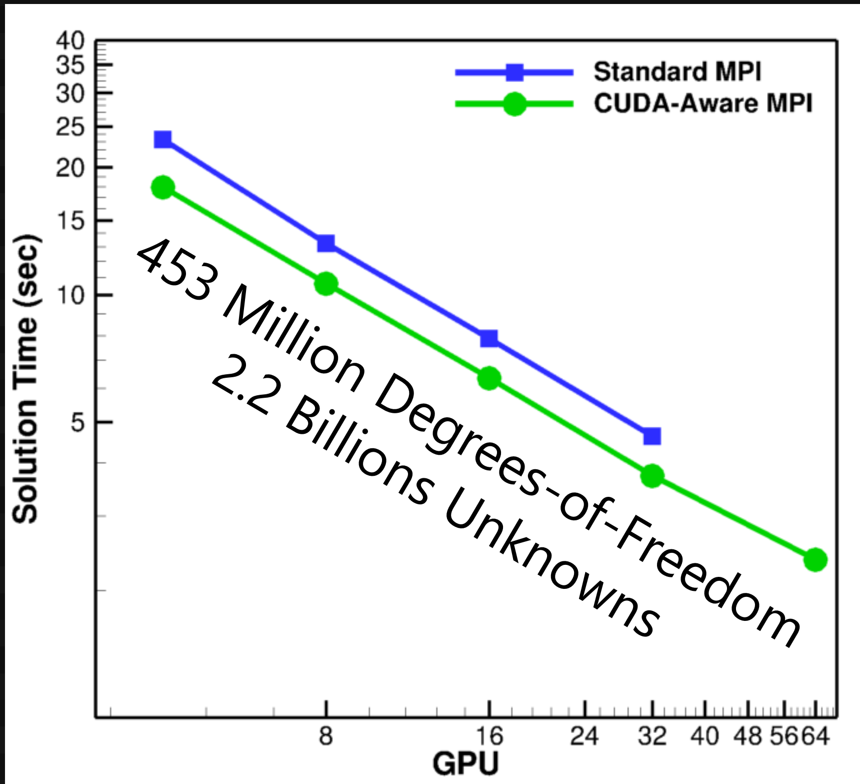
<u>Total</u>

- 256 GPUs
- #7 Green500



MIT Satori: CUDA-Aware MPI
Problem Size: 452,984,832 DOF.

	Solution Time (sec)		
GPUs	ON	OFF	Speedup
4	17.94	23.22	1.29x
8	10.60	13.22	1.25x
16	6.34	7.87	1.24x
32	3.73	4.62	1.24x



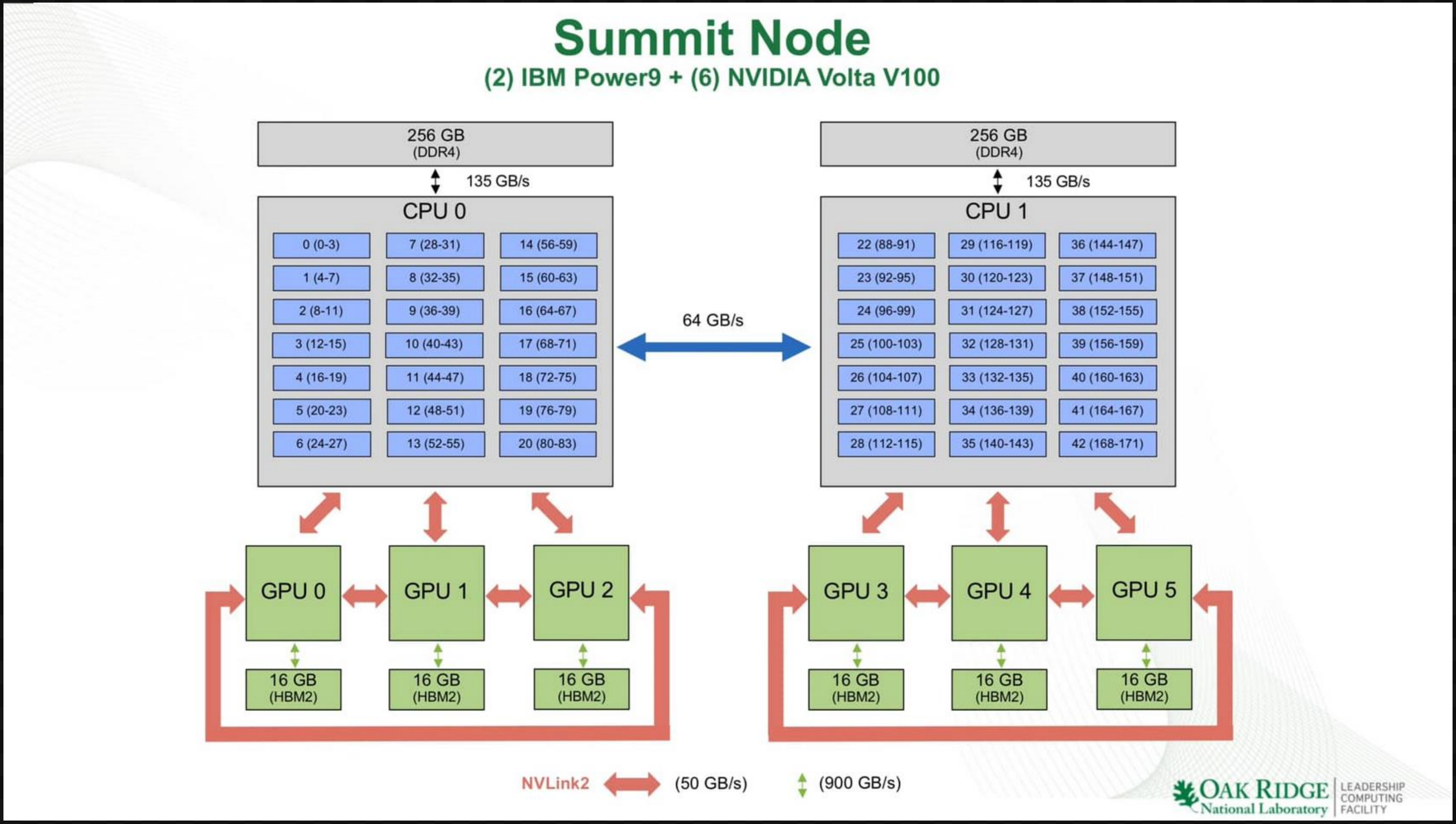
ORNL Summit

4,608 Nodes

- 2 IBM Power9 CPUs
- 6 NVIDIA Volta V100

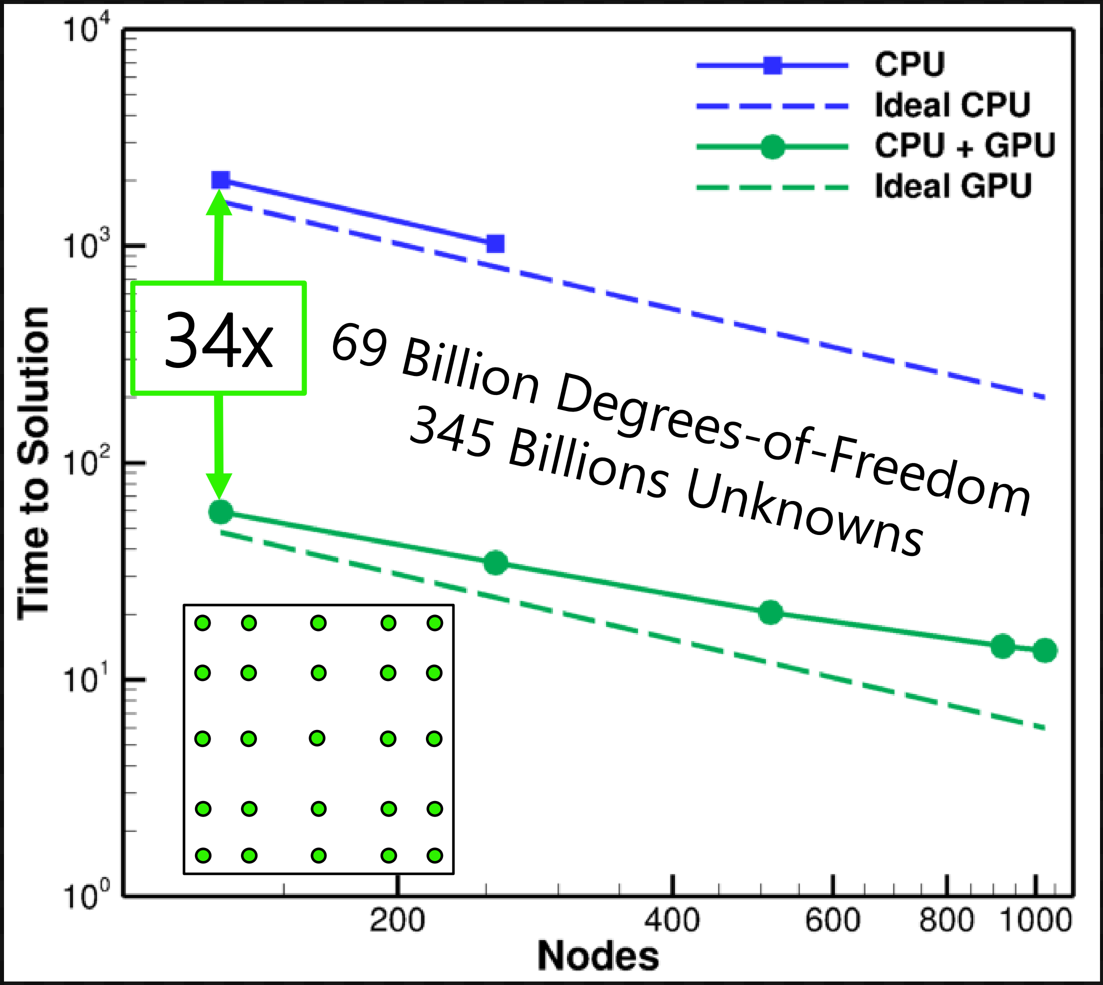
Total

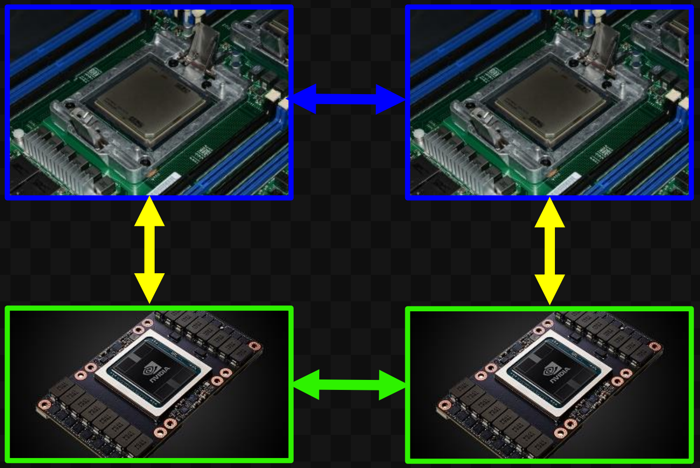
- 27,648 GPUs
- 200 PFLOPs



ORNL Summit

Non-CUDA-Aware MPI





30% Speedup with CUDA-Aware MPI

CUDA-Aware MPI		
Kernel	Time (sec)	
	ON	OFF
Volume	1.67	1.68
Surface	2.04	2.04
Update-Project	5.70	5.70
Communication	0.65	5.13
Overall	10.37	14.86

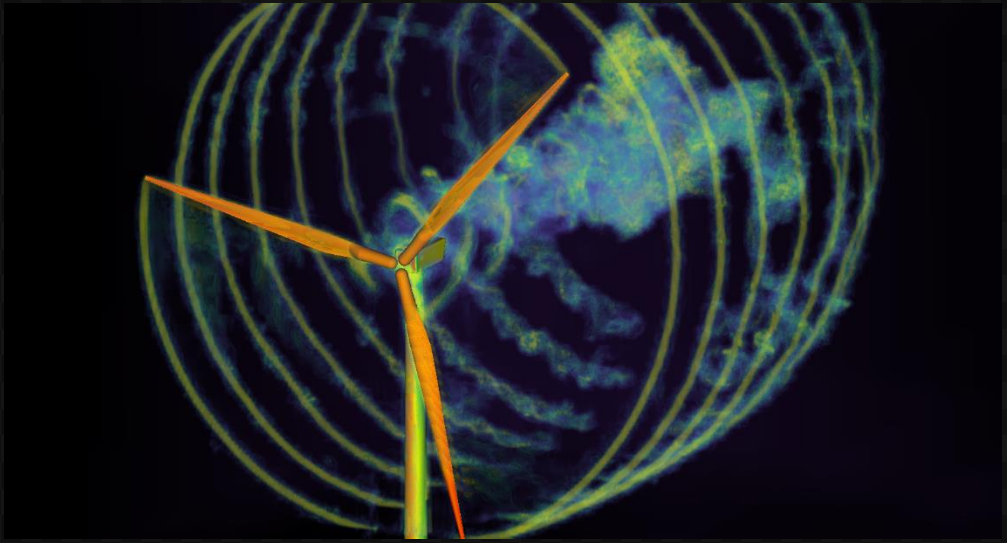
Conclusion

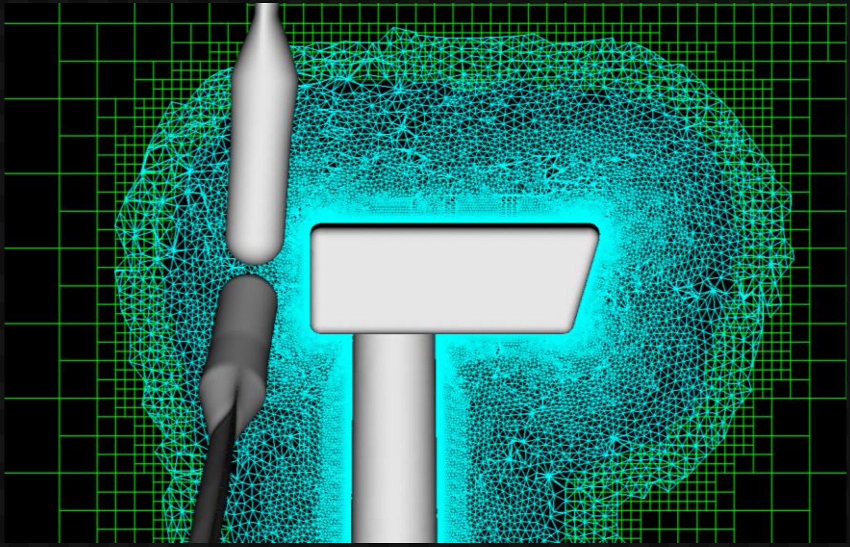
Results

- Demonstrated good performance on GPUs ~30x speedup
- Demonstrated 24-30% improvement using GPU-Direct MPI

Future Work

- Extend GPU implementation to Adaptive Mesh Refinement on unstructured hex/quad meshes
- Extend GPU implementation to overset mesh capabilities
- 1 Trillion DOF Simulations





Acknowledgements

Compute Time

- MIT: Satori Supercomputer
- ORNL: Summit Supercomputer

Office of Naval Research

- ONR Grant N00014-14-1-0045
- ONR Grant N00014-16-1-2737

U.S. Department of Energy, Office of Science, Basic Energy Sciences

- Award #DE-SC0012671

Thank You

Questions?