SAND2020-1624PE

Sandia National Laboratories

JOWOG 34 High-Order Working Group: Status and Performance of High-Order Methods in SPARC

PRESENTED BY

Travis Fisher and Jerry Watkins

Contributors: Jungyeoul Maeng, Cory Stack, and Michael Hansen

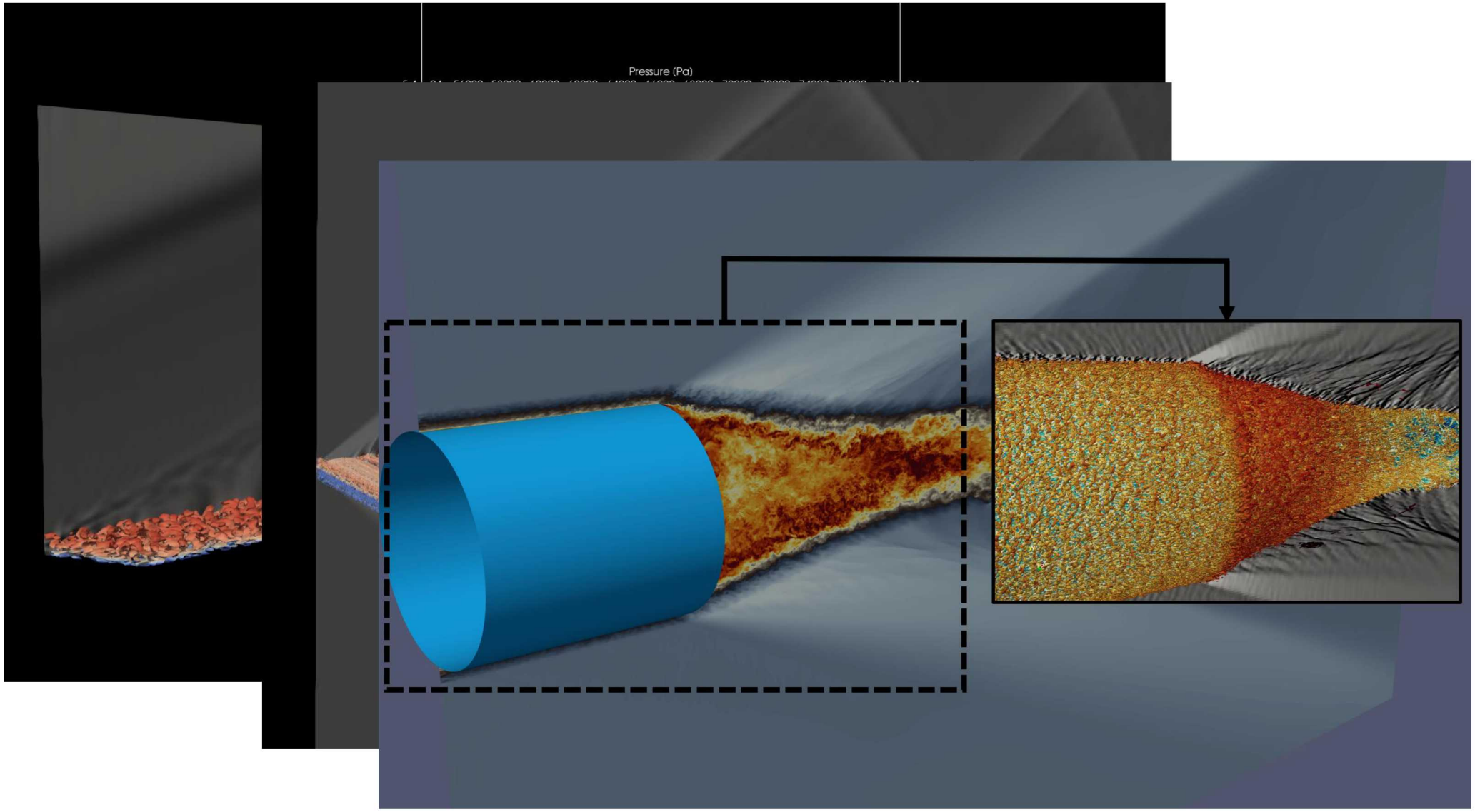
2020 JOWOG-34 Applied Computer Science Meeting
02/24/2020 - 02/27/2019
Livermore, CA

SAND

U.S. DEPARTMENT OF ENERGY NNSA

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

High-Order Methods Usage in SPARC



Wall-Modeled LES

High-Order Methods in SPARC

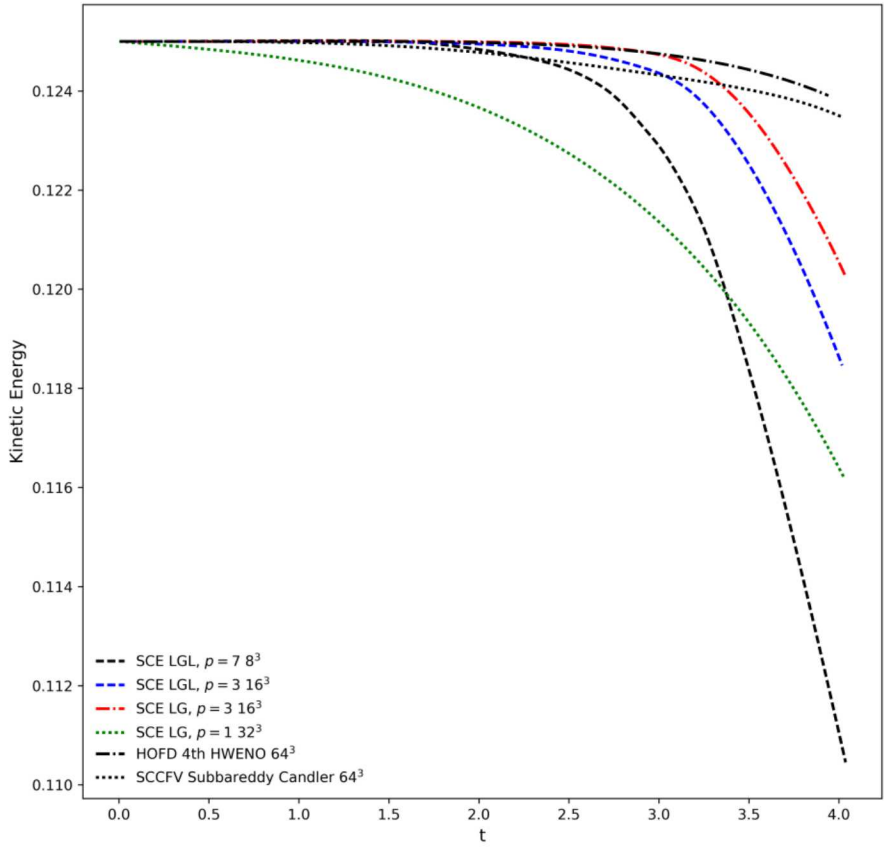
Summation-by-Parts Methods:

Entropy-Stable High-Order Finite Difference (Multi-block Structured)

- Hybrid WENO or Artificial Viscosity for shock capturing
- Cell-centered for strong inter-block coupling

Entropy-Stable Spectral Collocation Elements (Unstructured)

- No need for over-integration
- Tensor product elements
- Artificial Viscosity for shock capturing
- Legendre-Gauss or Legendre-Gauss-Lobatto solution points



Entropy-Stable Spectral Collocation Element Operators

Continuous

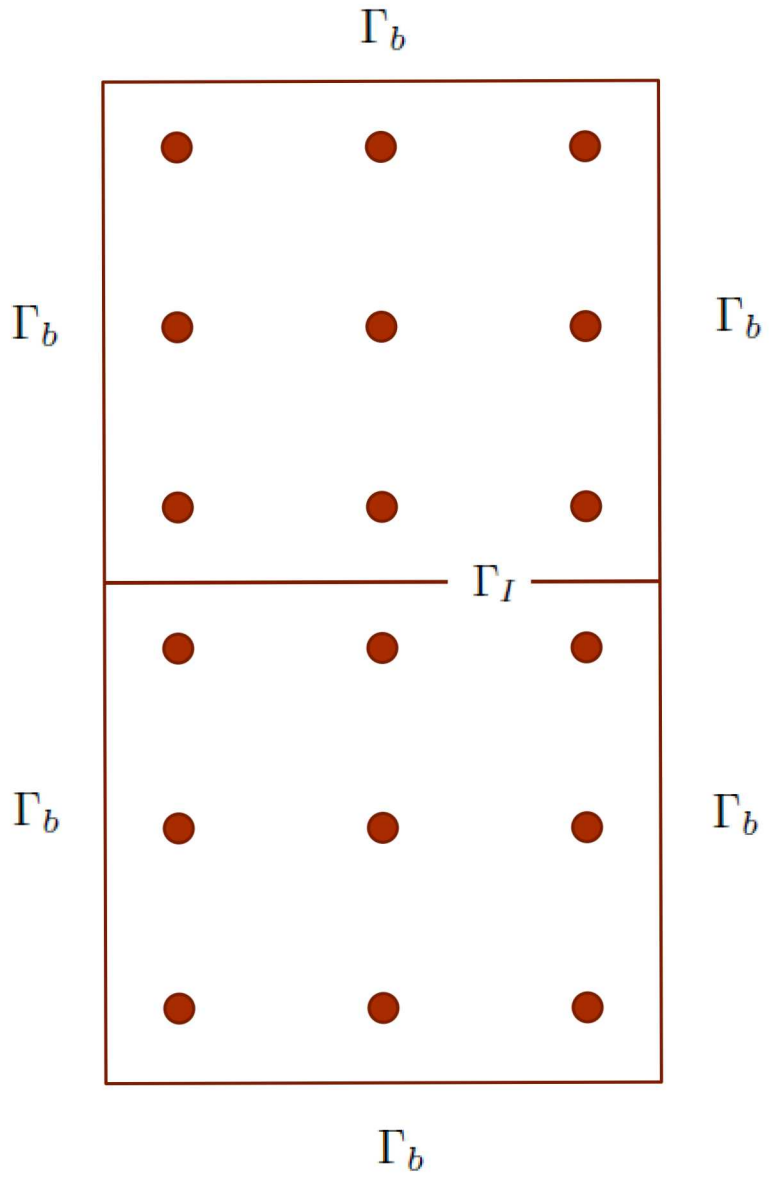
$$\partial_t q + \partial_{x_k} f_k(q) = \partial_{x_k} c_{kj} \partial_{x_j} q$$

Linearly Stable

$$\partial_t \mathcal{P}\mathbf{q} + \mathcal{Q}_k f_k(\mathbf{q}) = \mathcal{Q}_k c_{kj} \psi_j + g_b + g_I$$
$$\mathcal{P}\psi_\alpha = -\mathcal{Q}_\alpha^T \mathbf{q} + n_{\Gamma_I} b_{\alpha_{\Gamma_I}} \bar{q}_I(\mathbf{q}) + n_{\Gamma_b} b_{\alpha_{\Gamma_b}} \bar{q}_b(\mathbf{q})$$

$$\mathcal{Q}_\alpha + \mathcal{Q}_\alpha^T = \sum_\Gamma n_\Gamma b_{\alpha_\Gamma} b_{\alpha_\Gamma}^T$$

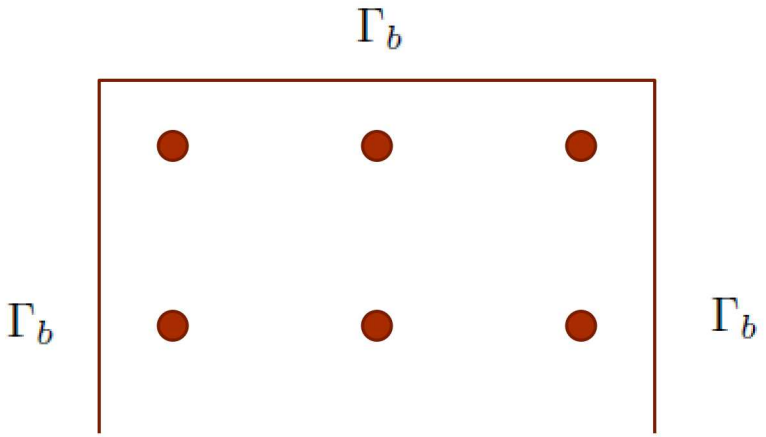
$$\partial_t \mathcal{P}\mathbf{q} + \left[\mathcal{Q}_k - \frac{1}{2} n_\Gamma b_{k_\Gamma} b_{k_\Gamma}^T\right] f_k(\mathbf{q}) = -\mathcal{Q}_k^T c_{kj} \psi_j + n_{\Gamma_b} b_{k_{\Gamma_b}} \bar{f}_b(b_{k_{\Gamma_b}}^T \mathbf{q}, b_{k_{\Gamma_b}}^T \psi_j)$$
$$+ n_{\Gamma_I} b_{k_{\Gamma_I}} \bar{f}_I(b_{k_{\Gamma_I}}^T \mathbf{q}, b_{k_{\Gamma_I}}^T \psi_j)$$



Entropy-Stable Spectral Collocation Element Operators

Continuous

$$\partial_t q + \partial_{x_k} f_k(q) = \partial_{x_k} c_{kj} \partial_{x_j} q$$



Entropy Stable

$$\partial_t \mathcal{P}\mathbf{q} + \left[2 \left(\mathcal{Q}_k - \frac{1}{2} n_\Gamma b_{k\Gamma} b_{k\Gamma}^T \right) \circ \hat{f}_k \right] \mathbf{1} = - \mathcal{Q}_k^T \hat{c}_{kj} \hat{\psi}_j$$
$$- n_{\Gamma_b} \left[b_{k\Gamma_b} b_{k\Gamma_b}^T \circ \hat{f}_{\Gamma_b} \right] \mathbf{1} + n_{\Gamma_b} b_{k\Gamma_b} g(b_{k\Gamma_b} \mathbf{w}, b_{k\Gamma_b} \hat{\psi}_j)$$
$$- n_{\Gamma_I} \left[b_{k\Gamma_I}^{(-)} b_{k\Gamma_I}^{(+)T} \circ \hat{f}_{\Gamma_I} \right] \mathbf{1} + n_{\Gamma_I} b_{k\Gamma_I}^{(-)} g(b_{k\Gamma_I}^{(-)} \mathbf{w}, b_{k\Gamma_I}^{(-)} \psi_j, b_{k\Gamma_I}^{(+)} \mathbf{w}, b_{k\Gamma_I}^{(+)} \psi_j)$$
$$\mathcal{P}\hat{\psi}_\alpha = - \mathcal{Q}_\alpha^T w(\mathbf{q}) + n_{\Gamma_I} b_{\alpha\Gamma_I}^{(-)} \bar{w}_I(\mathbf{q}) + n_{\Gamma_b} b_{\alpha\Gamma_I}^{(-)} \bar{w}_b(\mathbf{q})$$

Entropy-Stable Spectral Collocation Element Operators

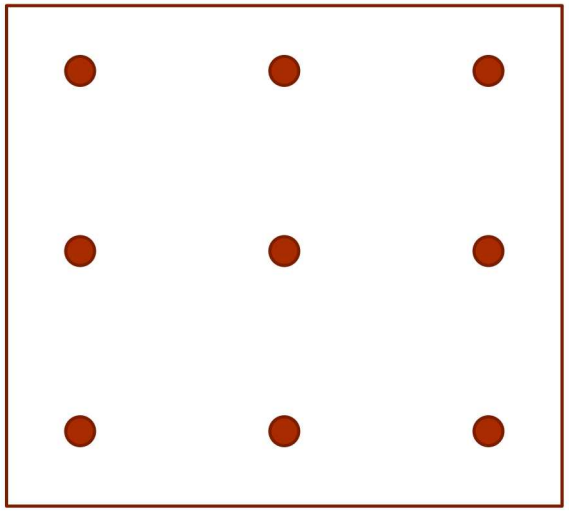
Volume terms (operators are sparse)

- Simple gradient complexity: 3*(p+1) per solution point [LG and LGL]

$$-\mathcal{Q}_\alpha^T w(\mathbf{q})$$

- Flux divergence complexity: 3*(p+1) inviscid and viscous flux evaluations per solution point [LG and LGL]

$$\left[2\left(\mathcal{Q}_k - \frac{1}{2} n_\Gamma b_{k_\Gamma} b_{k_\Gamma}^T\right) \circ \hat{f}_k\right] \mathbf{1} + \mathcal{Q}_k^T \hat{c}_{kj} \hat{\psi}_j \qquad \hat{f}_{k_{ij}} = \hat{f}_k(q_i, q_j)$$



Concurrency over (p+1)^3 solution points per element

Halo for ghosting

Mass matrix always diagonal: Exact for LG points, approximate for LGL points

Entropy-Stable Spectral Collocation Element Operators

Interface terms (operators are sparse)

- Simple gradient complexity: 2*(p+1) [LG] or 2 [LGL] per face point

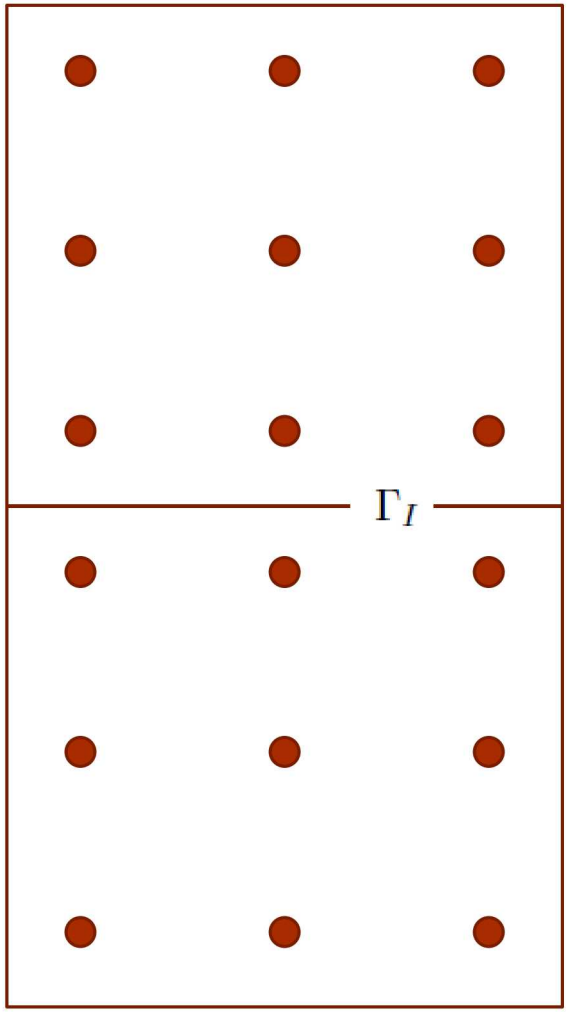
$$\sum b^{(-)}_{\alpha\Gamma_I} \bar{w}_b(\mathbf{q}) = n^{(L)}_{\Gamma_I} \left(b^{(L)}_{\alpha\Gamma_I} - b^{(R)}_{\alpha\Gamma_I} \right) \bar{w}_b \left(b^{(L)T}_{\alpha\Gamma_I} w(\mathbf{q}), b^{(R)T}_{\alpha\Gamma_I} w(\mathbf{q}) \right)$$

- Flux divergence complexity:
  - 2*(p+1)+1 [LG] or 1 [LGL] non-dissipative inviscid flux evaluations
  - 1 [LG and LGL] dissipative inviscid flux evaluation
  - 2*(p+1)+1 [LG] or 3 [LGL] viscous flux evaluations

$$- n^{(L)}_{\Gamma_I} \left[\left(b^{(L)}_{k\Gamma_I} b^{(R)T}_{k\Gamma_I} - b^{(R)}_{k\Gamma_I} b^{(L)T}_{k\Gamma_I} \right) \circ \hat{f}_{\Gamma_I} \right] \mathbf{1}$$
$$+ n^{(L)}_{\Gamma_I} \left(b^{(L)}_{k\Gamma_I} - b^{(R)}_{k\Gamma_I} \right) g \left(b^{(L)}_{k\Gamma_I} \mathbf{w}, b^{(L)}_{k\Gamma_I} \psi_j, b^{(R)}_{k\Gamma_I} \mathbf{w}, b^{(R)}_{k\Gamma_I} \psi_j \right)$$

Concurrency over (p+1)^2 face points per interface

Use face coloring to manage conflicts



Performance Analysis – Taylor-Green Vortex (TGV)

Methods:

- Structured cell-centered finite volume (**SCCFV**)
- Structured high-order finite difference (**CCHOFD**)
- Unstructured spectral collocation element (**SCE**)
  - Note: All cases using LGL points

Architectures:

- Intel Haswell (**HSW**) CPU (32 cores/node)
- NVIDIA Volta (**V100**) GPU (4 GPUs/node)

Configuration:

- Single node simulations
- 30s simulation time; explicit time step control
- 32(MPI+1OMP) vs. 2(MPI+32OMP)
  - High-order performed better with latter
  - OpenMP threads mapped to hardware-threads

3D Cartesian Meshes:

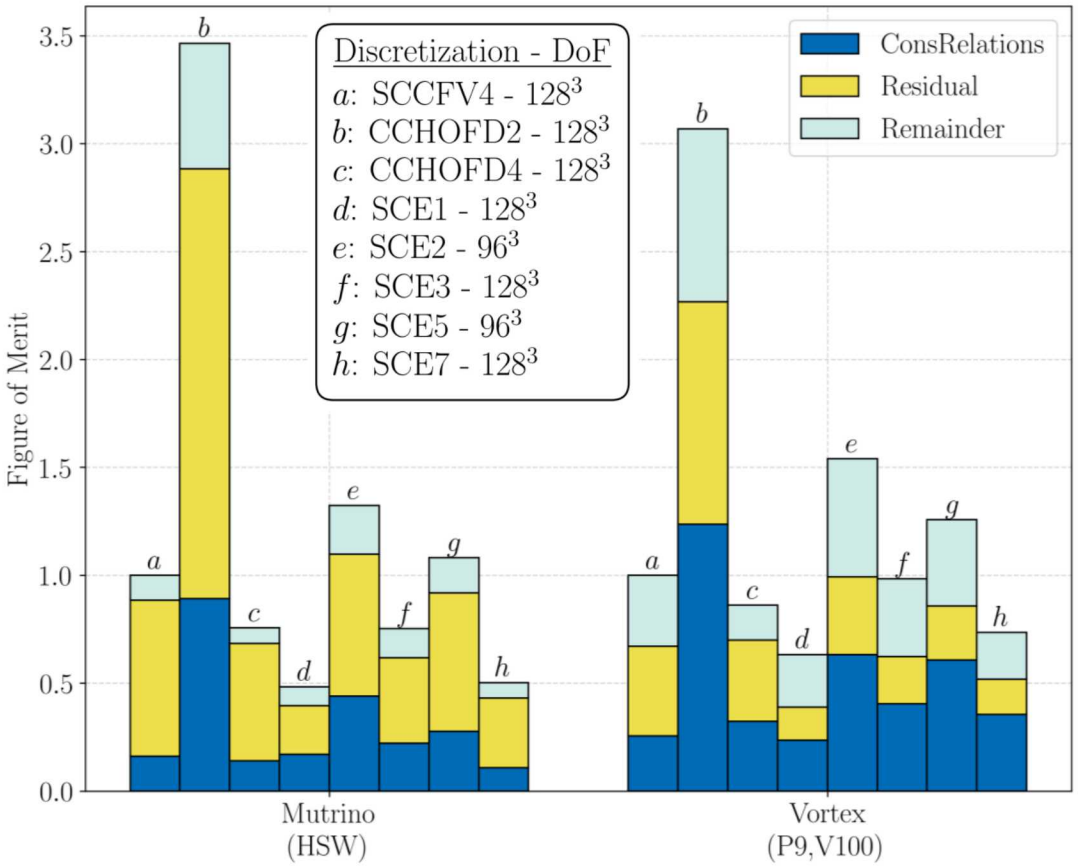
- 96^3, 885k DOFs
- 128^3, 2.1mil DOFs

Performance Analysis – TGV Figure of Merit

First attempt to quantify performance of high-order

$$Figure\ of\ Merit = \frac{t_{SCCFV} \cdot err^{\varepsilon}_{SCCFV}}{t \cdot err^{\varepsilon}}$$

- t - Wall-clock time for 30s simulation time (s)
- $err^{\varepsilon} = \int |\varepsilon(t) - \varepsilon_{ref}(t)|^2 dt$ - Enstrophy error
- Reference: Spectral element solution, 512^3
- Note: results are architecture independent



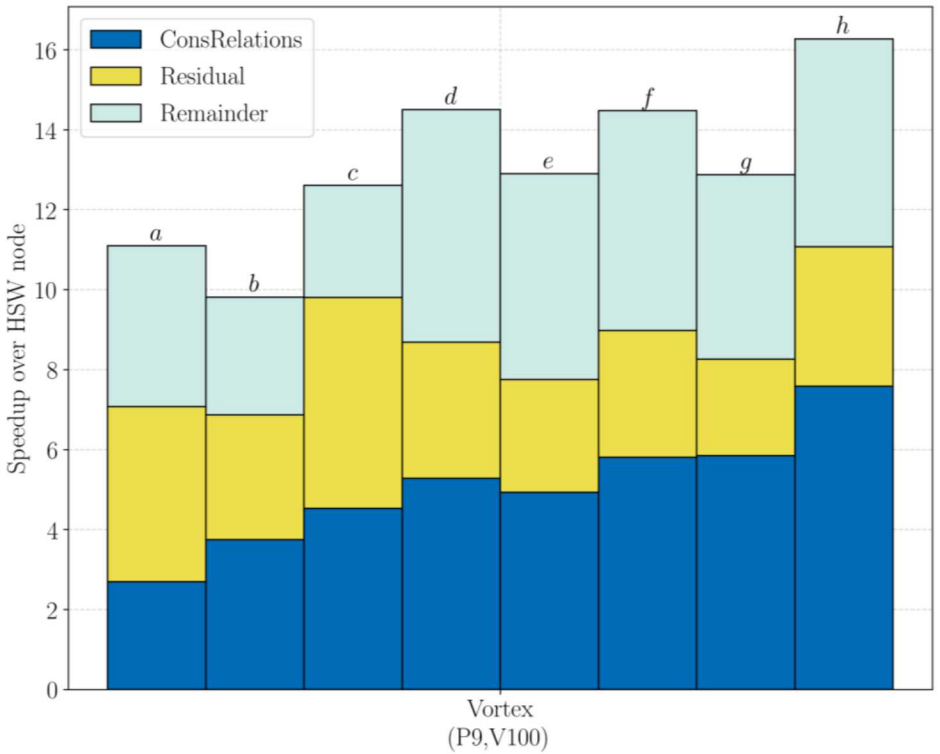
Analysis:

- Current optimal is near **CCHOFD2**, **SCE2**
- Better performance on GPUs
- Bottlenecks
  - CPU – Residual (computation)
  - GPU – ConsRelations (gradient, communication)
  - Remainder also needs more profiling/improvement

Performance Analysis – TGV Node Utilization

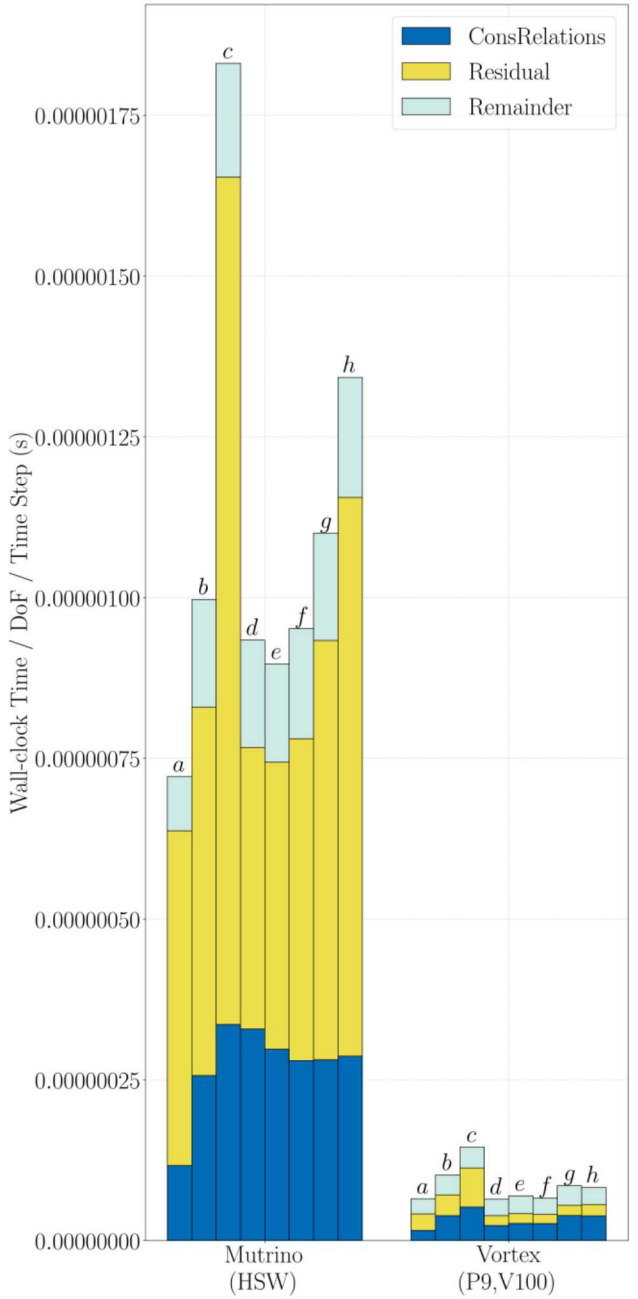
Analysis:

- ~10-16x GPU node speedup over HSW node
- Speedup is greater at high-order with more DoF
- **SCCFV** faster on CPU, lower order **SCE** matches on GPU



Discretization - DoF

a: SCCFV4 - 128^3
b: CCHOFD2 - 128^3
c: CCHOFD4 - 128^3
d: SCE1 - 128^3
e: SCE2 - 96^3
f: SCE3 - 128^3
g: SCE5 - 96^3
h: SCE7 - 128^3

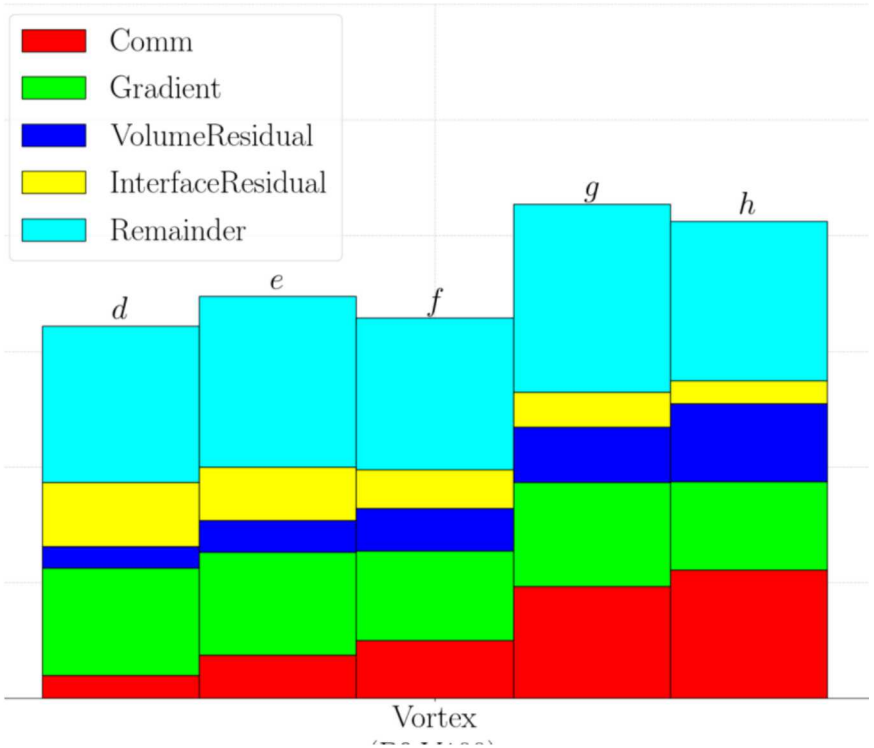


Performance Analysis – TGV Timer Breakdown

SCE Kernel Analysis:

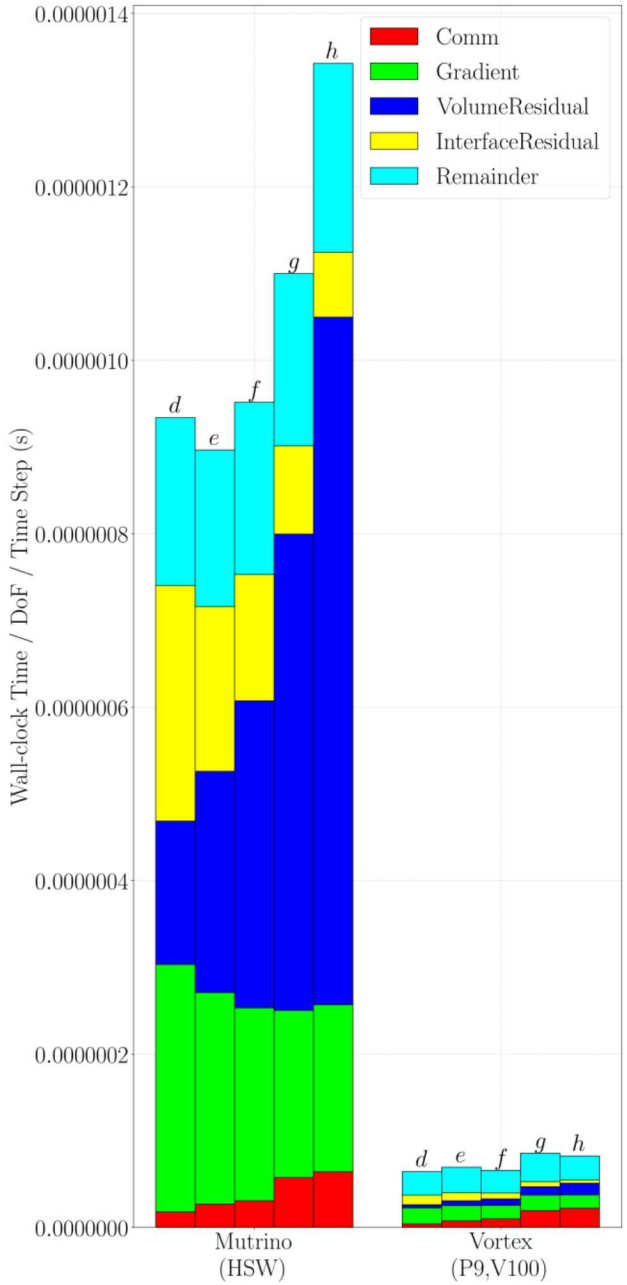
High-order trends:

- CPU – Volume increases, interface decreases
- GPU – similar, communication and remainder dominant



Discretization - DoF

a: SCCFV4 - 128^3
b: CCHOFD2 - 128^3
c: CCHOFD4 - 128^3
d: SCE1 - 128^3
e: SCE2 - 96^3
f: SCE3 - 128^3
g: SCE5 - 96^3
h: SCE7 - 128^3



Future Performance Work

Optimization for HOFD and SCE:

- More detailed profiling
- More work on hierarchical parallelism (memory layouts, shared memory)
- SIMD for CPUs
- Better communication pattern for SCE-LGL

Evaluate efficiency of high-order methods:

- More analysis/feedback on figure of merit
- Measure and compare communication cost and computational throughput
- Benchmark with open-source software

Performance analysis of more complex problems:

- Wall-modeled large eddy simulation
- High-speed boundary layer with chemical reactions