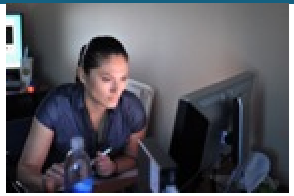
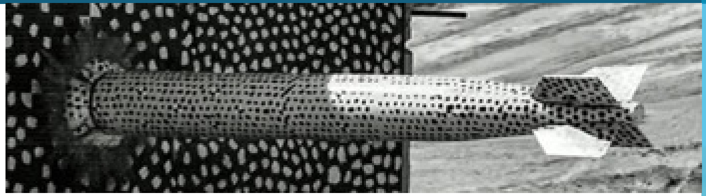
SAND2021-8923PE

Sandia National Laboratories

Tensors for Statistical Analyses of Scientific Data: A Turbulent Combustion Perspective





PRESENTED BY

Hemanth Kolla, 08753

U.S. DEPARTMENT OF ENERGY NNSA

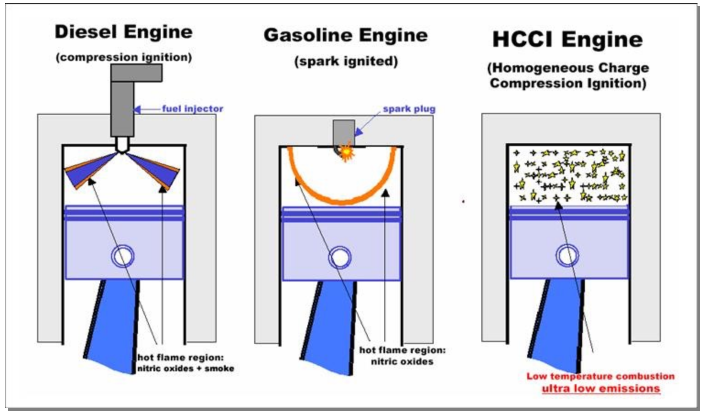
CSRI Summer Seminar Series, 27

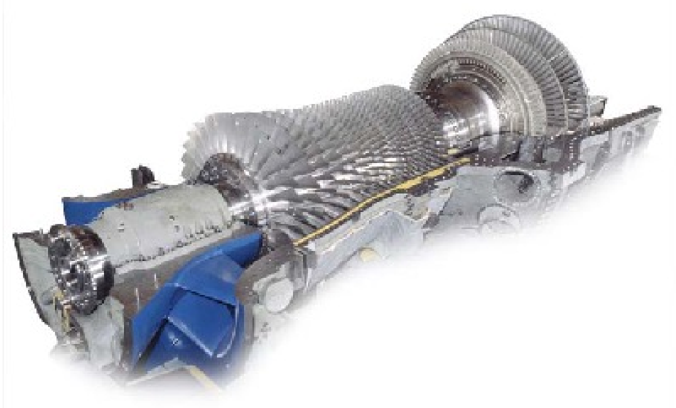
Outline

1. Turbulent flows with reactions.
   - Relevance.
   - Exascale computing.
   - Statistical analyses and learning.
2. Tensor decompositions.
   - Types & applications.
   - Use case 1: Data compression.
   - Use case 2: Rare event detection.

Many Applications Involve Turbulent Reacting Flows

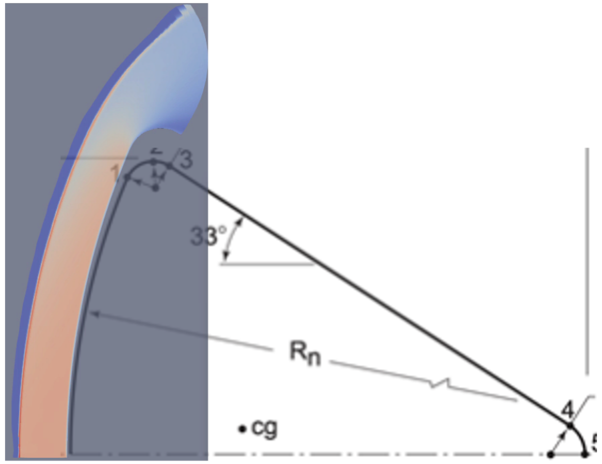
Energy & Transportation





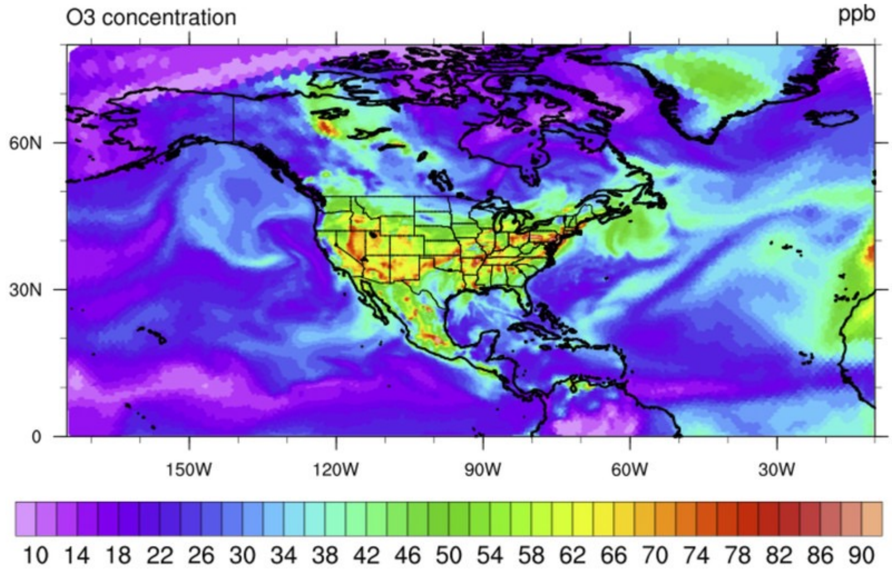
New generation systems: alternative fuels, nonconventional combustion regimes.

Aerospace



Hypersonic vehicles: thermal and structural integrity under extreme environments.

Earth Systems: Atmosphere



[Pfeister *et al.*, 2020, Bulletin of AMS, vol. 101](#)

Climate/Weather predictions: influence of natural and anthropogenic activity.

Understanding/Modelling interactions of turbulence and chemical reactions critical to design, operation, predictions.

Turbulence Cascade: A Foundational View

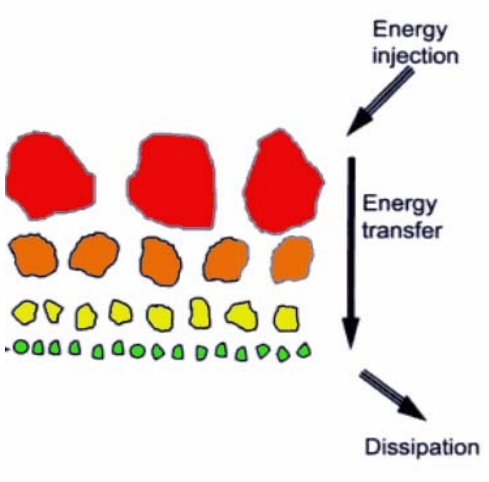
Big whorls have little whorls,

Which feed on their velocity;

And little whorls have lesser whorls,

And so on to viscosity.

- Lewis Richardson



Kolmogorov Theory for Large Reynolds Number (Re)

- Energy transfer is unidirectional: only from large to smaller scales.
- Only large scales "know" geometry. Small scales are agnostic.
- Small scale behaviour is statistically universal; governed only by energy dissipation rate (ε) and viscosity (ν).
- Intermediate scale (inertial range) statistics also universal; governed by ε.

Turbulence Cascade: A Foundational View

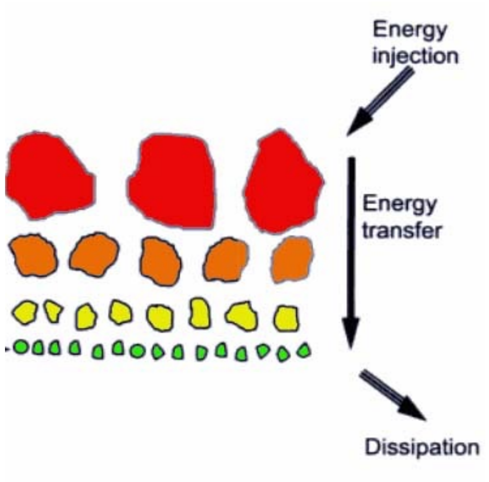
Big whorls have little whorls,

Which feed on their velocity;

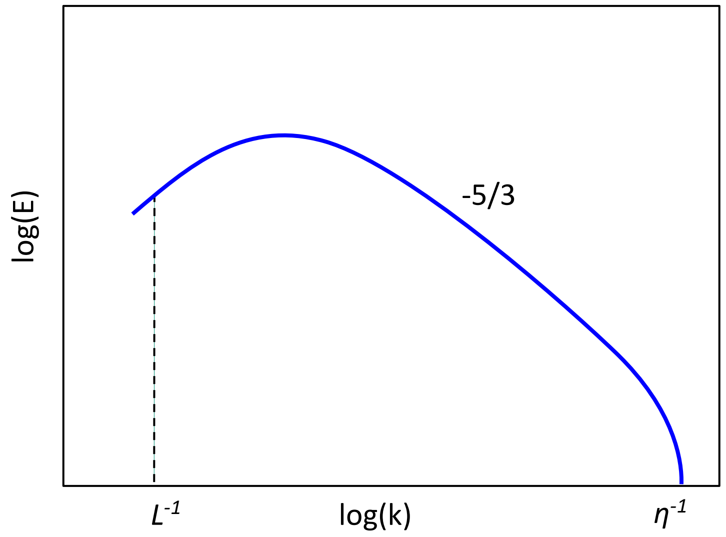
And little whorls have lesser whorls,

And so on to viscosity.

- Lewis Richardson



Kinetic Energy Spectrum



Kolmogorov Theory for Large Reynolds Number (Re)

- Energy transfer is unidirectional: only from large to smaller scales.
- Only large scales "know" geometry. Small scales are agnostic.
- Small scale behaviour is statistically universal; governed only by energy dissipation rate (ε) and viscosity (ν).
- Intermediate scale (inertial range) statistics also universal; governed by ε.

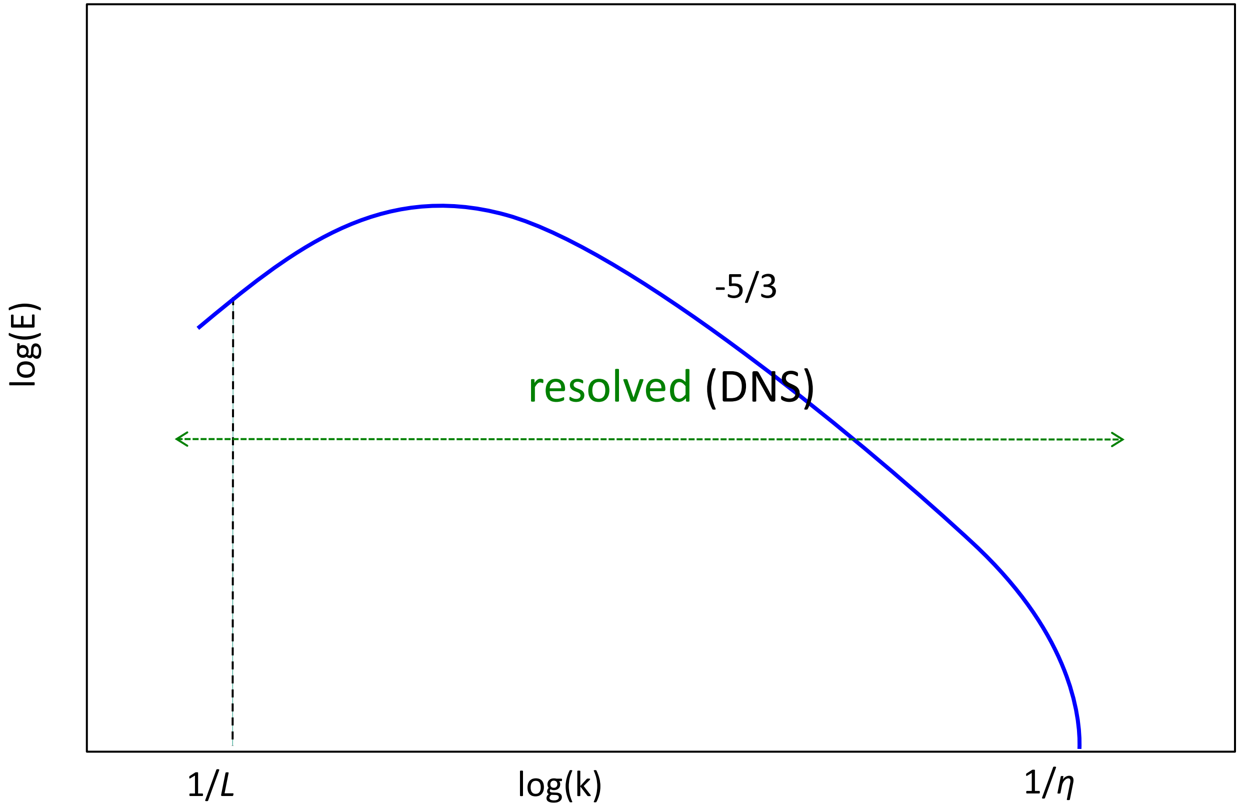
Length scales, $L/\eta \approx Re^{3/4}$

Time scales, $\tau_L/\tau_\eta \approx Re^{1/2}$

Simulating Turbulent Flows

$$\frac{\partial \boldsymbol{\phi}}{\partial t} = -(\vec{u} \cdot \nabla \boldsymbol{\phi}) - (\nabla \cdot \mathrm{j}_{\boldsymbol{\phi}})$$

Direct Numerical Simulations (DNS): Resolve all scales.



$L/\eta \approx \mathrm{Re}^{3/4}$

$\tau_L/\tau_\eta \approx \mathrm{Re}^{1/2}$

$\mathrm{cost} \approx \mathrm{Re}^{11/4}$

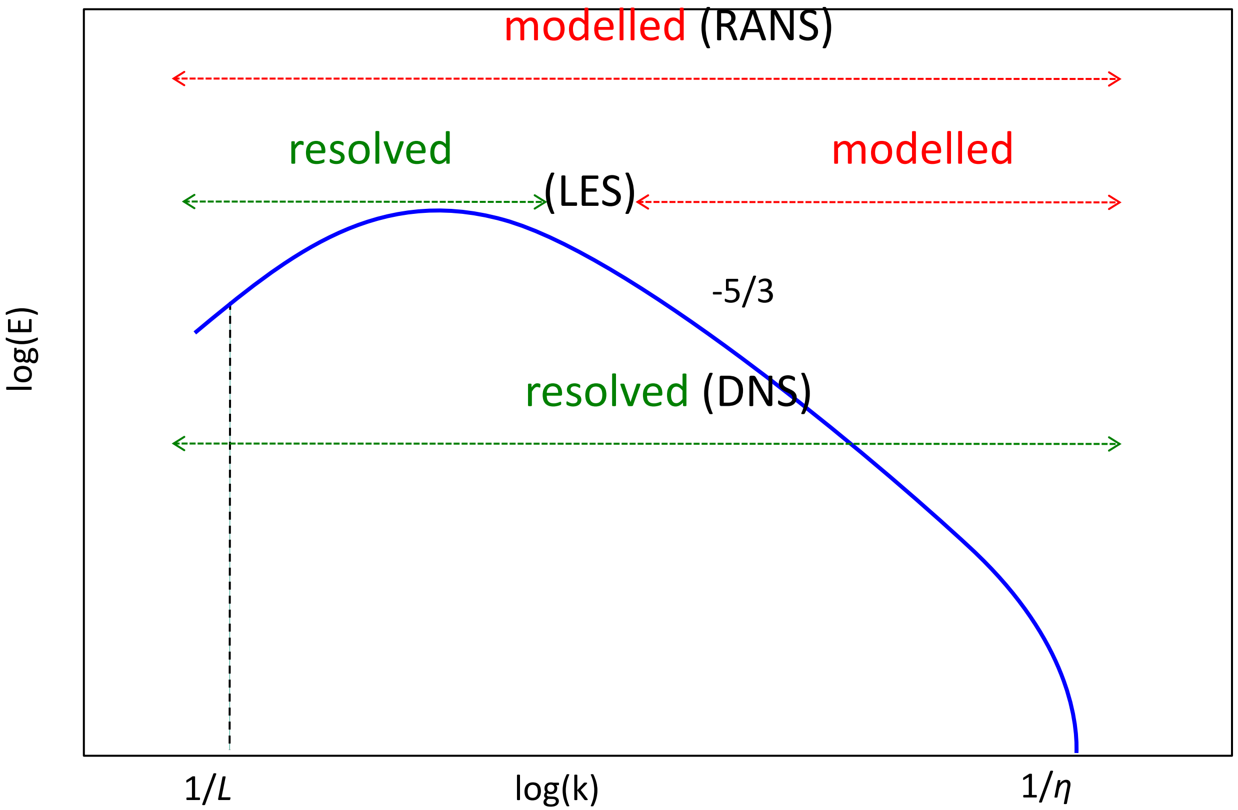
Simulating Turbulent Flows

$$\frac{\partial\langle\boldsymbol{\phi}\rangle}{\partial t} = -\left(\overrightarrow{\langle u\rangle}\cdot\nabla\langle\boldsymbol{\phi}\rangle\right) - \left(\nabla\cdot\langle \mathrm{j}_{\boldsymbol{\phi}}\rangle\right) + (\dots)$$

Direct Numerical Simulations (DNS): Resolve all scales.

Reynolds Averaged Navier Stokes (RANS): Model statistics of all scales (solve averaged form of NS equations).

Large Eddy Simulations (LES): Resolve 'energy containing' scales, model (statistics) of smaller scales.



$L/\eta \approx \mathrm{Re}^{3/4}$

$\tau_L/\tau_\eta \approx \mathrm{Re}^{1/2}$

$\mathrm{cost} \approx \mathrm{Re}^{11/4}$

Simulating Turbulent Flows with reactions

$$\frac{\partial \boldsymbol{\phi}}{\partial t} = -(\vec{u} \cdot \nabla \boldsymbol{\phi}) - (\nabla \cdot \mathrm{j}_{\boldsymbol{\phi}}) + \omega_{\boldsymbol{\phi}}$$

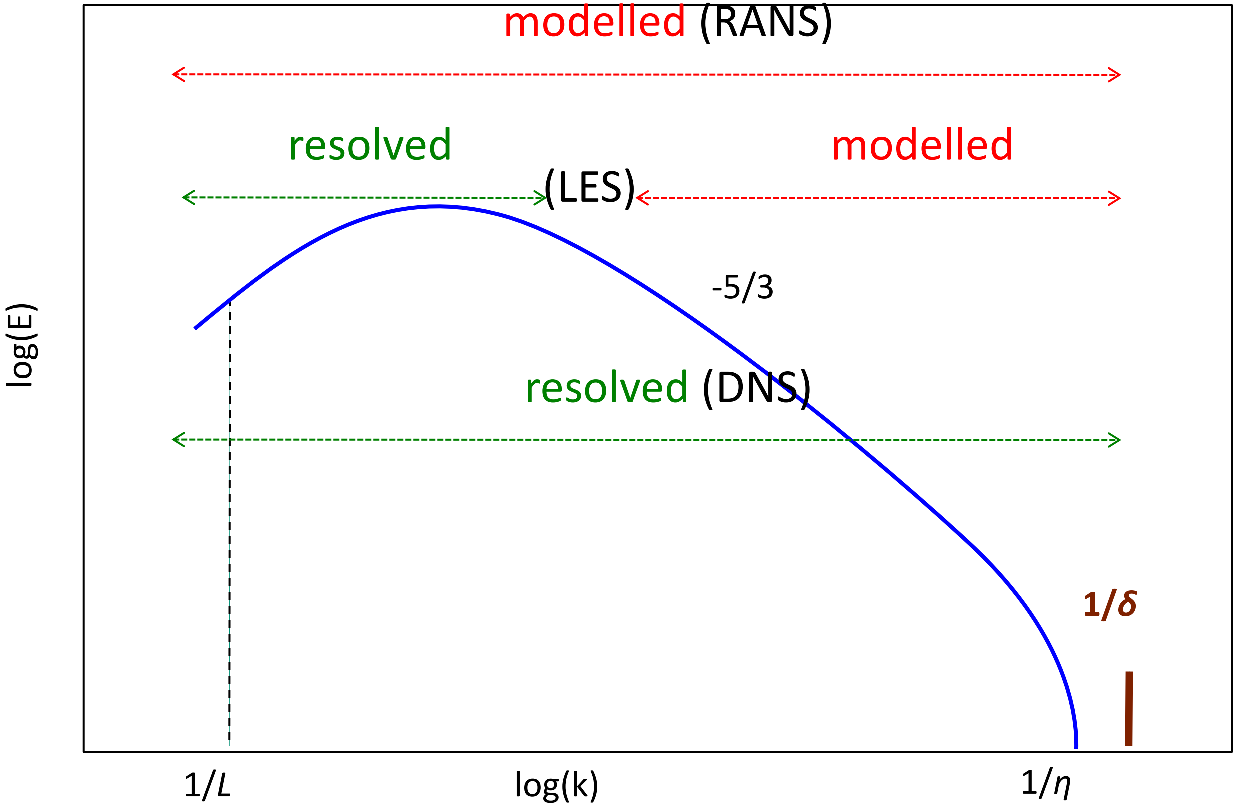
Direct Numerical Simulations (DNS): Resolve all scales.

Reynolds Averaged Navier Stokes (RANS): Model statistics of all scales (solve averaged form of NS equations).

Large Eddy Simulations (LES): Resolve 'energy containing' scales, model (statistics) of smaller scales.

Chemical reactions compound the difficulties:

- Introduce finer space-time scales.
- Additional physics to compute.
- Increase the PDE dimensionality; $\boldsymbol{\phi} \sim O(100)$.



$L/\eta \approx \mathrm{Re}^{3/4}$

$\tau_L/\tau_\eta \approx \mathrm{Re}^{1/2}$

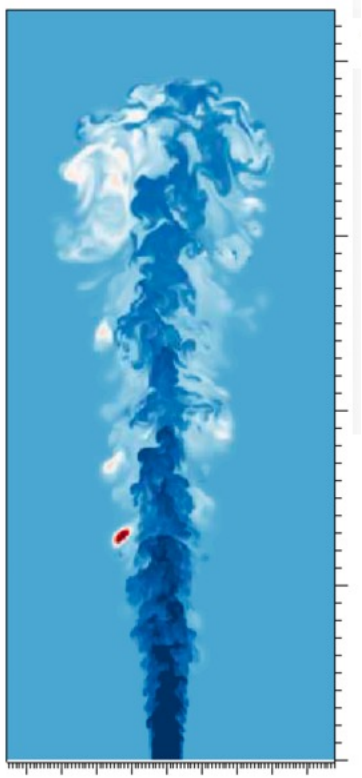
$\text{cost} \approx \mathrm{Re}^{11/4} \times (\eta/\delta)^3$

Turbulent Combustion DNS is an Exascale (+) Problem

- Conditions achieved at petascale: Re ≈ 10^4.
- Typical device-relevant conditions:
  - IC engines, Re ≈ 10^5.
  - Gas-turbine engines, Re ≈ 10^6.
- DNS cost scaling* ~ $Re^{11/4}$
- An order of magnitude increase in Re requires 560**x** more computing resources

Turbulent Combustion DNS is an Exascale (+) Problem

- Conditions achieved at petascale: Re ≈ 10^4.
- Typical device-relevant conditions:
  - IC engines, Re ≈ 10^5.
  - Gas-turbine engines, Re ≈ 10^6.
- DNS cost scaling* ~ $Re^{11/4}$
- An order of magnitude increase in Re requires 560**x** more computing resources



- Combustion-Pele part of ECP EXASCALE COMPUTING PROJECT
- PI, Jackie Chen (SNL, 08351). Co-PIs at LBL, NREL, ORNL, ANL, MIT.
- Advances on multiple fronts to attain a target problem: (1) performance portability to exascale architectures, (2) new load balancing and communication strategies, (3) communication avoiding linear solvers, (4) asynchronous execution, (5) in-situ analytics

Pele Exascale Target Problem

First-principles (DNS) and near-first principles (DNS/LES hybrids) simulations of the relevant processes in a low temperature reactivity controlled compression ignition (RCCI) internal combustion engine. The relevant processes include turbulence, mixing, spray vaporization, low-temperature ignition, flame propagation, and soot/radiation. As part of 10yr roadmap perform a hybrid LES/DNS simulation of a sector from a gas turbine for power generation burning hydrogen enriched natural gas.

Role of Statistical Analyses and Learning

Needs/Requirements

- Data management and organization.
- Surrogates, Reduced-order-models:
  - For unresolved scales.
  - For unresolved physics.
- Low-dimensional manifolds: identification, parametrization.
- Statistical inference (distributions, joint/conditional moments).
- Event/phenomena detection.
- Uncertainty quantification.

Role of Statistical Analyses and Learning

Needs/Requirements

- Data management and organization.
- Surrogates, Reduced-order-models:
  - For unresolved scales.
  - For unresolved physics.
- Low-dimensional manifolds: identification, parametrization.
- Statistical inference (distributions, joint/conditional moments).
- Event/phenomena detection.
- Uncertainty quantification.

Challenges

- Wide range of scales: observables span ~ 10 decades.
- Large dimensionality of state space (100s of features).
- Non-Gaussian multi-variate statistics; not always parametrizable.
- Robustness and Stability of reduced representations.
- Boundary conditions are integral part of the physics.

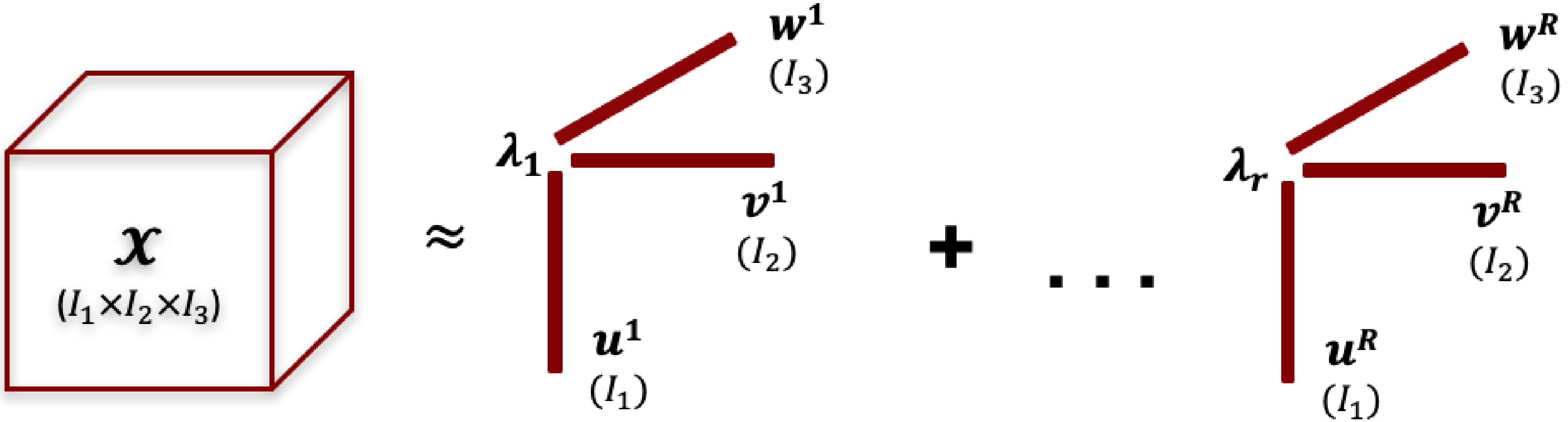
Outline

1. Turbulent flows with reactions.
   - Relevance.
   - Exascale computing.
   - Statistical analyses and learning.
2. Tensor decompositions.
   - Types & applications.
   - Use case 1: Data compression.
   - Use case 2: Rare event detection.

Tensors Are Versatile For High-Dimensional Learning

- Tensors are multidimensional (multiway) arrays; higher-order generalizations of vectors/matrices.
  - Order: Number of dimensions (e.g. a matrix is order-2 tensor).
  - Mode: A specific dimension (e.g. in a matrix rows are mode-1, columns are mode-2).
- Tensor decompositions – linear algebra in dimensions > 2.
  - Many formats in literature (Kolda & Bader, 2009, "Tensor Decompositions and Applications", SIAM Review, vol.51).
  - Not all concepts of matrix linear algebra generalize (e.g. rank, existence, uniqueness).
- Versatile for learning in high-dimensional settings:
  - Pattern identification.
  - Parameter importance.
  - Surrogates/Reduced representations…..
- Long history (~20yrs) of foundational math research at Sandia (lead by Tammy Kolda, others).
- Successful in applying to national security domains; recently extended for scientific data and HPC.

Canonical Polyadic (CP) Decomposition



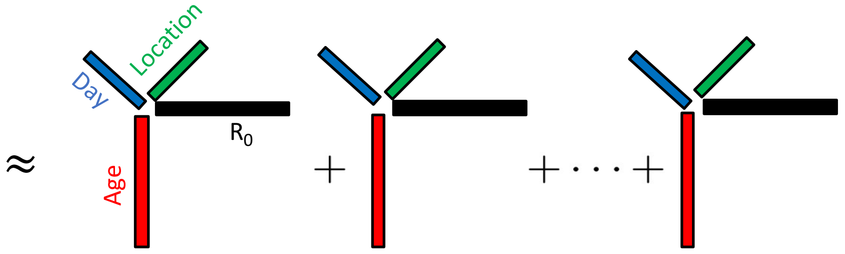
$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{u}_r \odot \mathbf{v}_r \odot \mathbf{w}_r \qquad\qquad x_{ijk} \approx \sum_{r=1}^{R} \lambda_r \, u_{ir} v_{jr} w_{kr}$$

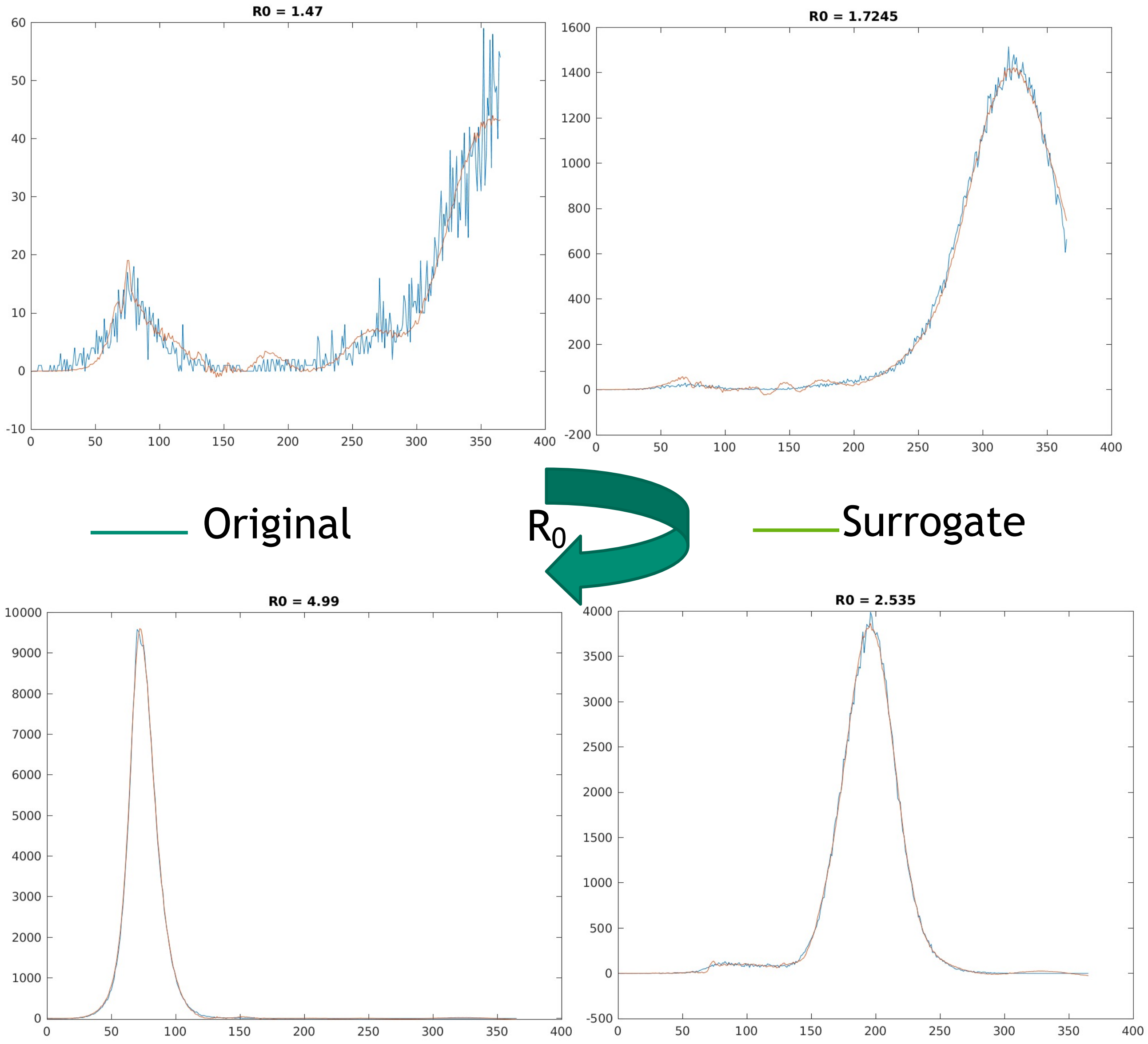
- Tensor approximated as "sum of R outer product of vectors".
- Compact, interpretable, computationally inexpensive. Complexity ~ $O(Rd)$.
- Recently generalized to different underlying distributions, loss functions (Hong, Kolda, Duersch, 2020, SIAM Rev).

Applications of CP Decomposition

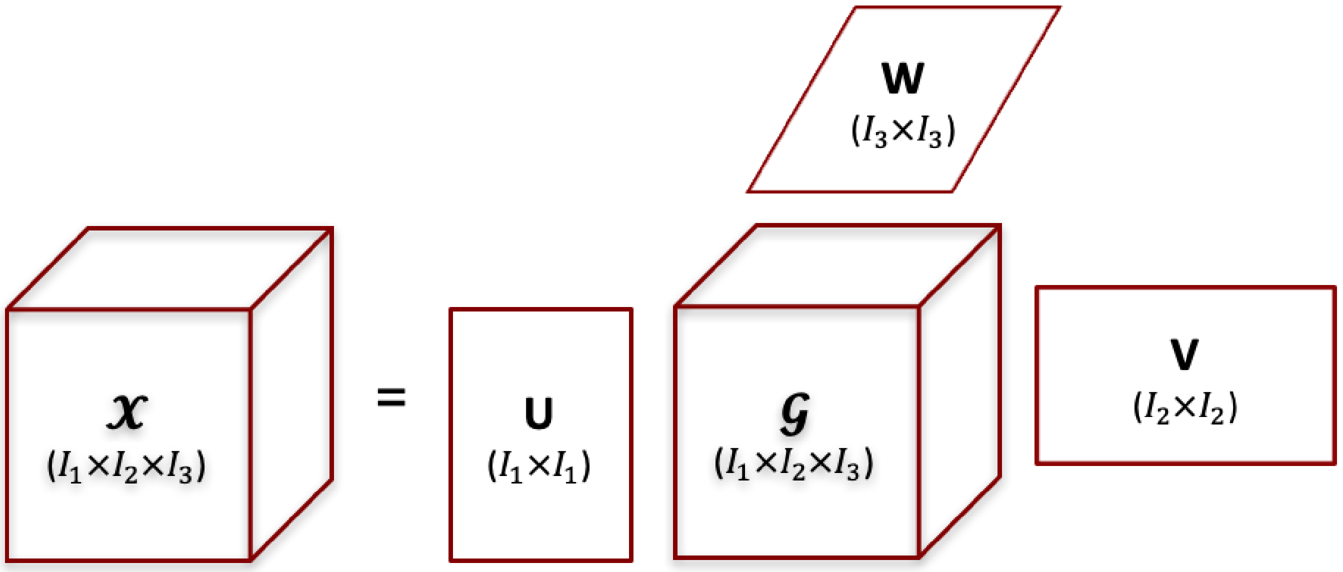
- SOLSTICE: DoD Advanced Computing Initiative.
  - GCP for decompositions of geographically distributed sensor data (algorithms and HPC software).
  - Sandia: Eric Phipps(PI), Drew Lewis, Rich Field, Richard Barrett, Kyle Gilman, Tammy Kolda (former PI). Georgia Tech: Rich Vuduc, Koby Hayashi, Chunxing Yin.
- ExaLearn: Co-design Center for Machine Learning Technologies.
  - Tensor decompositions for scientific data (algorithms and HPC software).
  - Sandia: Michael Wolf (PI), Eric Phipps, Hemanth Kolla, Ben Cobb (intern, GaTech), Zitong Li (intern, Wake Forest).
- Example: CP surrogate of epidemiological data

infections(age, location, day, R_0) $\approx$





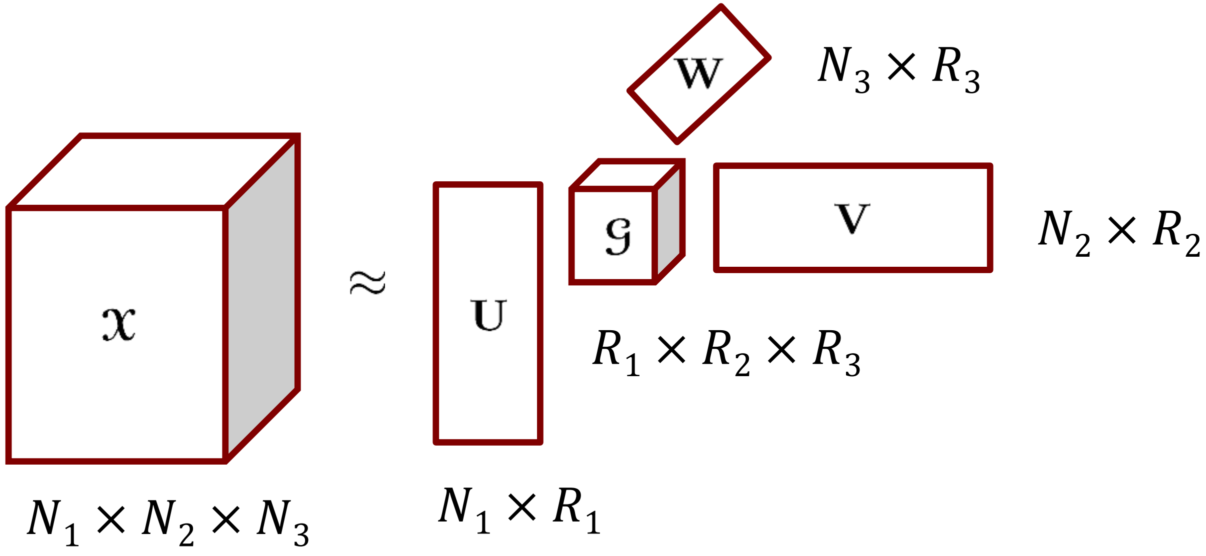
Higher-order Singular Value Decomposition (HOSVD)



$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}$$

- Multilinear generalization of matrix SVD (De Lathauwer, De Moor, Vandewalle, 2000, SIAM J. Matrix Anal. Appl., 21(4)).
- $\mathbf{U}, \mathbf{V}, \mathbf{W}$: ortho-normal bases of corresponding mode-spaces (left singular vectors of matricized tensor).
- Computed as matrix SVD by unfolding along each mode.

Tucker Decomposition: Truncated HOSVD



$$\mathcal{X} \approx \mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}$$

$$\|\mathcal{X} - (\mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W})\| \leq \epsilon \|\mathcal{X}\|$$

$$\|\mathcal{X}\|^2 - \|\mathcal{G}\|^2 \leq \epsilon^2 \|\mathcal{X}\|^2$$

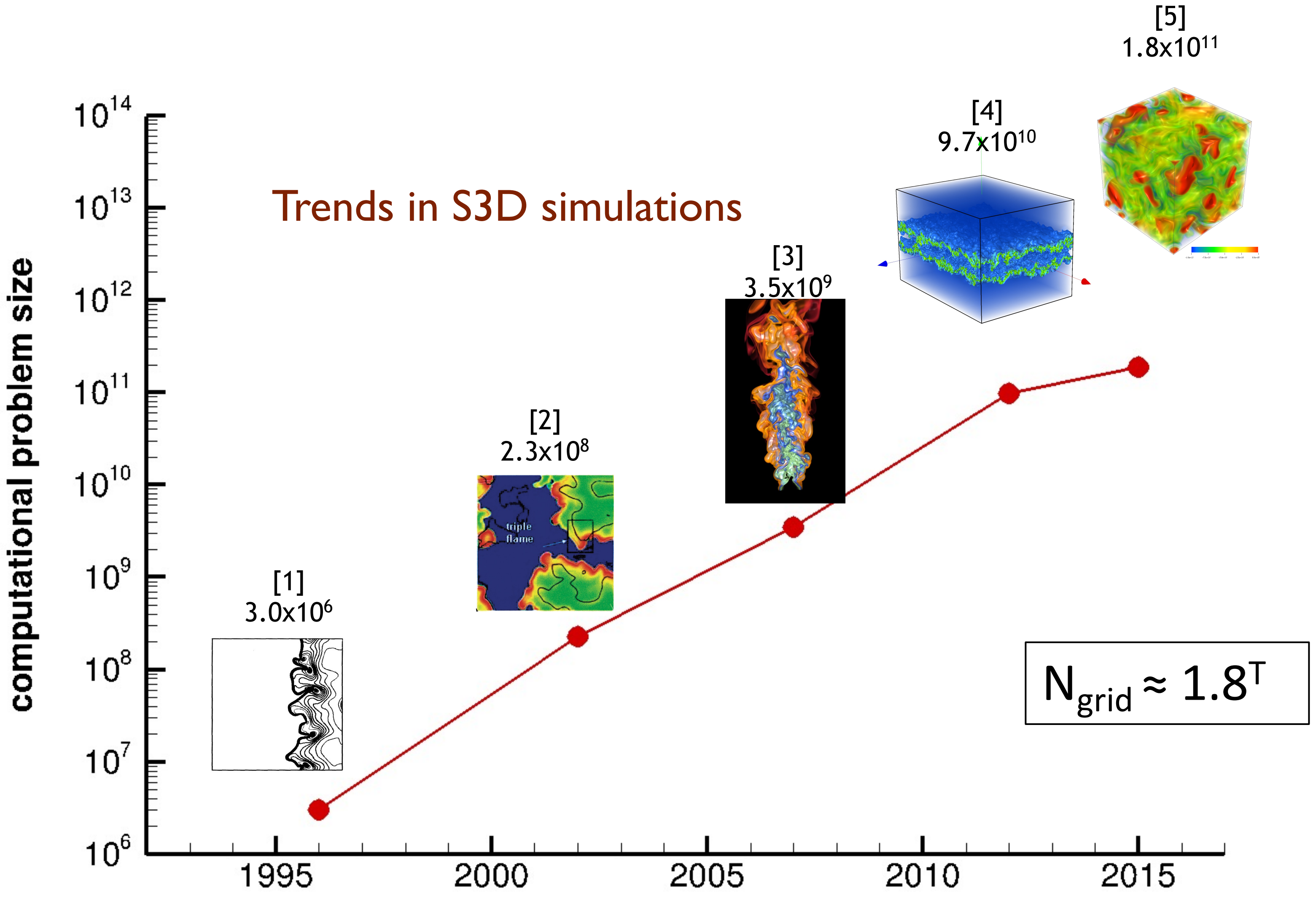
$$x(i_1, i_2, i_3) \approx \sum_{j_1, j_2, j_3} g(j_1, j_2, j_3)\, u(i_1, j_1)\, v(i_2, j_2)\, w(i_3, j_3)$$

- Approximation that exploits low-rank structure along each mode.
- **U**, **V**, **W**: orthogonal matrices spanning high variance subspaces (leading left singular vectors).
- Model complexity ~ $O(R^d)$. But still provides large compression ~ $O((N/R)^d)$.

Outline

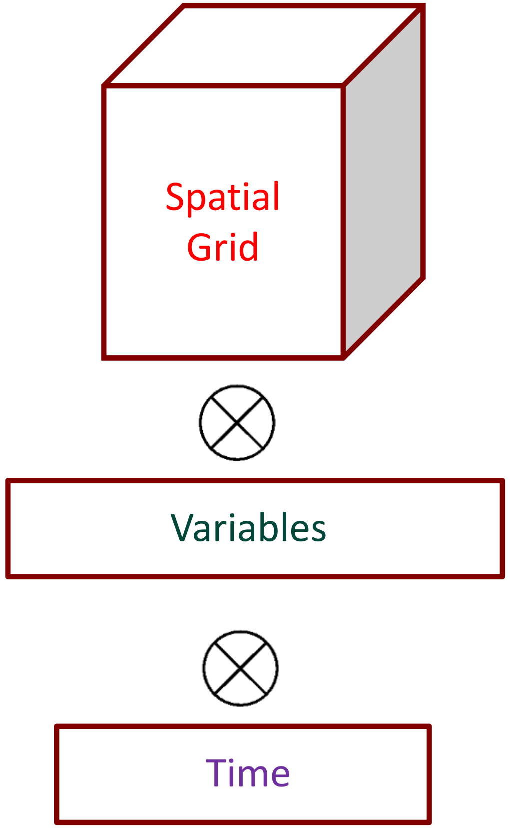
1. Turbulent flows with reactions.
   - Relevance.
   - Exascale computing.
   - Statistical analyses and learning.
2. Tensor decompositions.
   - Types & applications.
   - Use case 1: Data compression.
   - Use case 2: Rare event detection.

Scientific Data Volumes are Untenable



[1] T. Echekki, J.H. Chen, *Comb. Flame*, 1996, vol.106.
[2] T. Echekki, J.H, Chen, *Proc. Comb. Inst.*, 2002, vol. 29.
[3] R. Sankaran, E.R. Hawkes, J.H. Chen, *Proc. Comb. Inst.*, 2007, vol. 31.
[4] E.R. Hawkes, O. Chatakonda, H. Kolla, A.R. Kerstein, J.H. Chen, *Comb. Flame*, 2012 (online).
[5] Gordon Bell submission, 2015

Tucker Compression of Combustion Data



	Original	$\epsilon = 10^{-4}$	$\epsilon = 10^{-2}$
HCCI	672 x 672 x 32 x 626	330 x 310 x 31 x 199 **(14 X)**	111 x 105 x 22 x 46 **(760 X)**
SP	500 x 500 x 500 x 11 x 400	95 x 129 x 125 x 7 x 125 **(410 X)**	30 x 38 x 35 x 6 x 11 **(20000 X)**
JICF	1500 x 2080 x 1500 x 18 x 10	424 x 387 x 261 x 18 x 10 **(110 X)**	90 x 61 x 48 x 13 x 6 **(40000 X)**

TuckerMPI: Scalable Parallel Implementation

- C++/MPI library for distributed data compression (Ballard, Klinvex, Kolda, 2020, ACM Trans. Math. Soft., 46(2))

- Optimizing data layouts and communication for key kernels: modewise unfolding (matricization), Gram matrix computation, Eigensolve, Tensor×Matrix

- Open repository https://gitlab.com/tensors/TuckerMPI.

- Two algorithm variants:
  - Sequentially Truncated HOSVD: Matricize→SVD→Truncate/Reduce (backup slide)
  - Higher Order Orthogonal Iteration (HOOI): Iterate with ST-HOSVD as initial guess, until convergence (Austin, Ballard, Kolda, IPDPS, 2016).

- Partial reconstruction of tensor subset. Effective for data dissemination.

TuckerMPI timing profiles

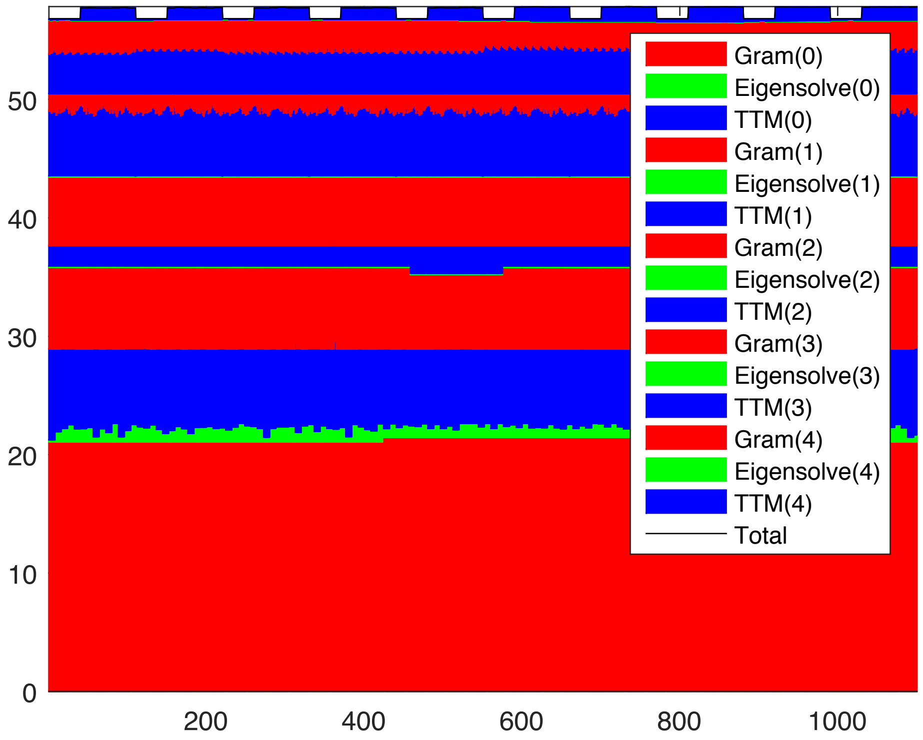
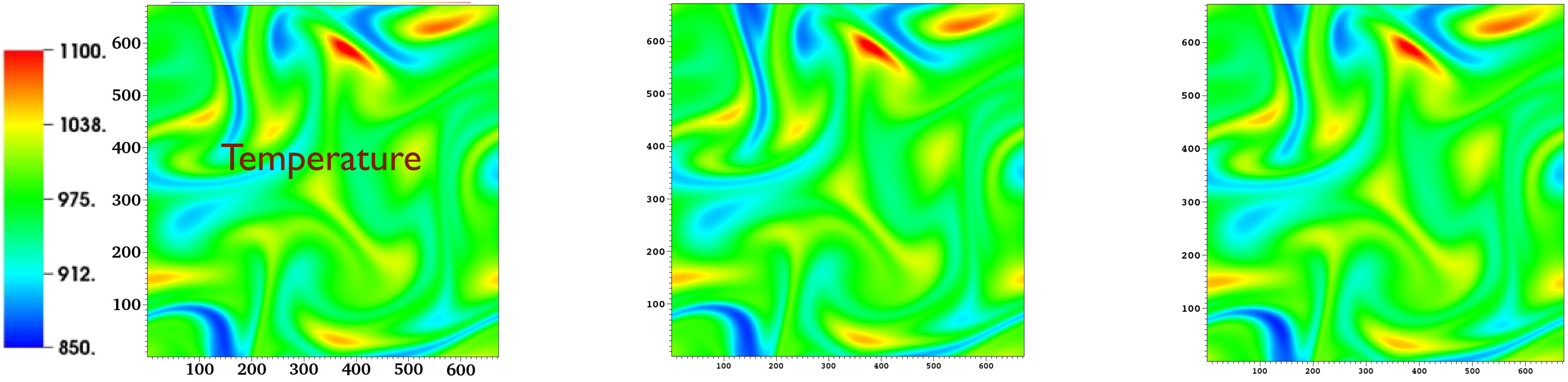


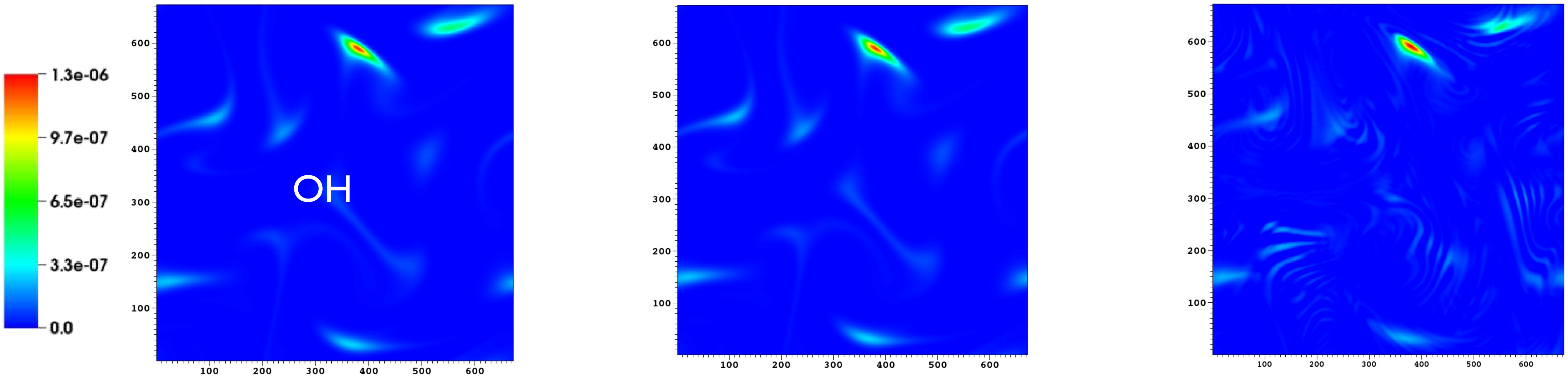
Image Courtesy: Alicia Klinvex

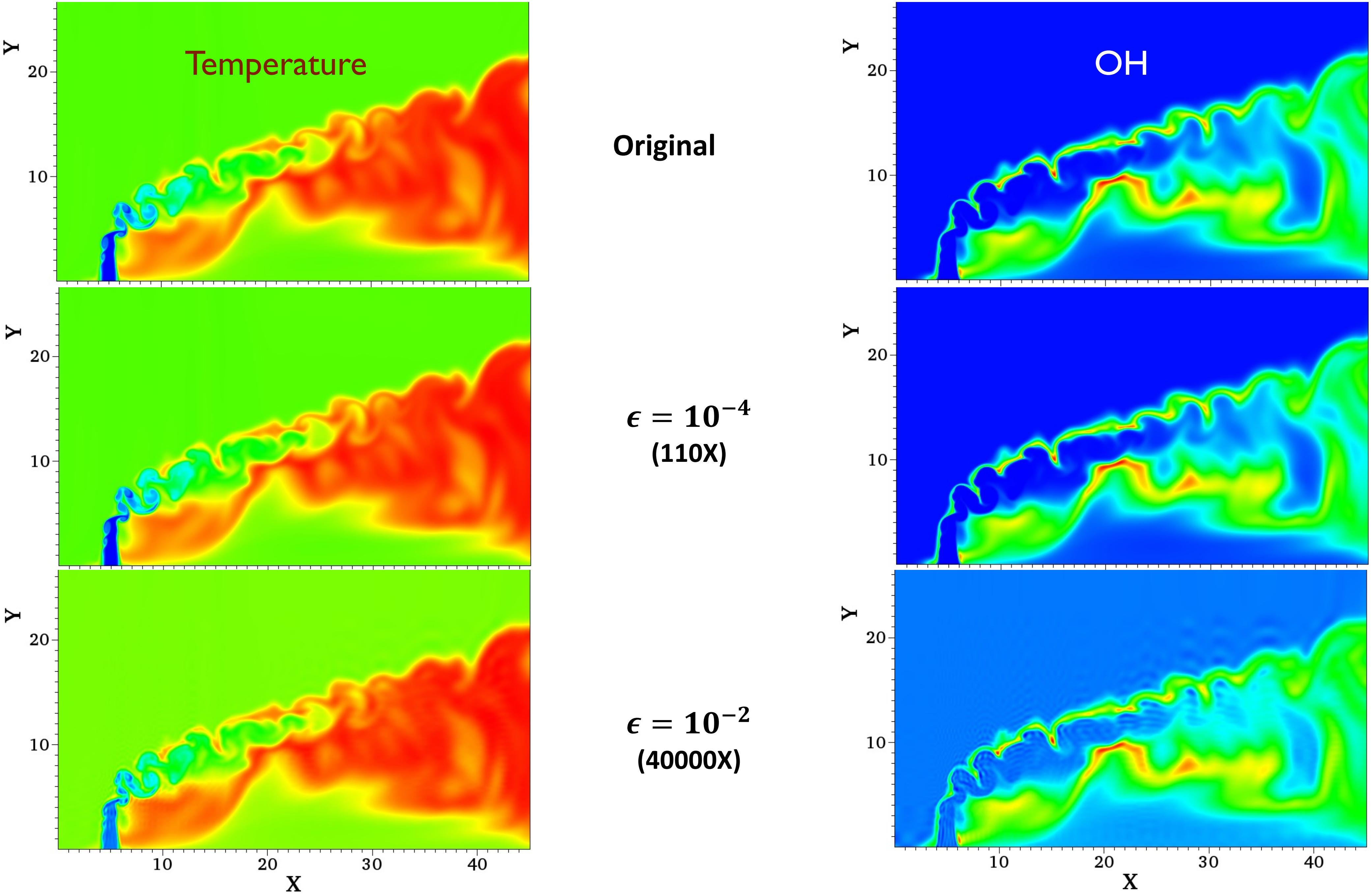
- HCCI data set, 4.4TB→10GB (410X).
- 1100 processors.
- Total time of 55s.

Tucker Compression: Error Distribution



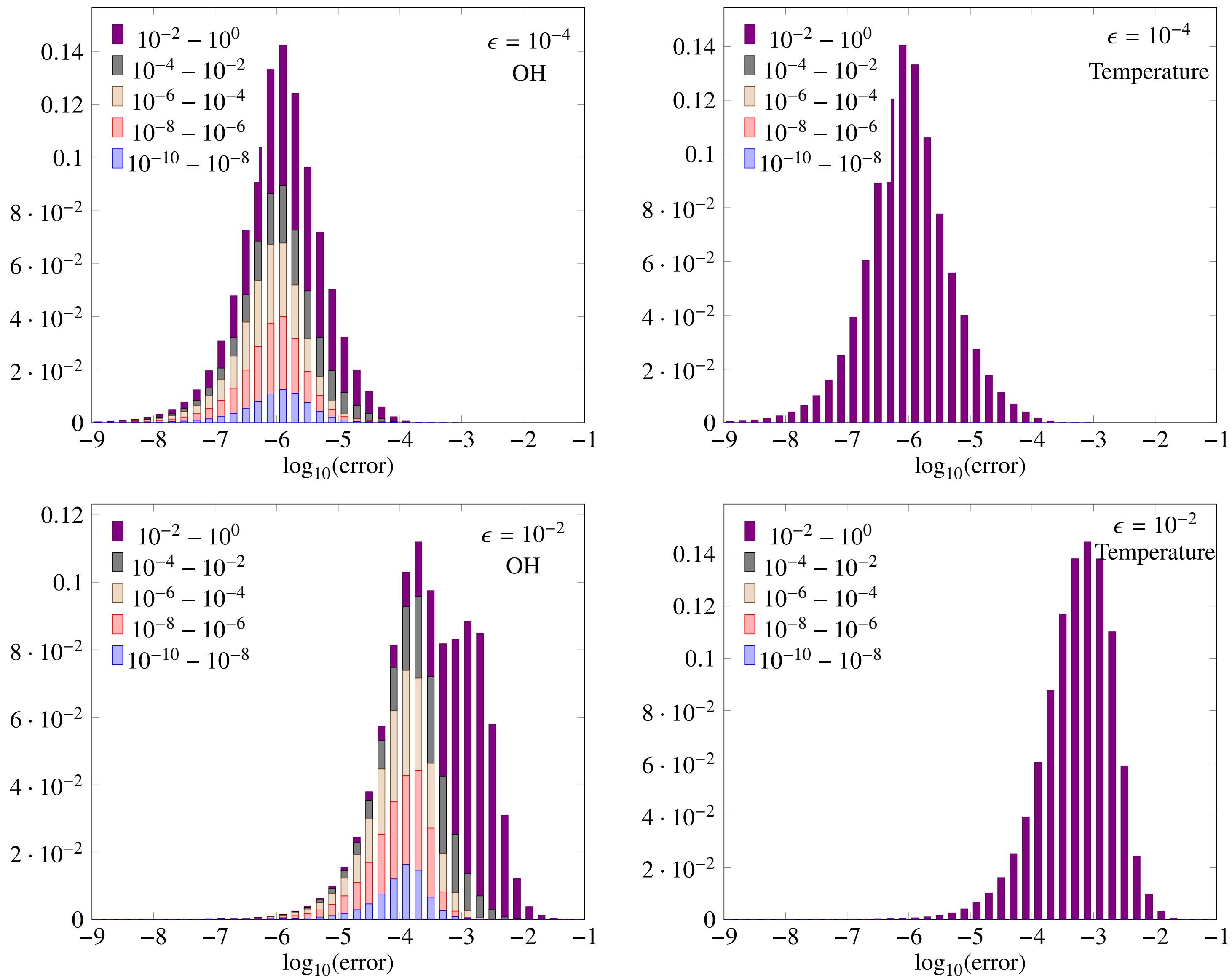
Original | $\epsilon = 10^{-4}$ **(14X)** | $\epsilon = 10^{-2}$ **(760X)**




Temperature
OH
Original
$\epsilon = 10^{-4}$
(110X)
$\epsilon = 10^{-2}$
(40000X)
Y
X
10
20
30
40

Tucker Compression: Error Distribution

- Elementwise error guarantees are difficult.
- Elements with small absolute values have large relative errors.
- “Minor” variables are bound to be more erroneous.



Outline

1. Turbulent flows with reactions.
   - Relevance.
   - Exascale computing.
   - Statistical analyses and learning.
2. Tensor decompositions.
   - Types & applications.
   - Use case 1: Data compression.
   - Use case 2: Rare/Anomalous event detection.

Problem Statement & Proposed Solution

Robust in-situ detection of rare events in distributed scientific simulations

Challenge

- *Scientific data*: continuous smoothly varying multi-variate non-Gaussian data.
- *Rare events*: group of physically valid extreme valued samples; hard to specify universal thresholds.
- *In-situ, distributed*: computational expense, scalability are important.

Key Idea

- Information of rare events is present in higher order joint moments, e.g. co-kurtosis.
- Identify rare events based on a "distinct" signature of joint moments.

Identification currently based on ad-hoc thresholds

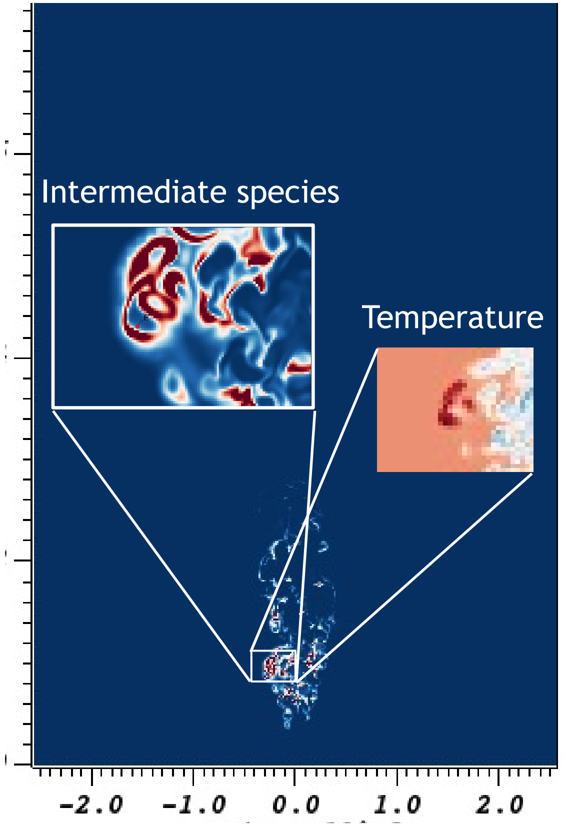
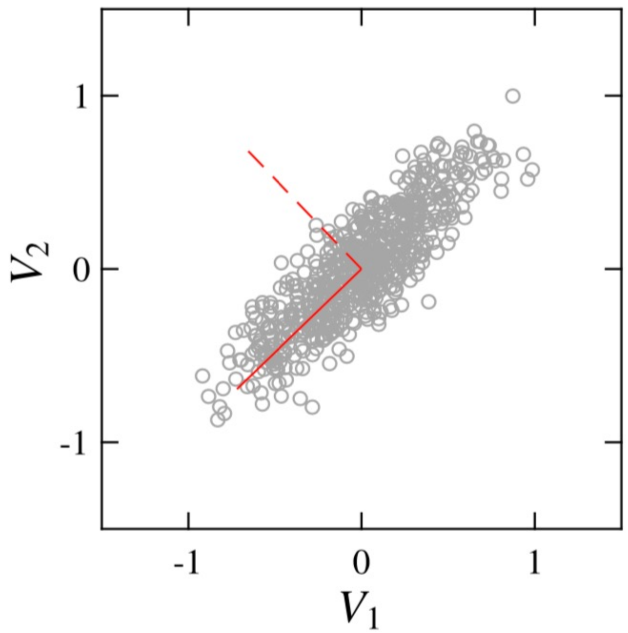


Image Courtesy: Martin Rieth, Marco Arienti, Matt Larsen

Information In Higher-Order Statistical Moments

For non-Gaussian multi-variate statistical processes higher-order joint moments are informative

(co-skewness is 3rd-order tensor, co-kurtosis is 4th-order tensor)

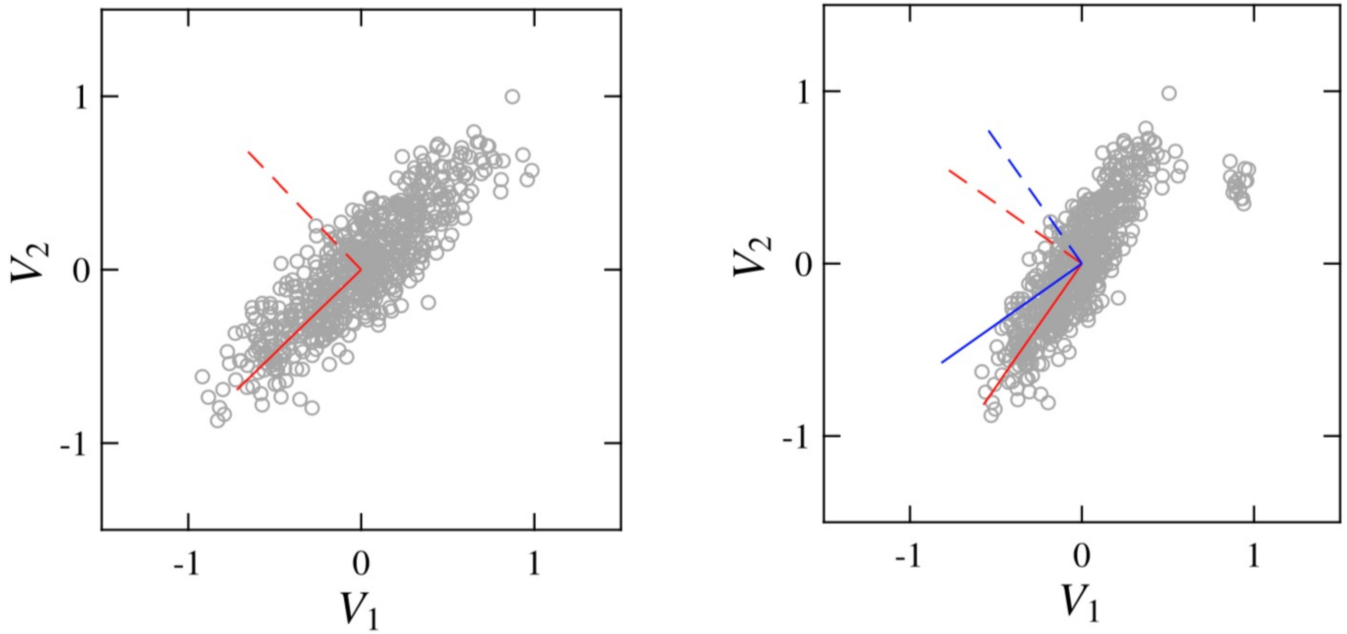


Red: Eigenvectors of Covariance (Principal Component Analysis). Denote directions of maximal variance

Information In Higher-Order Statistical Moments

For non-Gaussian multi-variate statistical processes higher-order joint moments are informative

(co-skewness is 3rd-order tensor, co-kurtosis is 4th-order tensor)

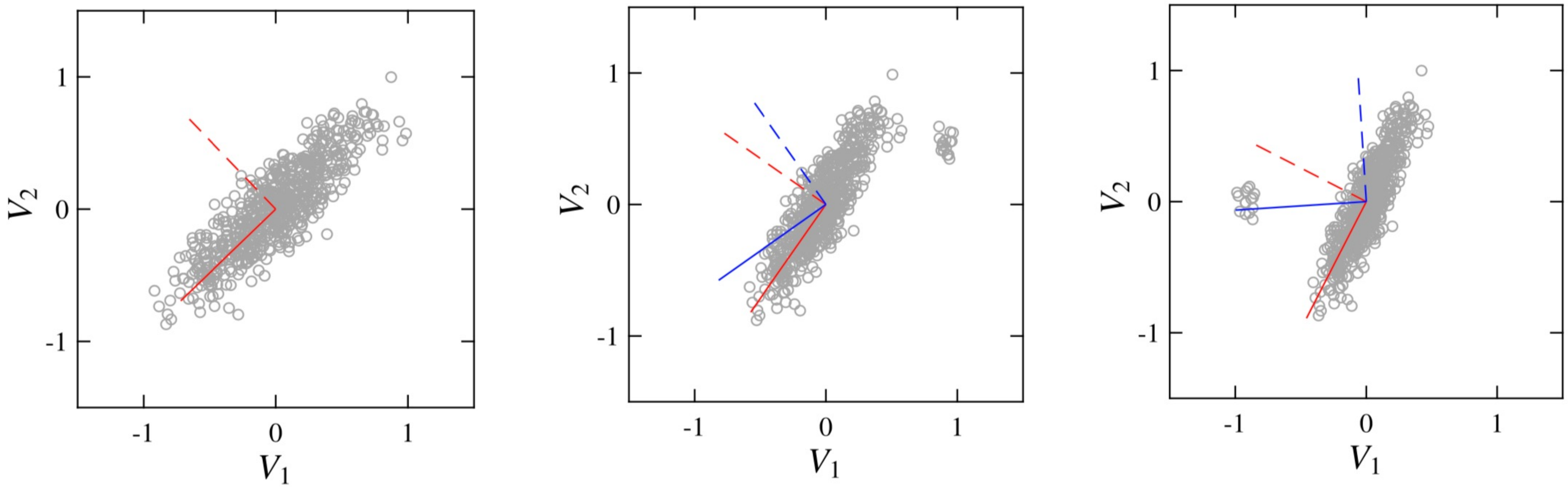


Red: Eigenvectors of Covariance (Principal Component Analysis). Denote directions of maximal variance

Blue: 'Principal Kurtosis Vectors'. Obtained through HOSVD of co-kurtosis tensor.

Information In Higher-Order Statistical Moments

PCA vectors not sensitive to outliers, Principal Kurtosis Vectors are.



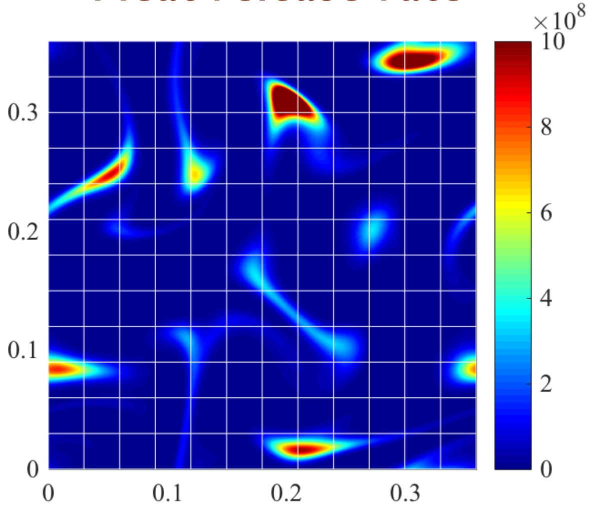
Red: Eigenvectors of Covariance (Principal Component Analysis). Denote directions of maximal variance

Blue: 'Principal Kurtosis Vectors'. Obtained through HOSVD of co-kurtosis tensor. (backup: connection to ICA)

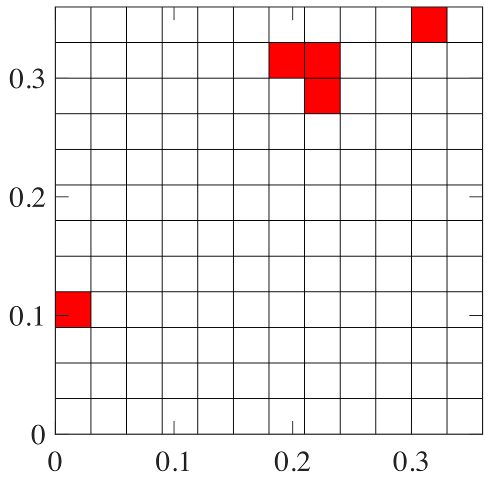
Formalizing Distributed Rare Event Detection

- Compute Principal Kurtosis Vectors on each data partition (e.g. processor).
- Compare the vectors amongst partitions in space and/or time:
  - Proposed Feature moment metrics (fraction of the kurtosis attributable to each variable) to quantify orientation of Kurtosis vectors.
  - FMMs sum to unity, akin to discrete distribution.
  - Divergence metric (Hellinger distance) to compare across partitions.
- Most computation (cokurtosis tensor and principal vectors) is local.
- Communication only of a small vector of numbers (FMMs).

Heat release rate



Anomalous partitions

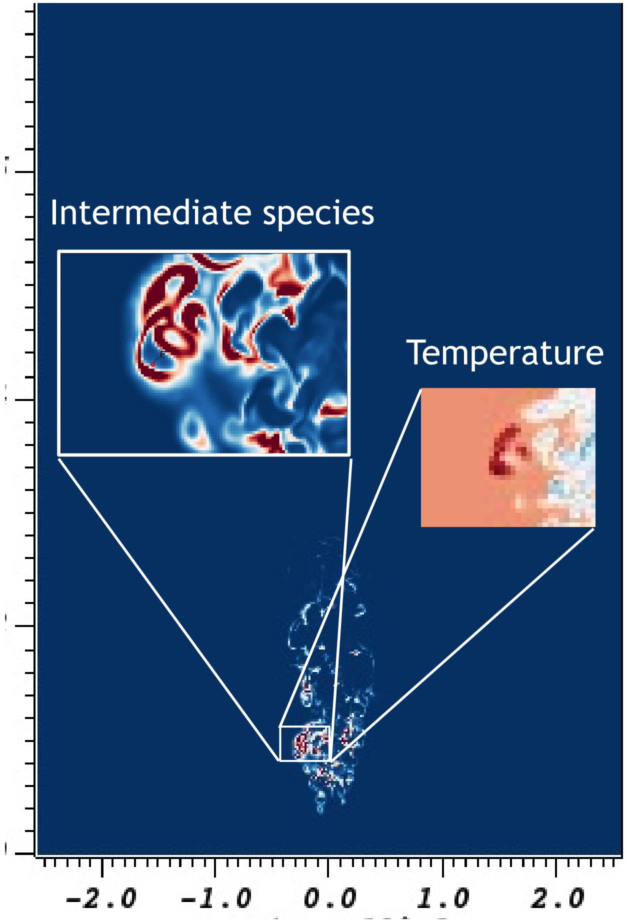


"*Anomaly detection in scientific data using joint statistical moments*." K. Aditya, H. Kolla, W.P. Kegelmeyer, T.M. Shead, J. Ling, W.L. Davis IV, Journal of Computational Physics, 2019.

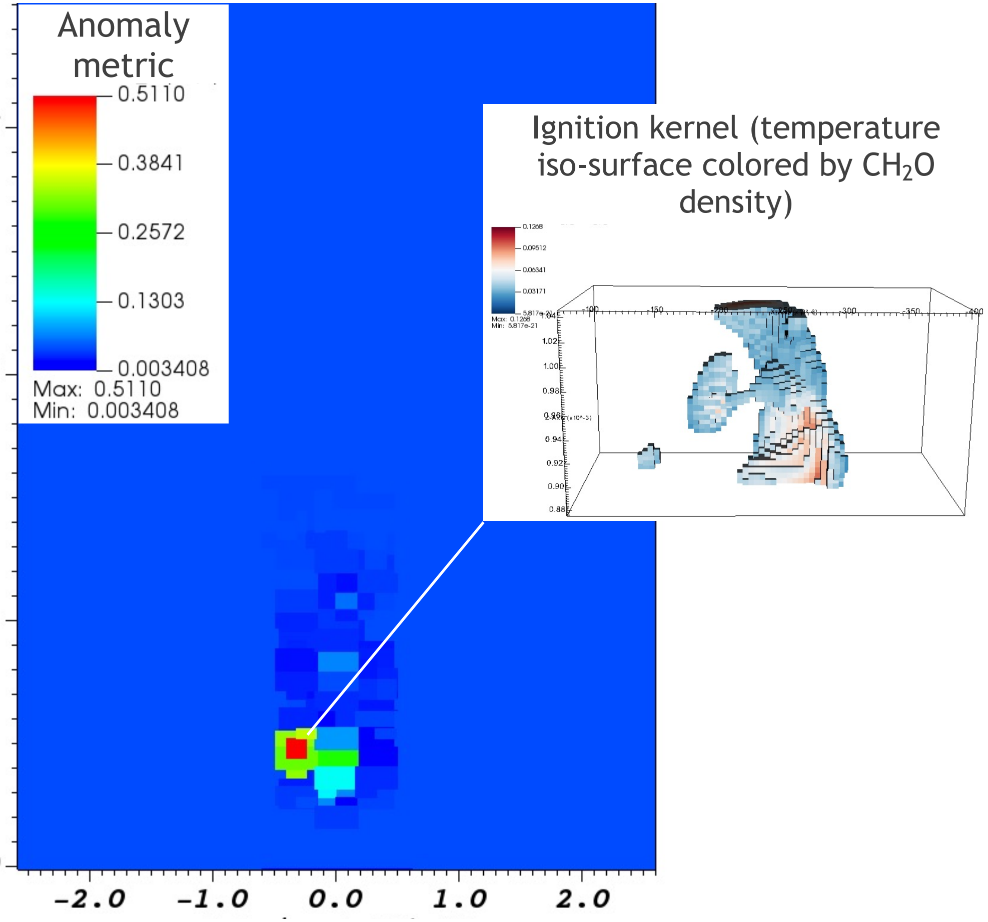
In-situ Deployment as ECP Cross-Cut Effort

- ExaLearn: GenTen, Software for generalized Tensor Decompositions.
- ALPINE: Ascent, flyweight in situ visualization and analysis infrastructure.
- Pele: PeleLM, adaptive-mesh low Mach number hydrodynamics code for reacting flows

Identification currently based on ad-hoc thresholds



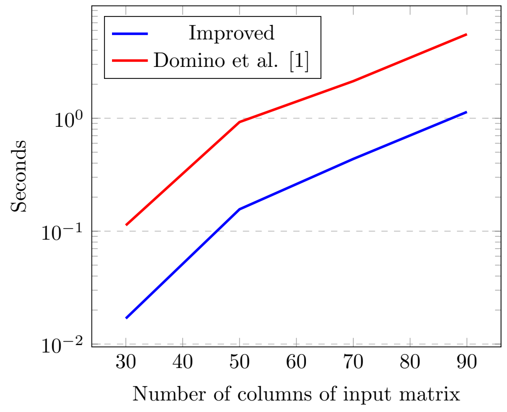
Validation: co-kurtosis tensor-based unsupervised anomaly detection

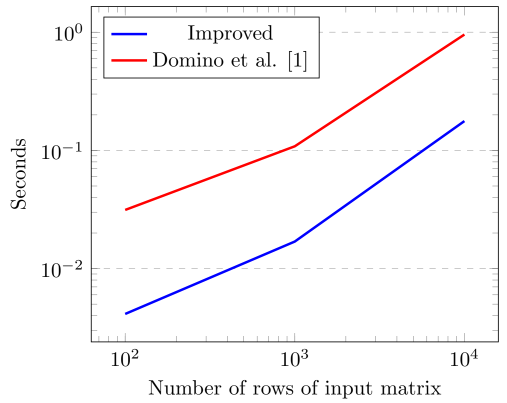


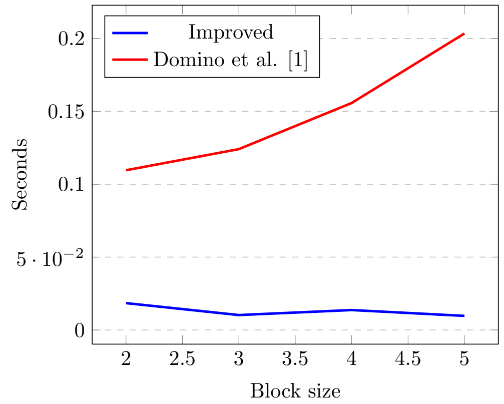
Contributors: Martin Rieth, Jackie Chen (Pele), Marco Arienti, Matt Larsen, Janine Bennett (ALPINE), Hemanth Kolla (ExaLearn)

Ongoing Performance Optimizations

- Higher-order joint moment/cumulant tensors can be expensive:
  - For a data set with N vars, co-kurtosis tensor is N^4 elements.
  - Each element is reduction over large number of samples.
- Idea: Leverage symmetry
  - Number of unique elements is $\binom{N+3}{4}$
- Efficient computation of hyper-triangular elements in sub-blocks: Domino, Gawron, Pawela, 2018, SIAM J. Sci. Comp., 40(3).
- We have identified further optimizations that are cache friendly, give ~5x speedup (work by summer intern Zitong Li, Wake Forest).







Thank You

Algorithm: ST-HOSVD

1. Choose $\mathbf{U}$ with projection rank R_1 such that: $\|\mathbf{X}_{(1)}\|^2 - \|\mathbf{U}'\mathbf{X}_{(1)}\|^2 \leq \epsilon^2 \|\mathcal{X}\|^2/3$
   a) Compute gram matrix: $\mathbf{X}_{(1)}\mathbf{X}_{(1)}'$
   b) Use eigendecomposition of N_1 x N_1 matrix to choose R_1
   c) Set $\mathbf{U}$ = R_1 leading eigenvectors of gram matrix
2. Shrink to size R_1 x N_2 x N_3: $\mathcal{Y} = \mathcal{X} \times_1 \mathbf{U}'$
3. Choose $\mathbf{V}$ with projection rank R_2 such that: $\|\mathbf{Y}_{(2)}\|^2 - \|\mathbf{V}'\mathbf{Y}_{(2)}\|^2 \leq \epsilon^2 \|\mathcal{X}\|^2/3$
   a) Compute gram matrix: $\mathbf{Y}_{(2)}\mathbf{Y}_{(2)}'$
   b) Use eigendecomposition of N_2 x N_2 matrix to choose R_2
   c) Set $\mathbf{V}$ = R_2 leading eigenvectors of gram matrix
4. Shrink to size R_1 x R_2 x N_3: $\mathcal{Z} = \mathcal{Y} \times_2 \mathbf{V}'$
5. Choose $\mathbf{W}$ with projection rank R_3 such that: $\|\mathbf{Z}_{(3)}\|^2 - \|\mathbf{W}'\mathbf{Z}_{(3)}\|^2 \leq \epsilon^2 \|\mathcal{X}\|^2/3$
   a) Compute gram matrix: $\mathbf{Z}_{(3)}\mathbf{Z}_{(3)}'$
   b) Use eigendecomposition of N_3 x N_3 matrix to choose R_3
   c) Set $\mathbf{W}$ = R_3 leading eigenvectors of gram matrix
6. Shrink to size R_1 x R_2 x R_3: $\mathcal{G} = \mathcal{Z} \times_3 \mathbf{W}'$

Vannieuwenhoven, Vandebril, Meerbergen (SISC 2012)

Formalizing anomaly detection

- Define Feature Moment (Kurtosis) Metric, FMM:
  - Quantifies contribution of feature, *i*, to the overall moment.

$$F_i^{j,n} = \frac{\sum_{k=1}^{N_f} \lambda_k \left(\hat{e}_i \cdot \hat{v}_k\right)^2}{\sum_{k=1}^{N_f} \lambda_k}$$

- FMMs sum to unity (over *i*): a.k.a, a distribution.
- *Anomalous events result in change the FMM distribution.*
- Use *f*-divergence metrics to quantify the change, signal an anomaly.

Independent Component Analysis (ICA)

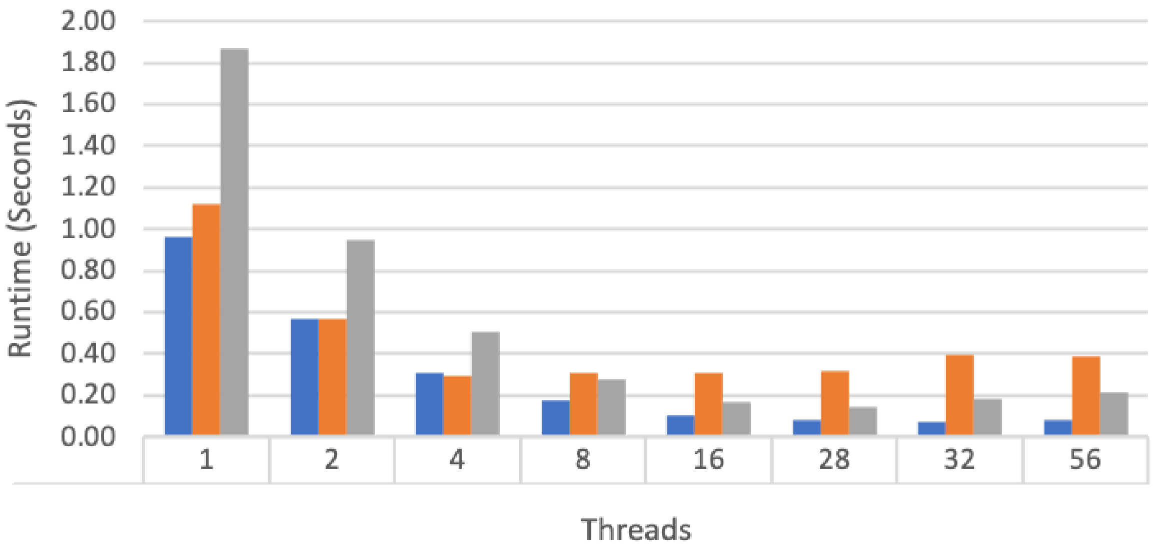
- Identifies non-Gaussian independent random variables that are linearly mixed:
  - $\boldsymbol{x} := \mathbf{A}\boldsymbol{s} + \boldsymbol{n}$. (***x***-observed vector; ***s***-independent sources, ***n***-Gaussian i.i.d noise)
- Specifically deals with fourth cumulant tensor (Lathauwer & Moore 2001, Comon & Jutten 2010, Anandkumar *et al.* 2014)
  - $\mathcal{M}_4 := \mathbb{E}[\boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x}] - \mathbb{E}[x_{i1}x_{i2}]\,\mathbb{E}[x_{i3}x_{i4}] - \mathbb{E}[x_{i1}x_{i3}]\,\mathbb{E}[x_{i2}x_{i4}] - \mathbb{E}[x_{i1}x_{i4}]\,\mathbb{E}[x_{i2}x_{i3}]$
  - $\mathcal{M}_4 = \sum_i \kappa_{s_i}\, a_i \otimes a_i \otimes a_i \otimes a_i$ (κ_{s_i}-excess Kurtosis of i^{th} source; a_i – columns of **A**)
- A simpler way to decompose $\mathcal{M}_4$: matricize and SVD (Anandkumar *et al.* 2014):
  - mat($\mathcal{M}_4$) = **M**= $\sum_i \kappa_{s_i}\, a_i \otimes vec(a_i \otimes a_i \otimes a_i)$
  - Caveats: repeated or close eigenvalues.

Algorithm Design: Accelerate Key Kernels

- **Focus on emerging hardware:**
  - Breadth of hardware spanning HPC (multi-core, heterogeneous).
  - Design for extreme heterogeneity (memory, compute, communication).
  - Explore algorithmic tradeoffs w.r.t. concurrency, parallelism, asynchrony, memory locality, latency.
- **Directly engage driver application(s) to define design space:**
  - Combustion Pele (ECP) is a direct customer, interested in anomaly detection and dimensionality reduction.
  - Exploring new customers e.g., Hardware-Software co-design, remote sensing, climate.
- **Leverage complementary capabilities within Sandia:**
  - ASCR Base Math funded research (PI: Tammy Kolda).
  - Kokkos (PI: Christian Trott).
  - Kokkos-Kernels (PI: Siva Rajamanickam).

Multi-threaded performance of Gram kernel of the HOSVD algorithm on Intel Xeon E5-2683

Image courtesy: Ben Cobb



- Three variants of exposing parallelism in the Gram matrix computation were investigated.
- The variants differ w.r.t. parallelism width, memory access patterns, extra storage.