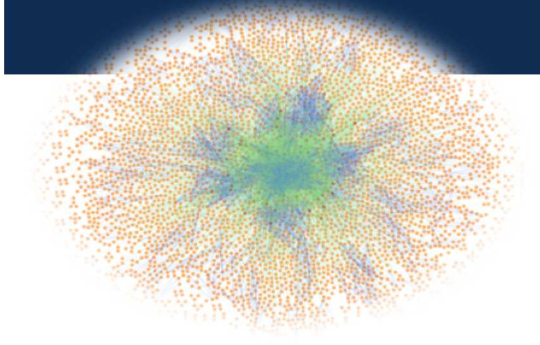
Exceptional service in the national interest

Sandia National Laboratories

SAND2019-4997C



Higher Order Moment Tensors for Combustion Analysis: GPU Acceleration

Hemanth Kolla, Jed Duersch, Aditya Konduri and Jackie Chen

Sandia National Laboratories, California, USA

U.S. DEPARTMENT OF ENERGY

NNSA

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed ... t DE-AC04-94AL85000.

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.

Acknowledgments

- US-DOE, Office of Science, Exascale Computing Project (ECP).
- US-DOE, ASCR funded project *“In-situ Machine Learning for Exascale”*.
- Warren Davis, Tim Shead, Prashant Rai, Philip Kegelmeyer, Tammy Kolda, Julia Ling.
- Oakridge Leadership Computing Facility (OLCF).

Problem context

- Combustion simulation data are multi-scale, multi-variate.
- Tensors are a very powerful abstraction; Tensor decomposition offer a rich set of analyses.
- Higher order tensors are usually large and expensive to compute/store.
- Tensor algebra can (often is) posed as “multi-linear algebra”.
- GPU acceleration, leveraging well-established linear algebra kernels, can provide great speedups.

Outline

Sandia National Laboratories

- Tensor-based analysis.
- Algorithm outline.
- Computational Challenge.
- GPU acceleration.

Outline

- Tensor-based analysis: anomalous event detection.
- Algorithm outline.
- Computational Challenge.
- GPU acceleration.

Hypothesis and proposed solution

- Information of anomalous events present in higher order statistical moments, e.g. kurtosis.

- For multi-variate non-Gaussian fields, joint moments (co-kurtosis) need to be analysed.

- Identify principal vectors of kurtosis (analogous to PCs in PCA) in the variable (a.k.a feature) space.

- *Anomalies manifest as principal kurtosis vectors (PKVs) that are "distinct".*

Hypothesis and proposed solution

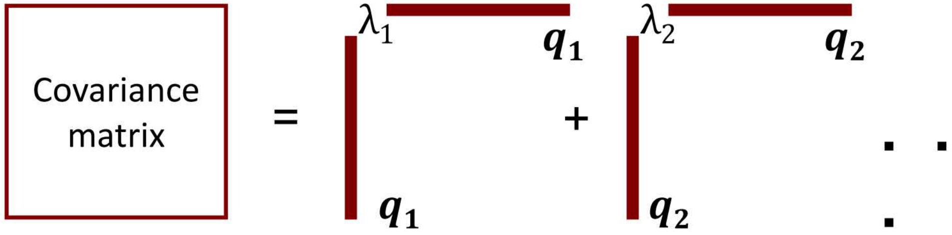
Sandia National Laboratories

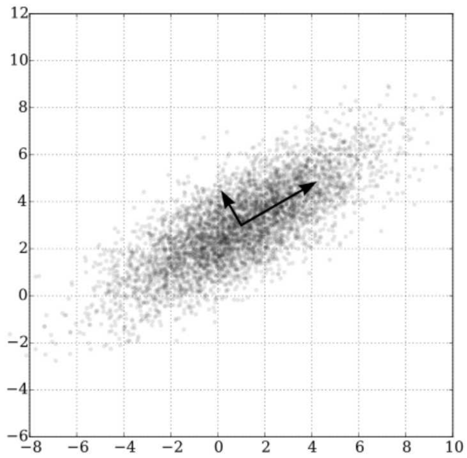
- Information of anomalous events present in higher order statistical moments, e.g. kurtosis.
- For multi-variate non-Gaussian fields, joint moments (co-kurtosis) need to be analysed.
- Identify principal vectors of kurtosis (analogous to PCs in PCA) in the variable (a.k.a feature) space.
- *Anomalies manifest as principal kurtosis vectors (PKVs) that are "distinct".*

Principal Component Analysis (PCA) Revisit

Sandia National Laboratories

- Eigen-decomposition of co-variance matrix.
  - $C = Q \Lambda Q^T$



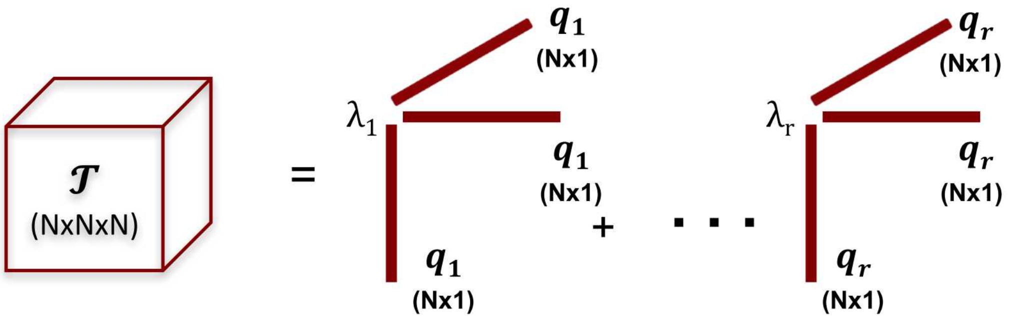


https://commons.wikimedia.org/w/index.php?curid=46871195

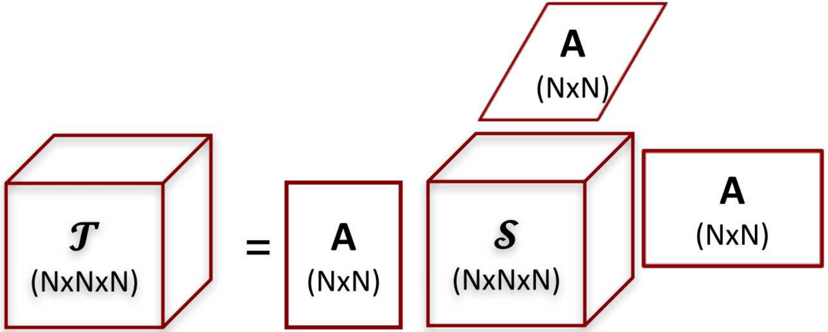
- Principal Components represent directions of variance in the data.
- By analogy, *we seek vectors that represent the higher moments*:
  - extend concept of PCA to higher joint moment (co-kurtosis) tensors.
  - becomes a *symmetric tensor decomposition problem*.

Symmetric Tensor Decomposition: Choices

- Canonical Polyadic (CP) decomp:
  - Sum of outer products of vectors.
  - What rank, ***r*** (Comon *et al.* 2008)?
  - Not orthogonal in general.



- Higher Order SVD (HOSVD, Lathauwer *et al.* 2000):
  - Symmetric tensor is a special case
  - Factor matrices (**A**) are orthogonal.



- Tensor Eigenpairs (Lim 2005, Qi 2005, Kolda & Mayo 2011) :
  - $\mathcal{T}\boldsymbol{x}^{m-1} = \lambda\boldsymbol{x}$

Outline

- Tensor-based analysis.
- Algorithm outline: Extract, Compare PKVs.
- Computational Challenge.
- GPU acceleration.

Simple Moment-Tensor Decomposition

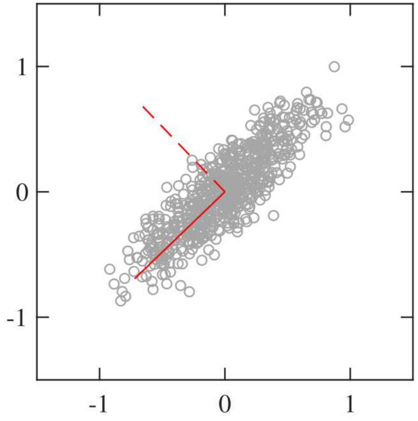
Sandia National Laboratories

- Motivated by connections to Independent Component Analysis (ICA).
- Operate on fourth cumulant tensor (Lathauwer & Moore 2001, Comon & Jutten 2010, Anandkumar *et al.* 2014)
  - $\mathcal{M}_4 \coloneqq \mathbb{E}[\boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x}] - \mathbb{E}[x_{i1}x_{i2}]\,\mathbb{E}[x_{i3}x_{i4}] - \mathbb{E}[x_{i1}x_{i3}]\,\mathbb{E}[x_{i2}x_{i4}] - \mathbb{E}[x_{i1}x_{i4}]\,\mathbb{E}[x_{i2}x_{i3}]$
- A simple way to decompose $\mathcal{M}_4$: matricize and SVD (Anandkumar *et al.* 2014):
  - mat($\mathcal{M}_4$) = **M** $= \sum_i \kappa_{s_i}\, a_i \otimes vec(a_i \otimes a_i \otimes a_i)$

Simple Moment-Tensor Decomposition

- Motivated by connections to Independent Component Analysis (ICA).
- Operate on fourth cumulant tensor (Lathauwer & Moore 2001, Comon & Jutten 2010, Anandkumar *et al.* 2014)
  - $\mathcal{M}_4 \coloneqq \mathbb{E}[\boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x}]$ - $\mathbb{E}[x_{i1}x_{i2}]\,\mathbb{E}[x_{i3}x_{i4}]$ - $\mathbb{E}[x_{i1}x_{i3}]\,\mathbb{E}[x_{i2}x_{i4}]$ - $\mathbb{E}[x_{i1}x_{i4}]\,\mathbb{E}[x_{i2}x_{i3}]$
- A simple way to decompose $\mathcal{M}_4$: matricize and SVD (Anandkumar *et al.* 2014):
  - mat($\mathcal{M}_4$) = **M**= $\sum_i \kappa_{s_i}\, a_i \otimes vec(a_i \otimes a_i \otimes a_i)$

PCA vecs

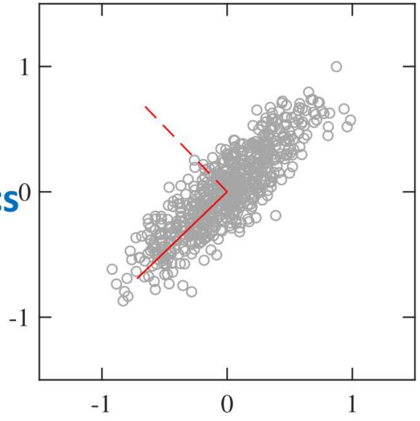


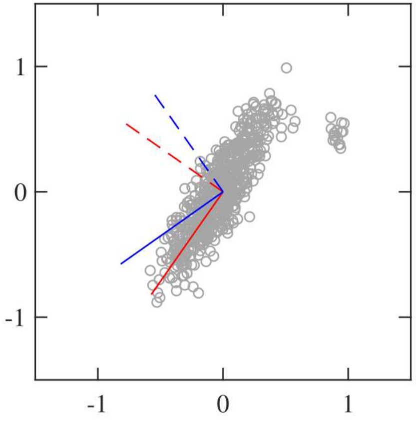
Simple Moment-Tensor Decomposition

- Motivated by connections to Independent Component Analysis (ICA).
- Operate on fourth cumulant tensor (Lathauwer & Moore 2001, Comon & Jutten 2010, Anandkumar *et al.* 2014)
  - $\mathcal{M}_4 \coloneqq \mathbb{E}[\boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x}] - \mathbb{E}[x_{i1}x_{i2}]\,\mathbb{E}[x_{i3}x_{i4}] - \mathbb{E}[x_{i1}x_{i3}]\,\mathbb{E}[x_{i2}x_{i4}] - \mathbb{E}[x_{i1}x_{i4}]\,\mathbb{E}[x_{i2}x_{i3}]$
- A simple way to decompose $\mathcal{M}_4$: matricize and SVD (Anandkumar *et al.* 2014):
  - mat($\mathcal{M}_4$) = **M**= $\sum_i \kappa_{s_i}\, a_i \otimes vec(a_i \otimes a_i \otimes a_i)$

PCA vecs

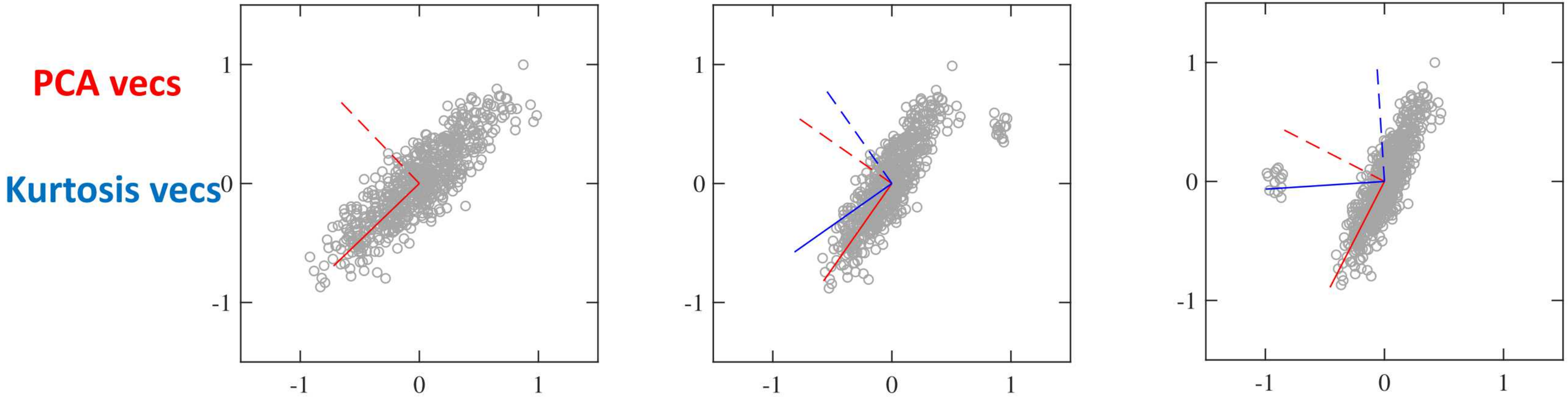
Kurtosis vecs





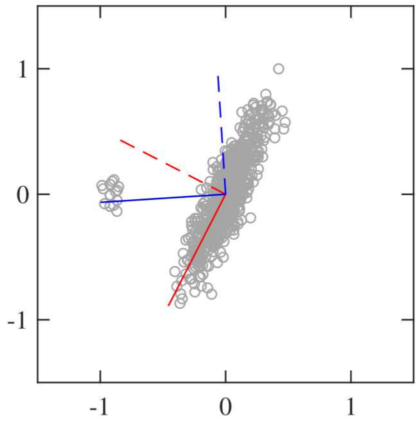
Simple Moment-Tensor Decomposition

- Motivated by connections to Independent Component Analysis (ICA).
- Operate on fourth cumulant tensor (Lathauwer & Moore 2001, Comon & Jutten 2010, Anandkumar *et al.* 2014)
  - $\mathcal{M}_4 := \mathbb{E}[\boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x}] - \mathbb{E}[x_{i1}x_{i2}]\,\mathbb{E}[x_{i3}x_{i4}] - \mathbb{E}[x_{i1}x_{i3}]\,\mathbb{E}[x_{i2}x_{i4}] - \mathbb{E}[x_{i1}x_{i4}]\,\mathbb{E}[x_{i2}x_{i3}]$
- A simple way to decompose $\mathcal{M}_4$: matricize and SVD (Anandkumar *et al.* 2014):
  - $\mathrm{mat}(\mathcal{M}_4) = \mathbf{M} = \sum_i \kappa_{s_i}\, a_i \otimes vec(a_i \otimes a_i \otimes a_i)$



Putting It Together: HCCI Data Set

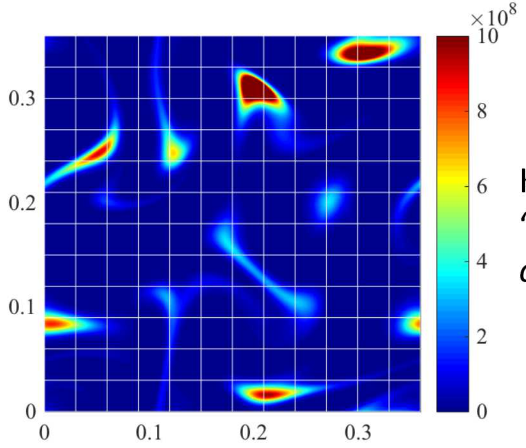
- Extract Principal Kurtosis Vectors (PKVs) on each MPI rank.
- Transform PKVs to a "moment (kurtosis) metric per feature (variable)".
  - Moment metrics quantify contribution of a feature to overall kurtosis.
  - Normalized (between 0-1), and also sum to 1 (like a discrete distribution).
- Compare moment metrics across MPI ranks (Hellinger distance).



feature X has a higher metric (close to 1) than feature Y (close to 0).

Putting It Together: HCCI Data Set

- Extract Principal Kurtosis Vectors (PKVs) on each MPI rank.
- Transform PKVs to a "moment (kurtosis) metric per feature (variable)".
  - Moment metrics quantify contribution of a feature to overall kurtosis.
  - Normalized (between 0-1), and also sum to 1 (like a discrete distribution).
- Compare moment metrics across MPI ranks (Hellinger distance).

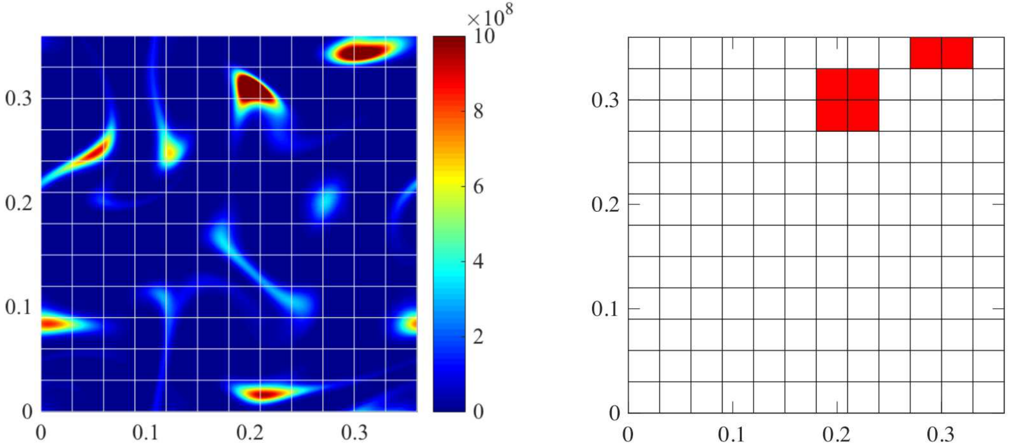


K. Aditya, H. Kolla, W. P. Kegelmeyer, T. M. Shead, J. Ling, Warren L Davis IV, 2019, "Anomaly detection in scientific data using joint statistical moments", *Journal of Computational Physics,* vol. 387, pp:522.

Putting It Together: HCCI Data Set

Sandia National Laboratories

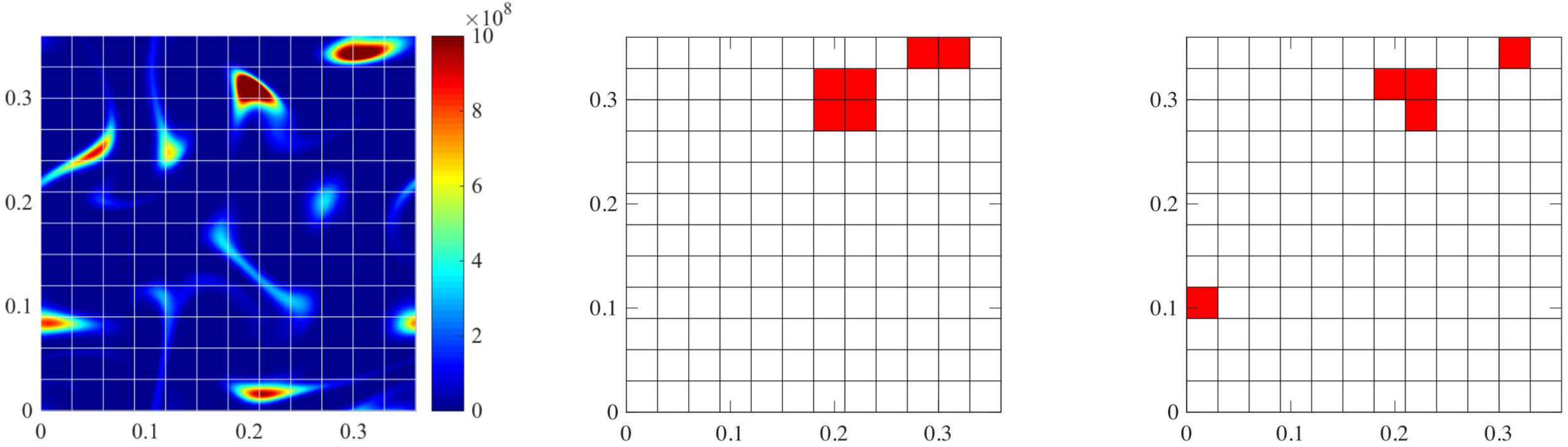
- Extract Principal Kurtosis Vectors (PKVs) on each MPI rank.
- Transform PKVs to a “moment (kurtosis) metric per feature (variable)”.
  - Moment metrics quantify contribution of a feature to overall kurtosis.
  - Normalized (between 0-1), and also sum to 1 (like a discrete distribution).
- Compare moment metrics across MPI ranks (Hellinger distance).



Putting It Together: HCCI Data Set

Sandia National Laboratories

- Extract Principal Kurtosis Vectors (PKVs) on each MPI rank.
- Transform PKVs to a “moment (kurtosis) metric per feature (variable)”.
  - Moment metrics quantify contribution of a feature to overall kurtosis.
  - Normalized (between 0-1), and also sum to 1 (like a discrete distribution).
- Compare moment metrics across MPI ranks (Hellinger distance).



Outline

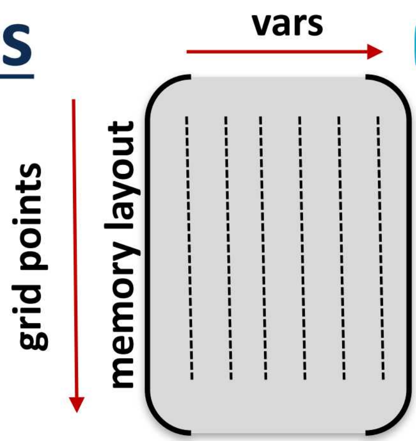
- Tensor-based analysis.
- Algorithm outline.
- Computational Challenge: Moment Tensors are expensive.
- GPU acceleration.

Computational Considerations

- Revisiting the fourth moment (cumulant) tensor:
  - $\mathcal{M}_4 := \mathbb{E}[\boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x}]$ - $\mathbb{E}[x_{i1}x_{i2}]\,\mathbb{E}[x_{i3}x_{i4}]$ - $\mathbb{E}[x_{i1}x_{i3}]\,\mathbb{E}[x_{i2}x_{i4}]$ - $\mathbb{E}[x_{i1}x_{i4}]\,\mathbb{E}[x_{i2}x_{i3}]$
- Co-kurtosis tensor is large ($\mathcal{O}(\text{nvars}^4)$).
- Very expensive to compute.

```
do L = 1, nvars
   do K = 1, nvars
      do J = 1, nvars
         do I = 1, nvars
            do N = 1,nx*ny*nz
                ......................
            enddo
         enddo
      enddo
   enddo
enddo
```

Solution: Refactored Lin-Alg Kernels



Typical layout of data matrix, $\boldsymbol{X}_{(\text{ngrids x nvars})}$

- <u>Key insight:</u> $\mathcal{M}_4$ can be expressed as sequence of operations on $\boldsymbol{X}^T$
  - $\mathcal{S} = \boldsymbol{X}^T \odot \boldsymbol{X}^T$; $\odot$ - Khatri-Rao product
  - mat($\mathcal{M}_4$) = $\mathcal{S}\ \mathcal{S}^T$; matrix-matrix multiplication
- Any moment tensor can be expressed as such sequence; in the limit covariance $\mathcal{C} = X^T X$
- This refactoring saves both memory and compute.

Khatri-Rao product

- Preliminary: Kronecker product $\mathbf{A} \otimes \mathbf{B}$

$$\mathbf{A} := \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}; \qquad \mathbf{A} \otimes \mathbf{B} := \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}$$

- Khatri-Rao product = column-wise Kronecker product

$$\mathbf{A} := \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}; \quad \mathbf{B} := [b_1 \dots . b_n]; \quad \mathbf{A} \odot \mathbf{B} := \begin{bmatrix} a_{11}b_1 & \cdots & a_{1n}b_n \\ \vdots & \ddots & \vdots \\ a_{m1}b_1 & \cdots & a_{mn}b_n \end{bmatrix}$$

Outline

- Tensor-based analysis.
- Algorithm outline.
- Computational Challenge.
- GPU acceleration: Handwritten + cuBLAS kernel

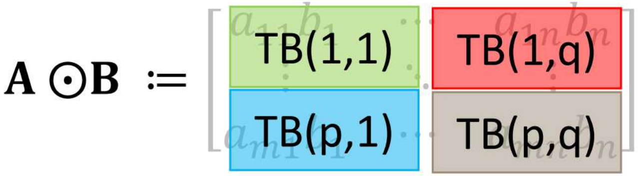
GPU-fication of Key Lin-Alg Kernels

- $\mathcal{S} = \boldsymbol{X}^T \odot \boldsymbol{X}^T$; Khatri-Rao product hand-optimised:
  - Coarse-grain parallelism.
  - Thread parallelism + memory coalescence.

$$\mathbf{A} \odot \mathbf{B} \coloneqq \begin{bmatrix} a_{11}b_1 & \cdots & a_{1n}b_n \\ \vdots & \ddots & \vdots \\ a_{m1}b_1 & \cdots & a_{mn}b_n \end{bmatrix}$$

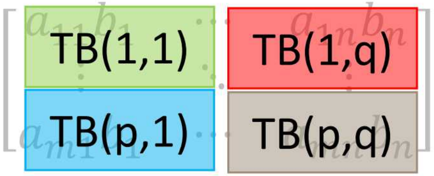
GPU-fication of Key Lin-Alg Kernels

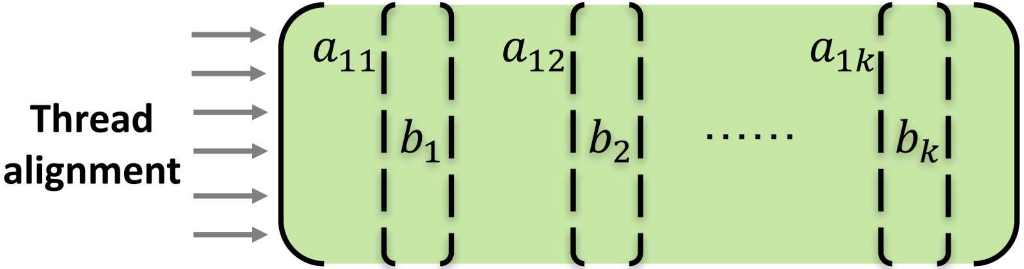
- $\mathcal{S} = \boldsymbol{X}^T \odot \boldsymbol{X}^T$; Khatri-Rao product hand-optimised:
  - Coarse-grain parallelism: 2D Thread Block grid
  - Thread parallelism + memory coalescence.



GPU-fication of Key Lin-Alg Kernels

- $\mathcal{S} = \boldsymbol{X}^T \odot \boldsymbol{X}^T$; Khatri-Rao product hand-optimised:
  - Coarse-grain parallelism: 2D Thread Block grid
  - Thread parallelism + memory coalescence.

$\mathbf{A} \odot \mathbf{B} :=$ 



- Threads operate on contiguous rows
- Reads (b_i) are contiguous vectors.
- Writes are also contiguous row elems.
- Threads in a block operate one column at a time.

- $\mathrm{mat}(\mathcal{M}_4) = \mathcal{S}\,\mathcal{S}^T$; `cublasDgemm` .

GPU-Speedup

- Tests for the 2D-HCCI data set:
  - 56 x 56 grid points per block (MPI rank), 28 species.
  - Comparisons of refactored + GPU-fied kernel vs naïve Fortran.
  - Time includes the cost of a one-time data transpose, and host-GPU copies.
- Runs on Rhea @OLCF:
  - Intel® Xeon® E5-2695 (14 cores) + K80 GPUs.
  - cuda/10.0.130.
- CPU version (naïve Fortran), 9.5s; GPU version, 0.3s, **30x speedup.**

Summary

- For anomaly detection in scientific data, statistical models based on higher moments may be promising.
- Use of "principal vectors of Kurtosis" as indicators of anomalous events.
- Metrics quantify change in the principal kurtosis vectors and identify anomalous subdomains.
- Construction of PKVs as a symmetric tensor decomposition problem.
- Refactored linear algebra kernels for tensor formation ported effectively to GPUs, with ~30x speedups.

Extra Slides

- 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)

- Identifies non-Gaussian independent random variables that are linearly mixed:

  - $\boldsymbol{x} := \mathbf{A}\boldsymbol{s} + \boldsymbol{n}$. ($\boldsymbol{x}$-observed vector; $\boldsymbol{s}$-independent sources, $\boldsymbol{n}$-Gaussian i.i.d noise)

- Specifically deals with fourth cumulant tensor (Lathauwer & Moore 2001, Comon & Jutten 2010, Anandkumar *et al.* 2014)

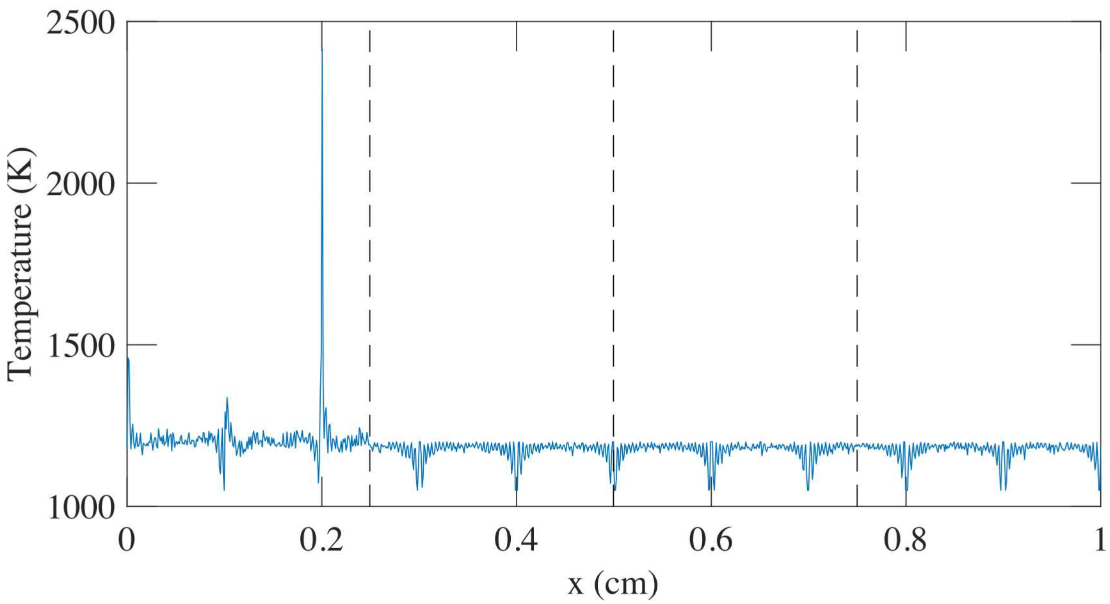
  - $\mathcal{M}_4 := \mathbb{E}[\boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{x}] - \mathbb{E}[x_{i1}x_{i2}]\,\mathbb{E}[x_{i3}x_{i4}] - \mathbb{E}[x_{i1}x_{i3}]\,\mathbb{E}[x_{i2}x_{i4}] - \mathbb{E}[x_{i1}x_{i4}]\,\mathbb{E}[x_{i2}x_{i3}]$

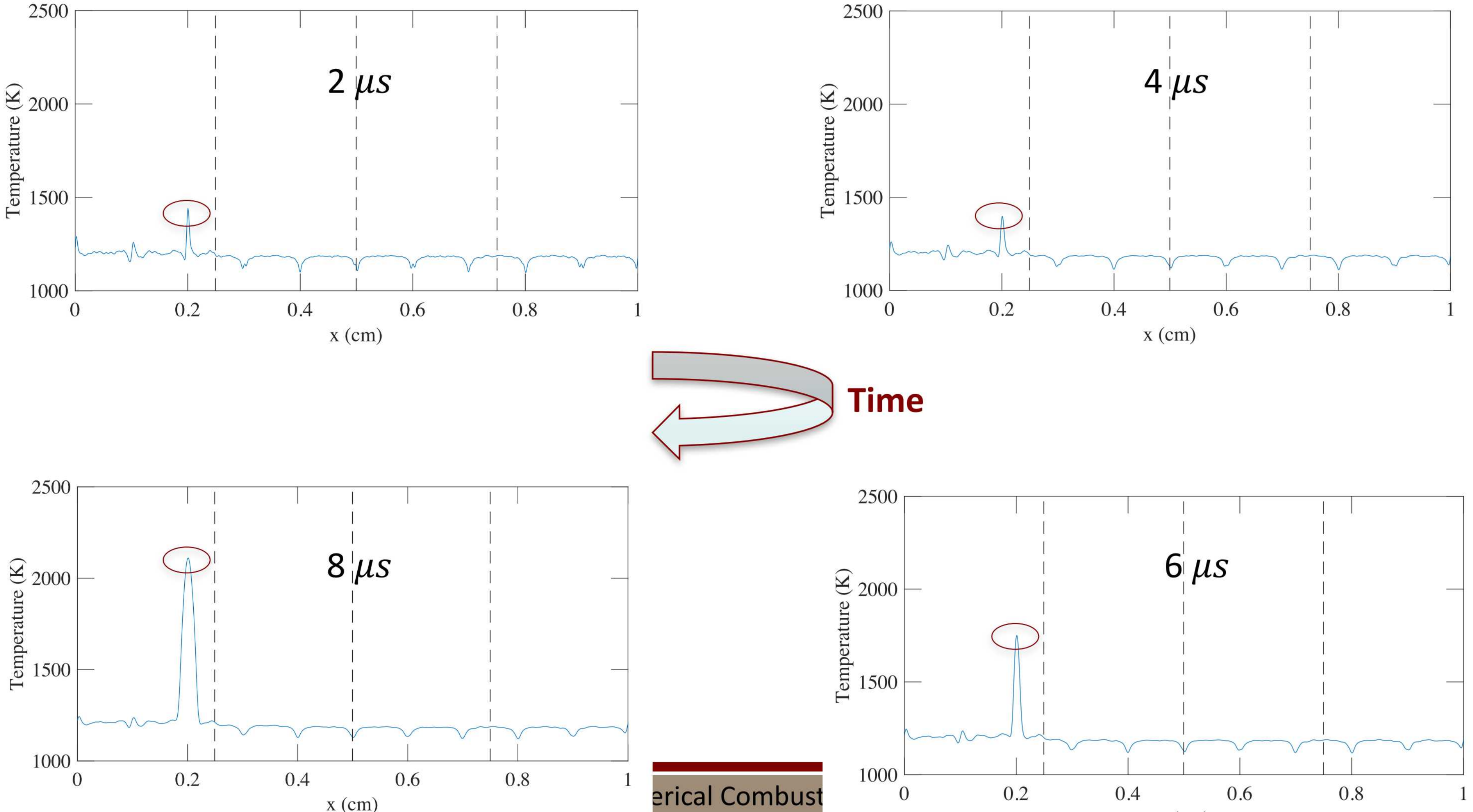
  - $\mathcal{M}_4 = \sum_i \kappa_{s_i}\, a_i \otimes a_i \otimes a_i \otimes a_i$ (κ_{s_i}-excess Kurtosis of ith source; a_i – columns of **A**)

Independent Component Analysis (ICA)

- Identifies non-Gaussian independent random variables that are linearly mixed:
  - $\boldsymbol{x} \;:= \mathbf{A}\boldsymbol{s} + \boldsymbol{n}$. ($\boldsymbol{x}$-observed vector; $\boldsymbol{s}$-independent sources, $\boldsymbol{n}$-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.

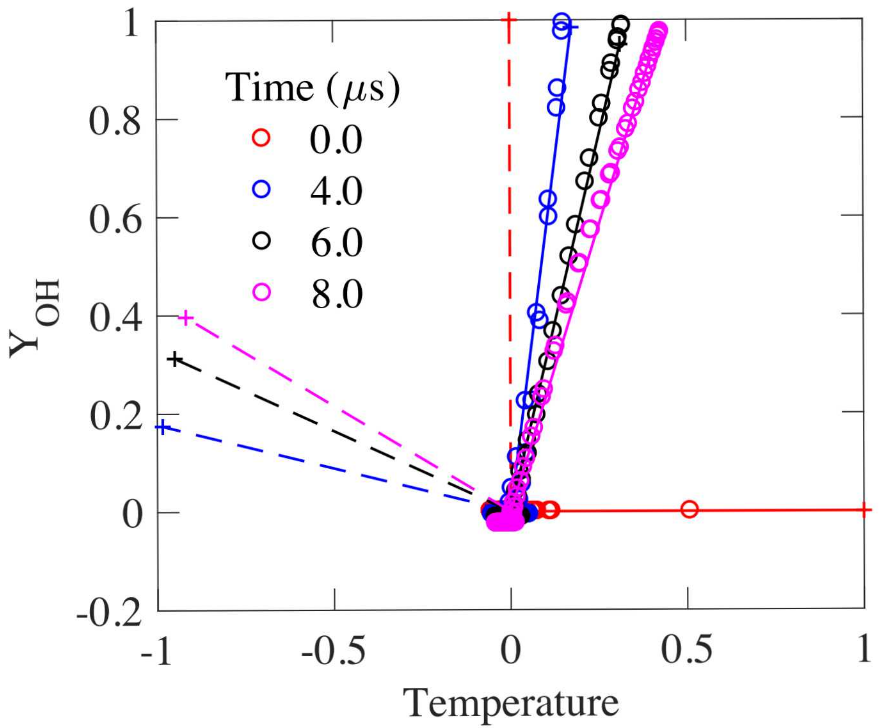
Simple 1D configuration with initial temperature inhomogeneity.



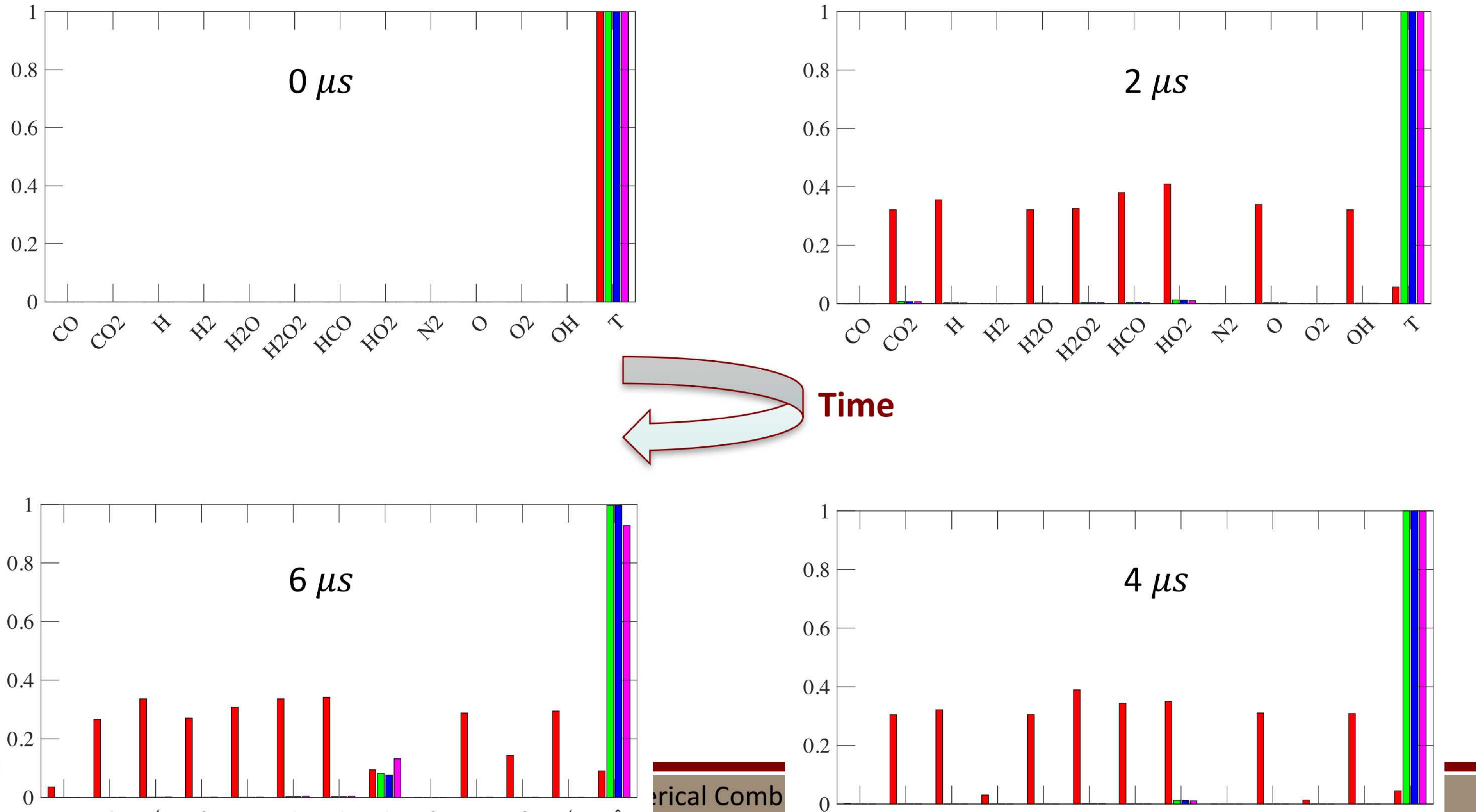

2 μs
4 μs
6 μs
8 μs
Time
Temperature (K)
x (cm)
2500
2000
1500
1000
0
0.2
0.4
0.6
0.8
1

Evolution of Kurtosis vectors.

Sandia National Laboratories



Feature Anomaly Metric (FAM_i) $= \sum_i \lambda_i(\vec{e_i} \cdot \vec{v_i}) \Big/ \sum_i \lambda_i$


0 μs
2 μs
4 μs
6 μs
Time
CO
CO2
H
H2
H2O
H2O2
HCO
HO2
N2
O
O2
OH
T
0
0.2
0.4
0.6
0.8
1

Defined per processor, aggregated over all features

$$M1_{p_j} = \sqrt{\frac{1}{n_v}\sum_{i=1}^{n_v}\frac{[\mathrm{FIM}_{i,p_j}(t) - \overline{\mathrm{FIM}_i(t)}]^2}{\overline{\mathrm{FIM}_i(t)}^2}}$$