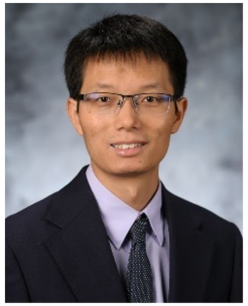
Progress of Tensor-Based High-Dimensional Uncertainty Quantification of Process Variations

Zichang He
Department of Electrical & Computer Engineering
University of California, Santa Barbara, CA, USA
zichanghe@ucsb.edu

Zichang He

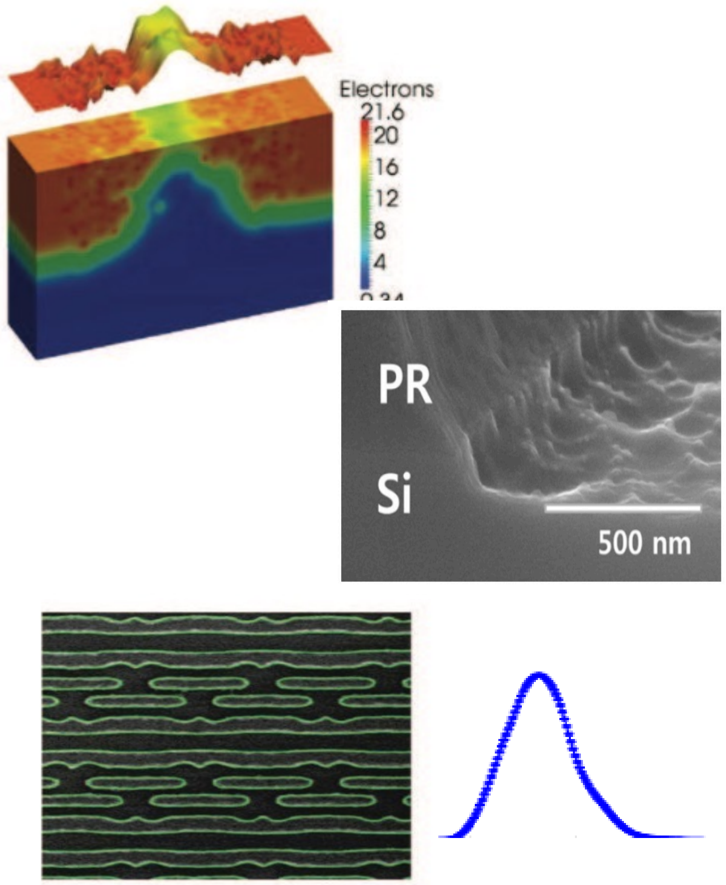


Zheng Zhang

ACES, Aug. 2021

Motivation: Uncertainty quantification

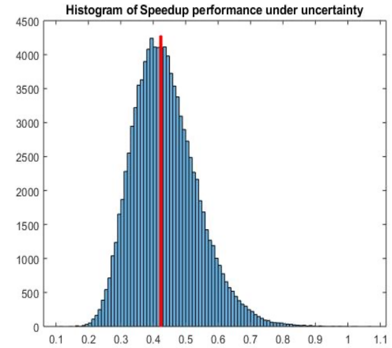
Process variations

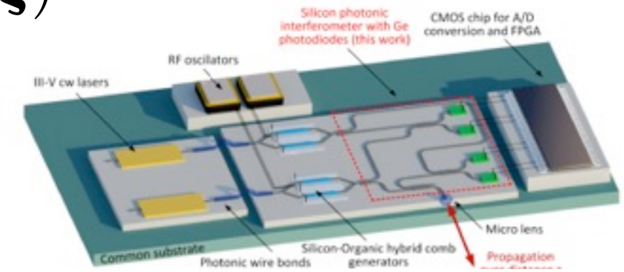


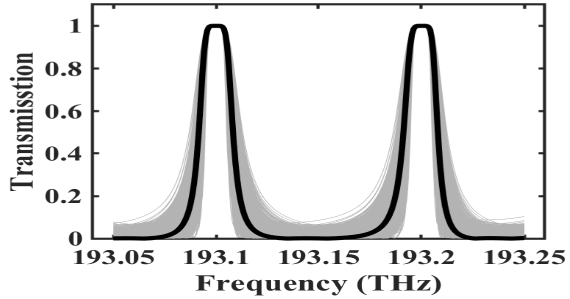
Device/ Circuit simulator

$\boldsymbol{\xi} \in \mathbb{R}^d \rightarrow y(\boldsymbol{\xi})$

Performance uncertainties







Detailed simulations are usually expensive!

Stochastic spectral methods

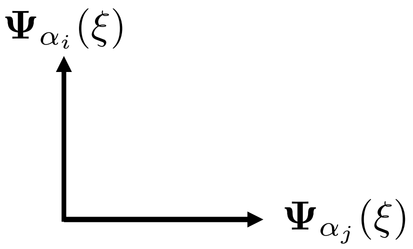
Given process variation random parameters

$$\boldsymbol{\xi} = [\xi_1, \ldots, \xi_d]$$

We want to find a surrogate model such that

$$y(\xi) \approx \sum_{\boldsymbol{\alpha} \in \Theta} c_{\boldsymbol{\alpha}} \boldsymbol{\Psi}_{\boldsymbol{\alpha}}(\xi)$$

- $\boldsymbol{\Psi}_{\boldsymbol{\alpha}}(\xi)$ is a predefined orthogonal and normalized polynomial basis.
- $c_{\boldsymbol{\alpha}}$ is the unknown coefficient.
- Θ is an index set to truncate the expansion. E.g., bound basis by a polynomial order *p*



Xiu, Dongbin. *Numerical methods for stochastic computations: a spectral method approach*. Princeton university press, 2010.

Challenges in surrogate modeling

Curse of dimensionality: Exponential complexity of needed samples

- Pseudo-projection-based stochastic collocation: $O(p+1)^d$

$$c_{\boldsymbol{\alpha}} = \int_{\mathbb{R}^d} y(\xi)\boldsymbol{\Psi}_{\boldsymbol{\alpha}}(\xi)\rho(\xi)d\xi = \sum_{1\le i_1,\ldots,i_d\le p+1} y(\xi_{i_1,\ldots,i_d})\boldsymbol{\Psi}_{\boldsymbol{\alpha}}(\xi_{i_1,\ldots,i_d})w_{i_1,\ldots,i_d}$$

- Regression-based stochastic collocation: $O\binom{p+d}{d} \approx O(d^p)$

$$c = \underset{c}{\operatorname{argmin}} \sum_n \left(y_n - \sum_{\boldsymbol{\alpha}\in\Theta} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\xi_n) \right)^2$$

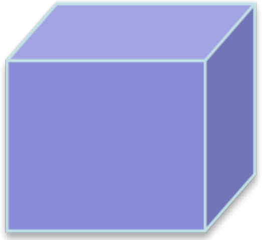
Compressive sensing (Li et al.), Hyperbolic regression (Roy et al.), ANOVA (Zhang et al.) …

Tensor background

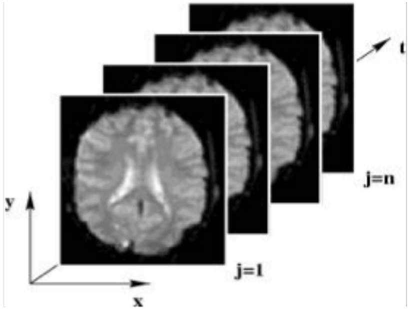
- matrix: 2-D data array

$$\mathbf{A} = [a_{i_1 i_2}] \in \mathbb{R}^{n_1 \times n_2}$$

- 3-D tensor



$$\mathcal{A} = [a_{i_1 i_2 i_3}] \in \mathbb{R}^{n_1 \times n_2 \times n_3}$$

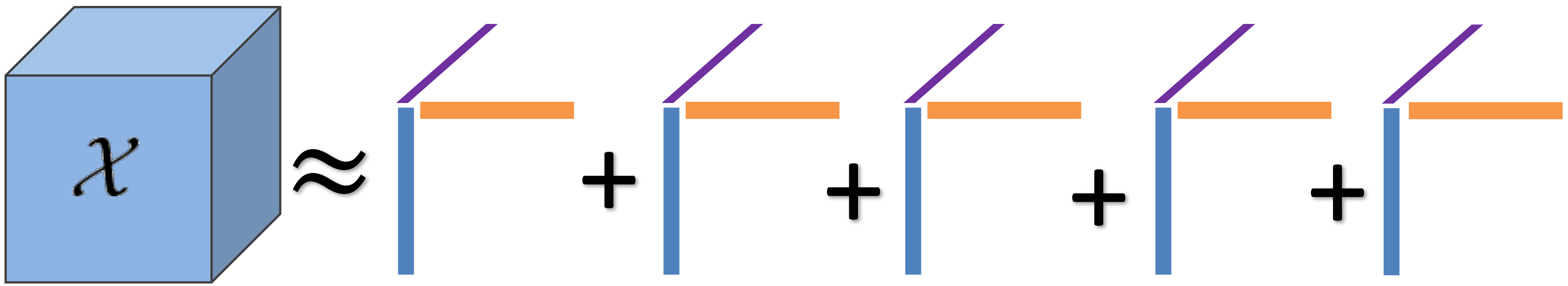


- General case: d-dimensional tensor

$$\mathcal{A} = [a_{i_1 \cdots i_d}] \in \mathbb{R}^{n_1 \times \cdots \times n_d}$$

Low-rank tensor decomposition

CP format: $$\mathcal{X} \approx \sum_{r=1}^{R} \mathbf{u}_r^{(1)} \circ \mathbf{u}_r^{(2)} \cdots \circ \mathbf{u}_r^{(d)} = [\![\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(d)}]\!]$$



Other popular tensor decomposition formats include Tucker, Tensor-train, Tensor-ring, etc.

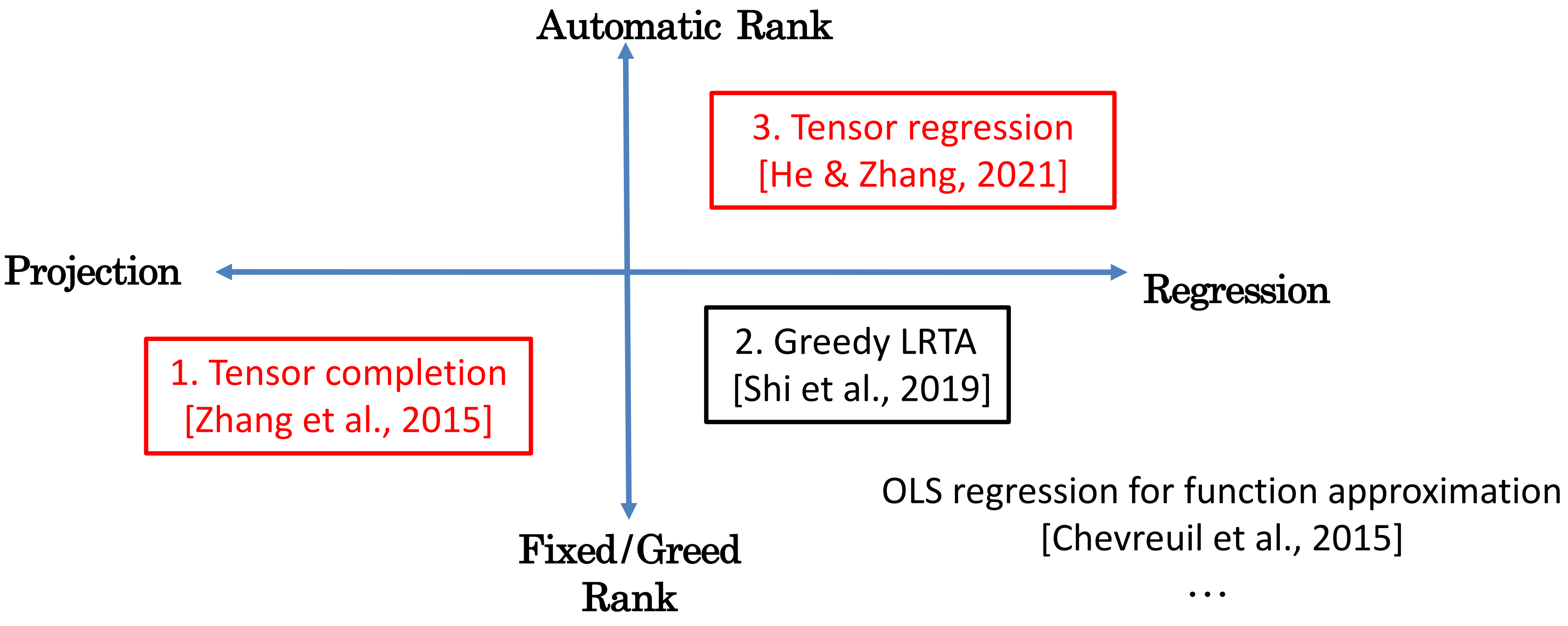
Low-rank modeling in $y(\xi) \approx \sum_{\boldsymbol{\alpha} \in \Theta} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\xi)$

Pseudo-projection-based SC:

1. Low rank modeling of the simulation output $y(\xi)$ on the quadrature grids: Tensor completion

Regression-based SC:

2. Low-rank modeling of the simulation output functional space $y(\xi)$: Greedy LRTA
3. Low-rank modeling of the coefficients $c_{\boldsymbol{\alpha}}$: Tensor regression



Tensor completion: Idea

$$c_{\boldsymbol{\alpha}} = \int_{\mathbb{R}^d} y(\xi)\boldsymbol{\Psi}_{\boldsymbol{\alpha}}(\xi)\rho(\xi)d\xi = \sum_{1\leq i_1,\ldots,i_d\leq p+1} y(\xi_{i_1,\ldots,i_d})\boldsymbol{\Psi}_{\boldsymbol{\alpha}}(\xi_{i_1,\ldots,i_d})w_{i_1,\ldots,i_d}$$

Tensor inner product form: $c_{\boldsymbol{\alpha}} = \langle \mathcal{Y}, \mathcal{W}_{\boldsymbol{\alpha}} \rangle$

$(p+1)^d$ simulations to fill the tensor $\mathcal{Y}$: Very expensive for detailed simulations

Solution:

- Assume the low-rank structure of $\mathcal{Y}$
- Based on limited simulated elements in $\mathcal{Y}$, to estimate the missing elements

Zhang et al., "Big-data tensor recovery for high-dimensional uncertainty quantification of process variations," TCPMT 2016.

Tensor completion: Formulation

$$\min_{\{\mathbf{u}_1^{(k)},\ldots,\mathbf{u}_R^{(k)}\}_{k=1}^{d}} \frac{1}{2}\|\mathbb{P}_\Omega(\sum_{r=1}^{R}\mathbf{u}_r^{(1)}\cdots\circ\mathbf{u}_r^{(d)}-\mathcal{Y})\|_F^2+\lambda\sum_{|\boldsymbol{\alpha}|=0}^{p}|\langle\sum_{r=1}^{R}\mathbf{u}_r^{(1)}\cdots\circ\mathbf{u}_r^{(d)},\mathcal{W}_{\boldsymbol{\alpha}}\rangle|$$

First term: To minimize the estimated error of the observed samples

Second term: A regularizor to enforce the sparsity of estimated coefficients

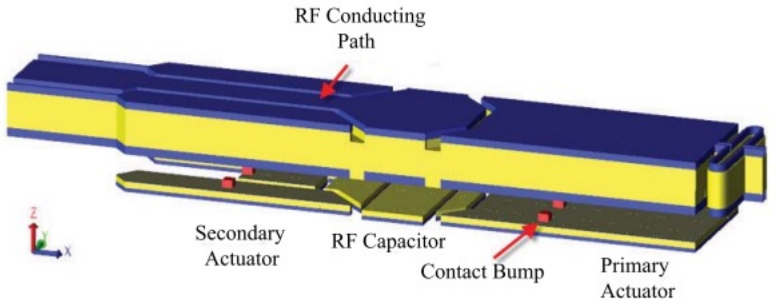
Remarks:

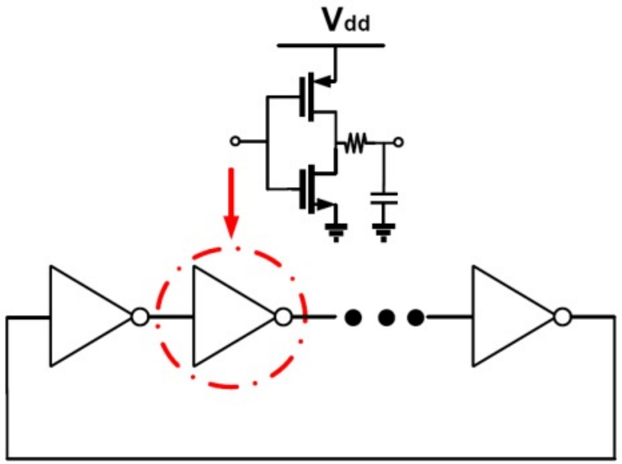
- The problem is efficiently solved via ADMM
- The tensor rank R is chosen via cross-validation
- The simulated samples are randomly chosen

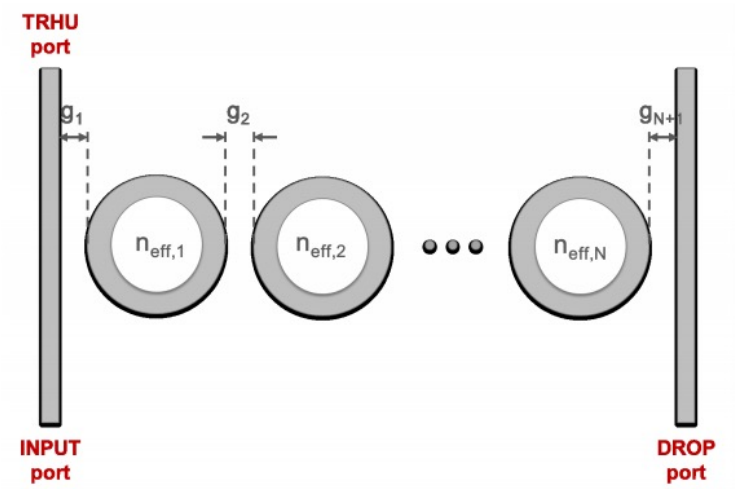
Tensor completion: Numerical results

Number of Simulation samples:

	Tensor product	Sparse grid	Proposed
MEMS (46-dim)	8.9 x 10^21	4512	300
Ring Oscillator (57-dim)	1.6 x 10^27	6844	500
Photonic bandpass filter (41-dim)	3.6 x 10^19	3445	500







Tensor regression: Idea

$$c = \arg\min_{c} \sum_{n} \left(y_n - \sum_{\boldsymbol{\alpha} \in \Theta} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\xi_n) \right)^2 \Rightarrow$$

$$\arg\min_{\{\mathbf{u}_1^{(k)}, \ldots, \mathbf{u}_R^{(k)}\}_{k=1}^{d}} \boxed{\sum_{n} \left(y(\xi_n) - \langle \sum_{r=1}^{R} \mathbf{u}_r^{(1)} \cdots \circ \mathbf{u}_r^{(d)}, \mathcal{B}(\xi_n) \rangle \right)^2}$$

Denoted as h(X)

We use the full tensor-product index set: (p+1)^d terms, leading to a coefficient tensor of size (p+1)^d.

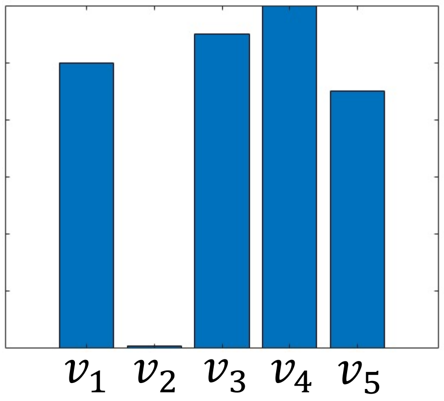
Based on the low-rank approximation, the number of unknown is reduced to linearly on dimensionality: (p+1)dR

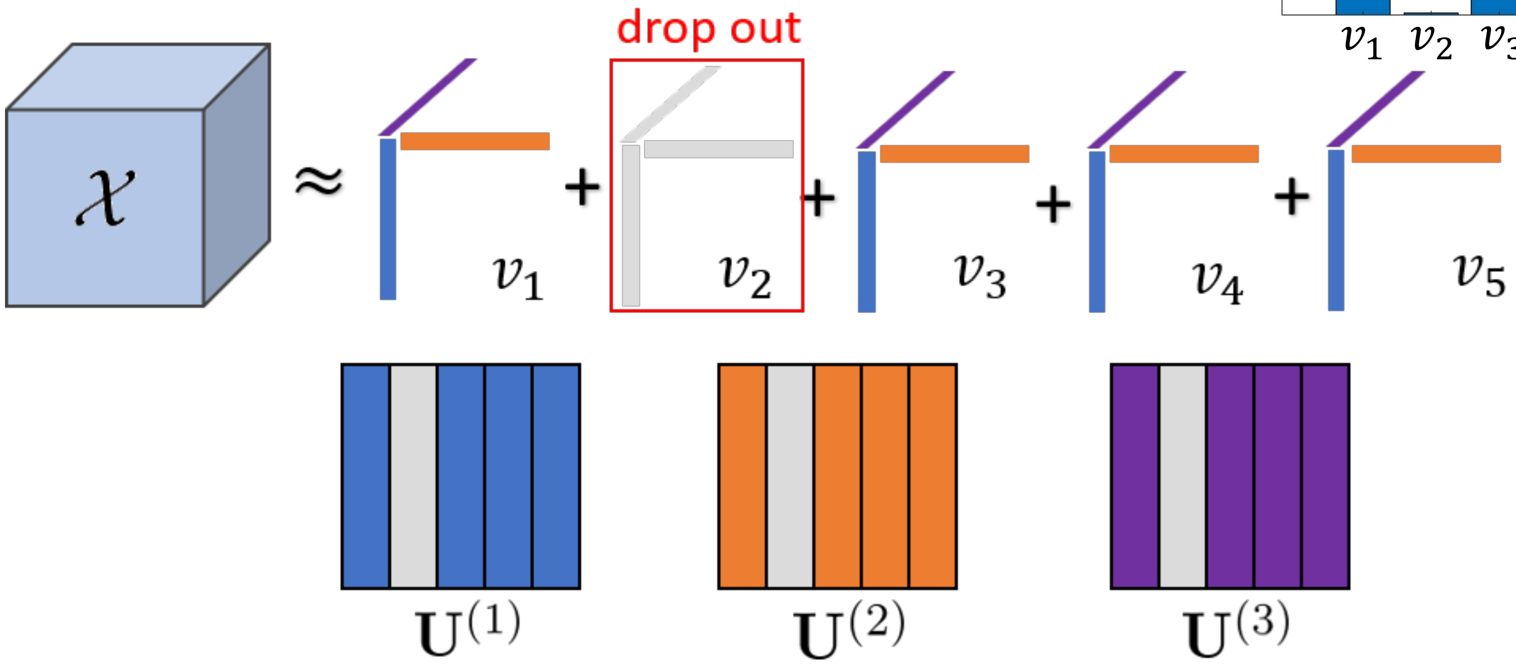
He and Zhang, "High-dimensional uncertainty quantification via tensor regression with rank determination and adaptive sampling," TCPMT 2021.

Tensor regression: Formulation

$$\min_{\{\mathbf{U}^{(k)}\}_{k=1}^{d}} f(\mathcal{X}) = h(\mathcal{X}) + \lambda g(\mathcal{X}).$$

$$g(\mathcal{X}) = \|\mathbf{v}\|_q, \mathbf{v} = \left(\sum_{k=1}^{d} \|\mathbf{u}_r^{(k)}\|_2^2\right)^{\frac{1}{2}}, \quad q \in (0, 1].$$





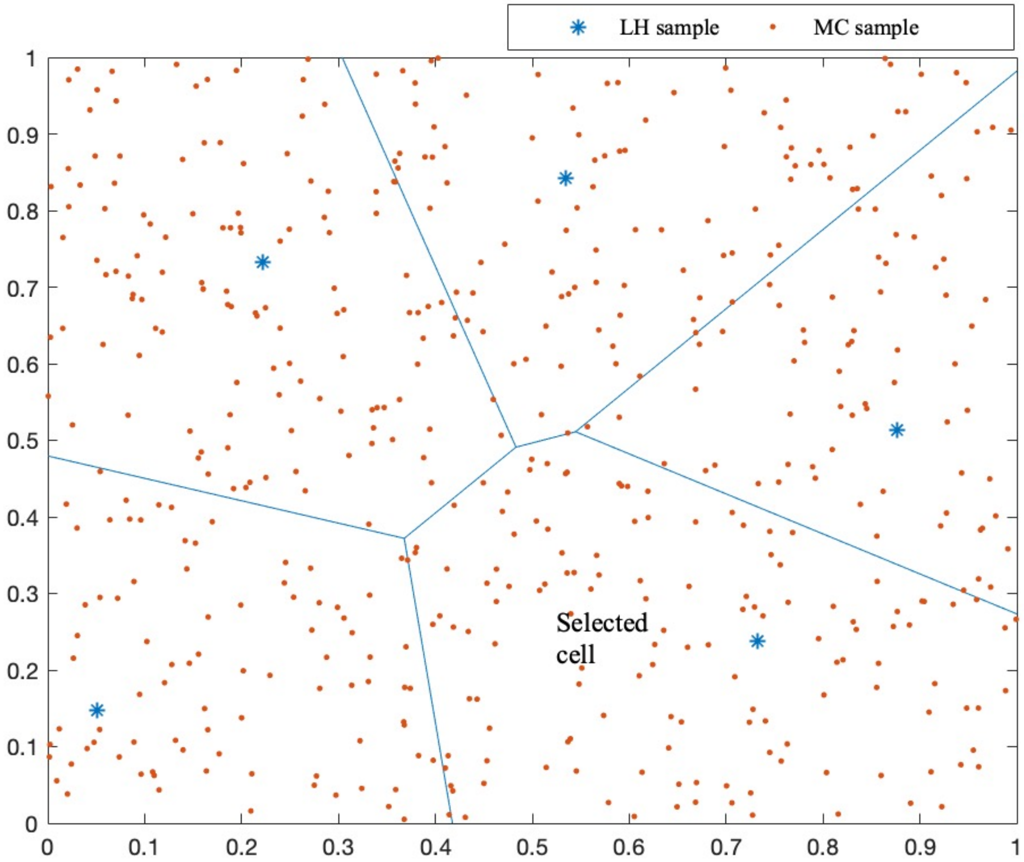
Tensor regression: Remarks

$$\min_{\{\mathbf{U}^{(k)}\}_{k=1}^{d}} f(\mathcal{X}) = h(\mathcal{X}) + \lambda g(\mathcal{X}).$$

Remarks:

- We set a large initial rank and the group sparsity can shrink the rank adaptively
- The non-convex & non-differentiable regularization term g(X) is reformulated via a variational equality
- The reformulated problem is efficiently solved via alternating solvers
- A space-filling-based method to guide the sampling of simulations adaptively

Tensor regression: Adaptive sampling



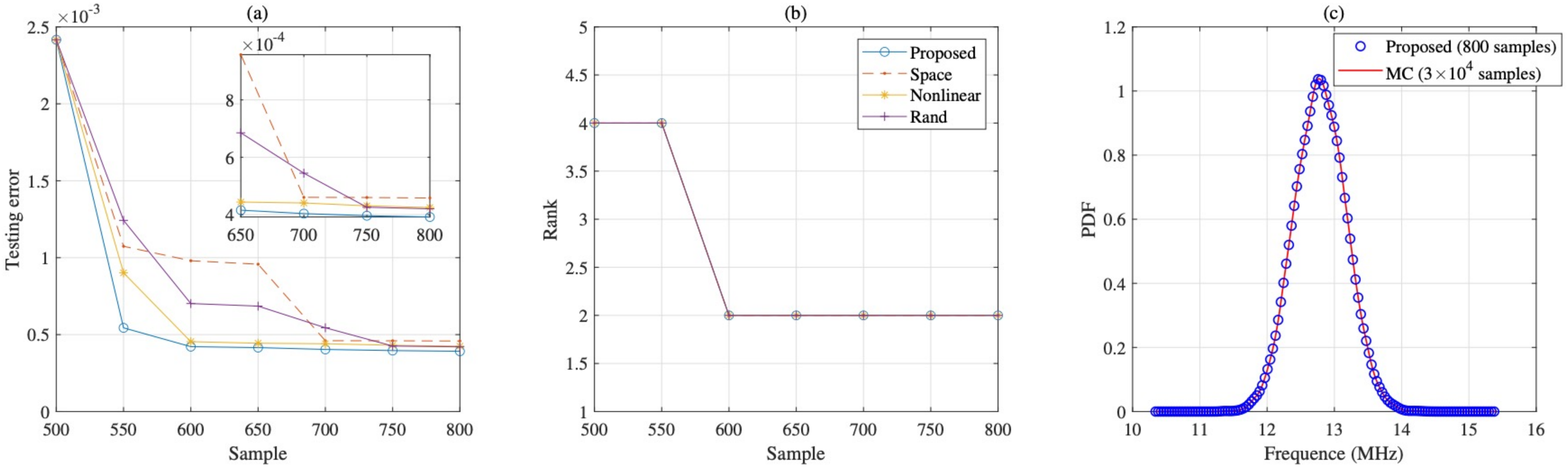
Initialize with Latin-Hypercube samples

Two-stage strategies:

- Based on Voronoi diagram, identify the cell region that is least-sampled
- Based on the nonlinearity measure, identify the sample that most nonlinear in the selected cell.

The most nonlinear sample in the least-sampled region: balancing exploration & exploitation.

Tensor regression: Numerical results



Uncertainty quantification of a 57-dim CMOS ring oscillator

Take-home message

Tensor-based methods for high-dimensional UQ is a great hammer for the curse of dimensionality.

Within the low-rank tensor-based modeling, some technical problems have been investigated in the recent paper

- Automatic rank determination
- Adaptive sampling

Future works

Application-specific optimal tensor network structure

Certified adaptive sampling methods in constructing tensor-based model

Uncertainty quantification of multiple interested metrics

…

Reference

- Z. He and Z. Zhang, "Progress of tensor-based high-dimensional uncertainty quantification of process variations," Applied Computational Electromagnetics Society Conference (ACES), August 2021.
- Z. Zhang, T.-W. Weng, and L. Daniel, “Big-data tensor recovery for high-dimensional uncertainty quantification of process variations,” IEEE Trans. Compon. Packag. Manuf. Technol., vol. 7, no. 5, pp. 687–697, 2016.
- Z. Zhang, K. Batselier, H. Liu, L. Daniel, and N. Wong, “Tensor computation: A new framework for high-dimensional problems in EDA,” IEEE Trans. Comput.-Aided Design Integr. Circuits Syst., vol. 36, no. 4, pp. 521–536, 2016.
- X. Shi, H. Yan, Q. Huang, J. Zhang, L. Shi, and L. He, “Metamodel based high-dimensional yield analysis using low-rank tensor approximation,” in Proc. Design Autom. Conf, 2019, pp. 1–6.
- Z. He and Z. Zhang, “High-dimensional uncertainty quantification via active and rank-adaptive tensor regression,” in Proc. Electr. Perform. Electron. Packag. Syst., 2020, pp. 1–3.
- Z. He and Z. Zhang, “High-dimensional uncertainty quantification via tensor regression with rank determination and adaptive sampling,” accepted by IEEE Trans. Compon. Packag. Manuf. Technol., doi: 10.1109/TCPMT.2021.3093432.