UNT

UNIVERSITY OF NORTH TEXAS

Effective and Scalable Uncertainty Evaluation for Large-Scale Complex System Applications

Junfei Xie, Yan Wan, Yi Zhou, *University of North Texas*

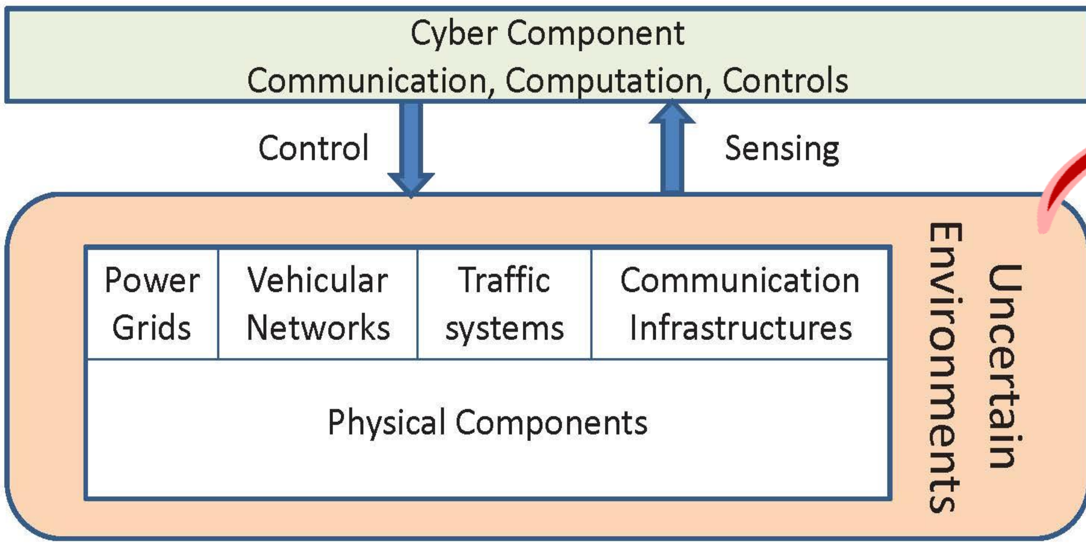
Kevin Mills, James J. Filliben, *NIST*

Yu Lei, *University of Texas at Arlington*

Winter Simulation Conference, 2014

MOTIVATION

- Modern large-scale cyber-physical systems (CPSs) involve a large number of uncertain parameters.



Uncertainties

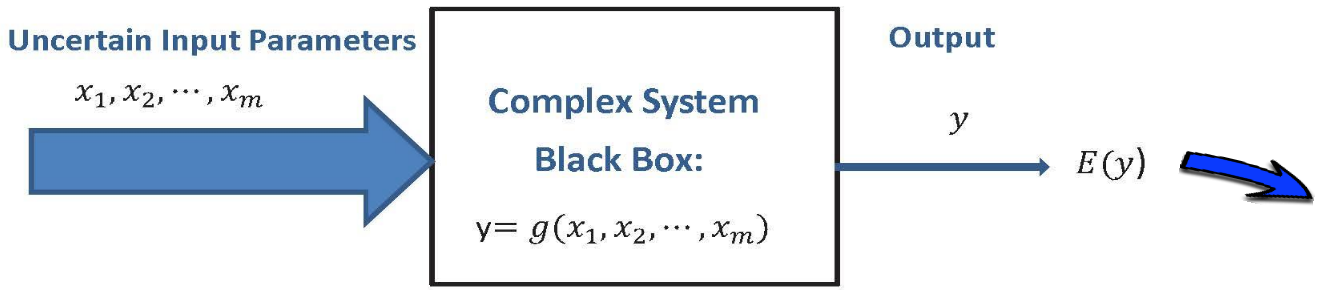
- Modulate system's **dynamics**
- Pose significant challenges for **real-time system evaluation** & **decision-support**

Management of physical dynamics must be designed in a way to achieve robust performance under the uncertainties

Effective uncertainty evaluation

PROBLEM FORMULATION

- **Problem Formulation**



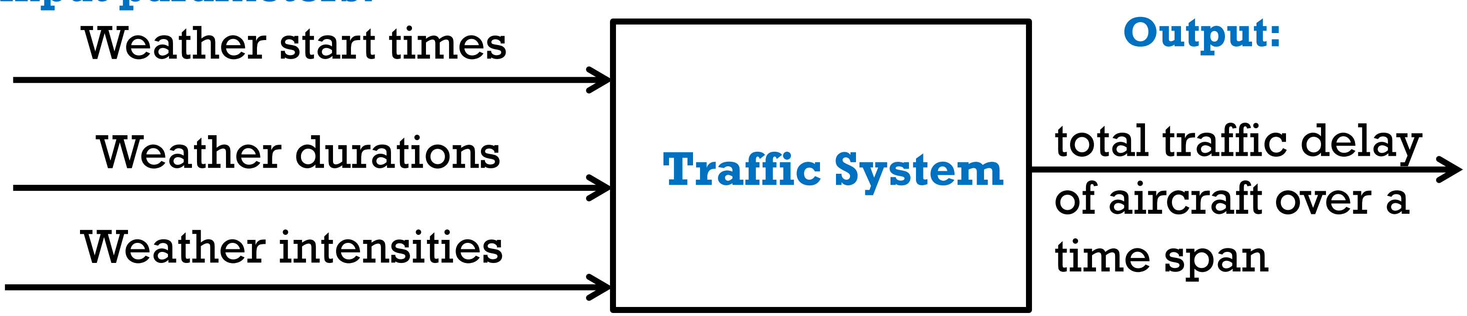
Goal: correctly estimate the mean output

- **Example Application**

Air Traffic Flow Management

Uncertain input parameters:

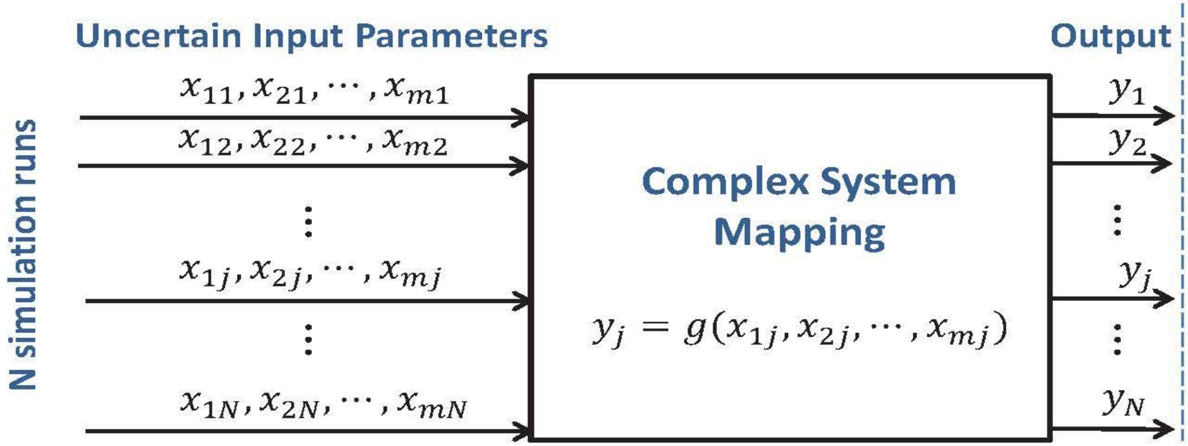
Output:



EXISTING METHODS

- **Monte Carlo Simulation Method**

High Computational Cost !!



Problem

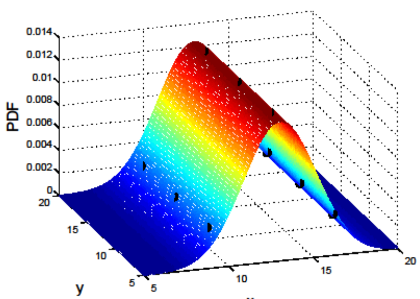
To calculate the mean output $E(y)$, the Monte Carlo method requires a large number of simulations to converge:

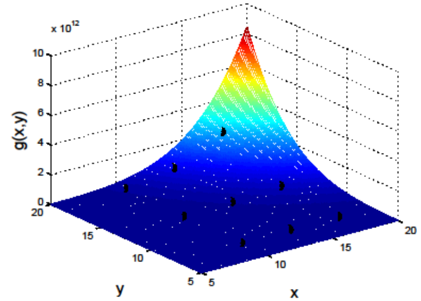
$$E(y) = \frac{1}{N} \sum_{j}^{N \to \infty} y_j$$

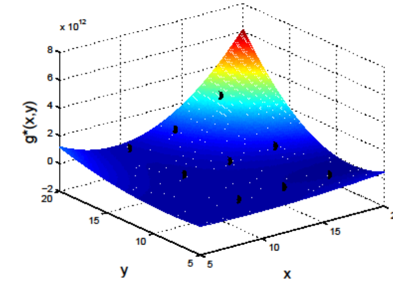
M-PCM

- **Multivariate Probabilistic Collocation Method (M-PCM)**

Knowledge of the two variables' probability distributions → Simulation of the original system at the selected PCM points → Reduced-order mapping with the same mean statistics







Original System Mapping:

$$g(x_1, x_2, \ldots, x_m) = \sum_{k_1=0}^{2n_1-1} \sum_{k_2=0}^{2n_2-1} \cdots \sum_{k_m=0}^{2n_m-1} \Psi_{k_1,\ldots,k_m} \prod_{i=1}^{m} x_i^{k_i}$$

Low-order Mapping:

$$g^*(x_1, x_2, \ldots, x_m) = \sum_{k_1=0}^{n_1-1} \sum_{k_2=0}^{n_2-1} \cdots \sum_{k_m=0}^{n_m-1} \Omega_{k_1,\ldots,k_m} \prod_{i=1}^{m} x_i^{k_i}$$

Reduce the number of simulations from

$$2^m \prod_{i=1}^{m} n_i \longrightarrow \prod_{i=1}^{m} n_i$$

Predict the correct mean output!

$$E[g(x_1, x_2, \ldots, x_m)] = E[g^*(x_1, x_2, \ldots, x_m)]$$

Y. Zhou, Y. Wan, S. Roy, C. Taylor, C. Wanke, D. Ramamurthy, J. Xie, "Multivariate Probabilistic Collocation Method for Effective Uncertainty Evaluation with Application to Air Traffic Management", IEEE Transactions on Systems, Man and Cybernetics: System, Vol. 44, No. 10, pp.1347-1363, 2014.

NEEDS

- **Limitation of M-PCM:** Not scalable with the number of parameters

Suppose $n_i = 2, i = 1,2,..,m$
Number of simulations: $\prod_{i=1}^{m} n_i = 2^m$

Computational load issue for real-time applications

	Number of parameters	Number of simulations
	m=2	2^2
	m=100	2^{100}

Still too large !!

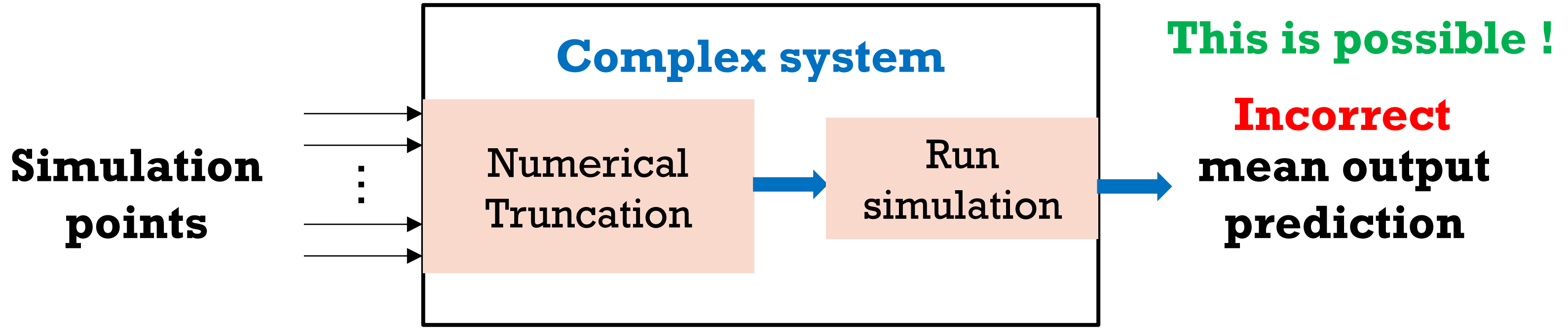
- **Possibility of Further Reduction**
  - M-PCM assumes that there exist cross-multiplication terms for all combinations of uncertain parameters of all degrees
  - Suppose $n_i = 1, i = 1,2,..,m$

$$g(x_1, x_2, \ldots, x_m) = a_0 + a_1 x_1 + \cdots + a_m x_m + a_{m+1} x_1 x_2 + \cdots + a_N x_1 x_2 \ldots x_m$$

Some of these terms may not exist ($a_i \approx 0$ or $a_i = 0$) in realistic applications

FURTHER REDUCTION

- **Challenge:** existence of a practical numerical issue
  - Many system simulations have constraints on the resolutions of input parameters



- **Approach:** Integration of M-PCM with the orthogonal fractional factorial design (OFFD)------ **M-PCM-OFFD**
  - OFFDs meet our need to reduce the number of simulations
  - Both OFFDs and our study are motivated by the same assumption

    ---high-order interactions among parameters are insignificant in real applications

PRELIMINARY: OFFD

- **Orthogonal Fractional Factorial Designs (OFFDs)**
  - Selects a subset of experimental combinations that best estimate the main effects of single factors and low-order interaction effects.

x_1, x_2, x_3 are factors (input parameters)

y is the output

- **Full Factorial Designs**

All possible combinations of levels of all factors.

'$-$' lower level

'$+$' higher level

	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	y
1	$-$	$-$	$-$	$+$	$+$	$+$	$-$	y_1
2	$+$	$-$	$-$	$-$	$-$	$+$	$+$	y_2
3	$-$	$+$	$-$	$-$	$+$	$-$	$+$	y_3
4	$+$	$+$	$-$	$+$	$-$	$-$	$-$	y_4
5	$-$	$-$	$+$	$+$	$-$	$-$	$+$	y_5
6	$+$	$-$	$+$	$-$	$+$	$-$	$-$	y_6
7	$-$	$+$	$+$	$-$	$-$	$+$	$-$	y_7
8	$+$	$+$	$+$	$+$	$+$	$+$	$+$	y_8

2^3 full factorial design

	x_1	x_2	x_3	y
1	$-$	$-$	$+$	y_5
2	$+$	$-$	$-$	y_2
3	$-$	$+$	$-$	y_3
4	$+$	$+$	$+$	y_8

	x_1	x_2	x_3	y
1	$-$	$-$	$-$	y_1
2	$+$	$+$	$-$	y_4
3	$-$	$+$	$+$	y_7
4	$+$	$-$	$+$	y_6

2^{3-1} OFFDs

PROPERTIES OF OFFD

- **Main Effect and Interaction**

Main effect ME_i of factor x_i

$$ME_i = (\bar{y} \text{ when } x_i \text{ is } +) - (\bar{y} \text{ when } x_i \text{ is } -)$$

Interaction effect ME_{ij} of $x_i x_j$

$$ME_{ij} = (\bar{y} \text{ when } x_i x_j \text{ is } +) - (\bar{y} \text{ when } x_i x_j \text{ is } -)$$

	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	y
1	−	−	−	+	+	+	−	y_1
2	+	−	−	−	−	+	+	y_2
3	−	+	−	−	+	−	+	y_3
4	+	+	−	+	−	−	−	y_4
5	−	−	+	+	−	−	+	y_5
6	+	−	+	−	+	−	−	y_6
7	−	+	+	−	−	+	−	y_7
8	+	+	+	+	+	+	+	y_8

Regression Model:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{i.j} x_i x_j + \epsilon$$

The least square estimators for β, denoted as $\hat{\beta}$ are:

$$\hat{\beta}_i = \frac{1}{2} ME_i, \qquad \hat{\beta}_{ij} = \frac{1}{2} ME_{ij}$$

e.g., $ME_3 = \dfrac{y_5+y_6+y_7+y_8}{4} - \dfrac{y_1+y_2+y_3+y_4}{4}$

$$ME_{23} = \frac{y_1 + y_2 + y_7 + y_8}{4} - \frac{y_3 + y_4 + y_5 + y_6}{4}$$

PROPERTIES OF OFFD (CONT.)

- **The main effects and interactions estimated by the subset of simulations selected by OFFDs are aliased.**

	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	y
1	−	−	−	+	+	+	−	y_1
2	+	−	−	−	−	+	+	y_2
3	−	+	−	−	+	−	+	y_3
4	+	+	−	+	−	−	−	y_4
5	−	−	+	+	−	−	+	y_5
6	+	−	+	−	+	−	−	y_6
7	−	+	+	−	−	+	−	y_7
8	+	+	+	+	+	+	+	y_8

$x_3 = x_1x_2$

Generator: $I = x_1x_2x_3$

	x_1	x_2	x_3	y
1	−	−	+	y_5
2	+	−	−	y_2
3	−	+	−	y_3
4	+	+	+	y_8

These are what OFFD estimates

$I = x_1x_2x_3 \rightarrow$	$\beta_0 + \beta_{1.2.3}$	
$x_1 = x_2x_3 \rightarrow$	$\beta_1 + \beta_{2.3}$	$\widetilde{ME_1} = ME_1 + ME_{2.3}$
$x_2 = x_1x_3 \rightarrow$	$\beta_2 + \beta_{1.3}$	$\widetilde{ME_2} = ME_2 + ME_{1.3}$
$x_3 = x_1x_2 \rightarrow$	$\beta_3 + \beta_{1.2}$	$\widetilde{ME_3} = ME_3 + ME_{1.2}$

Regression Model:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{i.j} x_i x_j + \epsilon$$

PROCEDURES OF APPLYING OFFD

Step 1	Step 2	Step 3
Generate the $P^{m-\gamma}$ full factorial design for $m-\gamma$ factors	Specify γ generators	Determine the levels of all other γ factors for each experimental run

P: the number of levels
m: the number of factors
γ: the fractionation constant

e.g., 2^{3-1} OFFD, $P=2$, $m=3$, $\gamma=1$,

2^2 full factorial design

x_1	x_2	x_3
$-$	$-$	
$-$	$+$	
$+$	$-$	
$+$	$+$	

Generator

$I=-x_1x_2x_3$

$\boxed{x_3=-x_1x_2}$

'$-$' lower level
'$+$' higher level

2^{3-1} OFFD

x_1	x_2	x_3
$-$	$-$	$-$
$-$	$+$	$+$
$+$	$-$	$+$
$+$	$+$	$-$

M-PCM-OFFD

- If we view all simulation points selected by M-PCM as a full factorial design, the OFFDs provide systematic procedures to select a subset of simulation points.
- **Design Procedures**

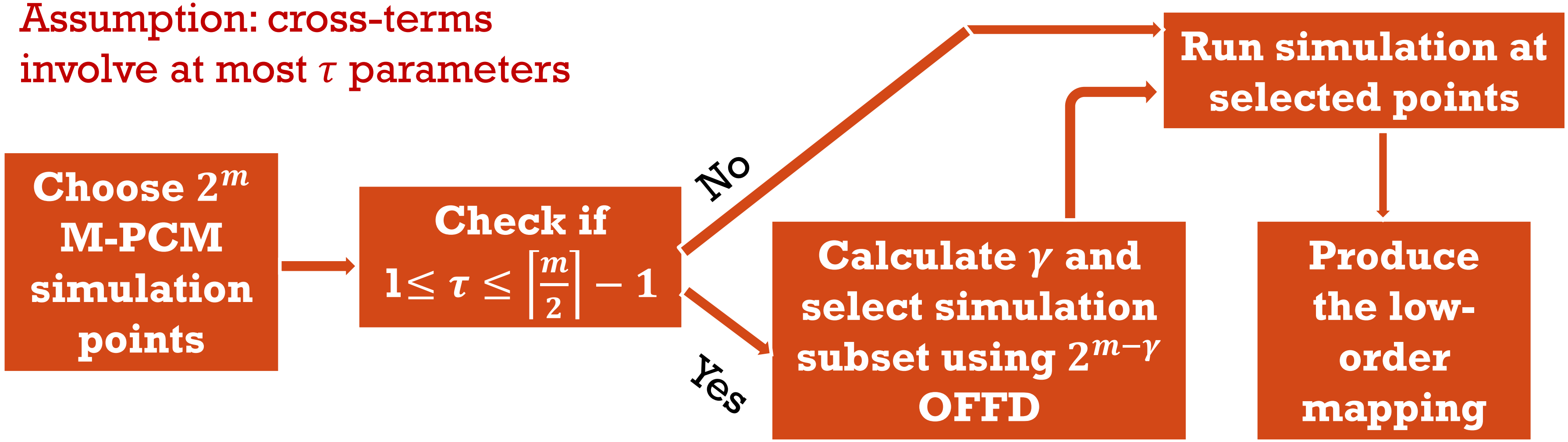
Original System Mapping:

$$g(x_1, x_2, \ldots x_m) = \sum_{k_1=0}^{3} \sum_{k_2=0}^{3} \cdots \sum_{k_m=0}^{3} \Psi_{k_1,\ldots,k_m} \prod_{i=1}^{m} x_i^{k_i}$$

Low-order mapping:

$$g^*(x_1, x_2, \ldots, x_m) = \sum_{k_1=0}^{1} \sum_{k_2=0}^{1} \cdots \sum_{k_m=0}^{1} \Omega_{k_1,\ldots,k_m} \prod_{i=1}^{m} x_i^{k_i}$$

Assumption: cross-terms involve at most τ parameters



ESTIMATION OF MEAN OUTPUT

- **Lemma 1:**

If the original system mapping $g(x_1, x_2, \dots, x_m)$ contains cross-terms of at most τ parameters

The low-order mapping $g^*(x_1, x_2, \dots, x_m)$ also contains cross-terms of at most τ parameters

- **Lemma 2:**

If $1 \le \boldsymbol{\tau} \le \left\lceil \frac{\boldsymbol{m}}{\boldsymbol{2}} \right\rceil - \boldsymbol{1}$

$2^{m-\gamma_{max}}$ OFFD can further reduce the number of simulations from 2^m to $2^{m-\gamma_{max}}$, where $\gamma_{max} = m - \left\lceil log_2\left(\sum_{i=0}^{\tau} \binom{i}{m}\right)\right\rceil$

- **Lemma 3:**

Constructed by selected simulation points to estimate the coefficients of the low-order mapping.

The **matrix $\boldsymbol{L \in R^{l_{offd}\times l}}$** constructe **rank**, and can be represented by $L = \mathrm{QU}$, where $Q \in R^{l_{offd}\times l}$ is an orthogonal matrix and $U \in R^{l\times l}$ is an upper triangular matrix.

Lemma 1

Lemma 2

Lemma 3

$$E[g(x_1, x_2, \dots, x_m)] = E[g^*(x_1, x_2, \dots, x_m)]$$

ROBUSTNESS TO NUMERICAL ERRORS

- **Problem Formulation**
  - The M-PCM-OFFD involves the calculation of L^{-1} or $(L^T L)^{-1} L^T$
  - L must be full column rank → Guaranteed using OFFD
  - Numerical errors may easily push L to lose rank and fail the computation
  - To facilitate the calculation and minimize the impact of such numerical error-induced disturbances, L needs to have a large margin to rank loss.

- **Metric: full-column-rank margin**
  - The full-column rank margin for matrix L to rank loss is

$$D(L) = min\{\|e\|_F \ | \ rank(L+e) < l\}$$

where $e \in R^{l_{offd} \times l}$ is a perturbation matrix

We proved that L matrix obtained using OFFD, denoted as L_{offd}, has the largest margin to rank loss, among all designs of the same size.

SIMULATION STUDY

- **Original Mapping:**

$$g(x_1,x_2,x_3) = x_1^3 + x_1^2 + x_1 + x_2^3 + x_2^2 + x_2 + x_3^3 + x_3^2 + x_3 + 1$$

$$x_1 \sim f_{X_1}(x_1) = 2e^{-2x_1};$$

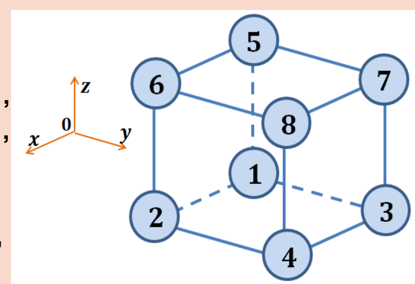
$$x_2 \sim f_{X_2}(x_2) = \frac{1}{15}, 5 \le x_2 \le 20;$$

$$x_3 \sim f_{X_3}(x_3) = \frac{1}{5}, 5 \le x_2 \le 10;$$

- **Illustration of Design Procedures**

Step 1: Choose 8 M-PCM points based on the pdf of each parameter

$p1 = (0.2929,8.1699,6.0566)$,
$p2 = (1.7071,8.1699,6.0566)$,
$p3 = (0.2929,16.8301,6.0566)$,
$p4 = (1.7071,16.8301,6.0566)$,
$p5 = (0.2929, 8.1699,8.9434)$,
$p6 = (1.7071,8.1699,8.9434)$,
$p7 = (0.2929,16.8301,8.9434)$,
$p8 = (1.7071,16.8301,8.9434)$



Step 2: Use 2_{III}^{3-1} OFFD to select 4 M-PCM points

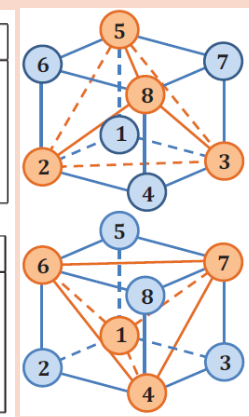
{***p2***, ***p3***, ***p5***, ***p8***}

	x_1	x_2	x_3	y
1	−	−	+	y_5
2	+	−	−	y_2
3	−	+	−	y_3
4	+	+	+	y_8

or

{***p1***, ***p4***, ***p6***, ***p7***}

	x_1	x_2	x_3	y
1	−	−	−	y_1
2	+	+	−	y_4
3	−	+	+	y_7
4	+	−	+	y_6



2_{III}^{3-1} OFFD design table

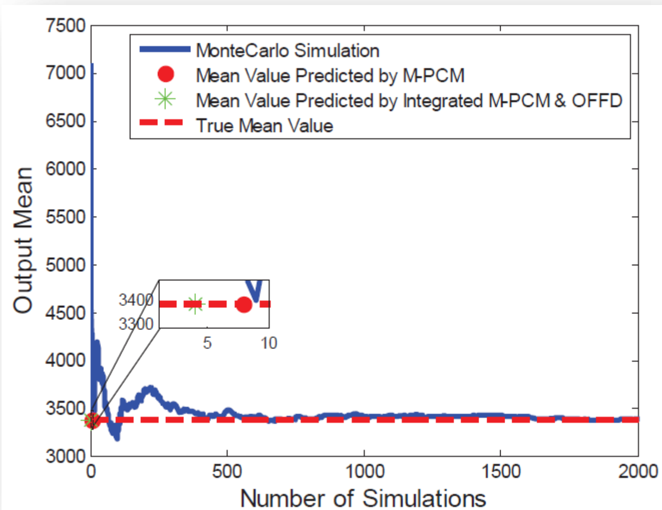
Step 3: Run simulations to evaluate $g(x_1, x_2, x_3)$ at these 4 M-PCM points

Step 4: Estimate the coefficients of the low-order mapping

$$g^*(x_1,x_2,x_3) = -4442.2 + 6.5x_1 + 513.5x_2 + 186.8x_3$$

SIMULATION STUDY (CONT.)

- **Illustration of Performance**



Estimation of Mean Output

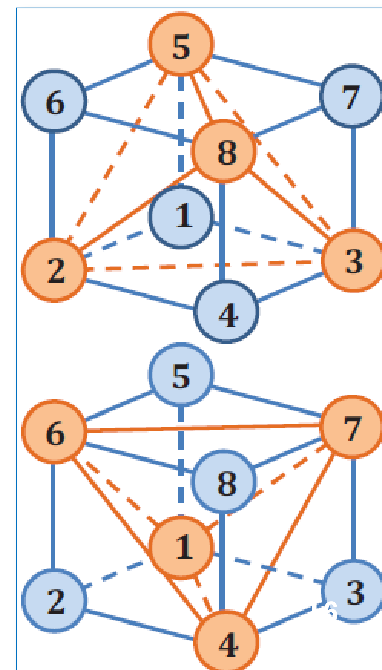
$$E[g(x_1, x_2, x_3)] = E[g^*(x_1, x_2, x_3)] = 3381.1$$

Robustness to Numerical Errors

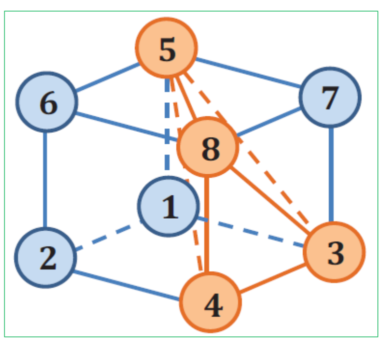
$$D(L_{offd}) = 1.4142 \quad D(L) = \{0, 0.866, 1.4142\}$$

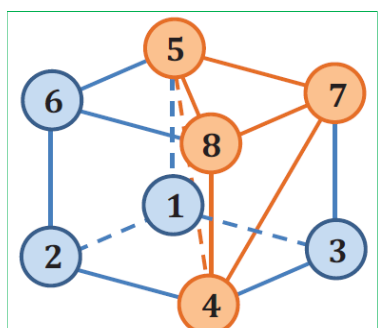
$$max(D(L)) = D(L_{offd})$$

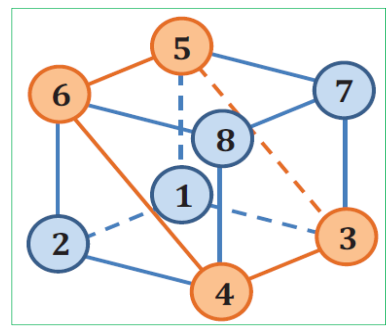
Selected by OFFD

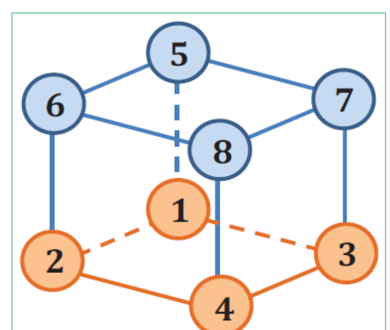


Other possible selections









CONCLUSION & FUTURE WORK

- An effective and scalable uncertainty evaluation method for large-scale complex systems

M-PCM-OFFD

- reduces the number of simulations from 2^{2m} to at most $2^{\lceil log_2(m+1)\rceil}$
- accurately predicts the output mean under broad assumptions
- is the most robust to numerical errors compared with designs of the same size

- New interpretations of the optimality of OFFDs
- **In the future work**
  - Generalize the degree of uncertain input parameters by exploring multiple-factor OFFDs
  - Exploit parameter dependency to further reduce the number of simulations required.

We thank National Institute of Standards and Technology for the support.