Uncertainty Quantification Tutorial for Computer Scientists

Michael S. Eldred

Optimization and Uncertainty Quantification Dept.

Sandia National Laboratories, Albuquerque, NM

Survey of core UQ methods:

- Sampling
- Local and global reliability
- Stochastic expansions: polynomial chaos, stochastic collocation
- Epistemic UQ

Overview of advanced topics (as time permits):

- Scalability: algorithmic and parallel
- Complexity: mixed uncertainties, multiphysics, multifidelity, …

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Sandia National Laboratories

Introduction

Many communities shifting from test-based to M&S-based approaches. For high consequence apps, we require confidence in the predictive accuracy of our M&S:

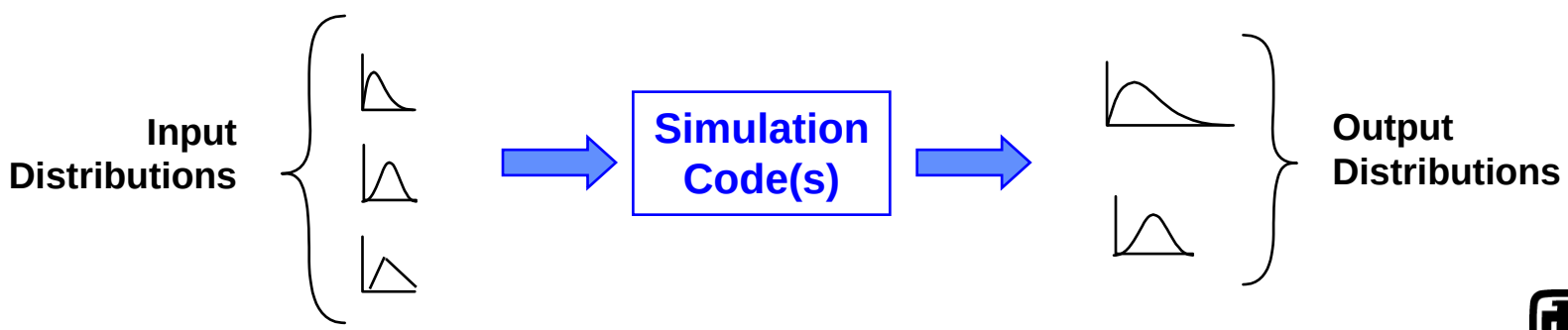
- ***Predictive simulations:*** **verified and validated for application of interest (“best estimate”)**
- ***Quantified uncertainties:*** **the effect of variation has been quantified (“plus uncertainty”)**

“Uncertainty quantification” generally includes (or is commonly integrated with):

- ***(Global) sensitivity analysis:*** **identification of input set w/ greatest influence on output QoIs**
- ***Uncertainty characterization:*** **fit or infer from observable data; parametric/non-parametric/KDE**
- ***Uncertainty propagation:*** **input distributions → output distributions**
- ***Decision making:*** **model validation, prediction ({interp,extrap}olation), design, …**

Uncertainty can be categorized to be one of two different types:

- ***Aleatory/irreducible:*** inherent variability with sufficient data → objective probabilistic models
- ***Epistemic/reducible:*** uncertainty from lack of knowledge → subjective probabilistic, nonprobabilistic



Introduction

Many communities shifting from test-based to M&S-based approaches. For high consequence apps, we require confidence in the predictive accuracy of our M&S:

- ***Predictive simulations:*** **verified and validated for application of interest (“best estimate”)**
- ***Quantified uncertainties:*** **the effect of variation has been quantified (“plus uncertainty”)**

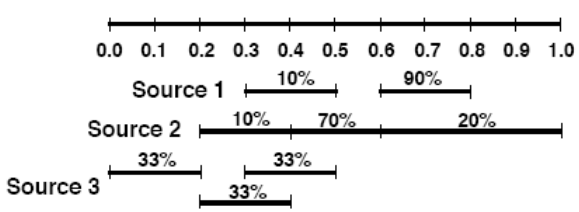
“Uncertainty quantification” generally includes (or is commonly integrated with):

- ***(Global) sensitivity analysis:*** **identification of input set w/ greatest influence on output QoIs**
- ***Uncertainty characterization:*** **fit or infer from observable data; parametric/non-parametric/KDE**
- ***Uncertainty propagation:*** **input distributions → output distributions**
- ***Decision making:*** **model validation, prediction ({interp,extrap}olation), design, …**

Uncertainty can be categorized to be one of two different types:

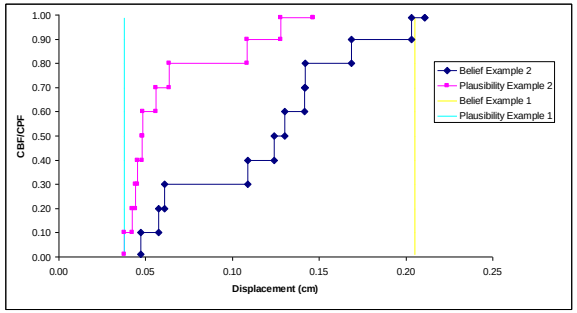
- ***Aleatory/irreducible:*** inherent variability with sufficient data → objective probabilistic models
- ***Epistemic/reducible:*** uncertainty from lack of knowledge → subjective probabilistic, nonprobabilistic

Input Intervals/BPAs



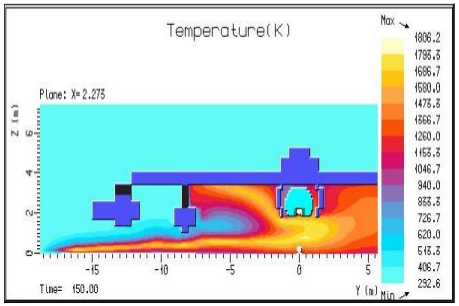
Simulation Code(s)

Output Intervals/Belief+Plausibility

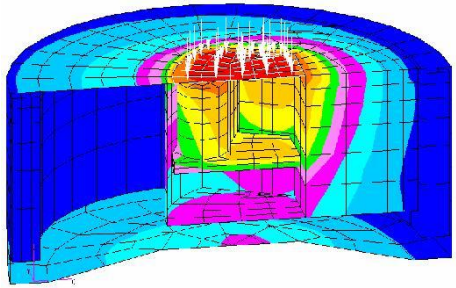


Example: Safety in Abnormal Environments

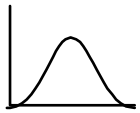
System level environment



Component level environment

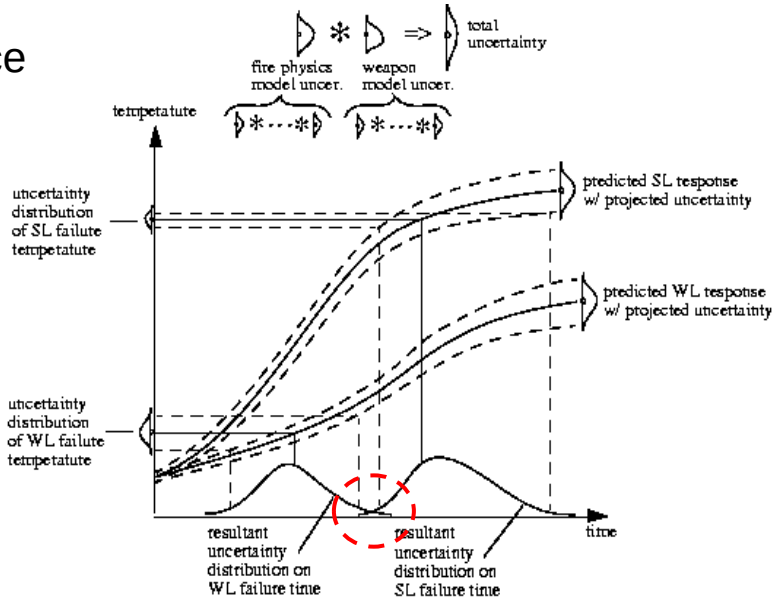


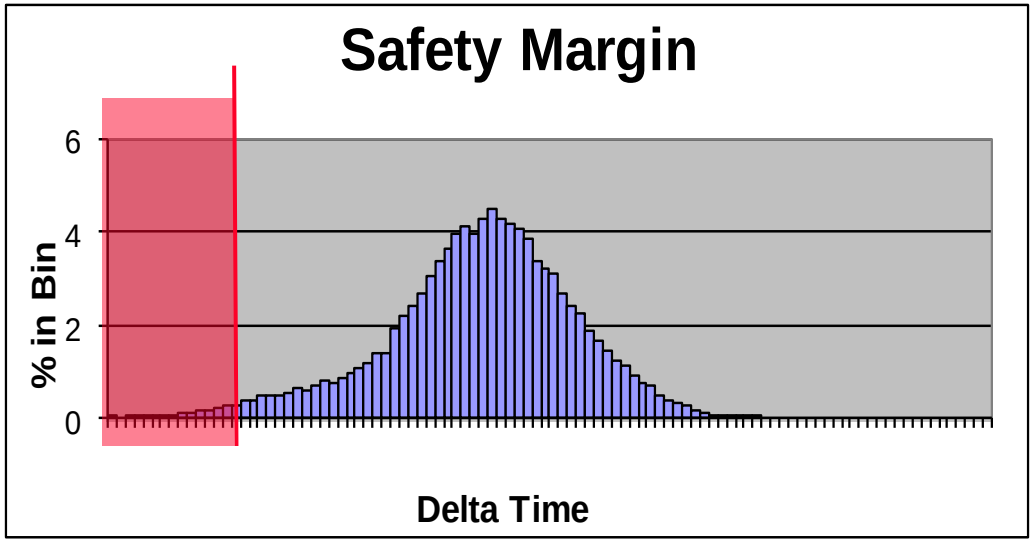
Uncertainties in B.C., matl. properties (density, conductivity), failure thesholds



Temperature time histories calculated by thermal code for B.C./matl. props

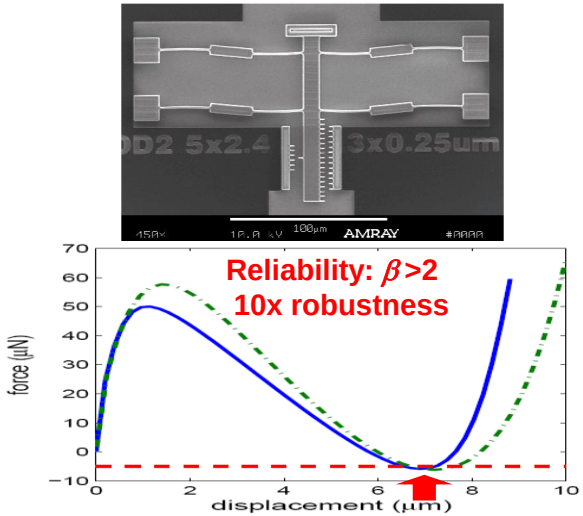
Thermal race



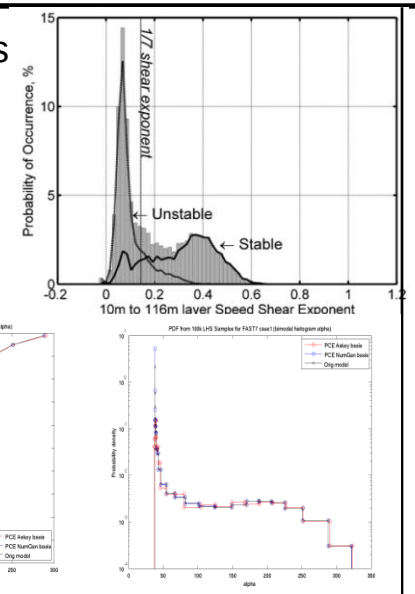


Additional Examples

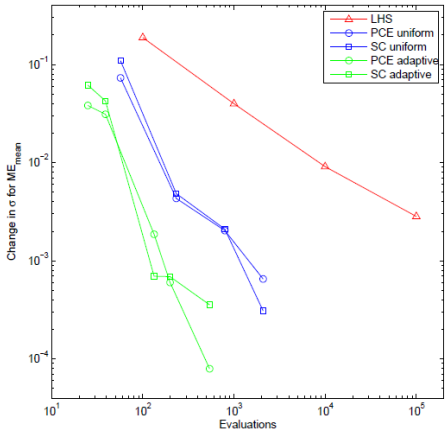
Stochastic design for bi-stable MEMS



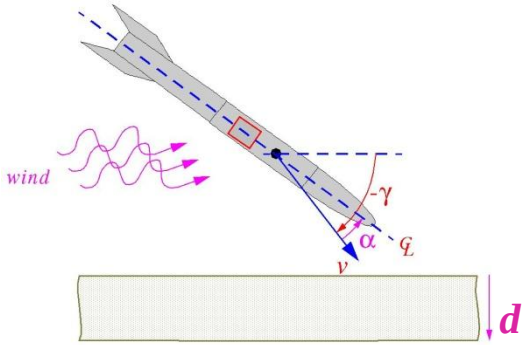
Wind turbines



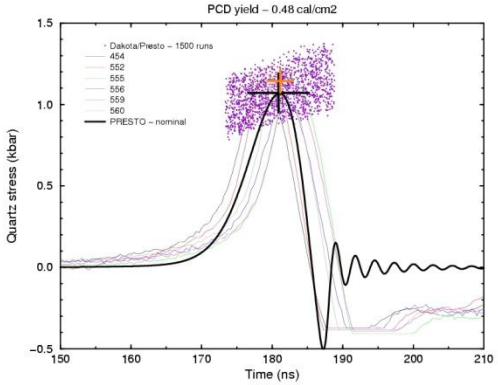
Nuclear reactors: conv. of CRUD statistics w/ adaptive UQ



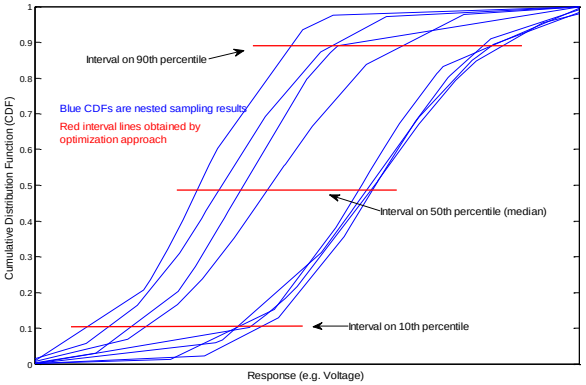
Penetrator performance



High energy density physics: Z-pinch model validation



Circuits/devices in radiation env.: Mixed aleatory-epistemic UQ



Connectivity of Exascale and UQ

UQ is a natural customer for exascale computing, both in terms of current mission needs and algorithmic opportunities

- ***Nonintrusive methods:*** coarse-grained, fully asynchronous w/ local refinement, sim. fault tolerant
- ***Embedded methods:*** rearrange stoch Galerkin for an outer-spatial/inner-stochastic ordering

DOE Program Context:

- **NNSA ASCI/ASC has invested significantly in predictive science since 1995**
  - Perform stockpile stewardship using advanced M&S on HPC with limited/no future stockpile testing
  - LANL, LLNL, SNL each have significant V&V programs within ASC, since ~2000
  - Academic alliances; current cycle focused on predictive science/V&V/UQ (2008-12); PSAAP II (2013-2017) in proposal phase; marries predictive science/V&V/UQ + exascale
- **ASCR has recently funded a program focused on UQ (~FY11-FY13):**
  - Alonso/Eldred/Iaccarino: Wind turbine UQ
  - Efendiev/Vassilevski: Bayesian UQ
  - Lin/Kevrekidis/Zabaras: Stochastic Multiscale
  - Shadid/Estep: Adjoint UQ/EE
  - Tong/Lee/Iaccarino: Multiphysics UQ
  - Xiu/Archibald/Deiterding: Adaptive Parallel UQ
- **SciDAC-3: math/CS enabling technology (~FY12-16) and application institutes (~FY13-17)**
  - UQ-centric enabling technologies center (QUEST; SNL/LANL/UT/USC/MIT/JHU, PI: Najm)
  - Application institute partnerships currently being defined
- **Exascale codesign**

Uncertainty Quantification Algorithms @ SNL: New methods bridge robustness/efficiency gap

	Production	New	Under dev.	Planned	Collabs.
Sampling	Latin Hypercube, Monte Carlo	Importance, Incremental		Bootstrap, Jackknife	FSU
Reliability	*Local:* Mean Value, First-order & second-order reliability methods (FORM, SORM)	*Global:* Efficient global reliability analysis (EGRA)	gradient-enhanced	recursive emulation, TGP	*Local:* Notre Dame, *Global:* Vanderbilt
Stochastic expansion		PCE and SC with uniform & dimension-adaptive p-/h-refinement	local h-refinement, gradient-enhanced	hp-adaptive, discrete, multi-physics	Stanford, Purdue, FSU
Other probabilistic		Random fields/ stochastic proc.		Dimension reduction	Cornell, Maryland
Epistemic	Interval-valued/ Second-order prob. (nested sampling)	Opt-based interval estimation, Dempster-Shafer	Bayesian	Imprecise probability	LANL, UT Austin
Metrics & Global SA	Importance factors, Partial correlations	Main effects, Variance-based decomposition	Stepwise regression		LANL

Uncertainty Quantification Algorithms @ SNL: New methods bridge robustness/efficiency gap

	Production	New	Under dev.	Planned	Collabs.
Sampling	Latin Hypercube, Monte Carlo	Importance, Incremental		Bootstrap, Jackknife	FSU
Reliability	*Local:* Mean Value, First-order & second-order reliability methods (FORM, SORM)	*Global:* Efficient global reliability analysis (EGRA)	gradient-enhanced	recursive emulation, TGP	*Local:* Notre Dame, *Global:* Vanderbilt
Stochastic expansion		PCE and SC with uniform & dimension-adaptive p-/h-refinement	local h-refinement, gradient-enhanced	hp-adaptive, discrete, multi-physics	Stanford, Purdue, FSU
Other probabilistic		Random fields/ stochastic proc.		Dimension reduction	Cornell, Maryland
Epistemic	Interval-valued/ Second-order prob. (nested sampling)	Opt-based interval estimation, Dempster-Shafer	Bayesian	Imprecise probability	LANL, UT Austin
Metrics & Global SA	Importance factors, Partial correlations	Main effects, Variance-based decomposition	Stepwise regression		LANL

Research: Adaptive Refinement, Gradient Enhancement

Adv. Deployment

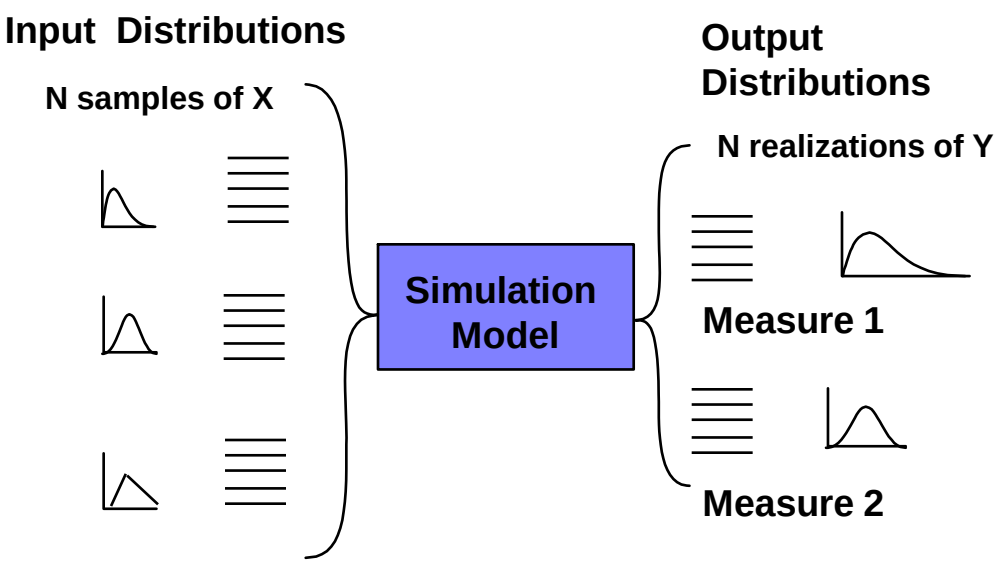
Fills Gaps

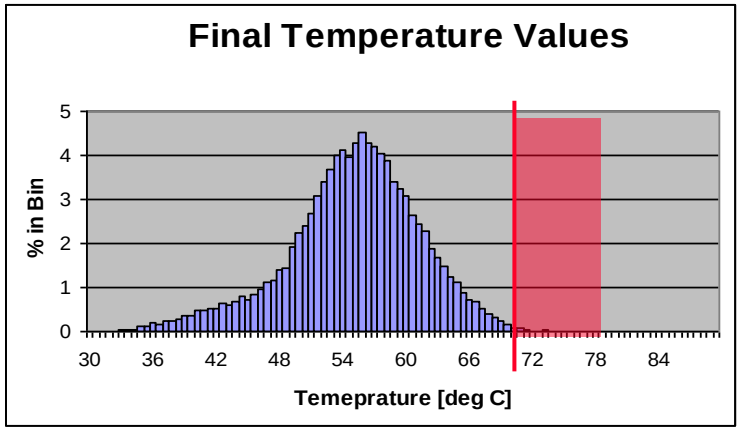
Sampling

Starting from distributions on the uncertain input values, draw observations from each distribution, pair samples, and execute the model for each pairing
→ ensemble of results yields distributions of the outputs

- Monte Carlo: **basic random sampling**
- Pseudo Monte Carlo: **Latin Hypercube Sampling (LHS)**
- Quasi Monte Carlo**:** **Halton, Hammersley, Sobol sequences**
- Orthogonal arrays, Centroidal Voronoi Tesselation (CVT), Importance Sampling

Sampling is not the most efficient UQ method, but is easy to implement, robust, and transparent.



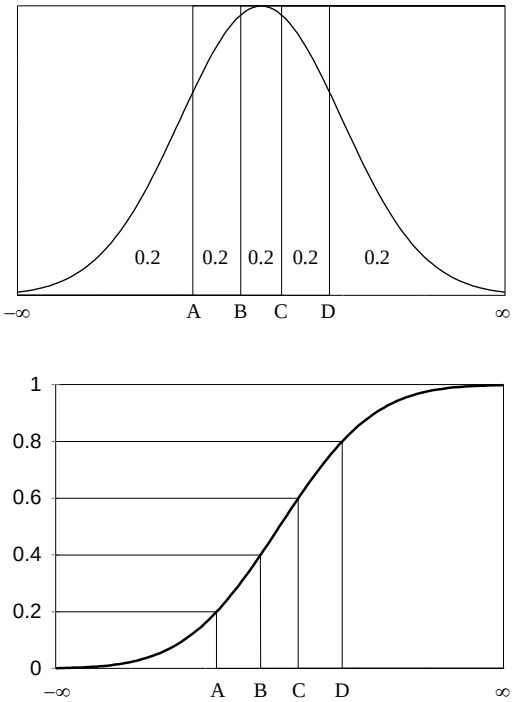


Goals: assess mean, variance, tail statistics, probability of failure

Sandia National Laboratories

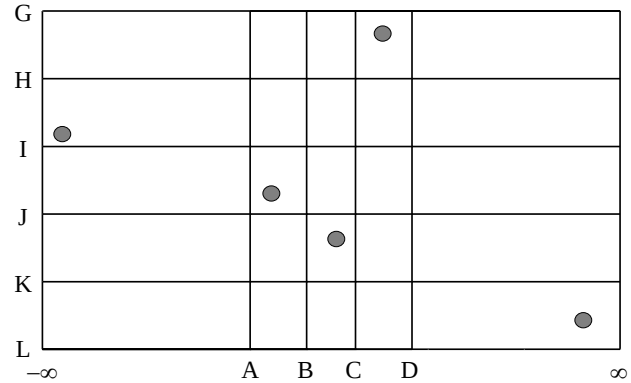
Latin Hypercube Sampling

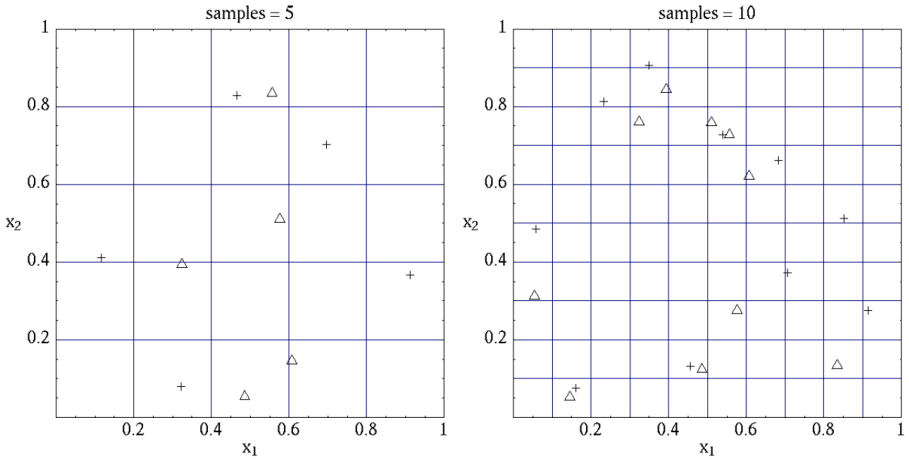
- **LHS is stratified random sampling among equal probability bins for all 1-D projections of an n-dimensional set of samples.**
  - **Early work by McKay and Conover**
  - **Restricted pairing by Iman → enforce prescribed input correlations**



Intervals Used with a LHS of Size N = 5 in Terms of the PDF and CDF for a Normal Random Variable

A possible LHS for n=2, N=5 with X1 = normal and X2 = uniform





Sandia National Laboratories

Reliability Methods for UQ

UQ with Reliability Methods

Mean Value Method

$$\mu_g = g(\mu_{\mathbf{x}})$$

$$\sigma_g^2 = \sum_i \sum_j Cov(i,j) \frac{dg}{dx_i}(\mu_{\mathbf{x}}) \frac{dg}{dx_j}(\mu_{\mathbf{x}})$$

$$\bar{z} \rightarrow p, \beta \begin{cases} \beta_{cdf} = \dfrac{\mu_g - \bar{z}}{\sigma_g} \\ \beta_{ccdf} = \dfrac{\bar{z} - \mu_g}{\sigma_g} \end{cases} \qquad \bar{p}, \bar{\beta} \rightarrow z \begin{cases} z = \mu_g - \sigma_g \bar{\beta}_{cdf} \\ z = \mu_g + \sigma_g \bar{\beta}_{ccdf} \end{cases}$$

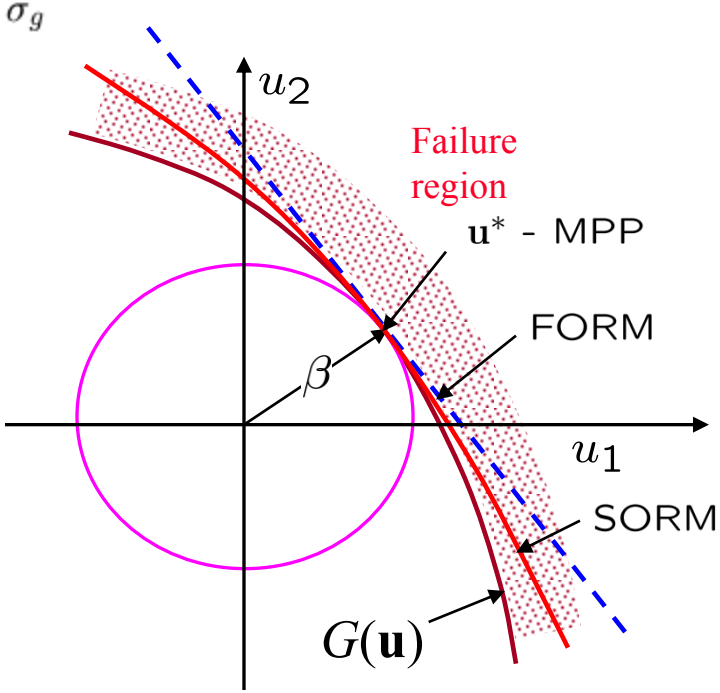
Rough statistics

MPP search methods

Reliability Index Approach (RIA)

$$\begin{aligned} \text{minimize} \quad & \mathbf{u}^T \mathbf{u} \\ \text{subject to} \quad & G(\mathbf{u}) = \bar{z} \end{aligned}$$

Find min dist to *G* level curve
Used for fwd map *z* → *p*/*β*



Performance Measure Approach (PMA)

$$\begin{aligned} \text{minimize} \quad & \pm G(\mathbf{u}) \\ \text{subject to} \quad & \mathbf{u}^T \mathbf{u} = \bar{\beta}^2 \end{aligned}$$

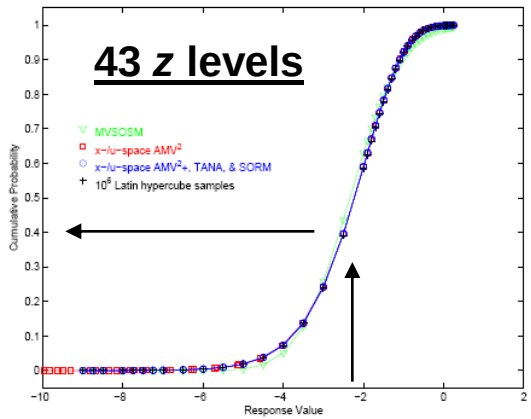
Find min *G* at *β* radius
Used for inv map *p*/*β* → *z*

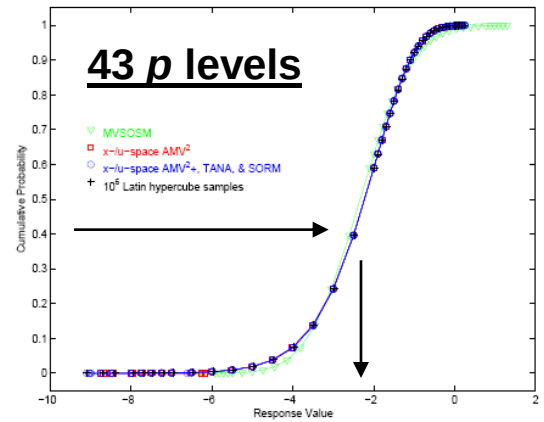
Nataf x → u: $\Phi(z_i) = F(x_i)$
$\mathbf{z} = \mathbf{L}\mathbf{u}$

Sandia National Laboratories

Reliability Algorithm Variations: Algorithm Performance Results

Analytic benchmark test problems: lognormal ratio, short column, cantilever





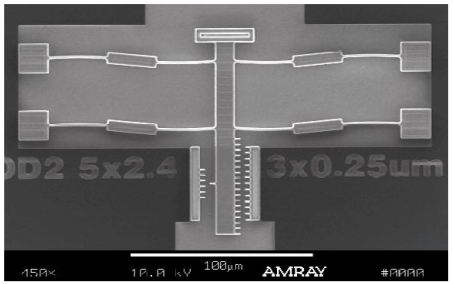
RIA Approach	SQP Function Evaluations	NIP Function Evaluations	CDF p Error Norm	Target z Offset Norm
MVFOSM	1	1	0.1548	0.0
MVSOSM	1	1	0.1127	0.0
x-space AMV	45	45	0.009275	18.28
u-space AMV	45	45	0.006408	18.81
x-space AMV^2	45	45	0.002063	2.482
u-space AMV^2	45	45	0.001410	2.031
x-space AMV+	192	192	0.0	0.0
u-space AMV+	207	207	0.0	0.0
x-space AMV^2+	125	131	0.0	0.0
u-space AMV^2+	122	130	0.0	0.0
x-space TANA	245	246	0.0	0.0
u-space TANA	296*	278*	6.982e-5	0.08014
FORM	626	176	0.0	0.0
SORM	669	219	0.0	0.0

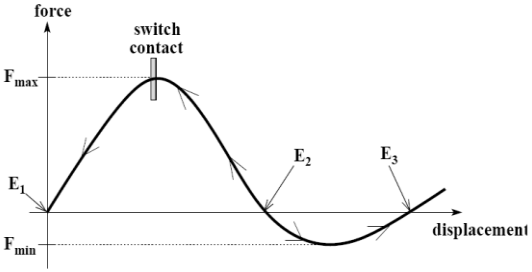
PMA Approach	SQP Function Evaluations	NIP Function Evaluations	CDF z Error Norm	Target p Offset Norm
MVFOSM	1	1	7.454	0.0
MVSOSM	1	1	6.823	0.0
x-space AMV	45	45	0.9420	0.0
u-space AMV	45	45	0.5828	0.0
x-space AMV^2	45	45	2.730	0.0
u-space AMV^2	45	45	2.828	0.0
x-space AMV+	171	179	0.0	0.0
u-space AMV+	205	205	0.0	0.0
x-space AMV^2+	135	142	0.0	0.0
u-space AMV^2+	132	139	0.0	0.0
x-space TANA	293*	272	0.04259	1.598e-4
u-space TANA	325*	311*	2.208	5.600e-4
FORM	720	192	0.0	0.0
SORM	535	191*	2.410	6.522e-4

Note: 2nd-order PMA with prescribed *p* level is harder problem → requires $\beta(p)$ update/inversion

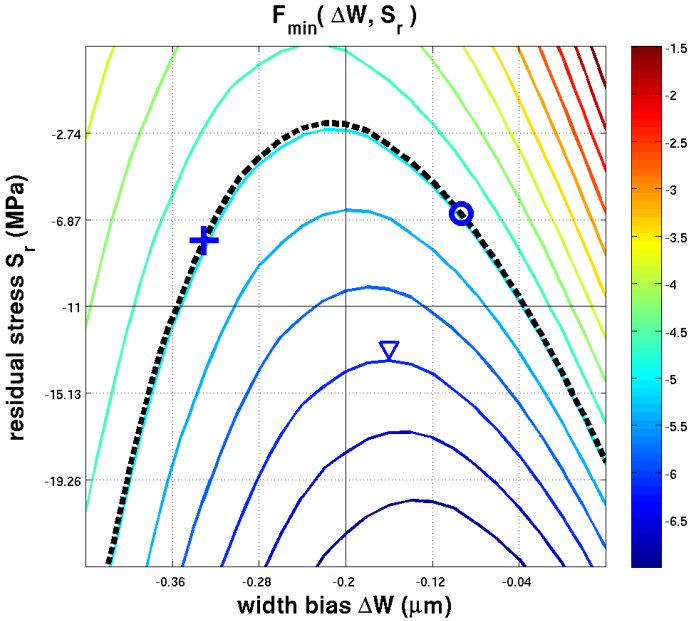
Solution-Verified Reliability Analysis and Design of MEMS

- **Problem:** MEMS subject to substantial variabilities
  - Material properties, manufactured geometry, residual stresses
  - Part yields can be low or have poor durability
  - Data can be obtained → aleatory UQ → probabilistic methods
- **Goal:** account for both uncertainties and errors in design
  - Integrate UQ/OUU (DAKOTA), ZZ/QOI error estimation (Encore), adaptivity (SIERRA), nonlin mech (Aria) → MESA application
  - Perform soln verification in automated, parameter-adaptive way
  - Generate fully converged UQ/OUU results at lower cost





- AMV²+ and FORM converge to different MPPs (+ and O respectively)
- Issue: high nonlinearity leading to multiple legitimate MPP solns.
- Challenge: design optimization may tend to seek out regions encircled by the failure domain. 1st-order and even 2nd-order probability integrations can experience difficulty with this degree of nonlinearity. Optimizers can/will exploit this model weakness.

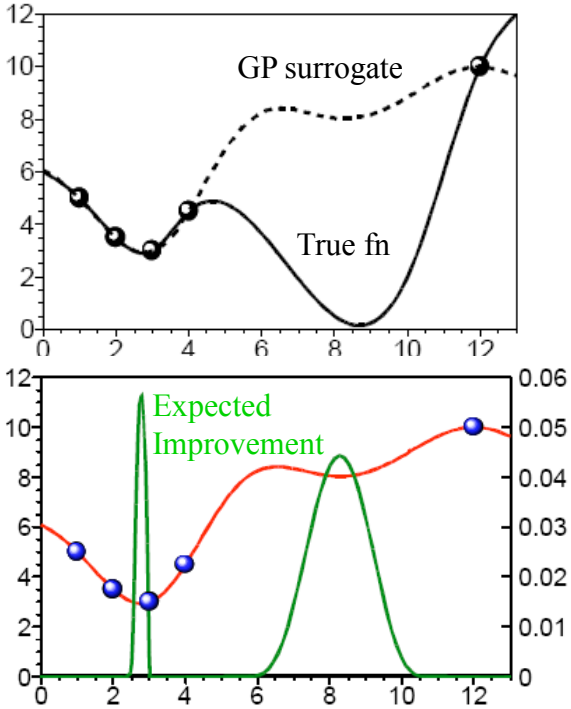


Parameter study over 3σ uncertain variable range for fixed design variables d_M*. ***Dashed black line denotes g(x) = $F_{min}(x)$ = -5.0.***

Sandia National Laboratories

Efficient Global Reliability Analysis (EGRA)

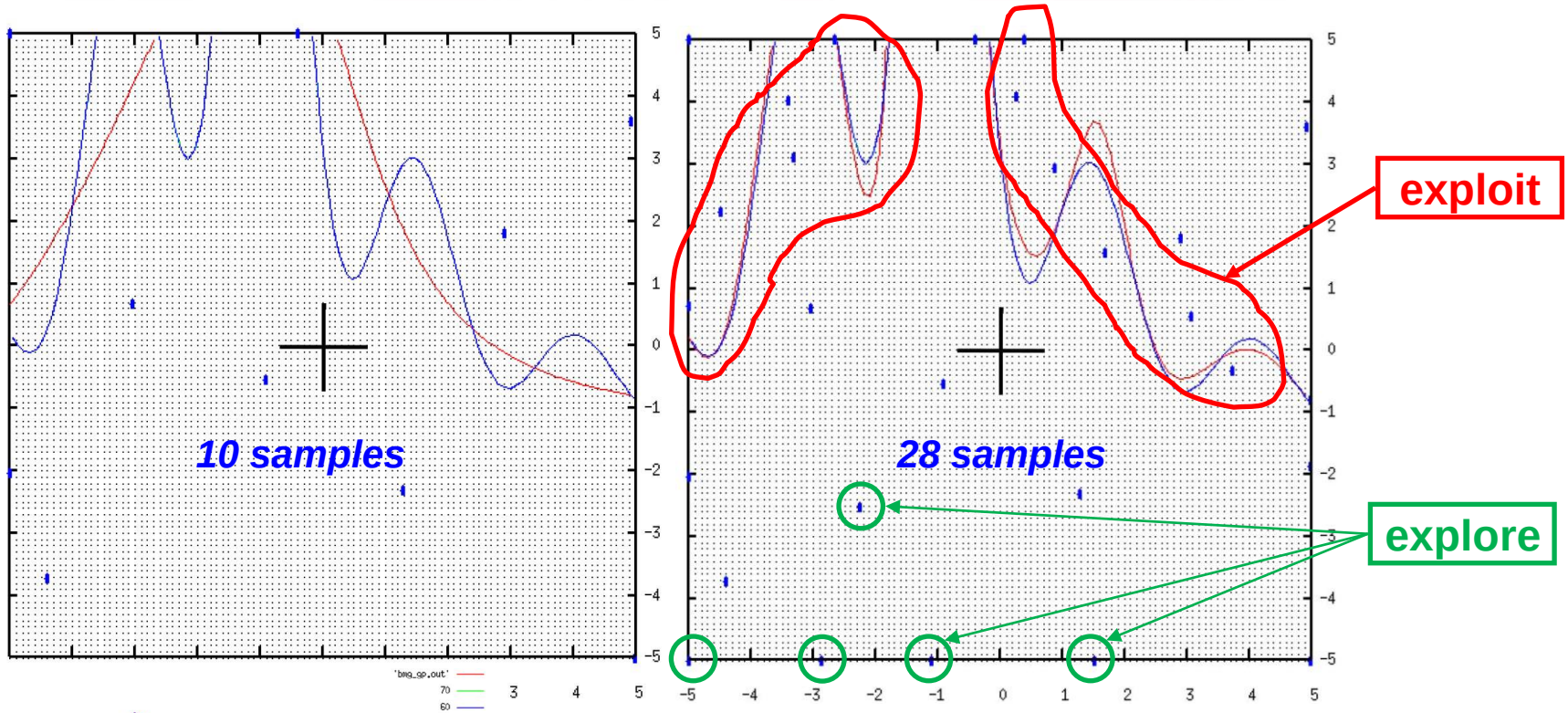
- **Address known failure modes of local reliability methods:**
  - Nonsmooth: fail to converge to an MPP
  - Multimodal: only locate one of several MPPs
  - Highly nonlinear: low order limit state approxs. fail to accurately estimate probability at MPP
- **Based on EGO (surrogate-based global opt.), which exploits special features of GPs**
  - Mean and variance predictions: formulate expected improvement (EGO) or expected feasibility (EGRA)
  - Balance explore and exploit in computing an optimum (EGO) or locating the limit state (EGRA)

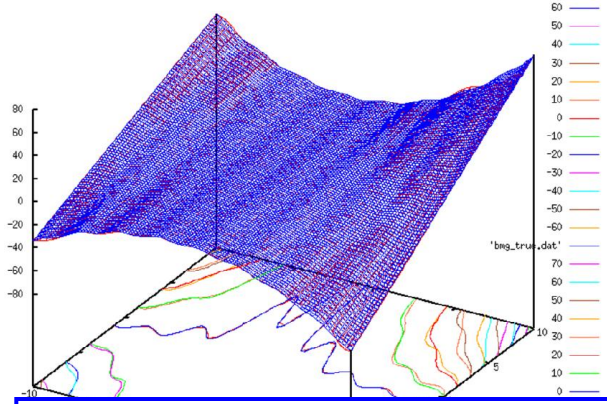


From Jones, Schonlau, Welch, 1998

Sandia National Laboratories

Efficient Global Reliability Analysis





Reliability Method	Function Evaluations	First-Order p_f (% Error)	Second-Order p_f (% Error)	Sampling p_f (% Error, Avg. Error)
No Approximation	70	0.11797 (277.0%)	0.02516 (-19.6%)	—
x-space AMV^2+	26	0.11797 (277.0%)	0.02516 (-19.6%)	—
u-space AMV^2+	26	0.11777 (277.0%)	0.02516 (-19.6%)	—
u-space TANA	131	0.11797 (277.0%)	0.02516 (-19.6%)	—
LHS solution	10k	—	—	0.03117 (0.385%, 2.847%)
LHS solution	100k	—	—	0.03126 (0.085%, 1.397%)
LHS solution	1M	—	—	0.03129 (truth , 0.339%)
x-space EGRA	35.1	—	—	0.03134 (0.155%, 0.433%)
u-space EGRA	35.2	—	—	0.03133 (0.136%, 0.296%)

Accuracy similar to exhaustive sampling at cost similar to local reliability assessment

Stochastic Expansion Methods for UQ

Polynomial Chaos Expansions (PCE)

Approximate response w/ spectral proj. using orthogonal polynomial basis fns

i.e. $$R = \sum_{j=0}^{P} \alpha_j \Psi_j(\boldsymbol{\xi})$$ using

$$\begin{aligned}
\Psi_0(\boldsymbol{\xi}) &= \psi_0(\xi_1)\,\psi_0(\xi_2) &&= 1\\
\Psi_1(\boldsymbol{\xi}) &= \psi_1(\xi_1)\,\psi_0(\xi_2) &&= \xi_1\\
\Psi_2(\boldsymbol{\xi}) &= \psi_0(\xi_1)\,\psi_1(\xi_2) &&= \xi_2\\
\Psi_3(\boldsymbol{\xi}) &= \psi_2(\xi_1)\,\psi_0(\xi_2) &&= \xi_1^2 - 1\\
\Psi_4(\boldsymbol{\xi}) &= \psi_1(\xi_1)\,\psi_1(\xi_2) &&= \xi_1\xi_2\\
\Psi_5(\boldsymbol{\xi}) &= \psi_0(\xi_1)\,\psi_2(\xi_2) &&= \xi_2^2 - 1
\end{aligned}$$

$$\alpha_j = \frac{\langle R, \Psi_j \rangle}{\langle \Psi_j^2 \rangle} = \frac{1}{\langle \Psi_j^2 \rangle} \int_\Omega R\,\Psi_j\,\varrho(\boldsymbol{\xi})\,d\boldsymbol{\xi}$$

$$\langle \Psi_j^2 \rangle = \prod_{i=1}^{n} \langle \psi_{m_i^j}^2 \rangle$$

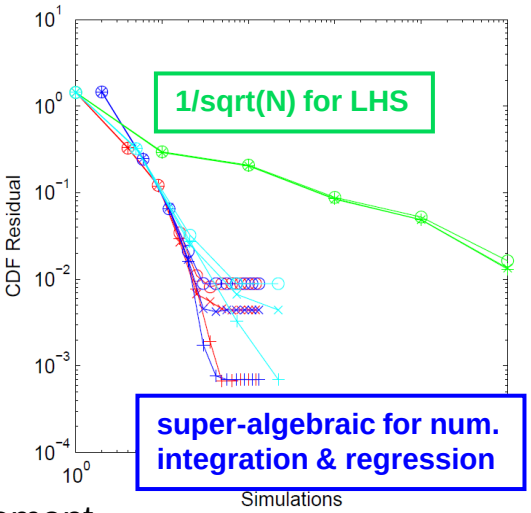
- Nonintrusive: estimate α_j using sampling, regression, tensor-product quadrature, sparse grids, or cubature

Generalized PCE (Wiener-Askey + numerically-generated)

- Tailor basis: selection of basis orthogonal to input PDF avoids additional nonlinearity

Distribution	Density function	Polynomial	Weight function	Support range
Normal	$\frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$	Hermite $He_n(x)$	$e^{\frac{-x^2}{2}}$	$[-\infty, \infty]$
Uniform	$\frac{1}{2}$	Legendre $P_n(x)$	1	$[-1, 1]$
Beta	$\frac{(1-x)^\alpha (1+x)^\beta}{2^{\alpha+\beta+1} B(\alpha+1, \beta+1)}$	Jacobi $P_n^{(\alpha,\beta)}(x)$	$(1-x)^\alpha (1+x)^\beta$	$[-1, 1]$
Exponential	e^{-x}	Laguerre $L_n(x)$	e^{-x}	$[0, \infty]$
Gamma	$\frac{x^\alpha e^{-x}}{\Gamma(\alpha+1)}$	Generalized Laguerre $L_n^{(\alpha)}(x)$	$x^\alpha e^{-x}$	$[0, \infty]$

Additional bases generated numerically (discretized Stieltjes + Golub-Welsch)



- Tailor expansion form:
  - Dimension p-refinement: anisotropic TPQ/SSG, generalized SSG
  - Dimension & region h-refinement: local bases with global & local refinement

Stochastic Collocation (SC)

$$R = \sum_{j=1}^{N_p} r_j \boldsymbol{L}_j(\boldsymbol{\xi})$$

Instead of estimating coefficients for known basis functions, form <u>interpolants</u> for known coefficients

- ***Global:*** Lagrange (values) or Hermite (values+derivatives)
- ***Local:*** linear (values) or cubic (values+gradients) splines

$$L_i = \prod_{\substack{j=1 \\ j \neq i}}^{m} \frac{x - x_j}{x_i - x_j}$$

$$R(\boldsymbol{\xi}) \cong \sum_{j_1=1}^{m_{i_1}} \cdots \sum_{j_n=1}^{m_{i_n}} r\left(\xi_{j_1}^{i_1}, \ldots, \xi_{j_n}^{i_n}\right) \left(L_{j_1}^{i_1} \otimes \cdots \otimes L_{j_n}^{i_n}\right)$$

Sparse interpolants formed using Σ of tensor interpolants

Advantages relative to PCE:

- Somewhat simpler (no expansion order to manage separately)
- Often less expensive (no integration for coefficients)
- Expansion only formed for sampling → probabilities (estimating moments of any order is straightforward)
- Adaptive h-refinement with hierarchical surpluses; explicit gradient-enhancement

Disadvantages relative to PCE:

- Less flexible/fault tolerant → structured data sets (tensor/sparse grids)
- Expansion variance not guaranteed positive (important in opt./interval est.)
- No direct inference of spectral decay rates

With sufficient care on PCE form, PCE/SC performance is essentially identical for many cases of interest (tensor/sparse grids with standard Gauss rules)

Approaches for forming PCE/SC Expansions

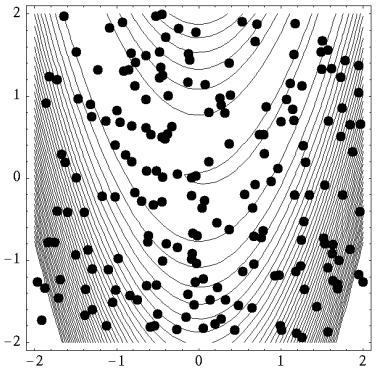
Random sampling: PCE

Expectation (sampling):

- Sample w/i distribution of ξ
- Compute expected value of product of R and each Ψ_j

Linear regression ("point collocation"):

- Sample w/i distribution of ξ
- Solves least squares data fit for all coefficients at once:



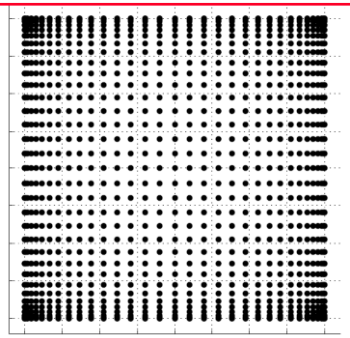
$$\Psi\alpha = R$$

Tensor-product quadrature: PCE/SC

$$\mathscr{U}^i(f)(\xi) = \sum_{j=1}^{m_i} f(\xi_j^i)\, w_j^i$$

$$\mathcal{Q}_{\mathbf{i}}^n f(\xi) = \left(\mathscr{U}^{i_1} \otimes \cdots \otimes \mathscr{U}^{i_n}\right)(f)(\boldsymbol{\xi}) = \sum_{j_1=1}^{m_{i_1}} \cdots \sum_{j_n=1}^{m_{i_n}} f\left(\xi_{j_1}^{i_1}, \ldots, \xi_{j_n}^{i_n}\right)\left(w_{j_1}^{i_1} \otimes \cdots \otimes w_{j_n}^{i_n}\right)$$

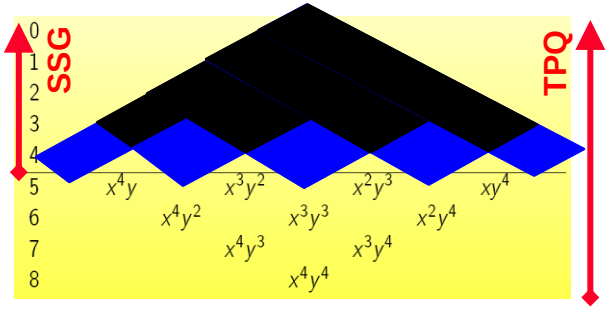
- **Every combination of 1-D rules**
- **Scales as m^n**
- **1-D Gaussian rule of order m → integrands to order $2m - 1$**
- **Assuming $R\,\Psi_j$ of order $2p$, select $m = p + 1$**

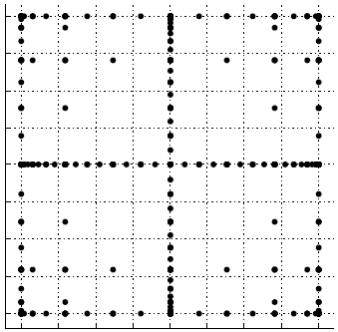


Smolyak Sparse Grid: PCE/SC

$$\mathscr{A}(\mathrm{w}, n) = \sum_{\mathrm{w}+1 \le |\mathbf{i}| \le \mathrm{w}+n} (-1)^{\mathrm{w}+n-|\mathbf{i}|} \binom{n-1}{\mathrm{w}+n-|\mathbf{i}|} \cdot \left(\mathscr{U}^{i_1} \otimes \cdots \otimes \mathscr{U}^{i_n}\right)$$

Pascal's triangle (2D):



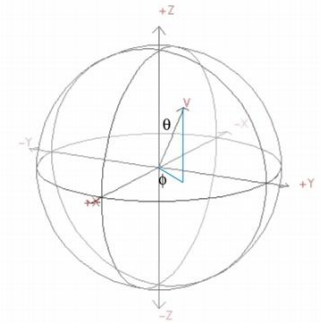


Cubature: PCE

Stroud and extensions (Xiu, Cools)

→ Low order PCE

→ global SA, anisotropy detection



Gaussian $i = 2$ → $p = 1$

$$x_{k,2r-1} = \sqrt{2}\cos\frac{2rk\pi}{n+1}, \qquad x_{k,2r} = \sqrt{2}\sin\frac{2rk\pi}{n+1}$$

Arbitrary PDF

$$t^{(k)} = \frac{1}{\gamma}\left[\sqrt{\gamma c_1}\, x^{(k)} - \delta\right]$$

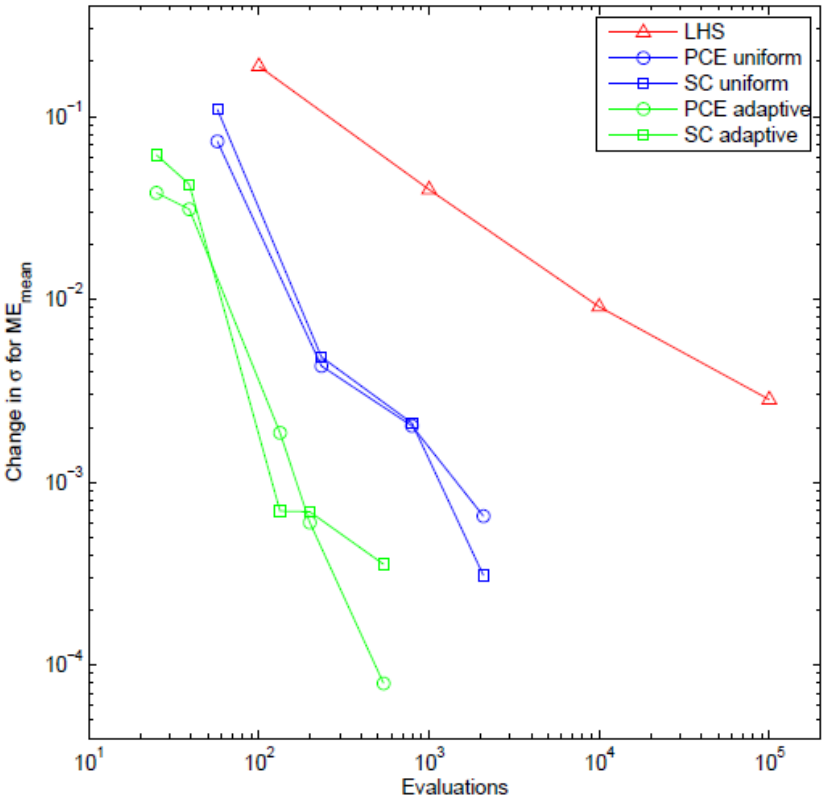
Sandia National Laboratories

Application Deployment (CASL)

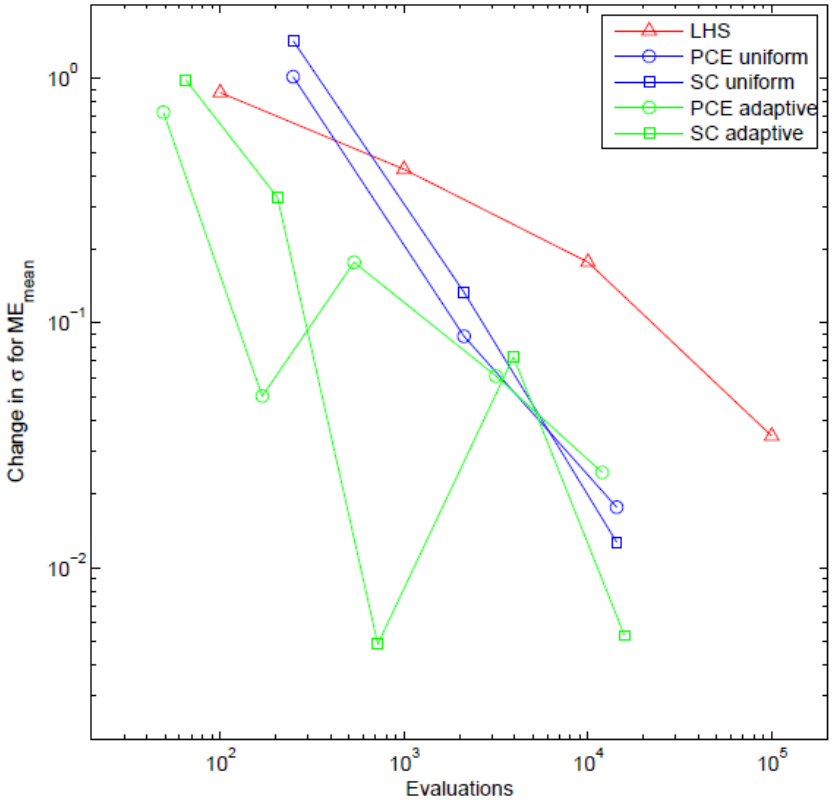
Application: Nuclear reactor cores experience localized boiling, which leads to CRUD (Chalk River Unidentified Deposit). These deposits result in undesirable power shifts (CIPS) within the core. Statistics of mass evaporation (ME) rate are of interest.

Methodology: PCE/SC with uniform/adaptive refinement compared to LHS

Plant A (n=4, smooth, mild anisotropy)



Plant B (n=10, discontinuous, high anisotropy)

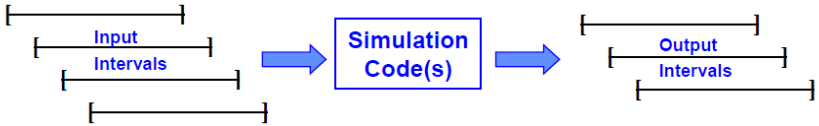


Epistemic UQ

Epistemic UQ: one does not know enough to specify probability distributions

Sometimes referred to as subjective, reducible, or lack of knowledge uncertainty

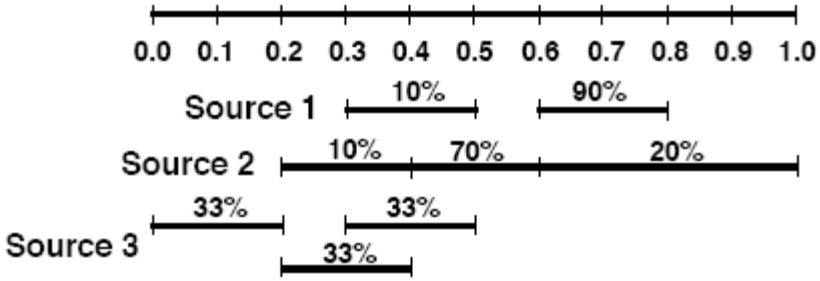
Interval analysis

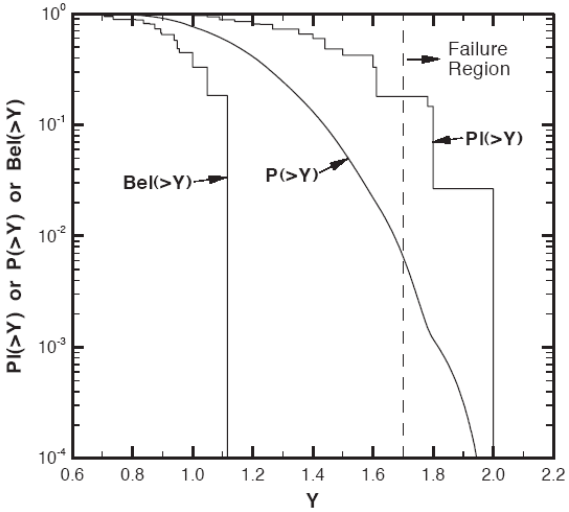


- Propagate input intervals to output intervals
- Intrusive interval methods (operation by operation propagation) have been investigated for several decades, but have not become mainstream (key issue: interval growth)
- Sampling methods (+ surrogate models if expensive evals) are commonly used
- Optimization methods are promising and some variants exploit data reuse

Dempster-Shafer theory of evidence

- Basic probability assignment (interval-based)
- Solve opt. problems (currently sampling-based) to compute belief/plausibility for output intervals





Imprecise probability (p-boxes), Info gap, …

Survey of Core UQ Algorithms:
strengths, weaknesses, research needs

Sampling (nongradient-based)

- ***Strengths:*** Simple and reliable, convergence rate is dimension-independent
- ***Weaknesses:*** $N^{-1/2}$ convergence → expensive for accurate tail statistics

Local reliability (gradient-based)

- ***Strengths:*** computationally efficient, widely used, scalable to large *n* (w/ efficient derivs.)
- ***Weaknesses:*** algorithmic failures for limit states with following features
  - Nonsmooth: fail to converge to an MPP
  - Multimodal: only locate one of several MPPs
  - Highly nonlinear: low order limit state approxs. insufficient to resolve probability at MPP

Global reliability (typically nongradient-based)

- ***Strengths:*** handles multimodal and/or highly nonlinear limit states
- ***Weaknesses:***
  - Conditioning, nonsmoothness → ensemble emulation (recursion, discretization)
  - Scaling to large *n* → adjoint gradient-enhancement, additional refinement bias

Stochastic expansions (typically nongradient-based)

- ***Strengths:*** functional representation, exponential convergence rates for smooth problems
- ***Weaknesses:***
  - Nonsmoothness → local h-refinement based on hierarchical error estimates
  - Scaling to large *n* → adaptive refinement, adjoint gradient-enhancement

Epistemic methods (typically nongradient-based)

- ***Strengths:*** extrema are point solutions instead of integrated quantities
- ***Weaknesses:*** high degrees of input structure (Dempster-Shafer) require *many* extrema (bridging intervals and distributions breaks down as continuum is approached discretely)

Advanced Topics

R&D Drivers:

Efficient/robust/scalable foundation of core UQ algorithms

- Adaptive refinement
- Adjoint enhancement
- Multilevel parallel UQ (circa 2000) → Exascale UQ efforts (intrusive, non-intrusive)

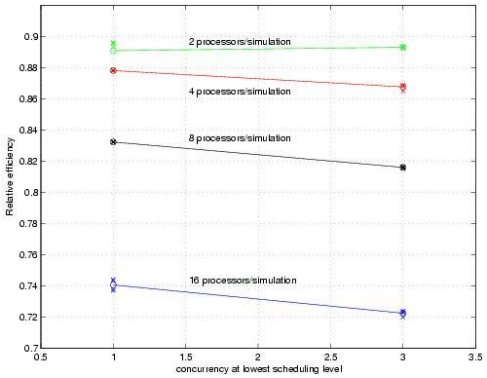
Build on foundation to address more complex random environments

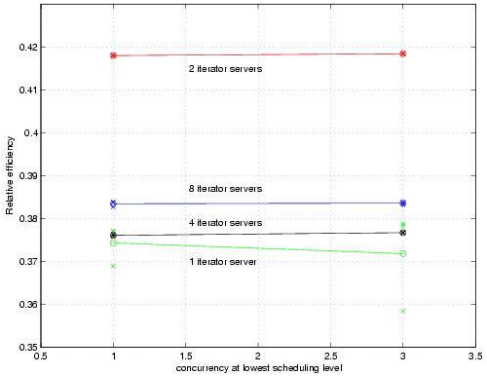
- Mixed aleatory-epistemic UQ
- Multiphysics UQ
- Multifidelity UQ

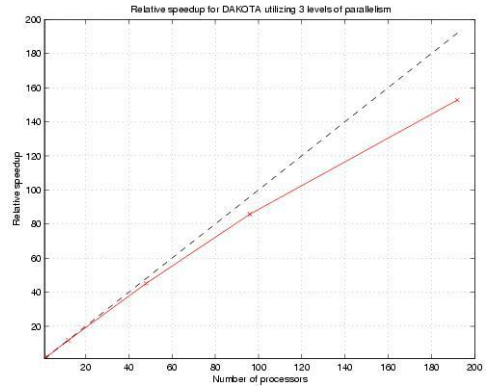
High-Performance Computing
Exploiting parallelism in Opt., UQ, …

1. ***Algorithmic coarse-grained***: **concurrency in data requests:**
   - **Iterators: Gradient-based, Nongradient-based, Surrogate-based**
   - **Strategies with concurrent Iterators: Multi-start, Pareto, Hybrid**
   - **Nested Models: OUU/MCUU, Mixed UQ**
2. ***Algorithmic fine-grained:*** **computing the internal linear algebra of an opt. algorithm in parallel (e.g., large-scale opt., SAND)**
3. ***Fn eval coarse-grained:*** **concurrent execution of separable simulations within a fn. eval. (e.g., multiple loading cases)**
4. ***Fn eval fine-grained:*** **parallelization of the solution steps within a single analysis code (e.g., Salinas, Aria (SIERRA), Xyce)**

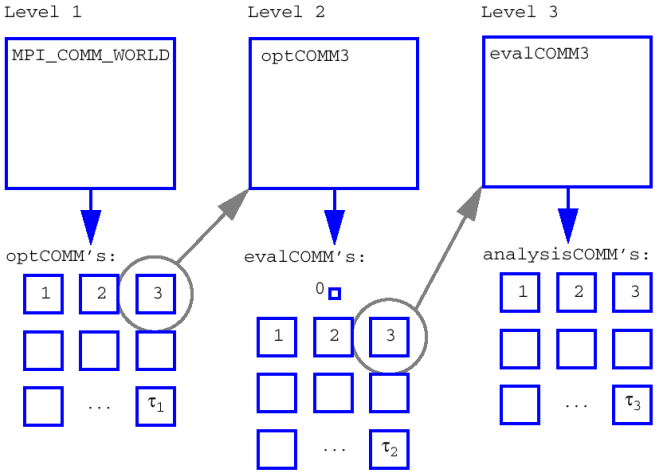
ASCI Red Scalability



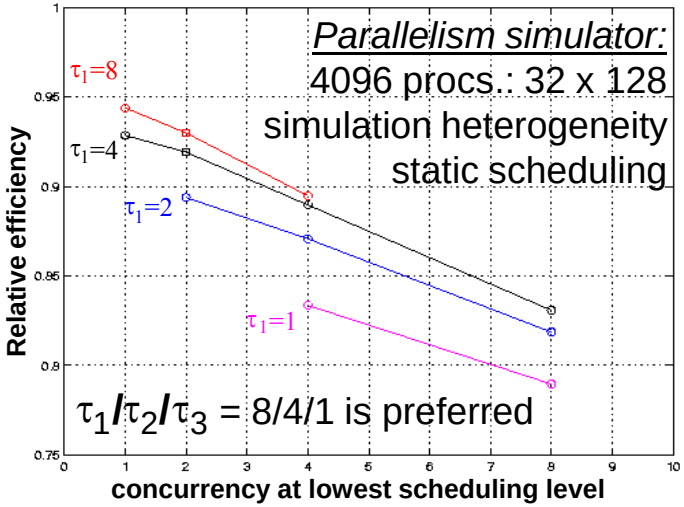




Recursive partitioning & scheduling



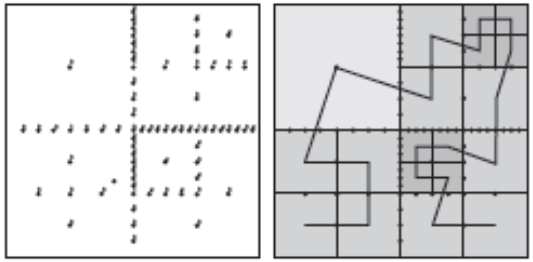
Math analysis & experiments



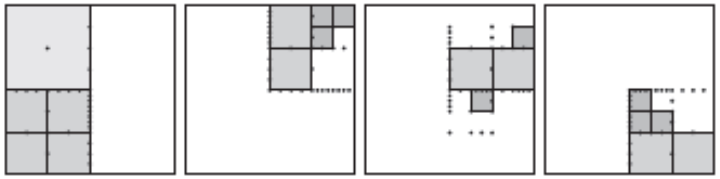
Extending adaptive sparse grids for stochastic collocation to hybrid architectures

Motivation

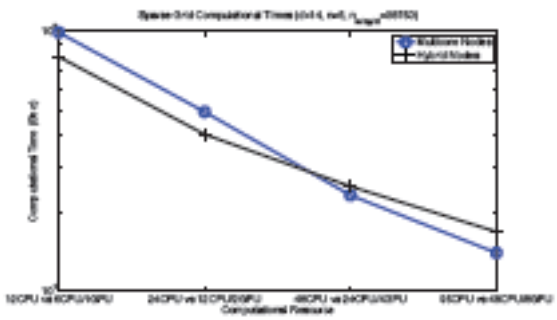
- Very large scale uncertainty quantification (UQ) of realistic problems require 'curse of dimensionality' limiting methods, such as sparse grids, and scalable UQ methods, such as stochastic collocation.
- Growing high-performance computing to the exascale will require high flop/watt computing devices, and next generation architectures are adopting these devices in hybrid CPU/GPU systems.



2D unbounded adapted sparse grid using five subspace levels. Left: each point indexed with subspace level used. Right: domains of the respective highest level and a generalized SFC used for decomposition.



Domain decomposition of above adaptive sparse grid to four nodes and/or thread blocks based on a generalized SFC.



Strong scaling of sparse grid construction on the hybrid test cluster. MPI only vs. hybrid MPI/CUDA parallelization.

Results

- Built an adaptive sparse grid library for stochastic collocation
- Extended this library so that parallel sparse grids can be scalable on small hybrid test clusters.

Courtesy of Ralf Deiterding, ORNL

OAK RIDGE National Laboratory

Embedded stochastic Galerkin (SG) uncertainty quantification method generates large block systems

- **Steady-state stochastic problem (for simplicity):**

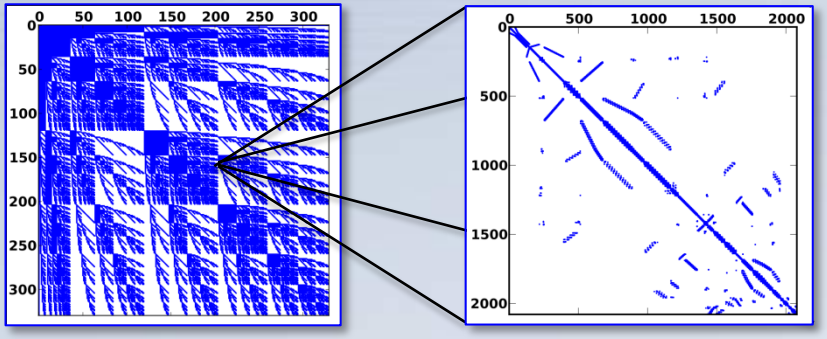
$$\text{Find } u(\xi) \text{ such that } f(u,\xi) = 0,\ \xi : \Omega \to \Gamma \subset R^M,\ \text{density } \rho$$

- **Stochastic Galerkin method (Ghanem and many, many others…):**

$$\hat{u}(\xi) = \sum_{i=0}^{P} u_i \psi_i(\xi) \to F_i(u_0, \ldots, u_P) = \int_\Gamma f(\hat{u}(y), y)\psi_i(y)\rho(y)dy = 0, \quad i = 0, \ldots, P$$

- **Method generates new coupled spatial-stochastic nonlinear problem (intrusive)**

$$0 = F(U) = \begin{bmatrix} F_0 \\ F_1 \\ \vdots \\ F_P \end{bmatrix}, \quad U = \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_P \end{bmatrix} \quad \frac{\partial F}{\partial U} :$$



Stochastic sparsity

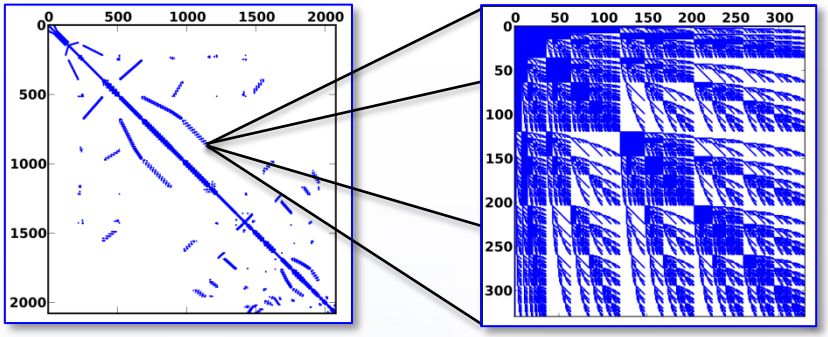
Spatial sparsity

- **Advantages:**
  - Many fewer stochastic degrees-of-freedom for comparable level of accuracy
- **Challenges:**
  - Computing SG residual and Jacobian entries in large-scale, production simulation codes
  - Solving resulting systems of equations efficiently

Courtesy of Eric Phipps, SNL

Exploit large stochastic blocks for multicore shared-memory parallelism

- Rearrange for an outer-spatial, inner-stochastic, ordering
  - Obtain very large, nearly dense blocks
  - Use sparse outer layout for distributed memory parallelism
  - Use dense inner blocks for on-node shared memory parallelism



Spatial sparsity Stochastic sparsity

- Requires heterogeneous multicore parallelism in complete forward uncertainty propagation calculation
  - Application fill
  - Iterative solver matrix-vector productions
  - Preconditioning

- FY12-14 SNL LDRD

Courtesy of Eric Phipps, SNL

Adaptive Collocation Methods

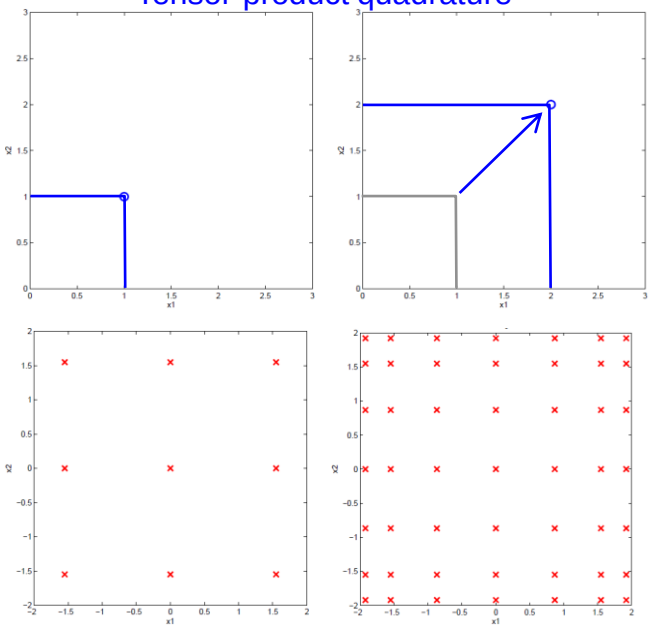
Drivers: Efficiency, robustness, *scalability* → adaptive methods, adjoint enhancement

Polynomial order (p-) refinement approaches:

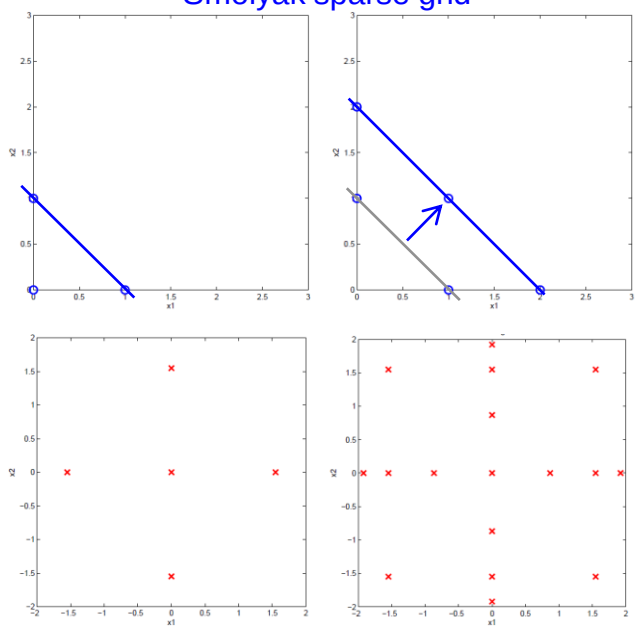
- **Uniform:** *isotropic* tensor/sparse grids
  - *Increment grid:* increase order/level, ensure change (restricted growth in nested rules)
  - *Assess convergence:* L^2 change in response covariance

$$\boxed{w+1 \leq |\mathbf{i}| \leq w+n}$$

Tensor-product quadrature



Smolyak sparse grid



Adaptive Collocation Methods

Drivers: Efficiency, robustness, *scalability* → adaptive methods, adjoint enhancement

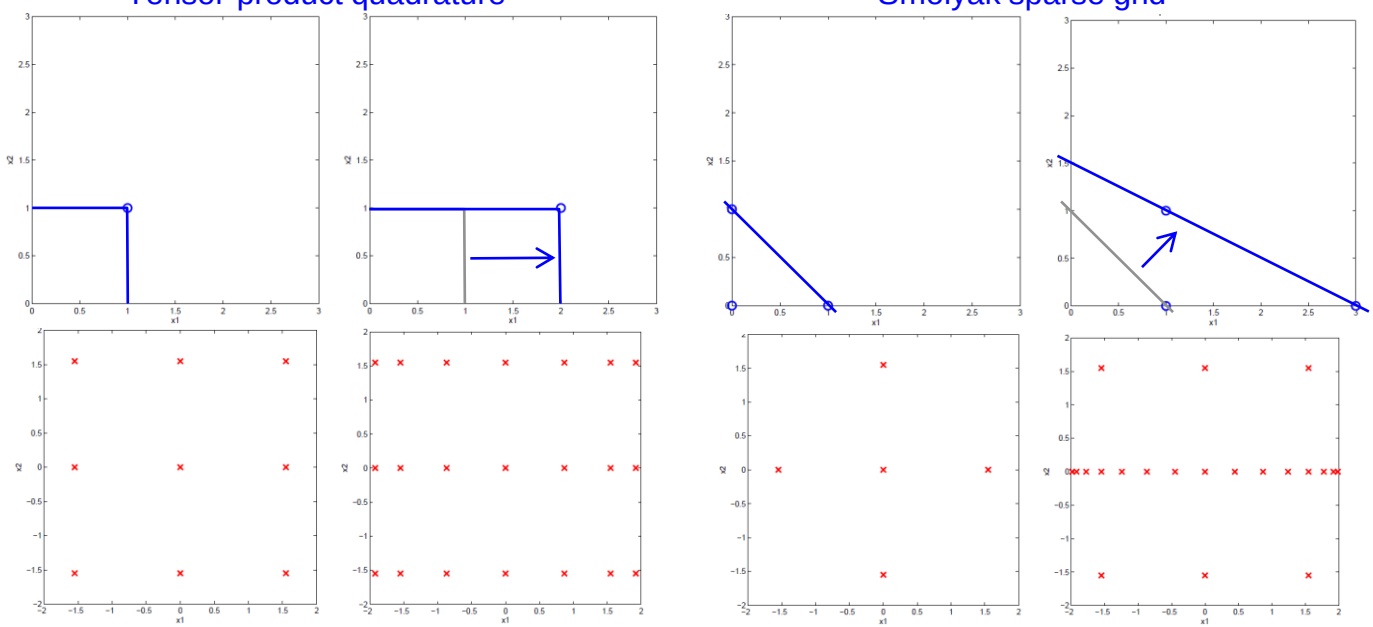
Polynomial order (p-) refinement approaches:

- **Uniform:** *isotropic* tensor/sparse grids
  - *Increment grid:* increase order/level, ensure change (restricted growth in nested rules)
  - *Assess convergence:* L^2 change in response covariance
- **Dimension-adaptive:** *anisotropic* tensor/sparse grids
  - **PCE/SC:** variance-based decomp. → total Sobol' indices → anisotropy (dimension preference)
  - **PCE:** spectral coefficient decay rates → anisotropy (index set weights)

$$\mathrm{w}\underline{\gamma} < \mathbf{i} \cdot \gamma \leq \mathrm{w}\underline{\gamma} + |\gamma|$$

Tensor-product quadrature

Smolyak sparse grid



Adaptive Collocation Methods

Drivers: Efficiency, robustness, *scalability* → adaptive methods, adjoint enhancement

Polynomial order (p-) refinement approaches:

- **Uniform:** *isotropic* tensor/sparse grids
  - *Increment grid:* increase order/level, ensure change (restricted growth in nested rules)
  - *Assess convergence:* L^2 change in response covariance
- **Dimension-adaptive:** *anisotropic* tensor/sparse grids $\mathrm{w}\underline{\gamma} < \mathbf{i} \cdot \gamma \leq \mathrm{w}\underline{\gamma} + |\gamma|$
  - **PCE/SC:** variance-based decomp. → total Sobol' indices → anisotropy
  - **PCE:** spectral coefficient decay rates → anisotropy
- **Goal-oriented dimension-adaptive:** *generalized* sparse grids
  - **PCE/SC:** change in QOI induced by trial index sets on active front

1. *Initialization:* Starting from reference grid (often w = 0 grid), define active index sets using admissible forward neighbors of all old index sets.

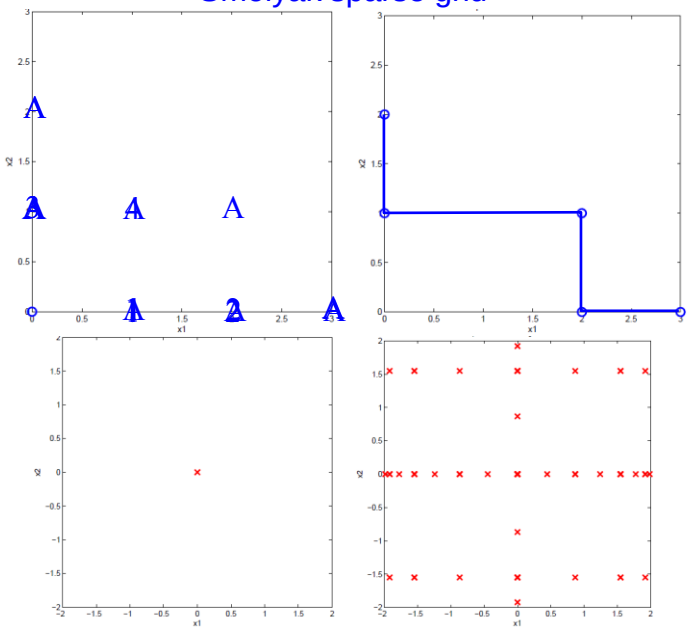
2. *Trial set evaluation:* For each trial index set, evaluate tensor grid, form tensor expansion, update combinatorial coefficients, and combine with reference expansion. Perform necessary bookkeeping to allow efficient restoration.

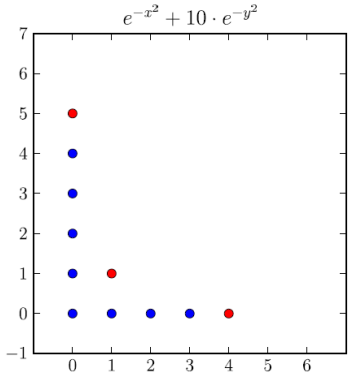
3. *Trial set selection:* Select trial index set that induces largest change in statistical QOI.

4. *Update sets:* If largest change > tolerance, then promote selected trial set from active to old and compute new admissible active sets; return to 2. If tolerance is satisfied, advance to step 5.

5. *Finalization:* Promote all remaining active sets to old set, update combinatorial coefficients, and perform final combination of tensor expansions to arrive at final result for statistical QOI.

Smolyak sparse grid





(Gerstner, 2003)

Fine-grained control: frontier not limited by prescribed shape of index set constraint

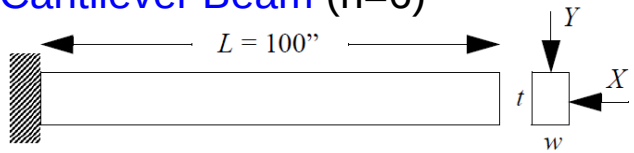
Algebraic Test Functions

Short Column (n=5)

$$g(\mathbf{x}) = 1 - \frac{4M}{bh^2Y} - \frac{P^2}{b^2h^2Y^2}$$

$b = U[5,15]$, $h = U[15,25]$,
$P = N(500, 100)$, $M = N(2000, 400)$,
$\rho_{P,M} = 0.5$, $Y = logN(5, 0.5)$

Cantilever Beam (n=6)



$$\begin{aligned} S &= \frac{600}{wt^2}Y + \frac{600}{w^2t}X \leq R \\ D &= \frac{4L^3}{Ewt}\sqrt{\left(\frac{Y}{t^2}\right)^2 + \left(\frac{X}{w^2}\right)^2} \leq D_0 \end{aligned}$$

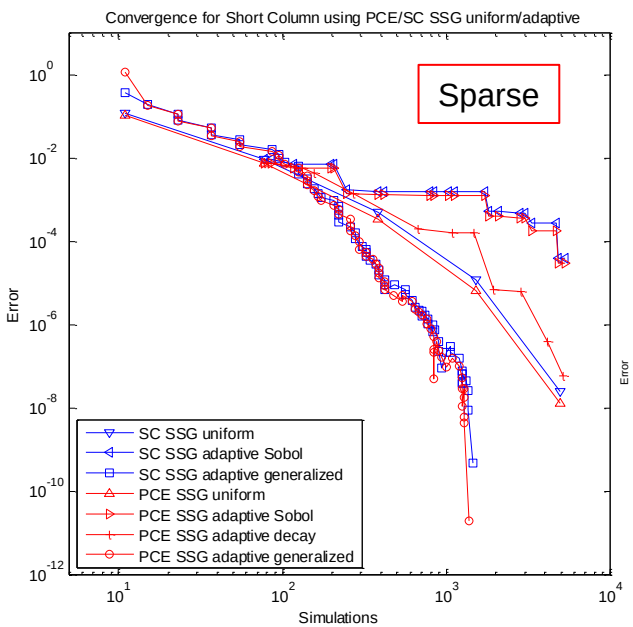
w, t, R, E, X, Y: $U[1,10]$, $U[1,10]$,
$N(4E4, 2E3)$, $N(2.9E7, 1.45E6)$,
$N(500, 100)$, $N(1E3, 100)$; $D_0 = 2.2535$"

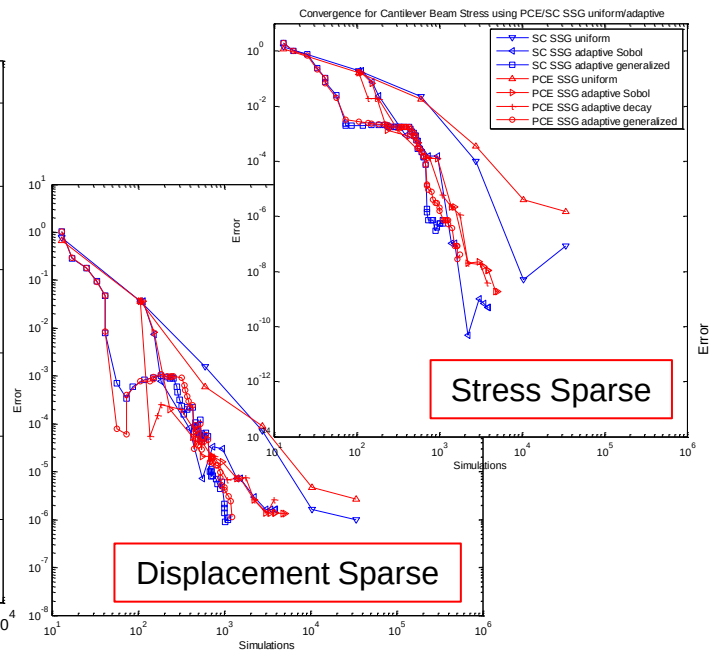
Ishigami (n=3)

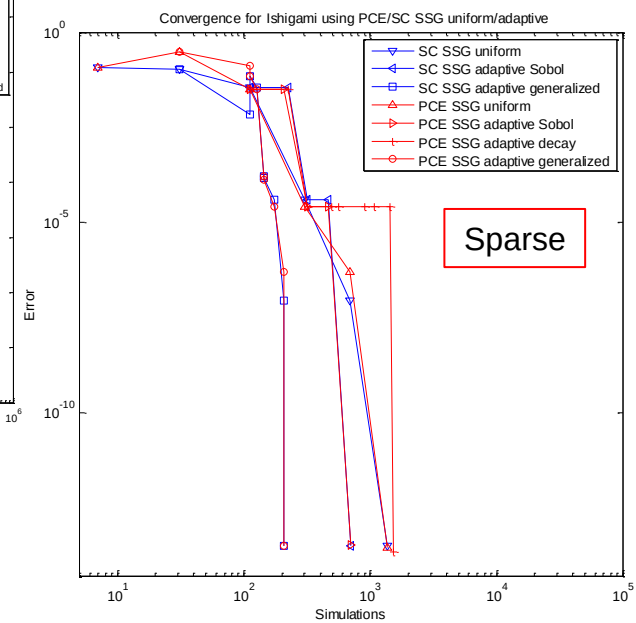
$$\begin{aligned} f(\mathbf{x}) &= \sin(2\pi x_1 - \pi) \\ &+ 7\sin^2(2\pi x_2 - \pi) \\ &+ 0.1(2\pi x_3 - \pi)^4 \sin(2\pi x_1 - \pi) \end{aligned}$$

x_1, x_2, x_3: *iid* $U[0, 1]$

- Designed to be challenging for global SA: term cancellations at mid-point & bounds
- Premature convergence in adaptive methods → start from higher-order grid







Extend Input Scalability through Adjoint Derivative-Enhancement

$$\begin{bmatrix} \vdots & \vdots & & \vdots \\ \pi_{0,j}(\vec{\xi}_i) & \pi_{1,j}(\vec{\xi}_i) & \cdots & \pi_{P,j}(\vec{\xi}_i) \\ \frac{\partial \pi_{0,j}}{\partial \xi_1}(\vec{\xi}_i) & \frac{\partial \pi_{1,j}}{\partial \xi_1}(\vec{\xi}_i) & \cdots & \frac{\partial \pi_{P,j}}{\partial \xi_1}(\vec{\xi}_i) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \pi_{0,j}}{\partial \xi_{n_\xi}}(\vec{\xi}_i) & \frac{\partial \pi_{1,j}}{\partial \xi_{n_\xi}}(\vec{\xi}_i) & \cdots & \frac{\partial \pi_{P,j}}{\partial \xi_{n_\xi}}(\vec{\xi}_i) \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{pmatrix} \vdots \\ \vec{u}^{(m,j)} \\ \vec{u}^{(m+1,j)} \\ \vdots \\ \vec{u}^{(m+n_\xi,j)} \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ \vec{u}_i \\ \frac{\partial \vec{u}_i}{\partial \xi_1} \\ \vdots \\ \frac{\partial \vec{u}_i}{\partial \xi_{n_\xi}} \\ \vdots \end{pmatrix}$$

PCE:

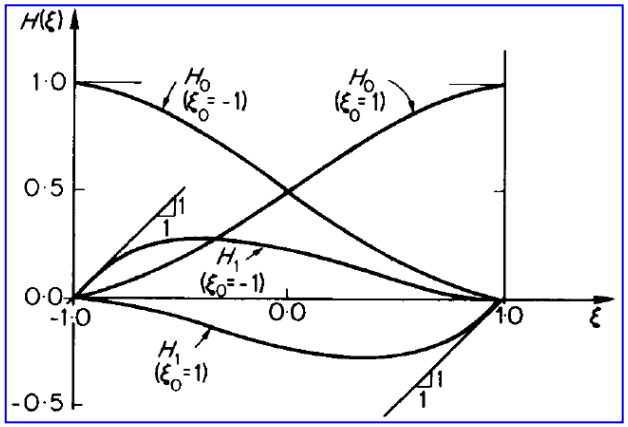
- **Linear regression with derivatives**
  - **Gradients/Hessians → addtnl. eqns.**

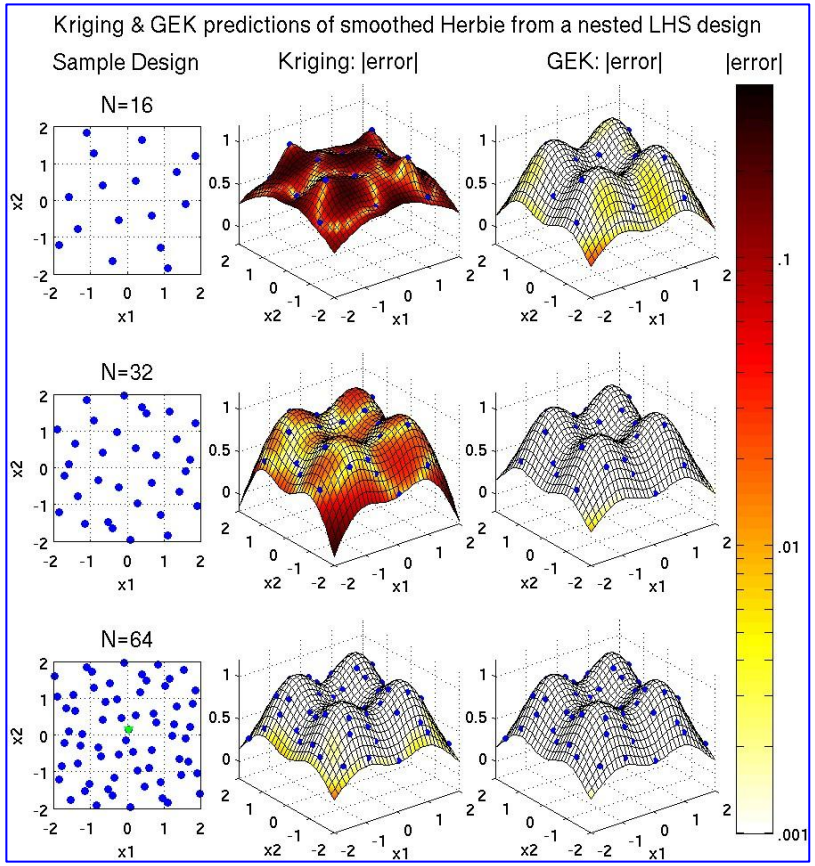
SC:

- **Gradient-enhanced interpolants**
  - **Local: cubic Hermite splines**
  - **Global: Hermite interpolation polynomials**

EGRA:

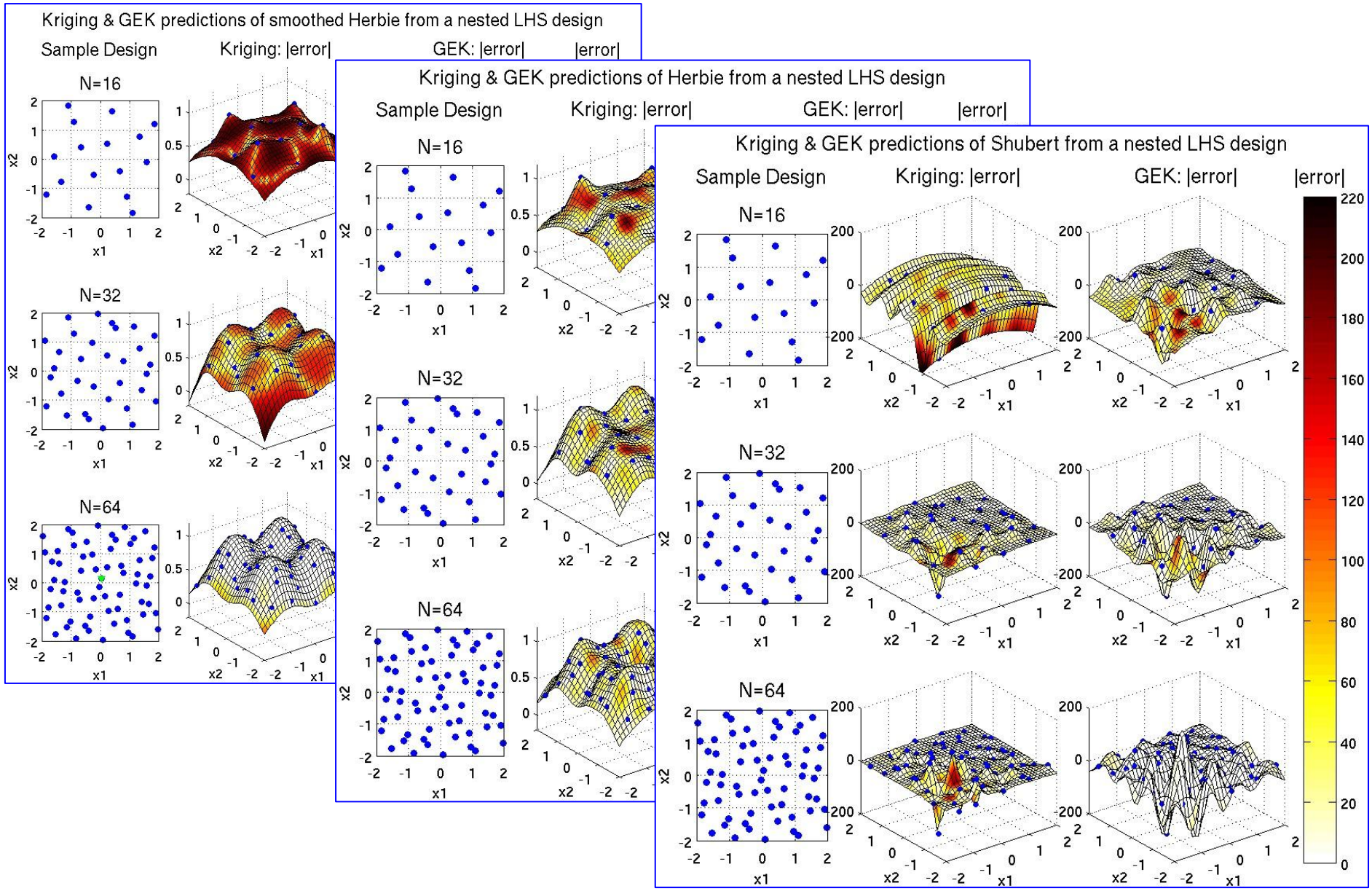
- **Gradient-enhanced kriging/cokriging**
  - **Interpolates function values and gradients**
  - **Scaling: n^2 → n**





Gradient-Enhanced Kriging (GEK)

Enabler for EGRA and Bayesian emulation



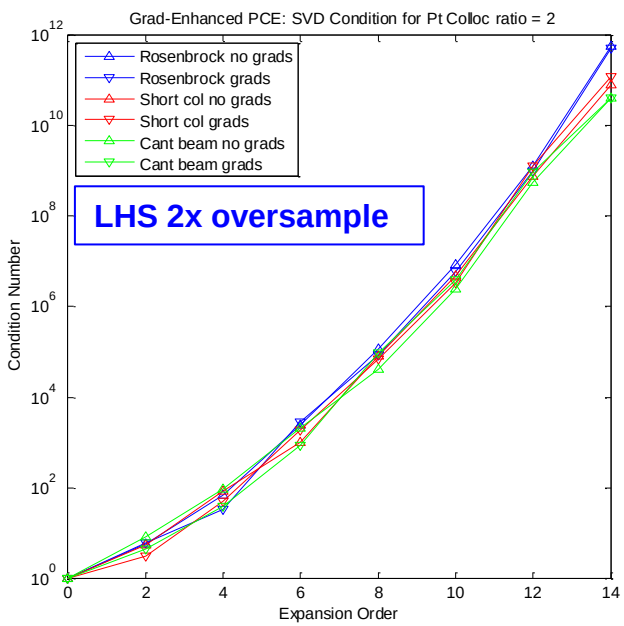
Gradient-Enhanced PCE

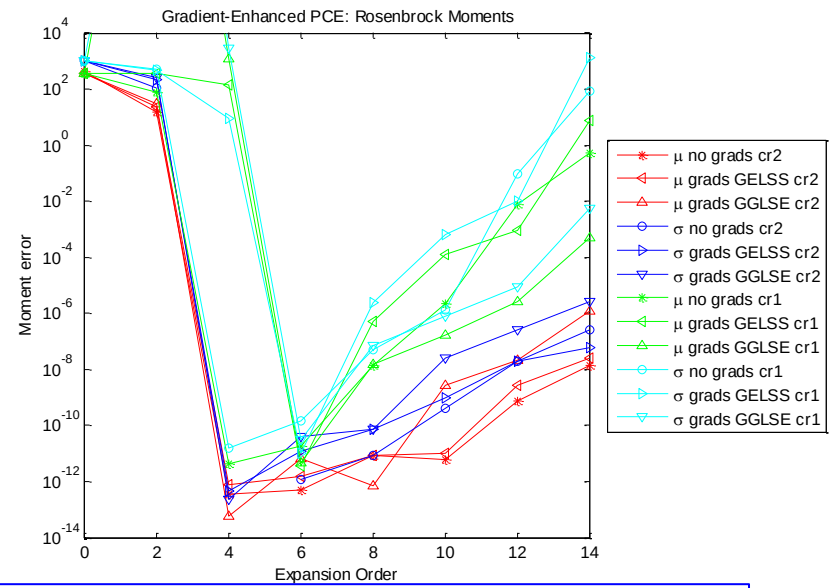
Straightforward regression approach:

$$\begin{bmatrix} \vdots & \vdots & & \vdots \\ \pi_{0,j}(\vec{\xi}_i) & \pi_{1,j}(\vec{\xi}_i) & \cdots & \pi_{P,j}(\vec{\xi}_i) \\ \frac{\partial \pi_{0,j}}{\partial \xi_1}(\vec{\xi}_i) & \frac{\partial \pi_{1,j}}{\partial \xi_1}(\vec{\xi}_i) & \cdots & \frac{\partial \pi_{P,j}}{\partial \xi_1}(\vec{\xi}_i) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \pi_{0,j}}{\partial \xi_{n_\xi}}(\vec{\xi}_i) & \frac{\partial \pi_{1,j}}{\partial \xi_{n_\xi}}(\vec{\xi}_i) & \cdots & \frac{\partial \pi_{P,j}}{\partial \xi_{n_\xi}}(\vec{\xi}_i) \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{pmatrix} \vdots \\ \vec{u}^{(m,j)} \\ \vec{u}^{(m+1,j)} \\ \vdots \\ \vec{u}^{(m+n_\xi,j)} \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ \vec{u}_i \\ \frac{\partial \vec{u}_i}{\partial \xi_1} \\ \vdots \\ \frac{\partial \vec{u}_i}{\partial \xi_{n_\xi}} \\ \vdots \end{pmatrix}$$

- unweighted LLS by SVD (LAPACK GELSS)
- equality constrained LLS by QR (LAPACK GGLSE) when under-determined by values alone

Vandermonde-like systems known to suffer from ill-conditioning





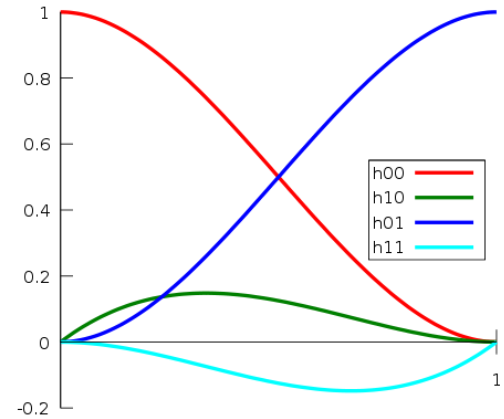
Error growth as we over-resolve exact solutions

Dimension-adaptive h-refinement with gradient-enhanced interpolants

Dimension-adaptive h-refinement for SC:

- *Local spline interpolants:* linear Lagrange (value-based), cubic Hermite (gradient-enhanced)
- *Global grids:* iso/aniso tensor, iso/aniso/generalized sparse
- *h-refinement:* uniform, adaptive, goal-oriented adaptive
- *Basis formulations:* nodal, hierarchical

Cubic shape fns: type 1 (value) & type 2 (gradient)



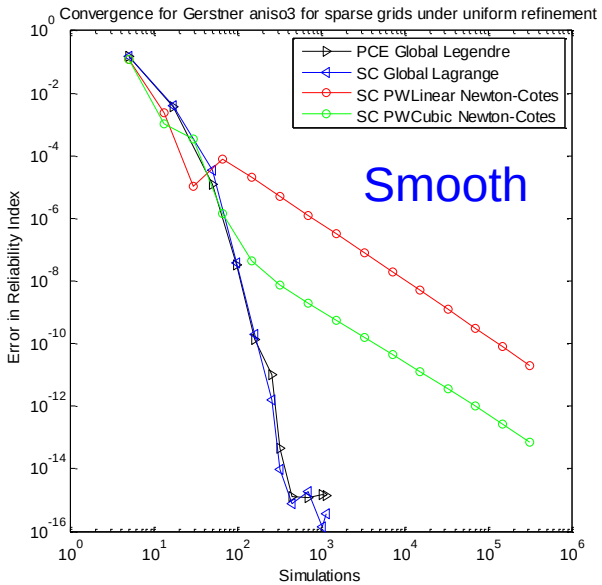
Multivariate tensor product to arbitrary derivative order (Lalescu):

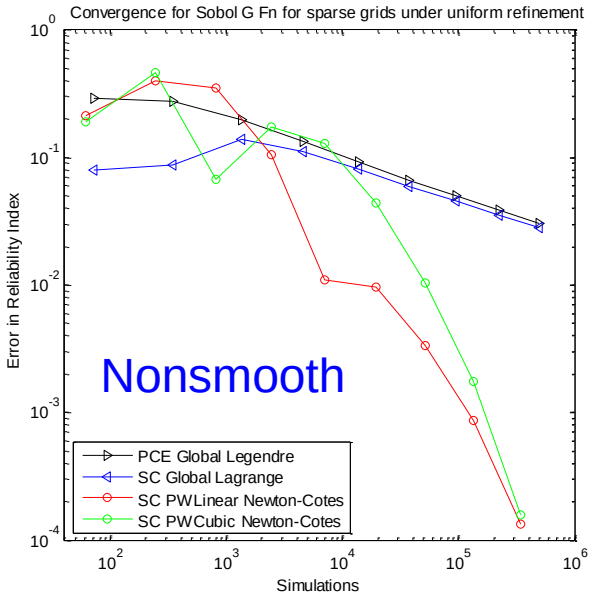
$$s^{(n)}(x_1, x_2, \ldots, x_D) = \sum_{l_1,\ldots,l_D=0}^{m} \sum_{i_1,\ldots,i_D=0,1} f^{(l_1,\ldots,l_D)}(i_1,\ldots,i_D) \prod_{k=1}^{D} \alpha_{i_k}^{(n,l_k)}(x_k)$$

$$f = \sum_{i=1}^{N} f_i H_i^{(1)}(x_1) H_i^{(1)}(x_2) H_i^{(1)}(x_3) + \sum_{i=1}^{N} \frac{df_i}{dx_1} H_i^{(2)}(x_1) H_i^{(1)}(x_2) H_i^{(1)}(x_3) + \sum_{i=1}^{N} \frac{df_i}{dx_2} H_i^{(1)}(x_1) H_i^{(2)}(x_2) H_i^{(1)}(x_3) + \sum_{i=1}^{N} \frac{df_i}{dx_3} H_i^{(1)}(x_1) H_i^{(1)}(x_2) H_i^{(2)}(x_3)$$

$$\mu = \sum_{i=1}^{N} f_i w_i^{(1)} w_i^{(1)} w_i^{(1)} + \sum_{i=1}^{N} \frac{df_i}{dx_1} w_i^{(2)} w_i^{(1)} w_i^{(1)} + \sum_{i=1}^{N} \frac{df_i}{dx_2} w_i^{(1)} w_i^{(2)} w_i^{(1)} + \sum_{i=1}^{N} \frac{df_i}{dx_3} w_i^{(1)} w_i^{(1)} w_i^{(2)}$$

and similar for higher-order moments





Build on efficient/scalable UQ core

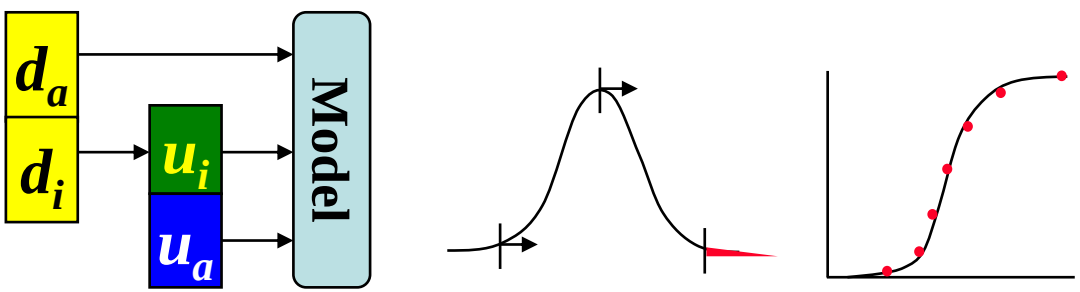
Stochastic sensitivity analysis

- Aleatory or combined expansions including nonprobabilistic dimensions *s*
  → sensitivities of moments w.r.t. design and/or epistemic parameters

$$R(\xi, s) = \sum_{j=0}^{P} \alpha_j \Psi_j(\xi, s) \quad \begin{cases} \dfrac{d\mu_R}{ds} = \langle \dfrac{dR}{ds} \rangle \\ \dfrac{d\sigma_R^2}{ds} = 2\sum_{j=1}^{P} \alpha_j \langle \dfrac{dR}{ds}, \Psi_j \rangle \end{cases}$$

$$R(\xi, s) = \sum_{j=0}^{P} \alpha_j(s) \Psi_j(\xi) \quad \begin{cases} \mu_R(s) = \sum_{j=0}^{P} \alpha_j \langle \Psi_j(\xi, s) \rangle_\xi \\ \sigma_R^2(s) = \sum_{j=0}^{P} \sum_{k=0}^{P} \alpha_j \alpha_k \langle \Psi_j(\xi, s) \Psi_k(\xi, s) \rangle_\xi - \mu_R^2(s) \end{cases}$$

Design and Model Calibration Under Uncertainty



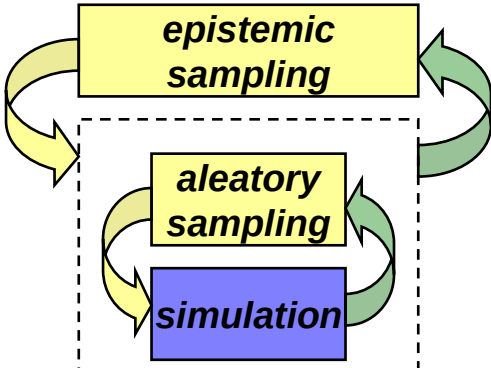
Add resp stats s_u *(μ, σ, z/β/p)*

$$\begin{aligned} \min \quad & f(d) + W s_u(d) \\ \text{s.t.} \quad & g_l \le g(d) \le g_u \\ & h(d) = h_t \\ & d_l \le d \le d_u \\ & a_l \le A_i s_u(d) \le a_u \\ & A_e s_u(d) = a_t \end{aligned}$$

Mixed Aleatory-Epistemic UQ

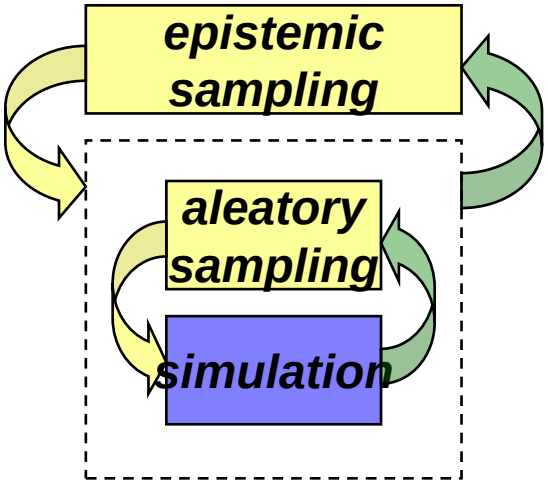
- Approaches that are more accurate/efficient than nested sampling
  - Interval-valued probability (IVP), *aka* PBA
  - Dempster-Shafer theory of evidence (DSTE)
  - Second-order probability (SOP), *aka* PoF

Increasing epistemic structure (stronger assumptions)



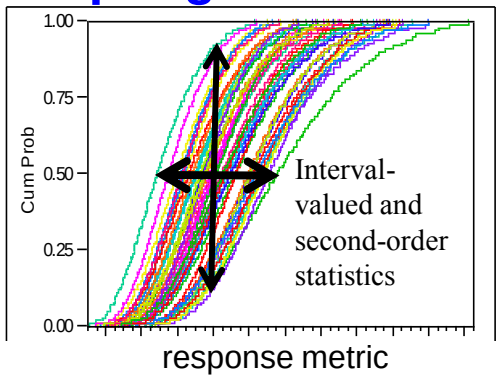
Mixed Aleatory-Epistemic UQ: IVP, DSTE, and SOP

Epistemic uncertainty (aka: subjective, reducible, lack of knowledge uncertainty): insufficient info to specify objective probability distributions



Traditional approach: nested sampling

- Expensive sims → under-resolved sampling (especially @ outer loop)
- Under-prediction of credible outcomes



Algorithmic approaches

- Interval-valued probability (IVP), *aka* probability bounds analysis (PBA)
- Dempster-Shafer theory of evidence (DSTE)
- Second-order probability (SOP), *aka* probability of frequency

Increasing epistemic structure (stronger assumptions)

Address accuracy and efficiency

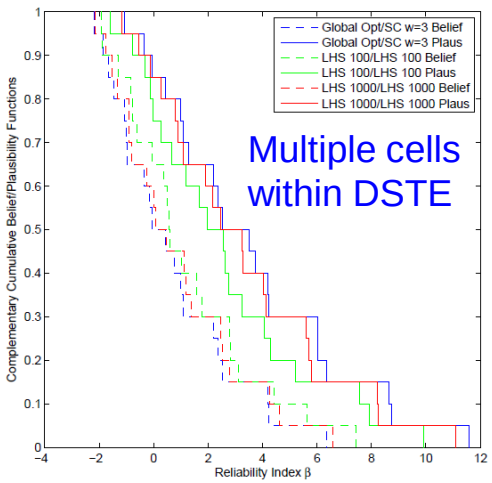
- Inner loop: stochastic exp. that are epistemic-aware (aleatory, combined)
- Outer loop:
  - IVP, DSTE: opt-based interval estimation, global (EGO) or local (NLP)
  - SOP: nested stochastic exp. (nested expectation is only post-processing in special cases)

$$\begin{array}{ll} \text{minimize} & M(s) \\ \text{subject to} & s_L \leq s \leq s_U \\ \\ \text{maximize} & M(s) \\ \text{subject to} & s_L \leq s \leq s_U \end{array}$$

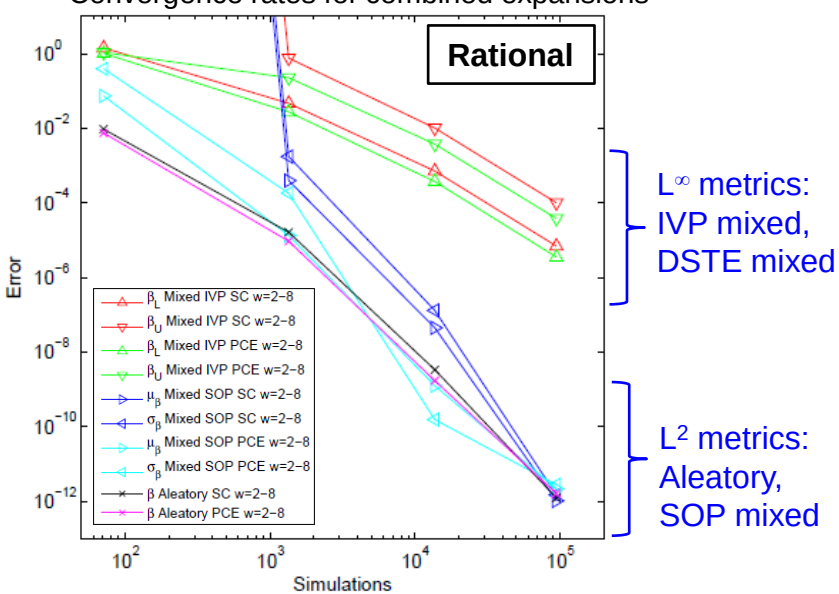
Mixed Aleatory-Epistemic UQ:
IVP, SOP, and DSTE based on Stochastic Expansions

Interv Est Approach	UQ Approach	Expansion Variables	Evaluations (Fn, Grad)	Area	β
IVP SC SSG Aleatory: β interval converged to 5-6 digits by 300-400 evals					
EGO	SC SSG w = 1	Aleatory	(84/91, 0/0)	[75.0002, 374.999]	[-2.26264, 11.8623]
EGO	SC SSG w = 2	Aleatory	(372/403, 0/0)	[75.0002, 374.999]	[-2.18735, 11.5900]
EGO	SC SSG w = 3	Aleatory	(1260/1365, 0/0)	[75.0002, 374.999]	[-2.18732, 11.5900]
EGO	SC SSG w = 4	Aleatory	(3564/3861, 0/0)	[75.0002, 374.999]	[-2.18732, 11.5900]
NPSOL	SC SSG w = 1	Aleatory	(21/77, 21/77)	[75.0000, 375.000]	[-2.26264, 11.8623]
NPSOL	SC SSG w = 2	Aleatory	(93/341, 93/341)	[75.0000, 375.000]	[-2.18735, 11.5901]
NPSOL	SC SSG w = 3	Aleatory	(315/1155, 315/1155)	[75.0000, 375.000]	[-2.18732, 11.5900]
NPSOL	SC SSG w = 4	Aleatory	(891/3267, 891/3267)	[75.0000, 375.000]	[-2.18732, 11.5900]
IVP nested LHS sampling: converged to 2-3 digits by 10^8 evals					
LHS 100	LHS 100	N/A	$(10^4/10^4, 0/0)$	[80.5075, 338.607]	[-2.14505, 8.64891]
LHS 1000	LHS 1000	N/A	$(10^6/10^6, 0/0)$	[76.5939, 368.225]	[-2.19883, 11.2353]
LHS 10^4	LHS 10^4	N/A	$(10^8/10^8, 0/0)$	[76.4755, 373.935]	[-2.16323, 11.5593]

Fully converged area interval = [75., 375.], β interval = [−2.18732, 11.5900]



Convergence rates for combined expansions



Impact: render mixed UQ studies practical for large-scale applications

Current:

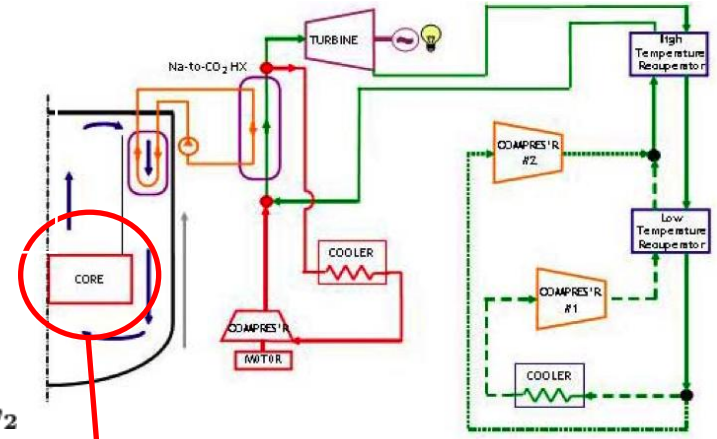
- **Global or local opt. for epistemic intervals**
  → accuracy *or* scaling w/ epistemic dimension
- **Global or local UQ for aleatory statistics**
  → accuracy *or* scaling w/ aleatory dimension

Future:

- ***adaptive* and *adjoint-enhanced* global methods**
  → accuracy *and* scaling

ASCR: Multi-Physics and Network-Coupled UQ

- Component-level UQ via stochastic expansions
- Stochastic dimension reduction at component interfaces (generate new bases orthogonal to (implied) output PDFs)
- Strongly coupled solver technology for coupled stochastic problems



Low-fidelity Network Plant Model

Component 1

$$v_2 = G_1(v_1, p_1) = g_1(u_1(v_1), p_1) \;\; \text{s.t.} \;\; f_1(u_1, v_1, p_1) = 0$$

Component 2

$$v_1 = G_2(v_2, p_2) = g_2(u_2(v_2), p_2) \;\; \text{s.t.} \;\; f_2(u_2, v_2, p_2) = 0$$

v_1 v_2

Nonlinear elimination

Equations

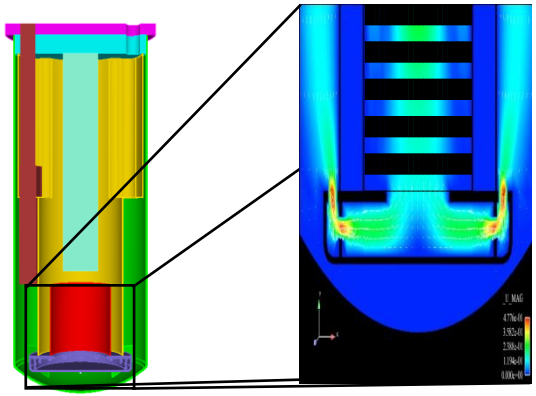
$$v_2 - G_1(v_1, p_1) = 0$$
$$v_1 - G_2(v_2, p_2) = 0$$

Newton Step

$$\begin{bmatrix} -dG_1/dv_1 & 1 \\ 1 & -dG_2/dv_2 \end{bmatrix} \begin{bmatrix} \Delta v_1 \\ \Delta v_2 \end{bmatrix} = - \begin{bmatrix} v_2 - G_1(v_1, p_1) \\ v_1 - G_2(v_2, p_2) \end{bmatrix}$$

$$\frac{dG_i}{dv_i} = -\frac{\partial g_i}{\partial u_i} \left(\frac{\partial f_i}{\partial u_i} \right)^{-1} \frac{\partial f_i}{\partial v_i}$$

$$\left.\begin{aligned} \frac{1}{\langle \Psi_j^2 \rangle} \langle (\hat{v}_2(\xi) - \hat{G}_1(\xi)) \Psi_j(\xi) \rangle = 0 \\ \frac{1}{\langle \Psi_j^2 \rangle} \langle (\hat{v}_1(\xi) - \hat{G}_2(\xi)) \Psi_j(\xi) \rangle = 0 \end{aligned}\right\} \implies \left.\begin{aligned} v_{2j} - G_{1j} = 0 \\ v_{1j} - G_{2j} = 0 \end{aligned}\right\}, \quad j = 0, \ldots, P$$



High-fidelity Multi-physics Component Model (Core)

Graphics courtesy: Rod Schmidt, BRISC project

Sandia National Laboratories

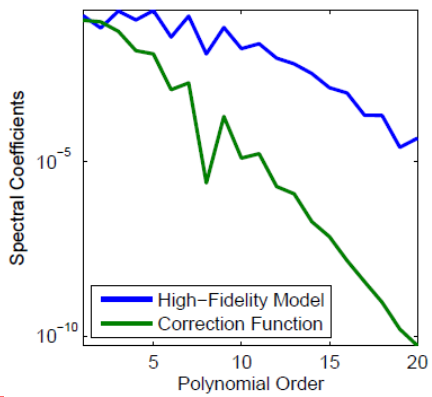
Multifidelity UQ using Stochastic Expansions

Multifidelity UQ through stochastic expansion of model discrepancy:

- Extension of multifidelity opt methods that converge to local HF optimum based on local corrections
- Converge to global HF statistics based on global corrections ($0^{th}/1^{st}$ consistency @HF collocation pts)

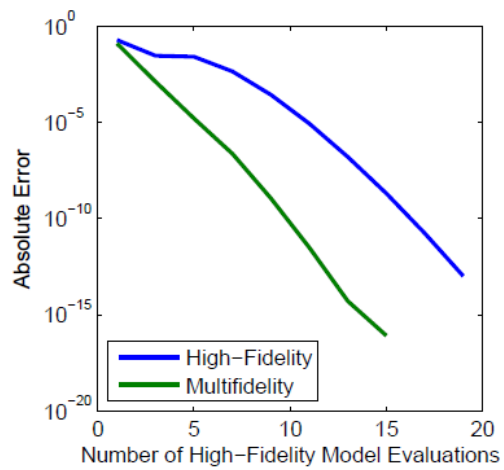
$$\hat{f}_{hi}(\xi) = \sum_{j=1}^{N_{lo}} f_{lo}(\xi_j) L_j(\xi) + \sum_{j=1}^{N_{hi}} \Delta f(\xi_j) L_j(\xi)$$

$$N_{lo} >> N_{hi}$$

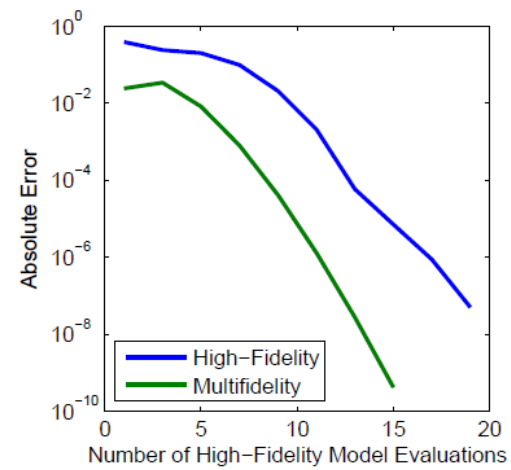


$$R_{\text{high}}(\xi) = e^{-0.05\xi^2} \cos 0.5\xi - 0.5 e^{-0.02(\xi-5)^2}$$

$$R_{\text{low}}(\xi) = e^{-0.05\xi^2} \cos 0.5\xi,$$



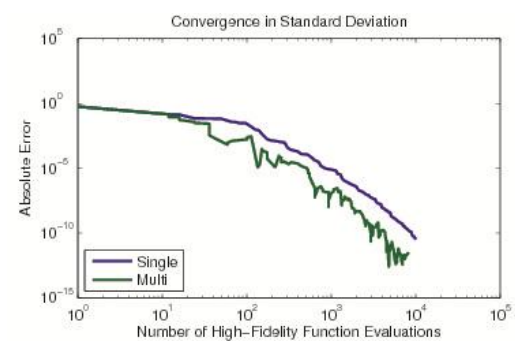
(a) Error in mean



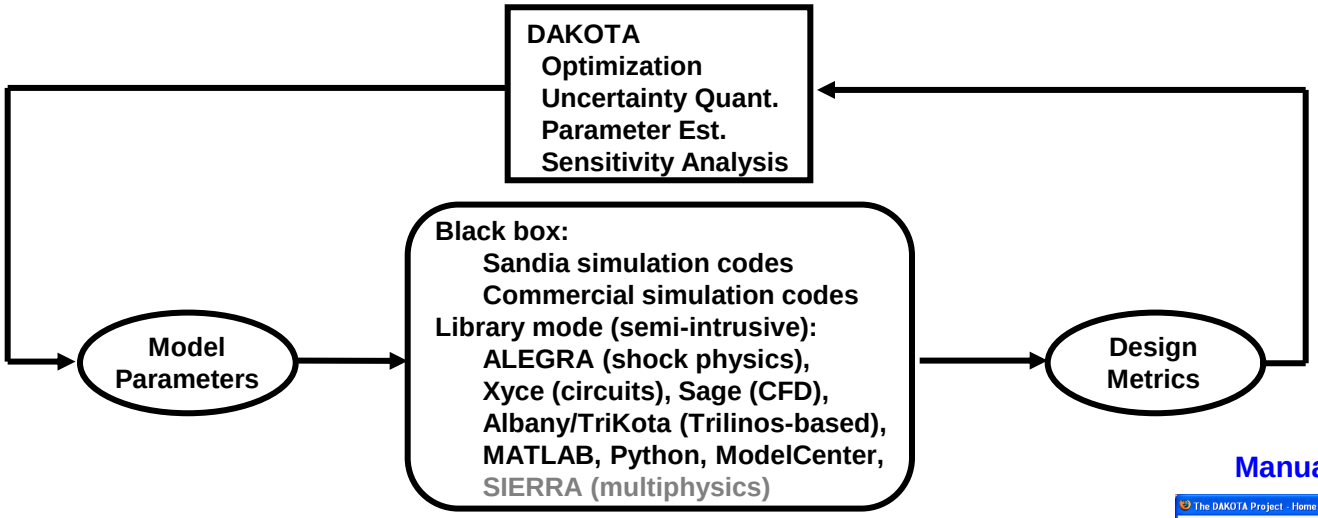
(b) Error in standard deviation

Adaptive algorithm balances LF/HF cost and targets regions where LF predictive capabilities break down:

- Greedy selection of index sets for LF *or* model discrepancy based on ΔQOI/ΔCost



DAKOTA Software



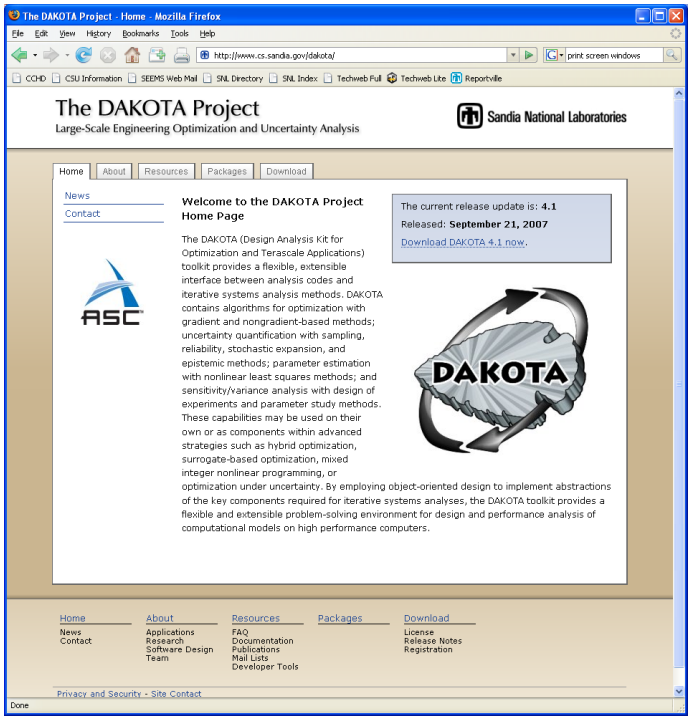
Iterative systems analysis

Multilevel parallel computing

Simulation management

http://dakota.sandia.gov

Manuals, Publications, Training matls. online



Releases: **Major/Interim, Stable/VOTD; *5.1 released 12/10***

Modern SQE: **Linux/Unix, Mac, Windows; Nightly builds/testing; subversion, TRAC, autotools/Cmake**

GNU LGPL: **free downloads worldwide (~9000 total ext. registrations, ~3500 distributions last yr.)**

Community development: **open checkouts now available**

Community support: **dakota-users, dakota-help**