DAKOTA/UQ: A Toolkit For Uncertainty Quantification in a Multiphysics, Massively Parallel Computational Environment

Steven F. Wojtkiewicz
sfwojtk@sandia.gov

Structural Dynamics Research, Dept. 9124
Sandia-Albuquerque

LANL Uncertainty Quantification Working Group

October 4, 2001

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company,for the United States Department of Energy under contract DE-AC04-94AL85000.

Sandia National Laboratories

Uncertainty Quantification

Real Physical Systems:

- Display random and systematic variation- geometry, materials, boundary conditions, initial conditions, excitations
- Vary from one realization to the next
- Display behavior that cannot be precisely measured

Uncertainty occurs in various forms:

- Irreducible, variability, aleatoric
- Reducible, epistemic, subjective, model form uncertainty

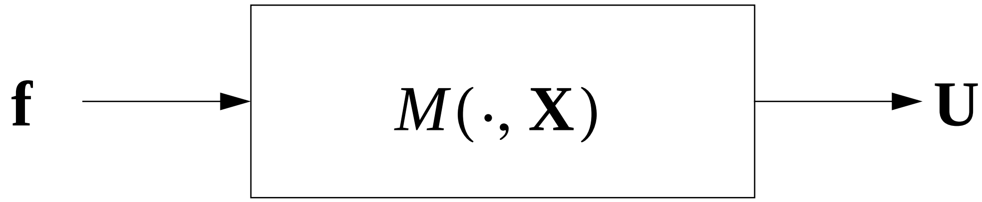
Uncertainty Quantification

Useful in:

- **Analysis and Design**
  - –To assess the reliability of physical systems.
  - –To establish designs that satisfy pre-established reliability requirements.
  - –To establish sensitivities to key uncertainties
- **Model validation, certification, and accreditation**
  - –As defined in the DOE Defense Programs (DOE/DP) ASCI Program Plan, validation is the process of determining the degree to which a computer model is an accurate representation of the real world from the perspective of the *intended model applications.*
  - –Convey confidence in predictions to decision makers

Uncertainty Quantification: General Framework

General Description:



$\mathbf{X}$: vector of uncertain parameters

M: a deterministic mapping

$\mathbf{U}$: output(s) of system

$\mathbf{f}$: input(s) to system

Statistical Approach:

- **Model components of $\mathbf{X}$ as *Random Variables or Fields*, and f as (possibly) *Random External Input***
- **Seek quantities such as $E[g(\mathbf{U})]$. However, what is actually obtained are *conditional* statistics $E[g(\mathbf{U})|M]$.**

Probabilistic/Statistical Approach: Essential Elements of a Statistical Approach:

Conclusion: Need a Generalized Outlook.

Essential Elements of a Statistical Approach:

- **Random External Inputs**
- **Propagation Techniques**
  - –Analytical Reliability Methods; Sampling; Response Surface Approximations; Stochastic Finite Element Methods.
- **Characterization of Models**
  - –Verification and Validation.

General functions of the response, u

Random external loads

Probabilistic characterization of the input parameters

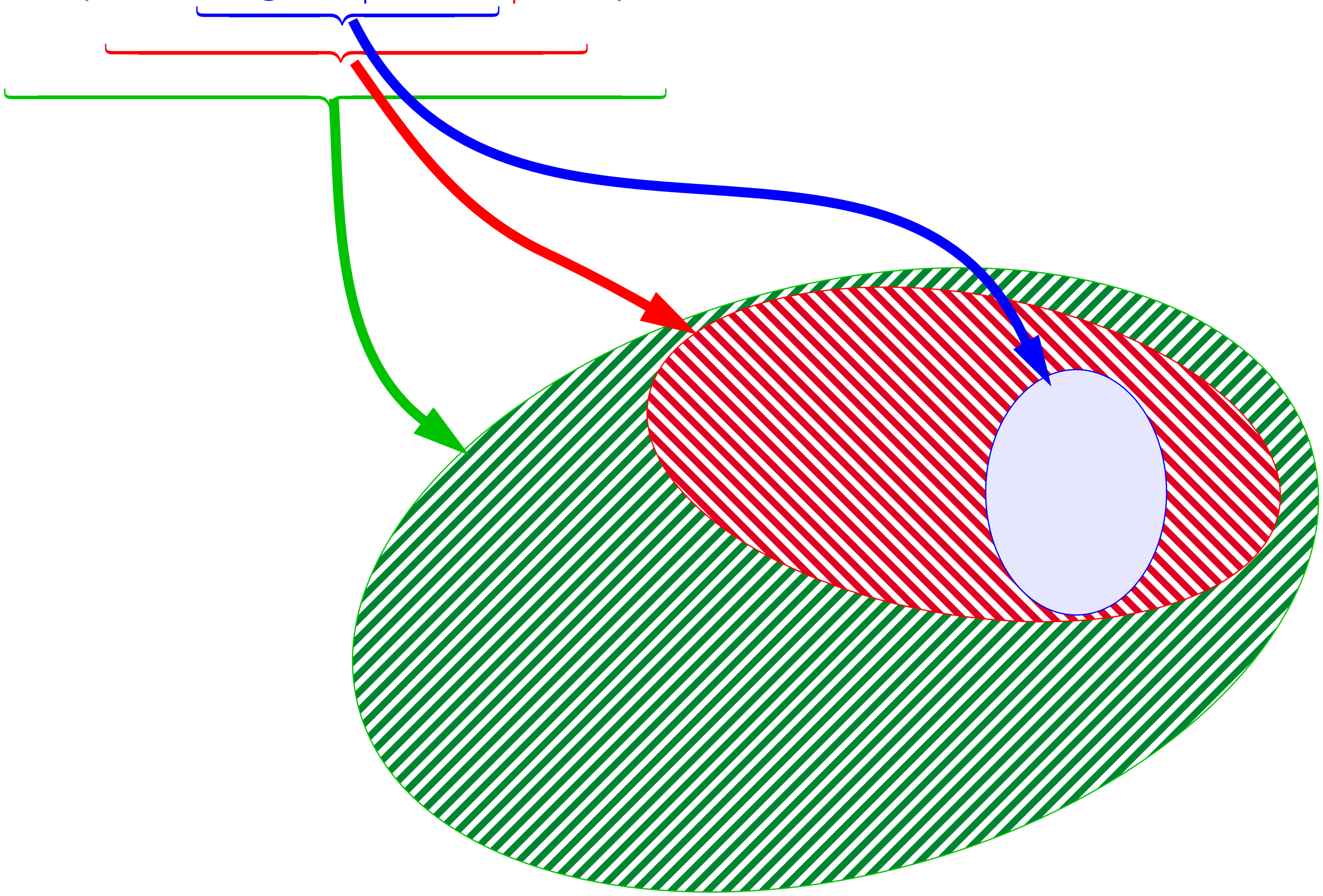
$$E[g(\mathbf{U})] = E\left\langle \underbrace{E\left\{ \overbrace{E[g(\mathbf{U})|M, \mathbf{X}]} \,|M\right\}} \right\rangle$$

Propagation Techniques (Note: Dependency on M)

Characterization of the Model, M

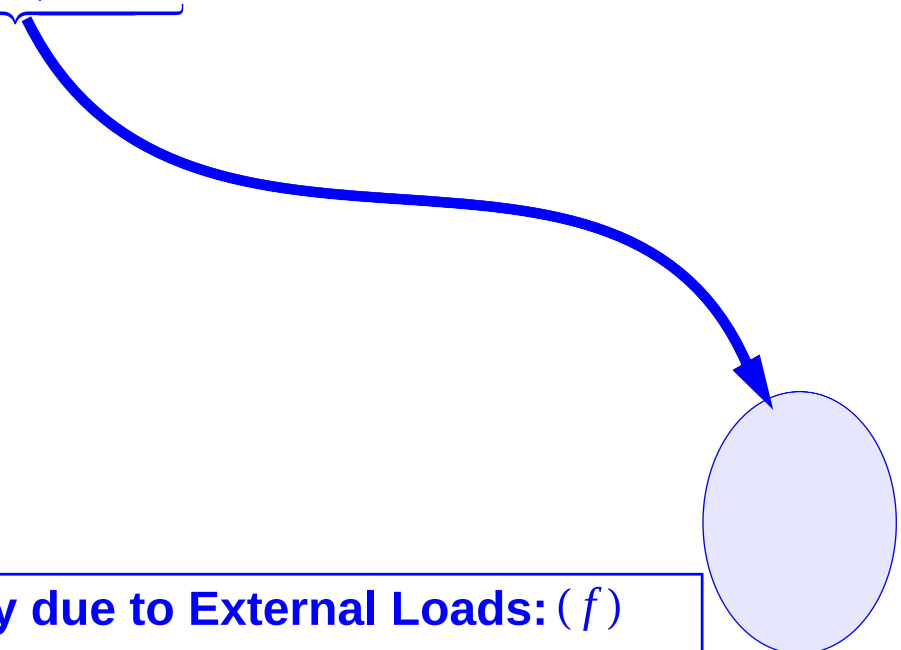
Anatomy of Global Uncertainty

$$E\langle\, E\{\, E[g(\mathbf{U})|M,\mathbf{X}]\, |M\}\,\rangle = E[g(\mathbf{U})]$$



Anatomy of Global Uncertainty

$$\underbrace{\boldsymbol{E}[\boldsymbol{g}(\mathbf{U})|M, \mathbf{X}]}$$

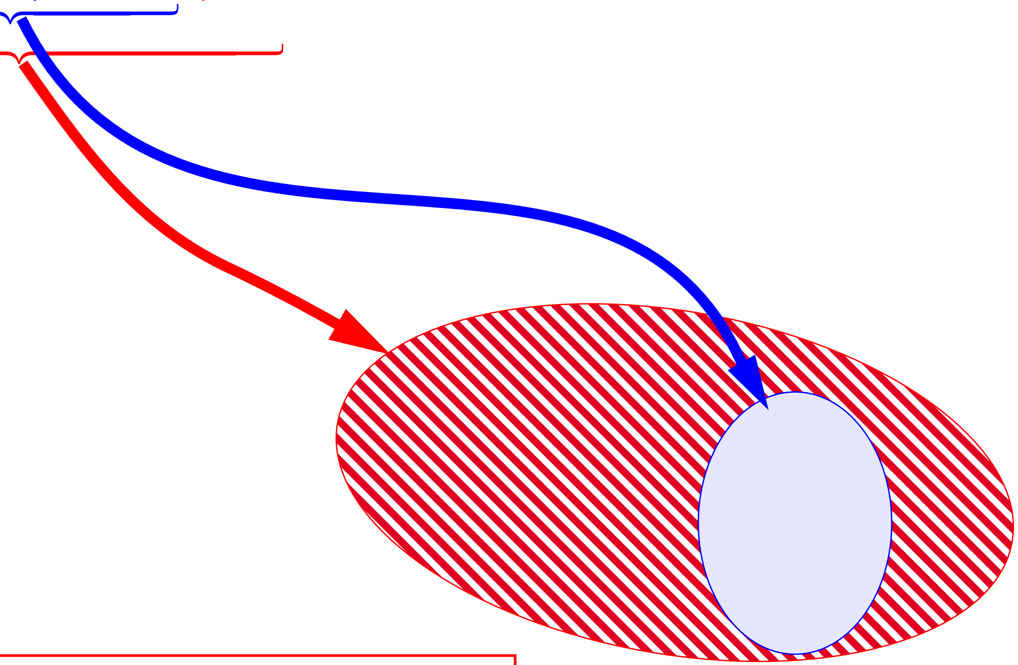


Uncertainty due to External Loads: (f)

- **Random Vibration**
- **Earthquake Engineering**
- **Ocean Engineering**
- **Weapons Applications:**
  Launch Shocks/Re-entry Loads,
  Penetration Loads,
  Hostile Environments

Anatomy of Global Uncertainty

$$E\{\ \underbrace{E[\boldsymbol{g}(\mathbf{U})|M,\mathbf{X}]}\ |M\}$$

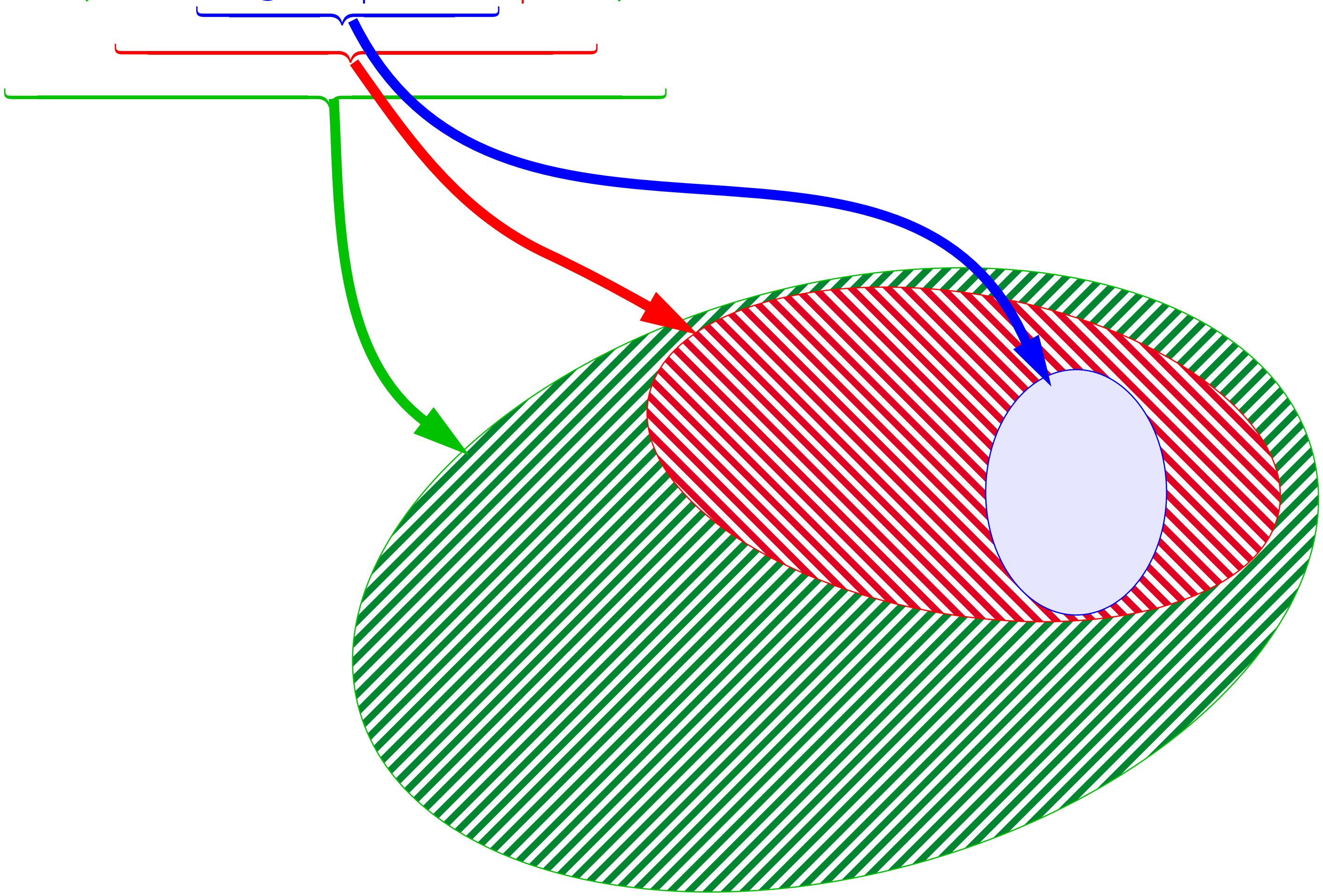


Uncertainty Propagation: $(\mathbf{X})$

- **Effects of parametric uncertainty: Intrinsic variabilities, Tolerances, Lack of repeatability**

Anatomy of Global Uncertainty

$$E\langle\ E\{\ \underbrace{E[g(\mathbf{U})|M,\mathbf{X}]}\ |M\}\ \rangle = E[g(\mathbf{U})]$$



Uncertainty Quantification at Sandia-NM

- **DAKOTA (Design Analysis Kit for OpTimizAion)/UQ**
  - Framework for multi-level, parallel computation: ASCI-level problems, optimization, nondeterministic analysis, response surface approximation, design of experiments, optimization under uncertainty
- **Polynomial Chaos and Stochastic Finite Elements**
  - Analysis of response of stochastic systems
- **Epistemic Uncertainty**
  - Non-Probabilistic Approach, Probabilistic Approach, Model Uncertainty
- **Sensitivity Analysis**

Objectives of Toolkit

Provide uncertainty quantification tools to the analyst community in a unified framework to be used in the design and certification processes.

- **Discipline independent**
- **ASCI (Accelerated Strategic Computing Initiative)-scale problems**
- **Minimize number of function evaluations**
- **Flexibility in uncertainty model**

Why tie UQ tools to the DAKOTA framework?

- **Existing, proven software framework**
- **Successfully linked with over 20 application codes**
- **Multilevel parallelism**
- **Extensive optimization algorithm library (gradient and non-gradient)**
- **Extensive selection of approximation strategies**

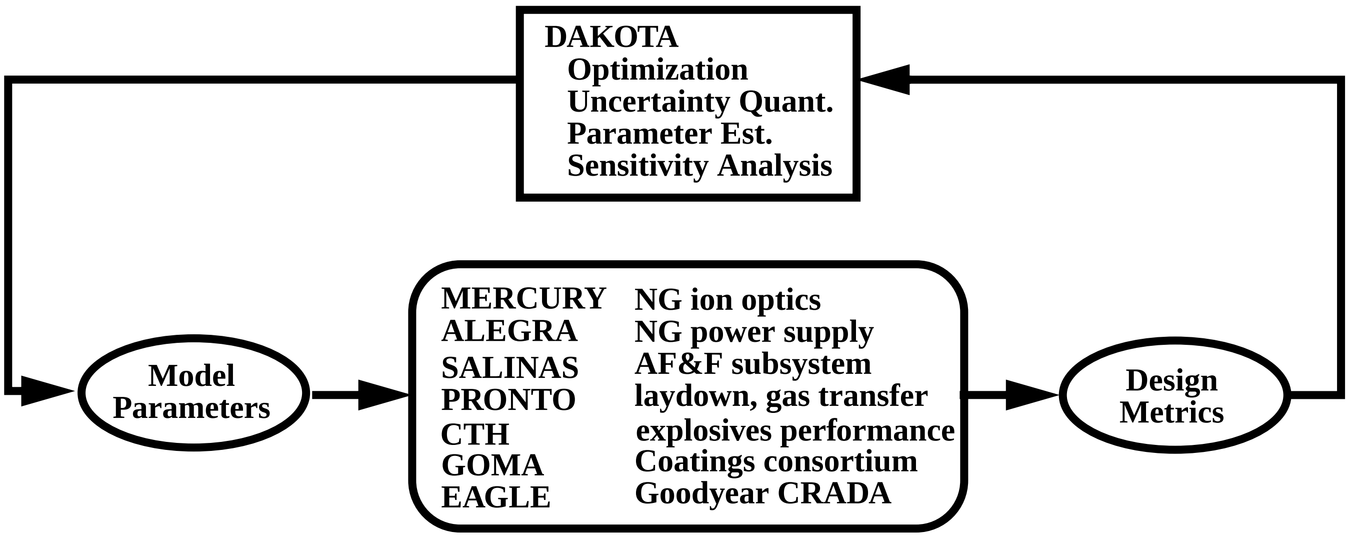
DAKOTA toolkit
Design optimization of engineering applications

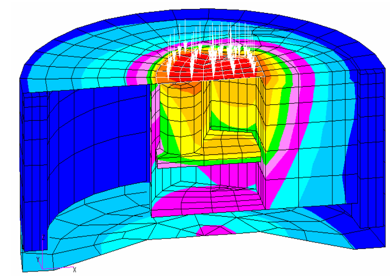
Sandia National Laboratories

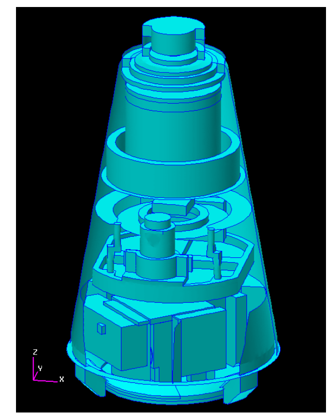
Engineering Sciences Center

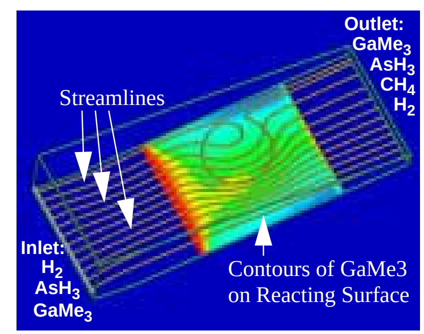
Answer fundamental engineering questions:

- What is the best design?
- How safe is it?
- How much confidence in my answer





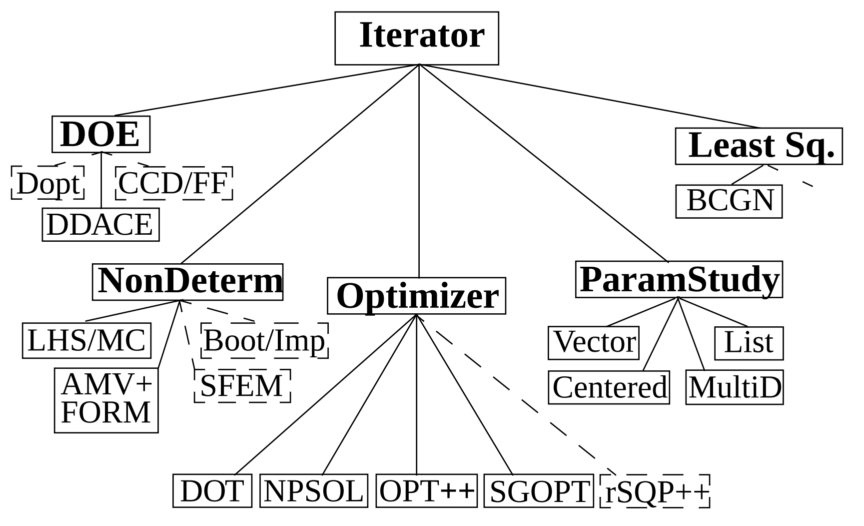




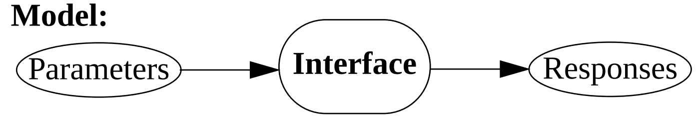
Additional motivations:

- Reuse tools and interfaces
- Leverage optimization, UQ, *et al.*
- Nonconvex, nonsmooth design spaces → state-of-the-art methodologies
- ASCI-scale applications and architectures → scalable parallelism
- Be a pathfinder in enabling M&S-based culture change at Sandia

Overview of DAKOTA framework



Model:



Design
- continuous
- discrete

Uncertain
- normal/logn
- uniform/logu
- weibull
- histogram

State
- continuous
- discrete

Application
- message passing
- system call, fork
  - synch & asynch
- direct
  - synch & asynch

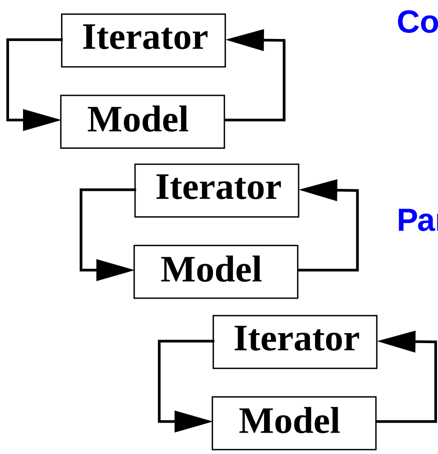
Approximation
- polynomial
- neural network
- splines
- kriging
- Taylor series
- hierarchical

Functions
- objectives
- constraints
- least sq. terms
- generic

Gradients
- numerical
- analytic
- mixed

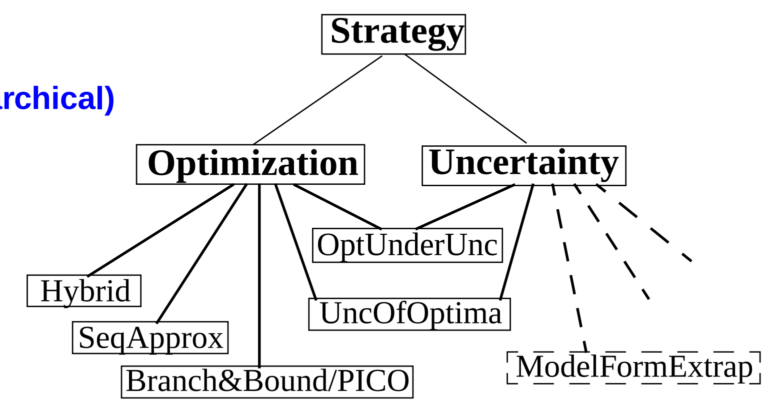
Hessians
- analytic

Strategy: control of multiple iterators and models



Coordination:
- Nested
- Layered (surrogates, hierarchical)
- Cascaded
- Concurrent
- Adaptive/Interactive

Parallelism:
- Master-slave (self-sched.)
- Peer (static, distr. sched.)
- 4 nested levels



Optimization/UQ Projects

DAKOTA project (optimization with engineering simulations):

Sandia manager - David Womble, 9211, dewombl@cs.sandia.gov, 845-7471
PI - Mike Eldred, 9211, mseldre@sandia.gov, 844-6479
Team members - Tony Giunta, Bill Hart, Bart van Bloemen Waanders
http://endo.sandia.gov/DAKOTA/

DAKOTA/UQ project (analytic reliability, sampling, and SFE UQ library):

Sandia manager - Martin Pilch, 9133, mpilch@sandia.gov, 845-3047
PI - Steve Wojtkiewicz, 9124, sfwojtk@sandia.gov, 284-5482
Team members - Mike Eldred, Rich Field, John Red-Horse, Angel Urbina

SGOPT project (stochastic global optimization):

Sandia manager - David Womble, 9211, dewombl@cs.sandia.gov, 845-7471
PI - Bill Hart, 9211, wehart@cs.sandia.gov, 844-2217
http://www.cs.sandia.gov/~wehart/main.html

PICO project (mixed integer programming, scheduling and logistics):

Sandia manager - David Womble, 9211, dewombl@cs.sandia.gov, 845-7471
PI - Cindy Phillips, 9211, caphill@cs.sandia.gov, 845-7296
Team members - Bob Carr, Jonathan Eckstein (Rutgers), Bill Hart, Vitus Leung
http://www.cs.sandia.gov/~caphill/proj/pico.html

OPT++/DDACE/APPS/IDEA projects (NLP, sampling, & pattern search libraries):

Sandia manager - Chuck Hartwig(acting), 8950, hartwi@ca.sandia.gov
PI - Juan Meza, 8950, meza@ca.sandia.gov, 294-2425
Team members - Paul Boggs, Patty Hough, Tamara Kolda, Leslea Lehoucq, Kevin Long, Monica Martinez-Canales, and Pam Williams
http://csmr.ca.sandia.gov/~meza/research.html

Current Dakota/UQ Capabilities

Sampling Techniques:

- **Random Sampling (Monte Carlo)**
- **Stratified Sampling (LHS (Latin Hypercube Sampling)**

Analytical Reliability Techniques:

- **Mean Value (MV), Advanced Mean Value (AMV/AMV+)**
- **FORM (First Order Reliability Method)/SORM (Second Order Reliability Method)**

Robustness Analysis

Stochastic Finite Element/ Polynomial Chaos Expansions

Response Surface Approximations:

- **Application of UQ tools to a surrogate function to minimize computational expense.**

Sampling Techniques

Monte Carlo-Style (Sampling-based) Analysis:

- **General, simple to implement and robust to size and discipline of problem being investigated**
- **Easily wrapped around current deterministic analysis capabilities**
- **Computationally expensive (many function evaluations)**
- **Two current options:**
  - –Traditional Monte Carlo
  - –Latin Hypercube Sampling
- **Under investigation:**
  - –Bootstrap Sampling
  - –Importance Sampling Techniques
  - –Quasi-Monte Carlo SImulation
  - –Markov Chain Monte Carlo

Overview of Analytically Based Reliability Methods

- Involve a transformation to unit variance, uncorrelated normal random variable space.
- Nataf Transformation used in DAKOTA/UQ.
- MV, AMV/AMV+, FORM all solve a constrained optimization problem where the objective function is always this minimum distance function with the constraint function depending on the method.
- MV and AMV/AMV+ work in the original random variable space.
- FORM/SORM work in the transformed space.
- Equivalent to Polynomial Response Surface Techniques about an “optimally” selected expansion point

Probabilistic Robustness Analysis

- **"Given the bounds on the input parameters, what range of output function is possible?"**

- **Pose two global optimization problems:**

$$g_{upper} = \max_{\mathbf{x}} g(M(f, \mathbf{x}))$$

$$g_{lower} = \min_{\mathbf{x}} g(M(f, \mathbf{x}))$$

such that

$$(x_i)_L \le x_i \le (x_i)_U \quad \forall \ i = 1 \ldots N$$

where N is the size of uncertain input vector, denote $\mathbf{x}_L$ and $\mathbf{x}_U$ its lower and upper bounds, respectively.

Probabilistic Robustness Analysis

Answer:

$$g(\mathbf{u}) \in [g_{lower}, g_{upper}]$$

- **Recently extended to mixed case of intervals and random variables of unknown dependence (to appear in Wojtkiewicz, AIAA SDM 2002)**

SFEM/Polynomial Chaos Techniques

- **Approximation of full stochastic representation**
- **Optimal approximation in inner product spaces,** L_2 **space of random variables.**
- **Represents a more general alternative to the Rosenblatt transforma-tion**
  - –avoid assuming full distribution when faced with limited input data
- **Estimating coefficients is the key issue**
  - –requires realizations of the function it replaces
- **Convergence issues**
  - –are there sufficient samples to compute coefficients?
  - –possibility of non-physical realizations
  - –mean square convergence

SFEM/Polynomial Chaos Techniques

Sandia National Laboratories

Engineering Sciences Center

- Consider PCE of general random process, u

$$u(x;\Phi) \approx u(x;\Phi)^{(P)} \equiv \sum_{i=0}^{P} u_i(x)\Gamma_i(\underline{\xi}), \text{ where } P = \sum_{s=1}^{q} \frac{1}{s!}\left\{\prod_{r=0}^{s-1}(m+r)\right\}$$

–qth order polynomial in $\underline{\xi}$, where $\underline{\xi} = \begin{bmatrix}\xi_1 & \xi_2 & \dots & \xi_m\end{bmatrix}^T$

–function of m underlying random variables

- Solve for the Fourier coefficients, $u_i(x)$

$$u_i(x) = \frac{\langle u(x, h[\underline{\xi}])\ \Gamma_i(\underline{\xi})\rangle}{\langle \Gamma_i^2(\underline{\xi})\rangle} = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} u(x, h[\underline{\xi}])\ \Gamma_i(\underline{\xi}) f_{\underline{\xi}}(\underline{\xi}) d\underline{\xi}}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} \Gamma_i^2(\underline{\xi}) f_{\underline{\xi}}(\underline{\xi}) d\underline{\xi}} = \frac{n_i(x)}{\delta_i(x)}$$

$\delta_i(x)$ can be solved in closed-form

Epistemic Uncertainty

Sandia National Laboratories

- Epistemic Uncertainty results from a lack of information.

- Epistemic Uncertainty manifests itself in several ways

  - Uncertainty in parameters for which statistically significant databases do not exist
  - The form of the model is not known exactly

Non-Probabilistic Approach

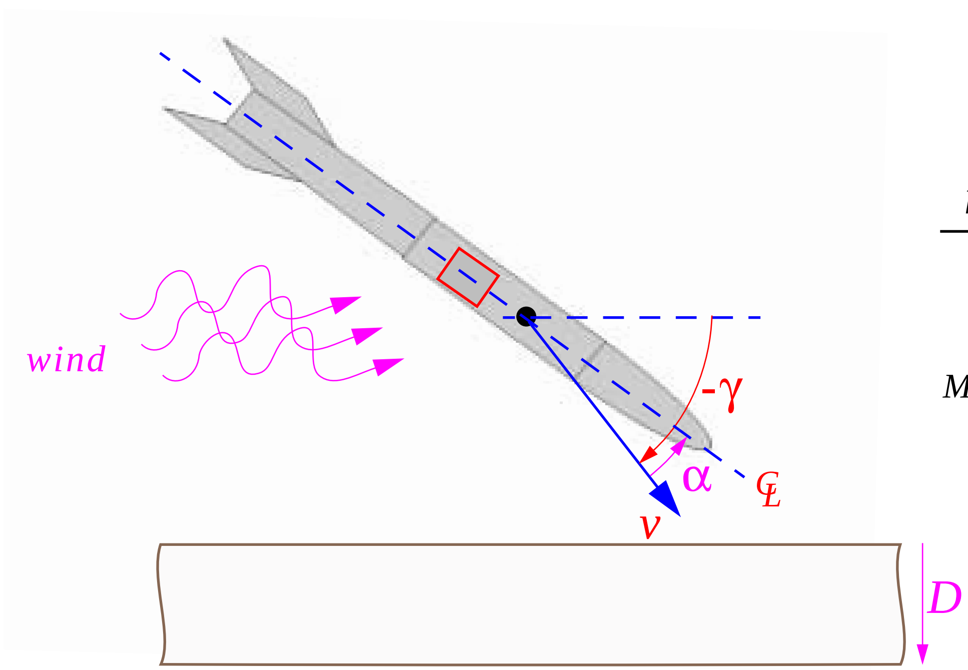
- Variety of approaches investigated:

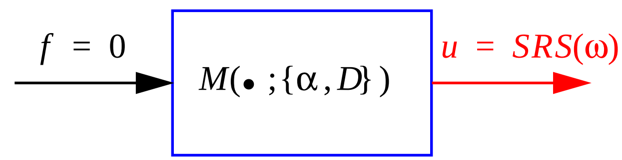
  - Interval analysis
  - Possibility Theory
  - Evidence Theory (Dempster-Shafer)
  - Imprecise Probability
  - Probability Bounds
  - Interval-valued Probability Distributions
  - Convex Sets of Probability Distributions

The Penetrator Problem

- **Problem statement**

–During the penetration event, predict the probability of component failure, P_f





M ~
ASCI FE codes
- *structure - nonlinear transient dynamics*
- *boundary - nonlinear soil mechanics*
- *SRS - nonlinear filter*

<u>Uncertain parameters</u>
- α, angle-of-attack
- D, soil parameter

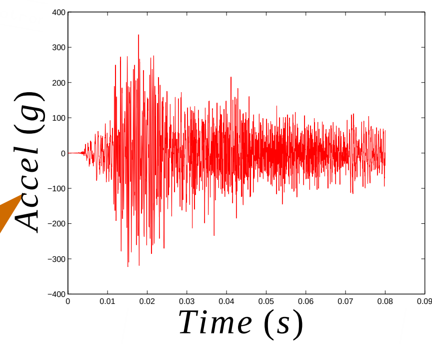
–Consider a nonlinear, full-body, 3D, coupled-physics simulation with simplified probabilistic properties.

The model, *M*: a complex, cascaded system

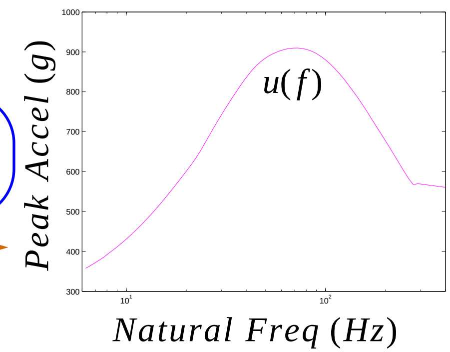
Sandia National Laboratories

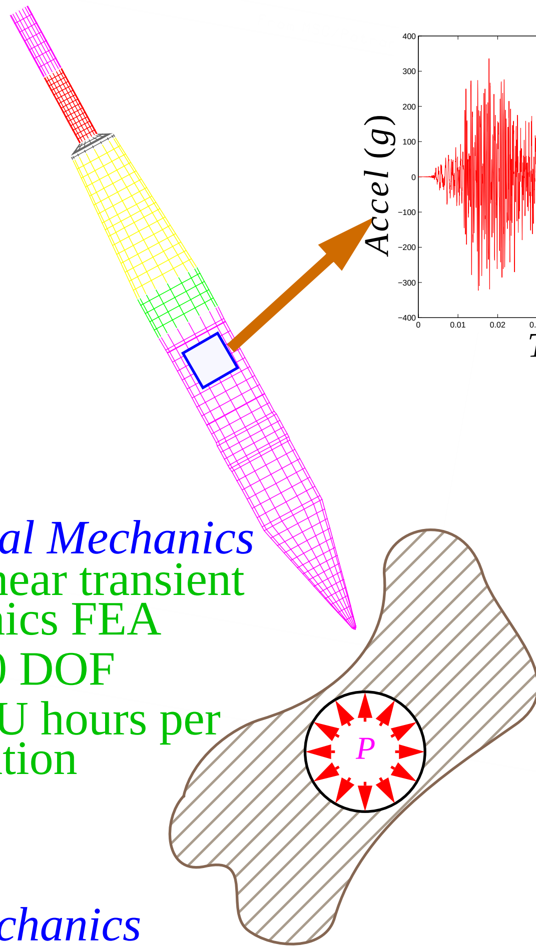
Engineering Sciences Center

Component Response



Filters
LP, SRS





Structural Mechanics

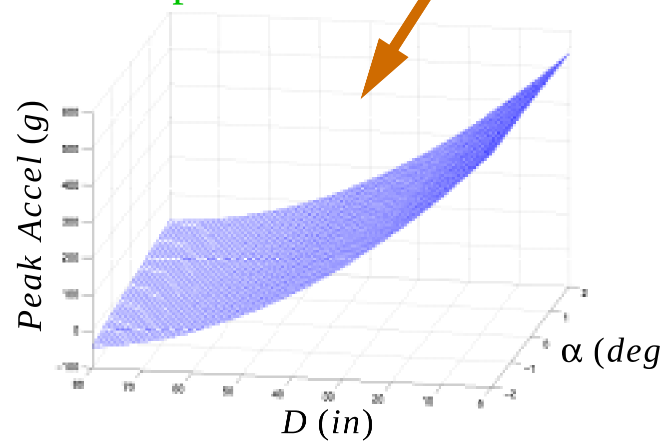
- ✓Nonlinear transient dynamics FEA
- ✓50,000 DOF
- ✓33 CPU hours per simulation

Soil Mechanics

- ✓Spherical cavity expansion
- ✓Loads couple with mechanics simulation

Approximate RS Models

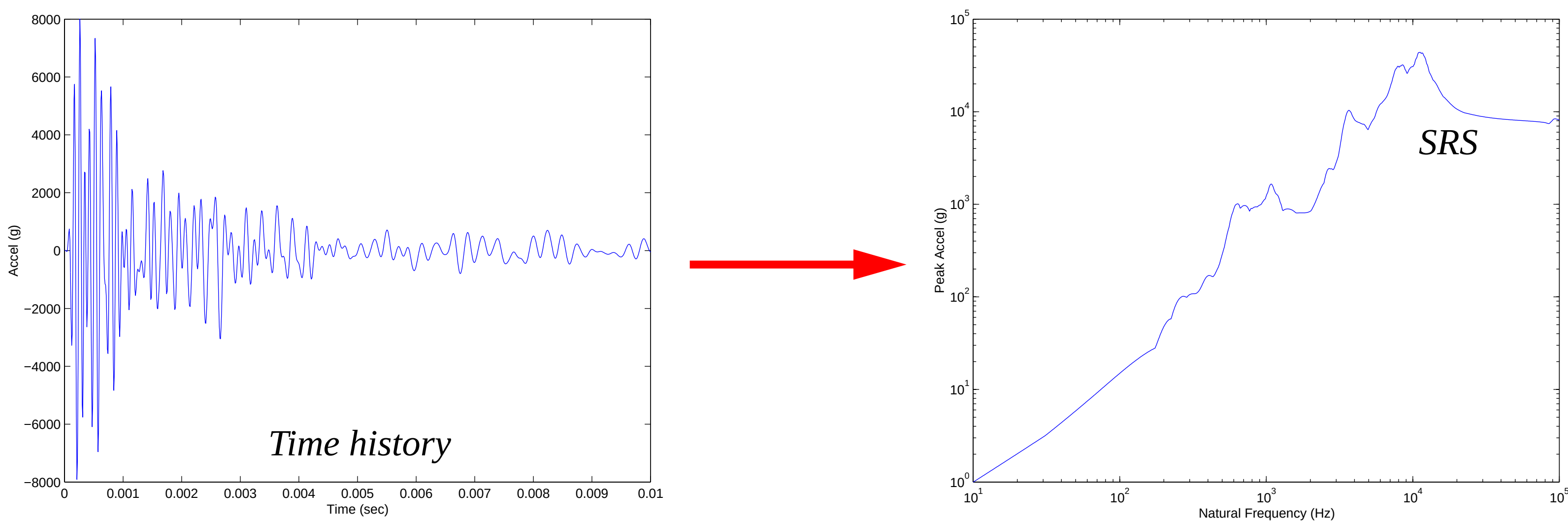
- ✓Performed simultaneously on network of Sun Ultra IIs
- ✓49 total runs performed



Overview of the Shock Response Spectrum (SRS)

Why use SRS?

- measure of shock severity; indicative of shock damage potential
- frequency-domain representation of shock response
- long history of use in weapon design; test-based spec
- used for component qualification - compare to SRS_{ref}



UQ analysis of Penetrator System

Sandia National Laboratories

- **Two design variables, X:**
  - –α, angle of attack is a normal random variable with mean 1 and standard deviation of 1.
  - –D, soil depth is a lognormal random variable with mean 25 and standard deviation 16.

$$\bar{u} = \min_i \left(SRS_{ref}(f_i) - SRS(f_i)\right)$$

$$Z = g(\bar{U}) = I(\bar{U})$$

$$P_f = P(Z \le 0) = 1 - E[g(\bar{U})]$$

- **Using the results from simulations, build a approximate model**

(response surface approximation) for $\bar{u}$.

UQ analysis of Penetrator System

- **Apply MC/LHS to these surrogate models to evaluate $F_Z(0)$:**

$$N_s = 1\times10^4 \text{ and } N_s = 5\times10^6$$

Response Surface Approximation Method	*MC*	*LHS*
Kriging	0.02000/0.02300	0.02000/0.02400
Splines	0.06900/0.06781	0.06720/0.06767
Neural Net	0.05024/0.05588	0.05500/0.05581
Quadratic Polynomial	0.04960/0.05077	0.05070/0.05071

Summary

Current Capabilities

Analytical Reliability Techniques:

- MV, AMV, AMV+, FORM/SORM

Sampling techniques:

- Pure Random Sampling (Monte Carlo)
- Stratified Sampling (LHS)

Probabilistic Robustness Analysis

Polynomial Chaos Expansions/Stochastic Finite Element Techniques

Future Capabilities:

Enhanced sampling methods:

- Importance Sampling, Bootstrap Sampling,

  Quasi-Monte Carlo Sampling, Markov Chain Monte Carlo Sampling

Non-traditional uncertainty methodologies