Challenges and Approaches for Simulation-Based Optimization Problems

Juan Meza

Lawrence Berkeley National Laboratory

Mónica Martínez-Canales
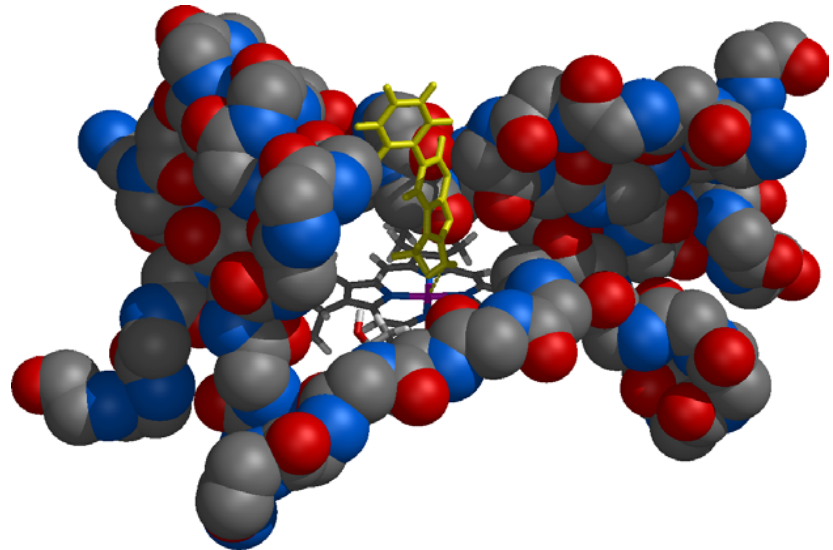
Sandia National Laboratories

Acknowledgements

- ❖ Leslea Lehoucq
- ❖ Kevin Long
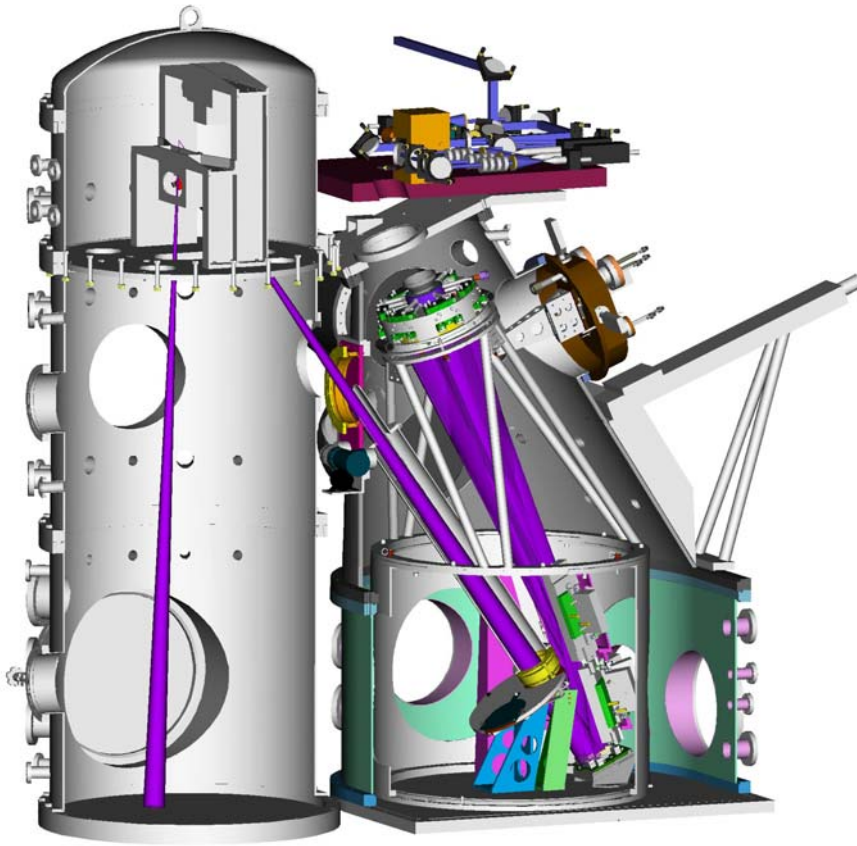- ❖ Patty Hough
- ❖ Pam Williams
- ❖ Chris Moen

Optimization problems arise in a wide variety of applications

Extreme Ultraviolet Lithography
10x10x10-foot engineering test stand

Target problem was parameter identification for extreme UV light source model

$$\min_{x} \quad \sum_{i=1}^{N} (T_i(x) - T_i^*)^2$$

$$\text{s.t.} \quad 0 \leq x \leq u$$

❖ Find model parameters, satisfying some bounds, for which the simulation matches the observed temperature profiles

❖ Computing objective function requires running thermal analysis code

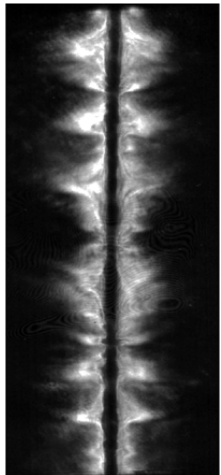# Z-Pinch Machine: matching simulations with experiments
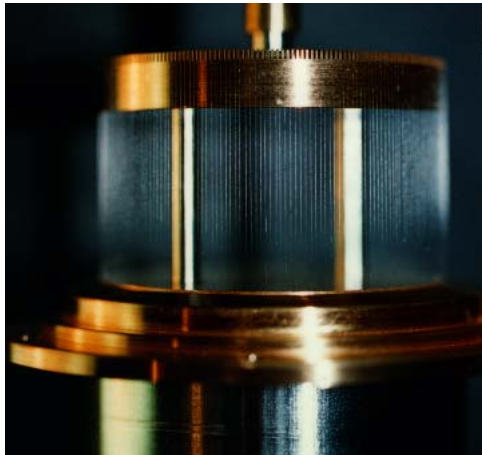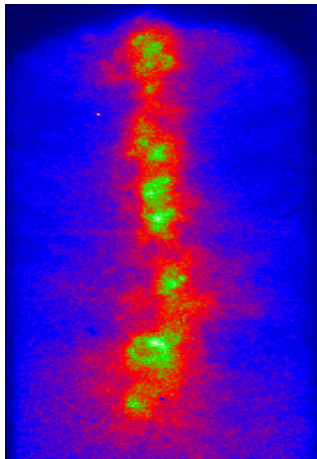

Wire Array for Z-machine

Goals:
- ❖ Improved models of the Z-machine
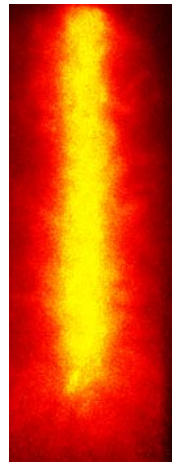- ❖ Optimize new designs

Current Focus:
- ❖ Methods for design with uncertainty

Wire Initiation

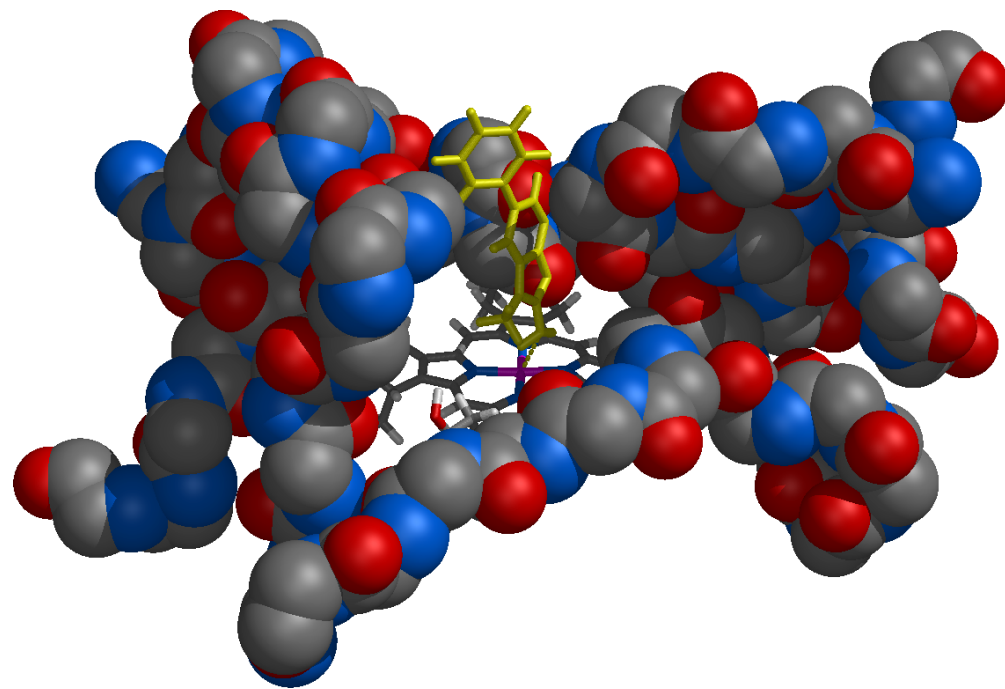**Load Implosion**

Load Stagnation

Developing New Drugs: an energy minimization problem

Docking model for environmental carcinogen bound in *Pseudomonas Putida* cytochrome P450

❖ A single new drug may cost over $500 million to develop and the design process typically takes more than 10 years

❖ There are thousands of parameters and constraints

❖ There are thousands of local minima

Example: Model-based Safety Assessments

- ❖ Problem: model accident scenarios to determine the worst-case response

- ❖ Challenges:
- ❖ Simulation of coupled sub-systems
- ❖ Need a family of solutions
- ❖ Characterize uncertainty in design safety

We have chosen to focus on particular classes of nonlinear optimization problems

❖ Expensive function evaluations
» CPU time is measured in hours (even on parallel computers)

❖ Variable digits of accuracy
» Usually a result of solving a PDE

❖ Gradient information is (usually) not available

❖ Small dimensional
» Number of variables ~ 10 - 100

Schnabel (1995) identified three levels for introducing parallelism into optimization

1. Parallelize evaluation of functions, gradients, and or constraints

2. Parallelize linear algebra

3. Parallelize optimization algorithm at a high level

Basic idea is to solve a nonstandard Trust-Region subproblem using PDS (TRPDS)
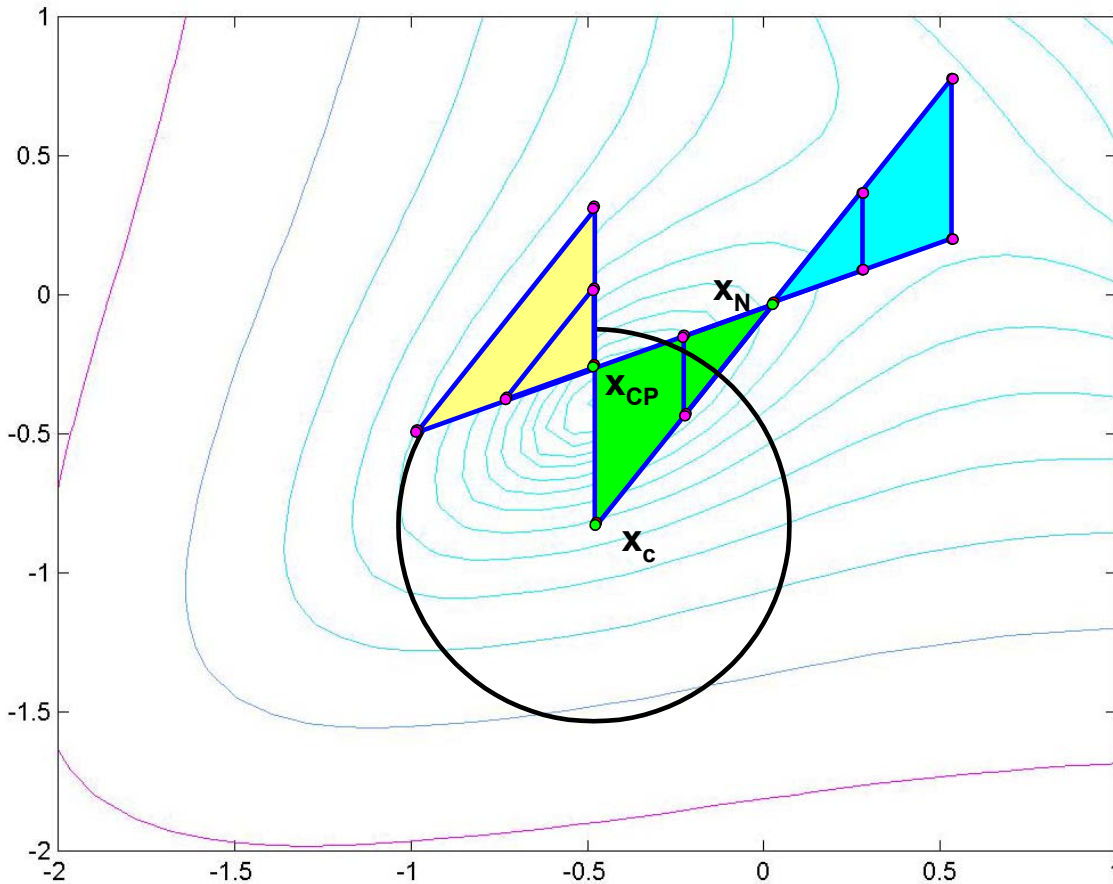
- ❖ Fast convergence properties of Newton method
- ❖ Good global convergence properties of trust region approach
- ❖ Inherent parallelism of PDS
- ❖ Ability to handle noisy functions

General statement of TRPDS algorithm

Given x_0, g_0, H_0, δ_0, and η

for k=0,1, … until convergence do

    1.   Solve $H_k s_N = -g_k$

    for i=0, 1, … until step accepted do

        2. Form initial simplex using s_N

        3. Compute s that approximately minimizes $f(x_k + s)$,   subject to trust region constraint

        if *ared/pred* > η **then**

            5. Set $x_{k+1} = x_k + s$; Evaluate g_{k+1}, H_{k+1}

        endif

        6. Update δ

    end for

end for

Convergence of TRPDS follows from theory of Alexandrov, Dennis, Lewis, and Torczon (1997)

❖ Assume

» Function uniformly continuously differentiable and bounded below; Hessian approximations uniformly bounded

» Approximation model satisfies the following conditions:
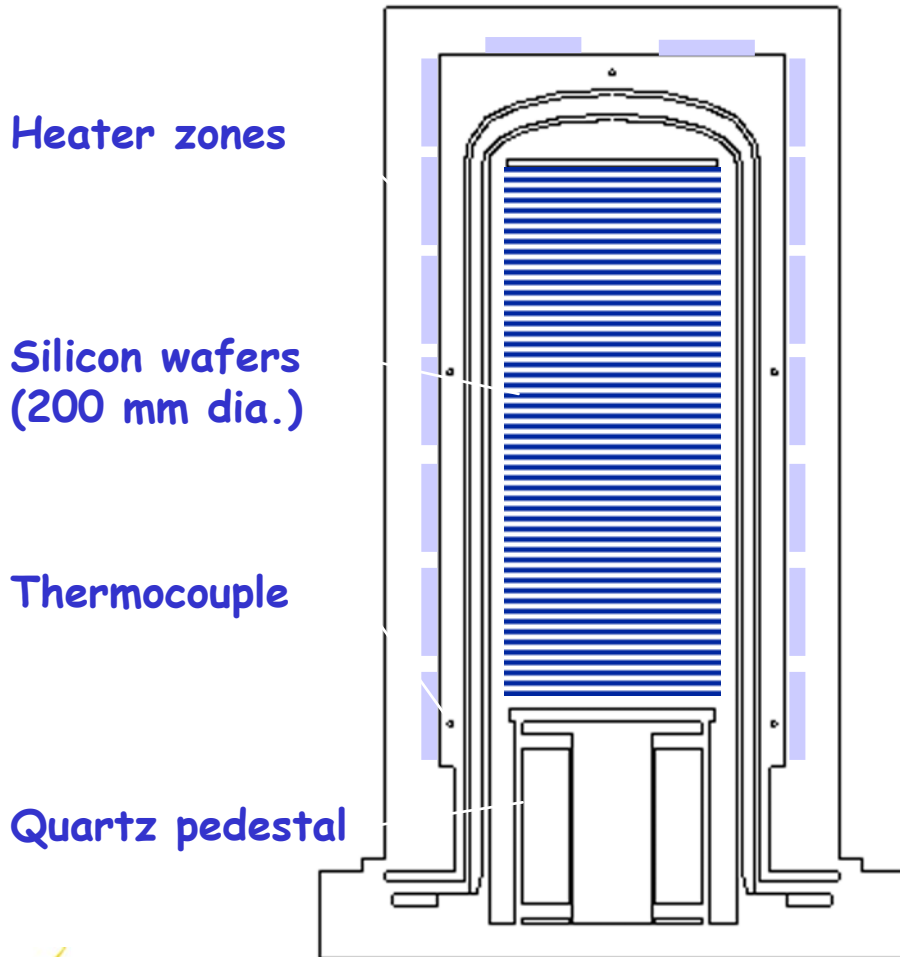
$$1. \quad a(x_k) = f(x_k)$$
$$2. \quad \nabla a(x_k) = \nabla f(x_k)$$

» Steps satisfy fraction of Cauchy decrease condition

❖ Then

» $\lim\inf\limits_{k \to \infty} \| \nabla f(x_k) \| = 0$

An application of TRPDS to the optimization of the performance of LPCVD furnaces

Heater zones

Silicon wafers (200 mm dia.)

Thermocouple

Quartz pedestal

❖ Temperature uniformity is critical
  » between wafers
  » across a wafer

❖ Independently controlled heater zones regulate temperature

❖ Wafers are radiatively heated

Computing the objective function requires the solution of a PDE
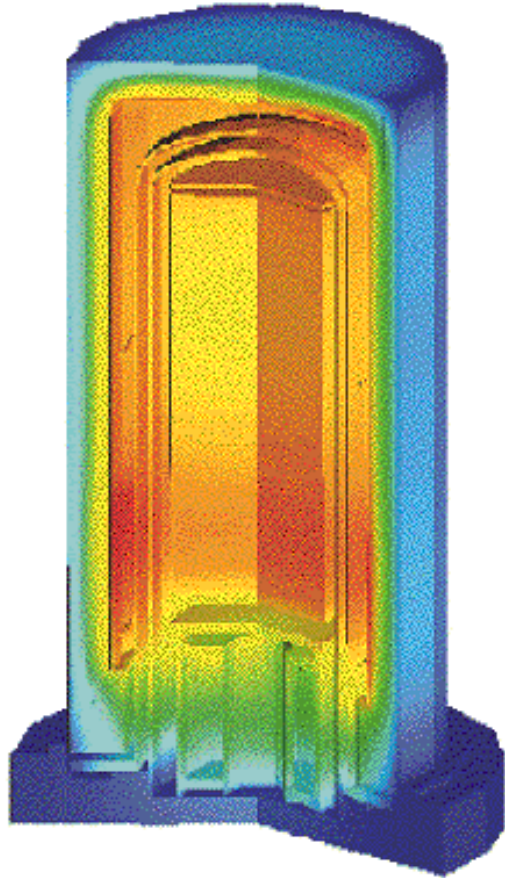
Temperature fields in a vertical, stacked-wafer, low-pressure, chemical-vapor-deposition furnace

- ❖ Finding temperatures involves solving a heat transfer problem with radiation

- ❖ Two-point boundary value problem solved by finite differences

- ❖ Adjusting tolerances in the PDE solution trades off noise with CPU time

  - » Larger tolerances lead to
    - – Less accurate PDE solutions
    - – Less time per function evaluation

The goal is to find heater powers that yield optimal uniform temperature

$$\min\ F(\mathbf{p}) = \sum_{i=1}^{N}\ (T_i(\mathbf{p}) - T^*)^2,$$
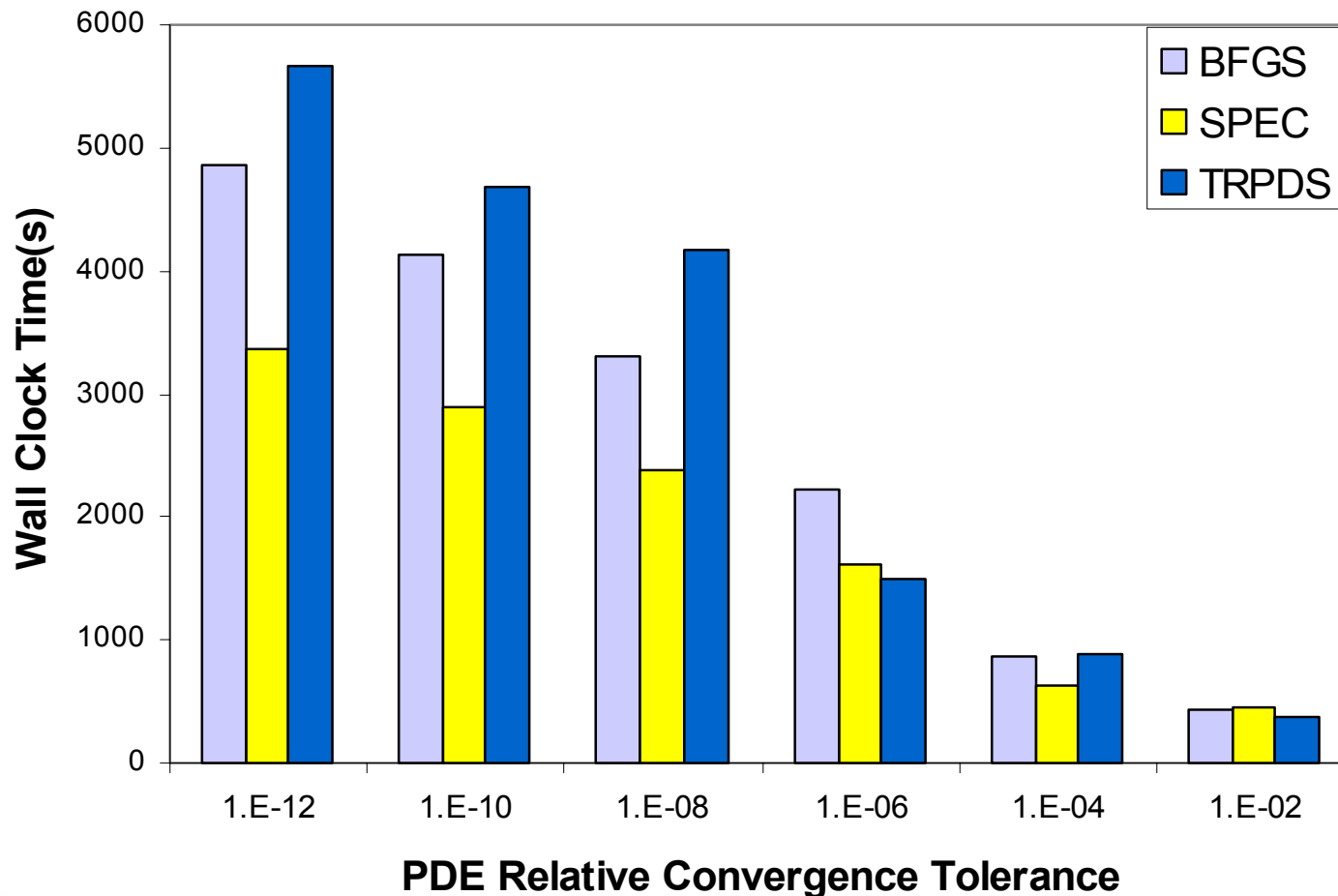
where $\mathbf{p}$ is the vector containing the heater powers,

$T_i(\mathbf{p})$ is the temperature at discretization point i given powers $\mathbf{p}$,
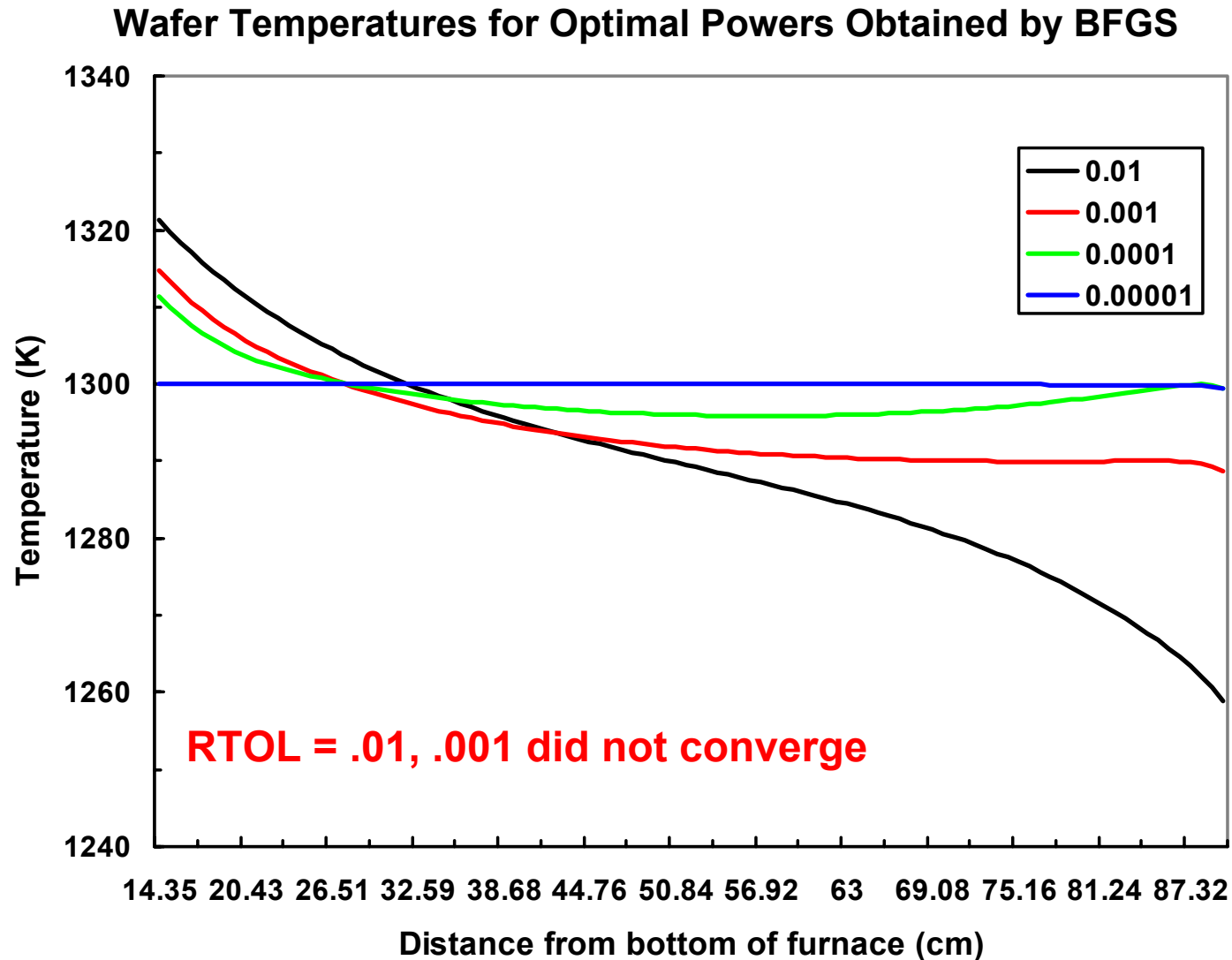
T^* is the target temperature, and

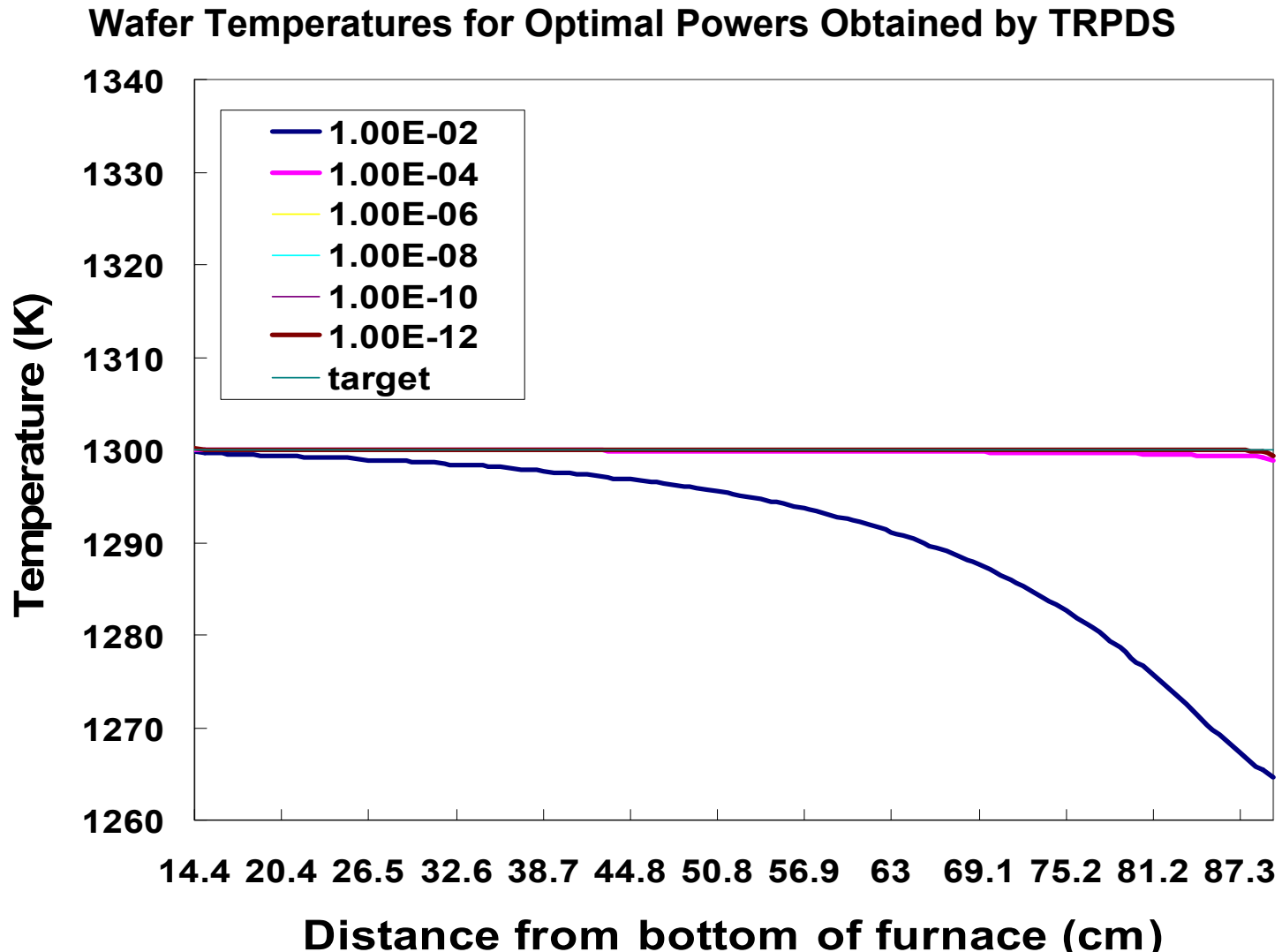N is the total number of discretization points

TRPDS becomes more competitive with standard methods as accuracy decreases

PDE Relative Convergence Tolerance

BFGS may not converge when simulations have fewer digits of accuracy

Wafer Temperatures for Optimal Powers Obtained by BFGS

RTOL = .01, .001 did not converge

TRPDS is more robust than standard methods when we have fewer digits of accuracy

Wafer Temperatures for Optimal Powers Obtained by TRPDS

Why Uncertainty Quantification?

"As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality"

Albert Einstein

Major goal is to develop new techniques for quantifying uncertainty in computer simulations


Experimental errors versus simulation errors
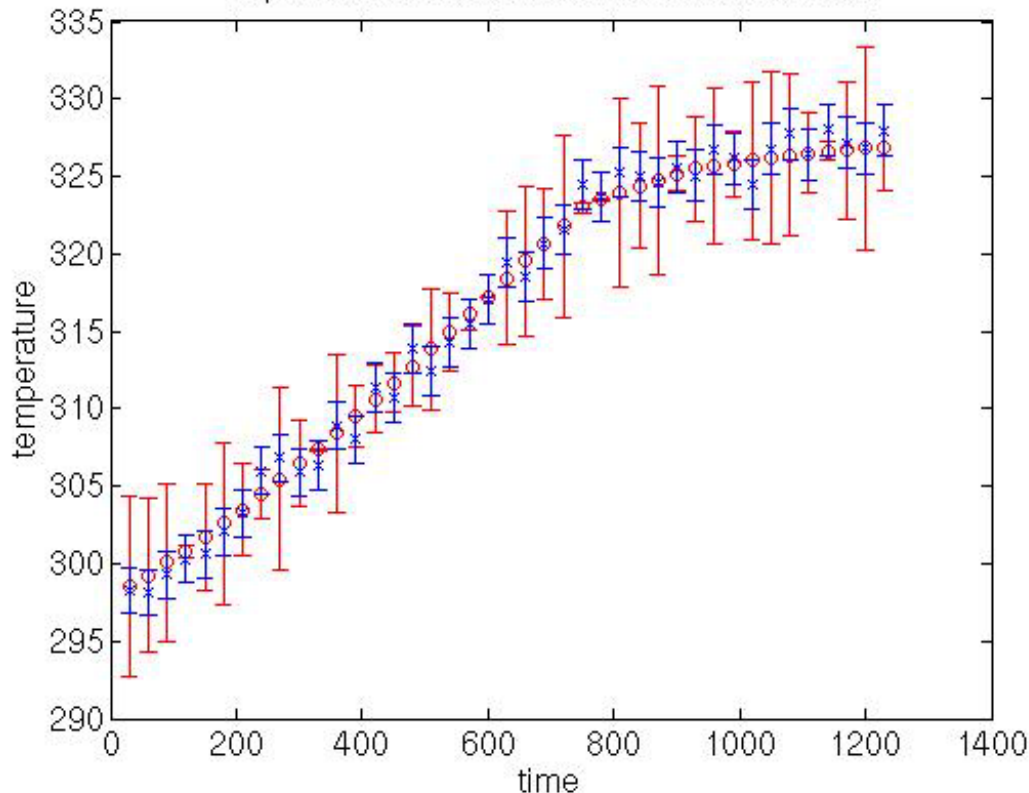
- ❖ Develop fast algorithms for computing uncertainty (error bars) in simulation results

- ❖ Implement parallel versions of algorithms

- ❖ Coordinate efforts with other UQ projects: Sphynx, DDace, OPT++, Dakota

EUVL Lamp model and experimental data
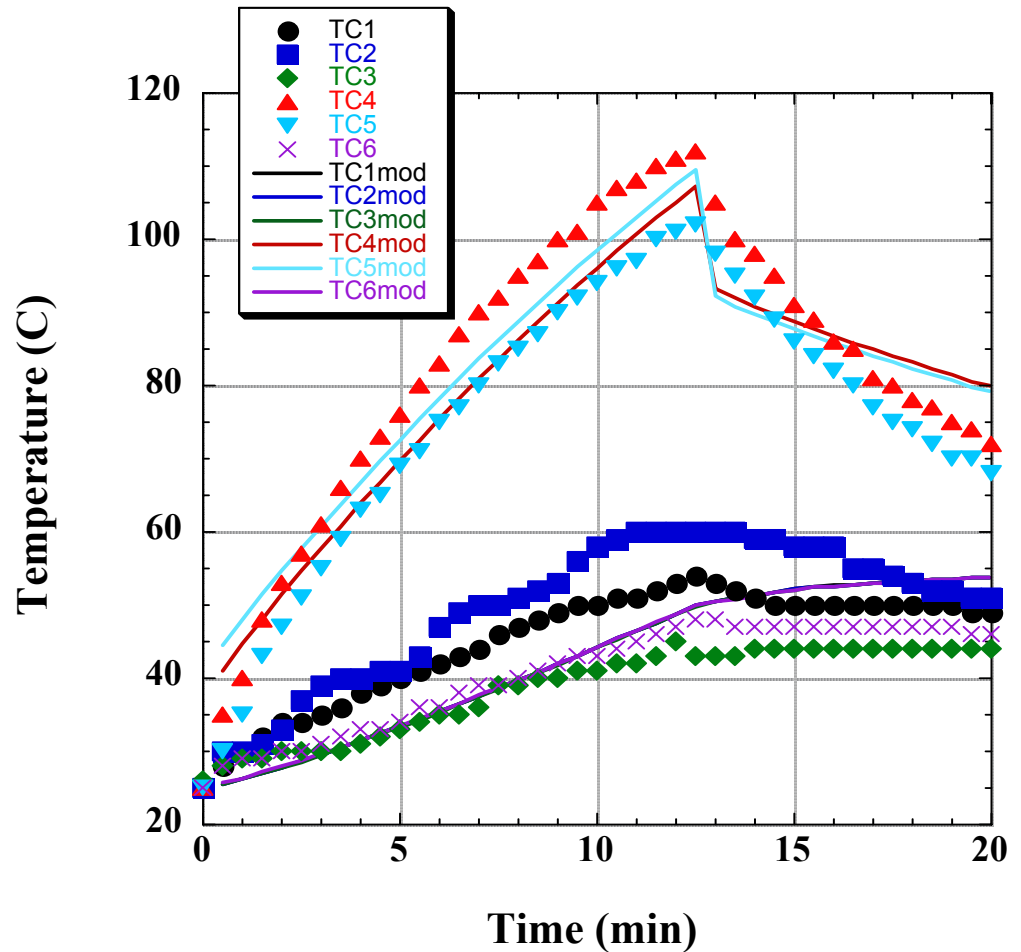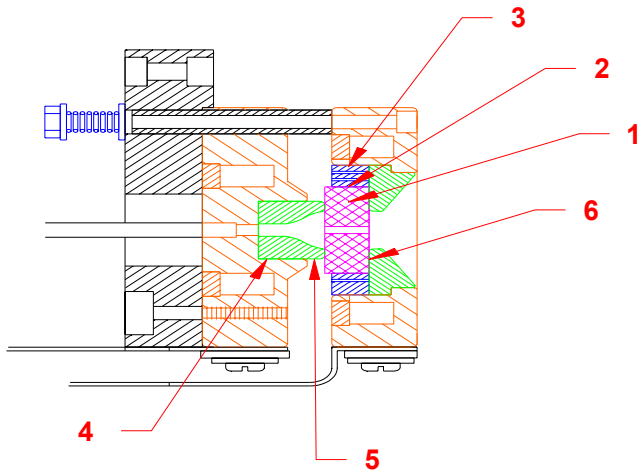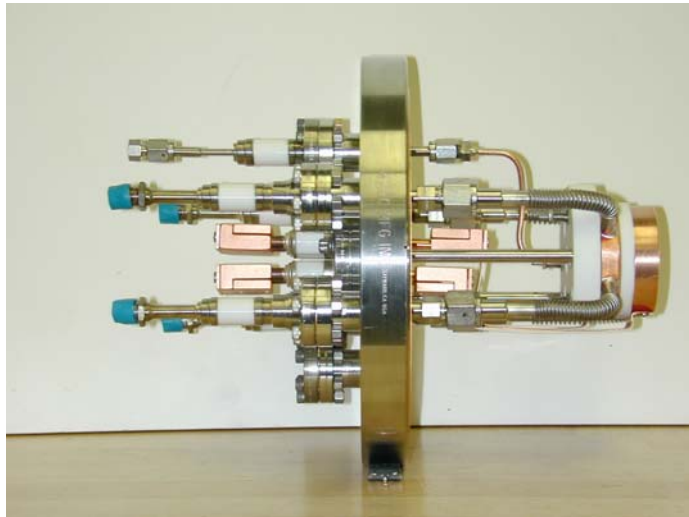
This optimization problem requires intensive computational resources


Maximum Temperature Difference
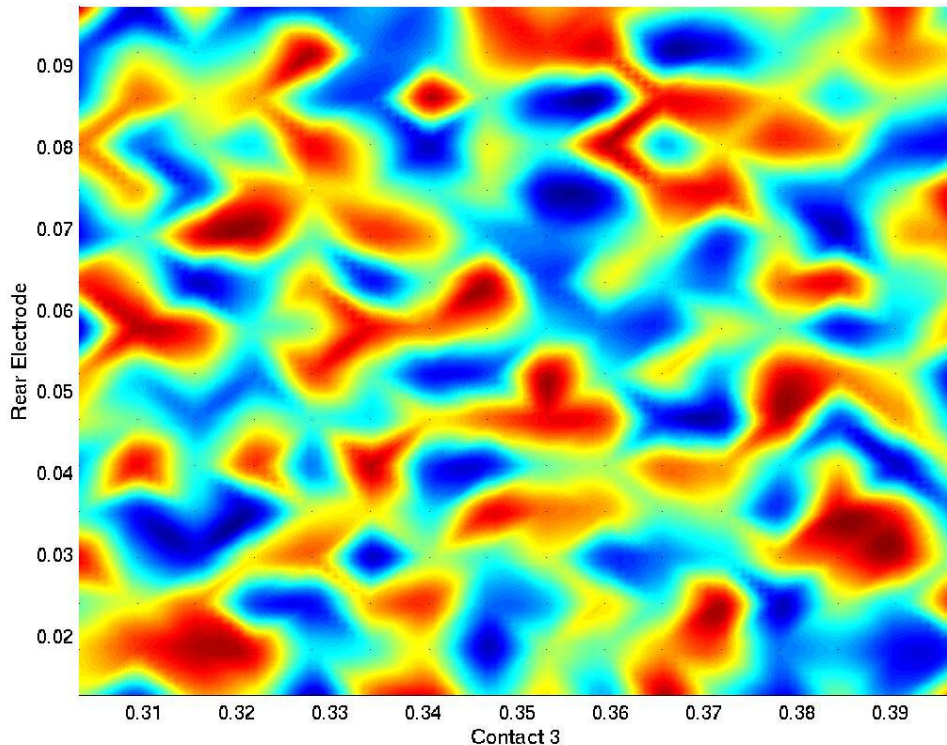
- ❖ Objective function consists of computing the maximum temperature difference over all 5 curves
- ❖ Each simulation requires approximately 7 hours on 1 processor
- ❖ The objective function has many local minima

Types of questions we would like to ask

- ❖ Which parameters are the "most important"?
- ❖ How sensitive are the simulation results to the parameters?
- ❖ Can we quantify the variability of our simulation results?
- ❖ For a given confidence interval how many simulations runs do I need to run?
- ❖ Can I build a reduced model that approximates the simulation?
- ❖ ....

Statistical analysis can yield insight into the behavior of the simulation
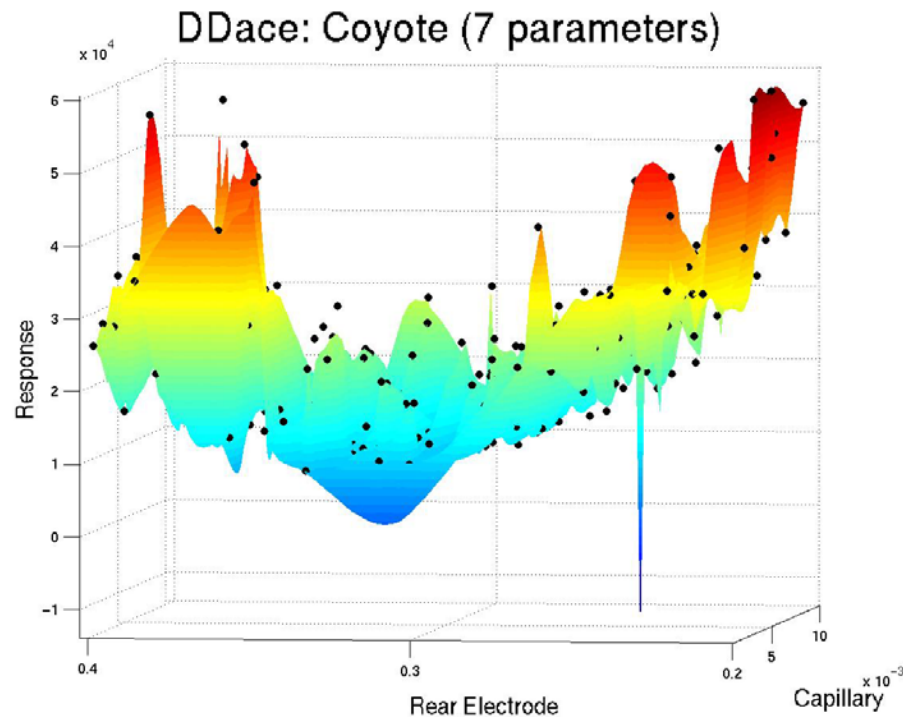
DDace results of LHS on the EUVL lamp model
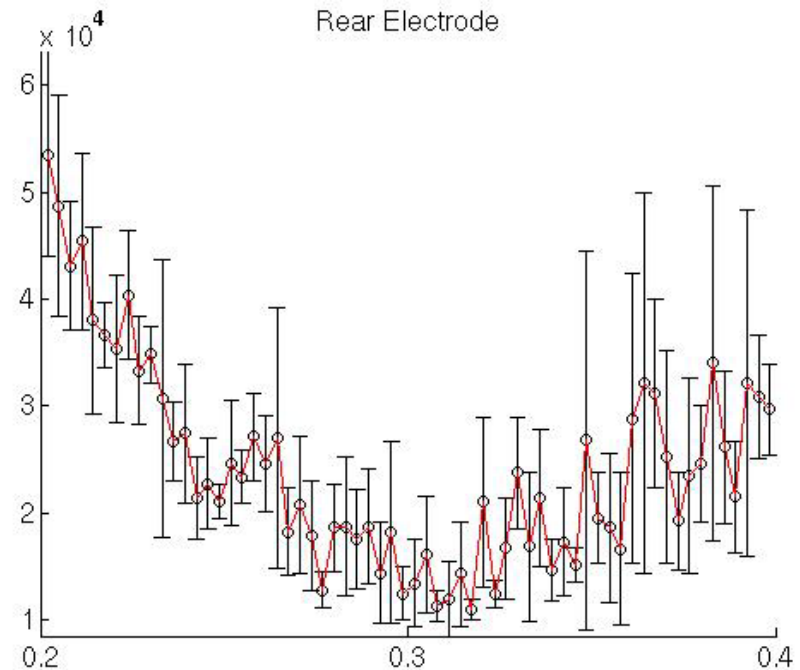
Mean and standard deviation of simulation results holding all but one parameter fixed

Global Sensitivity Analysis

Pearson's correlation ratio can be used as a measure of importance of a subset of the parameters

❖ Compute Pearson correlation ratio:

$$Corr = V(y^s)/V(y)$$

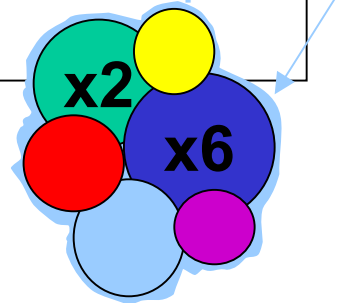V(y) = model prediction variance of y=model(x)

V(y^s) = restricted model prediction variance of y^s=E(y|x^s), the model prediction based on the parameter subset x^s.

❖ McKay et al 0th-iteration estimate of Pearson correlation ratio:

$$Corr(x_j) = SSB0/SST0$$

Unity Correlation Space

x2

x6

$$SSB0 = \sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{i\bullet} - \bar{y})^2 \qquad SST0 = \sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij} - \bar{y})^2$$

$$\bar{y}_{i\bullet} = \frac{1}{J}\sum_{j=1}^{J}y_{ij} \qquad \bar{y} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}y_{ij}$$

For the EUVL model, correlation ratios suggest that parameters 2 and 6 are more important
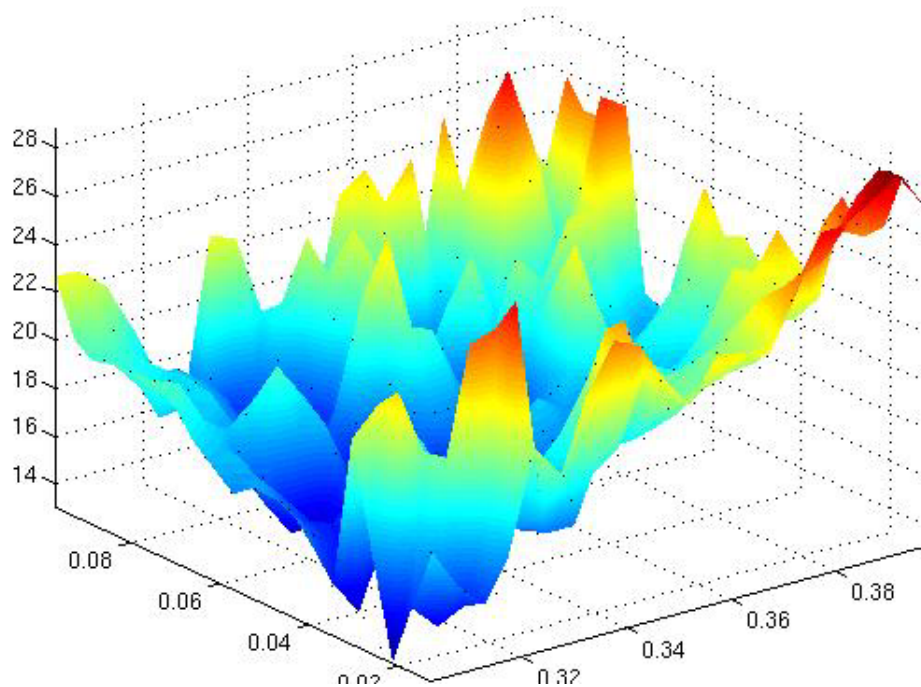
	Correlation	*Mean*	*Standard Deviation*	*Sensitivity*
x1	0.12	14.60	0.01	5.70E-06
x2	0.37	18.60	1.10	5.30E-06
x3	0.17	15.70	1.55	7.20E-07
x4	0.14	14.70	0.79	1.10E-05
x5	0.12	14.60	0.01	4.10E-04
x6	0.39	16.80	3.18	7.20E-06
x7	0.10	14.60	0.02	4.60E-05

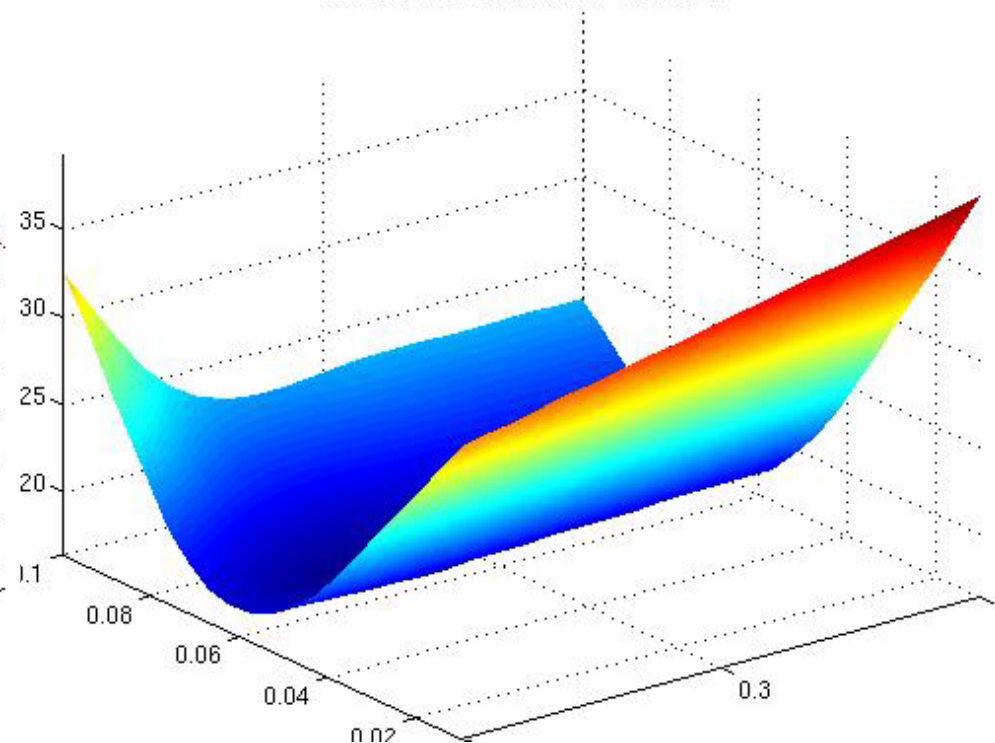x2 = heat flux to rear electrode
x6 = conductivity of contact 3

Model reduction captures trends but not variability
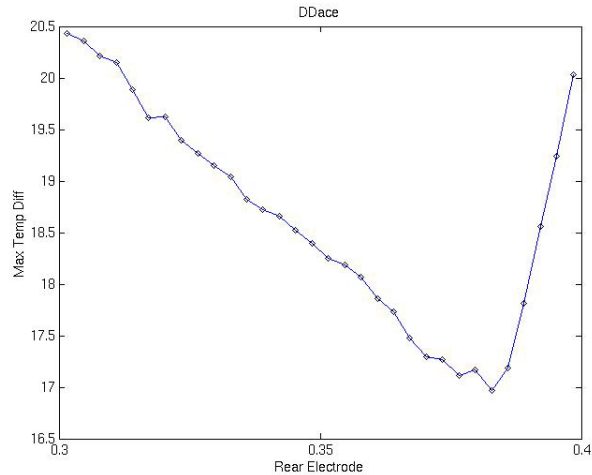

DDace/Coyote output results of EUVL model

MARS (Multi-variate Additive Regression Splines) response surface
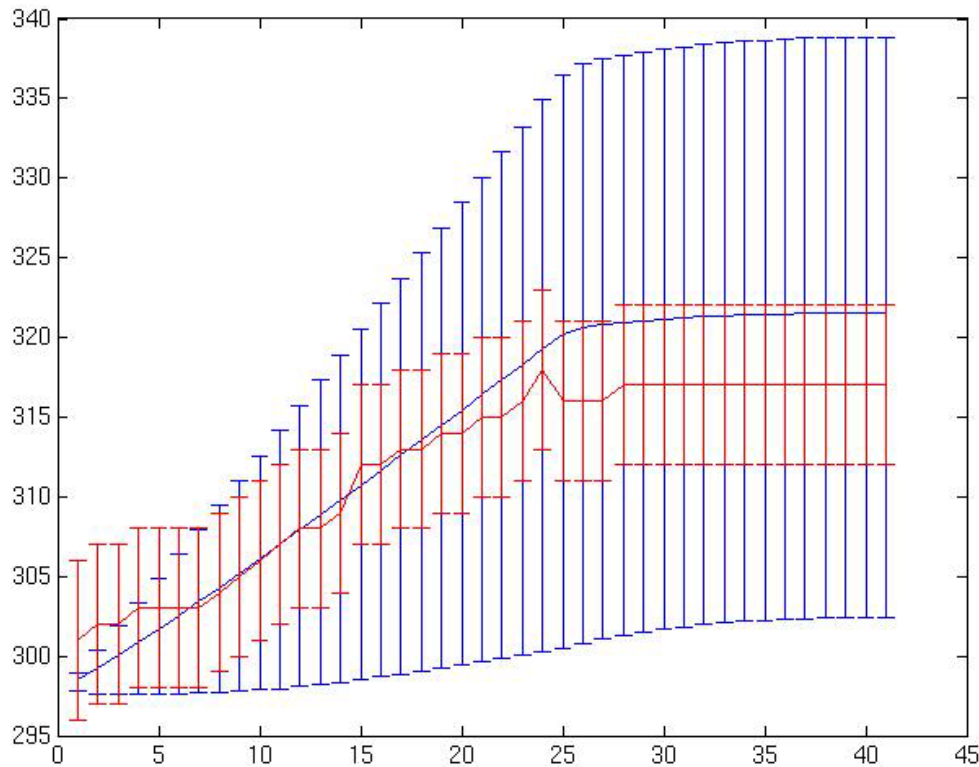
A Taste of Things to Come

The objective function is still offering us many challenges

- ❖ Some of the optimization parameters are well-behaved - others exhibit nastier behavior
- ❖ Computation of finite-difference gradients can be tricky
- ❖ Main effects analysis can be used to restrict the parameter space of interest
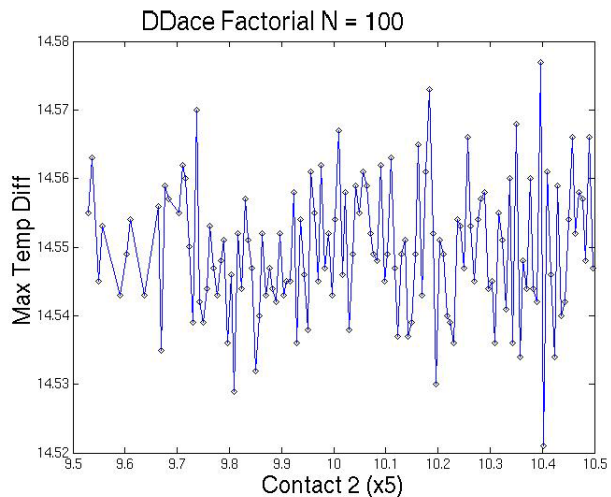
OShI – Overlap Shortfall Index

**Experimental Data vs. DDACE Simulation Results on EUVL Lamp Model**

❖ OShI is an index between 0 and 1. The closer to 1, the greater the overlap of the simulation and data ranges.

❖ OShI measures how well simulation output matches experimental data.

❖ OShI is also a mathematical measure

Summary and Future Work

- ❖ New class of parallel optimization methods
  - » Parallelism of pattern search combined with the good convergence properties of Newton methods
  - » Competitive with standard methods
- ❖ Greater robustness in applications that contain variable accuracy objective functions
- ❖ Develop methods for handling uncertainty in models and algorithm

The End

Summary

Confidence / Cost / Uncertainty chart showing "Improved tools" and "Current tools" curves
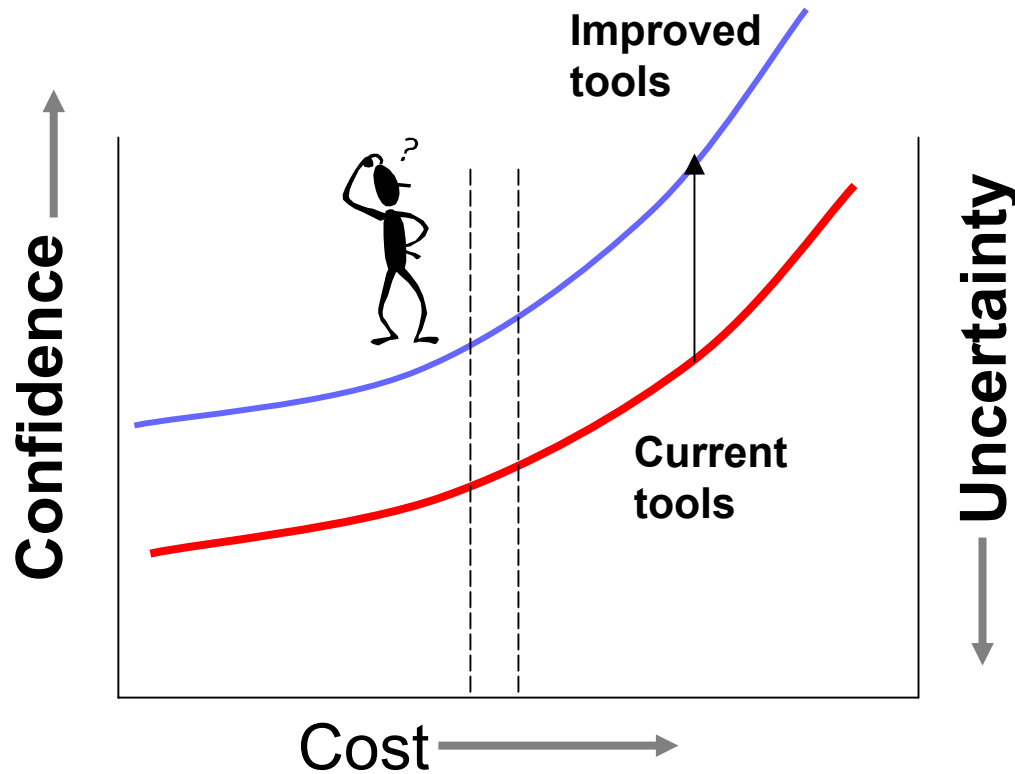
- ❖ UQ tools already being applied to some prototype problems
- ❖ UQ will help analysts make better decisions in the face of uncertainty
- ❖ Working towards more effective and easy to use decision support tools

Stochastic Response Surface

❖ Use Polynomial Chaos Expansions to construct a Stochastic Response Surface (SRS)

❖ Compare Response Surface Models:
  » MARS (currently in DDACE)
  » SSANOVA (R statistical package library)
  » Kriging (new capability to be added to DDACE)
  » SRS (new capability to be added to DDACE)

What we really need is a measure of the variability in the simulation

DDace results on EUVL model with 256 LHS run
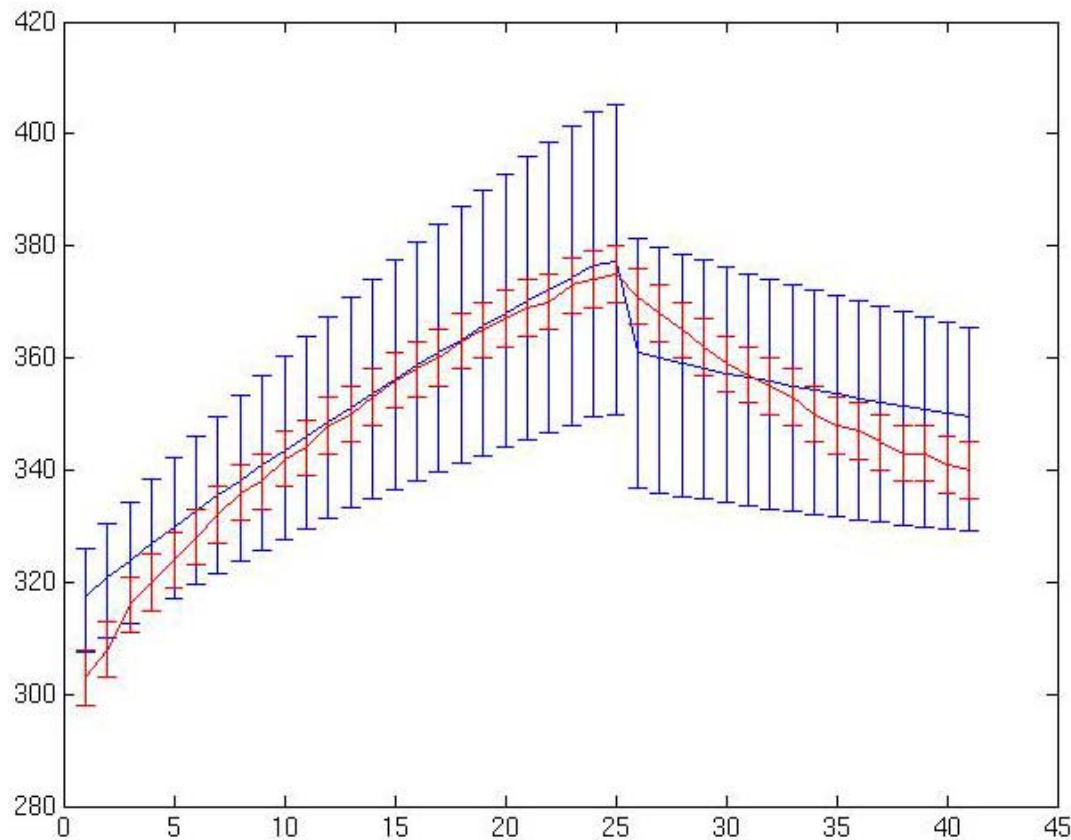
- ❖ Develop scalable algorithms for computing uncertainty in simulation results
- ❖ Develop optimization methods that take uncertainty into account
- ❖ Implement both into software toolkits

The ASCI V&V Program is the main driver for this project

❖ **The V&V vision was stated as follows:**

"Establish <u>confidence</u> in the simulations supporting the Stockpile Stewardship Program through systematic demonstration and documentation of the predictive capability of the codes and their underlying models."

❖ **The V&V Level 1 milepost states:**

"Demonstrate initial uncertainty quantification assessments of ASCI nuclear and nonnuclear simulation codes."

DDACE is a software package for designing computer experiments and analyzing the results

OA–based Latin Hypercube Samples (n=361)
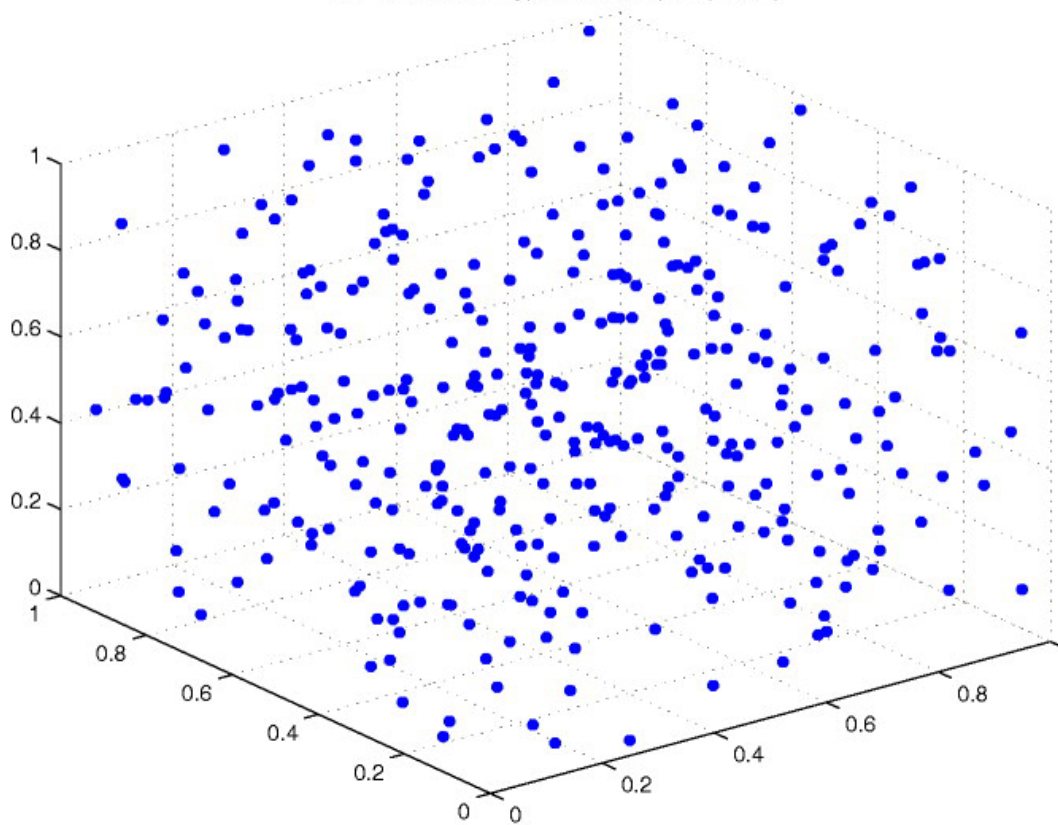
- ❖ Wide variety of distributions and sampling techniques
- ❖ Techniques for determining main effects
- ❖ DDACE integrated with IDEA and Dakota
- ❖ Parallel and serial versions
- ❖ XML interface

Current capabilities of DDace

❖ A collection of popular sampling strategies
  » Random
  » Full Factorial
  » Latin Hypercube
  » Orthogonal arrays (OA)
  » OA-based Latin hypercube
  » User-defined sampling strategy
  » Capability to generate function approximations using Multivariate Additive Regression Splines (MARS)

❖ Parallel and serial versions

❖ XML interface, GUI under development

Some Definitions

- ❖ Variability – inherent variation associated with the physical system under consideration

- ❖ Uncertainty – a *potential* deficiency in any phase or activity of the modeling process that is due to *lack of knowledge*

- ❖ Sensitivity Analysis – estimates changes to output with respect to changes of inputs

- ❖ Uncertainty Analysis – quantifies the degree of confidence in existing data and models

Must determine computation error ε_A using only computed function values

❖ Use difference table

$r(x)$

Δr_0

$r(x+h)$ $\Delta^2 r_0$

Δr_1 $\Delta^3 r_0$

$r(x+2h)$ $\Delta^2 r_1$ $\Delta^4 r_0$

Δr_2 $\Delta^3 r_1$

$r(x+3h)$ $\Delta^2 r_2$

Δr_3

$r(x+4h)$

❖ Compute error at k^{th} difference

$$\varepsilon_A^k \approx \frac{\max_i \left| \Delta^k r_i \right|}{\beta_k}, \ \beta_k = \sqrt{\frac{(2k)!}{(k!)^2}}$$

❖ Errors begin to converge

$$\varepsilon_A^k \to \varepsilon_A \ \text{ as } \ k \to \infty$$

❖ Also

$$r'''(\mu) \approx \frac{\Delta^3 r}{h^3}$$

Optimization algorithms can take advantage of sensitivity information

❖ Computing sensitivities requires a little bit of error analysis…

» Use centered differences on residuals

$$r'(x) = \frac{r(x+h) - r(x-h)}{2h}$$
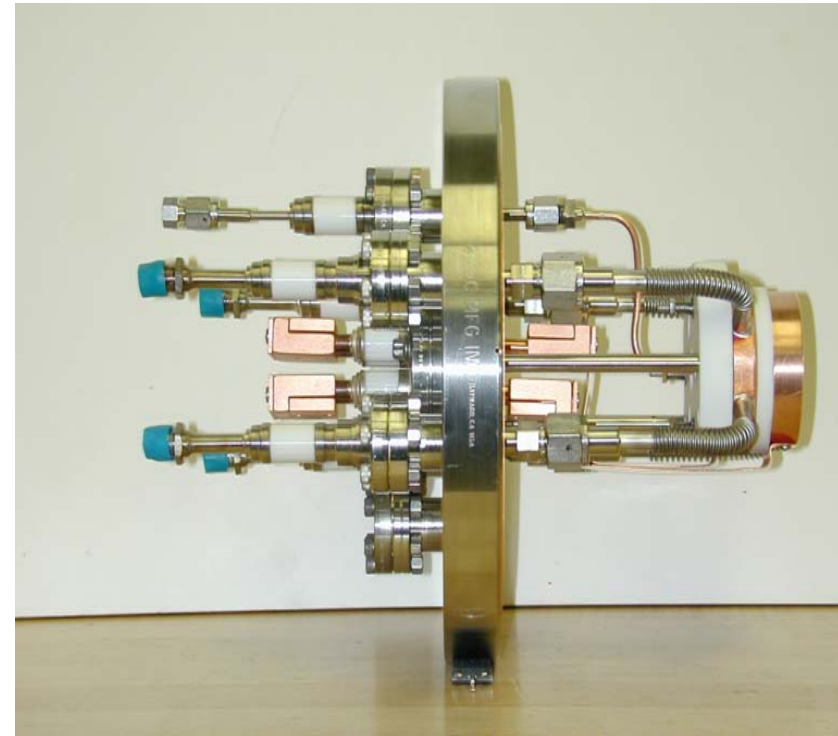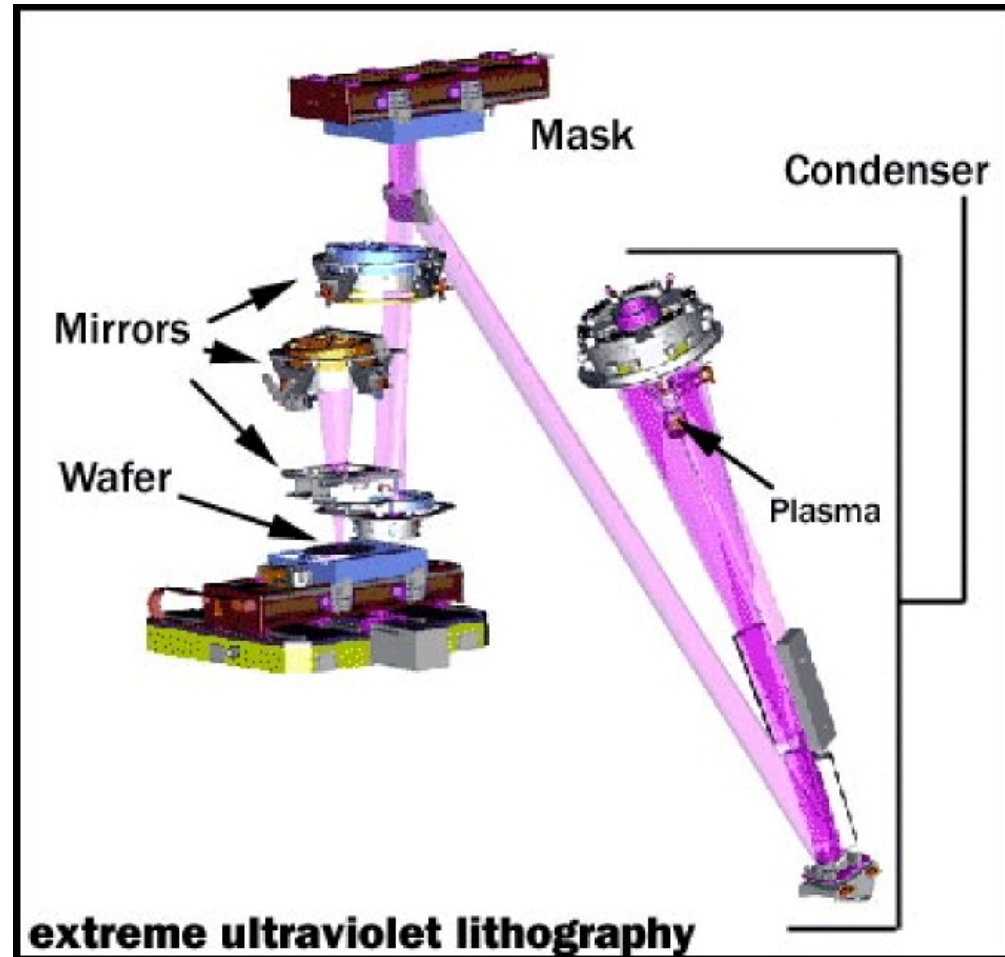
» Truncation and computation yields error

$$\varepsilon = \frac{1}{6} h^2 \left| r'''(\mu) \right| + \frac{\varepsilon_A}{h}, \quad \mu \in [x-h, x+h]$$

» Find the step size h that minimizes error

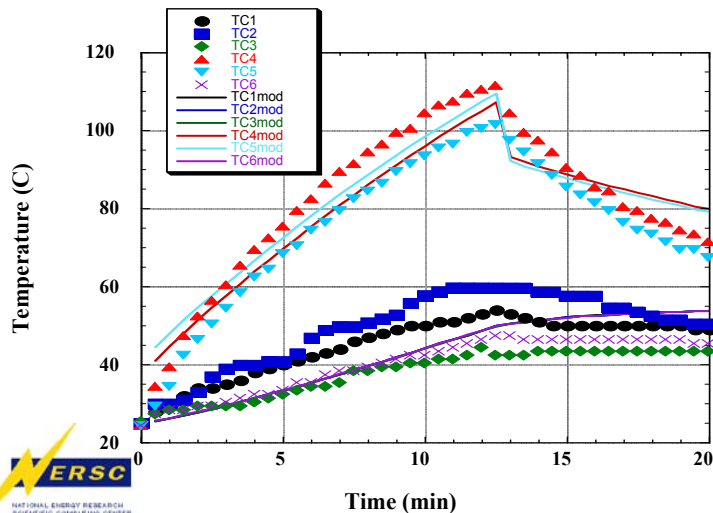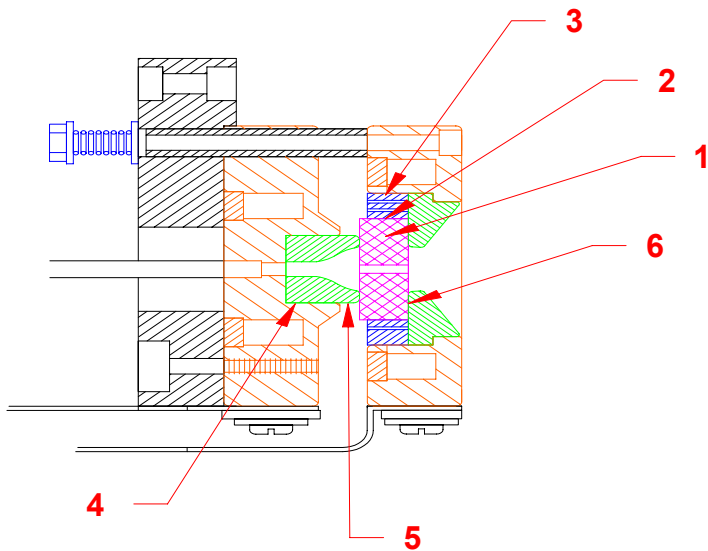$$h_{best} = \sqrt[3]{\frac{3\varepsilon_A}{\left| r'''(\mu) \right|}}$$

$$\nabla f(x) = \frac{f(x + h_{best}) - f(x)}{h_{best}}$$

Extreme Ultraviolet Lithography

extreme ultraviolet lithography

Mask

Condenser

Mirrors

Wafer

Plasma

Current EUVL Lamp

The model problem was taken from an EUVL design problem

- ❖ Find model parameters, satisfying some bounds, for which the simulation matches the observed temperature profiles
- ❖ Objective function consisted of computing the maximum temperature difference over all 6 curves.
- ❖ Each simulation required approximately 7 hours on 1 processor of Cplant.