New Directions in Model Based Data Assimilation

Gregory J. McRae

MIT
Chemical Engineering
Course 10

Outline of Presentation

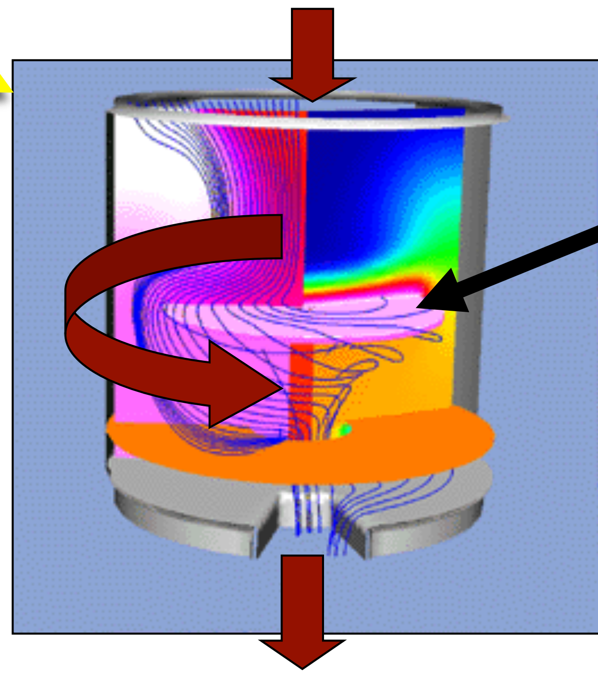
- Introduction to data assimilation
- Decision making in the presence of uncertainty
- Challenges and new opportunities
  - *Algorithms for uncertainty propagation*
  - *Data architectures and management*
  - *Standards for data exchange*
  - *Need for an interdisciplinary approach*
- Future directions and conclusions

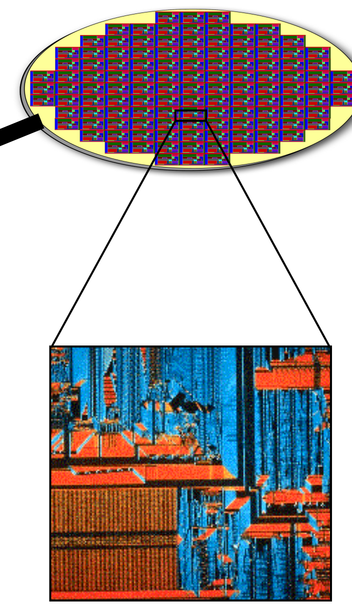
There is a critical need for a new approach to merging models / data

What is the Chemical Industry?

Traditional Processes

Semiconductor Manufacturing





"…Chemistry is the important and common ingredient…"

MIT Chemical Engineering

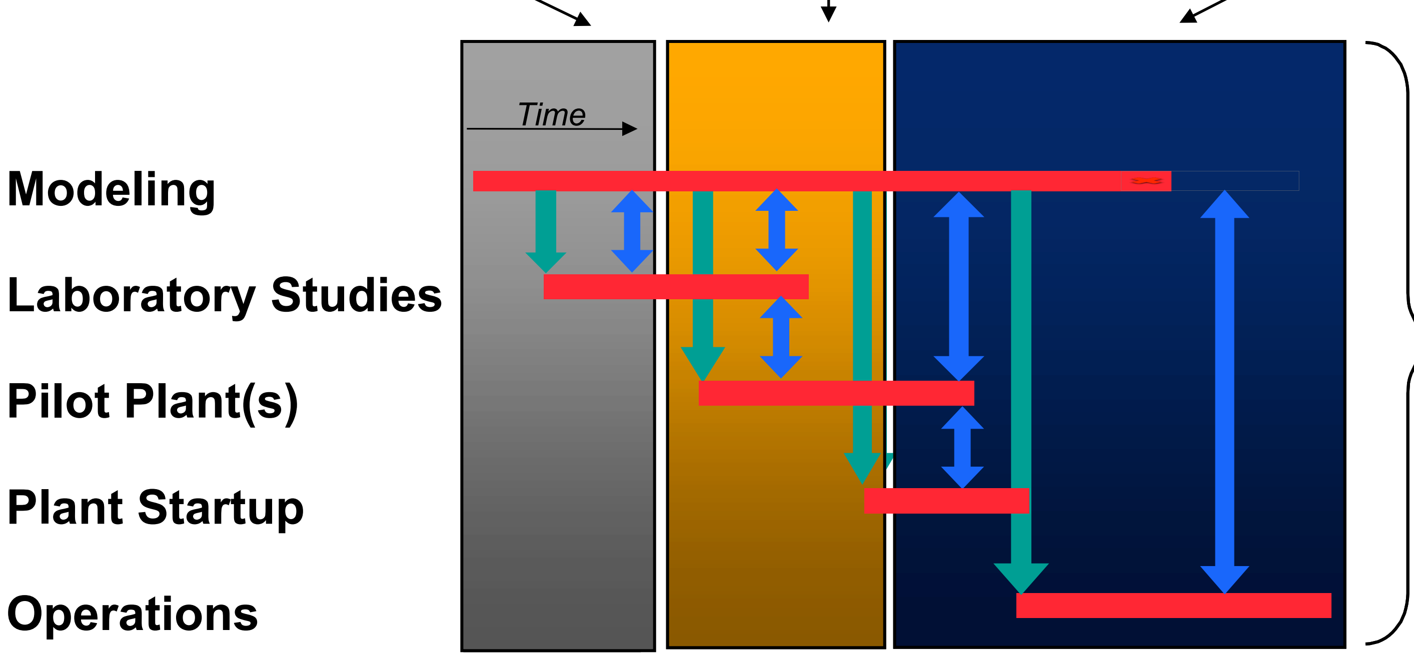
Models are Crucial in Chemical Engineering

1. Technical Feasibility → **2. Operational Feasibility** → **3. Performance Improvements**

1. Technical Feasibility
- *Synthesis routes*
- *By-products*
- *Separation issues*

2. Operational Feasibility
- *Safety*
- *Robustness*
- *Consistency*

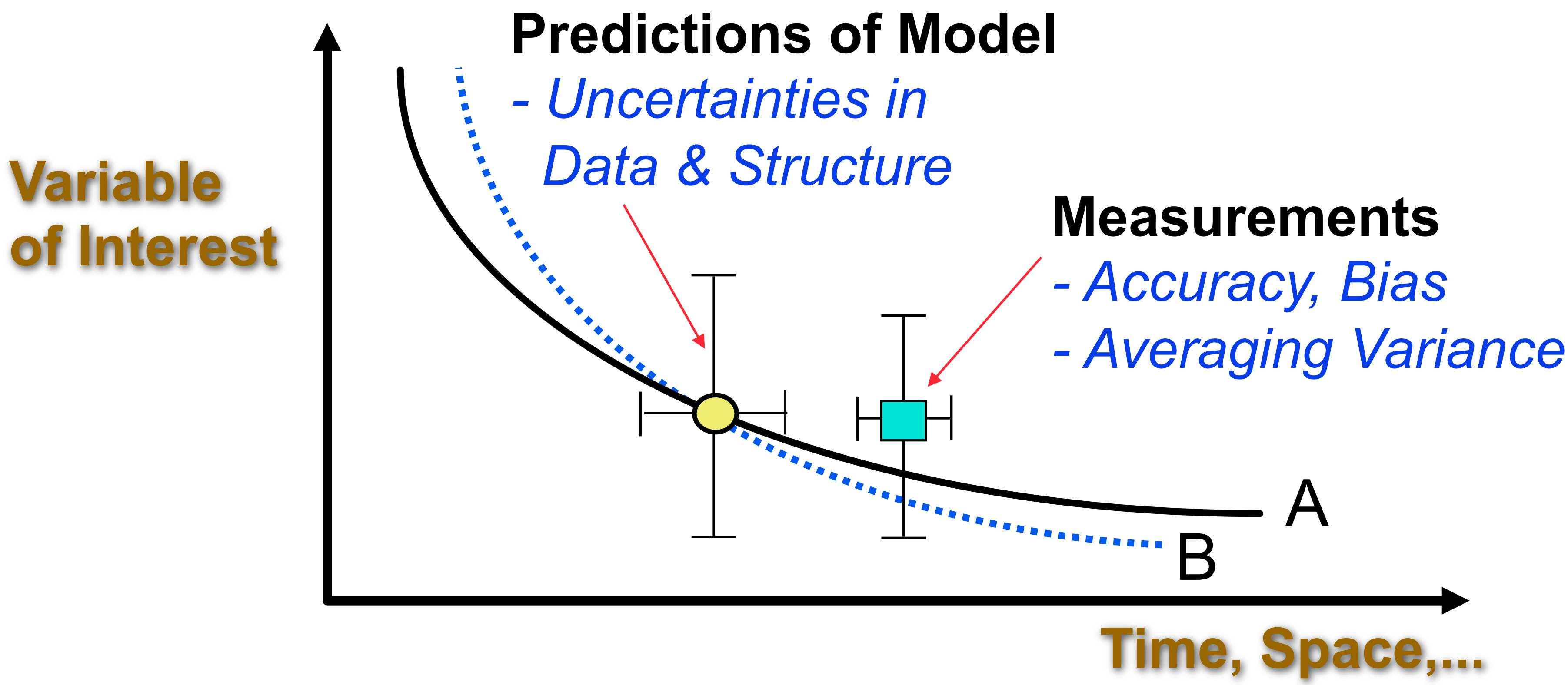
3. Performance Improvements
- *Yield*
- *Throughput*
- *Quality*



Models serve as the basic framework for knowledge management

MIT Chemical Engineering

Data Need: Model Verification / Discrimination

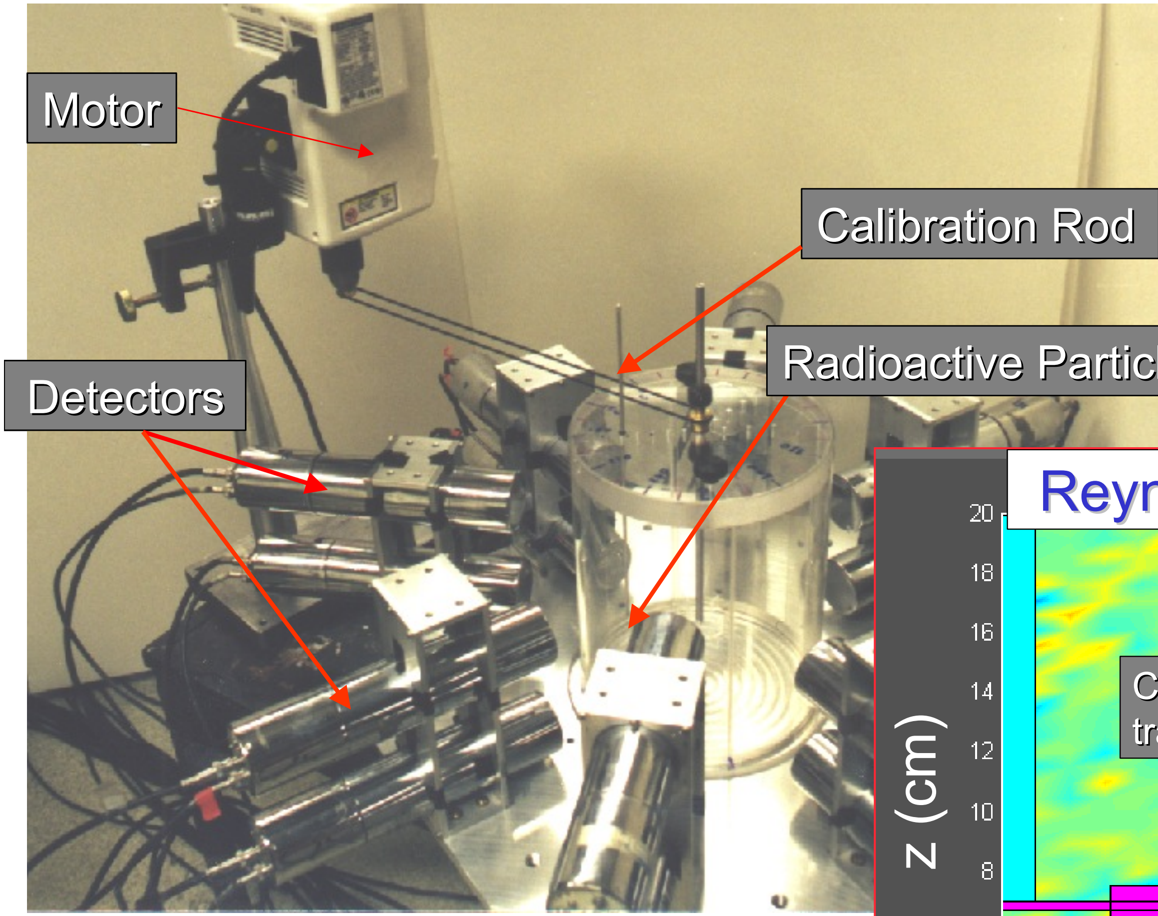


Meaningful comparisons requires estimates of uncertainties in prediction and observations

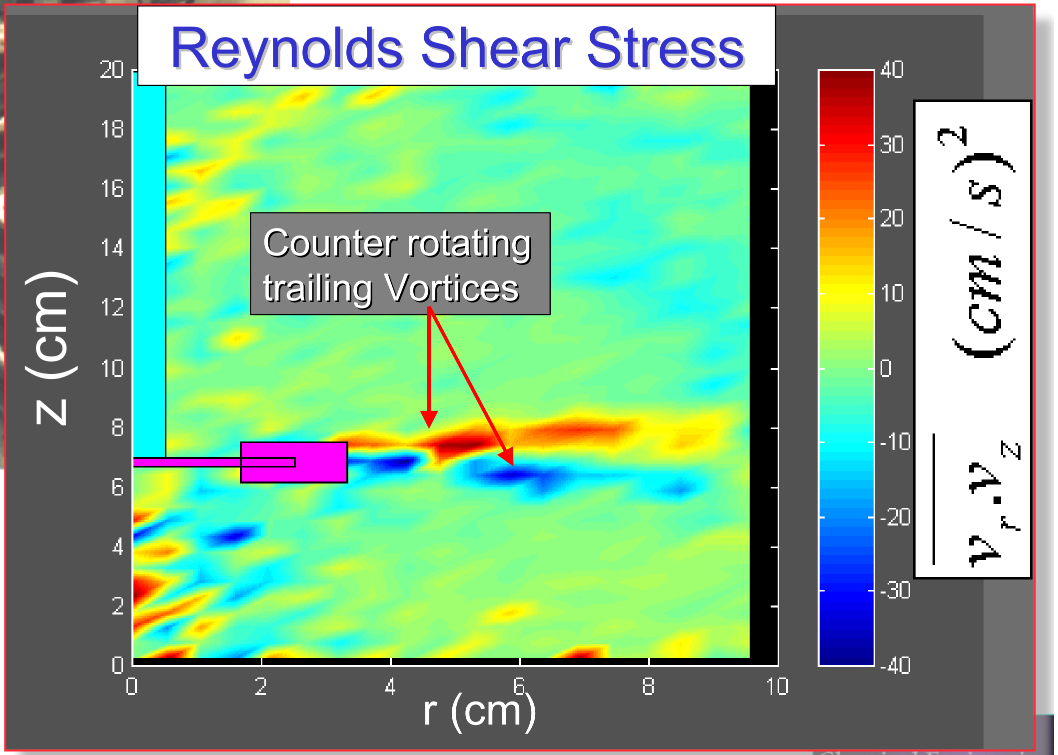
Other Dimensions of the Need for Data

- Model discrimination – *Model A vs. B*
- Hypothesis testing – Which parameter is "best"
- Model verification – What are the "stopping" rules
- Experimental design – Where to measure
- Optimization objectives – Fail-Safe vs. Safe-Fail
- Resource allocation – *Where to spend the money*
- *etc.*

High Bandwidth Data Assimilation -- *Mixing*



X-Ray Tomography



CREL Washington University

Measurements are Vital – *Historical View*

"…When you can <u>measure</u> what you are speaking about, and express it in numbers, you know something about it; but when you cannot … your knowledge is of a meager and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely in your thoughts, advanced the state of science…"

William Thompson, Lord Kelvin

MIT
Chemical Engineering

SUCCESS WILL DEPEND ON:

1. **NEW ALGORITHMS THAT CAN TACKLE NONLINEAR, COMPLEX AND LARGE-SCALE PROBLEMS**

Opportunities for New Algorithms/Theory

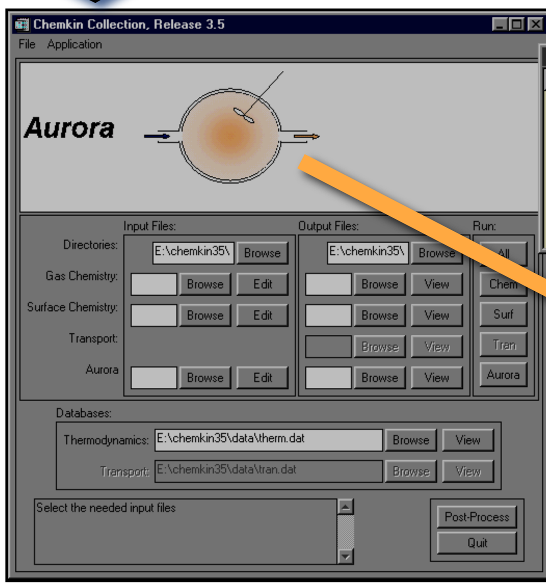
- Statistics (<u>Bayesian</u> based approaches)
  - *Model discrimination*
  - *Experimental design*
  - *Parameter and state estimation for ODE/PDE's*
- Numerical Algorithms
  - *Inverse problems for DAE/PDE's*
  - *Data assimilation of n-dimensional data*
  - *Solution of large scale optimization problems*
  - *Uncertainty propagation*
  - *Multiscale integration*
- Data analysis and visualization
  - *Immersive data mining and analysis*

MIT Chemical Engineering

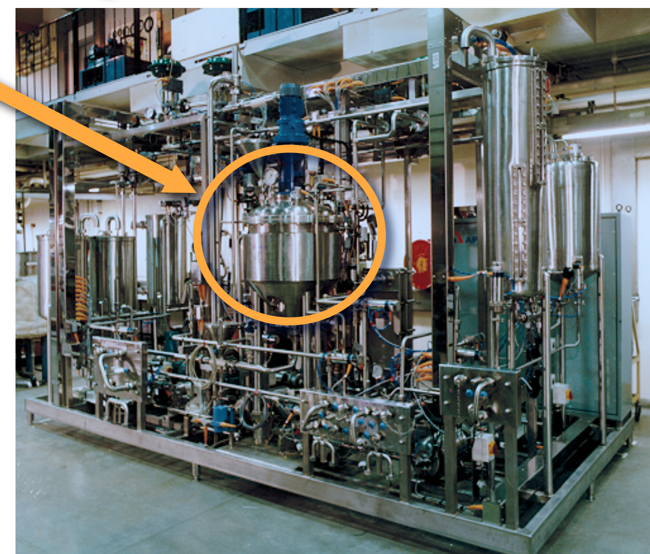
Overall Goal – *From Bench to Plant Scales*

Laboratory Experiment

CHEMKIN model



Chemical Plant



<u>Model based</u>
design/analysis
of experiments

Data from plant

MIT
Chemical Engineering

A Simple Example – Linear Balance Equations

Chemical Plant

Abstraction

(Mass & Energy Balances)

$$0 = Ax + w$$

↑ Error in modeling

Measurements

$$y = Cx + e$$

↑ Error in measurement

Given **A**,**C**,**w**,**e** and **y** find best estimate of state **x**

MIT Chemical Engineering

Solution to State Estimation Problem

1 Set up the constrained optimization problem (No model error)

$$\underbrace{\min}_{\mathbf{x}} \quad (\mathbf{y}-\mathbf{C}\mathbf{x})^T \mathbf{W}^{-1} (\mathbf{y}-\mathbf{C}\mathbf{x})$$

$$s.t. \quad \mathbf{A}\mathbf{x} = \mathbf{0}$$

2 Solve for the estimates of the state

$$\hat{\mathbf{x}} = \hat{\mathbf{x}}_0 - (\mathbf{C}^T \mathbf{W}^{-1} \mathbf{C})^{-1} \mathbf{A}^T [\mathbf{A}(\mathbf{C}^T \mathbf{W}^{-1} \mathbf{C})^{-1} \mathbf{A}^T]^{-1} \mathbf{A}\hat{\mathbf{x}}_0$$

$$\hat{\mathbf{x}}_0 = (\mathbf{C}^T \mathbf{W}^{-1} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{W}^{-1} \mathbf{y} \qquad (\text{OLS} - \text{No constraint})$$

3 Determine the variance in the estimates

$$\mathbf{V} = \mathbf{V}_0 - \underbrace{(\mathbf{C}^T \mathbf{W}^{-1} \mathbf{C})^{-1} \mathbf{A}^T [\mathbf{A}(\mathbf{C}^T \mathbf{W}^{-1} \mathbf{C})^{-1} \mathbf{A}^T]^{-1} \mathbf{A}(\mathbf{C}^T \mathbf{W}^{-1} \mathbf{C})^{-1}}$$

The use of model-based constraints reduces variance in estimates … BUT

Assumptions underlying Solution

1. Model is linear
2. Normally distributed errors in data and solution
3. Data are uncorrelated in time
4. <u>No errors in the model of the process!!</u>
5. Etc.

→ There is a critical need for a more realistic approach that deals with model uncertainty

Simple Problem: Kinetics of SiH_4 -> (Si) + $2H_2$

Model

$$\frac{dy(t)}{dt} = -k\, y(t) \quad ; y(0) = y_0, \qquad y(t) = [SiH_4(t)]$$

Solution

$$y(t) = y_0 e^{-k\,t}$$

Sensitivity to parameter variations

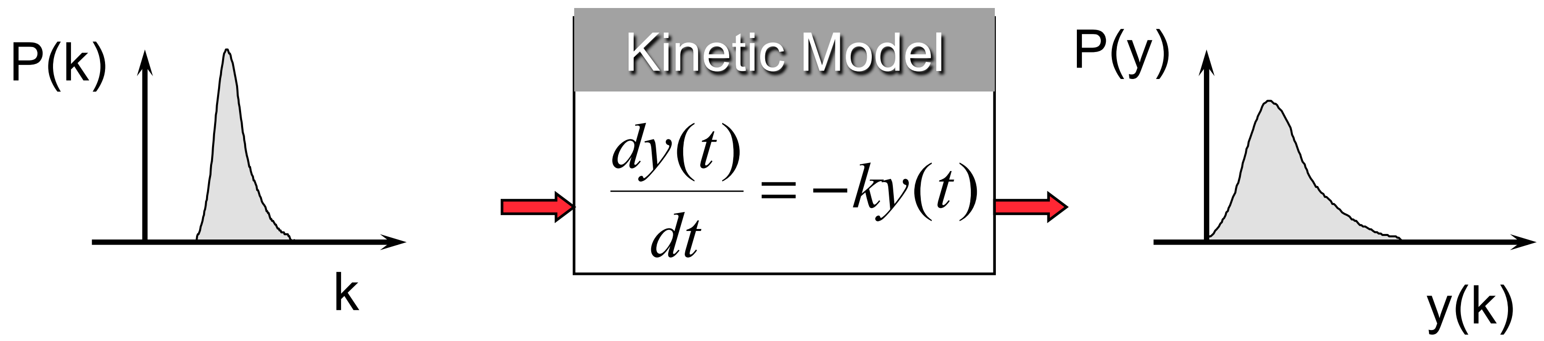
$$S = \left.\frac{\partial y(t)}{\partial k}\right|_{\bar{k}} = -t\, y_0 e^{-\bar{k}\,t}$$

➡ But, what if k is uncertain?

Solution in Presence of Uncertainty

Uncertain Rate Constant

Uncertain Output



Normal distribution for rate constant (k)

But

Lognormal distribution for solution at each time

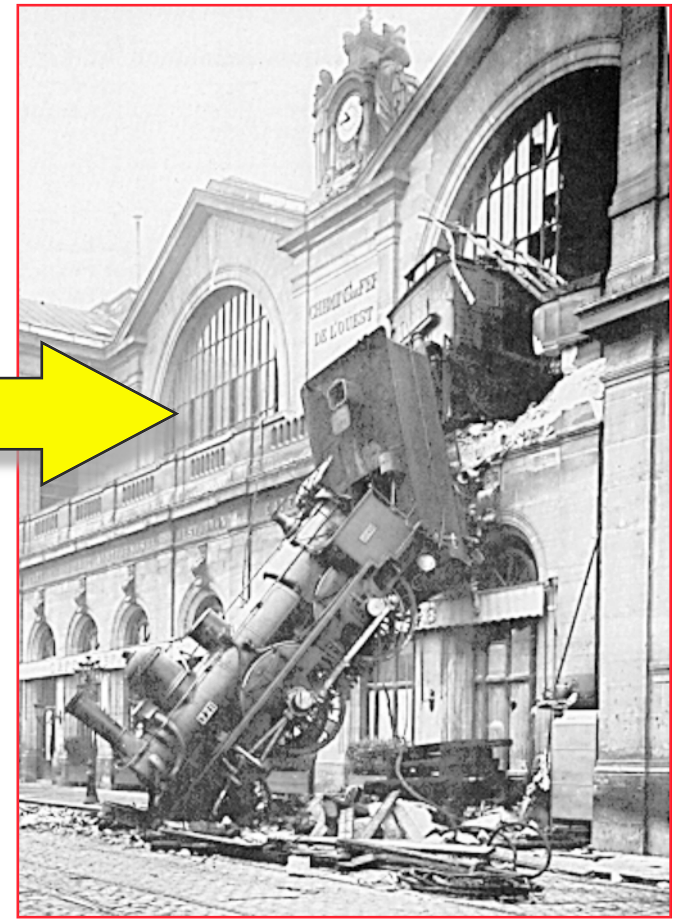
$$P(k) = \frac{1}{k_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{k-k_0}{k_1}\right)^2}$$

$$P(y) = \frac{1}{k_1 t y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(y/y_0)+tk_0}{tk_1}\right)^2}$$

MIT Chemical Engineering

Key Message – Outcomes are Important

"... While there are always lots of uncertainties, the key challenge in engineering is to find those problem components that contribute most to uncertainties in ***outcomes****..."*



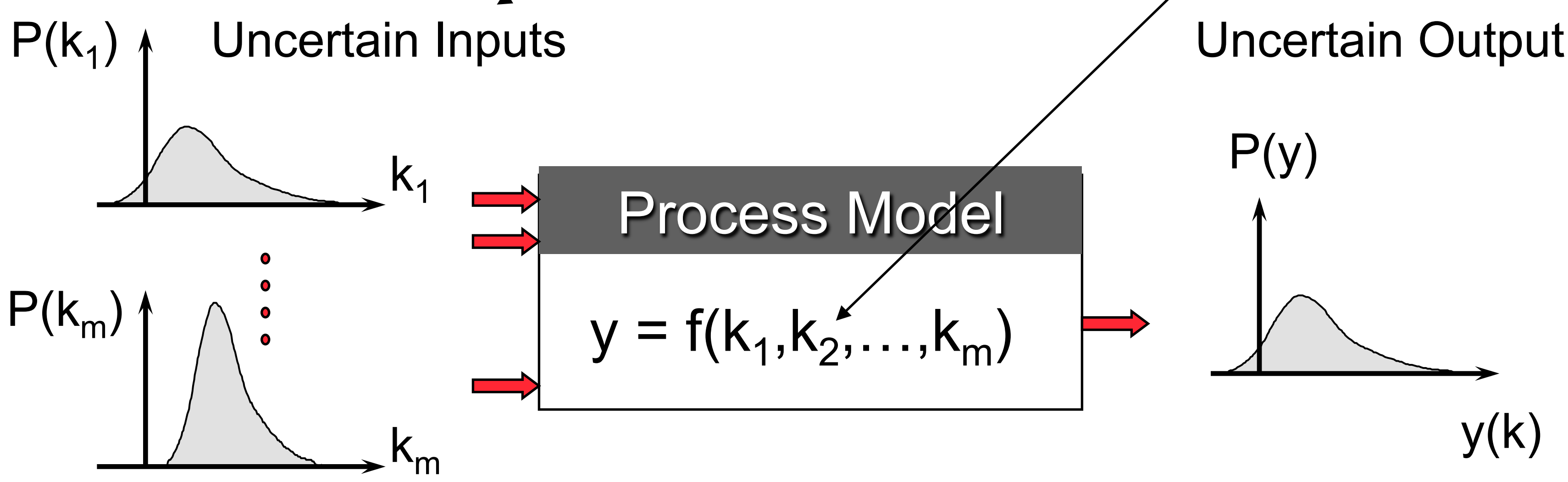
MIT Chemical Engineering

Example: Where to Allocate Resources

Physical, chemical and decision process models

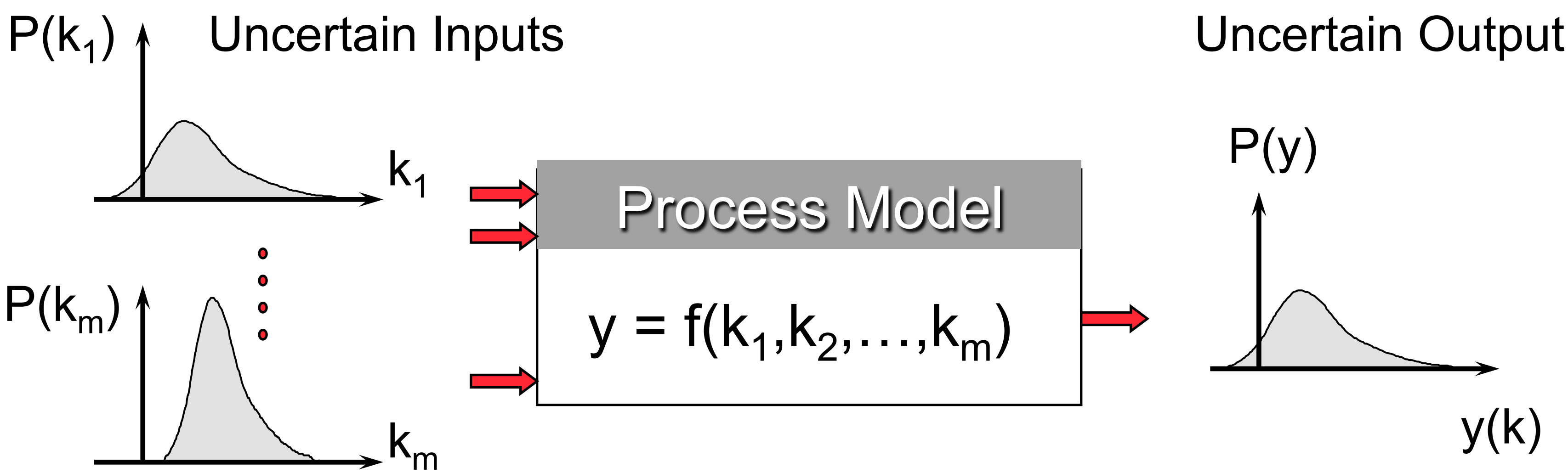
Uncertain Inputs
- *Literature*
- *Experiments*
- *Other models*
- *etc.*

Which parameters control outcomes?



MIT Chemical Engineering

How do Uncertain Inputs effect the Outputs?



Measures of Uncertainty (Expected value, variance, pdf, etc.)

e.g. $E[y(k)] = \int_y y(k)\,P[y(k)]\,dy(k)$

$$\equiv \underset{k}{\int\cdots\int} y(k)\,P[k]\,dk_1\cdots dk_m$$

Multi-dimensional integrals are computationally very expensive

MIT Chemical Engineering

Attributes of an Uncertainty Analysis System

- Compatible with existing modeling systems
- At least <u>four orders of magnitude faster</u> than Monte Carlo
- An ability to get the probability density function of outputs
- Be able to identify the key contributions to uncertainties in outcomes

How? → *"...By rethinking conventional methods and directly embedding uncertainty into the modeling process itself..."*

Incorporating Uncertainty at the Beginning

Fourier Series Representation of f(x)

$$f(x) = a_0 + \sum_{i=1}^{\infty} a_i \sin(\omega_i x) + b_i \cos(\omega_i x)$$

What happens if x is a random variable?

Polynomial Chaos Representation of f(w) (Wiener, 1947)

$$f(\omega) = \sum_{i=1}^{\infty} a_i \, H_i[\xi_1(\omega), \cdots, \zeta_m(\omega)]$$

Coefficients of expansion

Functional (e.g. Hermite Polynomial)

Known probability distributions (e.g. unit Normal N[0,1])

Curse of Dimensionality

Measures of Uncertainty *(expected value, variance, etc.)*

Definition

$$E\{y(\underline{\theta})\} = \int_{-\infty}^{\infty} y(\underline{\theta}) f_{y(\underline{\theta})}(y(\underline{\theta}))dy(\underline{\theta})$$

$f_{y(\underline{\theta})}(y(\underline{\theta}))$ unknown

Statistically equivalent integral definition

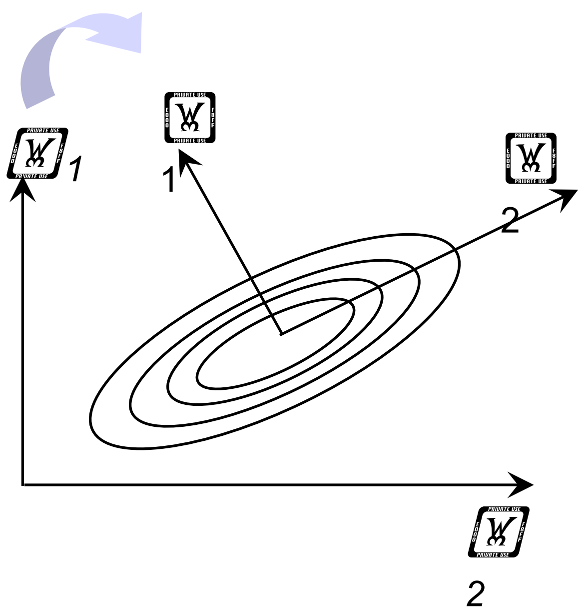
$$E\{y(\underline{\theta})\} = \underbrace{\int \cdot\cdot \int}_{n} y(\underline{\theta}) f_{\underline{\theta}}(\underline{\theta}) d\theta_1 \cdot\cdot d\theta_n$$

Multi-dimensional integrals are computationally very expensive

Projection into 1-D by orthogonal polynomials

$$\theta_i = \sum_j a_{ij} H_j(\underline{\xi}(\omega))$$

Polynomial chaos expansion



New definition (one-dimensional integrals)

$$E\{y(\underline{\theta})\} = \sum_j c_j \int y_j(\xi_1) f_{\xi_1}(\xi_1) d\xi_1 \cdots \int y_j(\xi_n) f_{\xi_n}(\xi_n) d\xi_n$$

MIT Chemical Engineering

Example of Uncertainty Analysis

Model

$$\frac{dy(t)}{dt} = -ky$$

Expansion of y(t)

$$y = \sum y_i H[\zeta_i(\omega)] \approx y_0(t) + y_1(t)\xi + y_2(\xi^2 - 1) + \cdots$$

Residual

$$R(t,\zeta) = \frac{dy(t,\zeta)}{dt} - k(\xi)y(t,\xi)$$

Residual Minimization (Galerkin or Collocation)

$$\int_{-\infty}^{+\infty} R(t,\xi) w_j(\xi) d\xi = 0 \ ; j = 0,1,\cdots,n$$

Uncertain coefficients (Galerkin $w_j = x_j$)

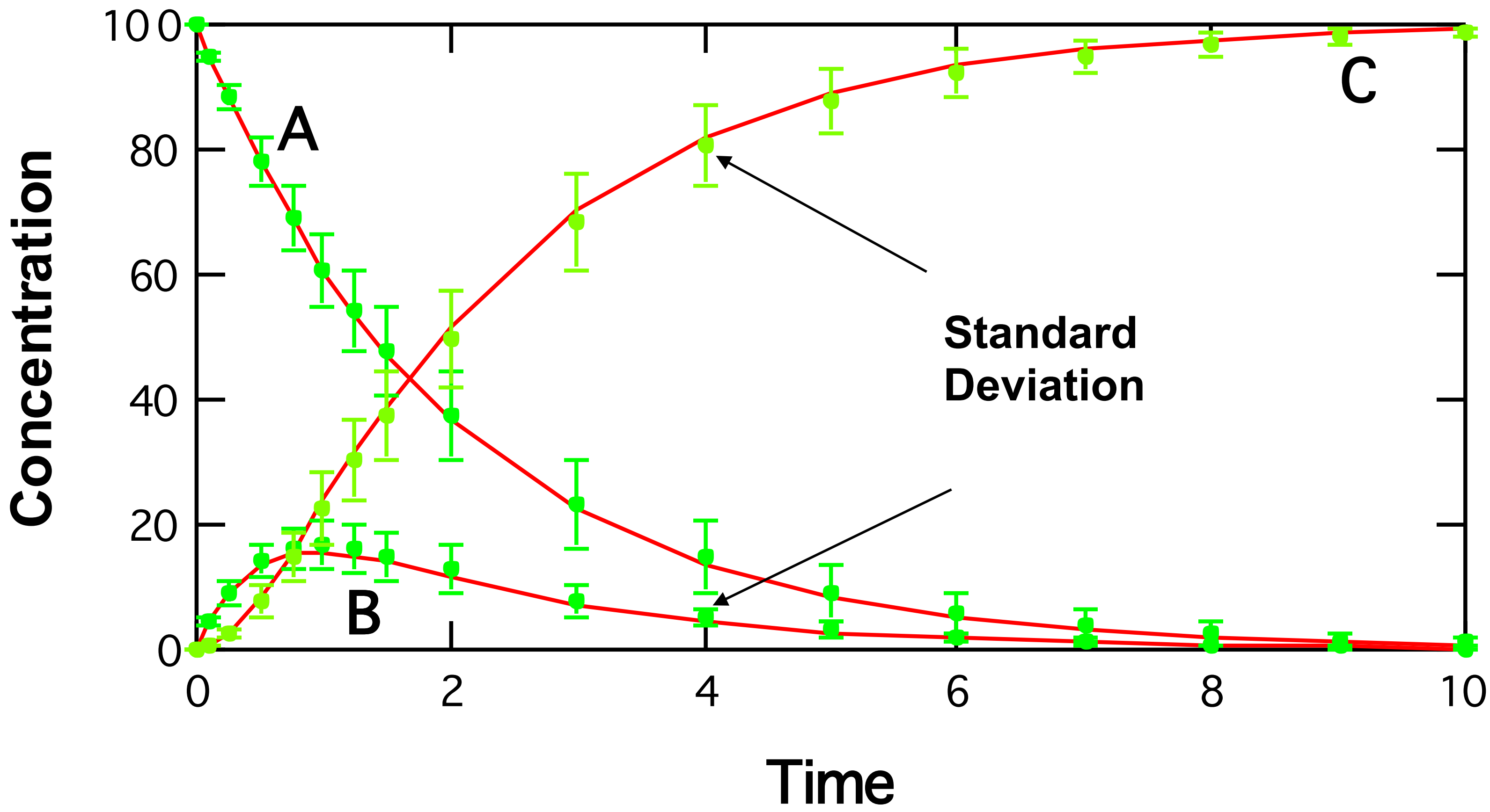
$$A\frac{dy(t)}{dt} + By(t) = 0$$

MIT Chemical Engineering

Simple Reaction Sequence $A \rightarrow B \rightarrow C$

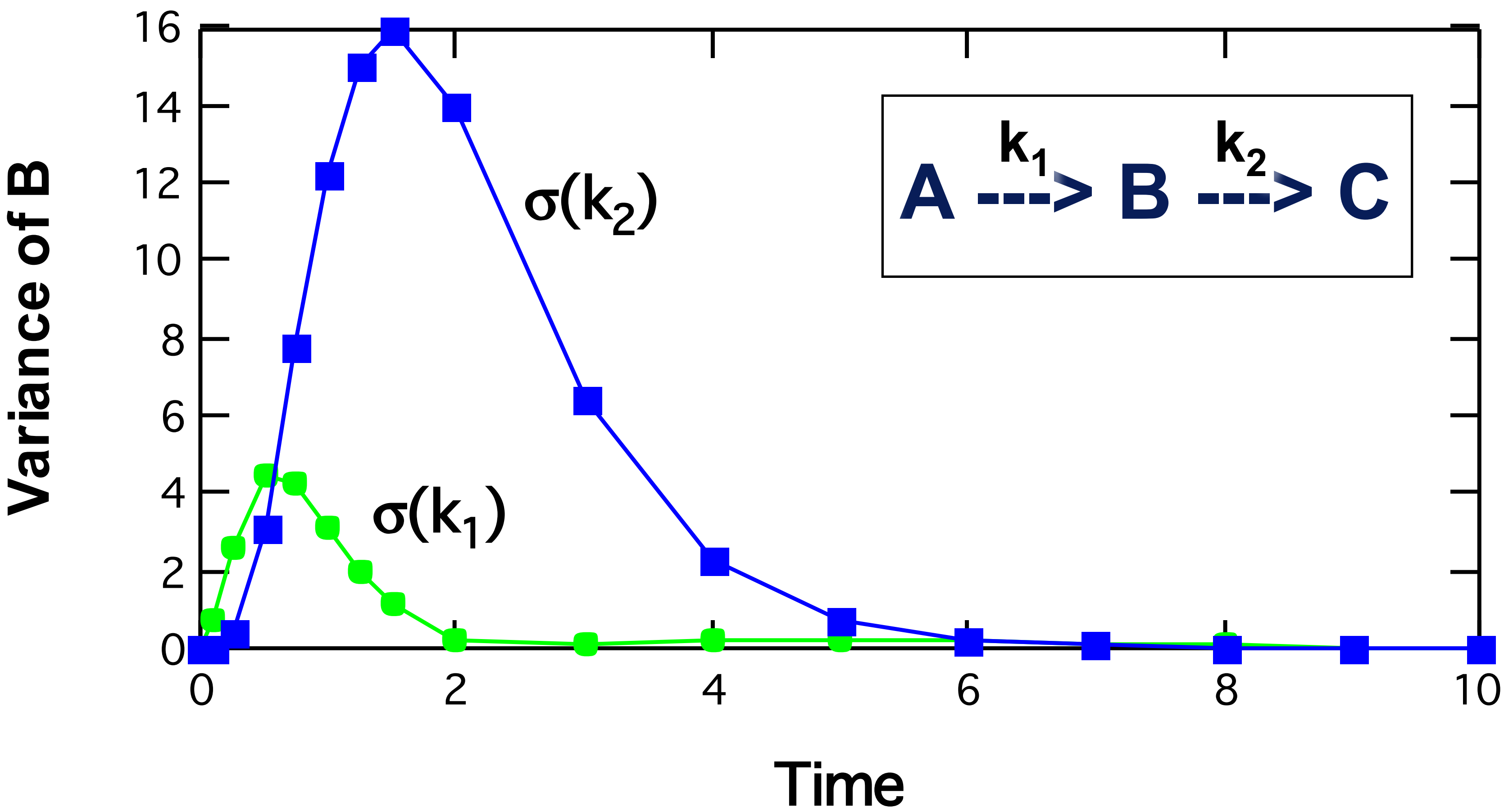
A --> B $\qquad k_1 = 0.5 \pm 0.1$

B --> C $\qquad k_2 = 2.0 \pm 0.5$

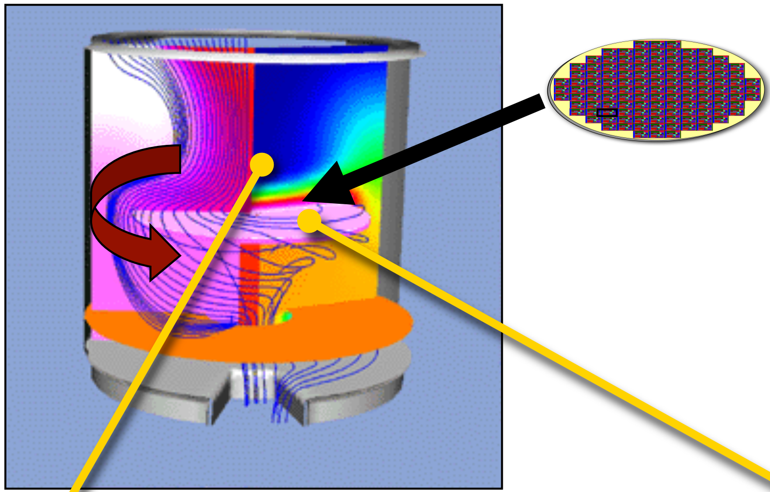


MIT Chemical Engineering

Effect of Parameter Uncertainty on Variance



AMAT Centura Chemical Vapor Deposition Reactor



Operating Conditions

Reactor Pressure	1 atm
Inlet Gas Temperature	698 K
Surface Temperature	1173 K
Inlet Gas-Phase Velocity	46.6 cm/sec

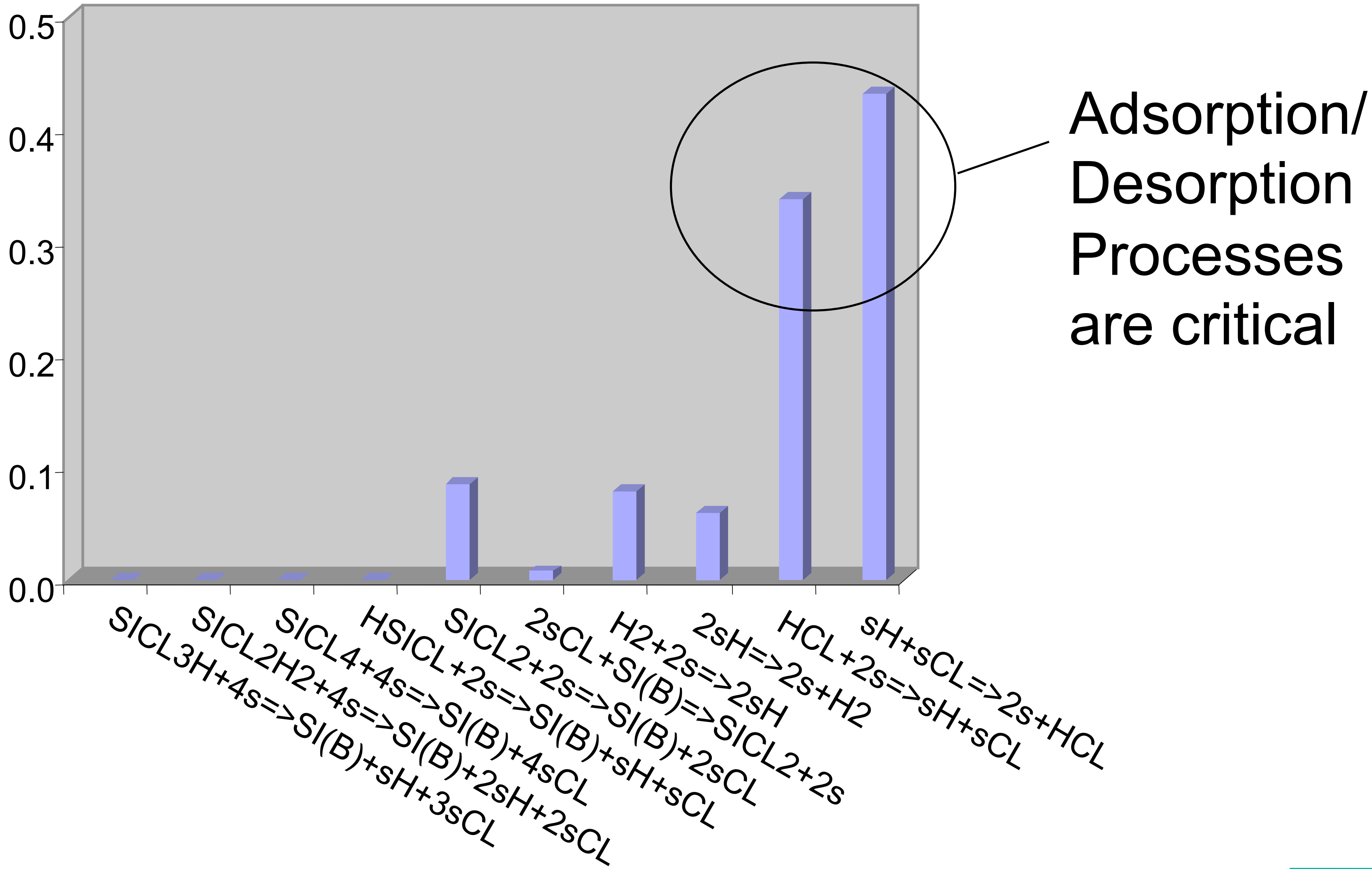
Surface Reactions

$SiCl_3H + 4s \rightarrow Si(B) + sH + 3sCl$
$SiCl_2H_2 + 4s \rightarrow Si(B) + 2sH + 2sCl$
$SiCl_4 + 4s \rightarrow Si(B) + 4sCl$
$HSiCl + 2s \rightarrow Si(B) + sH + sCl$
$SiCl_2 + 2s \rightarrow Si(B) + 2sCl$
$2sCl + Si(B) \rightarrow SiCl_2 + 2s$
$H_2 + 2s \rightarrow 2sH$
$2sH \rightarrow 2s + H_2$
$HCl + 2s \rightarrow sH + sCl$
$sH + sCl \rightarrow 2s + HCl$

Gas Phase Reactions

$SiCl_3H \rightarrow HCl + SiCl_2$
$SiCl_2H_2 \rightarrow SiCl_2 + H_2$
$SiCl_2H_2 \rightarrow HSiCl + HCl$
$H_2ClSiSiCl_3 \rightarrow SiCl_4 + SiH_2$
$H_2ClSiSiCl_3 \rightarrow SiCl_3H + HSiCl$
$H_2ClSiSiCl_3 \rightarrow SiCl_2H_2 + SiCl_2$
$Si_2Cl_5H \rightarrow SiCl_4 + HSiCl$
$Si_2Cl_5H \rightarrow SiCl_3H + SiCl_2$
$Si_2Cl_6 \rightarrow SiCl_4 + SiCl_2$

TCS Lower Wall Deposition Rate (ANOVA)



MIT Chemical Engineering

Research Opportunities in Uncertainty

- Uncertainty analysis is a fertile and much needed area for inter-disciplinary research
- There are many research opportunities
  - *Database design for representation of uncertainties*
  - *New algorithms for uncertainty propagation*
  - *Decision making metrics in the presence of uncertainties*
  - *Treatment of structural uncertainties*
  - *Valuation of cost of "safety factors"*
- Estimates of uncertainties in model inputs are desperately needed

Uncertainty [illegible] Ignorance

MIT Chemical Engineering

SUCCESS WILL DEPEND ON:

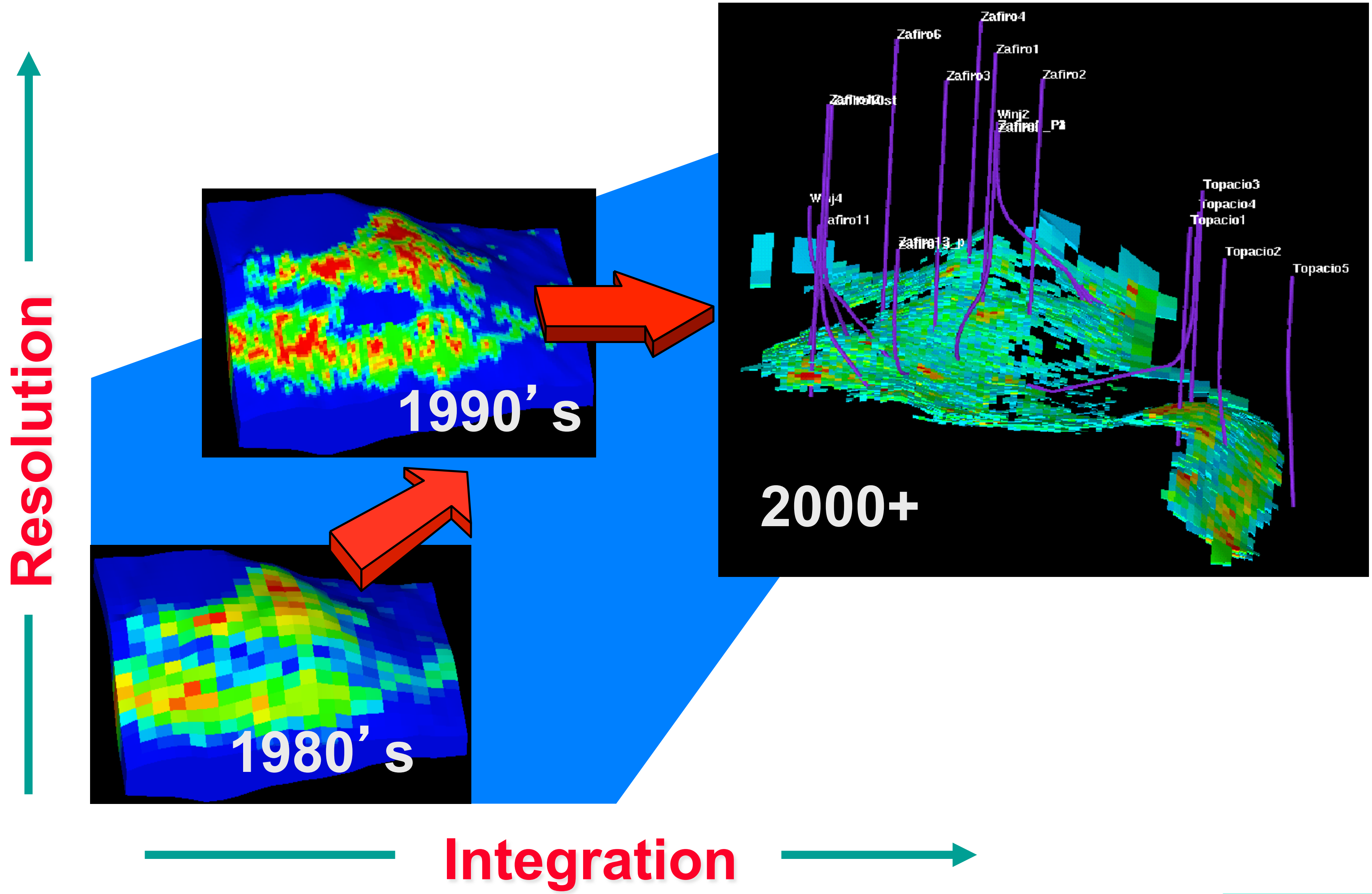
2. INTEGRATED APPROACHES FOR DATA MANAGEMENT, MODELING, SOLUTION, ANALYSIS AND VISUALIZATION

Opportunities in Computational Systems

- Data Modeling and Management
  - *Data exchange standards*
  - *Enterprise integration*
  - *Risk management*
  - *Intellectual capital (e.g. corporate knowledge)*
- Computing Environments
  - *Interoperability of commercial systems*
  - *Collaborative environments*
  - *Uncertainty propagation*
  - *Sensor technology*
- Visualization and Interpretation
  - *Immersive data analysis*

MIT Chemical Engineering

Role of Computing in Data Assimilation



Mobil

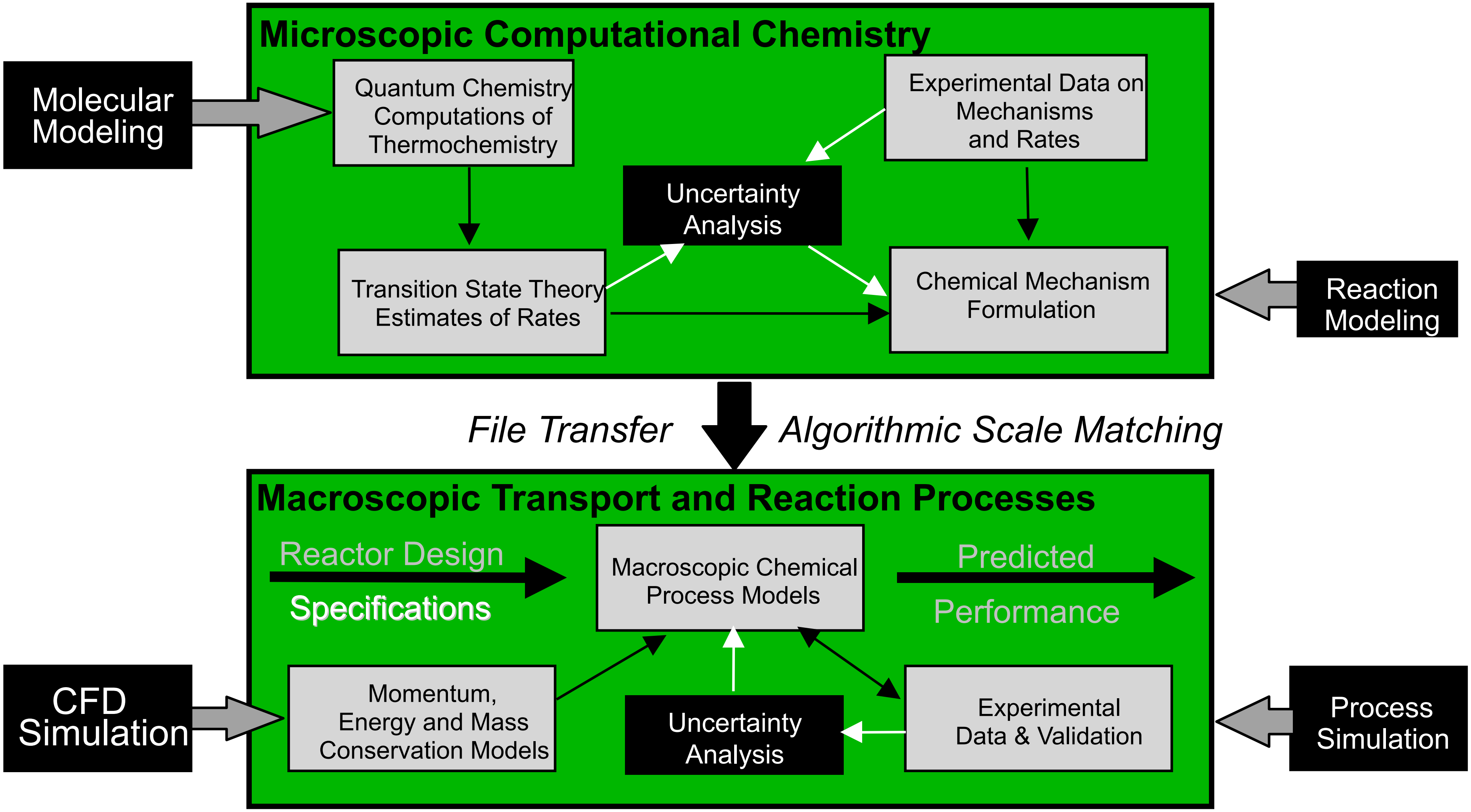
MIT Chemical Engineering

Visions for Software Architecture

A Computational System that is:

- An integrated web-based environment for chemical process modeling, control and optimization
- Able to link molecular, reactor and plant scale models for whole plant simulation
- Integration of corporate information repositories with experimental design

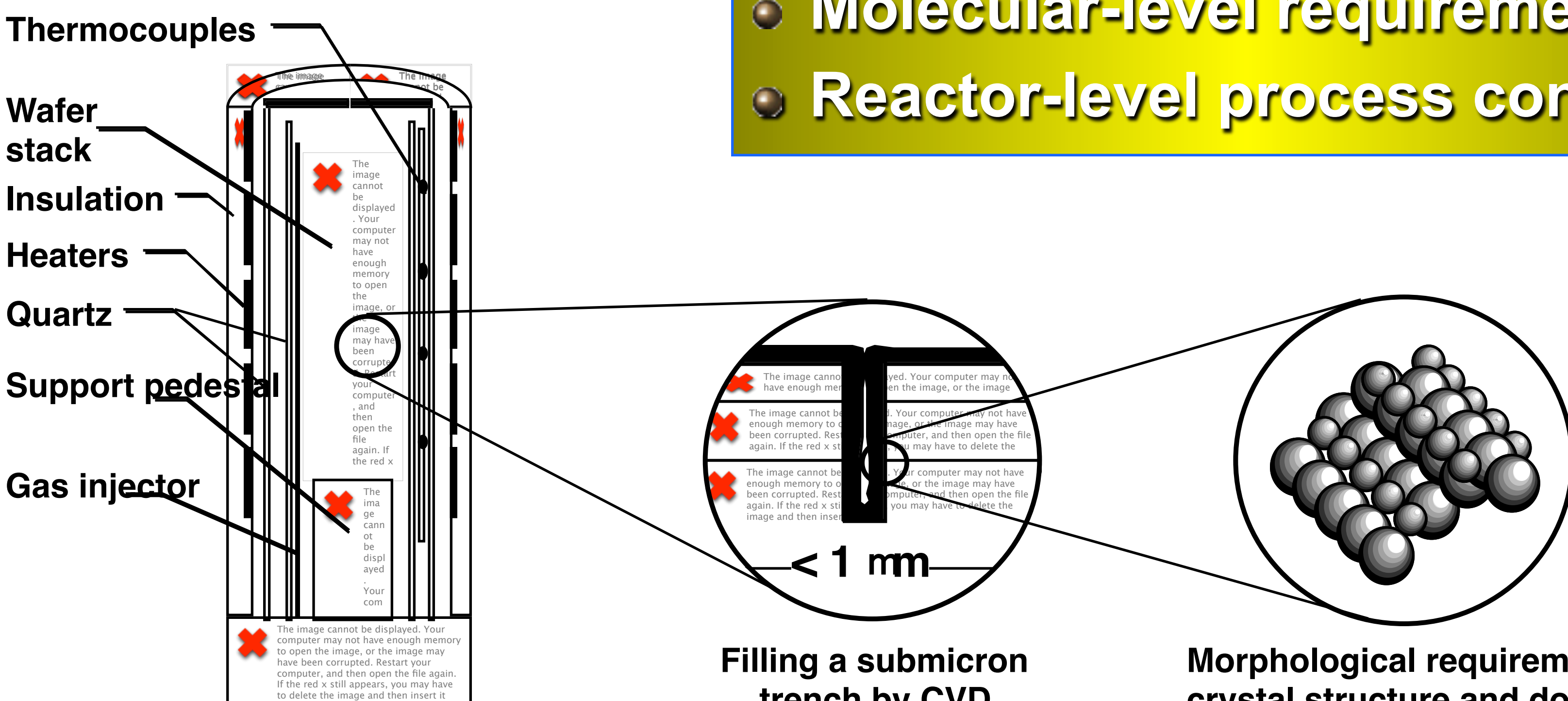
Chemical Engineer's Workbench



MIT Chemical Engineering

Need for Multi-scale Models and their Integration

- Molecular-level requirements
- Reactor-level process controls



Low-Pressure Chemical Vapor Deposition Furnace

Filling a submicron trench by CVD

Morphological requirements, crystal structure and doping

Macroscopic uniformity (at the wafer scale)

Microscopic features (at the device scale)

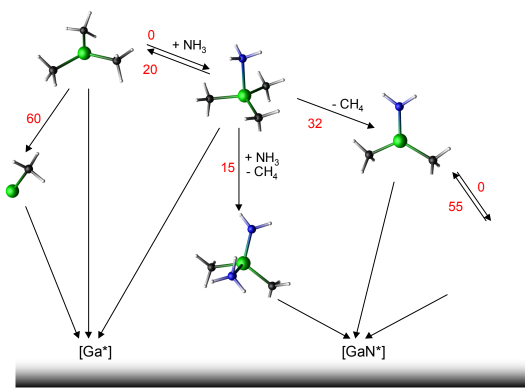
Atomic-scale features (at the surface scale)

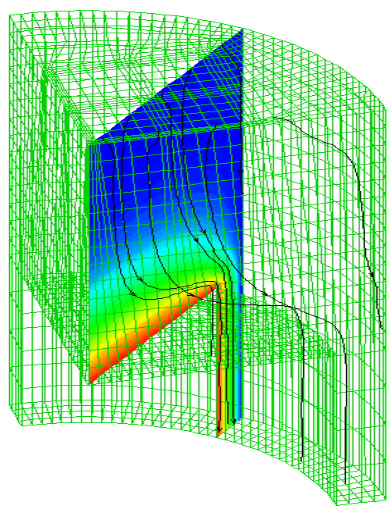
MIT Chemical Engineering

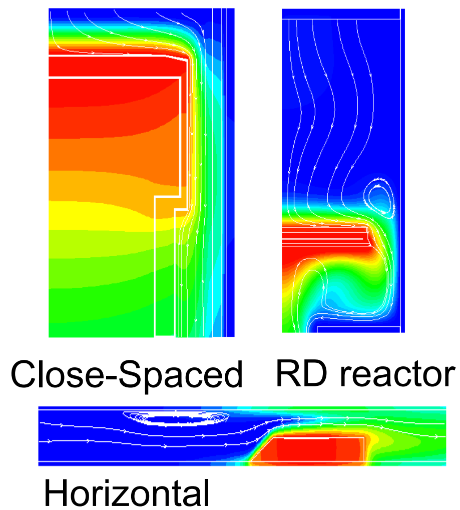
Multi-Scale Integration of Software Systems

XML As a standard for data exchange

Experimental Data & Quantum Chemistry (Gaussian, DFT) ⟷ CFD Model of Reactor Flow (STARCD, CFDRC,..) ⟷ Design Optimization (MINLP, Minos)





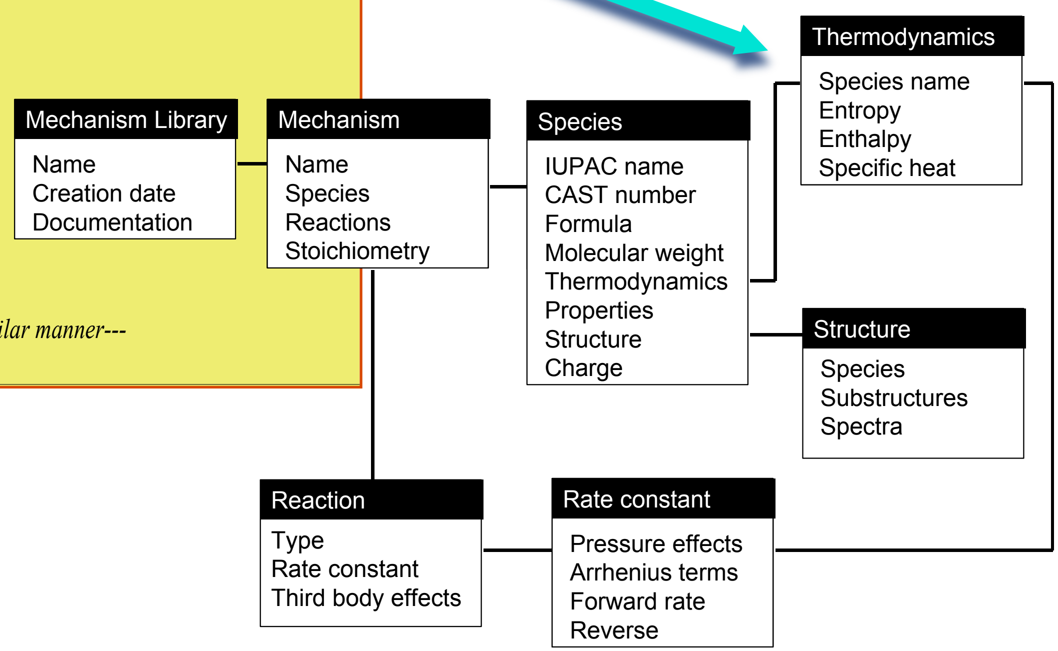


Distributed Computing Resources

Data Structures and XML Representations

eXtended Markup Language (XML)

```
<Thermodynamic_Data>
  <Species>
    <Chemical_Description>
        <Formula> (CH2O)3 </Formula> <Atoms>  C H O  </Atoms> <Phase> G </Phase>
    </Chemical_Description>
      <Date_Entered_Database>
          <Day>  5 </Day> <Month>   7 </Month> <Year> 1990 </Year>
      </Date_Entered_Database>
          <Polynomial_Coefficients>
              <Low_T_Range>
                  <a1>  0.01913678E+03  </a1>
                        …
                  <a7> -0.08432507E+03  </a7>
              </Low_T_Range>
          </Polynomial_Coefficients>
  </Species>
              --- Additional species would be added in a similar manner---
</Thermodynamic_Data>
```

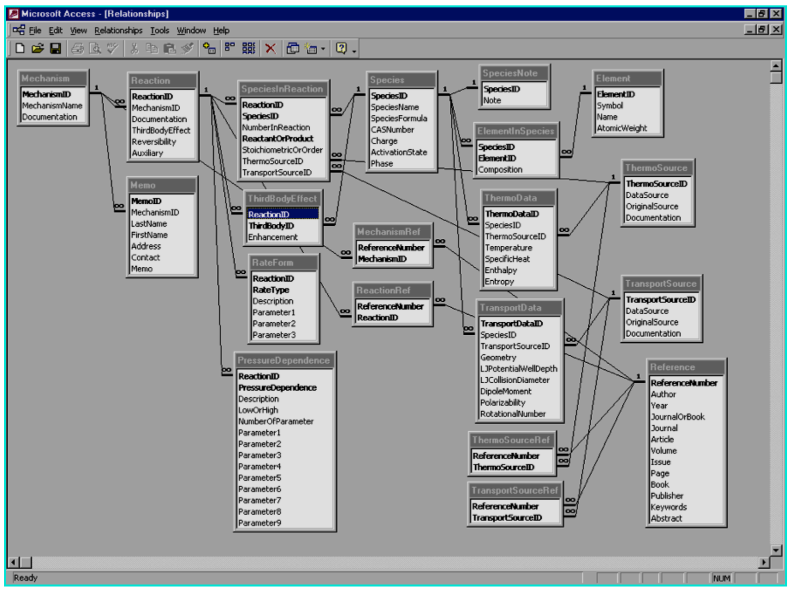


Data Structures

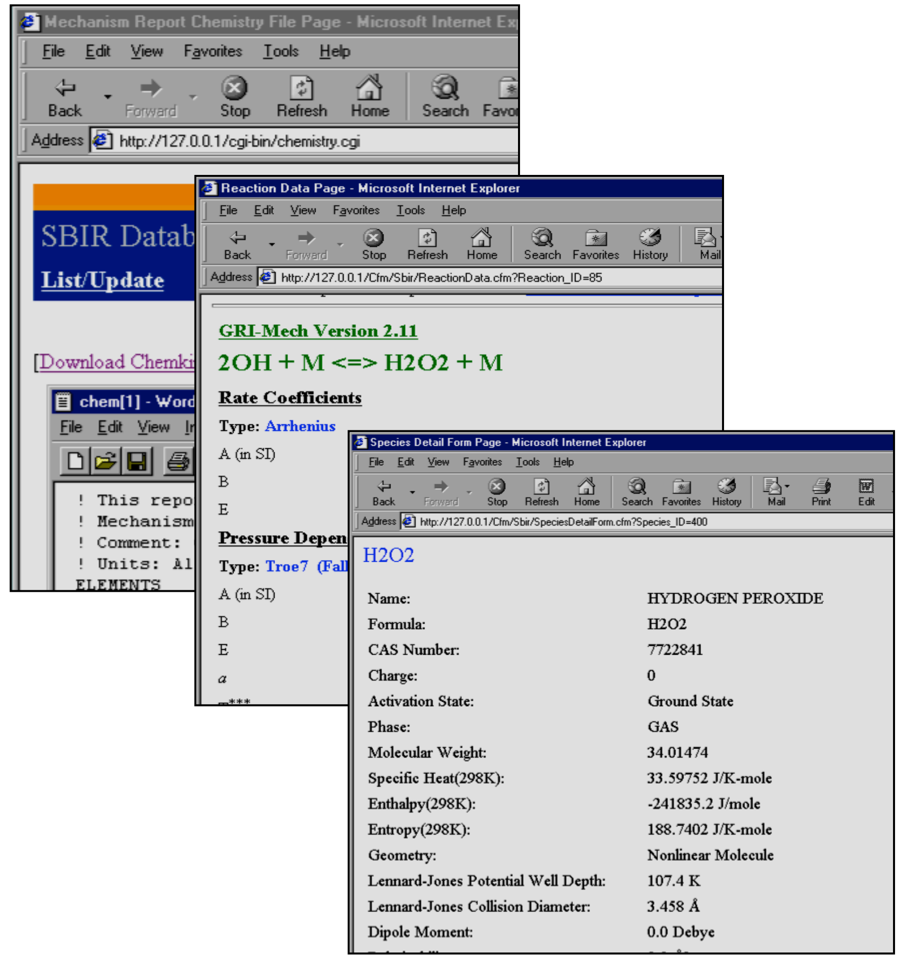
MIT Chemical Engineering

Reaction Mechanism Manager

Data Base Structure

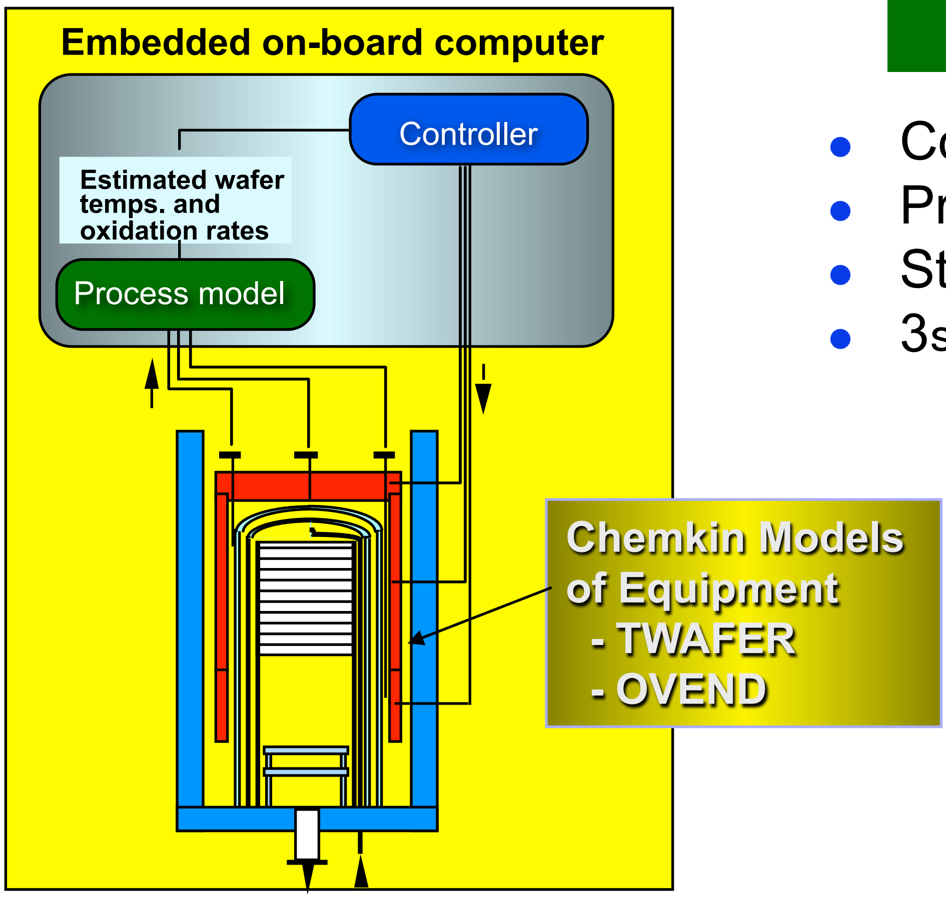


Sample Reports



MIT
Chemical Engineering

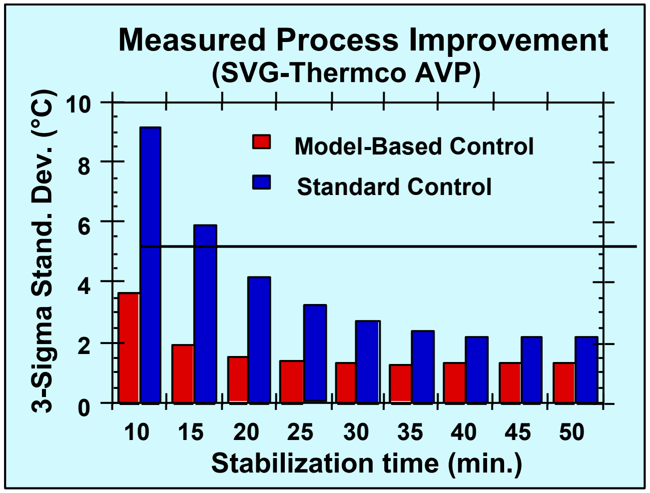
Benefits of Model Based Process Controllers



* Controller design and implementation by Relman, Inc.
* Cray test and evaluation by SEMATECH

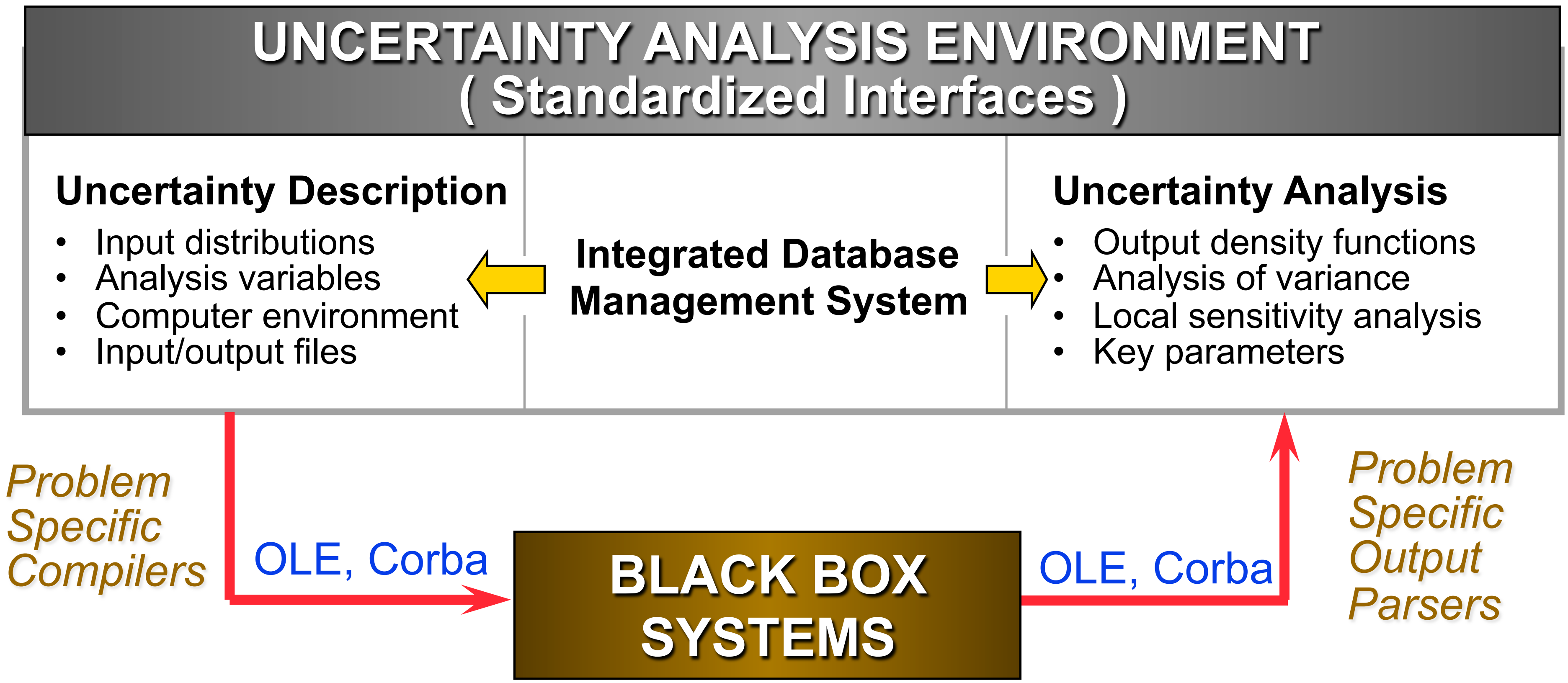
Economic Benefits

- Cost per deposition reduced by 25%
- Process cycle time cut by 20%.
- Stable operations 3 times faster
- 3s uniformity < 4% goal



MIT Chemical Engineering

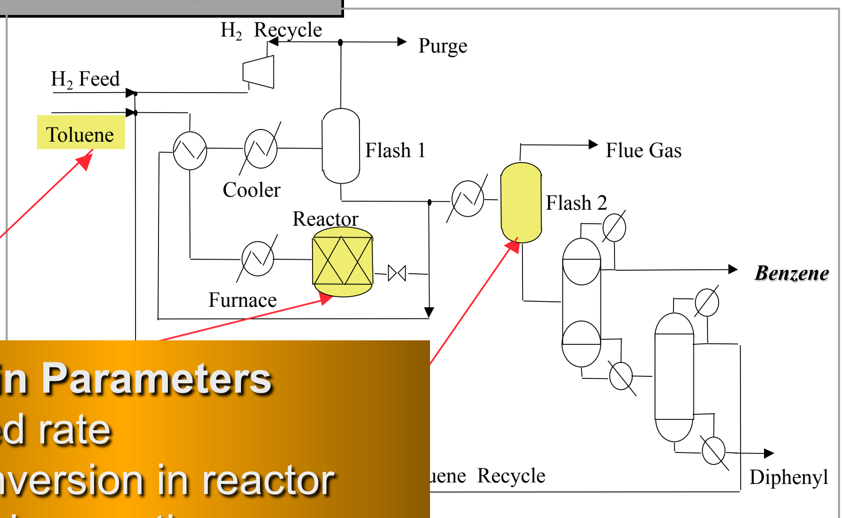
Interacting with "Black Box" Models



- Process Modeling Tools
- Chemical Modeling Tools
- Mathematical Modeling Tools

Example -- Aspen Process Simulation

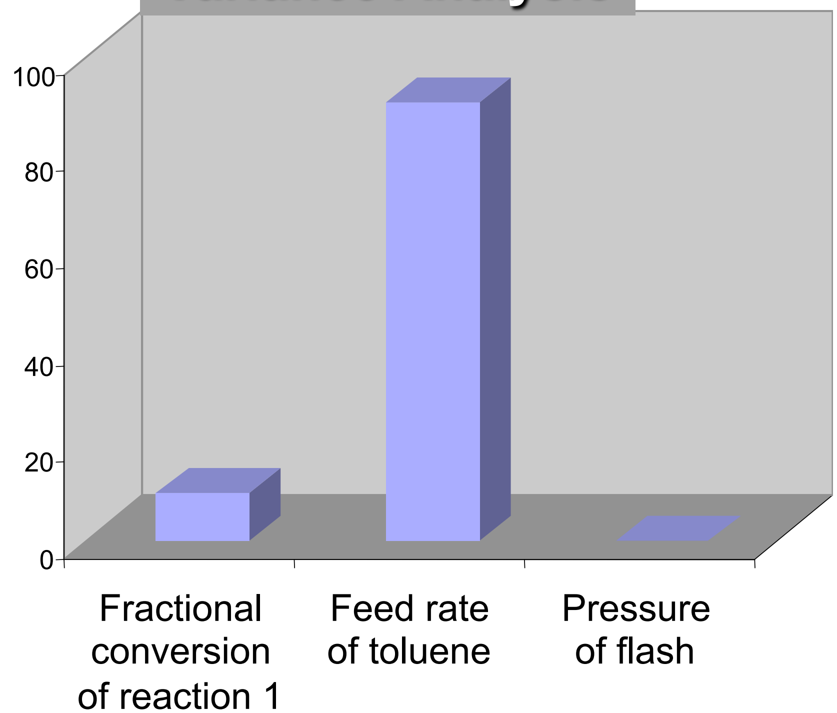
Process Flow Sheet



Uncertain Parameters
- Feed rate
- Conversion in reactor
- Flash operating pressure

Uncertainty Analysis

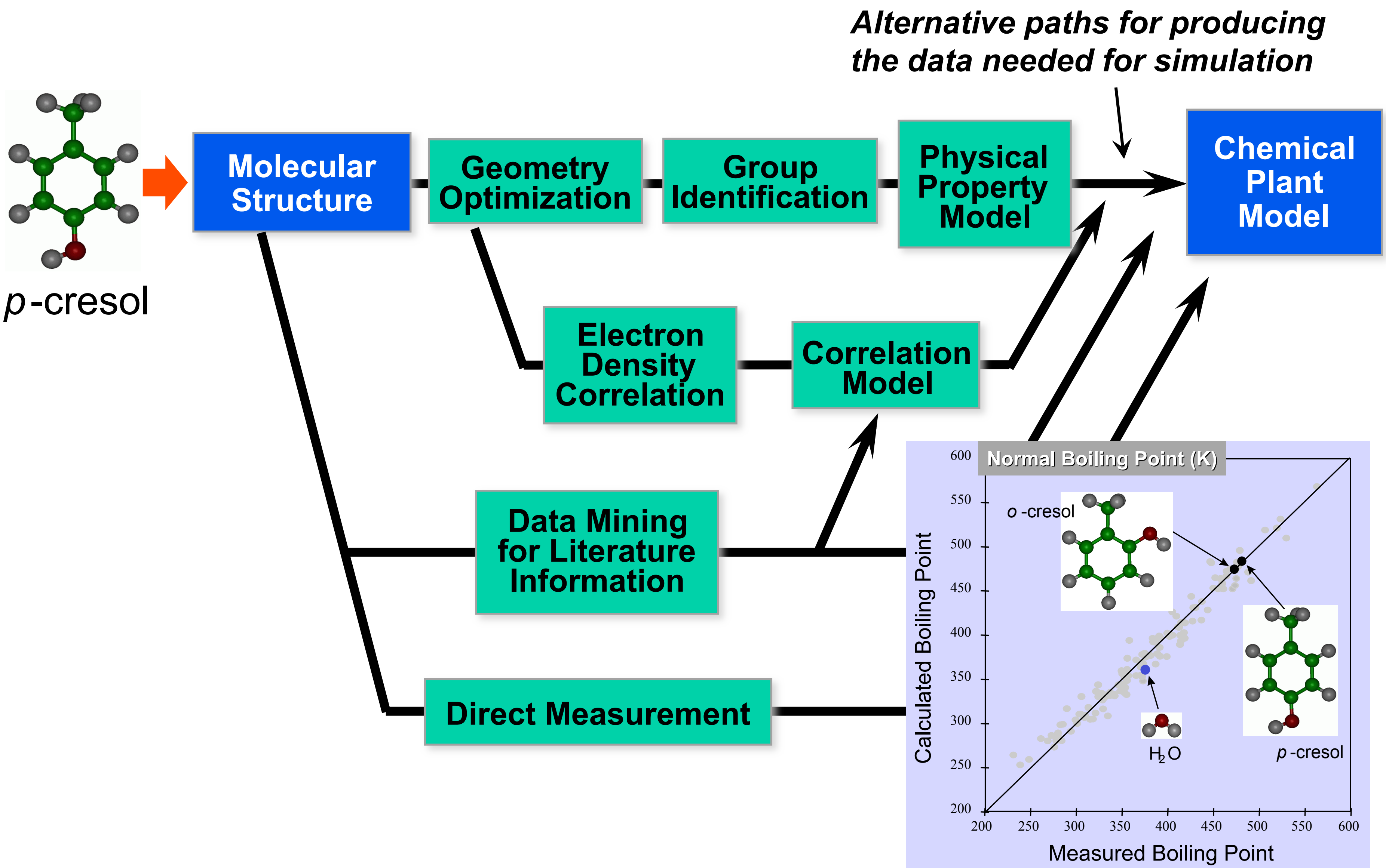
Variance Analysis



➡ Identification of key parameters for further work

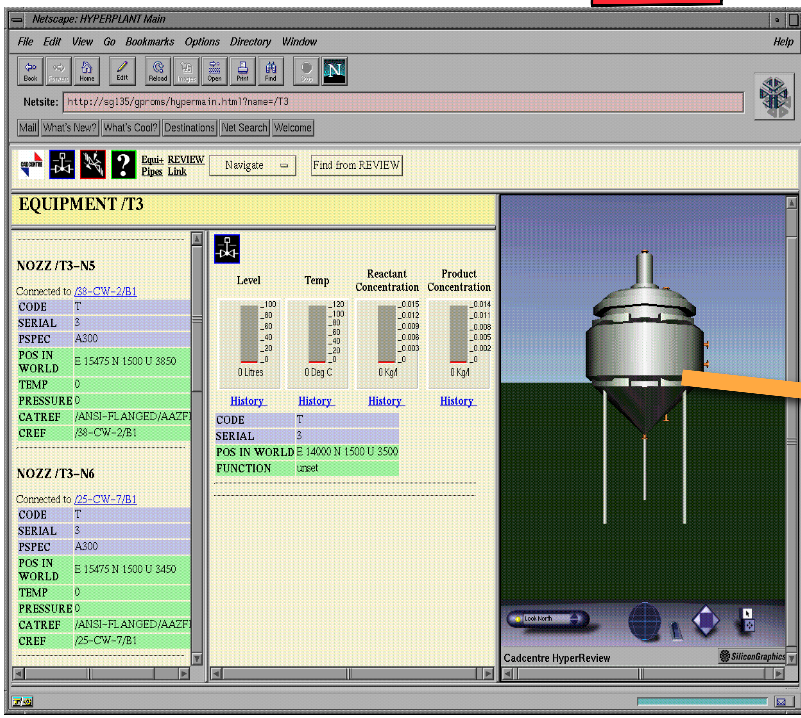
MIT Chemical Engineering

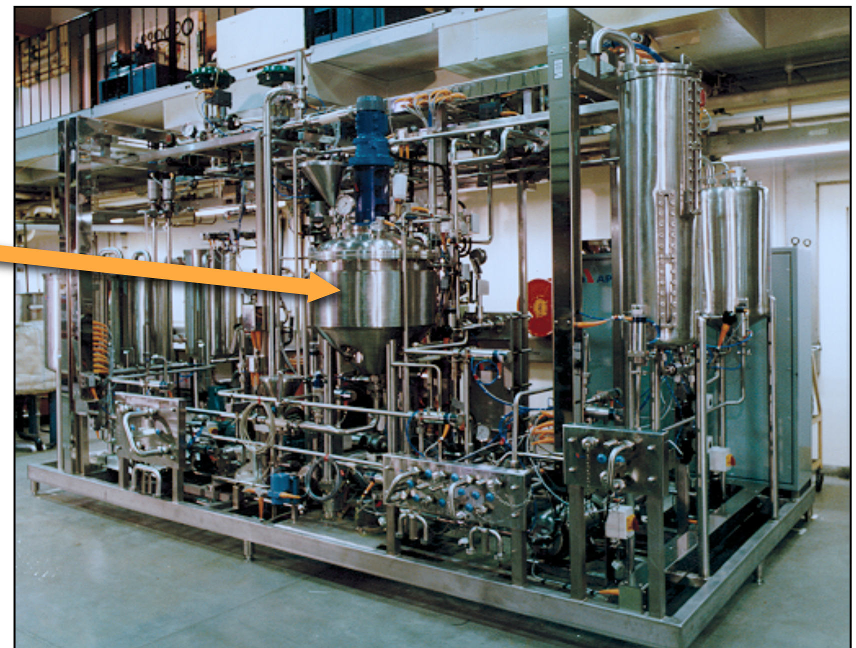
Hierarchical Information – *Physical Properties*



Environments to help Build Models

Model of plant

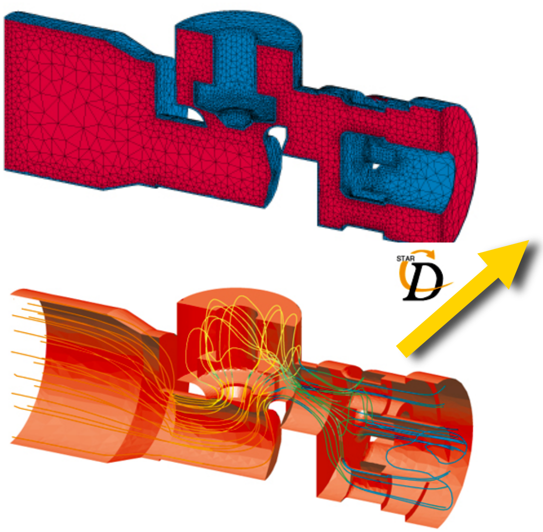


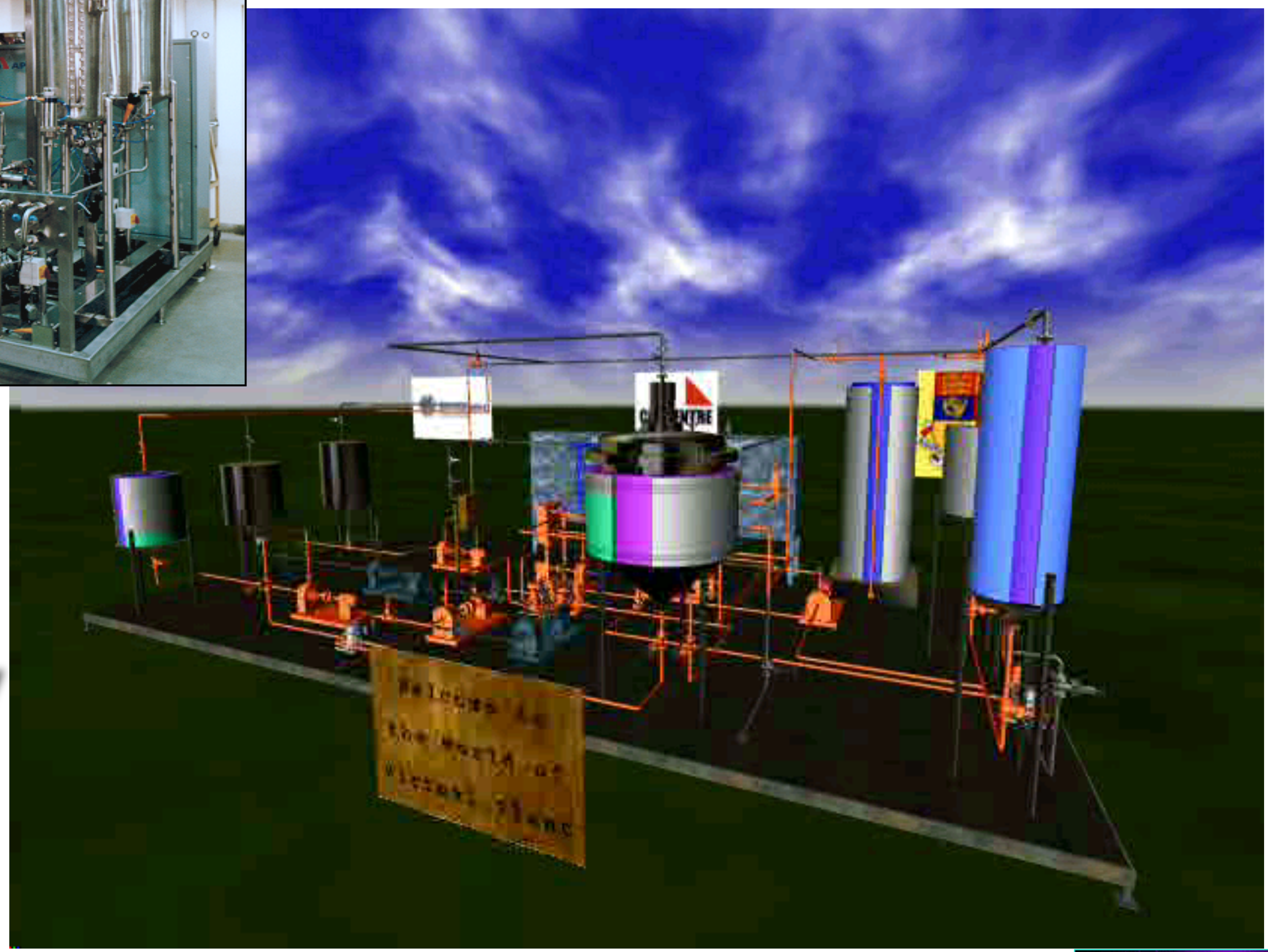


Data from plant

MIT Chemical Engineering

Virtual Plant Walkthrough





SGI, PSE, Adapco

Opportunities from Computational Systems

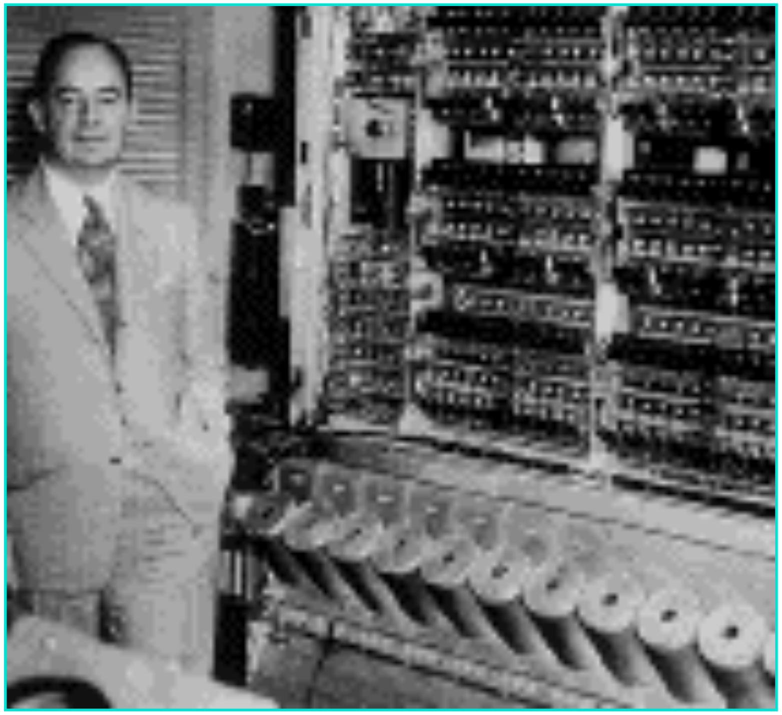
- Reducing the elapsed time for a “solution”
- Hierarchical / linked data bases
  - *Community standards for data exchange*
  - *Corporate knowledge repositories*
- Model based approaches to:
  - *Experimental design and optimization*
  - *Process controllers*
- Immersive data analysis and visualization for:
  - *Data mining and analysis*
  - *Immersive graphical interfaces*

Conclusions

"...While it is hard to predict the future, creating it is much easier…"

We have an exciting opportunity to shape an integrated approach to merging models and data

John von Neumann



Contributions

- Algorithms
- Software
- Hardware architecture
- Practical problems

B.S. Chemical Engineering
ETH Zurich

[redacted]