Machine Learning in Process Systems Engineering:
Opportunities and Challenges

Venkat Venkatasubramanian

Complex Resilient Intelligent Systems Laboratory

Department of Chemical Engineering

Columbia University

New York, NY

AI and ML in PSE

- AI & ML in PSE is a 35-year old idea with many success stories
- What is new and exciting is the availability of
  - Lots of data
  - New algorithms
  - New languages
  - Cheap & powerful computational/communication resources
- Barriers of entry lower now
  - My students used to program in Lisp to do AI & ML
- What took us months to accomplish can be done in days or hours now
- We have entered the "Watson" Era in PSE
- New Opportunities & Challenges

AI and ML in a "Big Data" Era: Importance of Hybrid Models

- Machine learning provides tools for learning Input-Output correlations
- But it's not the mere correlations that we want
  - Target vs ExxonMobil
  - Consumer behavior vs Catalyst design
  - Diaper-Beer vs Yield-Selectivity
- What are the fundamental physics/chemistry/biology principles and mechanisms?
- Desire in-depth knowledge of the phenomena
  - Mechanistic, first-principles based, understanding
  - Makes hybrid models very important: Mechanistic + Data Science
- To have confidence and explanatory power when we make decisions about design, control, optimization, and safety
- How do we go from Big Data to Deep Models and Insights quickly?

*V. Venkatasubramanian, "Drowning in Data: Informatics and Modelling Challenges in a Data-Rich Networked World", Perspective Article, AIChE J., **2009.***

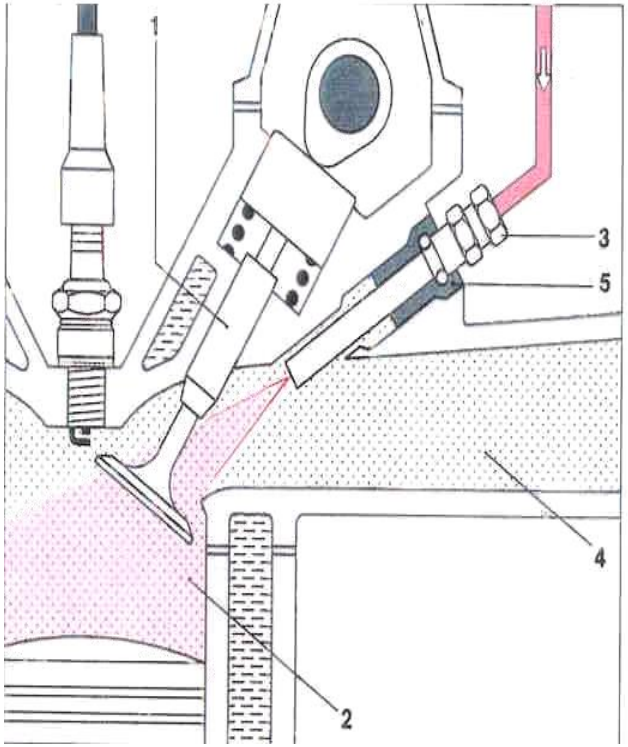
What is Watson?

- Question answering (QA) AI system built by IBM
- Advanced Natural Language Processing, Information Retrieval, Knowledge Representation and Reasoning, and Machine Learning technologies
- ~100 different techniques used to analyze natural language, generate hypotheses, score evidence, and merge and rank hypotheses

"Watson" for PSE?

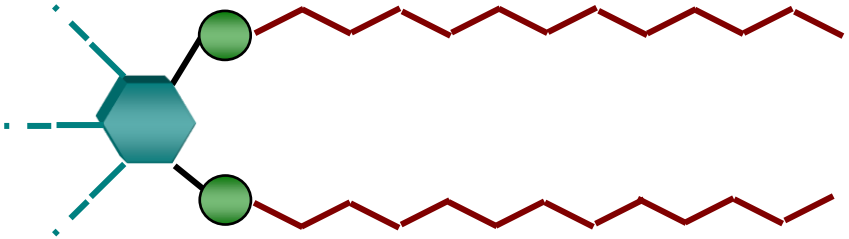
- What will it take to develop "Watson" for PSE?
- Can we develop "Watson" for fault diagnosis, risk analysis, materials design, etc.?
- More challenging than Watson for Jeopardy
  - Not just qualitative facts
  - Quantitative
  - Math Models
  - Charts, Tables, Spectra
  - Heuristic Knowledge

LUBRIZOL: AI for Fuel Additive Design (1995-99)



Mixture formation
1 Intake valve
2 Combustion chamber
3 Fuel-injection valve
4 Intake manifold (tube)
5 Heat-isolating mount

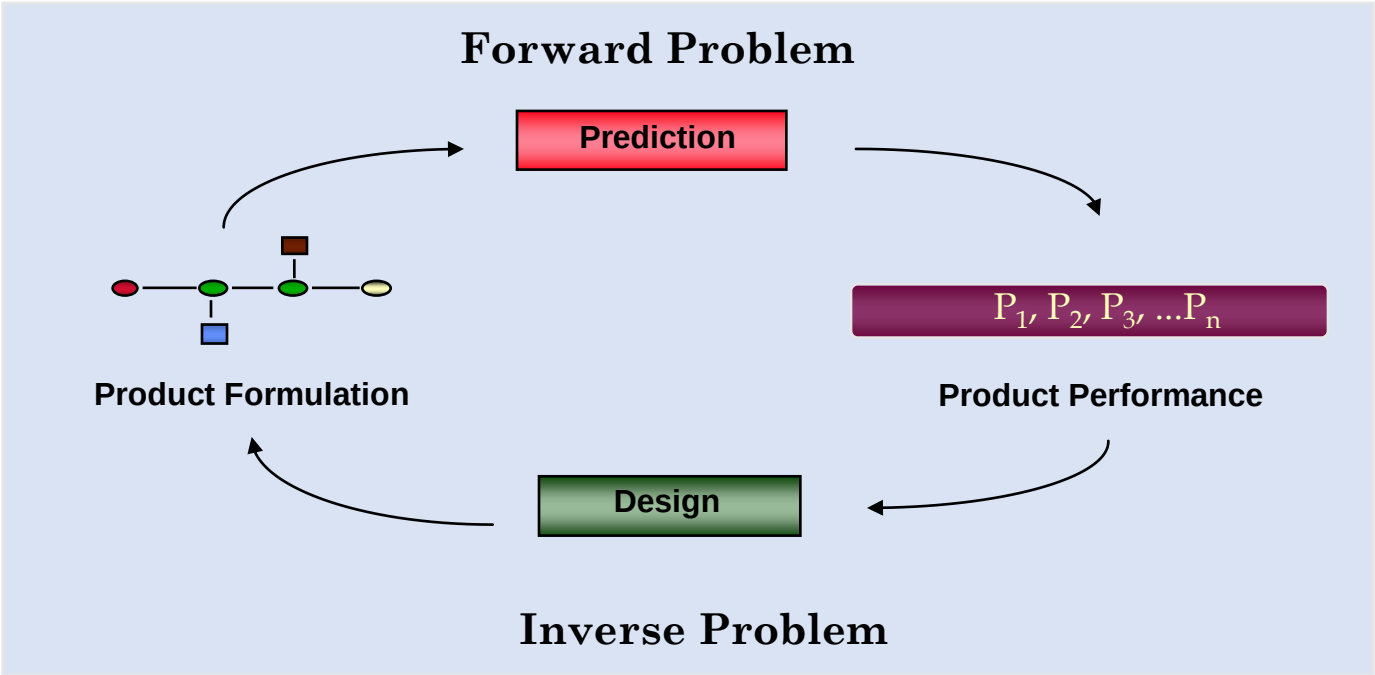
Intake Valve and Manifold



- **EPA Performance Measure**
  - BMW Test for Intake Valve Deposit
  - Stipulated to be less than 100 mg over a 10,000 mile road test
- Fuel-additives are added to gasoline to minimize IVD
- **Expensive testing**
  - Around $10K for a single datum
- **Molecular Product Design Problem:** ***Design fuel-additives that meet desired IVD performance levels***

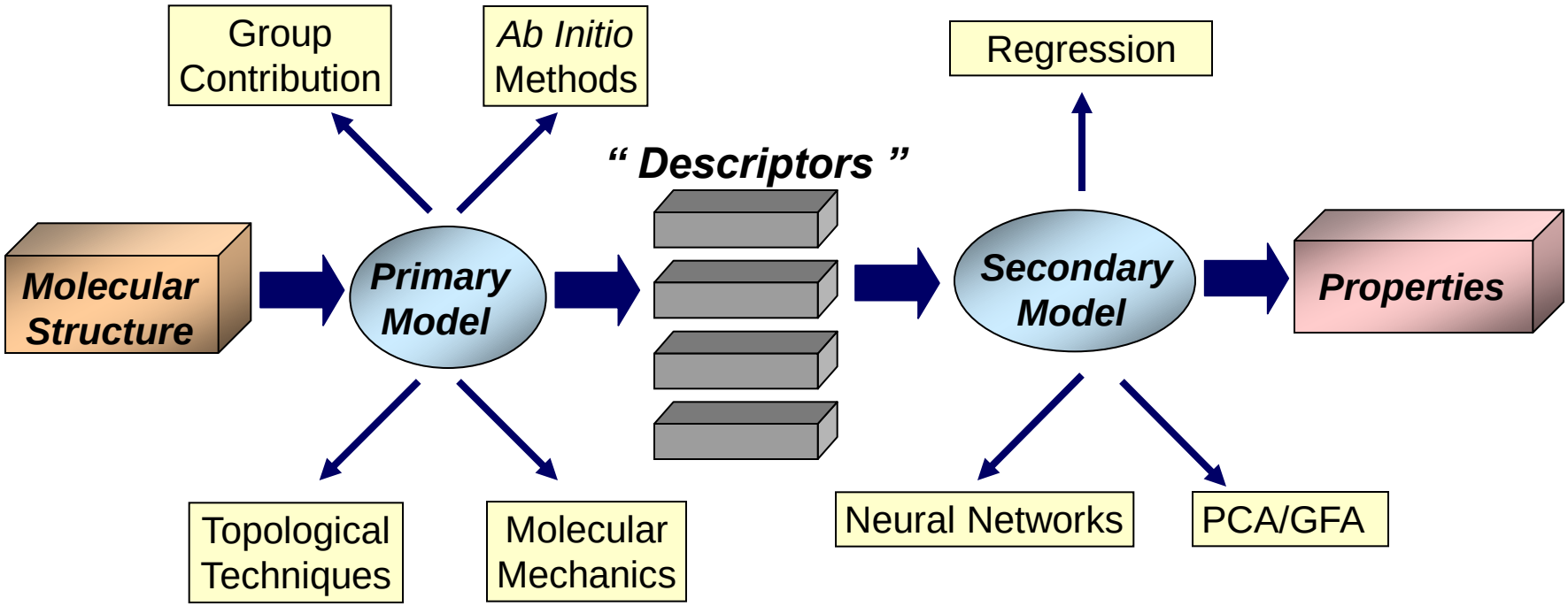
Sundaram, A., Ghosh, P., Caruthers, J.M. and Venkatasubramanian, V., "Design of Fuel Additives Using Neural Networks and Evolutionary Algorithms", *AICHE J.*, 47 (6), pp. 1387 – 1406, **2001.**

Product Formulation and Design

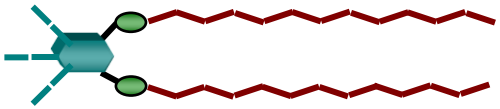


- Forward Problem: Prediction
  - Estimate Product Performance from Formulation
- Inverse Problem: Design
  - Determine a set of products that satisfy desired performance criteria

Hybrid Model: First-principles + Data Science

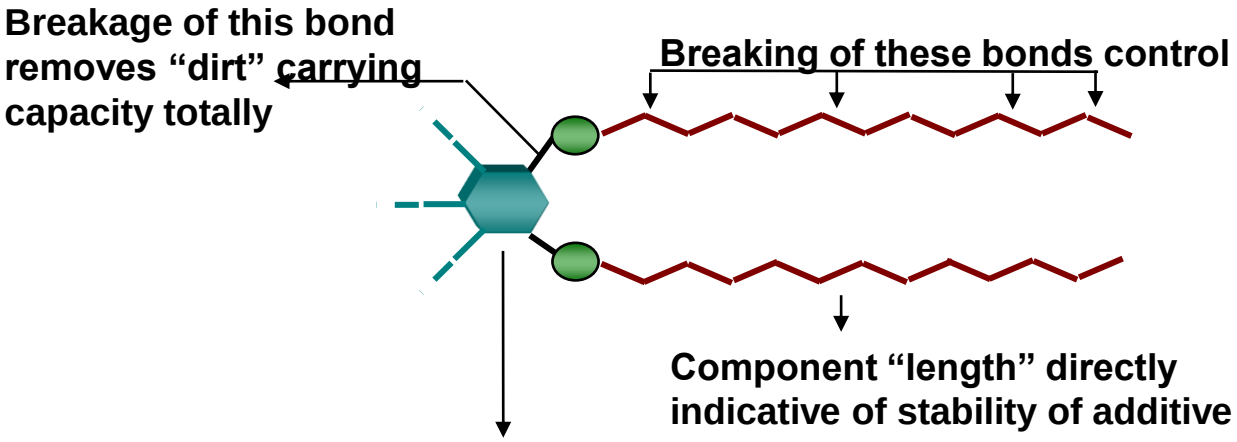


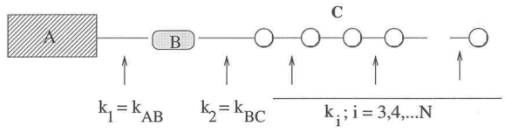
First-principles Model:
Polymer link breakage dynamics



Data Science Model

First-Principles Model for Additive Degradation: Population Balance Equations





- First-principles model
- Tracks the structural distribution of fuel-additive with time due to reactive degradation

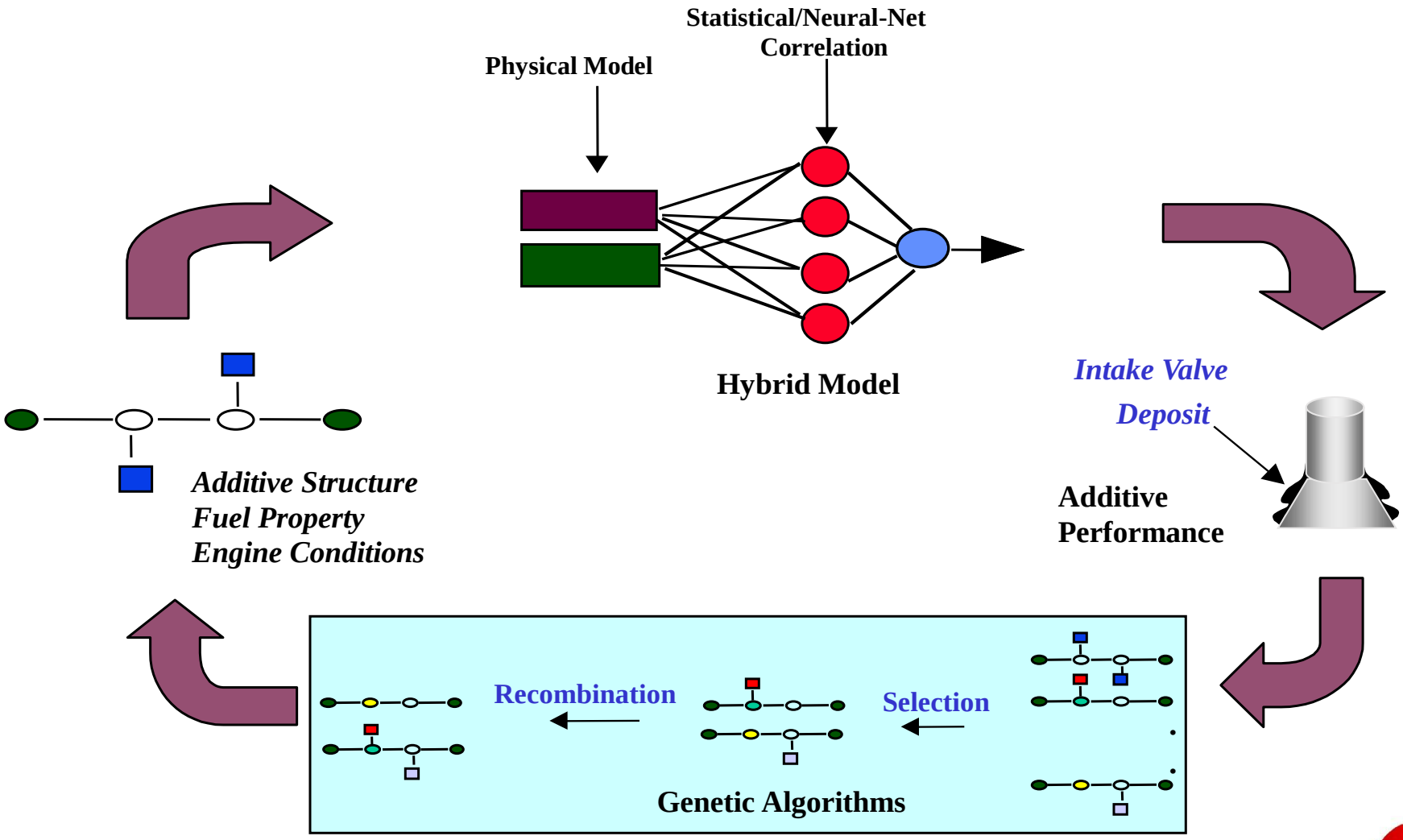
$$\frac{d\mathbf{X}}{d\tau} = A \bullet \mathbf{X};\ \mathbf{A} = \begin{bmatrix} -k_1 & k_2 & k_3 & \cdots & k_N \\ 0 & -k_1-k_2 & k_2 & \cdots & k_N \\ \cdot & 0 & -k_1-k_2-k_3 & k_3 & \cdots & k_N \\ \cdot & \cdot & 0 & \cdots & k_N \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & -\sum_{i\le j} k_i & k_{i+1} & \cdots & k_N \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & 0\ 0 & 0 & \cdots & -\sum_{i=1}^{N} k_i \end{bmatrix};\ \mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \cdot \\ X_j \\ \cdot \\ X_N \end{bmatrix}$$

$$X_0 = \sum_{i=1}^{N} k_i * X_i$$

$$\mathbf{X}(\tau = 0) = \mathbf{P}(\mu, \sigma)$$

Population Balance Eqns

Discovery Analytics for Fuel-Additives Design



Sundaram, A., Ghosh, P., Caruthers, J.M. and Venkatasubramanian, V., "Design of Fuel Additives Using Neural Networks and Evolutionary Algorithms", *AICHE J.*, 47 (6), pp. 1387 – 1406, **2001.**

CRIS.lab

Objective: Determine a structure with IVD < 10 mg

Rank/Identifier	Fitness	Predicted IVD	Structural Description
III-1	1.000	8.9 mg	**Novel Structure New Chemistry**
I-2	0.996	11.5 mg	**Novel Structure New Chemistry**
I-6	0.993	12.0 mg	Found by the traditional approach also

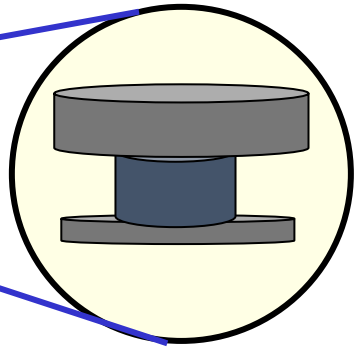
Found new candidates in minutes

Opened up new possibilities: New Insights

Think "Out of the Box"

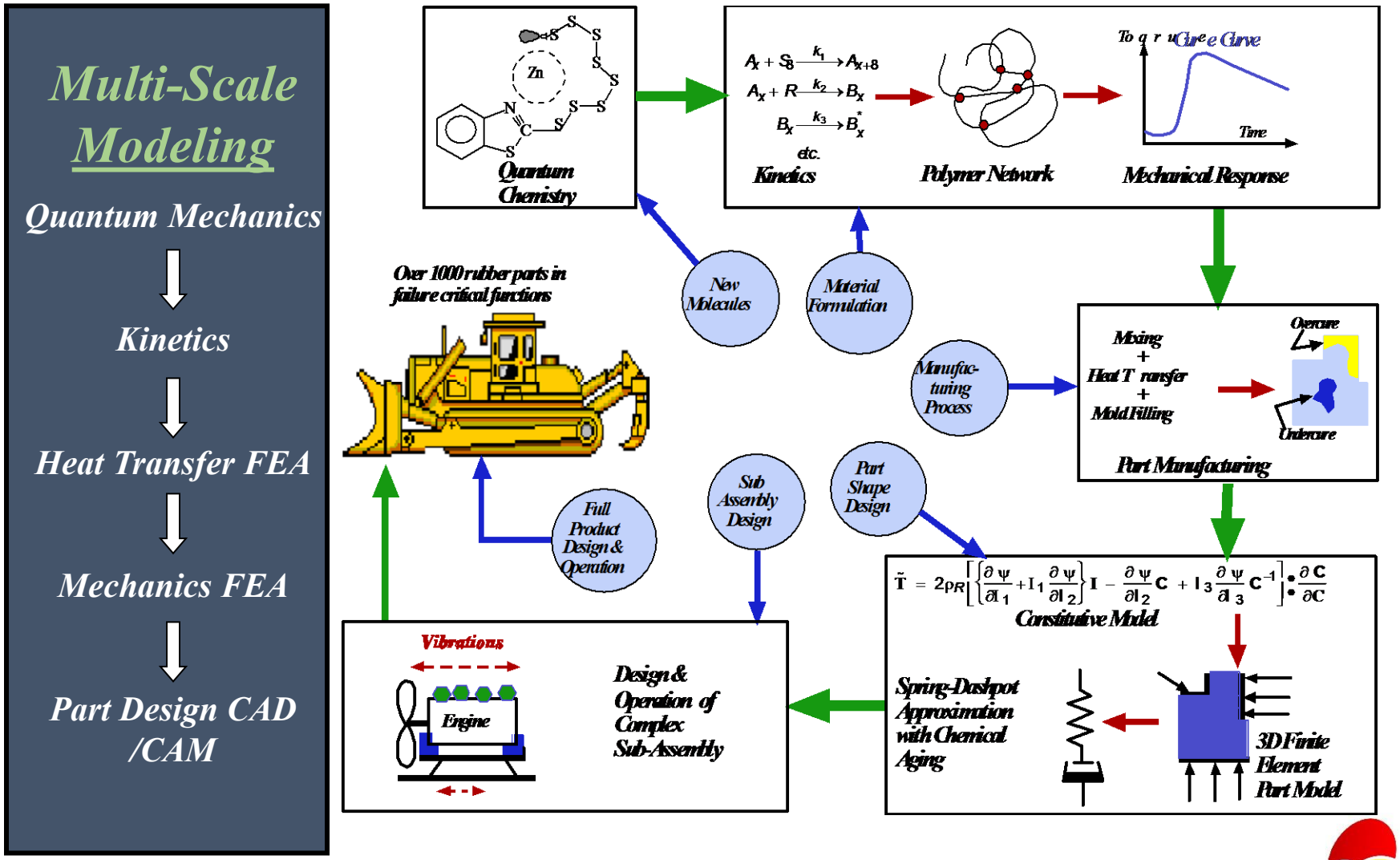
Sundaram, A., Ghosh, P., Caruthers, J.M. and Venkatasubramanian, V., "Design of Fuel Additives Using Neural Networks and Evolutionary Algorithms", *AICHE J.*, 47 (6), pp. 1387 – 1406, **2001.**

CATERPILLAR (1998-2001)

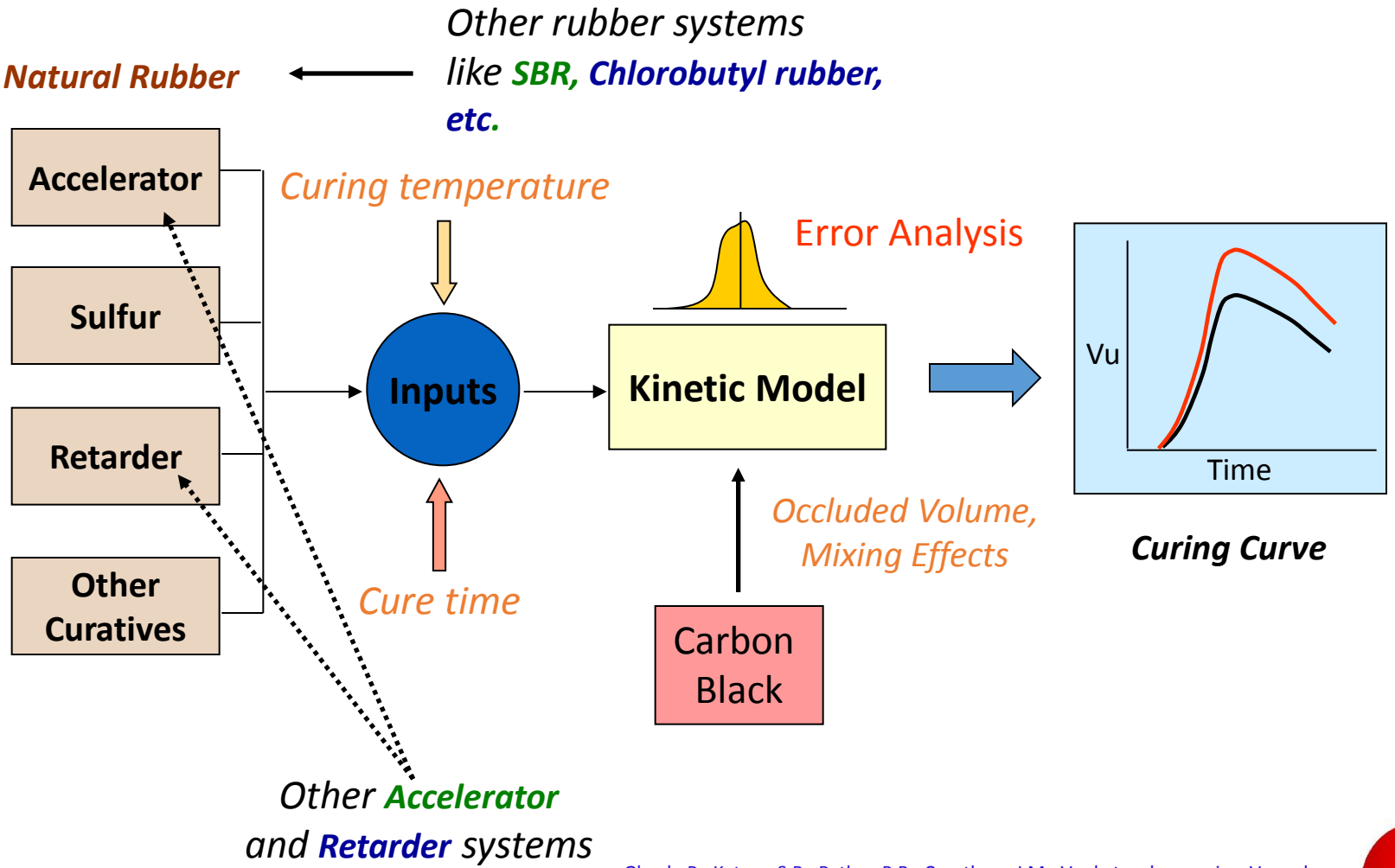


- About 1000 Rubber Parts in Failure Critical Functions
- Tires, Treads, Hoses, Shock Absorbers, O-rings, Gaskets, Mounts …
- Reliability and Warranty Problems

Caterpillar's Multi-Scale Modeling Challenge: Design of Formulated Rubber Parts



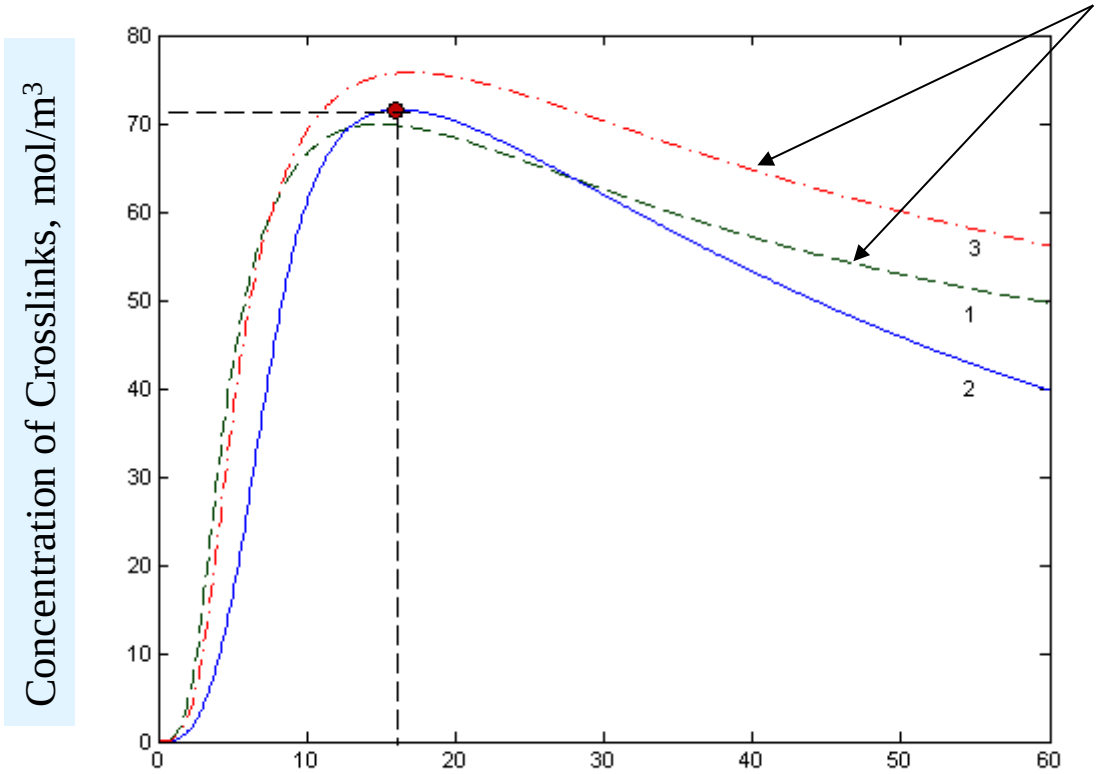
Forward Model: First-principles



Ghosh, P., Katare, S.R., Patkar, P.R., Caruthers, J.M., Venkatasubramanian, V., and Walker, K.A., "Sulfur Vulcanization of Natural Rubber for Benzothiazole Accelerated Formulations: From Reaction Mechanisms to a Rational Kinetic Model", *Rubber Chemistry and Technology*, 76 (3), pp. 592 – 693, **2003.**

Inverse Problem Solution

Formulation	Fitness	T_{max}	Vu_{max}	σ_{100}	σ_{200}
[1.75 1.25 0.05 0 315]	0.9999	15	69.92	0.91	1.63
[0.50 4.00 0.10 0 310]	0.9999	16	71.57	0.92	1.65
[1.75 1.50 0.05 0 310]	0.9996	17	75.75	0.97	1.75
[0.75 2.75 0.10 0 315]	0.9994	13	67.88	0.88	1.58



New Formulations

Better Formulations

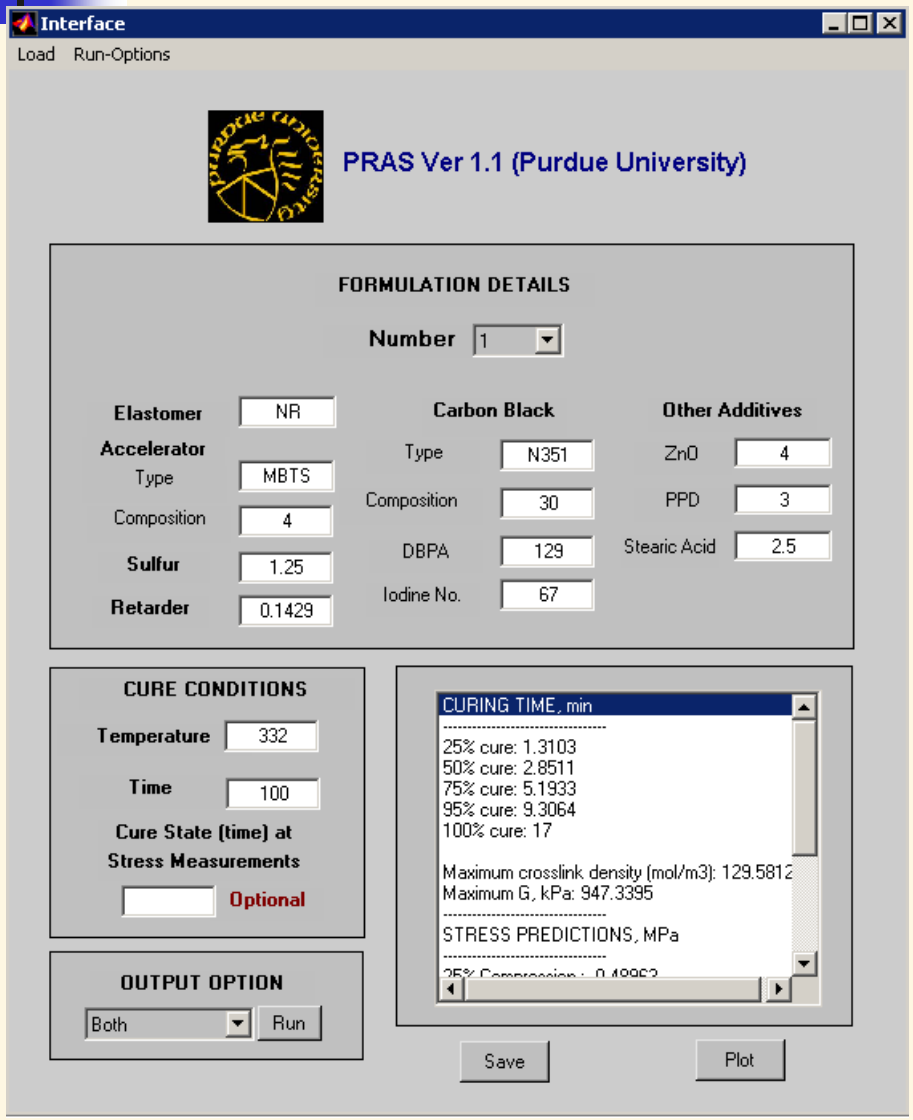
Designed in hours instead of about 2 weeks taken by human experts

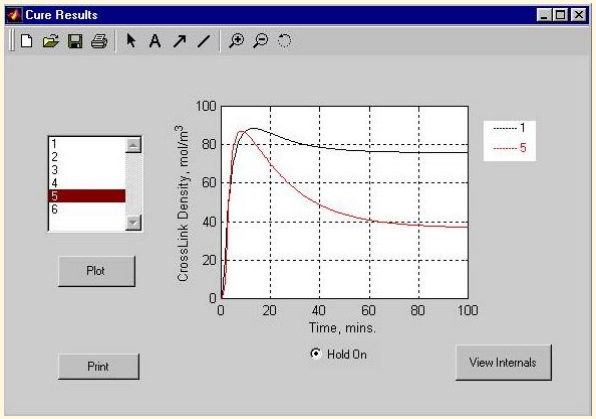
Used daily at Caterpillar

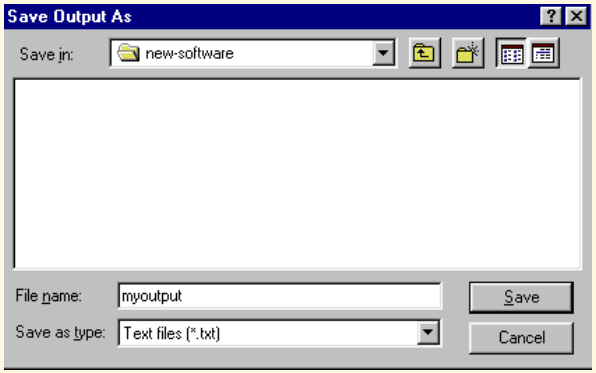
Ghosh, P., Katare, S.R., Patkar, P.R., Caruthers, J.M., Venkatasubramanian, V., and Walker, K.A., "Sulfur Vulcanization of Natural Rubber for Benzothiazole Accelerated Formulations: From Reaction Mechanisms to a Rational Kinetic Model", *Rubber Chemistry and Technology*, 76 (3), pp. 592 – 693, **2003.**

CRIS.lab

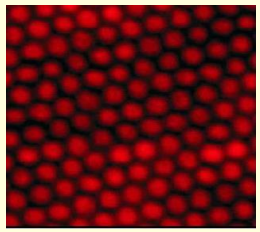
AI System Used at Caterpillar on a Daily Basis



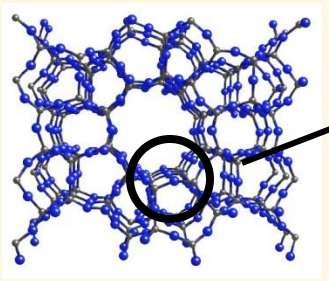


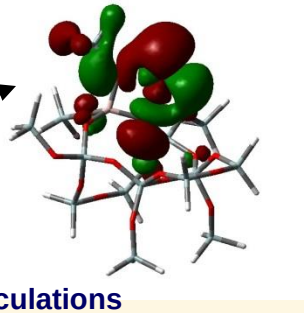


ExxonMobil's Grand Challenge: Catalyst Design via Combinatorial Chemistry



Catalyst Structures





Quantum calculations

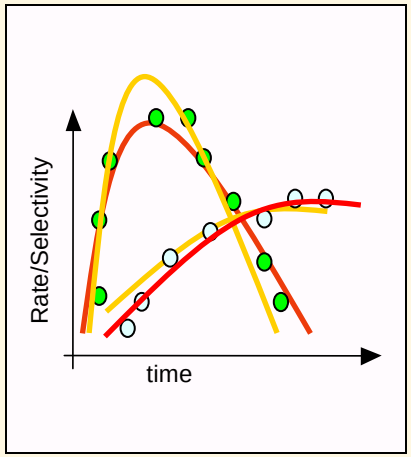
d band center
Electronegativity
⋮
Electronic/Structural descriptors

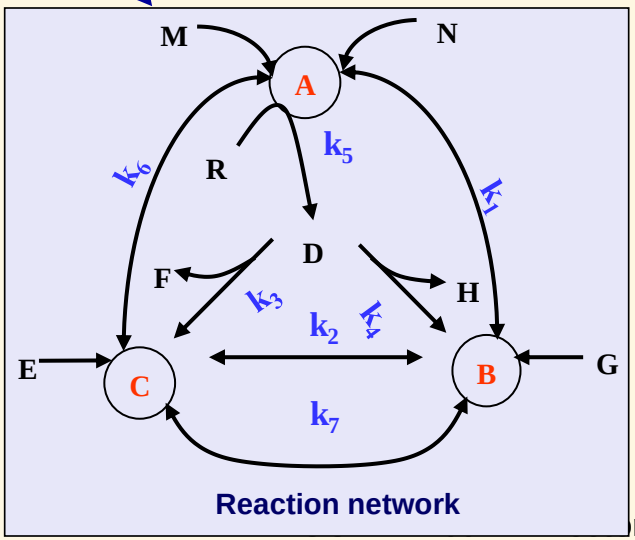
Fundamental Chemistry

Reaction kinetics

Performance measures

Synthesis Characterization





Reaction network

Paraffin Aromatization

- Identify a catalyst formulation for light paraffin aromatization that is superior to Ga/H-ZSM-5 in terms of:
  - Higher Benzene, Toluene, Xylene (B/T/X) selectivity
  - Higher Hydrogen selectivity

- Microkinetic model development for the kinetic description of the system

Caruthers, J.M., Lauterbach, J.A., Thomson, K.T., Venkatasubramanian, V., Snively, C.M., Bhan, A., Katare, S. and Oskarsdottir, G., "Catalyst Design: Knowledge Extraction from High Throughput Experimentation", *Journal of Catalysis*, vol. 216/1-2, pp. 98 – 109, **2003**.

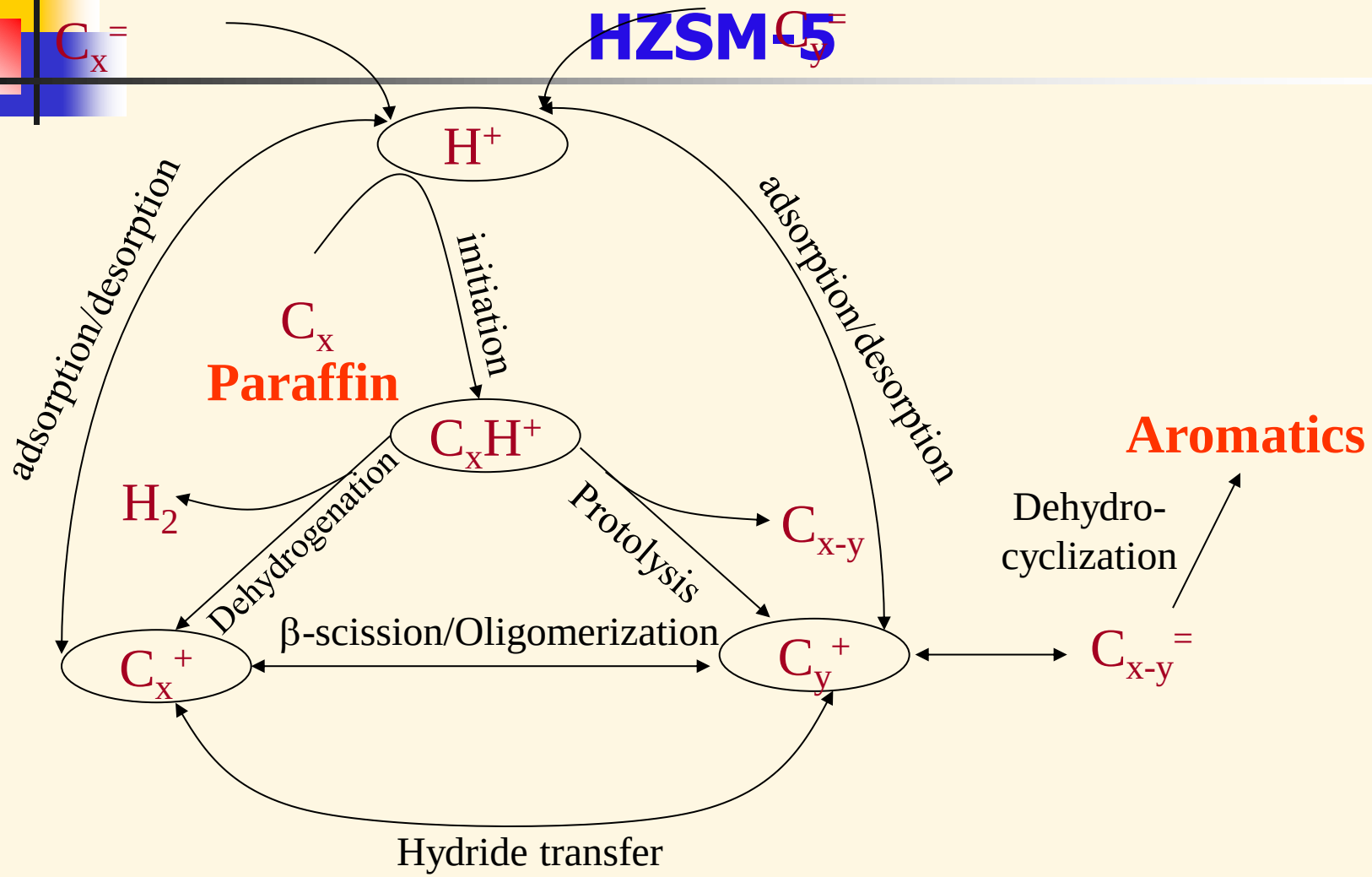
Katare, S., Caruthers, J.M., Delgass, W.N., and Venkatasubramanian, V., "An Intelligent System for Reaction Kinetic Modeling and Catalyst Design", *Ind. Eng. Chem. Res. and Dev.*, 43(14), pp. 3484 – 3512, **2004.**

Hsu, H. – S., Krishnamurthy, B., Rao, P., Zhao, C., Jagannathan, S. and Venkatasubramanian, V., "A Domain-specific Compiler Theory Based Framework for Automated Reaction Network Generation", *Comp. Chem. Eng.,* 32, pp. 2455 – 2470, **2008.**

Blau, G., Lasinski, M., Orcun, S., Hsu,S-H., Caruthers, J., Delgass, N. and Venkatasubramanian, V., "High Fidelity Mathematical Model Building with Experimental Data: A Bayesian Approach", *Comp. Chem. Eng.,* 32(2008), pp. 971 – 989, **2008.**

Hsu, S.–H.,Stamatis, S.D., Caruthers, J.M., Delgass, W.N., Venkatasubramanian, V., Blau, G.E., Lasinski, M. and Orcun, S., "Bayesian framework for building kinetic models of catalytic systems", *Ind. Eng. Chem. Res.*, 48 (10),cpp. 4768 – 4790., **2009.**

Reaction Network – Propane Aromatization on HZSM-5



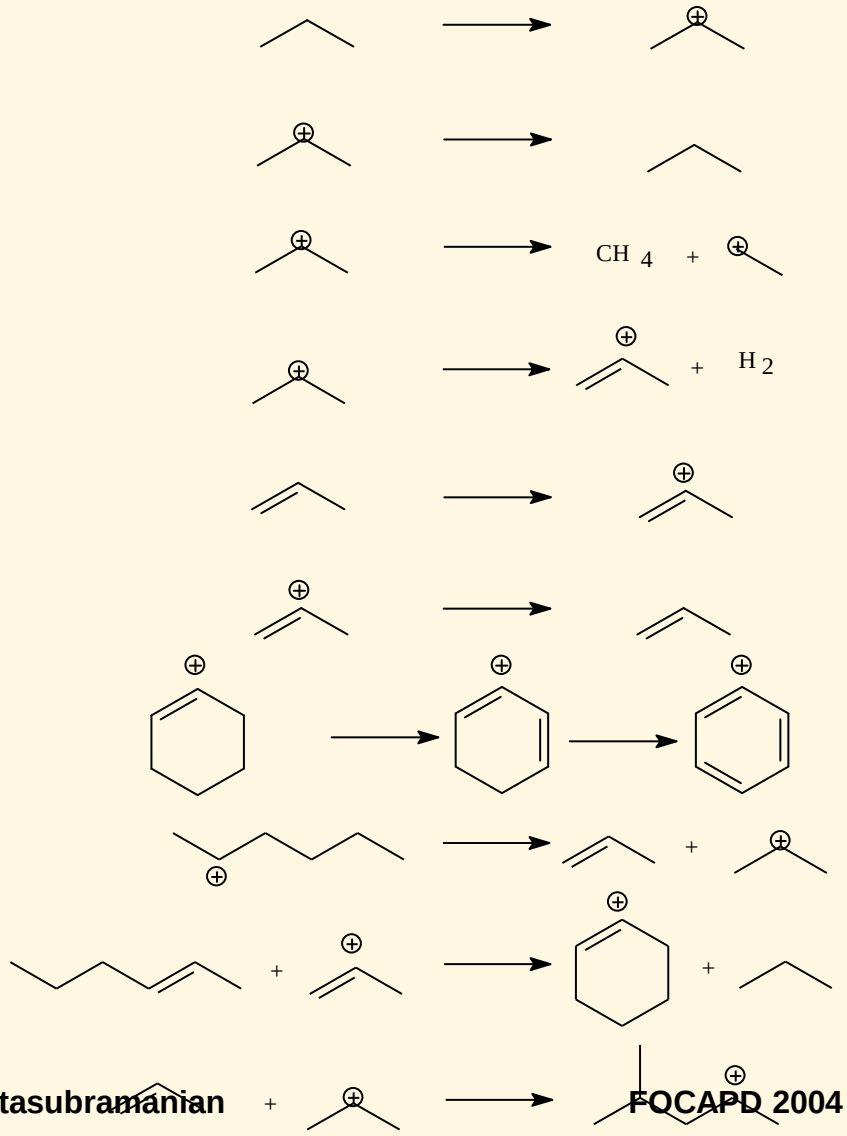
- 31 gas phase species + 29 surface species + 271 reaction steps
- Model with 31 ODEs, 29 algebraic equations
- 13 parameters with up to 10 orders of magnitude bounds on each

Chemistry Rules for Propane Aromatization on HZSM-5

Chemistry Rules

1. Alkane adsorption
2. Alkane desorption
3. Carbonium ion protolysis
4. Carbonium ion dehydrogenation
5. Olefin adsorption
6. Olefin desorption
7. Aromatization
8. Beta-Scission
9. Hydride Transfer
10. Oligomerization

Representative Chemical Reactions



Reaction Modeling Suite (RMS)

English Language Rules

Chemistry

8. Beta Scission
transforms a carbenium ion into a smaller carbenium ion and an olefin

Grouping

8. a. Formation of a secondary carbenium ion
is 20 times faster than a primary carbenium ion

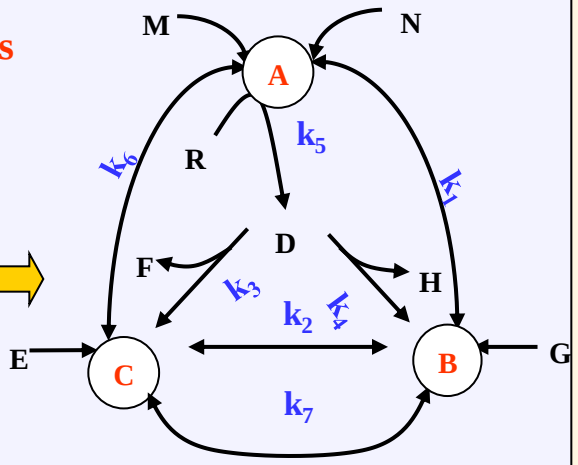
b. Formation of a tertiary carbenium ion
is 60 times faster than a primary carbenium ion

Reaction Description Language Plus

Beta Scission
Label-site c1+ (find positive carbon)
Label-site c2 (find neutral-carbon attached-to c1+)
Label-site c3 (find neutral-carbon attached-to c2)
Forbid (primary c3)
Forbid (less-than (size-of reactant) 9)
Disconnect c2 c3)
Increase-order-of (find bond connecting c1+ c2)
Add-charge c3
Subtract-charge c1+

Beta Scission
Label-site c1+ (find positive carbon)
Require (c1+ primary and product)
set-k k1
Label-site c2+ (find positive carbon)
Require (c2+ secondary and product)
set-k 20*k1
Label-site c3+ (find positive carbon)
Require (c2+ tertiary and product)
set-k 60*k1

Reaction Network



Model Generator

Mathematical Equations

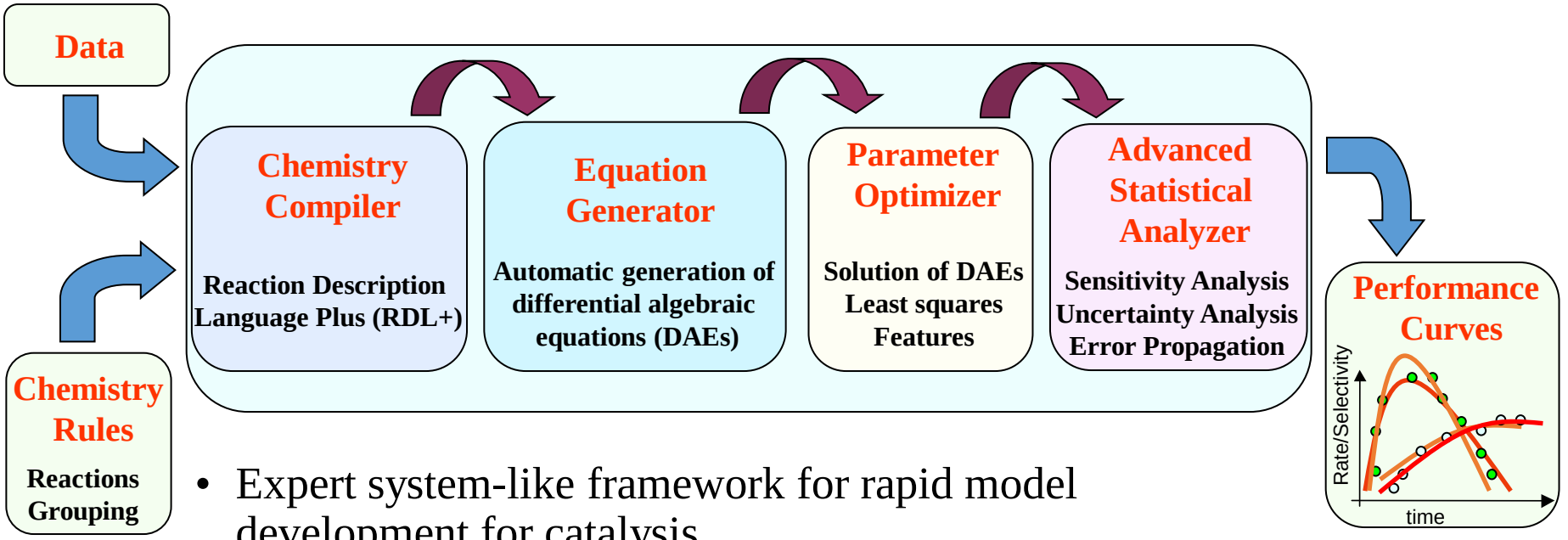
$$dC_A/dt = -k_1 C_A$$

$$dC_B/dt = k_1 C_A + k_4 D - k_5 B$$

$$\theta_A + \theta_B + \theta_C = 1$$

100's of DAE's

Discovery Analytics: AI-based Modeling Environment

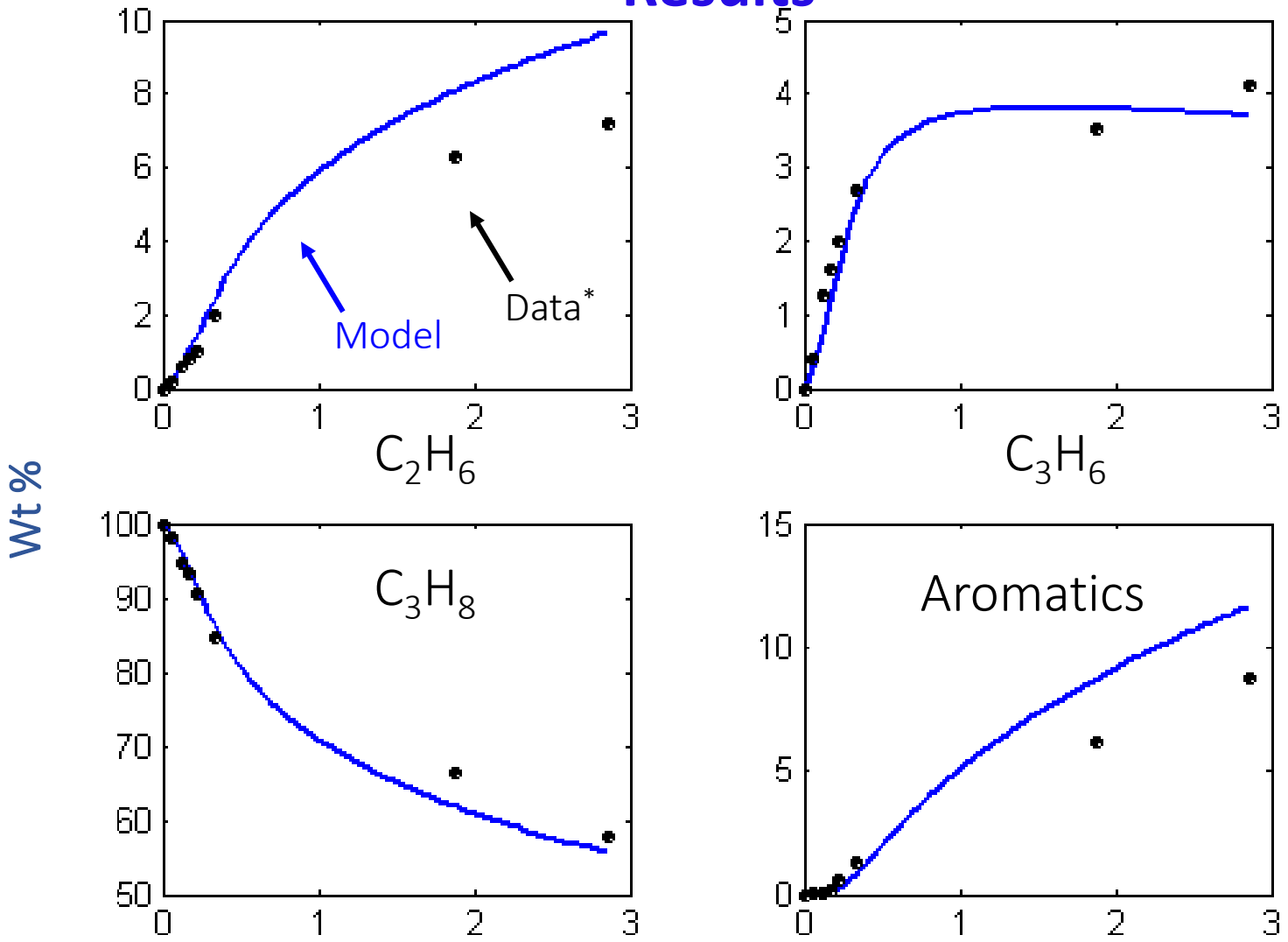


- Expert system-like framework for rapid model development for catalysis
- Differential-Algebraic Equation (DAE) System
- Dozens of equations and parameters
- Real-time model development

Katare, S., Caruthers, J.M., Delgass, W.N., and Venkatasubramanian, V., "An Intelligent System for Reaction Kinetic Modeling and Catalyst Design", *Ind. Eng. Chem. Res. and Dev.*, 43(14), pp. 3484 – 3512, **2004.**

Hsu, H. – S., Krishnamurthy, B., Rao, P., Zhao, C., Jagannathan, S. and Venkatasubramanian, V., "A Domain-specific Compiler Theory Based Framework for Automated Reaction Network Generation", *Comp. Chem. Eng.*, 32, pp. 2455 – 2470, **2008.**

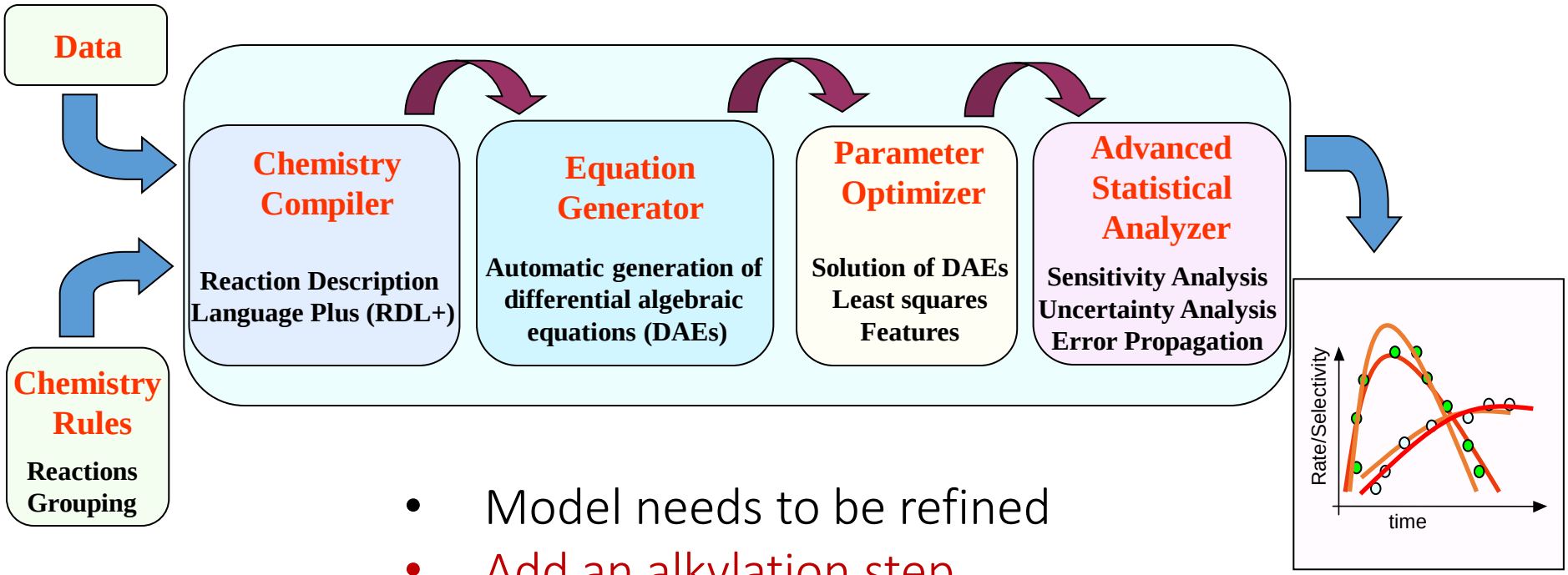
Results



Over prediction of C_2s and under prediction of C_3s

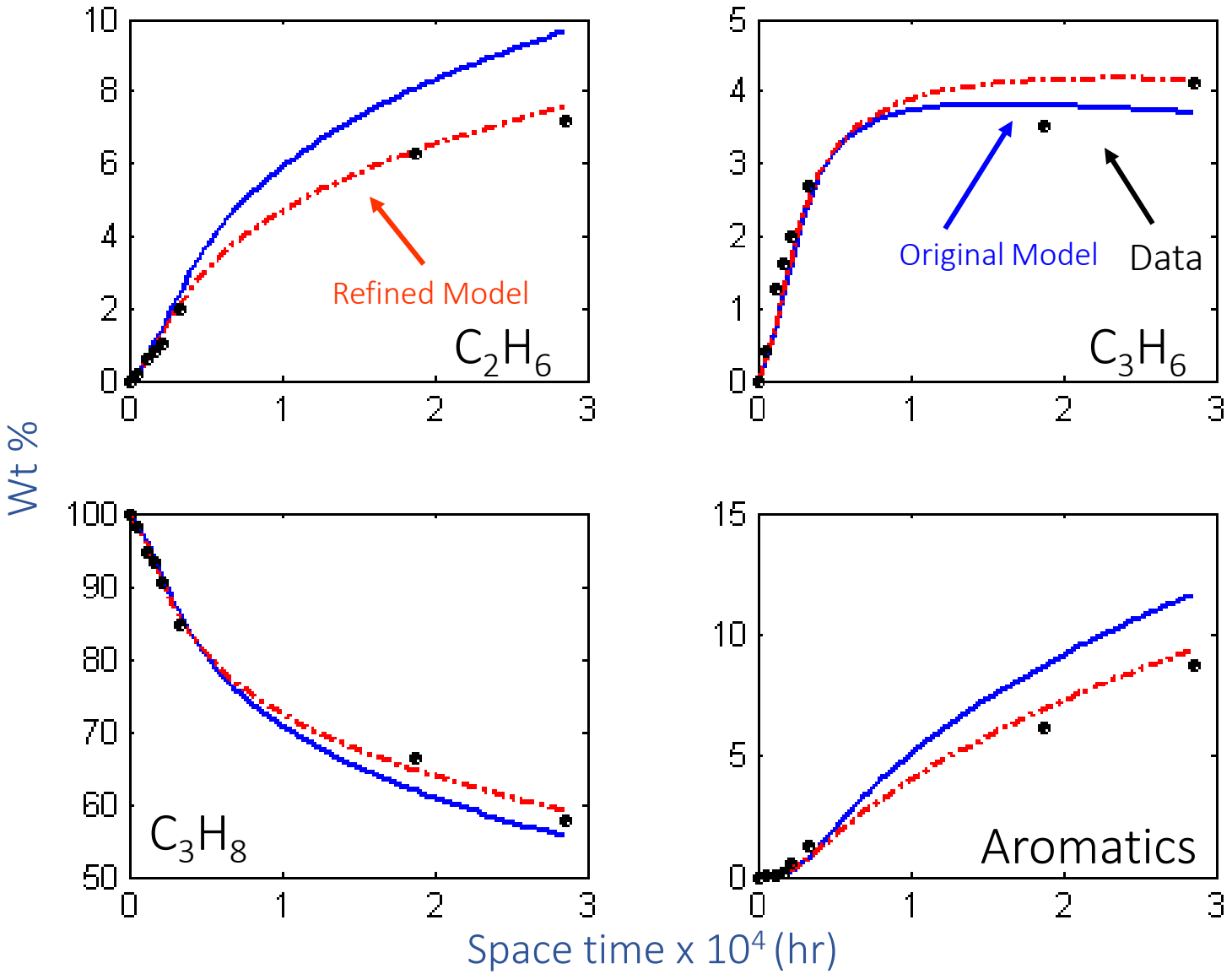
* Lukyanov et. al., Ind. Eng. Chem. Res. , 34, 516-523, 1995

Ontological Informatics Cyberinfrastructure: Reaction Modeling Suite (RMS++)



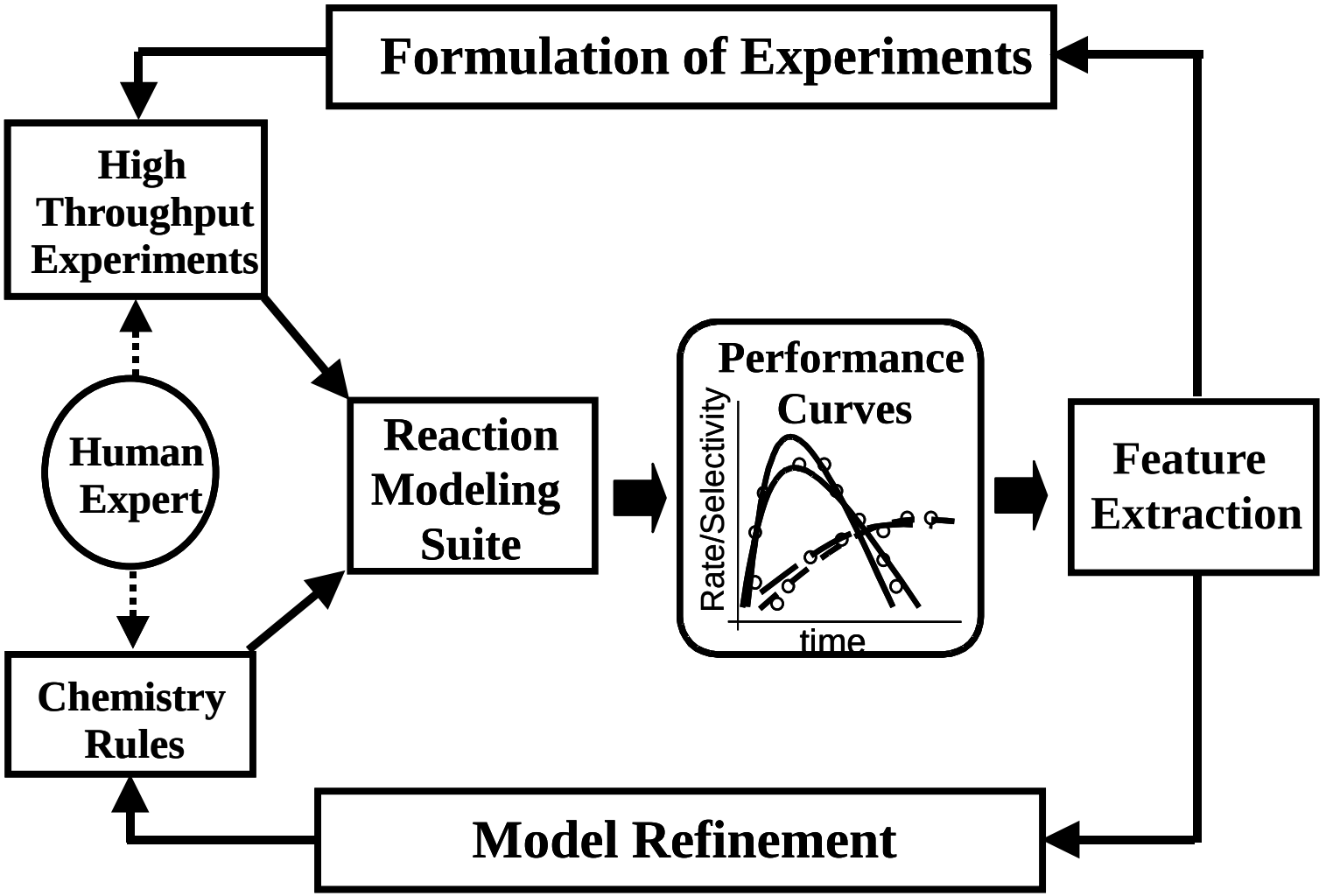
- Model needs to be refined
- Add an alkylation step
- To convert lighter alkanes to higher ones

Results – Model Refinement



Refined model adds alkylation step to convert lighter alkanes to higher ones

Materials Discovery Analytics Cycle: Active Learning



Caruthers, J.M., Lauterbach, J.A., Thomson, K.T., Venkatasubramanian, V., Snively, C.M., Bhan, A., Katare, S. and Oskarsdottir, G., "Catalyst Design: Knowledge Extraction from High Throughput Experimentation", *Journal of Catalysis*, vol. 216/1-2, pp. 98 – 109, **2003**.

Summary

- Vision: "Watson" for PSE
  - AI & ML in PSE is a 35-year old idea with many past success stories
  - What is new is the availability of great amounts of data, new algorithms, new languages, and cheap & powerful computational/communication resources
  - What took us months to accomplish can be done in days or hours now
- Most scientific & engineering applications likely require hybrid models: First-principles + Data Science
- Need to develop modeling environments to build hybrid models quickly
- Require progress in building domain-specific components
  - Ontologies, Compilers, Molecular Structure and Semantic Search Engines
  - Chemical entities extraction systems, Languages for modeling chemical reaction pathways, etc.
  - These are the next generation Aspen+ and gPROMS

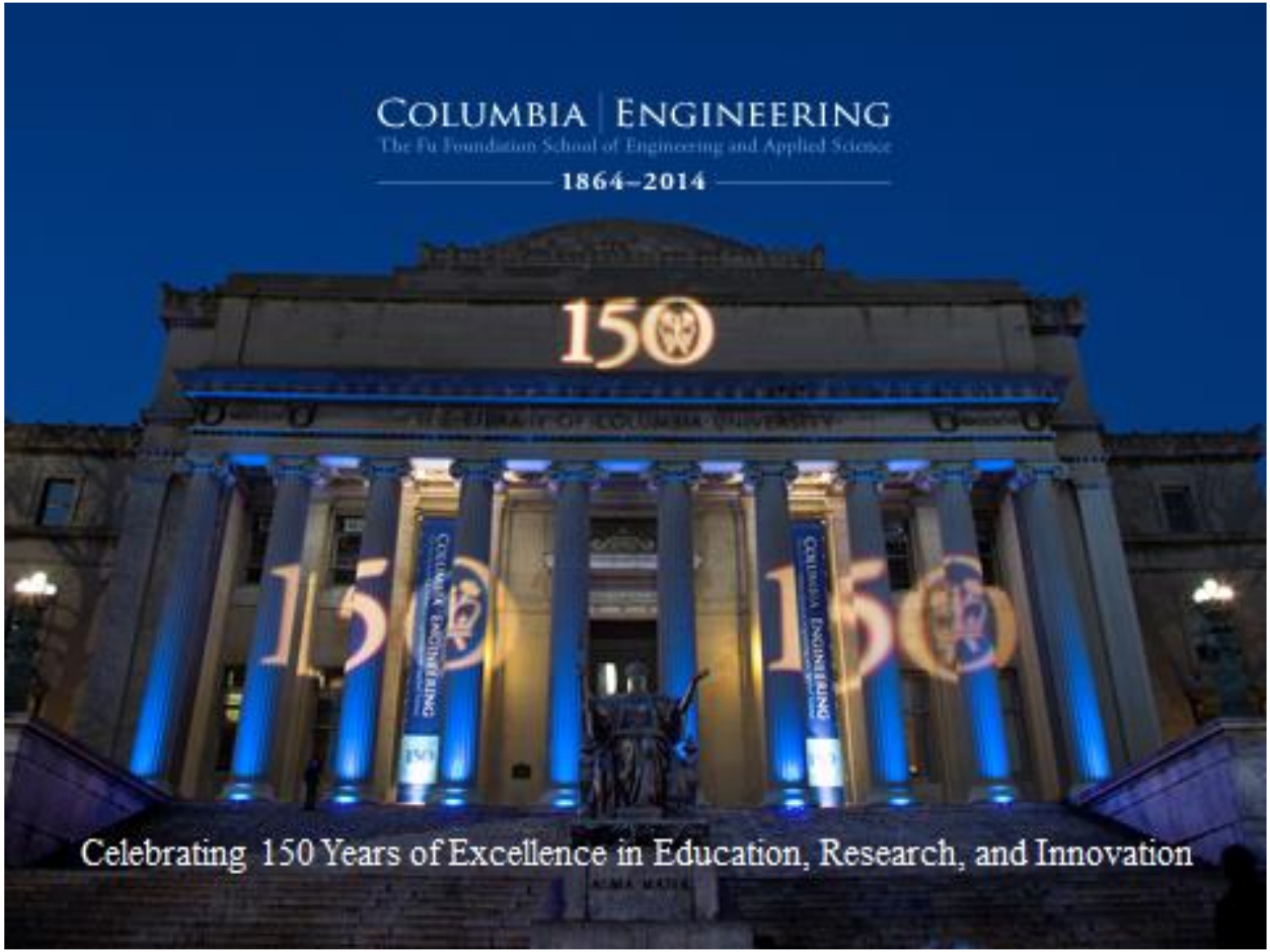
Acknowledgements

- **Prof. G. V. Reklaitis, Prof. K. Morris, Dr. P. Basu, Prof. J. M. Caruthers, Prof. W. N. Delgass**
- Dr. K. Chan , Dr. A. Sundaram , Dr. P. Ghosh , Dr. S. Katare, Dr. A. Sirkar, Dr. P. Patkar, Dr. S-H. Hsu, Dr. S. Syal, Dr. B. K. Balachandra, Z. Zhang, Y. Luo, Prof. B. Srinivasan, Prof. R. Rengaswamy, Prof. R. Srinivasan, Dr. S. Viswanathan, Dr. S. H. Rich, Dr. S. Kavuri, Dr. Rama Vaidyanathan, Dr. Ramesh Vaidyanathan, Dr. C. Zhao (Bayer), Dr. A. Jain, Dr. S-H. Hsu, Dr. L. Hailemariam, Dr. P. Akkisetty, Dr. P. Sureshbabu, Dr. I. Hamdan, Dr. A. Shukla, Dr. A. Giridhar, Dr. G. Joglekar

Funding Agencies

- NSF
- DOE
- Indiana 21st Century Research and Technology Fund
- GE, Bangalore, India
- Lubrizol
- Caterpillar
- ExxonMobil
- Equistar
- Cummins

Thank You for Your Attention!



Questions?