Cyberinfrastructure-enabled Molecular Products Design and Engineering: Challenges and Opportunities

Venkat Venkatasubramanian
Laboratory for Intelligent Process Systems
Purdue University
West Lafayette, IN, USA

Outline

- **Molecular Products Design**
  - Data, Information and Knowledge Modeling Challenges
- **Cyberinfrastructure**
  - Ontological Informatics
- **Industrial Case Studies**
  - Lubrizol: Fuel Additives Design
  - Caterpillar: Rubber Products Design
  - ExxonMobil: Catalyst Design
  - Eli Lilly: Seromycin Formulation for MDR-TB
- **Summary**

Acknowledgements

Materials Design

- **Prof. J. F. Caruthers and Prof. W. N. Delgass**
- Dr. K. Chan (Bear Stearns), Dr. A. Sundaram (ExxonMobil), Dr. P. Ghosh (ExxonMobil), Dr. S. Katare (Dow), Dr. A. Sirkar (Intel), Dr. P. Patkar (ZS), Dr. S-H. Hsu (Purdue), S. Syal (Purdue), B. K. Balachandra (Purdue)

Funding Agencies

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

Acknowledgements

Pharmaceutical Engineering

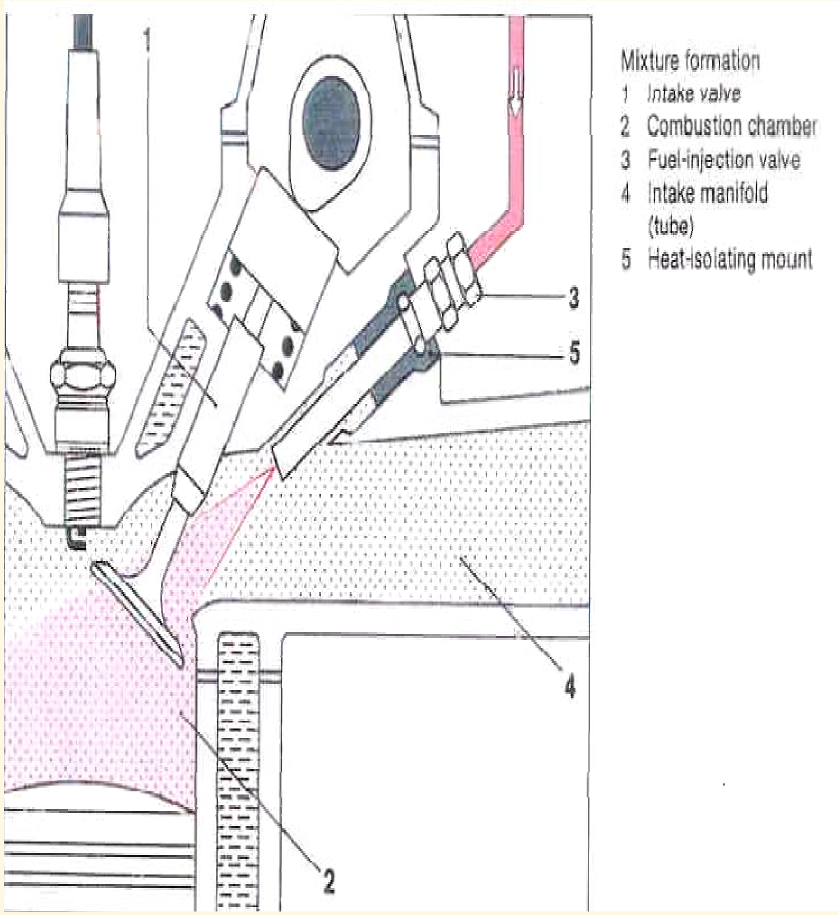
Dr. C. Zhao (Bayer)
Dr. A. Jain (United)
Dr. S-H. Hsu (Purdue)
L. Hailemariam (Purdue)
P. Akkisetty (Purdue)
P. Sureshbabu (Purdue)
I. Hamdan (Purdue)
M. Chaffee (Purdue)
A. Shukla (Purdue)

Dr. Prabir Basu
Dr. Girish Joglekar
Prof. Ken Morris
Prof. G. V. Reklaitis
Information Technology@Purdue

Funding Agencies:

NSF ERC
Indiana 21st Century R&T Fund
Eli Lilly
Pfizer
Abbott

LUBRIZOL: Fuel Additive Design



Intake Valve and Manifold

- **EPA requirement:** Minimize intake-valve deposits (IVD)
- **Approach:** Fuel Additives
- **Performance measure**
  - BMW Test for IVD
  - Stipulated to be less than 100 mg over a 10,000 mile road test
- **Expensive and time-consuming testing**
  - Around $8000 for a single datum
- **Problem:** ***Design fuel-additives that meet desired IVD performance levels***

CATERPILLAR

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

Multi-Scale Modeling

Quantum Mechanics

↓

Kinetics

↓

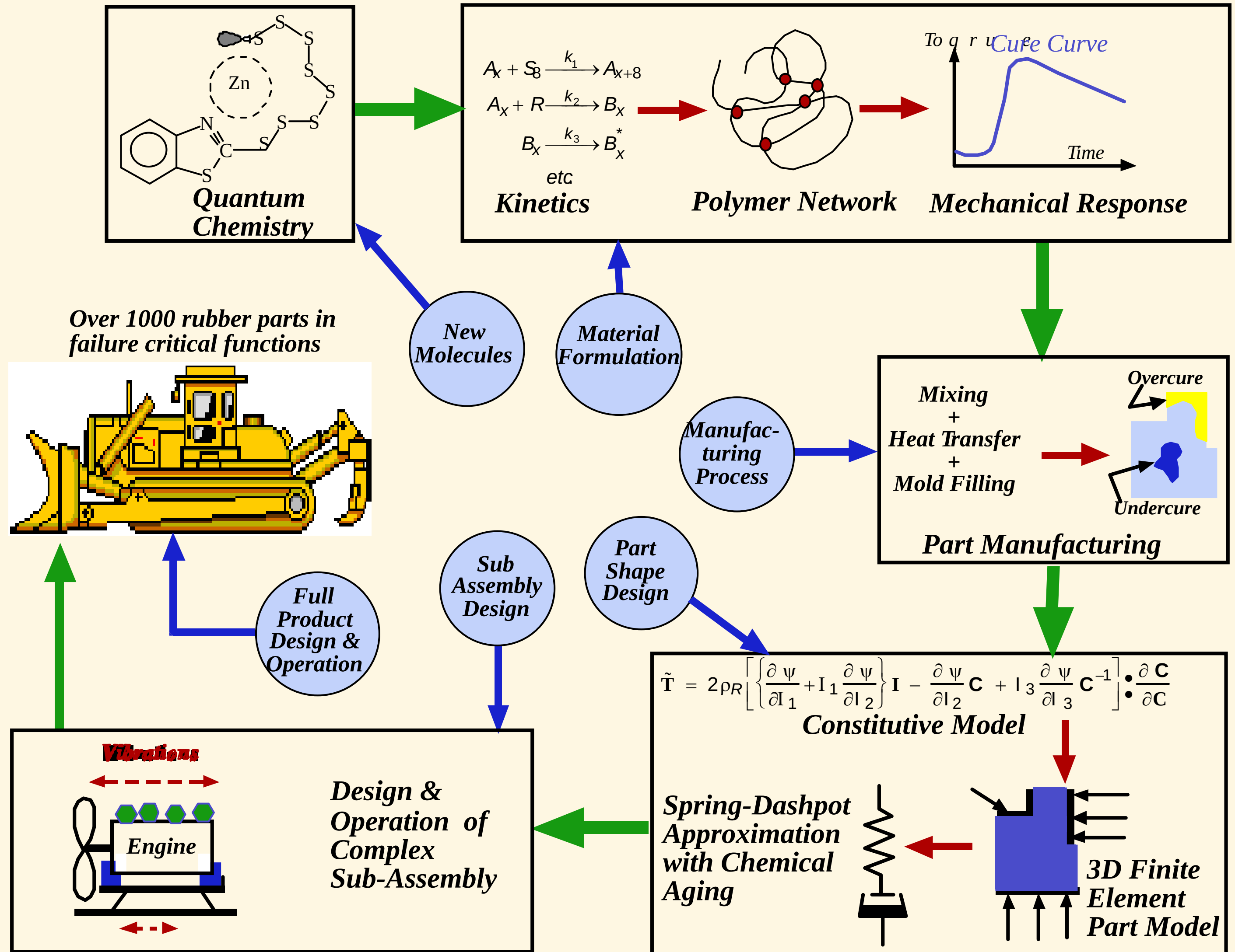
Heat Transfer FEA

↓

Mechanics FEA

↓

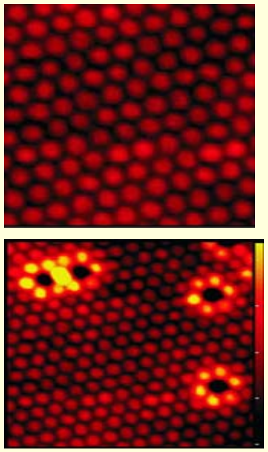
Part Design CAD /CAM



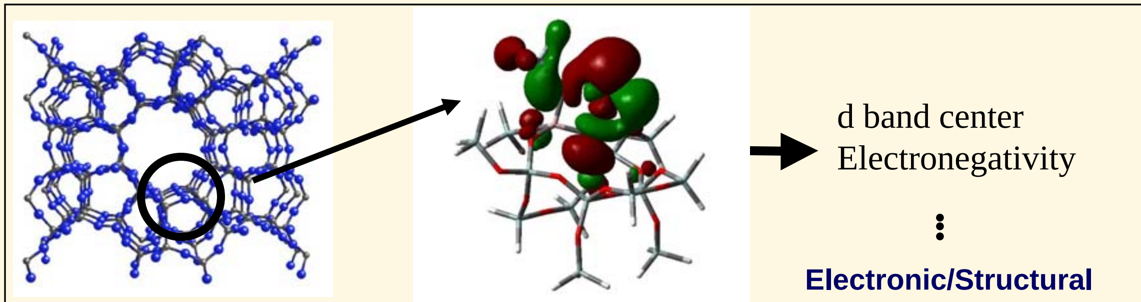
Venkat Venkatasubramanian, Plenary Lecture,
Foundations of Computer-Aided Process Design, Princeton, July 2004.

PURDUE UNIVERSITY

ExxonMobil's Grand Challenge: Catalyst Design via Combinatorial Chemistry



Catalyst Structures



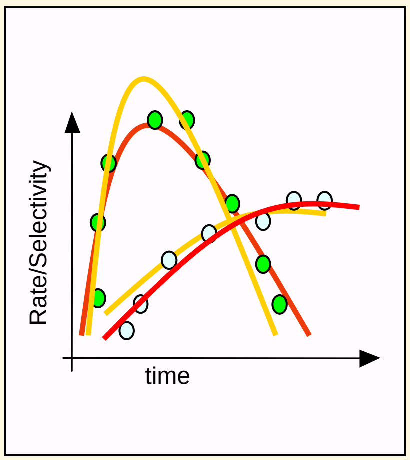
Quantum calculations

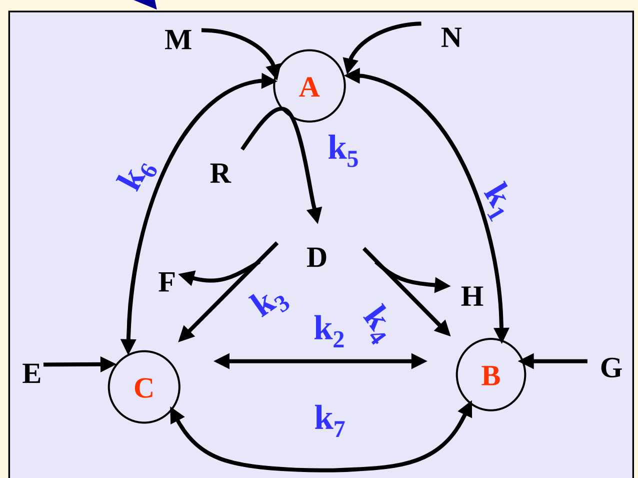
Fundamental Chemistry

Synthesis Characterization

Reaction kinetics

Performance measures

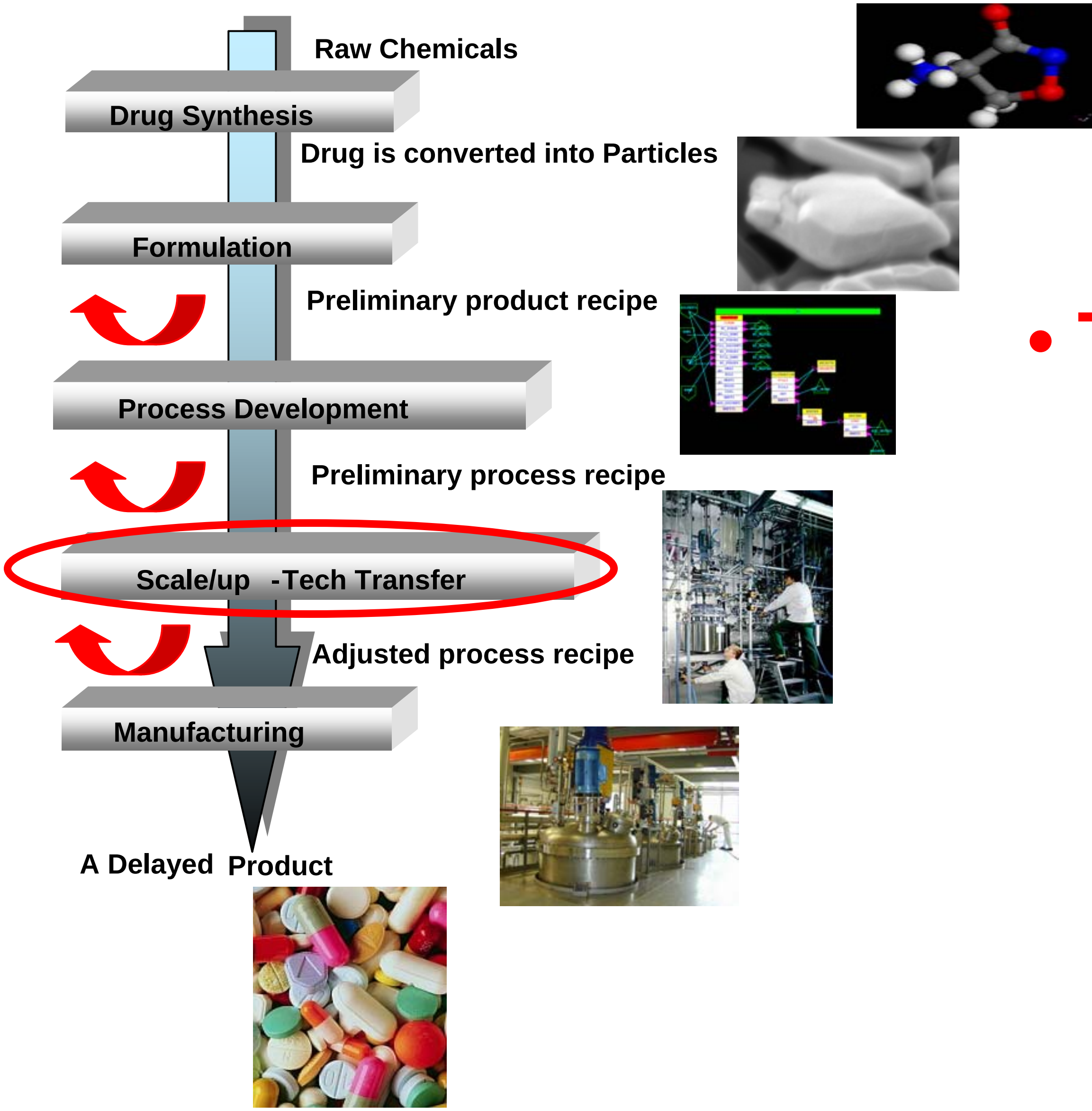




Reaction network

Venkat Venkatasubramanian, Plenary
Foundations of Computer-Aided Process Design, Princeton, July 2004.

Pharmaceutical Product Development and Engineering



- Two Products
  - Drugs
  - Documents

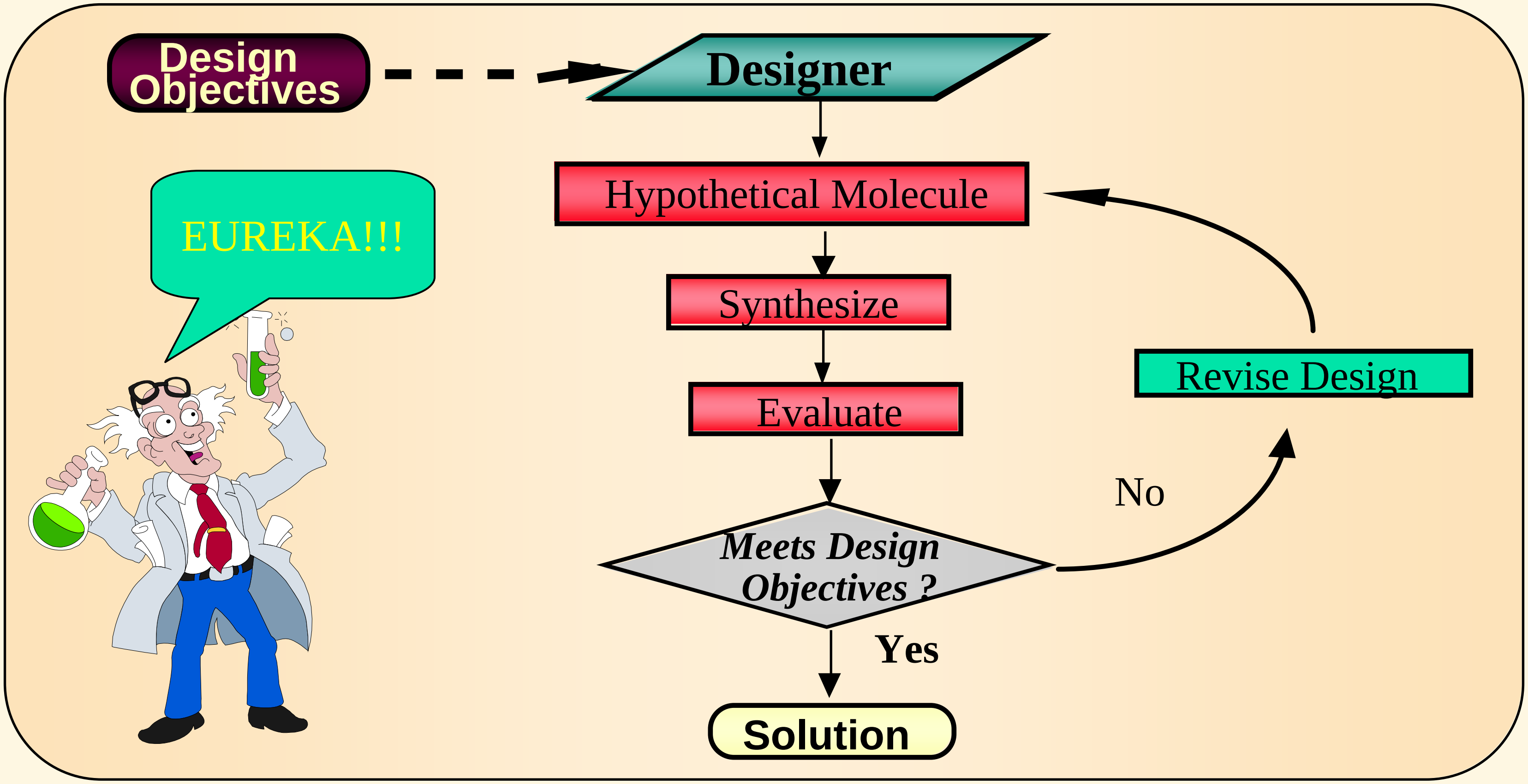
Molecular Products Design

- **Given a set of desired performance specifications, how does one rationally and efficiently identify the optimal product structures and formulations?**
- **$200+ Billion/yr industry: Major Opportunity for ChEs**
- **Areas of Application**
  - **Engineering Materials**
    - **Fuel and Oil Additives**
    - **Polymer Composites**
    - **Rubber Compounds**
    - **Catalysts**
    - **Solvents, Paints and Varnishes…..**
  - **Pharmaceuticals**
    - **Drug Design**
  - **Agricultural Chemicals**
    - **Pesticides**
    - **Insecticides**

Different Products, But Common Challenges

- Lots of Data
- Uncertain/Noisy Data
- Complex chemistry, lots of species
- Nonlinear systems and processes
- Incomplete, Uncertain, Mechanisms
- Combinatorially large search spaces
- Need Multi-scale Models
- Hundreds of Differential and Algebraic Equations to formulate and solve
- Laborious and time consuming model development process
- Limited human expertise

Traditional Approach



Drawback: Protracted and Expensive Design Cycle

Need a Rational, Automated Approach: Discovery Informatics

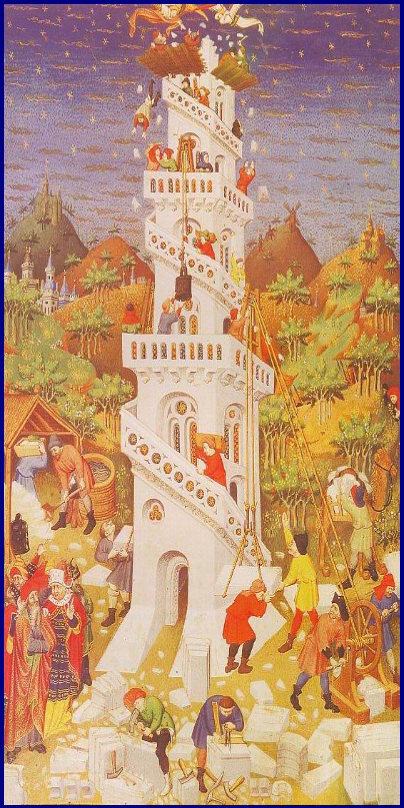
Need a New Paradigm

- **Cyberinfrastructure Methods and Tools**

- **Ontology-based Discovery Informatics**

Data, Information, and Knowledge

- Data
  - **What's** going on?
    - Raw, does not inform much
- Information
  - **How** are the variables related?
    - Correlations and relationships
- Knowledge
  - **Why** are they related?
    - Develop Mechanistic understanding
    - First-Principles Based Math Models
    - Heuristics



Tower of Babel of Data, Information and Models

What is an Ontology?

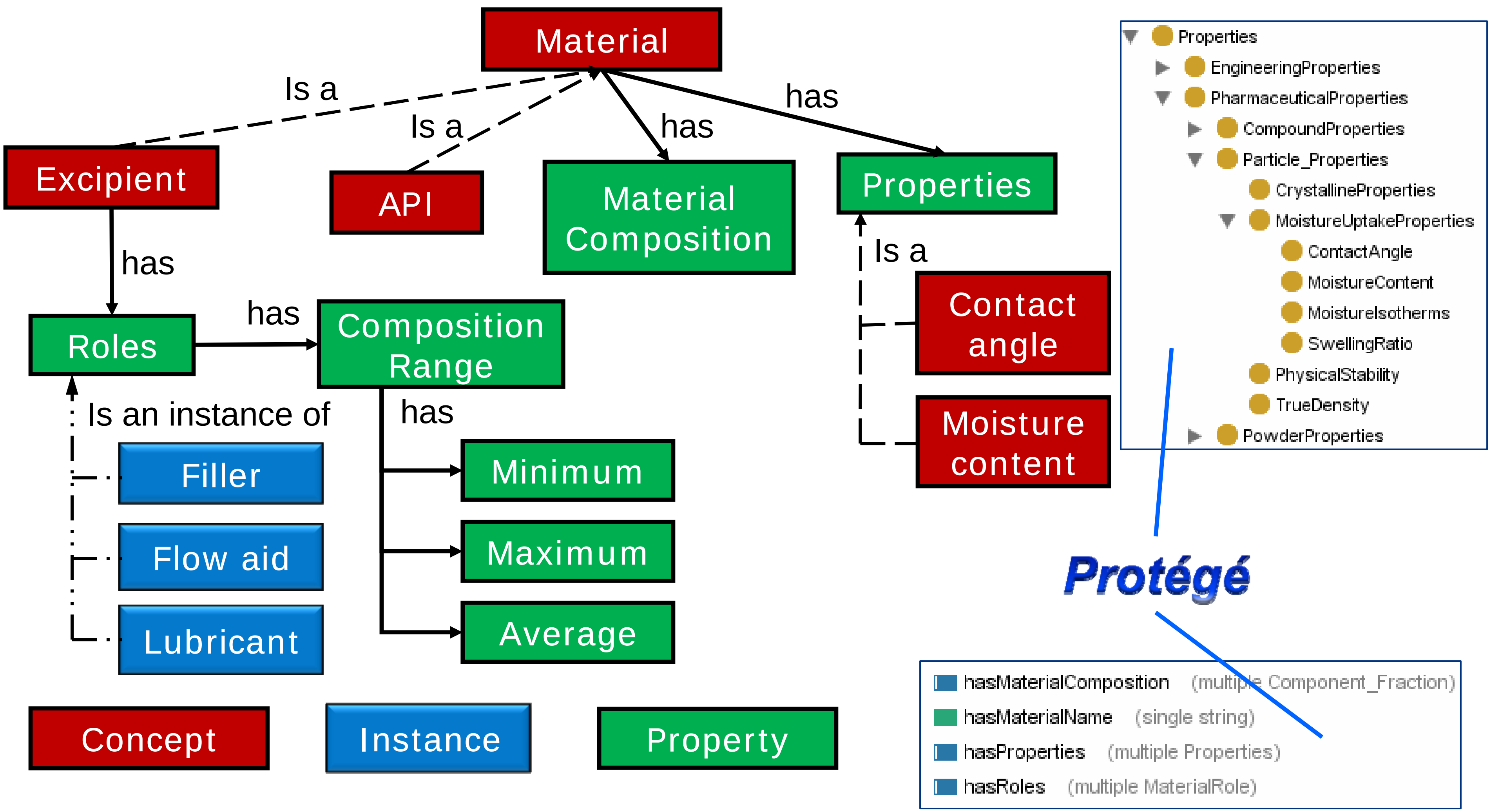
- Originated in Philosophy
  - Study of Existence
  - Not to be confused with Epistemology: Theory of Knowledge and Knowing
- Computer Science and Artificial Intelligence
  - A formal, explicit specification of a shared conceptualization
  - Knowledge representation in AI
    - Logic, Semantic Networks, Frames, Objects, **Ontology**
  - Web-driven development
- Semantic Richness: Description Logic (DL) provides foundation for formal reasoning
- Easily create, share and reuse knowledge
- Semantic Search

Based on Set Theory and First-order Logic

An ontology is a 5-tuple $O := (C, R, H^c, rel, A^o)$ consisting of

- Two disjoint sets C and R whose elements are called concept identifiers and relation identifiers, respectively
- A concept hierarchy H^C: H^C is a directed, transitive relation $H^C \subseteq C \times C$ which is called concept hierarchy or taxonomy. $H^C(C_1, C_2)$ means that C_1 is a subconcept of C_2.
- A function $rel : R \rightarrow C \times C$ that relates concepts non-taxonomically. The function dom: $R \rightarrow C$ with $dom(R) := \Pi_1(rel(R))$ gives the domain of R, and range: $R \rightarrow C$ with $range(R) := \Pi_2(rel(R))$ give its range. For $rel(R) = (C_1, C_2)$ we also write $R(C_1, C_2)$.
- A set of axioms A^o, expressed in an appropriate logic language, e.g. first order logic

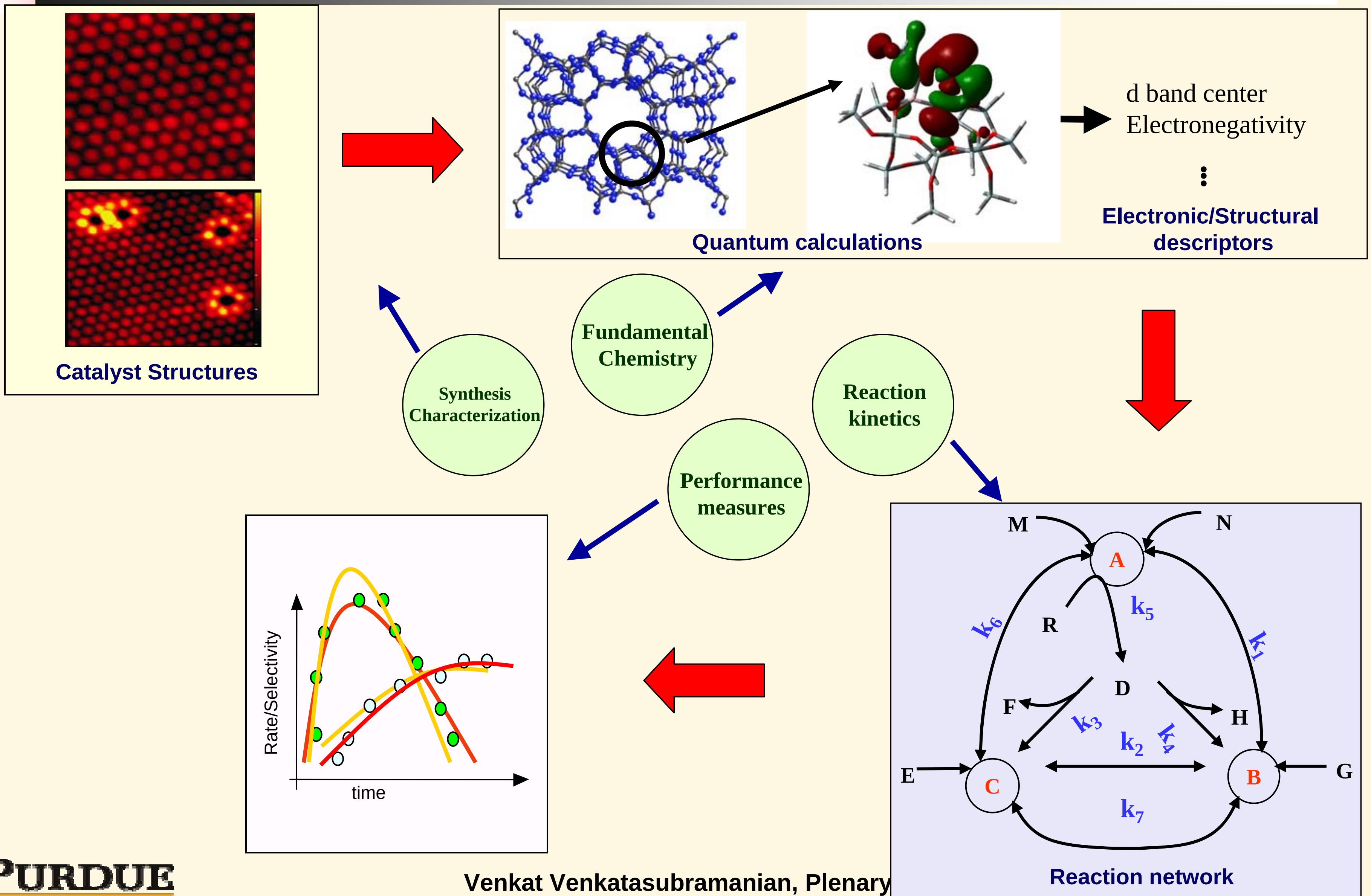
Ontology: A Simple Example



Web Ontology Language: OWL

ENGINEERING RESEARCH CENTER FOR
STRUCTURED ORGANIC PARTICULATE SYSTEMS
RUTGERS UNIVERSITY
PURDUE UNIVERSITY
NEW JERSEY INSTITUTE OF TECHNOLOGY
UNIVERSITY OF PUERTO RICO AT MAYAGÜEZ

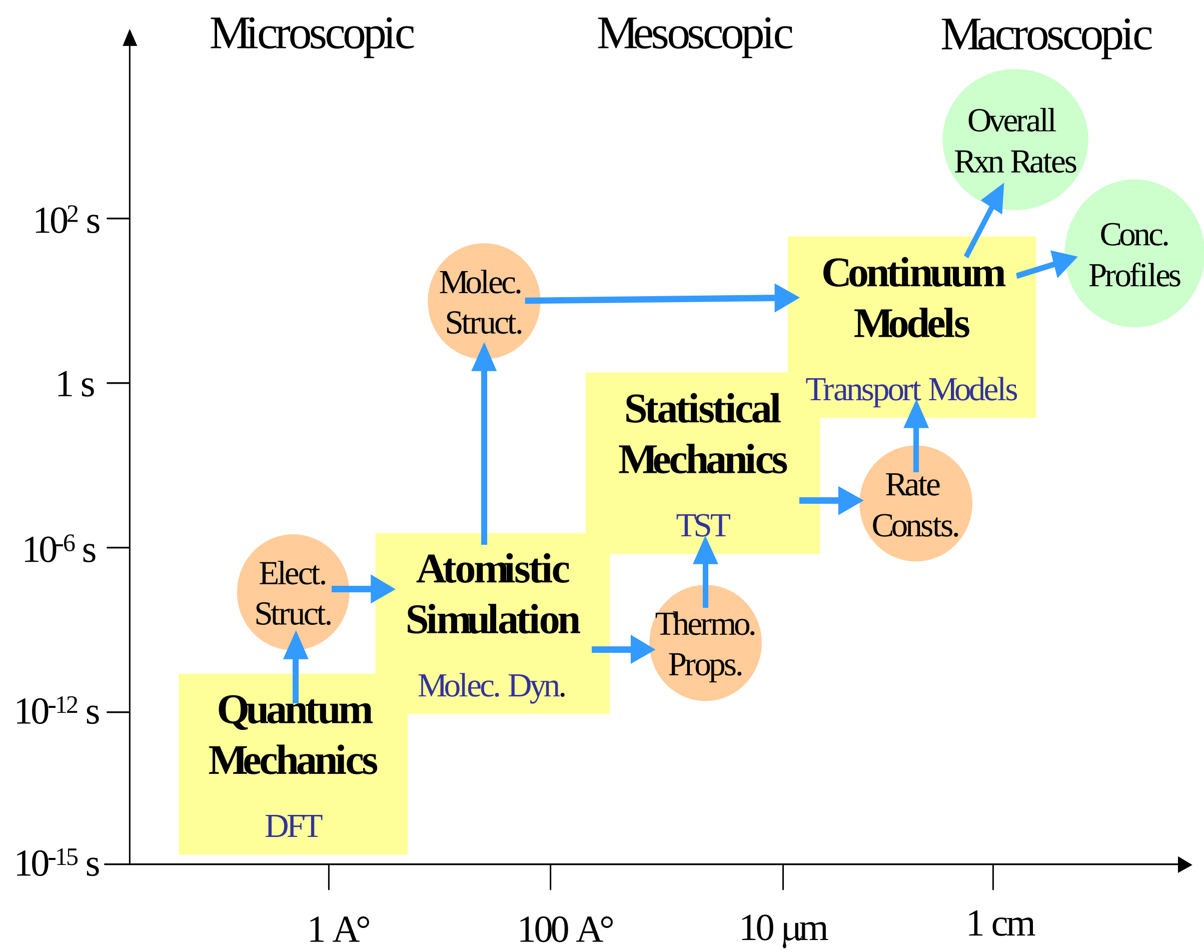
Catalyst Design Problem – The Grand Challenge



PURDUE UNIVERSITY

Venkat Venkatasubramanian, Plenary
Foundations of Computer-Aided Process Design, Princeton, July 2004.

Modeling at different Length and Time Scales



Venkat Venkatasubramanian, Plenary Lecture,
Foundations of Computer-Aided Process Design, Princeton, July 2004.

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

Previous Work

Microkinetic Analysis (Dumesic, 1993)

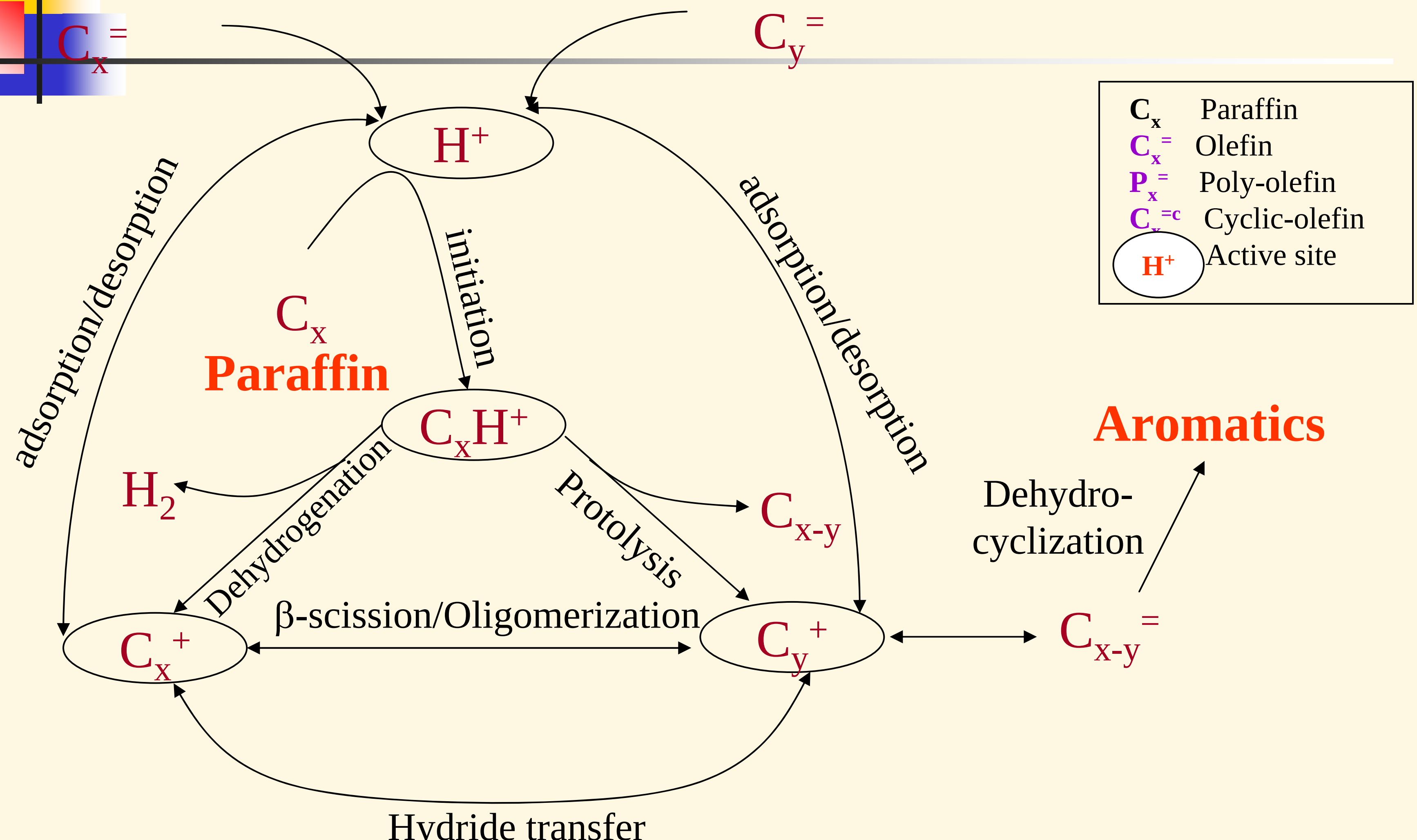
- First unified view of reaction engineering on catalytic surfaces
- Incomplete forward problem
- Lacks clear view of the design perspective (inverse problem)

Empirical Catalyst Design (Baerns, 2000)

- First attempt to "design" catalysts
- Completely based on "guided" experiments – time consuming and expensive
- No fundamental understanding of the system (forward problem)

Ertl, Rasmussen, Lauterbach: Surface Science Studies
Steve Jaffe: Composition based modeling of large systems
Jens Norskov: Computation based studies
Symyx, Novodynamics, Lauterbach: Combinatorial HTE

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

Ontological Informatics: Modeling Super Highway

Data

Chemistry Rules
Reactions
Grouping

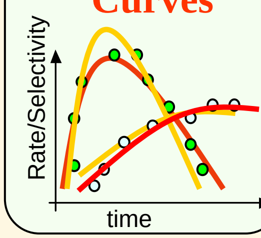
Chemistry Compiler
Reaction Description Language Plus (RDL+)

Equation Generator
Automatic generation of differential algebraic equations (DAEs)

Parameter Optimizer
Solution of DAEs
Least squares
Features

Advanced Statistical Analyzer
Sensitivity Analysis
Uncertainty Analysis
Error Propagation

Performance Curves

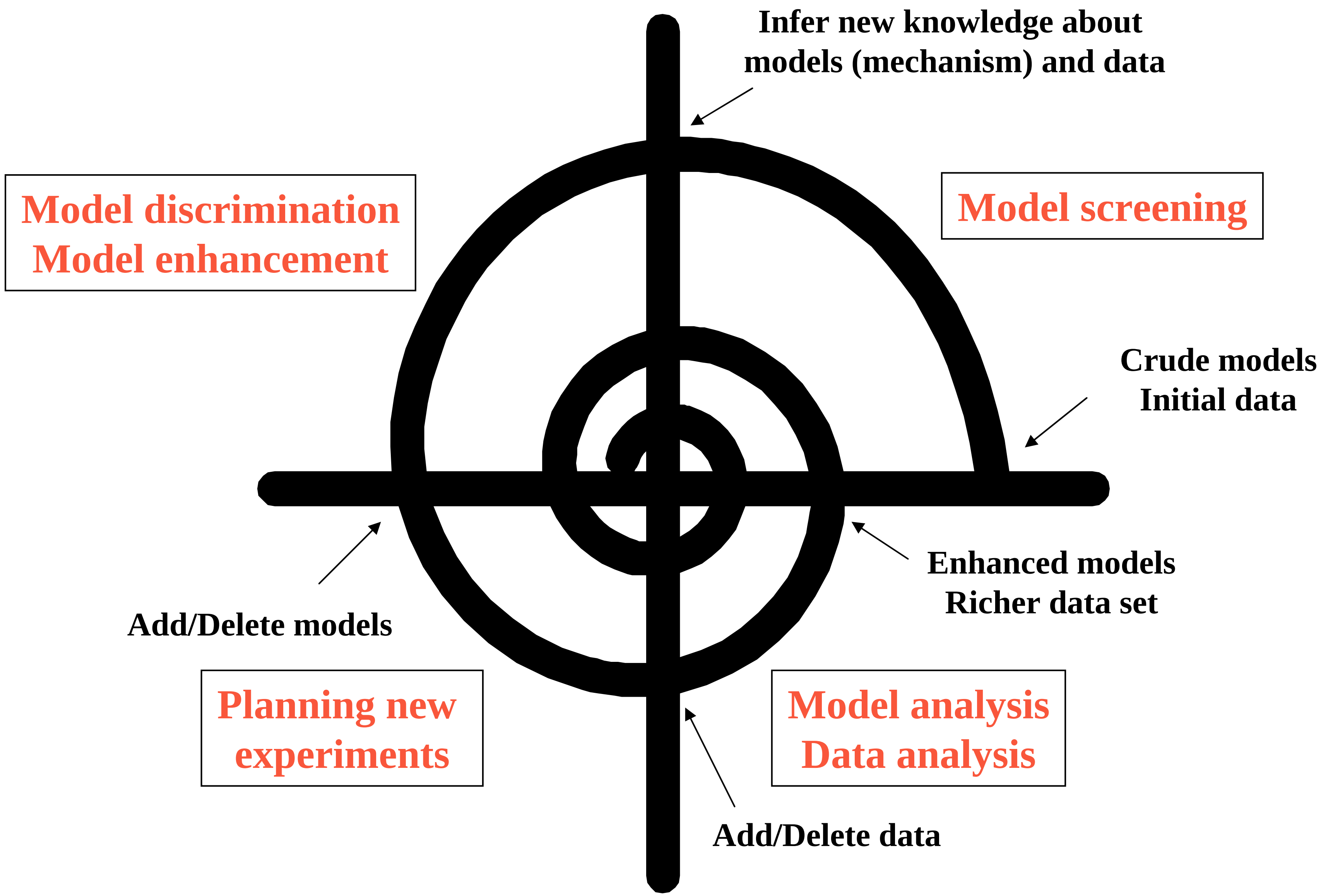


- DAE System: 30 to 1000 species and 10 to 50 parameters
- Parameter search space is highly nonlinear with numerous local minima
  GA based pseudo global parameter estimator
  Fast and efficient coverage of the parameter space
- Optimization: Traditional least-squares and features-based objectives

J.M. Caruthers, J.A. Lauterbach, K.T. Thomson, V. Venkatasubramanian, C.M. Snively, A. Bhan, S. Katare and G. Oskarsdottir, "Catalyst Design: Knowledge Extraction from High Throughput Experimentation", In "Understanding Catalysis from a Fundamental Perspective: Past, Present, and Future", A. Bell, M. Che and W.N. Delgass, Eds., Invited Paper for the 40th Anniversary Issue of the *Journal of Catalysis*, 2003.

Katare, S., Caruthers, J. M., Delgass, W. N., Venkatasubramanian, V., "An Intelligent System for Reaction Kinetic Modeling and Catalyst Design", *Ind. Eng. Chem. Res. and Dev.*, 2004.

Model Refinement Procedure

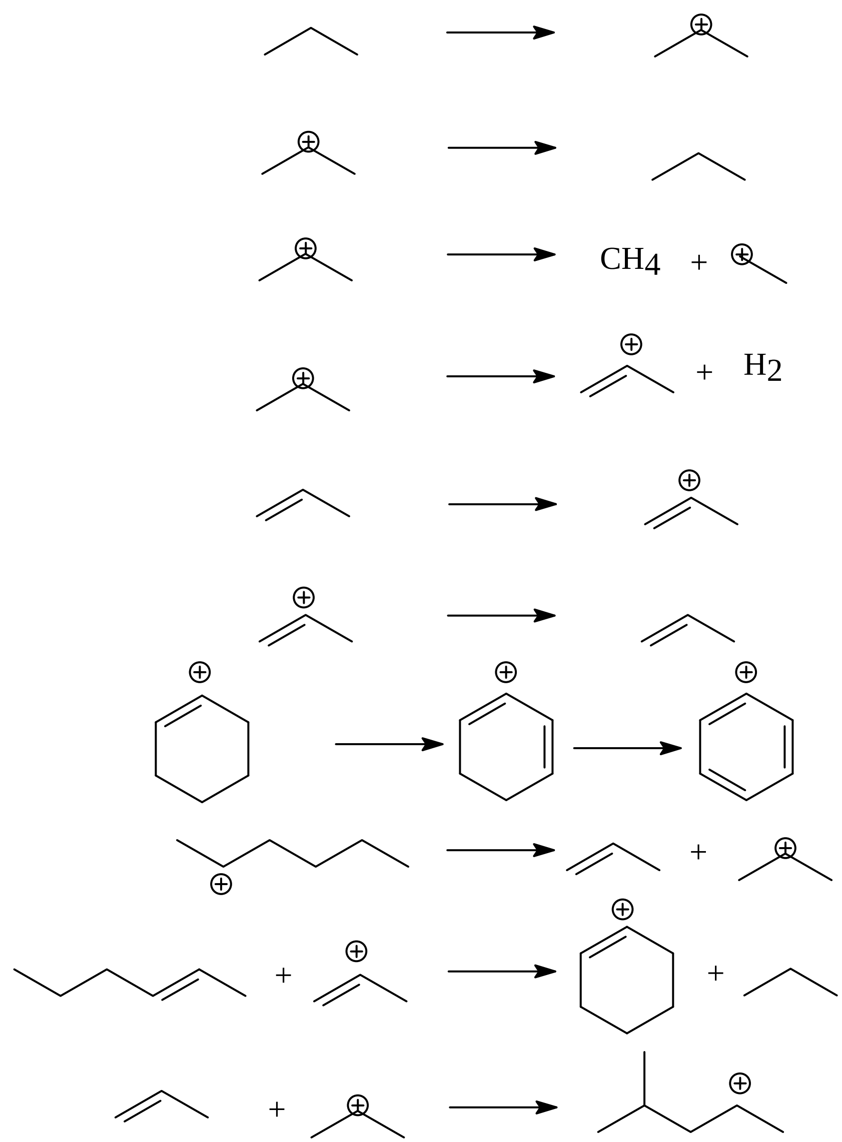


Chemistry Rules for Propane Aromatization on HZSM-5

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



Modeling Super Highway: 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

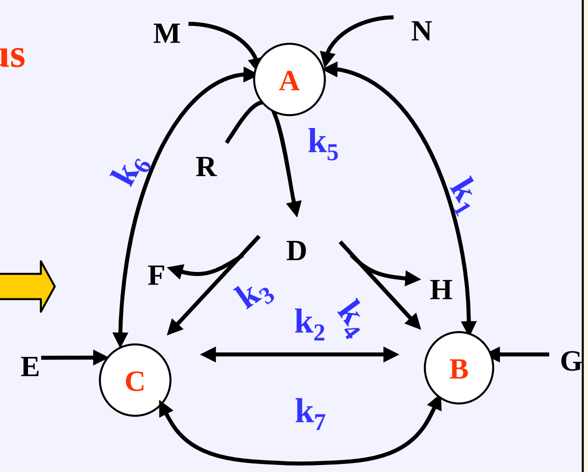
⋮

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

$$dC_A / dt = -k_1 C_A$$

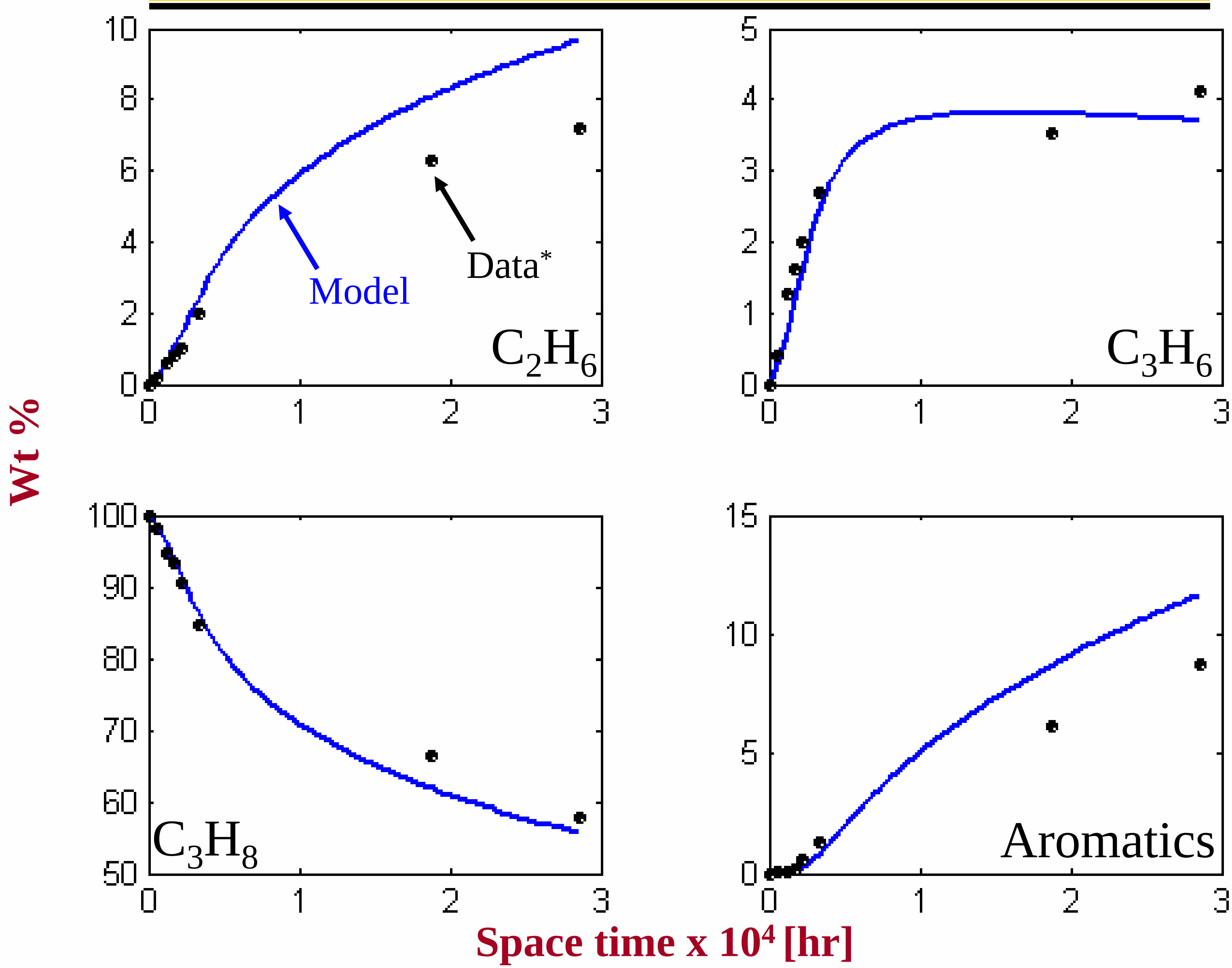
$$dC_B / dt = k_1 C_A + k_4 D - k_5 B$$

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

⋮

100's of DAE's

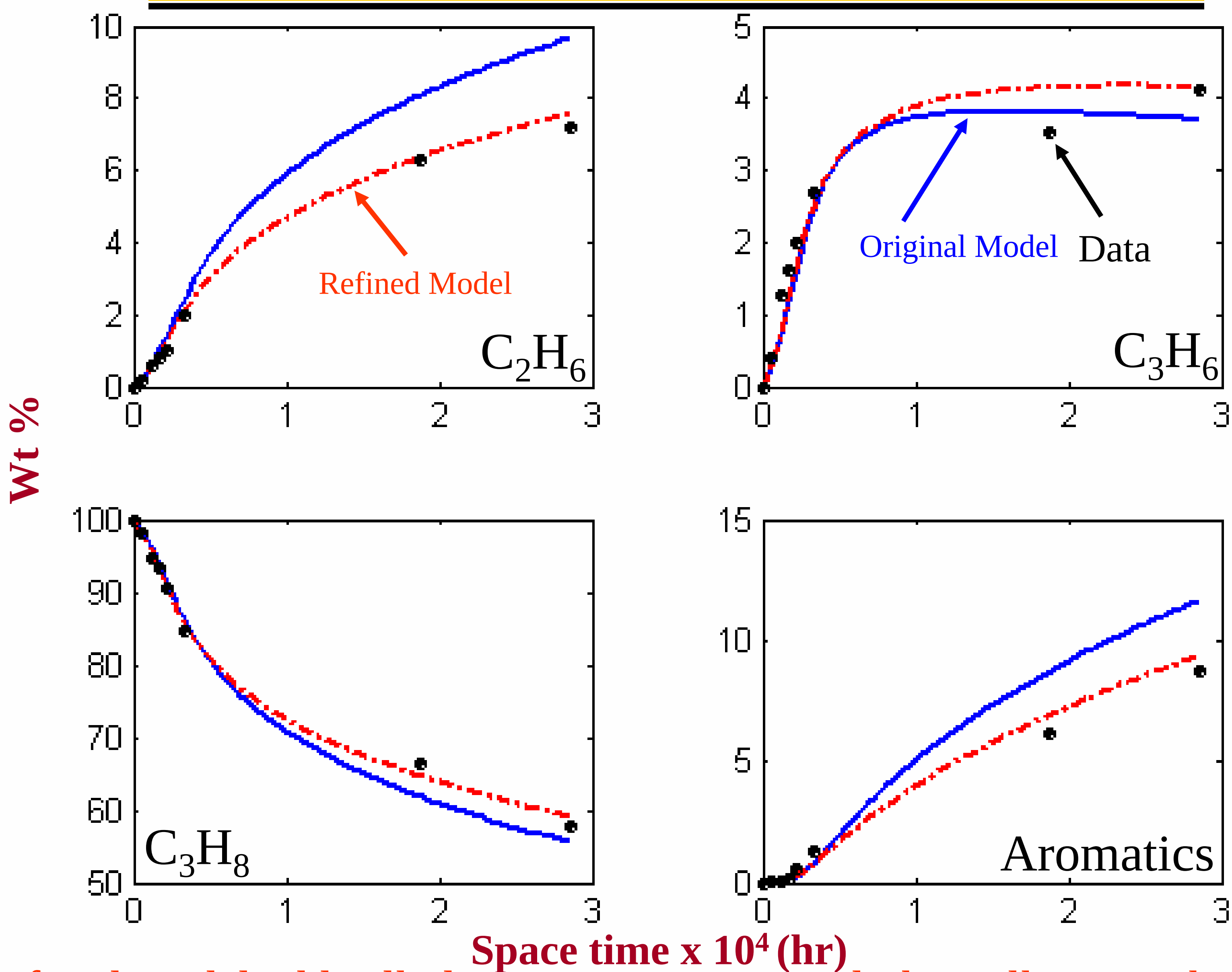
Results



Over prediction of C_2s and under prediction of aromatics

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

Results – Model Refinement



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

GA Based Pseudo Global Parameter Estimator

Performance comparison on test problems – worst case: 3 ODEs/5 parameters

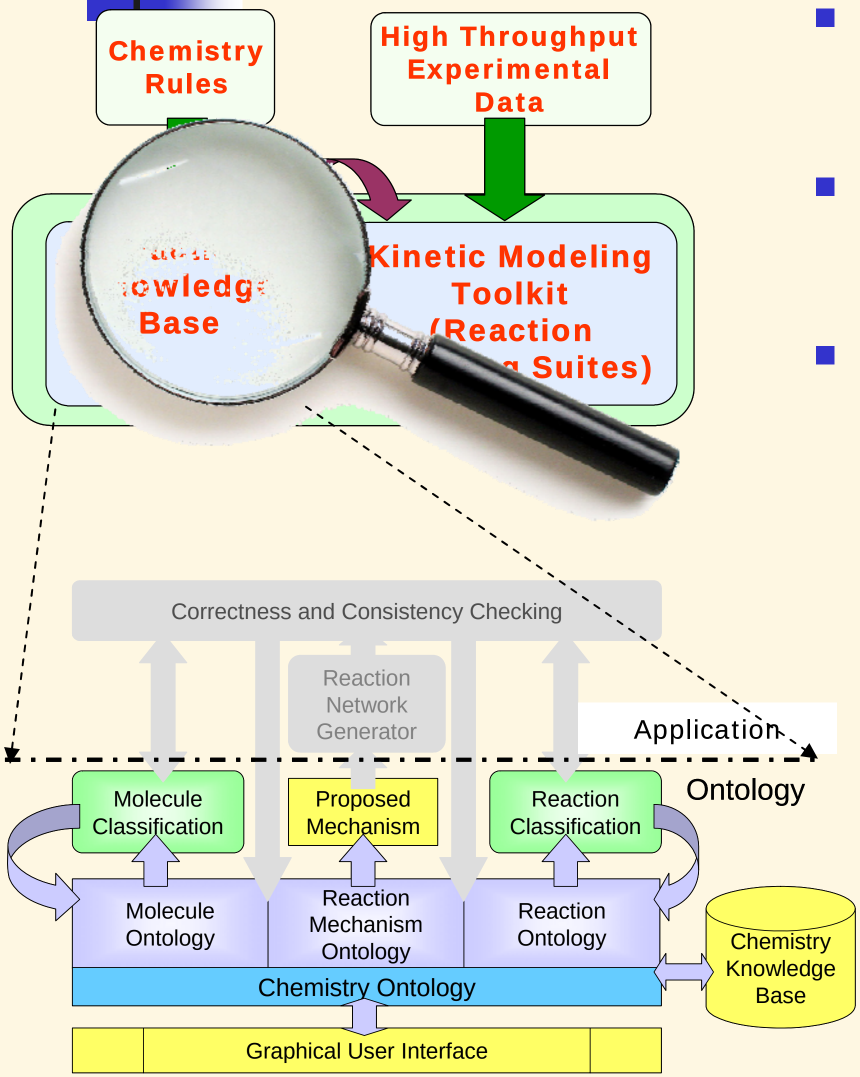
S.no.	Name	Esposito & Floudas (2000) Global optimizer			GA based procedure			
		υ	Objective	CPU s reported min max	υ	Objective	Time taken (CPU s)	Scaled CPU s Time taken *671/211 **(% Saving)**
1	First order irreversible chain reaction	5.0035 1.0000	1.18584×10^{-6}	2.92 20.05	5.0122 0.9976	9.25418×10^{-6}	0.24	0.76 **(74)**
2a	First order reversible chain reaction	4.0000 2.0000 40.013 20.007	1.8897×10^{-7}	568.44 6164.3	3.9813 1.9759 39.787 19.887	8.1313×10^{-6}	0.45	1.43 **(100)**
2b	Same as 2a but with error in data	4.021 2.052 39.45 19.62	1.586×10^{-3}	272.91 1899.82	4.001 2.027 39.22 19.50	1.589×10^{-3}	0.5	1.59 **(99)**
3	Catalytic cracking of gas oil	12.214 7.9798 2.2216	2.65567×10^{-3}	79.77 1185	12.246 7.9614 2.2351	2.68017×10^{-3}	0.38	1.21 **(98)**
4	Bellman's problem	4.5704×10^{-6} 2.7845×10^{-4}	22.03094	36.16 12222	4.5815×10^{-6} 2.7899×10^{-4}	22.2885	3.54	11.26 **(69)**
5	Methanol-to-hydrocarbon process	5.1981 1.2112 0 0 0	0.10652	1361.8 19125	5.2212 1.2320 1.6363×10^{-31} 4.0059×10^{-12} 0.004665	0.10586	2.08	6.61 **(100)**
6	Lotka-Volterra Problem	3.2434 0.9209	1.24924×10^{-3}	367.26 9689.67	3.1434 0.9583	2.39784×10^{-3}	0.56	1.78 **(100)**

Our zeolite model

31 ODEs, 29 Algebraic equations
13 parameters
9 concentration curves

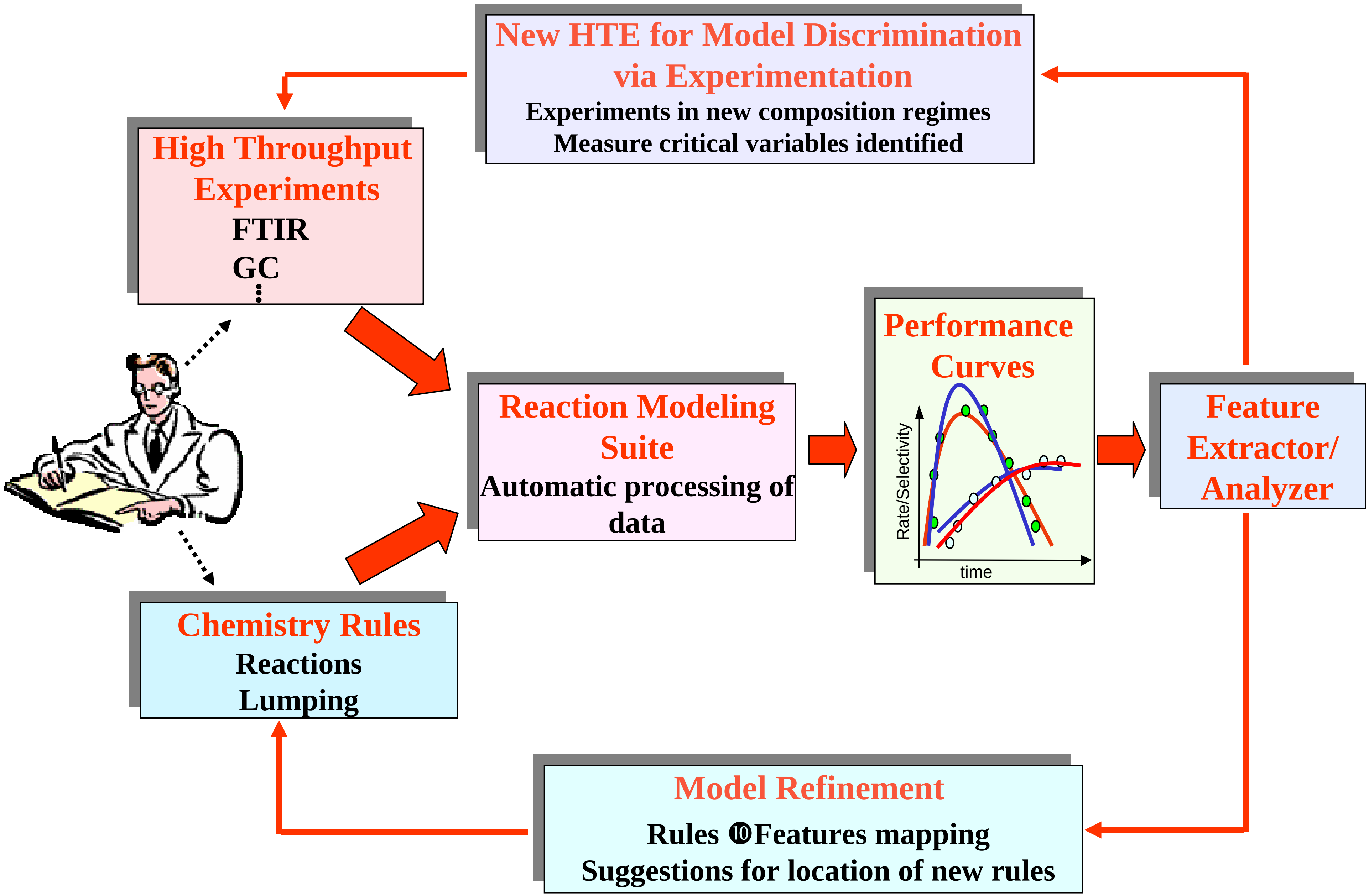
⇨ **Reasonable comparison to experimental data** ⇨ **2 hours on a Sun 400 MHz solaris machine**

Ontological Informatics for Catalyst Design: Summary

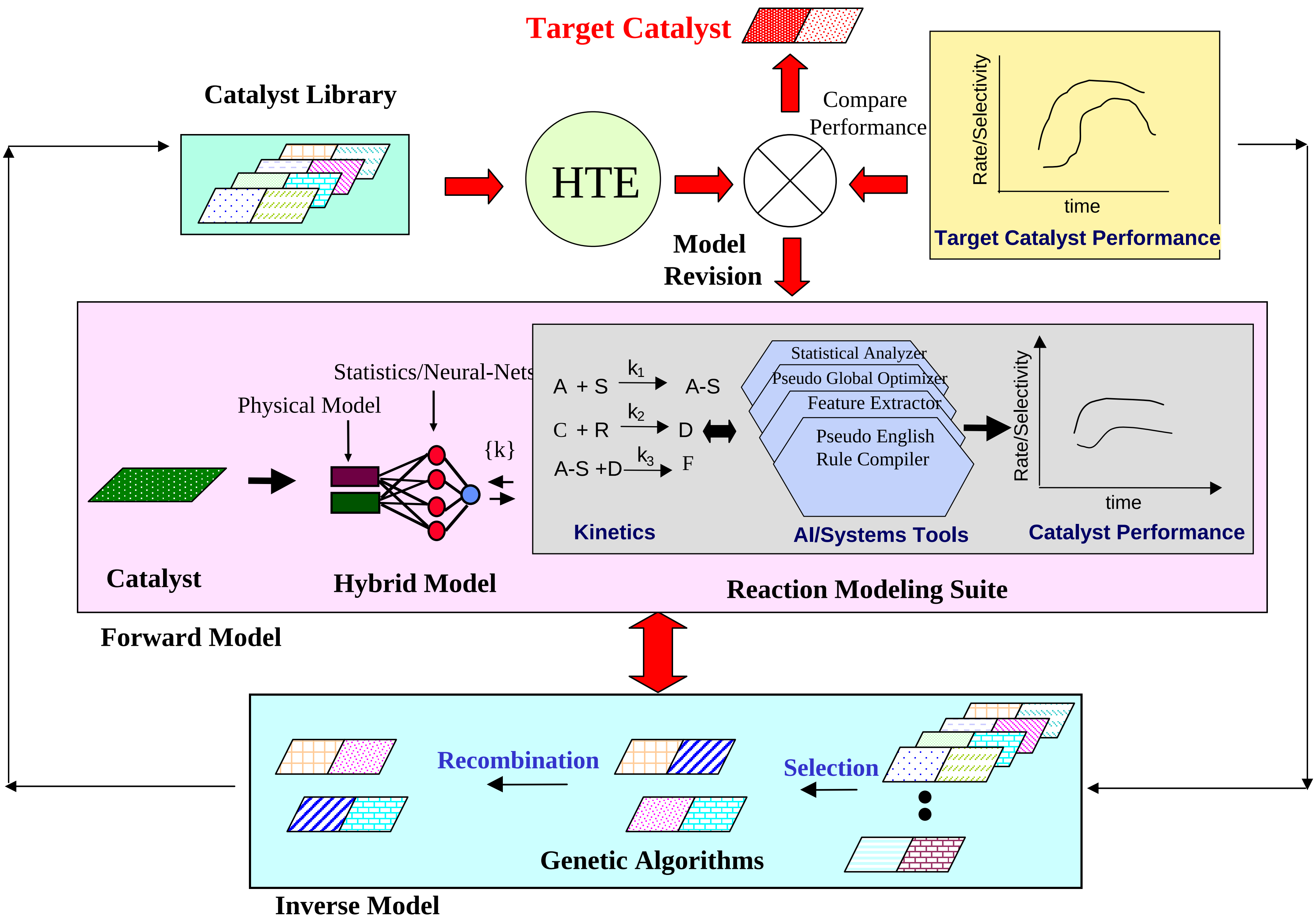


- Represents reaction chemistries explicitly
- Customizable for different catalyst chemistries
- Results
  - Real-time evaluation of 100s of reaction pathways and 1000s of DAEs
  - Saved months of model development effort
  - Improved mechanistic understanding and first principles models

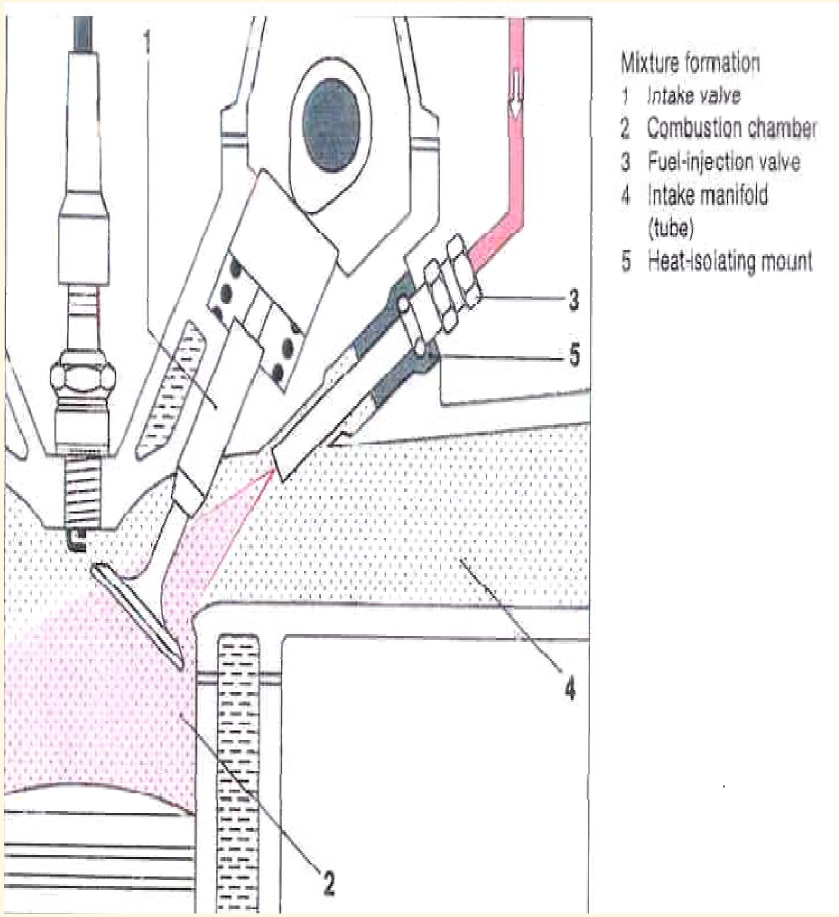
Proposed Framework for Knowledge Extraction from HTE – Kinetic Modeling



Catalyst Design Challenge



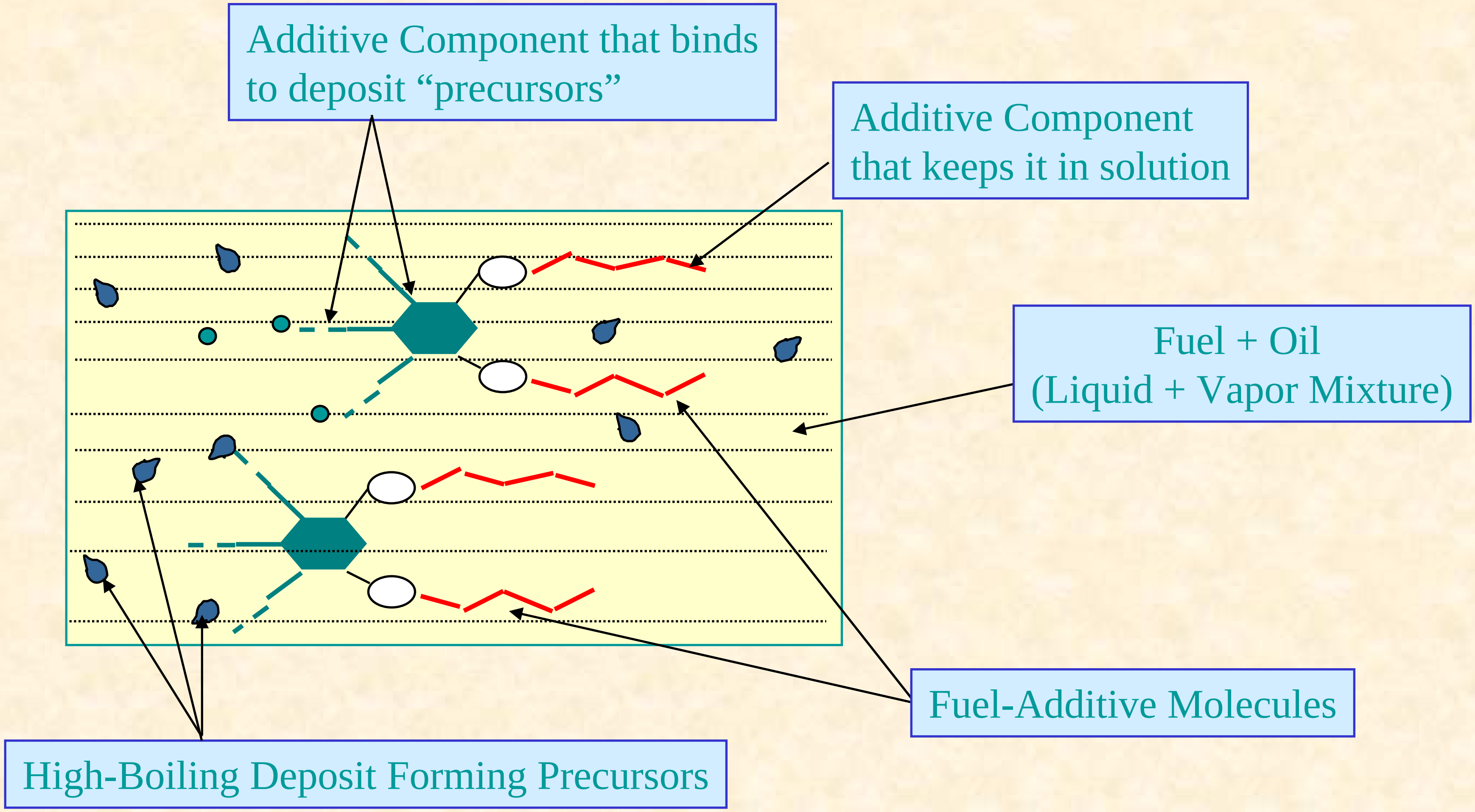
LUBRIZOL: Fuel Additive Design



Intake Valve and Manifold

- **EPA requirement:** Minimize intake-valve deposits (IVD)
- **Approach:** Fuel Additives
- **Performance measure**
  - BMW Test for IVD
  - Stipulated to be less than 100 mg over a 10,000 mile road test
- **Expensive and time-consuming testing**
  - Around $8000 for a single datum
- **Problem:** ***Design fuel-additives that meet desired IVD performance levels***

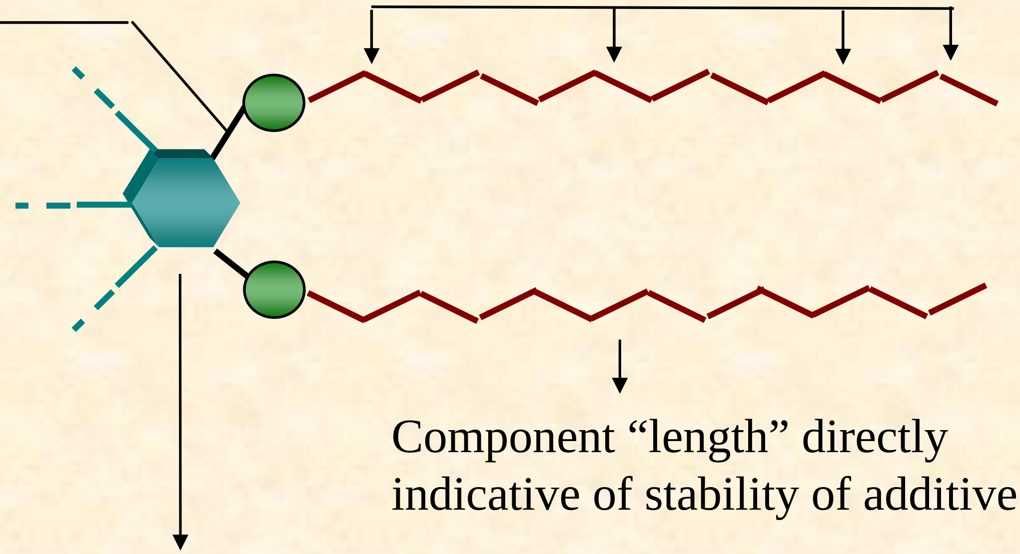
Fuel-Additives : A Functional Description



First-Principles Model for Additive Degradation

Breakage of this bond removes “dirt” carrying capacity totally

Breaking of these bonds control “length”

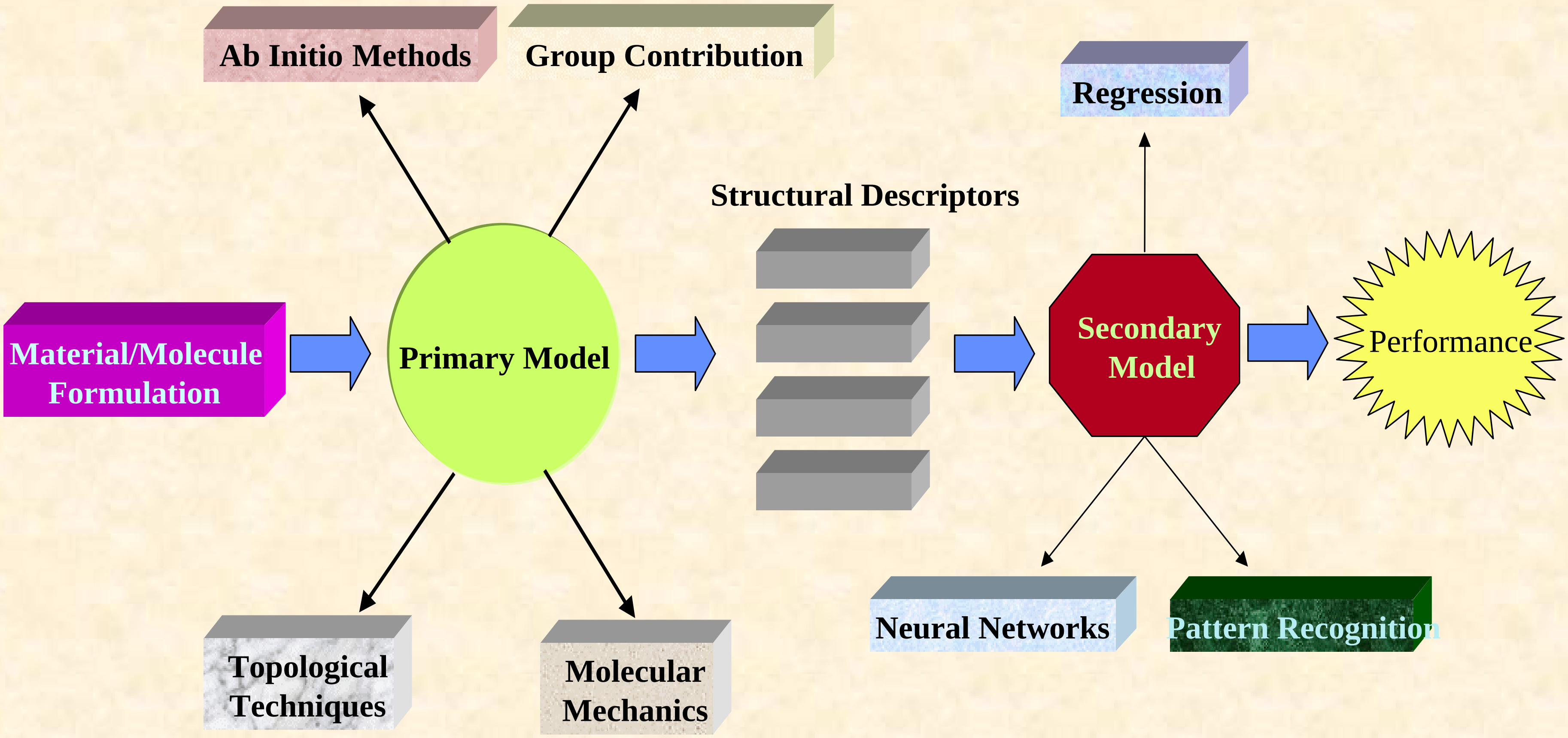


Component “length” directly indicative of stability of additive

Chemical nature of this component (polar/non-polar) controls “dirt” removing capacity

The first-principle model tracks the structural distribution of fuel-additive with time due to reactive degradation

General Framework of Predictor Models



Modeling Degradation

- Dynamic in nature
  - Modeled as first-order irreversible reactions for additive degradation
  - Rate constants in proportion to rates of pure thermal degradation

- Distribution of molecular species differing in structure
  - Identified by effective "length"
  - Population balance model tracks distribution with time

From Stability to IVD

- Define "**amount of active additive**"

  The fraction of the additive species that remains active (i.e. intact viable structure) in the fuel at any point of time

  - A dynamic quantity decreasing with time
  - Depends on additive distribution

- Solubility correlated to IVD via regression
  - Linear models
  - Neural-Network models
  - Input descriptors are the ***amounts of active additive*** at different times

Forward Problem Approaches

Data-driven/empirical models

Empirically speaking, a polynomial expansion fits my data............

Fundamental Physics/Chemistry

$$\frac{\partial(\rho v)}{\partial t} + v \bullet \nabla(\rho v) = -\nabla P + \nabla \bullet (\tau)$$

$$\frac{\partial \rho}{\partial t} = -\nabla \bullet (\rho v)$$

Expert Rules/ Heuristics

If A goes higher, B should go higher .....

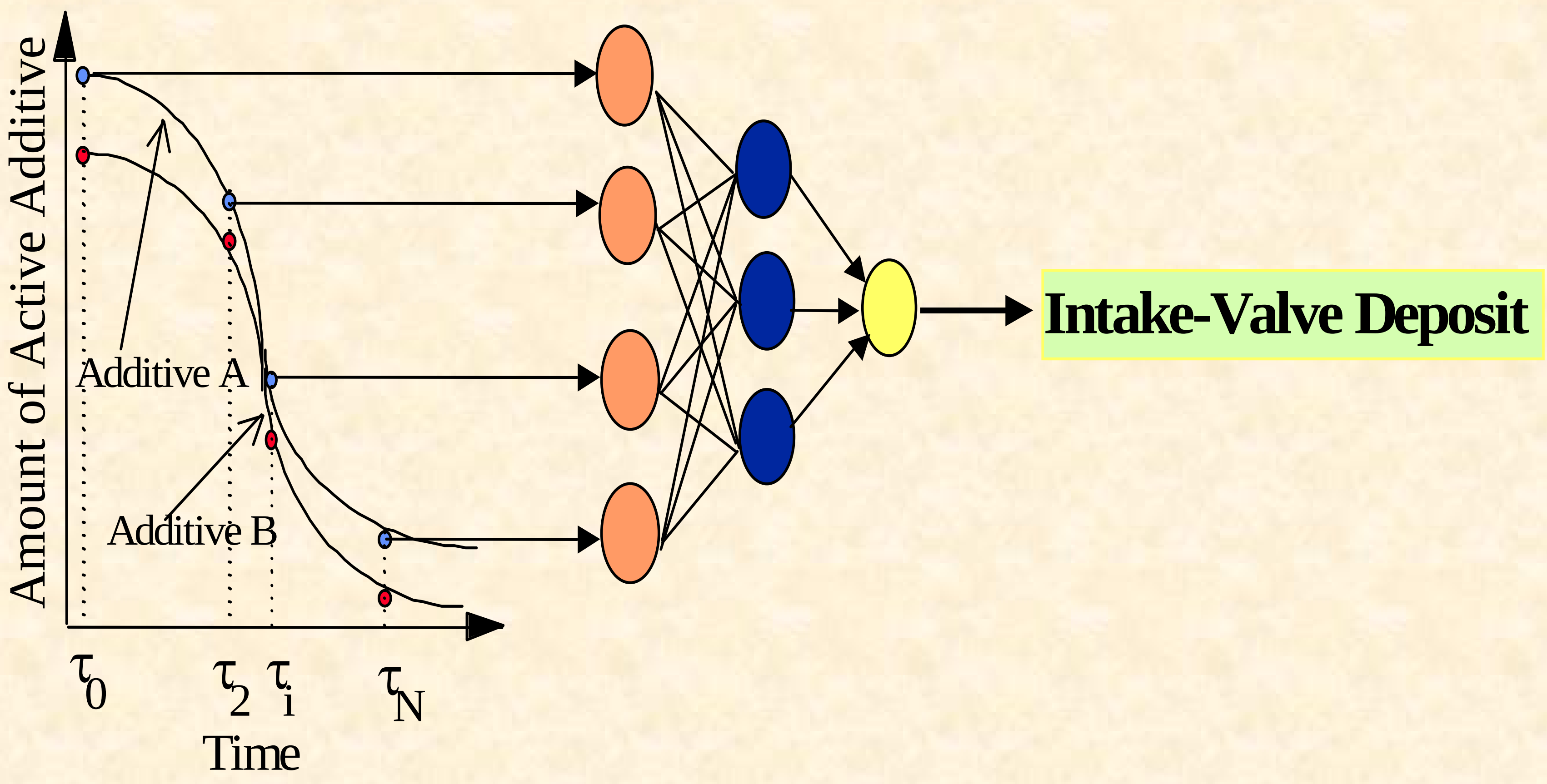
If C goes higher, D should go lower .....

How to integrate diverse scientific knowledge/information into a single unified knowledge architecture ?

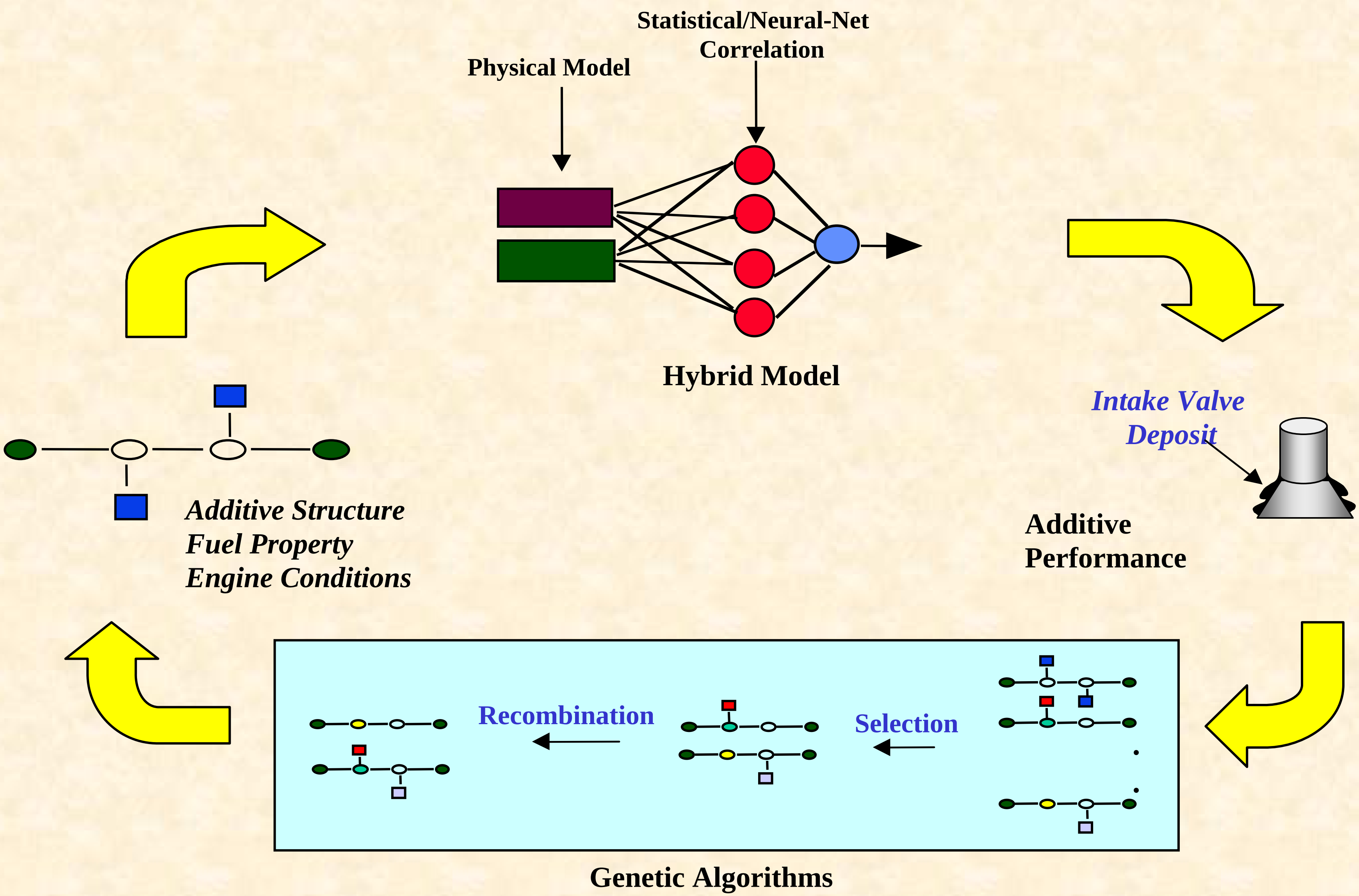
Modeling Philosophy

- **All models are wrong....**
  ... but some are useful!
  -- George Box
  (U. Wisconsin)
- **Hybrid Models**
  - **Combine First-principles and Data-driven models**

Hybrid Model for IVD Prediction



Computer-Aided Design of Fuel-Additives



Previous Approaches to Inverse Problem

- Previous Methodologies
  - **Random Search**
  - **Heuristic Enumeration**
  - **Math Programming**
  - **Knowledge-Based Systems**
  - **Graph Reconstruction**
- Disadvantages
  - **Combinatorial Complexity**
    - **Nonlinear Search Spaces**
    - **Local Minima Traps**
  - **Difficulties in Knowledge Acquisition**
  - **Difficulties in using high-level chemical/bio-chemical knowledge**

Overview of Genetic Algorithms

- Definition

  Genetic Algorithms are stochastic, evolutionary search procedures based on Darwinian model of natural selection

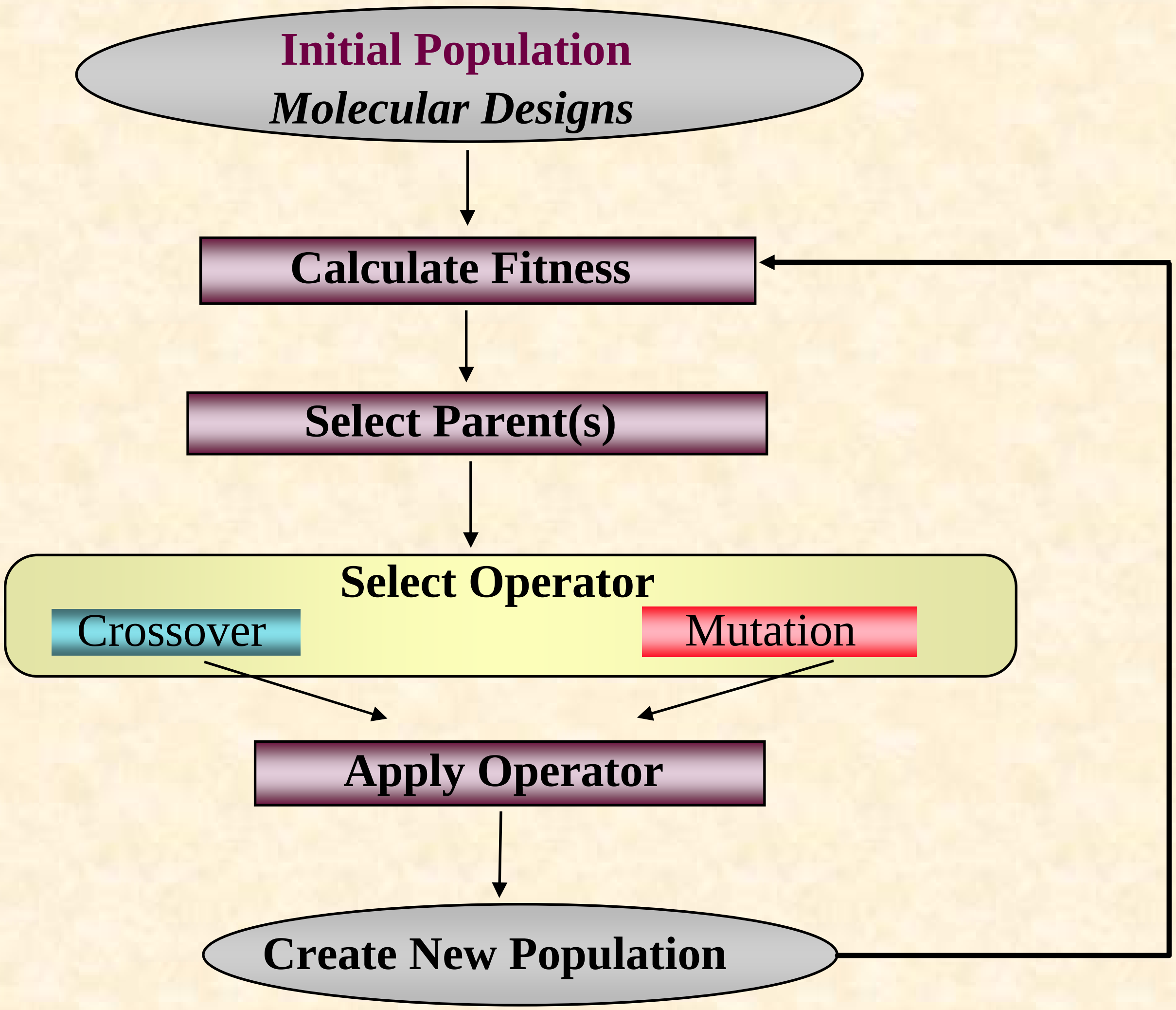
  Evolution = Random Change + Survival of the fittest

- Essential Components
  - Genetic Operators
    - Crossover
    - Mutation
  - Reproductive Plan
    - Fitness Proportionate Selection

Genetic Algorithms for Product Design

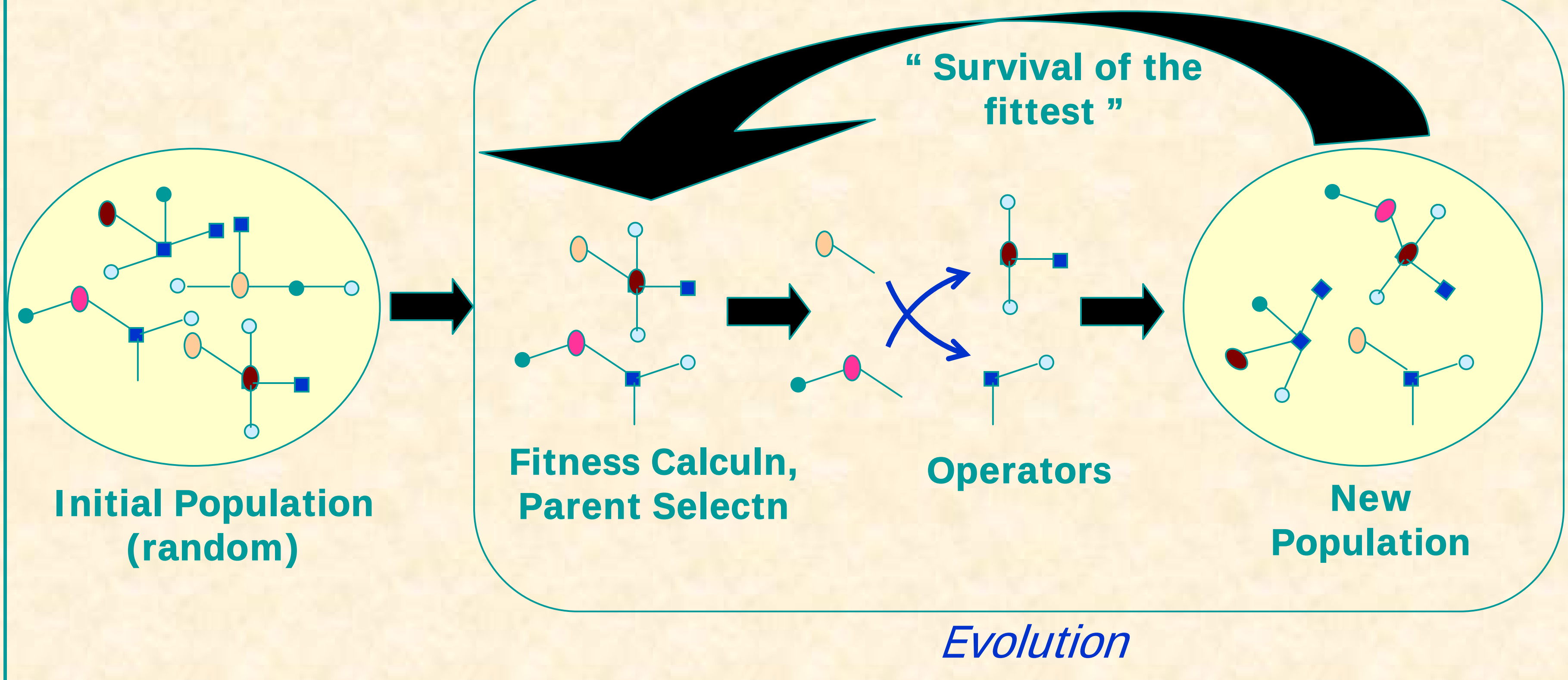
- Global Search
  - Diversity of solutions
  - High potential for novelty
  - Global Optima
- Development is de-coupled from forward problem
  - Robust to non-linearity
- Population based search
  - Ability to provide several near-optimal solutions
- Captures transparently the rich chemistry of the design problem

Overview of Genetic Algorithms



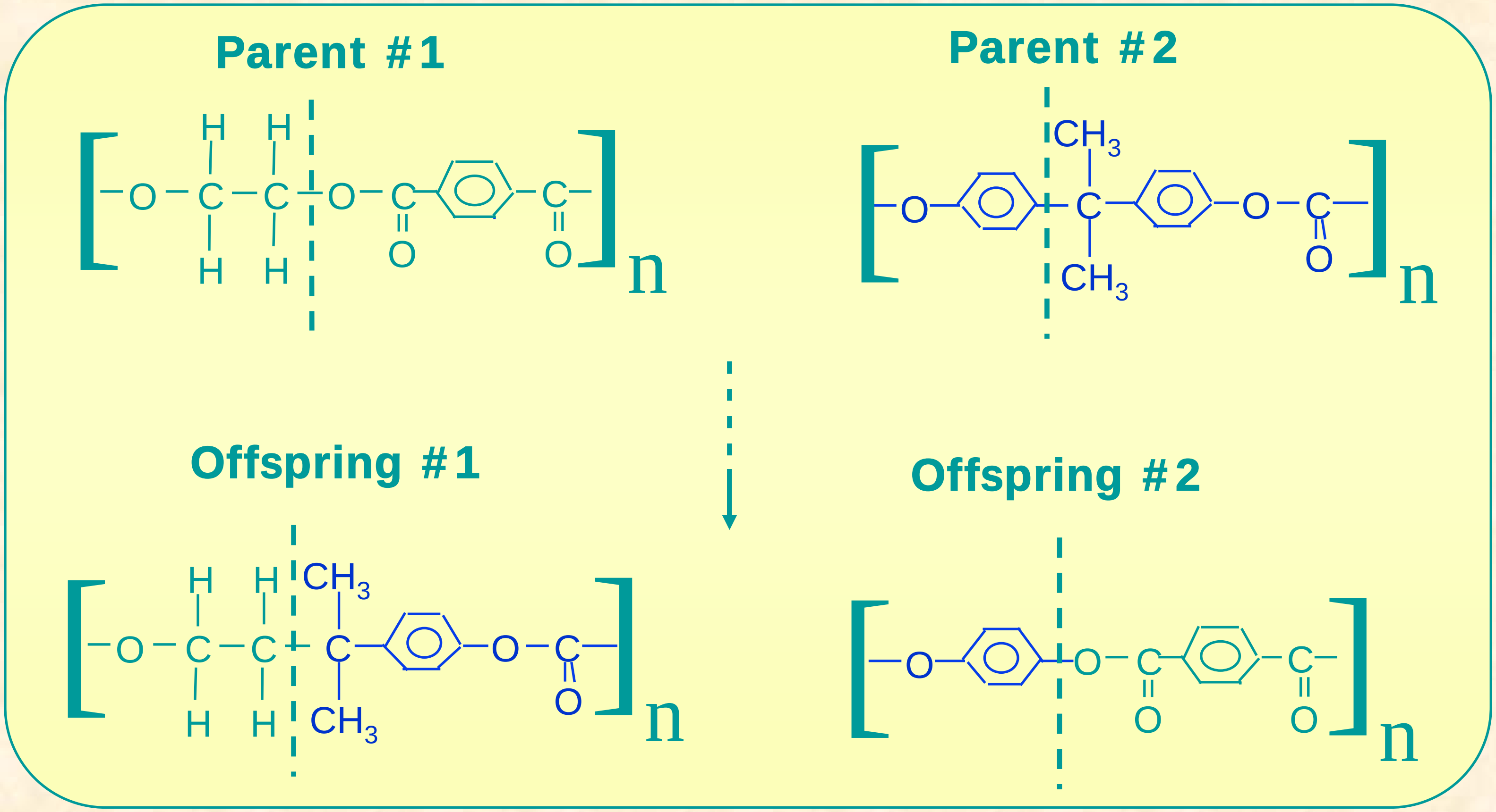
Genetic Algorithms (GA)

GAs are stochastic evolutionary search procedures based on the Darwinian model of natural selection



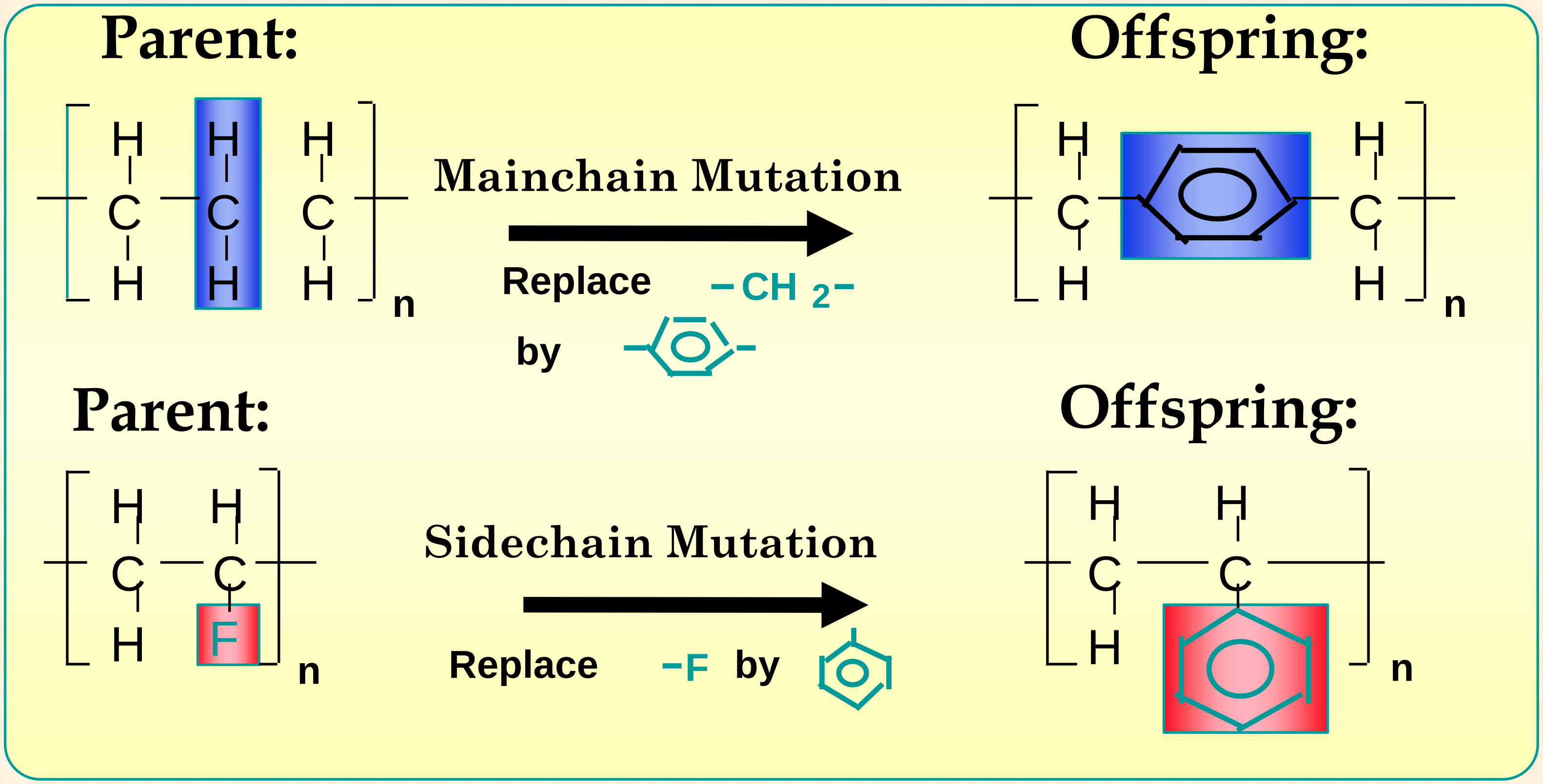
Genetic Operators: Crossover

- **Single-point Crossover**



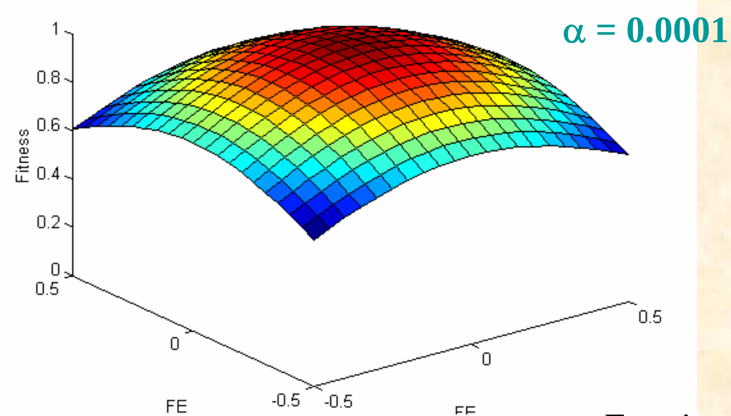
Mutation Operators

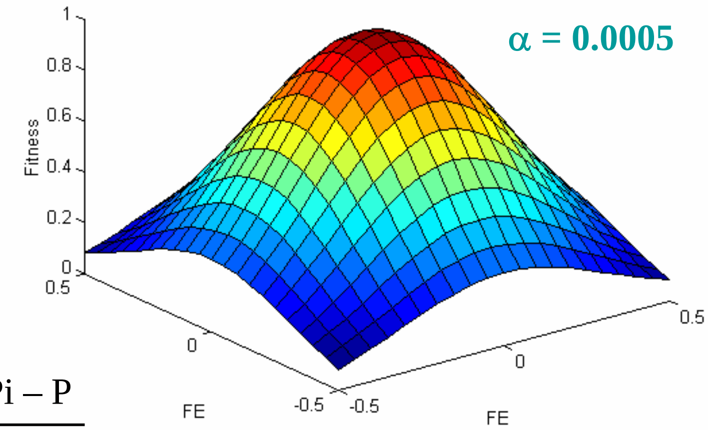
- Main-chain and Side-chain Mutation



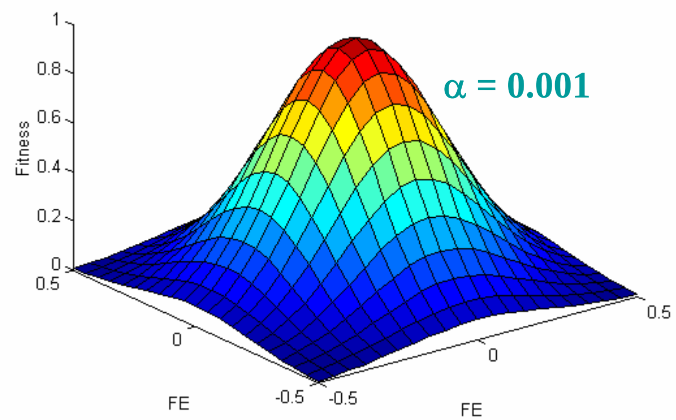
Fitness Function for GA-based CAMD

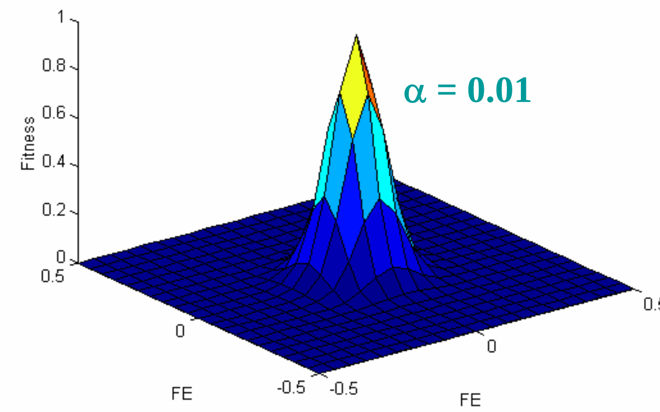
$$\text{Fitness (F)} = \exp\left(-\alpha\left[\left\{\frac{P_i - P}{P_{i,max} - P_{i,min}}\right\}^2\right]\right)$$





$$\text{Fractional Error (FE)} = \frac{Pi - P}{P}$$

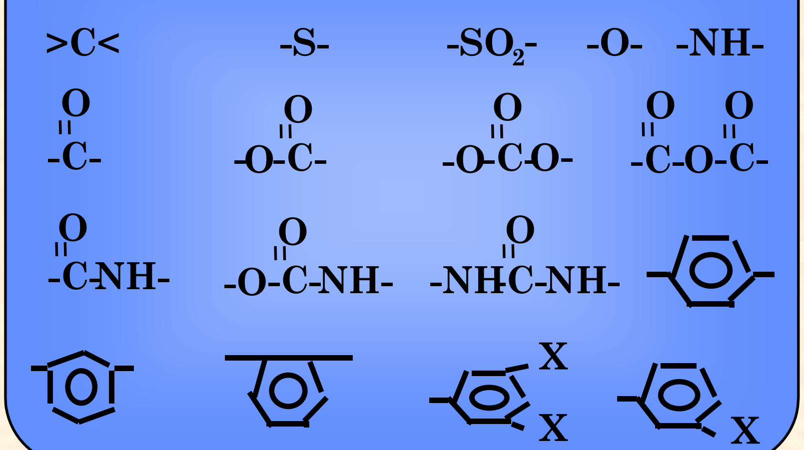




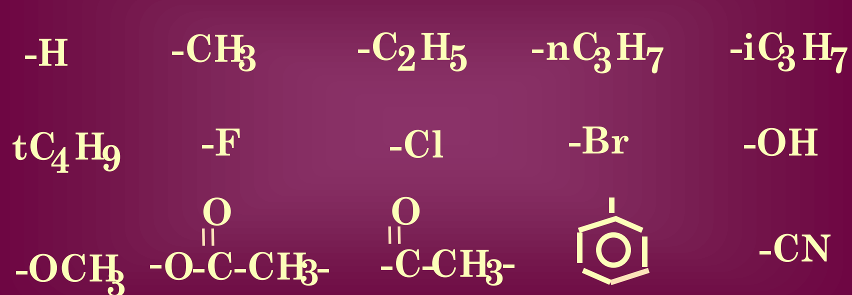
Polymer Design Case Study

- **Base Groups**

17 Main-chain Groups



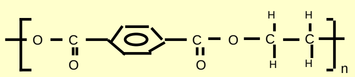
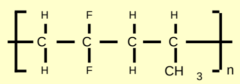
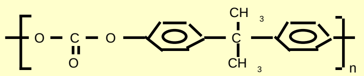
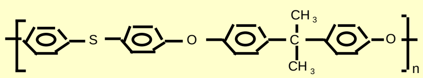
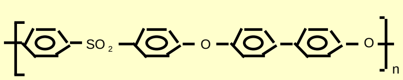
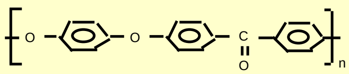
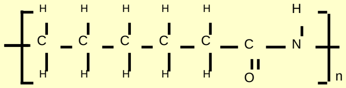
15 Side-chain Groups



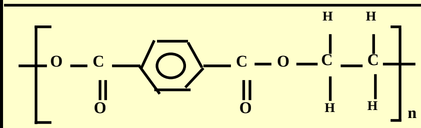
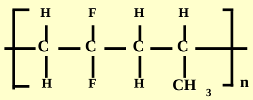
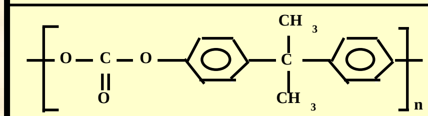
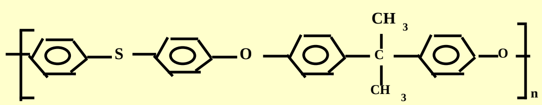
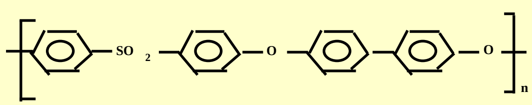
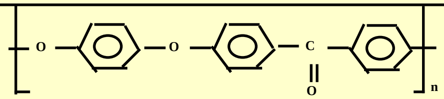
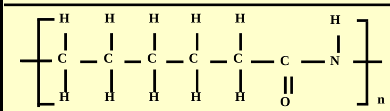
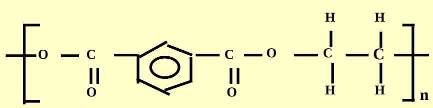
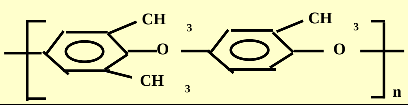
- **Target Properties**

• Density	• Glass Transition Temperature
• Thermal Expansion Coefficient	• Specific Heat Capacity
• Bulk Modulus	

Target Properties

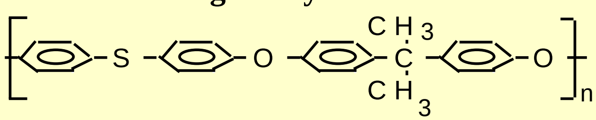
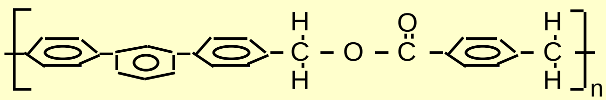
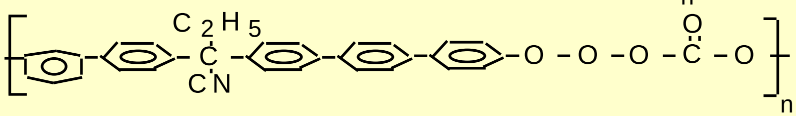
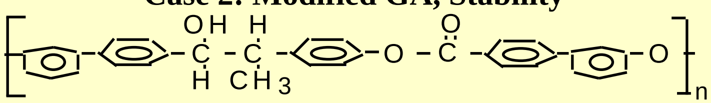
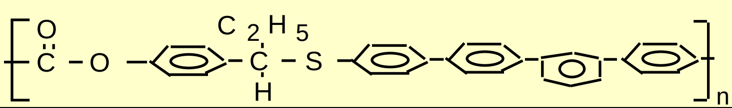
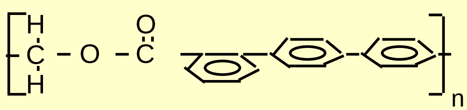
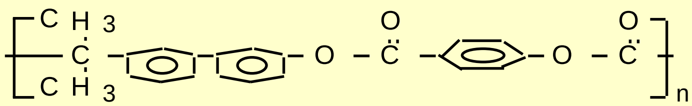
Target Polymer	Density (gm/cm³)	Glass Transition (K)	Thermal Expansion 1/K	Heat Capacity J/kg K	Bulk Modulus K N/m³
TP1: 	1.34	350.75	2.96×10^{-4}	1152.67	5.18×10^{9}
TP2: 	1.18	225.24	2.81×10^{-4}	1377.82	2.51×10^{9}
TP3: 	1.21	420.83	2.90×10^{-4}	1135.10	5.40×10^{9}
TP4: 	1.19	406.83	2.90×10^{-4}	1073.96	5.39×10^{9}
TP5: 	1.28	472.00	2.89×10^{-4}	995.95	5.31×10^{9}
TP6: 	1.25	421.12	2.90×10^{-4}	1016.55	6.12×10^{9}
TP7: 	1.06	322.55	2.98×10^{-4}	1455.90	3.85×10^{9}

Polymer Design Case Study: Results

Target Polymer	Random MC, SC	Random MC, SC Feasibility Constraints
	**12%** **184** **282**	**60%** **240** **213**
	**36%** **411** **6**	**48%** **522** **6**
	**0%** **-** **163**	**12%** **193** **74**
	**0%** **-** **861**	**0%** **-** **589**
	**32%** **400** **175**	**48%** **232** **142**
	**8%** **548** **199**	**32%** **632** **168**
	**100%** **61** **217**	**100%** **64** **198**
	**68%** **210** **162**	**88%** **109** **161**
	**8%** **382** **144**	**4%** **868** **70**

Legend: **Success Rate; Average Conv. Generation; Number with fitness >0.99**

Near Optimal Designs

Polymer Design	Overall Error	Fitness
Target Polymer: #4 	**0%**	**1.0**
Case 1: Standard GA 	**0.74%**	**0.995**
	**1.18%**	**0.991**
Case 2: Modified GA, Stability 	**0.25%**	**0.999**
	**1.10%**	**0.991**
Case 3: Modified GA, Stability & Complexity 	**0.21%**	**0.999**
	**0.83%**	**0.995**

Advantages of GAs for Product Design

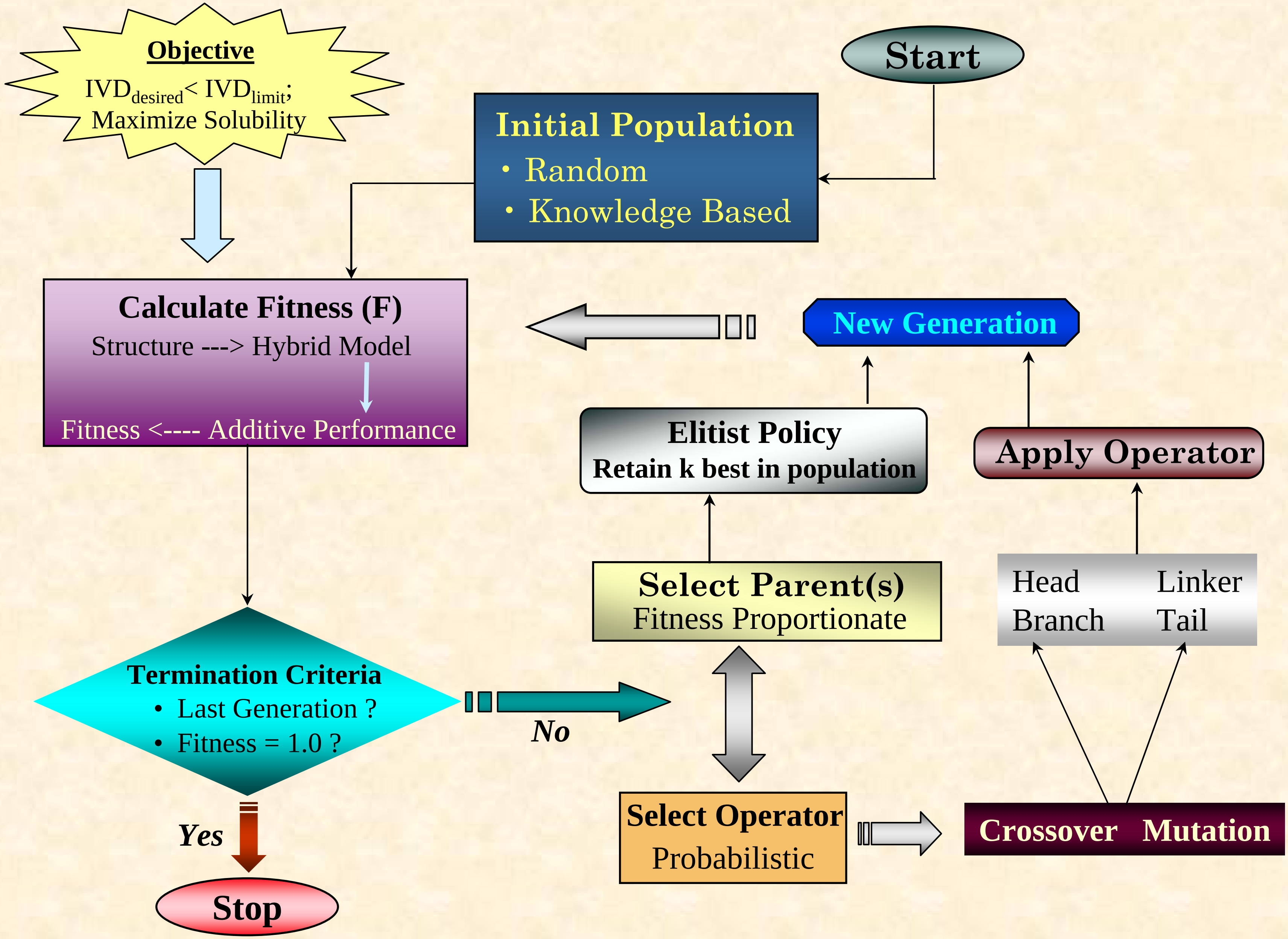
- Global Search
  - Diversity of solutions
  - High potential for novelty
  - Global Optima
- Development is de-coupled from forward problem
  - Robust to non-linearity
- Population based search
  - Ability to provide several near-optimal solutions
- Captures transparently the rich chemistry of the design problem

Drawbacks of GAs

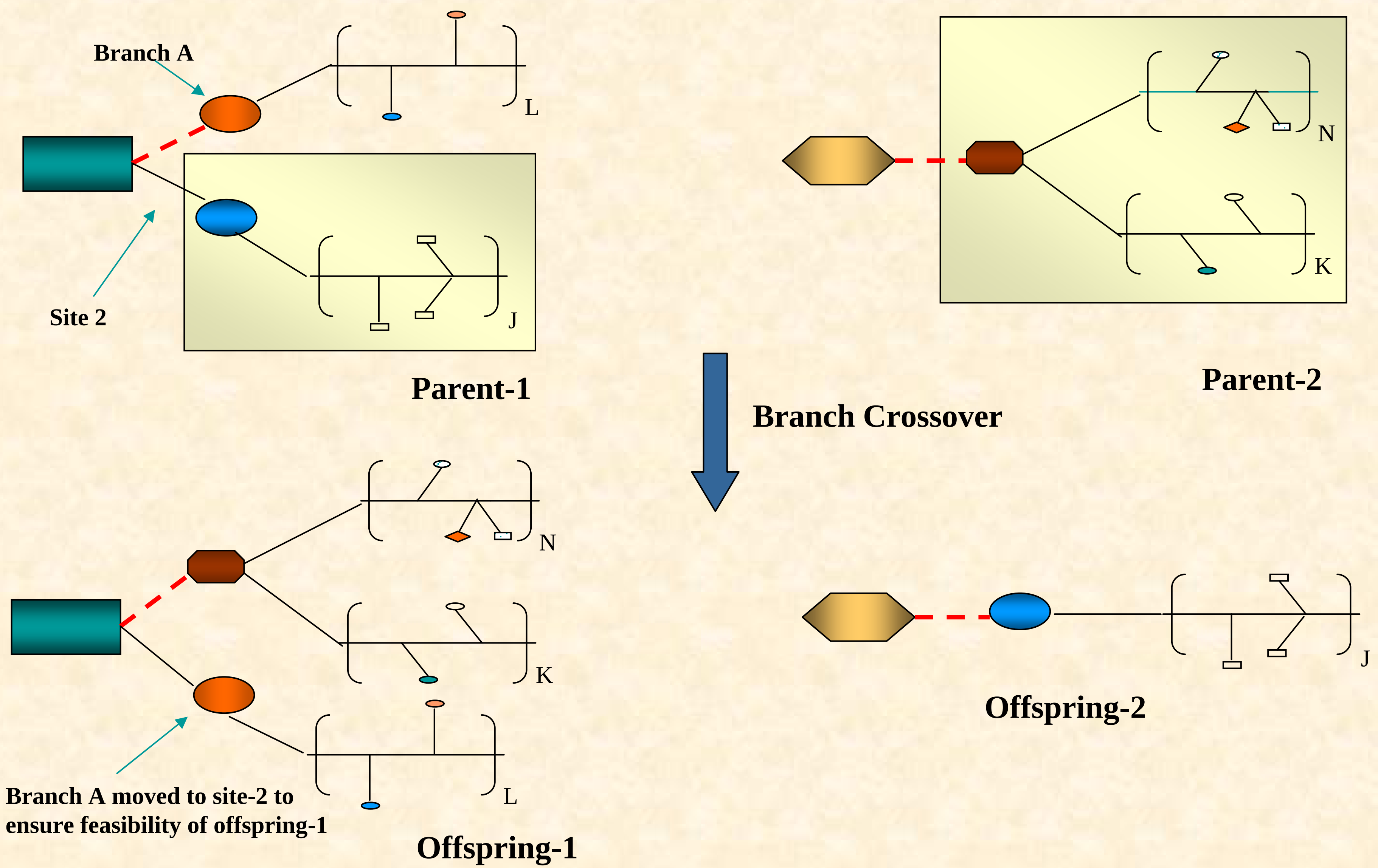
Powerful generic method with some drawbacks

- No convergence guarantees
- Performance sensitive to parameters
- Performance dependent on search space structure

Evolutionary Design of Fuel-Additives



Genetic Operators: Branch Crossover



Evolutionary Design of Fuel Additives: Results

Objective: Determine a structure with IVD < 10 mg
Dosage: 50 PTB; Population Size: 25; Generations: 25

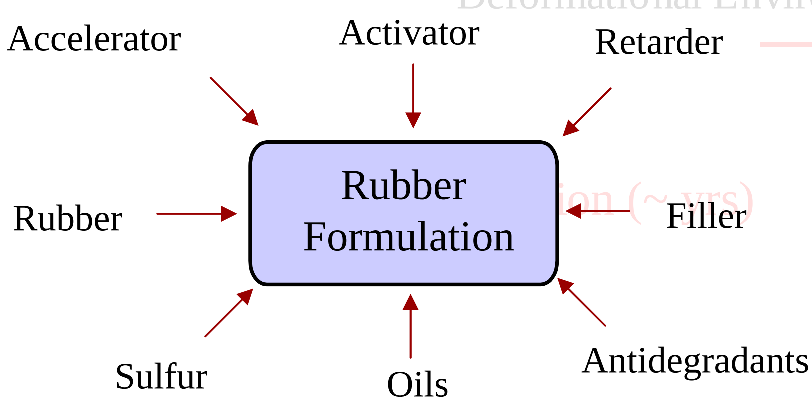
Run	Rank/Identifier	Fitness	Predicted IVD (PLS-NN Model)	Structural Description
I	*1, I-1*	0.997	11.4 mg	Novel Structure. Infrequently used linker.
	2, I-2	0.996	11.5 mg	Novel Structure. Same tails as best structure, different heads and linkers
	6, I-6	0.993	12.0 mg	Variant of structure found in the BMW database. Same head and linkers, different tails
II	*1, II-1*	0.999	10.1 mg	Novel Structure. Different from ***I-1***. Infrequently used linker component.
	2, II-2	0.989	12.6 mg	Slight variant of additive structure found in BMW and HONDA databases. Different tails but same head and linker
	4, II-4	0.983	13.2 mg	Minor variation of structure ***II-2*** above. Slight modification of the head
III	*1, III-1*	1.00	8.9 mg	Novel Structure. Different from ***I-1*** and ***II-1***. Commonly used components
	2, III-2	0.994	11.9 mg	Variant of ***III-1***. One linker and tail modified.
	3, III-3	0.993	12.1 mg	Variant of structure ***II-2*** above. Slight modification of head. A linker and tail inserted.

Rubber Parts in Service

Tires, Treads, Hoses, Shock Absorbers, O-rings, Gaskets, Mounts ……..

Problem Definition

Given the Desired Property/Performance of Interest, what is the Optimal Rubber Formulation?

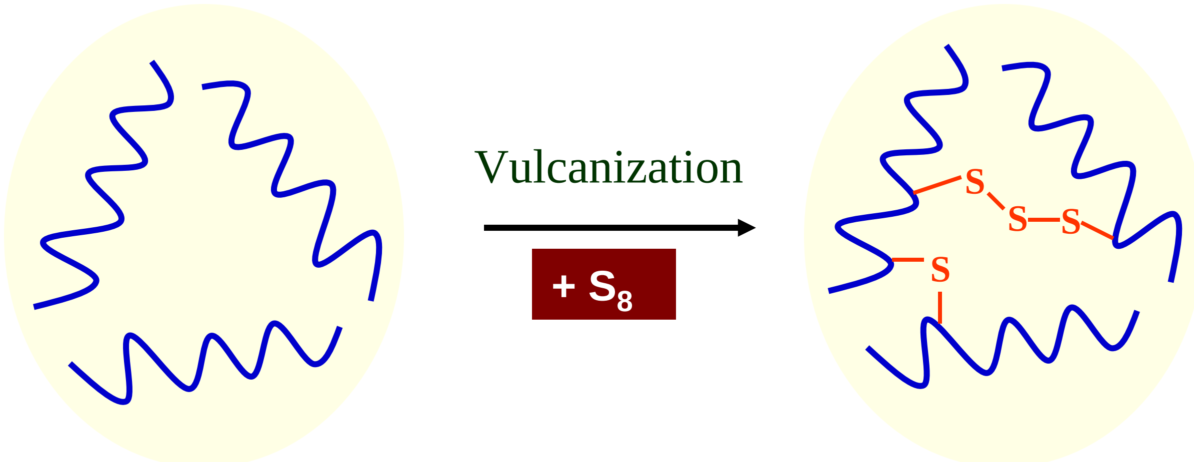


Which materials to include?

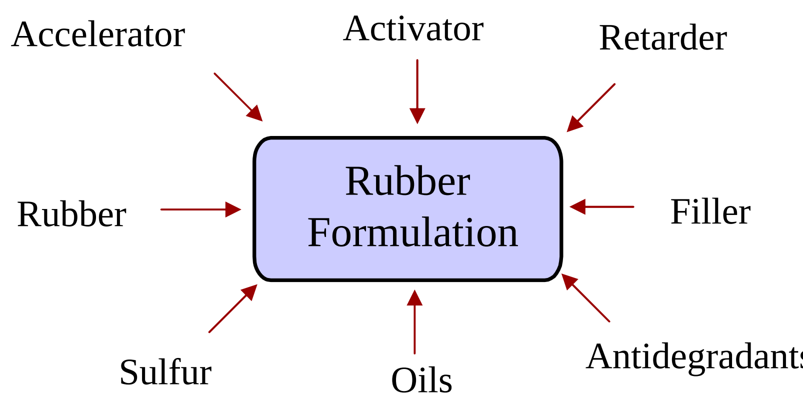
How much of them to include?

What should be the processing conditions?

What are Sulfur-Vulcanized Elastomers?

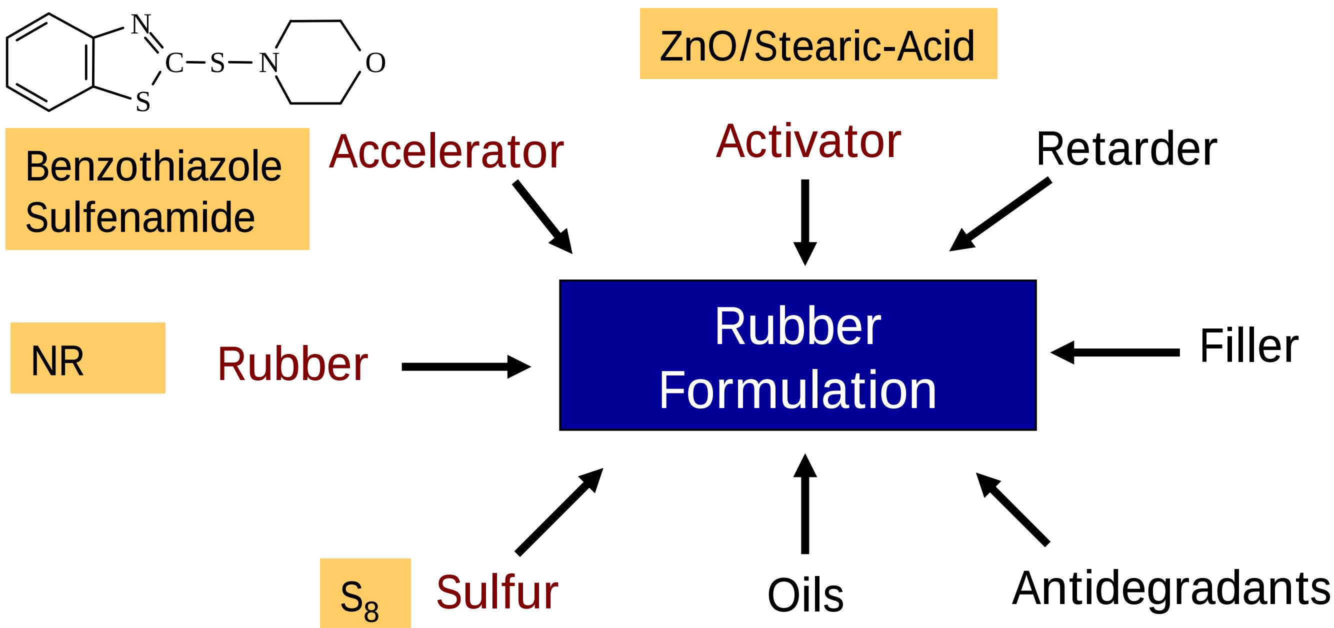


Accelerated Sulfur Vulcanization



OBJECTIVE OF THIS WORK

• Understand the Mechanistic Details of Accelerated-Sulfur-Vulcanization.



• Provide a Mathematical Description of the Kinetics of Accelerated-Sulfur Vulcanization consistent with Mechanistic Chemistry

COMPLEXITY OF VULCANIZATION REACTIONS

W.Scheele (1956) [1] Perhaps no-where in chemistry is there encountered a field which even in its literature alone shows so many uncertainties and (possibly only apparent) contradictions as that of the vulcanization of rubber.

[1]W. Scheele, O. Lorenz and W. Dummer, *Rubber Chemistry and Technology*, 29, 1 (1956)

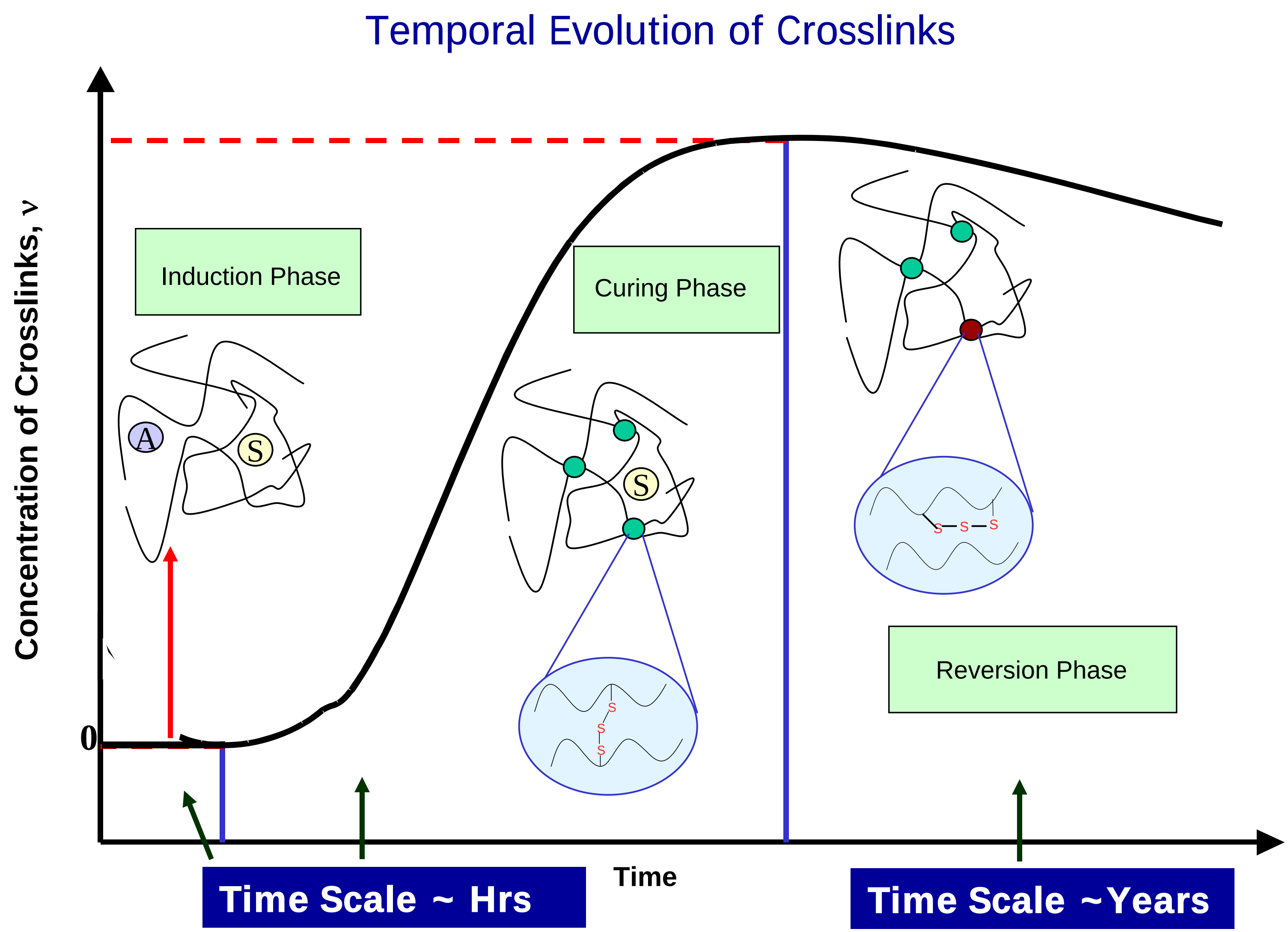
L.Bateman (1963) [2] Whilst it has long been appreciated, albeit intuitively, that sulfur vulcanization is a very complex chemical process, the actual complexity as revealed in the studies described above is probably far in excess of what has ever been envisaged.

[2] L. Bateman, C.G. Moore, M. Porter and B. Saville, *"The Chemistry and Physics of Rubber-Like Substances"*, L Bateman, Ed., Maclaren & Sons Ltd., London, 1963, pp 449.

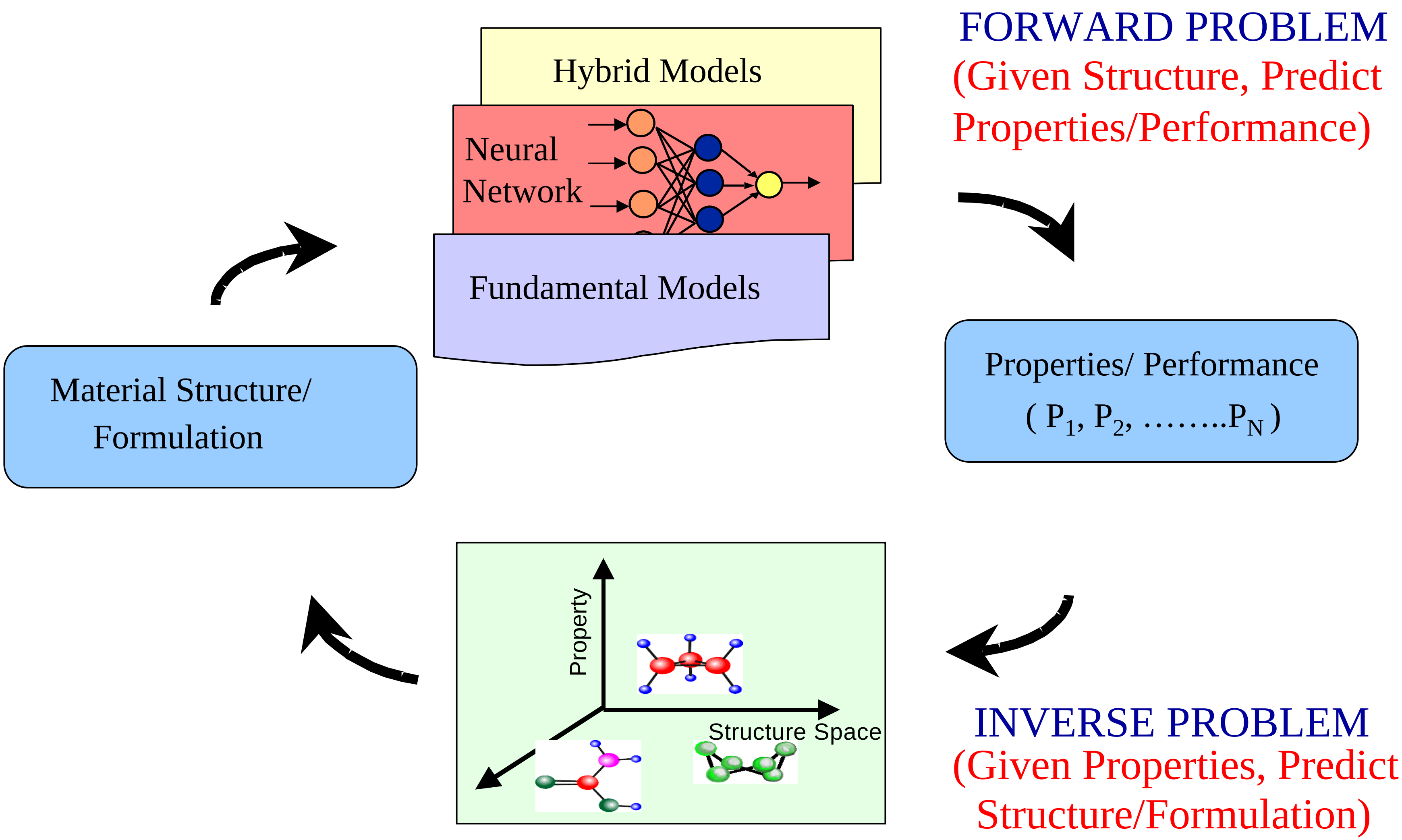
P.J. Nieuwenhuizen (1997) [3] It must be considered remarkable that despite all the efforts and progress in the field of vulcanization during the past decade, one has to conclude that the statements of Scheele and Bateman, made 30 to 40 years ago, are still true to a great extent.

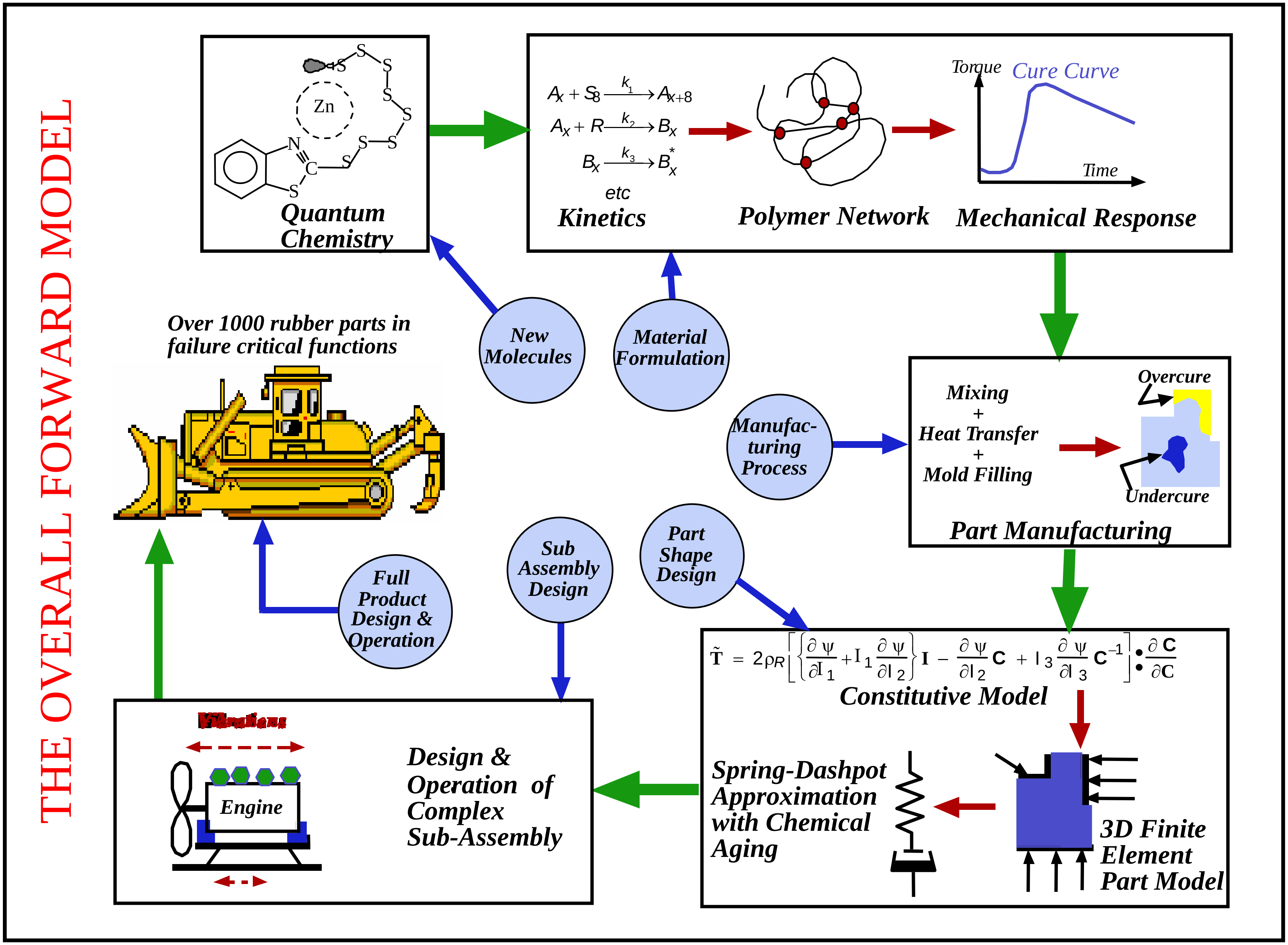
[3] P.J. Nieuwenhuizen, *Rubber Reviews, Rubber Chemistry and Technology*, 29, 70 (1997)

SCHEMATIC OF A VULCANIZATION CURVE



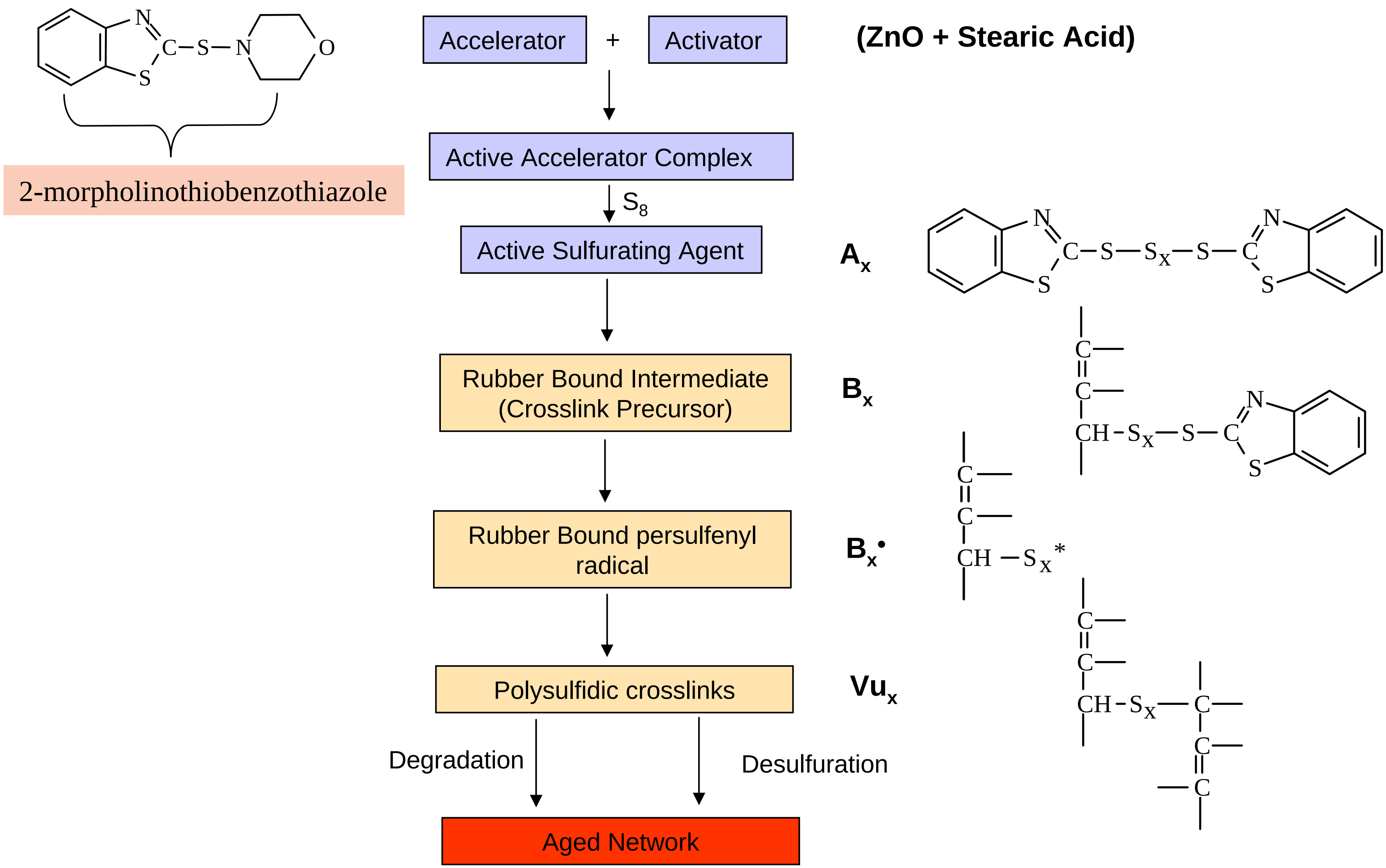
CAMD Sub-Problems





School of Chemical Engineering, Purdue University

Schematic of a Vulcanization Curve



Accelerator Chemistry

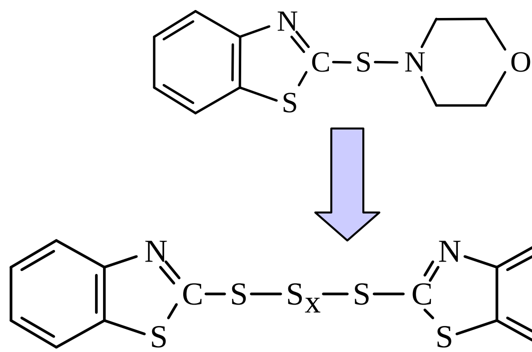
Accelerator + Activator

↓

Active Accelerator Complex

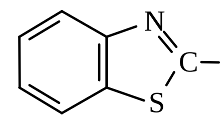
↓ S_8

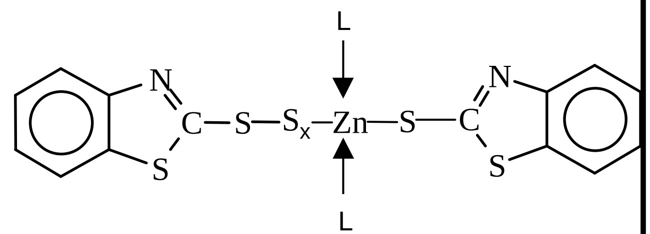
Active Sulfurating Agent



Without Activator

Bt =





With Activator

$$BtS\text{-}NR_2 \longrightarrow BtSH + R_2NH$$

$$BtS\text{-}NR_2 + BtSH \longrightarrow BtSSBt + R_2NH$$

~~$BtS\text{-}SBt + S_8 \longrightarrow BtS\text{-}S_8\text{-}SBt$~~ (✗)

$$BtS\text{-}SBt + S_8 \longrightarrow BtS\text{-}S\text{-}SBt + S_7 \longrightarrow BtS\text{-}S_2\text{-}SBt + S_6 \cdots$$

$$BtS\text{-}S_x\text{-}SBt + BtS\text{-}S_y\text{-}SBt \rightleftharpoons BtS\text{-}S_z\text{-}SBt + BtS\text{-}S_w\text{-}SBt$$

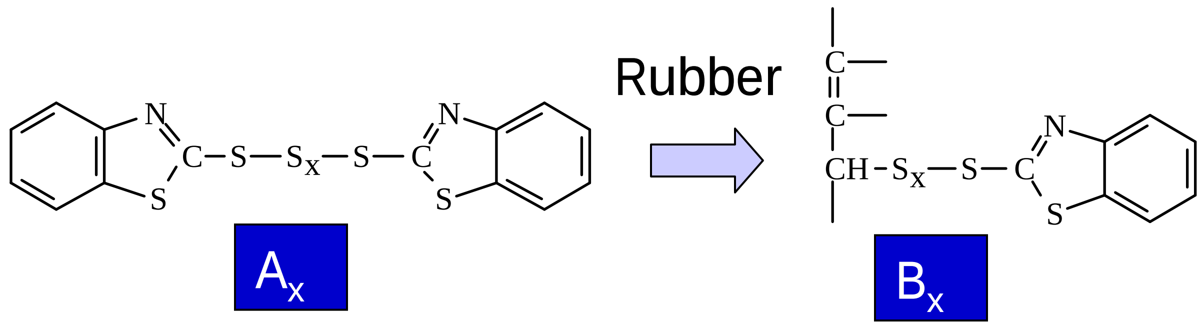
$$A_x = BtS\text{-}S_x\text{-}SBt$$

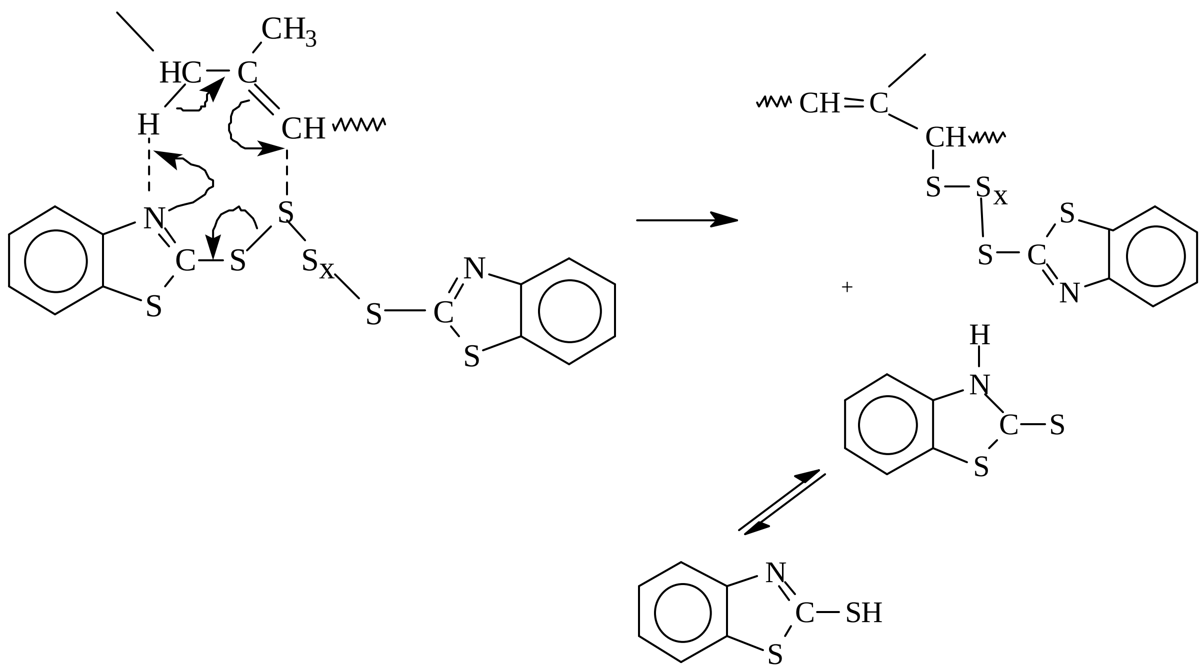
Crosslinking Chemistry (I)

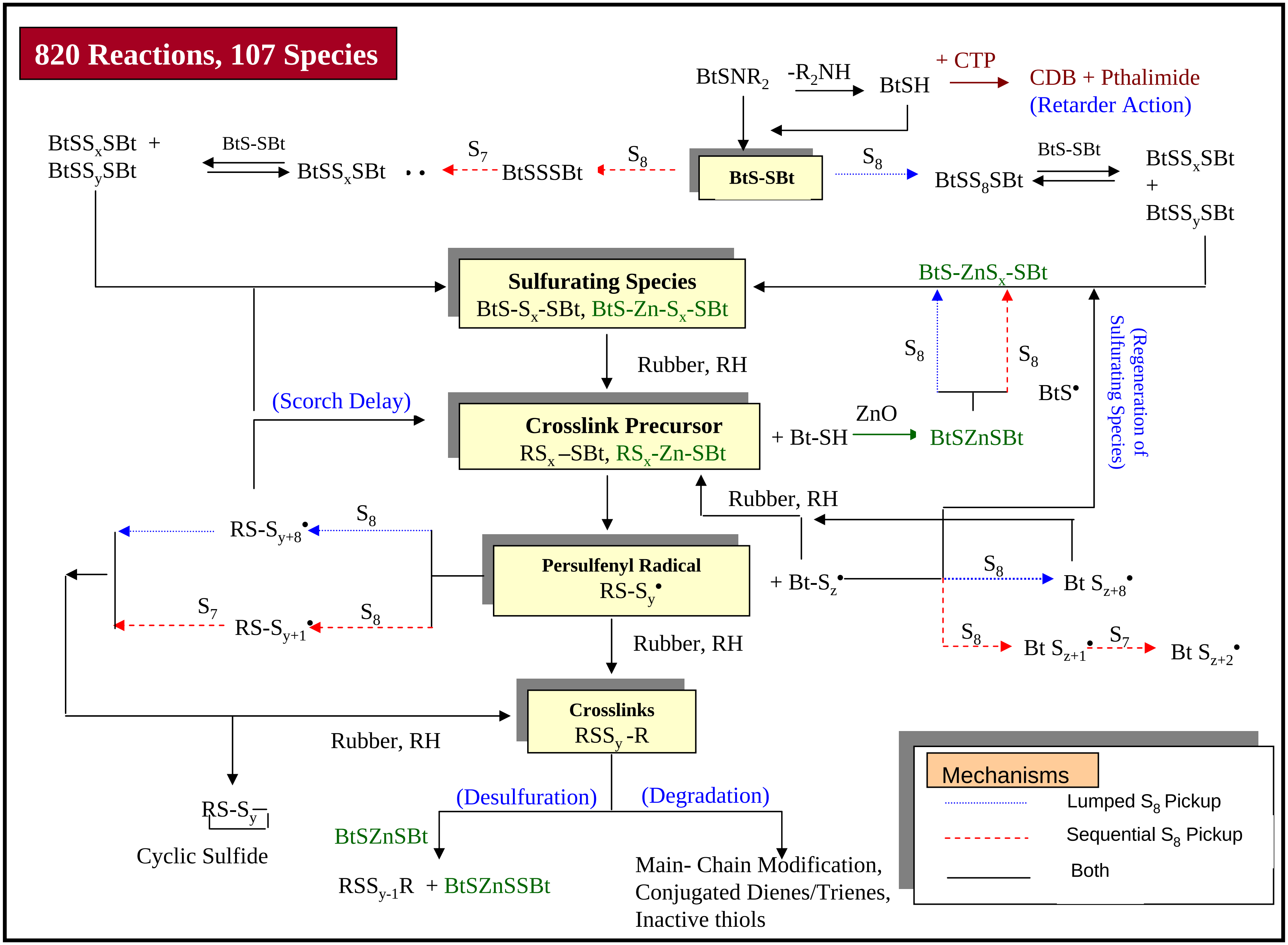
Active Sulfurating Agent

↓

Rubber Bound Intermediate
(Crosslink Precursor)







School of Chemical Engineering, Purdue University

POPULATION BALANCE EQUATIONS

Equations for MBTS and other sulfurating species

$$\frac{d}{dt}[A_0] = k_{MBS}[MBS] - k_{MBS\text{-}MBT}[MBT][MBS] - k_{A\text{-}A}[A_0]\sum_{r=2}^{14}(r-1)A_r - k_{A\text{-}S}[A_0][S_8]$$

$$-k_{A\text{-}R}[A_0] - k_{A\text{-}BST}[A_0]\sum_{r=1}^{16}B^*{}_r + 0.5k_{E\text{-}E}[E_0^*]^2 - k_{DESULF}[A_0]\sum_{r=2}^{16}Vu_r +$$

$$k_{A\text{-}A}\begin{bmatrix}[A_1]^2 + 2[A_1][A_2] + 2\left\{[A_1][A_3] + 0.5[A_2]^2\right\} + 2\left\{[A_1][A_4] + [A_2][A_3]\right\} + \\ 2\left\{[A_1][A_5] + [A_2][A_4] + 0.5[A_3]^2\right\} + 2\left\{[A_1][A_6] + [A_2][A_5] + [A_3][A_4]\right\} + \\ 2\left\{[A_1][A_7] + [A_2][A_6] + [A_3][A_5] + 0.5[A_4]^2\right\} + \\ 2\left\{[A_1][A_8] + [A_2][A_7] + [A_3][A_6] + [A_4][A_5]\right\} + \\ 2\left\{[A_1][A_9] + [A_2][A_8] + [A_3][A_7] + [A_4][A_6] + 0.5[A_5]^2\right\} + \\ 2\left\{[A_1][A_{10}] + [A_2][A_9] + [A_3][A_8] + [A_4][A_7] + [A_5][A_6]\right\} + \\ 2\left\{[A_1][A_{11}] + [A_2][A_{10}] + [A_3][A_9] + [A_4][A_8] + [A_5][A_7] + 0.5[A_6]^2\right\} + \\ 2\left\{[A_1][A_{12}] + [A_2][A_{11}] + [A_3][A_{10}] + [A_4][A_9] + [A_5][A_8] + [A_6][A_7]\right\} + \\ 2\left\{[A_1][A_{13}] + [A_2][A_{12}] + [A_3][A_{11}] + [A_4][A_{10}] + [A_5][A_9] + [A_6][A_8] + 0.5[A_7]^2\right\}\end{bmatrix}$$

$$\frac{d}{dt}[A_1] = -2k_{A\text{-}R}[A_1] - k_{A\text{-}S}[A_1][S_8] + 2k_{A\text{-}A}[A_0]\sum_{r=2}^{14}A_r + k_{DESULF}[A_0]\sum_{r=2}^{16}Vu_r$$

$$-2k_{A\text{-}A}[A_1]\sum_{r=1}^{13}A_r + k_{E\text{-}E}[E_0^*][E_1^*]$$

POPULATION BALANCE EQUATIONS

Equations for MBTS and other sulfurating species

$$\frac{d}{dt}[A_0] = k_{A2}[MBT][MBS] - k_{A3}[A_0][S_8] - k_{A4}[A_0]\left[\sum_{i=2}^{14}(i-1)A_i\right] - k_{C1}[A_0]$$

$$- k_{C4}[A_0]\left[\sum_{i=1}^{16}B_i*\right] - k_{R1}[A_0]\left[\sum_{i=1}^{16}Vu_i\right]$$

$$\frac{d}{dt}[A_1] = -2k_{C1}[A_1] - k_{A3}[A_1][S_8] - 2k_{A4}[MBTS][A_1] + k_{R1}[A_0]\left[\sum_{i=1}^{16}Vu_i\right]$$

$$\frac{d}{dt}[A_i] = -2k_{C1}[A_i] - k_{A3}[A_i]\begin{cases}[S_8], & 2 \le i \le 6 \\ 0, & i > 7\end{cases} - (i+1)k_{A4}[A_0][A_i] + k_{A3}[A_i]\begin{cases}0, & 2 \le i \le 6 \\ [S_8] & i > 7\end{cases}$$

Equations for Crosslink Precursors

$$\frac{d}{dt}[B_0] = (k_{C1}[MBTS] + k_{C7}[E_1*])$$

$$\frac{d}{dt}[B_i] = -k_{A4}[A_0][B_i*] + 2k_{C1}[A_i] - (i-1)k_{C2}[B_i] + k_{C7}[E_{i+1}*]$$

RATE-CONSTANT DETERMINATION

FACTS

Total of 820 different reactions
107 Coupled Ordinary Nonlinear Differential Equations
9 Optimizable Rate Constants

$$\min_{\mathrm{k}} \sum_{i=1}^{q} \sum_{j=1}^{m} \left[\frac{\nu_{i,j}(\mathrm{k}) - \nu_{i,j}^{\exp}}{\nu_{i,j}^{\exp}} \right]^2, \quad \mathrm{k} = [k_1, k_2, k_3 \ldots\ldots k_P]$$

$s.t.$

m time points, q concentrations

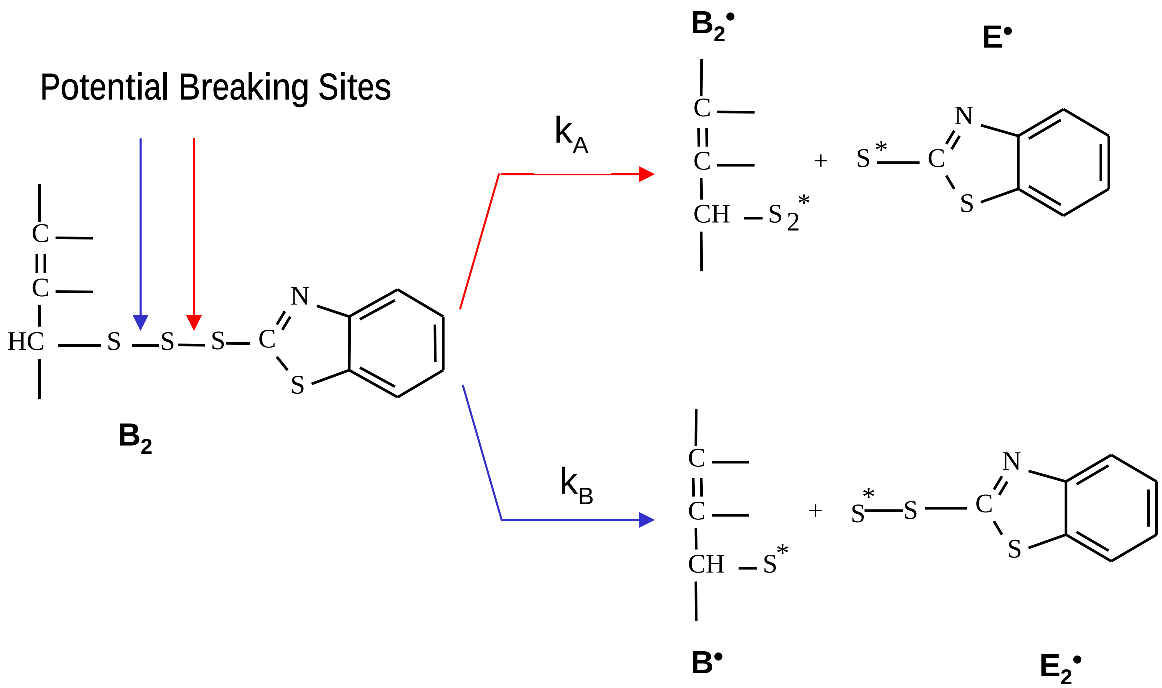
$$\frac{d}{dt}[\nu] = \phi(\nu, c, \mathrm{k})$$

$$\mathrm{k} \geq 0$$

$$\mathrm{k} = Rate\,Constants$$

$$\nu_{i,j}^{\exp} = Experimental\ Data\ Point\ at\ time\ i\ and\ conc\ j$$

Example Writing Population Balance Equations



$$\frac{d}{dt}[B_2] = -(k_A + k_B)[B_2], \quad k_A = k_A^0 \exp\left(-\frac{E_A}{RT}\right), \; k_B = k_B^0 \exp\left(-\frac{E_B}{RT}\right)$$

$$\frac{d}{dt}[B_2^\bullet] = k_A [B_2], \; \frac{d}{dt}[E^\bullet] = k_A [B_2]$$

$$\frac{d}{dt}[B^\bullet] = k_B [B_2], \; \frac{d}{dt}[E_2^\bullet] = k_B [B_2]$$

Population Balance Equations

Equations for MBTS and other sulfurating species

x = 0

$$\frac{d}{dt}[A_0] = k_{MBS\text{-}MBT}[MBT][MBS] - k_{A\text{-}A}[A_0]\sum_{r=2}^{14}(r-1)A_r - k_{A\text{-}S}[A_0]\sum_{y=1}^{8}[S_y]$$

$$-k_{A\text{-}R}[A_0] - k_{A\text{-}BST}[A_0]\sum_{r=1}^{16}B^*{}_r + \frac{1}{2}k_{E\text{-}E}[E_0^*]^2 - k_{DESULF}[A_0]\sum_{r=2}^{16}Vu_r$$

$$+k_{A\text{-}A}\sum_{x=1}^{13}\sum_{y=1}^{13-x}A_x A_y$$

x = 1

$$\frac{d}{dt}[A_1] = -2k_{A\text{-}R}[A_1] - k_{A\text{-}S}([A_1]\text{-}[A_0])\sum_{y=1}^{8}[S_y] + 2k_{A\text{-}A}[A_0]\sum_{r=2}^{14}A_r + k_{DESULF}[A_0]\sum_{r=2}^{16}Vu_r$$

$$-2k_{A\text{-}A}[A_1]\sum_{r=1}^{13}A_r + k_{E\text{-}E}[E_0^*][E_1^*]$$

2 ≤ x ≤ 13

$$\frac{d}{dt}[A_x] = -2k_{A\text{-}R}[A_x] - k_{A\text{-}S}([A_x]\text{-}[A_{x\text{-}1}])\sum_{y=1}^{8}[S_y] + 2k_{A\text{-}A}[A_0]\sum_{r=x+1}^{14}A_r$$

$$-(x-1)k_{A\text{-}A}[A_0][A_x] - 2k_{A\text{-}A}[A_x]\sum_{r=1}^{14-x}A_r + k_{E\text{-}E}[E_0^*][E_x^*]$$

$$+k_{A\text{-}A}\sum_{r=1}^{x\text{-}1}[A_r][A_{x\text{-}r}]$$

Final Set of Population Balance Equations

$$\frac{d}{dt}\begin{bmatrix} \vdots \\ Ax \\ \vdots \\ Bx \\ \vdots \\ Bx^* \\ \vdots \\ Vu_x \\ \vdots \\ S_8 \\ MBT \\ E \end{bmatrix} = \begin{bmatrix} \vdots \\ f_1(k_{01}, k_{02}, k_1, k_4, Ax) \\ \vdots \\ f_2(k_1, k_2, k_4, Ax, Bx, Bx^*) \\ \vdots \\ f_3(k_2, k_3, k_4, Ax, Bx, Bx^*) \\ \vdots \\ f_4(k_3, k_6, Bx^*, Vu_x) \\ \vdots \\ f_5(k_{01}, A_0, S_8) \\ f_6(k_1, Ax) \\ f_7(k_2, k_4, k_{51}, k_{52}, Ax, Bx, Bx^*, E) \end{bmatrix}$$

Initial Conditions: $A_0(0) = [MBS]_0$ $S_8(0) = [Sulfur]_0$

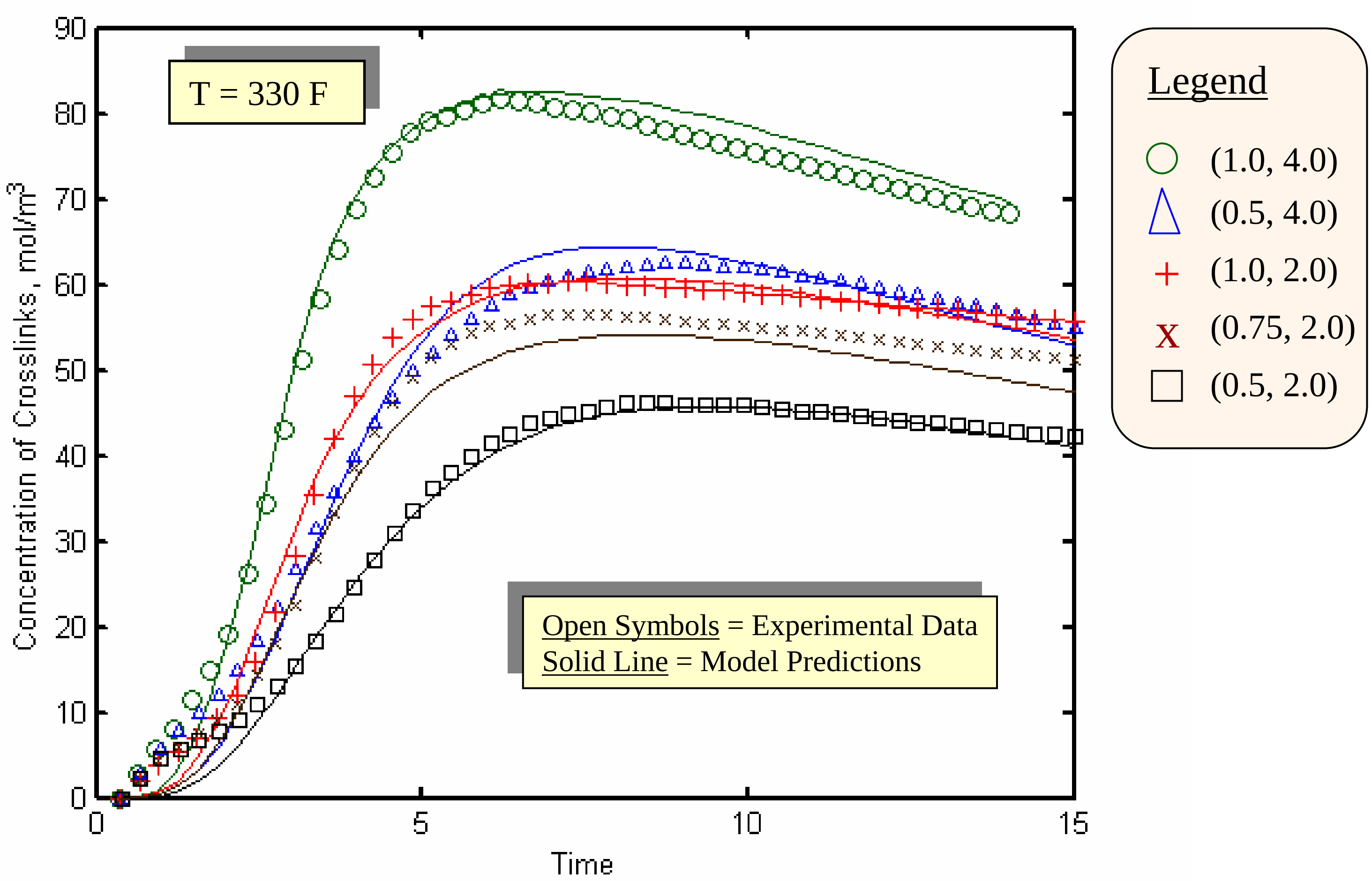
$A_{x(>0)}(0) = B_x(0) = B^*_x(0) = Vu_x(0) = MBT(0) = E(0) = 0$

Most simple kinetic model that explains all the important features of sulfur vulcanization.

Total of 820 different reactions considered with 107 coupled ordinary nonlinear differential equations and 9 optimizable rate-constants.

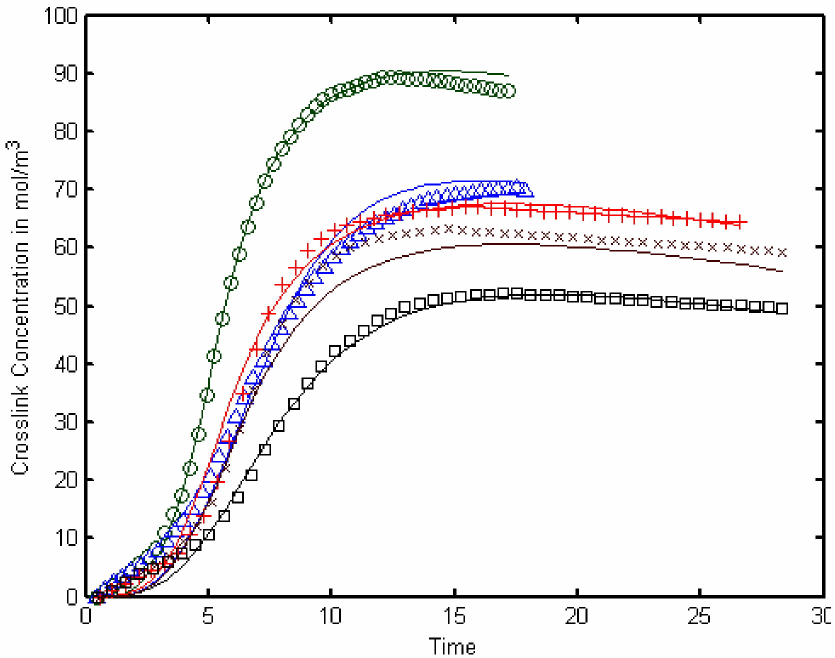
Population Balance Model - Predictions

Effect of Accelerator and Sulfur

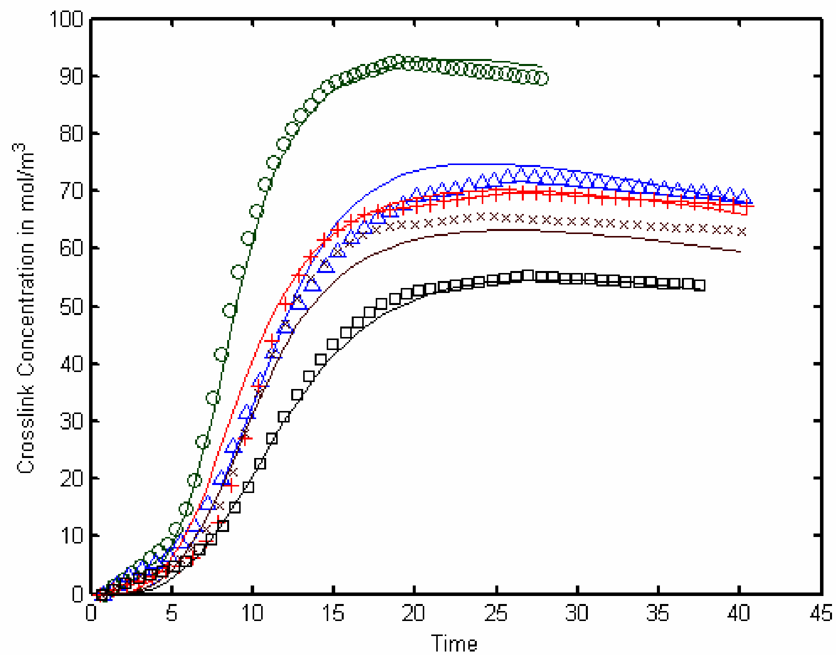


Population Balance Model – Different Temperatures

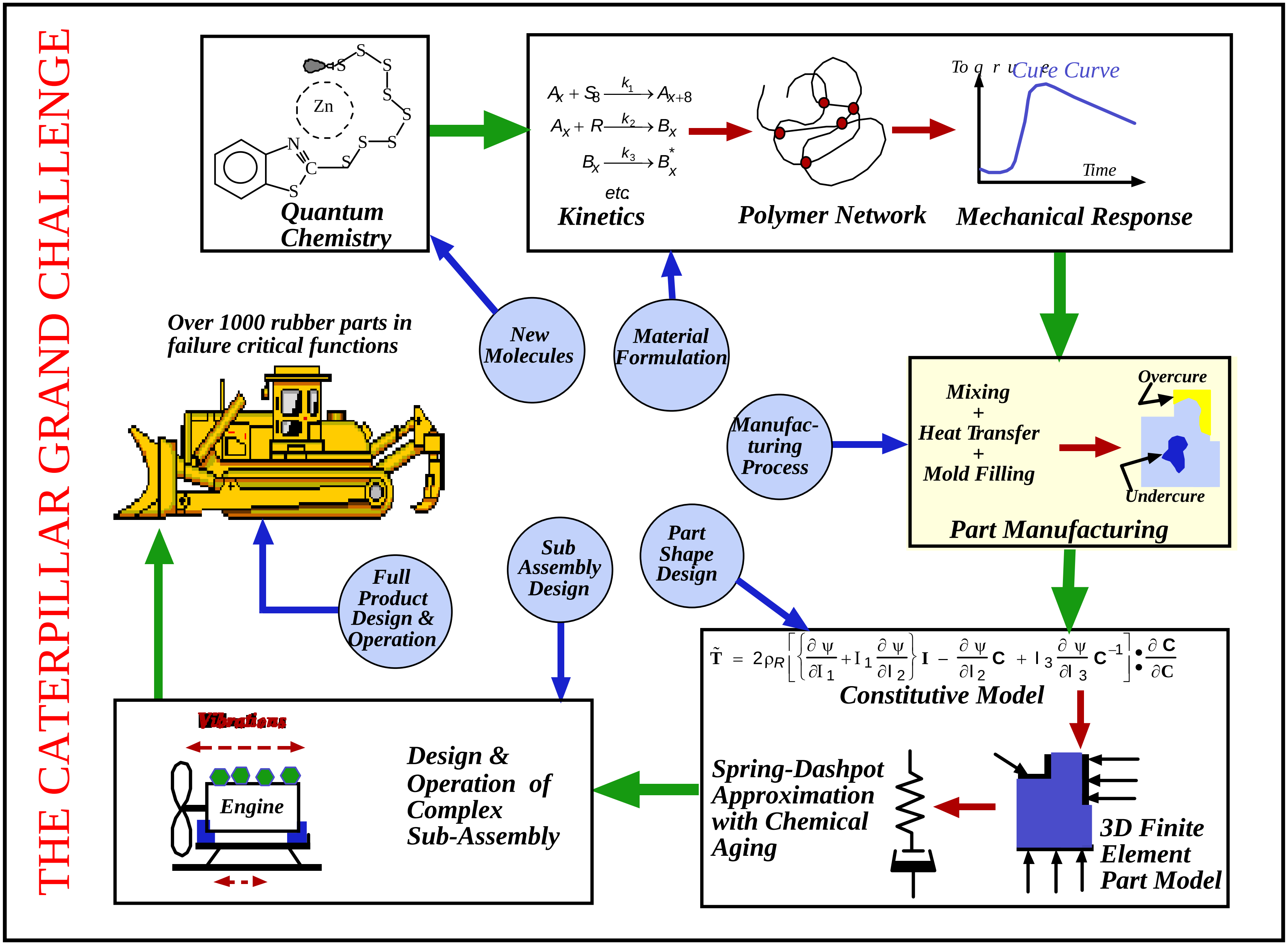
T = 310 F



T = 298 F



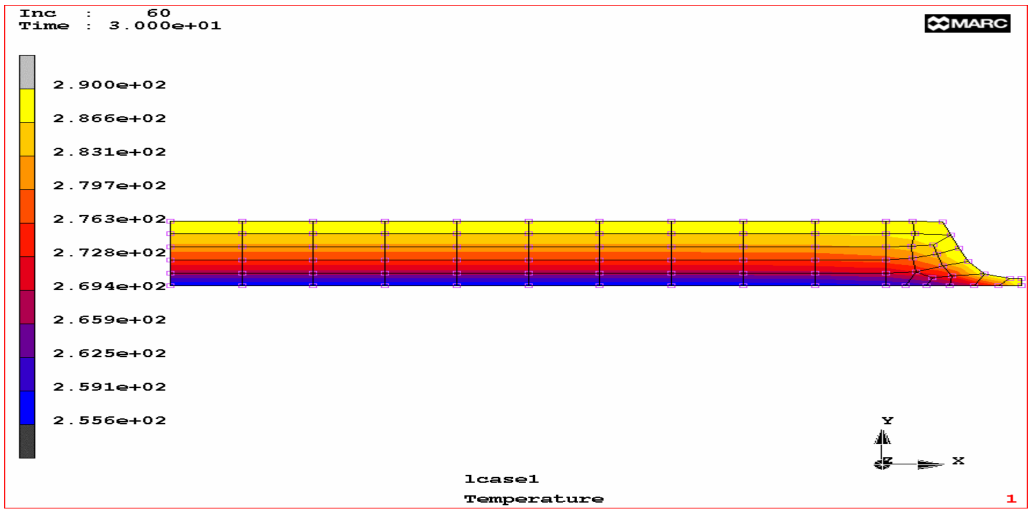
Open Symbols = Experimental Data
Solid Line = Model Predictions



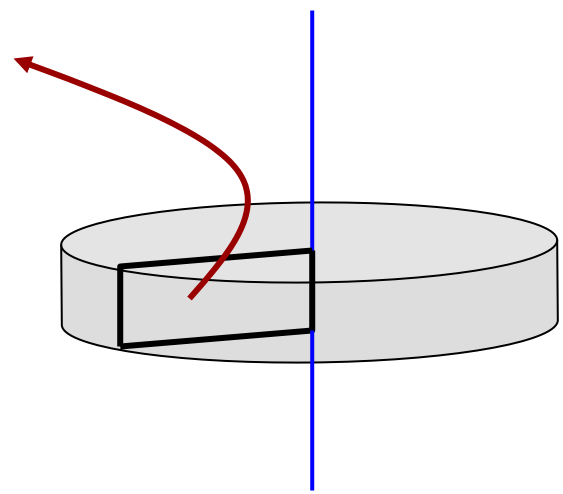
School of Chemical Engineering, Purdue University

Spatial Profile of the State-of-Cure for the Thermal History in the Mold

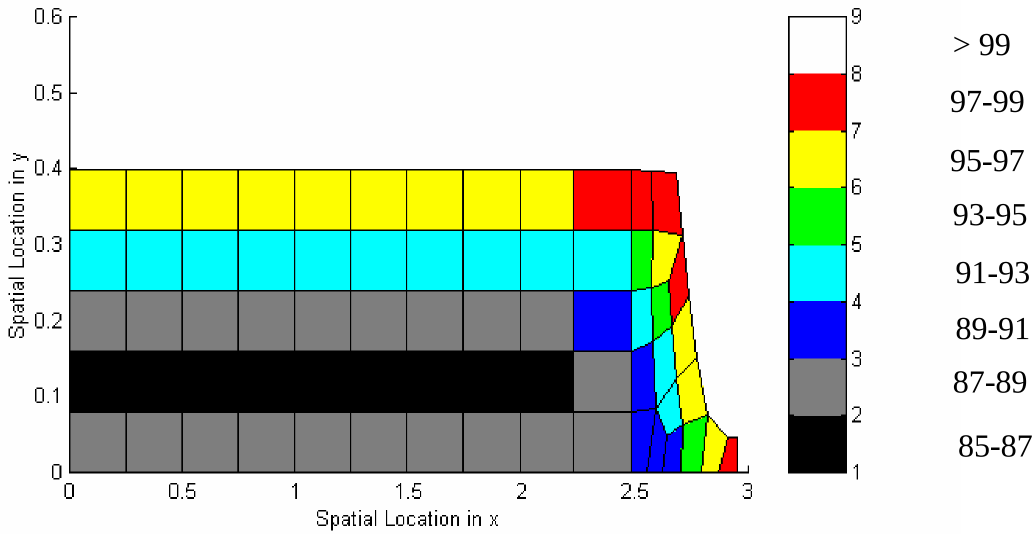
Thermal Profile



Actual Part Design



Cure Profile



Result Summary Vulcanization

Vulcanization Chemistry

Mechanistic Evaluation

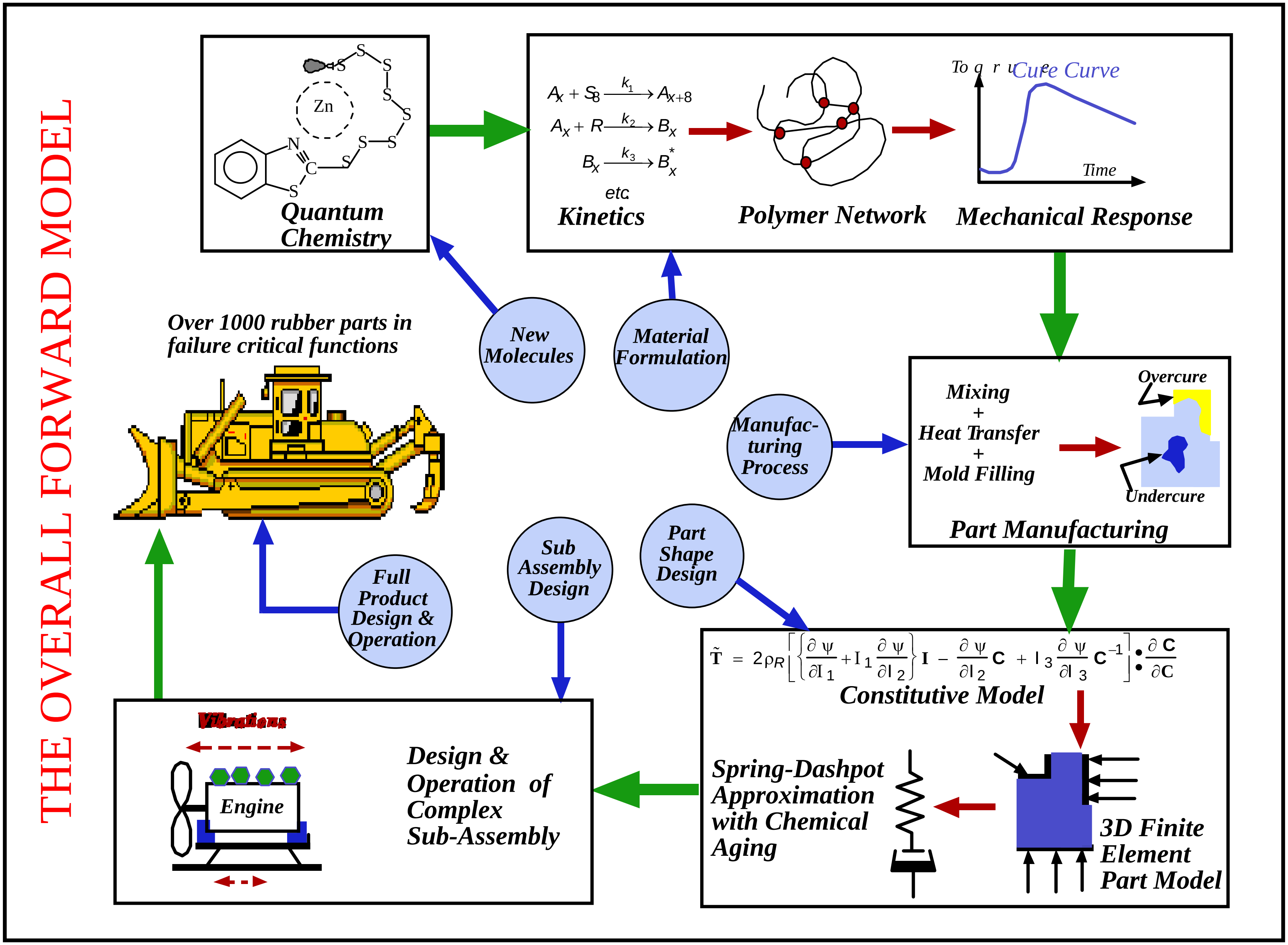
- Evaluated all possible mechanisms critically for accelerated sulfur vulcanization
- Identified the simplest set of reactions that describe all the important aspects of cure.

Mathematical Quantification

- Developed Population Balance Model based on mechanistic details to describe the different aspects of cure.
- Single set of parameters that predicts cure details for all formulations at all temperatures.
- Model extended for retarders, filled systems and other accelerator and elastomer classes.

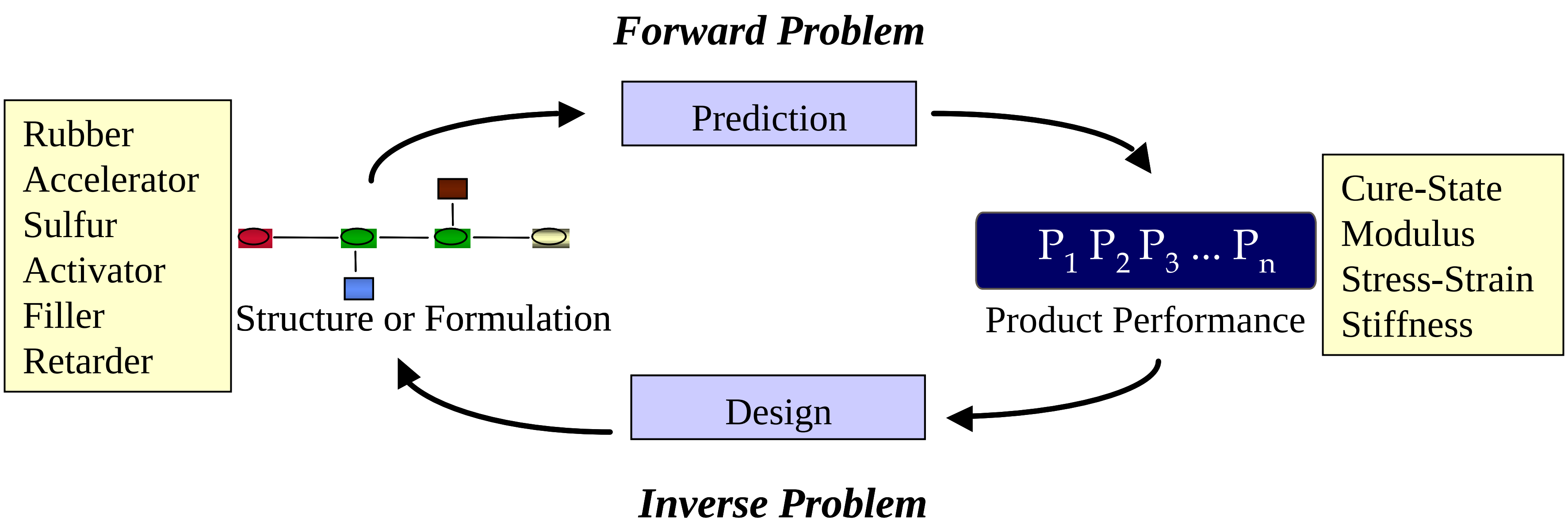
Insights into Actual Part Design

- Incorporation of Population Balance model with a finite element code to predict spatial cure profile.
- Identification of undercured and overcured regions visually.



School of Chemical Engineering, Purdue University

CAMD Framework



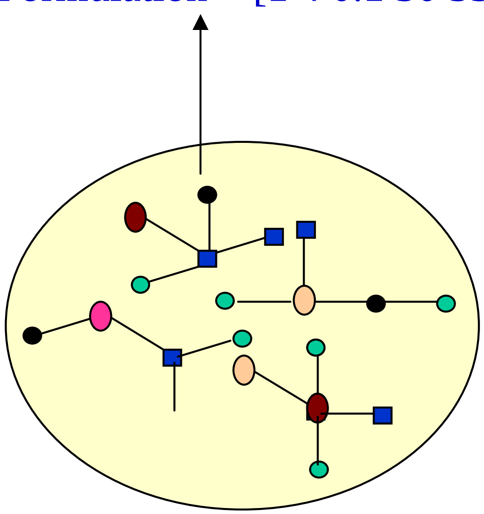
The solution of **INVERSE PROBLEM** involves searching for the optimal rubber formulation that has the desired macroscopic performance.

Genetic Algorithms (GA)

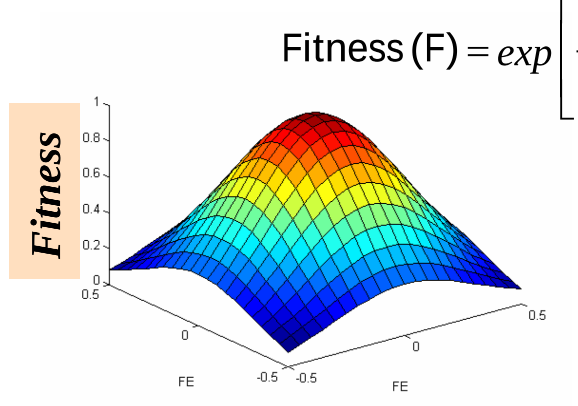
GAs are stochastic evolutionary search procedures based on the Darwinian model of natural selection

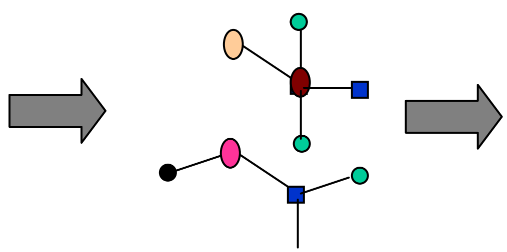
$$\text{Fitness (F)} = exp\left[-\alpha \sum \left(\frac{P_i - P_{i,\,des}}{P_{i,max} - P_{i,min}}\right)^2\right]$$

Formulation = [1 4 0.1 30 330]



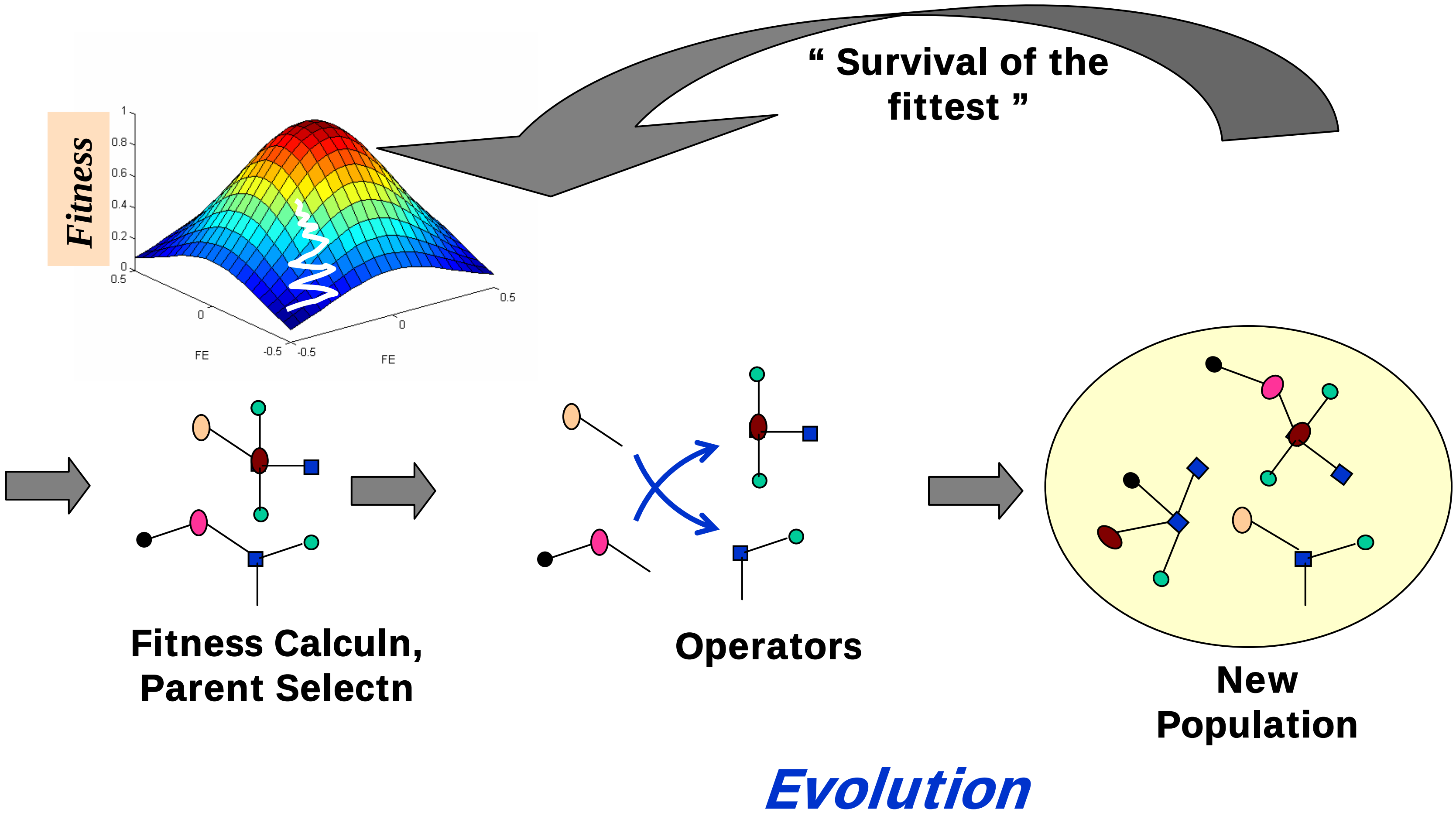
Initial Population (random)



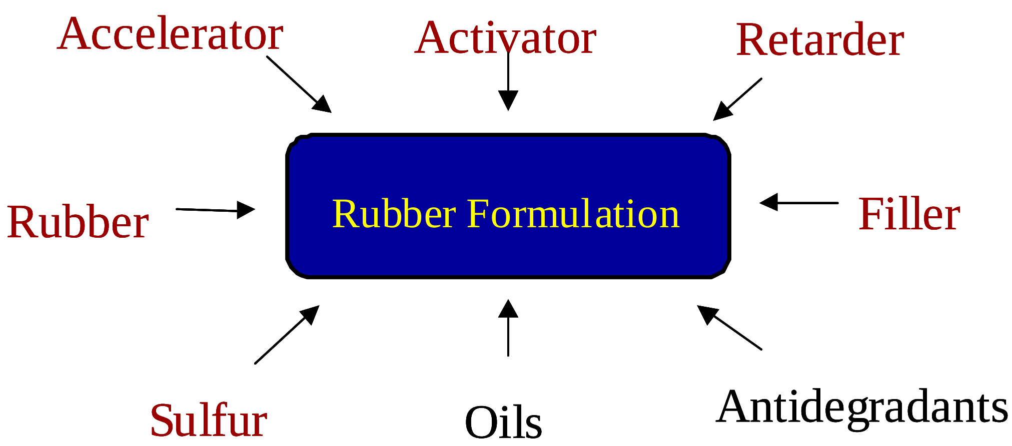


Fitness Calculn, Parent Selectn

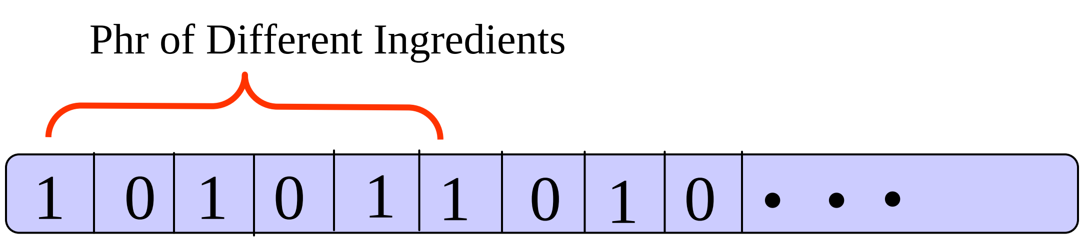
Genetic Algorithms (GA, contd)



Formulation Representation



Binary Representation



Genetic Operators

Crossover

Parent 1

| 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |

Parent 2

| 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |

Offspring 1

| 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |

Offspring 2

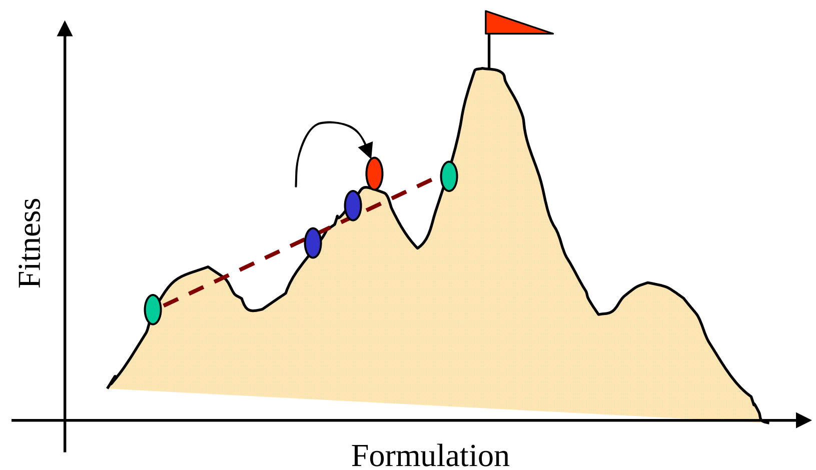
| 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 |

Mutation

| 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |

↑

| 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |



Inverse Problem (Results)

- Binary Representation
- Fitness Proportionate Selection
- Crossover Probability = 0.8, Mutation Probability = 0.2
- Population Size = 40, Number of Generations = 40, Elitism = 10%

Example 2

Desired Property

T_{max} (time to reach max cure) = 16 min

Vu_{max} (crosslink @ T_{max}) = 71.5 mol/m^3

Stress (100% Elongation) = 0.92 MPa

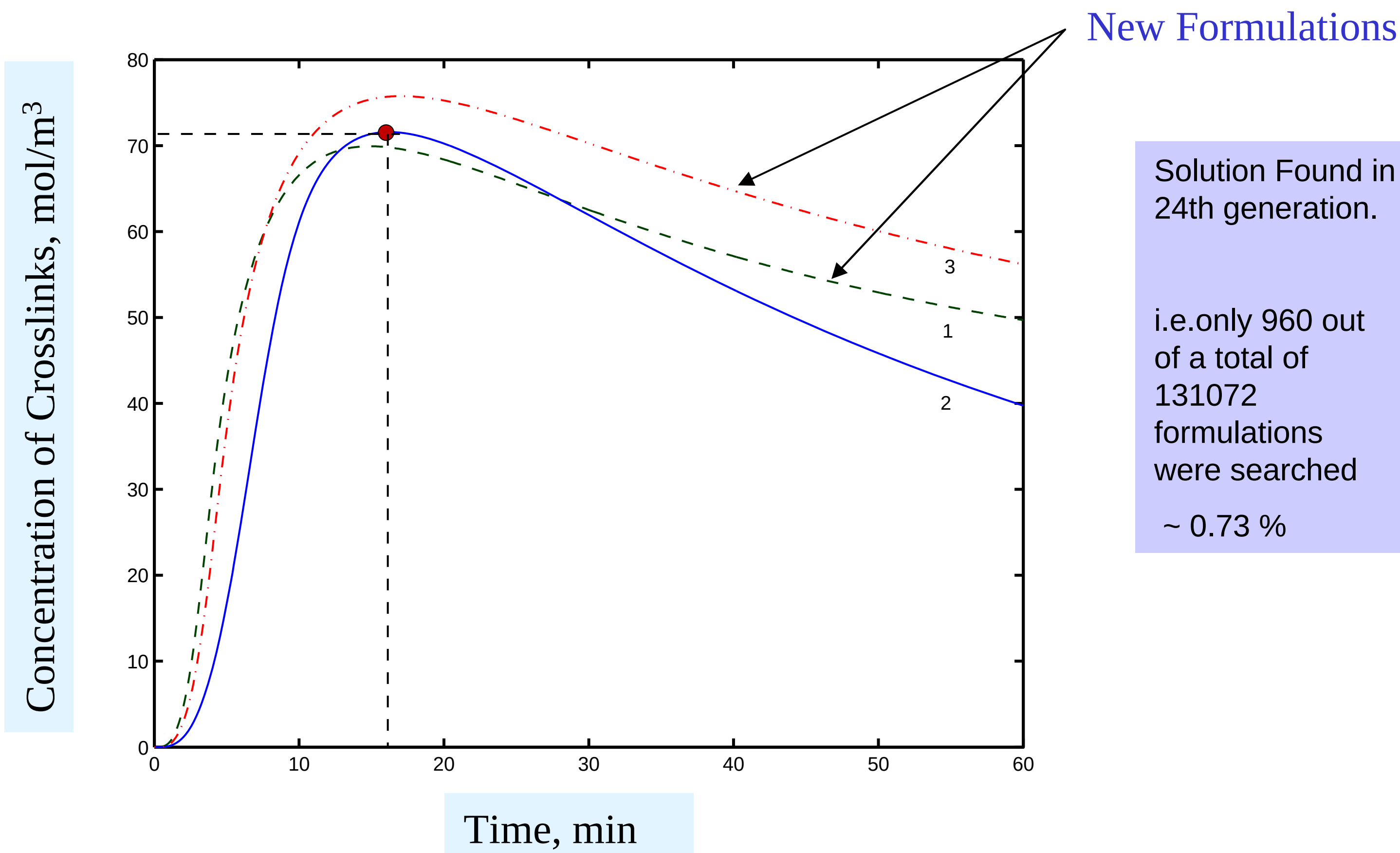
Stress (200% Elongation) = 1.65 MPa

Optimal Formulations

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
[3.00 0.75 0.10 0 315]	0.9991	17.5	65.34	0.85	1.52
[2.75 1.00 0.25 0 330]	0.9990	10.5	70.46	0.93	1.67

BEST THREE FORMULATIONS

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

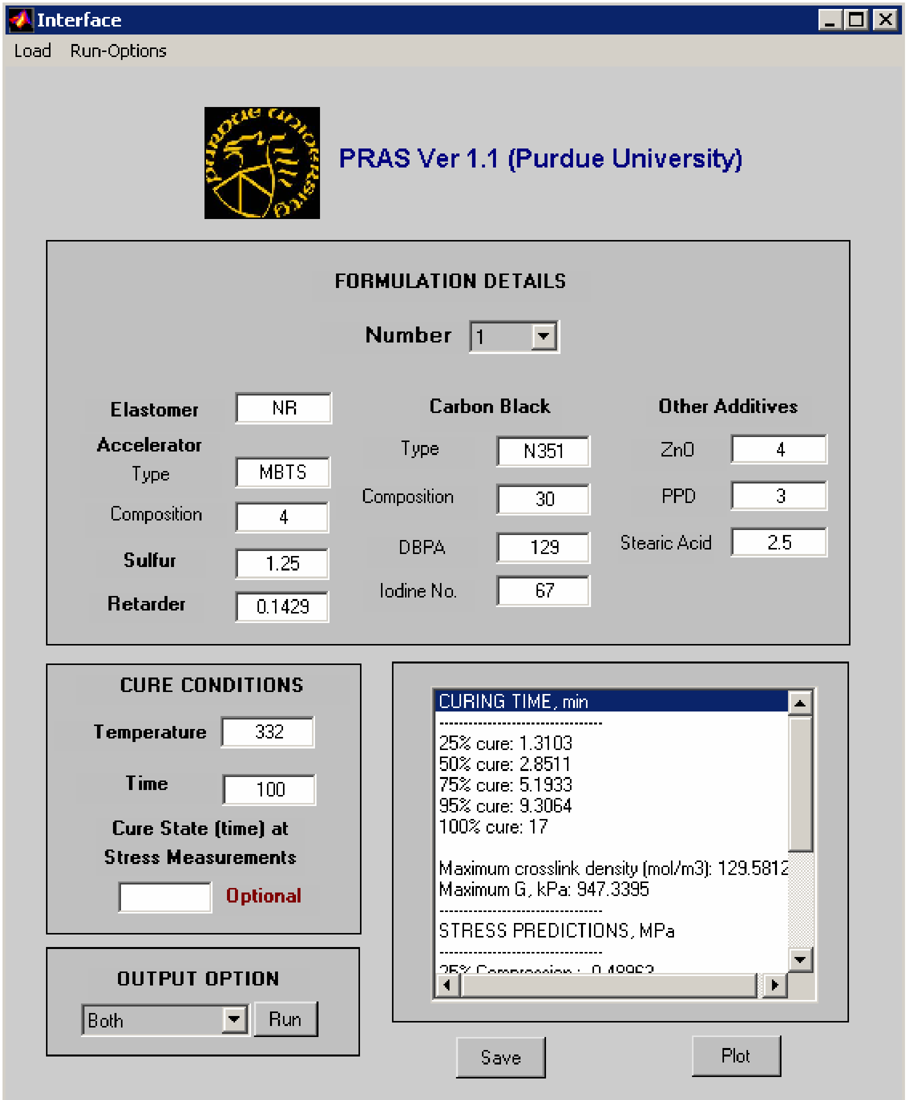


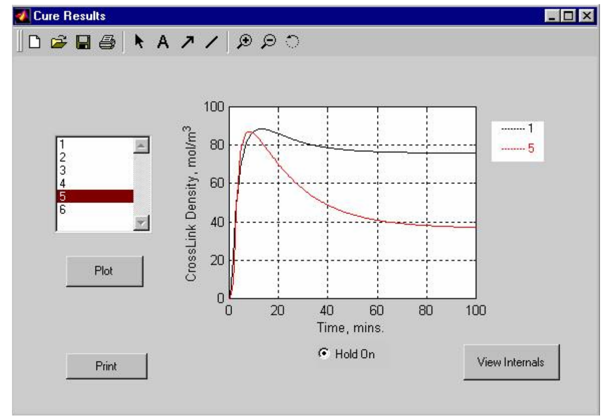
Solution Found in 24th generation.

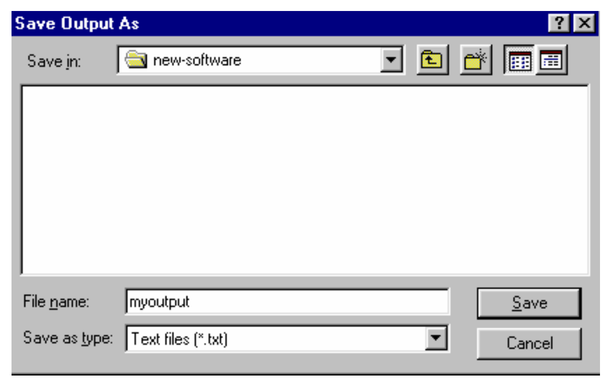
i.e.only 960 out of a total of 131072 formulations were searched

~ 0.73 %

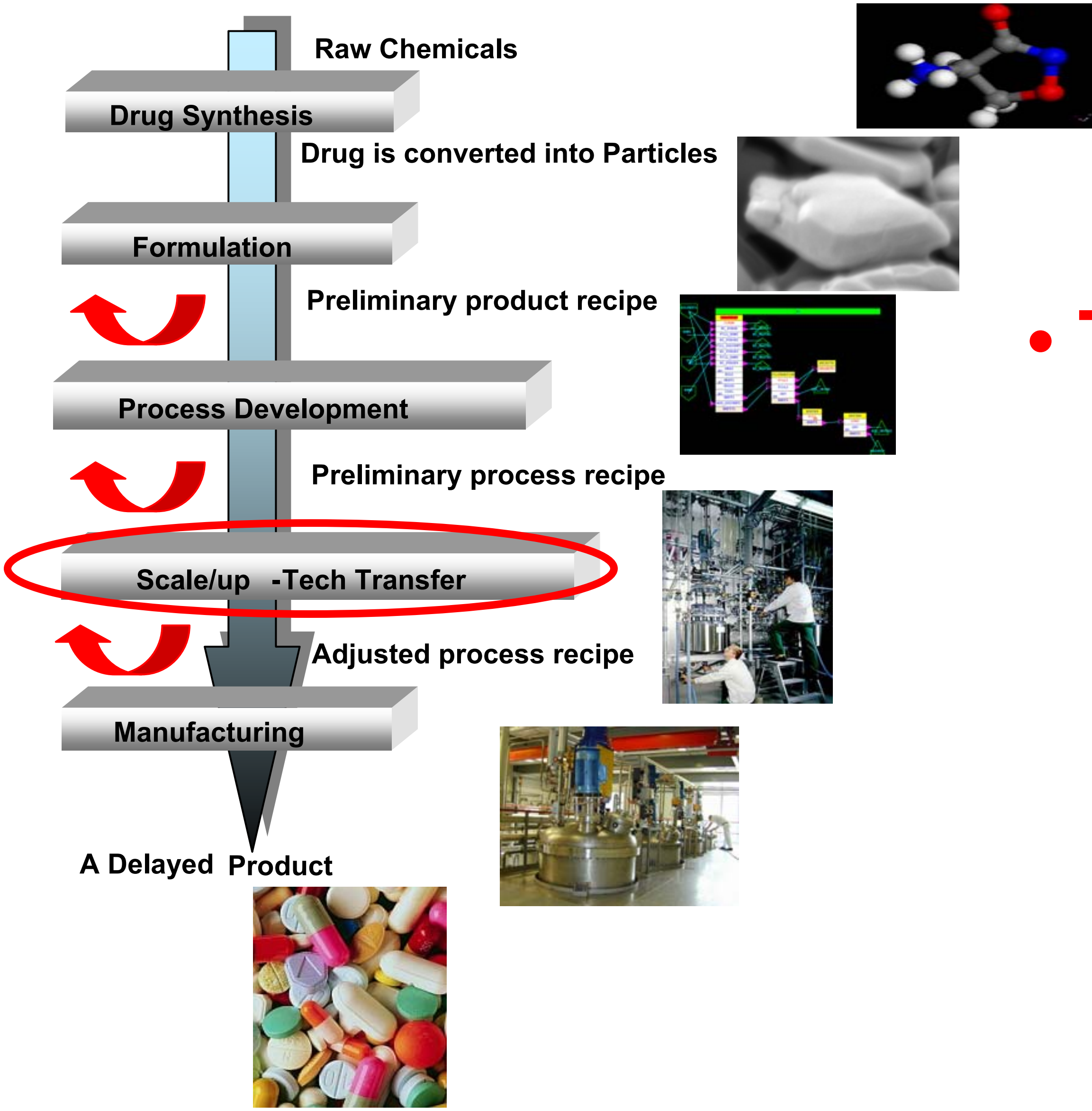
Interactive Software Used At Caterpillar on a Daily Basis







Pharmaceutical Product Development and Engineering



- **Two Products**
  - **Drugs**
  - **Documents**

THE WALL STREET JOURNAL.

WEDNESDAY, SEPTEMBER 3, 2003 - VOL. CCXLII NO. 45 - ★★★★ $1.00

Factory Shift

New Prescription For Drug Makers: Update the Plants

After Years of Neglect, Industry Focuses on Manufacturing; FDA Acts as a Catalyst

The Three-Story Blender

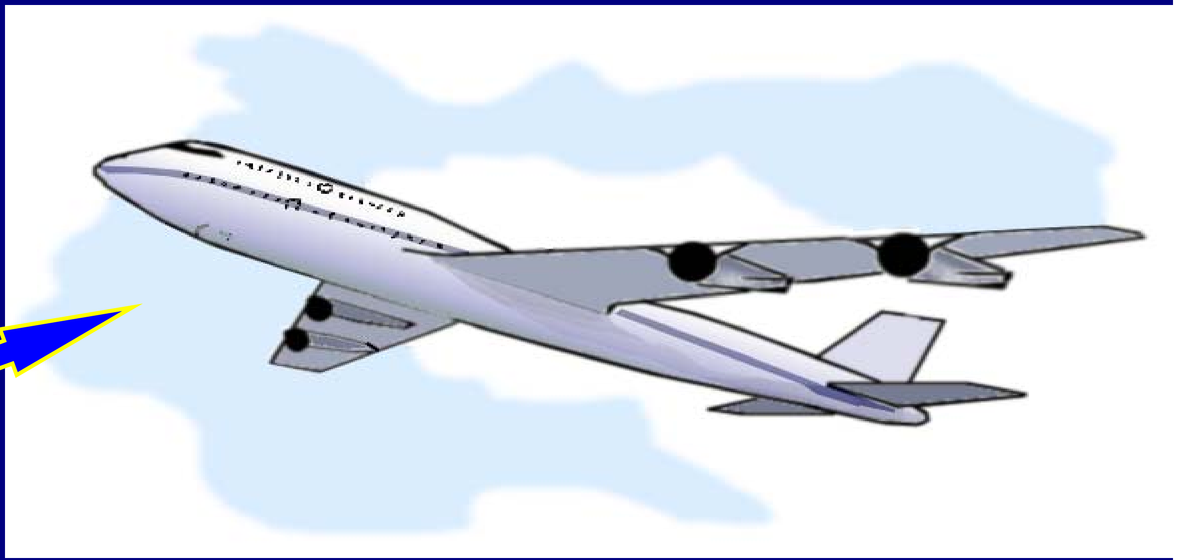
By Leila Abboud And Scott Hensley

- **Prescription drug recalls**
  - **176 in 1998**
  - **248 in 2001**
  - **354 in 2002**
- **Schering-Plough Corp. recalled 59 million asthma inhalers in 1999 and 2000 -- unknown number were shipped empty.**

•Semiconductor Manufacture	**6 σ**
•Chemicals	**5 to 5.5 σ**
•Pharma	**2.5 σ**

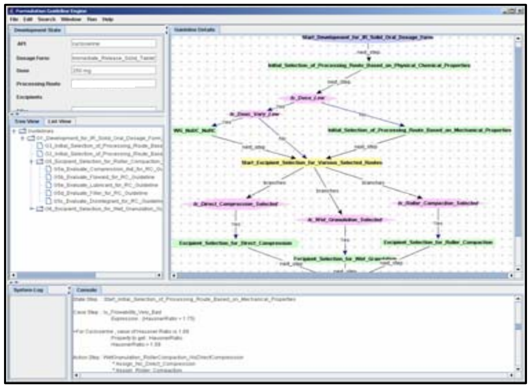
Pharma Industry's Scale-up Challenge

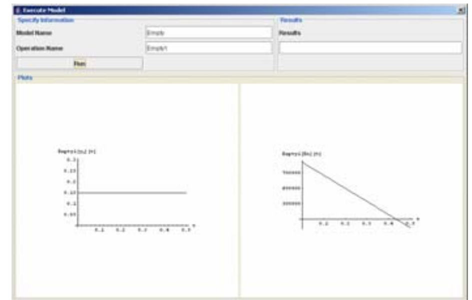
engineering



pharmaceutical

Purdue Ontology for Pharmaceutical Engineering (**POPE**)

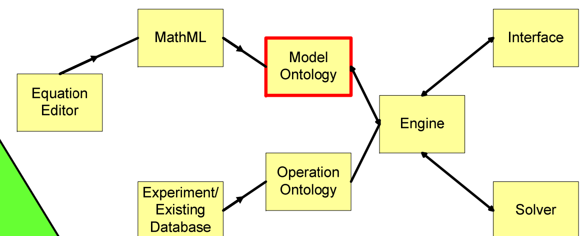




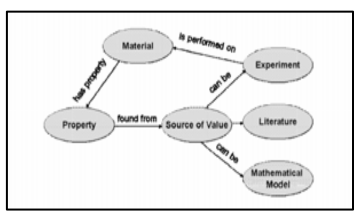
Decision trees and heuristics

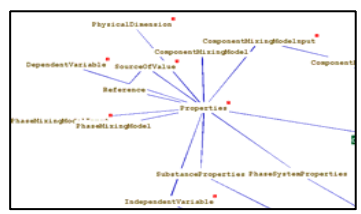
Mathematical models

Information



ModLAB
OntoMODEL

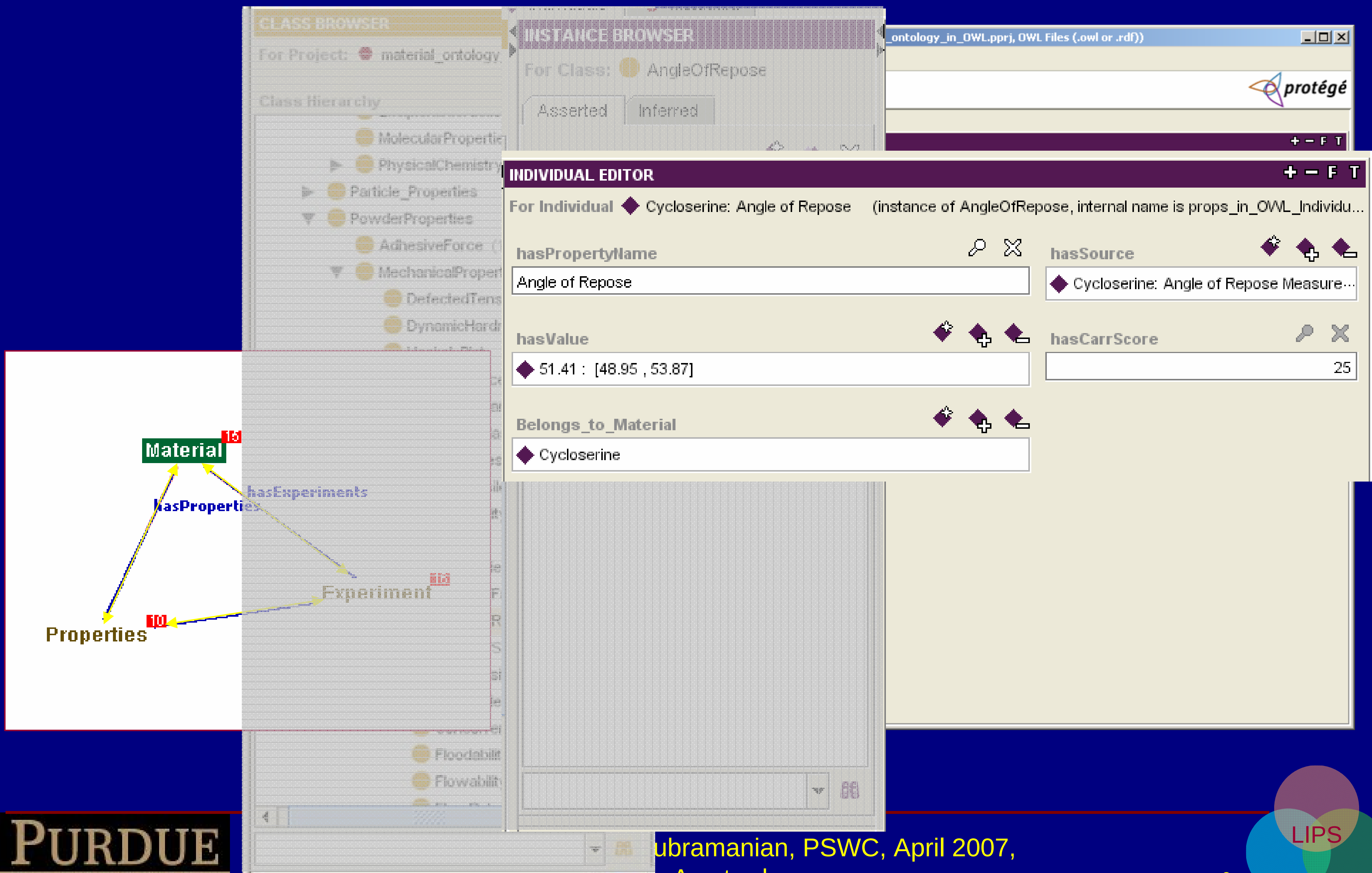




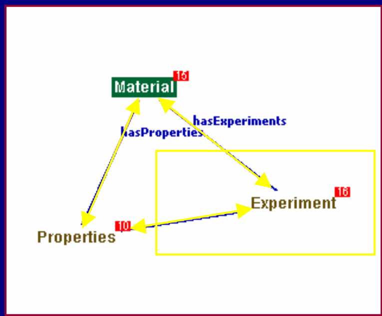
Ontologies Developed So Far in Our Project

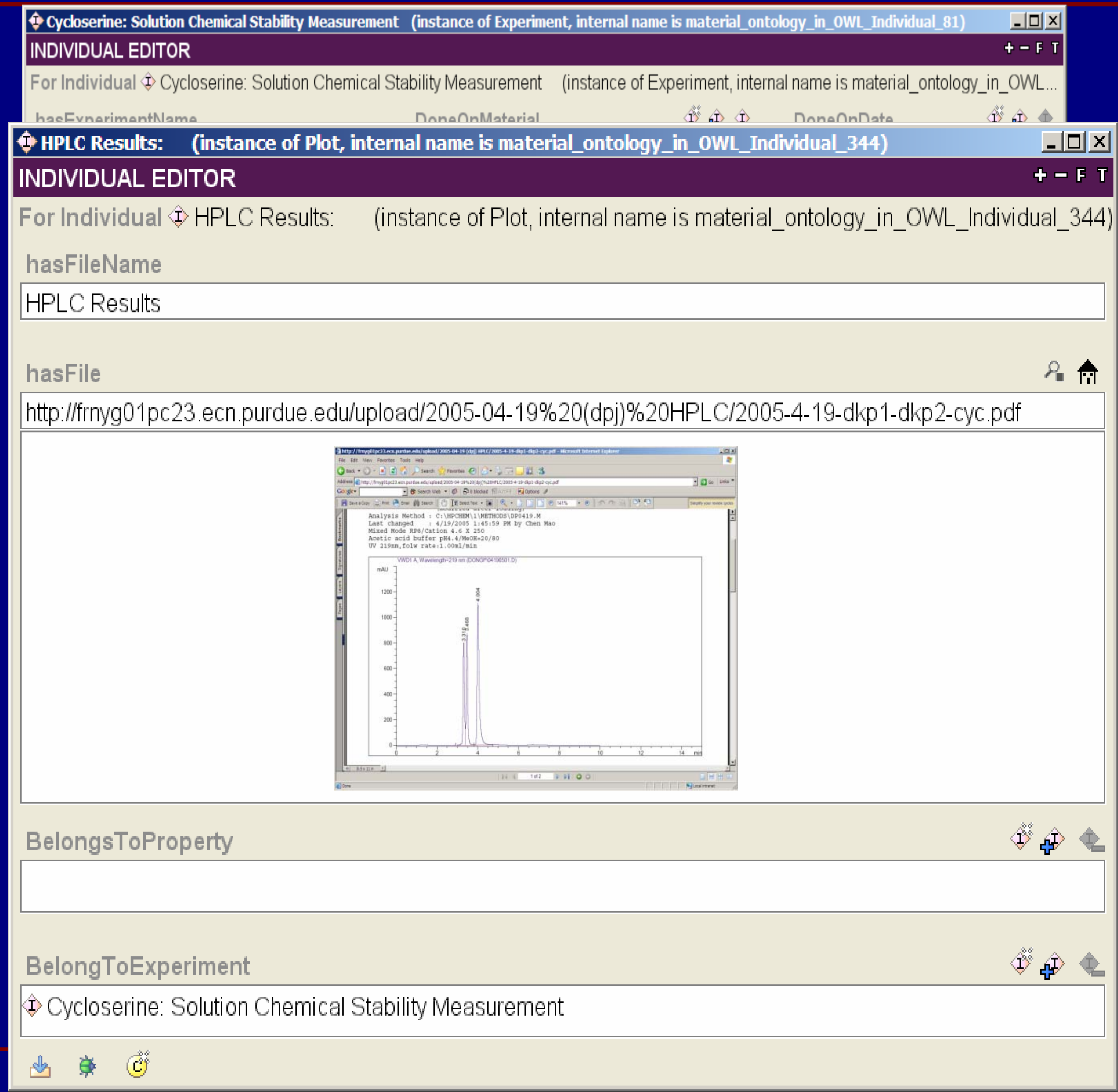
- Equipment ontology
  - Standards referenced: STEP, AP231, FIATECH
- General recipe ontology
  - RecipeElement, UnitProcedure, Operation etc.
  - Standards referenced: ISA S88, S95, OntoCAPE
- Process safety ontology
  - Deviation, Cause, Consequence etc.
- Material ontology for pharmaceutical product development
  - Flow property, Angle_of_Fall, Carrs_Index etc.
- Reaction mechanism ontology
  - Molecule, Atom, Bond, Reaction etc.
  - Referenced: Chemistry Development Kit
- Model ontology
- Guideline ontology
  - Referenced: GuideLine Interchange Format

Material Property Ontology in Protégé

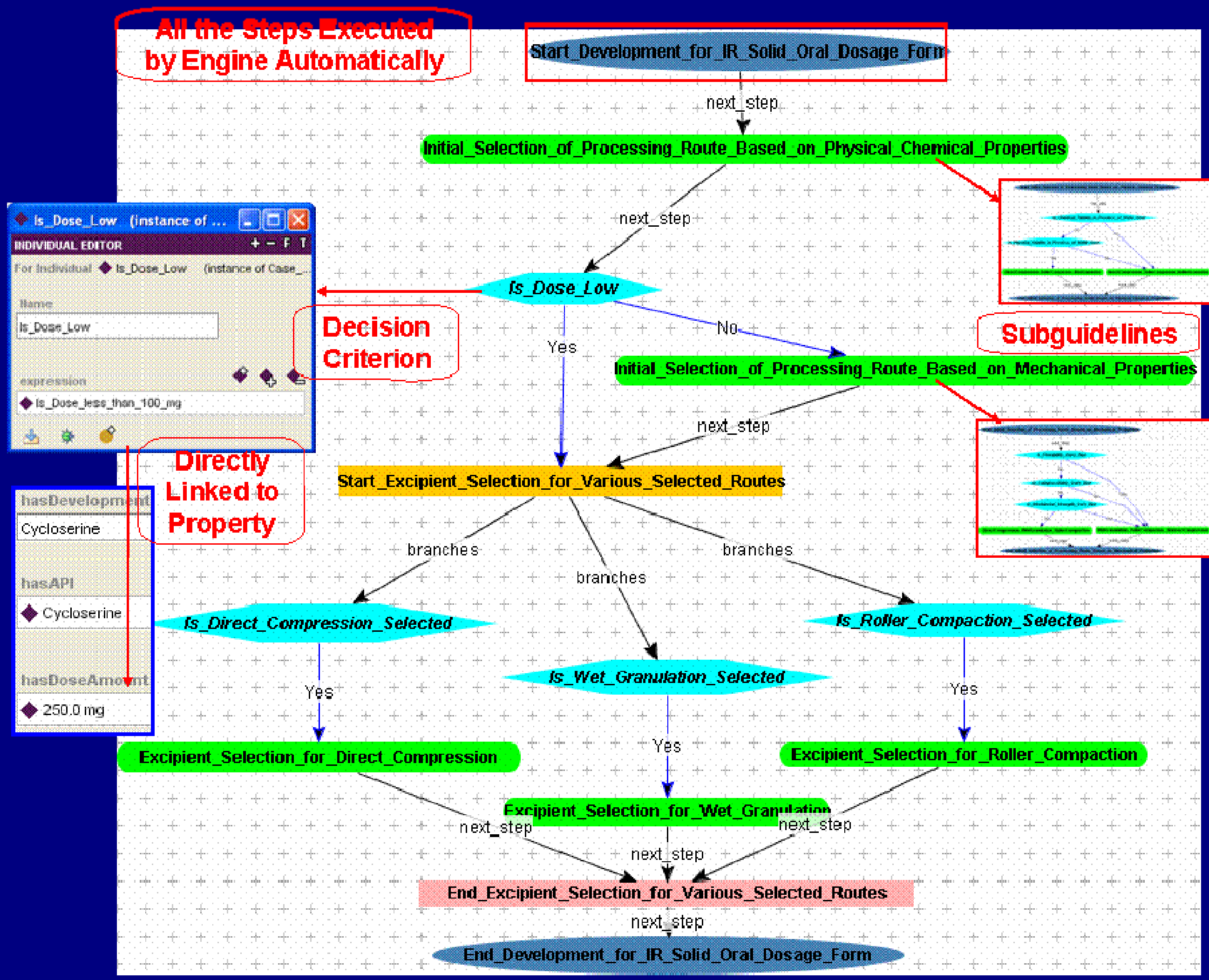


Experiment Ontology



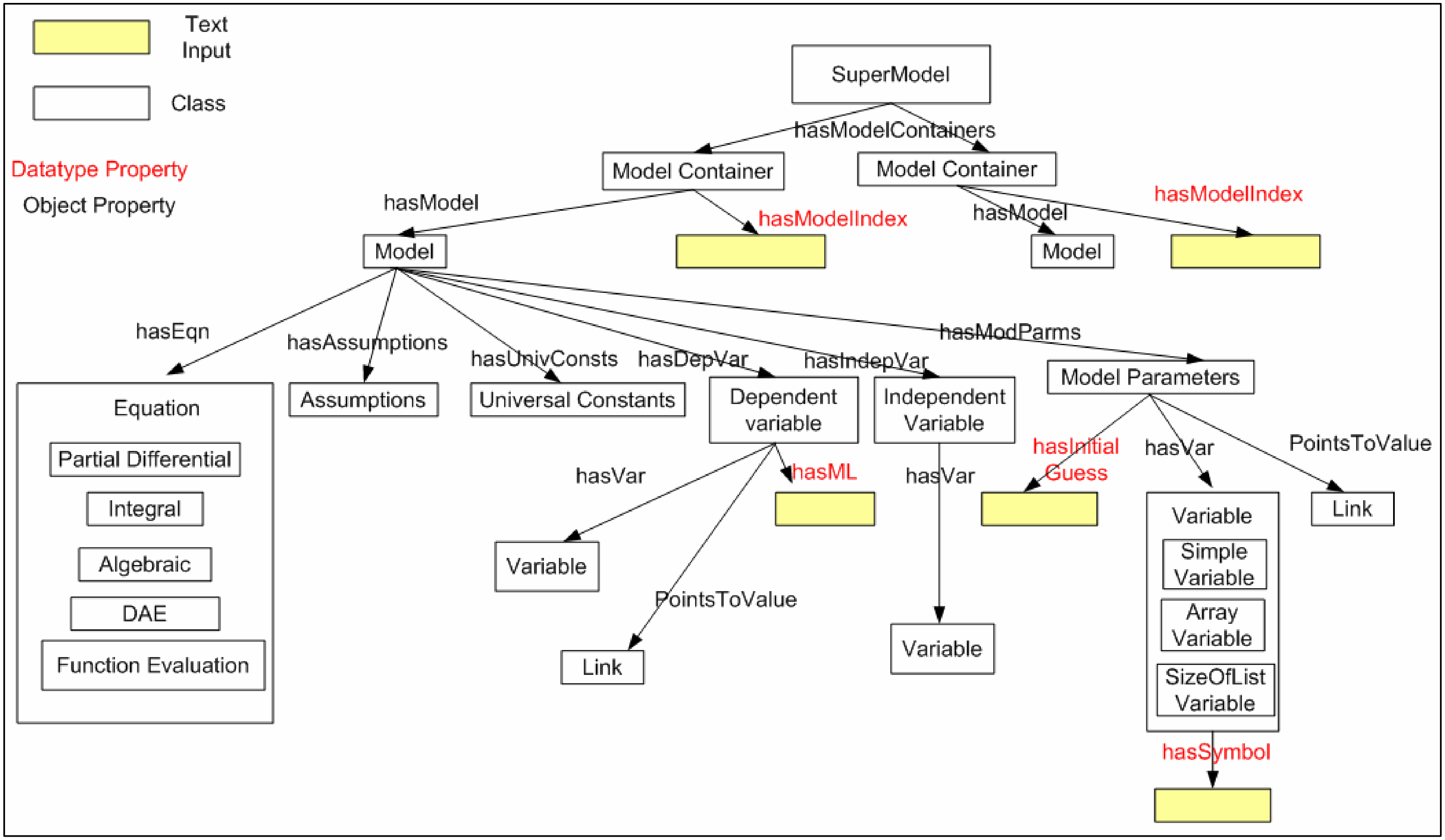


Guideline Ontology for Dosage Form Formulation

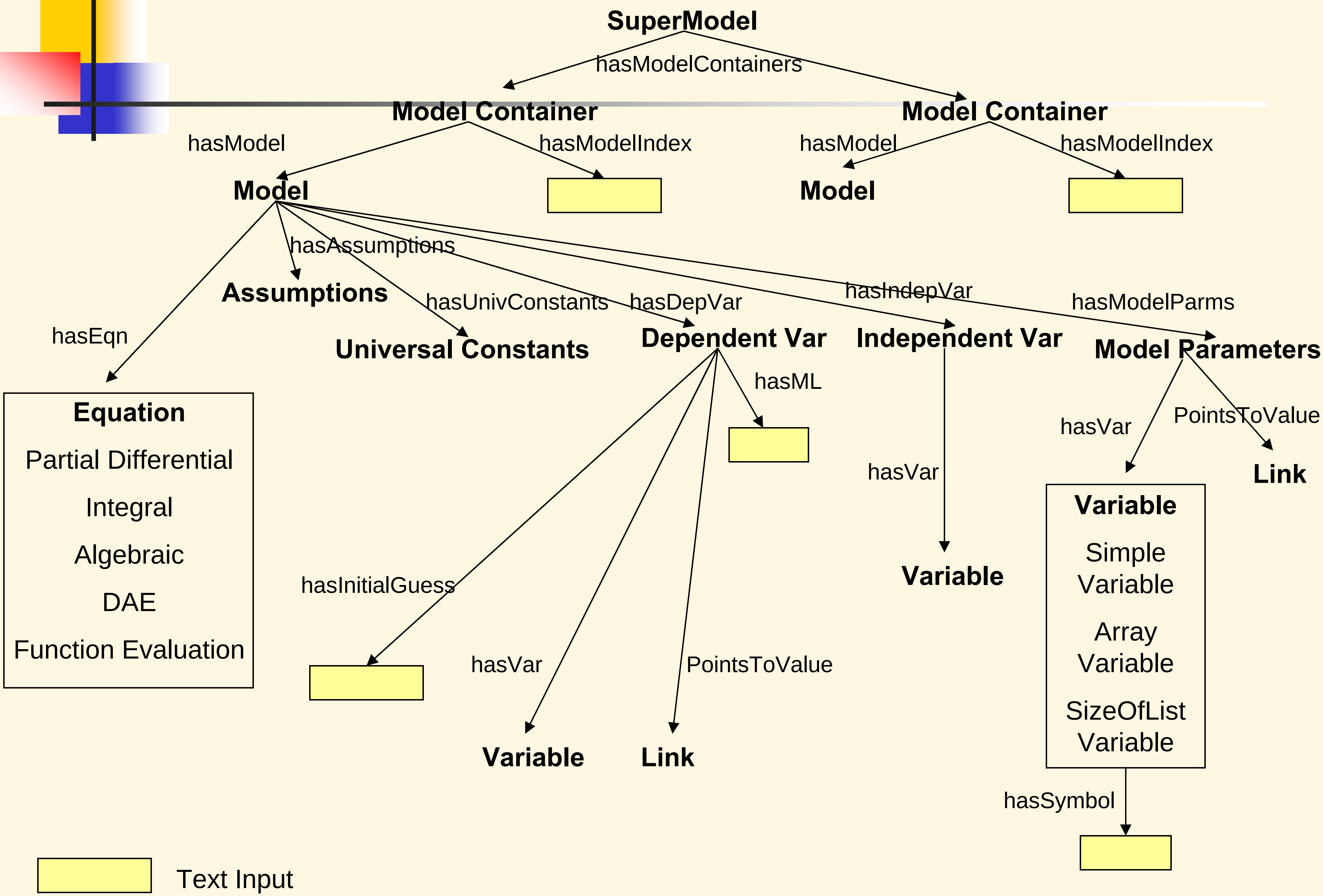


Follows GLIF (Guideline Interchange Format), A standard ontology for clinical guidelines

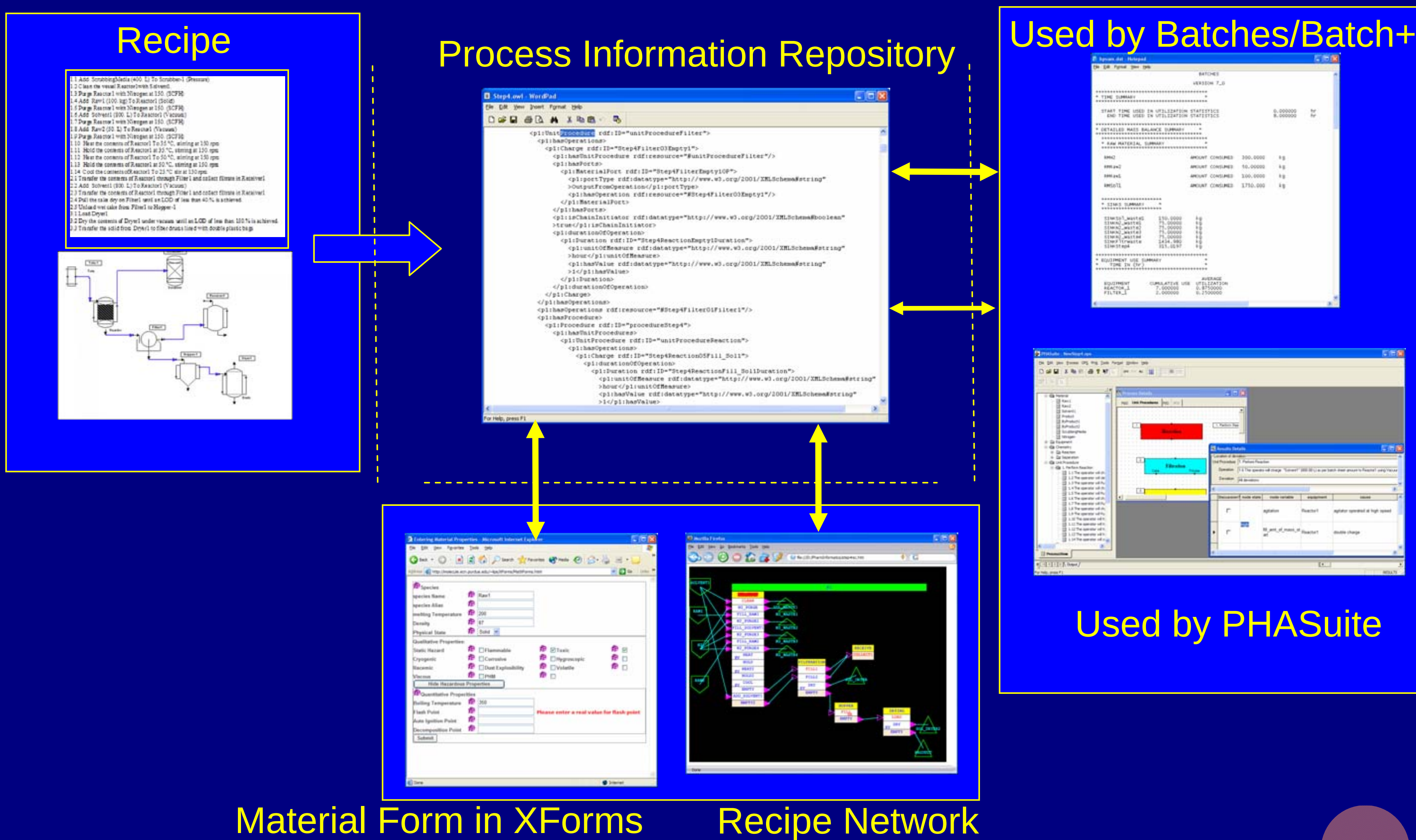
Model Ontology



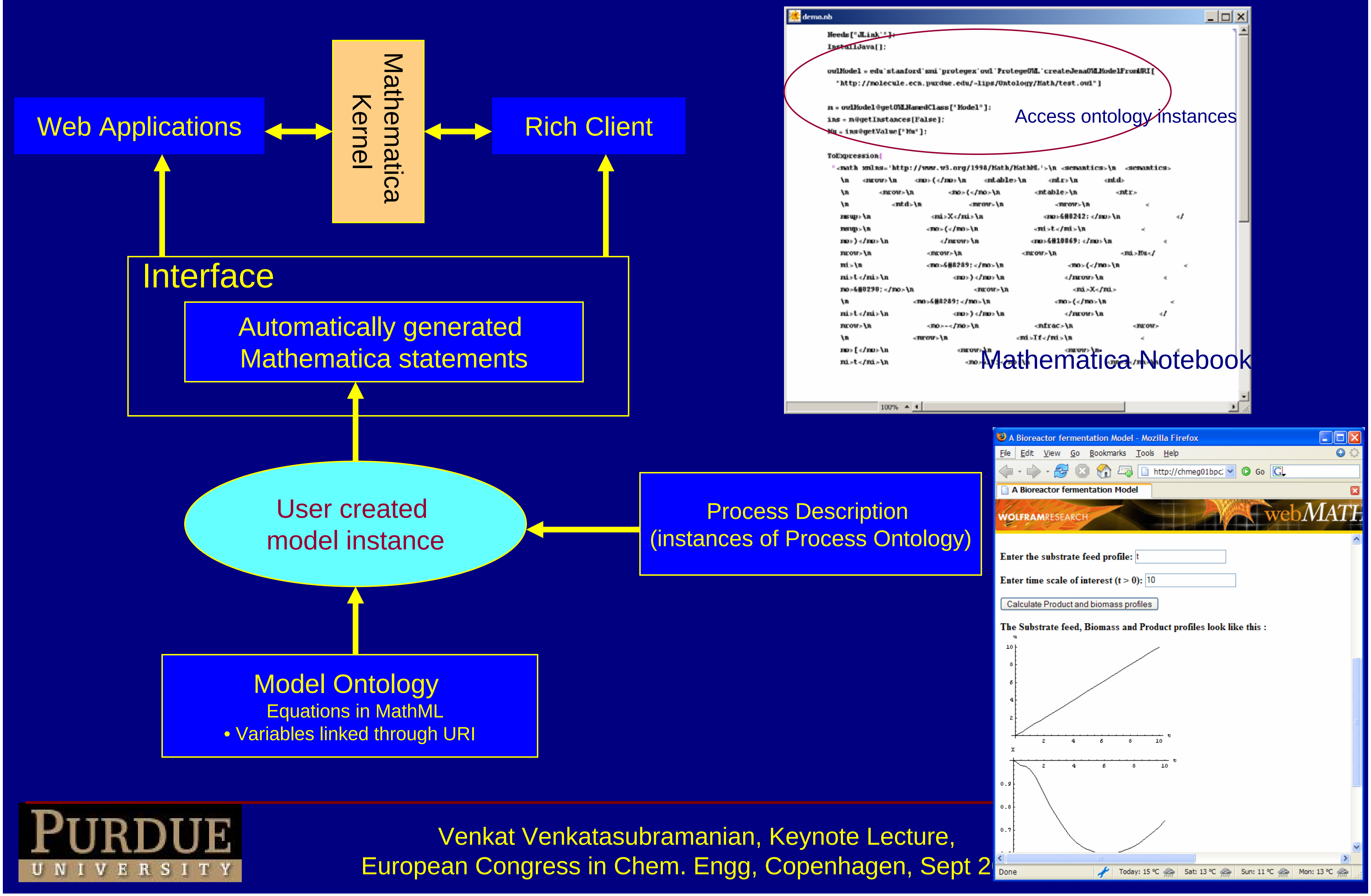
ENGINEERING RESEARCH CENTER FOR
STRUCTURED ORGANIC PARTICULATE SYSTEMS
RUTGERS UNIVERSITY
PURDUE UNIVERSITY
NEW JERSEY INSTITUTE OF TECHNOLOGY
UNIVERSITY OF PUERTO RICO AT MAYAGÜEZ


SuperModel
hasModelContainers
Model Container
Model Container
hasModel
hasModelIndex
hasModel
hasModelIndex
Model
Model
hasAssumptions
Assumptions
hasUnivConstants
hasDepVar
hasIndepVar
hasModelParms
hasEqn
Universal Constants
Dependent Var
Independent Var
Model Parameters
hasML
Equation
Partial Differential
Integral
Algebraic
DAE
Function Evaluation
hasVar
PointsToValue
Link
hasVar
Variable
Simple Variable
Array Variable
SizeOfList Variable
Variable
hasInitialGuess
hasVar
PointsToValue
Variable
Link
hasSymbol
Text Input

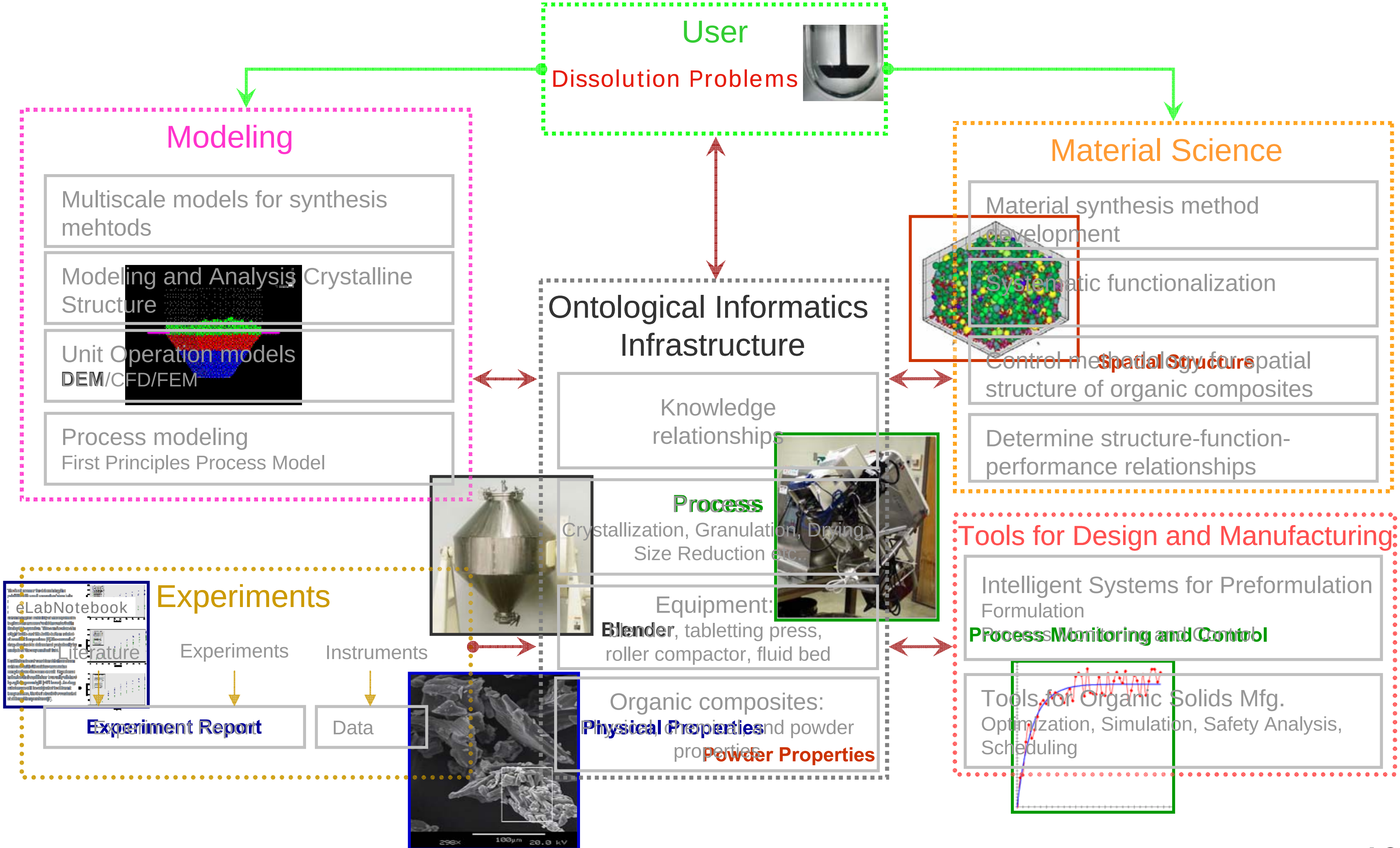
Integrated Model-based Decision Support



Integrating with Mathematical Knowledge



Cyberinfrastructure for Real-time Decision Support for Design, Control, and Optimization



Summary

- Reviewed Modeling and Informatics Challenges in Molecular Products Design and Engineering
- Need for Cyberinfrastructure Concepts, Methods, and Tools
- Ontological approach for information and knowledge modeling
  - Beyond ERP, SAP, Oracle, Expert Systems etc.
  - This is not just programming! Nor can it be done by CS folks alone!
- POPE: Purdue Ontology for Pharmaceutical Engineering
  - Conceptual foundation for next the generation, integrated, decision support tools environment
- This is only a beginning….Miles to go before we sleep….
- Huge opportunities for Process/Product Systems Engineers

Thank You for Your Attention!

Any Questions?