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

- **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

Intake Valve and Manifold
About 1000 Rubber Parts in Failure Critical Functions

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

Reliability and Warranty Problems
ExxonMobil’s Grand Challenge: Catalyst Design via Combinatorial Chemistry

Synthesis Characterization

Fundamental Chemistry

Reaction kinetics

Performance measures

Quantum calculations

d band center
Electronegativity

Electronic/Structural descriptors

Catalyst Structures

Electron density maps

Electronic/Structural descriptors

Rate/Selectivity vs. time

Reaction network

Electronic/Structural descriptors

Quantum calculations

Catalyst Structures
Pharmaceutical Product Development and Engineering

- Two Products
- Drugs
- Documents
Given a set of desired performance specifications, how does one rationally and efficiently identify the optimal product structures and formulations?

$200+\text{ Billion/yr industry: Major Opportunity for ChEs}$

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 $rel : R \to C \times C$ that relates concepts non-taxonomically. The function $dom : R \to C$ with $dom(R) := \Pi_1(rel(R))$ gives the domain of $R$, and range: $R \to C$ 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

**Concepts**
- Material
- Properties
- Roles
- Composition Range
- Filler
- Flow aid
- Lubricant
- Flow aid
- Lubricant
- Minimum
- Maximum
- Average

**Properties**
- Contact angle
- Moisture content

**Web Ontology Language: OWL**
Catalyst Design Problem – The Grand Challenge

Modeling at different Length and Time Scales

- **Quantum Mechanics**
- **DFT**
- **Atomistic Simulation**
- **Molec. Dyn.**
- **Statistical Mechanics**
- **TST**
- **Continuum Models**
- **Transport Models**
- **Rate Consts.**
- **Overall Rxn Rates**
- **Conc. Profiles**

Microscopic

Mesoscopic

Macroscopic

- **Elect. Struct.**
- **Molec. Struct.**
- **Thermo. Props.**

Length and Time Scales:
- $1 \text{Å}$
- $100 \text{Å}$
- $10 \text{µm}$
- $1 \text{cm}$
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

Hydride transfer

- Paraffin
- Olefin
- Poly-olefin
- Cyclic-olefin
- Active site
Ontological Informatics: Modeling Super Highway

- 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

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


Model Refinement Procedure

Crude models

Initial data

Model screening

Infer new knowledge about models (mechanism) and data

Model discrimination

Model enhancement

Planning new experiments

Add/Delete models

Enhanced models

Richer data set

Data analysis

Model analysis

Add/Delete data
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

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
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+

**Mathematical Equations**

\[ \frac{dC_A}{dt} = -k_1C_A \]
\[ \frac{dC_B}{dt} = k_3C_A + k_4D - k_5B \]
\[ \theta_i + \theta_t + \theta_c = 1 \]

100’s of DAE’s

---


Results

Over prediction of $C_2s$ and under prediction of aromatics

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

<table>
<thead>
<tr>
<th>S.no.</th>
<th>Name</th>
<th>Esposito &amp; Floudas (2000)</th>
<th>GA based procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Global optimizer</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>v</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Objective</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CPU s reported min max</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>v</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Objective</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Time taken (CPU s)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scaled CPU s Time taken</td>
<td>(%) Saving</td>
</tr>
<tr>
<td>1</td>
<td>First order irreversible chain reaction</td>
<td>5.0035</td>
<td>1.18584x10^6 2.92 20.05 5.0122 0.9976 9.25418x10^6 0.24 0.76 (74)</td>
</tr>
<tr>
<td>2a</td>
<td>First order reversible chain reaction</td>
<td>4.0000 2.0000 40.013 20.007</td>
<td>1.8897x10^-7 568.44 6164.3 3.9813 1.9759 39.787 19.887 8.1313x10^-6 0.45 1.43 (100)</td>
</tr>
<tr>
<td>2b</td>
<td>Same as 2a but with error in data</td>
<td>4.021 2.052 39.45 19.62</td>
<td>1.586x10^-3 272.91 1899.82 4.001 2.027 39.22 19.50 1.589x10^-3 0.5 1.59 (99)</td>
</tr>
<tr>
<td>3</td>
<td>Catalytic cracking of gas oil</td>
<td>12.214 7.9798 2.2216</td>
<td>2.65567x10^-3 79.77 1185 12.246 7.9614 2.2351 2.68017x10^-3 0.38 1.21 (98)</td>
</tr>
<tr>
<td>4</td>
<td>Bellman’s problem</td>
<td>4.5704x10^-6 2.7845x10^-4</td>
<td>22.03094 36.16 12222 4.5815x10^-6 2.7899x10^-4 22.2885 3.54 11.26 (69)</td>
</tr>
<tr>
<td>5</td>
<td>Methanol-to-hydrocarbon process</td>
<td>5.1981 1.2112 0 0</td>
<td>0.10652 1361.8 19125 5.2212 1.2320 1.6363x10^-3 4.0059x10^-12 0.004665 0.10586 2.08 6.61 (100)</td>
</tr>
<tr>
<td>6</td>
<td>Lotka-Volterra Problem</td>
<td>3.2434 0.9209 1.24924x10^-3</td>
<td>367.26 9689.67 3.1434 0.9583 2.39784x10^-3 0.56 1.78 (100)</td>
</tr>
</tbody>
</table>

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

- **High Throughput Experiments**
  - FTIR
  - GC

- **Reaction Modeling Suite**
  - Automatic processing of data

- **Chemistry Rules**
  - Reactions
  - Lumping

- **Model Refinement**
  - Rules & Features mapping
  - Suggestions for location of new rules

- **New HTE for Model Discrimination via Experimentation**
  - Experiments in new composition regimes
  - Measure critical variables identified

- **Performance Curves**
  - Rate/Selectivity vs. time

- **Feature Extractor/Analyzer**
Catalyst Design Challenge

Target Catalyst

Catalyst Library

HTE

Model Revision

Compare Performance

Target Catalyst Performance

Catalyst Library

Statistics/Neural-Nets

Physical Model

Hybrid Model

Forward Model

Inverse Model

Recombination

Selection

Genetic Algorithms

A + S \xrightarrow{k_1} A-S

C + R \xrightarrow{k_2} D

A-S + D \xrightarrow{k_3} F

Statistical Analyzer

Pseudo Global Optimizer

Feature Extractor

Pseudo English Rule Compiler

Statistical Analyzer

Al/Systems Tools

Reaction Modeling Suite

Rate/Selectivity

time

Rate/Selectivity

time

Catalyst Performance

Catalyst Performance

Kinetics

A+ S A- S

k1

C R D

k2

A-S +D k3

F
LUBRIZOL: Fuel Additive Design

- **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

- Additive Component that binds to deposit “precursors”
- Additive Component that keeps it in solution
- Fuel + Oil (Liquid + Vapor Mixture)
- Fuel-Additive Molecules
- High-Boiling Deposit Forming Precursors

- Additive Component that keeps it in solution
- Fuel + Oil (Liquid + Vapor Mixture)
Component “length” directly indicative of stability of additive

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

Primary Model

Ab Initio Methods
Group Contribution
Material/Molecule Formulation
Topological Techniques
Molecular Mechanics

Structural Descriptors

Secondary Model
Regression
Neural Networks
Pattern Recognition

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

Forward Problem Approaches

Fundamental Physics/Chemistry

\[ \frac{\partial (\rho v)}{\partial t} + v \cdot \nabla (\rho v) = -\nabla P + \nabla \cdot (\tau) \]

\[ \frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho v) \]

Data-driven/empirical models

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

Expert Rules/Heuristics

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

Additive A
Additive B
Intake-Valve Deposit

Amount of Active Additive

$\tau_0$  $\tau_2$  $\tau_1$  $\tau_N$

Time
Computer-Aided Design of Fuel-Additives

Hybrid Model

Statistical/Neural-Net Correlation

Physical Model

Intake Valve Deposit

Additive Performance

Additive Structure
Fuel Property
Engine Conditions

Genetic Algorithms

Recombination
Selection
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

1. Initial Population
   Molecular Designs

2. Calculate Fitness

3. Select Parent(s)

4. Select Operator
   - Crossover
   - Mutation

5. Apply Operator

6. Create New Population
Genetic Algorithms (GA)

GAs are stochastic evolutionary search procedures based on the Darwinian model of natural selection.
Genetic Operators: Crossover

- Single-point Crossover
Main-chain and Side-chain Mutation

- **Mainchain Mutation**
  - **Parent:** \[ \begin{array}{c}
  H \\
  C \\
  H \\
  H
  \end{array} \quad \begin{array}{c}
  H \\
  C \\
  H \\
  C
  \end{array} \quad \begin{array}{c}
  H \\
  C \\
  H \\
  H
  \end{array} \_n \]
  - Mainchain Mutation
  - Replace \(-\text{CH}_2-\) by \(\text{CH}_2\)
  - **Offspring:** \[ \begin{array}{c}
  H \\
  C \\
  H \\
  H
  \end{array} \quad \begin{array}{c}
  H \\
  C \\
  H \\
  C
  \end{array} \quad \begin{array}{c}
  H \\
  C \\
  H \\
  H
  \end{array} \_n \]

- **Sidechain Mutation**
  - **Parent:** \[ \begin{array}{c}
  H \\
  C \\
  H \\
  C
  \end{array} \_n \]
  - Sidechain Mutation
  - Replace \(-\text{F}\) by \(-\text{F}\)
  - **Offspring:** \[ \begin{array}{c}
  H \\
  C \\
  H \\
  C
  \end{array} \_n \]
Fitness Function for GA-based CAMD

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

Fractional Error (FE) = \[
\frac{P_i - P}{P}
\]

\[\alpha = 0.0001\]

\[\alpha = 0.0005\]

\[\alpha = 0.001\]

\[\alpha = 0.01\]
Polymer Design Case Study

• Base Groups

17 Main-chain Groups
- >C< - S- -SO₂- -O- -NH-
- O -O-C- -O-C-O- -C-O-C-
- O -O-C-NH- -O-C-NH- -NH-C-NH-

15 Side-chain Groups
- H -CH₃ -C₂H₅ -nC₃H₇ -iC₃H₇
tC₄H₉ -F -Cl -Br -OH
- OCH₃ -O-C-CH₃ -O-C-CH₃ -CN

• Target Properties

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

<table>
<thead>
<tr>
<th>Target Polymer</th>
<th>Density (gm/cm³)</th>
<th>Glass Transition (K)</th>
<th>Thermal Expansion 1/K</th>
<th>Heat Capacity J/kg K</th>
<th>Bulk Modulus K N/m³</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP1: [C-C=C-C CH₃]ₙ</td>
<td>2.96x10⁻⁴</td>
<td>350.75</td>
<td>2.96x10⁻⁴</td>
<td>1152.67</td>
<td>5.18x10⁹</td>
</tr>
<tr>
<td>TP2: [C-C=C-C CH₃]ₙ</td>
<td>1.34</td>
<td>225.24</td>
<td>2.81x10⁻⁴</td>
<td>1377.82</td>
<td>2.51x10⁹</td>
</tr>
<tr>
<td>TP3: [C-C=C-C CH₃]ₙ</td>
<td>1.18</td>
<td>420.83</td>
<td>2.90x10⁻⁴</td>
<td>1135.10</td>
<td>5.40x10⁹</td>
</tr>
<tr>
<td>TP4: [C-C=C-C CH₃]ₙ</td>
<td>1.21</td>
<td>406.83</td>
<td>2.90x10⁻⁴</td>
<td>1073.96</td>
<td>5.39x10⁹</td>
</tr>
<tr>
<td>TP5: [C-C=C-C CH₃]ₙ</td>
<td>1.19</td>
<td>472.00</td>
<td>2.89x10⁻⁴</td>
<td>995.95</td>
<td>5.31x10⁹</td>
</tr>
<tr>
<td>TP6: [C-C=C-C CH₃]ₙ</td>
<td>1.28</td>
<td>421.12</td>
<td>2.90x10⁻⁴</td>
<td>1016.55</td>
<td>6.12x10⁹</td>
</tr>
<tr>
<td>TP7: [C-C=C-C CH₃]ₙ</td>
<td>1.06</td>
<td>322.55</td>
<td>2.98x10⁻⁴</td>
<td>1455.90</td>
<td>3.85x10⁹</td>
</tr>
</tbody>
</table>
## Polymer Design Case Study: Results

<table>
<thead>
<tr>
<th>Target Polymer</th>
<th>Random MC, SC</th>
<th>Random MC, SC Feasibility Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Chemical Structure" /></td>
<td>12% 184 282</td>
<td>60% 240 213</td>
</tr>
<tr>
<td><img src="image2" alt="Chemical Structure" /></td>
<td>36% 411 6</td>
<td>48% 522 6</td>
</tr>
<tr>
<td><img src="image3" alt="Chemical Structure" /></td>
<td>0% - 163</td>
<td>12% 193 74</td>
</tr>
<tr>
<td><img src="image4" alt="Chemical Structure" /></td>
<td>0% - 861</td>
<td>0% - 589</td>
</tr>
<tr>
<td><img src="image5" alt="Chemical Structure" /></td>
<td>32% 400 175</td>
<td>48% 232 142</td>
</tr>
<tr>
<td><img src="image6" alt="Chemical Structure" /></td>
<td>8% 548 199</td>
<td>32% 632 168</td>
</tr>
<tr>
<td><img src="image7" alt="Chemical Structure" /></td>
<td>100% 61 217</td>
<td>100% 64 198</td>
</tr>
<tr>
<td><img src="image8" alt="Chemical Structure" /></td>
<td>68% 210 162</td>
<td>88% 109 161</td>
</tr>
<tr>
<td><img src="image9" alt="Chemical Structure" /></td>
<td>8% 382 144</td>
<td>4% 868 70</td>
</tr>
</tbody>
</table>

*Legend: Success Rate; Average Conv. Generation; Number with fitness >0.99*
## Near Optimal Designs

<table>
<thead>
<tr>
<th>Polymer Design</th>
<th>Overall Error</th>
<th>Fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Target Polymer: #4</strong></td>
<td>0%</td>
<td>1.0</td>
</tr>
</tbody>
</table>
| \[
\begin{array}{c}
\text{CH}_3 \\
\text{S} \\
\text{O} \\
\text{C} \\
\text{CH}_3
\end{array}
\] \_n | | |
| **Case 1: Standard GA** | 0.74% | 0.995 |
| \[
\begin{array}{c}
\text{CH}_3 \\
\text{H} \\
\text{O} \\
\text{C} \\
\text{H}
\end{array}
\] \_n | | |
| \[
\begin{array}{c}
\text{CH}_3 \\
\text{C} \\
\text{O} \\
\text{O} \\
\text{O}
\end{array}
\] \_n | | |
| **Case 2: Modified GA, Stability** | 0.25% | 0.999 |
| \[
\begin{array}{c}
\text{CH}_3 \\
\text{O} \\
\text{H}
\end{array}
\] \_n | | |
| \[
\begin{array}{c}
\text{CH}_3 \\
\text{O} \\
\text{C}
\end{array}
\] \_n | | |
| **Case 3: Modified GA, Stability & Complexity** | 0.21% | 0.999 |
| \[
\begin{array}{c}
\text{CH}_3 \\
\text{O} \\
\text{O}
\end{array}
\] \_n | | |
| \[
\begin{array}{c}
\text{CH}_3 \\
\text{O} \\
\text{C}
\end{array}
\] \_n | | |
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

Objective
- IVD_{desired} > IVD_{limit}
- Maximize Solubility

Initial Population
- Random
- Knowledge Based

Calculate Fitness (F)
- Structure --- Hybrid Model
- Fitness <---- Additive Performance

Termination Criteria
- Last Generation ?
- Fitness = 1.0 ?

New Generation

Elitist Policy
- Retain k best in population

Select Parent(s)
- Fitness Proportionate

Select Operator
- Probabilistic
- Crossover Mutation

Apply Operator

No

Start

Yes

Stop
Genetic Operators: Branch Crossover

Branch A moved to site-2 to ensure feasibility of offspring-1.
## Evolutionary Design of Fuel Additives: Results

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

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

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

Accelerator, Activator, Retarder, Rubber, Sulfur, Oils, Antidegradants, Filler

Rubber Formulation

Vulcanization + S$_8$
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
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.

1W. 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.


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)
Temporal Evolution of Crosslinks

Concentration of Crosslinks, $\nu$

Induction Phase

Curing Phase

Reversion Phase

Time Scale ~ Hrs

Time

Time Scale ~ Years
CAMD Sub-Problems

FORWARD PROBLEM
(Given Structure, Predict Properties/Performance)

INVERSE PROBLEM
(Given Properties, Predict Structure/Formulation)

Material Structure/Formulation

Hybrid Models

Neural Network

Fundamental Models

Properties/ Performance (P₁, P₂, ……Pₙ)

School of Chemical Engineering, Purdue University
THE OVERALL FORWARD MODEL

School of Chemical Engineering, Purdue University
Schematic of a Vulcanization Curve

Accelerator + Activator

Active Accelerator Complex

Active Sulfurating Agent

Rubber Bound Intermediate (Crosslink Precursor)

Rubber Bound persulfenyl radical

Polysulfidic crosslinks

Degradation

Desulfuration

Aged Network

2-morpholinothiobenzothiazole

(ZnO + Stearic Acid)
Accelerator Chemistry

Accelerator + Activator → Active Accelerator Complex

Active Accelerator Complex + $S_8$ → Active Sulfurating Agent

BtS-NR$_2$ → BtSH + R$_2$NH

BtS-NR$_2$ + BtSH → BtSSBt + R$_2$NH

BtS-SBt + $S_8$ → BtS-$S_x$-SBt

BtS-SBt + $S_8$ → BtS-SBt + $S_7$ → BtS-$S_2$-SBt + $S_6$ •••

BtS-$S_x$-SBt + BtS-$S_y$-SBt ↔ BtS-$S_z$-SBt + BtS-$S_w$-SBt

$A_x$ = BtS-$S_x$-SBt

School of Chemical Engineering, Purdue University
820 Reactions, 107 Species

- **Rubber, RH**
  - BtS-SBt
  - S8
  - S8
  - BtSZnSBt

- **Sulfurating Species**
  - BtS-Sx-SBt, BtS-Zn-Sx-SBt

- **Crosslink Precursor**
  - RSx-SBt, RSx-Zn-SBt

- **Persulfenyl Radical**
  - RSy•

- **Crosslinks**
  - RSSy-R

**Mechanisms**
- Lumped S8 Pickup
- Sequential S8 Pickup
- Both

**Desulfuration**
- RSSy-1R + BtSZnSSBt

**Degradation**
- Main-Chain Modification, Conjugated Dienes/Trienes, Inactive thiols

**Scorch Delay**
- BtSNR2
- ZnO

**Regeneration of Sulfurating Species**
- CDB + Pthalimide (Retarder Action)

**Cyclic Sulfide**
- BtS-SBt
- BtSNR2

**Crosslink Precursor**
- RSx-SBt, RSx-Zn-SBt

**Persulfenyl Radical**
- RSy•

**Crosslinks**
- RSSy-R

**Desulfuration**
- RSSy-1R + BtSZnSSBt

**Degradation**
- Main-Chain Modification, Conjugated Dienes/Trienes, Inactive thiols

**Mechanisms**
- Lumped S8 Pickup
- Sequential S8 Pickup
- Both
**POPULATION BALANCE EQUATIONS**

Equations for MBTS and other sulfating species

\[
\frac{d}{dt}[A_0] = k_{\text{MBS-MBS}}[\text{MBS}] - k_{\text{MBS-MBT}}[\text{MBT}] + k_{\text{A-A}}[A_0] \sum_{r=2}^{14} (r-1) A_r - k_{\text{A-S}}[A_0][S_8] \\
- k_{\text{A-R}}[A_0] - k_{\text{A-BST}}[A_0] \sum_{r=1}^{16} B_r^* + 0.5 k_{\text{E-E}}[E^*_0]^2 - k_{\text{DESULF}}[A_0] \sum_{r=2}^{16} V_u r + \\
\left[ \begin{array}{c} [A_1]^2 + 2[A_1][A_2] + 2 \{ [A_1][A_3] + 0.5[A_2]^2 \} + 2 \{ [A_1][A_4] + [A_2][A_3] \} + \\
2 \{ [A_1][A_5] + [A_2][A_4] + 0.5[A_3]^2 \} + 2 \{ [A_1][A_6] + [A_2][A_5] + [A_3][A_4] \} + \\
2 \{ [A_1][A_7] + [A_2][A_6] + [A_3][A_5] + 0.5[A_4]^2 \} + \\
2 \{ [A_1][A_8] + [A_2][A_7] + [A_3][A_6] + [A_4][A_5] \} \\
\end{array} \right] \\
\]

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

\[
\frac{d}{dt}[A_1] = -2k_{\text{A-R}}[A_1] - k_{\text{A-S}}[A_1][S_8] + 2k_{\text{A-A}}[A_0] \sum_{r=2}^{14} A_r + k_{\text{DESULF}}[A_0] \sum_{r=2}^{16} V_u r + \\
-2k_{\text{A-A}}[A_1] \sum_{r=1}^{13} A_r + k_{\text{E-E}}[E^*_0][E^*_1] \\
\]

School of Chemical Engineering, Purdue University
Equations for MBTS and other sulfurating species

\[
\frac{d}{dt}[A_0] = k_{A_2}[\text{MBT}][\text{MBS}] - k_{A_3}[A_0][S_8] - k_{A_4}[A_0] \left( \sum_{i=2}^{14} (i-1) A_i \right) - k_{C_1}[A_0] \\
- k_{C_4}[A_0] \left( \sum_{i=1}^{16} B_i \right) - k_{R_1}[A_0] \left( \sum_{i=1}^{16} V_u \right) \\
\frac{d}{dt}[A_1] = -2k_{C_1}[A_1] - k_{A_3}[A_1][S_8] - 2k_{A_4}[\text{MBTS}][A_1] + k_{R_1}[A_0] \left( \sum_{i=1}^{16} V_u \right) \\
\frac{d}{dt}[A_i] = -2k_{C_1}[A_i] - k_{A_3}[A_i] \begin{cases} [S_8], & 2 \leq i \leq 6 \\ 0, & i > 7 \end{cases} - (i+1)k_{A_4}[A_0][A_i] + k_{A_3}[A_i] \begin{cases} 0, & 2 \leq i \leq 6 \\ [S_8], & i > 7 \end{cases}
\]

Equations for Crosslink Precursors

\[
\frac{d}{dt}[B_0] = (k_{C_1}[\text{MBTS}] + k_{C_7}[E_1]) \\
\frac{d}{dt}[B_i] = -k_{A_4}[A_0][B_i] - 2k_{C_1}[A_i] - (i-1)k_{C_2}[B_i] + k_{C_7}[E_{i+1}]
\]
RATE-CONSTANT DETERMINATION

FACTS

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

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

\[ s.t. \quad m \text{ time points, } q \text{ concentrations} \]

\[ \frac{d}{dt} [v] = \phi(v, c, k) \]

k ≥ 0

k = Rate Constants

\[ v_{i,j}^{\exp} = \text{Experimental Data Point at time } i \text{ and conc } j \]
Example Writing Population Balance Equations

Potential Breaking Sites

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

\[ \frac{d}{dt} [B_2^\cdot] = k_A [B_2], \quad \frac{d}{dt} [E^\cdot] = k_A [B_2] \]

\[ \frac{d}{dt} [B^\cdot] = k_B [B_2], \quad \frac{d}{dt} [E_2^\cdot] = k_B [B_2] \]
Population Balance Equations

Equations for MBTS and other sulfurating species

\[
\frac{d}{dt} [A_0] = k_{MBS-MBT} [MBT][MBS] - k_{A-A} [A_0] \sum_{r=2}^{14} (r-1) A_r - k_{A-S} [A_0] \sum_{y=1}^{8} [S_y] \\
-k_{A-R} [A_0] - k_{A-BST} [A_0] \sum_{r=1}^{16} B^*_r + \frac{1}{2} k_{E-E} [E^*_0]^2 - k_{DESLF} [A_0] \sum_{r=2}^{16} V_u r \\
+k_{A-A} \sum_{x=1}^{13} \sum_{y=1}^{13-x} A_x A_y
\]

\[
\frac{d}{dt} [A_1] = -2k_{A-R} [A_1] - k_{A-S} ([A_1] - [A_0]) \sum_{y=1}^{8} [S_y] + 2k_{A-A} [A_0] \sum_{r=2}^{14} A_r + k_{DESLF} [A_0] \sum_{r=2}^{16} V_u r \\
-2k_{A-A} [A_1] \sum_{r=1}^{13} A_r + k_{E-E} [E^*_0][E^*_1]
\]

\[
\frac{d}{dt} [A_x] = -2k_{A-R} [A_x] - k_{A-S} ([A_x] - [A_{x-1}]) \sum_{y=1}^{8} [S_y] + 2k_{A-A} [A_0] \sum_{r=x+1}^{14} A_r \\
-(x-1)k_{A-A} [A_0][A_x] - 2k_{A-A} [A_x] \sum_{r=1}^{14-x} A_r + k_{E-E} [E^*_0][E^*_x] \\
+k_{A-A} \sum_{r=1}^{x-1} [A_r][A_{x-r}]
\]

School of Chemical Engineering, Purdue University
Final Set of Population Balance Equations

\[
\frac{d}{dt} \begin{bmatrix}
   Ax \\
   Bx \\
   V_u \\
   S_8 \\
   MBT \\
   E
\end{bmatrix} =
\begin{bmatrix}
   f_1 (k_{a_1}, k_{a_2}, k_{a_3}, k_{a_4}, Ax) \\
   f_2 (k_{a_1}, k_{a_2}, k_{a_3}, k_{a_4}, Ax, Bx, Bx) \\
   f_3 (k_{a_1}, k_{a_2}, k_{a_3}, k_{a_4}, Ax, Bx, Bx) \\
   f_4 (k_{a_1}, k_{a_2}, k_{a_3}, k_{a_4}, Bx, V_u) \\
   f_5 (k_{a_1}, A_0, S_8) \\
   f_6 (k_1, Ax) \\
   f_7 (k_2, k_4, k_{31}, k_{32}, Ax, Bx, Bx)
\end{bmatrix}
\]

Initial Conditions:
\[
A_0(0) = [MBS]_0 \quad S_8(0) = [Sulfur]_0 \\
A_{x(0)}(0) = B_x(0) = B^x_x(0) = V_u(0) = MBT(0) = E(0) = 0
\]

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.
Effect of Accelerator and Sulfur

Legend

- (1.0, 4.0)
- (0.5, 4.0)
- (1.0, 2.0)
- (0.75, 2.0)
- (0.5, 2.0)

Open Symbols = Experimental Data
Solid Line = Model Predictions
Population Balance Model – Different Temperatures

- T = 310 F
- T = 298 F

Open Symbols = Experimental Data
Solid Line = Model Predictions
Over 1000 rubber parts in failure critical functions

School of Chemical Engineering, Purdue University
Spatial Profile of the State-of-Cure for the Thermal History in the Mold

Thermal Profile

Cure Profile

Actual Part Design
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.
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

Formulation = [1 4 0.1 30 330]

Fitness (F) = \( \exp \left( -\alpha \sum \frac{P_i - P_i^{\text{des}}}{P_i^{\text{max}} - P_i^{\text{min}}} \right) \)
Genetic Algorithms (GA, contd)

“Survival of the fittest”

Fitness Calculation, Parent Selection

Operators

New Population

Evolution
Formulation Representation

Binary Representation

Phr of Different Ingredients

1 0 1 0 1 1 0 1 0 • • •
Genetic Operators

Crossover

Parent 1

00100110101

Parent 2

11101100000

Offspring 1

00100100000

Offspring 2

11101101010

Mutation

0100010101

1110010101
**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**

<table>
<thead>
<tr>
<th>Desired Property</th>
<th>Optimal Formulations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{\text{max}}$ (time to reach max cure) = 16 min</td>
<td>Formulation</td>
</tr>
<tr>
<td></td>
<td>[1.75 1.25 0.05 0 315]</td>
</tr>
<tr>
<td></td>
<td>[0.50 4.00 0.10 0 310]</td>
</tr>
<tr>
<td></td>
<td>[1.75 1.50 0.05 0 310]</td>
</tr>
<tr>
<td></td>
<td>[0.75 2.75 0.10 0 315]</td>
</tr>
<tr>
<td></td>
<td>[3.00 0.75 0.10 0 315]</td>
</tr>
<tr>
<td></td>
<td>[2.75 1.00 0.25 0 330]</td>
</tr>
</tbody>
</table>
NEW FORMULATIONS

Formulation       Fitness  $T_{\text{max}}$  $V_{u_{\text{max}}}$  $\sigma_{100}$  $\sigma_{200}$
[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

School of Chemical Engineering, Purdue University
Pharmaceutical Product
Development and Engineering

- Two Products
  Drugs
  Documents
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
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

Experiment Data Files

Experimental Methods Details

[Diagram of Experiment Ontology]

BelongsToProperty

BelongsToExperiment

Cycloserine: Solution Chemical Stability Measurement
Follows GLIF (Guideline Interchange Format), A standard ontology for clinical guidelines
Integrated Model-based Decision Support

Recipe

Process Information Repository

Used by Batches/Batch+

Material Form in XForms

Recipe Network

Used by PHASuite

Integrating with Mathematical Knowledge

- User created model instance
- Process Description (instances of Process Ontology)
- Model Ontology
  - Equations in MathML
  - Variables linked through URI
- Interface
  - Automatically generated Mathematica statements
- Web Applications
- Mathematica Kernel
- Rich Client

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?