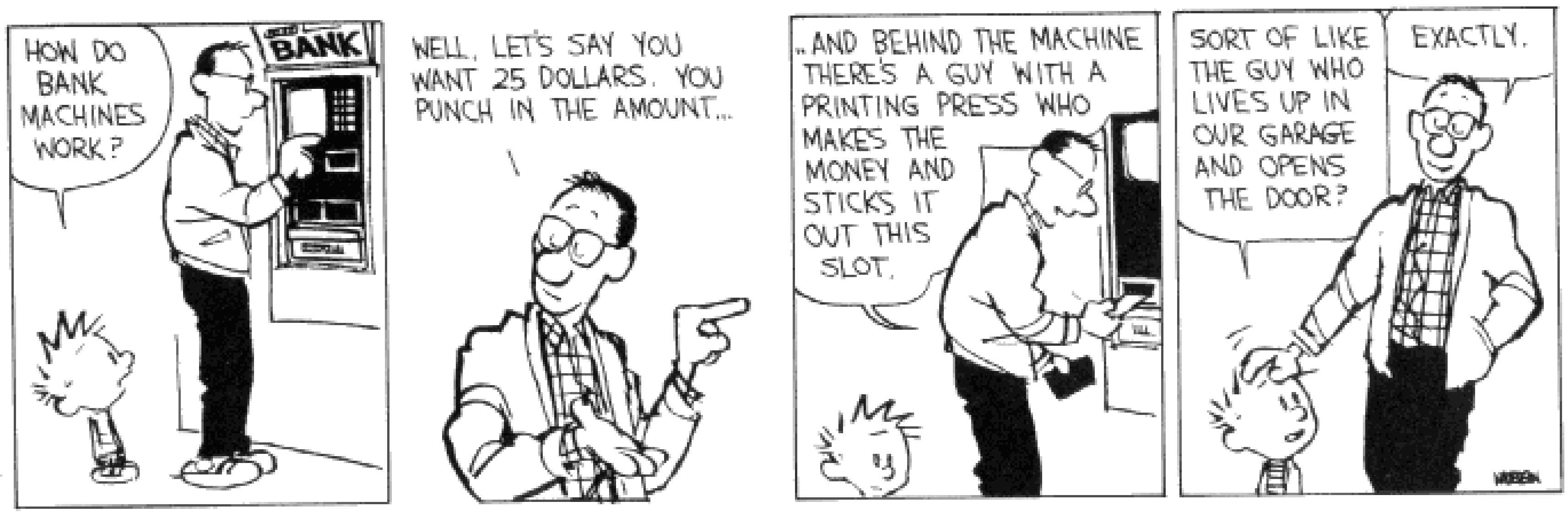
Analysis of complex reaction networks using mathematical programming approaches

Marianthi Ierapetritou

Department Chemical and Biochemical Engineering
Piscataway, NJ 08854-8058

RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Complex Process Engineering Systems?



General Motivation

Diverse complex systems spanning different scales

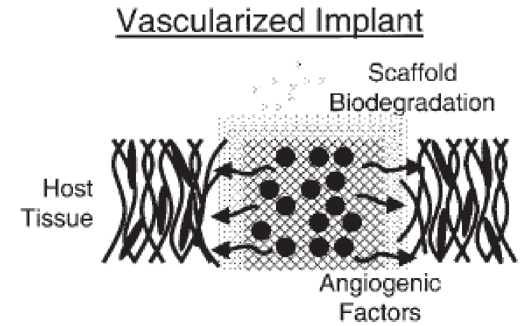
- Liver metabolism (molecular level)
- Combustion systems (process level)
- Scheduling of multiproduct-multipurpose plants (plant level)

Motivation -1: Liver Support Devices

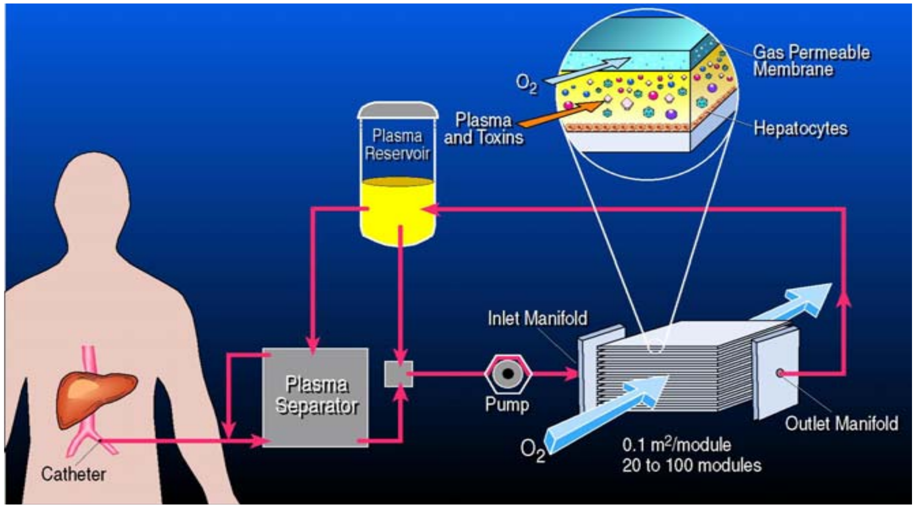
- Acute and chronic liver failure account for 30,000 deaths each year in the US
- A large number of liver diseases:
  - Alagille Syndrome
  - Alpha 1 - Antitrypsin Deficiency
  - Autoimmune Hepatitis
  - Biliary Atresia
  - Chronic Hepatitis
  - Cancer of the Liver
  - Cirrhosis
  - Cystic Disease of the Liver
  - Fatty Liver
  - Galactosemia
  - Hepatitis A, B, C
- Currently liver transplantation is primary therapeutic option. Scarcity of donor organs limits this treatment

Solutions

- Adjunct Internal Liver Support With Implantable Devices
  - Hepatocyte Transplantation
  - Implantable Devices
  - Encapsulated Hepatocytes



- Extracorporeal Temporary Liver Support
  - Nonbiological devices: hemodialysis, hemofiltration, plasma exchange units
  - Hepatocyte- and liver cell-based extracorporeal devices



Challenges

a) How to maximize long-term functional stability of hepatocytes in inhospitable environments

b) How to manufacture a liver functional unit that is scalable without creating transport limitations or excessive priming volume that must be filled by blood or plasma from the patient

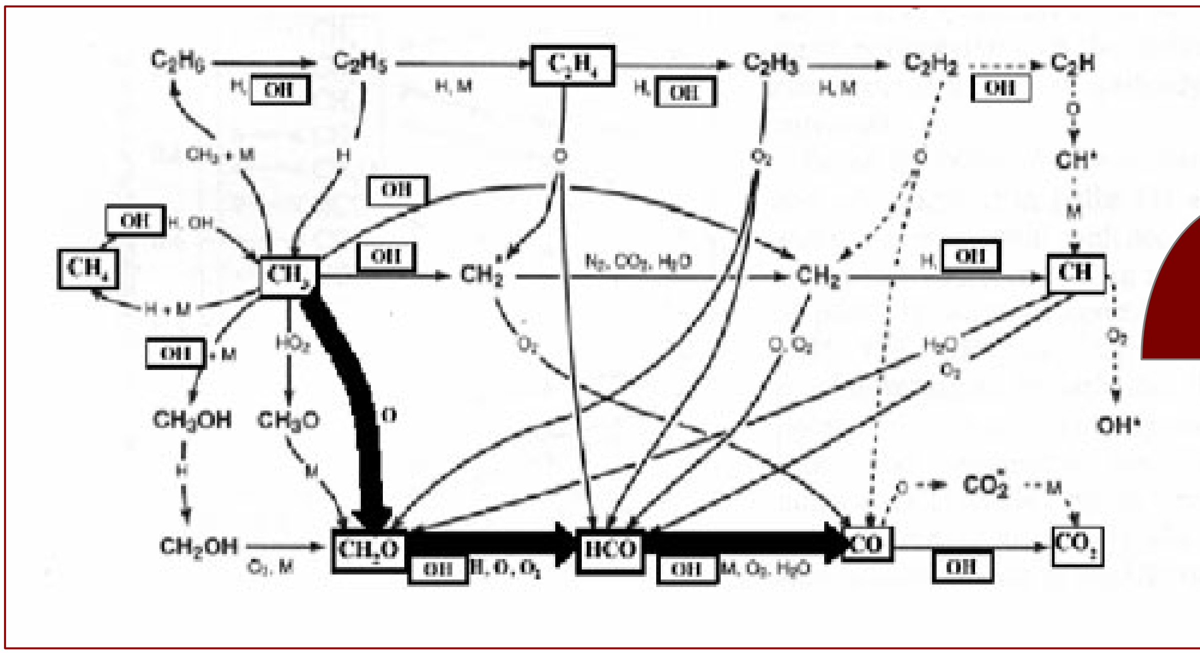
c) How to procure the large number of cells that is needed for a clinically effective device

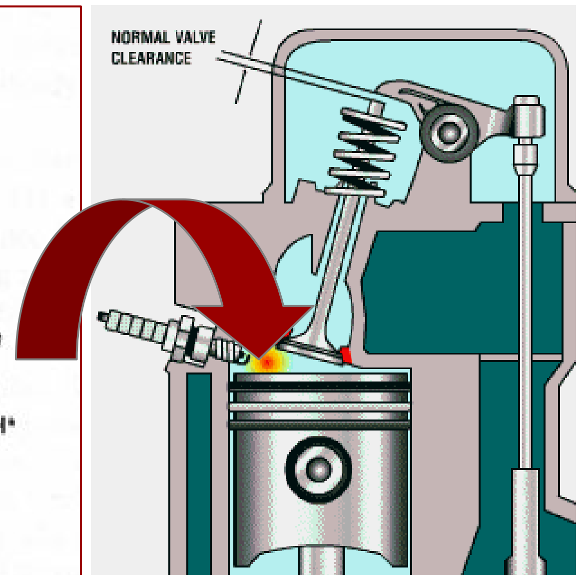
… and the Reality

- Problem complexity: System of large interconnectivity
- Large number of adjustable variables
- Uncertainty

Motivation – 2 : Combustion

Conversion of chemical energy to mechanical energy



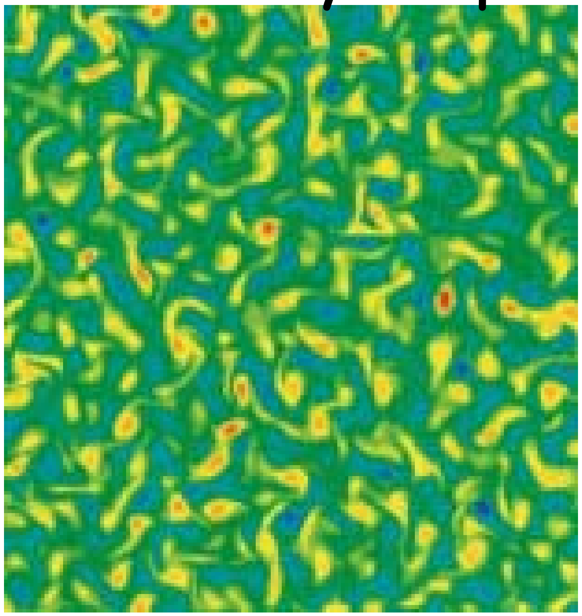


Require alternate representation of complex kinetic mechanism, without sacrificing accuracy

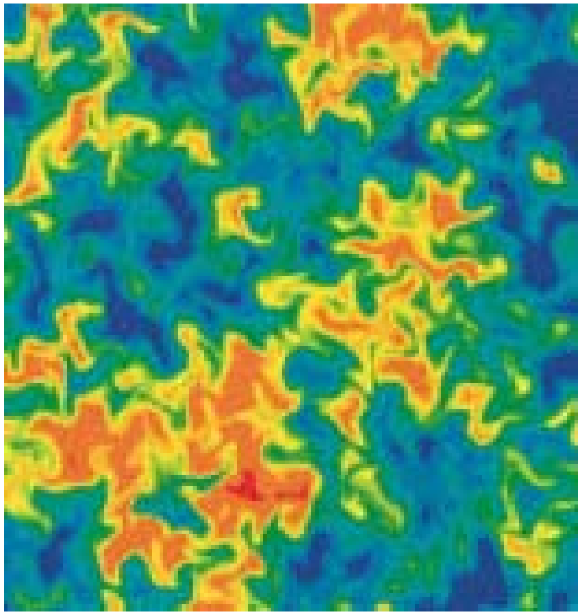
Challenges: Combine Flow and Chemistry

- How should these be combined ?

Velocity map



Composition map



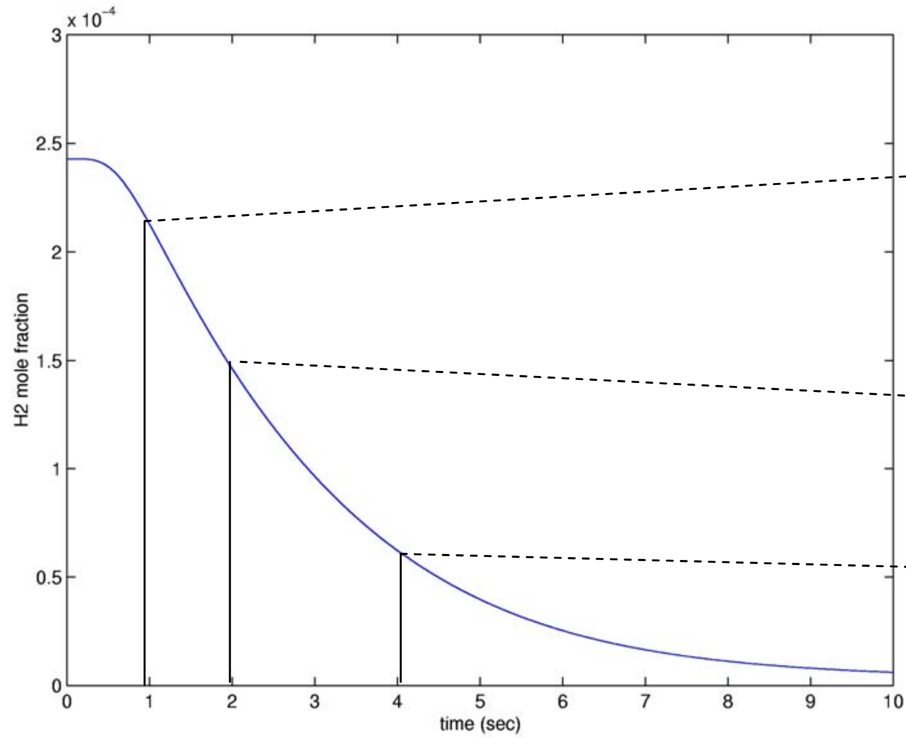
…and the Reality

Reaction mechanisms grow with molecular complexity:

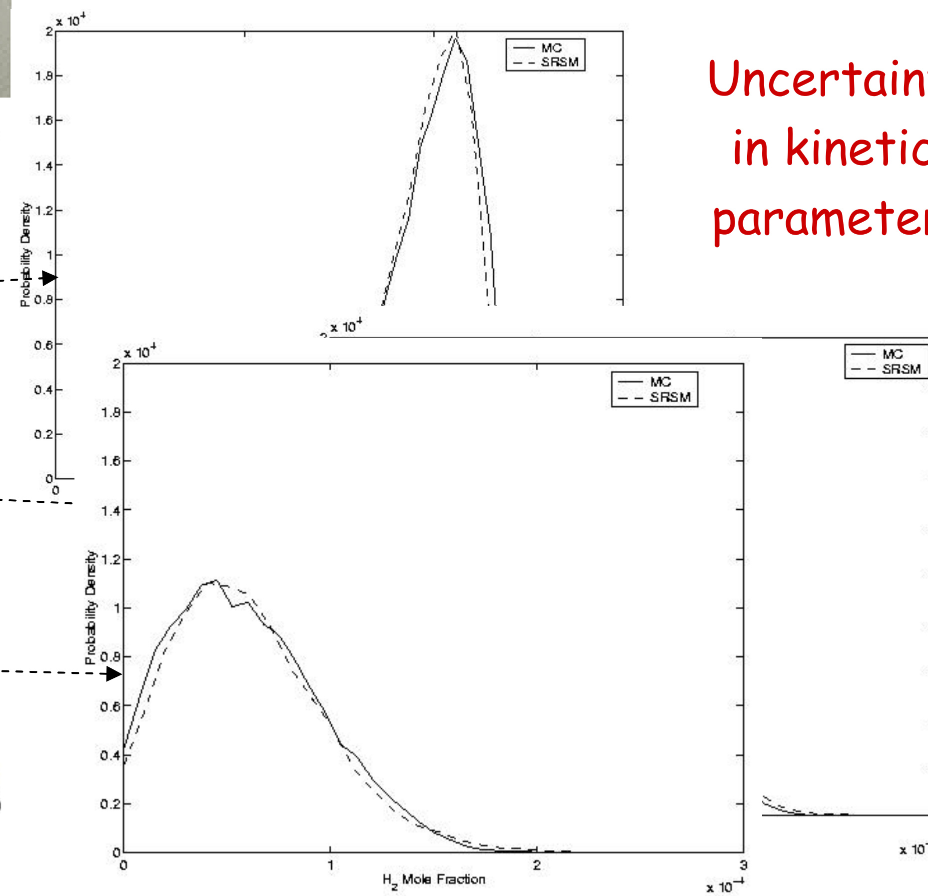
Fuel	H_2 (hydrogen)	CH_4 (methane)	C_3H_8 (propane)	C_6H_{14} (hexane)	$C_{16}H_{34}$ (cetane)
Number of species	7	30	100	450	1,200
Number of reactions	25	200	400	1,500	7,000

Complex kinetics
(LLNL Report, 2000)

Uncertainty in kinetic parameters



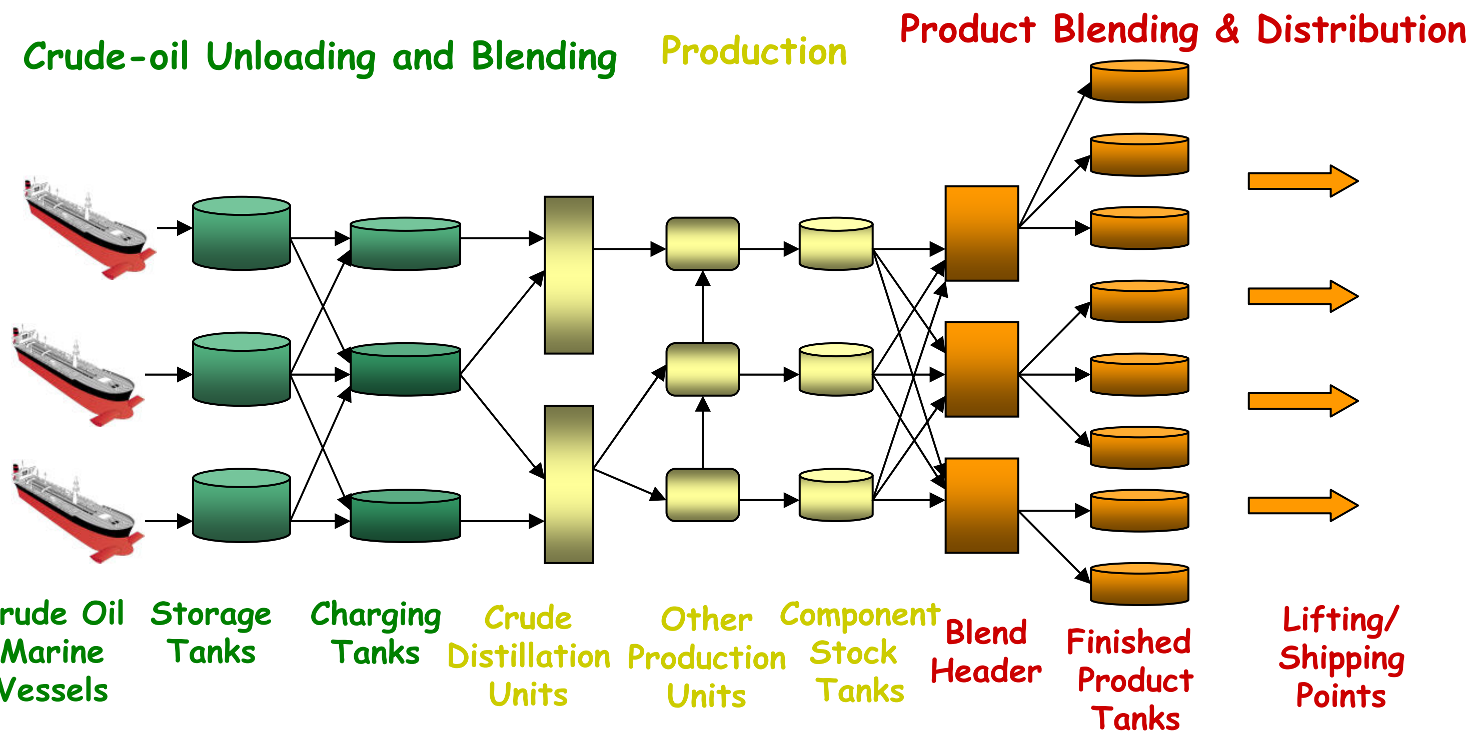
H2 mole fraction vs. time



Motivation-3: Large-Scale Process Operations

Goal: Address the optimization of large-scale short-term scheduling problem, specifically in the area of refinery operations

Add $100s of million/year profit by optimizing crude-oil-marketing enterprise



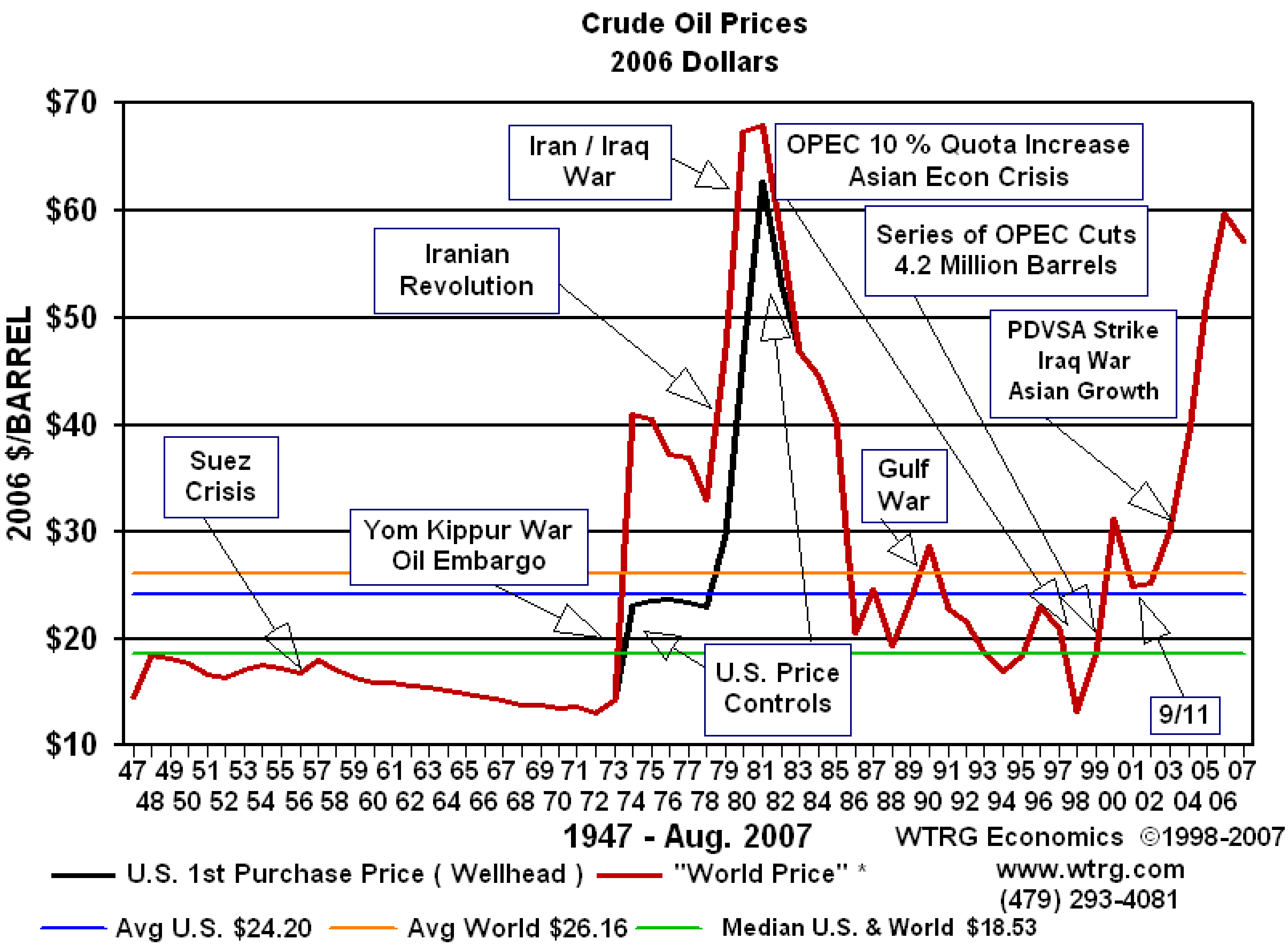
Max Profit

Subject to: Material Balance Constraints

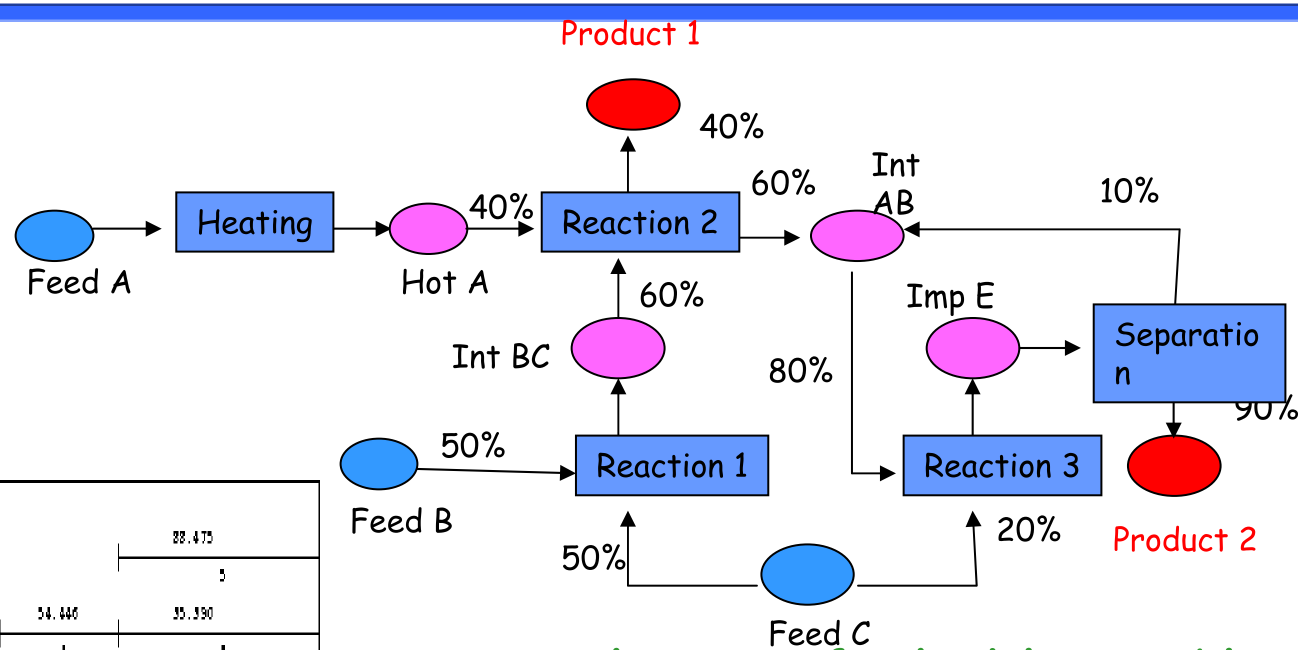
Allocation Constraints, Sequence Constraints

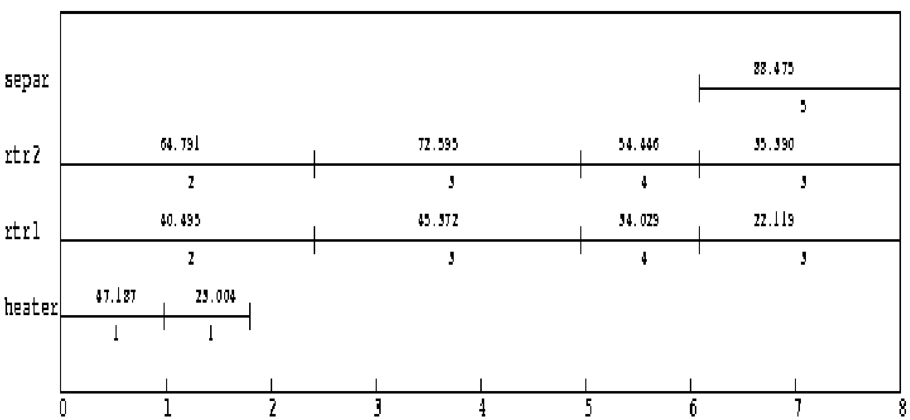
Duration Constraints, Demand Constraints …

Challenges: Parameter Fluctuations



…and the Reality





As time horizon of scheduling problem increases, the solution requires exponential computational time which makes the problem intractable.

Time	Number of Event Points	Objective function value	CPU time (sec)
8 hours	5	1498.19	0.47
16hours	9	3737.10	177.93
24 hours	13	6034.92	92367.94

Systems Approaches

- Mathematical programming

  Systematic consideration of variable dependences

  Continuous and discrete representation

- Sensitivity – parametric analysis

  Identification of important features and parameters

- Feasibility evaluation

  Conditions of acceptable operation

- Optimization

  Multiobjective since we have more than one objective to optimize

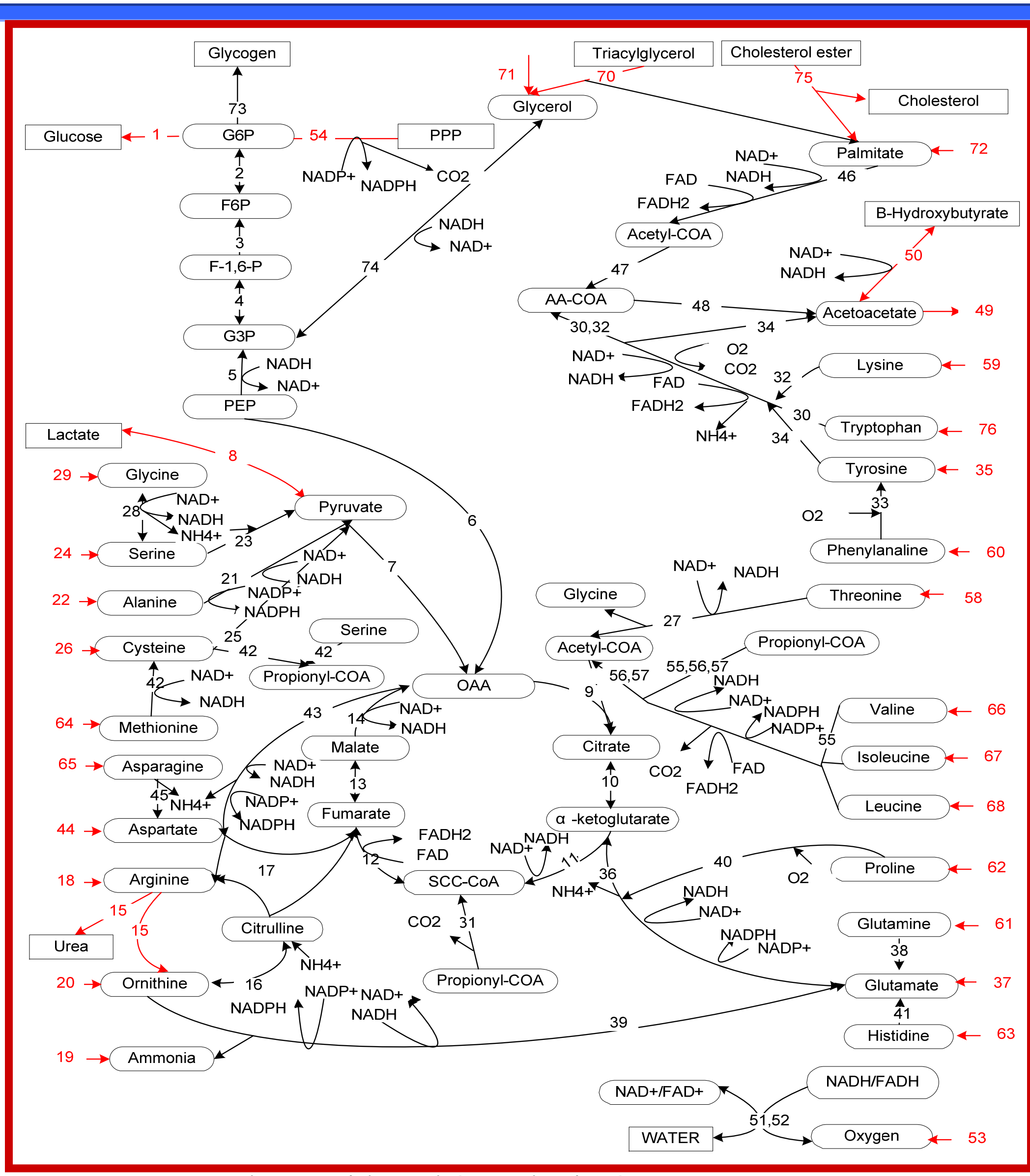
- Uncertainty

  Evaluation of solutions that are robust to highly fluctuating environment

Presentation Outline

- Complexity reduction using mathematical programming approaches
  - Optimization of hepatocyte functionality
  - Reduction of complex chemistry
- Uncertainty analysis & feasibility evaluation
- Analysis of alternative solutions

Hepatic Metabolic Network



Chan et al (2003) Biotechnol & Bioengineering

Main Assumptions

1) Gluconeogenic and fatty acid oxidation enzymes are active in plasma
2) Energy-requiring pathways are negligible
3) Metabolic pools are at pseudo-steady state.

Main Reactions

Glucose Metabolism (v_1-v_7)

Lactate Metabolites & TCA Cycle(v_8-v_{14})

Urea Cycle (v_{15}-v_{20})

Amino acid uptake & metabolism (v_{21}-v_{68}, ,v_{76})

Lipid & Fatty Acid Metabolism (v_{46}-v_{50},v_{71}-v_{75})

45 internal metabolites

76 reactions:

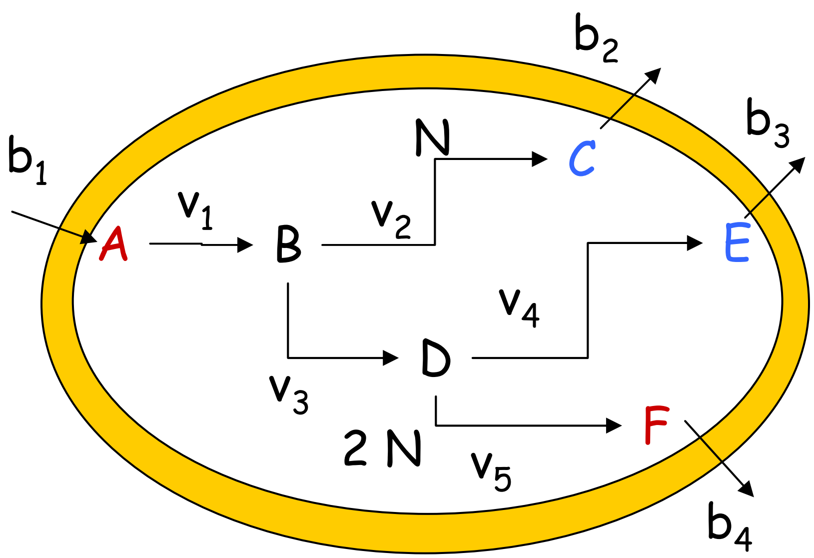
33 irreversible + 43 reversible

34 measured (red) + 42 unknown

Rationale for Metabolic Modeling

- Interpretation and coupling to experimental data.
- Gain insights into how cells adapt to environmental changes.
- To identify key pathways for hepatocyte function.

Metabolic **F**lux **A**nalysis (**MFA**) is developed to calculate unknown intracellular fluxes based on the extracellular measured fluxes.



$$\begin{bmatrix} dA/dt \\ dB/dt \\ dC/dt \\ dD/dt \\ dE/dt \\ dF/dt \\ dN/dt \end{bmatrix} = \begin{vmatrix} -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & -2 & 0 & 0 & 0 & 0 \end{vmatrix} \begin{vmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{vmatrix}$$

Pseudo-steady State $Sv = 0$

$S_u v_u = -S_m v_m$

- Measure 2 fluxes: Uniquely-determined system
- Measure 3 fluxes: Overdetermined System- Least Square method
- Measure 1 flux: Underdetermined System- Linear Programming

Optimization in Metabolic Networks

- **Single-level Optimization:** Optimize a single objective function (e.g. maximization of a single metabolic flux).

  Eward & Palsson (2000) PNAS
  Uygun et al., (2006) Ind. Eng. Chem. Res.
  Schilling. et al., (2001) Biotechnol Bioeng
  Lee S. et al (2000) Computer & Chem. Eng.

- **Multi-objective Optimization**: Several objective functions are simultaneously optimized (e.g. minimizing the toxicity and maximizing metabolic production).

  Sharma N.P. et al., (2005) Biotechnol Bioeng
  Nagrath D. et al. (2007) Annals of Biomedical Engineering

- **Multi-level Optimization**: Several objectives acting hierarchically to optimize their own objective function (e.g. Minimize the difference of predicted fluxes from experimentally observed values to optimize the cellular objective function).

  Segre D. et al (2002) PNAS
  Nolan R.P et al (2005) Metabolic Engineering
  Burgard & Maranas (2003) Biotechnol Bioeng
  Uygun et al., (2007) Biotechnol Bioeng

Multi-Objective Optimization $\neq$ Multi-level Optimization

Single-level Optimization: Maximize Urea Secretion

Aim: Identify the flux distributions for optimal urea production that can fulfill metabolites balances and flux constraints

$$Max : Z = v_{urea}$$

$$Subject \quad to : \sum_{j=1}^{N} S_{ij} v_j = 0 \qquad \forall i \in M$$

$$v_j^{\min} \le v_j \le v_j^{\max} \qquad \forall j \in K$$

Experimental Data*		Optimal Value	Increase
HIP	0.23±0.43	6.81	> 10 fold
HPAA	1.32±0.69		> 3 fold
LIP	0.17±0.24		> 15 fold
LPAA	2.35±0.52		> 2 fold

Unit: μmol/million cells/day

**Chan & Yarmush et al (2003) Biotechnol Prog*

Results for Single-Level Optimization

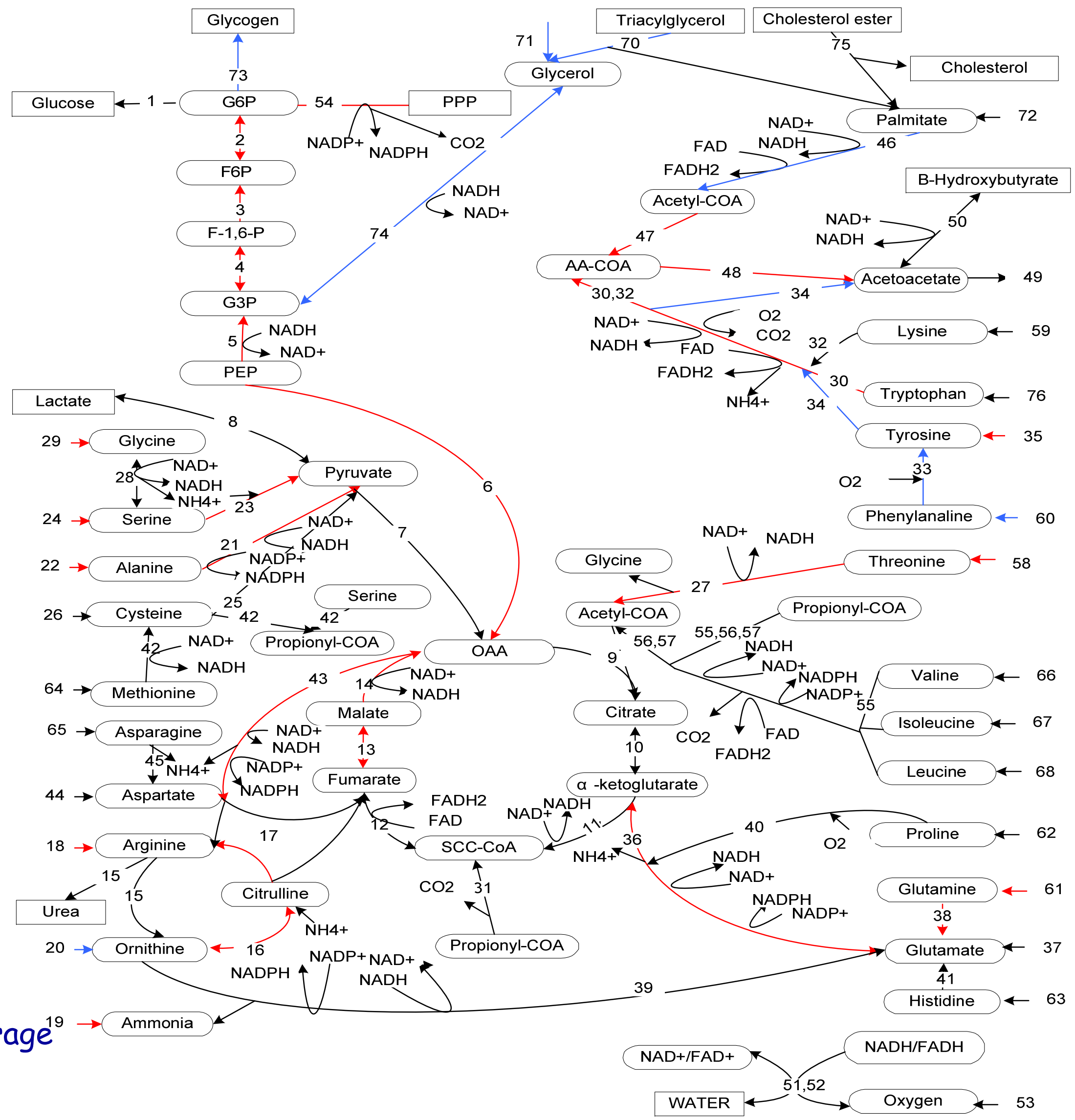
Fluxes significantly altered through the pathways (more than 30 % change)

Increased fluxes

- Gluconeogenesis (R2-R6)
- TCA Cycle (R13,R14)
- Urea Cycle (R16,R17)
- Amino Acid Catabolism (R21,R23,R27,R30,R36,R38,R43)
- Fatty Acid Metabolism (R47,R48)
- Pentose Phosphate Pathway (R54)
- Amino acid uptake fluxes (e.g: Arginine, Serine, Glycine.....)

Decreased fluxes

- Amino Acid Catabolism (R33,34)
- Fatty Acid Oxidation (R46)
- Glycerol uptake and metabolism, glycogen storage (R70,R71,R73,R74)



Marianthi Ierapetritou
RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

Multi-objective Optimization

ε-Constraint Method

$$\max \ v_{albumin}$$

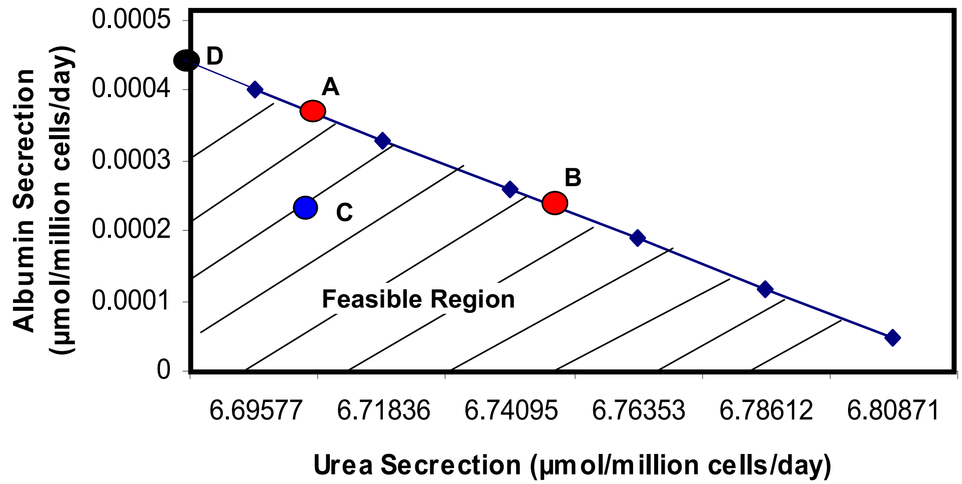
$$\text{subject to}: \sum_{j=1}^{N} S_{ij} v_j = b_i, \quad i = 1,\ldots, M$$

$$v_j^{min} \le v_j \le v_j^{max}$$

$$v_{urea} \ge \varepsilon$$

Different values of ε were used to calculate the maximum albumin Production = Pareto set

Results



- Pareto optimal set yields the feasible region for BAL operation
- Point A,D and B belong to the Pareto Set
- Both urea and albumin secretion can be improved at point C by moving towards the Pareto set

Sharma NS, Ierapetritou MG, Yarmush ML., Biotechnol Bioeng. 2005 Nov 5;92(3): 321-35.

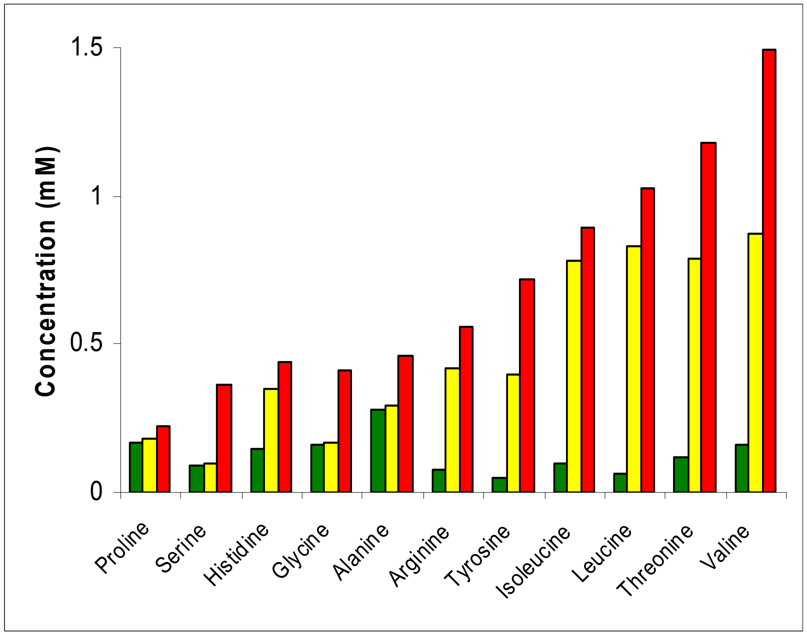
Optimal Amino Acid Supplementation

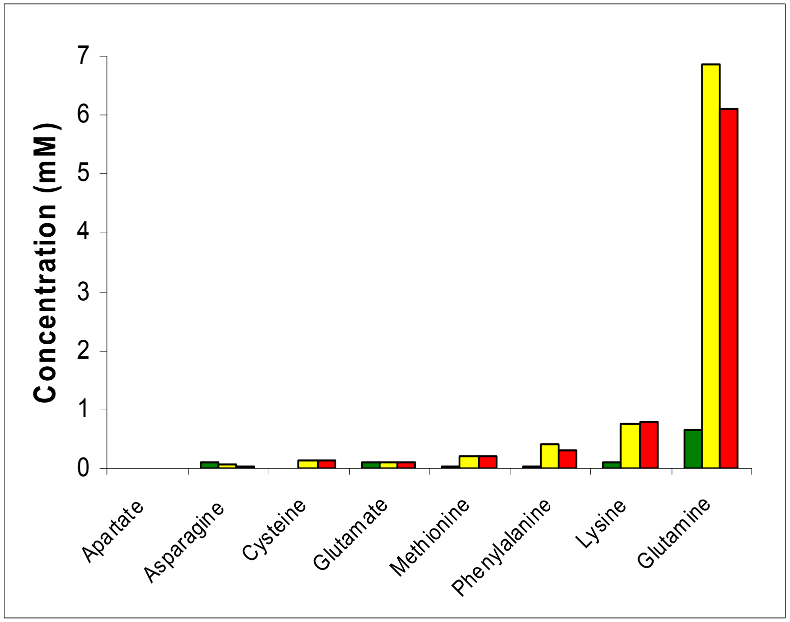
Optimal Flux Distribution

Assumption: Linear Relationship

Modulation by optimal AA. Supplementation

- Culture media supplementation to improve cellular function
- Advantage is no direct genetic intervention





Higher Amino Acid Supplementation

Similar & Low Amino Acid Supplementation

Plasma — AA supplementation in low-insulin — Optimal AA supplementation

Critical Pathways for Urea and Albumin Function

Aim: Compare the flux distributions between the wild-type and knock out condition and identify the essential reactions for target cell functions

Bi-level Optimization

Leader Objective: Minimize the difference between the native type and the knockout condition after reaction deletion

Follower Objective: Maximize the particular cell function (urea production)

v_j^{NA} : represents the flux distribution of native type determined from MILP model

$$
\begin{aligned}
Min: \quad & \sum_{j \in N} | v_j - v_j^{NA} | \\
s.t. \quad & Max: \quad v_{urea} \\
& s.t. \quad \sum_{j=1}^{N} S_{ij} v_j = 0, i = 1,2 \ldots\ldots M \\
& \qquad v_j^{\min} \le v_j \le v_j^{\max} \\
& \qquad v_d = 0
\end{aligned}
$$

Burgard et al., (2003) Biotechnol Bioeng; 84: 647-657

Segre D. et al (2002) PNAS; 99: 15112-15117

Important Pathways

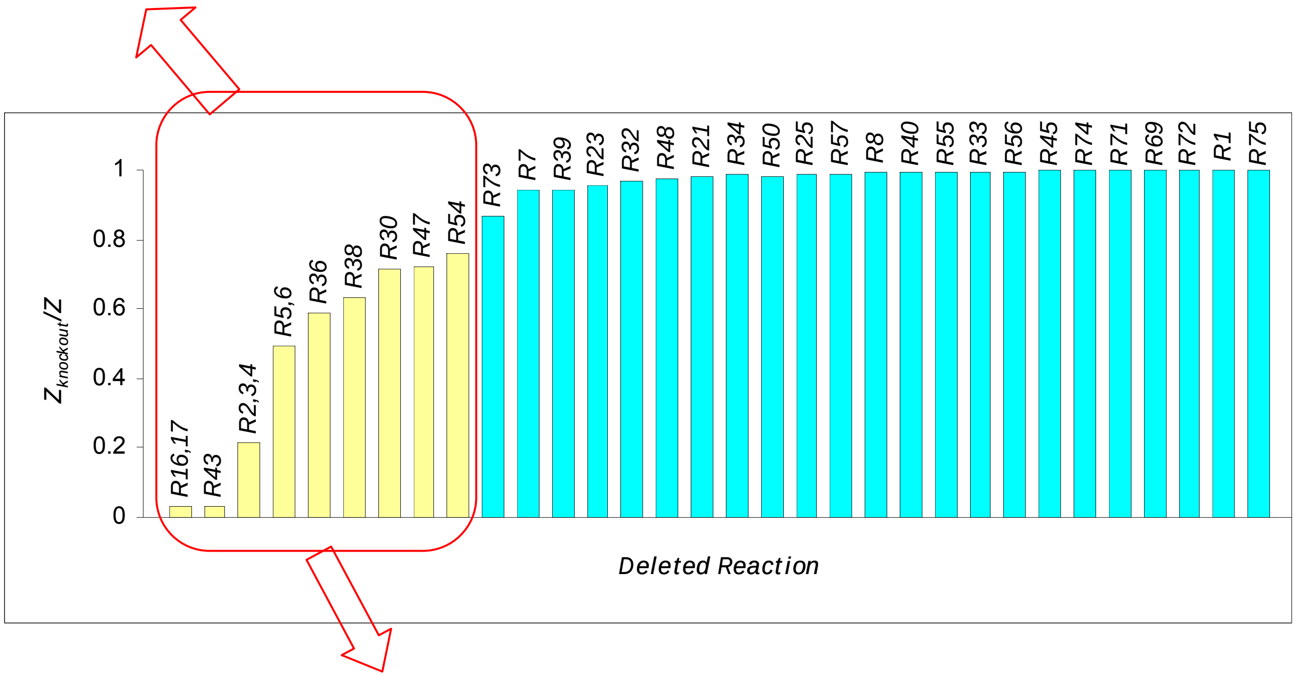
Z: optimal value of native type

$Z_{knockout}$: optimal value after individual reaction deletion

Using Extended Kuhn-Tucker Approach

C. Shi. (2005) Applied Mathematics & Computation

Gluconeogenesis(R2-R6), Urea Cycle (R16,R17)



Amino Acid Catabolism (R30,R36,R38,R43) Fatty Acid Metabolism (R47),

Pentose Phosphate Pathway (R54)

Comparison of Different Methodologies

Experiments:

High Insulin (0.5 U/ml)
- I) no supplements
- II) *AA* supplements

Low Insulin (50 μU/ml)
- III) no supplements
- IV) *AA* supplements

Model:

$$Min: \sum_{j \in Measured\ fluxes} | v_j - v_j^{NA} |$$

Subject to: $Max: \; v_{urea}$

Subject to: $$\sum_{J=1}^{N} S_{ij} v_j = 0, i = 1,2 \cdots\cdots M$$

$$v_j^{\min} \le v_j \le v_j^{\max}, \forall j \in k$$

Results:

Case 1: v_j^{NA} - Measured fluxes from Experiment LIP

Approach	Error	Urea
KKT	0.079	0.165
Primal-Dual	269.239	6.809

Case 2: v_j^{NA} - Measured fluxes from Experiment LPAA

Approach	Error	Urea
KKT	0.246	2.254
Primal-Dual	59.886	6.809

Hong, Roth, Ierapetritou, AIChE Annual Meeting, Nov 2007

Critical Pathways for Urea and Albumin Function

Logic based programming

$$\min_{\lambda_j} \Phi = \sum_{j=1}^{N} \lambda_j$$

$$\text{subject to: } \sum_{j=1}^{N} S_{ij} v_j = b_i \,, \quad i = 1, \ldots, M$$

$$v_j^{min} \lambda_j \leq v_j \leq v_j^{max} \lambda_j , \quad j = 1, \ldots, N$$

λ_j is a binary variable corresponding to the presence or absence of reaction (j) in the network.

Critical pathways for urea and albumin function

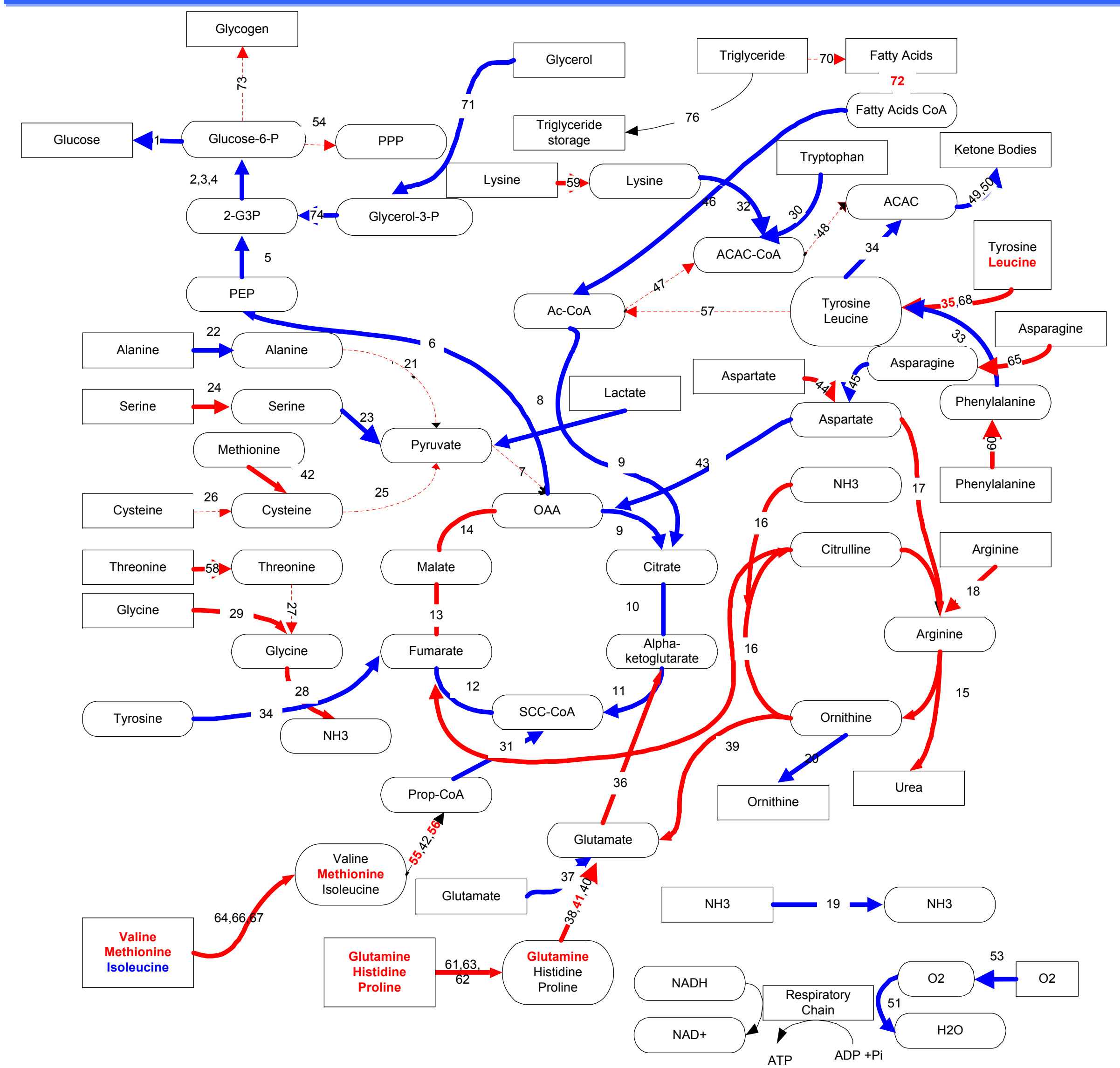
Different Conditions

Elucidate Insulin Effects

Elucidate AA Effects

Condition	Medium Pre-Conditioning	Plasma Supplementation
HIP	High Insulin	Unsupplemented
LIP	Low Insulin	Unsupplemented
LIPAA	Low Insulin	Amino Acids
Optimal	Low Insulin	Optimal Amino Acids

Optimal Condition vs. LIPAA



24 ARG + 32 ASP + 61 ALA + 24 SER + 35 CYS + 57 GLU + 17 GLY + 21 TYR + 33 **THR** + 53 **LYS** + 26 **PHE** + 25 GLN + 30 **PRO** + 15 **HIS** + 6 MET + 20 **ASN** + TRP + 35 **VAL** + 13 ISO + 56 **LEU** + 2332 ATP → **albumin** + 2332 ADP + 2332

- Compensatory effects in TCA cycle fluxes
- Lower gluconeogenic and lipid metabolism pathway fluxes.
- Higher urea cycle fluxes.
- Higher AA uptake rates

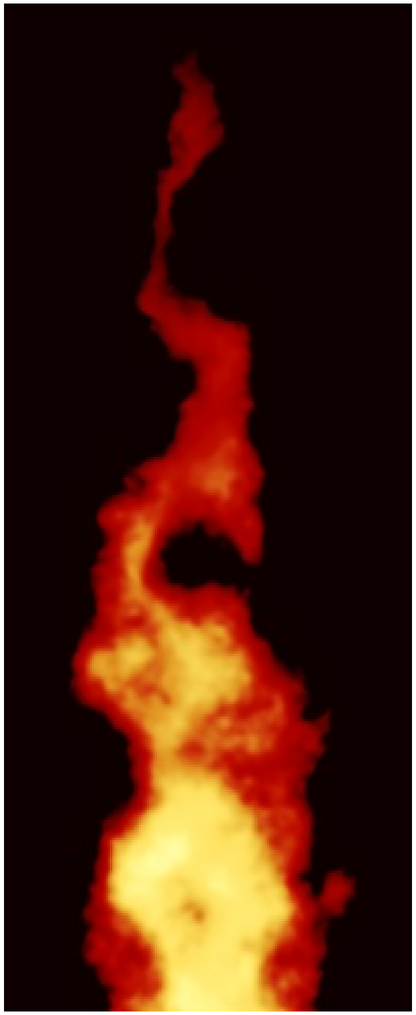
Thick red lines correspond to higher fluxes for optimal condition as compared to LIPAA.

Thick blue lines correspond to lower fluxes for optimal condition as compared to LIPAA.

Dotted red lines correspond to reactions not important in Optimal case for maximal urea and albumin function.

- Complexity reduction using mathematical programming approaches
    - Optimization of hepatocyte functionality
    - Reduction of complex chemistry
- Uncertainty analysis & feasibility evaluation
- Analysis of alternative solutions

Model Reduction Using Mathematical Programming



$H_2 + O_2 \Leftrightarrow 2OH$
$OH + H_2 \Leftrightarrow H_2O+H$
$O + OH \Leftrightarrow O_2+H$
$O + H_2 \Leftrightarrow OH+H$
$H + O_2 \Leftrightarrow HO_2$
$OH + HO_2 \Leftrightarrow H_2O+O_2$
$H + HO_2 \Leftrightarrow 2OH$
$O + HO_2 \Leftrightarrow O_2+OH$
$2OH \Leftrightarrow O+H_2O$
$H + H \Leftrightarrow H_2$
$H + H + H_2 \Leftrightarrow H2 + H_2$
$H + H + H_2O \Leftrightarrow H_2 + H_2O$
$H + OH \Leftrightarrow H_2O$
$H + O \Leftrightarrow OH$
$O + O \Leftrightarrow O_2$
$H + HO_2 \Leftrightarrow H_2 + O_2$
$HO_2 + HO_2 \Leftrightarrow H_2O_2 + O_2$
$H_2O_2 \Leftrightarrow OH + OH$
$H2O_2 + H \Leftrightarrow HO_2 + H_2$
$H_2O_2 + OH \Leftrightarrow H_2O + HO_2$

Reduction of complex kinetic mechanism to enable detailed flame simulation

Chemical Kinetic Model

Reaction mechanisms grow with molecular complexity:

Fuel	H_2 (hydrogen)	CH_4 (methane)	C_3H_8 (propane)	C_6H_{14} (hexane)	$C_{16}H_{34}$ (cetane)
Number of species	7	30	100	450	1,200
Number of reactions	25	200	400	1,500	7,000

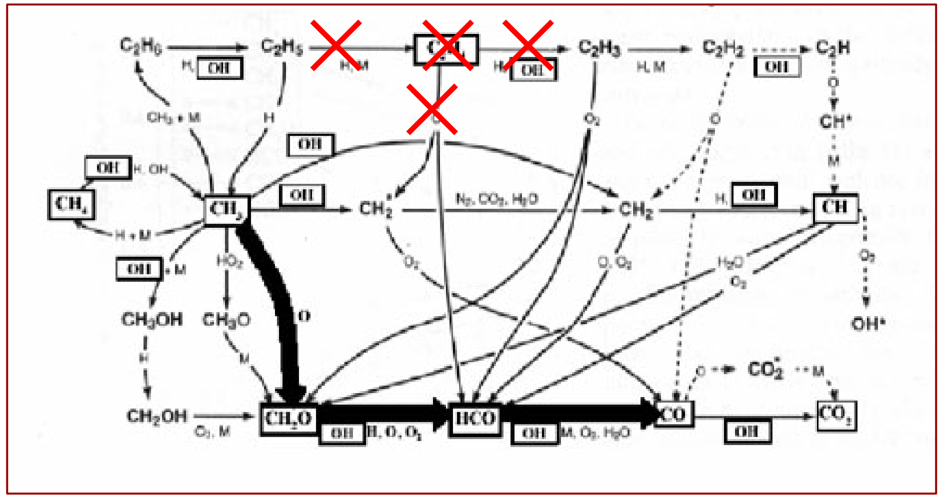
Detailed kinetic models are extremely complex

Optimization Based Reduction

Objective function : $min \sum_{i=1}^{N_S/N_R} \lambda_i$

Constraint : $\chi \leq \delta$

where χ is an error measure representing deviation of full profile from reduced profile



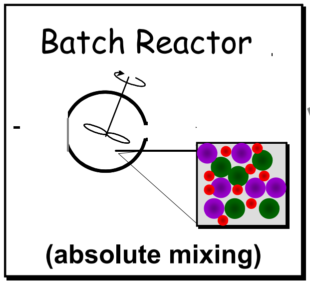
λ_i : Binary variable corresponding to i[th] reaction/species

$\sum_{i=1}^{N_S/N_R} \lambda_i$ represents total number of species / reactions

Constraint : retain desired system behavior within prescribed accuracy

Evaluation of Constraint Function

Constraint : $\chi \le \delta$

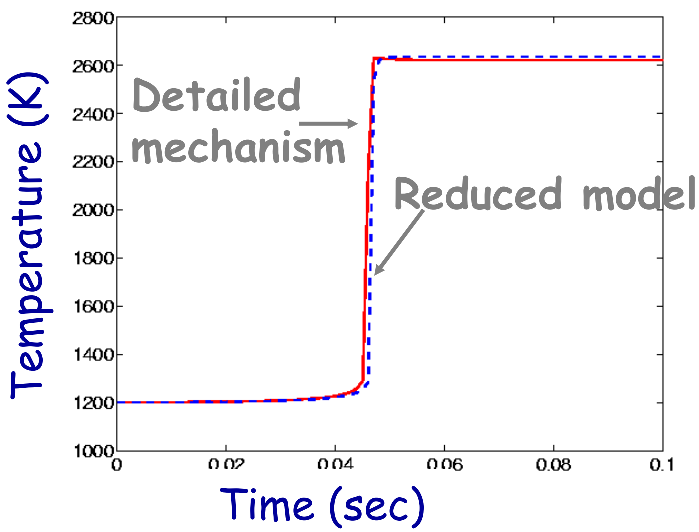


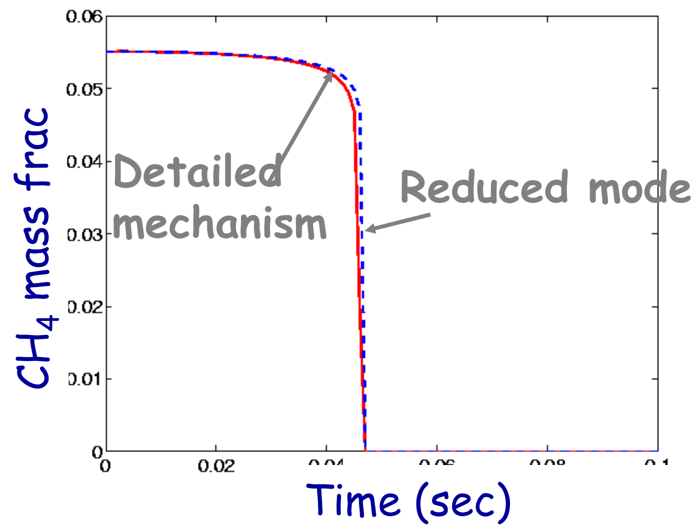
$$\frac{dy_k(t)}{dt} = \frac{R_k M_k}{\rho} \qquad k = 1, \ldots, N_s$$

$$\frac{dT}{dt} = -\sum_{k=1}^{N_s} \frac{R_k M_k h_k}{\rho \bar{C}_p} \qquad R_k = \sum_{i=1}^{N_r} \lambda_i (\gamma_{ki}{}^r - \gamma_{ki}{}^f) q_i$$

Initial Condition:
CH_4=0.055, O_2=0.19
T=1200 K
(Solved using DVODE)

Methane mechanism(GRI-3.0) : 53 species, 325 reactions
Reduced Model : 17 species, 59 reactions, δ = 0.085





$$\chi = \left(\sum_{k \in \kappa} \int_t \left(\frac{y_k{}^{reduced}(t) - y_k{}^{full}(t)}{y_k{}^{full}(t)} \right)^2 dt + \int_t \left(\frac{T^{reduced}(t) - T^{full}(t)}{T^{full}(t)} \right)^2 dt \right)^{1/2}$$

Constraint : $\chi \le \delta$

Two Step Solution Procedure

Mathematical Model: MINLP with embedded ODEs

Binary variables for species reduction (N_s) : 53
Binary variables for reaction reduction (N_r) : 325

Methane mechanism: GRI 3.0

Species reduction

↓

Eliminate reactions associated with removed species

↓

Generate initial reduced reaction set ($N_s < N_r$)

(17 species, 113 reactions)

↓

Perform reaction reduction with N_r binary variables

Final reduced model

(17 species,59 reactions)

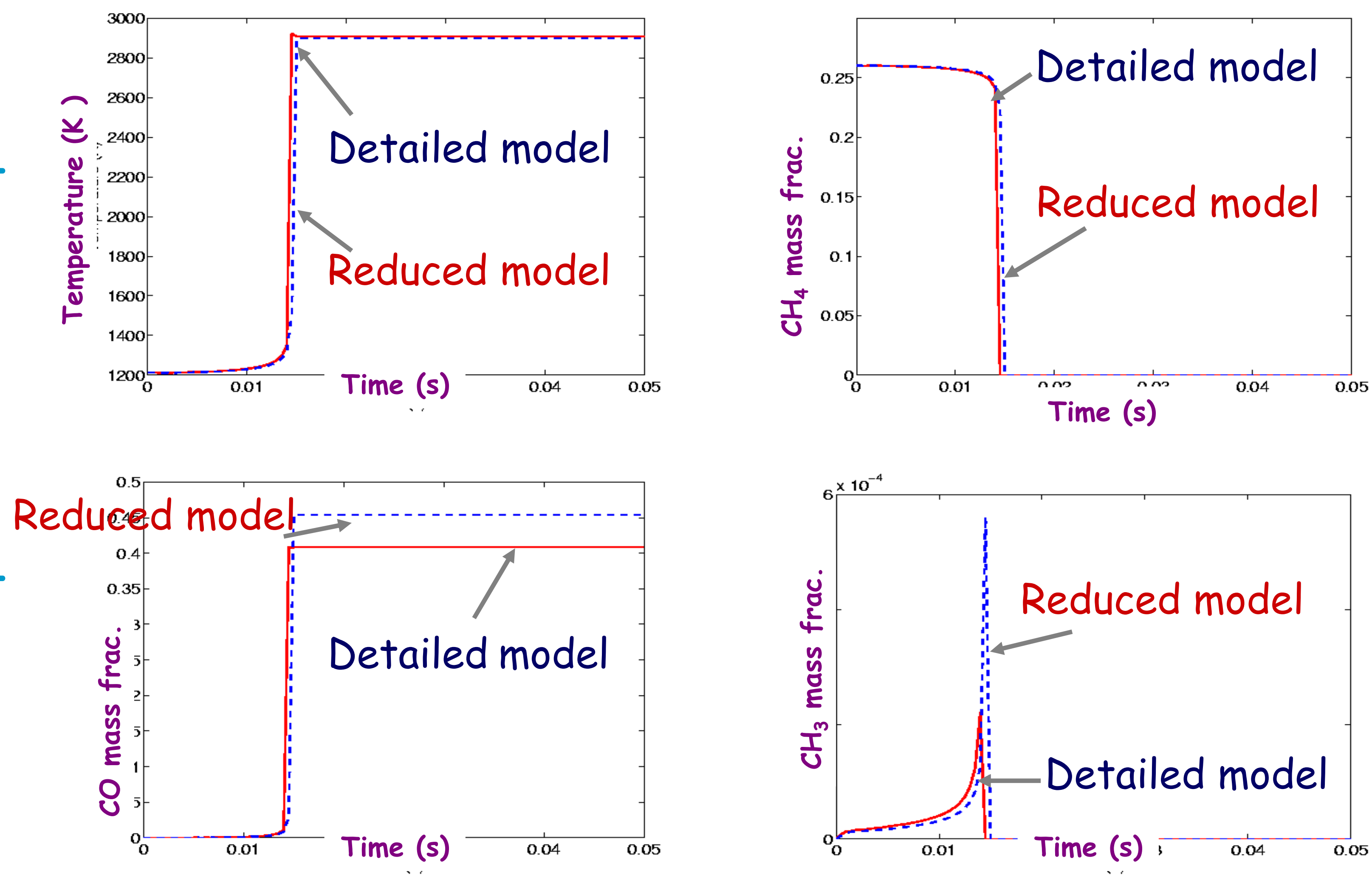
Banerjee and Ierapetritou, Chem. Eng. Sci, 8, 4537, 2003.

Performance of Reduced Models

Reduced Model : 17 species, 59 reactions, δ = 0.085 CH_4=0.26, O_2=0.086 T=1200 K

Watched Species

Non-Watched Species



Computational Savings by Reduction

Full Mechanism

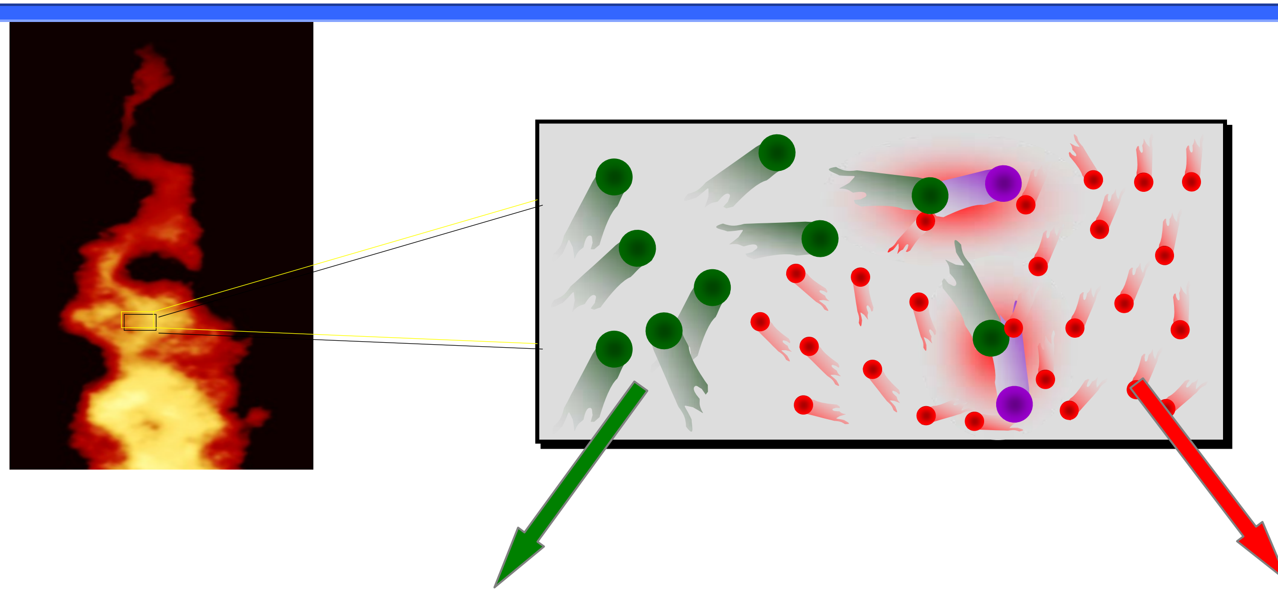
Reduced Mechanisms

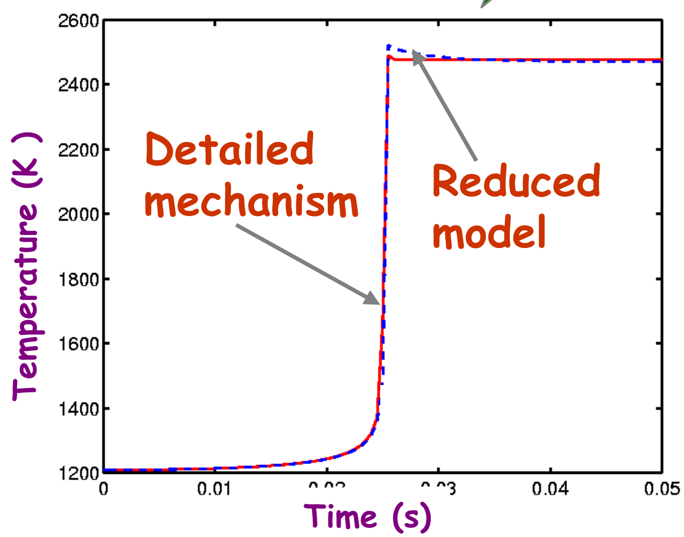
Ns	Nr	Sparsity	Cummulative(CPU)
53	325	1227 (0.07)	190.3
29	126	461 (0.126)	29.57
22	81	291 (0.163	13.87
22	35	131 (0.17)	5.67
19	59	210 (0.187)	5.75
20	30	112(0.187)	3.9
20	25	95 (0.19)	2.34
20	22	84 (0.19)	1.83

96%

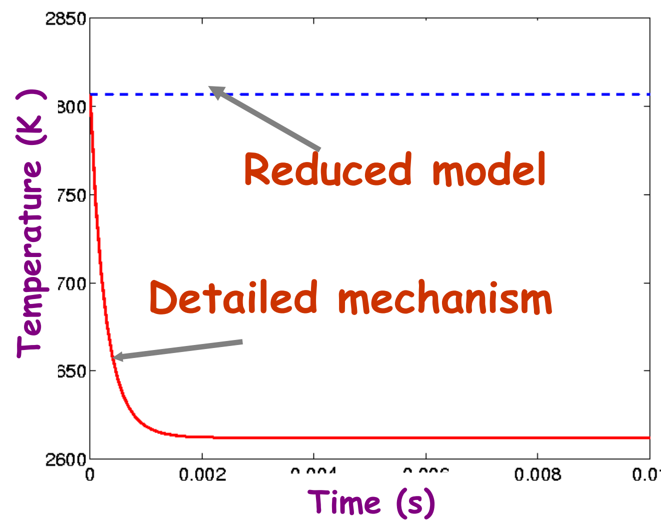
50%

Reduced Model has Limited Range of Validity





Nominal condition



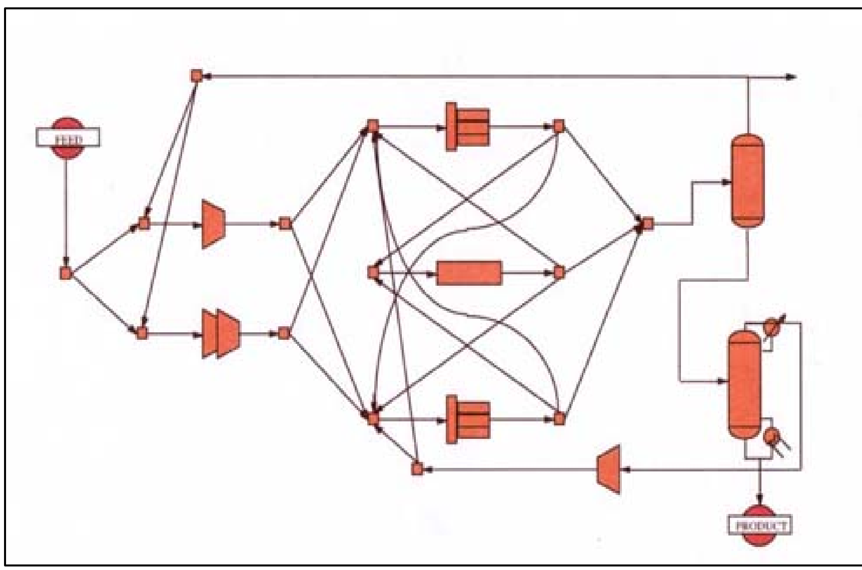
Outside feasible range

- Complexity reduction using mathematical programming approaches
  - Optimization of hepatocyte functionality
  - Reduction of complex chemistry
- Uncertainty analysis & feasibility evaluation
- Analysis of alternative solutions

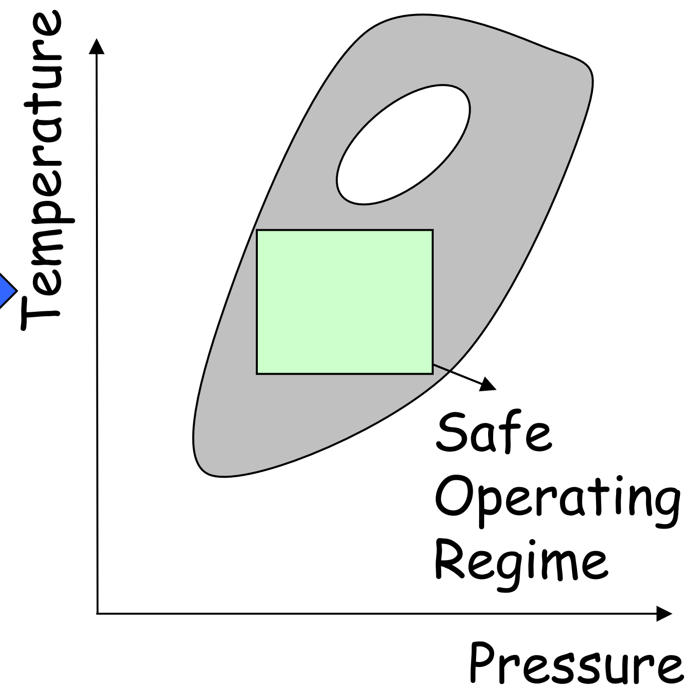
Feasibility Quantification

Given a design/plant or process

Determine the range operating conditions for safe and productive operations



Design

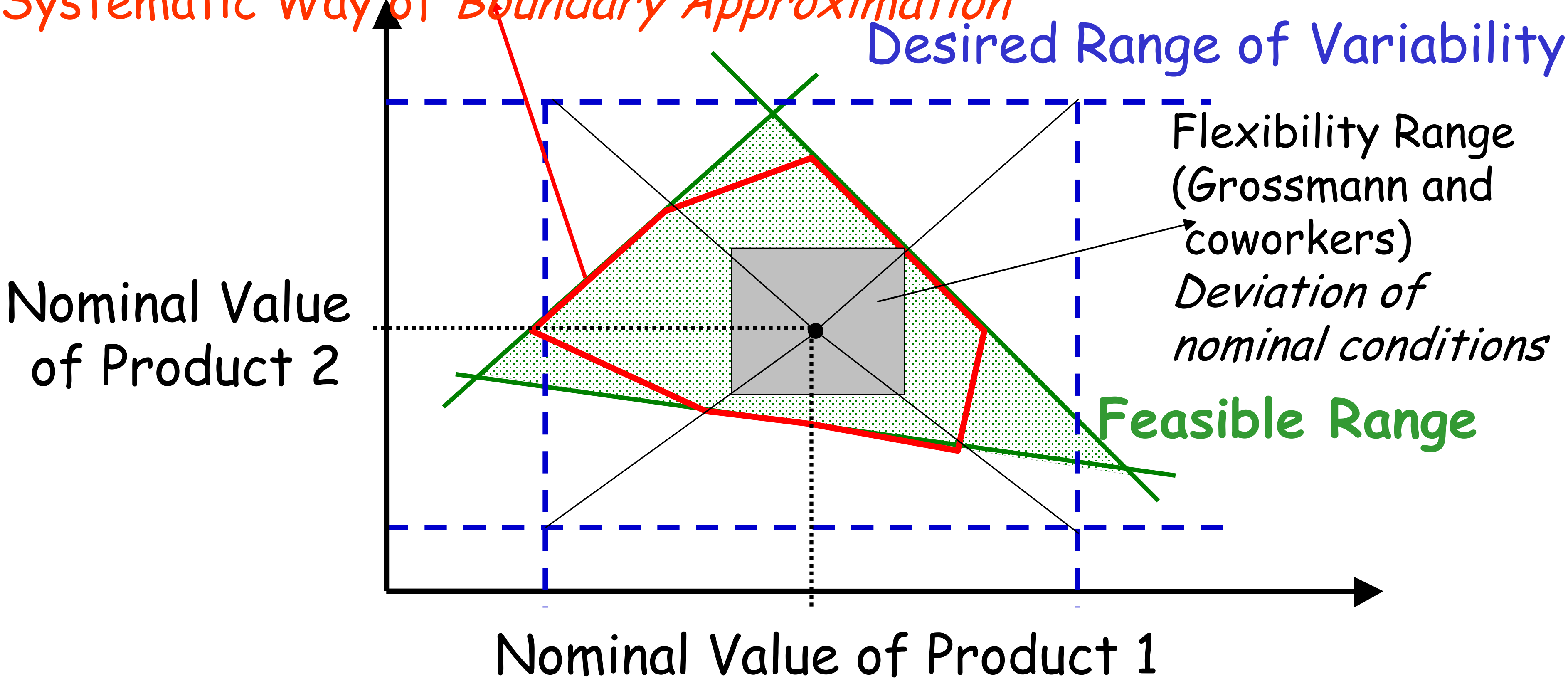


Feasibility Quantification

Convex Hull Approach

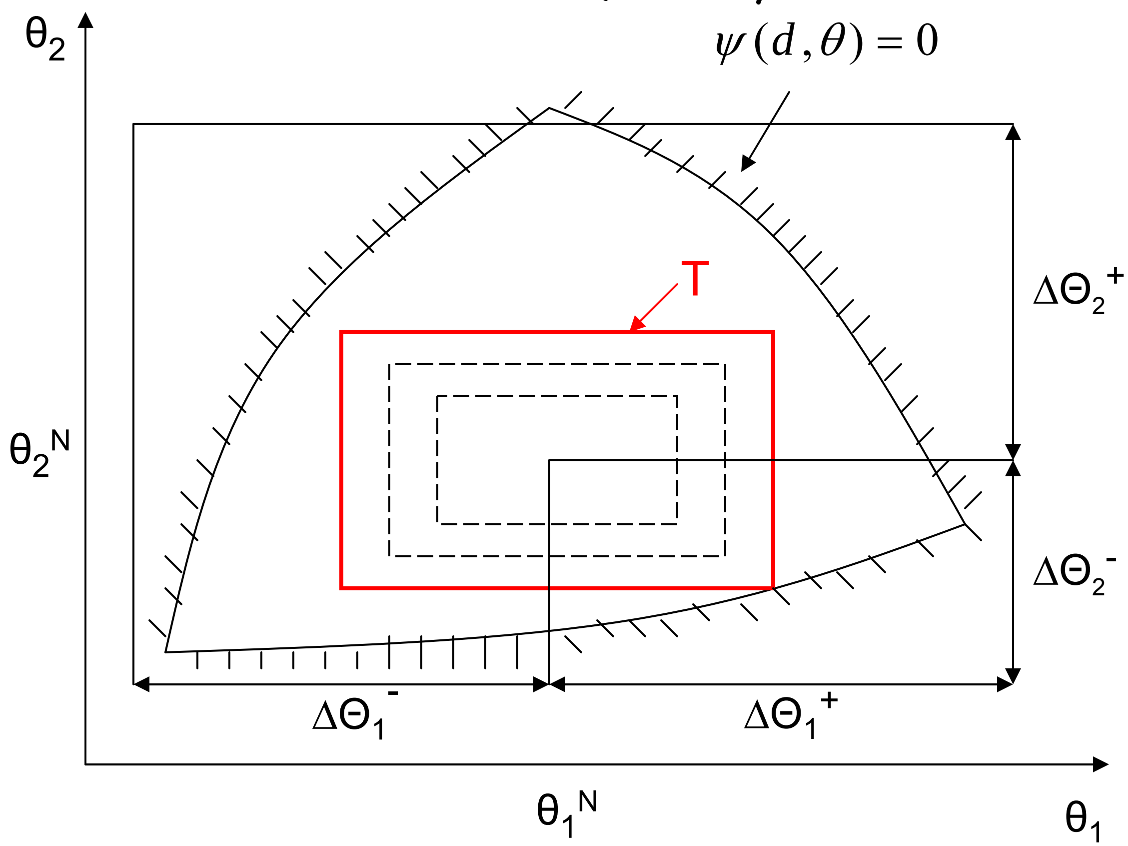
(Ierapetritou, *AIChE J.*, **47**, 1407, 2001)

Systematic Way of *Boundary Approximation*



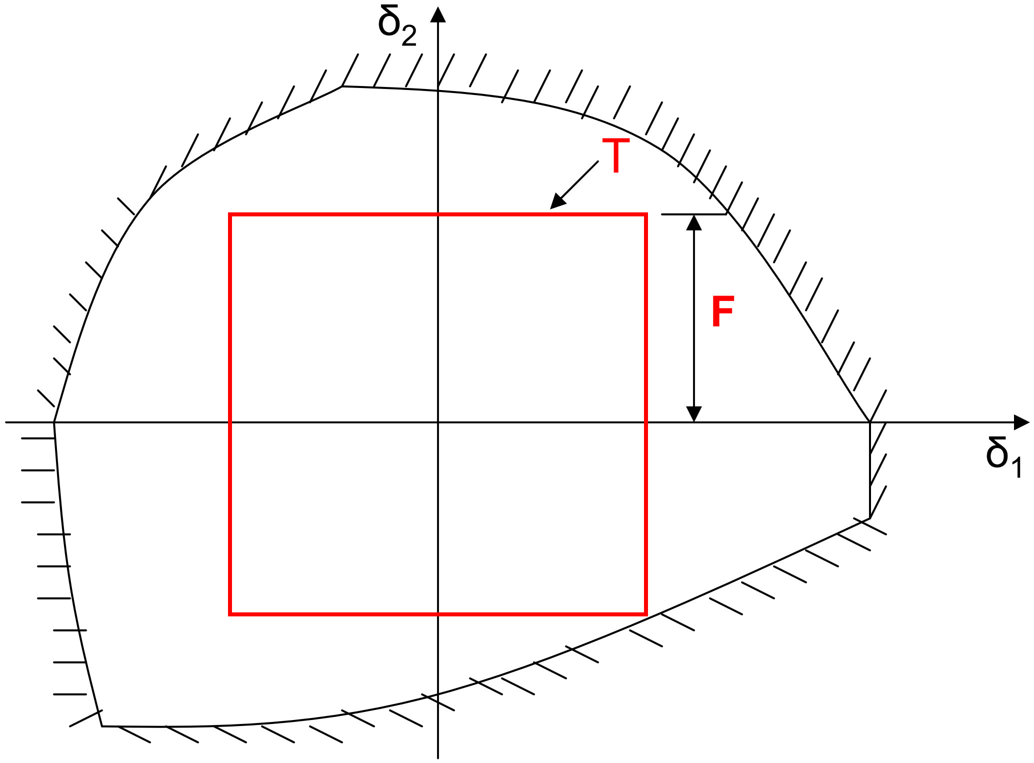
Process Flexibility

A powerful, approach available to identify the uncertainty ranges where the design, process or material is feasible to operate or function. (Swaney & Grossmann 1985)



$$\delta_j^+ = \frac{\theta_j - \theta_j^N}{\Delta\theta_j^+} \qquad \delta_j^- = \frac{\theta_j^N - \theta_j}{\Delta\theta_j^-} \qquad j = 1,..,\ p$$

Flexibility Index



Flexibility Index F – one-half the length of the side of hypercube T

Feasible operation can be guaranteed for $\theta^N - F\Delta\theta^- \le \theta \le \theta^N + F\Delta\theta^+$

Mathematical Formulation

(Swaney & Grossmann 1985)

$$s.t.\quad FI = \max \delta$$

$$\boxed{\max_{\theta \in T(\delta)} \min_{z} \max_{i \in I} \quad f_i(d, z, \theta) \le 0}$$ Feasibility test

$$T(\delta) = \{\theta \mid \theta^N - \delta\Delta\theta^- \le \theta \le \theta^N - \delta\Delta\theta^+\}$$

$$FI = \min \delta$$

$$s.t. \qquad \psi(d, \theta) = 0$$

$$\psi(d, \theta) = \min_{z} u$$

$$h_i(d, z, x, \theta) = 0, i \in I$$

$$g_j(d, z, x, \theta) \le u, j \in J$$

$$T(\delta) = \{\theta \mid \theta^N - \delta\Delta\theta^- \le \theta \le \theta^N - \delta\Delta\theta^+\}$$

$$\delta \ge 0$$

Active Set Strategy

$$FI = \min \delta$$ (Grossmann & Floudas 1987)

$$s.t. \quad h_i(d,z,x,\theta) = 0, i = 1,\ldots,I$$

$$g_j(d,z,x,\theta) + s_j = 0, j = 1,\ldots,J$$

Inner problem is replaced by KKT constraints

$$\boxed{\begin{aligned}
&\sum_{j=1}^{J} \lambda_j = 1 \\
&\sum_{j=1}^{J} \lambda_j \frac{\partial g_j}{\partial z} + \sum_{i=1}^{I} \mu_i \frac{\partial h_i}{\partial z} = 0 \\
&\lambda_j - w_j \le 0 \\
&s_j - U(1 - w_j) \le 0 \\
&\sum_{j=1}^{J} w_j = n_z + 1
\end{aligned}}$$

$$\theta^N - \delta\Delta\theta^- \le \theta \le \theta^N - \delta\Delta\theta^+$$

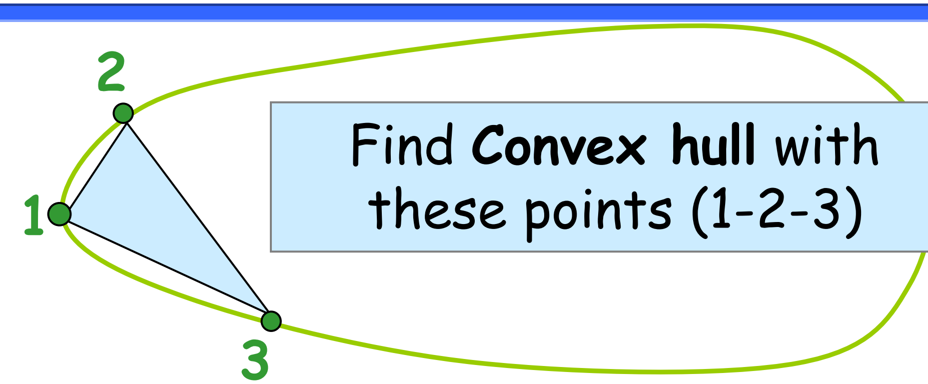
$$w_j = \{0,1\},$$

$$\delta, \lambda_j, s_j \ge 0$$

Simplicial Approximation

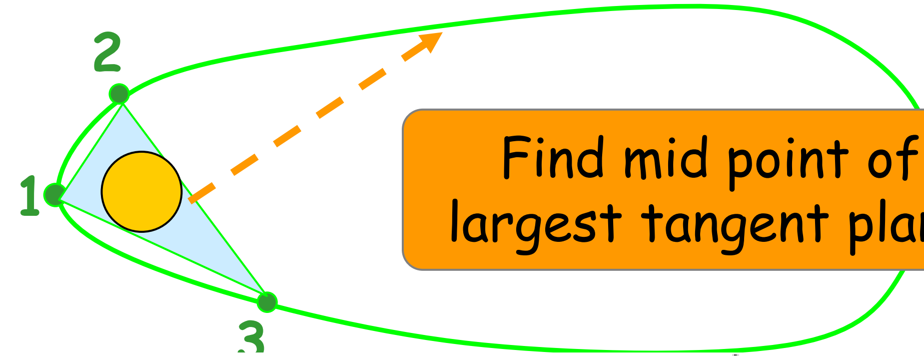
Choose $m \geq n+1$ points for n dimensions (points 1,2,3)

Find **Convex hull** with these points (1-2-3)

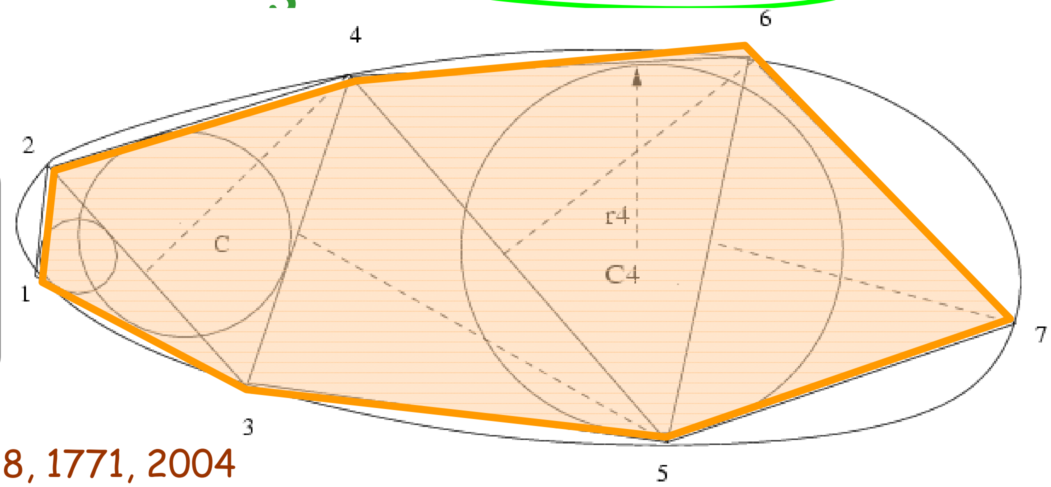


Insert the ***largest*** hypersphere in the convex hull

Find mid point of largest tangent plane

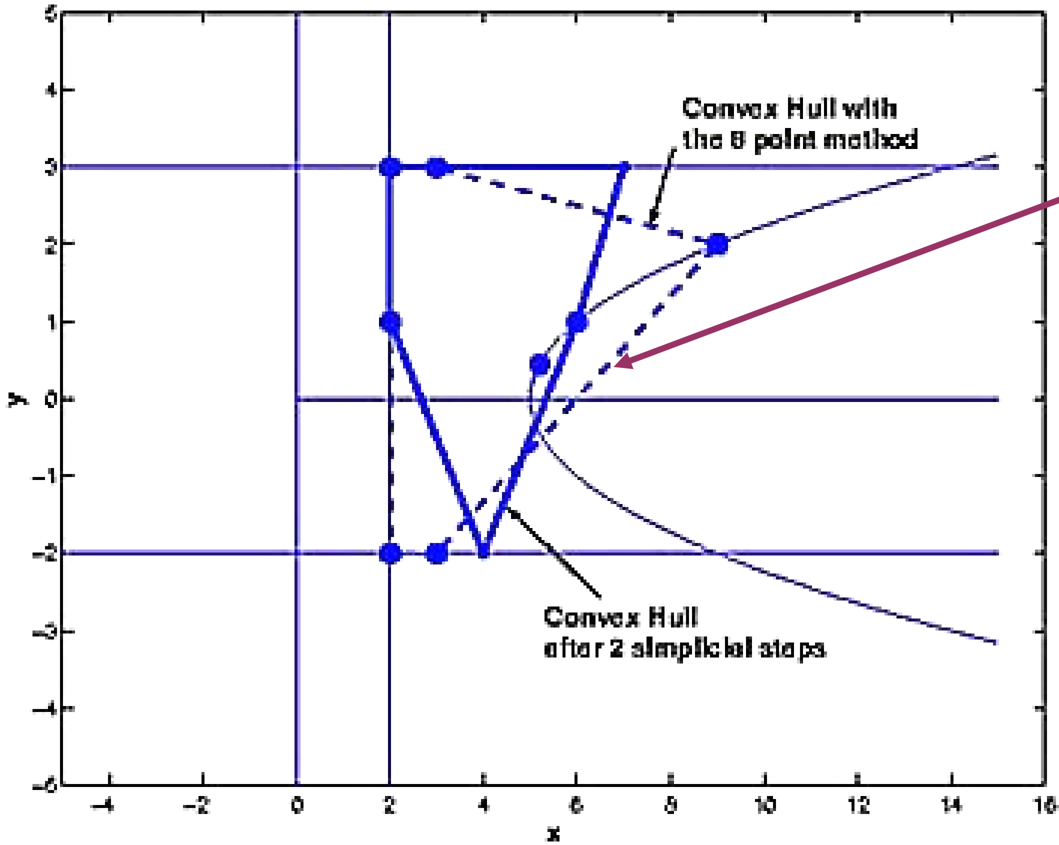


After 4 iterations Approximate Feasible Region1-2-3-4-5-6-7



Goyal and Ierapetritou, Comput. Chem. Engng. 28, 1771, 2004

Noncovex Problems: Need for Alternative Methods



Failure of Existing Methods due to Convexity Assumptions

Assumption: The Non-Convex Constraints can be identified a priori

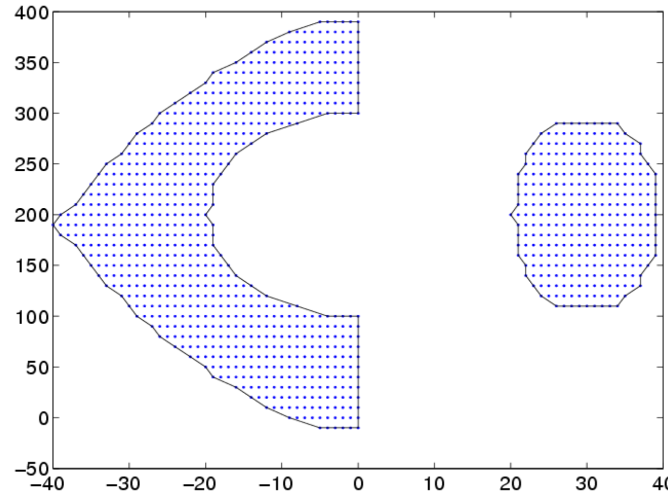
Improved Feasibility Analysis: Shape Reconsntruction

Problem definition :

Given a set of points (sample feasible points), determine mathematical representation of occupied space

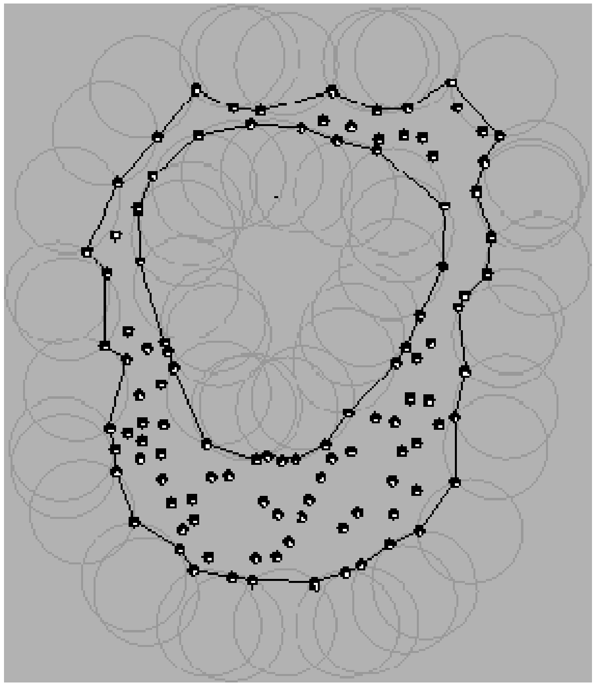
ion

shape formed by these points



Alpha-shape method:

Eliminate maximum possible circles of radius α without eliminating any data point

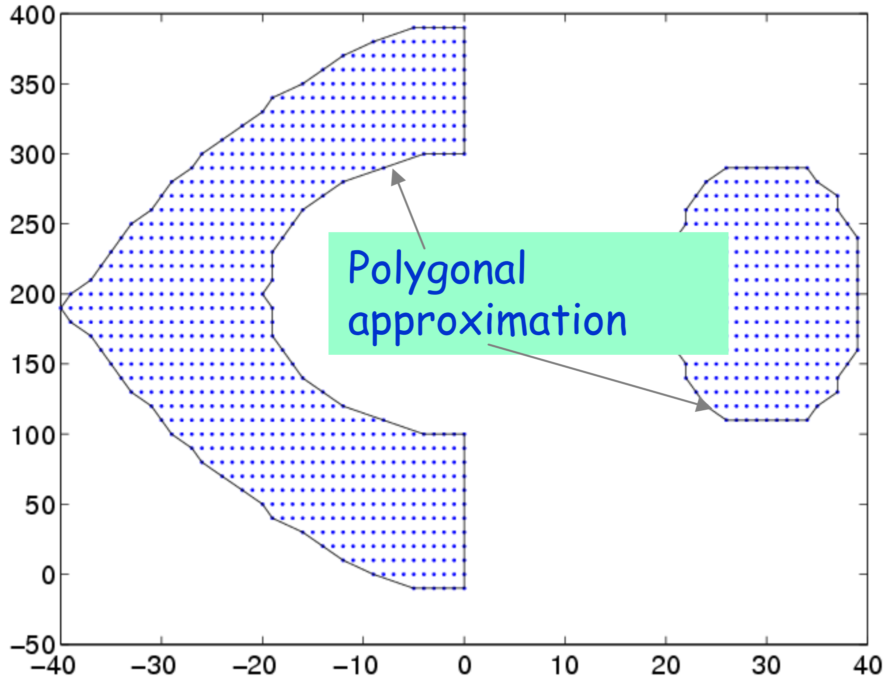


For $\alpha \rightarrow 0$ the α shape degenerates to the original point set

For $\alpha \rightarrow \infty$ the α shape is the convex hull of the original point set

(Ken Clarkson http://bell-labs.com/netlib/voronoi/hull.html)

Improved Feasibility Analysis by α - Shapes



Disjoint nonconvex object

Conventional techniques of inscribing hyper-rectangle or convex hull performs poorly

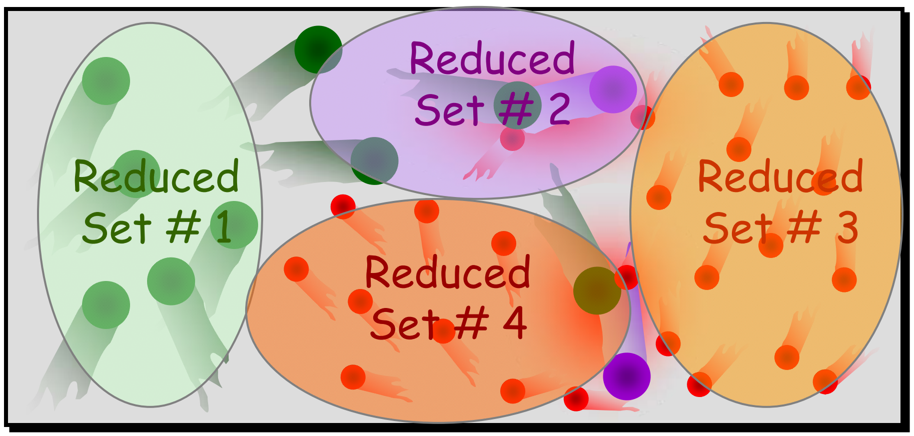
Identify boundary points using α shape

Connect boundary points to form a polygon

Banerjee and Ierapetritou, Ind. Eng. Chem. Res., 44, 3638, 2005.

Adaptive Reduction

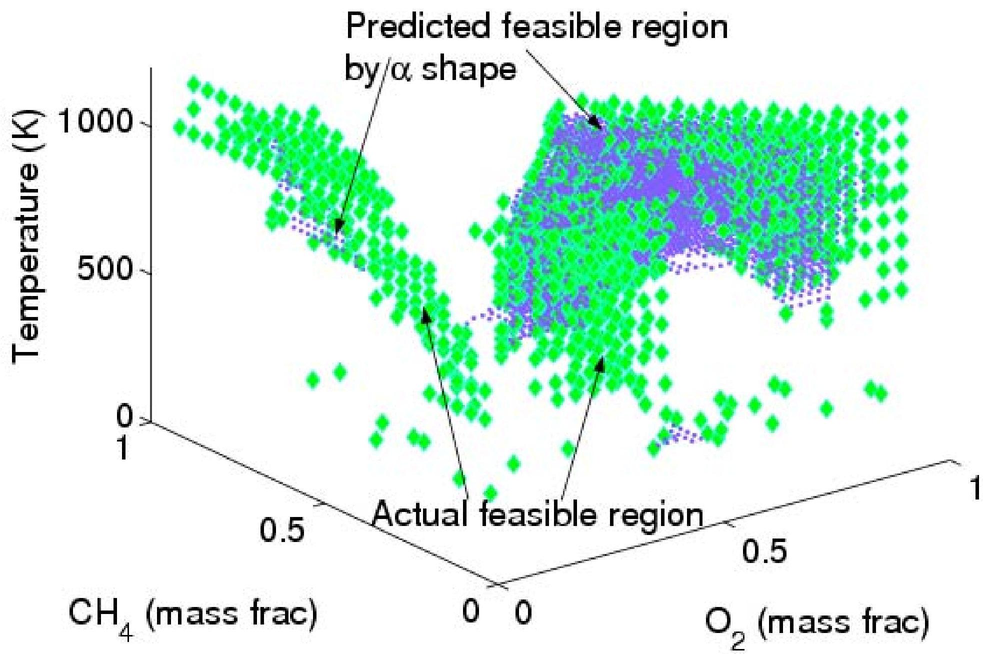
Broad range of species concentration and temperature encountered in flow simulation



Different reduced models for different conditions encountered in flow simulation

Banerjee and Ierapetritou, Comb. Flame, 144, 219, 2006.

Estimation of Feasible Region: α –shape

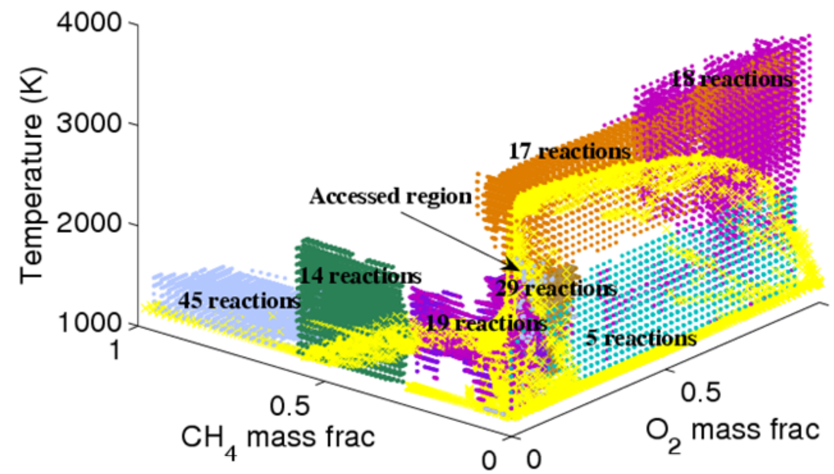


Sample the feasible space

Construct α – shape with the sampled points

Determine points forming the boundary of the feasible region

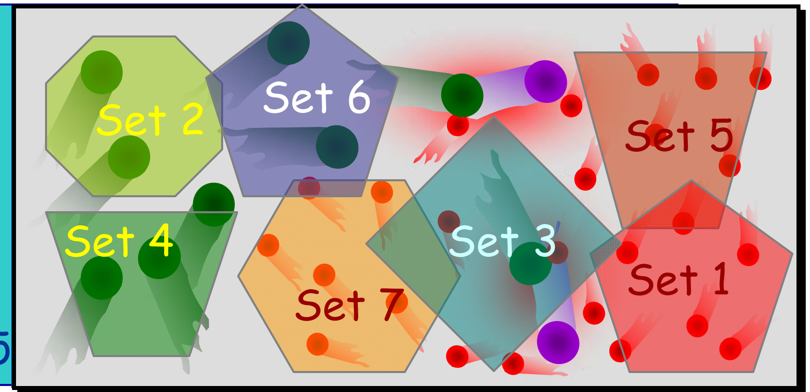
20 reduced sets



Generate Library of Reduced Model

Library of reduced models

Model 1	Species = 22, Reactions = 81
Model 2	Species = 19, Reactions = 59
Model 3	Species = 20, Reactions = 22
	:
	:
Model 20	Species = 53, Reactions = 325



Determine
$\lambda_1, \lambda_2, \ldots, \lambda_{ns}$

$y_1, y_2, \ldots, y_{ns}$, T
(Checks for a feasible model)

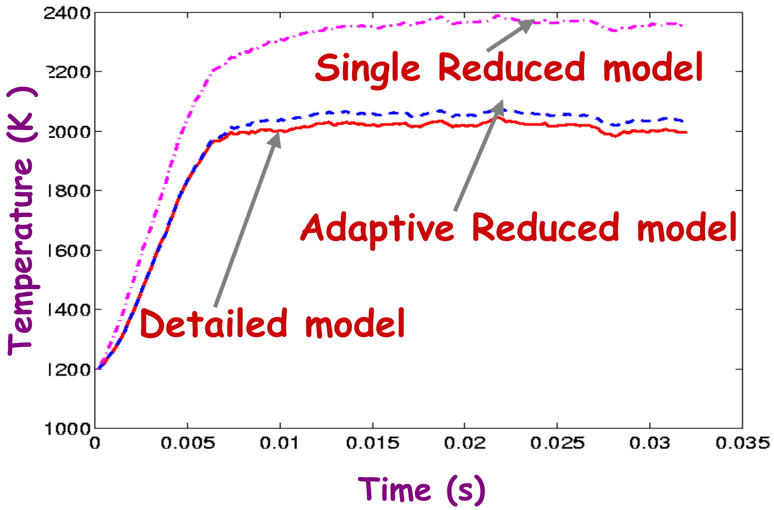
Reactive flow model

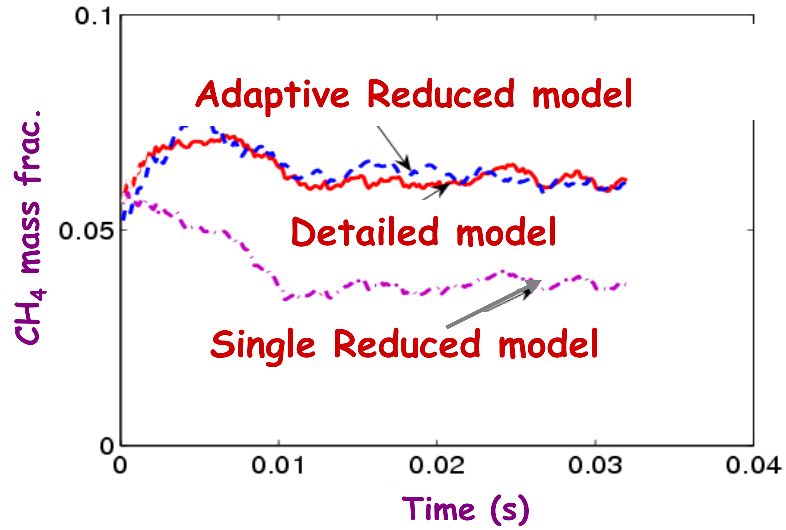
$$\frac{\partial W}{\partial t} + \frac{\partial F}{\partial x} = S,$$

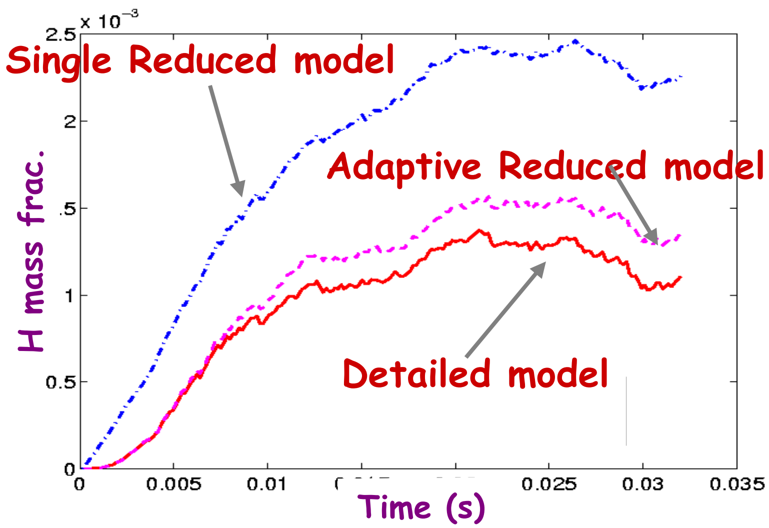
$$W = \begin{pmatrix} \rho \\ \rho u \\ E_t \\ \rho y_1 \\ \rho y_2 \\ \vdots \\ \rho y_{n_s} \end{pmatrix}, \quad F = \begin{pmatrix} \rho u \\ \rho u^2 + P \\ (E_t + P)u \\ \rho u y_1 \\ \rho u y_2 \\ \vdots \\ \rho u y_{n_s} \end{pmatrix}, \quad S = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \lambda_1 \rho \omega_1 \\ \lambda_2 \rho \omega_2 \\ \vdots \\ \lambda_{n_s} \rho \omega_{n_s} \end{pmatrix}$$

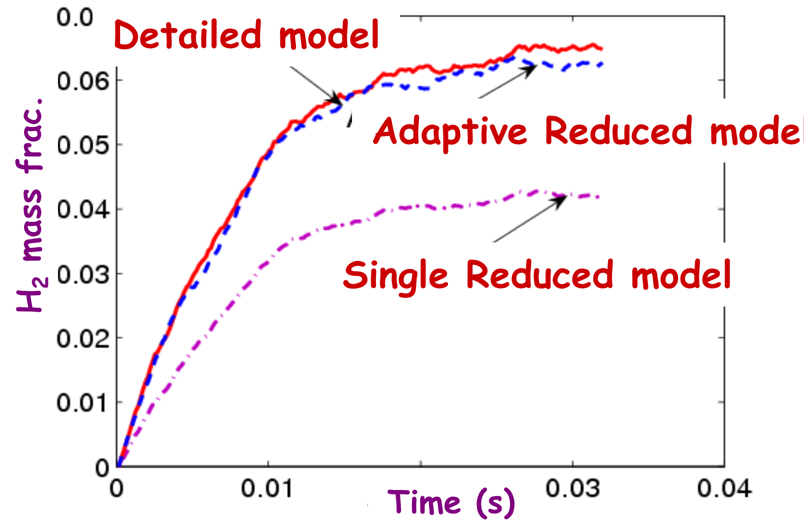
Adaptive Reduction Model in PMSR Simulation

Single reduced model : 38% error   Adaptive reduced model : 3% error









RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

Uncertainty in kinetic parameters

- Uncertainty inherent in kinetic parameter data

$$k_{f,i} = A_i T^{\beta_i} \exp\left(\frac{-E_{a,i}}{R_g T}\right)$$

- Commonly characterized by
  - Error bounds ($\Delta \log k_{f,i}$, ΔE_i etc.), confidence intervals/ranges.
  - Multiplicative Uncertainty Factor (UF $\geq$1)
    Upper bound = UF*$k_{f,i}$,
    Lower Bound = $k_{f,i}$/UF

Objective: Development of an accurate, systematic and efficient framework of analysis, that characterizes uncertainty in kinetic mechanisms

Representation of Uncertainty

- Classical/Rough Set Theory, Fuzzy Measure/Set Theory, Interval Mathematics and
- Probabilistic/Statistical Analysis
  - Sensitivity Testing Methods
  - Analytical Methods
    - Differential Analysis e.g. Perturbation Methods
    - Green's Function Method
    - Spectral Based Stochastic Finite Element Method forms the basis of the Stochastic Response Surface Method (SRSM)
  - Sampling Based Methods e.g.
    - Monte Carlo Methods
    - Latin Hypercube Methods

Stochastic Response Surface Method

- Extension of classical deterministic Response Surface Method and newer Deterministic Equivalent Modeling Method
- The outputs are represented as a polynomial chaos expansion in terms of Hermite polynomials:

$$U_1 = a_{0,1} + \sum_{i=1}^{n} a_{i,1}\xi_i \qquad \text{1}^{\text{st}} \text{ order}$$

$$U_2 = a_{0,2} + \sum_{i=1}^{n} a_{i,2}\xi_i + \sum_{i=1}^{n} a_{ii,2}(\xi_i^2 - 1) + \sum_{i=1}^{n-1}\sum_{j>i}^{n} a_{ij,2}\xi_i\xi_j \qquad \text{2}^{\text{nd}} \text{ order}$$

- Allows for direct and probabilistic evaluation of statistical parameters of the outputs e.g., for the second order output U_2: Mean = $\alpha_{0,2}$ Variance = $a_{1,2}^2 + 2a_{11,2}^2$

Stochastic Response Surface Method

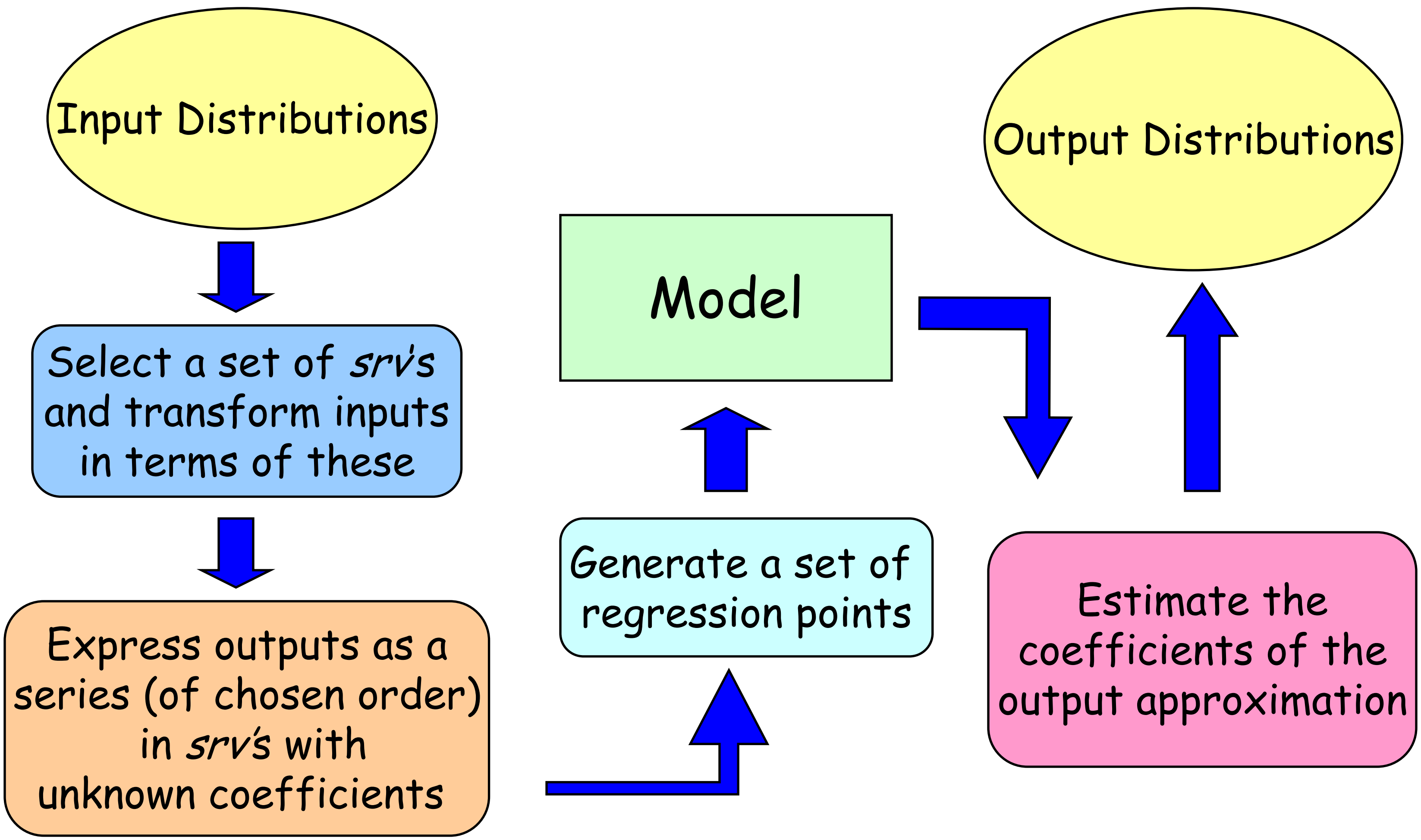
- Method Outline:
- Choice of order of expansion and transformation of the set of parametric input uncertainties in terms of a set of standard random variables (srv's) ξ's - Gaussian (N(0,1)). Commonly encountered transformations include :

Distribution Type	Transformation
Uniform (a,b)	$a+(b-a)\left(\frac{1}{2}+\frac{1}{2}erf(\xi/2)\right)$
Normal (μ,σ)	$\mu + \sigma\xi$
Lognormal (μ,σ)	$\exp(\mu+\sigma\xi)$
Exponential	$-\frac{1}{\lambda}\log\left(\frac{1}{2}+\frac{1}{2}erf(\xi/2)\right)$

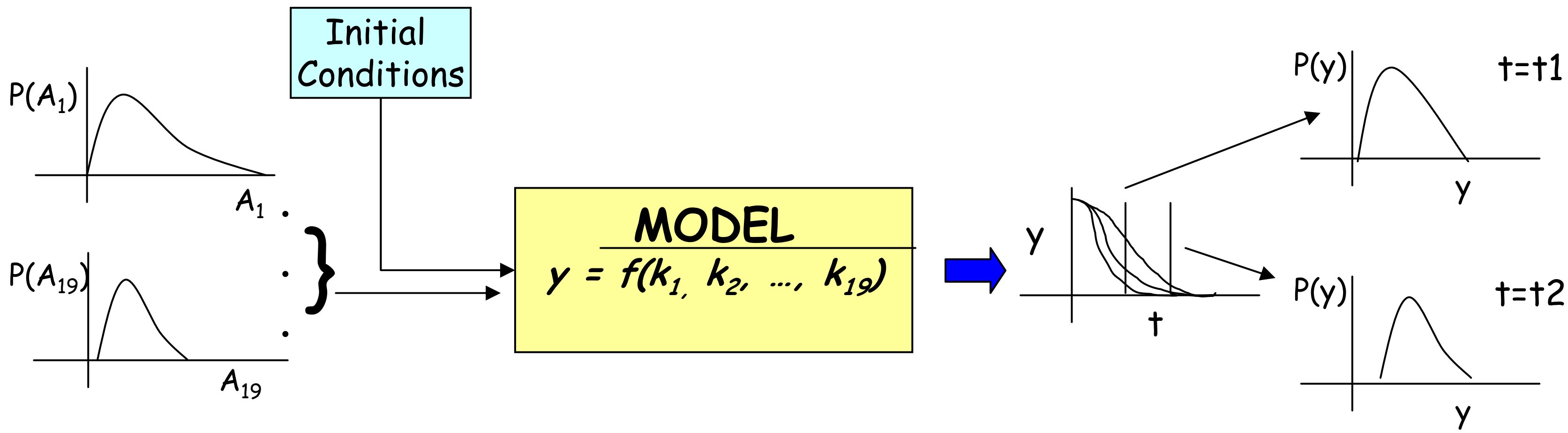
Stochastic Response Surface Method

- Generation of input points following the Efficient Collocation Method (ECM)
  - Points are selected from the roots of Hermite polynomials of higher order than the expansion
  - Borrows from Gaussian quadrature
- Application of the model to these input points and computation of relevant model outputs
- Estimation of the unknown coefficients of the expansion via regression using singular value decomposition (SVD)
- Statistical and direct analysis of the series expression of the outputs

SRSM - Algorithm



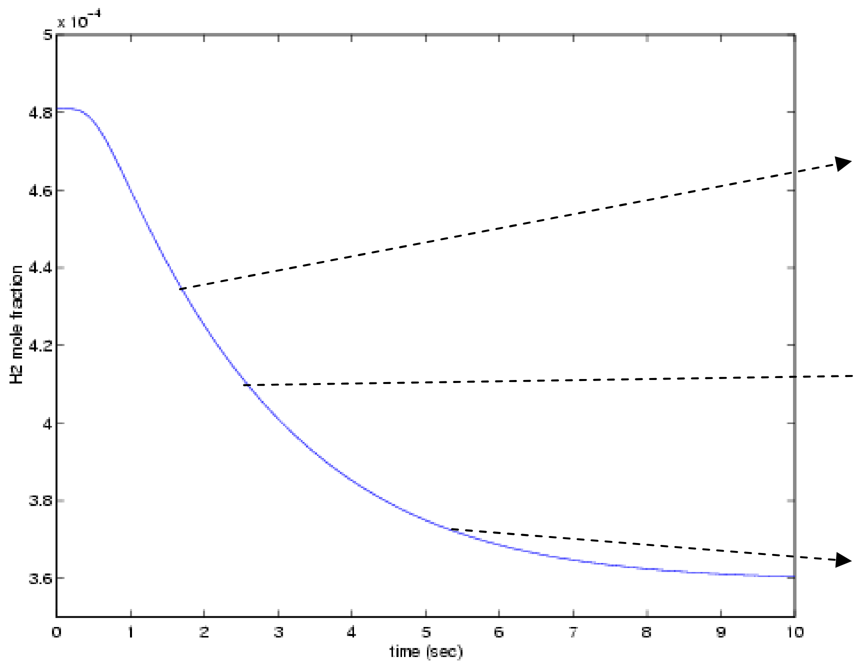
Implementation



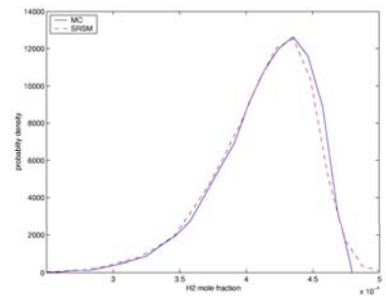
- Discretization of time interval
- 2nd order SRSM expansion fit for each output species at each time point

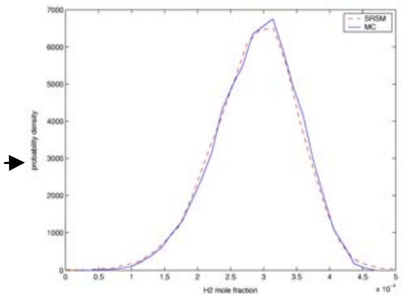
Uncertainty Propagation: Results

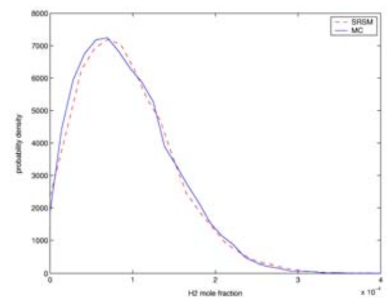
- Concentration profiles display time varying distributions
- Number of model simulations required by SRSM is orders of magnitude less than Monte Carlo (723 vs. 15,000)



Nominal H_2 mole fraction vs. time plot



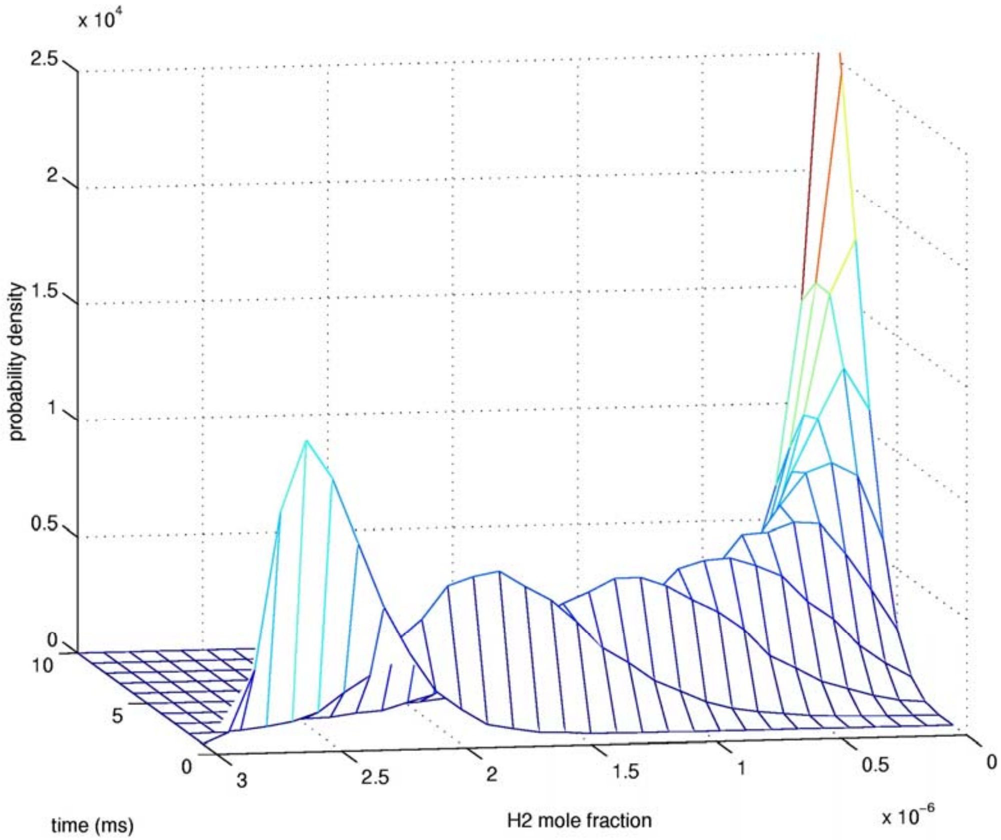




Distributions of H_2 at t=1, 2 and 5 seconds generated by Monte Carlo (MC) simulation and SRSM

Uncertainty Propagation: Results

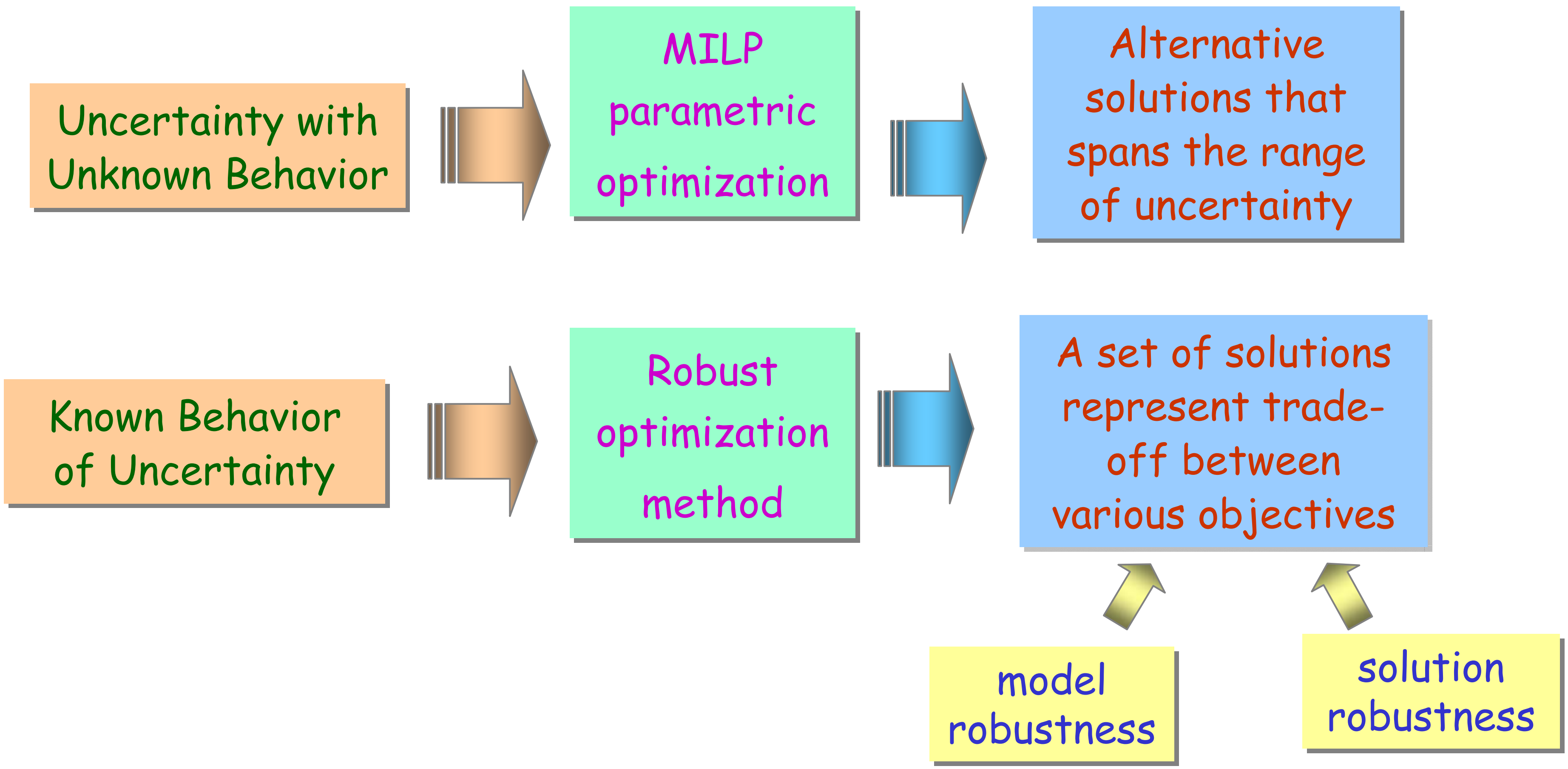
- Output distributions at each time point very well approximated by second order SRSM
- Sensitivity information easily obtained via expansion coefficients - aids understanding how the reaction sequence progresses
- Means for successfully preprocessing the **reduction** of the kinetic model taking into account uncertainty



Presentation Outline

- Complexity reduction using mathematical programming approaches
  - Optimization of hepatocyte functionality
  - Reduction of complex chemistry
- Uncertainty analysis & feasibility evaluation
- Analysis of alternative solutions

Determine a Set of Alternative Solutions



Li and Ierapetritou, Comp. Chem. Eng. in press, 2007 (doi:10.1016/j.compchemeng.2007.03.001).

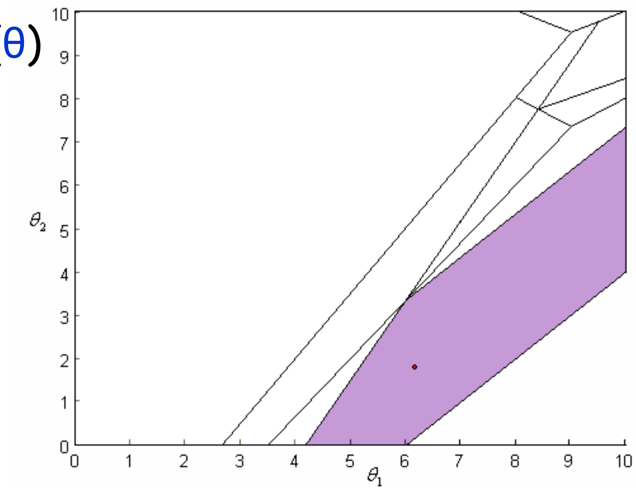
Multiparametric MILP (mpMILP) Approach

mpMILP problem generalized from scheduling under uncertainty

$$
\begin{aligned}
\min \quad & z = (c + \theta^T D)x \\
s.t. \quad & Ax \geq b + E\theta \\
& x \geq 0 \\
& \theta^l \leq \theta \leq \theta^u \\
& x_j \in \{0,1\},\ j = 1,\dots,k
\end{aligned}
$$

In any Critical Region of an mpMILP

* same integer solution
* same parametric objective: $z^* = f(\theta)$
* same parametric solution (continuous variable): $x^* = f(\theta)$

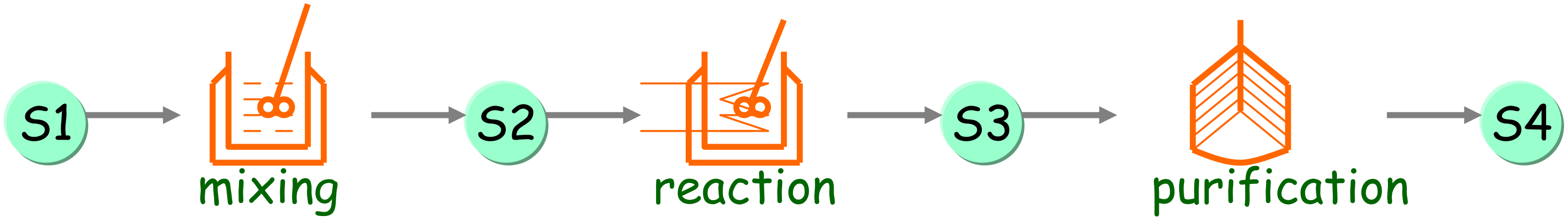


BASIC IDEA

* One critical region with one starting point
* Complete solution is retrieved with different starting points (parallelization)

Li and Ierapetritou, *AIChE Jl.* 53, 3183, 2007; Ind. Eng. Chem. Res. 46, 5141, 2007.

Illustrating Example



Given:

Raw Materials

Required Products

Production Recipe

Unit Capacity

Determine:

Task Sequence

Exact Amounts of material processed

Objective:

Maximize Profit

Ierapetritou MG, Floudas CA, 1998

$$\text{Profit} = 88.55 + 49.07\theta_1 - 0.25\theta_2 + 20\theta_1^2 - 1.2\theta_1\theta_2 + 0.01\theta_2^2$$

Demand and Price Only

Demand, Price, Processing Time Uncertainty

$$\text{Profit} = 88.55 + 44\theta_1 - 25.16\theta_2 - 20\theta_1^2$$

Uncertainty with Known Behavior: Robust Optimization

minimize H or maximize $\sum price(s)d(s,n)$ — Objective Function

subject to $\sum wv(i,j,n) \le 1$ — Allocation Constraints

$$st(s,n) = st(s,n-1) - d(s,n) + \sum \rho^P \sum b(i,j,n-1) + \sum \rho^c \sum b(i,j,n)$$
$$st(s,n) \le stmax(s)$$
— Material Balances

$$Vmin(i,j)wv(i,j,n) \le b(i,j,n) \le Vmax(i,j)wv(i,j,n)$$
— Capacity Constraints

$$\sum d(s,n) \ge r(s)$$
— Demand Constraints

$$Tf(i,j,n) = Ts(i,j,n) + \alpha(i,j)wv(i,j,n) + \beta(i,j)b(i,j,n)$$
$$Ts(i,j,n+1) \ge Tf(i,j,n) - U(1-wv(i,j,n))$$
$$Ts(i,j,n+1) \ge Tf(i',j,n) - U(1-wv(i',j,n))$$
$$Ts(i,j,n+1) \ge Tf(i',j',n) - U(1-wv(i',j',n))$$
$$Ts(i,j,n) \le H, \; Tf(i,j,n) \le H$$
— Duration Constraints

Ierapetritou MG, and Floudas CA, *Ind. & Eng. Chem. Res.*, 37, 11, 4341, 1998

- Scenario-based Robust Stochastic Programming
  - Requires some statistic knowledge of the input data
  - Optimization of expectations is a practice of questionable validity
  - Problem size will increase exponentially with the number of uncertain parameters

Robust Counterpart Optimization

Find solution which copes best with the various realizations of uncertain data

$$\tilde{a}_{lm} \in [a_{lm} - \hat{a}_{lm}, a_{lm} + \hat{a}_{lm}]$$

- **Soyster's,** Soyster (1973)

$$\sum_{m \notin M_l} a_{lm} x_m + \sum_{m \in M_l} (a_{lm} + \hat{a}_{lm}) u_m \le p_l - \hat{p}_l$$

- **Ben-Tal and Nemirovski's,** Ben-Tal and Nemirovski (2000); Lin, Janak et al. (2004)

$$\Pr\left\{ \sum_m a_{lm} x_m + \sum_k b_{lk} y_k \ge p_l + \delta \max[1, |p_l|] \right\} \le \kappa, \kappa = e^{-\Omega^2/2}$$

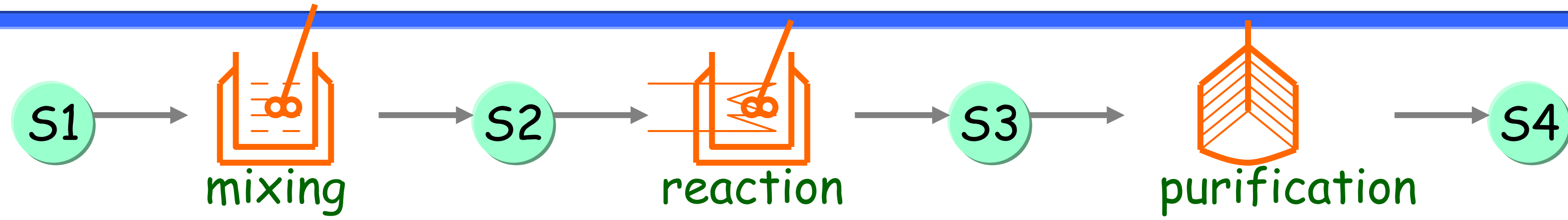
- **Bertsimas and Sim's,** Bertsimas and Sim, 2003

$$\sum_m a_{lm} x_m + \max_{\{S_l \cup \{t_l\} | S_l \subseteq M_l, |S_l| = \lfloor \Gamma_l \rfloor, t_l \in M_l \setminus S_l\}} \left\{ \sum_{m \in S_l} \hat{a}_{lm} |x_m| + (\Gamma_l - \lfloor \Gamma_l \rfloor) \hat{a}_{lt_l} |x_{t_l}| \right\} \le p_l$$

Soyster's	Ben-Tal and Nemirovski's	Bertsimas and Sim's
- Linear	- Nonlinear	- Linear
- No flexibility	- Flexibility	- Higher flexibility
- Most pessimistic	- Relative smaller number of variables and constraints	- Relative larger number of variables and constraints

Efficient alternative to scenario based robust stochastic programming

Illustration



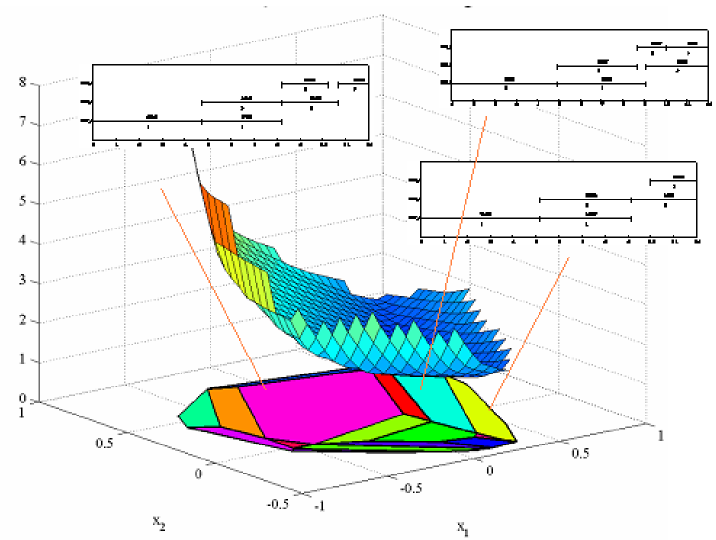
Comparison for the robust courterpart formulations for processing time uncertainty

	Deterministic	Soyster	Ben-Tal	Bertsimas and Sim		
				$\Gamma=0$	$\Gamma=0.5$	$\Gamma=1$
objective	1052.50	939.12	-	1052.50	1005.5	939.12
Probability of constraint violation	-	-	k=75%	p≤0.75	p≤0.625	p≤0.5
CPU time	4.2	150	infeasible	4.8	43.4	254
Continuous variable	625	625	625	700		
Binary variable	216	216	216	216		
constraints	1167	1167	1167	1239		

- 15% variability for all the processing time
- 72 hours horizon, 24 event points

Parametric and Robust Solution

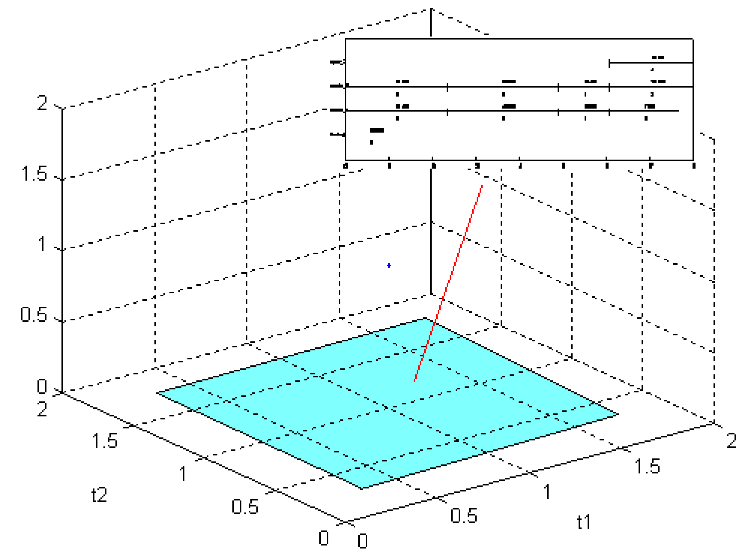
➢ **Parametric Solution**



∗Study the effect of the different uncertainties

∗ Provide an efficient way to look up the reactive schedule with the realization of uncertainty (e.g., rush order, machine breakdown)

➢ **Robust Counterpart Solution**

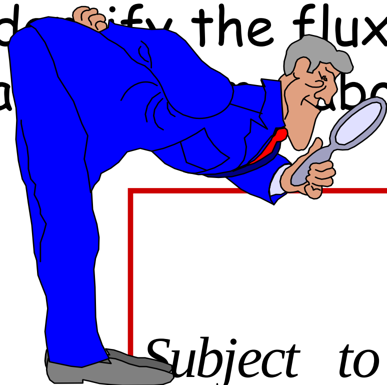


∗ Provide an effective way to generate robust preventive schedule with boundary information on uncertainty (e.g., processing time variability)

Uncertainty in Hepatocyte Functionality

- How can we use these techniques to deal with experimental variability?
  - In many cases experimental error is more than 100%
- How can we analyze the results?
  - Is the results an artifact of uncertainty?
- How can we move beyond experimental error?
  - Can we determine which parameters are more important and what experiment to do next?

Single-level Optimization: Maximize Urea Secretion



Aim: Id[illegible]ify the flux distributions for optimal urea production that ca[illegible]abolites balances and flux constraints

$$Max: Z = v_{urea}$$

$$Subject \quad to: \sum_{j=1}^{N} S_{ij} v_j = 0 \qquad \forall i \in M$$

$$v_j^{\min} \le v_j \le v_j^{\max} \qquad \forall j \in K$$

Experimental Data*		Optimal Value	Increase
HIP	0.23±0.43	6.81	> 10 fold
HPAA	1.32±0.69		> 3 fold
LIP	0.17±0.24		> 15 fold
LPAA	2.35±0.52		> 2 fold

Unit: μmol/million cells/day

**Chan & Yarmush et al (2003) Biotechnol Prog*

Example of Multiple Solution in 2-D

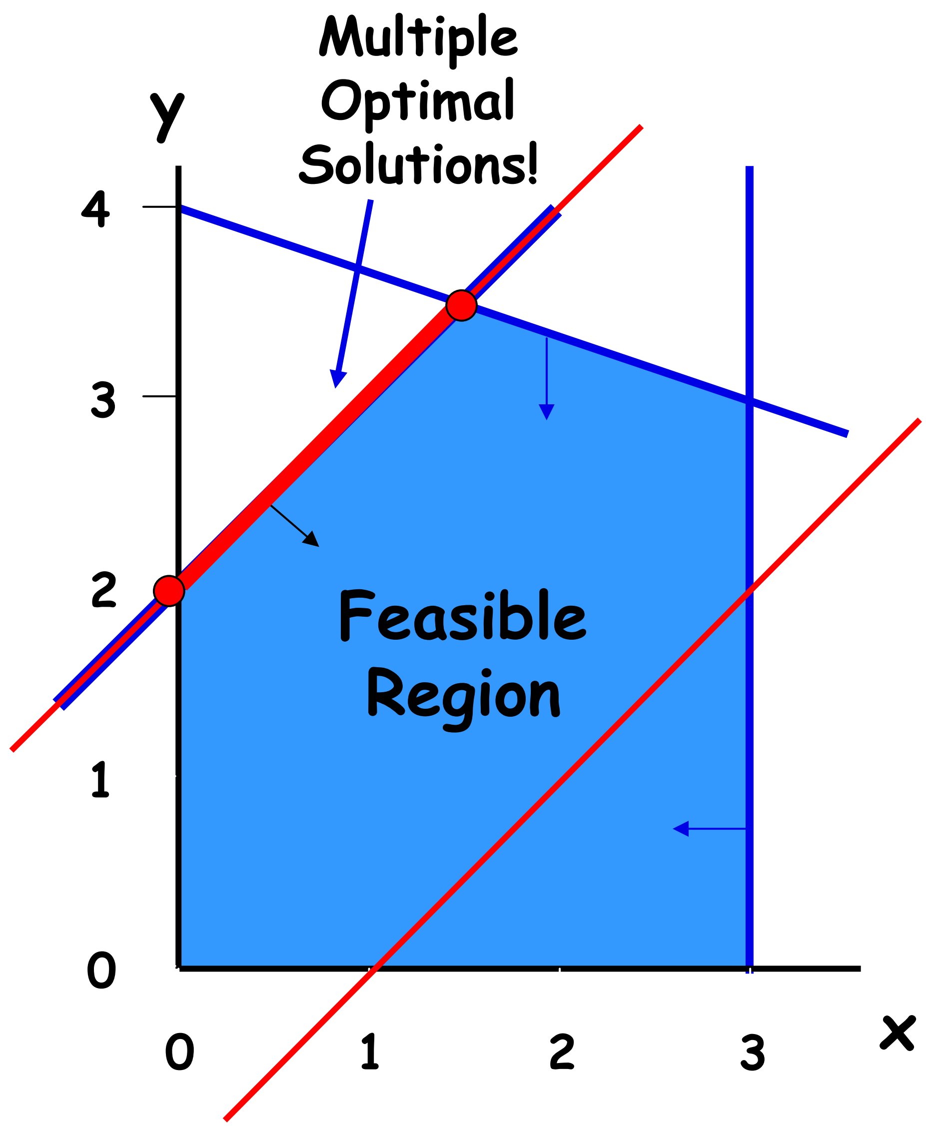
Minimize $\quad x - y$

Subject to: $\quad 1/3\ x + y \leq 4$

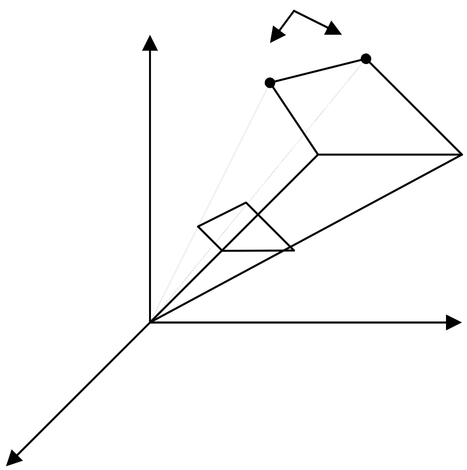
$-2\ x + 2\ y \leq 4$

$x \leq 3$

$x \geq 0 \quad y \geq 0$



Finding all Solutions



Question: How can you determine all solutions?

A recursive MILP problem that has a set of constrains for changing the basis and identifying a new extreme point

Nathan D, Price et al., 2004, Nature Review: Microbiology

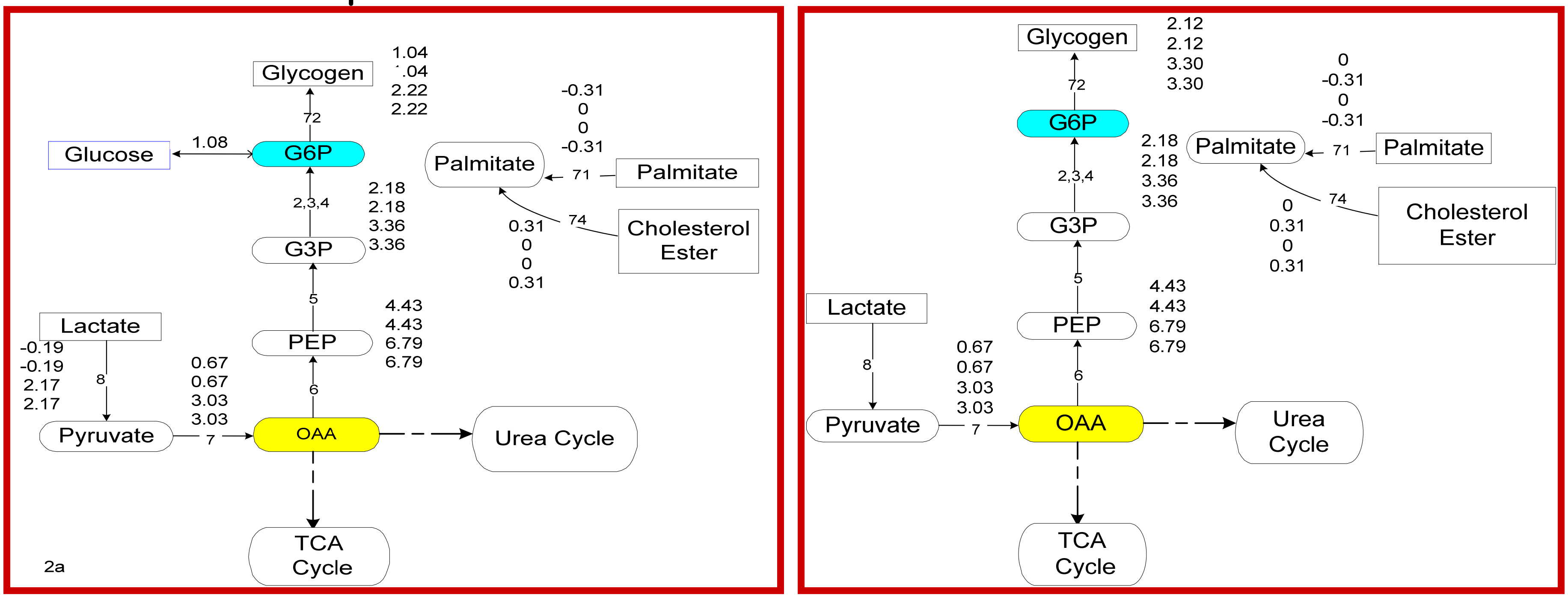
$$
\begin{aligned}
\min Z &= \alpha^T z \\
s.t. Bz &= q \\
z &\geq 0
\end{aligned}
$$

$$
\begin{aligned}
&\min Z^K = \alpha^T z \\
&s.t. Bz = q \\
&\sum_{i \in NZ^{K-1}} y_i \geq 1 \\
&\sum_{i \in NZ^k} w_i \leq \left|NZ^k\right| - 1, k = 1, 2, \ldots, K-1 \\
&0 \leq z_i \leq U w_i, i \in I \\
&y_i + w_i \leq 1, i \in NZ^{K-1} \\
&z \geq 0 \qquad \textbf{(MILP)}
\end{aligned}
$$

Lee et al., 2000, Computer and Chemical Engineering

MILP Model: Application to Hepatocytes

Enumerate Eight different flux distributions flux distributions that satisfy mass balance and all constraints with the same value of maximal urea production.



Flux distributions including glucose production (left) & without glucose production (right)

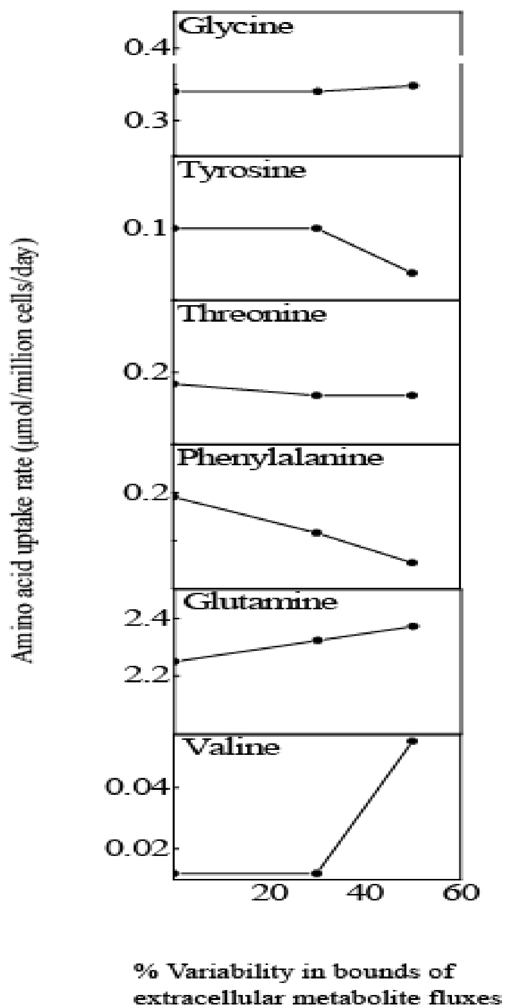
	D1	D2	D3	D4	D5	D6	D7	D8
R(1,72)	0.509	0.509	0.327	0.327	0	0	0	0
R6-R7	3.76	3.76	3.76	3.76	3.76	3.76	3.76	3.76
R(7,6)	0.178	0.178	0.806	0.806	0.178	0.178	0.806	0.806

$$R(7,6) = \frac{v_7}{v_6 - v_7}$$

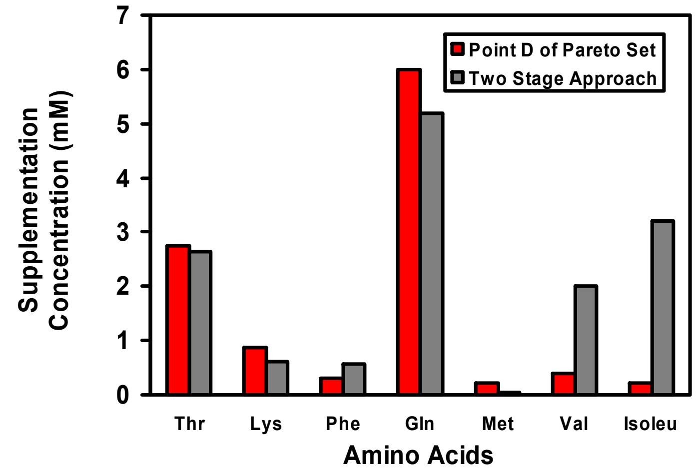
$$R(1,72) = \frac{v_1}{v_1 + v_{72}}$$

Develop Patient Specific Treatment

LINEAR VARIABILITY IN EXTRACELLULAR FLUXES



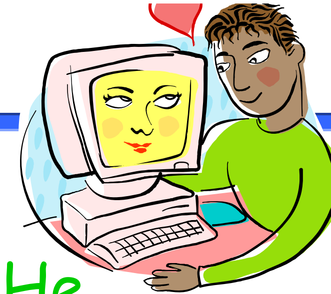
ROBUST SOLUTION CONSIDERING VARIABILITY



- All 19 amino acids are indispensable for maximum function
- Valine and Isoleucine are required at higher concentrations

** All fluxes are in μmol/million cells/day*

Acknowledgements



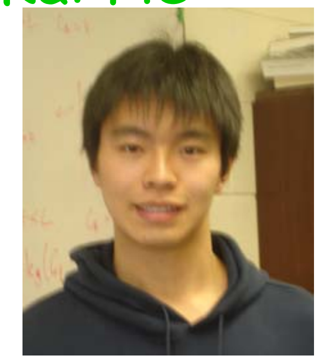
Beverly Smith

Hong Yang

Mehmet Orman

Eddie Davis

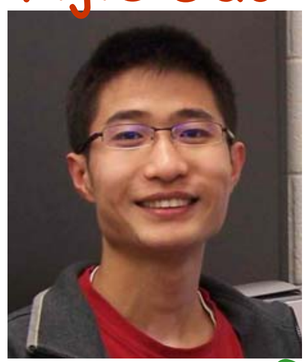
Kai He



Zukui Li

Patricia Portillo

Yijie Gao



George Saharidis

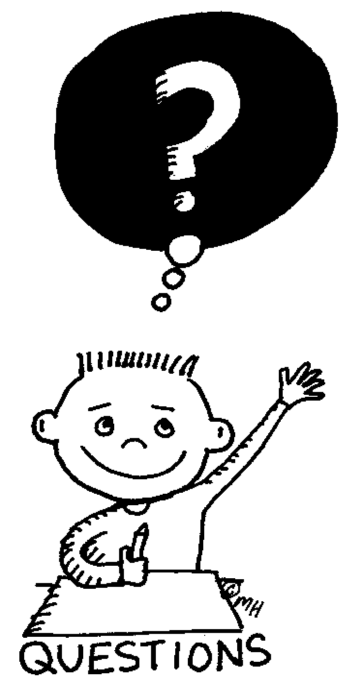
Zhenya Jia

Vidya Iyer

Financial Support:
NSF, ONR, PRF, ExxonMobil

Our Web Page:
http://sol.rutgers.edu/staff/marianth


Thanks!


QUESTIONS