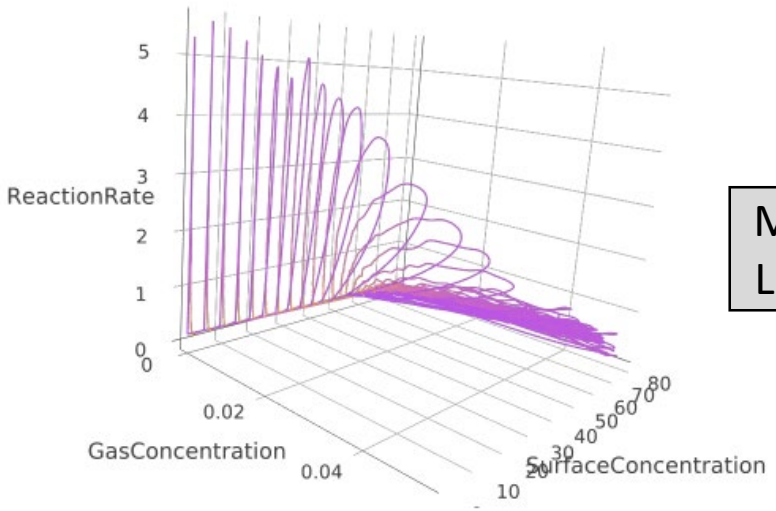
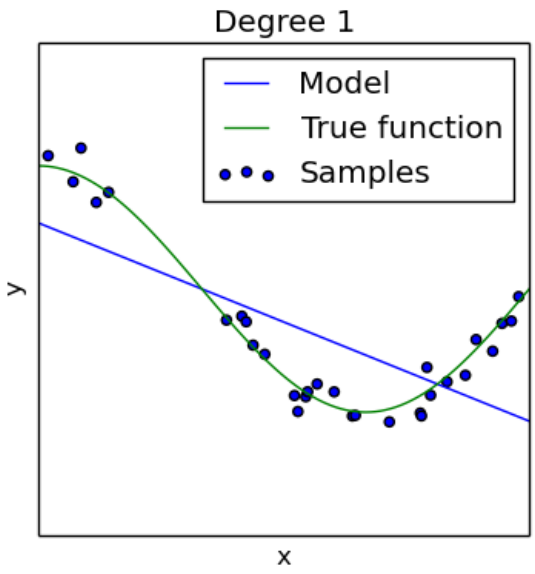
Extracting Knowledge for Industrial Catalysis through Machine Learning



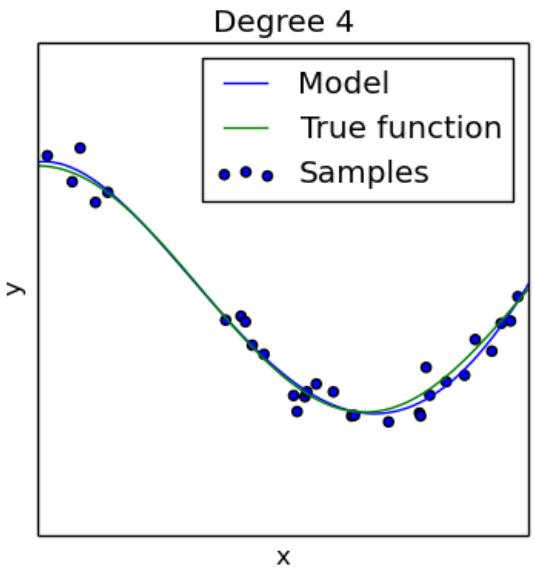
A.J. Medford
Assistant Professor
Dept. of Chemical & Biomolecular Engineering
Georgia Institute of Technology

02.26.20
Dynamic Catalyst Science Roundtable
University of Houston

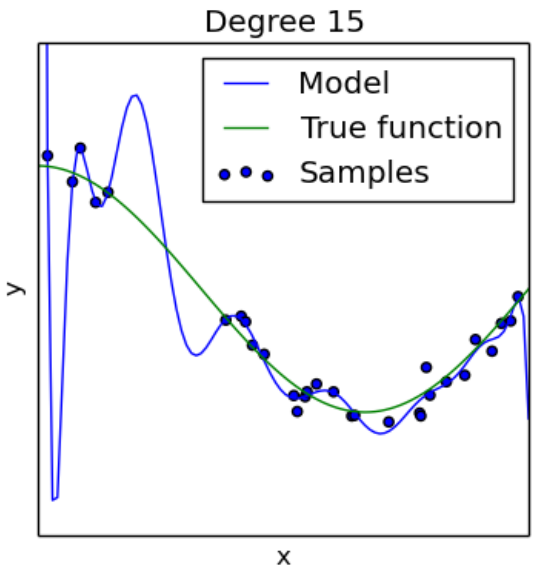
What is “machine learning”?



”Underfitting”
(Machine guessing)



Optimal complexity
(Machine learning)



”Overfitting”
(Machine memorizing)

Machine learning is the quantitative optimization of model complexity based on large amounts of observed data.

Naive applications of machine learning models are likely to fail in catalysis

Small diverse datasets with large uncertainty

Models with millions of parameters and no physics

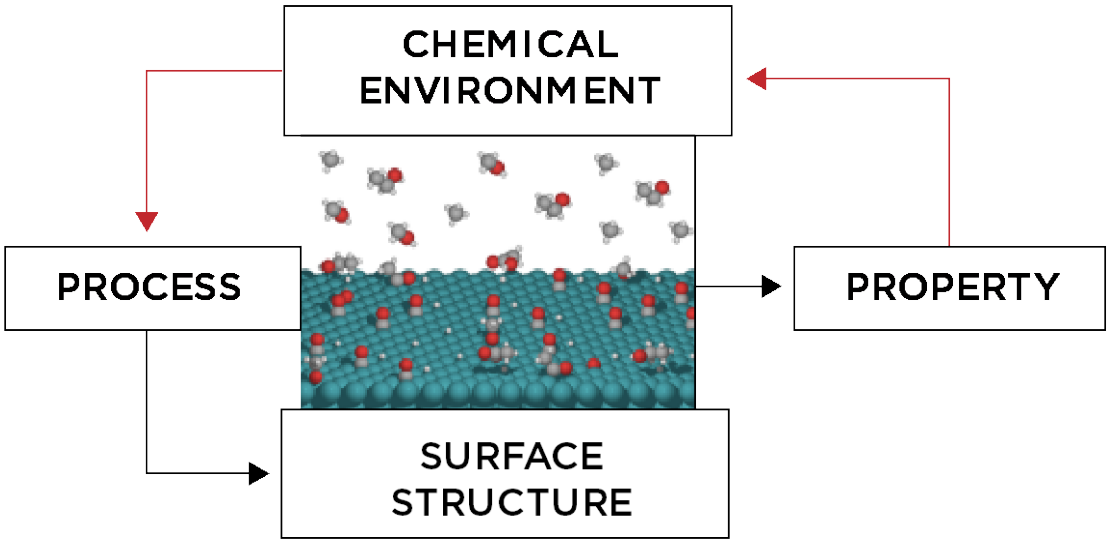
Magical discovery of new materials and processes

V	description	issue in catalysis?
volume	size of data	no
velocity	flux of data	no
variety	diversity of data	yes
veracity	uncertainty of data	yes
volatility	stability of data structure	yes

Large volumes of data from a consistent source are ideal for machine learning

A. J. Medford, M. R. Kunz, S. M. Ewing, T. Borders, and R. Fushimi, ACS Catalysis, vol. 8, no. 8, pp. 7403–7429, Jun. 2018.

Catalysis data has unique challenges



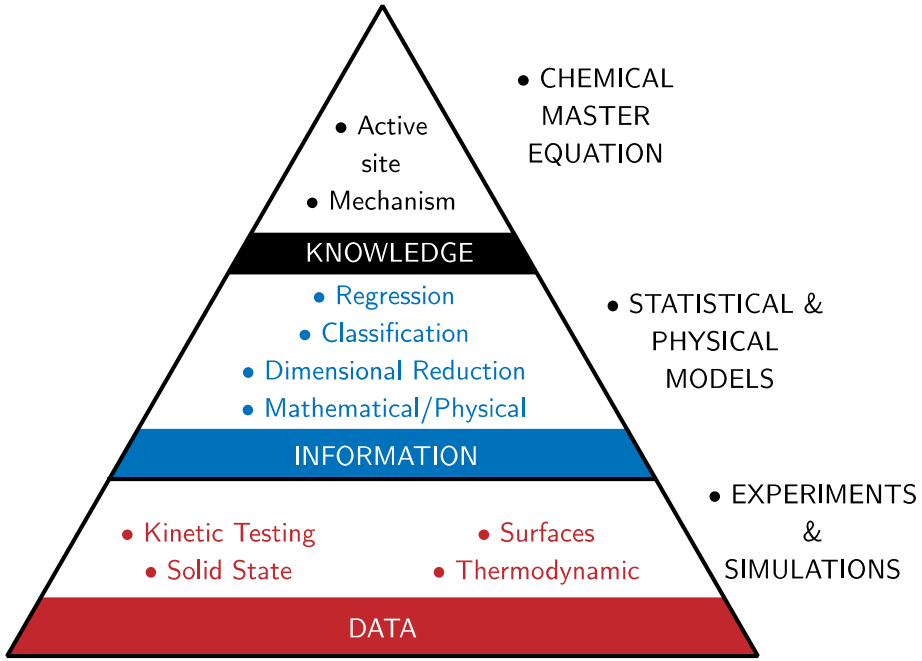
- Chemical environment data (cheminformatics)
  - Chemical potentials, molecular structures, etc.
- Reaction kinetics data
  - Reaction rates, activation barriers, etc.
- Surface science data
  - Oxidation state, adsorption energies, etc.
- Bulk materials data (materials informatics)
  - Material stability, composition, structure, etc.

- Process-Structure-Property paradigm fails for catalysis
  - Same material responds differently depending on environment
- Catalysis involves interaction of molecules + materials
  - Intersection of cheminformatics and materials informatics
- Catalysis is a dynamic phenomenon
  - Catalysts alter their environment, which can induce structure changes
- Catalytic active sites are often ”rare” defects
  - Catalytic activity can be sensitive to unknown or unquantified impurities

A. J. Medford, M. R. Kunz, S. M. Ewing, T. Borders, and R. Fushimi, ACS Catalysis, vol. 8, no. 8, pp. 7403–7429, Jun. 2018.

The chemical master equation quantifies catalysis knowledge

$$\frac{\mathrm{d}P(S_i)}{\mathrm{d}t} = \sum_j \left(A_{ji}P(S_j) - A_{ij}P(S_i)\right)$$



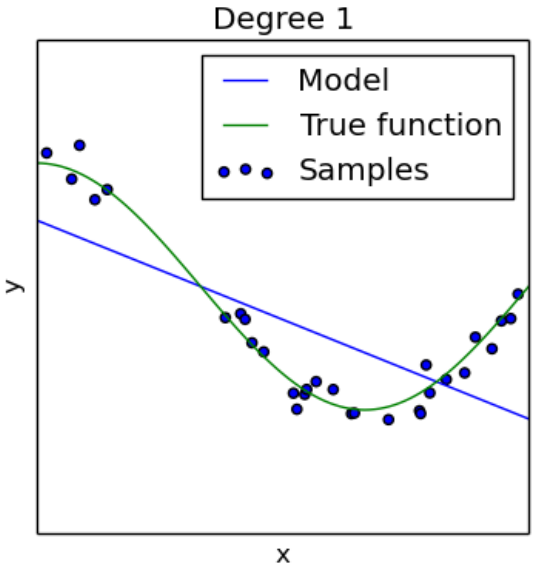
- Chemical master equation = knowledge

$$\frac{\mathrm{d}C_j}{\mathrm{d}t} = \underbrace{\sum_i s_{ij}}_{\text{mechanism}} \left(\overbrace{k_i^+}^{\text{active site}} \underbrace{\prod_r C_r}_{\text{mechanism}} - k_i^- \prod_p C_p \right)$$

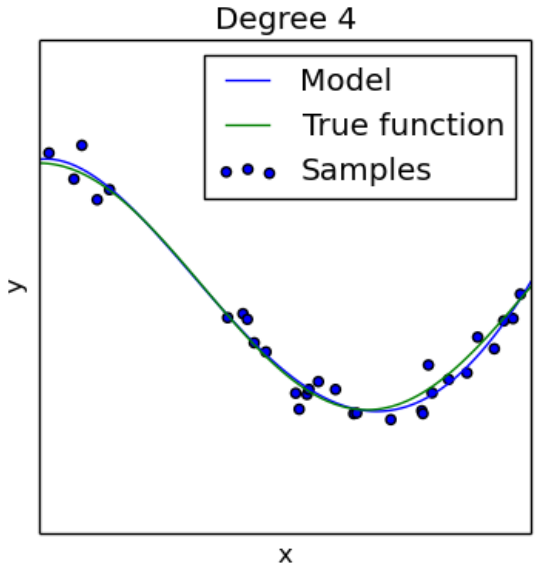
  - Active site(s)
  - Reaction mechanism(s)
- Data science and machine learning extract knowledge from data

A. J. Medford, M. R. Kunz, S. M. Ewing, T. Borders, and R. Fushimi, ACS Catalysis, vol. 8, no. 8, pp. 7403–7429, Jun. 2018.

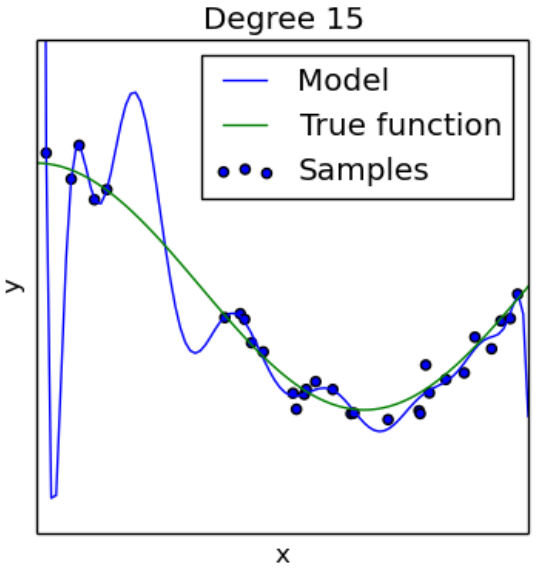
Machine learning for catalysis optimizes complexity of reaction model



Mechanism is too simple:
Data not fully explained



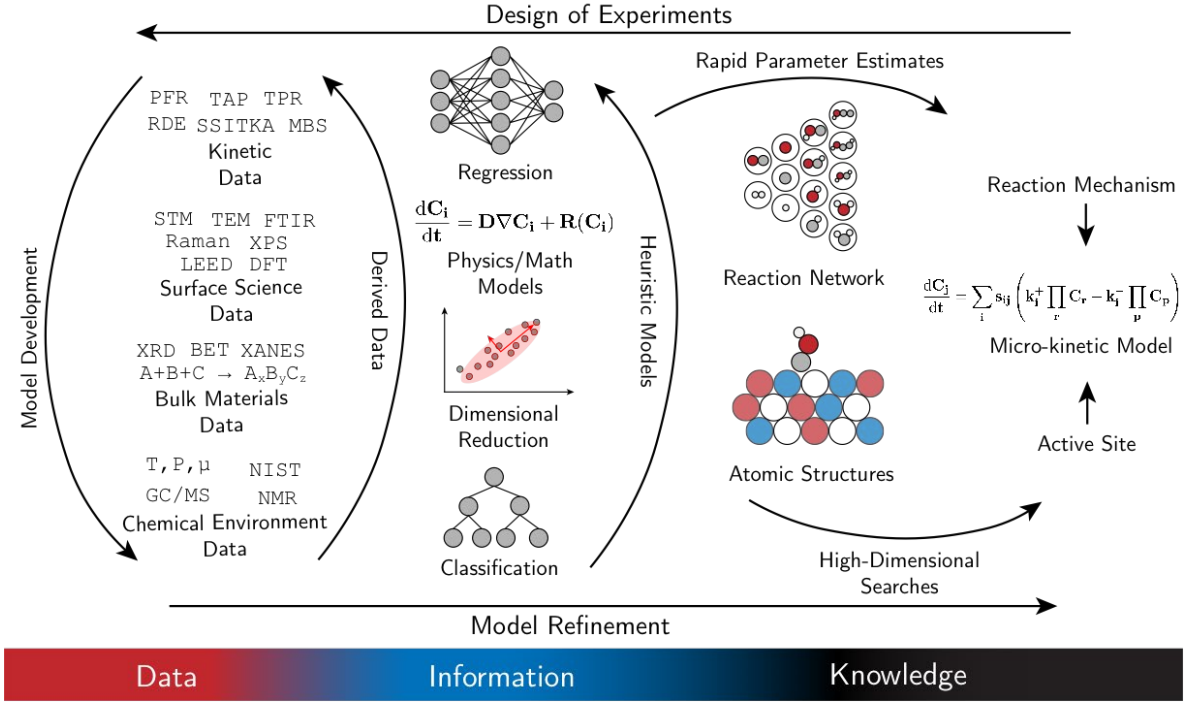
Mechanism correct:
Data is explained and model works at other conditions



Mechanism is too complex:
Model doesn't work at other conditions

Identifying (micro-)kinetic models with optimal complexity can facilitate robust process models that enable optimization of process conditions

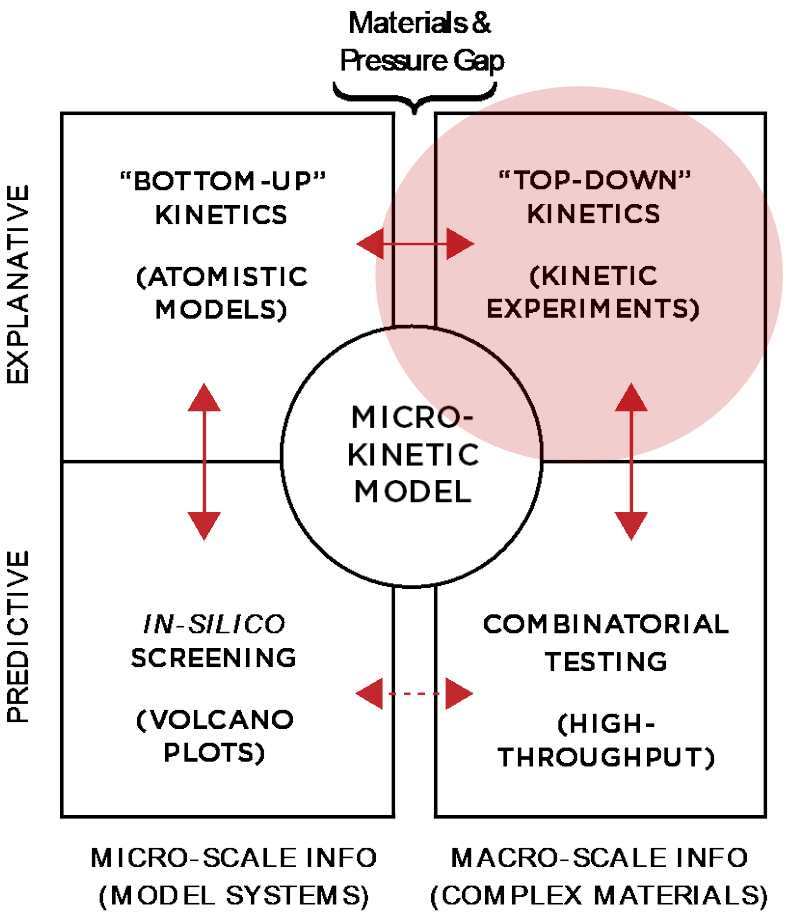
A variety of techniques and data are needed for atomic-scale knowledge



- Boundaries between data/info/knowledge are fuzzy
  - Derived data, TPR, etc.
- Knowledge extraction is a dynamic process
  - Design of experiments/calculations
  - Model refinement

A.J. Medford, R. Kunz, S. Ewing, T. Borders, R. Fushimi. ACS Catalysis - submitted

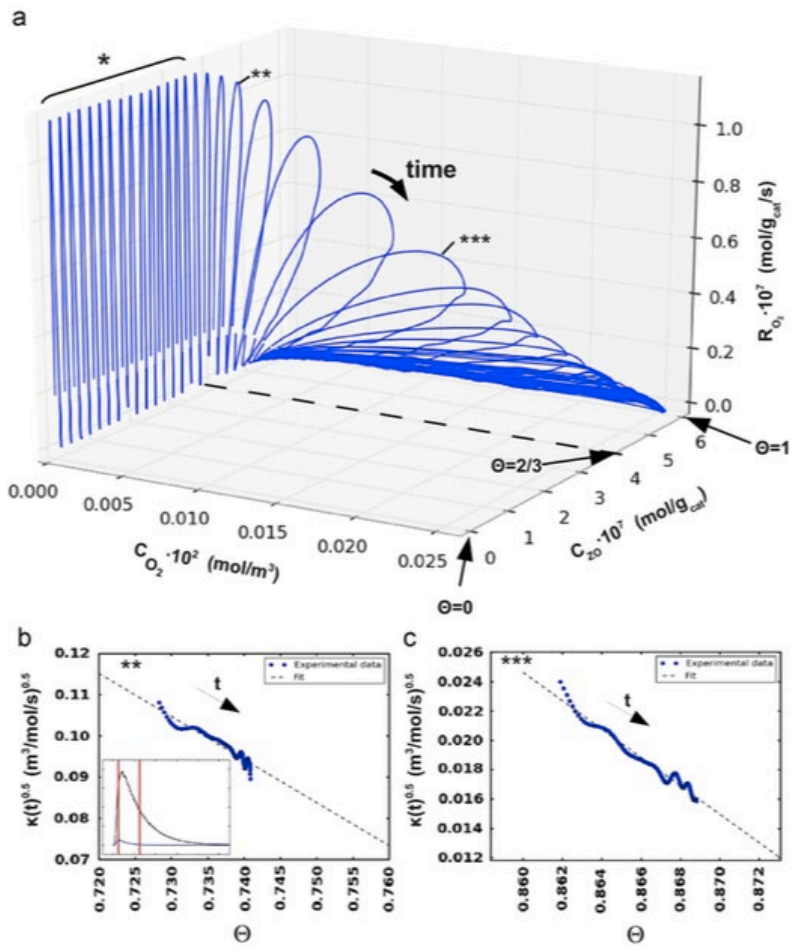
Catalysis “knowledge engines” seek to extract knowledge from data



- Catalysis “knowledge engines” were first proposed by Caruthers et. al. in 2004
  - Combine high-throughput experimentation with model fitting
- Recent advances improve feasibility
  - computational catalysis
  - machine learning
  - open-source development
  - data infrastructure
- Dynamic catalysis provides new opportunities for top-down kinetics

J. M. Caruthers, J. A. Lauterbach, K. T. Thomson, V. Venkatasubramanian, et. al., Journal of Catalysis, vol. 216, no. 1–2, pp. 98–109, May 2003.
A.J. Medford, R. Kunz, S. Ewing, T. Borders, R. Fushimi. ACS Catalysis - submitted

Dynamic catalysis provides large, consistent kinetic datasets for real catalysts



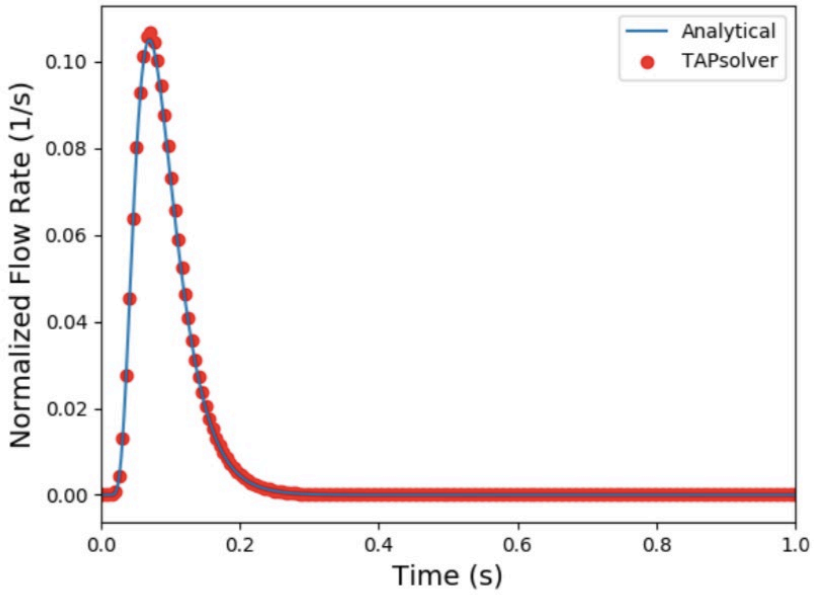
- TAP pulses are ~1s
  - Compare to hours for steady-state
- TAP pulses probe multiple rate constants
  - Only rate-limiting step for steady-state
- All measurements can be performed on a single sample
  - Steady state often requires re-starting with new samples
- TAP experiments work with real catalysts
  - Surface science typically requires single crystals

E. A. Redekop, G. S. Yablonsky, D. Constales, P. A. Ramachandran, C. Pherigo, and J. T. Gleaves, Chemical Engineering Science, 66, 24, 6441–6452, 2011.

Physics-based models facilitate learning of micro-kinetic models

$$\varepsilon \frac{\partial C_i(x,t)}{\partial t} - \frac{\partial}{\partial x} \cdot (D \, \nabla C_i(x,t)) = R(C_i, k_j, ...)$$

$$r_i = k_i^+ \prod_j \theta_{ij} \prod_j C_{ij} - k_i^- \prod_l \theta_{il} \prod_l C_{il}$$



- "TAPSolver" code developed in collaboration with INL
- Enables simulation of TAP reactor with micro-kinetic models
- Use of "automatic differentiation" enables efficient optimization
  - Similar to techniques used by TensorFlow, etc.

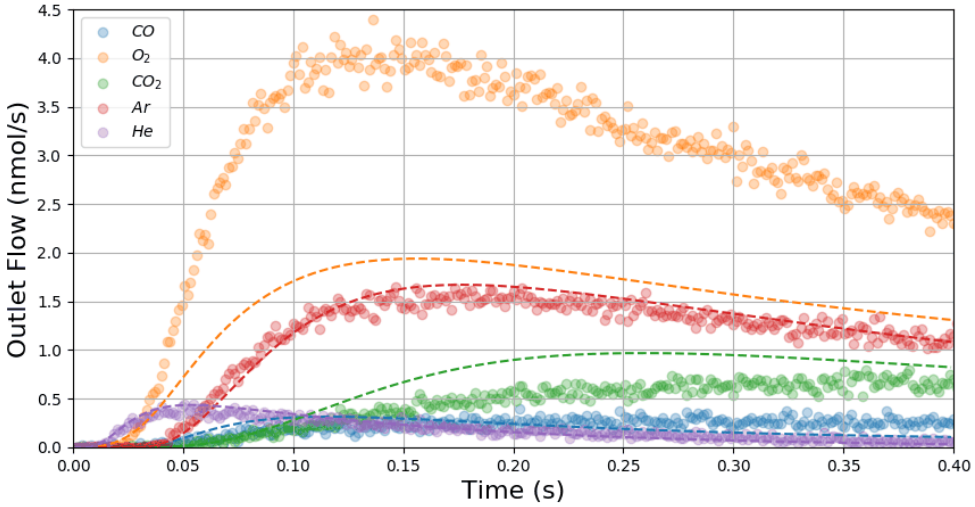
Dynamic catalysis data can be coupled with physics-based “learning” models

TAP data → Differentiable TAP simulator → Catalyst knowledge

Large consistent dataset

Parameters of model are rate constants

Intrinsic kinetic information

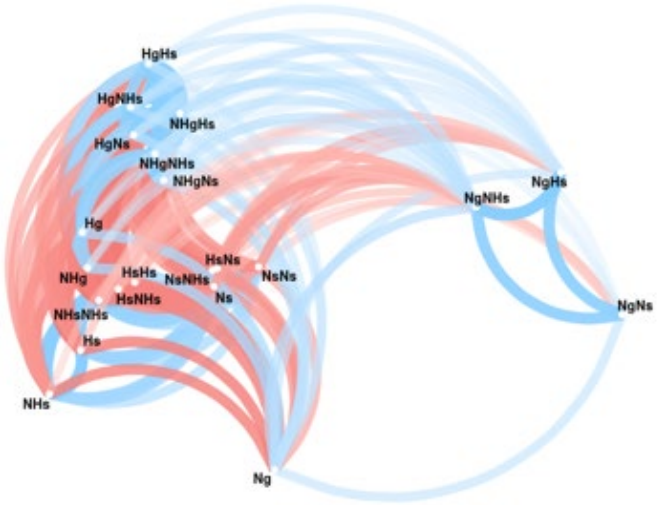


Example: Fitting TAP curve for CO oxidation on supported Pt particles

Statistical approaches can also provide kinetic insight through correlations

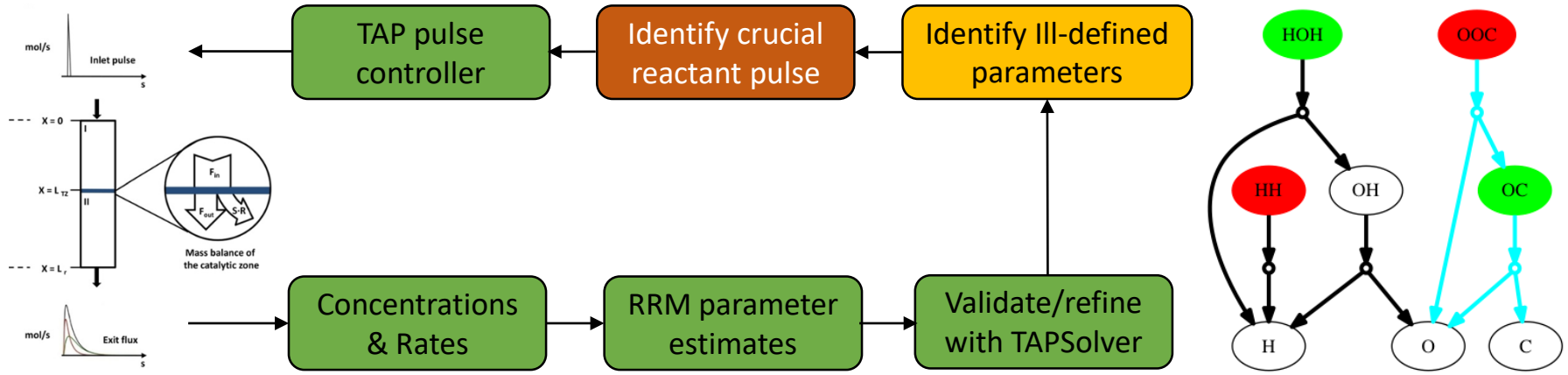
- "Rate Reactivity Model" uses correlations in transient concentrations.
  - Co-linearity can be removed with penalties (LASSO/SCAD)
- Correlation structure provides insight into mechanism
- Results are related to Jacobian of rate with respect to concentration.
  - Can be directly related to rate constants for simple cases
  - Provide initial guesses for rate constants

$$R_i(t) = \mu + \sum_k \beta_{i,k} C_k(t) + \sum_j \beta_{i,j} U_j(t) + \sum_k \sum_j \beta_{i,k,j} C_k(t) U_j(t) + \sum_j \sum_l \beta_{i,k,l} U_j(t) U_l(t) + \epsilon$$



G. S. Yablonsky, E. A. Redekop, D. Constales, J. T. Gleaves, and G. B. Marin, International Journal of Chemical Kinetics, vol. 48, no. 6, pp. 304–317, Mar. 2016.
Kunz et al, Fused Lasso and Tensor Covariance Learning with Robust Estimation. 2018

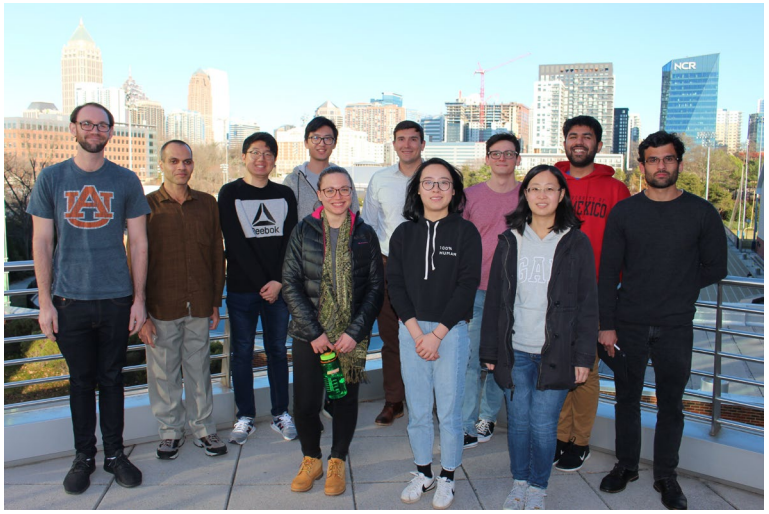
Rapid response time of TAP can enable adaptive approaches

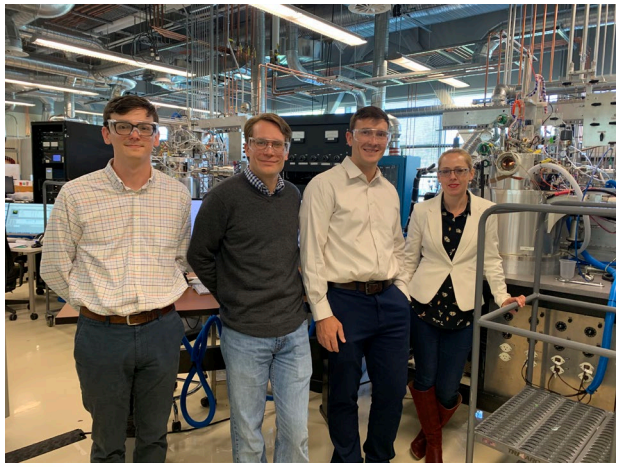


- Automated mechanism analysis
  - Modify pulse composition/frequency to distinguish between ambiguous mechanisms
- Automated condition optimization
  - Modify pulse composition/frequency to search for active/selective catalyst regions
- Need to couple with search/optimization algorithms
  - Integration with mechanism/kinetic model generation

Acknowledgements

- Georgia Tech
  - **Adam Yonge**
  - Gabriel Gusmão
  - Sean Najmi
  - Ben Comer
  - Chaoyi Chang
  - Suvadip Das
  - Ray Lei
  - Sihoon Choi
  - Nicole Hu
  - Jagriti Sahoo
  - Giada Innocenti
- Idaho National Labs
  - **Ross Kunz**
  - **Rebecca Fushimi**
- Funding
  - USDOE EERE Advanced Manufacturing Office Next Generation R&D Projects Contract no. DE-AC07- 05ID14517





U.S. DEPARTMENT OF ENERGY

INL Idaho National Laboratory

Georgia Institute of Technology