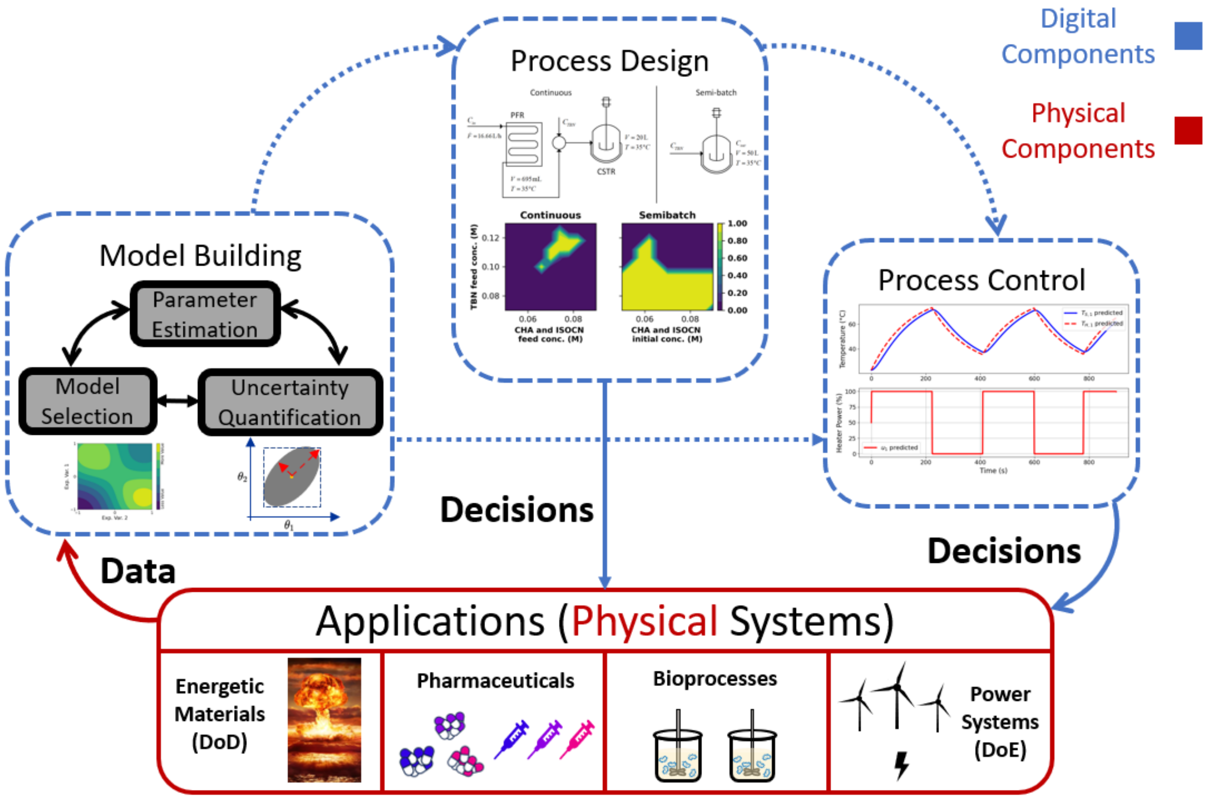
Digitalization in Chemical Engineering: Building Blocks of Fit-for-Purpose Digital Twins

Fit-for-Purpose Digital Twins



Daniel Laky, Ph.D.

Department of Chemical and Biomolecular Engineering

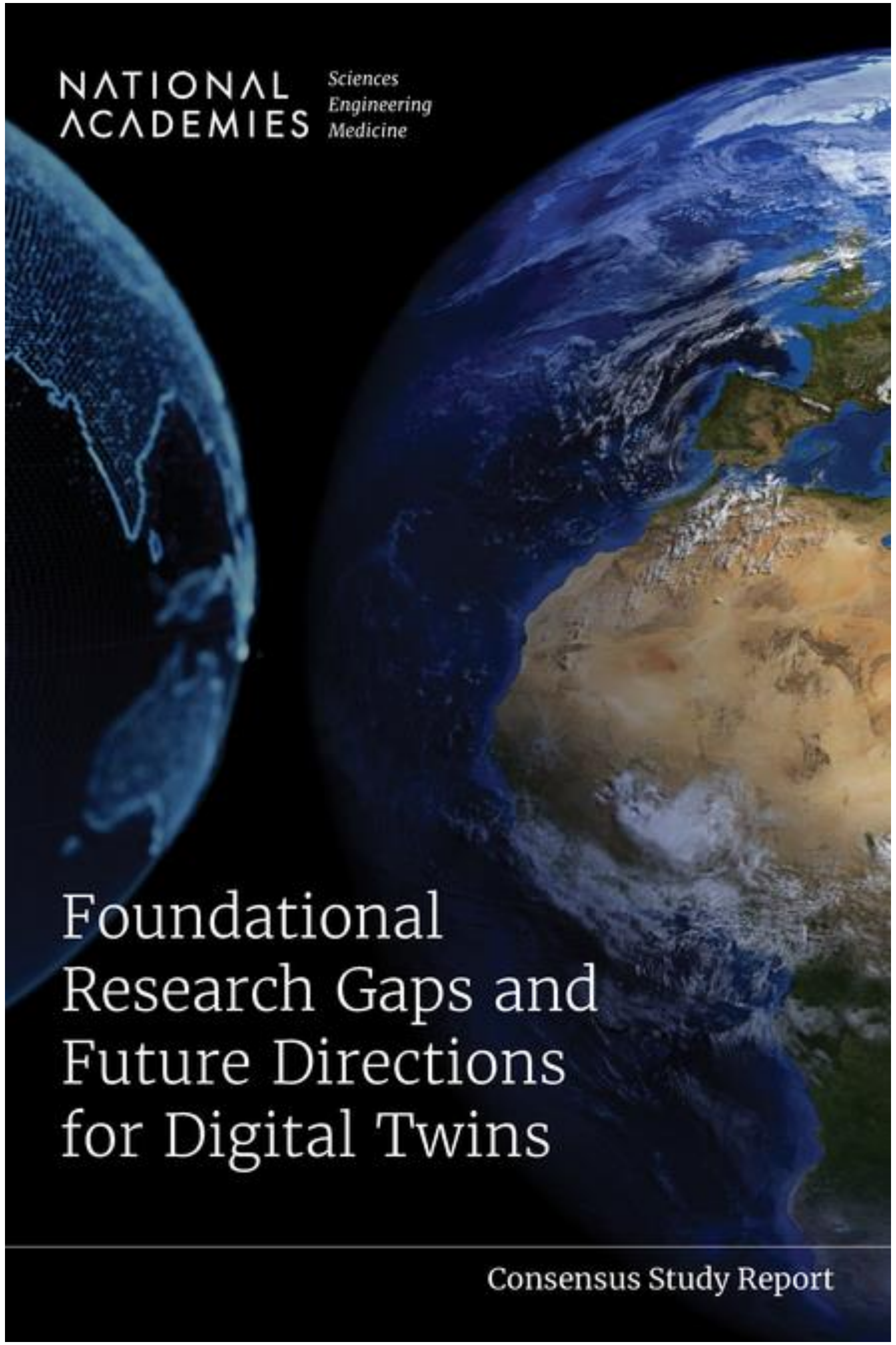
Computational Molecular Science & Engineering Laboratory

University of Notre Dame, Notre Dame, IN

September 6th, 2024

Definition of "Digital Twin"

*A **digital twin** is a set of **virtual information constructs** that mimics the structure, context, and behavior of a **natural**, **engineered**, or **social system** (or **system-of-systems**), is **dynamically updated** with data from its physical twin, has a **predictive capability**, and **informs decisions that realize value**. The **bidirectional interaction** between the virtual and the physical is central to the digital twin.*



Digital Twins Provide Opportunities to Accelerate Scientific Discovery and Improve Manufacturing

- High-throughput applications
  - Self-driving laboratories
  - Automated and adaptive experimental campaigns

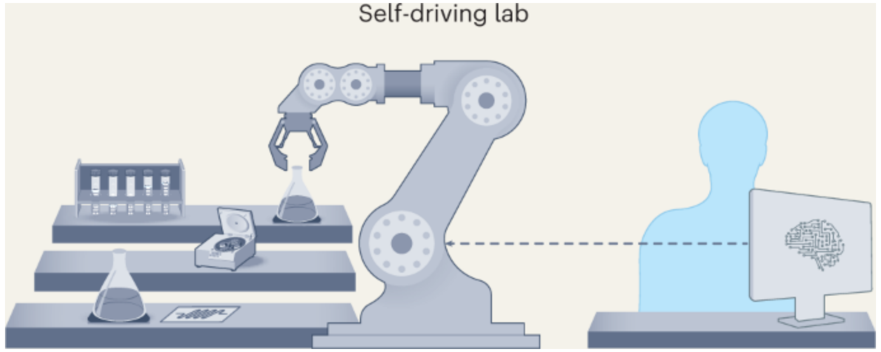


Figure from Abolhasani & Kumacheva (2023), *Nature Syn.*

Digital Twins Provide Opportunities to Accelerate Scientific Discovery and Improve Manufacturing

- High-throughput applications
  - Self-driving laboratories
  - Automated and adaptive experimental campaigns

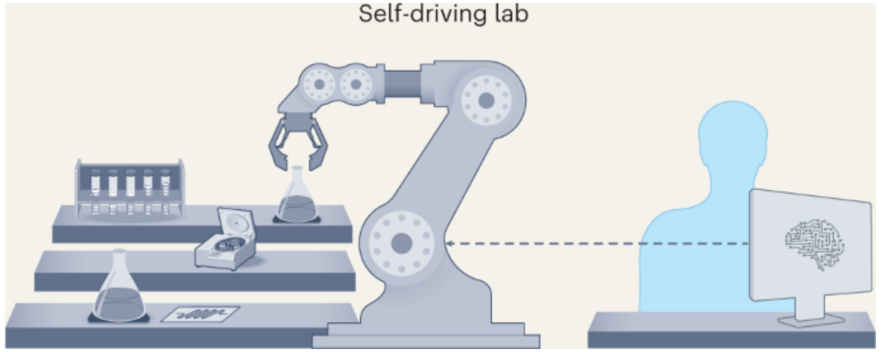
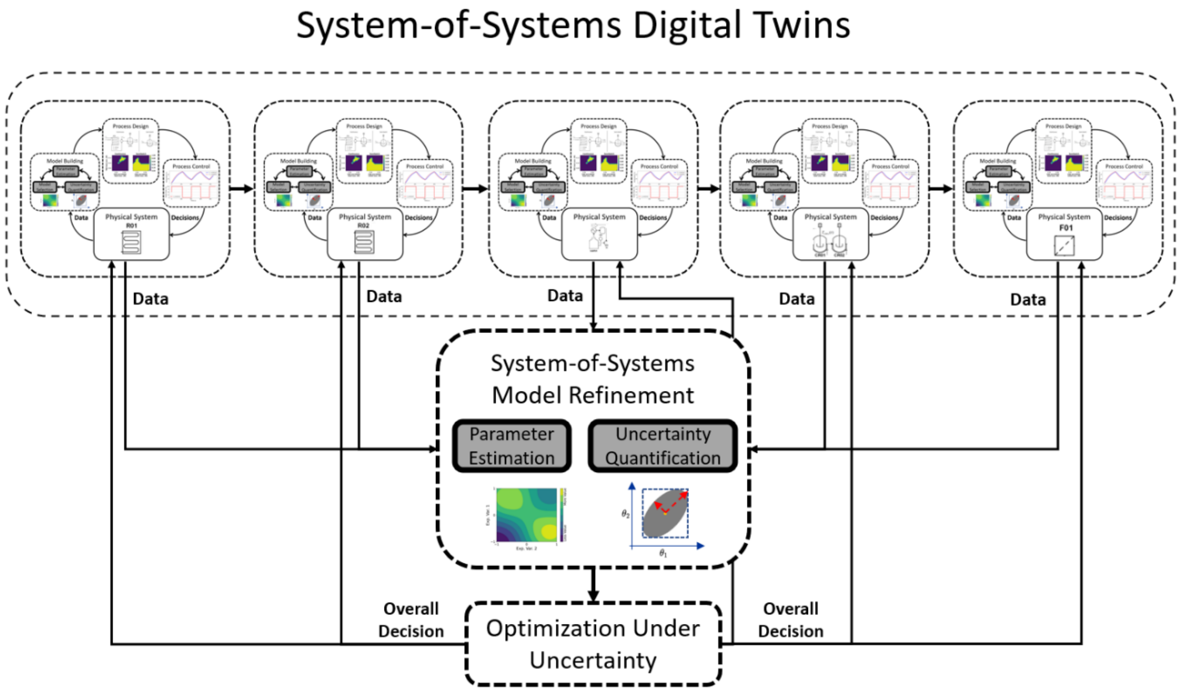


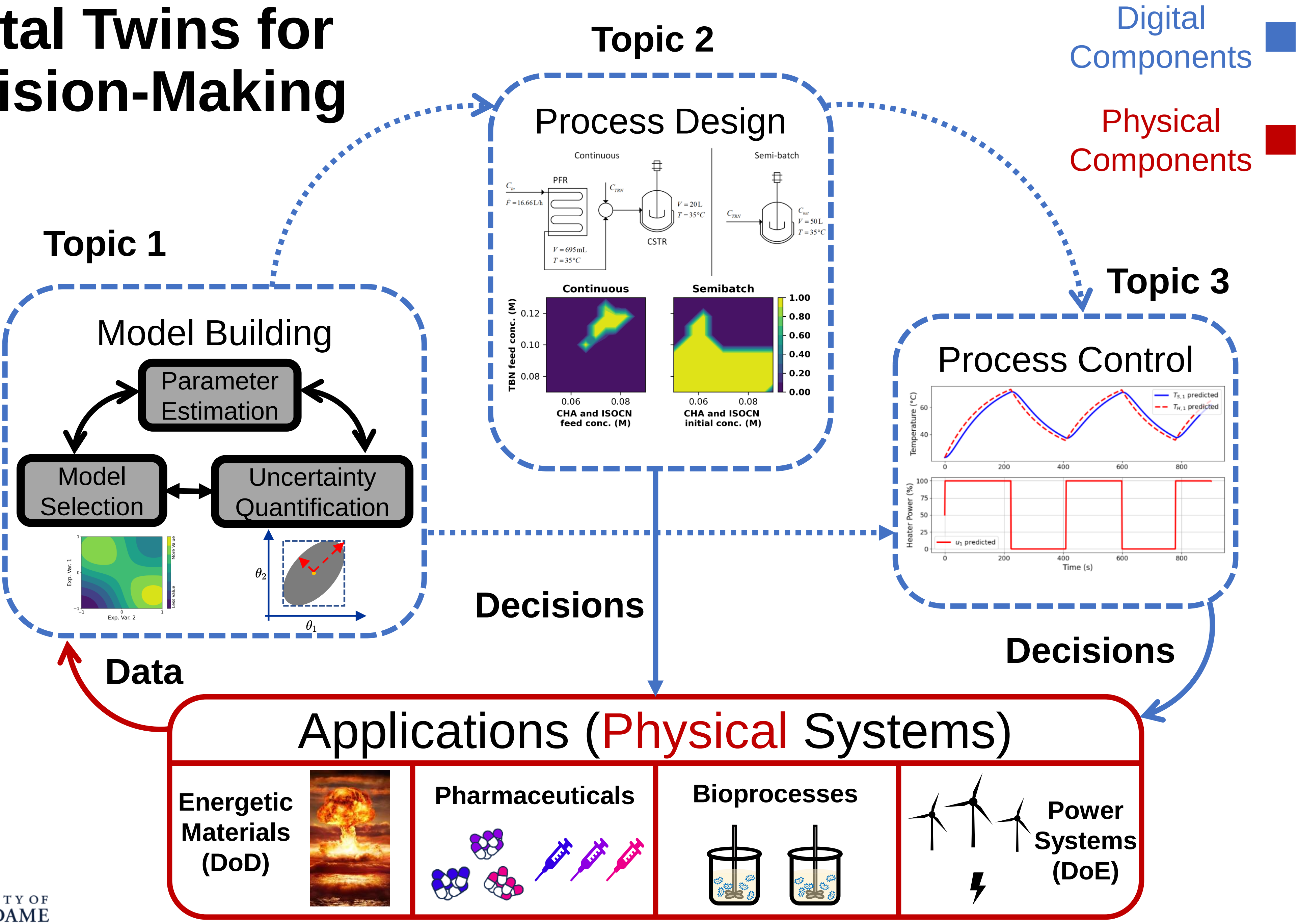
Figure from Abolhasani & Kumacheva (2023), *Nature Syn.*

- Robust digital twins enable optimal manufacturing
  - Optimal design
    - Traditional process design
    - Data flow optimization
  - Optimal control
    - Real-time decision making
  - Applications: pharmaceuticals and energy systems, among others[1,2,3,4,5]



1. Baniqued et al. (2024) *J. of Field Robotics*
2. Wang et al. (2024) *Robotics and Computer-Integrated Manu.*
3. Rajesh et al. (2024) *Journal of Power Sources*
4. Wang et al. (2024) *IEEE Transactions on Intelligent Vehicles*
5. Xames et al. (2024) *IEEE Access*

Digital Twins for Decision-Making

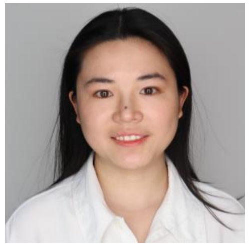


Science-Based Design of Experiments (SBDoE)

University of Notre Dame

Sandia National Lab

Prof. Alexander Dowling



Dr. Jialu Wang

Hailey Lynch

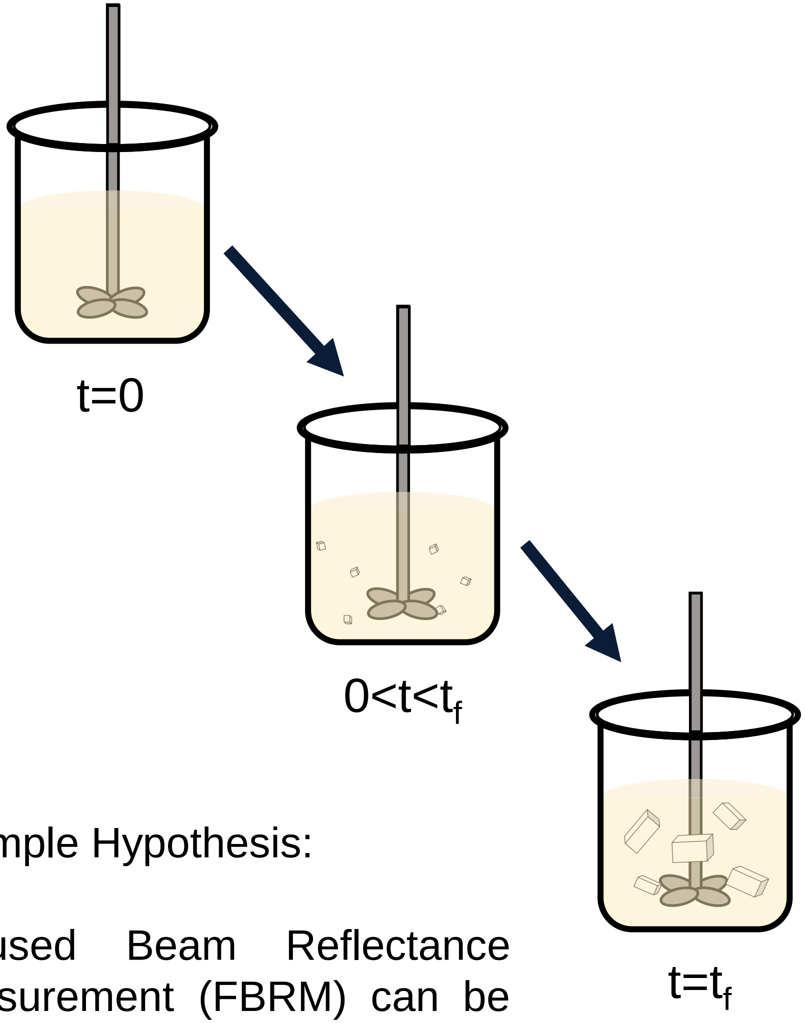
Dr. Bethany Nicholson

Dr. John Siirola

Dr. Shawn Martin

Katherine Klise

Scientific Method



Example Hypothesis:

Focused Beam Reflectance Measurement (FBRM) can be used to control particle size in crystallization

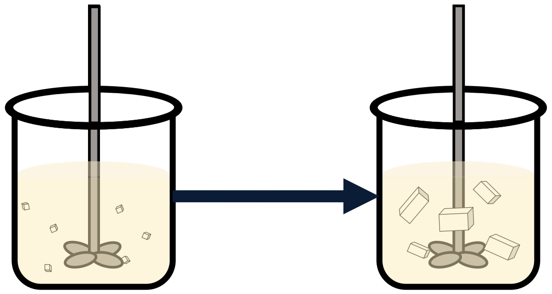
(Direct Nucleation Control [**DNC**])

How easy is it to “Test with Experiments”?

Test with Experiments

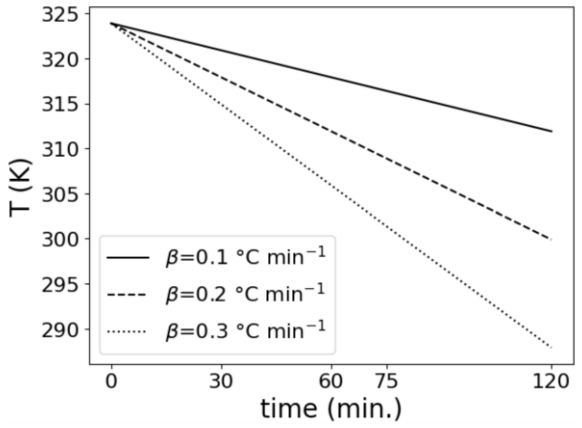
- How many experiments should I perform?
- Which experiment(s) will inform our problem best?
- How expensive are the experiments?
  - Material cost
  - Monetary cost
  - Time cost
- Are we developing a model for use in other engineering tasks?

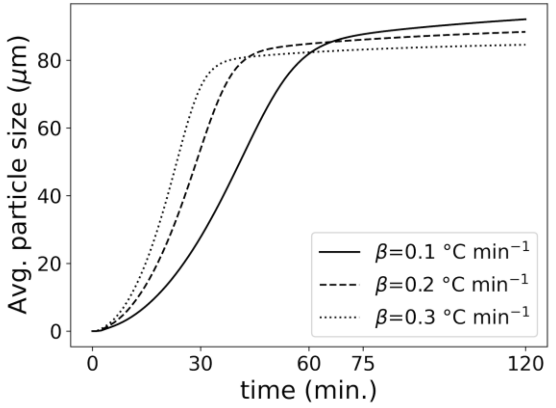
How easy is it to "Test with Experiments"?



Test with Experiments

- How many experiments should I perform?
- Which experiment(s) will inform our problem best?
- How expensive are the experiments?
  - Material cost
  - Monetary cost
  - Time cost
- Are we developing a model for use in other engineering tasks?

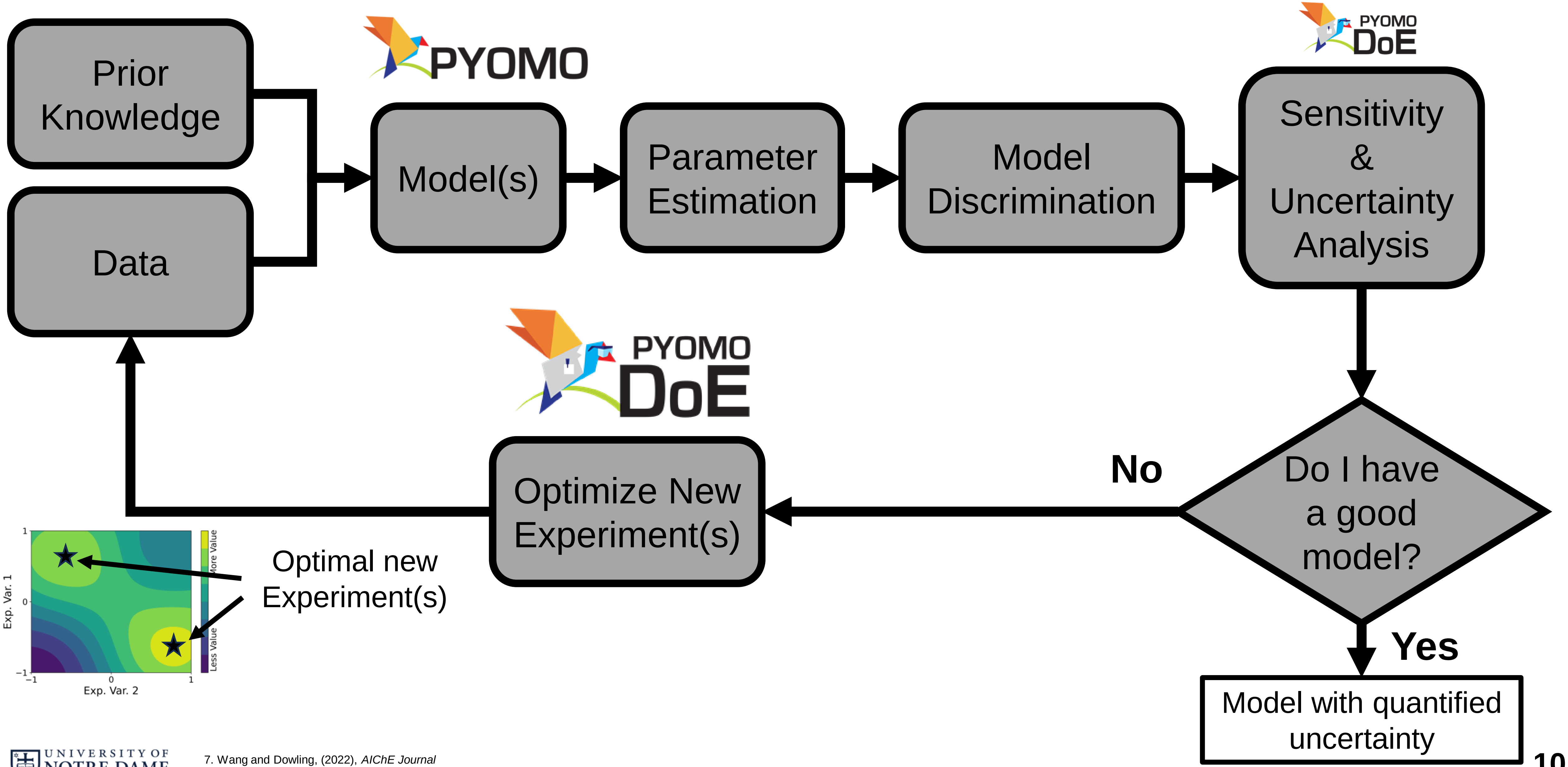




Figures from [6]

6. Lynch, H.G., Bjarnason, A., **Laky, D.J.**, Brown, C.J., and Dowling, A.W. (2024), *Foundations of Computer Aided Process Design (FOCAPD 2024)*

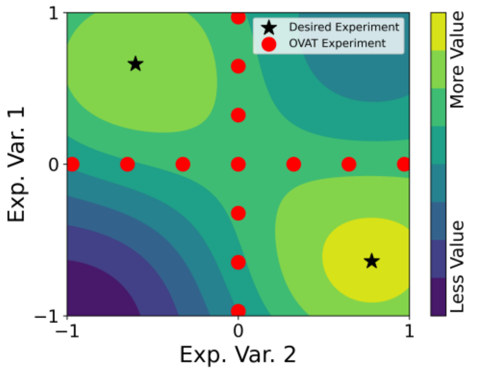
Sequential Optimal Experiment Design (Model-Building)[7]

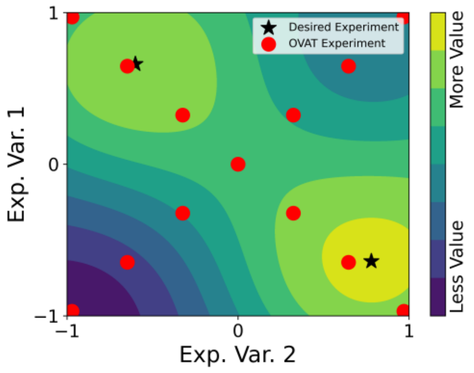


7. Wang and Dowling, (2022), *AIChE Journal*

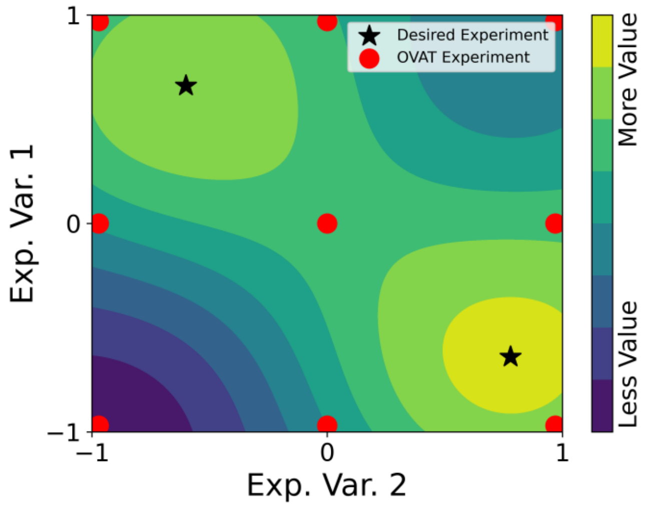
Traditional Experimental Design Strategies Miss Model Information

- One variable at a time[7]
  - Fix variable 1, adjust variable 2
  - Requires some knowledge of the surface to explore critical regions





- Factorial Design[8]
  - Set of values for each design variable
  - Number of experiments grows exponentially with number of design variables

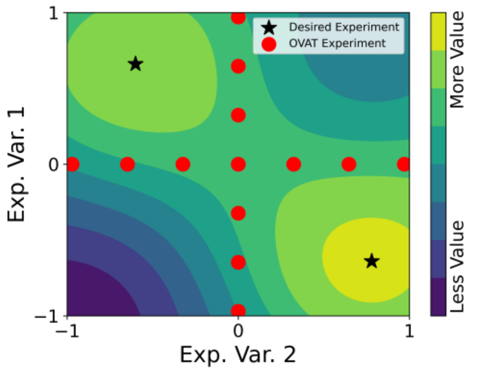


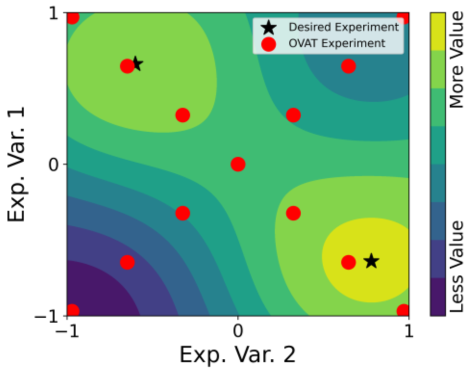
7. Williamson et al. (2022) *Chem. Mater.*

8. Szilyagi et al. (2021) *Crystal Growth & Design*

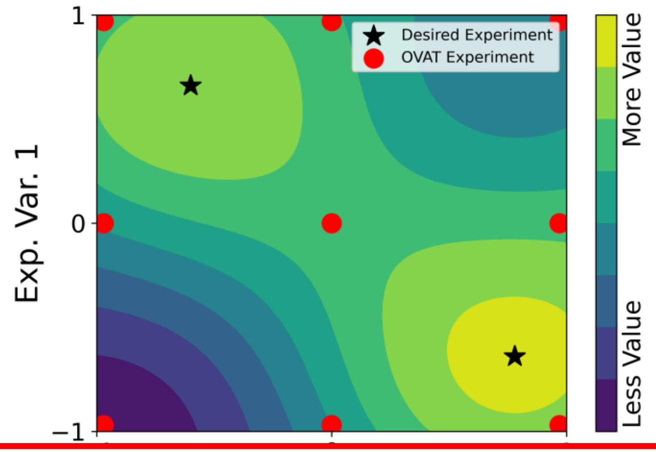
Traditional Experimental Design Strategies Miss Model Information

- One variable at a time[7]
  - Fix variable 1, adjust variable 2
  - Requires some knowledge of the surface to explore critical regions





- Factorial Design[8]
  - Set of values for each design variable
  - Number of experiments grows exponentially with number of design variables

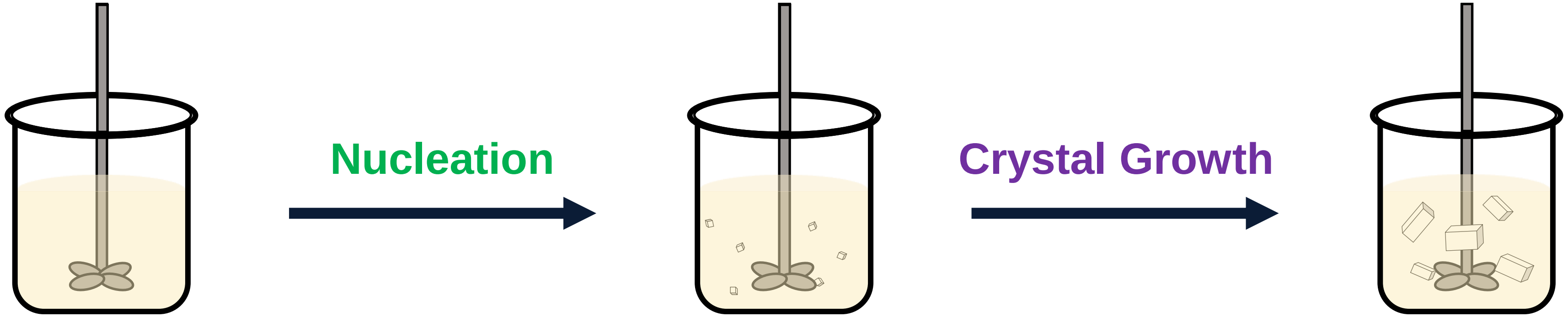


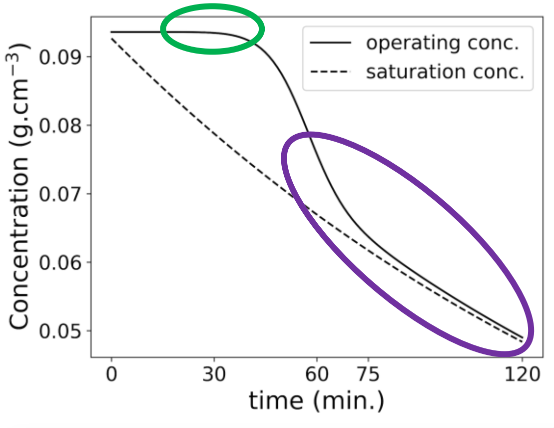
1. Many experiments required to explore design space
2. Often miss critical regions of the design space

7. Williamson et al. (2022) *Chem. Mater.*
8. Szilyagi et al. (2021) *Crystal Growth & Design*

Brief Crystallization Overview (Case Study)

$$\frac{d\mu_0}{dt} = B_p + B_s$$
$$\frac{d\mu_1}{dt} = G\mu_0$$
$$\frac{d\mu_2}{dt} = 2G\mu_1$$
$$\frac{d\mu_3}{dt} = 3G\mu_2$$
$$\frac{dT}{dt} = \alpha(t)$$
$$\frac{dC_{\mathrm{crys}}}{dt} = -\rho_{\mathrm{crys}} k_v \frac{d\mu_3}{dt}$$
$$B_p = k_{b_p} S^p$$
$$B_s = k_{b_s} S^{s_1} (k_v \mu_3)^{s_2}$$
$$G = k_g S^{g_1}$$
$$S = \frac{C - C_{\mathrm{sat}}}{C_{\mathrm{sat}}}$$
$$C_{\mathrm{sat}} = A + BT + CT^2$$
$$\mu_i(0) = \mu_{i_{\mathrm{seed}}} \qquad i \in (0,\ 1,\ 2,\ 3)$$
$$T(0) = T_0$$
$$C_{\mathrm{crys}}(0) = C_0$$





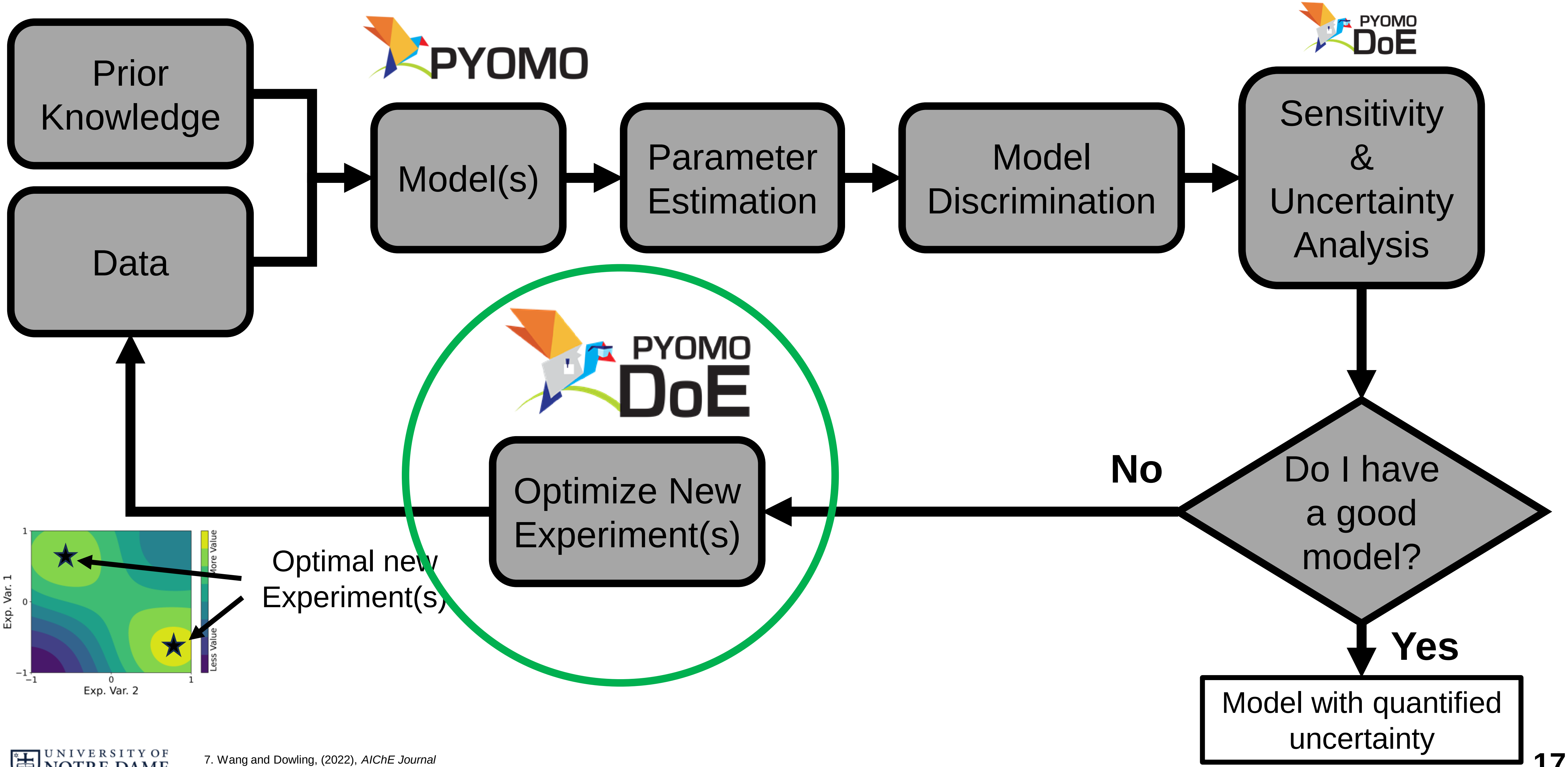
- Cooling crystallization: Solubility reduced by lowering temperature → causes supersaturation

$$S = \frac{C - C_{\mathrm{sat}}}{C_{\mathrm{sat}}}$$
$$C_{\mathrm{sat}} = A + BT + CT^2$$

- At some point, **nucleation** occurs, causing seed crystals to form
- **Growth** becomes the dominant mechanism as supersaturation depletes

6. Lynch, H.G., Bjarnason, A., **Laky, D.J.**, Brown, C.J., and Dowling, A.W. (2024), *Foundations of Computer Aided Process Design (FOCAPD 2024)*

Sequential Optimal Experiment Design (Model-Building)[7]



7. Wang and Dowling, (2022), *AIChE Journal*

Crystallization Model In General Format

$$\frac{d\mu_0}{dt} = B_p + B_s$$
$$\frac{d\mu_1}{dt} = G\mu_0$$
$$\frac{d\mu_2}{dt} = 2G\mu_1$$
$$\frac{d\mu_3}{dt} = 3G\mu_2$$
$$\frac{dT}{dt} = \alpha(t)$$
$$\frac{d\boxed{C_{\mathrm{crys}}}}{dt} = -\rho_{\mathrm{crys}} k_v \frac{d\mu_3}{dt}$$
$$B_p = \boxed{k_{b_p}} S^{\boxed{p}}$$
$$B_s = \boxed{k_{b_s}} S^{\boxed{s_1}} (k_v \mu_3)^{\boxed{s_2}}$$
$$G = \boxed{k_g} S^{\boxed{g_1}}$$
$$S = \frac{C - C_{\mathrm{sat}}}{C_{\mathrm{sat}}}$$
$$C_{\mathrm{sat}} = A + BT + CT^2$$

$$\mu_i(0) = \mu_{i_{\mathrm{seed}}} \qquad i \in (0,\ 1,\ 2,\ 3)$$
$$T(0) = T_0$$
$$C_{\mathrm{crys}}(0) = C_0$$

Measurement Estimates

Unknown Parameters

Experimental Design

- $\boldsymbol{f}$: differential equations
- $\boldsymbol{g}$: algebraic equations
- $\boldsymbol{h}$: measurement functions
- $\hat{\boldsymbol{y}}$: measurement response variable prediction
- $\boldsymbol{y}$: measurement response variable
- $\dot{\boldsymbol{x}}$: derivate of $\boldsymbol{x}$ *wrt* time
- $\boldsymbol{x}$: dynamic state variable
- $\boldsymbol{z}$: algebraic state variables
- $\boldsymbol{u}$: time-varying control variables
- $\bar{\boldsymbol{w}}$: constant control variables
- $\boldsymbol{t}$: control timepoints
- $\boldsymbol{\theta}$: unknown model parameters
- $\boldsymbol{\phi}$: experimental design vector

Sensitivity Matrix Used to Explain Information Content of Experimental Data

$$\mathbf{Q}_r = \begin{bmatrix} \left.\frac{\partial \hat{y}_r}{\partial \theta_1}\right|_{t_1} & \cdots & \left.\frac{\partial \hat{y}_r}{\partial \theta_p}\right|_{t_1} \\ \vdots & \ddots & \vdots \\ \left.\frac{\partial \hat{y}_r}{\partial \theta_1}\right|_{t_{n_r}} & \cdots & \left.\frac{\partial \hat{y}_r}{\partial \theta_p}\right|_{t_{n_r}} \end{bmatrix}$$

Sensitivity matrix

$$\mathbf{V}\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\phi}\right) \approx \left[\sum_{r}^{N_r} \sum_{r'}^{N_r} \tilde{\sigma}_{(r,r')} \mathbf{Q}_r^{\mathrm{T}} \mathbf{Q}_{r'} + \overbrace{\mathbf{V}_0\left(\hat{\boldsymbol{\theta}}\right)^{-1}}^{\text{Prior}}\right]^{-1}$$

Parameter covariance matrix

$$\mathbf{FIM} \approx \left[\mathbf{V}\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\phi}\right)\right]^{-1}$$

Fisher information matrix (FIM)

- Fisher Information Matrix[9]: Used to understand how an experiment or set of experiments contribute to “information” explained

9. Fisher (1949) *The Design of Experiments*
10. Atkinson (1982) *International Statistical Review*
11. Kiefer (1959) *J. of the Royal Statistical Society. Series B*
12. Box and Draper (1959) *J. of the American Statistical Association*

Alphabetic Design Criteria Measure Information Content

Figure adapted from: Franceschini, G., & Macchietto, S. (2008). *Chem. Eng. Sci.*, *63*(19), 4846-4872.

A-optimality

max trace(**FIM**)

enclosing box volume

poor choice for highly correlated **θ**

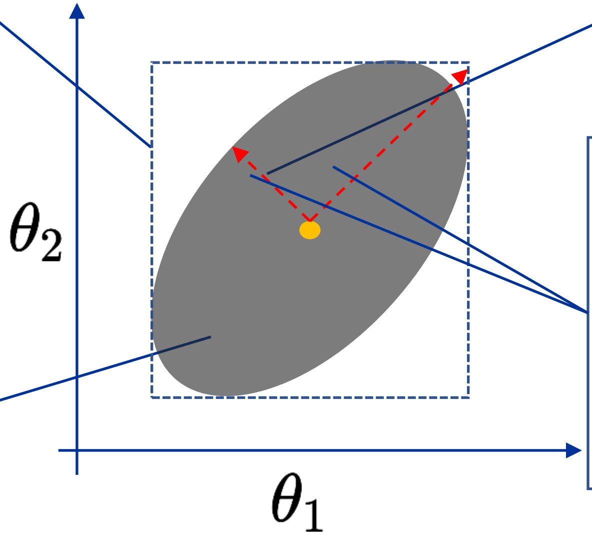
D-optimality

max det(**FIM**)

ellipsoid volume

robust to linear transformations

confidence ellipsoid for covariance matrix **V** = **FIM**$^{-1}$



E-optimality

max min(eig(**FIM**))

major axis

recommended if M is ill-conditioned

ME-optimality

min κ(**FIM**) = max(eig(**FIM**)) / min(eig(**FIM**))

ratio of major to minor axes

recommended if M is ill-conditioned

Model Discrimination

Hunter, W.G. and Reiner, A.M., 1965. Designs for discriminating between two rival models. *Technometrics*, *7*(3), pp.307-323.

Buzzi-Ferraris, G. and Forzatti, P., 1983. A new sequential experimental design procedure for discriminating among rival models. *Chemical engineering science*, *38*(2), pp.225-232.

Ferraris, G.B., Forzatti, P., Emig, G. and Hofmann, H., 1984. Sequential experimental design for model discrimination in the case of multiple responses. *Chemical engineering science*, *39*(1), pp.81-85.

Joint Parameter Precision and Model Discrimination

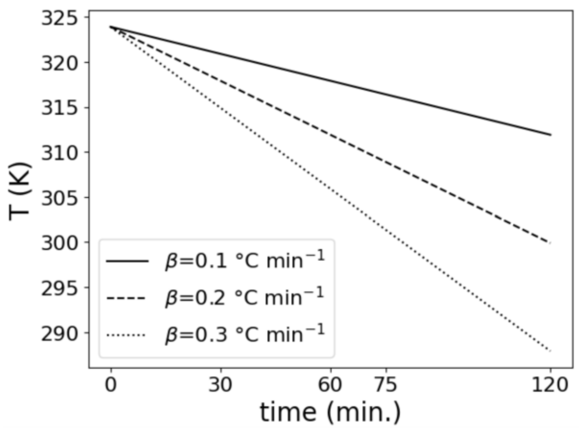
Alberton, A.L., Schwaab, M., Lobão, M.W.N. and Pinto, J.C., 2011. Experimental design for the joint model discrimination and precise parameter estimation through information measures. *Chemical Engineering Science*, *66*(9), pp.1940-1952.

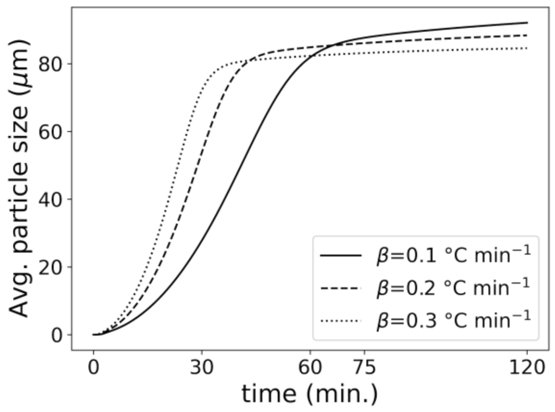
Galvanin, F., Cao, E., Al-Rifai, N., Gavriilidis, A. and Dua, V., 2016. A joint model-based experimental design approach for the identification of kinetic models in continuous flow laboratory reactors. *Computers & Chemical Engineering*, *95*, pp.202-215.

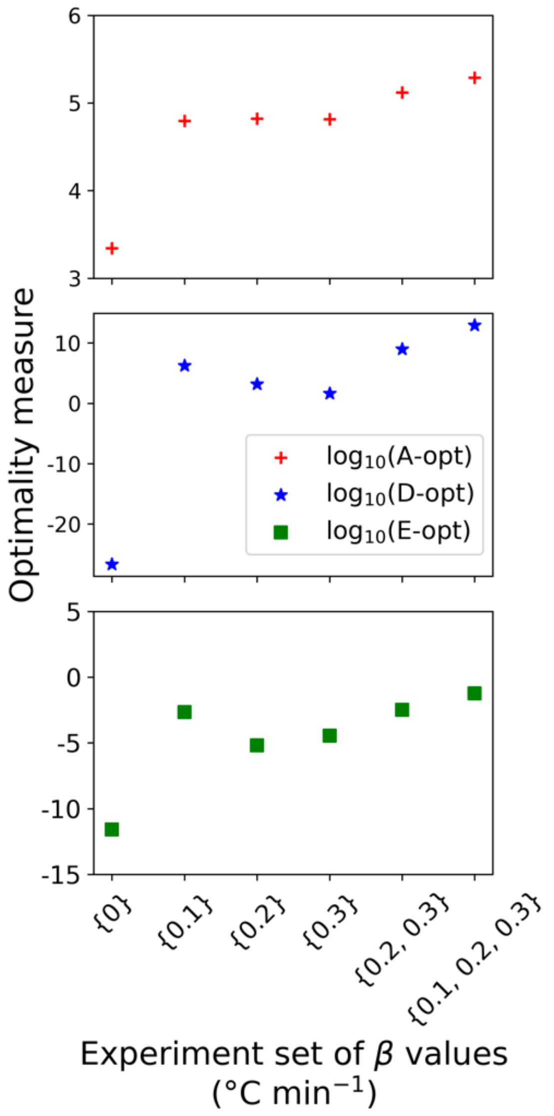
Galvanin, F., Cao, E., Al-Rifai, N., Dua, V. and Gavriilidis, A., 2015. Optimal design of experiments for the identification of kinetic models of methanol oxidation over silver catalyst. *Chimica Oggi-Chemistry Today*, *33*(3), pp.51-56.

Pankajakshan, A., Waldron, C., Quaglio, M., Gavriilidis, A. and Galvanin, F., 2019. A Multi-Objective Optimal Experimental Design Framework for Enhancing the Efficiency of Online Model Identification Platforms. *Engineering*, *5*(6), pp.1049-1059.

Most Experimental Information at Lowest Cooling Rate







- Explore information content of different cooling rates (0.1, 0.2, and 0.3 ºC·min^{-1})
  - Traditional “factorial” DoE target slower cooling rates for growth kinetics

- Low cooling rate is most information-rich experiment (aligns with expert intuition)

- Multiple experiments lead to higher E-optimality, A-optimality, and especially D-optimality
  - D-Optimality → Smaller Confidence Region

Hailey Lynch

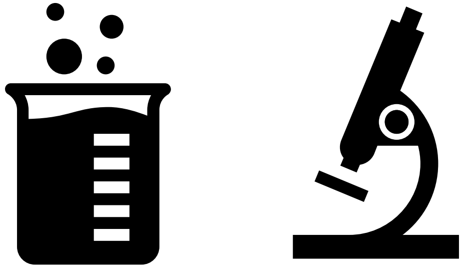
UNIVERSITY OF NOTRE DAME

6. Lynch, H.G., Bjarnason, A., **Laky, D.J.**, Brown, C.J., and Dowling, A.W. (2024), *Foundations of Computer Aided Process Design (FOCAPD 2024)*

"Experiment" Abstraction Streamlines Closed-Loop Experiment Design

Experiment Inputs (Conditions)

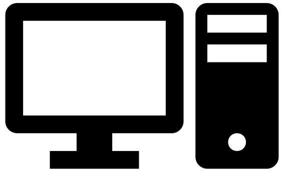
Experiment (Physical)



Experiment Outputs (Data) + Measurement Uncertainty

ϕ : experimental design vector

Experiment (Digital)



$$y_i = f(x, \phi, \theta) + \varepsilon_i$$

$\hat{y}$: measurement response variable prediction

Dr. Bethany Nicholson

Dr. John Siirola

Dr. Shawn Martin

Katherine Klise

“Experiment” Abstraction Streamlines Closed-Loop Experiment Design

PYOMO - parmest

Known Variables

$\boldsymbol{\phi}$: experimental design vector

$\boldsymbol{y}$: measurement response variable

Experiment
(Digital)

$y_i = f(x, \phi, \theta) + \varepsilon_i$

Unknown Variables

$\boldsymbol{\theta}$: unknown model parameters

$$\min_{\boldsymbol{\theta}} \sum_i (y_i - \hat{y}_i)^2$$

PYOMO DoE

$\boldsymbol{\theta}$: unknown model parameters

Experiment
(Digital)

$y_i = f(x, \phi, \theta) + \varepsilon_i$

$\boldsymbol{\phi}$: experimental design vector

$$\max_{\phi_l \leq \phi \leq \phi_u} \Psi\left(\mathbf{FIM}\left(\boldsymbol{\theta}, \boldsymbol{\phi}\right)\right)$$


Dr. Bethany Nicholson


Dr. John Siirola


Dr. Shawn Martin


Katherine Klise

Temperature Control Lab (TC-Lab) – Closed-Loop Experimental Design with New Experiment Abstraction

Parameter Estimation

$$C_p^H \frac{dT_{H,1}}{dt} = U_a(T_{amb} - T_{H,1}) + U_b(T_{S,1} - T_{H,1}) + \alpha P_1 u_1$$

$$C_p^S \frac{dT_{S,1}}{dt} = U_b(T_{H,1} - T_{S,1})$$

$$\theta \equiv \{C_P^H, C_P^S, U_a, U_b\}$$

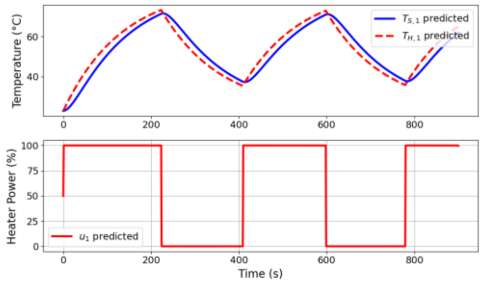
$$y \equiv \{T_{S,1}\}$$

$$\phi \equiv \{u_1\}$$

PYOMO - parmest

PYOMO DoE

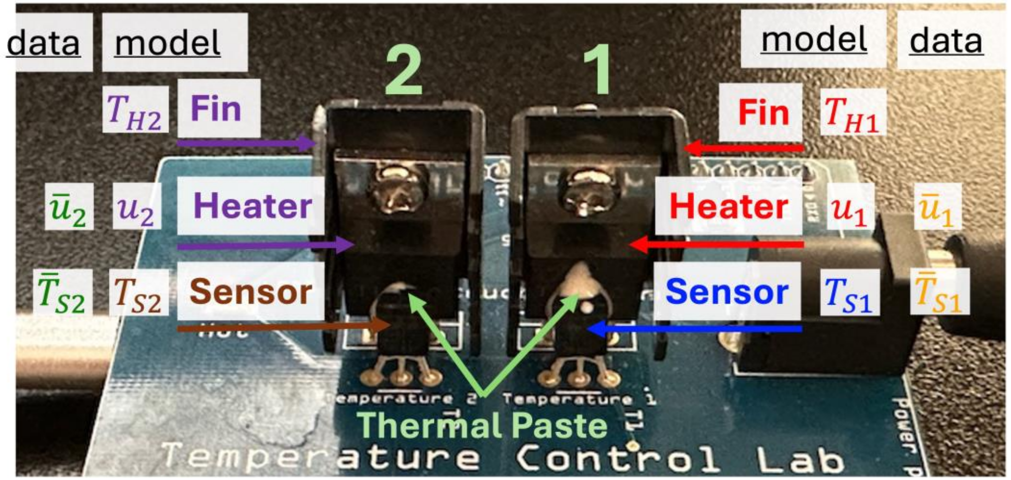
Optimal Experiment Design/Optimal Control



Digital Components

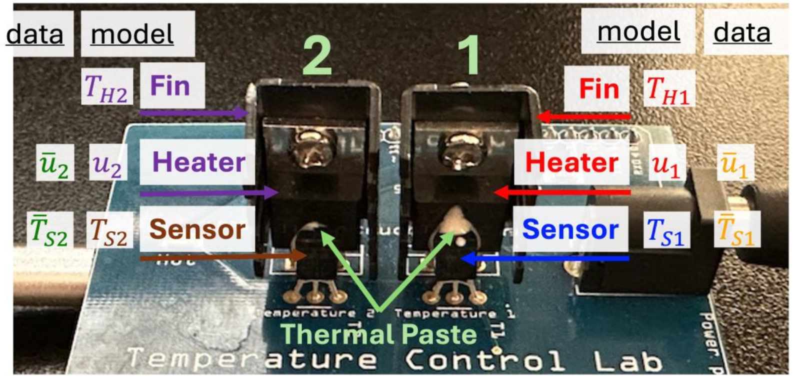
Physical Components

Physical System

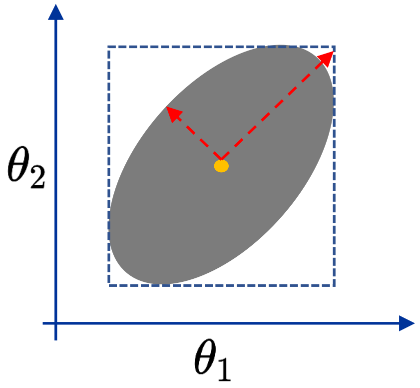


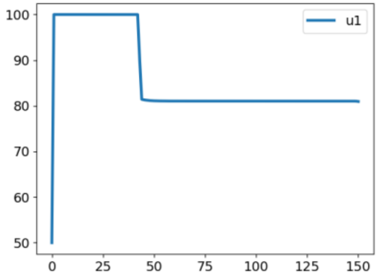
UNIVERSITY OF NOTRE DAME

Optimal Experiments Reduce Uncertainty Depending on FIM Criteria



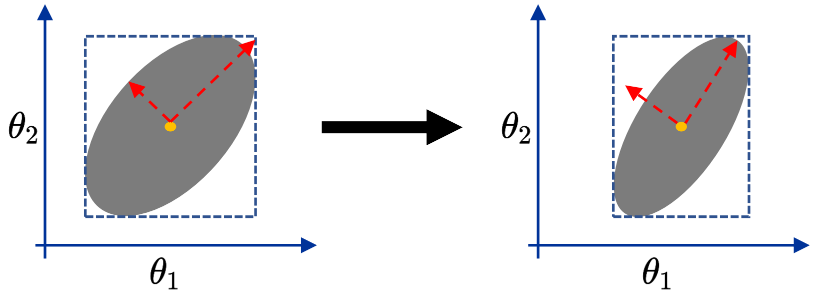
Initial Uncertainty





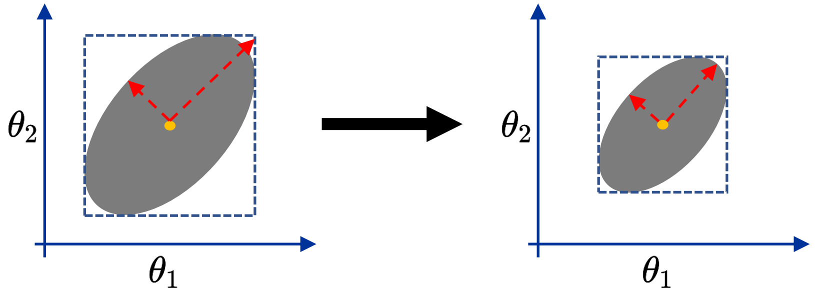
A-optimal

Reduced bounding box size

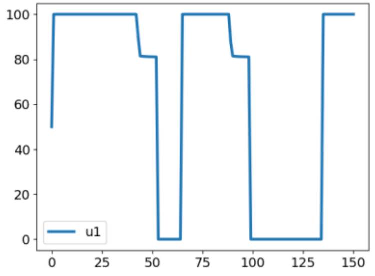


PYOMO DoE

Reduced ellipsoid hypervolume



D-optimal

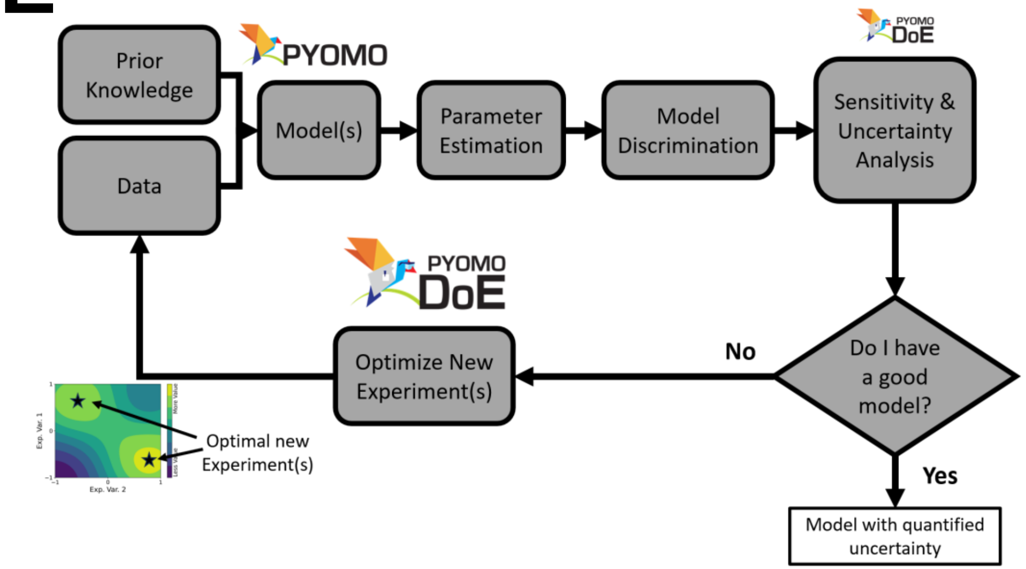


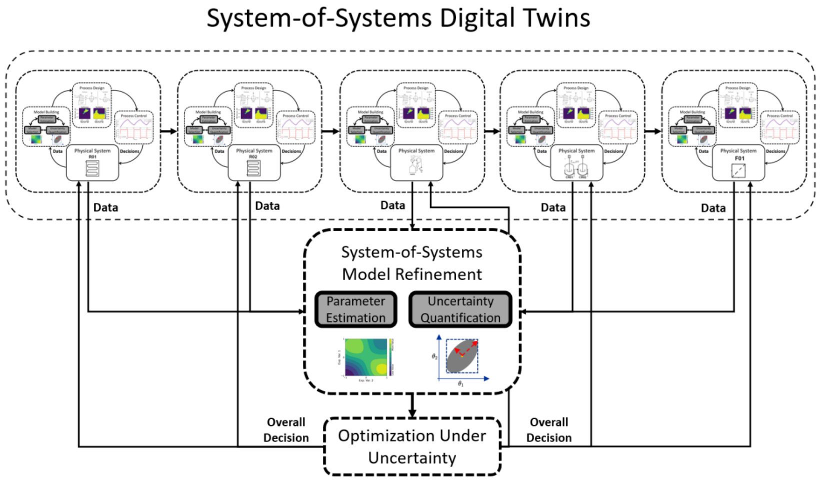
Summary and Future Work in SBDoE

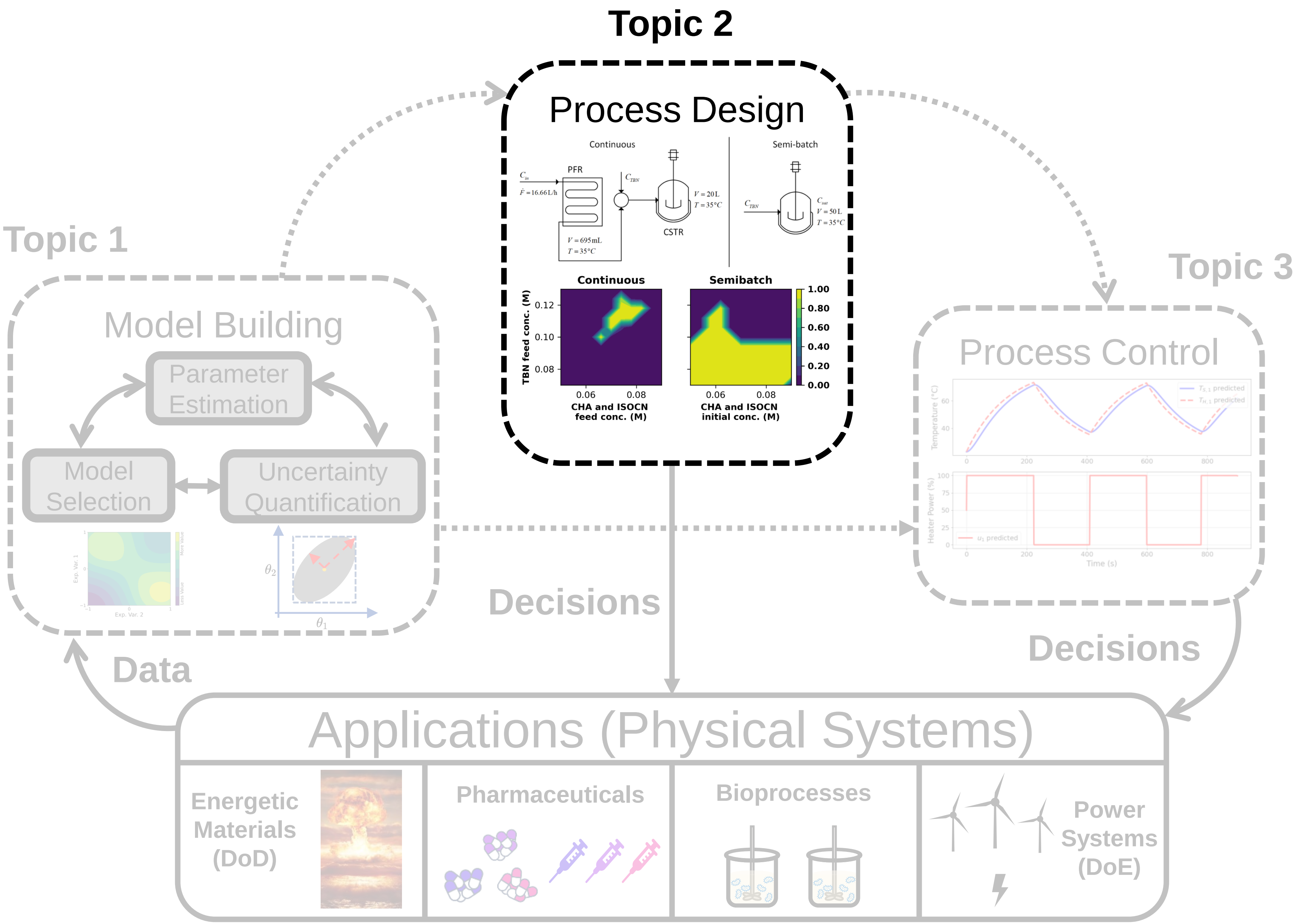
- Pyomo.DoE is a capable and easy-to-use tool for implementing science-based design of experiments (SBDoE) to build digital twins

- Abundant opportunities for future collaboration

- In the future we want to move past individual subsystems toward "system-of-systems" digital twins
  - Uncertainty propagation
  - Subsystem interactions





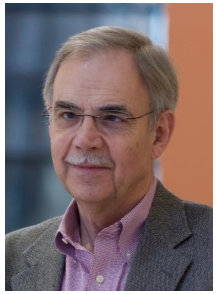

Topic 2
Process Design
Continuous
Semi-batch
PFR
CSTR
Continuous
Semibatch
TBN feed conc. (M)
CHA and ISOCN feed conc. (M)
CHA and ISOCN initial conc. (M)
Topic 1
Model Building
Parameter Estimation
Model Selection
Uncertainty Quantification
Topic 3
Process Control
Decisions
Decisions
Data
Applications (Physical Systems)
Energetic Materials (DoD)
Pharmaceuticals
Bioprocesses
Power Systems (DoE)

Optimal Process Design
(Pharmaceuticals and Energy Systems)

Purdue University

University of Notre Dame

Dr. Daniel Casas-Orozco



Prof. Gintaras Reklaitis

Prof. Zoltan Nagy

Prof. Alexander Dowling

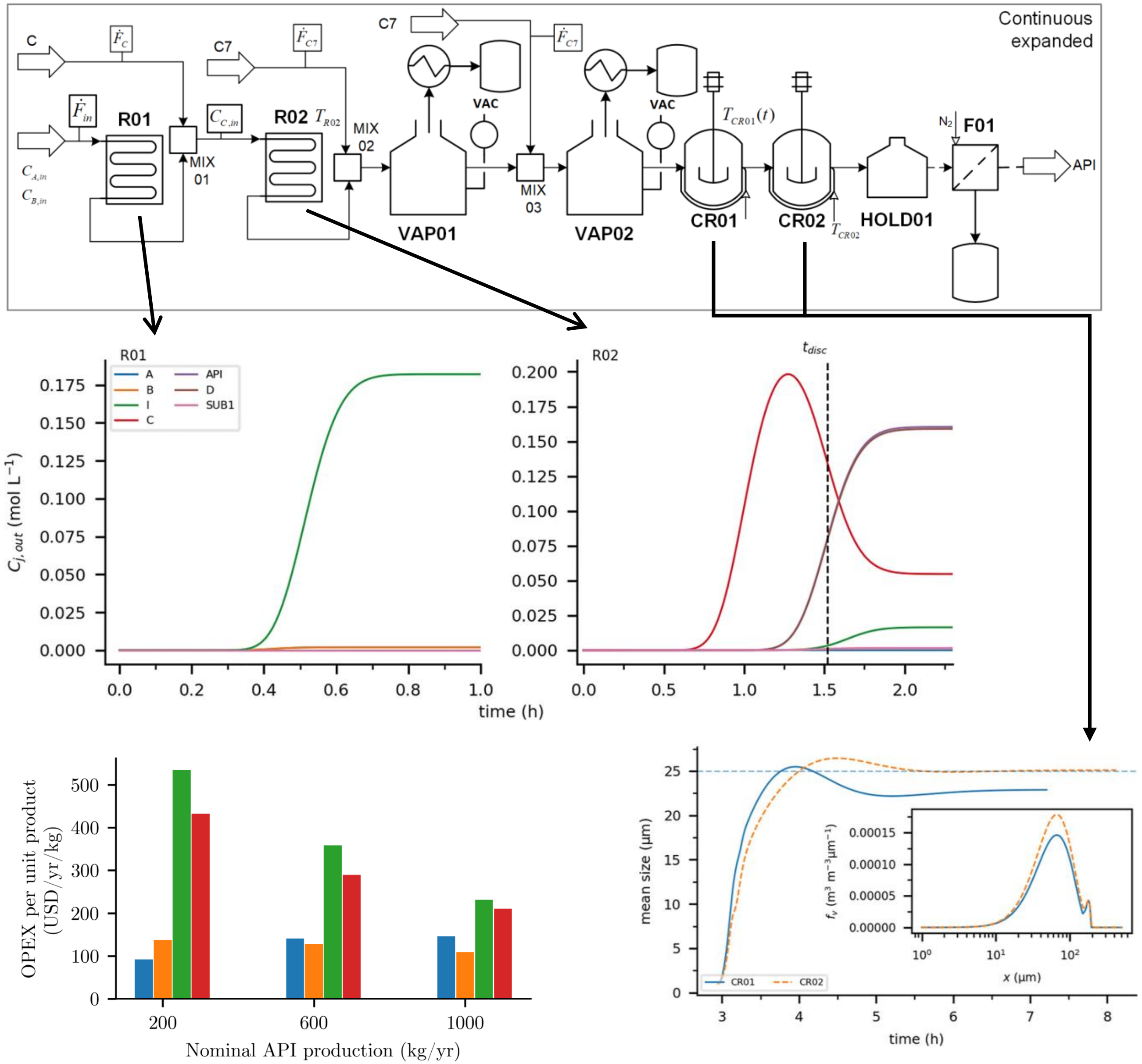
Nicole Cortes

PharmaPy: An Open-Source Process Simulator in Python

- Python-based modeling
  - Code-based UI

- Optimal process design and redesign
  - Modeling and simulation of manufacturing processes

- Evaluate economics for manufacturing processes



Motivating Example: Brain Cancer Treatment – Lomustine

- Lomustine is an anti-cancer drug, began being produced by Bristol-Myers Squibb in 1976
- Was re-released as Gleostine in 2013 by NextSource, after which the price dramatically increased over the next years [27]
- Our colleagues at Purdue have developed a new pathway for a two-step flow synthesis of Lomustine [28] in 2019

$$\mathrm{CHA} + \mathrm{ISOCN} \xrightarrow{k_1} \mathrm{INT}$$

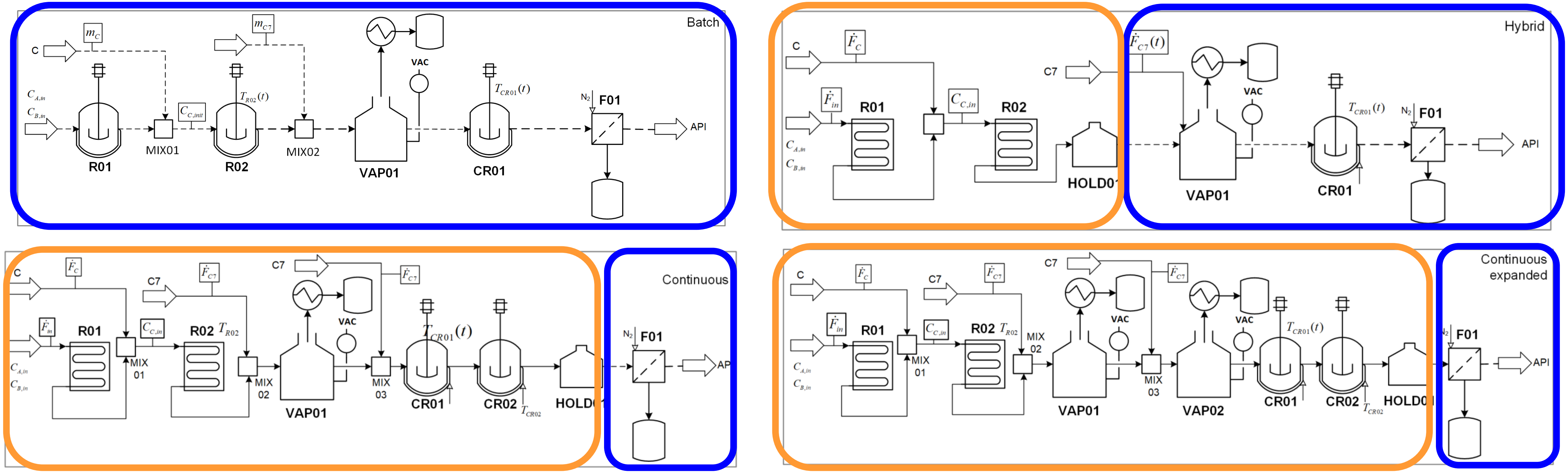
$$\mathrm{INT} + \mathrm{TBN} \xrightarrow{k_2} \mathrm{Lom} + \mathrm{TBOH}$$

- Reaction 1 occurs very quickly when compared to Reaction 2
- Goals:
  - Determine an efficient manufacturing route for Lomustine
  - Refine kinetic models for Lomustine synthesis
  - Identify flexible operating conditions for Lomustine synthesis under uncertainty

[27] Loftus, P., (2017) *The Wall Street Journal*,
[28] Jaman, Z. et. al. (2019) *Org. Process Res. Dev.*

PharmaPy Digital Twins Can Help Compare Manufacturing Candidates[2]

- Analyze production of Lomustine for different routes and scales
  - Scales: 200kg, 600kg, 1000kg
  - Modes: Batch, Hybrid, Continuous

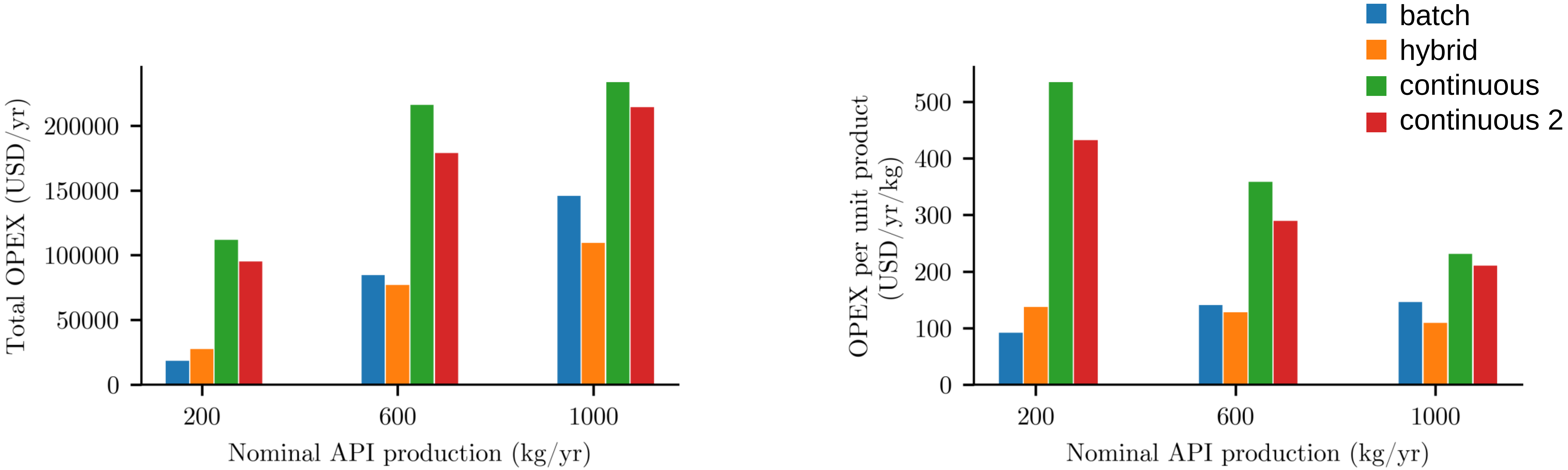


Continuous unit operations

Batchwise unit operations

2. Casas-Orozco et al. (2023) *AIChE Journal*

PharmaPy Digital Twins Can Help Compare Manufacturing Candidates[2]

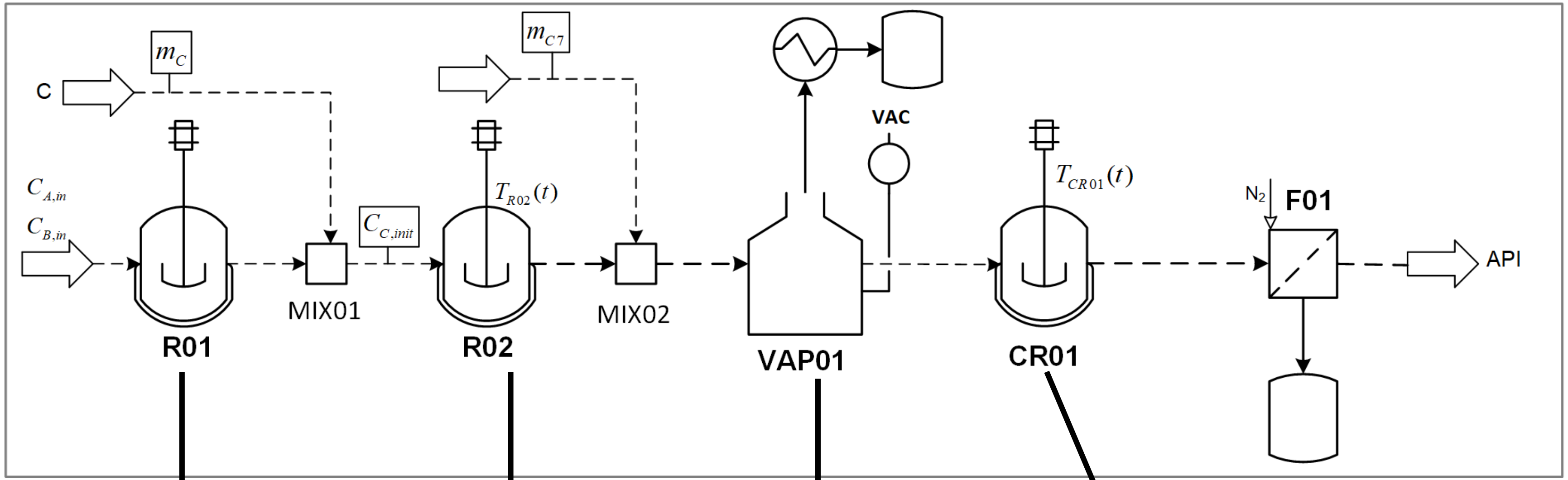


- Derivative free optimization (Nelder-Mead in scipy with penalties for constraint violation)
- Batch or Hybrid operating modes have the lowest cost for all production scenarios
- As scale increases, continuous is becoming more desirable

2. Casas-Orozco et al. (2023) *AIChE Journal*

PharmaPy Digital Twins Can Help Identify Optimal Manufacturing Routes[15]

- Generate API with the optimal manufacturing route
- Process includes: (1) Synthesis, (2) Solvent Switch (vaporization), (3) Crystallization, (4) Filtration



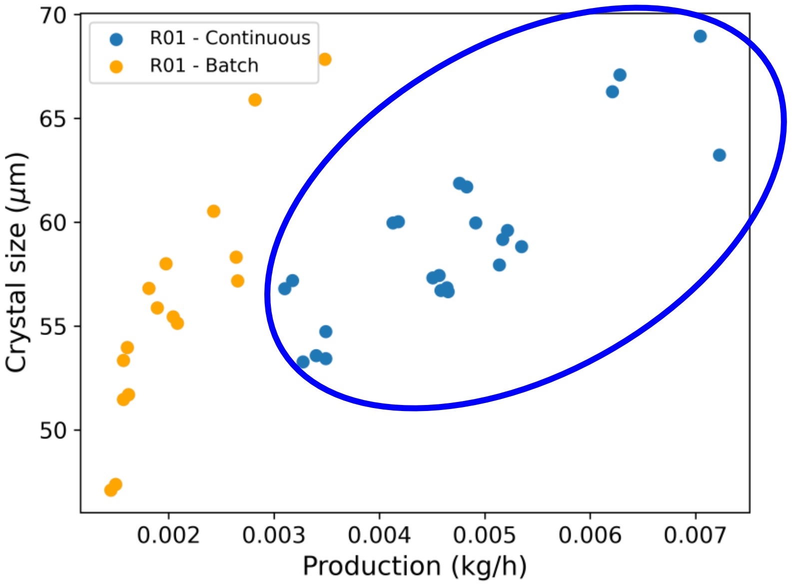
R01	R02	VAP01	CR01
Batch CSTR PFR Semibatch*	Batch CSTR PFR Semibatch*	Batch	Batch Cooling MSMPR 2 MSMPRs 3 MSMPRs

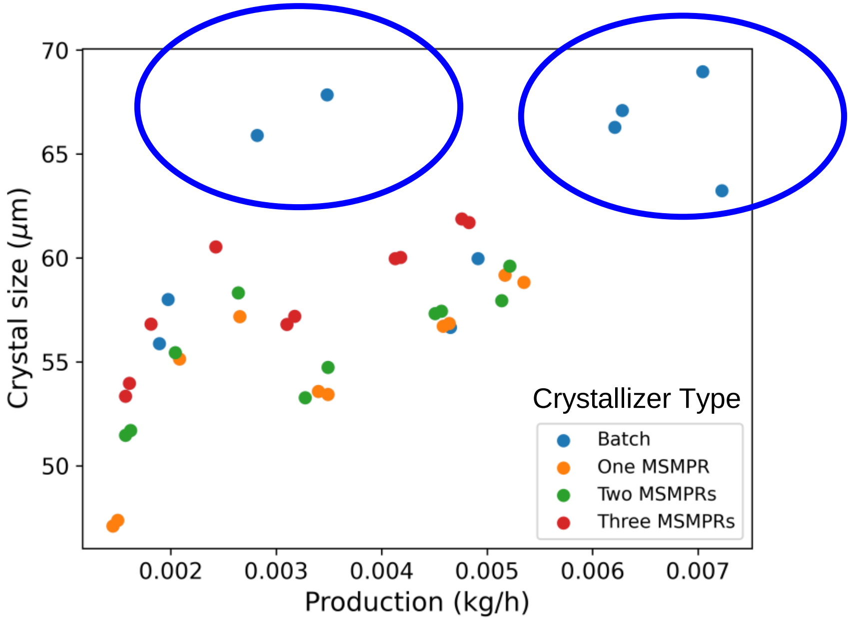
40 Manufacturing Pathways

15. Laky et al. (2022) *AIChE Annual Meeting, Phoenix. AZ*

PharmaPy Digital Twins Can Help Identify Optimal Manufacturing Routes[15]

- Major observations from pareto-optimal curves:
  1. First reactor should not be batch
  2. Batch cooling crystallization yields much better size and production than the continuous MSMPRs





- Heuristic potentially reduces computational time:
  - We know reaction 1 is fast, so choose continuous reactor 2 if reactor 1 is continuous (reduces computational requirement by 25%)

15. Laky et al. (2022) *AIChE Annual Meeting, Phoenix. AZ*

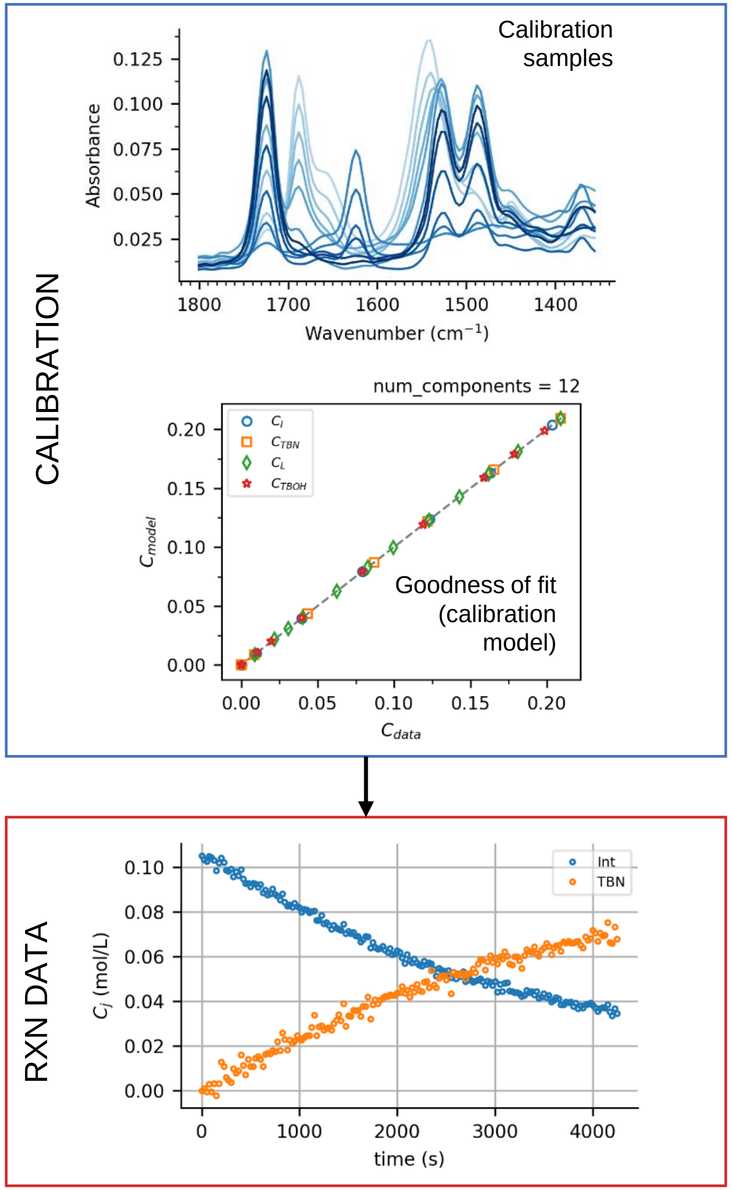
PharmaPy Digital Twins Can Aid in Uncertainty Identification[14]

- Estimate kinetic parameters Lomustine reaction mechanism
- Lomustine (L) synthesis:

$$I + TBN \xrightarrow{THF} \boldsymbol{L} + TBOH$$

- Partial Factorial DoE:

Experiment	$C_I : C_{TBN}$ ratio	T (°C)
1	1	15
2	2	15
3	1	25
4	2	25
5	1	35
6	2	35



14. Casas-Orozco et al. (2023), *Chem. Eng. Science*

PharmaPy Digital Twins Can Aid in Uncertainty Identification[14]

Best of 5 reaction mechanisms considered [14]

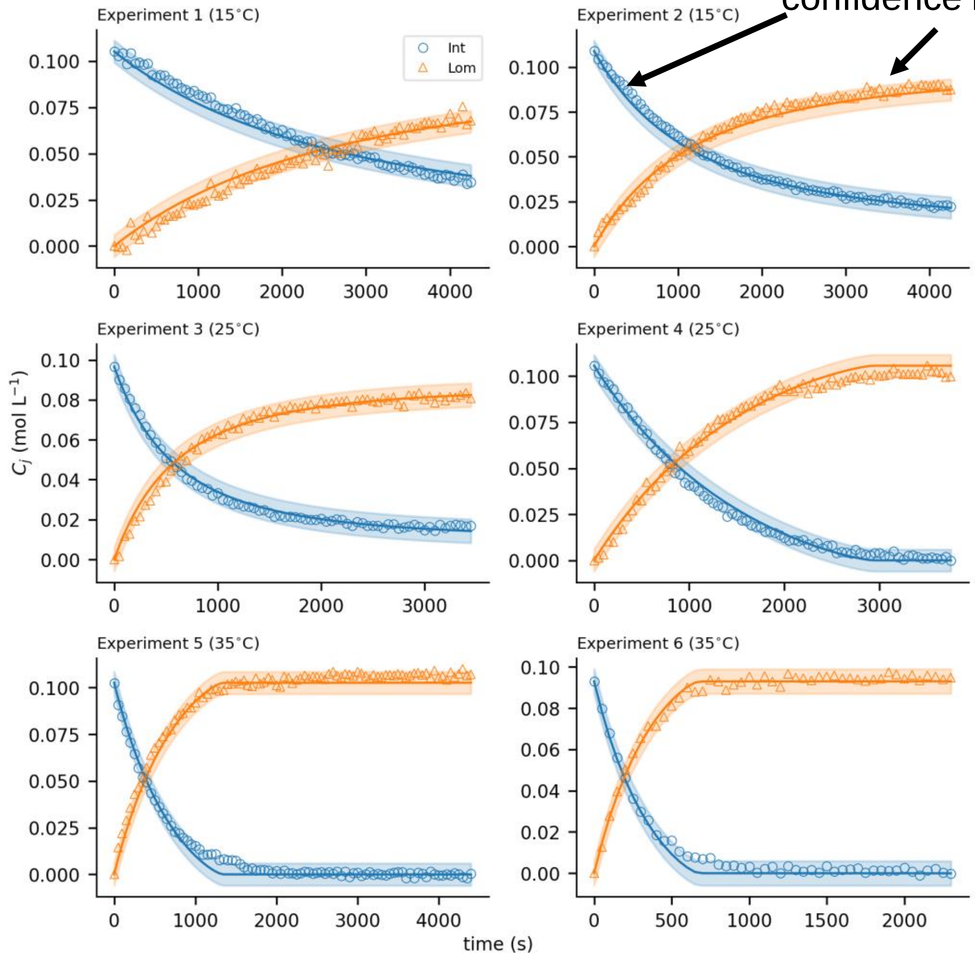
$$r = k\left(C_I\right)^{\alpha_I}\left(C_{TBN}\right)^{\alpha_{TBN}}$$

$$k = \exp\left(\phi_1 + \exp\left(\phi_2\right)\left(\frac{1}{T_{mean}} - \frac{1}{T}\right)\right)$$

$$\phi_1 = \ln\left(A\right) - \frac{E_a}{RT_{mean}}$$

$$\phi_2 = \ln\left(\frac{E_a}{R}\right)$$

95 % asymptotic confidence interval



- Partial Factorial DoE:

Experiment	C_I: C_{TBN} ratio	T (°C)
1	1	15
2	2	15
3	1	25
4	2	25
5	1	35
6	2	35

14. Casas-Orozco et al. (2023), *Chem. Eng. Science*

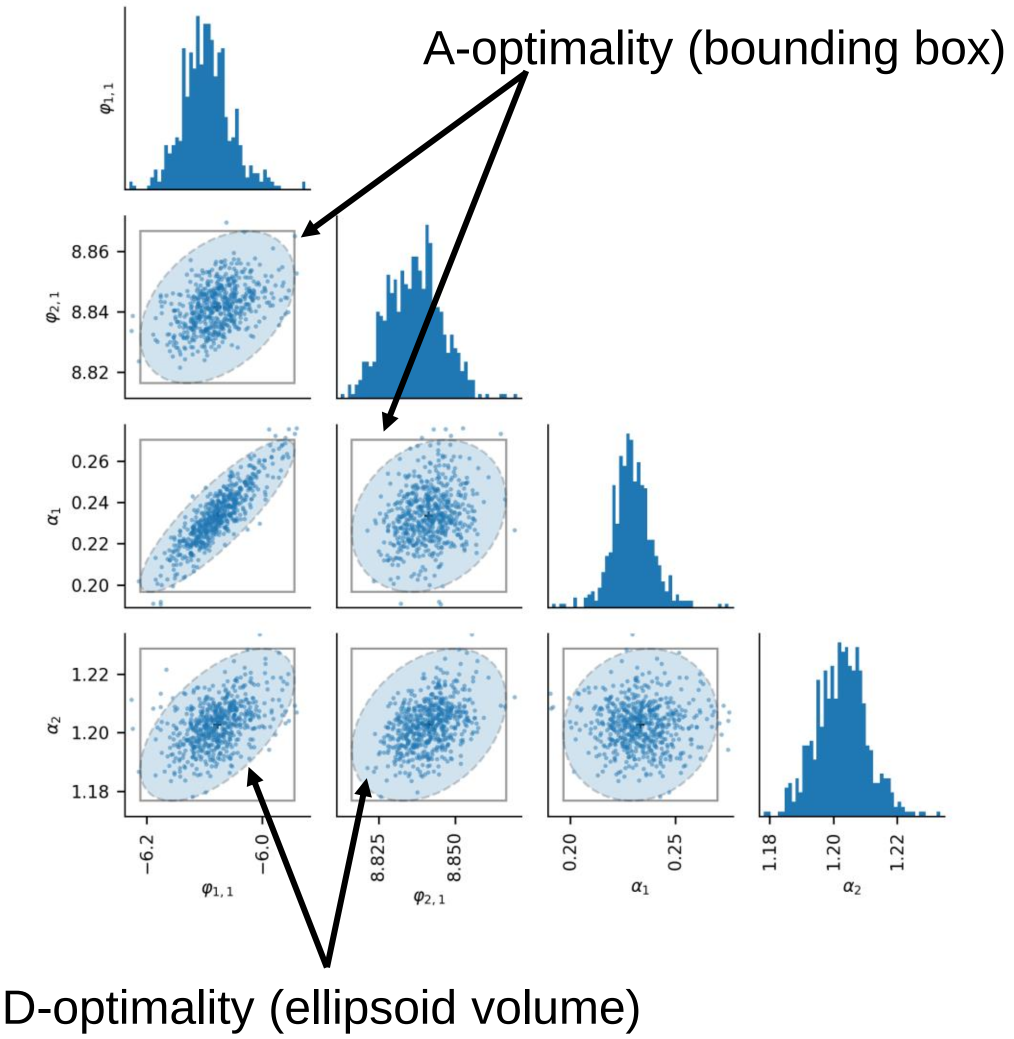
PharmaPy Digital Twins Can Aid in Uncertainty Identification[14]

- Bootstrapping with PharmaPy:

$$\mathbf{y}_{boot,k} = \mathbf{y}(\boldsymbol{\theta}^*) + \varepsilon_{boot,k}(\boldsymbol{\theta}^*), \qquad k = \{1, \ldots, n_{samples}\}$$

Final residual $\boldsymbol{\varepsilon}(\boldsymbol{\theta}^*)$ is sampled **with replacement** to produce a set of $\boldsymbol{\varepsilon}_{\boldsymbol{boot,k}}(\boldsymbol{\theta}^*)$

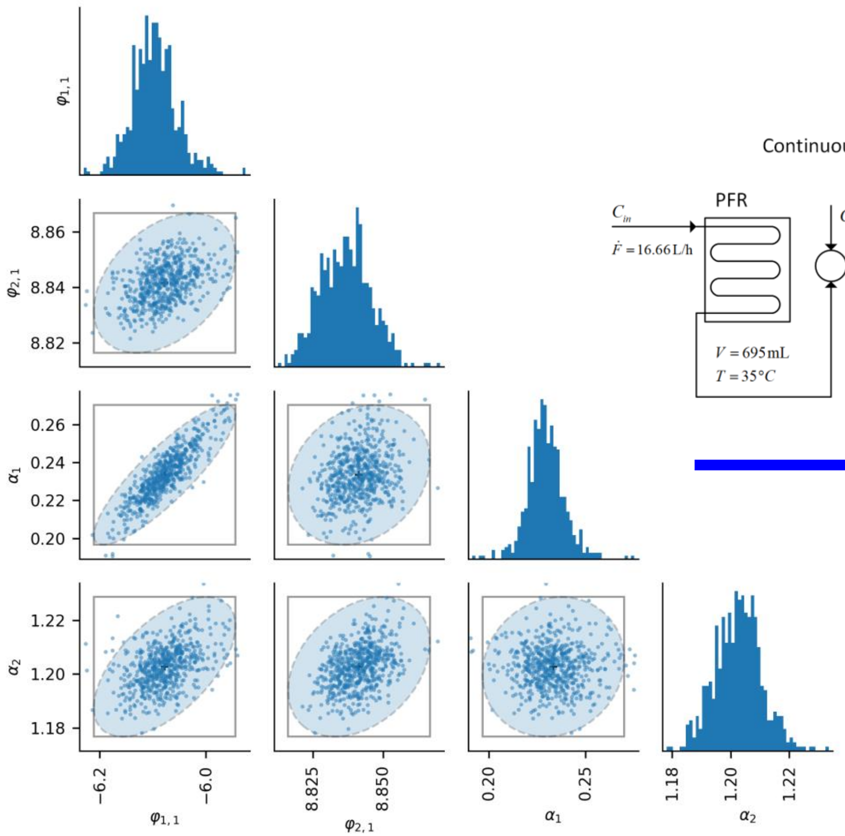
Parameter	Estimate	95% CI (asymptotic)
φ_1	-6.073	±1.39
φ_2	8.818	±0.18
α_I	0.246	±9.52
α_{TBN}	1.189	±1.38

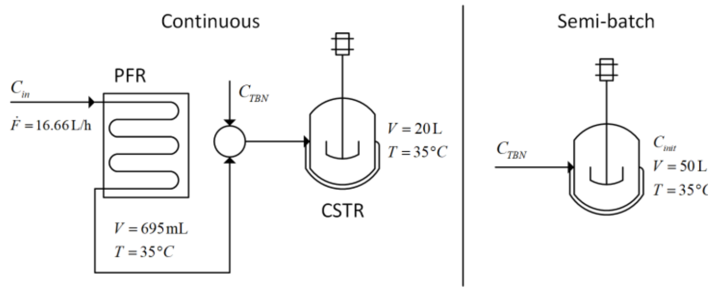


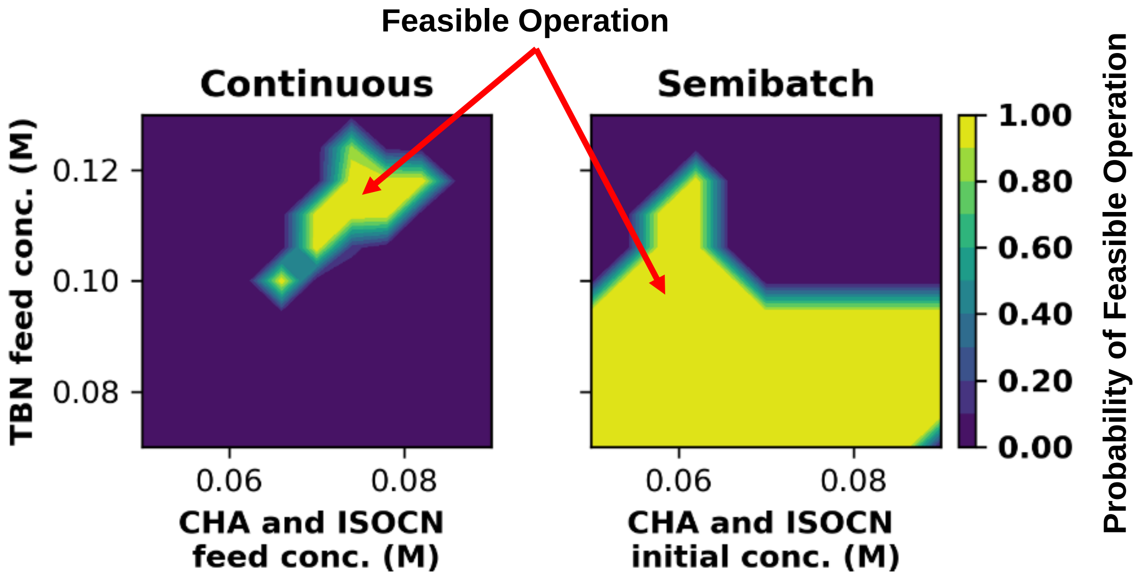
14. Casas-Orozco et al. (2023), *Chem. Eng. Science*

PharmaPy Digital Twins Can Incorporate Uncertainty for More Robust Optimization and Design Space Analysis[15]

- Using covariance information/confidence intervals, identify what operating regions are feasible (probabilistic design space)
- Semi-batch has a larger region of feasibility than the continuous process







Used process metrics (CQA's):

- API maximum **solubility**
- Intermediate **conversion**
- Intermediate:API **ratio**

15. Laky et al. (2022), *Comp. Aided Chem. Eng.*

Optimal Process Design
(Pharmaceuticals and Energy Systems)

Purdue University

University of Notre Dame

PYOMO

Dr. Daniel Casas-Orozco

Prof. Gintaras Reklaitis

Prof. Zoltan Nagy

IDAES
Institute for the Design of Advanced Energy Systems

NATIONAL ENERGY TECHNOLOGY LABORATORY

Prof. Alexander Dowling

Nicole Cortes

UNIVERSITY OF NOTRE DAME

PURDUE UNIVERSITY

Energy Systems Require Increased Flexibility and Must Adapt to Modern Sustainable Energy Objectives (e.g., Hydrogen)

- Increased renewable generation requires innovative energy demand solutions

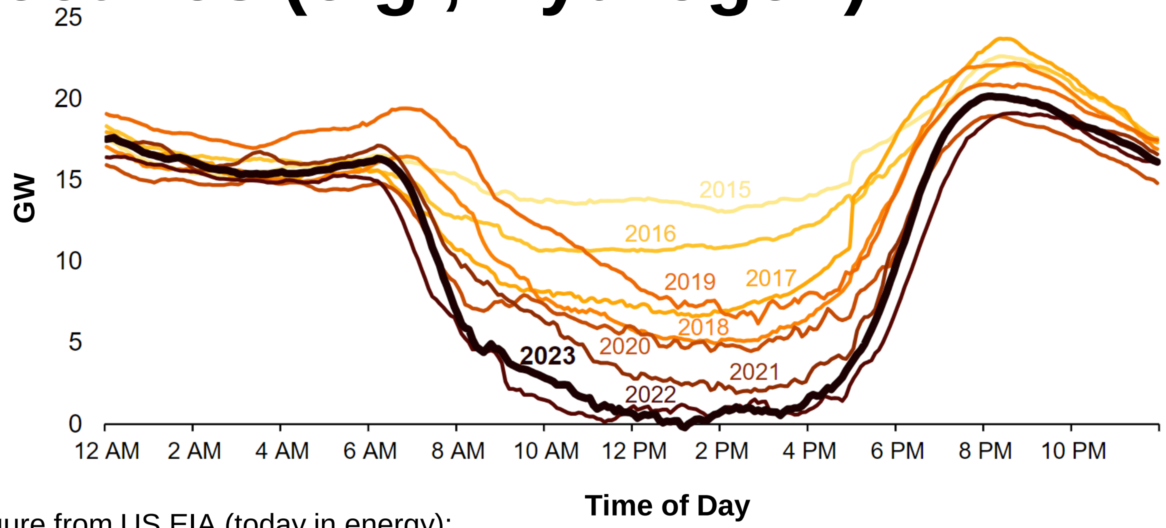


Figure from US EIA (today in energy): (https://www.eia.gov/todayinenergy/detail.php?id=56880)

- Hydrogen as an emerging fuel source needs technological evaluation paradigms to understand benefit

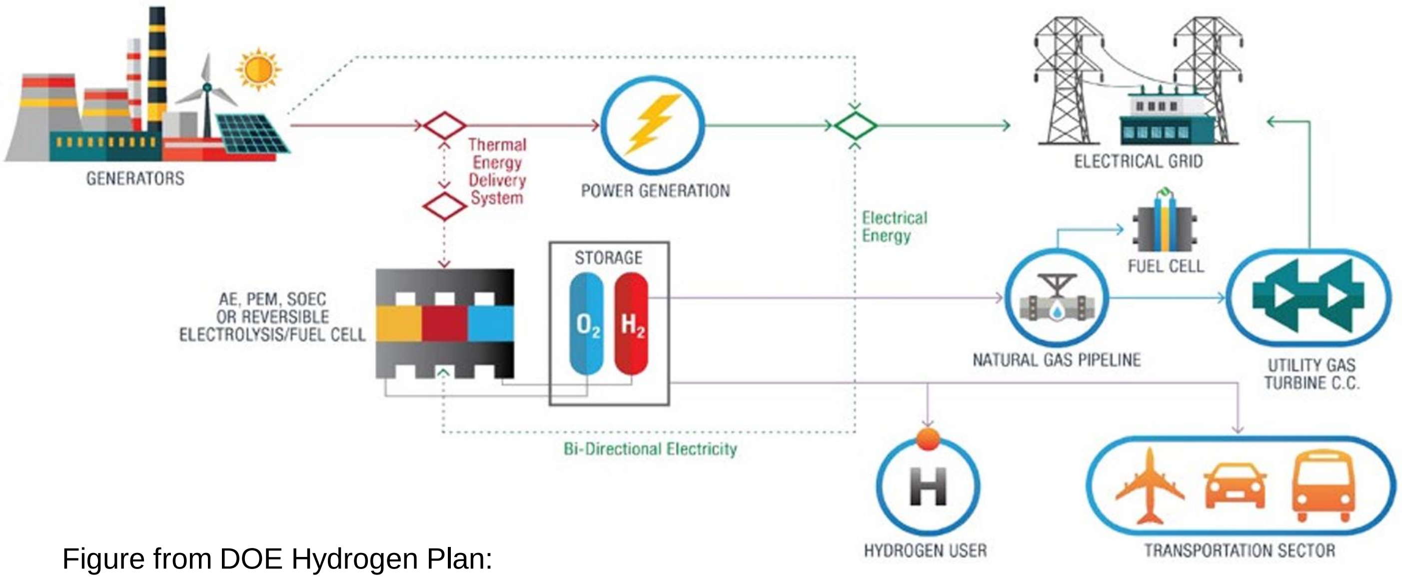


Figure from DOE Hydrogen Plan: (https://www.hydrogen.energy.gov/pdfs/hydrogen-program-plan-2020.pdf)

61 Markets Used to Evaluate Emerging Technologies

Figure from FERC (https://www.ferc.gov/electric-power-markets)

- 15 historical markets (2019)
- 15 "current" markets (2022)
- 16 forecasted scenarios (2030)
  - "Princeton"
- 10 forecasted scenarios (2035)
  - "NREL"
- 5 forecasted scenarios (2035)
  - "NETL"

Emerging Co-production Technologies Make a Profit in Most Scenarios

Percentage of scenarios that make profit at each hydrogen selling price

Process Concept	$1.00\$\cdot kg^{-1}$	$1.50\$\cdot kg^{-1}$	$2.00\$\cdot kg^{-1}$	$2.50\$\cdot kg^{-1}$	$3.00\$\cdot kg^{-1}$
NGCC	13%	13%	13%	13%	13%
SOFC	54%	54%	54%	54%	54%
NGCC + SOEC	8%	11%	16%	62%	80%
rSOC	54%	77%	97%	100%	100%
SOFC + SOEC	46%	79%	98%	100%	100%
SOEC	10%	49%	74%	87%	98%

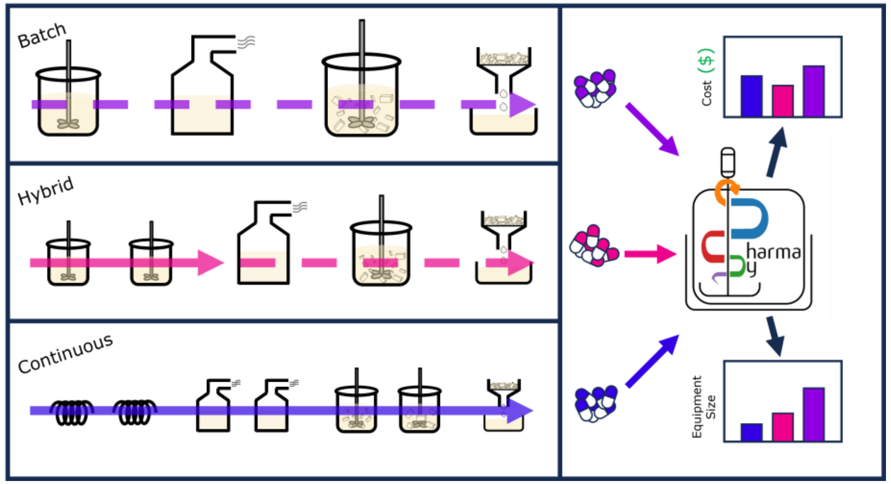
- At sufficient hydrogen price ($2.50 +), even the existing thermal generation technology with co-production (NGCC +SOEC) sees profit in over half of the market scenarios

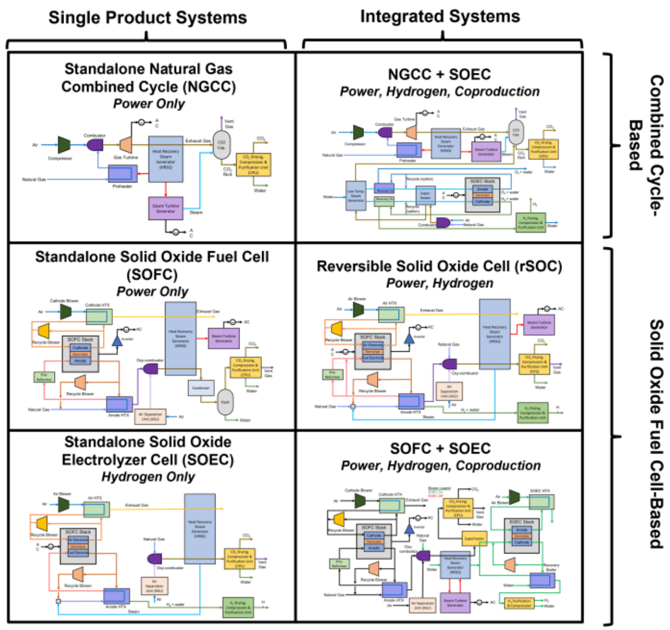
UNIVERSITY OF NOTRE DAME

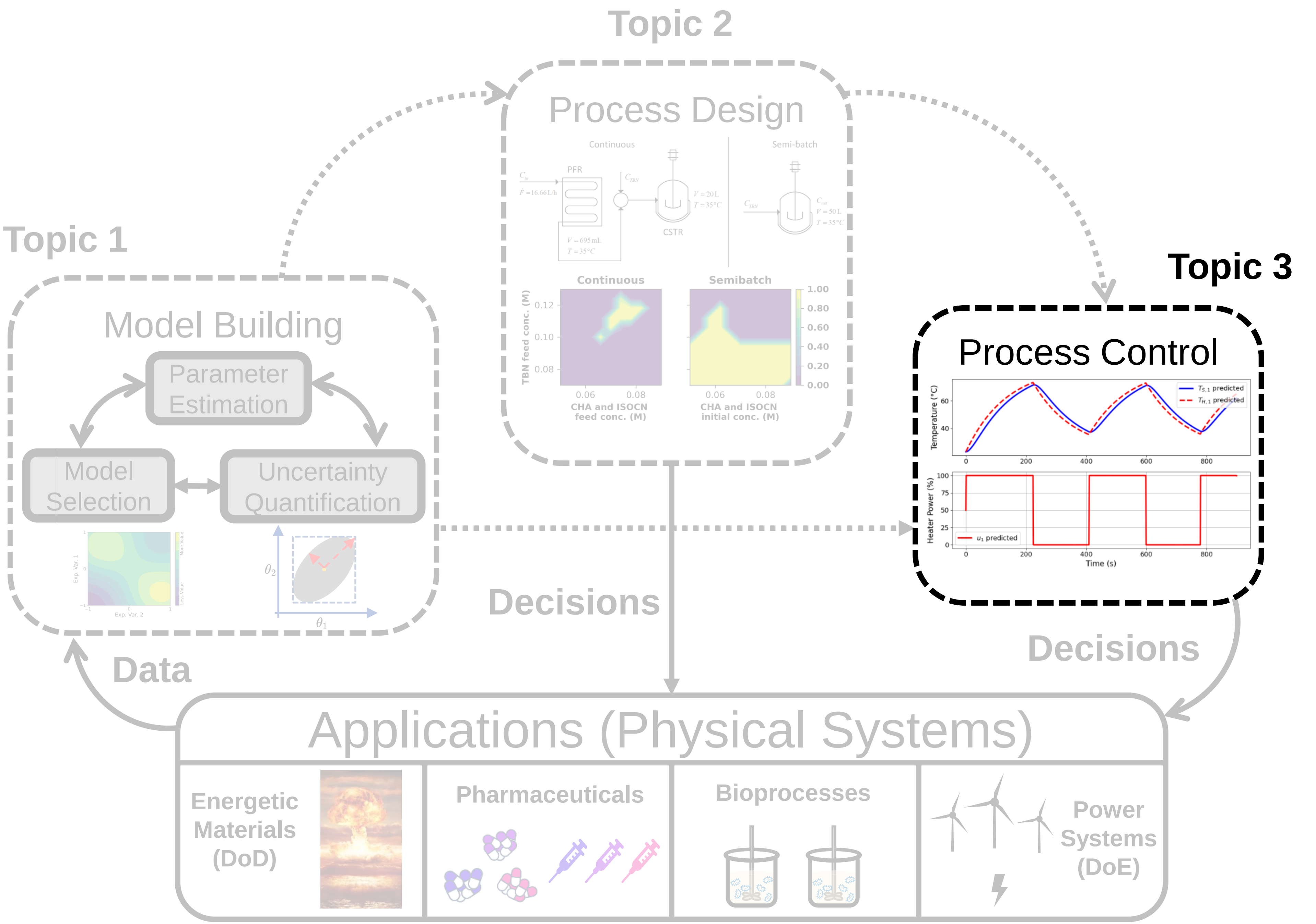
Optimal Design Conclusions

- PharmaPy can be leveraged to develop impactful digital twins
    1. Simulate process models
    2. Optimize process conditions
    3. Compare candidate models
    4. Generate and utilize model uncertainty
    5. Evaluate process superstructure to inform potential manufacturing routes

- Data-driven and mechanistic models can be combined to evaluate emerging energy technologies using digital twins





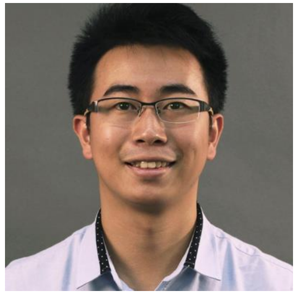

Topic 2
Process Design
Continuous
Semi-batch
PFR
CSTR
Continuous
Semibatch
TBN feed conc. (M)
CHA and ISOCN feed conc. (M)
CHA and ISOCN initial conc. (M)
Topic 1
Model Building
Parameter Estimation
Model Selection
Uncertainty Quantification
Topic 3
Process Control
Temperature (°C)
Heater Power (%)
Time (s)
Decisions
Decisions
Data
Applications (Physical Systems)
Energetic Materials (DoD)
Pharmaceuticals
Bioprocesses
Power Systems (DoE)

Data Reduction using Topological Data Analysis

University of Wisconsin-Madison

Prof. Victor Zavala

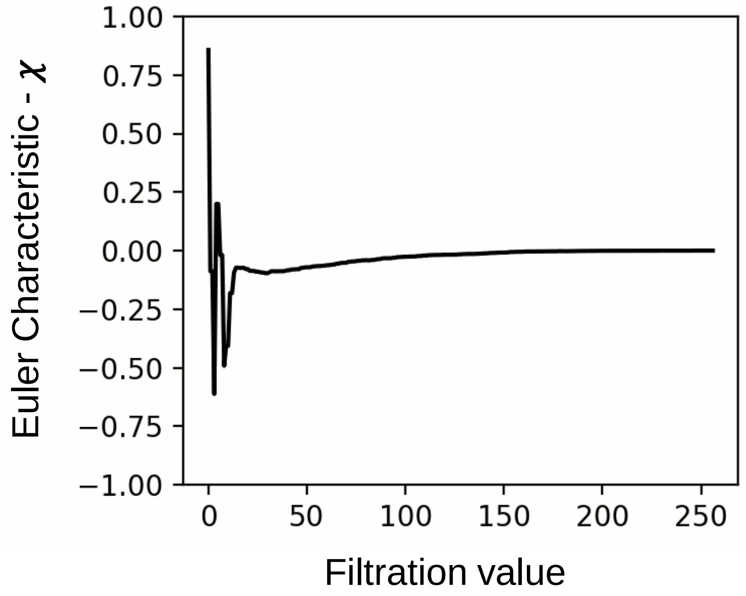
Dr. Alexander Smith



Dr. Shengli (Bruce) Jiang

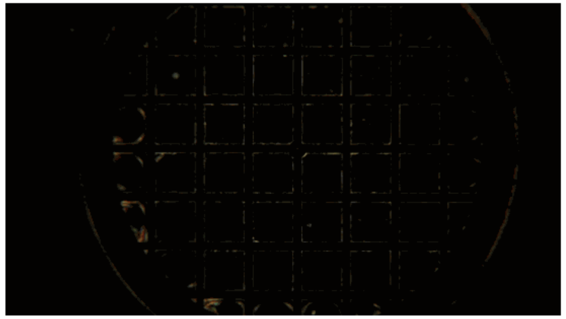
Topological Data Analysis for Classification and Quality Control

100x slower than real time

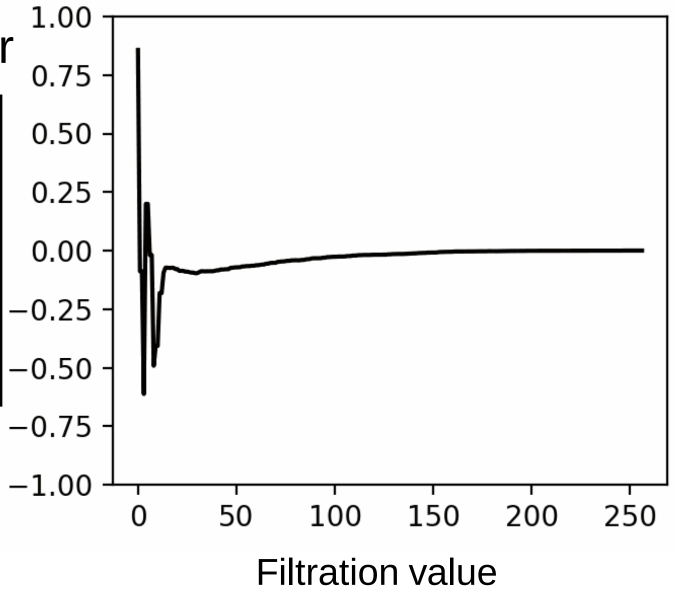


Computation with GUDHI
(state-of-the-art tool)

Liquid crystal chemical sensor



Real-time data processing



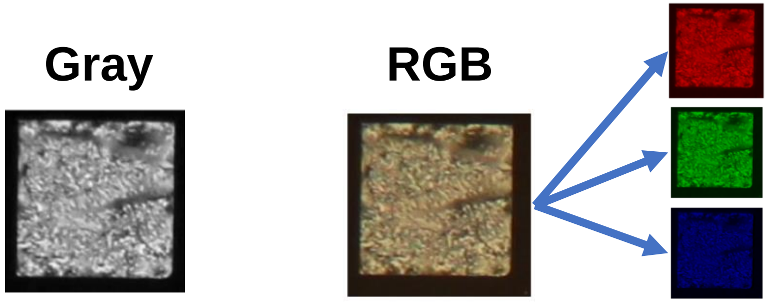
Computation with vertex contributions (**VC**)
(our implementation)

Data in Engineering is Often Represented as a Field

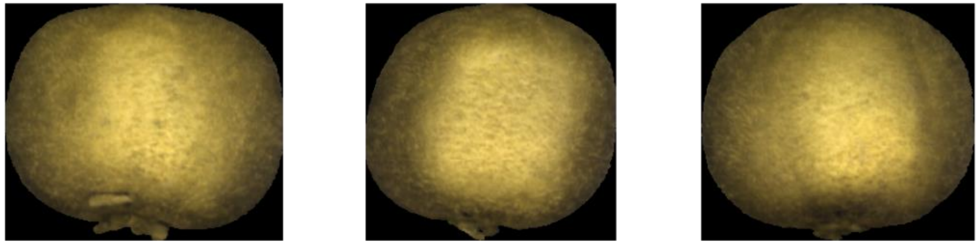
- Images
  - Grayscale (2D field)
  - RGB (~3D field)
  - Hyperspectral (3D field)
- Space-Time Data
  - Molecular dynamics
  - Computational fluid dynamics
  - *Any spatial data (e.g., GIS)*

Gray

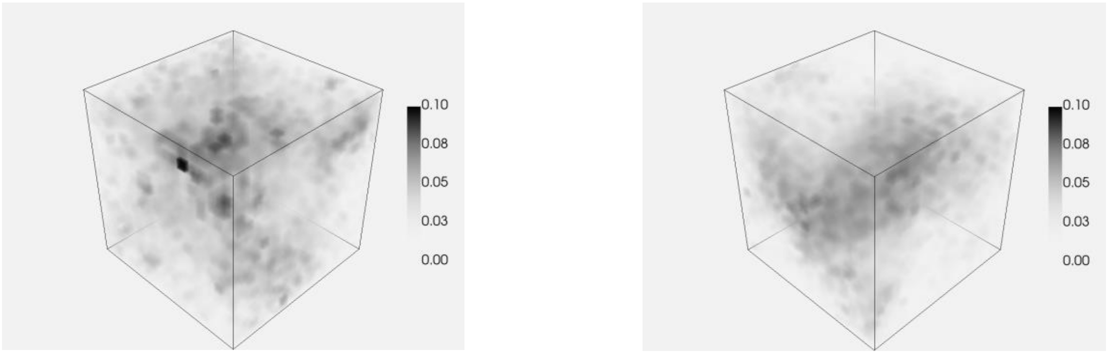
RGB



Hyperspectral

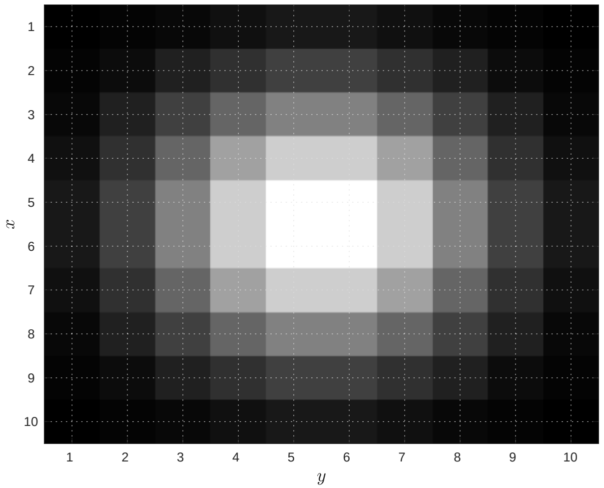


Molecular Dynamics



Field Data – Representation, Transformation, Information

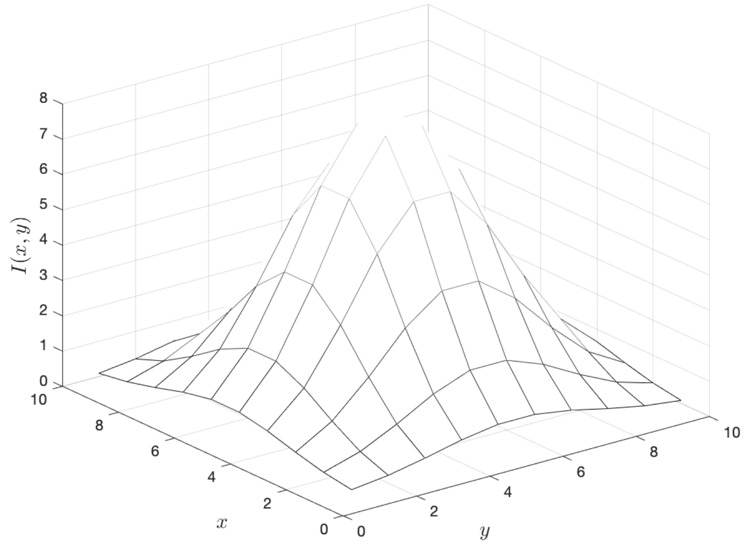
Image



Matrix

0	0	0	1	1	1	1	0	0	0
0	1	1	2	2	2	2	1	1	0
0	1	2	3	4	4	3	2	1	0
1	2	3	5	6	6	5	3	2	1
1	2	4	6	8	8	6	4	2	1
1	2	4	6	8	8	6	4	2	1
1	2	3	5	6	6	5	3	2	1
0	1	2	3	4	4	3	2	1	0
0	1	1	2	2	2	2	1	1	0
0	0	0	1	1	1	1	0	0	0

Function/Manifold



SVD $\mathbf{T} = \mathbf{X} \cdot \mathbf{W}$

Convolution $\mathbf{T} = \mathbf{W} * \mathbf{X}$

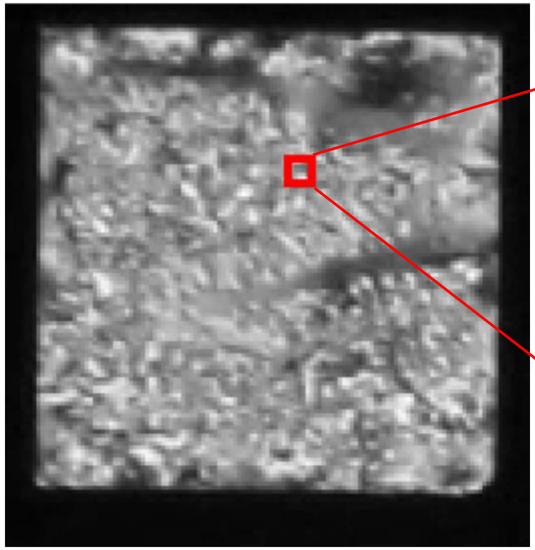
Fourier $\mathbf{T} = \mathcal{F}(\mathbf{X})$

Different representations lead to different transformations

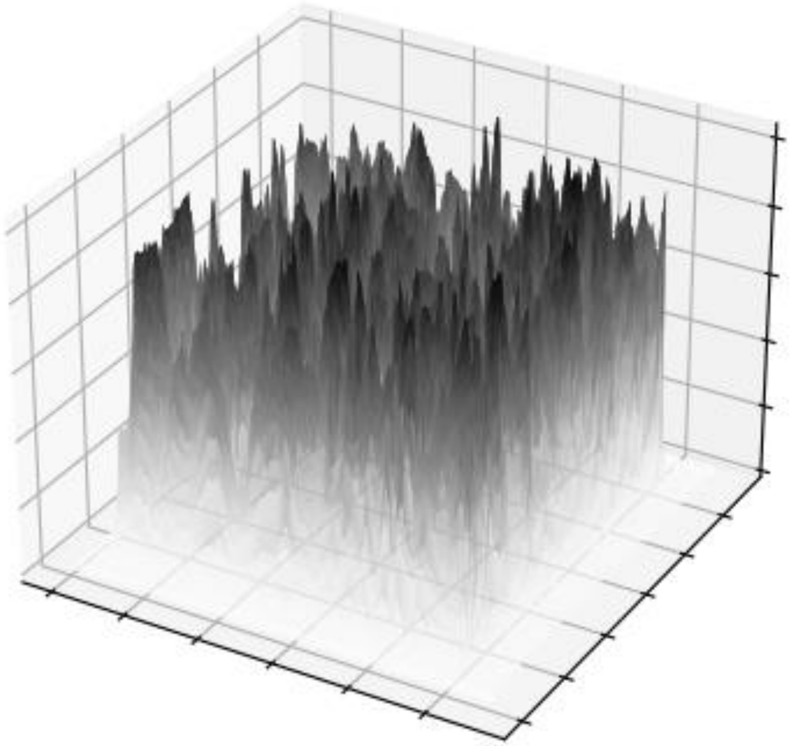
Different transformations extract different types of information

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Field Data – Topological Data Analysis (TDA)



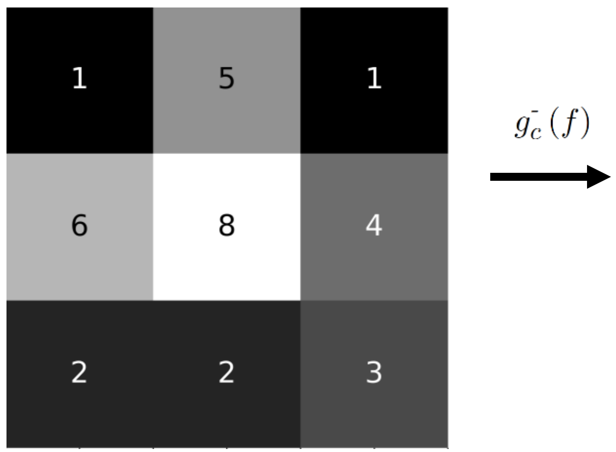
174	211	210	176	162	159
183	186	217	199	158	118
179	145	120	118	108	101
167	152	141	135	125	99
161	151	169	192	193	125
167	141	189	230	247	190

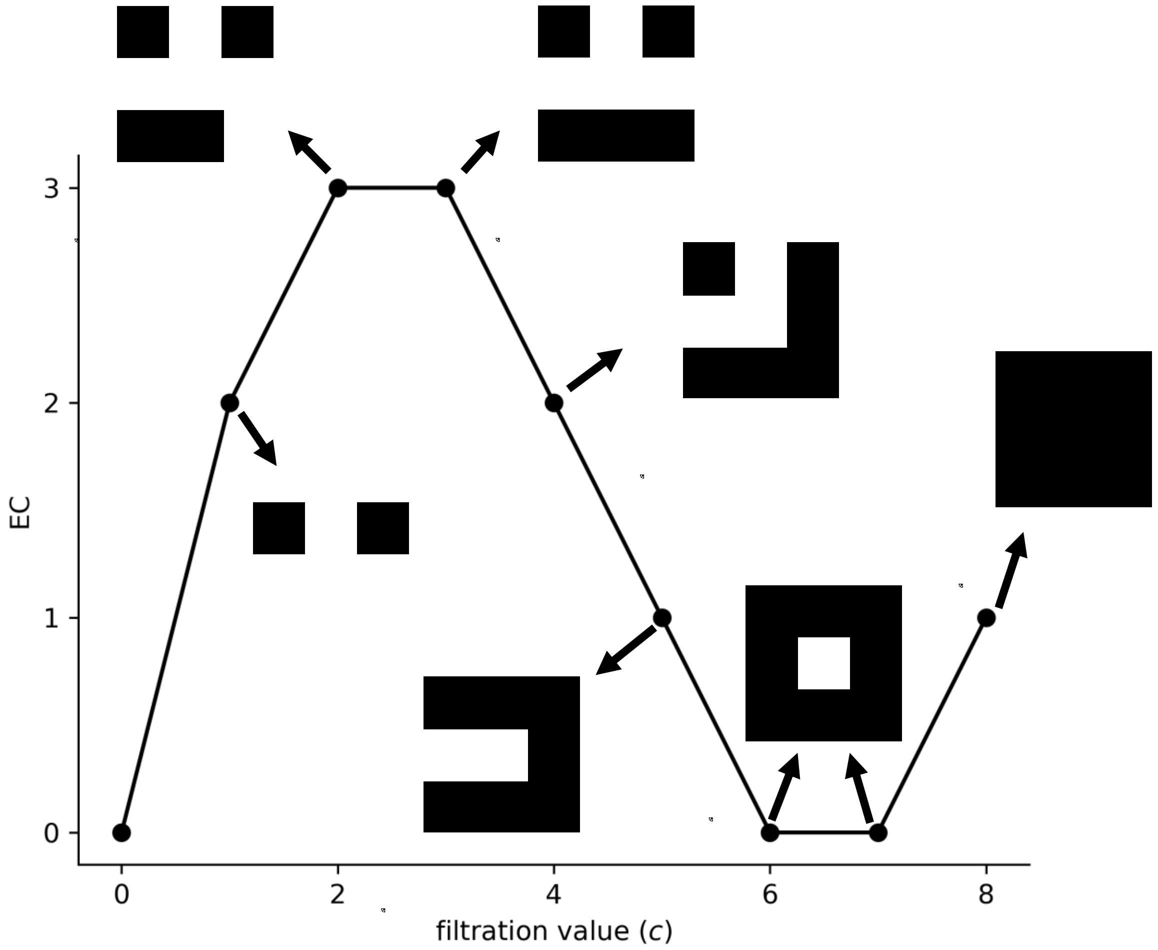


2D fields (i.e., images) reveal topographical features when represented as a 3D surface

Filtration – Euler Characteristic Curve

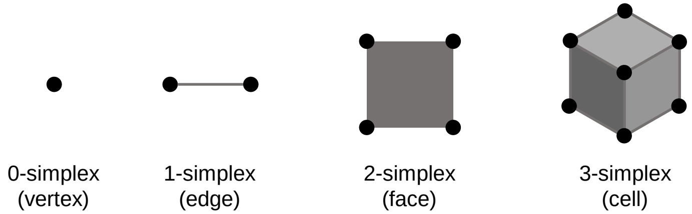
$$g_c^-(f) = \{(x_w, x_h) | f(x_w, x_h) \le c\}$$





Euler Characteristic (EC)

$$\chi = V - E + F - C$$



Filtration yields a series of binary fields for the EC computation

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

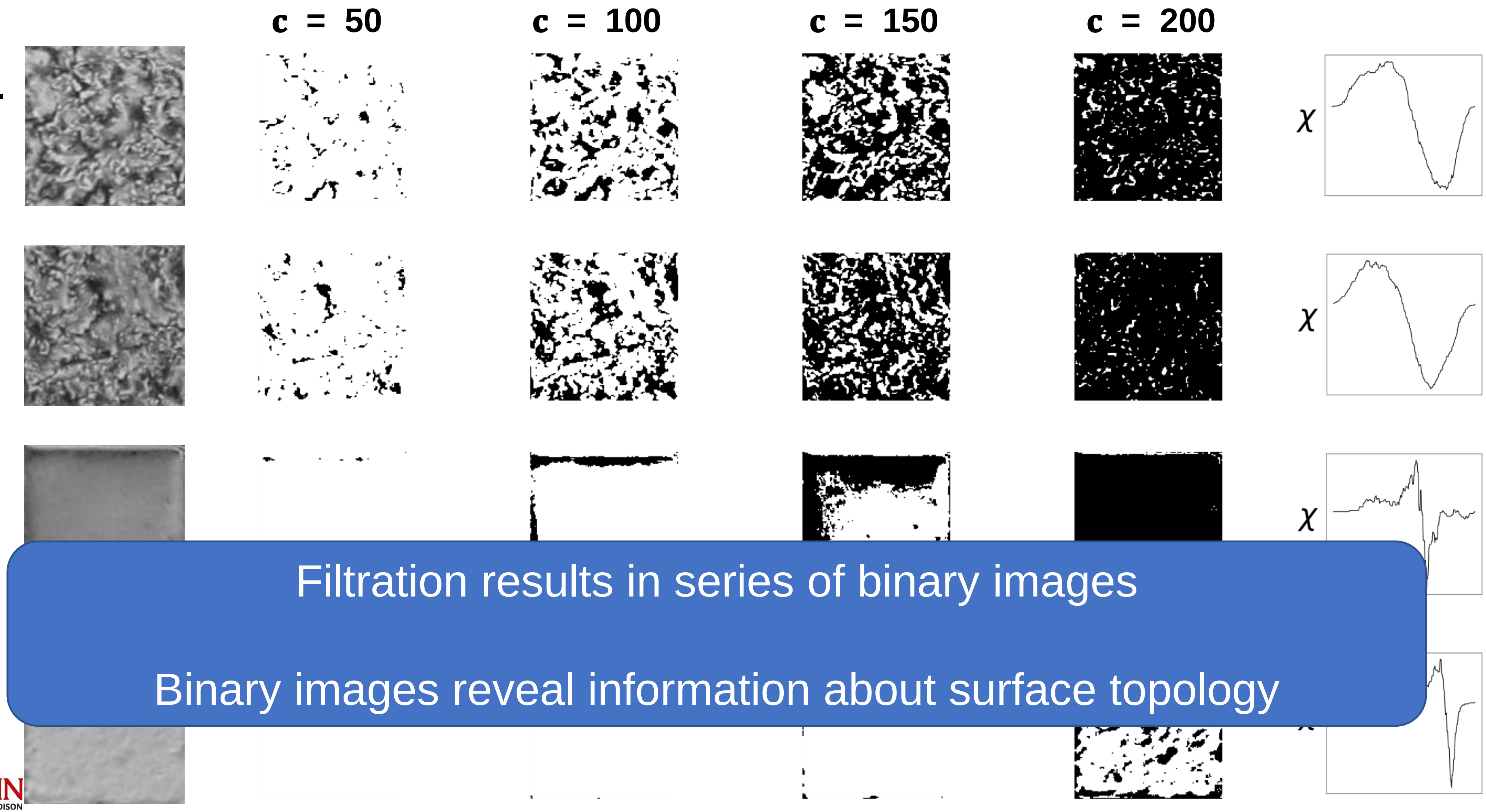
Filtration Example – Liquid Crystal Sensors

SO_2 Conc.
0.5 ppm

1 ppm

2 ppm

5 ppm



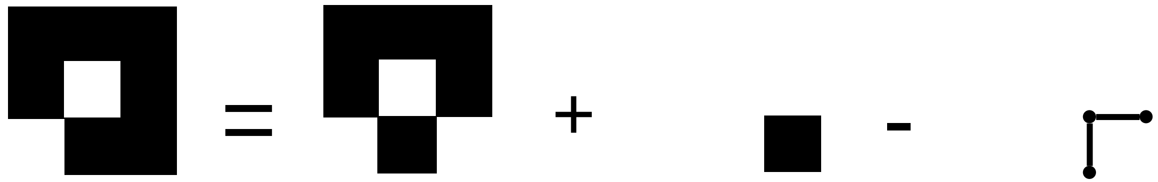
Filtration results in series of binary images

Binary images reveal information about surface topology

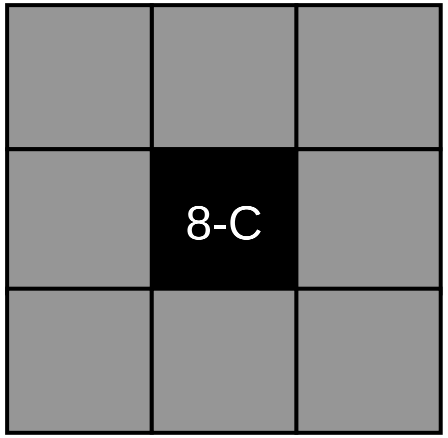
Euler Characteristic Generalization – Problems

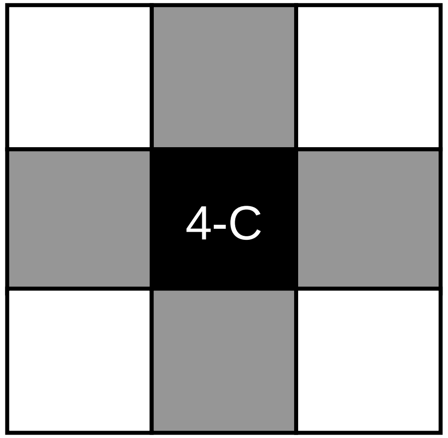
- Euler characteristic follows the **inclusion-exclusion principle**

$$\chi(A \cup B) = \chi(A) + \chi(B) - \chi(A \cap B)$$



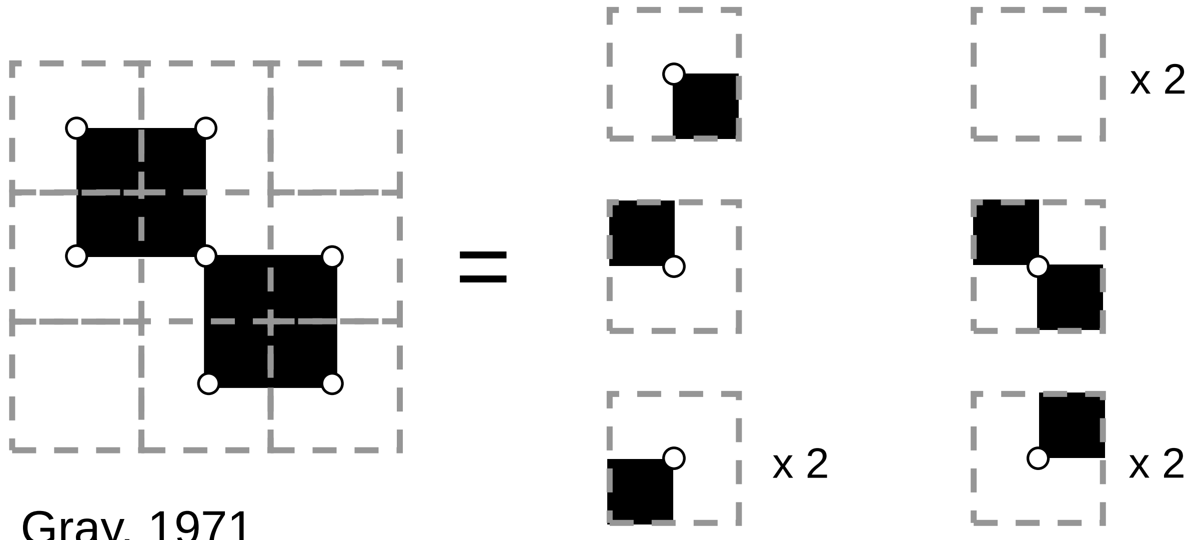
- Should faces be considered connected via edges (4-C), or via vertices (8-C)?





- This principle is key for enabling **parallel computation**

Avoiding Repetitive Evaluation through Abstraction



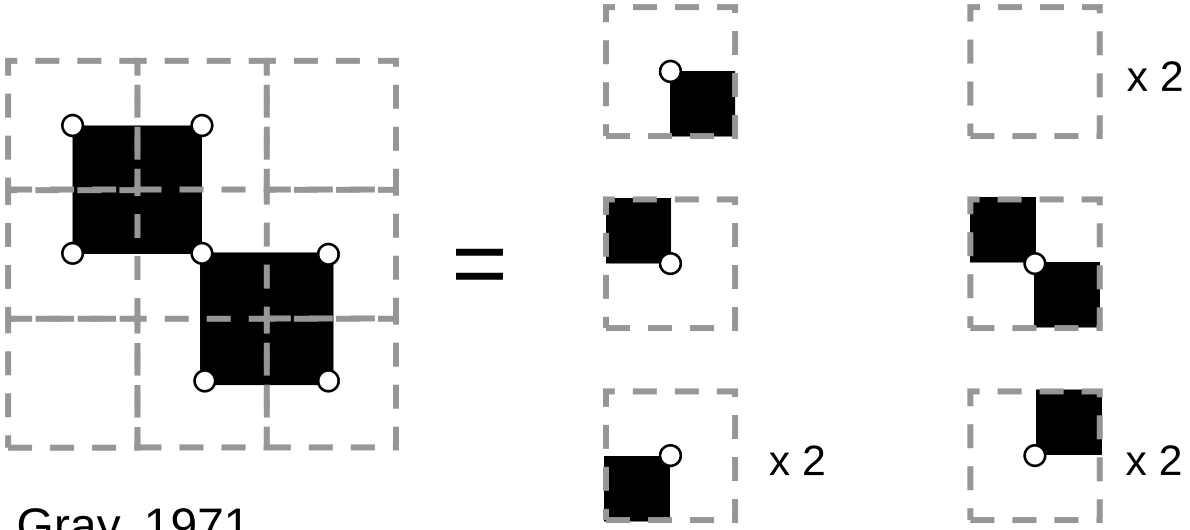
Gray, 1971

0

$$\chi\left(A \cup B\right) = \chi\left(A\right) + \chi\left(B\right) - \chi\left(A \cap B\right)$$

17. Gray, S. B. Local properties of binary images in two dimensions. IEEE Transactions on Computers, C-20(5):551–561, 1971.

Avoiding Repetitive Evaluation through Abstraction



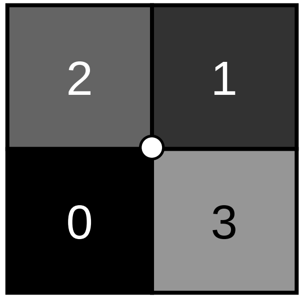
Gray, 1971

Type	V (8-C)	E (8-C)	F (8-C)	EC (8-C)	V (4-C)	E (4-C)	F (4-C)	EC (4-C)	Perimeter	Area
	0	0	0	0	0	0	0	0	0	0
	1	2 * (0.5)	1 * (0.25)	0.25	1	2 * (0.5)	1 * (0.25)	0.25	1	0.25
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	1	4 * (0.5)	3 * (0.25)	-0.25	1	4 * (0.5)	3 * (0.25)	-0.25	1	0.75
	1	4 * (0.5)	4 * (0.25)	0	1	4 * (0.5)	4 * (0.25)	0	0	1

Even with this method, we still must compute contributions at *each* filtration value (not scalable)

17. Gray, S. B. Local properties of binary images in two dimensions. IEEE Transactions on Computers, C-20(5):551–561, 1971.
18. Laky and Zavala, 2023, *Digital Discoery*.

Avoiding Repetitive Evaluation through Abstraction



birth – 0
death – 1

birth – 1
death – 2

birth – 2
death – 3

birth – 3
no death

$$
\begin{array}{c|cccccccccccc}
\text{index} & \text{birth} & \text{death} & \text{birth} & \text{death} & \text{birth} & \text{death} & \text{birth} & \text{death} & \text{birth} & \text{death} & \text{birth} & \text{death} \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0
\end{array}
$$

Only need local vertex neighborhood values → Highly parallelizable; Low memory requirements

$$\mathcal{O}(n_c l w) \longrightarrow \mathcal{O}(n_c + l w)$$

l - length of the field

w - width of the field

n_c - number of filtration values

19. Snidaro, L. and Foresti, G. Real-time thresholding with euler numbers. Pattern Recognition Letters, 24:1533–1544, 2003.

Abstraction for 3D fields

- Voxels in 3D digital images (or cells) [4]

- Voxel-sized neighborhood about a vertex yields the same savings

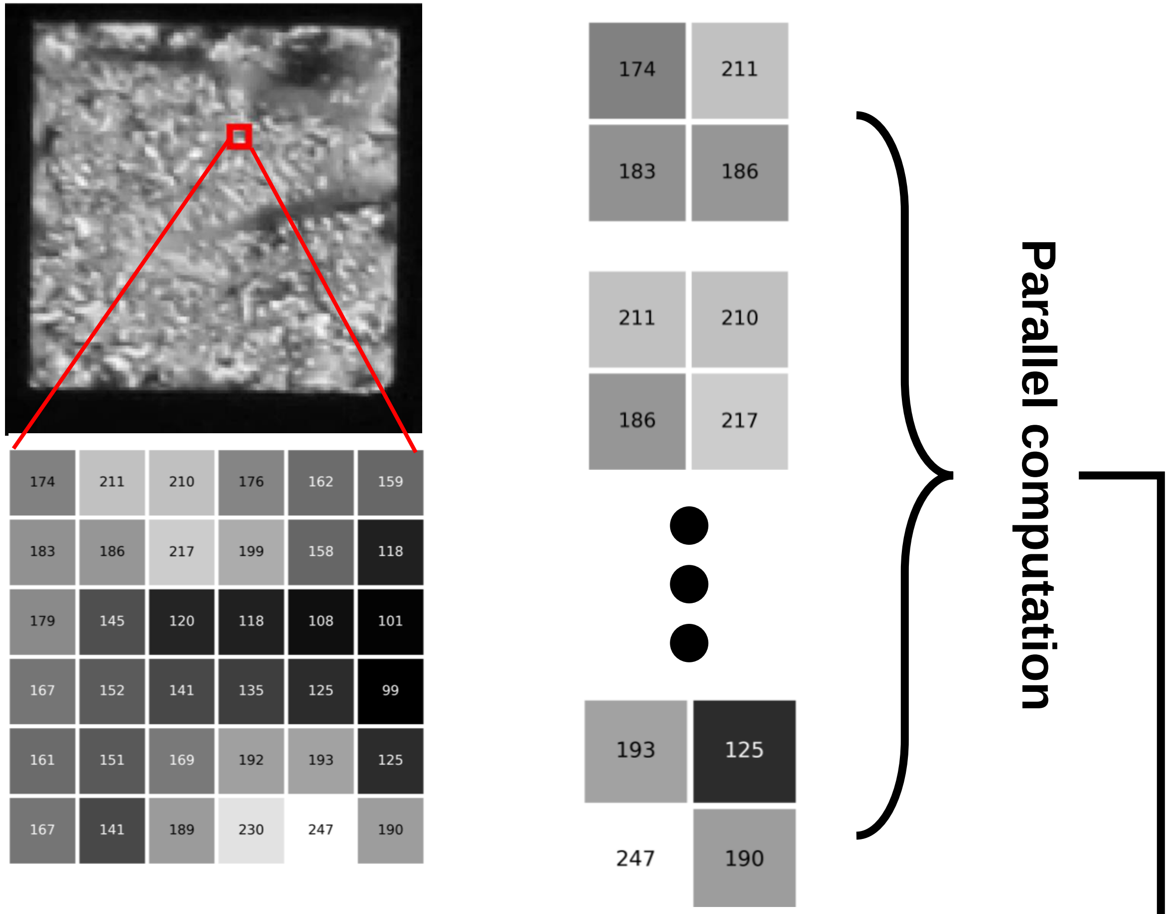
- More hassle distinguishing which pattern is represented for active cells

Type	f_{adj}	V (26-C)	E (26-C)	F (26-C)	Vox (26-C)	EC (26-C)	Perimeter	Area	Volume
	0	0	0	0	0	0	0	0	0
	0	1	3 * (0.5)	3 * (0.25)	1 * (0.125)	0.125	1.5	0.75	0.125
	1	1	4 * (0.5)	5 * (0.25)	2 * (0.125)	0	1	1	0.25
	1.414	1	5 * (0.5)	6 * (0.25)	2 * (0.125)	-0.25	3	1.5	0.25
	1. 732	1	6 * (0.5)	6 * (0.25)	2 * (0.125)	-0.75	3	1.5	0.25
	3.414	1	5 * (0.5)	7 * (0.25)	3 * (0.125)	-0.125	1.5	1.25	0.375
	4.146	1	6 * (0.5)	8 * (0.25)	3 * (0.125)	-0.375	2.5	1.75	0.375
	4.243	1	6 * (0.5)	9 * (0.25)	3 * (0.125)	-0.125	4.5	2.25	0.375
	6.828	1	5 * (0.5)	8 * (0.25)	4 * (0.125)	0	0	1	0.5
	7.243	1	6 * (0.5)	9 * (0.25)	4 * (0.125)	-0.25	3	1.75	0.5
	7.560	1	6 * (0.5)	9 * (0.25)	4 * (0.125)	-0.25	2	1.5	0.5
	7.975	1	6 * (0.5)	10 * (0.25)	4 * (0.125)	0	3	2	0.5
	8.293	1	6 * (0.5)	10 * (0.25)	4 * (0.125)	0	2	2	0.5
	8.485	1	6 * (0.5)	12 * (0.25)	4 * (0.125)	0.5	6	3	0.5
	$3.414^{\dagger}$	1	6 * (0.5)	10 * (0.25)	5 * (0.125)	-0.125	1.5	1.25	0.625
	$4.146^{\dagger}$	1	6 * (0.5)	11 * (0.25)	5 * (0.125)	0.125	2.5	1.75	0.625
	$4.243^{\dagger}$	1	6 * (0.5)	12 * (0.25)	5 * (0.125)	0.375	4.5	2.25	0.625
	$1^{\dagger}$	1	6 * (0.5)	11 * (0.25)	6 * (0.125)	0	1	1	0.75
	$1.414^{\dagger}$	1	6 * (0.5)	12 * (0.25)	6 * (0.125)	0.25	3	1.5	0.75
	$1.732^{\dagger}$	1	6 * (0.5)	12 * (0.25)	6 * (0.125)	0.25	3	1.5	0.75
	$0^{\dagger}$	1	6 * (0.5)	12 * (0.25)	7 * (0.125)	0.125	1.5	0.75	0.875
	$0^{\dagger}$	1	6 * (0.5)	12 * (0.25)	8 * (0.125)	0	0	0	1

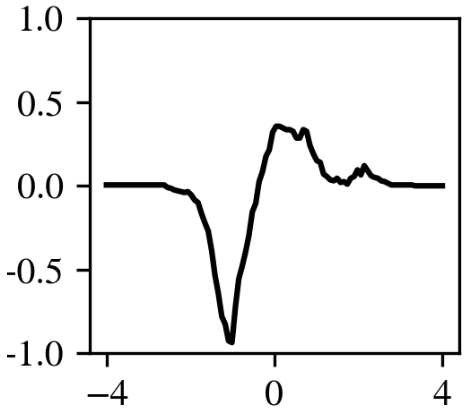
18. Laky and Zavala, 2023, *Digital Discovery*.
20. Toriwaki, J. and Yonekura, T. Euler number and connectivity indexes of a three dimensional digital picture. Forma, 17:183–209, 2002.

Software Implementation

- All code written in **C++** and called from an interface in Python
- Exploit high-powered computing hardware for parallelization
  - Intel Xeon E5-2697—2.7 GHz
    - 24 Cores
    - 256 GB RAM
- Implemented low-memory version for 2D, bitmap files (.BMP)



Type	V (8-C)	E (8-C)	F (8-C)	EC (8-C)
	0	0	0	0
	1	2 * (0.5)	1 * (0.25)	0.25
	1	3 * (0.5)	2 * (0.25)	0
	1	4 * (0.5)	2 * (0.25)	-0.5
	1	4 * (0.5)	3 * (0.25)	-0.25
	1	4 * (0.5)	4 * (0.25)	0



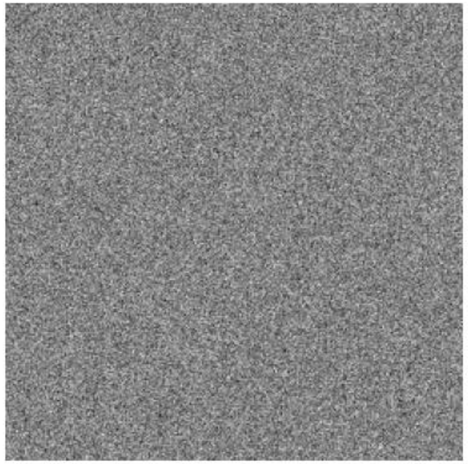
18. Laky and Zavala, 2023, *Digital Discovery*.
20. Toriwaki, J. and Yonekura, T. Euler number and connectivity indexes of a three dimensional digital picture. Forma, 17:183–209, 2002.

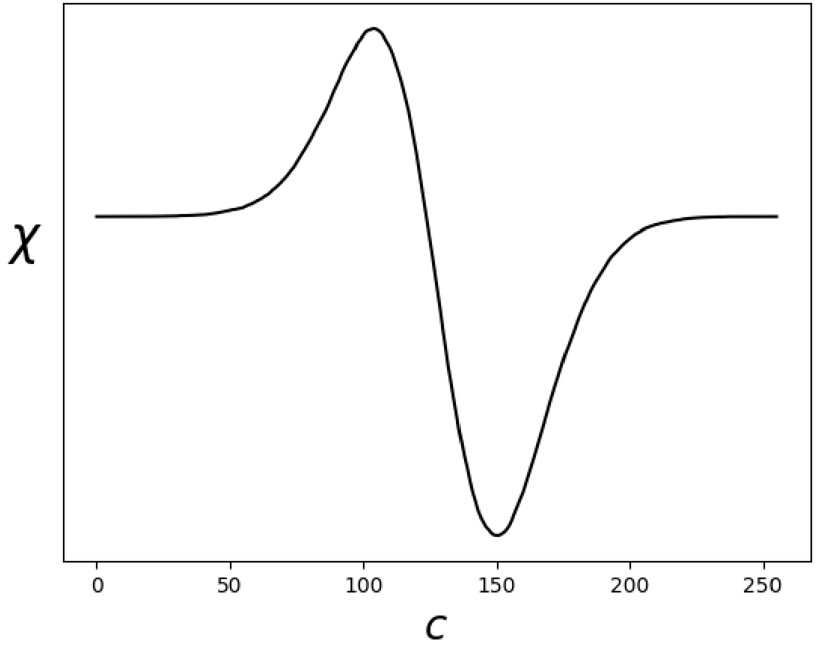
Baseline Case - Random Fields (2D and 3D)

- **“Standard” resolutions**
  - 1920x1080
  - 1280x720
  - 128x128x128
  - 256x256x256

- **Large, square systems**
  - 2048x2048
  - 4096x4096

- All field sizes tested with both **Uniform and Normal noise**





Random Fields (2D and 3D) – Significant Speedup

Type (2D)	GUDHI (s)	GUDHI (MP/s)	VC, 24 cores (s)	VC, 24 cores (MP/s)	Speedup
1280x720	5.1 ± 0.2	0.181 ± 0.005	0.0093 ± 0.0006	99.3 ± 4.9	**550**
1920x1080	14.6 ± 0.4	0.142 ± 0.004	0.0188 ± 0.0011	110.8 ± 4.6	**780**
2048x2048	31.6 ± 1.1	0.133 ± 0.004	0.0381 ± 0.0029	110.4 ± 6.0	**830**
4096x4096	145.2 ± 5.0	0.116 ± 0.003	0.210 ± 0.032	81.95 ± 13.2	**706**
Type (3D)	**GUDHI (s)**	**GUDHI (MV/s)**	**VC, 24 cores (s)**	**VC, 24 cores (MV/s)**	**Speedup**
128x128x128	42.1 ± 1.2	0.050 ± 0.001	0.0765 ± 0.0067	27.5 ± 1.2	**550**
2					

1. Significant speedup over traditional tools
   (2-3 orders of magnitude with 24-core parallelization)

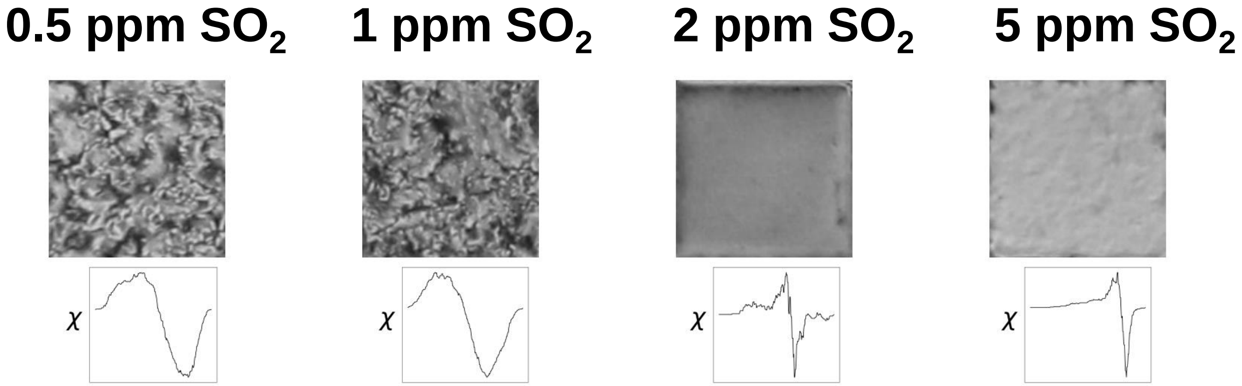
2. Can process ~100Megapixel or ~30Megavoxel images in **one second**

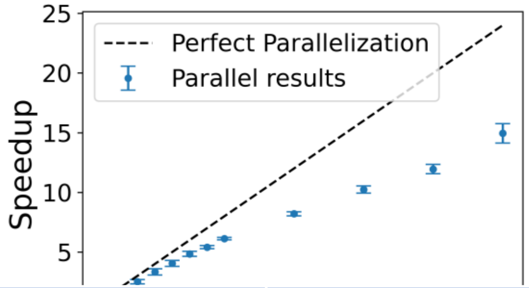
Liquid Crystal Sensor Analysis

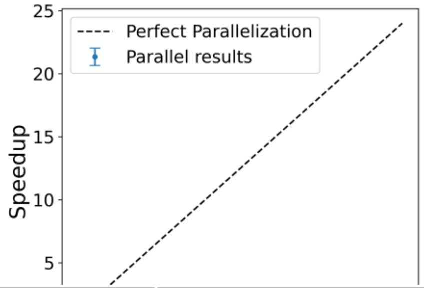
- LC micrograph responses to SO_2 under 40% RH

- Average image size of 134x134

- Small size drastically reduces efficiency of pa



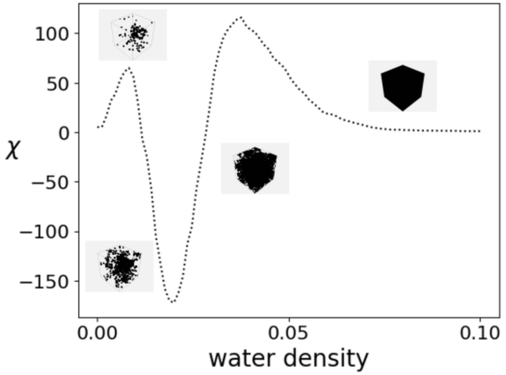




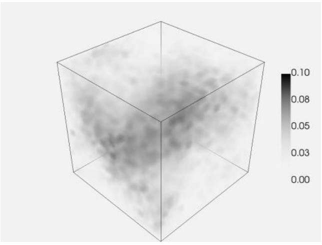
Type	GUDHI (MP/s)	VC, serial (MP/s)	VC, 12 cores (MP/s)	Speedup
SO_2 Images	0.332 ± 0.009	6.268 ± 0.171	11.33 ± 0.49	**34**
Random Field	0.271 ± 0.009	6.127 ± 0.132	11.20 ± 0.39	**41**

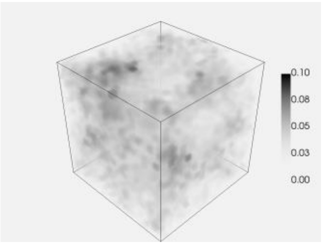
22. Jiang et al. Journal of Physical Chemistry C. 2023

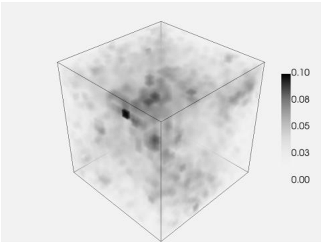
Molecular Dynamics Simulations



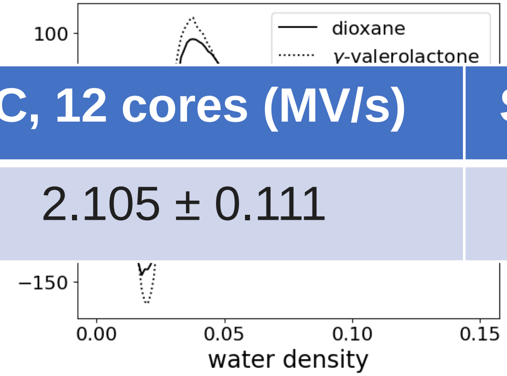
- Biomass (fructose) interaction in water-cosolvent systems
  - dioxane
  - γ-valerolactone
  - tetrahydrofuran


dioxane


γ-valerolactone


tetrahydrofuran

- Water molecule density can be represented as a 3D field



- Fie

Type	GUDHI (MV/s)	VC, serial (MV/s)	VC, 12 cores (MV/s)	Speedup
MD Fields	0.108 ± 0.004	1.065 ± 0.025	2.105 ± 0.111	**19.5**

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

23. Chew et al. Fast predictions of liquid-phase acid-catalyzed reaction rates using molecular dynamics simulations and convolutional neural networks. Chemical science, 11(46):12464–12476, 2020
24. Smith et al. Topological analysis of molecular dynamics simulations using the euler characteristic. J. Chem. Theory Comput., 19(5):1553–1567, 2023

Hyperspectral Imaging

- Determining kiwi ripeness without destructive sampling[26]

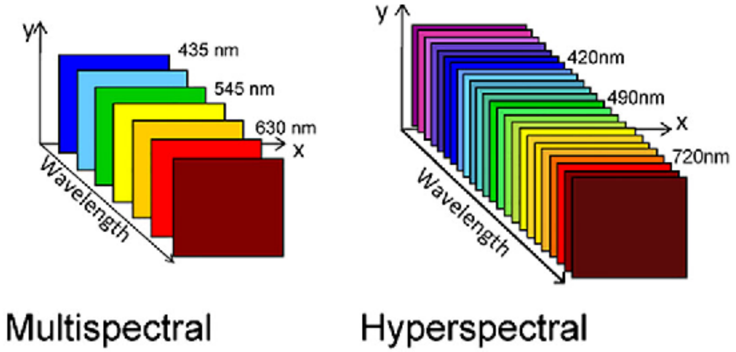
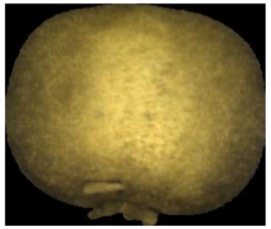
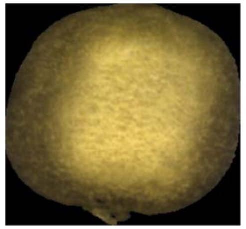


Figure from Mehta et al. 2018

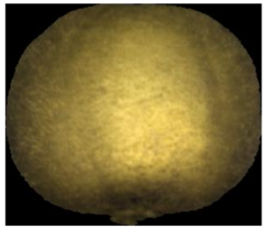
- RGB images look similar



under ripe



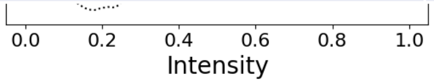
ripe



overripe

- EC
dif

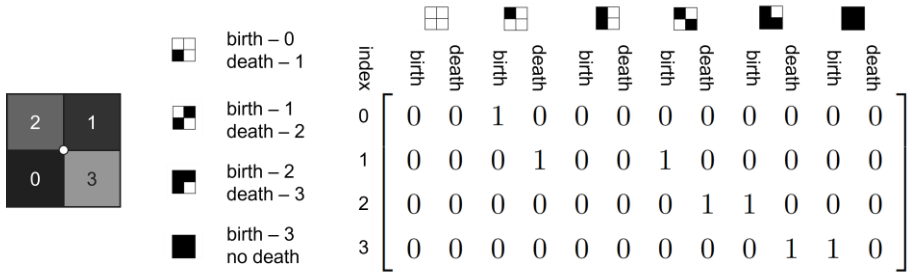
Type	GUDHI (MV/s)	VC, serial (MV/s)	VC, 24 cores (MV/s)	Speedup
Kiwi Images	0.0701 ± 0.0078	1.973 ± 0.112	34.57 ± 2.38	**493**
Random Field	0.0427 ± 0.0031	1.506 ± 0.006	28.19 ± 1.02	**660**



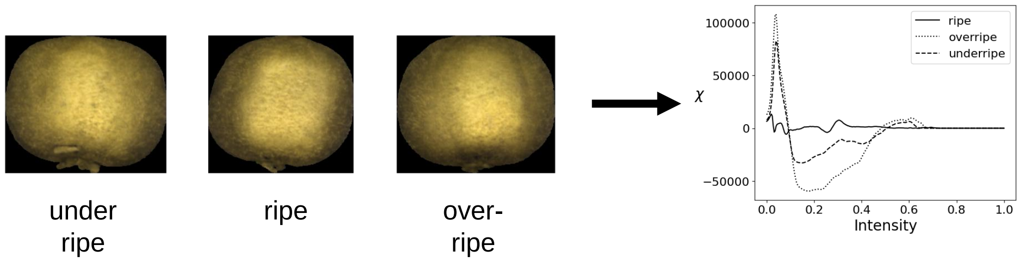
25. Mehta et al. (2018) Single-cell analysis using hyperspectral imaging modalities, J Biomech Eng, 140(2), 020802.
26. Varga et al. (2021) Measuring the Ripeness of Fruit with Hyperspectral Imaging and Deep Learning, arXiv:2104.09808v1

Conclusions

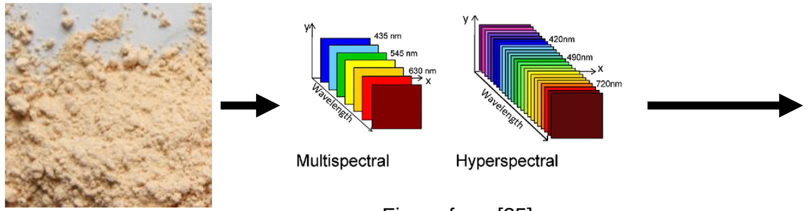
- Topology reveals order and patterns throughout field data



- How we process these patterns drastically impacts computational efficiency
  - Enabling real-time analysis and control



- Extend these methods to other applications
  - Pharmaceuticals
  - Other powder/melt-based processing



Formulation Powder (Pharmaceuticals)

Figure from [25]

Critical Quality Monitoring

Impurity content

Homogeneity of active ingredient in powder

Powder physical and chemical properties

25. Mehta et al. (2018) Single-cell analysis using hyperspectral imaging modalities, J Biomech Eng, 140(2), 020802.
26. Varga et al. (2021) Measuring the Ripeness of Fruit with Hyperspectral Imaging and Deep Learning, arXiv:2104.09808v1

Conclusions

Digital Components

Physical Components

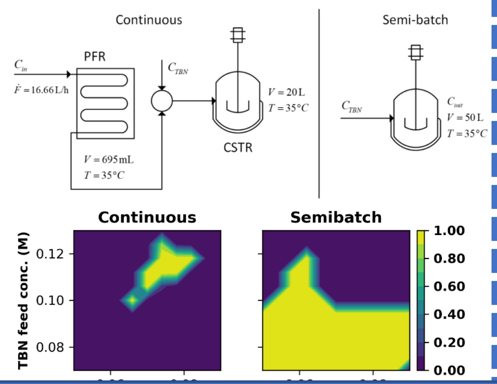
Topic 1

Model Building

Parameter

Topic 2

Process Design



Topic 3

Process Control

1. Flexible software tools enable building of fit-for-purpose digital twins through:
   - **Functionality (e.g., model building, optimal process design)**
   - **Computational tractability (real-time algorithms)**

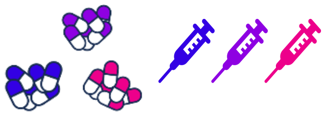
Data

Decisions

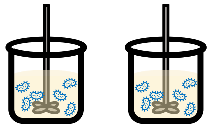
Applications (Physical Systems)

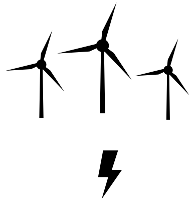
Energetic Materials (DoD)

Pharmaceuticals



Bioprocesses

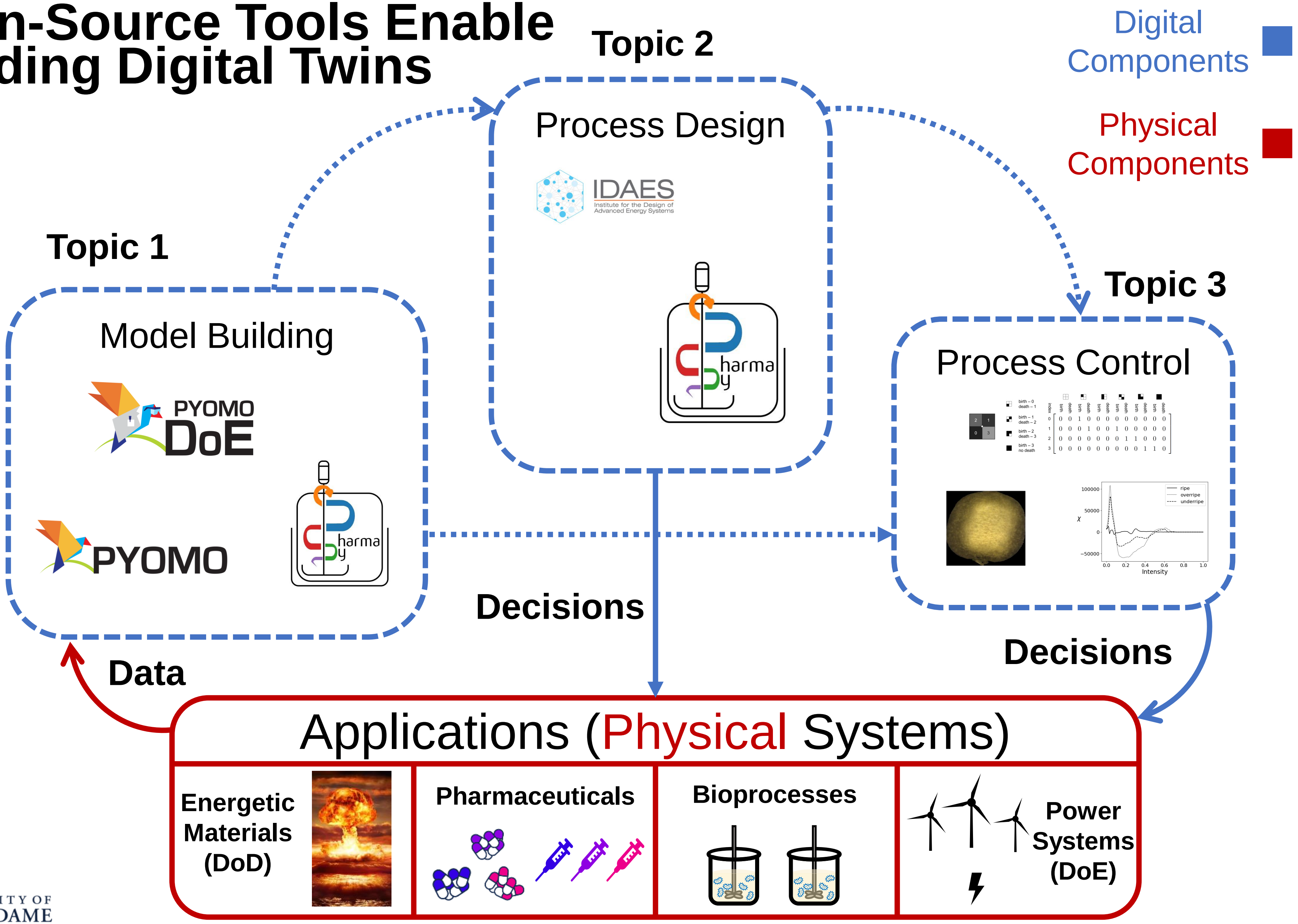




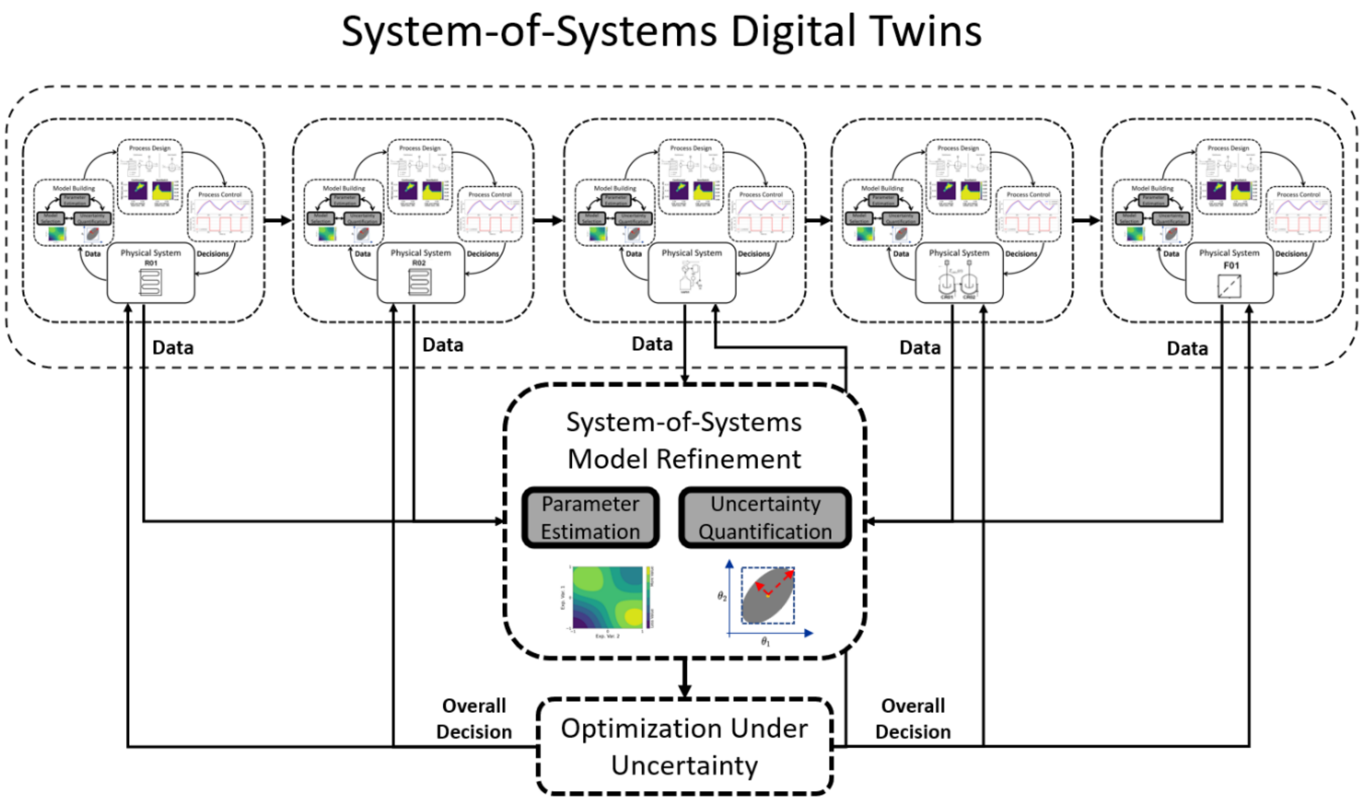
Power Systems (DoE)

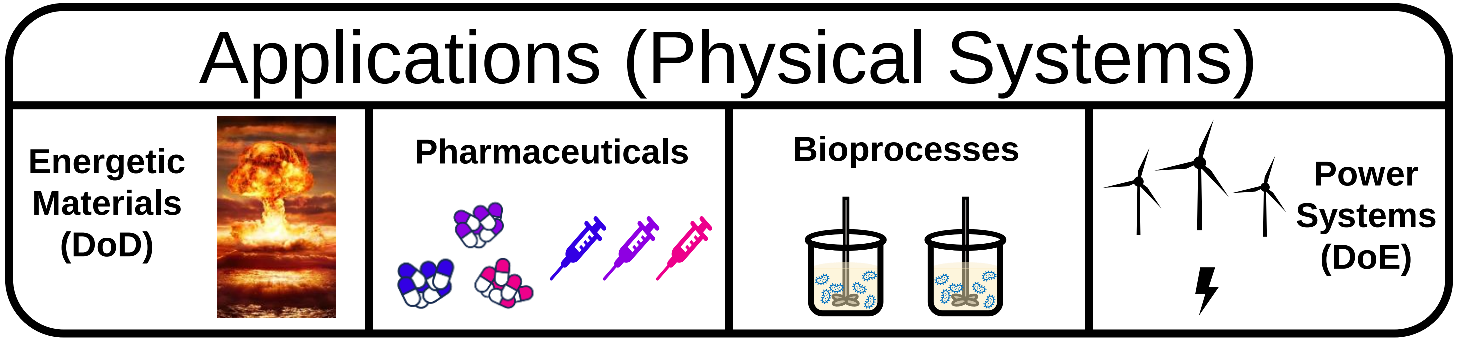
UNIVERSITY OF NOTRE DAME

Open-Source Tools Enable Building Digital Twins



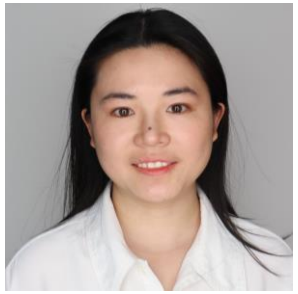
Research Opportunities for Fit-for-Purpose Digital Twins





Pyomo.DoE Acknowledgments

Prof. Alexander Dowling



Dr. Jialu Wang

Hailey Lynch

We graciously acknowledge funding from the U.S. Department of Energy, Office of Fossil Energy and Carbon Management, through the Carbon Capture Program.

This project was funded by the Department of Energy, National Energy Technology Laboratory an agency of the United States Government, through a support contract. Neither the United States Government nor any agency thereof, nor any of their employees, nor the support contractor, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Dowling Lab

Pyomo Team (SNL):

- Bethany Nicholson
- John Siirola
- Katherine Klise
- Shawn Martin




Crystallization (Strathclyde):

- Aaron Bjarnason
- Cameron Brown


CO2 Capture Case Study:

- Ryan Hughes (NETL)
- Debangsu Bhattacharyya (WVU)


UNIVERSITY OF NOTRE DAME

NATIONAL ENERGY TECHNOLOGY LABORATORY

PYOMO

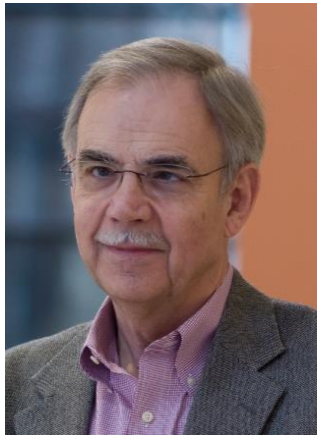
PYOMO DAE

PYOMO DoE

PharmaPy Acknowledgments

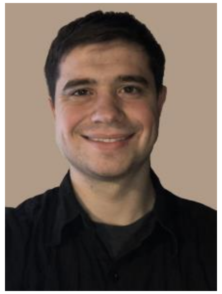
- The authors would like to acknowledge the financial support of the U.S. Food and Drug Administration (FDA), grant U01FD006738

- PharmaPy contributors:
  1. Dr. Daniel Casas-Orozco (Currently at Eli Lilly & Company)
  2. Dr. Varun Sundarkumar (Currently at Takeda)
  3. Inyoung Hur
  4. Dr. Francesco Destro (Currently at MIT)
  5. Jung-Soo Rhim
  6. Yash Barhate
  7. Montgomery Laky
  8. Zachary Hillman



Dr. Gintaras Reklaitis

Dr. Zoltan Nagy



Price Taker/IDAES Acknowledgments

We graciously acknowledge funding from the U.S. Department of Energy, Office of Fossil Energy and Carbon Management, through the Carbon Capture Program.

This project was funded by the Department of Energy, National Energy Technology Laboratory an agency of the United States Government, through a support contract. Neither the United States Government nor any agency thereof, nor any of their employees, nor the support contractor, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Prof. Alexander Dowling

Nicole Cortes

Dowling Lab

NETL Team:

- John C. Eslick
- Alexander A. Noring
- Naresh Susarla
- Chinedu Okoli
- Miguel A. Zamarripa
- Douglas A. Allan
- John H. Brewer
- Arun K.S. Iyengar
- Maojian Wang
- Anthony P. Burgard
- David C. Miller
- Radhakrishna Tumbalam Gooty
- Jaffer Ghouse

NATIONAL ENERGY TECHNOLOGY LABORATORY

IDAES Institute for the Design of Advanced Energy Systems

PYOMO

UNIVERSITY OF NOTRE DAME

Topology Acknowledgments

- Funding from UW-Madison MRSEC *(DMR-2309000)* and NSF BIGDATA (IIS-1837812)

- Thanks to all members of the ZavaLab

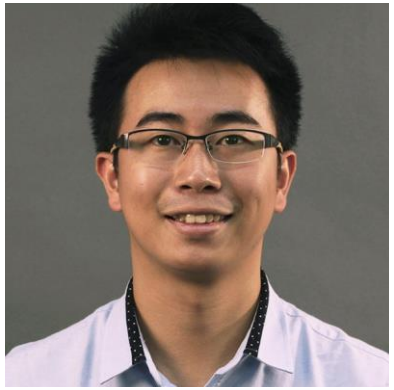
  - Samuel Hower (undergrad)
  - Celeste Mills (undergrad)
  - Leo Gonzalez
  - Shiyi (Amy) Qin
  - Jiaze Ma
  - Bo-Xun Wang
  - David Cole
  - Blake Lopez
  - Lisa Je
  - Jaron Thompson
  - J. Elvis Umaña
  - Ugochukwu Ikegwu
  - Dr. Alexander Smith
  - Dr. Shengli (Bruce) Jiang
  - Dr. Aurora Munguia (postdoc)
  - Dr. Brenda Cansino (postdoc)

Prof. Victor Zavala
(UW-Madison)

Prof. Nick Abbott
(Cornell)

Prof. Reid Van Lehn
(UW-Madison)

Dr. Alex Smith
(postdoc, Minnesota)



Dr. Shengli Jiang
(postdoc, Princeton)