Minitutorial: Multi-Modal Data driven and Physics-informed ML + UQ for Materials Applications

The Hunt for Material Fingerprints

Nat Trask
University of Pennsylvania
Sandia National Laboratories

www.github.com/natrask/PIMA_torch

Talk Overview

Material Fingerprinting

Our wish list for how to get ML working to integrate physics alongside physics-agnostic experimental modalities

Variational Inference 101

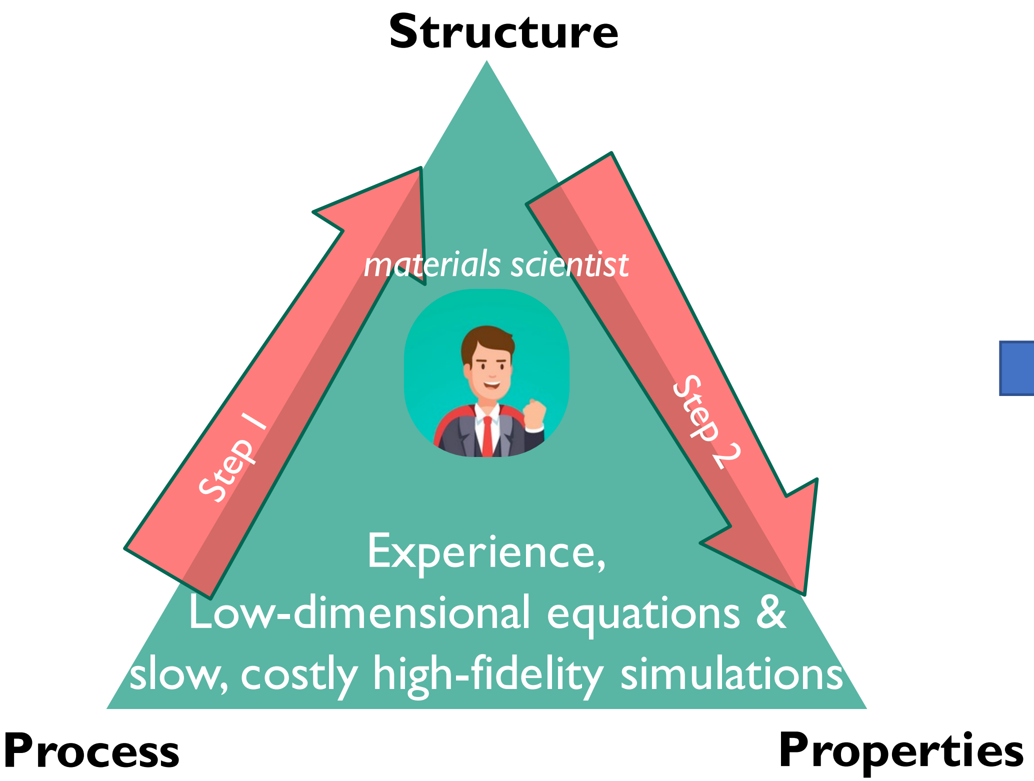
The basics of applying Bayesian reasoning in the context of deep learning models

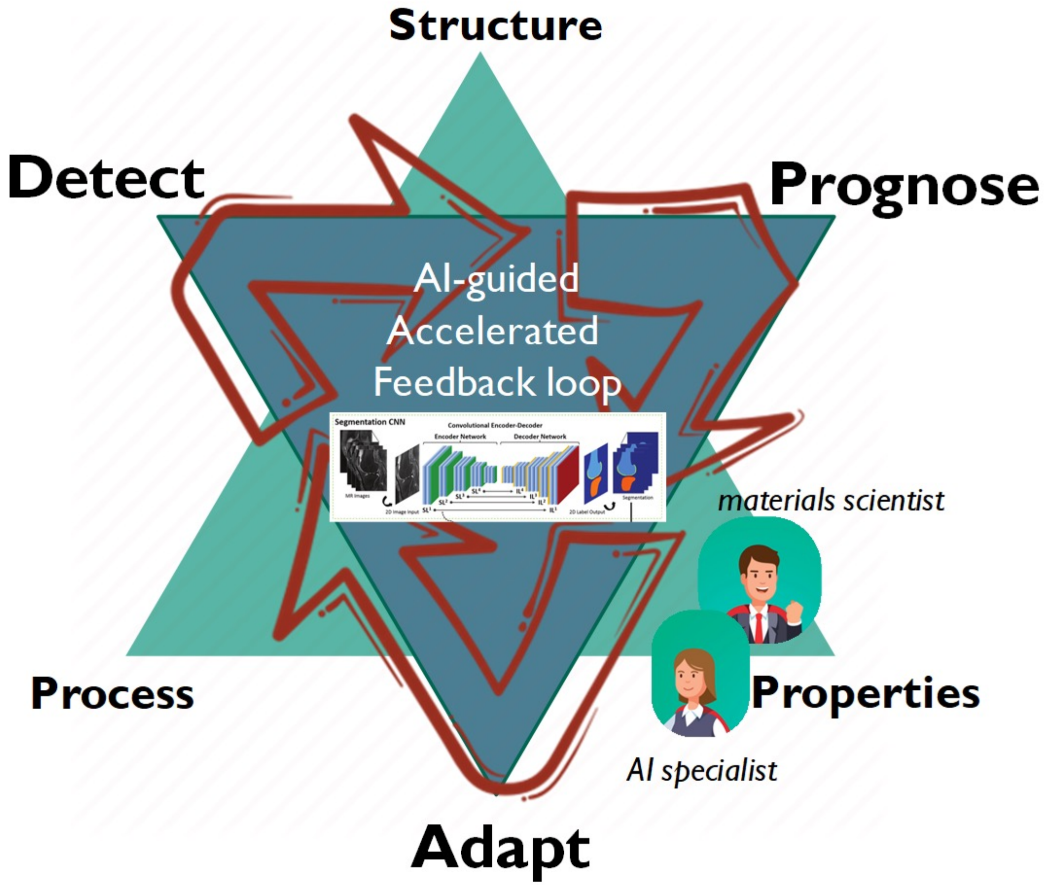
<u>From ELBO to VAE</u>
<u>Clustering with EM</u>
<u>Physics-based expert models</u>

Physics-Informed Multimodal Autoencoders (PIMA)

A scaled-up production framework for high-throughput materials optimization and discovery

Beyond bespoke experiments through **“self-driving” labs**

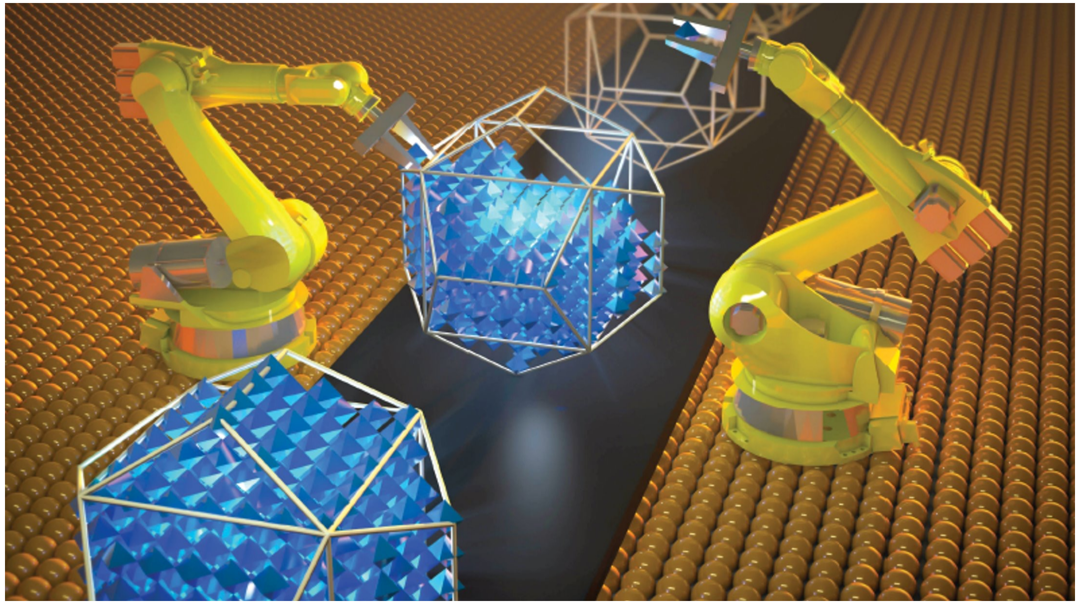


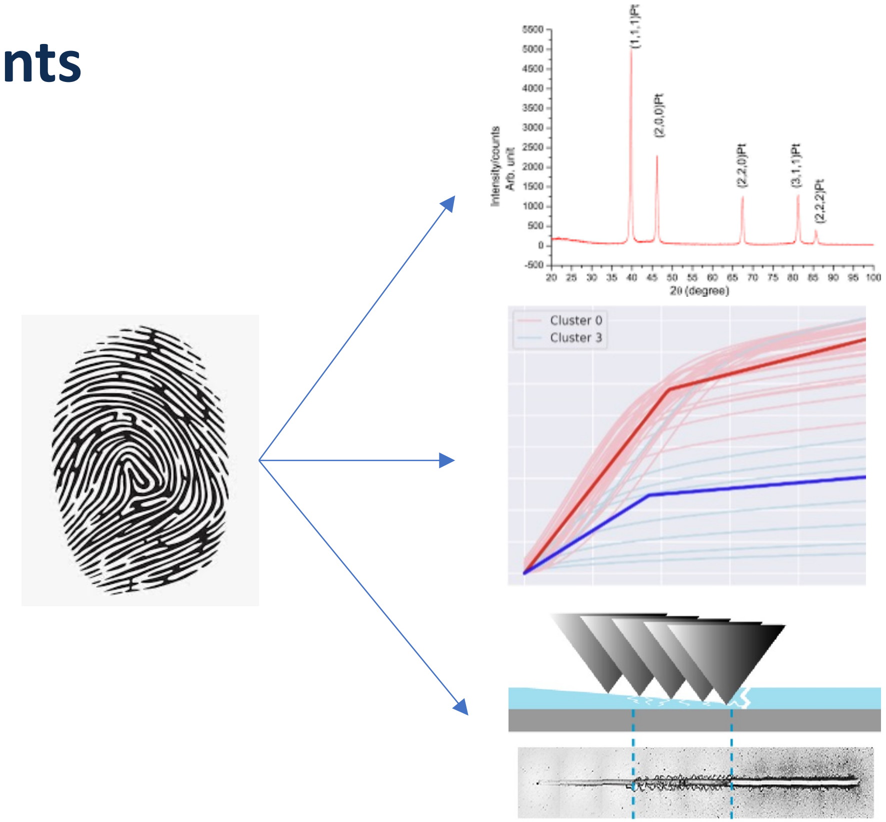


Traditional material science:
Small slow data, **trustworthy**
Bespoke single modality experiments
Campaigns of **1’s-10’s** of experiments

High-throughput self-driving labs:
Large fast data, noisy, **unlabeled**
Combinatorial, **multimodal** experiments
Campaigns of **thousands** of experiments

The hunt for comprehensive material **fingerprints**





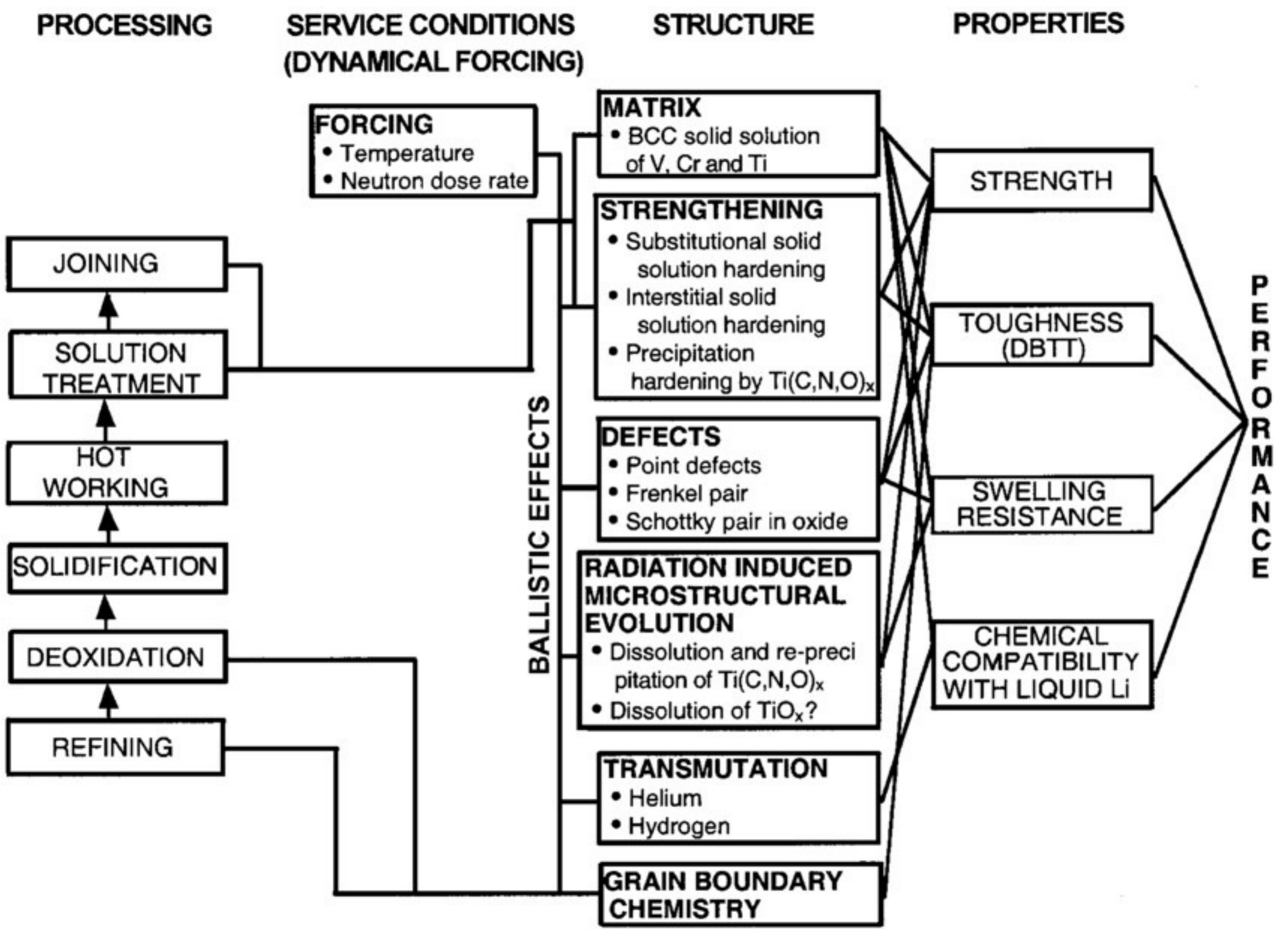
Fingerprint: low dimensional signal correlating with an underlying complex physical process in high dimensional data
State-of-the-Art: PCA, feature engineering, manifold learning

<u>Question:</u> How to move beyond task-agnostic dimension reduction tools to integrate physics?

Chakraborty, A., Nandi, P., and Chakraborty, B. Fingerprints of the quantum space-time in time-dependent quantum mechanics: An emergent geometric phase. *arXiv preprint arXiv:2110.04370*, 2021.

Hasselmann, K. Multi-pattern fingerprint method for detection and attribution of climate change. *Climate dynamics*, 13(9):601–611, 1997.

Toward process-structure-property-maps with **causal** & **mechanistic** learning

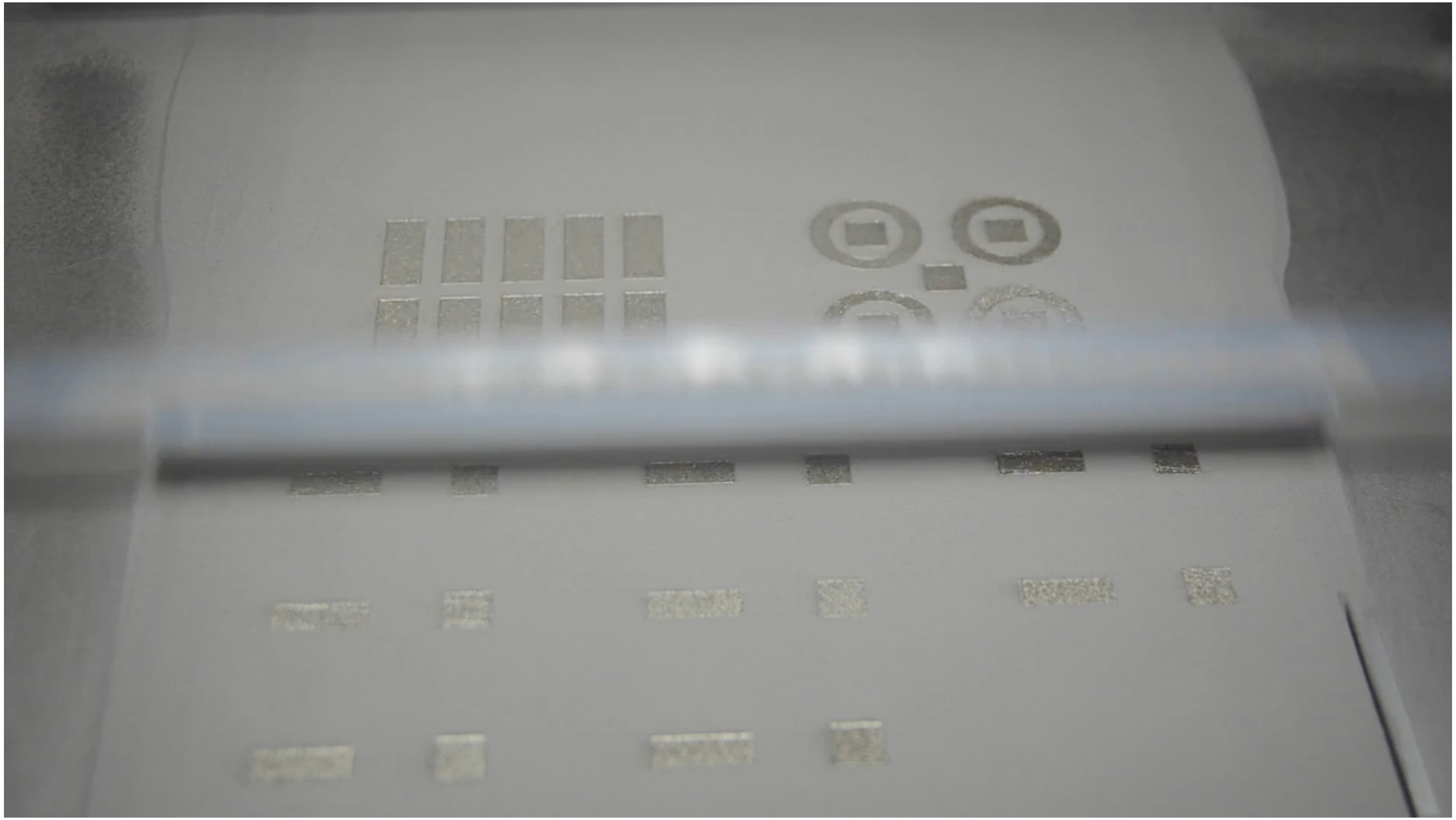


When people say AI (as opposed to ML) often they envision modular reasoning and cause/effect

(e.g. Olson Diagram)

Ghosh. *Journal of Phase Equilibria.* 2002

Unconventional sources of multimodal data

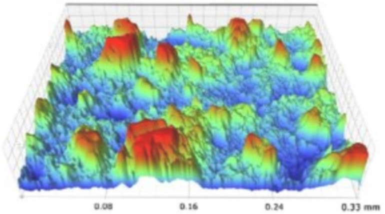


Process parameters: laser power, speed, path, powder composition
High-fidelity characterization: microscopy, XRF, XRD, TEM, SEM
Low-fidelity signals: light, sound, images, profilometry

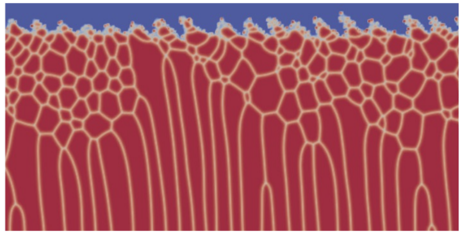
The **avalanche of modalities** available in material science applications

<u>Pre-process</u>

Stage	Composition	Concentration
Electroless nickel plating	$NiSO_4 \cdot 6H_2O$	30 g L^{-1}
	$NaH_2PO_2 \cdot H_2O$	30 g L^{-1}
	$Na_3C_6H_5O_7 \cdot 2H_2O$	20 g L^{-1}
Electroplating	$Ni(SO_3NH_2)_2 \cdot 4H_2O$	300 g L^{-1}
	$NiCl_2 \cdot 6H_2O$	15 g L^{-1}
	H_3BO_3	20 g L^{-1}

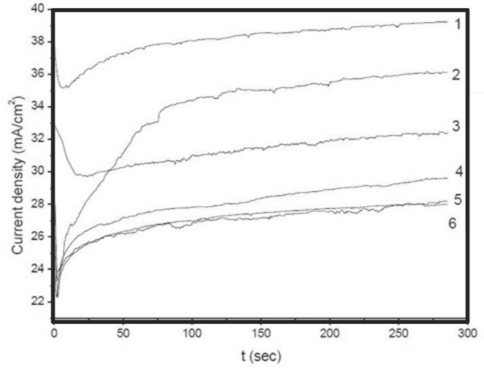


Precursor characterization

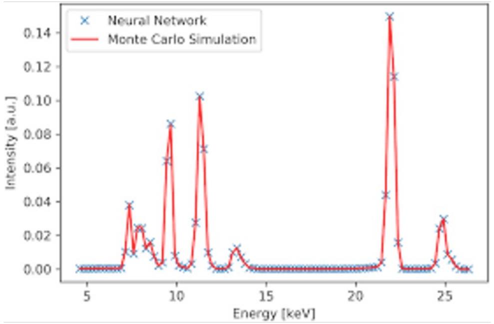


High-fidelity mod/sim

<u>In-process</u>

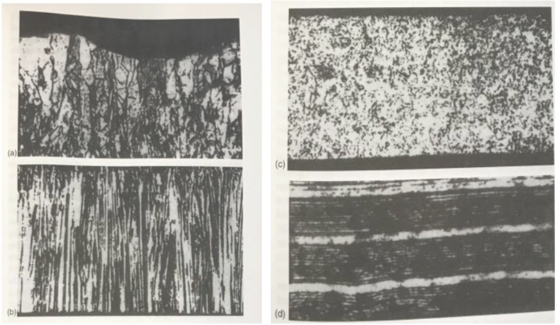


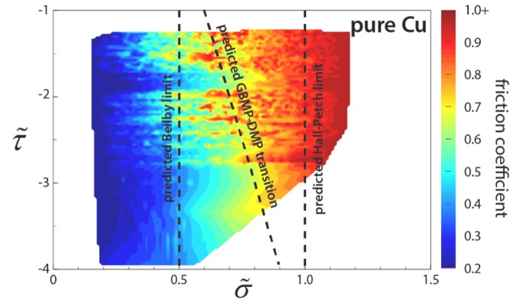
Time-series data



In-situ characterization

<u>Post-process</u>



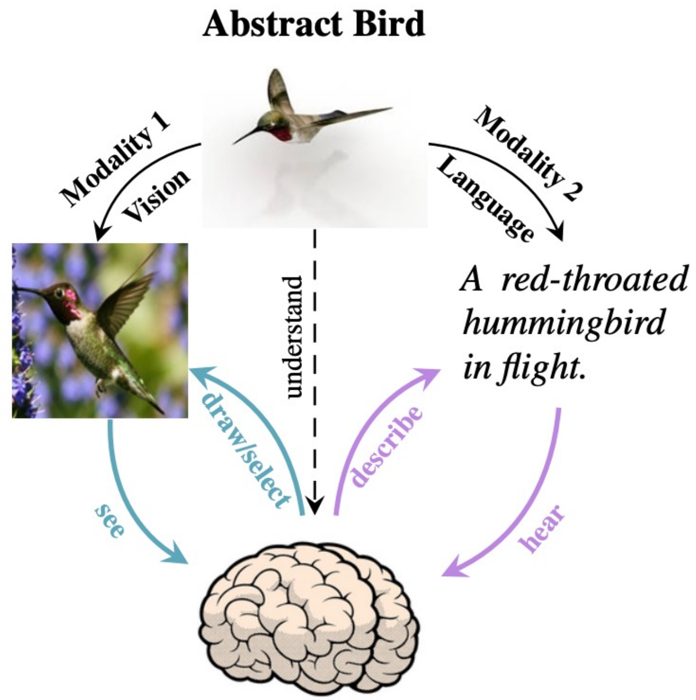


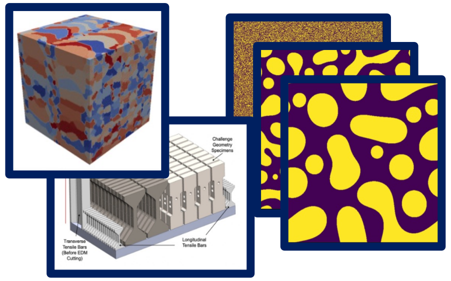
Property/Performance/Aging Measurements

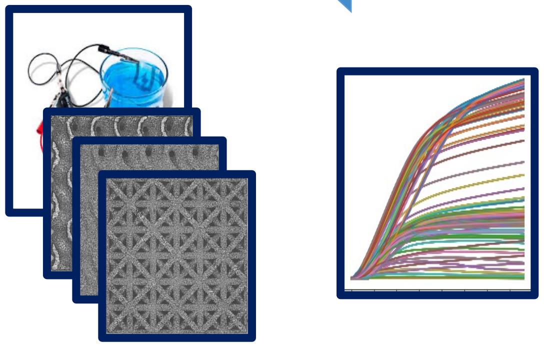
Sifting through a quantity of modalities beyond limits of human cognition

Shi, Y., Siddharth, N., Paige, B. and Torr, P.H., 2019. Variational mixture-of-experts autoencoders for multi-modal deep generative models. *arXiv preprint arXiv:1911.03393.*

What do we mean by **multimodal deep learning**?







Traditional Multimodality

Assimilate heterogeneous data sources so they're greater than sum of their parts

e.g. audio + text to generate automatic subtitles

High-Fidelity Modalities

Classical material characterization

e.g. microscopy (physical), direct simulation (synthetic)

High-Throughput Modalities

Unconventional measurements

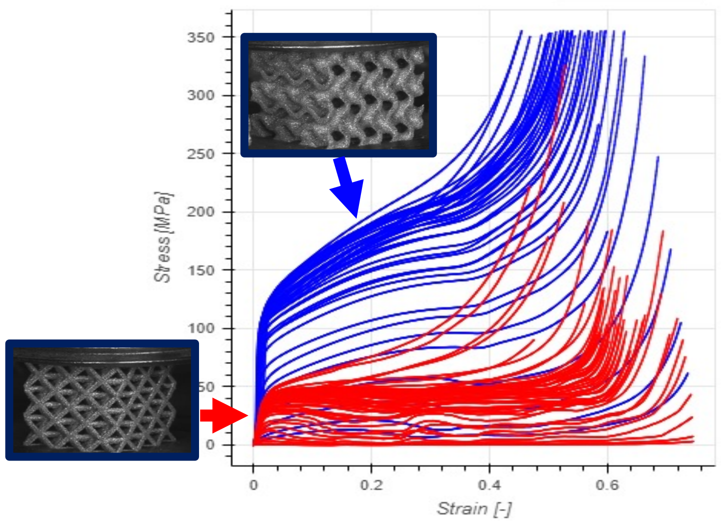
e.g. spectroscopy, electrical signals, sound, photographs

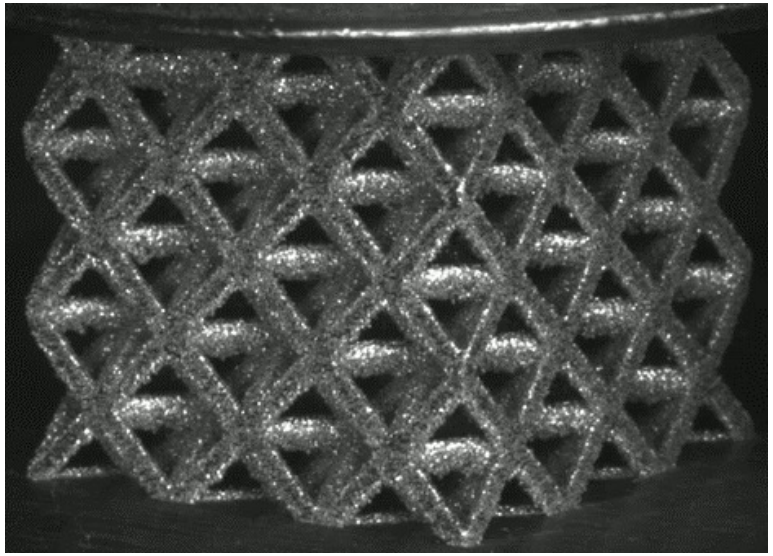
Shi, Yuge, Brooks Paige, and Philip Torr. "Variational mixture-of-experts autoencoders for multi-modal deep generative models." *Advances in neural information processing systems* 32 (2019).

AI4SS + AI@DOE workshops denote multimodal scientific data a top opportunity for scientific impact across DOE in next 5-10 years

Shi, Y., Siddharth, N., Paige, B. and Torr, P.H., 2019. Variational mixture-of-experts autoencoders for multi-modal deep generative models. *arXiv preprint arXiv:1911.03393.*

Building toward: Disentangling latent representations through physics



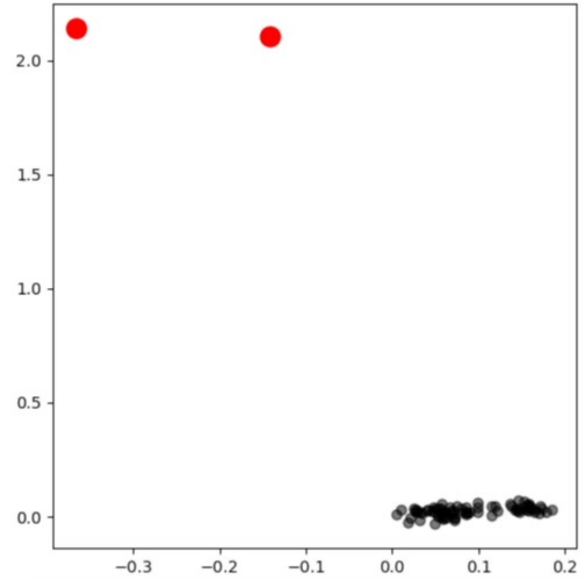


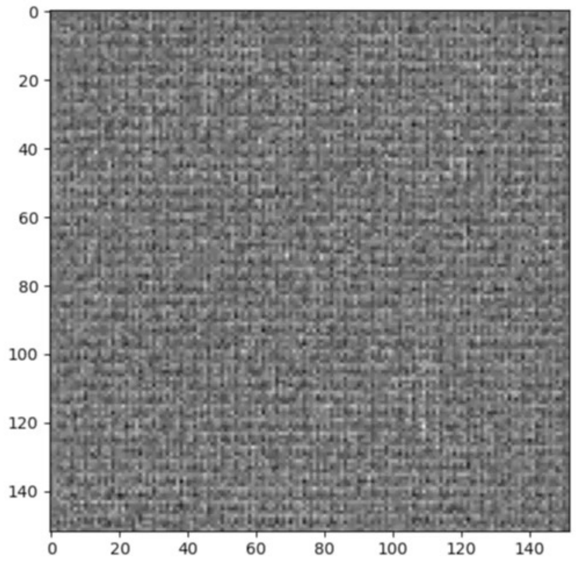
1. Images of lattices and stress/strain curves are disentangled in latent space

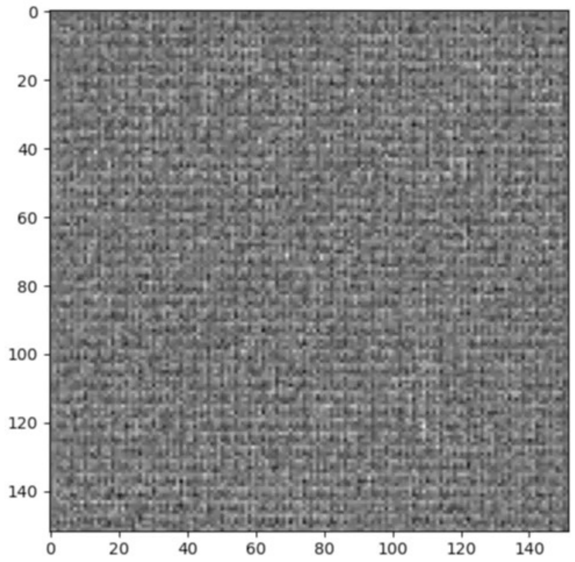
2. Once disentangled, images are calibrated to clusters

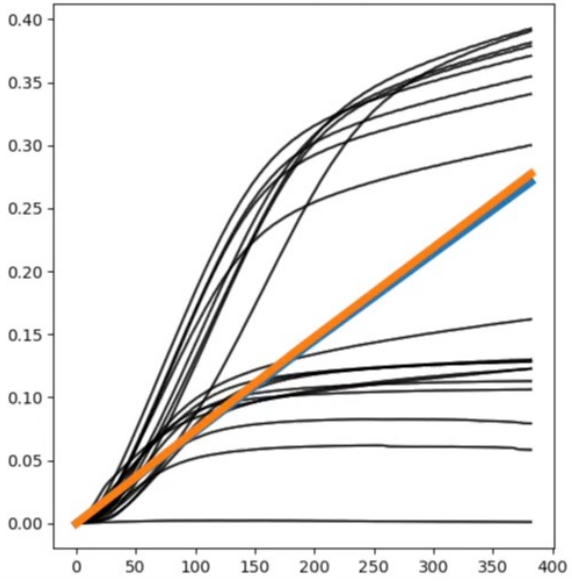
3. Once disentangled, expert models calibrate into two populations

4. Impact: Low-throughput uniaxial testing replaced w/ high-throughput imaging









Talk Overview

Material Fingerprinting

Our wish list for how to get ML working to integrate physics alongside physics-agnostic experimental modalities

Variational Inference 101

The basics of applying Bayesian reasoning in the context of deep learning models

<u>From ELBO to VAE</u>
<u>Clustering with EM</u>
<u>Physics-based expert models</u>

Physics-Informed Multimodal Autoencoders (PIMA)

A scaled-up production framework for high-throughput materials optimization and discovery

Tutorial 1: maximum likelihood estimates for supervised learning

$$\mathcal{L} = \log p(\mathbf{y};\theta) = \log \prod_d p(y_d;\theta)$$

Define loss function as assumed IID joint distribution

$$= \sum_d \log p(y_d;\theta)$$

Select distribution for marginal distributions

$$= \sum_d \log \mathcal{N}(y_d;\mu,\Sigma)$$

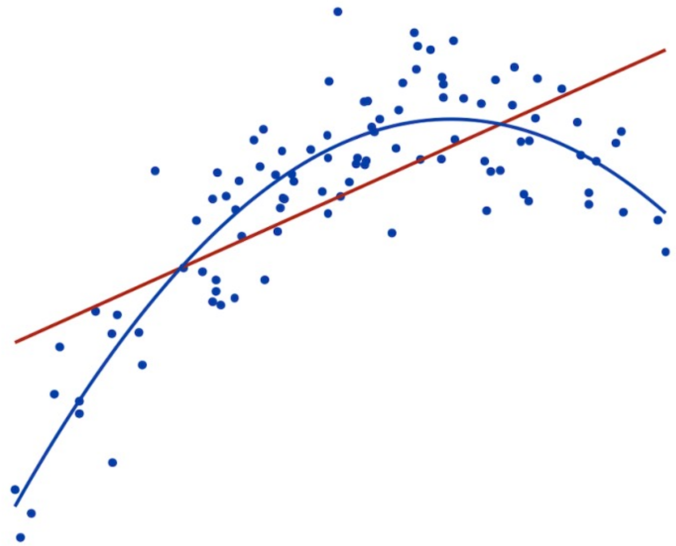
Solve for distribution parameters which maximize (log) likelihood of your dataset

Exercise 1: Recast polynomial fitting as an MLE problem

$$\Theta = \left\{ \mu(x) = \sum_\alpha c_\alpha \phi_\alpha(x), \quad \sigma^2 \right\}$$

$$p(y_d|\Theta) = \mathcal{N}\left(\mu(x_d), \sigma\right)$$

Show: $\mathcal{L} = C + N \log|\sigma| + \sum_{d=1}^{N} \left(\frac{y_d - \sum_\alpha c_\alpha \phi_\alpha(x_d)}{\sigma} \right)^2$



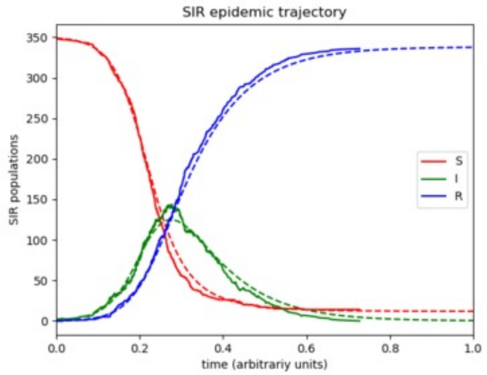
MLE is intractable for **hidden/latent** data

Given:
Hidden variables - Z
Observable variables - X

$$\mathcal{L} = -\sum_d \log p(X_d, Z_d)$$

Can't plug data in for Z!

$$\begin{cases} \frac{dS}{dt} = -\beta IS, \\ \frac{dI}{dt} = \beta IS - \gamma I, \\ \frac{dR}{dt} = \gamma I, \end{cases}$$



where S is the stock of susceptible population, I is the stock of infected, R is the stock of removed population (either by death or recovery), and N is the sum of these three.

Example: For compartment models during COVID, no direct measurements of I were available

Can we estimate the posterior distribution:

$$\underset{\text{Posterior Distribution}}{p(Z|X)} = \frac{p(X|Z)p(Z)}{p(X)} = \frac{p(X|Z)p(Z)}{\underbrace{\int p(X|Z)p(Z)dZ}_{\text{Computationally intractable integral}}}$$

Option 1: (Sampling)
Use data intensive sampling to estimate denominator – can't afford!

Option 2: (Variational Inference)
Solve an optimization problem to infer a "close as possible" posterior distribution that's "nice" to work with

Variational inference 101: First, replace log likelihood with something that doesn't depend on z

$$\mathcal{L} = -\sum_d \log p(y_d|\Theta)$$

Start with what we know

$$= -\sum_d \log \sum_{z_d} p(y_d, z_d|\Theta)$$

Marginalization / LoTP

$$= -\sum_d \log \sum_{z_d} q(z_d)\frac{p(y_d, z_d|\Theta)}{q(z_d)}$$

Introduce an arbitrary distribution q and multiply by 1

$$= -\sum_d \log \mathbb{E}_q \left[\frac{p(y_d, z_d|\Theta)}{q(z_d)}\right]$$

Reinterpret inner sum as an expectation with respect to the measure q

$$\leq -\sum_d \mathbb{E}_q \left[\log \frac{p(y_d, z_d|\Theta)}{q(z_d)}\right]$$

Apply Jensen's inequality

Evidence Lower BOund (ELBO)
Lower bound on likelihood considering all possible q's

$$= -\sum_d \mathbb{E}_q \left[\log p(y_d, z_d|\Theta)\right] - H(q_d)$$

Entropy of q

How do we make the ELBO as tight as possible of a lower bound on the LL?

Evidence Lower Bound (ELBo)
Lower bound on likelihood considering all possible q's

$$\epsilon(\Theta, q_d|y_d) = \sum_d \mathbb{E}_q\left[\log \frac{p(y_d, z_d|\Theta)}{q_d}\right]$$

$$= \sum_d \mathbb{E}_q\left[\log \frac{p(z_d|y_d, \Theta)p(y_d|\Theta)}{q_d}\right]$$

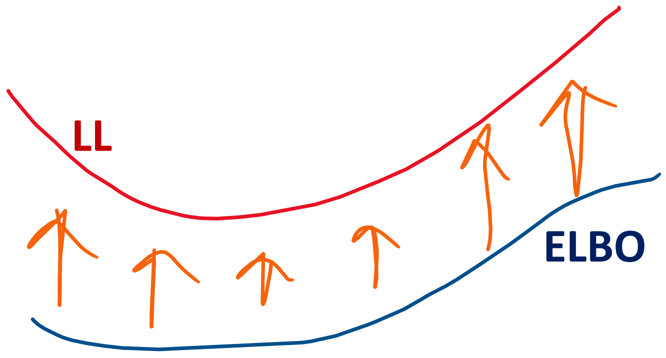
Law of total probability to separate marginal on y

$$= \sum_d \mathbb{E}_q\left[\log \frac{p(z_d|y_d, \Theta)}{q_d}\right] + \mathbb{E}_q\left[\log p(y_d|\Theta)\right]$$

Linear separability of logs of products give 2 terms

$$= \sum_d -\mathbb{KL}\left(p(z_d|y_d, \Theta)||q_d\right) + \mathbb{E}_q\left[\log p(y_d|\Theta)\right]$$

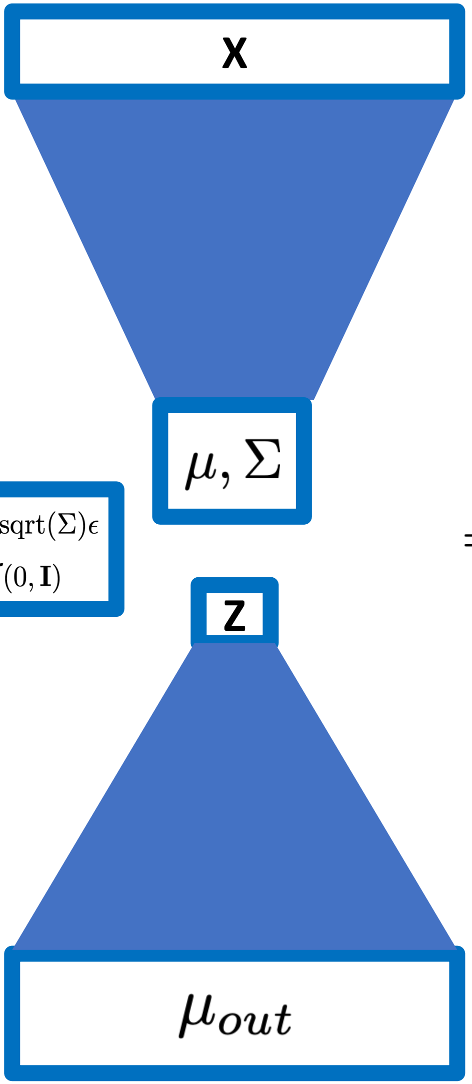
Get a Kullback-Leibler divergence (a quasi-metric on distributions)



Choose the q that makes ELBO as tight as possible
Since KL is non-neg, ELBO is biggest when

$$q_d = p(z_d|y_d, \Theta)$$

Now that we have an alternative loss, how can we use it?

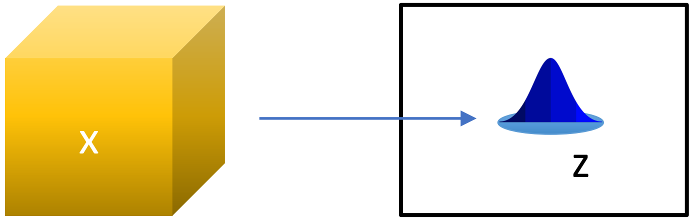


$$Z = \mu + \mathrm{sqrt}(\Sigma)\epsilon$$
$$\epsilon \sim \mathcal{N}(0, \mathbf{I})$$

$$p(Z) \sim \mathcal{N}(0, \mathbf{I})$$
$$q(Z|X) \sim \mathcal{N}(\mu, \Sigma)$$
$$p(X|Z) \sim \mathcal{N}(\mu_{out}, \mathbf{I})$$

$$\mathrm{ELBO}(q) = \sum_Z q(Z) \log \frac{p(Z, X)}{q(Z)}$$

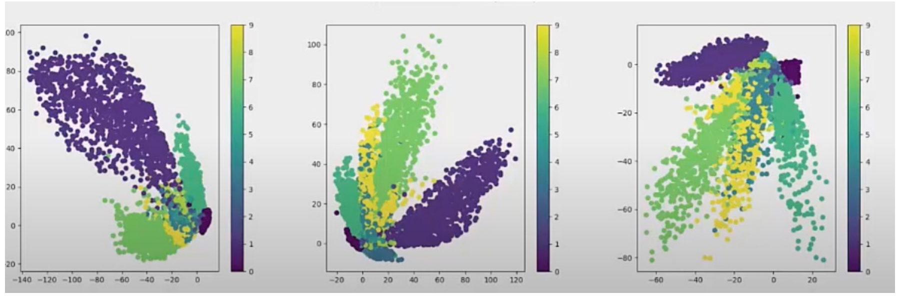
$$= \sum_Z \underbrace{\mathbb{E}_q[\log p(X|Z)]}_{\text{Reconstruction Error (MSE)}} - \underbrace{KL\left(q(Z|X) \,||\, p(Z)\right)}_{\text{Prior/Posterior Mismatch Penalty}}$$



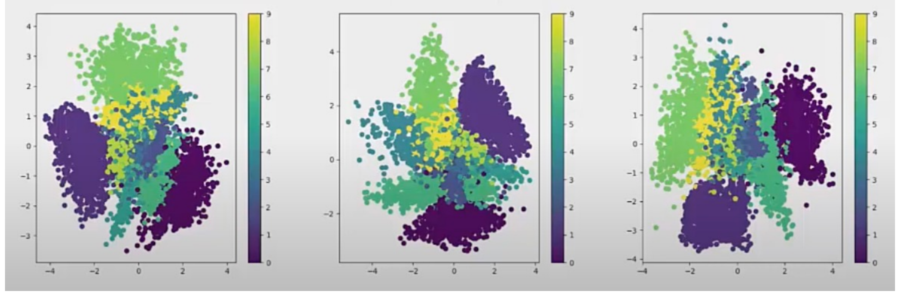
High-dim Data Space

Low-dim Latent Space

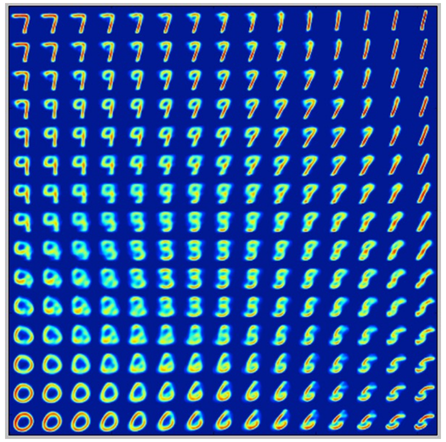
Vanilla Autoencoder



Variational Autoencoder (VAE)

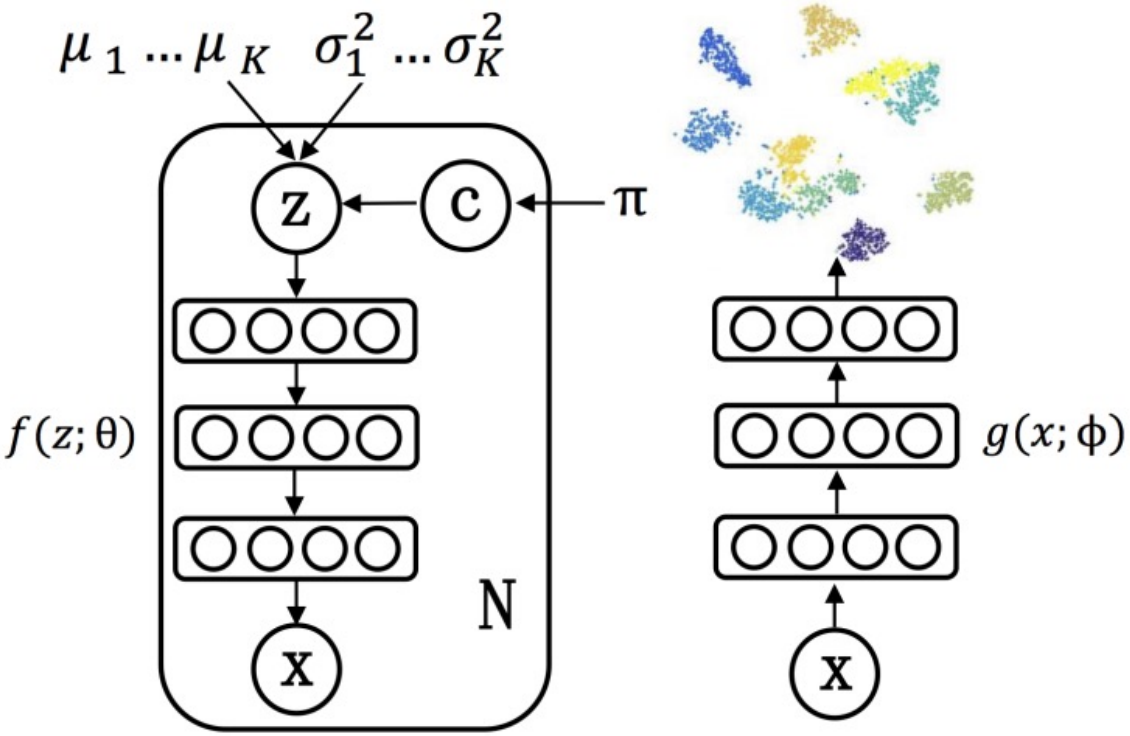


Choose prior to allow generative modeling



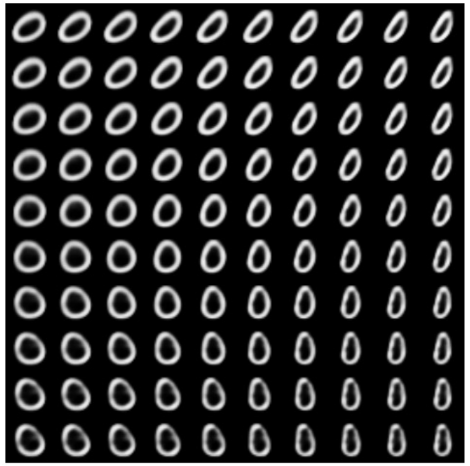
[Link to colab notebook]

There is a lot more we can do with the prior than make a plain VAE though



To obtain manipulable clusters, take GMM prior:

$$p(Z) = \sum_c \Pi_c \mathcal{N}(Z; \mu_c, \Sigma_c)$$



Explicitly access *disentangled* clusters describing semantically distinct clusters in dataset

Can we steer unsupervised clustering algorithms that are biased toward *physicically-biased* clusters encoding material fingerprints?

Jiang, Zhuxi, et al. "Variational deep embedding: An unsupervised and generative approach to clustering." *arXiv preprint arXiv:1611.05148* (2016).

Dilokthanakul, Nat, et al. "Deep unsupervised clustering with gaussian mixture variational autoencoders." *arXiv preprint arXiv:1611.02648* (2016).

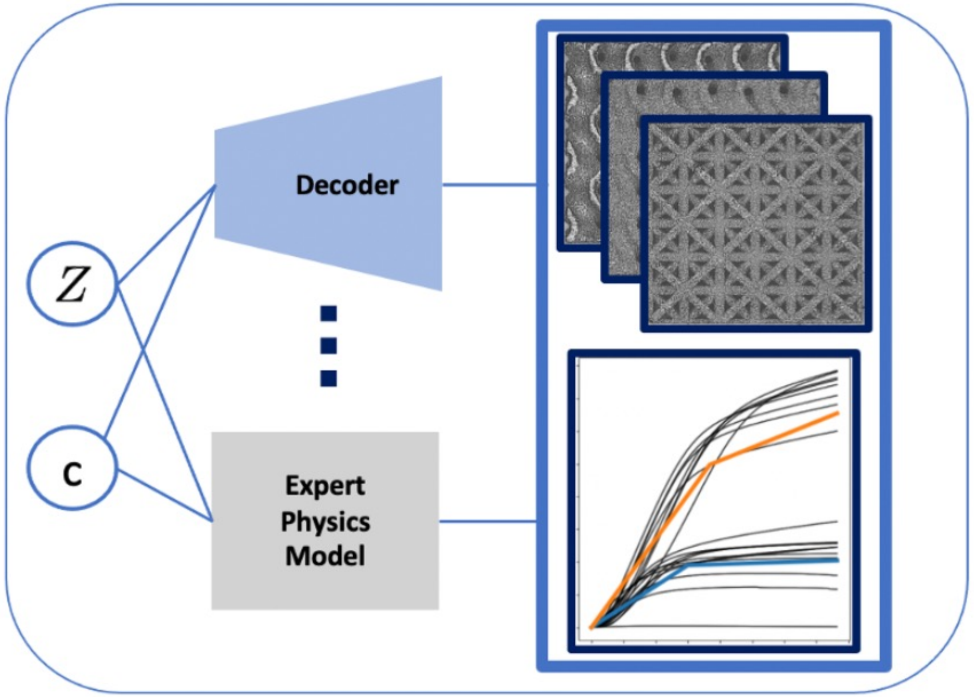
Tutorial 2: Mixture of expert models conditioned on Gaussian mixtures

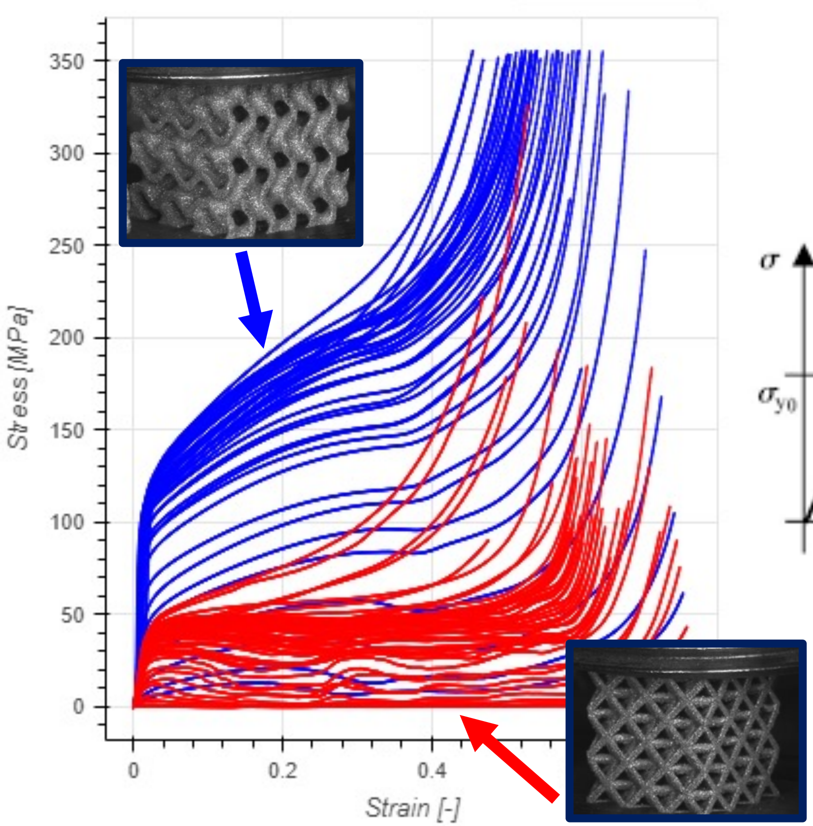
$$z \sim GMM\left(\pi, \left\{\mathcal{N}(\mu_c, \sigma_c^2)\right\}_{c=1}^{N}\right) \quad c \sim cat(\pi) \quad y|c,t \sim \mathcal{N}(\mu_c = f(t;\Theta_c), \sigma_c^2)$$

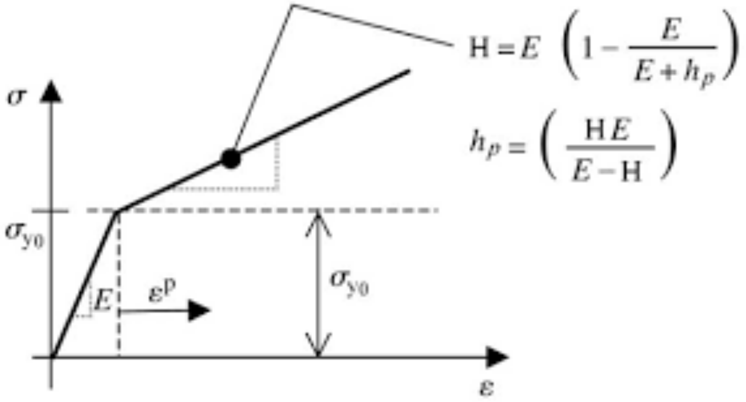
Prior Distribution on Latent Space

Latent Cluster Probability

Cluster-specific Physics Model
Each cluster can get its own custom tailored physics model



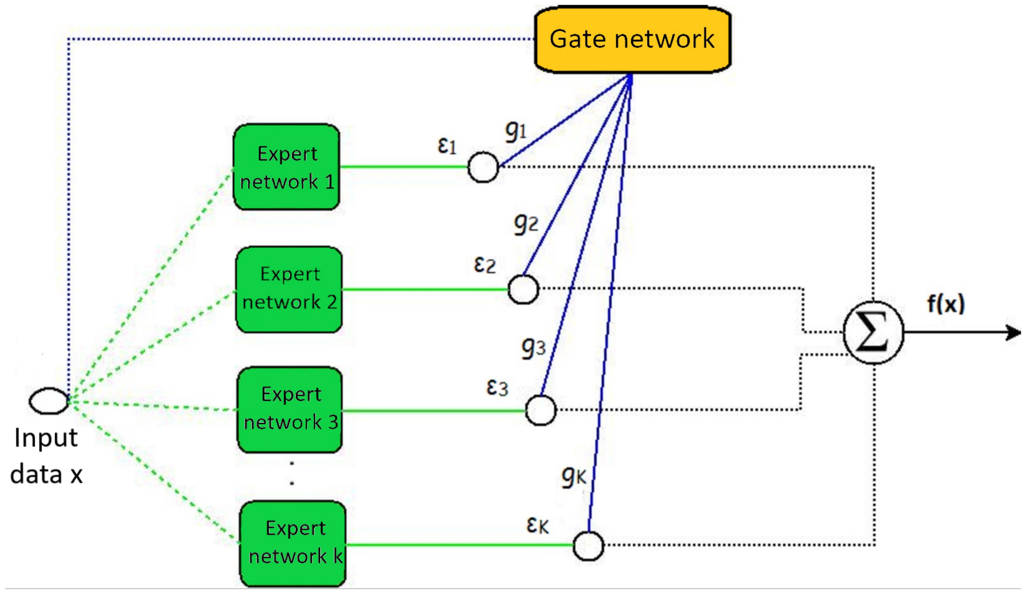




Tutorial 2: Mixture of expert models conditioned on Gaussian mixtures (simplified)

$$c \sim cat(\pi) \quad y|c,t \sim \mathcal{N}(\mu_c = f(t;\Theta_c), \sigma_c^2) \quad p(y;\pi,\Theta) = \sum_c \pi_c \mathcal{N}(f(t;\Theta_c), \sigma_c^2)$$

Latent Cluster Probability — **Cluster-specific Physics Model** — **Mixture of Experts Model**



Architecture for a contemporary MOE model

Mixture of Experts (MoE)
Blending expert models is an old idea getting a resurgence recently, as a powerful tool to blend smaller models and access very, very large models [e.g. 100B+ param (*Shazeer 2018)]*

As an exercise, we use VI to fix constant gates
In the process we will learn *expectation maximization (EM)*, a nice technique for problems with latent variables

Jordan, Michael I., and Robert A. Jacobs. "Hierarchical mixtures of experts and the EM algorithm." *Neural computation* 6.2 (1994): 181-214.
Peralta, Billy, Ariel Saavedra, Luis Caro, and Alvaro Soto. "Mixture of experts with entropic regularization for data classification." *Entropy* 21, no. 2 (2019): 190.

Exercise: Derive expectation-maximization algorithm for MoE model

E-step
Calculate q = p(c|y)

M-step
Optimize ELBO holding q from M-step constant

$$q = p(c|y) = \frac{p(c)p(y|c)}{\sum_{c'} p(c')p(y|c')}$$

Bayes rule

$$= \frac{\pi_c \mathcal{N}(f(t;\theta_c), \sigma_c^2)}{\sum_{c'} \mathcal{N}(f(t;\theta_{c'}), \sigma_{c'}^2)}$$

Estimate q

E-step
Using the current guesses for the model, we compute the posterior distribution q to identify which model is most likely to have generated the data

[Sample code and blog for EM algorithm]

Murphy, Kevin P. *Probabilistic machine learning: an introduction*. MIT press, 2022.

Exercise: Derive expectation-maximization algorithm for MoE model

E-step
Calculate q = p(c|y)

M-step
Optimize ELBO holding q from M-step constant

$$\min \sum_d \mathbb{E}_q \left[\log p(y_d, z_d|\Theta)\right]$$

Expand definition of ELBO

$$= \sum_{d,c} q_{dc} \left[\log p(y_d|z_d, \Theta_c) + \log p(z_d)\right]$$

LoTP

$$= \sum_{d,c} q_{dc} \left[\log \mathcal{N}(f(t;\Theta_c), \sigma_c^2) + \log \pi_c\right]$$

Defs from MoE

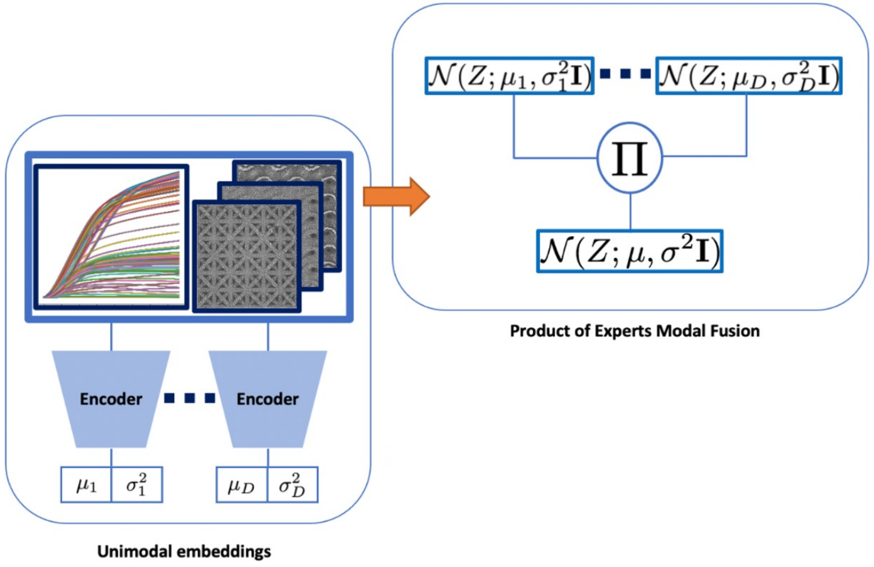
M-step
Calibration of **all** random variables follows from either explicit formulas or weighted L2 residuals
Note this for the remainder of the talks in this mini

$$\frac{\partial}{\partial \pi_c} = 0 \rightarrow \pi_c = \frac{\sum_d q_{dc}}{N}$$

$$\frac{\partial}{\partial \Theta_c} = 0 \rightarrow \theta_c = \text{argmin} \sum_d q_{dc} \left(y_d - f(t;\Theta_c)\right)^2$$

[Sample code and blog for EM algorithm]

Final Exercise: Enriching VAE with multimodal product of experts (PoE)



Key Idea

Use standard VAE encodings, and exploit the fact that product of Gaussian PDFs yields another Gaussian PDF

Derive through Bayes + IID assumption

$$q(Z|X_1,...,X_D) = q(Z)^{1-D}\Pi_{i=1}^{D} q(Z|X_i),$$

Uninformative prior

$$q(Z|X_1,...,X_D) \propto \Pi_{i=1}^{D} q(Z|X_i)$$

Closed form posterior

$$q(Z|X_1,...,X_D) = \mathcal{N}(\mu,\sigma^2\mathbf{I}),$$

$$\sigma^{-2} = \sum_{i=1}^{D} \sigma_i^{-2}, \qquad \frac{\mu}{\sigma^2} = \sum_{i=1}^{D} \frac{\mu_i}{\sigma_i^2},$$

Wu, M., & Goodman, N. (2018). Multimodal generative models for scalable weakly-supervised learning. *Advances in neural information processing systems*, *31*.

Talk Overview

Material Fingerprinting

Our wish list for how to get ML working to integrate physics alongside physics-agnostic experimental modalities

Variational Inference 101

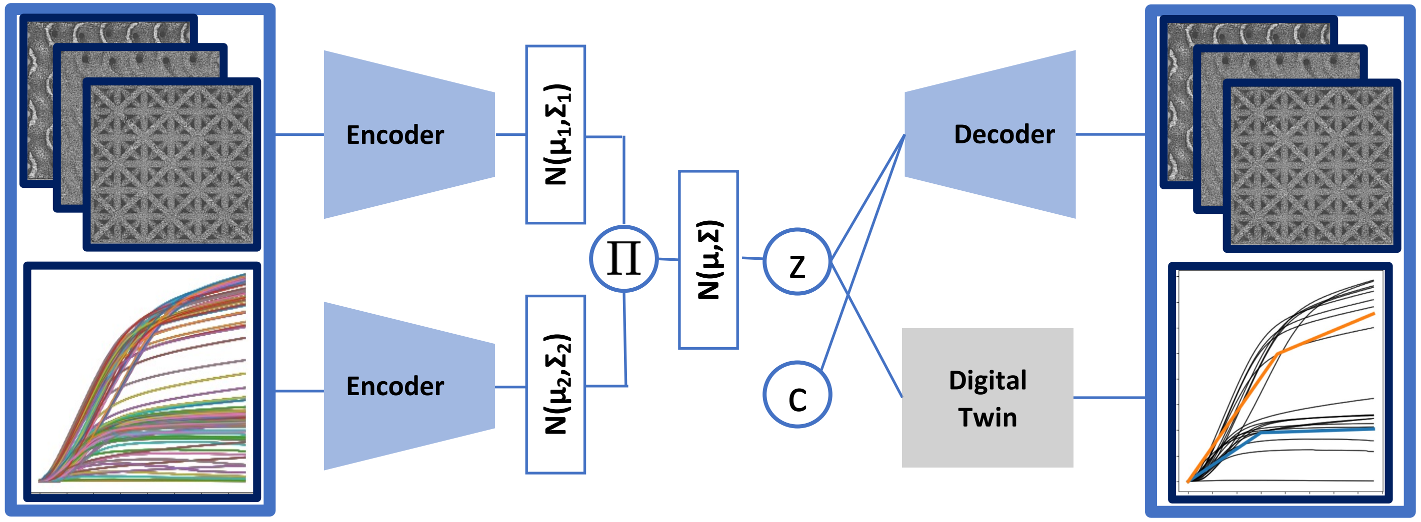
The basics of applying Bayesian reasoning in the context of deep learning models

From ELBO to VAE
Clustering with EM
Physics-based expert models

Physics-Informed Multimodal Autoencoders (PIMA)

A scaled-up production framework for high-throughput materials optimization and discovery

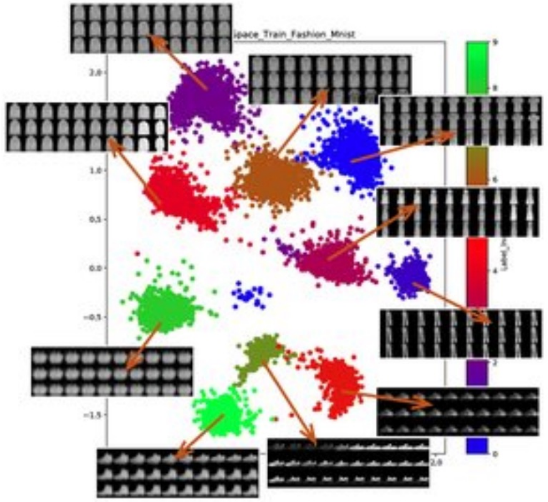
The complete framework: Physics-informed multimodal autoencoders (PIMA)



Walker, Elise, et al. "Unsupervised physics-informed disentanglement of multimodal data." *Foundations of Data Science* (2024): 0-0.

Trask, Nathaniel Albert, Carianne Martinez, and Brad Boyce. "Physics-informed multimodal autoencoder." U.S. Patent Application No. 17/743,160.

Informal Idea:
We discover a shared latent representation of data providing a Rosetta stone across modalities <u>w/ uncertainty estimation</u>

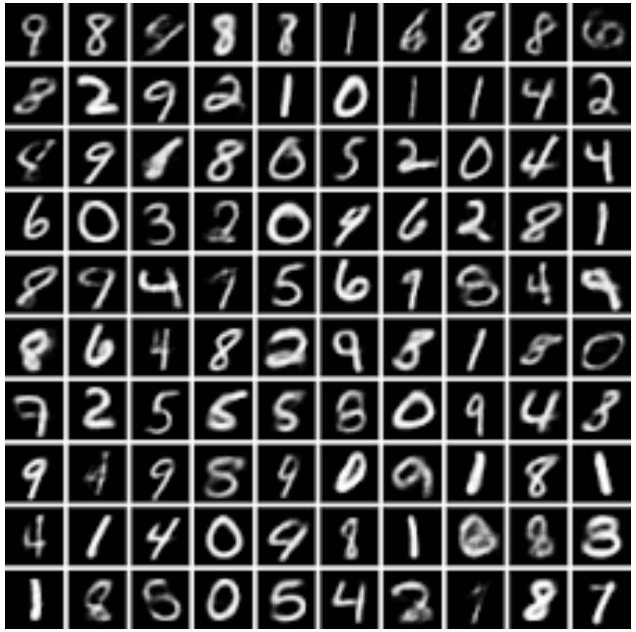


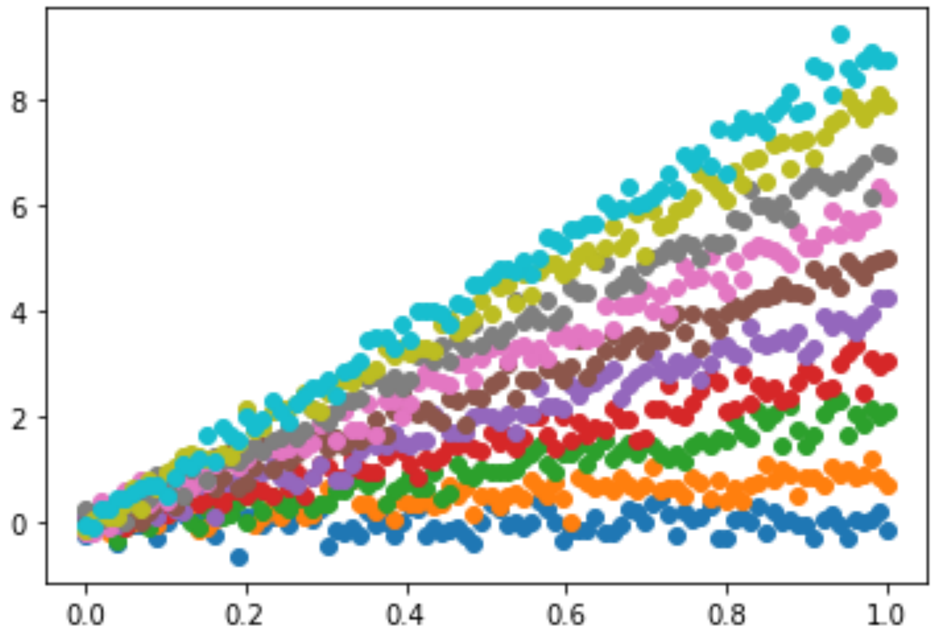
Formal Idea:

- **Gaussian product distribution gives deep posterior embedding for each modality**
- **Gaussian mixture prior in latent space identifies populations in data across modalities**
- ***Closed form expressions* for loss – no Monte Carlo**
- **Supports Bayesian inference across modalities**

Toy code: calibrating a simple affine expert model

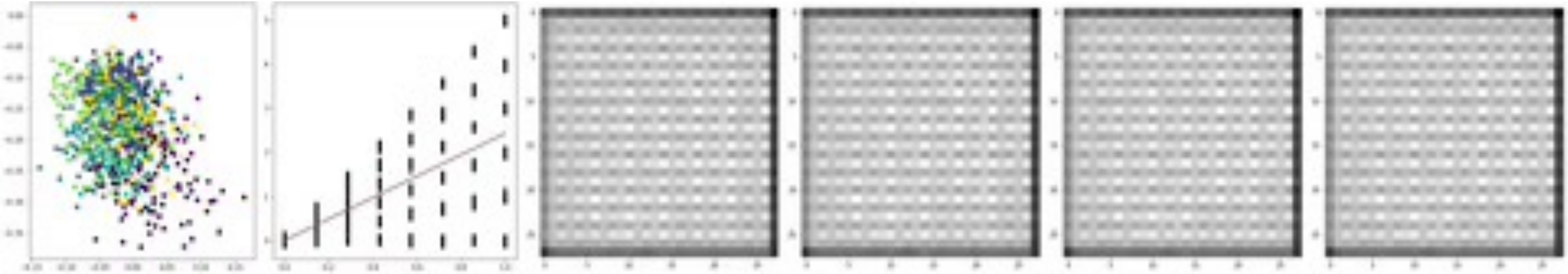
[Link to TF Code]



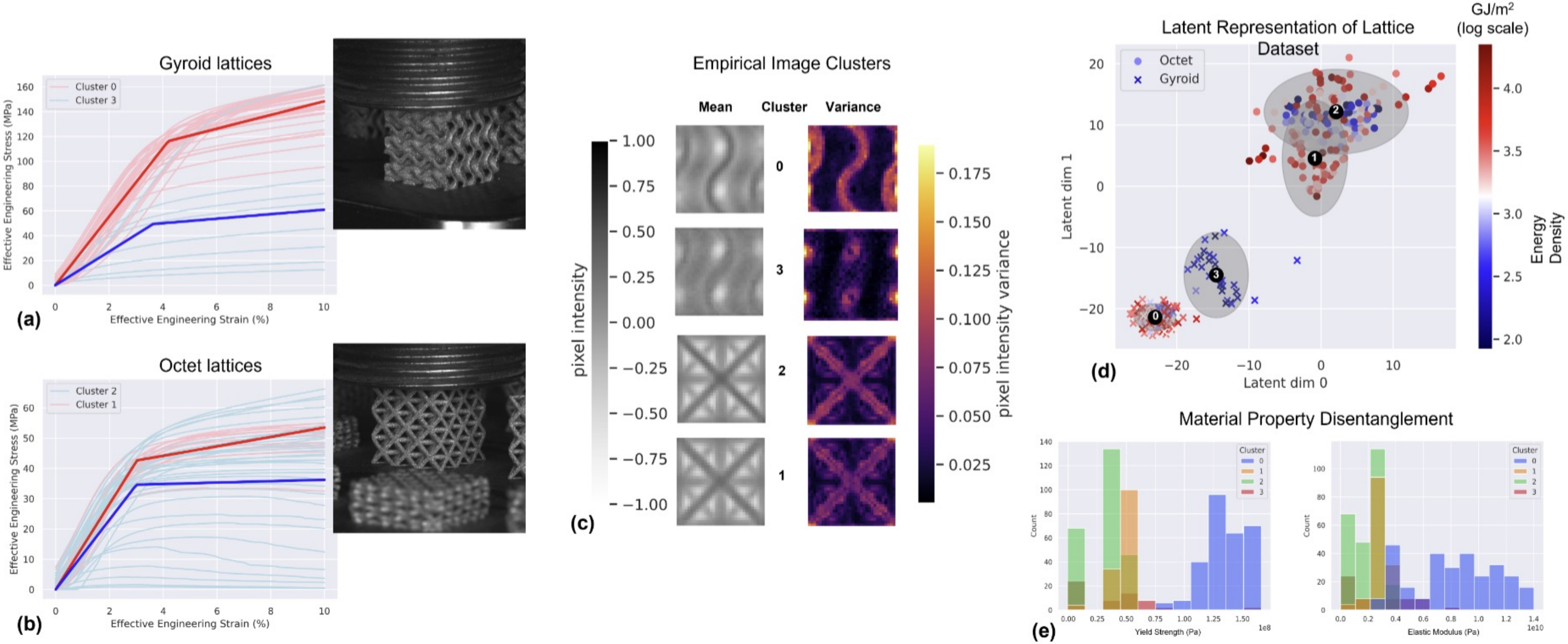


X_1 - Image of Z

X_2 = Z*x + noise

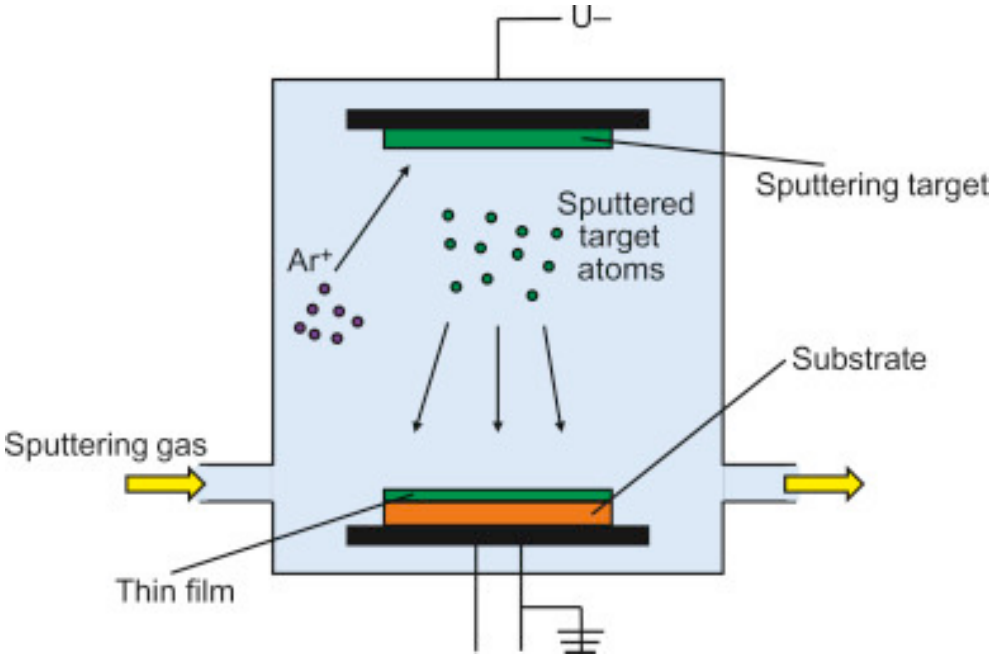


Application 1: Laser powderbed fusion (revisited)

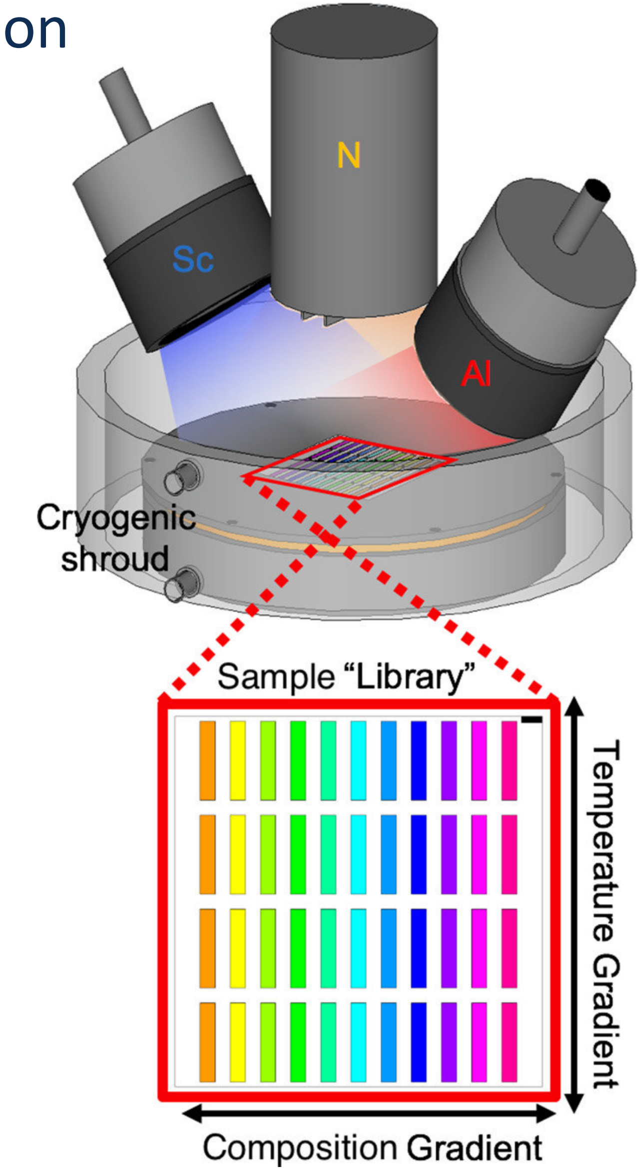


Physical model reveals qualitative regimes of performance, other modalities facilitate hypothesis generation for explainability

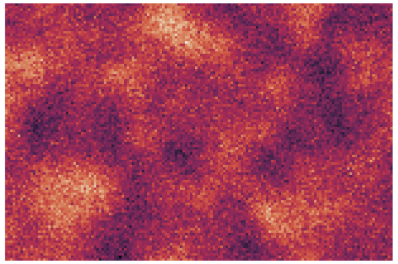
Application 2: Physical vapor deposition

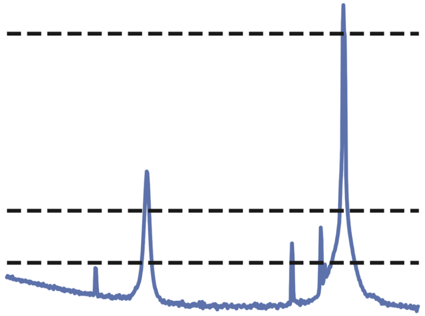


Standard PVD deposition



Combinatorial deposition to access large data



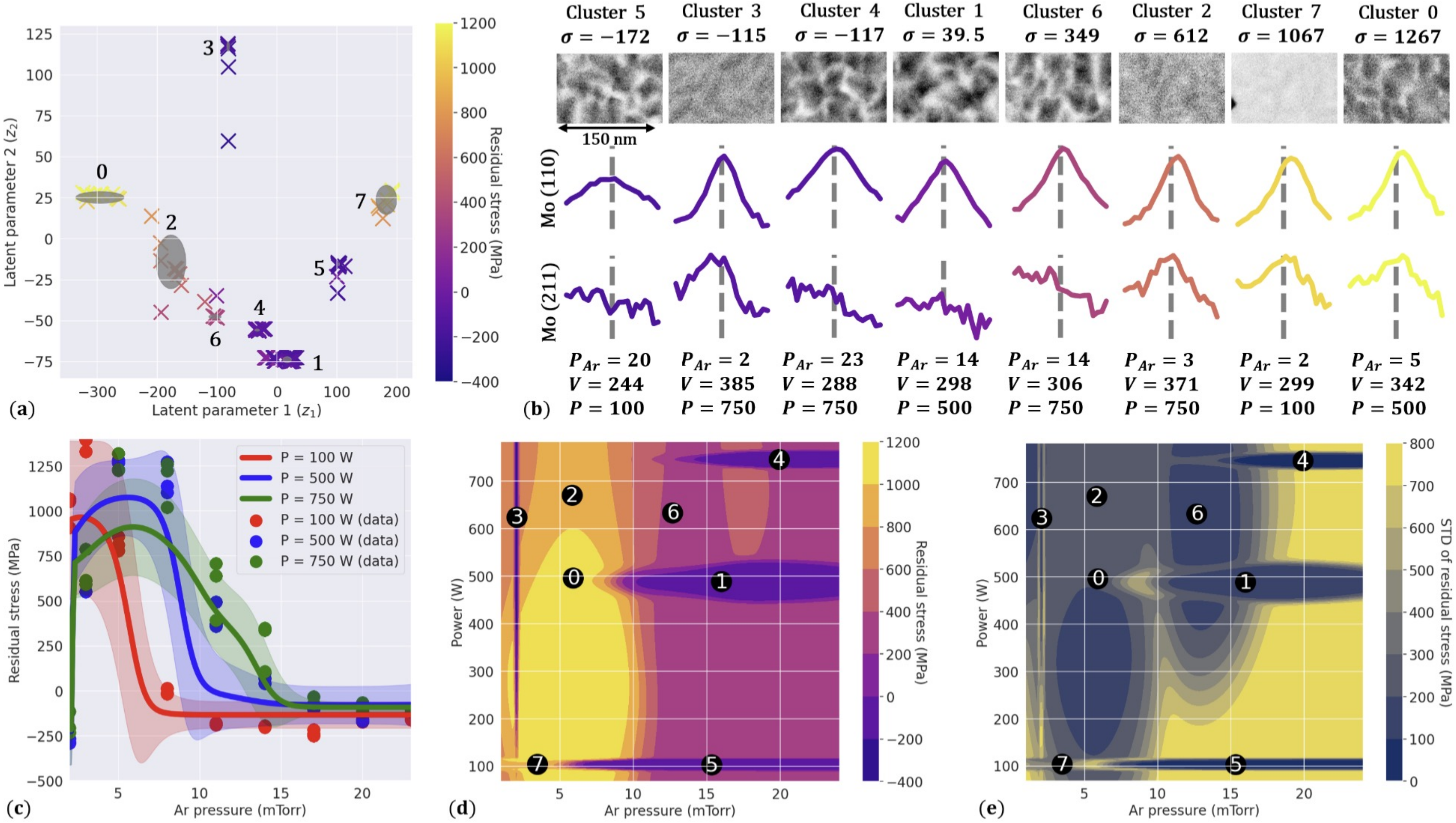


Multimodal characterization

- SEM images
- XRD profiles
- Residual stress measurements
- Process parameters

Implications of heterostructural alloying for enhanced piezoelectric performance of (Al,Sc)N
June 2018 Physical Review Materials 2(6)

Comprehensive multimodal embeddings of physical vapor deposition data

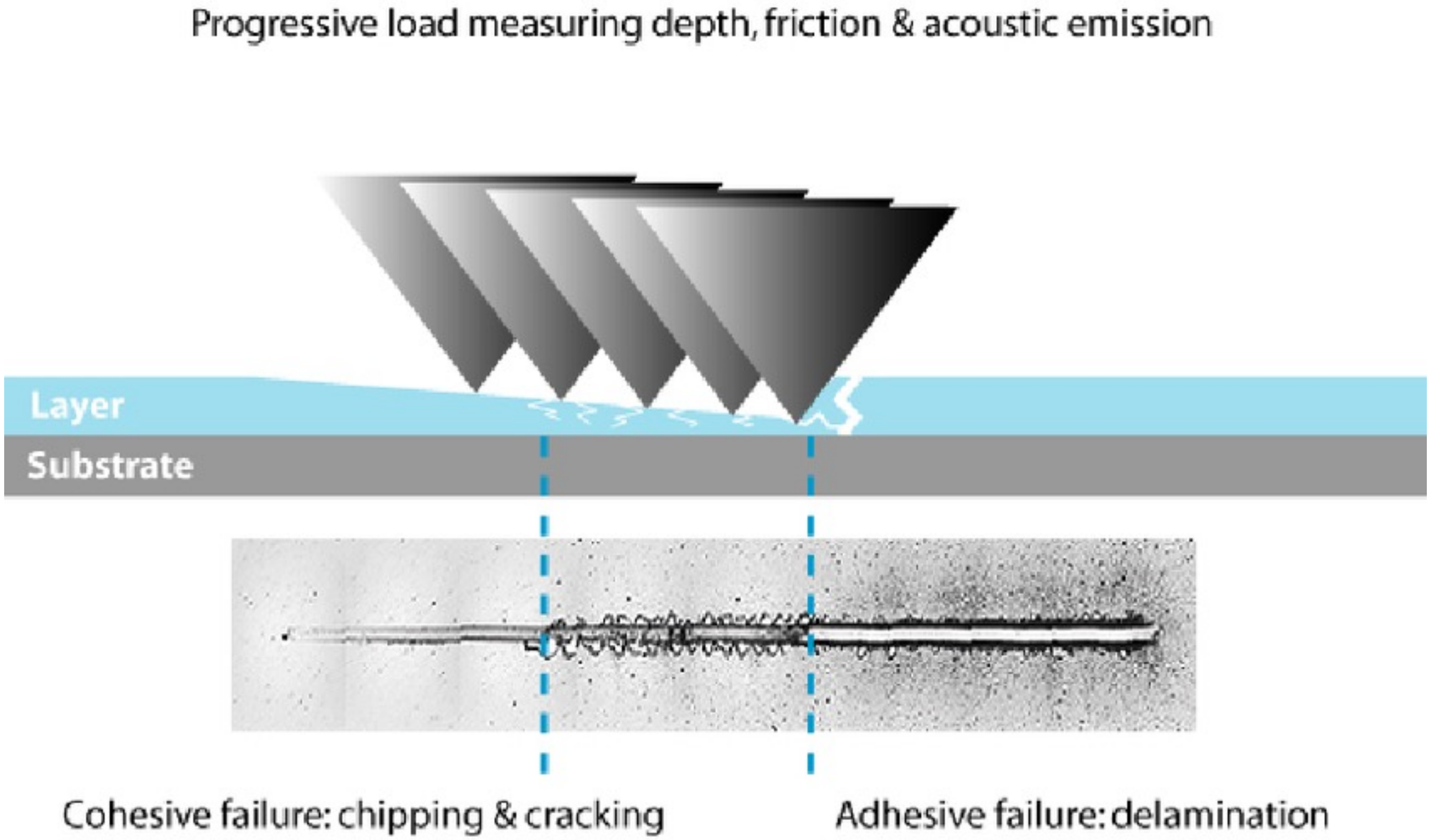


The real test: are we greater than the sum of the parts?

Input Modalities	Train	Test	Val.
(X_1, X_2, X_3, X_4)	1	1	1
(X_1, X_2, X_3)	1.225	1.310	1.413
(X_1, X_2)	1.827	2.745	3.728
(X_1, X_3)	1.286	1.323	1.473
(X_2, X_3)	1.286	1.441	1.466
(X_1)	2.736	3.528	4.057
(X_2)	2.247	2.779	3.618
(X_3)	1.640	1.419	1.600

Table 1. Averaged relative L_2 errors for 10 runs for crossmodal prediction of residual stress. Errors are presented as the ratio of the crossmodal prediction w.r.t. using all four modalities (i.e., all four modalities predict residual stress for the train/test/val datasets as 18.67%/23.36%/20.48%). Note: X_1 = SEM images, X_2 = XRD curves, X_3 = process parameters, X_4 = residual stress

Application 3: tribological testing of platinum gold films



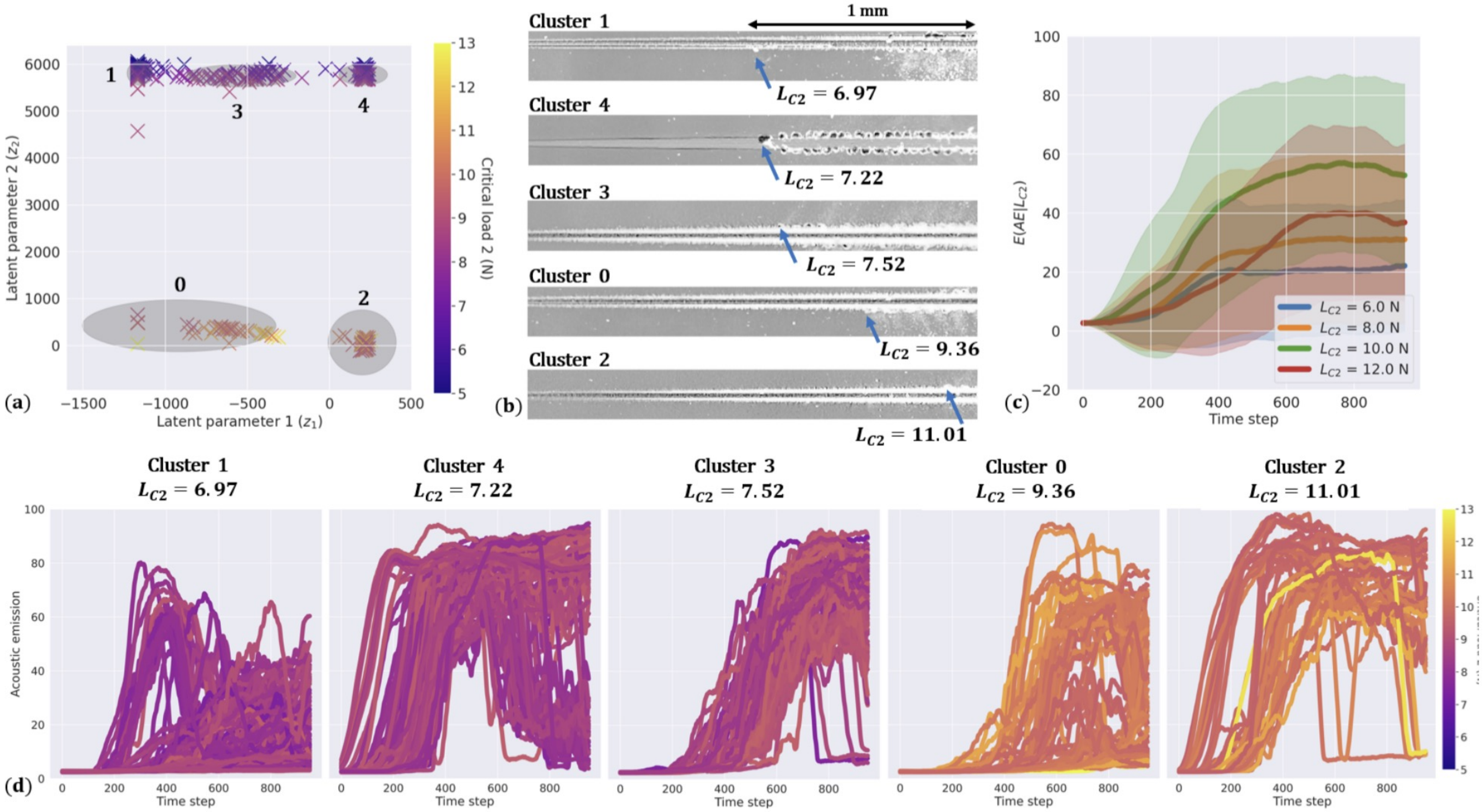
$$p(c|X_{m'_1}\ldots,X_{m'_N}) = \frac{\Pi_{i=1}^{N} p(X_{m'_i}|c)\pi_c}{\sum_{c'}\Pi_{i=1}^{N} p(X_{m'_i}|c')\pi_{c'}}$$

Science Question:
Can we use Bayesian estimate of cluster ownership to deduce LC2 from acoustic emission?

If we can, we can build high-throughput experiments to detect unconventional material fingerprints

Leroux, Pierre. (2014). AUTOMATED MULTI SAMPLE SCRATCH TESTING. 10.13140/RG.2.1.1022.8644.

Comprehensive multimodal embeddings of physical vapor deposition data



Closing thoughts

Embedding physics-informed models alongside physics agnostic data

Not just looking for good compression, but fingerprints which optimally describe relevant physics or support a given task

How can we use advances in SciML to embed PDE-based expert models?

See next two talks of session

Toward causality in material fingerprinting

How to identify root causes in fingerprints?

Modern architectures

Current SotA is data-hungry transformers – are there data-efficient alternatives to variational autoencoders?