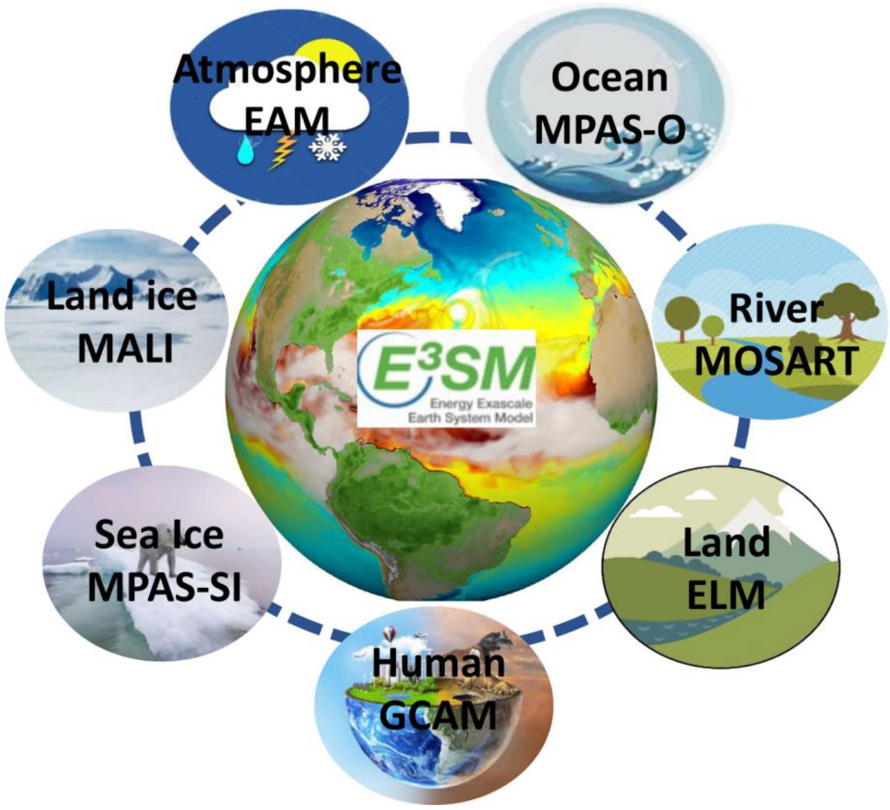
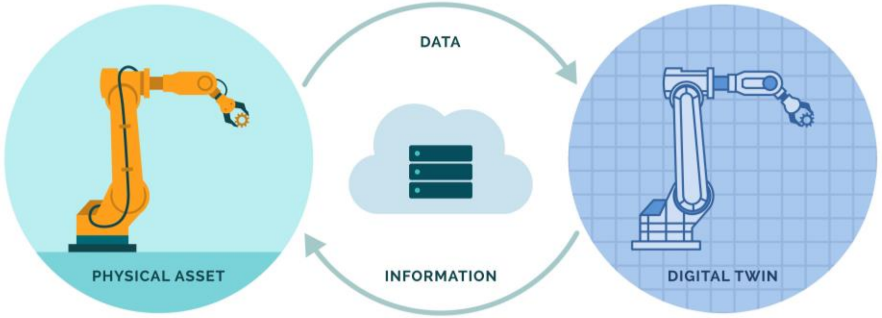
A data-driven exterior calculus for probabilistic digital twins

Nat Trask
University of Pennsylvania
Sandia National Laboratories

Rigorous digital twins for systems of systems



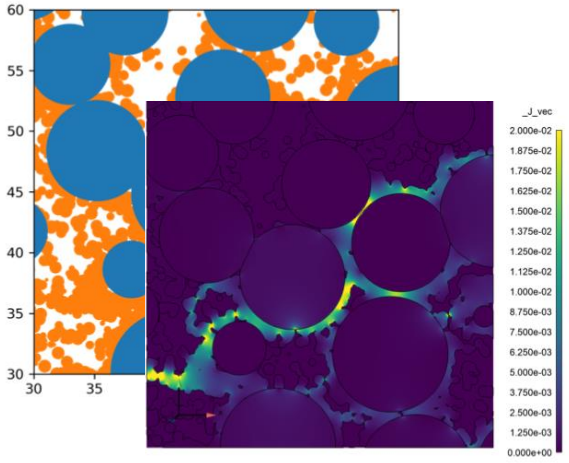


"The size and complexity of many systems being built for government, industry, and the military have reached a threshold where customary methods of analysis, design, implementation, and operation are no longer sufficiently reliable. Many of these large systems are properly described as "systems-of-systems" in that they are composed of many systems" (Dvorak 2005)

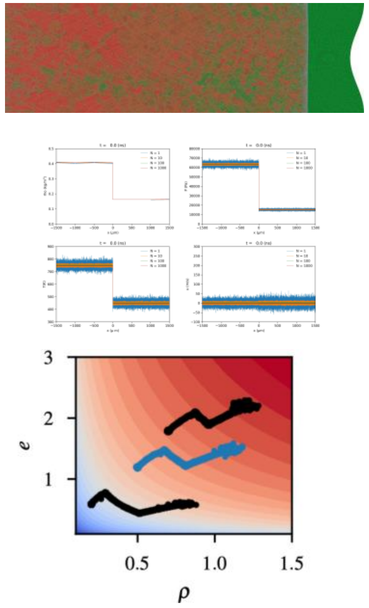
Mathematical Requirements

Near real-time prediction
Stable inter-model coupling
Efficient data assimilation
Physical structure preservation
Causal structure preservation
Support VV+UQ
Multimodal/multiscale data

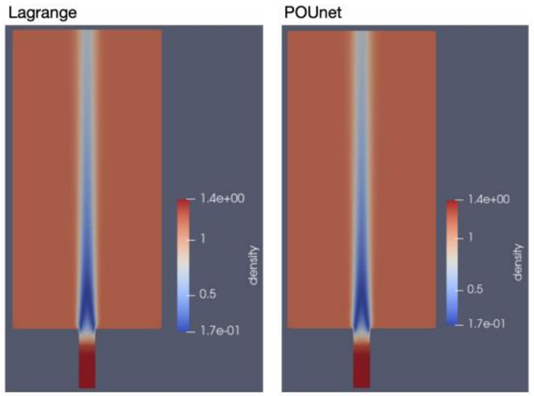
Data-driven models for high-consequence engineering



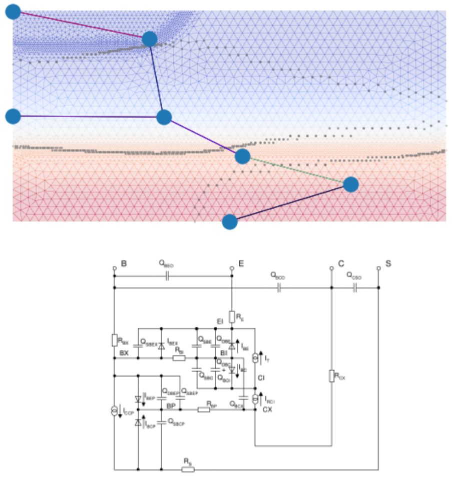
Data-driven multiscale FEM



Hyperbolicity preserving fluid closures



High-dimensional chemistry surrogates



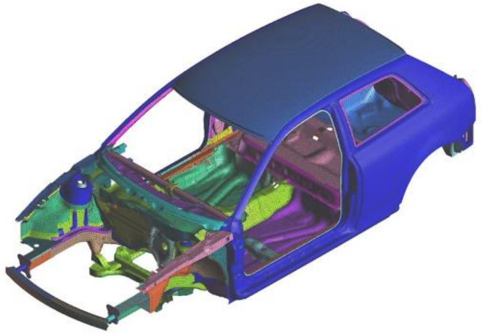
Robust sub-system surrogates in systems of systems

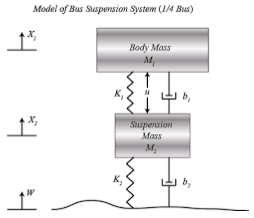
Extracting a physics-based model when first principles derivation is intractable

To **reliably** embed in **high-consequence** engineering applications – need guarantees

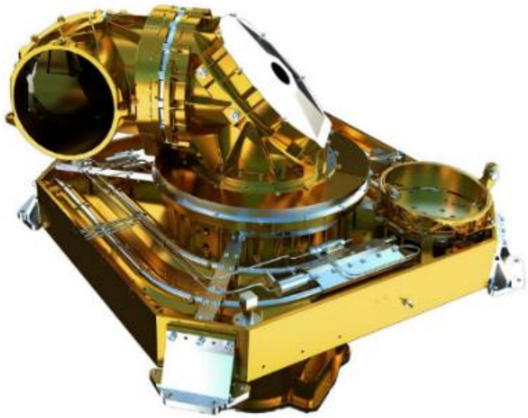
Systems of systems are ubiquitous in multiphysics/multiscale problems

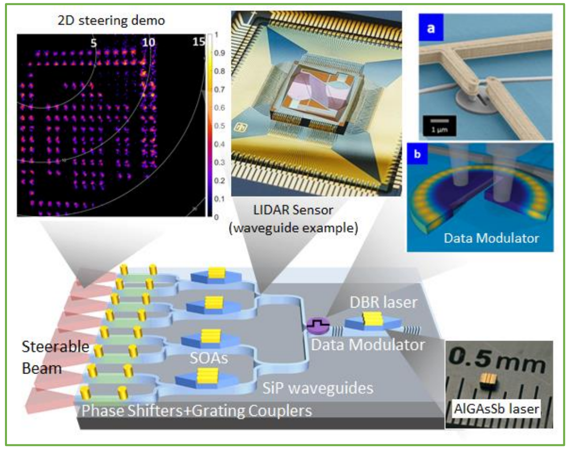
Mechanical assemblies
Multiphysics across coupled components



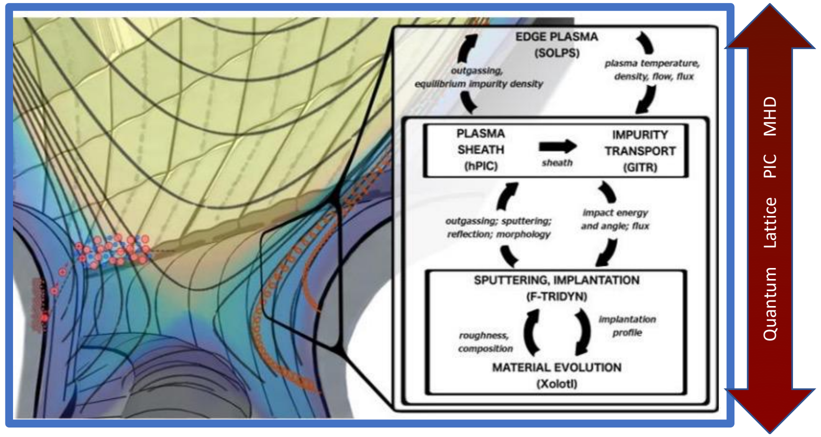


Microelectronics codesign
Two-way scale bridging between material, device, and circuit scales



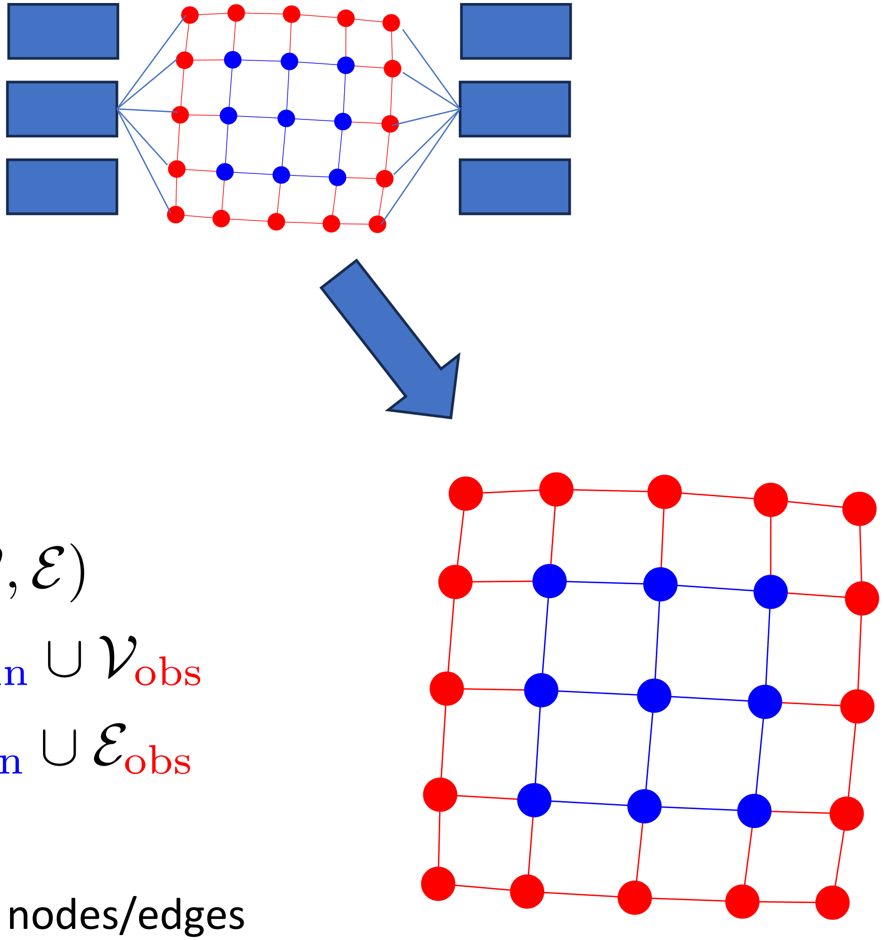


Fusion Power
Plasma/material interactions bridging scales and models



GOAL:
Rigorous multiphysics coupling preserving physical invariances with <u>guaranteed</u> performance

Why the exterior calculus? Structure preserving Dirichlet2Neumann maps



$$\mathcal{G} = (\mathcal{V}, \mathcal{E})$$
$$\mathcal{V} = \mathcal{V}_{\text{un}} \cup \mathcal{V}_{\text{obs}}$$
$$\mathcal{E} = \mathcal{E}_{\text{un}} \cup \mathcal{E}_{\text{obs}}$$

Red: Mortar nodes/edges
Blue: Internal nodes/edges

Definition
The *dirichlet2neumann* map for a hodge Laplacian problem is the unique linear map from potential functions on boundary nodes to currents at edges coincident on boundary nodes

We pose learning of surrogates on structure-preserving subgraphs coupled through flux/state relationships

$$F = d_0 u + \mathcal{N}[u; \theta]$$

$$\partial_t u + d_0^* F = f$$

DNN or GP

Graph discovery from full-field data – talk overview

Element	Circuit Analogy		Pipe Analogy		Flow Equation
Node	Junction		Junction		Total Flow = 0
Path	Wire		Rigid Pipe		Solve Directly
Resistance	Resistor		Aperture		$F = P/R$
Compliance	Capacitor		Diaphram		$F(t) = C\frac{dP(t)}{dt}$
Inertance	Inductor		Heavy Paddle		$F = \frac{1}{L}\int_{t_0}^{t} P\,dt + F(t_0)$
Switch	Switch		Gate Valve		Solve Directly
Valve	Diode		Check Valve		Solve Directly
Pressure Source	Voltage Source		Pump		Solve Directly
Flow Source	Current Source				$F = F$

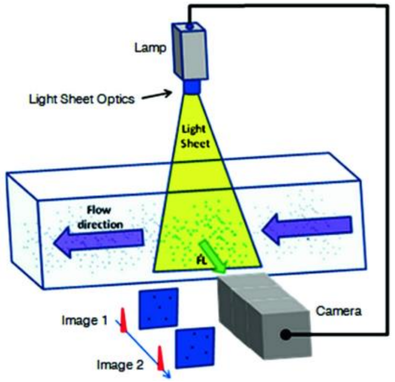
Systems admitting a circuit analogy admit a natural graph to "hang" a model on

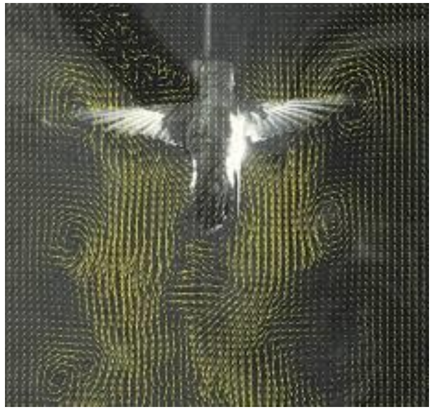
<u>Examples</u>

Electrical circuits

Subsurface fracture networks

Mesh from simulations





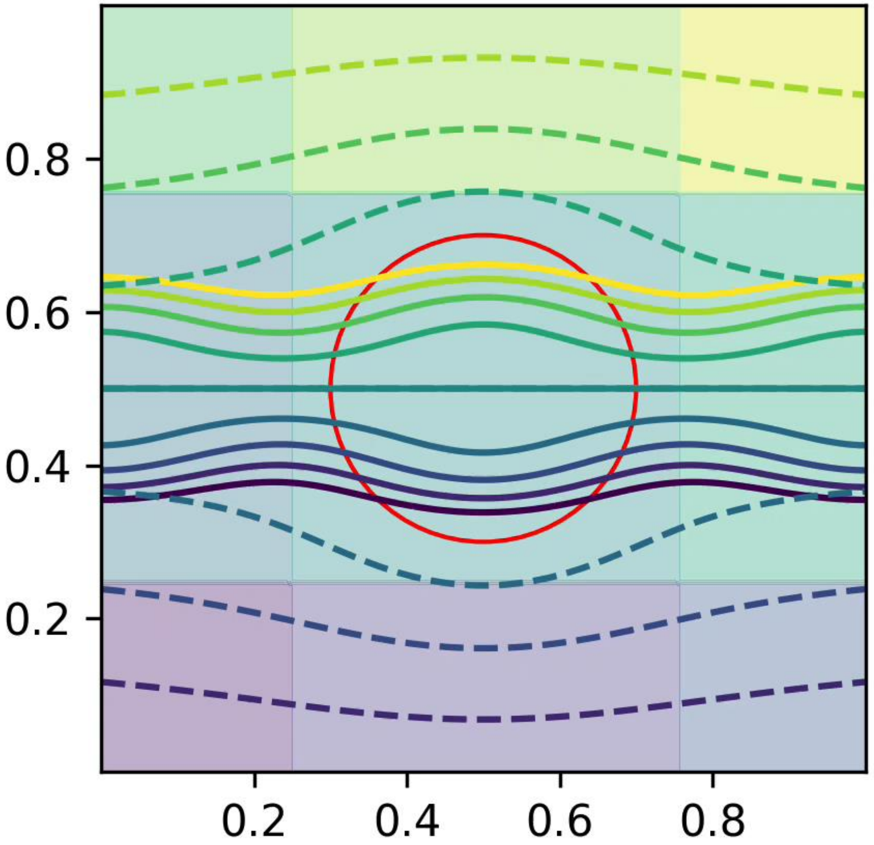
Full-field data has no natural graph

<u>Examples</u>

Particle imaging velocimetry

Digital imaging correlation

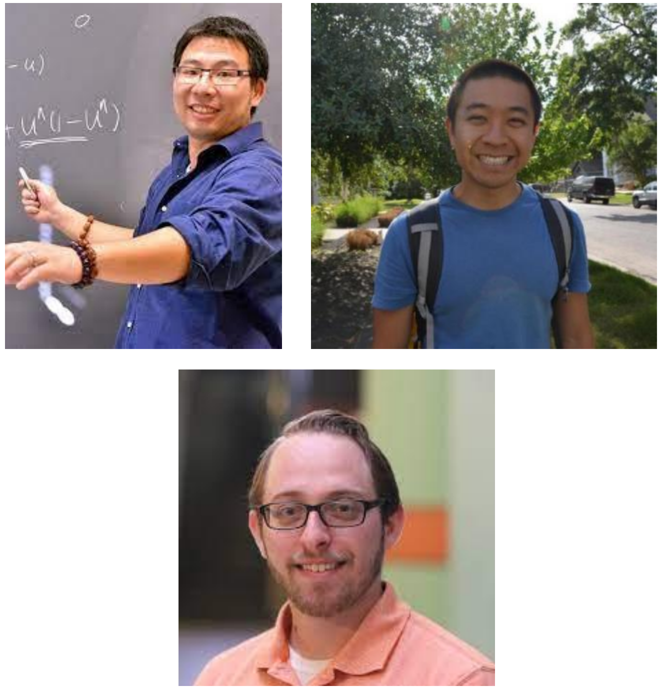
Homogenized particle systems



Use Whitney forms to exploit duality between geometry and graphs

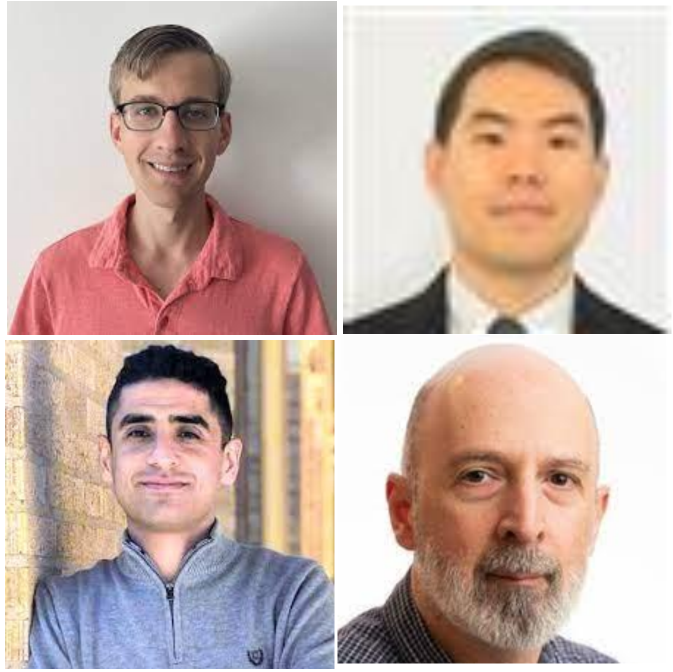
Repurpose classification networks to concurrently learn control volumes and balance laws

Collaborators



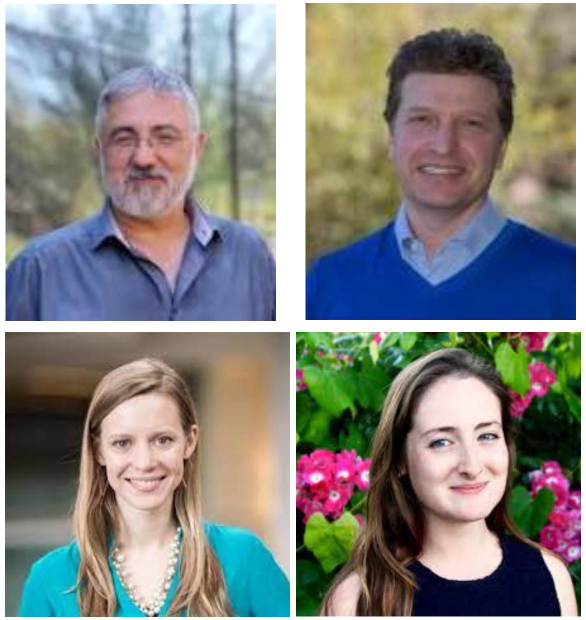
Data-driven exterior calculus

Xiaozhe Hu – Tufts
Andy Huang, Jonas Actor- SNL



Metriplectic bracket discovery

Anthony Gruber – SNL
Kookjin Lee - ASU
Erdi Kara – Spelman
Panos Stinis - PNNL



Optimal Recovery Problem

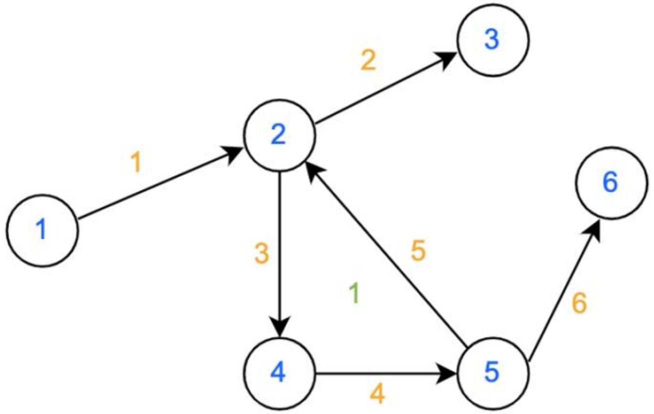
Houman Owhadi – Caltech
Daniel Tartakovsky, Adrienne Propp – Stanford
Jonas Actor, Elise Walker – SNL

Data-driven exterior calculus

div/grad/curl building blocks on graphs

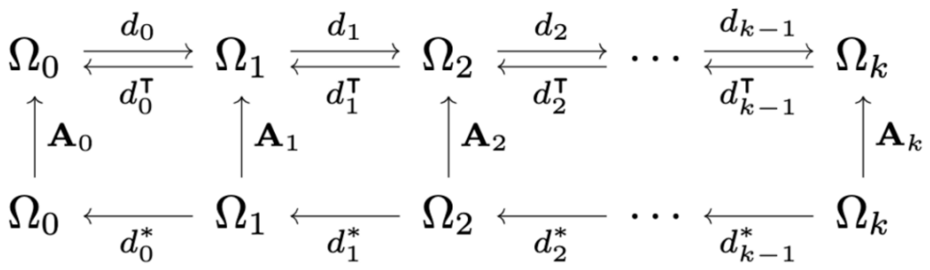
Mathematical preliminaries: Data-driven graph exterior calculus

Trask, Nathaniel, Andy Huang, and Xiaozhe Hu. "Enforcing exact physics in scientific machine learning: a data-driven exterior calculus on graphs." *Journal of Computational Physics* 456 (2022): 110969.



$$d_0 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

$$d_1 = \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$



d_k - matrix mapping from k to $k+1$ clique

d_k^* - matrix mapping from $k+1$ to k clique

A_k - machine learnable inner product

Choice of inner product induces a dual operator

$$\delta_{k+1} \circ \delta_k = 0 \qquad \delta_k^* \circ \delta_{k+1}^* = 0$$

De Rham complex encodes commuting diagram relationship

Definition: a **k-clique** is an ordered collection of nodes

Definition: a **graph coboundary operator** δ_k is a mapping from values associated with k-cliques to $(k+1)$-cliques

Graph Coboundary

$$d_k f_{i_0,\ldots,i_{k+1}} = \sum_{j=0}^{k+1} (-1)^j f\left(i_0,\ldots,\hat{i_j},\ldots,i_{k+1}\right)$$

Graph Gradient/Curl

$$d_0 u_{ij} = u_j - u_i$$

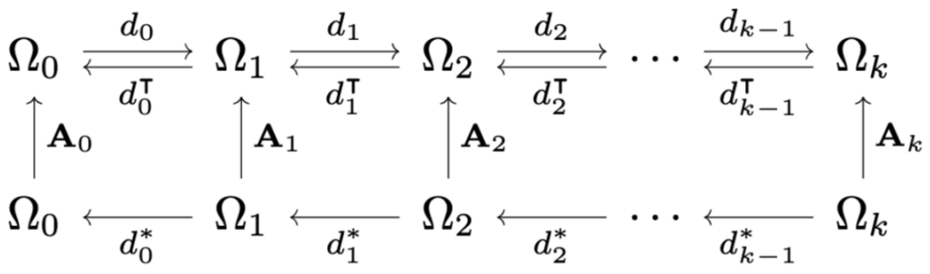
$$d_1 u_{ijk} = u_{ij} + u_{jk} + u_{ki}$$

IDEA

Traditional PDE: A_k are metrics from a mesh

Data-driven exterior calc: Fit to data w/ backprop

Result: Combinatorial Hodge + Lax Milgram theory for elliptic operators



Combinatorial Hodge Laplacian

$$\Delta_k = d_{k-1} d_{k-1}^* + d_k^* d_k$$

Obtain standard results from traditional finite element analysis:
- *Preserve exact sequence property*
- *Hodge decomposition*
- *Poincare inequality*
- *Lax-Milgram stability theory*
- *Conservation structure*

Theorem 3.1. *The discrete derivatives* d_k *in* (11) *form an exact sequence if the simplicial complex is exact, and in particular* $\mathsf{d}_{k+1} \circ \mathsf{d}_k = 0$. *In* $\mathbb{R}^3$, *we have* $CURL_h \circ GRAD_h = DIV_h \circ CURL_h = 0$.

Theorem 3.2. *The discrete derivatives* d_k^* *in* (11) *form an exact sequence of the simplicial complex is exact, and in particular* $\mathsf{d}_k^* \circ \mathsf{d}_{k+1}^* = 0$. *In* $\mathbb{R}^3$, $DIV_h^* \circ CURL_h^* = CURL_h^* \circ GRAD_h^* = 0$.

Theorem 3.3 (Hodge Decomposition)**.** *For* C^k, *the following decomposition holds*

$$C^k = \mathrm{im}(\mathsf{d}_{k-1}) \bigoplus_k \ker(\Delta_k) \bigoplus_k \mathrm{im}(\mathsf{d}_k^*), \tag{17}$$

where $\bigoplus_k$ *means the orthogonality with respect to the* $(\cdot,\cdot)_{\mathbf{D}_k \mathbf{B}_k^{-1}}$*-inner product.*

Theorem 3.4 (Poincaré inequality)**.** *For each* k, *there exists a constant* $c_{P,k}$ *such that*

$$\|\mathbf{z}_k\|_{\mathbf{D}_k \mathbf{B}_k^{-1}} \le c_{P,k} \|\mathsf{d}_k \mathbf{z}_k\|_{\mathbf{D}_{k+1} \mathbf{B}_{k+1}^{-1}}, \quad \mathbf{z}_k \in \mathrm{im}(\mathsf{d}_k^*),$$

and another constant $c_{P,k}^*$ *such that*

$$\|\mathbf{z}_k\|_{\mathbf{D}_k \mathbf{B}_k^{-1}} \le c_{P,k}^* \|\mathsf{d}_{k-1}^* \mathbf{z}_k\|_{\mathbf{D}_{k-1} \mathbf{B}_{k-1}^{-1}}, \quad \mathbf{z}_k \in \mathrm{im}(\mathsf{d}_{k-1}).$$

Thus, for $\mathbf{u}_k \in C^k$, *we have*

$$\inf_{\mathbf{h}_k \in \ker(\Delta_k)} \|\mathbf{u}_k - \mathbf{h}_k\|_{\mathbf{D}_k \mathbf{B}_k^{-1}} \le C \left(\|\mathsf{d}_k \mathbf{u}_k\|_{\mathbf{D}_{k+1} \mathbf{B}_{k+1}^{-1}} + \|\mathsf{d}_{k-1}^* \mathbf{u}_k\|_{\mathbf{D}_{k-1} \mathbf{B}_{k-1}^{-1}} \right),$$

where constant $C > 0$ *only depends on* $c_{P,k}$ *and* $c_{P,k}^*$.

Theorem 3.5 (Invertibility of Hodge Laplacian)**.** *The* k^{th}*-order Hodge Laplacian* Δ_k *is positive-semidefinite, with the dimension of its null-space equal to the dimension of the corresponding homology* $H^k = \ker(\mathsf{d}_k) / \mathrm{im}(\mathsf{d}_{k-1})$.

Enforcing exact physics via equality constrained QP

Trask, Nathaniel, Andy Huang, and Xiaozhe Hu. "Enforcing exact physics in scientific machine learning: a data-driven exterior calculus on graphs." *Journal of Computational Physics* 456 (2022): 110969.

Conservation structure
Exact physics treatment

$$\mathrm{d}_{k-1}\ \mathrm{d}_{k-1}^{*}\mathbf{u}_k + \mathrm{d}_k^{*}\mathbf{w}_{k+1} = \mathbf{f}_k$$

+

Generalized flux
Stabilize "black-box" physics with Hodge Laplacian

$$\mathbf{w}_{k+1} = d_k \mathbf{u}_k + \mathcal{N}[d_k \mathbf{u}_k; \theta]$$

Provides variational form
Conservation structure gives SBP formulas

$$a(\mathbf{u}, \mathbf{v}; \mathbf{A}) + N_{\mathbf{v}}[\mathbf{u}; \theta] = b(\mathbf{v})$$

Invertible bilinear form w/ metric params

Nonparametric estimator of "black-box" flux

IDEA
If constraint is feasible (guaranteed solvable) then we are guaranteed to obtain a model preserving structure independent of data/model fit

$$\underset{\mathbf{A},\theta}{\operatorname{argmin}} ||\mathbf{u} - \mathbf{u}_{data}||^2 + \epsilon^2 ||\mathbf{w} - \mathbf{w}_{data}||^2$$

$$\text{such that } a(\mathbf{u}, \mathbf{v}; \mathbf{A}) + N_{\mathbf{v}}[\mathbf{u}; \theta] = b(\mathbf{v}) \quad \forall \mathbf{v}$$

Enforcing exact physics via equality constrained QP

Trask, Nathaniel, Andy Huang, and Xiaozhe Hu. "Enforcing exact physics in scientific machine learning: a data-driven exterior calculus on graphs." *Journal of Computational Physics* 456 (2022): 110969.

$$a(\mathbf{u}, \mathbf{v}; \mathbf{A}) + N_{\mathbf{v}}[\mathbf{u}; \theta] = b(\mathbf{v})$$

Theorem 3.6. *The equation* (24) *has at least one solution* $\mathbf{u}_k \in \mathbb{V}$ *satisfies*

$$\|\mathbf{u}_k\| \leq \frac{\|\mathbf{f}\|}{(C_p - C_N)}. \quad (26)$$

Theorem 3.7. *If* $\frac{C_{\nabla N}\|\mathbf{f}\|}{C_p(C_p - C_N)} < 1$, *then the equation* (24) *has at most one solution in* $\mathbb{V}$.

A unique solution exists if the Hodge-Laplacian is sufficiently large relative to the nonlinear part, following standard elliptic PDE arguments

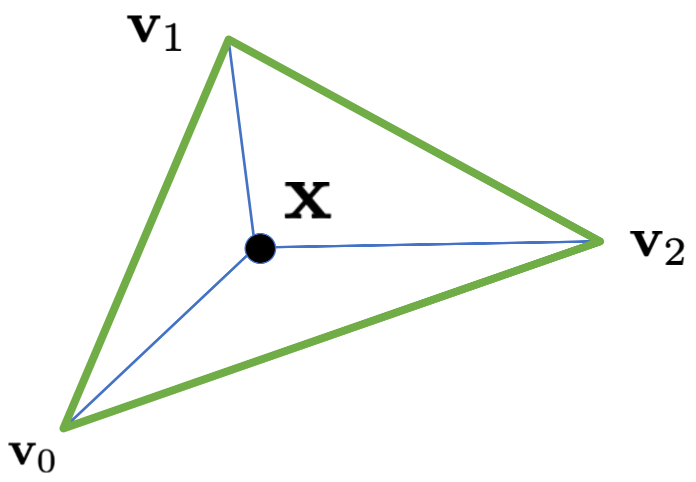
- Poincare constant easily estimated from matrix eigenvalues
- Lipschitz constant on nonlinearity straightforward for DNNs

Solvability constraint could be enforced during training if desired

Data-driven Whitney forms

Learning a finite element space which implicitly defines a graph

Mathematical preliminaries: barycentric coordinates



Consider the simplex $\mathcal{S} = \{\mathbf{v}_0, ..., \mathbf{v}_d\}$.

Definition: Barycentric coordinates $\{\lambda_i\}_i$ are the unique linear polynomials satisfying

- $\sum_i \lambda_i(\mathbf{x}) = 1$ for all $\mathbf{x}$ ← **Partition of Unity property**
- $\sum_i \mathbf{v}_i \lambda_i(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x}$

Definition: The barycentric interpolant is defined $f_B(\mathbf{x}) = \sum_i f(\mathbf{v}_i)\lambda_i(\mathbf{x})$

Mathematical preliminaries: low-order Whitney forms

Whitney Form Definition
Construction follows from barycentric interpolant

$$\mathcal{W}_{j_0 \cdots j_k} = k! \sum_{i=0}^{k} (-1)^i \lambda_{j_i} \mathbf{d}\lambda_{j_0} \wedge \cdots \widehat{\mathbf{d}\lambda_{j_i}} \wedge \cdots \wedge \mathbf{d}\lambda_{j_k}$$

P1 nodal Lagrange Elements

$$\text{span}\left\{1, x_1, x_2\right\}$$

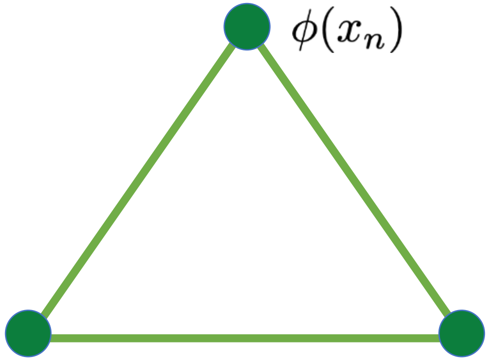
Nedelec Edge Elements

$$\text{span}\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}x_1\\-x_2\end{bmatrix}\right\}$$

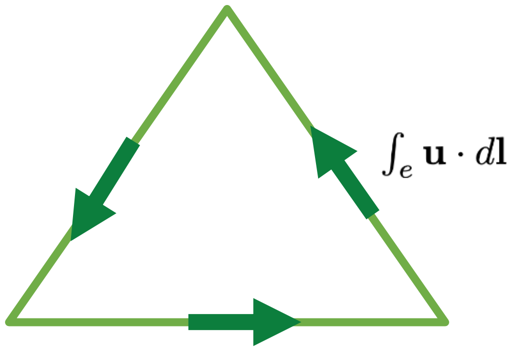
Raviart-Thomas Face Elements

$$\text{span}\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}x_1\\x_2\end{bmatrix}\right\}$$

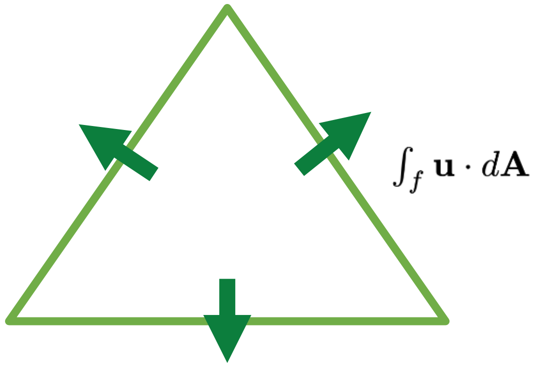
DOF are integral moments associated with mesh node/edge/face/cells



$$\mathcal{W}_i = \lambda_i$$

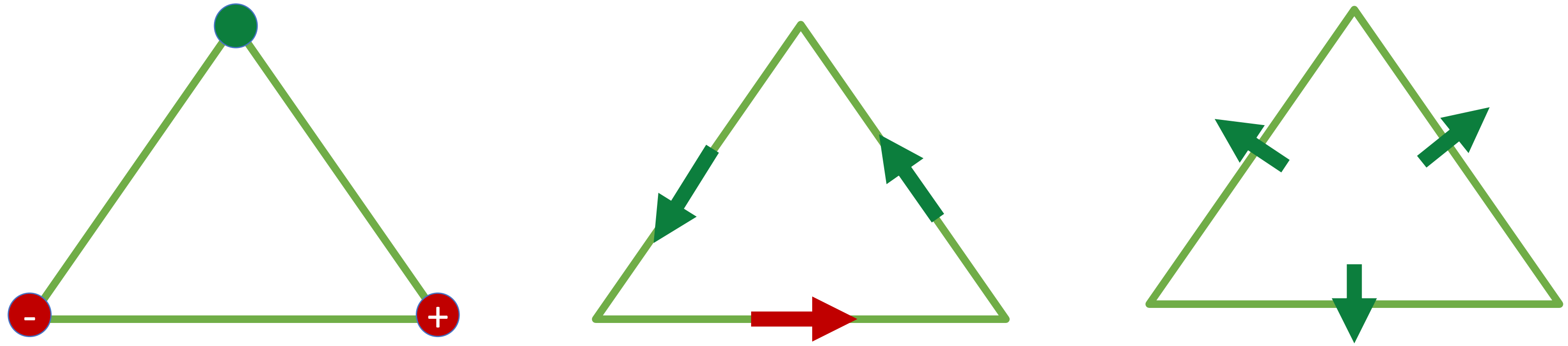


$$\mathcal{W}_{ij} = \lambda_i \nabla \lambda_j - \lambda_j \nabla \lambda_i$$



$$\mathcal{W}_{ijk} = \lambda_i \nabla \lambda_j \times \nabla \lambda_k + \lambda_j \nabla \lambda_k \times \nabla \lambda_i + \lambda_k \nabla \lambda_i \times \nabla \lambda_j$$

What's special about Whitney forms?



$$\mathcal{W}_i = \lambda_i$$

$$\mathcal{W}_{ij} = \lambda_i \nabla \lambda_j - \lambda_j \nabla \lambda_i$$

$$\mathcal{W}_{ijk} = \lambda_i \nabla \lambda_j \times \nabla \lambda_k + \lambda_j \nabla \lambda_k \times \nabla \lambda_i + \lambda_k \nabla \lambda_i \times \nabla \lambda_j$$

$$\subset H_1(grad;\Omega) \quad \subset H_1(curl;\Omega) \quad \subset H_1(div;\Omega)$$

$$\int_e \nabla \phi \cdot d\mathbf{l} = \phi(n_+) - \phi(n_-)$$

$$\int_f \nabla \times \mathbf{u} \cdot d\mathbf{A} = \sum_{e \in \partial f} \int_f \mathbf{u} \cdot d\mathbf{l}$$

$$\mathsf{DIV} = \delta_0^T \mathbf{M}_1, \quad \mathsf{CURL} = 2\,\delta_1^T \mathbf{M}_2, \quad \mathsf{GRAD} = 6\,\delta_2^T \mathbf{M}_3$$

Exterior derivative (DEC)

Graph coboundary

FEM mass matrix

Exact treatment of vector calculus operations
Image of exterior derivative is onto, giving pointwise exact sequence property

Explicit tie to conservation structure
Derivatives map differential forms directly using generalized Stokes theorems

Alternative view in the graph exterior calculus
Mass matrix can be viewed as imposing sparsity on a fully connected graph

Mathematical preliminaries: partition of unity

Traditional role:
Localizing approximation
Identifying charts of atlas

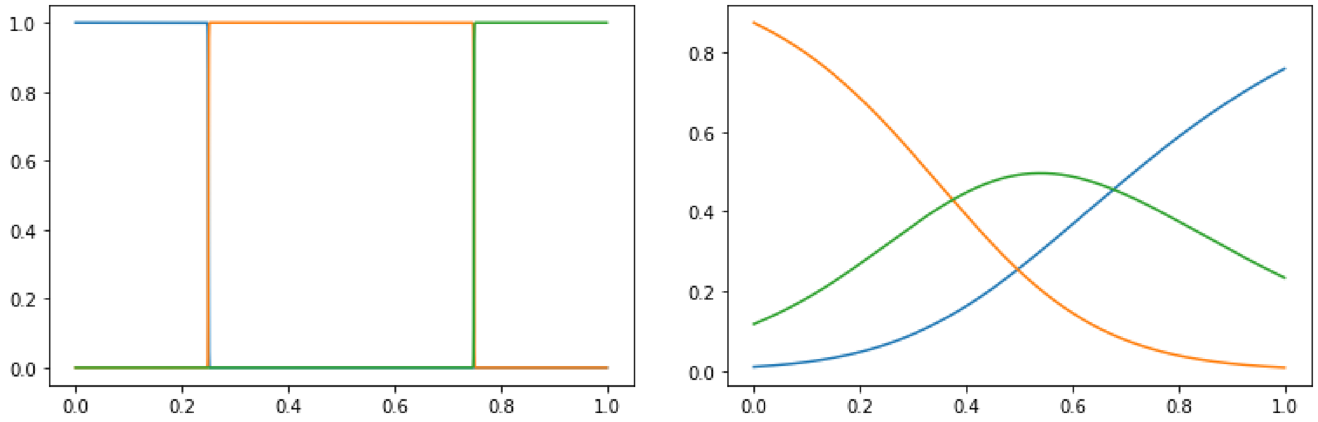
Our use:
Replace barycentric coordinates in Whitney form construction

Definition: *Partition of unity (POU)*
A collection of functions $\{\phi_i\}_{i=1,\ldots,N}$ satisfying

- $\phi_i > 0$
- $\sum_i \phi_i = 1$

Example:
Consider a partition of $\Omega \subset \mathbb{R}^d$ into disjoint cells $\Omega = \bigcup_i C_i$. Then the indicator functions $\phi_i(x) = \mathbb{1}_{C_i}(x)$ form a POU.

$$\phi_i(x) = \text{softmax} \circ NN(x;\theta)$$

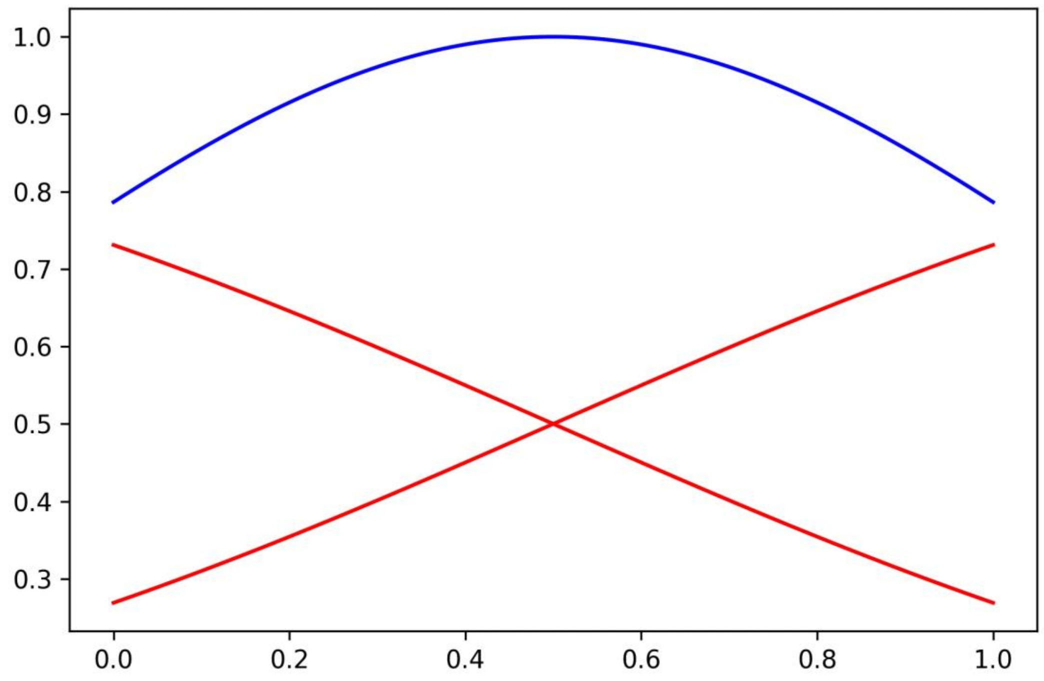


POU corresponding to Cartesian mesh vs learnable POU with non-disjoint support associated with a traditional logistic regression network for a categorical RV

Whitney forms define diffuse boundary operators

Red: POU on cells
Blue: Boundary of POUS

In limit of disjoint partitions, want to recover oriented Dirac distribution



$$\int_c \nabla \cdot \mathbf{u} dx = \int_{f \in \partial c} \mathbf{u} \cdot d\mathbf{A}$$

POUs generalize cell

Defining boundary operator provides exterior derivative

IDEA:
Replace barycentric coordinates with machine learnable POU and perform standard Whitney form construction

$$\mathcal{W}_i = \lambda_i$$

$$\mathcal{W}_{ij} = \lambda_i \nabla \lambda_j - \lambda_j \nabla \lambda_i$$

Using automatic differentiation, we obtain a fully differentiable ML architecture generalizing a traditional computational mesh

Whitney forms define diffuse boundary operators

Compare to: $V_i \nabla \cdot \mathbf{u}_i = \sum_j \mathbf{A}_{ij} \cdot \mathbf{u}_{ij}$

- Let $\psi_i = \phi_i$. Define a function space $V_0 = \{\sum_i c_i \psi_i(x) \,|\, c_i \in \mathbb{R}^{N_0}\}$.
- Integrating by parts we obtain

$$\int_\Omega \psi_i \nabla \cdot \mathbf{u} = -\int_\Omega \nabla \phi_i \cdot \mathbf{u} + \int \phi_i \mathbf{u} \cdot dA$$

POU property
Multiply by one

$$= -\sum_j \int_\Omega \phi_j \nabla \phi_i \cdot \mathbf{u} + \int \phi_i \mathbf{u} \cdot dA$$

Grad of POU property
Add zero

$$= \sum_{j \neq i} \int_\Omega (\phi_i \nabla \phi_j - \phi_j \nabla \phi_i) \cdot \mathbf{u} + \int \phi_i \mathbf{u} \cdot dA$$

$$= \sum_{j \neq i} \int_\Omega \psi_{ij} \cdot \mathbf{u} + \int \phi_i \mathbf{u} \cdot dA$$

where $\psi_{ij} = \phi_i \nabla \phi_j - \phi_j \nabla \phi_i$, and we note that $\psi_{ij} = -\psi_{ji}$.

Proceed by induction on arbitrary manifolds

Replace IBP with Leibniz rule:

$$\int_\Omega (d\omega_k) \wedge \omega_l = (-1)^{k+1} \int_\Omega \omega_k \wedge (d\omega_l) + \int_{\partial\Omega} \operatorname{tr} \omega_k \wedge \operatorname{tr} \omega_l$$

Inductively define Whitney form shape functions by mimicking construction:

$$\psi^k_{j_0 \cdots j_k} = k! \sum_{i=0}^{k} (-1)^i \phi_{j_i} d\phi_{j_0} \wedge \cdots \wedge \widehat{d\phi_{j_i}} \wedge \cdots \wedge d\phi_{j_k}$$

Obtain discrete "differential form" DOFs that induce coboundary operator:

$$U_{j_0 \cdots j_{k+1}} = \int_\omega u \wedge \psi^{k+1}_{j_0 \cdots j_{k+1}}$$

$$D_k(U)_{j_0 \cdots j_k} = (-1)^{n-1} \sum_{j_{k+1} \neq j_0, \cdots, j_k} U_{j_0 \cdots j_{k+1}} + \int \operatorname{tr} u \wedge \operatorname{tr} \psi^k_{j_0 \cdots j_k}$$

Preserve exact sequence property to induce de Rham complex:

$$D_k \circ D_{k-1}(U)_{j_0 \cdots j_{k-1}} = \int_\Omega d(du) \wedge \phi^k_{j_0 \cdots j_k} = 0$$

Requirement – how to build mass matrix

Continuous Galerkin treatment of Hodge operator

$$\begin{aligned} (\mathbf{F}, \mathbf{E}) - (\mathsf{d}^0 p_0, \mathbf{E}) &= (\mathsf{d}^0 p_D, \mathbf{E}) \\ -(\mathbf{F}, \mathsf{d}^0 q_0) &= (f, q_0) - (g_N, q_0)_{\Gamma_N}. \end{aligned}$$

Metric now comes from FEEC mass matrix

$$(F, \nabla q) \rightarrow DIV\ F = d_0^{\intercal} \underline{\mathbf{M}}$$

Only choice: how to specifically design architecture for initial POU?

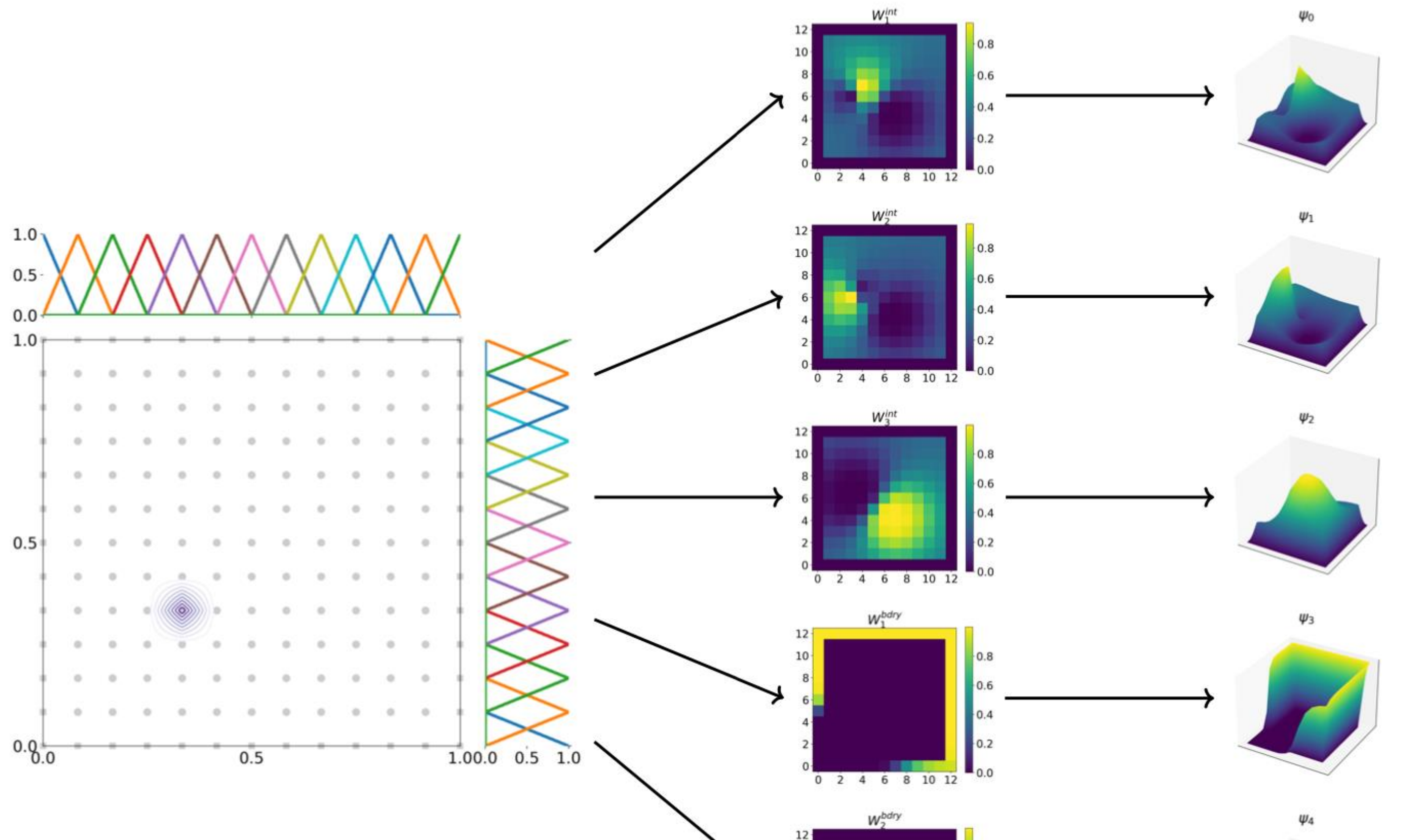
$$(F, \nabla q) \to DIV\ F = d_0^{\intercal}\mathbf{M}$$

Bilinear forms for conservation statements yield graph coboundary times mass matrices

If mass matrix is easily computable, we obtain a continuous Galerkin (AKA no variational crimes)

Option 1:
Convex combinations of B-splines

<u>Quadrature via pull-back to Cartesian mesh</u>



$$\phi_i(\boldsymbol{x}) = \sum_{k_1=1}^{N_{x_1}} \cdots \sum_{k_n=1}^{N_{x_n}} W_{k_1 \dots k_n i}\, \zeta_{k_1 \dots k_n}(\boldsymbol{x})$$

Define POU as convex combination of B1-splines

Only choice: how to specifically design architecture for initial POU?

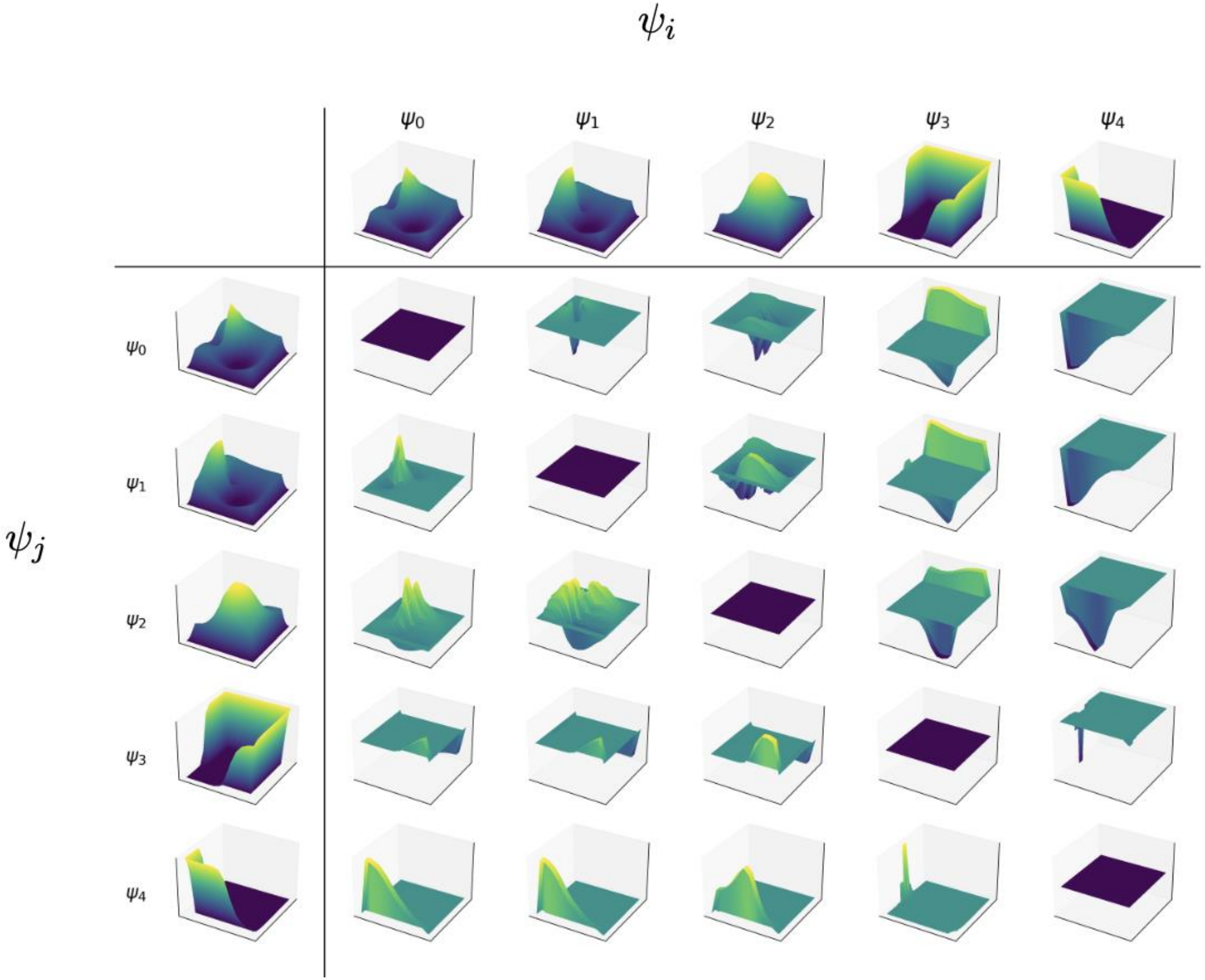
$$(F, \nabla q) \rightarrow DIV\ F = d_0^{\intercal}\mathbf{M}$$

Bilinear forms for conservation statements yield graph coboundary times mass matrices

If mass matrix is easily computable, we obtain a continuous Galerkin (AKA no variational crimes)

Option 1:
Convex combinations of B-splines

Quadrature via pull-back to Cartesian mesh



Only choice: how to specifically design architecture for initial POU?

$$(F, \nabla q) \rightarrow DIV\ F = d_0^\intercal \mathbf{M}$$

Bilinear forms for conservation statements yield graph coboundary times mass matrices

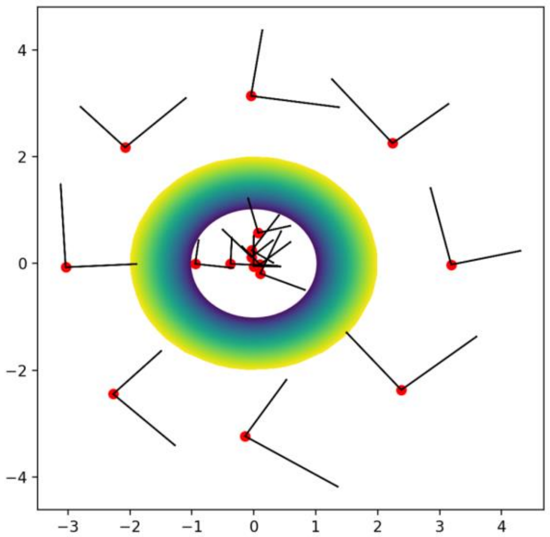
If mass matrix is easily computable, we obtain a continuous Galerkin (AKA no variational crimes)

Option 2:
Multivariate Gaussian PDF as POU

Analytic expressions for mass matrices

Choose $\psi_i = \mathcal{N}(\mathbf{m}_i, \mathbf{C}_i) = \phi_i$
Then $\psi_{ij} = \phi_i \nabla \phi_j = \phi_i \phi_j C_j^{-1}(\mathbf{m}_j - x)$
And mass matrix over infinite domain has closed form expression

$$\mathbf{M} = \int \psi_{ij}\psi_{kl} = C^{-1}\mathbb{E}_{x\sim\phi_i\phi_j}\left[x - \mathbf{m}\right]$$



Trainable mean and covariance allows for partitions to move and find optimal arrangement

Finally:

Continuous Galerkin treatment of Hodge operator

$$\begin{aligned} (\mathbf{F}, \mathbf{E}) - (\mathsf{d}^0 p_0, \mathbf{E}) &= (\mathsf{d}^0 p_D, \mathbf{E}) \\ -(\mathbf{F}, \mathsf{d}^0 q_0) &= (f, q_0) - (g_N, q_0)_{\Gamma_N}. \end{aligned}$$

Metric now comes from FEEC mass matrix

$$(F, \nabla q) \rightarrow DIV\ F = d_0^{\intercal} \mathbf{M}$$

$$a(\mathbf{u}, \mathbf{v}; \mathbf{A}) + N_{\mathbf{v}}[\mathbf{u}; \theta] = b(\mathbf{v})$$

$$\begin{gathered} \underset{\mathbf{A}, \theta}{\operatorname{argmin}} ||\mathbf{u} - \mathbf{u}_{data}||^2 + \epsilon^2 ||\mathbf{w} - \mathbf{w}_{data}||^2 \\ \text{such that } a(\mathbf{u}, \mathbf{v}; \mathbf{A}) + N_{\mathbf{v}}[\mathbf{u}; \theta] = b(\mathbf{v}) \quad \forall \mathbf{v} \end{gathered}$$

Data-driven Whitney forms

Applications

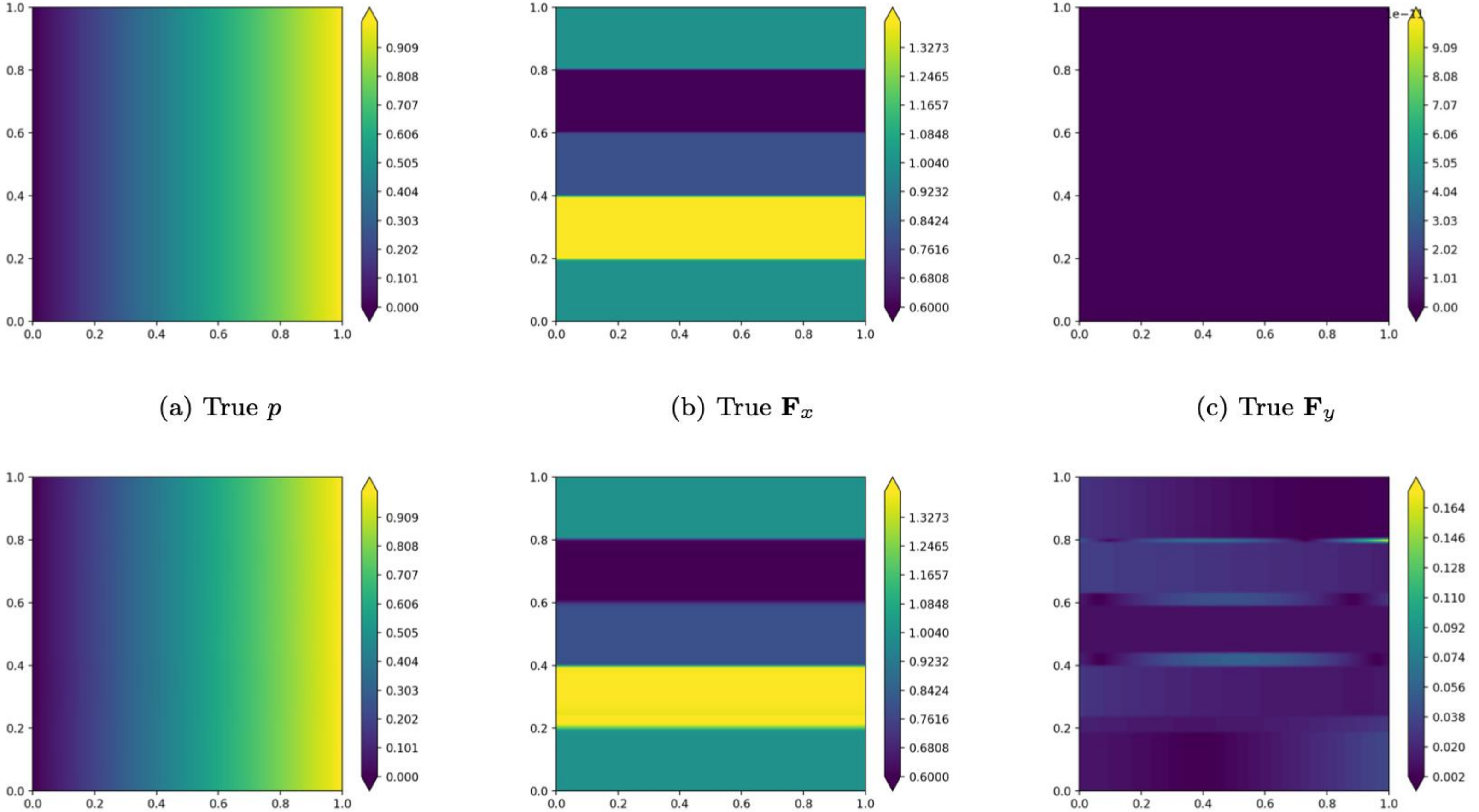
Applications: unsupervised identification of material properties

No nonlinear fluxes

$$\nabla \cdot \underbrace{\kappa(x)}\nabla\phi = f \qquad [\![\hat{n} \cdot \kappa\nabla\phi]\!] = 0$$

Discontinuity parallel to gradient

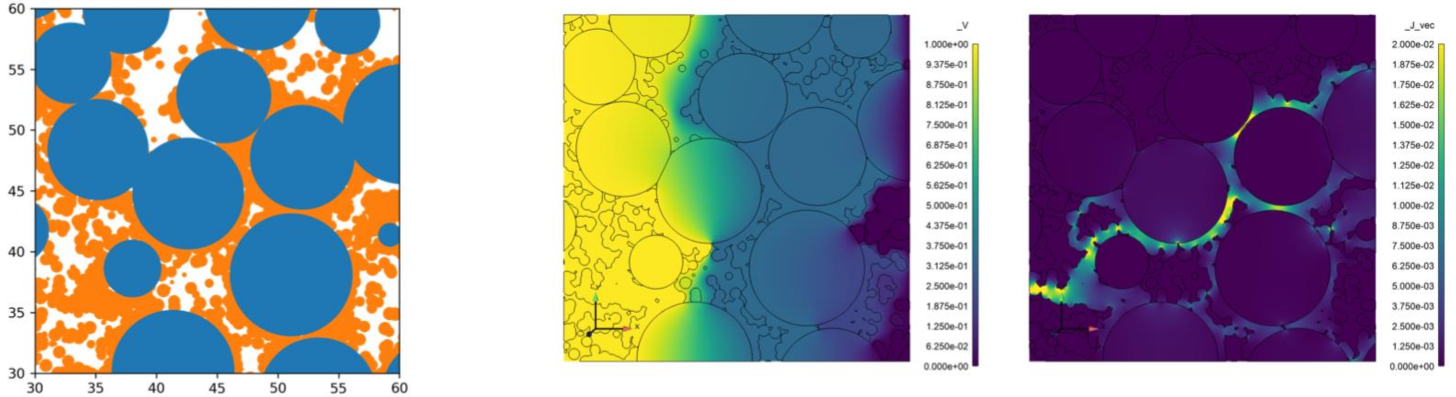
Mass conservation



(a) True p (b) True $\mathbf{F}_x$ (c) True $\mathbf{F}_y$

(d) Predicted p (e) Predicted $\mathbf{F}_x$ (f) Predicted $\mathbf{F}_y$

Applications: digital twins of as built Lithium-ion battery cathode

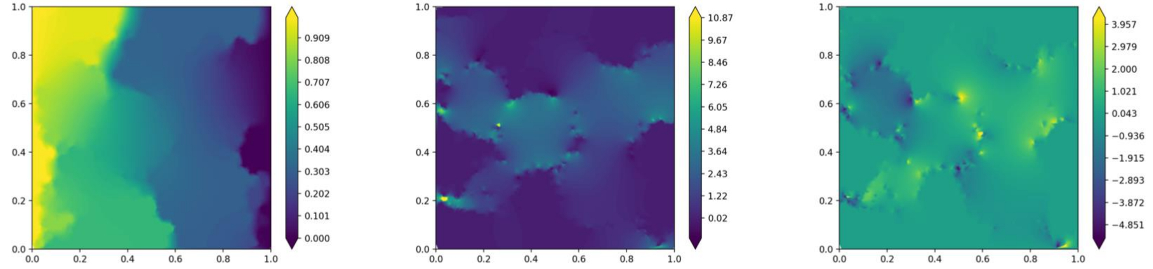


(a) Matrix layout

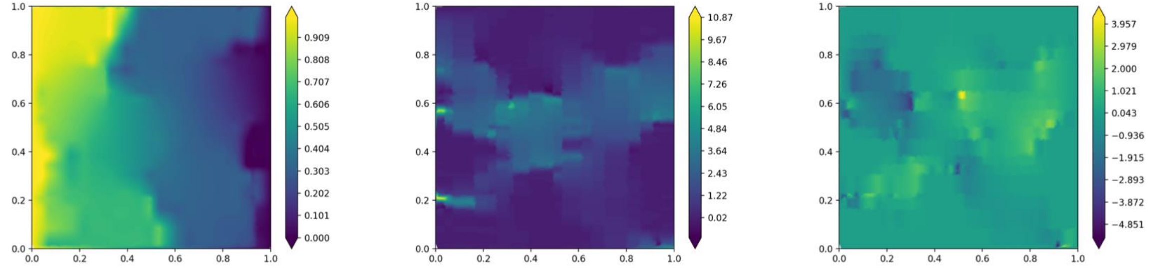
(b) Voltage

(c) Magnitude of current

Replace a 5.89M finite element simulation of as-built geometry with 8 data-driven elements w/ ~0.1% error **implemented in production FEM code**



(a) True p (b) True $\mathbf{F}_x$ (c) True $\mathbf{F}_y$



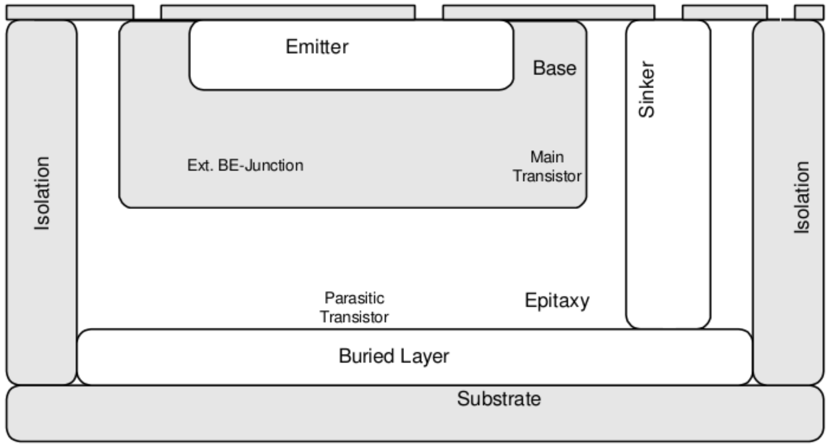
(d) Predicted p (e) Predicted $\mathbf{F}_x$ (f) Predicted $\mathbf{F}_y$

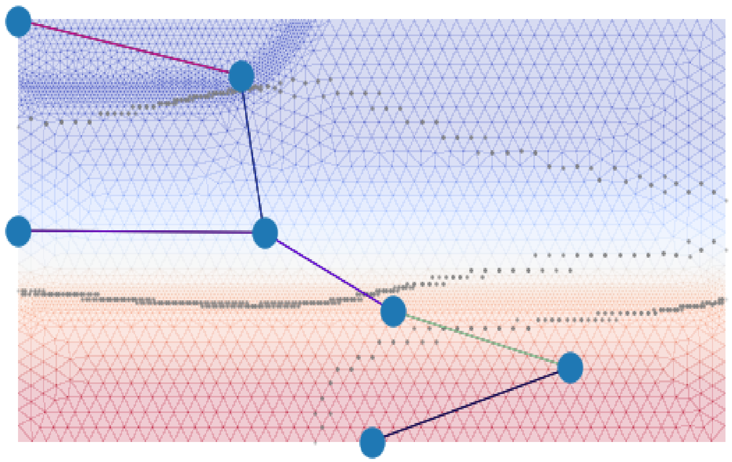
Applications: digital twins of bipolar junction transistor

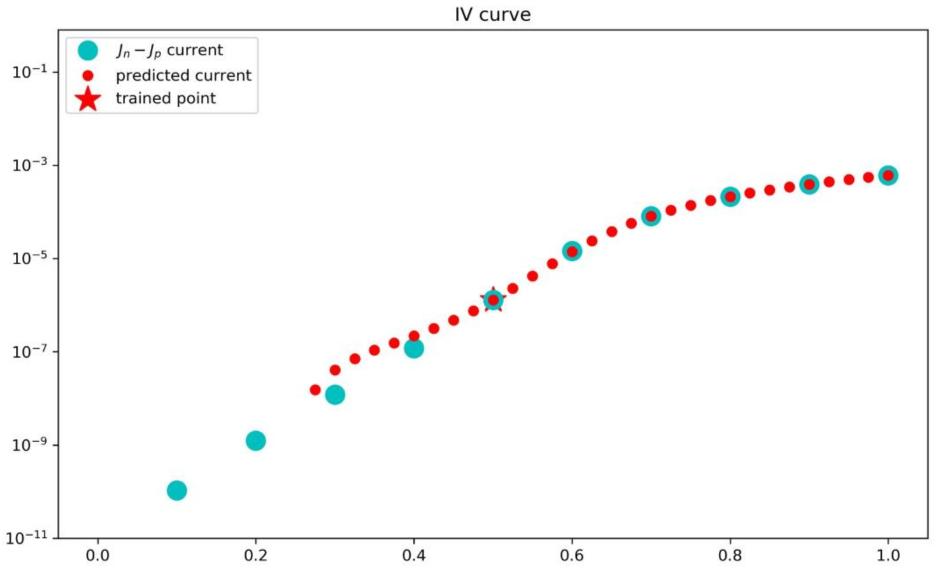
Nonlinear fluxes

Extrapolation from single data point calibration

(More on this one later)

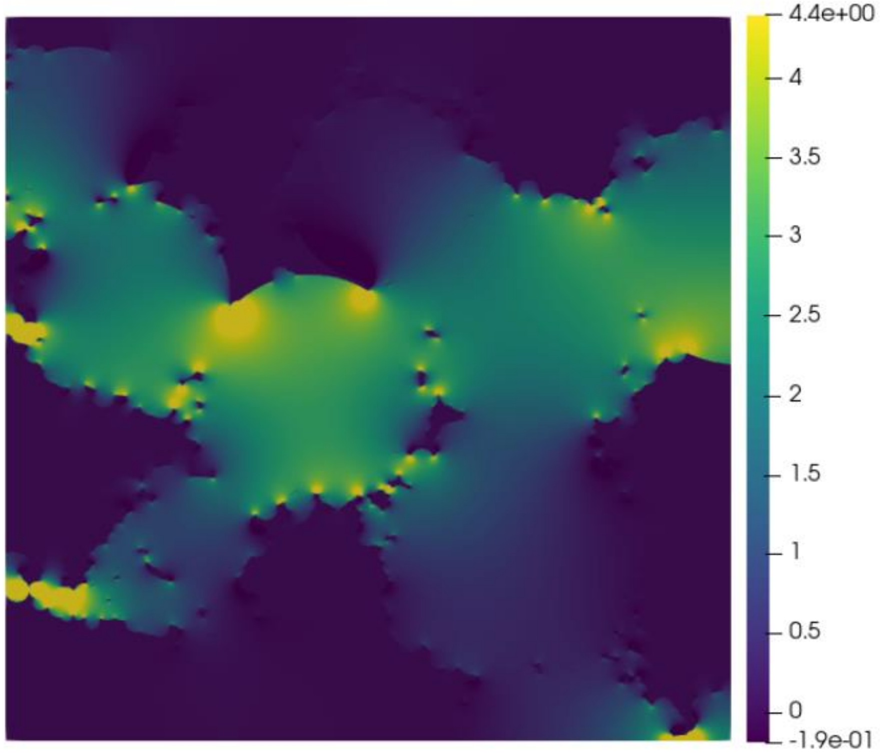


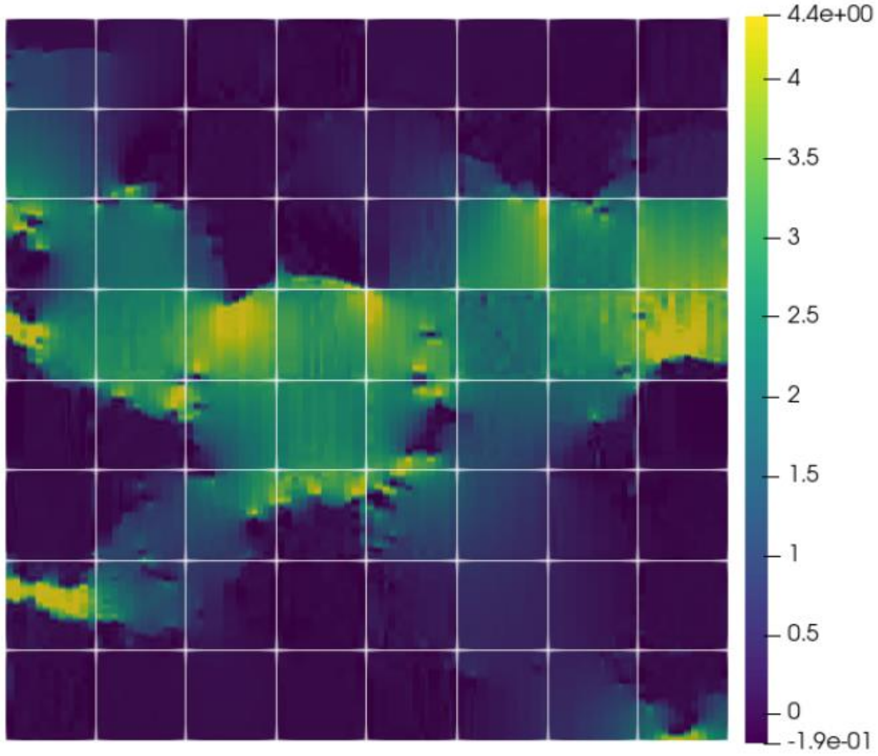




$$\mathbf{J}_\phi - \nabla\phi = \mathbf{0}$$
$$\mathbf{J}_n + \nabla n - n\nabla\phi = \mathbf{0}$$
$$\mathbf{J}_p + \nabla p - p\nabla\phi = \mathbf{0}$$
$$-\nabla\cdot\mathbf{J}_\phi + n - p = f_\phi$$
$$-\nabla\cdot\mathbf{J}_n + R(n,p) = f_n$$
$$-\nabla\cdot\mathbf{J}_p - R(n,p) = f_p$$

Scaling up: data-driven mortar methods





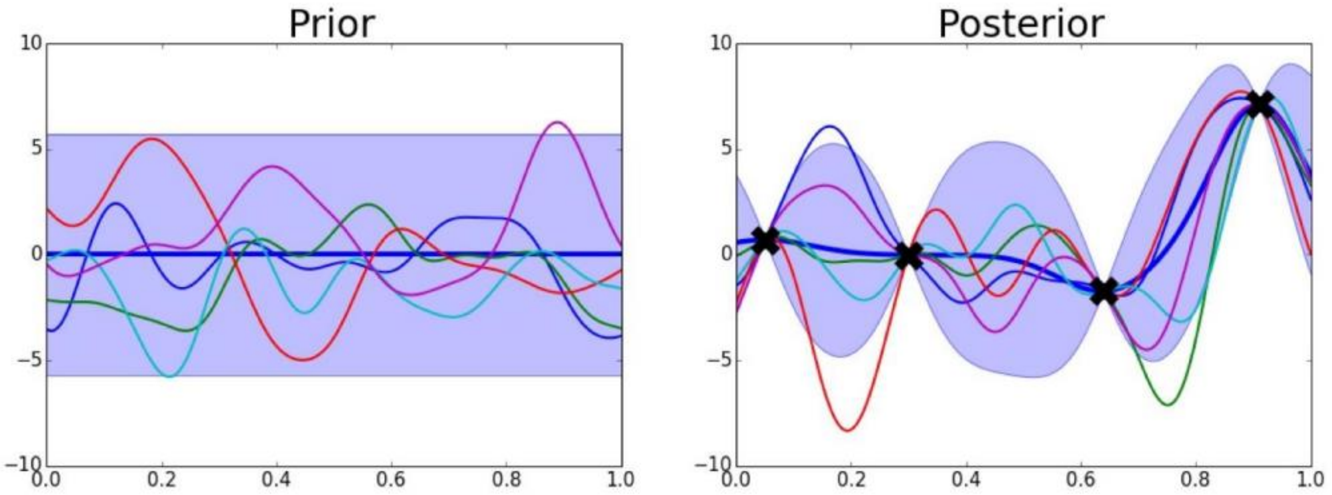
To appear

The optimal recovery problem for uncertain physics

How do we account for model form uncertainty

Mathematical preliminaries: Gaussian processes

$$y = f(\boldsymbol{x}) + \epsilon \qquad f \sim \mathcal{GP}(0, k(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\theta}))$$



Classical kriging process:

1. Assume an RBF kernel for spatial covariance
2. Perform MLE to scale kernel to data
3. Condition on available data to obtain closed form posterior expression

Schur complement gives posterior

Training via maximizing the marginal likelihood

$$\log p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\theta}) = -\frac{1}{2}\log|\boldsymbol{K} + \sigma_\epsilon^2 \boldsymbol{I}| - \frac{1}{2}\boldsymbol{y}^T(\boldsymbol{K} + \sigma_\epsilon^2 \boldsymbol{I})^{-1}\boldsymbol{y} - \frac{N}{2}\log 2\pi$$

Prediction via conditioning on available data

$$p(f_*|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{x}_*) = \mathcal{N}(f_*|\mu_*, \sigma_*^2),$$
$$\mu_*(\boldsymbol{x}_*) = \boldsymbol{k}_{*N}(\boldsymbol{K} + \sigma_\epsilon^2 \boldsymbol{I})^{-1}\boldsymbol{y},$$
$$\sigma_*^2(\boldsymbol{x}_*) = \boldsymbol{k}_{**} - \boldsymbol{k}_{*N}(\boldsymbol{K} + \sigma_\epsilon^2 \boldsymbol{I})^{-1}\boldsymbol{k}_{N*},$$

$$f(x) \sim GP(0, \kappa(x, x'; \theta))$$

$$\begin{bmatrix} f(x) \\ f(x') \end{bmatrix} \sim GP\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \kappa(x,x) & \kappa(x,x') \\ \kappa(x',x) & \kappa(x',x') \end{bmatrix}\right)$$

$$p(f(x^*|x,y) = \mathcal{N}\left(\underbrace{\kappa(x^*,x)K^{-1}y}_{\mu(x^*)}, \underbrace{\kappa(x^*,x^*) - \kappa(x^*,x)K^{-1}\kappa(x,x^*)}_{\Sigma(x^*,x^*)}\right)$$

Mathematical preliminaries: Optimal recovery problem

Recast Kriging as an optimal recovery for interpolant with minimal RKHS norm

Let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a symmetric positive definite bivariate kernel, and let $\mathcal{H}_K = \text{span}\{K(\cdot, x_i)\}$ be the induced RKHS space with accompanying RKHS norm $\|\cdot\|_K$. The **optimal recovery problem** consists of finding $f \in \mathcal{H}_K : \mathbf{X} \subset \mathcal{X} \mapsto \mathbf{Y} \subset \mathbb{R}$.

$$\min_{f \in \mathcal{H}_K} \|f\|_K^2 + \frac{1}{\epsilon} \|f(\mathbf{X}) - \mathbf{Y}\|_2^2 .$$

Expanding in terms of linear algebra

$$V^* = \arg\min_{V \in \mathbf{R}^N} V^T K(\mathbf{X}, \mathbf{X}) V + \frac{1}{\epsilon} \|K(\mathbf{X}, \mathbf{X}) V - \mathbf{Y}\|_2^2 .$$

We recover in expectation the traditional Gaussian process posterior

$$f^*(\cdot) = K(\cdot\, , \mathbf{X}) (\mathbf{K} + \epsilon I)^{-1} \mathbf{Y} .$$

Computational graph completion

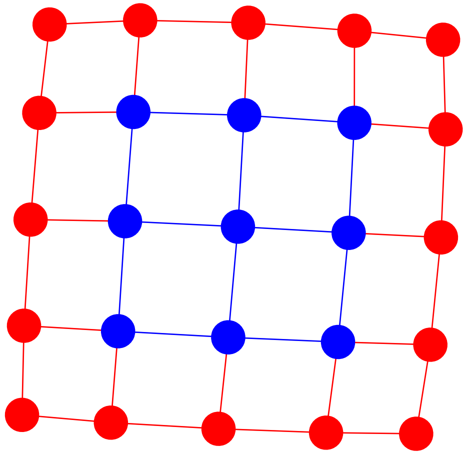
$$\min_{\theta, \mathbf{u}_{\text{un}}} \min_{\mathbf{F}_{\text{un}}} \sum_{e \in \mathcal{E}} \mathbf{F}_e^T (K_e(\delta_0 \mathbf{u}_e, \delta_0 \mathbf{u}_e) + \epsilon I)^{-1} \mathbf{F}_e + \log\det(K_e(\delta_0 \mathbf{u}_e, \delta_0 \mathbf{u}_e) + \epsilon I)$$

$$\text{s.t. } \mathbf{d}_k^\intercal \mathbf{F} = 0$$

IDEA:

- Introduce slack variables for unknown variables
  - Produce GP for every edge
  - Couple through conservation law
- Extract set of GPs with closed form posteriors

$$\mathcal{G} = (\mathcal{V}, \mathcal{E})$$
$$\mathcal{V} = \mathcal{V}_{\text{un}} \cup \mathcal{V}_{\text{obs}}$$
$$\mathcal{E} = \mathcal{E}_{\text{un}} \cup \mathcal{E}_{\text{obs}}$$



Red: Boundary nodes/edges
Blue: Internal nodes/edges

Owhadi, Houman. "Computational graph completion." *Research in the Mathematical Sciences* 9.2 (2022): 27.

Fast training with block coordinate descent

Note the equality constrained QP buried in the MLE problem

$$\min_{\theta,\mathbf{u}_{\text{un}}} \min_{\mathbf{F}_{\text{un}}} \sum_{e\in\mathcal{E}} \mathbf{F}_e^T (K_e(\delta_0\mathbf{u}_e,\delta_0\mathbf{u}_e)+\epsilon I)^{-1}\mathbf{F}_e + \log\det(K_e(\delta_0\mathbf{u}_e,\delta_0\mathbf{u}_e)+\epsilon I)$$

$$\text{s.t.} \qquad \mathbf{F}\delta_0 = 0$$

Solving KKT system allows gradient descent training on slack variables only

$$\min_{F;\lambda} F^T \widehat{K} F + \lambda^T (\widehat{D_0}^T F - b)$$

$$\begin{bmatrix} \widehat{K} & \widehat{D_0} \\ \widehat{D_0}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} F \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix},$$

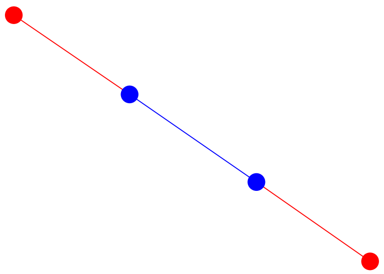
$$F = \text{vec}(\mathbf{F}_{\text{un}})$$

$$\widehat{K} = \text{diag}(K_e(\delta_0\mathbf{u}_e,\delta_0\mathbf{u}_e)+\epsilon I)^{-1}_{e\in\mathcal{E}_u}$$

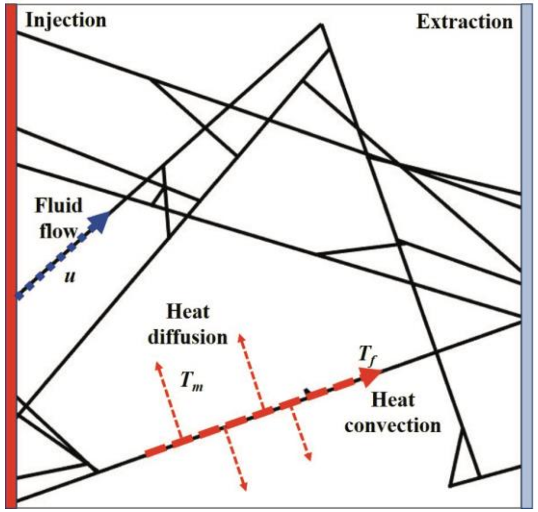
$$D_0 = \delta_{0\,\mathcal{E}_{\text{un}},\mathcal{V}_{\text{un}}}$$

$$\min_{\theta,\mathbf{u}_{\text{un}}} \sum_{e\in\mathcal{E}_{\text{un}}} \mathbf{b}_e^T \left(K_e(\delta_0\mathbf{u}_e,\delta_0\mathbf{u}_e)+\epsilon I\right)^{-1}\mathbf{b}_e + \sum_{e\in\mathcal{E}_{\text{obs}}} \mathbf{F}_e^T \left(K_e(\delta_0\mathbf{u}_e,\delta_0\mathbf{u}_e)+\epsilon I\right)^{-1}\mathbf{F}_e + \sum_{e\in\mathcal{E}} \log\det\left(K_e(\delta_0\mathbf{u}_e,\delta_0\mathbf{u}_e)+\epsilon I\right).$$

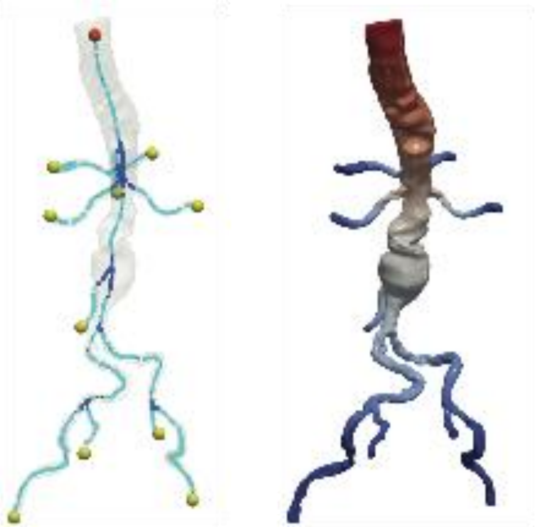
Examples: probabilistic circuit discovery



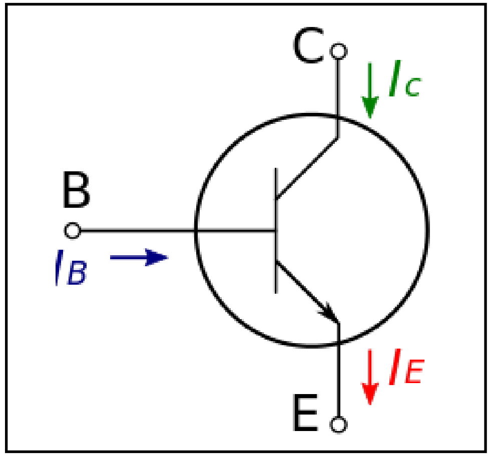
toy circuit



subsurface discrete fracture network[1] (linear)



arterial flow[2] (highly nonlinear)

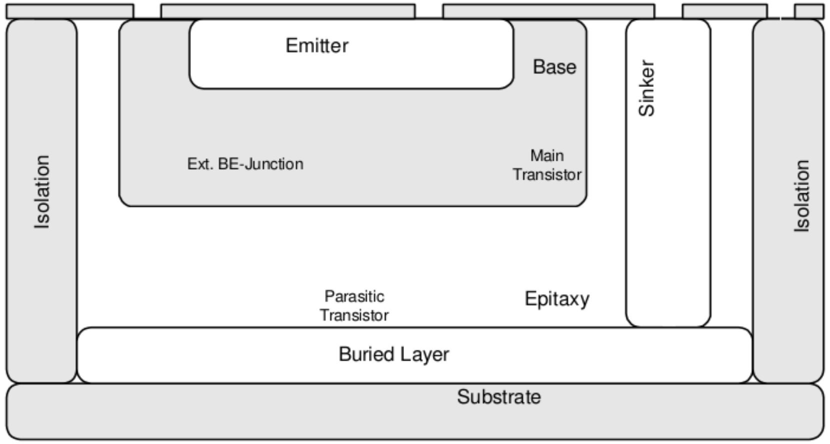


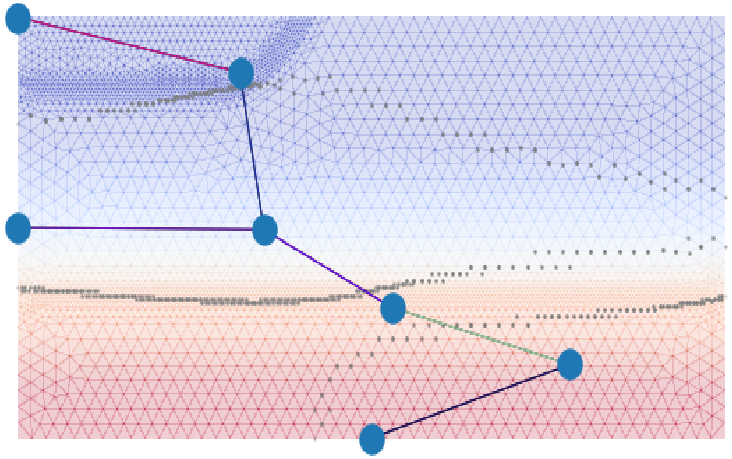
Bipolar junction transistor[3] (nonlinear & multiscale)

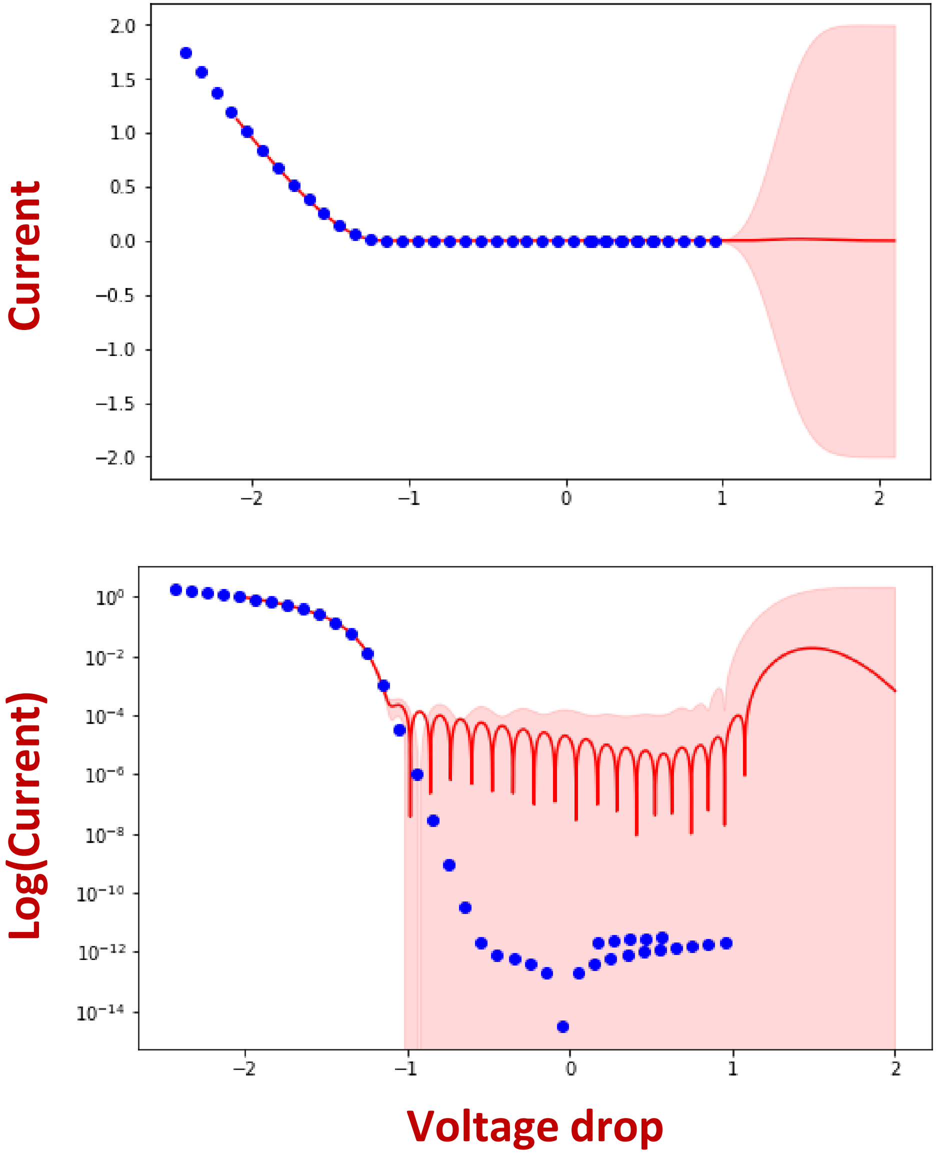
1. Song et al. "Surrogate models of heat transfer in fractured rock and their use in parameter estimation" (in review)
2. Pegolotti et al. "Learning Reduced-Order Models for Cardiovascular Simulations with Graph Neural Networks" (arXiv preprint)
3. Generated by Paul Kuberry (Sandia National Laboratories, 01442 Computational Mathematics)

Fast training with block coordinate descent

Epistemic uncertainty quantification in exponential regime







Structure preservation when discovering bracket dynamics

Moving toward dynamical systems and physics-inspired deep learning architectures

Beyond boundary value problems

How can we learn dynamical systems with structure preservation when the governing equations are unknown?

"Black box" – no required model | Structure-preserving ML | Strong physical priors

Neural ODE (NODE)

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 6572–6583, 2018.

Hamiltonian NN

Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Symplectic RNN

Zhengdao Chen, Jianyu Zhang, Martin Arjovsky, and Léon Bottou. Symplectic recurrent neural networks. In *International Conference on Learning Representations*, 2019.

SympNets

Pengzhan Jin, Zhen Zhang, Aiqing Zhu, Yifa Tang, and George Em Karniadakis. Sympnets: Intrinsic structure-preserving symplectic networks for identifying hamiltonian systems. *Neural Networks*, 132:166– 179, 2020.

Lagrangian NN

Miles Cranmer, Sam Greydanus, Stephan Hoyer, Peter Battaglia, David Spergel, and Shirley Ho. La- grangian neural networks. In *ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations*, 2020.

Universal DiffEq (UDE)

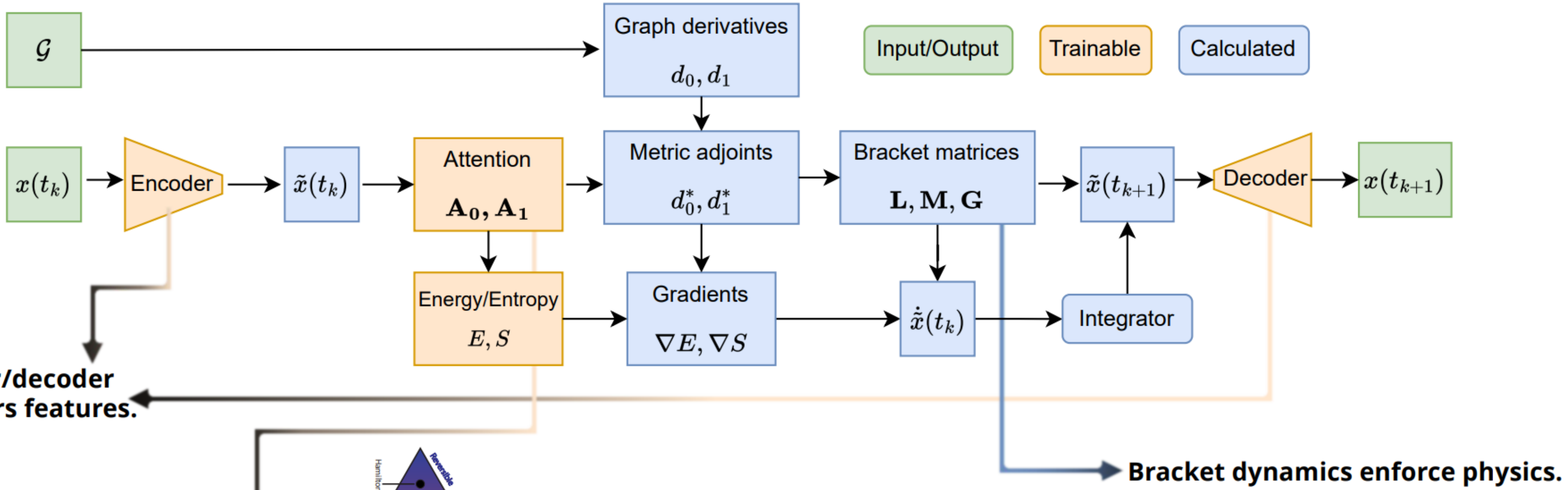
Christopher Rackauckas, Yingbo Ma, Julius Martensen, Collin Warner, Kirill Zubov, Rohit Supekar, Dominic Skinner, Ali Ramadhan, and Alan Edelman. Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*, 2020.

Dictionary (e.g SINDy)

Brunton, Steven L., Joshua L. Proctor, and J. Nathan Kutz. "Discovering governing equations from data by sparse identification of nonlinear dynamical systems." *Proceedings of the national academy of sciences* 113.15 (2016): 3932-3937.

Reversible Systems Only!

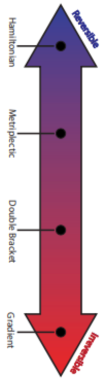
Structure preserving bracket dynamics



Encoder/decoder discovers features.

Inner products capture metric information.

Extends graph attention to include higher-order graph information.



Formalism	Equation	Requirements	Completeness	Character
Hamiltonian	$\dot{\mathbf{x}} = \{\mathbf{x}, E\}$	$\mathbf{L}^* = -\mathbf{L}$, Jacobi's identity	complete	conservative
Gradient	$\dot{\mathbf{x}} = -[\mathbf{x}, E]$	$\mathbf{M}^* = \mathbf{M}$	incomplete	totally dissipative
Double Bracket	$\dot{\mathbf{x}} = \{\mathbf{x}, E\} + \{\{\mathbf{x}, E\}\}$	$\mathbf{L}^* = -\mathbf{L}$	incomplete	partially dissipative
Metriplectic	$\dot{\mathbf{x}} = \{\mathbf{x}, E\} + [\mathbf{x}, S]$	$\mathbf{L}^* = -\mathbf{L}, \mathbf{M}^* = \mathbf{M}$, $\mathbf{L}\nabla S = \mathbf{M}\nabla E = \mathbf{0}$	complete	partially dissipative

Gruber, Anthony, Kookjin Lee, and Nathaniel Trask. "Reversible and irreversible bracket-based dynamics for deep graph neural networks." *arXiv preprint arXiv:2305.15616* (2023).
Accepted to NeurIPS

Structure preserving bracket dynamics

(SNL, Spelman, PNNL, Brown, UPenn)

$$\frac{d\mathbf{x}}{dt} = \mathbf{L}\,\nabla_{\mathbf{x}}\mathcal{E}(\mathbf{x}) + \mathbf{M}\,\nabla_{\mathbf{x}}\mathcal{S}(\mathbf{x})$$

$$\mathbf{L} = -\mathbf{L}^{\intercal} \qquad \mathbf{M} = \mathbf{M}^{\intercal}$$

$$\mathbf{L}\,\nabla_x\mathcal{S} = \mathbf{M}\,\nabla_x\mathcal{E} = 0$$

$$\frac{d\mathcal{E}}{dt} = 0 \qquad \frac{d\mathcal{S}}{dt} \geq 0$$

Classically, a model is derived from first principles and one notices GENERIC structure

We parameterize algebraic structure and discover dissipative model

First law of thermodynamics

$$\begin{aligned}\frac{d\mathcal{E}}{dt} &= \nabla_x\mathcal{E}^{\intercal}\frac{d\mathbf{x}}{dt}\\ &= \nabla_x\mathcal{E}^{\intercal}\left(\mathbf{L}\,\nabla_{\mathbf{x}}\mathcal{E}(\mathbf{x}) + \mathbf{M}\,\nabla_{\mathbf{x}}\mathcal{S}(\mathbf{x})\right)\\ &= \nabla_x\mathcal{E}^{\intercal}\,\mathbf{L}\,\nabla_{\mathbf{x}}\mathcal{E}(\mathbf{x}) + \nabla_x\mathcal{S}^{\intercal}\,\mathbf{M}\,\nabla_{\mathbf{x}}\mathcal{E}(\mathbf{x})\\ &= 0\end{aligned}$$

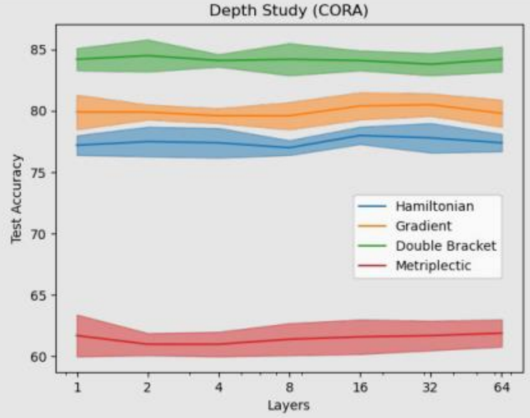
Second law of thermodynamics

$$\begin{aligned}\frac{d\mathcal{S}}{dt} &= \nabla_x\mathcal{S}^{\intercal}\frac{d\mathbf{x}}{dt}\\ &= \nabla_x\mathcal{S}^{\intercal}\left(\mathbf{L}\,\nabla_{\mathbf{x}}\mathcal{E}(\mathbf{x}) + \mathbf{M}\,\nabla_{\mathbf{x}}\mathcal{S}(\mathbf{x})\right)\\ &= -\nabla_x\mathcal{E}^{\intercal}\,\mathbf{L}\,\nabla_{\mathbf{x}}\mathcal{S}(\mathbf{x}) + \nabla_x\mathcal{S}^{\intercal}\,\mathbf{M}\,\nabla_{\mathbf{x}}\mathcal{S}(\mathbf{x})\\ &\geq 0\end{aligned}$$

How to enforce null-space condition on L and M?
Exploit exact sequence property of data-driven exterior calculus!

Science4AI and AI4Science

(SNL, Spelman, PNNL, Brown, UPenn)



Stable with increasing depth!

	CORA	CiteSeer	PubMed	CoauthorCS	Computer	Photo	HalfCheetah	Hopper	Swimmer
GAT	81.8 ± 1.3	71.4 ± 1.9	78.7 ± 2.3	90.5 ± 0.6	78.0 ± 19.0	85.7 ± 20.3	-	-	-
GDE	78.7 ± 2.2	71.8 ± 1.1	73.9 ± 3.7	91.6 ± 0.1	82.9 ± 0.6	92.4 ± 2.0	-	-	-
GRAND-nl	82.3 ± 1.6	70.9 ± 1.0	77.5 ± 1.8	92.4 ± 0.3	82.4 ± 2.1	92.4 ± 0.8	-	-	-
NODE+AE	-	-	-	-	-	-	0.106 ± 0.0011	0.0780 ± 0.0021	0.0297 ± 0.0036
Hamiltonian	76.2 ± 2.1	72.2 ± 1.9	76.8 ± 1.1	92.0 ± 0.2	84.0 ± 1.0	91.8 ± 0.2	0.0566 ± 0.013	0.0279 ± 0.0019	0.0122 ± 0.00044
Gradient	81.3 ± 1.2	72.1 ± 1.7	77.2 ± 2.1	92.2 ± 0.3	78.1 ± 1.2	88.2 ± 0.6	0.105 ± 0.0076	0.0848 ± 0.0011	0.0290 ± 0.0011
Double Bracket	83.0 ± 1.1	74.2 ± 2.5	78.2 ± 2.0	92.5 ± 0.2	84.8 ± 0.5	92.4 ± 0.3	0.0621 ± 0.0096	0.0297 ± 0.0048	0.0128 ± 0.00070
Metriplectic	59.6 ± 2.0	63.1 ± 2.4	69.8 ± 2.1	-	-	-	0.105 ± 0.0091	0.0398 ± 0.0057	0.0179 ± 0.00059

Graph analytics problems
(higher is better)

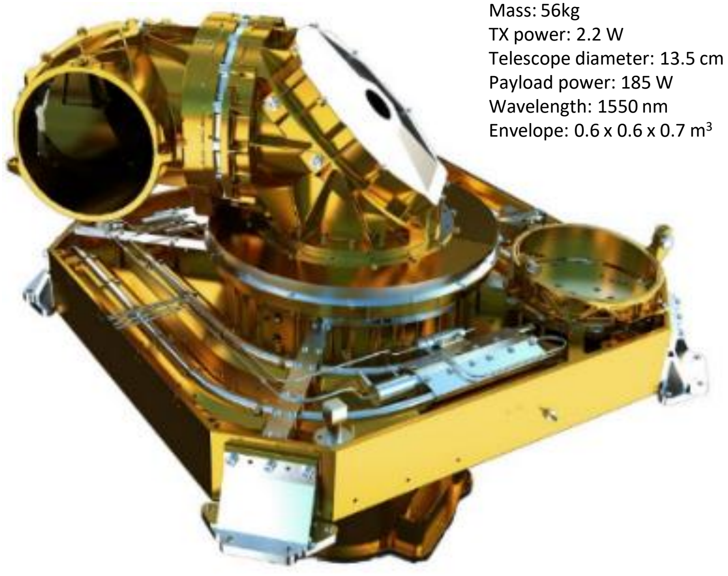
Physics simulation problems
(lower is better)

$$d\boldsymbol{x}_t = \left(L\frac{\partial E}{\partial \boldsymbol{x}} + M\frac{\partial S}{\partial \boldsymbol{x}} + k_B \frac{\partial}{\partial \boldsymbol{x}} \cdot M \right) dt + \sqrt{2 k_B M}\, dW_t$$

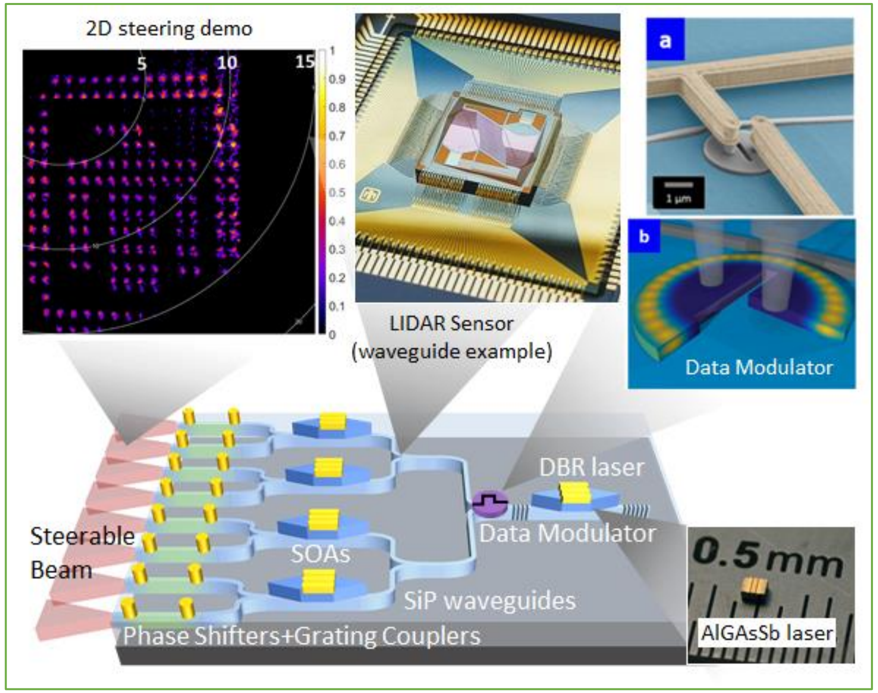
Reversible — Irreversible dissipation — Thermal noise

Key result: First O(N) method providing structure preserving dynamics!

Conclusions: Toward robust data-driven multiscale digital twins



Laser Communication Terminal (LCT)
- Bandwidth: 1.8 Gbps
- Range: 45,000 km
- Mass: 56kg
- TX power: 2.2 W
- Telescope diameter: 13.5 cm
- Payload power: 185 W
- Wavelength: 1550 nm
- Envelope: 0.6 x 0.6 x 0.7 m^3



<u>Big device O(1m^3)</u>
Individual systems
Large motor to point and steer laser

<u>Small device O(1cm^3)</u>
Integrated multiphysics
No mechanical system
Challenge: cross device coupling