Pacific Northwest NATIONAL LABORATORY

Smart Grid Edge Analytics Workshop
Georgia Tech Global Learning Center
June 4-5, 2019, Atlanta, GA, USA

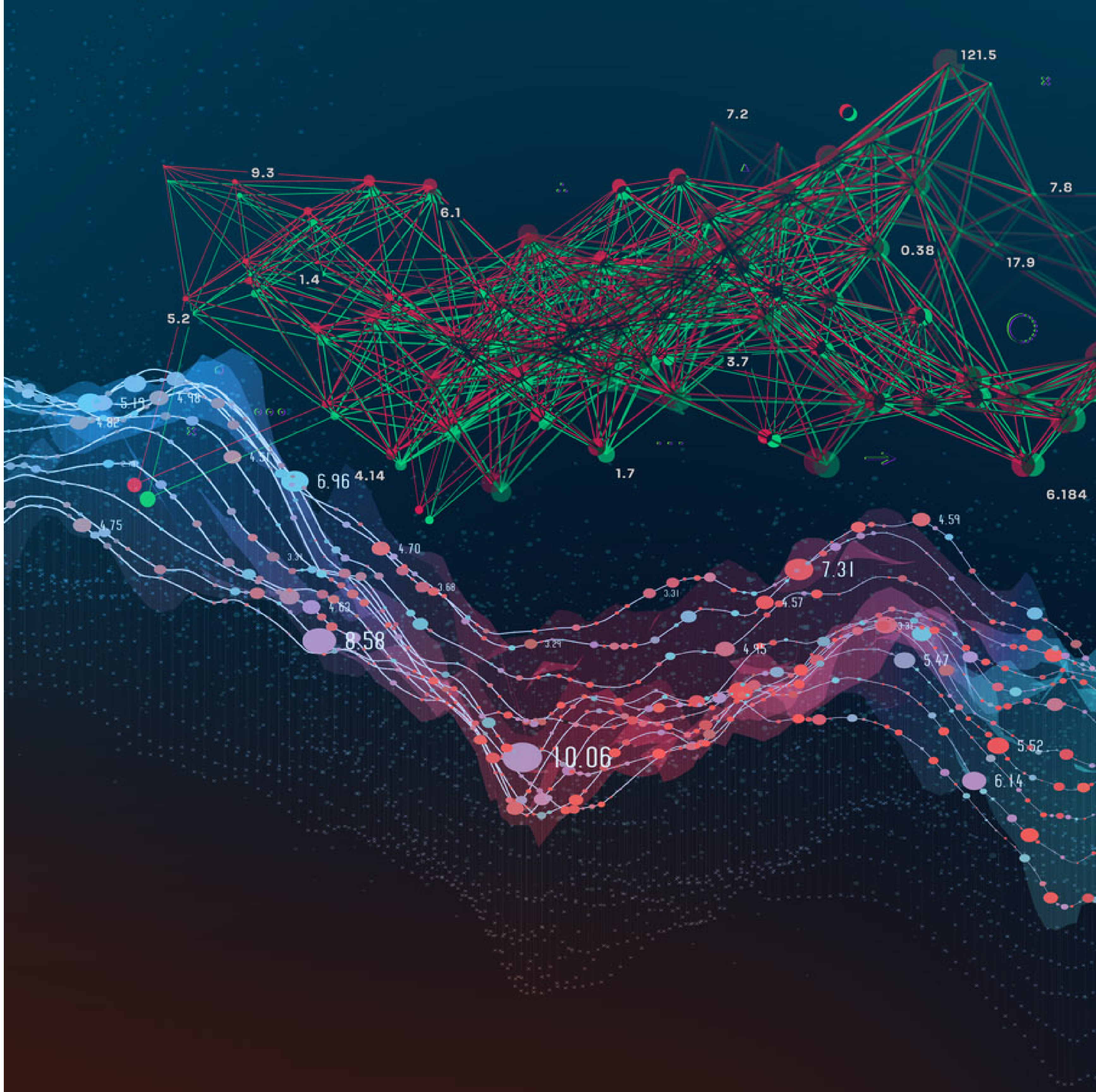
Domain-aware Machine Learning

Robert Rallo

Advanced Computing, Mathematics, and Data Division

U.S. DEPARTMENT OF ENERGY BATTELLE

PNNL is operated by Battelle for the U.S. Department of Energy



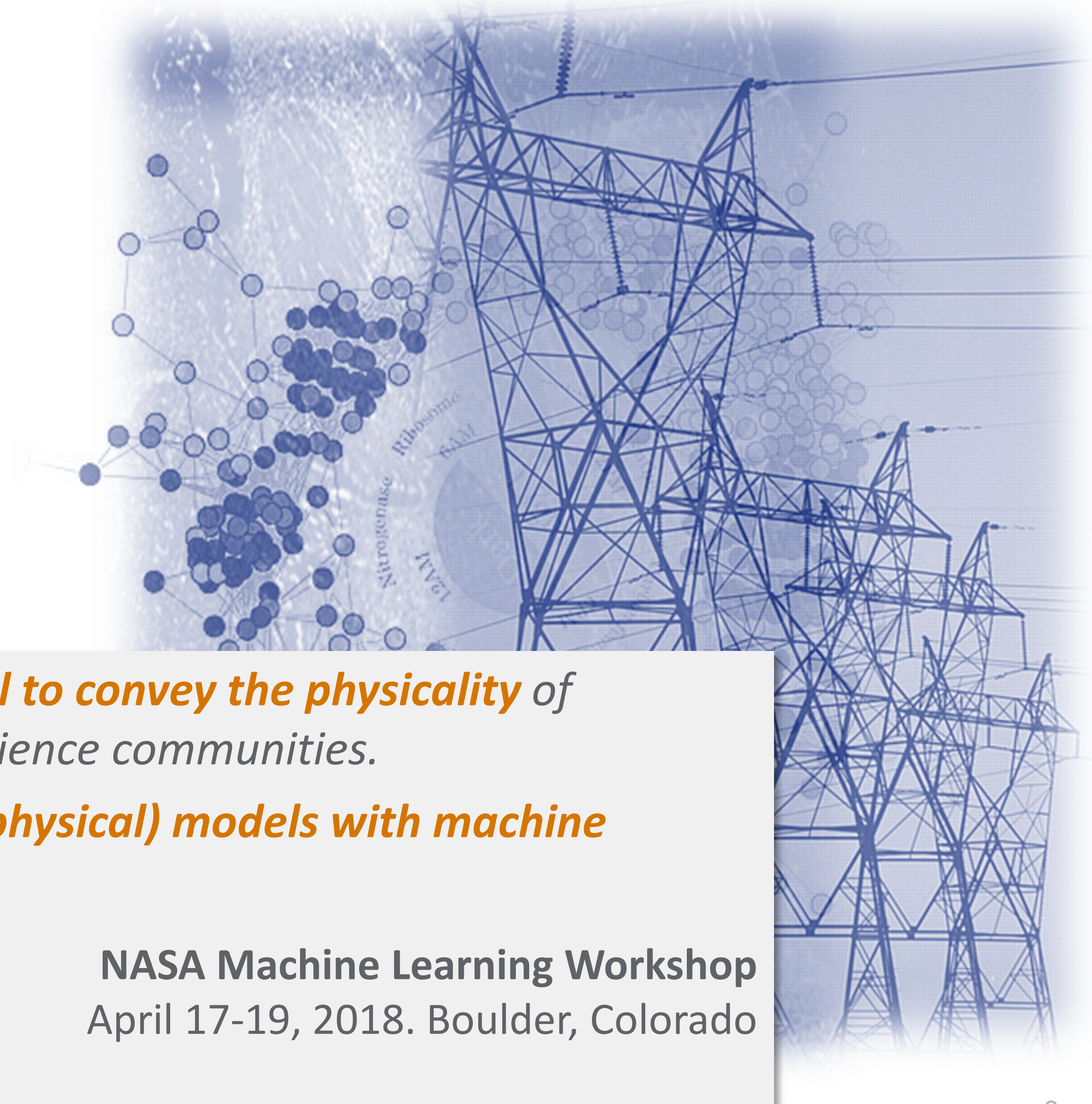
Machine Learning: challenges

- Guiding principles
  - What is the best approach for a given task?
  - What are the optimal model parameters?
- Interpretability
  - How to extract cause-effect relationships from complex non-linear models?
- Scalability
  - What is the optimal way of leveraging heterogeneous distributed computing resources?
- Validity
  - What are the best performance metrics to capture generalization power?
  - How to quantify uncertainty without increasing the computational complexity?

“Black Cube” John McCraken, 1971
Portland Art Museum

From modeling and simulation to big data analytics and machine learning

- High-fidelity models and simulations
  - Computational complexity
- Big data analytics
  - Require significant amounts of data
- Machine learning
  - Computational complexity
  - Large datasets of “acceptable quality”



*Regression and machine learning techniques generally **fail to convey the physicality** of processes being modeled and **lack acceptance** by some science communities.*

*Techniques and tools are needed for **combining process (physical) models with machine learning** models in a meaningful way.*

NASA Machine Learning Workshop
April 17-19, 2018. Boulder, Colorado

Scientific Machine Learning

- Machine Learning algorithms should be:
  - **Domain-aware**
  - Interpretable
  - Robust
  - Data-intensive
- Machine Learning must contribute to:
  - Enhance current modeling and simulation
  - Automation and decision support

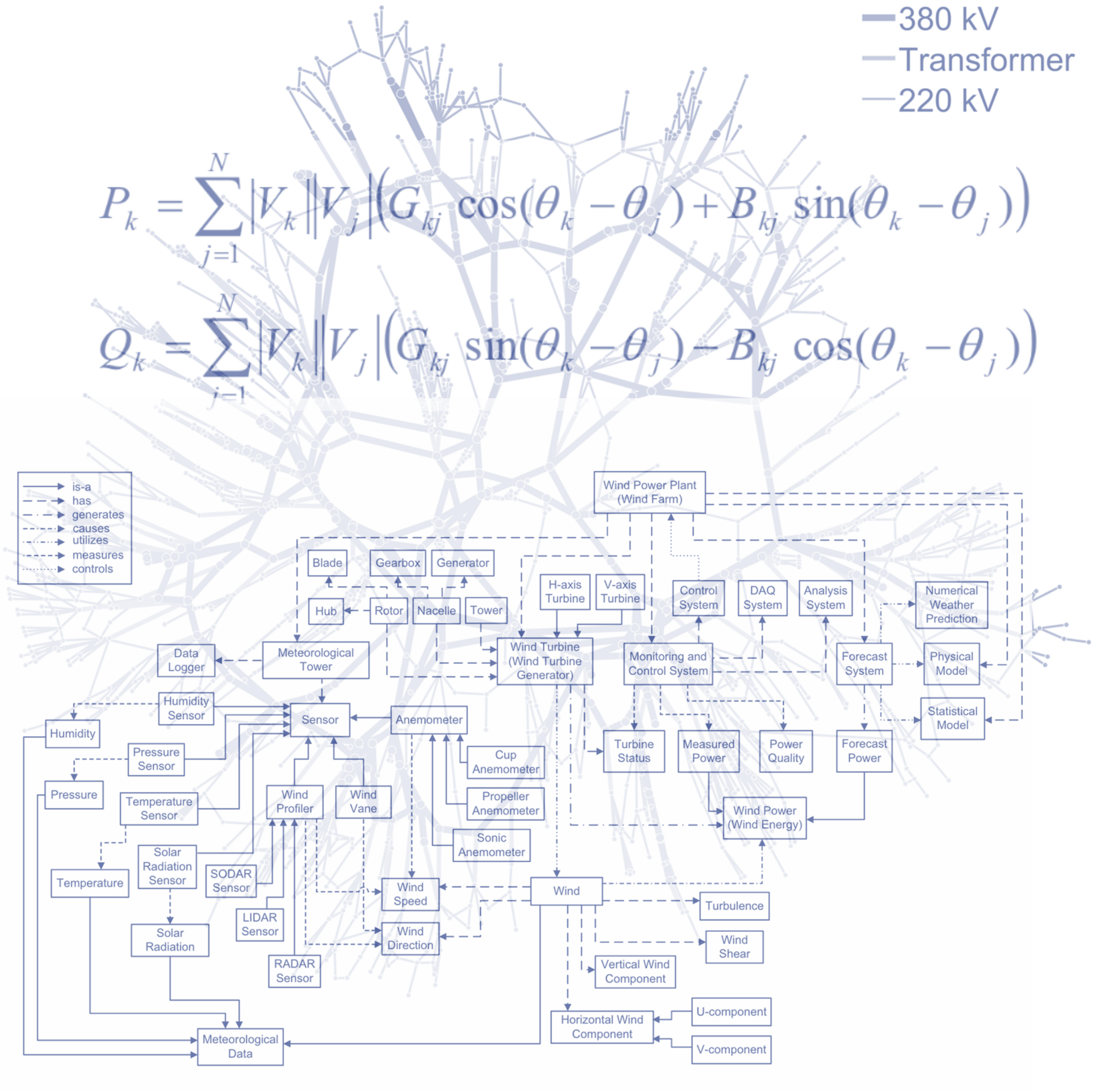
DATA (Features) → MODELS (Algorithms) ← SCIENTIFIC KNOWLEDGE

Pacific Northwest
NATIONAL LABORATORY

Domain-aware Machine Learning

- Embedding domain knowledge
  - Knowledge representations for ML
  - Integration of domain knowledge
  - Data-driven scientific discovery

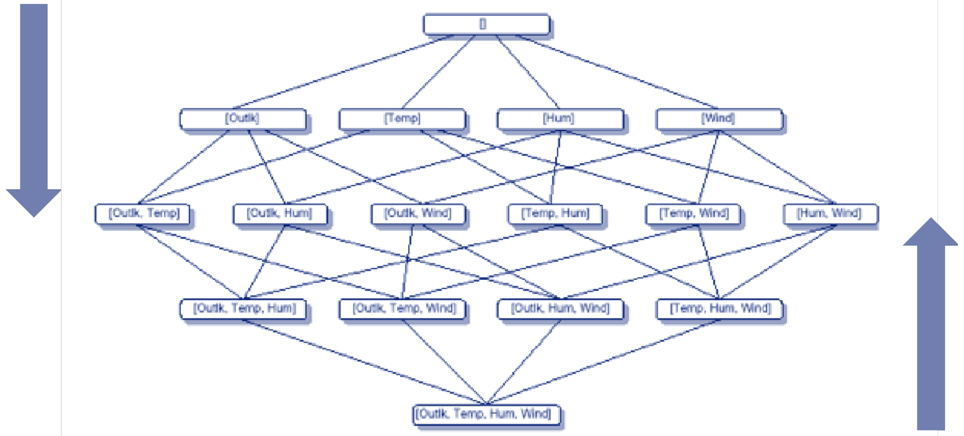
- Facilitate accelerated learning
  - Methods to accelerate the convergence and stability of ML algorithms when (labeled) data are limited
  - Tools to speed-up the tuning and optimization of domain-aware ML models



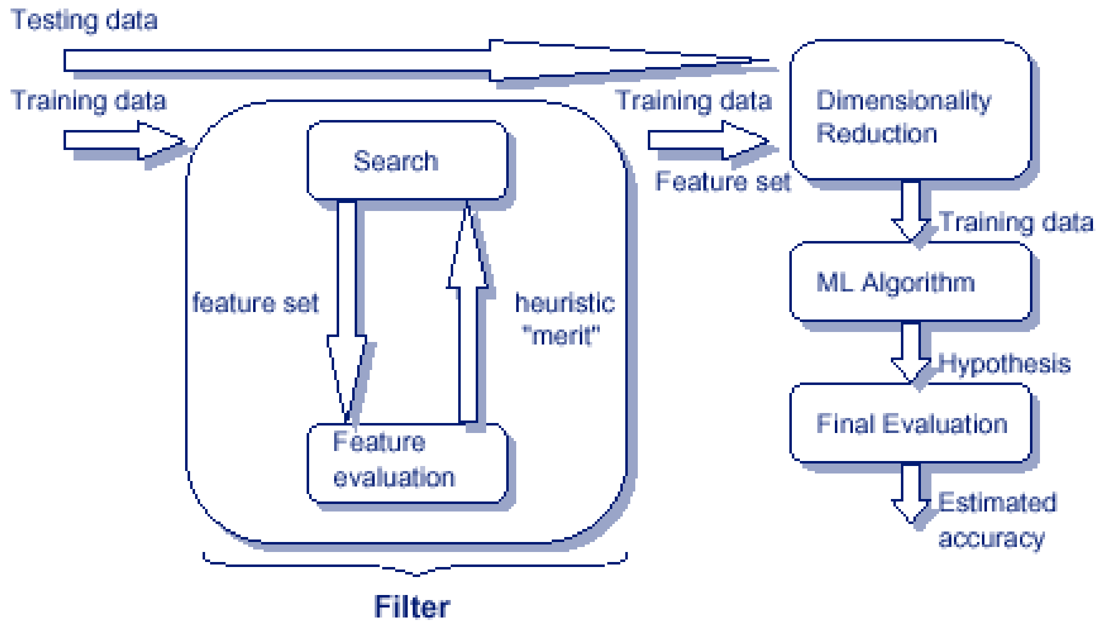
Data (Features)

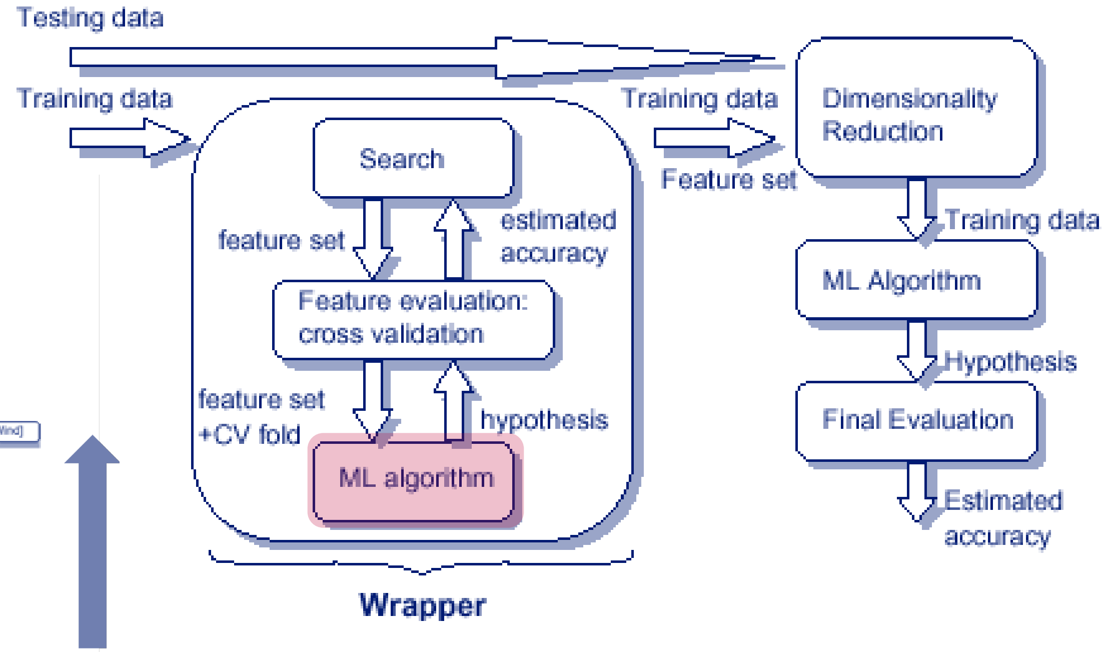
- Features must be:
  - Representative
    - ✓ Capture relevant information in data
  - Interpretable
    - ✓ Recognizable by human experts
  - Generalizable
    - ✓ Same results using different ML techniques
- Feature Selection
  - Expert criteria
  - Heuristic search
  - Filters
  - Wrappers

forward selection



Backward elimination





Pacific Northwest NATIONAL LABORATORY

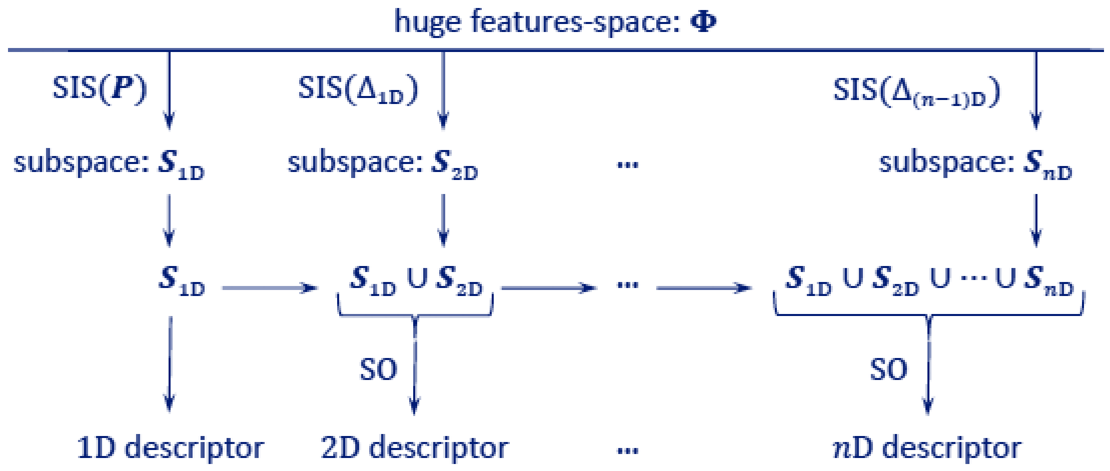
Feature engineering

SISSO (“Sure Independence Screening and Sparsifying Operator”)

- Generation of physically interpretable descriptors
- Based on compressing-sensing principles

- **Algorithm:**
  - Starting feature space: readily-available (and relevant properties): $\{\boldsymbol{\Phi}_0\}$
  - Operator set: $\hat{H}^{(m)} \equiv \{I, +, -, \times, exp, log, |-|, \sqrt{\ }, \ {}^{-1}, \ {}^{2}, \ {}^{3}\}[\phi_1, \phi_2]$
    - ✓ dimensional analysis; linear and non-linear operators
  - Recursive expansion of the feature space: $\boldsymbol{\Phi}_n \equiv \bigcup_{i=1}^{n} \hat{H}^{(m)}[\phi_1, \phi_2], \forall \phi_1, \phi_2 \in \boldsymbol{\Phi}_{i-1}$
  - **SIS**: scores each standardized feature with a metric and keeps only top ranked features
  - **SO**: finds optimal n-dimensional descriptor

- **Data requirements:**
  - $\# \, samples \ \geq k \cdot n \cdot \log(\#\boldsymbol{\Phi})$, where $k \sim 1 - 10$



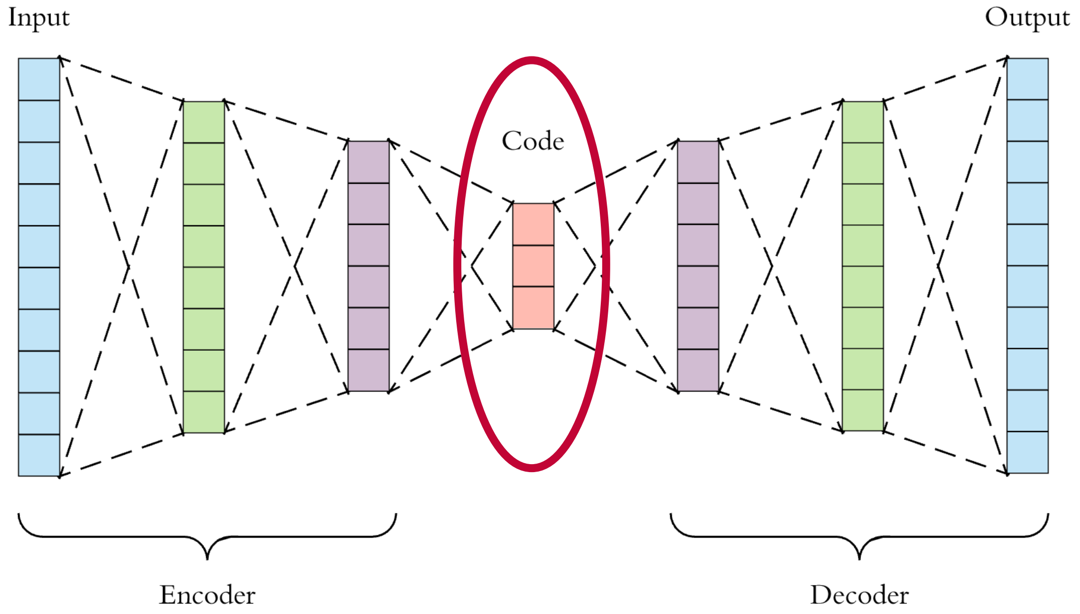
Ouyang R et al. 2017, arXiv preprint arXiv:1710.03319

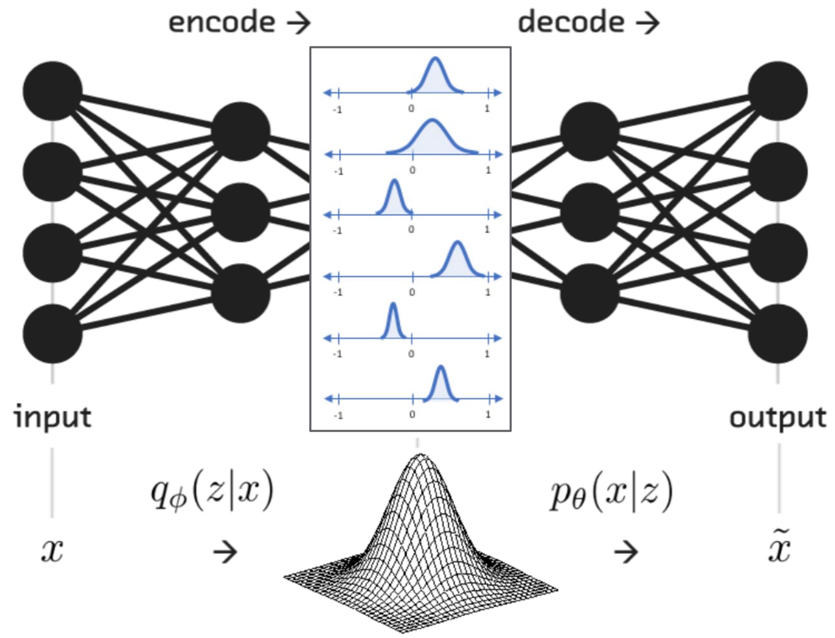
Representation Learning

- Autoencoders:
  - Unsupervised Learning
  - Bottleneck architecture
  - Compressed data representation
  - $\mathcal{L}(x, \hat{x}) + regularizer$
- Sparse autoencoders
  - Penalize node activations
  - Few units active at the same time
  - $\mathcal{L}(x, \hat{x}) + \lambda \sum_i |a_i|$
- Variational Autoencoders:
  - Generative
  - Latent gaussian models

$$\mathcal{L}(x, \hat{x}) + \sum_j KL\left(q_j(z|x) \,||\, p(z)\right)$$

assumed prior (gaussian)



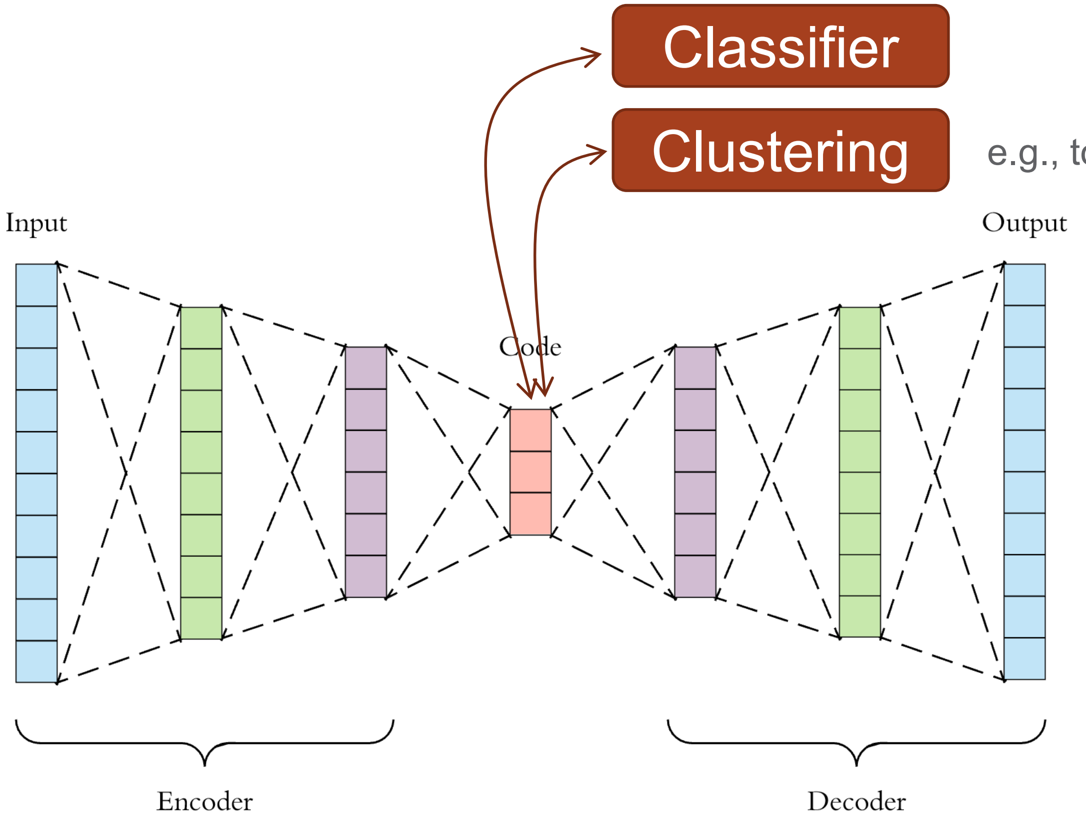


Pacific Northwest
NATIONAL LABORATORY

Domain-aware representations

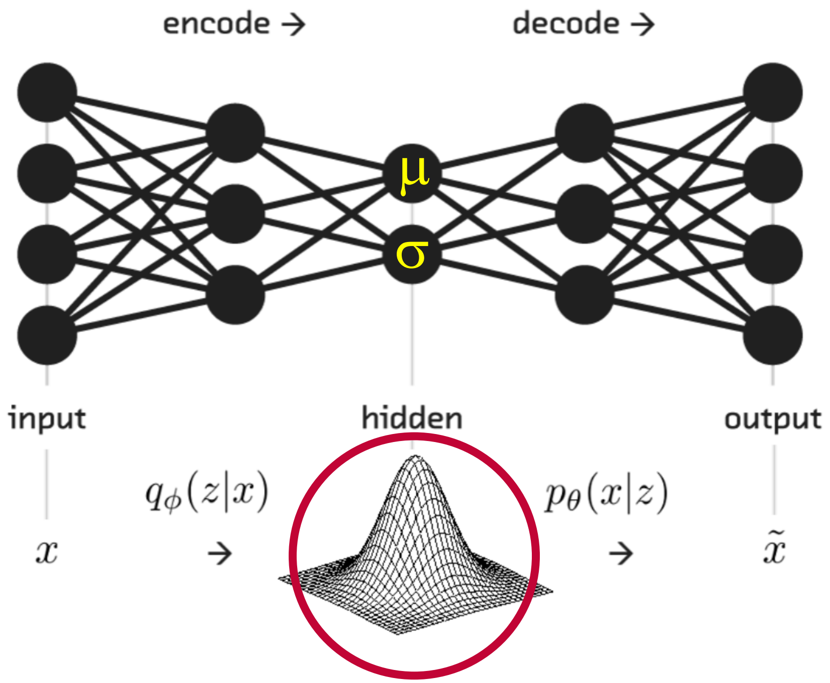
- Coupling autoencoders with:
  - Clustering algorithms
  - Classification algorithms

$$\mathcal{L}(x,\hat{x}) + \lambda_1 \sum_i |a_i| + \lambda_2 \mathcal{L}_{\substack{clustering \\ classifier}}$$



e.g., topology preservation

- Variational autoencoders with:
  - Predefined priors

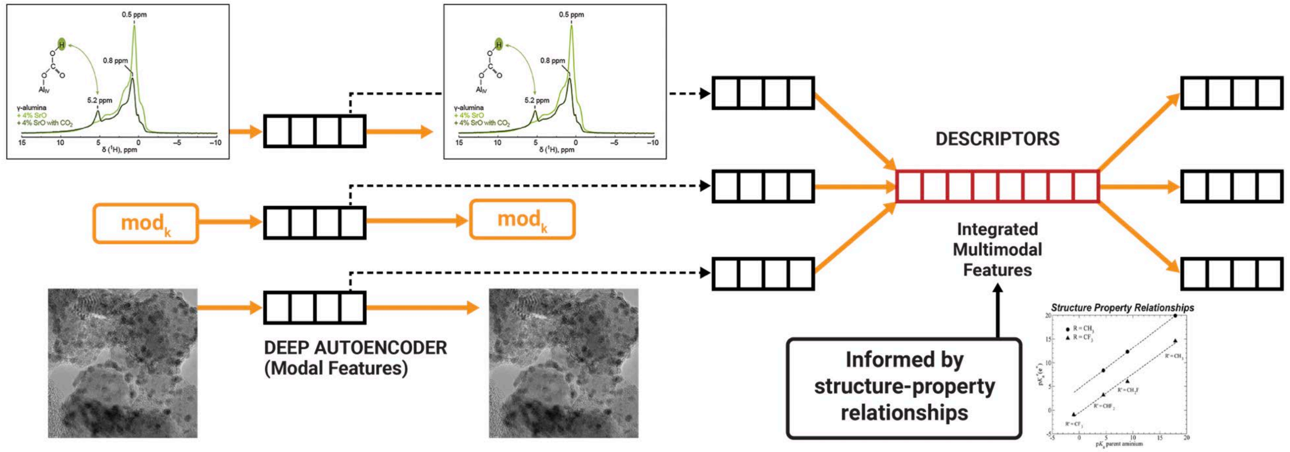


- Domain-informed prior
- KL term in the loss function enforces the desired probability distribution

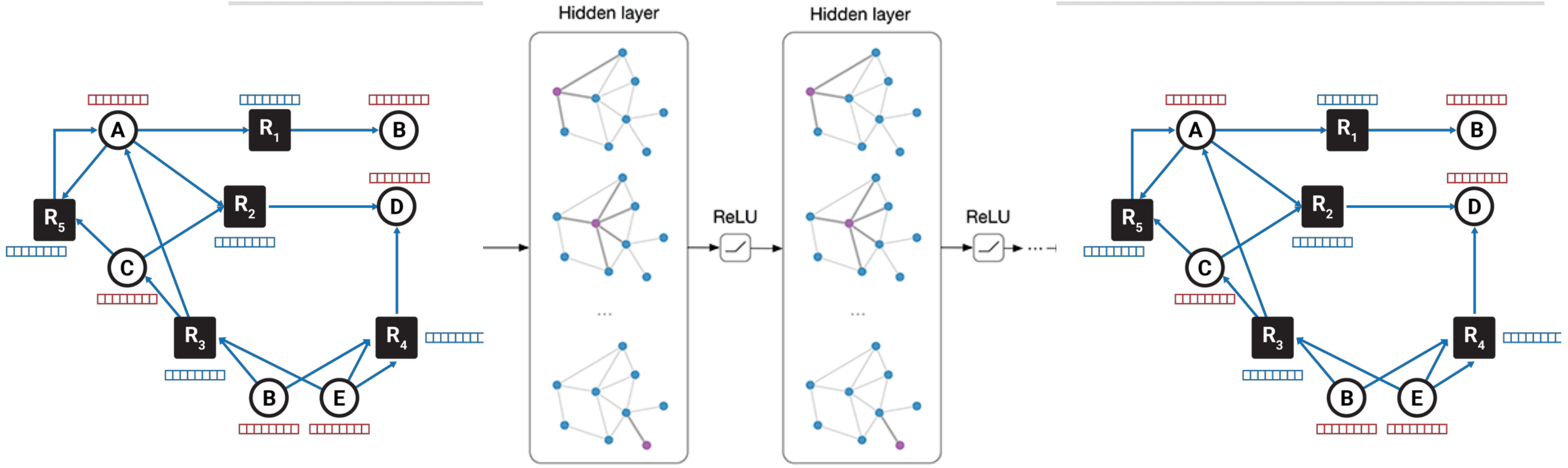
Domain-awareness in supervised ML

- Learning as a constrained optimization technique
  - Hard Constraints
    - ✓ During training
    - ✓ At the time of prediction
  - **Soft Constraints**
    - ✓ **Additional terms in the loss-function**
    - ✓ **Coupling with simulation codes**
  - Model form
    - ✓ Basis functions
    - ✓ Domain-specific kernels

Multimodal data integration



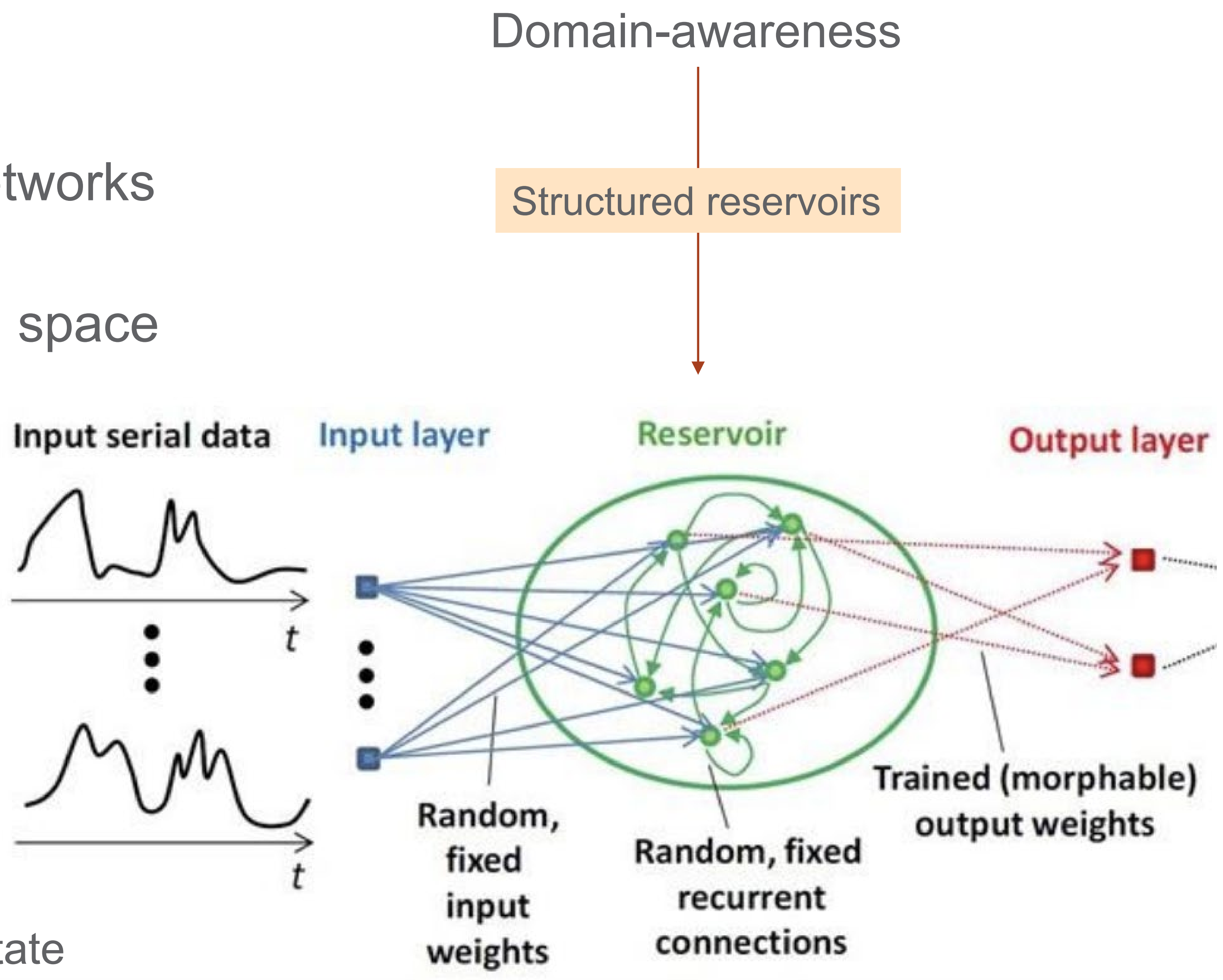
$L = L_r + L_{SPR}$



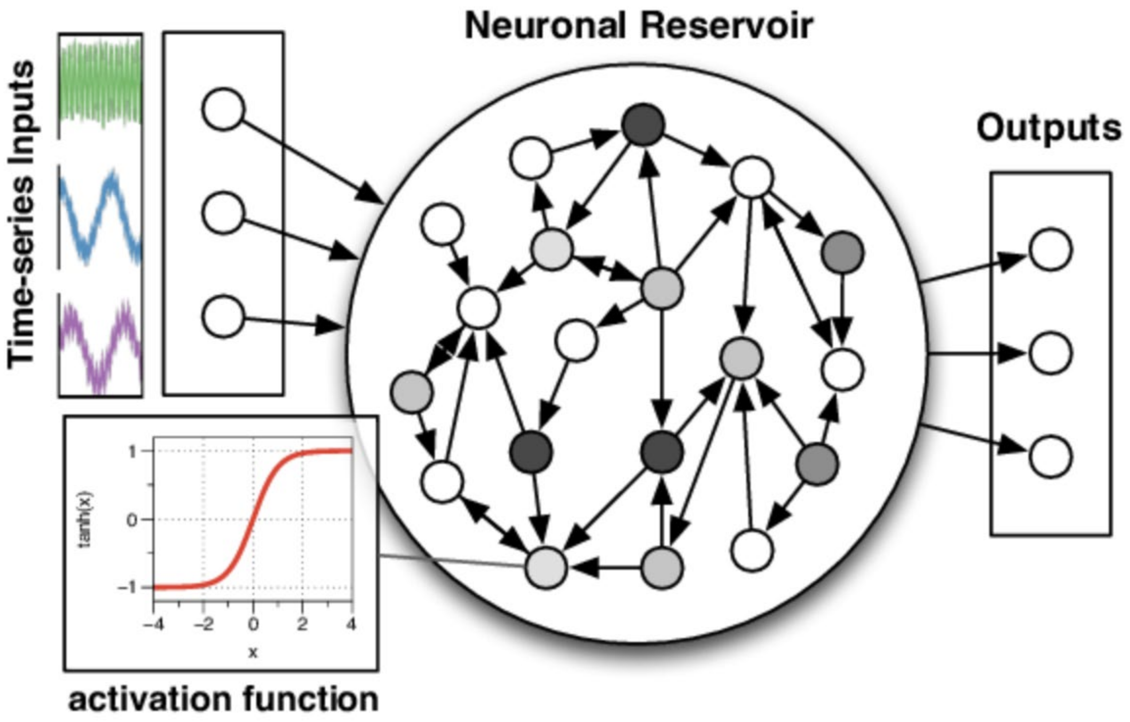
knowledge graph

Machine Learning at the Edge

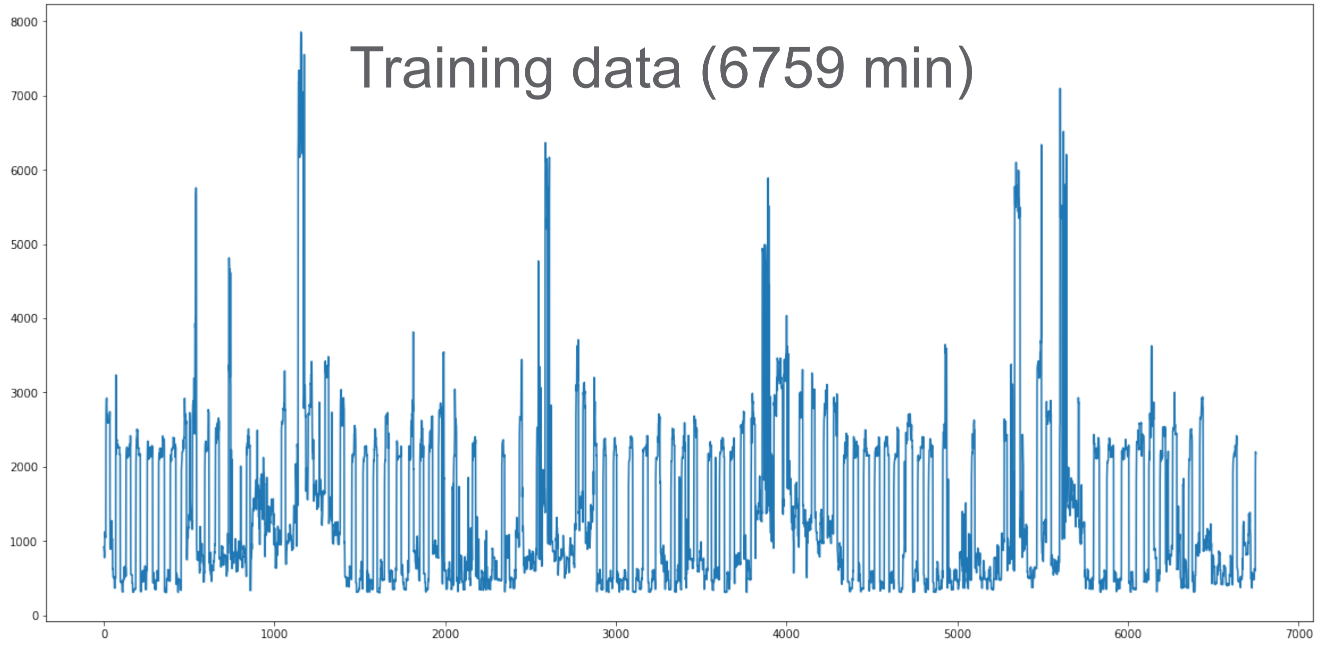
- Reservoir computing
  - Generalization of recurrent neural networks
  - Dynamical systems
  - Maps inputs onto a high-dimensional space
  - Hardware implementation
- Elements
  - Input layer
    - ✓ random weights
  - Reservoir
    - ✓ Random sparse connectivity
    - ✓ Non-linear activation
  - Readout layer
    - ✓ Linear transformation of the reservoir state
    - ✓ Fast adaptation using ridge regression

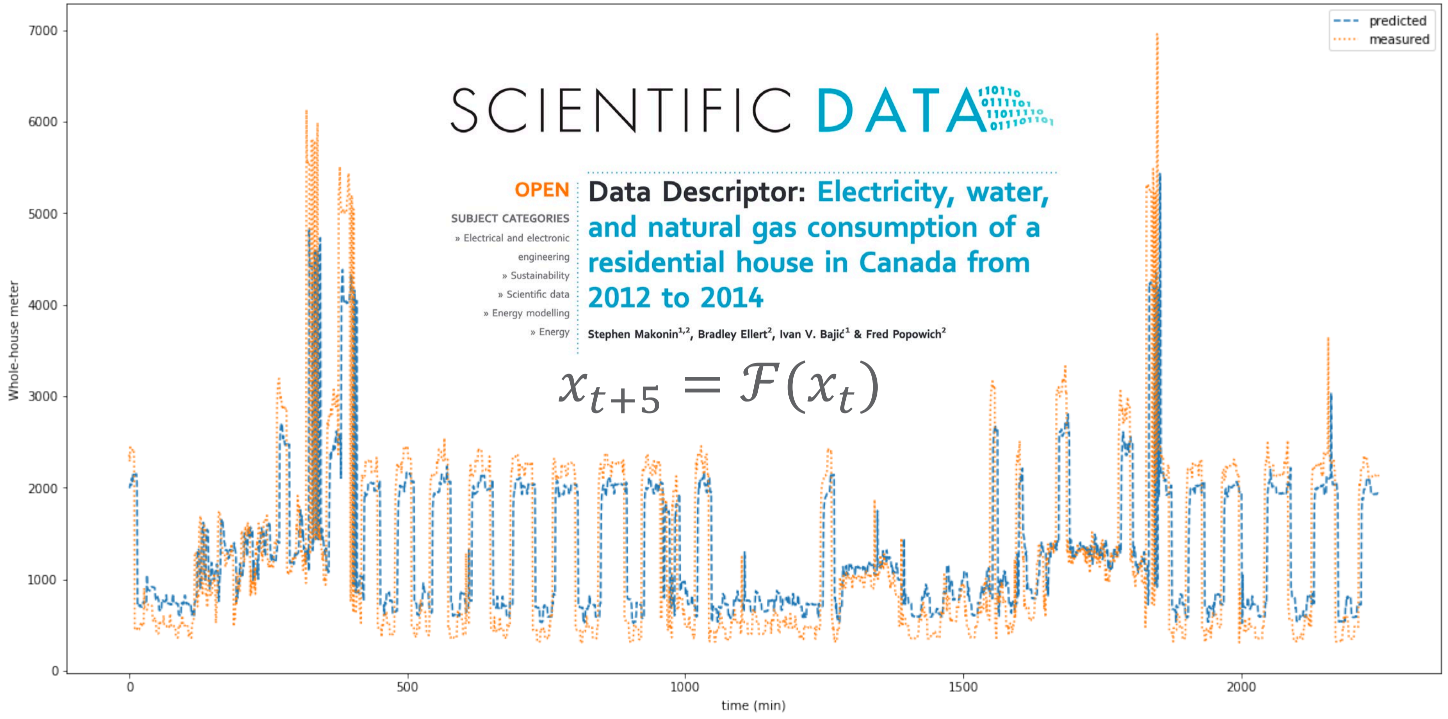


Echo State Machines

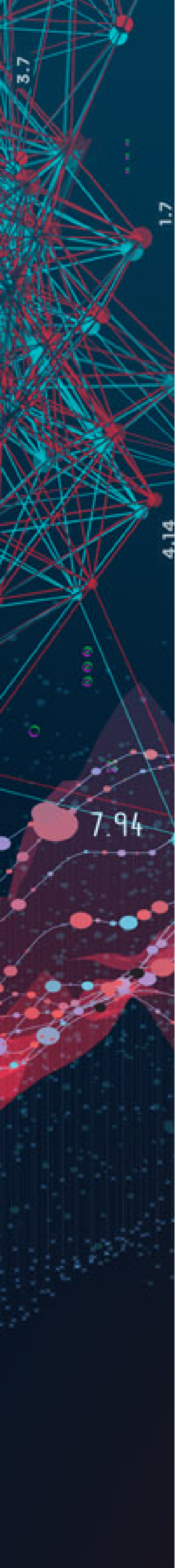


Reservoir size: 100 units
(15% density)





Online training and short-time prediction (5min)



Pacific Northwest
NATIONAL LABORATORY

Thank you