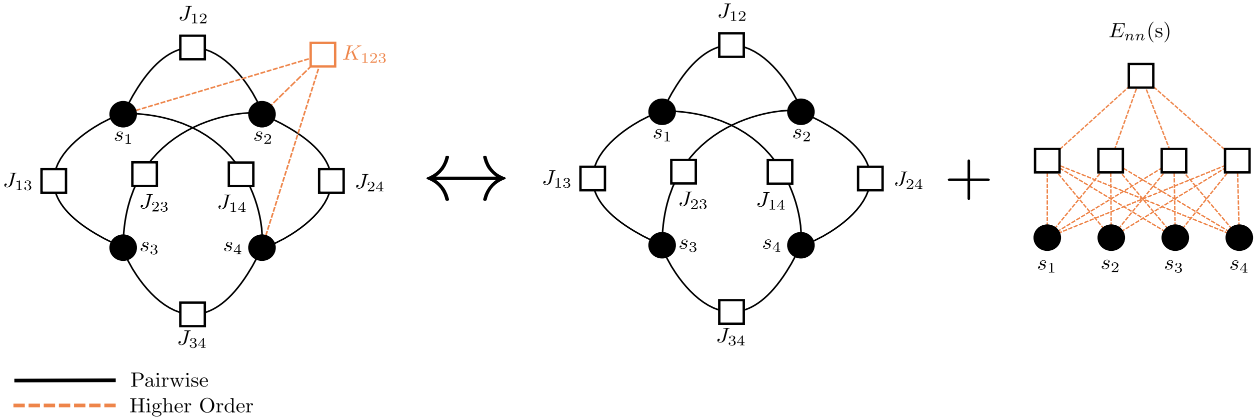
Paper Presentation, MSML 2021

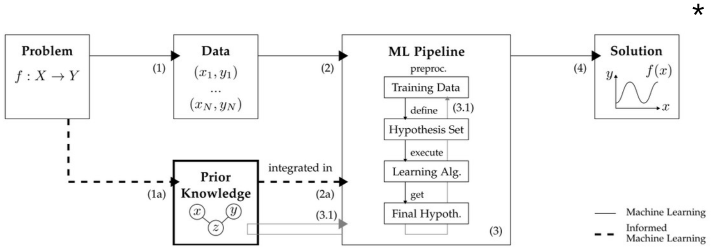
Reconstruction of Pairwise Interactions using Energy-Based Models

Christoph Feinauer & Carlo Lucibello
Bocconi University



One-Slide Summary

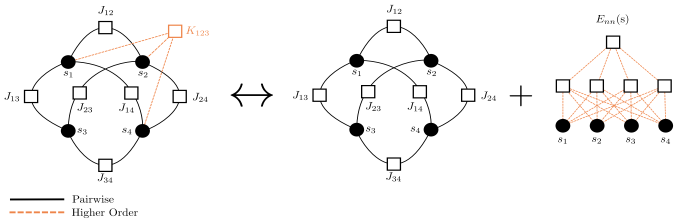
Knowledge Integration



Energy-Based Models

$$p_\theta(\mathbf{x}) = \frac{e^{-E_\theta(\mathbf{x})}}{Z_\theta}$$

Hybrid EBMs



Coupling Reconstruction

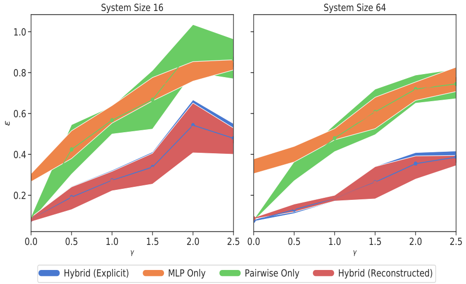
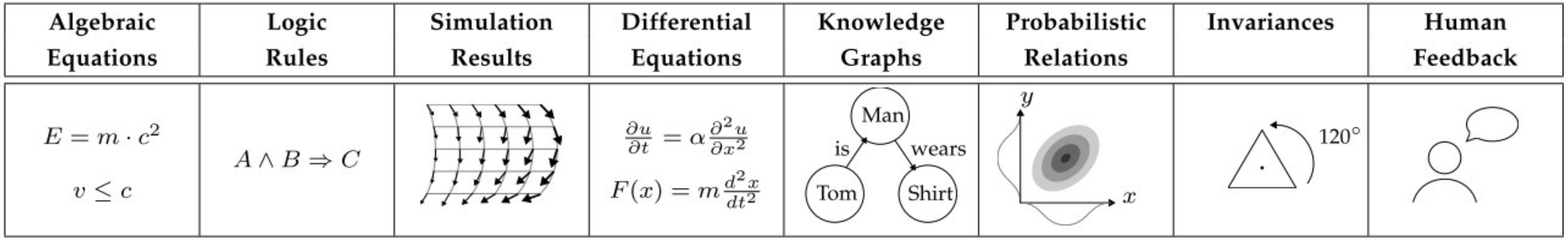


*Image from: von Rueden et al. 2019, *Informed Machine Learning – A Taxonomy and Survey.*

Knowledge Integration

- Great success in recent years of data-driven black-box approaches (mostly deep learning)
- Domain knowledge and modeling still very relevant though:
  - Ever increasing compute, but low data availability can be an issue sometimes
  - Interpretability & Fairness
  - By modeling and testing our models we increase our understanding
- Leverage both worlds:
  - hybrid modeling
  - knowledge integration
  - (physics) informed machine learning



*

*Image from: von Rueden et al. 2019, *Informed Machine Learning – A Taxonomy and Survey.*

Generative Modeling

Given a dataset $D = \{\mathbf{x}^\mu\}_{\mu=1}^M$ the task is to generate new samples consistent with the data distribution.

Prominent generative models:

- Variational Autoencoders
- Generative Adversarial Networks
- Autoregressive Neural Networks
- Energy-Based Models

Different trade-offs (complexity, generation quality, likelihood existence/tractability).

<u>Compositionality</u>

If we think of a distribution as a set of constraints, what if we have several such sets of constraints?

Energy Based Models

Density distributions with untractable normalization

$$p_\theta(\mathbf{x}) = \frac{e^{-E_\theta(\mathbf{x})}}{Z_\theta}$$

Maximum-Likelihood Objective

$$\nabla_\theta \log p_\theta(D) = -\mathbb{E}_{\mathbf{x}\sim D}\left[\nabla_\theta E_\theta(\mathbf{x})\right] + \mathbb{E}_{\mathbf{x}\sim p_\theta}\left[\nabla_\theta E_\theta(\mathbf{x})\right]$$

Energy Based Models

Pros

- Simplicity
- Flexibility
- **Compositionality**

$$E(x) = E_1(x) + E_2(x)$$

Cons

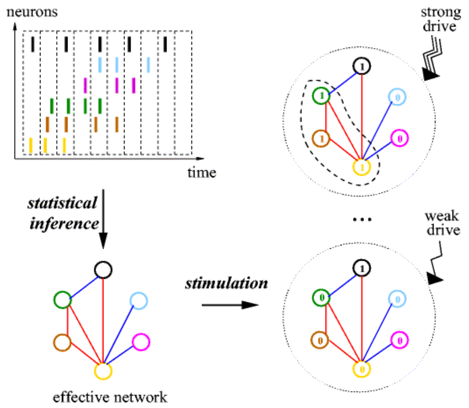
- Hard to train
- Hard to sample from
- Likelihood known only up to a normalization factor

Pairwise Models

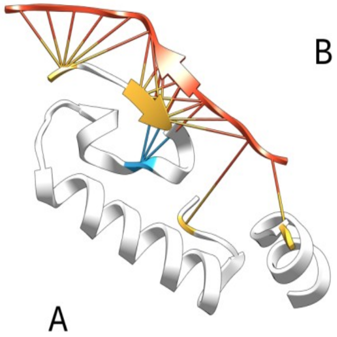
So called Boltzmann machines are a popular class of EBMs

$$E_{pw}(s) = -\sum_i h_i s_i - \sum_{i<j} J_{ij} s_i s_j \qquad s_i \in \{-1, +1\}$$

Such pairwise models already capture a good part of the data distribution in many applications.



*



**

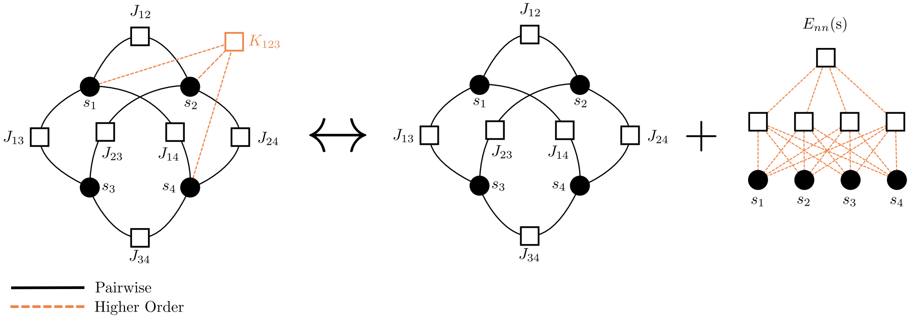
*Tavoni, Gaia, Simona Cocco, and Rémi Monasson. "Neural assemblies revealed by inferred connectivity-based models of prefrontal cortex recordings." Journal of computational neuroscience 41.3 (2016): 269-293.

**Feinauer, Christoph, and Martin Weigt. "Context-aware prediction of pathogenicity of missense mutations involved in human disease." arXiv preprint arXiv:1701.07246 (2017)

Hybrid models

In order to encode some prior structural knowledge of the data generating distribution, we use an hybrid approach (physics + black-box):

$$E_{hybrid}(s) = E_{pw}(s) + E_{nn}(s) = -\sum_{i<j} J_{ij} s_i s_j + E_{nn}(s)$$



Training EBMs: Pseudolikelihoods

Pseudolikehood maximization:

$$\mathcal{L}(\theta) = \frac{1}{B} \sum_{b=1}^{B} \sum_{i=1}^{N} \log p_\theta(s_i^b \mid s_{/i}^b)$$

Pseudo-likelihood can be easily computed for a generic energy model:

$$\log p_\theta(s_i \mid s_{/i}) = \log \frac{e^{-E_\theta(s_i, s_{/i})}}{\sum_{s_i'} e^{-E_\theta(s_i', s_{/i})}}$$

(2 forwards for binary)

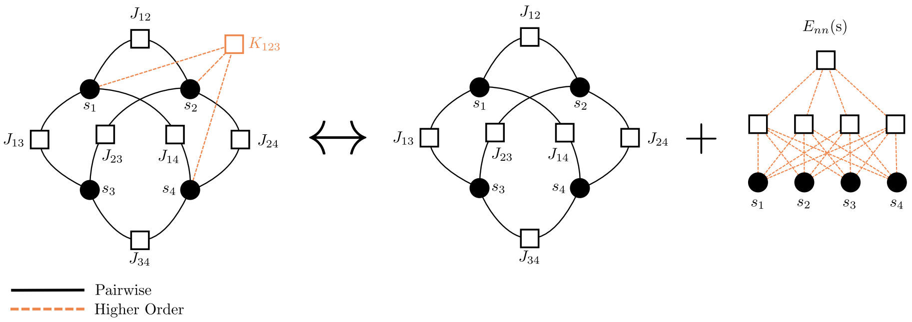
Coupling reconstruction

We define a synthetic coupling reconstruction task where the data is generated by the Hamiltonian

$$E_G(s) = -\sum_{i<j} J_{ij}^G s_i s_j - \sqrt{\gamma} \sum_{I \in \mathcal{I}_G} \xi_I^G \prod_{i \in I} s_i.$$

We consider N randomly sampled higher order interactions.

Training data is generated by a MCMC.



Coupling extraction

How to extract pairwise interactions from a generic learned energy function?

Our couplings' **estimators** are given by

$$\hat{J}_{ij} = -\frac{1}{2^N} \sum_{s} E(s)\, s_i s_j$$

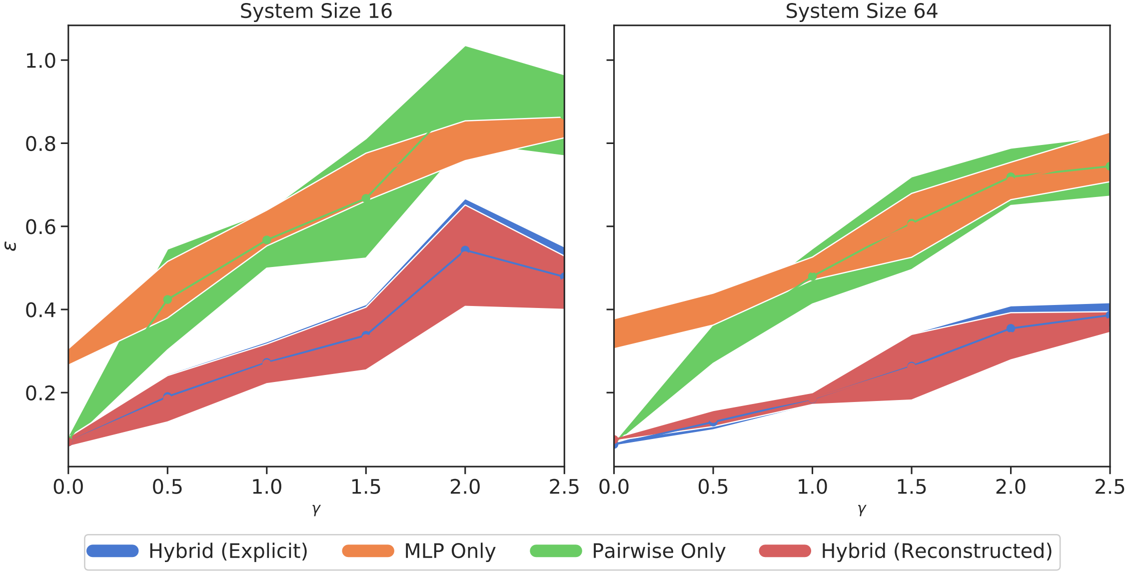
This can be approximated by uniform sampling at the end of training.

The **reconstruction error** is then given by

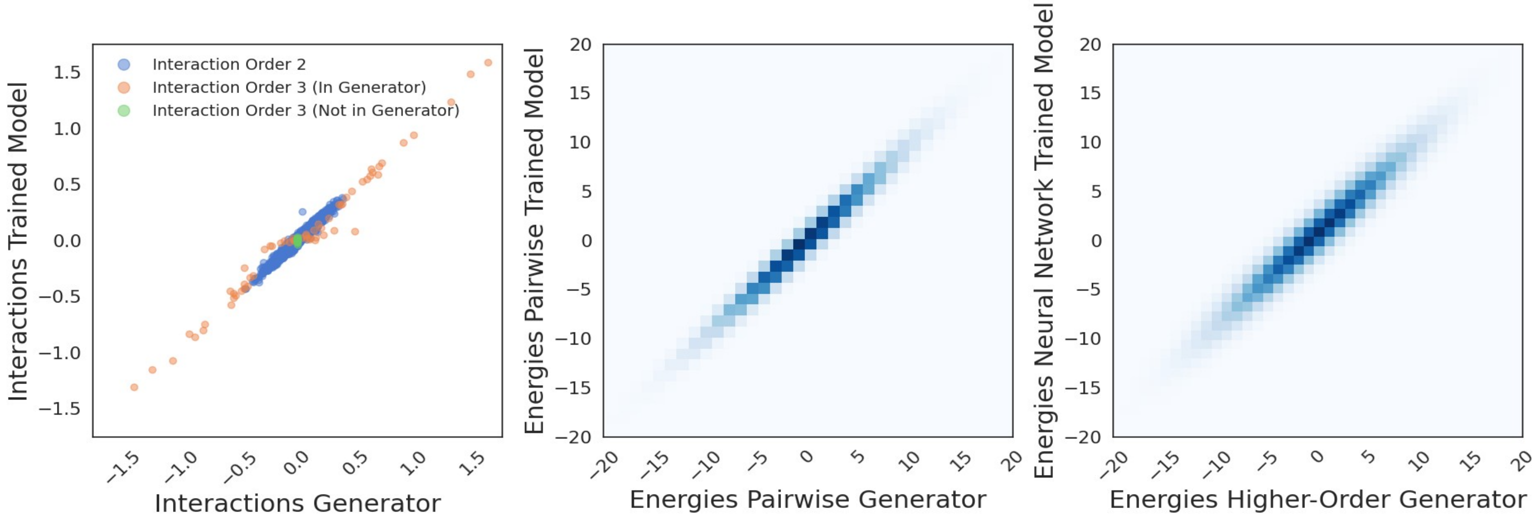
$$\epsilon = \sqrt{\frac{\sum_{i<j} \left(J_{ij}^G - \hat{J}_{ij} \right)^2}{\sum_{i<j} \left(J_{ij}^G \right)^2}}$$

Experiment Results I

Order 3 interactions, nhiddens=128, varying higher-order interaction strength parameter



Experiment Results II



Conclusions

- Energy Based Models are good candidates for knowledge integration
- Training is expensive and sensitive to hyperparameters, some tendency to overfitting (will try different training techniques)
- Currently working on proteins (some preliminary results in the paper)