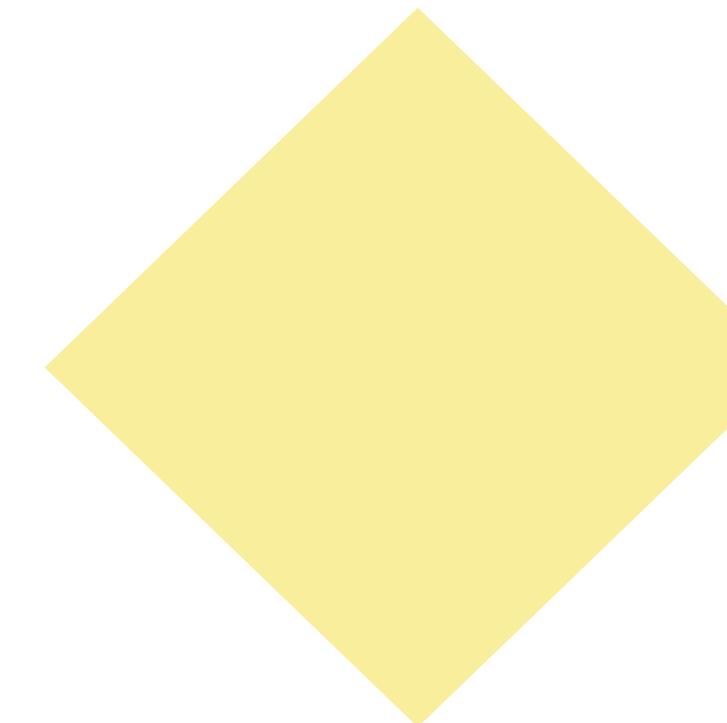
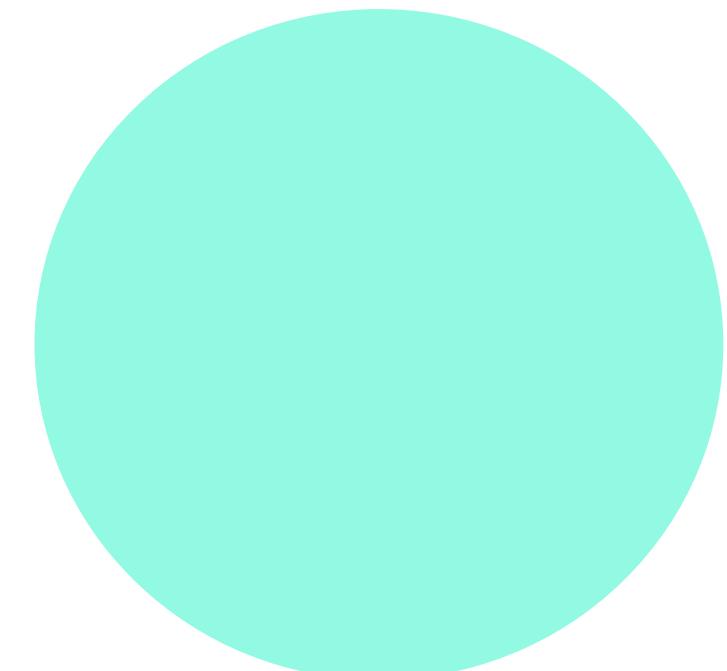


E(n) Equivariant Normalizing Flows

Victor Garcia Satorras, Emiel Hoogeboom, Fabian B. Fuchs, Ingmar Posner, Max Welling

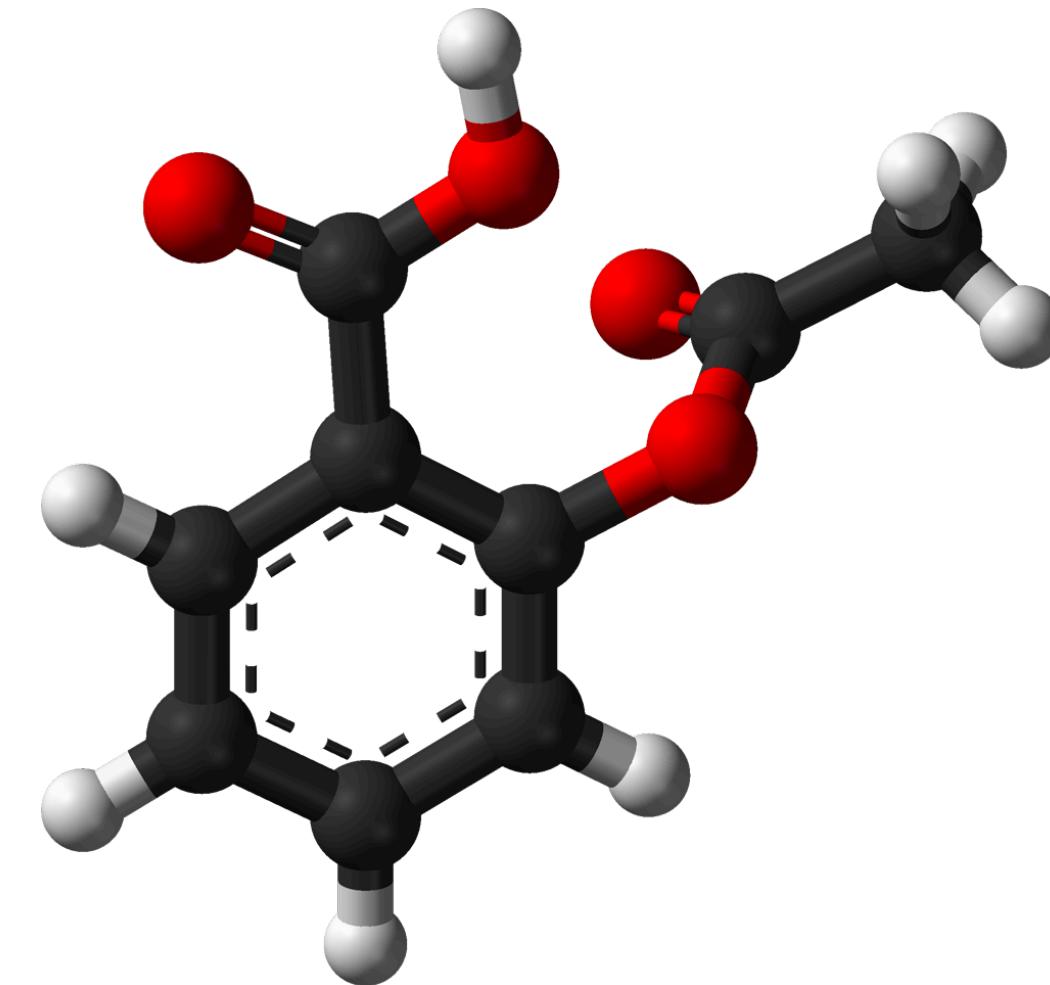
Prashant Govindarajan
4th November, 2022
COMP760



Motivation

Molecule Generation

- Consider the example of building a generative model for designing molecules
- Molecules can be modelled as graphs
 - **Nodes:** Atoms
 - **Edges:** Bonds



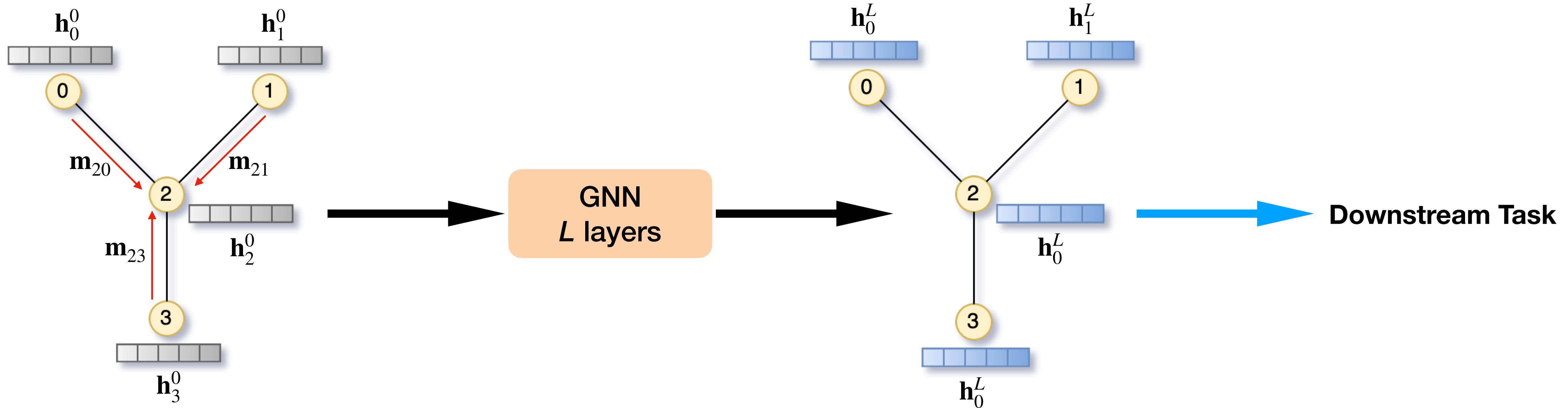
Node features

- x** Spatial position of atoms
h Atom features (e.g., atom type, charge, etc.)

- Expectations of the generative model
 - **Permutation** invariance/equivariance
 - Invariant/equivariant to **rigid body transformations** (rotations or translations)

Graph Neural Network

GNN Basics



Step 1: Message Passing

$$\mathbf{m}_{ij} = \phi_e (\mathbf{h}_i^l, \mathbf{h}_j^l, a_{ij}) \quad \text{Compute messages for all edges}$$

Step 2: Neighborhood Aggregation

$$\mathbf{m}_i = \sum_{j \in \mathcal{N}(i)} \mathbf{m}_{ij} \quad \text{Permutation invariant aggregation}$$

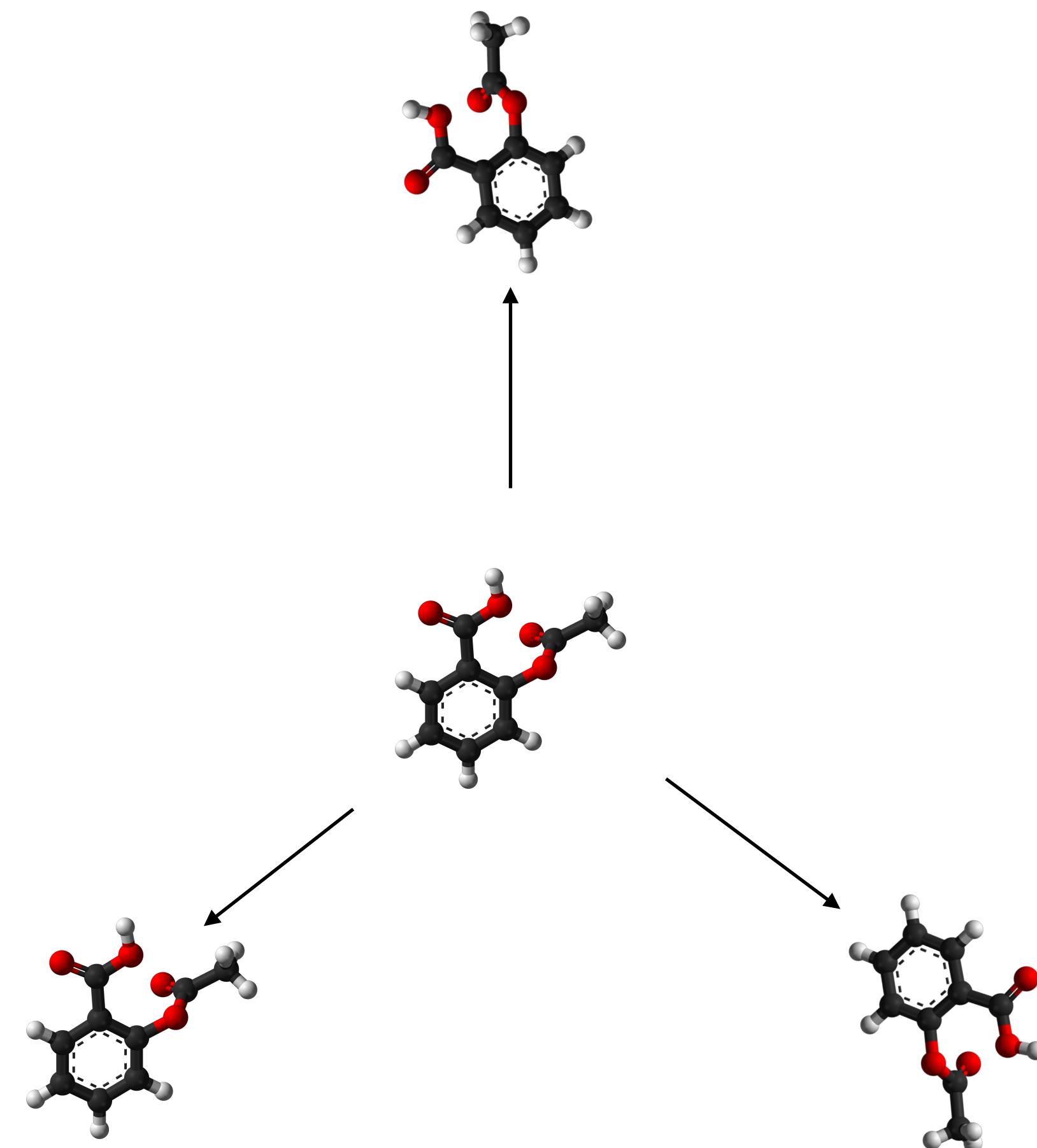
Step 3: Update node state

$$\mathbf{h}_i^{l+1} = \phi_h (\mathbf{h}_i^l, \mathbf{m}_i)$$

E(n) Equivariant GNN

E(n) Equivariance

- The Euclidean group, or E(n), is a group of transformations that **preserve the Euclidean distance** between two points
 - Translation
 - Rotation
 - Reflection
- Equivariance
$$\phi(T_g(\mathbf{x})) = S_g(\phi(\mathbf{x}))$$
- E(n) Equivariance
 - Rotation/reflection matrix
 - Translation vector
$$\mathbf{R}\mathbf{z}_x + \mathbf{t}, \mathbf{z}_h = f(\mathbf{R}\mathbf{x} + \mathbf{t}, \mathbf{h})$$



E(n) Equivariant GNN

(Satorras et al. 2021)

- E(n) invariant message passing step

$$\mathbf{m}_{ij} = \phi_e \left(\mathbf{h}_i^l, \mathbf{h}_j^l, \|\mathbf{x}_i^l - \mathbf{x}_j^l\|^2, a_{ij} \right)$$

Euclidean distance is
E(n) invariant

- E(n) equivariant position update step

$$\mathbf{x}_i^{l+1} = \mathbf{x}_i^l + C \sum_{j \neq i} (\mathbf{x}_i^l - \mathbf{x}_j^l) \phi_x(\mathbf{m}_{ij})$$

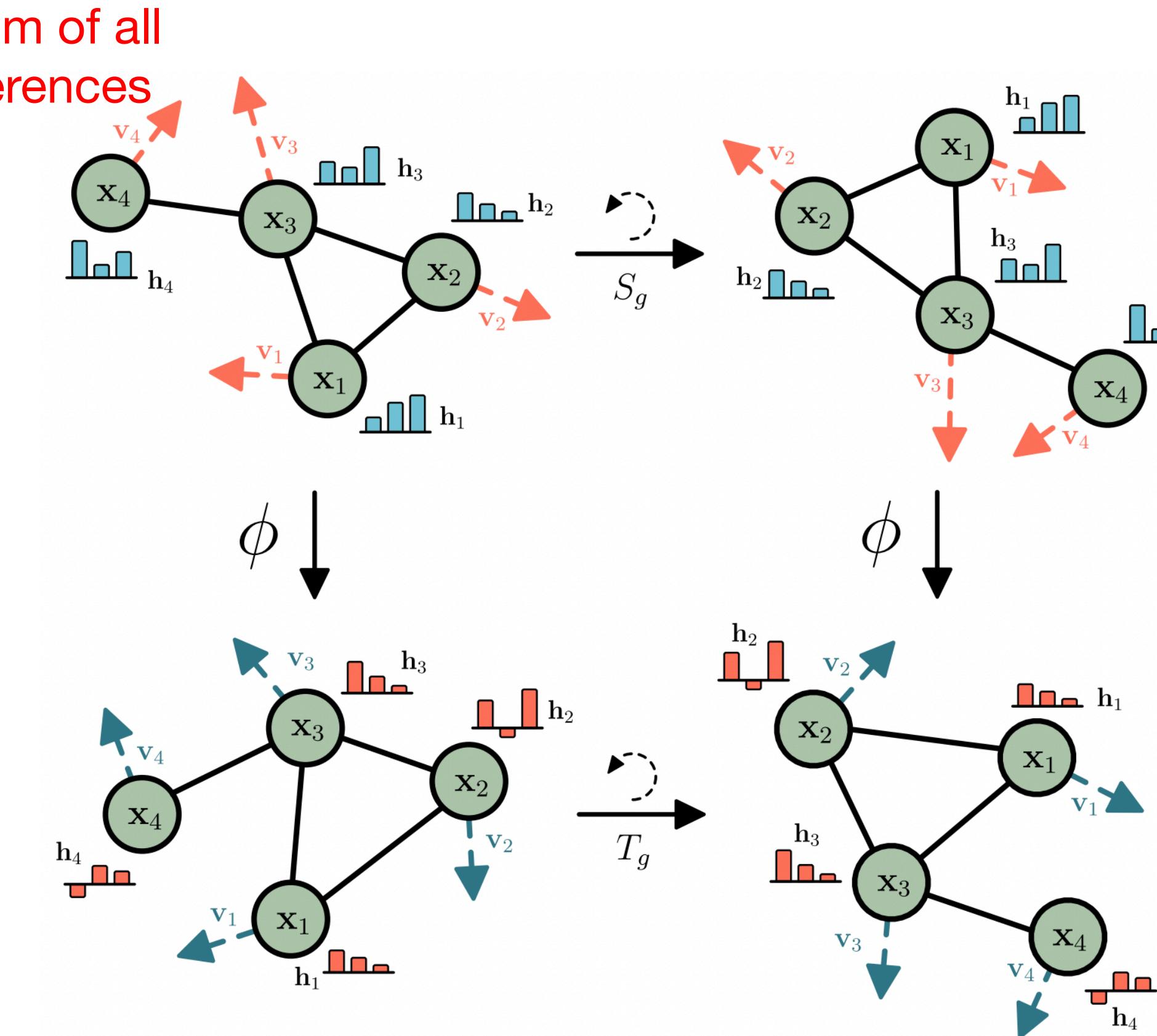
- Message aggregation

$$\mathbf{m}_i = \sum_{j \neq i} \mathbf{m}_{ij}$$

Summation can also
be performed over
just the neighbors

- Feature update

$$\mathbf{h}_i^{l+1} = \phi_h(\mathbf{h}_i^l, \mathbf{m}_i)$$



Continuous Normalizing Flows

Continuous Normalizing Flows

- **Normalizing Flow**

$$\mathbf{z} = f(\mathbf{x})$$

$$p_X(\mathbf{x}) = p_Z(\mathbf{z}) |\det J_f(\mathbf{x})|$$

$$\log p_X(\mathbf{x}) = \log p_Z(\mathbf{z}) + \log |\det J_f(\mathbf{x})|$$

- **Continuous Normalizing Flow**

$$\frac{d\mathbf{x}(t)}{dt} = \phi(\mathbf{x}(t))$$

$$\mathbf{z} = \mathbf{x}(0) + \int_0^1 \phi(\mathbf{x}(t)) dt \quad \text{Computed by ODE solver}$$

$$\log p_X(\mathbf{x}) = \log p_Z(\mathbf{z}) + \int_0^1 \text{Tr} J_\phi(\mathbf{x}(t)) dt$$

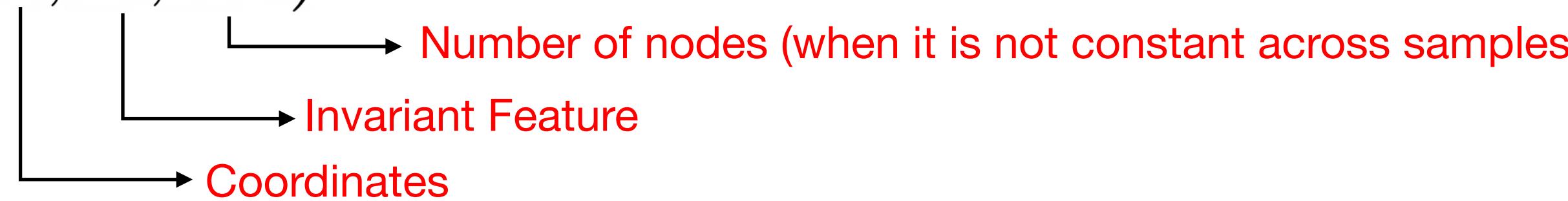


Trace of the
Jacobian

Methods

Workflow

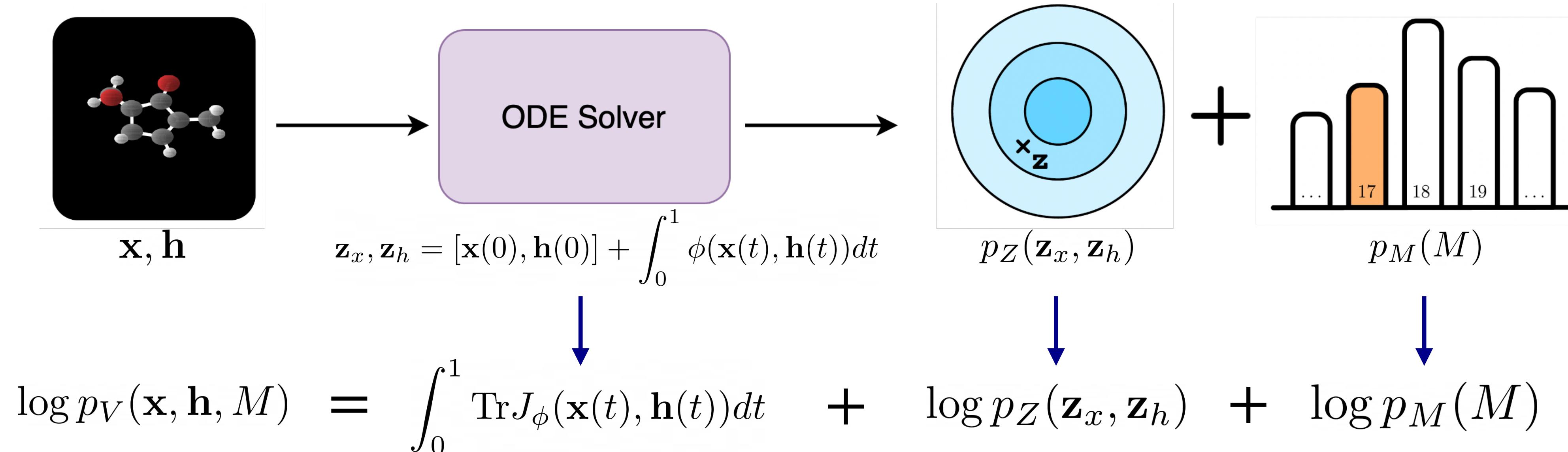
- **Goal:** To model $p_V(\mathbf{x}, \mathbf{h}, M)$



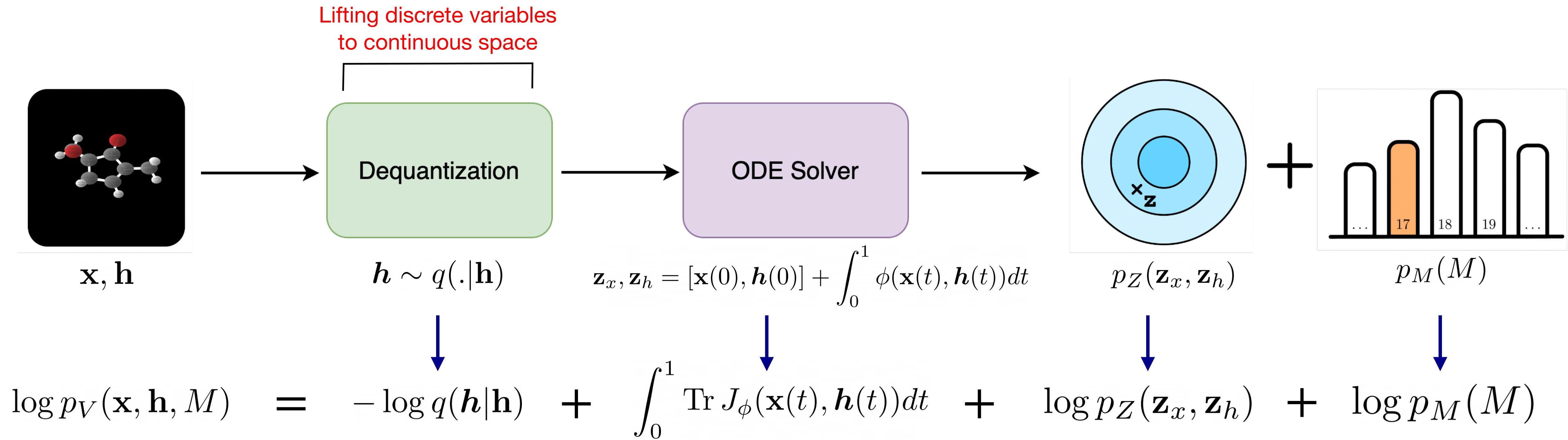
$$p_V(\mathbf{x}, \mathbf{h}, M) = p_V(\mathbf{x}, \mathbf{h}|M)p_M(M)$$

\mathbf{x}, \mathbf{h} Modelled using
Categorical
Equivariant NFs Distribution

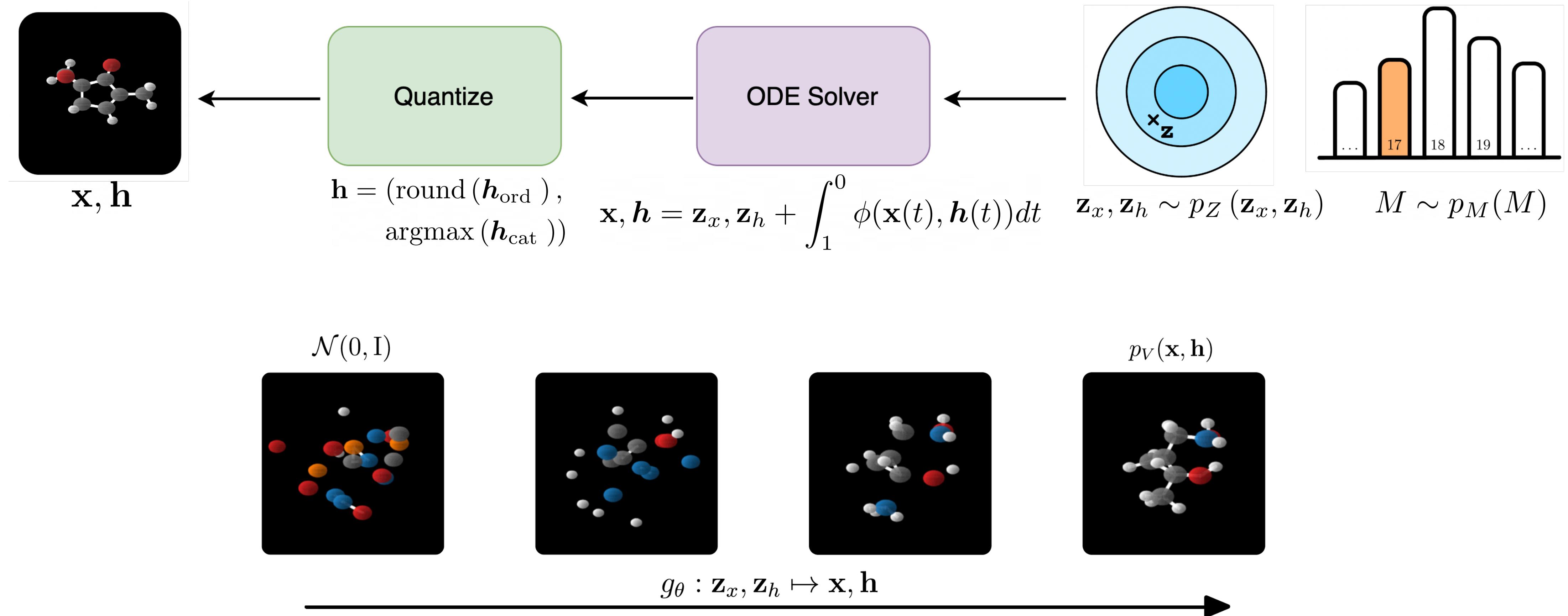
NOTE: In this approach, edges of the graph
are not modelled (no adjacency matrix).



Workflow



Generation



Normalizing Flow

$$\mathbf{z}_x, \mathbf{z}_h = [\mathbf{x}(0), \mathbf{h}(0)] + \int_0^1 \phi(\mathbf{x}(t), \mathbf{h}(t)) dt$$

- EGNN outputs
 $\mathbf{x}^L(t), \mathbf{h}^L(t) = \text{EGNN}[\mathbf{x}(t), \mathbf{h}(t)]$
- ODE for the invariant feature

$$\frac{d}{dt} \mathbf{h}(t) = \mathbf{h}^L(t)$$

- ODE for the positional feature

$$\frac{d}{dt} \mathbf{x}(t) = \mathbf{x}^L(t) - \mathbf{x}(t)$$



Velocity type invariance,
i.e., velocity is invariant
to translations

E(n) Equivariant GNN

- E(n) invariant message passing

$$\mathbf{m}_{ij} = \phi_e \left(\mathbf{h}_i^l, \mathbf{h}_j^l, \|\mathbf{x}_i^l - \mathbf{x}_j^l\|^2 \right)$$

Euclidean distance is
E(n) invariant

- E(n) equivariant position update

$$\mathbf{x}_i^{l+1} = \mathbf{x}_i^l + \sum_{j \neq i} \frac{(\mathbf{x}_i^l - \mathbf{x}_j^l)}{\|\mathbf{x}_i^l - \mathbf{x}_j^l\|^2 + 1} \phi_x (\mathbf{m}_{ij})$$

For stability reasons

- Message aggregation

$$\mathbf{m}_i = \sum_{j \neq i} \phi_{\text{inf}} (\mathbf{m}_{ij}) \mathbf{m}_{ij}$$

Edge Inference, inferring edges
using a parametrized network

- Feature update

$$\mathbf{h}_i^{l+1} = \phi_h (\mathbf{h}_i^l, \mathbf{m}_i)$$

Further Details

- **Dequantization of discrete features**

$\mathbf{h} = (\mathbf{h}_{\text{ord}}, \mathbf{h}_{\text{cat}})$ Ordinal (e.g. charge) and categorical (e.g. atom type)

$$\begin{aligned}\mathbf{h}_{\text{ord}} &= \mathbf{h}_{\text{ord}} + u, \quad u \sim q_{\text{ord}}(\cdot | \mathbf{h}_{\text{ord}}) \\ \mathbf{h}_{\text{cat}} &\sim q_{\text{cat}}(\cdot | \mathbf{h}_{\text{cat}})\end{aligned}$$

Normal Distributions; parameters learned by EGNN

$$\log p_{\mathbf{H}}(\mathbf{h}) \geq \mathbb{E}_{\mathbf{h} \sim q_{\text{ord}, \text{cat}}(\cdot | \mathbf{h})} [\log p_H(\mathbf{h}) - \log q_{\text{ord,cat}}(\mathbf{h} | \mathbf{h})]$$

$\mathbf{h} = (\text{round}(\mathbf{h}_{\text{ord}}), \text{argmax}(\mathbf{h}_{\text{cat}}))$ During generation

- **Translation Invariance**

- For the distribution to be translation invariant, $p_X(\mathbf{x} + \mathbf{t}) = p_X(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^{M \times n}$
- $p_X(\mathbf{x} + \mathbf{t})$ does not always integrate to 1
- Solution:
 - Mean center the positions and project them into an $(M - 1) \times n$ dimensional linear subspace
 - Jacobian in the subspace is related to the Jacobian in the ambient space

- **Base Distribution**

$$p(\mathbf{z}_x, \mathbf{z}_h) = p(\mathbf{z}_x)p(\mathbf{z}_h)$$

- Standard Gaussian priors

Experiments and Results

DW4 and LJ13

- **DW4** - 4 particle systems in 2 dimensional space governed by an energy function
- **LJ13** - approximation of inter-molecular pair that models repulsive and attractive interactions. 13 particle system in 3 dimensional space

Analysis of negative log-likelihood values

# Samples	DW4				LJ13			
	10 ²	10 ³	10 ⁴	10 ⁵	10	10 ²	10 ³	10 ⁴
GNF	11.93	11.31	10.38	7.95	43.56	42.84	37.17	36.49
GNF-att	11.65	11.13	9.34	7.83	43.32	36.22	33.84	32.65
GNF-att-aug	8.81	8.31	7.90	7.61	41.09	31.50	30.74	30.93
Simple dynamics	9.58	9.51	9.53	9.47	33.67	33.10	32.79	32.99
Kernel dynamics	8.74	8.67	8.42	8.26	35.03	31.49	31.17	31.25
E-NF	8.31	8.15	7.69	7.48	33.12	30.99	30.56	30.41

Comparisons

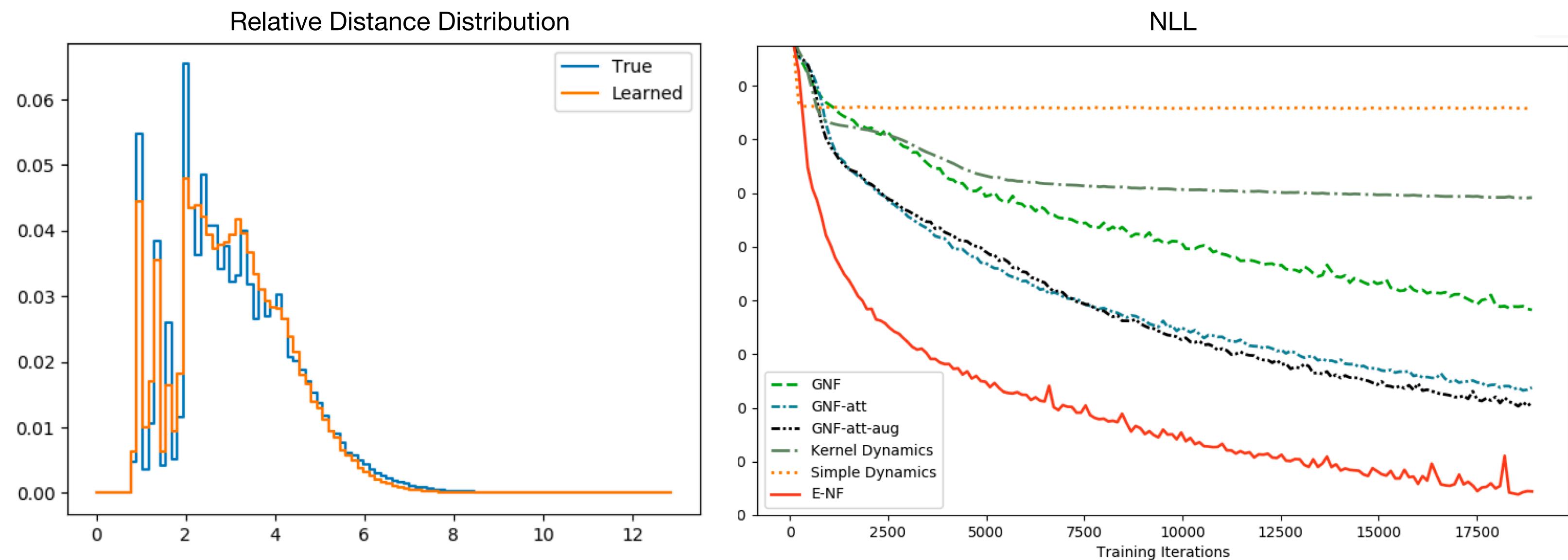
- **GNF** : Graph Normalizing Flow
- **GNF-att** : Graph Normalizing Flow with attention
- **GNF-att-aug** : Graph Normalizing Flow with attention and data augmentation
- **Simple & Kernel Dynamics** : Equivariant Flows that do not work with feature information

QM9 Positional

- **QM9 positional** - Subset of QM9 (small molecule dataset of chemical properties) consisting of only positional information
- Node features are absent

Analysis of negative log-likelihood values and Jensen-Shannon (JS) divergence

# Metrics	NLL	JS(rel. dist)
Simple dynamics	73.0	.086
Kernel dynamics	38.6	.029
GNF	-00.9	.011
GNF-att	-26.6	.007
GNF-att-aug	-33.5	.006
E-NF (ours)	-70.2	.006



QM9 Molecules

- **QM9** - small molecule dataset (up to 29 atoms per molecule)
- Node features: atom properties like atom type (categorical) and charge (ordinal)
- **Metrics:**

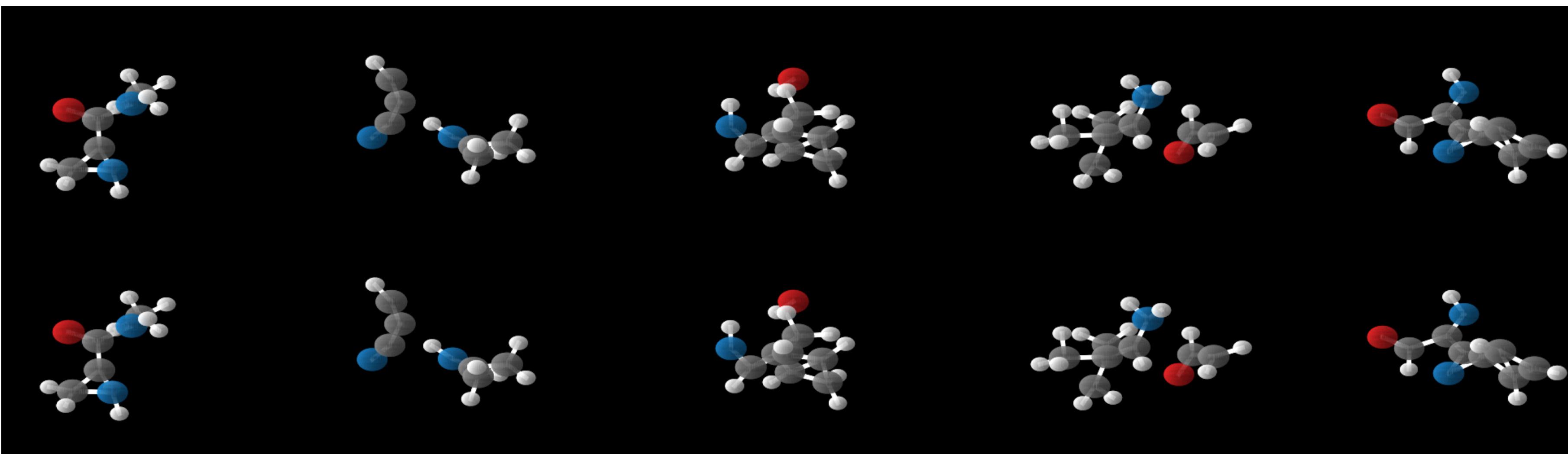
- Atom Stability (% of atoms obeying valence rules)
- Molecule Stability (% of molecules obeying valence rules)
- Novelty and uniqueness

# Metrics	NLL	Atom stability	Mol stable
GNF-attention	-28.2	72%	0.3%
GNF-attention-augmentation	-29.3	75%	0.5%
E-NF (ours)	-59.7	85%	4.9%
Data	-	99%	95.2%

93.28% novel
99.80% unique

Random Samples

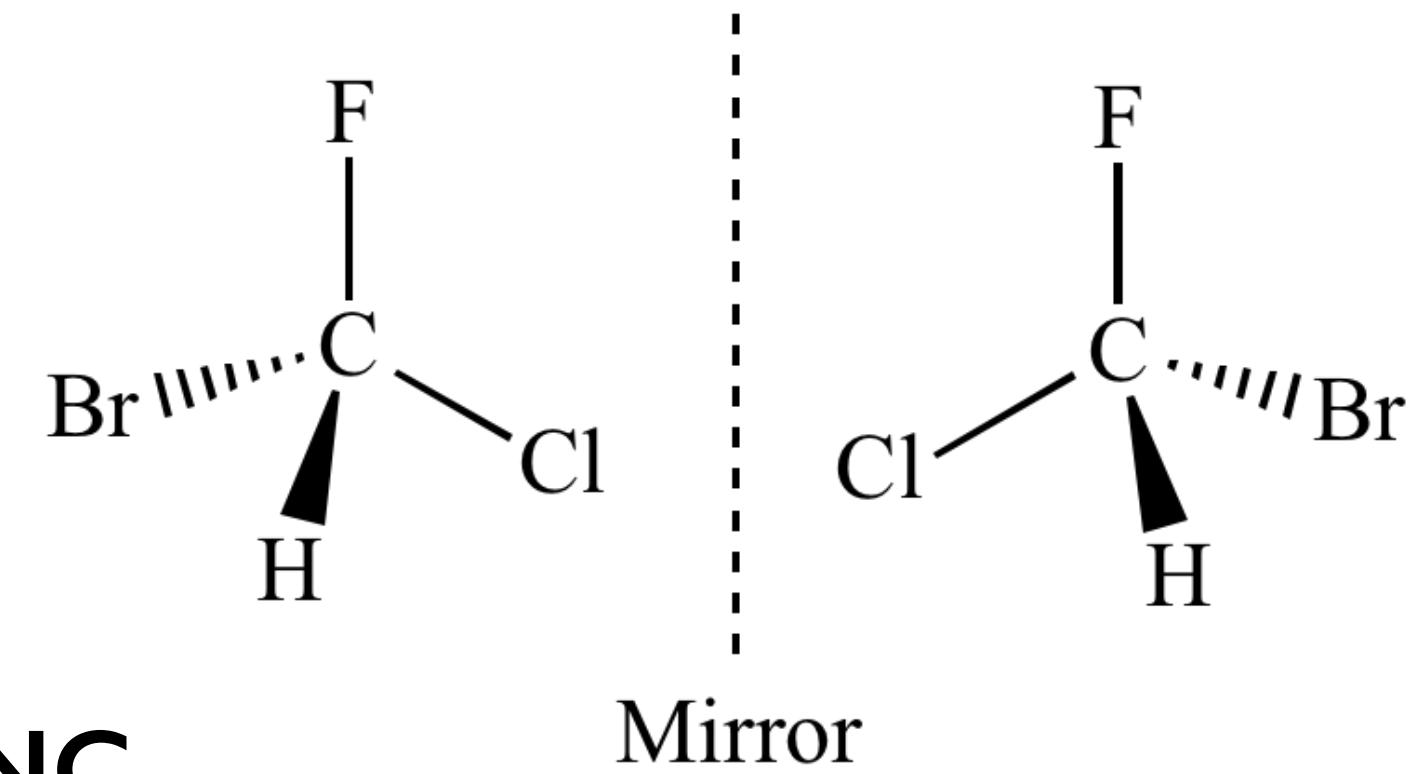
Stable Samples



Drawbacks and Limitations

Drawbacks and Limitations

- ODE type flow is **computationally expensive**
 - Moser Flows can be explored (?)
- **Stability issues** when combining ODE flows and EGNN
- Chiral molecules are not reflection invariant
 - **SE(3) equivariant networks** might be more appropriate for molecules
- The approach could have been evaluated on larger molecule datasets like **ZINC**
 - More number of nodes per molecule
 - Incorporating edge information
- Dequantization of discrete features produces **bias**
- **Energies** of generated molecules could have been evaluated



Interesting question to ponder: Are there real-world applications of this method beyond 3 dimensions?

**Thank You
Questions?**