

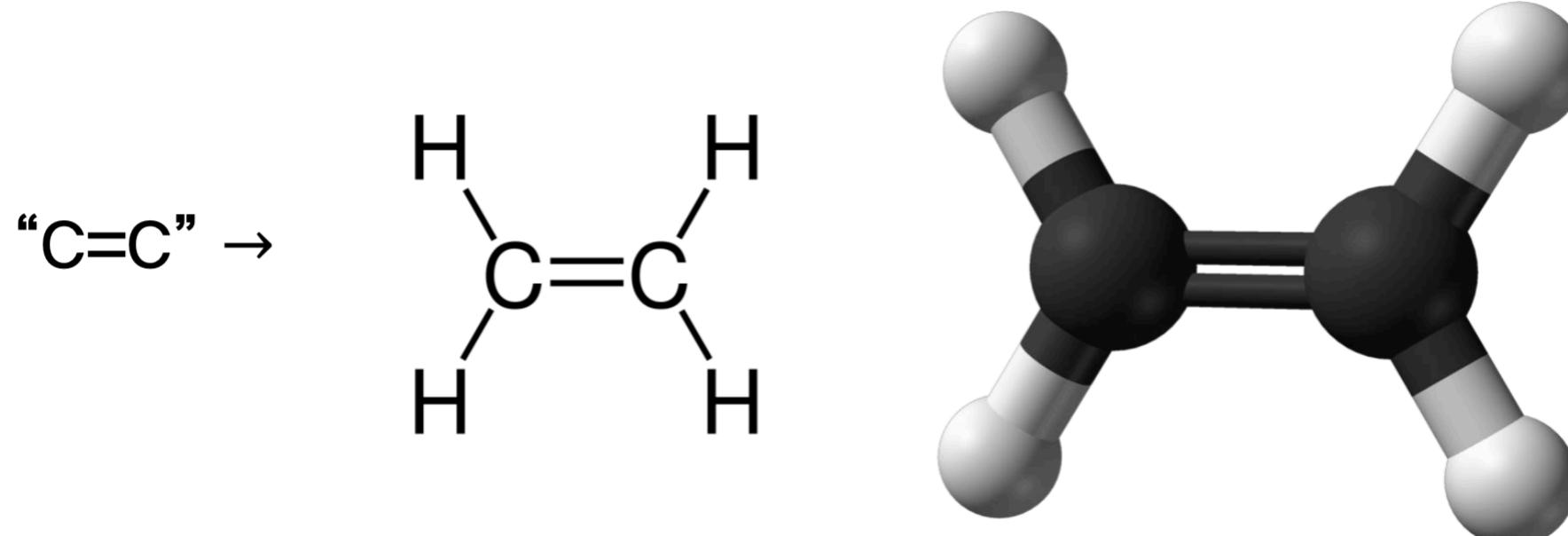
Reinforcement Learning for molecular design guided by quantum mechanics

김 준 태
Standigm

Task

3D molecular structure prediction

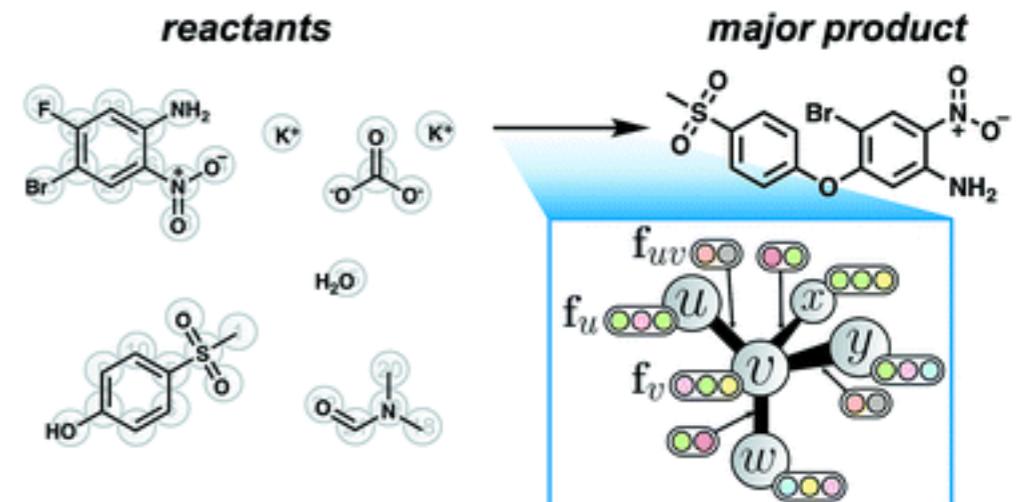
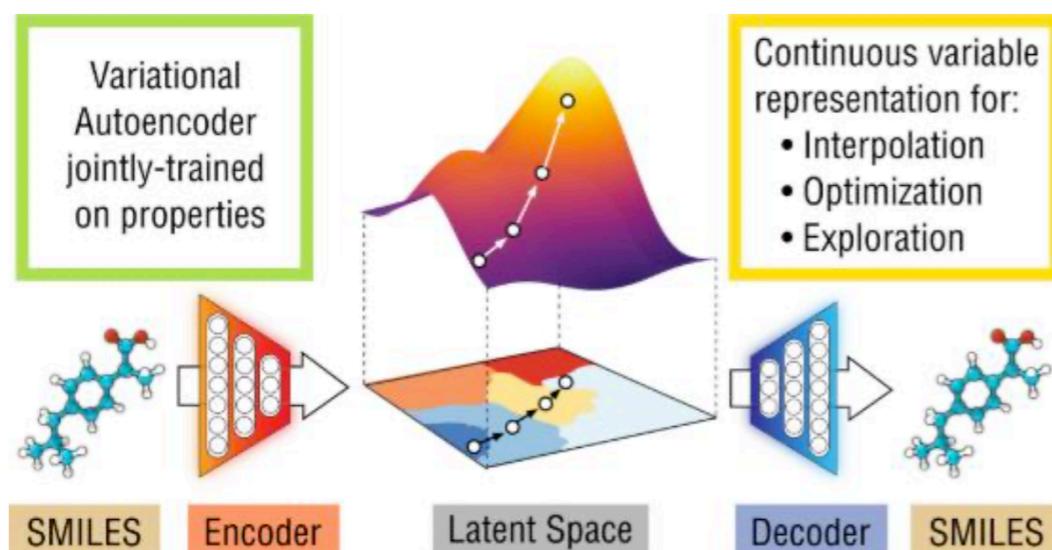
- Find a structure, which has minimum potential energy
- Smiles string -> optimal 3D conformation



Motivation

Drug Discovery

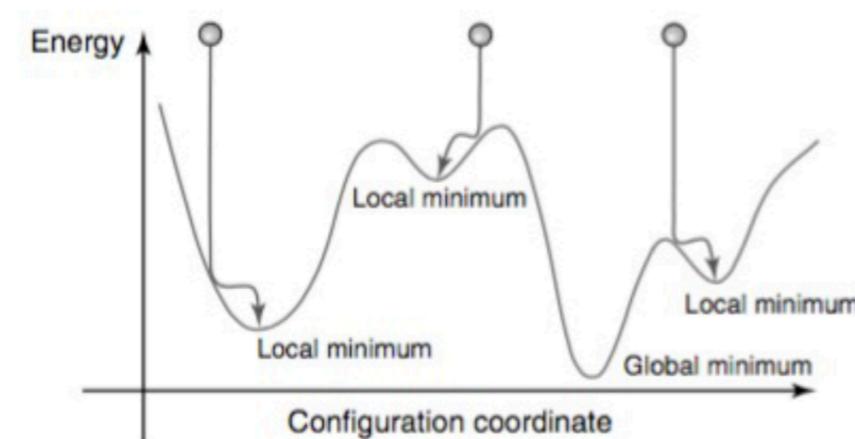
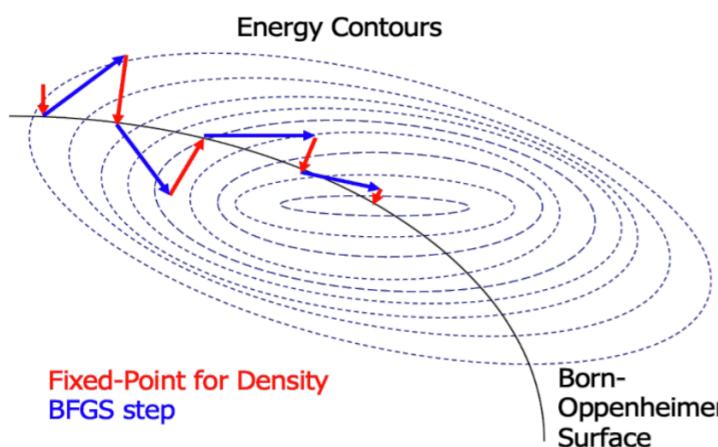
- Using SMILES string is quite good
- However, if we know 3D structure, we can do better
- And, We don't use Graph structure



Motivation

Conventional method

- Optimization is based on the gradient of energy
- Searching the lowest point of potential energy surface(PES) iteratively



Motivation

Conventional method

- Energy calculation requires massive amount of computational cost
- Density Functional Theory (DFT)

Reinforcement Learning to predict 3D structure of molecules

- Expectation: RL agent can find the optimal molecular structure

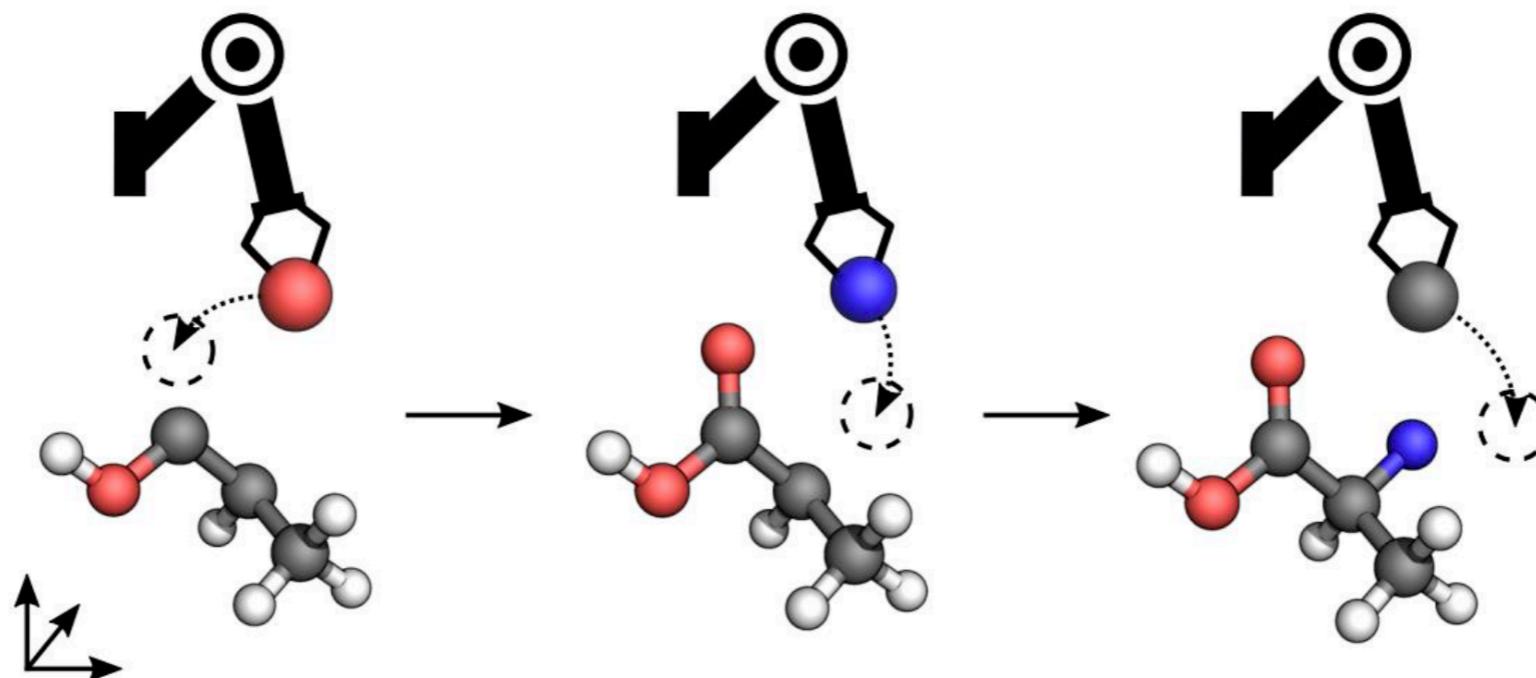
Proposed method

Contributions

- **The first work that proposed a novel RL formulation for general molecular Design in Cartesian coordinates**
- **Introduce proper reward function and policy network architecture**
- **Proposed MOLGYM, which is a RL environment for molecular design task**

Proposed method

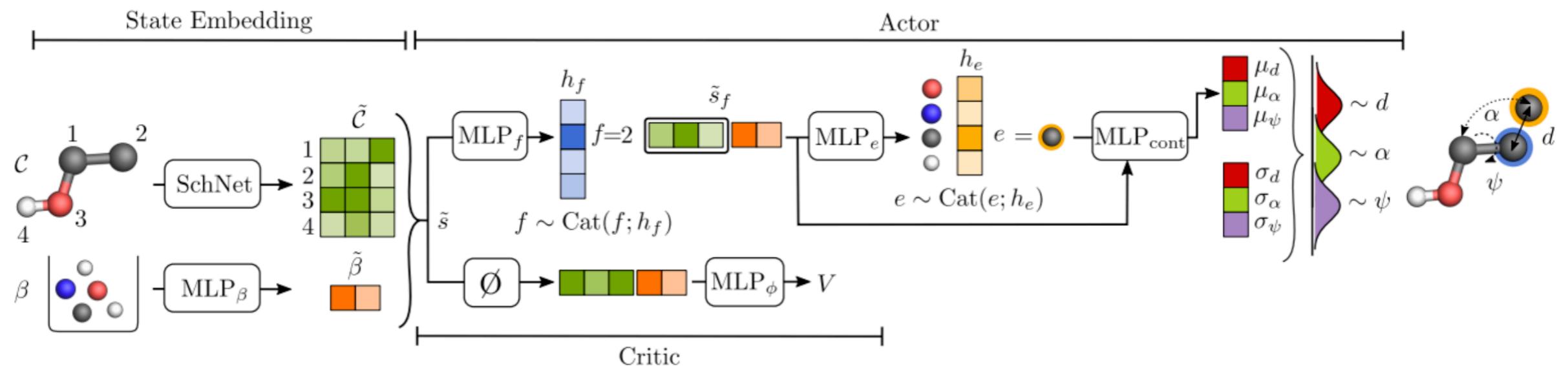
Idea: grab-a-prize of atoms



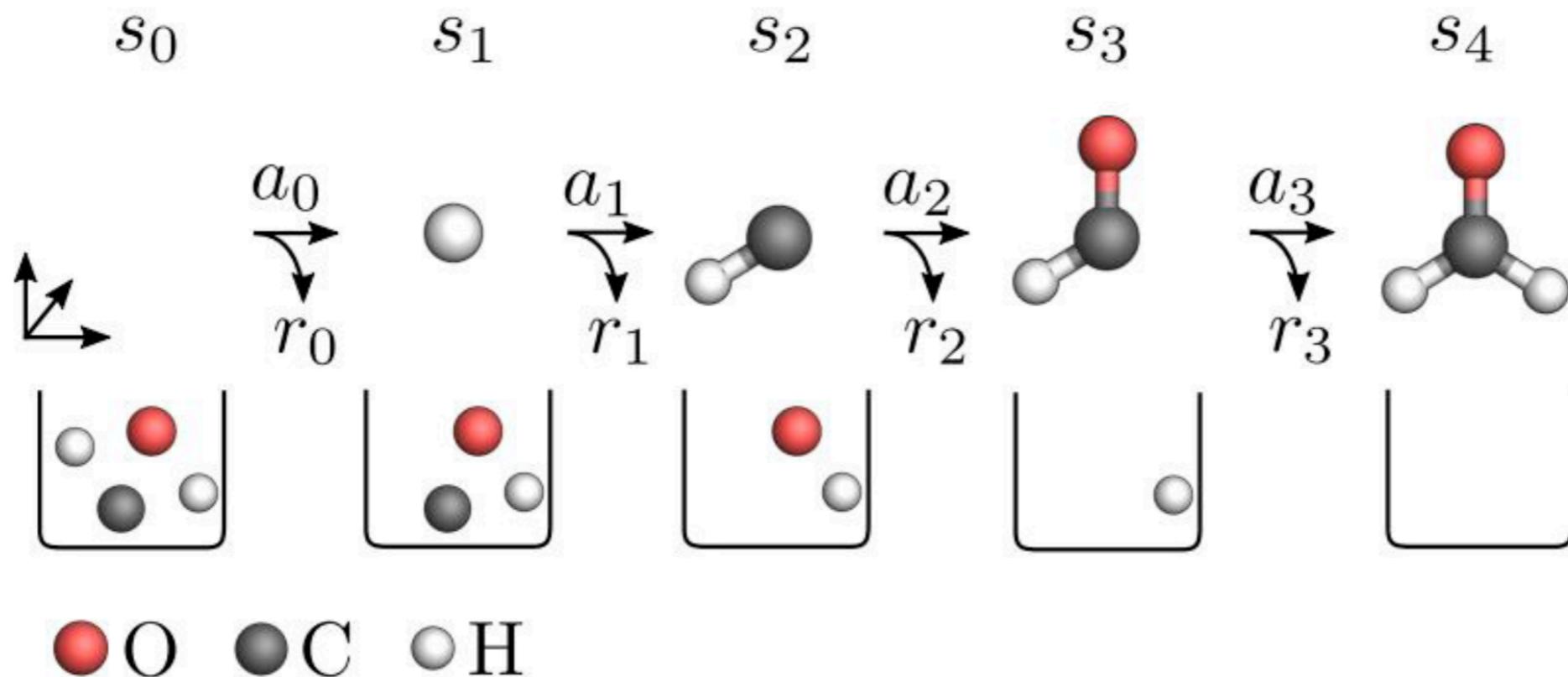
Proposed method

Reinforcement Learning

- Proximal Policy Gradient(PPO) used



Proposed method



Proposed method

State

$$s_t = (C_t, \beta_t)$$

Canvas $C_t = C_0 \cup \{(e_i, x_i)\}_{i=0}^{t-1}$, $C_0 = \emptyset$

Element $e_i \in \{H, C, N, O, \dots\}$ and position $x_i \in \mathbb{R}^3$

$$\text{Bag } \beta_t = \{(e, m(e))\}$$

Transition function

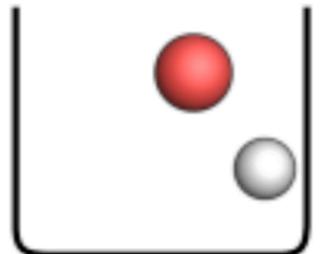
Deterministic $T(s_t, a_t)$

$$s_{t+1} = (C_{t+1}, \beta_{t+1}) \text{ with } \beta_{t+1} = \beta^t \setminus e_t$$

Canvas



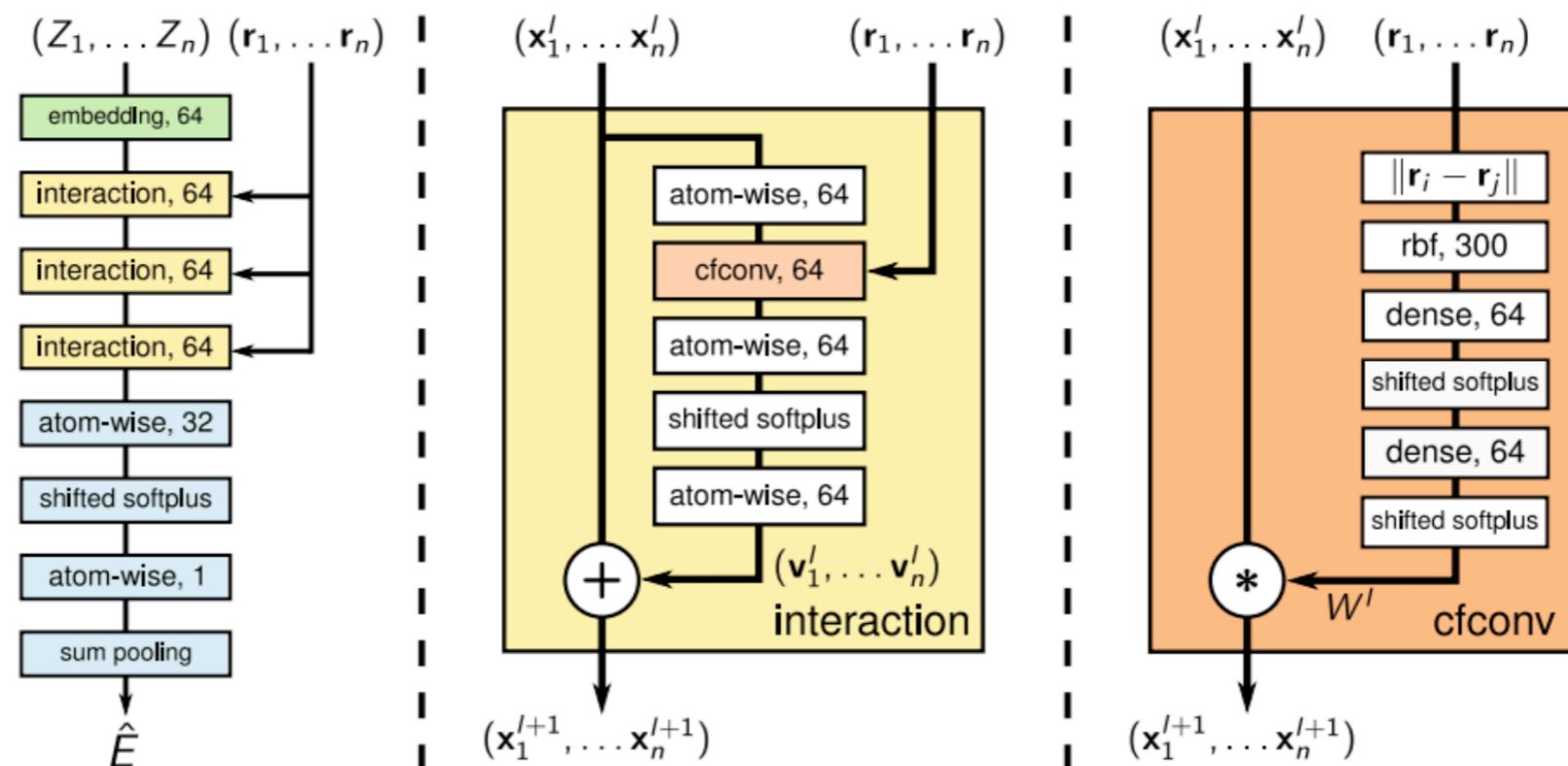
Bag



Proposed method

State Embedding

$$\tilde{s} = [\tilde{C}, \tilde{\beta}], \tilde{C} = SchNet(C), \tilde{\beta} = MLP_{\beta}(\beta)$$

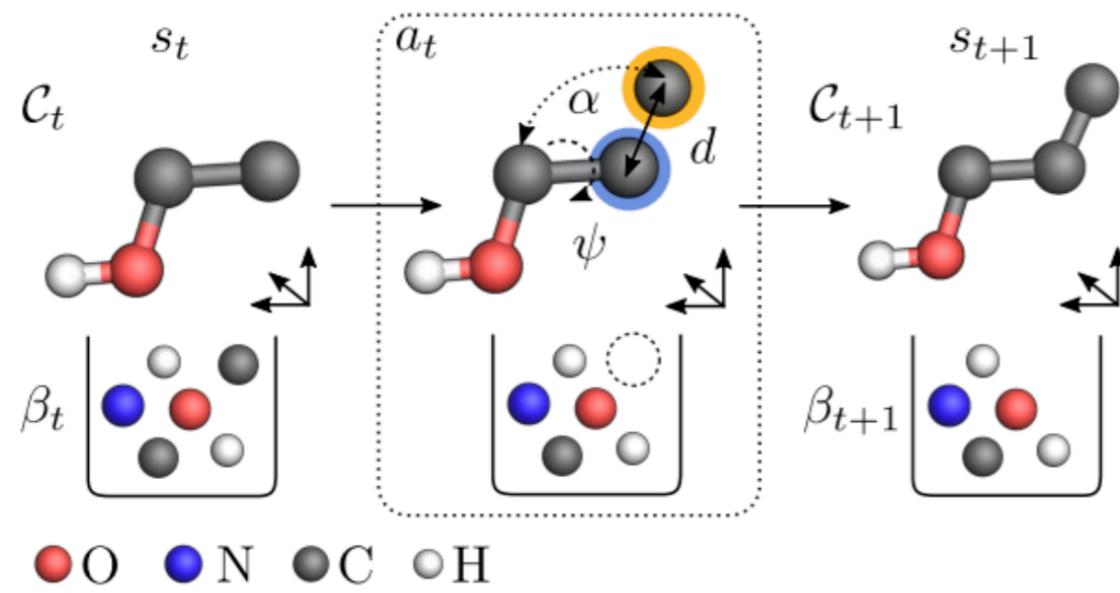


Proposed method

Action

- Select focal atom
 - Masking: set probabilities to zero and re-normalize the categorical distribution
 - Sample relative position

$$\pi_{\theta}(\psi, \alpha, d, e, f | s) = p(\psi, \alpha, d | e, f, s) \times p(e | f, s) \times p(f | s)$$



$$h_f = \text{MLP}_f(\tilde{s}), \quad (4)$$

$$h_e = \text{MLP}_e(\tilde{s}_f), \quad (5)$$

$$\mu_d, \mu_\alpha, \mu_\psi = \text{MLP}_{\text{cont}}(\tilde{s}_f, \mathbb{1}(e)), \quad (6)$$

$$f \sim \text{Cat}(f; h_f)$$

$$e \sim \text{Cat}(e; h_e)$$

$$d \sim N(d; \mu_d, \sigma_d^2)$$

$$\alpha \sim N(\alpha; \mu_\alpha, \sigma_\alpha^2)$$

$$\psi \sim N(\psi; \mu_\psi, \sigma_\psi^2)$$

Proposed method

Reward

$$\Delta E(s_t, a_t) = E(C_{t+1}) - [E(C_t) + E(e_t)]$$
$$r(s_t, a_t) = -\Delta E(s_t, a_t)$$

Energy calculation

- DFT is expensive
 - Take minutes or hours
- Semi-empirical method: SPARROW
 - Take milliseconds

Proposed method

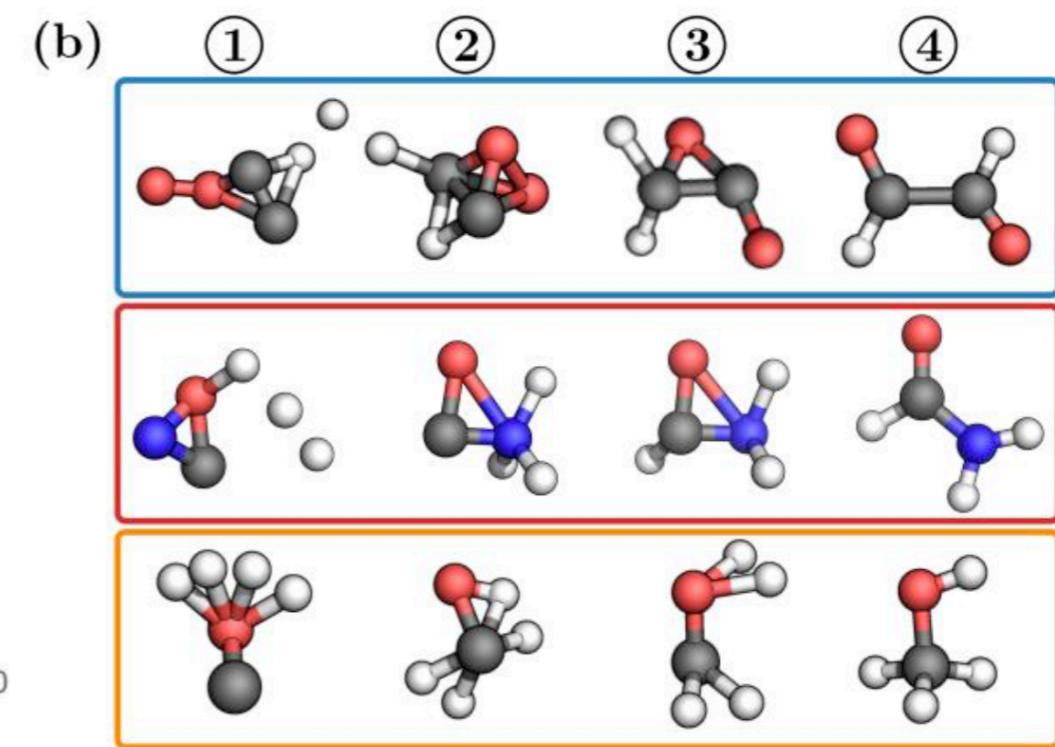
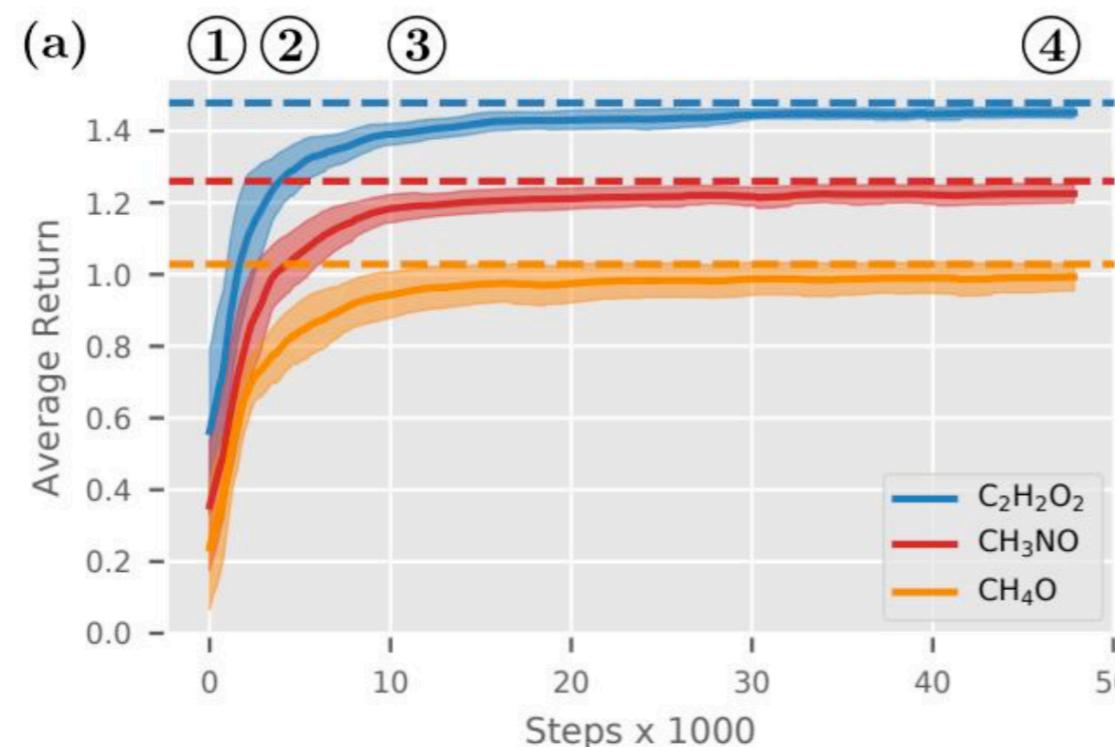
Critic

- Average over all atoms within the embeddings

$$V_\phi(s) = \text{MLP}_\phi\left(\frac{1}{n(C)} \sum_{i=1}^{n(C)} \tilde{s}_i\right)$$

Experiments

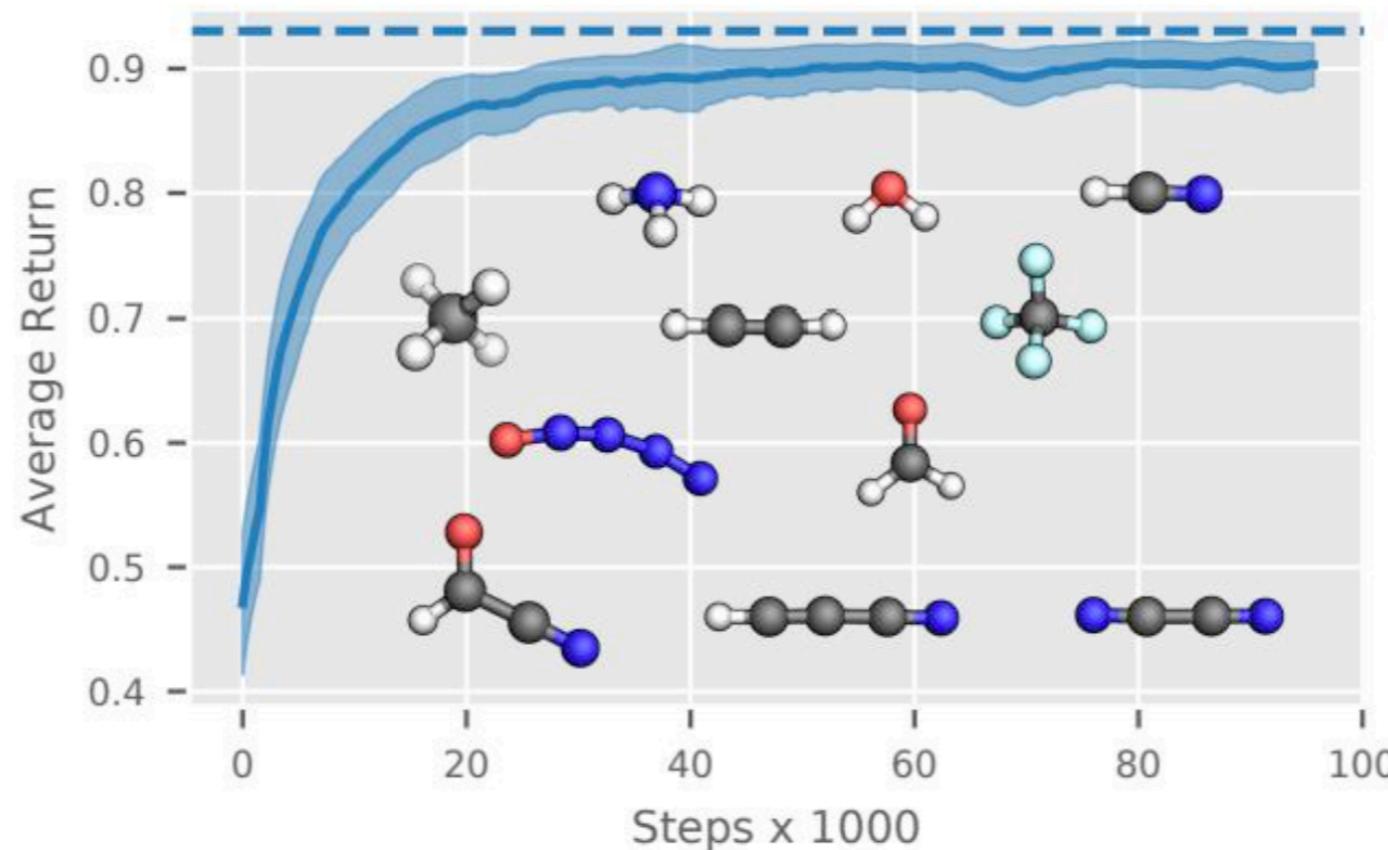
Single-bag



Experiments

Multi-bag

- Bag(molecule) is randomly selected before each episode



감사합니다