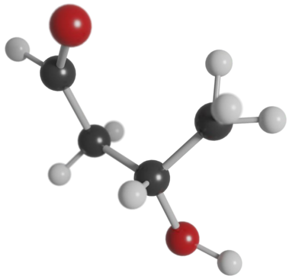
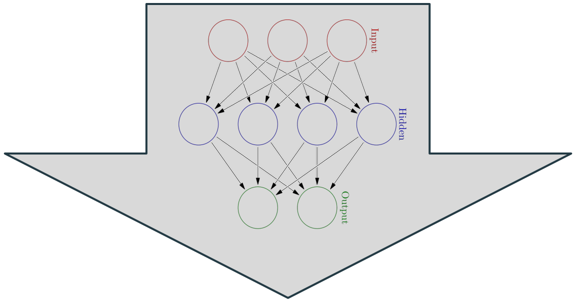
# LATTE: an atomic environment descriptor based on Cartesian tensor contractions

Franco Pellegrini, Stefano de Gironcoli, Emine Kucukbenli

SISSA

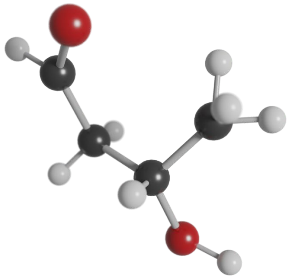
# Machine Learned Interatomic Potentials (MLIPs)

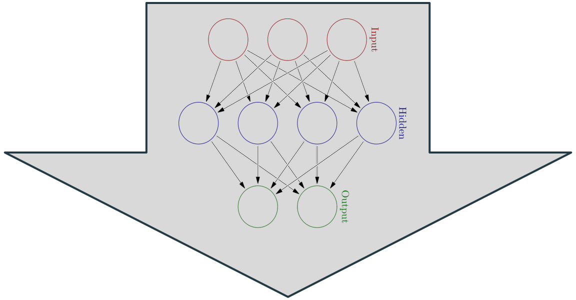





E(→F)

# Machine Learned Interatomic Potentials (MLIPs)

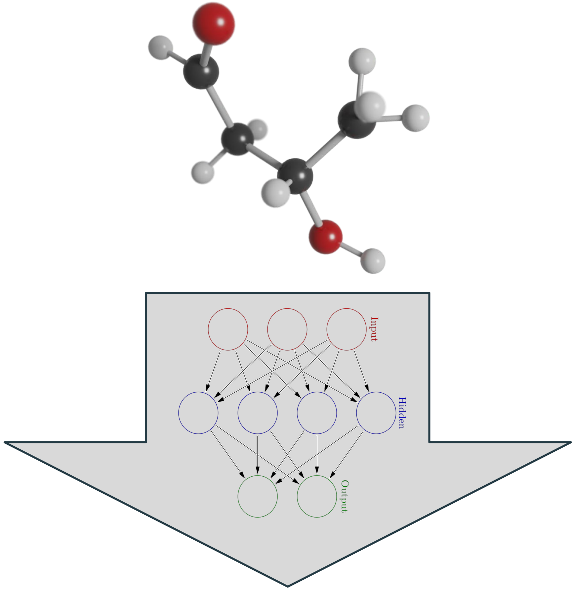





$E(\rightarrow F)$

- How to feed the structure to ML model?
    - It should extend to different systems
    - We want exact symmetries

# Machine Learned Interatomic Potentials (MLIPs)



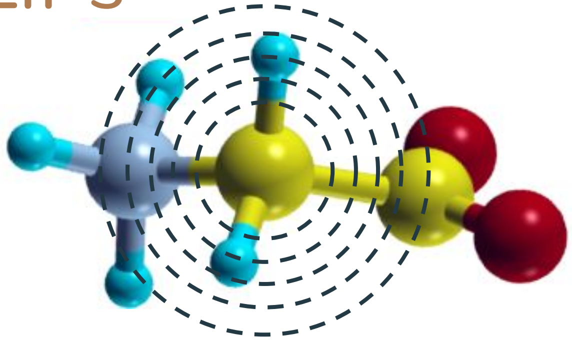

E(→F)

- How to feed the structure to ML model?
  - It should extend to different systems
  - We want exact symmetries

➢ Sum of atomic energies based on local environments
  - How to describe it?
  - How to fit?

# (biased) evolution of MLIPs

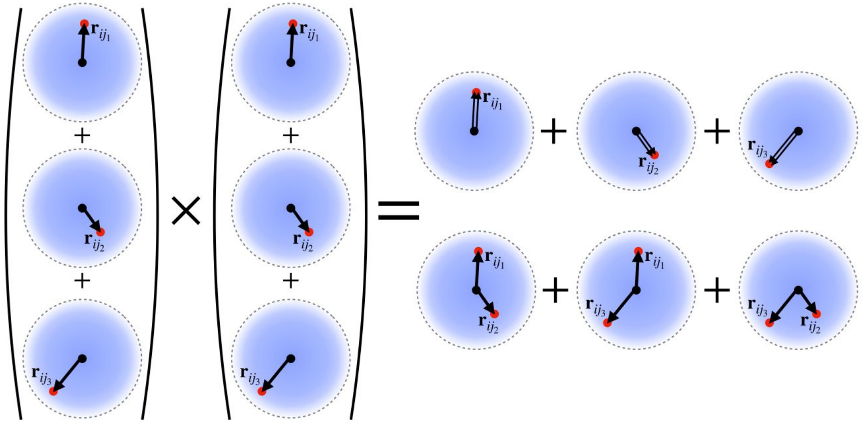
- Fixed descriptor + NN (Behler-Parrinello)
  - Sampling 2, 3 bodies (bad scaling)
  - Decent accuracy with enough data



# (biased) evolution of MLIPs

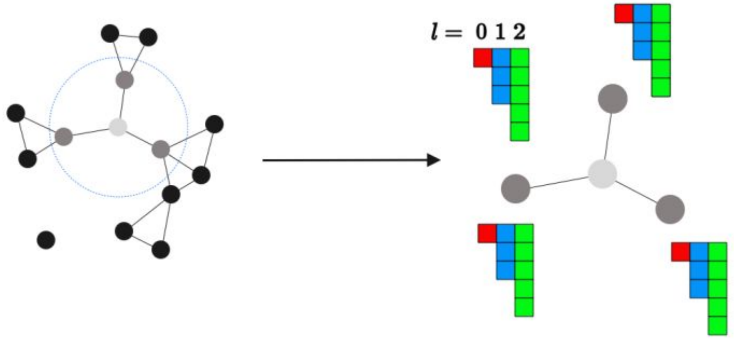
- Fixed descriptor + NN (Behler-Parrinello)
  - Sampling 2, 3 bodies (bad scaling)
  - Decent accuracy with enough data

- Better descriptors (ACE, MTP, …)
  - Higher body order at low cost
  - A linear model is enough



# (biased) evolution of MLIPs

- Fixed descriptor + NN (Behler-Parrinello)
  - Sampling 2, 3 bodies (bad scaling)
  - Decent accuracy with enough data
- Better descriptors (ACE, MTP, …)
  - Higher body order at low cost
  - A linear model is enough
- Equivariant graph NN (Nequip, MACE, …)
  - Message passing
  - Best accuracy/data efficiency but more expensive



# (biased) evolution of MLIPs

- Fixed descriptor + NN (Behler-Parrinello)
  - Sampling 2, 3 bodies (bad scaling)
  - Decent accuracy with enough data
- Better descriptors (ACE, MTP, …)
  - Higher body order at low cost
  - A linear model is enough
- Equivariant graph NN (Nequip, MACE, …)
  - Message passing
  - Best accuracy/data efficiency but more expensive

Do we really need message passing increasing receptive field? (see Allegro)
Can we do better with invariants + MLP?

# Better descriptors (but linear models):

## Atomic Cluster Expansion

## Moment Tensor Potential

Sample local environment with radial functions and

Spherical Harmonics

$$A_{z_i,znlm} = \langle \rho_i^z | \phi_{nlm}^{z_i z} \rangle = \sum_{\substack{j \\ \text{where } z_j = z}} \phi_{nlm}^{z_i z}(\mathbf{r}_{ji})$$

Cartesian Tensors

$$M_{\mu,\nu}(\mathfrak{n}_i) = \sum_j f_\mu(|r_{ij}|, z_i, z_j)\, \underbrace{\boldsymbol{r}_{ij} \otimes \ldots \otimes \boldsymbol{r}_{ij}}_{\nu \text{ times}};$$

Product of $N$ functions to obtain $N$+1 bodies terms:

$$A_{z_i\mathbf{v}} = \prod_{t=1}^{\nu} A_{z_i v_t}, \quad \mathbf{v} = (v_1, \ldots, v_\nu)$$

$$L_{\mu,\boldsymbol{\nu}} = \prod_{t=1}^{N} M_{\mu_t,\nu_t}$$

Symmetrize to have invariant contribution

$$B_{z_i\mathbf{v}} := \int_{\hat{R}\in O(3)} \prod_{t=1}^{\nu} A_{z_i v_t}(\{\hat{R}\mathbf{r}_{ij}\})\,d\hat{R} = \sum_{\mathbf{v}'} C_{\mathbf{v}\mathbf{v}'} A_{z_i\mathbf{v}'}$$

Contract indices to have invariant terms

$$L_\mu = \sum_{\boldsymbol{\nu}} L_{\mu,\boldsymbol{\nu}}$$

# Cartesian contractions: LATTE
## (Local Atomic Tensors Trainable Expansion)

Mixed species, radially localized tensor sum:

$$A_{i,u}^{\alpha\beta} = \sum_{j \in \mathcal{N}(i)} \sigma_u(s_i, s_j)\, f_u(r_{ij})\, r_{ij}^{\alpha} \otimes r_{ij}^{\beta}$$

# Cartesian contractions: LATTE

## (Local Atomic Tensors Trainable Expansion)

Mixed species, radially localized tensor sum:

$$A_{i,u}^{\alpha\beta} = \sum_{j \in \mathcal{N}(i)} \underbrace{\boxed{\sigma_u(s_i, s_j)}\,\boxed{f_u(r_{ij})}}_{\text{LEARNABLE}}\, r_{ij}^{\alpha} \otimes r_{ij}^{\beta}$$

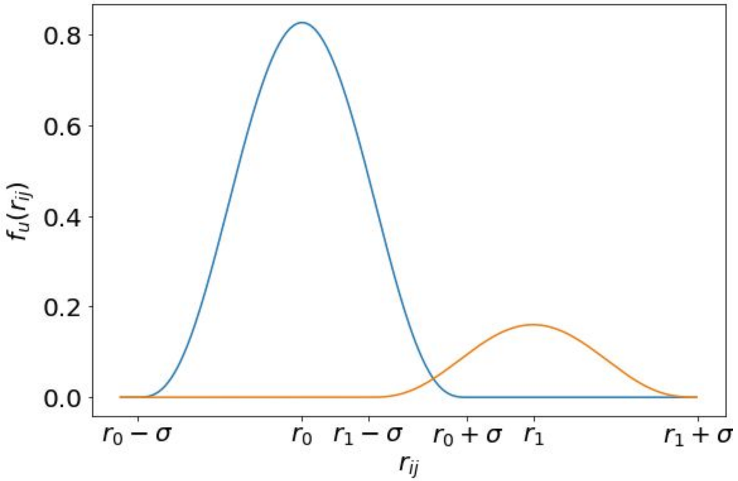
# Cartesian contractions: LATTE

## (Local Atomic Tensors Trainable Expansion)

Mixed species, radially localized tensor sum:

$$A^{\alpha\beta}_{i,u} = \sum_{j \in \mathcal{N}(i)} \boxed{\sigma_u(s_i, s_j)}\, f_u(r_{ij})\; r^{\alpha}_{ij} \otimes r^{\beta}_{ij}$$

$$f_u(r) = \frac{1}{r_u^2}\left[\mathrm{ReLU}\left(1 - \left(\frac{r - \boxed{r_u}}{\boxed{w_u}}\right)^2\right)\right]^3$$



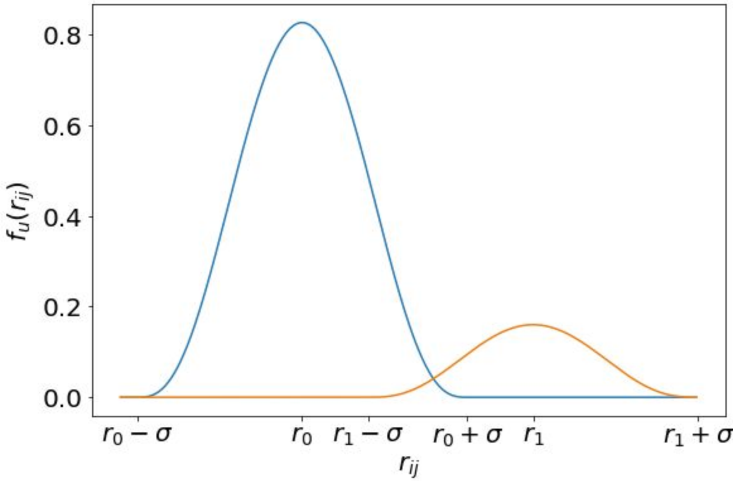
# Cartesian contractions: LATTE

## (Local Atomic Tensors Trainable Expansion)

Mixed species, radially localized tensor sum:

$$A_{i,u}^{\alpha\beta} = \sum_{j \in \mathcal{N}(i)} \sigma_u(s_i, s_j)\, f_u(r_{ij})\, r_{ij}^{\alpha} \otimes r_{ij}^{\beta}$$

$$f_u(r) = \frac{1}{r_u^2}\left[\mathrm{ReLU}\left(1 - \left(\frac{r - r_u}{w_u}\right)^2\right)\right]^3$$




Tensor contraction to invariant descriptor element:

$$B_{i,u}^{(\alpha,\beta,\alpha\beta)} = \sum_{\alpha\beta} A_{i,u_1}^{\alpha}\, A_{i,u_2}^{\beta}\, A_{i,u_3}^{\alpha\beta}$$

VECTOR

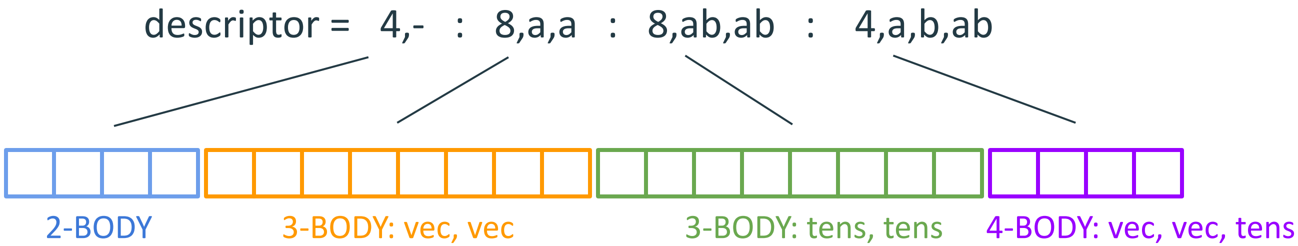
VECTOR

TENSOR

# Freely specify descriptor terms

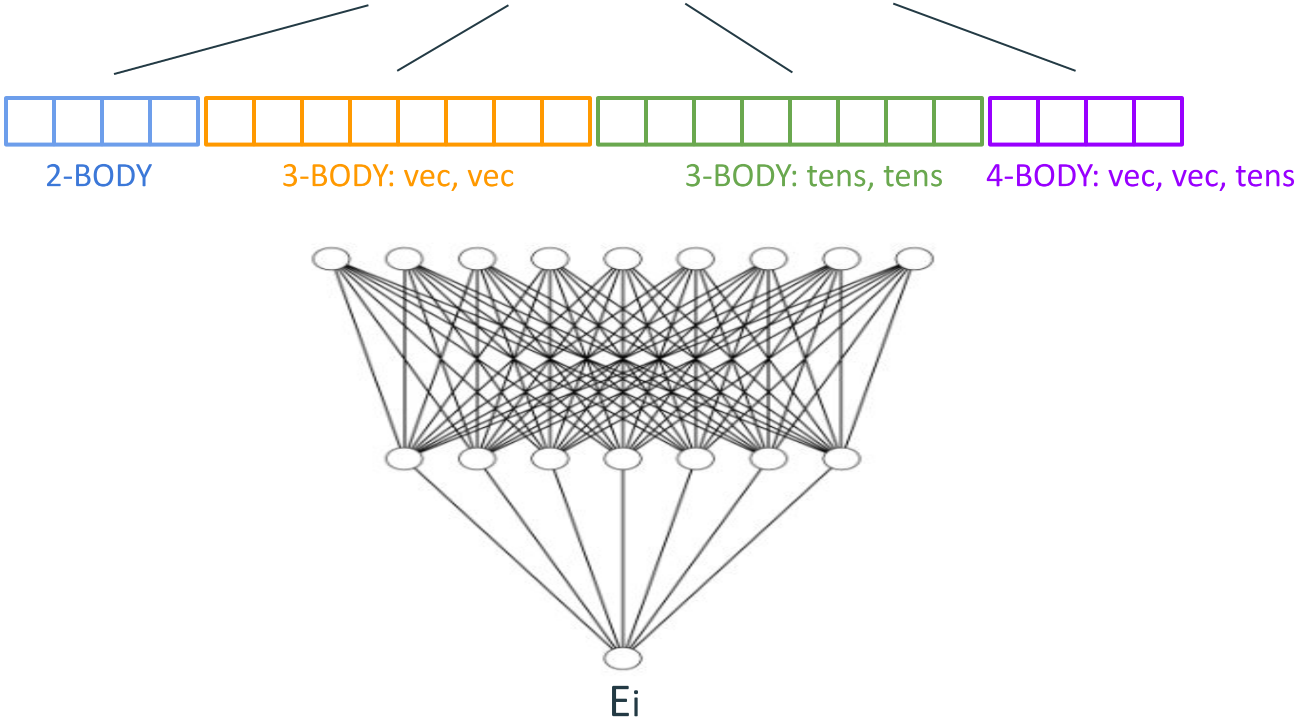
descriptor = 4,- : 8,a,a : 8,ab,ab : 4,a,b,ab

# Freely specify descriptor terms

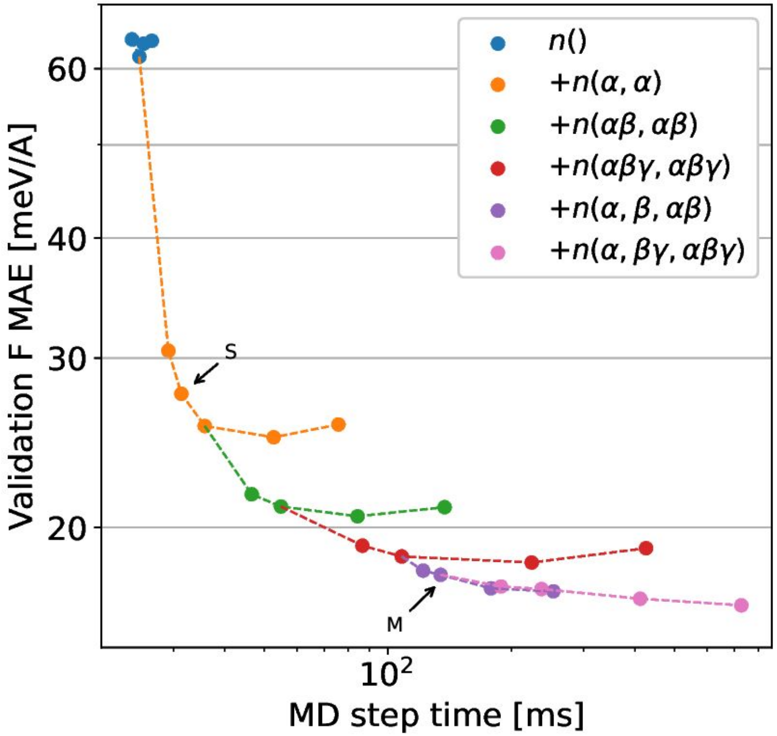


# Freely specify descriptor terms

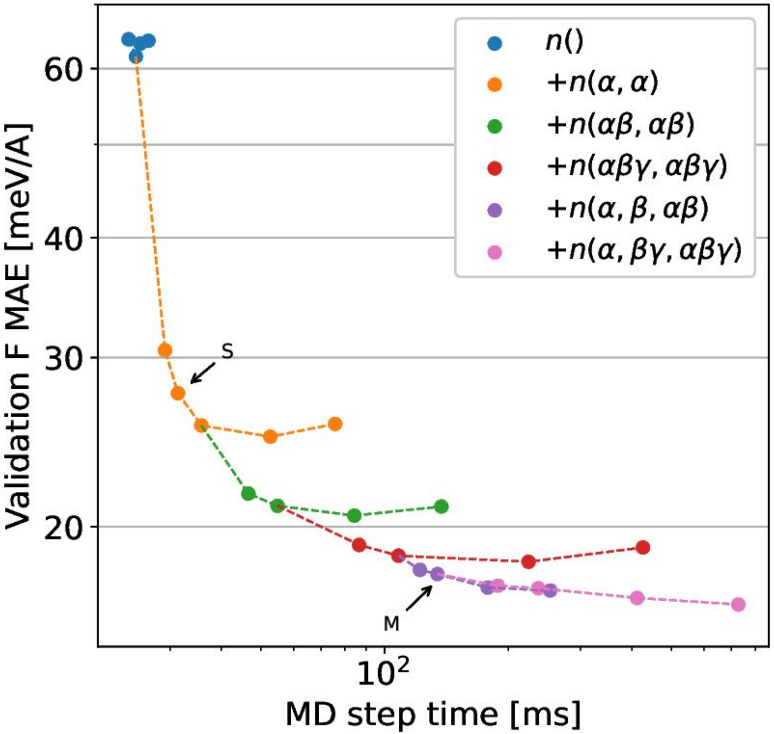
descriptor = 4,- : 8,a,a : 8,ab,ab : 4,a,b,ab



# Scaling with descriptor size/shape: rMD17



# Scaling with descriptor size/shape: rMD17




| | | ACSF[6] | ACE[20] | LATTE small | LATTE medium | SOTA |
|---|---|---|---|---|---|---|
| **Aspirin** | E | 10.6 | 6.1 | 9.5 | 5.7 | 2.2[25] |
| | F | 32.9 | 17.9 | 27.5 | 17.8 | 6.6 |
| **Azobenzene** | E | 5.8 | 3.6 | 7.0 | 3.3 | 1.2[26] |
| | F | 18.4 | 10.9 | 19.5 | 10.4 | 2.6 |
| **Benzene** | E | 1.0 | 0.04 | 0.73 | 0.42 | 0.3[26] |
| | F | 5.4 | 0.5 | 4.6 | 1.1 | 0.2 |
| **Ethanol** | E | 2.9 | 1.2 | 2.5 | 1.2 | 0.4[25,26] |
| | F | 16.5 | 7.3 | 14.2 | 6.3 | 2.1 |
| **Malonaldehyde** | E | 4.0 | 1.7 | 3.9 | 1.7 | 0.6[26] |
| | F | 24.3 | 11.1 | 20.4 | 10.6 | 3.6 |
| **Naphthalene** | E | 3.0 | 0.9 | 3.6 | 1.6 | 0.2[26] |
| | F | 13.2 | 5.1 | 15.1 | 6.3 | 0.9 |
| **Paracetamol** | E | 6.3 | 4.0 | 5.8 | 3.5 | 1.3[25] |
| | F | 22.0 | 12.7 | 20.9 | 12.6 | 4.8 |
| **Salicylic acid** | E | 4.1 | 1.8 | 4.1 | 2.4 | 0.9[26] |
| | F | 19.4 | 9.3 | 18.3 | 9.6 | 2.9 |
| **Toluene** | E | 3.9 | 1.1 | 3.6 | 1.5 | 0.5[25] |
| | F | 15.9 | 6.5 | 15.3 | 6.6 | 1.5 |
| **Uracil** | E | 2.4 | 1.1 | 3.2 | 1.1 | 0.6[26] |
| | F | 13.7 | 6.6 | 12.2 | 5.7 | 1.8 |

# Multi species: SPICE

- Drug-like molecules and peptides
- **15 species** (H, Li, C, N, O, F, Na, Mg, P, S, Cl, K, Ca, Br, I)
- 19k systems, **1.1M** configurations

- **SOTA: Allegro**: **25 meV/Å** validation force MAE, speed ~10 µs/step/atom
- **DeepMD** (transformer emb.): **233 meV/Å** validation force RMSE, speed ~20 µs/step/atom
- **LATTE+NN**: down to **45 meV/Å** validation force MAE
  - JAX-MD: DHFR (23k atoms) on 1 A100 @1.7 µs/step/atom

# Conclusions

- Simple and fast descriptor + NN can reach good accuracy
- Can handle multiple bodies and species with good performance
- Implemented in PANNA package (JAX)

# The future

- Improve performance
- Better species embedding (generalization, ask Mina)

Thank you

arXiv:2405.08137