Crystalformer: Infinitely Connected Attention for Periodic Structure Encoding

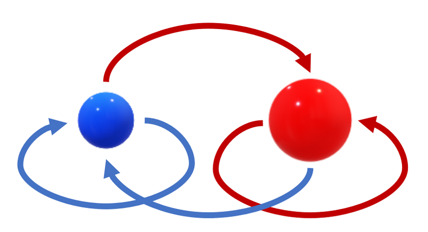
TL; DR: Propose a transformer-based encoder for crystal property prediction by mimicking energy calculations (interatomic potential summations) in physics via self-attention

Tatsunori Taniai [1] Ryo Igarashi [1] Yuta Suzuki [2] Naoya Chiba [3] Kotaro Saito [4,5] Yoshitaka Ushiku [1] Kanta Ono [5]

[1] OMRON SINIC X Corporation [2] Toyota Motor Corporation [3] Tohoku University [4] Randeft Inc. [5] Osaka University

Transformers are good for molecules

Key is **fully-connected self-attention** for finite atoms, with **relative position representations** (scalar ϕ and vector $\boldsymbol{\psi}$) encoding spatial relations between atom pairs.



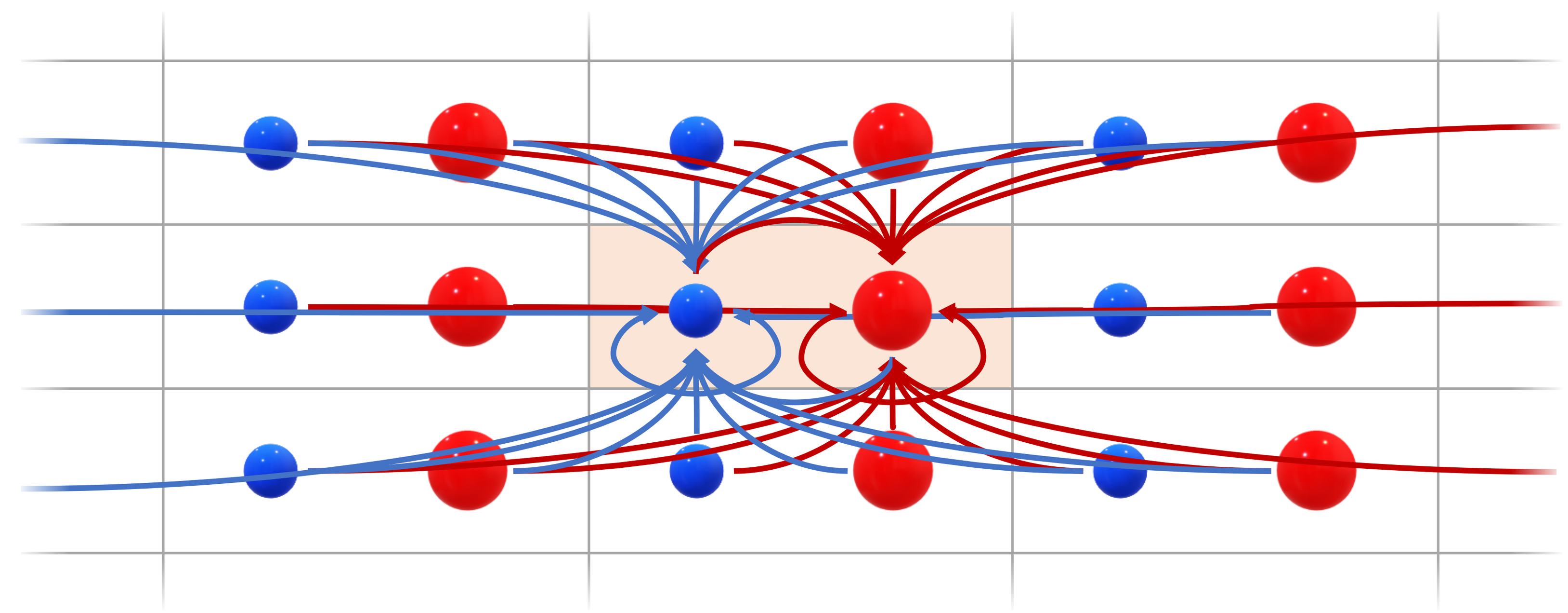
$$\boldsymbol{y}_i = \frac{1}{Z_i}\sum_{j=1}^{N} \exp\left(\boldsymbol{q}_i^T \boldsymbol{k}_j/\sqrt{d_K} + \phi_{ij}\right)\left(\boldsymbol{v}_j + \boldsymbol{\psi}_{ij}\right)$$

(Similar to *Graphormer* by Ying et al., 2021)

But transformers for crystal are very rare.

Why not use transformers for crystals?

Let finite atoms i in a unit cell **attend to infinite atoms** $j(\boldsymbol{n})$ in periodically repeated unit cells $\boldsymbol{n}$.



$$\boldsymbol{y}_i = \frac{1}{Z_i}\sum_{j=1}^{N}\sum_{\boldsymbol{n}\in\mathbb{Z}^3} \exp\left(\boldsymbol{q}_i^T \boldsymbol{k}_j/\sqrt{d_K} + \phi_{ij(\boldsymbol{n})}\right)\left(\boldsymbol{v}_j + \boldsymbol{\psi}_{ij(\boldsymbol{n})}\right)$$

We call it the ***infinitely connected attention***.

Infinitely connected attention can be

Interpreted as *Neural Potential Summation*
by introducing **distance decay attention**

$$\exp\left(\phi_{ij(\boldsymbol{n})}\right) = \exp\left(-\frac{\left\|\boldsymbol{p}_{j(\boldsymbol{n})} - \boldsymbol{p}_i\right\|^2}{2\sigma_i^2}\right)$$

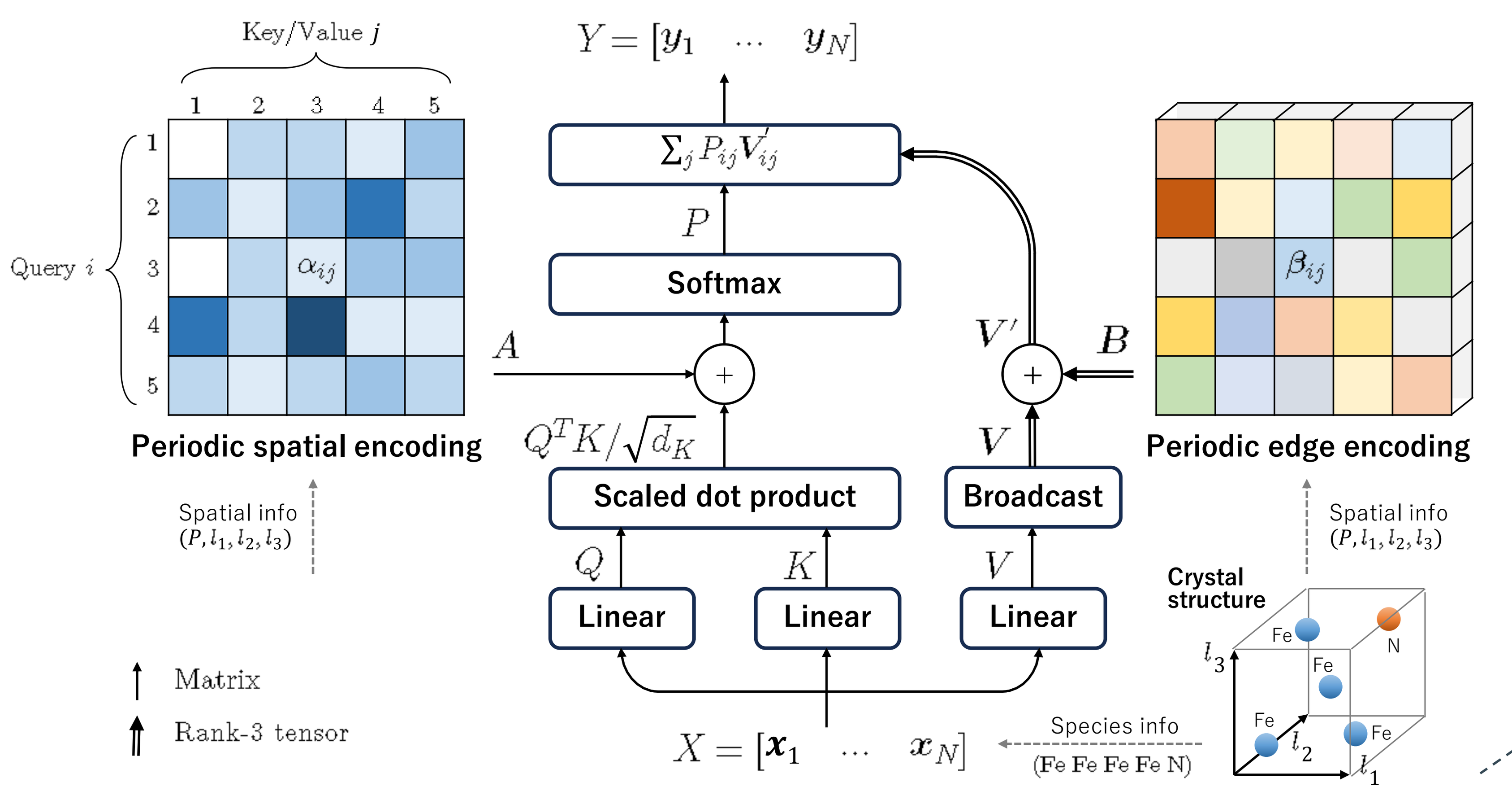
Performed just like standard self-attention

$$\boldsymbol{y}_i = \frac{1}{Z_i}\sum_{j=1}^{N} \exp\left(\boldsymbol{q}_i^T \boldsymbol{k}_j/\sqrt{d_K} + \alpha_{ij}\right)\left(\boldsymbol{v}_j + \boldsymbol{\beta}_{ij}\right)$$

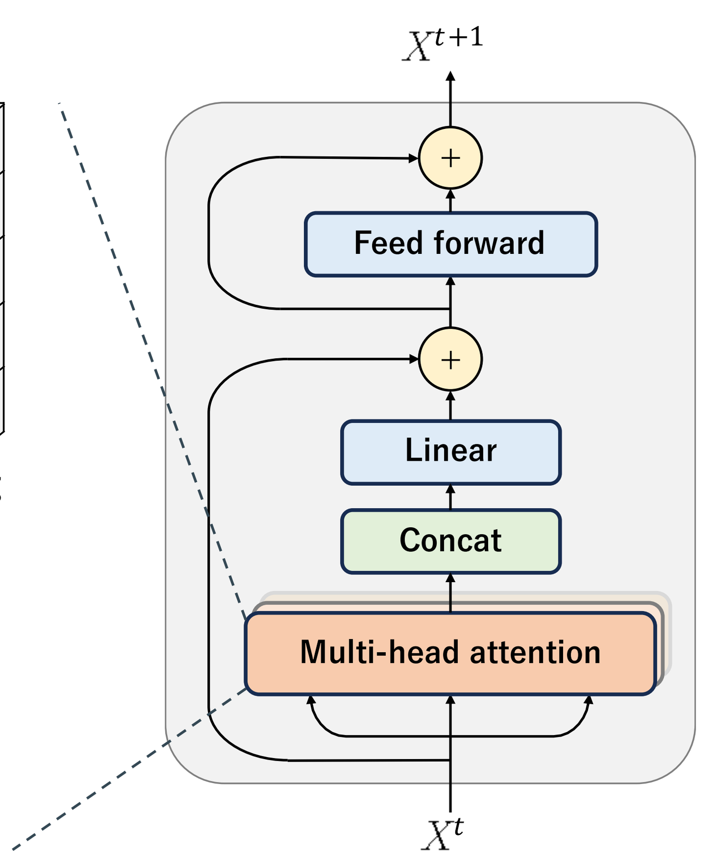
where $\alpha_{ij} = \log \sum_{\boldsymbol{n}} \exp\left(\phi_{ij(\boldsymbol{n})}\right)$

$$\boldsymbol{\beta}_{ij} = \sum_{\boldsymbol{n}} \exp\left(\phi_{ij(\boldsymbol{n})} - \alpha_{ij}\right)\boldsymbol{\psi}_{ij(\boldsymbol{n})}$$

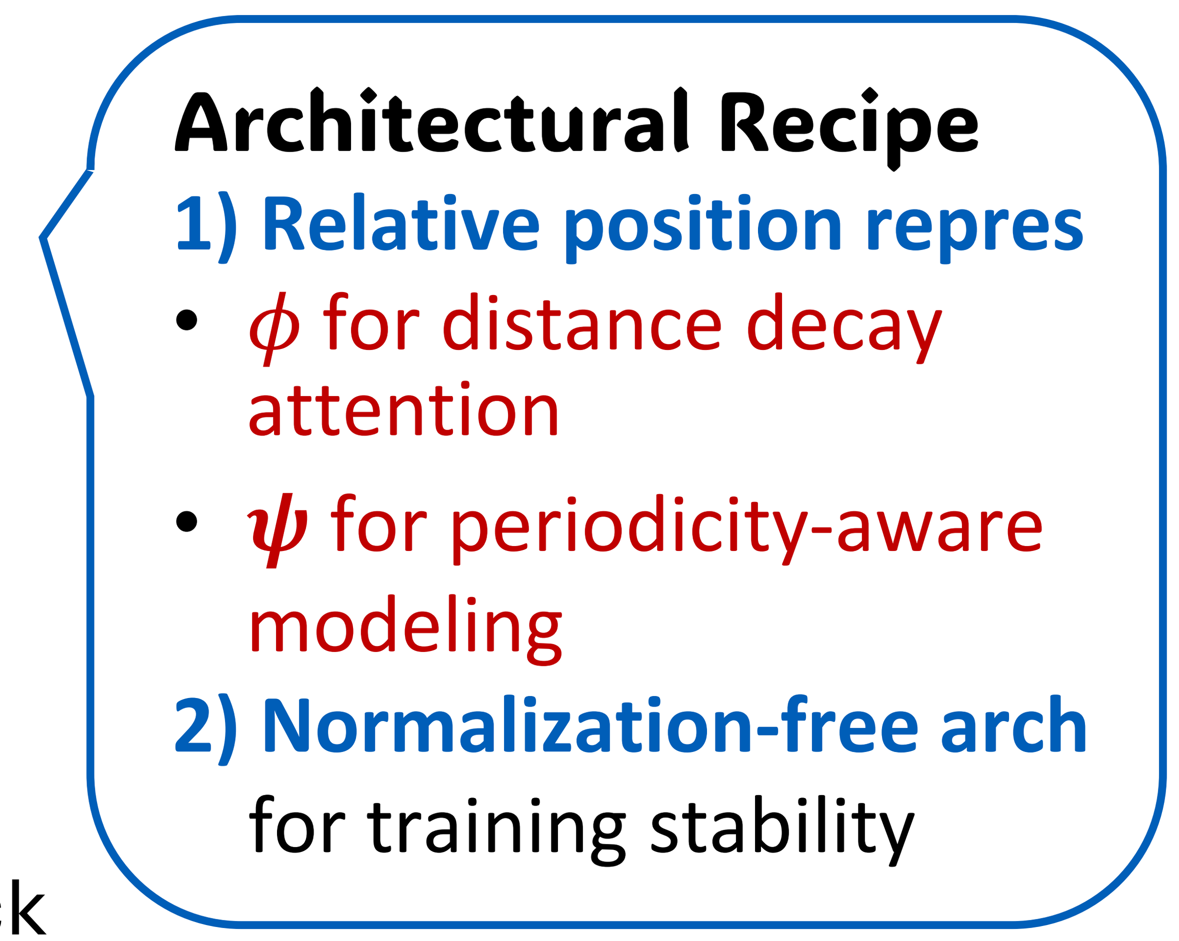
Closely follow original Transformer architecture



Proposed self-attention layer



Self-attention block



Architectural Recipe

1) Relative position repres

- ϕ for distance decay attention
- $\boldsymbol{\psi}$ for periodicity-aware modeling

2) Normalization-free arch for training stability

Results

Beats most existing methods!

	Materials Project (MEGNET's snapshot)				JARVIS-DFT 3D				
	E form eV/atom	BG eV	Bulk mod. log (GPa)	Shear mod. log (Gpa)	E form eV/atom	E total eV/atom	BG (OPT) eV	BG (MBJ) eV	E hull eV
CGCNN	0.031	0.292	0.047	0.077	0.063	0.078	0.20	0.41	0.17
SchNet	0.033	0.345	0.066	0.099	0.045	0.047	0.19	0.43	0.14
MEGNET	0.030	0.307	0.060	0.099	0.047	0.058	0.145	0.34	0.084
GATGNN	0.033	0.280	0.045	0.075	0.047	0.056	0.17	0.51	0.12
ALIGNN	0.022	0.218	0.051	0.078	0.0331	0.037	0.142	0.51	0.076
Matformer	0.021	0.211	0.043	0.073	0.0325	0.035	0.137	0.30	0.064
PotNet	**0.0188**	0.204	0.040	**0.065**	**0.0294**	**0.032**	**0.127**	**0.27**	0.055
Ours	0.0198	**0.201**	**0.0399**	0.0692	0.0319	0.0342	0.131	0.275	**0.0482**

More efficient and light-weight!

	Type	Time/ep	Test/mat.	# params	# blk. params
PotNet	GNN	**43 s**	313 ms	1.8 M	527 K
Matformer	Transformer	60 s	20.4 ms	2.9 M	544 K
Ours	Transformer	**32 s**	**6.6 ms**	**853 K**	**206 K**

What's more in paper

- Fourier-space attention for long-range interaction
- Importance of $\boldsymbol{\psi}$ term