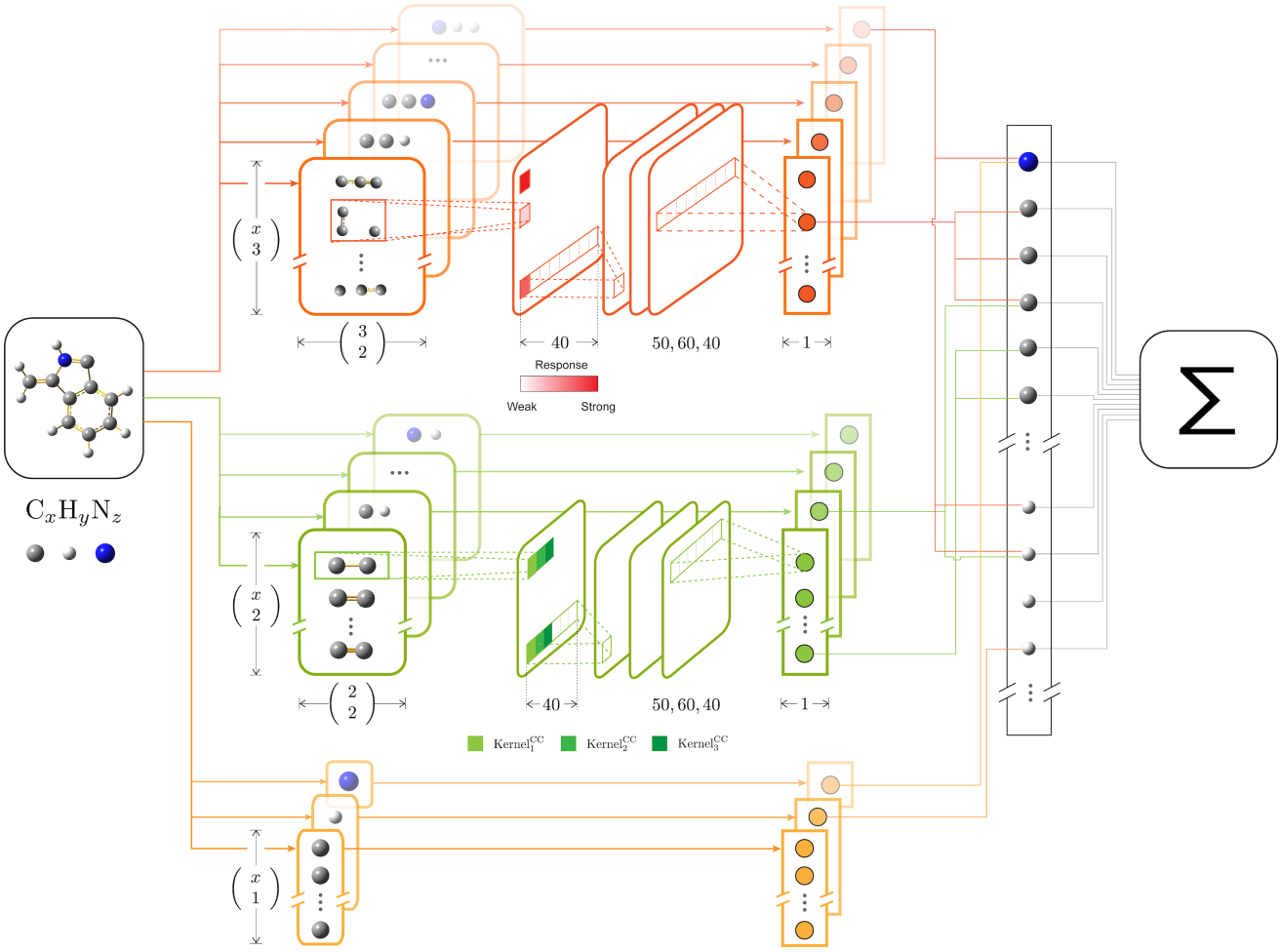
The theory and implementation of kCON

Xin Chen

October 18, 2017

1 Overview

kCON is scalable and transferable deep learning framework for chemistry with the ability to provide insight into atomistic structures of varying stoichiometry from small and scrap training sets. kCON is built upon convolutional neural networks, or more specifically, 1D-convolutional neural networks with 1x1 convolutions.



2 Energy

The total energy output by kCON is modeled under the many-body expression scheme:

$$\begin{aligned} E^{total} &= E^{(k=1)} + E^{(k=2)} + E^{(k=3)} + \cdots \\ &= \sum_{a}^{C_1^N} E_a^{(k=1)} + \sum_{a,b}^{C_2^N} E_{ab}^{(k=2)} + \sum_{a,b,c}^{C_3^N} E_{abc}^{(k=3)} + \cdots \\ &= \sum_{a}^{C_1^N} \mathbf{F}^{(k=1)}(A_a) + \sum_{a,b}^{C_2^N} \mathbf{F}^{(k=2)}(r_{ab}, A_a, A_b) + \sum_{a,b,c}^{C_3^N} \mathbf{F}^{(k=3)}(r_{ab}, r_{bc}, r_{ac}, A_a, A_b, A_c) + \cdots \end{aligned} \tag{1}$$

where r_{ab} denotes the interatomic distance between atom a and b, A_A represents the element of atom a, $\mathbf{F}^{(k)}$ is an arbitrary function that outputs the energy for k-body inputs.

For most cases, equation 1 can be truncated at $k = 3$ because higher order terms contribute far less to the total energy while require much more computational resources as C_k^N roughly scales as $\mathcal{O}(N)$ and the 2-body features alone (bonds) cannot uniquely describe structures. The interatomic distances, r_{ab}, range from 0 to $+\infty$. To normalize the distances, the Laplacian kernel is used:

$$z_{ab} = \exp\left(-\frac{r_{ab}}{L_a + L_b}\right) = \exp\left(-\frac{r_{ab}}{L_{ab}}\right) \tag{2}$$

where L_a is the covalent radius of element A_a. For non periodic molecules the Pÿÿkko radii are used as the covalent radii and for periodic structures the CSD covalent bonds should be used.

Then the MBE expression becomes:

$$E^{total} = \sum_{a}^{C_1^N} \mathbf{F}^{(k=1)}(A_a) + \sum_{a,b}^{C_2^N} \mathbf{F}^{(k=2)}(z_{ab}, A_a, A_b) + \sum_{a,b,c}^{C_3^N} \mathbf{F}^{(k=3)}(z_{ab}, z_{bc}, z_{ac}, A_a, A_b, A_c) \tag{3}$$

Now what we need is to model the three $\mathbf{F}^{(k)}$.

2.1 One-body

In kCON, the 1-body terms are expressed by a linear model:

$$\mathbf{F}^{(k=1)}(A_a) = E^{A_a} \tag{4}$$

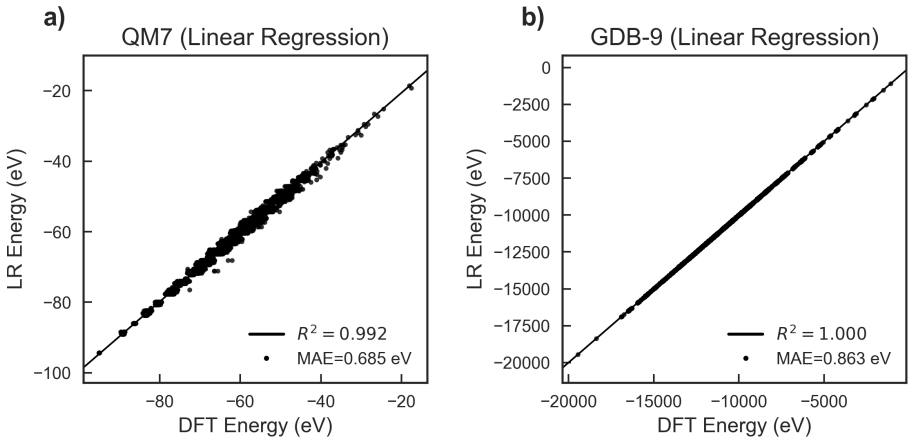
$$\sum_{a}^{C_1^N} \mathbf{F}^{(k=1)}(A_a) = \sum_{a}^{N} E^{A_a} = \sum_{A_a} n^{A_a} E^{A_a} \tag{5}$$

where E^{A_a} represents the learnable 1-body energy of element A_a and n^{A_a} is the total number of element A_a in the structure. As an example, the 1-body energy of a C_9H_7N molecule should be:

$$E^{(k=1)} = 9E^{\mathrm{C}} + 7E^{\mathrm{H}} + E^{\mathrm{N}} \tag{6}$$

where E^{C}, E^{H} and E^{N} are all trainable parameters.

The reason to use a simple linear model to express the 1-body term is that we observed that for chemically reasonable structures, for example the organic compounds in QM7 and GDB-9 datasets that can be synthesized, even the linear model can give acceptable energy predictions.



2.2 Two/three-body

Modeling the 2-body and 3-body interactions are more complicated because if we want our model transferable, the model must support accepting variable-length inputs so that a single model can be used to process atomistic structures of varying size and composition.

So here, we use 1D-convolutional neural networks to model $\mathbf{F}^{(k=2)}$ and $\mathbf{F}^{(k=2)}$. For each k-body term (k-body interactions, e.g. Carbon-Carbon-Carbon), an independent CNN is used. So kCON is indeed a model composed of a collection of k-body CNNs and a linear model.

$$E^{(k=2)} = \sum_{a,b}^{C_2^N} \mathrm{CNN}^{\mathrm{A}_a\mathrm{A}_b}(r_{ab}) \tag{7}$$

$$E^{(k=3)} = \sum_{a,b,c}^{C_3^N} \mathrm{CNN}^{\mathrm{A}_a\mathrm{A}_b\mathrm{A}_c}(r_{ab}, r_{ac}, r_{bc}) \tag{8}$$

As an example, the C_9H_7N system has five 2-body terms (CC, CH, CN, HH, HN) and seven 3-body terms (CCC, CCH, CCN, CHH, CHN, HHH, HHN). So the kCON model for C_9H_7N has 12 independent CNNs to train at the same time.

The activation function $\sigma(\cdot)$ used in kCON is leaky ReLU:

$$\sigma(x) = \begin{cases} x & \text{if } x \geq 0 \\ \alpha x & \text{else} \end{cases} \tag{9}$$

$$\frac{\partial\sigma(x)}{\partial x} = \begin{cases} 1 & \text{if } x \geq 0 \\ \alpha & \text{else} \end{cases} \tag{10}$$

$$\alpha = 0.2 \tag{11}$$

because leaky ReLU can give much better results compared with traditional activations functions like hyperbolic tangent (tanh) or sigmoid.

2.3 The input feature matrix

Now we can start building the inputs for kCON. Inputs for the 1-body part is very simple, so here we mainly focus on how to efficiently build input feature matrix for 2-body and 3-body terms.

As introduced before, kCON has a collection of k-body CNNs. But for any specific chemical system, the dimensions of its k-body terms should be fixed. For example, table 1 demonstrates the sizes of the k-body terms of the C_9H_7N system:

Instead of generating separated inputs for each k-body CNN, we can build a single input matrix and split it based on the dimensions. For a system composed of N atoms, the total number of input patterns should be:

Term	k	Expression	Dimension
CC	2	C_2^9	36
CH	2	$C_1^9 \cdot C_1^7$	63
CN	2	$C_1^9 \cdot C_1^1$	9
HH	2	C_2^7	21
HN	2	$C_1^7 \cdot C_1^1$	7
CCC	3	C_3^9	84
CCH	3	$C_2^9 \cdot C_1^7$	252
CCN	3	$C_2^9 \cdot C_1^1$	36
CHH	3	$C_1^9 \cdot C_2^7$	189
CHN	3	$C_1^9 \cdot C_1^7 \cdot C_1^1$	63
HHH	3	C_3^7	35
HHN	3	$C_2^7 \cdot C_1^1$	21
Total			816

Table 1: the dimensions of the k-body terms of C_9H_7N

$$C_3^N + C_2^N = C_3^{N+1} \tag{12}$$

and the total number of entries (element in the input feature matrix) should be:

$$C_3^N \cdot C_2^3 + C_2^N \cdot C_1^1 = \frac{N(N-1)^2}{2} \tag{13}$$

One can notice that for a system with N atoms the number of chemical patterns is C_3^{N+1}. So a more easy way can be used to construct the input feature matrix. This scheme is called **ghost atom scheme**. We can just temporarily append a ghost atom, denoted as **X**, to the original system and only keep the 3-body features. The cartesian coordinates of the ghost atom is always $(+\infty, +\infty, +\infty)$, so the interatomic distance between X and any real atom is $+\infty$. Thus, the Laplacian normalization, defined in equation 2, will transform all distances between **X** and real atoms to 0.

As an example shown, table 2 shows the dimensions of the 3-body terms of the C_9H_7NX system:

Term	k	Expression	Dimension
CCX	3	$C_2^9 \cdot C_1^1$	36
CHX	3	$C_1^9 \cdot C_1^7 \cdot C_1^1$	63
CNX	3	$C_1^9 \cdot C_1^1 \cdot C_1^1$	9
HHX	3	$C_2^7 \cdot C_1^1$	21
HNX	3	$C_1^7 \cdot C_1^1 \cdot C_1^1$	7
CCC	3	C_3^9	84
CCH	3	$C_2^9 \cdot C_1^7$	252
CCN	3	$C_2^9 \cdot C_1^1$	36
CHH	3	$C_1^9 \cdot C_2^7$	189
CHN	3	$C_1^9 \cdot C_1^7 \cdot C_1^1$	63
HHH	3	C_3^7	35
HHN	3	$C_2^7 \cdot C_1^1$	21
Total			816

Table 2: the dimensions of the k-body terms of C_9H_7NX

We can notice that the total number of input chemical patterns is unchanged. The total number of entries becomes:

$$C_3^{N+1} \cdot C_2^3 = \frac{N(N-1)^2}{2} + C_2^N \cdot 2 \tag{14}$$

so we need more space to store the input features but this is worthy as all features can be saved in a single matrix of shape $[C_3^{N+1}, C_2^3]$.

2.4 Permutational Invariance

The three spatial (translational, rotational and permutational) invariances must all be satisfied. Since kCON only uses interatomic distances r, the translational and rotational invariances are naturally kept and the 1-body and 2-body terms are also permutationally invariant. However, the 3-body terms cannot uphold this requirement because the orders of the parameters of convolutional kernels are fixed while (r_{ab}, r_{ac}, r_{bc}) in equation 8 must be inter-changeable.

To overcome this problem, the conditional sorting scheme is adopted: the columns of the input matrix (the input layer) of each k-body CNN are ordered according to the bond types, and for each k-body interaction (matrix row) we only sort the entries of the same atom types. Taking the examples of CHN, CCH and CCC:

- CHN: each column represents a unique bond type (C-H, C-N, H-N). Sorting is not needed.
- CCN: the entries corresponding to C-C of each row should be sorted.
- CCC: all three entries of each row should be sorted.

2.5 Loss

By default, kCON uses the root mean squared error as the total loss to minimize:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(E_i^{\text{kCON}} - E_i^{\text{True}}\right)^2} \tag{15}$$

kCON also supports exponentially-scaled RMSE if the contributions of the unstable structures should be minimized:

$$\text{esRMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\exp\left(-\frac{E_i^{\text{kCON}} - E_i^{\text{True}}}{k_B T}\right)\left(E_i^{\text{kCON}} - E_i^{\text{True}}\right)^2} \tag{16}$$

2.6 Atomic Energy

The concept of atomic energy was first proposed by Behler et al in 2007. Despite that many later machine learning models include the concept of atomic energy, very little work has been done on interpreting the chemical meaning of these contributions and utilizing them in chemical applications.

kCON is also capable of predicting effective atomic energies. The atomic energies learned from kCON perfectly agree with our chemical intuitions from valence-bond theory, thus providing us with a new approach to understand the local stabilities of atoms and molecular moieties. The usefulness of the atomic energies is further demonstrated by their ability to significantly speed up evolutionary global minimum structure searches.

In fact, atomic energy can be derived from the energy expression, equation 3, directly:

$$\begin{aligned}
E^{total} &= \sum_a^N E^{A_a} + \sum_a^N\sum_{b>a}^N \mathbf{F}^{(k=2)}(z_{ab}, A_a, A_b) + \sum_a^N\sum_{b>a}^N\sum_{c>b}^N \mathbf{F}^{(k=3)}(z_{ab}, z_{bc}, z_{ac}, A_a, A_b, A_c) \\
&= \sum_a^N E^{A_a} + \frac{1}{2!}\sum_a^N\sum_{b\neq a}^N \mathbf{F}^{(k=2)}(z_{ab}, A_a, A_b) + \frac{1}{3!}\sum_a^N\sum_{b\neq a}^N\sum_{c\neq a,b}^N \mathbf{F}^{(k=3)}(z_{ab}, z_{bc}, z_{ac}, A_a, A_b, A_c) \\
&= \sum_a^N \left(E^{A_a} + \frac{1}{2!}\sum_{b\neq a}^N \mathbf{F}^{(k=2)}(z_{ab}, A_a, A_b) + \frac{1}{3!}\sum_{b\neq a}^N\sum_{c\neq a,b}^N \mathbf{F}^{(k=3)}(z_{ab}, z_{bc}, z_{ac}, A_a, A_b, A_c) \right) \\
&= \sum_a^N E_a
\end{aligned} \tag{17}$$

where E_a is the atomic energy of atom a and the total energy E^{total} is the sum of all atomic energies.

3 Atomic forces

3.1 Overview

According to equation 3, for any specific stoichiometry, the total energy:

$$E^{total} = \underbrace{\sum_{a}^{N} E^{A_a}}_{\text{Constant}} + \underbrace{\sum_{a}^{N}\sum_{b>a}^{N} \mathbf{F}^{(k=2)}(z_{ab}, A_a, A_b) + \sum_{a}^{N}\sum_{b>a}^{N}\sum_{c>b}^{N} \mathbf{F}^{(k=3)}(z_{ab}, z_{bc}, z_{ac}, A_a, A_b, A_c)}_{\mathbf{NN}(R)} \tag{18}$$

is composed of two parts: the 1-body part which is constant and the 2/3-body part. which only depends on the coordinates of the atoms. So the total energy of kCON is conservative by construction. Thus, the kCON-derived atomic force should be the negative of the first-order derivative of E^{kCON} with respect to the atomic coordinates r:

$$f(r) = -\frac{\partial E^{\mathrm{kCON}}(r)}{\partial r} \tag{19}$$

So we can get:

$$\begin{aligned}
f(\{x,y,z\}_i) &= -\frac{\partial E^{total}}{\partial \{x,y,z\}_i} \\
&= -\left(\frac{\partial E^{(k=2)}}{\partial \{x,y,z\}_i} + \frac{\partial E^{(k=3)}}{\partial \{x,y,z\}_i}\right) \\
&= -\left(\frac{\partial \sum_a^N \sum_{b>a}^N \mathrm{CNN}^{\mathrm{A}_a\mathrm{A}_b}(z_{ab})}{\partial \{x,y,z\}_i} + \frac{\partial \sum_a^N \sum_{b>a}^N \sum_{c>b}^N \mathrm{CNN}^{\mathrm{A}_a\mathrm{A}_b\mathrm{A}_c}(z_{ab}, z_{ac}, z_{bc})}{\partial \{x,y,z\}_i}\right) \\
&= -\left(\sum_a^N \sum_{b>a}^N \frac{\partial \mathrm{CNN}^{\mathrm{A}_a\mathrm{A}_b}(z_{ab})}{\partial \{x,y,z\}_i} + \sum_a^N \sum_{b>a}^N \sum_{c>b}^N \frac{\partial \mathrm{CNN}^{\mathrm{A}_a\mathrm{A}_b\mathrm{A}_c}(z_{ab}, z_{ac}, z_{bc})}{\partial \{x,y,z\}_i}\right)
\end{aligned} \tag{20}$$

where $f(\{x,y,z\}_i)$ is the force component of atom i along the X/Y/Z direction. We also have:

$$\frac{\partial \mathrm{CNN}^{\mathrm{A}_a\mathrm{A}_b}(z_{ab})}{\partial x_i} = \frac{\partial \mathrm{CNN}^{\mathrm{A}_a\mathrm{A}_b}(z_{ab})}{\partial z_{ab}} \frac{\partial z_{ab}}{\partial r_{ab}} \frac{\partial r_{ab}}{\partial x_i} \tag{21}$$

$$\frac{\partial z_{ab}}{\partial r_{ab}} = -\frac{z_{ab}}{L_{A_a} + L_{A_b}} = -\frac{z_{ab}}{L_{ab}} \tag{22}$$

$$\frac{\partial r_{ab}}{\partial x_i} = \begin{cases} \frac{x_a - x_b}{r_{ab}} & \text{if } i = a \\ -\frac{x_a - x_b}{r_{ab}} & \text{if } i = b \\ 0 & \text{else} \end{cases} \tag{23}$$

Finally we get:

$$r_{ab} = -L_{ab} \log(z_{ab}) \tag{24}$$

$$d_{ab}^{x} = x_a - x_b \tag{25}$$

$$d_{ab}^{y} = y_a - y_b \tag{26}$$

$$d_{ab}^{z} = z_a - z_b \tag{27}$$

$$\frac{\partial z_{ab}}{\partial r_{ab}} \frac{\partial r_{ab}}{\partial \{x,y,z\}_i} = \begin{cases} z_{ab} d_{ab}^{\{x,y,z\}} / (L_{ab}^2 \log(z_{ab})) & \text{if } i = a \\ -z_{ab} d_{ab}^{\{x,y,z\}} / (L_{ab}^2 \log(z_{ab})) & \text{if } i = b \\ 0 & \text{else} \end{cases} \tag{28}$$

For three (or higher body) terms, the result is similar to Equation 11 (See the Appendix A for detailed derivation):

$$\frac{\partial \mathrm{CNN}^k\left(\{z\}\right)}{\partial x_i} = \sum_{ab}^{C_k^N} \frac{\partial \mathrm{CNN}^k\left(\{z\}\right)}{\partial z_{ab}} \frac{\partial z_{ab}}{\partial r_{ab}} \frac{\partial r_{ab}}{\partial x_i} \tag{29}$$

As kCON is built upon Google's TensorFlow, the calculations of the gradients above become far more easier as TensorFlow can output the these complicated derivatives automatically:

$$\frac{\partial \mathrm{CNN}^k\left(\{z\}\right)}{\partial z_{ab}} \tag{30}$$

4 Implementation of the forces

The implementation the atomic forces is a complicated though the theoretical analysis is clear because we must make all operations **vectorizable** so that we can take advantages of modern deep learning frameworks like TensorFlow or MXNet.

4.1 Dimension analysis

Suppose we have a system composed of N atoms with $k^{\max} = 3$, the total energy can be computed with the following equations:

$$E^{\mathrm{kCON}} = E^{(k=1)} + \mathrm{NN}(\boldsymbol{Z}) \tag{31}$$

$$\boldsymbol{Z} = \begin{bmatrix} \vec{z_1} \\ \vec{z_2} \\ \vec{z_3} \\ \vec{z_4} \\ \vec{z_5} \\ \vdots \\ \vec{z_n} \end{bmatrix} \tag{32}$$

$$n = C_3^{N+1} \tag{33}$$

$$\boldsymbol{L} = \begin{bmatrix} \vec{l_1} \\ \vec{l_2} \\ \vec{l_3} \\ \vec{l_4} \\ \vec{l_5} \\ \vdots \\ \vec{l_n} \end{bmatrix} \tag{34}$$

where $\mathbf{Z}$ is the input feature matrix with shape $[C_3^{N+1}, 3]$, $\vec{z_i}$ is a three-components vector representing the **conditionally sorted** features of a chemical pattern and $\mathbf{L}$ is the associated covalent radii matrix for $\mathbf{Z}$. $E^{(k=1)}$ does not depend on interatomic distances, so we can safely ignore it when computing atomic forces. Now TensorFlow can output the derivatives of E^{kCON} with respect to the input feature matrix $\mathbf{Z}$ directly:

$$\frac{\partial E^{\mathrm{kCON}}}{\partial \mathbf{Z}} = \begin{bmatrix} \partial E/\partial \vec{z_1} \\ \partial E/\partial \vec{z_2} \\ \partial E/\partial \vec{z_3} \\ \partial E/\partial \vec{z_4} \\ \partial E/\partial \vec{z_5} \\ \vdots \\ \partial E/\partial \vec{z_n} \end{bmatrix} \tag{35}$$

and the shape of $\partial E^{\mathrm{kCON}}/\partial \boldsymbol{Z}$ is also $[C_3^{N+1}, 3]$.

Now let's look into $\partial E/\partial \vec{z_i}$. Here we define $\vec{z_1} = [z_{12}, z_{13}, z_{23}]$ where z_{ab} is the scaled interatomic distance of atom a and b. According to equation 15, $\partial z_{12}/\partial x_i$ will be non -zero if and only if $i = a$ or $i = b$. Hence, $\partial E/\partial z_{12} \cdot \partial z_{12}/\partial x_i$ will only give effective contributions to **six** atomic force components: f_1^x, f_1^y, f_1^z, f_2^x, f_2^y and f_2^z. Thus, $\partial E^{\text{kCON}}/\partial \boldsymbol{Z}$ will produce $6 \cdot C_3^{N+1} \cdot 3 = 18C_3^{N+1}$ atomic force contributions but only $6 \cdot C_3^N \cdot 3 + 6 \cdot C_2^N \cdot 1$ of them are effective because the ghost atom should give zero contribution. Since we have N atoms, there will be 3N force components and each force component is the sum of $(N-1)^2$ force contributions.

4.2 Tiling

According to the dimension analysis, each entry of $\mathbf{Z}$, $\mathbf{L}$ and $\partial E^{\text{kCON}}/\partial Z$ corresponds to six force components. So repeating these matrices 6 times will make each entry correspond to only one force component. This can be achieved by tiling. The following example demonstrates the tiled $\mathbf{Z}$:

$$Z_{tiled} = \text{tile}(Z, (1, 6)) = \begin{bmatrix} \vec{z_1} & \vec{z_1} & \vec{z_1} & \vec{z_1} & \vec{z_1} & \vec{z_1} \\ \vec{z_2} & \vec{z_2} & \vec{z_2} & \vec{z_2} & \vec{z_2} & \vec{z_2} \\ \vec{z_3} & \vec{z_3} & \vec{z_3} & \vec{z_3} & \vec{z_3} & \vec{z_3} \\ \vec{z_4} & \vec{z_4} & \vec{z_4} & \vec{z_4} & \vec{z_4} & \vec{z_4} \\ \vec{z_5} & \vec{z_5} & \vec{z_5} & \vec{z_5} & \vec{z_5} & \vec{z_5} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vec{z_n} & \vec{z_n} & \vec{z_n} & \vec{z_n} & \vec{z_n} & \vec{z_n} \end{bmatrix} \tag{36}$$

After the tiling, the shapes of $\boldsymbol{Z}_{tiled}$, $\boldsymbol{L}_{tiled}$ and $(\partial E^{\text{kCON}}/\partial Z)_{tiled}$ now become $[C_3^{N+1}, 18]$.

4.3 Coordinates differences

The one last auxiliary matrix to compute is the differences of the atomic coordinates $d_{ab}^{\{x,y,z\}}$ introduced in equation 18. The matrix, denoted as $\mathbf{D}$, also has the shape of $[C_3^{N+1}, 18]$:

$$\mathbf{D} = \begin{bmatrix} \vec{d_1} & \vec{d_2} & \vec{d_3} & \vec{d_4} & \vec{d_5} & \dots & \vec{d_n} \end{bmatrix}^T \tag{37}$$

Suppose $\vec{z}_1 = [z_{12}, z_{13}, z_{23}]$ and $\vec{l_1} = [l_{12}, l_{13}, l_{23}]$. After the tiling, we have:

$$(\vec{z}_{1,tiled})^T = \begin{bmatrix} z_{12} \\ z_{13} \\ z_{23} \\ z_{12} \\ z_{13} \\ z_{23} \\ z_{12} \\ z_{13} \\ z_{23} \\ z_{12} \\ z_{13} \\ z_{23} \\ z_{12} \\ z_{13} \\ z_{23} \\ z_{12} \\ z_{13} \\ z_{23} \end{bmatrix}, \quad \left(\vec{l}_{1,tiled}\right)^T = \begin{bmatrix} l_{12} \\ l_{13} \\ l_{23} \\ l_{12} \\ l_{13} \\ l_{23} \\ l_{12} \\ l_{13} \\ l_{23} \\ l_{12} \\ l_{13} \\ l_{23} \\ l_{12} \\ l_{13} \\ l_{23} \\ l_{12} \\ l_{13} \\ l_{23} \end{bmatrix} \tag{38}$$

So, we can easily compute the corresponding $d_{ab}^{\{x,y,z\}}$:

$$\vec{d}_1 = \begin{matrix} \\ +d_{12}^x \\ +d_{13}^x \\ +d_{23}^x \\ +d_{12}^y \\ +d_{13}^y \\ +d_{23}^y \\ +d_{12}^z \\ +d_{13}^z \\ +d_{23}^z \\ -d_{12}^x \\ -d_{13}^x \\ -d_{23}^x \\ -d_{12}^y \\ -d_{13}^y \\ -d_{23}^y \\ -d_{12}^z \\ -d_{13}^z \\ -d_{23}^z \end{matrix} \begin{matrix} component \\ f_1^x \\ f_1^x \\ f_2^x \\ f_1^y \\ f_1^y \\ f_2^y \\ f_1^z \\ f_1^z \\ f_2^z \\ f_2^x \\ f_3^x \\ f_3^x \\ f_2^y \\ f_3^y \\ f_3^y \\ f_2^z \\ f_3^z \\ f_3^z \end{matrix} \tag{39}$$

4.4 Atomic forces

Finally we can compute atomic forces.

Appendix: an example of B_4

Consider a simple example of B_4 with coordinates:

$$\begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \\ x_4 & y_4 & z_4 \end{bmatrix} \tag{40}$$

This cluster has 4 atoms, so the total dimension of the input feature matrix should be $C_3^{4+1} = 10$ (equation 12), including C_3^4 3-body features (BBB) and C_2^4 2-body features (BBX). Its input feature matrix, $\mathbf{Z}^{(0)}$, can be expressed with the following equation:

$$\mathbf{Z}^{(0)} = \begin{bmatrix} \mathbf{Z^{(0)}}_{\mathrm{BBX}} \\ \mathbf{Z^{(0)}}_{\mathrm{BBB}} \end{bmatrix} \tag{41}$$

where:

$$\mathbf{Z}^{(0)}_{\mathrm{BBX}} = \begin{array}{c} \begin{matrix} \mathrm{BB} & \mathrm{BX} & \mathrm{BX} \end{matrix} \\ \begin{pmatrix} z_{11} & 0 & 0 \\ z_{21} & 0 & 0 \\ z_{31} & 0 & 0 \\ z_{41} & 0 & 0 \\ z_{51} & 0 & 0 \\ z_{61} & 0 & 0 \end{pmatrix} \end{array} = f\left(\begin{bmatrix} r_{12} & +\infty & +\infty \\ r_{13} & +\infty & +\infty \\ r_{14} & +\infty & +\infty \\ r_{23} & +\infty & +\infty \\ r_{24} & +\infty & +\infty \\ r_{34} & +\infty & +\infty \end{bmatrix} \right) \tag{42}$$

$$\mathbf{Z}^{(0)}_{\mathrm{BBB}} = \begin{array}{c} \begin{matrix} \mathrm{BB} & \mathrm{BB} & \mathrm{BB} \end{matrix} \\ \begin{pmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \\ z_{31} & z_{32} & z_{33} \\ z_{41} & z_{42} & z_{43} \end{pmatrix} \end{array} = f\left(\begin{bmatrix} r_{12} & r_{13} & r_{23} \\ r_{12} & r_{14} & r_{24} \\ r_{13} & r_{14} & r_{34} \\ r_{23} & r_{24} & r_{34} \end{bmatrix} \right) = f(\mathbf{R}) \tag{43}$$

and $f(\cdot)$ is the Laplacian normalization function defined in equation 2. The analysis will only consider the 3-body part for simplicity. **Note**: z_{ab} represents the entry of $\mathbf{Z}^{(0)}_{\mathrm{BBB}}$ at row a and column b, z_i denotes the coordinate of atom i along the cartesian direction Z.

Suppose the convolutional neural network for BBB has two hidden layers and each hidden layer has two kernels, then we can get the convolutional kernels:

$$\mathbf{W}^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \end{bmatrix} \tag{44}$$

$$\mathbf{b}^{(1)} = \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \end{bmatrix} \tag{45}$$

$$\mathbf{W}^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} \end{bmatrix} \tag{46}$$

$$\mathbf{b}^{(2)} = \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \end{bmatrix} \tag{47}$$

$$\mathbf{W}^{(3)} = \begin{bmatrix} w_{11}^{(3)} & w_{12}^{(3)} \end{bmatrix} \tag{48}$$

where $\mathbf{W}^{(l)}$ is the kernel matrix for layer l and each row of $\mathbf{W}^{(l)}$ represents a kernel. $\mathbf{b}^{(l)}$ are the biases for the kernels of layer l. One should notice that the last layer (the output layer) does not have a bias unit.
Now we can start the forward propagation. The results of the first layer, $\mathbf{Z}^{(1)}$, should be:

$$
\begin{aligned}
\mathbf{Z}^{(1)} &= \begin{bmatrix} \sigma(w_{11}^{(1)} z_{11} + w_{12}^{(1)} z_{12} + w_{13}^{(1)} z_{13} + b_1^{(1)}) & \sigma(w_{21}^{(1)} z_{11} + w_{22}^{(1)} z_{12} + w_{23}^{(1)} z_{13} + b_2^{(1)}) \\ \sigma(w_{11}^{(1)} z_{21} + w_{12}^{(1)} z_{22} + w_{13}^{(1)} z_{23} + b_1^{(1)}) & \sigma(w_{21}^{(1)} z_{21} + w_{22}^{(1)} z_{22} + w_{23}^{(1)} z_{23} + b_2^{(1)}) \\ \sigma(w_{11}^{(1)} z_{31} + w_{12}^{(1)} z_{32} + w_{13}^{(1)} z_{33} + b_1^{(1)}) & \sigma(w_{21}^{(1)} z_{31} + w_{22}^{(1)} z_{32} + w_{23}^{(1)} z_{33} + b_2^{(1)}) \\ \sigma(w_{11}^{(1)} z_{41} + w_{12}^{(1)} z_{42} + w_{13}^{(1)} z_{43} + b_1^{(1)}) & \sigma(w_{21}^{(1)} z_{41} + w_{22}^{(1)} z_{42} + w_{23}^{(1)} z_{43} + b_2^{(1)}) \end{bmatrix} \\
&= \begin{bmatrix} \sigma(a_{11}^{(1)}) & \sigma(a_{12}^{(1)}) \\ \sigma(a_{21}^{(1)}) & \sigma(a_{22}^{(1)}) \\ \sigma(a_{31}^{(1)}) & \sigma(a_{32}^{(1)}) \\ \sigma(a_{41}^{(1)}) & \sigma(a_{42}^{(1)}) \end{bmatrix} \\
&= \begin{bmatrix} z_{11}^{(1)} & z_{12}^{(1)} \\ z_{21}^{(1)} & z_{22}^{(1)} \\ z_{31}^{(1)} & z_{32}^{(1)} \\ z_{41}^{(1)} & z_{42}^{(1)} \end{bmatrix}
\end{aligned} \tag{49}
$$

where $\sigma(\cdot)$ is the Leaky ReLU activation function (equation 9). Then we can calculate the results of the second layer, $\mathbf{Z}^{(2)}$:

$$
\begin{aligned}
\mathbf{Z}^{(2)} &= \begin{bmatrix} \sigma(w_{11}^{(2)} z_{11}^{(1)} + w_{12}^{(2)} z_{12}^{(1)} + b_1^{(2)}) & \sigma(w_{21}^{(2)} z_{11}^{(1)} + w_{22}^{(2)} z_{12}^{(1)} + b_2^{(2)}) \\ \sigma(w_{11}^{(2)} z_{21}^{(1)} + w_{12}^{(2)} z_{22}^{(1)} + b_1^{(2)}) & \sigma(w_{21}^{(2)} z_{21}^{(1)} + w_{22}^{(2)} z_{22}^{(1)} + b_2^{(2)}) \\ \sigma(w_{11}^{(2)} z_{31}^{(1)} + w_{12}^{(2)} z_{32}^{(1)} + b_1^{(2)}) & \sigma(w_{21}^{(2)} z_{31}^{(1)} + w_{22}^{(2)} z_{32}^{(1)} + b_2^{(2)}) \\ \sigma(w_{11}^{(2)} z_{41}^{(1)} + w_{12}^{(2)} z_{42}^{(1)} + b_1^{(2)}) & \sigma(w_{21}^{(2)} z_{41}^{(1)} + w_{22}^{(2)} z_{42}^{(1)} + b_2^{(2)}) \end{bmatrix} \\
&= \begin{bmatrix} \sigma(a_{11}^{(2)}) & \sigma(a_{12}^{(2)}) \\ \sigma(a_{21}^{(2)}) & \sigma(a_{22}^{(2)}) \\ \sigma(a_{31}^{(2)}) & \sigma(a_{32}^{(2)}) \\ \sigma(a_{41}^{(2)}) & \sigma(a_{42}^{(2)}) \end{bmatrix} \\
&= \begin{bmatrix} z_{11}^{(2)} & z_{12}^{(2)} \\ z_{21}^{(2)} & z_{22}^{(2)} \\ z_{31}^{(2)} & z_{32}^{(2)} \\ z_{41}^{(2)} & z_{42}^{(2)} \end{bmatrix}
\end{aligned} \tag{50}
$$

The results of output layer, $\mathbf{Z}^{(3)}$, should be:

$$
\mathbf{Z}^{(3)} = \begin{bmatrix} z_{11}^{(2)} w_{11}^{(3)} + z_{12}^{(2)} w_{12}^{(3)} \\ z_{21}^{(2)} w_{11}^{(3)} + z_{22}^{(2)} w_{12}^{(3)} \\ z_{31}^{(2)} w_{11}^{(3)} + z_{32}^{(2)} w_{12}^{(3)} \\ z_{41}^{(2)} w_{11}^{(3)} + z_{42}^{(2)} w_{12}^{(3)} \end{bmatrix} \tag{51}
$$

The activation function will not be applied to the output layer. Each entry of $\mathbf{Z}^{(3)}$ represents the k-body energy of its corresponding input chemical pattern (row of the matrix) of $\mathrm{Z}^{(0)}$. The total energy E is just the sum of the entries of $\mathbf{Z}^{(3)}$:

$$
\begin{aligned}
E &= z_{11}^{(2)}w_{11}^{(3)} + z_{12}^{(2)}w_{12}^{(3)} + z_{21}^{(2)}w_{11}^{(3)} + z_{22}^{(2)}w_{12}^{(3)} + z_{31}^{(2)}w_{11}^{(3)} + z_{32}^{(2)}w_{12}^{(3)} + z_{41}^{(2)}w_{11}^{(3)} + z_{42}^{(2)}w_{12}^{(3)} \\
&= \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}^{\mathbf{T}} \begin{bmatrix} z_{11}^{(2)}w_{11}^{(3)} \\ z_{12}^{(2)}w_{12}^{(3)} \\ z_{21}^{(2)}w_{11}^{(3)} \\ z_{22}^{(2)}w_{12}^{(3)} \\ z_{31}^{(2)}w_{11}^{(3)} \\ z_{32}^{(2)}w_{12}^{(3)} \\ z_{41}^{(2)}w_{11}^{(3)} \\ z_{42}^{(2)}w_{12}^{(3)} \end{bmatrix} \\
&= \mathbf{I}^{\mathbf{T}} \begin{bmatrix} \sigma(w_{11}^{(2)}z_{11}^{(1)} + w_{12}^{(2)}z_{12}^{(1)} + b_1^{(2)})w_{11}^{3} \\ \sigma(w_{21}^{(2)}z_{11}^{(1)} + w_{22}^{(2)}z_{12}^{(1)} + b_2^{(2)})w_{12}^{3} \\ \sigma(w_{11}^{(2)}z_{21}^{(1)} + w_{12}^{(2)}z_{22}^{(1)} + b_1^{(2)})w_{11}^{3} \\ \sigma(w_{21}^{(2)}z_{21}^{(1)} + w_{22}^{(2)}z_{22}^{(1)} + b_2^{(2)})w_{12}^{3} \\ \sigma(w_{11}^{(2)}z_{31}^{(1)} + w_{12}^{(2)}z_{32}^{(1)} + b_1^{(2)})w_{11}^{3} \\ \sigma(w_{21}^{(2)}z_{31}^{(1)} + w_{22}^{(2)}z_{32}^{(1)} + b_2^{(2)})w_{12}^{3} \\ \sigma(w_{11}^{(2)}z_{41}^{(1)} + w_{12}^{(2)}z_{42}^{(1)} + b_1^{(2)})w_{11}^{3} \\ \sigma(w_{21}^{(2)}z_{41}^{(1)} + w_{22}^{(2)}z_{42}^{(1)} + b_2^{(2)})w_{12}^{3} \end{bmatrix} \\
&= \mathbf{I}^{\mathbf{T}} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \\ Y_8 \end{bmatrix} \\
&= \mathbf{I}^{\mathbf{T}}\mathbf{Y}
\end{aligned} \tag{52}
$$

where $\mathbf{Y}$ is:

$$
\mathbf{Y} = \begin{bmatrix} \sigma\left(w_{11}^{(2)}\sigma(w_{11}^{(1)}z_{11} + w_{12}^{(1)}z_{12} + w_{13}^{(1)}z_{13} + b_1^{(1)}) + w_{12}^{(2)}\sigma(w_{21}^{(1)}z_{11} + w_{22}^{(1)}z_{12} + w_{23}^{(1)}z_{13} + b_2^{(1)}) + b_1^{(2)}\right)w_{11}^{(3)} \\ \sigma\left(w_{21}^{(2)}\sigma(w_{11}^{(1)}z_{11} + w_{12}^{(1)}z_{12} + w_{13}^{(1)}z_{13} + b_1^{(1)}) + w_{22}^{(2)}\sigma(w_{21}^{(1)}z_{11} + w_{22}^{(1)}z_{12} + w_{23}^{(1)}z_{13} + b_2^{(1)}) + b_2^{(2)}\right)w_{12}^{(3)} \\ \sigma\left(w_{11}^{(2)}\sigma(w_{11}^{(1)}z_{21} + w_{12}^{(1)}z_{22} + w_{13}^{(1)}z_{23} + b_1^{(1)}) + w_{12}^{(2)}\sigma(w_{21}^{(1)}z_{21} + w_{22}^{(1)}z_{22} + w_{23}^{(1)}z_{23} + b_2^{(1)}) + b_1^{(2)}\right)w_{11}^{(3)} \\ \sigma\left(w_{21}^{(2)}\sigma(w_{11}^{(1)}z_{21} + w_{12}^{(1)}z_{22} + w_{13}^{(1)}z_{23} + b_1^{(1)}) + w_{22}^{(2)}\sigma(w_{21}^{(1)}z_{21} + w_{22}^{(1)}z_{22} + w_{23}^{(1)}z_{23} + b_2^{(1)}) + b_2^{(2)}\right)w_{12}^{(3)} \\ \sigma\left(w_{11}^{(2)}\sigma(w_{11}^{(1)}z_{31} + w_{12}^{(1)}z_{32} + w_{13}^{(1)}z_{33} + b_1^{(1)}) + w_{12}^{(2)}\sigma(w_{21}^{(1)}z_{31} + w_{22}^{(1)}z_{32} + w_{23}^{(1)}z_{33} + b_2^{(1)}) + b_1^{(2)}\right)w_{11}^{(3)} \\ \sigma\left(w_{21}^{(2)}\sigma(w_{11}^{(1)}z_{31} + w_{12}^{(1)}z_{32} + w_{13}^{(1)}z_{33} + b_1^{(1)}) + w_{22}^{(2)}\sigma(w_{21}^{(1)}z_{31} + w_{22}^{(1)}z_{32} + w_{23}^{(1)}z_{33} + b_2^{(1)}) + b_2^{(2)}\right)w_{12}^{(3)} \\ \sigma\left(w_{11}^{(2)}\sigma(w_{11}^{(1)}z_{41} + w_{12}^{(1)}z_{42} + w_{13}^{(1)}z_{43} + b_1^{(1)}) + w_{12}^{(2)}\sigma(w_{21}^{(1)}z_{41} + w_{22}^{(1)}z_{42} + w_{23}^{(1)}z_{43} + b_2^{(1)}) + b_1^{(2)}\right)w_{11}^{(3)} \\ \sigma\left(w_{21}^{(2)}\sigma(w_{11}^{(1)}z_{41} + w_{12}^{(1)}z_{42} + w_{13}^{(1)}z_{43} + b_1^{(1)}) + w_{22}^{(2)}\sigma(w_{21}^{(1)}z_{41} + w_{22}^{(1)}z_{42} + w_{23}^{(1)}z_{43} + b_2^{(1)}) + b_2^{(2)}\right)w_{12}^{(3)} \end{bmatrix} \tag{53}
$$

Now the forward propagation is finished and the output (total energy) is obtained. Then we can start the back propagation. The calculation of atomic forces can be done at the same time. To compute the atomic forces, we should first resolve the derivative of E with respect to z_{ab}. Taking the example of $\partial E/\partial z_{11}$, we have:

$$
\begin{aligned}
\frac{\partial E}{\partial z_{11}} &= \sum_{i=1}^{8} \frac{\partial Y_i}{\partial z_{11}} \\
&= \frac{\partial Y_1}{\partial z_{11}} + \frac{\partial Y_2}{\partial z_{11}} \qquad (54) \\
&= w_{11}^{(3)} \frac{\partial \sigma(a_{11}^{(2)})}{\partial a_{11}^{(2)}} \left(w_{11}^{(2)} \frac{\partial \sigma(a_{11}^{(1)})}{\partial a_{11}^{(1)}} w_{11}^{(1)} + w_{12}^{(2)} \frac{\partial \sigma(a_{12}^{(1)})}{\partial a_{12}^{(1)}} w_{21}^{(1)} \right) + \\
&\quad w_{12}^{(3)} \frac{\partial \sigma(a_{12}^{(2)})}{\partial a_{12}^{(2)}} \left(w_{21}^{(2)} \frac{\partial \sigma(a_{11}^{(1)})}{\partial a_{11}^{(1)}} w_{11}^{(1)} + w_{22}^{(2)} \frac{\partial \sigma(a_{12}^{(1)})}{\partial a_{12}^{(1)}} w_{21}^{(1)} \right) \qquad (55)
\end{aligned}
$$

Similarly, we can compute the derivatives of E with respect to all entries of $\mathbf{Z}_{\text{BBB}}^{(0)}$. Then we can calculate the derivative of E with respect to an arbitrary force component, e.g. x_1:

$$
\frac{\partial E}{\partial x_1} = \sum_{i=1}^{8} \sum_{a,b} \frac{\partial Y_i}{\partial z_{ab}} \cdot \frac{\partial z_{ab}}{\partial x_1} \qquad (56)
$$

Remember that:

$$
\begin{bmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \\ z_{31} & z_{32} & z_{33} \\ z_{41} & z_{42} & z_{43} \end{bmatrix} = f(\begin{bmatrix} r_{12} & r_{13} & r_{23} \\ r_{12} & r_{14} & r_{24} \\ r_{13} & r_{14} & r_{34} \\ r_{23} & r_{24} & r_{34} \end{bmatrix}) = f(\mathbf{R}) \qquad (57)
$$

and:

$$
r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \qquad (58)
$$

Only $\partial z_{11}/\partial x_1$, $\partial z_{12}/\partial x_1$, $\partial z_{21}/\partial x_1$, $\partial z_{22}/\partial x_1$, $\partial z_{31}/\partial x_1$ and $\partial z_{32}/\partial x_1$ are non-zero. Thus, equation 56 can be further simplified:

$$
\begin{aligned}
\frac{\partial E}{\partial x_1} &= \sum_{i=1}^{8} \sum_{a,b} \frac{\partial Y_i}{\partial z_{ab}} \cdot \frac{\partial z_{ab}}{\partial x_1} \\
&= \sum_{i=1}^{8} \left(\frac{\partial Y_i}{\partial z_{11}} \frac{\partial z_{11}}{\partial x_1} + \frac{\partial Y_i}{\partial z_{12}} \frac{\partial z_{12}}{\partial x_1} + \frac{\partial Y_i}{\partial z_{21}} \frac{\partial z_{21}}{\partial x_1} + \frac{\partial Y_i}{\partial z_{22}} \frac{\partial z_{22}}{\partial x_1} + \frac{\partial Y_i}{\partial z_{31}} \frac{\partial z_{31}}{\partial x_1} + \frac{\partial Y_i}{\partial z_{32}} \frac{\partial z_{32}}{\partial x_1} \right) \\
&= \left(\frac{\partial Y_1}{\partial z_{11}} \frac{\partial z_{11}}{\partial x_1} + \frac{\partial Y_2}{\partial z_{11}} \frac{\partial z_{11}}{\partial x_1} \right) + \left(\frac{\partial Y_1}{\partial z_{12}} \frac{\partial z_{12}}{\partial x_1} + \frac{\partial Y_2}{\partial z_{12}} \frac{\partial z_{12}}{\partial x_1} \right) + \\
&\quad \left(\frac{\partial Y_3}{\partial z_{21}} \frac{\partial z_{21}}{\partial x_1} + \frac{\partial Y_4}{\partial z_{21}} \frac{\partial z_{21}}{\partial x_1} \right) + \left(\frac{\partial Y_3}{\partial z_{22}} \frac{\partial z_{22}}{\partial x_1} + \frac{\partial Y_4}{\partial z_{22}} \frac{\partial z_{22}}{\partial x_1} \right) + \\
&\quad \left(\frac{\partial Y_5}{\partial z_{31}} \frac{\partial z_{31}}{\partial x_1} + \frac{\partial Y_6}{\partial z_{31}} \frac{\partial z_{31}}{\partial x_1} \right) + \left(\frac{\partial Y_5}{\partial z_{32}} \frac{\partial z_{32}}{\partial x_1} + \frac{\partial Y_6}{\partial z_{32}} \frac{\partial z_{32}}{\partial x_1} \right) \\
&= \frac{\partial E}{\partial z_{11}} \frac{\partial z_{11}}{\partial x_1} + \frac{\partial E}{\partial z_{12}} \frac{\partial z_{12}}{\partial x_1} + \frac{\partial E}{\partial z_{21}} \frac{\partial z_{21}}{\partial x_1} + \frac{\partial E}{\partial z_{22}} \frac{\partial z_{22}}{\partial x_1} + \frac{\partial E}{\partial z_{31}} \frac{\partial z_{31}}{\partial x_1} + \frac{\partial E}{\partial z_{32}} \frac{\partial z_{32}}{\partial x_1} \\
&= \begin{bmatrix} \partial E/\partial z_{11} & \partial E/\partial z_{12} & \partial E/\partial z_{21} & \partial E/\partial z_{22} & \partial E/\partial z_{31} & \partial E/\partial z_{32} \end{bmatrix}^T \begin{bmatrix} \partial z_{11}/\partial x_1 \\ \partial z_{12}/\partial x_1 \\ \partial z_{21}/\partial x_1 \\ \partial z_{22}/\partial x_1 \\ \partial z_{31}/\partial x_1 \\ \partial z_{32}/\partial x_1 \end{bmatrix} \qquad (59)
\end{aligned}
$$

If the element-wise matrix multiplication (Hadamard product) is denoted as $\circ$:

$$(A \circ B)_{i,j} = (A)_{i,j}(B)_{i,j} \tag{60}$$

for any two matrices A and B of the same dimension and **grandsum** is the sum of all elements of arbitrary matrix A of shape $[m, n]$:

$$\text{grandsum}(A) = \sum_{i}^{m} \sum_{j}^{n} (A)_{i,j} \tag{61}$$

Then equation 52 can be converted to the following form:

$$\begin{aligned} \frac{\partial E}{\partial x_1} &= \text{grandsum}\left(\begin{bmatrix} \partial E/\partial z_{11} & \partial E/\partial z_{12} & \partial E/\partial z_{13} \\ \partial E/\partial z_{21} & \partial E/\partial z_{22} & \partial E/\partial z_{23} \\ \partial E/\partial z_{31} & \partial E/\partial z_{32} & \partial E/\partial z_{33} \\ \partial E/\partial z_{41} & \partial E/\partial z_{42} & \partial E/\partial z_{43} \end{bmatrix} \circ \begin{bmatrix} \partial z_{11}/\partial x_1 & \partial z_{12}/\partial x_1 & 0 \\ \partial z_{21}/\partial x_1 & \partial z_{22}/\partial x_1 & 0 \\ \partial z_{31}/\partial x_1 & \partial z_{32}/\partial x_1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) \qquad (62) \\ &= \text{grandsum}\left(\partial E/\partial \mathbf{Z}^{(\mathbf{0})}_{\text{BBB}} \circ \partial \mathbf{Z}^{(0)}_{\mathbf{BBB}}/\partial x_1 \right) \qquad (63) \end{aligned}$$

TensorFlow can handle the complicated derivative $\partial E/\partial \mathbf{Z}^{(\mathbf{0})}_{\text{BBB}}$ and $\partial \mathbf{Z}^{(0)}_{\text{BBB}}/\partial x_1$ can be pre-computed because it doesn't depend on kernel weights. However, $\partial \mathbf{Z}^{(0)}_{\text{BBB}}/\partial x_1$ in equation 62 has 12 ($C_3^4 \cdot C_2^3$) entries but only 6 ($C_3^4 \cdot C_2^3 \cdot 6/(4 \cdot 3)$) of them are non-zero. To avoid unnecessary space waste, the coefficients matrix $\partial \mathbf{Z}^{(0)}_{\text{BBB}}/\partial\{x, y, z\}_i$ should be constructed in another way.

Since each entry of $\partial E/\partial \mathbf{Z}^{(0)}_{\text{BBB}}$ contributes to six force components, it's natural for us to tile $\partial E/\partial \mathbf{Z}^{(0)}_{\text{BBB}}$ six times so that each entry of the tiled matrix only contributes to one force component:

$$\left(\frac{\partial E}{\partial \mathbf{Z}^{(0)}_{\text{BBB}}} \right)_{\text{tiled}} = \begin{bmatrix} \partial E/\partial \mathbf{Z}^{(0)}_{\text{BBB}} & \partial E/\partial \mathbf{Z}^{(0)}_{\text{BBB}} & \partial E/\partial \mathbf{Z}^{(0)}_{\text{BBB}} & \partial E/\partial \mathbf{Z}^{(0)}_{\text{BBB}} & \partial E/\partial \mathbf{Z}^{(0)}_{\text{BBB}} & \partial E/\partial \mathbf{Z}^{(0)}_{\text{BBB}} \end{bmatrix} \tag{64}$$

and then calculate the coefficients matrix $\partial \mathbf{Z}^{(0)}_{\text{BBB}}/\partial\{x, y, z\}_i$:

$$\left(\frac{\partial \mathbf{Z}^{(0)}_{\text{BBB}}}{\partial\{x, y, z\}_i} \right)^{\mathbf{T}} = \begin{bmatrix} \partial z_{11}/\partial x_1 & \partial z_{21}/\partial x_1 & \partial z_{31}/\partial x_1 & \partial z_{41}/\partial x_2 \\ \partial z_{12}/\partial x_1 & \partial z_{22}/\partial x_1 & \partial z_{32}/\partial x_1 & \partial z_{42}/\partial x_2 \\ \partial z_{13}/\partial x_2 & \partial z_{23}/\partial x_2 & \partial z_{33}/\partial x_3 & \partial z_{43}/\partial x_3 \\ \partial z_{11}/\partial y_1 & \partial z_{21}/\partial y_1 & \partial z_{31}/\partial y_1 & \partial z_{41}/\partial y_2 \\ \partial z_{12}/\partial y_1 & \partial z_{22}/\partial y_1 & \partial z_{32}/\partial y_1 & \partial z_{42}/\partial y_2 \\ \partial z_{13}/\partial y_2 & \partial z_{23}/\partial y_2 & \partial z_{33}/\partial y_3 & \partial z_{43}/\partial y_3 \\ \partial z_{11}/\partial z_1 & \partial z_{21}/\partial z_1 & \partial z_{31}/\partial z_1 & \partial z_{41}/\partial z_2 \\ \partial z_{12}/\partial z_1 & \partial z_{22}/\partial z_1 & \partial z_{32}/\partial z_1 & \partial z_{42}/\partial z_2 \\ \partial z_{13}/\partial z_2 & \partial z_{23}/\partial z_2 & \partial z_{33}/\partial z_3 & \partial z_{43}/\partial z_3 \\ \partial z_{11}/\partial x_2 & \partial z_{21}/\partial x_2 & \partial z_{31}/\partial x_3 & \partial z_{41}/\partial x_3 \\ \partial z_{12}/\partial x_3 & \partial z_{22}/\partial x_4 & \partial z_{32}/\partial x_4 & \partial z_{42}/\partial x_4 \\ \partial z_{13}/\partial x_3 & \partial z_{23}/\partial x_4 & \partial z_{33}/\partial x_4 & \partial z_{43}/\partial x_4 \\ \partial z_{11}/\partial y_2 & \partial z_{21}/\partial y_2 & \partial z_{31}/\partial y_3 & \partial z_{41}/\partial y_3 \\ \partial z_{12}/\partial y_3 & \partial z_{22}/\partial y_4 & \partial z_{32}/\partial y_4 & \partial z_{42}/\partial y_4 \\ \partial z_{13}/\partial y_3 & \partial z_{23}/\partial y_4 & \partial z_{33}/\partial y_4 & \partial z_{43}/\partial y_4 \\ \partial z_{11}/\partial z_2 & \partial z_{21}/\partial z_2 & \partial z_{31}/\partial z_3 & \partial z_{41}/\partial z_3 \\ \partial z_{12}/\partial z_3 & \partial z_{22}/\partial z_4 & \partial z_{32}/\partial z_4 & \partial z_{42}/\partial z_4 \\ \partial z_{13}/\partial z_3 & \partial z_{23}/\partial z_4 & \partial z_{33}/\partial z_4 & \partial z_{43}/\partial z_4 \end{bmatrix} \tag{65}$$

Then we can obtain the force contributions matrix, $\partial E/\partial\{x, y, z\}_i$, with the following Hadamard product:

$$\frac{\partial E}{\partial\{x, y, z\}_i} = \frac{\partial \mathbf{Z}^{(0)}_{\text{BBB}}}{\partial\{x, y, z\}_i} \circ \left(\frac{\partial E}{\partial \mathbf{Z}^{(0)}_{\text{BBB}}} \right)_{\text{tiled}} \tag{66}$$

Expand $\partial E/\partial\{x, y, z\}_i$ we can have:

$$\frac{\partial E}{\partial \{x,y,z\}_i} = \begin{bmatrix} \partial E/\partial z_{11}\cdot\partial z_{11}/\partial x_1 & \partial E/\partial z_{21}\cdot\partial z_{21}/\partial x_1 & \partial E/\partial z_{31}\cdot\partial z_{31}/\partial x_1 & \partial E/\partial z_{41}\cdot\partial z_{41}/\partial x_2 \\ \partial E/\partial z_{12}\cdot\partial z_{12}/\partial x_1 & \partial E/\partial z_{22}\cdot\partial z_{22}/\partial x_1 & \partial E/\partial z_{32}\cdot\partial z_{32}/\partial x_1 & \partial E/\partial z_{42}\cdot\partial z_{42}/\partial x_2 \\ \partial E/\partial z_{13}\cdot\partial z_{13}/\partial x_2 & \partial E/\partial z_{23}\cdot\partial z_{23}/\partial x_2 & \partial E/\partial z_{33}\cdot\partial z_{33}/\partial x_3 & \partial E/\partial z_{43}\cdot\partial z_{43}/\partial x_3 \\ \partial E/\partial z_{11}\cdot\partial z_{11}/\partial y_1 & \partial E/\partial z_{21}\cdot\partial z_{21}/\partial y_1 & \partial E/\partial z_{31}\cdot\partial z_{31}/\partial y_1 & \partial E/\partial z_{41}\cdot\partial z_{41}/\partial y_2 \\ \partial E/\partial z_{12}\cdot\partial z_{12}/\partial y_1 & \partial E/\partial z_{22}\cdot\partial z_{22}/\partial y_1 & \partial E/\partial z_{32}\cdot\partial z_{32}/\partial y_1 & \partial E/\partial z_{42}\cdot\partial z_{42}/\partial y_2 \\ \partial E/\partial z_{13}\cdot\partial z_{13}/\partial y_2 & \partial E/\partial z_{23}\cdot\partial z_{23}/\partial y_2 & \partial E/\partial z_{33}\cdot\partial z_{33}/\partial y_3 & \partial E/\partial z_{43}\cdot\partial z_{43}/\partial y_3 \\ \partial E/\partial z_{11}\cdot\partial z_{11}/\partial z_1 & \partial E/\partial z_{21}\cdot\partial z_{21}/\partial z_1 & \partial E/\partial z_{31}\cdot\partial z_{31}/\partial z_1 & \partial E/\partial z_{41}\cdot\partial z_{41}/\partial z_2 \\ \partial E/\partial z_{12}\cdot\partial z_{12}/\partial z_1 & \partial E/\partial z_{22}\cdot\partial z_{22}/\partial z_1 & \partial E/\partial z_{32}\cdot\partial z_{32}/\partial z_1 & \partial E/\partial z_{42}\cdot\partial z_{42}/\partial z_2 \\ \partial E/\partial z_{13}\cdot\partial z_{13}/\partial z_2 & \partial E/\partial z_{23}\cdot\partial z_{23}/\partial z_2 & \partial E/\partial z_{33}\cdot\partial z_{33}/\partial z_3 & \partial E/\partial z_{43}\cdot\partial z_{43}/\partial z_3 \\ \partial E/\partial z_{11}\cdot\partial z_{11}/\partial x_2 & \partial E/\partial z_{21}\cdot\partial z_{21}/\partial x_2 & \partial E/\partial z_{31}\cdot\partial z_{31}/\partial x_3 & \partial E/\partial z_{41}\cdot\partial z_{41}/\partial x_3 \\ \partial E/\partial z_{12}\cdot\partial z_{12}/\partial x_3 & \partial E/\partial z_{22}\cdot\partial z_{22}/\partial x_4 & \partial E/\partial z_{32}\cdot\partial z_{32}/\partial x_4 & \partial E/\partial z_{42}\cdot\partial z_{42}/\partial x_4 \\ \partial E/\partial z_{13}\cdot\partial z_{13}/\partial x_3 & \partial E/\partial z_{23}\cdot\partial z_{23}/\partial x_4 & \partial E/\partial z_{33}\cdot\partial z_{33}/\partial x_4 & \partial E/\partial z_{43}\cdot\partial z_{43}/\partial x_4 \\ \partial E/\partial z_{11}\cdot\partial z_{11}/\partial y_2 & \partial E/\partial z_{21}\cdot\partial z_{21}/\partial y_2 & \partial E/\partial z_{31}\cdot\partial z_{31}/\partial y_3 & \partial E/\partial z_{41}\cdot\partial z_{41}/\partial y_3 \\ \partial E/\partial z_{12}\cdot\partial z_{12}/\partial y_3 & \partial E/\partial z_{22}\cdot\partial z_{22}/\partial y_4 & \partial E/\partial z_{32}\cdot\partial z_{32}/\partial y_4 & \partial E/\partial z_{42}\cdot\partial z_{42}/\partial y_4 \\ \partial E/\partial z_{13}\cdot\partial z_{13}/\partial y_3 & \partial E/\partial z_{23}\cdot\partial z_{23}/\partial y_4 & \partial E/\partial z_{33}\cdot\partial z_{33}/\partial y_4 & \partial E/\partial z_{43}\cdot\partial z_{43}/\partial y_4 \\ \partial E/\partial z_{11}\cdot\partial z_{11}/\partial z_2 & \partial E/\partial z_{21}\cdot\partial z_{21}/\partial z_2 & \partial E/\partial z_{31}\cdot\partial z_{31}/\partial z_3 & \partial E/\partial z_{41}\cdot\partial z_{41}/\partial z_3 \\ \partial E/\partial z_{12}\cdot\partial z_{12}/\partial z_3 & \partial E/\partial z_{22}\cdot\partial z_{22}/\partial z_4 & \partial E/\partial z_{32}\cdot\partial z_{32}/\partial z_4 & \partial E/\partial z_{42}\cdot\partial z_{42}/\partial z_4 \\ \partial E/\partial z_{13}\cdot\partial z_{13}/\partial z_3 & \partial E/\partial z_{23}\cdot\partial z_{23}/\partial z_4 & \partial E/\partial z_{33}\cdot\partial z_{33}/\partial z_4 & \partial E/\partial z_{43}\cdot\partial z_{43}/\partial z_4 \end{bmatrix} \tag{67}$$

Now, the one last thing to do is building an auxiliary matrix **IND** that can **re-order** the entries of $\partial E/\partial\{x,y,z\}_i$ to build a matrix of shape $[3\mathrm{N}, 6]$ so that all entries of each row corresponds to the same force component. **IND** can also be pre-computed.

$$\begin{aligned} &\left(\frac{\partial E}{\partial \{x,y,z\}_i}\right)_{\text{ordered}} \\ = \;& \mathbf{Reorder}\left(\frac{\partial E}{\partial \{x,y,z\}_i}, \mathbf{IND}\right) \\ = \;& \begin{array}{c} \textit{Component} \\ f_1^x \\ f_1^y \\ f_1^z \\ f_2^x \\ f_2^y \\ f_2^z \\ f_3^x \\ f_3^y \\ f_3^z \\ f_4^x \\ f_4^y \\ f_4^z \end{array} \begin{pmatrix} \partial E^{11}/\partial x_1 & \partial E^{12}/\partial x_1 & \partial E^{21}/\partial x_1 & \partial E^{22}/\partial x_1 & \partial E^{31}/\partial x_1 & \partial E^{32}/\partial x_1 \\ \partial E^{11}/\partial y_1 & \partial E^{12}/\partial y_1 & \partial E^{21}/\partial y_1 & \partial E^{22}/\partial y_1 & \partial E^{31}/\partial y_1 & \partial E^{32}/\partial y_1 \\ \partial E^{11}/\partial z_1 & \partial E^{12}/\partial z_1 & \partial E^{21}/\partial z_1 & \partial E^{22}/\partial z_1 & \partial E^{31}/\partial z_1 & \partial E^{32}/\partial z_1 \\ \partial E^{13}/\partial x_2 & \partial E^{11}/\partial x_2 & \partial E^{23}/\partial x_2 & \partial E^{21}/\partial x_2 & \partial E^{41}/\partial x_2 & \partial E^{42}/\partial x_2 \\ \partial E^{13}/\partial y_2 & \partial E^{11}/\partial y_2 & \partial E^{23}/\partial y_2 & \partial E^{21}/\partial y_2 & \partial E^{41}/\partial y_2 & \partial E^{42}/\partial y_2 \\ \partial E^{13}/\partial z_2 & \partial E^{11}/\partial z_2 & \partial E^{23}/\partial z_2 & \partial E^{21}/\partial z_2 & \partial E^{41}/\partial z_2 & \partial E^{42}/\partial z_2 \\ \partial E^{12}/\partial x_3 & \partial E^{13}/\partial x_3 & \partial E^{33}/\partial x_3 & \partial E^{31}/\partial x_3 & \partial E^{43}/\partial x_3 & \partial E^{41}/\partial x_3 \\ \partial E^{12}/\partial y_3 & \partial E^{13}/\partial y_3 & \partial E^{33}/\partial y_3 & \partial E^{31}/\partial y_3 & \partial E^{43}/\partial y_3 & \partial E^{41}/\partial y_3 \\ \partial E^{12}/\partial z_3 & \partial E^{13}/\partial z_3 & \partial E^{33}/\partial z_3 & \partial E^{31}/\partial z_3 & \partial E^{43}/\partial z_3 & \partial E^{41}/\partial z_3 \\ \partial E^{22}/\partial x_4 & \partial E^{23}/\partial x_4 & \partial E^{32}/\partial x_4 & \partial E^{33}/\partial x_4 & \partial E^{42}/\partial x_4 & \partial E^{43}/\partial x_4 \\ \partial E^{22}/\partial y_4 & \partial E^{23}/\partial y_4 & \partial E^{32}/\partial y_4 & \partial E^{33}/\partial y_4 & \partial E^{42}/\partial y_4 & \partial E^{43}/\partial y_4 \\ \partial E^{22}/\partial z_4 & \partial E^{23}/\partial z_4 & \partial E^{32}/\partial z_4 & \partial E^{33}/\partial z_4 & \partial E^{42}/\partial z_4 & \partial E^{43}/\partial z_4 \end{pmatrix} \end{aligned} \tag{68}$$

where:

$$\frac{\partial E^{ab}}{\partial \{x,y,z\}_i} = \frac{\partial E}{\partial z_{ab}} \cdot \frac{\partial z_{ab}}{\partial \{x,y,z\}_3} \tag{69}$$

Finally, we can get all atomic forces by summing up each row:

$$\vec{F}(\mathrm{B}_4) = \mathbf{Sum}\left(\left(\frac{\partial E}{\partial \{x,y,z\}_i}\right)_{\text{ordered}}, \quad \text{axis} = 1\right) \tag{70}$$