Features encoding structure also

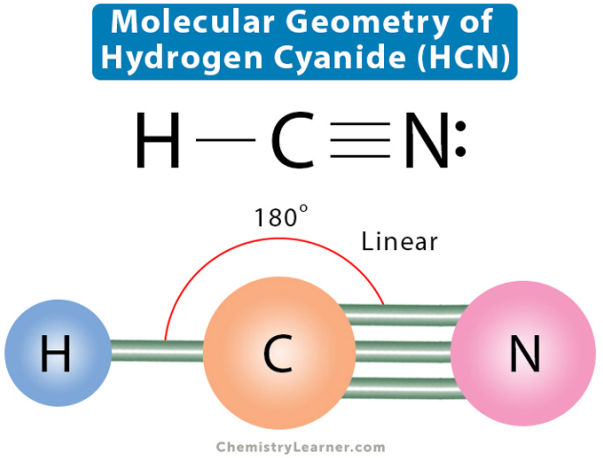
Global features

1. Coulomb matrix

- Relevant only for molecules. Still we discuss because the there are generalizations of this for periodic solids.

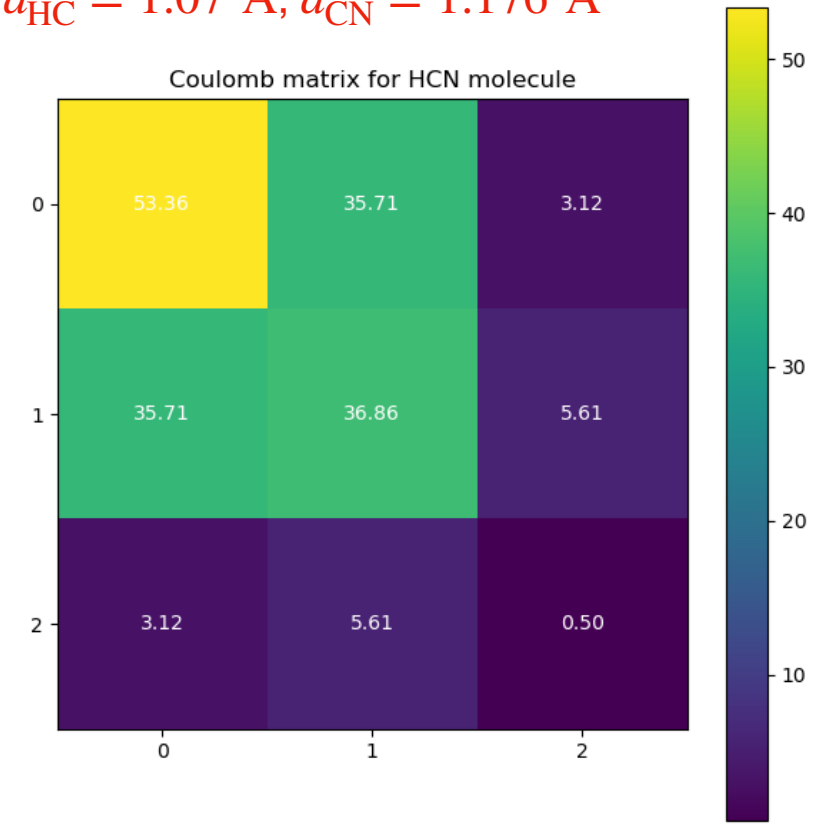
$$M_{ij}^{\text{Coulomb}} = \begin{cases} 0.5 Z_i^{2.4} & \forall \quad i = j \\ \dfrac{Z_i Z_j}{|\vec{R}_i - \vec{R}_j|} & \forall i \quad \neq j \,. \end{cases}$$

- Off-diagonal elements Coulomb interaction energy between nuclei i and j.
- Diagonal elements, atomic energies fitted to Z.



$$H - C \equiv N$$

$$d_{HC} = 1.07 \text{ Å}; d_{CN} = 1.176 \text{ Å}$$

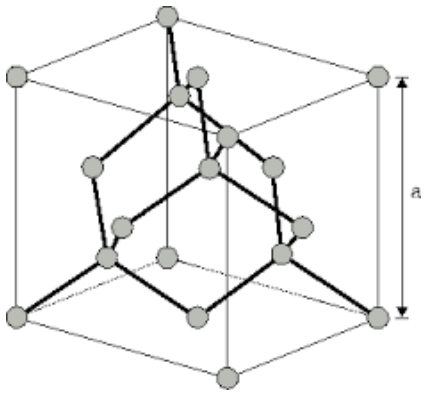


Arxiv:1904.08875

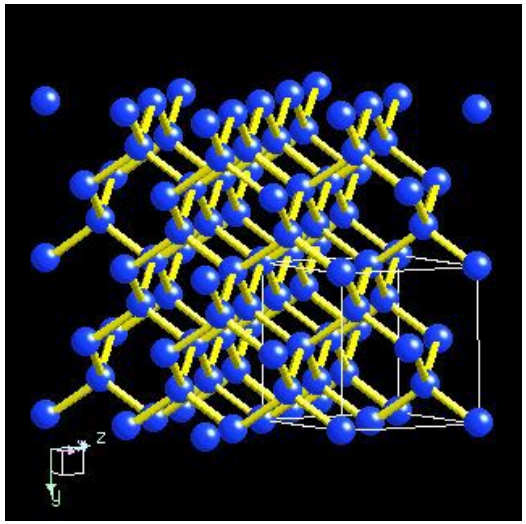
2. **Ewald matrix**

- Generalization of Coulomb matrix for an infinite, periodic solid
- $\phi_{ij} = Z_i Z_j \sum_{\vec{R}}{}' V(|\vec{R}_i - \vec{R}_j + \vec{R}|); \quad \vec{R} = h\vec{a}_1 + k\vec{a}_2 + l\vec{a}_3; \quad \text{for } \vec{R} = 0,\ i \neq j$
- If $|V(r)| \leq Ar^{-3-\epsilon}$ for $A,\ \epsilon > 0$, the infinite series as $h, k, l \in [0,\infty)$ **converges absolutely**
  - A series $S = \sum_1^\infty a_k$ converges absolutely iff $\sum_1^\infty |a_k| < \infty$.
- But Coulomb interaction is $\phi_{ij} = \sum_{\vec{R}}{}' \frac{Z_i Z_j}{|\vec{R}_i - R_j + \vec{R}|}, \quad V(r) \sim |r|^{-1}$.
- **Conditionally convergent** series for $h, k, l \in [0,\infty)$, value depends on the order of the terms
- S is said to converge conditionally iff S converges, but not absolutely, Example: $S = \sum_1^\infty \frac{(-1)^k}{k}$
- **Diverges** if system is not charge neutral,
- **Ewald summation** technique.

Arxiv:1904.08875



Diamond structure



3. **Sine matrix**

- Ewald matrix is computationally expensive to calculate.
- Can we construct a quantity akin to EM, yet easy to calculate? Sine matrix.
- $$M_{ij}^{\text{sine}} = \begin{cases} 0.5 Z_i^{2.4} & \forall i = j, \\ \phi_{ij} & \forall i \neq j, \end{cases}$$

Interaction of two charges $\phi_{ij} = \sum_{\vec{R}} \frac{Z_i Z_j}{|\vec{R}_i - \vec{R}_j + \vec{R}|}$ satisfy

(i) $\vec{R}_i \to \vec{R}_i + \vec{R}'$, ϕ_{ij} is unchanged;

(ii) $\vec{R}_i \to \vec{R}_j$, ϕ_{ij} diverges.

$$\phi_{ij} = Z_i Z_j \left|\left| \mathbf{B} \cdot \sum_{k=1}^{3} \hat{e}_k \sin^2(\pi \hat{e}_k \cdot \mathbf{B}^{-1} \cdot (\vec{R}_i - \vec{R}_j)) \right|\right|_2^{-1}.$$

This ϕ_{ij} satisfies above conditions.

- $\mathbf{B}$ is a 3×3 matrix whose columns are the lattice vectors written column-wise:

  $B_{11} = a_{11}$, $B_{21} = a_{12}$ and $B_{31} = a_{13}$.

- a_{ij} is component of $\vec{a}_i$ along Cartesian direction j.
- $\hat{e}_k$: Cartesian unit vector along direction k.

$$\vec{a}_i = \begin{pmatrix} a_{i1} \\ a_{i2} \\ a_{i3} \end{pmatrix} \quad i = 1, 2, 3.$$

$$\mathbf{B} = \left((\vec{a}_1) \quad (\vec{a}_2) \quad (\vec{a}_3) \right)$$

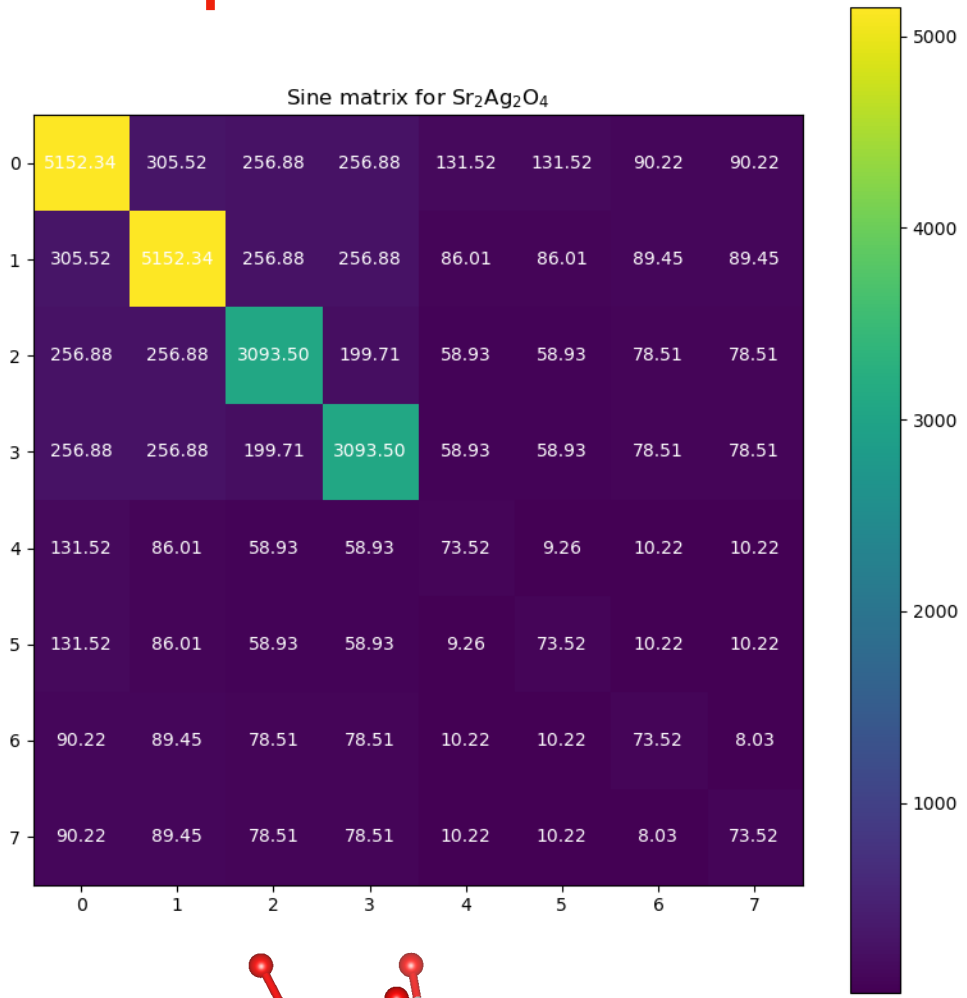
$$\mathbf{B} = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$$

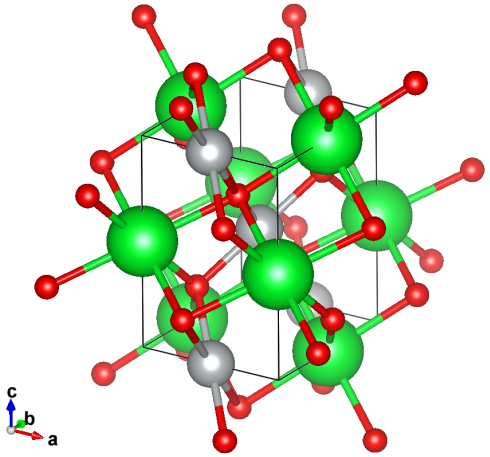
- Captures the infinite, periodic arrangement of the atoms.

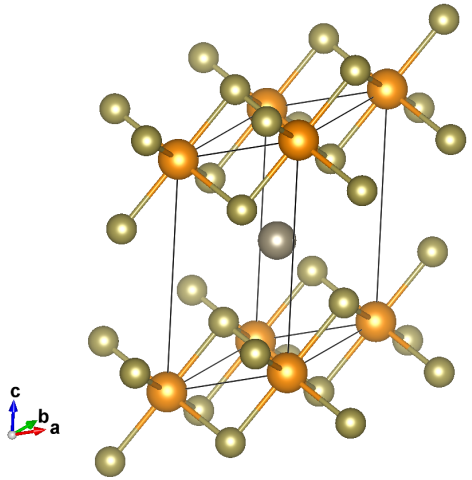
  No physical meaning.

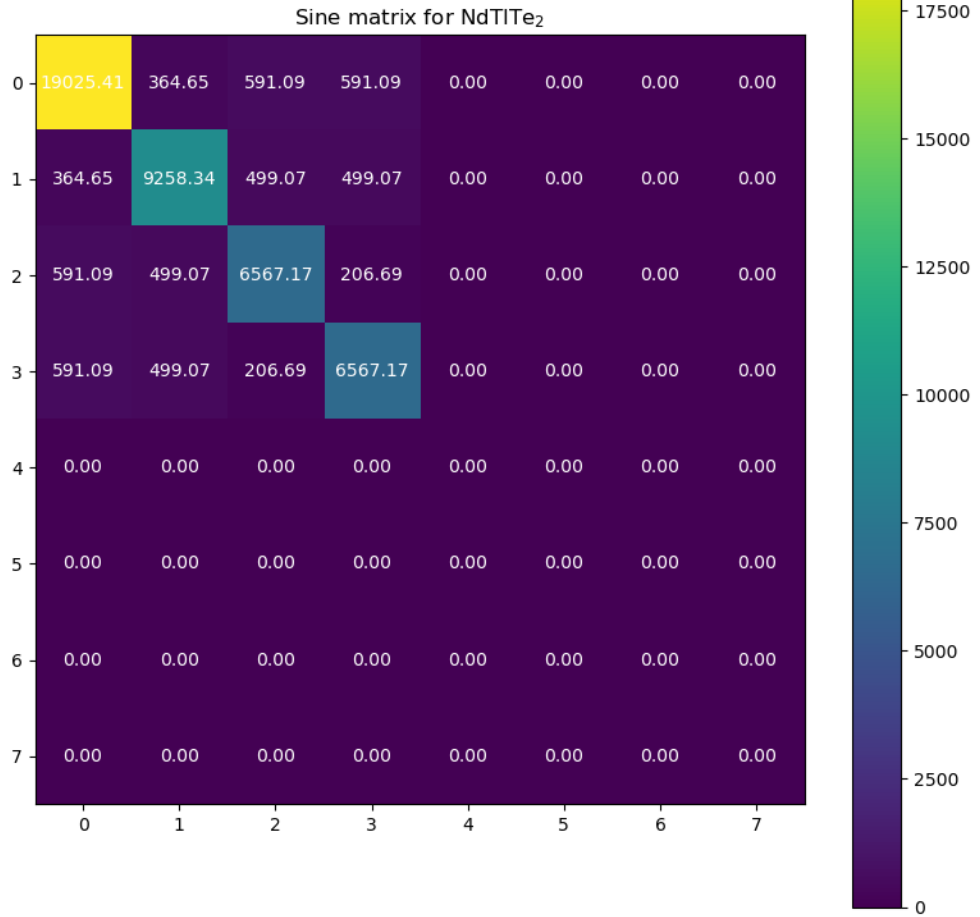
Arxiv:1904.08875

Examples of sine matrices





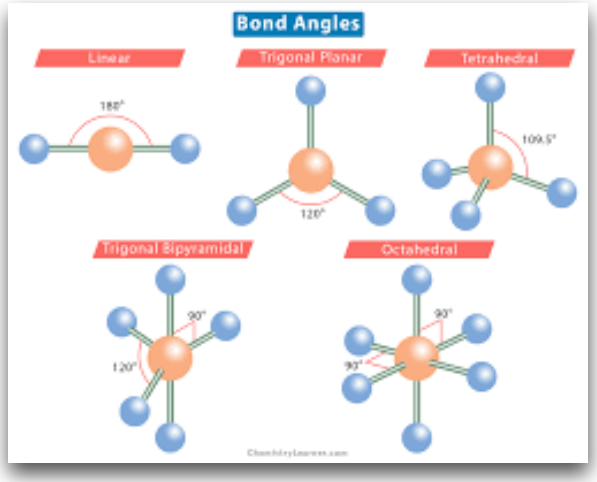




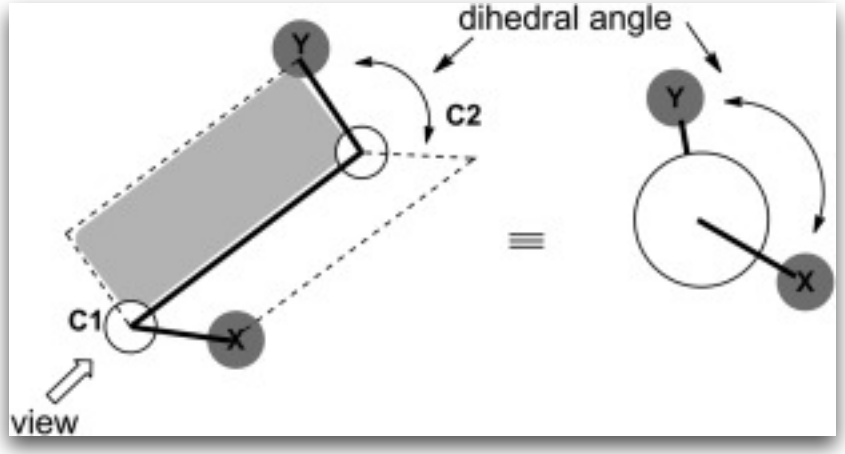
Global features

1. Many-body tensor representation

- Ewald matrix, sine matrix contain information about interatomic distances, but not bond angles or dihedrals.
- MBTR captures such information.
- Identify and encode information about motifs of different size, then group them according to their chemical composition.



https://www.chemistrylearner.com/bond-angles.html



https://www.sciencedirect.com/topics/chemistry/dihedral-angle

Arxiv:1904.08875

- Steps to construct MBTR

1. Identify structural motifs of k atoms involving specific chemical elements

(I) $k = 1, 2, 3$ etc.

(II) $k = 1$ motif for H or He or Fe etc. $k = 2$ motifs for H-M, M-O etc. $k = 3$ motifs: M-O-M, O-M-O, M-M-M etc.

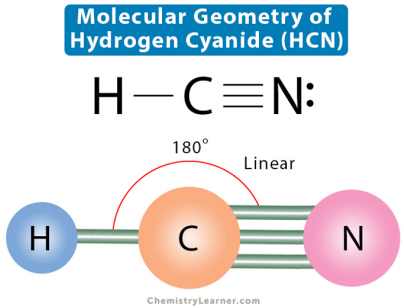
2. Map motif to a scalar function $g_k^{Z_1,\ldots,Z_k}(\vec{R}_1, \ldots \vec{R}_k)$.

(I) Commonly used functions:

(II) $g_1^{Z_i}(\vec{R}_i) = Z_i$, one for H, one of C, … one for each element present

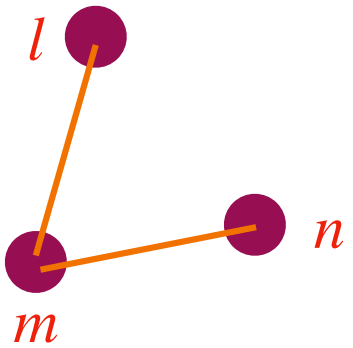
(III) $g_2(\vec{R}_i, \vec{R}_j) = \dfrac{1}{|\vec{R}_i - \vec{R}_j|}$, Z indices suppressed for now.

(IV) $g_3(\vec{R}_l, \vec{R}_m, \vec{R}_n) = \cos(\angle(\vec{R}_l - \vec{R}_m, \vec{R}_m - \vec{R}_n))$.



3. $g_k's$ are 'broadened' using a (Gaussian) kernel

(I) $g_k^{Z_1,\ldots,Z_k}(\vec{R}_1, \ldots \vec{R}_k) \rightarrow \mathscr{D}_k(x, g_k^{Z_1,\ldots,Z_k}(\vec{R}_1, \ldots \vec{R}_k))$



Motif → $g_k's$ → $\mathscr{D}_k$'s

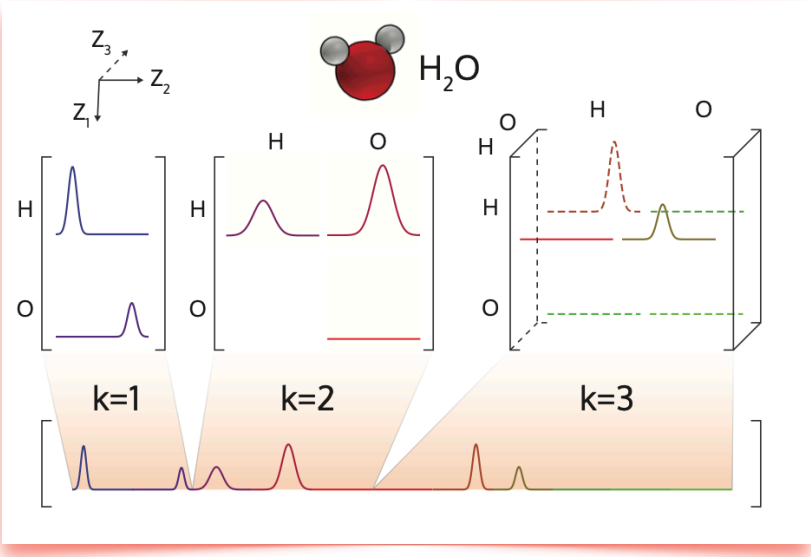
- More details of MBTR:
  - $g_1(\vec{R}_i) = Z_i, \quad g_2(\vec{R}_i, \vec{R}_j) = \frac{1}{|\vec{R}_i - \vec{R}_j|}.$
  - $g_3(\vec{R}_l, \vec{R}_m, \vec{R}_n) = \cos(\angle(\vec{R}_l - \vec{R}_m, \vec{R}_m - \vec{R}_n)).$
  - $\mathscr{D}_1^i(x) = \frac{1}{\sigma_1\sqrt{2\pi}} \exp(-\frac{(x - g_1(\vec{R}_i))^2}{2\sigma_1^2}),$
  - $\mathscr{D}_2^{i,j}(x) = \frac{1}{\sigma_2\sqrt{2\pi}} \exp(-\frac{(x - g_2(\vec{R}_i, \vec{R}_j))^2}{2\sigma_2^2}),$ etc.

Arxiv:1904.08875

- Now f_k's are the MBTRs
  - $f_1^{Z_1}(x) = \sum_i^{N_e} w_1^i \mathscr{D}_1(x)\delta_{Z_i,Z_1},$
  - $f_2^{Z_1,Z_2}(x) = \sum_{i,j}^{N_e} w_2^{i,j} \mathscr{D}_2^{i,j}(x)\delta_{Z_i,Z_1}\delta_{Z_j,Z_2},$
  - $f_3^{Z_1,Z_2,Z_3}(x) = \sum_{l,m,n}^{N_e} w_2^{l,m,n} \mathscr{D}_2^{i,j}(x)\delta_{Z_l,Z_1}\delta_{Z_m,Z_2}\delta_{Z_n,Z_3}.$

- MBTR continued:
  - Weights
  - $w_1^i = 1 \quad \forall i,$
  - $w_2^{i,j} = e^{-s_2|\vec{R}_i - \vec{R}_j|},$
  - $w_3^{l,m,n} = e^{-s_3(|\vec{R}_l - \vec{R}_m| + |\vec{R}_m - \vec{R}_n| + |\vec{R}_n - \vec{R}_l|)}.$
  - Exponential forms ensure that contributions of motifs of distant atoms are damped out.

MBTR of H2O molecule



Arxiv:1904.08875