Local features

2. Smooth Overlap of Atomic Orbitals (SOAP)

- Atomic density field $\rho^Z(\vec{r}) = \sum_i w_{Z_i}\delta(\vec{r} - \vec{R}_i)\delta_{Z_i,Z}$, or
- Gaussian smeared atomic density field $\rho^Z(\vec{r}) = \sum_i e^{-\frac{1}{2\sigma^2}|\vec{r}-\vec{R}_i|^2}\delta_{Z_i,Z}$.
- $\rho^Z(\vec{r})$ in a local region surrounding the point of interest as origin expanded as

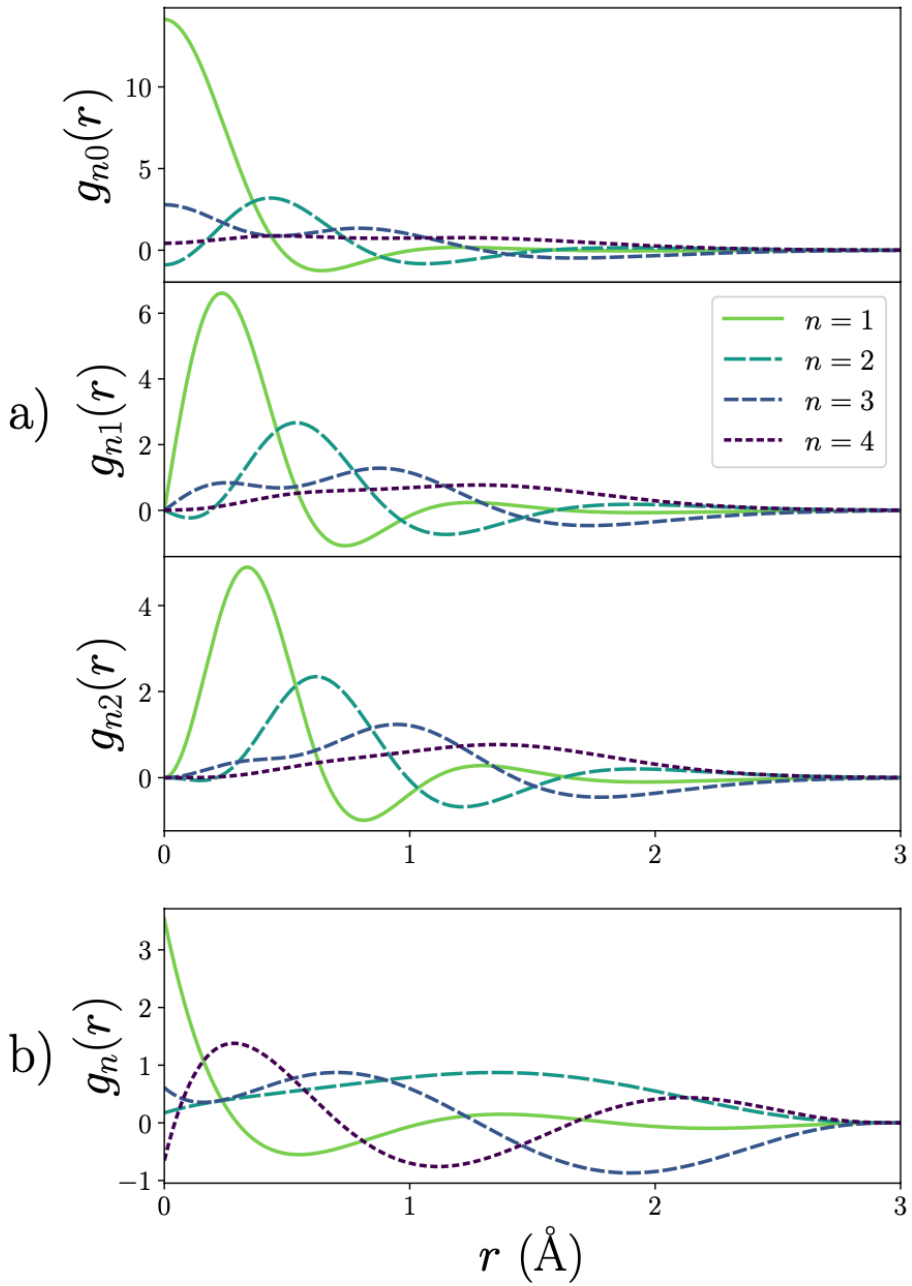
  $\rho^Z(\vec{r}) = \sum_{n=1}^{n_{\max}}\sum_{l=1}^{l_{\max}}\sum_{m=-l}^{l} c^Z_{nlm} g_{nl}(r) Y_{lm}(\theta,\phi)$. C^Z_{nlm}s are coefficients.

- $Y^l_m(\theta,\phi) = (-1)^l\sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}}P_{lm}(\cos\theta)e^{im\phi}$, angular basis functions.
- If g_{nl}'s are ortho-normal functions
  - $c_{nlm} = \int_V g_n(r) Y^*_{lm}(\theta,\phi)\rho(\vec{r})dv$.

Two types of radial basis functions:

- Spherical primitive Gaussian orbitals as
  - $g_{nl}(r) = \sum_{n'=1}^{n_{\max}} \beta_{nn'l} \Phi_{n'l}(r)$ with $\Phi_{nl}(r) = r^l e^{-\alpha_n r^2}$
  - α_n is the decay parameter, chosen such that each $\Phi_{nl}(r)$ decays to 10^{-3} at a cutoff radius r_{cut}.
  - $g_{nl}(r)$'s are not orthonormal
  - They can be made orthonormal by Lowdiwn orthonormalization procedure:
    - Overlap integral $S_{nn'} \equiv \int_0^\infty r^2 (\Phi_{nl}(r))(\Phi_{n'l}(r)) dr = \int_0^\infty r^2 (r^l e^{-\alpha_n r^2})(r^l e^{-\alpha_{n'} r^2}) dr$
    - Choose $\beta = \mathbf{S}^{-1/2}$
    - $\int r^2 g_{ml}(r) g_{nl}(r) dr = \int r^2 \Big(\sum_{m'} \beta_{mm'l} \Phi_{m'l}(r) \Big) \Big(\sum_{n'} \beta_{nn'l} \Phi_{n'l}(r) \Big) dr$
    - $\beta_{mm'l} \beta_{nn'l} \int r^2 \Phi_{m'l}(r) \Phi_{n'l}(r) dr = \beta_{mm'l} \beta_{nn'l} S_{m'n'} = \beta_{mm'l} S_{m'n'} \beta_{n'nl}$ matrices are real symmetric
    - $= S^{-1/2}_{mm'} S_{m'n'} S^{-1/2}_{n'n} = \delta_{mn}.$
  - DScribe (we will work with later) includes up to $l \leq 9$.

- Polynomial functions
  - $g_n(r) = \sum_{n'=1}^{n_{\max}} \beta_{nn'} \Phi_{n'}(r)$ with $\Phi_{nl}(r) = (r - r_{\text{cut}})^{n+2}$
  - α_n is the decay parameter, chosen such that each $\Phi_{nl}(r)$ decays to 10^{-3} at a cutoff radius r_{cut}.
  - These can be made orthonormal by Lowdiwn orthonormalization procedure: choose
    - $\mathbf{S}_{nn'} = \int_0^\infty r^2 (r - r_{\text{cut}})^{n+2} (r - r_{\text{cut}})^{n'+2} dr$ for each l
    - $\beta = \mathbf{S}^{-1/2}$.
  - DScribe, cubic or higher order polynomials, $l \leq 20$.



Radial basis functions for SOAP: (a) spherical gaussian, (b) polynomial

- Effect of rotation, rotation operator $\hat{R}$ acting on $\rho(\vec{r})$:

$$\hat{R}\rho(\vec{r}) = \hat{R}\sum_{nlm} c_{nlm} g_{nl}(r) Y_{lm}(\theta,\phi) \quad = \sum_{nlm} c_{nlm} g_{nl}(r)\left(\hat{R} Y_{lm}(\theta,\phi)\right)$$

  - Spherical harmonics under rotation $\hat{R}Y_{lm}(\theta,\phi) = D^{l}_{mm'}(\hat{R})Y_{lm'}$, $\mathbf{D}^{l}(\hat{R})$'s are $(2l+1)(2l+1)$ matrices, called Wigner matrix.

$$\hat{R}\rho(\vec{r}) = \sum_{n=1}^{n_{\max}} \sum_{l=0}^{l_{\max}} \sum_{m=-l}^{+l} \sum_{m'=-l}^{+l} c_{nlm} g_{nl}(r)\left(D^{l}_{mm'}(\hat{R}) Y_{lm'}(\theta,\phi)\right)$$

$$\hat{R}\rho(\vec{r}) = \sum_{n=1}^{n_{\max}} \sum_{l=0}^{l_{\max}} \sum_{m'=-l}^{+l} \left(\sum_{m=-l}^{+l} c_{nlm} D^{l}_{mm'}(\hat{R})\right) g_{nl}(r) Y_{lm'}(\theta,\phi)$$

$$= \sum_{n=1}^{n_{\max}} \sum_{l=0}^{l_{\max}} \sum_{m'=-l}^{+l} c_{nlm'} g_{nl}(r) Y_{lm'}(\theta,\phi); \text{ with } c_{nlm'} \equiv \sum_{m=-l}^{+l} c_{nlm} D^{l}_{mm'}(\hat{R})$$

- The rotation, thus, can be thought of in terms of the transformation of the coeficients c_{nlm}'s
- That is, the function is expressed in terms of an initial set of Y_{lm}'s (c_{nlm}) and a rotated set of Y_{lm}'s ($c_{nlm'}$).

- Wigner matrices are unitary
  - $\mathbf{D}^{l\dagger}\mathbf{D}^{l} = \mathbf{I}$, unit matrix.
- Therefore, considering c_{nlm} for fixed (n, l) as a column vector $\mathbf{c}_{nl} = \begin{pmatrix} c_{n,l,-l} \\ c_{n,l,-l+1} \\ \vdots \\ c_{n,l,l} \end{pmatrix}$,
  - $c_{nlm'} \equiv \sum_{m=-l}^{+l} c_{nlm} D^{l}_{mm'}(\hat{R})$ can be written as $\mathbf{c}^{'}_{nl} = \mathbf{D}^{l}\mathbf{c}_{nl}$.
- $\mathbf{c}^{'\dagger}_{nl}\mathbf{c}'_{nl} = (\mathbf{c}^{\dagger}_{nl}\mathbf{D}^{l\dagger})(\mathbf{D}^{l}\mathbf{c}_{nl}) = \mathbf{c}^{\dagger}_{nl}\mathbf{c}_{nl}$.
- Thus $\mathbf{c}^{\dagger}_{nl}\mathbf{c}_{nl}$ is a quantity that remains invariant under rotations.
- This property is used to construct a rotation-invariant feature vector
  - $p^{Z_1,Z_2}_{nn'l} = \sum_{m=-l}^{+l} (c^{Z_1}_{nlm})^{*}(c^{Z_2}_{n'lm})$.
  - Concatenating $p^{Z_1,Z_2}_{nn'l}$'s for all possible (Z_1, Z_2) pairs, and all possible (nl) combinations produce the SOAP feature vector.

- SOAP construction

  - Atomic density field
  - Expand in terms of radial and angular basis functions
  - Usually spherical primitive Gaussian and polynomial radial basis
  - Make them orthonormal
  - Once they are made orthonormal, the expansion coefficients

    $$c_{nlm} = \int_V g_n(r) Y^*(\theta, \phi) \rho(\vec{r}) dv$$

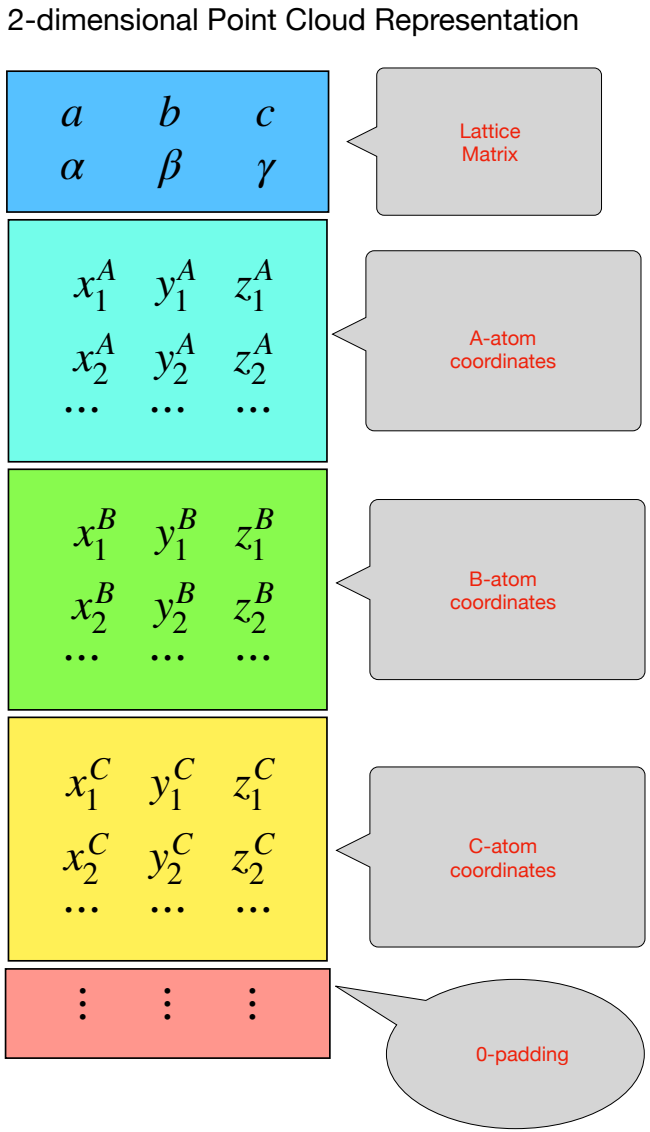
  - Once c's are known, construct $p_{nn'l}^{Z_1,Z_2} = \sum_{m=-l}^{+l} (c_{nlm}^{Z_1})^* (c_{n'lm}^{Z_2})$
  - Concatenation of p's is the SOAP

Global features

AnBmCl

- **Point cloud representation:** pictorial representations
  - 3D voxel image, values of some quantity on a 3D grid

ational resources



2D point cloud representation of a ternary material

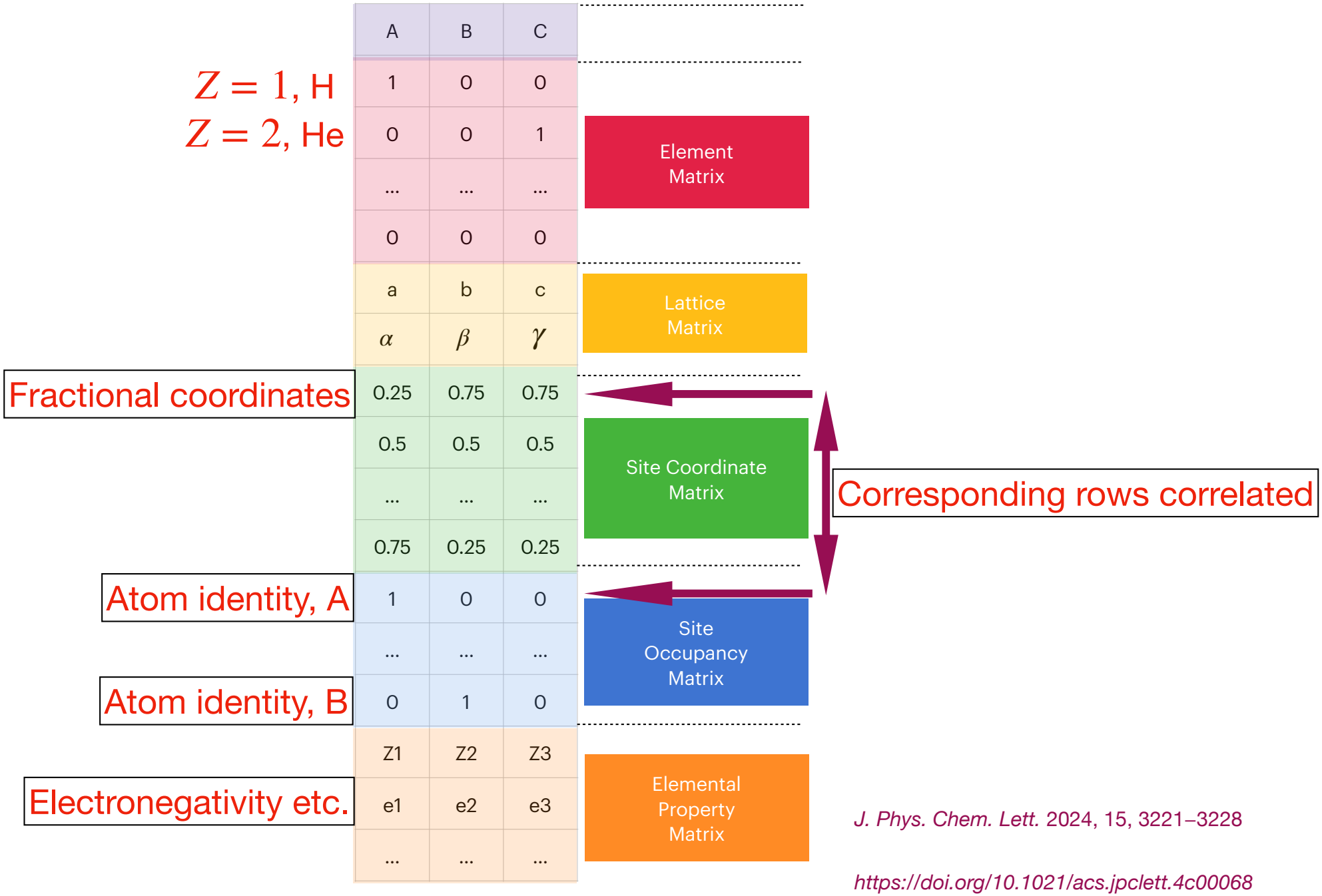
Accompanied by 'composition condition'

$(N_{\max} + 2) \times 3$ image

$N_{\max}$ is max number of atoms in the unit cell in the entire training set

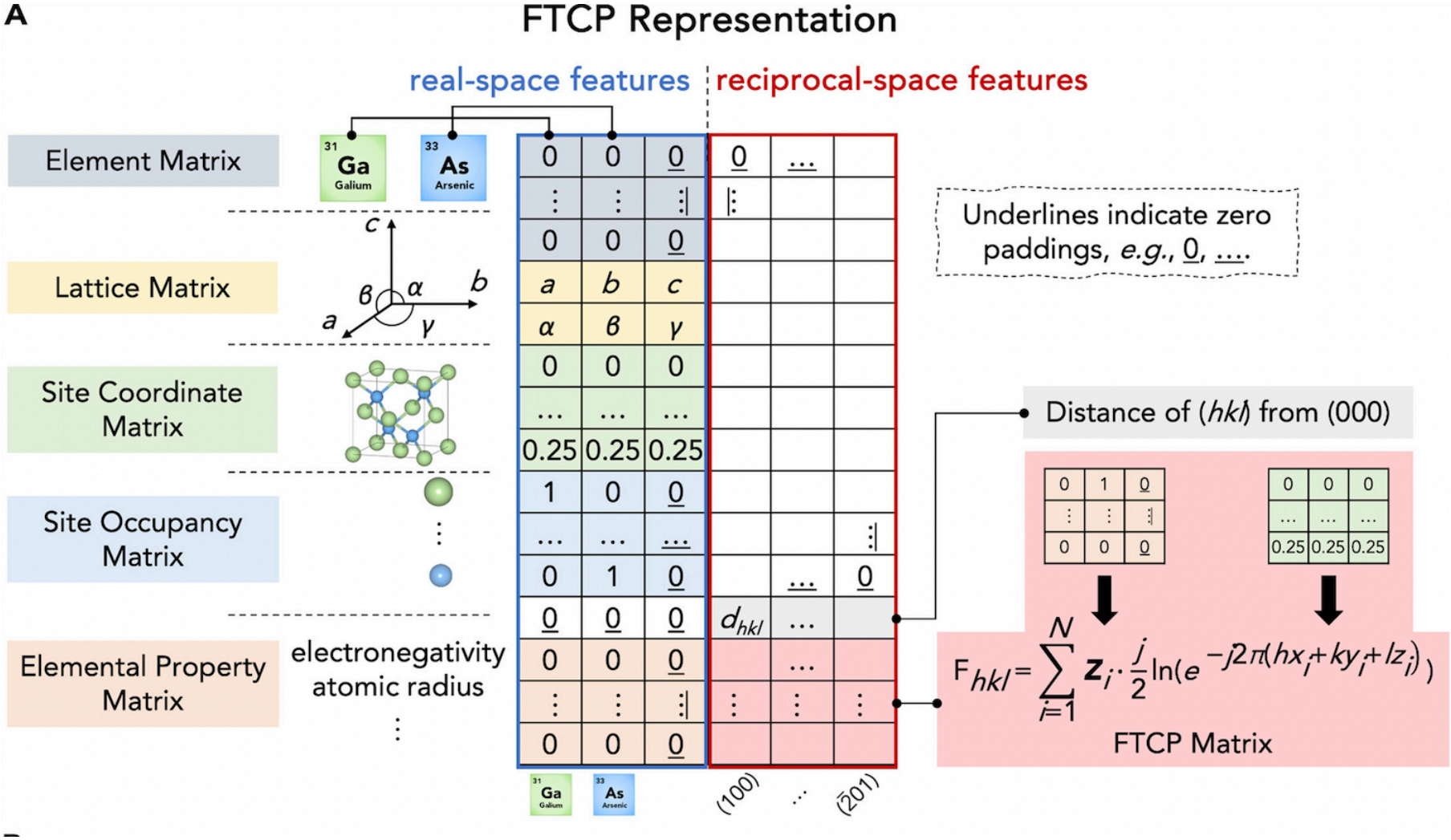
ACS Cent. Sci. 6 (8), 1412–1420. doi:10.1021/acscentsci.0c00426

IRCR Representation



J. Phys. Chem. Lett. 2024, 15, 3221–3228

https://doi.org/10.1021/acs.jpclett.4c00068



Ren et al., Matter 5, 314–335

https://doi.org/10.1016/j.matt.2021.11.032