A QSPR Methodology Based on the Dynamic Analysis of Crystal Lattice Voids
Mathematical and Computational Specification at the Voxel and Topological Level

Daniel M. Kamiński

July 2026

Introduction and Physical Principles

Modeling the thermodynamic stability of crystal lattices using quantum mechanical methods (e.g., DFT-D) is highly demanding computationally. The proposed QSPR (*Quantitative Structure-Property Relationship*) methodology offers a highly efficient alternative by evaluating the **dynamic scanning of lattice cavities (voids)** as a function of the van der Waals radius multiplier:

$$f_{vdW} \in [1.0,1.35]$$

This scanning procedure acts as a **quasi-temperature substitute**. As f_{vdW} increases, the virtual atomic spheres swell, mimicking the contraction of free configurational space during thermal cooling and the freezing of lattice vibrations. Tracking the geometric and physicochemical trends on the cavity boundary surfaces across this range provides a unique fingerprint of the crystal's potential energy surface.

—

Mathematical Formalism of the Data Representation

In the computational pipeline, the unit cell is discretized into a three-dimensional voxel grid (VoxelGrid).

Voxel Space Definition

Every identified cavity k at a given scanning step f_{vdW} consists of a set of discrete boundary surface voxels $i \in \{1,2,\dots,N_k\}$. A single voxel i is defined in the coordinate and property space as:

$$\mathbf{v}_i = \left(x_i, y_i, z_i, v_i, s_i, q_i, h_i, Z_{eff,i}\right)$$

Legend (single voxel data structure):

- x_i, y_i, z_i – Cartesian coordinates of the voxel's center of gravity [Å],

- v_i – voxel volume [Å^3]
- s_i – fractional surface area patch assigned to voxel i [Å^2] (e.g., generated via Marching Cubes),
- q_i – local partial charge assigned to the voxel [e], induced by the surrounding atomic environment,
- h_i – normalized hydrophobicity coefficient of the voxel, where $h_i \in [-1.0, 1.0]$,
- $Z_{eff,i}$ – effective atomic number mapping at the voxel's boundary position.

Mapping the Effective Atomic Number ($Z_{eff,i}$)

To capture the chemical polarizability of the atoms forming the cavity walls (e.g., distinguishing a heavy, polarizable chlorine or sulfur atom from a light hydrogen atom), each surface voxel i is assigned an effective atomic number $Z_{eff,i}$:

$$Z_{eff,i} = \sum_{j \in \mathcal{A}} \frac{Z_j}{\| \mathbf{r}_i - \mathbf{R}_j \|^2}$$

Legend:

- $\mathcal{A}$ – the set of atoms in the nearest neighborhood of the cavity (within a specified cutoff radius R_{cut}),
- Z_j – atomic number of the j-th atom (e.g., $Z_H = 1, Z_C = 6, Z_O = 8, Z_S = 16$),
- $\mathbf{r}_i$ – position vector of the voxel i: $\mathbf{r}_i = [x_i, y_i, z_i]^T$,
- $\mathbf{R}_j$ – position vector of the nucleus of the j-th atom: $\mathbf{R}_j = [X_j, Y_j, Z_j]^T$,
- $\| \mathbf{r}_i - \mathbf{R}_j \|$ – Euclidean distance between the voxel and the atomic nucleus [Å].

—

Detailed QSPR Descriptor Dictionary

The following descriptors are aggregated globally for the entire crystal structure based on the physical evolution of cavities during f_{vdW} modulation.

Category 1: Topological and Geometric Metrics

Network Fragmentation Index (I_{frag})

Measures the degree to which the free volume breaks down into narrow micro-fissures as the atomic spheres swell:

$$I_{frag} = \frac{\max_f N_{cav}(f)}{N_{atoms}}$$

Legend:

- $N_{cav}(f)$ – the total count of distinct cavities detected at the scanning step $f = f_{vdW}$,
- $\max_f N_{cav}(f)$ – the maximum recorded number of cavities within the range $f \in [1.0,1.35]$,
- N_{atoms} – total number of atoms in the unit cell (normalization parameter).

Physical Meaning: A high I_{frag} value indicates an unstable, heavily fractured packing arrangement, where small thermal fluctuations can easily disrupt lattice continuity.

Critical Collapse Scale ($f_{collapse}$)

Identifies the geometric limit at which the free volume of the crystal completely vanishes:

$$f_{collapse} = \min\left\{ f \;\middle|\; \frac{\sum_k V_k(f)}{V_{cell}} < \epsilon \right\}$$

Legend:

- $V_k(f)$ – volume of the k-th cavity at multiplier f [Å^3],
- V_{cell} – total volume of the crystal unit cell [Å^3],
- ϵ – numerical threshold for cavity extinction (default $\epsilon = 0.001$, i.e., 0.1% of free space remaining).

Physical Meaning: A low $f_{collapse}$ value (close to 1.0) demonstrates a highly optimized, dense, and rigid initial molecular packing.

—

Category 2: Electrostatic Metrics

Normalized Electrostatic Cavity Footprint (Ω_Q)

Measures the cumulative electrostatic potential energy "trapped" at the boundaries of unoccupied cavities:

$$\Omega_Q = \frac{1}{M_{cell}} \int_{1.0}^{1.35} \left(\sum_k \sigma_{q,k}^2(f) \cdot V_k(f) \right) df$$

Where the charge variance of cavity k at step f is defined as:

$$\sigma_{q,k}^2(f) = \frac{1}{N_k}\sum_{i=1}^{N_k}(q_i - \bar{q}_k)^2\,, \quad \bar{q}_k = \frac{1}{N_k}\sum_{i=1}^{N_k} q_i$$

Legend:

- $\sigma_{q,k}^2(f)$ – surface charge variance of the k-th cavity at step f [e^2],
- $\bar{q}_k$ – mean charge assigned to the walls of cavity k [e],
- $V_k(f)$ – volume of the k-th cavity [Å^3],
- M_{cell} – molar mass of the unit cell contents [g/mol] (mass-based normalization parameter),
- df – differential step size of the f_{vdW} scan.

Physical Meaning: High values of Ω_Q reveal the presence of highly polar, uncompensated charge regions directly facing empty space, driving strong electrostatic frustration within the lattice.

Maximum Dipole Sensitivity (α_{polar})

Determines the peak electrostatic responsiveness of cavities to minor atomic displacements:

$$\alpha_{polar} = \max_f \left(\max_k \left\| \frac{d\boldsymbol{\mu}_k(f)}{df} \right\| \right)$$

Where the dipole moment vector of cavity k at step f is:

$$\boldsymbol{\mu}_k(f) = \sum_{i=1}^{N_k} q_i \cdot (\mathbf{r}_i - \mathbf{r}_{cog,k})$$

Legend:

- $\boldsymbol{\mu}_k(f)$ – net dipole moment vector of cavity k [e · Å],
- $\mathbf{r}_{cog,k}$ – geometric center of gravity vector of cavity k [Å],
- $\left\| \frac{d\boldsymbol{\mu}_k(f)}{df} \right\|$ – norm of the derivative of the dipole vector with respect to the scaling parameter.

Physical Meaning: Shows how abruptly the electrostatic fields around a void expose polar anisotropy during lattice vibrations. High sensitivity values correlate with lower dynamic stability.

—

Category 3: Hydrophobic and Polarizability Metrics

Hydrophobic Frustration Density (Ψ_H)

Evaluates the chemical heterogeneity and domain mismatch on the cavity walls:

$$\Psi_H = \frac{1}{V_{cell}} \int_{1.0}^{1.35} \left(\sum_k \sigma_{h,k}^2 (f) \cdot S_k(f) \right) df$$

Where the hydrophobicity variance of cavity k is given by:

$$\sigma_{h,k}^2(f) = \frac{1}{N_k} \sum_{i=1}^{N_k} \left(h_i - \bar{h}_k\right)^2, \quad \bar{h}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} h_i$$

Legend:

- $\sigma_{h,k}^2(f)$ – variance of the hydrophobicity index on the walls of cavity k at step f,
- $S_k(f)$ – surface area of the k-th cavity [Å^2],
- h_i – hydrophobicity of the i-th voxel, $\bar{h}_k$ – average hydrophobicity of the cavity,
- V_{cell} – unit cell volume [Å^3] (volume-based normalization parameter).

Physical Meaning: According to solvophobic principles, stable networks favor clear boundary segregation between hydrophobic and hydrophilic regions. High Ψ_H values indicate poorly resolved, frustrated chemical surfaces.

Initial Mean Polarizability ($\langle Z_{cav} \rangle$)

Quantifies the stabilizing contribution of dispersion forces (London interactions) at the cavity boundaries:

$$\langle Z_{cav} \rangle = \left. \frac{\sum_k \sum_{i=1}^{N_k} Z_{eff,i}}{\sum_k N_k} \right|_{f_{vdW}=1.0}$$

Legend:

- $\sum_k N_k$ – sum of all surface voxels across all cavities at the unperturbed initial step ($f_{vdW} = 1.0$).
- $Z_{eff,i}$ – mapped effective atomic number of the i-th voxel from equation (3).

—

Model Normalization Matrix

To map these descriptors to lattice energies (E_{net}) derived from crystal systems of vastly different sizes and elemental compositions, three complementary normalization pathways are utilized:

Physical Evaluation of Descriptor Normalization Methods

Normalization	Formula & Dimensions	Physical Interpretation
Volume-based	$\bar{X}_V = \frac{X}{V_{cell}}$ $[\text{Unit} \cdot \text{Å}^{-3}]$	Defines the *spatial density* of a phenomenon. Essential for representing hydrophobic frustration (Ψ_H).
Atom-based	$\bar{X}_A = \frac{X}{N_{atoms}}$ $[\text{Unit} \cdot \text{atom}^{-1}]$	Represents the local structural overhead mapped to a single atomic site. Used for I_{frag}.
Mass-based	$\bar{X}_M = \frac{X}{M_{cell}}$ $[\text{Unit} \cdot \text{mol} \cdot \text{g}^{-1}]$	Standardizes the model relative to mass density. Crucial for comparing light organic polymorphs with heavy solvated structures (Ω_Q).

—

Lattice Stability Equation and Model Evaluation

The integrated QSPR model predicts the normalized lattice energy (E_{net}) or free enthalpy change (ΔG) through a multi-parameter linear regression framework:

$$\frac{E_{net}}{M_{cell}} = E_0 + w_V \cdot \left(\frac{V_{init}}{V_{cell}}\right) + w_Q \cdot \Omega_Q + w_H \cdot \Psi_H + w_{frag} \cdot I_{frag} - w_Z \cdot \langle Z_{cav} \rangle$$

Legend of weights (w_x):

- E_0 – reference energy for an ideally packed, void-free crystal lattice,
- $w_V > 0$ – geometric destabilization penalty (larger initial free volumes V_{init} increase the overall lattice energy),
- $w_Q > 0$ – energy cost of electrostatic frustration within the voids,
- $w_H > 0$ – thermodynamic penalty of hydrophobic/polar domain mismatch,
- $w_{frag} > 0$ – weight representing lattice vulnerability to dynamic fragmentation,

- $w_Z > 0$ – dispersion stabilization coefficient (the negative sign ensures that highly polarizable atoms with large Z_{eff} bordering the voids lower the net lattice energy, enhancing crystal stability).