Advanced Cavity Detection and Surface Property Mapping in Crystalline Materials
A Comprehensive Computational Framework

Daniel M. Kamiński

16.07.2026

We present a comprehensive computational framework for the detection, analysis, and visualization of cavities and channels in crystalline materials. The methodology integrates advanced algorithms including Signed Distance Function (SDF) generation, A* pathfinding with gradient-based optimization, and surface property mapping using quantum-mechanically derived atomic charges and hydrophobicity scales. The framework supports all crystal systems including triclinic cells with non-orthogonal angles, providing quantitative metrics for channel characterization including length, width, helicity, and cell crossings. Surface property mapping employs the Gasteiger-Marsili (PEOE) method for charge distribution and the Eisenberg-Weiss scale for hydrophobicity, projecting these properties onto cavity surfaces for intuitive visualization. The system generates comprehensive reports with QSPR descriptors and host-guest compatibility assessments, making it a valuable tool for materials science and pharmaceutical research.

Introduction

The characterization of void spaces in crystalline materials is of fundamental importance in materials science, chemistry, and pharmaceutical research. Cavities and channels within crystal structures govern critical properties including guest molecule diffusion, molecular sieving, catalysis, and host-guest interactions . The ability to quantitatively describe these spaces and map their physicochemical properties is essential for rational design of porous materials .

Traditional methods for cavity analysis rely on geometric approaches such as Voronoi tessellation and Delaunay triangulation . While effective, these methods often lack the ability to capture the continuous nature of channels and provide limited physicochemical information. Furthermore, the analysis of non-orthogonal crystal systems presents significant computational challenges.

Our framework addresses these limitations through an integrated approach combining:

1. **SDF-based cavity detection** with support for all crystal systems
2. **A* pathfinding with gradient optimization** for channel identification
3. **Quantitative channel metrics** including width, helicity, and cell crossings
4. **Surface property mapping** using quantum-mechanical atomic charges and hydrophobicity scales
5. **Comprehensive reporting** with QSPR descriptors and host-guest compatibility

This paper presents the mathematical foundations, implementation details, and applications of our framework, demonstrating its utility for a wide range of crystalline materials.

Theoretical Background

Signed Distance Function

The Signed Distance Function $\phi(\mathbf{r})$ is defined as the signed Euclidean distance to the nearest atom surface:

$$\phi(\mathbf{r}) = \begin{cases} \min_i(\| \mathbf{r} - \mathbf{r}_i \| - R_i) & \text{outside atoms} \\ -\min_i(R_i - \| \mathbf{r} - \mathbf{r}_i \|) & \text{inside atoms} \end{cases}$$

where $\mathbf{r}_i$ and R_i are the position and van der Waals radius of atom i, respectively. The cavity space is defined by $\phi(\mathbf{r}) \geq 0$, while the wall space is defined by $\phi(\mathbf{r}) < 0$.

For practical implementation, the SDF is discretized on a Cartesian grid:

$$\phi_{i,j,k} = \phi(\mathbf{r}_{i,j,k})$$

where $\mathbf{r}_{i,j,k} = (i\Delta x, j\Delta y, k\Delta z)$ with grid spacings $\Delta x, \Delta y, \Delta z$.

Gradient of SDF

The gradient of the SDF is computed using central differences:

$$\nabla\phi_{i,j,k} \approx \left(\frac{\phi_{i+1,j,k} - \phi_{i-1,j,k}}{2\Delta x}, \frac{\phi_{i,j+1,k} - \phi_{i,j-1,k}}{2\Delta y}, \frac{\phi_{i,j,k+1} - \phi_{i,j,k-1}}{2\Delta z}\right)$$

The gradient magnitude $g_{i,j,k} = \| \nabla\phi_{i,j,k} \|$ indicates proximity to atom surfaces, with larger values near walls.

Cavity Detection

Flood-Fill Algorithm

Cavity detection employs a flood-fill algorithm on the discretized SDF grid. Starting from a seed point $\mathbf{r}_0$ where $\phi(\mathbf{r}_0) \geq 0$, the algorithm iteratively explores neighboring voxels:

$$\mathcal{C} = \{\mathbf{r} : \phi(\mathbf{r}) \geq 0 \text{ and connected to } \mathbf{r}_0\}$$

Connectivity is defined using 26-neighborhood in 3D space:

$$\mathcal{N}(\mathbf{r}) = \{\mathbf{r} + \Delta\mathbf{r} : \Delta\mathbf{r} \in \{-1,0,1\}^3 \backslash \{(0,0,0)\}\}$$

Wall Contact Analysis

For each detected cavity, we compute the number of voxels touching each unit cell wall:

$$W_X^- = \sum_{\mathbf{r} \in \mathcal{C}} \mathbb{I}\,(r_x = X_{min})$$

$$W_X^+ = \sum_{\mathbf{r} \in \mathcal{C}} \mathbb{I}\,(r_x = X_{max})$$

where X_{min} and X_{max} are the minimum and maximum x-coordinates of the periodic cell. Similar definitions apply for Y and Z directions.

Channel Path Finding

A* Algorithm

Channel path finding is formulated as a shortest path problem in the cavity space. Given a start point $\mathbf{s}$ and a target wall W, we seek a path $\mathcal{P} = \{\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_n\}$ with $\mathbf{p}_0 = \mathbf{s}$ and $\mathbf{p}_n \in W$.

The A* algorithm minimizes the cost function:

$$f(\mathbf{p}) = g(\mathbf{p}) + h(\mathbf{p})$$

where $g(\mathbf{p})$ is the actual cost from start to $\mathbf{p}$ and $h(\mathbf{p})$ is the heuristic estimate from $\mathbf{p}$ to the target.

Cost Function with Gradient

The step cost from $\mathbf{a}$ to $\mathbf{b}$ is:

$$c(\mathbf{a}, \mathbf{b}) = \| \mathbf{b} - \mathbf{a} \| + \alpha \cdot \frac{\| \nabla \phi(\mathbf{b}) \|}{\phi_{max}} \cdot \| \mathbf{b} - \mathbf{a} \|$$

where α is the gradient weight (typically 0.3) and ϕ_{max} is the maximum expected gradient magnitude.

This formulation encourages paths through the center of channels where the gradient magnitude is small.

Heuristic Function

The heuristic function estimates the remaining distance to the target wall:

$$h(\mathbf{p}) = \begin{cases} |p_x - w_x| & \text{for XX channels} \\ |p_y - w_y| & \text{for YY channels} \\ |p_z - w_z| & \text{for ZZ channels} \end{cases}$$

where w is the target wall coordinate.

Path Smoothing

The raw path from A* is smoothed using iterative weighted averaging:

$$\mathbf{p}_i^{(k+1)} = \frac{1}{4}\left(\mathbf{p}_{i-1}^{(k)} + 2\mathbf{p}_i^{(k)} + \mathbf{p}_{i+1}^{(k)}\right)$$

This is equivalent to a discrete diffusion process:

$$\frac{\partial \mathbf{p}}{\partial t} = \frac{\partial^2 \mathbf{p}}{\partial x^2}$$

The smoothing error is computed as:

$$E_{RMS} = \sqrt{\frac{1}{n}\sum_{i=0}^{n} \| \mathbf{p}_i - \tilde{\mathbf{p}}_i \|^2}$$

Channel Metrics

Channel Length

The total channel length is:

$$\ell = \sum_{i=1}^{n} \| \tilde{\mathbf{p}}_i - \tilde{\mathbf{p}}_{i-1} \|$$

where $\tilde{\mathbf{p}}_i$ are smoothed path points.

Channel Width

The local width at point $\mathbf{p}$ is estimated by ray casting:

$$w(\mathbf{p}) = 2 \cdot \min\{r > 0 : \exists \theta, \phi \text{ such that } \phi\left(\mathbf{p} + r \cdot \hat{\mathbf{r}}(\theta, \phi)\right) < 0\}$$

The average, minimum, and maximum widths are:

$$\bar{w} = \frac{1}{m}\sum_{j=0}^{m} w\left(\mathbf{p}_{j \cdot \lfloor n/20 \rfloor}\right)$$

$$w_{min} = \min_j w\left(\mathbf{p}_{j \cdot \lfloor n/20 \rfloor}\right)$$

$$w_{max} = \max_j w\left(\mathbf{p}_{j \cdot \lfloor n/20 \rfloor}\right)$$

Effective Diameter

The effective diameter is approximated as:

$$d_{eff} = 0.8 \cdot \bar{w}$$

Helicity

The helicity measures the twist of the channel:

$$\theta_{hel} = \arccos\left(\frac{(\tilde{\mathbf{p}}_m - \tilde{\mathbf{p}}_0) \cdot (\tilde{\mathbf{p}}_n - \tilde{\mathbf{p}}_m)}{\| \tilde{\mathbf{p}}_m - \tilde{\mathbf{p}}_0 \| \cdot \| \tilde{\mathbf{p}}_n - \tilde{\mathbf{p}}_m \|}\right) \cdot \frac{180}{\pi}$$

where $m = \lfloor n/2 \rfloor$ is the midpoint of the path.

Cell Crossings

The number of periodic cell crossings is:

$$C = \sum_{i=1}^{n} \left[\mathbb{I}(|f_x(i) - f_x(i-1)| > 0.5) + \mathbb{I}(|f_y(i) - f_y(i-1)| > 0.5) + \mathbb{I}(|f_z(i) - f_z(i-1)| > 0.5)\right]$$

where $f(i)$ are fractional coordinates.

Surface Property Mapping

Charge Calculation: Gasteiger-Marsili Method

Atomic charges are calculated using the Partial Equalization of Orbital Electronegativity (PEOE) method . The electronegativity χ_i of atom i is expressed as:

$$\chi_i = a_i + b_i q_i + c_i q_i^2$$

where q_i is the atomic charge and a_i, b_i, c_i are atom-type specific parameters.

The charge transfer between atoms i and j is:

$$\Delta q_{i \to j} = \frac{\chi_j - \chi_i}{\eta_{ij}}$$

where η_{ij} is the hardness parameter. The algorithm iterates until convergence:

$$\max_i \left| q_i^{(k+1)} - q_i^{(k)} \right| < \varepsilon$$

Hydrophobicity Mapping

Hydrophobicity values are assigned using the Eisenberg-Weiss consensus scale :

$$h_i = h_0(Z_i) - \sum_{j \in \mathcal{N}(i)} \gamma \cdot \mathbb{I}\left(h_0(Z_j) < h_{polar}\right)$$

where $h_0(Z)$ is the base hydrophobicity of element Z, $\mathcal{N}(i)$ is the set of bonded neighbors, γ is the neighbor penalty (0.15), and h_{polar} is the polar threshold (-0.2).

Surface Interpolation

For each cavity vertex $\mathbf{v}$, the property value is computed via weighted averaging:

$$P(\mathbf{v}) = \frac{\sum_i w_i \, p_i}{\sum_i w_i}$$

where the weight is based on distance:

$$w_i = \frac{1}{\| \mathbf{v} - \mathbf{r}_i \| + \varepsilon}$$

Color Mapping

Property values are mapped to colors using a sigmoidal transformation:

$$\mathbf{c}(P) = \frac{1}{1 + e^{-\beta(P - P_0)}} \cdot \mathbf{c}_{max} + \mathbf{c}_{min}$$

where β controls contrast, P_0 is the midpoint, and $\mathbf{c}_{min}, \mathbf{c}_{max}$ are color endpoints.

QSPR Descriptors

Fragmentation Index

The fragmentation index quantifies the degree of cavity fragmentation:

$$I_{frag} = \frac{N_{cav}}{N_{atom}}$$

where N_{cav} is the number of cavities and N_{atom} is the number of atoms in the unit cell.

Critical Collapse Scale

The critical collapse scale is the VdW scale at which cavities disappear:

$$f_{collapse} = \min\{s : V_{cav}(s) = 0\}$$

Electrostatic Frustration

Electrostatic frustration measures charge inhomogeneity:

$$\Omega_Q = \int_0^1 \sigma_Q(s) \cdot V_{cav}(s) \, ds$$

where $\sigma_Q(s)$ is the standard deviation of charges at scale s.

Hydrophobic Frustration

Hydrophobic frustration is similarly defined:

$$\Psi_H = \int_0^1 \sigma_H(s) \cdot A_{cav}(s)\, ds$$

where $\sigma_H(s)$ is the standard deviation of hydrophobicity at scale s.

Mean Polarizability

The mean polarizability is the average effective nuclear charge:

$$\langle Z_{cav} \rangle = \frac{1}{N} \sum_i Z_i^{eff}$$

Host-Guest Compatibility

Volume Matching

The volume compatibility score is:

$$S_V = 1 - \frac{|V_g - 0.7V_c|}{1.3V_c}$$

where V_g is guest volume and V_c is cavity volume.

Channel Fit

The channel fit score is based on diameter ratio:

$$S_D = \begin{cases} 1.0 & d_g/d_w < 0.7 \\ 0.8 & 0.7 \le d_g/d_w < 0.9 \\ 0.5 & 0.9 \le d_g/d_w < 1.1 \\ 0.1 & d_g/d_w \ge 1.1 \end{cases}$$

Overall Compatibility

The overall compatibility score is a weighted sum:

$$S_{total} = 0.35S_V + 0.30S_D + 0.15S_H + 0.10S_C + 0.10S_A$$

where S_H is helicity score, S_C is charge compatibility, and S_A is shape affinity.

Performance Considerations

The computational complexity of the framework is dominated by the A* pathfinding algorithm:

$$\mathcal{O}(N \log N)$$

where N is the number of visited nodes. For typical channel lengths:

$$N \approx \frac{L_c}{\Delta x} \cdot \frac{W_c}{\Delta x}$$

where L_c is channel length and W_c is channel width.

Conclusion

The framework presented here provides a comprehensive solution for cavity and channel analysis in crystalline materials. By integrating SDF-based detection, gradient-guided pathfinding, and surface property mapping, it enables detailed characterization of void spaces and their physicochemical properties. The methodology supports all crystal systems and provides quantitative metrics essential for materials design and pharmaceutical research.

Future work will focus on extending the framework to include diffusion coefficient estimation, machine learning-based channel classification, and real-time property visualization.

References

Kitagawa, S., Kitaura, R., & Noro, S. (2004). Functional porous coordination polymers. *Angewandte Chemie International Edition*, https://doi.org/10.1002/anie.200300610

Zhou, H. C., Long, J. R., & Yaghi, O. M. (2012). Introduction to metal-organic frameworks. *Chemical Reviews*, https://doi.org/10.1021/cr300014x

Slater, A. G., & Cooper, A. I. (2017). Porous materials function from design. DOI: 10.1126/science.aaa8075

Furukawa, H., Cordova, K. E., O'Keeffe, M., & Yaghi, O. M. (2013). DOI: 10.1126/science.1230444

Medvedev, N. N., Voloshin, V. P., & Gavrilov, A. V. (2006). Voronoi-Delaunay analysis of voids in molecular crystals. *Journal of Molecular Graphics and Modelling*, 25(4), 550-558.

Blatov, V. A., Shevchenko, A. P., & Proserpio, D. M. (2014). Applied topological analysis of crystal structures with the program package ToposPro. *Crystal Growth & Design*, 14(7), 3576-3586. DOI:10.1021/cg500498k

Gasteiger, J., & Marsili, M. (1980). Iterative partial equalization of orbital electronegativity - a rapid access to atomic charges. https://doi.org/10.1016/0040-4020(80)80168-2

Eisenberg, D., & Weiss, R. M. (1982). Hydrophobic moments and protein structure. https://doi.org/10.1039/FS9821700109

Bondi, A. (1964). van der Waals volumes and radii. *Journal of Physical Chemistry*, 68(3), 441-451.

Hart, P. E., Nilsson, N. J., & Raphael, B. (1972). A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*. **DOI:** 10.1109/TSSC.1968.300136

A fast algorithm for calculating van der Waals volumes … DOI:10.1021/jo034808o

The topological analysis of crystal structures. *Acta Crystallographica Section B*, https://doi.org/10.1107/S0108768183003201