CHEMICAL REVIEWS

pubs.acs.org/CR

Review

Topological Data Analysis in Materials Science: Principles, Machine Learning Integration, and Application Landscapes

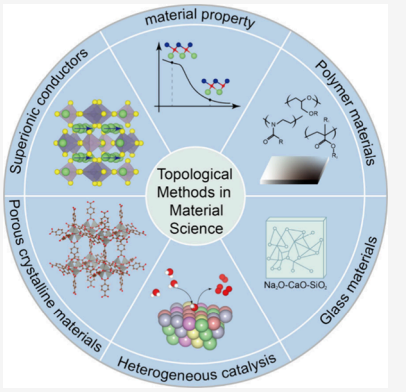
Shisheng Zheng,[#] Bingxu Wang,[#] Mengmeng Zhang, Jiawei Tan, Hang-Biao Lv, Ran Liu, Zhong-Zhang Shi, Xin Li, Ruoxuan Wang, Siwang Zhang, Jingyan Li, Yingru Zhao,* Jie Wu,* Jian-Feng Li,* Shing-Tung Yau,* and Feng Pan*

Cite This: *Chem. Rev.* 2026, 126, 7029−7067

Read Online

ACCESS | Metrics & More | Article Recommendations

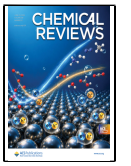
ABSTRACT: The data-driven paradigms are reshaping materials science, placing unprecedented emphasis on the digital characterization and representation of the material data. Topological Data Analysis (TDA) methods offer a promising and innovative approach to uncovering data patterns in complex material systems from a topological perspective. By focusing on topological invariants across multiscale resolution, TDA offers unique robustness, interpretability, and universality for the discovery of structure−property relationships. This review presents the mathematical foundations of TDA, with a focus on key methodologies such as persistent homology, persistent GLMY homology, and Euler characteristic curves. We further introduce the emerging field of topological learning, which integrates TDA with machine learning to enhance the development of predictive models. Representative applications of TDA are discussed in diverse material systems. In addition, we briefly highlight its emerging applications in related domains such as molecular science and biochemistry. Building on these examples, we summarize the unique advantages TDA offers in materials informatics, along with the critical challenges that remain. This review aims to provide a practical guide for materials scientists seeking to apply TDA in data-driven research, and to highlight its growing potential in advancing materials discovery and design.



CONTENTS

1. Introduction 7030
2. THE FOUNDATION OF TDA 7030
2.1. Persistent Homology 7031
2.1.1. Simplices and Simplicial Complexes 7031
2.1.2. Homology 7031
2.1.3. Filtration and Persistent Homology 7032
2.1.4. Visualization of Persistent Homology 7032
2.1.5. Practical Implementation of Persistent Homology 7033
2.2. GLMY Theory 7034
2.2.1. GLMY Homology 7035
2.2.2. Generalized GLMY Homology 7036
2.3. Euler Characteristic 7037
2.3.1. The Basis of Euler Characteristic 7037
2.3.2. Filtration And Euler Characteristic Curve 7038
2.4. Comparison of Topological Descriptors 7039
3. TOPOLOGICAL LEARNING 7040
3.1. Explicit Topological Representations 7040
3.2. Implicit Topological Regularization 7042
3.3. Topology-Aware Architecture 7042
4. The Applications of TDA in Materials Science 7043
4.1. Fundamental Material Property 7043
4.2. Polymer Materials 7045
4.3. Glassy Materials 7047
4.4. Heterogeneous Catalysis 7049
4.5. Porous Crystalline Materials 7051
4.6. Superionic Conductors 7052
4.7. Applications in Additional Material Systems 7054
4.8. Summary of Material Data Types for TDA 7054
Atomic Coordinates 7054
MD Trajectories 7054
Imaging Data 7054
Potential energy surfaces 7054
5. EXTENSION OF TDA TO MOLECULAR AND BIOCHEMICAL SYSTEMS 7055
5.1. Molecular Systems 7055
5.1.1. Molecular Energy Landscapes 7055
5.1.2. Molecular Organization 7055
5.1.3. Molecular Screening 7055
5.2. Biochemical Systems 7056

Received: December 16, 2025
Revised: May 25, 2026
Accepted: May 28, 2026
Published: June 9, 2026



Downloaded via UNIV TOWN OF SHENZHEN on June 26, 2026 at 05:56:19 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

ACS Publications

© 2026 American Chemical Society

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

5.2.1. Biochemical Structure and Interactions 7056
5.2.2. Biological Networks 7056
6. DISCUSSION AND OUTLOOK 7056
Author Information 7059
Corresponding Authors 7059
Authors 7059
Author Contributions 7059
Notes 7059
Biographies 7059
Acknowledgments 7061
References 7061

1. INTRODUCTION

In the realm of materials science, researchers of the past often relied on limited data, observations, and intuition to uncover fundamental patterns, e.g., the scaling relationship in heterogeneous catalysis.[1,2] With the advent of the data-driven era in modern research, scientists are now faced with vast amounts of material data. Extracting meaningful patterns from this immense and complex data by human intelligence has become a significant challenge.[3] In materials science, the data is not only voluminous but also highly complex.[4] This complexity is first reflected in the data format. While material structures are often represented as point clouds, their properties can exist as functions or even manifolds.[5] For example, the spatial distribution of continuous physical quantities such as electric fields, temperature fields, and density fields can be represented as scalar fields defined over space. Moreover, the structure of material data is intricately complex, spanning multiple levels, from atomic-level microstructural descriptions to macroscopic performance characteristics.[6] Features at these different scales are often interwoven and interdependent, forming highly complex, nonlinear, and high-dimensional data structures.[7]

The growing demand for analyzing complex data has driven rapid advancements in the field of machine learning. Today, researchers increasingly turn to various machine learning algorithms to extract insights from material data.[8] These analytical approaches can be unsupervised, uncovering the intrinsic structure and patterns within the data, or supervised, identifying potential relationships between the data and predefined targets.[9] Regardless of the approach, effective data characterization and representation remain crucial for successful modeling.[10−12] The goal of characterization is to understand the underlying structure and patterns within the data, while representation involves quantifying these features and transforming them into formats suitable for machine learning models, thus laying the groundwork for further computation and prediction.

Topology, as a major branch of mathematics, provides a natural and rigorous framework for characterization and representation of material data. Topology is dedicated to investigating the "shape" of objects and their intrinsic properties.[13] A range of topology-based methods have been developed for the characterization and representation of material data, such as knot theory and topological band theory.[14,15] In particular, topological data analysis (TDA), a prominent technique grounded in algebraic topology, has attracted increasing attention due to its ability to systematically enumerate and characterize geometric features in space with a solid mathematical foundation. By employing a suite of algebraic tools,[16] TDA can extract and quantify higher−order connectivity, loops and voids from complex data structures such as point clouds and manifolds.[17] When employed in material science, these topological invariants are often closely linked to key material properties, such as gas adsorption channels in porous materials, medium−range order (MRO) rings in glasses, and ion transport networks in superionic conductors.[18−21] Compared to traditional statistical descriptors, TDA offers enhanced robustness to noise and perturbation, as well as multiscale resolution of structural organization.[22] Notably, TDA−derived topological invariants serve as highly informative and interpretable features, facilitating their direct integration into ML pipelines for predictive modeling and establishing the structure−property relationships.[22−26] As a versatile data analysis technique, TDA has achieved significant success in fields such as molecular science, biochemistry, drug design, network science, and environmental science, with comprehensive reviews documenting its impact.[27−36] In addition, several reviews have focused on the capability of TDA to handle diverse types of chemistry data.[17,37] However, despite these developments across multiple disciplines, no systematic review currently exists to introduce TDA's principles, tools, and applications specifically tailored to materials science, impeding its broader adoption and advancement in this field.

Herein, we aim to bridge this gap by providing a conceptual and practical overview of TDA in material science. First, we introduce the mathematical foundations of TDA and several representative methodologies, including persistent homology, GLMY homology and Euler characteristic curve. We further introduce the emerging field of topological learning, which integrates TDA with machine learning to enhance the development of predictive models. We then review recent advances in the application of TDA in materials science, encompassing the prediction of fundamental properties and studies on polymers, glasses, crystalline porous materials, heterogeneous catalysts, superionic conductors, and other systems. Building on these examples, we highlight the advantages of TDA for structural characterization and representation, with particular emphasis on its interpretability, generalizability, and capacity for multiscale feature extraction. Finally, we discuss the current challenges in applying TDA to materials research and outline potential directions for future development.

2. THE FOUNDATION OF TDA

To understand how TDA can be applied to materials science, it is necessary to first introduce its mathematical foundations. TDA originates from topology, which is a branch of mathematics concerned with the qualitative properties of space that are preserved under continuous deformations such as stretching, bending, or compressing, but not tearing or gluing.[38] It focuses on so-called topological invariants, properties like connectedness, continuity, and the presence of holes, rather than precise geometric quantities like distances or angles. For example, a circle and a square are considered topologically equivalent because they can be continuously deformed into one another. Similarly, a torus and a coffee mug with a handle are topologically the same, as both contain one hole; a sphere and a cube are also equivalent, as neither has any holes. One of the central goals of topology is to classify spaces up to homeomorphism, and to characterize them using invariants such as the Euler characteristic, homology groups, and homotopy groups.[13] Two topological spaces are said to be

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

homeomorphic if there exists a homeomorphism between them, where a homeomorphism is a bijective continuous mapping whose inverse is also continuous, indicating that the two spaces are topologically equivalent. More comprehensive background on topology can be found in previous literatures.[13,39]

TDA leverages these ideas to study the shape and structure of data, offering a framework that is robust to noise and agnostic to coordinate systems.[40] Rather than treating data as individual points in space, TDA interprets data sets as topological spaces, allowing researchers to extract intrinsic structural features that persist across different scales and embeddings. At its core, TDA constructs topological representations (e.g., simplicial complexes) from data and extracts topological features for data analysis and characterization.

Within the TDA framework, a variety of techniques have been developed, including persistent homology,[41,42] persistent GLMY homology,[43,44] topological summaries like Euler characteristic curves,[44,45] and complementary tools such as the Mapper algorithm[46] and topological simplification methods.[47] Among these, persistent homology is the most widely adopted method in material science, enabling capture multiscale topological features from complex structures. GLMY homology, designed for directed or asymmetric data structures, offers a conceptual framework that may analogously capture certain directional characteristics observed in chemical bonding. Euler characteristic curves have also attracted increasing attention in material science due to their conceptual simplicity, intuitive interpretability, and low computational cost. Thus, we focus primarily on these three TDA techniques and systematically introduce their mathematical foundations.

2.1. Persistent Homology

Persistent homology, originally introduced by Edelsbrunner, Letscher, and Zomorodian and subsequently advanced by numerous researchers, provides a rigorous methodology for extracting and quantifying topological information from data.[37,41,42,48−50] Unlike traditional geometric descriptors that rely on predefined structural features, persistent homology characterizes the intrinsic topological organization of data across multiple length scales. In materials science, this multiscale capability is particularly valuable because many structural phenomena, such as atomic packing, pore connectivity, and network structures, naturally exhibit hierarchical organization. Application examples are presented throughout Section 4. In the following, we present a concise overview of the fundamental concepts underlying persistent homology.

2.1.1. Simplices and Simplicial Complexes. TDA employs simplices and simplicial complexes to represent geometric and topological structures in a mathematically tractable way. A k-simplex σ^k is a set of k + 1affinely independent points in $\mathbb{R}^n$, which can be represented by $\{v_0, v_1,\cdots,v_k\}$, where each v_i is a vertex of the simplex. A 0-simplex is a vertex, a 1-simplex is a line segment, a 2-simplex is a triangle, and a 3-simplex is a tetrahedron as shown in Figure 1. An m-face of simple σ^k is formed by selecting m+1 vertices of σ^k and taking their convex hull.

A simplicial complex K is a collection of simplices satisfying: (1) every face of a simplex in k is also in K; (2) the intersection of any two simplices is either empty or a common face; (3) The dimension of K is the highest dimension of its simplices. Geometrically, simplices provide the simplest building blocks

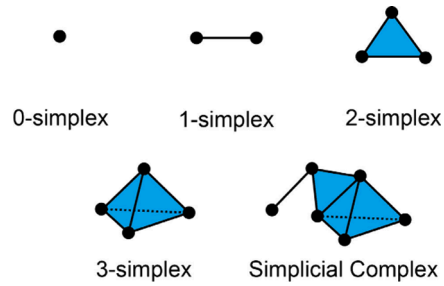


Figure 1. Examples of k-dimensional simplices for k = 0, 1, 2, 3. A simplicial complex must also contain all faces of the included simplices, i.e., their lower-dimensional simplices such as edges and vertices. 0-simplices are vertices (points), 1-simplices are edges, 2-simplices are triangles, and 3-simplices are tetrahedra.

for representing shapes of different dimensions. By assembling simplices of increasing dimension (vertices, edges, triangles, and tetrahedra) complex geometric structures can be approximated using purely combinatorial objects. This representation enables the application of algebraic topology to discrete data sets such as atomic coordinates or point clouds.

2.1.2. Homology. Homology provides an algebraic framework to quantify the number of multidimensional topological features in a simplicial complex.[51] For a simplicial complex K, a k-chain c_k is a formal sum of k-simplices in K, where k is less than or equal to the dimension of K. A k-chain is represented as $c_k = \sum \alpha_i\sigma^i$, with each σ^i denoting a k-simplex and each α_i being a coefficient from an algebraic field or ring. These coefficients α_i can generally be elements from a field such as $\mathbb{R}$, $\mathbb{Q}$, or $\mathbb{Z}_n$, but for simplicity, $\mathbb{Z}_2$ is often used. The set of all k-chains in K, denoted by C_k, forms an Abelian group $(C_k, \mathbb{Z}_2)$ under component-wise addition. This structure allows the boundary operator to be extended from simplices to chains. The boundary operator ∂_k maps k-chains to $(k$ - 1)-chains:

$$\partial_k[v_0, v_1, ..., v_k] = \sum_{i=0}^{k} (-1)^i[v_0, \cdots, \hat{v}_i, \cdots, v_k] \tag{1}$$

In this equation, $\hat{v}_i$ omits the i-th vertex. This definition extends linearly to all k-chains in the chain group $C_k(K)$, allowing us to build the following chain complex:

$$\cdots \xrightarrow{\partial_{n+1}} C_n(K) \xrightarrow{\partial_n} C_{n-1}(K) \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_1} C_0(K) \xrightarrow{\partial_0} 0 \tag{2}$$

For dimensions $k >$ dim (K) or $k < 0$, the chain groups are defined to be zero, and the corresponding boundary maps are also zero maps. The most fundamental topological property satisfied by the boundary operator is

$$\partial_{k-1}\circ\partial_k = 0 \tag{3}$$

In other words, the boundary of a boundary is always zero. This fundamental property enables the definition of homology groups by distinguishing cycles (chains with no boundary) from boundaries (chains that are the boundary of a higher-dimensional chain). Accordingly, we define the k-cycle group and k-boundary group as follows:

$$Z_k = \mathrm{Ker}(\partial_k),\ B_k = \mathrm{Im}(\partial_{k+1}) \tag{4}$$

The k-homology group H_k is then defined as the quotient group:

$$H_k = Z_k/B_k \tag{5}$$

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

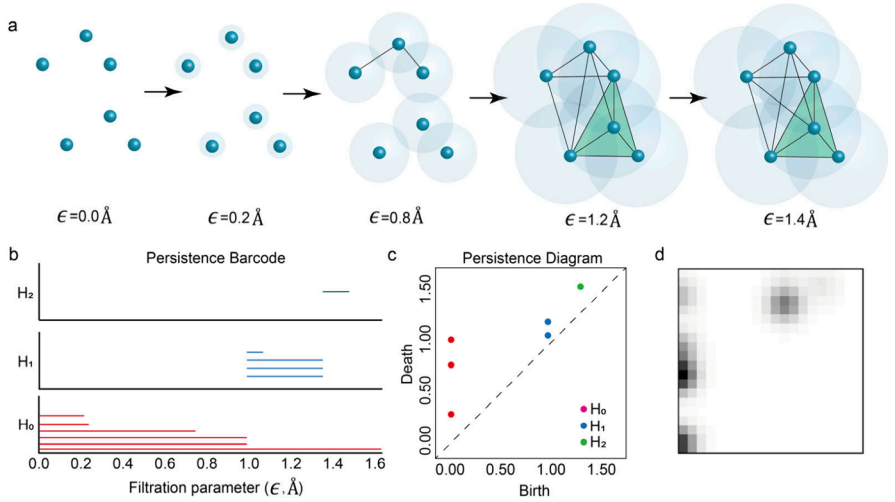


Figure 2. (a) Filtration of points $x_i \in X$ by a set of balls $B(x_i, \epsilon)$ with expanding ϵ. As ε is increased the topology evolves and this introduces holes and higher dimensional simplices. The (b) persistence barcode, (c) persistence diagram and (d) persistence image of the filtration process in (a) are presented.

The H_k measures k-dimensional holes in the simplicial complex.[52] H_0 homology groups correspond to connected components, H_1 homology groups correspond to loops, and H_2 homology groups correspond to higher-dimensional voids. The Betti number β_k, which is the rank of H_k, indicates the number of independent k-dimensional features. Specifically, β_0 describes the number of connected components, β_1 counts independent loops and β_2 measures independent voids.[52] Intuitively, homology identifies topological structures that correspond to "holes" of different dimensions. In the context of materials structures, these features often have direct physical interpretations. For example, β_0 corresponds to disconnected clusters, β_1 captures looplike structures such as ring networks in glasses or polymers, and β_2 describes enclosed voids or pores in porous materials.

2.1.3. Filtration and Persistent Homology. Persistent homology extends classical homology theory to a multiscale framework, enabling the capture of topological features that persist as a scale parameter (e.g.,) varies. Its introduction is motivated by the fact that computing the homology of an arbitrary subset of $\mathbb{R}^n$ is generally intractable, due to the continuous and potentially high-dimensional nature of such spaces. To overcome this difficulty, the space X is typically approximated by a discrete combinatorial object-most commonly a simplicial complex K-whose homological properties approximate those of X. This discretization allows the tools of simplicial homology to be applied directly to data.

The core mechanism enabling persistent homology is the use of a filtration, which defines a sequence of nested simplicial complexes:

$$\emptyset \subseteq K_0 \subseteq K_1 \ldots \subseteq K_n = K \tag{6}$$

where each K corresponds to a specific scale parameter ϵ_i. As the scale increases, more simplices are added to the complex, allowing the tracking of the appearance (birth) and disappearance (death) of topological features across scales.[16]

Intuitively, the filtration process can be understood as progressively "growing" the geometric structure by increasing the scale parameter. At small values of ϵ, only nearby points are connected, resulting in many disconnected components. As ϵ increases, additional simplices are added, leading to the formation of loops and higher-dimensional structures. Eventually, these features disappear when they become filled in by higher-dimensional simplices. This multiscale evolution allows persistent homology to distinguish between topological features that are stable across a wide range of scales and those that arise from noise or local fluctuations in the data.

This filtration induces a sequence of homology groups at each dimension k, and the persistent homology groups $H_k^{i,j}$ measure the topological features that are born at filtration step i and persist until step j. The corresponding persistent Betti number is defined as

$$\beta_k^{i,j} = rank(H_k^{i,j}) = rank(Z_k(K_i)/(B_k(K_j) \cap Z_k(K_i))) \tag{7}$$

where Z_k and B_k denote the cycle and boundary groups at dimension k, respectively.

Figure 2a illustrates a simple point cloud at varying thresholds ϵ and the corresponding persistent topological patterns, which is used to demonstrate the basic workflow of persistent homology.

2.1.4. Visualization of Persistent Homology. To facilitate interpretation and quantitative analysis, the results of persistent homology are typically represented using a range of complementary visualization and vectorization techniques, as outlined below, with the point cloud shown in Figure 2a serving as an illustrative example.

Persistence Barcode. The persistence barcode (Figure 2b) is a fundamental visualization tool in persistent homology that represents the lifetime of each topological feature across a filtration. Each bar corresponds to a homological feature and spans from its birth time (when it appears in the filtration) to its death time (when it is filled in or merges with other features). The barcode encodes the complete persistence information on the filtration, allowing direct visualization of topological features such as connected components, loops, and voids across different scales. Longer bars typically correspond to more persistent and topologically significant features, while shorter bars often arise from noise or minor fluctuations in the underlying data, such as transient perturbations in atomic coordinates. While persistence barcodes are highly informative for low-dimensional or small-scale data sets, they become

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

increasingly difficult to interpret as the number of features grows.

Persistence Diagram. A persistence diagram is a summary representation to visualize the lifespan of topological features, which is obtained by mapping each bar from the persistence barcode to a point in a two-dimensional Cartesian plane (Figure 2c).[48] Generally, each point in a persistence diagram corresponds to a homological feature, plotted as a coordinate pair (b, d), where b denotes the birth time and d the death time of the feature across a filtration. In this format, each point satisfies $d \geq b$ by definition and therefore lies on or above the diagonal line $b = d$. Alternatively, the diagram can be formulated using birth-lifespan pairs, $\{(\epsilon_i, \epsilon_k - \epsilon_i)$, where ϵ_i is the scale at which the feature appears, and $\epsilon_k - \epsilon_i$ is its persistence or lifespan. The persistence diagram provides a more compact and holistic view of the overall topological structure, particularly for large or high-dimensional data sets. Compared with persistence barcode, persistence diagram is generally more scalable and facilitate aggregate analysis and interpretation.

Distance between two persistence diagrams PD_1 and PD_2 can be calculated by the p^{th}-Wasserstein to compare their topological features. Each persistence diagram consists of a multiset of points (b_i, d_i), which represent the birth and death of homological features (e.g., connected components, loops, voids) across a filtration. The distance is defined as

$$d_{W_p}(PD_1, PD_2) = \left(\inf_{\gamma} \sum_{x \in PD_1} \|x - \gamma(x)\|_\infty^p \right)^{1/p} \tag{8}$$

Here, γ ranges over all possible bijections (one-to-one correspondences) from the points in PD_1 to those in PD_2, and $\|x - \gamma(x)\|_\infty$ denotes the maximum coordinate-wise difference (i.e., the L^∞ norm) between matched pairs x and $\gamma(x)$ This metric captures both geometric and topological differences. To accommodate persistence diagrams of different sizes, additional points are placed on the diagonal $b = d$, representing features that are born and die simultaneously (i.e., trivial features), enabling valid bijections between unequal sets. The Wasserstein distance thus provides a stable, interpretable, and scale-sensitive measure for comparing persistence diagrams.

Persistence Image. A persistence image is a vectorized representation of a persistence barcode constructed through kernel density estimation (Figure 2d).[53] It maps the birth-persistence coordinates of each topological feature into a smooth distribution over a fixed-size grid. Specifically, each point $(\epsilon_i, \epsilon_j - \epsilon_i)$ in the persistence barcode is denoted as (b, d), where $b = \epsilon_i$ is the birth time and $d = \epsilon_j - \epsilon_i$ is the persistence. Assuming the persistence image has resolution $\mathcal{M} \times \mathcal{N}$, the pixel value at position (m, n), where $m \in [1, \mathcal{M}]$ and $n \in [1, \mathcal{N}]$, is computed as

$$PI(m, n) = \sum_{k \in PD} \left[\Phi\left(\frac{x_m - b_k}{\sigma} \right) \Phi\left(\frac{y_n - d_k}{\sigma} \right) w(b_k, d_k) \right] \tag{9}$$

where σ is the standard deviation of the Gaussian kernel, $\Phi(\mathcal{Z})$ is the cumulative distribution function of the standard normal distribution, $\omega(b, d) = d/max(d)$ is a weighting function that emphasizes more persistent features. Finally, the resulting persistence image is flattened into a fixed-length vector, making it well-suited for input into machine learning models.

Persistence Entropy. Persistence entropy is a quantitative descriptor derived from persistence barcodes or persistence diagrams that measures the distribution and complexity of topological features across different filtration scales.[54,55] Given persistence intervals (b_i, d_i) with lifetimes $l_i = b_i - d_i$, persistence entropy is defined as $H = -\sum_{i=1}^{N} p_i \log p_i$ and $p_i = \frac{l_i}{\sum_{j=1}^{N} l_i}$, where p_i represents the normalized contribution of each topological feature.

Persistence Landscape. Persistence landscape provides a functional representation of persistence diagrams by mapping each persistence interval (b_i, d_i) to a piecewise-linear function $\lambda_i(t) = \max(\min(t - b_i, d_i - t), 0)$, and defining the persistence landscape as an ordered sequence of functions $\{\lambda_k(t)\}_{k=1}^{\infty}$, where $\lambda_k(t)$ denotes the k-th largest function value at each t. This functional formulation enables statistical analysis and seamless integration with machine learning methods.[56]

These representations provide a compact yet expressive summary of the multiscale topology of complex data sets, forming the basis for subsequent application of TDA in topological learning and materials informatics.

2.1.5. Practical Implementation of Persistent Homology. A critical component in the practical application of persistent homology is the construction of appropriate simplicial complexes that evolve with the scale parameter. Several methods have been developed to build such complexes from point cloud data, including the Vietoris−Rips (VR) complex, Čech complex, Alpha complex, among others.[57−59]

Among these, the VR complex is conceptually simple and computationally efficient. The following outlines the construction of the VR complex. Consider a point cloud $P = \{p_1, p_2, \ldots, p_N\}$, where each point $p_i = (x_i, y_i, z_i)$ lies in a Euclidean subspace of $\mathbb{R}^3$. The distance matrix D is defined by $D_{ij} = |p_i - p_j|_2$ representing the pairwise Euclidean distances between points in P. Given a non-negative distance threshold ϵ, the VR complex $VR(P, \epsilon)$ is constructed as an abstract simplicial complex in which a finite subset $\sigma = \{p_{i_0}, p_{i_1}, \cdots, p_{i_k}\} \subset P$ forms a k-simplex if the pairwise distances between all points in σ satisfy $\|p_i - p_l\| \leq \epsilon$ for all $j, l \in \{0, 1, \ldots, k\}$. The construction proceeds as follows: each point in P is treated as a 0-simplex (vertex), a pair of points (p_i, p_j) with $D_{ij} \leq \epsilon$ is connected by a 1-simplex (edge) and a triple $\{p_i, p_j, p_k\}$ with all pairwise distances $D_{ij}, D_{ik}, D_{jk} \leq \epsilon$ forms a 2-simplex (triangle), and so on. This process naturally generalizes to higher dimensions. The resulting simplicial complex includes vertices, edges, triangles, and higher-order simplices depending on the threshold ϵ. Varying ϵ yields different filtrations of the complex, enabling the analysis of geometric and topological structures of the data across multiple scales.

The Čech complex offers a more geometrically precise construction. Given a set of balls of radius $\epsilon/2$ centered at each data point, a k-simplex is included in the complex if and only if the corresponding $(k + 1)$ balls have a nonempty common intersection. This construction is closely tied to the Nerve Theorem,[60] and under suitable conditions, it ensures topological correctness, meaning that the resulting complex preserves the homotopy type of the underlying space.[61] The Nerve Theorem states that if a topological space is covered by a collection of sets whose finite intersections are either empty or contractible, then the nerve of the cover has the same

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

homotopy type as the original space. Here, homotopy refers to a continuous deformation between maps, and it provides the basis for defining when two spaces are equivalent up to such deformations.[51] In contrast to the VR complex, which only requires all pairwise distances between points in a simplex to be less than ϵ, the Čech complex imposes a stronger condition: a full intersection of all involved metric balls. Consequently, Čech complexes tend to provide a more faithful geometric approximation, but are also computationally more expensive, especially in higher dimensions due to the complexity of checking multiway intersections. In materials science applications, the points in the cloud typically correspond to atomic coordinates, and the filtration parameter ϵ can be interpreted as a distance threshold that gradually reveals structural connectivity at different length scales.

The Alpha complex provides a practical alternative by forming a subcomplex of the Delaunay triangulation. Delaunay triangulation is a geometric construction that partitions a set of points into simplices such that no point lies inside the circumcircle (or circumsphere in higher dimensions) of any simplex.[62] It captures both proximity and local geometric structure, offering a computationally efficient and geometrically meaningful approximation of the Čech complex. Due to its sparsity and dimensional constraints, the Alpha complex is particularly well-suited for low-dimensional data embedded in Euclidean space and often results in more compact and interpretable simplicial representations.

To illustrate how simplicial complexes are constructed and how topological features evolve along a filtration, we consider a simple example comparing the VR complex and the alpha complex constructed from the same point configuration. Both constructions generate filtrations parametrized by a distance scale, but they differ in how simplices are introduced. In the VR complex, simplices are added whenever all pairwise distances among a set of points are below the filtration threshold, making the construction purely distance-based and computationally convenient. In contrast, the alpha complex is derived from the Delaunay triangulation and includes simplices that correspond to intersections of balls centered at the data points, providing a geometric approximation of the underlying union of balls.

Figure 3 illustrates the evolution of topological features for a simple cubic arrangement of points under these two filtrations. As the filtration parameter increases, connected components merge, loops appear, and higher-dimensional voids may form before eventually being filled. The resulting birth and death of these features are summarized by the corresponding persistence barcodes. This example highlights how different filtrations may capture similar large-scale topological structures while differing in the precise scales at which features appear and disappear.

In scientific and engineering applications, researchers have developed several mature software libraries for conducting persistent homology analysis on data points. Among these, Ripser and GUDHI are the most widely utilized in material sciences.[63,64] Additional libraries, such as giotto-tda, Dionysus, JAVAPLEX, Topology Toolkit and PERSEUS, also provide robust tools for TDA in various contexts.[65−70]

From a theoretical perspective, the computational complexity of persistent homology is governed by the combinatorial growth of simplices in the constructed filtration. For VR filtrations, the number of k-dimensional simplices scales as $O(n^{k+1})$, while the subsequent boundary matrix reduction has

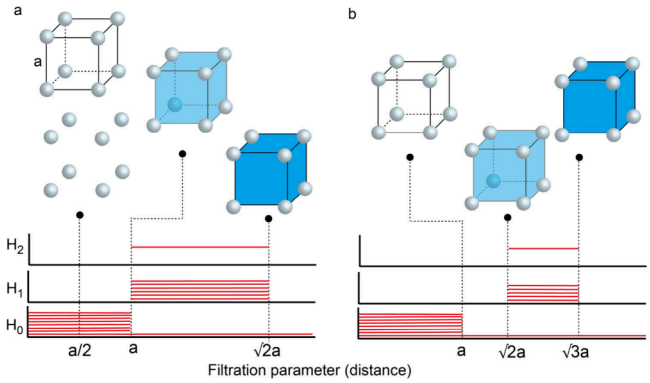


Figure 3. Comparison of VR and alpha complex filtrations. (a) VR filtration constructed from a simple cubic arrangement of points with lattice parameter. As the filtration parameter (distance threshold) increases, simplices are added whenever all pairwise distances among a set of points fall below the threshold. This purely distance-based rule leads to the gradual formation of higher-dimensional simplices. The corresponding evolution of topological features is summarized in the persistence barcode shown below, where H_0, H_1, and H_2 denote connected components, loops, and voids, respectively. (b) Alpha complex filtration constructed from the same point configuration. In contrast to the VR complex, the alpha complex is derived from the Delaunay triangulation and includes simplices corresponding to intersections of balls centered at the data points. This construction provides a geometrically faithful representation of the union of balls and typically yields a more compact simplicial complex. The resulting persistence barcode illustrates the birth and death of topological features along the filtration parameter.

cubic complexity in the number of simplices, yielding an overall theoretical complexity of $O(n^{3(k+1)})$. To provide a more intuitive understanding of the computational cost for material systems, we constructed a sodium chloride supercell, a very simple material, as an example, using VR complexes with the Ripser library. The computation was performed with maxdim = 2, meaning that topological features were evaluated up to two-dimensional homology. Timing benchmarks were performed on a node equipped with Intel Xeon Gold 5218 processor (2.30 GHz, 64 logical CPUs). The computational cost increases with system size, with runtimes of approximately 0.01, 0.08, 34.05, and 529.64 s for systems containing 10, 100, 500, and 1000 atoms, respectively. In typical materials research, a system containing 1000 atoms is already regarded as relatively large. Consequently, existing persistent homology packages are adequate for a wide range of material research tasks. Nevertheless, with the continuous advancement of computational power and machine learning technologies, the systems being studied are growing ever larger. In such cases, strategies to mitigate computational complexity are required. A detailed discussion of these strategies is provided in the Discussion and Outlook section.

2.2. GLMY Theory

The homology theory and persistent homology of simplicial complexes have demonstrated its strong power to extract structural information from data. However, in many real-world systems, particularly in chemical contexts, the interactions between units often exhibit directionality, such as the directionality and polarity of chemical bonds and reaction pathways. This directionality plays a crucial role in determining both the intrinsic structure and the functional behavior of materials.[71] In such cases, the directed graphs (digraphs)

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

provide a more precise model as they naturally are. A graph is a mathematical structure consisting of a set of vertices (nodes) and a set of edges that connect pairs of vertices.[72] Graph is widely used to represent pairwise relationships in complex systems.[73] If the edges in a graph have no orientation, it is referred to as an undirected graph; if they possess direction, it is referred to as a digraph.

Path homology theory, originally introduced by Grigor'yan, Lin, Muranov, and Yau in 2012 and now known as GLMY homology, was specifically designed for digraphs.[74] Moreover, Chowdhury and Mémoli introduced persistent path homology, proved its stability under small perturbations, and provided an algorithm for its computation.[75] On the other hand, theoretically, path complexes overcome the limitation in simplicial complexes that every face must be a simplex, thereby extending the classical homology theory and making it applicable to broader and more general practical problems. Driven by these facts, the digraph topology theory,[76−83] now commonly referred to as GLMY theory, has emerged as a unified framework. It encompasses the embedded homology theories of hypergraphs, superhypergraphs, and hyperdigraphs as generalizations of GLMY homology. Readers can refer to Section 4.1 to see the advantages of GLMY homology in distinguishing structural differences in materials, and to Section 4.4 (e.g., Figure 18) for how these advantages are further integrated with machine learning models to capture structure−property relationships.

2.2.1. GLMY Homology. Throughout this work, we assume that digraphs contain neither self-loops nor multiple arrows between any pair of vertices. Let V be a nonempty finite vertex set, typically written as $\{v_0,v_1,\cdots,v_m\}$. For an integer $n \geq 0$, an elementary n-path on V is a sequence $v_0\ v_1\cdots v_n$ and the collection of all elementary n-path is denoted by $\mathcal{P}_n(V)$. Examples are illustrated in Figure 4a. Elementary paths depend only on the vertex set and are independent of the arrow structure of a digraph. In particular, any pair of vertices forms an elementary 1-path even if no arrow exists between them (blue arrows in Figure 4b). Let $\mathbb{K}$ be a field, typically $\mathbb{Z}_2$. Define $\Lambda_n(V)$ to be the $\mathbb{K}$-linear space $\mathbb{K}\langle e_{v_0v_1\cdots v_n} | v_0v_1\cdots v_n \in \mathcal{P}_n(V)\rangle$ generated by all elementary n-paths. Here, $e_{v_0\ v_1\cdots v_n}$ denotes the generator of $\Lambda_n(V)$ corresponding to the elementary $v_0\ v_1\cdots v_n$. The family $\{e_{v_0v_1\cdots v_n} | v_0v_1\cdots v_n \in \mathcal{P}_n(V)\}$ forms a basis of $\Lambda_n(V)$. Consequently, any element $x \in \Lambda_n(V)$ admits a unique representation:

$$x = \sum_{v_0v_1\cdots v_n \in_n(V)} a_{v_0v_1\cdots v_n} e_{v_0v_1\cdots v_n},\ a_{v_0v_1\cdots v_n} \in \mathbb{K} \tag{10}$$

Define the boundary map:

$$\partial_n\colon \Lambda_n(V) \to \Lambda_{n-1}(V),\ e_{v_0v_1\cdots v_n} = \sum_{i=0}^{n} (-1)^i e_{v_0v_1\cdots \hat{v}_i\cdots v_n} \tag{11}$$

where $e_{v_0\ v_1\cdots\hat{v}_i\cdots v_n}$ denotes the generator of $\Lambda_{n-1}(V)$ corresponding to the elementary $(n-1)$-path obtained by deleting the i-th vertex from $v_0\ v_1\cdots v_n$. By convention, $\Lambda_{n-1}(V) = 0$, so ∂_0 is the zero map. Further, it is easy to compute that $\partial_{n-1}\ \partial_n = 0$. Thus, there is a chain complex $\Lambda_*(G)$

$$\Lambda_*(G)\colon \cdots \xrightarrow{\partial_{n+1}} \Lambda_n(V) \xrightarrow{\partial_n} \Lambda_{n-1}(V) \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_1} \Lambda_0(V) \xrightarrow{\partial_0} 0. \tag{12}$$

However, so far, we have considered only the vertex information, without including any arrow information on digraphs, as Figure 4b said. Let $G = (V_G,A_G)$ be a digraph with finite vertex set V_G and arrow set A_G. Taking the arrow information into account, we can define the allowed n-paths on G. An allowed n-path is a vertex sequence $v_0\ v_1\cdots v_n$ such that every adjacent pair satisfies $v_iv_{i+1} \in A_G$, i.e., v_iv_{i+1} is an arrow in G for $0 \leq i \leq n-1$. Figure 4c illustrates the allowed n-paths. The set of all allowed n-paths is denoted by $\mathcal{A}_n(G)$. Similarly, the $\mathbb{K}$-linear space $\mathbb{K}\langle e_{v_0v_1\cdots v_n} | v_0v_1\cdots v_n \in \mathcal{A}_n(G)\rangle$ generated by all allowed n-paths on G was defined, denoted by $A_n(G)$. Obviously, $A_n(G)$ is a $\mathbb{K}$-linear subspace of $\Lambda_n(V_G)$. Nevertheless, there maybe some element $x \in A_n(G)$ such that $\partial_n x \notin A_{n-1}(G)$, as shown in Figure 4d, which implies that the boundary map $\partial_n{:}A_nG \to A_{n-1}(G)$ is not closed.

To solve this problem, the ∂-invariant chain complex $\Omega_*(G)$

$$\Omega_*(G)\colon \cdots \xrightarrow{\partial_{n+1}} \Omega_n(G) \xrightarrow{\partial_n} \Omega_{n-1}(G) \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_1} \Omega_0(G) \xrightarrow{\partial_0} 0 \tag{13}$$

was introduced. Here, $\Omega_n(G) = A_n(G) \cap \partial_n^{-1}(A_{n-1}(G)) = \{x \in A_n(G) \mid \partial_n(x) \in A_{n-1}(G)\}$, $n \geq 0$ is the maximal $\mathbb{K}$-linear subspace of $A_n(G)$ such that ∂_n can be defined and $\Omega_{-1}(G) = 0$. For example (Figure 4e), although $\partial_2(e_{013}) \notin A_1(G)$ and $\partial_2(e_{023}) \notin A_1(G)$, the boundary $\partial_2(e_{013} - e_{023})$ of the linear combination $e_{013} - e_{023}$ does lie in $A_1(G_1)$. Therefore, $e_{013} - e_{023}$ generates $\Omega_2(G_1)$. For simplicity of practical applications, we record the following two results: $\Omega_1(G) = A_1(G)$ and $\Omega_0(G) = A_0(G)$.

Define the cycle and boundary groups as

$$Z_n(G) = \mathrm{Ker}\partial_n,\ B_n(G) = \mathrm{Im}\ \partial_{n+1} \tag{14}$$

Then the n-th GLMY homology group is

$$H_n(G) = \frac{Z_n(G)}{B_n(G)},\ n \geq 0. \tag{15}$$

Using the facts that $\Omega_1(G) = A_1(G)$ and $\Omega_0(G) = A_0(G)$, we obtain that $Z_1(G_1)$ is the $\mathbb{K}$-linear space

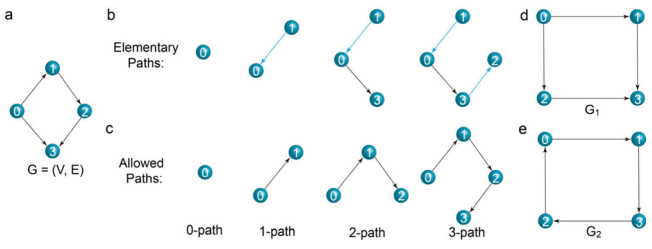


Figure 4. Elementary n-paths on vertex set V, and allowed n-paths on digraph G for $n = 0, 1, 2, 3$. (a) Digraph G. (b) Elementary 0-paths on V are vertices in V, elementary 1-paths are vertex sequence with two vertices in V, elementary 2-paths are vertex sequence with three vertices in V, and so on. (c) Allowed 0-paths are vertices in G, allowed 1-paths are arrows in G, allowed 2-paths are the 2-walks that follow the arrows in G. (d) For the allowed 2-path e_{013} on G_1, its boundary $\partial_2(e_{013}) = e_{13} - e_{03} + e_{01}$. However, the 1-path 03 is not allowed in G_1. Thus, $e_{03} \notin A_1(G_1)$ and $\partial_2(e_{013}) \notin A_1(G_1)$. The 2-path 013 is an allowed 2-path in G_1, but its subpath 03 is not an allowed 1-path in G_1. (e) $\Omega_2(G_1) = \mathbb{K}\langle e_{013} - e_{023}\rangle$ is the $\mathbb{K}$-linear space generated by the element $e_{013} - e_{023}$, and $\Omega_2(G_2) = 0$. $H_1(G_1) = 0$ and $H_1(G_2) = \mathbb{K}\langle e_{01} + e_{13} + e_{32} + e_{20}\rangle$ is the $\mathbb{K}$-linear space with a generator $e_{01} + e_{13} + e_{32} + e_{20}$.

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

$\mathbb{K}\langle e_{01} + e_{13} - e_{02} - e_{23}\rangle$ and $Z_1(G_2)$ is the $\mathbb{K}$-linear space $\mathbb{K}\langle e_{01} + e_{13} + e_{32} + e_{20}\rangle$ with the generator $e_{01} + e_{13} + e_{32} + e_{20}$, which represents the directed cycle in G_2. The 2-dimensional ∂-invariant chain groups of G_1 and G_2 are then computed and shown in Figures 4d and 4e, respectively. This demonstrates that the GLMY homology indeed distinguishes the arrow information on digraphs.

From a mathematical viewpoint, GLMY homology generalizes simplicial homology via path complexes. A path complex $\{P_n\}_{n\geq 0}$ on V satisfies $v_0v_1v_2\cdots v_n \in P_n$, then both $v_1v_2\cdots v_n \in P_{n-1}$ and $v_0v_1\cdots v_{n-1} \in P_{n-1}$. Obviously, the family $\{\mathcal{A}_n(G)\}_{n\geq 0}$ on G forms a path complex, called allowed path complex. Figure 4d shows that path complex overcomes the limitation of simplicial complex: every face of a simplex must belong to this simplicial complex.

Inspired by the idea of persistent homology theory, several groups have proposed algorithms to compute persistent GLMY homology in Python.[84] Similar to persistent homology for simplicial complexes, the procedure for persistent GLMY homology is as follows: we first construct a filtration $f\ P\to K$ from path complex $\mathcal{P}$ to $\mathbb{K}$ which induces an increasing sequence of path complexes parametrized by a $\in \mathbb{K}$. Then, we compute the GLMY homology groups along this increasing sequence of path complexes to obtain the persistent Betti numbers (Figure 5a). Subsequently, n-th persistent GLMY Betti numbers can be visualized by persistence barcodes (Figure 5b), persistence diagrams (Figure 5c), and so on.

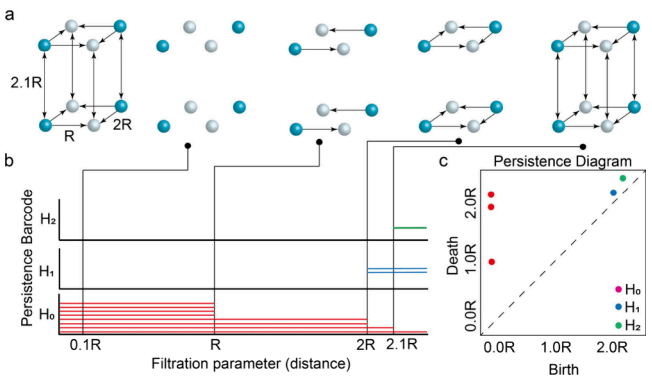


Figure 5. Persistent GLMY homology for digraph G. (a) A distance function w: $A_G \to \mathbb{K}$ is defined by sending each arrow to the distance between its end points. A sequence of increasing subdigraphs $G_{0.1R} \subseteq G_R \subseteq G_{2R} \subseteq G_{2.1R}$ are shown in the top row. (b) Persistence barcodes of persistent GLMY homology of G along the distance. (c) persistence diagrams of the filtration process.

A filtration $f: \mathcal{P} \to \mathbb{K}$ is a function from path complex $\mathcal{P}$ to $\mathbb{K}$ such that $f(\sigma) \leq f(\tau)$ for any path τ and its subpath σ. For $a \in K$, define $f_a = f^{-1}(-\infty, a)$, then this yields an increasing sequence of path complexes

$$\emptyset \subseteq f_{a_1} \subseteq f_{a_2} \subseteq \cdots \subseteq f_{a_m} = \tag{16}$$

Moreover, this increasing sequence induces a sequence of homology groups, where the arrows denote inclusion maps

$$0 = H_n(f_{a_1}) \to H_n(f_{a_2}) \to \cdots \to H_n(f_{a_m}) = H_n() \tag{17}$$

Finally, we define the n-th persistent GLMY homology $H_n^{i,j}(\mathcal{P})$ of this filtration born at step i and persist until step j to be $\mathrm{Im}(H_n(f_{a_i}) \to H_n(f_{a_j}))$, where the arrow denotes the inclusion map, and the corresponding n-th persistent Betti number is defined as

$$\beta_n^{i,j} = \dim H_n^{i,j}() \tag{18}$$

A frequent example is a finite digraph equipped with a distance (threshold) function w: $A_G \to \mathbb{K}$, as Figure 5a and 5b shown us. Denote by G_a the subdigraph of G consisting of those arrows whose distance are less than a, then the increasing sequence $\{G_a\}$ of digraphs gives a growing family of path complexes

$$\emptyset = \{\mathcal{A}_n(G_{a_1})\}_{n\geq 0} \subseteq \{\mathcal{A}_n(G_{a_2})\}_{n\geq 0} \subseteq \cdots \subseteq \{\mathcal{A}_n(G_{a_m})\}_{n\geq 0} = \{\mathcal{A}_n(G)\}_{n\geq 0} \tag{19}$$

and hence an increasing sequence of ∂-invariant chain complexes

$$0 = \Omega_n(G_{a_1}) < \Omega_n(G_{a_2}) < \cdots < \Omega_n(G_{a_m}) = \Omega_n(G),\ n \geq 0. \tag{20}$$

Finally, the n-th persistent Betti number $\beta_n^{i,j}$ from step i to step j is given by $\dim\mathrm{Im}(H_n(G_{a_i}) \to H_n(G_{a_j}))$.

2.2.2. Generalized GLMY Homology. Beyond simple graphs, higher-order interactions can be represented using hypergraphs, where edges, called hyperedges, can connect more than two vertices simultaneously. Extensions such as superhypergraphs and hyperdigraphs further generalize this framework by allowing more complex and directed higher-order relationships. In this framework, a superhypergraph is viewed as a generalization of a hypergraph that allows multiple hyperedges associated with the same set of vertices.[80,81,85,86] Inspired by the idea of GLMY homology theory of digraphs, more general homology theories of hypergraphs, superhypergraphs and hyperdigraphs have been developed to enable broader applications.[87,88]

As a generalization of graphs and path complexes, hypergraphs can naturally capture group interactions, not just pairwise interactions, in complex system without any constraints.[89,90] A (finite) hypergraph is a pair $\mathcal{H} = (V; E)$ of sets, where V denotes the (finite) vertex set, often written as $\{v_0,v_1,\cdots,v_m\}$, and E is a collection of subsets of V. Each element of E is called a hyperedge and an n-hyperedge can be represented as $\{v_{i_0}\ v_{i_1}\cdots v_{i_n}\}$. Figure 6 illustrates a hypergraph and its hyperedges.

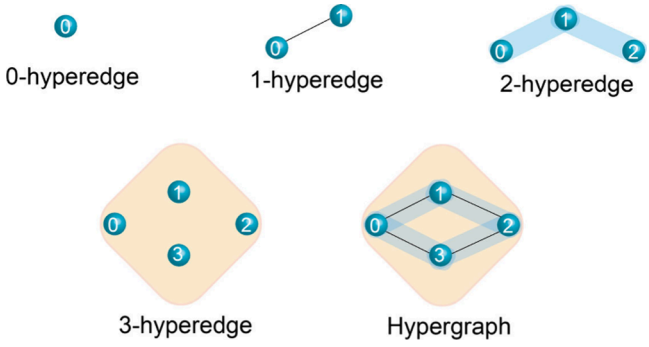


Figure 6. Hypergraph $\mathcal{H}$ with E = {{0},{1},{2},{3},{01},{12},{03},{32},{012}, {032},{0123}}. 0-hyperedges are vertices, 1-hyperedges are edges, 2-hyperedges are cliques with 3 vertices and 3-hyperedges are cliques with 4 vertices.

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

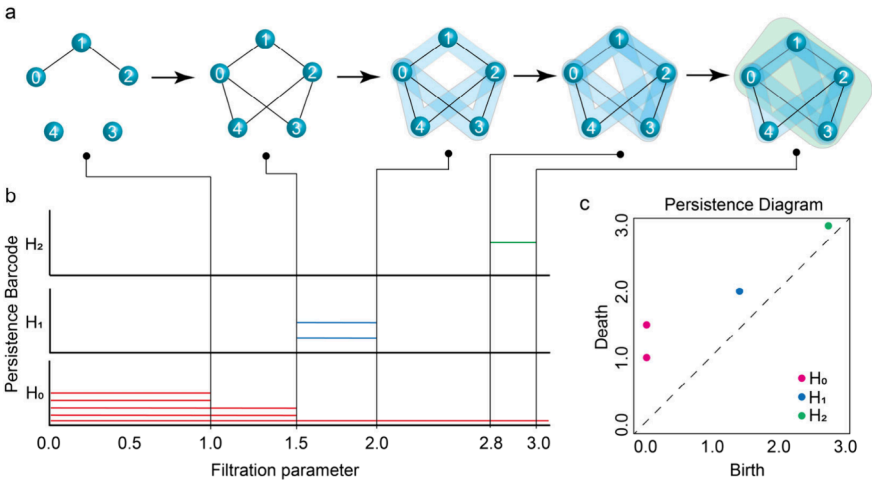


Figure 7. Persistent hypergraph homology for digraph Hypergraph $\mathcal{H}$. (a) Hypergraph $\mathcal{H}$ with E = {{0},{1},{2},{3},{4},{01},{12},{03},{32},{04},{42},{012},{032},{042},{132},{013},{0132}}. (b) Persistence barcodes of persistent hypergraph homology of $\mathcal{H}$ along the filtration parameter. (c) persistence diagrams of the filtration process.

Let $\mathbb{K}$ be a field. Denote by $P_n(\mathcal{H})$ the set of all n-hyperedges of $\mathcal{H}$ and by $C_n(\mathcal{H})$ the $\mathbb{K}$-linear space

$$\mathbb{K}\langle e_{v_0v_1\cdots v_n} | v_0v_1\cdots v_n \in P_n(\mathcal{H})\rangle = \{ae_{v_0v_1\cdots v_n} | v_0v_1\cdots v_n \in P_n(\mathcal{H}),\ a \in \mathbb{K}\} \tag{21}$$

generated by the basis $P_n(\mathcal{H})$, for integers $n \geq 0$. By convention, $C_{-1}(\mathcal{H})=0$ and ∂_0 is the zero map. So similarly to the allowed path complex of digraphs, the $\mathbb{K}$-linear boundary map

$$\partial_n : C_n(\mathcal{H}) \to C_{n-1}(\mathcal{H}),\ e_{v_0v_1\cdots v_n} \to \sum_{i=0}^{n} (-1)^i e_{v_0v_1\cdots \hat{v}_i\cdots v_n} \tag{22}$$

may not be closed.

Based on this fact and inspired by the ∂-invariant chain complex of digraphs, Bressan et al. introduced the maximal chain complex $\mathrm{Inf}_*(\mathcal{H})$ contained in $C_*(\mathcal{H})$ such that the boundary maps are well-defined

$$\mathrm{Inf}_*(\mathcal{H}): \cdots \xrightarrow{\partial_{n+1}} \mathrm{Inf}_n(\mathcal{H}) \xrightarrow{\partial_n} \mathrm{Inf}_{n-1}(\mathcal{H}) \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_1} \mathrm{Inf}_0(\mathcal{H}) \xrightarrow{\partial_0} 0 \tag{23}$$

and minimal chain complex $\mathcal{S}up_*(H)$ including the $C_*(\mathcal{H})$ such that the boundary maps are well-defined

$$\mathrm{Sup}_*(\mathcal{H}): \cdots \xrightarrow{\partial_{n+1}} \mathrm{Sup}_n(\mathcal{H}) \xrightarrow{\partial_n} \mathrm{Sup}_{n-1}(\mathcal{H}) \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_1} \mathrm{Sup}_0(\mathcal{H}) \tag{24}$$

where $\mathrm{Inf}_n(\mathcal{H}) = C_n(\mathcal{H}) \cap \partial_n^{-1}(C_{n-1}(\mathcal{H})) = \{x \in C_n(\mathcal{H}) | \partial_n(x) \in C_{n-1}(\mathcal{H})\}$ and $\mathrm{Sup}_n(\mathcal{H}) = C_n(\mathcal{H}) + \partial_{n+1}(C_{n+1}(\mathcal{H}))$ for $n \geq 0$. Following these, we have two n-th homology groups for hypergraph $\mathcal{H}$, denoted by $H_n(\mathrm{Inf}_*(\mathcal{H})) = \frac{\mathrm{Ker}\partial_n|_{\mathrm{Inf}_n(\mathcal{H})}}{\mathrm{Im}\,\partial_{n+1}|_{\mathrm{Inf}_{n+1}(\mathcal{H})}}$ and $H_n(\mathrm{Sup}_*(\mathcal{H})) = \frac{\mathrm{Ker}\partial_n|_{\mathrm{Sup}_n(\mathcal{H})}}{\mathrm{Im}\,\partial_{n+1}|_{\mathrm{Sup}_{n+1}(\mathcal{H})}}$ respectively. Fortunately, $H_n(\mathrm{Inf}_*(\mathcal{H})) \cong H_n(\mathrm{Sup}_*(\mathcal{H}))$ for integers $n \geq 0$. This result motivates the definition of the n-th embedded homology group of $\mathcal{H}$ as

$$H_n(\mathcal{H}) = H_n(\mathrm{Inf}_*(\mathcal{H})) \cong H_n(\mathrm{Sup}_*(\mathcal{H})),\ n \geq 0. \tag{25}$$

Due to this isomorphism, we compute the embedded homology of $\mathcal{H}$ by using of the chain complex $\mathrm{Inf}_*(\mathcal{H})$ instead of $\mathrm{Sup}_*(\mathcal{H})$ for simplifying the calculation (Figure 7a).

Subsequently, n-th persistent hypergraph Betti numbers can be visualized by persistence barcodes (Figure 7b), persistence diagrams (Figure 7c), and so on. Analogous to the homology theory of quivers, the embedded homology theory of superhypergraphs was introduced by Grbić et al., where the superhypergraph is the generalization of a hypergraph that allows multiple hyperedges among vertices.[82] By adding directions to hypergraphs, Gallo et al. introduced directed hypergraphs.[91] Later, Chen et al. developed the more general concept of hyperdigraphs, offering a more flexible model for practical applications and creating their embedded homology theory, generalizing both hypergraph homology and GLMY homology.[44] Beyond theoretical developments, GLMY-based topology has also been applied to complex systems involving higher-order interactions, including molecular, biological, and materials-related systems.[92] These studies demonstrate the potential of GLMY homology to capture directional and hypergraph structures that are difficult to represent using classical simplicial complexes.[74]

2.3. Euler Characteristic

The Euler characteristic is an integer that summarizes the shape of a topological space, which is defined by continuity and connectedness rather than distance or metrics. As a topological invariant, the Euler characteristic remains unchanged under deformations like stretching, rotation, or bending. Readers can refer to Sections 4.2 (e.g., Figure 14) and 5.1 for applications of the Euler characteristic in materials and molecular systems.

2.3.1. The Basis of Euler Characteristic. Here, we introduced the Euler characteristic in the context of data objects modeled as graphs and manifolds.[93,94] Graphs and manifolds are fundamental mathematical structures used to represent data. A graph, denoted as $G(V,E)$, is a two-dimensional topological space (i.e., a two-dimensional

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

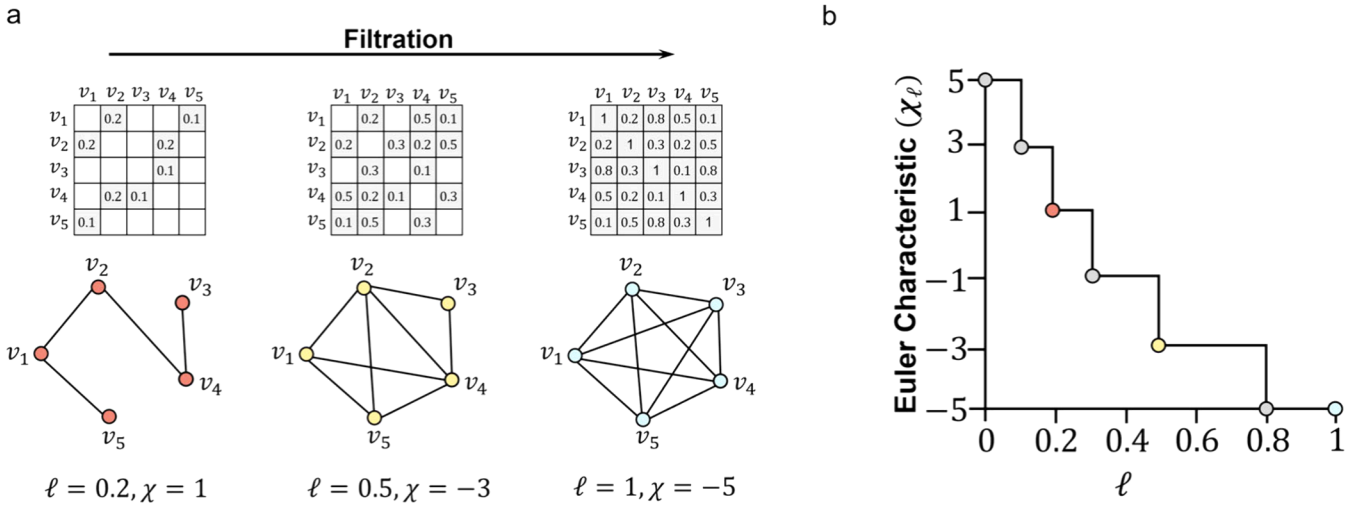


Figure 8. (a) Filtration of a correlation matrix, which is represented as an edge-weighted graph. The random variables are treated as nodes, and the correlation between the variables are treated as edge weights $w_E(e)$. As the level set of the filtration function f_E increases in value, edges that are below the given threshold are added into the resulting graph f_E^{-1}, revealing the topology of the correlation matrix over multiple correlation thresholds. (b) Resulting EC curve for the filtration of the edge-weighted graph in (a). Adapted with permission from ref 45. Copyright 2021 Elsevier.

polyhedron) whose topology is fully determined by the connectivity between its set of nodes V and edges E. In contrast, a manifold is a topological space that is locally homeomorphic to Euclidean space. Each local region can be described by a chart (X,f), where $X \subseteq \mathbb{R}^n$ is the domain and $f: X \to \mathbb{R}^n$ is a homeomorphism (a bijective, continuous function with a continuous inverse). This chart not only encodes the geometric structure of the manifold but also defines a field, which can be embedded into an $(n + 1)-$dimensional space to form a topological image.

In the two-dimensional case, the chart (X,f) can be interpreted as a continuous analogue of a node-weighted graph $G(V,E,w_V)$, where X represents the spatial positions of the nodes and f corresponds to the node weights. Such structures are commonly encountered in images, matrices, and numerical solutions of discretized partial differential equations, and collectively form what is known as a field graph.

For a general $(n + 1)$-dimensional topological space, the Euler characteristic is defined as the alternating sum of the Betti numbers across dimensions:

$$\chi = \sum_{i=0}^{n} (-1)^i \beta_i \tag{26}$$

where $\beta_i \in \mathbb{Z}_+$ denotes the i-th Betti number, representing the number of independent topological bases in dimension i (e.g., connected components, loops, voids). $\mathbb{Z}_+$ denotes the set of all nonnegative integers. This formulation provides a general definition of the Euler characteristic, applicable to a wide range of topological spaces, from simple surfaces to high-dimensional manifolds.

2.3.2. Filtration And Euler Characteristic Curve. Taking a graph as an example, how to compute the Euler characteristic is illustrated as below. For an undirected graph, the Euler characteristic is defined as follows:

$$\chi = |V| - |E| = \beta_0 - \beta_1 \tag{27}$$

where $|V| \in \mathbb{Z}_+$ denotes the number of vertices (nodes) and $|E| \in \mathbb{Z}_+$ denotes the number of edges. For each edge $e \in E$, we define $v(e),v'(e) \in V$ as its end point nodes. It can be shown that $|V| - |E|$ equals the number of connected components β_0 minus the number of cycles (or holes) β_1 in the graph.

The Euler characteristic is more often used to characterize edge-weighted graphs $G(V,E,w_E)$, where each edge $e \in E$ is associated with a scalar weight $w_E(e) \in \mathbb{R}$. Similarly, node-weighted graphs $G(V,E,w_V)$ can be characterized, where each node $v \in V$ is assigned a scalar weight $w_V(v) \in \mathbb{R}$. This characterization is also achieved through the filtration process.

The core idea of this process is to remove edges e from the graph whose weights are less than or equal to a given threshold $l \in \mathbb{R}$. The resulting graph is sparser than the original and is associated with a corresponding Euler characteristic value. By progressively increasing the threshold l, this filtration process can be repeated to generate a sequence of graphs, each with an associated Euler characteristic value. The process continues until the threshold reaches the maximum edge weight in the original graph, at which point the graph is fully recovered and its complete Euler characteristic value is computed.

To formalize this filtration process, a filtration function is defined as follows. For an undirected edge-weighted graph $G := G(V, E, w_E) \in \mathbb{G}$, where each edge $e \in E$ is associated with a scalar weight $w_E(e) \in \mathbb{R}$, we define the filtration function $f_E: \mathbb{G} \to \mathbb{R}$ such that

$$G_l = f_E^{-1}([-\infty, l]) \tag{28}$$

The preimage of the filtration function, denoted by $f_E^{-1}(l)$ for $l \in \mathbb{R}$, is defined as the graph $G_l := G(V, E_l, w_{E_l})$, where

$$E_l = \{e \in E | w_E(e) \le l\} \tag{29}$$

That is, the graph G_l consists of all edges from the original graph whose weights are less than or equal to the threshold l, resulting in a sparser subgraph.

The preimage of the filtration function is used to construct a sequence of nested subgraphs, defined via an increasing sequence of thresholds $l_1 < l_2 < \cdots < l_m$, yielding:

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

$$G_{l_1} \subseteq G_{l_2} \subseteq \cdots \subseteq G_{l_m} \subseteq G \tag{30}$$

If $l_m = \max_{e \in E} w_E(e)$, then the final graph in the sequence G_{l_m} recovers the original graph G. It can be seen that the graph density (i.e., the number of edges) increases with the threshold. Specifically, G_{l_1} is the sparsest graph in the sequence, while G_{l_m} is the densest. The filtration for node-weighted graphs and more general manifold objects can be defined in a similar manner. The related details are omitted here, and interested readers are referred to prior literature for further information.[45]

For each graph G_l in the node or edge filtration process, we compute and record its Euler characteristic value χ_l. This information is used to construct the Euler characteristic curve, which contains the pairs (l, χ_l). A simple example of graph edge filtration process is presented in Figure 8a and 8b. The Euler characteristic curve thus provides a topological descriptor for edge weighted graphs.

2.4. Comparison of Topological Descriptors

The methods discussed above provide complementary perspectives for characterizing topological structures in chemical and materials systems. Persistent homology offers a rigorous multiscale framework for identifying and tracking topological features such as connected components, loops, and voids across a filtration. Because it explicitly records the birth and death of these features, persistent homology provides a detailed description of structural organization and has therefore become one of the most widely used topological tools for analyzing molecular configurations, porous materials, and atomic structures.

GLMY homology extends classical homological constructions to more general complexes that can incorporate asymmetric relations or higher-order interactions. This flexibility enables the analysis of systems where interactions are not purely pairwise or where directionality plays an important role, such as reaction networks or generalized interaction complexes. However, due to its more specialized mathematical formulation, its applications remain less widespread compared with persistent homology.

In contrast, Euler characteristic–based descriptors provide a compact summary of topology by aggregating contributions from features of different dimensions into a single scalar quantity along the filtration. While this representation does not retain detailed information about individual topological features, it is computationally efficient and particularly suitable for feature construction in machine learning frameworks.

Taken together, these approaches should be viewed as complementary rather than competing tools. Persistent homology provides the most detailed multiscale topological information, GLMY homology extends topological analysis to generalized interaction structures, and Euler characteristic–based descriptors offer efficient global summaries suitable for large-scale data analysis. To clarify the distinctions among these approaches, Table 1 summarizes their mathematical foundations, advantages and limitations. In specific materials applications, the selection of methods should be determined by the nature of the task and the suitability of the corresponding tools.

Table 1. Comparison of topological descriptors used for chemical and materials systems

Method	Mathematical framework	Captured structural information	Advantages	Limitations
Persistent Homology	Simplicial complexes constructed from point clouds (e.g., VR or alpha complexes)	Multiscale topological features such as connected components, loops, and voids	Rigorous multiscale characterization; robust to noise; widely supported computational tools	Computational cost can increase rapidly with system size and dimension
GLMY Homology	Generalized complexes allowing directed or asymmetric interactions	Topological structures incorporating directional or higher-order relations	Extends homological analysis to systems with asymmetric or constraint-driven interactions	Mathematical framework less widely established; computational tools more specialized
Euler Characteristic–based Descriptors	Scalar topological invariants derived from complexes	Global topological summaries across filtration scales	Computationally efficient; compact descriptors suitable for machine learning	Less detailed structural information compared to full homology

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

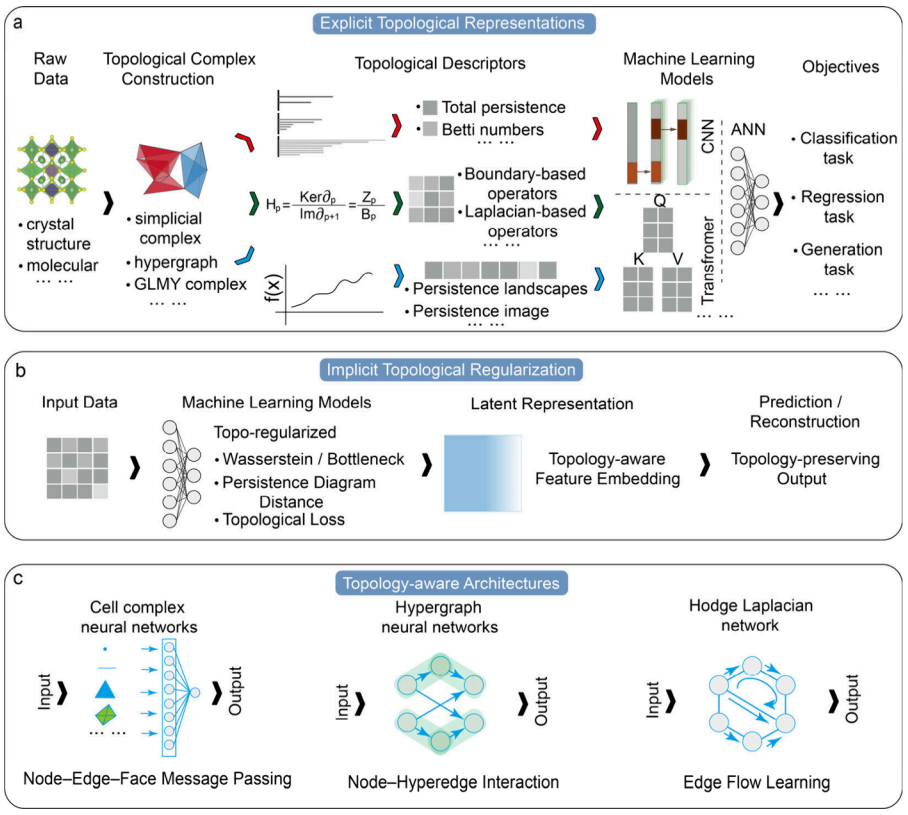


Figure 9. Schematic illustration of the three paradigms of topological learning: (a) Explicit topological representations. Raw structured data (e.g., crystal structures or molecular configurations) are first transformed into topological complexes such as simplicial complexes and hypergraphs. Topological descriptors are then extracted using tools from TDA. These descriptors include persistence-based summaries (e.g., persistence barcodes, total persistence, Betti numbers), operator-based representations derived from boundary or Laplacian operators, and functional summaries. The resulting descriptors are subsequently used as inputs for machine learning models to perform downstream tasks such as classification, regression, or generative modeling. (b) Implicit topological regularization. Instead of explicitly constructing topological descriptors, topology can be incorporated into machine learning models through topological regularization. In this paradigm, topological constraints are enforced during training via topological loss functions, typically defined using distances between persistence diagrams (e.g., Wasserstein or bottleneck distances). This encourages the learned latent representations to preserve the intrinsic topological structures of the input data, leading to topology-aware feature embeddings and topology-preserving predictions or reconstructions. (c) Topology-aware neural architectures. Some neural network architectures directly incorporate higher-order topological structures into their design. Cell complex neural networks perform message passing across nodes, edges, and higher-dimensional cells (e.g., faces) to capture multiscale topological interactions. Hypergraph neural networks model higher-order relationships through node−hyperedge interactions, enabling information propagation among groups of nodes. Networks based on the Hodge Laplacian operate on edge-based signals and learn flow-like structures through edge-level message passing and flow decomposition.

3. TOPOLOGICAL LEARNING

Topological learning has emerged as a unifying framework that integrates the mathematical foundations of TDA with the representational power of modern machine learning.[95] This paradigm leverages TDA to characterize the intrinsic structure of chemical and materials data, capturing global, multiscale, and connectivity-driven patterns that are often inaccessible to traditional descriptors.[41,51] Machine learning complements these capabilities by identifying nonlinear relationships, extracting latent features, and enabling predictive modeling across large and heterogeneous data sets.[8,96] The integration of these two methodologies creates structurally informed and data-efficient models with enhanced descriptive and predictive capacity.

We structure topological learning into three complementary parts (Figure 9). Section 3.1 describes how topological information is extracted and transformed into machine-readable numerical features, including summaries, operators, and vectorized descriptors. Section 3.2 introduces topology as a constraint in training objectives, where persistence-based distances and topology-preserving penalties guide model optimization without serving as explicit inputs. Section 3.3 highlights neural network designs whose structures are directly informed by topological operators and high-order connectivity.

3.1. Explicit Topological Representations

Topological invariants arise from rigorous mathematical theorems and are intrinsically defined on highly abstract spaces.[52] Consequently, any topological analysis of concrete physical structures can be understood as a procedure that maps raw structural data into the axiomatic framework of algebraic or combinatorial topology. As illustrated in Figure 9a, for a given structural system such as a crystalline lattice, molecular network, reaction pathway, or dynamical trajectory, the first step is to construct an appropriate topological complex.[16,51] Depending on the nature of the underlying data, the resulting complex may take the form of a simplicial complex, a path complex, a hypergraph, a directed hypergraph, or other generalized combinatorial complexes. Once constructed, the complex provides a formal mathematical object upon which various topological computations can be performed.

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

Topological summaries refer to scalar or low-dimensional descriptors that distill global structural information from a topological complex, as indicated by the red arrow in Figure 9a. These quantities compress the rich multiscale organization of the system into interpretable metrics that can be directly correlated with physical or chemical properties. Representative examples include Betti numbers, which count the number of connected components, independent rings, or higher-dimensional voids at a fixed filtration level.[97] These quantities quantify the basic topological skeleton of a structure, capturing connectivity patterns, ring motifs, and cage-like voids. Another family of descriptors consists of the Euler characteristic and related combinatorial invariants, obtained by taking alternating sums of cell counts or Betti numbers.[98] These measures offer a coarse but highly robust global summary of structural complexity and disorder. Beyond these classical invariants, persistent global descriptors such as total persistence, persistence entropy, integrated lifetime statistics, rank invariants, and Wasserstein-based summary scores provide richer scalar or low-dimensional characterizations.[99−103] Integrated lifetime statistics summarize the persistence of topological features by aggregating the lifetimes derived from persistence diagrams or barcodes, where the lifetime of a feature is defined as the difference between its death and birth times (d_i-b_i). Rank invariants are descriptors in persistent homology that quantify the number of homological features persisting between two filtration parameters, defined as the rank of the homomorphism induced between homology groups along the filtration. By aggregating information across the entire persistence diagram, they capture the overall stability, prominence, and scale distribution of the underlying topological features. Topological summaries remain popular in materials science, chemistry, and molecular modeling because they are explicitly interpretable, computationally efficient, and easily adapted to domain-informed invariants. However, their coarse nature inevitably collapses high-dimensional information into aggregated statistics, which are physically meaningful but may obscure fine structural nuances required by modern machine-learning pipelines. As a result, summaries often excel when interpretability is paramount but may lack the representational fidelity needed for tasks involving subtle geometric or chemical variation.

A second family of representations utilizes algebraic and spectral operators that arise during homology and persistent homology computations. These objects can be viewed as intermediate constructions because they are generated during the algebraic processing of the simplicial complex rather than serving as the final topological summaries. As indicated by the green arrow in Figure 9a, these topological operators retain a richer encoding of structure than scalar summaries because they preserve local and global incidence relationships within the complex. They are inherently matrix-based, making them immediately compatible with numerical linear algebra and machine-learning frameworks. Representative examples of such operator-based constructions include the boundary and coboundary operators, which map between k -and (k + 1) -dimensional faces and encode the fundamental adjacency structure that underpins homology computation.[49,97] The combinatorial Laplacian and the Hodge Laplacian are operators defined on graphs and simplicial complexes, respectively, that capture structural information across different dimensions and facilitate spectral analyses of the geometric and topological properties of complex systems.[104] Higher-order incidence operators designed for hyperedges, directed complexes, or multirelational structures provide algebraic representations suitable for modeling chemical bonding networks, reaction pathways, and directional transport processes. More general operator families such as those associated with GLMY complexes, quotient complexes, or directed complexes extend classical constructions to asymmetric, weighted, or constraint-driven settings that arise frequently in chemical and materials systems.[74,105−107] GLMY complexes are combinatorial constructions associated with digraphs that extend classical simplicial frameworks to digraph settings, forming the basis for the homology theory of directed networks. A quotient complex is obtained by identifying certain cells or substructures of a complex according to an equivalence relation, resulting in a reduced complex that preserves essential topological properties. A cell complex is a topological space constructed by attaching cells of different dimensions (such as points, line segments, polygons, and higher-dimensional analogues) via continuous boundary maps, forming a structured decomposition of the space.[52] A directed cell complex is a cell complex in which orientation or direction is assigned to its cells or incidence relations, enabling the representation of asymmetric interactions or flows within the structure.[108] In addition, spectral operators including Laplacian spectra, heat kernels, diffusion operators, and spectral embeddings reveal multiscale structural organization and establish natural connections between topological complexes and geometric or dynamical processes.[109−112] The Laplacian spectrum of a graph or simplicial complex refers to the set of eigenvalues of its Laplacian operator, which encode structural and topological properties of the underlying network or complex. The heat kernel is a function derived from the Laplacian operator that describes the diffusion of heat (or information) over a graph or topological space over time. Diffusion operators are operators derived from graph or simplicial Laplacians that model diffusion processes on networks or complexes, capturing how information propagates across the structure. Spectral embedding refers to a dimensionality reduction technique that maps data points into a low-dimensional space using eigenvectors of Laplacian or diffusion operators, preserving structural relationships in the data.

Because these operators are at their core matrices or linear maps, they have fixed or scalable dimensions and can be directly fed into graph neural networks, spectral learning algorithms, diffusion models, or operator-based machine-learning architectures. Importantly, they preserve more detail than summaries, enabling models to exploit high-order structural correlations in complex materials or molecular data sets. However, this expressive capacity comes at a cost. The physical interpretation of individual matrix entries or eigenmodes is often nontrivial, and the spectral signatures may mix geometric and topological information in a way that is difficult to disentangle. As a result, topological operators, despite their compatibility with advanced learning systems, can reinforce the "black-box" nature of predictive models if they are not paired with appropriate interpretability strategies.

Between the extremes of global summaries and high-dimensional operators lies a rapidly expanding class of topological vectorizations, which translate persistence-based information into structured numerical vectors suitable for statistical or deep-learning models. These approaches seek to preserve the interpretability and stability of topological

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

invariants while providing richer representational capacity than scalar summaries. A defining feature of vectorization is that the resulting representation is not merely a collection of independent summary values. Instead, it forms a continuous, numerically correlated sequence in which neighboring components encode smoothly varying geometric or homological information. This stands in contrast to stacked topological summaries, whose entries are typically independent and lack intrinsic ordering. As indicated by the blue arrow in Figure 9a, vectorization methods such as persistent images, landscapes, Betti sequences, and kernel-based embeddings therefore produce feature spaces with meaningful internal geometry, enabling models to exploit correlations across scales and capture finer structural variations within the underlying topological patterns.

Persistence landscapes encode each diagram as a sequence of piecewise-linear functions whose values are sampled on a predefined grid.[56] Persistent images map each topological feature to a point in the plane and apply Gaussian smoothing followed by rasterization, yielding an image-like array that preserves stability properties.[53] Betti curves record the evolution of Betti numbers across filtration scales, providing a coarse but interpretable summary of topological activity.[113] Silhouette functions aggregate persistence landscape layers into a single curve weighted by persistence magnitudes.[114] Silhouette functions provide a weighted functional summary of persistence diagrams by averaging persistence landscapes while assigning larger weights to topological features with longer lifetimes. Persistence entropy measures the Shannon entropy of normalized lifetimes and produces a single scalar that captures the distributional spread of topological features.[54] Landscape functions fall naturally into this category, but a broader class of approaches constructs kernel density estimates on the space of birth−death pairs, smoothed Betti functions, or functional transforms such as the cumulative persistence function.[56] Additional examples include scale-space kernels or other kernel density approximators that generate stable functional embeddings through controlled smoothing.[53]

As shown in Figure 9a, With the topological structures now encoded into machine-ready representations, these features can be fed into downstream machine-learning models. For example, these representations can be naturally processed by convolutional neural networks (CNNs), allowing CNNs to identify characteristic patterns associated with connectivity changes, cavity distributions, or multiscale geometric motifs.[115] Explicit encoding is straightforward to implement, computationally efficient, and highly compatible with conventional machine learning pipelines. Because topological features are precomputed, the learning algorithm operates on stable and low-dimensional summaries that enhance interpretability and offer strong inductive biases. However, the representation quality depends heavily on the choice, preprocessing, and vectorization of topological descriptors. Improper featurization may discard critical information, and fixed encodings may not adapt optimally to task-specific notions of similarity. Despite these limitations, explicit topological encoding remains one of the most practical and widely adopted strategies for integrating topological data analysis into materials modeling.

3.2. Implicit Topological Regularization

Implicit topological regularization refers to learning strategies in which topological information does not enter the model as an explicit input feature but instead influences the optimization objective. As shown in Figure 9b, the goal is to guide the learned latent space, energy landscape, or generative distribution toward representations that respect the underlying topology of the data. The model therefore does not consume topological descriptors directly; rather, topology shapes the training dynamics and restricts the family of admissible solutions.

A representative example is the use of topo−regularized autoencoders, in which the encoder−decoder network is trained not only to reconstruct the input but also to preserve its homological structure.[116,117] This is typically formulated by penalizing discrepancies between the persistence diagrams of the original and reconstructed data, so that the autoencoder learns a latent manifold that respects essential topological features. Another important class of methods introduces matching distance or Wasserstein distance between persistence diagrams as part of the optimization target.[118] When two samples are required to remain topologically similar, the network is encouraged to evolve its representations in a way that minimizes topological distortion across training batches. Contrastive learning provides a related mechanism where positive pairs share similar persistence signatures, while negative pairs differ in their cycle or connectivity structure. Contrastive learning is a self-supervised learning framework that learns feature representations by maximizing similarity between related samples while minimizing similarity between unrelated ones.[119] The contrastive loss implicitly aligns latent embeddings according to topological similarity, promoting representations that are sensitive to connectivity and scale interactions.[120,121] Implicit regularization is also used in generative models such as diffusion or transformer architectures, where penalties on the topological divergence between intermediate states and target distributions guide the generative trajectory. This ensures that the model maintains key homological patterns, such as pore connectivity or cycle distributions, even as it explores high dimensional chemical or structural spaces.

Implicit regularization offers several advantages. Because the topology is encoded through the loss function rather than the input space, the method is highly flexible and can be coupled with any differentiable architecture. It also preserves the original data representation, which is especially beneficial when geometric or physical fidelity must be retained. Furthermore, this strategy allows the model to learn a latent space that is sensitive to topological changes even when those changes are subtle or sparsely represented in the data. However, the approach also has limitations. Persistence-based losses can be computationally demanding and challenging to differentiate, which limits scalability for very large data sets or high-throughput applications. The effect of the topological constraint can be difficult to tune, and overly strong penalties may hinder optimization or trap the model in suboptimal minima. Finally, because topology is not explicitly represented in the input, interpretability can remain limited, and the learned representations may be hard to analyze without additional probing.

3.3. Topology-Aware Architecture

Topology−aware architecture represents the most explicit and structurally integrated form for topological learning. In this paradigm, as shown in Figure 9c, the neural network is constructed so that its computational primitives mirror the algebraic and combinatorial relations of the underlying

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

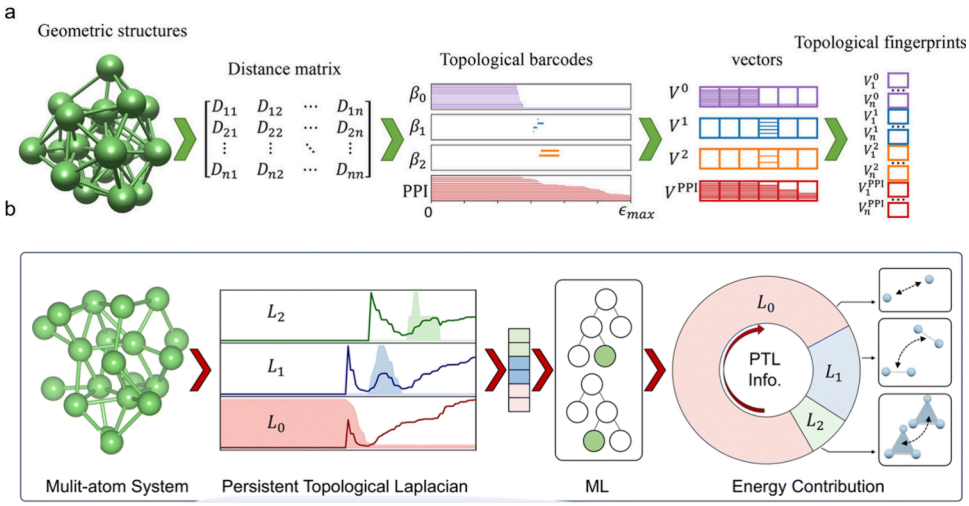


Figure 10. (a) Flowchart illustrates the workflow from a geometric structure, distance matrix, topological barcodes, and vectors of topological invariants to final topological fingerprints of a lithium cluster. Reproduced from ref 24. Copyright 2020 American Chemical Society. (b) Workflow for energy prediction in multiatom systems using the persistent topological Laplacian. The 20-atom Li cluster is treated as a simplicial complex, and the persistent topological Laplacian method is applied to capture system characteristics across different dimensions, specifically the 0-simplices (connected components), 1-simplices (loops), and 2-simplices (voids), represented by L_0, L_1, and L_2, respectively. Adapted from ref 129. Available under a CC-BY 3.0 license. Copyright 2025 The Author(s).

topological objects. Simplicial complex neural networks exemplify this strategy by defining learnable functions on vertices, edges, faces, and higher-dimensional simplices while coupling these signals through boundary and coboundary operators. This design allows information to propagate across different homological dimensions and enables the network to reason about cycles, cavities, and other nonpairwise interactions.

Cell complex neural networks generalize this idea by allowing heterogeneous cells with arbitrary incidence relations, providing a flexible mechanism to encode porous networks, crystal motifs, or heterogeneous molecular environments in which higher-order adjacency cannot be reduced to standard graph structure.[122] Hypergraph neural networks take a complementary approach in which interactions are modeled through hyperedges that link multiple nodes simultaneously.[123] This architecture captures cooperative or multibody structural patterns that are common in chemical bonding networks and framework materials. Networks built on the Hodge Laplacian introduce the spectral structure of discrete differential geometry directly into message passing, allowing the model to decompose information into gradient, curl, and harmonic components.[124,125] This provides a principled way to learn from directional transport channels, ring currents, or ionic migration pathways. More broadly, higher-order message-passing mechanisms unify these developments by enabling neural updates that depend on the configuration of simplices, cells, or hyperedges rather than individual edges alone. Topology-aware architectures therefore encode the relational grammar of the system directly into the model's backbone, yielding representations that are intrinsically sensitive to multiscale connectivity and homological organization. This deep integration offers strong interpretability and structural fidelity, although it often requires specialized data structures and may introduce additional computational complexity when applied to large materials data sets.

4. THE APPLICATIONS OF TDA IN MATERIALS SCIENCE

Having introduced the theoretical foundations of TDA and its integration with machine learning frameworks, we now turn to its practical applications in materials science. The ability of TDA to capture multiscale structural organization, connectivity patterns, and higher-order correlations makes it particularly suitable for analyzing complex material systems. We first present studies where TDA has been employed to describe fundamental material properties. We then turn to its applications in specific material systems. Among these, the most extensively explored systems include polymer materials, glassy materials, heterogeneous catalysts, nanoporous materials and superionic conductors. These systems represent the most rapidly evolving areas of TDA application and will therefore be discussed in detail. In addition, TDA has also shown initial progress in several other classes of material systems, which we briefly highlight. Finally, we summarize the various types of material data that can be characterized and represented by TDA.

4.1. Fundamental Material Property

As a key thermodynamic indicator of material stability, the formation energy is one of the most fundamental yet critical properties in materials science. It reflects whether a material can stably exist under given conditions and serves as a primary criterion for evaluating its feasibility of synthesis and application in real-world environments. In high-throughput materials screening and materials informatics, formation energy is commonly used as a primary filter to eliminate thermodynamically unstable candidates on a large scale.[126−128]

At the microscopic level, formation energy is strongly influenced by structural factors such as atomic packing and many-body interactions. These features are often difficult to fully capture using conventional descriptors or handcrafted features, yet they fall naturally within the representational strength of TDA. Chen et al. employed a topological learning framework to predict the formation energies of lithium clusters, aiming to identify their most stable configurations.[24]

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

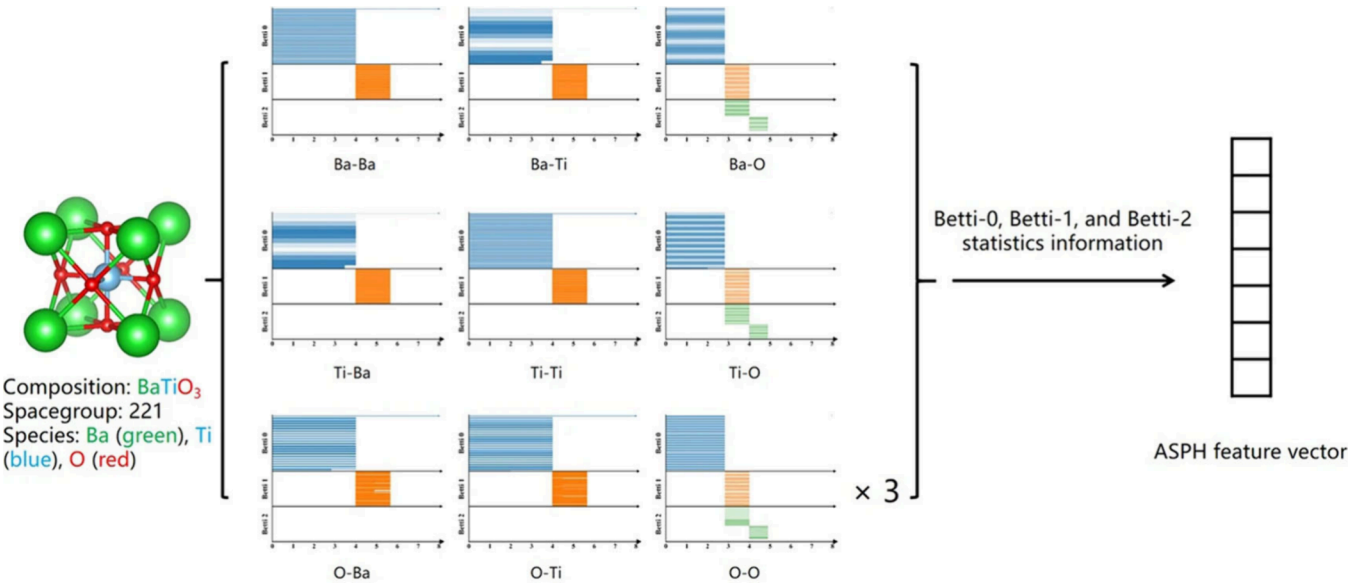


Figure 11. Construction process of $BaTiO_3$ topological feature. First, compute all barcodes of the combinations formed by each atom in the unit cell as a central atom and surrounded by various element types. The ASPH features are then generated from the statistics data of these barcodes. Since the unit cell of $BaTiO_3$ has three O atoms, the barcode data of this type in the figure is multiplied by three. Reproduced from ref 23, Available under a CC-BY 4.0 license. Copyright 2021 The Author(s).

They used persistent homology to extract topological features from each structure, which include traditional Betti-number features along with the proposed persistent pairwise independence (PPI) descriptor (Figure 10a). The PPI descriptor measures the independence of each pair of atoms during the filtration process: initially all atoms are disconnected and each independent atom pair corresponds to a PPI bar, while increasing the filtration parameter causes connected pairs to terminate their bars. Because it captures pairwise structural relationships, PPI provides complementary information beyond the β_0 persistence barcode. When combined with Gradient Boosting Regression (GBR), the topology-based model achieved strong predictive performance for Li_{20} clusters, with a Pearson correlation coefficient (PCC) of 0.993 between predicted and real binding energies. This performance outperformed traditional geometric descriptors based on local atomic density representations (PCC = 0.978), highlighting the potential of TDA-based descriptors in energy prediction. They further introduced the persistent topological Laplacian (PTL), an operator-level byproduct in persistent homology, to encode structural information across different topological dimensional, capturing complex many-body interactions (Figure 10b).[129] In the prediction task of the formation energy of Li_{20} clusters, the optimal performance achieved by this method is PCC = 0.982. Topological invariants such as Betti numbers and minimal nonzero eigenvalues were found to correlate with structural stability and atomic connectivity. Among them, 0-simplices had the strongest impact on energy prediction, while 1-simplices and 2-simplices features contributed to structure ranking, such as identifying low-energy configurations. This interpretability stems from TDA's ability to preserve multiscale topological signatures, allowing models not only to predict energies accurately but also to capture how structural features influence energetic stability.

In comparison, the two approaches differ fundamentally in their methodological roles. PPI descriptors can be viewed as higher-level statistical summaries constructed from Betti number–based topological information, which aggregate persistence patterns into optimized descriptors tailored for regression tasks, thereby yielding strong predictive accuracy. In contrast, PTL provides an operator-based representation rooted in algebraic topology, where harmonic spectra arise naturally as intermediate structures associated with homology and persistent homology computations. Although the PPI model achieves higher predictive accuracy for Li_{20} clusters, PTL features offer a more systematic and interpretable multiscale framework, enabling consistent predictions across multiple feature subsets and evaluation metrics. This comparison suggests that descriptor-level statistical aggregation favors predictive performance, whereas operator-based topological representations provide broader structural interpretability and extensibility for analyzing complex many-body systems.

TDA has been further applied to predicting the formation energies of crystalline materials.[23,130] Unlike clusters, crystals exhibit periodicity, which requires accounting for periodic interactions in feature representation. To address this, Jiang et al. proposed Atom-specific Persistent Homology (ASPH) for structural characterization.[23] The method treats each atom in the unit cell as a center and identifies neighboring atoms of different types within a supercell-expanded structure and a defined cutoff radius. Persistent homology is then applied to these atomic environments to generate persistence barcodes, from which topological summaries (e.g., minimum, maximum, mean, standard deviation, and sum of birth, death, and persistence values) are extracted and concatenated into global material representations. An example using $BaTiO_3$ is shown in Figure 11. Combined with GBR, ASPH achieved a prediction accuracy of 61 meV/atom on a data set of 31,912 crystals, outperforming traditional descriptors such as Voronoi tessellations and the Coulomb matrix.[10,131] Building on this, Fang et al. integrated ASPH features into various graph neural network (GNN) models, including the crystal graph neural network (CGCNN), the graph attention neural network (GAT), and graph transformer, to predict oxygen vacancy

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

formation energies in perovskite materials.[132] Incorporating persistent homology features, along with appropriate graph pooling strategies, significantly improved predictive performance, reducing the mean absolute error (MAE) by 55%. This improvement arises from the ability of TDA-based features to capture multiscale and high-dimensional geometric structures, whereas traditional graph representations are often limited to local connectivity. As a result, TDA offers a more expressive framework for modeling the complex structure−property relationships.

In recent years, the rapid development of machine learning interatomic potentials (MLIPs) has significantly advanced materials modeling.[133−135] Accurate prediction of formation energies is a core objective in MLIPs design. Constructing high-quality MLIPs requires input features that are invariant to physical symmetries such as rotation, translation, reflection, and atomic permutation.[136] Persistent homology, as a point cloud−based topological method, naturally satisfies rotational and translational invariance and offers strong robustness to reflection and permutation when appropriately designed. As such, it holds great potential for feature construction in MLIP frameworks. Minamitani et al. conducted an initial exploration in this direction.[137] Using amorphous carbon as an example, topological features such as loops were extracted as birth−death pairs, encoded into persistence diagrams, and converted into histograms representing their distribution across spatial subregions. These histograms were normalized into probability distributions and standardized via z-score to ensure size invariance. Combined with CNNs, the topological descriptors achieved a prediction error of 40.7 meV/atom. The performance, while comparable to traditional and modern state-of-the-art descriptors, still shows a modest gap: the classical hand-crafted SOAP achieves an accuracy of 16.4 meV/atom, whereas the graph neural network−based SchNet reaches 30.5 meV/atom. One key reason for this limitation is that persistent homology primarily encodes topological invariants and medium-range structural organization, but has limited sensitivity to fine geometric features such as subtle bond-angle distortions and bond-length variations, which can play a critical role in determining the energetics of amorphous systems. Nevertheless, topological-based descriptors exhibit distinct methodological advantages. They require minimal hyperparameter tuning, involve lower model complexity, and offer enhanced interpretability. For instance, inverse analysis enables direct identification of regions in the persistence diagram that contribute significantly to high− or low−energy configurations, which can be further mapped back to specific local structural environments, thereby providing a physically meaningful interpretation of the structure−energy relationship. Additionally, unlike most modern MLIP frameworks that rely on atom-centered descriptors to represent local atomic environments, the present approach introduces a global descriptor derived from topological representations of the entire structure.

Beyond formation energy prediction, TDA has shown broad applicability in modeling other fundamental material properties. In potential energy surfaces modeling, sub- or superlevel set persistent homology facilitates the extraction of multiscale topological features that capture subtle structural variations impacting the energy landscape.[138] For inherently high-dimensional electronic charge densities of crystalline compounds, persistent homology offers a compact, topology-based representation that preserves key spatial characteristics while reducing computational complexity.[26] In phase classification, persistent homology-derived invariants serve as structural fingerprints, enabling accurate phase distinction on par with state-of-the-art machine learning algorithms.[139] In band gap prediction, TDA-based descriptors, by capturing higher-order interactions and periodic boundary information, significantly outperform existing methods in predictive accuracy.[140] Persistent GLMY homology has been employed to characterize the extent of lattice distortion in materials induced by the Jahn−Teller effect.[71] These examples underscore TDA's unique ability to extract robust topological invariants from diverse material data sets, reinforcing its role as a powerful and general-purpose framework for data-driven materials modeling.

4.2. Polymer Materials

Polymer materials are a broad class of substances composed of long-chain macromolecules formed by the repeated linking of simpler molecular units called monomers.[141] Owing to their structural diversity and tunable physicochemical properties, polymers exhibit a wide range of mechanical flexibility, thermal stability, chemical resistance, and processability. Polymers often display amorphous or semicrystalline morphologies, leading to complex, disordered molecular arrangements that pose significant challenges for structural characterization and predictive modeling. Polymers are widely used across numerous fields, including packaging, additive manufacturing, biomedicine, electronics, and energy applications.[142] Their functionality can be tailored through molecular design, copolymerization, cross-linking, or blending, making them highly versatile for both structural and functional applications.[143] Given their inherent complexity and structural variability across multiple length scales, polymer systems offer a compelling platform for applying TDA to extract meaningful structural descriptors and uncover structure−property relationships.

TDA has been initially applied to characterize fundamental structural features in polymeric materials, such as voids, rings, and thread-like formations, and to track their evolution under varying environmental conditions. In a study by Ichinomiya et al., persistent homology was introduced to quantitatively analyze the formation of craze structures in glassy polymers under tensile deformation.[144] By constructing Persistent diagrams from coarse-grained molecular dynamics (CGMD) simulations, the authors identified the formation, growth, and coalescence of cavities (captured as H_2 homology features) in a parameter-free and global manner. While the number of voids initially increased and then decreased with strain, the average void volume expanded significantly, indicating the merging of small voids around the yield point into percolating pathways. This process suggests that material yielding can be interpreted as a void percolation transition, offering a novel topological perspective on polymer failure mechanisms. In another work, Shimizu et al. employed all−atom MD simulations to investigate the structural evolution of polar nitrile-butadiene rubber under varying simulation cell sizes.[145] Using persistent homology, the study extracted topological features, particularly H_2 cavities, and showed that unlike radial distribution functions (RDF), persistence diagrams captured the transition of cavity structures from discrete to continuous with increasing system size, and this behavior correlated strongly with the convergence of dielectric constant predictions. This work highlights the capability of persistent homology to capture microstructural topology that traditional geometric methods like RDF may overlook.

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

In ring polymer systems, topological constraints originate from chain noncrossability and ring entanglement. Threading, where one ring penetrates another, is a key mechanism that restricts chain mobility and governs relaxation dynamics. Landuzzi et al. proposed a persistent homology–based method to rigorously detect such threadings.[146] In this framework, "homological threading" (H–threading) refers to a class of topological constraints between ring polymers identified through persistent homology, where the threading relationship is determined by the homological properties of the configuration and can be uniquely detected without introducing additional parameters. The approach compares the persistence diagrams of individual rings with those of merged ring pairs. As illustrated in Figure 12a, when two rings are not

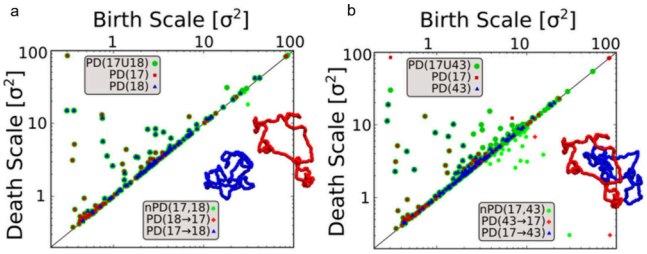


Figure 12. Persistence diagrams–based H–threading analysis: persistent diagrams of two rings PD(i) (red), PD(j) (blue), and persistence diagrams obtained from the union of two rings, i.e., twin configuration PD (i ∪ j) (green). (a) The persistence diagrams of two rings (i = 17 and j = 18) that have no H-threading between them; this is evident in the persistence diagrams since all the red and blue points are overlapped by green points, i.e., there are no lost points. Snapshot at bottom right provides a visible inspection of the spatial position of the two rings. (b) The ring j = 43 that is H-threading into i = 17 as mirrored by the presence of lost points, i.e., the red point in the top left (primal loop of ring (i) is not overlapped by a green one (the snapshot confirms this H-threading event). The red point in the lower side corresponds to this H-threading, and with them a cloud of new points is formed. The upper region shows the standard persistence diagrams (PDs in figure) of the two rings and their union configuration. The lower region displays the corresponding difference persistence diagrams (nPD in figure), highlighting new and lost persistence points, which are plotted below the diagonal for visual separation. Adapted from ref 146. Available under a CC-BY 4.0 license. Copyright 2020 The Author(s).

threaded, their topological features in the persistence diagram largely coincide. In contrast, as shown in Figure 12b, if a topological feature corresponding to one ring disappears in the merged diagram, this "lost point" is defined as a one-dimensional homology threading. This disappearance indicates that the ring's structural integrity has been disrupted by the penetration of another ring. This method overcomes the limitations of heuristic-based techniques and reveals that the number of H-threading events increases with chain length, highlighting the scalable and essential role of threading in controlling polymer diffusion and dynamics.

Leveraging the capability of TDA to effectively characterize polymer microstructures, topological learning has been introduced to build predictive models linking structure to properties. Sato et al. combined persistent homology with PCA to correlate microscopic structure with stress–strain behavior in polymer films under tensile deformation.[147] Through CGMD simulations of poly(ethylene oxide) films, they encoded topological ring structures into persistent images. The first principal component (PC1) derived from these images closely tracked the stress–strain curve (Figure 13a),

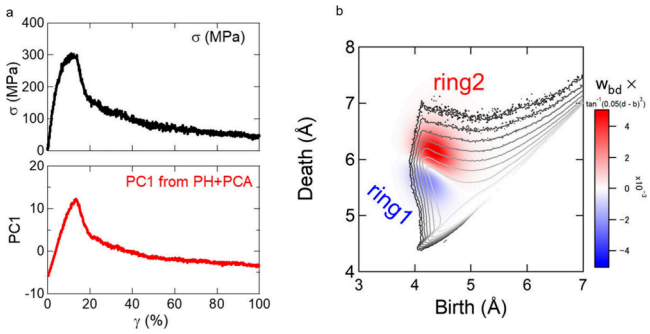


Figure 13. (a) (Top) Stress–strain curve and (bottom) PC1 of the Persistent images obtained by PCA during the uniaxial tensile MD simulation for the polymer film. (b) Distribution of $w'_{bd} = \tan^{-1}(0.05(d - b)^3)w_{bd}$, the coefficient for PC1, at each (b, d) mesh. As indicated by the color scale, red corresponds to positive and blue corresponds to negative coefficients for PC1. The contour lines in the figure represent the time-averaged Persistent diagrams during the MD simulation. Adapted from ref 147. Copyright 2024 American Chemical Society.

confirming that mechanical response is captured by topological evolution. Spatial mapping of PC1 coefficients (Figure 13b) revealed two competing ring types: ring2 (red), whose growth increases stress, and ring1 (blue), whose decline reduces it. Inverse analysis confirmed the dominance of small rings (≤10 beads) in governing mechanical behavior. Present in both the stable ring1 and transient ring2, the stress–strain response emerges from the competition and transformation between these small-ring-dominated structures during stretching. Similar unsupervised clustering methods with persistent homology have been employed to investigate the complex nanophase-separated structures in polymer electrolyte membranes, successfully establishing a correlation between network morphology and proton conductivity.[148] Kojima et al. proposed a method integrating persistent homology with Logistic Regression (LR) to identify critical microstructural features influencing the mechanical behavior of filled rubber.[149] Persistent homology was used to extract topological loop structures representing particle connectivity, which were vectorized into persistence images as inputs to the LR model. The learned coefficients revealed which topological features correlated with high or low stress. This enabled the extraction of stress–contributing aggregates, which were found to be anisotropic and elongated along the loading direction, highlighting their key role in stress enhancement. Xia et al. introduced a novel multi–cover persistence–based molecular representation for polymers, which utilizes Delaunay slices at different dimensions and Rhomboid tiling to characterize the complicated geometric and topological information.[150] When combined with gradient boosting tree models, the model achieves superior or comparable predictive performance for polymer properties to existing methods, particularly for large and complex monomer structures.

The phase separation, aggregation, and self-assembly of polymers can lead to gelation, resulting in functional materials with extensive applications in pharmaceuticals, foods, and construction.[151,152] These gels are also characterized by their structural complexity, exhibiting highly variable topology networks and porosity. Smith et al. employed the Euler

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029–7067

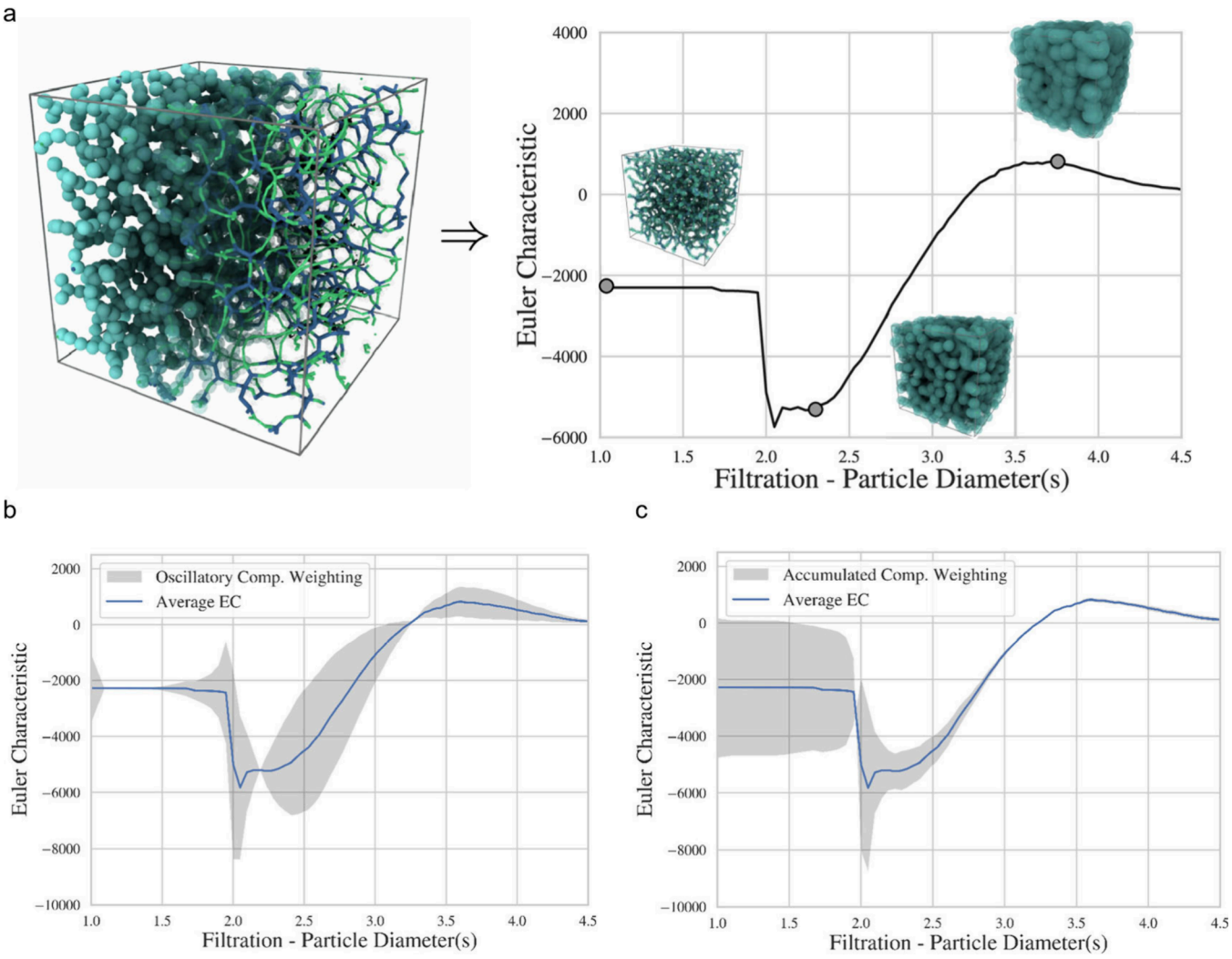


Figure 14. (a) The Euler characteristic curve for the gel simulation along with a visualization of the changes in the simulation topology at various filtration diameters. Visualization of the (b) oscillatory and (c) cumulative component weightings for the Euler characteristic curves. The area of the weightings (gray) represents higher emphasis placed on that particular portion of the Euler characteristic dynamics during oscillatory shear. Adapted from ref 153. Copyright 2024 American Chemical Society.

characteristic curves to track topological changes during gel self−assembly and shear deformation (Figure 14a).[153] Principal component analysis of the Euler characteristic curves for the gel structures successfully identified two key signals: an oscillatory component, representing rapid and reversible structural responses under external loading and primarily associated with mesoscale features such as pores and cavities (Figure 14b); and an accumulated component, reflecting slow, irreversible structural reconstruction, mainly linked to the breaking and reformation of local bonding networks (Figure 14c). These topological signals exhibit distinct behaviors across different strain amplitudes, enabling precise differentiation between linear, nonlinear, and flow deformation regimes. Further analysis revealed a strong correlation between the accumulated component and shear stiffening (R^2 = 0.97), indicating that local bond rearrangement is the primary microscopic mechanism of stiffening. In contrast, large−scale reorganization of the mesoscale structure in the flow regime mitigates the stiffening effect. These findings not only uncover the causal relationship between multiscale structural evolution and macroscopic mechanical performance but also demonstrate that topological signals can quantitatively predict gel mechanics under varying loading conditions.

4.3. Glassy Materials

Glassy materials are a class of widely used inorganic amorphous materials characterized by a lack of long−range atomic order, exhibiting a highly disordered structure.[154] They have unique physical and chemical properties such as transparency, excellent chemical stability, high mechanical strength, and good electrical insulation.[155] Glassy materials have broad applications in modern society, including in optics, construction and electronic packaging.[154] Typical glassy materials include silica glass, phosphate glass, and metallic glasses.

The primary use of TDA in glassy materials centers on the characterization and understanding of MRO structures in glassy systems.[20] In covalent glasses, MRO is typically defined as structural features on the length scale of 5 to 20 Å, including arrangements of connected polyhedra, superstructural units, and clustering around network−modifying species. Neutron and X−ray diffraction are widely employed to probe MRO by measuring the structure factor $S(q)$.[156] However, due to the

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

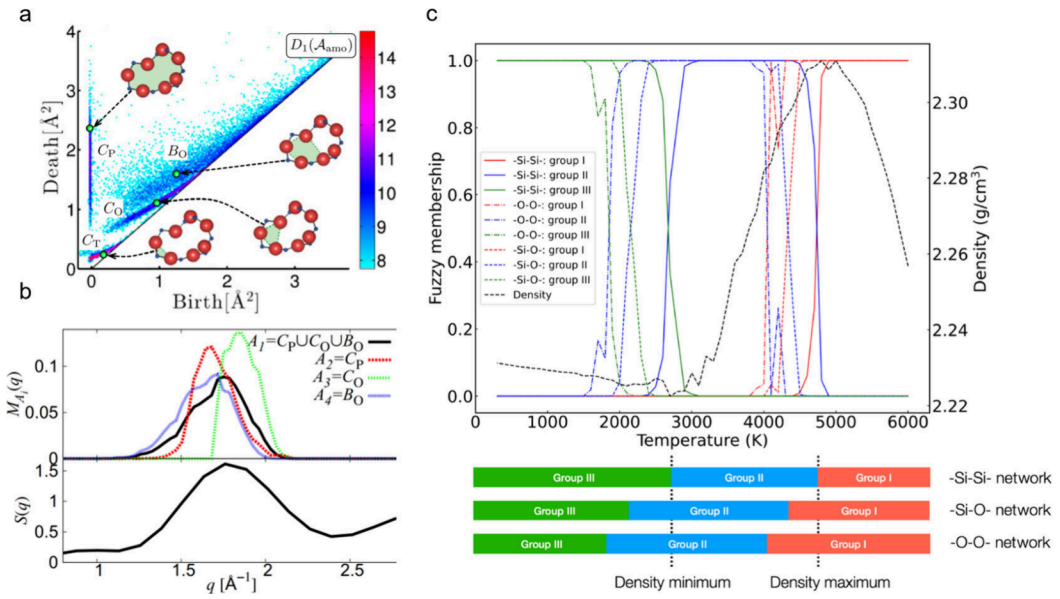


Figure 15. (a) Persistent diagrams of the amorphous state of silica. The three characteristic curves and one band region are labeled C_P, C_T, C_O, and B_O, respectively. (b) The upper panel is the distributions of M_{A_i} $(q)(i = 1,2,3,4)$ calculated from the C_P, C_T and C_O. The bottom panel is the structure factor S(q) around the FSDP of amorphous silica. Reproduced with permission from ref 157. Copyright 2016 National Academy of Sciences. (c) (Upper panel) Fuzzy memberships of the − Si−Si−, − O−O−, and − Si−O− networks of silica. The Persistent diagrams are divided into three different clusters (I−III) using the fuzzy spectral clustering algorithm. The intersection points for groups I (red) and II (blue) are located at 4735, 4048, and 4334 K for − Si−Si−, − O−O−, and − Si−O− networks, respectively. The intersection points for groups II (blue) and III (green) are located at 2674, 1856, and 2153 K for − Si−Si−, − O−O−, and − Si−O− networks, respectively. The density of the silica (15 000 atoms) is plotted on the right *y* axis as a function of temperature. (Lower panel) Visualization of the major groups of each network as a function of temperature as obtained from the TDA. Reproduced from ref 162. Copyright 2024 American Chemical Society.

reciprocal−space nature of $S(q)$, its direct structural interpretation is nontrivial. In particular, the microscopic origin of the First Sharp Diffraction Peak (FSDP) in $S(q)$ has been under long-standing debate, presenting significant challenges for accurately resolving MRO structures. Traditional descriptors such as bond angle distributions or ring statistics often fail to fully capture the multiscale geometrical characteristics of MRO. To address this limitation, Hiraoka et al. introduced the use of persistent homology to analyze atomic configurations by constructing persistence diagrams, which summarize atomic rings and cavities across multiple length scales[157] (Figure 15a). By extracting the birth and death scales of these features, the authors infer their real-space sizes and associate them with corresponding wavevector ranges in diffraction data. Through statistical analysis of the density distribution of specific topological structures in persistence diagrams, the study establishes a direct correspondence between MRO atomic rings and the FSDP observed in diffraction spectra, thereby revealing the real-space origin of this peak. The authors applied this methodology to silica glass, where the persistence diagrams successfully identified medium-range oxygen ring structures responsible for the FSDP (Figure 15b). Similar approaches have been proposed for investigating the microscopic origin of MRO in alkali-metal-doped silicate glasse,[19,158,159] phosphate glasse,[160] Ge-based glass.[161]

Persistent homology has been further applied to investigate the microstructural evolution of glasses under external fields. Pan et al. combined persistent homology analysis with large-scale MD simulations to study how ring and cavity structures in silicate glasses evolve under varying Na_2O concentrations and uniaxial strain.[163] To quantify topological changes, the authors introduced a topological similarity metric based on normalized 2D cross−correlation of persistence diagrams, enabling sensitive detection of strain-induced structural variations. They show that strain predominantly affects Si−O−Si ring structures, with minimal impact on cavities and Na distribution. Notably, despite local rearrangements under high strain, the glass network exhibits high topological reversibility after fracture, reflecting its structural resilience. Tirelli et al. applied persistent homology to explore the density anomaly in liquid silica during cooling, where the density exhibits a non−monotonic trend with both a maximum and a minimum at specific temperatures.[162] Although widely observed, the structural origin of this anomaly remained unclear. Their study revealed that topological transitions in the Si−Si network coincide with the temperatures of the density extrema, while changes in the Si−O and O−O networks occur at lower temperatures. Using Persistent diagrams across temperatures, they constructed a distance matrix and applied fuzzy spectral clustering to automatically classify topological states. This method effectively captured gradual structural changes and identified transition points via intersections of membership functions. The close match between Si−Si topology transitions and density extrema provides strong evidence linking network rearrangements to the anomaly (Figure 15c).

Topological learning is used to elucidate how the underlying structure arrangement, e.g., MRO, influence the properties of the glass. Minamitani et al. investigated the relationship between thermal conductivity and MRO in amorphous silicon.[164] By employing the persistence image framework to quantitatively describe the topological features of amorphous structures, and applying a ridge regression model, they achieved highly accurate predictions of thermal conductivity with R^2 at 0.99. Their analysis revealed that five-membered

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

ring structures serve as the fundamental MRO units governing thermal transport, whereas rapid quenching introduces four-membered rings at around 3 Å that disrupt MRO. (Figure 16a and 16b). The perturbation of the MRO further leads to a broader peak of the dynamical matrix elements in the fast-cooled samples within the range of 0.4−0.7 eV amu^{-1} Å^{-2} (Figure 16c). Such enhanced inhomogeneity in the force constants promotes stronger vibrational localization, which in turn suppresses the mode diffusivity and ultimately results in a reduction of the thermal conductivity. This study uncovers a direct link between local topological motifs and heat conduction, offering a novel theoretical and methodological foundation for tuning the properties of amorphous materials through structural topology. Interestingly, in amorphous Ge, five− and even six−membered ring structures have also been suggested to contribute to the extension of the mean free path of thermal vibrations, thereby enhancing thermal conductivity.[165] In amorphous alumina (a−Al_2O_3), high−density systems exhibit more six-vertex cavity structures linked to stable, face-sharing AlO_6 networks, enabling efficient heat conduction. In contrast, low−density systems show network collapse and increased distorted or five-coordinated AlO_6 clusters, limiting diffusion transport and lowering thermal conductivity.[166]

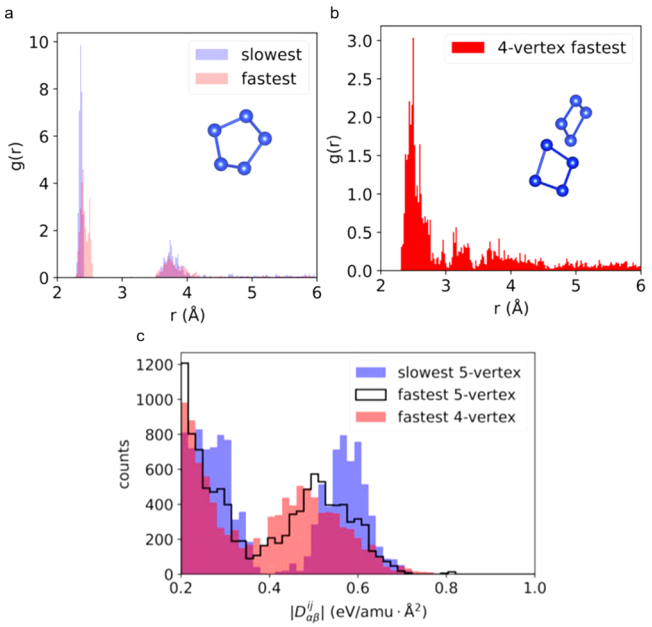


Figure 16. (a) Calculation results of RDF restricted to five-vertex cycles in volume−optimal cycles. The inset shows an example of the cycle considered in this plot. (b) Calculation results of RDF restricted to the four-vertex cycles in the volume−optimal cycles of the children's birth−death pairs in the fastest-cooled models. The inset shows an example of the cycle considered in this plot. (c) Histograms of the absolute value of the dynamical matrix elements correlated with the atoms in the five-vertex cycles in both models and the four-vertex children's cycles in the fastest-cooled model. Adapted from ref 164. Available under a CC-BY 4.0 license. Copyright 2022 The Author(s).

4.4. Heterogeneous Catalysis

Heterogeneous catalysis, owing to its outstanding energy conversion efficiency, reaction selectivity, and environmental sustainability, has demonstrated tremendous potential across critical domains such as energy transformation, energy storage, and fine chemical synthesis.[167,168] A range of challenging reactions, including green hydrogen production, carbon dioxide (CO_2) reduction, nitrogen-based resource conversion, and fuel cell processes, rely on the involvement of efficient heterogeneous catalysts.[169−171]

Adsorption, a critical step in heterocatalysis, involves electronic reorganization between adsorbates and catalyst surfaces, driving energy redistribution and structural match-

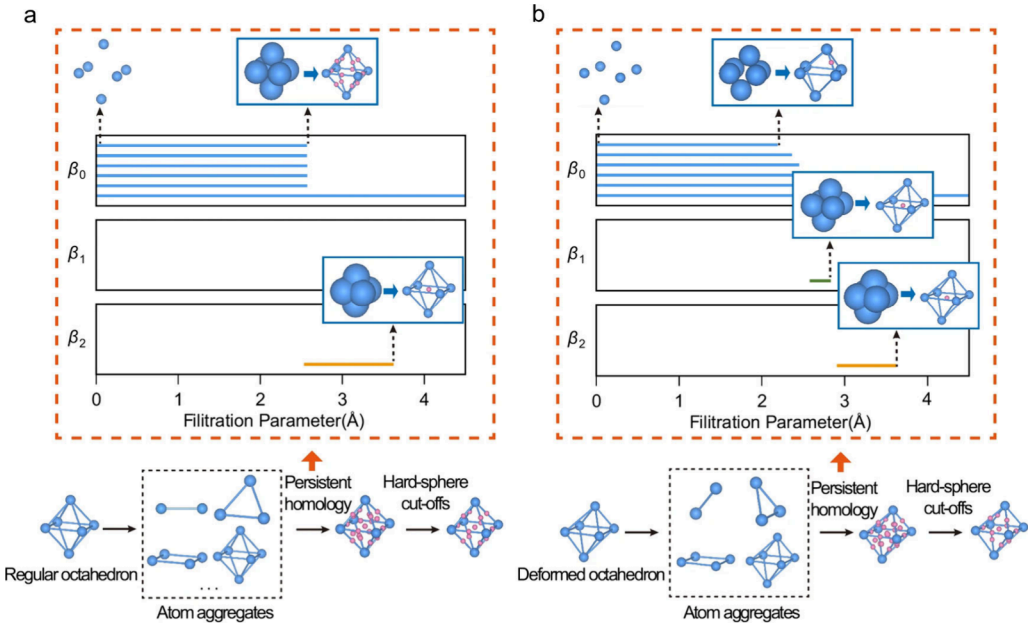


Figure 17. Illustration of persistent homology-based sampling algorithm (PH−SA). The determination of possible interactions sites by PH−SA with (a) regular octahedron and (b) deformed octahedron unit structure as examples. The blue balls are the atoms and red balls indicate the possible interaction sites. The barcodes of β_0, β_1 and β_2 record the persistent homology process. Each bar of β_0 corresponds to an isolated atom (point). The β_1 corresponds to the existence of loop. The β_2 corresponds to the existence of cavity. Reproduced from ref 178. Available under a CC-BY 4.0 license. Copyright 2025 Author(s).

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

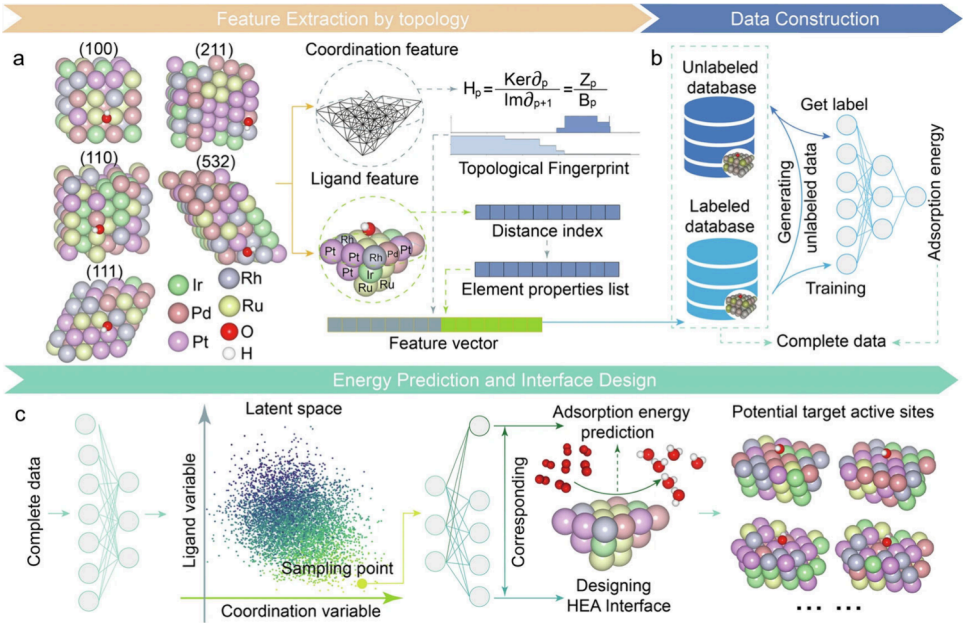


Figure 18. Overview of feature extraction, data set construction, and workflow for energy prediction and interface design. (a) Schematic illustration of feature construction. Coordination features are extracted using persistent GLMY homology, while ligand features are represented based on element properties. (b) Schematic of data set construction using DFT and semisupervised learning. A GBR model was initially trained on DFT−calculated adsorption energies and subsequently employed to predict adsorption energies for additional simulated active sites, enabling the construction of an expanded, pseudolabeled data set for model training. (c) Framework of PGH−VAEs, including modules for encoding, latent space visualization, latent space sampling, and decoding to generate potential active sites. Reproduced from ref 182. Available under a CC-BY 4.0 license. Copyright 2025 The Author(s).

ing.[172,173] Under practical conditions, adsorbates are invariably present at the interface and even subsurface with finite coverage, giving rise to coverage effects that can significantly influence the catalytic reaction mechanism.[174−177] Identifying stable adsorption geometries in certain coverage remains challenging due to the multitude of possible surface sites and adsorbate arrangements. To address this, a persistent homology-based automatic active phase exploration framework was proposed, combining persistent homology analysis with MLIPs for efficient configuration sampling.[178] The core algorithm, PH−SA, identifies potential adsorption or embedding sites of catalyst structure by tracking the spatial locations at which topological features vanish across filtration scales (Figure 17). These points, being approximately equidistant from surrounding atoms, minimize repulsion and favor energetically stable bonding with adsorbates. PH−SA enables unbiased exploration of surface, subsurface, and bulk interactions without reliance on predefined morphologies, making it applicable to both crystalline and amorphous systems. The effectiveness of this framework was demonstrated in two systems: hydrogen absorption in Pd, where hydrogen penetrates subsurface layers and the bulk, inducing a "hex" reconstruction critical for CO_2 electroreduction, explored through 50,000 sampled configurations; and the oxidation dynamics of Pt clusters, where oxygen incorporation renders the clusters less active during oxygen reduction reactions, investigated through 100,000 sampled configurations. In both cases, the predicted active phases and their impacts on catalytic mechanisms were found to be consistent with previous experimental observations. In this case, TDA enables the identification of geometrically and topologically favorable adsorption or embedding sites for adsorbates, which determine the formation of specific active phases and consequently influence the reaction pathways and catalytic mechanisms.

In heterogeneous catalysis, accurate prediction of adsorption energies is crucial following the enumeration of adsorption configurations. While density functional theory (DFT) enables high-throughput calculations, the growing scale of data-driven catalysis has motivated the adoption of ML for more efficient predictions.[179,180] A key challenge in ML models is the effective representation of active site geometry, particularly when three-dimensional spatial variations influence reactivity. A persistent homology−based curvature descriptor was introduced to characterize subtle geometric differences in metal−nitrogen−carbon single-atom catalysts, where curvature modulation plays a critical role but does not affect bonding connectivity. Conventional graph neural networks like CGCNN struggled with such purely spatial variations.[126] To overcome this, PH−CGCNN was developed, which integrates persistent homology into the CGCNN framework.[181] This model improved adsorption energy predictions for nitric oxide reduction intermediates by nearly 30% and enabled the identification of promising catalyst candidates, providing a powerful pathway for curvature effect exploration. Further, a topology−based representation using persistent GLMY homology was developed for the inverse design of active sites in high-entropy alloys for the oxygen reduction reaction[182] (Figure 18a−18c). Unlike conventional coordination number descriptors, this method captured spatial information and showed a strong linear correlation with average *OH adsorption energies across different site types. With this feature, a persistent GLMY homology−guided variational autoencoder framework (PGH−VAEs) was proposed, enabling interpretable inverse design of catalytic active sites. The model elucidated how coordination and ligand

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

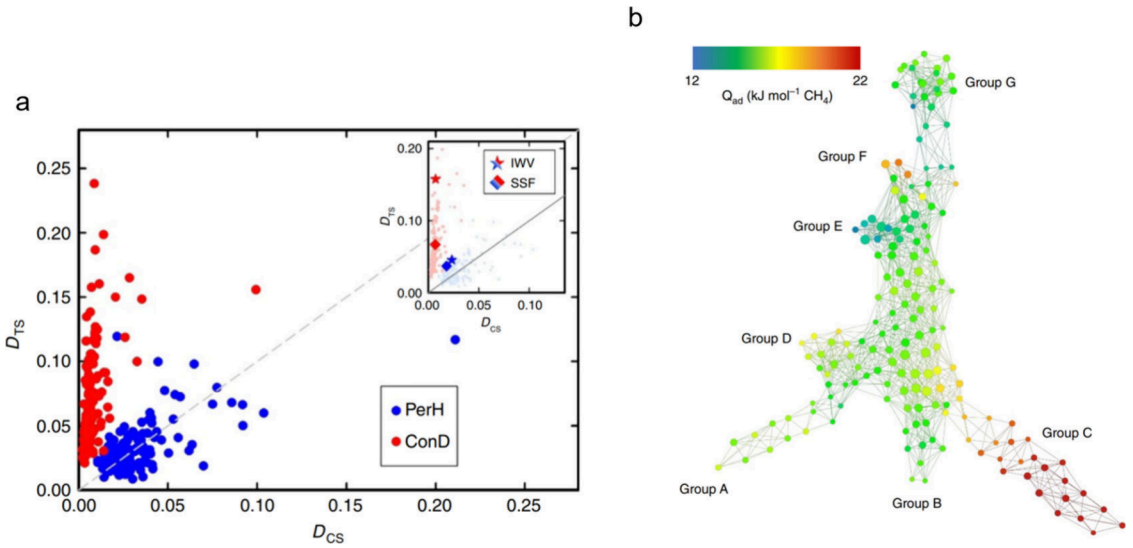


Figure 19. (a) For each known zeolite, the two sets of four most similar structures, once selected using the TDA descriptor (PerH, one blue dot for a set of four) and once selected by the conventional descriptor (ConD, one red dot per set) are compared. This is done by plotting their average distances D_{CS} in conventional space (x−axis) and their average distance D_{TS} in the barcode space (y axis) to the reference zeolite. The distances are normalized by the largest pairwise distance in the database. (b) Mapper plot of the best zeolites (top 1%) for methane storage. Nodes in the network represent clusters of materials with similar pore shapes, and edges connect nodes that contain structures in common. Each material is represented by a combination of its Persistent barcodes, and the metric in this space is D_{TS}. Nodes are colored by the average value of the heats of adsorption of the materials in a cluster (Red: high value, Blue: low value). Adapted from ref 18. Available under a CC-BY 4.0 license. Copyright 2017 Author(s).

effects, including those from distant, noncontacting atoms, governed *OH adsorption and provided targeted strategies for catalyst optimization.

Persistent homology also emerges as a promising tool for structural analysis in heterogeneous catalysis, particularly in systems exhibiting amorphous and compositionally complex atomic arrangements. Kumara et al. applied this method to Ir_xCu_{1-x} alloy nanoparticles, revealing pronounced topological disorder through persistent diagrams with broad distributions but no long-lived features, indicating a lack of extended networks or cavities and only transient local rings.[183] This contrasted sharply with ideal crystalline models and reflected atomic-scale structural perturbations. While not directly measuring catalytic activity, persistent homology complemented other metrics, such as bond orientational order and coordination number, in highlighting disordered, low-coordination surface environments. These structural features were correlated with enhanced OER activity, notably in $Ir_{0.9}Cu_{0.1}$.

4.5. Porous Crystalline Materials

Porous crystalline materials are a class of solids characterized by a periodic atomic structure and an interconnected network of pores, typically with dimensions below 100 nm. They exhibit exceptionally high surface areas, well−defined and tunable pore architectures, and customizable chemical functionalities.[184] Representative classes include zeolites, metal−organic frameworks (MOFs) and covalent organic frameworks (COFs). These materials have emerged as a versatile platform for a wide array of advanced applications, ranging from gas separation and storage, catalysis, and molecular recognition, to drug delivery, sensing, energy storage, and environmental remediation.[185−187] The modularity and flexibility of their chemistry enable the synthesis of an essentially unlimited number of new structures. This vast design space offers unparalleled opportunities to engineer materials with finely tuned properties.[188−190]

The gas adsorption performance of zeolites and MOFs is critically governed by pore geometry, yet a general and precise mathematical framework for quantifying shape similarity has been lacking. Lee et al. addressed this by applying persistent homology to compute H_0, H_1, and H_2 Betti numbers, which are encoded as persistence barcodes, transformed into persistence landscapes, and compared via a weighted L^2 distance across dimensions to define a novel similarity metric—D_{TS}.[18] Compared with conventional geometric descriptors (D_{CS}), the topological fingerprint method can more accurately identify nanoporous materials that exhibit similarity in both geometry and pore shape. (Figure 19a). Crucially, it identifies materials with similar pore shapes regardless of differences in crystal structure or composition. Among top−performing CH_4 adsorbents, pore morphologies are topologically diverse, each class associated with a distinct optimal adsorption enthalpy, challenging the long−held assumption of a universal optimum and underscoring the need for structure-specific optimization strategies (Figure 19b). They further used persistent homology to construct geometric landscapes of crystal structures derived from single organic building blocks, enabling a high−resolution topological clustering of thousands of predicted crystal polymorphs. By mapping these into energy−structure−function space and integrating kernel ridge regression models, the study not only captured geometrical similarity but also enabled accurate prediction of methane deliverable capacities using minimal training data.[191]

Wei et al. further expanded the above framework by introducing category−specific topological learning (CSTL).[192] This method not only quantifies the pore morphology and topological architecture of MOFs but also embeds chemical information directly into the topological space by grouping all elements into eight categories according to their valence-electron configurations and chemical function-

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

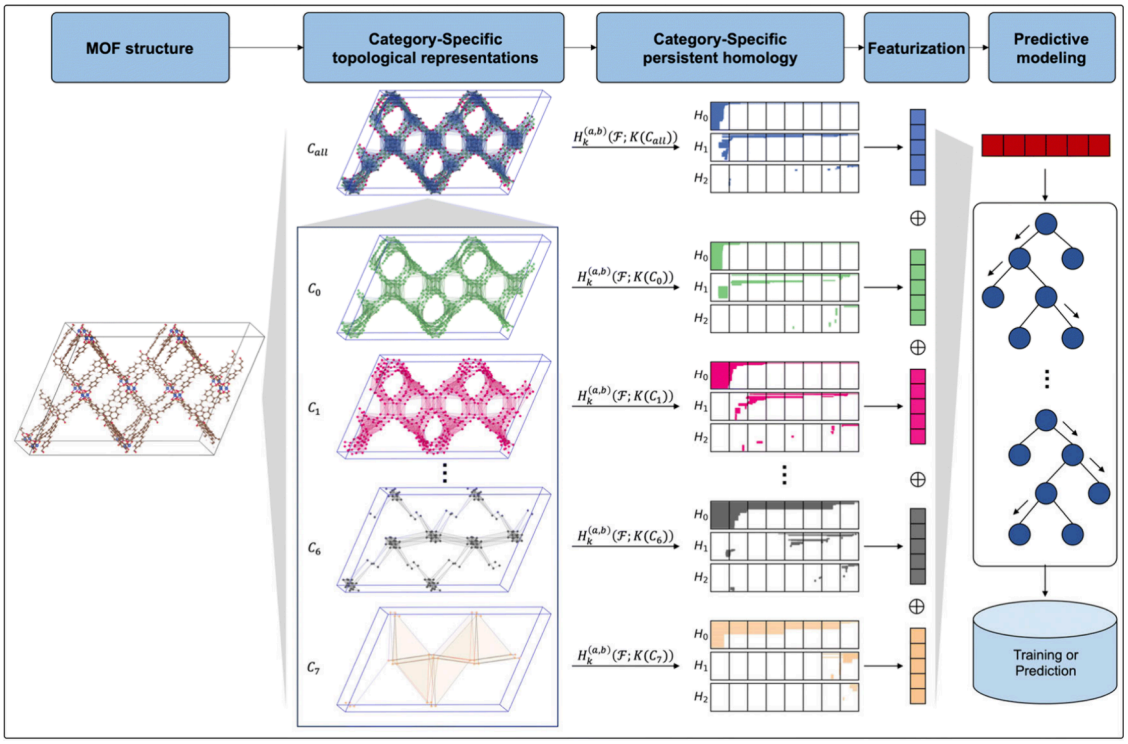


Figure 20. Overview of the category-specific topological models used for predicting properties of MOFs. Given MOF structures (first column), category-specific topological representations are constructed, including a simplicial complex for all structures (Call) and subcomplexes based on categories (C0 to C7). The persistent homology method is applied to each category to generate barcode representations. A featurization vector extracts features from these barcodes, which are then used to construct a gradient boosting tree model for predictions on specific data sets. Reproduced from ref 192. Available under a CC-BY-NC 3.0 license. Copyright 2025 The Author(s).

alities (Figure 20). By constructing simplicial complexes for each element category and computing their corresponding H_0/H_1/H_2 persistent homology features, CSTL simultaneously captures multiscale geometric−chemical characteristics, including local coordination environments, ring motifs, and cavity distributions, thereby generating a physically meaningful, chemically resolved, high-dimensional topological descriptor. When combined with a GBR model, CSTL not only surpasses state-of-the-art deep learning approaches across multiple MOF property data sets (e.g., Henry coefficients, adsorption capacities, and diffusivities), but also provides interpretability through feature-importance analyses, elucidating the roles of different element categories and topological dimensions in structure−property relationships. Recently, they further developed an interaction topological transformer for multiscale learning in porous materials, achieving state−of−the art, accurate, and transferable predictions for adsorption, transport, and stability properties.[193] Beyond serving as an independent tool for directly characterizing MOF structures, persistent homology can also be integrated with conventional geometric descriptors to enhance their structural representational power.[194,195] Furthermore, it can be combined with techniques such as word embeddings to enable unified multimodal feature representations, thereby improving the accuracy of MOF adsorption-performance predictions.[196]

4.6. Superionic Conductors

Solid−state lithium−ion batteries (SSLBs) have garnered significant attention in recent years due to their numerous potential advantages in terms of safety, durability, energy density, and structural design flexibility.[197,198] These attributes position SSLBs as promising candidates for critical applications such as electric vehicles and smart grid technologies. To realize such systems, a variety of highly ion−conductive solid electrolytes, collectively referred to as superionic conductors, have been developed.[199−201] These materials often exhibit intricate migration behaviors, and a fundamental understanding of their transport mechanisms is essential for the rational design of next−generation superionic conductors. At its core, ionic transport in solid−state materials can be regarded as the trajectory of mobile ions traversing a stationary crystal framework, where structural features such as voids and channels play a crucial role in facilitating or hindering ion mobility.[127,202,203] In this context, TDA may be capable of capturing the topological characteristics of loops and voids within complex structures, holds significant promise for elucidating ion migration pathways, identifying effective diffusion networks, and enabling data-driven materials discovery.

Sato et al. were the first to apply persistent homology to the study of superionic conductors, using α-AgI as a model system to investigate its complex Ag^+ ion migration mechanisms.[21] Traditional methods struggle to reveal cooperative migration and local structural dynamics, prompting the use of persistence diagrams derived from MD. By expanding equal-radius spheres around atomic positions and tracking the formation and disappearance of topological loops, the diagrams representing the birth and death of ring structures can be constructed (Figure 21a). These diagrams capture the evolution of local atomic configurations across multiple scales. Analysis at elevated temperatures showed that Ag^+ migration predominantly occurs through four-membered rings composed of two Ag and two I atoms, with ∼ 39% involving cooperative Ag^+

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

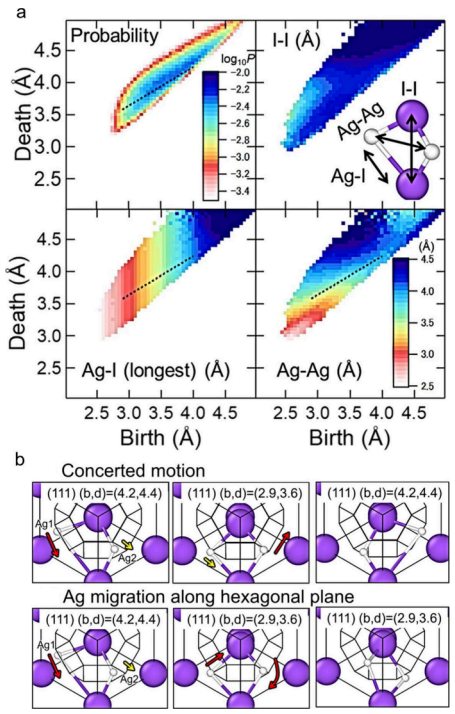


Figure 21. (a) Persistence diagram only showing four-membered rings and the distribution of averaged I−I, Ag−I, and Ag−Ag distances in each (b, d) for these four-membered rings in the persistence diagram. (b) The illustration of cooperative Ag^+ motion with two different modes. Reproduced with permission from ref 21. Copyright 2023 AIP Publishing.

motion (Figure 21b). The deformation energy of these rings (~0.12 eV) closely matches the activation energy for conduction. Compared to random migration models, persistent homology more effectively reveals how ion−ion interactions influence migration pathways. Jalem et al. employed persistent homology to investigate lithium-ion diffusion pathways at grain boundaries and amorphous/crystalline interfaces in β−Li_3PS_4 solid electrolytes.[204] By analyzing MD−derived Li^+ trajectories, they quantified the topology of migration channels using f_{PD}—the fraction of persistent loops with lifetimes exceeding 1.0 Å. The study found a strong positive correlation between f_{PD}, structural disorder, and room-temperature ionic conductivity, highlighting topological richness as a key factor for fast ion transport. Particularly, twisted grain boundaries and amorphous/crystalline interfaces showed higher f_{PD} values and superior conductivity, rivaling or surpassing those of fully amorphous regions.

The topological features revealed by persistent homology not only enable direct analysis of superionic conductor structures to investigate ion migration mechanisms but can also serve as effective criteria for screening new superionic conductor materials. Chen et al. proposed an innovative multiscale topological learning framework that integrates algebraic topology with unsupervised learning to discovery lithium solid ionic conductor (LSIC) (Figure 22).[205] By decomposing crystal structures into lithium-only (Li−only) and lithium−free (Li−free) substructures, the model extracts two key topological descriptors: the minimum connectivity distance ($r_{connected}$) and the cycle density (ρ_{cycles}), which evaluate the continuity of lithium migration pathways and the geometric suitability of the surrounding framework, respec-

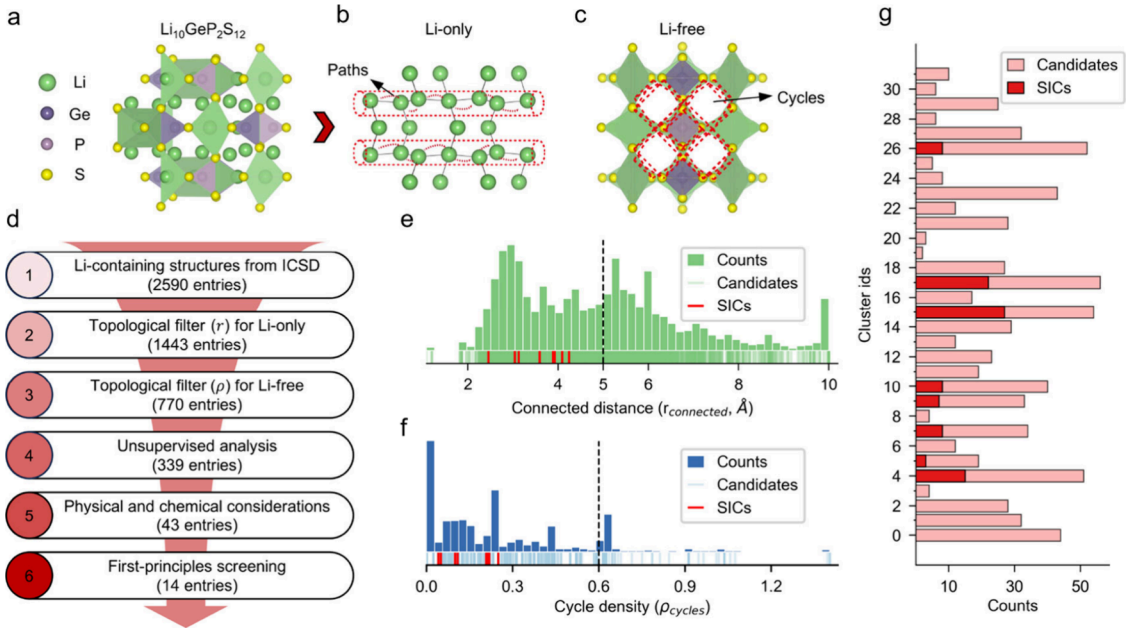


Figure 22. (a) Crystal structure of the solid ionic conductor (LSIC) $Li_{10}GeP_2S_{12}$, used as an example. (b) Li−only substructure extracted from the LSIC. (c) Li−free substructure is derived from the same material. (d) Overview of the materials discovery workflow showing six stages with the corresponding number of materials filtered at each stage. (e) Distribution of the minimum connectivity distances ($r_{connected}$) for Li−only substructures. The red lines in the rug plot highlight the known LSICs, and the dashed line marks the threshold of $r_{connected}$ = 5 Å used in the filtering process. (f) Distribution of the pore occupancy index (ρ_{cycles}) for Li−free substructures. The red lines in the rug plot indicate known LSICs, and the dashed line denotes the filtering threshold ρ_{cycles} = 0.6. (g) Results from the affinity propagation algorithm. The lighter red bars represent all materials retained after topological prefiltering, while the darker red bars indicate the known LSICs. The horizontal axis corresponds to the number of structures in each cluster. Reproduced from ref 205. Copyright 2025 American Chemical Society.

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

tively (Figure 22b and 22c). Specifically, $r_{connected}$ quantifies the minimal spatial threshold required to form a fully connected lithium network within the Li−only substructure, where lower values indicate tighter connectivity and facilitate concerted ion migration, while ρ_{cycles} assesses the density of topological cycles in the Li−free substructure, balancing structural openness with framework stability. The integration of these two metrics ensures that screened materials not only possess efficient ion diffusion channels but also maintain structural robustness. Through persistent homology, the authors extract these features across multiple spatial scales and apply empirical thresholds ($r_{connected}$ < 5 Å, ρ_{cycles} < 0.6) to filter 2590 lithium−containing compounds, narrowing the candidate pool to 1443 materials. From these, 22-dimensional multiscale topological features were derived from persistence barcodes and used as input for Affinity Propagation clustering, resulting in 339 candidates topologically similar to known LSICs. Subsequent chemical screening and ab initio MD simulations further refined the selection, ultimately leading to the identification of 14 novel LSICs (Figure 22d-22g). Among these, four materials, including Li_7SbO_6, $Li_{10}Zn_4O_9$, $LiAlSiO_4$, and $NaLi_2PO_4$, have been experimentally or previously validated and exhibit room−temperature ionic conductivities on the order of 10^{-2} S/cm and diffusion activation energies between 0.2 and 0.9 eV. This work presents a highly efficient and scalable paradigm for the discovery of solid−state electrolytes, significantly reducing computational costs and enabling data−driven design of functional materials.

4.7. Applications in Additional Material Systems

TDA has established a diverse set of applications and methodological frameworks in the aforementioned material systems. In several other domains, TDA is also beginning to exhibit methodological potential, although current studies remain sporadic and have not yet matured into comprehensive or systematic application paradigms.[206−214] For example, topological learning has enabled automated and high−precision classification and regression of topological states from high−dimensional phase−field simulation data of oxide superlattices, allowing strain−electric field phase diagrams to be generated within seconds from high−throughput data sets.[206] In another application, topological learning has been used to construct microstructure−property linkages in dual−phase steels, where persistent images serve as effective microstructural descriptors and artificial neural networks successfully predict the corresponding stress−strain curves.[207] Persistent homology has been employed to quantitatively characterize the topological ordering and many−body structural correlations in two−dimensional materials, revealing structural features inaccessible to traditional two−body metrics such as structure factors.[208] The persistent homology can be used to characterize time−dependent structural evolution in complex soft−matter systems, as demonstrated in a nematic liquid crystal−gold nanoparticle composite undergoing phase transitions. Time-resolved microscopy images are converted into persistence diagrams that encode mesoscopic structural features, from which structural heterogeneity and pairwise diagram distances are derived. The temporal sequence of diagrams defines a trajectory in topological space, enabling analysis of evolution pathways. This approach allows identification of key transition stages, coexistence regimes, and multistep dynamics during heating and cooling, highlighting the capability of persistent homology to extract dynamical information beyond static structural characterization.[213]

4.8. Summary of Material Data Types for TDA

In the examples above, we have demonstrated the application of TDA to various material systems. In practice, despite the differences among these systems, the types of data involved share certain similarities. In the following, we summarize the commonly encountered data type in materials science that can be characterized and represented by TDA. To avoid redundancy, we cite only a selection of representative studies here. Additional examples can be found in the preceding sections.

Atomic Coordinates. Atomic coordinates represent the most common form of structural data in materials science, as the majority of material systems can be described in this manner. They can be naturally represented as point clouds, from which simplicial complexes are constructed, followed by the application of persistent homology to extract topological features. These features, including connected components, loops, and voids, can be related to coordination environments, pore structures, and MRO in materials.[19,178,192] In addition to persistent homology, the GLMY homology and Euler characteristic can also be employed to characterize geometric and topological properties of atomic configurations, providing comprehensive description of structural organization across multiple scales.[71,153,182]

MD Trajectories. MD Trajectory data arise from simulations of system evolution at finite temperature and over finite time scales, and can be viewed as a sequence of time−dependent atomic configurations. The most common approach is to represent atomic coordinates at each time step as point clouds, thereby treating MD trajectories as time-dependent point cloud sequences. Based on this representation, TDA can be applied to individual frames or ensembles of configurations to extract topological features.[21,162,164] In addition, field-based representations (e.g., density distributions) can be constructed and analyzed using sublevel-set filtrations along the MD Trajectory.[215,216]

Imaging Data. Material structures are typically characterized using various microscopy techniques in experiments. The resulting imaging data (e.g., transmission electron microscopy) contain spatial information on the material's microstructure. Such data can be represented as pixel-based scalar fields (e.g., intensity or height distributions), enabling the topological analysis. On this basis, TDA can extract key morphological features, including domains, interfaces, and connectivity, thereby providing valuable insights into the relationship between microstructure and material properties.[165,207,217]

Potential energy surfaces. The potential energy surface describes the relationship between atomic configurations and system energy. A wide range of chemical processes, including conformational changes, reaction pathways, phase transitions, and interfacial reconstructions, can be viewed as dynamical evolution on a high−dimensional potential energy surfaces. Its topology directly governs system stability and kinetics: local minima correspond to stable or metastable states, saddle points to transition states, and the connectivity between minima defines possible reaction pathways. Within the framework of TDA, the potential energy surfaces can be formulated as a scalar field over configuration space and analyzed via sublevel or superlevel set filtrations. Persistent homology enables the

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

identification of key topological features across multiple scales, including energy basins, barrier structures, and connectivity between configurational regions, thereby revealing multistability and transition pathways without predefined reaction coordinates. Furthermore, TDA provides a global, multiscale characterization of the potential energy surfaces, facilitating the identification of multichannel reaction mechanisms and rare events that are difficult to capture with conventional approaches.[218,219]

5. EXTENSION OF TDA TO MOLECULAR AND BIOCHEMICAL SYSTEMS

Beyond materials science, TDA has also been widely applied in molecular and biochemical systems. As these systems are likewise governed by atomic−scale chemical interactions, insights from these fields provide valuable guidance for understanding structure−function relationships in materials at the same scale. In this section, we present selected studies from these domains to highlight the cross−disciplinary utility of TDA and, in particular, to discuss how these approaches can inform its further application in materials science. Notably, this review primarily focuses on TDA in materials science. Molecular and biochemical systems are therefore discussed only briefly, and readers are referred to existing reviews for more comprehensive treatments.[30,31,35,220−222]

5.1. Molecular Systems

5.1.1. Molecular Energy Landscapes. TDA can be used to characterize the potential energy landscape and reaction processes of molecules. By analyzing the energy landscape of molecular conformations, TDA can identify different energy minima (representing stable states) and energy maxima (representing energy barriers).[138,218,219] This high−dimensional energy landscape information can be compactly encoded as a barcode, revealing how molecules transition between different conformations, particularly how they overcome energy barriers during chemical reactions.

Such complex energy landscapes are particularly evident in biomolecular folding processes, where conformational transitions often involve multiple intermediate states rather than simple two-state dynamics. Early studies on RNA folding revealed the presence of complex conformational landscapes with multiple metastable states.[223] These characteristics make molecular simulation data well suited for topological analysis, where persistent homology can capture metastable states and transition pathways from high-dimensional conformational data sets. More recently, topology−based frameworks have been combined with machine learning to extract transition pathways and rare events from molecular dynamics trajectories, enabling efficient analysis of large-scale simulation data sets.[224] Beyond conformational dynamics, persistent homology has also been applied to analyze electron density and reduced density gradient fields, revealing topological signatures of chemical bonding and intermolecular interactions.[225]

These developments provide a conceptual framework for materials systems, where energy landscapes are more complex due to collective phenomena Similar to molecular system, materials energy landscapes can also be viewed as high−dimensional manifolds with multiple metastable states, transition pathways, and energy barriers, which can be compactly encoded by topological descriptors. Persistent homology can identify distinct basins corresponding to stable or metastable phases and their connectivity, revealing transition mechanisms beyond conventional reaction−coordinate descriptions. This perspective may enable the systematic characterization of phase transitions, surface reconstructions, and multicomponent interactions.

5.1.2. Molecular Organization. Understanding the collective organization of molecules under specific constraints is a cornerstone of molecular science. A prime example is water, one of the most essential substances in nature, whose complex network architecture remains a perennial challenge in the field.[226,227] Recently, TDA has emerged as a powerful tool for characterizing the evolution of intricate structures and spatial organizational features of water under varying temperatures, pressures, and solute conditions.[228−233] For example, persistent homology was employed to uncover MRO in glassy water, focusing on two canonical phases: low-density (LDA) and high-density (HDA) glassy water.[234] Persistent homology analysis identified looplike atomic structures as the real space origin of MRO. Through persistence diagrams, the authors characterized these rings in terms of size, planarity, and atom count, revealing that LDA rings are generally larger, flatter, and involve fewer atoms than their HDA counterparts. Crucially, these structures accurately reproduced the FSDP in both position and width, bridging the gap between abstract topological features and physical scattering signatures.

Beyond studies of bulk water organization, recent works have extended to interfacial water at material surfaces, whose structural organization directly influences material functionality.[235,236] TDA provides a powerful tool to extract structural features of interfacial water and to elucidate the relationship between its organization and material properties.[216,237,238] A representative case is the use of Euler characteristic to characterize the structural distribution of interfacial water, and, when combined with linear regression models, to successfully predict the hydration free energies of various self-assembled monolayer surfaces, a key descriptor closely related to surface hydrophobicity.[215] Despite these advances, the application of TDA to interfacial water in materials remains at an early stage, with relatively limited studies to date, indicating a highly promising direction for future research.

5.1.3. Molecular Screening. Topology−based molecular fingerprints have also been applied to large-scale molecular screening. For example, persistence image representations have been used to search molecular databases and identify candidate molecules with enhanced CO_2 adsorption properties.[239] Persistence image descriptors combined with machine learning have also been used for high-throughput screening of Fe(IV)−oxo complexes, enabling efficient prediction of C−H activation barriers and spin states.[240] Similar topology−derived descriptors have been applied to materials−related molecular systems such as porous organic cages, where machine learning models can predict structural stability and cavity properties across large molecular databases.[241] Topological features have also been incorporated into deep generative models for molecular design, improving the generation of chemically valid and structurally diverse molecules.[242] These screening approaches effectively treat molecules as functional entities, where their properties are evaluated in a manner analogous to materials discovery. In this sense, topology−based molecular descriptors bridge molecular modeling and materials design by enabling structure−property relationships to be explored across large chemical spaces.

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

5.2. Biochemical Systems

5.2.1. Biochemical Structure and Interactions. TDA has been increasingly extended to biochemical systems, providing a versatile toolkit to investigate structural, dynamical, and functional properties of proteins and their complexes. At the structural level, persistent homology and its extensions have been widely used to generate protein topological fingerprints for characterizing folding stability, conformational flexibility, and protein−protein or protein−ligand binding interfaces, and further facilitating ligand and drug design.[243−247]

Several representative studies further demonstrate how topology−based descriptors can be applied to analyze protein structures and their functional variations. For example, topology−based representations have been developed for biomolecular complexes, including multicomponent and multilevel persistent homology, which preserve chemical information while enabling accurate prediction of protein−ligand binding affinity.[248] Persistent homology has also been used to characterize protein conformational changes through persistence landscapes, allowing statistical comparison of different functional states and identification of topological features associated with ligand binding.[249] In addition, topology−based mutation predictors have been proposed to estimate protein stability changes upon mutation, where element−specific persistent homology encodes structural and biochemical information for machine learning models.[250] Topological descriptors have further been used as machine learning features for biomolecular modeling. For example, persistence image representations combined with conventional cheminformatics descriptors have demonstrated improved performance in predicting protein−ligand interactions and bioactivity.[251]

Beyond static structural analysis, TDA has also been integrated into data−driven strategies for rapid response to viral evolution and drug discovery. By combining AlphaFold−assisted models with multi task topological Laplacians, researchers have achieved accurate prediction of mutational impacts on protein−protein interactions, thereby supporting fast−tracked antibody and vaccine design.[252] Similarly, multiscale topology−based frameworks have facilitated drug repurposing for complex diseases such as addiction, bridging transcriptomic signals, protein−protein interaction networks, and pharmacological screening.[253] In addition, topology−based deep learning frameworks have been developed to model complex drug−target interactions, where multiscale topological Laplacians are integrated with neural network architectures.[254] Persistent homology has also been explored for ligand−based virtual screening, where topology−derived similarity metrics complement traditional cheminformatics fingerprints in identifying potential drug candidates.[255] Importantly, the underlying methodological principles of these studies are highly transferable to materials science. Element−specific and multiscale topological representations provide a natural way to encode chemically heterogeneous environments, enabling materials structures to be decomposed into element− or coordination−resolved substructures. In solid−state electrolytes, such representations can capture ion migration pathways and framework connectivity, providing topology−derived descriptors related to diffusion channels and collective transport behavior. Similar formulations can also be extended to heterogeneous catalysis to describe adsorption configurations within a unified topology−aware framework.

5.2.2. Biological Networks. Furthermore, statistical physics models combined with GLMY homology offer a rigorous means to capture higher order topological alterations in metabolic and hypernetwork models, providing novel insights into disease-associated dysbiosis and system-level interactions.[256−258] These approaches demonstrate that GLMY homology, built on path and hypergraph representations, enables the explicit characterization of directional reaction flows, asymmetric interactions, and higher order cooperative relationships that commonly arise in complex biochemical. Research on high order interactions in biochemical networks provides important insights for understanding complex network behavior in materials science. In catalytic systems, for example, reaction networks often involve multispecies cooperative effects that go beyond simple pairwise interactions. By leveraging GLMY homology, reaction networks can be characterized from a topological perspective, moving beyond traditional single pathway−based analyses to enable a global topological representation of the reaction space. This approach may facilitate the identification of key high−order functional motifs governing reaction selectivity, as well as topological bottlenecks that limit reaction flux. Such a framework holds significant potential for elucidating complex reaction mechanisms and guiding materials design, particularly in multistep, competing processes such as the CO_2 electroreduction to multi−carbon products.

6. DISCUSSION AND OUTLOOK

With the rapid advancement of computational capabilities and machine learning, materials science is increasingly adopting data−driven paradigms. Central to this transformation is the digital characterization and representation of material data, where feature extraction plays a crucial role in elucidating structure−property relationships and enabling machine learning−driven tasks such as property prediction and inverse design. Throughout this review, we have introduced TDA as a powerful topology−based framework for extracting structural information from complex materials data and discussed its applications in structural characterization, materials modeling, and topological learning. In this section, we synthesize these developments by outlining the key advantages of TDA in materials science, together with the current challenges that limit its broader adoption. By highlighting both the strengths and open problems of TDA−based approaches, we aim to provide perspectives that may guide future methodological developments and inspire new research directions in topology−driven materials science.

TDA is an algebraic topology−based method for extracting and analyzing information from data, particularly excelling at capturing the shape of the data. In materials science, TDA is inherently suited because it allows for the mapping of simplices of various dimensions to atomic clusters at different scales. For example, the β_0 reflects isolated atoms, and β_1 and β_2 capturing higher-order many-body interactions between atoms.[129] These capabilities make TDA uniquely capable of revealing intricate structural features that are often challenging to capture using traditional methods. The following outlines its distinct advantages and applicability in materials science.

First, TDA achieves scale robustness through the use of filtration, enabling the tracking of topological features across multiple geometric scales and thus revealing hierarchical structural organization without reliance on a specific resolution. This contrasts sharply with traditional fixed−cutoff

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

approaches. For instance, conventional graph−based methods require a predefined cutoff radius to determine atomic connectivity,[259−261] which may lead to different graph structures and analysis outcomes with varying cutoffs. In contrast, TDA systematically summarizes connectivity patterns and voids over the entire range of the filtration parameter (such as distance), yielding hierarchical and global topological invariants. These invariants capture complex geometric correlations beyond local two− or three−body interactions and their evolution with scale, allowing for a more comprehensive and accurate characterization of multiscale structural features in materials.[23,129,130,144]

Second, TDA is model−agnostic, meaning it does not rely on specific physical models or material types. Instead, it analyzes the structure of data using mathematical tools from topological algebra to reveal the intrinsic topological features of the data. Therefore, in principle, TDA can effectively analyze any material system, including crystalline, amorphous, composite, and even complex multiphase systems. In fact, TDA shows greater advantages when dealing with complex systems. Traditional structural analysis methods, such as those based on geometric descriptors or local order, often fail to capture high−dimensional, multiscale topological features, particularly in materials with long-range disorder or strong many-body interactions. For example, in glassy materials, traditional edge length and angle analyses struggle to distinguish the structural organization of glasses with different cooling rates, whereas TDA−based classification methods are able to achieve it.[164] Interestingly, with appropriate processing, TDA−derived results can also yield information analogous to traditional structural descriptors, such as radial distribution functions,[233] highlighting TDA's potential as a unified framework to characterize both local and global structures in material systems.

Third, TDA exhibits inherent robustness to noise. Persistent homology, the representative tool of TDA, focuses on tracking the "persistence" of topological features across filtration scales, that is, how long such features survive as the parameter varies. Features that appear only briefly, typically arising from thermal fluctuations, experimental errors, or numerical noise, are distinguished from physically meaningful structural motifs that persist over a wide range of scales. By quantifying persistence, TDA automatically suppresses spurious signals induced by noise and retains only stable, significant features. This noise robustness makes TDA especially suitable for analyzing data from molecular dynamics simulations, Monte Carlo sampling, or low−resolution experimental measurements, where transient fluctuations and sampling errors are common.[146,164,165,231] Traditional statistical or graph−based descriptors often struggle to separate genuine structure from noise, whereas TDA's persistence analysis enables clear discrimination, thereby improving the accuracy, stability, and interpretability of structural analysis and machine learning models.

Fourth, TDA is capable of characterizing multi−modal data in materials science. In materials science, data comes in various forms. One of the most fundamental is point cloud data representing material structures, which is derived from the atomic arrangement of materials. This data can be directly used to establish structure−property relationships and is currently one of the most widely used representations in the field. However, other types of data also exist in materials science, such as charge density, potential energy surfaces, images, which are manifold data.[26,138,165] Compared to structural data, these data carry deeper physical and chemical significance, directly reflecting certain key properties of materials. For instance, the distribution and interaction of charges can reveal bonding information. However, such data are often high−dimensional and redundant, and there are still few effective methods for characterizing and analyzing them. TDA, which focuses on extracting the intrinsic topological features of data, is well−suited to extracting key information from these complex, high−dimensional data sets, thus providing a foundation for their application in machine learning.[26,138]

Lastly, TDA yields descriptors (e.g., persistence images and persistence diagrams) that are inherently invariant to translation, rotation, and permutation.[137] These descriptors compactly encode complex, multi−scale, high−dimensional structural information into low−dimensional yet information−rich feature vectors, preserving the essential topological and geometric signals that dominate macroscopic properties. This ability renders TDA−derived features suitable as fingerprints for materials machine learning. More crucially, TDA confers intrinsic physical interpretability to machine learning models. Because TDA directly tracks the birth and death of topological features in the filtration process, each high weight element in a persistence barcodes or persistence images can be traced back to specific geometric/topological events (e.g., multiatom loops, caged voids, or interconnected channel networks), which effectively alleviates the black-box issue of the resulting models. For example, when predicting high thermal conductivity in an amorphous material, one can pinpoint exactly which persistent loop−like features contribute most;[164] When screening MOFs for adsorption, the dominant factors, three−dimensional cavities versus one−dimensional channel networks, can be interpreted.[192] This rigorous traceability from prediction to mechanistic understanding constitutes the core distinction of TDA−driven materials machine learning compared with conventional machine learning. In an era that increasingly demands interpretable and physics−guided machine learning, TDA stands as one of the most elegant bridges linking rigorous mathematics, microscopic structure, and macroscopic properties.

TDA exhibits distinctive potential for applications in material systems, yet its broader development remains constrained by several critical challenges. TDA excels at capturing geometric and topological features in structural modeling, but it often overlooks critical chemical information, such as atomic types, bond properties, or electronegativity differences. Consequently, systems with distinct chemical properties but similar topological structures may not be well distinguished in predictive models, posing a key bottleneck for topological learning. Three main strategies are developed to address this limitation: First, chemical information can be incorporated directly into the filtration process, for example, by constructing weighted simplicial complexes,[262] assigning initial radius to point sets,[160,234] or applying filtration strategies that are tailored to individual elements.[192] These approaches aim to ensure that the resulting persistence images or persistence barcodes possess sufficient chemical discriminability, but systematic design and generalization of such methods remain challenging. Second, structural topological features and chemical descriptors can be extracted separately and then integrated within a machine learning framework, enabling their complementary information to be fused in the latent space of deep learning models.[182] This strategy involves the develop-

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

ment of multi−branch neural network architectures and the design of specific loss functions for latent space alignment, facilitating a more comprehensive structure−property mapping. Third, continuous scalar fields reflecting chemical properties, e.g., electron density, can be treated as data objects and analyzed via sublevel set filtrations in TDA to extract topological invariants.[26] While this method can, in principle, directly capture chemical characteristics, it also faces practical challenges such as high data dimensionality and strong sensitivity to grid resolution, and the difficulty of correlating topological features with explicit quantum observables.

Further, interpreting TDA results in materials systems poses significant interdisciplinary challenges. While topological invariants such as loops or voids provide clear geometric interpretability, their functional significance in specific chemical contexts remains far from established. For example, a persistent one−dimensional homology class often corresponds to a molecular loop structure. In predictive modeling, such loops may exhibit strong correlations with properties like reactivity, binding sites, or thermal stability. However, the mechanisms by which these topological features influence macroscopic properties, potentially mediated by electronic structure, remain poorly understood, and there is a lack of systematic theoretical or methodological frameworks to bridge this gap. How electronic structure and material properties are represented and transmitted through topological invariants requires further investigation and methodological advancement for modeling high-level structure−function relationships.

Topological learning in the materials domain is still in the early stages of exploration. Now, most works remain focused on explicit topological representations extracted by TDA and using them as input features for traditional machine learning models or neural networks. Topological features can independently achieve excellent predictive performance.[23,24] While in certain cases, topological features alone are insufficient, and optimal performance is achieved only when combined with classical chemical descriptors.[194,195] This suggests that topology−based and traditional descriptors capture complementary aspects of the underlying systems. Therefore, an important future direction is the development of hybrid descriptors that integrate TDA−derived features with complementary global or local structural descriptors, enabling richer and more informative representations for machine learning workflows across different tasks and material systems. Furthermore, the TDA feature extraction heavily relies on manually designed filtration functions, such as element weighting,[262] localized,[23] or sub/superlevel set persistent homology,[216] most driven by chemical insight and inspiration of researchers, severely limits the automation of TDA in large scale scenarios such as high throughput virtual screening and massive database mining. Consequently, there is an urgent need to develop a task adaptive topological representation framework that can automatically sense the prediction target and material system characteristics, dynamically optimize filtration strategies, feature extraction pathways, and vector representation methods during end−to−end training. Additionally, topological learning approaches based on implicit topological regularization and topology-aware architectures remain nearly unexplored in the field of materials science. Unlike conventional approaches that merely use TDA−derived features as input, these two methodologies embed TDA directly into the neural network architecture itself. This integration paves the way for the emergence of entirely new architectures, fundamentally different from current mainstream frameworks such as graph neural networks and thus represents a highly promising direction for future research.

Applying TDA for post hoc interpretation of structure−property relationships in existing systems is highly effective. However, directly performing material inverse design based on topological features remains a challenging and emerging problem. The main difficulty lies in the fact that topological representations are typically compressed mappings of the original structures, leading to information loss and non−uniqueness when reconstructing atomic configurations from topological space. One potential route is to integrate TDA with generative models. Topological descriptors, such as persistence diagrams or their vectorized representations, can be embedded into the latent space of deep generative models (e.g., variational autoencoders or diffusion models), thereby establishing a bidirectional mapping between topological features and structural configurations. By specifying target topological features, such as desired pore structures, connectivity, or multiscale topological patterns, these models can be guided to generate candidate structures satisfying the imposed topological constraints. Another approach is to incorporate topological features directly into structure-search algorithms as constraints or regularization terms, allowing the structural evolution process to be steered toward target topological characteristics.

Currently, the application of TDA in materials science is still dominated by persistent homology, which has proven effective in capturing multiscale topological features of complex structures. In contrast, the use of GLMY homology and the Euler characteristic remains at an early stage in this field, yet both exhibit significant potential for future development. Notably, GLMY homology offers unique advantages in analyzing directional structures and networks, providing richer topological information than conventional homology. It has already found broad applications in molecular and biochemical systems. Given the prevalence of asymmetric and heterogeneous features in material systems, GLMY homology holds promise for applications such as structural characterization, interface analysis, and defect detection, and merits further exploration in materials modeling and data-driven analysis. On the other hand, the Euler characteristic, as a topological summary, is simple in form and computationally efficient. Although it lacks the expressiveness of higher-order homological tools, it may be particularly advantageous in specific scenarios such as high-throughput screening, rapid feature extraction, or preliminary classification. Its low computational cost and clear interpretability make it a promising candidate for feature engineering, model filtering, and fast evaluation in machine learning workflows within materials science.

As a final remark, recent advances in simulation capabilities have enabled the study of increasingly large materials systems, making scalability a critical challenge for TDA, particularly for data sets such as point clouds containing millions of atoms. As system size grows, the computational cost of constructing complexes increases combinatorially, rendering direct application to large-scale systems impractical.[263] To address this issue, several scalable strategies have been developed. One important approach is based on localization and decomposition, where large systems are partitioned into smaller subregions or local environments for analysis.[243] For example, inspired by k−mer−like representations, the system can be divided into fixed-size local structural units (e.g., atomic neighborhoods), for

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

which topological features are computed individually and subsequently aggregated to reconstruct global information.[264] This strategy preserves local structural details while significantly reducing computational cost. Another key direction is multiscale topology, which analyzes systems across multiple spatial resolutions.[265] By combining coarse grained descriptions that capture global connectivity with fine scale analyses of local coordination and structural motifs, this hierarchical framework enables efficient characterization of large systems while revealing structure−property relationships across scales.

In conclusion, TDA offers a profound and broadly applicable methodological framework for characterizing and representing data in materials science, one that remains far from fully explored. The unique strength of TDA lies in its mathematically rigorous and unified approach to systematically extracting and encoding multiscale, complex topological features within material data, thereby opening new avenues for understanding and modeling the fundamental structure−property relationships that underpin material behavior. While challenges remain, TDA has already demonstrated remarkable theoretical value and application potential across the field. As algorithms mature and interdisciplinary collaborations deepen, TDA is poised to accelerate the discovery of structure−function knowledge and the intelligent design of materials, thereby driving transformative advances.

AUTHOR INFORMATION

Corresponding Authors

Yingru Zhao − *College of Energy, State Key Laboratory of Physical Chemistry of Solid Surfaces, iChEM, College of Chemistry and Chemical Engineering, College of Materials, Institute of Artificial Intelligence, College of Physical Science and Technology, Xiamen University, Xiamen 361000, China*; Email: yrzhao@xmu.edu.cn

Jie Wu − *Beijing Institute of Mathematical Sciences and Applications, Beijing 101408, China*; Email: wujie@bimsa.cn

Jian-Feng Li − *College of Energy, State Key Laboratory of Physical Chemistry of Solid Surfaces, iChEM, College of Chemistry and Chemical Engineering, College of Materials, Institute of Artificial Intelligence, College of Physical Science and Technology, Xiamen University, Xiamen 361000, China*; orcid.org/0000-0003-1598-6856; Email: Li@xmu.edu.cn

Shing-Tung Yau − *Beijing Institute of Mathematical Sciences and Applications, Beijing 101408, China; Qiuzhen College, Tsinghua University, Beijing 100084, China; Yau Mathematical Sciences Center, Tsinghua University, Beijing 100191, China*; Email: styau@tsinghua.edu.cn

Feng Pan − *School of Advanced Materials, Peking University, Shenzhen Graduate School, Shenzhen 518055, China*; orcid.org/0000-0002-8216-1339; Email: panfeng@pkusz.edu.cn

Authors

Shisheng Zheng − *College of Energy, State Key Laboratory of Physical Chemistry of Solid Surfaces, iChEM, College of Chemistry and Chemical Engineering, College of Materials, Institute of Artificial Intelligence, College of Physical Science and Technology, Xiamen University, Xiamen 361000, China*; orcid.org/0009-0000-4529-7440

Bingxu Wang − *School of Advanced Materials, Peking University, Shenzhen Graduate School, Shenzhen 518055, China*

Mengmeng Zhang − *Beijing Institute of Mathematical Sciences and Applications, Beijing 101408, China*

Jiawei Tan − *College of Energy, State Key Laboratory of Physical Chemistry of Solid Surfaces, iChEM, College of Chemistry and Chemical Engineering, College of Materials, Institute of Artificial Intelligence, College of Physical Science and Technology, Xiamen University, Xiamen 361000, China*

Hang-Biao Lv − *College of Energy, State Key Laboratory of Physical Chemistry of Solid Surfaces, iChEM, College of Chemistry and Chemical Engineering, College of Materials, Institute of Artificial Intelligence, College of Physical Science and Technology, Xiamen University, Xiamen 361000, China*

Ran Liu − *Beijing Institute of Mathematical Sciences and Applications, Beijing 101408, China; School of Mathematical Sciences, Beihang University, Beijing 100000, China*

Zhong-Zhang Shi − *College of Energy, State Key Laboratory of Physical Chemistry of Solid Surfaces, iChEM, College of Chemistry and Chemical Engineering, College of Materials, Institute of Artificial Intelligence, College of Physical Science and Technology, Xiamen University, Xiamen 361000, China*

Xin Li − *College of Energy, State Key Laboratory of Physical Chemistry of Solid Surfaces, iChEM, College of Chemistry and Chemical Engineering, College of Materials, Institute of Artificial Intelligence, College of Physical Science and Technology, Xiamen University, Xiamen 361000, China*

Ruoxuan Wang − *College of Energy, State Key Laboratory of Physical Chemistry of Solid Surfaces, iChEM, College of Chemistry and Chemical Engineering, College of Materials, Institute of Artificial Intelligence, College of Physical Science and Technology, Xiamen University, Xiamen 361000, China*

Siwang Zhang − *College of Energy, State Key Laboratory of Physical Chemistry of Solid Surfaces, iChEM, College of Chemistry and Chemical Engineering, College of Materials, Institute of Artificial Intelligence, College of Physical Science and Technology, Xiamen University, Xiamen 361000, China*

Jingyan Li − *Beijing Institute of Mathematical Sciences and Applications, Beijing 101408, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chemrev.5c01098

Author Contributions

#Shisheng Zheng and Bingxu Wang contributed equally to this work.

Notes

The authors declare no competing financial interest.

Biographies

Shisheng Zheng received his Ph.D. degree at Peking University in 2023 under the supervisor of Prof. Feng Pan. Presently, he is an assistant professor at Xiamen University. His research interests focus on AI-driven investigation of electrochemical materials.

Bingxu Wang is a Ph.D. student in the School of Advanced Materials at Peking University, Shenzhen Graduate School. He obtained his B.S. degree from the School of Mathematical Science, Jiangsu University, in 2021. He is currently conducting his doctoral research under Prof. Feng Pan at Peking University. His research primarily centers on materials design and property prediction for energy storage

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

applications, using topological data analysis and deep-learning techniques.

Mengmeng Zhang is currently a postdoctoral researcher at the Beijing Institute of Mathematical Sciences and Applications (BIMSA), supervised by Prof. Shing-Tung Yau and Prof. Jie Wu. She received her Ph.D. degree in Mathematics from Hebei Normal University in 2024 under the supervision of Prof. Jie Wu. She has been devoted to exploring the homology and homotopy theory of digraphs, and also seeking further possibilities to apply these theories across various fields.

Jiawei Tan is a Ph.D. student at the College of Energy, Xiamen University (XMU). He obtained his B.S. degree from Northeastern University, China, in 2020 and his M.S. degree from Nanyang Technological University, Singapore, in 2022. He is currently studying under the supervision of Prof. Yingru Zhao at XMU. His research focuses on the design, and optimization of energy systems.

Hang-Biao Lv is a Master student at Xiamen University(XMU). He obtained his B.S. degree from XMU in 2023. He is currently working under the supervision of Assistant Professor Shisheng Zheng and Professor Jian-Feng Li. His research interests mainly focus on the applications of topology in machine learning and prediction.

Ran Liu is a Ph.D. candidate jointly affiliated with Beihang University (BUAA) and Beijing Institute of Mathematical Sciences and Applications (BIMSA). She received her B.S. (2020) and her M.S. (2023) degrees from Hebei Normal University. Her doctoral studies are supervised by Prof. Jie Wu at BIMSA and Prof. Xue Luo at BUAA. Her research focuses on the topological analysis of higher-order interactions in complex systems and the application of GLMY analysis in cancer networks.

Zhong-Zhang Shi is a Master student at Xiamen University (XMU). He obtained his B.S. degree from Xiamen University in 2023. He is currently studying under the supervision of Prof. Jian-Feng Li at XMU. His research focuses on DFT calculations, machine learning, and the exploration of electrocatalysts.

Xin Li is a Ph.D candidate at Xiamen University. He received his bachelor's degree from Hunan University of Arts and Sciences in 2021 and his master's degree from the School of Science at Harbin University of Science and Technology in 2024, under the supervision of Associate Professor Shu Wang. Currently, he is conducting doctoral research under the guidance of Assistant Professor Shisheng Zheng and Professor Jian-Feng Li at Xiamen University. His research focuses on the machine learning model for electrocatalytic interfaces.

Ruoxuan Wang is a Master student at Xiamen University (XMU). She obtained her B.S. degree from Nankai University in 2021. She is currently working under the supervision of Assistant Professor Shisheng Zheng and Professor Jian-Feng Li. Her research interests mainly focus on the applications of machine learning models in predicting electrochemical properties.

Siwang Zhang is a Master student in the School of Chemistry and Chemical Engineering at Xiamen University. He obtained his B.S. degree from Xi'an University of Architecture and Technology in 2023. He is currently studying under the guidance of Assistant Professor Shisheng Zheng and Professor Jian-Feng Li at Xiamen University. His research focuses on theoretical computational studies of the reaction mechanisms in CO_2 electroreduction.

Jingyan Li is an assistant professor at the Beijing Institute of Mathematical Sciences and Applications (BIMSA). She obtained her Ph.D. degree from Hebei Normal University. Her research focuses on the homotopy and homology theories of digraphs, particularly the development and application of GLMY theory. She is actively exploring the application of path homology and related topological methods to interdisciplinary fields such as materials science and computational biology. She has coauthored a substantial body of work, including an upcoming book with Prof. Shing-Tung Yau, and has published extensively in journals spanning pure mathematics, applied topology, and cross-disciplinary sciences.

Yingru Zhao is Professor in Energy Efficiency Engineering at College of Energy, Xiamen University. She holds a PhD in Theoretical Physics and a Bachelor's degree in Physics both from Xiamen University. Between 2008 and 2011 she worked as a Research Associate at Imperial College London. She has comprehensive experience in modelling, simulation and optimization of complex systems, with particular focus on fuel cells, poly generation, supercritical CO_2 power cycle, etc.

Jie Wu earned his Ph.D. in Mathematics from the University of Rochester and subsequently conducted postdoctoral research at the Mathematical Sciences Research Institute (MSRI), University of California, Berkeley. He previously held a tenured professorship in the Department of Mathematics at the National University of Singapore. In December 2021, he joined the Beijing Institute of Mathematical Sciences and Applications (BIMSA) as a professor. His research focuses on algebraic topology and applied topology. His main achievements in algebraic topology are to establish the fundamental relations between homotopy groups and the theory of braids, and the fundamental relations between loop spaces and modular representation theory of symmetric groups. In applied topology, he has achieved several significant results related to topological methods in data science. He received the Singapore National Science Award in 2007 and the Outstanding Senior Talent Award at the International Congress of Basic Science (ICBS) in 2025.

Jian-Feng Li is a full Professor of Chemistry at Xiamen University. He received his BSc degree in Chemistry from Zhejiang University in 2003, and his PhD degree in Chemistry from Xiamen University in 2010. He worked as a postdoctoral researcher at the University of Bern and ETH Zurich in Switzerland from 2011 to 2014. His research interests include core−shell nanomaterials, surface-enhanced Raman spectroscopy, in situ characterization and AI-driven electrochemical research. He has published more than 370 peer-reviewed papers, with a total citation count of over 28000. He has received numerous prestigious awards, including the National Natural Science Award (Second Prize, China), the National Science Fund for Distinguished Young Scholars, the Xplorer Prize, and the Advances in Measurement Science Lectureship Award (ACS).

Shing-Tung Yau, a member of NAS, AAAS, and a foreign academician of CAS, has made significant contributions to differential geometry. In 1976, he proved the Calabi Conjecture. In 1979, Yau and Richard Schoen solved the Positive Mass Conjecture in General Relativity. He currently serves as a chair professor at Tsinghua University, the Director of Yau Mathematical Sciences Center and Qiuzhen College at Tsinghua University and the Dean of Beijing Institute of Mathematical Sciences and Applications. In 1982, he was awarded the Fields Medal. He continued to be recognized via the Veblen Prize in Geometry, MacArthur Fellowship, Crafoord Prize, US National Medal of Science, Wolf Prize, Marcel Grossmann Award and Shaw Prize in Mathematical Sciences.

Feng Pan is a Chair Professor at Peking University, Founding Dean of the School of Advanced Materials, and former Vice President of the Peking University Shenzhen Graduate School. He received his Ph.D. in 1994 from the University of Strathclyde and conducted postdoctoral research at ETH Zurich, following his M.S. from the

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

Chinese Academy of Sciences and B.S. from Peking University. He has long been dedicated to advancing structural chemistry methodology and its application to the research and development of new materials. His contributions include establishing a structural chemistry framework grounded in topology, developing in situ dynamic structural characterization systems based on large-scale scientific facilities such as neutron and synchrotron radiation, and elucidating material genes and structure−function relationships. He has made significant breakthroughs in addressing key scientific challenges in lithium battery research, including energy density, power density, and stability.

■ ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (22402163, 22525042, 52576024, 92472206, 92472203, 22361132532 and T2293692), Natural Science Foundation of Xiamen, China (3502Z202472001), the Fundamental Research Funds for the Central Universities (20720250026) and Fujian Provincial Natural Science Foundation of China (2025I0002). The Liaoning Binhai Laboratory (Grant No. 2024-05), the State Key Laboratory of Fine Chemicals, Dalian University of Technology (KF 2401), Guangdong Key Laboratory of Design and calculation of New Energy Materials (2017B030301013), National Center for International Research of Power Batteries and Materials (2015B01015), Shenzhen Key Laboratory of New Energy Resources Genome Preparation and Testing (ZDSYS201707281026184), the Key Project of Basic Research of the Shenzhen Natural Science Foundation (JCYJ20241202125901003).

■ REFERENCES

(1) Abild-Pedersen, F.; Greeley, J.; Studt, F.; Rossmeisl, J.; Munter, T. R.; Moses, P. G.; Skulason, E.; Bligaard, T.; Norskov, J. K. Scaling properties of adsorption energies for hydrogen-containing molecules on transition-metal surfaces. *Phys. Rev. Lett.* **2007**, *99* (1), 016105.

(2) Calle-Vallejo, F.; Martinez, J. I.; Garcia-Lastra, J. M.; Rossmeisl, J.; Koper, M. T. Physical and chemical nature of the scaling relations between adsorption energies of atoms on metal surfaces. *Phys. Rev. Lett.* **2012**, *108* (11), 116103.

(3) Wei, J.; Chu, X.; Sun, X. Y.; Xu, K.; Deng, H. X.; Chen, J.; Wei, Z.; Lei, M. Machine learning in materials science. *InfoMat* **2019**, *1* (3), 338−358.

(4) Pollice, R.; Dos Passos Gomes, G.; Aldeghi, M.; Hickman, R. J.; Krenn, M.; Lavigne, C.; Lindner-D'Addario, M.; Nigam, A.; Ser, C. T.; Yao, Z.; et al. Data-Driven Strategies for Accelerated Materials Design. *Acc. Chem. Res.* **2021**, *54* (4), 849−860.

(5) Himanen, L.; Geurts, A.; Foster, A. S.; Rinke, P. Data-Driven Materials Science: Status, Challenges, and Perspectives. *Adv. Sci. (Weinh)* **2019**, *6* (21), 1900808.

(6) Bruix, A.; Margraf, J. T.; Andersen, M.; Reuter, K. First-principles-based multiscale modelling of heterogeneous catalysis. *Nature Catalysis* **2019**, *2* (8), 659−670.

(7) Wang, Z.; Sun, Z.; Yin, H.; Liu, X.; Wang, J.; Zhao, H.; Pang, C. H.; Wu, T.; Li, S.; Yin, Z.; et al. Data-Driven Materials Innovation and Applications. *Adv. Mater.* **2022**, *34* (36), No. e2104113.

(8) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559* (7715), 547−555.

(9) Mai, H.; Le, T. C.; Chen, D.; Winkler, D. A.; Caruso, R. A. Machine Learning for Electrocatalyst and Photocatalyst Design and Discovery. *Chem. Rev.* **2022**, *122* (16), 13478−13515.

(10) Li, S. N.; Liu, Y. J.; Chen, D.; Jiang, Y.; Nie, Z. W.; Pan, F. Encoding the atomic structure for machine learning in materials science. *Wires Comput. Mol. Sci.* **2022**, *12* (1), No. e1558.

(11) Gao, C.; Min, X.; Fang, M.; Tao, T.; Zheng, X.; Liu, Y.; Wu, X.; Huang, Z. Innovative Materials Science via Machine Learning. *Adv. Funct. Mater.* **2022**, *32* (1), 2108044.

(12) Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine learning in materials informatics: recent applications and prospects. *npj Computational Materials* **2017**, *3* (1), 54.

(13) Kelley, J. L. *General topology*; Courier Dover Publications, 2017.

(14) Horner, K. E.; Miller, M. A.; Steed, J. W.; Sutcliffe, P. M. Knot theory in modern chemistry. *Chem. Soc. Rev.* **2016**, *45* (23), 6432−6448.

(15) Narang, P.; Garcia, C. A. C.; Felser, C. The topology of electronic band structures. *Nat. Mater.* **2021**, *20* (3), 293−300.

(16) Carlsson, G. Topology and data. *Bulletin of the American Mathematical Society* **2009**, *46* (2), 255−308.

(17) Bilsky, J.; Clark, A. E. Understanding the shape of chemistry data-Applications with persistent homology. *J. Chem. Phys.* **2025**, *163* (9), 090901.

(18) Lee, Y.; Barthel, S. D.; Dlotko, P.; Moosavi, S. M.; Hess, K.; Smit, B. Quantifying similarity of pore-geometry in nanoporous materials. *Nat. Commun.* **2017**, *8*, 15396.

(19) Sorensen, S. S.; Biscio, C. A. N.; Bauchy, M.; Fajstrup, L.; Smedskjaer, M. M. Revealing hidden medium-range order in amorphous materials using topological data analysis. *Sci. Adv.* **2020**, *6* (37), No. eabc2320.

(20) Sørensen, S. S.; Du, T.; Biscio, C. A. N.; Fajstrup, L.; Smedskjaer, M. M. Persistent homology: A tool to understand medium-range order glass structure. *Journal of Non-Crystalline Solids: X* **2022**, *16*, 100123.

(21) Sato, R.; Akagi, K.; Takagi, S.; Sau, K.; Kisu, K.; Li, H.; Orimo, S. I. Topological Data analysis of Ion Migration Mechanism. *J. Chem. Phys.* **2023**, *158* (14), 144116.

(22) Wang, B.; Feng, B.; Lv, L.; Li, S.; Pan, F. Structural Feature Extraction via Topological Data Analysis. *J. Phys. Chem. Lett.* **2025**, *16*, 8056−8067.

(23) Jiang, Y.; Chen, D.; Chen, X.; Li, T.; Wei, G. W.; Pan, F. Topological representations of crystalline compounds for the machine-learning prediction of materials properties. *NPJ. Comput. Mater.* **2021**, *7*, 28.

(24) Chen, X.; Chen, D.; Weng, M.; Jiang, Y.; Wei, G. W.; Pan, F. Topology-Based Machine Learning Strategy for Cluster Structure Prediction. *J. Phys. Chem. Lett.* **2020**, *11* (11), 4392−4401.

(25) Zheng, S.; Ding, H.; Li, S.; Chen, D.; Pan, F. Application of topology-based structure features for machine learning in materials science. *Chin. J. Struc Chem.* **2023**, *42* (7), 100120.

(26) Szymanski, N. J.; Smith, A.; Daoutidis, P.; Bartel, C. J. Topological Descriptors for the Electron Density of Inorganic Solids. *ACS Materials Letters* **2025**, *7* (6), 2158−2164.

(27) Carlsson, G. Topological methods for data modelling. *Nature Reviews Physics* **2020**, *2* (12), 697−708.

(28) Singh, Y.; Farrelly, C. M.; Hathaway, Q. A.; Leiner, T.; Jagtap, J.; Carlsson, G. E.; Erickson, B. J. Topological data analysis in medical imaging: current state of the art. *Insights Imaging* **2023**, *14* (1), 58.

(29) Bukkuri, A.; Andor, N.; Darcy, I. K. Applications of Topological Data Analysis in Oncology. *Front Artif Intell* **2021**, *4*, 659037.

(30) Loughrey, C. F.; Fitzpatrick, P.; Orr, N.; Jurek-Loughrey, A. The topology of data: opportunities for cancer research. *Bioinformatics* **2021**, *37* (19), 3091−3098.

(31) Amezquita, E. J.; Quigley, M. Y.; Ophelders, T.; Munch, E.; Chitwood, D. H. The shape of things to come: Topological data analysis and biology, from molecules to organisms. *Dev. Dyn.* **2020**, *249* (7), 816−833.

(32) Ver Hoef, L.; Adams, H.; King, E. J.; Ebert-Uphoff, I. A Primer on Topological Data Analysis to Support Image Analysis Tasks in Environmental Science. *Artificial Intelligence for the Earth Systems* **2023**, *2* (1), No. e220039.

(33) Xu, X.; Drougard, N.; Roy, R. N. Topological Data Analysis as a New Tool for EEG Processing. *Front Neurosci* **2021**, *15*, 761703.

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

(34) Hernandez-Lemus, E.; Miramontes, P.; Martinez-Garcia, M. Topological Data Analysis in Cardiovascular Signals: An Overview. *Entropy (Basel)* **2024**, *26*, 67.
(35) Skaf, Y.; Laubenbacher, R. Topological data analysis in biomedicine: A review. *J. Biomed Inform* **2022**, *130*, 104082.
(36) Wee, J.; Jiang, J. A Review of Topological Data Analysis and Topological Deep Learning in Molecular Sciences. *J. Chem. Inf Model* **2025**, *65* (23), 12691−12706.
(37) Smith, A. D.; Dłotko, P.; Zavala, V. M. Topological data analysis: Concepts, computation, and applications in chemical engineering. *Comput. Chem. Eng.* **2021**, *146*, 107202.
(38) Manetti, M. *Topology*; Springer Nature, 2023.
(39) Morris, S. A. *Topology without tears*; University of New England, 1989.
(40) Chazal, F.; Michel, B. An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists. *Front Artif Intell* **2021**, *4*, 667963.
(41) Ghrist, R. Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society* **2008**, *45* (01), 61−76.
(42) Carlsson, G.; Zomorodian, A.; Collins, A.; Guibas, L. J. Persistence Barcodes for Shapes. *International Journal of Shape Modeling* **2005**, *11* (02), 149−187.
(43) Grigor'yan, A. A.; Lin, Y.; Muranov, Y. V.; Yau, S.-T. Path Complexes and their Homologies. *Journal of Mathematical Sciences* **2020**, *248* (5), 564−599.
(44) Chen, D.; Liu, J.; Wu, J.; Wei, G. W. Persistent Hyperdigraph Homology and Persistent Hyperdigraph Laplacians. *Found Data Sci.* **2023**, *5* (4), 558−588.
(45) Smith, A.; Zavala, V. M. The Euler characteristic: A general topological descriptor for complex data. *Comput. Chem. Eng.* **2021**, *154*, 107463.
(46) Gurjeet Singh, F. M. a. G. C. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. *Eurographics* **2007**, *2*, 91−100.
(47) Bajaj, C. L.; Schikore, D. R. Topology preserving data simplification with error bounds. *Computers & Graphics* **1998**, *22* (1), 3−12.
(48) Edelsbrunner, H.; Letscher, D.; Zomorodian, A. Topological persistence and simplification. *Discrete & computational geometry* **2002**, *28* (4), 511−533.
(49) Zomorodian, A.; Carlsson, G. Computing Persistent Homology. *Discrete & Computational Geometry* **2005**, *33* (2), 249−274.
(50) Zomorodian, A.; Zamanian, M. Designing and Evaluating an Innovative Solar Air Collector with Transpired Absorber and Cover. *ISRN Renewable Energy* **2012**, *1*, 282538.
(51) Edelsbrunner, H.; Harer, J. *Computational topology: an introduction*; American Mathematical Soc., 2010.
(52) Hatcher, A. *Algebraic Topology*; Cambridge University Press, 2002.
(53) Henry Adams, T. E.; Kirby, M.; Neville, R.; Peterson, C.; Shipman, P.; Chepushtanova, S.; Hanson, E.; Motta, F.; Ziegelmeier, L. Persistence Images: A Stable Vector Representation of Persistent Homology. *Journal of Machine Learning Research* **2017**, *18*, 1−35.
(54) Rucco, M.; Castiglione, F.; Merelli, E.; Pettini, M. Characterisation of the idiotypic immune network through persistent entropy. In *Proceedings of ECCS 2014: European conference on complex systems*, 2016; Springer: pp 117−128.
(55) Chintakunta, H.; Gentimis, T.; Gonzalez-Diaz, R.; Jimenez, M.-J.; Krim, H. An entropy-based persistence barcode. *Pattern Recognition* **2015**, *48* (2), 391−401.
(56) Bubenik, P. Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.* **2015**, *16* (1), 77−102.
(57) Gromov, M. Hyperbolic Groups. In *Essays in Group Theory*; Mathematical Sciences Research Institute Publications, 1987; pp 75−263.
(58) Dantchev, S.; Ivrissimtzis, I. Efficient construction of the Čech complex. *Computers & Graphics* **2012**, *36* (6), 708−713.
(59) Kim, D.; Lee, M.; Cho, Y.; Kim, D. S. Beta-complex versus Alpha-complex: Similarities and Dissimilarities. *IEEE Trans Vis Comput. Graph* **2019**, *26* (4), 1686−1701.
(60) Borsuk, K. On the imbedding of systems of compacta in simplicial complexes. *Fundamenta Mathematicae* **1948**, *35* (1), 217−234.
(61) Bauer, U.; Kerber, M.; Roll, F.; Rolle, A. A unified view on the functorial nerve theorem and its variations. *Expositiones Mathematicae* **2023**, *41* (4), 125503.
(62) De Berg, M.; Cheong, O.; Van Kreveld, M.; Overmars, M. *Computational geometry: algorithms and applications*; Springer, 2008.
(63) Maria, C.; Boissonnat, J.-D.; Glisse, M.; Yvinec, M. The Gudhi Library: Simplicial Complexes and Persistent Homology. In *Mathematical Software - ICMS 2014*; Springer Berlin Heidelberg, 2014; Hong, H., Yap, C., Eds.; Berlin, Heidelberg, pp 167−174.
(64) Tralie, C.; Saul, N.; Bar-On, R. Ripser.py: A Lean Persistent Homology Library for Python. *Journal of Open Source Software* **2018**, *3* (29), 925.
(65) Morozov, D. . *Dionysus, a C++ library for computing persistent homology***2007**..
(66) Adams, H.; T, A.; Vejdemo-Johansson, M.. JavaPlex: A research software package for persistent (co) homology. In *International congress on mathematical software*; Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.
(67) Otter, N.; Porter, M. A.; Tillmann, U.; Grindrod, P.; Harrington, H. A. A roadmap for the computation of persistent homology. *EPJ. Data Sci.* **2017**, *6* (1), 17.
(68) Mischaikow, K.; Nanda, V. Morse Theory for Filtrations and Efficient Computation of Persistent Homology. *Discrete & Computational Geometry* **2013**, *50* (2), 330−353.
(69) Tierny, J.; Favelier, G.; Levine, J. A.; Gueunet, C.; Michaux, M. The topology toolkit. *IEEE transactions on visualization and computer graphics* **2018**, *24* (1), 832−842.
(70) Tauzin, G.; Lupo, U.; Tunstall, L.; Pérez, J. B.; Caorsi, M.; Medina-Mardones, A. M.; Dassatti, A.; Hess, K. giotto-tda:: A topological data analysis toolkit for machine learning and data exploration. *Journal of Machine Learning Research* **2021**, *22* (39), 1−6.
(71) Chen, D.; Liu, J.; Wu, J.; Wei, G. W.; Pan, F.; Yau, S. T. Path Topology in Molecular and Materials Sciences. *J. Phys. Chem. Lett.* **2023**, *14* (4), 954−964.
(72) West, D. B. *Introduction to graph theory*; Prentice hall Upper Saddle River, 2001.
(73) Garcia-Domenech, R.; Galvez, J.; de Julian-Ortiz, J. V.; Pogliani, L. Some new trends in chemical graph theory. *Chem. Rev.* **2008**, *108* (3), 1127−1169.
(74) Grigor'yan, A.; Lin, Y.; Muranov, Y.; Yau, S.-T. Homologies of path complexes and digraphs. *arXiv* **2012**, DOI: 10.48550/arXiv.1207.2834.
(75) Chowdhury, S.; Mémoli, F. Persistent path homology of directed networks. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2018; SIAM: pp 1152−1169.
(76) Grigor'yan, A.; Lin, Y.; Muranov, Y.; Yau, S.-T. Homotopy theory for digraphs. *arXiv* **2014**, DOI: 10.48550/arXiv.1407.0234.
(77) Grigor'yan, A.; Muranov, Y.; Vershinin, V.; Yau, S.-T. Path homology theory of multigraphs and quivers. In *Forum mathematicum*, 2018; De Gruyter: Vol. *30*, pp 1319−1337.
(78) Grigor'yan, A.; Lin, Y.; Muranov, Y.; Yau, S.-T. Cohomology of digraphs and (undirected) graphs. *Asian J. Math* **2015**, *19* (5), 887−931.
(79) Di, S.; Ivanov, S. O.; Mukoseev, L.; Zhang, M. On the path homology of Cayley digraphs and covering digraphs. *Journal of Algebra* **2024**, *653*, 156−199.
(80) Bressan, S.; Li, J.; Ren, S.; Wu, J. The embedded homology of hypergraphs and applications. *Asian J. Math* **2019**, *23* (3), 479−500.
(81) Grigor'yan, A.; Jimenez, R.; Muranov, Y.; Yau, S.-T. Homology of path complexes and hypergraphs. *Topology and its Applications* **2019**, *267*, 106877.

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

(82) Grbić, J.; Wu, J.; Xia, K.; Wei, G.-W. Aspects of topological approaches for data science. *Foundations of data science (Springfield, Mo.)* **2022**, *4* (2), 165.
(83) Muranov, Y.; Szczepkowska, A.; Vershinin, V. Homology of weighted path complexes and directed hypergraphs. *Portugaliae Mathematica* **2023**, *80* (1), 67−80.
(84) Dey, T. K.; Li, T.; Wang, Y. An efficient algorithm for 1-dimensional (persistent) path homology. *Discrete & Computational Geometry* **2022**, *68* (4), 1102−1132.
(85) Gasparovic, E.; Purvine, E.; Sazdanovic, R.; Wang, B.; Wang, Y.; Ziegelmeier, L. A survey of simplicial, relative, and chain complex homology theories for hypergraphs. *arXiv* **2024**, DOI: 10.48550/arXiv.2409.18310.
(86) Chen, D.; Liu, J.; Wu, J.; Wei, G.-W. Persistent hyperdigraph homology and persistent hyperdigraph Laplacians. *Foundations of data science (Springfield, Mo.)* **2023**, *5* (4), 558.
(87) Ren, S.; Wu, J. The stability of persistent homology of hypergraphs. *arXiv* **2020**, DOI: 10.48550/arXiv.2002.02237.
(88) Liu, X.; Feng, H.; Wu, J.; Xia, K. Computing hypergraph homology. *Foundations of Data Science* **2024**, *6* (2), 172−194.
(89) Tavakoli, M.; Shmakov, A.; Ceccarelli, F.; Baldi, P. Rxn hypergraph: a hypergraph attention model for chemical reaction representation. *arXiv* **2022**, DOI: 10.48550/arXiv.2201.01196.
(90) Battiston, F.; Cencetti, G.; Iacopini, I.; Latora, V.; Lucas, M.; Patania, A.; Young, J.-G.; Petri, G. Networks beyond pairwise interactions: Structure and dynamics. *Physics reports* **2020**, *874*, 1−92.
(91) Gallo, G.; Longo, G.; Pallottino, S.; Nguyen, S. Directed hypergraphs and applications. *Discrete applied mathematics* **1993**, *42* (2−3), 177−201.
(92) Chen, D.; Liu, J.; Wei, G.-W. Multiscale topology-enabled structure-to-sequence transformer for protein-ligand interaction predictions. *Nature Machine Intelligence* **2024**, *6* (7), 799−810.
(93) Adler, R. J. Some new random field tools for spatial analysis. *Stochastic Environmental Research and Risk Assessment* **2008**, *22* (6), 809−822.
(94) Laky, D. J.; Zavala, V. M. A fast and scalable computational topology framework for the Euler characteristic. *Digital Discovery* **2024**, 3 (2), 392−409.
(95) Hensel, F.; Moor, M.; Rieck, B. A Survey of Topological Machine Learning Methods. *Frontiers in Artificial Intelligence* **2021**, *4*, 681108.
(96) LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521* (7553), 436−444.
(97) Munkres, J. R. *Elements of algebraic topology*; CRC Press, 2018.
(98) Turner, K.; Mukherjee, S.; Boyer, D. M. Persistent homology transform for modeling shapes and surfaces. *Information and Inference: A Journal of the IMA* **2014**, *3* (4), 310−344.
(99) Cohen-Steiner, D.; Edelsbrunner, H.; Harer, J.; Mileyko, Y. Lipschitz functions have L p-stable persistence. *Foundations of computational mathematics* **2010**, *10* (2), 127−139.
(100) Atienza, N.; Gonzalez-Diaz, R.; Rucco, M. Persistent entropy for separating topological features from noise in vietoris-rips complexes. *Journal of Intelligent Information Systems* **2019**, *52* (3), 637−655.
(101) Cohen-Steiner, D.; Edelsbrunner, H.; Harer, J. Stability of persistence diagrams. *Proceedings of the twenty-first annual symposium on Computational geometry* **2005**, 263−271.
(102) Carlsson, G.; Zomorodian, A. The theory of multidimensional persistence. *Proceedings of the twenty-third annual symposium on Computational geometry* **2007**, 184−193.
(103) Turner, K.; Mileyko, Y.; Mukherjee, S.; Harer, J. Fréchet means for distributions of persistence diagrams. *Discrete & Computational Geometry* **2014**, *52* (1), 44−70.
(104) Lim, L.-H. Hodge Laplacians on graphs. *Siam Review* **2020**, *62* (3), 685−715.
(105) Chambers, E. W.; De Silva, V.; Erickson, J.; Ghrist, R. Vietoris-rips complexes of planar point sets. *Discrete & Computational Geometry* **2010**, *44* (1), 75−90.
(106) Grady, L. J.; Polimeni, J. R. *Discrete calculus: Applied analysis on graphs for computational science*; Springer, 2010.
(107) Hanik, M. *Geometric data analysis: Advancements of the statistical methodology and applications*; Freie Universitaet Berlin (Germany), 2023.
(108) Grandis, M. *Directed algebraic topology: models of non-reversible worlds*; Cambridge University Press, 2009.
(109) Chung, F. R. *Spectral graph theory*; American Mathematical Soc., 1997.
(110) Sun, J.; Ovsjanikov, M.; Guibas, L. A concise and provably informative multi-scale signature based on heat diffusion. In *Computer graphics forum*, 2009; Wiley Online Library: Vol. *28*, pp 1383−1392.
(111) Coifman, R. R.; Lafon, S.; Lee, A. B.; Maggioni, M.; Nadler, B.; Warner, F.; Zucker, S. W. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (21), 7426−7431.
(112) Von Luxburg, U. A tutorial on spectral clustering. *Statistics and computing* **2007**, *17* (4), 395−416.
(113) Chevyrev, I.; Nanda, V.; Oberhauser, H. Persistence paths and signature features in topological data analysis. *IEEE transactions on pattern analysis and machine intelligence* **2020**, *42* (1), 192−202.
(114) Chazal, F.; Fasy, B.; Lecci, F.; Michel, B.; Rinaldo, A.; Wasserman, L. Robust topological inference: Distance to a measure and kernel distance. *Journal of Machine Learning Research* **2018**, *18* (159), 1−40.
(115) Som, A.; Choi, H.; Ramamurthy, K. N.; Buman, M. P.; Turaga, P. Pi-net: A deep learning approach to extract topological persistence images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020; pp 834−835.
(116) Moor, M.; Horn, M.; Rieck, B.; Borgwardt, K. Topological autoencoders. In *International conference on machine learning*, 2020; PMLR: pp 7045−7054.
(117) Gabrielsson, R. B.; Nelson, B. J.; Dwaraknath, A.; Skraba, P. A topology layer for machine learning. In *International Conference on Artificial Intelligence and Statistics*, 2020; PMLR: pp 1553−1563.
(118) Hofer, C.; Kwitt, R.; Niethammer, M.; Uhl, A. Deep learning with topological signatures. *arXiv* **2017**, DOI: 10.48550/arXiv.1707.04041.
(119) Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, 2020; PmLR: pp 1597−1607.
(120) Chen, Y.; Frias, J.; Gel, Y. R. TopoGCL: Topological graph contrastive learning. In *Proceedings of the AAAI conference on artificial intelligence*, 2024; Vol. *38*, pp 11453−11461.
(121) Luo, Y.; Shi, L.; Thost, V. Improving self-supervised molecular representation learning using persistent homology. *Advances in Neural Information Processing Systems* **2023**, *36*, 34043−34073.
(122) Hajij, M.; Istvan, K.; Zamzmi, G. Cell complex neural networks. *arXiv* **2020**, DOI: 10.48550/arXiv.2010.00743.
(123) Feng, Y.; You, H.; Zhang, Z.; Ji, R.; Gao, Y. Hypergraph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, 2019; Vol. *33*, pp 3558−3565.
(124) Park, J.; Hwang, Y.; Kim, M.; Chung, M. K.; Wu, G.; Kim, W. H. Convolving directed graph edges via Hodge Laplacian for brain network analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2023; Springer: pp 789−799.
(125) Anand, D. V.; Chung, M. K. Hodge Laplacian of brain networks. *IEEE transactions on medical imaging* **2023**, *42* (5), 1563−1573.
(126) Xie, T.; Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.* **2018**, *120* (14), 145301.
(127) Yang, G.; Liang, X.; Zheng, S.; Chen, H.; Zhang, W.; Li, S.; Pan, F. Li-rich channels as the material gene for facile lithium diffusion in halide solid electrolytes. *eScience* **2022**, *2* (1), 79−86.

(128) Peterson, G. G. C.; Brgoch, J. Materials discovery through machine learning formation energy. *Journal of Physics: Energy* **2021**, *3* (2), 022002.
(129) Chen, D.; Wang, R.; Wei, G. W.; Pan, F. Enhancing energy predictions in multi-atom systems with multiscale topological learning. *J. Mater. Chem. A Mater.* **2025**, *13* (27), 21555−21563.
(130) Anand, D. V.; Xu, Q.; Wee, J.; Xia, K.; Sum, T. C. Topological feature engineering for machine learning based halide perovskite materials design. *npj Computational Materials* **2022**, *8* (1), 203.
(131) Blatov, V. A. Voronoi-dirichlet polyhedra in crystal chemistry: theory and applications. *Crystallography Reviews* **2004**, *10* (4), 249−318.
(132) Fang, Z.; Yan, Q. Leveraging Persistent Homology Features for Accurate Defect Formation Energy Predictions via Graph Neural Networks. *Chem. Mater.* **2025**, *37* (4), 1531−1540.
(133) Deringer, V. L.; Caro, M. A.; Csanyi, G. Machine Learning Interatomic Potentials as Emerging Tools for Materials Science. *Adv. Mater.* **2019**, *31* (46), No. e1902765.
(134) Friederich, P.; Hase, F.; Proppe, J.; Aspuru-Guzik, A. Machine-learned potentials for next-generation matter simulations. *Nat. Mater.* **2021**, *20* (6), 750−761.
(135) Mueller, T.; Hernandez, A.; Wang, C. Machine learning for interatomic potential models. *J. Chem. Phys.* **2020**, *152* (5), 050902.
(136) Behler, J. Four Generations of High-Dimensional Neural Network Potentials. *Chem. Rev.* **2021**, *121* (16), 10037−10072.
(137) Minamitani, E.; Obayashi, I.; Shimizu, K.; Watanabe, S. Persistent homology-based descriptor for machine-learning potential of amorphous structures. *J. Chem. Phys.* **2023**, *159* (8), 084101.
(138) Mirth, J.; Zhai, Y.; Bush, J.; Alvarado, E. G.; Jordan, H.; Heim, M.; Krishnamoorthy, B.; Pflaum, M.; Clark, A.; Z, Y.; et al. Representations of energy landscapes by sublevelset persistent homology: An example with n-alkanes. *J. Chem. Phys.* **2021**, *154* (11), 114114.
(139) Wang, A.; Zou, L. Persistent homology for structural characterization in disordered systems. *Phys. Rev. E* **2025**, *111* (4−2), 045306.
(140) Hu, C. S.; Mayengbam, R.; Xia, K.; Sum, T. C. Quotient Complex (QC)-Based Machine Learning for 2D Hybrid Perovskite Design. *J. Chem. Inf Model* **2025**, *65* (2), 660−671.
(141) Yang, Y.; Urban, M. W. Self-healing polymeric materials. *Chem. Soc. Rev.* **2013**, *42* (17), 7446−7467.
(142) Brazel, C. S.; Rosen, S. L. *Fundamental principles of polymeric materials*; John Wiley & Sons, 2012.
(143) Jochum, F. D.; Theato, P. Temperature- and light-responsive smart polymer materials. *Chem. Soc. Rev.* **2013**, *42* (17), 7468−7483.
(144) Ichinomiya, T.; Obayashi, I.; Hiraoka, Y. Persistent homology analysis of craze formation. *Phys. Rev. E* **2017**, *95* (1−1), 012504.
(145) Shimizu, Y.; Kurokawa, T.; Arai, H.; Washizu, H. Higher-order structure of polymer melt described by persistent homology. *Sci. Rep* **2021**, *11* (1), 2274.
(146) Landuzzi, F.; Nakamura, T.; Michieletto, D.; Sakaue, T. Persistence homology of entangled rings. *Physical Review Research* **2020**, *2* (3), 033529.
(147) Sato, R.; Kawakami, S.; Ejima, H.; Ujii, T.; Sato, K.; Ichiki, T.; Shibuta, Y. Dynamic Correlation Analysis between Stress-Strain Curve and Polymer Film Structure Using Persistent Homology. *J. Chem. Theory Comput* **2024**, *20* (24), 10798−10806.
(148) Higashi, Y.; Okuwaki, K.; Mochizuki, Y.; Fujigaya, T.; Kato, K. Feature vectorization of microphase-separated structures in polymeric materials using dissipative particle dynamics and persistent homology for machine learning applications. *Digital Discovery* **2025**, *4* (5), 1339−1351.
(149) Kojima, T.; Washio, T.; Hara, S.; Koishi, M.; Amino, N. Analysis on Microstructure-Property Linkages of Filled Rubber Using Machine Learning and Molecular Dynamics Simulations. *Polymers (Basel)* **2021**, *13* (16), 2683.
(150) Zhang, Y.; Shen, C.; Xia, K. Multi-Cover Persistence (MCP)-based machine learning for polymer property prediction. *Brief Bioinform* **2024**, *25* (6), bbae465.
(151) Tanaka, F. *Polymer physics: applications to molecular association and thermoreversible gelation*; Cambridge University Press, 2011.
(152) Pekcan, Ö.; Kara, S. Gelation Mechanisms. *Mod. Phys. Lett. B* **2012**, *26* (27), 1230019.
(153) Smith, A. D.; Donley, G. J.; Del Gado, E.; Zavala, V. M. Topological Data Analysis for Particulate Gels. *ACS Nano* **2024**, *18* (42), 28622−28635.
(154) Scholze, H. *Glass: nature, structure, and properties*; Springer Science & Business Media, 2012.
(155) Axinte, E. Glasses as engineering materials: A review. *Materials & Design* **2011**, *32* (4), 1717−1732.
(156) Amann-Winkel, K.; Bellissent-Funel, M. C.; Bove, L. E.; Loerting, T.; Nilsson, A.; Paciaroni, A.; Schlesinger, D.; Skinner, L. X-ray and Neutron Scattering of Water. *Chem. Rev.* **2016**, *116* (13), 7570−7589.
(157) Hiraoka, Y.; Nakamura, T.; Hirata, A.; Escolar, E. G.; Matsue, K.; Nishiura, Y. Hierarchical structures of amorphous solids characterized by persistent homology. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (26), 7035−7040.
(158) Firooz, A. F.; Christensen, R.; Biscio, C. A. N.; Smedskjaer, M. M. Characterizing medium-range order structure of binary silicate glasses using ring analysis and persistent homology. *J. Am. Ceram. Soc.* **2024**, *107* (12), 7739−7750.
(159) Nakamura, T.; Hiraoka, Y.; Hirata, A.; Escolar, E. G.; Nishiura, Y. Persistent homology and many-body atomic structure for medium-range order in the glass. *Nanotechnology* **2015**, *26* (30), 304001.
(160) Xiao, Y.; Du, T.; Sørensen, S. S.; Chen, Z.; Biscio, C. A. N.; Fajstrup, L.; Bauchy, M.; Smedskjaer, M. M. Deciphering the hierarchical structure of phosphate glasses using persistent homology with optimized input radii. *Physical Review Materials* **2023**, *7* (6), 065602.
(161) Tavanti, F.; Calzolari, A. Multi-technique Approach to Unravel the (Dis)order in Amorphous Materials. *ACS Omega* **2022**, *7* (27), 23255−23264.
(162) Tirelli, A.; Nakano, K. Topological Data Analysis for Revealing the Structural Origin of Density Anomalies in Silica Glass. *J. Phys. Chem. B* **2023**, *127* (14), 3302−3311.
(163) Pan, Z.; Atila, A.; Bitzek, E.; Wondraczek, L. Topology of anisotropic glasses from persistent homology analysis. *J. Non-Cryst. Solids* **2024**, *627*, 122801.
(164) Minamitani, E.; Shiga, T.; Kashiwagi, M.; Obayashi, I. Topological descriptor of thermal conductivity in amorphous Si. *J. Chem. Phys.* **2022**, *156* (24), 244502.
(165) Wu, Y.-J.; Akagi, K.; Goto, M.; Xu, Y. Topological data analysis of TEM-based structural features affecting the thermal conductivity of amorphous Ge. *Int. J. Heat Mass Transfer* **2024**, *221*, 125012.
(166) Shiga, T.; Minamitani, E.; Yamashita, Y.; Yagi, T.; Taketoshi, N.; Shigesato, Y.; Kashiwagi, M. Thermal transport and topological analyses of the heat-carrying modes and their relevant local structures in variously dense amorphous alumina. *Appl. Phys. Lett.* **2024**, *125* (1), 012201.
(167) Seh, Z. W.; Kibsgaard, J.; Dickens, C. F.; Chorkendorff, I.; Norskov, J. K.; Jaramillo, T. F. Combining theory and experiment in electrocatalysis: Insights into materials design. *Science* **2017**, *355* (6321), No. eaad4998.
(168) Zheng, R.; Liu, Z.; Wang, Y.; Xie, Z.; He, M. The future of green energy and chemicals: Rational design of catalysis routes. *Joule* **2022**, *6* (6), 1148−1159.
(169) Zheng, S.; Yang, X.; Shi, Z. Z.; Ding, H.; Pan, F.; Li, J. F. The Loss of Interfacial Water-Adsorbate Hydrogen Bond Connectivity Position Surface-Active Hydrogen as a Crucial Intermediate to Enhance Nitrate Reduction Reaction. *J. Am. Chem. Soc.* **2024**, *146* (39), 26965−26974.
(170) Wang, Y.-H.; Zheng, S.; Yang, W.-M.; Zhou, R.-Y.; He, Q.-F.; Radjenovic, P.; Dong, J.-C.; Li, S.; Zheng, J.; Yang, Z.-L.; et al. In situ Raman spectroscopy reveals the structure and dissociation of interfacial water. *Nature* **2021**, *600* (7887), 81−85.

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

(171) Wu, J.; Huang, Y.; Ye, W.; Li, Y. CO2 Reduction: From the Electrochemical to Photochemical Approach. *Adv. Sci. (Weinh)* **2017**, *4* (11), 1700194.
(172) Fung, V.; Hu, G.; Ganesh, P.; Sumpter, B. G. Machine learned features from density of states for accurate adsorption energy prediction. *Nat. Commun.* **2021**, *12* (1), 88.
(173) Ock, J.; Badrinarayanan, S.; Magar, R.; Antony, A.; Barati Farimani, A. Multimodal language and graph learning of adsorption configuration in catalysis. *Nature Machine Intelligence* **2024**, *6* (12), 1501−1511.
(174) Wei, P.; Gao, D.; Liu, T.; Li, H.; Sang, J.; Wang, C.; Cai, R.; Wang, G.; Bao, X. Coverage-driven selectivity switch from ethylene to acetate in high-rate CO(2)/CO electrolysis. *Nat. Nanotechnol* **2023**, *18* (3), 299−306.
(175) Chen, B. W. J.; Mavrikakis, M. Modeling the impact of structure and coverage on the reactivity of realistic heterogeneous catalysts. *Nature Chemical Engineering* **2025**, *2* (3), 181−197.
(176) Zheng, S.; Liang, X.; Pan, J.; Hu, K.; Li, S.; Pan, F. Multi-Center Cooperativity Enables Facile C-C Coupling in Electrochemical CO2 Reduction on a Ni2P Catalyst. *ACS Catal.* **2023**, *13* (5), 2847−2856.
(177) Zhang, S.; Lv, H.-B.; Shi, Z.-Z.; Wang, R.; Zheng, S.; Li, J.-F. Subsurface hydrogen as a hidden driver of copper surface reconstruction in CO_2 electroreduction. *Natl. Sci. Rev.* **2026**, *13*, nwag128.
(178) Zheng, S.; Zhang, X.-M.; Liu, H.-S.; Liang, G.-H.; Zhang, S.-W.; Zhang, W.; Wang, B.; Yang, J.; Jin, X. a.; Pan, F.; et al. Active phase discovery in heterogeneous catalysis via topology-guided sampling and machine learning. *Nat. Commun.* **2025**, *16* (1), 2542.
(179) Andersen, M.; Reuter, K. Adsorption Enthalpies for Catalysis Modeling through Machine-Learned Descriptors. *Acc. Chem. Res.* **2021**, *54* (12), 2741−2749.
(180) Mou, L. H.; Han, T.; Smith, P. E. S.; Sharman, E.; Jiang, J. Machine Learning Descriptors for Data-Driven Catalysis Study. *Adv. Sci. (Weinh)* **2023**, *10* (22), No. e2301020.
(181) Liang, G.-H.; Liu, H.-S.; Zhang, X.-M.; Li, J.-F.; Zheng, S. Topology-based machine learning for predicting curvature effects in metal-nitrogen-carbon single-atom catalysts. *Journal of Energy Chemistry* **2025**, *105*, 608−616.
(182) Wang, B.; Zheng, S.; Wu, J.; Li, J.; Pan, F. Inverse design of catalytic active sites via interpretable topology-based deep generative models. *npj Computational Materials* **2025**, *11* (1), 147.
(183) Kumara, L. S. R.; Seo, O.; Kim, J.; Wang, F.; Chen, Y.; Gueye, I.; Song, C.; Ohara, K.; Kusada, K.; Kitagawa, H.; et al. Persistent Homology and Bond Orientational Order in Ir-Cu Solid-Solution Alloy Nanoparticles: Implications for Electrocatalysts. *ACS Applied Nano Materials* **2023**, *6* (18), 16653−16661.
(184) Bennett, T. D.; Coudert, F. X.; James, S. L.; Cooper, A. I. The changing state of porous materials. *Nat. Mater.* **2021**, *20* (9), 1179−1187.
(185) Huang, D.-S.; Qiu, X.-F.; Huang, J.-R.; Mao, M.; Liu, L.; Han, Y.; Zhao, Z.-H.; Liao, P.-Q.; Chen, X.-M. Electrosynthesis of urea by using Fe2O3 nanoparticles encapsulated in a conductive metal-organic framework. *Nature Synthesis* **2024**, *3* (11), 1404−1413.
(186) Zhu, Q.; Gu, Y.; Liang, X.; Wang, X.; Ma, J. A Machine Learning Model To Predict CO2 Reduction Reactivity and Products Transferred from Metal-Zeolites. *ACS Catal.* **2022**, *12* (19), 12336−12348.
(187) Wang, X.; Ma, T.; Ma, J.-G.; Cheng, P. Integration of devices based on metal-organic frameworks: A promising platform for chemical sensing. *Coord. Chem. Rev.* **2024**, *518*, 216067.
(188) Kang, Y.; Kim, J. ChatMOF: an artificial intelligence system for predicting and generating metal-organic frameworks using large language models. *Nat. Commun.* **2024**, *15* (1), 4705.
(189) Yao, Z.; Sánchez-Lengeling, B.; Bobbitt, N. S.; Bucior, B. J.; Kumar, S. G. H.; Collins, S. P.; Burns, T.; Woo, T. K.; Farha, O. K.; Snurr, R. Q.; et al. Inverse design of nanoporous crystalline reticular materials with deep generative models. *Nature Machine Intelligence* **2021**, *3* (1), 76−86.
(190) Kim, N.; Min, K. Accelerated Discovery of Zeolite Structures with Superior Mechanical Properties via Active Learning. *J. Phys. Chem. Lett.* **2021**, *12* (9), 2334−2339.
(191) Moosavi, S. M.; Xu, H.; Chen, L.; Cooper, A. I.; Smit, B. Geometric landscapes for material discovery within energy-structure-function maps. *Chem. Sci.* **2020**, *11* (21), 5423−5433.
(192) Chen, D.; Chen, C. L.; Wei, G. W. Category-specific topological learning of metal-organic frameworks. *J. Mater. Chem. A Mater.* **2025**, *13* (13), 9292−9303.
(193) Chen, D. L. J.; Chen, C.-L.; Wei, G.-W. Interaction topological transformer for multiscale learning in porous materials. *arXiv* **2025**, DOI: 10.48550/arXiv.2509.18573.
(194) Yang, Y.; Guo, S.; Li, S.; Wu, Y.; Qiao, Z. Topological Data Analysis Combined with High-Throughput Computational Screening of Hydrophobic Metal-Organic Frameworks: Application to the Adsorptive Separation of C3 Components. *Nanomaterials (Basel)* **2024**, *14* (3), 298.
(195) Shekhar, S.; Chowdhury, C. Topological data analysis enhanced prediction of hydrogen storage in metal-organic frameworks (MOFs). *Materials Advances* **2024**, *5* (2), 820−830.
(196) Krishnapriyan, A. S.; Montoya, J.; Haranczyk, M.; Hummelshoj, J.; Morozov, D. Machine learning with persistent homology and chemical word embeddings improves prediction accuracy and interpretability in metal-organic frameworks. *Sci. Rep* **2021**, *11* (1), 8888.
(197) Bates, A. M.; Preger, Y.; Torres-Castro, L.; Harrison, K. L.; Harris, S. J.; Hewson, J. Are solid-state batteries safer than lithium-ion batteries? *Joule* **2022**, *6* (4), 742−755.
(198) Ke, X.; Wang, Y.; Ren, G.; Yuan, C. Towards rational mechanical design of inorganic solid electrolytes for all-solid-state lithium ion batteries. *Energy Storage Materials* **2020**, *26*, 313−324.
(199) Yang, Q.; Deng, N.; Zhao, Y.; Gao, L.; Cheng, B.; Kang, W. A review on 1D materials for all-solid-state lithium-ion batteries and all-solid-state lithium-sulfur batteries. *Chem. Eng. J.* **2023**, *451*, 138532.
(200) Chang, X.; Zhao, Y.-M.; Yuan, B.; Fan, M.; Meng, Q.; Guo, Y.-G.; Wan, L.-J. Solid-state lithium-ion batteries for grid energy storage: opportunities and challenges. *Science China Chemistry* **2024**, *67* (1), 43−66.
(201) Li, B.; Chao, Y.; Li, M.; Xiao, Y.; Li, R.; Yang, K.; Cui, X.; Xu, G.; Li, L.; Yang, C.; et al. A Review of Solid Electrolyte Interphase (SEI) and Dendrite Formation in Lithium Batteries. *Electrochemical Energy Reviews* **2023**, *6* (1), 7.
(202) Zhang, W.; Weng, M.; Zhang, M.; Chen, Z.; Wang, B.; Li, S.; Pan, F. Rapid Mining of Fast Ion Conductors via Subgraph Isomorphism Matching. *J. Am. Chem. Soc.* **2024**, *146* (27), 18535−18543.
(203) Yang, H.; Wu, N. Ionic conductivity and ion transport mechanisms of solid-state lithium-ion battery electrolytes: A review. *Energy Science & Engineering* **2022**, *10* (5), 1643−1671.
(204) Jalem, R.; Chandrappa, M. L. H.; Qi, J.; Tateyama, Y.; Ong, S. P. Lithium dynamics at grain boundaries of β-Li3PS4 solid electrolyte. *Energy Advances* **2023**, *2* (12), 2029−2041.
(205) Chen, D.; Wang, B.; Li, S.; Zhang, W.; Yang, K.; Song, Y.; Wei, G. W.; Pan, F. Superionic Ionic Conductor Discovery via Multiscale Topological Learning. *J. Am. Chem. Soc.* **2025**, *147* (24), 20888−20898.
(206) Du, G.; Zhou, L.; Huang, Y.; Wu, Y.; Hong, Z. Topological data analysis assisted machine learning for polar topological structures in oxide superlattices. *Acta Mater.* **2025**, *282*, 120467.
(207) Wang, Z. L.; Ogawa, T.; Adachi, Y. Property Predictions for Dual-Phase Steels Using Persistent Homology and Machine Learning. *Advanced Theory and Simulations* **2020**, *3* (3), 1900227.
(208) Ormrod Morley, D.; Salmon, P. S.; Wilson, M. Persistent homology in two-dimensional atomic networks. *J. Chem. Phys.* **2021**, *154* (12), 124109.
(209) Saadatfar, M.; Takeuchi, H.; Robins, V.; Francois, N.; Hiraoka, Y. Pore configuration landscape of granular crystallization. *Nat. Commun.* **2017**, *8*, 15082.

(210) Madeleine, T.; Podoliak, N.; Buchnev, O.; Membrillo Solis, I.; Orlova, T.; van Rossem, M.; Kaczmarek, M.; D'Alessandro, G.; Brodzki, J. Topological Learning for the Classification of Disorder: An Application to the Design of Metasurfaces. *ACS Nano* **2024**, *18* (1), 630−640.
(211) Rasul, A.; Hossain, M. S.; Dastider, A. G.; Roy, H.; Hasan, M. Z.; Khosru, Q. D. M. A machine learning based classifier for topological quantum materials. *Sci. Rep* **2024**, *14* (1), 31564.
(212) Xu, Y. Efficient Classification Method for Topological Quantum Materials Based on Graph Neural Networks and Persistent Homology Theory. *Computers and Artificial Intelligence* **2025**, *2* (2), 53−65.
(213) Membrillo Solis, I.; Orlova, T.; Bednarska, K.; Lesiak, P.; Woliński, T. R.; D'Alessandro, G.; Brodzki, J.; Kaczmarek, M. Tracking the time evolution of soft matter systems via topological structural heterogeneity. *Communications Materials* **2022**, *3* (1), 1−10.
(214) Chen, Z.; Li, Y.; Wang, L.; Zhang, J.; Dang, R.; Yu, H.; Wang, B.; Xu, C.; Li, H.; Li, Y.; et al. Durable alloy anode for Na-ion batteries with high volumetric energy density. *Nature Energy* **2026**, *11* (3), 449−459.
(215) Smith, A.; Runde, S.; Chew, A. K.; Kelkar, A. S.; Maheshwari, U.; Van Lehn, R. C.; Zavala, V. M. Topological Analysis of Molecular Dynamics Simulations using the Euler Characteristic. *J. Chem. Theory Comput* **2023**, *19* (5), 1553−1567.
(216) Kumar, N.; Clark, A. E. Adsorbate Organization Characterized by Sublevelset Persistent Homology. *J. Chem. Theory Comput* **2023**, *19* (11), 3303−3312.
(217) Yamaguchi, S.; Li, H.; Funayama, N.; Kohno, T.; Imazato, S. Topological features of lithium disilicate glass-ceramics uncovered through materials informatics. *Dent. Mater.* **2026**, *42* (1), 16−22.
(218) Manuchehrfar, F.; Li, H.; Tian, W.; Ma, A.; Liang, J. Exact Topology of the Dynamic Probability Surface of an Activated Process by Persistent Homology. *J. Phys. Chem. B* **2021**, *125* (18), 4667−4680.
(219) Story, B.; Sadhu, B.; Adams, H.; Clark, A. E. Additive energy functions have predictable landscape topologies. *J. Chem. Phys.* **2023**, *158* (16), 164104.
(220) Wee, J.; Jiang, J. A review of topological data analysis and topological deep learning in molecular sciences. *J. Chem. Inf. Model.* **2025**, *65* (23), 12691−12706.
(221) Amézquita, E. J.; Quigley, M. Y.; Ophelders, T.; Munch, E.; Chitwood, D. H. The shape of things to come: Topological data analysis and biology, from molecules to organisms. *Dev. Dyn.* **2020**, *249* (7), 816−833.
(222) Skaf, Y.; Laubenbacher, R. Topological data analysis in biomedicine: A review. *Journal of Biomedical Informatics* **2022**, *130*, 104082.
(223) Bowman, G. R.; Huang, X.; Yao, Y.; Sun, J.; Carlsson, G.; Guibas, L. J.; Pande, V. S. Structural insight into RNA hairpin folding intermediates. *J. Am. Chem. Soc.* **2008**, *130* (30), 9676−9678.
(224) Donati, L.; Chewle, S.; St Pierre, D.; Natarajan, V.; Weber, M. Topological analysis reveals multiple pathways in molecular dynamics. *J. Chem. Theory Comput.* **2025**, *21* (20), 10385−10397.
(225) Olejniczak, M.; Severo Pereira Gomes, A.; Tierny, J. A topological data analysis perspective on noncovalent interactions in relativistic calculations. *Int. J. Quantum Chem.* **2020**, *120* (8), No. e26133.
(226) Dargaville, B. L.; Hutmacher, D. W. Water as the often neglected medium at the interface between materials and biology. *Nat. Commun.* **2022**, *13* (1), 4222.
(227) Shih, A. J.; Monteiro, M. C. O.; Dattila, F.; Pavesi, D.; Philips, M.; da Silva, A. H. M.; Vos, R. E.; Ojha, K.; Park, S.; van der Heijden, O.; et al. Water electrolysis. *Nature Reviews Methods Primers* **2022**, *2* (1), 84.
(228) Pirashvili, M.; Steinberg, L.; Belchi Guillamon, F.; Niranjan, M.; Frey, J. G.; Brodzki, J. Improved understanding of aqueous solubility modeling through topological data analysis. *J. Cheminform* **2018**, *10* (1), 54.
(229) Steinberg, L.; Russo, J.; Frey, J. A new topological descriptor for water network structure. *J. Cheminform* **2019**, *11* (1), 48.
(230) Hu, Y.; Ounkham, P.; Marsalek, O.; Markland, T. E.; Krishmoorthy, B.; Clark, A. E. Persistent Homology Metrics Reveal Quantum Fluctuations and Reactive Atoms in Path Integral Dynamics. *Frontiers in Chemistry* **2021**, *9*, 624937.
(231) Xia, K. Persistent homology analysis of ion aggregations and hydrogen-bonding networks. *Phys. Chem. Chem. Phys.* **2018**, *20* (19), 13448−13460.
(232) Xia, K.; Anand, D. V.; Shikhar, S.; Mu, Y. Persistent homology analysis of osmolyte molecular aggregation and their hydrogen-bonding networks. *Phys. Chem. Chem. Phys.* **2019**, *21* (37), 21038−21048.
(233) Anand, D. V.; Meng, Z.; Xia, K.; Mu, Y. Weighted persistent homology for osmolyte molecular aggregation and hydrogen-bonding network analysis. *Sci. Rep* **2020**, *10* (1), 9685.
(234) Hong, S.; Kim, D. Medium-range order in amorphous ices revealed by persistent homology. *J. Phys.: Condens. Matter* **2019**, *31* (45), 455403.
(235) Qiu, Y. R.; Peng, W.; Zhang, R.; Wang, Y. H.; Guo, W.; Li, J. F. Frontier interfacial water properties characterization and applications. *Natl. Sci. Rev.* **2025**, *12* (11), nwaf284.
(236) Gonella, G.; Backus, E. H. G.; Nagata, Y.; Bonthuis, D. J.; Loche, P.; Schlaich, A.; Netz, R. R.; Kuhnle, A.; McCrum, I. T.; Koper, M. T. M.; et al. Water at charged interfaces. *Nat. Rev. Chem.* **2021**, *5* (7), 466−485.
(237) Li, C.; Shen, Z. Role of Solvents in Oriented Attachment of Ag Nanoparticles: Insights from Molecular Dynamics Simulations and Topological Analysis. *Langmuir* **2024**, *40* (33), 17423−17429.
(238) Liu, H. S.; Zhang, X. M.; Liang, G. H.; Zheng, S. S.; Li, J. F. Investigation of water structure and proton transfer within confined graphene by molecule dynamics and multiscale data analysis. *Chin. J. Struc Chem.* **2025**, *44* (6), 100596.
(239) Townsend, J.; Micucci, C. P.; Hymel, J. H.; Maroulas, V.; Vogiatzis, K. D. Representation of molecular structures with persistent homology for machine learning applications in chemistry. *Nat. Commun.* **2020**, *11* (1), 3230.
(240) Jones, G. M.; Smith, B. A.; Kirkland, J. K.; Vogiatzis, K. D. Data-driven ligand field exploration of Fe (iv)-oxo sites for C-H activation. *Inorganic Chemistry Frontiers* **2023**, *10* (4), 1062−1075.
(241) Turcani, L.; Greenaway, R. L.; Jelfs, K. E. Machine learning for organic cage property prediction. *Chem. Mater.* **2019**, *31* (3), 714−727.
(242) Schiff, Y.; Chenthamarakshan, V.; Hoffman, S. C.; Ramamurthy, K. N.; Das, P. Augmenting molecular deep generative models with topological data analysis representations. In *ICASSP 2022−2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022; IEEE: pp 3783−3787.
(243) Xia, K.; Wei, G. W. Persistent homology analysis of protein structure, flexibility, and folding. *International journal for numerical methods in biomedical engineering* **2014**, *30* (8), 814−844.
(244) Meng, Z.; Xia, K. Persistent spectral-based machine learning (PerSpect ML) for protein-ligand binding affinity prediction. *Science advances* **2021**, *7* (19), No. eabc5329.
(245) Wang, Y.; Liu, X.; Zhang, Y.; Wang, X.; Xia, K. Join Persistent Homology (JPH)-Based Machine Learning for Metalloprotein-Ligand Binding Affinity Prediction. *J. Chem. Inf. Model.* **2025**, *65* (6), 2785−2793.
(246) Han, B.; Zhang, Y.; Li, L.; Gong, X.; Xia, K. TopoQA: a topological deep learning-based approach for protein complex structure interface quality assessment. *Briefings in Bioinformatics* **2025**, *26* (2), bbaf083.
(247) Nguyen, D. D.; Cang, Z.; Wei, G.-W. A review of mathematical representations of biomolecular data. *Phys. Chem. Chem. Phys.* **2020**, *22* (8), 4343−4367.
(248) Cang, Z.; Mu, L.; Wei, G.-W. Representability of algebraic topology for biomolecules in machine learning based scoring and virtual screening. *PLoS computational biology* **2018**, *14* (1), No. e1005929.

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067

(249) Kovacev-Nikolic, V.; Bubenik, P.; Nikolić, D.; Heo, G. Using persistent homology and dynamical distances to analyze protein binding. *Statistical applications in genetics and molecular biology* **2016**, *15* (1), 19−38.
(250) Cang, Z.; Wei, G.-W. Analysis and prediction of protein folding energy changes upon mutation by element specific persistent homology. *Bioinformatics* **2017**, *33* (22), 3549−3557.
(251) Campbell, J. W.; Vogiatzis, K. D. Augmenting maccs keys with persistent homology fingerprints for protein-ligand binding classification. *J. Chem. Inf. Model.* **2025**, *65* (15), 8113−8126.
(252) Wee, J.; Wei, G.-W. Rapid response to fast viral evolution using AlphaFold 3-assisted topological deep learning. *Virus Evolution* **2025**, *11* (1), veaf026.
(253) Du, H.; Wei, G.-W.; Hou, T. Multiscale topology in interactomic network: from transcriptome to antiaddiction drug repurposing. *Briefings in Bioinformatics* **2024**, *25* (2), bbae054.
(254) Zhu, Z.; Dou, B.; Cao, Y.; Jiang, J.; Zhu, Y.; Chen, D.; Feng, H.; Liu, J.; Zhang, B.; Zhou, T.; et al. Tidal: topology-inferred drug addiction learning. *J. Chem. Inf. Model.* **2023**, *63* (5), 1472−1489.
(255) Keller, B.; Lesnick, M.; Willke, T. L. Persistent homology for virtual screening. *ChemRxiv* **2018**, DOI: 10.26434/chemrxiv.6969260.v3.
(256) Wu, S.; Liu, X.; Dong, A.; Gragnoli, C.; Griffin, C.; Wu, J.; Yau, S.-T.; Wu, R. The metabolomic physics of complex diseases. *Proc. Natl. Acad. Sci. U. S. A.* **2023**, *120* (42), No. e2308496120.
(257) Feng, L.; Gong, H.; Zhang, S.; Liu, X.; Wang, Y.; Che, J.; Dong, A.; Griffin, C. H.; Gragnoli, C.; Wu, J.; et al. Hypernetwork modeling and topology of high-order interactions for complex systems. *Proc. Natl. Acad. Sci. U. S. A.* **2024**, *121* (40), No. e2412220121.
(258) Feng, L.; Yang, D.; Wu, S.; Xue, C.; Sang, M.; Liu, X.; Che, J.; Wu, J.; Gragnoli, C.; Griffin, C.; et al. Network modeling and topology of aging. *Physics reports* **2025**, *1101*, 1−65.
(259) Bougueroua, S.; Spezia, R.; Pezzotti, S.; Vial, S.; Quessette, F.; Barth, D.; Gaigeot, M. P. Graph theory for automatic structural recognition in molecular dynamics simulations. *J. Chem. Phys.* **2018**, *149* (18), 184102.
(260) Li, S.; Chen, Z.; Wang, Z.; Weng, M.; Li, J.; Zhang, M.; Lu, J.; Xu, K.; Pan, F. Graph-based discovery and analysis of atomic-scale one-dimensional materials. *Natl. Sci. Rev.* **2022**, *9* (6), 28.
(261) Weng, M.; Wang, Z.; Qian, G.; Ye, Y.; Chen, Z.; Chen, X.; Zheng, S.; Pan, F. Identify crystal structures by a new paradigm based on graph theory for building materials big data. *Science China Chemistry* **2019**, *62* (8), 982−986.
(262) Townsend, J.; Micucci, C. P.; Hymel, J. H.; Maroulas, V.; Vogiatzis, K. D. Representation of molecular structures with persistent homology for machine learning applications in chemistry. *Nat. Commun.* **2020**, *11* (1), 3230.
(263) Otter, N.; Porter, M. A.; Tillmann, U.; Grindrod, P.; Harrington, H. A. A roadmap for the computation of persistent homology. *EPJ. Data Science* **2017**, *6* (1), 17.
(264) Suwayyid, F.; Hozumi, Y.; Feng, H.; Zia, M.; Wee, J.; Wei, G.-W. CAKL: Commutative algebra k-mer learning of genomics. *arXiv* **2025**, DOI: 10.48550/arXiv.2508.09406.
(265) Hiraoka, Y.; Nakamura, T.; Hirata, A.; Escolar, E. G.; Matsue, K.; Nishiura, Y. Hierarchical structures of amorphous solids characterized by persistent homology. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (26), 7035−7040.

https://doi.org/10.1021/acs.chemrev.5c01098
Chem. Rev. 2026, 126, 7029−7067