# Quantification of Crystal Packing Similarity from Spherical Harmonic Transform

Qiang Zhu,[1, a)] Weilun Tang,[1] and Shinnosuke Hattori[2]
[1)]*Department of Physics and Astronomy, University of Nevada Las Vegas, NV 89154, USA*
[2)]*Advanced Research Laboratory, R&D Center, Sony Group Corporation, 4-14-1 Asahi-cho, Atsugi-shi 243-0014, Japan*

(Dated: 11 August 2022)

In this work, we present a new computational approach to characterize and classify molecular packing in the solid states. The key idea is to project each neighboring molecule (or short contact) from the centered molecule into a unit sphere according to the interaction energy. Consequently, the similarity between two spherical images can be evaluated from the spherical harmonics expansion based on the maximum cross-correlation. We apply this approach to successfully reproduce the previous packing assignment on a small amount of data with an improved categorization. Furthermore, we conduct a packing similarity analysis over 2000 hydrocarbon crystal data sets and uncover a set of abundant packing motifs. Unlike the previous approaches based on the subjective visual comparison at the real space, our approach provides a more robust way to measure the packing similarity, thus paving the way for a rapid classification of large scale crystal data.

## I. INTRODUCTION

Molecular solids have been extensively used as key ponents in chemical industries such as medicine[1], fe ers, dyes[2], pesticides[3], and high energy explosives[4], as as electronics industry since the discovery of bulk co tivity in polycyclic aromatic compounds in 1950s[5]. some molecules can aggregate with no particular order as amorphous solids, most organic solids are crystallin their physical properties largely depend on regularly r ing intermolecular packing. Nevertheless, understandi packing of molecules is a rather elusive subject due to th ety of molecular shapes. Unlike the packing of spheres atomic crystals, geometry analysis on the irregularly s molecules appears to be more of an art than a science visually difficult to capture the pattern even for experi researchers[6].

To ease the visual challenge, most early works focused on the simple shapes. Among them, Robertson first proposed to divide planar aromatic hydrocarbons (PAH) into two categories based on the ratio of molecular area to the thickness[7]. It was found that the disk-like molecules tend to stack together via rigid translation (*stack-promoting*), while the molecules with smaller areas prefer the glide-like reflections (*glide-promoting*). Using energetic as well as geometrical criteria, Desiraju and Gavezotti[8] extended the categorization of PAHs into four groups and used them as a guide for prediction of crystal packing for new molecules. As shown in Fig. 1, they include (i)*herringbone*, in which each molecular column has interactions with the nonparallel neighboring molecules; (ii) *sandwich*, consisting of the herringbone motif with sandwich-type diads; (iii) $\gamma$, the flattened-out herringbone; and (iv) $\beta$, a layered structure made up of graphite-like planes.

The packing motif concept has been applied to establish the correlations between these materials' packing and

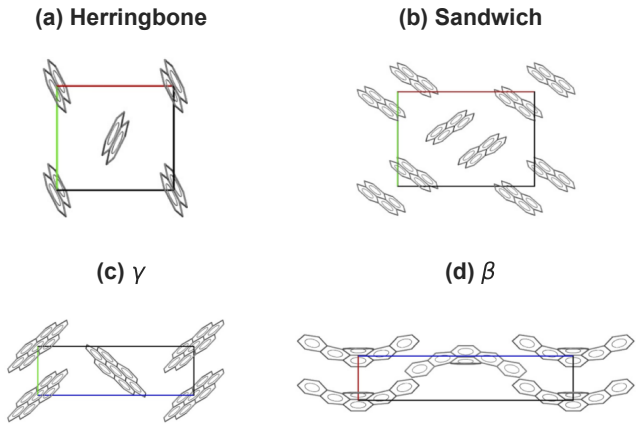


FIG. 1. Four packing motifs on the crystals made of planar aromatic hydrocarbons (PAHs).

observed physical properties (e.g., charge transport for organic semiconductors[9,10] and insensitivity for molecular explosives[11]). Despite its popularity in crystal analysis, the definition of four patterns, like many other chemical nomenclatures, lacks a mathematical rigor. From the definition, it is hard to define the boundary between $\gamma$ and $\beta$. Even within the herringbone group, some crystals clearly have different packing as compared to others. Although several tools have also been developed to automate the assignment of packing motif[9,12,13], they sometimes yield inconsistent results due to the subjective choice of projection plane or simply due to the ambiguity of the definition itself.

In addition, the Hirshfield fingerprint has been commonly used to analyze the complex information contained in a molecular crystal structure into a single, unique full colour plot[14]. Derived from the Hirshfeld surface, these 2D-fingerprint plots provide a visual summary of the frequency of each combination of internal and external distances across the surface of a molecule, so they do not only indicate which

[a)]Electronic mail: qiang.zhu@unlv.edu

intermolecular interactions are present, but also the rela area of the surface corresponding to each kind of interact Other analysis techniques[15–17] were also proposed recer However, the interpretation is either too abstract or too cı bersome, and they also lack a simple criterion to judg two crystals are truly similar or not. On the other hand, COMPACK algorithm[18] is commonly used to check if structures are identical or not when they consist of the sim molecules. Nevertheless, it is not suitable to compare the s ilarity between two crystals with different constituents, wh is important in many crystal engineering applications.

To address the aforementioned challenge, we introduce image approach in combination with spherical harmonics pansion to quantify the packing similarity of organic cryst This essentially involves a two-step process. First, we construct the spherical images to represent the molecular packing and energy distribution in a crystal. Second, the similarity of two spherical images (i.e., crystal packing) is computed by the maximum normalized cross correlation in the Fourier space. In the following, we present the methodology and its application to a set of hydrocarbon crystal data. Due to its mathematical rigor, this approach can be fully automated for a rapid classification of distinct crystal packing motifs in a large data set, regardless of the molecular choices.

## II. COMPUTATIONAL METHODOLOGY

### A. Image Representation of Molecular Packing

Although it is highly subjective to describe the crystal packing, there is a general consensus that at least two factors are of critical importance. First, the geometry has been widely used to understand the packing. The simplest analysis is to count the coordination number as a function of the cutoff distance. Moreover, the spatial distribution of neighboring molecules can also analyzed with more sophisticated tools[19,20]. However, crystal packing is not simply a geometry problem. The intermolecular energy should be taken into account as well, as described by Desiraju and Gavezotti[8], as well as the 2D-fingerprint of the Hirshfeld surface[14]. Using the naphthalene crystal (referred as NAPHTA in the Cambridge Crystal Database) as an example, a molecule can be generally considered to have 16 neighbors (see Fig. 2). Based on the intermolecular interactions, these molecules can divided to four tiers, colored in yellow (6), green (6), light blue (2) and deep blue (2). The last tier is often omitted because their interactions are much weaker. One can consider that the most important yellow tier molecules form the equator while the less important green and blue molecules are distributed on the Southern and Northern Hemispheres by following some pattern. Therefore, a good representation should be able to probe such characteristics of the spatial distribution of both molecules and energetics.

Inspired by those previously developed approaches, we construct the spherical image to represent the packing as follows,

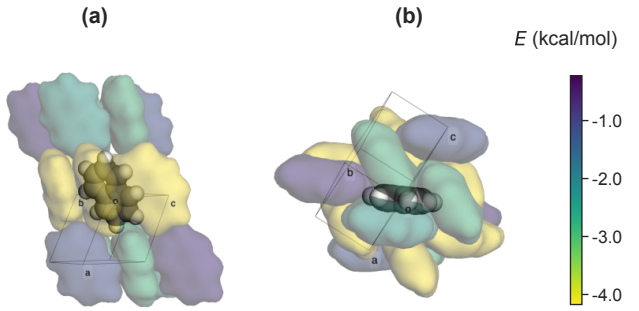


FIG. 2. Different projections of NAPHTA crystal from the top (a) and side (b) views. The center molecule is shown as spheres while the surrounding molecules are shown by their van de Waals surface colored by the interaction energy.

1. Choose the center molecule ($M_0$) and find the neighboring molecules or intermolecular short contacts $\{M_i\}$.

2. For each $M_i$, compute the bonding energies ($\{E_i\}$) with respect to $M_0$. If $M_i$ denotes a molecule, $E_i$ is the sum of all interactions between $M_i$ and $M_0$.

3. Project each $M_i$ onto the unit sphere centered at $M_0$ with the spherical coordinates $d_i = (\theta_i, \phi_i)$. If $M_i$ denotes a short contact (e.g., C$\cdots$C or C$\cdots$H pairs), the center of the atomic pair is used as the reference point to determine $d_i$. When $M_i$ denotes a molecule, the molecular center shall be used as the reference.

4. For each $M_i$, place a Gaussian based on $E_i \exp \frac{(d-d_i)^2}{2\sigma^2}$ in the unit sphere grids.

Therefore, we come up with two definitions of spherical functions as follows:

$$f = \begin{cases} \sum_{i=1}^{\text{molecules}} E_i \exp \frac{(d-d_i)^2}{2\sigma^2} & \text{Coarse grained model} \\ \sum_{i=1}^{\text{contacts}} E_i \exp \frac{(d-d_i)^2}{2\sigma^2} & \text{Fine resolution model} \end{cases} \quad (1)$$

In general, $f_{\text{molecule}}$ can be considered as a coarse-grained descriptor for the molecular packing. Fig. 3a shows the representation for a naphthalene crystal (see Fig. 2). Unlike other descriptors such as Hirshfield fingerprint plots[14], the spherical image can be easily interpreted since it retains the spatial information. For instance, the center of each hot spot denotes the position of the neighboring molecule. In NAPHTA, one can clearly find six bright spots at the equator plane, and four less bright spots at each of the hemispheres. This information is consistent with our analysis from Fig. 2. It is essentially the arrangement of hot spots in the sphere. For such a spherical image, it is also common to plot its cylindrical projection as used in the map of the Earth. On the other hand, $f_{\text{contact}}$, shown in Fig. 3b is a break-down version of $f_{\text{molecule}}$ or a projection of the Hirshfield surface into the unit sphere[16]. It can capture more details when describing the molecular packing. As one can see in the projection map in Fig. 3b, some bright

spots becomes rather diffusive, indicating that the intermolecular interaction are from multiple short contacts between the center and surrounding molecules, while the very hot spots suggest that there exist very strong interactions such as C···H bonds for the PAH system, or hydrogen bonds for the gen-

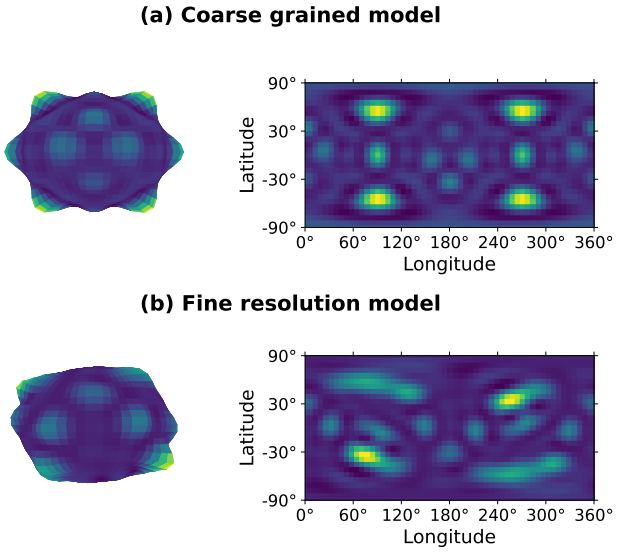


FIG. 3. Two sphere image representations for NAPHTA, (a) the coarse grained model and (b) the fine resolution model.

When computing the intermolecular energy, one can use different approaches from empirical force fields to more advanced electronic structure methods. In our case, we follow an empirical atom-atom potential[21] due to its simplicity. In addition, it is important to note that the choice of Gaussian smearing is to ensure a smooth distribution when comparing two distinct local environments. As such, it can be also used to characterize the packing environments in an aggregate, amorphous solid, or even liquid. In the following, we will keep $\sigma$ = 0.1 Å at the unit sphere for all the subsequent calculations. To construct the spherical image, we uniformly sampled 10,000 grids on the unit sphere according to the Fibonacci method, which is more efficient than the evenly spaced grid sampling on $(\theta,\phi)$.

### B. Real Spherical Harmonic Expansion

For a real valued spherical function $f(\theta,\phi)$, it can be expressed as a series of spherical harmonic functions,

$$f(\theta,\phi)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l} f_{lm}Y_{lm}(\theta,\phi), \tag{2}$$

where $f_{lm}$ is the spherical harmonic coefficient, $Y_{lm}$ is the spherical harmonic function, and $l,m$ are the spherical harmonic degree and order. The details of $Y_{lm}$ are given in the appendix.

Similar to the Fourier transform of a one-dimension spectrum, the purpose of spherical harmonic expansion is to efficiently extract the features from the frequency domain with only a few $f_{lm}$ coefficients. It is straightforward to show that $f_{lm}$ can be calculated by the integral

$$f_{lm}=\frac{1}{4\pi}\int_{\Omega} f(\theta,\phi)Y_{lm}(\theta,\phi)d\Omega \tag{3}$$

In general, $f_{lm}$ consists of $2\times l\times l$ real valued numbers. Fig. 4a shows the power spectrum of $f_{lm}$ obtained from the naphthalene crystal. Compared to the grids in the real space, these $f_{lm}$ coefficients provide an economical way to store all information to reconstruct the images. With the increase of spherical harmonic degree $l$, it is expected that more details of the function can be described. However, the $f_{lm}$ at very high order simply means the noise. Therefore, there should be a cutoff when it is found that further increase of $l$ no longer brings any notable improvement.

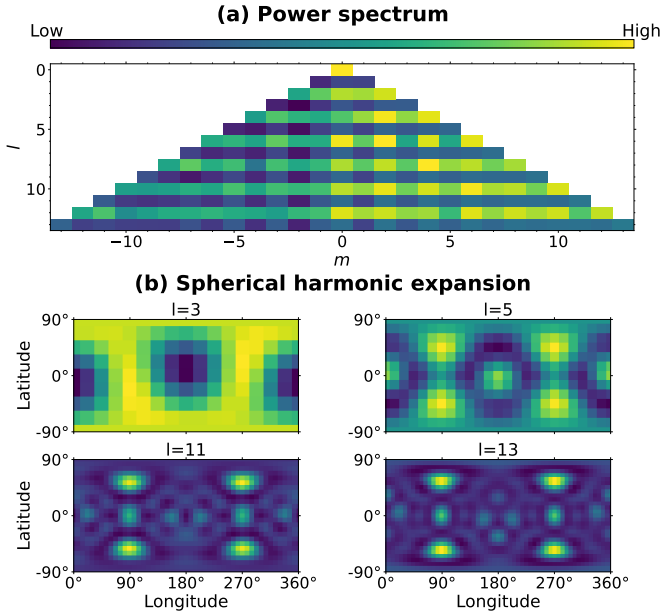


FIG. 4. Spherical harmonic expansion. (a) shows the spectrum as a function of spherical harmonic degree ($l$) and order ($m$. (b) plots the reconstructed images from different cutoff values of $l$.

In practice, we construct the spherical image by sampling 10,000 grids on the unit sphere according to the Fibonacci method, and compute the expansion coefficients from eq. 3. Fig. 4b plots the recovered images from different $l_{\max}$ values. In this work, we choose $l_{\max}$ = 13 that should be sufficient to recover all spherical images in high accuracy.

### C. Cross-correlation on the Sphere

Having established the compact description for one crystal packing ($f$), we can apply the same scheme to describe another crystal named $g$. Obviously, we arrive at the point to address the original question, namely, how to quantify the

similarity between $f$ and $g$? It is clear that the use of expansion coefficients is advantageous to reconstruct the original function. Therefore, we wish to compute the similarity based on these coefficients.

One possible solution is to derive a set of rotation-invariant arrays. For instance, the power spectrum has been popularly used in distinguishing the local environments of an atomic crystal[19,22]. However, it can lead to a substantial dimensionality reduction from the original coefficients ($2 \times l \times l$) to a power spectrum ($l+1$), This indicates that a lot of information is lost during this process. Indeed, we found this is not sufficient since many different packing patterns may share a similar power spectrum.

To avoid unnecessary information loss, we aim to seek a better metric by considering all $f_{lm}$ coefficients when comparing the similarity. For image similarity analysis, a common way is to analyze the cross-correlation spectrum spectrum between two functions $f$ and $g$,

$$C_l(f,g) = \int f(\Omega) g(\Omega) \tag{4}$$

From the cross-correlation, we can further define a scalar metric at the Fourier space that is bounded between 0 and 1,

$$S(f,g) = \frac{\sum_{lm} f_{lm} g_{lm}}{\sqrt{\sum_{lm} f_{lm}^2 \sum_{lm} g_{lm}^2}} \tag{5}$$

According to the definition, $f$ and $g$ is identical if $S = 1$. When $S$ has a strong deviation from 1, it means $f$ and $g$ are less similar. However, it is important to note that the coefficients are subject to change under a rotation ($R$, which can also be denoted by a set of Euler angles $\{\alpha, \beta, \gamma\}$). Therefore, we seek to maximize $S$ by sampling all possible rotational space of the SO(3). In practice, we perform deterministic quasi random sampling from the low-discrepancy Sobol sequence[23] to generate a uniform grids of $\{\alpha, \beta, \gamma\}$ on the sphere. Then we rotate $g$ accordingly, and $S$ is further maximized based on the derivative-free optimization method as implemented in the Scipy.optimize tool box[24]. Finally, the search returns the $S_{\max}(f,g)$, a scalar number between 0 and 1, as well as the rotation ($R$) on $g$ that achieves the best match. Fig. 5 shows the initial and optimized rotation for PHENAN when it is compared with naphthalene by following this procedure. For the initial state in Fig. 5a, it returns $S = 0.253$. However, the true $S$ becomes 0.931 after the best rotation is found.

## III. RESULTS

With this new image approach, we proceed to build the first link between inorganic and organic crystallography. Considering hydrocarbon molecules as the simplest category, like elemental allotropes in the atomic crystals, can we classify them by some packing motif? If so, what will be the most common motifs, like body-centered-cubic (bcc), face-centered-cubic (fcc), hexagonal-close-packing (hcp) and diamond types? In the following, we will first validate the approach in a small set of previously investigated systems and then present our own categorization for a more extended data consisting of over 2000 crystals.

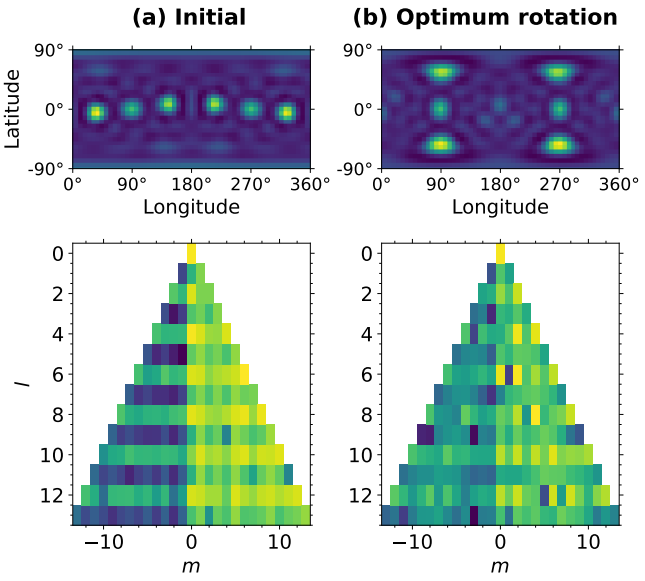


FIG. 5. The spherical function's projection and power spectrum of PHENAN before (a) and after (b) rotation with respect to NAPTHA.

### A. Regroup of 30 PAH crystals

In 1989, Desaraju and Gavezotti have defined four prototypical packing for about 30 crystals of disk-shaped PAH molecules[8]. Their pioneering work has significantly influenced the following studies in crystal engineering. With this new image approach, we revisited this data set. We calculated the similarity function for each structure pair. From the computed similarity matrix, we performed an unsupervised hierarchical agglomerative clustering, Fig. 6 displays the resulting dendrogram plot of the 30 PAH crystals. Clearly, our results largely agree with the previous assignment[8]. Out of 30 samples, we found that 24 of them are closely clustered into three groups, corresponding to the previously assigned herringbone, sandwich and $\gamma/\beta$ types. This is encouraging, given that our calculation is fully automated without any supervision. From these results, we can also analyze each cluster in detail.

**Herringbone** is the most common group. Using NAPHTA as the reference, we found that most crystals within this group (colored in red in Fig. 6) have rather high $S$ values, including CRYSEN (0.975), ANTCEN (0.974), ZZZOYC01 (0.945), QUPHEN (0.941), BIPHEN (0.938), PHENAN (0.931), BEANTR (0.896).

**Sandwich** was the least common group with only five examples in its original proposal[8]. Compared to the herringbone group, the $S$ values between each member are smaller. However, they can be clearly grouped to the same big branch (colored in black in Fig. 6) with the unsupervised clustering method, indicating a strong correlation. According to our

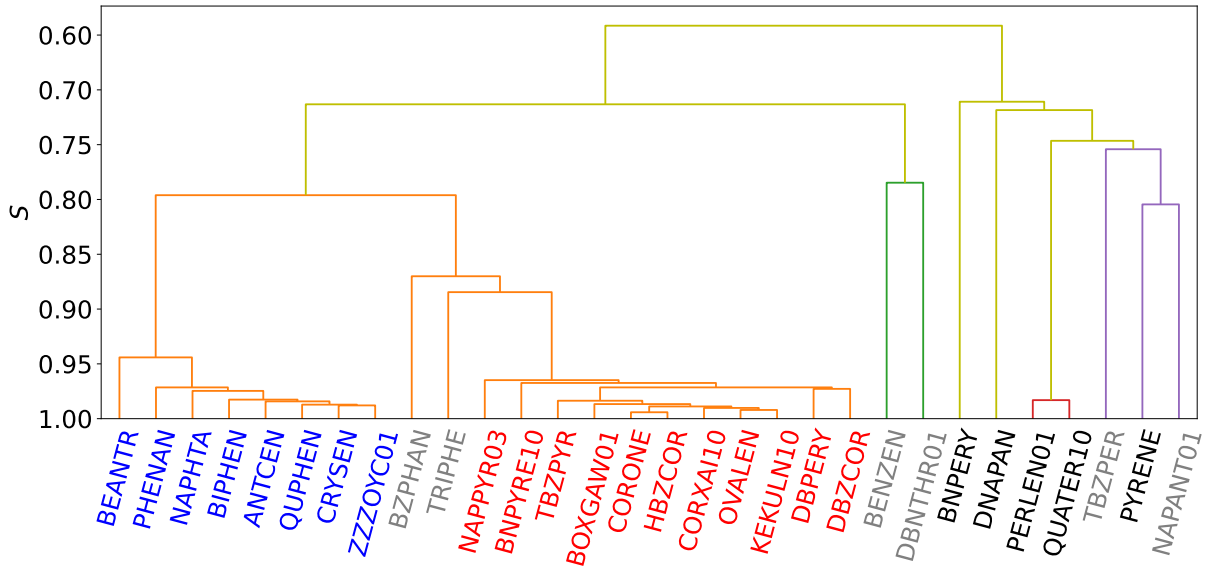


FIG. 6. The hierarchical clustering of 30 PAH crystals. The structures belonging to herringbone, sandwich are colored in blue and black, respectively. The $\gamma$ and $\beta$ types are indistinguishable, and hence they are all colored in red. In addition, several structures that cannot be matched to the previous assignments are colored in grey.

analysis, there actually exist two types of sandwich crystals. The details will be discussed in the following section.

***γ/β*** are almost identical if one only focuses on the arrangement of molecular centers in Figs. 1c-d. They can be only distinguished by counting the ratio of C$\cdots$C and C$\cdots$H interactions[8]. Indeed, the converted coarse grained tends to group them together (colored in red in Fig. 6) with the high threshold value among three major clusters. However, one can always use the fine resolution model to search for really matched crystals.

Despite the overall agreement, some structures, colored in grey in Fig. 6 cannot be assigned to the same groups as suggested in the previous literature. For instance, the benzene crystal (BENZEN) was grouped to the herringbone type, but the authors also mentioned that it was more like an outlier[8]. Similarly, several other crystals (including BZPHAN, TRIPHE, DBNTHR01, TBZPER, NAPANT01) clearly exhibit different patterns from any of four types as shown in Fig. 1. The differences can be directly detected by a trivial visual analysis. Therefore, our image approach is more advantageous since it can provide a robust way to detect such outliers without ambiguity.

## B. Common prototypes on a larger data set

Encouraged by the successful application on the 30 PAH crystals, we systematically inquired all crystals from the Cambridge Structural Database by searching for the systems containing only C and H elements and no more than one molecule in the asymmetric unit. After removing the duplicate entries, we obtained 2007 crystals (see more details in the supplementary materials). This data set serves as the test bed to further validate our image approach.

### 1. *The naphthalene family*

In the early days of small molecule crystallography, the naphthalene crystal was considered to be the common salt of organic crystal[25]. This type of crystal packing was assigned to the herringbone group, which was found to occur more often than any other group. Fig. 7a plots the histogram of the distances between NAPHTA with all other hydrocarbon crystals. The entire distribution can be roughly described by some normal distribution with a mean around 0.6-0.7 (more details can be found in the supplementary materials). Here we focus on the region with high $S$ values. Among the entire data set, we found CEKWEU (Fig. 7b) has the best match with NAPHTA. Although the molecule in CEKWEU is not flat, the mapping is obvious as they share the same space group ($P2_1/c$) and Wykcoff position, as well as similar molecular orientation. In addition, Figs. 7c-e displays several structures with different $S$ values. By analyzing these patterns as well as the 3D structure, we decide to set a threshold of 0.892, leading to 86 crystals belonging to the naphthalene family.

It is important to note that our similarity calculation does impose any symmetry constraints. As shown in Table I, several space groups, e.g., $P2_1/c$, $Pbca$, $P\bar{1}$, $P2_1$, occur more often. Clearly, all of them follow group-subgroup relations. For instance, the molecules containing four or more aromatic rings crystallize in the $P\bar{1}$ with Z=2 (e.g., TENCEN, PENCEN), but the overall packing is still close to that of NAPHTA. It would be interesting to collect more data and use them to predict the possible phase transition.

Finally, we emphasize that the coarse-grained image model

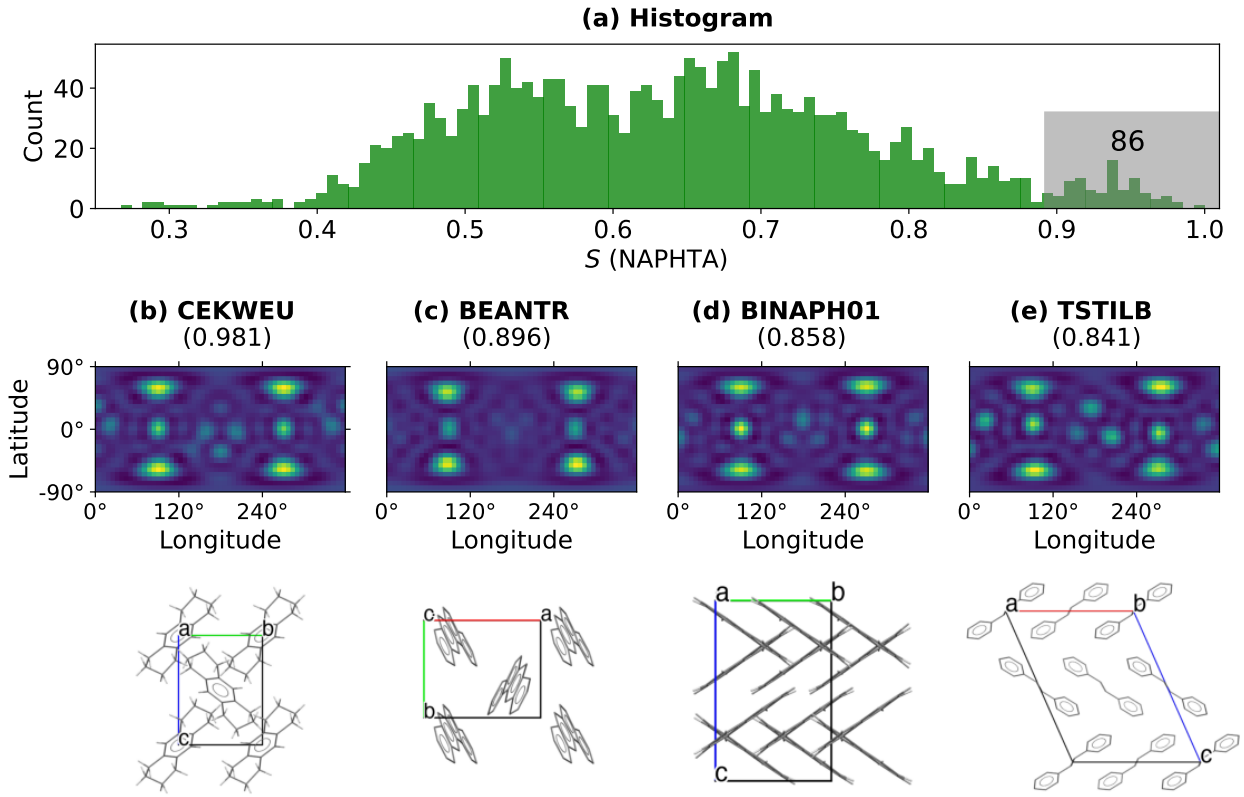


FIG. 7. The naphthalene crystal. (a) is the similarity ($S$) distribution of all PAH crystals with respect to NAPHTA. (b)-(e) show several representative crystals with different $S$ values.

TABLE I. The space group distribution of 86 naphthalene crystals.

| Space group | # of molecules per cell | Occurrence |
|---|---|---|
| $P2_1/c$ | 2 | 34 |
| $Pbca$ | 4 | 14 |
| $P2_1/c$ | 4 | 9 |
| $P\bar{1}$ | 2 | 7 |
| $P2_1$ | 2 | 5 |
| $Pnma$ | 4 | 5 |
| $P2_12_12_1$ | 4 | 3 |
| $Aea2$ | 4 | 2 |
| $C2/c$ | 4 | 2 |
| $Pbcn$ | 4 | 2 |

is still limited by its molecular approximation. Therefore, it is not able to detect the difference in terms of molecular shape. This can be corrected by the fine-resolution model. If we simulate the similarity based on $f_{\text{contact}}$, only three crystals ANTCEN (0.917), CEKWEU (0.834), DMANTR (0.833) can match NAPHTA well. Therefore, we suggest the use of $f_{\text{contact}}$ when it is necessary to find a really good match on a reference crystal.

### 2. *Other common prototypes*

We also checked other common prototypes. The results are summarized in Fig. 8.

**BENZEN**. The crystal structure of benzene is an outlier of herringbone and was less investigated in terms of packing. It was considered to have a pseudo-fcc arrangement due to strong X-ray reflections at the (111) plane[26]. However, from the projection image as shown in Fig. 8a, such a type can also be explained to have four strongly interacted molecules surrounding the center molecule forming the equator plane, while four secondary interactions are evenly distributed in each of the Northern and Southern Hemispheres. In our explanation, we take into account the fact that the 12 neighboring molecules do not have the same interactions with the centering molecules, which is different from a typical fcc atomic crystal. To our knowledge, there are no previous reports on the crystal sharing the same packing pattern with BENZEN. Here our analysis suggests that at least 14 crystals belong to this family with the threshold distance value of 0.89, such as NUNCUW (0.966) and FAPZOL (0.962).

**Sandwich**. As shown in Figs. 8b-c, both PYRENE and PERlEN01 has one very bright spot in the upper left of the projection image, however, other less bright spots notably differ from each other. Such subtle differences can be easily ignored when one just focuses on the common projec-

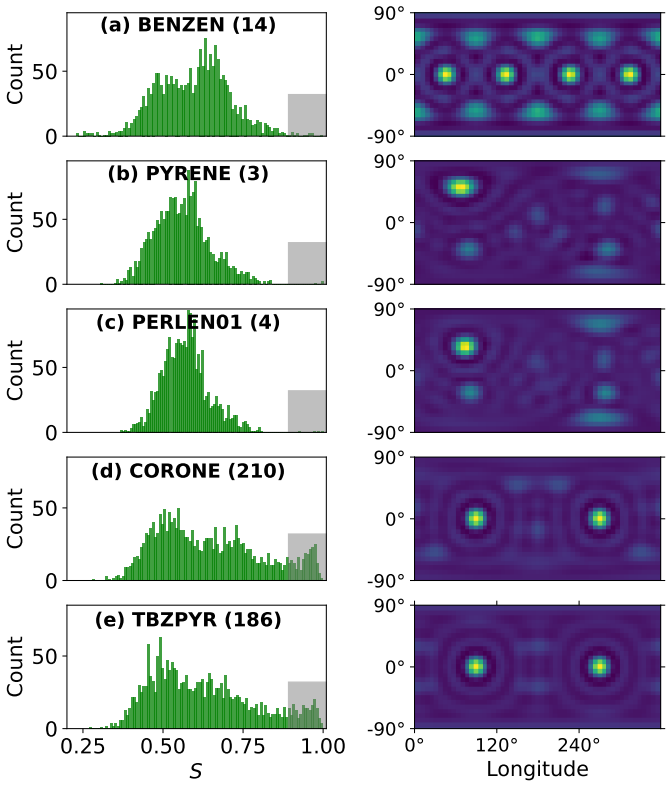


FIG. 8. Other representative crystals, including (a) benzene type, (b) sandwich type 1, (c) sandwich type 2, (d) $\gamma$ and (e) $\beta$. The numbers in the parenthesis indicate the number of similar structures identified in the 2007 data set.

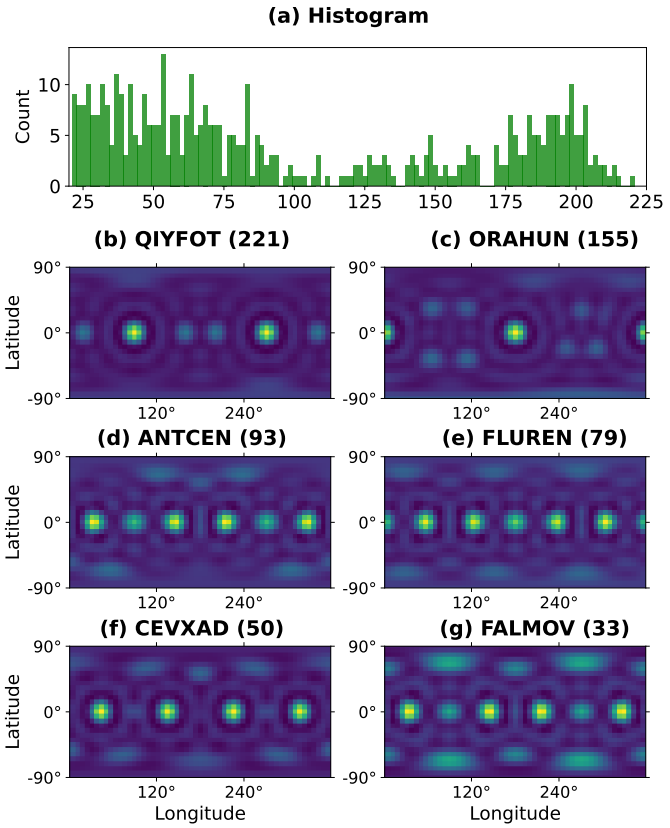


FIG. 9. The complete clustering results and several abundant packing motifs. (a) is the histogram of similar structures for each crystal with a threshold of 0.88. (b)-(h) show six representative structures' patterns featured by two, four, and six bright spots in the projected equatorial plane. The numbers in the parenthesis indicate the number of similar structures identified in the 2007 data set.

tion plane. The $S$ between PYRENE and PERLEN01 is only 0.753. Among the entire dataset, such packings do not appear often, with only 3 PYRENE and 4 PERLEN03 types.

***$\gamma/\beta$***. Two representative crystals (CORONE and TBZPYR), as shown in Figs. 8d-e, are nearly identical in terms of the strong characteristics. Both CORONE (210) and TBZPYR (186) can find a lot of structures sharing the same packing pattern. The $S$ value between CORONE and TBZPYR is 0.975. Therefore, these two families are expected to have a lot of overlaps.

### 3. *A complete categorization*

We also performed a systematic pairwise distance analysis for the entire 2007 crystal data set. Using the threshold of 0.88, we count the number of similar structures for each crystal as shown in Fig. 9a. From the distribution, it is clear that many structures possessing the pattern of two strong bright spots (see Figs. 9b-c) are most abundant. This pattern is also consistent with what we have observed in $\gamma/\beta$. However, we note that other neighbors in the crystal can adopt different packing sequences. Therefore, there exist many subtypes within this family. The second large group is featured by the pattern of six bright spots forming the equatorial plane (see Figs. 9d-e)). Therefore, the most important feature of this group is that the molecules form close packing layers with strong intermolecular interactions[27]. Within this group, ANTCEN belongs to the common naphthalene family. However, the projection map in Fig. 9d is rotated to put the six bright spots in the equatorial plane. Another frequent-occurring subgroup is shown in Fig. 9e of FLUREN. The structure differs from ANTCEN by the relative shifting of molecules in the adjacent layers. As shown in Fig. 10, the strong similarity between ANTCEN and FLUREN can be easily detected in real space. However, it can also be seen that ANTCEN follows the herringbone pattern and FLUREN adopts the sandwich type from other choices of projection. This drastic difference clearly suggests the limitation of interpreting structure by visual projection. On the other hand, our spherical image representation provides more rigorous comparison by capturing both similarity and dissimilarity at all directions. Finally, we also observe another group of structures with four bright spots in the equatorial plane (see Figs. 9g-h). This is less common. Each structure of this group has only 30-50 similar structures within the 2007 data set.

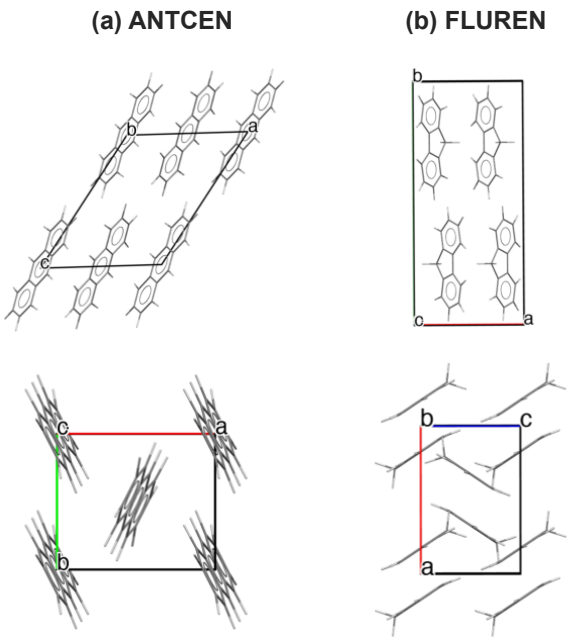


FIG. 10. Crystal packing comparison between (a) ANTCEN and (b) FLUREN.

## IV. CONCLUSIONS

In sum, we present an image approach to characterize and classify molecular packing from the local neighboring environment. The similarity between two crystals (i.e., images) can be evaluated from the spherical harmonics expansion based on the maximum cross-correlation. We apply this approach to investigate the packing similarity analysis over 2000 C-H crystal data sets and categorize them by similarity. Compared to the previous approaches based on visual comparison in the real space, our spherical harmonics expansion provides a robust way to measure the packing similarity. It is suitable for a rapid search for similar crystal structures by crystal packing from a large data set. For instance, there is a general perception that herringbone is high-mobility[28–30]. Our tool can clearly be applied to search for new candidates with a similar packing pattern. In the future, a more extensive crystal packing motif library can be built by using this criterion. When such templates becomes available, it can map the relation between molecule properties and crystal packing to guide the design of new materials. Alternatively, it can also be used to complement the crystal structure prediction by chemical substitution of the well defined structural prototypes[31].

## ACKNOWLEDGMENTS

Q.Z. acknowledge the NSF (DMR-2142570) and Sony Group Corporation for their financial supports. The computing resources are provided by XSEDE (TG-DMR180040).

## Appendix A: Real Spherical Harmonics

The real spherical harmonics are defined as

$$Y_{lm}(\theta,\phi) = \begin{cases} \bar{P}_{lm}(\cos\theta)\cos m\phi, & \text{if } m \geq 0 \\ \bar{P}_{l|m|}(\cos\theta)\sin|m|\phi, & \text{if } m < 0 \end{cases} \tag{A1}$$

where normalized associated Legendre functions with the $4\pi$-normalized spherical harmonic functions are given by

$$\bar{P}_{lm}(\mu) = \sqrt{(2-\delta_{m0}(2l+1)\frac{(l-m)!}{(l+m)!}}P_{lm}(\mu) \tag{A2}$$

and $\delta_{ij}$ is the Kronecker delta function.

The unnormalized associated Lengendre functions are from the following relations

$$\begin{aligned} P_{lm}(\mu) &= (1-\mu^2)^{m/2}\frac{d^m}{d\mu^m}P_l(\mu) \\ P_l(\mu) &= \frac{1}{2^l l!}\frac{d^l}{d\mu^l}(\mu^2-1)^l \end{aligned} \tag{A3}$$

In this work, we strictly follow the definitions used in the `pyshtools` package[32]. The image conversion and optimization functions have been implemented in the `PyXtal` package[33].

## CONTRIBUTIONS

O.Z. conceived the idea, wrote the code, and performed the simulation; W.T. participated in code development, Q.Z and H.S wrote the paper.

## DATA AVAILABILITY

The data sets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

## CODE AVAILABILITY

The codes used to calculate the results of this study are available in https://github.com/qzhu2017/PyXtal.

## CONFLICT OF INTEREST

Q.Z. have received research funding from Sony Group Corporation.

## REFERENCES


[1] A. Y. Lee, D. Erdemir, and A. S. Myerson, "Crystal polymorphism in chemical process development," Annu. Rev. Chem. Biomol. Eng. **2**, 259–280 (2011).

[2] Z. Zhuo, C. Wei, M. Ni, J. Cai, L. Bai, H. Zhang, Q. Zhao, L. Sun, J. Lin, W. Liu, et al., "Organic molecular crystal with a high ultra-deep-blue emission efficiency of 85% for low-threshold laser," Dyes and Pigments , 110425 (2022).

[3] J. Yang, C. Hu, X. Zhu, Q. Zhu, M. D. Ward, and B. Kahr, "Ddt polymorphism and the lethality of crystal forms," Angew. Chem. **129**, 10299–10303 (2017).

[4] G. Liu, R. Gou, H. Li, and C. Zhang, "Polymorphism of energetic materials: a comprehensive study of molecular conformers, crystal packing, and the dominance of their energetics in governing the most stable polymorph," Cryst. Growth Des. **18**, 4174–4186 (2018).

[5] H. Kallmann and M. Pope, "Bulk conductivity in organic crystals," Nature **186**, 31–33 (1960).

[6] G. M. Day and W. S. Motherwell, "An experiment in crystal structure prediction by popular vote," Cryst. Growth Des. **6**, 1985–1990 (2006).

[7] J. M. Robertson, "The measurement of bond lengths in conjugated molecules of carbon centres," Proc. R. Soc. London Ser. A **207**, 101–110 (19511).

[8] G. R. Desiraju and A. Gavezzotti, "Crystal structures of polynuclear aromatic hydrocarbons. classification, rationalization and prediction from molecular structure," Acta Cryst. B **45**, 473–482 (1989).

[9] J. E. Campbell, J. Yang, and G. M. Day, "Predicted energy–structure–function maps for the evaluation of small molecule organic semiconductors," J. Mater. Chem. C **5**, 7574–7584 (2017).

[10] N.-X. Zhang, A.-M. Ren, L.-F. Ji, S.-F. Zhang, and J.-F. Guo, "Theoretical investigations on molecular packing motifs and charge transport properties of a family of trialkylsilylethynyl-modified pentacenes/anthradithiophenes," J. Phys. Chem. C **122**, 18880–18894 (2018).

[11] D. Mathieu, "Sensitivity of energetic materials: Theoretical relationships to detonation performance and molecular structure," Ind. Eng. Chem. Res. **56**, 8191–8201 (2017).

[12] D. Loveland, B. Kailkhura, P. Karande, A. M. Hiszpanski, and T. Y.-J. Han, "Automated identification of molecular crystals' packing motifs," J. Chem. Inf. Model. **60**, 6147–6154 (2020).

[13] D. Ito, R. Shirasawa, Y. Iino, S. Tomiya, and G. Tanaka, "Estimation and prediction of ellipsoidal molecular shapes in organic crystals based on ellipsoid packing," Plos one **15**, e0239933 (2020).

[14] J. J. McKinnon, F. P. Fabbiani, and M. A. Spackman, "Comparison of polymorphic molecular crystal structures through hirshfeld surface analysis," Cryst. Growth Des. **7**, 755–769 (2007).

[15] W. Motherwell, "Molecular shape and crystal packing: a database study," CrystEngComm **12**, 3554–3570 (2010).

[16] P. R. Spackman, S. P. Thomas, and D. Jayatilaka, "High throughput profiling of molecular shapes in crystals," Sci. Rep. **6**, 22204 (2016).

[17] O. Carugo, O. A. Blatova, E. O. Medrish, V. A. Blatov, and D. M. Proserpio, "Packing topology in crystals of proteins and small molecules: A comparison," Sci. Rep. **7**, 1–12 (2017).

[18] J. A. Chisholm and S. Motherwell, "Compack: a program for identifying crystal structure similarity using distances," J. Appl. Crystallogr. **38**, 228–231 (2005).

[19] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, "Bond-orientational order in liquids and glasses," Phys. Rev. B **28**, 784–805 (1983).

[20] J. D. Honeycutt and H. C. Andersen, "Molecular dynamics study of melting and freezing of small lennard-jones clusters," J. Phys. Chem. **91**, 4950–4963 (1987).

[21] A. Gavezzotti, "Are crystal structures predictable?" Acc. Chem. Res. **27**, 309–314 (1994).

[22] A. P. Bartók, R. Kondor, and G. Csányi, "On representing chemical environments," Phys. Rev. B **87**, 184115 (2013).

[23] S. Joe and F. Y. Kuo, "Constructing sobol sequences with better two-dimensional projections," SIAM Journal on Scientific Computing **30**, 2635–2654 (2008).

[24] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, İ. Polat, Y. Feng, E. W. Moore, J. Vand erPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and S. . . Contributors, "SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python," arXiv e-prints , arXiv:1907.10121 (2019).

[25] A. Kitaigorodsky, Molecular crystals and molecules, Vol. 29 (Elsevier, 2012).

[26] E. G. Cox, "Crystal structure of benzene," Rev. Mod. Phys. **30**, 159–162 (1958).

[27] A. Kitaigorodskii, "Organic chemical crystallography, con sultants bureau: New york, 1961 (originally published in russian by the press of the academy of sciences of the ussr, moscow, 1955); spek, al, single-crystal structure validation with the program platon," J. Appl. Crystallogr **36**, 7–13 (2003).

[28] C. C. Mattheus, A. B. Dros, J. Baas, A. Meetsma, J. L. d. Boer, and T. T. M. Palstra, "Polymorphism in pentacene," Acta Cryst. C **57**, 939–941 (2001).

[29] O. D. Jurchescu, A. Meetsma, and T. T. M. Palstra, "Low-temperature structure of rubrene single crystals grown by vapor transport," Acta Cryst. B **62**, 330–334 (2006).

[30] T. Izawa, E. Miyazaki, and K. Takimiya, "Molecular ordering of high-performance soluble molecular semiconductors and re-evaluation of their field-effect transistor characteristics," Adv. Mater. **20**, 3388–3392 (2008).

[31] A. M. Reilly, R. I. Cooper, C. S. Adjiman, S. Bhattacharya, A. D. Boese, J. G. Brandenburg, P. J. Bygrave, R. Bylsma, J. E. Campbell, R. Car, et al., "Report on the sixth blind test of organic crystal structure prediction methods," Acta Cryst. B **72**, 439–459 (2016).

[32] M. A. Wieczorek and M. Meschede, "Shtools: Tools for working with spherical harmonics," Geochemistry, Geophysics, Geosystems **19**, 2574–2592 (2018).

[33] S. Fredericks, K. Parrish, D. Sayre, and Q. Zhu, "Pyxtal: A python library for crystal structure generation and symmetry analysis," Comput. Phys. Comm. **261**, 107810 (2021).