Ronin Institute for Independent Scholarship
New Jersey, United States
Lightning Talks, October 2023

# Transcript of *Dissecting the Hirshfeld Surface*

Thaicia Stona de Almeida, Ph.D., Ronin Scholar

**Abstract of Transcribed Presentation**

The Hirshfeld surface was introduced by Spackman and Byrom in 1997, based on the Hirshfeld partitioning scheme for electronic contribution in molecular crystals. The Hirshfeld surface analysis and its fingerprint have become since then popular tools for visualization of intermolecular interactions and crystal packing. In this talk, I would like to go over some aspects of the Hirshfeld surface protocol and raise some questions about the method.

## Transcript

[Talk starts, recording starts]

I would like to thank the Ronin institute for this opportunity. I will make a short introduction to myself first, since I have become a scholar just a few months ago, and am still getting introduced to my colleagues.

I am an applied mathematician of formation and experience, master and doctor in Materials Science Informatics. I was born and raised in Brazil, in Sao Paulo, where I studied Mathematics at the University of Sao Paulo. Brazil is a country sitting on ancient, eroded land, which makes it rich in crystal mines. Emeralds and gold can be found in Brazil, amethyst and topaz, and lots of clear quartz. I first visited a crystal mine when I was young, could dig with my bare hands a soft area, rich in perfect little quartz crystals. Later, at the university, I fell in love with Differential Geometry, which is also a national pride in the Brazilian academic scene.

I wanted then to join these two things, crystals and Differential Geometry, but I eventually learned that there was no such road ready for me to take and go in this direction. Maybe it

would not even be of interest. The typical use of Mathematics in Chemistry are algebraic concepts for crystallography, and some Topology for representing chemical structures. It is not so obvious to find continuous, deterministic, differential mathematics, and this is understandable: at atomic scale, the closest theoretical tools one can find are probabilistic versions of simpler classic concepts. The Hirshfeld surface is one of such tools that tries to estimate a geometric locus from quantum principles.

The Hirshfeld surface was proposed by Spackman and Byrom in 1997, based on the Hirshfeld partitioning scheme for electronic contribution in molecular crystals. The method tries to estimate a surface within molecules or molecular crystals, which separate electronic contributions from the delimited area to the whole molecule. For example, a molecule of water could be considered the larger molecule, and the Hirshfeld surface could be used to estimate the region inside the molecule that is likely to contain electrons that originally belong to the oxygen atom. The Hirshfeld surface splits the electronic density into two groups of 50 percent contribution. One of the 50 percent contributed electronic charge is the inside of the surface, and the remaining estimated molecular shape minus the inside of the Hirshfeld surface is the other half. This is a useful tool when the molecule studied is very large and the chosen subset of atoms inside of it is not as simple as one oxygen.

A series of assumptions and approximations were made before the Hirshfeld surface was proposed. Many layers of theories and quantum chemistry approximations support it, theories developed over many decades, starting from the Born-Oppenheimer approximation for the Schrödinger wave function, to the Roothaan-Hartree-Fock atomic wavefunction, which treats a confined set of connected atoms as a linear combination of their orbitals. Then, the Hirshfeld partition of electronic density was developed, and finally, the Hirshfeld surface method was built on all of these previous works.

[slide shows equation 1]

$$w(r) = \frac{\rho_{promolecule}}{\rho_{procrystal}} = \frac{\sum\limits_{A\in\ promolecule} \rho_A(r)}{\sum\limits_{A\in\ procrystal} \rho_A(r)} \tag{1}$$

The molecule or molecular crystal considered is called the procrystal, and its target subset is called the promolecule. The Hirshfeld method computes the spherical atomic electron contribution of the atoms that belong to the promolecule, as well as the total spherical atomic electron contribution from all atoms to the procrystal. Each spherical atomic electron contribution is a function of the radius considered. The density weight function is then defined as the ratio between the promolecule contribution and the procrystal contribution. This density weight function is also a function of the radius, which can then be used to implicitly define the Hirshfeld surface for the isovalue 50 percent.

[slide shows figure 1]

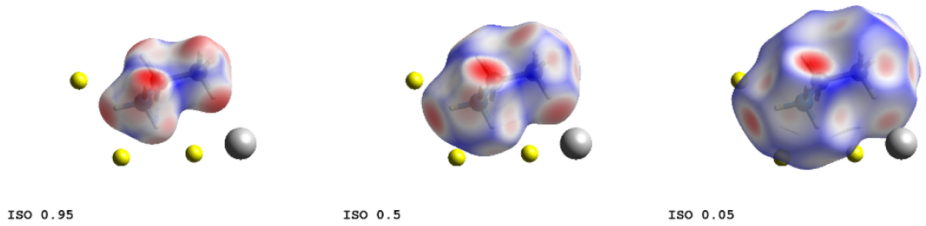


Figure 1: From reference [1], isosurfaces for isovalues 95%, 50%, and 5%.

Note that the implicit isosurface for density isovalue 95 percent is a much smaller surface, very close to the atomic nuclei in the promolecule. The implicity isosurface for the density isovalue of 5 percent on the contrary is very large, since it estimates the region that should, statistically, represent the sum of orbits with less than 5 percent of the contributions original from the promolecule's electrons.

In practice, the surface is a set of thousands of 3-dimensional points that can be interpolated to render a surface. The shortest distances from the points of the surface to the promolecule atomic nuclei are computed, they called internal distances $d_i$. The shortest distances from the points of the surface to the external atomic nuclei $d_e$ are also computed, they are the external distances. We obtain pairs $d_i$ $d_e$, of distances associated to each point of the surface, which are used to build the fingerprint. The fingerprint is a 2-dimensional 2 variable histogram of the distances where the frequency in the z-axis is represented in colors, so that it can be seen in a 2-dimensional plot. In other words, the fingerprint is the color-coded projection of the histogram.

[slide shows figure 2]

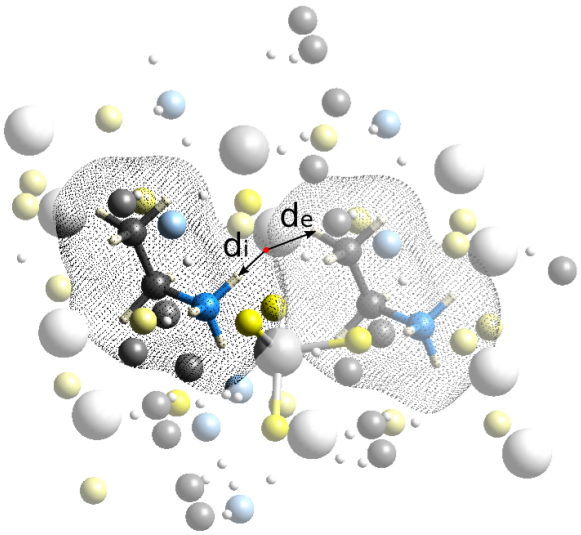


Figure 2: From reference [1], thousands of vertices of a Hirshfeld surface mesh.

The Hirhsfeld surface analysis and its fingerprint have become popular tools for visualization of intermolecular interactions and crystal packing. Rich electronegative areas indicate strong chemical bonding zones. Most applications of the method will only use the surface and the fingerprints as supporting information, visual aid, intuitive confirmation of hypotheses proven by other methods. And even though the official Hirshfeld surface software includes also color-coded geometric properties to be added to the rendered surface, such as curvedness and shape index, there is not much development in the direction of analytical and statistical properties and methods to leverage surface information. There have been a few neural-network based proposed works to try and correlate the fingerprint to material properties.

I have experimented with the marginal distributions of distances in my doctoral thesis, where I correlated the marginals and tested against material properties. It was possible to predict properties within certain classes of perovskites. But the analysis is incomplete and an approach that takes the complete sample should be used.

[slide shows figure 3]

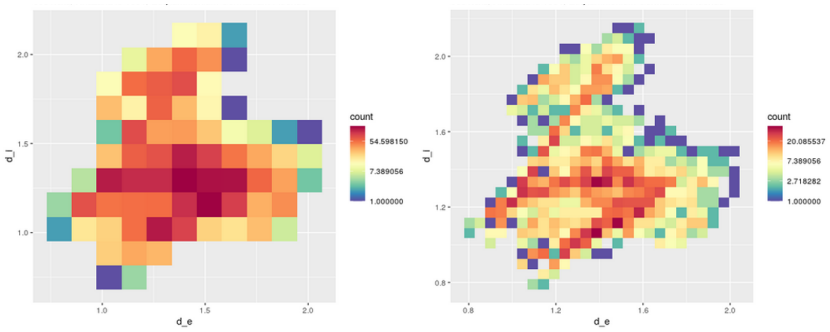


Figure 3: From reference [1], fingerprints generated for different binning numbers.

I started the study with a large database of perovskite crystal structures, several different types of perovskites. A group of hybrid organic inorganic perovskites was especially interesting. These hybrid perovskites are promising, low cost and highly efficient in photovoltaics. Since the first perovskite solar cells were developed, in 2009, the technology has been rapidly improving, reaching the unprecedented power conversion efficiency above 20 percent in 2016. I studied circa 15 hundred hybrid perovskites, computed their surfaces and fingerprints, and tried to find correlations with the materials properties, such as bandgap, atomization energy, crystal unit cell volume, molecular weight, polar area, density, refractive index, and other structural parameters of the crystal.

I had already applied principal component analysis to try to predict the formation of such crystals using only structural parameters and ratios, particularly I used the tolerance factors of Goldschmidt and the octahedral factor. It was not successful, and I tried then to use the factors to predict the materials properties. When such a large set of crystals in the same bag, the predictor was not satisfactory, it gave a very coarse-grained rule to map the properties.

What I did then was to look at the marginal distributions of the fingerprint and try to group them as curves, by simple correlation and clustering. This method worked much more successfully for some classes of materials, but was still indifferent to others. The fingerprint statistics helped predicting the formation of inorganic perovskites. The fact that

it did not work for the whole sample, which contained 1500 crystals, is not remarkable. It is remarkable though that, even though I only picked a slice of the sample histogram, when it worked, it worked well enough to predict formation for a subset of crystals.

I am working on two papers about formal questions on this method, and how to establish protocols to test similarity of surfaces and fingerprints using analytical tools instead of neural networks. There are questions about numerical methods to be asked, it can be observed that the surface points will often be distributed non-parsimoniously, sometimes including several surface points within a very small area, which can potentially predict biased statistics if the overdetermined points are not removed. There is a question of choosing the bin size of the fingerprint, which affect the comparison of Hirshfeld statistics. There is the obvious question of actually trying to treat the Hirshfeld surface as a differentiable manifold: what could we gain from it?

I look back at my first impression of the Hirshfeld surface and still ask myself the same questions, the questions that a mathematician would like to answer. What theorems can I prove here? What are easier ways to parametrize the surface, how reliable is it compared to the true quantum distribution? But my questions might lead to more mathematical beauty and enjoyment than to useful results for materials scientists and chemists.

Biomathematics has been established and so has Mathematical Physics. I think: why is it that Mathematical Chemistry does not occupy the same magnitude of importance in the Applied Mathematics sciences? For now, I will keep working on the beautiful side of this question and worry later about the potential uses.

Thank you for watching my presentation.

[end of recording]

## References

[1] Stona de Almeida, T. (2021). *Structure-Bonding Relationships in Perovskites: Statistics of Hirshfeld Surfaces* [Doctoral dissertation, State University of New York]. ProQuest.