

# パーシステントホモロジーによる材料科学データ解析

理化学研究所革新知能統合研究センター

大林 一平

## Materials data analysis using persistent homology

**Ippei Obayashi**

*Center for Advanced Intelligence Project, RIKEN*

### 1. はじめに

本稿ではデータの幾何的特徴の抽出と定量化に役立つ、パーシステントホモロジー (PH) という手法 [1] について紹介したい。PH は数学のトポロジーという概念にもとづく位相的データ解析 [2] [3] の一手法で、連結性、穴、リング、空隙、といった構造を用いてデータの形をマルチスケールに特徴付ける。

本稿では以下のような疑問に答えることを目標とする。

- PH はどのようにしてデータの幾何的特徴を抽出し定量化するのか？
- この手法の材料科学データ解析への適用例は？
- PH は自分の持っているデータの解析に有用

そうか？

- PH はどのようなデータの解析が得意か？
- PH によるデータ解析ツールはどのようなものがあるか？

そこで本稿では以下のような構成で PH について紹介する。まずは PH の材料科学への適用例について筆者が関連してきたものを中心に紹介する。特にデータの種類について注目して紹介する。2 節では PH の理論的基礎について解説する。3 節では 2 つの活用事例についてより詳しく解説する。これらの事例により PH による実際的な材料データ解析について知ることができよう。4 節では筆者の大林が中心となって開発しているデータ解析ツール「HomCloud」について紹介し、5 節は PH による材料データ解析において直面するいくつかの課題について検討する。最後に 6 節で本稿のまとめとする。

#### 1.1 PH による材料科学データ解析の事例

PH によるデータ解析の事例として代表的な

---

〒606-8501  
京都府京都市左京区吉田本町京都大学情報学研究所  
人工知能研究ユニット  
TEL 075-753-5073  
E-mail: ippei.obayashi@riken.jp

ものとしてはガラス（アモルファス）の原子レベルでの構造解析 [4] [5] [6] [7] が挙げられる。これらの研究では基本的に分子動力学（MD）シミュレーションや逆モンテカルロ法などによって計算された原子配置データを用いている。

論文 [8] では粉体の秩序-無秩序転移について解析している。この研究では基本的には粒状のビーズを容器に詰め込み X 線 CT で計測した空間データを用いている。粉体モデルの数値シミュレーションの結果も PH で解析し、2 つの PH 解析結果の整合性も調べている。

論文 [9] ではガラス質ポリマーの変形によるひびの発生とその成長について解析している。こちらの研究では粗視化 MD によってデータを得ている。

材料科学というよりは生化学の研究であるが、論文 [10] ではタンパク質の構造からの分類問題を取り扱っている。X 線結晶構造解析で得た原子レベルでの構造を入力データとし、機械学習手法も併用して問題に挑んでいる。[11] も同様にタンパク質の構造を PH と機械学習で解析している。

画像データへの適用事例もある。焼結鈷の X 線 CT 画像の解析 [12] やカー顕微鏡で計測した磁気材料の磁区構造 [13] などの事例が挙げられる。

以上のように、様々な材料のデータ解析に PH が用いられている。2 次元 3 次元の構造の解析に主に活用されている。ガラスや粉体の転移現象に利用されているように、結晶のような強い秩序を持つ構造の解析よりは乱れた構造の解

析に大きな力を発揮する。

本稿では 3 節で論文 [4] と [9] の研究についてより詳しく紹介する。また論文 [6] [7] などの研究は本特集の別の論文によってより詳しく解説されているのでそちらも参考にして欲しい。

## 2. パーシステントホモロジー

PH の基礎となっているのは、ホモロジーという数学である。ホモロジーはトポロジーという数学の一分野である。ホモロジーは深い内容を持った数学であるが、その基本的な目標は穴を数えることである。ただ、穴を数えるだけではデータ解析のツールとしては力不足であり、よりデータ解析に向けたツールとして PH が考え出された。

### 2.1 パーシステント図

ここでは図 1 (a) のような点の集合（ポイントクラウド）を例題として、PH について解説する。このポイントクラウドには大小 2 つのリング構造があるように「見える」が、点同士が繋がっているわけではない。そこで図 1 (b) - (e) のように円を各点に置くことで点同士が繋がりリング構造が実際に現れる。

ここで問題なのがこの円の半径で、図の通り半径が小さすぎれば繋がりが切れるし、半径が大きすぎればリングの内側が埋まってしまうリングでなくなってしまう。そこでどうするかというと半径を 1 つに決めるのではなく全ての半径での構造を考えるのである。半径が 0 から大きくなるに従ってリングが発生したり消滅した

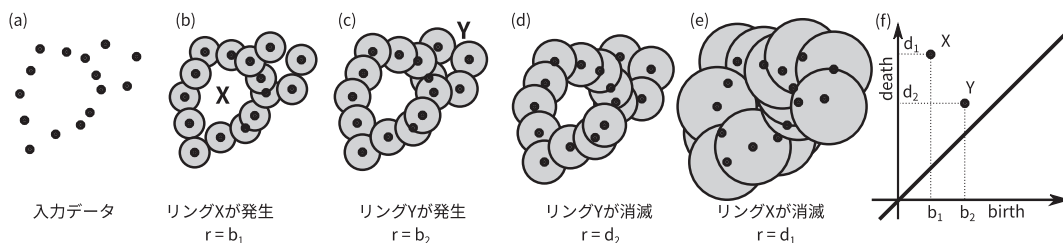


図 1 パーシステント図

りするのを調べるのがPHの重要なアイデアである。この図の場合には (b) で発生したリング X が (e) で消滅し、(c) で発生したリング Y が (d) で消滅する。リングが発生したタイミングの半径を発生時間 (birth time)、消滅の半径を消滅時間 (death time) と呼び、この2つのペアを birth-death pair と呼ぶ。すべてのペアを平面上にプロットしたものをパーシステント図 (PD) と呼ぶ (図1 (f))<sup>1</sup>。

このように半径を徐々に大きくするという過程を考えることで形のマルチスケールな情報を引き出すことができる。PDのX軸はリングを構成する点の密度の情報、Y軸にはリングの大きさの情報、をそれぞれ表現している。対角線から離れた birth-death pair は発生から消滅までの間隔が長いので他のものより重要性の高い構造である<sup>2</sup>。

2次元では円を考えたが3次元では球を考えるとPDが同様に計算できる。実は3次元の図形にはホモロジー理論的には2種類の穴があり、それに対応してPDも2種類定義される。リング的な通り抜けできる構造を1次元の穴、空洞や風船的な構造を2次元の穴と呼び、PDも対応して1次元と2次元のものがある。

### 3. PHの材料科学への応用について

PHを材料科学に応用する際には、入力データとしてMDシミュレーションなどで得た原子配置のデータや各種観測画像などを用いる。原子配置データを入力に使う場合にはポイントクラウドとして取り扱う。

#### 3.1 アモルファスの特徴付け

まずはアモルファスの原子配置の特徴付けに関する研究 [4] について紹介する。この論文では主にMDシミュレーションで得られたシリカの原子配置データを、アモルファス、液体、結晶の3つの状態に対して解析を行っている。

読者はよくご存知のことだと思うが、アモルファスは結晶のような周期的な構造を取らず、原子配置だけでは液体と区別するのは簡単ではない。構造因子における First Sharp Diffraction Peak などからアモルファスは中距離秩序と呼ばれる秩序構造を持つと考えられている。この構造を理解可能な形で記述したい、というのがこの論文の目標である。

図2はこの論文から引用したそれぞれの原子配置から計算された1次元のPDである。この3つのPDはそれぞれ固有の特徴を持ち、この

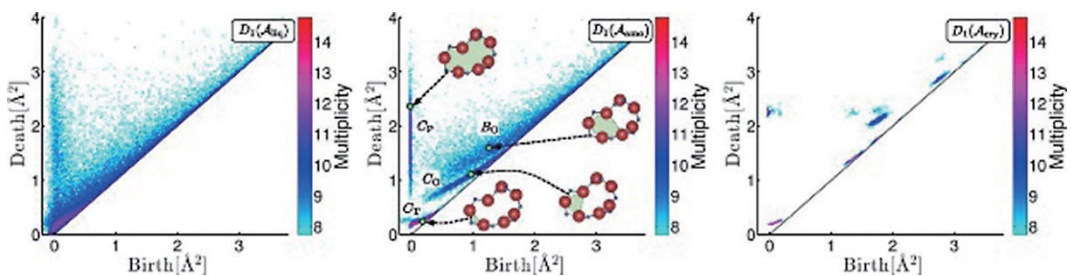


図2 左から液体/アモルファス/結晶のシリカの1次元PD。[4]より引用。

<sup>1</sup> 数学的にこのようなペアリングが常に可能であること、およびペアの取り方にあいまいさがないこと、が証明されている。

<sup>2</sup> 数学的にもこの事実は証明されている。対角線から離れた birth-death pair は入力データが変動しても簡単には消滅しないことが数学的に保証されている。一方対角線に近い birth-death pair は小さな入力データの変動で消滅する可能性がある。

birth-death pair の分布がそれぞれに特徴的な幾何構造を表している。

まず、結晶の PD は孤立した島状に点が分布している。これは原子の相対的な位置関係にほとんど自由度がないことを意味している。一方アモルファスの PD には筋状の特徴的な分布が見てとれる。これは以下に詳しく説明するように原子の相対的な位置関係に強い拘束があるが、自由度もあることを示している。液体の PD は幅広い分布を持っている。これはそういった拘束が他に比べて弱いことを示している。

さらにアモルファスの PD には逆解析<sup>3</sup>で得られたそれぞれの筋状の分布に対応する特徴的なリング構造が図示されている。 $C_T$  は隣接する 3 原子 (O-Si-O) が構成する三角形に対応している<sup>4</sup>。これはいわゆる近距離秩序に対応する構造で、図 2 の 3 つの PD に共通してこの部分にピークを持っている。 $C_p$  は O と Si の間の結合によって構成されるリングで、アモルファスの場合に特にはっきりと見えている。垂直な分布となっているのは結合間距離のばらつきが小さいこととリングの形が多様であることによる。

$C_o$  は間に Si 原子を挟んだ 3 つの O 原子による三角形に対応している。筋状の構造が見えるということは、この 3 原子の相対位置に何らかの強い拘束があることを示している。この論文ではより具体的にその拘束を調べており、実際に低次元の自由度しか持たないことを確認している。このような構造がアモルファスの中距離秩序を記述していると考えられる。間に Si を挟んだ 3 つの O 原子の 3 体相関の上の拘束を直感で発見したり構造因子のような標準的な方法で発見したりするのは難しく、こういった構造をうまく発見できる所に PH の強みがある。

また、この論文では金属ガラスの原子配置についても同様に解析している。詳しくは説明し

ないが、興味深いこととしては金属ガラスの解析では 2 次元の PD のほうが 1 次元より解析に有用だったという事実である。シリカの場合にはネットワーク構造が重要だったので 1 次元の PD のほうが有用で、金属ガラスではパッキングによってできる隙間の形の情報が重要なため 2 次元の PD が有用なのである。

### 3.2 ガラス質ポリマーの変形による微小なひびの形成

次に粗視化 MD シミュレーションによって計算したガラス質ポリマーの変形による破壊の過程を解析した論文 [9] を紹介する。ポリマーをビーズでモデル化 (ビーズを鎖状に連結したものを考える) し、計算機上でガラス状にした後降伏するまで変形した。この変形で微小なひびが大きくなり最終的に降伏に至る。この研究の目標はこのひびの成長過程の解析である。結論としては微小なひびが合併することで大きくなり、降伏に至るという示唆する結果を得ている。

これを MD シミュレーションなどで実証するにはひびの同定が問題になり、その目的で PH が有効活用されている。モデル上ではこの物質は絡まった鎖として表現されているため、ひびというものを効果的かつ効率的に同定することは簡単ではない。ひびを直接モデルで表現する方法も考えられるが、汎用的ではない。PH の役割はこのひびの同定と定量化である。2 次元の PD を用い、各 birth-death pair の発生時間や消滅時間からそのサイズ (小さな発生時間、大きな消滅時間を持つ pair が大きなひびである) の決定が、さらに逆解析を用いることでその同定が、可能となる。

この研究では大きなひびを降伏直後の PD から同定し、そのひびの領域を時間逆向きに追跡

<sup>3</sup> PD 上のそれぞれの birth-death pair に対応するリング / 空隙を元データ上から見つけ出す手法。

<sup>4</sup> この三角形では Si と O の間には結合があるが、2 つの O の間にはない。PH は空間的な近接性によってリング構造を見付けだすため、化学的結合がない所にもリング構造を見付け出すことができる。



することでひびの成長を調べてる。逆解析を活用することでその領域に現われた小さなひびの同定も同様に行われている。これによって微小なひびが合併して大きなひびが形成される様子が観察された。

#### 4. ソフトウェア

PDを使ったデータ解析のためのソフトウェアは世界各地の研究者によって様々なものが開発されている。本稿では我々が開発しているHomCloud<sup>5</sup>を紹介させていただく。HomCloudは3.2で紹介したガラス質ポリマーの論文や、その他様々な研究 [5] [6] [7] [12] [13] ですでに活用されている。

HomCloudは筆者(大林)が中心となり開発しているPHによるデータ解析ソフトウェアである。応用,特に材料科学に対する応用にフォーカスして開発を進めている。ポイントクラウド,白黒画像,グレイスケール画像といったデータを対象として解析することができる。データからパーシステント図を計算することはもちろん,可視化,逆解析,機械学習との連携,といった機能に力を入れている<sup>6</sup>。

特に逆解析 [14] は他のPHソフトウェアにはない, HomCloud特有の機能である。図1 (f)の2点は入力データのXとYの2つのリング構造に対応しているが, PDはその対応関係を捨象することで形の情報を定量的に表現している。一方実際の解析ではこの対応を調べたいことも多い。逆解析はPDの各点に対応する幾何構造を再現する機能である。PDによる解析結果を直感的に可視化して理解するのも有用であるし, 逆解析結果をさらに別の方法で解析することもできる。

HomCloudはPythonによるインターフェー

スとコマンドラインインターフェースの2つのインターフェースを持っている。PythonインターフェースはPythonの科学技術計算エコシステムを活用したデータ解析が可能となるため,特に機械学習を併用したい場合にはこれを使ったほうが良いだろう。HomCloudのインストールや使い方についてはHomCloudのWebサイトにドキュメント,チュートリアルが準備されている。

#### 5. 材料科学への応用に関する課題

本節ではPHを材料化学に適用するにあたっての課題とその対策について議論する。

##### 5.1 原子の種類等の情報の活用

まずは原子配置データにPHを適用する際に原子の種類などの情報を反映させるにはどうすればよいかについて議論する。

一つの方法としては初期半径を原子ごとに割り当てる方法である。2節の解説では半径0から徐々に大きくしたが,実はその初期半径を指定することが可能である。ファンデルワールス半径やイオン半径など,系にとって重要な半径を各原子固有の半径とするとよい。論文 [4] ではこの方法が使われている。

また別の方法として,数種類の原子のうち,特定の種類の原子だけを用いる方法もある。論文 [5] [6] [7] などではこの方法が使われている。

##### 5.2 どの次元のPDを使うべきか

どの次元のPDを使うかは問題に依存する。

3.1節で紹介した研究ではガラスの種類(ネットワークガラス/金属ガラス)によってどの次元のPDが有効かが異なった。3.2節で紹介した研

<sup>5</sup> [https://www.wpi-aimr.tohoku.ac.jp/hiraoka\\_lab/homcloud/](https://www.wpi-aimr.tohoku.ac.jp/hiraoka_lab/homcloud/)

<sup>6</sup> 一方PDの計算のためにはDipha, Phat [16], Ripserといった高速なソフトウェアを組み込んでいる。  
<https://github.com/DIPHA/dipha> <https://bitbucket.org/phat-code/phat/> <https://github.com/Ripser/ripser>などを参照。

究では微小なひびについて調べるため2次元のPDが自然であった。

そういった先験的な知識がないならば、すべての次元のPDを計算して比較するのが妥当であろう。実際論文[4]ではそういった検討がなされている。

### 5.3 PHと機械学習の組み合わせ

3節で紹介した2つの事例は共にPDを目で見えて解析していた。シリカの事例では結晶、アモルファス、液体、の3つのPDは目で見えてははっきりした違いがあった。ガラス質ポリマーの事例でも大きなひびを特定するためにはPDの動画を目で見えて特徴的なbirth-death pairを探していた。どちらの場合もはっきりした特徴がPDに現れるのでこれで問題なかった。

しかし、データの種類によってはPDを目で見てもどの部分に注目すべきかわからない事例も多い。そういった場合にはPDをさらに加工して解析する。機械学習はこういった場合に有効である。機械学習はデータに隠されたパターンを見つけ出すことができ、これとPHを組み合わせることでデータの幾何学的パターンを見つけることができる。機械学習の手法の入力通常ベクトルであるので、何らかの方法でPDをベクトル化することでこの組み合わせを実現する。このベクトルをPDからどのように構築するかには、様々な研究がある。

材料科学データへのこの組み合わせの応用事例としては[10][11][12]などが挙げられる。[12]の研究ではPHの逆解析も組み合わせたデータ解析が行われている。

## 6. おわりに

本稿で紹介したPHはデータの形の情報を定量的に特徴付けることが可能で、MDでシミュレーション得られた原子配置データなどに応用可能である。本稿では説明しなかったが画像にも適用可能で、特にCT画像のような3次元データを直接解析可能な点でも強力である。

図1のように円(球)を各点に貼り付けて半径を大きくしていき、リングや空隙の発生と消滅を調べるのがPHの重要なアイデアである。発生と消滅のペアを作りその半径をプロットしたものがPDである。画像の場合も似たような方法でピクセル領域の増大によるリングや空隙の発生消滅を解析する。

PHをさらに詳しく知りたければサーベイ論文[15]が良い。この論文では数学的な詳細の説明や様々なソフトウェアの性能比較などがなされている。機械学習のためのベクトル化の手法についても簡単にはあるがサーベイされている。

PHが得意とする「形の情報」とはガラス、ポリマー、タンパク質、磁性材料の磁区構造、焼結鉄、と乱れた、非一様な構造であると考えられる。こういった構造は定量化が簡単ではなく、PHという汎用的ツールが役立つと考えられる。

## 謝辞

本稿で紹介した研究の一部はJST CREST JPMJCR15Dの支援を受けたものである。またHomCloudの開発にはJST CREST JPMJCR15D, JST PRESTO JPMJPR1923, JSPS 科研費 P 16K17638 および JP 19H00834, 総合科学技術・イノベーション会議のSIP(戦略的イノベーション創造プログラム)「統合型材料開発システムによるマテリアル革命」の支援を受けている。

## 参考文献

- [1] H. Edelsbrunner, D. Letscher, A. Zomorodian, "Topological Persistence and Simplification," *Discrete and Computational Geometry*, Vol. 28, pp. 511-533, 2002.
- [2] G. Carlsson, "Topology and data," *Bull. Amer. Math. Soc.*, Vol. 46, pp. 255-308, 2009.
- [3] H. Edelsbrunner, J. Harer, *Computational topology: an introduction*, American Mathematical Soc., 2010.

- [4] Y. Hiraoka, T. Nakamura, A. Hirata, E. G. Escobar, K. Matsue, Y. Nishiura, "Hierarchical structures of amorphous solids characterized by persistent homology," *Proceedings of the National Academy of Sciences*, Vol. 113, pp. 7035-7040, 2016.
- [5] M. Murakami, S. Kohara, N. Hirao, T. Musso, A. S. Foster, Y. Idemoto, O. Sakata, Y. Ohishi, N. Kitamura, J. Akola, H. Inoue, A. Hirata, Y. Hiraoka, Y. Onodera, I. Obayashi, J. Kalikka, "Ultra-high-pressure form of SiO<sub>2</sub> glass with dense pyrite-type crystalline homology," *Phys. Rev. B*, Vol. 99, p. 045153, 2019.
- [6] Y. ONODERA, S. KOHARA, K. OHARA, A. MIZUNO, O. SAKATA, S. TAHARA, A. MASUNO, H. INOUE, M. SHIGA, A. HIRATA, K. TSUCHIYA, Y. HIRAOKA, I. OBAYASHI, "Understanding diffraction patterns of glassy, liquid and amorphous materials via persistent homology analyses," *Journal of the Ceramic Society of Japan*, Vol. 127, pp. 853-863, 2019.
- [7] Y. Onodera, Y. Takimoto, H. Hijiya, T. Taniguchi, S. Urata, S. Inaba, S. Fujita, I. Obayashi, Y. Hiraoka, S. Kohara, "Origin of the mixed alkali effect in silicate glass," *NPG Asia Materials*, Vol. 11, p. 75, 2019.
- [8] M. Saadatfar, H. Takeuchi, V. Robins, N. Francois, Y. Hiraoka, "Pore configuration landscape of granular crystallization," *Nature Communications*, Vol. 8, p. 15082, 2017.
- [9] T. Ichinomiya, I. Obayashi, Y. Hiraoka, "Persistent homology analysis of craze formation," *Phys. Rev. E*, Vol. 95, p. 012504, 2017.
- [10] C. Zixuan, M. Lin, W. Kedi, O. Kristopher, X. Keli, W. Guo-Wei, *A topological approach for protein classification*, Vol. 3, 2015.
- [11] Z. Cang, G.-W. Wei, "TopologyNet: Topology based deep convolutional and multi-task neural networks for biomolecular property predictions," *PLOS Computational Biology*, Vol. 13, p. e1005690, 2017.
- [12] M. Kimura, I. Obayashi, Y. Takeichi, R. Mura, Y. Hiraoka, "Non-empirical identification of trigger sites in heterogeneous processes using persistent homology," *Scientific Reports*, Vol. 8, p. 3553, 2018.
- [13] 拓. 山田, 雄. 鈴木, 千. 三俣, 寛. 小野, 哲. 上野, 一. 大林, 裕. 平岡, 真. 小副, "パーシステントホモロジーを用いた迷路状磁区構造におけるトポロジカル欠陥の可視化," *表面と真空*, Vol. 62, pp. 153-160, 2019.
- [14] I. Obayashi, "Volume-Optimal Cycle: Tightest Representative Cycle of a Generator in Persistent Homology," *SIAM Journal on Applied Algebra and Geometry*, Vol. 2, pp. 508-534, 2018.
- [15] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, H. A. Harrington, "A roadmap for the computation of persistent homology," *EPJ Data Science*, Vol. 6, p. 17, 2017.
- [16] U. Bauer, M. Kerber, J. Reininghaus, H. Wagner, "Phat - Persistent Homology Algorithms Toolbox," *Journal of Symbolic Computation*, Vol. 78, pp. 76-90, 2017.