Discovery of new microscopic structures in surface water on graphene using data science

Koichiro Kato1,2, Yuki Maekawa3, Naoki Watanabe1,2, Kenji Sasaoka2, and Takahiro Yamamoto2,3*

1Science Solutions Division, Mizuho Information & Research Institute, Inc., 2-3 Kanda Nishiki-cho, Chiyoda, Tokyo 101-8442, Japan
2Water Frontier Science & Technology Research Center, RIST, Tokyo University of Science, 6-3-1 Niijuku, Katsushika, Tokyo 125-8585, Japan
3Department of Liberal Arts (Physics), Faculty of Engineering, Tokyo University of Science, 6-3-1 Niijuku, Katsushika, Tokyo 125-8585, Japan

E-mail: takahiro@rs.tus.ac.jp

Received October 31, 2019; revised December 17, 2019; accepted December 24, 2019; published online January 14, 2020

We present the first-ever application of a statistical data analysis technique based on a persistent homology method combined with molecular dynamics simulations to determine the microscopic structure of water on graphene. This work assessed the rapid transition in the water structure on going from surface water to free water with increasing distance from the graphene surface, which is one of the most fundamental issues to be resolved in the field of water research. The crossover distance was unexpectedly short at approximately 1 nm, equivalent to three water layers. Within this 1 nm thick layer of surface water, the rotational freedom of tetrahedral water clusters was almost zero due to the presence of the graphene surface. Thus, water on hydrophobic surfaces such as graphene was found to contain a newly-discovered type of water cluster. © 2020 The Japan Society of Applied Physics

1. Introduction

Water is a very familiar substance and is necessary for life as we know it, and so both bulk water and water on surfaces or confined in nanospaces have been studied in detail. Surface water is especially important because it can greatly affect the physical and chemical properties of material surfaces, such as wettability and friction.1) Graphite surfaces have been widely used to research such effects because graphite represents a typical hydrophobic surface.2) Surface water on carbon nanomaterials such as graphene3) and carbon nanotubes (CNTs)4) has also attracted much attention. Because these carbon nanomaterials have non-polar, atomically flat surfaces, they are ideal for the study of the microscopic structure of surface water. Recent theoretical and experimental investigations showed that both graphene and CNT surfaces can be covered with several layers of water molecules even though these surfaces are hydrophobic.5–7) This surface water can be directly observed using high-speed frequency modulation atomic force microscopy (FM-AFM).8,9) In addition, recent theoretical studies based on first-principles calculations have demonstrated the existence of several stable types of polygonal water clusters in bulk water based on hydrogen bonding.10–12) In the case of surface water on graphene, several planar water clusters were also identified in work by Maekawa et al.,13) who established that the probability of the formation of such clusters on surfaces is different from that in bulk water.

However, the differences between surface and bulk water have yet to fully established, because it is difficult to analyze the transition between the two scenarios using conventional simulation methods such as trajectory analysis and snapshot structural observations. Thus, this type of analysis remains one of the simplest but most important issues related to the study of water on material surfaces. In the present work, we employed a data science approach based on a persistent homology (PH) analysis14,15) combined with molecular dynamics (MD) simulations. In contrast to the previous application of PH to a solid (i.e. silica) by Hiraoka et al.,16,17) this study represents the first-ever use of PH to assess a liquid (i.e. water).18,19)

2. Simulation methods

2.1. MD simulations

The structure of water on graphene, on which the present statistical analysis was based, was constructed using classical MD simulations. The MS ForcitePlus18) module in the Materials Studio 8.0 software package was employed for this purpose, with force fields calculated using COMPASS II.19,20) The cutoff for short-range interactions in these simulations was 1.0 nm while long-range interactions were calculated using the Ewald method.21) The optimized structure of water molecules generated by the COMPASS II package included an O–H bond length of 0.11 nm and an O–H–O bond angle of 109.4°, and this molecular structure was fixed for use in the present MD simulations in conjunction with a time step of 1.0 fs. The MD temperature was controlled by a Nose thermostat and calculations were carried out at both room temperature (RT) and low temperature (LT).

Since it is known that the MD temperature of \( T_{\text{MD}} = 250 \, \text{K} \) corresponds to the actual temperature of 300 K based on the evaluation of the diffusion coefficient in the COMPASS II force field, this value was used for RT simulation. On the other hand, a value of \( T_{\text{MD}} = 100 \, \text{K} \) was used for LT simulation. The same initial configuration was used for simulations at both LT and RT. The simulations were performed over a time span of 100 ns at each temperature and the structure obtained from the last 10 000 steps was used for statistical analysis.

2.2. PH analysis

During the statistical analysis, the two-dimensional hydrogen bond network of the first layer on graphene was examined using a polygon extraction algorithm, whereas a three-dimensional structural analysis was carried out during the PH analysis (see appendix A for details regarding the polygon extraction algorithm). PH is a method for the
characterization of data structures including graphics and images in a multi-scale manner, and represents a type of topological data analysis. This method, which was first proposed by Edelsbrunner and Haerete,15) has recently been applied in a wide range of fields, and the details of PH have been described in a number of publications.22–26) In brief, this technique is based on finding features in data by searching for so-called holes, such as rings and cavities. Together, the times at which a hole is born and dies comprise a persistence pair, and a plot of the birth and death times on the horizontal and vertical axes is termed a persistence diagram (PD). A PD plot in which the line is further from the diagonal suggests longer hole survival times. In the field of materials science, PD has been applied by regarding atom positions (specifically oxygen positions in the present work) as a collection of points in space.16,17) Herein, we applied this technique to the structural analysis of water molecules for the first time ever. In order to simplify the computational treatment, a periodic boundary was not used in our MD simulations. CGAL,27) Dionysus,28) and Diode29) were used in the preparation of the PD data. As in our previous work,13) we used a cell with \( x = 2.13 \text{ nm}, y = 2.214 \text{ nm} \) and \( z = 4.15 \text{ nm} \), in which 60, 100 or 160 water molecules were distributed. Since 60, 100 and 160 molecules formed what were essentially single, double and triple layers of water molecules, these are referred to as the 1L, 2L and 3L models, respectively. Similar calculations and analyses were also performed for a model in which 400, 250 and 150 water molecules were dispersed in a cell having dimensions of \((x, y, z) = (3.409, 3.444, 6.134) \text{ nm}\). The results of such calculations were found not to be affected by the MD cell size, and so the analytical results for the 60, 100 and 160 molecular systems are shown in the following sections.

3. Results and discussion

3.1. Density distribution of water molecules perpendicular to graphene surface

The structures of the 1L, 2L and 3L models and the distribution functions for oxygen atoms in the \( z \)-axis direction (perpendicular to the graphene surface) obtained from MD calculations are shown in Fig. 1. From these data, it is evident that water layer structures were formed at both RT and LT, particularly for the 1L and 2L models. In contrast, in the case of the 3L model, the formation of the first layer on the graphene surface is clearly visible but the second layer has a lower peak than the first while the third layer is broader and has no distinct peak structure. At RT, the second and third peaks connect smoothly. It should also be noted that subpeaks (at 0.5 and 0.8 \text{ nm}) can be seen slightly below the main peaks (at 0.6 and 1 \text{ nm}), indicating the appearance of second and third layers at LT. There has, to date, been little discussion of this complex peak structure, but its origin will be explored later (Sect. 3.5) in this paper.

3.2. Inner-layer hydrogen bonding network on graphene surface

Here, we discuss the analysis of the in-plane hydrogen bonding network structure of the first layer on the graphene surface. The network structure in the planes of 1L and 2L models have been discussed in detail by Maekawa et al.13) and the results of the present work are in good agreement with these prior data. Figure 2 shows the number of polygons (from triangles to octagons) in the hydrogen bonding networks in the first layers on the graphene for a total of 10 000 steps sampled from MD production runs. Using the 1L and 2L models, tetragons appeared most frequently at both RT and LT and there is a characteristic increased number of appearance of tetragons at LT in the 1L model. In the 1L model, the formation of tetragons is the most efficient and stable means of covering the surface because all OH groups can interact via intra-layer hydrogen bonds. Other polygons, such as pentagons and hexagons, are less efficient. In the case that the water molecules in the layer adopt these other shapes, some OH groups must orient themselves vertically relative to the graphene surface, and so cannot connect with other water molecules by hydrogen bonding. Interestingly, the results obtained from the 3L model show the same trends as in the 1L and 2L models at RT, but the number of pentagons is highest at LT, which differs from the other scenarios.

3.3. Analyses of PH

The changes in the in-plane network structure of the first water layer on the graphene can be clearly seen from the results of the PD analysis for each model in Fig. 3. In these plots, the units for both axes are angstroms so that the data are easier to compare to the lengths of hydrogen bonds. At a distance of approximately 1.4 \text{ Å} above the birth axis, the dark region is seen to move upwards (that is, in the direction in which the death time becomes longer) as the model transitions from 1L to 3L. If we consider a regular polygon for which the length of one side equals the distance between hydrogen bonds, the birth value is approximately 1.4 \text{ Å} while the death values are approximately 2.0, 2.4 and 2.8 \text{ Å} for a regular square, pentagon and hexagon, respectively, and the polygon moves upward along the same birth line (appendix B). Thus, it is evident that the change in the thick portion of the PD plot at which the birth value is 1.4 \text{ Å} reflects the change in the in-plane cluster structure of the water in the first layer on the graphene. This result is also qualitatively consistent with the change in the proportion of polygons in the cluster structure formed by the water in the first layer on the graphene surface, as seen in Fig. 2.

3.4. Novel three-dimensional water structure discovered by PH

Additional features can be obtained from the PD analysis, including a structure that appears in the vicinity of birth and death values of 2.2 \text{ Å} and is prominent at LT. In the 1L model, there is no persistence pair near this point. Rather, this feature begins to appear in the 2L model and is not only more prominent in the 3L model but has a definite pinnate structure. When the origin of these persistence pairs was investigated by considering their relationship with the hydrogen bond distance, it was found that they were caused by a distorted triangle with two oxygen atoms at its apex and a water molecule at the center of gravity of a tetrahedron. The data also showed that the pinnate part represents a persistence pair derived from a polygon formed by the oxygen atoms located at the center of gravity of the tetrahedron and at its second nearest neighbors [Fig. 4(a)]. Considering the efficient formation of the hydrogen bonding network, in the case of the 2L model, energetic stability is increased by closing the hydrogen bonding network through forming polygonal clusters in and between planes. That is, in the 2L model, tetrahedra are formed in a limited manner under the condition.
Fig. 1. (Color online) Molecular structures and oxygen distributions along z-axis of (a) 1L model (60 molecules), (b) 2L model (100 molecules) and (c) 3L model (160 molecules). The backgrounds of each figure are the structures from the final MD step at LT. Red, white and gray represent oxygen, hydrogen and carbon, respectively. Graphene, shown in gray, overlaps the horizontal axis.

Fig. 2. (Color online) Distributions of cluster structures in 10 000 snapshot structures at 100 K for (a) 1L, (b) 2L and (c) 3L models, and at room temperature for (d) 1L, (e) 2L and (f) 3L models.
that the direction of water molecules is restricted at the graphene-water and water-vacuum interfaces. Conversely, in the 3L model, since the third layer exists, more tetrahedra can be formed by increasing the degrees of freedom related to the direction of movement of water molecules in the intermediate layer. Therefore, the difference in the total number of tetrahedral structures formed suggests that no pinnate peak should appear in the 2L model at the same LT. At RT, this peak is buried in a broad distribution, likely because of the frequent rearrangement of the hydrogen bonding network due to thermal fluctuations and the presence of water molecules in various situations, including tetrahedral structures. The PD differences between the three models at RT appear small. However, an analysis of the tetrahedral order\(^{30,31}\) shows an increase in tetrahedral structures in the 3L model compared to the other two models. Therefore, the behavior at LT evidently demonstrates that the formation of tetrahedral structures is more pronounced at RT. See appendix C for details.

3.5. Crossover between surface and free water
Considering the direction of formation of the tetrahedra, it is conceivable that any one of the triangular faces, sides and apexes may be situated in the first layer of the graphene surface. However, since the apexes of a tetrahedron are not hydrogen bonded and the distance between the apexes is about twice as long as the hydrogen bond distance, if the base of a tetrahedron is situated in the first layer, the gain in energy associated with hydrogen bonding will be lost. That is, from the viewpoint of energy theory, it is expected that tetrahedra whose apexes are located in the first layer on the graphene surface will be predominant. Thus, the oxygen atom at the center of gravity takes a position closer to the second layer than to the middle of the first layer and the second layer. The oxygen at the center of gravity of this anisotropic tetrahedral structure is the origin of the subpeak just below the second layer in the distribution of oxygen atoms along the \(z\)-axis shown in Fig. 1. The anisotropy of the tetrahedra resulting

---

Fig. 3. (Color online) Persistence diagrams at 100 K for (a) 1L, (b) 2L and (c) 3L models, and at room temperature for (d) 1L, (e) 2L and (f) 3L models.

Fig. 4. (Color online) (a) The anisotropic tetrahedral structure formed by surface-affected water. Red, blue and green denote the tetrahedral apex, the water in the second nearest neighbor of the tetrahedral centroid, respectively. (b) The relation among subpeak values of both 0.5 and 0.8 nm and anisotropic tetrahedral structures. The centroid oxygen of the tetrahedral structures shown in blue and deep green is considered to be the origin of the first and second subpeaks, respectively.
from the effect of the surface sharpens this subpeak. In addition, the subpeak seen under the third layer is also considered to be derived from the oxygen at the center of gravity of the tetrahedron. The broadening of the peak is thought to be due to the weakening of the anisotropy caused by the reduced effect of the surface and the approach to free water. Figure 4(b) shows some tetrahedral structures in the vicinity of graphene in a snapshot structure used in the analysis. In the actual snapshot structure, it can be seen that the oxygen of the tetrahedral center of gravity is located at the height corresponding to each subpeak. Although it has been reported that ice-like tetrahedral structures are formed in free water,32) the number of water molecules required for the development of such three-dimensional structures is unknown. The present results indicate that the presence of tetrahedral structures is already significant when three layers of water molecules are stacked on the graphene surface, and that a two-dimensional hydrogen bonding network is transformed into a three-dimensional network. The data also suggest that tetrahedral structures between the first and second layers will be highly anisotropic, while tetrahedral structures between the second and third layers show lower anisotropy and approach free water. That is, the crossover between surface water and free water occurs within only three layers of water.

4. Conclusions

This work applied PH, which has become an important data analysis method in recent years, to the analysis of water structures for the first time ever. This analysis confirmed a crossover between surface water and free water, which has thus far been considered difficult to establish. This is a good example of the application of modern data analysis techniques, which have provided new and important insights. The crossover demonstrated herein is based on the anisotropic formation of tetrahedral structures that could possibly be directly observed using FM-AFM or by spectroscopic means such as sum frequency and infrared absorption spectroscopy. It is expected that these structures will be confirmed experimentally in the future. Graphene has a charge neutral, atomically flat surface and so it is not difficult to imagine that water on a polar surface or a more highly structured surface would have a more complicated structure. Even so, this study provides a starting point for discussions of more realistic surface effects, and we expect it will lead to the control of surface properties through a deeper understanding of wettabiliy and friction.

Acknowledgments

This work was supported in part by Grants-in-Aid for Exploratory Research (Nos. 17H02756 and 16H02079) from the Japan Society for the Promotion of Science (JSPS). This work was also supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) under the Program for the Strategic Research Foundation at Private Universities, 2015–2019.

Appendix A: Polygon extraction algorithm

The top of the network is defined as the position of the oxygen atoms of the water molecules, while lines connecting the vertices of hydrogen bonded water molecules form the edges of the network. Polygons contained in the overall network of water molecules formed by these vertices and edges were extracted using an algorithm. This algorithm searched for polygons ranging from triangles to octagons, starting at the point where they were most likely to be located. By recursively calling subroutines, all polygons were searched.

Initially, a vertex having two connecting sides at the network end was searched for [Fig. A·1(a)]. From this vertex, the side connected to the vertex was traced in one direction and the n-gon was extracted from the number of sides, \(N\), that must be traced to return to the original vertex. As an example, \(N = 4\) is shown in Fig. A·1(b). After the polygon was extracted, the vertices of the network ends and the two sides connected thereto were deleted [Fig. A·1(c)]. The n-gon was extracted from the number of sides, \(N\), that were traced from one vertex of the remaining side to the side connected to the vertex in one direction until returning to the original vertex. \(N = 3\) is shown in Fig. A·1(d). After the n-gons were extracted, the first side was deleted [Fig. A·1(e)] and the n-gons were extracted from the remaining sides in the same manner [Figs. A·1(f)–A·1(h)]. If an n-gon was not found in the above procedure, the vertex at the other network end was located and the n-gon was searched for by repeating the above procedure [Fig. A·1(a)].
Appendix B: PD for the planar polygon formed by hydrogen bonds

If we consider persistent pairs in a regular polygon with sides of length \( L \), as shown in Fig. B·1(a), in the case of an equilateral triangle birth equals \( L/2 \) and death is \( L/\sqrt{3} \). In contrast, in Fig. B·1(b), in the case of a square, birth is \( L/2 \) and death is \( L/\sqrt{2} \). Similarly, for regular equilateral polygons, birth is \( L/2 \) and death is \( L/2 \sin(\pi/N) \). In the PD diagram, birth is plotted on the \( L/2 \) line, such that a greater number of corners increases the death value.

Appendix C: Analysis of tetrahedral order

The tetrahedral order is one of the most commonly used order parameters for a water structure, and is determined based on the following equation

\[
q = 1 - \frac{3}{8} \sum_{j=1}^{3} \sum_{k=j+1}^{4} \left( \cos \psi_{jk} + \frac{1}{3} \right)^2.
\]

A tetrahedral structure will have a value of 1, and Fig. C·1 shows the tetrahedral order values for the 2L and 3L models at RT. These data can be divided into two regions, consisting of a region having a dense population of water molecules (0.0–0.4 nm) and a region in which water molecules are sparse (0.4–0.55 nm). The results suggest that tetrahedral structures were formed in both model regions. However, in the sparse region, structures associated with \( q \) values smaller than 0.7 were decreased while those with \( q \) values larger than 0.7 were increased. This change is most pronounced in the 3L model, which demonstrates the presence of tetrahedral structures in the sparse regions even at RT.

Fig. B·1. (Color online) Persistence pairs in (a) regular triangle and (b) square.

Fig. C·1. (Color online) Tetrahedral order for 2L and 3L models at room temperature. Arrows in the figure indicate increases in the tetrahedral structure in regions in which water molecules are sparse (0.4–0.55 nm) as the structure changes from a 2-layer to a 3-layer morphology.

ORCID iDs
Koichiro Kato © https://orcid.org/0000-0003-4392-8741

1) P. G. de Gennes, Rev. Mod. Phys. 57, 827 (1985).
2) F. M. Fowkes and W. D. Harkins, J. Am. Chem. Soc. 62, 3377 (1940).
3) K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, Science 306, 666 (2004).
4) S. Iijima, Nature 354, 600 (1990).
5) T. A. Ho and A. Striolo, J. Chem. Phys. 138, 054117 (2013).
6) A. Akaishi, T. Yonemaru, and J. Nakamura, ACS Omega 2, 2184 (2017).
7) Y. Homma, S. Chiashi, T. Yamamoto, K. Kono, D. Matsumoto, J. Shitaba, and S. Sato, Phys. Rev. Lett. 110, 157402 (2013).
8) T. Fukuma, Y. Ueda, S. Yoshioka, and H. Asakawa, Phys. Rev. Lett. 104, 016101 (2010).
9) K. Kobayashi, N. Oyabu, K. Kimura, S. Ido, K. Suzuki, T. Imai, K. Tagami, M. Tsukada, and H. Yamada, J. Chem. Phys. 138, 184704 (2013).
10) R. M. Shields, B. Temelso, K. Archer, T. E. Morrell, and G. C. Shields, J. Phys. Chem. A 114, 11725 (2010).
11) M. E. Dunn, E. K. Pokon, and G. C. Shields, Int. J. Quantum Chem. 100, 1065 (2004).
12) M. B. Day, K. N. Kirschner, and G. C. Shields, J. Phys. Chem. A 109, 6773 (2005).
13) Y. Maekawa, K. Sasaoka, and T. Yamamoto, Jpn. J. Appl. Phys. 57, 035102 (2018).
14) H. Edelsbrunner and J. Harer, Contemp. Math. 453, 257 (2008).
15) A. Zomorodian and G. Carlsson, Discrete Comput. Geom. 33, 249 (2005).
16) T. Nakamura, Y. Hiraoka, A. Hirata, E. G. Escolar, and Y. Nishitaa, Nanotechnology 26, 304001 (2015).
17) Y. Hiraoka, T. Nakamura, A. Hirata, E. G. Escolar, K. Matsue, and Y. Nishitaa, Proc. Natl. Acad. Sci. 113, 7035 (2016).
18) Biovia, BIOVIA Materials Studio 8.0 Dassault Systèmes (Biovia, Cambridge, 2015).
19) H. Sun, J. Phys. Chem. B 102, 7338 (1998).
20) H. Sun and D. Rigby, Spectrochim. Acta A 53, 1301 (1997).
21) P. P. Ewald, Ann. Phys. 369, 253 (1921), [in German].
22) H. Edelsbrunner and J. Harer, Computational Topology: An Introduction (American Mathematical Society, Providence, RI, 2010).
23) G. Carlsson, Bull. Am. Math. Soc. 46, 255 (2009).
24) P. Babušek and P. T. Kim, Homotopy Appl. 9, 337 (2007).
25) R. J. Adler, O. Bobrowski, M. S. Borman, E. Subag, and S. Weinberger, Inst. Math. Stat. Collect. 6, 124 (2010).
26) F. Chazal, V. de Silva, and S. Oudot, Geom. Dedicata 173, 193 (2014).
27) CGAL, Computational Geometry Algorithms Library [https://cgal.org].
28) D. Morozov, Dionysus 2: a computational topology package focused on persistent homology [http://mrzv.org/software/dionysus2].
29) D. Morozov, DioDe: alpha shape generator using CGAL [https://github.com/mrzv/diode].
30) P. L. Chau and A. J. A. Hardwick, Mol. Phys. 93, 511 (1998).
31) J. R. Errington and P. G. Debenetti, Nature 409, 318 (2001).
32) J. D. Smith, C. D. Cappa, K. R. Wilson, B. M. Messer, R. C. Cohen, and R. J. Saykally, Science 306, 851 (2004).