Research Article

Yisi Liu, Ke Xu, Yihua Xu, Jinjie Liu, Jianyang Wu*, and Zhisen Zhang*

HTR: An ultra-high speed algorithm for cage recognition of clathrate hydrates

Abstract: Clathrate hydrates find diverse significant applications including but not limited to future energy resources, gas storage and transport, gas separation, water desalination, and refrigeration. Studies on the nucleation, growth, dissociation, and micro/nanoscale properties of clathrate hydrates that are of utmost importance for those applications are challenging by experiments but can be accessible by molecular simulations. By this method, however, identification of cage structures to extract useful insights is highly required. Herein, we introduce a hierarchical topology ring (HTR) algorithm to recognize cage structures with high efficiency and high accuracy. The HTR algorithm can identify all types of complete cages and is particularly optimized for hydrate identification in large-scale systems composed of millions of water molecules. Moreover, topological isomers of cages and $n \times$ guest@cage can be uniquely identified. Besides, we validate the use of HTR for the identification of cages of clathrate hydrates upon mechanical loads to failure.

Keywords: clathrate hydrate, ultra-high speed, cage recognition algorithm, molecular dynamic simulation

* **Corresponding author: Jianyang Wu,** Department of Physics, Research Institute for Biomimetics and Soft Matter, Jiujiang Research Institute and Fujian Provincial Key Laboratory for Soft Functional Materials Research, Xiamen University, Xiamen 361005, China; Department of Structural Engineering, NTNU Nanomechanical Lab, Norwegian University of Science and Technology (NTNU), Trondheim 7491, Norway, e-mail: jianyang@xmu.edu.cn
* **Corresponding author: Zhisen Zhang,** Department of Physics, Research Institute for Biomimetics and Soft Matter, Jiujiang Research Institute and Fujian Provincial Key Laboratory for Soft Functional Materials Research, Xiamen University, Xiamen 361005, China, e-mail: zhangzs@xmu.edu.cn
Yisi Liu, Ke Xu, Yihua Xu, Jinjie Liu: Department of Physics, Research Institute for Biomimetics and Soft Matter, Jiujiang Research Institute and Fujian Provincial Key Laboratory for Soft Functional Materials Research, Xiamen University, Xiamen 361005, China

1 Introduction

Clathrate hydrates are solid compounds composed of water and small molecules termed as guest molecules (such as methane, carbon dioxide, and so on), in which guest molecules are encased in solid cage-like water structures. According to conservative estimates, the energy, mainly methane, contained in the global accumulation of natural gas hydrate is twofold as that of currently recoverable fossil fuels [1] on the earth. Hydrates mainly occur in the deep-sea sediments, where high-pressure and low-temperature conditions facilitate their formation [2–4]. With the continuous maturity of technology, the mining of gas inclusion compounds has gradually become a hot topic, and these mineral resources are expected to become an important part of the energy supply in the world. Inclusion complex hydrates are also involved in various scientific fields such as the storage of greenhouse gases [5–8], the storage of energy gases [9], and the study of antifreeze protein principles [10–12]. Due to the important role of clathrate hydrates in a series of important scientific and technical issues in the field of energy, environment, and geological disaster prevention, the research of gas hydrates has become a lasting and important topic [1]. In addition, the formation of hydrates is likely to cause a series of problems such as pipeline blockage during natural gas transportation [13]; therefore, further research on the thermal and mechanical aspects of clathrate hydrates is of essential importance.

Currently, the researches of clathrate hydrate are mainly carried out through experiments and molecular dynamics (MD) simulations. In the research on the nucleation and mechanical properties of clathrate hydrates, it is not easily accessible to obtain molecular information on the nanometer length scale and microsecond time scale through experimental methods. With the aid of computational modeling, some phenomena that cannot be observed in experiments may be discovered [14]. Therefore, it is necessary to conduct clathrate hydrate research by MD simulations [15]. Debenedetti and Sarupria [16] realized the spontaneous nucleation of hydrates by MD, and it is inspired by the use of MD methods to study the nucleation and growth of

hydrates, making the MD simulation popular in the study of clathrate hydrate [17–20]. In addition, although great efforts have been made in hydrate nucleation, the path is not clear, and nucleation seems to occur through multiple competitive paths to varying degrees [21]. Consequently, it is essential to recognize the cage structure form in the MD trajectories to extract useful insights. As a result, there have been several cage recognition algorithms proposed. For instance, Jacobson *et al.* [8] connected particles by identifying rings by treating water as vertices in an undirected graph. After determining the relationship between rings, the algorithm searches a five-element ring, which connects five other pentagonal rings at each edge. This structure is a half dodecahedral cage, which is called a "cup." A complete dodecahedral cage was obtained by merging these cups. Guo *et al.* [22,23] designed an algorithm called face-saturated incomplete cage analysis (FSICA) and identified 1,258 complete cages and 7,015 face-saturated incomplete cages (FSICs). The recognition accuracy of FSICA is quite high, and FSICA can recognize the surface-unsaturated cage, which cannot be achieved by other algorithms. Nguyen and Molinero [24] designed the CHILL + algorithm that uses staggered and overlapping water bond numbers to identify water molecules in ice and clathrate hydrates. Mahmoudinobar and Dias [25] presented the GRADE algorithm to identify the hydrates of the 5^{12} and $5^{12}6^{2}$ cage structures through the combination of the upper and lower cup-shaped half cages or the combination of four cup-shaped cages to identify the $5^{12}6^{4}$ structure. Recently, Hao *et al.* [26] proposed the iterative cup overlap (ICO) algorithm, which can identify hydrate cages by combining multiple semicage structures. However, it still remains deficiencies for these algorithms yet, such as fewer types of recognizable cages, low recognition efficiency, and inability to accurately recognize deformed cages.

To address these issues, we proposed the HTR algorithm. The HTR algorithm is written in Java that can identify any type of water cage. Its advantage is that it has excellent cross-platform performance while maintaining high-speed code execution. The compiled ".class" file package can be run directly on all platforms where the Java virtual machine is installed without repeated compilation. By comparing with other algorithms, HTR shows higher efficiency advantages on the basis of accurate identification, particularly for large systems composed of hundreds of thousands or even tens of millions of water molecules. In addition, HTR only requires the least hardware cost in the identification process because it does not need to read the atomic information in the entire system into the memory at one time when identifying the cage structure for a larger hydrate system.

2 Algorithm

To improve the maintainability of the program and reduce the degree of coupling between functions, the HTR algorithm divides the hydrate identification into three steps, including (i) the segmentation of the clathrate hydrate system, (ii) the identification of the rings through the water molecule, and (iii) the identification of the cages through the identified ring (refer to Figure 1 for more details).

2.1 Identification of the rings

The ring recognition algorithm by the GRADE [25] is on the basis of the breadth-first search (BFS) method, which should be the most accurate method for ring recognition at present. However, the problem brought by this method is that it will identify many severely deformed rings. GRADE filters out severely deformed rings because they seldom form a cage, and only identifies the relatively stable cages of 5^{12}, $5^{12}6^{2}$, and $5^{12}6^{4}$. At the same time, a phenomenon was discovered in our study of the structure of hydrate cages, that is, severely deformed rings usually appear in some unconventional cages. Therefore, filtering out this part of the rings will inevitably reduce the accuracy of recognition. Moreover, these unconventional cages are of great importance in the studies of clathrate hydrate nucleation and thermodynamic properties. After referring to the various ring deformation conditions proposed by Matsumoto *et al.* [27], HTR retains all deformed rings when calculating cages to ensure the accuracy of identifying unconventional cages. According to the current general judgment conditions for hydrogen bond formation, HTR uses a cutoff of 3.5 Å and ∠HOO at 45° as the hydrogen bond formation conditions. The reason for setting the hydrogen bond judgment condition is that Laage and Hynes [28] proposed that the cutoff should not be greater than 4.0 Å and the bond angle should be less than 50°, and GRADE [25] also adopted this set of parameters. Using 45° as the bond angle can minimize the inaccurate identification caused by the serious deformation of the cage in the process of hydrate nucleation. The cage identification is based on the MD simulations with coarse-grained model of water, where the bond length is determined only by the distance [29,30].

In this study, the hydrate cage studied mainly contains two characteristics, one of which is that the cage only contains four-, five-, and six-membered rings, and the other is that the cage is surface saturated, as shown in Figure 1. Regardless of the type of clathrate hydrate, it usually contains only four-, five-, and six-membered rings.

2.2 Identification of the cages

The process of HTR recognition of cages can be mainly divided into two parts: (i) the recognition of the cup-shaped half-cage structure and (ii) the topology of rings. Jacobson *et al.* [8] defined a cup as a half-cage composed of five-membered and six-membered rings sharing some hydrogen bonds. Since the HTR algorithm needs to identify all types of cages, the four-membered ring is also included in the definition of the cup.

2.2.1 Recognition of the cup

According to the order in which the rings are stored in the memory, each ring must be used as a bottom ring (Figure 2a). After selecting a certain ring as the bottom ring, the ring
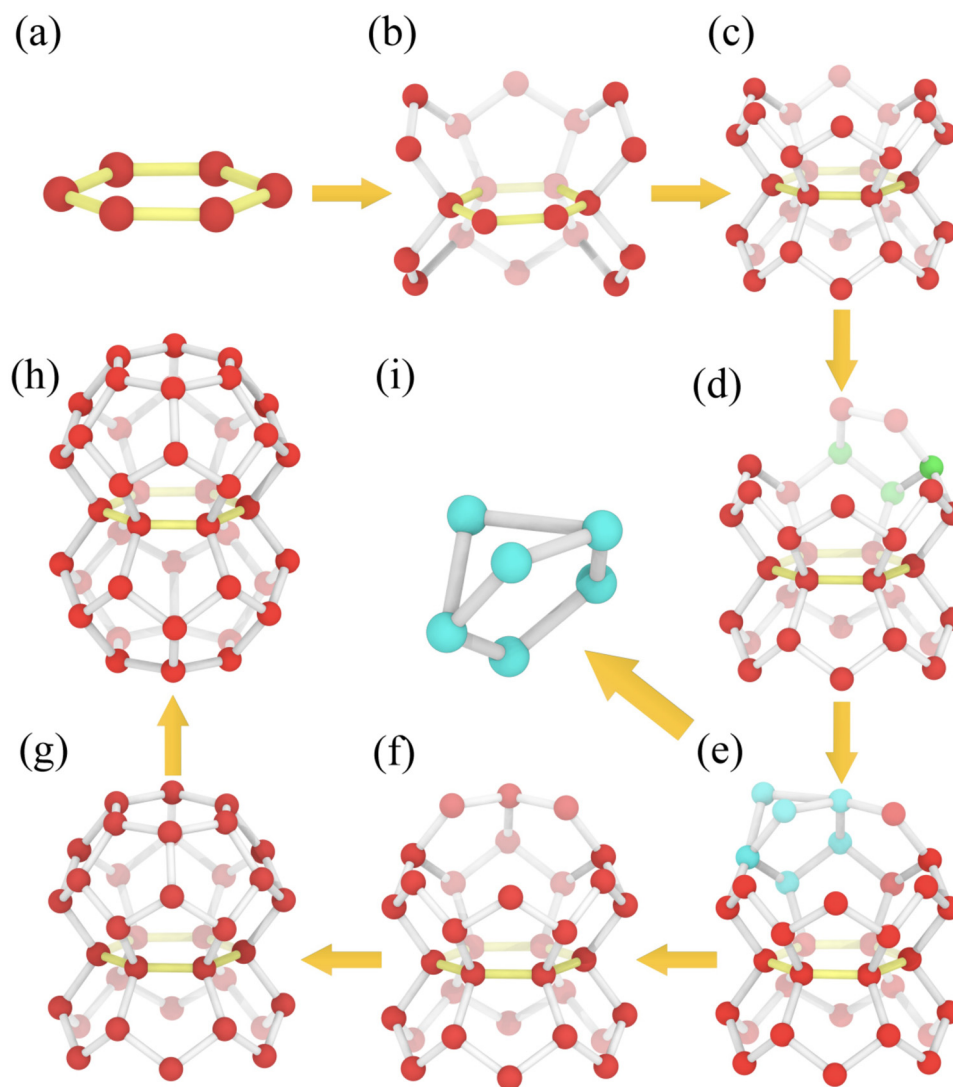
library that shares a hydrogen bond with the bottom ring is generated. All the rings can identify the upper and lower cups that share the same ring (Figure 2c).

2.2.2 Topology of the ring

From the geometrical perspective, two straight lines that intersect at one point can uniquely determine a plane, and similarly, it can be uniquely positioned to a ring through two hydrogen bonds formed by three adjacent water molecules. When this happens, the HTR algorithm will first use one of the rings to perform topology calculations (Figure 2f). If the current ring cannot be topped with a complete cage, it will switch to another ring to perform the topology again. After repeating the aforementioned ring topology process, when the cage is finally closed
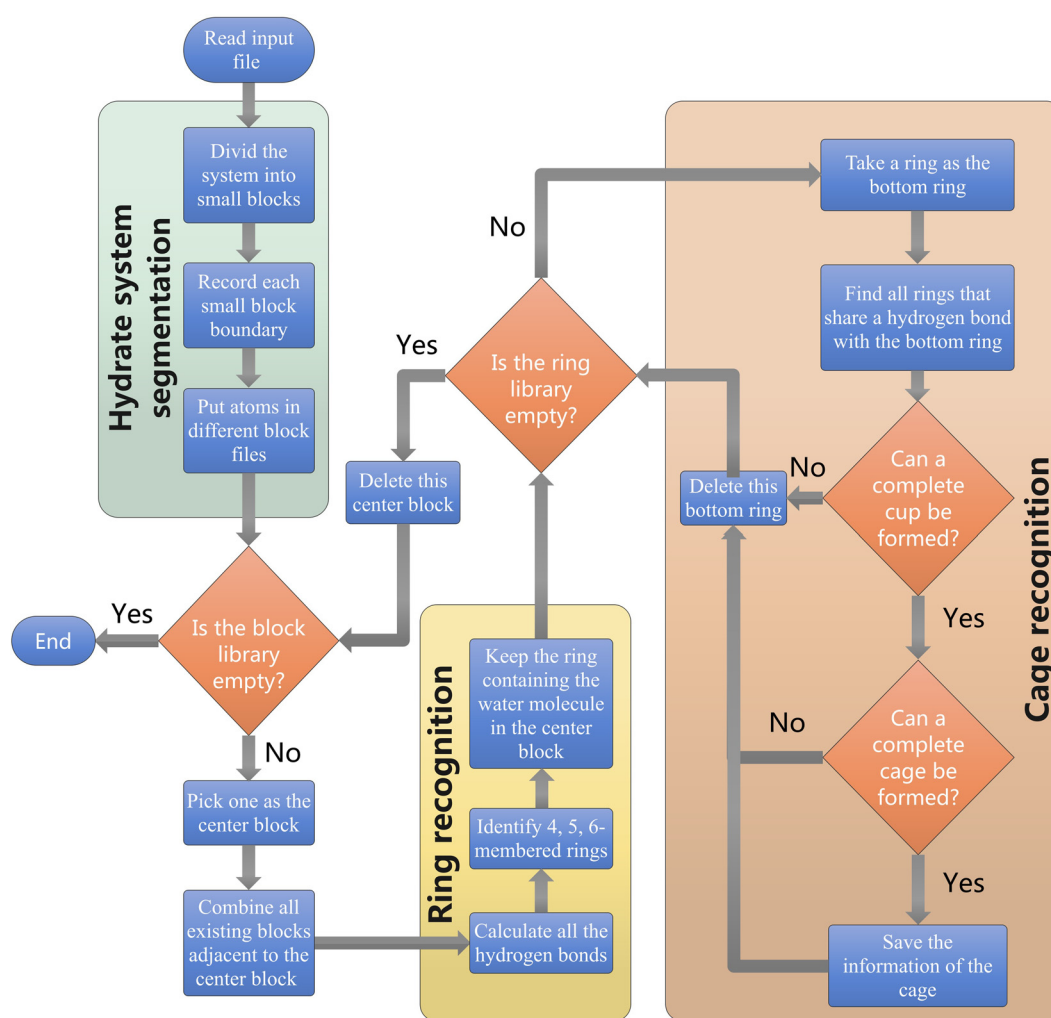
Figure 1: General flow chart of HTR. The algorithm is mainly divided into three processes: (i) hydrate system segmentation, (ii) ring recognition, and (iii) cage recognition.

(a)   (b)   (c)

(h)   (i)   (d)

(g)   (f)   (e)

Figure 2: Scheme of HTR algorithm. Choosing a ring (a) as the bottom ring, we can find all the rings that share a hydrogen bond with this ring. Through the process (b), the algorithm can find all the rings that share a hydrogen bond with this ring and obtain the two half-cage structures that share the same ring in (c). Select one of the half cages for the topology of rings, and the three green water molecules at the edge of the upper half of the cage in (d) can generally be located in the unique ring from the identified ring library. Repeat the ring topology process, and if the cage cannot be closed, go back to step (e) to use another ring for topology. If the cage is closed, as shown in (g), repeat the aforementioned topological process for the lower half of the cage and finally get a double-cage structure sharing a ring in (h). If the two rings in (i) share more than three water molecules, the HTR algorithm will first select one of the rings to obtain the structure (f) and then proceed to the next topology.

(Figure 2g), the HTR then determined that the identification of the cage is completed, otherwise it is considered that it is not a complete cage. Then, the aforementioned topological process for the lower cup is repeated, and one can get the structure of two cages sharing a ring, as shown in Figure 2h. At this point, the bottom ring has completed its mission and will no longer appear in other cages (because a ring can only belong to two cages at most).

The hydrate cage is a polyhedral structure, and the contact surface of two convex polyhedral cages is a ring.

Since a cage contains multiple rings, the selection of different rings on the same cage as the bottom ring for cage topology will surely find the same cage. To ensure that the same cage will not be identified repeatedly, a very clever algorithm to prevent misidentification has been added to the HTR. In Figure 3, as the left and middle cages are identified by the left yellow ring, the yellow ring is shielded from the ring library. When the middle cage is identified by the red ring on the right, the yellow ring on the left can no longer be obtained, thereby the
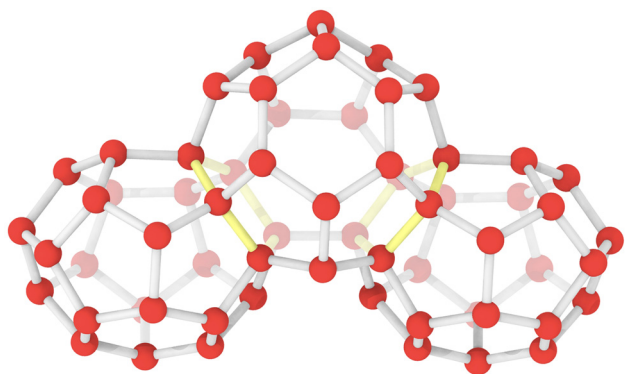
Figure 3: Mechanisms of HTR algorithm to prevent duplicate identification. The yellow ring is at the intersection of two adjacent cages. When the middle cage is recognized by the yellow ring on the left, it can no longer be recognized by the yellow ring on the right.

cage is deemed incomplete and the problem of the repeated identification is avoided.

2.3 Acceleration of HTR algorithm

After referring to the previous hydrate cage identification algorithms, a common feature can be summarized, that is, hydrates in large systems cannot be effectively identified, especially for the MD trajectories with a large number of frames and water molecules. The efficiency issue of large-scale system recognition is that the recognition of each cage needs to traverse the spatial coordinates of all water molecules to determine which are adjacent water molecules. This means that when a larger system is presented, more

particles are needed to be traversed to identify a cage. Therefore, a power or exponential relationship between the number of particles and time is shown, which means that it will be very hard to recognize the cage structures for large systems.

To solve this issue, HTR divides the system into multiple blocks, identifies the cage structures separately, and hence greatly improves the recognition efficiency. As illustrated in Figure 4, a clathrate hydrate system is divided into $4 \times 4 \times 4$ identical subblocks. By the way, the number of subblocks to be split is determined according to the size of the system, and the larger the hydrate system, the more times it will be split. All atomic information of each small piece after dividing is saved in the local hard disk and can be identified individually, therefore, the pressure one the running memory is greatly reduced when HTR running on a large system. However, it should be noticed that the segmentation will destroy the completeness of the cages at the cut surface, which will cause the cages at the incision to not be accurately identified. For example, as illustrated by the green block in Figure 4, to accurately identify the cage structure, the HTR algorithm connects all the neighboring 26 blocks of the green block and combines them into a whole and only identifies the cage structure in the green block (Figure 4). By this method, all the cages at the incision of the green block can be correctly identified. Once the cages in one block are recognized, this block is discarded and will not be called during the recognition process for other blocks. When HTR running on one neighboring block, all the cages at the cut surface adjacent to the previous block will be identified as broken cages, thereby ensuring that the
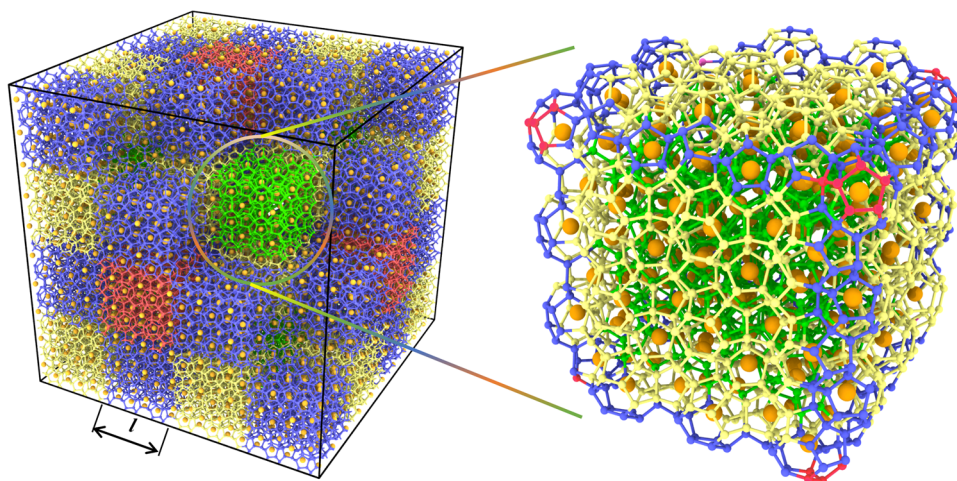
Figure 4: Cross-sectional view of segmentation of the clathrate hydrate system by HTR algorithm. Adjacent blocks are distinguished by different colors. The water molecules of different colors in the partially enlarged view come from the blocks adjacent to the central green block.
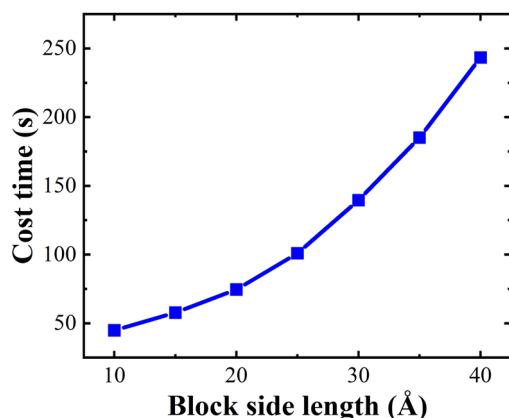
704 — Yisi Liu *et al.*

DE GRUYTER

cages at the incision are only one time recognized. As a result, the recognition time of each subblock is identical, and in principle, there is a linear relationship between the number of molecules and the recognition time for an MD simulation system. It should be reminded that to avoid missing the large-volume cages, the block size should be larger than the maximum diameter of the cages in the simulation systems.

3 Results and discussion

3.1 High efficiency of HTR on cage recognition of clathrate hydrate

According to the algorithm section, when HTR running on the green subblock in Figure 4, the time to identify all the rings in the block is t_r, and the time to identify all the cages in the small block through the rings is t_c. Then, the recognition time of the rings of the 27 subblocks centered on the green block is $27t_r$, and the recognition time of the cages is t_c (only cages in the green block are recognized). The total cost time on the green subblock is $27t_r + t_c$. Assuming that the hydrate system is divided into n blocks, the total identification time will be $n(27t_r + t_c)$. The recognition time of the entire system can be accurately predicted only by the number of subblocks and the recognition time of a subblock. A perfect $5 \times 8 \times 10$ supercell of sI-type clathrate hydrate system was selected for testing, which is shown in Figure 5.

Figure 5 shows that there is a strong relationship between the block size and the cost time on cage recogni-

tion for the clathrate hydrate system by HTR. After try-error tests, a subblock size of 12 Å was chosen. The numbers of 5^{12} and $5^{12}6^2$ cages in a $5 \times 8 \times 10$ supercell of sI-type clathrate hydrate system are recognized to be 8,000 and 24,000, respectively, verifying the accuracy of the HTR algorithm.

To verify the recognition efficiency of the HTR algorithm, the sI-type clathrate hydrate systems with different dimensions were identified. The cost time of each system as a function of water molecules was plotted in Figure 6, which is approximately linear with a slope of 1.38×10^{-4}. The HTR algorithm exhibits an extremely high efficiency and linear advantages for the cage identification of clathrate hydrate large systems. Here, the computer configuration parameters are as follows: operating system: Ubuntu 18.04.5 LTS; hardware architecture: ×86_64; Kernel: Linux 5.4.0-81-generic; CPU: Intel(R) Xeon(R) Gold 6128 CPU@3.40 GHz 6 cores; memory: 32 G.

In the research area of clathrate hydrates, nucleation and growth are important research topics. HTR algorithm was first applied to the nucleation and growth process of clathrate hydrates. Figure 7 shows the variation in the number of cages recognized by HTR with the MD simulation time in a homogeneous nucleation process of methane hydrate systems at a temperature of 250 K and confining pressure of 500 bar, in which the simulation system is composed of 1,242 water molecules and 216 methane molecules. As is indicated, the cage recognition speed by HTR is very efficient, that is, the recognition time of each frame of MD trajectory is in the millisecond level. To be rigorous, the simulation system was expanded to be $5 \times 6 \times 6$ times for the cage recognition by HTR algorithm. It is tested that the cost time for cage recognition does not change with the cage numbers, suggesting that the system size (the number of water molecules) is the only factor for the computational efficiency of HTR algorithm.
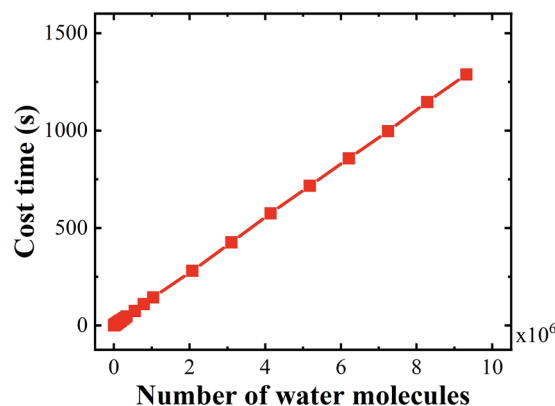

Figure 5: Cage recognition cost time by HTR algorithm as a function of block size (system: $5 \times 8 \times 10$ sI-type clathrate hydrate, 18,400 water molecules).
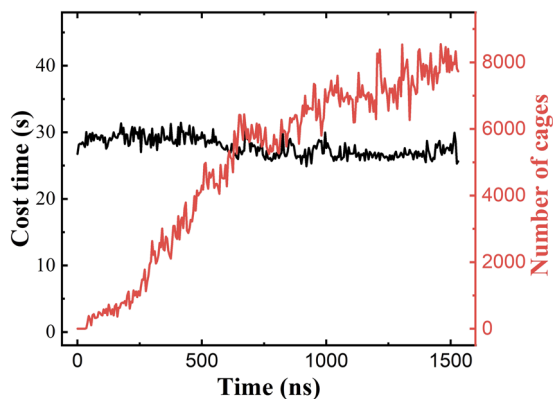

Figure 6: Cost time of cage recognition by HTR as a function of water molecules in sI-type clathrate hydrate systems.

Figure 7: Cost time (black) and number of cages (red) by HTR algorithm along with the simulation time of clathrate hydrate nucleation.

It is summarized from Figure 7 that HTR is able to accurately predict the time required for cage recognition as the input files are different. Table 1 lists the cost time of cage recognition for a similar system by HTR, as well as literature-reported FSICA [22], GRADE [25], and ICO [26] algorithms. Note that the MD trajectory with 50 frames is shared by FSICA, GRADE, and ICO for cage recognition. The information data such as the time required for cage recognition and the number of cages are provided by Hao *et al.* and Zhang *et al.* Because we are not able to have the same MD trajectory and the FSICA, GRADE, and ICO source codes, a system with similar molecular composition and 50-frame MD trajectories was achieved by us for the comparison test of the performance of HTR. The number of water molecules in the system by HTR is 3,726, while the number of water molecules in the system by FSICA, GRADE, and ICO algorithms is 3,487. It is found that the time consumed to identify cage recognition by HTR is not sensitive to the number of cages in the system.

As indicated by Table 1, HTR shows dozens of times higher efficiency in the cage recognition for small systems than FSICA, GRADE, and ICO, although it shows high efficiency in the cage recognition for large systems that are challenging for FSICA, GRADE, and ICO. It is worth

Table 1: In a hydrate system with basically the same number of water molecules, the comparison of the time is required for the four algorithms to identify. Among them, HTR is the data of this work, and the other three sets of data are from other articles

Algorithm	Total cages	Cost time (s)
HTR	15,724	26
FSICA	13,691	16,480 [26]
ICO	13,695	691 [31]
GRADE	4,410	449 [31]

noting that HTR and other algorithms (FSICA, GRADE, and ICO) do not use the same MD trajectory. The issues of recognition accuracy and identifiable cage types will be discussed further in the following section.

3.2 Accuracy of HTR algorithm on cage recognition of clathrate hydrate

To verify the accuracy of cage recognition by HTR, the hydrate nucleation simulation system composed of 1,242 water molecules and 216 methane molecules was used for HTR and GRADE to analyze the simulation trajectory with 1,000 ns. HTR was evaluated by comparing the results of cage recognition with GRADE. As is known, the performance of GRADE on cage recognition also depends on the setting of filter conditions $\delta1$ and $\delta2$. To ensure the accuracy of recognition, $\delta1$ and $\delta2$ are set to be 3 Å [8]. Figure 8a compares the variation in the numbers of 5^{12}, $5^{12}6^2$, $5^{12}6^4$, and other cages identified by HTR and GRADE with MD simulation time. For GRADE, there is no curve for "other cages" because GRADE only recognizes three types of cages including 5^{12}, $5^{12}6^2$, and $5^{12}6^4$. However, HTR recognizes more types of cages, that is, besides the common 5^{12}, $5^{12}6^2$, and $5^{12}6^4$ cages, there are other cages that can be identified by HTR.

To check whether there is over recognition of cages by HTR, the snapshots of cages recognized by HTR and GRADE are captured in Figure 8. Apparently, it can be seen that all the cages identified by GRADE can be also identified by HTR. However, there are more cages that are recognized by HTR. This is because GRADE does not consider the periodic boundary conditions and thereby is not able to identify the cages across the boundaries of the simulation box, while HTR recognizes the cages across the boundaries. This also further verifies the accuracy of HTR recognition.

3.3 Application of HTR algorithm in nucleation and micro/nanomechanics of clathrate hydrate

Figure 9 shows the snapshots of cage structures recognized by HTR for a methane hydrate nucleation system. As shown in Figure 9a, a $4^15^{10}6^2$ cage appeared at 30 ns and disappeared within a short simulation time due to the fluctuation of the system before the nucleation process. The number of cages gradually increases along with
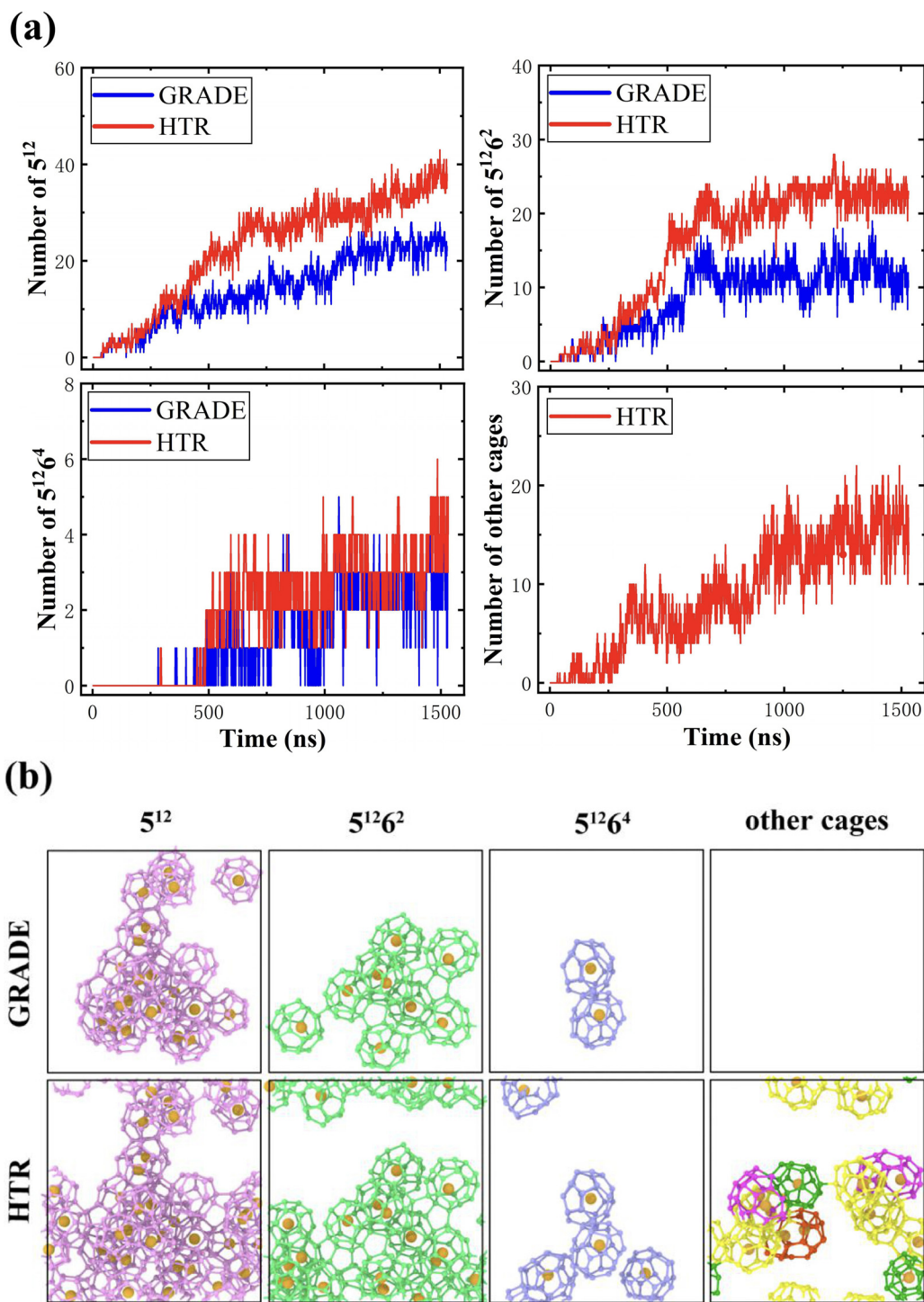
Figure 8: (a) Numbers of 5^{12}, $5^{12}6^2$, $5^{12}6^4$, and other cages identified by the HTR algorithm and the GRADE algorithm in the entire trajectory. (b) A snapshot of each cage identified by HTR and GRADE at 1,000 ns.

the simulation time. At about 250–300 ns, the fluctuation in the number of newly formed cages decreased rapidly and plateaued (Figure 9d), indicating that the nucleation process has been completed. After the simulation of 500 ns, the number of cages reached a plateau (Figure 9d) due to

the fact that nearly all the space of the system were occupied by the cages (Figure 9c). It is noticed that the unconventional cages of hydrate appeared throughout the trajectory, especially for the period of 250–300 ns (shown in Figure 8a), where the number of newly formed unconventional cages
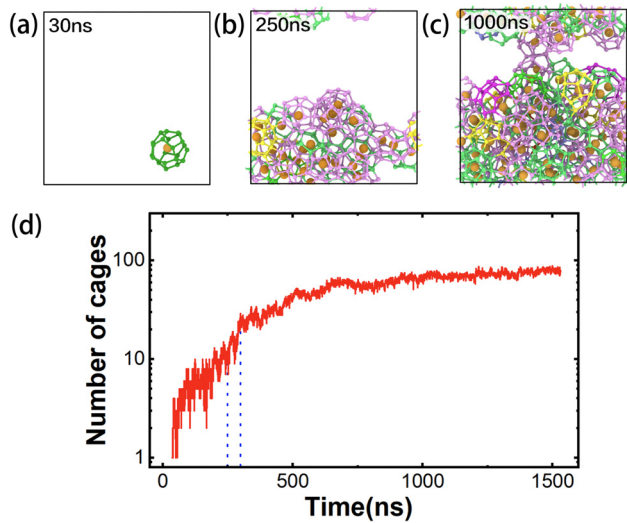
Figure 9: (a)–(c) snapshots at different times during the nucleation of clathrate hydrate. The different types of cage recognized by the HTR algorithm are represented by different colors: light green, pink, Roland purple, and other colors for 5^{12}, $5^{12}6^2$, $5^{12}6^4$, and other types of cages, respectively. (d) Number of newly formed cages in the system along with time.

increased significantly. These results suggest that the unconventional cages are closely related to the nucleation process of hydrate, which is consistent with the previous report: the nucleation/growth path can be well predicted according to some unconventional cage structures [32]. It should be pointed out that due to the complexity of the hydrate nucleation mechanism, one cannot draw a clear conclusion about the nucleation of hydrate based on such a simple test. While,

despite the importance of the unconventional cages on hydrate nucleation, there is no effective algorithm for the recognition of unconventional cages. Thus, the HTR algorithm can provide crucial data for the nucleation mechanism study of hydrate.

Micro/nanomechanics is another hot research topic in the study of the clathrate hydrate system. To expand the scope of the application of HTR, cage recognition of mechanically deformed sI methane hydrate system with dimensions of ~46.5 Å³ × 418.8 Å³ × 374.6 Å³, containing 36,864 methane molecules and 211,968 water molecules, was performed. Figure 10a shows the side views of the system at four uniaxial strains of 0.00, 0.22, 0.23, and 0.24. As the uniaxial strain is imposed from 0.00 to 0.22, the methane hydrate underwent elastic deformation. As a result, there is no dissociation of 5^{12} and $5^{12}6^2$ cages. In this deformational stage, both 5^{12} and $5^{12}6^2$ cages in the system are deformed. As indicated by Figure 10b, although they are mechanically deformed, they can be identified by HTR. For example, the number of 5^{12} and $5^{12}6^2$ cages identified by HTR before failure (strain varying from 0.00 to 0.22) remains to be 9,216 and 27,648, indicating that HTR can be applied to recognize highly deformed clathrate cages, which is of importance in the application of micro/nanomechanics study of clathrate hydrate. As the strain is applied to around 0.23, the methane hydrate fails *via* brittle cracks in the crystal. The brittle cracks can be characterized by local dissociation by both 5^{12} and $5^{12}6^2$ cages in the system, which is explained by the sudden drop in the number of 5^{12} and $5^{12}6^2$ cages, as illustrated in Figure 10b. Intriguingly, other types of clathrate cages form at the crack
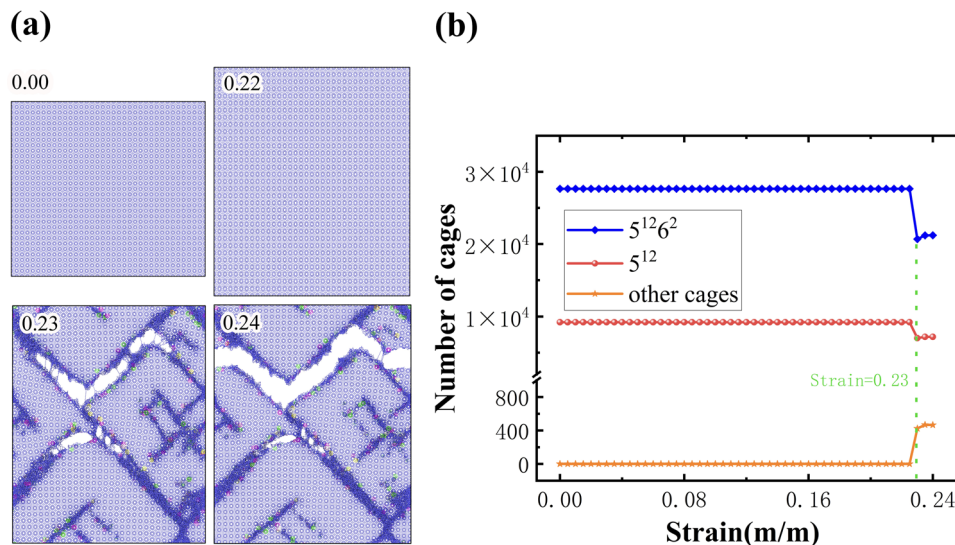
Figure 10: (a) Snapshots of strains ε = 0.00, 0.22, 0.23, and 0.24 of the sI-type clathrate hydrate system (with 211,968 water molecules). (b) Number of different cages identified by the HTR algorithm as a function of strains.
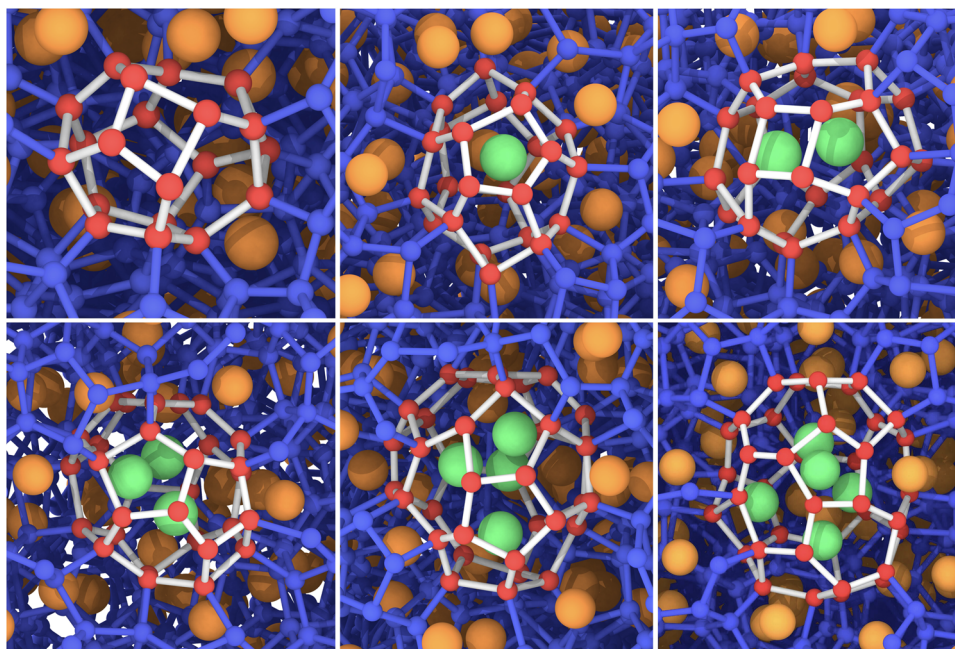
Figure 11: Snapshots from empty cages to quadruple-occupied cages in clathrate hydrate systems.

boundaries, which can be indicated by Figure 10a and b. (In Figure 10a, blue indicates 5^{12} and $5^{12}6^2$ cages, and other different colors indicate different types of newly generated cages.) As the uniaxial strain is imposed to 0.24, the system tends to be stable because the number of all kinds of clathrate cages remains constant.

3.4 Recognition of special cage structures in clathrate hydrate by HTR algorithm

The movement and distribution of guest molecules in clathrate cages play a vital role in the study of thermodynamics of clathrate hydrates. For example, Liang *et al.* [33] revealed the jumping mechanism of CO_2 hydrate near the dissociation temperature (310–320 K at 100 MPa) by MD simulations. Lo *et al.* [34] discovered the migration of guest molecules from one cage to another at a much lower temperature under melting conditions. Lu *et al.* [35] studied the behavior of guest molecules in the nucleation and growth of CO_2 hydrate in the presence of soluble ionic organic matter. To address this topic, HTR also integrates the function of tracking and locating the guest molecules, by calculating the center of mass (CoM) coordinates of each cage and searching for guest molecules in a sphere centered on the CoM of each cage. With this function, HTR can identify cages encapsulate none or multiple guest molecules. Here, a sI-type polycrystalline methane hydrate

composed of 314,664 coarse-grained water molecules was generated to test this function. As illustrated by Figure 11, one clathrate hydrate cage at the grain boundaries of a polycrystal is able to encapsulate different numbers of guest molecules (from 0 to 5). Note that multiple guest molecules@cage is instantaneously captured, and it is a metastable structure.

Another integrated function in HTR is the recognition of isomers cages. Commonly, the naming of one clathrate hydrate cage is defined according to the number of four-, five-, and six-membered rings in the cage, while the topology of the rings is ignored. From the topological perspective, a cage shows several possible arrangements of four-, five-, and six-membered rings, although it is composed of a specific number of polygonal rings. As such, a cage with identical polygonal compositions but different topology is termed as topological isomers of the cage. In HTR, the arrangement of the rings neighboring the four- and six-membered rings is the key to recognize isomers. Since HTR recognizes cages based on the topology of the ring, the arrangement of the rings around each ring can be easily determined, and thereby isomers of cages can be identified. In this study, the aforementioned large-scale polycrystalline methane hydrate sample subjected to uniaxial tension was employed to explore the isomers of the clathrate cage. As is known for 5^{12} cage, there is no isomer of the clathrate cage due to the fact that it is only composed of five-membered rings.

(a)

| $4^15^{10}6^2$ | $5^{12}6^4$ | $5^{12}6^5$ | $5^{12}6^8$ |

(b)
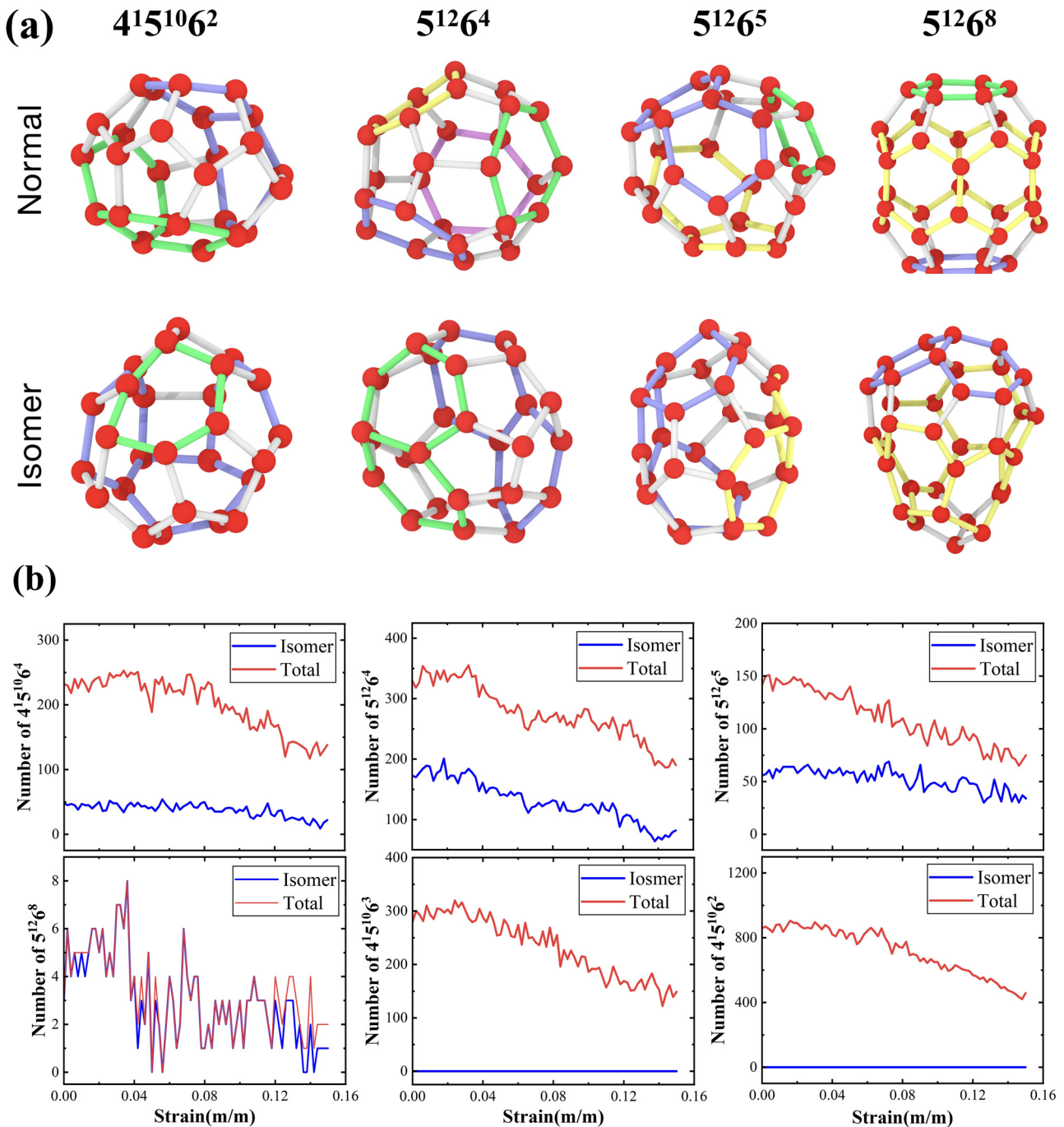
Figure 12: HTR identified isomer cages in the polycrystalline stretching process of a system of 314,664 water molecules. (a) The colored rings are all six-membered rings, and two six-membered rings that share a hydrogen bond are represented by the same color. (b) Number of isomers and total cages along with the stretching process.

Figure 12a shows the topological isomers of several clathrate cages ($4^15^{10}6^4$, $5^{12}6^4$, $5^{12}6^5$, and $5^{12}6^8$) identified from the deforming polycrystalline sample. For clarification of topological isomers in the cages, the six-membered rings are highlighted, and the two adjacent six-membered rings that share a hydrogen bond are identically colored.

Apparently, for each type of the aforementioned clathrate cages, there are topological isomers. Figure 12b shows the variations in the total number of cages and the number of isomers of cages in the polycrystalline system subjected to uniaxial strain. Apparently, there is a reduction tendency in the total number of all types of clathrate cages. For the

case of the isomers of cages, there is a negligible change in the number of topological isomers of $4^1 5^{10} 6^4$ and $5^{12} 6^8$ clathrate cages, while the number of topological isomers of $5^{12} 6^4$ and $5^{12} 6^5$ clathrate cage is changeable with uni-axial strain, and the change range of the number of isomers of $5^{12} 6^5$ is small. However, it is identified that there are no topological isomers of $4^1 5^{10} 6^2$ and $4^1 5^{10} 6^3$ clathrate cages. By comparing with the conventional cages, the topological isomers of $5^{12} 6^8$ clathrate cages are predominated, while the topological isomers of $4^1 5^{10} 6^4$ clathrate cages are the least.

4 Conclusion

In this study, an efficient, accurate, and multifunctional HTR algorithm is introduced for the recognition of clathrate cages in a hydrate-based system. HTR is based on ring topology to identify cage structures, and the efficiency of cage recognition by HTR is innovatively improved *via* the splitting method, leading to the recognition time–the system size relationship changing from exponential to linear one. As a result, HTR can be effectively used for cage recognition of large-scale systems. HTR avoids over-identification of clathrate cages, and the accuracy and efficiency of cage identification by HTR were verified with other popular algorithms. Uniquely, HTR is able to identify multiple guests@cages and topological isomer structures of clathrate cages, as well as clathrate cages subjected to mechanical loads. The presented HTR algorithm is accessible for extracting helpful insights into the information of clathrate cages in a large-scale system of MD simulations.

References

[1] Sloan Jr ED, Koh CA. Clathrate hydrates of natural gases. Boca Raton: CRC Press; 2007.

[2] Macdonald GJ. The future of methane as an energy resource. Annu Rev Energy. 1990;15:53–83.

[3] Koh CA, Sum AK, Sloan ED. Gas hydrates: unlocking the energy from icy cages. J Appl Phys. 2009;106:9.

[4] Sum AK, Koh CA, Sloan ED. Clathrate hydrates: from laboratory science to engineering practice. Ind Eng Chem Res. 2009;48:7457–65.

[5] Kim SM, Lee JD, Lee HJ, Lee EK, Kim Y. Gas hydrate formation method to capture the carbon dioxide for pre-combustion process in IGCC plant. Int J Hydrogen Energ. 2011;36:1115–21.

[6] Dashti H, Thomas D, Amiri A. Modeling of hydrate-based CO_2 capture with nucleation stage and induction time prediction capability. J Clean Prod. 2019;231:805–16.

[7] Park Y, Kim DY, Lee JW, Huh DG, Park KP, Lee J, et al. Sequestering carbon dioxide into complex structures of naturally occurring gas hydrates. Proc Natl Acad Sci U S A. 2006;103:12690–4.

[8] Jacobson LC, Hujo W, Molinero V. Thermodynamic stability and growth of guest-free clathrate hydrates: a low-density crystal phase of water. J Phys Chem B. 2009;113:10298–307.

[9] Mao WL, Mao HK, Goncharov AF, Struzhkin VV, Guo QZ, Hu JZ, et al. Hydrogen clusters in clathrate hydrate. Science. 2002;297:2247–9.

[10] ten Wolde PR, Frenkel D. Enhancement of protein crystal nucleation by critical density fluctuations. Science. 1997;277:1975–8.

[11] Maddah M, Maddah M, Peyvandi K. Investigation on structural properties of winter flounder antifreeze protein in interaction with clathrate hydrate by molecular dynamics simulation. J Chem Thermodyn. 2021;152:106267.

[12] Pandey HD, Leitner DM. Thermodynamics of hydration water around an antifreeze protein: a molecular simulation study. J Phys Chem B. 2017;121:9498–507.

[13] Mu L, Ramløv H, Søgaard TMM, Jørgensen T, de Jongh WA, von Solms N. Inhibition of methane hydrate nucleation and growth by an antifreeze protein. J Petrol Sci Eng. 2019;183:106388.

[14] Yin ZY, Chong ZR, Tan HK, Linga P. Review of gas hydrate dissociation kinetic models for energy recovery. J Nat Gas Sci Eng. 2016;35:1362–87.

[15] Liu H, Kumar SK, Douglas JF. Self-assembly-induced protein crystallization. Phys Rev Lett. 2009;103:018101.

[16] Debenedetti PG, Sarupria S. Chemistry. Hydrate molecular ballet. Science. 2009;326:1070–1.

[17] Walsh MR, Rainey JD, Lafond PG, Park DH, Beckham GT, Jones MD, et al. The cages, dynamics, and structuring of incipient methane clathrate hydrates. Phys Chem Chem Phys. 2011;13:19951–9.

[18] Christiansen RL, Sloan ED. Mechanisms and kinetics of hydrate formation. Ann Ny Acad Sci. 1994;715:283–305.

[19] Sloan ED, Fleyfel F. A molecular mechanism for gas hydrate nucleation from ice. Aiche J. 1991;37:1281–92.

[20] Yin ZY, Khurana M, Tan HK, Linga P. A review of gas hydrate growth kinetic models. Chem Eng J. 2018;342:9–29.

[21] Khurana M, Yin ZY, Linga P. A review of clathrate hydrate nucleation. ACS Sustain Chem Eng. 2017;5:11176–203.

[22] Guo GJ, Zhang YG, Li M, Wu CH. Can the dodecahedral water cluster naturally form in methane aqueous solutions? A molecular dynamics study on the hydrate nucleation mechanisms. J Chem Phys. 2008;128:194504.

[23] Guo GJ, Zhang YG, Liu CJ, Li KH. Using the face-saturated incomplete cage analysis to quantify the cage compositions and cage linking structures of amorphous phase hydrates. Phys Chem Chem Phys. 2011;13:12048–57.

[24] Nguyen AH, Molinero V. Identification of clathrate hydrates, hexagonal ice, cubic ice, and liquid water in simulations: the CHILL + algorithm. J Phys Chem B. 2015;119:9369–76.

[25] Mahmoudinobar F, Dias CL. GRADE: a code to determine clathrate hydrate structures. Comput Phys Commun. 2019;244:385–91.

[26] Hao Y, Xu Z, Du S, Yang X, Ding T, Wang B, et al. Iterative cup overlapping: an efficient identification algorithm for cage structures of amorphous phase hydrates. J Phys Chem B. 2021;125:1282–92.

[27] Matsumoto M, Baba A, Ohmine I. Topological building blocks of hydrogen bond network in water. J Chem Phys. 2007;127:134504.

[28] Laage D, Hynes JT. A molecular jump mechanism of water reorientation. Science. 2006;311:832–5.

[29] Dias CL, Ala-Nissila T, Grant M, Karttunen M. Three-dimensional "Mercedes-Benz" model for water. J Chem Phys. 2009;131:054505.

[30] Mahmoudinobar F, Dias CL, Zangi R. Role of side-chain interactions on the formation of alpha-helices in model peptides. Phys Rev E. 2015;91:032710.

[31] Zhang Z, Guo G-J. Comment on "iterative cup overlapping: an efficient identification algorithm for cage structures of amorphous phase hydrates". J Phys Chem B. 2021;125:5451–3.

[32] Li LW, Zhong J, Yan YG, Zhang J, Xu JF, Francisco JS, et al. Unraveling nucleation pathway in methane clathrate formation. Proc Natl Acad Sci U S A. 2020;117:24701–8.

[33] Liang S, Liang D, Wu N, Yi L, Hu G. Molecular mechanisms of gas diffusion in CO_2 hydrates. J Phys Chem C. 2016;120:16298–304.

[34] Lo H, Lee M-T, Lin S-T. Water vacancy driven diffusion in clathrate hydrates: molecular dynamics simulation study. J Phys Chem C. 2017;121:8280–9.

[35] Lu Y, Sun L, Guan D, Yang L, Zhang L, Song Y, et al. Molecular behavior of CO_2 hydrate growth in the presence of dissolvable ionic organics. Chem Eng J. 2022;428:131176.