PCCP

View Online
Dynamic Article Links

Cite this: *Phys. Chem. Chem. Phys.*, 2011, **13**, 12048–12057

www.rsc.org/pccp

PAPER

Using the face-saturated incomplete cage analysis to quantify the cage compositions and cage linking structures of amorphous phase hydrates†

Guang-Jun Guo,*[a] Yi-Gang Zhang,[b] Chan-Juan Liu[a] and Kai-Hua Li[b]

Received 8th January 2011, Accepted 26th April 2011
DOI: 10.1039/c1cp20070d

Recent advances in the molecular dynamics simulations of spontaneous nucleation and growth of methane hydrate show that an amorphous phase of the hydrate is first reached. However, the amorphous hydrate has not been well described, due to the insufficient identification of cage structures. Here, we develop a method, called "face-saturated incomplete cage analysis", which can identify all face-saturated cages in a given system. As a result, it is found that thousands of cage types and abundant occupancy states are present in the amorphous hydrate. Moreover, the crystallinity of amorphous hydrate is evaluated according to the quantitative calculation of cage linking structures, and the critical nucleus of hydrate is also estimated on the basis of clustering analysis for all face-saturated cages.

1. Introduction

Gas hydrates are ice-like crystalline compounds comprised of guests (*e.g.*, methane, CO_2) trapped in a clathrate, hydrogen-bond network of water. Because of their roles in a range of important scientific and technical problems in the fields of energy resources, the environment, chemical industries and geological disaster prevention, they have become an enduring research topic.[1] Hydrates are known to have three main clathrate structures, sI, sII, and sH, in which five types of polyhedral water cages combine to fully tessellate in three-dimensional space by sharing cage faces.[2] However, many questions remain concerning the formation of these structures. There has been some success in simulating ice freezing from water;[3] however no simulation has yet succeeded in forming highly crystalline hydrate structures from a gas–water mixture.

Recently, using different calculative strategies, several groups observed the spontaneous nucleation and growth of methane hydrate *via* classical molecular dynamics (MD) simulations. For example, Rodger and co-workers[4,5] utilized the very high concentration of methane at a water–methane interface, Walsh *et al.*[6] carried out very long MD simulations on the scale of microseconds, Jacobson *et al.*[7,8] adopted a coarse-grained model of water molecules to speed up calculation, and Vatamanu and Kusalik[9] used a steady-state condition during simulations. Their results indicate that the methane–water mixture evolves first into an amorphous phase of methane hydrate. However, their descriptions for the amorphous phase are incomplete and qualitative due to the insufficient cage identification. For example, Rodger and co-workers[5] identified only three types of cages, including 5^{12}, $5^{12}6^2$, and $5^{12}6^4$. Walsh *et al.*[6] and Jacobson *et al.*[7,8] identified one more cage type of $5^{12}6^3$ than Rodger and co-workers on the basis of a cage searching algorithm.[10] Vatamanu and Kusalik[9] recorded decades of cage types by visual observation. Obviously, these methods cannot ensure that all cage types in the system are identified, and so cannot fully characterise the amorphous hydrates and how it differs from crystalline hydrates.

In this study we report a method called the face-saturated incomplete cage analysis (FSICA). The FSICA can identify all possible face-saturated cage compositions in a system and thus gives a complete picture of system evolution. Here, the face-saturated cages include not only the polyhedral complete cages but also the face-saturated incomplete cages, whose definition can be found below. We use the method to re-analyze the nucleation trajectories previously reported by Walsh *et al.*[6] with the purpose of verifying our conjecture that these nucleation trajectories actually reach the amorphous hydrate phase, which had not been well identified by the authors. The calculation of linking structures among all existing face-saturated cages leads to a quantitative description of the amorphous hydrate phase and enables us to conclude that the formed phase is still a long distance from the crystalline hydrate phase. Furthermore, clustering analysis for all face-saturated cages lets us estimate the critical nucleus of hydrate.

2. Methods

First, the concept of incomplete cages, defined elsewhere,[11] is introduced briefly. A perfect polyhedron has two important

[a] *Key Laboratory of the Earth's Deep Interior, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, People's Republic of China. E-mail: guogj@mail.igcas.ac.cn*
[b] *State Key Laboratory of Lithospheric Evolution, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, People's Republic of China*

† Electronic supplementary information (ESI) available: Some FSICA results for T2μs and T5μs, including Fig. S1–S8 and Tables S1–S3. See DOI: 10.1039/c1cp20070d

features, edge-saturation and face-saturation. The former means that every vertex is shared with at least three edges, and the latter means that every edge is shared with exactly two faces. If a polyhedron-like cage structure meets the edge-saturated and face-saturated conditions at the same time, it will be a complete cage (CC), otherwise it will be an incomplete cage (IC). In this work, we only study the cages meeting at least the face-saturated condition, which include CCs and face-saturated incomplete cages (FSICs), but exclude other ICs. The importance of face-saturated cages is two-fold. First, each face-saturated cage is free of holes, so that the cage can adsorb dissolved gas molecules (like methane) on its faces and keep the adsorbed methane from contacting the encaged methane. Because this process favors the occurrence of clathrate structures rather than methane bubbles, the face-saturated cages may act as the precursors to hydrate nucleation. Second, the surface area and volume of face-saturated cages can be well defined so that they can be treated as building blocks of the amorphous hydrate and their linking relationship can be calculated without ambiguity.

Based on the IC concept, we have established a method called incomplete cage analysis (ICA) to quantify the cage-like degree of the hydration shell of aqueous methane and to search for the cage structures in a methane solution. However, the earlier ICA method could not search empty cages, because it requires the position of an existing methane molecule to test for the existence of a surrounding cage. In this study, we greatly improve the ICA method to be able to search for empty cages by feeding it all specified grid points in the space of system one by one. Because we set the resolution of grid points to be smaller than the minimum cage diameter, the method will not miss any face-saturated cages that occur in the simulated system unless their size exceeds the given lower- and upper-limits or their shapes are bent (*e.g.*, considering a tire-like structure as an extreme case). In this work, we denote the improved ICA as the FSICA because only face-saturated cages are searched.

The method for implementing FSICA is as follows. In a pure water system, which forms a network structure through H-bonds, there are typically voids existing in the network. Our algorithm aims to find each void and judge whether the water molecules enclosing the void form a polyhedron-like cage structure. Obviously, some geometric elements of H-bond topological structure should be identified first, including vertices (*i.e.*, oxygen atoms), edges (*i.e.*, H-bonds), and faces (*i.e.*, water rings). Then, the whole system is scanned using a 3-dimensional spatial grid. If the separation between neighboring grid points is small enough relative to the size of void, several grid points will be located within each void. Taking one of them as a target, one can consider the relationship of all nearby water molecules to the void. If the sight line between a water molecule and the target crosses any faces, the water molecule must be located outside the void in the cage structure. After those external water molecules are removed, the remaining H-bond topological structure is calculated again and those water molecules not belonging to any faces are further abandoned. Then, any water molecules left behind are the cage makers that contribute directly to enclosing the void.

To judge whether cage makers form a polyhedron-like structure, their edge-saturated index (ζ_V) and face-saturated index (ζ_E) are evaluated by

$$\zeta_V = \frac{n_V^{E3+}}{n_V}, \quad (1)$$

where n_V is the total number of cage makers and n_V^{E3+} is the number of cage makers shared with at least three edges, and by

$$\zeta_E = \frac{n_E^{F2}}{n_E}, \quad (2)$$

where n_E is the total number of edges and n_E^{F2} is the number of edges shared with two faces. The cage-like degree of the structure that these cage makers form is then defined as,

$$\zeta_C = \zeta_V \zeta_E. \quad (3)$$

$\zeta_C = 1$ corresponds to a CC, and $\zeta_C \neq 1$ but $\zeta_E = 1$ corresponds to an FSIC. Other cases signify an IC. In this study, only CCs and FSICs, both meeting the face-saturated condition, are considered for further study.

The above is the basic procedure of FSICA. There are a number of points about the implementation procedure that are worth noting.

(1) H-bond: we adopt a widely used geometric definition of the hydrogen bond,[12] *i.e.*, $r_{OO} < 3.5$ Å and $\angle HOO < 30°$, where r_{OO} is the distance between oxygen atoms and $\angle HOO$ is the angle between the OH bond and the OO vectors.

(2) Ring searching: we search only simple rings in which 3 or 4 arbitrary sequential vertices cannot form a triangular or quadrangular ring. The largest ring considered in our implementation of the FSICA is the hexagonal ring.

(3) Cage faces: we define only 3-, 4-, 5-, and 6-membered water rings as cage faces; 7- and larger membered rings are considered cage holes rather than faces. The physical meaning of faces and holes is that faces can keep methane from crossing them while holes cannot. A detailed discussion of this definition is given later in this paper.

(4) Hunting sites: these are the sites we use to hunt for possible cage structures. Besides all system grid points, hunting sites include some water sites where each water molecule forms at most two H-bonds. When hunting for cages at these water sites, the involved water molecules will not be considered as possible cage makers. In such a way, water-filled cages can be found. In addition, the distance between neighboring grid points is set at 1 Å. Methane sites need not be considered.

(5) Cage type: the cage type is determined with five characteristic parameters, *i.e.*, the numbers of triangular (T), quadrangular (Q), pentagonal (P), and hexagonal (H) faces, and the number of vertices with only two edges (Z). These structural elements are thus denoted as $[3^T4^Q5^P6^H]_Z$. From this symbol one can further calculate the numbers of vertices (V), edges (E), and faces (F) of each cage, as well as the cage-like degree (ζ_C) through

$$F = T + Q + P + H, \quad (4)$$

$$E = (3T + 4Q + 5P + 6H)/2, \quad (5)$$

$$V = E - F + 2, \quad (6)$$

$$\zeta_C = 1 - Z/V. \quad (7)$$

(6) Cage size: as for the lower-limit, a cage must accommodate at least a sphere with a diameter of 2.82 Å (*i.e.*, the average length of H-bonds). Thus, the minimum CC expected is a cubic cage. As for the upper-limit, although the FSICA can search large cages without limit, we set the maximum radius of cage to 12 Å with consideration to the calculation speed. Using a value of 14 Å has been tried for some frames of trajectories, and yielded the same results as using an upper limit of 12 Å.

(7) Repeated cage: if the distance between two cages is smaller than 2.82 Å, we identify them as the same cage, in which case only the one with a larger value of ζ_C is recorded.

(8) Merging triangular faces: because of the deformation of cage faces, some large faces may transform into two smaller faces. For example, a pentagon changes to a quadrangle and a triangle. Triangular faces can be either independent or dependent. Merging a triangular face with its neighboring faces will reduce the ζ_C value of cage for the independent but not for the dependent faces. In the FSICA, each dependent triangular face is merged with its smallest neighboring face to avoid outputting too many triangle-bearing cages.

(9) Cage area and volume: by using the face center, each face can be divided into sub-triangles, so the cage area is the sum of all sub-triangle areas. By further using the cage center, each sub-triangle corresponds to a sub-tetrahedron of the cage, so the cage volume is the sum of the volume of all sub-tetrahedrons.

(10) Cage vertex- and face-radii: the vertex-radius corresponds to the average distance between the cage center and vertices, and the face-radius corresponds to the average distance between the cage center and face centers.

(11) Convex and concave cage criteria: we divide a cage into two parts using the extended plane of a cage face, and count the number of vertices for each part. Then, the vertex fraction for the larger part is calculated, and the minimum value over all faces is found. Experientially, the cage is convex if its minimum vertex fraction is larger than 0.8; otherwise it is concave.

(12) Guest methane and water: if the "line of sight" between a methane molecule and the cage center does not cross any cage faces, then the methane molecule is a guest for that cage. Guest water molecules are found in a corresponding manner.

(13) Adsorbed methane: each face of a cage has an adsorption site. We define that the vector from the cage face to its adsorption site is perpendicular to the face and the length of the vector is 3 Å.[13] Methane molecules separating within 3 Å of an adsorption site are labeled as the adsorbed.

3. Results and discussion

We use the FSICA to re-analyze the two classical MD simulation trajectories showing methane hydrate nucleation and growth reported by Walsh *et al.*[6] Both simulation systems consisted of 512 united atom methane molecules and 2944 TIP4P-ice water molecules. Their initial configurations were prepared by melting sI hydrate until the phase separation of gaseous methane and liquid water was complete. One run lasted 2 μs at 250 K and 50 MPa, and the other lasted 5 μs at 260 K and 45 MPa (denoted as T2μs and T5μs hereafter).

Table 1 Results of total cage types and cage numbers in T2μs and T5μs[a]

	T2μs	T5μs
CC type	732	1258
FSIC type	5975	7015
CC number	983 863	5 520 072
FSIC number	1 421 434	2 565 521
Frame number	19 999	49 999

[a] These data are accumulated from trajectory frames sampled at an interval of 100 ps.

Surprisingly, the FSICA identifies thousands of cage types within T2μs and T5μs (see Table 1). In these cage types, *i.e.*, $[3^T4^Q5^P6^H]_Z$, the ranges of characteristic parameters are found to be $0 \le T \le 2$, $0 \le Q \le 12$, $0 \le P \le 42$, $0 \le H \le 24$, and $0 \le Z \le 17$. Because the triangle-bearing cages account for only about 0.3% of total cage numbers, we do not discuss them in the following despite that they cover up to 17.6% of the total cage types in T5μs. The notation for CC types (but not IC) can be further reduced to the conventional form $4^Q5^P6^H$ since the Z value is zero for CCs.

Now we describe the overall features of CCs. The smallest one is an empty 4^55^2 (Fig. 1A) found in both T2μs and T5μs. The largest one is a concave $4^25^{36}6^{14}$ (Fig. 1B) in T2μs, equivalent to four ICs fused together through sharing their holes. The occupancy states of cages also vary. In addition to being empty or filled with a single methane molecule, the cages can also be filled with multiple methane and/or water molecules. This is easily understood because of the existence of complex, concave cages (*e.g.* Fig. 1B). However, we also observed a convex $5^{12}6^8$ cage containing three methane molecules (Fig. 1C) and even a convex $4^25^86^5$ filled with three water molecules (Fig. 1D). In this study, the face arrangements of the five standard hydrate cages have been checked according to Table 2, and it was found that 5^{12} and $5^{12}6^2$ were in their most common respective structures without exception, but that other sub-types of $5^{12}6^4$ and $4^35^66^3$ were observed (Fig. 1E and F) in addition to the standard structures. We did not find the standard $5^{12}6^8$ hydrate cage in either T2μs or T5μs; the one that is presented (Fig. 1C) is a sub-type because all of its eight hexagonal faces connect together, different from the standard structure that contains two isolated hexagonal faces. The 10 most abundant cage types for T2μs and T5μs are listed in Table 3 and Table S1 (ESI†), respectively. Note that the $4^15^{10}6^2$ cage (Fig. 1G) is obviously an important CC with a rank of 2 in T2μs and of 3 in T5μs. To compare with the results of Walsh *et al.*, we plot the evolution of four CCs they monitored: 5^{12}, $5^{12}6^2$, $5^{12}6^3$, and $5^{12}6^4$ (Fig. S1 and S2, ESI†). One can see that the FSICA reproduces their results (see Fig. 3 and S3 in ref. 6) very well. The evolution of other selected CCs in T2μs is shown in Fig. 2A. Although these cages start to increase from the nucleation time, *i.e.*, ~1.2 μs, the first CC formed in the system is a 4^55^2 cage (Fig. 1A) occurring at 0.0656 μs and the first 5^{12} cage occurs at 0.361 μs in T2μs, both substantially earlier than 1.2 μs.

As for the FSICs, their features are similar to those of CCs in size and occupancy, but not in occurrence. FSICs can be classified into two groups. One group, including examples such

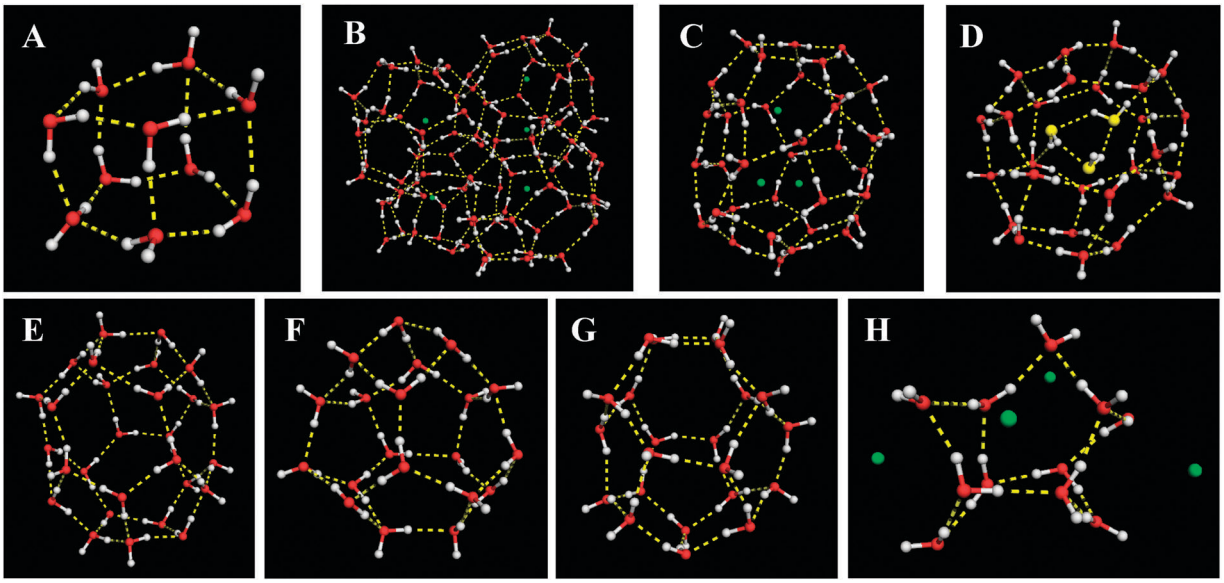

Fig. 1 Cage snapshots from T2μs and T5μs. The red balls are O atoms, the white are H atoms, and the green are CH_4 (methane). The yellow dashed lines are H-bonds. (A) The smallest CC, $4^5 5^2$. (B) The largest CC, $4^2 5^{36} 6^{14}$. (C) A $5^{12} 6^8$ filled with three CH_4 molecules. (D) A $4^2 5^8 6^5$ filled with three H_2O molecules, whose O atoms (yellow) form a triangular ring. (E) A sub-type of $5^{12} 6^4$ in which connected hexagonal faces exist, different from the $5^{12} 6^4$ in sII hydrate. (F) A sub-type of $4^3 5^6 6^3$ in which connected hexagonal faces exist, different from the $4^3 5^6 6^3$ in sH hydrate. (G) A $4^1 5^{10} 6^2$ cage. (H) A $[5^2 6^3]_5$ cage which adsorbs four CH_4 molecules.

Table 2 Face arrangement criteria for identifying standard hydrate cages[a]

Type	Name	Criterion of face arrangement
5^{12}	D	$N_P(p5) = 12$
$5^{12}6^2$	T	$N_P(p4 + h1) = 12$, $N_H(p6) = 2$
$5^{12}6^4$	H	$N_P(p3 + h2) = 12$, $N_H(p6) = 4$
$5^{12}6^8$	L	$N_P(p2 + h3) = 12$, $N_H(p4 + h2) = 6$, $N_H(p6) = 2$
$4^3 5^6 6^3$	S	$N_Q(p2 + h2) = 3$, $N_P(q1 + p2 + h2) = 6$, $N_H(q2 + p4) = 3$

[a] In the last column, N_Q, N_P, and N_H correspond to the numbers of quadrangular, pentagonal, and hexagonal faces with specific types in a cage. The face types are described in brackets, where q, p, and h mean the quadrangular, pentagonal, and hexagonal neighbouring faces of the corresponding face types, followed by their numbers.

Table 3 The top 10 CCs in T2μs[a]

Type	Distribution	R_{vertex}/Å	R_{face}/Å	Area/Å^2	Volume/Å^3
5^{12}	0.150	3.899	3.096	160.2	163.4
$4^1 5^{10} 6^2$	0.056	4.130	3.359	181.5	195.7
$5^{12}6^2$	0.054	4.309	3.562	200.4	230.5
$5^{12}6^3$	0.037	4.494	3.765	220.7	268.7
$4^1 5^{10} 6^3$	0.022	4.333	3.595	201.8	229.5
$4^1 5^{10} 6^4$	0.015	4.506	3.786	221.4	266.9
$5^{12}6^4$	0.013	4.670	3.959	241.1	309.6
$4^2 5^8 6^2$	0.010	3.958	3.167	162.9	162.2
$4^2 5^8 6^4$	0.009	4.351	3.609	203.4	231.6
$5^{12}6^5$	0.005	4.866	4.187	262.3	349.0

[a] In Tables 3 and 4, the cage distribution is calculated as the fraction of cage number to total cage number (including both FSICs and CCs). R_{vertex} and R_{face} mean the vertex- and face-radii of cage, respectively.

as $[5^2 6^3]_5$, $[5^2 6^4]_5$, $[4^1 5^2 6^3]_4$, is characterized by a small number of cage vertices and empty occupancy. They form at the very beginning of trajectories and dominate before the nucleation time in T2μs and T5μs (Fig. 2B and Fig. S3B, ESI†). The other group, including $[5^{10} 6^2]_1$, $[5^{10} 6^4]_1$, and $[4^1 5^8 6^3]_1$, is characterized by a large number of cage vertices. Their behavior is similar to that of CCs, and they mainly occur after hydrate nucleation. The top 10 FSIC types by occurrence for T2μs and T5μs are listed in Table 4 and Table S2 (ESI†), respectively. In fact, the occurrence of small FSICs is an inherent feature of liquid water, as demonstrated by Matsumoto *et al.*[14] Those authors found that the H-bond network in water can be tessellated into "quasipolyhedrons", similar to FSICs. Two of them, "J" and "K" in Fig. 4 of ref. 14, correspond to the $[5^2 6^3]_5$ (Fig. 1H) and $[6^4]_6$ cages found in this work. Interestingly, the small, empty FSICs play an important role in dissolving methane into water. The inset of Fig. 2C shows that about 50% of dissolved methane is adsorbed by FSICs before hydrate nucleation in T2μs.

We have also sorted CCs and FSICs into different cage groups by size. Each group contains many cages, covering different cage types, but these cages are characteristic of the same number of vertices (V). We found the distribution of cage groups to show an unexpected feature. For groups with smaller values of V, such as $V \leq 33$, the probability of finding CC-groups with the even V is always more intense than those neighbor groups with the odd V by 1–2 orders of magnitude (Fig. 3 and Fig. S4, ESI†). However, the opposite occurs for

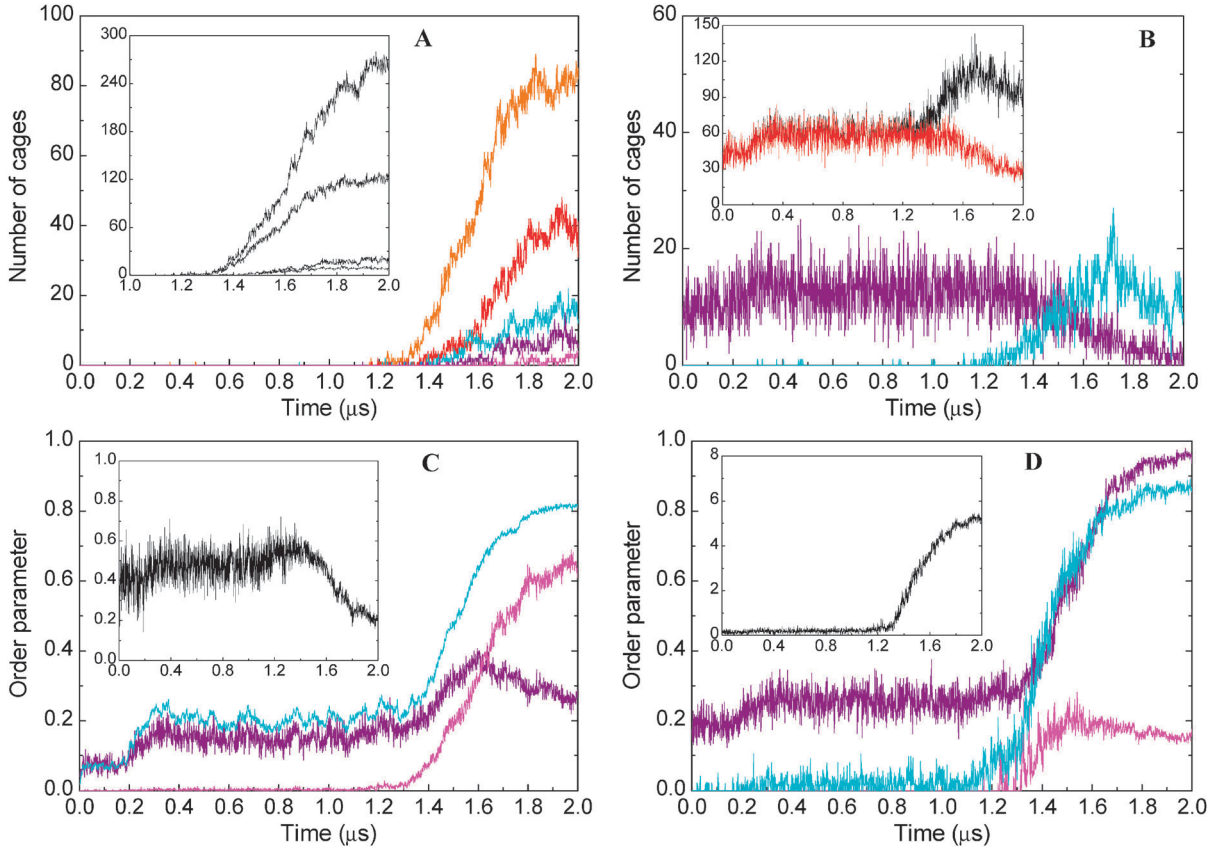

Fig. 2 (A) Evolution of selected CCs, including 5^{12} (orange), $4^1 5^{10} 6^2$ (red), $4^1 5^{10} 6^3$ (cyan), $4^2 5^8 6^2$ (purple), and $4^3 5^6 6^3$ (magenta). The four black lines in the inset, from the top down, are for CCs, standard hydrate cages, empty CCs, and empty standard hydrate cages. (B) Evolution of selected FSICs, including $[5^2 6^3]_5$ (purple) and $[5^{10} 6^2]_1$ (cyan). In the inset, the black and red lines refer to FSICs and empty FSICs, respectively. (C) Evolution of several order parameters, including fractions of different types of methane to total methane, *i.e.*, the aqueous methane (cyan), the guest methane (magenta), and the adsorbed methane (purple). The inset shows the fraction of the adsorbed aqueous methane to total aqueous methane. In this study, a methane molecule is classified as aqueous if it has 16 or more hydration water molecules, otherwise it belongs to the gaseous class. The radius of the hydration shell of a methane molecule is set at 5.4 Å. Note: the peak at ~1.6 μs in the curve for adsorbed methane actually results from using the periodic boundary conditions (PBCs). During the growth of a cage cluster, the surface of the cluster will reduce and some adsorbed methanes have to change to guest methanes when the cage cluster spans the simulation box. (D) Evolution of several order parameters, including the cage occupancy (cyan) which is the fraction of occupied cages to total cages, the fraction of cage water to total water (purple), and the hydrate crystallinity (magenta) which is the ratio of the number of standard hydrate cage links (including sI, sII, and sH) to total cage links. In the inset, the black line represents the links per cage, *i.e.*, the ratio of total cage links to total cages, and so measures the aggregate degree of cages. Note: there is an unusual peak at ~1.55 μs in the curve of hydrate crystallinity. We infer that it is probably caused by the PBC mentioned above because the cages induced by the PBC may disturb the natural growth of the cage cluster, but we know little about how the hydrate crystallinity really changes with time if the PBC can be removed.

Table 4 The top 10 FSICs in T2μs

Type	Distribution	R_{vertex}/Å	R_{face}/Å	Area/Å^2	Volume/Å^3
$[5^2 6^3]_5$	0.084	3.051	1.687	79.7	37.6
$[5^2 6^4]_5$	0.068	3.288	2.047	98.1	54.4
$[4^1 5^2 6^3]_4$	0.034	3.154	1.929	88.5	47.7
$[5^{10} 6^2]_1$	0.031	4.028	3.192	171.9	176.0
$[5^2 6^5]_5$	0.022	3.510	2.362	116.7	73.9
$[5^4 6^3]_4$	0.015	3.359	2.227	105.9	66.5
$[5^{10} 6^3]_1$	0.013	4.231	3.429	191.2	205.6
$[4^1 5^2 6^4]_4$	0.011	3.391	2.259	107.0	65.9
$[5^4 6^4]_4$	0.010	3.599	2.544	124.9	85.4
$[6^5]_6$	0.010	3.201	1.761	91.6	47.7

FSIC-groups, albeit with a few exceptions and a smaller difference (less than one order of magnitude). These phenomena may be related to another interesting topic concerned with

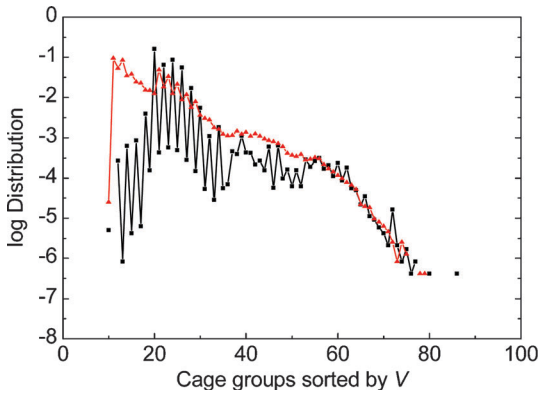

Fig. 3 Distribution of CC-(black squares) and FSIC-groups (red triangles) sorted by the number of cage vertices (V) for T2μs.

Table 5 Characteristic cage links in the unit cells of three hydrate structures[a]

Structure	Cage composition	Cage links	Links number	Links per cage	Hydrate crystallinity
sI	D_2T_6	D-5-T	24	6.75	1
		T-5-T*	24		
		T-6-T	6		
sII	$D_{16}H_8$	D-5-D	48	6.67	1
		D-5-H	96		
		H-6-H	16		
sH	$D_3S_2L_1$	S-4-S*	3	6.67	1
		S-5-D	12		
		S-6-L*	6		
		D-5-D	6		
		D-5-L	12		
		L-6-L*	1		

[a] In the second and third columns, the capital characters are the names of hydrate cages as listed in Table 2. The numbers in the second column are the numbers of cages, and those in the third column are the size of linking faces. Those cage links requiring specific orientational restraints, labelled with asterisks (*), are listed below with the detailed explanations. T-5-T*: only one vertex in the linking face is shared with two hexagonal faces. S-4-S*: only two edges in the linking face are shared with two hexagonal faces, respectively. S-6-L*: only two edges in the linking face are shared with one quadrangular and one hexagonal face, respectively. L-6-L*: all of the six edges in the linking face are shared with two pentagonal faces in each.

cage transformation that is beyond the scope of this study. For example, when a water molecule is inserted into the edge of a 5^{12} CC with $V = 20$, the cage changes to a $[5^{10}6^2]_1$ FSIC with $V = 21$ (Fig. 2B and Fig. S3B, ESI†; Table 4 and Table S2, ESI†).

Because we have identified all possible face-saturated cages in the two tested trajectories, we are able to analyze the linking relationships among them for each time frame. This is a routine analysis similar to (but with a higher level than) the calculation of the H-bond linking of bulk water. We consider two cages to be linked if they share at least one cage face. Such a structure is called the cage link, and such a cage face is called the linking face. Correspondingly, those cage faces that do not link other cages are called surface faces. Obviously, a cage may link many other cages *via* its linking faces and may also contact aqueous solution or gaseous methane directly *via* its surface faces. In this study, we count all cage links and distinguish standard hydrate structures among them according to the characteristic cage links of hydrates, including sI, sII, and sH (Table 5). Note that not all cage links between T cages represent the sI structure because the orientation between T cages must be considered, especially for the cage links with the pentagonal linking faces (T-5-T*). Additionally, the cage link of D-5-D belongs to both sII and sH. Because the standard $5^{12}6^8$ hydrate cage does not occur in either T2μs or T5μs, and other sH cage links are rare, we treat the D-5-D as an sII cage link. Based on the calculation of cage links, we therefore define hydrate crystallinity as the fraction of standard hydrate cage links to total cage links found in a guest–water system. Fig. 2D shows that at the end of T2μs, although the water molecules with a fraction of 0.95 have become cage makers, the total crystallinity (including sI, sII, and sH) accounts for only 0.16, far away from hydrate crystalline phases. In fact, these cage makers build an H-bond network containing 348 cages covering 90 types with a cage occupancy value of 0.862 (Fig. 2D). These results are unambiguous and provide quantitative evidence to demonstrate that the system reaches an amorphous phase of methane hydrate, taking the Fig. 1F in ref. 6 as a typical snapshot. "*It is a miscellaneous packing of various face-saturated cages, including complete, incomplete, empty, and water-filled cages. Certainly, the standard hydrate cages could also occur in the amorphous phase and even form small fragments of sI, sII, sH, and other hydrate structures*".[13] The amorphous phase is markedly different from the solution phase in many aspects (Table 6), mainly including the number of cages, the types of predominant cages, the occupancy of cages, the degree to which cages aggregate, and the hydrate crystallinity. Moreover, another inherent difference is that the water rings of the amorphous phase become more planar than that of the solution phase, which is clearly shown by the fraction of planar rings to total rings (λ) in the system water (Fig. 4). The case in T5μs is similar to that in T2μs but showing a little higher extent of crystallization (Fig. S3D and Table S3, ESI†). Additionally, we emphasize the fact that the amorphous phases in both T2μs and T5μs are far away from the perfect crystalline phase, cannot be recognized clearly by using other common order parameters, such as F_4 (Fig. 1 in ref. 6), λ (Fig. 4), and the fraction of cage water to total water (Fig. 2D).

Further, the dynamic feature of the amorphous hydrate is also evaluated. We had studied the dynamic profile of one 5^{12} cage and its liquid surroundings, and found that the self-diffusion coefficient of cage water can be up to 10 times smaller than that of the liquid water when the 5^{12} cage adsorbs 12 methane molecules.[15] Therefore, one can infer that the difference in water mobility between an amorphous hydrate consisting of many cages and its neighboring solution phase must be far larger than 10-fold, and thus the amorphous hydrate should be solid-like. In this work, the calculation of self-diffusion coefficient of system water (Fig. 5) proves this inference because the final mobility of water is 2–3 orders of magnitude smaller than the initial mobility observed in either T2μs or T5μs. These results are not only consistent with the identification of the solid-like amorphous phase by Vatamanu and Kusalik,[9] but also clearer than theirs because the corresponding difference in water mobility is only about 30 times in their results.

By using the cage linking relationship, the critical nucleation of hydrate can potentially be observed. We first define that a

Table 6 Average values of some parameters showing the structural difference between the solution phase and amorphous phase in T2μs[a]

Parameters	Solution phase	Amorphous phase
Number of FSIC	60.0(5)	98(2)
Number of CC	0.03(2)	249(7)
Number of hydrate cage	0.02(2)	118(1)
Occupancy for all cages	0.021(2)	0.861(3)
Fraction of cage water	0.255(2)	0.945(5)
Fraction of guest methane	0.003(0)	0.62(1)
Fraction of adsorbed methane	0.151(4)	0.28(1)
Top 10 cage types	$[5^2 6^3]_5$/0.213	5^{12}/0.232
	$[5^2 6^4]_5$/0.175	$4^1 5^{10} 6^2$/0.111
	$[4^1 5^2 6^3]_4$/0.083	$5^{12} 6^2$/0.090
	$[5^2 6^5]_5$/0.054	$5^{12} 6^3$/0.055
	$[5^4 6^3]_4$/0.035	$4^1 5^{10} 6^3$/0.042
	$[4^1 5^2 6^4]_4$/0.027	$[5^{10} 6^2]_1$/0.031
	$[6^5]_6$/0.027	$4^1 5^{10} 6^4$/0.028
	$[5^4 6^4]_4$/0.025	$5^{12} 6^4$/0.024
	$[4^2 5^2 6^2]_3$/0.020	$4^2 5^8 6^2$/0.020
	$[5^2 6^6]_5$/0.017	$4^2 5^8 6^4$/0.019
Fraction of planar ring	0.155(1)	0.815(9)
Links per cage	0.202(3)	5.09(7)
Total crystallinity	Trace	0.156(4)
sI crystallinity	0	0.067(1)
sII crystallinity	Trace	0.089(3)
sH crystallinity	0	0.0001(1)
Sampling region	0.4–0.6 μs	1.8–2.0 μs
Total CC numbers	52	498177
Total FSIC numbers	120086	195263
Total CC types	12	605
Total FSIC types	1339	2917

[a] The numbers after the slashes are the fractions of the cage number for each type to the total number of all cages. The numbers in parentheses are the standard errors of the mean estimated from 4 blocks with each covering 500 frames.

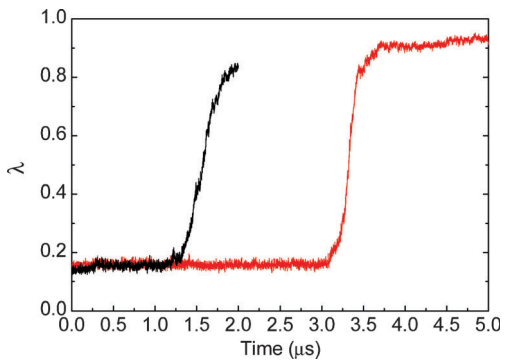

Fig. 4 Evolution of the fraction of planar rings to total rings in the system water, λ, for T2μs (black) and T5μs (red). The original definition of λ can be found elsewhere,[15] and only 4-, 5-, and 6-membered rings are counted. Note: primitive rings are used in the original definition but simple rings are used here. This change causes the λ value of pure water to decrease from the original 0.2 to the present 0.14. However, the change does not affect the λ value of hydrates, which is 1 in both cases since all rings in hydrate structures are planar.

cage cluster consists of at least two cages linked together and then define that a hydrate nucleus occurs as a special cage cluster composed of arbitrary numbers of FSICs plus at least one CC. Obviously, the hydrate nucleus occupies a space equal to the total volume of its member-cages and separates itself

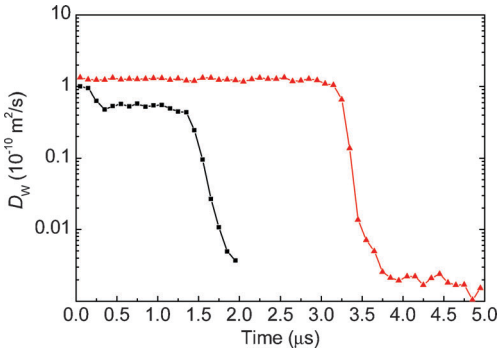

Fig. 5 Evolution of the self-diffusion coefficients of system water, D_W, for T2μs (black squares) and T5μs (red triangles). The trajectories were first divided into blocks with 100 ns in length. Then, the D_W of each block was calculated from the slope of the MSD curve between 5 and 25 ns.

from liquid and gas phases by all surface faces of cages. According to this definition, one can imagine that a hydrate nucleus could shrink and grow in size due to the fluctuation of its surroundings, and could change its cage composition and cage links to survive in the system. If the nucleus shrinks and disappears, the nucleation event will have failed. In contrast, if the nucleus grows to such a size that can withstand fluctuations, (*i.e.*, it may shrink a little but never totally disappear), it must have reached the critical size. Fig. 6A and B show the evolution of the size of a nucleus (*i.e.*, the number of member-cages, N); the curve may be divided into three stages—the induction stage in which many failed nucleation events randomly occur, the effective nucleation stage in which the critical nucleus occurs, and the growth stage in which steady increase of nucleus size begins. Obviously, the maximum value of N in the induction stage corresponds to the unreachable lower-limit of the critical nucleus because a nucleus with such a size is still insufficient to maintain its existence in solution. In contrast, the maximum value of N in the effective nucleation stage means the upper-limit of critical nucleus. Therefore, we estimate the critical size of hydrate nucleus to be $25 < N_c \leq 28$ in T2μs (Fig. 6B). Fig. 6D shows the upper-limit of the critical nucleus in T2μs, which is a cage cluster composed of 5 CCs and 23 FSICs (Table 7). As for the molecular composition, it contains 352 water and 16 single-occupied methane molecules. The cluster's maximum radius is 20.6 Å, and its specific surface area (*i.e.*, surface area/volume) is 0.66 Å^{-1} (Fig. 6C). The critical nucleus in T5μs is similar to this one with an estimation of critical size as $21 < N_c \leq 26$ (Fig. S5, ESI† and Table 7). Compared with other estimates of hydrate critical nuclei, the present result is reasonable. For example, the critical radius of CO_2 hydrate is estimated to be 14.5 Å using MD simulations[16] and that of methane hydrate is estimated to be 32 Å using classical nucleation theory.[17] At present, we still cannot identify the critical nucleus accurately but can constrain it to be within a narrow range using the FSICA. Additionally, because the shape of the critical nucleus may be rather irregular (taking Fig. 6D and Fig. S5D, ESI† as examples), its cage number, cage composition, cage link, and specific surface area are more important properties than its radius and molecular number usually considered.

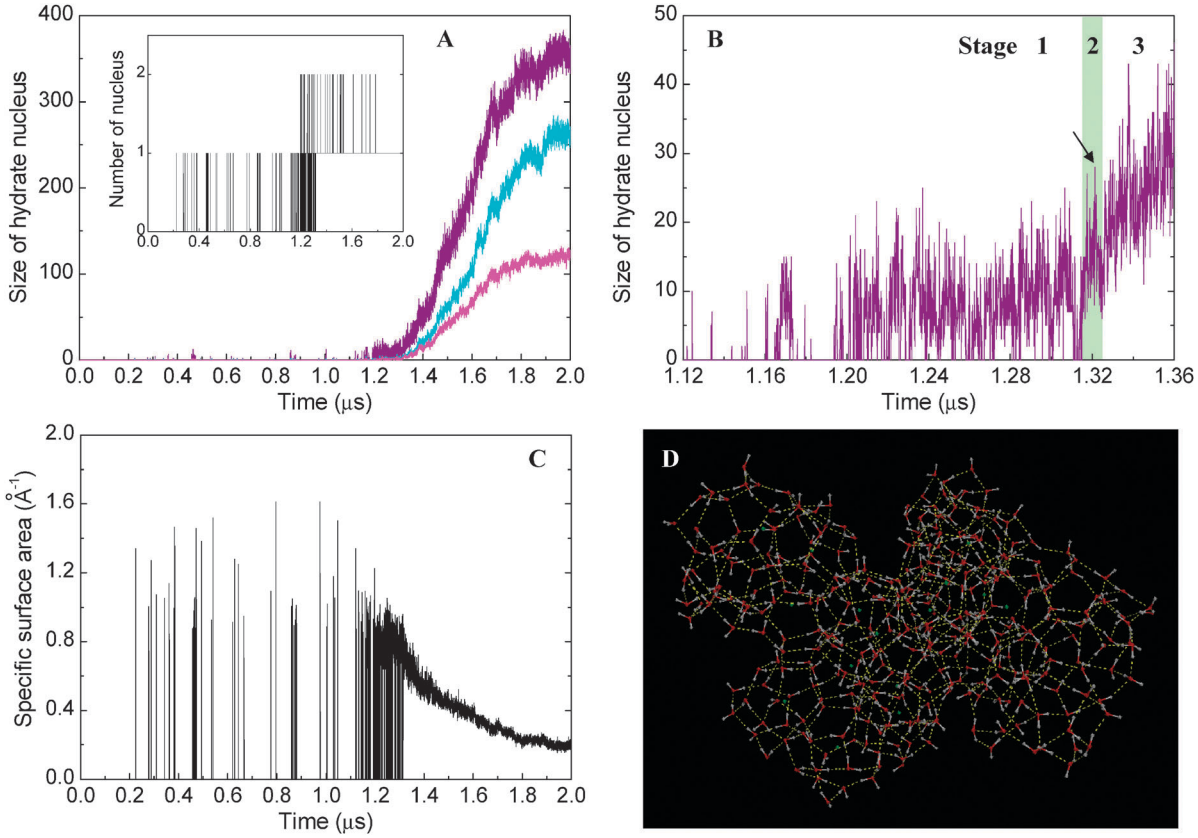

Fig. 6 (A) Evolution of the size of the hydrate nucleus. The coloured lines correspond to total cages (purple), CCs (cyan), and standard hydrate cages (magenta). The inset shows the number of nuclei, in which the value of 2 actually reflects the intermittent breakdown of bridging cages in the large nucleus and does not represent another independent nucleus. In the case of two nuclei occurring, only the larger one was considered. (B) The magnification of the purple line in (A). Stages 1–3 correspond to the induction, the effective nucleation, and the growth of a hydrate nucleus. Stage 2 (green region) begins from the last time the nucleus reached zero size, and extends to the last time the nucleus reached the minimum value after the start of the stage. (C) The change of the specific surface area of nucleus with time. (D) A snapshot of the upper-limit that the critical nucleus takes at the time indicated by the arrow in (B).

Table 7 Parameters of upper-limit critical nucleus for T2μs and T5μs

Parameters	T2μs	T5μs
Occurrence time/μs	1.3213	3.1265
Number of FSIC (with type)	23 (19)	21 (17)
Number of CC (with type)[a]	5 (3)	5 (2)
Number of hydrate cage (with type)	4 (2)	4 (1)
Occupancy for all cages	0.571	0.538
Specific surface area/Å^{-1}	0.664	0.651
Maximum radius/Å	20.6	22.8
Cage water	352	326
Guest methane	16	14
Adsorbed methane	35	42
Links per cage	1.964	2.269
Total crystallinity	0.091	0.051
sI crystallinity	0.055	0
sII crystallinity	0.036	0.051
sH crystallinity	0	0

[a] The CCs in T2μs include three 5^{12}, one $5^{12}6^2$, and one $5^{12}6^3$. The CCs in T5μs include four 5^{12} and one $5^{12}6^3$.

Recently, on the basis of classical MD simulations, a two-step mechanism was proposed by both Jacobsen *et al.*[7,8] and Vatamanu and Kusalik[9] to describe the nucleation and growth of hydrates. According to this mechanism, a metastable, solid-like, amorphous hydrate forms first, and subsequently transforms into crystalline hydrates, a step considered to be an annealing process. The two-step mechanism is actually consistent with our previously proposed cage adsorption hypothesis (CAH) for hydrate nucleation.[13] Although the second step is just a rough conjecture in the CAH, the first step is described with more substantial details in the CAH than in the two-step mechanism. The original statements of the CAH can be found elsewhere,[13] and here, we emphasise several main viewpoints which are further supported by the new evidence found in this work. First, face-saturated cages can form in methane solution spontaneously, especially for small, empty FSICs (Fig. 2B). We had not thought of this case when we initially proposed the CAH because we could not search empty cages at that time. Second, we previously showed that a strong attractive force exists between a water cage and an aqueous methane molecule.[13] Although the cage we studied was a 5^{12} prototype, we infer that other cages can also adsorb methane. Thus, the CAH suggests that a cage adsorbing methane molecules is a dominant process causing methane to preferentially aggregate toward hydrate formation rather than the formation of methane bubbles. In this work, we indeed observe some

phenomena attributable to the cage–methane adsorption interaction. For example, Fig. 1H shows the $[5^26^3]_5$ cage adsorbing four methane molecules and the inset of Fig. 2C shows that about 50% of the dissolved methane is adsorbed primarily by FSICs. Third, the CAH suggests that hydrate formation should reach an amorphous phase in which various cages link together miscellaneously. Here, T2μs and T5μs provide unambiguous evidence (Fig. 2 and 6; Fig. S3 and S5, ESI†; Tables 3, 4, 6 and 7; Tables S1–S3, ESI†) to support this viewpoint. Moreover, the amorphous phase is undoubtedly solid-like in character (Fig. 5), and its H-bond network is mainly composed of planar water rings (Fig. 4). Additionally, we emphasize that cage formation is an independent behavior of water molecules; the presence of methane is not a necessary condition for cage formation, but it is a favorable one. The ordered arrangement of methane surrounding a cage is the consequence of the cage formation—unless the cage acts as a bridging cage between other existing cages. In T2μs and T5μs, we have observed that the numbers of adsorbed methane for many FSICs can be zero, and those for CCs can take a value as small as one. These phenomena show that when a cage forms, few methane molecules need be present in its surrounding fluid. Certainly, the more methane molecules that are found in its vicinity, the larger the chance of forming the cage.

At last, we discuss the definition of cage faces and cage holes in detail. It is crucial to differentiate them in the FSICA because those cages including holes, so called ICs, are screened out in this study. In fact, whether a water ring permits a guest to pass through it does depend on the size and speed of the guest molecule as well as the shape of the ring. Because we mainly consider the static structure of the system, we do not consider the factor of speed. In this work, we choose the 6-membered ring as the largest cage face based on three aspects. First, the radius of a regular 6-membered ring (= 2.82 Å, the average length of H-bond) is equivalent to the minimum distance between an aqueous methane molecule and a water oxygen atom (= 2.88 Å) recorded in the studied system. This means that a methane molecule is very unlikely to traverse the ring, and that it is almost impossible if the ring is irregular and nonplanar. Second, the computational cost for searching up to a 6-membered ring is affordable, but there is a major increase in cost if larger rings are considered. Third, it is the smallest choice to ensure that all standard hydrate cages can be found. Recently, Vatamanu and Kusalik reported several cage structures including a 7-membered ring.[9] It is true that nonplanar 7-membered rings can also keep methane from traversing it, as can nonplanar 8-, and larger rings. When a 5^{12} cage breaks one of its edges, it becomes a $[5^{10}8^1]_2$. To judge whether the 8-membered ring being a cage face or a cage hole is a difficult problem. From a practical view, using the ring size rather than the ring shape to resolve the problem is simple and efficient. Interestingly, although we define the 6-membered ring as the maximum cage face and output only FSICs, the 7-, and larger membered rings are not fully excluded from the structure of amorphous hydrate. They occur in the complex CCs and FSICs that can be looked as the combination of several ICs by sharing the 7-, and larger membered holes (Fig. 1B and Fig. S6–S8, ESI†). These complex cages indeed can be divided into independent simpler cages by choosing a larger maximum ring size than 6 to distinguish cage faces and holes, but it requires unaffordable increase in computational costs and the selection of 7-membered ring cannot approach this aim without considering the existence of 8- and larger membered holes (Fig. S7 and S8, ESI†). Even if this aim is achieved, all of the newly divided cages are not standard hydrate cages and they cannot qualitatively affect the identification of amorphous hydrates through our crystallinity parameter. Additionally, for guests being smaller (or larger) than methane in size, such as H_2 (or oil), the cage face can certainly be redefined according to the purpose of the study.[11] However, we do not suggest to define the 5-membered ring as the maximum cage face when studying the H_2–water system because the jump diffusion of H_2 among cages bearing 6-membered rings deserves to be investigated in detail. Finally, although we do not consider ICs in this study, it does not mean ICs are unimportant. Some ICs with large values of cage-like degree are expected to build the interface between the face-saturated cage cluster and the aqueous solution, and must play their particular roles during the nucleation and growth of hydrates.

4. Conclusions

Because the FSICA can identify all face-saturated cage compositions in the H-bond network of water–guest (or pure water) systems, it is a powerful tool in hydrate research based on computer simulations. By applying the FSICA, thousands of cage types were found in previously reported MD trajectories.[6] The evolution of hydrate crystallinity quantitatively demonstrates that the system reaches an amorphous phase (with average crystallinity values of 0.156 in T2μs and 0.212 in T5μs), being still a long way from the crystalline phases. Moreover, the critical nuclei of hydrate are also bounded within reasonable limits, *i.e.*, 25–28 cages in T2μs and 21–26 cages in T5μs. In the future, the structural transition from the amorphous to crystalline phase will be the next challenge. More in-depth studies on critical nuclei are also expected.

Acknowledgements

We thank M. R. Walsh, C. A. Koh, E. D. Sloan, A. K. Sum, and D. T. Wu for providing the raw data for the two trajectories (T2μs and T5μs), and for helpful discussion with them. We also appreciate Prof. P. M. Rodger for his help during this work. This work is supported by the National Basic Research Program of China (Program 973, Grant No. 2009CB219503).

References

1 E. D. Sloan and C. A. Koh, *Clathrate Hydrates of Natural Gases*, CRC Press, Boca Raton, 3rd edn, 2008.
2 E. D. Sloan, *Nature*, 2003, **426**, 353.
3 M. Matsumoto, S. Saito and I. Ohmine, *Nature*, 2002, **416**, 409.
4 C. Moon, P. C. Taylor and P. M. Rodger, *J. Am. Chem. Soc.*, 2003, **125**, 4706.
5 R. W. Hawtin, D. Quigley and P. M. Rodger, *Phys. Chem. Chem. Phys.*, 2008, **10**, 4853.
6 M. R. Walsh, C. A. Koh, E. D. Sloan, A. K. Sum and D. T. Wu, *Science*, 2009, **326**, 1095.

7 L. C. Jacobson, W. Hujo and V. Molinero, *J. Am. Chem. Soc.*, 2010, **132**, 11806.
8 L. C. Jacobson, W. Hujo and V. Molinero, *J. Phys. Chem. B*, 2010, **114**, 13796.
9 J. Vatamanu and P. G. Kusalik, *Phys. Chem. Chem. Phys.*, 2010, **12**, 15065.
10 L. C. Jacobson, W. Hujo and V. Molinero, *J. Phys. Chem. B*, 2009, **113**, 10298.
11 G.-J. Guo, Y.-G. Zhang, M. Li and C.-H. Wu, *J. Chem. Phys.*, 2008, **128**, 194504.
12 D. Laage and J. T. Hynes, *Science*, 2006, **311**, 832.
13 G.-J. Guo, M. Li, Y.-G. Zhang and C.-H. Wu, *Phys. Chem. Chem. Phys.*, 2009, **11**, 10427.
14 M. Matsumoto, A. Baba and I. Ohmine, *J. Chem. Phys.*, 2007, **127**, 134504.
15 G.-J. Guo, Y.-G. Zhang and H. Liu, *J. Phys. Chem. C*, 2007, **111**, 2595.
16 R. Radhakrishnan and B. L. Trout, *J. Chem. Phys.*, 2002, **117**, 1786.
17 M. A. Larson and J. Garside, *Chem. Eng. Sci.*, 1986, **41**, 1285.