9-14-2016

Nonlinear Machine Learning and Design of Reconfigurable Digital Colloids

Andrew W. Long
University of Illinois at Urbana-Champaign

Carolyn L. Phillips
Neurensic, Inc.

Eric Jankowski
Boise State University

Andrew L. Ferguson
University of Illinois at Urbana-Champaign

Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Nonlinear machine learning and design of reconfigurable digital colloids†

Andrew W. Long,[a] Carolyn L. Phillips,[b] Eric Jankowksi,[c] and Andrew L. Ferguson[a,d,*]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Digital colloids, a cluster of freely rotating “halo" particles tethered to the surface of a central particle, were recently proposed as ultra-high density memory elements for information storage. Rational design of these digital colloids for memory storage applications requires a quantitative understanding of the thermodynamic and kinetic stability of the configurational states within which information is stored. We apply nonlinear machine learning to Brownian dynamics simulations of these digital colloids to extract the low-dimensional intrinsic manifold governing digital colloid morphology, thermodynamics, and kinetics. By modulating the relative size ratio between halo particles and central particles, we investigate the size-dependent configurational stability and transition kinetics for the 2-state tetrahedral (N=4) and 30-state octahedral (N=6) digital colloids. We demonstrate the use of this framework to guide the rational design of a memory storage element to hold a block of text that trades off the competing design criteria of memory addressability and volatility.

[a] *Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, 61801, USA*
[b] *Neurensic, Inc., Chicago, IL, 60604, USA*
[c] *Micron School of Materials Science and Engineering, Boise State University, Boise, ID, 83725, USA*
[d] *Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, 61801, USA*
[*] *E-mail: alf@illinois.edu; Phone: (217) 300-2354; Fax: (217) 333-2736*

† Electronic Supplementary Information (ESI) available: Two movies highlighting the structure and topography of the low-dimensional intrinsic manifolds for the $N=4$ and $N=6$ digital colloids, a figure showing the transition network for the $N=6$ digital colloid, and a movie tracking a $N=6$ transition event over the intrinsic manifold. See DOI: 10.1039/b000000x/

1 Introduction

Digital colloids[1], reconfigurable clusters of lock-and-key colloidal particles[2], have recently been proposed as a novel soft matter-based substrate for high-density information storage. A digital colloid comprises freely-rotating but tethered "halo" particles that are bound to the surface of a central particle (Figure 1). Information can be stored within distinguishable configurations of the halo particles around the central particle. These structures have been experimentally synthesized via depletion mediated binding of dimpled halo particles to the surface of a central spherical particle[1,2], as well as through programmable DNA interactions[3]. These halo particles may be distinguished through a variety of labeling techniques[4], such as functionalization with complementary Förster Resonant Energy Transfer (FRET) pairs[5] or through binding of short oligonucleotides[1], thereby enabling the unique identification of distinct metastable halo particle configurations that mediate information storage and retrieval.

Storing information in micron-scale colloids has the potential to advance computing in unconventional environments. A prime example of which is DNA-based computing in solution, wherein arbitrary digital circuitry can be mapped onto reaction networks of DNA strands[6]. One practical limitation of DNA-based computing is the storage of information in the concentrations of DNA strands, which are essentially uniform throughout the computing volume. Digital colloids have potential as compartmentalized, high-density storage elements that can diffuse throughout a computing volume, and which can in principle be reconfigured during computations. In order to realize the potential of digital colloids as unconventional computing elements, we require a fundamental understanding of the configurational transitions between storage states.

The number of distinguishable halo particle configurations in the digital colloid ground state determines the number of unique "bit states" of a cluster, and therefore its information storage capacity. For example, the halo particles of the N=4 digital colloid exist in a tetrahedral ground state, meaning – after the elimination of trivial rotational symmetries – it possesses exactly two distinguishable arrangements of the four halo particles (Figure 1c). These two available states mean that the N=4 colloid is a binary storage element that can encode exactly one bit of information. In general, the number of storable states is specified by the $N!$ halo particle permutations divided by the symmetry operator for the ground state configurational arrangement, Υ_N. The symmetry operator for an N-particle colloid is determined by the spherical code solution for the arrangement of the halo particles around the central particle[1,7–10]. This arrangement is the solution to the Tammes problem[11] – a special case of the Thomson problem[12] – corresponding to the configuration of N halo particles that maximizes the minimum distance (or equivalently minimum angle) between the halo particle centers[7–10,13]. This configuration determines both the point group of the packing structure and the multiplicity of the symmetry operator. Putatively optimal spherical code solutions up to N=130 are reported in Ref. [14]. The spherical code solution for N=4 is a tetrahedron (Figure 1a) and for N=6 an octahedron (Figure 1b) with corresponding point groups, rotational symmetries, and number of storable states listed in Figure 1d. The number of bit states that can be written in a digital colloid is given by the base-2 logarithm of the number of storable states, $\log_2(N!/\Upsilon_N)$. The information storage capacity of digital colloids increases as $O(\log_2(N!))$, making them ideal for high-density information storage[1]. For example, a single N=6 octahedral cluster is capable of storing 4.9 bits (0.613 bytes), and a N=12 icosahedral cluster 22.9 bits (2.87 bytes).

A key feature of these digital colloids is that the halo particles, although strongly tethered to the central particle, are free to move over the surface of the central particle. Defining the diameter ratio between the central and halo particles as $\Lambda = d_{\text{central}}/d_{\text{halo}}$, the spherical code solution imposes a minimum value Λ_{SC} for a given system of N halo particles, below which the N particles cannot bind to the surface without overlapping. For Λ above this value but below some threshold ratio, $\Lambda_{\text{SC}} \leq \Lambda < \Lambda_{\text{T}}$, each halo particle is confined within a cage composed of its neighbors, and the digital colloid is locked into a single bit state. This threshold ratio, Λ_{T}, is given by the minimum value of Λ needed for collective halo particle rearrangements to access the transition state, and has been previously computed for N=4-12[1]. At $\Lambda \geq \Lambda_{\text{T}}$ there is sufficient free volume for the halo particles to transition between bit states and the digital colloid becomes unlocked. By actuating halo particle transitions between bit states, information can be written. By observing the halo particle bit state, information can be read.

Above the unlocking transition, the digital colloid is no longer confined to oscillate around the spherical code configuration, and can explore a rich and complex morphological landscape. The design of information storage elements with tailored free energy barriers sufficiently low to actuate the halo particle configuration (i.e., a low write free energy) but sufficiently high to prevent thermally-activated spontaneous bit state flips (i.e., low memory volatility) requires an accounting of the thermally accessible configurational state space, transition state free energy barriers, and mean first passage times as a function of N and Λ.

Furthermore, designing reliable digital colloids requires an understanding of how sensitive bit transitions are to the manufacturing tolerances of their constituent parts.

The configurational state of a digital colloid exists in a $2N$-dimensional phase space defined by the azimuthal and polar coordinates of the N halo particles on the surface of the central sphere. Excluded volume and any other interactions between the halo particles means that their motions and locations are correlated, and these cooperative couplings are expected to define an emergent low-dimensional subspace to which the digital colloid dynamics are effectively restrained[15]. This latent low-dimensional manifold supports the free energy surface containing the thermally-accessible system configurations, stable and metastable states, and transition pathways between them. Determining this landscape provides quantitative understanding of the system morphology, thermodynamics, and kinetics.

Dimensionality reduction techniques have demonstrated good success in recovering these low-dimensional manifolds for both single molecule[16–18] and many-body self-assembling[15,19,20] systems. Recently, we (A.W.L. and A.L.F.) developed an extension of the diffusion map nonlinear dimensionality reduction technique[21–23] for many-body colloidal systems to infer low-dimensional assembly landscapes from both simulation[15] and experimental particle tracking data[19]. Building upon these foundations, in this work we combine Brownian dynamics simulations, nonlinear machine learning, and first passage time analysis to quantify the stable and transition state morphologies and determine the transition mechanisms, free energy barriers, and transition rates for bit state switching for N=4 and N=6 digital colloids. We use this new quantitative understanding to computationally design addressable digital colloids with tailored kinetic stability and actuation free energies, providing a step towards the engineering of high-density digital colloid information storage media.

2 Materials & Methods

We study the morphology, thermodynamics, and kinetics of digital colloids comprising N=4 and N=6 halo particles using Brownian dynamics simulations, nonlinear dimensionality reduction, and first passage time analysis.

2.1 Brownian Dynamics simulations of digital colloids

Following the protocol previously detailed in Refs. [1,24], we performed Brownian dynamics simulations of N=4 and N=6 digital colloids using the HOOMD-blue GPU-based molecular simulation package (`http://glotzerlab.engin.umich.edu/hoomd-blue/`)[25–27]. The spherical halo particles are constrained to glide over the surface of the spherical central particle using a radial restoring force perpendicular to the free particle direction of motion. This restoring force provides a rigid constraint binding the halo particles to the central particle surface[1], implemented via the HOOMD `constrain.sphere` module (`http://hoomd-blue.readthedocs.io/en/stable/module-md-constrain.html`). The diameter of the central particle d_{central} regulates the geometric confinement of these halo particles, which is defined by the ratio of central to halo particle diameters, $\Lambda = d_{\text{central}}/d_{\text{halo}}$. The central particle is treated as static and inert, while halo particles i and j interact through a surface-shifted Weeks-Chandler-Anderson (WCA) pair potential to account for excluded volume interactions[28],

$$u(r_{ij}) = \begin{cases} 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}-\Delta}\right)^{12} - \left(\frac{\sigma}{r_{ij}-\Delta}\right)^{6}\right] + \varepsilon, & r_{ij} < 2^{(1/6)}\sigma + \Delta \\ 0, & r_{ij} \geq 2^{(1/6)}\sigma + \Delta \end{cases} \quad (1)$$

where r_{ij} is the center of mass separation between the halo particles, ε is the interaction strength, σ is the size parameter, and the parameter $\Delta = 2\sigma$ shifts the WCA potential to an effective surface of the halo particles. We perform our simulations in a dimensionless gauge such that energy is measured in units of ε and distance in units of σ. The reduced temperature is defined as $T^* = k_B T/\varepsilon$, and reduced time as $t^* = t/\sqrt{m\sigma^2/\varepsilon}$ where m is the mass of a halo particle.

The surface-shifted WCA potential is a convenient computational approximation for particles possessing hard core interactions[29,30], and can be mapped to a hard sphere fluid using a perturbative generalization of the Rowlinson scheme developed by Barker and Henderson[31–33]. Employing Eqn. 12 in Ref. [31], we determine an effective hard sphere diameter of the halo particles of d_{halo} = 3.0785σ. The central particle diameter is defined as $d_{\text{central}} = 2d_{\text{central-halo}} - d_{\text{halo}}$, where $d_{\text{central-halo}}$ is the constrained center of mass distance between the central and halo particles. This mapping of our WCA halo particles to effective hard spheres of diameter $d_{\text{halo}} = 3.0785\sigma$ enables us to calculate the spherical code diameter ratio Λ_{SC} and threshold diameter ratio Λ_{T} based on purely geometric considerations[1,7]. These hard sphere values do not hold precisely for our soft particles, but the steepness of the WCA potential and moderate temperatures employed in our calculations mean that they do behave as approximate hard spheres. Specifically, we observe no pair of halo particles i and j to approach closer than $r_{ij} = 2.95\sigma$ corresponding to a pairwise interaction energy $u(r_{ij}) = 2.96\varepsilon$. The values of Λ_{SC} and Λ_{T} reported for each of the digital colloid architectures should therefore be interpreted as approximations based on the effective hard sphere mappings.

Simulations were performed at a reduced temperature of T^*=0.1 and the Brownian dynamics (overdamped Langevin dynamics) equations of motion numerically integrated using velocity verlet in the NVE ensemble and a Langevin thermostat[34] with a time step of dt^*=10^{-3}[35]. Each simulation was initialized with $\Lambda_0 >> \Lambda_T > \Lambda_{SC}$ and the halo particles randomly distributed and free to move over unencumbered over the central particle surface. The central particle was then gradually shrunk to the desired value of Λ over the course of 1.1×10^5 time steps to allow the system to equilibrate before conducting a 10^9 time step production run.

To make contact between our dimensionless simulations and real units, we adopt a characteristic interaction strength of $\varepsilon = 10\ k_BT$ at 298 K between micron-sized halo particles with $\sigma = 0.915\ \mu$m ($d_{\text{halo}} = 2.82\ \mu$m) fabricated from 3-methacryloxypropyl trimethoxysilane (TPM) with density $\rho = 1.228$ g/cm^3[36]. Under this choice the reduced temperature $T^* = 0.1$ maps to $T = 298$ K, the reduced time step $dt^* = 10^{-3}$ to $dt = 0.54\ \mu$s, and the reduced simulation time $t^* = 10^9 \times 10^{-3} = 10^6$ to $t = 540$ s.

2.2 Digital colloid dimensionality reduction

Simulations of the digital colloid dynamics provide all particle coordinates as a function of time, furnishing, in principle, all of the morphological, thermodynamic, and kinetic details governing transitions between different configurations of the distinguishable halo particles defining the information storage bit states. In practice, extracting this information from the trajectories is challenging due to the high dimensionality of the system: a particular configuration of N halo particles rolling around the surface of a central particle exists as a point in a $2N$-dimensional phase space defined by the surface locations of each halo particle. For sufficiently small values of Λ the accessible phase space volume is, however, drastically reduced by excluded volume interactions that give rise to cooperative halo particle rearrangements[1]. These cooperative couplings are expected to yield a separation of time scales, wherein local caged oscillations within a stable bit state give rise to high-frequency motions, and transitions between bit states are rare events with slow characteristic time scales. We anticipate that these cooperative transition pathways correspond to a small number of slow collective modes of the halo particle dynamics[15,19,37]. Extracting these slow modes from the simulation trajectory reveals the multi-body transition pathways governing the digital colloid dynamics, and also presents kinetically meaningful collective coordinates in which to construct low-dimensional free energy surface mapping out the accessible morphologies, stable states, and dynamical transition pathways[15,16,18,19,38–41].

Nonlinear dimensionality reduction provides a means to determine the low-dimensional hypersurface – the so-called "intrinsic manifold" – within the high-dimensional phase space to which the multi-body digital colloid dynamics are effectively restrained[15,19]. Geometrically, the intrinsic manifold is the low-dimensional phase space volume parameterized by a small number of collective variables that contains the stable colloidal bit states and the dynamic transition pathways between them[15,19,38,39]. Temporally, the existence of the intrinsic manifold can be viewed through the Mori-Zwanzig formalism as the emergence of a small number of slowly-evolving collective coordinates to which the remaining system degrees of freedom are slaved as effective noise[15,19,23,38,39,42]. Compared to linear dimensionality reduction techniques such as principal components analysis (PCA)[43] or multidimensional scaling (MDS)[44], nonlinear methods provide greater flexibility in extracting collective modes that are nonlinear combinations of the particle degrees of freedom[16,39,45], enabling the recovery of more general and complex intrinsic manifolds.

Path-based approaches such as transition path sampling[46], finite temperature string[47], and the nudged elastic band[48] seek to compute high-probability (low-free energy) reactive pathways between two metastable system configurations. These paths contain information on the reaction coordinate, transition state ensemble, isocommittor surfaces, and reaction mechanism for the structural transition between the metastable basins[49]. In contrast, nonlinear dimensionality reduction approaches seek to discover within high-dimensional simulation trajectories – including ensembles of simulations performed under different conditions[18] or biased simulations to which artificial restraining potentials are applied to enhance sampling[50] – low-dimensional projections of the configurational phase space containing the metastable states and transition pathways between them. These reduced dimensional maps are valuable in identifying the metastable states, dynamical pathways and transition mechanisms, and also furnish good order parameters describing the emergent collective motions governing the long-time dynamics of the system[39]. Accordingly, the aims and utility of the two classes of methods are slightly different. Path methods are preferred where there exists an unambiguous definition of the reactant and product states and the objective is to obtain a precise characterization of the transition path ensemble, transition mechanisms, or reaction rates between these two basins[49]. Nonlinear dimensionality reduction methods may be preferred where good order parameters distinguishing the metastable states are sought

or it is the objective to obtain a global overview of the available metastable states, their mutual connectivity, and slow collective structural rearrangements. Of course only those metastable states within the simulation trajectories to which the dimensionality reduction methodology is applied can be resolved, and although accelerated sampling techniques can facilitate barrier crossing, path-based methods are typically more adept at surmounting high free energy barriers and may be more efficient at resolving transition pathways within rugged free energy surfaces. Finally, nonlinear dimensionality reduction may be profitably combined with path sampling as a preprocessing step to identify metastable states, deliver good collective variables with which to distinguish them, and determine initial transition paths with which to initialize path sampling[38,39].

We (A.W.L. and A.L.F.) recently reported a novel approach extending the application of the nonlinear diffusion map technique to multi-body systems to extract aggregate morphology and assembly mechanisms from computer simulations of patchy colloid self-assembly[15] and experimental particle tracking data of Janus colloid self-assembly[19]. In this work, we apply this technique to Brownian dynamics simulations of digital colloids to extract the intrinsic manifold governing digital colloid morphology and transitions. We present a complete description of the many-body diffusion map in Ref. [15], but we briefly detail the approach below, including system specific adaptations for the digital colloid system.

2.2.1 Microstate pairwise distances

The metastable bit states of the digital colloids are morphologically identical in that they correspond to the same configurational arrangement of halo particles around the central sphere[1]. Information storage is only possible insofar as all (or some) of the halo particles are distinguishable, thereby permitting unambiguous discrimination between all (or some) of the bit state configurations. In this work we consider all N halo particles to be distinguishable such that there are $N!$ relabelings cluster that are reduced by the rotational symmetry of the bit state Υ_N to yield an information storage capacity of $\log_2(N!/\Upsilon_N)$ bits.

Our nonlinear dimensionality reduction algorithm requires as input a measure of similarity distances between different digital colloid configurations observed over the course of the simulation. When we measure the distance between different halo particle configurations, we assume the halo particles are indistinguishable, but retain the halo particle label information for *a posteriori* identification of the distinguishable bits states within the low-dimensional projection. By separating morphology and labeling when sampling the configurational phase space we achieve an important practical advantage. Since each of the $N!$ bit states are morphologically identical, treating the halo particles as indistinguishable allows for comprehensive sampling of the identical metastable basin of all bit states within relatively short simulation trajectories. In contrast, treating halo particles as distinguishable would result in the emergence of $(N!/\Upsilon_N)$ identical metastable free energy wells within the intrinsic manifold, many of which would be incompletely sampled, and some never sampled at all. For small digital colloids containing few states, this is unlikely to present difficulties, but for larger colloids containing hundreds of thousands of states, sampling the intrinsic manifold without this simplification presents a practical challenge.

For digital colloids comprising N halo particles we compile a library $\mathscr{L}(N,\{\Lambda\})$ of all digital colloid microstates observed in our simulation trajectories over a range of central to halo particle diameter ratios Λ. We compute χ the N-by-N pairwise distance matrix between all pairs of halo particles in a particular digital colloid microstate in $\mathscr{L}(N,\{\Lambda\})$. We normalize pairwise distances by the minimum distance between vertices in the idealized regular polyhedron with equispaced halo particle vertices for a given value of Λ. For a central to halo particle distance $d_{\text{central-halo}} = d_{\text{halo}}(\Lambda+1)/2$, the edge length of the idealized tetrahedron in the $N=4$ system is $d'_{\text{halo-halo}} = \sqrt{8/3}d_{\text{central-halo}}$, and that of the idealized octahedron in the $N=6$ system is $d'_{\text{halo-halo}} = \sqrt{2}d_{\text{central-halo}}$. In each case, we define the normalized pairwise distance matrix $\chi' = \chi/d'_{\text{halo-halo}}$. By normalizing with respect to the ideal polyhedral edge length, we enable direct comparisons between digital colloid microstates across different Λ values.

Using the χ' computed for each digital colloid microstate in $\mathscr{L}(N,\{\Lambda\})$, we can define a fast, rotationally invariant, graph-based measure of structural similarity between pairs of digital colloid microstates i and j,

$$\mathbf{d}_{ij} = \min_{\mathbf{H}} ||\chi'_i - \mathbf{H}^T\chi'_j\mathbf{H}||, \tag{2}$$

where $\mathbf{H}$ is a permutation matrix rearranging the rows and columns of χ'_j. That is, our measure is the minimum of the permuted entrywise $L_{1,1}$ norm between halo particle distance matrices. Minimizing over permutations is equivalent to minimizing over relabelings of halo particles. For the $N=4$ and $N=6$ systems considered in this work, the number of permutations is quite small, $N! = 24$ and $N! = 720$, permitting us to determine the minimum permuted distance by brute force enumeration. Exhaustive enumeration will become intractable for sufficiently large systems, and it will be necessary to resort to approximate graph matching approaches such as those we have previ-

ously employed in Ref. [15].

2.2.2 Diffusion maps

Given our measure of structural similarity between all pairs of digital colloid microstates in $\mathscr{L}(N,\{\Lambda\})$, we apply the diffusion map nonlinear dimensionality reduction technique[21–23,39,51] to extract the collective order parameters governing structural transitions between digital colloid morphologies. These collective coordinates are associated with the slow modes governing the long-time structural evolution of the system to which the remaining motions are effectively slaved. The diffusion map approach was first introduced by Coifmann and coworkers[21–23,51], and has been used to determine the underlying pathways and mechanisms underpinning both molecular folding[18,40,45] and colloidal assembly[15,19,20].

In the diffusion map approach, a random walk is constructed over the high-dimensional data with transition probabilities based on structural similarity. Then a spectral analysis of the random walk is performed to identify the dimensionality and collective coordinates of a low-dimensional intrinsic manifold to which the system dynamics are effectively restrained[15,19,38,39]. The approach proceeds by forming the kernel matrix $\mathbf{A}$ by convoluting our pairwise distance matrix with a Gaussian kernel, $\mathbf{A}_{ij} = \exp\left(-\mathbf{d}_{ij}^2/2\varepsilon\right)$, where ε is a soft-thresholding bandwidth that restricts transitions between structures to those within an ε-neighborhood in the high dimensional space. This procedure serves to model structural transitions between microstates as a diffusion process over the discrete point cloud of digital colloid microstates in the high-dimensional space[21]. The value of ε is specified for each system using an automated procedure detailed in Ref. [52].

We next form the right-stochastic Markov matrix $\mathbf{M} = \mathbf{D}^{-1}\mathbf{A}$, where $\mathbf{D}$ is the diagonal matrix $\mathbf{D}_{ii} = \sum_k \mathbf{A}_{ik}$. The element M_{ij} can be interpreted as the transition probability from microstate i to microstate j, and therefore $\mathbf{M}$ describes a random walk over the data[21,22]. $\mathbf{M}$ is positive-semidefinite, right stochastic, and adjoint to a symmetric matrix $\mathbf{M}_S = \mathbf{D}^{\frac{1}{2}}\mathbf{M}\mathbf{D}^{-\frac{1}{2}}$, so possesses an orthogonal set of real right eigenvectors $\{\vec{\Psi}_i\}$ and associated eigenvalues $\{0 \leqslant \lambda_i \leqslant 1\}$ with the trivial top pair $\vec{\Psi}_1 = \vec{1}$ and $\lambda_1 = 1$[38]. These eigenvectors are discrete approximations to the eigenfunctions of the Fokker-Planck equation describing the time evolution of the probability density function over the data[21,22], with eigenvalues corresponding to the frequency of these collective harmonic modes. Large eigenvalues correspond to the slow (e.g. long time) relaxation modes of the Markov process, and small eigenvalues correspond to fast modes. A gap in the eigenvalue spectrum corresponds to a separation of relaxation time scales, with collective modes above the gap corresponding to the slow subspace modes governing long time structural evolution to which the remaining fast modes are restrained[23,38,42]. We employ the L-method of Salvador and Chan to systematically identify the location of this spectral gap[15,18,53].

Identifying a gap in the eigenvalue spectrum after $\lambda_{(k+1)}$ prompts construction of a low-dimensional landscape in the top k nontrivial eigenvectors furnished by the diffusion map $\{\vec{\Psi}_i\}_{i=2}^{k+1}$. The low-dimensional projection of the i^{th} microstate in our library $\mathscr{L}(N,\{\Lambda\})$ is specified by the "diffusion mapping" into the i^{th} component of these top k eigenvectors,

$$\text{microstate}_i \mapsto \left(\vec{\Psi}_2(i), \vec{\Psi}_3(i), ..., \vec{\Psi}_{k+1}(i)\right). \qquad (3)$$

This projection defines the nonlinear dimensionality reduction from the $2N$-dimensional phase space of halo particle angular coordinates to the k-dimensional intrinsic manifold.

One of the primary limitations of the diffusion map, and nonlinear dimensionality reduction methods in general – with the exception of autoencoders[54] – is that the technique does not furnish a mapping for these collective variables to the input features of the original high dimensional data. While automated techniques exist to sieve and approximate these top eigenvectors from a candidate feature pool[55,56], these algorithmic features can be so complex that attributing physical characteristics to these eigenvectors can still be exceedingly difficult. Further, there is no guarantee that a concise physical mapping exists for the nonlinear order parameters describing these many-body systems[15,39]. To aid in physical interpretation, we correlate the diffusion map variables by physical "bridge" variables such as the polyhedral volume or the distance of a configuration from an idealized polyhedron.

"Diffusion distances" in the high-dimensional space – defined precisely in Refs. [21,51] – may be colloquially interpreted as the ease with which two microstates may interconvert, and are approximated by Euclidean distances in the low-dimensional manifold. This property of diffusion map embeddings provides a direct kinetic interpretation of distances between microstates in the low-dimensional embedding, which holds under the mild assumptions that (i) the dynamics of the system may be described by a diffusion process and (ii) that our measure of pairwise similarity effectively captures short time diffusive motions of the system[15,21,38,39,51]. Regarding the first condition, we expect that the coupled motions of the halo particles and cooperative motion required for transition events between bit states will lead to the existence of a small set of slow collective modes describing the long time dynamical evolution.

This hypothesis is supported *post-hoc* by the emergence of a spectral gap in $\mathbf{M}$ and the small number of modes existing above this gap. This separation of time scales into slow and fast collective modes suggests that the dynamics of this digital colloid system can be modeled under the Mori-Zwanzig formalism as a set of coupled stochastic differential equations in the slow collective modes to which the remaining fast modes couple as noise[42,57]. Accordingly, the system dynamics are well approximated as a diffusion process. Regarding the second condition, we have shown in Refs. [15,19] that our graph-based distance metric for many-body systems provides a good measure for structural proximity over short time scales, and therefore captures the short time diffusive motions.

2.3 Free energy surfaces

We construct free energy landscapes over the k-dimensional intrinsic landscape defined by the diffusion map embedding (Eqn. 3) for each digital colloid. The free energy landscape is constructed by collecting histograms over all microstates collected over the course of the simulation projected into the low-dimensional projection, and employing the statistical mechanical relationship[58,59],

$$F^*(\Lambda,\vec{\xi})/T^* = -\ln\hat{P}(\Lambda,\vec{\xi}) + C(\Lambda), \tag{4}$$

where Λ is the diameter ratio, $T^* = k_BT/\varepsilon$ is the reduced temperature at which we conduct our simulations, $\vec{\xi}$ is a k-dimensional vector specifying a point on the k-dimensional intrinsic manifold defined by the diffusion map, $\hat{P}(\Lambda,\vec{\xi})$ is a histogram approximation to the probability density of digital colloid microstates at $\vec{\xi}$ for a specific value of Λ, $F^*(\Lambda,\vec{\xi})$ is the reduced free energy of the configurations at location $\vec{\xi}$ and diameter ratio Λ, and $C(\Lambda)$ is an arbitrary additive constant that may differ for each diameter ratio. At constant temperature and volume, $F^* = F/\varepsilon$ is identifiable as the reduced Helmholtz free energy. In applying Eqn. 4 to estimate the equilibrium free energy surface from the empirical probability distribution computed from the Brownian dynamics trajectories, we assume that our simulations are sufficiently long to return converged estimates of the probability distribution over the low-dimensional manifold. Computing $F^* = F/\varepsilon$ from our simulations at a reduced temperature $T^* = 0.1$, we identify $\beta F = \varepsilon F^*/k_BT$. By converting these units to room temperature $T = \varepsilon T^*/k_B = 298$ K, and a characteristic interaction energy of $\varepsilon = 10\ k_BT$, we can calculate $\beta F = \varepsilon F^*/k_BT = F^*/T^* = 10F^*$ in units of k_BT at $T = 298$ K.

Importantly, although each histogram $\hat{P}(\Lambda,\vec{\xi})$ is constructed from an independent simulation trajectory performed at a different value of Λ, the diffusion map was constructed from a library of microstates $\mathscr{L}(N,\{\Lambda\})$ collected over all values of Λ. This permits each individual simulation trajectory to be projected into a shared intrinsic manifold spanned by the same nonlinear basis vectors $\{\vec{\Psi}_i\}_{i=2}^{k+1}$, and allowing us to quantify the impact of Λ upon the underlying free energy landscape of the digital colloid of fixed size N. We have previously referred to this approach as the composite diffusion map approach[18]. We choose the arbitrary constant $C(\Lambda)$ for each value of Λ so as to shift the global free energy minimum to zero (i.e., $C(\Lambda) : \min_{\vec{\xi}} F(\Lambda,\vec{\xi}) = 0$). Uncertainties in the free energy landscape were computed by performing 10 bootstrap resamples (random resampling with replacement) of the data used to compile the histograms at each Λ.

2.4 Determining the digital colloid bit state

Up until now, the halo particles in the calculation have been treated as indistinguishable. To discern one bit state from another, we must now perform a *post hoc* accounting of the distinguishable states of each of the digital colloid microstructures projected into the intrinsic manifold. For the $N = 4$ and $N = 6$ digital colloids considered in this work, we employ a simple procedure to identify the distinguishable bit state of a given microstate based on the N-by-N halo particle adjacency matrix $\mathbf{G}$ and the halo particle positions. We compute the symmetric binary adjacency matrix $\mathbf{G}$ between the halo particles of any particular digital colloid configuration from the pairwise distances matrix χ by specifying a threshold separation of 110% of the halo particle separation in the idealized polyhedral configuration. Halo particles i and j closer than this distance are defined as bonded ($G_{ij} = 1$) and those at larger separations are not ($G_{ij} = 0$). Thus the adjacency matrix maps the set of all possible microstates to a finite set of structures by applying a proximity criterion between halo particles. We have verified that our results are robust to choices of cutoff over the range 105% to 130% of the ideal polyhedron halo particle separation. The symmetric binary adjacency matrix $\mathbf{G}$ cannot, however, distinguish between two states that rotationally distinguishable but related through a mirror reflection, and distinguishing between these states requires additional analyses described below.

2.4.1 N=4 bit states

The tetrahedral structure formed by the N=4 digital colloid belongs to the T_d point group, with rotational symmetry $\Upsilon_4 = 12$. Given the $N! = 24$ possible permutations of halo particle labelings, we have $N!/\Upsilon_N = 2$ rotationally distinguishable states that can be stored in the configuration of these halo particles, corresponding to the two chiral enantiomers of the tetrahedral bit state. We distinguish

the two bit states through their chirality (i.e., handedness) by computing the signed volume, V, of the parallelepiped connecting the halo particle centers of mass using the triple product. The sign of this volume, $\mathrm{sgn}(V)$, determines the handedness of the tetrahedron, allowing us to distinguish between left- and right-handed structures and discriminate between the two bit states.

2.4.2 N=6 bit states

The octahedral structure defined by the N=6 digital colloid belongs to the O_h point group with rotational symmetry $\Upsilon_6 = 24$, giving a total of $N!/\Upsilon_N = 30$ rotationally distinguishable bit states. These 30 bit states correspond to 15 possible unique permutations of the adjacency matrix of an octahedron, each of which has two distinct chiralities owing to a left- or right-handed arrangement of the halo particles occupying the mid-plane of the octahedron. To identify a digital colloid as belonging to a particular bit state, we first classify the structure as containing one of the 15 unique adjacency matrices. To distinguish the two chiral alternatives within each of these 15 matrix classes, we tag the particle corresponding to the first row of the adjacency matrix and specify that particle to reside at the top of the digital colloid (e.g., the red halo particle in Figure 1b). We then identify the three halo particles corresponding to the first three non-zero entries of the first row of the adjacency matrix as three neighboring particles residing in the mid-plane of the octahedron (e.g., the green, yellow, and orange particles in Figure 1b). The signed volume, V, of the parallelepiped connecting these four halo particle centers of mass distinguishes between a clockwise or counter-clockwise arrangement of halo particles in the mid-plane, permitting unambiguous identification of the right- and left-handed chiral bit states. By pairing the adjacency matrix and chirality information, the instantaneous configuration of an arbitrary sized digital colloid can be unambiguously assigned to one of its distinguishable bit states.

2.5 Bit state transition kinetics

Having determined the rotationally distinguishable bit state of each digital colloid microstructure following the approach detailed in Section 2.4, we can track transition events as changes in colloidal bit state over time. To quantify the transition kinetics and the dwell time (or volatility) of each bit state, we compute from our simulations the mean first passage time (MFPT) as the average duration for which a bit state is occupied before a transition occurs[42,60]. The minimum first passage time identified in our simulation trajectories is 80-times larger that the sampling period, assuring that the identified transition events are "elementary transitions" in the sense that they correspond to a single structural rearrangement events. We define for each bit state a core region comprising an ensemble of configurations around the idealized bit state structure. The phase space between core regions comprises a "no man's land" belonging to none of the bit states. The digital colloid is considered to occupy a particular bit state when it passes into its corresponding core region, and to have left that bit state only when it passes into the core region of a new bit state. Transient excursions into the no man's land and then back into the same bit state core region are not considered transitions. The use of non-contiguous core regions separated by a no man's land leads to improved MFPT estimates by eliminating diffusive recrossing events of the separatrix between bit states that can artificially depress the MFPT estimate[61]. The particular definition of the core regions depends on the architecture of the digital colloid under consideration and is detailed in Section 3. Having defined the core regions we analyze our Brownian dynamics production run to extract for each of the $k = 1...K$ bit states all of the $i = 1...P(k)$ first passage times of the system out of that bit state $\{FPT_i^k\}_{i=1}^{P(k)}$. The MFPT for bit state k then follows straightforwardly as $MFPT^k = \frac{1}{P(k)}\sum_{i=1}^{P(k)} FPT_i^k$. Moreover, since the bit states are all structurally identical, we can average the estimate over all N equivalent bit state transitions $MFPT = \frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{P(k)} FPT_i^k$. Similarly, we compute the variance of the first passage time as $\sigma^2 = \frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{P(k)}\left(FPT_i^k - MFPT\right)^2$. Using the central limit theorem, we estimate 95% confidence intervals from the sampling distribution of the MFPT as $\left[MFPT - 1.96\frac{\sigma}{\sqrt{N}}, MFPT + 1.96\frac{\sigma}{\sqrt{N}}\right]$.

The reciprocal of the MFPT is the escape rate $\Gamma = 1/\mathrm{MFPT}$[42], which, since all bit states for a particular digital colloid are structurally equivalent, is identical for all bit states. For the N=4 system, the escape rate is precisely equal to the transition rate between the two chiral bit states. For the N=6 case, the escape rate is equal to the transition rate from any particular bit state to any one of eight other rotationally distinguishable bit states accessible by a single halo particle rearrangement event (Figure S1). This quantification of mean first passage times and escape rates can be straightforwardly extended to digital colloids of arbitrary size.

As detailed in Section 2.1, we adopt experimentally relevant parameter values of $\varepsilon = 10\ k_BT$ at 298 K, $\sigma = 0.915$ μm, and $\rho = 1.228$ g/cm^3 [1,36]. This maps our reduced time step $dt^* = 10^{-3}$ to $dt = 0.54$ μs, and permits us to report the MFPT in seconds and the escape rate in inverse seconds.

3 Results & Discussion

3.1 N=4 (2-state) digital colloids

Using the approach detailed in Section 2.1, we performed Brownian dynamics simulations of the tetrahedral N=4 digital colloid for $\Lambda = \{0.4000, 0.4142, 0.4400, 0.5000, 0.7500, 1.0000, 1.1256\}$. In each case we held the halo particle diameter fixed and varied the size of the central particle. These values of Λ span three regimes: (I) $(\Lambda_{SC} \approx 0.2247) \leq \Lambda < (\Lambda_T = 0.4142)$ where the central particle diameter is sufficiently small that the digital colloid bit state is effectively locked since the halo particles possess insufficient free volume to transition to a new bit state, (II) $(\Lambda_T = 0.4142) \leq \Lambda < (\Lambda_U = 0.5361)$ where the central particle is sufficiently large such that the cluster is unlocked and transitions between bit states are permitted, and (III) $\Lambda \geq (\Lambda_U = 0.5361)$ where the central particle is so large that bit state transitions are effectively barrierless and the digital colloid remains in constant flux between states. $\Lambda_{SC} \approx 0.2247$ is the minimum diameter ratio at which the halo particles can exist on the surface of the central particle without overlaps[1] under the effective hard sphere mapping with $d_{halo} = 3.0785\sigma$. (As observed above, the soft WCA potential does admit a closer distance of approach but we never observe two halo particles to approach closer than 2.95σ over the course of any of our calculations.) The threshold $\Lambda_T = 0.4142$ is the transition threshold diameter ratio below which bit state transitions are geometrically forbidden under the effective hard sphere mapping, and was previously computed from geometric considerations in Ref. [1]. The threshold $\Lambda_U = 0.5361$ is defined by the value of Λ for which the free energy barrier height for bit state transitions falls to 1 k_BT and the transition free energy barrier is comparable to the size of thermal fluctuations (cf. Figure 4a).

From each simulation we extract 2×10^6 digital colloid microstates saved every 500 time steps, and aggregate these into a single composite library $\mathscr{L}(N = 4, \Lambda = \{0.4000, 0.4142, 0.4400, 0.5000, 0.7500, 1.0000, 1.1256\})$ comprising 1.4×10^7 microstates. Constructing diffusion maps over such a large number of microstates is computationally intractable, so we reduce the number of microstates used to construct the diffusion map embedding by subsampling the data. In principle, this could be done uniformly, but to ensure that we have good coverage of the structural conformations we instead take advantage of the vast number of redundant (i.e., nearly identical) microstates in our library. We identify from each simulation trajectory the set of unique adjacency matrices $\mathbf{G}$ and assign each conformation in the trajectory as belonging to one of these prototypical architectures. Aggregating the different architecture lists across all values of Λ yields a set of 34 prototypical architectures. To obtain good coverage over this range of structures, we sample 30 configurations from each architecture (when there are less than 30 configurations, we select all observed microstates) to define a representative set of 941 digital colloid microstates over the various values of Λ.

We then applied diffusion maps to these 941 microstates using the procedure detailed in Section 2.2.2, employing a soft-thresholding bandwidth of $\varepsilon = \exp(3)$ determined via the method defined in Ref. [52]. Using the L-method[53], we identified a gap in the eigenvalue spectrum beyond λ_3, prompting the construction of a two-dimensional landscape in the top two non-trivial diffusion map eigenvectors (Ψ_2, Ψ_3). The remaining $(1.4 \times 10^7 - 941)$ configurations were projected into this two-dimensional intrinsic manifold using the Nyström extension[62–64]. By constructing a diffusion map over microstates collected from simulations spanning all values of Λ, we can project each simulation trajectory into a shared intrinsic manifold and quantify the impact of Λ upon the morphology, stability, and kinetics of the digital colloid within a unified low-dimensional basis[18].

Morphology and stability. We present in Figure 2a the reduced ensemble of digital colloid structures observed at the transition threshold value of $\Lambda_T = 0.4142$ in the two dimensional intrinsic manifold spanned by (Ψ_2, Ψ_3). Each point in this manifold corresponds to a single digital colloid configuration observed in our simulation. To aid in visualization, we have superimposed onto the manifold renderings of representative snapshots visualized using VMD[65]. Given a particular labeling of the distinguishable halo particles (A – red, B – yellow, C – blue, D – green) illustrated in Figure 2c, we compute for each digital colloid microstate the vector triple product $V = \hat{AB} \cdot (\hat{AC} \times \hat{AD})$ defining the signed volume of a parallelepiped connecting the halo particle centers of mass, where $\hat{XY}$ denotes the unit vector pointing from X to Y. By employing unit vectors, the magnitude of V is bounded on the interval [0,1] irrespective of Λ and can be interpreted as a measure of the deviation of the halo particle configuration from planarity: $|V| = 0$ corresponds to all halo particles residing in a single plane, and $|V| = \sqrt{2}/2$ to an idealized tetrahedral arrangement. The sign of V distinguishes the chirality of the two tetrahedral bit states: $V > 0$ corresponds to right-handed tetrahedra and $V < 0$ to left-handed configurations.

Coloring each point in Figure 2a by $|V|$ reveals a strong correlation between Ψ_3 and $|V|$, indicating that Ψ_3 was discovered by the diffusion map as a collective variable measuring the relative planarity of the halo particle configuration. The minimum value of $|V| = 0$ is attained at the top of the tear-shaped intrinsic manifold in the region of

$(\Psi_2 \approx 0.01, \Psi_3 \approx 0.085)$ where the digital colloid adopts a square planar geometry. The largest values of $|V| \approx \sqrt{2}/2$ are reached at $(\Psi_2 \approx -0.03, \Psi_3 \approx -0.04)$ where the digital colloid adopts approximately tetrahedral geometries. The square planar geometry defines the transition state for interconversions between bit states[1]. At the transition threshold value of $\Lambda_T = 0.4142$ the transition state comprises essentially a single point within the intrinsic manifold since the free volume available to the halo particles at the transition state is so small as to admit essentially a unique square planar configuration. Contrariwise, since $(\Lambda_T = 0.4142) > \Lambda_{SC}$ the tetrahedral structures occupy a relatively larger volume at the bottom of the tear-shaped intrinsic manifold corresponding to a diversity of structures defining the metastable bit state rattling around the caged volume centered on the idealized tetrahedral geometry.

As a result of our choice to construct diffusion map embeddings without regard to halo particle distinguishability (Section 2.2.1), the two rotationally distinguishable chiral bit states accessible to the N=4 digital colloid are collapsed together onto the intrinsic manifold in Figure 2a. In order to separate these two bit states we present in Figure 2b the diffusion map embedding augmented with a third axis containing the signed triple product $V = \vec{AB} \cdot \left(\vec{AC} \times \vec{AD}\right)$ as a measure of both microstate morphology and chiral bit state. This augmented embedding separates the two chiral bit states and makes clear their structural equivalence, with each occupying a topographically identical volume of the intrinsic manifold and connected by the square planar transition structure occupying the slim bottleneck at $V = 0$.

In Figure 3, we present the augmented diffusion map embeddings for all seven values of Λ. In each case we present both the low-dimensional diffusion map projection into (Ψ_2, Ψ_3, V) of the 2×10^6 microstates recorded over the course of this simulation trajectory, and the free energy surface over the intrinsic manifold $F(\Psi_2, \Psi_3, V)$ defined by collecting histograms over the embedding. As detailed in Section 2.3, by assuming a characteristic interaction strength between halo particles of $\varepsilon = 10\ k_BT$ at 298 K (cf. Eqn. 1), we report βF as a dimensionless free energy measured in units of k_BT at $T = 298$ K. In Movie S1 in the Electronic Supplementary Information (ESI), we present rotating images of the images in Figure 3 that more clearly illustrate the shape and topography of the intrinsic manifolds and free energy surfaces.

For $(\Lambda = 0.4000) < (\Lambda_T = 0.4142)$ (Figure 3a,b), structural transitions are effectively forbidden and the digital colloid is locked into the bit state in which it was initialized. The square planar transition state at $V = 0$ is effectively inaccessible due to excluded volume interactions and is never visited over the course of our simulations.

At $\Lambda = (\Lambda_T = 0.4142)$ (Figure 3c,d), the digital colloid is at the unlocking threshold and is able to access the square planar transition state $|V| = 0$ and make transitions between the two bit states. The transition bottleneck exists at the top of a high $\Delta(\beta F) = (9.07 \pm 0.23)$ free energy barrier. Accordingly, bit state transitions are an activated process and a rare event, and we observe only three bit state transitions over the course of the simulation. The structural equivalence of the two chiral bit states means that the topography of the free energy surface within each metastable basin should be identical. The observed differences in the empirical free energy values computed within the two basins is a consequence of kinetic trapping and rare transition events leading to poor sampling of the accessible phase space.

For $\Lambda > (\Lambda_T = 0.4142)$ (Figure 3e-n), the volume of the transition state ensemble at $|V| = 0$ expands as the increased free volume available to the halo particles admits a greater diversity of square planar transition structures. There is a commensurate drop in the transition barrier to $\Delta(\beta F) = (5.15 \pm 0.07)$ at $\Lambda = 0.4400$, to $\Delta(\beta F) = (1.93 \pm 0.02)$ at $\Lambda = 0.5000$, before the transitions become essentially barrierless with $\Delta(\beta F) < 1$ for $\Lambda > (\Lambda_U = 0.5361)$.

We present in Figure 4a the free energy barrier as a function of Λ. An exponential fit of the form $\Delta(\beta F) = C_0 \exp(-C_1 \Lambda)$ provides a good fit to the data, with best fit parameters $\{C_0 = (1.29 \pm 0.89) \times 10^4,\ C_1 = (17.65 \pm 1.55)\}$ and an adjusted $R^2 = 0.98$. Quantification of the free energy barrier for the transition provides a measure of the reversible work required to actuate a transition between bit states, or equivalently the size of a concerted thermal fluctuation necessary to induce a spontaneous transition. This stands as one of the key variables for the design of stable but addressable information storage elements.

To ascertain the change in the phase space volume and structural morphologies of the transition state as a function of Λ, we present in Figure 4b a superposition of 2D (Ψ_2, Ψ_3) sections through the 3D free energy landscapes at $V = 0$, where the halo particles occupy planar configurations at the transition between the two chiral bit states. The broadening of the transition region bottleneck with increasing Λ is a result of the increased free volume available to the halo particles moving over the surface of increasingly large central particles, permitting access to a greater diversity of planar transition structures and a commensurate reduction in the free energy barrier between the two bit states.

Kinetics. The depression of the transition state free energy with increasing Λ leads to more frequently observed transitions between the two chiral bit states. To quantify

the transition rate as a function of Λ we report in Figure 5 the mean first passage time (MFPT) computed from each simulation as the average duration for which a bit state is occupied before a transition occurs[42,60]. MFPT calculations were conducted for Λ ={0.4142, 0.4200, 0.4300, 0.4400, 0.4500, 0.4750, 0.5000, 0.5100, 0.5200, 0.5300, 0.5400, 0.5500, 0.6000, 0.6500, 0.7000, 0.7500, 1.000, 1.1256}. To mitigate the impact of possible recrossing events at the summit of the free energy barrier leading to artificially depressed MFPT estimates, we define a core regions within each bit state separated by a no man's land[61]. We define the core region of the right-handed bit state to be the region of the intrinsic manifold for which $V > 0.5$, and that of the left-handed bit state to be the region for which $V < -0.5$. The digital colloid is defined to enter a particular bit state the first time it passes into the corresponding core region, and is defined to have transitioned out of that bit state only when it passes into the core region of a new bit state. Excursions into the interstitial no man's land (i.e., $-0.5 \leqslant V \leqslant 0.5$) followed by re-entry into the core region of the same bit state are not considered transition events.

We present in Figure 5 the calculated MFPT as a function of Λ and associated error bars delimiting 95% confidence intervals. A shifted exponential of the form $\ln(\text{MFPT}) = B_0 \exp(-B_1\Lambda) + B_2$ provides a good fit to the data, with best fit parameters $\{B_0 = (2.03 \pm 0.91) \times 10^3, B_1 = (14.08 \pm 0.92), B_2 = (-2.53 \pm 0.04)\}$ and an adjusted $R^2 = 0.97$.

For $\Lambda < (\Lambda_T = 0.4142)$, the cluster is effectively locked and the MFPT is undefined since transitions are sterically forbidden. For $\Lambda \geq (\Lambda_T = 0.4142)$, the cluster unlocks and MFPT rapidly decreases with increasing Λ as the free volume available to the halo particles increased and the transition free energy barrier decreases (cf. Figure 4). The MFPT approaches an approximate asymptotic limit at $\Lambda > (\Lambda_U = 0.5361)$ corresponding to a central particle so large that the motions of the halo particles are largely decorrelated and transitions are effectively barrierless. The escape rate $\Gamma = 1/\text{MFPT}$ defines the transition rate from one bit state to the other[42] and is a critical design variable in the practical realization of digital colloids for information storage.

3.2 N=6 (30-state) digital colloids

Using an analogous approach, we conducted Brownian dynamics simulations of an N=6 digital colloid for $\Lambda = \{0.4142, 0.5100, 0.5275, 0.5500, 0.5750, 0.6000, 0.6500\}$, again holding the halo particle size fixed and varying the center particle diameter. There are three regimes of Λ: (I) $(\Lambda_{SC} \approx 0.4142) \leq \Lambda < (\Lambda_T = 0.5275)$ where the digital colloid configuration is effectively locked into a particular octahedral bit state configuration, and halo particle transitions between bit states are sterically forbidden, (II) $(\Lambda_T = 0.5275) \leq \Lambda < (\Lambda_U = 0.7021)$ where the configuration becomes unlocked as halo particles are able to collectively transition between bit states, and (III) $\Lambda \geq (\Lambda_U = 0.7021)$ where transitions between bit states are effectively barrierless, and the digital colloid configuration is in constant flux between states. The spherical code diameter ratio $\Lambda_{SC} \approx 0.4142$ and unlocking transition threshold $\Lambda_T = 0.5275$ were previously reported in Ref. [1] for the effective hard sphere mapping. The barrierless transition threshold $\Lambda_U = 0.7021$ is defined as the value of Λ at which the (extrapolated) free energy barrier for bit state transitions drops to 1 k_BT (cf. Figure 7a).

We extract from our simulations 2×10^6 microstates saved every 500 time steps for each value of Λ and compile these into a composite library $\mathscr{L}(N = 6, \Lambda = \{0.4142, 0.5100, 0.5275, 0.5500, 0.5750, 0.6000, 0.6500\})$ comprising 1.4×10^7 microstates. We extract a restricted set of 1,932 representative microstates spanning the 110 prototypical cluster architectures observed in the simulations, and apply diffusion maps employing a soft-thresholding bandwidth of $\varepsilon = \exp(2)$[52]. Using the L-method we identify a gap in the eigenvalue spectrum above λ_4[53], prompting construction of a three-dimensional embedding into the top three non-trivial diffusion map eigenvectors (Ψ_2, Ψ_3, Ψ_4). We identify three collective variables to parameterize the intrinsic manifold of the N=6 system compared to only two for the N=4 case, reflecting the richer configurational dynamics available to the larger system. The remaining (1.4×10^7-1,932) configurations were projected into this reduced embedding via the Nyström extension.

Morphology and stability. We present in Figure 6 the three-dimensional diffusion map embeddings for all seven values of Λ, constructed by projecting the 2×10^6 microstates recorded from each simulation into (Ψ_2, Ψ_3, Ψ_4). We also report the associated free energy landscapes $F(\Psi_2, \Psi_3, \Psi_4)$ computed from histograms over the low-dimensional projections. Rotating movies of these figures are provided in Movie S2 of the ESI.

Our diffusion map embeddings treat all halo particles as indistinguishable, collapsing together the structurally identical $(N!/\Upsilon_N) = 30$ bit states within the intrinsic manifold spanned by (Ψ_2, Ψ_3, Ψ_4). In the case of N=4 we were able to distinguish the two chiral bit states by augmenting our low-dimensional projection with the signed volume V of the parallelepiped mapped out by the halo particles. For the N=6 case, no simple order parameters are available to separate out the 30 rotationally distinguishable bit states within our low dimensional embedding. Accordingly, we do not augment our embeddings but appreciate – by virtue

of the structural equivalence of all bit states – that the low-dimensional projections and free energy surfaces pertain equally to all bit states. The connectivity graph illustrating the bit states that are mutually accessible by single collective rearrangement events of the halo particles are illustrated in Figure S1 in the ESI.

To gain insight into the physical interpretation of the diffusion map variables, we project representative microstates onto the low-dimensional manifolds in Figure 6 and color each microstate according to the $L_{1,1}$ norm between its normalized halo particle pairwise distances matrix and that of an idealized octahedron $||\chi' - \chi'_{\text{oct}}||$ (cf. Section 2.2.1). The collective variable Ψ_2 is well-correlated with $||\chi' - \chi'_{\text{oct}}||$ indicating that it provides a measure of the octahedral character of the digital colloid architecture. The minimum value of $||\chi' - \chi'_{\text{oct}}|| = 0$ is found in the region of $(\Psi_2 \approx 0.05, \Psi_3 \approx 0.03, \Psi_4 \approx -0.03)$ where the digital colloid adopts an idealized octahedral geometry. Each of the low-dimensional manifolds in Figure 6 is funnel-shaped, with the narrow tip containing the transition state ensemble and wide base corresponding to the "rattling" of halo particles about a particular stable bit state. The size of the tip of the funnel increases with Λ as the increased free volume accessible to the halo particles opens access to a greater diversity of transition state configurations.

At $\Lambda = (\Lambda_{\text{SC}} \approx 0.4142)$ (Figure 6a,b) – corresponding to the minimum diameter ratio for which effective hard sphere halo particles can exist on the surface of the central particle without overlapping[1,24] – the digital colloid is confined to a small set of possible configurations localized around the ideal octahedron. At $(\Lambda = 0.5100) < (\Lambda_{\text{T}} = 0.5275)$ (Figure 6c,d), the digital colloid remains effectively locked in a single bit state, but the increased free volume permits the halo particles to rattle around around the ideal octahedron and fill a greater volume of the intrinsic manifold.

At the unlocking threshold of $\Lambda = (\Lambda_{\text{T}} = 0.5275)$ (Figure 6e,f), a structure very close to the triangular prism transition state between octahedral bit states emerges at $(\Psi_2 \approx -0.03, \Psi_3 \approx -0.03, \Psi_4 \approx -0.07)$[24]. Nevertheless, bit state transitions remain sterically disfavored and – in line with Figure SI-I-3 in Ref. [1] – no transition events are observed over the course of our simulations.

For $(\Lambda = 0.5500) > (\Lambda_{\text{T}} = 0.5275)$ (Figure 6g-h), there is sufficient free volume for the halo particles to realize the triangular prism transition structure, and we observed multiple bit state transitions over the course of our calculations. The free energy surface develops a funnel-like structure with the caged octahedral structures occupying a broad low-free energy region at the bottom of the funnel-shaped intrinsic manifold, and the triangular prism transition state occupying a small high-free energy region at the narrow end of the funnel. We define microstates as belonging to the transition state ensemble by thresholding the $L_{1,1}$ norm between the normalized halo particle pairwise distance matrix and that of the idealized triangular prism, given as $||\chi' - \chi'_{\text{prism}}|| < 1.75$. These configurations possess an average 5% deviation in the normalized pairwise distances matrix away from that of the idealized triangular prism. Productive transitions between bit states occur when the digital colloid enters the transition state and the halo particles undergo a collective rearrangement into a new rotationally distinguishable configuration (cf. Figure S1 in the ESI). The free energy barrier between the bottom of the octahedral free energy well and the transition state is $\Delta(\beta F) = (8.22 \pm 0.17)$.

At larger values of Λ (Figure 6i-n), the volume of the intrinsic manifold increases with the elevated free volume available to the halo particles. The transition state ensemble becomes markedly enlarged since the two planar layers of the triangular prism are able to move more freely over the center particle surface, leading to an expansion of the intrinsic manifold toward $(\Psi_2 \approx -0.10, \Psi_3 \approx 0.00, \Psi_4 \approx -0.05)$ due to a greater diversity of transition structures. There is a commensurate decrease in the transition barrier height from $\Delta(\beta F) = (5.52 \pm 0.07)$ at $\Lambda = 0.5750$, to $\Delta(\beta F) = (3.99 \pm 0.06)$ at $\Lambda = 0.6000$, to $\Delta(\beta F) = (2.14 \pm 0.04)$ at $\Lambda = 0.6500$.

We report the dependence of the transition state free energy barrier as a function of Λ in Figure 7a. An exponential fit of the form $\Delta(\beta F) = C_0 \exp(-C_1 \Lambda)$ provides an excellent fit to the data, with best fit parameters $\{C_0 = (1.41 \pm 0.58) \times 10^4,\ C_1 = (13.60 \pm 0.70)\}$ and an adjusted $R^2 = 0.99$. In the next section, we will use this dependence to design of the digital colloids as information storage elements that have sufficiently low free energy barriers so as to be writable (i.e., their bit state can be changed by doing external work) but sufficiently high to be stable to thermally-activated spontaneous changes in the bit state.

Kinetics. We present in Figure 7b the MFPT as a function of Λ. MFPT calculations were conducted for $\Lambda = \{0.5480, 0.5490, 0.5500, 0.5550, 0.5600, 0.5700, 0.5750, 0.5800, 0.5900, 0.6000, 0.6100, 0.6200, 0.6300, 0.6400, 0.6500\}$. The MFPT corresponds to the mean residence time of the digital colloid within a bit state before it makes a transition to one of the eight other bit states accessible by a single collective rearrangement of the halo particles (Figure S1 in the ESI). To eliminate artificial depression of the MFPT due to recrossing events, we define the core region of a bit state to be that region of the intrinsic manifold containing configurations possessing a halo particle adjacency matrix equivalent to that of the ideal octahedron. A transition is

defined when the digital colloid adjacency matrix departs from the octahedral architecture corresponding to one rotationally distinguishable bit state then reforms the octahedral adjacency matrix in different bit state. An illustration of a productive transition event for $\Lambda = 0.5750$ is illustrated in Movie S3 of the ESI, where we simultaneously visualize the digital colloid configuration, follow its location on the low dimensional manifold, and map the transition between a pair of the 30 rotationally distinguishable bit states.

A shifted exponential of the form $\ln(\text{MFPT}) = B_0 \exp(-B_1\Lambda) + B_2$ provides a good fit to the data, with best fit parameters $\{B_0 = (2.30 \pm 0.16) \times 10^6, B_1 = (23.17 \pm 0.12), B_2 = (-1.91 \pm 0.08)\}$ and an adjusted $R^2 = 0.93$. In the effectively locked region $\Lambda < (\Lambda_T = 0.5275)$ the MFPT is undefined, before dropping steeply beyond the unlocking threshold and attaining an approximate plateau at $\Lambda > (\Lambda_U = 0.7021)$ where transitions become effectively barrierless. The escape rate $\Gamma = 1/\text{MFPT}$ quantifies the volatility of a digital colloid memory storage element due to thermally-driven bit state flips. In the next section, we will combine this design variable with the transition free energy barrier to design $N=6$ digital colloid memory elements.

3.3 Rational design of $N=6$ digital colloid memory storage elements

We now proceed to use the calculated dependencies of the transition barrier height $\Delta(\beta F)$ and escape rate Γ on Λ to rationally design a $N=6$ digital colloid with desired memory volatility and/or actuation free energy. In Figure 8 we present a parametric plot of the escape rate $\Gamma = 1/\text{MFPT}$ against the transition barrier height $\Delta(\beta F)$ for the four diameter ratios at which bit state transitions were observed $\Lambda = \{0.5500, 0.5750, 0.6000, 0.6500\}$. We may interpret the transition barrier height as the reversible work required to actuate a halo particle rearrangement and intentionally write a new bit state into the digital colloid. The escape rate can be conceived as the rate at which errors are introduced due to thermally-activated spontaneous bit state transitions. An ideal storage medium would be both nonvolatile and cheap to address, implying a low write free energy and a low error rate. Since the transition barrier height controls both of these properties, they are competing design constraints and our design objective becomes to identify the value of Λ that minimizes the write free energy while maintaining the error rate below a user-specified tolerance. As an illustrative example, we consider the storage of a short text block for the period of one hour, and illustrate the tradeoff between the expected error rate and write free energy. Our analysis indicates that relatively large transition barriers are required to preserve data integrity over any significant time period. Determining the tradeoff between volatility and ease of writing will depend largely on the specifics of the system, but this approach can help guide the design process by imposing the engineering requirements of the write element and intended storage duration.

4 Conclusions

We show how to use low-dimensional embeddings to characterize the stability and dynamical transition pathways for digital colloids. By applying diffusion maps to simulation-generated trajectories of the digital colloid dynamics, we discover a nonlinear projections of the digital colloid dynamics into low-dimensional embeddings. These projections of the dynamics into a low-dimensional space reveal the metastable bit states and transition pathways between them. The free energy surfaces supported by these low-dimensional embeddings quantify the stability of the bit states and height of the transition barriers for bit state interconversions as a measure of the reversible work required to write a new bit state into the digital colloid. The low-dimensional embeddings also reveal the diversity of morphologies within the metastable bit state minima and transition state ensemble, and a means to robustly identify transitions between bit states. Mean first passage times calculated from our simulations provide a measure of the average dwell time within a bit state, and a measure of memory volatility and expected rates of memory degradation due to thermally-activated bit state transitions.

We apply our method to the $N=4$ and $N=6$ digital colloids capable of storing 0.125 bytes and 0.613 bytes, respectively. Using our calculated barrier heights and escape rates for the $N=6$ system, we calculate the tradeoff between the reversible work required to write a bit state change and the memory degradation rate due to thermal fluctuations. The results of this calculation guide the rational design of the relative halo and central particle diameters for a memory storage application.

We also describe how to extend our analysis to larger systems. Our exhaustive graph matching methods to compute the microstate pairwise distances required by the diffusion map are limited by their high computational cost. We are currently exploring and developing approximate methods with sufficiently high efficiency and fidelity to robustly recover the low-dimensional embeddings for larger digital colloids with higher information storage capacity, such as the 2.87 byte $N=12$ system[1].

The double-exponential dependence of mean first passage times on Λ shows that rearrangement kinetics of digital colloids are extremely sensitive to Λ near the locking transition. This fact highlights opportunities for designing

digital colloids that can "lock" and "unlock" in response to relatively small changes in the central particle size, and also provides guidelines for manufacturaing digital colloids that satisfy long-term data storage requirements. Examples of $N=4$ digital colloids have been realized by Grier, Pine and coworkers using holographic optical tweezers[1], but it remains a challenge to produce larger clusters and fabricate them at scale. Based on the findings of this work, we anticipate next steps for manufacturing bulk quantities of digital colloids will depend on controlling the monodispersity of central particles and developing mechanisms for shrinking and swelling the central particles controllably. Moreover, robust and scalable methods to read, store, and write information to the colloids must be developed before this can be used as a practical information storage substrate. Techniques based on DNA functionalization[66], FRET barcoding[5], and swellable central particles[67] have been suggested, and this remains an area of active research. We anticipate that the mutually beneficial feedback between simulation, machine learning-enabled data analysis, and advanced experimental synthesis and imaging techniques will improve our understanding and control of this novel and potentially ultra-high capacity information storage medium.

Acknowledgements

This material is based upon work supported by a National Science Foundation CAREER Award to A.L.F. (Grant No. DMR-1350008).

References

1 C. L. Phillips, E. Jankowski, B. J. Krishnatreya, K. V. Edmond, S. Sacanna, D. G. Grier, D. J. Pine and S. C. Glotzer, *Soft Matter*, 2014, **10**, 7468–7479.

2 S. Sacanna, W. T. M. Irvine, P. M. Chaikin and D. J. Pine, *Nature*, 2010, **464**, 575–578.

3 J. T. McGinley, I. Jenkins, T. Sinno and J. C. Crocker, *Soft Matter*, 2013, **9**, 9119–9128.

4 L. Wang, W. Zhao and W. Tan, *Nano Research*, 2008, **1**, 99–115.

5 L. Wang, and W. Tan, *Nano Letters*, 2006, **6**, 84–88.

6 L. Qian and E. Winfree, *Science*, 2011, **332**, 1196–1201.

7 C. L. Phillips, E. Jankowski, M. Marval and S. C. Glotzer, *Physical Review E*, 2012, **86**, 041124.

8 T. W. Melnyk, O. Knop and W. R. Smith, *Canadian Journal of Chemistry*, 1977, **55**, 1745–1761.

9 L. Whyte, *The American Mathematical Monthly*, 1952, **59**, 606–611.

10 J. Edmundson, *Acta Crystallographica Section A: Foundations of Crystallography*, 1992, **48**, 60–69.

11 R. Tammes, *Recueil des Travaux Botaniques Neerlandais*, 1930, **27**, 1–84.

12 J. J. Thomson, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1904, **7**, 237–265.

13 D. Weaire and T. Aste, *The pursuit of perfect packing*, CRC Press, 2008.

14 W. D. S. N. J. A. Sloane, with the collaboration of R. H. Hardin *et al.*, *Tables of Spherical Codes, published electronically at NeilSloane.com/packings/*.

15 A. W. Long and A. L. Ferguson, *J. Phys. Chem. B*, 2014, **118**, 4228–4244.

16 P. Das, M. Moll, H. Stamati, L. E. Kavraki and C. Clementi, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 9885–9890.

17 M. Ceriotti, G. A. Tribello and M. Parrinello, *Proceedings of the National Academy of Sciences*, 2011, **108**, 13023–13028.

18 R. A. Mansbach and A. L. Ferguson, *J. Chem. Phys.*, 2015, **142**, 105101.

19 A. W. Long, J. Zhang, S. Granick and A. L. Ferguson, *Soft Matter*, 2015, **11**, 8141–8153.

20 M. A. Bevan, D. M. Ford, M. A. Grover, B. Shapiro, D. Maroudas, Y. Yang, R. Thyagarajan, X. Tang and R. M. Sehgal, *J. Process Contr.*, 2015, **27**, 64 – 75.

21 R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner and S. W. Zucker, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 7426–7431.

22 R. R. Coifman and S. Lafon, *Appl. Comput. Harm. Anal.*, 2006, **21**, 5–30.

23 R. Coifman, I. Kevrekidis, S. Lafon, M. Maggioni and B. Nadler, *Multiscale Model. Simul.*, 2008, **7**, 842–864.

24 C. L. Phillips, E. Jankowski, M. Marval and S. C. Glotzer, *Phys. Rev. E*, 2012, **86**, 041124.

25 C. L. Phillips, J. A. Anderson and S. C. Glotzer, *J. Comput. Phys.*, 2011, **230**, 7191 – 7201.

26 J. A. Anderson, C. D. Lorenz and A. Travesset, *J. Comput. Phys.*, 2008, **227**, 5342–5359.

27 J. Glaser, T. D. Nguyen, J. A. Anderson, P. Lui, F. Spiga, J. A. Millan, D. C. Morse and S. C. Glotzer, *Comput. Phys. Commun.*, 2015, **192**, 97 – 107.

28 J. D. Weeks, D. Chandler and H. C. Andersen, *J. Chem. Phys.*, 1971, **54**, 5237–5247.

29 H. C. Andersen, J. D. Weeks and D. Chandler, *Phys. Rev. A*, 1971, **4**, 1597.

30 L. Filion, R. Ni, D. Frenkel and M. Dijkstra, *J. Chem. Phys.*, 2011, **134**, 134901.

31 J. A. Barker and D. Henderson, *J. Chem. Phys.*, 1967, **47**, 4714–4721.

32 J. Rowlinson, *Mol. Phys.*, 1964, **7**, 349–361.

33 J. Rowlinson, *Mol. Phys.*, 1964, **8**, 107–115.

34 C. L. Phillips, J. A. Anderson and S. C. Glotzer, *Journal of Computational Physics*, 2011, **230**, 7191 – 7201.

35 T. Schlick, *Molecular Modeling and Simulation: An Interdisciplinary Guide*, Springer Science & Business Media, 2010.

36 S. Sacanna, W. T. Irvine, L. Rossib and D. J. Pinea, *Soft Matter*, 2011, **7**, 1631–1634.

37 M. K. Transtrum, B. B. Machta, K. S. Brown, B. C. Daniels, C. R. Myers and J. P. Sethna, *J. Chem. Phys.*, 2015, **143**, 010901.

38 A. L. Ferguson, A. Z. Panagiotopoulos, P. G. Debenedetti and I. G. Kevrekidis, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 13597–13602.

39 A. L. Ferguson, A. Z. Panagiotopoulos, I. G. Kevrekidis and P. G. Debenedetti, *Chem. Phys. Lett.*, 2011, **509**, 1 – 11.

40 M. A. Rohrdanz, W. Zheng, M. Maggioni and C. Clementi, *J. Chem. Phys.*, 2011, **134**, 124116.

41 H. Stamati, C. Clementi and L. E. Kavraki, *Proteins: Struct., Funct., Bioinf.*, 2010, **78**, 223–235.

42 R. Zwanzig, *Nonequilibrium Statistical Mechanics*, Oxford University Press, USA, 2001.

43 I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, 2nd edn, 2002.

44 T. Cox and M. Cox, *Multidimensional Scaling, Second Edition*, CRC Press, 2000.

45 A. L. Ferguson, S. Zhang, I. Dikiy, A. Z. Panagiotopoulos, P. G. Debenedetti and A. J. Link, *Biophys. J.*, 2010, **99**, 3056 – 3065.

46 P. G. Bolhuis, D. Chandler, C. Dellago and P. L. Geissler, *Annual Review of Physical Chemistry*, 2002, **53**, 291–318.

47 L. Maragliano, A. Fischer, E. Vanden-Eijnden and G. Ciccotti, *Journal of Chemical Physics*, 2006, **125**, 024106.

48 H. Jónsson, G. Mills and K. W. Jacobsen, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, ed. B. J. Berne, G. Ciccotti and D. F. Coker, World Scientific, Singapore, 1998, ch. Nudged elastic band method for finding minimum energy paths of transitions, pp. 385–404.

49 W. E and E. Vanden-Eijnden, *Annual Review of Physical Chemistry*, 2010, **61**, 391–420.

50 A. L. Ferguson, A. Z. Panagiotopoulos, P. G. Debenedetti and I. G. Kevrekidis, *Journal of Chemical Physics*, 2011, **134**, 135103.

51 B. Nadler, S. Lafon, R. R. Coifman and I. G. Kevrekidis, *Advances in Neural Information Processing Systems 18: Proceedings of the 2005 Conference (Neural Information Processing)*, The MIT Press, 2006, pp. 955–962.

52 R. Coifman, Y. Shkolnisky, F. Sigworth and A. Singer, *IEEE Trans. Image Process.*, 2008, **17**, 1891–1899.

53 S. Salvador and P. Chan, Tools with Artificial Intelligence, 2004. ICTAI 2004. 16th IEEE International Conference on, 2004, pp. 576–584.

54 M. Scholz, M. Fraunholz and J. Selbig, *Principal Manifolds for Data Visualization and Dimension Reduction*, Springer, 2008, pp. 44–67.

55 A. Ma and A. R. Dinner, *J. Phys. Chem. B*, 2005, **109**, 6769–6779.

56 B. Peters and B. L. Trout, *J. Chem. Phys.*, 2006, **125**, 054108.

57 J. Xing and K. S. Kim, *J. Chem. Phys.*, 2011, **134**, 044132.

58 M. S. Shell, in *Thermodynamics and Statistical Mechanics: An Integrated Approach*, Cambridge University Press, 2015, ch. Chapter 16.

59 M. S. Shell, in *Thermodynamics and Statistical Mechanics: An Integrated Approach*, Cambridge University Press, 2015, ch. Chapter 21.

60 S. Redner, *A Guide to First-Passage Processes*, Cambridge University Press, 2001.

61 G. R. Bowman, in *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, ed. G. R. Bowman, V. S. Pande and F. Noé, Springer Science & Business Media, 2013, vol. 797, ch. 2. An overview and practical guide to building Markov state models, pp. 20–21.

62 C. T. Baker and C. Baker, *The Numerical Treatment of Integral Equations*, Clarendon Press, Oxford, 1977, vol. 13.

63 C. R. Laing, T. A. Frewen and I. G. Kevrekidis, *Nonlinearity*, 2007, **20**, 2127.

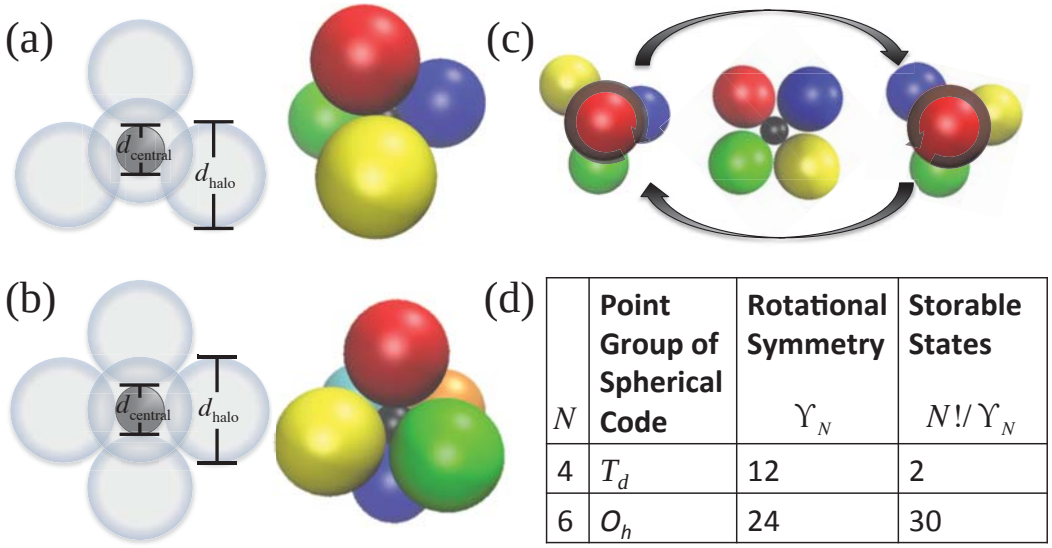
64 B. E. Sonday, M. Haataja and I. G. Kevrekidis, *Phys. Rev. E*, 2009, **80**, 031102.

65 W. Humphrey, A. Dalke and K. Schulten, *Journal of Molecular Graphics*, 1996, **14**, 33 – 38.

66 N. Geerts and E. Eiser, *Soft Matter*, 2010, **6**, 4647–4660.

67 K. Kratz, T. Hellweg and W. Eimer, *Colloids Surf., A*, 2000, **170**, 137 – 149.

68 S. Chatterjee and A. S. Hadi, *Regression Analysis by Example*, John Wiley & Sons, Hoboken, New Jersey, 5th edn, 2012.



(d)

N	Point Group of Spherical Code	Rotational Symmetry Υ_N	Storable States $N!/\Upsilon_N$
4	T_d	12	2
6	O_h	24	30

Fig. 1 Schematic of the digital colloid architecture with N distinguishable halo particles with diameter d_{halo} tethered to the surface of a central particle with diameter d_{central}. The diameters exist in a ratio $\Lambda = d_{\text{central}}/d_{\text{halo}}$. The *spherical code* structure defines the densest packing of halo particles around the central particle at the minimum diameter ratio Λ_{SC} for which the halo particles do not overlap [1], and defines the idealized structural arrangement of halo particles defining a distinguishable *bit state*. At $\Lambda = \Lambda_{\text{SC}}$, the halo particles of a (a) N=4 digital colloid define the vertices of a tetrahedron, and (b) an octahedron in a N=6 digital colloid. For $\Lambda_{\text{SC}} < \Lambda < \Lambda_{\text{T}}$ there is sufficient free volume for the halo particles to rattle around the stable spherical code structure, but not transition between different structures. Above a critical diameter ratio $\Lambda > \Lambda_{\text{T}} > \Lambda_{\text{SC}}$, there is sufficient free volume for the halo particles to cooperatively transition between different distinguishable configurations of the halo particles. (c) It is the availability of distinguishable bit states and transition pathways between them that makes it possible to store information within digital colloids. The N=4 digital colloid, for example, possesses two distinguishable tetrahedral bit states corresponding to left- and right-handed chiral arrangements. Transitions between these two chiral configurations mediated by a structural transition through a square planar arrangement of halo particles. (d) Table of the symmetry Υ_N and information storage capability of the N=4 and N=6 digital colloids. The more rotationally distinguishable bit states available the higher the information storage capacity. By using configurations for information storage, these digital colloids are capable of ultra-high information storage density. For example, N=6 digital colloids composed of 3-methacryloxypropyl trimethoxysilane (TPM) colloidal particles with density ρ = 1.228 g/cm^3 [36] and particle diameters of 1 μm possess an estimated memory density of 1.1 TB / g.

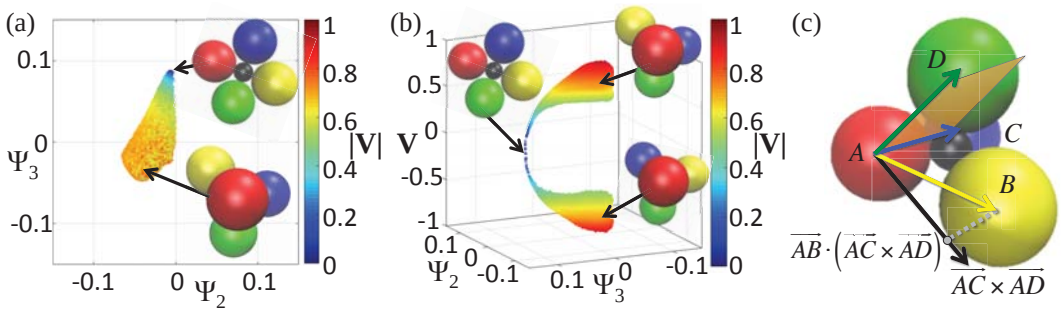


Fig. 2 Diffusion map embeddings of the N=4 tetrahedral digital colloid at the unlocking threshold $\Lambda_{\mathrm{T}} = 0.4142$. (a) Embedding into the top two collective modes (Ψ_2, Ψ_3) furnished by the diffusion map. Each point represents one of the 2×10^6 microstates recorded over the course of this simulation trajectory. To assist in interpretation of the embedding, we present visualizations of representative microstates and color points by the absolute volume of the parallelepiped $|V| = |\vec{AB} \cdot \left(\vec{AC} \times \vec{AD}\right)|$ connecting the centers of mass of the halo particles. The magnitude of V provides a measure of the deviation of the halo particle configuration from planarity: $|V| = 0$ corresponds to all halo particles residing in a single plane, and $|V| = \sqrt{2}/2$ to an idealized tetrahedral arrangement. (b) Augmentation of the 2D intrinsic manifold in panel (a) with a third axis containing the signed volume of the parallelepiped V. The sign of V distinguishes the chirality of the two tetrahedral bit states: $V > 0$ corresponds to right-handed tetrahedra and $V < 0$ to left-handed configurations. (c) Graphical illustration of the triple product defining the signed parallelepiped volume $V = \vec{AB} \cdot \left(\vec{AC} \times \vec{AD}\right)$.

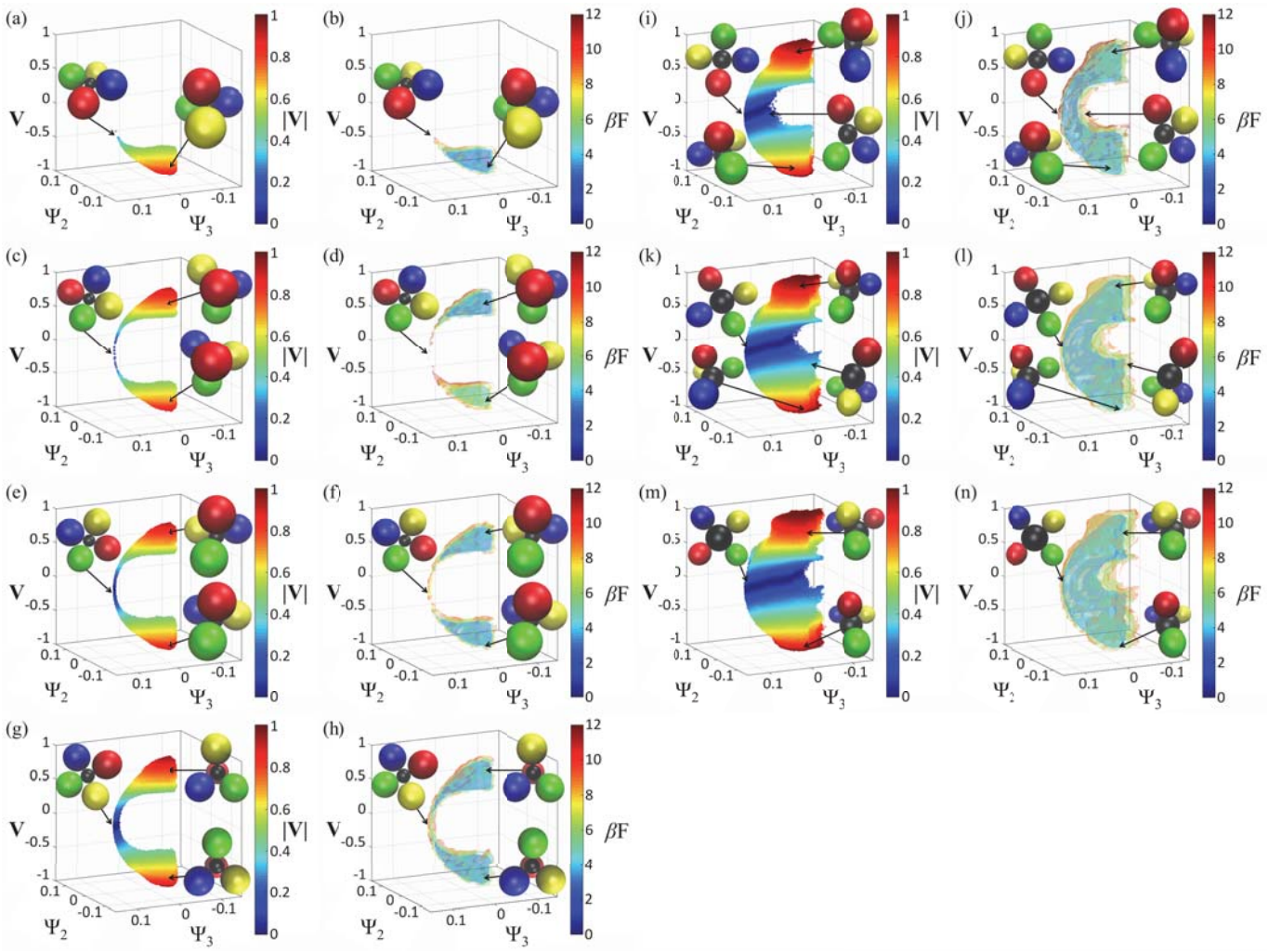


Fig. 3 Diffusion map embeddings and the free energy surfaces they support of the N=4 tetrahedral digital colloid as a function of Λ. Embeddings are constructed in (Ψ_2, Ψ_3, V) where (Ψ_2, Ψ_3) are the top two non-trivial eigenvectors discovered by diffusion map corresponding to the slowest collective modes of the digital colloid dynamical rearrangements and V is the signed volume of a parallelepiped connecting the halo particle centers of mass (Figure 2c). The sign of V distinguishes the chirality of the two tetrahedral bit states: $V > 0$ corresponds to right-handed tetrahedra and $V < 0$ to left-handed configurations. The magnitude of V provides a measure of the deviation of the halo particle configuration from planarity: $|V| = 0$ corresponds to all halo particles residing in a single plane, and $|V| = \sqrt{2}/2$ to an idealized tetrahedral arrangement. In each pair of panels we present on the left the low-dimensional projection of the 2×10^6 microstates harvested from the Brownian dynamics simulation at a particular value of Λ into the intrinsic manifold, and on the right the free energy surface over the intrinsic manifold $F(\Psi_2, \Psi_3, V)$ computed by application of Eqn. 4 employing rectilinear histogram bins of size $[\Delta\Psi_2, \Delta\Psi_3, \Delta V] = [0.01, 0.01, 0.05]$. (a,b) $\Lambda = 0.4000$, (c,d) $\Lambda = 0.4142$, (e,f) $\Lambda = 0.4400$, (g,h) $\Lambda = 0.5000$, (i,j) $\Lambda = 0.7500$, (k,l) $\Lambda = 1.0000$, (m,n) $\Lambda = 1.1256$. Points in the diffusion map embeddings are colored by $|V|$. Isosurfaces are plotted at $\beta F = \{0, 2, 4, 6, 8, 10, 12\}$, with the arbitrary zero of free energy defined by the most populated voxel of the embedding. Representative microstates have been projected over the embeddings to illustrate the halo particle configurations in each region of the intrinsic manifold.

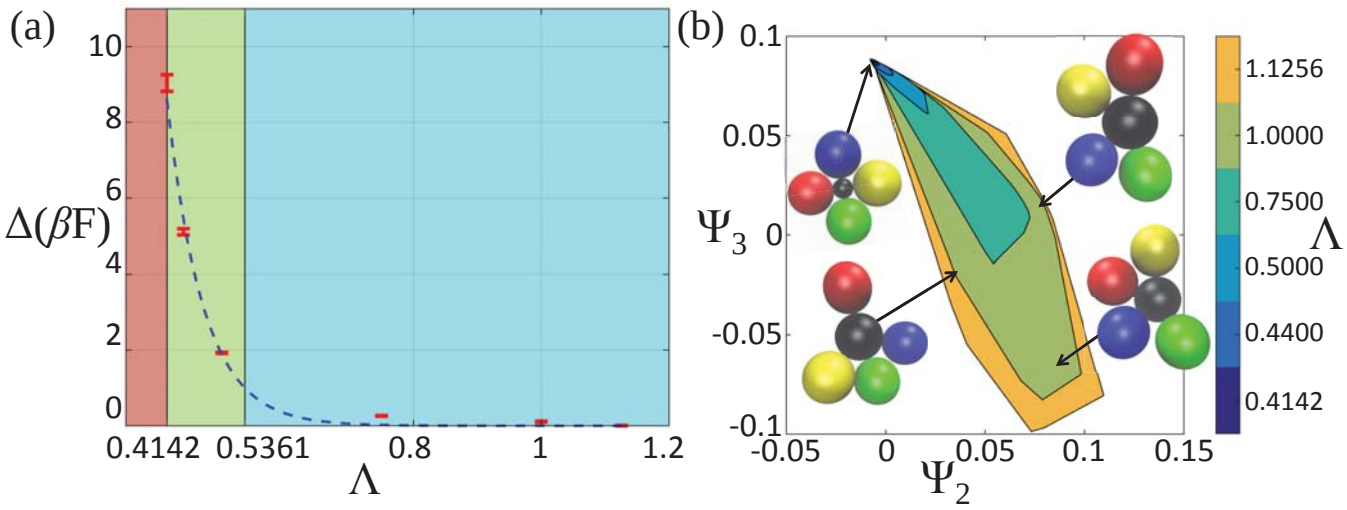


Fig. 4 Free energy barriers for bit state transitions and cross sections of the transition region from the N=4 free energy surfaces in Figure 3 as a function of Λ. (a) Free energy barrier height $\Delta(\beta F)$ computed as the free energy difference between the global minimum of the free energy surface and the minimum free energy within the transition state region defined by $V \in [-0.1, 0.1]$. Error bars correspond to 95% confidence bounds computed by 10 rounds of bootstrapping. The blue dashed-line corresponds to an exponential of the form $\Delta(\beta F) = C_0 \exp(-C_1 \Lambda)$ with parameters $\{C_0 = (1.29 \pm 0.89) \times 10^4, C_1 = (17.65 \pm 1.55)\}$ and an adjusted $R^2 = 0.98$. Best fit parameters were computed by weighted least squares regression in which we confronted the observed heteroskedasticity in the data by weighting each free energy measurement in inverse proportion to its variance[68]. The colored regions distinguish the different regimes of Λ: $(\Lambda_{SC} \approx 0.2247) \leq \Lambda < (\Lambda_T = 0.4142)$ (red, locked), $(\Lambda_T = 0.4142) \leq \Lambda < (\Lambda_U = 0.5361)$ (green, unlocked), and $\Lambda \geq (\Lambda_U = 0.5361)$ (blue, barrierless transitions). $\Lambda_T = 0.4142$ defines the unlocking transition at which there is sufficient free volume for the halo particles to transition between bit states through the square planar transition state. $\Lambda_U = 0.5361$ defines the transition to effectively unconstrained barrierless transitions between bit states where $\Delta(\beta F) = 1$. (b) Sections through the 3D free energy surfaces in $F(\Psi_2, \Psi_3, V)$ constructed from the subset of approximately planar halo particle configurations with $V \in [-0.1, 0.1]$. At $\Lambda_T = 0.4142$ the cluster is at the transition threshold and there exists an essentially unique square planar transition state. With increasing Λ the increased free volume available to the halo particles permits access to a greater diversity of planar configurations including trapezoidal and kite geometries. This increase in the phase space volume of the transition region leads to a commensurate decrease in the transition state free energy barrier due to reduced energetic and entropic barriers.

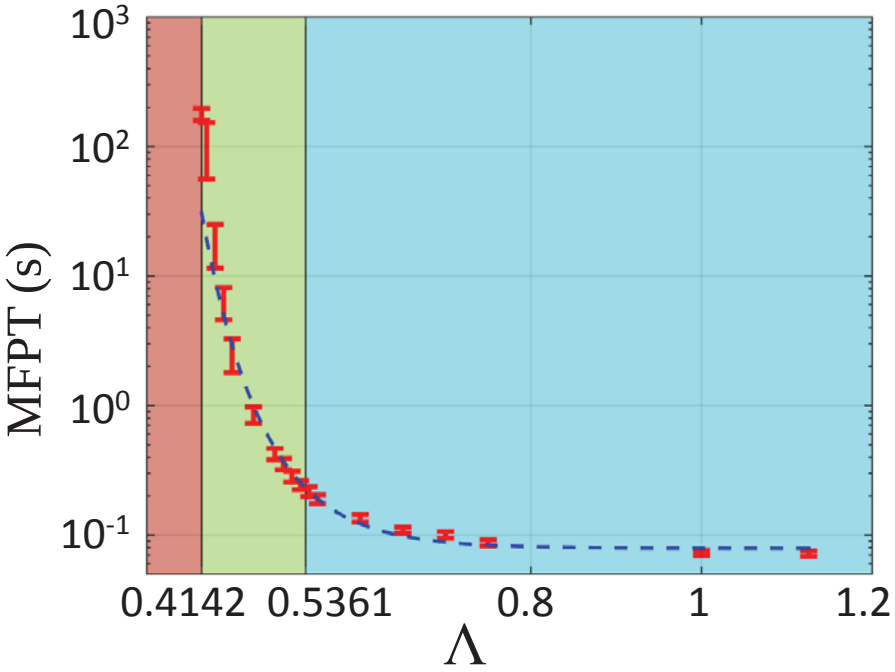


Fig. 5 Mean first passage times (MFPT) for the N=4 digital colloid as a function of Λ. Errorbars denote 95% confidence intervals on the mean value computed from the observed distribution of first passage times. The blue dashed-line corresponds to a shifted exponential of the form $\ln(\mathrm{MFPT}) = B_0 \exp(-B_1 \Lambda) + B_2$, with parameters $\{B_0 = (2.03 \pm 0.91) \times 10^3$, $B_1 = (14.08 \pm 0.92)$, $B_2 = (-2.53 \pm 0.04)\}$ and an adjusted $R^2 = 0.97$. Best fit parameters were computed by weighted least squares regression in which each MFPT measurement is weighted in inverse proportion to its variance to confront the observed heteroskedasticity in the data due to elevated uncertainties in the low-Λ regime where bit state transitions are rare events[68]. $\Gamma = 1/\mathrm{MFPT}$ defines the transition rate from one chiral bit state to the other.

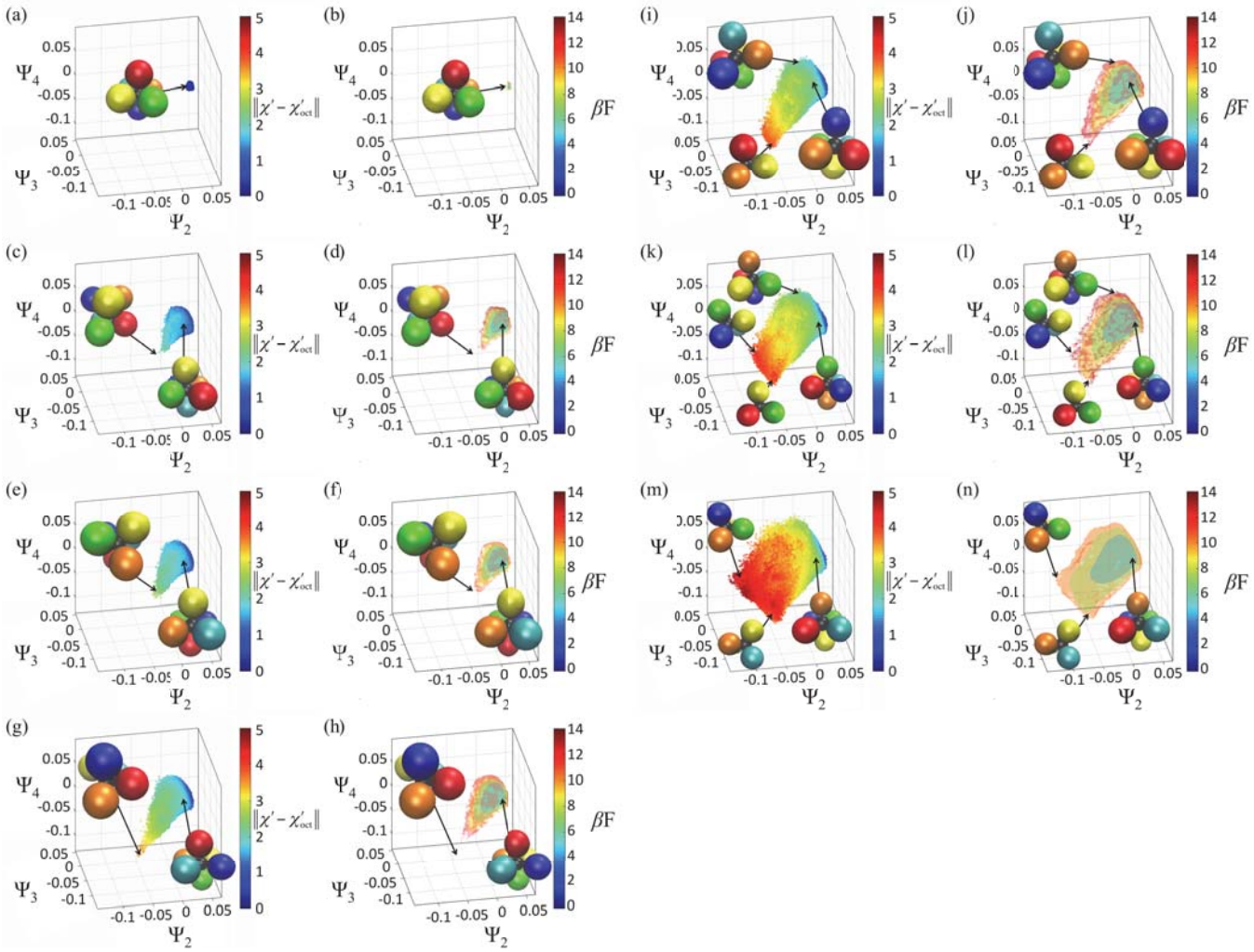


Fig. 6 Diffusion map embeddings and free energy surfaces for the N=6 octahedral digital colloid as a function of Λ. Embeddings are constructed in the top three non-trivial eigenvectors discovered by diffusion maps corresponding to the slowest collective modes of the digital colloid dynamical rearrangements (Ψ_2, Ψ_3, Ψ_4). In each pair of panels we present on the left the low-dimensional projection of the 2×10^6 microstates harvested from the Brownian dynamics simulation performed at a particular value of Λ into the intrinsic manifold, and on the right the free energy surface over the intrinsic manifold $F(\Psi_2, \Psi_3, \Psi_4)$ computed by application of Eqn. 4 employing rectilinear histogram bins of size $[\Delta\Psi_2, \Delta\Psi_3, \Delta\Psi_4] = [0.005, 0.005, 0.005]$. (a,b) $\Lambda = 0.4142$, (c,d) $\Lambda = 0.5100$, (e,f) $\Lambda = 0.5275$, (g,h) $\Lambda = 0.5500$, (i,j) $\Lambda = 0.5750$, (k,l) $\Lambda = 0.6000$, (m,n) $\Lambda = 0.6500$. Points in the diffusion map embedding are colored by the $L_{1,1}$ norm between the normalized halo particle pairwise distances matrix of the corresponding microstate and that of an idealized octahedron $||\chi' - \chi'_{\text{oct}}||$. Isosurfaces are plotted at $\beta F = \{0, 2, 4, 6, 8, 10, 12, 14\}$, with the arbitrary zero of free energy defined by the most populated voxel of the embedding. Representative microstates have been projected over the embeddings to illustrate the halo particle configurations in each region of the intrinsic manifold. The viewing angles of the triangular prism configurations within the transition state ensemble were selected to emphasize the stacking of the two triangular layers of halo particles such that those closer to the viewer eclipse from view those that are further away.

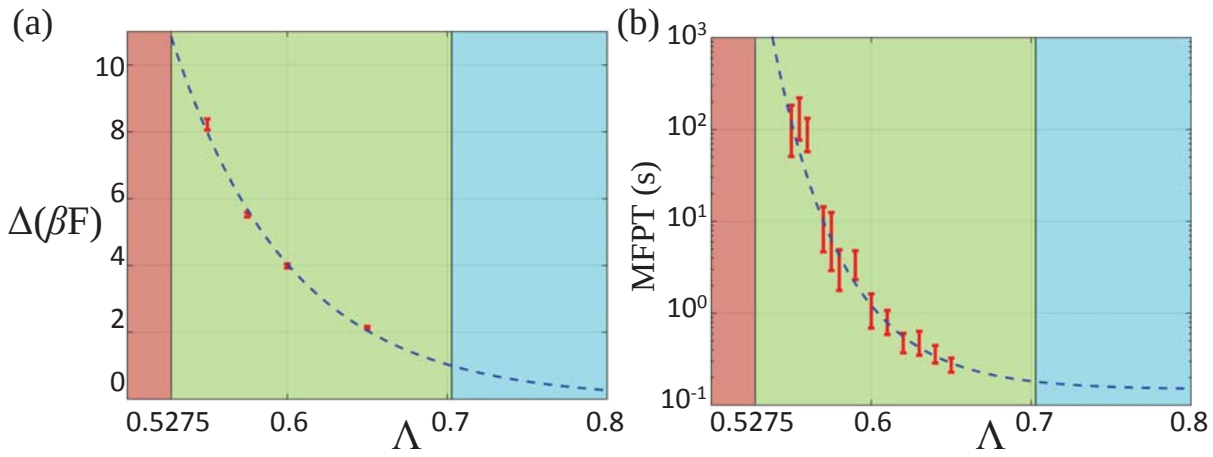


Fig. 7 Free energy barriers for bit state transitions and mean first passage time as a function of Λ for the N=6 digital colloid. (a) Free energy barrier height $\Delta(\beta F)$ computed as the free energy difference between the global minimum of the free energy surface and the minimum free energy within the transition state region defined by the voxels containing microstates with $||\chi' - \chi'_{\text{prism}}|| < 1.75$, where χ'_{prism} is the normalized halo pairwise distances matrix for the idealized triangular prism transition state structure. This threshold identifies microstates possessing an average 5% deviation away from the normalized distance matrix for the idealized triangular prism. Error bars correspond to 95% confidence bounds computed by 10 rounds of bootstrapping. The blue dashed-line corresponds to an exponential of the form $\Delta(\beta F) = C_0 \exp(-C_1 \Lambda)$ with parameters $\{C_0 = (1.41 \pm 0.58) \times 10^4, C_1 = (13.60 \pm 0.70)\}$ and an adjusted $R^2 = 0.99$. Best fit parameters were computed by weighted least squares regression in which we confronted the observed heteroskedasticity in the data by weighting each free energy measurement in inverse proportion to its variance[68]. The colored regions distinguish the different regimes of Λ: $(\Lambda_{\text{SC}} \approx 0.4142) \leq \Lambda < (\Lambda_{\text{T}} = 0.5275)$ (red, locked), $(\Lambda_{\text{T}} = 0.5275) \leq \Lambda < (\Lambda_{\text{U}} = 0.7021)$ (green, unlocked), and $\Lambda \geq (\Lambda_{\text{U}} = 0.7021)$ (blue, barrierless transitions). $\Lambda_{\text{T}} = 0.5275$ defines the unlocking transition at which there is sufficient free volume for the halo particles to transition between bit states through the square planar transition state. $\Lambda_{\text{U}} = 0.7021$ defines the (extrapolated) transition to effectively unconstrained barrierless transitions between bit states where $\Delta(\beta F) = 1$. (b) Mean first passage time computed as the average dwell time within a bit state before transitioning to a new bit state by a collective structural rearrangement of the halo particles. Errorbars denote 95% confidence intervals on the mean value computed from the observed distribution of first passage times. The blue dashed-line corresponds to a shifted exponential of the form $\ln(\text{MFPT}) = B_0 \exp(-B_1 \Lambda) + B_2$, with parameters $\{B_0 = (2.30 \pm 0.16) \times 10^6, B_1 = (23.17 \pm 0.12), B_2 = (-1.91 \pm 0.08)\}$ and an adjusted $R^2 = 0.93$ Best fit parameters were computed by weighted least squares regression in which each MFPT measurement is weighted in inverse proportion to its variance[68]. $\Gamma = 1/\text{MFPT}$ is the escape rate from any particular bit state to any one of eight other rotationally distinguishable bit states accessible by a single halo particle rearrangement event (Figure S1).

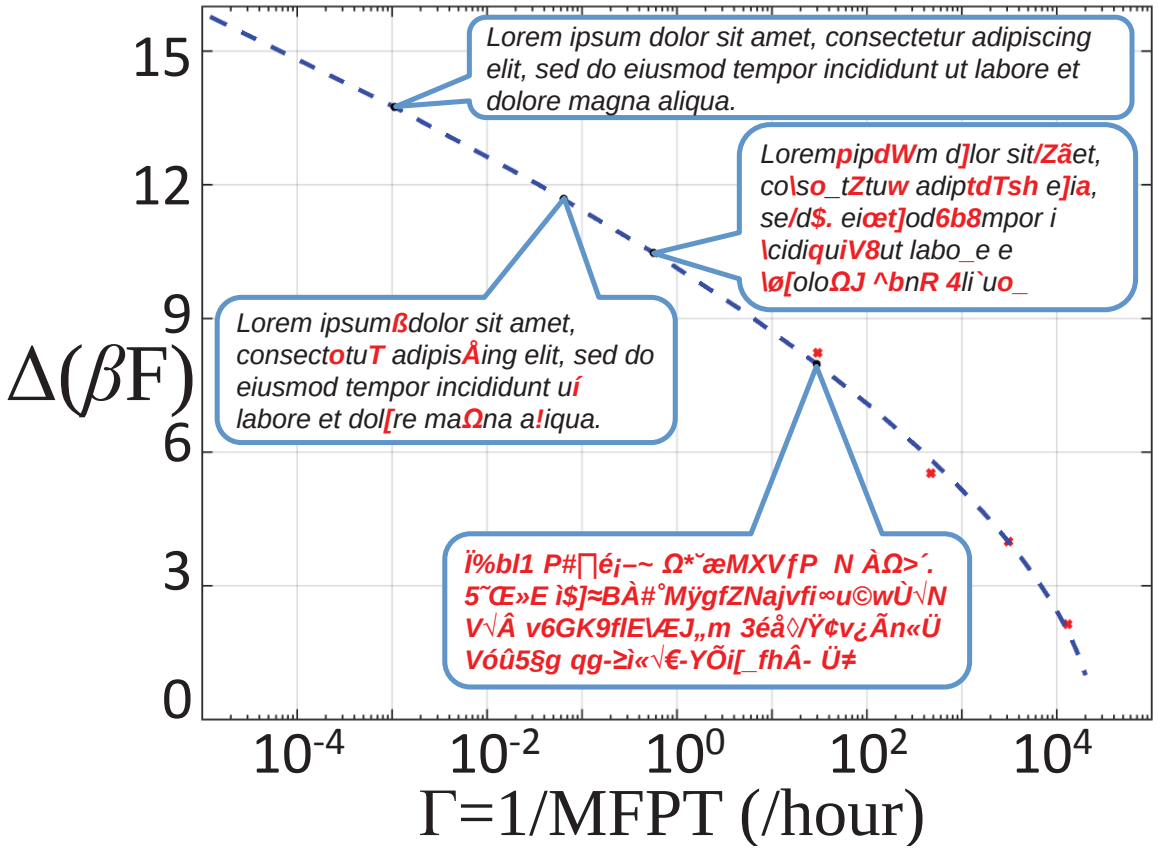


Fig. 8 Rational design of a N=6 digital colloid with tailored write free energy and memory volatility. Red points correspond to the calculated transition barrier heights $\Delta(\beta F)$ and transition rates $\Gamma = 1/\mathrm{MFPT}$ for the four N=6 digital colloid architectures with $\Lambda = \{0.5500, 0.5750, 0.6000, 0.6500\}$. The dashed blue line is a parametric fit of the best fit shifted exponential for the MFPT and the best fit exponential for $\Delta(\beta F)$ (cf. Figure 7) constructed by eliminating Λ to obtain an expression for $\Delta(\beta F)$ as a function of Γ. The ratio of the coefficients B_1 and C_1 of Λ in the two best fit exponentials controls the nonlinearity of the parametric fit. At a ratio of unity, the parametric fit corresponds to Arrhenius behavior and the dashed blue line would appear a straight line in the semi-logarithmic axes. The actual ratio is $B_1/C_1 = (1.70 \pm 0.09)$ leading to non-Arrhenius behavior and the observed non-linearity of the dashed blue curve. $\Delta(\beta F)$ is the reversible work required to actuate a halo particle rearrangement and write a new bit state into the digital colloid. Γ is the expected rate at which errors are introduced due to thermally-activated bit state transitions. The lower the write free energy, the higher the error rate. The design problem is to minimize the work required to write to the digital colloid elements while maintaining a the error rate above a particular error tolerance threshold. As a simple example, the overlaid text boxes illustrate the information storage fidelity of a short text block over a one hour duration. A system of 246 N=6 digital colloids would be required to encode this text using two digital colloids per 8-bit ASCII character. Red text is illustrative of the expected number of errors due to thermally-activated bit state transitions at the end of one hour of storage.

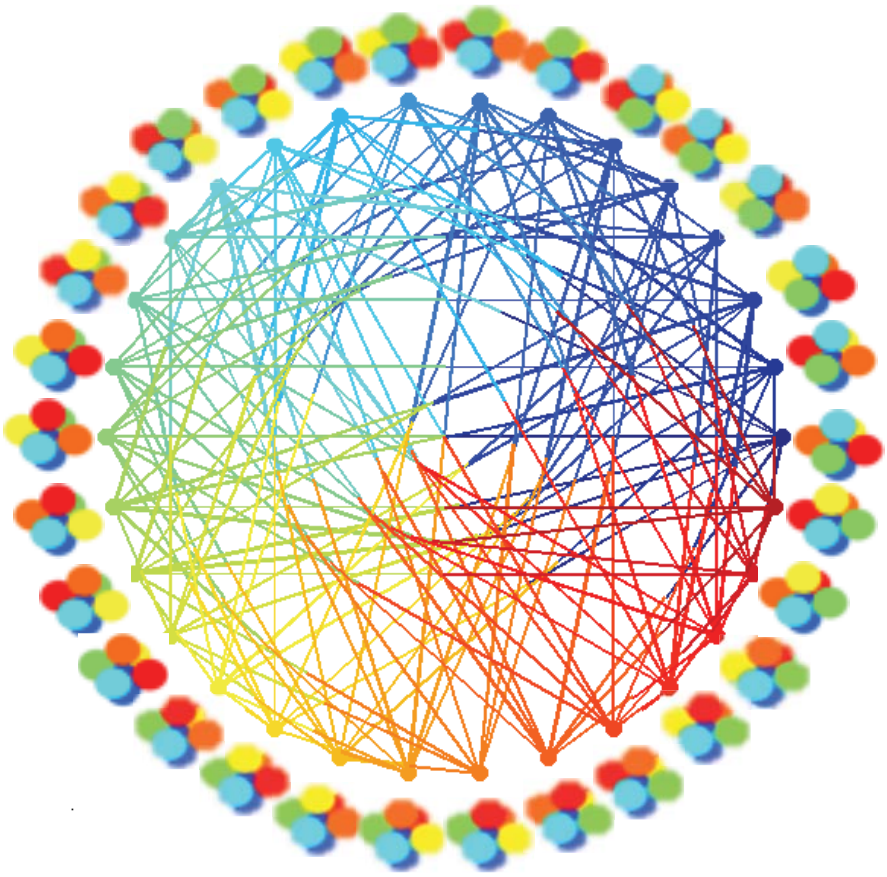
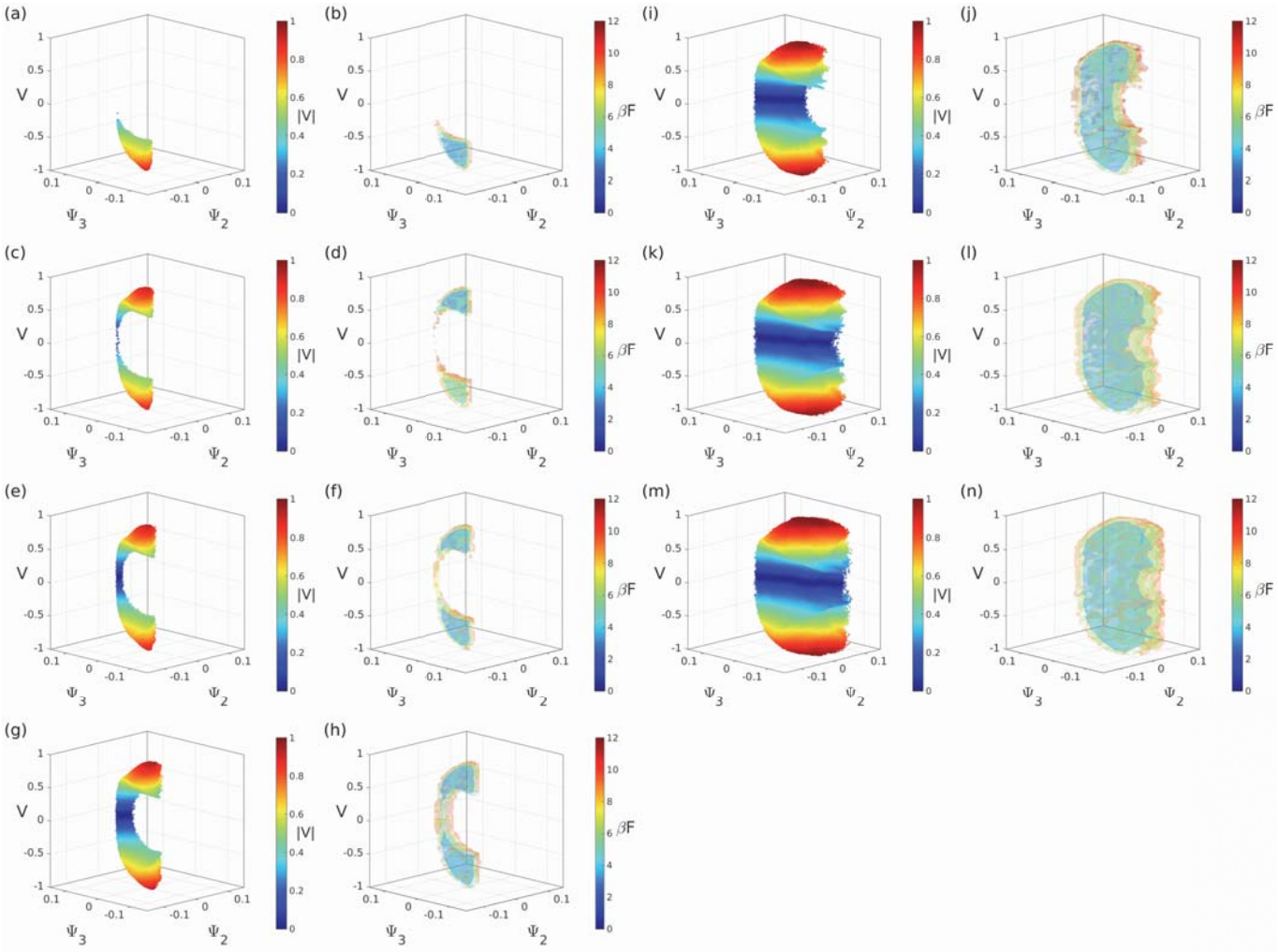
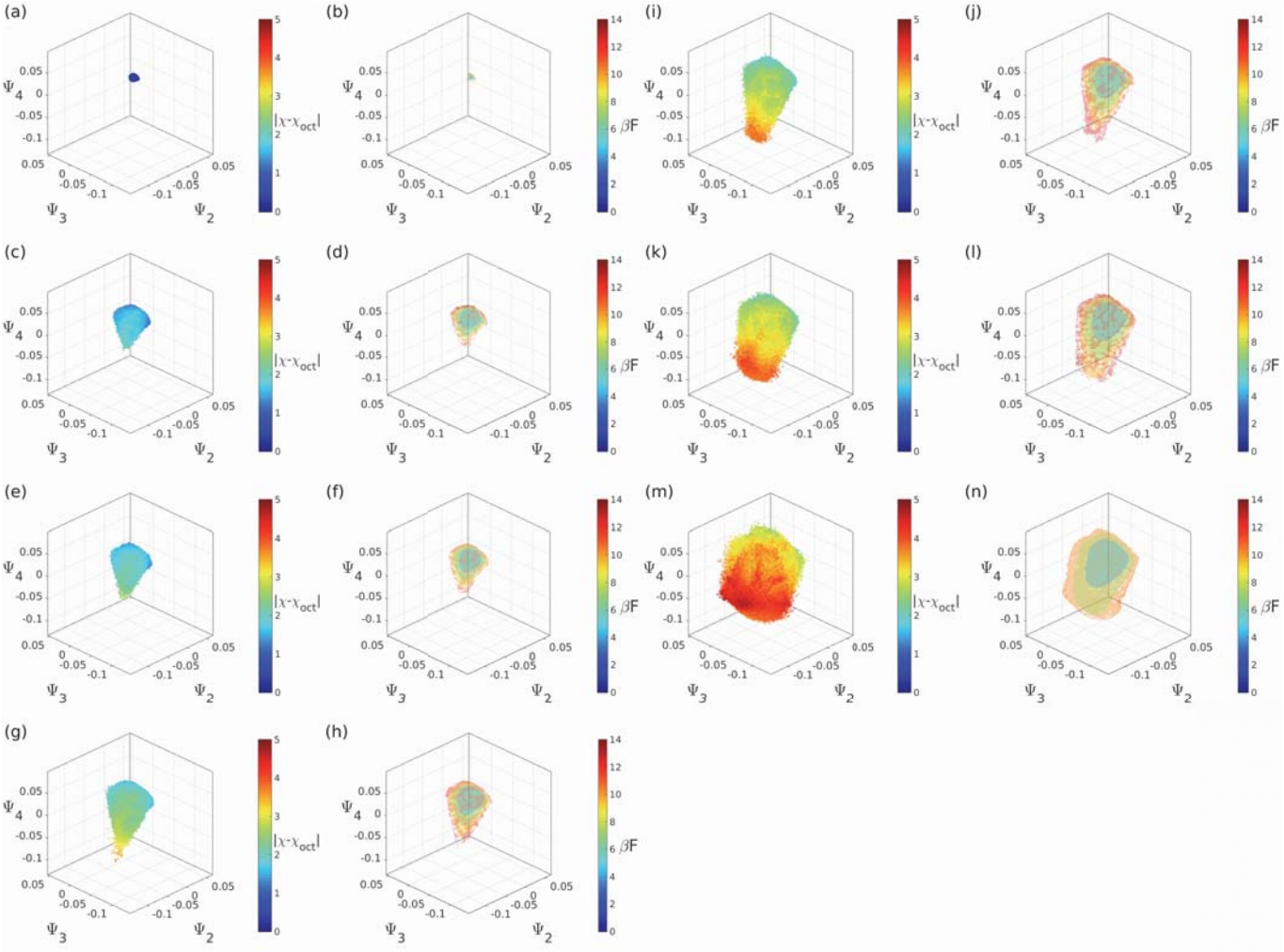


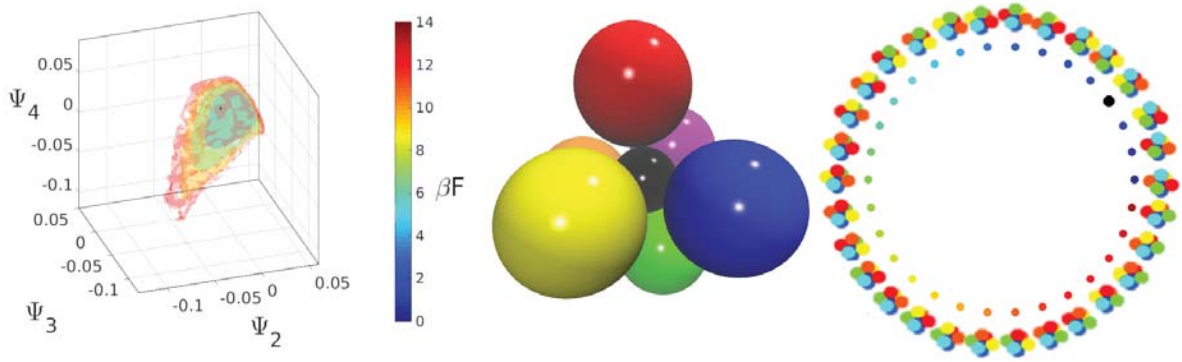
Fig. S1 Bit state transition map for the N=6 digital colloid illustrating the 30 rotationally distinguishable bit states as nodes around the exterior of graph and structural transition events linking bit states as edges through the interior. Nodes and edges are color coded to aid in interpretation. Each bit state is able to access eight others through a single halo particle collective transition event. The diameter of the transition network is such that a maximum of three transition events are required to transition between any pair of bit states.



Movie S1 Rotating movies of the diffusion map embeddings and free energy landscapes for the N=4 tetrahedral digital colloid given in Figure 3 of the main text. In each pair of panels we present on the left the low-dimensional projection of the 2×10^6 microstates harvested from the Brownian dynamics simulation at a particular value of Λ into the intrinsic manifold, and on the right the free energy surface over the intrinsic manifold $F(\Psi_2, \Psi_3, V)$. (a,b) $\Lambda = 0.4000$, (c,d) $\Lambda = 0.4142$, (e,f) $\Lambda = 0.4400$, (g,h) $\Lambda = 0.5000$, (i,j) $\Lambda = 0.7500$, (k,l) $\Lambda = 1.0000$, (m,n) $\Lambda = 1.1256$. Points in the diffusion map embeddings are colored by $|V|$. Isosurfaces are plotted at $\beta F = \{0, 2, 4, 6, 8, 10, 12\}$, with the arbitrary zero of free energy defined by the most populated voxel of the embedding.



Movie S2 Rotating movies of the diffusion map embeddings and free energy landscapes for the N=6 octahedral digital colloid given in Figure 6 of the main text. In each pair of panels we present on the left the low-dimensional projection of the 2×10^6 microstates harvested from the Brownian dynamics simulation performed at a particular value of Λ into the intrinsic manifold, and on the right the free energy surface over the intrinsic manifold $F(\Psi_2, \Psi_3, \Psi_4)$. (a,b) $\Lambda = 0.4142$, (c,d) $\Lambda = 0.5100$, (e,f) $\Lambda = 0.5275$, (g,h) $\Lambda = 0.5500$, (i,j) $\Lambda = 0.5750$, (k,l) $\Lambda = 0.6000$, (m,n) $\Lambda = 0.6500$. Points in the diffusion map embedding are colored by the $L_{1,1}$ norm between the normalized halo particle pairwise distances matrix of the corresponding microstate and that of an idealized octahedron $||\chi' - \chi'_{\text{oct}}||$. Isosurfaces are plotted at $\beta F = \{0, 2, 4, 6, 8, 10, 12, 14\}$, with the arbitrary zero of free energy defined by the most populated voxel of the embedding.



Movie S3 Tracking a N=6 octahedral digital colloid transition event through the intrinsic manifold and transition network at $\Lambda = 0.5750$. **(middle)** The digital colloid undergoes thermally-driven fluctuations that lead to a collective rearrangement of the halo particles from one of the 30 rotationally distinguishable octahedral bit states to another via a triangular prism transition state. **(left)** We track the structural evolution of the digital colloid as a black point moving over the low-dimensional intrinsic manifold discovered by diffusion maps previously presented in Figure 6j. The digital colloid initially resides in the broad octahedral basin until fluctuations lead to a cooperative transition event wherein the halo particles rearrange into the triangular prism transition structure that ultimately leads to a productive transition to a new rotationally distinguishable bit state. **(right)** The transition event can be visualized as a transformation from one of the 30 rotationally distinguishable octahedral bit states to another as illustrated by the pathway connecting the two nodes in the bit state network. Each bit state is able to access eight others through a single halo particle collective transition event (cf. Figure S1).