Cryptographic hashing using chaotic hydrodynamics

William Gilpin[a,1]

[a]Department of Applied Physics, Stanford University, Stanford, CA 94305

Edited by David A. Weitz, Harvard University, Cambridge, MA, and approved March 26, 2018 (received for review December 15, 2017)

Fluids may store and manipulate information, enabling complex applications ranging from digital logic gates to algorithmic self-assembly. While controllable hydrodynamic chaos has previously been observed in viscous fluids and harnessed for efficient mixing, its application to the manipulation of digital information has been sparsely investigated. We show that chaotic stirring of a viscous fluid naturally produces a characteristic signature of the stirring process in the arrangement of particles in the fluid, and that this signature directly satisfies the requirements for a cryptographic hash function. This includes strong divergence between similar stirring protocols' hashes and avoidance of collisions (identical hashes from distinct stirs), which are facilitated by noninvertibility and a broad chaotic attractor that samples many points in the fluid domain. The hashing ability of the chaotic fluidic map implicates several unexpected mechanisms, including incomplete mixing at short time scales that produces a hyperuniform hash distribution. We investigate the dynamics of hashing using interparticle winding statistics, and find that hashing starts with large-scale winding of kinetically disjoint regions of the chaotic attractor, which gradually gives way to smaller scale braiding of single-particle trajectories. In addition to providing a physically motivated approach to implementing and analyzing deterministic chaotic maps for cryptographic applications, we anticipate that our approach has applications in microfluidic proof-of-work systems and characterizing large-scale turbulent flows from sparse tracer data.

nonlinear dynamics | fluid dynamics | encryption | mixing | braiding

Recent experimental work has highlighted the ability of fluids to encode and store information (1, 2), motivating reciprocal inquiry into the role of computational rules in shaping the behavior of fluids in the natural world (3, 4). Such work has established applications in improving complex microfluidic devices and for characterizing large-scale complex flows using sparse data (5, 6), but it has broader implications for understanding constraints that shape active matter and self-assembly schemes (7). At the same time, studying logical operations and algorithmic performance in physical systems allows analytical tools borrowed from physics to be applied to traditional digital information systems, a line of inquiry that traces from Wheeler's original "it from bit" conjecture to Landauer's arguments about the role of physical representation on information processing (8, 9).

A potential new avenue for such inquiries is chaos, which has been widely investigated in digital applications due to the rich statistical structure it affords deterministic (and thus manipulable) dynamical systems (10). While hydrodynamic systems have been shown to exhibit chaos—both ubiquitously in turbulent flows (11, 12) but also unexpectedly in viscous flows via elegant analogies to classical dynamical nonintegrability (13)—the implications of chaos for digital fluid physics remain mostly unexplored.

Here, we exploit recent advances in the field of chaotic hydrodynamics to show how well-understood properties of chaotic maps can encode information about the underlying flow dynamics into the relative arrangements of advected particles. We show that this operation satisfies all of the properties of cryptographic hash functions that typically appear in digital security applications, including noninvertible compression of arbitrary inputs to fixed-length outputs, strong divergence between the hashes of similar inputs, and resistance to collisions between the hashes of two distinct inputs (14). We show that these unexpected properties arise naturally from the time scale-dependent dynamics of stirring a viscous liquid, implying potential new analysis techniques and applications at the interface of nonlinear dynamics and cryptography.

Model

Our cryptographic hashing scheme is based on chaotic advection at low Reynolds number. Given a time-varying flow and a small set of particles being advected, our hash consists of a short digest containing the relative ordering of the particles along one dimension. In order for the hashing scheme to be effective, this digest must be unique to the specific flow, but the original flow itself should not be easily computed from the hash—a property that naturally emerges in chaotic flows.

Under our approach, if the flow being studied has a known, finite set of governing parameters (such as jet speeds or stirring rates), then the time-dependent flow itself may be denoted by a discrete sequence of L vectors of parameter values $\boldsymbol{\sigma}$, which we refer to as a "stirring protocol" for the flow. The time step between parameter changes is arbitrary and may even be infinitesimal (corresponding to an analog signal); however, we assume that dissipation is large enough (and thus the Reynolds number and inertia are small enough) that the stirring protocol fully and invertibly specifies the dynamics of particles in the flow. We associate the specific stirring protocol $\boldsymbol{\sigma}$ with a "message" of length L that we wish to encrypt.

We then specify M labeled particles at known initial positions and allow the flow to advect these particles for L time steps with the step-wise parameters specified by $\boldsymbol{\sigma}$. The final arrangement of these particles is discretized by recording their

Significance

An essential component of digital communication is hashing, in which a complex piece of information (a document, video, etc.) is mathematically transformed into a unique signature that can later be used to identify the original piece of data. Here, we show that this process bears strong similarity to the chaotic behavior of certain types of flows observed when ordinary fluids mix, such as the stirring of dye into water. We use this analogy between rearranging information and stirring a fluid to construct a fluid-based hash function with comparable properties to traditional algorithms. Our work bears direct relevance to cases in which the physical representation of information affects its transmission, including in microfluidic self-assembly schemes and characterizing complex natural flows.

Author contributions: W.G. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission.

[1] Email: wgilpin@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1721852115/-/DCSupplemental.

ordering, $\boldsymbol{\mu}$, along the x axis, which is taken as the hash of the message $\boldsymbol{\sigma}$. This last step ensures that the hash $\boldsymbol{\mu}$ is noninvertible, since it discards information about the final coordinates of the particles and thus ensures that knowledge of the initial conditions and the hash is insufficient to determine the stirring process, a necessary condition for a hash function. Our approach thus differs from standard block-cipher approaches to discrete chaotic encryption in that the original message is encoded in a time-varying set of map parameters rather than in the initial conditions of the particles (15, 16). Instead, our approach generalizes stream ciphers that encode information in the parameters of a chaotic map (15); however, the continuous-domain hydrodynamic map allows the message bandwidth to be made arbitrarily high using an arbitrarily long stirring sequence. The latter property satisfies the requirement that a hash function maps arbitrary-length inputs to fixed-length outputs (14).

Our approach is illustrated in Fig. 1, in which we demonstrate our hashing scheme using the classical "blinking vortex" flow, a canonical and early chaotic fluidic map that has been shown both theoretically and experimentally to demonstrate strong chaos over a wide range of parameter values (6, 13, 17, 18). This map consists of a unit circular domain with no-slip boundary conditions, which contains two vortices at positions $b = \pm 0.5$. At every fixed time interval T, one vortex shuts off and the other turns on, causing particles in the domain to trace circular arcs around the vortex center until another interval T has elapsed. T thus controls the degree of overall chaos in the flow, and when $T > 1$, the map is strongly chaotic with an attractor occupying nearly all of the map's domain—making the flow a canonical example of mixing in the absence of inertia. We use the strongly chaotic value $T = 1.5$ throughout the text, except in a figure below where we specifically vary T to change the strength of chaos in the system. The dynamical equations for the map are given explicitly in *SI Appendix*, section 1.

The blinking vortex system has several properties that render it particularly useful for hashing. The two-vortex map has analytic, circular particle trajectories, which do not require numerical integration to compute and thus retain numerical precision after many iterations of the system. Moreover, the two-vortex flow is isomorphic to the widely studied linked twist map, and so it exhibits phenomena such as domain-wide streamline crossing that are broadly applicable to many other classes of chaotic maps (17).

Here, we exploit the fact that the only relevant parameter describing the map at a given time step is which of the two vortices is currently "on," which causes the message vector $\boldsymbol{\sigma}$ to reduce to a binary encoding of the message information, with the message $\boldsymbol{\sigma} = 10101010$ representing $L = 8$ iterations of the classical two-vortex flow in which the vortices always alternate. We apply each message/stirring protocol to the same set of initial conditions for all messages; these initial conditions are public and arbitrary, and throughout this study, we assign the M hash particles initial positions in the circular domain using a 2D Fibonacci sunflower spiral, which provides a deterministic, quasi-uniform initial packing that maximizes the initial interparticle displacements and thus improves mixing. We contract the radius of the sunflower by an empirical increment to keep particles away from the boundary (which inhibits transport). Unlike previous chaotic hashing protocols (15, 16), these initial conditions are not secret; in fact, they could in principle be customized by a rotation angle and then used as a "public key" in an asymmetric encryption system.

While we use the quasi-uniform sunflower packing for different Ms throughout this work, we note that a more optimal set of initial conditions would take into account information about the specific chaotic map's dynamics when assigning initial particle locations, to maximize mixing. For example, the distribution of principal eigenvalues associated with local stretching of the flow could be used to weigh initial locations; more generally, the set of maximal finite-time Lyapunov exponents corresponding to the message length L could be used to maximize mixing over that specific time scale (19). Conveniently, the blinking vortex flow already has a quasi-uniform spatial stretching distribution for $L \approx 10$, and so we expect the quasi-uniform packing used here to be sufficient (*SI Appendix*, section 9). However, for more complex empirical mixing systems arising in experimental applications, preanalysis of mixing regions may be necessary to assign appropriate initial conditions.

We denote the hash associated with a message $\boldsymbol{\sigma}$ as $\boldsymbol{\mu}$, which represents an ordered set of integers corresponding to the final locations of the stirred hash particles, ranked by their relative positions along the x axis. For example, for a hash of length $M = 5$, five particles are initially labeled based on the relative x coordinates of their initial positions so that $\boldsymbol{\mu}_0 = 12345$. After L iterations of the chaotic map, particles will have traded relative positions along the x axis, leading to a final hash $\boldsymbol{\mu}_L = 34512$. In cryptography, the size of the hash space determines the difficulty that an attacker will have computing all possible hashes–a brute force approach that would allow incorrect information to masquerade as correct in an application such as password storage. Thus, the size of the hash (as measured in bits) provides an upper bound on the algorithm's security; for example, the frequently used digital security function SHA-256 has a hash space of size $2^{256} \sim 10^{77}$. Because the hash space in our system grows combinatorially ($M!$) with the number of particles M, in principle our system achieves comparable security when $M = 58$ particles ($58! \sim 10^{78}$); however, in fluidic implementations, M or L may be limited by diffusivity, precision loss, or other irreversible processes (*SI Appendix*, section 6).

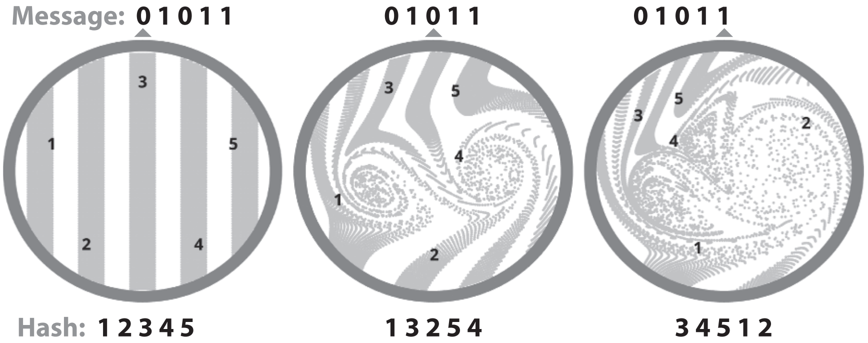

Fig. 1. Message hashing using a chaotic stirring protocol. The hashing procedure is shown at three different time points, including the known initial conditions and the final configuration and hash. The labeled particles used to compute the hash are indicated with numeric labels; additional points are underlain in gray to illustrate the chaotic mixing patterns characteristic of the two-vortex "eggbeater" flow.

Results

Having defined our system, we next seek to characterize how well it performs in generating uncorrelated hashes $\boldsymbol{\mu}, \boldsymbol{\mu}'$ from correlated messages $\boldsymbol{\sigma}, \boldsymbol{\sigma}'$. We define a variation of the permutation Lyapunov exponent originally proposed by Amigó et al. (16),

$$\lambda_D = \left\langle \log_2 \left(\frac{d(\boldsymbol{\mu}, \boldsymbol{\mu}')}{d(\boldsymbol{\sigma}, \boldsymbol{\sigma}')} \right) \right\rangle_{\boldsymbol{\sigma}, \boldsymbol{\sigma}'}, \qquad \textbf{[1]}$$

where $d(.,\,.)$ represents an appropriate distance function. Here, we use the edit distance between sets, which counts the minimum number of reorderings, swaps, or deletions necessary to transform one ordered list of symbols into another. When applied to binary messages, this corresponds to a generalization of the Hamming distance that includes insertions and deletions, and when applied to the hash (an ordered set of distinct symbols), it corresponds to the minimal permutation distance. Due to the boundedness of the chaotic domain, in practice the denominator $d(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ in Eq. **1** should be kept as small as possible, due to system-size saturation of the potential interparticle distances after applying the map.

We generate an ensemble of one million random binary messages, each with a distinct length L, hash length M, and mixing period T. For each message $\boldsymbol{\sigma}$, a second message is randomly generated via mutation, insertion, or deletion of a random bit, to generate a dual message $\boldsymbol{\sigma}'$ exactly one edit distance away from the original $[d(\boldsymbol{\sigma}, \boldsymbol{\sigma}') = 1]$. The distance between each pair's hashes $d(\boldsymbol{\mu}, \boldsymbol{\mu}')$ is then computed to generate a parameter-dependent estimate of the permutation Lyapunov exponent $\lambda(M, L, T)$.

Fig. 2*A* shows λ_D as a function of L and T. In general, increasing T (and thus the true dynamical Lyapunov exponent associated with the map) leads to larger values of λ_D due to stronger mixing of particles within the chaotic flow. This is consistent with recent results for discrete, pseudochaotic maps that suggest asymptotic equivalence between the discrete Lyapunov exponent and true Lyapunov exponent in the limit of high cardinality (and thus M) (20). Surprisingly, however, increasing the message length L has a nearly identical effect to increasing T, suggesting that a more chaotic (albeit computationally demanding) map may compensate for redundant or short messages. For all conditions tested, λ_D approaches an asymptotic maximum value 5.25 comparable to that of traditional cryptographic hashing functions based on bitwise operations (16). That increasing L and T have similar effects suggests that the primary determinant of the mixing is the total time that the particles are stirred: Shorter blinking cycles T may be compensated by using more cycles L. Thus, it is intuitive that the strongest mixing and chaotic behavior (and thus most effective hashing) occurs when the relative frequencies of 0 and 1 in the message are equal; otherwise, long, multicycle runs of a single symbol occur and undermine chaotic advection—for example, the message 000000... would generate a flow with no chaotic advection at all. For encoding specific information into a flow (for example, binary representations of English characters), achieving equal symbol distribution is a matter of choosing a highly compressed binary representation or padding to the message with fixed-length string of alternating bits.

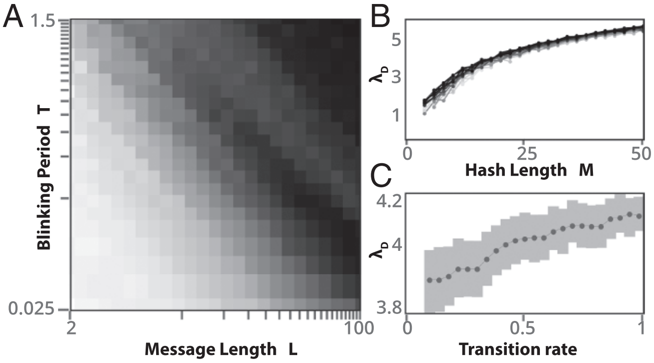

Fig. 2. Chaos spreads hashes. (*A*) The average permutation Lyapunov exponent λ_D as a function of the chaotic parameter of the underlying map T and the message length L. The color map runs between 0 (white) and 8 (black). (*B*) Mean λ_D as a function of the hash size M, for message lengths L running from 10 (light gray) to 100 (black). (*C*) Mean λ_D with SE bars for ensembles of messages with different switching rates. The rightmost edge corresponds to continuous alternation between the stirrers (10101010…), as in the standard blinking vortex flow, and the leftmost edge consists of a single transition (00…01…11). For this figure, averages were computed from an ensemble of 10^6 message pairs. For *B* and *C*, $T = 1.5$; for *C*, $M = 20$ and $L = 30$.

However, while it is intuitive that the relative symbol ratio in the stirring protocol (here, the ratio of 1s and 0s) should be nearly 1, we also observe (Fig. 2*C*) that more switches between the symbols generally increase λ_D: 101010 has a larger λ_D than 111000. This parameter is an invariant of the information type being encrypted (21), and it suggests that the distribution of "runs" of consecutive identical symbols affects the hashing process. Overall, we conclude that frequent and uniform switches between symbols in the message, long messages, and a strongly chaotic mapping function all lead to less predictable hashing in which the hashes of two similar messages are unrelated—jointly endowing our hashing protocol with "preimage" resistance, a fundamental requirement for cryptography (14).

While the final projection of the coordinates onto the *x* axis ensures that the deterministic hashing procedure is noninvertible, a stirring-based hashing scheme may be subject to a second preimage or "birthday" attack, wherein two random inputs produce the same hash, resulting in a "collision" that renders the original inputs indistinguishable. To investigate the frequency of hash collisions under various conditions, we compute an ensemble of messages and determine how M and L affect the collision density ϕ_M, defined as

$$\phi_M \equiv 1 - \frac{N_u}{N_s}, \qquad \textbf{[2]}$$

where N_u is the number of unique hashes found, and N_s is the total number searched. Because an M particle hash has $M!$ possible values, we expect a transition between most hashes being unique ($\phi_M \sim 0$) and most hashes being duplicates ($\phi_M \sim 1$) to occur when $N_s \approx M!$. Fig. 3*A* shows ϕ_M for several different values of M and L (black lines), illustrating a sharp sigmoidal transition that occurs at $N_s \sim M!$ across many values of L. However, at small L, the map undergoes too few iterations to satisfactorily sample the chaotic attractor, leading to a lower bound on ϕ that decreases to zero as L grows.

In addition to avoiding collisions, a strong cryptographic hashing scheme distributes hashes uniformly across the space of possible values. If distributions of collision locations are not uniform, then neighboring hashes are correlated and a single collision will compromise many messages. We assess hash uniformity by generating N_s and then tabulating the total number of identical pairs, triples, etc. present in the sample. We then order the N_s hashes by their rarity and compute a running sum of the total number of unique hashes contained within a given $n < N_s$ of the observed hashes. The resulting hash coverage plots are shown in Fig. 3*B* for various values of M and N_s. For small numbers of computed hashes, the fraction of unique hashes grows linearly with the sample size because $N_u = N_s$, and the duration of the linear growth increases with M. However, as more hashes are sampled, duplicates begin to appear, causing

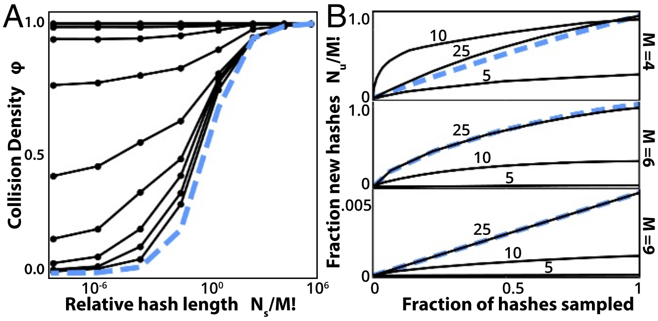

Fig. 3. Collision probability and uniformity of hash values. (*A*) Mean hash collision density ϕ_M as the hash size M is varied, for message lengths $L = 2, 4, 6, \ldots, 20$ (black traces). Traces from top to bottom correspond to the smallest to largest message lengths, and the topmost curves for $L = 2$ and $L = 4$ overlap. The hash space is rescaled by the number of hashes sampled, N_s, and the analytic solution for ϕ_M under uniform random sampling is underlain (blue dashed trace). (*B*) Cumulative fraction of new hashes found as a function of total hashes computed, normalized by total possible unique hashes ($M!$) (black traces), for several different message lengths ($L = 5$, 10, 25, annotated). Underlain on each subplot is the analytic result for random sampling (blue dashed trace). All subpanels contain a fixed number of samples $N_s = 2000$, leading to a linear trend in the lowest subpanel where $N_s \ll M!$.

sublinear growth and gradual saturation of the fraction of unique hashes observed.

We compare these curves to our expectation for a "perfect" hash function, which we assume would uniformly randomly sample (with replacement) the $M!$ elements of hash space when given random messages as inputs. In this case, whether a given hash is new or a previously observed value is given by a binomial process. If N_s hashes are randomly sampled (with replacement) from this set, then the probability of drawing exactly U unique hashes is given by

$$P_U = \frac{M!!}{M!^{N_s}(M!-U)!}\left(\frac{1}{U}\sum_{q=0}^{U}(-1)^{U-q}\binom{U}{q}q^{N_s}\right), \tag{3}$$

where the parenthetical term is the explicit formula for the Stirling number of the second kind $S(N_s, U)$, which is a summation over the number of singly drawn hashes (q = 1), pairs (q = 2), etc. for a binomial (uncorrelated) hash selection process. The cumulative distribution over these pairs, $\sum_U P_U$, is underlain as dashed lines in Fig. 3*B*, and it is apparent that, at large message lengths L and hash sizes M, the chaotic hash function approaches this upper bound in performance. Calculation of the expectation value of Eq. **3**, $\sum_U UP_U$, produces the expected number of unique hashes N_u from a uniform sample of size N_s from $M!$ possible hashes,

$$N_u = M!\left(1-\left(1-\frac{1}{M!}\right)^{N_s}\right). \tag{4}$$

Inserting this equation into Eq. **2** provides a null estimate of the collision density, underlain as a dashed line in Fig. 3*A*. As expected, this function is strongly sigmoidal, predicting a sharp transition from $\phi = 1$ to $\phi = 0$ as $N_s \sim M!$.

The analytic estimates provided by Eqs. **3** and **4** illustrate a surprising property of chaotic hashing: for small M and L, the hash function initially encounters fewer collisions than would be expected if it uniformly sampled the hash space. This effect is apparent as the concave transients above the blue trace in Fig. 3*B*. This anomalous scaling is consistent with hyperuniformity (22), a phenomenon in fluids and granular media in which long-wavelength density fluctuations are suppressed (22–24). In our system, its occurrence at small M and L implies a dynamical transient wherein the particles retain information of their initial conditions: The particles have an initial separation $\sim 1/\sqrt{M}$, which typically takes a certain number of time steps L to traverse. As a result, the time scale of hyperuniformity depends on the particular value of M; at very small M in Fig. 3, the sampling of new hashes is also bounded from below due to a secondary time scale in which particles travel too little to ever exchange ordering, leading to a plateau in the number of distinct hashes found. Hyperuniformity thus occurs at intermediate L when particles travel for enough steps to exchange positions but not to fully explore the chaotic attractor.

This apparent oversampling of new hashes at relatively short message lengths introduces a tradeoff: Hyperuniformity improves the hash function's tolerance to random collisions, but it undermines the hash function's cryptographic security because an attacker intending to infer a message's hash from that of a similar message could exploit the finite set of locations sampled by a given particle over small L. Thus, for general process validation applications, hyperuniformity at small L and M may be desirable; however, applications with strongly correlated messages would require cryptographic hash security. These limits represent a form of tradeoff between fault tolerance (random collisions) and attack tolerance (collisions informed by prior information about the dynamics), a dichotomy previously observed in random networks (25).

We next investigate the origin of this hyperuniform transient, which we suspect arises from short-time mixing dynamics due to stretching associated with subregions in the fluid's chaotic attractor that kinetically isolates distinct hash particles. To identify the signature of these dynamics, we study the short time pairwise winding statistics of particle trajectories, which recent work on topological chaos and braiding has shown to represent a characteristic invariant of the dynamical system (6, 26). Using the notation of Caussin and Bartolo (27), we define the pairwise linking number $w_{ij}(L)$ as the total amount that two particles rotate around each other during L iterations of the map,

$$w_{ij}(L) = \frac{1}{2}\sum_{a}^{L}\varepsilon_a^{ij}.$$

Here, a indexes all cases in which particles exchange relative positions along the x axis. $\varepsilon_a^{ij} = -1$ if particle i crosses above particle j during the a^{th} intersection; conversely, $\varepsilon_a^{ij} = 1$ if particle j is above i. The total winding across an M particle hash is given by $W(L) = \sum_{(i,j)} w_{ij}(L)$, and it may be taken as characteristic global property of the mixing process that produced the hash: Linear growth of $W(L)$ (continuous winding) may be induced by static vortices, whereas sublinear growth is indicative of dynamical effects like random winding and unwinding of particles' worldlines (27, 28).

Fig. 4*A* shows the root-mean-square of the total winding $\sqrt{\langle W(L)^2\rangle}$ for several different values of M, with the average taken over an ensemble of 10^6 random messages. The total winding shows a clear transition from ballistic $[W(L) \sim L]$ to diffusive $[W(L) \sim L^{1/2}]$ scaling at $L \approx 10$, with the transition point only weakly varying with M (see Fig. 4*A*, *Inset*). This suggests that, over short times, mixing in braid space is governed primarily by convergent behavior of hash particles with nearby initial conditions, which explore the chaotic attractor locally but have insufficient time to wander the full domain. These transient dynamical barriers partition the chaotic attractor into subsets of trajectories, which wind together as clusters over short time scales that gradually shrink as particles have more time to explore the attractor and intercalate their worldlines' windings. This transition is illustrated in Fig. 4*C*, which shows for a

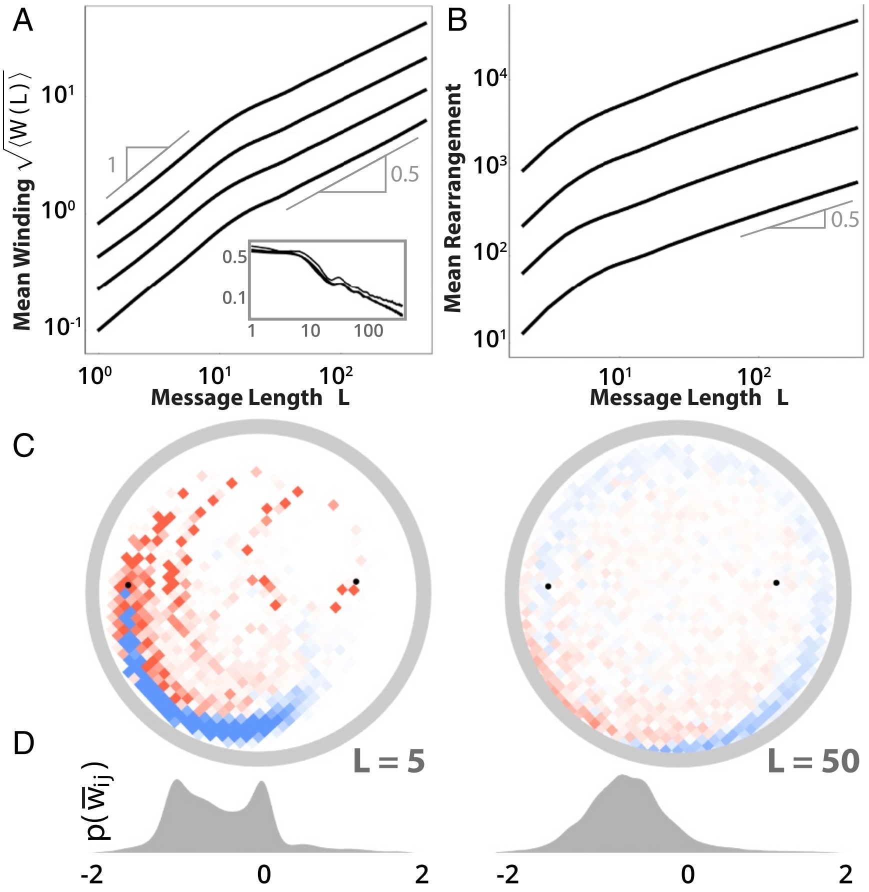

Fig. 4. Short-time particle braiding disperses trajectories across hash space. (*A*) The root-mean-square winding index, $\sqrt{\langle W(L)^2\rangle}$, over time for different hash lengths M (traces from topmost to bottommost correspond to $M = 10, 20, 40, 80$). *Inset* shows the relative slope, as measured by the normalized differences $(\sqrt{\langle W(L+1)^2\rangle} - \sqrt{\langle W(L)^2\rangle})/\sqrt{\langle W(1)^2\rangle}$. (*B*) The mean rearrangement index as a function of the same values of M and L given in the first panel. (*C*) For two hash particles (initial locations marked by black dots), the relative frequencies that each one visits different points in the domain over short (*Left*) and long (*Right*) time scales. White regions are visited equally by both points. (*D*) The distribution of message-averaged pairwise linking numbers, $p(\bar{w}_{ij}(L))$, across all particle pairs, at the same short and long time scales as *C*.

single pair of hash points the relative density of their future positions in the domain over short and long time scales (Fig. 4*C*, *Left* and *Right*, respectively). Red (blue) regions correspond to those occupied by potential trajectories originating from the left (right) hash point, and white regions correspond to regions that trajectories for both particles explore to equal degrees (see *SI Appendix*, section 3 for more detail). Clustering indicates that over short time scales the initial positions strongly determine the regions of the attractor explored by a given hash particles, but over longer time scales, both particles sample domain relatively equally. Over short time scales, the large-scale winding of these clusters of kinetically accessible positions dominates the total winding, apparent as a strongly bimodal distribution of ensemble-averaged linking numbers across all possible particle pairs $p(\bar{w}_{ij}(L))$ (Fig. 4*D*, *Left*). Once the two point clouds fully intercalate, the average linking distribution becomes Gaussian (Fig. 4*D*, *Right*) and further growth in total winding becomes diffusive. For both time scales, the mean linkage is less than zero: because both vortices have positive rotation direction, particles over time tend to travel anticlockwise around the domain, inducing a global net twisting of particle worldlines that produces positive mean winding (27).

Analysis of the hashing process confirms the role that chaos plays in hyperuniformly randomizing the hash over short time scales: In early steps of hashing, large-scale transpositions of groups of neighboring particles within the hash (e.g., $\boldsymbol{\mu}_0 = 12345678 \rightarrow \boldsymbol{\mu}_1 = 56781234$) tend to occur much more frequently than would be expected for random rearrangements. This may be quantified by defining a "rearrangement index" $R(L)$ that quantifies how many initially nearest neighbors cease to be neighbors under L iterations of the map (see *SI Appendix*, section 4 for details). For all tested values of M, the root-mean-square growth of $R(L)$ shows similar dynamics to the winding, with an initial transient period lasting ~ 10 iterations in which nearest neighbors stay together (Fig. 4*B*). The two regimes of mixing and their effect on the hash function are thus analogous to initially cutting a deck of playing cards several times and then gradually riffling the halves of the deck back together.

Discussion

In this article, we have presented a physically motivated chaotic cryptography system implemented using a widely studied, canonical hydrodynamical system. The two-vortex map we study explicitly describes fluid flow at low Reynolds number, allowing interpretation of the properties of the hash function in terms of familiar Lagrangian concepts from dynamical systems such as braiding, particle dispersion, and mixing. Our work suggests a role for tools from dynamical systems theory in constructing and characterizing cryptographic functions as well as potential utility for cryptographic concepts (e.g., collision tolerance and attack resistance) in understanding hydrodynamical systems. The design of modern digital hash functions for computer security is an active area of research (29, 30), and by demonstrating a simple model in which hashing arises from mathematically characterizable chaos, we anticipate our work could inform future efforts to design nonheuristic hash functions from first principles based in dynamical systems theory (15, 16).

However, we anticipate that the most direct application of our work arises in microfluidics and mixing theory, particularly in process verification for particle self-assembly (7, 27) and microfluidic proof-of-work systems (31). Such microscale applications raise the question of whether, in practice, chaos-based hashes are too fragile to external perturbations or irreversible interactions to guarantee repeatability (32). In *SI Appendix*, section 6, we perform extensive characterization of how our hashing scheme's reproducibility is affected by random noise—either arising from numerical precision loss in digital implementations or from diffusion of particles across streamlines in a physical system. We parametrize the relative strength of noise using the dimensionless Péclet number Pe, which specifies the relative strength of advection relative to diffusive noise in the system: A higher Pe corresponds to larger-scale or faster-stirred physical systems, or computational settings with low error rates per operation. We find general repeatability when $\mathrm{Pe} \sim 10^5$, which is comparable to the parameter ranges for contemporary microfluidic mixing systems (33, 34). Reversibility against diffusion has previously been tested for some microfluidic mixing geometries (35), and in our simulations, we find that it primarily breaks down at very large message or hash lengths, thus creating a practical constraint on message size. However, higher Péclet numbers (and thus repeatability) could be achieved in microscale systems coupled to external forces, such as magneto- and electrofluidic mixers (1, 34, 36). Additionally, we have performed similar analyses to quantify separately potential limitations due to inertial effects and particle collisions, and we find that experimentally realistic values $\mathrm{Re} \sim 10^{-3}$ and tracer particle size $d < 10\,\mu m$ would allow consistent hashing within the two-vortex system (*SI Appendix*, sections 7 and 8).

Such constraints would not impede application of our hashing system to other, larger-scale problems in mixing theory, such as the identification and classification of coherent structures in complex flows using topological metrics (6, 37). Such efforts are often motivated by the characterization of large-scale ocean flows from sparse buoy data (38), but they have found diverse uses ranging from the biomechanics of swimming to jet dynamics (19, 39). Our work suggests that complex flows with chaotic dynamics may leave distinguishable signatures in their arrangement of advected particles, suggesting future uses for hash distributions in characterizing arbitrary flows from tracer trajectories.

Materials and Methods

All simulations were carried out using the Mathematica software package (Version 11.0.1.0; Wolfram Research, Inc.). When possible, all map parameters (such as the domain radius and stirrer coordinates) were expressed using fractional representations (i.e., $T = 3/2$ instead of $T = 1.5$; $b = 1/2$ instead of $b = 0.5$) to allow the program to retain numerical precision under repeated iterations of the map. Quantities requiring explicit decimal representations, such as the locations of advected particles under the map, we set to a precision limit of 100 places after the decimal, which was found to be sufficient for this system. Precision retention is further characterized in *SI Appendix*, section 5.

ACKNOWLEDGMENTS. I thank the US Department of Defense for supporting independent research through the National Defense Science & Engineering Fellowship program.

1. Katsikis G, Cybulski JS, Prakash M (2015) Synchronous universal droplet logic and control. *Nat Phys* 11:588–596.
2. Fuerstman MJ, Garstecki P, Whitesides GM (2007) Coding/decoding and reversibility of droplet trains in microfluidic networks. *Science* 315:828–832.
3. Woodhouse FG, Fawcett JB, Dunkel J (2018) Information transmission and signal permutation in active flow networks. *New J Phys* 20:035003.
4. Attanasi A, et al. (2014) Information transfer and behavioural inertia in starling flocks. *Nat Phys* 10:691–696.
5. Toepke MW, Abhyankar VV, Beebe DJ (2007) Microfluidic logic gates and timers. *Lab Chip* 7:1449–1453.
6. Thiffeault JL (2005) Measuring topological chaos. *Phys Rev Lett* 94:084502.
7. Schneider TM, Mandre S, Brenner MP (2011) Algorithm for a microfluidic assembly line. *Phys Rev Lett* 106:094503.
8. Wheeler JA (1992) Recent thinking about the nature of the physical world: It from bit. *Ann N Y Acad Sci* 655:349–364.
9. Landauer R (1996) The physical nature of information. *Phys Lett A* 217:188–193.
10. Ott E, Grebogi C, Yorke JA (1990) Controlling chaos. *Phys Rev Lett* 64:1196, and erratum (1990) 64:2837.
11. Ottino J (1990) Mixing, chaotic advection, and turbulence. *Annu Rev Fluid Mech* 22:207–254.
12. Brandstäter A, et al. (1983) Low-dimensional chaos in a hydrodynamic system. *Phys Rev Lett* 51:1442–1445.
13. Aref H (1984) Stirring by chaotic advection. *J Fluid Mech* 143:1–21.
14. Paar C, Pelzl J (2009) *Understanding Cryptography: A Textbook for Students and Practitioners* (Springer Science & Business Media, Berlin).
15. Amigo J, Kocarev L, Szczepanski J (2007) Theory and practice of chaotic cryptography. *Phys Lett A* 366:211–216.
16. Amigó J, Szczepanski J, Kocarev L (2005) Discrete chaos and cryptography. *Proceedings of International Symposium on Nonlinear Theory and its Applications (NOLTA2005)* (The Institute of Electronics, Information and Communication Engineers, Bruges, Belgium), pp 461–464.
17. Sturman R, Ottino JM, Wiggins S (2006) *The Mathematical Foundations of Mixing. The Linked Twist Map as a Paradigm in Applications: Micro to Macro, Fluids to Solids* (Cambridge Univ Press, Cambridge, UK), Vol 22.
18. Aref H, et al. (2017) Frontiers of chaotic advection. *Rev Mod Phys* 89:025007.
19. Haller G (2015) Lagrangian coherent structures. *Annu Rev Fluid Mech* 47:137–162.
20. Kocarev L, Szczepanski J (2004) Finite-space lyapunov exponents and pseudochaos. *Phys Rev Lett* 93:234101.
21. Sinha K, Sinha BP (2009) On the distribution of runs of ones in binary strings. *Comput Math Appl* 58:1816–1829.
22. Torquato S, Stillinger FH (2003) Local density fluctuations, hyperuniformity, and order metrics. *Phys Rev E* 68:041113.
23. Weijs JH, Jeanneret R, Dreyfus R, Bartolo D (2015) Emergent hyperuniformity in periodically driven emulsions. *Phys Rev Lett* 115:108301.
24. Jack RL, Thompson IR, Sollich P (2015) Hyperuniformity and phase separation in biased ensembles of trajectories for diffusive systems. *Phys Rev Lett* 114:060601.
25. Albert R, Jeong H, Barabási AL (2000) Error and attack tolerance of complex networks. *Nature* 406:378–382.
26. Boyland PL, Aref H, Stremler MA (2000) Topological fluid mechanics of stirring. *J Fluid Mech* 403:277–304.
27. Caussin JB, Bartolo D (2015) Braiding a flock: Winding statistics of interacting flying spins. *Phys Rev Lett* 114:258101.
28. Berger MA (2001) Topological invariants in braid theory. *Lett Math Phys* 55:181–192.
29. Indesteege S (2010) Analysis and design of cryptographic hash functions. PhD thesis (Katholieke Universiteit Leuven, Leuven, Belgium).
30. Preneel B (2010) The first 30 years of cryptographic hash functions and the NIST SHA-3 competition. *Cryptographers' Track at the RSA Conference* (Springer, Berlin), pp 1–14.
31. De Leo E, Galluccio L, Lombardo A, Morabito G (2012) Networked labs-on-a-chip (NLoC): Introducing networking technologies in microfluidic systems. *Nano Commun Networks* 3:217–228.
32. Michaelides EE (1997) The transient equation of motion for particles, bubbles, and droplets. *J Fluids Eng* 119:233–247.
33. Stroock AD, et al. (2002) Chaotic mixer for microchannels. *Science* 295:647–651.
34. Friend J, Yeo LY (2011) Microscale acoustofluidics: Microfluidics driven via acoustics and ultrasonics. *Rev Mod Phys* 83:647–704.
35. Tabeling P, Chabert M, Dodge A, Jullien C, Okkels F (2004) Chaotic mixing in cross-channel micromixers. *Philos Trans R Soc A* 362:987–1000.
36. Posner JD, Pérez CL, Santiago JG (2012) Electric fields yield chaos in microflows. *Proc Natl Acad Sci USA* 109:14353–14356.
37. Schlueter-Kuck KL, Dabiri JO (2017) Coherent structure colouring: Identification of coherent structures from sparse data using graph theory. *J Fluid Mech* 811:468–486.
38. Wiggins S (2005) The dynamical systems approach to Lagrangian transport in oceanic flows. *Annu Rev Fluid Mech* 37:295–328.
39. Peng J, Dabiri J (2009) Transport of inertial particles by Lagrangian coherent structures: Application to predator–prey interaction in jellyfish feeding. *J Fluid Mech* 623:75–84.