Critical Dynamics of Natural Time-Varying Images

Brandon Munn and Pulin Gong*
School of Physics and ARC Centre of Excellence for Integrative Brain Function, University of Sydney, Sydney, New South Wales 2006, Australia

(Received 19 March 2018; revised manuscript received 1 June 2018; published 1 August 2018)

A deep understanding of the dynamical properties of natural time-varying images is essential for interpreting how they are efficiently processed in the brain. Here we examine natural time-varying images from the perspective of their spatiotemporal patterns and find evidence of dynamical thermodynamic criticality. We further demonstrate that these spatiotemporal patterns ubiquitously organize as localized, propagating patterns. By studying these propagating patterns and their spreading dynamics over time, we demonstrate that the critical dynamics of natural time-varying images belong to the universality class of directed percolation. These critical dynamics have important implications for understanding neural processing of time-varying stimuli.

DOI: 10.1103/PhysRevLett.121.058101

Understanding the dynamical properties of natural scenes is a problem of long-standing interest in physics [1], neuroscience [2], psychology [3], and computer science [4,5]. In previous studies of natural scenes, it has been found that the spatial power spectrums of pixel intensities [6] and contrast values [1], along with clusters of pixels with similar intensity values [7,8], possess scale-invariant features. These physical properties of natural scenes have been revealed by treating contrast or intensity values of pixels as states of 2D lattice systems; despite the success of this approach, only static natural images have been considered in these previous studies. The visual system, however, is evolutionarily honed to adaptively and rapidly process dynamical, time-varying visual inputs. Studies of natural time-varying images (NTVI) have found nontrivial scaling properties between spatial and temporal power spectra [9] and higher-order spatiotemporal correlations from various local motion signals [10–12]; nevertheless, the intrinsic spatiotemporal properties of NTVI remain unknown.

In this study, we investigate NTVI from the perspective of their spatiotemporal patterns. We first show that the probability distribution of these patterns is inhomogeneous, and, by adopting a thermodynamic approach that treats time as an extra dimension [13], we demonstrate that the collective dynamics of the spatiotemporal patterns exhibit dynamical criticality. We then reveal the nonequilibrium nature of the criticality by examining the statistical and propagating properties of clustered patterns that are embedded in natural time-varying images and find that their collective dynamics belong to the universality class of directed percolation (DP).

To do so, we analyze an ensemble of NTVI belonging to the UCF-50 action recognition data set [14] that have been recorded using a dynamic vision sensor (DVS) imaging device [15]. Unlike traditional video cameras that capture entire frames at a fixed frequency, each pixel of a DVS produces an output only when the light intensity impinging upon a pixel changes beyond a threshold amount, with latencies and temporal precision down to microseconds [15]. One advantage of this feature is that the DVS produces a sparse, yet temporally precise representation of a visual scene by removing scene redundancies, such as static backgrounds [9,16]. DVS imaging devices are therefore targeted towards and widely used in dynamic computer vision tasks requiring high temporal precision, high dynamic range, and low power consumption [16,17]. This sparse DVS data set allows us to analyze a large ensemble of movies, extending upon previous analyses that have been restricted to small samples due to the use of computationally dense frame-based data [9,18]. Furthermore, the UCF-50 data set includes a range of natural videos such as skiing, horse riding, diving, and gymnastics. The advantage of this data set is that it consists of an ensemble of naturalistic action tracking scenes unlike cinematic movies possessing inherently unnatural statistics [19]. Thus, the UCF-50 action data set recorded using a DVS camera represents a pragmatic ensemble of NTVI.

The resolution of the DVS is 240×180 pixels, and the monitor displaying the movie has a refresh rate of $f = 60$ Hz [16]. To characterize events, we use a temporal resolution of $\Delta t = 1/f$. We use this window to remove any effects from the screen refreshing; however, the results are robust in the range $\Delta t = 1/f \pm 0.8 \times 1/f$. We focus on a randomly sampled subset of the ensemble, ~20% from each of the 50 categories, covering over 1000 movies. Compared to traditional fixed frame-rate movies [Fig. 1(a)], the DVS outputs a sparse representation [Fig. 1(b)]. To describe NTVI recorded by a DVS with M pixels, we assign a binary variable $\sigma_{n,t} = 1$ if an event occurs within

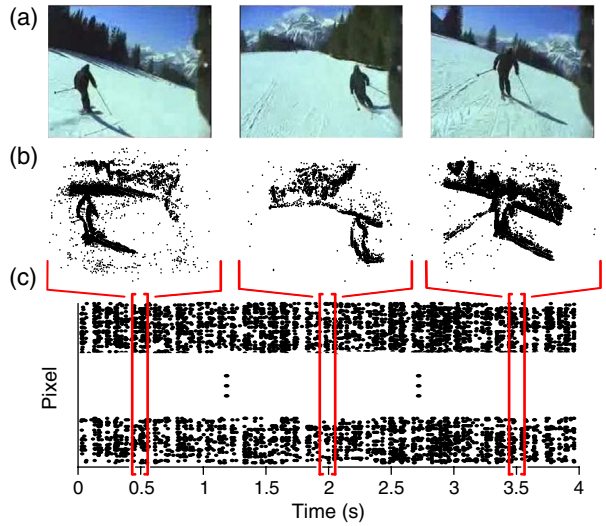

FIG. 1. Natural time-varying images recorded using a DVS. (a) Three frames from the UCF-50 action data set of a skier, (b) the corresponding DVS events, and (c) the responses of the DVS pixels across the entire movie (DVS events are represented by black dots).

the window t at pixel n and otherwise $\sigma_{n,t} = 0$, where $n = 1$ to M; an example of these binary variables is demonstrated in Fig. 1(c) from the corresponding natural movie shown in Fig. 1(a).

To include the temporal dynamics of the NTVI as in Ref. [13], we consider these binary variables of the DVS signal across L consecutive time steps. We define a state of a NTVI as the sequence of binary variables between $t = 1$ and $t = L$ consecutive time steps $\sigma = \{\sigma_{n,t}\}$. We find that the mean state probability distribution is inhomogeneous and heavy tailed [Fig. 2(a), left] and that the most common states consist of few events while the rarer states consist of multiple events [Fig. 2(a), right] (averaged across individual NTVI sampled with $M = 2500$ pixels across $L = 4$ time steps, and shaded areas are SEM). Figure 2(b) demonstrates the corresponding spatiotemporal patterns from three differently ranked states where events are colored for $L = 4$ time steps. As NTVI predominantly consist of a static background, the most common state is always that of quiescence (no events), accounting for ~50% of the observed activity. The great inhomogeneity shown in the state probabilities suggests that the movies exhibit nontrivial dynamics. To reveal these dynamics, we apply a thermodynamic approach treating time as an extra dimension and calculate the time-normalized specific heat [13].

In this thermodynamic framework, the probability of a given state can be related to an energy as $E(\sigma) = -\ln P(\sigma)$. We can then introduce a nonparametric control parameter analogous to the temperature, τ, to probe this inhomogeneous distribution, and the temperature-dependent probability distribution can be written as

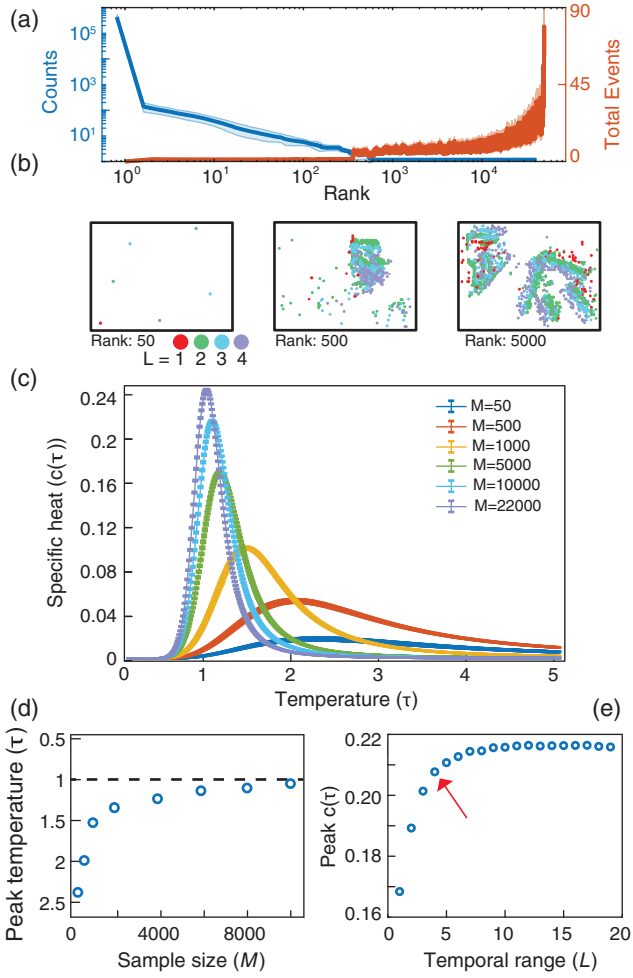

FIG. 2. Diverging specific heat of natural time-varying images. (a) The mean state count (left axis) and mean number of events (DVS output) (right axis) within a state ($M = 2500$, $L = 4$) ranked in order of their occurrence (averaged across individual NTVI, and shaded areas are SEM). (b) Spatiotemporal DVS patterns from three differently ranked states (events are colored by their time step). (c) Mean specific heat [$c(\tau)$] curves for different pixel samples (averaged across all NTVI; error bars are SEM and $L = 4$). (d) Specific heat peak temperature against sample size M (circles) approaches $\tau = 1$ ($L = 4$). (e) Maximum specific heat against temporal range L (circles); beyond $L = 4$ (red arrow), the increase plateaus ($M = 10\,000$).

$P_\tau(\sigma) = (1/Z_\tau)\exp[-E(\sigma)/\tau]$, where Z_τ is the partition function. First, the entropy can be defined as $H(\tau) = -\sum_\sigma P_\tau(\sigma)\ln P_\tau(\sigma)$ and the heat capacity as $C(\tau) = \tau\partial H(\tau)/\partial\tau \propto \mathrm{Var}[E_\tau(\sigma)]/\tau^2$ [13]. As $C(\tau)$ is an extrinsic property, it scales with the sample size [13]; consequently, we define the specific heat $c(\tau)$ as the heat capacity normalized across pixels and time, $c(\tau) = C(\tau)/ML$, which is an intrinsic measure. An increasing specific heat with an increasing pixel sample size suggests that the specific heat is divergent, indicating the presence of criticality [13,20].

Therefore, we test for criticality by analyzing how the specific heat evolves as we increase the pixel sample size. We find that the specific heat diverges as the sample size increases [Fig. 2(c)]; here the mean specific heat is calculated across individual NTVI for an increasing pixel sampling repeated 50 times for each movie and sample size, with $L = 4$. We find that the mean specific heat diverges as the sample size increases [Fig. 2(c)] and that the peak temperature of the specific heat curves approaches 1 from above with an increasing sample size [Fig. 2(d)], indicating that the observed NTVI ($\tau = 1$) are critical.

We find that including spatial and temporal aspects of NTVI is essential for a pronounced divergence of the specific heat. As in Fig. 2(c), for dense sampling (22 000 pixels covering ~50% of the available pixels) the criticality is more distinct, whereas for sparse sampling (50 pixels covering ~0.1% of the available pixels) the phenomenon is obscured. Similarly, Fig. 2(e) demonstrates the importance of including temporal dynamics for a pronounced divergence of the specific heat and that including timescales beyond $L = 4$ temporal dimensions leads to diminished increases of specific heat.

The thermal criticality identified above is based on the inhomogeneous probability distribution of states from NTVI. We find an inverse relationship between a state probability and the number of events within the state [Fig. 2(a), left] and that rarer states correspond to propagating patterns [Fig. 2(b)]. These patterns exhibit variable spatiotemporal dynamics ranging from localized bursts to long-lived wavelike activity and resemble percolating processes propagating through space and time observed in other complex physical systems [21,22] (Fig. 3). By investigating the dynamical properties of these patterns, we now demonstrate the nonequilibrium nature of the criticality belonging to the universality class of DP. We first detect these patterns based on their spatial and temporal

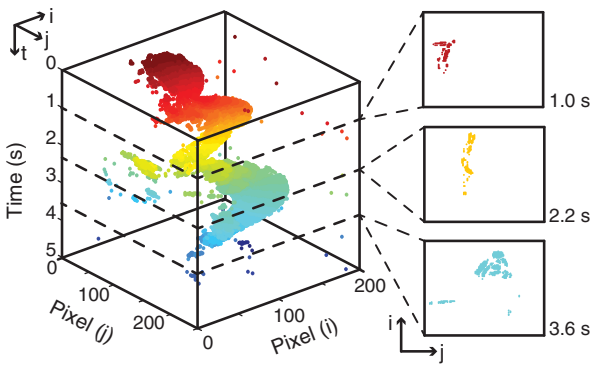

FIG. 3. Spatiotemporal patterns of events from natural time-varying images resemble a percolation process. An illustrative example of DVS events from a diver (left) (events are colored by their time step, progressing from red to blue as the time advances). Three snapshots of the DVS events at different moments: 1 s (red), beginning their dive; 2.2 s (yellow), jumping on the board; and 3.6 s (blue), flipping (right).

contiguity, with each of them being referred to as a cascade. Specifically, for each time step, events are clustered within a radius r_S, and a cascade is defined as a contiguous set of clusters with a center of mass within a radius r_T between successive time steps [23]; this method allows multiple cascades to be detected at any given time. A cascade is quantified by three quantities: (i) size S, i.e., the number of events within a cascade; (ii) duration T, i.e., the number of successive time steps a cascade is active; and (iii) mean squared displacement (MSD), i.e., the displacement from a cascade point of origin.

We now derive a nonparametric characterization of DP based on the cascade duration distribution, size distribution, MSD, and a scaling relation. This nonparametric characterization is derived from the standard characterization of DP based on a control parameter p and its distance to the critical point p_c, defined as

$$\rho \sim (p - p_c)^{\beta}, \qquad \xi_{\perp} \sim (p - p_c)^{-\nu_{\perp}}, \qquad \xi_{\parallel} \sim (p - p_c)^{-\nu_{\parallel}}, \tag{1}$$

where ρ is the density of events, $\xi_{\perp}$ and $\xi_{\parallel}$ are the spatial and temporal correlation lengths, respectively, and β, $\nu_{\perp}$, and $\nu_{\parallel}$ are the three critical bulk exponents [24]. In DP, the dimension d corresponds to the number of neighbors of each point. The upper critical dimension of DP is $d = 4$, and for $d \geq 4$ the scaling behavior is described by the mean field theory [24,25]. To account for occlusion effects, the cascades of NTVI are spatiotemporally contiguous within $r_S > 1$; thus, the system is in the mean field regime, and the three bulk exponents [24] are

$$\beta = 1, \qquad \nu_{\perp} = 1/2, \qquad \nu_{\parallel} = 1. \tag{2}$$

However, this standard method of analysis cannot be used for our system, as there is no control parameter in our data [Eq. (1)]; nevertheless, DP can also be characterized by a set of spreading exponents: η, δ, and z [25]. These are defined as

$$N(t) \sim t^{\eta}, \qquad Q(t) \sim t^{-\delta}, \qquad \mathrm{MSD}(t) \sim t^{z}, \tag{3}$$

where $N(t)$ is the mean number of active events at time t, $Q(t)$ is the survival probability, and $\mathrm{MSD}(t)$ is the mean squared displacement [25]. Scaling relations can be found between the spreading exponents and the bulk exponents in Eq. (1). Because of scale invariance, at criticality, if time is scaled by a factor b, space must be similarly scaled by $b^{\nu_{\perp}/\nu_{\parallel}}$; thus, $z = 2\nu_{\perp}/\nu_{\parallel}$. For brevity, we note that the remaining spreading exponents can be found from the well-known scaling relation $\beta = \delta\nu_{\parallel}$ and the DP specific hyperscaling relation $\eta + 2\delta = dz/2$ [24–26]. Thus, the mean field spreading exponents [25] are

$$\eta = 0, \qquad \delta = 1, \qquad z = 1. \tag{4}$$

These exponents are related to the cascade size S and duration T probability distribution exponents, which are expected to be self-similar at criticality: $P(S) \sim S^{-\gamma_S}$ and $P(T) \sim T^{-\gamma_T}$. We first note that the probability of a cascade surviving to t, $Q(t)$, is the sum of all duration probabilities with an equivalent or greater duration, that is, $Q(t) = \int_{T=t}^{T=\infty} P(T)dT$; thus,

$$P(T) = -\frac{dQ(T)}{dT} \sim T^{-(1+\delta)}. \tag{5}$$

Furthermore, the average size of a cascade is related to its duration through

$$\langle S \rangle \sim \int_{t=0}^{t=T} \frac{N(t)}{Q(t)} dt \sim T^{1+\eta+\delta}, \tag{6}$$

and, if the fluctuations are small around this proportionality, then $P(S)dS \sim P(T)dT$. Confining this relation [Eqs. (5) and (6)] gives

$$P(S) \sim T^{-(1+\eta+2\delta)} \sim S^{[-(1+\eta+2\delta)/(1+\eta+\delta)]}. \tag{7}$$

Finally, using the mean field DP exponents [Eq. (4)], the scaling relations for the exponents γ_T and γ_S characterizing the duration and size of the cascade, respectively, are

$$\gamma_T = 1 + \delta = 2 \tag{8}$$

and

$$\gamma_S = \frac{1 + \eta + 2\delta}{1 + \eta + \delta} = 3/2. \tag{9}$$

We now show that the observed pattern dynamics of NTVI satisfy these critical exponents and belong to the universality class of DP. The size, duration, and mean squared displacement of all cascades detected in individual natural time-varying images are pooled across the ensemble. As shown in Figs. 4(a) and 4(b), the cascade size and duration probability densities (red circles) appear to follow a power law relation spanning 4 and 2 orders of magnitude, respectively. The finite-size cutoff at $T \sim 10^3 \times \Delta t \sim 15$ s suggests that some cascades span an entire movie. We further demonstrate that the result is not due to our analysis techniques by replacing each pixel with a Poisson process, destroying the power law relationship [Figs. 4(a) and 4(b), blue crosses]. Furthermore, the close fit to a power law is confirmed by the complementary cumulative distribution function $P(X \geq x)$, which also shows a power law distribution [Figs. 4(a), inset and 4(b), inset].

We test these distributions by implementing maximum likelihood estimators [27] to fit the exponents and find $\gamma_S = 1.500 \pm 0.001$ and $\gamma_T = 2.1 \pm 0.1$ in agreement with the derived DP exponents of $\gamma_S = 3/2$ and $\gamma_T = 2$ [24,25]. As the maximum likelihood estimator provides only an estimate of the exponent, we compare the log-likelihood ratios

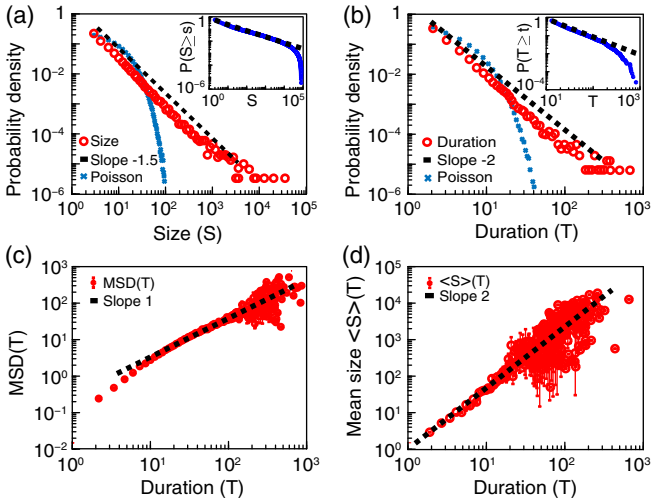

FIG. 4. Spatiotemporal cascades of natural time-varying images belong to the universality class of DP. (a),(b) Probability density functions for all detected cascade sizes and durations (red circles) detected in individual movies and pooled across the ensemble. Inset, complementary cumulative density functions of the same data. The distributions closely follow the DP mean field exponents $\gamma_S = 1.5$ and $\gamma_T = 2$ (dotted black line). Replacing each DVS pixel with a Poisson process destroys the power law relationship (blue crosses). We present results for $r_S = 15$ pixels and $r_T = 30$ pixels; however, the results are robust to parameter variation. (c) Scatter plots of MSD(T) against T; SEM are within the size of plotted points (red error bars) and well fitted to the DP exponent $z = 1$ (dotted black line). (d) Scatter plots of mean cascade size $\langle S \rangle(T)$ against T (red error bars) are well fitted to the DP scaling relation $\langle S \rangle \sim T^2$ (dotted black line).

(LLRs) between a power law and log-normal distribution. We find that the LLRs are positive for both size and duration distributions, suggesting that the distributions are very heavy tailed, indicative of a power law ($p < 0.001$, Vuong test). Thus, the cascade size and duration distributions closely coincide with those of the derived DP exponents.

The aforementioned analyses demonstrate that cascades of activity in NTVI possess power law distributions. It is interesting to note that the same size and duration exponents have been found in neuronal avalanches, which can be viewed as a critical branching process [28,29]. The branching process of neuronal avalanches has been related to a diverging specific heat [13]. As our analysis of NTVI focuses on propagating, spatiotemporally contiguous patterns, by characterizing these patterns, we can use the hyperscaling relation involving MSD, $\eta + 2\delta = dz/2$, which is unique to DP, to distinguish between other critical phenomena [24,25]. Thus, to complete the characterization of DP, we ask whether the spatiotemporally contiguous cascades satisfy the theoretical DP MSD exponent and scaling relation between the cascade size and duration [Eq. (6)]. As shown in Figs. 4(c) and 4(d), the MSD(T) and mean cascade size for a given duration both appear to

follow a power law relationship. We find $\mathrm{MSD}(T) \sim T^z$, where $z = 1.0$, and $\langle S \rangle \sim T^\alpha$, where $\alpha = 1.95$ (nonlinear least-squares fitting), both consistent with the derived mean field DP values of $z = 1$ and $\alpha = 2$, respectively.

In summary, by uniquely focusing on the spatiotemporal patterns of NTVI, our study presents the first observation that the collective spatiotemporal dynamics of NTVI are critical and belong to the universality class of DP.

The DP property of natural stimuli indicates that there exist spatiotemporal patterns with a range of sizes, particularly large percolating clusters. These dynamical patterns as found in the DVS recordings are much larger than local, Gabor-like filters often used in understanding signal processing within the visual system [6]. Hence, segmenting natural time-varying stimuli into such dynamical patterns may be an efficient signal representation mechanism. In addition, it has long been hypothesized that neural systems are adapted to efficiently exploit the spatiotemporal structures of natural signals [30,31] such that the internal dynamics of neural population activity is similar to that of external environments [2,31]. Propagating patterns belonging to DP represent a novel, dynamical property for quantifying whether and how internal dynamics of the brain match external environments. Indeed, clustered activity patterns with similar propagating properties to those found in NTVI have been observed at both mesoscopic [32,33] and macroscopic levels [34,35] of neural systems. These experimental observations suggest that it would be interesting to analyze neural population activity and test whether the spatiotemporal dynamics of neural systems belong to the same universality class as natural movies.

Finally, the ubiquity of such patterns in both natural images and neural spatiotemporal activity leads us to hypothesize that they might be essential for external inputs to efficiently modulate internal neural activity to optimally represent dynamical external environments. Such efficient modulation may explain why criticality has been clearly observed in the collective activity of retinal neurons responding to naturalistic movies compared with other artificial stimuli [20], suggesting that such driven retinal activity might also exhibit DP properties.

This work was supported by the Australian Research Council (Grants No. DP160104316 and No. DP160104368).

*pulin.gong@sydney.edu.au

[1] D. L. Ruderman and W. Bialek, Phys. Rev. Lett. **73**, 814 (1994).
[2] E. P. Simoncelli and B. A. Olshausen, Annu. Rev. Neurosci. **24**, 1193 (2001).
[3] K. K. Evans and A. Treisman, J. Exp. Psychol. Hum. Percept. Perform. **31**, 1476 (2005).
[4] S. Roth and M. J. Black, Int. J. Comput. Vis. **82**, 205 (2009).
[5] R. W. Buccigrossi and E. P. Simoncelli, IEEE Trans. Image Process. **8**, 1688 (1999).
[6] D. J. Field, J. Opt. Soc. Am. A **4**, 2379 (1987).
[7] S. Saremi and T. J. Sejnowski, Proc. Natl. Acad. Sci. U.S.A. **110**, 3071 (2013).
[8] G. J. Stephens, T. Mora, G. Tkačik, and W. Bialek, Phys. Rev. Lett. **110**, 018701 (2013).
[9] D. W. Dong and J. J. Atick, Netw., Comput. Neural Syst. **6**, 345 (1995).
[10] E. H. Adelson and J. R. Bergen, J. Opt. Soc. Am. **2**, 284 (1985).
[11] C. Chubb and G. Sperling, J. Opt. Soc. Am. **5**, 1986 (1988).
[12] Q. Hu and J. D. Victor, J. Vision **10**, 9 (2010).
[13] T. Mora, S. Deny, and O. Marre, Phys. Rev. Lett. **114**, 078105 (2015).
[14] K. K. Reddy and M. Shah, Machine Vision and Applications **24**, 971 (2013).
[15] P. Lichtsteiner, C. Posch, and T. Delbruck, IEEE J. Solid-State Circuits **43**, 566 (2008).
[16] Y. Hu, H. Liu, M. Pfeiffer, and T. Delbruck, Front. Neurosci. **10**, 405 (2016).
[17] P. O'Connor, D. Neil, S.-C. Liu, T. Delbruck, and M. Pfeiffer, Front. Neurosci. **7**, 178 (2013).
[18] E. I. Nitzany and J. D. Victor, J. Vision **14**, 10 (2014).
[19] A. Torralba and A. Oliva, Netw., Comput. Neural Syst. **14**, 391 (2003).
[20] G. Tkačik, T. Mora, O. Marre, D. Amodei, S. E. Palmer, M. J. Berry, and W. Bialek, Proc. Natl. Acad. Sci. U.S.A. **112**, 11508 (2015).
[21] K. A. Takeuchi, M. Kuroda, H. Chaté, and M. Sano, Phys. Rev. Lett. **99**, 234503 (2007).
[22] G. Lemoult, L. Shi, K. Avila, S. V. Jalikop, M. Avila, and B. Hof, Nat. Phys. **12**, 254 (2016).
[23] G. De Soete and J. D. Carroll, *New Approaches in Classification and Data Analysis* (Springer, Heidelberg, 1994).
[24] H. Hinrichsen, Adv. Phys. **49**, 815 (2000).
[25] M. A. Munoz, R. Dickman, A. Vespignani, and S. Zapperi, Phys. Rev. E **59**, 6175 (1999).
[26] F. Landes, A. Rosso, and E. A. Jagla, Phys. Rev. E **86**, 041150 (2012).
[27] A. Clauset, C. R. Shalizi, and M. E. Newman, SIAM Rev. **51**, 661 (2009).
[28] J. M. Beggs and D. Plenz, J. Neurosci. **23**, 11167 (2003).
[29] T. Petermann, T. C. Thiagarajan, M. A. Lebedev, M. A. Nicolelis, D. R. Chialvo, and D. Plenz, Proc. Natl. Acad. Sci. U.S.A. **106**, 15921 (2009).
[30] H. B. Barlow, in *Sensory Communication*, edited by W. A. Rosenblith (MIT Press, Cambridge, MA, 1961).
[31] P. Berkes, G. Orbán, M. Lengyel, and J. Fiser, Science **331**, 83 (2011).
[32] G. Scott, E. D. Fagerholm, H. Mutoh, R. Leech, D. J. Sharp, W. L. Shew, and T. Knöpfel, J. Neurosci. **34**, 16611 (2014).
[33] R. G. Townsend, S. S. Solomon, S. C. Chen, A. N. Pietersen, P. R. Martin, S. G. Solomon, and P. Gong, J. Neurosci. **35**, 4657 (2015).
[34] E. Tagliazucchi, P. Balenzuela, D. Fraiman, and D. R. Chialvo, Front. Oral Physiol. **3**, 15 (2012).
[35] A. Mitra, A. Z. Snyder, T. Blazey, and M. E. Raichle, Proc. Natl. Acad. Sci. U.S.A. **112**, E2235 (2015).