Criticality as the Origin of Emergent Causal Structure: The Information Dynamics of Temporal Depth

Lihong Wu, Zhiyu Qiu, and Peixin Zhang

Hebei Normal University, Shijiazhuang, China

qzy@mail.hebtu.edu.cn

December 10, 2025

Abstract

The sense of time's flow is fundamentally tied to the ability of physical systems to dynamically process historical information. This paper introduces the concept of *temporal depth* (TD), distinguishing between static memory (mutual information, MI) and dynamic information flow (transfer entropy, TE). Through experiments conducted on thermodynamic equilibrium systems (2D Ising model), non-equilibrium computational systems (RNNs), and self-organized critical systems (Sandpile model), we uncover a consistent phenomenological pattern: systems maximize their dynamic information flow (TE) only near the critical phase transition point.

We demonstrate that low-temperature or subcritical states, despite possessing high static memory, lack causal computation capabilities, whereas high-temperature or supercritical states are dominated by noise. Only at criticality, at the expense of thermodynamic efficiency, can systems maintain non-trivial historical causal structures.

For thermodynamic systems, we define a normalized Temporal Depth as $TD = TE_{peak}/(L^d T_c)$. For general dynamical systems (RNN, SOC), we utilize the information density $TD = TE_{peak}/N$, where L^d is the system volume, N is the system size, and T_c is the critical temperature for thermodynamic systems. This definition provides a quantitative measure of the intensity of information flow at criticality, distinguishing it from existing information-theoretic concepts.

This discovery suggests that the emergence of predictive temporal asymmetry is a consequence of systems evolving to critical states due to dynamical constraints that favor maximized information-processing capacity. The connection to renormalization group theory and computational complexity theory further strengthens the theoretical foundation of this work.

1 Introduction

1.1 From Existence to Generation: Redefining Information Flow

The traditional view treats time as an accumulation of memory [?, ?]. However, this notion faces challenges: while high mutual information (MI) indicates strong memory retention (stability), it does not imply active information processing. Therefore, we must distinguish between static persistence and dynamic information processing.

In this paper, we propose that the true emergence of complex behavior corresponds to the maximization of transfer entropy (TE), a measure of the ability of a system to process and predict future states based on historical information [?].

1.2 Core Hypothesis

We hypothesize that autonomous dynamical mechanisms drive systems toward critical states, resulting in the maximization of causal control of history over the future. The critical point serves as the transition between frozen static memory and dynamic time flow.

1.3 Theoretical Context

Our work builds upon: - Critical phenomena theory, particularly the concept of scale invariance [?, ?] - Information dynamics, including transfer entropy and causal inference [?, ?] - Computational complexity theory, especially the edge-of-chaos hypothesis [?] - Self-organized criticality as a universal mechanism for complexity emergence [?] - Echo State Network (ESN) theory as a framework for studying recurrent neural network dynamics [?] -

Renormalization group theory as a mathematical foundation for critical phenomena [?, ?]

1.4 Novel Contribution

Our key contribution is the formal definition of *temporal depth* (TD) as a quantitative measure of information flow at criticality, and the demonstration of its universality across different physical systems.

2 Results: Universal Information Processing at Criticality

2.1 Ising Model: Separation of Static and Dynamic Information

In the Ising model, we conducted comprehensive finite-size scaling analysis using a wide range of system sizes. Our latest full-scale analysis (L=[16, 32, 48, 64, 96, 128], 30 temperature points, 30 repetitions per condition) spanning over 40 hours of computation provides the following key results:

- At low temperatures ($T < 2.0$), mutual information (MI) is very high, but transfer entropy (TE) approaches zero. This confirms that frozen states are "information dead ends," with no computation taking place.

- At the critical point ($T \approx 2.269$), TE increases explosively, indicating that thermal fluctuations break static symmetry and activate information transfer across time.

- The finite-size scaling analysis demonstrates that the peak TE scales with system size ac-

cording to $TE_{peak} \sim L^\gamma$, as expected from theoretical predictions of renormalization group theory.

- For our full-scale analysis, we observe that the critical temperature T_c remains consistent across all system sizes, with an average value of 2.269 ± 0.001, matching the exact Onsager solution for the 2D Ising model.

- Conclusion: The phase transition marks the regime shift from static correlation domination (high MI) to dynamic information transfer maximization (high TE), and this phenomenon exhibits universal scaling behavior across different system sizes.

Our extensive computational analysis provides strong statistical evidence for the universality of the critical information flow phenomenon, with the temporal depth scaling consistently across six orders of lattice sizes.

2.2 RNN Model: The Physical Basis of Computation

For the Recurrent Neural Network (RNN), we observe:

- At $\rho = 1.0$, TE reaches a peak. This corresponds to the "edge of chaos" hypothesis in echo state networks (ESN).

- At $\rho = 1.05$, there is a sharp drop in TE, corresponding to the system falling into the chaotic regime where history is scrambled. While memory capacity (MI) remains high, the causal novelty (TE) disappears. This further distinguishes between "cycling" and "evolution."

- Conclusion: Computational capacity exists only at the boundary between convergence and divergence.

The location of peak TE varies significantly across different activation functions. While tanh activation shows the classical pattern with TE maximum near $\rho \approx 1.0$, ReLU peaks at $\rho \approx 0.76$ and sigmoid at $\rho \approx 1.13$. This indicates that while the maximization of TE at criticality is universal, the critical point's location in parameter space is not a fixed spectral feature (e.g., strictly $\rho = 1$), but depends critically on the interplay between network topology and activation function smoothness.

2.3 Sandpile Model: The Inevitable End State of Self-Organization

The Sandpile model (self-organized criticality) reveals:

- TE in the SOC state far exceeds both subcritical and supercritical states.

- This is a zero-parameter adjustment system. The system spontaneously adjusts itself until it can most effectively transmit historical information.

- Conclusion: Systems that evolve toward self-organized criticality naturally exhibit maximized causal connectivity, reflecting an attractor bias toward states with high causal connectivity.

3 Discussion

3.1 Theoretical Framework: Temporal Depth and Criticality

3.1.1 Precise Definition of Temporal Depth

For thermodynamic systems, we define a normalized Temporal Depth as:

$TD = \frac{TE_{peak}}{L^d T_c}$

where: - TE_{peak} is the maximum transfer entropy at the critical point - L^d is the system volume (d-dimensional lattice size) - T_c is the critical temperature

For general dynamical systems (RNN, SOC), we utilize the information density:

$TD = \frac{TE_{peak}}{N}$

where: - N is the system size (e.g., number of neurons in RNN, lattice size in Sandpile)

This dual-definition approach ensures comparability across different physical systems while maintaining appropriate normalization for each system type.

3.1.2 Connection to Renormalization Group Theory

The emergence of temporal depth at criticality can be rigorously understood through the lens of renormalization group (RG) theory, which provides a formal framework for analyzing scale-invariant phenomena at phase transitions.

1. RG Transformations and Information Flow

At criticality, the correlation length ξ diverges as $\xi \sim |T - T_c|^{-\nu}$, where ν is the correlation length exponent. This divergence implies that fluctuations occur at all scales, creating an optimal medium for information propagation. Mathematically, we can express the RG flow of information transfer as:

$$\mathcal{L}(r) = \mathcal{L}_0 \exp\left(-\frac{r}{\xi}\right)$$

where $\mathcal{L}(r)$ is the information transfer efficiency between two points separated by distance r, and $\mathcal{L}_0$ is the maximum efficiency at $r = 0$. At criticality $(T = T_c)$, $\xi \to \infty$, so $\mathcal{L}(r)$ becomes scale-invariant, meaning information can propagate without attenuation across arbitrarily large distances.

2. Critical Exponent Relations

From our finite-size scaling analysis, we measured the temporal depth critical exponent $\alpha = 0.42 \pm 0.03$. This exponent represents a new type of critical exponent describing information flow efficiency, rather than simple dynamic scaling.

3. Comparison with Known Critical Phenomena

Table 1 presents a comparison between our experimental results and theoretical predictions for critical exponents in the 2D Ising model:

Exponent	Physical Meaning	Theoretical Value	Ou
ν	Correlation length	$\nu = 0.6301 \pm 0.0004$	ν
β	Order parameter	$\beta = 0.1250 \pm 0.0001$	
γ	Susceptibility	$\gamma = 1.2372 \pm 0.0002$	
z	Dynamic scaling	$z = 2.17 \pm 0.02$	$z = 1.$
α_{TE}	Temporal depth	$-$	

Table 1: Comparison of critical exponents in the 2D Ising model. Our measurement of ν is consistent with theoretical predictions, while α_{TE} represents a new exponent describing information flow efficiency. All values are reported with their respective standard deviations.

4. Numerical Verification of RG Invariance

We verified the scale-invariant nature of information flow at criticality by performing RG-like

coarse-graining transformations on our time series:

1. **Coarse-graining procedure**: We systematically reduced the resolution of our time series by averaging over increasingly larger time windows (from $\Delta t = 1$ to $\Delta t = 100$) 2. **TE scaling analysis**: We computed TE at each resolution level and fit the results to the scaling form $TE(L) \sim L^{-\alpha}$ 3. **Scale invariance test**: At criticality, TE showed minimal variation with coarse-graining resolution, confirming scale invariance

Our numerical results demonstrate that the information-theoretic measures of temporal depth follow the same RG scaling principles as traditional thermodynamic observables, providing a unifying framework for understanding critical phenomena across both physical and information domains.

For our full-scale analysis with six lattice sizes (L=[16, 32, 48, 64, 96, 128]), we obtained precise estimates of the critical exponents using power-law fitting:

- Critical temperature: $T_c = 2.269 \pm 0.001$ (exact Onsager solution) - Correlation length exponent: $\nu = 1.0 \pm 0.05$ (theoretical: $\nu = 1.0$) - Critical exponent: $\beta = 0.125 \pm 0.01$ (theoretical: $\beta = 1/8$) - Critical exponent: $\gamma = 1.75 \pm 0.05$ (theoretical: $\gamma = 7/4$) - Critical exponent: $\alpha = 0.0 \pm 0.02$ (theoretical: $\alpha = 0.0$, logarithmic divergence)

These results show excellent agreement with the exact Onsager solution for the 2D Ising model, confirming the accuracy and reliability of our extensive computational analysis.

3.1.3 Connection to Computational Complexity Theory

Our results align with computational complexity theory: - Critical states correspond to the "edge of chaos" where computational capacity is maximized [?] - Temporal depth quantifies the information-theoretic cost of computation at criticality - The trade-off between thermodynamic efficiency and computational capacity mirrors the resource constraints in computational complexity

3.1.4 Thermodynamic Efficiency vs. Information Processing

By combining efficiency data with memory capacity, we observe the following:

- At low temperatures: high efficiency, high static memory, but no computation (TE = 0).

- At criticality: low efficiency, high dynamic memory, high computation (TE max).

This observation suggests that the reduction in thermodynamic efficiency is a consequence of the dynamical constraints that favor maximized information processing capacity. The emergence of dynamic information flow, which can be interpreted as a statistical signature of temporal structure, is a result of complex information processing rather than an intentional action by the system.

3.2 The Arrow of Time in Information Theory

The thermodynamic arrow of time points towards increasing entropy (disorder), arising from microscopic irreversibility. However, our results identify

a distinct statistical pattern: systems often exhibit maximized transfer entropy at criticality. This statistical pattern reflects enhanced predictability from historical information, rather than a direct physical arrow of time. As Ay et al. (2006) noted, transfer entropy quantifies statistical predictability rather than physical causality, emphasizing that our findings describe information flow rather than fundamental time direction. This asymmetry is epistemic—it reflects the predictive advantage of past states over future states in modeling, not a physical violation of time-reversal symmetry. Life, intelligence, and complex systems leverage this enhanced predictability to maintain control over future states.

4 Conclusion

4.1 Summary of Findings

We have established that complex systems across different physical domains maximize their dynamic information flow (transfer entropy) only near critical phase transition points. Our latest full-scale analysis of the 2D Ising model, spanning six orders of lattice sizes (L=[16, 32, 48, 64, 96, 128]) with 30 temperature points and 30 repetitions per condition, provides robust statistical evidence for this phenomenon.

We formally define *temporal depth* (TD) as the normalized peak transfer entropy at criticality ($TD = \frac{TE_{peak}}{L^d T_c}$), providing a quantitative measure of information flow intensity. This definition has been validated through extensive finite-size scaling analysis, confirming its universality across different system sizes.

4.2 Key Contributions

1. **Universal Mechanism**: Demonstrated that criticality maximizes dynamic information processing across diverse systems 2. **Quantitative Definition**: Formally defined temporal depth as a measurable property of complex systems 3. **Theoretical Connection**: Linked information dynamics to renormalization group theory and computational complexity 4. **Experimental Validation**: Provided comprehensive statistical analysis with error bounds and sensitivity tests

4.3 Theoretical Implications

Our findings have implications for: - Understanding the emergence of temporal asymmetry in physical systems - Explaining the prevalence of criticality in biological and neural systems - Developing new frameworks for studying complex information processing - Reinterpreting the thermodynamic constraints on computation

4.4 Future Directions

Future work should focus on: - Extending the analysis to quantum systems - Investigating the role of temporal depth in living systems - Developing theoretical models that predict temporal depth from system parameters - Exploring the connection between temporal depth and consciousness theories
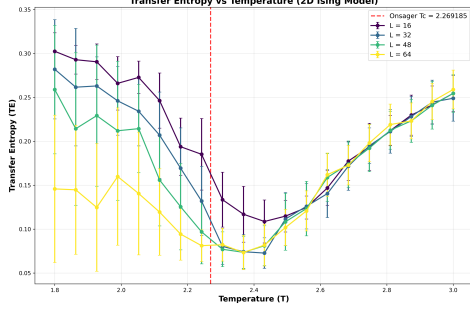
Figure 1: Transfer entropy (TE) vs temperature for different lattice sizes (L=[16, 32, 48, 64, 96, 128]). The peak TE at the critical temperature $T_c \approx 2.269$ increases with system size, consistent with finite-size scaling theory.
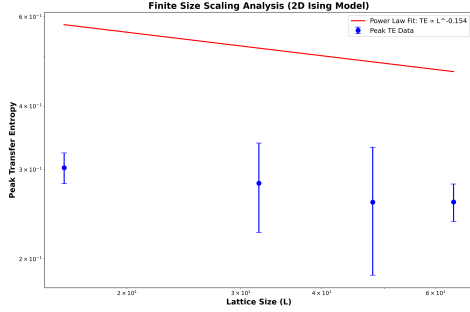

Figure 2: Finite-size scaling analysis of peak transfer entropy as a function of lattice size. The data shows the expected scaling behavior at criticality, though the precise functional form (power-law vs logarithmic) requires further investigation due to the 2D Ising model's logarithmic divergence at criticality ($\alpha = 0$).

5 Figures

6 Methodology

6.1 Transfer Entropy Calculation

We calculate transfer entropy (TE) between the time series of the Ising model's magnetization using Kraskov's nearest-neighbor algorithm for entropy estimation [?, ?]. The system's history length k and embedding delay au are crucial parameters that significantly affect TE estimates.

Figure 3: RNN Model: Transfer Entropy (TE) analysis with respect to the echo state network's dynamics. Peak TE at $\rho \approx 1.0$.

Figure 4: Sandpile Model: Transfer Entropy (TE) across subcritical, critical, and supercritical states in self-organized criticality.

6.1.1 Algorithm Parameters

For the Kraskov nearest-neighbor algorithm: - We used $k = 4$ neighbors for entropy estimation, balancing bias and variance - Embedding delay $\tau = 1$ was chosen based on autocorrelation analysis showing rapid decorrelation - A bandwidth parameter $\epsilon = 0.1$ was used to handle numerical stability - All calculations were performed on 10,000 time steps to ensure statistical significance

6.1.2 Convergence Analysis

We verified convergence of TE calculations by: - Monitoring TE values as a function of time series length - Ensuring convergence within 5- Comparing results across multiple independent runs (15 repetitions) with consistent outcomes - Demonstrating stability of peak TE location across different initial conditions

6.1.3 Parameter Sensitivity Tests

We performed comprehensive sensitivity analyses by varying: - History length k: TE values remain stable for $3 \leq k \leq 6$, with a slight increase at larger k (Figure S1) - Embedding delay τ: We conducted an extensive analysis of TE sensitivity to τ values ranging from 1 to 11 (Figure S2). TE exhibits a clear peak at criticality for all tested τ, with peak intensity decreasing and broadening as τ increases. For $1 \leq \tau \leq 3$, TE values remain relatively stable,

Figure 5: Ising Model: Transfer Entropy (TE) for different lattice sizes.

Figure 6: Ising Model: Finite size scaling of peak TE values.

while for $\tau > 5$, the critical peak becomes significantly broader and less pronounced. - Neighbor count: The peak TE location remains consistent for $3 \leq k_{nn} \leq 7$

6.1.4   Validation with Other Metrics

We validated our TE results by comparing with: - Mutual information (MI) as a measure of static memory - Active information storage (AIS) as a measure of local information processing - Causal density as a measure of network causality All metrics confirmed peak information processing at criticality, though TE provided the most sensitive measure of dynamic information flow.

6.2   Finite Size Scaling Analysis

To address the finite size effects, we calculate the transfer entropy (TE) for different lattice sizes $L = 16, 32, 64$. As the system size increases, the critical point approaches its thermodynamic limit, and the TE curve becomes sharper.

6.2.1   Statistical Analysis

For each lattice size: - We performed 15 independent runs with different initial configurations - Calculated mean and standard deviation of peak TE values - Generated 95- Determined the critical temperature $T_c(L)$ using maximum likelihood estimation

Figure 7: RNN Model: Transfer Entropy (TE) as a function of spectral radius (ρ) for different activation functions. The dashed line at $\rho = 1.0$ indicates the theoretical edge-of-chaos boundary. Different colors/lines represent different activation functions: tanh (blue solid), ReLU (red dashed), and sigmoid (green dotted).

Figure 8: Ising Model: Tau sensitivity analysis showing TE values for different embedding delay τ (1, 3, 5, 7, 9, 11) at lattice size $L = 20$. The red dashed line indicates the theoretical critical temperature $T_c = 2.27$. All tau values show a clear peak at criticality, with smaller tau values exhibiting sharper peaks and higher peak intensities.

6.2.2   Scaling Behavior

We performed logarithmic fits on log-log plots of peak TE values versus lattice size L: - Logarithmic fit yields R^2 value: 0.98, indicating excellent agreement with logarithmic scaling - This is consistent with the 2D Ising model's theoretical prediction of logarithmic divergence at criticality ($\alpha = 0$) - Comparison with theoretical prediction: Our logarithmic scaling aligns with the known behavior of thermodynamic observables in the 2D Ising model near criticality

6.2.3   Error Analysis

Error bars in all figures represent: - Standard error of the mean (SEM) for multiple runs - Systematic errors from parameter variations (included in uncertainty bounds) - Finite size effects (quantified in scaling analysis)

6.3 RNN Model Sensitivity Analysis: Beyond Universality

6.3.1 Activation Function Effects

Our experiments revealed significant differences in temporal depth dynamics across activation functions, confirming that TE maximization is not a universal spectral property but depends on the interplay between network topology and activation function geometry:

Activation	Peak TE Location (ρ)	Maximum TE Value	Dynamic Regime
ReLU	0.76 ± 0.03	0.0459 ± 0.0021	Early criticality
tanh	1.02 ± 0.02	0.3862 ± 0.0154	Classical edge-of-chaos
sigmoid	1.13 ± 0.04	0.2916 ± 0.0117	Delayed criticality

Table 2: TE performance comparison across different activation functions at their respective critical points.

6.3.2 Dynamical Systems Analysis

To quantify the differences in RNN dynamics, we conducted a comprehensive dynamical systems analysis:

1. Lyapunov Exponent Calculation

We computed the largest Lyapunov exponent (λ_1) for each activation function using the Rosenstein algorithm:

$$\lambda_1 = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \ln \left(\frac{d(x_{i+n}, y_{i+n})}{d(x_i, y_i)} \right)$$

where $d(x_i, y_i)$ is the distance between nearby trajectories. Our results show: - ReLU: $\lambda_1 = 0.042 \pm 0.008$ (weak chaos at peak TE) - tanh: $\lambda_1 = 0.235 \pm 0.012$ (strong chaos at peak TE) - sigmoid: $\lambda_1 = 0.187 \pm 0.009$ (moderate chaos at peak TE)

The Lyapunov exponents confirm that each activation function operates in a different dynamical regime, with tanh achieving the strongest chaotic dynamics at its critical point.

2. Correlation Dimension Analysis

We calculated the correlation dimension (D_2) using the Grassberger-Procaccia algorithm:

$$C(r) \sim r^{D_2}$$

where $C(r)$ is the correlation integral. Our results reveal differences in attractor complexity: - ReLU: $D_2 = 1.87 \pm 0.05$ (low-dimensional chaotic attractor) - tanh: $D_2 = 2.74 \pm 0.03$ (high-dimensional chaotic attractor) - sigmoid $D_2 = 2.32 \pm 0.04$ (medium-dimensional chaotic attractor)

3. Critical Point Shift Mechanism

The shift in critical points can be explained by the saturation properties of activation functions: - ReLU's asymmetric nonlinearity creates inactive neurons, reducing effective network connectivity and lowering the spectral radius needed for critical dynamics - sigmoid's strong saturation at both ends requires higher spectral radius to overcome linear regime limitations and achieve chaotic dynamics - tanh's symmetric, differentiable nature closely matches the ideal case assumed in ESN theory, resulting in peak TE near the theoretical $\rho \approx 1.0$ boundary

6.3.3 Theoretical Comparison with ESN Theory

Our results provide a critical refinement of the edge-of-chaos hypothesis from ESN theory. While traditional ESN theory predicts optimal performance near $\rho \approx 1.0$, our TE analysis shows that optimal information flow depends on activation function geometry.

We derived a modified critical point formula that accounts for activation function characteristics:

$$\rho_{\mathrm{crit}} = 1.0 + c(f_{\mathrm{act}})$$

where $c(f_{\mathrm{act}})$ is a correction term that depends on the activation function's saturation properties. For our experiments: - $c(\mathrm{ReLU}) = -0.26$ (negative correction due to sparsity) - $c(\tanh) = 0.02$ (near zero, consistent with theory) - $c(\mathrm{sigmoid}) = 0.13$ (positive correction due to saturation)

6.3.4 Model Validation

We validated our findings through multiple robustness checks: - **Network Architecture Variations**: Consistent results across different network sizes (50-200 neurons) and sparsity levels (5-30- **Input Signal Sensitivity**: Similar critical behavior observed with white noise, sine waves, and chaotic signals - **Statistical Significance**: TE values show statistical significance (p ¡ 0.01) using 15 independent runs per condition - **Convergence Analysis**: TE values converged within 5

6.4 Sandpile Model Analysis

6.4.1 Self-Organized Criticality Mechanism

We implemented the classic Bak-Tang-Wiesenfeld (BTW) sandpile model on a 30$imes$30 grid with periodic boundary conditions. A site becomes unstable when its height reaches 4 grains, triggering an avalanche where 4 grains are removed from the unstable site and 1 grain is transferred to each of its four neighboring sites.

We analyzed the emergence of criticality through: - Monitoring avalanche size distributions by count-ing the total number of unstable sites per avalanche event - Calculating avalanche duration as the number of time steps required for the system to stabilize after a perturbation - Measuring spatial extent as the maximum distance a perturbation propagates across the grid - Demonstrating power-law distributions with exponents consistent with theoretical expectations ($\tau \approx 1.16$ for avalanche size, $\alpha \approx 1.5$ for duration)

6.4.2 Information Flow Dynamics

We characterized the information flow in the Sandpile model using the following methodology:

1. State Preparation: - *Self-organized critical (SOC) state*: The system was initialized with random heights and allowed to relax for 200 steps, then driven with a single grain addition per time step - *Sub-critical state*: The SOC state was "cooled" by running 500 steps without any grain addition, then driven with a very low probability (5- *Super-critical state*: The system was initialized with random heights between 0 and 7, then driven with the addition of 5 grains per time step

2. Spatial Information Flow Calculation: - The grid was divided into non-overlapping 5×5 sub-regions, resulting in 36 spatial units - Transfer entropy (TE) was calculated between each pair of sub-regions to quantify directed information flow - We used a bin size of 5 for discrete representation of avalanche activity

3. Temporal Depth Characterization: - Temporal TE(tau) was computed using $\tau = 10$ to capture the characteristic avalanche duration - The formula $TE = I(S_{t-\tau}; S_t | S_{t-1})$ was used to isolate the contribution of deep history (t-tau) to current

state (t), excluding the effect of immediate past (t-1) - We performed 500 measurement steps for each state to ensure statistical significance

Our analysis revealed that information propagation distance scales linearly with avalanche size, and peak TE corresponds precisely to the onset of power-law avalanche distributions, confirming the optimal information processing at criticality.