

Memory-induced long-range order drag

Yuan-Hang Zhang,^{1,*} Chesson Sippling,^{1,†} and Massimiliano Di Ventra^{1,‡}

¹*Department of Physics, University of California San Diego, La Jolla, CA 92093*

Recent research has shown that memory, in the form of slow degrees of freedom, can induce a phase of long-range order (LRO) in locally-coupled fast degrees of freedom, producing power-law distributions of avalanches. In fact, such memory-induced LRO (MILRO) arises in a wide range of physical systems. Here, we show that MILRO can be transferred to coupled systems that have no memory of their own. As an example, we consider a stack of layers of spins with local feedforward couplings: only the first layer contains memory, while downstream layers are memory-free and locally interacting. Analytical arguments and simulations reveal that MILRO can indeed drag across the layers, enabling downstream layers to sustain intra-layer LRO despite having neither memory nor long-range interactions. This establishes a simple, yet generic mechanism for propagating collective activity through media without fine tuning to criticality, with testable implications for neuromorphic systems and laminar information flow in the brain cortex.

I. INTRODUCTION

Long-range order (LRO) is typically associated with systems tuned to a critical point, where correlations span the entire size and collective dynamics emerge [1–3]. Recent work, however, has shown that LRO can also occur away from criticality when memory—in the form of slow, auxiliary degrees of freedom—interacts with fast, locally coupled variables [4]. This memory-induced LRO (MILRO) produces robust power-law correlations in the fast degrees of freedom, and has been observed in diverse contexts including spin-glass models [4], quantum spin chains [5], neuromorphic devices [6], models of neural activity [7], memcomputing machines [8, 9], and feedback-induced quantum phase transitions [10]. In these systems, slow memory variables act as an effective long-range interaction *in time*, inducing spatial LRO in the fast degrees of freedom over a wide range of parameters, and offering a mechanism for collective dynamics that does not require fine parameter tuning.

It is worth noticing that in all the above examples, the memory degrees of freedom share the same spatial region as the fast degrees of freedom. An open question is then whether such MILRO can propagate even beyond spatial regions of the system where the memory resides. Answering this question is not just an academic exercise. Rather, it is motivated by the fact that in many natural and engineered systems, functional units are arranged in modules or layers that interact through local couplings. A prototypical example is the brain architecture, where the neocortex is organized into distinct layers with structured feedforward and feedback connectivity, enabling hierarchical and laminar information processing [11–14]. If a subset of these layers contains memory-rich dynamics, hence possibly supporting MILRO, can their order “drag” downstream into memory-free layers,

allowing the entire stack to sustain collective, scale-free activity? Furthermore, could this effect explain the experimentally observed scale-free dynamics in the firing of neurons [15]?

Here, we address these questions by considering a stack of spin-glass layers with unidirectional, local feedforward couplings; see Fig. 1. Only the base layer contains slow memory variables; all other layers evolve via memory-free spin-glass dynamics with short-range in-plane interactions. We find that, for a wide range of parameters, the MILRO in the driving layer propagates to deeper layers, which themselves exhibit robust intra-layer LRO despite lacking memory and long-range couplings.

We term this phenomenon (memory-induced) *long-range order drag* in analogy with the Coulomb drag [16–18]—where momentum transfer between nearby electron layers induces a current in the “passive” layer—and the spin drag, where spin currents in one spin species induce currents in another via some interactions [19, 20]. In our case, the “dragged” quantity is not a current but *an entire ordered state*, transmitted through local couplings.¹

This mechanism suggests a general route for propagating collective dynamics across layered media without fine tuning to criticality. Beyond its theoretical interest, LRO drag may offer design principles for multi-layer neuromorphic architectures [21–23] and provide a potential explanation for sustained inter-laminar correlations in cortical circuits [11–14].

II. ANALYTICAL UNDERSTANDING OF THE MILRO DRAG

As anticipated, in order to study this phenomenon within the simplest system possible, we consider the

* yuz092@ucsd.edu; Equal contributions.

† csippling@ucsd.edu; Equal contributions.

‡ [email: diventra@physics.ucsd.edu](mailto:diventra@physics.ucsd.edu)

¹ We point out that we expect such a phenomenon to occur even if the origin of LRO in the base layer is not due to memory, and such LRO “seed” is not destroyed by the interaction with the other layers.

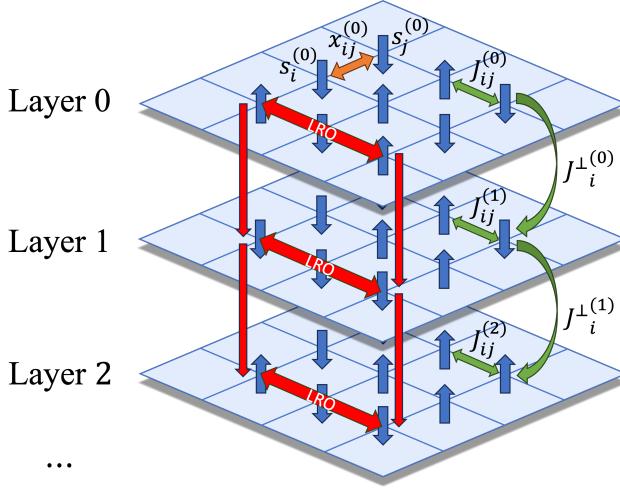


FIG. 1. Schematic of the memory-induced LRO drag phenomenon in our system. In all layers, continuously relaxed spins interact via traditional spin-glass interactions. In the base layer $k = 0$ layer, spins also experience memory interactions, where the memory x_{ij} is an auxiliary dynamical field which depends on the state of neighboring spins $s_i^{(0)}$ and $s_j^{(0)}$ and their mutual interaction $J_{ij}^{(0)}$. In deeper layers, memory does not appear, but instead, unidirectional, inter-layer interactions exist which enable the memory-induced LRO in layer 0 to be “dragged” through deeper layers.

model system depicted in Fig. 1. Each layer contains continuously relaxed spins with either ferromagnetic or anti-ferromagnetic interactions with their intra-layer neighbors. In the base layer, an additional memory degree of freedom (DOF) exists. Each memory DOF is defined on a lattice edge and tracks the recent history of its corresponding pair of spins; namely, it slowly grows or decays based on the degree of satisfaction of the relevant spin-glass interaction. In all other layers, these memories are not present. Instead, a unidirectional spin-glass interaction between neighboring *inter*-layer spins does exist, which enables layers that are closer to the base to dynamically influence deeper ones.²

A. Model dynamics

Mathematically, the above system can be modeled by the following dynamical equations, which are an extension of the ones used in Ref. [4]:

$$\begin{aligned}\dot{s}_i^{(0)} &= \sum_{\langle ij \rangle} J_{ij}^{(0)} s_j^{(0)} - g \sum_{\langle ij \rangle} x_{ij} s_i^{(0)}, \\ s_i^{(k)} &= \sum_{\langle ij \rangle} J_{ij}^{(k)} s_j^{(k)} + J_i^{\perp(k)} s_i^{(k-1)} \quad (k \geq 1), \\ x_{ij} &= \gamma(C_{ij} - \delta), \quad C_{ij} = (J_{ij}^{(0)} s_i^{(0)} s_j^{(0)} + 1)/2.\end{aligned}\quad (1)$$

Above, each 2D layer $k \geq 0$ consists of the standard, nearest-neighbor spin-glass interactions $J_{ij}^{(k)} = \pm J$, uniformly randomized in each layer. In the base layer $k = 0$ *only* do we introduce an additional degree of freedom (DOF) x_{ij} which grows or decays based on the state of its two accompanying spins, $s_i^{(0)}$ and $s_j^{(0)}$. In all other layers $k \geq 1$, there is an additional unidirectional ferromagnetic/anti-ferromagnetic interaction $J_i^{\perp(k)} = \pm J^{\perp}$ (again, assigned randomly) which couples $s_i^{(k)}$ to $s_i^{(k-1)}$. We clamp these dynamics so that $s_i^{(k)} \in [-1, 1]$ and $x_{ij} \in [0, 1]$. As in Ref. [4] and unless otherwise specified, we fix $\delta = 3/4$, $g = 2$ and $J = 1$ for the remainder of this work. A schematic representing this model can be found in Fig. 1.

B. MILRO Drag

Crucially, in the regime where γ is small relative to the other characteristic frequencies of the system, x_{ij} acts as a slowly-varying memory DOF which tracks the recent history of $s_i^{(0)}$ and $s_j^{(0)}$. It is precisely this memory DOF which has already been shown to induce LRO in the base layer $k = 0$ considered as a stand-alone layer [4]. Now, let us apply the same iterative integration approach used in [4] to Eqn. 1. We will work in the limit where the frequency of the memory DOFs, γ , is much smaller than all other characteristic frequencies in the problem (i.e., $\gamma \ll g$, $\gamma \ll J$, and $\gamma \ll J^{\perp}$) to infer how, under the appropriate conditions, long-range *intra*-layer correlations between high- k spins can be induced.

First, we reiterate the results of the previous work [4] establishing MILRO in a single spin-glass layer with memory. The iterative integration scheme can be performed on the first layer in isolation (due to the unidirectional nature of the inter-layer interactions), on both the memory DOFs (from Eqn. 1):

$$\begin{aligned}x_{ij}(1/\gamma) &\approx x_{ij}(0) + \int_0^{1/\gamma} dt \gamma(C_{ij} - \delta), \\ &\approx x_{ij}(0) + (J_{ij}^{(0)} \overline{s_{i,0}^{(0)} s_{j,0}^{(0)}} + 1)/2 - \delta,\end{aligned}\quad (2)$$

and the spins:

² For simplicity, in this work we consider only a *feedforward* inter-layer interaction, inspired by feedforward neural networks [24]. However, we expect such a phenomenon to also occur if feedback connectivity is present, provided this feedback does not destroy the base MILRO.

$$\begin{aligned}
s_i^{(0)}(1/\gamma) &\approx s_i^{(0)}(0) + \int_0^{1/\gamma} dt \left(\sum_{\langle ij \rangle} J_{ij}^{(0)} s_j^{(0)} - g \sum_{\langle ij \rangle} x_{ij} s_i^{(0)} \right), \\
&\approx s_i^{(0)}(0) - \frac{g}{\gamma} \sum_{\langle ij \rangle} x_{ij}(0) \overline{s_{i,0}^{(0)}} + \dots, \\
s_i^{(0)}(2/\gamma) &\approx s_i^{(0)}(1/\gamma) - \frac{g}{\gamma} \sum_{\langle ij \rangle} x_{ij}(1/\gamma) \overline{s_{i,1}^{(0)}} + \dots, \\
&\approx -\frac{g}{2\gamma} \overline{s_{i,0}^{(1)}} \sum_{\langle ij \rangle} J_{ij}^{(0)} \overline{s_{i,0}^{(0)}} \overline{s_{j,0}^{(0)}} + \dots,
\end{aligned} \tag{3}$$

where \bar{f} denotes the average of some quantity f over the interval $t \in [l/\gamma, (l+1)/\gamma]$. As this process is iterated further, additional, strong couplings begin to emerge between spins further and further from each other in the lattice. This induces an *effective* long-range interaction which emerges due to the presence of memory alongside the explicitly local spin-glass interactions [4].

Next, let us consider how $s_i^{(k)}$ evolves in time for $k \geq 1$, concentrating on its correlation to spins in other layers at the same lattice position i . After a duration of time $T = 1/\gamma$, we can approximate $s_i^{(k)}(1/\gamma)$ as follows:

$$\begin{aligned}
s_i^{(k)}(1/\gamma) &= s_i^{(k)}(0) + \int_0^{1/\gamma} dt \left(\sum_{\langle ij \rangle} J_{ij}^{(k)} s_j^{(k)} + J_i^{\perp(k)} s_i^{(k-1)} \right), \\
&\approx s_i^{(k)}(0) + \frac{1}{\gamma} \left(\sum_{\langle ij \rangle} J_{ij}^{(k)} \overline{s_{j,0}^{(k)}} + J_i^{\perp(k)} \overline{s_{i,0}^{(k-1)}} \right).
\end{aligned} \tag{4}$$

Importantly, when J^\perp is sufficiently strong, the unidirectional inter-layer interactions will begin to dominate over the intra-layer ones. The necessary strength of the J^\perp interaction which yields strong, inter-layer correlations will depend heavily on the degree of frustration within each spin-glass layer. In the most extreme case, where all of a spin's intra-layer neighbors are in energetic agreement with the corresponding spin-glass interaction $J_{ij}^{(k)}$ (i.e., where $J_{ij}^{(k)} s_j^{(k)} = 1$ for all nearest-neighbors j of i), we must have $J^\perp/J > 4$ for a flip in layer $k-1$ to induce a flip at the corresponding site in layer k . Only then can the one inter-layer interaction overpower the four nearest-neighbor intra-layer interactions. However, in general, not all spins will necessarily align in this manner. We could also have a net of two spins being energetically aligned (the other two canceling each other out), a net zero alignment (complete cancellation), a net of two anti-aligning, or all four anti-aligning.

Since $J_{ij}^{(k)} = \pm J$ are themselves chosen in a uniformly random manner (and because the spins spend most of their time at extremal values $s_i^{(k)} = \pm 1$; i.e., they flip rapidly relative to other dynamical timescales), let us assume the four nearest-neighbor $J_{ij}^{(k)} s_j^{(k)}$'s are drawn from

a uniform distribution $\{-J, J\}$. In this case, the *strength* of the net intra-layer interaction $|J_{ij}^{(k)} s_j^{(k)}|$ takes values $\{0, 2J, 4J\}$ with probabilities $\{6/16, 8/16, 2/16\}$. This gives a mean strength of $\langle |\sum_{ij} J_{ij}^{(k)} s_i^{(k)}| \rangle = 3J/2$. Thus, $J^\perp/J > 3/2$ may be a more reasonable lower bound beyond which inter-layer interactions dominate at a majority of sites, if $s_i^{(k-1)} = \pm 1$ as well.

Of course, the exact lower bound of J^\perp over which inter-layer correlations can easily spread to deeper layers, *propagating the entire collective state*, should be slightly higher than this value. That is because we anticipate some supermajority ($\gg 50\%$) of sites must be dominated by the inter-layer interaction in order for *system-wide* avalanches to propagate between layers; otherwise, the maximum number of spins which can participate in an avalanche will be capped far below the system size L^2 . This will also depend on the system's dynamics and the precise distribution of $J_{ij}^{(k)}$ in each lattice. We will indeed confirm that this is the case in our numerical experiments reported in Sec. III.

Regardless, it is clear that when J^\perp is sufficiently large, we expect that $s_i^{(k)}$ and $s_i^{(k')}$ will begin to correlate strongly after some transient duration, even at moderate inter-layer distances $|k - k'| > 1$. Therefore, after a sufficiently large number N of timesteps of duration $1/\gamma$, when the inter-layer interaction is strong enough, a spin $s_i^{(k)}(N/\gamma)$ will correlate strongly with the average value of the spin $s_i^{(k-1)}(t)$ in the interval $t \in [0, N/\gamma]$:

$$\begin{aligned}
s_i^{(k)}(N/\gamma) &\approx \frac{J_i^{\perp(k)}}{\gamma} \left(\overline{s_{i,N-1}^{(k-1)}} + \overline{s_{i,N-2}^{(k-1)}} + \dots + \overline{s_{i,0}^{(k-1)}} \right) + \dots \\
&\approx \frac{J_i^{\perp(k)}}{\gamma} \overline{s_i^{(k-1)}}.
\end{aligned} \tag{5}$$

Iterating this reasoning, we can infer that $s_i^{(k)}$ should also correlate with the corresponding spin in the *base* layer $s_i^{(0)}$, albeit more weakly than with the spin $s_i^{(k-1)}$ in the adjacent layer.

Now, if this result is combined with the fact that the memory interactions induce long-range order in the base $k = 0$ layer, it is now possible for the spins in layers $k \geq 1$ to correlate with each other over long distances due to their mutual correlation with the base layer. Essentially, after MILRO is established within spins in the base layer, the strong inter-layer coupling "drags" this collective state to deeper and deeper layers. Altogether, this constitutes a *multi-step* correlation process whereby LRO is transferred from a base layer *with* memory to additional layers *without* memory.

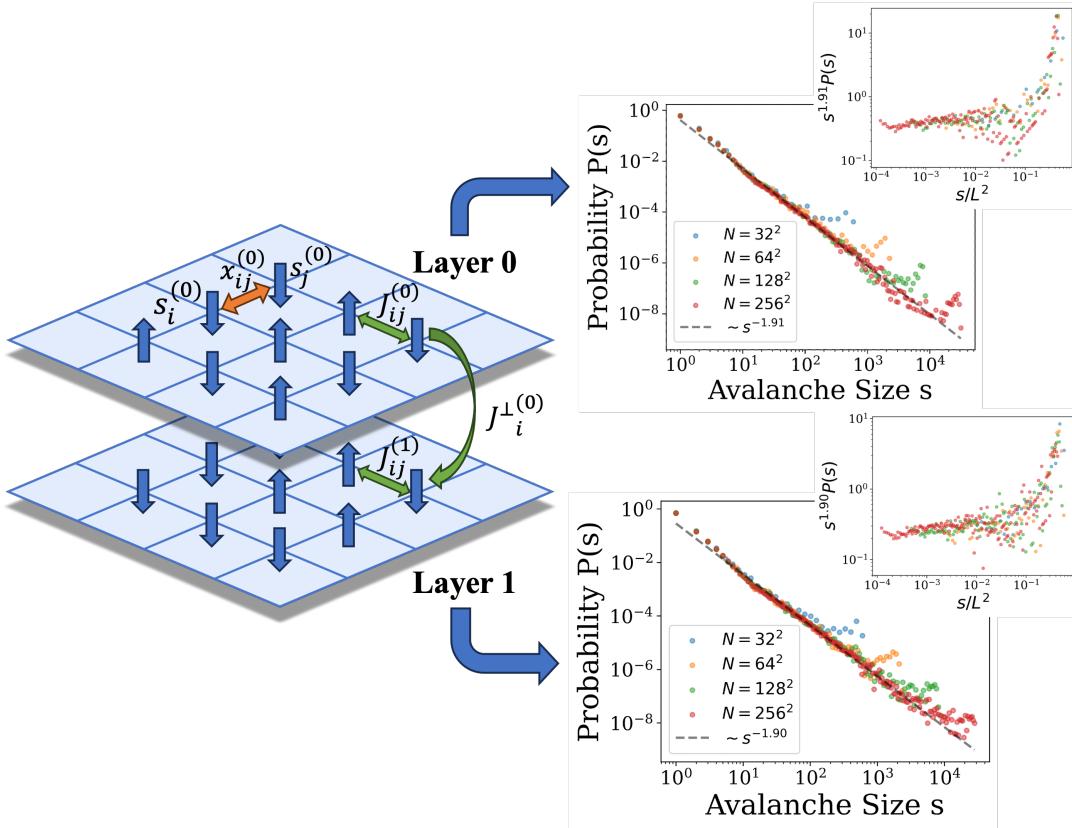


FIG. 2. Schematic of two layers of continuously relaxed spins alongside avalanche size s probability distributions $P(s)$. We notice that, in both layers, scale-free avalanche distributions exist which collapse well under finite-size scaling (shown in the insets). Furthermore, the scaling exponents in these distributions are nearly identical. This suggests that the collective, LRO state is effectively transferred from the base layer $k = 0$ to the $k = 1$ layer due to the inter-layer interactions. Both layers are simulated over 100 instances for $T = 200$ simulation time (a.u.), with $\gamma = 0.2$ and $J^\perp = 3.5$.

III. NUMERICAL VERIFICATION OF THE MILRO DRAG

We now demonstrate the existence of MILRO drag numerically using avalanche statistics, supporting the previous theoretical analysis. Avalanches are defined as spatiotemporal bursts of spin flips separated by quiescent intervals; the size s is the total number of flips in a burst. More numerical details are reported in Appendices B and C.

A. Two-layer test and finite-size scaling

We first simulate Eqn. 1 on square lattices of size $L = \{32, 64, 128, 256\}$ for the driving layer ($k = 0$) and the first downstream layer ($k = 1$); see Fig. 2. The distributions $P_k(s)$ of avalanche size s exhibit clear power

laws over several decades, with finite-size characteristic bumps at large $s \lesssim L^2$. Assuming scale invariance,

$$P_k(s; L) = s^{-\tau_k} f_k\left(\frac{s}{L^D}\right), \quad (6)$$

should hold [1]. Using the τ_k extracted from Fig. 2, we do indeed obtain excellent collapses of $s^{\tau_k} P_k(s)$ versus s/L^D with $D = 2$ for both $k = 0$ and $k = 1$ (insets of Fig. 2). This shows that the downstream, memory-free layer exhibits intra-layer scale-free avalanches with an area-controlled cutoff, which is a direct evidence of LRO drag.

B. Phase diagram

Having shown that MILRO can drag to an adjacent layer from the base one, we now want to understand

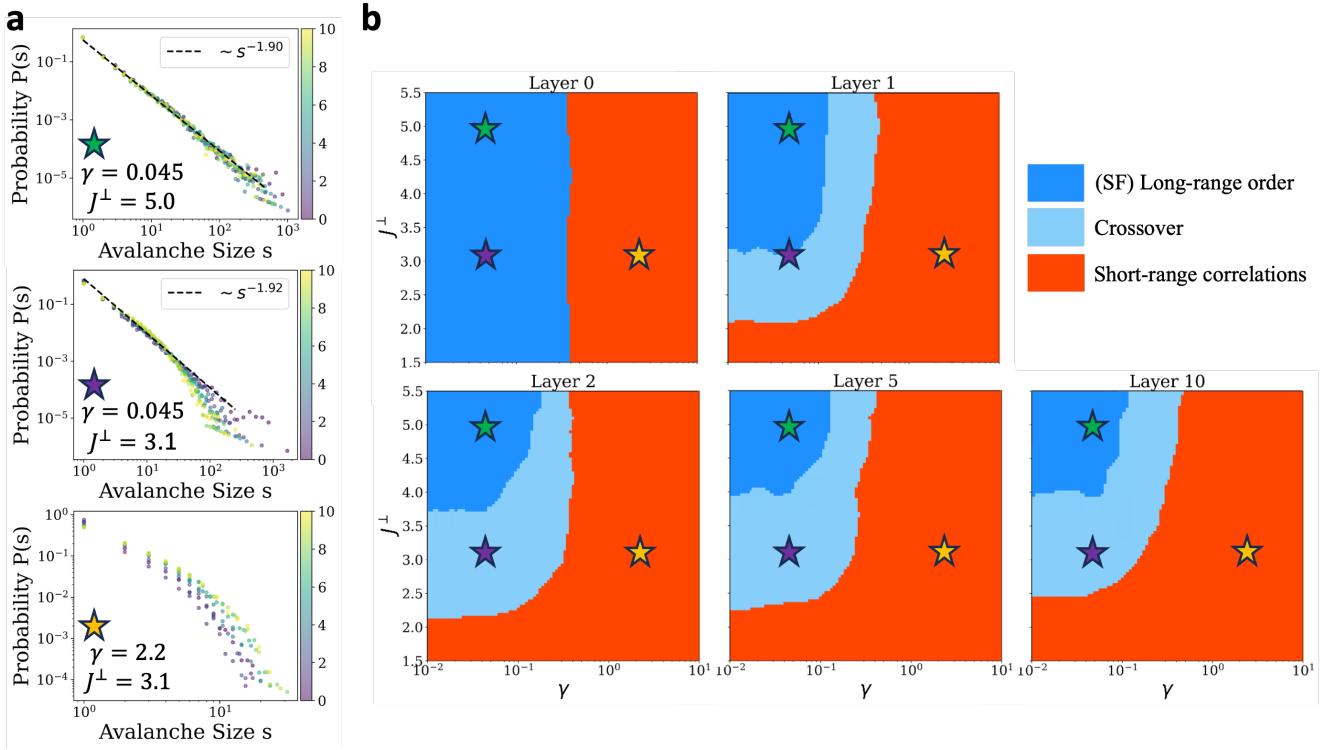


FIG. 3. Depth-resolved phase structure and representative avalanche statistics. (a) Avalanche size distributions $P(s)$ for $\{\gamma, J^\perp\} = \{0.045, 5.0\}, \{0.045, 3.1\}, \{2.2, 3.1\}$ (stars correspond to locations in the phase diagram in panel b). Dashed lines indicate power-law guides with exponents $\gamma \simeq 1.90$ and 1.92 , respectively. (b) Phase diagrams in the $\{\gamma, J^\perp\}$ plane at depths $k = 0, 1, 2, 5, 10$. Colors denote long-range order, crossover, and short-range correlations. Stars mark the parameter points used in (a). All points in the parameter space are simulated over 25 instances for $T = 200$ simulation time (a.u.) (with initial transient dynamics removed) on a 64×64 lattice.

how deep that drag can go, and how it depends on the main parameters in the system, in particular the frequency of the slow (memory) degrees of freedom, γ , and the strength of the inter-layer coupling, $J_i^{\perp(k)}$. We then consider the base driving layer coupled to ten downstream layers. Scanning the memory frequency scale γ and feedforward strength J^\perp yields the depth-resolved phase structure in Fig. 3: (i) an LRO phase with scale-free avalanches in every layer; (ii) a crossover band where power laws persist but the range shrinks with depth; and (iii) a short-range phase with only small avalanches.

Representative cuts at $\{\gamma, J^\perp\} = \{0.045, 5.0\}$, $\{0.045, 3.1\}$, and $\{2.2, 3.1\}$ illustrate these regimes, with avalanche distributions shown in Fig. 3(a). When γ is sufficiently low (to introduce strongly correlated behavior in the base layer $k = 0$) and $J^\perp > 4$ (to ensure that spins in deeper layers are always highly correlated with analogous spins in the base layer), LRO is detected up to the deepest layers tested. The size of this region appears to saturate, so we anticipate that such LRO behavior should persist in this parameter regime to arbitrarily deep layers. This corroborates our intuition from Sec. II B; once $J^\perp/J > 4$, inter-layer interactions will dominate over intra-layer ones at every site, enabling LRO to be dragged

to arbitrary depth without decay (if it exists in the base layer). On the other hand, when γ is small but J^\perp is not so large, LRO exists in the base layer, but inter-layer interactions must compete with intra-layer ones. In the regime where $2 < J^\perp/J < 4$, LRO can be dragged between layers, but the effectiveness of this drag phenomenon decays with depth. As more instances of frustration are encountered while moving through to deeper layers, correlations between the k^{th} layer and the base layer will weaken with increasing k . Finally, if γ is so large that LRO is never induced in the base layer at all, then the deeper layers only exhibit short-range correlations, as there is no collective state to “drag” at all.

The depth evolution reflects a simple competition. Memory in the base layer generates extended correlations; feedforward couplings transmit them downstream before they decay. In a band $J^\perp/J \in (2, 4)$ the inter-layer field is typically stronger than the net in-plane constraint at a site but does not fully dominate.³ In this regime, large-scale avalanches are still generally able to propa-

³ As anticipated in Sec. II B, the lower bound for J^\perp/J to observe LRO in deeper layers is indeed larger than $3/2$.

gate between layers, leading to wide phases of LRO which only shrink slowly when moving between subsequent layers (Fig. 3).

It is worth mentioning that even the crossover regime, which certainly lacks scale-freeness, can be considered to have an effective long-range interaction, induced by this drag effect (since avalanches with up to a few hundred participating spins exist). This is in stark contrast with the short-ranged region, in which avalanches above size 10 are already quite rare; see Fig. 3(a). Additionally, this crossover region appears to shrink quite slowly when comparing deeper layers.

Lastly, we must emphasize that for $J^\perp/J > 4$ the inter-layer term *always* dominates and downstream layers largely follow the base layer’s dynamics, causing inter-layer synchronizations. This is precisely why the LRO region saturates in size after a few layers (see the similarity between Layer 5 and 10 in Fig. 3(b)). This is also why we expect the phase diagram not to change much if we were to simulate even deeper layers. On the contrary, for $J^\perp/J < 2$ the system cannot overcome typical in-plane basins of attraction and layers quickly freeze. See Appendix A for a more rigorous analytical understanding of this point, and in Appendix D, we demonstrate the phase structures when we consider correlations within the entire three-dimensional stack instead of intra-layer only.

C. Transient build-up of order

Finally, to visualize how drag develops in time across layers, we first initialize each layer in its ground state and track the correlation length $\xi_k(t)$, defined as the mean in-plane separation of two events belonging to the same avalanche (excluding system-spanning avalanches to avoid trivial divergences, in analogy with percolation observables [25]) in layer k , from the start of the simulation up to time t . As shown in Fig. 4, for $\{\gamma, J^\perp\} = \{0.045, 5.0\}$ on a 64×64 lattice with $J = 1$, simulated over 100 instances for 4096 time steps, where each time step $\Delta t = 0.017$. All layers exhibit a rapid rise over $\sim 10^2$ time steps followed by a plateau. Mild overshoots and subsequent relaxation reflect the exclusion of system-wide events from ξ .

gion of parameters, the driven layers display power-law avalanche statistics with finite-size scaling indistinguishable from the driving layer, establishing MILRO drag as a

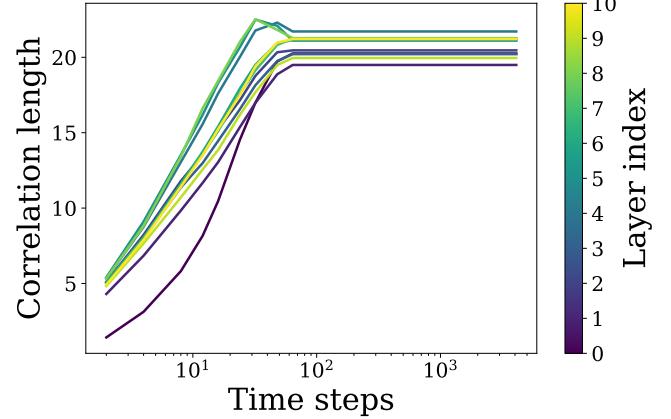


FIG. 4. Transient build-up of long-range order across layers. Time evolution of the in-plane correlation length $\xi_k(t)$ for the driving layer ($k = 0$) and ten downstream layers ($k = 1 - 10$) at $\{\gamma, J^\perp\} = \{0.045, 5.0\}$ on a 64×64 lattice with $J = 1$, simulated over 100 instances for 4096 time steps, where each time step $\Delta t = 0.017$. All layers exhibit a rapid rise over $\sim 10^2$ time steps followed by a plateau. Mild overshoots and subsequent relaxation reflect the exclusion of system-wide events from ξ .

robust, interaction-mediated propagation of order rather than a transport effect.

These results suggest practical principles for multilayer neuromorphic systems: a memory-rich “source” layer can broadcast coherent, scale-free dynamics to passive layers without fine tuning. This mechanism further provides a concrete hypothesis for inter-laminar correlations in the brain cortex: slow adaptive processes, confined to specific layers, can sustain depth-wide coherence through local pathways. Based on these results, we expect this MILRO drag effect could be not just a mechanism of theoretical interest but also a design tool for layered, adaptive matter and computation.

Data availability—The code for reproducing the results in this paper can be found at https://github.com/chessonipling/LRO_drag-main.

Acknowledgements.—We thank Kaushik Madapati for useful discussions. This work was funded by the National Science Foundation via grant No. ECCS-2229880. M.D. also acknowledges funding by the Alexander von Humboldt Stiftung through the Humboldt Research Award.

M.D. suggested and supervised the work. Y.-H.Z. and C.S. contributed equally to this work. Y.-H.Z. performed many numerical simulations, and C.S. the theoretical analysis and part of the numerical simulations. All authors have read and contributed to the writing of the paper.

IV. CONCLUSION

We have shown that slow memory degrees of freedom can seed long-range order (LRO) in a locally coupled layer and, via purely local feedforward couplings, drag this order into downstream layers that possess neither memory nor long-range interactions. Across a broad re-

-
- [1] U. C. Täuber, *Critical dynamics: a field theory approach to equilibrium and non-equilibrium scaling behavior* (Cambridge University Press, 2014).
- [2] G. Pruessner, *Self-organised criticality: theory, models and characterisation* (Cambridge University Press, 2012).
- [3] H. Hinrichsen, Non-equilibrium critical phenomena and phase transitions into absorbing states, *Advances in physics* **49**, 815 (2000).
- [4] C. Sippli, Y.-H. Zhang, and M. Di Ventra, Memory-induced long-range order in dynamical systems, *Physical Review E* **112**, 014124 (2025).
- [5] M. Weber, D. J. Luitz, and F. F. Assaad, Dissipation-induced order: the $s=1/2$ quantum spin chain coupled to an ohmic bath, *Physical Review Letters* **129**, 056402 (2022).
- [6] Y.-H. Zhang, C. Sippli, E. Qiu, I. K. Schuller, and M. Di Ventra, Collective dynamics and long-range order in thermal neuristor networks, *Nature Communications* **15**, 6986 (2024).
- [7] J. Sun, C. Sippli, Y.-H. Zhang, and M. Di Ventra, Memory in neural activity: long-range order without criticality, *arXiv preprint arXiv:2409.16394* (2024).
- [8] M. Di Ventra, *MemComputing: fundamentals and applications* (Oxford University Press, Oxford, 2022).
- [9] C. Sippli, Y.-H. Zhang, and M. Di Ventra, Phase-space engineering and dynamical long-range order in memcomputing, *arXiv preprint arXiv:2506.10149* (2025).
- [10] D. Ivanov, T. Y. Ivanova, S. Caballero-Benitez, and I. Mekhov, Feedback-induced quantum phase transitions using weak measurements, *Physical review letters* **124**, 010603 (2020).
- [11] K. D. Harris and G. M. Shepherd, The neocortical circuit: themes and variations, *Nature neuroscience* **18**, 170 (2015).
- [12] R. J. Douglas and K. A. Martin, Neuronal circuits of the neocortex, *Annu. Rev. Neurosci.* **27**, 419 (2004).
- [13] J. F. Mejias, J. D. Murray, H. Kennedy, and X.-J. Wang, Feedforward and feedback frequency-dependent interactions in a large-scale laminar network of the primate cortex, *Science advances* **2**, e1601335 (2016).
- [14] D. J. Felleman and D. C. Van Essen, Distributed hierarchical processing in the primate cerebral cortex., *Cerebral cortex* (New York, NY: 1991) **1**, 1 (1991).
- [15] J. M. Beggs and D. Plenz, Neuronal avalanches in neocortical circuits, *Journal of neuroscience* **23**, 11167 (2003).
- [16] B. Narozhny and A. Levchenko, Coulomb drag, *Reviews of Modern Physics* **88**, 025003 (2016).
- [17] A.-P. Jauho and H. Smith, Coulomb drag between parallel two-dimensional electron systems, *Physical Review B* **47**, 4420 (1993).
- [18] J. Li, T. Taniguchi, K. Watanabe, J. Hone, A. Levchenko, and C. Dean, Negative coulomb drag in double bilayer graphene, *Physical review letters* **117**, 046802 (2016).
- [19] I. D'Amico and G. Vignale, Theory of spin coulomb drag in spin-polarized transport, *Physical Review B* **62**, 4853 (2000).
- [20] C. P. Weber, N. Gedik, J. Moore, J. Orenstein, J. Stephens, and D. Awschalom, Observation of spin coulomb drag in a two-dimensional electron gas, *Nature* **437**, 1330 (2005).
- [21] C. Li, L. Han, H. Jiang, M.-H. Jang, P. Lin, Q. Wu, M. Barnell, J. J. Yang, H. L. Xin, and Q. Xia, Three-dimensional crossbar arrays of self-rectifying si/sio₂/si memristors, *Nature communications* **8**, 15666 (2017).
- [22] D. B. Strukov and R. S. Williams, Four-dimensional address topology for circuits with stacked multilayer crossbar arrays, *Proceedings of the National Academy of Sciences* **106**, 20155 (2009).
- [23] S. J. Kim, H.-J. Lee, C.-H. Lee, and H. W. Jang, 2d materials-based 3d integration for neuromorphic hardware, *npj 2D Materials and Applications* **8**, 70 (2024).
- [24] G. Bebis and M. Georgopoulos, Feed-forward neural networks, *Ieee Potentials* **13**, 27 (2002).
- [25] K. Christensen, Percolation theory, *Imperial College London* **1**, 87 (2002).
- [26] S. R. Bearden, F. Sheldon, and M. Di Ventra, Critical branching processes in digital memcomputing machines, *Europhysics Letters* **127**, 30005 (2019).

Appendix A: Analytical Derivation of Average Spin Flipping Time

To clarify the relevant timescales involved in our model, we will explicitly calculate the time Δt over which a single spin $s_i^{(k)}$ flips, in both the base layer $k = 0$ and deeper layers $k \geq 1$. We will assume that all neighbors of the flipping spin, $s_j^{(k)}$ and $s_i^{(k-1)}$, are constant over this time interval, and that $s_i^{(k)}$ changes approximately linearly over this flipping interval. Furthermore, without loss of generality, we will assume that $s_i^{(0)}$ flips from -1 to $+1$, so that $\Delta s_i^{(0)} = 2$. Note that the analysis would proceed identically, up to some factors of -1 , if this flip was from $+1$ to -1 .

First, since the change in a spin's magnitude must be 2, we can relate the flipping time Δt to the derivative $\dot{s}_i^{(k)}$:

$$\dot{s}_i^{(k)}(\Delta t) = s_i^{(k)}(0) + \dot{s}_i^{(k)} \cdot \Delta t \implies \dot{s}_i^{(k)} \cdot \Delta t = 2. \quad (\text{A1})$$

From here, we can use the functional form for $\dot{s}_i^{(k)}$, in Eqn. 1, to estimate Δt . We reiterate it here:

$$\begin{aligned} \dot{s}_i^{(0)} &= \sum_{\langle ij \rangle} J_{ij}^{(0)} s_j^{(0)} - g \sum_{\langle ij \rangle} x_{ij} s_i^{(0)}, \\ \dot{s}_i^{(k)} &= \sum_{\langle ij \rangle} J_{ij}^{(k)} s_j^{(k)} + J_i^{\perp(k)} s_i^{(k-1)} \quad (k \geq 1), \\ \dot{x}_{ij} &= \gamma(C_{ij} - \delta), \quad C_{ij} = (J_{ij}^{(0)} s_i^{(0)} s_j^{(0)} + 1)/2. \end{aligned} \quad (\text{A2})$$

Let us consider the base layer case, $k = 0$, first. There, the two types of interactions are the intra-layer spin-glass interactions and the memory interactions. Since x_{ij} and

$s_i^{(0)}$ are both dynamical, we can estimate their contributions to $\dot{s}_i^{(k)}$ by taking the average value of their product $\overline{x_{ij}s_i^{(0)}}$:

$$\dot{s}_i^{(0)} \approx \sum_{\langle ij \rangle} J_{ij}^{(0)} s_j^{(0)} - g \sum_{\langle ij \rangle} \overline{x_{ij}s_i^{(0)}}. \quad (\text{A3})$$

Since $s_i^{(k)}$ varies linearly in time (and we have assumed all nearest-neighbors s_j to be fixed during this interval of width Δt), $x_{ij}(t)$ will follow some parabolic trajectory, which can be solved for:

$$\begin{aligned} \dot{x}_{ij} &= \frac{\gamma}{2} J_{ij}^{(0)} \left(\frac{2}{\Delta t} t - 1 \right) s_j^{(0)} - \gamma(\delta - 1/2), \\ \Rightarrow x_{ij} &= \int dt \left(\frac{\gamma J_{ij}^{(0)} s_j^{(0)}}{\Delta t} t + \gamma \left(-\frac{J_{ij}^{(0)} s_j^{(0)}}{2} - \delta^* \right) \right), \\ &= \frac{\gamma J_{ij}^{(0)} s_j^{(0)}}{2\Delta t} t^2 + \left(-\frac{\gamma J_{ij}^{(0)} s_j^{(0)}}{2} - \gamma \delta^* \right) t + x_{ij}(0), \\ &\equiv \alpha t^2 + \beta t + x_{ij}(0). \end{aligned} \quad (\text{A4})$$

where $\delta^* \equiv \delta - 1/2$. Now that we have an expression for $x_{ij}(t)$, we can perform a simple integration to find $\overline{x_{ij}s_i^{(0)}}$:

$$\begin{aligned} \overline{x_{ij}s_i^{(0)}} &= \frac{1}{\Delta t} \int_0^{\Delta t} dt x_{ij} s_i^{(0)}, \\ &= \frac{1}{\Delta t} \int_0^{\Delta t} dt \left(\alpha t^2 + \beta t + x_{ij}(0) \right) \left(\frac{2}{\Delta t} t - 1 \right), \\ &= \frac{1}{\Delta t} \left(\frac{2\alpha}{4\Delta t} (\Delta t)^4 + \left(\frac{2\beta}{3\Delta t} - \frac{\alpha}{3} \right) (\Delta t)^3 \right. \\ &\quad \left. + \left(\frac{2x_{ij}(0)}{2\Delta t} - \frac{\beta}{2} \right) (\Delta t)^2 - x_{ij}(0)(\Delta t) \right), \\ &= \frac{\alpha}{6} (\Delta t)^2 + \frac{\beta}{6} (\Delta t), \\ &= \frac{\Delta t}{6} \left(\frac{\gamma J_{ij}^{(0)} s_j^{(0)}}{2} - \frac{\gamma J_{ij}^{(0)} s_j^{(0)}}{2} - \gamma \delta^* \right) = -\frac{\gamma \delta^* \Delta t}{6}, \end{aligned} \quad (\text{A5})$$

giving us a more concrete expression for $\dot{s}_i^{(0)}$ which now only depends on free parameters and Δt .

We can now write Eqn. A1 as a quadratic equation for Δt :

$$\begin{aligned} \frac{2}{\Delta t} &= \sum_{\langle ij \rangle} \left[J_{ij}^{(0)} s_j^{(0)} + \frac{g\gamma\delta^*\Delta t}{6} \right], \\ \Rightarrow 0 &= \frac{2g\gamma\delta^*}{3} (\Delta t)^2 + J_4(\Delta t) - 2, \\ \Rightarrow \Delta t &= \frac{3}{4g\gamma\delta^*} \left(-J_4 + \sqrt{J_4^2 + \frac{16g\gamma\delta^*}{3}} \right), \end{aligned} \quad (\text{A6})$$

where $J_4 \equiv \sum_{\langle ij \rangle} J_{ij}^{(k)} s_j^{(k)}$. As in the main text, the four nearest-neighbor $J_{ij}^{(0)} s_j^{(0)}$'s will each take values $\{-J, J\}$. Thus, $J_4 \equiv \sum_{\langle ij \rangle} J_{ij}^{(0)} s_j^{(0)}$ will take the values $\{-4J, -2J, 0, 2J, 4J\}$. In Fig. 5(a) we plot this flipping time Δt for each of these five cases as a function of γ , with $g = 2$, $\delta = 3/4$, and $J = 1$ fixed. As expected, Δt always decreases monotonically as a function of γ , since the faster the memory frequency γ , the faster $s_i^{(k)}$ will be able to flip. All cases tend towards the power law decay $\Delta t \sim \gamma^{-1/2}$ in the large- γ limit, when the memory interactions will dominate and the strength of the spin-glass interactions becomes irrelevant. Cases where $J_4 < 0$ follow $\Delta t \sim 1/\gamma$ for smaller γ . This makes sense, as these configurations would be stable in the absence of memories, so the slowly-varying memories must oppose the spin-glass interactions for a longer duration before a flip can be completed. On the other hand, $J_4 > 0$ cases tend towards constant Δt at small γ , as these flips are almost entirely instigated by the spin-glass interactions themselves if γ is not too large. The boundary case, $J_4 = 0$, corresponds to a flip in $s_i^{(0)}$ leaving the energy of the spin-glass configuration (ignoring the memory variables) unchanged. In this case, the solution to Eqn. A6 takes the simple form

$$\Delta t = \sqrt{\frac{3}{g\gamma\delta^*}}. \quad (\text{A7})$$

We can take this one step further by averaging over the entire lattice to find an average flipping time $\langle \Delta t \rangle$, considering that J_4 is itself a random variable on the lattice. As in the main text, we will draw the four nearest-neighbor $J_{ij}^{(0)} s_j^{(0)}$'s from the *uniform* distribution $\{-J, J\}$ so that J_4 takes the values $\{-4J, -2J, 0, 2J, 4J\}$ with probabilities $\{1/16, 4/16, 6/16, 4/16, 1/16\}$. This enables us to take the average value of any function of J_4 which appears in the numerical expression for Δt to find the average flipping time $\langle \Delta t \rangle$. See Fig. 6(a) for a plot of this average flipping time. Unsurprisingly, it is fitted well by a polynomial decay which begins near $\langle \Delta t \rangle \sim \gamma^{-1}$ for small γ and then gradually becomes more shallow until $\langle \Delta t \rangle \sim \gamma^{-1/2}$ for sufficiently large γ .

This procedure can also be applied to other layers with $k \geq 1$. A similar expression to Eqn. A1 holds, but this time $\dot{s}_i^{(k)}$ takes a different functional form:

$$\dot{s}_i^{(k)} = \sum_{\langle ij \rangle} J_{ij}^{(k)} s_j^{(k)} + J_i^{\perp(k)} s_i^{(k-1)}. \quad (\text{A8})$$

Again, we can now write an equation for Δt :

$$\frac{2}{\Delta t} = \sum_{\langle ij \rangle} J_{ij}^{(k)} s_j^{(k)} + J_i^{\perp(k)} s_i^{(k-1)} \implies \Delta t = \frac{2}{J_4 + J_1^{\perp}}, \quad (\text{A9})$$

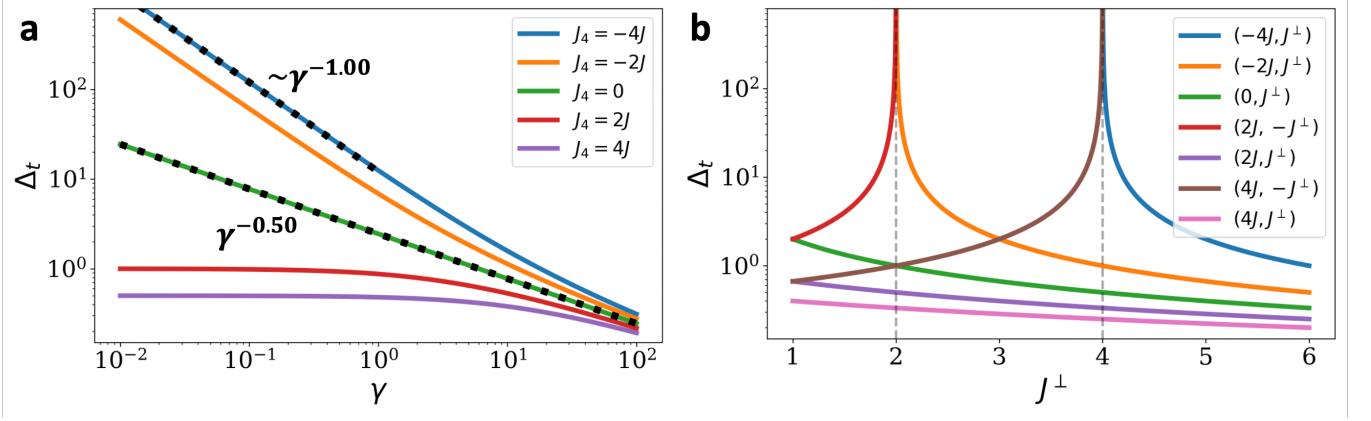


FIG. 5. The flipping time Δt over which a single spin $s_i^{(k)}$ flips, in both a) the base layer $k = 0$ and b) other layers $k \geq 1$. This timescale varies depending on the nature of the relevant spin-glass interactions. a) When $J_4 = 0$, $\Delta t = \sqrt{3/g\gamma\delta^*}$. When $J_4 < 0$, the flipping time decays as $\Delta t \sim 1/\gamma$, until $\gamma \sim O(1)$. When $J_4 > 0$, the flipping time remains constant at values near 1 for $J_4 = 2J$ and $1/2$ for $J_4 = 4J$, until $\gamma \sim O(1)$. b) Δt decays monotonically when both $J_4 > 0$ and $J_1^\perp > 0$. Divergences appear at $J^\perp = 2J$ and $J^\perp = 4J$ when the inter-layer and intra-layer spin-glass interactions are in opposition, and in these cases, flips can only occur within a particular range of J^\perp . When $J_4 < 0$ and $J_1^\perp < 0$, no flip will ever occur, so these curves cannot be plotted. Parameters chosen: $g = 2$, $\delta^* \equiv (\delta - 1/2) = 1/4$, $J = 1$.

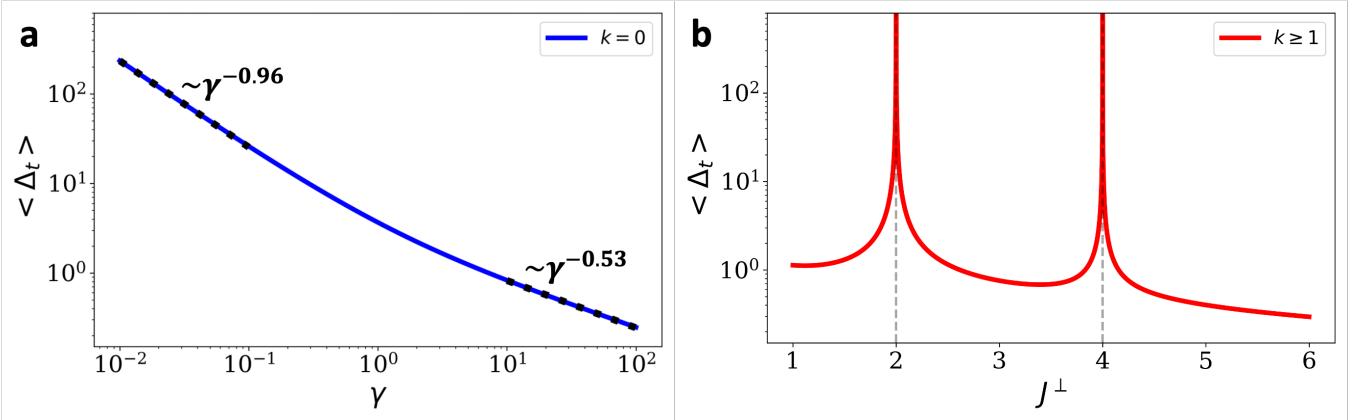


FIG. 6. The average time $\langle \Delta t \rangle$ over which a single spin $s_i^{(k)}$ flips, in both a) the base layer $k = 0$ and b) other layers $k \geq 1$. In both cases, all $\langle \Delta t \rangle$ shown in Fig. 5 are averaged according to the relative likelihood of occurrence, when both $J_{ij}^{(0)} s_i^{(0)}$ and $J_i^{\perp(k)} s_i^{(k-1)}$ are taken to be drawn from uniform distributions $\{-J, J\}$ and $\{-J^\perp, J^\perp\}$, respectively. a) The curve $\langle \Delta t \rangle$ is fitted to a polynomial decay for $\gamma \in [10^{-2}, 10^{-1}]$, yielding $\Delta t \sim \gamma^{-0.96}$, and $\gamma \in [10^1, 10^2]$, yielding $\Delta t \sim \gamma^{-0.53}$. If this plot were to be extended, $\langle \Delta t \rangle$ would tend further towards $\Delta t \sim \gamma^{-1}$ for lower γ and $\Delta t \sim \gamma^{-1/2}$ for higher γ . Parameters chosen: $g = 2$, $\delta^* \equiv (\delta - 1/2) = 1/4$, $J = 1$.

where J_4 is defined as above and $J_1^\perp \equiv J_i^{\perp(k)} s_i^{(k-1)} = \{-J^\perp, J^\perp\}$, with probabilities $\{1/2, 1/2\}$. We plot these flipping times as a function of J^\perp for each spin-glass configuration in Fig. 5(b). We notice quickly that the behavior of $\Delta t(J^\perp)$ depends strongly on the relevant spin-glass configuration. Cases $\{J_4, J_1^\perp\} = \{-4J, -J^\perp\}$, $\{-2J, -J^\perp\}$, or $\{0, -J^\perp\}$ are not plotted, as no flips would occur in any of these cases. When J_4 and J_1^\perp are instead strictly positive, Δt decays monotonically, and in these cases, its value is dramatically lower than in the base layer $k = 0$ for most values of γ . However, when one of J_4 and J_1^\perp is positive and one is negative, flip-

ping only occurs in specific regimes. Divergences appear at $J^\perp = 2J$ and $J^\perp = 4J$, corresponding to transition points beyond which either J_4 or J_1^\perp become too weak to initiate a flip. Interestingly, $J^\perp = 2J$ and $J^\perp = 4J$ both seem to correspond well with certain phase boundaries in Fig. 3. When $J^\perp < 2J$, the inter-layer interaction is consistently weaker than the intra-layer one, yielding short-range effective interactions in layers $k \geq 1$, regardless of the memory effect in layer $k = 0$. When $J^\perp > 4J$, a flip in an adjacent layer will always induce a flip relatively quickly, enabling the collective state to be dragged

to arbitrarily deep layer.⁴

Lastly, we can take the average value of the expression in Eqn. A9, given all possible values of J_4 and J_1^\perp , to produce an expression for $\langle \Delta t \rangle$. This average flipping time $\langle \Delta t \rangle$ as a function of J^\perp is plotted in Fig. 6(b), with $J = 1$ fixed. As expected, $\langle \Delta t \rangle$ is less than 1 when J^\perp is not too close to $2J$ or $4J$, but near these values, the flipping timescale diverges.

Appendix B: Numerical details

We simulate Eqn. 1 on k square layers of size $L \times L$ stacked along the feedforward direction. Unless stated otherwise, open boundaries are used in-plane and all parameters are homogeneous within a layer. Time evolution is integrated with a fixed-step, fourth-order Runge–Kutta scheme using $\Delta t = 0.048\gamma^{1/3}$, which resolves both the fast spin dynamics and the slow memory dynamics over the range of γ explored. To eliminate transients, we discard the first 10^3 steps and record the subsequent trajectory for avalanche analysis. For each $\{\gamma, J^\perp\}$, avalanches are collected over multiple disorder realizations of $\{J_{ij}^{(k)}, J_i^{\perp(k)}\}$ and independent random seeds.

An *avalanche* is a connected burst of spin flips in space and time. We detect flips by monitoring sign changes of $s_i^{(k)}(t)$; each flip is stored with its lattice site and timestamp. Two flips are deemed connected if they occur on nearest-neighbor sites and within a temporal proximity window Δt_w . We construct avalanches by recursively merging connected flips (single-linkage clustering on the event graph). Avalanche size s is the number of flips in the cluster. All distributions in Figs. 2 and 3 are obtained with this event-linked procedure.

The event-linked algorithm scales poorly at large L or when activity is dense, because the number of potential space-time links grows rapidly. For heavier workloads we therefore use a coarse-grained alternative: we discretize time into bins of width Δt_w and mark a site as “active” in bin n if at least one flip occurs at that site in $[n\Delta t_w, (n+1)\Delta t_w]$. This maps detection to a 3D (2D space + time) site-percolation problem with lattice spacing $(1, 1, \Delta t_w)$ and diagonal connections; avalanches correspond to connected clusters in this 3D grid.

Across all parameter sets tested, the two methods yield almost identical power-law regimes and consistent cutoffs within statistical error. However, explicit time binning breaks continuous time translation invariance and can introduce slightly different power-law exponents, as well as artifacts near the tail. We therefore report all main text results using the more physically reasonable event-linked

algorithm and reserve the binned 3D percolation method for large sweeps where its lower complexity is advantageous. See Fig. 7 for a comparison of these two avalanche extraction approaches.

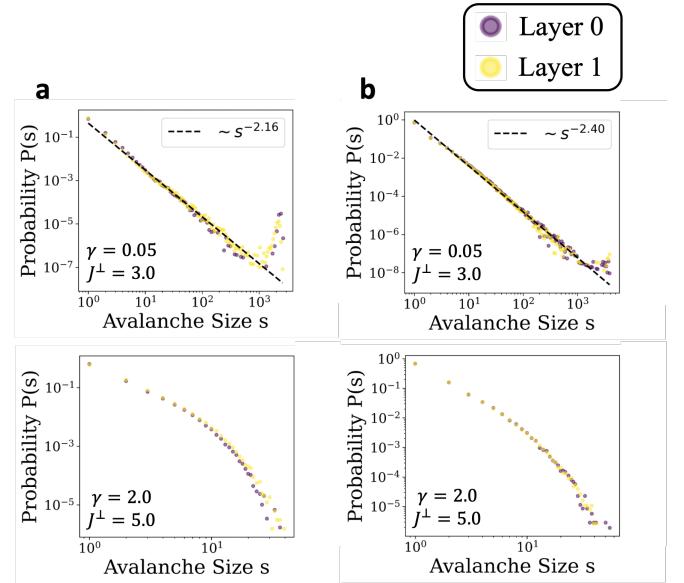


FIG. 7. Comparison of avalanche distributions extracted using the event-linked algorithm (a) and the coarse-grained alternative (b). Both extraction procedures produce similar phase structure, including both following power-law distributions in the LRO regime (e.g., $\{\gamma, J^\perp\} = \{0.05, 3.0\}$) and both only yielding small-scale avalanches in the short-range correlated regime (e.g., $\{\gamma, J^\perp\} = \{2.0, 5.0\}$). For each distribution, 1000 instances are simulated over $T = 200$ simulation time (a.u.) on a 64×64 lattice, with $\Delta_{tw}/dt = 75$.

Appendix C: Choosing the temporal proximity window

The avalanche definition involves a single free parameter, the temporal proximity window Δt_w . As with other common avalanche definitions (e.g., thresholding or fixed-time binning) [15, 26], results are typically robust over a finite range of this parameter. Here we describe our procedure for selecting Δt_w in a data-driven and physically motivated manner.

Because memory evolves on the slow timescale $1/\gamma$ and inter-layer transmission is controlled by J^\perp , the optimal Δt_w should be large enough to combine causally related flips within a burst, yet small enough to avoid merging distinct events. We therefore choose Δt_w to maximize an in-plane correlation length $\xi(\Delta t_w)$, defined as the mean spatial separation of two events within the same avalanche (excluding system-spanning avalanches to prevent a trivial divergence). This criterion favors windows that preserve the largest coherent structures without over-aggregation.

⁴ As mentioned in Sec. III, the size of the LRO region in phase space seems to saturate, persisting in the deepest layers simulated when $\gamma < 0.5$ and $J^\perp > 4$.

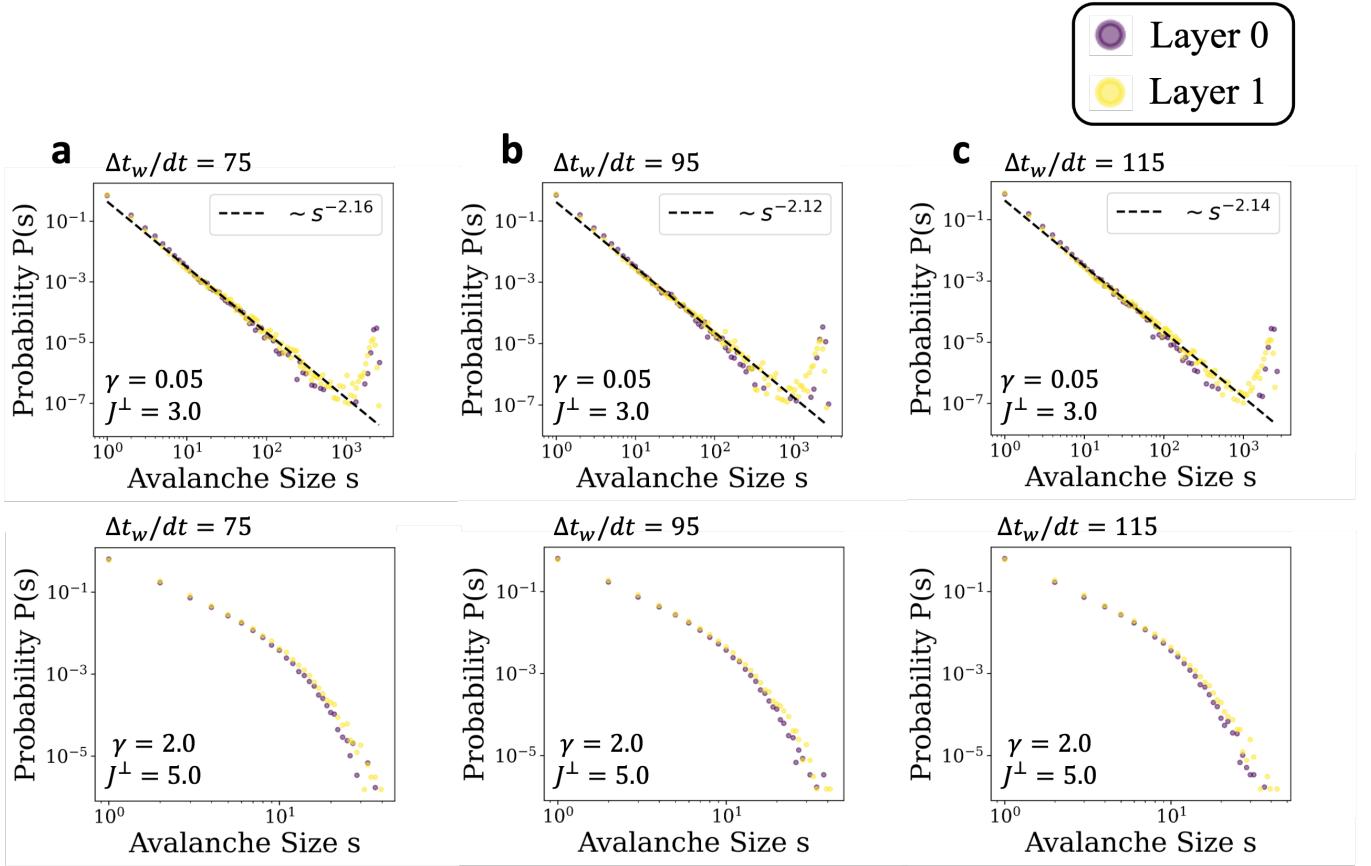


FIG. 8. Avalanche distributions over a variety of time windows. Avalanches are extracted at both $\{\gamma, J^\perp\} = \{0.05, 3.0\}$ and $\{2.0, 5.0\}$ with a maximum of 75, 95, and 115 integration timesteps (in subplots a, b, and c, respectively) between successive flips in a single avalanche (the number of integration timesteps can be scaled by the simulation timestep dt to find Δ_{tw} in physical, simulation time). This region represents a plateau in Δ_{tw} -space at which different time windows yield nearly identical distributions at both points in physical parameter space, suggesting the time window does not need to be fine-tuned to yield the distributions we observe. Avalanches are extracted over 1000 instances, each simulated for $T = 200$ simulation time (a.u.), on a 64×64 lattice.

Figure 8 shows $P(s)$ for several Δ_{tw} at $\{\gamma, J^\perp\} = \{0.05, 3.0\}$ (a point in the LRO regime) and $\{2.0, 5.0\}$ (a point in the short-range correlated regime) on a 64×64 lattice. Only very minor changes are visible in the distributions as Δ_{tw} is varied, suggesting that a broad plateau of near-optimal windows yields indistinguishable power-law regimes and essentially identical exponents. We observe similar plateaus at many other points in parameter space. We use the maximizer of $\xi(\Delta_{tw})$ within these plateaus to select an “optimal” window Δ_{tw} .

This optimal Δ_{tw} depends on $\{\gamma, J^\perp\}$: faster memory (γ large) typically favors smaller windows, whereas stronger drag (J^\perp large) can tolerate slightly larger windows due to tighter cross-layer coupling. For each parameter pair we sweep Δ_{tw} over a wide range, identify the maximizer of $\xi(\Delta_{tw})$, and verify that the inferred exponent τ is stable across the plateau.

The resulting optimal windows are summarized in Fig. 9. All phase-diagram measurements in Fig. 2 and Fig. 3 use these Δ_{tw} values.

Appendix D: Three-dimensional avalanche structures

The analysis in the main text focused on *in-plane* avalanches, i.e., clusters built from nearest-neighbor connectivity within a single layer. To probe *inter-layer* correlations directly, we extend the definition of an avalanche to *three dimensions*: two flip events are considered connected if they occur on nearest-neighbor sites in the full (x, y, k) lattice and within the temporal window Δ_{tw} (cf. Appendix B). Therefore clusters may percolate both laterally and across layers.

Fig. 10 shows the resulting $\{\gamma, J^\perp\}$ phase diagram together with representative distributions at three starred points: $\{0.1, 4.5\}$ (rigid), $\{0.1, 2.0\}$ (LRO), and $\{1.0, 2.0\}$ (crossover). Compared with the *in-plane* diagrams of Fig. 3, the main difference is the introduction of a vast rigid region. When J^\perp is large, feedforward coupling synchronizes flips across layers; many avalanches that are distinct within individual planes merge through the ver-

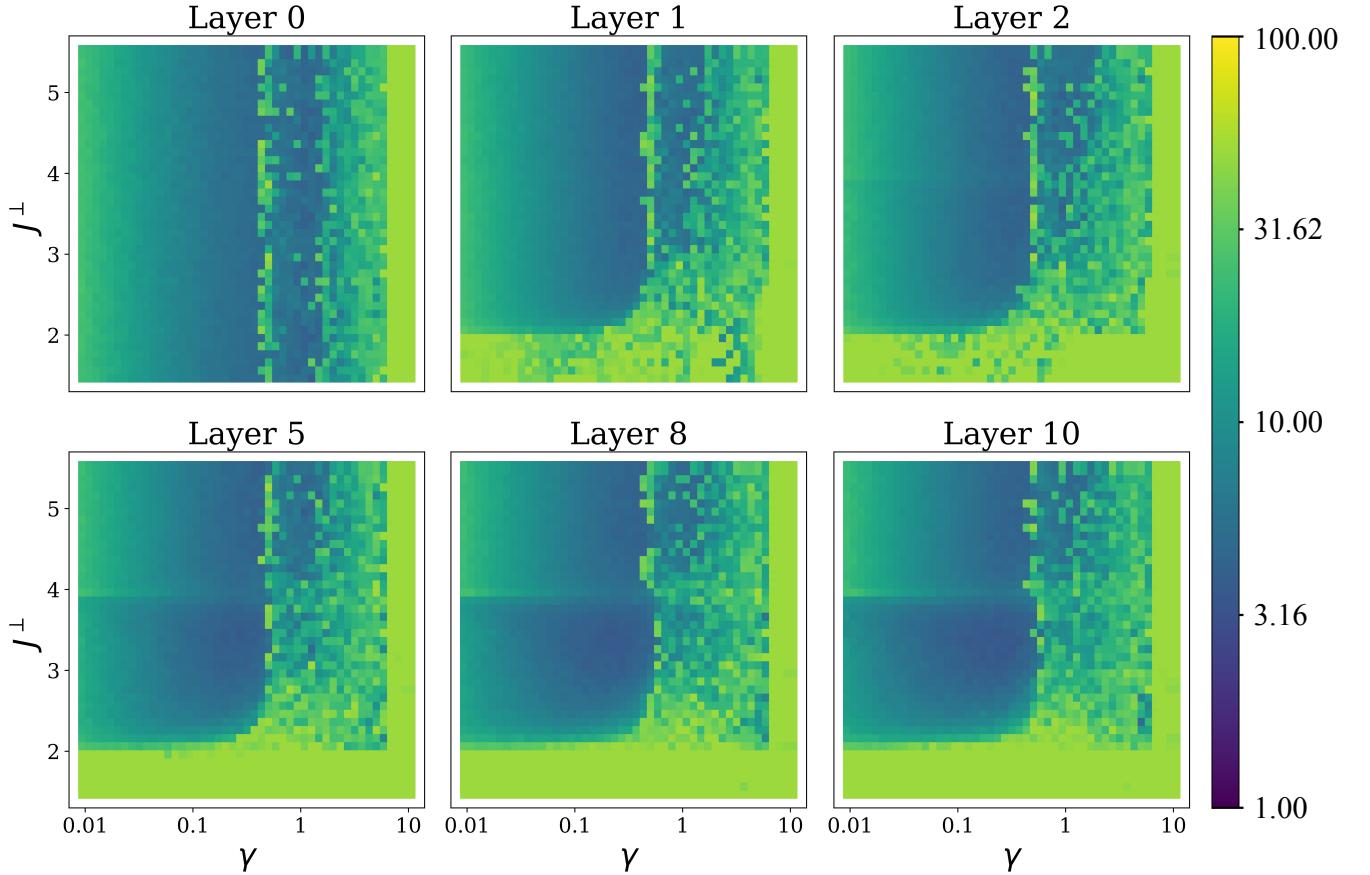


FIG. 9. The optimal temporal proximity windows Δt_w for different parameter configurations. Each data point is obtained by simulating the dynamics of 1 instance on a 64×64 lattice over 8000 time steps ($T = 8000\Delta t$, $\Delta t = 0.048\gamma^{1/3}$), extracting the avalanches using 20 different values of Δt_w and selecting the one that maximizes the correlation length. We can loosely see the phase structures presented in Fig. 3. We also note a line of large Δt_w at $J^\perp = 4$, corresponding to the diverging average spin flip time derived in Appendix A.

tical links into a single 3D cluster that spans the stack. Consequently, power-law regimes seen in 2D are replaced by distributions dominated by system-wide events in 3D.

At intermediate couplings, two trends appear: For $J^\perp \lesssim 2$, vertical links are too weak to fully synchronize layers. Avalanches remain predominantly lateral, and the 3D distributions recover a broad algebraic regime (LRO), similar to the in-plane case. For $J^\perp > 2$ but $\gamma \sim 1$, memory weakens while vertical links are moderately strong; partial inter-layer merging also produces power-law distributions of avalanches corresponding to LRO. As γ increases further, activity becomes sparse, producing a crossover regime which eventually gives way to short-range behavior when γ is too large.

Taken together, the 3D clusters make explicit the mechanism behind MILRO drag: vertical coupling links together correlated planes. When J^\perp is strong, this effect yields stack-spanning events (rigidity); when moderate, it transmits order without full synchronization (LRO/crossover); and when weak, layers decouple and only small clusters survive.

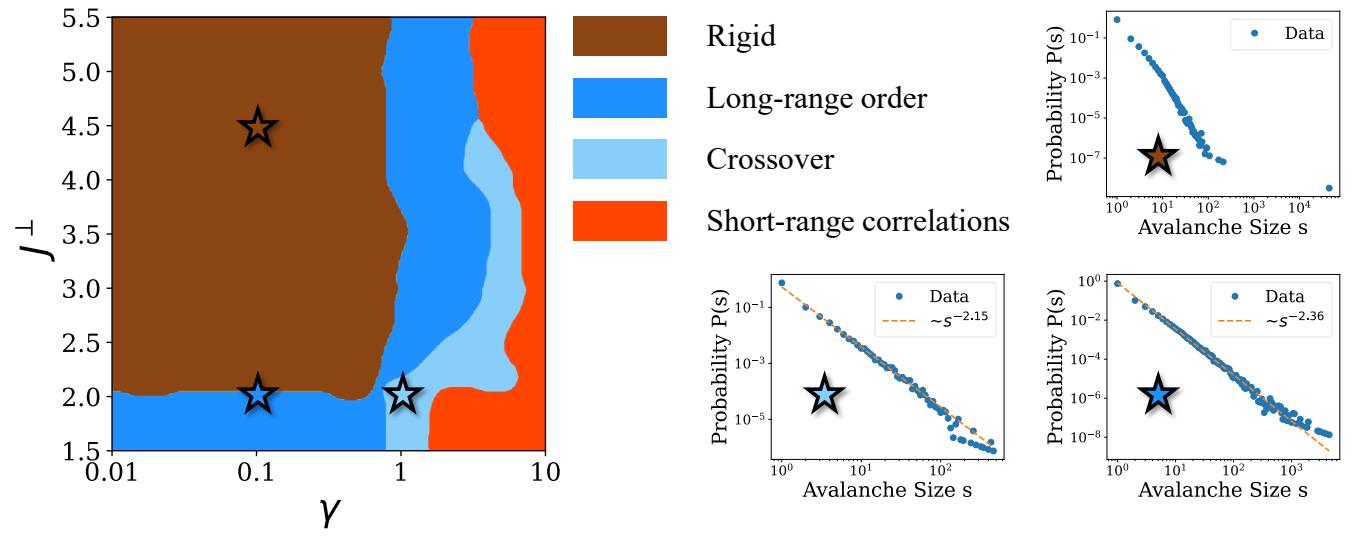


FIG. 10. Three-dimensional avalanche structure and phase map. Left: Phase diagram in the $\{\gamma, J^\perp\}$ plane when avalanches are defined with 3D nearest-neighbor connectivity in space (across (x, y, k)) and the temporal window Δt_w . Colors denote rigidity (stack-spanning avalanches), long-range order (LRO), a crossover regime, and short-range correlations. Compared to the in-plane analysis, a rigid region now exists because strong feedforward coupling merges layer-by-layer bursts into single 3D clusters. Right: Representative avalanche size distributions $P(s)$ at the starred points $\{0.1, 4.5\}$ (rigid; dominated by system-wide events), $\{0.1, 2.0\}$ (LRO; broad power-law regime), and $\{1.0, 2.0\}$ (crossover; truncated tails). Increasing γ suppresses activity, while decreasing J^\perp reduces inter-layer merging, recovering LRO-like statistics. Each data point is obtained using 1 instance simulated over 8000 time steps on a 64×64 lattice.