arXiv:2611.29880v1 [physics.bio-ph] 23 Nov 2026

神经动力学中的红外弛豫模与集体有序的涌现

Byung GyuChae[1]韩国大田市儒城区*Gajeong-ro 218*号，邮编*34129*，电子与电信研究院*

我们建立了一个描述相互作用的非线性神经系统中集体有序形成的微观场论。对微观神经自由度进行积分消去后，可得到一个有效作用量，其二次核决定有序阈值，而高阶连通响应则决定非线性稳定化。谱分解表明，弛豫率和集体模重叠控制着集体不稳定性与记忆动力学。增强的红外谱权重会降低临界耦合并产生长时记忆，揭示出集体有序与持久动力学具有共同的谱起源。受控数值构造验证了集体不稳定性与响应的谱依赖性，而非线性平均场标度则源自微观朗道展开。Transformer 雅可比谱进一步揭示了学习诱导的红外组织以及与类临界响应相符的瞬态记忆增强。我们的研究建立了一个将神经相互作用、集体有序和长时记忆联系起来的微观框架。

引言。— 长期以来，研究者一直通过神经元网络模型和统计力学方法研究集体神经动力学；实验研究揭示了皮层活动中的低维群体动力学和临界性特征 [1–4]。Wilson–Cowan 动力学及后续的神经场表述描述了相互作用的神经群体 [5–7], , 而 Hopfield 网络则阐明了微观相互作用如何产生集体吸引子和联想记忆 [8–10]。生成泛函和场论方法进一步刻画了非线性神经动力学、涨落、时间相关性和延迟相互作用 [11–15]。

一个不同的问题是，集体有序与神经活动的时间持续性有何关联。稳定的集体状态并不能唯一确定其弛豫谱，而仅有慢弛豫也不足以证明有序相的存在。对循环网络进行的谱分析和稳定性分析揭示了长时依赖、门控和慢动力学模的重要性 [16–20]。Chen 和 Bialek 进一步研究了随机神经网络中长时间尺度的出现，以及形成宽广慢模谱所需满足的约束 [21]。结构化状态空间模型和学习得到的神经架构的雅可比矩阵分析等相关进展，促使我们考察红外弛豫 [22, 23]。然而，仅凭特征值密度无法确定集体响应；还必须考虑模态与集体相互作用通道之间的重叠。

在此，我们建立一种微观集体场描述，将时间弛豫与集体有序联系起来。从相互作用的神经元出发，我们采用 MSRJD 响应形式和动态 Hubbard–Stratonovich 变换 [24–28] 推导出有效集体作用。其二次核决定集体易感性和失稳阈值，而高阶连通响应则产生非线性集体顶角。同一微观响应还提供了一个由观测量加权的红外弛豫谱所支配的因果记忆核。

我们表明，加权红外谱 $\rho_O(\lambda) \propto \lambda^{\beta}$ 在相应的标度区间内产生 $K(t) \propto t^{-(1+\beta)}$，而其逆速率谱矩决定了集体不稳定性。特别是，近乎平坦的红外谱将近似 $1/t$ 记忆核与有序阈值的对数红外敏感性联系起来。受控的非对角神经模型展示了这些关系及其相关的集体软化。这些结果通过共同的微观响应，将时间持续性与集体有序联系起来，同时区分了慢模积累与自发有序。

微观动力学与集体作用。—考虑 N 具有互易相互作用 $W_{ij} = W_{ji}$的真实神经活动 $x_i(t)$。按照基于连续状态能量的神经模型 [9], , 单调可逆的激活函数允许定义满足 H, $\sigma_i^{-1}\ x_i$ ( )

( )的局部势函数。我们分离出一个集体相互作用通道，$W_{ij} = W_{(0)\,ij} + Jc_ic_j/N$，其中 $J > 0$，而 $W_{(0)}$保留其余的微观相互作用。定义$H_0(\boldsymbol{x}) = \sum_i U_i(x_i)$ $-\frac{1}{2}\sum_{ij} W_{ij}^{(0)} x_i x_j$和 $M = \sum_i c_i x_i$，微观能量变为

$$H(\boldsymbol{x}) = H_0(\boldsymbol{x}) - \frac{J}{2N} M^2, \qquad m = M/N. \tag{1}$$

所选的秩一通道并不将完整连接限制为秩一。

我们考虑带有加性噪声的过阻尼梯度动力学，

$$\dot{x}_i = -\partial_i H_0(\boldsymbol{x}) + Jc_i m + \eta_i, \tag{2}$$

其中 $\langle \eta_i(t)\eta_j(s)\rangle = 2D\delta_{ij}\delta(t-s)$。互易模型的平稳分布为 $P_{\rm st} \propto e^{-H/D}$，前提是其可归一化。

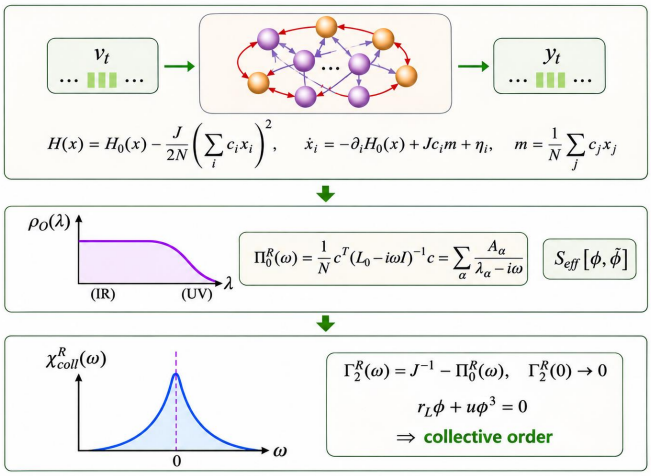


图1。集体神经动力学的微观框架。具有秩一集体耦合的高维神经动力学通过 MSRJD 和 Hubbard–Stratonovich公式产生弛豫模谱和有效集体场作用量。所得的集体响应由 $\Gamma_2^R(\omega) = J^{-1} - \Pi_0^R(\omega)$ 控制，其中不稳定性条件 $\Gamma_2^R(0) \to 0$ 标志着集体软化，并可能在非线性相互作用的稳定作用下导致宏观有序。

化。这一平衡关系并不会自动推广到一般的神经激活率动力学。

在实响应 Itô MSRJD约定 [24–26], 中，微观作用量为

$$S[\boldsymbol{x}, \tilde{\boldsymbol{x}}] = \int dt \sum_i \left[\tilde{x}_i(\dot{x}_i + \partial_i H_0) - D\tilde{x}_i^2\right] - \frac{J}{N}\int dt\, \widetilde{M}(t) M(t), \tag{3}$$

其中，$M = \sum_i c_i \tilde{x}_i$ 是集体响应变量。最后一项将集体活动与其响应对应变量耦合起来。

动力学 Hubbard–Stratonovich 变换[27, 28] 引入了集体驱动 ϕ 及其共轭的响应场 $\tilde{\phi}$。对完整的参考网络 ϕ 积分后得到

$$S_{\text{eff}}[\phi, \tilde{\phi}] = \frac{N}{J}\int dt\, \tilde{\phi}\phi - \ln\left\langle e^{\mathcal{A}}\right\rangle_0, \tag{4}$$

其中 $\mathcal{A} = \int dt\,(\tilde{\phi} M + \phi \widetilde{M})$，而 $\langle\cdots\rangle_0$ 表示相互作用参考网络 H_0 的归一化 MSRJD 测度，该测度不必分解为相互独立的神经元。辅助场积分路径实现了因果HS恒等式，并且在物理鞍点处有 $\phi = J\langle m\rangle$ 。

式（4）在路径积分层面上是形式上精确的。将对数按连通累积量展开，$\ln\langle e^{\mathcal{A}}\rangle_0 = \sum_{n=1}^{\infty}\langle\mathcal{A}^n\rangle_{0,c}/n!$ ，会生成一系列集体顶点。在参考 MSRJD 测度具有反射对称性的条件下，奇数总阶数项均为零。

二次扇区决定集体易感性和延迟记忆核，而四阶连通响应会生成非线性延迟顶点和涨落顶点。鞍点近似和高斯近似要求适当的大-N 标度及稳定性条件。

二次扇区：延迟集体核——对于满足$\langle m\rangle_0 = 0$的稳态反射对称参考态，由 H_0 支配的网络对集体驱动 $\phi\boldsymbol{c}$ 的延迟响应为

$$\Pi_0^R(t-s) = \left.\frac{\delta\langle m(t)\rangle_\phi}{\delta\phi(s)}\right|_{\phi=0} = \frac{1}{N}\langle M(t)\widetilde{M}(s)\rangle_{0,c}. \tag{5}$$

因果性要求 $\Pi_0^R(t-s) = 0$ 适用于 $t < s$。将式 (4) 展开至二次阶，得到延迟集体顶点 $\Gamma_2^R(\omega) = J^{-1} - \Pi_0^R(\omega)$。相应的高斯物理通道磁化率为

$$\chi_m^R(\omega) = \frac{\Pi_0^R(\omega)}{1 - J\Pi_0^R(\omega)}. \tag{6}$$

辅助场传播子与物理磁化率并不相同，但二者具有相同的高斯不稳定性分母。二次作用量还包含一个与连通相关函数$\langle M(t)M(s)\rangle_{0,\,c}$成正比的集体噪声顶点；仅凭延迟核无法确定涨落。

对称高斯鞍点在$J_c\Pi_0^R(0) = 1$处变为边缘稳定，等价地，$r_L \equiv J^{-1} - \Pi_0^R(0) = 0$。此条件确定的是平均场不稳定性，并不一定意味着连续相变。参考响应Π_0^R是针对完整非线性参考动力学定义的，而式 (6) 对高斯集体动力学是精确的，在其他情况下则代表高斯/RPA近似。

记忆核的微观起源。—在稳定定点 $\boldsymbol{x}_*$ 附近对确定性参考动力学进行线性化，并假设海森矩阵为正定，得到 $\delta\dot{\boldsymbol{x}} = -L_0\delta\boldsymbol{x} + \boldsymbol{c}\,\delta\phi$，其中 $L_0 = \nabla\nabla H_0(\boldsymbol{x}_*)$ 。对于互易梯度动力学，$L_0 = L_0^{\mathsf{T}} > 0$，其特征对为（$\lambda_\alpha$, $\boldsymbol{v}_\alpha$）。随后，集体响应可写成谱形式

$$\Pi_0^R(\omega) = \frac{1}{N}\boldsymbol{c}^{\mathsf{T}}(L_0 - i\omega I)^{-1}\boldsymbol{c} = \sum_\alpha \frac{A_\alpha}{\lambda_\alpha - i\omega}, \tag{7}$$

其中 $A_\alpha = |\boldsymbol{c}^{\mathsf{T}}\boldsymbol{v}_\alpha|^2/N \geq 0$。这些特征模描述的是整个 N 个神经元状态空间的弛豫，而非单个神经元的弛豫。

未加权的弛豫率密度为 $\rho(\lambda) = N^{-1}\sum_\alpha \delta(\lambda - \lambda_\alpha)$，而集体可观测量探测的是 $\rho_O(\lambda) = \sum_\alpha A_\alpha\delta(\lambda - \lambda_\alpha)$。由于集体通道不一定与每个模态都有重叠，这些分布可能具有不同的红外行为。慢模只有通过与相互作用通道的非零重叠，才会对集体不稳定性产生贡献

。加权谱决定了频率响应和因果时域核，

$$\Pi_0^R(\omega)=\int_0^\infty d\lambda\,\frac{\rho_O(\lambda)}{\lambda-i\omega},\quad K(t)=\theta(t)\int_0^\infty d\lambda\,\rho_O(\lambda)e^{-\lambda t}. \tag{8}$$

在高斯层面，由此得到的集体场满足线性化鞍点方程

$$\frac{1}{J}\phi(t)-\int_{-\infty}^{t} ds\,K(t-s)\phi(s)=0, \tag{9}$$

在不存在外部集体源时。因此，非局域核 $K(t)=\Pi_0^R(t)$ 是通过积分消去微观神经自由度而产生的，而非现象学地施加的。它是 $J^{-1}-\Pi_0^R$ 中本已包含的参考响应；其静态贡献不得再次作为独立自能计入。

对于有限噪声下的非线性参考动力学，精确的随机响应未必等于确定性雅可比矩阵的预解式。非互易网络需要考虑左右特征向量重叠，并可能呈现复弛豫模态。

红外标度与集体阈值。— 假设存在正的加权谱 $\rho_O(\lambda)=A\lambda^\beta$，其范围为 $\lambda_{\rm IR}\le\lambda\le\Lambda$，且 $A>0$ 和 $\beta>-1$，并且该窗口之外的贡献可以忽略。对于 $\Lambda^{-1}\ll t\ll\lambda^{-1}$ 红外，记忆核满足

$$K(t)\simeq A\Gamma(1+\beta)t^{-(1+\beta)}. \tag{10}$$

对于边缘指数 $\beta=0$，有限窗口表达式为 $K(t)=A(e^{-\lambda_{\rm IR}t}-e^{-\Lambda t})/t$，在中间区间内得到 $K(t)\sim A/t$ 。

同一个加权谱通过 $J_c^{-1}=\Pi_0^R(0)=A\int_{\lambda_{\rm IR}}^{\Lambda}d\lambda\,\lambda^{\beta-1}$ 决定高斯有序阈值，从而得到

$$J_c^{-1}=\begin{cases}\dfrac{A}{\beta}(\Lambda^\beta-\lambda_{\rm IR}^\beta), & \beta\neq 0,\\ A\ln(\Lambda/\lambda_{\rm IR}), & \beta=0.\end{cases} \tag{11}$$

在固定 A 和 Λ 时，减小 $\lambda_{\rm IR}$ 会导致：对于 $-1<\beta<0$，$J_c\propto\lambda^{-\beta}$ 红外；对于 $\beta=0$，为 $J_c\propto[\ln(\Lambda/\lambda_{\rm IR})$；而对于 $\beta>0$，则得到一个有限的极限阈值。

边缘情形 $\beta=0$ 将幂律红外发散与有限的静态响应分隔开来，并在边界处出现对数增强。因此，$1/t$ 记忆机制和静态响应的对数红外增强源自同一个加权弛豫谱。有限截断会使这些渐近行为正则化，而谱归一化或模态权重的变化则可能改变其对截断的依赖关系。

非线性顶点与集体有序。— 四阶连通 MSRJD 展开产生一个推迟顶点 $R_0^{(3)}(t;\,s_1,\,s_2,\,s_3)=\langle M(t)\,\widetilde{M}(s_1)\,\widetilde{M}(s_2)\,\widetilde{M}(s_3)\rangle_{0,\,c}$。在物理鞍点 $\phi=0$ 处，直至三次阶的集体方程为

$$0=J^{-1}\phi(t)-\int ds\,\Pi_0^R(t-s)\phi(s)-\frac{1}{3!N}\int d^3s\,R_0^{(3)}(t;\boldsymbol{s})\prod_{a=1}^{3}\phi(s_a)+O(\phi^5). \tag{12}$$

其他四阶顶点控制非线性噪声和涨落；式（12）是一个鞍点方程，而不是关于集体场的精确随机方程。在此鞍点处，$\phi=J\langle m\rangle$。

如果三次响应是正则的，且相对于缓慢的集体动力学而言寿命较短，则其卷积可近似为 $u\phi^3(t)$，其中 $u=-(6N)^{-1}\int d^3s\,R_0^{(3)}(t;\,\boldsymbol{s})$。将线性响应核的静态贡献分离出来后，得到

$$0=r_L\phi(t)+\int_0^\infty d\tau\,K(\tau)[\phi(t)-\phi(t-\tau)]+u\phi^3(t), \tag{13}$$

其中 $r_L=J^{-1}-\Pi_0^R(0)$ 和 $K(t)=\Pi_0^R(t)$。

局部一阶导数近似还要求有限时间矩 $Z_1=\int_0^\infty d\tau\,\tau K(\tau)=\int d\lambda\,\rho_O(\lambda)/\lambda^2$。对于 $\rho_O(\lambda)\propto\lambda^\beta$，该时间矩在 $\beta\le 1$ 时发散为 $\lambda_{\rm IR}\to 0$ ，因此无法进行一致局部导数展开。如果三次响应本身具有长寿命，则必须保留完整的非局域方程。

对于互易平衡参考网络，静态HS自由能密度可直接由微观配分函数得到，

$$f_{\rm HS}(\phi)=\frac{\phi^2}{2J}-\frac{D}{N}\ln z_0(\phi), \tag{14}$$

其中 $z_0(\phi)=\int d^Nx\exp[-(H_0(\boldsymbol{x})-\phi M)/D]$。假设具有 $M\mapsto -M$ 对称性，其连通累积量展开给出

$$f_{\rm HS}(\phi)=f_0+\frac{r_L}{2}\phi^2+\frac{u}{4}\phi^4+O(\phi^6), \tag{15}$$

其中 $r_L=J^{-1}-\kappa_2(M)/(ND)$，且 $u=-\kappa_4(M)/(6ND^3)$，其中 $\kappa_n(M)=\langle M^n\rangle_{0,c}$。平衡态静态涨落耗散关系给出$\Pi_0^R(0)=\kappa_2(M)/(ND)$。对于正则平衡静态响应，积分三次动力学响应等于 $\kappa_4(M)/D^3$，从而确保静态和动态四次系数的一致性。相应的响应恒等式以及局部非线性约化的条件详见补充材料的 S8 节。因此，非线性集体顶点由微观连通关联决定，而非通过现象学方式引入。

非线性稳定化并非自动发生。连续对称相变要求 $u>0$，等价地，

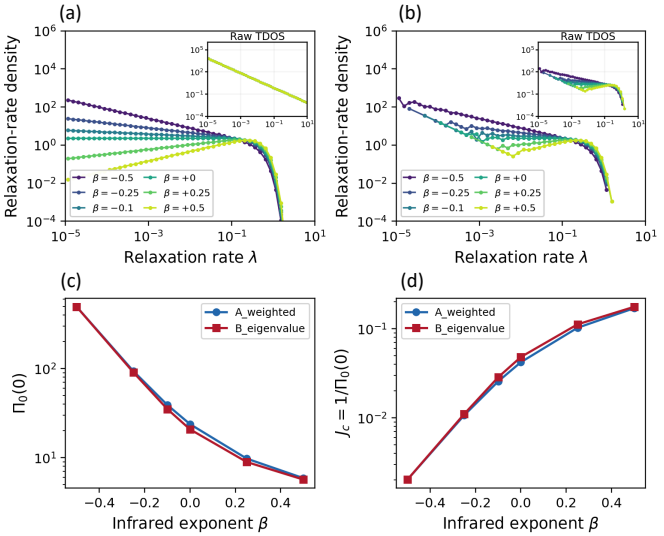


FIG.微2观弛豫矩阵中的谱组织与集体响应。比较了具有 $N = 2048$ 的两种对称非对角算子构造。（a，b）六种红外指数 $\beta = -0.5$、-0.25、-0.1、0、0。25、0。5 对应的响应加权弛豫谱 $\rho_O(\lambda)$。构造A在特征值固定时改变集体模重叠，而构造B在均匀重叠下改变特征值密度。插图显示相应的未加权弛豫速率密度。（c）静态极化 $\Pi_0^R(0)$。（d）高斯有序阈值 $J_c = 1/\Pi_0^R(0)$。

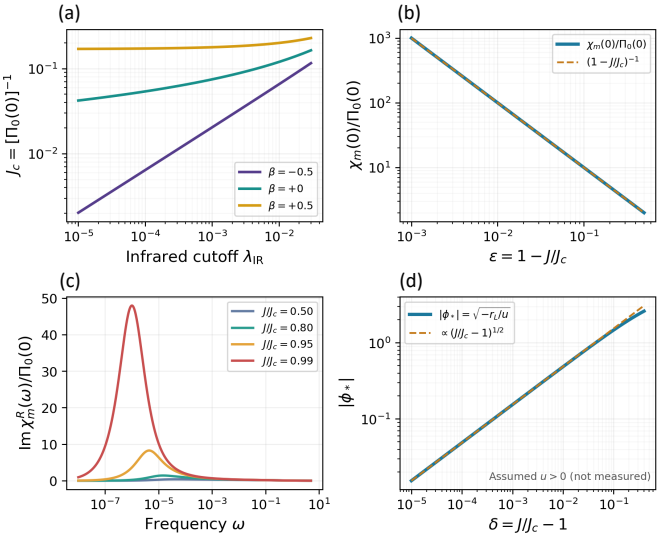


图 3. 集体临界性的红外谱调控。(a) 对于具有代表性的红外指数，高斯有序阈值 $J_c = [\Pi_0^R(0)]^{-1}$ 随红外截止 λ_{IR} 的变化。(b) 归一化静态集体磁化率 $\chi_m(0)/\Pi_0^R(0)$ 在 $J = J_c$ 处趋近高斯失稳。(c) $\beta = 0$ 时集体响应的低频虚部，展示了随着 J/J_c 趋近于 1，动力学软化的情况。(d) 朗道序参量标度示意，$|\phi_*| = \sqrt{-r_L/u} \propto (J/J_c - 1)^{1/2}$，假设 $u > 0$；四次系数并未通过微观四阶累积量单独评估。

$\kappa_4(M) < 0$，并且高阶项是正则的。对于固定的参考网络和 $J > J_c$，局部稳定分支满足

$$\phi_*^2 = \frac{J - J_c}{uJJ_c} + O((J - J_c)^2), \tag{16}$$

给出平均场序参量指数 $\beta_{\mathrm{L}} = 1/2$ 和静态磁化率 $\chi_\phi^{\mathrm{stat}} \propto |J - J_c|^{-1}$，其中$\gamma_{\mathrm{L}} = 1$。朗道指数 β_{L} 不同于红外谱指数 β。有序态曲率为 $f''_{\mathrm{HS}}(\phi_*) = -2r_L + O(r_L^2) > 0$，但这一静态回复曲率本身并不能决定动力学遗忘速率，后者取决于有序态的迟滞响应。

数值测试与启示.—为检验集体临界性的谱表述，我们构造了维度为 $N = 2048$ 的对称非对角弛豫算子：$L = Q\mathrm{diag}(\lambda_\alpha)Q^{\mathsf{T}}$，其中 Q 由高斯随机矩阵的QR分解得到。对于集体通道 $\boldsymbol{c}$，模态权重 $a_\alpha = |\boldsymbol{c}^{\mathsf{T}}\boldsymbol{v}_\alpha|^2/N$ 定义了响应加权谱 $\rho_O(\lambda) = \sum_\alpha a_\alpha \delta(\lambda - \lambda_\alpha)$。我们指定归一化目标测度 $\rho_O(\lambda) \propto \lambda^\beta e^{-(\lambda/\lambda_{\mathrm{UV}})^2}$，，其中$\lambda_{\mathrm{UV}} = 0.5$、$\lambda_{\mathrm{IR}} = 10^{-5}$、$\lambda_{\max} = 10$，以及六个指数$\beta = -0.5, -0.25, -0.1,$ 0,0.25, 0.5。谱构造、本征模诊断、集体极化和时间记忆核的详情见《补充材料》第 S1–S6 节。

图 2 对比了两种互补的构造。在构造A中，固定按对数间隔排列的弛豫率，同时改变集体模重叠。因此，未加权密度与 β 无关，而加权密度在红外区遵循 $\rho_O(\lambda) \propto \lambda^\beta$ [图 2(a)]。在构造B中，弛豫率取自目标分布的分位数，并采用均匀权重$a_\alpha = 1/N$，因此两种密度都呈现出预设的红外标度 [图 2(b)]。这两种构造通过不同的有限尺寸离散化逼近相同的加权连续测度，其中构造B在红外区的偏差更大。

静态极化和高斯阈值由微观本征模重构为 $\Pi_0^R(0) = \sum_\alpha a_\alpha/\lambda_\alpha$ 和 $J_c = [\Pi_0^R(0)]^{-1}$。图 2(c,d) 显示，慢模权重的增加会增强极化并降低集体阈值。对于 $\beta = -0.5$，构造 A 和 B 分别得到 $\Pi_0^R(0) \simeq 493.12$ 和 492.72，对应于 $J_c \simeq 2.03 \times 10^{-3}$。在$\beta = 0$处，相应的极化降至 23.77 和 20.78；在 $\beta = 0.5$ 处则降至 5.89 和 5.68。除有限尺寸离散化效应外，这些结果表明，集体响应由加权谱决定，而非仅由未加权的弛豫率密度决定。

图3(a)显示了J_c 在 $\beta = -0.5$，0，和 0。5 时对红外截止的依赖，分别呈现出幂律增

强的极化、对数敏感性和红外收敛。与固定振幅渐近行为不同，这些计算保留了预设测度的紫外正则化和依赖于截止的归一化。

根据式（6）计算集体易感性。定义 $\epsilon = 1 - J/J_c$，归一化静态易感性满足 $\chi_m^R(0)/\Pi_0^R(0) = \epsilon^{-1}$ [图 3(b)]，，对应高斯指数 $\gamma_{\mathrm{L}} = 1$，而非独立测得的涨落修正指数。

图 3(c) 展示了 $\beta = 0$ 时的动态响应 $\mathrm{Im}\,\chi_m^R(\omega)/\Pi_0^R(0)$。随着 J/J_c 从 0.50 增至 0.99，峰值频率从约 4.3×10^{-5} 降至 1.0×10^{-6}，而归一化振幅从 0.39 增至 48，这与集体临界慢化一致。峰值频率表征响应尺度，但未必对应一个孤立的弛豫极点。平衡态涨落分析还进一步区分了由第一逆速率谱矩决定的集体有序阈值，与同样取决于第二逆速率矩的时间持续性（见补充材料第 S7 节）。

最后，图 3(d) 展示了静态 HS 展开所预测的非线性有序化。对于假定为正的四次系数，有序解在阈值附近满足 $|\phi_*| = \sqrt{r_L/u} \propto (J/J_c - 1)^{1/2}$，说明了平均场指数 $\beta_{\mathrm{L}} = 1/2$，而无需独立计算决定 u 的微观四阶累积量。

图 2 和图 3 共同建立了微观弛豫谱、集体模重叠、红外增强极化以及静态和动态集体响应之间受控的联系。谱分布是预先设定的，而非通过学习生成；与此同时，有限红外截断和离散模采样限制了可观测的幂律时间窗口。

学习得到的神经架构可表示为隐藏状态之间的高维非线性映射，$\mathbf{h}_{\ell+1} = \mathcal{F}_\ell(\mathbf{h}_\ell)$；其完整雅可比矩阵能够捕捉局部线性化响应，而无需分解其内部计算操作。近期关于 Transformer 和 Mamba 的研究促使我们将谱与本文的微观框架进行比较 [23, 29]。我们的预测涉及特定集体可观测量的响应加权 低速率谱。Transformer 层间雅可比矩阵描述的是计算传播；要确定物理弛豫速率，则需要经过验证的有效演化定律。非正规算子或非互易算子也需要适当的左右模态重叠。仅凭近乎平坦的无权重 TDOS，既不能确立物理时间上的 $1/t$ 记忆核，也不能证明自发集体有序。直接的扰动响应测量、可控反馈变化，以及对学习得到的网络与未训练网络进行比较，都是

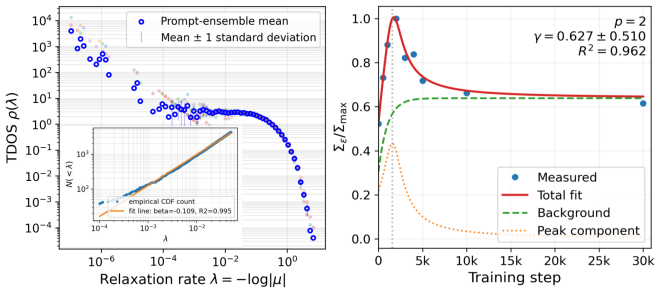


FIG. 3. Pythia-410M Transformer 中学习诱导的红外谱组织与瞬态类临界响应。(a) 对五个不同输入提示进行平均得到的总态密度。误差条表示提示集合中的一个标准差。插图显示了累积模计数及其幂律拟合，得到红外谱指数 $\beta = -0.109$。(b) 学习过程中第 $5 \to 6$ 层归一化红外记忆自能的演化。实线表示总的现象学拟合，虚线和点线分别表示背景贡献和瞬态贡献。竖直点线表示瞬态响应的中心 s_c。拟合得到的有效响应指数为 $\gamma = 0.627 \pm 0.510$。

需要检验这些预测。

学习诱导的红外有序与集体临界性。— 近期对 Transformer 雅可比矩阵谱的测量揭示了稳健的近边缘红外分布，$\rho(\lambda) \sim \lambda^\beta$ 其中 $\beta \simeq -0.1$，同时还发现红外记忆自能在学习早期出现显著的瞬态峰值 [23]。尽管训练期间谱发生了大幅重分布，红外标度在不同学习阶段和输入实现之间仍近似保持不变。

图 4(a) 展示了 Pythia-410M Transformer 在五个不同输入提示上的平均 TDOS。尽管存在依赖于提示的涨落，这些谱仍呈现出共同的红外结构。累积模态数遵循 $N(\lambda) \sim \lambda^{1+\beta}$，从而得到 $\beta = -0.109$，其中 $R^2 = 0.995$ [见图 4(a) 的插图]。这些观察表明，学习诱导的红外软化可能增强集体极化，并降低宏观有序化所需的临界耦合。

图 4(b) 展示了归一化红外记忆自能 $\Sigma(s) = \Sigma_\epsilon(0;s)/\Sigma_{\max}$ 的演化，其中 s 表示训练步。我们以现象学方式将观测到的自能分解为不断演化的谱背景和瞬态临界型贡献，

$$\overline{\Sigma}(s) = \overline{\Sigma}_f + (\overline{\Sigma}_i - \overline{\Sigma}_f)e^{-s/\tau} + \frac{B}{[1 + ((s - s_c)/w)^2]^\gamma}. \quad (17)$$

指数背景描述了既有弛豫谱的逐渐重组，而第二项则源于临界标度 $\chi_m(0) \sim \delta^{-\gamma}$，以及沿学习

轨迹到集体不稳定性的距离的局部二次展开。拟合得到有效响应指数$\gamma = 0.627 \pm 0.510$，其中 $R^2 = 0.962$；在拟合不确定性范围内，该结果与平均场值 $\gamma_{\mathrm{MF}} = 1$ 一致。理论依据、拟合程序和标度表示详见补充材料第 S9 节。

为估计这种红外增强对集体有序可能产生的影响，我们假设一种唯象关系 $\Pi_0(0; s) = \Pi_{\mathrm{bg}} + A\Sigma(s)$，其中$A > 0$。由此得到的临界耦合满足$J_c(s)/J_c(s_{\mathrm{peak}}) = (b+1)/[b+\Sigma(s)]$,，其中 $b = \Pi_{\mathrm{bg}}/A$且 $\Sigma(s_{\mathrm{peak}}) = 1$。在这一假设下，观测到的自能最大值对应于估算临界耦合的最小值，表明集体有序的阈值降低。

尽管雅可比松弛谱本身无法确定微观静态极化或集体易感性，但这些数值观察结果与学习诱导的红外谱组织之上叠加的瞬态类临界响应相一致。

讨论——我们从随机神经动力学出发，推导出一条微观路径，可得到有效集体场、其因果响应核以及有序阈值。其共同核心是参考网络的集体通道易感性：它的频率依赖性生成记忆核，而其零频值则控制静态集体不稳定性。对于可观测量加权红外谱 $\rho_O(\lambda) \propto \lambda^{\beta}$，边缘情形 $\beta = 0$ 会产生一个 $1/t$ 记忆窗口，并使静态集体响应呈对数增强。平衡涨落分析进一步表明，集体有序与时间持续性对逆速率谱矩的依赖各不相同，从而将二者区分开来。

动力学HS表述将二次集体不稳定性与同一微观展开中的高阶非线性响应顶角联系起来。在常规平衡态静态极限下，三次动力学响应再现了四次朗道系数，从而确定连续集体有序的条件。这些结果是利用既有的场论方法得到的，而无需现象学地引入记忆核或非线性集体顶角。

数值构造验证了高斯不稳定性与集体响应的谱依赖性，而非线性平均场标度则源于微观朗道展开。仅有红外累积并不足以意味着自发有序：集体相互作用必须达到失稳阈值，且非线性响应必须提供适当的稳定化。

Transformer 测量为理解学习过程中红外谱组织的涌现提供了互补视角。稳健的近边缘标度关系以及红外记忆自能的瞬态峰值表明，学习可能会动态调节有利于集体有序形成的谱条件。尽管观测到的类临界响应与所提出的框架一致，但仍需直接测量集体通道响应，才能确立其潜在机制。同一机制是否支配生物神经系统中的集体临界性，仍是一个有待实证检验的问题。

总之，微观理论确定了一种共同的响应结构，它支配着长时间集体记忆以及趋向宏观有序的失稳。响应加权弛豫谱决定了时间记忆核、集体失稳阈值，以及在平衡态下集体涨落的时间持续性，而更高阶的连通响应则决定了支配局部稳定化的非线性顶点。由此得到的框架提供了定量预测，用以检验微观弛豫模态及其通过学习形成的组织如何促成生物与人工神经系统中的持久集体动力学。

致谢—本研究部分得到了由韩国政府（MSIT）资助的信息与通信技术规划评估院（IITP）项目（ IITP-RS-2025-02214780）的支持。

* bgchae@etri.re.kr

[1] W. Gerstner、W. M. Kistler、R. Naud 和 L. Paninski，《神经元动力学：从单个神经元到网络与认知模型》（剑桥大学出版社，剑桥，2014年）。[2] V. Mante、D. Sussillo、K. V. Shenoy 和 W. T. Newsome，“前额叶皮层中由循环动力学实现的情境依赖计算”，《自然》**503**，78–84（2013年）。[3] D. R. Chialvo，“涌现的复杂神经动力学”，《自然·物理》**6**，744–750（2010年）。[4] J. M. Beggs 和 D. Plenz，“新皮层回路中的神经元雪崩”，《神经科学杂志》**23**，11167–11177（2003年）。[5] H. R. Wilson 和 J. D. Cowan，“模型神经元局部群体中的兴奋性与抑制性交互作用”，《生物物理学杂志》**12**，1–24（1972年）。[6] S. Amari，“侧向抑制型神经场中的图样形成动力学”，《生物控制论》**27**，77–87（1977年）。[7] B. J. Cook、A. D. H. Peterson、W. Woldman 和 J. R. Terry，“神经场模型：数学综述与统一框架”，《数学神经科学应用》**2**，1–67（2022年）。[8] J. J. Hopfield，“具有涌现集体计算能力的神经网络与物理系统”，《美国国家科学院院刊》**79**，2554–2558（1982年）。[9] J. J. Hopfield，“具有分级响应的神经元拥有类似双态神经元的集体计算特性”，《美国国家科学院院刊》**81**，3088–3092

（1984年）。[10] D. J. Amit 和 H. Gutfreund，“在神经网络自旋玻璃模型中存储无限数量的模式”，《物理评论快报》55，1530–1533（1985年）。[11] M. A. Buice 和 J. D. Cowan，“神经网络涨落效应的场论方法”，《物理评论E》75，051919（2007年）。[12] A. C. C. Coolen 和 D. Sherrington，“接近饱和时全连接吸引子神经网络的动力学”，《物理评论快报》71，3886–3889（1993年）。[13] A. Crisanti 和 H. Sompolinsky，“随机神经网络的路径积分方法”，《物理评论E》98，062120（2018年）。[14] M. Dick、A. van Meegen 和 M. Helias，“通过重整化场论关联网络层面与神经元层面的相关性”，《物理评论研究》6，033264（2024年）。[15] K. Mimura，“重新审视连续Hopfield模型的并行动力学”，J. Phys. Soc. Jpn. 78，033001（2009年）。[16] Y. Bengio、P. Simard 和 P. Frasconi，“使用梯度下降学习长期依赖关系十分困难”，IEEE Trans. Neural Netw. 5，157–166（1994年）。[17] R. Pascanu、T. Mikolov 和 Y. Bengio，“关于训练循环神经网络的困难”，载于第30届国际机器学习会议论文集，《机器学习研究论文集》28，1310–1318（2013年）。[18] G. Wainrib 和 M. N. Galtier，“通过最大李雅普诺夫指数得到局部回声状态性质”，Neural Networks 76，39–45（2016年）。[19] T. Can、K. Krishnamurthy 和 D. J. Schwab，“门控在门控循环单元和长短期记忆网络中产生慢模并控制相空间复杂度”，载于首届数学与科学机器学习会议论文集，《机器学习研究论文集》107，476–511（2020年）。[20] M. Chen、J. Pennington 和 S. S. Schoenholz，“动力学等距性与循环神经网络的平均场理论：门控使循环神经网络中的信号传播成为可能”，载于第35届国际机器学习会议论文集，《机器学习研究论文集》80，873–882（2018年）。[21] X. Chen 和 W. Bialek，“无需精细调节即可寻找长时间尺度”，《物理评论E》110，034407（2024年）。[22] A. Gu、K. Goel 和 C. Ré，“利用结构化状态空间高效建模长序列”，载于国际学习表征会议（2022年）。[23] B. G. Chae，“Transformer动力学中的红外弛豫谱与时间组织”，arXiv预印本 arXiv:2607.10923v4（2026年）。[24] P. C. Martin、E. D. Siggia 和 H. A. Rose，“经典系统的统计动力学”，Phys. Rev. A 8，423–437（1973年）。[25] H. K. Janssen，“关于经典场动力学的拉格朗日量及动力学临界性质的重整化群计算”，Z. Phys. B 23，377–380（1976年）。[26] C. De Dominicis，“场论重整化技术与临界现象动力学”，J. Phys. Colloq. 37，247–253（1976年）。[27] J. Hubbard，“配分函数的计算”，《物理评论快报》3，77–78（1959年）。[28] R. L. Stratonovich，“一种计算量子分布函数的方法”，Sov. Phys. Dokl. 2，416（1957年）。[29] B. G. Chae，“Transformer与状态空间架构中集体动力学的红外普适性”，arXiv预印本 arXiv:2608.18592v1（2026年）。

补充材料

S1. 数值设置与谱目标

我们数值构造有限维对称弛豫算子，以考察微观弛豫特征值及其与预设集体通道的重叠如何决定集体谱测度及相关的静态极化。计算使用 $N = 2048$ 个自由度和归一化目标密度

$$p_\beta(\lambda) = \frac{\lambda^\beta \exp[-(\lambda/\lambda_{\mathrm{UV}})^2]}{\displaystyle\int_{\lambda_{\mathrm{IR}}}^{\lambda_{\max}} ds\, s^\beta \exp[-(s/\lambda_{\mathrm{UV}})^2]}, \qquad \lambda_{\mathrm{IR}} \le \lambda \le \lambda_{\max}, \tag{S1.1}$$

其中 $\lambda_{\mathrm{IR}} = 10^{-5}$、 $\lambda_{\max} = 10$、 $\lambda_{\mathrm{UV}} = 0.5$ 和 $\beta \in \{-0.5, -0.25, -0.1, 0, 0.25, 0.5\}$。累积分布$F_\beta(\lambda) = \int_{\lambda_{\mathrm{IR}}}^{\lambda} p_\beta(s)\, ds$ 在包含 30,001 个点的对数网格上，使用梯形积分进行计算。所有矩阵计算均采用双精度。对伪随机高斯 $N \times N$ 矩阵进行 QR分解正交化，以获得实正交矩阵 Q，伪随机种子为 20261110。在 A/B比较中始终使用相同的 Q 。

对于每种构造，我们定义

$$L = Q\,\mathrm{diag}(\lambda_1, \ldots, \lambda_N) Q^{\mathsf{T}}, \qquad \boldsymbol{c} = \sqrt{N}\, Q \left(\sqrt{w_1} \;\; \cdots \;\; \sqrt{w_N}\right)^{\mathsf{T}}, \tag{S1.2}$$

其中 $w_\alpha \ge 0$ 和 $\sum_\alpha w_\alpha = 1$。集体响应及其谱权重定义为

$$\Pi_0^R(\omega) = \frac{1}{N} \boldsymbol{c}^{\mathsf{T}} (L - i\omega I)^{-1} \boldsymbol{c} = \sum_{\alpha=1}^{N} \frac{a_\alpha}{\lambda_\alpha - i\omega}, \qquad a_\alpha = \frac{1}{N} \left| \boldsymbol{c}^{\mathsf{T}} \boldsymbol{v}_\alpha \right|^2, \tag{S1.3}$$

其中 $L\boldsymbol{v}_\alpha = \lambda_\alpha \boldsymbol{v}_\alpha$。无权谱测度和集体加权谱测度分别为 ,

$$\rho_{\mathrm{raw}}(\lambda) = \frac{1}{N} \sum_\alpha \delta(\lambda - \lambda_\alpha), \qquad \rho_{\mathrm{coll}}(\lambda) = \sum_\alpha a_\alpha \delta(\lambda - \lambda_\alpha). \tag{S1.4}$$

这两种测度通常并不相同。式（S1.2）中的构造旨在使指定的 w_α能够精确地实现为特征向量重叠权重，误差仅为浮点误差；它并非对自发谱组织的模拟。

S2. 两种互补的微观构造

S2.1. A：固定对数特征值网格与可变集体权重

在 A_加权构造中， $N+1$ 分箱边界在 λ_{IR} 和 $\lambda_{\max}$之间按对数间隔排列。特征值位于这些分箱的几何中心，

$$\lambda_\alpha^{(A)} = \sqrt{e_\alpha e_{\alpha+1}}, \qquad w_\alpha^{(A)} = F_\beta(e_{\alpha+1}) - F_\beta(e_\alpha), \tag{S2.1}$$

其中 e_α 表示分箱边界。因此，随着 β 变化，特征值和矩阵 $L^{(A)}$ 保持不变，而集体耦合向量 $\boldsymbol{c}^{(A)}$ 则随 β变化。在红外区间， $w_\alpha^{(A)} \simeq p_\beta(\lambda_\alpha)\Delta\lambda_\alpha \propto \lambda_\alpha^{\beta+1}$，因为$\Delta\lambda_\alpha \propto \lambda_\alpha$ 位于对数网格上。因此，单个权重的幂律指数不必与单位 λ密度的幂律指数相同。

S2.2. B：目标分布特征值与均匀集体权重

在 B_特征值构造中，特征值被放置在目标分布的中点分位数处，

$$\lambda_\alpha^{(B)} = F_\beta^{-1}\left(\frac{\alpha - \frac{1}{2}}{N}\right), \qquad w_\alpha^{(B)} = \frac{1}{N}, \qquad \alpha = 1, \ldots, N. \tag{S2.2}$$

因此，矩阵 $L^{(B)}$ 会随 β变化，而集体耦合向量保持固定，因为 Q 和权重均未改变。对于 $\beta = 0$ ，且不存在紫外包络时，分位数网格在 λ中呈线性间隔；一般而言，它既不是线性间隔，也不是对数间隔。该构造具有 $\rho^{(B)}$ 集体$= \rho^{(B)}_{\mathrm{raw}}$ 作为离散测度。

A 和 B 都近似于同一个目标集体谱测度，但采用了不同的有限-N 离散化。特别是，两种构造在给定红外区间内的特征值数量并不相同；因此，必须将积分观测量的一致性与分箱谱密度的逐点一致性区分开来。

S3. 直接对角化与重叠验证

每个 L 均经过显式组装，并使用对称特征值求解器 `scipy.linalg.eigh`独立对角化。令 $\lambda_\alpha^{\mathrm{diag}}$和 $\boldsymbol{v}_\alpha^{\mathrm{diag}}$分别表示按特征值排序后得到的特征对。恢复的集体权重根据特征向量计算，而非从其输入中直接复制：

$$a_\alpha^{\mathrm{diag}} = \frac{1}{N}\left|\boldsymbol{c}^{\mathsf{T}}\boldsymbol{v}_\alpha^{\mathrm{diag}}\right|^2. \tag{S3.1}$$

特征值误差的最大绝对值、权重误差的绝对值和相对误差分别为

$$\epsilon_\lambda = \max_\alpha |\lambda_\alpha^{\mathrm{diag}} - \lambda_\alpha^{\mathrm{input}}|, \tag{S3.2}$$

$$\epsilon_a = \max_\alpha |a_\alpha^{\mathrm{diag}} - w_\alpha|, \tag{S3.3}$$

$$\epsilon_a^{\mathrm{rel}} = \max_\alpha \frac{|a_\alpha^{\mathrm{diag}} - w_\alpha|}{\max(w_\alpha, 10^{-15})}. \tag{S3.4}$$

$\epsilon_a^{\mathrm{rel}}$中的分母下限遵循实现方式，用于防止对极小的紫外尾部权重进行除法；因此，所报告的量是设有下限的相对误差诊断，并非权重低于 10^{-15} 时的常规相对误差。

表 S1. 最大特征值 e 以及两种构造的集体权重恢复误差。数值通过直接 $N = 2048$ 对角化得到。

β	构造	ϵ_λ	ϵ_a	$\epsilon_a^{\mathrm{rel}}$
−0.50	A_加权	1.24×10^{-14}	5.92×10^{-14}	3.35×10^{-9}
	B_特征值	1.33×10^{-15}	7.28×10^{-15}	1.49×10^{-11}
−0.25	A_加权	1.24×10^{-14}	6.28×10^{-15}	5.15×10^{-9}
	B_特征值	1.33×10^{-15}	1.02×10^{-15}	2.09×10^{-12}
−0.10	A_加权	1.24×10^{-14}	1.54×10^{-15}	4.16×10^{-9}
	B_特征值	1.33×10^{-15}	4.71×10^{-16}	9.65×10^{-13}
0.00	A_加权	1.24×10^{-14}	1.16×10^{-15}	5.13×10^{-9}
	B_特征值	1.44×10^{-15}	1.58×10^{-15}	3.23×10^{-12}
0.25	A_加权	1.24×10^{-14}	1.10×10^{-15}	6.50×10^{-9}
	B_特征值	1.55×10^{-15}	1.86×10^{-15}	3.80×10^{-12}
0.50	A_加权	1.24×10^{-14}	1.31×10^{-15}	7.08×10^{-9}
	B_特征值	1.55×10^{-15}	7.43×10^{-16}	1.52×10^{-12}

图 S1 展示了针对 $\beta = -0.5$ 的特征向量重叠检验。在 A 中，在已解析范围内，指定的模权重通过独立对角化得到的特征向量得以复现。在较大 λ处，权重的陡降反映了施加的紫外包络；该尾部中极小的数值容易受到浮点舍入误差的影响。在 B 中，恢复的权重仍接近 $1/N$，相对偏差接近特征值求解器的数值精度。

S4. 静态极化与有限尺寸考量

对于每个对角化后的矩阵，零频集体极化及相应的二次失稳阈值计算如下：

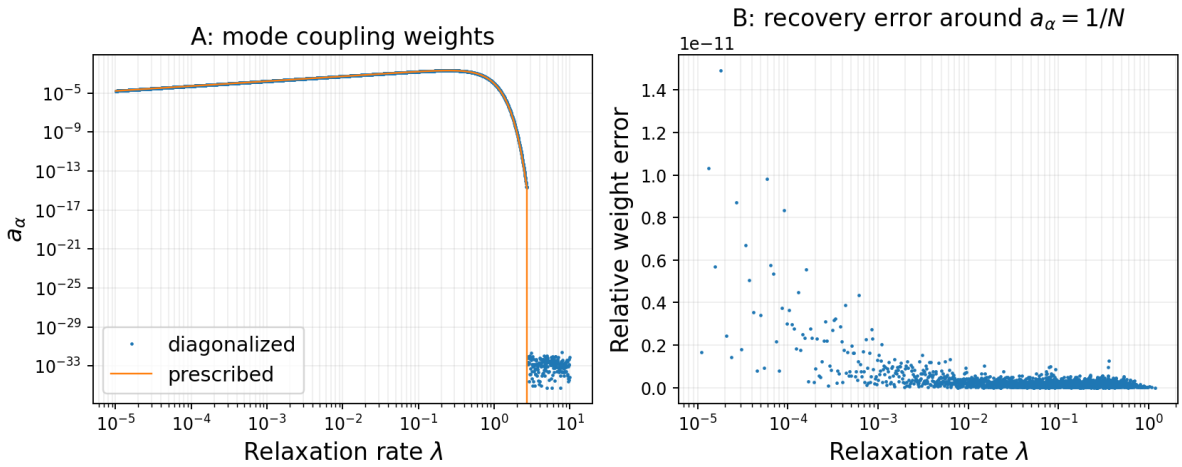


图 S1。**在 $\beta = -0.5$ 处对集体模重叠权重进行数值验证**。左：预设权重 w_α（线）以及通过直接对角化恢复的权重 $a_\alpha^{\mathrm{diag}} = |\boldsymbol{c}^{\mathsf{T}}\boldsymbol{v}_\alpha^{\mathrm{diag}}|^2/N$（点），对应 A_加权。右：恢复权重相对于 $w_\alpha = 1/N$ 的相对偏差，对应 B_ 特征值。这种一致性验证了构造矩阵、集体耦合向量和特征值求解器的一致性；但并不能证明目标分布是动态形成的。

$$\Pi_0(0) = \sum_\alpha \frac{a_\alpha^{\mathrm{diag}}}{\lambda_\alpha^{\mathrm{diag}}}, \qquad J_c = \frac{1}{\Pi_0(0)}. \tag{S4.1}$$

这些是正文中进行比较的量。A 和 B 结果的相似性反映了它们基于相同的归一化目标集体测度构造，但在有限 N 下，两者并不完全一致。具体而言，A 中的对数间距用许多低权重模态解析红外区域，而 B 中的分位数间距在同一区间内可能只保留较少的特征值。例如，在 $\beta = 0$ 和 $\lambda < 10^{-3}$ 处，A 包含 683 个模态，而 B 包含 5 个模态。它们的静态极化分别为 $\Pi_0^{(A)}(0) = 23.766879$ 和 $\Pi_0^{(B)}(0) = 20.782395$。因此，表 S1 中较小的对角化误差不应与连续谱积分的离散化误差混淆；后者可能被式 (S4.1) 中的因子 $1/\lambda$ 放大。

这些检验验证了预设有限维算符及重叠权重的代数构造和数值恢复。它们既不构成系统的 $N \to \infty$ 收敛性研究，也不证明目标红外分布会自发形成。在上述限制范围内，这些检验为正文中使用的集体谱表示提供了独立的矩阵层面一致性检验。

S5. 特征向量重叠与红外对集体极化的贡献

前文各节介绍了微观弛豫矩阵的数值构造，并验证了其特征值和集体模权重的准确性。现在，我们将考察特征向量重叠如何决定集体响应的谱分布，以及不同的红外指数如何将静态极化重新分配到各弛豫模中。

我们采用方程 (S1.3) 和 (S1.4) 中定义的集体模重叠权重和集体加权谱测度。

S5.1. 模态重叠与特征向量参与度

图 S2 展示了 $N = 2048$ 和 $\beta = 0$的特征向量及集体通道诊断。A 加权构造采用按对数间隔排列的弛豫速率，并使用非均匀集体模权重；而 B 特征值构造则根据规定的目标谱分配弛豫率，并赋予均匀权重 $a_\alpha = 1/N$。

尽管这两种构造中的单个模态权重差异很大，但它们的累积重叠分布却高度吻合。这种吻合是意料之中的，因为这两种构造都近似于同一个归一化集体谱测度。

利用式 (S4.1) 中的静态极化，我们通过定义来确定速率低于给定值λ 的弛豫模所作的贡献

$$F_{\Pi}(\lambda) = \frac{\Pi_0(<\lambda)}{\Pi_0(0)} = \frac{\sum_{\lambda_\alpha<\lambda} a_\alpha/\lambda_\alpha}{\sum_{\alpha} a_\alpha/\lambda_\alpha}. \tag{S5.1}$$

图 S2 的中间面板比较了这种累积极化与模态分辨贡献 a_α/λ_α。两种构造之间的微小差异反映了它们各自不同的有限尺寸离散化；当谱测度以 $1/\lambda_\alpha$ 加权时，这种差异尤为显著。

此外，我们　　使用逆参与率表征特征向量的空间延展程度

$$\mathrm{IPR}_\alpha = \sum_{i=1}^{N} |v_{i\alpha}|^4 \tag{S5.2}$$

以及归一化参与率

$$p_\alpha = \frac{1}{N\,\mathrm{IPR}_\alpha}. \tag{S5.3}$$

对于扩展的实随机特征向量，预期的逆参与率约为 $3/(N+2)$，对应于在 N 较大时接近 $1/3$ 的归一化参与率。

数值结果与这些预期一致，表明特征向量是空间扩展的随机特征向量，而非强烈局域化的模态。由于两种构造使用相同的正交矩阵，特征向量参与统计量并非空间组织的独立实现。相反，这种比较分离出了特征值配置和集体通道重叠的影响。

S5.2. 静态极化的红外再分布

接下来，我们考察红外指数 β 如何控制集体极化的谱分布。规定的集体加权谱具有如下形式

$$\rho_{\mathrm{coll}}(\lambda) \propto \lambda^{\beta} \exp\left[-\left(\frac{\lambda}{\lambda_{\mathrm{UV}}}\right)^2\right], \tag{S5.4}$$

在有限区间 $\lambda_{\mathrm{IR}} \le \lambda \le \lambda_{\max}$ 内。
在红外区间内，静态极化的贡献正比于

$$\Pi_0(0) \sim \int_{\lambda_{\mathrm{IR}}} \lambda^{\beta-1}\, d\lambda. \tag{S5.5}$$

因此，当 $\beta = 0$ 时，极化呈现对数红外敏感性；当 $\beta < 0$ 时，则呈现更强的幂律敏感性，而正的 β 会抑制最慢弛豫模态的相对贡献。

图 S3 比较了六种红外指数下的 $F_\Pi(\lambda)$，分别为 $\beta = -0.5,\ -0.25,\ -0.1,\ 0,\ 0.25,$ 和 0.5。在两种微观构造中，随着 β 减小，累积极化系统性地向较小的弛豫速率偏移。当 β 为负时，总静态响应的相当大一部分会在红外区内累积。

A 加权结果与 B特征值结果相对于 β 呈现相同的总体排序。当 B 特征值构造中的红外区采样稀疏时，定量差异最为明显。这些差异源于有限尺寸谱分辨率，而非预设的连续极限不同。

结果表明，集体模谱权重的红外分布直接控制静态集体极化，进而决定式 (S4.1) 中的临界耦合。这一数值比较以受控的方式从微观层面展示了红外弛豫谱与集体不稳定性之间的联系。本构造中预先设定了红外分布；并未假定它们会在微观学习或相互作用动力学过程中自发形成。

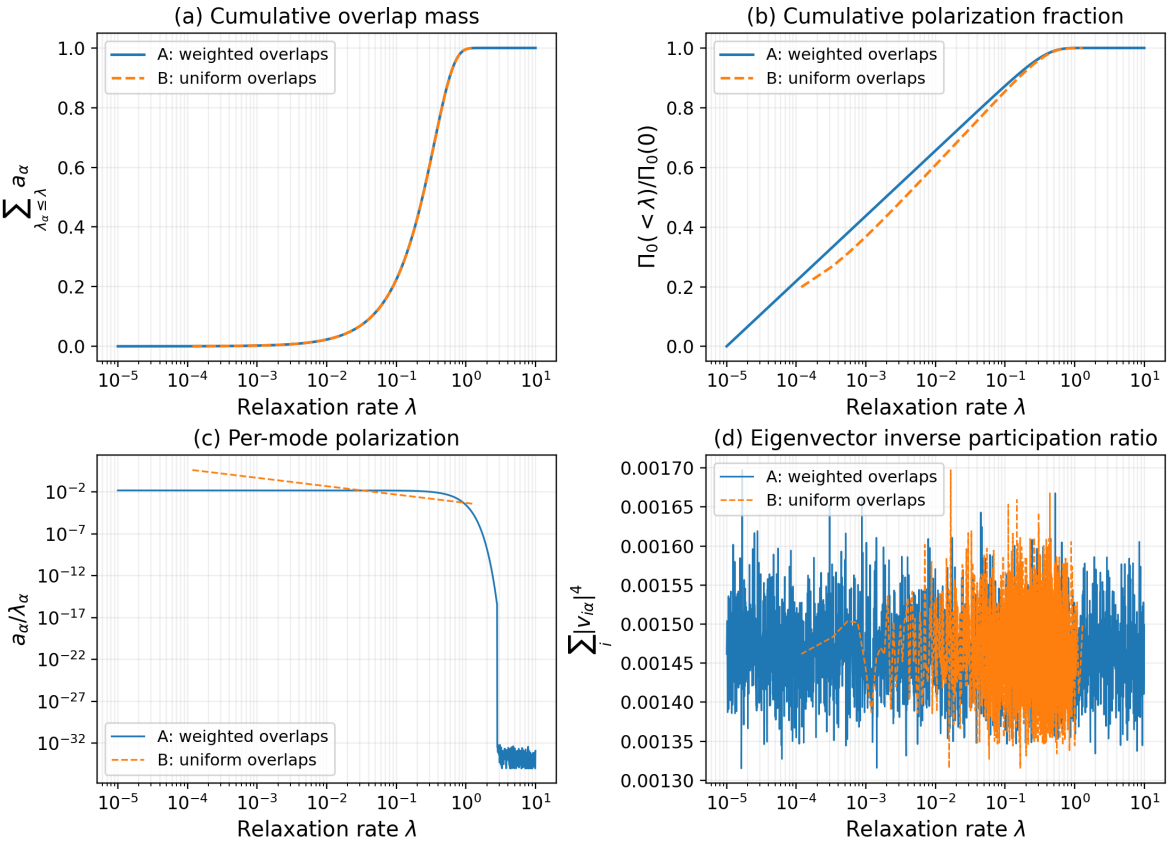


图 S2. $N = 2048$ 和 $\beta = 0$的特征向量及集体通道诊断。(a) 累积集体重叠质量$\sum_{\lambda_\alpha \le \lambda} a_{\alpha\circ}$ (b) 归一化累积静态极化 $F_\Pi(\lambda) = \Pi_0(< \lambda)/\Pi_0(0)$。(c) 单个模态对静态极化的贡献， $a_\alpha/\lambda_{\alpha\circ}$ (d) 特征向量逆参与率，$\mathrm{IPR}_\alpha = \sum_i |v_{i\alpha}|^4$。A 加权构造和 B 特征值构造通过不同的离散化近似相同的集体谱测度。两种构造的参与统计量均与扩展随机特征向量相符。

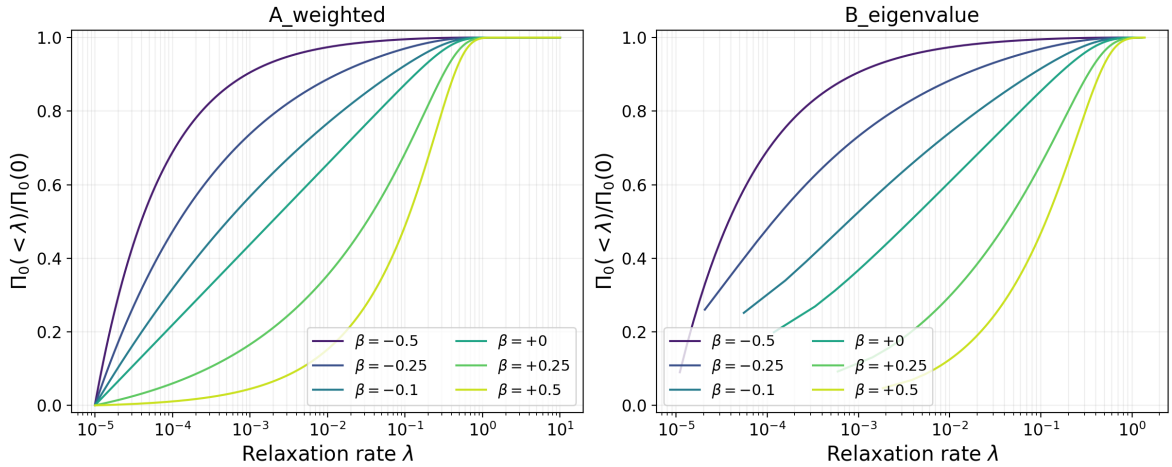


图 S3. 六种红外指数 β 在 $N = 2048$处的归一化累积静态极化 $F_\Pi(\lambda) = \Pi_0(< \lambda)/\Pi_0(0)$。左：A 加权构造，采用对数间隔的弛豫率和非均匀集体重叠权重。右：B 特征值构造，采用依赖谱的弛豫率和均匀集体重叠权重。随着 β 减小，静态集体极化中源自慢红外模态的比例不断增加。两种构造之间的差异反映了有限尺寸离散化效应。

S6. 时间记忆核与有限尺寸红外分辨率

对称过阻尼弛豫算符 $L = Q\mathrm{diag}(\lambda_\alpha)Q^{\mathsf{T}}$ 的延迟极化具有如下谱表示

$$\Pi_0^R(\omega) = \sum_{\alpha=1}^{N} \frac{a_\alpha}{\lambda_\alpha - i\omega}, \qquad a_\alpha = \frac{1}{N}\left|\mathbf{c}^{\mathsf{T}}\mathbf{v}_\alpha\right|^2. \tag{S6.1}$$

相应的因果响应为 $\Pi_0^R(t) = \Theta(t)K_O(t)$，其中

$$K_O(t) = \sum_{\alpha=1}^{N} a_\alpha e^{-\lambda_\alpha t} = \int d\lambda\, \rho_O(\lambda) e^{-\lambda t}, \tag{S6.2}$$

其中 $\rho_O(\lambda) = \sum_\alpha a_\alpha \delta(\lambda - \lambda_\alpha)$ 是集体加权弛豫谱。

对于红外密度 $\rho_O(\lambda) \simeq A_\beta \lambda^\beta$，当 $\beta > -1$时，中间时间核满足

$$K_O(t) \simeq A_\beta \int_{\lambda_{\mathrm{IR}}}^{\Lambda} d\lambda\, \lambda^\beta e^{-\lambda t} \simeq A_\beta \Gamma(1+\beta) t^{-(1+\beta)}, \tag{S6.3}$$

当 $\Lambda^{-1} \ll t \ll \lambda^{-1}$ 红外 时。因此，边缘情形 $\beta = 0$ 给出 $K_O(t) \propto t^{-1}$。

我们使用第 S1–S2 节中的 A 加权构造和 B 特征值构造，结合 $N = 2048$、$\lambda_{\mathrm{IR}} = 10^{-5}$、 $\lambda_{\max} = 10$ 和 $\lambda_{\mathrm{UV}} = 0.5$，对离散核进行评估。红外指数为 $\beta = -0.5,\ -0.25,\ -0.1,$ 0， 0。25， 0。5。图 S4(a,b) 显示，尽管两种构造的特征值分布和重叠分布不同，但它们均表现出相当的中间时间幂律行为。

有效衰减指数直接根据离散谱进行评估：

$$\alpha_{\mathrm{eff}}(t) \equiv -\frac{d\ln K_O(t)}{d\ln t} = t\frac{\sum_\alpha a_\alpha \lambda_\alpha e^{-\lambda_\alpha t}}{\sum_\alpha a_\alpha e^{-\lambda_\alpha t}}. \tag{S6.4}$$

如图 S4(c,d) 所示，两种构造都形成了与 $\alpha_{\mathrm{eff}} \simeq 1+\beta$ 一致的中间时间平台。对于 $\beta \le 0$，二者的一致性尤其接近。

对于 $\beta > 0$，构造B 表现出更早的有限尺寸交叉，因为其基于分位数的特征值采样在深红外区变得稀疏。相比之下，构造A 保留了对数间隔的红外特征值，同时降低了集体重叠权重。这一差异影响的是幂律区间的持续时间，而不是连续极限标度指数。

对于任何弛豫速率严格为正的有限谱，长时间响应由具有非零集体权重的最慢模态主导，

$$K_O(t) \sim a_{\alpha_*} e^{-\lambda_{\alpha_*} t}, \qquad \lambda_{\alpha_*} = \min_{\alpha: a_\alpha > 0} \lambda_\alpha, \tag{S6.5}$$

得到 $\alpha_{\mathrm{eff}}(t) \sim \lambda_{\alpha_*} t$。

时间核由集体加权弛豫谱决定，而非仅由未加权特征值密度决定。

S7. 红外　　集体时间相关性的谱依赖性　　S

正文确立了集体有序阈值由投影到集体相互作用通道上的弛豫谱决定。在此，我们考察相变无序侧相应的时间涨落。我们将分析限定在过阻尼动力学的平衡高斯极限，此时涨落耗散定理（FDT）将集体响应与其连通关联函数联系起来。下文的数值结果是对由此产生的谱依赖性的可控示例，而非对非线性有序的独立模拟。

对于集体　　变量 $m(t) = N^{-1}\mathbf{c}^{\mathsf{T}}\mathbf{x}(t)$，我们定义经 N 重标度的连通关联函数

$$C_m^{(J)}(t) = N\left[\langle m(t)m(0)\rangle - \langle m\rangle^2\right]. \tag{S7.1}$$

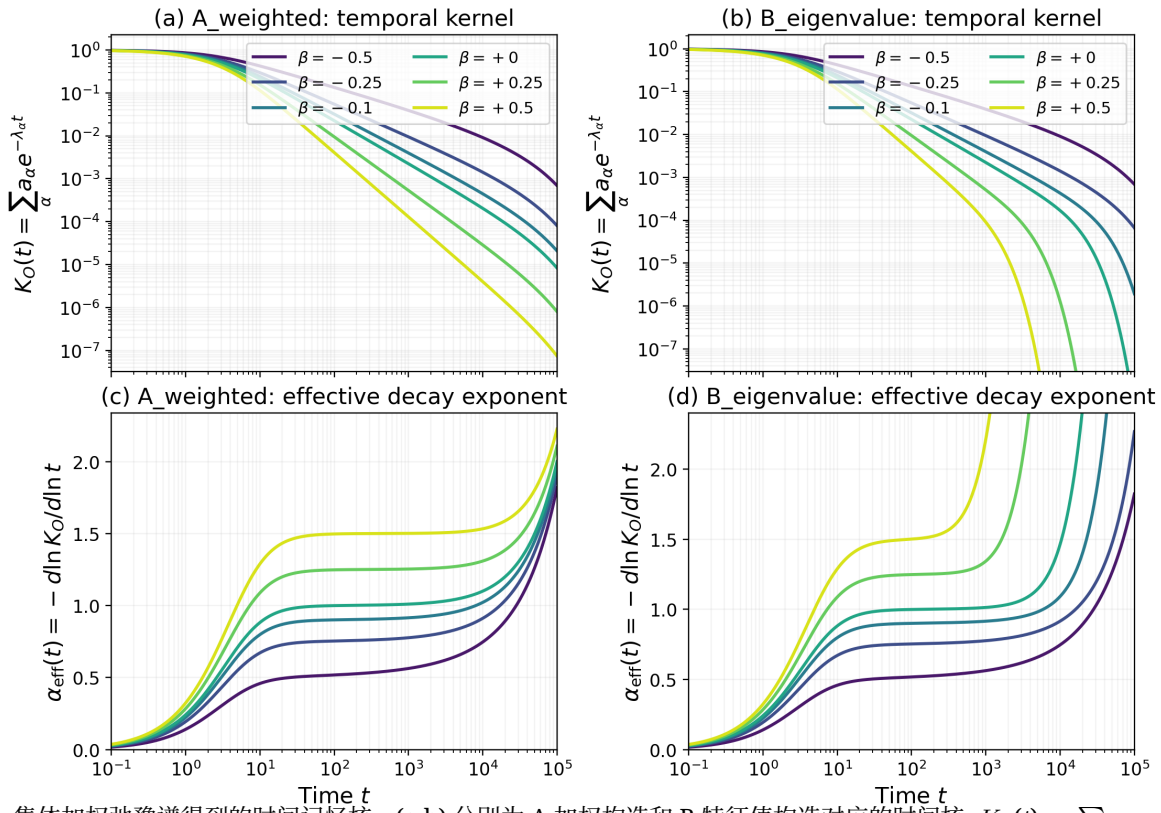


图 S4. 集体加权弛豫谱得到的时间记忆核。(a,b) 分别为 A 加权构造和 B 特征值构造对应的时间核 $K_O(t) = \sum_\alpha a_\alpha e^{-\lambda_\alpha t}$，其中 $N = 2048$ 和 $\beta = -0.5,\ -0.25,\ -0.1,\ 0,\ 0。25,\ 0。5$。(c,d) 对应的有效衰减指数 $\alpha_{\rm eff}(t) = -d\ln K_O/d\ln t$。两种构造均表现出中间时间范围内的幂律弛豫，其 $\alpha_{\rm eff} \simeq 1 + \beta$。对于正的 β，由于基于特征值的离散化对红外模的采样较为稀疏，构造B会更早出现有限尺寸交叉。

参考极化和相互作用修饰的易感性为

$$\Pi_0^R(\omega) = \sum_\alpha \frac{a_\alpha}{\lambda_\alpha - i\omega}, \qquad \chi_m^R(\omega) = \frac{\Pi_0^R(\omega)}{1 - J\Pi_0^R(\omega)}, \tag{S7.2}$$

其中 $a_\alpha = N^{-1}|\mathbf{c}^{\mathsf{T}}\mathbf{v}_\alpha|^2$ 和 $\lambda_\alpha > 0$。

在 e 平衡白噪声约定 $\langle \eta_i(t)\eta_j(t')\rangle = 2D\delta_{ij}\delta(t-t')$ 下，经典涨落耗散定理给出 s

$$S_m^{(J)}(\omega) = \frac{2D}{\omega}\,\mathrm{Im}\,\chi_m^R(\omega), \qquad C_m^{(J)}(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, e^{-i\omega t} S_m^{(J)}(\omega). \tag{S7.3}$$

公式（S7.1）中的归一化与对以 $-Nh(t)m(t)$ 形式耦合的场所产生的集体响应一致。公式（S7.3）假设采用平衡梯度朗之万模型。对于一般的非平衡神经动力学，涨落谱还取决于有效噪声核。

为区分静态谱敏感性与时间谱敏感性，引入集体通道加权态密度的逆速率矩，

$$\rho_O(\lambda) = \sum_\alpha a_\alpha \delta(\lambda - \lambda_\alpha), \qquad M_n = \int d\lambda\, \frac{\rho_O(\lambda)}{\lambda^n} = \sum_\alpha \frac{a_\alpha}{\lambda_\alpha^n}. \tag{S7.4}$$

有序阈值为 $J_c^{-1} = \Pi_0^R(0) = M_1$。定义 $\epsilon = 1 - JM_1 = 1 - J/J_c > 0$ 后，等时关联可由静态涨落耗散定理得出：

$$C_m^{(J)}(0) = D\chi_m^R(0) = \frac{DM_1}{\epsilon}. \tag{S7.5}$$

低频展开 $\Pi_0^R(\omega) = M_1 + i\omega M_2 + O(\omega^2)$ 意味着

$$\lim_{\omega\to 0} \frac{\mathrm{Im}\,\chi_m^R(\omega)}{\omega} = \frac{M_2}{\epsilon^2}. \tag{S7.6}$$

由于 $S_m^{(J)}(0) = 2\int_0^\infty dt\, C_m^{(J)}(t)$，方程 (S7.3) 和 (S7.6) 给出

$$\int_0^\infty dt\, C_m^{(J)}(t) = \frac{DM_2}{\epsilon^2}. \tag{S7.7}$$

因此，积分相关时间满足精确的高斯关系

$$\tau_{\mathrm{int}}(J) \equiv \int_0^\infty dt\, \frac{C_m^{(J)}(t)}{C_m^{(J)}(0)} = \frac{M_2}{M_1}\frac{1}{1 - J/J_c}. \tag{S7.8}$$

因此， M_1 决定集体有序阈值，而额外的逆速率矩 M_2 则控制时间持续性。特别是，对于不同的红外谱，相对于失稳的相同距离并不意味着相关时间相同。

我们通过具有 $\rho_O(\lambda) \propto \lambda^\beta$的有限维谱构造来说明这种区别，其中 $\beta = -1/2$、0、$1/2$，它们具有共同的速率区间 $0.01 \le \lambda \le 1$和相同的集体通道谱权重。对于每个谱， J_c 根据其离散的 M_1计算，相关性则在 $J/J_c = 0$、0。5、0。9、0。98 处评估。

图 S5 显示，在每个固定的相对耦合下，红外增强谱呈现出更慢的归一化时间衰减。进一步增大 J/J_c 会增强相关振幅，并产生高斯临界慢化。在有限 N下，由于离散谱具有正的最小弛豫速率，最终的长时间衰减呈指数形式。

这种有限尺寸行为与图 S4 所示的区别一致。尽管 A 加权构造和 B 特征值构造近似于同一个给定的连续谱测度，但它们的有限尺寸红外分辨率可能存在显著差异。具体而言，基于分位数的 B 构造所采样的最小弛豫速率可能远高于名义红外截止，而对数间隔的 A 构造则保留了接近该截止的模态。因此，即使它们的连续极限一致，其时间核在有限 N 下也可能呈现不同的交叉时间。

因此，观测到的 A–B 差异与有限尺寸红外离散化效应相符，而不必归因于不同的底层连续记忆定律。要定量确认收敛性，需要增大 N，并将两种构造与连续谱积分进行比较。

图 S6 总结了相应的等时涨落和积分关联时间。在高斯近似下，二者均按 $(1 - J/J_c)^{-1}$ 增长，但其谱前因子不同：$C_m^{(J)}(0)$ 与 M_1 成正比，而 τ_{int} 与 M_2/M_1 成正比。在 $J/J_c = 0.9$ 下，代表性构造分别给出 $\tau_{\mathrm{int}} \simeq 370$，215、100（对于 $\beta = -0.5$、0、0。5）。这些结果直观展示了同一投影弛豫谱所产生的时间后果，而该谱控制着正文中的集体不稳定性。高斯发散本身并不能确立非线性有序相或一般非平衡神经涨落的行为。

S8. 动力学与静态四次顶点的一致性

动力学 Hubbard–Stratonovich 有效作用量通过连通 MSRJD 响应函数生成非线性集体顶角。对于反射对称参考态，集体响应场的一次项与集体驱动的三次项构成的项为

$$W_0^{(4)} \supset \frac{1}{3!}\int dt\, d^3\boldsymbol{s}\, \tilde{\phi}(t) R_0^{(3)}(t;\boldsymbol{s}) \prod_{a=1}^{3} \phi(s_a), \tag{S8.1}$$

其中

$$R_0^{(3)}(t;\boldsymbol{s}) = \left\langle M(t)\widetilde{M}(s_1)\widetilde{M}(s_2)\widetilde{M}(s_3)\right\rangle_{0,c}. \tag{S8.2}$$

因子 $1/3!$源自第四阶连通累积量的组合学。

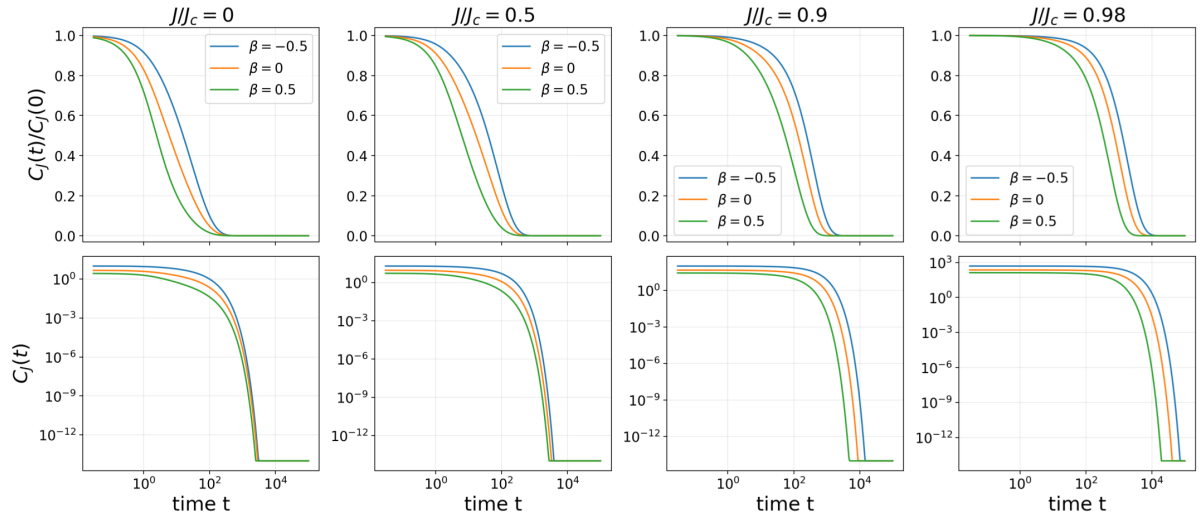


图 S5. 三种加权红外谱指数下的集体时间关联 $\beta = -0.5$，0，0.5，在固定相对耦合 $J/J_c = 0$，0.5，0.9，0.98 时的结果。上排面板显示 $C_m^{(J)}(t)/C_m^{(J)}(0)$，下排面板显示 $C_m^{(J)}(t)$。红外谱权重越大，关联越持久；而增大 J/J_c 会增强集体涨落并减缓其衰减。下排面板中接近数值绘图下限的数值不具有物理意义。

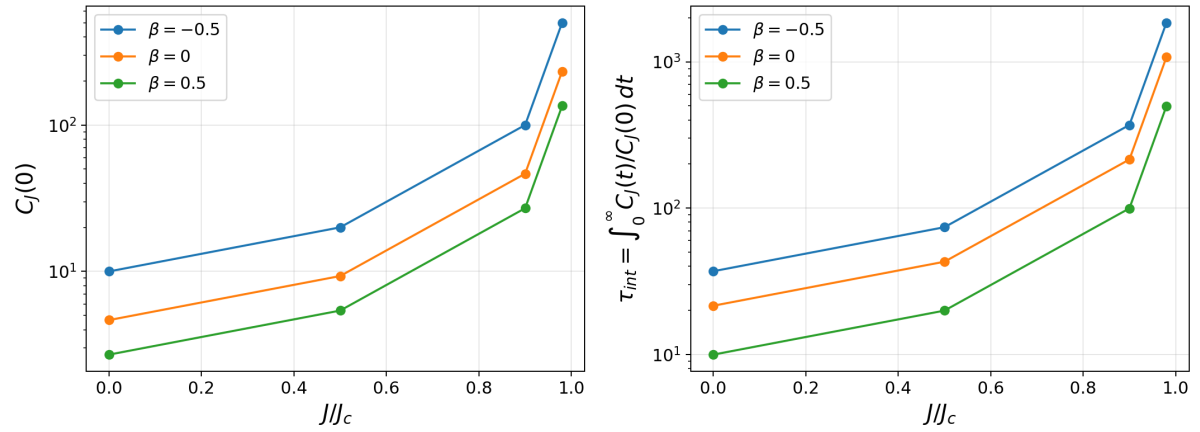


图 S6. 三种红外谱指数下，等时集体涨落 $C_m^{(J)}(0)$（左）和积分相关时间 τ_{int}（右）随 J/J_c 的变化。高斯关系 $C_m^{(J)}(0) = DM_1/(1 - J/J_c)$ 和 $\tau_{\mathrm{int}} = (M_2/M_1)/(1 - J/J_c)$ 解释了临界增强及不同的谱前因子。

~ ~在物理鞍点 $\phi = 0$处，对ϕ求变分会产生三次贡献–$(3!N)^{-1}\int d^3\boldsymbol{s}\, R_0^{(3)}(t;\, \boldsymbol{s})\, \prod_a \phi(s_a)$。

对于平稳平衡动力学，假设静态非线性响应是正则的，且相关时间积分收敛。于是，积分后的三阶响应满足

$$\begin{aligned}\mathcal{R}_3 &= \int d^3\boldsymbol{s}\, R_0^{(3)}(t;\boldsymbol{s}) \\ &= \left.\frac{\partial^3\langle M\rangle_\phi}{\partial\phi^3}\right|_{\phi=0} = \frac{\kappa_4(M)}{D^3}.\end{aligned} \tag{S8.3}$$

最后一个等式 源自平衡生成函数 $z_0(\phi) = \int d^N x \exp[-(H_0 - \phi M)/D$].

如果集体场相对于三阶响应的时间支撑变化缓慢，三次卷积就会变为局域形式，其表达式为

$$u_{\mathrm{dyn}} = -\frac{\mathcal{R}_3}{6N} = -\frac{\kappa_4(M)}{6ND^3} = u_{\mathrm{stat}}. \tag{S8.4}$$

因此，局域动力学三次系数与静态 HS 自由能的四次系数一致。

此等式涉及零频非线性响应，但并不意味着完整的三阶响应在时间上是局域的。如果非线性响应保留了长寿命记忆，就必须保留其非局域卷积。此外，包含两个或更多集体响应场的顶点会对非线性涨落产生贡献，且无法由确定性鞍点方程表示。

S9. 学习诱导的谱组织与瞬态临界响应

我们考察 Transformer 学习过程中观察到的红外记忆自能的瞬态增强。受正文中推导的微观集体响应启发，我们引入一种唯象描述，将背景弛豫谱的渐进组织与学习过程中的瞬态类临界增强区分开来。

S9.1. 学习前的弛豫谱

神经动力学的一个重要特征是，学习开始前就已存在弛豫谱。设相邻层之间内部神经状态的传播由下式描述

$$\boldsymbol{x}_{l+1} = F_l(\boldsymbol{x}_l; \boldsymbol{\theta}_s), \tag{S9.1}$$

其中 $\boldsymbol{\theta}_s$ 表示 训练步 s 时的网络参数。相应的局域传播雅可比矩阵 is

$$\mathcal{J}_l(s) = \frac{\partial F_l(\boldsymbol{x}_l; \boldsymbol{\theta}_s)}{\partial \boldsymbol{x}_l}. \tag{S9.2}$$

设$\mathcal{J}_l(s)$的特征值为 $\zeta_\alpha(s)$，有效弛豫率定义为 s

$$\lambda_\alpha(s) = -\frac{1}{\Delta l} \ln |\zeta_\alpha(s)|, \tag{S9.3}$$

其中 $\Delta l = 1$ 用于 相邻层传播。相应的时间尺度态密度（总态密度） is

$$\rho(\lambda; s) = \frac{1}{N} \sum_{\alpha=1}^{N} \delta\big(\lambda - \lambda_\alpha(s)\big), \tag{S9.4}$$

其中 N 是传播雅可比矩阵的维度。

由于随机初始化网络已经具有定义良好的传播雅可比矩阵，因此即使在 $s = 0$ 时，其弛豫谱也已存在。因此，学习并非从原本缺失的动力学结构中创造弛豫谱，而是通过网络参数的演化重组已有的谱。

特别是，谱权重向红外区的累积，体现了学习诱导的内部弛豫模重新分布。

S9.2. 背景与瞬态谱分量

为刻画学习诱导的谱重组，我们以现象学方式将总态密度分解为背景贡献和瞬态临界型贡献，

$$\rho(\lambda; s) = \rho_{\mathrm{bg}}(\lambda; s) + \Delta\rho_{\mathrm{crit}}(\lambda; s). \tag{S9.5}$$

背景分量描述了学习过程中弛豫谱的渐进组织，而 $\Delta\rho_{\mathrm{crit}}(\lambda; s)$ 则描述了与慢弛豫模增强相关的额外瞬态谱权重重新分布。

后者是一种谱修正，不必在每个弛豫速率处都为正。对于归一化的完整谱测度，谱权重的重新分配必须保持总归一化。

对于稳定弛豫扇区，正则化静态谱记忆自能定义为

$$\Sigma_\epsilon(0;s) = g^2 \int_0^\Lambda d\lambda \, \frac{\rho(\lambda;s)}{\lambda+\epsilon}, \tag{S9.6}$$

其中， g 表示有效模态无关耦合， $\epsilon > 0$ 是红外正则化参数， Λ 是谱上截止。

将式 (S9.5) 代入式 (S9.6)，可得

$$\begin{aligned}\Sigma_\epsilon(0;s) = g^2 \int_0^\Lambda d\lambda \, \frac{\rho_{\mathrm{bg}}(\lambda;s)}{\lambda+\epsilon} \\ + g^2 \int_0^\Lambda d\lambda \, \frac{\Delta\rho_{\mathrm{crit}}(\lambda;s)}{\lambda+\epsilon}.\end{aligned} \tag{S9.7}$$

因此，静态记忆自能可分解为

$$\Sigma_\epsilon(0;s) = \Sigma_{\mathrm{bg}}(s) + \Delta\Sigma_{\mathrm{crit}}(s), \tag{S9.8}$$

with

$$\Sigma_{\mathrm{bg}}(s) = g^2 \int_0^\Lambda d\lambda \, \frac{\rho_{\mathrm{bg}}(\lambda;s)}{\lambda+\epsilon}, \tag{S9.9}$$

$$\Delta\Sigma_{\mathrm{crit}}(s) = g^2 \int_0^\Lambda d\lambda \, \frac{\Delta\rho_{\mathrm{crit}}(\lambda;s)}{\lambda+\epsilon}. \tag{S9.10}$$

这种加性结构直接源于谱记忆自能对总态密度的线性依赖。然而，这种分解本身是唯象的，无法由测量得到的总谱唯一确定。

S9.3. 背景谱的有效学习动力学

背景谱的渐进组织过程可用训练步空间中的一阶弛豫过程近似。作为一种最简唯象近似，我们假设

$$\frac{\partial\rho_{\mathrm{bg}}(\lambda;s)}{\partial s} = -\frac{1}{\tau}\left[\rho_{\mathrm{bg}}(\lambda;s) - \rho_f(\lambda)\right], \tag{S9.11}$$

其中， $\rho_f(\lambda)$ 表示有效晚期训练背景谱， τ 是一个特征学习尺度。

这里假设背景谱演化具有单一的有效学习尺度。该近似并不要求各个微观弛豫速率完全相同。

引入谱偏差

$$u(\lambda;s) = \rho_{\mathrm{bg}}(\lambda;s) - \rho_f(\lambda), \tag{S9.12}$$

式 (S9.11) 变为

$$\frac{\partial u(\lambda;s)}{\partial s} = -\frac{1}{\tau}u(\lambda;s). \tag{S9.13}$$

解为

$$u(\lambda;s) = u(\lambda;0)e^{-s/\tau}. \tag{S9.14}$$

将初始背景谱定义为

$$\rho_i(\lambda) \equiv \rho_{\mathrm{bg}}(\lambda;0), \tag{S9.15}$$

我们得到

$$\rho_{\mathrm{bg}}(\lambda;s) = \rho_f(\lambda) + \left[\rho_i(\lambda) - \rho_f(\lambda)\right] e^{-s/\tau}. \tag{S9.16}$$

等价地，背景谱可以表示为

$$\rho_{\mathrm{bg}}(\lambda; s) = \rho_i(\lambda) e^{-s/\tau} + \rho_f(\lambda)(1 - e^{-s/\tau}). \tag{S9.17}$$

该表达式描述了背景谱从初始状态连续演化至训练后期的有效背景谱。

代入 式 (S9.16) 的背景自能谱表示后，得到 s

$$\begin{aligned}\Sigma_{\mathrm{bg}}(s) = g^2 \int_0^\Lambda d\lambda \, \frac{\rho_f(\lambda)}{\lambda + \epsilon} \\ + e^{-s/\tau} g^2 \int_0^\Lambda d\lambda \, \frac{\rho_i(\lambda) - \rho_f(\lambda)}{\lambda + \epsilon}.\end{aligned} \tag{S9.18}$$

定义

$$\Sigma_i = g^2 \int_0^\Lambda d\lambda \, \frac{\rho_i(\lambda)}{\lambda + \epsilon} \tag{S9.19}$$

and

$$\Sigma_f = g^2 \int_0^\Lambda d\lambda \, \frac{\rho_f(\lambda)}{\lambda + \epsilon}, \tag{S9.20}$$

我们得到背景记忆自能

$$\Sigma_{\mathrm{bg}}(s) = \Sigma_f + (\Sigma_i - \Sigma_f) e^{-s/\tau}. \tag{S9.21}$$

对于 $\Sigma_f > \Sigma_i$，该表达式描述了背景记忆响应逐渐增强并随后达到饱和。

特征学习尺度 τ 应与单个模态的内禀弛豫时间 λ_α^{-1}区分开来。前者刻画学习过程中谱结构的演化，后者则刻画给定训练步下的内部弛豫动力学。

S9.4. 沿学习轨迹的临界响应

正文中推导的微观集体相互作用给出静态二次核

$$\Gamma_2(0; s) = \frac{1}{J_{\mathrm{eff}}(s)} - \Pi_0^R(0; s), \tag{S9.22}$$

其中， J_{eff}（s）和 Π_0^R（0； s）分别表示有效集体耦合和静态参考极化。相应的瞬时二次失稳阈值为

$$J_c(s) = \frac{1}{\Pi_0^R(0; s)}. \tag{S9.23}$$

我们引入到集体不稳定性的无量纲距离，

$$\delta(s) = 1 - J_{\mathrm{eff}}(s) \Pi_0^R(0; s) = 1 - \frac{J_{\mathrm{eff}}(s)}{J_c(s)}. \tag{S9.24}$$

在失稳的无序侧， δ（s） > 0。在高斯层面，静态集体磁化率为

$$\chi_m^R(0; s) = \frac{\Pi_0^R(0; s)}{1 - J_{\mathrm{eff}}(s) \Pi_0^R(0; s)} = \frac{\Pi_0^R(0; s)}{\delta(s)}. \tag{S9.25}$$

如果分子在失稳附近保持有限且足够光滑，则主导的临界响应满足

$$\chi_m^R(0; s) \propto \delta(s)^{-1}. \tag{S9.26}$$

更一般地，引入易感性临界指数 γ，我们写作

$$\chi_m^R(0;s) \sim \delta(s)^{-\gamma}, \tag{S9.27}$$

其中

$$\gamma_{\mathrm{MF}} = 1 \tag{S9.28}$$

对应于高斯平均场预测。

现在考虑一条趋近集体不稳定性、随后又远离它但未越过失稳阈值的学习轨迹。令 s_c 表示距离 $\delta(s)$ 达到局部最小值时的训练步。

假设学习轨迹在 s_c 附近可以进行光滑插值，我们进行泰勒展开，

$$\begin{aligned}\delta(s) = \delta(s_c) &+ \delta'(s_c)(s-s_c) \\ &+ \frac{1}{2}\delta''(s_c)(s-s_c)^2 \\ &+ \frac{1}{3!}\delta'''(s_c)(s-s_c)^3 + \cdots .\end{aligned} \tag{S9.29}$$

在光滑非退化极小值处，

$$\delta'(s_c) = 0, \qquad \delta''(s_c) > 0. \tag{S9.30}$$

因此，在最低阶近似下，

$$\delta(s) \simeq \delta_{\min} + a(s-s_c)^2, \tag{S9.31}$$

其中

$$\delta_{\min} \equiv \delta(s_c) > 0, \qquad a \equiv \frac{1}{2}\delta''(s_c) > 0. \tag{S9.32}$$

将式 (S9.31) 代入式 (S9.27)，得到

$$\chi_m^R(0;s) \sim \frac{C}{[\delta_{\min} + a(s-s_c)^2]^{\gamma}}, \tag{S9.33}$$

其中 C 是一个正则振幅，在局部临界区域内近似为常数 n.

引入特征宽度

$$w = \sqrt{\frac{\delta_{\min}}{a}} \tag{S9.34}$$

以及有效响应振幅

$$B_\chi = C\delta_{\min}^{-\gamma}, \tag{S9.35}$$

临界响应变为

$$\chi_m^R(0;s) \simeq \frac{B_\chi}{\left[1 + \left(\frac{s-s_c}{w}\right)^2\right]^{\gamma}}. \tag{S9.36}$$

对于 $\gamma = 1$，这简化为洛伦兹形平均场响应，

$$\chi_m^R(0;s) \simeq \frac{B_\chi}{1 + \left(\frac{s-s_c}{w}\right)^2}. \tag{S9.37}$$

重要的是要区分学习步长依赖性中的二次幂与易感性临界指数 γ。前者源于光滑非退化极小值的局部泰勒展开，而后者刻画了临界响应随与失稳的距离变化的特征。

因此，在这一近似下 中，二次幂是固定的，而非视为独立的拟合指数 .

S9.5. 红外记忆自能的现象学拟合

受上述推导的临界响应启发，我们将红外记忆自能中的瞬态贡献近似为

$$\Delta\Sigma_{\mathrm{crit}}(s) \simeq \frac{B_\Sigma}{\left[1+\left(\frac{s-s_c}{w}\right)^2\right]^\gamma}. \tag{S9.38}$$

这一唯象近似假设，瞬态谱增强沿学习轨迹呈现类临界响应。它并未将记忆自能等同于集体易感性。结合式 (S9.21) 和式 (S9.38)，可得总谱记忆自能：

$$\begin{aligned}\Sigma_\epsilon(0;s) = \Sigma_f + (\Sigma_i - \Sigma_f)e^{-s/\tau} \\ + \frac{B_\Sigma}{\left[1+\left(\frac{s-s_c}{w}\right)^2\right]^\gamma}.\end{aligned} \tag{S9.39}$$

为与数值观测进行比较，我们引入归一化记忆自能

$$\overline{\Sigma}(s) = \frac{\Sigma_\epsilon(0;s)}{\Sigma_{\max}}, \tag{S9.40}$$

其中 $\Sigma_{\max}$ 是可用参考训练轨迹上测得的最大自能。因此，归一化拟合函数为

$$\begin{aligned}\overline{\Sigma}(s) = \overline{\Sigma}_f + (\overline{\Sigma}_i - \overline{\Sigma}_f)e^{-s/\tau} \\ + \frac{B}{\left[1+\left(\frac{s-s_c}{w}\right)^2\right]^\gamma},\end{aligned} \tag{S9.41}$$

其中

$$\overline{\Sigma}_i = \frac{\Sigma_i}{\Sigma_{\max}}, \qquad \overline{\Sigma}_f = \frac{\Sigma_f}{\Sigma_{\max}}, \qquad B = \frac{B_\Sigma}{\Sigma_{\max}}. \tag{S9.42}$$

式 (S9.41) 中的第一项描述逐渐形成的背景谱组织，而第二项描述瞬态类临界增强。

由于瞬态贡献在 $s = 0$ 处通常仍为有限值，初始测得自能满足

$$\overline{\Sigma}(0) = \overline{\Sigma}_i + \frac{B}{\left[1+\left(\frac{s_c}{w}\right)^2\right]^\gamma}. \tag{S9.43}$$

因此，当将初始测得值设为拟合约束时，初始背景振幅由下式确定

$$\overline{\Sigma}_i = \overline{\Sigma}(0) - \frac{B}{\left[1+\left(\frac{s_c}{w}\right)^2\right]^\gamma}. \tag{S9.44}$$

因此，$\overline{\Sigma}_i$ 表示初始背景贡献，而非初始总自能。类似地，s_c 表示瞬态分量的中心，不一定与总自能的最大值重合，因为背景本身会在学习过程中演化。

S9.6. 使用 Transformer 训练检查点进行数值示例

作为一个具有代表性的数值示例，我们分析了 Pythia-410M-deduped Transformer 模型中从第 5 层到第 6 层的相邻层传播。选择这对层，是因为其瞬态增强持续

的训练步区间相对较宽，因此可以利用现有的公开检查点更清晰地解析峰形。

正则化谱自能根据稳定弛豫模计算如下：

$$\Sigma_\epsilon(0;s) = g^2 \sum_{\alpha:\lambda_\alpha(s)>0} \frac{1}{\lambda_\alpha(s)+\epsilon}, \tag{S9.45}$$

数值分析中始终使用 $g = 1$ 和 $\epsilon = 10^{-3}$。

公式 (S9.45) 使用未归一化的经验谱测度。对于维度固定为 N 的雅可比矩阵，该测度与公式 (S9.4) 中的归一化总态密度表示相差一个整体因子 N。若使用相同的正速率区段和谱上截止，并按测得的最大自能进行归一化，该因子便会抵消。此处的截止包含经验谱中的所有正速率。

与公式 (S1.3) 中定义的集体通道极化不同，这一数值观测量不包含显式的特征向量重叠权重。因此，它提供的是红外记忆增强的谱测度，而非对投影集体易感性的直接测量。

我们比较两种基于公式 (S9.41) 的拟合模型：

$$\text{Model I:} \quad \gamma = 1, \tag{S9.46}$$

$$\text{Model II:} \quad \gamma \text{ free}. \tag{S9.47}$$

在两种模型中，$(s - s_c)/w$ 的二次依赖由临界距离的局部展开确定。初始总自能受公式 (S9.44) 约束，而有效的训练后期背景振幅 Σ_f 则固定为同一数据导出的训练后期估计值。

拟合在训练区间 $0 \le s \le 30000$ 内进行，并使用该区间内可用的九个数值测量。所得拟合参数汇总于表 S2。

表 S2. Pythia-410M-deduped 第 $5 \to 6$ 层的归一化红外记忆自能的现象学拟合参数。模型 I 使用平均场指数 $\gamma = 1$，而模型 II 允许 γ 变化。γ 的不确定性是由局部拟合得到的条件标准误。

参数	模型 I	模型 II
$\overline{\Sigma}_i$	0.2063	0.3046
$\overline{\Sigma}_f$	0.6386	0.6386
τ	1532.8	1016.0
B	0.4994	0.4331
s_c	1430.3	1578.3
w	1884.8	1123.7
γ	1（固定）	0.63 ± 0.51
RMSE	0.02792	0.02679
R^2	0.95895	0.96219

两个模型都能以相近的拟合精度描述红外记忆自能的瞬态增强。当允许响应指数变化时，所得有效指数为

$$\gamma_{\text{fit}} = 0.63 \pm 0.51. \tag{S9.48}$$

尽管不确定性较大，拟合得到的指数与平均场值 $\gamma_{\text{MF}} = 1$ 相符。此外，允许指数变化只能略微降低均方根拟合误差。

瞬态增强具有有限宽度，这与学习轨迹上圆化的类临界响应相一致。然而，仅凭测得的自能，无法证明系统正在趋近真正的集体不稳定性。

S9.7. 标度形式及其局限性

现象学拟合函数也可以表示为无量纲标度形式。减去背景贡献并引入缩放训练坐标

$$x = \frac{s - s_c}{w}, \tag{S9.49}$$

我们得到

$$\frac{\overline{\Sigma}(s)-\overline{\Sigma}_{\mathrm{bg}}(s)}{B}=\mathcal{F}_{\gamma}(x), \tag{S9.50}$$

其中

$$\mathcal{F}_{\gamma}(x)=\frac{1}{(1+x^{2})^{\gamma}}. \tag{S9.51}$$

相应的平均场标度函数为

$$\mathcal{F}_{\mathrm{MF}}(x)=\frac{1}{1+x^{2}}. \tag{S9.52}$$

在中间渐近区间内，$|x|\gg 1$，只要局部二次展开仍然有效，标度函数就呈现出幂律行为

$$\mathcal{F}_{\gamma}(x)\sim|x|^{-2\gamma}. \tag{S9.53}$$

这种表示为未来比较不同层、模型规模和学习轨迹之间的标度关系提供了一种可能的框架。若使用同一组数据确定拟合参数后得到标度塌缩，这将构成内部一致性检验，而非对普适性的独立证明。

公开可用的训练检查点限制了对瞬态增强的时间分辨率，尤其是在其达到最大值附近。要定量确定临界指数，需要采样更密集的训练轨迹、对距集体不稳定性的距离进行独立估计，并直接测量集体通道易感性。

因此，本分析旨在对微观临界响应框架进行唯象数值示例，而非确定 Transformer 学习的普适类。

尽管如此，所观察到的红外记忆自能的瞬态增强，与叠加在逐步演化的谱背景之上的类平均场临界响应是一致的。