VANISHING INITIALIZATION TURNS DIAGONAL TENSOR GRADIENT FLOW INTO GREEDY MATCHING

Anonymous authors
Paper under double-blind review

ABSTRACT

We study gradient-flow (GF) dynamics of canonical polyadic (CP) tensor factorization from vanishing initialization. For order-ν models with $\nu \geq 3$ and targets diagonal in mode-wise orthonormal bases, we prove that the limiting training trajectory is governed by greedy bipartite matching. Each atom–sector pair acquires a random activation clock determined by its microscopic initialization; processing these clocks in increasing order, subject to atom and sector exclusivity, determines the successive learning events. Away from activation times, the tensor output converges to the corresponding staircase trajectory. Complex modular addition is an exact equal-sector instance after Fourier diagonalization, so the activation clocks are i.i.d. Combining this reduction with known Random Edge asymptotics yields an explicit distribution-free large-system loss curve in the edge-exposure parametrization; we further derive its asymptotic conversion to physical training time. For real weights, conjugate Fourier modes instead form two-dimensional sectors. We characterize their rank and border-rank cascade and conjecture that the full GF dynamics is organized by the corresponding interleaved cascades.

1 INTRODUCTION [Appendix A1]

Very small initialization can qualitatively reorganize gradient-flow training in homogeneous factorized models, producing incremental or saddle-to-saddle dynamics: long plateaus near low-complexity states separated by rapid transitions. In matrix factorization and linear networks, such dynamics can successively learn singular modes, coordinates, or increasing-rank solutions (Saxe et al., 2014; Gissin et al., 2020; Li et al., 2021; Jacot et al., 2021; Pesme & Flammarion, 2023; Jin et al., 2023), with complete limiting trajectories characterized in some settings.

In contrast to matrix factorization and linear networks, complete small-initialization gradient-flow dynamics in higher-order tensor factorization remain much less understood. Tensors of order $\nu \geq 3$ lack a canonical analog of the matrix SVD, have more complicated low-rank geometry, including degeneracies and border-rank phenomena, and exhibit nonlinear near-origin dynamics. Existing work has nevertheless established important forms of greedy low-rank and deflation-like behavior (Razin et al., 2021; Ge et al., 2021), but a complete characterization of the vanishing-initialization trajectory remains largely missing.

Higher-order tensor learning problems arise naturally in structured multi-input tasks, such as learning binary or higher-arity operations. A finite binary operation, for example, can be represented by an order-three target tensor. One of the simplest and most studied examples is *modular addition*. Its gradient-based learning has attracted substantial recent attention, especially in connection with grokking and Fourier feature learning (Power et al., 2022; Nanda et al., 2023; Mohamadi et al., 2024; He et al., 2026a). Despite this progress, we are not aware of a rigorous asymptotic characterization of the full loss trajectory for modular addition in the small-initialization feature-learning regime.

In the present paper, we approach this problem through a more general theory of diagonal tensor learning. We consider targets that decompose into mutually orthogonal rank-one components, or *scalar sectors*, and prove that the vanishing-initialization limit of ordinary CP gradient flow reduces to *greedy bipartite matching* between model atoms and target sectors. The candidate activation times, which we call *activation clocks*, are determined by the microscopic random initialization. We then show that modular addition with *complex* weights falls exactly into this framework after Fourier diagonalization. This yields both a finite-system description of its learning trajectory and an

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

explicit large-system loss curve after a natural time reparametrization. Unlike prior deflation-style or auxiliary event-driven descriptions, our result determines the complete finite-system sequence of activation events for ordinary asymmetric CP gradient flow.

The resulting limit is strikingly different from the familiar $\nu = 2$ matrix factorization picture. There, small-initialization dynamics is typically organized by deterministic singular directions and logarithmic escape times. For $\nu \geq 3$, after the natural time rescaling, information from the microscopic Gaussian initialization survives the small amplitude limit as a collection of finite random activation times. Thus even for a completely symmetric equal-amplitude target, the order in which components are learned need not become deterministic. At the same time, this apparently more random higher-order dynamics admits an explicit analytic description, and in the large-system limit its CDF-reparametrized loss converges to a universal deterministic curve.

The choice between complex and real weights is crucial for modular addition. With complex weights, Fourier transformation separates the target into scalar sectors and makes the preceding theory directly applicable. With real weights, conjugate Fourier modes merge into two-dimensional blocks, and the simple one-stage sector picture is replaced by multi-stage rank and border-rank cascades, accompanied by large parameter growth. The resulting dynamics is substantially more complicated. We characterize the geometry of these real blocks and conjecture that the full GF evolution reduces to the corresponding interleaved cascades.

Taken together, our results rigorously establish the pipeline

CP gradient flow $\longrightarrow$ random activation clocks $\longrightarrow$ greedy matching $\longrightarrow$ large-system loss.

Contributions. Our main contributions are:

- We prove that vanishing-initialization gradient flow for diagonal order-ν CP tensor factorization, $\nu \geq 3$, reduces to greedy bipartite matching driven by microscopic activation clocks (Theorem 1).
- For equal-amplitude sectors, we derive an explicit large-system loss trajectory through Random Edge matching. For complex modular addition, we additionally obtain the asymptotic conversion to physical gradient-flow time (Theorems 2 and 3).
- Beyond scalar sectors, we characterize the rank and border-rank cascade of a two-dimensional real Fourier block (Proposition 2) and conjecture that real modular addition exhibits interleaved multi-stage versions of this cascade.

2 SETUP AND DIAGONAL TARGETS [Appendix A2]

Model and diagonal targets. Let the field $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and $\nu \geq 3$. We consider the task of fitting an order-ν tensor $T \in (\mathbb{K}^p)^{\otimes \nu}$ by a CP model with H atoms:

$$\widehat{T}(U) = \sum_{k=1}^{H} \mathbf{u}_k^{(1)} \otimes \cdots \otimes \mathbf{u}_k^{(\nu)}, \qquad U = (\mathbf{u}_k^{(a)})_{a\in[\nu],\, k\in[H]} \in (\mathbb{K}^p)^{\nu H} \cong \mathbb{K}^{\nu\times H\times p}. \tag{1}$$

The weights U are learned by gradient flow (GF) with the squared Frobenius loss:

$$L(U) = \frac{1}{2}\|T - \widehat{T}(U)\|_F^2, \qquad \frac{d\mathbf{u}_k^{(a)}}{dt} = -\nabla_{\mathbf{u}_k^{(a)}} L(U). \tag{2}$$

For $\mathbb{K} = \mathbb{C}$, gradients are taken with respect to the underlying real Euclidean structure on $\mathbb{C}^p$; the equivalent Wirtinger formulation is given in Appendix A2.

The summand $\mathbf{u}_k^{(1)} \otimes \cdots \otimes \mathbf{u}_k^{(\nu)}$ will be called the kth *atom*, and each $a \in [\nu]$ indexes a *mode*. We initialize the weights independently as $\mathbf{u}_k^{(a)}(0) = \sigma \mathbf{g}_k^{(a)}$, with $\mathbf{g}_k^{(a)} \sim \mathcal{N}_{\mathbb{K}}(0, I_p)$. For $\mathbb{K} = \mathbb{C}$, we use the standard circularly symmetric convention $Z = (X + iY)/\sqrt{2}$, where $X, Y \sim \mathcal{N}(0, 1)$ independently, so that $\mathbb{E}|Z|^2 = 1$.

We study diagonal targets

$$T = \sum_{i=1}^{p} \lambda_i \mathbf{e}_i^{(1)} \otimes \cdots \otimes \mathbf{e}_i^{(\nu)}, \qquad \lambda_i > 0, \tag{3}$$

where $\{\mathbf{e}_i^{(a)}\}_{i=1}^p$ is an orthonormal basis in each mode a. We assume all diagonal coefficients are nonzero; their signs or complex phases can be absorbed into the corresponding basis vectors, giving $\lambda_i > 0$ without loss of generality. The orthonormal bases $\{\mathbf{e}_i^{(a)}\}$ need not coincide with the original coordinate bases; this will be important for modular addition, where they are Fourier bases.

We refer to the ith summand in (3) as the ith *scalar sector*. For each atom k and sector i, we denote the coordinates in the diagonalizing bases as $x_{ki}^{(a)}(t) = \langle \mathbf{e}_i^{(a)}, \mathbf{u}_k^{(a)}(t) \rangle$ and $g_{ki}^{(a)} = \langle \mathbf{e}_i^{(a)}, \mathbf{g}_k^{(a)} \rangle$. By isotropy, the $g_{ki}^{(a)}$ are independent standard scalar Gaussians over $\mathbb{K}$. Throughout the finite-dimensional vanishing-initialization limit, p, H, ν are fixed and we use the natural slow-time rescaling $\theta = \sigma^{\nu-2}t$; its origin from the near-origin GF equations is reviewed in Appendix A2.

Activation clocks. At the beginning of training, while all coordinates remain of order σ and no target sector has yet been learned, the leading dynamics of each atom–sector pair decouples into a ν-dimensional auxiliary flow. For $\mathbf{g} = (g^{(1)}, \dots, g^{(\nu)}) \in \mathbb{K}^\nu$, let $\mathbf{z}(s; \mathbf{g}) = (z^{(1)}(s), \dots, z^{(\nu)}(s))$ solve

$$\frac{dz^{(a)}}{ds} = \prod_{b \neq a} \overline{z^{(b)}}, \qquad z^{(a)}(0) = g^{(a)}, \qquad \tau_\nu(\mathbf{g}) := \sup\{s > 0 : \|\mathbf{z}(s; \mathbf{g})\| < \infty\}, \tag{4}$$

where complex conjugation is immaterial over $\mathbb{R}$. Here s is only the time parameter of the isolated unit-amplitude flow. Its blow-up time $\tau_\nu(\mathbf{g})$ marks when the corresponding near-origin approximation leaves the small regime and enters a fast activation.

For atom k and target sector i, let $\mathbf{g}_{ki} = (g_{ki}^{(1)}, \dots, g_{ki}^{(\nu)})$. A sector of amplitude λ_i rescales the auxiliary time by $s = \lambda_i \theta$, giving the candidate activation time

$$\tau_{ki} = \frac{\tau_\nu(\mathbf{g}_{ki})}{\lambda_i}. \tag{5}$$

We call τ_{ki} the *activation clock* of the pair (k, i): it is the slow time at which this atom–sector pair would activate if it evolved in isolation. Thus each clock is fixed by the random initialization, while competition between the clocks determines which pairs actually activate.

For Gaussian $\mathbf{g}$, $\tau_\nu(\mathbf{g})$ is finite almost surely and has an atomless distribution; these properties and the scalar-flow analysis are given in A2. Consequently, the Hp candidate clocks are independent and almost surely finite and distinct; when the sector amplitudes λ_i are equal, they are i.i.d.

The pairs (k, i) can now be viewed as the edges of the complete bipartite graph $K_{H,p}$ between H model atoms and p target sectors, each carrying the clock (5). Section 3 shows that GF processes these clocks greedily as $\sigma \to 0$.

3 Vanishing initialization yields greedy matching [Appendix A3]

Greedy clock process. Recall that the candidate pairs (k, i) form the edges of $K_{H,p}$, each carrying the activation clock τ_{ki} from (5). Process these edges in increasing order of their clocks. An edge is accepted if neither endpoint is incident to an earlier accepted edge; otherwise it is discarded. Let $\mathcal{M}(\theta)$ be the matching formed by accepted edges with clock at most θ, and let $\mathcal{S}(\theta) = \{i : (k, i) \in \mathcal{M}(\theta) \text{ for some } k\}$ be the corresponding set of selected target sectors. The greedy loss trajectory is

$$L_*(\theta) = \frac{1}{2} \sum_{i \notin \mathcal{S}(\theta)} \lambda_i^2. \tag{6}$$

Thus each accepted edge removes exactly one scalar sector from the residual.

For $\sigma > 0$, write $\widehat{T}^\sigma(\theta) = \widehat{T}(U^\sigma(\sigma^{-(\nu-2)}\theta))$ and $L^\sigma(\theta) = L(U^\sigma(\sigma^{-(\nu-2)}\theta))$, where U^σ is the GF solution initialized at scale σ. Let $\mathcal{T}_*$ denote the finite set of clocks of the edges accepted by the greedy process.

Theorem 1 (Greedy matching limit)**.** *Under the setup of Section 2, for almost every realization of the Gaussian initialization, the following holds. For every compact $K \subset [0, \infty) \setminus \mathcal{T}_*$,*

$$\sup_{\theta \in K} \Big\| \widehat{T}^\sigma(\theta) - \sum_{i \in \mathcal{S}(\theta)} \lambda_i \mathbf{e}_i^{(1)} \otimes \cdots \otimes \mathbf{e}_i^{(\nu)} \Big\|_F \longrightarrow 0 \qquad \text{as } \sigma \to 0, \tag{7}$$

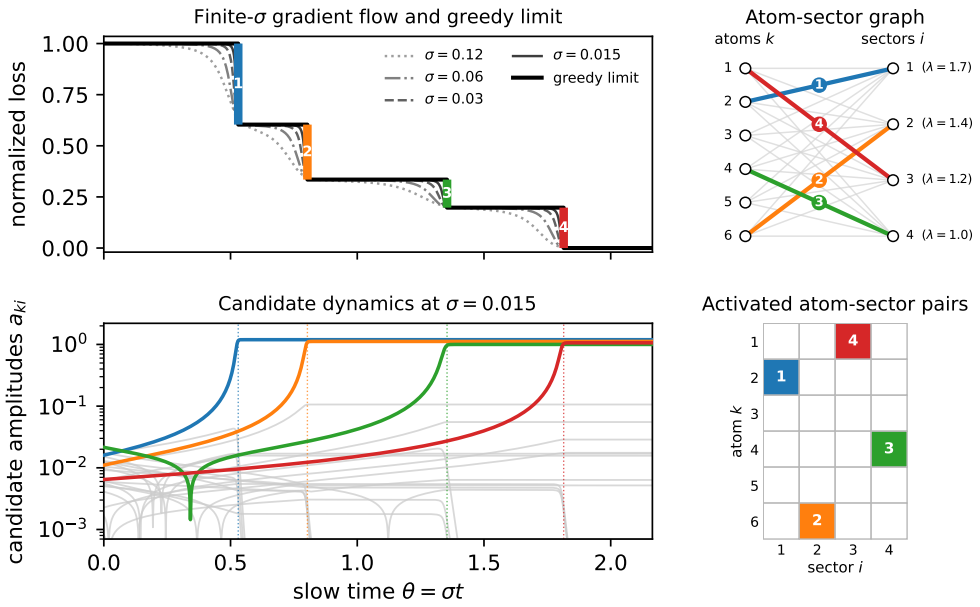


Figure 1: **Vanishing-initialization learning as greedy matching for a diagonal tensor.** Example with $p = 4$, $H = 6$, $\nu = 3$, and unequal sector amplitudes $\lambda = (1.7, 1.4, 1.2, 1.0)$. *Top left:* normalized GF loss for several initialization scales σ, plotted against the slow time $\theta = \sigma t$, together with the predicted greedy-matching staircase. As σ decreases, the finite-σ transitions sharpen around the greedy activation times. *Bottom left:* candidate amplitudes $a_{ki} = \left|\prod_{\mu=1}^{\nu} x_{ki}^{(\mu)}\right|^{1/\nu}$ for $\sigma = 0.015$. Selected atom–sector pairs are highlighted in color and losing candidates in gray. The third selected trajectory initially decreases because $A_0 = P(0) < 0$ in its isolated scalar flow, producing the turning behavior described in (43)–(46). *Top right:* atom–sector graph, with accepted greedy matches colored and numbered according to their activation order. *Bottom right:* the same accepted pairs shown as cells of the $H \times p$ atom–sector grid. Colors and event numbers agree across all panels.

and consequently

$$\sup_{\theta \in K} \left| L^{\sigma}(\theta) - L_*(\theta) \right| \longrightarrow 0. \tag{8}$$

Hence, away from activation times, the vanishing-initialization loss converges to the staircase (6) determined by greedy matching of the activation clocks.

Figure 1 illustrates this limiting process. Since all candidate clocks are finite almost surely, the greedy process eventually produces a maximal matching of $K_{H,p}$. Hence, if $H \geq p$, it selects every sector and $L_*(\theta)$ eventually vanishes; if $H < p$, it selects exactly H sectors. For unequal λ_i, these need not be the H largest coefficients because the outcome depends on the random initialization.

Proof idea. Before the first activation, all atom coordinates remain microscopic. After writing $x_{ki}^{(a)} = \sigma z_{ki}^{(a)}$ and using the slow time $\theta = \sigma^{\nu-2} t$, the exact equations reduce, to leading order, to

$$\frac{dz_{ki}^{(a)}}{d\theta} = \lambda_i \prod_{b \neq a} \overline{z_{ki}^{(b)}} + o(1), \tag{9}$$

so all candidate pairs follow their isolated scalar flows until the smallest clock is reached. Near that clock, the winning atom becomes macroscopic over a vanishing interval of slow time and aligns with the corresponding scalar sector, while the other unactivated atoms remain microscopic. The learned sector is thereby removed from the residual, and the winning atom becomes asymptotically committed to that sector.

Because the diagonal sectors are orthogonal, this event does not change the leading microscopic dynamics in the still-unclaimed sectors. Crucially, candidates whose atom and sector remain available continue along their original isolated trajectories: their clocks are not reset. The next activation is therefore the smallest remaining clock whose two endpoints are still available. Iterating gives exactly the greedy matching process above. The full argument, including uniform control of the microscopic regime and the fast activation transitions, is given in Appendix A3.

4 COMPLEX MODULAR ADDITION [Appendix A4]

Let $\mathbb{Z}_p = \{0, \dots, p-1\}$ and consider the modular-addition tensor $T_p^{\mathrm{add}} \in \mathbb{R}^{p\times p\times p}$,

$$(T_p^{\mathrm{add}})_{x,y,z} = \mathbf{1}\{x + y = z \pmod p\}.$$

Writing $\omega = e^{2\pi i/p}$ and $\boldsymbol{\varphi}_r(x) = p^{-1/2}\omega^{rx}$, the Fourier representation of the modular Kronecker delta gives

$$T_p^{\mathrm{add}} = \sqrt{p} \sum_{r\in\mathbb{Z}_p} \boldsymbol{\varphi}_r \otimes \boldsymbol{\varphi}_r \otimes \overline{\boldsymbol{\varphi}_r}. \tag{10}$$

For $\nu = 3$, CP model (1) can equivalently be viewed as a shallow network with H multiplicative hidden units, $f_U(x,y)_z = \sum_{k=1}^{H} u_k^{(1)}(x) u_k^{(2)}(y) u_k^{(3)}(z)$, trained full-batch on all p^2 input pairs (x,y); its summed squared output loss is exactly $L(U)$. Thus our results concern the training dynamics of this full-data model, rather than train/test generalization or grokking.

Over $\mathbb{C}$, (10) is exactly an equal-amplitude diagonal order-3 tensor in the sense of Section 2, with $\mathbf{e}_r^{(1)} = \mathbf{e}_r^{(2)} = \boldsymbol{\varphi}_r$, $\mathbf{e}_r^{(3)} = \overline{\boldsymbol{\varphi}_r}$, and $\lambda_r = \sqrt{p}$.

By unitary invariance of the complex Gaussian initialization, the microscopic atom–sector coordinates are i.i.d. $\mathbf{g}_{kr} \sim \mathcal{N}_{\mathbb{C}}(0, I_3)$. Setting $\xi_{kr} := \tau_3(\mathbf{g}_{kr})$, their activation clocks are $\tau_{kr} = \xi_{kr}/\sqrt{p}$. We therefore use the rescaled time $\vartheta = \sqrt{p}\,\theta = \sqrt{p}\,\sigma t$, in which the candidate activation times are ξ_{kr}. Let $S_p(\vartheta)$ be the set of selected Fourier sectors and $M_p(\vartheta) = |S_p(\vartheta)|$.

Write $\widehat{T}_p^{\mathrm{add},\sigma}(\vartheta)$ and $L_p^{\mathrm{add},\sigma}(\vartheta)$ for the GF output and loss at $\theta = \vartheta/\sqrt{p}$.

Corollary 1 (Greedy Fourier-sector learning)**.** *Fix p and H. For almost every complex Gaussian initialization, away from the accepted activation times,*

$$\lim_{\sigma\to 0} \widehat{T}_p^{\mathrm{add},\sigma}(\vartheta) = \sqrt{p} \sum_{r\in S_p(\vartheta)} \boldsymbol{\varphi}_r \otimes \boldsymbol{\varphi}_r \otimes \overline{\boldsymbol{\varphi}_r}, \qquad \frac{2L_{*,p}^{\mathrm{add}}(\vartheta)}{p^2} = 1 - \frac{M_p(\vartheta)}{p}, \tag{11}$$

where $L_{*,p}^{\mathrm{add}}(\vartheta) = \lim_{\sigma\to 0} L_p^{\mathrm{add},\sigma}(\vartheta)$.

Here $\|T_p^{\mathrm{add}}\|_F^2 = p^2$, so (11) is the normalized loss. Equal sector amplitudes make the edge clocks i.i.d. and continuous, hence uniformly ordered; their distribution determines only the time parametrization, while sector competition is governed by Random Edge matching.

5 LARGE-SYSTEM LIMIT [Appendix A5]

Random Edge matching. Let $H_p/p \to \alpha \in (0,\infty)$ and consider the complete bipartite graph $K_{H_p,p}$ between H_p atoms and p sectors. Assign each edge an independent uniform priority in $[0,1]$, process the edges in increasing order of priority, and accept an edge whenever both endpoints are still unmatched. Following Arnosti (2022), we call this the *Random Edge* algorithm.

At edge exposure $s \in [0,p]$, we have processed all edges with priority at most s/p. The exposed edges form the bipartite Erdős–Rényi graph $G_{H_p,p,s/p}$; the expected exposed degree is s on the atom side and $(H_p/p)s \to \alpha s$ on the sector side. Let $M_p^{\mathrm{RE}}(s)$ denote the number of edges accepted by this stage. Any i.i.d. continuous edge-clock distribution yields the same process after applying its CDF, which transforms the clocks into independent uniform priorities.

Proposition 1 (Random Edge matching limit)**.** *Suppose that $H_p/p \to \alpha \in (0,\infty)$. Then, for every fixed $S < \infty$,*

$$\sup_{0\le s\le S}\left|\frac{M_p^{\mathrm{RE}}(s)}{p} - \big(1-\ell_\alpha(s)\big)\right| \xrightarrow{\mathbb{P}} 0, \qquad \ell_\alpha(s) = \begin{cases} \dfrac{(\alpha-1)e^{-(\alpha-1)s}}{\alpha - e^{-(\alpha-1)s}}, & \alpha \neq 1, \\ \dfrac{1}{1+s}, & \alpha = 1. \end{cases} \tag{12}$$

Thus $\ell_\alpha(s)$ is the limiting fraction of unmatched sectors. Here and below, $\xrightarrow{\mathbb{P}}$ denotes convergence in probability. The explicit formula in Proposition 1 is the bipartite Random Edge asymptotic of Arnosti (2022); related balanced results appeared earlier in Dyer et al. (1993). The process-level statement also follows from the standard differential-equation method (Wormald, 1995; 1999). We use this known random-graph limit as an input to the tensor-dynamics reduction, rather than as a new combinatorial result.

Equal diagonal sectors. Consider the equal-amplitude diagonal target

$$T_p = \lambda_p \sum_{i=1}^{p} \mathbf{e}_i^{(1)} \otimes \cdots \otimes \mathbf{e}_i^{(\nu)}, \qquad \frac{H_p}{p} \to \alpha, \tag{13}$$

for fixed $\nu \ge 3$ and $\lambda_p > 0$. Let $U_{p,\sigma}(t)$ denote the GF solution with initialization scale σ, and let

$$F_\nu(x) := \mathbb{P}(\tau_\nu(\mathbf{g}) \le x), \qquad \mathbf{g} \sim \mathcal{N}_{\mathbb{K}}(0, I_\nu),$$

be the CDF of the scalar activation clock. The candidate clocks are $\tau_{ki} = \xi_{ki}/\lambda_p$, where $\xi_{ki} = \tau_\nu(\mathbf{g}_{ki})$ have common CDF F_ν. Thus, the variables $F_\nu(\xi_{ki})$ from the same Gaussian initialization are i.i.d. uniform priorities, giving an exact coupling to the Random Edge process above.

At physical training time t, the slow-time threshold is $\theta = \sigma^{\nu-2}t$. Hence the corresponding edge exposure and its inverse are

$$s_{p,\sigma}(t) = pF_\nu\big(\lambda_p\sigma^{\nu-2}t\big), \qquad t_{p,\sigma}(s) = \frac{1}{\sigma^{\nu-2}\lambda_p}F_\nu^{-1}\left(\frac{s}{p}\right), \tag{14}$$

where F_ν^{-1} denotes the inverse of F_ν (it is well-defined by Proposition 5).

Theorem 2 (Large-system loss trajectory for equal diagonal sectors)**.** *Let $H_p/p \to \alpha \in (0,\infty)$ and fix $\nu \ge 3$. For every compact $K \subset [0,\infty)$ and every $\varepsilon > 0$,*

$$\lim_{p\to\infty}\limsup_{\sigma\downarrow 0}\mathbb{P}\left(\sup_{s\in K}\left|\frac{2L(U_{p,\sigma}(t_{p,\sigma}(s)))}{p\lambda_p^2} - \ell_\alpha(s)\right| > \varepsilon\right) = 0, \tag{15}$$

where ℓ_α is given in (12).

Thus, after reparametrization by the common clock CDF, the limiting loss in edge exposure depends only on the width ratio α. The tensor order and microscopic clock law enter only through the conversion (14) back to physical training time. Figure 2 illustrates this separation: panel (b) shows the Random Edge loss dynamics in edge exposure, panel (a) shows the corresponding physical-time trajectory, and panel (c) shows the dependence on α.

Explicit training time for complex modular addition. For complex modular addition, $\nu = 3$ and every Fourier sector has amplitude $\lambda_p = \sqrt{p}$. Hence (14) specializes to $s_{p,\sigma}(t) = pF_3(\sqrt{p}\,\sigma t)$. Although this exact conversion involves the full cubic clock distribution, in the large-system limit it admits an explicit asymptotic inverse.

Theorem 3 (Explicit large-system loss trajectory for complex modular addition)**.** *Assume $H_p/p \to \alpha \in (0,\infty)$ and define*

$$t_{p,\sigma}^{\mathrm{as}}(s) = \frac{\sqrt{3}}{\sigma\sqrt{p}}\left[\log p + \frac{1}{2}\log\log p + \log A_3 - \log s\right]^{-1/2}, \qquad A_3 = \frac{5\sqrt{21\pi}}{27}. \tag{16}$$

Then, uniformly for s in compact subsets of $(0,\infty)$,

$$s_{p,\sigma}\big(t_{p,\sigma}^{\mathrm{as}}(s)\big) \longrightarrow s \qquad \text{as } p \to \infty. \tag{17}$$

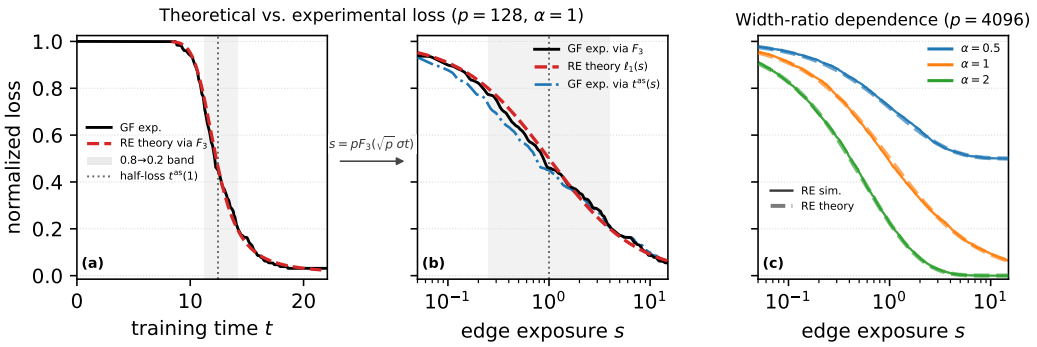


Figure 2: **Large-system loss dynamics for complex modular addition.** *(a)* Normalized GF loss for $p = 128$, $H = p$ ($\alpha = 1$), and $\sigma = 0.005$ in physical training time t. The dashed curve is the Random Edge (RE) prediction $\ell_1(s) = (1+s)^{-1}$ transported to physical time using the numerically evaluated clock CDF F_3. The shaded region is the explicit-asymptotic $0.8 \to 0.2$ loss-drop interval, and the dotted line marks the corresponding half-loss time $t^{\mathrm{as}}_{p,\sigma}(1)$. *(b)* The same experiment in edge exposure $s = pF_3(\sqrt{p}\,\sigma t)$. The reparameterized trajectory closely follows the limiting Random Edge curve $\ell_1(s)$; using the explicit asymptotic inverse $t^{\mathrm{as}}_{p,\sigma}(s)$ gives a similarly close reparameterization. The shaded region is $s \in [1/4, 4]$, where $\ell_1(s)$ decreases from 0.8 to 0.2, and the dotted line marks $s = 1$. *(c)* Dependence on the width ratio at $p = 4096$. Simulated finite-p Random Edge trajectories for $\alpha = 1/2, 1, 2$ are compared with the limiting curves $\ell_\alpha(s)$.

Consequently, for every compact $K \subset (0,\infty)$ and every $\varepsilon > 0$,

$$\lim_{p\to\infty} \limsup_{\sigma\downarrow 0} \mathbb{P}\left(\sup_{s\in K} \left| \frac{2L\big(U_{p,\sigma}\big(t^{\mathrm{as}}_{p,\sigma}(s)\big)\big)}{p^2} - \ell_\alpha(s) \right| > \varepsilon \right) = 0, \tag{18}$$

where ℓ_α is the Random Edge loss curve from Proposition 1.

The proof is given in Appendix A5. It derives the lower-tail asymptotics of the complex cubic activation clock and uses them to invert the exact exposure-time conversion (14). Thus, in the learning window, the full clock CDF can be replaced by the explicit physical-time formula (16).

For every fixed $s \in (0,\infty)$, the corresponding physical training time satisfies

$$t^{\mathrm{as}}_{p,\sigma}(s) \sim \frac{\sqrt{3}}{\sigma\sqrt{p\log p}}. \tag{19}$$

The physical-time separation between two fixed exposure levels is smaller. More precisely, for fixed $0 < s_- < s_+ < \infty$,

$$t^{\mathrm{as}}_{p,\sigma}(s_+) - t^{\mathrm{as}}_{p,\sigma}(s_-) = \frac{\sqrt{3}\,\log(s_+/s_-)}{2\sigma\sqrt{p}\,(\log p)^{3/2}}\big(1 + o(1)\big). \tag{20}$$

At balanced width, $\ell_1(s) = (1+s)^{-1}$, so $s = 1$ corresponds to the asymptotic half-loss time.

6 Beyond Scalar Sectors: Real Fourier Blocks [Appendix A6]

The use of complex weights is crucial for our exact treatment of modular addition in Sections 4–5. Over $\mathbb{C}$, Fourier decomposition separates the target into scalar rank-one sectors. Over $\mathbb{R}$, the conjugate frequencies r and $-r$ must instead be combined into a two-dimensional block, so real modular addition falls outside the scalar-sector theory of Theorem 1.

For such a two-dimensional Fourier block, suitable cosine–sine coordinates give, up to an overall amplitude,

$$B = eee - eff - fef - ffe = \frac{1}{2}\left[(\mathbf{e} + i\mathbf{f})^{\otimes 3} + (\mathbf{e} - i\mathbf{f})^{\otimes 3}\right]. \tag{21}$$

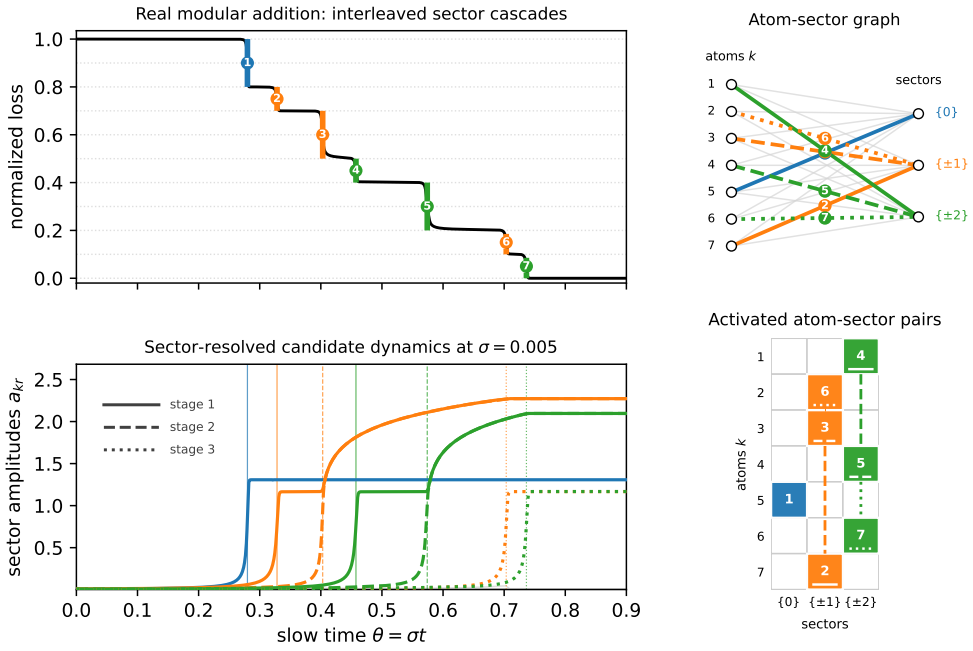


Figure 3: **Interleaved learning of non-scalar Fourier sectors in real modular addition.** Gradient flow for $p = 5$, $H = 7$, and $\sigma = 0.005$, with real Fourier sectors $\{0\}$, $\{\pm 1\}$, and $\{\pm 2\}$. *Top left:* normalized loss; numbered markers show the empirically identified activation sequence, with drops occurring near multiples of 0.1. *Bottom left:* sector-resolved amplitudes $a_{kr} = \big(\prod_{\mu=1}^{3} \|P_r^{(\mu)} \mathbf{u}_k^{(\mu)}\|_2\big)^{1/3}$. Colors identify sectors and line styles distinguish successive stages. Unlike in a scalar sector, previously activated atoms continue to evolve as later atoms enter the same sector. *Top right:* atom–sector graph showing the corresponding activation assignments. *Bottom right:* the same assignments as an atom–sector grid; numbers give the global activation order, while dashed and dotted connections link the first–second and second–third stages of each non-scalar cascade. The two three-stage cascades are visibly interleaved.

Here, for example, eff abbreviates $\mathbf{e}^{(1)} \otimes \mathbf{f}^{(2)} \otimes \mathbf{f}^{(3)}$. Over $\mathbb{C}$, this is a sum of two conjugate rank-one sectors; over $\mathbb{R}$, the same tensor has CP rank three.

For a real tensor X, let $\mathrm{rank}_{\mathbb{R}}(X)$ denote its real CP rank and define

$$d_q(B)^2 := \inf_{\mathrm{rank}_{\mathbb{R}}(X) \le q} \|B - X\|_F^2.$$

Proposition 2 (Rank cascade of a real Fourier block)**.** *For the canonical block B of modular addition in (21),*

$$d_0(B)^2 = 4, \qquad d_1(B)^2 = 3, \qquad d_2(B)^2 = 1, \qquad d_3(B)^2 = 0. \tag{22}$$

The rank-two infimum is not attained.

More precisely, there is a phase-dependent family (E_Θ, F_Θ) such that

$$B = E_\Theta - F_\Theta, \qquad \|E_\Theta\|_F^2 = 1, \qquad \|F_\Theta\|_F^2 = 3, \tag{23}$$

where E_Θ is rank one and F_Θ is a sum of three mutually orthogonal rank-one tensors. Moreover, $-F_\Theta$ lies in the closure of the real rank-two tensors and can be approached only through degenerating rank-two representations with diverging factors.

The definitions and proof are given in Appendix A6. The proposition identifies the residual cascade

$$B \longrightarrow -F_\Theta \longrightarrow E_\Theta \longrightarrow 0, \qquad \frac{\|R\|_F^2}{\|B\|_F^2} : 1 \longrightarrow \frac{3}{4} \longrightarrow \frac{1}{4} \longrightarrow 0. \tag{24}$$

The first intermediate residual $-F_\Theta$ has real rank three but border rank two: approaching it with rank-two tensors requires diverging factors even though the represented tensor remains bounded.

Figure 3 shows one representative small-initialization GF trajectory for real modular addition. Motivated by this behavior, which we believe to be typical, we conjecture that GF forms interleaved multi-stage cascades: successive atoms enter a sector while previously activated atoms reorganize, and cascades from different Fourier sectors interleave. The realized phase Θ depends on the microscopic initialization. For $p = 5$, each of the two nonzero sectors carries $2/5$ of the target energy, so the relative levels in (24) produce global loss changes in multiples of $1/10$, as seen in the figure.

7 DISCUSSION

Results and significance. We rigorously characterize the vanishing-initialization dynamics of diagonal order-ν CP tensor factorization for $\nu \geq 3$. We prove that the limiting learning trajectory is governed by greedy bipartite matching between model atoms and target sectors. This gives an explicit trajectory-level description of a nonlinear higher-order factorization problem. For equal-amplitude sectors, our results further yield an explicit large-system loss law: after reparametrization by edge exposure, the limiting normalized loss depends only on the atom-to-sector ratio. To our knowledge, such an exact matching representation of the GF trajectory, together with the resulting large-system loss law, has not appeared in previous analyses. As a concrete application, complex modular addition falls exactly within this framework, yielding an explicit description of its large-p loss dynamics.

Why tensor learning differs from the matrix case. In matrix and deep-linear settings, small-initialization dynamics is typically organized by spectral structure. In our tensor case, the elementary learning events are instead indexed by atom–sector pairs, and their ordering retains microscopic initialization randomness in the vanishing-initialization limit. The resulting global trajectory is therefore combinatorial rather than purely spectral.

Microscopic randomness and universality in edge exposure. Despite the microscopic randomness, equal-amplitude sectors have a distribution-free large-system limit after reparametrization by the common activation-clock CDF. The normalized loss converges to the Random Edge curve ℓ_α, which depends only on the asymptotic atom-to-sector ratio α. The tensor order ν affects only the conversion from physical time to edge exposure through the activation-clock distribution. More broadly, we expect the same matching law in other factorized models with i.i.d. continuous atom–sector activation times and the same one-atom/one-sector exclusivity structure.

Beyond scalar sectors. Our analysis reveals a sharp distinction between complex and real modular addition. With complex weights, Fourier diagonalization reduces the problem exactly to the scalar-sector theory; with real weights, conjugate modes form two-dimensional sectors with richer internal geometry. We characterize their rank and border-rank cascade and conjecture that gradient flow follows these stages through successive atom activations, reorganization of earlier atoms, and potentially large parameter growth while the represented tensor remains bounded. This suggests that a broader theory should replace one clock per sector by a description of its internal cascade and its interleaving with other sectors.

Limitations and open directions. Our exact theory assumes diagonal scalar sectors, gradient flow, and squared tensor loss, and takes the vanishing-initialization limit before the large-system limit. This iterated limit gives a clean formulation while retaining the nontrivial macroscopic loss law, but it does not quantify how small σ must be as p grows. Obtaining such a condition would require estimates uniform in p, in particular control of increasingly close activation events. Moreover, exact agreement of every event order may be stronger than necessary for the large-system loss curve: finite-σ reorderings of near-tied events may disappear macroscopically. Determining the weakest scaling $\sigma = \sigma(p)$ sufficient for the same loss limit is an interesting open problem, along with discrete-time optimization and multidimensional-sector dynamics.

AI USE STATEMENT

In this work, we used generative AI tools to assist with research brainstorming, the formulation and refinement of mathematical arguments, mathematical derivations and proof development, the design and analysis of numerical experiments, literature search, and the drafting and revision of the manuscript. Generative AI was also used to assist with experimental and figure-generation code and to improve the clarity, organization, and presentation of the text.

AI-generated mathematical arguments were treated as provisional: the authors independently checked the derivations, assumptions, theorem statements, and proofs used in the final manuscript. Numerical results and AI-assisted code were likewise inspected and verified by the authors. References suggested by generative AI were checked against the original publications before inclusion. We reviewed all AI-assisted work and take full responsibility for the final content of this paper, including all mathematical claims, experimental results, citations, and conclusions.

REPRODUCIBILITY STATEMENT

The main theoretical results are stated with their assumptions in the main text, and complete proofs are provided in the appendices. For the numerical experiments, we will include with the submission the Python scripts used to generate all experimental results and figures in the paper. The scripts specify the model dimensions, initialization scales, random seeds, gradient-flow simulation parameters, and plotting procedures needed to reproduce the reported figures. All experimental targets are generated directly from their mathematical definitions, so no external datasets or pretrained models are required.

REFERENCES

Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M. Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, 15(80): 2773–2832, 2014. URL https://jmlr.org/papers/v15/anandkumar14b.html.

Nick Arnosti. Greedy matching in bipartite random graphs. *Stochastic Systems*, 12(2):133–150, 2022. doi: 10.1287/stsy.2021.0082.

Sanjeev Arora, Nadav Cohen, Wei Hu, and Yuping Luo. Implicit regularization in deep matrix factorization. In *Advances in Neural Information Processing Systems*, volume 32, pp. 7411–7422, 2019.

Raphaël Berthier. Incremental learning in diagonal linear networks. *Journal of Machine Learning Research*, 24(171):1–26, 2023. URL https://jmlr.org/papers/v24/22-1395.html.

Ada Boralevi, Jan Draisma, Emil Horobeţ, and Elina Robeva. Orthogonal and unitary tensor decomposition from an algebraic perspective. *Israel Journal of Mathematics*, 222(1):223–260, 2017. doi: 10.1007/s11856-017-1588-6.

Bilal Chughtai, Lawrence Chan, and Neel Nanda. A toy model of universality: Reverse engineering how networks learn group operations. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 6243–6267. PMLR, 2023.

Martin E. Dyer, Alan M. Frieze, and Boris Pittel. The average performance of the greedy matching algorithm. *The Annals of Applied Probability*, 3(2):526–552, 1993. doi: 10.1214/aoap/1177005436.

Rong Ge, Yunwei Ren, Xiang Wang, and Mo Zhou. Understanding deflation process in over-parametrized tensor decomposition. In *Advances in Neural Information Processing Systems*, volume 34, pp. 1299–1311, 2021.

Daniel Gissin, Shai Shalev-Shwartz, and Amit Daniely. The implicit bias of depth: How incremental learning drives generalization. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=H1lj0nNFwB.

Andrey Gromov. Grokking modular arithmetic. *arXiv preprint arXiv:2301.02679*, 2023.

Jianliang He, Leda Wang, Siyu Chen, and Zhuoran Yang. On the mechanism and dynamics of modular addition: Fourier features, lottery ticket, and grokking, 2026a. URL https://arxiv.org/abs/2602.16849.

Jianliang He, Leda Wang, Fengzhuo Zhang, Siyu Chen, and Zhuoran Yang. Neural networks provably learn spectral representations for group composition, 2026b. URL https://arxiv.org/abs/2606.02993.

Arthur Jacot, François Ged, Berfin Şimşek, Clément Hongler, and Franck Gabriel. Saddle-to-saddle dynamics in deep linear networks: Small initialization training, symmetry, and sparsity, 2021. URL https://arxiv.org/abs/2106.15933.

Jikai Jin, Zhiyuan Li, Kaifeng Lyu, Simon Shaolei Du, and Jason D. Lee. Understanding incremental learning of gradient descent: A fine-grained analysis of matrix sensing. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 15200–15238. PMLR, 2023. URL https://proceedings.mlr.press/v202/jin23a.html.

Daniel Kunin, Giovanni Luca Marchetti, Feng Chen, Dhruva Karkada, James B. Simon, Michael R. DeWeese, Surya Ganguli, and Nina Miolane. Alternating gradient flows: A theory of feature learning in two-layer neural networks. In *Advances in Neural Information Processing Systems*, volume 38, pp. 4951–4998, 2025.

Zhiyuan Li, Yuping Luo, and Kaifeng Lyu. Towards resolving the implicit bias of gradient descent for matrix factorization: Greedy low-rank learning. In *International Conference on Learning Representations*, 2021.

Giovanni Luca Marchetti, Daniel Kunin, Adele Myers, Francisco Acosta, and Nina Miolane. Sequential group composition: A window into the mechanics of deep learning. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2602.03655.

Mohamad Amin Mohamadi, Zhiyuan Li, Lei Wu, and Danica J. Sutherland. Why do you grok? a theoretical analysis on grokking modular addition. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 35934–35967. PMLR, 2024.

Depen Morwani, Benjamin L. Edelman, Costin-Andrei Oncescu, Rosie Zhao, and Sham M. Kakade. Feature emergence via margin maximization: Case studies in algebraic tasks. In *International Conference on Learning Representations*, 2024.

Neel Nanda, Lawrence Chan, Tom Lieberum, Jess Smith, and Jacob Steinhardt. Progress measures for grokking via mechanistic interpretability. In *International Conference on Learning Representations*, 2023.

Scott Pesme and Nicolas Flammarion. Saddle-to-saddle dynamics in diagonal linear networks. In *Advances in Neural Information Processing Systems*, volume 36, pp. 7475–7505, 2023.

Alethea Power, Yuri Burda, Harri Edwards, Igor Babuschkin, and Vedant Misra. Grokking: Generalization beyond overfitting on small algorithmic datasets. *arXiv preprint arXiv:2201.02177*, 2022.

Noam Razin, Asaf Maman, and Nadav Cohen. Implicit regularization in tensor factorization. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8913–8924. PMLR, 2021.

Noam Razin, Asaf Maman, and Nadav Cohen. Implicit regularization in hierarchical tensor factorization and deep convolutional neural networks. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 18422–18462. PMLR, 2022. URL https://proceedings.mlr.press/v162/razin22a.html.

Andrew M. Saxe, James L. McClelland, and Surya Ganguli. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. In *International Conference on Learning Representations*, 2014.

Xiang Wang, Chenwei Wu, Jason D. Lee, Tengyu Ma, and Rong Ge. Beyond lazy training for over-parameterized tensor decomposition. In *Advances in Neural Information Processing Systems*, volume 33, pp. 21934–21944, 2020.

Nicholas C. Wormald. Differential equations for random processes and random graphs. *The Annals of Applied Probability*, 5(4):1217–1235, 1995.

Nicholas C. Wormald. The differential equation method for random graph processes and greedy algorithms. In Michał Karoński and Hans Jürgen Prömel (eds.), *Lectures on Approximation and Randomized Algorithms*, pp. 73–155. PWN, Warsaw, 1999.

Ziqian Zhong, Ziming Liu, Max Tegmark, and Jacob Andreas. The clock and the pizza: Two stories in mechanistic explanation of neural networks. In *Advances in Neural Information Processing Systems*, volume 36, pp. 27223–27250, 2023.

CONTENTS

1 Introduction **1**

2 Setup and diagonal targets **2**

3 Vanishing initialization yields greedy matching **3**

4 Complex modular addition **5**

5 Large-system limit **5**

6 Beyond scalar sectors: real Fourier blocks **7**

7 Discussion **9**

References **10**

A1 Introduction: context and related work **15**

A1.1 Small initialization and incremental trajectories 15

A1.2 Greedy and deflation dynamics in tensor factorization 16

A1.3 Event-driven feature learning and structured operations 17

A1.4 Greedy matching and the large-system limit 18

A2 Setup and diagonal targets: technical preliminaries **18**

A2.1 Gradient flow with complex weights 18

A2.2 Exact GF equations and microscopic scaling 20

A2.3 The isolated scalar flow 21

A2.4 Distribution of the candidate clocks 23

A3 Vanishing initialization yields greedy matching: proof and details **25**

A3.1 Greedy event sequence and induction scheme 25

A3.2 Dynamics in a locked background 29

A3.3 Slow propagation between activation events 38

A3.4 Entrance into the next fast layer 42

A3.5 Fast capture and renewal of the induction hypotheses 50

A3.6 Completion of the proof of the greedy matching limit 63

A4 Complex modular addition: Fourier details **67**

A5 Large-system limit: proofs and asymptotics **68**

A5.1 Large-system limit of Random Edge matching 68

A5.2 From Random Edge matching to the equal-sector tensor limit 71

A5.3 Lower tail of the complex cubic activation clock 73

A5.4 Asymptotic inversion and physical training time 75

A6 Beyond scalar sectors: real modular addition details **78**

A6.1 From complex Fourier sectors to real Fourier blocks 78

A6.2 Geometry of the canonical two-dimensional block 79

A6.3 Rank and border-rank cascade 80

The appendix mirrors the organization of the main text: Appendix Aj provides proofs, technical details, and additional material for Section j.

A1 INTRODUCTION: CONTEXT AND RELATED WORK [Section 1]

This appendix places our results in the literature on small-initialization dynamics, incremental low-rank learning, tensor factorization, and structured algorithmic tasks. The closest precedents establish several ingredients of the picture studied here: small initialization can produce incremental or saddle-to-saddle learning; matrix and tensor factorizations can exhibit greedy low-rank behavior; and Fourier features can emerge sequentially in models trained on modular and group operations. Our contribution is not any of these phenomena in isolation. The distinctive point is that, for diagonal order-ν CP tensor learning with $\nu \geq 3$, we resolve the full vanishing-initialization trajectory into random microscopic activation events, identify their global competition exactly with greedy bipartite matching, and then solve the corresponding equal-sector large-system loss trajectory.

A1.1 SMALL INITIALIZATION AND INCREMENTAL TRAJECTORIES

Prior work. Factorized matrix and linear models have long provided tractable settings for studying optimization dynamics and implicit regularization. Exact solutions for deep linear networks already exhibit plateaus, rapid transitions, and sequential learning of spectral modes (Saxe et al., 2014). Related work on deep matrix factorization connected gradient-based optimization to implicit low-rank structure (Arora et al., 2019). Small initialization was subsequently connected more directly to incremental learning. Gissin et al. (2020) study the role of depth in producing incremental learning, while Li et al. (2021) interpret infinitesimal-initialization matrix factorization through *Greedy Low-Rank Learning*, in which optimization proceeds through solutions of increasing rank. A particularly close dynamical comparison is Jacot et al. (2021), who analyze deep linear networks near vanishing initialization. They rigorously characterize the first escape from the zero saddle along an effectively low-rank path and propose a subsequent saddle-to-saddle dynamics of increasing bottleneck rank. Importantly for our setting, they also identify the qualitative distinction between depth two and larger depth: for $L = 2$ the escape time is logarithmic in the inverse initialization scale, whereas for $L > 2$ the higher-order homogeneous degeneracy produces a polynomial scale of order $\alpha^{-(L-2)}$.

More recent work gives complete incremental trajectories in several degree-two settings. For diagonal linear regression, Berthier (2023) establish successive coordinate activation, and Pesme & Flammarion (2023) characterize the full vanishing-initialization limiting trajectory by an explicit recursive saddle-to-saddle algorithm, including the limiting jump times. In matrix sensing, Jin et al. (2023) rigorously track all successive rank-learning phases and show that gradient descent approaches solutions of increasing rank until the target rank is reached. Thus neither incremental learning nor a complete small-initialization trajectory is new in itself.

Our contribution and differences. The local higher-order escape mechanism in our tensor model is closely related to that of deep linear networks of depth greater than two, rather than being specific to tensors. For an order-ν CP model, the leading balanced scalar dynamics near the origin has the form $\dot{a} \propto a^{\nu-1}$, so activations occur on the scale $\sigma^{-(\nu-2)}$, directly analogous to the $\alpha^{-(L-2)}$ scale in depth-L linear networks (Jacot et al., 2021). The distinction therefore lies not in the higher-order homogeneity itself, but in how the resulting escape processes are organized globally.

Microscopic candidate clocks. For a diagonal tensor target, the near-origin dynamics decomposes into candidate atom–sector pairs. After rescaling time by $\sigma^{\nu-2}$, the normalized Gaussian initialization of each pair survives as a finite random activation clock. Hence the singular limit retains an entire family of competing random candidate times. This differs, for example, from the first-escape result of Jacot et al. (2021), where under a simple leading singular value the globally fastest escape direction is selected by the top singular mode, up to the internal symmetries of the factorization.

Competition becomes matching. The candidate events are not independent at the macroscopic level: once an atom learns a sector, both that atom and that sector become unavailable to later candidates. We prove that these two exclusivity constraints resolve the complete sequence of activations exactly as greedy bipartite matching. The limiting trajectory is therefore not only an ordered sequence of

ranks, coordinates, or saddles, but an explicit combinatorial process generated by the microscopic initialization.

From a finite event process to an explicit large-system loss trajectory. Finally, in the equal-sector regime the candidate clocks are i.i.d., and their ordering turns the matching process into Random Edge on the complete atom–sector bipartite graph. Taking the large-system limit then produces an explicit deterministic loss curve depending only on the asymptotic atom-to-sector ratio. Neither the matrix and diagonal-linear trajectory results above nor the deep-linear saddle analysis derive an analogous random matching representation together with a solvable large-system loss trajectory.

A1.2 Greedy and deflation dynamics in tensor factorization

Prior work. The diagonal targets in (3) belong to the standard orthogonally decomposable (*odeco*) class over $\mathbb{R}$ and to its unitary analogue (*udeco*) over $\mathbb{C}$; our setting allows the orthonormal factors to depend on the mode (Boralevi et al., 2017). The symmetric odeco case underlies classical tensor-power and designed deflation methods (Anandkumar et al., 2014). A closely related line of work studies how small initialization can induce incremental low-rank behavior in tensor factorization. Razin et al. (2021) analyze CP tensor factorization dynamically and show that small components evolve much more slowly than larger ones, leading to a greedy low tensor-rank bias. Under suitable conditions, they prove that the induced tensor trajectory remains close to a rank-one trajectory over prescribed finite time or distance scales, and support a broader component-by-component learning picture experimentally. Their later work extends this implicit low-rank perspective to hierarchical tensor factorizations and deep convolutional networks (Razin et al., 2022).

A complementary line interprets overparameterized tensor decomposition as an implicit deflation process. Wang et al. (2020) show that suitably designed gradient-based methods can exploit low-rank tensor structure beyond the lazy regime. More directly, Ge et al. (2021) study the symmetric odeco specialization with a symmetric CP parameterization. They show that a slightly modified gradient flow follows a deflation-like process and recovers the target components, providing a rigorous tensor analogue of greedy low-rank learning.

These works already establish the main qualitative phenomenon that motivates our analysis: small initialization can cause tensor models to learn low-rank structure incrementally rather than fitting all components simultaneously. We therefore do not claim novelty for greedy tensor learning, for sequential recovery of tensor components, or for deflation-like behavior by itself.

Our contribution and differences. Our result instead treats the nonsymmetric odeco/udeco setting with mode-dependent orthonormal factors and resolves the complete event structure of ordinary asymmetric CP gradient flow.

From a dominant component to all activation events. For each atom–sector pair, the microscopic initialization determines a candidate activation clock. We prove the singular-limit dynamics through every accepted activation, including the persistence of already learned pairs and the suppression of candidates made inadmissible by earlier events. Thus the theorem determines the entire finite-system sequence rather than only the emergence of an initial low-rank component or a qualitative deflation order.

Random ordering instead of amplitude ordering. The distinction is especially clear for equal-amplitude targets. In a classical deflation picture, a larger target component naturally has an advantage and may be learned first. In our equal-sector setting there is no such deterministic hierarchy: all target sectors have the same amplitude, and the order remains random in the limit because it is inherited from the microscopic Gaussian initialization. The diagonal structure makes this randomness explicit through the collection of activation clocks.

Atom–sector competition rather than rank growth alone. The model contains multiple atoms competing for multiple target sectors. Once an atom has captured one sector, that atom cannot independently capture another, while a learned sector is no longer available to other atoms. These two-sided resource constraints are not described by a scalar notion of tensor rank or by ordinary deflation. We prove that they organize the complete trajectory exactly as greedy bipartite matching.

A solvable many-sector limit. For equal-amplitude sectors, the matching representation can itself be analyzed as the number of atoms and sectors grows. After transforming the i.i.d. clocks to uniform

priorities, the finite-system process becomes Random Edge on a complete bipartite graph, and its large-system limit yields an explicit deterministic loss trajectory. To our knowledge, prior tensor factorization results do not derive such a random matching representation of the full gradient-flow trajectory or an explicit corresponding large-system loss law.

A1.3 Event-driven feature learning and structured operations

Prior work. A broader line of work studies small-initialization training as a sequence of discrete feature-learning events. Most directly, Kunin et al. (2025) introduce *Alternating Gradient Flows* (AGF), an effective description in which dormant neurons first compete for useful directions and activation is followed by a faster loss-minimization stage. This framework predicts successive activation times and loss drops across several models. For diagonal linear networks, the authors rigorously relate AGF to the vanishing-initialization limit of gradient flow, while in more general architectures AGF serves as an effective approximation to the observed staircase dynamics.

Related work studies representation-theoretic structure in learned group operations. Chughtai et al. (2023) characterize circuits based on irreducible representations for finite-group composition, while Marchetti et al. (2026) obtain an irrep-by-irrep activation picture within an AGF-based analysis. Other work shows that random initialization can drive individual neurons toward spectral representations of finite groups (He et al., 2026b). These results are particularly relevant to our discussion beyond scalar sectors, since non-Abelian irreducible representations naturally lead to multidimensional Fourier blocks.

Modular addition itself has become a standard setting for studying grokking, Fourier representations, and learning dynamics (Power et al., 2022; Nanda et al., 2023; Mohamadi et al., 2024). Gromov (2023) derive analytic feature maps and weight configurations solving modular-arithmetic tasks in simple two-layer networks and show that gradient-based training can recover such representations. Morwani et al. (2024) theoretically characterize the emergence of Fourier features in stylized models through margin maximization, while He et al. (2026a) show dynamically that Fourier frequencies compete within neurons and that the winning frequency depends on initialization-dependent spectral magnitude and phase. At the same time, Zhong et al. (2023) show that neural networks can learn qualitatively different algorithms for modular addition under changes of initialization and hyperparameters. Thus neither Fourier learning nor initialization-dependent competition is specific to our model, and the Fourier-sector mechanism need not describe arbitrary architectures.

Our contribution and differences. Our use of structured operations is narrower but more explicit: modular addition is represented directly as a target tensor, so its Fourier structure can be inserted into the exact CP gradient-flow analysis.

An exact reduction of ordinary gradient flow. The event-driven viewpoint itself is not new, and AGF provides a general framework for it. Our result instead proves, for the diagonal CP model, that the original gradient flow converges in the vanishing-initialization limit to a specific discrete process through all activation events. No auxiliary alternating dynamics is introduced: the activation clocks and the subsequent matching rule are derived directly from the GF equations.

Fourier sectors as an exactly diagonal tensor problem. For complex modular addition, Fourier transformation does more than reveal features learned by the network: it diagonalizes the target tensor exactly into equal-amplitude scalar rank-one sectors. The modular-addition result is therefore a direct corollary of the general diagonal-tensor theorem. In this representation, the competition among Fourier modes is not merely observed or inferred from trained weights; it is identified with the competition among explicitly defined atom–sector clocks.

From feature competition to a solvable large-system trajectory. Initialization-dependent Fourier competition appears in prior work, but in our setting the equal amplitudes make the microscopic ordering purely random and the atom–sector exclusivity turns it into Random Edge matching. This additional combinatorial structure allows us to pass from individual feature selection events to an explicit large-system loss trajectory. To our knowledge, prior work on modular or group operations does not derive an analogous explicit large-system loss law from a random matching representation of the training dynamics.

Complex versus real representations. The exact scalar-sector reduction also clarifies why the choice of field matters. Over $\mathbb{C}$, the Fourier characters of modular addition give independent one-dimensional sectors. Over $\mathbb{R}$, conjugate frequencies must be combined into two-dimensional blocks, leading to the multi-stage rank and border-rank cascades of Section 6. This provides a concrete example of the multidimensional internal sector dynamics that also arise naturally in representation-theoretic analyses of more general group operations.

A1.4 Greedy matching and the large-system limit

Prior work. Once the tensor dynamics has been reduced to a matching process, the remaining large-system problem connects to classical probabilistic combinatorics. The *Random Edge* algorithm processes graph edges in a uniformly random order and accepts an edge whenever both endpoints are still unmatched. Average-case behavior of random-order greedy matching was studied already by Dyer et al. (1993), and sparse bipartite asymptotics for Random Edge and related algorithms were derived by Arnosti (2022). A standard way to obtain such deterministic large-system trajectories is the differential-equation method for random processes developed by Wormald (1995; 1999).

Accordingly, we do not claim novelty for Random Edge itself, for random-order greedy matching, or for the sparse bipartite matching asymptotic used in Section 5. Proposition 1 is stated and proved in the normalization needed for our application, but its closed form agrees with the known sparse bipartite Random Edge formula of Arnosti (2022).

Our contribution and differences. The novel connection runs in the opposite direction: we show that a singular limit of nonlinear tensor gradient flow generates this random combinatorial process.

From activation clocks to Random Edge. For equal-amplitude tensor sectors, the candidate atom–sector clocks are i.i.d. Applying their common CDF converts them into independent uniform edge priorities without changing their ordering. The full vanishing-initialization tensor trajectory is therefore exactly Random Edge on the complete atom–sector bipartite graph, observed as the priority threshold increases. Thus the random matching process is not introduced as an approximation or external algorithm; it is derived from the microscopic GF dynamics.

An explicit large-system loss trajectory. When $H_p/p \to \alpha$, the fraction of unmatched sectors converges to a deterministic function $\ell_\alpha(s)$ of the edge exposure parameter. Consequently, the normalized tensor loss has an explicit limiting trajectory that depends only on the asymptotic atom-to-sector ratio. The tensor order and the microscopic clock distribution disappear from the matching law itself and enter only through the conversion between physical training time and edge exposure.

Explicit physical time for modular addition. For complex modular addition, we go one step further. The Fourier sectors are equal-amplitude and the scalar clocks are cubic. By deriving the lower tail of this clock distribution and asymptotically inverting the exposure map, we obtain an explicit physical learning-time scale in addition to the universal matching trajectory.

Novelty boundary. To our knowledge, previous work does not characterize a vanishing-initialization CP tensor gradient-flow trajectory as a random greedy matching process, nor derive from such a reduction an explicit large-system loss law. Our contribution is therefore the chain

$$\text{CP GF} \longrightarrow \text{random activation clocks} \longrightarrow \text{greedy matching} \longrightarrow \text{large-system loss},$$

rather than any of these ingredients individually.

A2 Setup and diagonal targets: technical preliminaries [Section 2]

This appendix provides the technical facts underlying the setup of Section 2. We first specify the GF convention for complex parameters, then derive the microscopic rescaling and the decoupled scalar flows, and finally establish the properties of the activation clocks used in Section 3.

A2.1 Gradient flow with complex weights

Real Euclidean convention. When $\mathbb{K} = \mathbb{C}$, the loss is real valued but the learned weights are complex. We define GF by regarding a complex vector as a real vector of twice the dimension.

Namely, if $\mathbf{u} = \mathbf{x} + i\mathbf{y} \in \mathbb{C}^p$, with $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$, we identify

$$\mathbb{C}^p \cong \mathbb{R}^{2p}, \qquad \mathbf{u} = \mathbf{x} + i\mathbf{y} \longleftrightarrow (\mathbf{x}, \mathbf{y}).$$

For $\mathbf{u}, \mathbf{v} \in \mathbb{C}^p$, let

$$\langle \mathbf{u}, \mathbf{v} \rangle_{\mathbb{C}} := \mathbf{u}^* \mathbf{v} = \sum_{j=1}^{p} \overline{u_j} v_j,$$

where $\mathbf{u}^*$ denotes the conjugate transpose of $\mathbf{u}$. The real Euclidean inner product under the above identification is therefore

$$\langle \mathbf{u}, \mathbf{v} \rangle_{\mathbb{R}} := \mathrm{Re} \langle \mathbf{u}, \mathbf{v} \rangle_{\mathbb{C}}.$$

Equivalently, if $u_j = x_j + iy_j$ and $v_j = r_j + is_j$, then

$$\langle \mathbf{u}, \mathbf{v} \rangle_{\mathbb{R}} = \sum_{j=1}^{p} (x_j r_j + y_j s_j).$$

Let $f : \mathbb{C}^p \to \mathbb{R}$ be differentiable when regarded as a function of the $2p$ real variables $(\mathbf{x}, \mathbf{y})$. For a direction $\mathbf{h} \in \mathbb{C}^p$, its directional derivative is

$$Df(\mathbf{u})[\mathbf{h}] := \frac{d}{d\varepsilon} f(\mathbf{u} + \varepsilon \mathbf{h}) \Big|_{\varepsilon = 0}, \qquad \varepsilon \in \mathbb{R}.$$

The gradient $\nabla_{\mathbf{u}} f \in \mathbb{C}^p$ is the unique vector satisfying

$$Df(\mathbf{u})[\mathbf{h}] = \mathrm{Re} \langle \nabla_{\mathbf{u}} f, \mathbf{h} \rangle_{\mathbb{C}} \qquad \text{for every } \mathbf{h} \in \mathbb{C}^p. \tag{25}$$

Writing $u_j = x_j + iy_j$, this is simply

$$(\nabla_{\mathbf{u}} f)_j = \frac{\partial f}{\partial x_j} + i \frac{\partial f}{\partial y_j}.$$

Hence

$$\frac{d\mathbf{u}}{dt} = -\nabla_{\mathbf{u}} f$$

is exactly the ordinary real GF for $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{2p}$, written in complex notation.

For comparison with the commonly used Wirtinger notation, define

$$\frac{\partial}{\partial u_j} := \frac{1}{2} \left(\frac{\partial}{\partial x_j} - i \frac{\partial}{\partial y_j} \right), \qquad \frac{\partial}{\partial \overline{u_j}} := \frac{1}{2} \left(\frac{\partial}{\partial x_j} + i \frac{\partial}{\partial y_j} \right).$$

Then, for a real-valued function f,

$$\nabla_{\mathbf{u}} f = 2 \frac{\partial f}{\partial \overline{\mathbf{u}}}, \qquad \frac{\partial f}{\partial \overline{\mathbf{u}}} := \left(\frac{\partial f}{\partial \overline{u_j}} \right)_{j=1}^{p}. \tag{26}$$

We use the real-gradient definition (25) throughout; no additional complex optimization convention is required.

Basic identities. We record a few elementary consequences that will be used below. First, for $f(\mathbf{u}) = \frac{1}{2} \|\mathbf{u} - \mathbf{v}\|_2^2$, where $\mathbf{v} \in \mathbb{C}^p$ is fixed,

$$Df(\mathbf{u})[\mathbf{h}] = \mathrm{Re} \langle \mathbf{u} - \mathbf{v}, \mathbf{h} \rangle_{\mathbb{C}}, \qquad \nabla_{\mathbf{u}} f = \mathbf{u} - \mathbf{v}.$$

Thus the familiar gradient of a squared Euclidean distance is unchanged when the variables are complex, provided the gradient is understood in the real sense above.

The same convention applies to tensors. For complex tensors A, B of the same shape, we use the Hermitian Frobenius inner product

$$\langle A, B \rangle_F := \sum_{i_1, \ldots, i_\nu} \overline{A_{i_1 \cdots i_\nu}} B_{i_1 \cdots i_\nu}, \qquad \|A\|_F^2 = \langle A, A \rangle_F.$$

Consequently, if $R(\varepsilon)$ is a differentiable tensor-valued curve, then

$$\frac{d}{d\varepsilon}\frac{1}{2}\|R(\varepsilon)\|_F^2\Big|_{\varepsilon=0} = \mathrm{Re}\left\langle R(0),\, \frac{dR}{d\varepsilon}\Big|_{\varepsilon=0}\right\rangle_F. \tag{27}$$

We will also repeatedly differentiate a rank-one tensor with respect to one of its factors. If

$$\Phi(\mathbf{u}^{(1)},\dots,\mathbf{u}^{(\nu)}) = \mathbf{u}^{(1)}\otimes\cdots\otimes\mathbf{u}^{(\nu)},$$

then perturbing only the mode-a factor gives

$$D_{\mathbf{u}^{(a)}}\Phi[\mathbf{h}] = \mathbf{u}^{(1)}\otimes\cdots\otimes\mathbf{u}^{(a-1)}\otimes\mathbf{h}\otimes\mathbf{u}^{(a+1)}\otimes\cdots\otimes\mathbf{u}^{(\nu)}. \tag{28}$$

Finally, for an order-ν tensor $A\in(\mathbb{C}^p)^{\otimes\nu}$ and vectors $\mathbf{w}^{(b)}\in\mathbb{C}^p$, define the mode-$a$ contraction

$$A\times_{b\neq a}\overline{\mathbf{w}^{(b)}}\in\mathbb{C}^p$$

by

$$\Big[A\times_{b\neq a}\overline{\mathbf{w}^{(b)}}\Big]_{i_a} := \sum_{\{i_b:\, b\neq a\}} A_{i_1\cdots i_\nu}\prod_{b\neq a}\overline{w^{(b)}_{i_b}}. \tag{29}$$

With this definition,

$$\left\langle A,\, \mathbf{w}^{(1)}\otimes\cdots\otimes\mathbf{w}^{(a-1)}\otimes\mathbf{h}\otimes\mathbf{w}^{(a+1)}\otimes\cdots\otimes\mathbf{w}^{(\nu)}\right\rangle_F = \left\langle A\times_{b\neq a}\overline{\mathbf{w}^{(b)}},\mathbf{h}\right\rangle_{\mathbb{C}}. \tag{30}$$

Equations (27), (28), and (30) are the only complex-calculus identities needed to derive the CP GF equations below.

A2.2 Exact GF equations and microscopic scaling

Using the contraction defined in (29), differentiating the loss in (2) gives the exact GF equation

$$\frac{d\mathbf{u}_k^{(a)}}{dt} = T\times_{b\neq a}\overline{\mathbf{u}_k^{(b)}} - \sum_{j=1}^{H}\mathbf{u}_j^{(a)}\prod_{b\neq a}\left\langle\mathbf{u}_k^{(b)},\mathbf{u}_j^{(b)}\right\rangle_{\mathbb{K}}. \tag{31}$$

Here the first term is the contraction of the target with the other $\nu-1$ factors of atom k, while the second is the corresponding contraction of the current model output. We use $\langle\cdot,\cdot\rangle_{\mathbb{K}}$ for the Euclidean inner product over $\mathbb{R}$ and the Hermitian inner product over $\mathbb{C}$.

Set

$$\mathbf{u}_k^{(a)}(t) = \sigma\mathbf{v}_k^{(a)}(\theta),\qquad \theta=\sigma^{\nu-2}t.$$

Substituting into (31) and dividing by $\sigma^{\nu-1}$ yields

$$\frac{d\mathbf{v}_k^{(a)}}{d\theta} = T\times_{b\neq a}\overline{\mathbf{v}_k^{(b)}} - \sigma^{\nu}\sum_{j=1}^{H}\mathbf{v}_j^{(a)}\prod_{b\neq a}\left\langle\mathbf{v}_k^{(b)},\mathbf{v}_j^{(b)}\right\rangle_{\mathbb{K}}. \tag{32}$$

This explains the slow-time scaling used in the main text: while the rescaled factors remain $O(1)$, the target-driven dynamics is $O(1)$ in θ, whereas interaction through the current model output is suppressed by σ^{ν}.

For the diagonal target (3), define the rescaled sector coordinates

$$z_{ki}^{(a)}(\theta) = \left\langle\mathbf{e}_i^{(a)},\mathbf{v}_k^{(a)}(\theta)\right\rangle_{\mathbb{K}}.$$

Projecting (32) onto $\mathbf{e}_i^{(a)}$ gives the exact coordinate equation

$$\frac{dz_{ki}^{(a)}}{d\theta} = \lambda_i\prod_{b\neq a}\overline{z_{ki}^{(b)}} - \sigma^{\nu}\sum_{j=1}^{H}z_{ji}^{(a)}\prod_{b\neq a}\left\langle\mathbf{v}_k^{(b)},\mathbf{v}_j^{(b)}\right\rangle_{\mathbb{K}}. \tag{33}$$

Thus, before any rescaled factor becomes large, the second term vanishes as $\sigma\to0$, and different atom–sector pairs evolve independently to leading order.

More precisely, let $\mathbf{z}(s;\mathbf{g})$ denote the isolated flow (4) and define

$$\widetilde{z}_{ki}^{(a)}(\theta) = z^{(a)}(\lambda_i\theta;\mathbf{g}_{ki}), \tag{34}$$

$$\tau_{\min} := \min_{k,i}\frac{\tau_\nu(\mathbf{g}_{ki})}{\lambda_i}. \tag{35}$$

Proposition 3 (Microscopic decoupling)**.** *Fix p, H, ν and a realization of the initialization for which $\tau_{\min} > 0$. For every $\Theta < \tau_{\min}$,*

$$\max_{k,i,a} \sup_{0\le\theta\le\Theta} \left| z_{ki}^{(a)}(\theta) - \widetilde{z}_{ki}^{(a)}(\theta) \right| \longrightarrow 0 \qquad \text{as } \sigma \to 0. \tag{36}$$

Proof. On $[0, \Theta]$, all isolated trajectories $\widetilde{\mathbf{z}}_{ki}$ remain bounded because $\Theta < \tau_{\min}$. Since the mode-wise bases are complete, define

$$\widetilde{\mathbf{v}}_k^{(a)}(\theta) = \sum_{i=1}^{p} \widetilde{z}_{ki}^{(a)}(\theta)\mathbf{e}_i^{(a)}.$$

These vectors remain bounded on $[0, \Theta]$ and satisfy the $\sigma = 0$ version of (32).

Collect all rescaled factors into $\mathbf{v}^\sigma$ and the corresponding limiting factors into $\widetilde{\mathbf{v}}$. The right-hand side of (32) has the form

$$\mathcal{F}(\mathbf{v}) + \sigma^\nu \mathcal{G}(\mathbf{v}),$$

where $\mathcal{F}$ and $\mathcal{G}$ are polynomial vector fields. On a fixed bounded neighborhood of the limiting trajectory, $\mathcal{F}$ is Lipschitz and $\mathcal{G}$ is uniformly bounded. Standard continuous dependence of ODE solutions, or directly Gronwall's inequality, therefore gives

$$\sup_{0\le\theta\le\Theta} \|\mathbf{v}^\sigma(\theta) - \widetilde{\mathbf{v}}(\theta)\| = O(\sigma^\nu).$$

Projecting onto the diagonalizing basis vectors proves (36). □

Proposition 3 shows directly why the candidate clock associated with (k, i) is $\tau_\nu(\mathbf{g}_{ki})/\lambda_i$: after the change of variable $s = \lambda_i\theta$, its limiting microscopic dynamics is precisely the unit-amplitude flow (4). The proposition applies only strictly before the first activation. Through and after an activation, the rescaled factors become large enough that the σ^ν interaction term can no longer be neglected; this regime is analyzed in Appendix A3.

Atom-wise balance invariant. The full interacting CP flow has an exact balance invariant that will be used later. For every atom k and every pair of modes $a, b \in [\nu]$,

$$\left\|\mathbf{u}_k^{(a)}(t)\right\|^2 - \left\|\mathbf{u}_k^{(b)}(t)\right\|^2 = \left\|\mathbf{u}_k^{(a)}(0)\right\|^2 - \left\|\mathbf{u}_k^{(b)}(0)\right\|^2. \tag{37}$$

Indeed, the exact GF equation gives

$$\frac{d}{dt}\left\|\mathbf{u}_k^{(a)}\right\|^2 = 2\,\mathrm{Re}\left\langle T - \widehat{T}, \bigotimes_{c=1}^{\nu} \mathbf{u}_k^{(c)} \right\rangle_F, \tag{38}$$

whose right-hand side is independent of a. Subtracting the equations for two modes proves (37). Under the Gaussian initialization,

$$\left\|\mathbf{u}_k^{(a)}(t)\right\|^2 - \left\|\mathbf{u}_k^{(b)}(t)\right\|^2 = \sigma^2\left(\left\|\mathbf{g}_k^{(a)}\right\|^2 - \left\|\mathbf{g}_k^{(b)}\right\|^2\right). \tag{39}$$

A2.3 The isolated scalar flow

We next record the properties of the isolated flow

$$\frac{dz^{(a)}}{ds} = \prod_{b\neq a} \overline{z^{(b)}}, \qquad z^{(a)}(0) = g^{(a)}, \tag{40}$$

that are needed in the main argument. Let $\mathbf{z}(s) = (z^{(1)}(s), \dots, z^{(\nu)}(s))$, $\mathbf{g} = (g^{(1)}, \dots, g^{(\nu)})$, and

$$P(s) = \prod_{a=1}^{\nu} z^{(a)}(s).$$

A direct differentiation gives, for every a,

$$\frac{d}{ds}|z^{(a)}|^2 = 2\,\mathrm{Re}\,P(s), \qquad \frac{dP}{ds} = \sum_{a=1}^{\nu}\prod_{b\neq a}|z^{(b)}|^2. \tag{41}$$

In particular,

$$|z^{(a)}(s)|^2 - |z^{(b)}(s)|^2 = |g^{(a)}|^2 - |g^{(b)}|^2 \tag{42}$$

for every a, b, while $\mathrm{Im}\,P(s)$ is constant and $\mathrm{Re}\,P(s)$ is nondecreasing.

These identities reduce the evolution of the magnitudes to one scalar variable. Write

$$\rho_a = |g^{(a)}|^2, \qquad P(0) = A_0 + iB.$$

By (42), there is a scalar function $q(s)$ such that

$$|z^{(a)}(s)|^2 = \rho_a + q(s) \qquad \text{for all } a, \tag{43}$$

with $q(0) = 0$. Since $q' = 2\,\mathrm{Re}\,P$ and $|P|^2 = \prod_a(\rho_a + q)$,

$$\big(q'(s)\big)^2 = 4D_{\mathbf{g}}(q(s)), \qquad D_{\mathbf{g}}(q) := \prod_{a=1}^{\nu}(\rho_a + q) - B^2. \tag{44}$$

For generic initial data, this gives an explicit integral representation of the activation clock. Let $\rho_{\min} = \min_a \rho_a$. If $A_0 > 0$, then q increases from zero to infinity and

$$\tau_\nu(\mathbf{g}) = \frac{1}{2}\int_0^\infty \frac{dq}{\sqrt{D_{\mathbf{g}}(q)}}. \tag{45}$$

If $A_0 < 0$, then q first decreases to the unique turning point $q_* \in [-\rho_{\min}, 0)$ satisfying $D_{\mathbf{g}}(q_*) = 0$, and then increases to infinity, so

$$\tau_\nu(\mathbf{g}) = \frac{1}{2}\int_{q_*}^0 \frac{dq}{\sqrt{D_{\mathbf{g}}(q)}} + \frac{1}{2}\int_{q_*}^\infty \frac{dq}{\sqrt{D_{\mathbf{g}}(q)}}. \tag{46}$$

For complex Gaussian initialization, $\mathrm{Im}\,P(0) \neq 0$ almost surely, so $q_* > -\rho_{\min}$ in the second case. For real Gaussian initialization, $B = 0$ and, almost surely, the minimum $\rho_{\min}$ is attained at a unique mode, giving $q_* = -\rho_{\min}$.

Proposition 4 (Properties of the scalar clock)**.** *Let $\nu \geq 3$ and $\mathbf{g} \sim \mathcal{N}_{\mathbb{K}}(0, I_\nu)$. Then*

$$0 < \tau_\nu(\mathbf{g}) < \infty \qquad \textit{almost surely}.$$

Moreover, for every $c > 0$,

$$\tau_\nu(c\mathbf{g}) = c^{-(\nu-2)}\tau_\nu(\mathbf{g}). \tag{47}$$

Proof. Gaussian initialization almost surely has nonzero coordinates and pairwise distinct values of $|g^{(a)}|^2$. In the complex case, $A_0 = 0$ is also a null event. Hence the integral representations (45)–(46) apply almost surely.

The only possible divergence relevant to forward blow-up occurs as $q \to \infty$. Since

$$D_{\mathbf{g}}(q) = q^\nu(1 + o(1)),$$

the integrand is asymptotic to $q^{-\nu/2}$. Therefore

$$\int^\infty q^{-\nu/2}\,dq < \infty$$

precisely when $\nu > 2$. The turning-point integral in (46) is also finite: almost surely the root q_* is simple, so $D_{\mathbf{g}}(q) \asymp q - q_*$ near q_*. Thus $q(s)$ diverges to infinity at a finite forward time, and hence $\|\mathbf{z}(s)\| \to \infty$ at the same time. Local existence gives $\tau_\nu(\mathbf{g}) > 0$.

For homogeneity, if $\mathbf{z}(s; \mathbf{g})$ solves (40), then

$$\widetilde{\mathbf{z}}(s) = c\,\mathbf{z}(c^{\nu-2}s; \mathbf{g})$$

solves the same equation with initial condition $c\mathbf{g}$. Its blow-up time is therefore $c^{-(\nu-2)}\tau_\nu(\mathbf{g})$. □

Why $\nu \geq 3$ is different from matrix factorization. For $\nu = 2$, the isolated system is linear: $\dot{z}^{(1)} = \overline{z^{(2)}}$ and $\dot{z}^{(2)} = \overline{z^{(1)}}$. Equivalently, the large-q integrand in (45) behaves as q^{-1}, whose integral diverges. Thus the isolated matrix dynamics grows exponentially rather than blowing up at a finite rescaled time, and escape from a small initialization occurs on a logarithmic scale. The finite random activation clock underlying our tensor reduction is therefore specific to $\nu \geq 3$.

A2.4 Distribution of the candidate clocks

It remains to justify the probabilistic properties stated after (5). By Gaussian isotropy and independence of the initial weights, the vectors

$$\mathbf{g}_{ki} = \big(g_{ki}^{(1)}, \dots, g_{ki}^{(\nu)}\big), \qquad k \in [H], \quad i \in [p],$$

are independent standard Gaussian vectors in $\mathbb{K}^\nu$.

Proposition 5 (Distribution of the scalar and candidate clocks)**.** *For $\nu \geq 3$, the CDF*

$$F_\nu(x) := \mathbb{P}(\tau_\nu(\mathbf{g}) \leq x)\,, \qquad \mathbf{g} \sim \mathcal{N}_{\mathbb{K}}(0, I_\nu),$$

is a continuous strictly increasing bijection

$$F_\nu : [0, \infty) \longrightarrow [0, 1).$$

Consequently, the candidate clocks

$$\tau_{ki} = \frac{\tau_\nu(\mathbf{g}_{ki})}{\lambda_i}$$

are independent, finite, and almost surely pairwise distinct. If all λ_i are equal, they are i.i.d.

Proof. Write the Gaussian vector in radial form

$$\mathbf{g} = R\boldsymbol{\omega}, \qquad R = \|\mathbf{g}\|, \qquad \boldsymbol{\omega} = \frac{\mathbf{g}}{\|\mathbf{g}\|}.$$

For both real and complex standard Gaussian vectors, R has a density that is strictly positive on $(0, \infty)$, and R is independent of $\boldsymbol{\omega}$.

By Proposition 4,

$$0 < \tau_\nu(\mathbf{g}) < \infty \qquad \text{almost surely,}$$

and the clock homogeneity (47) gives

$$\tau_\nu(\mathbf{g}) = R^{-(\nu-2)}\tau_\nu(\boldsymbol{\omega}). \tag{48}$$

Since $R \in (0, \infty)$ almost surely, it follows that

$$0 < \tau_\nu(\boldsymbol{\omega}) < \infty \qquad \text{for almost every } \boldsymbol{\omega}. \tag{49}$$

We now establish the claimed properties of the CDF

$$F_\nu(x) = \mathbb{P}(\tau_\nu(\mathbf{g}) \leq x)\,.$$

Step 1: atomlessness and continuity. Fix $x \in (0, \infty)$. Conditional on any direction $\boldsymbol{\omega}$ satisfying (49), the radial representation (48) gives

$$\tau_\nu(\mathbf{g}) = R^{-(\nu-2)}\tau_\nu(\boldsymbol{\omega}).$$

As a function of $R \in (0, \infty)$, the right-hand side is continuous and strictly decreasing. Hence

$$\tau_\nu(\mathbf{g}) = x$$

can occur only at the single radial value

$$R = \left(\frac{\tau_\nu(\boldsymbol{\omega})}{x}\right)^{1/(\nu-2)}.$$

Because R has an absolutely continuous distribution,

$$\mathbb{P}(\tau_\nu(\mathbf{g}) = x \,|\, \boldsymbol{\omega}) = 0$$

for almost every $\boldsymbol{\omega}$. Averaging over $\boldsymbol{\omega}$ gives

$$\mathbb{P}(\tau_\nu(\mathbf{g}) = x) = 0. \tag{50}$$

Also, $\mathbb{P}(\tau_\nu(\mathbf{g}) = 0) = 0$ because the clock is strictly positive almost surely. Thus the distribution of $\tau_\nu(\mathbf{g})$ is atomless, and consequently its CDF F_ν is continuous on $[0, \infty)$.

Step 2: strict monotonicity. Let

$$0 \le x < y < \infty.$$

We show that

$$F_\nu(y) > F_\nu(x).$$

First suppose $x > 0$. Conditional on a direction $\boldsymbol{\omega}$ satisfying (49), the event

$$x < \tau_\nu(\mathbf{g}) \le y$$

is, by the radial representation (48), equivalent to

$$\left(\frac{\tau_\nu(\boldsymbol{\omega})}{y}\right)^{1/(\nu-2)} \le R < \left(\frac{\tau_\nu(\boldsymbol{\omega})}{x}\right)^{1/(\nu-2)}. \tag{51}$$

Because $x < y$ and $\tau_\nu(\boldsymbol{\omega}) \in (0,\infty)$, this is a nonempty interval of positive length contained in $(0,\infty)$. Since the density of R is strictly positive on $(0,\infty)$,

$$\mathbb{P}(x < \tau_\nu(\mathbf{g}) \le y \,|\, \boldsymbol{\omega}) > 0 \tag{52}$$

for almost every $\boldsymbol{\omega}$. Averaging over the direction gives

$$F_\nu(y) - F_\nu(x) = \mathbb{P}(x < \tau_\nu(\mathbf{g}) \le y) > 0.$$

If $x = 0$, then for almost every admissible direction,

$$\tau_\nu(\mathbf{g}) \le y \quad \Longleftrightarrow \quad R \ge \left(\frac{\tau_\nu(\boldsymbol{\omega})}{y}\right)^{1/(\nu-2)}.$$

The right-hand event has strictly positive conditional probability because the radial Gaussian distribution has positive mass above every finite threshold. Hence

$$F_\nu(y) > F_\nu(0).$$

Since the clock is strictly positive almost surely,

$$F_\nu(0) = 0. \tag{53}$$

Thus F_ν is strictly increasing on $[0,\infty)$.

Step 3: the range of F_ν. Since $\tau_\nu(\mathbf{g}) < \infty$ almost surely,

$$F_\nu(x) \longrightarrow 1 \qquad \text{as } x \to \infty. \tag{54}$$

On the other hand, $F_\nu(x) < 1$ for every finite x. Indeed, for any $x > 0$ and almost every direction satisfying (49),

$$\tau_\nu(\mathbf{g}) > x$$

whenever

$$0 < R < \left(\frac{\tau_\nu(\boldsymbol{\omega})}{x}\right)^{1/(\nu-2)}. \tag{55}$$

This radial interval has positive probability because the density of R is strictly positive on $(0,\infty)$. Hence

$$\mathbb{P}(\tau_\nu(\mathbf{g}) > x) > 0,$$

and therefore

$$F_\nu(x) < 1 \qquad \text{for every } x < \infty. \tag{56}$$

Combining (53), the continuity and strict monotonicity established above, (54), and (56), we conclude that

$$F_\nu : [0,\infty) \longrightarrow [0,1) \tag{57}$$

is a continuous strictly increasing bijection. In particular, its ordinary inverse

$$F_\nu^{-1} : [0,1) \to [0,\infty)$$

is well defined.

Step 4: candidate clocks. By Gaussian isotropy and independence of the initialization, the vectors

$$\mathbf{g}_{ki}, \qquad k \in [H], \quad i \in [p],$$

are independent standard Gaussian vectors in $\mathbb{K}^\nu$. Hence the scalar clocks $\tau_\nu(\mathbf{g}_{ki})$ are independent. Since $\lambda_i > 0$ is deterministic,

$$\tau_{ki} = \frac{\tau_\nu(\mathbf{g}_{ki})}{\lambda_i}$$

are also independent and finite almost surely.

Each τ_{ki} has an atomless distribution, because multiplication by the positive deterministic factor λ_i^{-1} preserves atomlessness. Therefore, for any two distinct pairs $(k,i) \neq (\ell,j)$,

$$\mathbb{P}(\tau_{ki} = \tau_{\ell j}) = 0.$$

Since there are only finitely many candidate pairs, a union bound gives that all Hp candidate clocks are almost surely pairwise distinct.

Finally, if all sector amplitudes λ_i are equal, the deterministic rescaling is the same for every pair. Since the vectors $\mathbf{g}_{ki}$ are i.i.d., the candidate clocks τ_{ki} are then i.i.d. as well. □

Given a sector i with amplitude λ_i, the clock of an edge incident to this sector has CDF $F_\nu(\lambda_i\theta)$. For equal-amplitude sectors, all edges therefore carry i.i.d. continuous clocks. This is the property that will later connect the tensor dynamics to Random Edge matching.

A3 Vanishing initialization yields greedy matching: proof and details [Section 3]

We now prove Theorem 1. Throughout this appendix, we fix a realization of the Gaussian initialization belonging to the probability-one event on which all candidate clocks are finite and pairwise distinct, as established in Appendix A2. All constants below may depend on this fixed realization, as well as on p, H, ν and the target coefficients λ_i, but not on σ.

A3.1 Greedy event sequence and induction scheme

Exact and limiting dynamical variables. For each $\sigma > 0$, let $U^\sigma(t) = (\mathbf{u}_k^{\sigma,(a)}(t))_{a\in[\nu],\, k\in[H]}$ denote the physical-time GF solution of (2). Since the proof is carried out in the slow time $\theta = \sigma^{\nu-2}t$, we write

$$\mathbf{u}_{k,\sigma}^{(a)}(\theta) := \mathbf{u}_k^{\sigma,(a)}\big(\sigma^{-(\nu-2)}\theta\big). \tag{58}$$

Thus $\mathbf{u}_{k,\sigma}^{(a)}(\theta) \in \mathbb{K}^p$ is the exact learned weight vector, with only the time variable reparametrized.

Its exact coordinate in target sector i is

$$x_{ki,\sigma}^{(a)}(\theta) := \big\langle \mathbf{e}_i^{(a)}, \mathbf{u}_{k,\sigma}^{(a)}(\theta)\big\rangle. \tag{59}$$

While this coordinate is microscopic, it is natural to divide by the initialization scale and define

$$y_{ki,\sigma}^{(a)}(\theta) := \frac{x_{ki,\sigma}^{(a)}(\theta)}{\sigma}. \tag{60}$$

The quantities $\mathbf{u}_{k,\sigma}^{(a)}$, $x_{ki,\sigma}^{(a)}$, and $y_{ki,\sigma}^{(a)}$ are all exact σ-dependent GF quantities.

We reserve the letter z for the σ-independent isolated scalar flow. For $\mathbf{g} = (g^{(1)}, \dots, g^{(\nu)}) \in \mathbb{K}^\nu$, write

$$\mathbf{z}(s;\mathbf{g}) = \big(z^{(1)}(s;\mathbf{g}), \dots, z^{(\nu)}(s;\mathbf{g})\big)$$

for the solution of

$$\frac{dz^{(a)}}{ds} = \prod_{b\neq a} \overline{z^{(b)}}, \qquad z^{(a)}(0;\mathbf{g}) = g^{(a)}, \qquad 0 \le s < \tau_\nu(\mathbf{g}). \tag{61}$$

For a candidate edge (k,i), let

$$\mathbf{g}_{ki} = \big(g_{ki}^{(1)}, \dots, g_{ki}^{(\nu)}\big) \in \mathbb{K}^{\nu}$$

and define its isolated profile in the global slow time by

$$\boldsymbol{\zeta}_{ki}(\theta) := \mathbf{z}(\lambda_i \theta; \mathbf{g}_{ki}), \qquad 0 \le \theta < \tau_{ki}, \qquad \tau_{ki} = \frac{\tau_\nu(\mathbf{g}_{ki})}{\lambda_i}. \tag{62}$$

Thus $\boldsymbol{\zeta}_{ki}$ is independent of σ. Before the first activation, Proposition 3 states precisely that the exact microscopic coordinates $y_{ki,\sigma}^{(a)}$ converge to the corresponding components $\zeta_{ki}^{(a)}$ of these isolated profiles.

Greedy event sequence. Let

$$(k_1, i_1), \dots, (k_M, i_M), \qquad M = \min\{H, p\}, \tag{63}$$

be the edges accepted by the greedy clock process, listed in their order of acceptance. Denote their clock times by

$$\theta_m^* := \tau_{k_m i_m}, \qquad 0 < \theta_1^* < \cdots < \theta_M^*, \qquad \theta_0^* := 0, \qquad \theta_{M+1}^* := +\infty. \tag{64}$$

After m accepted events, define the claimed atoms and sectors by

$$\mathcal{A}_m := \{k_1, \dots, k_m\}, \qquad \mathcal{S}_m := \{i_1, \dots, i_m\}, \qquad \mathcal{A}_0 = \mathcal{S}_0 = \varnothing, \tag{65}$$

and the candidate edges whose two endpoints are still available by

$$\mathcal{E}_m := \big([H] \setminus \mathcal{A}_m\big) \times \big([p] \setminus \mathcal{S}_m\big). \tag{66}$$

By the definition of greedy matching,

$$(k_{m+1}, i_{m+1}) = \operatorname*{arg\,min}_{(k,i)\in\mathcal{E}_m} \tau_{ki}, \qquad \theta_{m+1}^* = \min_{(k,i)\in\mathcal{E}_m} \tau_{ki}, \qquad 0 \le m < M. \tag{67}$$

The minimizer is unique because all candidate clocks are distinct.

A useful consequence of the greedy construction is

$$\tau_{ki} > \theta_m^* \qquad \text{for every } (k,i) \in \mathcal{E}_m. \tag{68}$$

Indeed, if an edge in $\mathcal{E}_m$ had clock at most θ_m^*, then neither of its endpoints had been claimed when that edge was processed. The greedy algorithm would therefore have accepted it, a contradiction. In particular, every profile $\boldsymbol{\zeta}_{ki}$ with $(k,i) \in \mathcal{E}_m$ remains finite in a right neighborhood of θ_m^*.

The tensor represented by the first m accepted sectors and the corresponding residual target are

$$T_m := \sum_{r=1}^{m} \lambda_{i_r} \mathbf{e}_{i_r}^{(1)} \otimes \cdots \otimes \mathbf{e}_{i_r}^{(\nu)}, \qquad R_m := T - T_m. \tag{69}$$

The limiting greedy trajectory predicts the tensor T_m throughout the plateau between the mth and $(m+1)$st accepted events.

Captured-atom notation and the exact balance invariant. For an accepted edge (k_r, i_r), define its main coordinates, their product, and the component orthogonal to its selected sector by

$$q_{r,\sigma}^{(a)}(\theta) := x_{k_r i_r,\sigma}^{(a)}(\theta), \qquad \pi_{r,\sigma}(\theta) := \prod_{a=1}^{\nu} q_{r,\sigma}^{(a)}(\theta), \tag{70}$$

and

$$\mathbf{h}_{r,\sigma}^{(a)}(\theta) := \mathbf{u}_{k_r,\sigma}^{(a)}(\theta) - q_{r,\sigma}^{(a)}(\theta) \mathbf{e}_{i_r}^{(a)}. \tag{71}$$

By construction,

$$\big\langle \mathbf{e}_{i_r}^{(a)}, \mathbf{h}_{r,\sigma}^{(a)} \big\rangle = 0. \tag{72}$$

A key tool throughout the proof is the exact atom-wise balance invariant derived in (39). In the present slow-time notation it reads

$$\left\|\mathbf{u}_{k,\sigma}^{(a)}(\theta)\right\|^2 - \left\|\mathbf{u}_{k,\sigma}^{(b)}(\theta)\right\|^2 = \sigma^2 \left(\left\|\mathbf{g}_k^{(a)}\right\|^2 - \left\|\mathbf{g}_k^{(b)}\right\|^2\right) \tag{73}$$

for every atom k, every $a, b \in [\nu]$, and every time for which the GF solution exists. Since the initialization realization is fixed, there is a constant $C_{\mathrm{bal}} < \infty$ such that

$$\max_{\substack{k\in[H]\\ a,b\in[\nu]}} \left| \left\|\mathbf{u}_{k,\sigma}^{(a)}(\theta)\right\|^2 - \left\|\mathbf{u}_{k,\sigma}^{(b)}(\theta)\right\|^2 \right| \le C_{\mathrm{bal}}\sigma^2 \tag{74}$$

uniformly in time.

For a captured atom, the orthogonal decomposition (71) therefore gives

$$\left| |q_{r,\sigma}^{(a)}|^2 - |q_{r,\sigma}^{(b)}|^2 \right| \le C_{\mathrm{bal}}\sigma^2 + \left\|\mathbf{h}_{r,\sigma}^{(a)}\right\|^2 + \left\|\mathbf{h}_{r,\sigma}^{(b)}\right\|^2. \tag{75}$$

This relation will be used throughout the proof to control the balance of the mode-wise factors.

Reset times. The proof proceeds inductively through the accepted edges. Between successive accepted clock times, the unclaimed atoms remain microscopic and their coordinates in unclaimed sectors follow the original isolated profiles $\boldsymbol{\zeta}_{ki}$. Near the next accepted clock, the winning profile becomes large and the microscopic approximation breaks down. The associated atom then undergoes a fast transition and captures the corresponding target sector.

For each $m \ge 1$, we will construct a *reset time* ρ_m^σ after the mth fast transition such that

$$\rho_m^\sigma \longrightarrow \theta_m^* \qquad \text{as } \sigma \to 0, \tag{76}$$

and set

$$\rho_0^\sigma := 0. \tag{77}$$

Since the accepted clock times are distinct, for every fixed $m \in \{1, \dots, M\}$ and all sufficiently small σ,

$$\theta_{m-1}^* < \rho_m^\sigma < \theta_{m+1}^*. \tag{78}$$

The reset times are checkpoints in the proof; the GF dynamics itself is not restarted. In particular, surviving microscopic candidates continue to be compared with the isolated profiles $\boldsymbol{\zeta}_{ki}(\theta)$ that started from the original initialization at $\theta = 0$.

Notation used in the proof. Table 1 summarizes the principal notation that persists throughout the argument. Quantities specific to the entrance and fast-transition layers will be introduced locally when needed.

Quantitative induction hypothesis. For each fixed $m \in \{0, \dots, M\}$, we will construct a reset time ρ_m^σ, numbers $\varepsilon_{m,\sigma} > 0$, and a constant $C_m < \infty$, independent of σ, such that

$$\varepsilon_{m,\sigma} \longrightarrow 0 \qquad \text{as } \sigma \to 0, \tag{79}$$

and the following four properties hold at ρ_m^σ.

($\mathbf{I}_m$) Captured coefficients. For every accepted edge (k_r, i_r) with $r \le m$, the product of its main coordinates represents the corresponding target coefficient with the quantitative accuracy

$$\max_{1\le r\le m} |\pi_{r,\sigma}(\rho_m^\sigma) - \lambda_{i_r}| \le C_m\sigma^\nu. \tag{80}$$

($\mathbf{II}_m$) Microscopic surviving atoms. Every atom whose index has not yet been claimed remains at the initialization scale:

$$\max_{\substack{k\notin\mathcal{A}_m\\ a\in[\nu]}} \left\|\mathbf{u}_{k,\sigma}^{(a)}(\rho_m^\sigma)\right\| \le C_m\sigma. \tag{81}$$

Quantity	Meaning	Definition
$\mathbf{u}_{k,\sigma}^{(a)}(\theta)$	Exact mode-a weight vector of atom k, expressed in slow time.	(58)
$x_{ki,\sigma}^{(a)}(\theta)$	Exact scalar coordinate of atom k in target sector i.	(59)
$y_{ki,\sigma}^{(a)}(\theta)$	Microscopic coordinate $x_{ki,\sigma}^{(a)}(\theta)/\sigma$.	(60)
$\mathbf{z}(s;\mathbf{g})$	Generic σ-independent isolated unit-amplitude scalar-sector flow.	(61)
$\boldsymbol{\zeta}_{ki}(\theta)$	Isolated profile of candidate (k,i) in global slow time.	(62)
θ_m^*	Clock time of the mth edge accepted by the greedy process.	(64)
$\mathcal{A}_m, \mathcal{S}_m$	Atoms and sectors claimed after the first m accepted events.	(65)
$\mathcal{E}_m$	Candidate edges whose atom and sector are both still available.	(66)
$T_m,\ R_m$	Tensor already learned after m events and the corresponding residual target.	(69)
$q_{r,\sigma}^{(a)},\ \pi_{r,\sigma}$	Main coordinate of captured atom k_r in sector i_r, and the product of its ν main coordinates.	(70)
$\mathbf{h}_{r,\sigma}^{(a)}$	Off-sector component of captured atom k_r in mode a.	(71)
ρ_m^σ	Checkpoint after the mth fast transition, converging to θ_m^*.	(76)

Table 1: Principal notation used in the proof of Theorem 1.

(III$_m$) Preservation of the available candidate profiles. For every edge whose atom and sector are both still available, the exact microscopic coordinate remains close to its original isolated profile:

$$\max_{\substack{(k,i)\in\mathcal{E}_m\\ a\in[\nu]}} \left| y_{ki,\sigma}^{(a)}(\rho_m^\sigma) - \zeta_{ki}^{(a)}(\rho_m^\sigma) \right| \le \varepsilon_{m,\sigma}. \tag{82}$$

Thus surviving candidates are always compared with the profiles started from the original Gaussian initialization at $\theta = 0$; their clocks are not reset after earlier captures.

(IV$_m$) Blocking of claimed endpoints. Every captured atom is quantitatively confined to its assigned sector:

$$\max_{\substack{1\le r\le m\\ a\in[\nu]}} \left\| \mathbf{h}_{r,\sigma}^{(a)}(\rho_m^\sigma) \right\| \le C_m \sigma^\nu. \tag{83}$$

At the same time, every surviving atom remains microscopic inside every already claimed sector:

$$\max_{\substack{k\notin\mathcal{A}_m,\, i\in\mathcal{S}_m\\ a\in[\nu]}} \left| x_{ki,\sigma}^{(a)}(\rho_m^\sigma) \right| \le C_m \sigma, \tag{84}$$

or equivalently,

$$\max_{\substack{k\notin\mathcal{A}_m,\, i\in\mathcal{S}_m\\ a\in[\nu]}} \left| y_{ki,\sigma}^{(a)}(\rho_m^\sigma) \right| \le C_m. \tag{85}$$

These four properties constitute the induction hypothesis at ρ_m^σ.

Consequences of the induction hypotheses. We record two consequences that will be used repeatedly. First, the captured factors are asymptotically balanced. Indeed, (75) and ($\mathrm{IV_m}$) give

$$\max_{\substack{1\le r\le m\\ a,b\in[\nu]}} \left| |q_{r,\sigma}^{(a)}(\rho_m^\sigma)|^2 - |q_{r,\sigma}^{(b)}(\rho_m^\sigma)|^2 \right| = O(\sigma^2). \tag{86}$$

On the other hand, ($\mathrm{I_m}$) implies

$$\prod_{a=1}^{\nu} \left| q_{r,\sigma}^{(a)}(\rho_m^\sigma) \right| = \lambda_{i_r} + o(1). \tag{87}$$

Combining (86) and (87) yields

$$\max_{\substack{1\le r\le m\\ a\in[\nu]}} \left|\left|q_{r,\sigma}^{(a)}(\rho_m^\sigma)\right| - \lambda_{i_r}^{1/\nu}\right| \longrightarrow 0. \tag{88}$$

Second, the represented tensor at the reset time converges to the partial target T_m. For every captured atom, $(\mathrm{IV_m})$ and (88) give

$$\mathbf{u}_{k_r,\sigma}^{(1)}(\rho_m^\sigma) \otimes \cdots \otimes \mathbf{u}_{k_r,\sigma}^{(\nu)}(\rho_m^\sigma) \longrightarrow \lambda_{i_r}\mathbf{e}_{i_r}^{(1)} \otimes \cdots \otimes \mathbf{e}_{i_r}^{(\nu)}. \tag{89}$$

At the same time, $(\mathrm{II_m})$ implies that every surviving atom contributes $O(\sigma^\nu)$ in Frobenius norm. Since H is fixed,

$$\widehat{T}\Big(U^\sigma\Big(\sigma^{-(\nu-2)}\rho_m^\sigma\Big)\Big) \longrightarrow T_m. \tag{90}$$

Base case. For $m = 0$, $(\mathrm{I_0})$ and the captured-atom part of $(\mathrm{IV_0})$ are vacuous. Property $(\mathrm{II_0})$ follows directly from the Gaussian initialization because p, H, ν are fixed, while $(\mathrm{III_0})$ holds exactly since

$$y_{ki,\sigma}^{(a)}(0) = g_{ki}^{(a)} = \zeta_{ki}^{(a)}(0).$$

The claimed-sector part of $(\mathrm{IV_0})$ is also vacuous.

Proof roadmap. The induction step $m \to m+1$ is organized as follows.

1. *Stability of the captured background.* In Subsection A3.2, we show that, as long as the surviving atoms remain microscopic, the previously captured atoms stay locked to their sectors, available atom–sector pairs retain their isolated dynamics to leading order, and claimed sectors exert no leading-order drive on surviving atoms. These statements are summarized in Proposition 6.
2. *Propagation up to the next activation.* In Subsection A3.3, we close the microscopic bootstrap and propagate the induction state through every fixed interval lying strictly before the next accepted clock θ_{m+1}^*. Proposition 7 shows in particular that the available candidates continue to follow their original isolated profiles $\boldsymbol{\zeta}_{ki}$, so their clocks are not restarted after earlier captures.
3. *Entrance into the fast layer.* In Subsection A3.4, we approach θ_{m+1}^* on a σ-dependent mesoscopic scale. Proposition 8 constructs an entrance time $\eta_{m+1}^\sigma \to \theta_{m+1}^*$ at which the winning atom is much larger than the initialization scale but still vanishing in the original variables, while all competing atoms remain microscopic.
4. *Fast capture and renewal.* In Subsection A3.5, Lemma 6 carries the winning atom from the mesoscopic scale to a fixed macroscopic scale, and Lemma 7 then establishes locking of the enlarged captured set. Proposition 9 combines these estimates to construct a new reset time $\rho_{m+1}^\sigma \to \theta_{m+1}^*$ at which $(\mathrm{I_{m+1}})$–$(\mathrm{IV_{m+1}})$ hold.

Throughout the induction step, (73) controls the relative mode-wise magnitudes, while (82) carries the original candidate profiles through successive captures.

Completion of the proof. After iterating the induction through all accepted edges, Subsection A3.6 controls the plateau following the final capture. Lemma 8 gives the required terminal propagation estimate, and the resulting plateau bounds yield the tensor and loss convergence stated in Theorem 1.

A3.2 Dynamics in a locked background

We next isolate the estimates that describe the dynamics after some atoms have already captured distinct target sectors. The role of this subsection is deliberately conditional. Suppose that, on a given slow-time interval, the previously captured atoms start well locked to their selected sectors and the remaining atoms stay microscopic, of size $O(\sigma)$. We show that the captured atoms then form a stable background: they remain locked, available atom–sector pairs continue to follow their original isolated dynamics up to a vanishing perturbation, and already claimed sectors exert no leading-order drive on the surviving atoms.

The assumption that the surviving atoms remain microscopic will be proved in the next subsection by a separate slow-propagation argument. Keeping these two steps separate avoids mixing the stability of the captured background with the bootstrap that controls the next generation of candidate atoms.

Notation and exact equations. For $k, j \in [H]$ and $a \in [\nu]$, define the mode-wise Gram entries

$$G^{(a)}_{kj,\sigma}(\theta) := \left\langle \mathbf{u}^{(a)}_{k,\sigma}(\theta), \mathbf{u}^{(a)}_{j,\sigma}(\theta) \right\rangle. \tag{91}$$

In the slow time $\theta = \sigma^{\nu-2}t$, the exact GF equation can be written, using the diagonal form of the target, as

$$\frac{d\mathbf{u}^{(a)}_{k,\sigma}}{d\theta} = \sigma^{-(\nu-2)} \left[\sum_{i=1}^{p} \lambda_i \mathbf{e}^{(a)}_i \prod_{b\neq a} \overline{x^{(b)}_{ki,\sigma}} - \sum_{j=1}^{H} \mathbf{u}^{(a)}_{j,\sigma} \prod_{b\neq a} G^{(b)}_{kj,\sigma} \right]. \tag{92}$$

Projecting the exact vector equation (92) onto the sector direction $\mathbf{e}^{(a)}_i$ gives

$$\frac{dx^{(a)}_{ki,\sigma}}{d\theta} = \sigma^{-(\nu-2)} \left[\lambda_i \prod_{b\neq a} \overline{x^{(b)}_{ki,\sigma}} - \sum_{j=1}^{H} x^{(a)}_{ji,\sigma} \prod_{b\neq a} G^{(b)}_{kj,\sigma} \right]. \tag{93}$$

Finally, for convenience, we denote the exact predicted tensor in slow time by

$$\widehat{T}_\sigma(\theta) := \sum_{k=1}^{H} \mathbf{u}^{(1)}_{k,\sigma}(\theta) \otimes \cdots \otimes \mathbf{u}^{(\nu)}_{k,\sigma}(\theta). \tag{94}$$

A bootstrap continuation principle. We will use the same elementary continuation argument several times in the proof. We record it once here.

Lemma 1 (Bootstrap continuation)**.** *Let $I_\sigma = [\theta^\sigma_0, \theta^\sigma_1]$, let $E_\sigma : I_\sigma \to [0,\infty)$ be continuous, and let $a_\sigma > 0$. Fix constants*

$$B_{\mathrm{init}} > 0, \qquad B_{\mathrm{imp}} > 0, \qquad B_{\mathrm{boot}} > 0, \tag{95}$$

independent of σ, such that

$$B_{\mathrm{init}} < B_{\mathrm{boot}}, \qquad B_{\mathrm{imp}} < B_{\mathrm{boot}}. \tag{96}$$

No relation between B_{init} and B_{imp} is assumed.

Suppose that the initial error satisfies

$$E_\sigma(\theta^\sigma_0) \le B_{\mathrm{init}} a_\sigma. \tag{97}$$

Assume moreover that, for every $\theta \in [\theta^\sigma_0, \theta^\sigma_1]$ for which

$$\sup_{\vartheta \in [\theta^\sigma_0, \theta]} E_\sigma(\vartheta) \le B_{\mathrm{boot}} a_\sigma, \tag{98}$$

the dynamical estimates imply the sharper bound

$$\sup_{\vartheta \in [\theta^\sigma_0, \theta]} E_\sigma(\vartheta) \le B_{\mathrm{imp}} a_\sigma. \tag{99}$$

Then

$$\sup_{\theta \in I_\sigma} E_\sigma(\theta) \le B_{\mathrm{imp}} a_\sigma. \tag{100}$$

Proof. Define the maximal bootstrap time

$$\theta^\sigma_{\mathrm{boot}} := \sup \left\{ \theta \in I_\sigma : \sup_{\vartheta \in [\theta^\sigma_0, \theta]} E_\sigma(\vartheta) \le B_{\mathrm{boot}} a_\sigma \right\}. \tag{101}$$

By the initial bound (97) and the strict slack $B_{\mathrm{init}} < B_{\mathrm{boot}}$ in (96), continuity of E_σ implies

$$\theta^\sigma_{\mathrm{boot}} > \theta^\sigma_0. \tag{102}$$

On every compact subinterval of $[\theta_0^\sigma, \theta_{\mathrm{boot}}^\sigma)$, the bootstrap assumption (98) holds. Hence the improvement estimate (99), followed by continuity at the endpoint, gives

$$\sup_{\vartheta \in [\theta_0^\sigma, \theta_{\mathrm{boot}}^\sigma]} E_\sigma(\vartheta) \le B_{\mathrm{imp}} a_\sigma < B_{\mathrm{boot}} a_\sigma. \tag{103}$$

If $\theta_{\mathrm{boot}}^\sigma < \theta_1^\sigma$, the strict inequality in (103) and continuity would allow the bootstrap bound to remain valid on a slightly larger interval. This contradicts the definition of $\theta_{\mathrm{boot}}^\sigma$. Therefore

$$\theta_{\mathrm{boot}}^\sigma = \theta_1^\sigma. \tag{104}$$

Applying the improvement estimate (99) on the full interval gives (100). □

In the present application, the quantity to be bootstrapped will be the off-sector locking error

$$E_{\mathrm{lock},\sigma}(\theta) := \max_{\substack{r \le m \\ a \in [\nu]}} \big\| \mathbf{h}_{r,\sigma}^{(a)}(\theta) \big\|. \tag{105}$$

The continuation principle applies because GF is a smooth finite-dimensional ODE, hence its solution and all quantities in (105) are continuous as long as the weights remain bounded. On the bootstrap intervals considered below, boundedness follows from the coarse bounds on the captured coordinates, the microscopic bound on the surviving atoms, and the bootstrap bound itself. The constants in the estimates below depend only on these coarse bounds and not on the endpoint of the bootstrap interval.

Main locked-background statement.

Proposition 6 (Reduction in a locked background)**.** *Fix $m \in \{0, \dots, M\}$ and an interval $I_\sigma = [\theta_0^\sigma, \theta_1^\sigma]$ whose length is bounded uniformly in σ. Suppose that the following conditions hold.*

First, all surviving atoms remain microscopic on I_σ:

$$\max_{\substack{k \notin \mathcal{A}_m \\ a \in [\nu]}} \sup_{\theta \in I_\sigma} \big\| \mathbf{u}_{k,\sigma}^{(a)}(\theta) \big\| \le B\sigma. \tag{106}$$

Second, the main coordinates of the captured atoms stay in a fixed compact annulus:

$$c \le \big| q_{r,\sigma}^{(a)}(\theta) \big| \le C, \qquad r \le m, \quad a \in [\nu], \quad \theta \in I_\sigma, \tag{107}$$

for constants $0 < c < C < \infty$ independent of σ. Third, at the left endpoint the captured atoms are quantitatively locked:

$$\max_{\substack{r \le m \\ a \in [\nu]}} \big\| \mathbf{h}_{r,\sigma}^{(a)}(\theta_0^\sigma) \big\| + \max_{r \le m} \big| \pi_{r,\sigma}(\theta_0^\sigma) - \lambda_{i_r} \big| \le B\sigma^\nu. \tag{108}$$

Then there is a constant $C_{\mathrm{bg}} < \infty$, independent of σ, such that the following conclusions hold for all sufficiently small σ.

1. Stability of captured atoms. *The off-sector and product errors remain of order σ^ν:*

$$\max_{\substack{r \le m \\ a \in [\nu]}} \sup_{\theta \in I_\sigma} \big\| \mathbf{h}_{r,\sigma}^{(a)}(\theta) \big\| + \max_{r \le m} \sup_{\theta \in I_\sigma} \big| \pi_{r,\sigma}(\theta) - \lambda_{i_r} \big| \le C_{\mathrm{bg}} \sigma^\nu. \tag{109}$$

Moreover, the exact full-flow balance invariant gives

$$\max_{\substack{r \le m \\ a,b \in [\nu]}} \sup_{\theta \in I_\sigma} \Big| |q_{r,\sigma}^{(a)}(\theta)|^2 - |q_{r,\sigma}^{(b)}(\theta)|^2 \Big| \le C_{\mathrm{bg}} \sigma^2. \tag{110}$$

2. Available sectors retain the isolated dynamics. *For every available edge $(k,i) \in \mathcal{E}_m$,*

$$\frac{dy_{ki,\sigma}^{(a)}}{d\theta} = \lambda_i \prod_{b \ne a} \overline{y_{ki,\sigma}^{(b)}} + r_{ki,\sigma}^{(a)}(\theta), \tag{111}$$

where

$$\max_{\substack{(k,i) \in \mathcal{E}_m \\ a \in [\nu]}} \sup_{\theta \in I_\sigma} \big| r_{ki,\sigma}^{(a)}(\theta) \big| \le C_{\mathrm{bg}} \sigma^\nu. \tag{112}$$

3. Claimed sectors are dynamically cancelled. *If $k \notin \mathcal{A}_m$ and $i = i_r \in \mathcal{S}_m$, then*

$$\frac{dy^{(a)}_{ki_r,\sigma}}{d\theta} = \big(\lambda_{i_r} - \pi_{r,\sigma}\big) \prod_{b\neq a} \overline{y^{(b)}_{ki_r,\sigma}} + \widehat{r}^{(a)}_{ki_r,\sigma}(\theta), \tag{113}$$

with

$$\max_{\substack{k\notin\mathcal{A}_m,\ i_r\in\mathcal{S}_m \\ a\in[\nu]}} \sup_{\theta\in I_\sigma} \big|\widehat{r}^{(a)}_{ki_r,\sigma}(\theta)\big| \le C_{\mathrm{bg}}\sigma^{\nu}. \tag{114}$$

Consequently,

$$\max_{\substack{k\notin\mathcal{A}_m,\ i\in\mathcal{S}_m \\ a\in[\nu]}} \sup_{\theta\in I_\sigma} \left|\frac{dy^{(a)}_{ki,\sigma}}{d\theta}\right| \le C_{\mathrm{bg}}\sigma^{\nu}. \tag{115}$$

4. Tensor-level approximation. *The locked background represents the already selected part of the target:*

$$\sup_{\theta\in I_\sigma} \big\|\widehat{T}_\sigma(\theta) - T_m\big\|_F \le C_{\mathrm{bg}}\sigma^{\nu}. \tag{116}$$

We prove Proposition 6 through three elementary lemmas. The first is an exact decomposition of the dynamics of a captured coordinate. The second uses this decomposition and the bootstrap principle to prove stability of the captured atoms. The third derives the effective dynamics of the surviving atoms.

Captured-coordinate decomposition.

Lemma 2 (Exact equation for a captured coordinate)**.** *Assume the hypotheses (106) and (107). Let $J_\sigma \subseteq I_\sigma$ be an interval beginning at θ^σ_0 on which, for some fixed $B_* > 0$,*

$$\max_{\substack{r\le m \\ a\in[\nu]}} \sup_{\theta\in J_\sigma} \big\|\mathbf{h}^{(a)}_{r,\sigma}(\theta)\big\| \le B_*\sigma^{\nu}. \tag{117}$$

Then every captured main coordinate satisfies

$$\frac{dq^{(a)}_{r,\sigma}}{d\theta} = \sigma^{-(\nu-2)}\big(\lambda_{i_r} - \pi_{r,\sigma}\big) \prod_{b\neq a} \overline{q^{(b)}_{r,\sigma}} + R^{(a)}_{r,\sigma}(\theta), \tag{118}$$

where

$$\max_{\substack{r\le m \\ a\in[\nu]}} \sup_{\theta\in J_\sigma} \big|R^{(a)}_{r,\sigma}(\theta)\big| = O(\sigma^2). \tag{119}$$

Proof. Fix $r \le m$, set $i = i_r$, and abbreviate $q^{(a)} = q^{(a)}_{r,\sigma}$, $\pi = \pi_{r,\sigma}$, and $\mathbf{h}^{(a)} = \mathbf{h}^{(a)}_{r,\sigma}$.

Taking $k = k_r$ and $i = i_r$ in the exact sector-coordinate equation (93) gives

$$\frac{dq^{(a)}}{d\theta} = \sigma^{-(\nu-2)} \left[\lambda_i \prod_{b\neq a} \overline{q^{(b)}} - \sum_{j=1}^{H} x^{(a)}_{ji,\sigma} \prod_{b\neq a} G^{(b)}_{k_r j,\sigma}\right]. \tag{120}$$

We separate the self-interaction $j = k_r$ from the remaining terms.

Step 1: the self-interaction. Since the captured-atom decomposition (71) gives $\mathbf{u}^{(b)}_{k_r,\sigma} = q^{(b)}\mathbf{e}^{(b)}_i + \mathbf{h}^{(b)}$, and the two components are orthogonal by (72),

$$G^{(b)}_{k_r k_r,\sigma} = \big\|\mathbf{u}^{(b)}_{k_r,\sigma}\big\|^2 = |q^{(b)}|^2 + \|\mathbf{h}^{(b)}\|^2. \tag{121}$$

Hence the self-interaction equals

$$q^{(a)} \prod_{b\neq a} \Big(|q^{(b)}|^2 + \|\mathbf{h}^{(b)}\|^2\Big). \tag{122}$$

Its leading part can be rewritten using $\pi = \prod_c q^{(c)}$ as

$$q^{(a)} \prod_{b \neq a} |q^{(b)}|^2 = \pi \prod_{b \neq a} \overline{q^{(b)}}. \tag{123}$$

Therefore the correction caused by the off-sector components is

$$R^{\mathrm{self},(a)}_{r,\sigma} := -\sigma^{-(\nu-2)} q^{(a)} \left[\prod_{b \neq a} \left(|q^{(b)}|^2 + \|\mathbf{h}^{(b)}\|^2 \right) - \prod_{b \neq a} |q^{(b)}|^2 \right]. \tag{124}$$

By the off-sector bootstrap (117) and the coarse boundedness of the captured coordinates in (107),

$$R^{\mathrm{self},(a)}_{r,\sigma} = O(\sigma^{\nu+2}). \tag{125}$$

Step 2: interactions with surviving atoms. The contribution of all $j \notin \mathcal{A}_m$ is

$$R^{\mathrm{surv},(a)}_{r,\sigma} := -\sigma^{-(\nu-2)} \sum_{j \notin \mathcal{A}_m} x^{(a)}_{ji,\sigma} \prod_{b \neq a} G^{(b)}_{k_r j,\sigma}. \tag{126}$$

By the assumed microscopic bound on surviving atoms (106),

$$|x^{(a)}_{ji,\sigma}| = O(\sigma), \qquad |G^{(b)}_{k_r j,\sigma}| = O(\sigma). \tag{127}$$

Each summand before the slow-time prefactor is therefore $O(\sigma^\nu)$. Since H is fixed,

$$R^{\mathrm{surv},(a)}_{r,\sigma} = O(\sigma^2). \tag{128}$$

Step 3: interactions with the other captured atoms. For $s \leq m$, $s \neq r$, we have $i_s \neq i_r = i$. Hence the coordinate of captured atom k_s in the wrong sector is contained entirely in its off-sector component:

$$x^{(a)}_{k_s i,\sigma} = \left\langle \mathbf{e}^{(a)}_i, \mathbf{h}^{(a)}_{s,\sigma} \right\rangle = O(\sigma^\nu). \tag{129}$$

Moreover, orthogonality of the sector directions $\mathbf{e}^{(b)}_{i_r}$ and $\mathbf{e}^{(b)}_{i_s}$, together with the decomposition (71), gives

$$\begin{aligned} G^{(b)}_{k_r k_s,\sigma} = \overline{q^{(b)}_{r,\sigma}} \left\langle \mathbf{e}^{(b)}_{i_r}, \mathbf{h}^{(b)}_{s,\sigma} \right\rangle + q^{(b)}_{s,\sigma} \left\langle \mathbf{h}^{(b)}_{r,\sigma}, \mathbf{e}^{(b)}_{i_s} \right\rangle \\ + \left\langle \mathbf{h}^{(b)}_{r,\sigma}, \mathbf{h}^{(b)}_{s,\sigma} \right\rangle = O(\sigma^\nu). \end{aligned} \tag{130}$$

Thus the interaction with the other captured atoms,

$$R^{\mathrm{cap},(a)}_{r,\sigma} := -\sigma^{-(\nu-2)} \sum_{\substack{s \leq m \\ s \neq r}} x^{(a)}_{k_s i,\sigma} \prod_{b \neq a} G^{(b)}_{k_r k_s,\sigma}, \tag{131}$$

satisfies

$$R^{\mathrm{cap},(a)}_{r,\sigma} = O\left(\sigma^{\nu^2-\nu+2} \right). \tag{132}$$

Combining the three contributions, define

$$R^{(a)}_{r,\sigma} := R^{\mathrm{self},(a)}_{r,\sigma} + R^{\mathrm{surv},(a)}_{r,\sigma} + R^{\mathrm{cap},(a)}_{r,\sigma}. \tag{133}$$

Substituting the leading self-interaction identity (123) into the exact equation (120) now gives (118). Since $\nu \geq 3$, the bounds (125), (128), and (132) imply the remainder estimate (119). The dominant perturbation is the interaction with the still-microscopic atoms. □

Why the captured state is stable. Before proving the estimates, it is useful to describe the underlying vector-field geometry. Ignoring the vanishing interactions with the other atoms, the configurations

$$\mathbf{u}^{(a)} = q^{(a)}\mathbf{e}_i^{(a)}, \qquad \prod_{a=1}^{\nu} q^{(a)} = \lambda_i, \tag{134}$$

form a manifold of equilibria for a captured sector. The vector field is contracting in the two directions relevant to locking. The product error $\pi - \lambda_i$ obeys, to leading order,

$$\frac{d}{d\theta}(\pi - \lambda_i) = -\sigma^{-(\nu-2)} S(\mathbf{q})(\pi - \lambda_i), \qquad S(\mathbf{q}) = \sum_{a=1}^{\nu} \prod_{b\neq a} |q^{(b)}|^2, \tag{135}$$

while an off-sector perturbation satisfies

$$\frac{d\mathbf{h}^{(a)}}{d\theta} = -\sigma^{-(\nu-2)}\gamma^{(a)}(\mathbf{q})\mathbf{h}^{(a)} + \text{perturbing terms}, \qquad \gamma^{(a)}(\mathbf{q}) = \prod_{b\neq a} |q^{(b)}|^2. \tag{136}$$

The coarse lower bound $|q^{(a)}| \geq c$ from (107) therefore gives

$$S(\mathbf{q}) \geq \nu c^{2(\nu-1)}, \qquad \gamma^{(a)}(\mathbf{q}) \geq c^{2(\nu-1)} =: c_\gamma > 0. \tag{137}$$

Thus the equilibrium manifold is uniformly attracting in the product and off-sector directions. The remaining directions tangent to the manifold correspond to redistributing scale and phase among the mode-wise factors. The relative factor magnitudes are controlled globally by the exact atom-wise balance invariant (73), while the phase directions remain neutral.

Stability of the captured atoms.

Lemma 3 (Stability of a captured background)**.** *Under the assumptions of Proposition 6, the locking estimate (109) and the balance estimate (110) hold.*

Proof. We use Lemma 1 with the locking error (105) and scale $a_\sigma = \sigma^\nu$.

Choose a fixed bootstrap constant $B_* > B$. On the maximal interval on which

$$E_{\mathrm{lock},\sigma}(\theta) \leq B_*\sigma^\nu, \tag{138}$$

the captured-coordinate decomposition of Lemma 2 applies. We derive estimates that improve this bound.

Step 1: stability of the captured product. For a fixed r, differentiate $\pi_{r,\sigma} = \prod_a q_{r,\sigma}^{(a)}$ and use the effective captured coordinate equation (118). This gives

$$\frac{d\pi_{r,\sigma}}{d\theta} = \sigma^{-(\nu-2)}\big(\lambda_{i_r} - \pi_{r,\sigma}\big) S_{r,\sigma}(\theta) + R^{\pi}_{r,\sigma}(\theta), \tag{139}$$

where

$$S_{r,\sigma}(\theta) := \sum_{a=1}^{\nu} \prod_{b\neq a} |q^{(b)}_{r,\sigma}(\theta)|^2 \tag{140}$$

and

$$R^{\pi}_{r,\sigma}(\theta) := \sum_{a=1}^{\nu} R^{(a)}_{r,\sigma}(\theta) \prod_{c\neq a} q^{(c)}_{r,\sigma}(\theta) = O(\sigma^2). \tag{141}$$

The coarse annulus bounds (107) imply

$$0 < c_S \leq S_{r,\sigma}(\theta) \leq C_S < \infty \tag{142}$$

uniformly on the bootstrap interval.

Writing $e^{\pi}_{r,\sigma} := \pi_{r,\sigma} - \lambda_{i_r}$, variation of constants in the contracting product equation (139) yields

$$\big|e^{\pi}_{r,\sigma}(\theta)\big| \leq e^{-c_S(\theta-\theta_0^\sigma)/\sigma^{\nu-2}}\big|e^{\pi}_{r,\sigma}(\theta_0^\sigma)\big| + C\sigma^2 \int_{\theta_0^\sigma}^{\theta} e^{-c_S(\theta-s)/\sigma^{\nu-2}}\, ds. \tag{143}$$

Since the fast exponential kernel satisfies

$$\int_{\theta_0^\sigma}^{\theta} e^{-c_S(\theta-s)/\sigma^{\nu-2}}\,ds \le \frac{\sigma^{\nu-2}}{c_S}, \tag{144}$$

the initial quantitative locking assumption (108) gives

$$\sup_{\theta}\left|\pi_{r,\sigma}(\theta)-\lambda_{i_r}\right| \le C_\pi \sigma^{\nu} \tag{145}$$

on the bootstrap interval.

Step 2: stability of the off-sector components. Let

$$Q_{i_r}^{(a)} := I - \mathbf{e}_{i_r}^{(a)}\big(\mathbf{e}_{i_r}^{(a)}\big)^* \tag{146}$$

be the orthogonal projector onto the complement of the captured sector. By the definition of the off-sector component (71),

$$\mathbf{h}_{r,\sigma}^{(a)} = Q_{i_r}^{(a)}\mathbf{u}_{k_r,\sigma}^{(a)}. \tag{147}$$

Applying this projector to the exact vector GF equation (92) gives

$$\frac{d\mathbf{h}_{r,\sigma}^{(a)}}{d\theta} = -\sigma^{-(\nu-2)}\gamma_{r,\sigma}^{(a)}(\theta)\mathbf{h}_{r,\sigma}^{(a)} + \mathbf{F}_{r,\sigma}^{(a)}(\theta), \tag{148}$$

where

$$\gamma_{r,\sigma}^{(a)}(\theta) := \prod_{b\neq a}\left\|\mathbf{u}_{k_r,\sigma}^{(b)}(\theta)\right\|^2 \tag{149}$$

and

$$\begin{aligned}\mathbf{F}_{r,\sigma}^{(a)} := {}& \sigma^{-(\nu-2)}\sum_{\ell\neq i_r}\lambda_\ell \mathbf{e}_\ell^{(a)}\prod_{b\neq a}\overline{x_{k_r\ell,\sigma}^{(b)}} \\ & - \sigma^{-(\nu-2)}\sum_{j\neq k_r} Q_{i_r}^{(a)}\mathbf{u}_{j,\sigma}^{(a)}\prod_{b\neq a} G_{k_rj,\sigma}^{(b)}.\end{aligned} \tag{150}$$

The coarse lower bound on the captured main coordinates in (107) implies

$$\gamma_{r,\sigma}^{(a)}(\theta) \ge c^{2(\nu-1)} =: c_\gamma > 0. \tag{151}$$

We next estimate the forcing in (150). The interaction with surviving atoms is $O(\sigma^2)$ in slow time, by the same scaling as in the surviving-atom remainder estimate (128). The target components orthogonal to i_r contain $\nu-1$ off-sector coordinates of the same captured atom, while the interaction with another captured atom contains $\nu-1$ small cross-Gram factors. Hence, on the locking bootstrap (138),

$$\max_{\substack{r\le m\\ a\in[\nu]}}\left\|\mathbf{F}_{r,\sigma}^{(a)}(\theta)\right\| \le C_0\sigma^2 + C_1(B_*)\,\sigma^{\nu^2-2\nu+2}. \tag{152}$$

For $\nu \ge 3$, the second exponent is strictly larger than 2.

Variation of constants in the damped off-sector equation (148), together with the uniform damping bound (151), gives

$$\begin{aligned}\left\|\mathbf{h}_{r,\sigma}^{(a)}(\theta)\right\| \le {}& e^{-c_\gamma(\theta-\theta_0^\sigma)/\sigma^{\nu-2}}\left\|\mathbf{h}_{r,\sigma}^{(a)}(\theta_0^\sigma)\right\| \\ & + \int_{\theta_0^\sigma}^{\theta} e^{-c_\gamma(\theta-s)/\sigma^{\nu-2}}\left\|\mathbf{F}_{r,\sigma}^{(a)}(s)\right\|\,ds.\end{aligned} \tag{153}$$

Using the fast-kernel bound (144), with c_γ in place of c_S, gives

$$\sup_{\theta}\left\|\mathbf{h}_{r,\sigma}^{(a)}(\theta)\right\| \le (B + C_h + o(1))\,\sigma^{\nu}, \tag{154}$$

where C_h is independent of the bootstrap constant B_*.

Step 3: closing the locking bootstrap. Choose

$$B_* > B + C_h + 1. \tag{155}$$

For sufficiently small σ, the improved off-sector estimate (154) is then a strict improvement over the bootstrap assumption (138). The continuation principle of Lemma 1 therefore shows that the bootstrap interval is all of I_σ. Combining this conclusion with the product estimate (145) proves the locking bound (109).

Step 4: balance from the exact full-flow invariant. It remains to prove the balance estimate (110). The exact atom-wise invariant (73) gives, for the captured atom k_r, a difference of squared full mode norms of order $O(\sigma^2)$ at every time. Using the orthogonal decomposition (71), this was rewritten in (75) as

$$\left| |q_{r,\sigma}^{(a)}|^2 - |q_{r,\sigma}^{(b)}|^2 \right| \le C_{\mathrm{bal}}\sigma^2 + \left\| \mathbf{h}_{r,\sigma}^{(a)} \right\|^2 + \left\| \mathbf{h}_{r,\sigma}^{(b)} \right\|^2 .$$

The off-sector stability already proved in (109) gives $\|\mathbf{h}_{r,\sigma}^{(a)}\| = O(\sigma^\nu)$ uniformly on I_σ. Since $\nu \ge 3$, the last two terms are $O(\sigma^{2\nu})$ and can be absorbed into the $O(\sigma^2)$ bound. This proves (110). □

Effective dynamics of the surviving atoms.

Lemma 4 (Surviving atoms in a locked background)**.** *Under the assumptions of Proposition 6, the available-sector estimates (111)–(112) and the claimed-sector estimates (113)–(115) hold.*

Proof. For $k \notin \mathcal{A}_m$, divide the exact coordinate equation (93) by σ and use the microscopic rescaling $x_{ki,\sigma}^{(a)} = \sigma y_{ki,\sigma}^{(a)}$. This gives

$$\frac{dy_{ki,\sigma}^{(a)}}{d\theta} = \lambda_i \prod_{b\neq a} \overline{y_{ki,\sigma}^{(b)}} - \sigma^{-(\nu-1)} \sum_{j=1}^{H} x_{ji,\sigma}^{(a)} \prod_{b\neq a} G_{kj,\sigma}^{(b)}. \tag{156}$$

Step 1: an unclaimed sector. Suppose $(k,i) \in \mathcal{E}_m$, so $i \notin \mathcal{S}_m$.

If $j \notin \mathcal{A}_m$, then both atoms k and j are microscopic by the survivor hypothesis (106). Hence

$$|x_{ji,\sigma}^{(a)}| = O(\sigma), \qquad |G_{kj,\sigma}^{(b)}| = O(\sigma^2), \tag{157}$$

and the corresponding term in the exact microscopic equation (156), after multiplication by $\sigma^{-(\nu-1)}$, is $O(\sigma^\nu)$.

If instead $j = k_r$ is captured, then $i \neq i_r$. The off-sector locking estimate (109) therefore gives

$$|x_{k_r i,\sigma}^{(a)}| = O(\sigma^\nu), \qquad |G_{k k_r,\sigma}^{(b)}| = O(\sigma). \tag{158}$$

The corresponding rescaled interaction is again $O(\sigma^\nu)$. Summing over the fixed number of atoms proves the isolated-flow equation (111) and its remainder estimate (112).

Step 2: a claimed sector. Now let $i = i_r \in \mathcal{S}_m$. The leading interaction in the exact microscopic equation (156) comes from the captured atom k_r.

Using the captured-atom decomposition (71) and its locking estimate (109),

$$\begin{aligned} G_{k k_r,\sigma}^{(b)} &= \left\langle \mathbf{u}_{k,\sigma}^{(b)}, q_{r,\sigma}^{(b)} \mathbf{e}_{i_r}^{(b)} + \mathbf{h}_{r,\sigma}^{(b)} \right\rangle \\ &= \sigma q_{r,\sigma}^{(b)} \overline{y_{k i_r,\sigma}^{(b)}} + O(\sigma^{\nu+1}). \end{aligned} \tag{159}$$

Consequently,

$$\sigma^{-(\nu-1)} x_{k_r i_r,\sigma}^{(a)} \prod_{b\neq a} G_{k k_r,\sigma}^{(b)} = \pi_{r,\sigma} \prod_{b\neq a} \overline{y_{k i_r,\sigma}^{(b)}} + O(\sigma^\nu). \tag{160}$$

Thus the model contribution of the captured atom cancels the target drive in its claimed sector up to the product error $\lambda_{i_r} - \pi_{r,\sigma}$.

The remaining surviving atoms contribute $O(\sigma^\nu)$ by the small–small estimate (157), while the other captured atoms contribute $O(\sigma^\nu)$ by the same cross-sector estimate as in (158). Substituting these bounds into the exact microscopic equation (156) yields the claimed-sector equation (113) and the remainder bound (114).

Finally, the microscopic survivor hypothesis (106) implies

$$\max_{\substack{k\notin\mathcal{A}_m\\ i\in[p]\\ a\in[\nu]}} |y^{(a)}_{ki,\sigma}| \le B. \tag{161}$$

Together with the product locking estimate (109), the claimed-sector equation (113), and its remainder estimate (114), this gives the near-freezing bound (115). □

Proof of the locked-background proposition.

Proof of Proposition 6. We verify the four conclusions in the order in which they were stated.

1. Captured atoms. Lemma 3 gives the quantitative locking estimate (109) and, using the exact full-flow invariant, the balance estimate (110).

2. Available sectors. The first part of Lemma 4 gives the perturbed isolated-flow equation (111) and its $O(\sigma^\nu)$ remainder estimate (112).

3. Claimed sectors. The second part of Lemma 4 gives the target–model cancellation equation (113), the corresponding remainder bound (114), and hence the near-freezing estimate (115).

4. Tensor-level approximation. For a captured atom k_r, expand its tensor product using the decomposition (71):

$$\bigotimes_{a=1}^{\nu} \mathbf{u}^{(a)}_{k_r,\sigma} = \bigotimes_{a=1}^{\nu} \left(q^{(a)}_{r,\sigma} \mathbf{e}^{(a)}_{i_r} + \mathbf{h}^{(a)}_{r,\sigma} \right). \tag{162}$$

The term containing no off-sector factor equals

$$\pi_{r,\sigma} \mathbf{e}^{(1)}_{i_r} \otimes \cdots \otimes \mathbf{e}^{(\nu)}_{i_r}. \tag{163}$$

Every other term contains at least one $\mathbf{h}^{(a)}_{r,\sigma}$ and is therefore $O(\sigma^\nu)$ in Frobenius norm by the captured-atom locking estimate (109). Combining this with the product estimate in the same equation gives

$$\left\| \bigotimes_{a=1}^{\nu} \mathbf{u}^{(a)}_{k_r,\sigma} - \lambda_{i_r} \mathbf{e}^{(1)}_{i_r} \otimes \cdots \otimes \mathbf{e}^{(\nu)}_{i_r} \right\|_F = O(\sigma^\nu). \tag{164}$$

For a surviving atom, the microscopic bound (106) gives

$$\left\| \mathbf{u}^{(1)}_{k,\sigma} \otimes \cdots \otimes \mathbf{u}^{(\nu)}_{k,\sigma} \right\|_F = O(\sigma^\nu). \tag{165}$$

Since H is fixed in the present $\sigma \downarrow 0$ limit, summing the captured-atom errors (164) and the surviving-atom errors (165) over the finitely many atoms proves the tensor-level approximation (116). □

Proposition 6 is conditional on the surviving atoms remaining $O(\sigma)$ and on the already captured main coordinates remaining inside the coarse annulus (107). The next subsection closes precisely these conditions. Starting from a reset time ρ^σ_m, we use another continuation argument together with the available-sector dynamics (111) to show that the surviving microscopic coordinates remain close to their original isolated profiles $\boldsymbol{\zeta}_{ki}$ throughout the slow interval preceding the next accepted clock. At the same time, the captured-atom stability (109) and the exact balance estimate (110) keep the previously captured atoms inside the required coarse bounds.

A3.3 Slow propagation between activation events

We now remove the conditional assumption in Proposition 6. Starting from a reset time ρ_m^σ, we show that the surviving atoms remain microscopic and continue to follow their original isolated profiles throughout every compact subinterval lying strictly before the next accepted clock θ_{m+1}^*. At the same time, the previously captured atoms remain locked. This provides the uniform control on the plateaus between successive activation events.

For convenience, collect the microscopic coordinates of a candidate (k, i) into the vector

$$\mathbf{y}_{ki,\sigma}(\theta) := \big(y_{ki,\sigma}^{(1)}(\theta), \dots, y_{ki,\sigma}^{(\nu)}(\theta)\big) \in \mathbb{K}^\nu. \tag{166}$$

Recall that its limiting isolated profile is $\boldsymbol{\zeta}_{ki}(\theta)$ from (62). Define the scalar-sector vector field $\mathbf{F}_i : \mathbb{K}^\nu \to \mathbb{K}^\nu$ by

$$\big[\mathbf{F}_i(\mathbf{z})\big]^{(a)} := \lambda_i \prod_{b \neq a} \overline{z^{(b)}}. \tag{167}$$

Then the isolated profile satisfies

$$\frac{d}{d\theta}\boldsymbol{\zeta}_{ki}(\theta) = \mathbf{F}_i\big(\boldsymbol{\zeta}_{ki}(\theta)\big), \qquad 0 \le \theta < \tau_{ki}. \tag{168}$$

Proposition 7 (Slow propagation before the next clock)**.** *Fix $m \in \{0, \dots, M-1\}$ and suppose that the induction hypotheses $(\mathrm{I_m})$–$(\mathrm{IV_m})$ hold at the reset time ρ_m^σ. In particular, we will use the captured-product estimate (80), the microscopic survivor bound (81), the preservation of the available profiles (82), and the two claimed-endpoint bounds (83) and (85). We will also use the derived balance consequence (88).*

Fix a constant δ satisfying

$$0 < \delta < \frac{1}{2}\big(\theta_{m+1}^* - \theta_m^*\big). \tag{169}$$

Define

$$I_{m,\delta}^\sigma := \big[\rho_m^\sigma, \theta_{m+1}^* - \delta\big]. \tag{170}$$

Since $\rho_m^\sigma \to \theta_m^*$, this interval is nonempty for all sufficiently small σ.

Then the following statements hold.

1. Uniform preservation of the candidate profiles. *For every available edge,*

$$\max_{\substack{(k,i)\in\mathcal{E}_m \\ a\in[\nu]}} \sup_{\theta\in I_{m,\delta}^\sigma} \Big|y_{ki,\sigma}^{(a)}(\theta) - \zeta_{ki}^{(a)}(\theta)\Big| \longrightarrow 0 \qquad \text{as } \sigma \to 0. \tag{171}$$

2. Surviving atoms remain microscopic. *There exists a constant $C_{m,\delta} < \infty$, independent of σ, such that*

$$\max_{\substack{k\notin\mathcal{A}_m \\ a\in[\nu]}} \sup_{\theta\in I_{m,\delta}^\sigma} \big\|\mathbf{u}_{k,\sigma}^{(a)}(\theta)\big\| \le C_{m,\delta}\sigma. \tag{172}$$

   In already claimed sectors, their rescaled coordinates change only by $O(\sigma^\nu)$:

$$\max_{\substack{k\notin\mathcal{A}_m,\ i\in\mathcal{S}_m \\ a\in[\nu]}} \sup_{\theta\in I_{m,\delta}^\sigma} \Big|y_{ki,\sigma}^{(a)}(\theta) - y_{ki,\sigma}^{(a)}(\rho_m^\sigma)\Big| \le C_{m,\delta}\sigma^\nu. \tag{173}$$

3. Captured atoms remain locked and balanced. *Uniformly on $I_{m,\delta}^\sigma$,*

$$\max_{\substack{r\le m \\ a\in[\nu]}} \sup_{\theta\in I_{m,\delta}^\sigma} \big\|\mathbf{h}_{r,\sigma}^{(a)}(\theta)\big\| + \max_{r\le m} \sup_{\theta\in I_{m,\delta}^\sigma} \big|\pi_{r,\sigma}(\theta) - \lambda_{i_r}\big| = O(\sigma^\nu), \tag{174}$$

   and

$$\max_{\substack{r\le m \\ a\in[\nu]}} \sup_{\theta\in I_{m,\delta}^\sigma} \Big|\big|q_{r,\sigma}^{(a)}(\theta)\big| - \lambda_{i_r}^{1/\nu}\Big| \longrightarrow 0. \tag{175}$$

4. Uniform plateau approximation. *The exact tensor output satisfies*

$$\sup_{\theta\in I^{\sigma}_{m,\delta}} \left\|\widehat{T}_{\sigma}(\theta) - T_m\right\|_F \longrightarrow 0, \tag{176}$$

and consequently

$$\sup_{\theta\in I^{\sigma}_{m,\delta}} \left| L\Big(U^{\sigma}\Big(\sigma^{-(\nu-2)}\theta\Big)\Big) - \frac{1}{2}\sum_{i\notin\mathcal{S}_m}\lambda_i^2 \right| \longrightarrow 0. \tag{177}$$

Proof. The main point is to justify, rather than assume, the microscopic survivor bound required by Proposition 6. We do this by a continuation argument with three simultaneous bootstrap conditions: one for the available candidate profiles, one for coordinates in already claimed sectors, and one for the main coordinates of the captured atoms.

Step 1: boundedness of the isolated profiles. By the greedy characterization of the next accepted edge (67),

$$\tau_{ki} \geq \theta^*_{m+1} \qquad \text{for every } (k,i)\in\mathcal{E}_m, \tag{178}$$

with equality only for the next accepted edge (k_{m+1}, i_{m+1}). Thus every available isolated profile is regular on a fixed neighborhood of the interval between θ^*_m and $\theta^*_{m+1} - \delta$.

To also cover the possibly σ-dependent starting point ρ^{σ}_m, define

$$\underline{\theta}_{m,\delta} := \max\{0, \theta^*_m - \delta\}. \tag{179}$$

Since $\rho^{\sigma}_m \to \theta^*_m$, for all sufficiently small σ,

$$\rho^{\sigma}_m \geq \underline{\theta}_{m,\delta}. \tag{180}$$

Because there are only finitely many available candidates, the quantity

$$Z_{m,\delta} := \max_{\substack{(k,i)\in\mathcal{E}_m\\ a\in[\nu]}} \sup_{\theta\in[\underline{\theta}_{m,\delta},\,\theta^*_{m+1}-\delta]} \left|\zeta^{(a)}_{ki}(\theta)\right| \tag{181}$$

is finite. This fixed bound will control the available microscopic coordinates throughout the bootstrap interval.

Step 2: bootstrap interval. We introduce three bootstrap conditions. First, every available candidate is required to remain within unit distance of its isolated profile:

$$\max_{\substack{(k,i)\in\mathcal{E}_m\\ a\in[\nu]}} \left|y^{(a)}_{ki,\sigma}(\theta) - \zeta^{(a)}_{ki}(\theta)\right| \leq 1. \tag{182}$$

Second, in every already claimed sector we require

$$\max_{\substack{k\notin\mathcal{A}_m,\ i\in\mathcal{S}_m\\ a\in[\nu]}} \left|y^{(a)}_{ki,\sigma}(\theta)\right| \leq C_m + 1. \tag{183}$$

Third, every captured main coordinate is required to remain in the coarse annulus

$$\frac{1}{2}\lambda_{i_r}^{1/\nu} \leq \left|q^{(a)}_{r,\sigma}(\theta)\right| \leq 2\lambda_{i_r}^{1/\nu}, \qquad r\leq m, \quad a\in[\nu]. \tag{184}$$

For $m = 0$, the last two conditions are vacuous.

At $\theta = \rho^{\sigma}_m$, the available-profile preservation hypothesis (82), the claimed-sector bound (85), and the derived captured-factor balance (88) show that all three bootstrap conditions hold with strict slack for sufficiently small σ. By continuity, they therefore hold on a nontrivial interval beginning at ρ^{σ}_m.

Let

$$\theta^{\sigma}_{\mathrm{boot}} := \sup\left\{\theta\in I^{\sigma}_{m,\delta} : (182), (183), (184) \text{ hold on } [\rho^{\sigma}_m, \theta]\right\}. \tag{185}$$

We will show that each of the three bootstrap bounds is improved strictly on $[\rho^{\sigma}_m, \theta^{\sigma}_{\mathrm{boot}}]$. The same continuation argument as in Lemma 1 will then force $\theta^{\sigma}_{\mathrm{boot}}$ to equal the right endpoint of $I^{\sigma}_{m,\delta}$.

Step 3: the bootstrap conditions imply microscopic surviving atoms. For an unclaimed sector $i \notin \mathcal{S}_m$, the profile bootstrap (182) and the fixed isolated-profile bound (181) give

$$\left|y^{(a)}_{ki,\sigma}(\theta)\right| \le Z_{m,\delta} + 1. \tag{186}$$

For a claimed sector, the claimed-coordinate bootstrap (183) gives the corresponding bound $C_m + 1$. Since $x^{(a)}_{ki,\sigma} = \sigma y^{(a)}_{ki,\sigma}$ and $\{\mathbf{e}^{(a)}_i\}_{i=1}^p$ is an orthonormal basis,

$$\begin{aligned}\left\|\mathbf{u}^{(a)}_{k,\sigma}(\theta)\right\|^2 &= \sum_{i=1}^{p} \left|x^{(a)}_{ki,\sigma}(\theta)\right|^2 \\ &= \sigma^2 \sum_{i=1}^{p} \left|y^{(a)}_{ki,\sigma}(\theta)\right|^2 \le B^2_{m,\delta}\sigma^2\end{aligned} \tag{187}$$

for a constant $B_{m,\delta} < \infty$ independent of σ.

The captured-coordinate bootstrap (184) supplies the uniform lower and upper bounds required by the coarse annulus hypothesis (107) of Proposition 6. Its quantitative locking hypothesis at the left endpoint follows directly from the induction hypotheses. Indeed, the captured-product condition (80) gives

$$\max_{r\le m} |\pi_{r,\sigma}(\rho^\sigma_m) - \lambda_{i_r}| \le C_m \sigma^\nu, \tag{188}$$

while the captured off-sector condition (83) gives

$$\max_{\substack{r\le m\\ a\in[\nu]}} \left\|\mathbf{h}^{(a)}_{r,\sigma}(\rho^\sigma_m)\right\| \le C_m\sigma^\nu. \tag{189}$$

Finally, the bootstrap interval has uniformly bounded length because it is contained in $I^\sigma_{m,\delta}$ and $\rho^\sigma_m \to \theta^*_m$. Thus all assumptions of Proposition 6 hold on $[\rho^\sigma_m, \theta^\sigma_{\mathrm{boot}}]$.

Step 4: available candidates remain close to their original profiles. By the available-sector conclusion (111)–(112) of Proposition 6, every available candidate satisfies

$$\frac{d\mathbf{y}_{ki,\sigma}}{d\theta} = \mathbf{F}_i(\mathbf{y}_{ki,\sigma}) + \mathbf{r}_{ki,\sigma}(\theta), \qquad \sup_\theta \left\|\mathbf{r}_{ki,\sigma}(\theta)\right\| \le C_{m,\delta}\sigma^\nu. \tag{190}$$

The corresponding isolated profile solves the unperturbed equation (168).

Both $\mathbf{y}_{ki,\sigma}$ and $\boldsymbol{\zeta}_{ki}$ remain in a fixed bounded subset of $\mathbb{K}^\nu$ on the bootstrap interval. Since the scalar-sector vector field $\mathbf{F}_i$ in (167) is polynomial, viewed as a smooth real vector field when $\mathbb{K} = \mathbb{C}$, there is a constant $L_{m,\delta} < \infty$ such that

$$\left\|\mathbf{F}_i(\mathbf{v}) - \mathbf{F}_i(\mathbf{w})\right\| \le L_{m,\delta}\|\mathbf{v} - \mathbf{w}\| \tag{191}$$

for all $\mathbf{v}, \mathbf{w}$ in this bounded set and all $i \notin \mathcal{S}_m$.

Set

$$\mathbf{d}_{ki,\sigma}(\theta) := \mathbf{y}_{ki,\sigma}(\theta) - \boldsymbol{\zeta}_{ki}(\theta). \tag{192}$$

Subtracting the isolated profile equation (168) from the perturbed exact equation (190) and using the Lipschitz bound (191) gives

$$\frac{d}{d\theta}\|\mathbf{d}_{ki,\sigma}(\theta)\| \le L_{m,\delta}\|\mathbf{d}_{ki,\sigma}(\theta)\| + C_{m,\delta}\sigma^\nu \tag{193}$$

in the standard upper-Dini-derivative sense at points where the norm vanishes.

Gronwall's inequality and the initial profile error (82) therefore give

$$\sup_{\theta\in[\rho^\sigma_m,\theta^\sigma_{\mathrm{boot}}]} \|\mathbf{d}_{ki,\sigma}(\theta)\| \le C'_{m,\delta}\left(\varepsilon_{m,\sigma} + \sigma^\nu\right). \tag{194}$$

Since $\varepsilon_{m,\sigma} \to 0$, the right-hand side tends to zero uniformly over the finitely many available candidates. In particular, for all sufficiently small σ, it is strictly smaller than the unit threshold in the profile bootstrap (182).

Step 5: claimed coordinates remain nearly frozen. The claimed-sector near-freezing estimate (115) from Proposition 6 gives

$$\max_{\substack{k\notin\mathcal{A}_m,\ i\in\mathcal{S}_m\\ a\in[\nu]}}\left|\frac{dy^{(a)}_{ki,\sigma}}{d\theta}\right|\le C_{m,\delta}\sigma^{\nu} \tag{195}$$

throughout the bootstrap interval. Since $I^{\sigma}_{m,\delta}$ has uniformly bounded length, integration gives

$$\max_{\substack{k\notin\mathcal{A}_m,\ i\in\mathcal{S}_m\\ a\in[\nu]}}\sup_{\theta\in[\rho^{\sigma}_m,\theta^{\sigma}_{\mathrm{boot}}]}\left|y^{(a)}_{ki,\sigma}(\theta)-y^{(a)}_{ki,\sigma}(\rho^{\sigma}_m)\right|\le C'_{m,\delta}\sigma^{\nu}. \tag{196}$$

Together with the claimed-sector reset bound (85), this is a strict improvement of the bootstrap condition (183) for sufficiently small σ.

Step 6: captured coordinates remain inside the coarse annulus. The captured-product stability (109) from Proposition 6 gives

$$\max_{r\le m}\sup_{\theta\in[\rho^{\sigma}_m,\theta^{\sigma}_{\mathrm{boot}}]}|\pi_{r,\sigma}(\theta)-\lambda_{i_r}|=O(\sigma^{\nu}). \tag{197}$$

At the same time, the exact balance consequence (110) gives

$$\max_{\substack{r\le m\\ a,b\in[\nu]}}\sup_{\theta\in[\rho^{\sigma}_m,\theta^{\sigma}_{\mathrm{boot}}]}\left||q^{(a)}_{r,\sigma}(\theta)|^2-|q^{(b)}_{r,\sigma}(\theta)|^2\right|=O(\sigma^2). \tag{198}$$

Thus the mode-wise magnitudes of each captured atom are uniformly asymptotically equal throughout the bootstrap interval.

On the other hand, the product control (197) implies

$$\prod_{a=1}^{\nu}\left|q^{(a)}_{r,\sigma}(\theta)\right|=\left|\pi_{r,\sigma}(\theta)\right|=\lambda_{i_r}+o(1) \tag{199}$$

uniformly on the same interval. Combining the asymptotic equality of the mode-wise magnitudes (198) with the product identity (199) yields

$$\max_{\substack{r\le m\\ a\in[\nu]}}\sup_{\theta\in[\rho^{\sigma}_m,\theta^{\sigma}_{\mathrm{boot}}]}\left|\left|q^{(a)}_{r,\sigma}(\theta)\right|-\lambda_{i_r}^{1/\nu}\right|\longrightarrow 0. \tag{200}$$

For sufficiently small σ, every captured main coordinate therefore lies strictly inside the coarse annulus (184).

Step 7: closure of the continuation argument. The profile estimate (194), the claimed-sector estimate (196), and the captured-factor estimate (200) strictly improve, respectively, the three bootstrap conditions (182), (183), and (184). By the same continuity argument as in Lemma 1, the bootstrap interval cannot terminate before the right endpoint of $I^{\sigma}_{m,\delta}$. Hence

$$\theta^{\sigma}_{\mathrm{boot}}=\theta^{*}_{m+1}-\delta. \tag{201}$$

The Gronwall estimate (194) now gives the uniform profile convergence (171) on the full interval, while the integrated near-freezing estimate (196) gives (173). The microscopic coordinate bounds used in (187) yield the survivor estimate (172). Finally, the captured-background stability (109) and the balanced-magnitude estimate (200) give (174) and (175).

Step 8: tensor and loss convergence. The tensor-level conclusion (116) of Proposition 6, now valid on the full interval $I^{\sigma}_{m,\delta}$, gives

$$\sup_{\theta\in I^{\sigma}_{m,\delta}}\left\|\widehat{T}_{\sigma}(\theta)-T_m\right\|_F=O(\sigma^{\nu}). \tag{202}$$

This proves the uniform tensor convergence (176).

Since the loss is one half of the squared Frobenius residual, the tensor estimate (202) implies

$$\sup_{\theta\in I^{\sigma}_{m,\delta}}\left|\frac{1}{2}\big\|T-\widehat{T}_{\sigma}(\theta)\big\|_F^2-\frac{1}{2}\|T-T_m\|_F^2\right|\longrightarrow 0. \tag{203}$$

Finally, orthonormality of the diagonal target components and the definition of the partial target T_m in (69) give

$$\|T-T_m\|_F^2=\sum_{i\notin\mathcal{S}_m}\lambda_i^2. \tag{204}$$

Substitution into (203) proves the loss convergence (177). □

Proposition 7 controls the dynamics uniformly on every fixed interval that stays a positive distance before the next accepted clock. It does not yet describe a shrinking neighborhood of θ^*_{m+1}, where the winning isolated profile $\boldsymbol{\zeta}_{k_{m+1}i_{m+1}}$ becomes large and the microscopic approximation eventually ceases to be valid. The next subsection therefore constructs a σ-dependent entrance time approaching θ^*_{m+1} from below. At that time, the winning candidate will already be much larger than the initialization scale while still vanishing in the original variables, whereas all competing candidates remain microscopic. This provides the initial state for the subsequent fast-capture analysis.

A3.4 Entrance into the next fast layer

Proposition 7 controls the dynamics on every fixed compact interval lying strictly before the next accepted clock. We now extend this control to a σ-dependent time approaching that clock.

The main idea is a diagonal argument. For every fixed large amplitude R, the winning isolated profile reaches amplitude R at a time strictly before its blow-up, so Proposition 7 applies up to that time. We then let the entrance level diverge sufficiently slowly as $\sigma\to 0$. The resulting exact winning atom is much larger than the initialization scale but still vanishing in the original variables. This provides the entrance state for the fast-capture analysis of the next subsection.

We first record the asymptotic form of an isolated profile near its activation clock.

Lemma 5 (Approach to an isolated activation clock)**.** *Let (k,i) be any candidate and let $\boldsymbol{\zeta}_{ki}(\theta)$ be its isolated profile (62), with activation clock τ_{ki}. Define its geometric-mean amplitude by*

$$A_{ki}(\theta):=\left(\prod_{a=1}^{\nu}\left|\zeta_{ki}^{(a)}(\theta)\right|\right)^{1/\nu}. \tag{205}$$

Then, as $\theta\uparrow\tau_{ki}$,

$$A_{ki}(\theta)^{\nu-2}\big(\tau_{ki}-\theta\big)\longrightarrow\frac{1}{(\nu-2)\lambda_i}. \tag{206}$$

Moreover, the factor magnitudes become asymptotically balanced:

$$\frac{\left|\zeta_{ki}^{(a)}(\theta)\right|}{A_{ki}(\theta)}\longrightarrow 1,\qquad a\in[\nu]. \tag{207}$$

There also exist phases $\omega_{ki}^{(a)}\in\mathbb{K}$ with $|\omega_{ki}^{(a)}|=1$ such that

$$\frac{\zeta_{ki}^{(a)}(\theta)}{|\zeta_{ki}^{(a)}(\theta)|}\longrightarrow\omega_{ki}^{(a)},\qquad a\in[\nu], \tag{208}$$

and these limiting phases satisfy

$$\prod_{a=1}^{\nu}\omega_{ki}^{(a)}=1. \tag{209}$$

For $\mathbb{K}=\mathbb{R}$, the limiting phases $\omega_{ki}^{(a)}$ belong to $\{-1,1\}$.

Proof. Recall from the scalar-flow analysis in Appendix A2 that the isolated unit-amplitude flow satisfies

$$|z^{(a)}(s)|^2 = \rho_a + q(s), \qquad \rho_a := |g^{(a)}|^2, \tag{210}$$

where, on the final increasing branch approaching the blow-up time,

$$q'(s) = 2\sqrt{\prod_{a=1}^{\nu}\big(\rho_a + q(s)\big) - B^2}, \qquad B := \operatorname{Im}\prod_{a=1}^{\nu} g^{(a)}. \tag{211}$$

As $q \to \infty$, the scalar equation (211) gives

$$q'(s) = 2q(s)^{\nu/2}\left(1 + O(q(s)^{-1})\right). \tag{212}$$

Writing $r(s) := \sqrt{q(s)}$, the large-amplitude equation (212) becomes

$$r'(s) = r(s)^{\nu-1}\left(1 + O(r(s)^{-2})\right). \tag{213}$$

Equivalently,

$$\frac{d}{ds} r(s)^{-(\nu-2)} = -(\nu-2)\left(1 + O(r(s)^{-2})\right). \tag{214}$$

Integrating the reciprocal-amplitude equation (214) up to the blow-up time $\tau_\nu(\mathbf{g})$ gives

$$r(s)^{\nu-2}\big(\tau_\nu(\mathbf{g}) - s\big) \longrightarrow \frac{1}{\nu-2}. \tag{215}$$

For the candidate (k, i), the representation (210) reads

$$|\zeta_{ki}^{(a)}(\theta)|^2 = \rho_a + q(\lambda_i\theta),$$

and its activation clock satisfies $\tau_\nu(\mathbf{g}_{ki}) = \lambda_i\tau_{ki}$. Since the geometric-mean amplitude is asymptotic to $\sqrt{q(\lambda_i\theta)}$, the unit-flow blow-up law (215) gives the clock asymptotic (206).

The same representation shows that

$$|\zeta_{ki}^{(a)}(\theta)|^2 - |\zeta_{ki}^{(b)}(\theta)|^2 = \rho_a - \rho_b \tag{216}$$

is constant. Since all factor magnitudes diverge as $\theta \uparrow \tau_{ki}$, the fixed differences in (216) become negligible relative to their common magnitude. This proves the asymptotic balance (207).

It remains to identify the limiting phases. Define the scalar product

$$P_{ki}(\theta) := \prod_{a=1}^{\nu} \zeta_{ki}^{(a)}(\theta). \tag{217}$$

The isolated equations imply that $\operatorname{Im} P_{ki}(\theta)$ is constant, while $\operatorname{Re} P_{ki}(\theta) \to +\infty$ as the profile approaches its clock. Therefore

$$\frac{P_{ki}(\theta)}{|P_{ki}(\theta)|} \longrightarrow 1. \tag{218}$$

For any mode a, once $\zeta_{ki}^{(a)}(\theta) \neq 0$, the isolated flow gives

$$\frac{d}{d\theta}\arg\zeta_{ki}^{(a)}(\theta) = -\lambda_i \frac{\operatorname{Im} P_{ki}(\theta)}{|\zeta_{ki}^{(a)}(\theta)|^2}. \tag{219}$$

The numerator in the phase equation (219) is constant, whereas its denominator diverges. Hence

$$\frac{d}{d\theta}\arg\zeta_{ki}^{(a)}(\theta) \longrightarrow 0 \qquad \text{as } \theta \uparrow \tau_{ki}. \tag{220}$$

In particular, the phase derivative is bounded in a left neighborhood of the finite time τ_{ki} and is therefore integrable there. Thus every factor phase has a finite limit, proving (208). Taking the product of the limiting phases and using the product-phase limit (218) gives (209). □

We now specialize to the next accepted edge. Write

$$k_* := k_{m+1}, \qquad i_* := i_{m+1}, \qquad \theta_* := \theta^*_{m+1} = \tau_{k_* i_*}. \tag{221}$$

Since all candidate clocks are distinct and (k_*, i_*) is the unique minimum over the available set $\mathcal{E}_m$, every other available candidate has a strictly later clock. Because $\mathcal{E}_m$ is finite, there is a positive clock gap

$$\Delta_m := \min_{\substack{(k,i)\in\mathcal{E}_m \\ (k,i)\neq(k_*,i_*)}} \big(\tau_{ki} - \theta_*\big) > 0, \tag{222}$$

with the convention $\Delta_m = +\infty$ if $\mathcal{E}_m = \{(k_*, i_*)\}$. Consequently, all nonwinning available profiles remain uniformly bounded up to the winning clock:

$$B^{\mathrm{nw}}_m := \max_{\substack{(k,i)\in\mathcal{E}_m \\ (k,i)\neq(k_*,i_*)}} \max_{a\in[\nu]} \sup_{0\le\theta\le\theta_*} \left|\zeta^{(a)}_{ki}(\theta)\right| < \infty, \tag{223}$$

where the maximum over the empty set is interpreted as zero.

For the winning isolated profile, define

$$A_*(\theta) := A_{k_* i_*}(\theta). \tag{224}$$

By the isolated blow-up lemma, Lemma 5, $A_*(\theta) \to \infty$ as $\theta \uparrow \theta_*$. Hence, for every sufficiently large R, there is a unique time on the final increasing branch, denoted by $\eta_{m+1}(R)$, such that

$$A_*\big(\eta_{m+1}(R)\big) = R. \tag{225}$$

Indeed, the scalar representation (210) gives

$$A_*(\theta) = \left[\prod_{a=1}^{\nu} \big(\rho_a + q(\lambda_{i_*}\theta)\big)\right]^{1/(2\nu)}, \tag{226}$$

and the scalar equation (211) shows that q is strictly increasing on the final branch. Thus sufficiently large amplitude levels are crossed exactly once.

The blow-up-rate estimate (206) furthermore gives

$$\theta_* - \eta_{m+1}(R) = \frac{1 + o_R(1)}{(\nu-2)\lambda_{i_*} R^{\nu-2}}, \qquad R \to \infty. \tag{227}$$

Proposition 8 (Mesoscopic entrance into the next fast layer)**.** *Fix $m \in \{0, \dots, M-1\}$ and suppose that the induction hypotheses* ($\mathrm{I_m}$)–($\mathrm{IV_m}$) *hold at the reset time ρ^σ_m.*

Then one can choose an initialization-dependent scale $R_{m,\sigma}$ and define

$$\eta^\sigma_{m+1} := \eta_{m+1}(R_{m,\sigma}) \tag{228}$$

such that the following properties hold.

1. Mesoscopic scale. *The scale can be chosen so that*

$$R_{m,\sigma} \longrightarrow \infty, \qquad \sigma R^q_{m,\sigma} \longrightarrow 0 \quad \textit{for every fixed } q > 0. \tag{229}$$

Define the corresponding factor scale in the original variables by

$$r_{m,\sigma} := \sigma R_{m,\sigma}. \tag{230}$$

Then

$$\sigma \ll r_{m,\sigma} \ll 1. \tag{231}$$

2. The entrance time approaches the next clock. *We have*

$$\eta^\sigma_{m+1} \longrightarrow \theta_*, \tag{232}$$

and, more precisely,

$$R^{\nu-2}_{m,\sigma}\big(\theta_* - \eta^\sigma_{m+1}\big) \longrightarrow \frac{1}{(\nu-2)\lambda_{i_*}}. \tag{233}$$

3. The winning atom has reached mesoscopic size. *Let*

$$\boldsymbol{\omega}_* = \big(\omega_*^{(1)}, \dots, \omega_*^{(\nu)}\big)$$

be the limiting phase vector of the winning isolated profile from Lemma 5. Then

$$\max_{a\in[\nu]} \left| \frac{x^{(a)}_{k_* i_*,\sigma}(\eta^\sigma_{m+1})}{r_{m,\sigma}} - \omega_*^{(a)} \right| \longrightarrow 0, \tag{234}$$

where

$$|\omega_*^{(a)}| = 1, \qquad \prod_{a=1}^{\nu} \omega_*^{(a)} = 1. \tag{235}$$

Thus the winning main coordinates are asymptotically balanced and phase aligned.

4. The winning atom is localized in the winning sector. *There exists a constant $C_m^{\mathrm{ent}} < \infty$, independent of σ, such that*

$$\max_{a\in[\nu]} \left\| \mathbf{u}^{(a)}_{k_*,\sigma}(\eta^\sigma_{m+1}) - x^{(a)}_{k_* i_*,\sigma}(\eta^\sigma_{m+1}) \mathbf{e}^{(a)}_{i_*} \right\| \le C_m^{\mathrm{ent}} \sigma \tag{236}$$

for all sufficiently small σ. In view of the scale separation (231), this off-sector component is $o(r_{m,\sigma})$.

5. All other surviving atoms remain microscopic. *After increasing C_m^{ent} if necessary,*

$$\max_{\substack{k\notin\mathcal{A}_m \\ k\neq k_* \\ a\in[\nu]}} \left\| \mathbf{u}^{(a)}_{k,\sigma}(\eta^\sigma_{m+1}) \right\| \le C_m^{\mathrm{ent}} \sigma \tag{237}$$

for all sufficiently small σ.

6. The previously captured background remains quantitatively locked. *After increasing C_m^{ent} once more if necessary, the product and off-sector errors of every previously captured atom satisfy*

$$\max_{r\le m} \left| \pi_{r,\sigma}(\eta^\sigma_{m+1}) - \lambda_{i_r} \right| + \max_{\substack{r\le m \\ a\in[\nu]}} \left\| \mathbf{h}^{(a)}_{r,\sigma}(\eta^\sigma_{m+1}) \right\| \le C_m^{\mathrm{ent}} \sigma. \tag{238}$$

Their factor magnitudes remain asymptotically balanced:

$$\max_{\substack{r\le m \\ a\in[\nu]}} \left| \left| q^{(a)}_{r,\sigma}(\eta^\sigma_{m+1}) \right| - \lambda_{i_r}^{1/\nu} \right| \longrightarrow 0. \tag{239}$$

Consequently, the tensor output immediately before the fast transition satisfies

$$\widehat{T}_\sigma(\eta^\sigma_{m+1}) \longrightarrow T_m. \tag{240}$$

Proof. The proof is a diagonalization of the fixed-distance estimates from Proposition 7. The important point is that the constants in that proposition may depend on the distance from the next clock. We therefore keep the entrance level fixed while taking $\sigma \to 0$, and only afterwards let the entrance level diverge. This produces a single σ-dependent scale for which all estimates needed by the subsequent fast analysis hold simultaneously.

Step 1: fixed entrance levels. Choose a strictly increasing sequence

$$1 < R_1 < R_2 < \cdots, \qquad R_n \longrightarrow \infty. \tag{241}$$

By the clock asymptotic and phase-balance conclusions of Lemma 5, we may choose this sequence so that, if

$$\eta_n := \eta_{m+1}(R_n), \tag{242}$$

then

$$\left| R_n^{\nu-2}\big(\theta_* - \eta_n\big) - \frac{1}{(\nu-2)\lambda_{i_*}} \right| \le \frac{1}{n} \tag{243}$$

and

$$\max_{a\in[\nu]}\left|\frac{\zeta^{(a)}_{k_*i_*}(\eta_n)}{R_n}-\omega^{(a)}_*\right|\le\frac{1}{n}. \tag{244}$$

By increasing R_n further if necessary, we may also ensure

$$0<\theta_*-\eta_n<\frac{1}{2}\big(\theta^*_{m+1}-\theta^*_m\big). \tag{245}$$

Since $\theta_*=\theta^*_{m+1}$, the last inequality implies $\eta_n>\theta^*_m$.

Define the fixed distance from the next clock by

$$\delta_n:=\theta_*-\eta_n. \tag{246}$$

For each fixed n, $\delta_n>0$ is independent of σ, so the slow-propagation result Proposition 7 applies with $\delta=\delta_n$. Moreover, since $\rho^\sigma_m\to\theta^*_m$ and $\eta_n>\theta^*_m$, the interval $[\rho^\sigma_m,\eta_n]$ is nonempty for all sufficiently small σ.

Step 2: estimates at a fixed entrance level. Fix n. Applying the available-profile conclusion (171) of Proposition 7 with $\delta=\delta_n$ gives

$$\max_{\substack{(k,i)\in\mathcal{E}_m\\ a\in[\nu]}}\sup_{\theta\in[\rho^\sigma_m,\eta_n]}\left|y^{(a)}_{ki,\sigma}(\theta)-\zeta^{(a)}_{ki}(\theta)\right|\longrightarrow 0\qquad\text{as }\sigma\to 0. \tag{247}$$

For the microscopic coordinates of surviving atoms inside already claimed sectors, the near-freezing conclusion (173) gives a finite constant $C^{\mathrm{cl}}_n<\infty$, depending on n through δ_n but not on σ, such that

$$\max_{\substack{k\notin\mathcal{A}_m,\,i\in\mathcal{S}_m\\ a\in[\nu]}}\sup_{\theta\in[\rho^\sigma_m,\eta_n]}\left|y^{(a)}_{ki,\sigma}(\theta)-y^{(a)}_{ki,\sigma}(\rho^\sigma_m)\right|\le C^{\mathrm{cl}}_n\sigma^\nu \tag{248}$$

for all sufficiently small σ.

The captured-background part of Proposition 7 is also quantitative. In particular, the locking estimate (174) gives, for some finite constant $C^{\mathrm{old}}_n<\infty$,

$$\max_{r\le m}|\pi_{r,\sigma}(\eta_n)-\lambda_{i_r}|+\max_{\substack{r\le m\\ a\in[\nu]}}\left\|\mathbf{h}^{(a)}_{r,\sigma}(\eta_n)\right\|\le C^{\mathrm{old}}_n\sigma^\nu \tag{249}$$

for all sufficiently small σ.

Finally, the captured-factor balance conclusion (175) gives

$$\max_{\substack{r\le m\\ a\in[\nu]}}\left|\left|q^{(a)}_{r,\sigma}(\eta_n)\right|-\lambda_{i_r}^{1/\nu}\right|\longrightarrow 0\qquad\text{as }\sigma\to 0. \tag{250}$$

The constants C^{cl}_n and C^{old}_n may diverge as $n\to\infty$, because $\delta_n\downarrow 0$. We do not require any uniformity in n. Instead, the next step absorbs these constants into the diagonal choice of the initialization scale.

Step 3: diagonal choice of σ_n and $R_{m,\sigma}$. We recursively choose a strictly decreasing sequence $\sigma_n\downarrow 0$ such that

$$\sigma_{n+1}<\frac{1}{2}\sigma_n,\qquad\sigma_n<\frac{1}{n}. \tag{251}$$

For each fixed n, choose σ_n sufficiently small that, whenever $0<\sigma\le\sigma_n$, all the fixed-n estimates from Step 2 apply and the following additional bounds hold.

First, the available-profile approximation at the fixed entrance time η_n satisfies

$$\max_{\substack{(k,i)\in\mathcal{E}_m\\ a\in[\nu]}}\left|y^{(a)}_{ki,\sigma}(\eta_n)-\zeta^{(a)}_{ki}(\eta_n)\right|\le\frac{1}{n}. \tag{252}$$

Second, the balance error of the previously captured atoms satisfies

$$\max_{\substack{r\le m\\ a\in[\nu]}} \left| \left| q_{r,\sigma}^{(a)}(\eta_n) \right| - \lambda_{i_r}^{1/\nu} \right| \le \frac{1}{n}. \tag{253}$$

We also absorb the possibly n-dependent constants in the claimed-sector and old-background estimates by requiring

$$C_n^{\mathrm{cl}} \sigma_n^{\nu} \le \frac{1}{n} \tag{254}$$

and

$$C_n^{\mathrm{old}} \sigma_n^{\nu-1} \le 1. \tag{255}$$

Finally, we require the entrance level to grow sufficiently slowly relative to σ_n:

$$\sigma_n R_n^n \le \frac{1}{n}. \tag{256}$$

All these requirements can be imposed simultaneously because, at this stage of the construction, n, R_n, η_n, C_n^{cl}, and C_n^{old} are fixed and only σ_n is being made smaller.

For

$$\sigma_{n+1} < \sigma \le \sigma_n, \tag{257}$$

define

$$R_{m,\sigma} := R_n, \qquad \eta_{m+1}^{\sigma} := \eta_n. \tag{258}$$

Since the block index n tends to infinity as $\sigma \to 0$, we immediately obtain

$$R_{m,\sigma} \longrightarrow \infty. \tag{259}$$

Now fix $q > 0$. For all sufficiently large n, we have $n \ge q$, and since $R_n > 1$,

$$\sigma R_{m,\sigma}^q \le \sigma_n R_n^q \le \sigma_n R_n^n \le \frac{1}{n}. \tag{260}$$

Thus

$$\sigma R_{m,\sigma}^q \longrightarrow 0 \qquad \text{for every fixed } q > 0, \tag{261}$$

which proves the mesoscopic-scale condition (229). In particular, with $r_{m,\sigma} = \sigma R_{m,\sigma}$,

$$\frac{r_{m,\sigma}}{\sigma} = R_{m,\sigma} \longrightarrow \infty, \qquad r_{m,\sigma} = \sigma R_{m,\sigma} \longrightarrow 0. \tag{262}$$

This is precisely the scale separation (231).

Step 4: location of the entrance time. If σ belongs to the nth diagonal block (257), then $\eta_{m+1}^{\sigma} = \eta_n$ and $R_{m,\sigma} = R_n$. Therefore the fixed-level clock estimate (243) gives

$$\left| R_{m,\sigma}^{\nu-2}\big(\theta_* - \eta_{m+1}^{\sigma}\big) - \frac{1}{(\nu-2)\lambda_{i_*}} \right| \le \frac{1}{n}. \tag{263}$$

Since $n \to \infty$ as $\sigma \to 0$, this proves the precise entrance-time asymptotic (233). Since $R_{m,\sigma} \to \infty$, the same relation also gives

$$\theta_* - \eta_{m+1}^{\sigma} \longrightarrow 0, \tag{264}$$

and hence the convergence of the entrance time (232).

Step 5: the winning main coordinates. For σ in the nth diagonal block, the diagonal profile estimate (252) gives

$$\max_{a\in[\nu]} \left| y_{k_* i_*,\sigma}^{(a)}(\eta_{m+1}^{\sigma}) - \zeta_{k_* i_*}^{(a)}(\eta_n) \right| \le \frac{1}{n}. \tag{265}$$

Using $x^{(a)}_{k_*i_*,\sigma} = \sigma y^{(a)}_{k_*i_*,\sigma}$ and $r_{m,\sigma} = \sigma R_n$, we obtain

$$\begin{aligned}\left|\frac{x^{(a)}_{k_*i_*,\sigma}(\eta^\sigma_{m+1})}{r_{m,\sigma}} - \omega^{(a)}_*\right| &\le \frac{1}{R_n}\left|y^{(a)}_{k_*i_*,\sigma}(\eta_n) - \zeta^{(a)}_{k_*i_*}(\eta_n)\right| \\ &\quad + \left|\frac{\zeta^{(a)}_{k_*i_*}(\eta_n)}{R_n} - \omega^{(a)}_*\right| \\ &\le \frac{1}{nR_n} + \frac{1}{n}.\end{aligned} \tag{266}$$

The right-hand side tends to zero. This proves the winning-coordinate asymptotic (234). The phase condition (235) is exactly the phase-product conclusion of Lemma 5 applied to the winning candidate.

Step 6: uniform control of the winning off-sector coordinates. We next show that all microscopic coordinates of the winning atom outside sector i_* remain uniformly bounded. This is where the diagonal control of the n-dependent constants from Step 3 is needed.

First suppose that $i \notin \mathcal{S}_m$ and $i \neq i_*$. Then $(k_*, i) \in \mathcal{E}_m$ is a nonwinning available edge. By the positive clock gap (222), its isolated profile is regular up to θ_*, and the nonwinner bound (223) gives

$$\left|\zeta^{(a)}_{k_*i}(\eta_n)\right| \le B^{\mathrm{nw}}_m. \tag{267}$$

Combining this bound with the diagonal profile approximation (252) gives

$$\left|y^{(a)}_{k_*i,\sigma}(\eta^\sigma_{m+1})\right| \le B^{\mathrm{nw}}_m + 1 \tag{268}$$

for all sufficiently small σ.

Now suppose that $i \in \mathcal{S}_m$. At the reset time, the blocked-sector part of the induction hypothesis (85) gives

$$\left|y^{(a)}_{k_*i,\sigma}(\rho^\sigma_m)\right| \le C_m. \tag{269}$$

If σ belongs to the nth diagonal block, the fixed-level near-freezing estimate (248) and the diagonal choice (254) imply

$$\left|y^{(a)}_{k_*i,\sigma}(\eta_n) - y^{(a)}_{k_*i,\sigma}(\rho^\sigma_m)\right| \le C^{\mathrm{cl}}_n \sigma^\nu \le \frac{1}{n}. \tag{270}$$

Together with the reset bound (269), this gives

$$\left|y^{(a)}_{k_*i,\sigma}(\eta^\sigma_{m+1})\right| \le C_m + 1. \tag{271}$$

Combining the unclaimed-sector estimate (268) and the claimed-sector estimate (271), there is a constant $C^{\mathrm{off}}_m < \infty$, independent of σ, such that

$$\max_{\substack{i \neq i_* \\ a \in [\nu]}} \left|y^{(a)}_{k_*i,\sigma}(\eta^\sigma_{m+1})\right| \le C^{\mathrm{off}}_m. \tag{272}$$

Since the sector vectors $\{\mathbf{e}^{(a)}_i\}_{i=1}^p$ form an orthonormal basis,

$$\begin{aligned}&\left\|\mathbf{u}^{(a)}_{k_*,\sigma}(\eta^\sigma_{m+1}) - x^{(a)}_{k_*i_*,\sigma}(\eta^\sigma_{m+1})\mathbf{e}^{(a)}_{i_*}\right\|^2 \\ &\qquad = \sum_{i \neq i_*} \left|x^{(a)}_{k_*i,\sigma}(\eta^\sigma_{m+1})\right|^2 \le (p-1)\big(C^{\mathrm{off}}_m\big)^2\sigma^2.\end{aligned} \tag{273}$$

Thus the winning off-sector component is $O(\sigma)$ uniformly in σ. Since $\sigma/r_{m,\sigma} = R^{-1}_{m,\sigma} \to 0$, it is also $o(r_{m,\sigma})$ relative to the winning main coordinates.

Step 7: uniform control of the other surviving atoms. Let $k \notin \mathcal{A}_m$ with $k \neq k_*$. If $i \notin \mathcal{S}_m$, then (k,i) is an available edge and, because $k \neq k_*$, it is not the winning edge. The same nonwinner-profile argument used in (267)–(268) therefore gives

$$\left| y^{(a)}_{ki,\sigma}(\eta^\sigma_{m+1}) \right| \le B^{\mathrm{nw}}_m + 1. \tag{274}$$

If instead $i \in \mathcal{S}_m$, the reset bound (85), the fixed-level near-freezing estimate (248), and the diagonal choice (254) give

$$\left| y^{(a)}_{ki,\sigma}(\eta^\sigma_{m+1}) \right| \le C_m + 1. \tag{275}$$

Consequently, for a constant $C^{\mathrm{surv}}_m < \infty$ independent of σ,

$$\max_{\substack{k \notin \mathcal{A}_m,\ k \neq k_* \\ i \in [p],\ a \in [\nu]}} \left| y^{(a)}_{ki,\sigma}(\eta^\sigma_{m+1}) \right| \le C^{\mathrm{surv}}_m. \tag{276}$$

Using orthonormality of the sector basis once again,

$$\left\| \mathbf{u}^{(a)}_{k,\sigma}(\eta^\sigma_{m+1}) \right\| \le \sqrt{p}\, C^{\mathrm{surv}}_m \sigma \tag{277}$$

uniformly over $k \notin \mathcal{A}_m$, $k \neq k_*$, and $a \in [\nu]$. Thus every surviving atom other than the winner is still microscopic at the entrance time.

Step 8: the previously captured background. Suppose that σ belongs to the nth diagonal block. The quantitative fixed-level locking estimate (249) gives

$$\begin{aligned} &\max_{r \le m} \left| \pi_{r,\sigma}(\eta^\sigma_{m+1}) - \lambda_{i_r} \right| \\ &\qquad + \max_{\substack{r \le m \\ a \in [\nu]}} \left\| \mathbf{h}^{(a)}_{r,\sigma}(\eta^\sigma_{m+1}) \right\| \le C^{\mathrm{old}}_n \sigma^\nu. \end{aligned} \tag{278}$$

Since $\sigma \le \sigma_n$, the diagonal condition (255) implies

$$C^{\mathrm{old}}_n \sigma^\nu = \sigma C^{\mathrm{old}}_n \sigma^{\nu-1} \le \sigma C^{\mathrm{old}}_n \sigma_n^{\nu-1} \le \sigma. \tag{279}$$

Thus the product and off-sector errors of the old captured atoms satisfy

$$\max_{r \le m} \left| \pi_{r,\sigma}(\eta^\sigma_{m+1}) - \lambda_{i_r} \right| + \max_{\substack{r \le m \\ a \in [\nu]}} \left\| \mathbf{h}^{(a)}_{r,\sigma}(\eta^\sigma_{m+1}) \right\| \le \sigma. \tag{280}$$

For their main-coordinate magnitudes, the diagonal balance estimate (253) gives

$$\max_{\substack{r \le m \\ a \in [\nu]}} \left| \left| q^{(a)}_{r,\sigma}(\eta^\sigma_{m+1}) \right| - \lambda_{i_r}^{1/\nu} \right| \le \frac{1}{n}. \tag{281}$$

Since the diagonal block index n tends to infinity with $\sigma \to 0$, this proves the old-background balance conclusion (239).

We can now choose one constant for the three quantitative entrance bounds. Let C^{off}_m and C^{surv}_m be the constants from Steps 6 and 7 and set

$$C^{\mathrm{ent}}_m := \max\left\{ \sqrt{p}\, C^{\mathrm{off}}_m, \sqrt{p}\, C^{\mathrm{surv}}_m, 1 \right\}. \tag{282}$$

Then the winning localization estimate (273), the spectator norm bound (277), and the old-background locking estimate (280) yield, respectively, (236), (237), and (238).

Step 9: tensor output at the entrance time. It remains to prove the tensor-level conclusion (240).

For a previously captured atom k_r, the decomposition (71) reads at the entrance time

$$\mathbf{u}^{(a)}_{k_r,\sigma} = q^{(a)}_{r,\sigma} \mathbf{e}^{(a)}_{i_r} + \mathbf{h}^{(a)}_{r,\sigma}. \tag{283}$$

The term in its tensor product containing no off-sector component is

$$\pi_{r,\sigma} \mathbf{e}^{(1)}_{i_r} \otimes \cdots \otimes \mathbf{e}^{(\nu)}_{i_r}. \tag{284}$$

By the old-background product estimate (280), this leading term converges to

$$\lambda_{i_r} \mathbf{e}_{i_r}^{(1)} \otimes \cdots \otimes \mathbf{e}_{i_r}^{(\nu)}. \tag{285}$$

Every remaining term in the multilinear expansion contains at least one off-sector factor $\mathbf{h}_{r,\sigma}^{(a)}$. These factors are $O(\sigma)$ by (280), while the balance estimate (281) keeps all main coordinates $q_{r,\sigma}^{(a)}$ uniformly bounded. Consequently,

$$\bigotimes_{a=1}^{\nu} \mathbf{u}_{k_r,\sigma}^{(a)}(\eta_{m+1}^{\sigma}) \longrightarrow \lambda_{i_r} \mathbf{e}_{i_r}^{(1)} \otimes \cdots \otimes \mathbf{e}_{i_r}^{(\nu)}. \tag{286}$$

For the winning atom, its mesoscopic main-coordinate estimate (234) and off-sector localization (236) imply

$$\max_{a\in[\nu]} \left\| \mathbf{u}_{k_*,\sigma}^{(a)}(\eta_{m+1}^{\sigma}) \right\| = O(r_{m,\sigma}), \tag{287}$$

because $\sigma = o(r_{m,\sigma})$. Therefore the tensor represented by the winner is still vanishing at the entrance:

$$\left\| \bigotimes_{a=1}^{\nu} \mathbf{u}_{k_*,\sigma}^{(a)}(\eta_{m+1}^{\sigma}) \right\|_F = O(r_{m,\sigma}^{\nu}) = o(1). \tag{288}$$

Finally, every other surviving atom is $O(\sigma)$ in every mode by the spectator estimate (237). Hence

$$\left\| \bigotimes_{a=1}^{\nu} \mathbf{u}_{k,\sigma}^{(a)}(\eta_{m+1}^{\sigma}) \right\|_F = O(\sigma^{\nu}) \tag{289}$$

for $k \notin \mathcal{A}_m \cup \{k_*\}$.

Since H is fixed, summing the old captured-atom limits (286), the vanishing winning-atom contribution (288), and the microscopic spectator contributions (289) gives

$$\widehat{T}_{\sigma}(\eta_{m+1}^{\sigma}) \longrightarrow \sum_{r=1}^{m} \lambda_{i_r} \mathbf{e}_{i_r}^{(1)} \otimes \cdots \otimes \mathbf{e}_{i_r}^{(\nu)} = T_m. \tag{290}$$

This proves the tensor-level entrance statement (240) and completes the proof. □

The entrance state now has precisely the scale separation needed for the fast transition. The winning atom has factor size $r_{m,\sigma}$, with $r_{m,\sigma} \to 0$ but $r_{m,\sigma}/\sigma = R_{m,\sigma} \to \infty$, and its mass is concentrated asymptotically in sector i_*. Its main coordinates are balanced and phase aligned, whereas every other surviving atom remains only $O(\sigma)$. The previously captured atoms remain asymptotically locked to their assigned sectors.

Moreover, the entrance-time asymptotic (233) shows that the remaining slow time to the limiting clock is of order $R_{m,\sigma}^{-(\nu-2)} \to 0$. The next subsection analyzes the ensuing fast growth and saturation of the winning atom inside this vanishing slow-time layer and constructs the next reset time ρ_{m+1}^{σ}.

A3.5 Fast capture and renewal of the induction hypotheses

We now analyze the shrinking slow-time layer following the entrance time η_{m+1}^{σ} constructed in Proposition 8. Throughout this subsection, m is fixed, and we write

$$k_* := k_{m+1}, \qquad i_* := i_{m+1}, \qquad \lambda_* := \lambda_{i_*}, \qquad \theta_* := \theta_{m+1}^{*}. \tag{291}$$

For readability, we abbreviate the stage-dependent mesoscopic scales from Proposition 8 by

$$R_{\sigma} := R_{m,\sigma}, \qquad r_{\sigma} := r_{m,\sigma} = \sigma R_{\sigma}. \tag{292}$$

Thus

$$R_{\sigma} \longrightarrow \infty, \qquad \sigma \ll r_{\sigma} \ll 1, \qquad \sigma R_{\sigma}^{q} \longrightarrow 0 \quad \text{for every fixed } q > 0. \tag{293}$$

The fast transition is most naturally studied in the original physical-time scale. We therefore introduce a local physical time $\xi \geq 0$ by

$$\Theta_\sigma(\xi) := \eta_{m+1}^\sigma + \sigma^{\nu-2}\xi. \tag{294}$$

Thus $\Theta_\sigma(\xi)$ is the slow time corresponding to a physical-time increment ξ after the entrance. Whenever quantities associated with the previously captured atoms are written below as functions of ξ, they are understood to be evaluated at the slow time $\Theta_\sigma(\xi)$.

For the winning atom, define its main coordinates and their product by

$$q_\sigma^{(a)}(\xi) := x^{(a)}_{k_* i_*,\sigma}\big(\Theta_\sigma(\xi)\big), \qquad \pi_\sigma(\xi) := \prod_{a=1}^{\nu} q_\sigma^{(a)}(\xi), \tag{295}$$

and its off-sector component by

$$\mathbf{h}_\sigma^{(a)}(\xi) := \mathbf{u}^{(a)}_{k_*,\sigma}\big(\Theta_\sigma(\xi)\big) - q_\sigma^{(a)}(\xi)\mathbf{e}^{(a)}_{i_*}. \tag{296}$$

At $\xi = 0$, the mesoscopic entrance estimate (234) gives

$$\frac{q_\sigma^{(a)}(0)}{r_\sigma} \longrightarrow \omega_*^{(a)}, \qquad |\omega_*^{(a)}| = 1, \qquad \prod_{a=1}^{\nu} \omega_*^{(a)} = 1, \tag{297}$$

while the winning-sector localization (236) gives

$$\max_{a\in[\nu]} \big\|\mathbf{h}_\sigma^{(a)}(0)\big\| = O(\sigma). \tag{298}$$

The capture is divided into two stages. During the first stage, the winning coordinates grow from the mesoscopic scale r_σ to a fixed small macroscopic amplitude. During the second stage, all modewise factors of the winner are bounded away from zero, and the vector field becomes uniformly contracting toward the captured equilibrium manifold.

Growth from the mesoscopic scale. Choose once and for all a constant $a_0 > 0$ sufficiently small that

$$4^\nu a_0^\nu \leq \frac{\lambda_*}{4}. \tag{299}$$

The numerical constant is unimportant. Its role is to ensure that, while the winning factors have magnitude at most a fixed multiple of a_0, the residual coefficient $\lambda_* - \pi_\sigma$ still drives the winning sector strongly in the positive-growth direction.

Define the average squared magnitude and the maximal squared-magnitude imbalance of the winning coordinates by

$$Q_\sigma(\xi) := \frac{1}{\nu}\sum_{a=1}^{\nu} |q_\sigma^{(a)}(\xi)|^2, \qquad D_\sigma(\xi) := \max_{a,b\in[\nu]} \Big| |q_\sigma^{(a)}(\xi)|^2 - |q_\sigma^{(b)}(\xi)|^2 \Big|. \tag{300}$$

Lemma 6 (Growth to a fixed macroscopic scale)**.** *Under the assumptions of Proposition 8, define*

$$E_{\mathrm{old},\sigma}(\xi) := \max_{r\leq m} \big|\pi_{r,\sigma}\big(\Theta_\sigma(\xi)\big) - \lambda_{i_r}\big| + \max_{\substack{r\leq m\\ a\in[\nu]}} \Big\|\mathbf{h}_{r,\sigma}^{(a)}\big(\Theta_\sigma(\xi)\big)\Big\|. \tag{301}$$

For $m = 0$, this quantity is understood to be identically zero.

Then, for all sufficiently small σ, there exists a finite first hitting time

$$\xi_{\mathrm{grow}}^\sigma := \inf\big\{\xi > 0 : Q_\sigma(\xi) = a_0^2\big\}, \tag{302}$$

satisfying

$$\xi_{\mathrm{grow}}^\sigma \leq C_{\mathrm{grow}} r_\sigma^{-(\nu-2)}. \tag{303}$$

At the hitting time,

$$\frac{a_0}{2} \leq |q_\sigma^{(a)}(\xi_{\mathrm{grow}}^\sigma)| \leq 2a_0, \qquad a \in [\nu]. \tag{304}$$

Moreover, throughout $0 \le \xi \le \xi^{\sigma}_{\mathrm{grow}}$,

$$\max_{a\in[\nu]} \left\| \mathbf{h}^{(a)}_{\sigma}(\xi) \right\| \le C_{\mathrm{grow}}\sigma, \tag{305}$$

and

$$\max_{\substack{k\notin \mathcal{A}_m\cup\{k_*\}\\ a\in[\nu]}} \left\| \mathbf{u}^{(a)}_{k,\sigma}\big(\Theta_\sigma(\xi)\big) \right\| \le C_{\mathrm{grow}}\sigma. \tag{306}$$

The previously captured background satisfies

$$\sup_{0\le\xi\le\xi^{\sigma}_{\mathrm{grow}}} E_{\mathrm{old},\sigma}(\xi) \le C_{\mathrm{grow}}\sigma, \tag{307}$$

and the exact balance invariant gives the stronger estimate

$$\sup_{0\le\xi\le\xi^{\sigma}_{\mathrm{grow}}} D_\sigma(\xi) = O(\sigma^2) = o(r_\sigma^2). \tag{308}$$

Finally, the winning product remains asymptotically phase aligned:

$$\sup_{0\le\xi\le\xi^{\sigma}_{\mathrm{grow}}} \left| \frac{\pi_\sigma(\xi)}{|\pi_\sigma(\xi)|} - 1 \right| \longrightarrow 0. \tag{309}$$

Proof. We use a stopping-time argument. The purpose of the bootstrap is to keep the winner inside a small macroscopic neighborhood while controlling its off-sector component, all spectator atoms, and the previously captured background. The exact balance invariant will then supply balance of the winner automatically.

Entrance estimates. Set

$$C_{\mathrm{ent}} := C_m^{\mathrm{ent}}, \tag{310}$$

where C_m^{ent} is the constant from Proposition 8. The quantitative old-background locking estimate (238) gives

$$E_{\mathrm{old},\sigma}(0) \le C_{\mathrm{ent}}\sigma. \tag{311}$$

The winning localization estimate (236) gives

$$\max_{a\in[\nu]} \|\mathbf{h}^{(a)}_{\sigma}(0)\| \le C_{\mathrm{ent}}\sigma, \tag{312}$$

while the spectator estimate (237) gives

$$\max_{\substack{k\notin \mathcal{A}_m\cup\{k_*\}\\ a\in[\nu]}} \left\| \mathbf{u}^{(a)}_{k,\sigma}\big(\Theta_\sigma(0)\big) \right\| \le C_{\mathrm{ent}}\sigma. \tag{313}$$

For the winning main coordinates, the entrance asymptotic (297) implies

$$Q_\sigma(0) = r_\sigma^2(1+o(1)), \tag{314}$$

and, using $\prod_a \omega_*^{(a)} = 1$,

$$\frac{\pi_\sigma(0)}{r_\sigma^{\nu}} \longrightarrow 1. \tag{315}$$

Finally, the balance conclusion for the old captured background (239) gives

$$\max_{\substack{r\le m\\ a\in[\nu]}} \left| \left| q^{(a)}_{r,\sigma}(0) \right| - \lambda_{i_r}^{1/\nu} \right| \longrightarrow 0. \tag{316}$$

Step 1: bootstrap interval. Choose constants

$$B_{\mathrm{fast}} > C_{\mathrm{ent}} + 2, \qquad 0 < \varepsilon_{\mathrm{old}} < 1, \tag{317}$$

where $\varepsilon_{\mathrm{old}}$ will be chosen sufficiently small below.

By the old-background balance at the entrance (316), we may fix

$$0 < c_{\mathrm{old}} < C_{\mathrm{old}} < \infty \tag{318}$$

such that every $\lambda_{i_r}^{1/\nu}$ lies strictly inside $(c_{\mathrm{old}}, C_{\mathrm{old}})$ and, for all sufficiently small σ,

$$2c_{\mathrm{old}} \le |q_{r,\sigma}^{(a)}(0)| \le \frac{1}{2} C_{\mathrm{old}}, \qquad r \le m, \quad a \in [\nu]. \tag{319}$$

Choose a fixed constant $K > 0$, to be specified in Step 9, and define the comparison horizon

$$T_{\mathrm{cmp},\sigma} := K r_\sigma^{-(\nu-2)}. \tag{320}$$

We consider the maximal interval $[0, \xi_{\mathrm{boot}}^\sigma]$, stopped no later than $T_{\mathrm{cmp},\sigma}$, on which either the level $Q_\sigma = a_0^2$ has already been reached or all of the following strict bootstrap conditions hold:

$$\frac{1}{2} Q_\sigma(0) < Q_\sigma(\xi) < a_0^2, \tag{321}$$

$$\operatorname{Re} \pi_\sigma(\xi) > \frac{1}{2} |\pi_\sigma(\xi)|, \tag{322}$$

$$\max_{a \in [\nu]} \|\mathbf{h}_\sigma^{(a)}(\xi)\| < B_{\mathrm{fast}} \sigma, \tag{323}$$

$$\max_{\substack{k \notin \mathcal{A}_m \cup \{k_*\} \\ a \in [\nu]}} \left\| \mathbf{u}_{k,\sigma}^{(a)}\big(\Theta_\sigma(\xi)\big) \right\| < B_{\mathrm{fast}} \sigma, \tag{324}$$

$$E_{\mathrm{old},\sigma}(\xi) < \varepsilon_{\mathrm{old}}, \tag{325}$$

$$c_{\mathrm{old}} < |q_{r,\sigma}^{(a)}(\xi)| < C_{\mathrm{old}}, \qquad r \le m, \quad a \in [\nu]. \tag{326}$$

The entrance estimates above show that all bootstrap conditions hold initially with strict slack for sufficiently small σ.

Step 2: stability of the previously captured background. We first improve the two bootstrap conditions concerning the previously captured atoms.

Define

$$P_\sigma(\xi) := \max_{r \le m} \left| \pi_{r,\sigma}\big(\Theta_\sigma(\xi)\big) - \lambda_{i_r} \right|, \qquad H_\sigma(\xi) := \max_{\substack{r \le m \\ a \in [\nu]}} \left\| \mathbf{h}_{r,\sigma}^{(a)}\big(\Theta_\sigma(\xi)\big) \right\|, \tag{327}$$

so that

$$E_{\mathrm{old},\sigma} = P_\sigma + H_\sigma. \tag{328}$$

Product stability. Fix a previously captured edge (k_r, i_r) and set

$$e_{r,\sigma} := \pi_{r,\sigma} - \lambda_{i_r}. \tag{329}$$

Repeating the captured-coordinate decomposition from Lemma 2, now in physical time, gives

$$\frac{dq_{r,\sigma}^{(a)}}{d\xi} = -e_{r,\sigma} \prod_{b \ne a} \overline{q_{r,\sigma}^{(b)}} + \mathcal{R}_{r,\sigma}^{(a)}. \tag{330}$$

Uniformly on the bootstrap interval,

$$\left| \mathcal{R}_{r,\sigma}^{(a)} \right| \le C \left(H_\sigma^2 + \sigma^\nu \right). \tag{331}$$

To see this, the exact self Gram factors satisfy

$$G_{k_r k_r, \sigma}^{(b)} = |q_{r,\sigma}^{(b)}|^2 + \|\mathbf{h}_{r,\sigma}^{(b)}\|^2, \tag{332}$$

so the error between the exact self-interaction and the pure one-sector interaction is $O(H_\sigma^2)$ under the old-annulus bootstrap (326). If $s \neq r$ is another old captured atom, orthogonality of the assigned sectors gives

$$|x^{(a)}_{k_s i_r,\sigma}| \le H_\sigma, \qquad |G^{(b)}_{k_r k_s,\sigma}| \le C H_\sigma, \tag{333}$$

and therefore an $O(H_\sigma^\nu)$ contribution.

For the interaction with the new winner, $i_* \neq i_r$. The winner off-sector bootstrap (323) and the boundedness $Q_\sigma < a_0^2$ give

$$|x^{(a)}_{k_* i_r,\sigma}| \le C\sigma, \qquad |G^{(b)}_{k_r k_*,\sigma}| \le C(\sigma + H_\sigma), \tag{334}$$

and hence an error bounded by $C(H_\sigma^2 + \sigma^\nu)$. Every other surviving atom is $O(\sigma)$ by the spectator bootstrap and contributes only $O(\sigma^\nu)$. This proves (331).

Differentiating the product $\pi_{r,\sigma} = \prod_a q^{(a)}_{r,\sigma}$ in the effective old-coordinate equation (330) gives

$$\frac{de_{r,\sigma}}{d\xi} = -S_{r,\sigma} e_{r,\sigma} + \mathcal{R}^\pi_{r,\sigma}, \tag{335}$$

where

$$S_{r,\sigma} := \sum_{a=1}^{\nu} \prod_{b \neq a} |q^{(b)}_{r,\sigma}|^2. \tag{336}$$

The old-annulus bootstrap gives the uniform coercivity

$$S_{r,\sigma} \ge \nu c_{\mathrm{old}}^{2(\nu-1)} =: c_\pi > 0, \tag{337}$$

while the remainder estimate (331) implies

$$|\mathcal{R}^\pi_{r,\sigma}| \le C\left(H_\sigma^2 + \sigma^\nu\right). \tag{338}$$

Taking upper Dini derivatives and maximizing over r therefore yields

$$D^+ P_\sigma \le -c_\pi P_\sigma + C\left(H_\sigma^2 + \sigma^\nu\right). \tag{339}$$

Transverse stability. For an old captured edge (k_r, i_r), let

$$\mathcal{P}_r^{(a),\perp} := I - \mathbf{e}^{(a)}_{i_r} \left(\mathbf{e}^{(a)}_{i_r}\right)^* \tag{340}$$

be orthogonal projection onto the complement of its assigned sector. By the definition of the off-sector residual,

$$\mathbf{h}^{(a)}_{r,\sigma} = \mathcal{P}_r^{(a),\perp} \mathbf{u}^{(a)}_{k_r,\sigma}. \tag{341}$$

Because the local variable ξ is physical time, applying $\mathcal{P}_r^{(a),\perp}$ to the exact GF equation gives

$$\frac{d\mathbf{h}^{(a)}_{r,\sigma}}{d\xi} = -\gamma^{(a)}_{r,\sigma} \mathbf{h}^{(a)}_{r,\sigma} + \mathbf{N}^{(a)}_{r,\sigma}, \tag{342}$$

where

$$\gamma^{(a)}_{r,\sigma} := \prod_{b \neq a} \left\| \mathbf{u}^{(b)}_{k_r,\sigma} \right\|^2 \ge c_{\mathrm{old}}^{2(\nu-1)} =: c_\gamma > 0. \tag{343}$$

The remainder $\mathbf{N}^{(a)}_{r,\sigma}$ consists exactly of the off-sector target terms and the interactions with atoms other than k_r. The same estimates used above give

$$\|\mathbf{N}^{(a)}_{r,\sigma}\| \le C\left(H_\sigma^{\nu-1} + \sigma^{\nu-1}\right). \tag{344}$$

Hence

$$D^+ H_\sigma \le -c_\gamma H_\sigma + C\left(H_\sigma^{\nu-1} + \sigma^{\nu-1}\right). \tag{345}$$

Set

$$c_0 := \min\{c_\pi, c_\gamma\} > 0. \tag{346}$$

Adding the product estimate (339) and the transverse estimate (345), and using $E_{\mathrm{old},\sigma} = P_\sigma + H_\sigma$, gives

$$D^+ E_{\mathrm{old},\sigma} \le -c_0 E_{\mathrm{old},\sigma} + C\left(H_\sigma^2 + H_\sigma^{\nu-1}\right) + C\sigma^{\nu-1}. \tag{347}$$

Choose $\varepsilon_{\mathrm{old}}$ in the bootstrap (325) sufficiently small that

$$C\left(\varepsilon_{\mathrm{old}} + \varepsilon_{\mathrm{old}}^{\nu-2}\right) \le \frac{1}{2}c_0. \tag{348}$$

Since $H_\sigma \le E_{\mathrm{old},\sigma} < \varepsilon_{\mathrm{old}}$, we obtain

$$D^+ E_{\mathrm{old},\sigma} \le -\frac{1}{2}c_0 E_{\mathrm{old},\sigma} + C\sigma^{\nu-1}. \tag{349}$$

Variation of constants, together with the entrance estimate (311), yields

$$E_{\mathrm{old},\sigma}(\xi) \le e^{-c_0\xi/2} E_{\mathrm{old},\sigma}(0) + C\sigma^{\nu-1} \le C'_{\mathrm{old}}\sigma. \tag{350}$$

Thus the old-background error bootstrap is strictly improved.

It remains to preserve the old captured annulus. Applying the exact main-coordinate balance consequence (75) to each old captured atom, and using the just-established bound $H_\sigma = O(\sigma)$, gives

$$\max_{\substack{r\le m\\ a,b\in[\nu]}} \sup_{0\le\xi\le\xi^\sigma_{\mathrm{boot}}} \left| |q^{(a)}_{r,\sigma}(\xi)|^2 - |q^{(b)}_{r,\sigma}(\xi)|^2 \right| = O(\sigma^2). \tag{351}$$

At the same time, the product part of (350) gives

$$\prod_{a=1}^{\nu} |q^{(a)}_{r,\sigma}(\xi)| = \lambda_{i_r} + O(\sigma) \tag{352}$$

uniformly on the bootstrap interval. Combining these two estimates yields

$$\max_{\substack{r\le m\\ a\in[\nu]}} \sup_{0\le\xi\le\xi^\sigma_{\mathrm{boot}}} \left| |q^{(a)}_{r,\sigma}(\xi)| - \lambda_{i_r}^{1/\nu} \right| = O(\sigma). \tag{353}$$

Since each $\lambda_{i_r}^{1/\nu}$ lies strictly inside $(c_{\mathrm{old}}, C_{\mathrm{old}})$, this strictly improves the old-annulus bootstrap (326).

Step 3: balance of the winning factors from the exact invariant. The winning off-sector bootstrap (323) and the exact balance relation (75), now applied to atom k_* and sector i_*, give

$$D_\sigma(\xi) \le \left(C_{\mathrm{bal}} + 2B^2_{\mathrm{fast}}\right)\sigma^2 \tag{354}$$

throughout the bootstrap interval.

On the other hand, the lower amplitude bootstrap (321) and the entrance asymptotic (314) give

$$Q_\sigma(\xi) \ge \frac{1}{2}Q_\sigma(0) = \frac{1}{2}r_\sigma^2(1+o(1)). \tag{355}$$

Therefore

$$\frac{D_\sigma(\xi)}{Q_\sigma(\xi)} \le \frac{C}{R_\sigma^2} \longrightarrow 0 \tag{356}$$

uniformly on the bootstrap interval. In particular, for sufficiently small σ,

$$\frac{3}{4}Q_\sigma \le |q^{(a)}_\sigma|^2 \le \frac{5}{4}Q_\sigma, \qquad a\in[\nu]. \tag{357}$$

This also already proves the balance conclusion (308), conditional only on closing the off-sector bootstrap later in the argument.

Step 4: the winning main-coordinate equation. Separating the self-interaction of the winning atom in the exact physical-time coordinate equation gives

$$\frac{dq_\sigma^{(a)}}{d\xi} = \big(\lambda_* - \pi_\sigma\big) \prod_{b\neq a} \overline{q_\sigma^{(b)}} + \mathcal{R}_\sigma^{(a)}(\xi). \tag{358}$$

Let

$$M_\sigma(\xi) := \max_{a\in[\nu]} |q_\sigma^{(a)}(\xi)|. \tag{359}$$

Using the winner off-sector bootstrap, the spectator bootstrap, and the old-background estimate (350), the remainder in (358) satisfies

$$\left|\mathcal{R}_\sigma^{(a)}(\xi)\right| \le C\left(\sigma^\nu + \sigma^2 M_\sigma(\xi)^{2\nu-3}\right). \tag{360}$$

Indeed, the correction to the winning self-interaction caused by the off-sector residual is

$$O\big(\sigma^2 M_\sigma^{2\nu-3}\big), \tag{361}$$

because $\|\mathbf{h}_\sigma^{(b)}\| = O(\sigma)$. An interaction with another surviving atom contains one $O(\sigma)$ coordinate and $\nu-1$ Gram factors of size $O(\sigma M_\sigma + \sigma^2)$ and is therefore $O(\sigma^\nu)$ while M_σ remains bounded. Finally, by the old-background bound (350), every interaction with a previously captured atom contains an $O(\sigma)$ wrong-sector coordinate and $\nu-1$ cross-Gram factors of size $O(\sigma)$, giving another $O(\sigma^\nu)$ contribution.

Step 5: positive growth of the winning amplitude. The balance estimate (357) implies

$$c_\nu Q_\sigma^{\nu/2} \le |\pi_\sigma| \le C_\nu Q_\sigma^{\nu/2}. \tag{362}$$

Averaging the squared-magnitude equations obtained from (358) gives

$$\frac{dQ_\sigma}{d\xi} = 2\,\mathrm{Re}\left[\big(\lambda_* - \pi_\sigma\big)\overline{\pi_\sigma}\right] + \mathcal{R}_\sigma^Q, \tag{363}$$

where the main-coordinate remainder estimate (360) gives

$$|\mathcal{R}_\sigma^Q| \le C M_\sigma\left(\sigma^\nu + \sigma^2 M_\sigma^{2\nu-3}\right). \tag{364}$$

Under the phase bootstrap (322),

$$\begin{aligned} 2\,\mathrm{Re}\left[(\lambda_* - \pi_\sigma)\overline{\pi_\sigma}\right] &= 2\left(\lambda_*\,\mathrm{Re}\,\pi_\sigma - |\pi_\sigma|^2\right) \\ &\ge \lambda_*|\pi_\sigma| - 2|\pi_\sigma|^2. \end{aligned} \tag{365}$$

The amplitude bootstrap $Q_\sigma < a_0^2$, the factor comparability (357), and the fixed choice of a_0 in (299) imply

$$|\pi_\sigma| \le \frac{\lambda_*}{4}. \tag{366}$$

Hence

$$2\,\mathrm{Re}\left[(\lambda_* - \pi_\sigma)\overline{\pi_\sigma}\right] \ge \frac{\lambda_*}{2}|\pi_\sigma|. \tag{367}$$

The lower amplitude bound (355) and factor comparability give

$$M_\sigma \ge c r_\sigma. \tag{368}$$

Therefore

$$\frac{|\mathcal{R}_\sigma^Q|}{|\pi_\sigma|} \le C\left(\frac{\sigma^\nu}{M_\sigma^{\nu-1}} + \sigma^2 M_\sigma^{\nu-2}\right) \le C\left(\frac{\sigma}{R_\sigma^{\nu-1}} + \sigma^2\right) \longrightarrow 0. \tag{369}$$

For sufficiently small σ, the positive leading term therefore dominates the perturbation, and

$$\frac{dQ_\sigma}{d\xi} \ge c_{\mathrm{grow}} Q_\sigma^{\nu/2}. \tag{370}$$

In particular, Q_σ is strictly increasing, so

$$Q_\sigma(\xi) \ge Q_\sigma(0), \tag{371}$$

which strictly improves the lower bound in the amplitude bootstrap (321).

Step 6: phase alignment is preserved. Differentiating the winning product $\pi_\sigma = \prod_a q_\sigma^{(a)}$ in the effective main-coordinate equation (358) gives

$$\frac{d\pi_\sigma}{d\xi} = S_\sigma(\xi)\big(\lambda_* - \pi_\sigma\big) + \mathcal{E}_\sigma^\pi(\xi), \tag{372}$$

where

$$S_\sigma := \sum_{a=1}^{\nu} \prod_{b\neq a} |q_\sigma^{(b)}|^2 \tag{373}$$

satisfies, by factor comparability,

$$c_\nu Q_\sigma^{\nu-1} \le S_\sigma \le C_\nu Q_\sigma^{\nu-1}. \tag{374}$$

The product remainder obeys

$$|\mathcal{E}_\sigma^\pi| \le C M_\sigma^{\nu-1}\left(\sigma^\nu + \sigma^2 M_\sigma^{2\nu-3}\right). \tag{375}$$

Using the lower bound $M_\sigma \ge c r_\sigma$ from (368), we obtain

$$\frac{|\mathcal{E}_\sigma^\pi|}{\lambda_* S_\sigma} \le C\left(\frac{\sigma^\nu}{M_\sigma^{\nu-1}} + \sigma^2 M_\sigma^{\nu-2}\right) =: \varepsilon_\sigma, \qquad \varepsilon_\sigma \longrightarrow 0. \tag{376}$$

Define

$$A_\sigma(\xi) := \int_0^\xi S_\sigma(s)\, ds. \tag{377}$$

Variation of constants in the product equation (372) gives

$$\begin{aligned}\pi_\sigma(\xi) = e^{-A_\sigma(\xi)}\pi_\sigma(0) + \lambda_*\big(1 - e^{-A_\sigma(\xi)}\big) \\ + \int_0^\xi e^{-[A_\sigma(\xi)-A_\sigma(s)]}\mathcal{E}_\sigma^\pi(s)\, ds.\end{aligned} \tag{378}$$

The relative remainder bound (376) implies

$$\left|\int_0^\xi e^{-[A_\sigma(\xi)-A_\sigma(s)]}\mathcal{E}_\sigma^\pi(s)\, ds\right| \le \varepsilon_\sigma \lambda_*\big(1 - e^{-A_\sigma(\xi)}\big). \tag{379}$$

By the entrance product asymptotic (315), there exists $\delta_\sigma \to 0$ such that

$$|\pi_\sigma(0) - r_\sigma^\nu| \le \delta_\sigma r_\sigma^\nu. \tag{380}$$

Define the positive real quantity

$$B_\sigma(\xi) := e^{-A_\sigma(\xi)} r_\sigma^\nu + \lambda_*\big(1 - e^{-A_\sigma(\xi)}\big). \tag{381}$$

Then (378)–(380) give

$$|\pi_\sigma(\xi) - B_\sigma(\xi)| \le \big(\delta_\sigma + \varepsilon_\sigma\big) B_\sigma(\xi). \tag{382}$$

Since $B_\sigma(\xi) > 0$, we conclude that

$$\sup_{0\le\xi\le\xi_{\mathrm{boot}}^\sigma} \left|\frac{\pi_\sigma(\xi)}{|\pi_\sigma(\xi)|} - 1\right| \longrightarrow 0. \tag{383}$$

For sufficiently small σ, this strictly improves the phase bootstrap (322).

Step 7: the winning off-sector component remains $O(\sigma)$. Let

$$\mathcal{P}_*^{(a),\perp} := I - \mathbf{e}_{i_*}^{(a)}\big(\mathbf{e}_{i_*}^{(a)}\big)^*. \tag{384}$$

By the definition of the winning off-sector residual (296),

$$\mathbf{h}_\sigma^{(a)} = \mathcal{P}_*^{(a),\perp}\mathbf{u}_{k_*,\sigma}^{(a)}. \tag{385}$$

Applying this projector to the exact physical-time GF equation gives

$$\frac{d\mathbf{h}_{\sigma}^{(a)}}{d\xi} = -\gamma_{*,\sigma}^{(a)}\mathbf{h}_{\sigma}^{(a)} + \mathbf{N}_{*,\sigma}^{(a)}, \tag{386}$$

where

$$\gamma_{*,\sigma}^{(a)} := \prod_{b\neq a} \left\| \mathbf{u}_{k_*,\sigma}^{(b)} \right\|^2 \geq 0. \tag{387}$$

The remainder contains the off-sector target terms and all interactions with atoms other than k_*. Under the winner off-sector bootstrap (323), the spectator bootstrap (324), and the old-background bound (350), every such term is $O(\sigma^{\nu-1})$. Hence

$$\max_{a\in[\nu]} \|\mathbf{N}_{*,\sigma}^{(a)}\| \leq C\sigma^{\nu-1}. \tag{388}$$

Dropping the nonpositive self-damping term from the norm inequality gives

$$D^+\left(\max_{a\in[\nu]} \|\mathbf{h}_{\sigma}^{(a)}\|\right) \leq C\sigma^{\nu-1}. \tag{389}$$

Using the entrance bound (312) and the comparison horizon (320),

$$\max_{a\in[\nu]} \|\mathbf{h}_{\sigma}^{(a)}(\xi)\| \leq C_{\mathrm{ent}}\sigma + C\sigma^{\nu-1}T_{\mathrm{cmp},\sigma}. \tag{390}$$

The mesoscopic scale gives

$$\sigma^{\nu-1}T_{\mathrm{cmp},\sigma} = K\frac{\sigma}{R_{\sigma}^{\nu-2}} = o(\sigma). \tag{391}$$

Therefore

$$\max_{a\in[\nu]} \sup_{0\leq\xi\leq\xi_{\mathrm{boot}}^{\sigma}} \|\mathbf{h}_{\sigma}^{(a)}(\xi)\| \leq \big(C_{\mathrm{ent}} + o(1)\big)\sigma, \tag{392}$$

which strictly improves the winner off-sector bootstrap (323). In particular, the balance estimate derived conditionally in Step 3 is now valid throughout the full bootstrap interval.

Step 8: the other surviving atoms remain microscopic. Let $k \notin \mathcal{A}_m \cup \{k_*\}$. The target part of its physical-time GF is $O(\sigma^{\nu-1})$ under the spectator bootstrap. Interactions with other microscopic atoms are smaller, while interactions with an old captured atom contain $\nu - 1$ Gram factors of size $O(\sigma)$ and are also $O(\sigma^{\nu-1})$.

The only potentially larger interaction is with the growing winner:

$$\mathbf{u}_{k_*,\sigma}^{(a)} \prod_{b\neq a} G_{kk_*,\sigma}^{(b)}. \tag{393}$$

Since the spectator atom is $O(\sigma)$ and the winner has main magnitude M_{σ} and off-sector part $O(\sigma)$,

$$\left| G_{kk_*,\sigma}^{(b)} \right| \leq C\sigma M_{\sigma}. \tag{394}$$

Therefore

$$\left\| \mathbf{u}_{k_*,\sigma}^{(a)} \prod_{b\neq a} G_{kk_*,\sigma}^{(b)} \right\| \leq C\sigma^{\nu-1}M_{\sigma}^{\nu} \leq C\sigma^{\nu-1}, \tag{395}$$

because M_{σ} remains bounded while $Q_{\sigma} < a_0^2$.

Hence

$$\max_{\substack{k\notin\mathcal{A}_m\cup\{k_*\}\\ a\in[\nu]}} \left\| \frac{d}{d\xi}\mathbf{u}_{k,\sigma}^{(a)}\big(\Theta_{\sigma}(\xi)\big) \right\| \leq C\sigma^{\nu-1}. \tag{396}$$

Integrating over the comparison horizon and using the entrance estimate (313) gives

$$\max_{\substack{k\notin\mathcal{A}_m\cup\{k_*\}\\ a\in[\nu]}} \left\| \mathbf{u}_{k,\sigma}^{(a)}\big(\Theta_{\sigma}(\xi)\big) \right\| \leq \big(C_{\mathrm{ent}} + o(1)\big)\sigma. \tag{397}$$

This strictly improves the spectator bootstrap (324).

Step 9: existence of the hitting time. Before the level $Q_\sigma = a_0^2$ is reached, Steps 2–8 strictly improve all auxiliary bootstrap conditions for sufficiently small σ. Therefore the bootstrap interval cannot terminate except by reaching the level $Q_\sigma = a_0^2$ or the comparison horizon $T_{\mathrm{cmp},\sigma}$.

The growth inequality (370) gives

$$\frac{d}{d\xi} Q_\sigma(\xi)^{-(\nu-2)/2} \le -\frac{\nu-2}{2} c_{\mathrm{grow}}. \tag{398}$$

Consequently, as long as $Q_\sigma < a_0^2$,

$$Q_\sigma(\xi)^{-(\nu-2)/2} \le Q_\sigma(0)^{-(\nu-2)/2} - \frac{\nu-2}{2} c_{\mathrm{grow}} \xi. \tag{399}$$

The target level must therefore be reached no later than

$$\frac{2}{(\nu-2)c_{\mathrm{grow}}} \left[Q_\sigma(0)^{-(\nu-2)/2} - a_0^{-(\nu-2)} \right]. \tag{400}$$

Since the entrance amplitude satisfies $Q_\sigma(0) = r_\sigma^2(1+o(1))$, there is a constant $C_{\mathrm{grow}} < \infty$ such that this upper bound is at most

$$C_{\mathrm{grow}} r_\sigma^{-(\nu-2)} \tag{401}$$

for sufficiently small σ. Choosing $K > C_{\mathrm{grow}}$ in the comparison horizon (320) shows that the level $Q_\sigma = a_0^2$ is reached before the horizon. This proves the existence and time estimate (302)–(303).

Step 10: state at the hitting time. At $\xi = \xi_{\mathrm{grow}}^\sigma$,

$$Q_\sigma(\xi_{\mathrm{grow}}^\sigma) = a_0^2. \tag{402}$$

The exact-invariant balance estimate (354) gives

$$D_\sigma(\xi_{\mathrm{grow}}^\sigma) = O(\sigma^2) = o(1). \tag{403}$$

Because Q_σ is the average of the squared factor magnitudes,

$$\max_{a\in[\nu]} \left| |q_\sigma^{(a)}(\xi_{\mathrm{grow}}^\sigma)|^2 - a_0^2 \right| \le D_\sigma(\xi_{\mathrm{grow}}^\sigma) = o(1). \tag{404}$$

Hence, for sufficiently small σ, (304) holds.

The improved winner off-sector estimate (392), the spectator estimate (397), and the old-background estimate (350) give (305), (306), and (307), respectively. The exact-invariant bound (354) gives (308), and the phase estimate (383) gives (309). This completes the proof. □

Uniform locking after the winning factors become macroscopic. Once the winning factors satisfy (304), their magnitudes are bounded away from zero. The new captured sector therefore enters the same uniformly coercive regime as the previously captured sectors. The next lemma records the contraction needed to reach the quantitative $O(\sigma^\nu)$ locking scale required by the induction hypothesis.

Lemma 7 (Uniform locking of the enlarged captured set)**.** *Suppose that at a physical time ξ_0^σ the following conditions hold.*

1. *The first m captured atoms satisfy*

$$\max_{r\le m} |\pi_{r,\sigma} - \lambda_{i_r}| + \max_{\substack{r\le m \\ a\in[\nu]}} \|\mathbf{h}_{r,\sigma}^{(a)}\| = o(1), \tag{405}$$

   and their main-coordinate magnitudes converge to $\lambda_{i_r}^{1/\nu}$.

2. *The new winning atom satisfies*

$$c_0 \le |q_\sigma^{(a)}(\xi_0^\sigma)| \le C_0, \qquad a \in [\nu], \tag{406}$$

   for fixed $0 < c_0 < C_0 < \infty$, and its product is asymptotically aligned with the positive real axis:

$$\frac{\pi_\sigma(\xi_0^\sigma)}{|\pi_\sigma(\xi_0^\sigma)|} \longrightarrow 1. \tag{407}$$

3. The new winner has off-sector residual $O(\sigma)$:

$$\max_{a\in[\nu]} \|\mathbf{h}_\sigma^{(a)}(\xi_0^\sigma)\| = O(\sigma). \tag{408}$$

4. Every atom whose index is outside $\mathcal{A}_m \cup \{k_*\}$ is $O(\sigma)$.

Then there exists a constant $C_{\mathrm{lock}} > 0$, independent of σ, such that after the additional physical time

$$\xi_{\mathrm{lock}}^\sigma := C_{\mathrm{lock}} \log \frac{1}{\sigma} \tag{409}$$

all $m+1$ captured atoms satisfy

$$\max_{1\le r\le m+1} |\pi_{r,\sigma} - \lambda_{i_r}| \le C\sigma^\nu, \tag{410}$$

and

$$\max_{\substack{1\le r\le m+1\\ a\in[\nu]}} \left\|\mathbf{h}_{r,\sigma}^{(a)}\right\| \le C\sigma^\nu. \tag{411}$$

Every atom whose index is outside $\mathcal{A}_{m+1}$ remains $O(\sigma)$ throughout this interval.

Proof. The proof is another continuation argument, now in a fixed macroscopic neighborhood of the enlarged captured manifold.

Step 1: preservation of a coarse captured annulus. At ξ_0^σ, the old captured factors are uniformly bounded above and away from zero by assumption 1, and the new factors satisfy the fixed bounds (406). Choose a slightly larger compact annulus containing all these values and bootstrap that every main coordinate of the $m+1$ captured atoms remains inside this annulus.

On this bootstrap interval, each captured product obeys

$$\frac{d}{d\xi}\big(\pi_{r,\sigma} - \lambda_{i_r}\big) = -S_{r,\sigma}(\xi)\big(\pi_{r,\sigma} - \lambda_{i_r}\big) + \mathcal{E}_{r,\sigma}^\pi(\xi), \tag{412}$$

with

$$S_{r,\sigma}(\xi) \ge c_S > 0. \tag{413}$$

The product forcing satisfies

$$|\mathcal{E}_{r,\sigma}^\pi(\xi)| \le C\left(\sigma^\nu + E_{\perp,\sigma}(\xi)^{\nu-1}\right), \tag{414}$$

where

$$E_{\perp,\sigma}(\xi) := \max_{\substack{1\le r\le m+1\\ a\in[\nu]}} \left\|\mathbf{h}_{r,\sigma}^{(a)}(\xi)\right\|. \tag{415}$$

For the old captured atoms, $\pi_{r,\sigma}(\xi_0^\sigma) = \lambda_{i_r} + o(1)$. For the new winner, the phase assumption (407) and the lower factor bound in (406) imply that its initial product has real part bounded below by a positive constant for sufficiently small σ. Variation of constants in (412) therefore shows, as long as $E_{\perp,\sigma} = o(1)$, that every captured product remains bounded above and away from zero.

The exact atom-wise balance invariant (73), together with the decomposition into main and off-sector components, then implies that the squared magnitudes of the factors belonging to the same atom differ by $O(\sigma^2) + O(E_{\perp,\sigma}^2)$. Hence the product bounds force every individual main-coordinate magnitude to stay inside a slightly smaller annulus. This closes the coarse-annulus continuation once the transverse estimate below is established.

Step 2: transverse contraction. Projecting the GF equation orthogonally to the assigned sector of each captured atom gives

$$D^+ E_{\perp,\sigma} \le -c_\gamma E_{\perp,\sigma} + C E_{\perp,\sigma}^{\nu-1} + C\sigma^\nu \tag{416}$$

for some $c_\gamma > 0$ independent of σ. The positive damping constant comes from the coarse lower bounds on all captured main coordinates.

Since $\nu \geq 3$, the nonlinear term is of higher order. Choosing the coarse neighborhood sufficiently small gives

$$CE_{\perp,\sigma}^{\nu-1} \leq \frac{c_\gamma}{2} E_{\perp,\sigma}, \tag{417}$$

and therefore

$$D^+ E_{\perp,\sigma} \leq -\frac{c_\gamma}{2} E_{\perp,\sigma} + C\sigma^\nu. \tag{418}$$

Variation of constants yields

$$E_{\perp,\sigma}(\xi) \leq e^{-c_\gamma(\xi-\xi_0^\sigma)/2} E_{\perp,\sigma}(\xi_0^\sigma) + C\sigma^\nu. \tag{419}$$

The initial transverse error is $o(1)$ by assumptions 1 and 3. Choosing C_{lock} sufficiently large in (409) makes the exponentially decaying term $O(\sigma^\nu)$ at $\xi = \xi_0^\sigma + \xi_{\mathrm{lock}}^\sigma$. This proves (411).

Step 3: contraction of the product errors. Substituting the transverse estimate (419) into the product equation (412) gives

$$|\pi_{r,\sigma}(\xi) - \lambda_{i_r}| \leq e^{-c_S(\xi-\xi_0^\sigma)} |\pi_{r,\sigma}(\xi_0^\sigma) - \lambda_{i_r}| + C\sigma^\nu. \tag{420}$$

The initial product errors are uniformly bounded: they are $o(1)$ for the old captured atoms and $O(1)$ for the new winner. Increasing C_{lock} if necessary makes the first term $O(\sigma^\nu)$ at the end of the locking interval. This proves (410).

Step 4: the remaining atoms stay microscopic. Every still-unclaimed atom has physical-time derivative $O(\sigma^{\nu-1})$ while the captured factors remain in the fixed annulus. Over the logarithmic locking interval, its total change is therefore

$$O\left(\sigma^{\nu-1} \log \frac{1}{\sigma}\right) = o(\sigma), \tag{421}$$

because $\nu \geq 3$. Since these atoms are $O(\sigma)$ at ξ_0^σ, they remain $O(\sigma)$ throughout the locking stage. This also completes the continuation argument used in Step 1. □

Completion of one fast transition.

Proposition 9 (Fast capture and renewal)**.** *Suppose that the induction hypotheses* $(\mathrm{I_m})$–$(\mathrm{IV_m})$ *hold at* ρ_m^σ*, and let* η_{m+1}^σ *be the entrance time supplied by Proposition 8.*

Then there exists a reset time ρ_{m+1}^σ *such that*

$$\rho_{m+1}^\sigma - \theta_{m+1}^* \longrightarrow 0, \tag{422}$$

and the induction hypotheses $(\mathrm{I_{m+1}})$–$(\mathrm{IV_{m+1}})$ *hold at* ρ_{m+1}^σ.

Proof. Apply Lemma 6 and set

$$\xi_0^\sigma := \xi_{\mathrm{grow}}^\sigma. \tag{423}$$

At this time the winning-factor bounds (304), the phase alignment (309), the winner off-sector estimate (305), the old-background estimate (307), and the spectator estimate (306) verify the hypotheses of Lemma 7.

Define the total physical duration of the fast layer by

$$T_{\mathrm{fast},\sigma} := \xi_{\mathrm{grow}}^\sigma + \xi_{\mathrm{lock}}^\sigma \tag{424}$$

and set

$$\rho_{m+1}^\sigma := \eta_{m+1}^\sigma + \sigma^{\nu-2} T_{\mathrm{fast},\sigma}. \tag{425}$$

The growth-time estimate (303) and the logarithmic locking time (409) give

$$\begin{aligned} \sigma^{\nu-2} T_{\mathrm{fast},\sigma} &= O\Big(\sigma^{\nu-2} r_\sigma^{-(\nu-2)}\Big) + O\left(\sigma^{\nu-2} \log \frac{1}{\sigma}\right) \\ &= O\Big(R_\sigma^{-(\nu-2)}\Big) + o(1) \longrightarrow 0. \end{aligned} \tag{426}$$

Together with the entrance-time convergence $\eta_{m+1}^\sigma \to \theta_{m+1}^*$ from (232), this proves the reset-time convergence (422).

It remains to verify the four induction hypotheses at ρ_{m+1}^σ.

1. Verification of $(\mathrm{I}_{\mathrm{m}+1})$**.** The product-locking conclusion (410) of Lemma 7 gives

$$\max_{1\le r\le m+1}\left|\pi_{r,\sigma}(\rho^{\sigma}_{m+1})-\lambda_{i_r}\right|\le C_{m+1}\sigma^{\nu}. \tag{427}$$

This is exactly $(\mathrm{I}_{\mathrm{m}+1})$.

2. Verification of $(\mathrm{II}_{\mathrm{m}+1})$**.** The spectator conclusions of Lemmas 6 and 7 show that every atom whose index is not in $\mathcal{A}_{m+1}$ remains at the initialization scale throughout the fast layer. Hence

$$\max_{\substack{k\notin\mathcal{A}_{m+1}\\ a\in[\nu]}}\left\|\mathbf{u}^{(a)}_{k,\sigma}(\rho^{\sigma}_{m+1})\right\|\le C_{m+1}\sigma, \tag{428}$$

which is $(\mathrm{II}_{\mathrm{m}+1})$.

3. Verification of $(\mathrm{III}_{\mathrm{m}+1})$**.** Let $(k,i)\in\mathcal{E}_{m+1}$. Then $k\neq k_*$ and $i\neq i_*$, so this candidate is a spectator throughout the fast transition.

Its physical-time derivative is $O(\sigma^{\nu-1})$ in every mode. Since $y^{(a)}_{ki,\sigma}=x^{(a)}_{ki,\sigma}/\sigma$, its microscopic coordinate therefore satisfies

$$\left|\frac{d}{d\xi}y^{(a)}_{ki,\sigma}\big(\Theta_\sigma(\xi)\big)\right|\le C\sigma^{\nu-2}. \tag{429}$$

Integrating over the total fast duration and using the vanishing slow width (426) gives

$$\max_{\substack{(k,i)\in\mathcal{E}_{m+1}\\ a\in[\nu]}}\left|y^{(a)}_{ki,\sigma}(\rho^{\sigma}_{m+1})-y^{(a)}_{ki,\sigma}(\eta^{\sigma}_{m+1})\right|\longrightarrow 0. \tag{430}$$

Every candidate in $\mathcal{E}_{m+1}$ has clock strictly larger than θ^{*}_{m+1} by the future-clock property (68). Hence its isolated profile is smooth in a fixed neighborhood of θ^{*}_{m+1}. Since both η^{σ}_{m+1} and ρ^{σ}_{m+1} converge to this time,

$$\max_{\substack{(k,i)\in\mathcal{E}_{m+1}\\ a\in[\nu]}}\left|\zeta^{(a)}_{ki}(\rho^{\sigma}_{m+1})-\zeta^{(a)}_{ki}(\eta^{\sigma}_{m+1})\right|\longrightarrow 0. \tag{431}$$

Finally, the diagonal profile estimate (252), established in the proof of Proposition 8, gives

$$\max_{\substack{(k,i)\in\mathcal{E}_{m+1}\\ a\in[\nu]}}\left|y^{(a)}_{ki,\sigma}(\eta^{\sigma}_{m+1})-\zeta^{(a)}_{ki}(\eta^{\sigma}_{m+1})\right|\longrightarrow 0. \tag{432}$$

Combining the exact fast-layer change (430), the isolated-profile change (431), and the entrance comparison (432) proves $(\mathrm{III}_{\mathrm{m}+1})$.

4. Verification of $(\mathrm{IV}_{\mathrm{m}+1})$**.** The off-sector locking conclusion (411) gives

$$\max_{\substack{1\le r\le m+1\\ a\in[\nu]}}\left\|\mathbf{h}^{(a)}_{r,\sigma}(\rho^{\sigma}_{m+1})\right\|\le C_{m+1}\sigma^{\nu}, \tag{433}$$

which is the captured-atom part of $(\mathrm{IV}_{\mathrm{m}+1})$.

For a surviving atom $k\notin\mathcal{A}_{m+1}$, the microscopic norm bound $(\mathrm{II}_{\mathrm{m}+1})$ just proved implies that every scalar coordinate is $O(\sigma)$. In particular,

$$\max_{\substack{k\notin\mathcal{A}_{m+1},\ i\in\mathcal{S}_{m+1}\\ a\in[\nu]}}\left|x^{(a)}_{ki,\sigma}(\rho^{\sigma}_{m+1})\right|\le C_{m+1}\sigma, \tag{434}$$

or, equivalently, the corresponding rescaled coordinates are uniformly $O(1)$. Thus $(\mathrm{IV}_{\mathrm{m}+1})$ holds.

The renewed product condition $(\mathrm{I}_{\mathrm{m}+1})$ and off-sector locking $(\mathrm{IV}_{\mathrm{m}+1})$ also imply, through the exact balance invariant (75), the derived captured-factor balance

$$\max_{\substack{1\le r\le m+1\\ a\in[\nu]}}\left|\left|q^{(a)}_{r,\sigma}(\rho^{\sigma}_{m+1})\right|-\lambda_{i_r}^{1/\nu}\right|\longrightarrow 0, \tag{435}$$

in agreement with the general consequence (88).

We have therefore renewed $(\mathrm{I}_{\mathrm{m}+1})$–$(\mathrm{IV}_{\mathrm{m}+1})$ at a reset time satisfying (422). This completes the induction step. □

Renewal of the induction. The renewed profile-preservation hypothesis $(\mathrm{III}_{\mathrm{m+1}})$ compares every surviving candidate with the same isolated profile $\boldsymbol{\zeta}_{ki}$ defined from the original initialization in (62). Thus earlier captures do not restart the candidate clocks.

By the future-clock property (68), now with m replaced by $m+1$,

$$\tau_{ki} > \theta^*_{m+1} \qquad \text{for every } (k,i) \in \mathcal{E}_{m+1}. \tag{436}$$

If $m+1 < M$, the greedy next-edge rule (67) therefore gives

$$(k_{m+2}, i_{m+2}) = \operatorname*{arg\,min}_{(k,i)\in\mathcal{E}_{m+1}} \tau_{ki}, \qquad \theta^*_{m+2} = \min_{(k,i)\in\mathcal{E}_{m+1}} \tau_{ki}. \tag{437}$$

Since

$$\rho^\sigma_{m+1} - \theta^*_{m+1} \longrightarrow 0 \tag{438}$$

and the next clock has the fixed positive gap $\theta^*_{m+2} - \theta^*_{m+1} > 0$, we have, for all sufficiently small σ,

$$\rho^\sigma_{m+1} < \theta^*_{m+2}. \tag{439}$$

Proposition 7 can therefore be applied again with m replaced by $m+1$. The induction continues with the same original isolated profiles and candidate clocks.

A3.6 Completion of the proof of the greedy matching limit

We now assemble the preceding estimates to complete the induction and prove Theorem 1. The only additional ingredient is control of the terminal plateau after the last accepted edge. At that point there is no subsequent activation clock, so Proposition 7 is no longer the appropriate propagation statement. Instead, the exhaustion of one side of the matching allows us to propagate the final locked state directly.

Lemma 8 (Propagation after the final capture)**.** *Suppose that the induction hypotheses* $(\mathrm{I_M})$–$(\mathrm{IV_M})$ *hold at the final reset time* ρ^σ_M*. Then, for every fixed* $\Theta > \theta^*_M$*,*

$$\sup_{\theta\in[\rho^\sigma_M,\Theta]} \left\|\widehat{T}_\sigma(\theta) - T_M\right\|_F \longrightarrow 0 \qquad \text{as } \sigma \to 0. \tag{440}$$

Proof. Since $M = \min\{H,p\}$, after the final greedy acceptance one side of the bipartite graph is exhausted. Equivalently,

$$\mathcal{A}_M = [H] \qquad \text{or} \qquad \mathcal{S}_M = [p]. \tag{441}$$

Thus either there are no surviving atoms, or every target sector has already been claimed.

Fix $\Theta > \theta^*_M$. The reset-time convergence supplied by the final application of Proposition 9,

$$\rho^\sigma_M \longrightarrow \theta^*_M, \tag{442}$$

implies that $\rho^\sigma_M < \Theta$ for all sufficiently small σ.

We use a continuation argument on $[\rho^\sigma_M, \Theta]$. For the captured atoms, impose the coarse annulus bootstrap

$$\frac{1}{2}\lambda_{i_r}^{1/\nu} < \left|q^{(a)}_{r,\sigma}(\theta)\right| < 2\lambda_{i_r}^{1/\nu}, \qquad r \le M, \quad a \in [\nu]. \tag{443}$$

If surviving atoms remain, then the exhaustion property (441) forces $\mathcal{S}_M = [p]$. In this case we also impose

$$\max_{\substack{k\notin\mathcal{A}_M \\ i\in[p],\ a\in[\nu]}} \left|y^{(a)}_{ki,\sigma}(\theta)\right| < C_M + 1. \tag{444}$$

If $\mathcal{A}_M = [H]$, there are no surviving atoms and the second bootstrap condition is vacuous.

The captured-factor balance at the reset time, (88) with $m = M$, gives strict initial slack in the annulus (443). If surviving atoms remain, then all sectors are claimed, and the blocked-sector induction bound (85) with $m = M$ gives strict initial slack in (444). Hence the bootstrap conditions hold on a nontrivial interval beginning at ρ^σ_M.

Step 1: the bootstrap assumptions give the hypotheses of the locked-background proposition. Suppose first that surviving atoms remain. The survivor bootstrap (444) and orthonormality of the sector basis give

$$\begin{aligned}\left\|\mathbf{u}_{k,\sigma}^{(a)}(\theta)\right\|^2 &= \sigma^2 \sum_{i=1}^{p} \left|y_{ki,\sigma}^{(a)}(\theta)\right|^2 \\ &\le p(C_M+1)^2\sigma^2.\end{aligned} \tag{445}$$

Thus every surviving atom is uniformly $O(\sigma)$ on the bootstrap interval. If no surviving atoms remain, this condition is automatic.

The captured annulus (443) supplies the coarse lower and upper bounds required by Proposition 6. Its quantitative initial-locking assumption follows from the captured-product induction estimate (80) and the captured off-sector induction estimate (83). Finally, $[\rho_M^\sigma, \Theta]$ has uniformly bounded length by the reset convergence (442). Hence all assumptions of Proposition 6 hold on the current bootstrap interval.

Step 2: the captured atoms remain locked and balanced. The captured-background stability conclusion (109) of Proposition 6 gives

$$\max_{r\le M} \sup_\theta |\pi_{r,\sigma}(\theta) - \lambda_{i_r}| + \max_{\substack{r\le M\\ a\in[\nu]}} \sup_\theta \left\|\mathbf{h}_{r,\sigma}^{(a)}(\theta)\right\| = O(\sigma^\nu) \tag{446}$$

on the bootstrap interval.

At the same time, the absolute main-coordinate balance estimate (110), which follows from the exact atom-wise balance invariant, gives

$$\max_{\substack{r\le M\\ a,b\in[\nu]}} \sup_\theta \left||q_{r,\sigma}^{(a)}(\theta)|^2 - |q_{r,\sigma}^{(b)}(\theta)|^2\right| = O(\sigma^2). \tag{447}$$

The product part of the locking estimate (446) also gives

$$\prod_{a=1}^{\nu} \left|q_{r,\sigma}^{(a)}(\theta)\right| = \left|\pi_{r,\sigma}(\theta)\right| = \lambda_{i_r} + o(1) \tag{448}$$

uniformly on the bootstrap interval. Combining the asymptotic equality of the squared magnitudes (447) with the product constraint (448) yields

$$\max_{\substack{r\le M\\ a\in[\nu]}} \sup_\theta \left|\left|q_{r,\sigma}^{(a)}(\theta)\right| - \lambda_{i_r}^{1/\nu}\right| \longrightarrow 0. \tag{449}$$

For sufficiently small σ, this is a strict improvement of the captured-annulus bootstrap (443).

Step 3: any remaining atoms stay microscopic. This step is needed only if surviving atoms remain. In that case, $\mathcal{S}_M = [p]$ by the exhaustion property (441), so every coordinate of every surviving atom lies in a claimed sector.

The claimed-sector near-freezing estimate (115) from Proposition 6 therefore gives

$$\max_{\substack{k\notin\mathcal{A}_M\\ i\in[p],\ a\in[\nu]}} \sup_\theta \left|\frac{d}{d\theta} y_{ki,\sigma}^{(a)}(\theta)\right| \le C\sigma^\nu \tag{450}$$

on the bootstrap interval. Since $\Theta - \rho_M^\sigma$ is uniformly bounded, integrating the derivative bound (450) and using the blocked-sector reset estimate (85) gives

$$\max_{\substack{k\notin\mathcal{A}_M\\ i\in[p],\ a\in[\nu]}} \sup_{\theta\in[\rho_M^\sigma,\Theta]} \left|y_{ki,\sigma}^{(a)}(\theta)\right| \le C_M + o(1). \tag{451}$$

For sufficiently small σ, this strictly improves the survivor bootstrap (444).

Step 4: closure of the continuation argument. Steps 2 and 3 strictly improve every nonvacuous bootstrap condition. The same maximal-interval argument as in Lemma 1 therefore shows that the bootstrap interval extends to all of $[\rho_M^\sigma, \Theta]$.

We may consequently apply the tensor-level conclusion (116) of Proposition 6 on the full interval, obtaining

$$\sup_{\theta\in[\rho_M^\sigma,\Theta]} \left\|\widehat{T}_\sigma(\theta) - T_M\right\|_F = O(\sigma^\nu). \tag{452}$$

This proves the terminal propagation statement (440). □

Proof of Theorem 1. We work throughout on the probability-one initialization event fixed at the beginning of Appendix A3, on which all candidate clocks are finite and pairwise distinct.

Step 1: completion of the induction. The base case $m = 0$ was verified after the induction hypotheses were introduced. At the initial reset time $\rho_0^\sigma = 0$ from (77), (I_0) and the captured-atom part of (IV_0) are vacuous, (II_0) follows directly from the Gaussian initialization, and (III_0) holds exactly because the exact microscopic coordinates and the isolated profiles have the same initial data.

Suppose now that $(\mathrm{I_m})$–$(\mathrm{IV_m})$ hold at ρ_m^σ for some $m < M$. Proposition 7 propagates this state through every fixed interval before the next accepted clock. The mesoscopic-entrance construction of Proposition 8 then produces an entrance time $\eta_{m+1}^\sigma \to \theta_{m+1}^*$, and Proposition 9 carries the resulting fast transition through to a new reset time ρ_{m+1}^σ at which $(\mathrm{I_{m+1}})$–$(\mathrm{IV_{m+1}})$ hold. In particular, its reset-time estimate gives

$$\rho_{m+1}^\sigma - \theta_{m+1}^* \longrightarrow 0. \tag{453}$$

If $m + 1 < M$, the next accepted clock has the fixed positive gap

$$\theta_{m+2}^* - \theta_{m+1}^* > 0. \tag{454}$$

Indeed, this follows from the distinctness of the accepted candidate clocks, or equivalently from the future-clock property (68) and the greedy next-edge rule (67). The reset convergence (453) therefore implies

$$\rho_{m+1}^\sigma < \theta_{m+2}^* \tag{455}$$

for all sufficiently small σ.

Crucially, the renewed profile-preservation property $(\mathrm{III_{m+1}})$ compares every surviving edge in $\mathcal{E}_{m+1}$ with the same isolated profile $\boldsymbol{\zeta}_{ki}$ defined from the original initialization in (62). Thus the candidate clocks are preserved through the fast transition rather than restarted. The induction can therefore be repeated with the next edge selected by the same original greedy clock process.

Iterating this argument constructs reset times $\rho_1^\sigma, \dots, \rho_M^\sigma$ such that

$$\rho_m^\sigma \longrightarrow \theta_m^*, \qquad 1 \le m \le M, \tag{456}$$

and the hypotheses $(\mathrm{I_m})$–$(\mathrm{IV_m})$ hold at every ρ_m^σ.

Step 2: uniform convergence on every nonterminal plateau. Recall the accepted times $\theta_1^*, \dots, \theta_M^*$ from (64), and define the open plateaus between them by

$$J_0 := [0, \theta_1^*), \qquad J_m := (\theta_m^*, \theta_{m+1}^*), \quad 1 \le m < M, \qquad J_M := (\theta_M^*, \infty). \tag{457}$$

By the greedy construction, the selected-sector set is constant on each plateau and equals $\mathcal{S}_m$. Hence, using the partial target T_m from (69),

$$\sum_{i\in\mathcal{S}(\theta)} \lambda_i \mathbf{e}_i^{(1)} \otimes \cdots \otimes \mathbf{e}_i^{(\nu)} = T_m, \qquad \theta \in J_m. \tag{458}$$

Let

$$K \subset [0, \infty) \setminus \mathcal{T}_* \tag{459}$$

be compact, as in the statement of Theorem 1, and set

$$K_m := K \cap J_m, \qquad 0 \le m \le M. \tag{460}$$

Because K is compact and avoids the finite activation set $\mathcal{T}_* = \{\theta_1^*, \dots, \theta_M^*\}$, every nonempty K_m lies a positive distance from the activation times bounding its plateau.

Fix $m < M$ with $K_m \neq \varnothing$. Choose $\delta_m > 0$ such that

$$0 < \delta_m < \frac{1}{2}\big(\theta_{m+1}^* - \theta_m^*\big), \qquad \sup K_m < \theta_{m+1}^* - \delta_m. \tag{461}$$

For $m \geq 1$, the positive separation of K_m from the left activation time also gives

$$\inf K_m > \theta_m^*. \tag{462}$$

The reset convergence (456) therefore implies

$$\rho_m^\sigma < \inf K_m \tag{463}$$

for all sufficiently small σ. For $m = 0$, the same conclusion is automatic because $\rho_0^\sigma = 0$. Hence

$$K_m \subset \big[\rho_m^\sigma, \theta_{m+1}^* - \delta_m\big] \tag{464}$$

for every nonempty K_m with $m < M$ and all sufficiently small σ.

Applying the uniform plateau approximation (176) from Proposition 7 on the containing interval (464) gives

$$\sup_{\theta \in K_m} \big\|\widehat{T}_\sigma(\theta) - T_m\big\|_F \longrightarrow 0, \qquad 0 \leq m < M. \tag{465}$$

Step 3: uniform convergence on the terminal plateau. If $K_M = \varnothing$, there is nothing to prove. Otherwise, because $K_M \subset (\theta_M^*, \infty)$ and K_M is separated from θ_M^*, we have

$$\inf K_M > \theta_M^*. \tag{466}$$

Choose a fixed

$$\Theta > \sup K_M. \tag{467}$$

The final reset convergence $\rho_M^\sigma \to \theta_M^*$ from (456) and the left gap (466) imply, for sufficiently small σ,

$$K_M \subset [\rho_M^\sigma, \Theta]. \tag{468}$$

Applying the terminal propagation lemma, specifically its output estimate (440), gives

$$\sup_{\theta \in K_M} \big\|\widehat{T}_\sigma(\theta) - T_M\big\|_F \longrightarrow 0. \tag{469}$$

Step 4: conclusion for the tensor trajectory. There are only finitely many plateaus. Combining the nonterminal convergence (465), the terminal convergence (469), and the identification of the greedy tensor on each plateau (458) gives

$$\sup_{\theta \in K} \left\| \widehat{T}_\sigma(\theta) - \sum_{i \in \mathcal{S}(\theta)} \lambda_i \mathbf{e}_i^{(1)} \otimes \cdots \otimes \mathbf{e}_i^{(\nu)} \right\|_F \longrightarrow 0. \tag{470}$$

By the slow-time definitions (58) and (94), $\widehat{T}_\sigma(\theta)$ is precisely $\widehat{T}^\sigma(\theta)$ in the notation of the main theorem. Therefore (470) is exactly the tensor convergence (7) asserted in Theorem 1.

Step 5: conclusion for the loss trajectory. For $\theta \in J_m$, the greedy loss (6) and orthogonality of the diagonal sector tensors give

$$L_*(\theta) = \frac{1}{2}\|T - T_m\|_F^2 = \frac{1}{2}\sum_{i \notin \mathcal{S}_m} \lambda_i^2. \tag{471}$$

For $\theta \in K_m$, the elementary difference-of-squares identity yields

$$\begin{aligned} |L^\sigma(\theta) - L_*(\theta)| &= \frac{1}{2}\left| \|T - \widehat{T}^\sigma(\theta)\|_F^2 - \|T - T_m\|_F^2 \right| \\ &\leq \frac{1}{2}\Big(\|T - \widehat{T}^\sigma(\theta)\|_F + \|T - T_m\|_F \Big) \big\|\widehat{T}^\sigma(\theta) - T_m\big\|_F. \end{aligned} \tag{472}$$

The tensor convergence already proved in (470) makes the second factor tend to zero uniformly on each K_m, while the first factor remains uniformly bounded. Since there are only finitely many nonempty plateau pieces K_m, we obtain

$$\sup_{\theta \in K} |L^\sigma(\theta) - L_*(\theta)| \longrightarrow 0. \tag{473}$$

This is precisely the loss convergence (8). The proof of Theorem 1 is complete. □

A4 COMPLEX MODULAR ADDITION: FOURIER DETAILS [Section 4]

We give the Fourier calculation underlying Section 4 and derive Corollary 1 directly from Theorem 1.

Fourier diagonalization. For $q \in \mathbb{Z}_p$, the discrete Fourier identity

$$\mathbf{1}\{q = 0 \pmod p\} = \frac{1}{p} \sum_{r \in \mathbb{Z}_p} \omega^{rq}$$

gives, with $q = x + y - z$,

$$\mathbf{1}\{x + y = z \pmod p\} = \frac{1}{p} \sum_{r \in \mathbb{Z}_p} \omega^{r(x+y-z)} = \sqrt{p} \sum_{r \in \mathbb{Z}_p} \boldsymbol{\varphi}_r(x) \boldsymbol{\varphi}_r(y) \overline{\boldsymbol{\varphi}_r(z)}.$$

This proves (10). Moreover, $\{\boldsymbol{\varphi}_r\}_{r \in \mathbb{Z}_p}$ and $\{\overline{\boldsymbol{\varphi}_r}\}_{r \in \mathbb{Z}_p}$ are orthonormal bases of $\mathbb{C}^p$. Hence (10) is exactly of the diagonal form (3), with the mode-wise orthonormal bases

$$\mathbf{e}_r^{(1)} = \mathbf{e}_r^{(2)} = \boldsymbol{\varphi}_r, \qquad \mathbf{e}_r^{(3)} = \overline{\boldsymbol{\varphi}_r},$$

and equal sector amplitudes $\lambda_r = \sqrt{p}$. Thus all assumptions of Theorem 1 concerning the target geometry are satisfied.

In particular, the sector tensors $\boldsymbol{\varphi}_r \otimes \boldsymbol{\varphi}_r \otimes \overline{\boldsymbol{\varphi}_r}$ are orthonormal in the Frobenius inner product. Therefore

$$\|T_p^{\mathrm{add}}\|_F^2 = p^2.$$

Gaussian coordinates and activation clocks. For atom k and Fourier sector r, define the microscopic coordinate triple

$$\mathbf{g}_{kr} := \left(\langle \boldsymbol{\varphi}_r, \mathbf{g}_k^{(1)} \rangle,\ \langle \boldsymbol{\varphi}_r, \mathbf{g}_k^{(2)} \rangle,\ \langle \overline{\boldsymbol{\varphi}_r}, \mathbf{g}_k^{(3)} \rangle \right).$$

For each mode, the change from the original coordinates to the corresponding Fourier basis is unitary. Since a standard circularly symmetric complex Gaussian vector is invariant under unitary transformations, its Fourier coordinates are again independent standard complex Gaussians. Independence across atoms and modes is preserved as well. Consequently,

$$\{\mathbf{g}_{kr} : k \in [H],\ r \in \mathbb{Z}_p\} \quad \text{are i.i.d. with} \quad \mathbf{g}_{kr} \sim \mathcal{N}_{\mathbb{C}}(0, I_3).$$

Applying the general clock formula (5) with $\lambda_r = \sqrt{p}$ gives

$$\tau_{kr} = \frac{\tau_3(\mathbf{g}_{kr})}{\sqrt{p}}.$$

Thus, after the change of time $\vartheta = \sqrt{p}\,\theta$, the candidate activation times are $\xi_{kr} := \tau_3(\mathbf{g}_{kr})$. By Proposition 5, the ξ_{kr} are i.i.d. and atomless.

Proof of Corollary 1. Let $\mathcal{T}_*$ be the set of accepted activation times in the slow-time variable θ, and let

$$\mathcal{T}_{*,p}^{\mathrm{add}} := \sqrt{p}\,\mathcal{T}_*$$

be the corresponding set in the parametrization $\vartheta = \sqrt{p}\,\theta$. Equivalently, $\mathcal{T}_{*,p}^{\mathrm{add}}$ consists of the values ξ_{kr} associated with the edges accepted by the greedy process.

For any compact $K \subset [0, \infty) \setminus \mathcal{T}_{*,p}^{\mathrm{add}}$, the rescaled set $K/\sqrt{p}$ is a compact subset of $[0, \infty) \setminus \mathcal{T}_*$. Therefore Theorem 1, applied with $\nu = 3$, $\lambda_r = \sqrt{p}$, and $\theta = \vartheta/\sqrt{p}$, gives

$$\sup_{\vartheta \in K} \left\| \widehat{T}_p^{\mathrm{add},\sigma}(\vartheta) - \sqrt{p} \sum_{r \in S_p(\vartheta)} \boldsymbol{\varphi}_r \otimes \boldsymbol{\varphi}_r \otimes \overline{\boldsymbol{\varphi}_r} \right\|_F \longrightarrow 0.$$

This is the tensor convergence stated in Corollary 1.

Since the Fourier sectors are orthonormal, the limiting residual at time ϑ is

$$\sqrt{p} \sum_{r \notin S_p(\vartheta)} \boldsymbol{\varphi}_r \otimes \boldsymbol{\varphi}_r \otimes \overline{\boldsymbol{\varphi}_r},$$

and therefore

$$L^{\mathrm{add}}_{*,p}(\vartheta) = \frac{p}{2}\big(p - M_p(\vartheta)\big).$$

Dividing by $p^2/2$ yields

$$\frac{2L^{\mathrm{add}}_{*,p}(\vartheta)}{p^2} = 1 - \frac{M_p(\vartheta)}{p},$$

which proves (11). □

Finite-p Random Edge ordering. Because the Hp variables ξ_{kr} are i.i.d. and atomless, every strict ordering of the edges of $K_{H,p}$ is equally likely. Hence, for fixed p and H, the accepted-edge sequence has exactly the same law as that obtained by processing the edges of $K_{H,p}$ in a uniformly random order and accepting an edge whenever both endpoints are still unmatched. The clock distribution determines when the events occur, but not their combinatorial ordering. This is precisely the Random Edge process used in Section 5.

A5 Large-system limit: proofs and asymptotics [Section 5]

A5.1 Large-system limit of Random Edge matching

We begin by describing the random matching process used in Proposition 1 in more detail. The algorithm is simply greedy matching in a random edge order. Given the complete bipartite graph $K_{H_p,p}$ between H_p atom vertices and p sector vertices, choose a uniformly random ordering of its $H_p p$ edges. Inspect the edges one at a time in this order. An edge is accepted if neither of its endpoints has appeared in any previously accepted edge; otherwise it is rejected. Equivalently, an edge is accepted if both of its endpoints are still unmatched. Following Arnosti (2022), we call this the *Random Edge* algorithm.

For the large-system analysis, it is useful to realize the same random ordering continuously. Equivalently, assign each edge an independent *random* priority, uniformly distributed on $[0, 1]$, and inspect the edges in increasing order of priority. Since the priority distribution is continuous, the priorities are almost surely distinct, and their ordering is a uniformly random permutation of the edges.

Edge exposure. For $s \in [0, p]$, we say that an edge is *exposed by level s* if its priority is at most s/p. Thus an exposed edge is simply an edge that has already been reached and processed by the algorithm; it may have been either accepted or rejected. Each edge is independently exposed with probability s/p, so the graph formed by the exposed edges is the bipartite Erdős–Rényi graph $G_{H_p,p,s/p}$, in which each of the $H_p p$ possible edges is present independently with probability s/p. We call s the *edge exposure parameter*.

The scaling s/p is the sparse scaling on which the matching process has a nontrivial large-p evolution. Indeed, every atom vertex has p possible incident edges, and therefore its expected exposed degree at exposure s is

$$p\frac{s}{p} = s.$$

Likewise, every sector vertex has H_p possible incident edges, so its expected exposed degree is

$$H_p\frac{s}{p} = \frac{H_p}{p}s \longrightarrow \alpha s.$$

Thus fixed $s = O(1)$ corresponds to a sparse graph in which both vertex classes have $O(1)$ exposed degree and the matching process has a nontrivial large-system evolution. By contrast, if the priority threshold were $o(p^{-1})$, almost all vertices would have no exposed edges, whereas a fixed positive priority threshold would give exposed degrees of order p and would place the graph in a dense regime.

The factor p^{-1} is the essential scaling. Using H_p^{-1} instead would produce an equivalent parametrization differing asymptotically only by the constant factor α. Our convention makes s the expected exposed degree on the atom side and is convenient for the tensor application below.

Let $M_p^{\mathrm{RE}}(s)$ denote the number of edges accepted by the time all edges with priority at most s/p have been processed. Since each accepted edge matches exactly one sector vertex,

$$1 - \frac{M_p^{\mathrm{RE}}(s)}{p}$$

is the fraction of sector vertices that remain unmatched at exposure s. The large-system limit below identifies the deterministic limit of this quantity.

The limiting matching trajectory is stated in Proposition 1. We write $X_p \xrightarrow{\mathbb{P}} X$ for convergence in probability: for every $\varepsilon > 0$, $\mathbb{P}(|X_p - X| > \varepsilon) \to 0$ as $p \to \infty$. The formula for $\ell_\alpha(s)$ in Proposition 1 is the sparse bipartite Random Edge asymptotic derived by Arnosti (2022, Appendix B.1); we use this known random-graph limit as a consequence of the tensor-dynamics reduction, rather than as a new combinatorial result. Related random-order greedy matching results include Dyer et al. (1993), while the differential-equation method used in the proof below is part of the general framework developed by Wormald (1995; 1999).

We now give a self-contained proof of Proposition 1. Besides recovering the known closed form, the argument directly controls the matching trajectory on every fixed exposure interval, which is the form needed below.

Proof of Proposition 1. Write

$$\alpha_p := \frac{H_p}{p}, \qquad \alpha_p \longrightarrow \alpha.$$

It is enough to prove uniform convergence on an arbitrary fixed interval $[0, S]$.

Step 1: the random-order process. First ignore the numerical values of the edge priorities and retain only their relative ordering. Since the priorities are i.i.d. and continuous, this ordering is a uniformly random permutation of the $H_p p$ edges of $K_{H_p,p}$.

Let $\widetilde{M}_{p,n}$ denote the number of edges accepted after the first n edges of this random ordering have been processed, and let $\mathcal{F}_{p,n}$ be the history up to that point. If $\widetilde{M}_{p,n} = m$, then $H_p - m$ atom vertices and $p - m$ sector vertices remain unmatched. Moreover, every edge joining two currently unmatched vertices is still unprocessed: had such an edge appeared earlier, it would have been accepted when processed. Hence exactly $(H_p - m)(p - m)$ of the $H_p p - n$ unprocessed edges join two unmatched vertices. Therefore, if $J_{p,n+1}$ is the indicator that the next edge is accepted,

$$\mathbb{P}(J_{p,n+1} = 1 \,|\, \mathcal{F}_{p,n}) = \frac{(H_p - \widetilde{M}_{p,n})(p - \widetilde{M}_{p,n})}{H_p p - n}. \tag{474}$$

Set $m_{p,n} := \widetilde{M}_{p,n}/p$ and $s_n := n/H_p$. Then (474) can be written as

$$\mathbb{E}[m_{p,n+1} - m_{p,n} \,|\, \mathcal{F}_{p,n}] = \frac{1}{H_p} f_p(s_n, m_{p,n}), \qquad f_p(s,m) := \frac{(\alpha_p - m)(1-m)}{1 - s/p}. \tag{475}$$

For every fixed $S < \infty$, as $p \to \infty$,

$$f_p(s,m) \longrightarrow f(m) := (\alpha - m)(1-m) \tag{476}$$

uniformly for $0 \le s \le S$ and for all values of m attainable by the process.

Step 2: concentration around the fluid equation. Define the martingale

$$Z_{p,n} := \frac{1}{p} \sum_{j=0}^{n-1} \left(J_{p,j+1} - \mathbb{E}[J_{p,j+1} \mid \mathcal{F}_{p,j}] \right). \tag{477}$$

Its increments have absolute value at most $1/p$. Since $H_p/p \to \alpha$, Azuma–Hoeffding gives, for every $\varepsilon > 0$,

$$\mathbb{P}\left(\max_{0\le n\le H_pS} |Z_{p,n}| > \varepsilon\right) \le 2\exp\left(-\frac{\varepsilon^2 p^2}{2H_pS}\right) \longrightarrow 0. \tag{478}$$

Define the piecewise-constant interpolation

$$\widehat{m}_p(s) := \frac{\widetilde{M}_{p,\lfloor H_p s\rfloor}}{p}. \tag{479}$$

Summing (475) and using (478), together with the uniform convergence (476), gives

$$\sup_{0\le s\le S}\left|\widehat{m}_p(s) - \int_0^s f\big(\widehat{m}_p(r)\big)\,dr\right| \xrightarrow{\mathbb{P}} 0. \tag{480}$$

Indeed, apart from the martingale term, the sum is a Riemann sum with mesh H_p^{-1}; on $[0, S]$, the functions f_p are uniformly bounded and Lipschitz in their second argument.

Let m be the solution of

$$m'(s) = (\alpha - m(s))(1 - m(s)), \qquad m(0) = 0. \tag{481}$$

The vector field in (481) is Lipschitz on the bounded interval containing all possible matching fractions. Subtracting the integral form of (481) from (480) and applying Gronwall's inequality therefore yields

$$\sup_{0\le s\le S} |\widehat{m}_p(s) - m(s)| \xrightarrow{\mathbb{P}} 0. \tag{482}$$

This is the standard differential-equation argument in this particularly simple one-dimensional setting; compare Wormald (1995; 1999).

Step 3: from the number of processed edges to edge exposure. We now return to the actual uniform priorities used in the definition of $M_p^{\mathrm{RE}}(s)$. Let

$$\mathcal{N}_p(s) := \#\{\text{edges with priority at most } s/p\}. \tag{483}$$

For fixed s, $\mathcal{N}_p(s)$ is binomial with parameters H_pp and s/p, so $\mathbb{E}\mathcal{N}_p(s) = H_ps$. Moreover,

$$\sup_{0\le s\le S}\left|\frac{\mathcal{N}_p(s)}{H_p} - s\right| \xrightarrow{\mathbb{P}} 0. \tag{484}$$

For completeness, this uniform statement follows by fixing a finite mesh $0, \delta, 2\delta, \dots, S$, applying a binomial concentration bound at each mesh point, and then using monotonicity of $\mathcal{N}_p(s)$ between successive mesh points. First letting $p \to \infty$ and then $\delta \downarrow 0$ proves (484).

Processing all edges with priority at most s/p is the same as processing the first $\mathcal{N}_p(s)$ edges of the random priority ordering. Hence, pathwise,

$$\frac{M_p^{\mathrm{RE}}(s)}{p} = \widehat{m}_p\left(\frac{\mathcal{N}_p(s)}{H_p}\right). \tag{485}$$

Combining (482), (484), and the uniform continuity of m on compact intervals gives

$$\sup_{0\le s\le S}\left|\frac{M_p^{\mathrm{RE}}(s)}{p} - m(s)\right| \xrightarrow{\mathbb{P}} 0. \tag{486}$$

Step 4: solution of the limiting equation. Let

$$\ell_\alpha(s) := 1 - m(s), \tag{487}$$

so that ℓ_α is the limiting fraction of unmatched sector vertices. Equation (481), which describes the limiting matching fraction, becomes

$$\ell_\alpha'(s) = -\ell_\alpha(s)\big(\ell_\alpha(s) + \alpha - 1\big), \qquad \ell_\alpha(0) = 1. \tag{488}$$

For $\alpha = 1$, this gives $\ell_1'(s) = -\ell_1(s)^2$, and therefore

$$\ell_1(s) = \frac{1}{1+s}. \tag{489}$$

If $\alpha \neq 1$, separation of variables in (488) gives

$$\frac{\ell_\alpha(s)}{\ell_\alpha(s) + \alpha - 1} = \frac{1}{\alpha} e^{-(\alpha-1)s}, \tag{490}$$

and hence

$$\ell_\alpha(s) = \frac{(\alpha-1)e^{-(\alpha-1)s}}{\alpha - e^{-(\alpha-1)s}}. \tag{491}$$

Together with (486), this is precisely (12). Since $S < \infty$ was arbitrary, the convergence is locally uniform for $s \geq 0$. Together with (486), this proves (12) and hence Proposition 1. □

A5.2 From Random Edge matching to the equal-sector tensor limit

We now prove Theorem 2. There is one minor technical point in passing from the fixed-dimensional greedy limit to the large-system trajectory. Theorem 1, specifically the loss convergence in (8), controls GF away from the accepted activation times, whereas Theorem 2 takes a supremum over an entire exposure interval, including the fast transition layers. For equal-amplitude sectors, each jump of the normalized greedy loss has size exactly $1/p$. Together with monotonicity of the GF loss, this makes the transition layers negligible as $p \to \infty$.

For the comparison with GF, we realize the Random Edge process on the same probability space as the Gaussian initialization. For each candidate atom–sector pair (k, i), let

$$\xi_{ki} := \tau_\nu(\mathbf{g}_{ki}).$$

and assign the corresponding edge the priority $F_\nu(\xi_{ki})$, where F_ν is the CDF of the scalar activation clock. Since F_ν is continuous, the probability-integral transform implies that these priorities are i.i.d. uniform on $[0, 1]$. Thus they generate exactly the Random Edge process of Proposition 1, but with its randomness coupled to the same microscopic initialization that determines the GF trajectory.

Since F_ν is nondecreasing and the transformed priorities are almost surely distinct, their ordering agrees almost surely with the ordering of the original clock variables ξ_{ki}. Hence the two greedy processes accept exactly the same sequence of atom–sector pairs. Throughout this subsection, $M_p^{\mathrm{RE}}(s)$ denotes this coupled realization; no additional randomness is introduced.

For fixed p, define the finite-p Random Edge loss trajectory

$$\ell_p^{\mathrm{RE}}(s) := 1 - \frac{M_p^{\mathrm{RE}}(s)}{p}, \tag{492}$$

and define the normalized GF loss in the same exposure parametrization by

$$\ell_{p,\sigma}^{\mathrm{GF}}(s) := \frac{2L(U_{p,\sigma}(t_{p,\sigma}(s)))}{p\lambda_p^2}. \tag{493}$$

Here $t_{p,\sigma}(s)$ is the exposure-to-training-time conversion defined in (14). In what follows we will use the inverse F_ν^{-1}, which is well-defined on $[0, 1)$ by Proposition 5.

The following observation bridges the fast transition layers.

Lemma 9 (Uniform finite-p comparison)**.** *Fix p and $S < \infty$, with $S < p$, and couple the Random Edge process to the Gaussian initialization as above. Then, for almost every realization of that initialization,*

$$\limsup_{\sigma \downarrow 0} \sup_{0 \leq s \leq S} \left| \ell_{p,\sigma}^{\mathrm{GF}}(s) - \ell_p^{\mathrm{RE}}(s) \right| \leq \frac{1}{p}. \tag{494}$$

Proof. Under the coupling above, the Random Edge priorities and the tensor activation clocks have the same ordering. At an exposure value s that is not one of the finitely many edge priorities,

processing all priorities up to s/p therefore selects exactly the same greedy edges as processing the tensor activation clocks up to

$$\theta_p(s) := \frac{1}{\lambda_p} F_\nu^{-1}\left(\frac{s}{p}\right). \tag{495}$$

Consequently, away from the accepted exposure levels, the normalized greedy loss associated with the tensor clocks is exactly $\ell_p^{\mathrm{RE}}(s)$.

By Theorem 1, specifically (8), for every s that is not an accepted exposure level,

$$\ell_{p,\sigma}^{\mathrm{GF}}(s) \longrightarrow \ell_p^{\mathrm{RE}}(s) \qquad \text{as } \sigma \downarrow 0. \tag{496}$$

We now use monotonicity to extend this control through the transition layers. Along GF,

$$\frac{d}{dt} L(U_{p,\sigma}(t)) = -\sum_{\substack{k\in[H_p]\\ a\in[\nu]}} \left\| \nabla_{\mathbf{u}_k^{(a)}} L(U_{p,\sigma}(t)) \right\|^2 \le 0. \tag{497}$$

Since $t_{p,\sigma}(s)$ is nondecreasing in s, $\ell_{p,\sigma}^{\mathrm{GF}}(s)$ is nonincreasing in s. The finite Random Edge trajectory $\ell_p^{\mathrm{RE}}(s)$ is also nonincreasing, and every one of its jumps has size exactly $1/p$, because each accepted edge matches exactly one additional sector.

There are only finitely many accepted exposure levels in $[0, S]$. Choose a finite mesh

$$0 = s_0 < s_1 < \cdots < s_N = S \tag{498}$$

such that no mesh point is an accepted exposure level and every interval $[s_j, s_{j+1}]$ contains at most one accepted exposure level. For fixed S, the endpoints 0 and S are themselves almost surely not accepted exposure levels; if necessary, the same argument may be applied on a slightly larger interval and then restricted to $[0, S]$.

By (496), at the finitely many mesh points,

$$\max_{0\le j\le N} \left|\ell_{p,\sigma}^{\mathrm{GF}}(s_j) - \ell_p^{\mathrm{RE}}(s_j)\right| \longrightarrow 0. \tag{499}$$

Now fix $s \in [s_j, s_{j+1}]$. Monotonicity gives

$$\ell_{p,\sigma}^{\mathrm{GF}}(s_j) \ge \ell_{p,\sigma}^{\mathrm{GF}}(s) \ge \ell_{p,\sigma}^{\mathrm{GF}}(s_{j+1}), \tag{500}$$

and similarly

$$\ell_p^{\mathrm{RE}}(s_j) \ge \ell_p^{\mathrm{RE}}(s) \ge \ell_p^{\mathrm{RE}}(s_{j+1}). \tag{501}$$

Because $[s_j, s_{j+1}]$ contains at most one accepted exposure level,

$$0 \le \ell_p^{\mathrm{RE}}(s_j) - \ell_p^{\mathrm{RE}}(s_{j+1}) \le \frac{1}{p}. \tag{502}$$

Using (500)–(502), we obtain

$$\left|\ell_{p,\sigma}^{\mathrm{GF}}(s) - \ell_p^{\mathrm{RE}}(s)\right| \le \frac{1}{p} + \max\left\{\left|\ell_{p,\sigma}^{\mathrm{GF}}(s_j) - \ell_p^{\mathrm{RE}}(s_j)\right|, \left|\ell_{p,\sigma}^{\mathrm{GF}}(s_{j+1}) - \ell_p^{\mathrm{RE}}(s_{j+1})\right|\right\}. \tag{503}$$

Taking the supremum over $s \in [0, S]$ and then $\limsup_{\sigma\downarrow 0}$, the mesh-point convergence (499) yields (494). □

Lemma 9 is the only additional ingredient needed to combine the fixed-p tensor limit with the large-p Random Edge limit. Notice that the error $1/p$ does not estimate the width of a fast transition in time. Rather, it bounds its possible effect on the *normalized loss*, because a single equal-amplitude sector contributes exactly $1/p$ of the total target energy.

Proof of Theorem 2. Fix a compact set $K \subset [0, \infty)$ and $\varepsilon > 0$. Choose $S < \infty$ such that $K \subset [0, S]$. Since the theorem takes $p \to \infty$, we may restrict to sufficiently large p so that $S < p$.

By the triangle inequality,

$$\sup_{s\in K}\left|\ell^{\mathrm{GF}}_{p,\sigma}(s)-\ell_\alpha(s)\right| \le \sup_{0\le s\le S}\left|\ell^{\mathrm{GF}}_{p,\sigma}(s)-\ell^{\mathrm{RE}}_{p}(s)\right| + \sup_{0\le s\le S}\left|\ell^{\mathrm{RE}}_{p}(s)-\ell_\alpha(s)\right|. \tag{504}$$

For sufficiently large p, $1/p < \varepsilon/3$. Then Lemma 9, specifically the finite-p comparison (494), implies

$$\lim_{\sigma\downarrow 0}\mathbb{P}\left(\sup_{0\le s\le S}\left|\ell^{\mathrm{GF}}_{p,\sigma}(s)-\ell^{\mathrm{RE}}_{p}(s)\right|>\frac{\varepsilon}{3}\right)=0. \tag{505}$$

On the other hand, Proposition 1, namely the large-system Random Edge limit (12), gives

$$\mathbb{P}\left(\sup_{0\le s\le S}\left|\ell^{\mathrm{RE}}_{p}(s)-\ell_\alpha(s)\right|>\frac{2\varepsilon}{3}\right)\longrightarrow 0 \qquad \text{as } p\to\infty. \tag{506}$$

Combining (504)–(506) therefore yields

$$\lim_{p\to\infty}\limsup_{\sigma\downarrow 0}\mathbb{P}\left(\sup_{s\in K}\left|\ell^{\mathrm{GF}}_{p,\sigma}(s)-\ell_\alpha(s)\right|>\varepsilon\right)=0. \tag{507}$$

Finally, substituting the definition (493) of $\ell^{\mathrm{GF}}_{p,\sigma}$ gives exactly (15), completing the proof. □

A5.3 Lower tail of the complex cubic activation clock

We now analyze the lower tail of the scalar activation clock for $\nu = 3$ and complex Gaussian initialization. Let

$$F_3(x) := \mathbb{P}(\tau_3(\mathbf{g})\le x), \qquad \mathbf{g}\sim\mathcal{N}_{\mathbb{C}}(0,I_3). \tag{508}$$

The goal of this subsection is to prove the asymptotic

Proposition 10 (Lower tail of the complex cubic clock)**.** *As $x\downarrow 0$,*

$$F_3(x)=A_3\,\frac{\sqrt{3}}{x}\exp\left(-\frac{3}{x^2}\right)\big(1+o(1)\big), \qquad A_3=\frac{5\sqrt{21\pi}}{27}. \tag{509}$$

The exponential factor comes from the most favorable direction of the microscopic initialization, while the factor x^{-1} and the constant A_3 result from Gaussian fluctuations around this direction.

Radial reduction. Write the complex Gaussian vector in polar form

$$\mathbf{g}=R\boldsymbol{\omega}, \qquad R=\|\mathbf{g}\|, \qquad \|\boldsymbol{\omega}\|=1. \tag{510}$$

For a standard complex Gaussian vector, R and $\boldsymbol{\omega}$ are independent, ω is uniform on the unit sphere in $\mathbb{C}^3$, and R^2 has the Gamma distribution with shape 3 and scale 1. Hence

$$\mathbb{P}(R\ge r)=e^{-r^2}\left(1+r^2+\frac{r^4}{2}\right). \tag{511}$$

By the homogeneity (47), specialized to $\nu=3$, $\tau_3(R\boldsymbol{\omega})=R^{-1}\tau_3(\boldsymbol{\omega})$. Conditioning on $\boldsymbol{\omega}$ therefore gives the exact representation

$$F_3(x)=\mathbb{E}_{\boldsymbol{\omega}}\left[\exp\left(-\frac{\tau_3(\boldsymbol{\omega})^2}{x^2}\right)\left(1+\frac{\tau_3(\boldsymbol{\omega})^2}{x^2}+\frac{\tau_3(\boldsymbol{\omega})^4}{2x^4}\right)\right]. \tag{512}$$

Thus the lower-tail problem reduces to locating and expanding the minimum of $\tau_3(\boldsymbol{\omega})$ on the unit sphere.

The optimal direction. For a unit vector $\boldsymbol{\omega} \in \mathbb{C}^3$, write

$$\rho_a := |\omega^{(a)}|^2, \qquad \rho_1 + \rho_2 + \rho_3 = 1, \qquad \psi := \arg\Big(\omega^{(1)}\omega^{(2)}\omega^{(3)}\Big). \tag{513}$$

Only the three magnitudes and the total product phase ψ affect the scalar clock.

Recall from (44) that, with $B = \mathrm{Im}(\omega^{(1)}\omega^{(2)}\omega^{(3)})$,

$$D_{\boldsymbol{\omega}}(q) = (q+\rho_1)(q+\rho_2)(q+\rho_3) - B^2. \tag{514}$$

For every $q \geq 0$, the arithmetic–geometric mean inequality gives

$$D_{\boldsymbol{\omega}}(q) \leq (q+\rho_1)(q+\rho_2)(q+\rho_3) \leq \left(q + \frac{1}{3}\right)^3. \tag{515}$$

If the initial real part of the product is nonnegative, the clock formula (45) therefore gives

$$\tau_3(\boldsymbol{\omega}) \geq \frac{1}{2}\int_0^\infty \left(q+\frac{1}{3}\right)^{-3/2} dq = \sqrt{3}. \tag{516}$$

If the initial real part is negative, (46) contains an additional positive turning-point contribution, so the same lower bound holds strictly.

Equality in (516) requires equality in (515) for all $q \geq 0$. Hence

$$\rho_1 = \rho_2 = \rho_3 = \frac{1}{3}, \qquad B = 0, \qquad \mathrm{Re}\big(\omega^{(1)}\omega^{(2)}\omega^{(3)}\big) > 0. \tag{517}$$

Equivalently, the minimizing directions have equal magnitudes $|\omega^{(a)}| = 3^{-1/2}$ and total phase $\psi = 0 \pmod{2\pi}$. Thus

$$\min_{\|\boldsymbol{\omega}\|=1} \tau_3(\boldsymbol{\omega}) = \sqrt{3}. \tag{518}$$

Quadratic expansion around the minimizing directions. Set

$$\rho_a = \frac{1}{3} + \delta_a, \qquad \delta_1 + \delta_2 + \delta_3 = 0, \tag{519}$$

and take ψ to be the representative of the total phase near zero. For $r = q + 1/3$,

$$\prod_{a=1}^3 (q+\rho_a) = r^3 - \frac{r}{2}\sum_{a=1}^3 \delta_a^2 + \delta_1\delta_2\delta_3, \tag{520}$$

while

$$B^2 = \rho_1\rho_2\rho_3 \sin^2\psi = \frac{\psi^2}{27} + o\left(\sum_{a=1}^3 \delta_a^2 + \psi^2\right). \tag{521}$$

Consequently,

$$D_{\boldsymbol{\omega}}(q) = r^3 - \frac{r}{2}\sum_{a=1}^3 \delta_a^2 - \frac{\psi^2}{27} + o\left(\sum_{a=1}^3 \delta_a^2 + \psi^2\right), \tag{522}$$

uniformly on compact q-intervals, with the resulting integral remainder also uniform.

In a neighborhood of the minimizers the product has positive real part, so we may use (45). Expanding $D_{\boldsymbol{\omega}}(q)^{-1/2}$ to second order and integrating gives

$$\tau_3(\boldsymbol{\omega}) = \sqrt{3} + \frac{9\sqrt{3}}{20}\sum_{a=1}^3 \delta_a^2 + \frac{\sqrt{3}}{14}\psi^2 + o\left(\sum_{a=1}^3 \delta_a^2 + \psi^2\right). \tag{523}$$

Indeed, the two required integrals are

$$\int_{1/3}^\infty r^{-7/2}\, dr = \frac{2}{5}\, 3^{5/2}, \qquad \int_{1/3}^\infty r^{-9/2}\, dr = \frac{2}{7}\, 3^{7/2}. \tag{524}$$

Squaring (523) yields the form convenient for Laplace asymptotics:

$$\tau_3(\boldsymbol{\omega})^2 = 3 + \frac{27}{10}\sum_{a=1}^3 \delta_a^2 + \frac{3}{7}\psi^2 + o\left(\sum_{a=1}^3 \delta_a^2 + \psi^2\right). \tag{525}$$

Proof of Proposition 10. For a uniform direction $\boldsymbol{\omega}$ on the unit sphere in $\mathbb{C}^3$, the squared magnitudes (ρ_1, ρ_2, ρ_3) have the $\mathrm{Dirichlet}(1,1,1)$ distribution. Thus their density on

$$\Delta_2 := \{(\rho_1, \rho_2) : \rho_1 > 0,\ \rho_2 > 0,\ \rho_1 + \rho_2 < 1\} \tag{526}$$

is the constant 2. The coordinate phases are independent and uniform, so the total product phase ψ is uniform on $[0, 2\pi)$ and independent of the magnitudes. Hence the directional expectation in (512) is integration against

$$\frac{1}{\pi}\, d\rho_1\, d\rho_2\, d\psi. \tag{527}$$

By (518), the exponent $\tau_3(\boldsymbol{\omega})^2$ has its minimum value 3 precisely at $\rho_1 = \rho_2 = \rho_3 = 1/3$ and $\psi = 0 \pmod{2\pi}$. The integral formulas for the clock imply continuity near this minimizing set, and the strict inequality in (516) shows that the contribution outside any fixed neighborhood of it is exponentially smaller than e^{-3/x^2}. It therefore suffices to use the local expansion (525).

Write $\delta_3 = -\delta_1 - \delta_2$. The quadratic part in (525) is

$$Q(\delta_1, \delta_2, \psi) := \frac{27}{10}\left(\delta_1^2 + \delta_2^2 + (\delta_1 + \delta_2)^2\right) + \frac{3}{7}\psi^2. \tag{528}$$

Under the rescaling $\delta_1 = xu_1$, $\delta_2 = xu_2$, $\psi = xv$, the directional volume element contributes a factor x^3. Moreover, in (512), the dominant polynomial term is

$$\frac{\tau_3(\boldsymbol{\omega})^4}{2x^4} = \frac{9}{2x^4}\big(1 + o(1)\big) \tag{529}$$

on the $O(x)$ neighborhood contributing to the Laplace integral. The other two terms in (512) are lower by powers of x.

It follows by the standard Laplace rescaling and dominated convergence that

$$F_3(x) \sim \frac{9}{2\pi x} e^{-3/x^2} \int_{\mathbb{R}^3} e^{-Q(u_1,u_2,v)}\, du_1\, du_2\, dv. \tag{530}$$

The Gaussian integral factors. Since

$$u_1^2 + u_2^2 + (u_1 + u_2)^2 = \begin{pmatrix} u_1 & u_2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \tag{531}$$

and the matrix has determinant 3,

$$\int_{\mathbb{R}^2} \exp\left[-\frac{27}{10}\big(u_1^2 + u_2^2 + (u_1 + u_2)^2\big)\right] du_1\, du_2 = \frac{10\pi}{27\sqrt{3}}. \tag{532}$$

Similarly,

$$\int_{\mathbb{R}} e^{-3v^2/7}\, dv = \sqrt{\frac{7\pi}{3}}. \tag{533}$$

Substituting (532) and (533) into (530) gives

$$F_3(x) \sim \frac{5\sqrt{7\pi}}{9x} e^{-3/x^2}. \tag{534}$$

Finally,

$$\frac{5\sqrt{7\pi}}{9} = \sqrt{3}\,\frac{5\sqrt{21\pi}}{27} = \sqrt{3}\,A_3, \tag{535}$$

which proves (509). □

A5.4 Asymptotic inversion and physical training time

We now use the cubic clock tail from Proposition 10 to prove Theorem 3. For complex modular addition, $\nu = 3$ and $\lambda_p = \sqrt{p}$, so the exact edge exposure corresponding to physical training time t is

$$s_{p,\sigma}(t) = pF_3(\sqrt{p}\,\sigma t).$$

The first step is to asymptotically invert this relation for fixed $O(1)$ exposure levels.

For $s > 0$, set

$$\beta_p(s) := \log p + \frac{1}{2}\log\log p + \log A_3 - \log s. \tag{536}$$

Then the explicit time from (16) can be written as

$$t^{\mathrm{as}}_{p,\sigma}(s) = \frac{\sqrt{3}}{\sigma\sqrt{p}\,\sqrt{\beta_p(s)}}. \tag{537}$$

Lemma 10 (Uniform asymptotic inversion)**.** *For every compact $K \subset (0,\infty)$,*

$$\sup_{s\in K} \left| s_{p,\sigma}\big(t^{\mathrm{as}}_{p,\sigma}(s)\big) - s \right| \longrightarrow 0 \qquad \text{as } p \to \infty. \tag{538}$$

The expression on the left is independent of σ.

Proof. For $\beta_p(s)$ defined in (536), the explicit time (16) satisfies

$$x_p(s) := \sqrt{p}\,\sigma t^{\mathrm{as}}_{p,\sigma}(s) = \frac{\sqrt{3}}{\sqrt{\beta_p(s)}}. \tag{539}$$

In particular, $x_p(s)$ is independent of σ.

If $K \subset (0,\infty)$ is compact, then

$$\sup_{s\in K} \left| \frac{\beta_p(s)}{\log p} - 1 \right| \longrightarrow 0, \tag{540}$$

because $\log s$ is bounded on K. Hence $x_p(s) \to 0$ uniformly for $s \in K$.

We may therefore apply the lower-tail asymptotic (509) uniformly along $x_p(s)$. Thus

$$F_3(x_p(s)) = A_3 \frac{\sqrt{3}}{x_p(s)} \exp\left(-\frac{3}{x_p(s)^2}\right)\big(1 + r_p(s)\big), \tag{541}$$

where

$$\sup_{s\in K} |r_p(s)| \longrightarrow 0. \tag{542}$$

Using

$$\frac{\sqrt{3}}{x_p(s)} = \sqrt{\beta_p(s)}, \qquad \frac{3}{x_p(s)^2} = \beta_p(s),$$

we obtain

$$pF_3(x_p(s)) = pA_3\sqrt{\beta_p(s)}e^{-\beta_p(s)}\big(1 + r_p(s)\big). \tag{543}$$

By the definition of $\beta_p(s)$,

$$e^{-\beta_p(s)} = \frac{s}{pA_3\sqrt{\log p}}. \tag{544}$$

Therefore

$$pF_3(x_p(s)) = s\sqrt{\frac{\beta_p(s)}{\log p}}\big(1 + r_p(s)\big). \tag{545}$$

Since

$$s_{p,\sigma}\big(t^{\mathrm{as}}_{p,\sigma}(s)\big) = pF_3(x_p(s)),$$

it follows from (545) that

$$\begin{aligned} &\left| s_{p,\sigma}\big(t^{\mathrm{as}}_{p,\sigma}(s)\big) - s \right| \\ &\qquad = s\left| \sqrt{\frac{\beta_p(s)}{\log p}}\big(1 + r_p(s)\big) - 1 \right|. \end{aligned} \tag{546}$$

Because K is compact, $\sup_{s\in K} s < \infty$. Moreover, (540) implies

$$\sup_{s\in K} \left| \sqrt{\frac{\beta_p(s)}{\log p}} - 1 \right| \longrightarrow 0,$$

while (542) gives $\sup_{s\in K} |r_p(s)| \to 0$. Taking the supremum over $s \in K$ in (546) therefore yields (538). □

Lemma 10 proves (17). We next transfer the large-system loss trajectory from exact exposure to the explicit physical time.

Proof of Theorem 3. The explicit formula (16) is exactly (537), so it remains to prove the loss convergence.

For notational convenience, define the actual exposure reached at the proposed physical time by

$$\widehat{s}_p(s) := s_{p,\sigma}\big(t^{\mathrm{as}}_{p,\sigma}(s)\big)\,. \tag{547}$$

Although the notation on the right contains σ, the quantity $\widehat{s}_p(s)$ is independent of σ, because $\sqrt{p}\,\sigma t^{\mathrm{as}}_{p,\sigma}(s) = x_p(s)$. By Lemma 10,

$$\sup_{s\in K} |\widehat{s}_p(s) - s| \longrightarrow 0 \tag{548}$$

for every compact $K \subset (0,\infty)$.

Fix such a compact set K. Choose another compact interval $K' \subset (0,\infty)$ containing both K and $\widehat{s}_p(K)$ for all sufficiently large p. Since $\lambda_p = \sqrt{p}$, the normalized loss at $t^{\mathrm{as}}_{p,\sigma}(s)$ can be written as the normalized equal-sector loss at exposure $\widehat{s}_p(s)$. Hence

$$\begin{aligned}
&\sup_{s\in K}\left|\frac{2L\big(U_{p,\sigma}\big(t^{\mathrm{as}}_{p,\sigma}(s)\big)\big)}{p^2} - \ell_\alpha(s)\right| \\
&\qquad \le \sup_{r\in K'}\left|\frac{2L(U_{p,\sigma}(t_{p,\sigma}(r)))}{p^2} - \ell_\alpha(r)\right| + \sup_{s\in K}\left|\ell_\alpha(\widehat{s}_p(s)) - \ell_\alpha(s)\right|.
\end{aligned} \tag{549}$$

The first term is controlled by Theorem 2, specifically the equal-sector large-system convergence (15). The second term converges deterministically to zero by (548) and uniform continuity of ℓ_α on K'. Therefore, for every $\varepsilon > 0$,

$$\lim_{p\to\infty}\limsup_{\sigma\downarrow 0}\mathbb{P}\left(\sup_{s\in K}\left|\frac{2L\big(U_{p,\sigma}\big(t^{\mathrm{as}}_{p,\sigma}(s)\big)\big)}{p^2} - \ell_\alpha(s)\right| > \varepsilon\right) = 0, \tag{550}$$

which is (18) and completes the proof. □

Location and width of the learning transition. The explicit inverse also gives the physical-time scales stated after Theorem 3. For every fixed $s > 0$, $\beta_p(s) \sim \log p$, and therefore

$$t^{\mathrm{as}}_{p,\sigma}(s) \sim \frac{\sqrt{3}}{\sigma\sqrt{p\log p}}, \tag{551}$$

which gives the physical location of every fixed exposure level.

The separation between two fixed exposure levels is smaller. Let $0 < s_- < s_+ < \infty$. Writing

$$b_p := \log p + \frac{1}{2}\log\log p + \log A_3, \tag{552}$$

we have

$$t^{\mathrm{as}}_{p,\sigma}(s_\pm) = \frac{\sqrt{3}}{\sigma\sqrt{p}}\big(b_p - \log s_\pm\big)^{-1/2}. \tag{553}$$

Applying the mean value theorem to $x \mapsto x^{-1/2}$ gives

$$t^{\mathrm{as}}_{p,\sigma}(s_+) - t^{\mathrm{as}}_{p,\sigma}(s_-) = \frac{\sqrt{3}\,\log(s_+/s_-)}{2\sigma\sqrt{p}\,(\log p)^{3/2}}\big(1 + o(1)\big), \tag{554}$$

because every intermediate point between $b_p - \log s_+$ and $b_p - \log s_-$ is asymptotic to $\log p$.

Thus the physical location of a fixed exposure level is of order $(\sigma\sqrt{p\log p})^{-1}$, whereas the separation between two fixed exposure levels is smaller by an additional factor of order $(\log p)^{-1}$. At balanced width, $\alpha = 1$ and $\ell_1(s) = (1+s)^{-1}$, so the exposure level $s = 1$ corresponds to the asymptotic half-loss time.

A6 BEYOND SCALAR SECTORS: REAL MODULAR ADDITION DETAILS [Section 6]

A6.1 FROM COMPLEX FOURIER SECTORS TO REAL FOURIER BLOCKS

We begin by deriving the real Fourier decomposition underlying Section 6. Recall from (10) that

$$T_p^{\mathrm{add}} = \sqrt{p} \sum_{r \in \mathbb{Z}_p} \boldsymbol{\varphi}_r \otimes \boldsymbol{\varphi}_r \otimes \overline{\boldsymbol{\varphi}_r}, \qquad \boldsymbol{\varphi}_r(x) = p^{-1/2} e^{2\pi i r x/p}. \tag{555}$$

Over $\mathbb{C}$, each summand is a scalar rank-one sector. Over $\mathbb{R}$, however, the sectors indexed by r and $-r$ must be combined, because

$$\boldsymbol{\varphi}_{-r} = \overline{\boldsymbol{\varphi}_r}.$$

We now make the resulting real block structure explicit.

For a frequency satisfying $r \not\equiv -r \pmod p$, define the normalized real cosine and sine vectors

$$\mathbf{c}_r := \sqrt{2}\, \mathrm{Re}\, \boldsymbol{\varphi}_r, \qquad \mathbf{s}_r := \sqrt{2}\, \mathrm{Im}\, \boldsymbol{\varphi}_r. \tag{556}$$

Then $\mathbf{c}_r, \mathbf{s}_r$ are orthonormal and

$$\boldsymbol{\varphi}_r = \frac{\mathbf{c}_r + i\mathbf{s}_r}{\sqrt{2}}, \qquad \boldsymbol{\varphi}_{-r} = \frac{\mathbf{c}_r - i\mathbf{s}_r}{\sqrt{2}}. \tag{557}$$

For the three tensor modes, choose the real bases

$$\mathbf{e}_r^{(1)} = \mathbf{e}_r^{(2)} = \mathbf{e}_r^{(3)} := \mathbf{c}_r, \qquad \mathbf{f}_r^{(1)} = \mathbf{f}_r^{(2)} := \mathbf{s}_r, \qquad \mathbf{f}_r^{(3)} := -\mathbf{s}_r. \tag{558}$$

The sign in the third mode reflects the conjugation of the third Fourier factor in (555). With this convention,

$$\boldsymbol{\varphi}_r \otimes \boldsymbol{\varphi}_r \otimes \overline{\boldsymbol{\varphi}_r} = \frac{1}{2\sqrt{2}} \big(\mathbf{e}_r^{(1)} + i\mathbf{f}_r^{(1)}\big) \otimes \big(\mathbf{e}_r^{(2)} + i\mathbf{f}_r^{(2)}\big) \otimes \big(\mathbf{e}_r^{(3)} + i\mathbf{f}_r^{(3)}\big), \tag{559}$$

while the $-r$ term is its complex conjugate. Adding the two terms therefore gives

$$\begin{aligned} &\sqrt{p}\,(\boldsymbol{\varphi}_r \otimes \boldsymbol{\varphi}_r \otimes \overline{\boldsymbol{\varphi}_r} + \boldsymbol{\varphi}_{-r} \otimes \boldsymbol{\varphi}_{-r} \otimes \overline{\boldsymbol{\varphi}_{-r}}) \\ &\qquad = \sqrt{\frac{p}{2}}\, B_r, \end{aligned} \tag{560}$$

where

$$B_r := \mathbf{e}_r^{(1)}\mathbf{e}_r^{(2)}\mathbf{e}_r^{(3)} - \mathbf{e}_r^{(1)}\mathbf{f}_r^{(2)}\mathbf{f}_r^{(3)} - \mathbf{f}_r^{(1)}\mathbf{e}_r^{(2)}\mathbf{f}_r^{(3)} - \mathbf{f}_r^{(1)}\mathbf{f}_r^{(2)}\mathbf{e}_r^{(3)}. \tag{561}$$

Thus every conjugate pair $\{r, -r\}$ becomes a two-dimensional real Fourier block, and after suppressing the mode and frequency indices, (561) is exactly the canonical tensor

$$B = eee - eff - fef - ffe$$

used in (21).

The only exceptions are the self-conjugate frequencies $r \equiv -r \pmod p$. The frequency $r = 0$ is always self-conjugate, and when p is even so is $r = p/2$. Their Fourier vectors are real, so each gives a one-dimensional scalar sector. Let

$$\mathcal{S}_p := \{r \in \mathbb{Z}_p : 2r = 0 \pmod p\}, \tag{562}$$

and let $\mathcal{R}_p$ contain one representative from each remaining pair $\{r, -r\}$. Then the full real decomposition can be written as

$$T_p^{\mathrm{add}} = \sqrt{p} \sum_{r \in \mathcal{S}_p} \boldsymbol{\varphi}_r \otimes \boldsymbol{\varphi}_r \otimes \boldsymbol{\varphi}_r + \sqrt{\frac{p}{2}} \sum_{r \in \mathcal{R}_p} B_r. \tag{563}$$

The blocks in (563) are mutually orthogonal. Indeed, distinct conjugate pairs span orthogonal real Fourier subspaces, while the four rank-one terms in each B_r are themselves mutually orthogonal. Consequently,

$$\|B_r\|_F^2 = 4, \qquad \left\| \sqrt{\frac{p}{2}}\, B_r \right\|_F^2 = 2p. \tag{564}$$

Since $\|T_p^{\mathrm{add}}\|_F^2 = p^2$, every two-dimensional block carries a fraction $2/p$ of the total target energy, whereas each scalar self-conjugate sector carries a fraction $1/p$.

This decomposition makes precise why the use of complex weights was essential in Sections 4–5. Complex weights keep the frequencies r and $-r$ as separate scalar rank-one sectors. Restricting the weights to be real forces them into the coupled block B_r, whose internal rank geometry is analyzed next.

A6.2 Geometry of the canonical two-dimensional block

We now study the canonical tensor

$$B = eee - eff - fef - ffe \in \mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \mathbb{R}^2, \tag{565}$$

where $(\mathbf{e}, \mathbf{f})$ is an orthonormal basis in each mode. As noted in (564),

$$\|B\|_F^2 = 4. \tag{566}$$

The block has a continuous phase symmetry that will account for the family (E_Θ, F_Θ) appearing in Proposition 2.

For $\theta \in \mathbb{R}$, define the rotated basis

$$\mathbf{e}_\theta := \cos\theta\, \mathbf{e} + \sin\theta\, \mathbf{f}, \qquad \mathbf{f}_\theta := -\sin\theta\, \mathbf{e} + \cos\theta\, \mathbf{f}. \tag{567}$$

Then

$$\mathbf{e}_\theta + i\mathbf{f}_\theta = e^{-i\theta}(\mathbf{e} + i\mathbf{f}). \tag{568}$$

Since

$$B = \mathrm{Re}\big[(\mathbf{e} + i\mathbf{f}) \otimes (\mathbf{e} + i\mathbf{f}) \otimes (\mathbf{e} + i\mathbf{f})\big], \tag{569}$$

independent rotations of the three modes leave B unchanged whenever their angles sum to zero modulo 2π.

More precisely, let

$$\Theta = (\theta_1, \theta_2, \theta_3), \qquad \theta_1 + \theta_2 + \theta_3 \equiv 0 \pmod{2\pi}, \tag{570}$$

and define

$$E_\Theta := \mathbf{e}_{\theta_1} \otimes \mathbf{e}_{\theta_2} \otimes \mathbf{e}_{\theta_3}, \tag{571}$$

$$F_\Theta := \mathbf{e}_{\theta_1} \otimes \mathbf{f}_{\theta_2} \otimes \mathbf{f}_{\theta_3} + \mathbf{f}_{\theta_1} \otimes \mathbf{e}_{\theta_2} \otimes \mathbf{f}_{\theta_3} + \mathbf{f}_{\theta_1} \otimes \mathbf{f}_{\theta_2} \otimes \mathbf{e}_{\theta_3}. \tag{572}$$

Expanding (569) in the rotated bases gives

$$B = E_\Theta - F_\Theta. \tag{573}$$

The four rank-one tensors on the right-hand side are mutually orthogonal. Consequently,

$$\|E_\Theta\|_F^2 = 1, \qquad \|F_\Theta\|_F^2 = 3, \qquad \langle E_\Theta, F_\Theta \rangle = 0. \tag{574}$$

We next record two elementary properties of B that will be useful in the rank calculation. Define its spectral norm by

$$\|B\|_{\mathrm{sp}} := \sup_{\|\mathbf{x}_1\|=\|\mathbf{x}_2\|=\|\mathbf{x}_3\|=1} |\langle B, \mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \mathbf{x}_3 \rangle|\,. \tag{575}$$

Lemma 11 (Basic geometry of the real Fourier block)**.** *For the tensor B in (565):*

1. $\|B\|_{\mathrm{sp}} = 1$.
2. *For every unit vector $\mathbf{x}$ in any one of the three modes, the contraction of B against $\mathbf{x}$ is a 2×2 orthogonal matrix. In particular, both of its singular values are equal to 1.*

Proof. Every unit vector in $\mathbb{R}^2$ can be written as

$$\mathbf{x}(\alpha) = \cos\alpha\, \mathbf{e} + \sin\alpha\, \mathbf{f}.$$

A direct expansion of (565) gives

$$\langle B, \mathbf{x}(\alpha) \otimes \mathbf{x}(\beta) \otimes \mathbf{x}(\gamma) \rangle = \cos(\alpha + \beta + \gamma). \tag{576}$$

Its absolute value is at most 1, and equality is attained, for example, at $\alpha = \beta = \gamma = 0$. Hence $\|B\|_{\mathrm{sp}} = 1$.

For the second claim, by symmetry it is enough to contract the first mode. Writing $\mathbf{x}(\alpha) = \cos\alpha\,\mathbf{e} + \sin\alpha\,\mathbf{f}$, the resulting bilinear form has matrix

$$B(\mathbf{x}(\alpha), \cdot, \cdot) = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ -\sin\alpha & -\cos\alpha \end{pmatrix} \tag{577}$$

in the bases $(\mathbf{e}, \mathbf{f})$ of the second and third modes. This matrix satisfies

$$B(\mathbf{x}(\alpha), \cdot, \cdot)^\top B(\mathbf{x}(\alpha), \cdot, \cdot) = I_2,$$

so it is orthogonal and has singular values $1, 1$. □

The phase decomposition (573) already identifies a family of best rank-one approximants to B. Indeed, (566) and Lemma 11 imply that the smallest squared error of a rank-one approximation is

$$\|B\|_F^2 - \|B\|_{\mathrm{sp}}^2 = 4 - 1 = 3, \tag{578}$$

and each E_Θ attains this value, with residual

$$B - E_\Theta = -F_\Theta. \tag{579}$$

The next subsection shows that this residual lies on the boundary of the real rank-two tensors and uses the contraction property above to determine the complete rank cascade.

A6.3 Rank and border-rank cascade

We now prove Proposition 2. Recall that

$$d_q(B)^2 = \inf_{\mathrm{rank}_{\mathbb{R}}(X) \le q} \|B - X\|_F^2.$$

The values for ranks zero and one follow directly from the geometry established above. The rank-two case is more subtle: its optimal error is an infimum that is not attained.

Proof of Proposition 2.

Ranks zero and one. By (566),

$$d_0(B)^2 = \|B\|_F^2 = 4.$$

For a unit rank-one tensor $\mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \mathbf{x}_3$, optimizing also over its scalar coefficient gives squared approximation error

$$\|B\|_F^2 - |\langle B, \mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \mathbf{x}_3 \rangle|^2 .$$

Since Lemma 11 gives $\|B\|_{\mathrm{sp}} = 1$, we obtain

$$d_1(B)^2 = 4 - 1 = 3. \tag{580}$$

Moreover, for every admissible phase Θ from (570), the rank-one tensor E_Θ attains this value. By (573),

$$B - E_\Theta = -F_\Theta, \tag{581}$$

so $-F_\Theta$ is the residual after the first stage of the cascade.

Rank-two lower bound. Let X be any real tensor of CP rank at most two. Write

$$X = \mathbf{a}_1 \otimes \mathbf{b}_1 \otimes \mathbf{c}_1 + \mathbf{a}_2 \otimes \mathbf{b}_2 \otimes \mathbf{c}_2. \tag{582}$$

Choose a unit vector $\mathbf{u}$ in the first mode orthogonal to $\mathbf{a}_1$. Then

$$X(\mathbf{u}, \cdot, \cdot) = \langle \mathbf{u}, \mathbf{a}_2 \rangle\, \mathbf{b}_2 \otimes \mathbf{c}_2$$

has matrix rank at most one. On the other hand, Lemma 11 shows that $B(\mathbf{u}, \cdot, \cdot)$ is a 2×2 orthogonal matrix, whose two singular values are both equal to one. Its Frobenius distance from the set of rank-one matrices is therefore exactly one. Hence

$$\|B - X\|_F \ge \|B(\mathbf{u}, \cdot, \cdot) - X(\mathbf{u}, \cdot, \cdot)\|_F \ge 1. \tag{583}$$

Thus

$$d_2(B)^2 \geq 1. \tag{584}$$

In fact, equality in (583) cannot hold for any rank-two tensor X. To see this, suppose that $\|B - X\|_F = 1$ and write $R := B - X$.

If $\mathbf{a}_1$ and $\mathbf{a}_2$ in (582) are linearly dependent, choose a unit vector $\mathbf{u}$ orthogonal to their common span. Then $X(\mathbf{u}, \cdot, \cdot) = 0$, whereas $B(\mathbf{u}, \cdot, \cdot)$ is orthogonal and therefore has Frobenius norm $\sqrt{2}$. This would give $\|R\|_F \geq \sqrt{2}$, a contradiction.

Suppose instead that $\mathbf{a}_1$ and $\mathbf{a}_2$ are linearly independent, and let $\mathbf{u}_j$ be a unit vector orthogonal to $\mathbf{a}_j$, $j = 1, 2$. Applying (583) with $\mathbf{u}_1$ and $\mathbf{u}_2$ gives

$$\|R(\mathbf{u}_j, \cdot, \cdot)\|_F \geq 1 = \|R\|_F, \qquad j = 1, 2.$$

Contraction by a unit vector cannot increase the Frobenius norm, so equality must hold in both cases. If $\mathbf{v}_j$ is a unit vector orthogonal to $\mathbf{u}_j$, then

$$\|R\|_F^2 = \|R(\mathbf{u}_j, \cdot, \cdot)\|_F^2 + \|R(\mathbf{v}_j, \cdot, \cdot)\|_F^2$$

implies $R(\mathbf{v}_j, \cdot, \cdot) = 0$. Thus the first-mode support of R is contained in $\operatorname{span}\{\mathbf{u}_j\}$ for both $j = 1$ and $j = 2$. Since $\mathbf{u}_1$ and $\mathbf{u}_2$ are not parallel, this forces $R = 0$, again a contradiction. Therefore the rank-two lower bound is never attained.

Rank-two upper bound and border rank. It remains to show that the lower bound can be approached. It is enough to consider $\Theta = 0$, for which

$$E = eee, \qquad F = eff + fef + ffe, \qquad B = E - F. \tag{585}$$

For $\varepsilon \neq 0$, define

$$X_\varepsilon := -\frac{(\mathbf{f} + \varepsilon \mathbf{e})^{\otimes 3} - \mathbf{f}^{\otimes 3}}{\varepsilon}. \tag{586}$$

Each X_ε has real CP rank at most two. Expanding the numerator gives

$$X_\varepsilon = -F - \varepsilon\big(eef + efe + fee\big) - \varepsilon^2 eee, \tag{587}$$

and therefore

$$X_\varepsilon \longrightarrow -F \qquad \text{as } \varepsilon \to 0. \tag{588}$$

Consequently,

$$B - X_\varepsilon \longrightarrow B + F = E,$$

so

$$d_2(B)^2 \leq \|E\|_F^2 = 1. \tag{589}$$

Together with (584), this proves

$$d_2(B)^2 = 1. \tag{590}$$

Since the lower bound is not attained, the infimum in (590) is not attained.

The same construction applies to every admissible phase Θ after rotating the three mode-wise bases. Hence $-F_\Theta$ lies in the closure of the real rank-two tensors, and the corresponding limiting residual is

$$B - (-F_\Theta) = E_\Theta. \tag{591}$$

Degeneration of the rank-two representation. For completeness, we verify that the border-rank approximation above necessarily involves diverging rank-one terms. First observe that F_Θ has real CP rank three. By the phase symmetry, it is enough to consider F from (585). Its first-mode slices are

$$F(\mathbf{e}, \cdot, \cdot) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \qquad F(\mathbf{f}, \cdot, \cdot) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{592}$$

A general matrix in their two-dimensional span is

$$\begin{pmatrix} 0 & y \\ y & x \end{pmatrix},$$

whose determinant is $-y^2$. Hence this slice space contains only one rank-one direction, namely the multiples of $F(\mathbf{e},\cdot,\cdot)$. If F had CP rank at most two, its two-dimensional slice space would be spanned by two rank-one matrices, which is impossible. Thus

$$\operatorname{rank}_{\mathbb{R}}(F_\Theta) = 3. \tag{593}$$

Now let

$$Y_n = A_n + C_n \longrightarrow -F_\Theta,$$

where A_n and C_n are rank-one tensors. If both sequences of rank-one summands were bounded in Frobenius norm, then after passing to a subsequence they would converge to tensors A and C of rank at most one, because the rank-one cone is closed in finite dimension. This would give $-F_\Theta = A+C$ and hence $\operatorname{rank}_{\mathbb{R}}(F_\Theta) \le 2$, contradicting (593).

In fact, both summand norms must diverge. If, for example, $\|A_n\|_F$ did not tend to infinity, then along some subsequence it would remain bounded. Since $Y_n \to -F_\Theta$ is bounded, the identity $C_n = Y_n - A_n$ would make $\|C_n\|_F$ bounded along the same subsequence, contradicting the preceding argument. The same reasoning applies with A_n and C_n interchanged. Therefore

$$\|A_n\|_F \longrightarrow \infty, \qquad \|C_n\|_F \longrightarrow \infty.$$

Thus every rank-two sequence converging to $-F_\Theta$ must degenerate through cancellation of two unbounded rank-one terms. The explicit sequence (586) exhibits this behavior directly, with both summands of order $|\varepsilon|^{-1}$.

Rank three. Finally, the identity

$$B = 2\mathbf{e}^{\otimes 3} - \frac{1}{2}(\mathbf{e}+\mathbf{f})^{\otimes 3} - \frac{1}{2}(\mathbf{e}-\mathbf{f})^{\otimes 3} \tag{594}$$

shows that $\operatorname{rank}_{\mathbb{R}}(B) \le 3$. Since $d_2(B)^2 = 1 > 0$, the rank cannot be at most two. Hence

$$\operatorname{rank}_{\mathbb{R}}(B) = 3, \qquad d_3(B)^2 = 0.$$

Together with (580) and (590), this proves (22) and completes the proof. □

The proposition therefore gives the sequence of infimal residual states

$$B \longrightarrow -F_\Theta \longrightarrow E_\Theta \longrightarrow 0.$$

The second transition is fundamentally different from an ordinary finite-rank approximation step: E_Θ is reached only as the residual of a degenerating rank-two representation. This algebraic feature provides a natural geometric explanation for the large parameter growth observed during the middle stage of the real-sector cascades in Figure 3.