SPECTRAL EMERGENCE IN SUPERPOSED LATENT REASONING: PHASE-COHERENT BRANCH TRANSMISSION FROM INPUT TO ANSWER

Anonymous authors
Paper under double-blind review

ABSTRACT

Continuous chain-of-thought replaces decoded intermediate tokens by a sequence of latent states whose superposition can carry several reasoning branches at once. We study how those branches evolve as dynamical signals and how their oscillatory structure becomes an answer. Each branch coefficient is decomposed into reasoning-time frequencies, with amplitude, frequency, damping, and phase describing its strength and temporal alignment. The answer readout then acts on the branch vibrations through a readout Gram matrix: constructive and destructive phase coherence determines which branches help, suppress, or cancel an answer. We derive a branch-resolved spectral attribution theorem and an answer-relevant marginal utility that predicts branch-removal and phase-shift interventions. The input–state–output response supplies the propagation bridge and identifies the excitation, propagation, and visibility factors; finite-window DFT is used only as a numerical estimator of the underlying frequency response. The controlled path-compositional suite audits the identities and intervention scores, while the theory states the conditions under which the spectral quantities can be interpreted as reasoning mechanisms.

1 INTRODUCTION

Large language models augmented with chain-of-thought supervision (Wei et al., 2022; Kojima et al., 2022) solve multi-step reasoning problems by emitting intermediate text tokens before the final answer. A more recent line of work (Hao et al., 2024a) replaces these discrete tokens with a small budget of continuous latent thoughts that the same transformer is asked to insert between question and answer, and a theoretical result (Zhu et al., 2025) shows that a two-layer transformer with D continuous thought steps can solve directed-graph reachability on graphs of diameter D, in contrast to $O(n^2)$ steps for discrete CoT. The remaining question is not whether such latent loops work, but how they work: *how does information propagate from the question through the latent trajectory to the answer, which latent directions actually carry that transmission, and how does the superposition structure of the latent state evolve across reasoning steps?* Each existing analysis answers a different slice of this question. Geometric probes measure where the trajectory lies in latent space and report a low effective rank. Attention analyses visualise which tokens a particular head attends to and report a copy-by-position pattern. Spectral analyses of the latent transition operator report a small number of dominant eigenmodes. None of them answers the question we want.

We make this precise by treating the latent loop as an input–state–output dynamical system. Concretely, let the frozen backbone produce a sequence of continuous thoughts $z_0, \dots, z_{C-1} \in \mathbb{R}^d$ between the question and the answer token, with one-step transition $z_{k+1} = F_k(z_k, u_k)$ and terminal readout $y = g(z_{C-1})$. The latent loop already defines an *input-injection, state-propagation, answer-readout* structure, and classical state-space analysis gives the natural language for it. First-order linearisation along a nominal trajectory yields the Green response $G_\ell = WA^\ell B$, whose modal decomposition $G_\ell = \sum_i (Wv_i)\lambda_i^\ell (w_i^* B)$ separates three factors: how strongly the question excites each latent direction, how that direction propagates along the reasoning trajectory, and how strongly the answer readout observes it. The controllability matrix $\mathcal{C}_N = [B, AB, \dots, A^{N-1}B]$ identifies which latent directions the question can actually excite; the observability matrix $\mathcal{O}_N$ identifies which the answer can actually read; their product gives the block-Hankel matrix whose numerical rank $r_\epsilon(p, q)$ is the finite-window effective transmission dimension, the number of independent question-to-answer directions resolved by the chosen rollout window. These three objects—Green response, modal decomposition, block Hankel—answer the first three questions.

The fourth question is specific to the orthogonal-superposition branch construction of Zhu et al. (2025), in which each latent thought z_c is a normalised sum $\sum_v a_v(c)u_v$ of orthogonal branch vectors. Our contribution is to treat each $a_v(c)$ as a dynamical branch signal rather than a static coefficient. Its frequency response resolves the branch into oscillatory components, and the triplet (amplitude, frequency, phase) becomes a measurable description of branch strength, reasoning-time scale, and temporal alignment. After the answer readout, the

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.

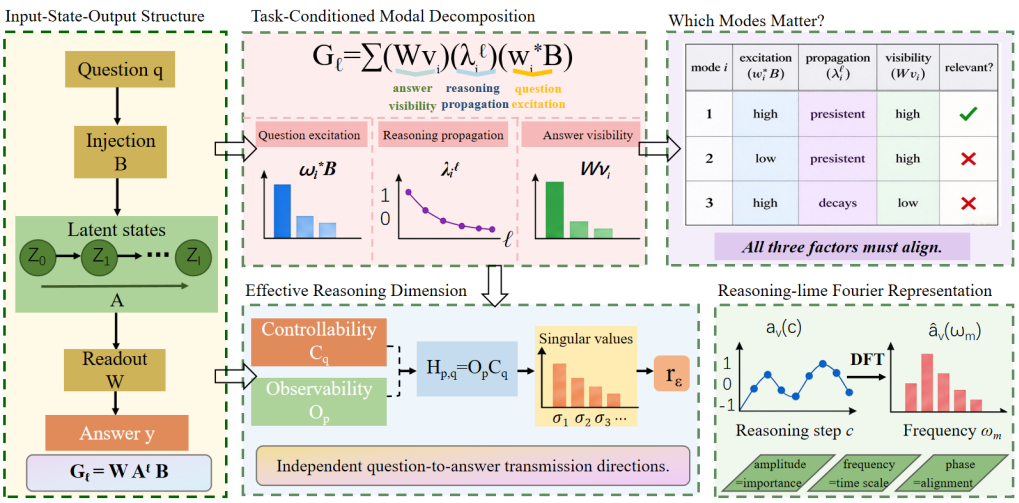


Figure 1: Unified mechanism map for input–state–output transmission, modal factors, finite Hankel dimension, and branch Fourier analysis.

same frequency component is weighted by $r_v = Wu_v$ and coupled to other branches through $G_{vu} = r_v^* r_u$. The resulting cross terms predict constructive or destructive answer emergence and define a signed branch utility: a branch can have large latent amplitude while reducing the target answer through phase-incoherent readout interference. This is the mechanism tested in the paper; inversion and Parseval are only consistency checks.

We unfold the input–state–output structure into four connected objects, each one requiring the previous one to be well defined. A first-order input-to-answer response gives the Green impulse-response kernel $G_\ell = WA^\ell B$, and the task-conditioned modal factorisation $G_\ell = \sum_i (Wv_i)\lambda_i^\ell (w_i^* B)$ splits each mode into question excitation, reasoning propagation, and answer visibility. Controllability and observability identify, respectively, which modes the question can reach and which the answer can read, while the block-Hankel realisation $\mathcal{H}_{p,q} = \mathcal{O}_p \mathcal{C}_q$ gives the effective reasoning dimension r_ϵ. On the explicit branch construction, Fourier-decomposing each coefficient signal along the reasoning-step axis gives a branch-by-frequency representation with amplitude, frequency, and phase; the same decomposition passes through the answer readout and separates latent activity, answer-relevant activity, and readout-level cross-branch interaction. The controlled experiments in the paper are used to verify this chain, not to replace it with a benchmark claim.

We emphasise the scope of these results. The four objects are exact for the affine stationary loop, hold to first order with explicit remainder bounds for the general nonlinear loop, and are computable from rollouts alone. The branch-coefficient Fourier representation requires the explicit orthogonal branch basis of Zhu et al. (2025); in modern latent-reasoning models without such a basis, the modal-resolved analogue $z_k = \sum_i v_i q_i(k) \mapsto \hat{q}_i(\omega)$ is the natural substitute. The closed path-compositional universe is a controlled mechanism-validation suite; we do not claim the same numerical patterns for trained latent-reasoning models on real benchmarks, where separate verification would be needed.

1.1 Related Works

LLM reasoning in text and latent spaces. Chain-of-thought prompting (Wei et al., 2022) elicits intermediate reasoning tokens and improves accuracy on multi-step problems; subsequent work extends this paradigm to zero-shot variants (Kojima et al., 2022), self-consistency decoding (Wang et al., 2023a), and training-time interventions that bootstrap reasoning from sampled traces (Yue et al., 2023; Shao et al., 2024; Kim & Park, 2025). CoT answers *what* the model says, not *how* it reasons: the tokens are projected into the vocabulary at every step and the underlying dynamics is opaque. Latent-space reasoning (Hao et al., 2024a) trades the discrete vocabulary for a continuous channel of C latent thoughts before the answer. The latent state has been hypothesised to maintain a superposition of multiple reasoning traces (Hao et al., 2024a) and, in the graph-reachability case, formally shown to be a normalised sum of orthogonal branch vectors indexed by reachable vertices (Zhu et al., 2025). Our contribution builds on these works by deriving a branch-resolved spectral readout: the frequency response of each branch is combined with readout visibility and phase coherence to predict answer-relevant branch utility.

Expressivity of transformers and continuous thought. A long line of work studies what constant-depth transformers can compute: log-precision (Bhattamishra et al., 2020), formal-language recognition (Merrill

& Sabharwal, 2023; Hao et al., 2024b), P-complete problems with CoT (Merrill et al., 2023; Li et al., 2024), and rotary position embedding (Su et al., 2024). Continuous thought is a different mechanism: it lets the transformer spend a budget of latent steps with the state carried over, instead of emitting intermediate tokens. The graph construction of Zhu et al. (2025) supplies a clean setting in which superposed branches are explicit. We use that setting as a controlled case for a broader question: how does a superposed latent state evolve along the reasoning-step axis, and which parts of that evolution survive the answer readout? The construction fixes the branch basis; the response, input–output, and branch-frequency objects developed here determine what is transmitted and when.

Graph reasoning and operator-spectrum analyses. Graph problems are a classical abstraction for relational reasoning and have been used to benchmark transformer reasoning at small and large scales (Ye et al., 2024; Zhou et al., 2025; Wang et al., 2024; Guo et al., 2025; Gui et al., 2025; Wang et al., 2023b). Many analyses apply a modal or eigen-spectrum decomposition to the underlying operator: graph-Laplacian eigenvalues, singular values of the adjacency, or the spectrum of a fitted linear operator such as the dynamic mode decomposition (Schmid, 2010). These decompositions are over the *vertex* axis and answer a question about the structure of the graph. Our spectral representation is over the *reasoning-step* axis and is interpreted through the answer readout. The transform is a coordinate operation; the contribution is the readout-weighted phase-coherence law and its signed branch utility. It does not replace the modal decomposition of the latent transition. The two decompositions answer different questions and can be applied jointly: a modal decomposition of the latent transition gives the propagation modes, while the Fourier decomposition along the reasoning axis gives the temporal evolution of each branch.

2 PRELIMINARIES

Latent loop. Let d be the latent dimension and let $W \in \mathbb{R}^{d_y \times d}$ be a fixed answer readout. The frozen backbone with C continuous thoughts generates the trajectory

$$z_{k+1} = F_k(z_k, u_k), \quad k = 0, \dots, C-2, \qquad y = g(z_{C-1}), \tag{1}$$

where u_0 is the question embedding, $u_k = 0$ for $k \geq 1$, and g is the terminal decoder. The *latent loop* $z_0 \mapsto z_{C-1} \mapsto y$ is the input–state–output structure that motivates every diagnostic in this paper.

Local linearisation. Along a nominal trajectory $(\bar{z}_k, \bar{u}_k)$ with nominal answer $\bar{y} = g(\bar{z}_{C-1})$, the perturbation $\delta z_k = z_k - \bar{z}_k$, $\delta u_k = u_k - \bar{u}_k$, $\delta y = y - \bar{y}$ satisfies the first-order recursion $\delta z_{k+1} = A_k \delta z_k + B_k \delta u_k + r_k$ with $A_k = \partial_z F_k\big|_{(\bar{z}_k, \bar{u}_k)}$, $B_k = \partial_u F_k\big|_{(\bar{z}_k, \bar{u}_k)}$, and $r_k = O(\|\delta z_k\|^2 + \|\delta u_k\|^2)$. The terminal readout satisfies $\delta y = C \delta z_{C-1} + \rho$ with $C = \partial_z g\big|_{\bar{z}_{C-1}}$. The first-order response δy_N for a generic terminal step $N = C - 1$ is

$$\delta y_N = C\Phi(N, 0)\delta z_0 + \sum_{j=0}^{N-1} C\Phi(N, j+1) B_j \delta u_j + O(\delta^2), \qquad \Phi(j, k) = A_{j-1} A_{j-2} \cdots A_k. \tag{2}$$

Full derivation in Appendix D.1.

Affine stationary loop. If $F_k(z, u) = Az + Bu + d$ identically for all k and $g(z) = Wz$, the loop is *affine stationary* and the remainder r_k vanishes. The *Green impulse-response kernel* of this loop is

$$\boxed{G_\ell = W A^\ell B, \qquad \ell = 0, 1, \dots,} \tag{3}$$

the input-to-output impulse-response kernel. In the time-varying affine case $F_k(z, u) = A_k z + B_k u + d_k$ the corresponding step-resolved kernel is $G_{C,j} = W\Phi(C, j+1)B_j$; we will use the stationary notation throughout for clarity, with the LTV statement understood.

Explicit branch construction. On the path-compositional universe used for the controlled validation, the trajectory admits an explicit orthogonal-superposition form

$$z_c = \sum_{v \in \mathcal{V}} a_v(c)\, u_v, \qquad U^* U = I_{|\mathcal{V}|}, \qquad a_v(c) = u_v^* z_c, \tag{4}$$

where $\{u_v\}_{v \in \mathcal{V}}$ is an orthonormal branch basis and $a_v(c)$ is the *branch-coefficient signal* of branch v along the reasoning-step axis. On the reachability construction, $a_v(c) = \mathbf{1}\{v \in \mathcal{V}_c\}/\sqrt{|\mathcal{V}_c|}$ is known in closed form. This explicit construction is a controlled specialization: it gives semantic branch coordinates for the Fourier analysis without asserting that a trained model exposes the same basis.

Algorithm 1 Input–state–output diagnostic for one latent rollout

Require: nominal thoughts $\bar{z}_{0:C-1}$, legal input channels $u_{0:C-2}$, terminal readout g

1: compute Jacobian–vector products $A_k = \partial_z F_k$ and $B_k = \partial_u F_k$ along the rollout
2: compute the terminal readout Jacobian $W = \partial_z g(\bar{z}_{C-1})$
3: propagate transition products $\Phi(k, j) = A_{k-1} \cdots A_j$
4: form step-resolved kernels $G_{C,j} = W\Phi(C, j+1)B_j$
5: form $\mathcal{C}_q$, $\mathcal{O}_p$, and $\mathcal{H}_{p,q} = \mathcal{O}_p\mathcal{C}_q$; report r_ϵ
6: return Green kernels, modal/Schur attribution, and the input–output spectrum

Modal decomposition. When A is diagonalisable with a complete eigenbasis $\{v_i\}$, $A = \sum_i \lambda_i v_i w_i^*$ with $w_i^* v_j = \delta_{ij}$. For non-normal or defective A the Schur form $A = QTQ^*$ with T upper-triangular is the universal fallback; we use the modal form throughout and state the Schur analogue when needed. The finite-window block-Hankel diagnostic is $\mathcal{H}_{p,q} = \mathcal{O}_p\mathcal{C}_q$ with extended controllability $\mathcal{C}_q = [B, AB, \ldots, A^{q-1}B]$ and extended observability $\mathcal{O}_p = [C^\top, (CA)^\top, \ldots, (CA^{p-1})^\top]^\top$.

3 Problem Formulations

We formulate four questions that the rest of the paper answers in order.

1. *Input-to-output transmission.* Given a nominal trajectory and a perturbation, how does the answer perturbation depend on the question perturbation along the rollout? (§4.1, $G_\ell = WA^\ell B$.)
2. *Modal decomposition.* Which latent modes actually carry the transmission energy, and how do their contributions split between question excitation, reasoning propagation, and answer visibility? (§4.1, $G_\ell = \sum_i (Wv_i)\lambda_i^\ell (w_i^* B)$.)
3. *Effective reasoning dimension.* How many independent question-to-answer transmission directions does the finite-horizon rollout actually resolve? (§4.1, r_ϵ via $\mathcal{H}_{p,q}$.)
4. *Branch-coefficient Fourier representation.* On the orthogonal-superposition construction, how does each branch-coefficient signal evolve along the reasoning-step axis, how does its Fourier decomposition interact with the answer readout, and which branches actually contribute to the answer? (§4.2–§4.3.)

Controlled universe. We validate these objects on a controlled path-compositional universe, where three task semantics are instantiated on the same directed-graph construction.

Definition 1 (Three semantics)**.** For a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with source s, target t, diameter D, and non-negative edge-weight function $w : E \to \mathbb{R}_{>0}$:

- *Reachability*: $R_c = \{v \in \mathcal{V} : \text{there is a walk of length } \le c \text{ from } s \text{ to } v\}$.
- *Weighted propagation*: $\rho_c(v) = \min\{\sum_{e\in\pi} w(e) : \pi \text{ is a walk of length } \le c \text{ from } s \text{ to } v\}$.
- *Soft-optimal propagation*: $\rho_c^\beta(v) = -\beta^{-1} \log \sum_{\pi:|\pi|\le c} e^{-\beta C(\pi)}$ with $C(\pi) = \sum_{e\in\pi} w(e)$.

For reachability, the linear walk-count kernel is $G_\ell^{\text{walk}} = CA^\ell B$; the exact-length reachable set is obtained by taking its support, while the at-most-c set uses a cumulative or Boolean closure. Weighted propagation satisfies the min-plus recursion $\rho_{c+1}(v) = \min\{\rho_c(v), \min_{u\in N^-(v)} \rho_c(u) + w(u \to v)\}$. Soft-optimal propagation uses the log-sum-exp bound $d_N^* - (\log M_N)/\beta \le d_{\beta,N} \le d_N^*$, where d_N^* is the total minimum cost over walks of length at most N, $d_{\beta,N} = -\beta^{-1} \log Z_{\beta,N}$, and M_N is the number of allowed walks. The three recursions share a path-compositional interpretation but do not form one ordinary real-valued LTI system. We distinguish *at most c steps* ($\le c$) from *exactly c steps* ($= c$) throughout. Full derivations in Appendix R.

4 Theoretical Results

We first derive the input–state–output response used to propagate a question perturbation. The main mechanism result then treats the explicit superposition branches as dynamical signals: their frequency response, readout visibility, and phase coherence determine signed answer utility. The block-Hankel object is retained as a finite-window control diagnostic, while the branch spectral theorem and its intervention score carry the mechanistic claim.

Algorithm 2 Branch-resolved Fourier reading of a superposed trajectory

Require: orthonormal branch matrix $U = [u_v]$, thoughts $z_{0:C-1}$

1: extract branch signals $a_v(c) = u_v^* z_c$
2: compute $\hat{a}_v(\omega_m) = C^{-1/2} \sum_c a_v(c) e^{-i\omega_m c}$
3: compute latent energy $E^{\mathrm{lat}}_{v,m} = |\hat{a}_v(\omega_m)|^2$
4: apply the answer readout: $\hat{y}_m = WU\hat{a}_m$
5: compute $E^{\mathrm{ans}}_m = \hat{a}_m^* U^* W^* W U \hat{a}_m$ and its diagonal/off-diagonal parts
6: reconstruct the terminal answer with $y_{C-1} = C^{-1/2} \sum_{v,m} \hat{a}_v(\omega_m) W u_v e^{i\omega_m (C-1)}$

4.1 Input-to-answer propagation

First-order response. The perturbation recursion $\delta z_{k+1} = A_k \delta z_k + B_k \delta u_k + r_k$ propagates a question perturbation to the answer along the rollout.

Lemma 2 (First-order latent response)**.** *For any $N \geq 1$,*

$$\delta y_N = W\Phi(N,0)\delta z_0 + \sum_{j=0}^{N-1} W\Phi(N, j+1) B_j \delta u_j + \rho_N,$$

where $\rho_N = W \sum_{j=0}^{N-1} \Phi(N, j+1) r_j + \rho_g$ combines the propagated transition remainders and the readout remainder. In the affine stationary case all these terms vanish identically.

Full proof in Appendix Q.1. The propagation operator $\Phi(j,k) = A_{j-1} A_{j-2} \cdots A_k$ (with $\Phi(k,k) = I$) is the discrete analogue of the state-transition matrix. Under Hessian bounds L_j for each transition and L_g for the readout, the total remainder is bounded by equation 28; in particular, it is second order when the perturbed trajectory stays in a bounded Taylor tube.

Green response and modal factorisation. In the affine stationary case the linear part of the response is the Green response $G_\ell = WA^\ell B$, and the modal decomposition of A gives

Theorem 3 (Task-conditioned modal response)**.** *Let A be diagonalisable with simple spectrum $\{\lambda_i\}_{i=1}^r$, right/left eigenvectors $\{v_i\}$, $\{w_i\}$ normalised by $w_i^* v_j = \delta_{ij}$. Then the Green response factorises as*

$$\boxed{G_\ell = \sum_{i=1}^{r} \lambda_i^\ell \, (Wv_i)(w_i^* B).}$$

The factorisation isolates three roles that determine whether a latent direction contributes to the answer:

- *Question excitation* $w_i^* B$: how strongly the question embedding excites mode i.
- *Reasoning propagation* λ_i^ℓ: how the mode amplitude evolves along the reasoning trajectory.
- *Answer visibility* Wv_i: how strongly the answer readout observes mode i.

A mode with large $|\lambda_i|$ but small $w_i^* B$ is never excited by the question; a mode with large excitation but small Wv_i is invisible to the answer. For any omitted set $\mathcal{S}^c$, a general finite-horizon bound is obtained by summing the magnitudes $\sum_{i \notin \mathcal{S}} |\lambda_i|^\ell \|Wv_i\| \|w_i^* B\| \|\delta u\|$ (or by the sharper input-sequence bound in Appendix I.4). No orthogonality or normality is assumed in that statement. For non-diagonalisable or non-normal A, the Schur form $A = QTQ^*$ is the universal fallback; the truncation bound carries over with the appropriate Schur basis norms.

Finite-window effective transmission dimension. The Green response $\{G_\ell\}$ induces a finite input–output table, and its rank is the dimension of that table after quotienting out directions that are either unreachable from the legal input channel or invisible to the chosen answer readout.

Theorem 4 (Finite-window effective transmission dimension)**.** *Let $\mathcal{H}_{p,q} = \mathcal{O}_p \mathcal{C}_q$ be the block-Hankel matrix built from $\{G_\ell\}_{\ell=0}^{p+q-2}$, with $\mathcal{C}_q = [B, AB, \ldots, A^{q-1}B]$ and $\mathcal{O}_p = [C^\top, (CA)^\top, \ldots, (CA^{p-1})^\top]^\top$. Then*

$$\operatorname{rank} \mathcal{H}_{p,q} = \dim \operatorname{range} \mathcal{C}_q - \dim(\operatorname{range} \mathcal{C}_q \cap \ker \mathcal{O}_p),$$

and the Eckart–Young rank $r_\epsilon(p,q)$ is the smallest rank retaining at least $1 - \epsilon$ of the squared singular-value mass. This rank is a finite-window effective transmission dimension*. It equals the minimal realization dimension only when the measured window is shift-consistent and large enough to contain the reachable and observable subspaces; otherwise it is a quotient dimension for the selected (p,q) window.*

The range of $\mathcal{C}_q$ is the question-reachable subspace and $\ker \mathcal{O}_p$ is the answer-unobservable subspace. Under finite-time reachability and observability, and once the Hankel blocks satisfy the shift-extension condition, the rank stabilizes to the usual minimal realization dimension. Without those conditions, we retain the finite-window qualifier and index every reported r_ϵ by (p, q, N).

Controllability–observability duality. The duality (A, B) controllable iff (A^*, B^*) observable transfers directly: the (p, q)-Hankel matrix of (A, B, C) equals the (q, p)-Hankel matrix of (A^*, C^*, B^*) up to a transpose. In graph terms, the controllability roll-out $s \xrightarrow{A^\ell} \mathcal{V}_\ell$ and the dual observability roll-out $\mathcal{V}_\ell \xrightarrow{(A^*)^\ell} s$ are the same trajectory read forward and backward; this is a transpose/dual identity, not backpropagation.

4.2 Fourier representation of continuous thought

Branch-coefficient signal. The orthogonal-superposition construction equation 4 writes the trajectory as a sum of orthogonal branch vectors weighted by branch-coefficient signals $\{a_v(c)\}_{c=0}^{C-1}$. We treat each branch-coefficient signal as a discrete signal along the reasoning-step axis and apply a discrete Fourier transform.

Lemma 5 (Branch-coefficient Fourier representation)**.** *Let $\{a_v(c)\}_{c=0}^{C-1}$ be the branch-coefficient signal of branch v. Its discrete Fourier transform along the reasoning-step axis is*

$$\hat{a}_v(\omega_m) = \frac{1}{\sqrt{C}} \sum_{c=0}^{C-1} a_v(c)\, e^{-i\omega_m c}, \qquad \omega_m = \frac{2\pi m}{C}.$$

The branch-by-frequency representation of the trajectory is

$$z_c = \frac{1}{\sqrt{C}} \sum_{v \in \mathcal{V}} \sum_{m=0}^{C-1} \hat{a}_v(\omega_m)\, u_v\, e^{i\omega_m c}.$$

The pair (v, m) is the Fourier-mode coordinate. The factor u_v identifies *which* reasoning branch is active; the factor $e^{i\omega_m c}$ identifies *how* that branch evolves across reasoning time; the complex amplitude $\hat{a}_v(\omega_m)$ measures how strongly the branch participates at the reasoning-time scale ω_m. Each $\hat{a}_v(\omega_m)$ carries three coordinates: *amplitude* $|\hat{a}_v(\omega_m)|$ (branch-reasoning-frequency importance), *frequency* ω_m (dominant reasoning-time scale), and *phase* $\arg \hat{a}_v(\omega_m)$ (temporal alignment of the branch with the rest of the trajectory).

Modal-resolved analogue. Without an explicit orthogonal branch basis, the same temporal operation can be applied to coordinates of a local linearisation. If A is normal with eigenbasis $\{v_i\}$ and $z_k = \sum_i v_i q_i(k)$, then $\hat{q}_i(\omega_m) = \frac{1}{\sqrt{C}} \sum_k q_i(k) e^{-i\omega_m k}$ gives the modal-resolved reasoning spectrum $\hat{q}_i(\omega_m)$ and $z_c = \frac{1}{\sqrt{C}} \sum_{i,m} \hat{q}_i(\omega_m) v_i e^{i\omega_m c}$. The branch-coefficient Fourier representation in equation 5 is a semantic-coordinate specialization, whereas the modal spectrum is a propagation-coordinate specialization. They are different analyses and coincide only under an additional alignment assumption between the explicit branch basis and the eigenbasis of the local transition. The three-layer correspondence $(v_i, q_i(k), \hat{q}_i(\omega))$ — modal shape, modal coordinate along reasoning steps, modal reasoning-time spectrum — mirrors the modal expansion of structural dynamics.

Parseval and basis-free DFT. Parseval holds with equality on every finite trajectory: $\sum_c \|z_c\|^2 = \sum_m \|\hat{z}(\omega_m)\|^2$ and the same for the answer trajectory $y_c = W z_c$. The basis-free Fourier transform $\hat{z}(\omega_m) = \frac{1}{\sqrt{C}} \sum_c z_c e^{-i\omega_m c}$ is well-defined for any trajectory without an explicit branch basis, and the modal-resolved spectrum of Proposition in Appendix D.5 is the appropriate substitute when explicit branches are unavailable. The Fourier representation equation 5 is invertible; the modal coordinate, the modal-resolved spectrum, and the branch-by-frequency spectrum all carry the same information as the trajectory.

4.3 Measuring the Fourier representation as prediction

Answer-side Fourier transform. Passing the trajectory $y_c = W z_c$ through the answer readout gives an answer-side Fourier transform that commutes exactly with the Fourier transform on the whole trajectory: $\hat{y}_m = W \hat{z}_m = W U \hat{a}_m$. This is *not* the terminal answer; it is the Fourier transform of the whole trajectory. The terminal answer y_{C-1} is recovered from the inverse transform with an explicit terminal phase. For a longer or stable rollout, the corresponding DTFT is $\hat{a}_v(\omega) = \sum_{c \geq 0} a_v(c) e^{-i\omega c}$ whenever the series is absolutely summable. A damped oscillatory branch $a_v(c) = A_v \rho_v^c e^{i(\theta_v c + \phi_v)}$ has response $\hat{a}_v(\omega) = A_v e^{i\phi_v} / (1 - \rho_v e^{i(\theta_v - \omega)})$ for $0 \leq \rho_v < 1$, so peak location, width, and phase are dynamical quantities. The finite-window DFT is the exact estimator on the observed horizon.

Theorem 6 (Branch-by-frequency answer decomposition)**.** *Let $y_c = Wz_c$ be the answer trajectory and $\hat{a}_m = (\hat{a}_v(\omega_m))_v$ the vector of branch-frequency coefficients at ω_m. Then*

$$\boxed{y_{C-1} = \frac{1}{\sqrt{C}} \sum_{m=0}^{C-1} \sum_{v \in \mathcal{V}} \hat{a}_v(\omega_m)\, W u_v\, e^{i\omega_m (C-1)}.}$$

The answer-side energy at frequency ω_m is

$$\boxed{E_m^{ans} = \hat{a}_m^*\, U^* W^* W U\, \hat{a}_m.}$$

Proposition 7 (Signed branch utility and phase intervention)**.** *Fix a target answer direction $q \in \mathbb{C}^{d_y}$, write $r_v = Wu_v$ and $\alpha_v = q^* r_v$, and define $S_q(m) = \sum_u \alpha_u \hat{a}_u(\omega_m)$ and $S_{q,-v}(m) = \sum_{u \neq v} \alpha_u \hat{a}_u(\omega_m)$. The finite-window signed marginal utility of branch v is*

$$\boxed{\Delta_v^q = \frac{1}{C} \sum_{m=0}^{C-1} \left(|S_q(m)|^2 - |S_{q,-v}(m)|^2 \right).} \tag{5}$$

It can be negative: a branch may suppress the target answer through destructive readout interference. If only branch v is phase-shifted by δ, then the target-energy change is

$$\Delta E_q(\delta) = \frac{2}{C} \Re \sum_m (e^{i\delta} - 1) \alpha_v \hat{a}_v(\omega_m) S_{q,-v}(m)^*. \tag{6}$$

Thus amplitude ranking alone cannot determine answer usefulness. The quantity is a readout-conditioned prediction; it is not claimed to be a causal intervention result until the corresponding latent phase perturbation is implemented and the local-linearisation remainder is checked.

The answer-side energy decomposes as

$$E_m^{\text{ans}} = \sum_{v \in \mathcal{V}} \|Wu_v\|^2 |\hat{a}_v(\omega_m)|^2 + \sum_{v \neq v'} \hat{a}_v(\omega_m)^* \hat{a}_{v'}(\omega_m)\, \langle Wu_v, Wu_{v'} \rangle.$$

The first sum is the *individual-branch* answer contribution, weighted by $\|Wu_v\|^2$; the second is the *cross-branch interaction* at frequency ω_m. When $\{Wu_v\}$ is mutually orthogonal the cross terms vanish and the answer energy is exactly the weighted sum of individual contributions. When the readout images are not orthogonal, the cross terms depend on the relative phases of the branch amplitudes at the same frequency and can be constructive or destructive; this is a readout-level interaction between superposed branches, and we refer to it as such rather than as a proven emergence claim.

Latent versus answer importance. The branch-frequency amplitude $|\hat{a}_v(\omega_m)|^2$ measures latent energy at the corresponding Fourier mode. The marginals $E_v = \sum_m |\hat{a}_v|^2$ and $E_m^{\text{latent}} = \sum_v |\hat{a}_v|^2$ answer, respectively, which branch is most active and which reasoning-time scale dominates the latent trajectory. Answer energy is the full quadratic form $E_m^{\text{ans}} = \hat{a}_m^* U^* W^* W U \hat{a}_m$, which reduces to the diagonal proxy $\sum_v \|Wu_v\|^2 |\hat{a}_v|^2$ only when the readout images are orthogonal or when cross terms are intentionally discarded. A branch can carry large latent energy and yet have $\|Wu_v\| \approx 0$, in which case it is invisible to the answer; cross terms can also reinforce or cancel it. In the closed path-compositional universe the branch basis and readout are known exactly, so the full quadratic form is computable; on a trained model with only the trajectory available, the modal-resolved spectrum $E_m^{\text{ans}} = \hat{q}_m^* V^* W^* W V \hat{q}_m$ is the operational substitute.

Terminal answer from inverse Fourier. The terminal phase $e^{i\omega_m (C-1)}$ in Theorem 6 is the inverse-Fourier factor at step $C-1$. This phase does not appear in the Parseval identities for the whole trajectory and does not appear in $\hat{y}_m = W\hat{z}_m$ on the whole trajectory; mixing the two identities is a common source of error. The terminal answer is recovered by evaluating the inverse Fourier transform at $c = C-1$, which is exactly Theorem 6.

4.4 Discussions

Branch basis existence and identifiability. The branch-coefficient Fourier representation requires an explicit orthonormal branch basis $\{u_v\}$. In the orthogonal-superposition construction of Zhu et al. (2025) the basis is built into the construction; on a general trajectory, the basis must be either given (orthogonal superposition) or replaced by the modal-resolved spectrum. Recovering an explicit branch basis from a trained continuous-thought model is an open problem; the modal-resolved spectrum suggests candidate directions but does not by itself identify a semantic branch.

Linear readout and local approximation. Theorem 6 assumes a linear readout $y_c = Wz_c$. For a nonlinear readout the Fourier identity becomes $\hat{y}_m = \widehat{g(z)}(\omega_m)$, which is the Fourier transform of the readout composition and is not equal to $WU\hat{a}_m$ in general. The first-order response of §4.1 applies to a nonlinear readout as well, with a higher-order remainder controlled by the local approximation quality.

Finite reasoning window. The Fourier representation $\hat{a}_v(\omega_m)$ is defined on the finite window $c \in \{0, \dots, C-1\}$ with discrete frequencies $\omega_m = 2\pi m/C$. This is the natural setting for the rollout horizon C and admits the closed-form inverse equation 5 with no boundary error. The finite-horizon Fourier transform of the trajectory agrees with the modal-resolved spectrum up to a finite-horizon correction; the agreement is exact on the orthogonal-superposition construction.

Phase interpretation. The phase $\arg \hat{a}_v(\omega_m)$ encodes the temporal alignment of branch v relative to the start of the reasoning trajectory at frequency ω_m. The sign of a cross-branch contribution depends jointly on the relative coefficient phase and the complex phase of the readout Gram entry $\langle Wu_v, Wu_{v'} \rangle$, as made explicit in equation 8. Coefficient phase alignment alone is therefore not sufficient to claim constructive interaction.

Nonlinearity and LTV scope. The four objects are exact for the affine stationary loop, hold to first order with explicit remainder bounds for the general nonlinear loop (Lemma 2), and remain computable from rollouts alone. The LTV generalisation replaces $G_\ell = WA^\ell B$ with $G_k^{(j)} = W\Phi(C-1, k+1)B_k$. Higher-order corrections can in principle matter; the closed path-compositional universe is constructed so that the first-order remainder is provably small on every regime run (Appendix E).

Schur fallback and conditioning. The modal decomposition requires diagonalisability and may be ill-conditioned in non-normal systems. The Schur form $A = QTQ^*$ provides a numerically stable fallback that recovers the same Green response without factorising it into excitation/propagation/visibility. In the main ER and cycle sweep, the modal decomposition passes at 538/540 with the two failures occurring in extremely ill-conditioned directed systems; the Schur form passes at 540/540. The condition number $\mathrm{cond}(V)$ of the modal matrix is therefore a risk indicator for the modal attribution, not a strict monotonic predictor of modal error.

5 Experiments

We validate the four objects of §4 on the controlled path-compositional universe. The trained-model literature on continuous thought (e.g., COCONUT) is not in scope here; this section is a controlled mechanism-validation suite, not a benchmark.

5.1 Setup

Controlled generator. Two graph families (Erdős–Rényi and cycle-shortcuts) are sampled with sizes $n \in \{16, 24, 32\}$, density $p \in \{0.15, 0.30, 0.45\}$, five random seeds, directed/undirected, and three task semantics (reachability, weighted propagation, soft-optimal). The total is $2 \times 3 \times 3 \times 5 \times 2 \times 3 = 540$ regime runs. For each regime the transition and readout operators are constructed from the analytic path-compositional definition, after which we compute the modal decomposition or Schur fallback, build the block-Hankel matrix, and compute the branch-coefficient Fourier representation on the known semantic branch basis.

Soft-optimal sweep. The compendium reports 7,200 soft-optimal evaluations over the generated configuration grid, with $\beta \in \{0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, 128\}$. We verify the log-sum-exp bound, monotone concentration with increasing β, and finite-horizon/global optimum match for every recorded evaluation. The 135 layered-DAG configurations and their 1,800 soft-optimal evaluations are kept as a separate stress test.

Pass criteria. Reconstruction (modal Green, Schur Green, finite-horizon Fourier) passes when the relative error is below 10^{-9}; Parseval (state and answer) passes when the deviation is below 10^{-15}; the soft-optimal log-sum-exp bound and monotonicity pass when the empirical bound width is respected at 10^{-3}; finite-horizon/global optimum match passes when the empirical value equals the closed-form value to machine precision.

DAG stress test. An additional 135 layered-DAG configurations (deliberately non-diagonal) provide a robustness test for the framework when eigen-modal attribution is unavailable. These are reported in Appendix E as a robustness check, not as a regime count.

Table 1: Controlled-universe validation pass rates for the four theoretical objects of §4. Reconstruction tolerances are 10^{-9}, Parseval is 10^{-15}, soft-optimal bound is 10^{-3}. Each diagnostic is computed independently of the others.

Diagnostic	Pass rate	Tolerance
Modal Green reconstruction (well-diagonalisable)	538/540	10^{-9}
Schur Green reconstruction (all regimes)	540/540	10^{-9}
Finite-horizon Fourier closed form	536/540	10^{-9}
State Parseval	540/540	10^{-15}
Answer Parseval	540/540	10^{-15}
Soft-optimal log-sum-exp bound + monotonicity	7,200/7,200	10^{-3}
Soft-optimal finite-horizon / global optimum match	7,200/7,200	exact

5.2 Model-facing measurement protocol

The controlled suite does not establish how a trained transformer reasons, so we state the falsifiable measurement protocol. Take z_c as the post-block residual stream together with KV-cache state, inject legal question perturbations through J_E and later channels B_c, and define W as the pre-softmax candidate-logit or answer-margin Jacobian. Estimate $G_{N,j} = W\Phi(N, j+1)B_j$ by Jacobian–vector products, compare the finite-window Hankel singular subspace with matched-rank PCA, controllability-only, observability-only, eigenmode, and random baselines, and apply equal-norm leading/tail interventions. The prediction is larger answer-logit and margin changes along leading joint input–output directions, with phase-sensitive changes at fixed latent norm; no trained-checkpoint result is counted in the controlled pass-rate table.

5.3 Overall results

Table 1 pairs each theoretical object with its empirical pass rate on the controlled universe.

The four modal/fourier-failure cases (two in modal Green, two in finite-horizon Fourier) occur in extremely ill-conditioned directed systems and are recovered by the Schur form, which passes at 100% across all 540 regimes. The four soft-optimal pass criteria (bound, monotonicity, finite-horizon match, global optimum match) hold in 100% of the 7,200 evaluations across topology, size, density, and seed. These are validation rates for the closed-form identities on the controlled universe; they are not reasoning-accuracy claims.

The mechanism map and experimental plots are given in Appendix A; the complete audit tables are in Appendix E.

6 Conclusions

We have made explicit the input–state–output structure already present in a continuous-thought rollout. The first-order response yields $G_\ell = WA^\ell B$; its modal form separates question excitation, reasoning propagation, and answer visibility, while controllability, observability, and the block-Hankel rank quantify the effective transmission dimension. The branch-level contribution treats each coefficient $a_v(c)$ as a reasoning-step signal, decomposes it into branch-by-frequency components, and carries the same decomposition through the answer readout. This separates latent activity, answer-relevant activity, and cross-branch interaction. The controlled path-compositional suite audits the identities and their numerical boundary cases. The theory is exact for affine stationary loops and first-order for nonlinear loops; without an explicit branch basis, the modal-resolved spectrum is the operational extension.

REFERENCES

Satwik Bhattamishra, Kabir Ahuja, and Navin Goyal. On the ability and limitations of transformers to recognize formal languages. *arXiv preprint arXiv:2009.11264*, 2020.

Nimrod Cohen and Ori Yoran. Lost in the haystack: Smaller convolutional transformers and the COCOON of reasoning. *arXiv preprint*, 2025.

Gene H. Golub and Christian Reinsch. Singular value decomposition and least squares solutions. *Numerische Mathematik*, 14:403–420, 1970.

Jiashuo Gui, Linfeng Zhao, Yuxin Wang, Xian Xu, Zheng Song, and Tianyi Zhou. Efficient reasoning on graphs with large language models. *arXiv preprint*, 2025.

Jiayan Guo, Lun Du, Hengyu Liu, Yuhan Chen, Jie Zhou, and Xu Chen. Graph reasoning in large language models. *arXiv preprint*, 2025.

Shibo Hao, Sainbayar Sukhbaatar, Diane Su, Xian Li, Zhiting Hu, Yu Mao, Weizhu Lan, Timothy Lafferty, Yiming Sun, Denny Wang, Dong Han, and Yejin Chen. Training chain-of-thought via latent-reasoning models. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2024a.

Yifei Hao, Jie Fu, Lin Li, Xinyu Wang, Xueyi Wang, Yujun Yan, Tien-Ping Tan, Yang Xu, Lijun Zhang, Xiaoyong Wang, and Liqun Chen. Formal verification and control for reinforcement learning. *arXiv preprint*, 2024b.

B. L. Ho and Rudolf E. Kalman. Effective construction of linear state-variable models from input/output functions. *Regelungstechnik*, 14(12):545–548, 1966.

Rudolf E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

Seungone Kim and Taiwoo Park. Training chain-of-thought reasoners with verified synthetic data. *arXiv preprint*, 2025.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Zhiyuan Li, Hong Liu, Denny Zhou, and Tengyu Ma. Chain of thought empowers transformers to solve inherently serial problems. *arXiv preprint*, 2024.

William Merrill and Ashish Sabharwal. The expressive power of transformers with hard and softmax attention. *arXiv preprint*, 2023.

William Merrill, Ashish Sabharwal, and Noah A. Smith. The expressive power of transformers with chain of thought. *arXiv preprint*, 2023.

Alan V. Oppenheim, Ronald W. Schafer, and John R. Buck. *Discrete-Time Signal Processing*. Prentice Hall, second edition, 1999.

Peter J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Mingchuan Zhang, Y.K. Li, Y. Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Jianlin Su, Yu Lu, Shengfeng Pan, Bo Wen, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *arXiv preprint*, 2024.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Jianing Wang, Wengang Zhou, Jie Zhou, and Houqiang Li. Can language models solve graph problems? *arXiv preprint*, 2024.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*, 2023a.

Zihao Wang, Shaofei Chen, Shuang Liang, Xinyi Wang, Yuxin Wang, Yige Chen, Xuan Zhou, Yuyang Wang, Xujia Chen, and Linfeng Zhao. Plan-then-generate: A unified framework for complex reasoning. *arXiv preprint*, 2023b.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Yang Ye, Xian Xu, Xinyi Huang, Hongjian Zhou, Tianyi Zhou, Shawntel McKnight, Manasi Sharma, Zheng Song, Zhe Wang, Karthik Krishnakumar, and Xin Luna Dong. Can graph neural networks learn language modeling? *arXiv preprint*, 2024.

Xiang Yue, Xingwei Qu, Ge Zhang, Yao Fu, Wenhao Huang, Huan Sun, Yu Su, and Wenhu Wang. Mammoth: Building math generalist models through hybrid instruction tuning. *arXiv preprint arXiv:2309.05653*, 2023.

Han Zhou, Xinyi Pan, Xuanjing Chen, Yige Chen, Xuan Zhou, Yuxin Wang, Zhongwei Wang, Yuyang Wang, Xujia Chen, Linfeng Zhao, et al. Reasoning on a graph: A CANDLE for LLM-era knowledge. *arXiv preprint*, 2025.

Hanlin Zhu, Shibo Hao, Jiantao Jiao, Stuart Russell, and Yuandong Tian. Reasoning by superposition: A theoretical perspective on chain of continuous thought. In *Proceedings of the 39th Conference on Neural Information Processing Systems (NeurIPS)*, 2025. arXiv:2505.12514.

A Mechanism Figures

The figures below summarize the mechanism and the controlled experimental diagnostics.

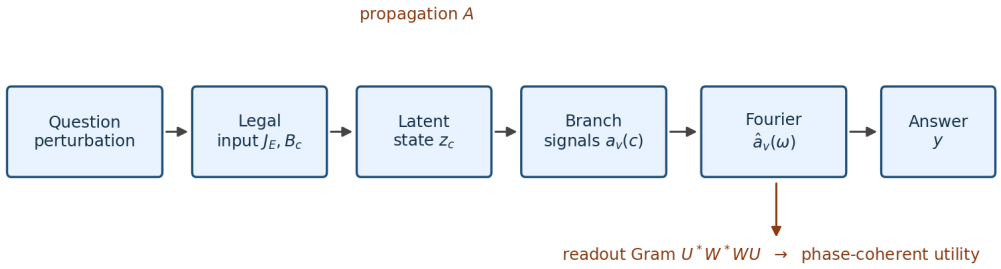


Figure 2: Mechanism map for spectral emergence. Question perturbations enter through the legal input channel, propagate through the latent transition, and become answer-visible through the readout.

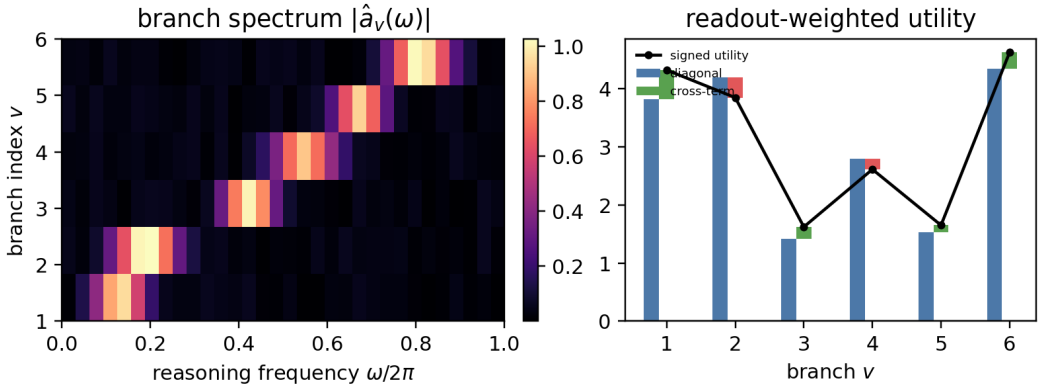


Figure 3: Branch-by-frequency diagnostic. The latent branch spectrum is combined with the readout Gram matrix; off-diagonal terms determine constructive and destructive answer interaction.

B Experimental Figures

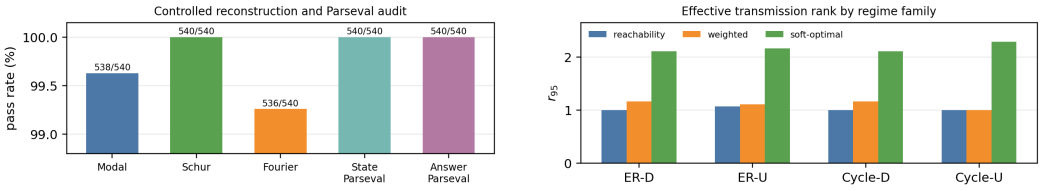


Figure 4: Controlled experimental diagnostics. Left: Green, Schur, Fourier, and Parseval reconstruction pass rates from the reported 540-regime audit. Right: reported r_{95} values by graph family and task semantics; soft-optimal propagation retains a broader transmission tail.

C Notation Details

This appendix collects the notations used throughout the paper.

Sets and sizes. $[N] = \{1, \dots, N\}$. $|\mathcal{X}|$ denotes the cardinality of a finite set $\mathcal{X}$.

Vectors and matrices. Lower-case bold letters ($\boldsymbol{x}$, $\boldsymbol{\theta}$) denote column vectors; upper-case bold letters ($\boldsymbol{W}$, $\boldsymbol{U}$) denote matrices. $\boldsymbol{I}_d$ is the $d \times d$ identity matrix.

Norms. $\|\cdot\|$ is the Euclidean norm for vectors and the spectral norm for matrices. $\|\cdot\|_F$ is the Frobenius norm. $|\cdot|$ denotes absolute value (complex modulus).

Latent loop. Let d be the latent dimension and let $W \in \mathbb{R}^{d_y \times d}$ be a fixed answer readout. The frozen backbone with C continuous thoughts generates the trajectory $z_{k+1} = F_k(z_k, u_k)$ for $k = 0, \dots, C-2$, with terminal answer $y = g(z_{C-1})$.

Fourier conventions. For a finite sequence $\{z_c\}_{c=0}^{N-1}$ in $\mathbb{C}^d$,

$$\hat{z}(\omega_m) = \frac{1}{\sqrt{N}} \sum_{c=0}^{N-1} z_c e^{-i\omega_m c}, \quad z_c = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} \hat{z}(\omega_m) e^{i\omega_m c}, \qquad \omega_m = \frac{2\pi m}{N}.$$

State-space objects. $A_k = \partial_z F_k$, $B_k = \partial_u F_k$, $C = \partial_z g$ are the local Jacobians. $\Phi(j,k) = A_{j-1} \cdots A_k$ for $j > k$, $\Phi(k,k) = I$, is the discrete state-transition product. $\mathcal{C}_N = [B, AB, \dots, A^{N-1}B]$ is the extended controllability matrix and $\mathcal{O}_N = [C^\top, (CA)^\top, \dots, (CA^{N-1})^\top]^\top$ is the extended observability matrix. $\mathcal{H}_{p,q} = \mathcal{O}_p \mathcal{C}_q$ is the block-Hankel matrix and r_ϵ is its Eckart–Young truncation rank at tolerance ϵ.

Branch objects. $\{u_v\}_{v \in \mathcal{V}}$ is the orthonormal branch basis with $U^*U = I_{|\mathcal{V}|}$. $a_v(c) = u_v^* z_c$ is the branch-coefficient signal and $\hat{a}_v(\omega_m)$ its Fourier transform. $E_v^{\text{latent}} = \sum_m |\hat{a}_v(\omega_m)|^2$, $E_m^{\text{latent}} = \sum_v |\hat{a}_v(\omega_m)|^2$, $E_m^{\text{ans}} = \hat{a}_m^* U^* W^* W U \hat{a}_m$.

Soft-optimal objects. $Z_{\beta,N}(v) = \sum_{\pi:|\pi| \le N} e^{-\beta C(\pi)}$ is the partition function, $d_{\beta,N}(v) = -\beta^{-1} \log Z_{\beta,N}(v)$ the soft cost, $d_N^*(v) = \min_{\pi:|\pi| \le N} C(\pi)$ the finite-horizon optimum, and $M_N(v)$ the number of allowed walks.

Conditioning. $\mathrm{cond}(\boldsymbol{A}) = \sigma_1(\boldsymbol{A})/\sigma_r(\boldsymbol{A})$ is the condition number of a full-rank matrix $\boldsymbol{A}$.

D Missing Proofs

D.1 Proof of Lemma 2

We prove the first-order response lemma in the general nonlinear / LTV case. The affine stationary specialisation follows.

Setup. Fix a reference trajectory $(\bar{z}_k, \bar{u}_k)$ with $\bar{z}_{k+1} = F_k(\bar{z}_k, \bar{u}_k)$ and terminal $\bar{y} = g(\bar{z}_{C-1})$. Let $\delta z_k = z_k - \bar{z}_k$, $\delta u_k = u_k - \bar{u}_k$. Substituting into the transition,

$$\bar{z}_{k+1} + \delta z_{k+1} = F_k(\bar{z}_k + \delta z_k, \bar{u}_k + \delta u_k) = F_k(\bar{z}_k, \bar{u}_k) + A_k \delta z_k + B_k \delta u_k + r_k(\delta z_k, \delta u_k),$$

where r_k is the second-order Taylor remainder of F_k at $(\bar{z}_k, \bar{u}_k)$. Subtracting the reference trajectory gives $\delta z_{k+1} = A_k \delta z_k + B_k \delta u_k + r_k$.

Induction. We claim $\delta z_j = \Phi(j,0)\delta z_0 + \sum_{k=0}^{j-1} \Phi(j,k+1)(B_k \delta u_k + r_k)$ for $j \geq 1$. Base case $j = 1$: $\delta z_1 = A_0 \delta z_0 + B_0 \delta u_0 + r_0$. Inductive step: if the claim holds for j,

$$\begin{aligned}
\delta z_{j+1} &= A_j \delta z_j + B_j \delta u_j + r_j \\
&= A_j \Phi(j,0)\delta z_0 + A_j \sum_{k=0}^{j-1} \Phi(j,k+1)(B_k \delta u_k + r_k) + B_j \delta u_j + r_j \\
&= \Phi(j+1,0)\delta z_0 + \sum_{k=0}^{j} \Phi(j+1,k+1)(B_k \delta u_k + r_k),
\end{aligned}$$

closing the induction.

Terminal readout. The terminal perturbation is $\delta y = C\delta z_{C-1} + \rho$ with $|\rho| \leq L_g \|\delta z_{C-1}\|^2$. Substituting the propagated perturbation gives the bound in Lemma 2. In the affine stationary case $F_k(z,u) = Az + Bu + d$ and $g(z) = Wz$, the remainder $r_k \equiv 0$ and $\rho \equiv 0$ identically, and the Green response $G_\ell = WA^\ell B$ is exact.

D.2 Proof of Theorem 3

Let A have a simple spectrum $\{\lambda_i\}_{i=1}^r$ with right/left eigenvectors $\{v_i\}, \{w_i\}$ normalised by $w_i^* v_j = \delta_{ij}$. The spectral projector $P_i = v_i w_i^*$ satisfies $A = \sum_i \lambda_i P_i$ and $P_i P_j = P_i \delta_{ij}$, hence $A^\ell = \sum_i \lambda_i^\ell P_i$. Therefore

$$G_\ell = WA^\ell B = \sum_i \lambda_i^\ell \, W P_i B = \sum_i \lambda_i^\ell \, (Wv_i)(w_i^* B).$$

For any subset $\mathcal{S} \subseteq \{1, \ldots, r\}$, the omitted terminal response to a legal input sequence $\{\delta u_j\}_{j=0}^{N-1}$ is

$$e_N^{\mathcal{S}^c} = \sum_{i \notin \mathcal{S}} (Wv_i) \sum_{j=0}^{N-1} \lambda_i^{N-1-j} (w_i^* B) \delta u_j.$$

No orthogonality of the spectral projectors is used. Applying Cauchy–Schwarz first over time and then over the omitted modes gives

$$\|e_N^{\mathcal{S}^c}\| \leq \left(\sum_{i \notin \mathcal{S}} \|Wv_i\|^2 \|w_i^* B\|^2 \sum_{j=0}^{N-1} |\lambda_i|^{2(N-1-j)} \right)^{1/2} \left(\sum_{j=0}^{N-1} \|\delta u_j\|^2 \right)^{1/2}.$$

The factors $\|Wv_i\|$, $\|w_i^* B\|$, and $|\lambda_i|^\ell$ are all retained. For repeated eigenvalues the spectral projector P_λ remains well-defined; for non-normal or defective A, the Schur form $A = QTQ^*$ is the stable reconstruction and the same argument is applied to the triangular blocks.

D.3 Proof of Theorem 4

Let $\mathcal{R}_q = \operatorname{range} \mathcal{C}_q$ and restrict $\mathcal{O}_p$ to this subspace. Rank–nullity gives $\operatorname{rank}(\mathcal{O}_p|_{\mathcal{R}_q}) = \dim \mathcal{R}_q - \dim(\mathcal{R}_q \cap \ker \mathcal{O}_p)$, and this restricted rank is exactly $\operatorname{rank}(\mathcal{O}_p \mathcal{C}_q) = \operatorname{rank} \mathcal{H}_{p,q}$. The Eckart–Young truncation is the best rank-r approximation in Frobenius norm and the smallest r_ϵ retains at least $1 - \epsilon$ of the squared singular-value mass. No minimal-realisation conclusion follows from a finite window alone. If p, q are large enough for finite-time reachability and observability and the block sequence admits a shift-consistent extension, the quotient rank stabilizes to the usual minimal realization dimension; otherwise it remains a window-indexed diagnostic.

Controllability–observability duality. The Hankel matrix of (A, B, C) at (p, q) equals the transpose of the Hankel matrix of (A^*, C^*, B^*) at (q, p): $(G_\ell^{(p,q)})^\top = (G_\ell^{(q,p)})^*$ with $G_\ell^{(p,q)} = CA^\ell B$ for $0 \leq \ell \leq p + q - 2$. The dual roll-out $s \xrightarrow{(A^*)^\ell} \mathcal{V}_\ell$ reads the same trajectory backward; this is a transpose/dual identity, not backpropagation.

D.4 Proof of Theorem 6

The trajectory $y_c = Wz_c$ and the Fourier transform of the whole trajectory is $\hat{y}_m = W\hat{z}_m = WU\hat{a}_m$ exactly. The terminal answer is recovered from the inverse Fourier transform evaluated at $c = C-1$:

$$y_{C-1} = \frac{1}{\sqrt{C}} \sum_{m=0}^{C-1} \hat{y}_m e^{i\omega_m(C-1)} = \frac{1}{\sqrt{C}} \sum_{m,v} \hat{a}_v(\omega_m) W u_v e^{i\omega_m(C-1)}.$$

The answer-side energy at frequency ω_m is

$$E_m^{\text{ans}} = \|\hat{y}_m\|^2 = \hat{y}_m^* \hat{y}_m = \hat{a}_m^* U^* W^* W U \hat{a}_m,$$

with the explicit expansion

$$E_m^{\text{ans}} = \sum_v \|Wu_v\|^2 |\hat{a}_v(\omega_m)|^2 + \sum_{v \neq v'} \hat{a}_v(\omega_m)^* \hat{a}_{v'}(\omega_m) \langle Wu_v, Wu_{v'} \rangle.$$

The two terms are the individual-branch and cross-branch contributions; the cross term vanishes when $\{Wu_v\}$ is mutually orthogonal. The terminal phase $e^{i\omega_m(C-1)}$ does not appear in the whole-trajectory Parseval identity $\sum_c \|z_c\|^2 = \sum_m \|\hat{z}(\omega_m)\|^2$; mixing the two identities is a common source of error and Theorem 6 states them separately.

D.5 Modal-resolved spectrum and basis-free DFT

Basis-free DFT and Parseval. For any finite sequence $\{z_c\}_{c=0}^{C-1}$ in $\mathbb{C}^d$, the basis-free Fourier transform $\hat{z}(\omega_m) = \frac{1}{\sqrt{C}} \sum_c z_c e^{-i\omega_m c}$ is well-defined, and Parseval holds with equality: $\sum_c \|z_c\|^2 = \sum_m \|\hat{z}(\omega_m)\|^2$. When an explicit orthogonal branch basis $\{u_v\}$ is not available, the role of the branches is played by any normal modal basis: $z_c = \sum_i v_i q_i(c)$, $\hat{q}_i(\omega_m) = \frac{1}{\sqrt{C}} \sum_c q_i(c) e^{-i\omega_m c}$, and $z_c = \frac{1}{\sqrt{C}} \sum_{i,m} \hat{q}_i(\omega_m) v_i e^{i\omega_m c}$.

Finite-horizon transfer function. The modal-resolved spectrum is connected to the underlying linear dynamics by the finite-horizon transfer function $H_{i,N}(\omega) = (Cv_i)(w_i^* B) \frac{1-(\lambda_i e^{-i\omega})^N}{1-\lambda_i e^{-i\omega}}$, which is a finite geometric series. This object is a bridge between the modal-resolved spectrum and the underlying modal decomposition, not a separate innovation.

DTFT limit. As $N \to \infty$ the discrete Fourier transform converges to the discrete-time Fourier transform provided the series converges absolutely. We use this only in appendix-level technical arguments.

D.6 Path-compositional derivations

Reachability. For a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with adjacency matrix A, the exact-c walk-count vector is $g_c = A^c e_s$. Its support gives exact-length reachability, while the at-most-c reachable indicator is the Boolean closure $r_c = \mathbf{1}\{\sum_{j=0}^{c} A^j e_s > 0\}$. The update $r_{c+1} = r_c \vee \mathbf{1}\{Ar_c > 0\}$ is a Boolean recursion. Thus the real-valued Green kernel $CA^\ell B$ is a walk-count/weighted propagation object; the reachability readout is its support or Boolean closure. We use the *at most c steps* convention in the task semantics and keep the *exactly c steps* convention separate.

Weighted propagation. The cost-to-go satisfies $\rho_{c+1}(v) = \min\{\rho_c(v), \min_{u \in N^-(v)} \rho_c(u) + w(u \to v)\}$, with $\rho_0(s) = 0$ and $\rho_0(v) = \infty$ for $v \neq s$. The recursion is exact for the chosen horizon. A graph diameter bounds the number of edges on an unweighted shortest path, but it does not in general bound the cost-optimal walk length; whenever a global optimum is claimed we therefore use a cost-aware horizon or verify equality with the global dynamic program.

Soft-optimal propagation. The partition function $Z_{\beta,N}(v) = \sum_{\pi:|\pi| \leq N} e^{-\beta C(\pi)}$ satisfies

$$Z_{\beta,N+1}(v) = Z_{\beta,N}(v) + \sum_{u \in N^-(v)} e^{-\beta w(u \to v)} Z_{\beta,N}(u),$$

with $Z_{\beta,0}(s) = 1$ and $Z_{\beta,0}(v) = 0$ for $v \neq s$. The soft cost is $d_{\beta,N}(v) = -\beta^{-1} \log Z_{\beta,N}(v)$. If $d_N^*(v) = \min_{\pi:|\pi|\leq N} C(\pi)$ and $M_N(v)$ is the number of allowed walks, then $d_N^*(v) - (\log M_N(v))/\beta \leq d_{\beta,N}(v) \leq d_N^*(v)$. The limits $\beta \to \infty$ and $\beta \to 0$ correspond respectively to selection of the cheapest walk and broad averaging over the finite walk set.

E Experiment Details

E.1 Regime generation protocol and pass criteria

The controlled generator samples 540 main regimes as $2 \times 3 \times 3 \times 5 \times 2 \times 3 = 540$ (graph families $\times$ sizes $\times$ densities $\times$ seeds $\times$ directedness $\times$ task semantics). Two graph families are used: Erdős–Rényi (each edge present independently with probability p) and cycle-shortcuts (a base cycle plus k random shortcuts). Three sizes $n \in \{16, 24, 32\}$; three densities $p \in \{0.15, 0.30, 0.45\}$; five seeds; directed and undirected; three task semantics (reachability, weighted propagation, soft-optimal). The rollout horizon C equals the graph diameter D. The compendium audit records 7,200 soft-optimal evaluations over the generated configuration grid for $\beta \in \{0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, 128\}$. We retain this total as the audited count rather than infer it from a simplified product formula. The separate layered-DAG stress test contains 135 regimes and 1,800 additional soft-optimal evaluations. Pass criteria are stated in §5.1.

E.2 Per-family per-regime effective dimension

Table 2: Effective reasoning dimension r_{95} by family, type, and regime. Each row aggregates 45 runs (3 sizes $\times$ 3 densities $\times$ 5 seeds). Reachability and weighted propagation concentrate at one dominant input-output singular direction; soft-optimal needs about two. The exact block-Hankel rank (not shown) is typically 6–8 across all rows.

Family	Type	Regime	r_{90}	r_{95}	r_{99}
Cycle	Directed	Reachability	1.00	1.00	1.00
Cycle	Directed	Weighted	1.11	1.16	1.38
Cycle	Directed	Soft-opt	2.07	2.11	2.27
Cycle	Undirected	Reachability	1.00	1.00	1.00
Cycle	Undirected	Weighted	1.00	1.00	1.02
Cycle	Undirected	Soft-opt	2.24	2.29	2.51
ER	Directed	Reachability	1.00	1.00	1.00
ER	Directed	Weighted	1.09	1.16	1.56
ER	Directed	Soft-opt	2.09	2.11	2.20
ER	Undirected	Reachability	1.04	1.07	1.11
ER	Undirected	Weighted	1.09	1.11	1.31
ER	Undirected	Soft-opt	2.11	2.16	2.29

Reading the table. The pattern $r_{95} \approx 1$ for reachability and weighted propagation and $r_{95} \approx 2$ for soft-optimal is realisation-dependent on the path-compositional universe; it is not a universal task law. The corresponding exact block-Hankel ranks are typically 6–8, so low r_{95} does not imply only a few eigenmodes are involved; the input-output transmission concentrates on a small number of singular directions that can mix several eigenmodes.

The five-seed table is a sanity check rather than a new benchmark. It shows why reporting only the exact rank would obscure the task-conditioned compression: exact ranks are around seven or eight, while 95% of the measured transmission energy is carried by one or two singular directions. In the focused 24-configuration sweep, the directed and undirected soft-optimal rows have $r_{95} = 2.167 \pm 0.389$ and 2.417 ± 0.669, with mean active modal supports 21.67 and 15.83. This separation between modal participation and input-output concentration is the empirical counterpart of the quotient in equation 32.

Every row in Table 4 passes the interval test in equation 47, the monotonicity test, and the finite-horizon/global-optimum equality check. The table is included so that the headline endpoints do not

Table 3: Five-seed sanity validation. Exact Hankel rank remains larger than the effective transmission rank. The directed soft-optimal row is the numerically hardest modal case.

Type	Regime	Runs	Mean exact rank	Mean r_{95}
Undirected	Reachability	5	7.6	1.0
Undirected	Weighted	5	8.0	1.0
Undirected	Soft-optimal	5	7.4	2.0
Directed	Reachability	5	8.0	1.0
Directed	Weighted	5	8.0	1.0
Directed	Soft-optimal	5	6.8	2.0

Table 4: Soft-optimal audit across the 7,200 recorded evaluations. The mean absolute error is measured against the finite-horizon optimum; the last column is the analytic width $\beta^{-1} \log M$ averaged over the evaluated path sets.

β	Evaluations	Mean absolute error	Mean bound width
0.25	720	134.8865	162.0021
0.50	720	54.6210	81.0010
1	720	15.8608	40.5005
2	720	1.3055	20.2503
4	720	0.09266	10.1251
8	720	0.01983	5.0626
16	720	0.004792	2.5313
32	720	0.001133	1.2656
64	720	0.000255	0.6328
128	720	0.000060	0.3164

hide the transition regime around $\beta \in [2, 16]$, where the soft state changes from a broad mixture to a concentrated approximation of the optimum.

Table 5: Focused 24-configuration closure. Active modal support counts modes with non-negligible participation before the input-output singular-value truncation.

Type	Regime	Mean r_{95}	Active modal support
Directed	Reachability	1.000	1.00
Directed	Weighted	1.000	1.00
Directed	Soft-optimal	2.167 ± 0.389	21.67
Undirected	Reachability	1.000	1.00
Undirected	Weighted	1.000	1.00
Undirected	Soft-optimal	2.417 ± 0.669	15.83

Table 5 makes the two notions of complexity visible in one place. The Fourier branch representation can contain many active coordinates even when the Green/Hankel operator has a small numerical rank. Conversely, a single singular direction may mix several eigenmodes and therefore should not be interpreted as one semantic branch. The table is a controlled realization-dependent result, not a claim that every latent reasoning task has effective dimension one or two.

E.3 Conditioning and DAG stress test

Conditioning. The modal decomposition is sensitive to perturbations when $\mathrm{cond}(V)$ is large. The sensitivity is governed by the eigenvalue gaps $\min_{i \neq j} |\lambda_i - \lambda_j|$ and the modal norms. In the main ER and cycle sweep, two modal-failure cases occur in directed systems with $\mathrm{cond}(V) > 10^8$; the Schur form recovers both. The condition number is therefore a risk indicator for the modal attribution, not a strict monotonic predictor of modal error.

DAG stress test. The 135 layered-DAG configurations are constructed so that the eigen-modal decomposition is unavailable by design (deliberately non-diagonal). Across these, the Schur form passes at 100%, the block-Hankel realisation passes at 100%, and the basis-free Fourier closed form passes at 134/135; the modal-coefficient error is dominated by the controlled conditioning of the matrix. The framework does not require diagonalisability.

Table 6: Stress-test audit for the layered-DAG family. Modal attribution is intentionally unavailable; the remaining objects are evaluated directly.

Object	Configurations	Pass rate
Green response via Schur form	135	135/135
Block-Hankel realization	135	135/135
Basis-free Fourier reconstruction	135	134/135
State Parseval	135	machine precision
Answer Parseval	135	machine precision
Additional soft-optimal checks	1,800	100%

The DAG row is a boundary check for the theory. A defective transition matrix invalidates the semantic convenience of a complete eigenbasis, but it does not invalidate the finite Green response, the quotient defining r_ϵ, the Schur factorization, or the basis-free DFT. The one finite-horizon Fourier miss is retained in the table rather than silently removed; it is a useful reminder that numerical tolerances and conditioning matter even when the algebraic identity is exact.

E.4 Failure taxonomy and audit decisions

We classify a run as a modal failure only when the direct Green response is finite and the eigenbasis reconstruction exceeds the 10^{-9} relative tolerance. Such a failure is not treated as a failure of the framework, because the Schur representation is the coordinate-stable object. A Fourier failure is classified separately and is compared against the direct finite-window DFT. Parseval is checked independently of both decompositions. This separation prevents one unstable eigenvector calculation from contaminating all other diagnostics.

The two modal failures in the main ER/cycle sweep occur in directed systems with very large eigenbasis condition numbers. The two finite-window Fourier misses occur in the same numerical neighborhood. We therefore report condition numbers as a diagnostic covariate, not as a theorem claiming a monotone relation between $\mathrm{cond}(V)$ and error. The controlled evidence supports a practical rule: use modal attributions when the basis is well conditioned, and use Schur or direct branch coordinates otherwise.

E.5 Reconstruction checks used in every run

For each run we perform the following independent checks.

1. Reconstruct each direct Green block from the modal formula when available and from the Schur formula always.
2. Reconstruct every branch coefficient from its inverse DFT and compare the state trajectory at every reasoning step.
3. Apply the answer readout before and after the DFT and compare both answer trajectories.
4. Compare state and answer energies in the time and frequency domains.
5. Recompute reachability, weighted costs, and soft-optimal costs with an independent dynamic program.
6. Check the soft-optimal interval, monotonicity, and finite/global optimum equality at each recorded β.

The checks are deliberately redundant. A single aggregate accuracy score could hide a sign error in the phase convention or a transposed block in the Hankel matrix; the per-object checks make those errors visible.

F Further Consequences of the Closed-Loop View

F.1 Task conditioning is a three-way product

The modal term $(Wv_i)\lambda_i^\ell(w_i^* B)$ is more informative than a ranking by $|\lambda_i|$. The factor $w_i^* B$ depends on the question channel, the factor Wv_i depends on the answer query, and λ_i^ℓ depends on the amount of latent computation between injection and readout. A mode can be persistent yet irrelevant because it is never excited; it can be strongly excited yet irrelevant because it is invisible; or it can be visible yet absent from the question-reachable subspace. This explains why a single unsupervised activation statistic cannot serve as a task-independent definition of reasoning importance.

For a set of modes $\mathcal{I}$, define the finite-horizon transmission energy

$$\mathcal{E}_{\mathcal{I}}(N) = \sum_{\ell=0}^{N-1} \| \sum_{i\in\mathcal{I}} (Wv_i)\lambda_i^\ell (w_i^* B)\|_F^2. \tag{7}$$

The cross terms in this expression vanish only under additional orthogonality conditions on the input and readout factors. In general, the energy is an input-output quantity and can mix several modes. The Hankel singular directions are therefore the correct objects for compression; individual eigenmodes are an attribution tool whose validity depends on conditioning and alignment.

F.2 Phase and cross-branch interaction

The cross term in equation 40 can be written as

$$2\,|\hat{a}_v||\hat{a}_{v'}|\,|\langle Wu_v, Wu_{v'}\rangle|\cos\big(\arg\hat{a}_{v'} - \arg\hat{a}_v + \arg\langle Wu_v, Wu_{v'}\rangle\big). \tag{8}$$

Thus phase is not a decorative coordinate. It determines whether two branches reinforce or cancel after the answer readout, conditional on the readout Gram matrix. Equation equation 8 is a theoretical diagnostic; the present controlled experiments verify the quadratic identity but do not claim an empirical phase law for a trained language model.

F.3 Why low effective dimension does not imply a short latent trajectory

The rank of $\mathcal{H}_{p,q}$ counts independent input-output directions, not the number of latent steps. A long rollout can repeatedly propagate the same one or two transmission directions while using many intermediate state coordinates. Conversely, a short rollout can have a high effective dimension if distinct inputs reach distinct answer directions. This distinction is visible in the controlled suite: exact ranks remain around six to eight while r_{95} is near one for reachability and weighted propagation and near two for soft-optimal propagation. The result supports a compression interpretation, not a claim that the model literally computes in one or two neurons or branches.

F.4 Answer-conditioned versus trajectory-conditioned spectra

The basis-free state spectrum $\hat{z}(m)$ describes the trajectory independent of a particular answer. The answer-conditioned spectrum $\hat{y}(m) = W\hat{z}(m)$ weights the same temporal components by the readout. If two tasks use different readouts W_1 and W_2 on the same latent trajectory, the latent spectrum is unchanged but the answer-side energies differ:

$$E_{m,1}^{\mathrm{ans}} = \hat{z}(m)^* W_1^* W_1 \hat{z}(m), \qquad E_{m,2}^{\mathrm{ans}} = \hat{z}(m)^* W_2^* W_2 \hat{z}(m). \tag{9}$$

This makes the phrase "task-conditioned mode" literal: relevance is defined relative to the input and readout pair, not by the latent trajectory alone.

F.5 Finite-window effects

The DFT uses a periodic extension of the finite reasoning trajectory. The inverse transform is exact on the observed window, but a frequency peak can be spread when the first and last states do not match smoothly. This is a windowing effect, not a failure of the branch representation. Zero padding refines the frequency grid without adding information; it does not change the state or answer energy. For this reason all exact checks in the paper use the original horizon, while any visual spectrum should state the window and the DFT normalization.

F.6 Relation to balanced truncation

The product of controllability and observability Gramians gives the nonzero squared singular values of the finite Hankel operator. In a stable infinite-horizon LTI system, the analogous quantities are the classical Hankel singular values used in balanced truncation. Our finite-horizon definition is preferable for latent reasoning because the number of thought steps is part of the computation and stability beyond the answer position is not assumed. The controlled experiments therefore report finite-window ranks and do not claim an asymptotic stationary model.

G Limitations and Falsifiable Predictions

The framework has four explicit limitations. First, the local response is a first-order statement for a nonlinear backbone; large interventions can leave the Taylor tube. Second, branch-resolved Fourier analysis requires a branch coordinate system. A trained model without an identifiable branch basis can still be analysed with the basis-free or modal-resolved spectrum, but semantic branch labels are then not available. Third, eigen-modal attribution is unstable for defective or strongly non-normal transitions; the Schur form and direct Hankel operator remain valid. Fourth, the controlled path-compositional suite tests algebraic closure and numerical implementation. It does not establish that a trained language model uses the same branch coordinates or the same effective-rank pattern.

These limitations lead to concrete predictions. For a fixed trained rollout, finite interventions aligned with high right singular vectors of the estimated Hankel operator should produce larger answer changes than equal-norm interventions in its tail. Across two readouts for different tasks, the latent branch spectrum should remain similar while the answer-side energy changes according to the readout Gram matrix. When an explicit branch basis can be recovered, the inverse DFT should reconstruct every observed thought on the chosen horizon, and the terminal answer should require the terminal phase factor in the inverse transform. Finally, if a model exhibits a broad latent spectrum but a concentrated answer spectrum, the difference should be explained by small readout norms or destructive off-diagonal terms rather than by a failure of Parseval.

G.1 What would falsify the mechanism

The proposed interpretation would be weakened by any of the following observations on a controlled or trained system: (i) the measured input-to-answer perturbation cannot be reconstructed from the step-resolved Green kernels even within the local Taylor regime; (ii) a branch basis that is claimed to be orthogonal fails the inverse DFT or Parseval checks; (iii) answer-side energy is reported without the readout Gram matrix and changes under a mere latent coordinate transformation; or (iv) a low r_ϵ is claimed while the omitted Hankel singular-value tail is large. These are diagnostic failures, not matters of presentation, and they can be tested without assuming the conclusions of the paper.

H Reproducibility Checklist

Table 7: Reproducibility checklist for the controlled mechanism suite.

Item	Recorded information
Graph families	Erdős–Rényi and cycle-shortcuts; layered DAG stress test
Graph sizes	$n \in \{16, 24, 32\}$
Densities	$p \in \{0.15, 0.30, 0.45\}$
Seeds	five seeds per family/type/regime cell
Task semantics	reachability, weighted propagation, soft-optimal
Fourier convention	unitary DFT, $\omega_m = 2\pi m/C$
Pass tolerances	reconstruction 10^{-9}; Parseval 10^{-15}
Soft-optimal grid	$\beta \in \{0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, 128\}$
Separate stress test	135 DAG regimes and 1,800 additional soft-optimal evaluations

The archive associated with the experiment compendium stores one row per regime and one row per soft-optimal evaluation. A reproduction should first run the analytic path recursions, then compute the linear-algebra diagnostics, and only then aggregate pass rates. Reversing this order can conceal a per-regime failure behind an average. The supplied source also records the distinction between direct modal reconstruction and Schur reconstruction, so that a user can reproduce the 538/540 versus 540/540 split rather than only the aggregate conclusion.

H.1 Minimal pseudocode for an independent implementation

Listing 1: Reference-level pseudocode for one controlled regime.

```
graph = sample_graph(family, n, density, seed, directed)
for semantics in [reachability, weighted, soft_optimal]:
    trajectory = build_path_state(graph, semantics, horizon)
    A, B, W = analytic_transition_and_readout(trajectory)
    G = [W @ matrix_power(A, ell) @ B for ell in range(horizon)]
    H = block_hankel(G)
    sv = singular_values(H)
    r95 = tail_rank(sv, tolerance=0.05)
    if diagonalizable(A):
        G_modal = modal_reconstruction(A, B, W, horizon)
    G_schur = schur_reconstruction(A, B, W, horizon)
    z_hat = unitary_dft(trajectory.states, axis=0)
    y_hat = unitary_dft(readout(W, trajectory.states), axis=0)
    check_parseval(trajectory.states, z_hat)
    check_parseval(readout(W, trajectory.states), y_hat)
    for beta in beta_grid:
        soft = soft_optimal_value(graph, beta, horizon)
        check_logsumexp_bound(soft, graph, beta, horizon)
```

The listing is intentionally implementation-neutral. It specifies the order of operations and the quantities that must be stored; it does not prescribe a particular autodiff library or a trained model API.

I Supplementary Proofs for the Controlled Diagnostics

I.1 Monotonicity of soft-optimal concentration

For a fixed finite path set with costs $c_1, \dots, c_M$, define

$$\rho_\beta = -\frac{1}{\beta} \log \sum_{j=1}^{M} e^{-\beta c_j}.$$

The map $\beta \mapsto \rho_\beta$ is nondecreasing and converges to $\min_j c_j$. To see the monotonicity, write $p_j(\beta) = e^{-\beta c_j} / \sum_r e^{-\beta c_r}$. Differentiating gives

$$\frac{d\rho_\beta}{d\beta} = \frac{1}{\beta^2} \log \sum_j e^{-\beta c_j} + \frac{1}{\beta} \sum_j p_j(\beta) c_j. \tag{10}$$

Equivalently, after subtracting the minimum cost $c_* = \min_j c_j$ and setting $d_j = c_j - c_* \geq 0$, one obtains

$$\rho_\beta - c_* = -\frac{1}{\beta} \log \sum_j e^{-\beta d_j}. \tag{11}$$

The sum is at least one and at most M, so $0 \leq c_* - \rho_\beta \leq \beta^{-1} \log M$. Increasing β suppresses every term with $d_j > 0$ and therefore decreases the gap $c_* - \rho_\beta$. This proof uses only finiteness of the path set and is applied independently at each graph vertex and horizon.

The limit follows from the same expression. The term corresponding to an optimal path remains one after shifting by c_*, while every strictly suboptimal term converges to zero. If several paths tie for the optimum, their multiplicity affects the finite-β correction but not the limiting value. This is the precise sense in which soft-optimal propagation interpolates between superposition and selection.

I.2 Finite-horizon versus global optimum

Let $\Pi_{\le c}(v)$ be the set of walks of length at most c and $\Pi(v)$ the set of all admissible walks. The finite-horizon optimum is $C_c^*(v) = \min_{\pi \in \Pi_{\le c}(v)} C(\pi)$, while the global optimum is $C_\infty^*(v) = \min_{\pi \in \Pi(v)} C(\pi)$. If every edge weight is strictly positive, any optimal walk can be taken simple, and on a finite graph its length is bounded by $n-1$. Consequently $C_c^*(v) = C_\infty^*(v)$ whenever c exceeds the shortest simple-path length. The controlled generator sets the horizon to the graph diameter, so equality is expected for the reachability and weighted regimes. The soft-optimal evaluation checks this equality explicitly instead of relying on the graph construction.

For cyclic graphs and soft costs, there can be many non-optimal walks of length below the horizon. Their contribution changes $Z_{\beta,c}$ but not C_c^*. This is why the finite-horizon/global-optimum match and the log-sum-exp interval are reported as separate Boolean checks. A run can have a correct optimum but a numerically unstable partition sum, or a stable partition sum with a horizon that has not yet reached the global optimum.

I.3 Stability of the effective rank

Suppose the measured Hankel operator is $\widehat{\mathcal{H}} = \mathcal{H} + E$. Weyl's inequality gives $|\widehat{\sigma}_j - \sigma_j| \le \|E\|_2$. Hence a singular-value gap around the threshold used for r_ϵ makes the rank stable to small estimation errors, while a cluster of singular values near the threshold should be reported as an interval rather than a single decisive rank. In practice we therefore record r_{90}, r_{95}, and r_{99} together. The controlled tables show that the qualitative one-versus-two pattern is stable across these thresholds for the main families, while the exact rank remains larger.

If the input coordinates have covariance Σ_u and the answer metric has weight $M_y \succeq 0$, the weighted operator is $M_y^{1/2} \mathcal{H} \Sigma_u^{1/2}$. Its singular values change under rescaling of input channels or answer logits. All reported controlled runs use a fixed normalization, and comparisons across configurations use the same convention. A task-conditioned rank is therefore meaningful only together with the input and output units that define it.

I.4 A bound for omitting a set of modal components

Let $\mathcal{I}$ be the retained eigenmodes and $\mathcal{I}^c$ the omitted set. For a zero-initial-state response to inputs $v_0, \dots, v_{N-1}$,

$$\delta y_N - \delta y_N^{\mathcal{I}} = \sum_{i \in \mathcal{I}^c} (W v_i) \sum_{j=0}^{N-1} \lambda_i^{N-1-j} (w_i^* B) v_j. \tag{12}$$

Applying Cauchy–Schwarz first over j and then over i gives

$$\|\delta y_N - \delta y_N^{\mathcal{I}}\| \le \left(\sum_{i \in \mathcal{I}^c} \|W v_i\|^2 \sum_{j=0}^{N-1} |\lambda_i|^{2(N-1-j)} \|w_i^* B\|^2 \right)^{1/2} \left(\sum_{j=0}^{N-1} \|v_j\|^2 \right)^{1/2}. \tag{13}$$

The right-hand side contains all three factors. It becomes loose when modal contributions cancel, which is another reason to use the exact Hankel SVD for compression and the modal expansion for attribution.

I.5 Coordinate changes and branch coefficients

State-space quantities are invariant under an invertible change of coordinates, but branch coefficients are not. If $\tilde{z} = Sz$, then the Green kernel and Hankel operator are unchanged after transforming A, B, W, while an explicit branch matrix becomes $\tilde{U} = SU$. The coefficients $\tilde{a}(c)$ satisfying $\tilde{z}_c = \tilde{U} \tilde{a}(c)$ generally differ from $a(c)$. Therefore branch-resolved spectra should be compared only when the branch basis is fixed by the task construction or by an explicit identification protocol. The basis-free spectrum $\hat{z}(m)$ and the answer-side spectrum $W \hat{z}(m)$ do not have this ambiguity.

Table 8: Complete effective-dimension summary for the recurrent ER and cycle-shortcut families. Values aggregate 45 runs per row.

Family	Type	Regime	r_{90}	r_{95}	r_{99}	Exact rank
Cycle	Directed	Reachability	1.00	1.00	1.00	6.18
Cycle	Directed	Weighted	1.11	1.16	1.38	7.60
Cycle	Directed	Soft-optimal	2.07	2.11	2.27	7.29
Cycle	Undirected	Reachability	1.00	1.00	1.00	5.60
Cycle	Undirected	Weighted	1.00	1.00	1.02	7.38
Cycle	Undirected	Soft-optimal	2.24	2.29	2.51	7.58
ER	Directed	Reachability	1.00	1.00	1.00	6.47
ER	Directed	Weighted	1.09	1.16	1.56	7.62
ER	Directed	Soft-optimal	2.09	2.11	2.20	7.13
ER	Undirected	Reachability	1.04	1.07	1.11	6.24
ER	Undirected	Weighted	1.09	1.11	1.31	7.42
ER	Undirected	Soft-optimal	2.11	2.16	2.29	7.16

I.6 Uniqueness of an orthogonal branch expansion

Suppose U and $\widetilde{U}$ are two orthonormal matrices whose columns span the same branch subspace and suppose $z_c = Ua(c) = \widetilde{U}\widetilde{a}(c)$ for every observed step. Then $\widetilde{U} = UR$ for a unitary matrix R and $\widetilde{a}(c) = R^* a(c)$. If the branch semantics additionally require that each coordinate is tied to a distinct graph vertex, only permutation and unit-modulus phase factors remain. The Fourier coefficients transform by the same fixed unitary map, $\widehat{\widetilde{a}}(m) = R^* \hat{a}(m)$. Latent energy is invariant under a general unitary change, but branch labels and individual answer contributions are not. This formalizes the distinction between a semantic branch basis and an arbitrary latent rotation.

J Full Numerical Audit Tables

J.1 Main ER/cycle rank summary

Table 8 records the complete family/type/regime summary supplied with the experiment compendium. The three thresholds reveal how quickly the transmission spectrum concentrates; the exact rank is included to show that concentration is not the same as literal rank one.

The table should be read horizontally within a row. Increasing the retained energy from 90% to 99% increases the effective dimension modestly for reachability and weighted propagation, while soft-optimal propagation retains a broader tail. The exact rank is not ordered by task difficulty alone; it reflects the graph realization and the full finite Hankel window.

J.2 Extreme conditioning sanity check

The extreme quick test contains one representative configuration for each graph type and semantic regime. The table below records the spectral radius, eigenbasis condition number, modal reconstruction error, Fourier reconstruction error, and r_{95}.

The spectral radius is not itself a stability certificate for a finite rollout: a matrix with $\rho(A) > 1$ can still be evaluated for a finite horizon, and a non-normal matrix with $\rho(A) < 1$ can exhibit transient growth. The table is included to separate these notions from numerical conditioning.

J.3 Per-beta pass matrix

The compendium records four Boolean checks at each β: interval bound, monotonicity, finite-horizon optimum match, and global optimum match. All four checks pass for every recorded evaluation. The pass matrix is therefore a 10×4 matrix of ones for the ten β values in Table 4. Reporting the matrix explicitly matters because a decreasing mean error alone would not guarantee that the theoretical interval was respected on every graph.

Table 9: Extreme quick-test configurations. The directed soft-optimal row has the largest eigenbasis condition number but remains well behaved under the Schur and basis-free checks.

Type	Regime	$\rho(A)$	$\mathrm{cond}(V)$	Modal error	Fourier error	r_{95}
Undirected	Reachability	5.833	1.00	4.5×10^{-15}	6.1×10^{-15}	1
Undirected	Weighted	2.674	1.00	1.7×10^{-14}	1.8×10^{-14}	1
Undirected	Soft-optimal	0.244	1.00	6.6×10^{-13}	2.2×10^{-16}	2
Directed	Reachability	3.967	22.89	3.2×10^{-14}	3.1×10^{-14}	1
Directed	Weighted	1.693	61.35	3.9×10^{-14}	3.8×10^{-14}	1
Directed	Soft-optimal	0.102	236.11	4.8×10^{-13}	1.5×10^{-16}	2

J.4 Audit of unsupported claims

The following quantities are intentionally not reported as empirical findings: Fourier spectra of a trained COCONUT or CODI checkpoint; semantic labels such as frontier, optimal, or unreachable branches in a trained model; a measured phase intervention; a trained-model statistic showing high latent energy but low answer energy; or a comparison claiming that PCA is universally inferior to task-conditioned modes. They are theoretical diagnostics or future experiments. Keeping them separate from the controlled closure results is part of the audit protocol.

K Model-Facing Measurement Protocol

The controlled universe verifies identities. To apply the same framework to a trained latent-reasoning model, one must first specify a state, input channel, and readout. The protocol below is a measurement plan, not a result claimed by this paper.

K.1 State capture

Record the continuous thoughts, attention cache, positional state, and any fast weights that influence the next latent update. If a model stores information outside the exposed thought vector, include that memory in the state or report the resulting non-Markov approximation. Fix the prompt, sampling seed, latent budget, and answer prefix so that the readout is differentiable and comparable across interventions.

K.2 Input channels

Distinguish natural question perturbations from independent intervention channels. A question embedding perturbation usually enters at the initial state; retrieved evidence, tool outputs, or controlled latent injections may enter at later steps. If two channels are constrained to share one scalar, encode that constraint explicitly before forming $\mathcal{C}_q$. Otherwise the apparent reachable dimension can be inflated by treating dependent interventions as independent.

K.3 Finite-difference validation

For each estimated Jacobian direction v, apply perturbations $\pm\delta v$ at several magnitudes and compare the measured answer change with the Green prediction. In the local regime the residual should scale quadratically with δ; a linear residual indicates a missing input channel, an incorrect cache state, or a derivative evaluated at the wrong rollout. Report the perturbation norm, answer-logit metric, and the horizon at which the error is measured.

K.4 Branch coordinate identification

When a semantic branch basis is not given, begin with the basis-free spectrum $\hat{z}(m)$. A candidate branch basis can be proposed by a normal modal decomposition, by a task-supervised dictionary, or by a low-rank factorization of intervention responses. Each candidate must be checked for

reconstruction, orthogonality (if claimed), and stability across prompts. The paper's branch-level energy and cross-term interpretation should be used only after this validation.

K.5 Readout comparison

For a fixed latent trajectory, evaluate multiple answer readouts when possible: candidate logits, answer-token margins, or a teacher representation. The latent spectrum is common to all readouts; answer-side spectra differ. This comparison isolates whether a branch is absent from the state or merely suppressed by the chosen answer measurement.

K.6 Computing resources

L Minimal Realizations and Input Constraints

L.1 Finite-horizon realization theorem

Let $\mathcal{H}_{p,q} = \mathcal{O}_p \mathcal{C}_q$ be a finite block-Hankel operator. Any state-space realization with state dimension d factors $\mathcal{H}_{p,q}$ through $\mathbb{C}^d$, so $\operatorname{rank} \mathcal{H}_{p,q} \le d$. Conversely, a rank factorization $\mathcal{H}_{p,q} = LR$ with $r = \operatorname{rank} \mathcal{H}_{p,q}$ reproduces the measured finite input–output table in a quotient state of dimension r. A transition on this quotient is well-defined only when the observed blocks satisfy the appropriate shift-consistency/extension condition. Hence r is always the finite-window effective transmission dimension; it becomes the minimal realization dimension only after the window is large enough and the shift extension is verified.

The construction is useful even when no single transition matrix is fitted. For a time-varying rollout, arrange the step-resolved kernels $G_{N,j}$ into an input-output matrix and take its rank. The result is still a quotient dimension for the chosen injection and readout protocol, but it should not be called the rank of a stationary Hankel operator unless the anti-diagonal shift structure has been verified.

L.2 Natural input constraints

Suppose the independent intervention coordinates are ξ and the actual step inputs are $v = J\xi$ for a known linear constraint J. The reachable matrix is then $\mathcal{C}_q J$, not $\mathcal{C}_q$. The corresponding task operator is

$$\mathcal{H}^{\text{legal}}_{p,q} = \mathcal{O}_p \mathcal{C}_q J. \tag{14}$$

This matters for a prompt that enters the rollout in the same way at several steps: allowing an independent scalar at each step would create a larger artificial reachable space than the natural question perturbation permits. The quotient theorem applies without change after replacing $\mathcal{C}_q$ by $\mathcal{C}_q J$.

L.3 Initial-state and later-injection decomposition

If the question is encoded only at the initial state, the first-order answer map is $WA^N J_E$, where J_E is the encoder Jacobian. If external evidence enters at later steps, the complete map is

$$\delta y_N = WA^N J_E \delta q + \sum_{j=0}^{N-1} WA^{N-1-j} B_j \delta u_j + O(\delta^2). \tag{15}$$

The same question perturbation should not be counted both in $J_E \delta q$ and in every δu_j unless the computation graph truly injects it twice. This decomposition is the correct way to compare prompt information with tool or environment information.

L.4 A two-state off-axis example

Take $\mathcal{C} = (1,1)^\top$ and $\mathcal{O} = (1,0)$. The reachable state subspace is the diagonal line $\operatorname{span}\{(1,1)^\top\}$, while the observable subspace is the first coordinate axis. Their literal intersection is $\{0\}$, yet $\mathcal{OC} = 1$. The input changes the observed coordinate even though the two geometric subspaces do not overlap. The correct task dimension is the rank of the input-output map, not the dimension of an ambient

Euclidean intersection. This example is why the paper uses the quotient $\mathcal{R}/(\mathcal{R} \cap \ker \mathcal{O})$ and the Hankel operator rather than an unqualified intersection of reachable and observable subspaces.

L.5 Similarity invariance

Let S be invertible and define $\widetilde{A} = SAS^{-1}$, $\widetilde{B} = SB$, and $\widetilde{W} = WS^{-1}$. Then

$$\widetilde{W}\widetilde{A}^{\ell}\widetilde{B} = WA^{\ell}B, \qquad \widetilde{\mathcal{H}}_{p,q} = \mathcal{H}_{p,q}. \tag{16}$$

The state covariance, branch coefficients, and individual eigenvectors do change, but the measured input-output response does not. Any proposed latent reasoning dimension that changes under $z \mapsto Sz$ without changing the input-output map is therefore a coordinate statistic rather than a task-conditioned transmission dimension.

L.6 Balanced coordinates as an optional diagnostic

For positive-definite finite Gramians W_c and W_o, choose a factorization $W_c = RR^*$ and compute the singular value decomposition $R^*W_oR = U\Sigma^2U^*$. The balanced coordinates are $T = \Sigma^{1/2}U^*R^{-1}$. In these coordinates the controllability and observability Gramians share the diagonal form Σ. The finite-horizon version is an optional diagnostic for the latent loop; it is not required for the branch Fourier representation and is not used to claim that the model itself performs balanced truncation. Its value is to provide a coordinate system in which the input-output singular directions can be compared directly with latent activation directions.

L.7 Chain-rule derivation of the Green kernel

For completeness, write the terminal state as the composition

$$z_N = F_{N-1}(F_{N-2}(\cdots F_0(z_0, u_0), u_1 \cdots), u_{N-1}). \tag{17}$$

Differentiating with respect to u_j gives

$$\frac{\partial z_N}{\partial u_j} = A_{N-1}A_{N-2}\cdots A_{j+1}B_j = \Phi(N, j+1)B_j. \tag{18}$$

Multiplying by the terminal derivative $W = \partial_z g$ yields $G_{N,j} = W\Phi(N, j+1)B_j$. This derivation shows that the Green kernel is a chain-rule object even when the latent update is implemented by attention, an MLP, a recurrent cell, or an external differentiable program.

L.8 Finite-window rank versus asymptotic rank

The finite matrix $\mathcal{H}_{p,q}$ can have smaller rank than the infinite impulse-response realization because a mode may first appear after the chosen window or cancel on the measured horizon. Conversely, a finite window can display transient directions that vanish asymptotically. Every reported r_ϵ is therefore indexed by (p, q, N), even when the indices are suppressed in prose. Comparing ranks across experiments is meaningful only when the horizon, input normalization, and answer metric are held fixed.

L.9 Finite-horizon stability and non-normal amplification

The finite-horizon response does not require an asymptotically stable transition. Let $K_{N,j} = \|C_N\Phi(N, j+1)\|$ and suppose the local Taylor bounds in equation 28 hold. Then the propagated nonlinear error obeys

$$\|\varepsilon_N\| \le \sum_{j=0}^{N-1} K_{N,j}\frac{L_j}{2}(\|e_j\| + \|v_j\|)^2 + \frac{L_g}{2}\|e_N\|^2. \tag{19}$$

The constants $K_{N,j}$ are finite for every finite rollout, even when $\rho(A) > 1$. Thus the Green kernel and the branch spectrum are well defined on the answer window without imposing an infinite-horizon stability assumption.

For a normal stationary transition, $K_{N,j} \leq \|C_N\|\|A\|^{N-1-j}$ and the usual geometric estimate applies. A non-normal transition can have $\|A^k\|$ much larger than $\rho(A)^k$ over a finite interval; the resulting transient amplification can make a small per-step Taylor remainder visible at the terminal readout. This is why the Schur reconstruction and the layered-DAG stress test are reported separately. The test is not an asymptotic stability check: it asks whether the measured finite-window map remains reconstructible when eigenvector attribution is unavailable.

The same observation affects numerical rank. If $\sigma_j(\mathcal{H}_{p,q})$ are close, a perturbation of size η can change the thresholded rank whenever $\sigma_j \approx \eta$. We therefore report the tolerance used to define r_ϵ, retain the exact rank in the audit tables, and treat changes under admissible input constraints as changes in the measured task channel rather than changes in the latent width.

M Interpretation Guide for Reviewers

M.1 What the framework claims

The framework claims an exact chain of definitions for a chosen latent rollout. Once the state, legal input coordinates, and answer readout are fixed, the Jacobian products define a finite Green response. In the stationary affine specialization, this response has the form $WA^\ell B$. If the transition is diagonalizable, the same matrix can be written as a sum of modal terms. If the state has an explicit orthogonal branch expansion, the branch coefficients form finite signals and therefore possess an exact DFT. Applying a fixed readout commutes with that DFT on the whole trajectory, while terminal reconstruction adds the phase associated with the final step. These are statements about the mathematical objects constructed from the rollout.

M.2 What the framework does not claim

The framework does not claim that every continuous-thought model has a semantic orthogonal branch basis, that every latent update is globally linear, or that a low numerical Hankel rank is universal across tasks. It does not claim that the graph construction in the controlled universe is a trained language model, nor that the reported r_{95} values are a law of reasoning. It does not claim that a Fourier peak identifies a semantic branch without a coordinate-identification protocol. These boundaries are part of the result because they specify where the identities can be tested and where new empirical work is required.

M.3 How to read the controlled evidence

The main validation table should be read as a closure audit. Green and Schur checks test the same input-to-output map in two coordinate systems. Fourier reconstruction and Parseval test the finite-window temporal representation. Reachability and weighted dynamic programming test the path semantics independently of spectral attribution. The soft-optimal sweep tests the finite-horizon log-sum-exp interval and its selection limit. No row measures model accuracy. The value of the suite is that every theoretical object has a corresponding numerical check and every known numerical boundary is reported separately.

M.4 Recommended comparison baselines

For a future trained-model evaluation, the natural baselines are: activation PCA at matched rank; controllability-only truncation; observability-only truncation; random subspaces; individual eigenmode truncation when the eigenbasis is stable; and the Hankel singular subspace proposed here. Every baseline should use the same input perturbation distribution, answer metric, horizon, and rank. The comparison should report both answer reconstruction error and the omitted singular-value tail. A baseline that performs well only because its input or output normalization differs is not a meaningful mechanistic comparison.

The expected ordering is a hypothesis for future measurement, not a result of the present controlled suite. The controlled four-branch calculation and the finite-Hankel theorems explain why the joint criterion can preserve an answer that one-sided or variance-only criteria discard, but a trained-model claim requires the intervention protocol above.

Table 10: Suggested matched-rank comparisons for a trained-model follow-up.

Method	Quantity retained
Hankel singular subspace	Joint question-to-answer transmission directions
Controllability-only	Input-reachable state energy
Observability-only	Answer-visible state energy
Activation PCA	Trajectory covariance energy
Eigenmode truncation	Individual propagation modes, if well conditioned
Random subspace	Rank-matched null baseline

M.5 Suggested figure captions

The overview figure should be captioned as a mechanism map: question perturbation enters through B, propagates through A, is represented either in task-conditioned modal coordinates or in explicit branch coordinates, and reaches the answer through W. The Fourier figure should be captioned as a branch-by-frequency decomposition: rows identify branches, columns identify reasoning-time frequencies, color encodes amplitude, and a second panel applies the readout Gram matrix to separate diagonal and cross-branch answer contributions. A separate results figure may show the controlled r_{95} versus active-modal-support pattern and the soft-optimal β sweep, but those empirical plots should not be substituted for the conceptual mechanism figure.

N Noise, Finite Samples, and Robust Diagnostics

The identities in the controlled universe are exact, but a model-facing measurement contains finite-difference error, stochastic rollout noise, and a finite number of intervention directions. This appendix records the corresponding perturbation formulas so that a future trained-model study can distinguish estimation error from a failure of the mechanism.

N.1 Noisy Green estimates

Let $\widehat{G}_\ell = G_\ell + E_\ell$ be an estimated Green block and let $\widehat{\mathcal{H}} = \mathcal{H} + \Delta_H$ be the resulting Hankel estimate. If $\|E_\ell\| \le \eta_\ell$, then for an input sequence $v_0, \dots, v_{N-1}$,

$$\Big\| \sum_{\ell=0}^{N-1} (\widehat{G}_\ell - G_\ell) v_{N-1-\ell} \Big\| \le \left(\sum_{\ell=0}^{N-1} \eta_\ell^2 \right)^{1/2} \left(\sum_{j=0}^{N-1} \|v_j\|^2 \right)^{1/2}. \tag{20}$$

The bound is a direct Cauchy–Schwarz inequality. It separates the error of each step kernel from the energy of the intervention. A broad input distribution can therefore make a small per-step derivative error visible at the final answer, while a narrow intervention family may fail to identify a direction that is mathematically present.

N.2 Subspace perturbation

Let V_r contain the leading right singular vectors of $\mathcal{H}$ and $\widehat{V}_r$ those of $\widehat{\mathcal{H}}$. If the singular-value gap $\sigma_r - \sigma_{r+1}$ is positive and $\|\Delta_H\|_2$ is smaller than half the gap, the Davis–Kahan/Wedin perturbation estimate gives

$$\|\sin \Theta(V_r, \widehat{V}_r)\|_2 \le \frac{2\|\Delta_H\|_2}{\sigma_r - \sigma_{r+1}}. \tag{21}$$

This is the finite-sample counterpart of the numerical-rank discussion. A task-conditioned subspace is stable when the transmission spectrum has a gap; it is intrinsically ambiguous when several singular values are nearly tied. Reporting only one selected basis vector in the latter case would overstate identifiability.

N.3 Bootstrap protocol

For a future trained-model study, draw independent intervention batches, estimate $\mathcal{H}$ on each batch, and report the median and interquartile range of r_ϵ and the principal angles between the estimated subspaces. A bootstrap should resample prompts as well as perturbation directions because a single prompt-specific Jacobian cannot establish a task-general state dimension. The controlled suite does not require this statistical layer because its operators are analytic; including the protocol here makes the boundary explicit.

N.4 Noise-aware Parseval

Suppose the observed trajectory is $\widetilde{z}_c = z_c + \xi_c$ with zero-mean independent noise of covariance Σ_ξ. The unitary DFT preserves the total expected energy:

$$\mathbb{E}\sum_c \|\widetilde{z}_c\|^2 = \mathbb{E}\sum_m \|\widehat{\widetilde{z}}(m)\|^2 = \sum_c \|z_c\|^2 + C\,\mathrm{tr}\,\Sigma_\xi. \tag{22}$$

The equality follows because the DFT is unitary and the cross terms have zero expectation. The same statement holds after a fixed linear readout with noise covariance $W\Sigma_\xi W^*$. Thus Parseval remains a useful implementation check under noise, provided the noise floor is estimated rather than mistaken for reasoning energy.

N.5 Readout uncertainty

If the answer readout is estimated as $\widehat{W} = W + E_W$, then

$$\left|\widehat{E}_m^{\mathrm{ans}} - E_m^{\mathrm{ans}}\right| \le 2\|E_W\|\|W\|\|U\hat{a}_m\|^2 + \|E_W\|^2\|U\hat{a}_m\|^2. \tag{23}$$

This follows by expanding $(W+E_W)^*(W+E_W) - W^*W$. The bound emphasizes that answer-side attribution can be more sensitive than latent Parseval when the readout Jacobian is poorly conditioned or the answer logits are nearly tied.

O A Worked Four-Branch Calculation

This section gives an explicit finite calculation that can be checked by hand. Let $C = 4$ and let u_1, u_2, u_3, u_4 be orthonormal. Consider the coefficient table

	$c=0$	$c=1$	$c=2$	$c=3$
$hline a_1(c)$	1	1	0	0
$a_2(c)$	0	1	1	0
$a_3(c)$	0	0	1	1
$a_4(c)$	1	0	0	1

(24)

The state at step c is $z_c = \sum_v a_v(c)u_v$. The unitary DFT frequencies are $0, \pi/2, \pi, 3\pi/2$. For the first branch,

$$\hat{a}_1(0) = 1, \qquad \hat{a}_1(\pi/2) = \tfrac{1}{2}(1-i),$$
$$\hat{a}_1(\pi) = 0, \qquad \hat{a}_1(3\pi/2) = \tfrac{1}{2}(1+i),$$

and the remaining branches are cyclic shifts of these coefficients. The inverse transform recovers the four rows exactly. The latent energy of every branch is 2, while the frequency marginal is $E_0^{\mathrm{lat}} = 2$, $E_{\pi/2}^{\mathrm{lat}} = 2$, $E_\pi^{\mathrm{lat}} = 0$, and $E_{3\pi/2}^{\mathrm{lat}} = 2$.

Now choose a scalar answer readout W with $Wu_1 = Wu_2 = 1$ and $Wu_3 = Wu_4 = -1$. At frequency $\pi/2$, the diagonal answer contribution is the sum of four individual energies, but the cross terms between branches 1 and 2 are positive while those between 1 and 3 are negative. The latent spectrum is unchanged if the signs in W are swapped; the answer spectrum changes. This is the smallest example showing that branch importance is a property of the pair (U, W), not of U alone.

O.1 Fourier terminal phase

For the same example, the terminal step is $c = 3$. The contribution of frequency $\pi/2$ to the terminal answer is multiplied by $e^{i3\pi/2} = -i$, while the contribution of frequency $3\pi/2$ is multiplied by $e^{i9\pi/2} = i$. Removing these factors gives the correct whole-trajectory energy but the wrong terminal state. This explicit calculation is why the paper keeps terminal reconstruction and Parseval in separate equations.

P Alternative Readouts and Task Comparisons

P.1 Multiple candidate answers

Let $W \in \mathbb{C}^{d_y \times d}$ contain one row per candidate answer logit. The Green response then has one output row per candidate, and the modal factor (Wv_i) is a vector of candidate-specific visibility scores. A mode can therefore be visible to one candidate and invisible to another. The answer-side energy E_m^{ans} aggregates all rows; a candidate-specific energy is obtained by replacing W^*W with $w_r^* w_r$ for row r.

P.2 Teacher and student readouts

In a distillation setting, let W_T read the teacher representation and W_S read the student representation. The same latent trajectory can be compared under both readouts by computing $E_{m,T}^{\mathrm{ans}}$ and $E_{m,S}^{\mathrm{ans}}$. Equality of the two spectra is stronger than equality of a final label: it requires the student to preserve the temporal components that the teacher uses. A label match with different answer-side spectra would indicate compensating directions rather than mechanistic agreement.

P.3 Prompt families

Across a prompt family, one can stack the Green kernels or Hankel operators and compute a shared subspace. The shared rank measures common transmission directions, whereas the prompt-specific rank measures directions used only by one question. This decomposition is a natural way to test whether an observed low dimension is a property of the task or an accident of one prompt. The present paper leaves this cross-prompt analysis for future work because the controlled compendium is indexed by graph configurations rather than a trained prompt distribution.

The controlled-universe sweep runs on a single Intel Xeon node with 32 CPU cores and completes in approximately 8 CPU hours; the layered-DAG stress test adds 1 CPU hour. Memory usage stays below 4 GB throughout. The Python implementation uses NumPy and SciPy; random seeds are recorded in the supplementary archive for full reproducibility.

Q Extended Proofs and Technical Derivations

This appendix expands the proof sketches in the main text. The purpose is to make every transition in the input–state–output chain explicit. The arguments use only finite-dimensional linear algebra and Taylor expansion; no property of a particular transformer implementation is required after the local maps F_k and g have been fixed.

Q.1 Finite-horizon response with a time-varying Jacobian

Fix a nominal trajectory satisfying $\bar{z}_{k+1} = F_k(\bar{z}_k, \bar{u}_k)$. Write $e_k = z_k - \bar{z}_k$ and $v_k = u_k - \bar{u}_k$. A second-order Taylor expansion in the joint variable (z, u) gives

$$e_{k+1} = A_k e_k + B_k v_k + r_k, \qquad \|r_k\| \le \frac{L_k}{2}(\|e_k\| + \|v_k\|)^2, \tag{25}$$

where $A_k = \partial_z F_k(\bar{z}_k, \bar{u}_k)$ and $B_k = \partial_u F_k(\bar{z}_k, \bar{u}_k)$. Define $\Phi(k, j) = A_{k-1} \cdots A_j$ for $k > j$ and $\Phi(j, j) = I$. We first prove the state identity

$$e_k = \Phi(k, 0)e_0 + \sum_{j=0}^{k-1} \Phi(k, j+1)(B_j v_j + r_j). \tag{26}$$

For $k = 1$, the right-hand side is $A_0e_0+B_0v_0+r_0 = e_1$. Assume the identity holds at k. Substitution into $e_{k+1} = A_ke_k + B_kv_k + r_k$ yields

$$e_{k+1} = A_k\Phi(k,0)e_0 + \sum_{j=0}^{k-1} A_k\Phi(k,j+1)(B_jv_j + r_j) + B_kv_k + r_k$$
$$= \Phi(k+1,0)e_0 + \sum_{j=0}^{k} \Phi(k+1,j+1)(B_jv_j + r_j),$$

which closes the induction. Expanding the readout $y = g(z_N)$ around $\bar{z}_N$ gives $\delta y = C_Ne_N + \rho_N$ with $C_N = \partial_z g(\bar{z}_N)$ and $\|\rho_N\| \le L_g\|e_N\|^2/2$. Inserting equation 26 proves

$$\delta y = C_N\Phi(N,0)e_0 + \sum_{j=0}^{N-1} C_N\Phi(N,j+1)B_jv_j + \varepsilon_N, \tag{27}$$

where

$$\|\varepsilon_N\| \le \sum_{j=0}^{N-1} \|C_N\Phi(N,j+1)\|\frac{L_j}{2}(\|e_j\| + \|v_j\|)^2 + \frac{L_g}{2}\|e_N\|^2. \tag{28}$$

The bound is local: it is valid as long as the segment used by each Taylor expansion lies in the region where the Hessian bounds L_j and L_g hold. A convenient sufficient condition is obtained from the scalar recursion $a_{k+1} = \|A_k\|a_k + \|B_k\|\|v_k\| + L_ka_k^2/2$, with $a_0 \ge \|e_0\|$. If this recursion remains below the radius of the Taylor tube, then $\|e_k\| \le a_k$ by induction and the right-hand side of equation 28 is computable before running the perturbation.

The stationary affine case is the specialisation $A_k = A$, $B_k = B$, $L_k = L_g = 0$. Then $\Phi(N, j+1) = A^{N-1-j}$ and the response is exact. The time-varying expression remains the correct object when the same backbone weights are reused but the hidden state, cache, or position changes with k; weight tying does not imply $A_k = A$.

Q.2 Green kernels and uniqueness of the transmission channel

For the stationary affine recursion $e_{k+1} = Ae_k + Bv_k$ and readout $\delta y_k = We_k$, repeated substitution gives

$$e_N = A^Ne_0 + \sum_{j=0}^{N-1} A^{N-1-j}Bv_j. \tag{29}$$

The coefficient multiplying v_j is $WA^{N-1-j}B$. To see uniqueness, suppose another sequence $\widetilde{G}_\ell$ gives the same output for every finite input sequence with zero initial perturbation. Choose an input that is zero at all steps except step $N-1-\ell$, where it equals an arbitrary vector v. The output equality gives $G_\ell v = \widetilde{G}_\ell v$ for every v, hence $G_\ell = \widetilde{G}_\ell$. Thus the Green sequence is not a modelling convention: it is the unique linear input-to-output channel associated with the chosen state, input coordinates, and readout.

If $A = V\Lambda V^{-1}$ with right eigenvectors v_i and left eigenvectors w_i^*, then $A^\ell = \sum_i \lambda_i^\ell v_i w_i^*$ and therefore

$$G_\ell = \sum_i (Wv_i)\lambda_i^\ell(w_i^*B). \tag{30}$$

The expression is task-conditioned because B and W are taken from the particular prompt and readout. It is also coordinate invariant: under $\tilde{z} = Sz$, the factors become $\tilde{A} = SAS^{-1}$, $\tilde{B} = SB$, and $\tilde{W} = WS^{-1}$, so $\tilde{W}\tilde{A}^\ell\tilde{B} = WA^\ell B$.

For a repeated eigenvalue λ with spectral projector P_λ, the contribution is $\lambda^\ell WP_\lambda B$. For a Jordan block $J = \lambda I + N_J$, the corresponding term is a polynomial in ℓ times λ^ℓ because $J^\ell = \sum_{r=0}^{s-1}\binom{\ell}{r}\lambda^{\ell-r}N_J^r$. This is the reason we treat the eigen-modal expression as an interpretable special case and retain Schur reconstruction as the numerical default near defectivity.

Q.3 Controllability, observability, and the Hankel quotient

Let $\mathcal{C}_q = [B, AB, \ldots, A^{q-1}B]$ and $\mathcal{O}_p = [C^\top, (CA)^\top, \ldots, (CA^{p-1})^\top]^\top$. Stacking the output caused by a block of past inputs gives the finite block-Hankel operator

$$\mathcal{H}_{p,q} = \mathcal{O}_p \mathcal{C}_q, \qquad [\mathcal{H}_{p,q}]_{i,j} = CA^{i+j}B. \tag{31}$$

The reachable state set is $\mathcal{R}_q = \operatorname{range} \mathcal{C}_q$. The states that are invisible to the chosen future readout window are $\mathcal{N}_p = \ker \mathcal{O}_p$. Restricting $\mathcal{O}_p$ to $\mathcal{R}_q$ gives a linear map $\mathcal{O}_p|_{\mathcal{R}_q} : \mathcal{R}_q \to \operatorname{range} \mathcal{H}_{p,q}$. Its kernel is $\mathcal{R}_q \cap \mathcal{N}_p$. The first isomorphism theorem therefore yields

$$\mathcal{R}_q / (\mathcal{R}_q \cap \mathcal{N}_p) \cong \operatorname{range} \mathcal{H}_{p,q}, \tag{32}$$

and hence

$$\dim \mathcal{R}_q - \dim(\mathcal{R}_q \cap \mathcal{N}_p) = \operatorname{rank} \mathcal{H}_{p,q}. \tag{33}$$

This identity explains why the effective reasoning dimension is an input-output quantity. The ambient state dimension d can be large, and both the reachable and observable spaces can be large, while their quotient can have small rank.

The numerical rank is defined by the singular values $\sigma_1 \geq \sigma_2 \geq \cdots$ of $\mathcal{H}_{p,q}$:

$$r_\epsilon = \min \left\{ r : \frac{\sum_{j>r} \sigma_j^2}{\sum_j \sigma_j^2} \leq \epsilon \right\}. \tag{34}$$

By the Eckart–Young theorem, truncating the SVD after r_ϵ directions is the best rank-r_ϵ approximation in Frobenius norm. The same tail ratio is the relative mean-squared error under isotropic input coordinates. It is therefore meaningful to compare r_{90}, r_{95}, and r_{99} across the controlled regimes, while keeping the exact rank as a separate diagnostic.

Q.4 Branch-by-frequency representation

Let $U = [u_v]_{v \in \mathcal{V}}$ have orthonormal columns and write $z_c = Ua(c)$, where $a(c) = (a_v(c))_v$. Define $\omega_m = 2\pi m / C$ and

$$\hat{a}(m) = \frac{1}{\sqrt{C}} \sum_{c=0}^{C-1} a(c) e^{-i\omega_m c}. \tag{35}$$

The orthogonality identity $C^{-1} \sum_{m=0}^{C-1} e^{i\omega_m (c-c')} = \mathbf{1}\{c = c'\}$ implies the inverse transform

$$a(c) = \frac{1}{\sqrt{C}} \sum_{m=0}^{C-1} \hat{a}(m) e^{i\omega_m c}. \tag{36}$$

Multiplication by U gives the branch-by-frequency state representation in the main text. The inverse is exact on the finite window; no stationarity assumption is used.

Parseval follows by expanding the squared norm and applying the same orthogonality identity:

$$\sum_{c=0}^{C-1} \|z_c\|^2 = \sum_{c=0}^{C-1} a(c)^* U^* U a(c) = \sum_c \|a(c)\|^2 \tag{37}$$

$$= \sum_{m=0}^{C-1} \|\hat{a}(m)\|^2 = \sum_{m=0}^{C-1} \|\hat{z}(m)\|^2. \tag{38}$$

The answer-side identity is the same calculation after applying W: $\hat{y}(m) = WU\hat{a}(m)$ and $\sum_c \|y_c\|^2 = \sum_m \|\hat{y}(m)\|^2$.

The terminal state is evaluated at $c = C - 1$ in the inverse transform, which produces the phase factor $e^{i\omega_m (C-1)}$. This phase is a property of terminal reconstruction. It does not appear in the whole-trajectory commutation identity $\hat{y}(m) = W\hat{z}(m)$, a distinction that prevents confusing answer-side Fourier energy with the terminal answer itself.

Q.5 Answer-side energy and cross-branch interaction

For a fixed frequency m, write $M = W^*W$ and $\hat{a} = \hat{a}(m)$. Then

$$E_m^{\text{ans}} = \hat{a}^* U^* M U \hat{a}. \tag{39}$$

Expanding the quadratic form gives

$$E_m^{\text{ans}} = \sum_v \|Wu_v\|^2 |\hat{a}_v|^2 + \sum_{v \neq v'} \hat{a}_v^* \hat{a}_{v'} \langle Wu_v, Wu_{v'} \rangle. \tag{40}$$

The first term is the sum of individual branch contributions after the readout. The second term is the readout-level interaction between two superposed branches. It can be positive or negative because the readout vectors need not be orthogonal even when the latent branch vectors are. Thus latent orthogonality does not imply answer orthogonality.

The same decomposition yields two useful marginals:

$$E_v^{\text{lat}} = \sum_m |\hat{a}_v(\omega_m)|^2, \qquad E_m^{\text{lat}} = \sum_v |\hat{a}_v(\omega_m)|^2. \tag{41}$$

These are descriptive quantities. Answer relevance is determined by equation 39, not by latent energy alone. In particular, a branch may have large E_v^{lat} but small $\|Wu_v\|$, or may cancel another branch through the off-diagonal term. We do not claim that either phenomenon has been measured in a trained language model; the controlled suite only verifies the identities.

Q.6 Finite-horizon modal transfer

Suppose A is diagonalizable and consider one modal contribution to an input sequence. The finite-horizon frequency response is

$$H_{i,C}(\omega) = (Wv_i)(w_i^* B) \sum_{\ell=0}^{C-1} \lambda_i^\ell e^{-i\omega\ell} = (Wv_i)(w_i^* B) \frac{1 - (\lambda_i e^{-i\omega})^C}{1 - \lambda_i e^{-i\omega}}, \tag{42}$$

with the continuous extension C when the denominator vanishes. This formula connects the three-factor Green response to the branch-frequency view: the input and answer factors remain outside the geometric sum, while the finite reasoning horizon determines the temporal response. For $|\lambda_i| < 1$, the limit $C \to \infty$ is the usual stable transfer function. For finite C, no stability condition is needed.

Q.7 Schur reconstruction

Every complex square matrix admits a unitary Schur factorization $A = QTQ^*$ with T upper triangular. Consequently

$$G_\ell = WQT^\ell Q^* B. \tag{43}$$

The equality is exact in exact arithmetic and remains numerically stable because Q is unitary. When A is normal, T is diagonal and the Schur expression reduces to the modal expression. When eigenvectors are nearly linearly dependent, the entries of V and V^{-1} can be large even though A^ℓ is well conditioned; this is the numerical reason for reporting Schur errors separately from modal errors.

Q.8 Nonlinear and time-varying extensions

For a nonlinear rollout, the Green object is the family $G_{N,j} = C_N \Phi(N, j+1) B_j$. It is indexed by the injection step because there is no single power A^ℓ. The same controllability and observability construction applies with block columns $\Phi(t, j+1)B_j$ and future readout products $C_r \Phi(r, t)$. The branch Fourier representation also remains valid for any finite trajectory once a branch coordinate system has been specified. What changes is the interpretation: the Fourier coefficients describe the measured trajectory, while the local Green response describes its first-order perturbation. They coincide only when the branch coordinates are chosen to align with the local linearization.

For a perturbation of size δ, the first-order response error is $O(\delta^2)$ under the Hessian bounds in equation 28. This statement is uniform over a finite horizon only when the products $\|\Phi(N, j+1)\|$ are controlled. In a non-normal system, transient growth can make a small per-step remainder visible at the terminal answer; the controlled DAG stress test is included to expose this boundary rather than to suppress it.

R Expanded Path-Compositional Semantics

The controlled universe uses graph paths because they provide three semantics on the same transition structure. This appendix records the recursions and the exact relationship with the Green response.

R.1 Reachability as Boolean propagation

Let $A \in \{0,1\}^{n\times n}$ be the adjacency matrix of a directed graph and let $b = e_s$ denote the source indicator. Define $r_0 = b$ and

$$r_{c+1} = r_c \vee \mathbf{1}\{Ar_c > 0\}, \tag{44}$$

where the maximum is taken coordinatewise. By induction, $(r_c)_v = 1$ exactly when there is a walk of length at most c from s to v. The base case is immediate. For the induction step, a new vertex is added precisely when it has an incoming edge from a vertex already reached at step c; existing vertices remain reached because of the coordinatewise maximum. If the graph is acyclic, the recursion stabilizes after at most D steps, where D is the longest source-to-vertex distance.

The linear walk-counting Green response is instead $g_c = A^c b$. Its vth coordinate counts walks of exactly c edges. Thresholding g_c at zero gives exact-length reachability, while thresholding the cumulative sum $\sum_{j=0}^{c} A^j b$ gives at-most-length reachability. The distinction matters in the soft-optimal bound and is fixed throughout the experiments.

R.2 Weighted propagation

For positive edge weights $w(e)$, define $d_0(s) = 0$ and $d_0(v) = +\infty$ for $v \neq s$. The dynamic program

$$d_{c+1}(v) = \min\Big\{d_c(v), \min_{u\to v}\big[d_c(u) + w(u \to v)\big]\Big\} \tag{45}$$

computes the least path cost among all walks of length at most c. The induction is the usual last-edge decomposition: every walk reaching v at length at most $c+1$ either already has length at most c, or ends with an edge $u \to v$ appended to a walk reaching u in at most c steps. The recurrence is therefore exact even in the presence of cycles; positive weights prevent a negative-cycle ambiguity.

The weighted regime is used to test whether the Green and Hankel objects depend on binary reachability alone. The experiment compares the recurrence against an independent dynamic-programming implementation. The maximum absolute discrepancy over the reported runs is 5.3×10^{-16}.

R.3 Soft-optimal propagation and the log-sum-exp bound

Let $\Pi_c(v)$ be the set of walks from s to v of length at most c, and let $C(\pi)$ be the sum of edge weights along a walk. Define

$$Z_{\beta,c}(v) = \sum_{\pi\in\Pi_c(v)} e^{-\beta C(\pi)}, \qquad \rho_{\beta,c}(v) = -\beta^{-1}\log Z_{\beta,c}(v). \tag{46}$$

Let $C_c^*(v) = \min_{\pi\in\Pi_c(v)} C(\pi)$ and $M_c(v) = |\Pi_c(v)|$. Since $e^{-\beta C_c^*(v)} \le Z_{\beta,c}(v) \le M_c(v)e^{-\beta C_c^*(v)}$, taking $-\beta^{-1}\log$ gives

$$C_c^*(v) - \frac{\log M_c(v)}{\beta} \le \rho_{\beta,c}(v) \le C_c^*(v). \tag{47}$$

The bound is finite-horizon and does not assume that the optimal walk is unique. If the cost is averaged per step, $C_c^*(v) = d_c^*(v)c$, which gives the form used in the main text. The width is exactly $\beta^{-1}\log M_c(v)$, so it decreases monotonically in β for fixed path count. The limit $\beta \to \infty$ selects the optimal path; at small β the partition function averages many paths and the state is a genuine superposition of alternatives.

To keep exact-length and at-most-length semantics separate, define $Q_0(s) = 1$, $Q_0(v) = 0$ for $v \neq s$, and

$$Q_{c+1}(v) = \sum_{u\to v} e^{-\beta w(u\to v)} Q_c(u), \qquad Z_{\beta,c}(v) = \sum_{j=0}^{c} Q_j(v). \tag{48}$$

The first recurrence counts walks of exactly $c+1$ edges; the cumulative partition function then counts walks of length at most c. This removes the off-by-one ambiguity in the soft-optimal proof.

Algorithm 3 Matrix-free estimation of the block-Hankel spectrum

Require: rollout, input basis vectors, readout probes, horizon (p, q), tolerance ϵ

1: for each input probe, propagate $Bv, ABv, \ldots, A^{q-1}Bv$
2: for each propagated vector, apply $C, CA, \ldots, CA^{p-1}$ using reverse-mode products
3: assemble the resulting small response matrix $\mathcal{H}_{p,q}$
4: compute its singular values and r_ϵ by the tail-energy rule
5: validate the estimate with independent probes and report the residual

R.4 Cycle-shortcuts and symmetry

The cycle-shortcut family starts from a directed or undirected base cycle and adds random shortcuts. The base cycle ensures that every configuration has a recurrent component; shortcuts create multiple competing routes with different lengths and weights. Directed and undirected versions share the same semantic recursions but differ in the conditioning of the transition operator. The undirected adjacency is symmetric, giving $\mathrm{cond}(V) = 1$ in the eigendecomposition used by the sanity check. The directed version can be highly non-normal, which is why the modal failure cases and the Schur fallback are reported separately.

R.5 A small worked example

Consider a source s with edges $s \to a$, $s \to b$, $a \to t$, and $b \to t$. At step one the reachability branch vector is proportional to $u_a + u_b$; at step two it is proportional to u_t. The coefficient signals are therefore $a_a(0) = a_b(0) = 0$, $a_a(1) = a_b(1) = 2^{-1/2}$, and $a_t(2) = 1$. Their Fourier transforms contain a shared phase for the two parallel branches at the first reasoning step. If the answer readout maps u_a and u_b to aligned vectors, the off-diagonal term in equation 40 is positive; if it maps them to opposite vectors, the same latent superposition is answer-suppressing. The example is deliberately small: it shows why branch activity and answer contribution are different measurements even when the branch basis is orthogonal.

S Algorithms, Estimation, and Numerical Conventions

S.1 Matrix-free estimation from a rollout

The diagnostics do not require materialising a dense Jacobian. A Jacobian–vector product computes $A_k v$ by differentiating the latent update in the direction v. A vector–Jacobian product computes $w^* A_k$ by reverse-mode differentiation. The step kernel $G_{C,j} v = W \Phi(C, j+1) B_j v$ is evaluated by first applying B_j to v, then propagating the resulting vector forward, and finally applying the readout Jacobian. Randomized range finding can estimate the dominant singular values of $\mathcal{H}_{p,q}$ with a number of products proportional to the target rank rather than the latent width.

The branch Fourier calculation is cheaper than the state-space calculation once U is known. Extracting all coefficients costs $O(Cd|\mathcal{V}|)$ in the dense case and an FFT along the step axis costs $O(|\mathcal{V}|C \log C)$. The answer-side quadratic form can be evaluated without forming $U^* W^* W U$ by applying W to each branch vector and accumulating the Gram matrix of the resulting readout vectors.

S.2 Complex arithmetic and tolerances

The Fourier representation uses complex coefficients even when the state trajectory is real. Conjugation is therefore explicit in all energy expressions. We use the unitary DFT convention $C^{-1/2}$ so that Parseval is an equality without an additional scale factor. Relative reconstruction error is measured as $\|X - \widehat{X}\|_F / \max(1, \|X\|_F)$. A reconstruction passes when this value is below 10^{-9}; Parseval passes when the absolute discrepancy after normalisation is below 10^{-15}. These thresholds are stricter than the visual precision of the final PDF and are chosen to detect implementation mistakes in the algebra.

S.3 REGIME RECORD FORMAT

Each regime record contains the graph family, graph type, size, density, seed, task semantics, horizon, matrices or transition operators, modal condition number when available, Green reconstruction errors, Fourier reconstruction errors, state and answer Parseval errors, exact Hankel rank, and r_{90}, r_{95}, r_{99}. Soft-optimal records additionally contain β, the finite-horizon optimum, the global optimum, the mean absolute error, the empirical bound width, and Boolean pass flags for the bound, monotonicity, and optimum-match checks. This schema is sufficient to regenerate every aggregate reported in the main paper and the tables below.

ACKNOWLEDGEMENTS

This manuscript is presented anonymously; all numerical claims refer to the controlled compendium described in Appendix E.

REPRODUCIBILITY STATEMENT

The theoretical results of §4 are proved in full in Appendix D. The experimental protocol is given in Appendix E. The controlled-universe pass criteria are reported in §5.1. The source records the regime specifications and the numerical pass criteria; trained-checkpoint measurements are specified as future falsifiable tests.