HARVARD UNIVERSITY

Guest Lecture · Applied and Computational Mathematics Seminar

Transport on Simplices: Coarse-to-Fine Maps for One-Step Language Generation

Hierarchical Beckmann transport: exact flow fields, convergence for every schedule, and staggered phase transitions

Johannes Herter Harvard University

Dartmouth Department of Mathematics
Kemeny Hall, Hanover, NH

Thursday, October 8, 2026

6 Theory I: the exact field and the flow

Prop. 6.1 (structure of the field). With $c_i^{(k)}(x) = \int_0^1 \sum_{j:j_k=i} w_j \frac{p_k}{m} dt \geq 0$:

$$b^{(k)} = \sum_i c_i^{(k)} (e_i - x^{(k)}),$$

$$\frac{d}{dx}\left(\mathbf{1}^\top X_x^{(k)} - 1\right) = -a_k M \left(\mathbf{1}^\top X_x^{(k)} - 1\right),$$

and $x_i^{(k)} = 0 \Rightarrow b_i^{(k)} = c_i^{(k)} \geq 0$.

Each level is pulled towards its own vertices, drawn onto its affine hull, and never leaves $\prod_k \Delta^{n_k - 1}$.

Lemma 6.2 (one common speed). With $M(x) = \mathbb{E}\left[\frac{1}{1-\tau} \mid I_t = x\right] \geq 1$ and $\bar{p}^{(k)}(x) \in \Delta^{n_k-1}$ the tilted posterior mean,

$$b^{(k)}(x) = a_k\, M(x)\, (\bar{p}^{(k)}(x) - x^{(k)}).$$

All blocks share the scalar speed M; the level only sets the rate a_k.

Prop. 6.3 (ambient dimension). $L\sum_k v_k \geq LKV^{1/K}$, and $KV^{1/K}$ is minimal at $K = \ln V$ with value $e\ln V$.

GPT-2: 50,257 *coordinates per position become* 111 *($K = 3$, $v_k = 37$); the floor is* 29.4.

Theorem 6.4 (convergence for every schedule). For any v_k, σ_k and $a_k > 0$, for μ_0-a.e. x_0 the flow reaches M_1 at a finite $\tau(x_0)$, and

$$\mathbb{E}_{\mu_0}\tau \leq 1, \qquad T_{\#}\mu_0 = \mu_1, \qquad J_T b = 0 \text{ a.e.},$$

$$\mathbb{E}_{\mu_0}\int_0^\tau \phi(X_s)\, ds \leq \int \phi q \quad (\phi \geq 0).$$

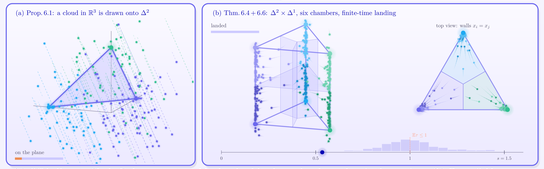
Near a vertex, $V(x) = \max_k \|x^{(k)} - e^{(k)}\|^{1/a_k}$, $M \simeq 1/V$:

$$V(x_0)/C \leq \tau(x_0) \leq V(x_0)/c,$$

$$\|X_s^{(k)} - e^{(k)}\| \leq \left(C(\tau - s)\right)^{a_k}.$$

Proof: local Lyapunov bound; Smirnov decomposition [5] of the current $q\,b$, $\nabla\cdot(qb) = \mu_0 - \mu_1$.

The one-step map exists for every schedule, including the self-stopping $a_k > 1$ that DBTM's theorem does not cover.



(a) Off-hull parts (dashed) decay at rate $a_k M$; the cloud lands on the plane, then in the vertices. (b) Exact field. $a = (5, 1)$: walls $x_i = x_j$ are invariant, each particle stays in its chamber and lands at its vertex in finite time; clusters resolve first. Timeline: landing times vs. $\mathbb{E}\tau \leq 1$.

Theorem 6.5 (noise memory). In the flow time $\varsigma = \int M\, ds$, for *any* data,

$$X^{(k)} = e^{-a_k\varsigma} x_0^{(k)} + \left(1 - e^{-a_k\varsigma}\right) y^{(k)}(\varsigma), \quad y^{(k)} \in \Delta^{n_k - 1},$$

so $\lambda_\varsigma = \lambda_0^{a_\varsigma/a_f}$ and level k crosses its threshold at $\varsigma_k^\star = -\ln(1 - t_k^\star)$.

The flow carries exactly the prior noise $\alpha_k(t)x_0$ of the interpolant at $t = 1 - e^{-\varsigma}$: each level is denoised on its own schedule.

Theorem 6.6 (the flow keeps the order). If μ_1 is invariant under swapping symbols i, i' of a block, the hyperplane $x_i = x_{i'}$ is invariant. For product-uniform data the order of every block's coordinates is kept, and

$$T(x_0)^{(k)} = e_{\arg\max_i x_{0,i}^{(k)}}, \qquad T_{\#}\mu_0 = \mu_1.$$

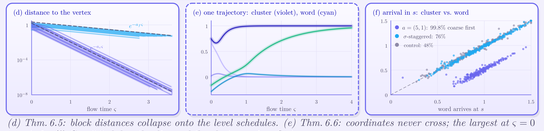
Each decision is fixed at time 0; the flow decides when *it becomes visible, level by level.*

Theorem 6.7 (arrival times). With $K_k = \int_0^\infty a_k e^{a_k s} \|\bar{p}^{(k)} - e^{(k)}\| ds < \infty$ a.s.,

$$\frac{1}{a_k}\ln\frac{\mathbf{1}^\top x_0^{(k)} - 1}{\varepsilon\sqrt{v_k}} \leq \varsigma_k^\star \leq \frac{1}{a_k}\ln\frac{\|x_0^{(k)} - e^{(k)}\| + K_k}{\varepsilon}.$$

Cor. 6.8. If $a_c > a_f$, for μ_0-a.e. x_0 and all small ε every coarse block reaches its vertex before the fine block of its position.

Coarse-to-fine is a property of the flow itself, not only of the interpolant posterior.



(d) Thm. 6.5: block distances collapse onto the level schedules. (e) Thm. 6.6: coordinates never cross; the largest at $\varsigma = 0$ goes to 1. (f) Cor. 6.8: below the diagonal the cluster arrives first.

1 Background: Beckmann transport

A token is a vertex of Δ^{V-1}, a sequence a point of $(\Delta^{V-1})^L$; data live on $M_1 = \{e_1, \ldots, e_V\}^L$. DBTM [1,2] learns a *one-step* map from a *time-independent* flow.

$$I_t = \alpha_t x_0 + \beta_t x_1, \quad \alpha_t = (1-t)^a, \quad x_0 \sim \mathcal{N}(0, \sigma^2 I)$$

$$b(x) := \mathbb{E}_{t, x_0, x_1}\left[\dot{I}_t \mid I_t = x\right], \qquad \dot{X}_s = b(X_s)$$

Prop. 3.1 [1] (time average). $b = \frac{\int_0^1 \mu_t b_t\, dt}{\int_0^1 \mu_t\, dt}$, $\nabla\cdot j = \mu_0 - \mu_1$, $j = \int_0^1 \mu_t b_t\, dt$.

Thm. 3.1 [1] (convergence). For μ_0-a.e. x_0: absorption; $X_s(x_0) \to x^* \in M_1$; for $a = 1$ the hitting time $\tau(x_0) = \inf\{s : X_s \in M_1\}$ is finite, as $\lim_{r\downarrow 0} b(e_j + r\omega) = -\kappa(\omega)\omega$.

$$T(x_0) := \lim_{s\to\tau(x_0)} X_s(x_0), \qquad T_{\#}\mu_0 = \mu_1, \qquad T \circ T = T$$

Prop. 3.2 [1,2] (conservation). $b\cdot\nabla X_s = \partial_s X_s$, and T uniquely solves

$$b(x)\cdot\nabla T(x) = 0 \quad (x \notin M_1), \qquad T(x) = x \quad (x \in M_1).$$

2 Training DBTM, and its limitation

$$X^{(n+1)} = X^{(n)} + \eta\, b\cdot\nabla X^{(n)}$$

$$\mathcal{L}_{\text{transport}} = \mathbb{E}\sum_{\ell\notin C}\left\|T_\theta(I_t)^\ell - \mathrm{sg}(T_\theta(I_t) + \dot{I}_t\cdot\nabla T_\theta(I_t))^\ell\right\|^2$$

$$\mathcal{L}_{\text{semi}} = \mathbb{E}\sum_{k,\ell}\mathrm{CE}(T_\theta(I_t)^\ell, \mathrm{sg}(T_\theta^{(k)}(I_t)^\ell)), \quad \mathcal{L}_{\text{bnd}} = \mathbb{E}\sum_\ell \mathrm{CE}(T_\theta(x_1)^\ell, x_1^\ell)$$

$$\mathcal{L}_{\text{anchor}} = \mathbb{E}\sum_\ell \mathbf{1}[t > t^\star]\,\mathrm{CE}(T_\theta(I_t)^\ell, x_1^\ell)$$

$$I_t^\ell(C) = x_1^\ell \ (\ell \in C), \quad \alpha_t^\ell x_0^\ell + (1-\alpha_t^\ell)x_1^\ell \ (\ell \notin C); \quad q^\ell = (T_\theta(x), \hat{x}_1)$$

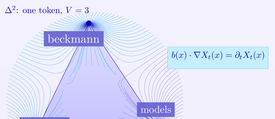
$$\Delta C_r = \{\ell \in R_r : q^\ell \geq \kappa\} \cup \mathrm{top}_{n_r}(R_r; q), \quad n_r = \lceil |R_r|/(k - r + 1)\rceil$$

Thm. 4.1 [1] (phase transition, asymptotic in V). The interpolant commits to a vertex at

$$t^\star = 1 - \left(1 + \sigma\sqrt{2\log V}\right)^{-1/a},$$

a REM-type transition sharpening to a step as $V \to \infty$.

Limitation: one flat simplex. $t^\star$ grows with $\log V$: at $V = 50{,}257$ every token is unresolved until one late, sharp transition; all tokens are equidistant; each position lives in $\mathbb{R}^V$.

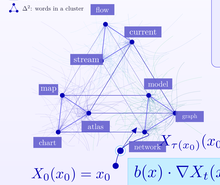


DBTM [1] on one simplex Δ^2 ($V = 3$), as in [1, Fig. 1], computed exactly from $b(x) = \mathbb{E}[\dot{I}_t \mid I_t = x]$: every streamline flows straight to a single vertex, all decided in one transition. Compare ours (centre), which decides the cluster first.

3 Ours: coarse-to-fine transport on hierarchical simplices

Factor the vocabulary $V = v_1 \cdots v_K$ (cluster $\to \cdots \to$ word). A token becomes a tuple of vertices of small simplices, token $\mapsto (e^{(1)}, \ldots, e^{(K)}) \in \prod_k \Delta^{v_k - 1}$, and each level gets its own schedule $\alpha_t^{(k)} = (1-t)^{a_k}$ with coarse levels faster, $a_1 > \cdots > a_K$.

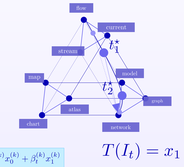
coarse-to-fine autonomous flow and transport map, computed exactly (Prop. 1)



$$b(x)\cdot\nabla X_t(x) = \partial_t X_t(t)$$

Coarse-to-Fine Map $T : \prod_{k=1}^K (\Delta^{v_k - 1})^L \to M_1$

$$t_k^\star = 1 - \left(1 + \sigma\sqrt{2\log v_k}\right)^{-1/a_k}$$



$$I_t^{(k)} = \alpha_t^{(k)} x_0^{(k)} + \beta_t^{(k)} x_1^{(k)}$$

$$T(I_t) = x_1$$

$$T = \arg\min \mathbb{E}_{t,x_0,x_1}\sum_{k=1}^K\left[\left\|T_\theta^{(k)}(I_t) - \mathrm{sg}(T_\theta^{(k)}(I_t) + \dot{I}_t\cdot\nabla T_\theta^{(k)}(I_t))\right\|^2 + \mathbf{1}[t > t_k^\star]\,\mathrm{CE}(T_\theta^{(k)}(I_t), x_1^{(k)})\right]$$

Our method on *Δ^2(cluster) $\times$ Δ^2(word), the 3-3 duoprism (9 words; its 4D hull drawn by a linear map to 3D). Left: the exact coarse-to-fine autonomous field (Prop. 1); every streamline is first pulled into a cluster's word triangle (violet) and only then resolves a word. Right: the coarse-to-fine interpolant leaves the flow cluster, commits at $t_1^\star$ (cluster), then $t_2^\star$ (word), and T maps every point on it to x_1 = network.*

Proposition 1 (exact field). For $\mu_1 = \sum_j \pi_j \delta_{e_j}$, $\mu_0 = \mathcal{N}(0, \sigma^2 I)$:

$$b(x) = \int_0^1 \sum_j w_j(t, x)\frac{a_k\left(e_j^{(k)} - x^{(k)}\right)}{1-t}\Big|_{k=1\ldots K} dt,$$

$$w_j \propto \pi_j \prod_k \left(\alpha_i^{(k)}\right)^{-v_k}\exp\left(-\frac{\|x^{(k)} - (1-\alpha_t^{(k)})e_j^{(k)}\|^2}{2\sigma^2(\alpha_t^{(k)})^2}\right)$$

normalised over (j, t). *Proof:* Eq. (7) of [1] with Gaussian conditionals.

Proposition 2 (factorisation). If $\mu_1 = \bigotimes_k \mu_1^{(k)}$:

$p(x_1 \mid I_t) = \prod_k p(x_1^{(k)} \mid I_t^{(k)})$, with

$\log p(x_1^{(k)} = e_i \mid I_t^{(k)}) = x) = c + \frac{(1-\alpha_t^{(k)})x_i}{(\alpha_t^{(k)})^2\sigma^2}$ for uniform $\mu_1^{(k)}$.

Algorithm 1 (coarse-to-fine refinement). $x \leftarrow x_0$, $C \leftarrow \emptyset$; for $r = 1..k$: $\hat{x} \leftarrow T_\theta(x)$; commit each (position, level) with $q \geq \kappa$, coarse levels first (at least n_r per round); renoise the uncommitted levels; return the per-level argmax. One NFE per round.

Theorem 3 (staggered commitment). For product μ_1, the sharp threshold (Thm. 7.2) holds per level, independently across blocks, at

$$t_k^\star = 1 - \left(1 + \sigma_k\sqrt{2\log(v_k - 1)}\right)^{-1/a_k},$$

$$a_1 > \cdots > a_K, \ \sigma_k s_{v_k} \uparrow, \ v_k \geq 3 \implies t_1^\star < \cdots < t_K^\star.$$

Each transition involves $\log v_k$, not $\log V$; Monte Carlo matches the prediction for every V (right).

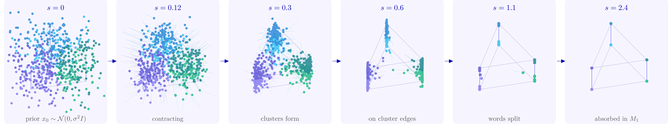
Level-wise objective. Transport loss per level, anchor after each level's own $t_k^\star$:

$$\mathcal{L} = \mathbb{E}\sum_k\left[\left\|T_\theta^{(k)} - \mathrm{sg}(T_\theta^{(k)} + \dot{I}_t\cdot\nabla T_\theta^{(k)})\right\|^2 + \mathbf{1}[t > t_k^\star]\,\mathrm{CE}(T_\theta^{(k)}, x_1^{(k)})\right]$$

$$\dim \text{ per position}: \ V \ \longrightarrow \ \sum_k v_k \qquad (50{,}257 \to 3\cdot 37 = 111)$$

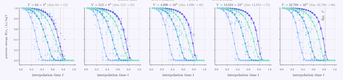
Every schedule converges (Thm. 6.4). For unequal v_k and every $a_k > 0$ the flow is absorbed in finite time, $T_{\#}\mu_0 = \mu_1$ and $\mathbb{E}\tau \leq 1$; DBTM's theorem covers equal simplices and $a = 1$.

(a) Particles under the autonomous flow $\dot{X}_s = b(X_s)$



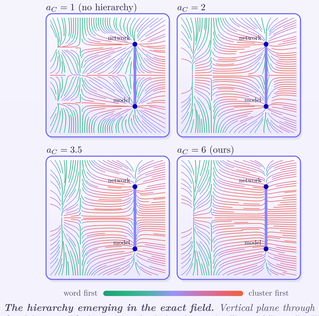
Our flow in motion. *900 prior particles under $\dot{X}_s = b(X_s)$ (Prop. 1): they collapse onto the cluster edges first, then split into words and are absorbed in M_1.*

4 Results

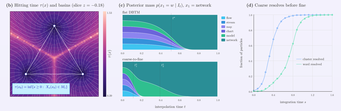


Posterior entropy $H(x_1 \mid I_t)$, flat DBTM vs. three levels ($a = 3, 1.6, 1$; $\sigma = 1$). Dotted: Thm. 3's $t_k^\star$, matched by simulation for every V.

V	flat $t^\star$	$t_1^\star$	$t_2^\star$	$t_3^\star$	dim	ours
64	0.74	0.28	0.46	0.62	64	12
512	0.78	0.31	0.50	0.67	512	24
4,096	0.80	0.33	0.53	0.70	4,096	48
13,824	0.81	0.34	0.54	0.72	13,824	72
32,768	0.82	0.35	0.55	0.72	32,768	96



The hierarchy emerging in the exact field. *Vertical plane through the {model, network} cluster edge (violet); colour = $\log(|b_{\text{cluster}}|/|b_{\text{word}}|)$. With one schedule ($a_C = 1$) flows head straight for the words; as a_C grows they first travel to the cluster edge, then along it.*



Dynamics of the coarse-to-fine flow: (b) hitting time $\tau(x)$ and basins; (c) posterior mass, flat DBTM vs. ours; (d) clusters resolve before words.

Conclusion. (1) One late commitment becomes K staggered ones, governed by $\log v_k$, along the interpolant *and* along the flow. (2) The flow converges for every schedule; the state shrinks from V to $\sum_k v_k$. (3) At matched energy the hierarchy gives no free lunch; what it buys is a state that fits through the input layer: width d reads $\sum_k v_k \leq d$ coordinates without loss, $V \gg d$ not (Sec. 5).

References. [1] S. Tang, S. Wang. Discrete Beckmann Transport Models... arXiv:2609.15903, 2026. [2] C.-K. Lee *et al.* Beckmann transport models. arXiv:2608.01692, 2026. [3] M.S. Albergo, E. Vanden-Eijnden. Stochastic interpolants. ICLR 2023. [4] B. Derrida. Random-energy model. PRL 1980. [5] S.K. Smirnov. Decomposition of solenoidal

7 Theory II: commitment, correlation, training

Lemma 7.1 (posterior of the truth). One simplex, n vertices, $g = x_0/\sigma$, SNR $\rho = \frac{1-\alpha}{\alpha\sigma}$:

$$\pi_\star = \left(1 + \sum_{j\neq i^\star} e^{\rho(g_j - g_{i^\star}) - \rho^2}\right)^{-1}.$$

Whether a token is decided depends on ρ alone.

Theorem 7.2 (sharp threshold, any n). $s_n = \sqrt{2\log(n-1)}$.

$$\rho \geq s_n + 2u: \ \mathbb{P}(1 - \pi_\star \leq e^{u' - \rho(\rho - s_n - 2u)}) \geq 1 - 2e^{-\frac{u^2}{2}} - e^{-u'},$$

$$\rho \leq s_n - h - u - \eta: \ \mathbb{P}(\pi_\star \leq e^{-\rho\eta}) \geq 1 - e^{-\frac{u^2}{2}} - e^{-\Lambda_n(h)}.$$

DBTM's random-energy heuristic, made a two-sided theorem with explicit probabilities.

Cor. 7.3 (sharpness). As $n \to \infty$: $\rho \geq (1+\varepsilon)s_n \Rightarrow \pi_\star \to 1$, $\rho \leq (1-\varepsilon)s_n \Rightarrow \pi_\star \to 0$ in probability. **Cor. 7.4 (time).** For $\alpha = (1-t)^a$:

$$t^\star(n, a, \sigma) = 1 - \left(1 + \sigma\sqrt{2\log(n-1)}\right)^{-1/a},$$

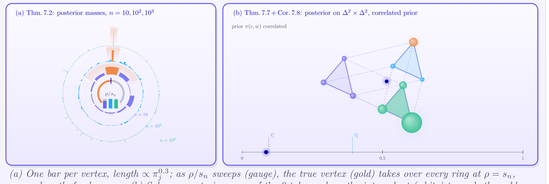
increasing in n, σ, decreasing in a.

Small simplices and fast schedules commit early.

Cor. 7.5 (staggering by noise). With $a_k = 1$, $v_k \geq 3$ and $\sigma_k s_{v_k}$ increasing, $t_k^\star = \frac{\sigma_k s_{v_k}}{1 + \sigma_k s_{v_k}}$ is increasing. **Prop. 7.6 (no free signal).** Scaling vertices by $c = K^{-1/2}$ matches a flat token's energy; then the finest level commits at

$$\sqrt{2K\log(v-1)} \ \leq \ \sqrt{2\log(V-1)}, \qquad \frac{K\log(v-1)}{\log(v^K - 1)} \to 1.$$

The hierarchy re-times decisions; it does not lower the total signal needed.



(a) One bar per vertex, length $\propto \pi^{0.3}$; as ρ/s_n sweeps (gauge), the true vertex (gold) takes over every ring at $\rho = s_n$, more abruptly for larger n. (b) Spheres: posterior mass of the 9 tokens along the interpolant (white) towards the gold token under a correlated prior; the prior first favours the green cluster, the true cluster's triangle lights up at $t_1^\star$, the word at $t_2^\star$.

Theorem 7.7 (correlated data). For any μ_1, block B's log-odds are $\rho(g_j - g_{i^\star}) - \rho^2 + \Delta_j$, with offsets independent of B's noise and $|\Delta_j| \leq \log\Gamma_{B|F} + D_F$; the threshold holds with the error times $\bar{w} = \frac{1}{m}\sum_j e^{\Delta_j}$, a shift of $\frac{\log\bar{w}}{\rho}$. Exactly, $\pi_\star^B = \sum_o \mathbb{P}(o \mid I_t)\,\pi_\star(o)$ over contexts o.

Correlation costs a logarithm; frequent symbols commit early, rare ones late.

Cor. 7.8 (two levels, correlated). Under bounded odds Γ_1, Γ_2,

$$1 - \pi_\star^{\text{tok}} \leq \Gamma_1 e^{b_\star} e^{u' - \rho_1(\rho_1 - s_{v_1} - 2u)},$$

$$1 - \pi_\star^{\text{tok}} \leq (1 - \pi_c^\star) + \Gamma_2 e^{u' - \rho_2(\rho_2 - s_{v_2} - 2u)}.$$

Given its cluster, a word competes only with the $n_c - 1$ valid words of that cluster.

Prop. 7.9 (regression step). The minimiser of $\mathbb{E}\|T(I_t) - T(I_t) + J_T I_t\|^2$ is $T^* = T + J_T b$, blockwise; T is a fixed point iff $J_T b = 0$ off the data. **Prop. 7.10 (anchoring).** $\mathbb{E}\,\mathrm{CE}(q, x_1^{(k)})$ is minimised by the posterior $p^{(k)}$, which differs from the anchor target by $\mathbb{E}\|p^{(k)} - x_1^{(k)}\|_1 = 2\,\mathbb{E}[1 - \pi_\star^{(k)}]$.

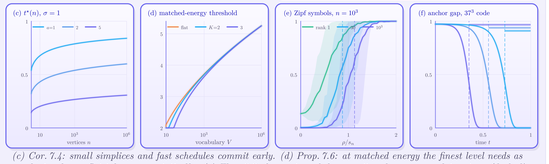
Training keeps DBTM's principle; anchor level k only after its own $t_k^\star$.

Theorem 7.11 (from loss to samples). For a learned T_θ,

$$W_1(T_{\theta\#}\mu_0, \mu_1) \leq \mathbb{E}\|T_\theta - T\| \leq \varepsilon_{\text{bnd}} + \mathbb{E}_q\|J_{T_\theta} b\|,$$

$(\mathbb{E}_q\|J_{T_\theta} b\|)^2 \leq$ the excess of the transport loss over one regression step, and the expected number of wrong tokens is $\leq \frac{\sqrt{2K\varepsilon}}{c}(\varepsilon_{\text{bnd}} + R)$.

Proof: integrate $\frac{d}{dt}T_\theta(X_s) = J_{T_\theta} b$ along the flow; its occupation is at most q.



(c) Cor. 7.4: small simplices and fast schedules commit early. (d) Prop. 7.6: at matched energy the finest level needs as much signal as a flat token, asymptotically. (e) Thm. 7.7: frequent symbols commit early, rare ones late, at p_w (dashed). (f) Prop. 7.10: each level's gap vanishes after its own $t_k^\star$, where its anchor switches on (bars).

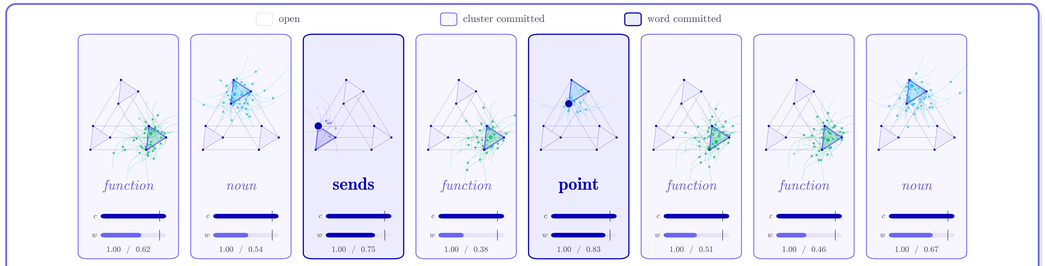
5 Coarse-to-fine refinement, and what the hierarchy does and does not buy

Every token is a point of $\Delta^2_{\text{cluster}} \times \Delta^2_{\text{word}}$ (one duoprism per card). After each evaluation of T_θ, every level whose confidence exceeds κ is committed, coarse levels first; the rest are renoised and the same map is applied again. Committing a cluster restricts the exact field to that cluster's word triangle. *Shown: "the map sends each point to the word", exact per-level posteriors, $\kappa = 0.9$, $k = 4$.*

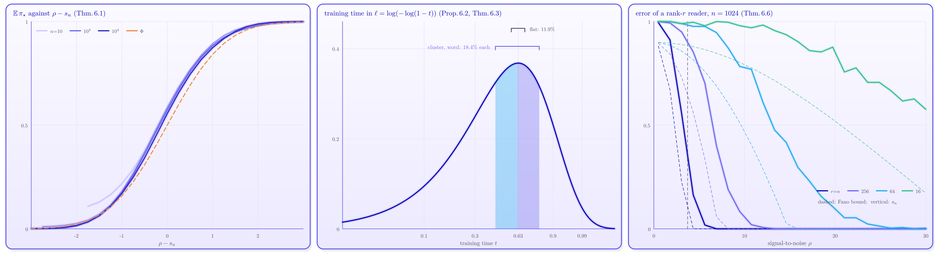
model	schedule	$r = 1$	2	3	4	accuracy
flat DBTM	$a = 1$	3.0	6.0	9.5	12.0	68.3%
flat DBTM	$a = 5$	9.4	12.0	12.0	12.0	**98.9%**
ours	$a = (5, 1)$	0.0	3.7	10.4	12.0	93.7%
ours, energy-matched	$a = (5, 1)$	0.0	1.8	7.4	12.0	92.4%
ours, energy-matched	$a = (1, 1)$	0.0	0.1	6.5	12.0	49.7%

Exact posteriors, 12 tokens, 2,000 draws: tokens committed after round r, final accuracy. A hierarchical token carries K times the signal energy; at matched energy and schedule it is not better, and a flat token on the fast schedule $a = 5$ wins. With exact posteriors, accuracy tracks the schedule, not the code (Prop. 7.6).

Algorithm 1 (coarse-to-fine refinement). $x \leftarrow x_0$, $C \leftarrow \emptyset$; for $r = 1..k$: $\hat{x} \leftarrow T_\theta(x)$; commit every block with confidence $q \geq \kappa$ whose coarser blocks are committed, and at least $n_r = \lceil |R_r|/(k - r + 1)\rceil$ frontier blocks; renoise the rest; return the per-block argmax. Terminates in exactly k evaluations and never commits a word before its cluster.



When does the hierarchy help? Commitment shape, training time per decision, and the embedding bottleneck (paper Sec. 6)



Thm. 6.1 (commitment profile). $\mathbb{E}\,\pi_\star(s_n + x) \to \Phi(x)$ uniformly on compacts, and $\lambda/s_n - (x + g_\star) \to 0$: every commitment has the same Gaussian shape in ρ, centred at $s_n = \sqrt{2\log(n-1)}$; a single sample commits within $O(1/s_n)$.

Thm. 6.3, Prop. 6.4 (training time). Uniform t is Gumbel in ℓ; the exponent moves a window, the scale sets its length $< L_s = \log\frac{a+s}{a-s}$.

Share $< M(L) = e^{-\frac{L}{a-1}} - e^{-\frac{L_s C}{a}} \approx \frac{2L}{C}$. $V = 50{,}257$: flat 11.9%; each level of a 2-level code 18.4%, $a = (1.29, 0.76)$.

Thms. 6.6–6.7 (bottleneck). Through rank r: $\mathbb{P}(\hat{i} \neq i^\star) \geq 1 - \frac{\frac{1}{2}\log(1+\rho^2/n) + \log 2}{\log n}$, so commitment needs $\rho \approx s_n\sqrt{n/r}$, and the flat map $T(x_0) = e_{\arg\max x_0}$ is recovered w.p. $\leq \frac{2r\log V}{V} + o(1)$: $\leq 15.7\%$ for DBTM ($V = 50{,}257$, width 768). Coarse-to-fine (2×225 coordinates) is read without loss.