TOWARDS SOLVING THE UNMASKING ORDER IN DLLMS VIA CONSTRAINED MAXENTROPY SAMPLING

Anonymous authors
Paper under double-blind review

ABSTRACT

A masked diffusion language model (dLLM) must choose which masked position to denoise at every step. Various rules exist for this choice, such as committing the most confident position (confidence order) and sampling positions in proportion to a tempered confidence (tempered selection). They are stated as separate procedures with no common objective, and tempered selection keeps every position eligible, so it cannot defer a commitment. We present **PACE**, a *position sampling strategy with adaptive confidence eligibility*, which keeps the odds of tempered selection but defers the least confident positions under an entropy budget relative to tempered selection. PACE follows from one constrained optimization problem: find the highest-entropy distribution over positions whose expected *damage*, the cost of a commitment, stays within a budget. Its solution is a Gibbs distribution over positions, and each rule we study corresponds to one choice of damage function, eligible position set and Lagrange multiplier, which is the inverse position temperature. A damage that depends on the committed token's probability and adds over successive commitments must be the surprisal, which gives the odds of tempered selection, and we prove that, among rules with these odds that defer only the least confident positions and meet the budget, PACE is the unique one that defers the fewest. We also show that selecting positions by the tempered confidence of their drawn tokens sharpens the distribution of written tokens toward an effective temperature that combines the token and position temperatures. On two model families and four benchmarks at token temperature 0.8, PACE improves single-sample accuracy on all eight model–benchmark pairs at the same compute, by 4.1 to 9.1 points over confidence order and 10.2 to 28.1 points over tempered selection, and leads at every number of samples up to 32; its single-sample gain over tempered selection holds at a second token temperature.

1 INTRODUCTION

Masked diffusion language models (dLLMs) generate text by repeatedly unmasking positions of a fixed-length sequence, the *canvas*. The order is set at decoding time rather than fixed left to right (Nie et al., 2025; Ye et al., 2025). This freedom is the defining feature of the architecture. It also creates a decision that autoregressive models do not face: at each step, the decoder must choose which masked position to fill next. We say a position is *committed* once its token is fixed.

This choice matters because every later prediction conditions on the committed tokens, so the order shapes what the model can still produce. A poor order reduces accuracy (Ni et al., 2026). On reasoning tasks the freedom to choose an order can itself be harmful (Yeom et al., 2026). Changing the order alone also changes the diversity of what the model produces (Olausson et al., 2026).

Limitations of existing rules. Prior work has proposed many heuristics based on the model's confidence at each position. Representative ones commit the most confident position or every position above a threshold (Wu et al., 2026), sample in proportion to a tempered confidence (Chang et al., 2022; Olausson et al., 2026), restrict eligibility to a sliding frontier (Yeom et al., 2026), gate on cross-step stability (Wang et al., 2026), or admit the largest set whose summed token entropies stay under a threshold (Ben-Hamu et al., 2025). Each rule is motivated separately, with no common objective under which to compare them and no explanation of why their hyperparameter values work.

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.

Tempered selection, which samples positions, also keeps every position eligible at every step, so it cannot defer a commitment and some of its commitments fall on positions the model rates poorly; its single-sample accuracy is 3 to 23 points below confidence order (Table 6).

Our approach. We write position selection as one constrained optimization: choose the position as randomly as possible, measured by entropy, subject to a budget on the expected *damage* $D(k)$ of committing at position k. The solution is $q_\lambda(k) \propto e^{-\lambda D(k)}$. Here $\lambda \geq 0$, the Lagrange multiplier of the damage budget, is the price of damage, and the position temperature T_{pos} practitioners tune is $1/\lambda$. The rules listed above can be written in this form (Table 1), in two parts: the odds between any two eligible positions, and which positions are eligible. For the odds, the damage is the surprisal $-\log c_k$ of the committed token, where c_k is the model's probability of the token drawn at k. Summed over a generation, the surprisal is the negative log-likelihood the model assigns to its own output along the order it was written (Equation (2) and Lemma 2). PACE keeps these odds and defers the least confident positions (Figure 2). At every step it takes the longest prefix of the confidence order whose entropy is at most a fraction α of the entropy of tempered selection; this is the maximum-entropy deferral rule (Proposition 4). A deferred position is redrawn at the next step rather than excluded, so every output keeps positive probability. PACE thus addresses both limitations above: it follows from a stated objective, and it defers the low-confidence commitments that tempered selection makes.

Main results. PACE is training-free and costs no extra forward passes. It adds one parameter to tempered selection, the budget α. At token temperature $T_{\text{tok}} = 0.8$ and the same compute, it improves single-sample accuracy over confidence order and tempered selection on all eight model–benchmark pairs (Figure 1), and it leads at every number of samples up to 32.

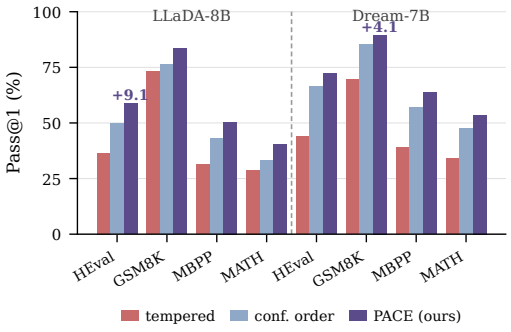

Figure 1: Single-sample accuracy (Pass@1) at $T_{\text{tok}} = 0.8$ on all eight model–benchmark pairs at the same number of forward passes; the two labels are PACE's smallest and largest gain over confidence order. Values from Table 6.

Contributions. (i) The position–token coupling: because a candidate is drawn at every masked position before a position is chosen, selecting positions in proportion to a tempered confidence of their drawn tokens places the committed token's odds between those at T_{tok} and those at an effective temperature T_{eff}, with $1/T_{\text{eff}} = 1/T_{\text{tok}} + 1/T_{\text{pos}}$, while a position-only score leaves them at T_{tok} (Theorem 5), consistent with the collapse we measure at high token temperature. (ii) A maximum-entropy account of unmasking order in two parts, odds and eligibility, in which the surprisal is the damage whose sum is the output's negative log-likelihood (Section 4), and PACE is the maximum-entropy deferral rule under a budget relative to the entropy of tempered selection (Propositions 4 and 10). (iii) Why PACE defers instead of changing the temperature: deferral removes the tail at the reference odds, while a temperature that suppresses the tail must distort the odds between the leading positions, with an exact exponent (Proposition 6); at every fixed state, PACE's committed confidence stochastically dominates that of tempered selection (Proposition 7). (iv) Experiments on two models and four benchmarks: gains on all eight pairs, a replication of the gain over tempered selection on five of them at a second token temperature with 164 to 200 problems, and ablations that isolate the effect of the budget at a fixed position temperature (Section 7).

2 Related Work

Order selection in dLLMs. Confidence-ordered unmasking is the default in LLaDA's reference decoder (Nie et al., 2025); Dream's decoder also offers margin and entropy orderings (Ye et al., 2025), and parallel variants commit every position above a threshold (Wu et al., 2026). Stochastic order dates back to MaskGIT's tempered confidence (Chang et al., 2022) and has been adapted to dLLMs (Olausson et al., 2026). Later rules restrict the order to a frontier (Yeom et al., 2026), condition it on cross-step stability (Wang et al., 2026), rank positions by margin (Kim et al., 2025), admit a set of positions under a bound on summed token entropies (Ben-Hamu et al., 2025), or weigh a

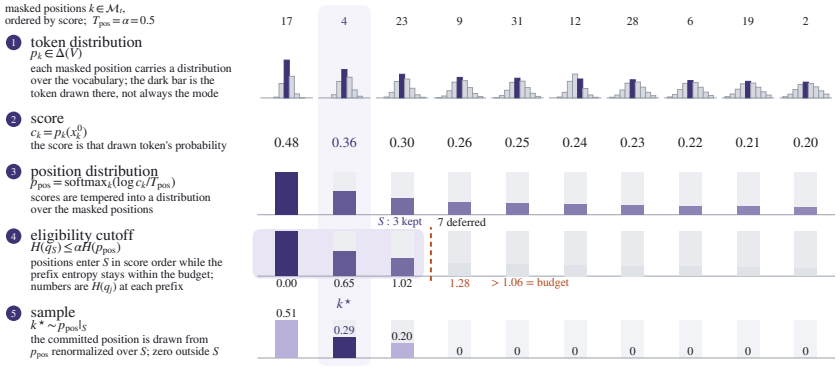

Figure 2: One step of PACE (Algorithm 1) on ten masked positions, $T_{\text{pos}} = \alpha = 0.5$. Positions join S in confidence order while the prefix entropy stays below $\alpha\mathbb{H}(p_{\text{pos}}) = 1.06$; adding the fourth would give 1.28, so three are eligible and seven deferred (row 4).

commitment's effect on the positions that remain (Yang et al., 2026). Order has also been planned or learned (Liu et al., 2025; Peng et al., 2025; Jazbec et al., 2026). In contrast to P2 (Peng et al., 2025), which unifies samplers through a planner inside an expanded evidence lower bound, we separate each rule into its odds and its eligible set, give the odds a likelihood meaning (Equation (2)), and show how position selection acts on the committed token (Theorem 5).

Quality and exploration. Fang et al. (2026) trade off single-sample accuracy against coverage with an independent Metropolis–Hastings sampler that corrects each one-token decoding step using batched lookahead proposals, and Ni et al. (2026) post-train with left-to-right exploration. In contrast, PACE changes only the per-step distribution over positions, at inference, with no proposals and no extra forward passes.

Truncation of the vocabulary. Token samplers truncate the vocabulary before sampling: nucleus sampling (Holtzman et al., 2020), η-sampling (Hewitt et al., 2022), min-p (Nguyen et al., 2025) and perplexity control (Basu et al., 2021). Truncating tokens removes outputs and is justified as undoing the model's smoothing (Hewitt et al., 2022; Finlayson et al., 2024). PACE runs the scan of entropy-bounded truncation (Baghaei Potraghloo et al., 2025) on the position distribution (Section 4.3). On positions, it removes no output, since a deferred position is committed later in the block (Remark 13), and it acts on the committed token only by choosing among drawn candidates.

3 Preliminaries

A masked dLLM starts from a canvas whose response region is fully masked. It then repeatedly runs a forward pass, selects a set of masked positions, and writes a token at each. Let V be the vocabulary, L the generation length, and $\mathcal{M}_t$ the set of positions of the current block still masked at step t. Decoding proceeds in blocks of length B; with K positions per step, a block costs B/K forward passes. All rules share this schedule and differ only in how they choose among $\mathcal{M}_t$.

At each step the model outputs, for every masked position k, a distribution $p_k \in \Delta(V)$ over the vocabulary. At every masked position a *candidate token* x_k^0 is drawn from the tempered distribution $\text{softmax}(\ell_k/T_{\text{tok}})$, where ℓ_k are the logits and T_{tok} is the token temperature. Following prior work (Wu et al., 2026; Olausson et al., 2026), the *confidence* at position k is the probability the model assigns to that candidate, $c_k = p_k(x_k^0)$. A position *rule*, or sampler, maps the confidences $\{c_k\}$ to a distribution q over $\mathcal{M}_t$. The committed position is drawn from q and receives the candidate token already drawn there. Two rules are used throughout: **confidence order**, which puts all probability on $\arg\max_k c_k$, and **tempered selection** with position temperature T_{pos}, $q(k) \propto c_k^{1/T_{\text{pos}}}$.

We work in the sampling regime $T_{\text{tok}} > 0$, in which several samples per problem differ and Pass@k is informative. We report Pass@k, the percentage of problems solved by at least one of k samples,

Table 1: Representative published position rules as Gibbs distributions (Proposition 1), in two parts: the odds among eligible positions, set by D and λ, and the eligible set. D is the damage on eligible positions and $+\infty$ off the eligible set; set-valued rules commit every eligible position at once instead of sampling one. The rules that defer either take $\lambda \to \infty$ or commit their whole eligible set, and tempered selection samples with a finite λ but never defers. PACE samples with a finite λ from an eligible set whose cutoff adapts at every step. Set-valued rules use no odds, and stability gating's damage depends on each position's trajectory. The tempered row fixes the temperature of the Gumbel-perturbed confidence ranking of Chang et al. (2022), who anneal it to zero across steps; the row follows from $e^{-\lambda D(k)} = c_k^{1/T_{\text{pos}}}$.

Rule	$D(k)$, eligible k	Eligible set	λ
Confidence order (Nie et al., 2025)	$-\log c_k$	$\mathcal{M}_t$	$\to \infty$
Confidence threshold (Wu et al., 2026)	$-\log c_k$	$\{k : c_k > \tau_{\text{CT}}\}$	set-valued
Tempered selection (Olausson et al., 2026)	$-\log c_k$	$\mathcal{M}_t$	$1/T_{\text{pos}}$
Frontier gating (Yeom et al., 2026)	$-\log c_k$	window at the leftmost mask	$\to \infty$
Stability gating (Wang et al., 2026)	decreasing in readiness	stable positions	set-valued
Entropy-bounded set (Ben-Hamu et al., 2025)	$\mathbb{H}(p_k)$	admitted prefix	set-valued
Uniform order	any	$\mathcal{M}_t$	$\to 0$
PACE (this work)	$-\log c_k$	top $j^\star$ by c_k, $j^\star$ adaptive	$1/T_{\text{pos}}$

computed with the unbiased estimator of Chen et al. (2021), and $\text{corr}/s = s \cdot \text{Pass@1}/100$, the mean number of correct samples per problem out of s. Notation is summarized in Section A.

4 Unmasking order as maximum-entropy sampling

4.1 The objective

Fix a denoising step. A sampler q over $\mathcal{M}_t$ has two goals. It should explore by spreading its choice over several positions, so that the order is not fixed by the candidate tokens alone. It should also avoid commitments that damage what the model can still produce. Write $D(k) \in [0, \infty]$ for the *damage* of committing at position k. A sampler should then be as random as possible, measured by its entropy $\mathbb{H}(q)$, subject to a budget γ on the expected damage:

$$\max_{q \in \Delta(\mathcal{M}_t)} \mathbb{H}(q) \qquad \text{s.t.} \qquad \mathbb{E}_{k \sim q}\big[D(k)\big] \;\le\; \gamma. \tag{1}$$

Proposition 1 (Maximum entropy under a damage budget; Jaynes, 1957)**.** *Let $F = \{k : D(k) < \infty\}$ be non-empty, let $\bar{D}$ be the mean of D over F, and let $\gamma > \min_k D(k)$. Then Equation (1) has a unique solution, $q_\lambda(k) = e^{-\lambda D(k)} \big/ \sum_{j \in \mathcal{M}_t} e^{-\lambda D(j)}$ with $e^{-\lambda \cdot \infty} = 0$ for every $\lambda \ge 0$. The multiplier is $\lambda = 0$ if $\gamma \ge \bar{D}$, and otherwise the unique $\lambda > 0$ with $\mathbb{E}_{q_\lambda}[D] = \gamma$. Moreover, for every $\lambda \ge 0$, q_λ is the unique maximizer of $\mathbb{H}(q) - \lambda\, \mathbb{E}_q[D]$ over $\Delta(F)$, because $\mathbb{H}(q) - \lambda\, \mathbb{E}_q[D] = \log Z_\lambda - \mathrm{KL}(q \,\|\, q_\lambda)$ with $Z_\lambda = \sum_{j \in F} e^{-\lambda D(j)}$.*

This is the standard maximum-entropy result applied to positions; we call distributions of this form the *Gibbs family*. The penalized form holds at every step with no budget to track: a fixed λ is a fixed price of damage, and the budget it enforces at a step is $\gamma_t = \mathbb{E}_{q_\lambda}[D]$. Confidence order is the limit $\gamma \downarrow \min_k D(k)$, that is $\lambda \to \infty$, and uniform order is $\lambda = 0$. Any full-support distribution has this form for some D, so the form constrains a rule only through what it requires of D. We require the odds between two eligible positions to depend only on those two commitments, through a fixed function, and let the eligible set depend on the whole step. A rule is then a damage for the odds and an eligible set (Table 1): the first part excludes temperatures that change from step to step (Section 5.2), and the second part contains PACE's cutoff. Section 4.2 derives the damage.

4.2 The damage function

Damage is defined per commitment. A step writes a particular token at a particular position, so its damage is a property of that *committed pair*. This rules out every summary of p_k on its own: its

Algorithm 1 PACE, one denoising step ($\varepsilon = 10^{-12}$).

input: logits at $\mathcal{M}_t$; T_{tok}, T_{pos}, α
$x_k^0 \leftarrow$ Gumbel-max(logits$_k$, T_{tok}); $c_k \leftarrow p_k(x_k^0)$
$w_k \leftarrow c_k^{1/T_{\text{pos}}}$; $h \leftarrow \alpha\,\mathbb{H}\big(w/\sum_i w_i\big)$
order $\mathcal{M}_t$ by c descending: $k_1, \ldots, k_n$
$W \leftarrow w_{k_1}$; $H \leftarrow 0$; $j \leftarrow 1$
while $j < n$ **do**
  $\theta \leftarrow W/(W + w_{k_{j+1}})$; $H' \leftarrow \theta H + H_{\text{b}}(\theta)$
  if $H' > h + \varepsilon$ **then break**
  $W \leftarrow W + w_{k_{j+1}}$; $H \leftarrow H'$; $j \leftarrow j + 1$
$\tau_t \leftarrow c_{k_j}$ (adaptive cutoff)
commit x_k^0 at one position drawn from w on $\{k_1, \ldots, k_j\}$

maximum, the margin between its top two entries (Kim et al., 2025), its Shannon or Rényi entropy, and its varentropy. Two positions with identical p_k can offer commitments that differ greatly in quality, depending on which candidate token was drawn. We therefore let the damage depend on the committed pair through the probability the model assigns to it, $D(k) = \varphi(c_k)$.

Summed damage is the output's negative log-likelihood. Step t commits the candidate at one position k_t, and c_{k_t} is the probability the model assigns to that token given the current canvas. Over a generation of N steps with order $\Pi = (k_1, \ldots, k_N)$, the product of the confidences is the probability the model's conditional distributions assign to the output y along Π, $p_\theta(y \mid \Pi) = \prod_t p_{k_t}(y_{k_t} \mid \text{canvas at } t)$, which for a masked diffusion model depends on Π:

$$\prod_{t=1}^{N} c_{k_t} \;=\; p_\theta(y \mid \Pi), \qquad \sum_{t=1}^{N} -\log c_{k_t} \;=\; -\log p_\theta(y \mid \Pi). \tag{2}$$

Two successive commitments with confidences a and b thus have joint probability ab under the model, and scoring them as one composite commitment or as two should give the same damage, $\varphi(ab) = \varphi(a) + \varphi(b)$. With $\varphi \geq 0$, this fixes φ up to a non-negative factor.

Lemma 2 (The damage of a commitment is its surprisal)**.** *Let $\varphi : (0, 1] \to [0, \infty)$ satisfy $\varphi(ab) = \varphi(a) + \varphi(b)$ for all $a, b \in (0, 1]$. Then $\varphi(c) = -\kappa \log c$ for some $\kappa \geq 0$.*

This is the classical characterization of self-information; the proof reduces it to Cauchy's functional equation (Section B). Only the product λD enters Proposition 1, so κ is absorbed into λ and we write

$$D(k) \;=\; -\log p_k(x_k^0) \;=\; -\log c_k. \tag{3}$$

Corollary 3 (Tempered selection)**.** *With D as in Equation (3) on every masked position and $\min_k D(k) < \gamma < \bar{D}$, the solution of Equation (1) is $q_\lambda(k) \propto c_k^\lambda$, tempered selection at $T_{\text{pos}} = 1/\lambda$, with $\lambda > 0$ given by Proposition 1.*

Tempered selection is thus the Gibbs rule with the surprisal damage and no deferral. By Equation (2), its multiplier weighs each step's contribution to the output's negative log-likelihood against the entropy of the step's choice (Section B.4 relates the damage budget to a cross-entropy). Tempered selection never defers: every masked position keeps positive probability at every step. With the damage fixed, two degrees of freedom remain within the Gibbs family: the multiplier, which a temperature schedule may vary by step (Section 5.2), and deferral, setting $D = +\infty$ on some positions. Section 4.3 determines which positions to defer and how many.

4.3 Deferral under an exploration budget

Deferral. Write $p_{\text{pos}}(k) = c_k^{1/T_{\text{pos}}}/Z$ for tempered selection, which we call the *reference distribution*. A *deferral rule* keeps the surprisal on an eligible set $S \subseteq \mathcal{M}_t$ and sets the damage to $+\infty$ off it,

$$D_S(k) \;=\; \begin{cases} -\log c_k & k \in S, \\ +\infty & \text{otherwise.} \end{cases} \tag{4}$$

By Proposition 1 at $\lambda = 1/T_{\text{pos}}$ the rule samples from q_S, the restriction of p_{pos} to S, renormalized. Whenever two positions are both eligible, their odds are $(c_i/c_j)^{1/T_{\text{pos}}}$, the odds of tempered selection, whatever the other positions are: the rule is Luce's choice model with the eligible set as the choice set (Luce, 1959). A position with $D = +\infty$ at a step is *deferred*: it cannot be committed at that step and

stays masked for later ones. We require S to be closed upward in the confidence order, ties broken by a fixed order, so that a position is deferred only if every position after it in that order is deferred too. With positions ordered $c_{k_1} \geq \cdots \geq c_{k_n}$, the eligible set is then a prefix $S_j = \{k_1, \ldots, k_j\}$, and choosing it is choosing a cutoff $\tau = c_{k_j}$.

The exploration budget. The cutoff is set by a budget on how spread out the choice may be. We require the sampler's entropy to be at most a fixed fraction $\alpha \in (0, 1)$ of the reference entropy,

$$\mathbb{H}(q) \;\leq\; \alpha\, \mathbb{H}(p_{\mathrm{pos}}), \qquad \text{equivalently} \qquad N_{\mathrm{eff}}(q) \;\leq\; N_{\mathrm{eff}}(p_{\mathrm{pos}})^{\alpha}, \tag{5}$$

where $N_{\mathrm{eff}}(p) = e^{\mathbb{H}(p)}$ is the effective number of positions p spreads over. The budget is relative because the number of masked positions runs $B, B-1, \ldots, 1$ within every block: a fixed budget keeps every masked position once few remain, while the relative one keeps $\lfloor n^{\alpha} \rfloor$ of n equally confident positions, ties broken by a fixed order, at every n (Proposition 10).

Maximum entropy within the budget. Following the principle of Equation (1), we take the most exploratory deferral rule that meets the budget:

$$\max_{1 \leq j \leq n} \; \mathbb{H}\big(q_{S_j}\big) \quad \text{s.t.} \quad \mathbb{H}\big(q_{S_j}\big) \leq \alpha\, \mathbb{H}(p_{\mathrm{pos}}). \tag{6}$$

Because the entropy of a prefix grows with its length (Proposition 4(i)), this is the longest prefix that meets the budget.

Proposition 4 (Characterization of PACE). *Fix $T_{\mathrm{pos}} > 0$ and $\alpha \in (0, 1)$, let $n = |\mathcal{M}_t| \geq 2$ with every $c_k > 0$, and order the positions by confidence, ties broken by a fixed order, $c_{k_1} \geq \cdots \geq c_{k_n}$, with $S_j = \{k_1, \ldots, k_j\}$.*

- *(i) $\mathbb{H}(q_{S_j})$ is strictly increasing in j, so the eligible sets that meet Equation (5) are exactly $S_1, \ldots, S_{j^\star}$ for some $j^\star$.*
- *(ii) The budget always binds: $1 \leq j^\star < n$, so tempered selection, the rule with $S = \mathcal{M}_t$, violates it.*
- *(iii) Equation (6) has the unique solution $S_{j^\star}$, with cutoff $\tau^\star = c_{k_{j^\star}}$. Among the prefixes that meet the budget, the same set is also the unique one that maximizes the retained reference mass $p_{\mathrm{pos}}(S)$, that defers the fewest positions, and that minimizes every f-divergence from q_S to p_{pos} with f strictly convex and $f(0) < \infty$, such as KL and Jensen–Shannon.*
- *(iv) Algorithm 1 returns $S_{j^\star}$ with one sort and one scan, in $O(n \log n)$ time, up to the floating-point tolerance ε.*

We call the rule of Proposition 4 PACE. Its damage is D_S with $S = S_{j^\star}$: the surprisal, with a confidence cutoff $\tau_t^\star$ that adapts to each step (Figure 2 works through one step). Among prefixes, entropy, retained mass, number of deferrals and divergence from the reference select the same set. Upward closure is needed: without it the mass-maximizing set under the budget can defer the second most confident position while keeping the least confident ones (Remark 14). Algorithm 1 evaluates each prefix in $O(1)$ with the grouping identity $\mathbb{H}(q_{S_{j+1}}) = \theta\, \mathbb{H}(q_{S_j}) + H_{\mathrm{b}}(\theta)$, where $\theta = p_{\mathrm{pos}}(S_j)/p_{\mathrm{pos}}(S_{j+1})$ and H_{b} is the binary entropy. PACE adds no forward passes, so every comparison below is at the same number of forward passes.

5 Theory

This section first shows what position selection does to the committed token, then compares deferral with a temperature that changes at every step. Write $\pi_1 \geq \pi_2 \geq \cdots \geq \pi_n$ for the probabilities of p_{pos} in decreasing order.

5.1 Position selection sharpens the committed token's distribution

Theorem 5 (Position selection sharpens the committed token's distribution). *Fix a step and a masked position k whose candidate $x_k^0 \sim r_k = \mathrm{softmax}(\ell_k / T_{\mathrm{tok}})$ is drawn independently of the other candidates. Let a rule select k with probability $w(c_k)/(w(c_k) + R)$, where $w : (0, 1] \to (0, \infty)$ is non-decreasing and $R \geq 0$ depends only on the other candidates, and let $\psi(u) = \mathbb{E}[u/(u + R)]$.*

(i) Given that k is selected, the committed token has law $\tilde{r}_k(v) \propto r_k(v)\,\psi\big(w(p_k(v))\big)$.

(ii) For any tokens with $p_k(v) \geq p_k(v')$,

$$\frac{r_k(v)}{r_k(v')} \;\leq\; \frac{\tilde{r}_k(v)}{\tilde{r}_k(v')} \;\leq\; \frac{r_k(v)\,w(p_k(v))}{r_k(v')\,w(p_k(v'))}.$$

(iii) For tempered selection, $w(c) = c^{1/T_{\text{pos}}}$, the upper bound is the odds of $\mathrm{softmax}(\ell_k/T_{\text{eff}})$ with $1/T_{\text{eff}} = 1/T_{\text{tok}} + 1/T_{\text{pos}}$. Conditional on the other candidates, the committed token has law $\propto r_k(v)\,u_v/(u_v + R)$ with $u_v = w(p_k(v))$; if $R \geq M\,w(\max_v p_k(v))$ for some $M > 0$, its odds for tokens ordered as in (ii) are at least $M/(M+1)$ times those of $\mathrm{softmax}(\ell_k/T_{\text{eff}})$.

(iv) If the probability of selecting k does not depend on x_k^0, then $\tilde{r}_k = r_k$.

Under tempered selection, the written token's distribution is at least as sharp as at T_{tok}, and close to that at T_{eff} when the other positions' total weight is large relative to k's. Olausson et al. (2026) describe the position temperature as a parameter complementary to the token temperature, with T_{tok} governing what is sampled at each position and T_{pos} the order. Because the position score is the probability of the drawn token, T_{pos} also shapes which tokens are committed. This is consistent with their Figure 9 (LLaDA-8B-Instruct, HumanEval), where autoregressive order, which ignores the draw and so writes at T_{tok} by (iv), degrades severely for $T_{\text{tok}} > 1$ while tempered selection does not. At their T_{pos}=1, the temperature of the sharpest law the theorem allows moves only from $T_{\text{eff}} = 0.23$ to 0.60 as T_{tok} moves from 0.3 to 1.5. The theorem does not cover PACE, whose eligible set depends on c_k jointly with the other candidates; PACE adds a cutoff below which a candidate is redrawn rather than written (Remark 13).

5.2 Deferral versus an adaptive temperature

A position temperature can also meet the budget of Equation (5), by changing at every step. When it can, the distribution closest to p_{pos} in KL that meets the budget is $q \propto p_{\text{pos}}^{\beta_t}$ with $\beta_t > 1$ (Proposition 11, by Lagrangian duality); we call this rule *adaptive-β selection*. It is closer to p_{pos} in KL than any deferral rule that meets the budget, but it loses the reference odds.

Proposition 6 (Suppressing the tail with a temperature distorts the head)**.** *Fix $1 \leq m < n$ with $\pi_1 > \pi_2$ and $\pi_m > \pi_{m+1}$. For a distribution q let $\Lambda(q) = q_{(1)}/q_{(2)}$ be the odds between its two largest masses and $\Theta_m(q) = \sum_{j>m} q_{(j)}$ its mass outside the top m, where $q_{(j)}$ is the j-th largest mass. Along $q^\beta \propto p_{\text{pos}}^\beta$, $\beta > 0$:*

(i) $\Lambda(q^\beta) = (\pi_1/\pi_2)^\beta$ is strictly increasing and $\Theta_m(q^\beta)$ is strictly decreasing in β;

(ii) for $\epsilon \in (0, 1/n)$, the smallest head odds at which a temperature reaches $\Theta_m \leq \epsilon$, $\Lambda^\star(\epsilon) = \min\{\Lambda(q^\beta) : \Theta_m(q^\beta) \leq \epsilon\}$, satisfies

$$(n\epsilon)^{-\rho} \;\leq\; \Lambda^\star(\epsilon) \;\leq\; \Big(\frac{n-m}{\epsilon}\Big)^{\rho}, \qquad \rho = \frac{\log(\pi_1/\pi_2)}{\log(\pi_1/\pi_{m+1})} \in (0, 1],$$

so $\Lambda^\star(\epsilon) = \epsilon^{-\rho+o(1)}$ as $\epsilon \to 0$;

(iii) for $m \geq 2$, the deferral rule with $S = S_m$ reaches $\Theta_m = 0$ with $\Lambda = \pi_1/\pi_2$, the reference odds.

A temperature that pushes the mass outside the leading m positions below ϵ must raise the odds between the two leading positions to a value that grows as $\epsilon^{-\rho}$, with an exponent that is exact as $\epsilon \to 0$ at fixed n. Deferral removes the tail at the reference odds. Adaptive-β selection distorts the odds at every step at which the budget binds: it multiplies those of any two positions with $c_i > c_j$ by $(c_i/c_j)^{(\beta_t - 1)/T_{\text{pos}}} > 1$. It also cannot meet the budget at steps where the ν positions tied for the largest reference probability have $\log \nu \geq \alpha \mathbb{H}(p_{\text{pos}})$.

5.3 PACE COMMITS MORE CONFIDENTLY THAN TEMPERED SELECTION

Proposition 7 (Stochastic dominance of the committed confidence)**.** *Fix a state, its drawn candidates and T_{pos}, and let $c^{(\alpha)}$ be the confidence of the position PACE commits at budget α, with $\alpha \geq 1$ meaning tempered selection. For every ξ, $\mathbb{P}\big(c^{(\alpha)} < \xi\big)$ is non-decreasing in α. For $\alpha < 1$ it is strictly smaller than under tempered selection whenever $\min_k c_k < \xi \leq \max_k c_k$. Equivalently, the expectation of every non-increasing function of $c^{(\alpha)}$, such as the surprisal $-\log c$ or a calibrated error probability, is non-decreasing in α.*

At every step, the confidence of what PACE writes first-order stochastically dominates that of tempered selection and is dominated by that of confidence order, the limit $\alpha \to 0$. Table 5 measures the difference on real trajectories, and Section D.4 measures what the extra low-confidence commitments cost.

6 EXPERIMENTAL SETUP

Models. LLaDA-8B-Instruct (Nie et al., 2025) and Dream-v0-Instruct-7B (Ye et al., 2025), the two open masked dLLMs used by prior work on decoding order, run in `bfloat16` on a single A6000 GPU (details in Section E).

Benchmarks. GSM8K (Cobbe et al., 2021), MATH-500 (Hendrycks et al., 2021; Lightman et al., 2024), HumanEval (Chen et al., 2021) and MBPP (Austin et al., 2021). We call each of the eight model–benchmark pairs a *cell*. Code must pass every assertion in the reference tests (10 s timeout, no partial credit). Math uses exact match after light normalization, with a shared answer extractor.

Protocol. We use generation length 256, block length 32 and one position per step (K=1). Every method therefore uses the same 256 forward passes per sample. The main protocol uses $T_{\text{tok}} = 0.8$, 100 problems and 32 samples per problem (the first 100 of HumanEval and of the MBPP test split, a random subset with seed 0 of GSM8K and MATH-500). A replication uses $T_{\text{tok}} = 0.6$, 8 samples and 164 to 200 problems. Table 2 lists every run. PACE uses $T_{\text{pos}} = 0.5$ and $\alpha = 0.5$ throughout, fixed once on 60 HumanEval problems and never re-tuned.

Baselines. **Confidence order** (AO in Figure 3 and tables; LLaDA's default), **tempered selection** (TLC) at $T_{\text{pos}} = 0.5$, and **autoregressive (AR) order**, which unmasks left to right within the same model and block schedule. Confidence thresholding (Wu et al., 2026) and its tempered form (Olausson et al., 2026) commit a variable number of positions per step, so we compare them separately (Section D.6).

Statistics. Every interval we report is a paired bootstrap over problems within a single run (10^4 resamples), and $*$ marks a 95% interval that excludes zero. All methods in a table are run together on one problem set, so that pairing removes problem difficulty, the largest source of variance.

7 RESULTS

7.1 MAIN RESULT

PACE gains $+4.1$ to $+9.1$ points of single-sample accuracy over confidence order and $+10.2$ to $+28.1$ over tempered selection. In Figure 3 and Table 6 PACE leads at every k on every cell. The gain over confidence order persists as k grows: at k=32 it is $+2.0$ to $+13.0$ points, and at least as large as at k=1 on five of eight cells (Table 7). On the six cells outside HumanEval, which are out of sample for α and T_{pos}, the k=1 gains are $+4.1$ to $+7.3$ and $+10.2$ to $+24.5$.

Replication at a second protocol. At $T_{\text{tok}} = 0.6$, with 8 samples and 164 to 200 problems, the Pass@1 gain over tempered selection is $+6.10\,[+3.1, +9.2]^*$ (LLaDA/HumanEval, 164), $+3.56\,[+1.3, +5.8]^*$ (LLaDA/GSM8K, 200), $+6.12\,[+3.9, +8.4]^*$ (LLaDA/MBPP, 200), $+11.59\,[+8.2, +15.2]^*$ (Dream/HumanEval, 164) and $+8.88\,[+6.1, +11.6]^*$ (Dream/MBPP, 200). In a separate LLaDA run with 60 problems and 64 samples (Table 8), PACE is within 0.9 points of

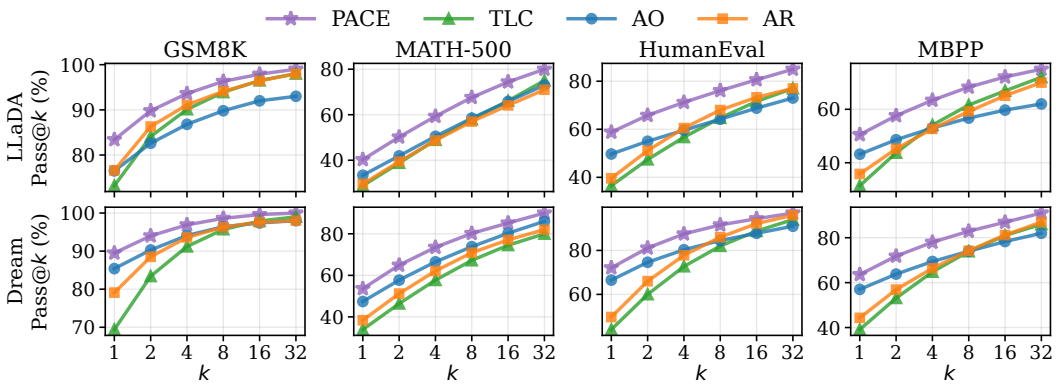

Figure 3: Pass@k at $T_{\text{tok}} = 0.8$ (100 problems, 32 samples) for LLaDA (top) and Dream (bottom); values in Table 6. Labels as in Section 6.

confidence order at k=1 and ahead of it at every larger k; tempered selection and AR order overtake PACE by k=8.

7.2 Varying the budget

On LLaDA/MATH-500 (150 problems, 8 samples), not used to set α, every budget from α=0.1 to 0.8 gives 34.6 to 40.1 Pass@1, above both limits: 30.4 with an inactive budget (α=1.05, tempered selection at the same T_{pos}) and 33.7 for confidence order ($\alpha \to 0$). The peak is at α=0.5, the value set on HumanEval (Figure 4).

7.3 Mechanism

Fewer low-confidence commitments. When both rules are evaluated on the same trajectories, tempered selection makes 10.5 expected commitments per generation at positions with $c < 0.5$, and PACE makes 2.3. The ratio is 4.1 to 5.0 at three thresholds (Table 5), in the direction Proposition 7 guarantees. At $T_{\text{tok}} = 0.8$ the extra commitments do not improve coverage: on LLaDA/HumanEval tempered selection reaches Pass@32 of 77.0 against 85.0 for PACE.

Low-confidence commitments are costly. Forcing low-confidence commitments into otherwise confidence-ordered decoding lowers Pass@1 monotonically, from 57.5 with none to 47.5, 40.0 and 37.5 with 1.8, 3.8 and 6.0 commitments achieved (Table 3); the drop is significant from 3.8 onward. Both findings concern the low-confidence tail that PACE removes; they do not support committing only the most confident position. Committing only that position, as confidence order does, commits final answers before their reasoning (Yeom et al., 2026).

The score must depend on the drawn token. Scoring positions in PACE by a statistic of p_k alone, with the committed token held fixed, lowers Pass@1 by 5.8 to 10.0 points (three of four significant). At $T_{\text{tok}} = 1.6$, the accuracy of scoring by the maximum probability drops to 0.0 against 37.3, consistent with Theorem 5(iv) (Section D.5).

8 Conclusion

We presented PACE, which keeps the odds of tempered selection and defers the least confident positions under an entropy budget relative to tempered selection. It improves single-sample accuracy over confidence order and tempered selection on every cell of the main protocol at equal compute. PACE follows from a maximum-entropy view of unmasking order with two parts: the odds among eligible positions, set by the surprisal, whose sum over a generation is the output's negative log-likelihood, and the eligible set. The same view shows that, because candidates are drawn before positions are chosen, selecting positions by the tempered confidence of the drawn tokens also sharpens the distribution of what is written, which links the token and position temperatures.

ETHICS STATEMENT

The method changes decoding only, adds no trained parameters and no forward passes, and removes no output from what the model can generate. We see no application-specific risk beyond those already present in the underlying models, and all benchmarks used are public.

AI USE STATEMENT

We used generative AI tools to help formulate several mathematical statements, to draft sections of the paper, to edit the manuscript for clarity, grammar, organization and LaTeX presentation, to find related work, to review analysis code, and to give feedback on the research methodology and experiments. The authors reviewed the AI-assisted revisions, checked every statement and proof, and verified the code. The research questions, experiments and conclusions are the authors'. We take full responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Algorithm 1 specifies PACE completely. Section E gives the token sampler, the Dream logit shift and the commit convention, Table 2 lists every run with its model, token temperature, and problem and sample counts, and Section B contains all proofs. The implementation and reproduction instructions will be released upon acceptance.

REFERENCES

János Aczél. *Lectures on Functional Equations and Their Applications*. Academic Press, 1966.

Jacob Austin, Augustus Odena, Maxwell Nye, et al. Program synthesis with large language models. *arXiv preprint arXiv:2108.07732*, 2021.

Erfan Baghaei Potraghloo, Seyedarmin Azizi, Souvik Kundu, and Massoud Pedram. Top-H decoding: Adapting the creativity and coherence with bounded entropy in text generation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. arXiv:2509.02510.

Sourya Basu, Govardana Sachithanandam Ramachandran, Nitish Shirish Keskar, and Lav R. Varshney. Mirostat: A neural text decoding algorithm that directly controls perplexity. In *International Conference on Learning Representations (ICLR)*, 2021.

Heli Ben-Hamu, Itai Gat, Daniel Severo, Niklas Nolte, and Brian Karrer. Accelerated sampling from masked diffusion models via entropy bounded unmasking. *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Huiwen Chang, Han Zhang, Lu Jiang, Ce Liu, and William T. Freeman. MaskGIT: Masked generative image transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11315–11325, 2022.

Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, et al. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*, 2021.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, et al. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

Liancheng Fang, Aiwei Liu, Henry Peng Zou, Yankai Chen, Enze Ma, Leyi Pan, Chunyu Miao, Wei-Chieh Huang, Xue Liu, and Philip S. Yu. Locally confident, globally stuck: The quality–exploration dilemma in diffusion language models. In *Conference on Language Modeling (COLM)*, 2026.

Matthew Finlayson, John Hewitt, Alexander Koller, Swabha Swayamdipta, and Ashish Sabharwal. Closing the curious case of neural text degeneration. In *International Conference on Learning Representations (ICLR)*, 2024.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the MATH dataset. In *NeurIPS Datasets and Benchmarks Track*, 2021.

John Hewitt, Christopher D. Manning, and Percy Liang. Truncation sampling as language model desmoothing. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pp. 3414–3427, 2022.

Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. The curious case of neural text degeneration. In *International Conference on Learning Representations (ICLR)*, 2020.

Edwin T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–630, 1957.

M. Jazbec, T. X. Olausson, L. Béthune, P. Ablin, M. Kirchhof, J. Monteiro, V. Turrisi, J. Ramapuram, and M. Cuturi. Learning unmasking policies for diffusion language models. In *International Conference on Machine Learning (ICML)*, 2026.

Jaeyeon Kim, Kulin Shah, Vasilis Kontonis, Sham M. Kakade, and Sitan Chen. Train for the worst, plan for the best: Understanding token ordering in masked diffusions. In *International Conference on Machine Learning (ICML)*, 2025.

Hunter Lightman, Vineet Kosaraju, Yura Burda, Harri Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. In *International Conference on Learning Representations (ICLR)*, 2024.

Sulin Liu, Juno Nam, Andrew Campbell, Hannes Stärk, Yilun Xu, Tommi Jaakkola, and Rafael Gómez-Bombarelli. Think while you generate: Discrete diffusion with planned denoising. In *International Conference on Learning Representations (ICLR)*, 2025.

R. Duncan Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, 1959.

Minh Nhat Nguyen, Andrew Baker, Clement Neo, Allen Roush, Andreas Kirsch, and Ravid Shwartz-Ziv. Turning up the heat: Min-p sampling for creative and coherent LLM outputs. In *International Conference on Learning Representations (ICLR)*, 2025.

Zanlin Ni, Shenzhi Wang, Yang Yue, Tianyu Yu, Weilin Zhao, Yeguo Hua, Tianyi Chen, Jun Song, Cheng Yu, Bo Zheng, and Gao Huang. The flexibility trap: Rethinking the value of arbitrary order in diffusion language models. In *International Conference on Machine Learning (ICML)*, 2026. arXiv:2601.15165.

Shen Nie, Fengqi Zhu, Zebin You, Xiaolu Zhang, Jingyang Ou, Jun Hu, Jun Zhou, Yankai Lin, Ji-Rong Wen, and Chongxuan Li. Large language diffusion models. *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Theo X. Olausson, Metod Jazbec, Xi Wang, Armando Solar-Lezama, Christian A. Naesseth, Stephan Mandt, and Eric Nalisnick. A tale of two temperatures: Simple, efficient, and diverse sampling from diffusion language models. In *Conference on Language Modeling (COLM)*, 2026.

Fred Zhangzhi Peng, Zachary Bezemek, Sawan Patel, Jarrid Rector-Brooks, Sherwood Yao, Avishek Joey Bose, Alexander Tong, and Pranam Chatterjee. Path planning for masked diffusion model sampling. *arXiv preprint arXiv:2502.03540*, 2025.

Chengcheng Wang, Tingzhang Luo, Wenhao Li, Jianyuan Guo, and Chang Xu. TACG: Trajectory-aware commit gating for diffusion language model decoding. *arXiv preprint arXiv:2607.03236*, 2026.

Chengyue Wu, Hao Zhang, Shuchen Xue, Zhijian Liu, Shizhe Diao, Ligeng Zhu, Ping Luo, Song Han, and Enze Xie. Fast-dLLM: Training-free acceleration of diffusion LLM by enabling KV cache and parallel decoding. *International Conference on Learning Representations (ICLR)*, 2026.

Kaisen Yang, Jayden Teoh, Kaicheng Yang, Yitong Zhang, and Alex Lamb. Improving sampling for masked diffusion models via information gain. In *International Conference on Machine Learning (ICML)*, 2026.

Jiacheng Ye, Zhihui Xie, Lin Zheng, Jiahui Gao, Zirui Wu, Xin Jiang, Zhenguo Li, and Lingpeng Kong. Dream 7B: Diffusion large language models. *arXiv preprint arXiv:2508.15487*, 2025.

Jewon Yeom, Jaewon Sok, Seonghyeon Park, Jeongjae Park, Hwiyeong Lee, and Taesup Kim. Answer first, reason later: When commitment order costs accuracy in diffusion language models. *arXiv preprint arXiv:2608.05687*, 2026.

A NOTATION

Symbol	Meaning
V, L, B	vocabulary, generation length, block length
$\mathcal{M}_t$	masked positions at denoising step t
K	positions committed per step
$p_k \in \Delta(V)$	model distribution at position k
x_k^0	candidate token drawn at k before position selection
$c_k = p_k(x_k^0)$	per-position confidence
T_{tok}, T_{pos}	token and position temperature
T_{eff}	$1/T_{\text{eff}} = 1/T_{\text{tok}} + 1/T_{\text{pos}}$ (Theorem 5)
Π, N	decoding order $(k_1, \dots, k_N)$ of a generation of N steps
p_{pos}	reference distribution over positions, $\propto c^{1/T_{\text{pos}}}$
q	sampler's distribution over positions
$D(k)$	damage of committing at k
φ	damage as a function of confidence
λ	multiplier of the damage budget; $\lambda = 1/T_{\text{pos}}$
γ	damage budget
S, Γ_S	eligible set and its reference mass
α	exploration budget, $\mathbb{H}(q_S) \le \alpha \mathbb{H}(p_{\text{pos}})$
τ_t	adaptive confidence cutoff
$N_{\text{eff}}(p) = e^{\mathbb{H}(p)}$	effective number of positions
$\pi_1 \ge \cdots \ge \pi_n$	masses of p_{pos} in descending order, $n = \lvert\mathcal{M}_t\rvert$
$H_{\text{b}}(\theta)$	binary entropy
$\Lambda(q)$, $\Theta_m(q)$	head odds $q_{(1)}/q_{(2)}$; mass outside the top m
ρ	exponent of Proposition 6
β	sharpening exponent on p_{pos}; β=1 is tempered selection

B PROOFS

B.1 PROOFS FOR SECTION 4

Proof of Proposition 1. A feasible q puts no mass off F, so we work on $\Delta(F)$. For $q_\lambda \propto e^{-\lambda D}$ on F, $\mathrm{KL}(q\|q_\lambda) = -\mathbb{H}(q) + \lambda \mathbb{E}_q[D] + \log Z_\lambda$, which is the identity in the statement; since $\mathrm{KL}(q\|q_\lambda) \ge 0$ with equality only at $q = q_\lambda$, q_λ uniquely maximizes the penalized objective. Let $g(\lambda) = \mathbb{E}_{q_\lambda}[D]$. It is continuous, $g(0) = \bar{D}$, $g(\lambda) \to \min_k D(k)$ as $\lambda \to \infty$, and $g'(\lambda) = -\operatorname{Var}_{q_\lambda}[D]$, which is negative unless D is constant on F. If $\gamma \ge \bar{D}$, take $\lambda^\star = 0$; otherwise $\min_k D(k) < \gamma < \bar{D}$ forces D to be non-constant on F, and there is a unique $\lambda^\star > 0$ with $g(\lambda^\star) = \gamma$. In both cases $\lambda^\star(g(\lambda^\star) - \gamma) = 0$. For any feasible q,

$$\mathbb{H}(q) \le \mathbb{H}(q) - \lambda^\star\big(\mathbb{E}_q[D] - \gamma\big) = \log Z_{\lambda^\star} + \lambda^\star\gamma - \mathrm{KL}(q\|q_{\lambda^\star}) \le \log Z_{\lambda^\star} + \lambda^\star\gamma = \mathbb{H}(q_{\lambda^\star}),$$

where the first step uses $\lambda^\star \ge 0$ and $\mathbb{E}_q[D] \le \gamma$, and the last uses $\lambda^\star(g(\lambda^\star) - \gamma) = 0$. Equality throughout requires $\mathrm{KL}(q\|q_{\lambda^\star}) = 0$, so $q_{\lambda^\star}$ is the unique solution. □

Proof of Lemma 2. Let $f(x) = \varphi(e^{-x})$ for $x \ge 0$. Then $f(x+y) = f(x) + f(y)$, which is Cauchy's functional equation on $[0, \infty)$ (Aczél, 1966). It gives $f(0) = 0$ and, by induction, $f(rx) = rf(x)$ for rational $r \ge 0$. Since $f \ge 0$, $f(x+y) = f(x) + f(y) \ge f(x)$, so f is non-decreasing. For real $x \ge 0$ and rationals $r_1 \le x \le r_2$, $r_1 f(1) \le f(x) \le r_2 f(1)$, and letting $r_1, r_2 \to x$ gives $f(x) = \kappa x$ with $\kappa = f(1) \ge 0$. Hence $\varphi(c) = f(-\log c) = -\kappa \log c$. □

Lemma 8 (Prefix entropy is strictly increasing; cf. Baghaei Potraghloo et al., 2025)**.** *Let $p_1 \geq p_2 \geq \cdots \geq p_n > 0$ and let q_j be the top-j prefix renormalized. Then $\mathbb{H}(q_{j+1}) > \mathbb{H}(q_j)$ for every $j < n$.*

Proof. Write $C_j = \sum_{i \leq j} p_i$ and $\theta = C_j / C_{j+1} \in (0, 1)$. Splitting the sum over the first j terms and the $(j+1)$-th,

$$\mathbb{H}(q_{j+1}) = \theta\, \mathbb{H}(q_j) + H_b(\theta), \qquad H_b(\theta) = -\theta \log \theta - (1-\theta) \log(1-\theta).$$

So the claim is $H_b(\theta) > (1-\theta)\mathbb{H}(q_j)$. Descending order, with ties allowed, gives $p_{j+1} \leq C_j / j$, hence $1 - \theta = p_{j+1}/C_{j+1} \leq 1/(j+1)$. Therefore $H_b(\theta) \geq (1-\theta) \log \frac{1}{1-\theta} \geq (1-\theta) \log(j+1) > (1-\theta)\mathbb{H}(q_j)$, where the last step holds because $\mathbb{H}(q_j) \leq \log j < \log(j+1)$ and $1 - \theta > 0$. □

Our implementation computes the entropy of every prefix at once from cumulative sums and keeps each position whose prefix meets the budget; it agrees with Algorithm 1 on 2×10^4 random distributions.

Proposition 9 (Divergence of a truncation)**.** *Let D_f be any f-divergence with $f(0) < \infty$. For $\Gamma_S = p_{\text{pos}}(S) > 0$ and $q_S = p_{\text{pos}}|_S / \Gamma_S$,*

$$D_f(q_S \,\|\, p_{\text{pos}}) \;=\; \Gamma_S\, f\big(1/\Gamma_S\big) + (1 - \Gamma_S)\, f(0),$$

a function of Γ_S alone, non-increasing for every convex f and strictly decreasing for every strictly convex f, including JSD, KL, χ^2 and squared Hellinger; total variation, $1 - \Gamma_S$, is also strictly decreasing. Hence, among eligible sets that meet the budget, minimizing any of these divergences is equivalent to maximizing Γ_S. This generalizes the reduction of Jensen–Shannon minimization to mass maximization in Baghaei Potraghloo et al. (2025).

Proof of Proposition 9. With $q_S = p_{\text{pos}}|_S / \Gamma_S$, the ratio $q_S(k)/p_{\text{pos}}(k)$ equals $1/\Gamma_S$ on S and 0 off it, so $D_f(q_S \| p_{\text{pos}}) = \sum_k p_{\text{pos}}(k) f(q_S(k)/p_{\text{pos}}(k)) = \Gamma_S f(1/\Gamma_S) + (1-\Gamma_S) f(0)$. Writing $t = 1/\Gamma$, the derivative in Γ is $f(t) - t f'(t) - f(0)$, which is at most zero because convexity gives $f(0) \geq f(t) - t f'(t)$, with strict inequality for strictly convex f (use a subgradient where f is not differentiable). As examples, for JSD, substituting the mixture $\frac{1}{2}(q_S + p_{\text{pos}})$ gives $\frac{1}{2}\big(2 \log 2 + \Gamma \log \Gamma - (1+\Gamma) \log(1+\Gamma)\big)$, whose derivative $\frac{1}{2} \log(\Gamma/(1+\Gamma))$ is negative; KL gives $-\log \Gamma$, total variation $1 - \Gamma$, χ^2 gives $1/\Gamma - 1$, and squared Hellinger $1 - \sqrt{\Gamma}$, each strictly decreasing. Reverse KL is infinite for every $S \subsetneq \mathcal{M}_t$ and cannot order eligible sets. □

Proof of Proposition 4. (i) is Lemma 8 applied to p_{pos} in the confidence order, ties broken by a fixed order, along which p_{pos} is non-increasing because it is increasing in c; strict monotonicity makes the feasible indices a down-set $\{1, \ldots, j^\star\}$. (ii) $\mathbb{H}(q_{S_1}) = 0 \leq \alpha \mathbb{H}(p_{\text{pos}})$, so $j^\star \geq 1$. For $n \geq 2$, p_{pos} has full support on at least two positions, so $\mathbb{H}(p_{\text{pos}}) > 0$ and $\mathbb{H}(q_{S_n}) = \mathbb{H}(p_{\text{pos}}) > \alpha \mathbb{H}(p_{\text{pos}})$, so $j^\star < n$. (iii) The eligible sets allowed are the prefixes $S_1, \ldots, S_n$. The entropy, the retained mass $p_{\text{pos}}(S_j)$ and the number j of eligible positions are strictly increasing in j, so each is uniquely maximized over the feasible indices at $j^\star$. By Proposition 9, every f-divergence of q_{S_j} from p_{pos} with f strictly convex and $f(0) < \infty$ is a strictly decreasing function of $p_{\text{pos}}(S_j)$, so it is uniquely minimized at $j^\star$ as well. (iv) Algorithm 1 visits $j = 1, 2, \ldots$ in order, computing $\mathbb{H}(q_{S_{j+1}}) = \theta \mathbb{H}(q_{S_j}) + H_{\text{b}}(\theta)$ with $\theta = p_{\text{pos}}(S_j)/p_{\text{pos}}(S_{j+1})$, the grouping property of entropy, and stops at the first j with $\mathbb{H}(q_{S_{j+1}}) > \alpha \mathbb{H}(p_{\text{pos}})$; by (i) this j is $j^\star$. Sorting costs $O(n \log n)$ and the scan $O(n)$. The tolerance ε admits prefixes that meet the budget up to ε, as when n^α is an integer in Proposition 10. □

In bfloat16, confidences can tie, for instance when they round to one; every statement above holds for any fixed tie-breaking order.

Proposition 10 (A relative budget scales with the masked set)**.** *Consider budgets $\mathbb{H}(q) \leq f(\mathbb{H}(p_{\text{pos}}))$ with $f : [0, \infty) \to [0, \infty)$. On n positions with equal confidences, ties broken by a fixed order, the scan of Algorithm 1 with budget f keeps $\min\big(n, \lfloor e^{f(\log n)} \rfloor\big)$ positions: $\lfloor n^\alpha \rfloor$ for $f(h) = \alpha h$ with $\alpha < 1$, and $\lfloor e^b \rfloor$ for a constant budget $f \equiv b$, which keeps every position once $n \leq e^b$.*

Proof. Each prefix of the tie-breaking order is uniform on j positions, with entropy $\log j$, so it meets the budget exactly when $j \leq e^{f(\log n)}$, and by Lemma 8 the scan stops at the largest such j, capped at n. Substituting the two forms of f gives the two cases. □

B.2 Proofs for Section 5

Proof of Theorem 5. (i) Given all candidates, k is selected with probability $w(c_k)/(w(c_k)+R)$, where $c_k = p_k(x_k^0)$. Since R depends only on the other candidates, which are independent of x_k^0, $\mathbb{P}(x_k^0 = v,\ k \text{ selected}) = r_k(v)\,\mathbb{E}\big[w(p_k(v))/(w(p_k(v))+R)\big] = r_k(v)\,\psi\big(w(p_k(v))\big)$, and normalizing over v gives $\tilde{r}_k$. (ii) For $u \geq u' > 0$ and every $R \geq 0$, $u/(u+R) \geq u'/(u'+R)$ and $1/(u+R) \leq 1/(u'+R)$. Taking expectations, ψ is non-decreasing and $\psi(u)/u = \mathbb{E}[1/(u+R)]$ is non-increasing. With $u = w(p_k(v)) \geq u' = w(p_k(v'))$, the ratio $\tilde{r}_k(v)/\tilde{r}_k(v')$ equals $r_k(v)/r_k(v')$ times $\psi(u)/\psi(u') \geq 1$, which is the left inequality, and equals $r_k(v)u/(r_k(v')u')$ times $(\psi(u)/u)/(\psi(u')/u') \leq 1$, which is the right one. (iii) With $w(c) = c^{1/T_{\text{pos}}}$ and $p_k \propto e^{\ell_k}$, $r_k(v)\,w(p_k(v)) \propto e^{\ell_{k,v}/T_{\text{tok}}}\,e^{\ell_{k,v}/T_{\text{pos}}}$, which is $\mathrm{softmax}(\ell_k/T_{\text{eff}})$. Given the other candidates, R is fixed, and the argument of (i) without the expectation gives the conditional law $\propto r_k(v)\,u_v/(u_v+R)$. Its odds between v and v' are those of $\mathrm{softmax}(\ell_k/T_{\text{eff}})$ times $(u_{v'}+R)/(u_v+R)$. If $R \geq M\,w(\max_v p_k(v))$, then $u_v \leq R/M$, so this factor is at least $R/(R+R/M) = M/(M+1)$. (iv) If the selection probability does not depend on x_k^0, the event that k is selected is independent of x_k^0, and conditioning on it leaves the law of x_k^0 unchanged. □

Proposition 11 (The KL projection onto the budget has an adaptive temperature)**.** *Over all distributions, the distribution closest to p_{pos} in $\mathrm{KL}(\cdot \,\|\, p_{\text{pos}})$ that meets Equation (5) is $q \propto p_{\text{pos}}^{\beta}$, with $\beta > 1$ the unique root of $\mathbb{H}(p_{\text{pos}}^{\beta}/Z_\beta) = \alpha\mathbb{H}(p_{\text{pos}})$, whenever $\log\nu < \alpha\mathbb{H}(p_{\text{pos}})$, where ν is the number of positions tied at $\max_k p_{\text{pos}}(k)$. When $\log\nu \geq \alpha\mathbb{H}(p_{\text{pos}})$, no temperature meets the budget.*

Proof of Proposition 11. The Lagrangian of $\mathrm{KL}(q\|p_{\text{pos}}) + \eta\big(\mathbb{H}(q) - \alpha\mathbb{H}(p_{\text{pos}})\big)$ is $(1-\eta)\sum_k q_k \log q_k - \sum_k q_k \log p_{\text{pos}}(k)$ up to constants, convex in q for $\eta < 1$, with stationary point $q \propto p_{\text{pos}}^{1/(1-\eta)}$. Writing $\beta = 1/(1-\eta)$, complementary slackness makes β the root of $\mathbb{H}(p_{\text{pos}}^{\beta}/Z_\beta) = \alpha\mathbb{H}(p_{\text{pos}})$. Since $\frac{d}{d\beta}\mathbb{H}(p_{\text{pos}}^{\beta}/Z_\beta) = -\beta\,\mathrm{Var}_{p_{\text{pos}}^{\beta}/Z_\beta}[\log p_{\text{pos}}]$, the entropy is continuous and strictly decreasing whenever p_{pos} is not uniform on its support, falling from $\mathbb{H}(p_{\text{pos}})$ at $\beta{=}1$ to $\log\nu$ as $\beta \to \infty$. A root therefore exists, and is unique, exactly when $\log\nu < \alpha\mathbb{H}(p_{\text{pos}}) < \mathbb{H}(p_{\text{pos}})$; the second inequality holds for $\alpha < 1$, and a uniform p_{pos} has $\log\nu = \mathbb{H}(p_{\text{pos}})$ and violates the first. When $\log\nu \geq \alpha\mathbb{H}(p_{\text{pos}})$ the same monotonicity keeps the entropy above $\alpha\mathbb{H}(p_{\text{pos}})$ at every finite $\beta > 0$, so no temperature meets the budget. At the root $\beta^\star > 1$, so $\eta^\star = 1 - 1/\beta^\star \in (0,1)$: the multiplier is dual feasible and $L(\cdot,\eta^\star)$ is strictly convex. For any feasible q, weak duality gives $\mathrm{KL}(q\|p_{\text{pos}}) \geq \min_{q'} L(q',\eta^\star) = \mathrm{KL}(q^\star\|p_{\text{pos}})$, so the stationary point is the global optimum although the feasible set is not convex, and it is the unique one because equality in this chain forces q to minimize the strictly convex $L(\cdot,\eta^\star)$. □

Proof of Proposition 6. (i) $q_{(j)}^{\beta} \propto \pi_j^{\beta}$ gives the head odds. For the leak, the odds $\Theta_m/(1-\Theta_m) = \sum_{j>m}\pi_j^{\beta}\big/\sum_{i\leq m}\pi_i^{\beta}$ have log-derivative in β equal to $\mathbb{E}_{\text{tail}}[\log\pi] - \mathbb{E}_{\text{head}}[\log\pi]$ under the β-tilted restrictions to $\{j>m\}$ and $\{i\leq m\}$. Every tail value is at most $\log\pi_{m+1}$ and every head value at least $\log\pi_m > \log\pi_{m+1}$, so the derivative is negative. (ii) From $\pi_{m+1}^{\beta} \leq \sum_{j>m}\pi_j^{\beta} \leq (n-m)\pi_{m+1}^{\beta}$ and $\pi_1^{\beta} \leq \sum_k \pi_k^{\beta} \leq n\pi_1^{\beta}$, $\frac{1}{n}r^{\beta} \leq \Theta_m(q^{\beta}) \leq (n-m)r^{\beta}$ with $r = \pi_{m+1}/\pi_1 < 1$. As $\beta \to 0$ the leak tends to $(n-m)/n \geq 1/n > \epsilon$, and it tends to 0 as $\beta \to \infty$; by (i) there is a unique $\beta_\epsilon > 0$ with $\Theta_m(q^{\beta_\epsilon}) = \epsilon$, the feasible set is $[\beta_\epsilon,\infty)$, and $\Lambda^\star(\epsilon) = (\pi_1/\pi_2)^{\beta_\epsilon}$. The lower bracket at β_ϵ gives $r^{\beta_\epsilon} \leq n\epsilon$, so $\beta_\epsilon \geq \log(1/(n\epsilon))/\log(1/r)$. The upper bracket shows that $\beta_u = \log((n-m)/\epsilon)/\log(1/r)$ is feasible, so $\beta_\epsilon \leq \beta_u$. Substituting both into $(\pi_1/\pi_2)^{\beta} = \exp\big(\beta\log(\pi_1/\pi_2)\big)$ and using $\log(1/r) = \log(\pi_1/\pi_{m+1})$ gives the two bounds. $\rho > 0$ because $\pi_1 > \pi_2$, and $\rho \leq 1$ because $\pi_2 \geq \pi_{m+1}$. Taking logarithms, $\log\Lambda^\star(\epsilon)/\log(1/\epsilon) \to \rho$. (iii) For $m \geq 2$, restricting p_{pos} to its top m positions sets the leak to zero and keeps the ratio of the two largest masses at π_1/π_2. □

Lemma 12 (Nested eligible sets)**.** *Let $S(\alpha)$ be the set returned by Algorithm 1. If $\alpha_1 \leq \alpha_2$ then $S(\alpha_1) \subseteq S(\alpha_2)$.*

Proof. Positions are added in a fixed order (descending p_{pos}) independent of α, so $S(\alpha)$ is always a prefix of that order and α only sets where it stops. Adding the i-th position is permitted exactly

when $\mathbb{H}(q_i) \le \alpha\mathbb{H}(p_{\mathrm{pos}})$, and the threshold increases in α; hence any i permitted at α_1 is permitted at $\alpha_2 \ge \alpha_1$. Both sets being prefixes of the same order gives inclusion. □

Proof of Proposition 7. Let $\mathcal{L} = \{k : c_k < \xi\}$ and $y = p_{\mathrm{pos}}(\mathcal{L})$. The eligible set at budget α is a prefix A of the confidence order, and by Lemma 12 the prefixes are nested in α; tempered selection is the full prefix $\mathcal{M}_t$. For a prefix A, $\mathbb{P}(c^{(\alpha)} < \xi) = q_A(\mathcal{L}) = p_{\mathrm{pos}}(A \cap \mathcal{L})/p_{\mathrm{pos}}(A)$. If $A \cap \mathcal{L} = \emptyset$ this is 0. Otherwise A contains a position with $c < \xi$, hence every position with $c \ge \xi$, so $p_{\mathrm{pos}}(A) = (1 - y) + x_A$ with $x_A = p_{\mathrm{pos}}(A \cap \mathcal{L})$, and $q_A(\mathcal{L}) = x_A/(1 - y + x_A)$, which is increasing in x_A. Both cases give $q_A(\mathcal{L}) \le q_{A'}(\mathcal{L})$ for prefixes $A \subseteq A'$, which is the monotonicity. For the strict claim, $\alpha < 1$ gives $A \subsetneq \mathcal{M}_t$ by Proposition 4(ii), and $\mathcal{M}_t \setminus A \subseteq \mathcal{L}$ is non-empty whenever $A \cap \mathcal{L} \neq \emptyset$. If $A \cap \mathcal{L} = \emptyset$, then $q_A(\mathcal{L}) = 0 < y$ because $\mathcal{L} \neq \emptyset$ for $\xi > \min_k c_k$. Otherwise $x_A < y$, and $x_A/(1-y+x_A) < y$ is equivalent to $x_A(1-y) < y(1-y)$, which holds because $y < 1$ for $\xi \le \max_k c_k$. The equivalence with expectations of non-increasing functions is the standard characterization of first-order stochastic dominance. □

Remark 13 (Exhaustion)**.** Fix $T_{\mathrm{tok}} > 0$ and let every p_k have full support, as a softmax does. For any position rule whose per-step eligible set $S_t \subseteq \mathcal{M}_t$ is non-empty, every position is committed before the block completes, and every string in V^L keeps strictly positive probability. In autoregressive decoding with token-level truncation to an allowed set $A_t \subsetneq V$ at step t, every string whose t-th token lies outside A_t has probability zero.

Proof of Remark 13. Each step commits at least one position and removes it from $\mathcal{M}_t$, which is strictly decreasing and empty at block end. Gumbel-max at $T_{\mathrm{tok}} > 0$ gives every token positive probability at every position. For a target string y, the event that at every step every candidate token equals the corresponding token of y has positive probability, and on it every sequence of selections writes y. The token-axis claim is immediate: the t-th token is drawn from a distribution supported on A_t. □

B.3 Why upward closure is needed

Remark 14 (An example without upward closure)**.** Over all subsets rather than upward-closed ones, the retained-mass maximizer need not be upward closed. On $p_{\mathrm{pos}} = (0.5508, 0.3825, 0.0272, 0.0220, 0.0174)$ at α=0.5 the budget is 0.4743; Algorithm 1 returns $\{1\}$ with $\Gamma = 0.5508$, while the all-subsets maximizer is $\{1, 3, 4, 5\}$ with $\Gamma = 0.6174$. That set discards the second most confident position and keeps the three least confident, which no upward-closed eligible set does. Upward closure in Proposition 4 excludes such sets.

B.4 The damage budget as a cross-entropy

Substituting Equation (3) into Equation (1) relates the damage budget to the cross-entropy between the sampler and the reference distribution over positions. The two are not equal, because the confidences $\{c_k\}$ are token probabilities at distinct positions and do not sum to one, so a normalizing constant enters. Writing $p_{\mathrm{pos}}(k) = c_k^{1/T_{\mathrm{pos}}}/Z$ with $Z = \sum_j c_j^{1/T_{\mathrm{pos}}}$,

$$\mathbb{E}_{k\sim q}[D] \;=\; T_{\mathrm{pos}}\Big(H(q, p_{\mathrm{pos}}) - \log Z\Big), \qquad H(q, p_{\mathrm{pos}}) = \mathbb{H}(q) + \mathrm{KL}(q \,\|\, p_{\mathrm{pos}})\,. \tag{7}$$

The offset $\log Z$ depends on the step but not on q. At a fixed step, minimizing expected damage is therefore exactly minimizing cross-entropy to p_{pos}, and the damage budget penalizes the sampler both for its own entropy and for its departure from the model's preferences. Across steps the correspondence between T_{pos} and the damage budget is monotone but not exact: at every step the enforced budget γ_t is non-decreasing in T_{pos}, but a fixed T_{pos} fixes the multiplier, not the damage budget in nats, because Z varies with the confidences. We use T_{pos} as a multiplier throughout and make no claim that γ is held constant.

C Protocol inventory

The paper reports runs at several problem counts and numbers of samples, chosen by what each measurement needed and what it cost. We list them all here so that the protocol behind every table is

stated, and so that a reader can see which comparisons share a run. Only methods within one row share a run, and every statistical comparison is made within a row.

Table 2: Every run reported in the paper. "Cells" counts model–benchmark pairs. All runs use generation length 256, block length 32, and one position committed per step unless the row says otherwise.

Run	Where	Model	T_{tok}	probs × samples	cells	Arms
Main	Tables 6 and 7, Figure 3	both	0.8	100×32	8	PACE, TLC, AO, AR
Replication	Section 7	both	0.6	164–200×8	5	PACE, TLC
Full k curves	Table 8	LLaDA	0.6	60×64	2	PACE, TLC, AO, AR
Budget sweep	Figure 4	LLaDA	0.8	150×8	1	7 values of α, AO
Adaptive width	Table 4	LLaDA	0.8	60×8	1	CT, TCT (variable positions per step), PACE
Dose–response	Table 3	LLaDA	0.8	40×1	1	doses $0, 2, 5, 10$
Position-only stats	Section H	LLaDA	0.8	60×8	1	surprisal vs. 4 others
Token-temp. probe	Section D.5	LLaDA	1.6	60×8	1	surprisal vs. $\max_v p_k(v)$
Case study	Table 5	LLaDA	0.8	8 gens	1	exact expectations

D Further results

This appendix holds measurements the main text refers to: the budget sweep, further results at the main protocol, curves to k=64 at $T_{\text{tok}} = 0.6$, the cost of commitments at low-confidence positions, which statistic may enter D, and adaptive-width baselines.

D.1 The budget sweep

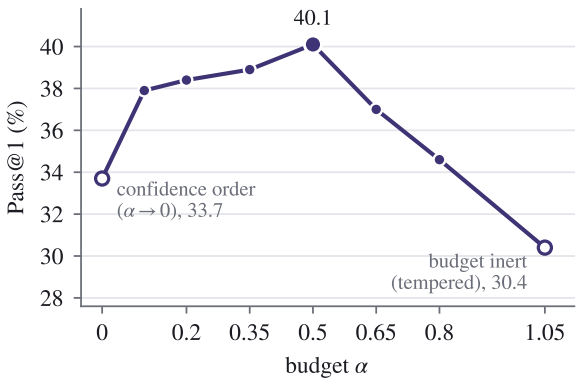

Figure 4: Pass@1 against the budget α on LLaDA/MATH-500 (150 problems × 8 samples, $T_{\text{tok}} = 0.8$, $T_{\text{pos}} = 0.5$). Hollow markers are the two limits: confidence order ($\alpha \to 0$) and the inactive budget ($\alpha = 1.05$), which is tempered selection at the same position temperature. Every binding budget exceeds both limits.

D.2 Further results at the main protocol

Largest gain. The largest gain, +28.1 on Dream/HumanEval, is on the pair Olausson et al. (2026) report as the weakest for tempered selection. On this pair, tempered selection is 22.5 points below confidence order, while PACE is 5.6 points above it.

D.3 CURVES TO k=64 AT A SECOND TOKEN TEMPERATURE

Full curves to k=64 at $T_{\text{tok}} = 0.6$ on two benchmarks, with 60 problems and 64 samples, are given in Table 8. Against tempered selection, PACE has the higher Pass@1 ($+6.28^*$ on HumanEval, $+4.22^*$ on MBPP). At this token temperature the advantage is concentrated at small k: tempered selection keeps every position eligible, the two are equal near k=4, and tempered selection leads beyond it on both benchmarks. At $T_{\text{tok}} = 0.8$ the advantage over tempered selection shrinks with k but stays positive at every $k \leq 32$ in all eight cells (Table 6).

D.4 COST OF THE EXTRA EXPLORATION

Table 5 shows that tempered selection makes several times as many commitments at low-confidence positions as PACE, and Table 6 shows that at $T_{\text{tok}} = 0.8$ the extra commitments do not improve coverage. Here we measure their cost. Because this cost builds up over many commitments rather than coming from a single one, we force a controlled *number* of such commitments and measure accuracy as that number grows.

At dose d, the first d diffuse steps of a generation (most confident position below 60% of the position mass, tail below c=0.5 holding at least 20%) commit at a position drawn from the tail instead of at the most confident position. Every other step commits by confidence order, and every dose shares the same prompt, model and number of steps, so the curve isolates the effect of the forced commitments alone. Dose 0 is pure confidence order.

Table 3: Accuracy against the number of forced commitments at low-confidence positions, 40 HumanEval problems with LLaDA-8B-Instruct. 'Achieved' is how many diffuse steps the generation actually offered, which saturates below the requested dose. $*$ marks a 95% paired interval against dose 0 excluding zero.

Forced	Achieved	Pass@1	vs. dose 0
0	0.0	**57.5**%	—
2	1.8	47.5%	-10.0 $[-25.0, +5.0]$
5	3.8	40.0%	-17.5 $[-32.5, -5.0]^*$
10	6.0	37.5%	-20.0 $[-35.0, -5.0]^*$

Accuracy falls monotonically with the number of forced commitments, and the decline is significant from 3.8 achieved commitments onward. Over the measured range each costs roughly 3.3 points. A run at half the problem count gave the same pattern (55.0, 45.0, 35.0, 25.0).

D.5 THE STATISTIC USED IN D

Section 4.2 argues that D must be a function of the committed pair, not of the position alone, and that among such functions the surprisal is distinguished. The first claim can be tested directly.

Because a candidate token is drawn at every masked position before any position is selected, scoring by the drawn token's probability makes selection prefer positions whose drawn token has high probability. Tempered selection based on the drawn token shifts the committed token toward likelier values (Theorem 5(ii)), whereas a position-only statistic leaves the draw independent of the selection (Theorem 5(iv)).

Therefore, scoring by the drawn token's confidence is robust to token temperature in a way that position-only statistics cannot be. At T_{tok}=1.6, many drawn tokens have low probability. Replacing c_k with $\max_v p_k(v)$ still commits them, and accuracy drops to a Pass@1 of 0.0 against 37.3 on the same problems; the realized entropy of the written tokens reaches 7.9 nats, about $e^{7.9} \approx 2{,}700$ effective tokens per commitment. A statistic that does not depend on the drawn token cannot avoid committing a low-probability token.

The commit convention matters for this comparison. Every method in this section holds the drawn token x_k^0 fixed and varies only the position score, so that methods differ in the statistic alone rather than in the statistic and the committed token together. The alternative, in which each rule commits the token its own score refers to, would make the $\max_v p_k(v)$ rule write the mode at every position; the

mode does not depend on T_{tok}, so that rule would be insensitive to token temperature by construction and the comparison could not test the statistic. Holding the drawn token fixed is what allows this test: the rule is given a low-probability token and cannot detect it.

Substituting the margin, the scale-free margin, or the maximum after token-level truncation for the confidence costs 5.8 to 10.0 points of Pass@1 at T_{tok}=0.8 for the same reason (Section H).

D.6 Adaptive-width baselines

Confidence thresholding (Wu et al., 2026) commits every position with $c_k > \tau_{\text{CT}}$, and its tempered form (Olausson et al., 2026) replaces that test with a Bernoulli draw through a sigmoid. Both commit a variable number of positions per step, so their cost depends on the data, and they must be compared by the number of forward passes, not only by k.

Table 4: Adaptive-width rules against PACE, HumanEval, LLaDA-8B-Instruct, 60 problems $\times$ 8 samples. The three occupy different points on the cost–accuracy plane rather than dominating one another.

	NFE	Pass@1	Pass@4	corr/8	gap at k=1
CT, τ_{CT}=0.9	**43.4**	57.3	63.8	4.58	-2.08 $[-8.1, +3.3]$
TCT, τ_{CT}=0.9, T_{pos}=0.1	66.8	56.0	64.5	4.48	-3.33 $[-7.9, +0.6]$
PACE	256.0	**59.4**	**69.5**	**4.75**	—

Neither gap at k=1 is distinguishable from zero at this sample size. Thresholding chooses how many positions to commit per step and PACE chooses which positions are eligible, so the two are complementary: a threshold rule can draw its committed set from PACE's eligible set.

E Implementation details

Gumbel-max sampler. We use the token sampler from the reference LLaDA decoder for comparability with prior work. It computes $\arg\max_v \exp(\ell_v)/(-\log u_v)^T$ with $u_v \sim \text{Uniform}(0,1)$. Taking logarithms gives $\ell_v + T\, g_v$ with $g_v = -\log(-\log u_v)$ a standard Gumbel variable, so this is the Gumbel-max trick at temperature T, sampling from $\text{softmax}(\ell/T)$; it reduces to greedy decoding at T=0.

Commit convention for position-scoring ablations. Every method in the main results scores positions by $c_k = p_k(x_k^0)$ and commits x_k^0, so the score and the committed token refer to the same object. The ablations of Section D.5 and Section H replace that score with a statistic of p_k alone, which leaves open what the method writes. All of them hold the drawn token fixed and vary only the score, so that a method differs from the reference in exactly one respect and the comparison isolates the statistic. This is also the only convention under which the token-temperature probe is informative: a rule that committed the token its own score refers to would write the mode, which does not depend on T_{tok}.

Position sampling. Positions are ordered by `torch.sort`, and tied confidences keep the order the sort returns; the statements of Sections 4 and 5 hold for any tie-breaking order.

Dream logit shift. Dream was initialized from an autoregressive model and emits the distribution for position i at index $i-1$. We shift the logits accordingly, and verify numerically that the shift is applied for Dream and not for LLaDA.

Reproducibility. Every run records, per method, the model identifier, benchmark, problem count, sample count, temperatures, block length, step count and a timestamp, and retains all generated text and extracted answers. Output filenames carry the same identifiers.

F MOTIVATING CASE STUDY

Tempered selection commits at low-confidence positions on purpose. Its premise is that deterministic confidence order fills easy positions first, so that when an uncertain position is reached, its context is already fixed and its options are gone. A random order restores these options. We accept this premise. Our question is whether a temperature uses exploration well, not whether exploration is worth having.

The extra exploration does not improve coverage. Exploration has a cost: every commitment at a position the model rates poorly is a token that all later predictions condition on. A sampler should therefore be judged on what its exploration gains. The number of such commitments can be computed exactly. At each step t a rule induces a distribution q_t over masked positions, so the expected number of commitments made at positions rated below ξ over a 256-step generation is 256 times the mean over steps of $\sum_{k:c_k<\xi} q_t(k)$, and the expected confidence of what is written at step t is $\sum_k q_t(k)c_k$. Evaluating both on the *same* trajectory compares the rules step for step.

Table 5: Exploratory commitments per 256-step generation, measured over 8 HumanEval generations with LLaDA-8B-Instruct. In each row, both entries are exact expectations on the same trajectory, which follows confidence order; steps with fewer than four masked positions are excluded. The ratio lies between 4.1 and 5.0 at the three thresholds.

	tempered	PACE (ours)	ratio
Mean confidence of the committed position	0.93	**0.97**	—
Commitments at $c < 0.3$, per generation	4.0	**0.8**	5.0×
Commitments at $c < 0.5$, per generation	10.5	**2.3**	4.6×
Commitments at $c < 0.7$, per generation	20.4	**5.0**	4.1×

Tempered selection makes 4.1 to 5.0× as many exploratory commitments as PACE at the thresholds tested. If these extra commitments improved coverage, the improvement would appear when many samples are drawn. This is not the case: at $T_{\text{tok}} = 0.8$ on LLaDA/HumanEval, tempered selection reaches $\text{Pass@32} = 77.0$ against 85.0 for PACE (Table 6), while single-sample accuracy differs by 22.3 points in favor of PACE.

Summary. Exploration itself is not the problem with tempered selection. The problem is that a temperature keeps every position eligible at every step, so it uses exploration on commitments that the model rates poorly. Proposition 7 shows that at every step the confidence of the position PACE commits stochastically dominates that of the position tempered selection commits, and Table 5 measures the size of the difference. Section D.4 measures what the extra low-confidence commitments cost.

G FULL Pass@k CURVES

This appendix gives the per-cell values behind Figures 1 and 3: the full Pass@k table at $T_{\text{tok}} = 0.8$ (Table 6), the gaps over both baselines (Table 7), and the curves to k=64 at $T_{\text{tok}} = 0.6$ (Table 8).

Table 6: Pass@k and mean correct samples per problem at $T_{\text{tok}} = 0.8$, 100 problems, 32 samples. **PACE** is Algorithm 1; **TLC** is tempered selection at T_{pos}=0.5; **AO** is confidence order; **AR** is left-to-right order. Bold marks the best entry in each column within a cell. Figure 3 plots the same values.

		Pass@k						
Cell	Method	1	2	4	8	16	32	corr/32
LLaDA / HumanEval	**PACE**	**58.8**	**65.8**	**71.2**	**76.1**	**80.6**	**85.0**	**18.82**
	TLC	36.5	47.4	56.7	64.7	71.6	77.0	11.69
	AO	49.7	55.0	59.6	64.2	68.8	73.0	15.91
	AR	39.6	51.1	60.5	68.0	73.3	77.0	12.68
LLaDA / GSM8K	**PACE**	**83.4**	**89.8**	**93.6**	**96.3**	**97.9**	**99.0**	**26.69**
	TLC	73.2	84.1	90.1	93.9	96.5	98.0	23.44
	AO	76.5	82.6	86.8	89.8	92.0	93.0	24.47
	AR	76.6	86.3	91.1	94.1	96.4	98.0	24.52
LLaDA / MBPP	**PACE**	**50.5**	**57.6**	**63.4**	**68.3**	**72.1**	**75.0**	**16.16**
	TLC	31.4	43.7	54.1	61.6	66.9	72.0	10.06
	AO	43.2	48.6	53.0	56.7	59.7	62.0	13.83
	AR	35.8	45.2	52.7	59.2	65.1	70.0	11.44
Dream / HumanEval	**PACE**	**72.1**	**81.1**	**87.5**	**91.6**	**94.4**	**97.0**	**23.07**
	TLC	44.0	60.0	72.7	82.0	88.8	94.0	14.07
	AO	66.5	74.6	80.3	84.5	87.8	91.0	21.28
	AR	49.7	66.0	77.8	86.0	92.2	96.0	15.91
Dream / GSM8K	**PACE**	**89.5**	**94.0**	**96.9**	**98.6**	**99.6**	**100.0**	**28.64**
	TLC	69.4	83.4	91.2	95.7	97.9	99.0	22.21
	AO	85.4	90.3	94.1	96.4	97.4	98.0	27.33
	AR	79.1	88.5	93.6	96.3	97.6	98.0	25.32
Dream / MATH-500	**PACE**	**53.6**	**64.9**	**73.5**	**80.2**	**85.2**	**90.0**	**17.15**
	TLC	33.9	46.4	57.7	67.2	74.5	80.0	10.86
	AO	47.4	57.7	66.5	73.8	80.3	86.0	15.17
	AR	38.4	51.2	62.1	70.9	77.1	82.0	12.30
LLaDA / MATH-500	**PACE**	**40.3**	**50.2**	**59.3**	**67.6**	**74.4**	**80.0**	**12.90**
	TLC	28.8	38.8	48.9	58.0	66.1	75.0	9.22
	AO	33.4	41.9	50.5	58.5	65.8	73.0	10.68
	AR	29.6	39.4	48.7	57.0	64.1	71.0	9.47
Dream / MBPP	**PACE**	**63.6**	**71.8**	**78.0**	**82.9**	**86.9**	**91.0**	**20.35**
	TLC	39.1	53.1	64.8	73.9	81.0	86.0	12.52
	AO	57.0	63.8	69.4	74.0	78.3	82.0	18.23
	AR	44.3	56.9	66.4	74.2	81.2	87.0	14.17

Table 7: Gap of Algorithm 1 over the two surprisal-damage baselines: confidence order (AO, $\lambda \to \infty$) and tempered selection (TLC, λ=$1/T_{\text{pos}}$). Cells are ordered by the gap over TLC at k=1; the ordering by the gap over AO is different, and the two columns are not interchangeable. The advantage over TLC decays with k in every cell while the advantage over AO does not on five of eight cells. Entries are differences of the per-cell means in Table 6; gaps in corr/32 are 0.32 times the k=1 gaps and are omitted.

	vs. confidence order		vs. tempered selection			
Cell	k=1	k=32	k=1	k=2	k=4	k=32
Dream / HumanEval	+5.6	+6.0	+28.1	+21.1	+14.8	+3.0
Dream / MBPP	+6.6	+9.0	+24.5	+18.7	+13.2	+5.0
LLaDA / HumanEval	+9.1	+12.0	+22.3	+18.4	+14.5	+8.0
Dream / GSM8K	+4.1	+2.0	+20.1	+10.6	+5.7	+1.0
Dream / MATH-500	+6.2	+4.0	+19.7	+18.5	+15.8	+10.0
LLaDA / MBPP	+7.3	+13.0	+19.1	+13.9	+9.3	+3.0
LLaDA / MATH-500	+6.9	+7.0	+11.5	+11.4	+10.4	+5.0
LLaDA / GSM8K	+6.9	+6.0	+10.2	+5.7	+3.5	+1.0

Table 8: Full k curves to k=64, LLaDA-8B-Instruct, $T_{\text{tok}} = 0.6$, 60 problems, 64 samples, tempered selection at $T_{\text{pos}} = 0.5$. *gap* is PACE minus tempered selection, paired bootstrap.

		Pass@k							
Cell	Method	1	2	4	8	16	32	64	corr/64
HumanEval	**PACE**	**60.0**	**65.5**	69.4	72.7	75.2	77.1	78.3	**38.38**
	TLC	53.7	63.1	69.4	73.7	77.3	80.1	81.7	34.37
	AO	59.1	62.5	65.3	68.1	70.8	72.7	73.3	37.83
	AR	55.5	64.4	**71.2**	**76.5**	**80.3**	**83.7**	**86.7**	35.52
	gap	$+6.28^{*}$	$+2.35$	$+0.07$	-1.02	-2.02	-3.02^{*}	-3.33	$+4.02^{*}$
MBPP	**PACE**	40.3	46.2	50.4	53.2	55.5	57.9	60.0	25.80
	TLC	36.1	45.0	51.4	56.5	61.9	68.2	**73.3**	23.10
	AO	**40.8**	44.0	46.5	49.1	51.7	53.9	55.0	**26.13**
	AR	39.0	**46.8**	**53.2**	**58.9**	**63.8**	**68.6**	**73.3**	24.98
	gap	$+4.22^{*}$	$+1.21$	-1.01	-3.31^{*}	-6.46^{*}	-10.23^{*}	-13.33^{*}	$+2.70^{*}$

H ADDITIONAL RESULTS

Position-only statistics. Section D.5 reports the collapse of the rule that scores by $\max_v p_k(v)$ at high token temperature. At $T_{\text{tok}}{=}0.8$ on 60 HumanEval problems the Pass@1 costs of replacing c_k in PACE are: maximum probability -5.83 $[-11.2, -0.4]^*$, margin $p_1 - p_2$ -5.83 $[-11.7, -0.8]^*$, scale-free margin $1 - p_2/p_1$ -6.25 $[-13.3, +0.8]$, and maximum after token-level truncation -10.00 $[-16.2, -3.8]^*$.