# SELF-REPULSIVE SAMPLING FOR DIFFUSION LANGUAGE MODELS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Sampling several responses and voting over their answers can improve a language model's accuracy, but repeated answers limit the benefit of additional samples. Raising temperature increases diversity at a potential cost to per-sample accuracy. We introduce Self-Repulsion (SR), a sampler for masked diffusion language models that uses peer commitments to diversify the pool. At each penalized denoising step, each path lowers a token's logit according to how many peers have committed that token at the same position. Paths share a batched forward pass and then commit in sequence, so later paths observe choices made earlier in the same step. This coupling requires no training or additional forward or backward pass and can produce distinct paths even at temperature zero. When all paths commit a position together from identical logits, the update exactly maximizes total logit minus a convex duplication cost. On LLaDA-8B-Instruct with ten paths and 128 denoising steps, deterministic SR reaches $80.38\%$ plurality accuracy on GSM8K, compared with $70.17\%$ for the unpenalized greedy decoder. At temperature $0.6$ and matched model-evaluation budgets, the count penalty improves over self-consistency by $2.06$ percentage points in blocks of 32 and $14.50$ under pure diffusion. Experiments on GSM8K, MATH and TruthfulQA show that voting gains arise mainly from higher coverage of correct answers, with gains that vary by benchmark and decoding regime.

## 1 INTRODUCTION

Sampling several reasoning paths and voting over their answers is a standard way to improve a language model's accuracy with additional inference computation. Self-consistency (Wang et al., 2023) samples chain-of-thought paths independently (Wei et al., 2022) and returns the most frequent final answer, without training a verifier. Its success depends on the samples available to the vote. In particular, *coverage*, the probability that the pool contains a correct answer, upper-bounds the accuracy of any selector over that pool. Repeated answers can limit the benefit of additional samples, but disagreement alone is insufficient: a more diverse pool need not yield a more accurate vote (Bay & Yearick, 2026; Olausson et al., 2026). The sampling problem is therefore to find more correct answers while retaining enough evidence for the vote to select them.

Temperature offers a simple way to change the pool. At low temperature, samples concentrate on a few answers; at higher temperature, they explore more alternatives but can lose per-sample accuracy. Temperature and truncation act on each sample's own distribution, regardless of what other samples have already produced. Couplings based on shared randomness can spread the draws while preserving their marginal distributions (Vilnis et al., 2023; Amballa et al., 2025). We investigate a different signal: the tokens that other paths have already committed while the ensemble is being generated. Can this information improve coverage without relying on additional sampling noise?

Masked diffusion language models (MDLMs), a prominent form of discrete diffusion, are an emerging alternative to autoregressive generation (Li et al., 2025). A response begins fully masked and the decoder runs $T$ denoising steps, each a single forward pass, or *network function evaluation* (NFE). Each step predicts every still-masked position in parallel (Ye et al., 2025) and commits the most confident (Nie et al., 2025). Generation costs $T$ NFEs rather than one forward pass per token, which industrial systems increasingly build on (Inception Labs et al., 2025; Song et al., 2025; DiffusionGemma Team, 2026). This structure makes an alternative source of ensemble diversity



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.

available at almost no cost. When $K$ samples decode the same positions on a shared schedule, each reads its peers' current commitments from the batch the decoder already holds, while the pool is still forming; an any-order decoder can then defer a contested position and repair around a substitution, which a fixed-order decoder cannot. This signal has a timing. Between steps the pool is fixed, so a penalty read at the *start* of a step sees only the commitments of earlier steps. Read *within* a step, as the paths commit in turn, it also sees the tokens peers commit earlier in that same step. **Released decoders leave this signal unused;** in the case of LLaDA they even use greedy (argmax) decoding by default (Nie et al., 2025).

We introduce Self-Repulsion (SR), a sampler that uses this within-step information. At each penalized step, a path lowers a token's logit in proportion to the number of peers that have committed that token at the same position. All paths share one batched forward pass, then commit in a fixed order; each path reads the updated peer counts before making its choices. We call this ordered update the *cascade*. The confidence-based commit rule and final plurality vote remain unchanged (Figure 1, Section 2). The penalty requires no training, learned potential, or additional forward or backward pass.

Temperature zero isolates the role of the cascade. With identical initial canvases, a shared deterministic update and path-independent tie-breaking, paths that read the pool only at the start of a step remain identical (Proposition 1). The cascade can break this symmetry because later paths observe earlier commitments. Deterministic ensembles and diversity penalties have precedents in CoT-decoding, Diverse Beam Search and Orthogonal Diverse Diffusion (ODD) (Wang & Zhou, 2024; Vijayakumar et al., 2016; Lamont et al., 2026); our contribution is to couple masked-diffusion paths through their current commitments at each position and evaluate the resulting pools under a fixed voting rule.

**Contributions:**

- **Self-Repulsion (Section 2).** A drop-in sampler that draws ensemble diversity from the pool, not from noise: at each still-masked position it lowers each token's logit in proportion to how strongly the peers have committed to that token, adding no extra forward or backward pass on a shared schedule.
- **Ensembles without randomness.** At temperature zero a symmetric start-of-step sampler is degenerate: its $K$ samples coincide, so the vote cannot beat a single decode. Reading the peers within the step can make the paths differ. When all paths commit a position together from identical logits, the cascade maximizes total logit minus a convex cost of duplication. Measured, they vote above self-consistency with no randomness, on GSM8K reaching $80.38$ against self-consistency's $77.05$ and the greedy decode's $70.17$ (128 steps).
- **Experiments on masked diffusion (Section 3).** On LLaDA-8B-Instruct over GSM8K, MATH and TruthfulQA, SR matches or exceeds the diversity methods we reproduce at matched computation, the gain in coverage, not the vote, and in blocks 2.46 points above a peer-blind control at matched disagreement (Appendix D.3). Its gain over self-consistency is larger where duplication is worse: on GSM8K, $+2.06$ points in blocks of 32 and $+14.50$ under pure diffusion, at the same 1280 evaluations.

## 2 Self-Repulsion

In the following, we characterize Self-Repulsion from three angles. As a *sampler*, it is a single additive term on the logits, by which the $K$ paths of one parallel decode repel one another through their commitments, with no added randomness and no extra forward or backward pass (Section 2.1). As a *penalty*, at temperature zero it *allocates* the paths at a position they commit together from identical logits, the cascade maximizing an explicit quality-minus-duplication objective (Section 2.2). As one *sampler among others*, its asymmetry is the information order alone, the samples reading one another as the pool forms; at temperature zero a peer-blind ensemble collapses, which reading the peers within the step escapes (Section 2.3).

### 2.1 Sampler

Self-Repulsion runs $K$ masked-diffusion *paths* on one prompt in parallel. Each path fills its own *canvas* and the $K$ canvases together form the *pool*. The paths are coupled through their logits: at every

denoising step, at each still-masked position $j$, the law of each path is tilted away from the tokens *peers*, the other paths, have committed there. Within a step the paths commit in turn, each reading the tokens the paths ahead of it committed; we call this ordered commit the *cascade*. Writing $p$ for the law the path would otherwise sample from and $\Delta$ for a non-negative statistic of the peers' commitments at $j$, for $\tau > 0$ the path samples from

$$q(v) \;\propto\; p(v)\,\exp\Big(-\tilde{\lambda}\,\Delta_v\Big)\,. \tag{1}$$

The $K$ answers are then aggregated by a plurality vote. Self-consistency is the $\lambda = 0$ member of this family, with the penalty disabled; this section asks what a positive strength provides.

**Cascade.** When each path reads its peers is the one choice in the sampler that is not forced, and it is what makes the method work. The natural implementation reads $\Delta$ once after the batched forward pass and commits all $K$ paths in parallel; then $\Delta$ is fixed at the start of the step, identical across the paths, so at $\tau = 0$ they commit the same tokens and never diverge (Proposition 1). The cascade instead reads $\Delta$ within the per-path loop, so a path sees what the paths ahead of it committed at this same step. Formally, the $K$ paths share a common initial canvas, each is scored on its own, and each commits by argmax under an index-free tie-break after one shared update from a statistic $\Delta$ of the pool.

**Proposition 1** (Degeneracy at zero temperature)**.** *At $\tau = 0$, if each path's $\Delta$ is one function, shared across paths, of its own canvas and its peers' canvases, read at the start of the step, the canvases coincide at every step and the pool holds one distinct sample; self-consistency at $\tau = 0$ is the case $\Delta \equiv 0$.*

**Setup.** The sampler adds a peer-count penalty to the logits and reads these counts within the ordered commitment loop, while retaining the original decoder's confidence-based commit rule and denoising schedule. With raw logits $\boldsymbol{\ell}$ and strength $\lambda \geq 0$, every path applies $\ell_k[j, v] \leftarrow \ell_k[j, v] - \frac{\lambda}{K-1}\,\Delta_v$. This splits the strength evenly across the $K-1$ peers, so a unanimous pool costs $\lambda$ whatever $K$ is. For $\tau > 0$, dividing by the temperature $\tau$ carries the penalty into the exponent and recovers equation (1). We distinguish three related quantities:

$$\text{strength } \lambda, \qquad \text{gate } g = \lambda/(K-1), \qquad \text{tilt } \tilde{\lambda} = g/\tau \quad (\tau > 0). \tag{2}$$

At the deployed $K = 10$ and $\tau = 0.6$, the gate is $g = 7.11$ (strength $\lambda = 64$, tilt $\tilde{\lambda} = 11.85$). Across $K$ we hold the strength $\lambda$, not the gate: $gK = \lambda K/(K-1)$ tends to $\lambda$, so the fractional allocation proportions approach a $K$-independent limit at fixed $\lambda$ and flatten toward uniform at fixed $g$ (Appendix C.3); the ablation tables report $\lambda$ and $g$ together.

The penalty is subtracted from the raw logits, before dividing by $\tau$, so for $\tau > 0$, $q$ is an exponential tilt of $p$ that strengthens as $\tau$ falls. The logit penalty remains defined at $\tau = 0$. At $\tau = 0$ the decoder takes $\arg\max_v(\ell_v - g\,c_v)$, with $c_v$ the number of peers committed to token $v$ at $j$, so the gate is a margin in raw logits: the top token loses once $g$ times its excess peer count over a competitor exceeds its raw-logit lead over that competitor. At the deployed $\tau = 0.6$ the tilt is $\tilde{\lambda} = 11.85$, so a token even one peer holds is down-weighted by $e^{-\tilde{\lambda}} \approx 7 \times 10^{-6}$, close to hard exclusion unless its raw-logit lead over the best untaken token exceeds the gate (Appendix D.1).

**Penalty statistic.** We deploy the *count* statistic, $\Delta_v = c_k[j, v]$, the number of peers committed to token $v$ at position $j$ when path $k$ reads the pool. We suppress $k$ and $j$ below. Appendices C.4 and C.5 give two variants that read the peers' predicted distributions instead, which can down-weight a token before any peer commits it but degenerate at $\tau = 0$, where those distributions are start-of-step state (Proposition 1); each induces equation (1). Nothing here is specific to masked diffusion beyond what an any-order decoder does with a pushed-off path, which a fixed-order decoder cannot. It can defer a contested position: a path pushed off a peer-held token scores that position by its replacement's lower untilted confidence and so defers it, and at $\tau = 0$ and gate 8 only 3.00% of the commits made under the penalty in blocks of 32 differ from the path's own argmax, each where a peer already holds that argmax. And it can repair around a substitution, the neighbors still masked and attention bidirectional, so a perturbation stays local (Appendix D.1); the confidence order helps both, penalizing early structural commits and releasing before the late uncertain ones. Masked diffusion is the home of the method because it can defer and repair, not because it keeps the logits shared. In a fixed-order decoder this within-step count read is Diverse Beam Search's Hamming penalty at one path per group (Vijayakumar et al., 2016); the two properties above are what any-order decoding adds.

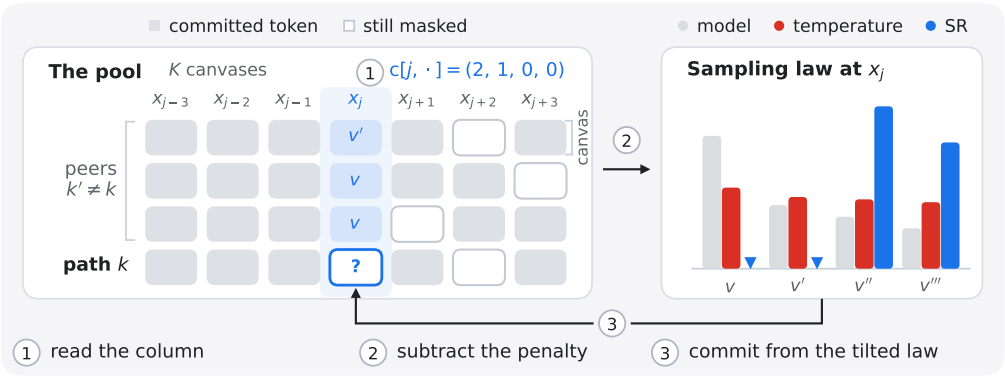


Figure 1: One denoising step of Self-Repulsion, four paths, illustrative values. **Left:** the $K$ paths decode in parallel; the column at a still-masked position $j$ is the count vector $\mathbf{c}_k[j, \cdot]$, the peers committed to each token (here $(2, 1, 0, 0)$). The step runs in three parts: (1) read the column, (2) subtract the penalty $g\,\mathbf{c}_k[j, \cdot]$ from the logits, (3) commit from the tilted law. **Right:** the sampling law at $x_j$ under the model, under temperature and under SR. The tilt down-weights tokens peers hold and preserves probability ratios between tokens with equal counts; temperature reshapes every token, the taken ones included. Values are schematic except the effective tilt $e^{-11.85}$ per peer.

**Algorithm 1** Self-Repulsion. Shading marks the peer-count read and logit penalty, being the only modification to self-consistency. Counts are read *inside* the cascade, so later paths see earlier commitments from the same step. $s$ is the prompt, $j$ indexes positions, and $v$ indexes tokens. Token updates apply only at still-masked positions.

**Require:** model $f_\theta$, prompt $s$, paths $K > 1$, gate $g = \lambda/(K-1)$, temperature $\tau$, steps $T$, penalty scope $\rho$ (active during the first fraction $\rho$ of each block)

$\mathbf{x}_k \leftarrow (\texttt{[MASK]}, \ldots, \texttt{[MASK]})$ for every path $k \le K$
**for** $t = 1, \ldots, T$ **do**
  $\boldsymbol{\ell}_k \leftarrow f_\theta(s, \mathbf{x}_k)$ for all $k$ ▷ one batched forward pass
  **for** $k = 1, \ldots, K$ **do** ▷ cascade, fixed path order
    $c_k[j, v] \leftarrow \#\{k' \neq k : x_{k'}[j] = v \neq \texttt{[MASK]}\}$ at masked $j$ ▷ sees this step's earlier commits
    $\boldsymbol{\ell}'_k \leftarrow \boldsymbol{\ell}_k - g\,\mathbf{c}_k$ if the penalty is active, else $\boldsymbol{\ell}_k$
    $\tilde{x}_k[j] \leftarrow \arg\max_v \ell'_k[j, v]$ if $\tau = 0$, else $\tilde{x}_k[j] \sim \mathrm{softmax}(\boldsymbol{\ell}'_k[j, \cdot]/\tau)$
    commit $\tilde{x}_k[j]$ where the original confidence rule selects (scored by untilted probability); other positions stay masked
  **end for**
**end for**
**return** $\arg\max_a \sum_k \mathbf{1}\{\mathrm{answer}(\mathbf{x}_k) = a\}$, ties uniform ▷ plurality vote

**Cost.** SR adds no extra forward pass, no backward pass, no learned potential and no training. The paths share a canvas layout and block schedule, so peer counts can be computed directly from the commitments already stored in the batch. The added computation consists of reading these counts and subtracting the corresponding penalty from each path's logits. The original decoder's commit and remasking rules are unchanged: positions are ranked by the untilted probability of the selected token. Changing that token can therefore change when a position is committed. The penalty is active during the first three-quarters of each block's denoising steps and disabled thereafter. Algorithm 1 gives the full procedure.

The closest analogue, ODD, spends a gradient step on the logits per step, orthogonalizing each sample in a pooled feature space (Lamont et al., 2026); particle guidance needs a fixed or learned potential (Corso et al., 2023); determinantal beam search needs a similarity kernel and greedy log-determinant maximization (Meister et al., 2021). SR needs only the resident batch column its paths already share, without additional model evaluations.

The processing order inside the cascade, held fixed across steps, is a free choice that does not change the pool: at temperature zero every such order yields the same multiset (Remark A.1, Appendix A).

### 2.2 PENALTY

At temperature zero the penalty *allocates*: where the paths commit a position together from identical logits, the cascade maximizes total logit minus a duplication penalty. Proofs are in Appendices A.2 and A.3.

**Allocation at a shared position.** Suppose all $K$ paths commit the same position in one step and face identical logits. Here $c_v$ counts the paths allocated to token $v$. Its successive choices have scores $\ell_v,\ \ell_v - g,\ \ell_v - 2g, \ldots$: each commitment lowers that token's next score by $g$. The cascade takes the $K$ largest scores across these descending sequences. The *water level* is a common cutoff separating selected from unselected scores; it is determined directly by the resulting allocation (Figure 4 in Appendix).

**Proposition 2** (Allocation at a shared position)**.** *At $\tau = 0$, at a step where all $K$ paths commit the same position, none having committed it before, and face the same logits $\boldsymbol{\ell}$, the commitment counts $\mathbf{c}$ exactly maximize $\sum_v c_v \ell_v - g \sum_v \binom{c_v}{2}$ over $\{\mathbf{c} \in \mathbb{Z}^{\mathcal{V}}_{\geq 0} : \sum_v c_v = K\}$, by water-filling against a common level: the penalty allocates the paths by trading total logit against the pool's duplication, one gate $g$ per agreeing pair. Away from shared logits each path still commits $\arg\max_v(\ell_k[j, v] - g\, c_v)$, but the global optimality guarantee no longer applies.*

*Reading.* A path switches tokens when another token has a higher penalized score. The allocation need not make every path distinct. The result holds for the count statistic, which charges only committed tokens; the distribution-reading variants are not covered. Its hypothesis is narrow: in the decodes above, all $K$ paths commit a position at one step at only 4.9% of positions, fewer still under full canvas, the regime with the largest gain.

**Corollary 3** (Sparsemax form of the fractional allocation)**.** *For $g > 0$, allow the counts in Proposition 2 to take nonnegative real values while retaining $\sum_v c_v = K$. The unique optimum is*

$$\mathbf{c}^\star = K\,\mathrm{sparsemax}\Big(\frac{\boldsymbol{\ell}}{gK}\Big),$$

*where* sparsemax *denotes Euclidean projection onto the probability simplex $\{\mathbf{p} \geq 0 : \sum_v p_v = 1\}$ (Martins & Astudillo, 2016). The cascade's integer counts satisfy $|c_v - c_v^\star| < 1$ for every token $v$ (Appendix A.3).*

*Reading.* The fractional solution is $c_v^\star = \max\{0, (\ell_v - \mu)/g\}$, where the threshold $\mu$ is chosen so that the counts sum to $K$. Tokens at or below $\mu$ receive zero allocation; above it, allocation is proportional to the excess logit. The cascade assigns whole paths, within one path per token of this fractional solution. The continuous threshold $\mu$ need not equal the integer cutoff described above.

**Tilt.** With randomness, unlike temperature, the tilt raises untaken tokens and lowers those of the largest count, preserving equal-penalty ratios (Figure 1, right; Appendix A.4). Like any exponential tilt, it is the I-projection of $p$ onto laws with at most its expected peer agreement (Lemma A.2; Csiszár, 1975).

### 2.3 DISTINCTION TO OTHER SAMPLERS

**Sources of asymmetry.** For the $K$ paths to decode differently, the decode must depend on which path is running. At temperature zero, with no randomness, only three parts of the decode can carry that dependence:

- the *rule* each path applies; different rules give an ensemble of configurations (Lee et al., 2025), or an index-ordered penalty breaks the symmetry (Lamont et al., 2026);
- the *assignment* of tokens to paths; handing out shared candidates by rank is a search (Wang & Zhou, 2024);
- the *information order*: what each path reads of its peers and when.

A *drop-in* sampler, one shared configuration and one additive penalty, forgoes the first two; the cascade realizes the information order by committing the paths in sequence, each reading its predecessors, though not uniquely (Appendix D.4).

**Peer-blind boundary.** A *peer-blind* sampler transforms each sample without reading its peers, drawing its randomness independently across samples. A pool can differ from an independent self-consistency pool in only two ways: a different per-sample law, or correlation. Writing $P$ for the joint law, $P_k$ for its marginals and $p$ for one untilted sample, the divergence splits,

$$\mathrm{KL}\big(P \,\big\|\, p^{\otimes K}\big) \;=\; \sum_k \mathrm{KL}(P_k \,\|\, p) \;+\; \mathrm{KL}(P \,\|\, \textstyle\bigotimes_k P_k)\,,$$

into a *marginal* term (the samples' own laws) and a *dependence* term (their correlation, a total correlation). A peer-blind sampler leaves the dependence term at zero: a product channel carries a product law to a product law, so its $K$ draws are i.i.d. from one law, and their mean pairwise answer agreement within a pool equals the mean across independent pools, $\mathbb{E}[M_{\mathrm{within}}] = \mathbb{E}[M_{\mathrm{cross}}]$. The split is a statement about the sampled configurations; against a deterministic pool its marginal terms are infinite, and $\tau = 0$ is covered by Proposition 1 instead. What a peer-blind sampler can still move is the marginal term, and coupling the draws at fixed marginals raises coverage by at most a bounded amount.

**Proposition 4** (Coupling ceiling on coverage)**.** *At fixed per-sample marginals, re-coupling raises coverage above its independent value by at most* $(1 - 1/K)^K < e^{-1}$.

*Reading.* Correlating the draws without changing their laws adds at most $0.3487$ coverage per problem at $K = 10$ (Appendix A.6), so any gain above that ceiling must change the laws. Temperature, the truncation family and any self-computed penalty ($\Delta = p$) are peer-blind, their pools i.i.d.; Section 3.4 tests whether a self-computed penalty tracks temperature.

## 3 Experiments

We test whether peer commitments can diversify deterministic decoding, how repulsion compares with increasing temperature, and whether the resulting pools improve plurality accuracy. We compare with self-consistency and reproduced diversity methods at matched numbers of model evaluations, and use controls to examine the role of peer information. Additional results are in Appendix D.

**Metrics.** *Plurality accuracy* scores the most frequent parsed answer, averaging uniformly over tied answers. *Pass@$k$* measures whether a subset of $k$ paths contains a correct answer; pass@1 is per-sample accuracy and pass@10 is coverage. *Disagreement* is the rate at which two paths return different parsed answers. Metrics are computed per problem and then averaged (Appendix B.1); scores and differences are reported as percentages and percentage points, respectively.

**Setup.** We evaluate on GSM8K (Cobbe et al., 2021), MATH (Hendrycks et al., 2021) and TruthfulQA (Lin et al., 2022). All ensemble methods use LLaDA-8B-Instruct in BF16 with $K = 10$ paths and generation length 256. Appendix C.8 adds a second checkpoint, LLaDA-1.5 (Zhu et al., 2025), with similar GSM8K gains over self-consistency (coverage $+6.75$, plurality $+2.46$). Unless specified otherwise, SR uses the count statistic with strength $\lambda = 64$ (gate $g = 7.11$; recommended settings in Table 10). We compare pure diffusion, which treats the response as one block, with semi-autoregressive blocks of 32, reported to improve LLaDA-Instruct on GSM8K and MATH (Nie et al., 2025). Strength was selected on 200 problems with one seed; the penalty scope was fixed beforehand to the first three-quarters of each block's steps ($\rho = 0.75$; Appendix C). Sampled configurations use $\tau = 0.6$ unless a temperature sweep is stated. The released decoder instead uses greedy decoding with length and step count 512 (Nie et al., 2025). Table 2 uses 128 denoising steps per path (1280 evaluations in total); the temperature sweeps, Table 13 and Table 5 use 256 steps (2560 evaluations). These budgets count model evaluations and exclude sampler-specific overhead.

**Baselines.** TAPS perturbs prompt embeddings with annealed noise (Wu et al., 2026); ODD diversifies samples through a penalty on their pooled features (Lamont et al., 2026). We hold the model, prompt, length, schedule, path count, step count and aggregation fixed, and vary the sampler. This protocol excludes the schedule ensemble of Lee et al. (2025), which varies decoding configurations. All baseline numbers are our reproductions. We also vary each family's temperature (Appendix D.2); Appendix B documents the configurations and deviations from released code.

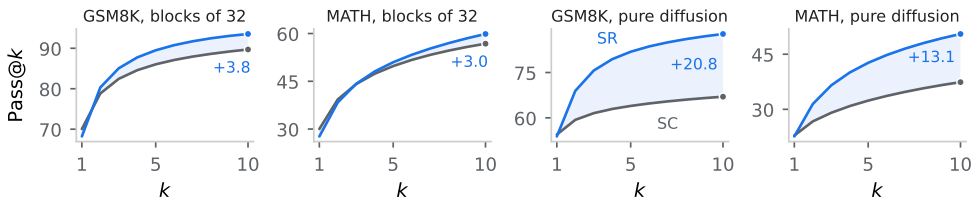


Figure 2: Pass@$k$, self-consistency against SR (count statistic), Table 2's campaigns; shading is the gap, the printed number its size at $k$=10. SR is below self-consistency only at $k \leq 2$, the per-sample end of the curve, and above it at every larger $k$, including MATH; all methods and values in Table 17.

### 3.1 Deterministic decoding

With the penalty disabled, the ten paths at temperature zero are identical. On GSM8K with 256 steps and blocks of 32, they reproduce the greedy decoder's 73.00 accuracy. Enabling the count penalty at gate 8 ($\lambda = 72$) raises coverage to 92.33 and plurality accuracy to 78.60 without token-sampling noise (Table 13; an example decode in Figure 5). This is 0.33 points below the sampled SR configuration's five-seed mean of 78.93 (Table 4). The 128-step results in Table 2 show the same qualitative effect.

### 3.2 Temperature and repulsion

Under pure diffusion with 256 steps, we sweep temperature with repulsion disabled, then sweep penalty strength at $\tau = 0.6$. Table 1 compares the resulting trade-offs among disagreement, per-sample accuracy, coverage and plurality accuracy. These runs predate the end-of-sequence correction used in Table 2 (Appendix D.9).

Table 1: The two dials swept over their usable range: GSM8K under pure diffusion, $K$=10, 2560 evaluations, $n$=600 over two seeds. Panel B's $\lambda$=0 row is Panel A's $\tau$=0.6 row. Higher is better in the accuracy columns; disagreement is a rate, not an objective. Bold marks temperature's best rung on each metric; $\star$ our deployed setting ($\lambda = 64$, gate $g = 7.11$), which exceeds temperature's best on plurality and coverage (Panel B unmarked; Appendix D.7).

*A. Temperature, $\lambda = 0$*

| $\tau$ | Disagr. | Per-s. | Plur. | Cover. |
|---|---|---|---|---|
| 0.6 | 15.02 | 53.42 | 54.53 | 63.17 |
| 0.8 | 20.87 | **54.21** | 56.31 | 66.83 |
| 1.5 | 34.86 | 51.95 | 57.83 | 73.08 |
| 2.0 | 50.61 | 47.77 | 60.72 | 79.92 |
| 2.5 | 64.33 | 41.34 | **62.00** | **81.25** |
| 3.0 | 88.87 | 16.73 | 35.62 | 62.75 |

*B. Count penalty, $\tau = 0.6$*

| $\lambda$ | Disagr. | Per-s. | Plur. | Cover. |
|---|---|---|---|---|
| 0 | 15.02 | 53.42 | 54.53 | 63.17 |
| 4 | 33.46 | 54.48 | 60.78 | 77.17 |
| 16 | 42.63 | 56.22 | 69.00 | 85.50 |
| 32 | 43.87 | 57.31 | 71.65 | 87.75 |
| $\star$ 64 | 46.60 | 56.19 | 73.09 | 86.83 |
| 128 | 47.80 | 55.36 | 71.66 | 88.33 |
| 384 | 53.46 | 53.30 | 73.10 | 90.00 |

Tuning the baseline's own dial does not reach the penalty. The deployed setting exceeds every temperature rung on plurality and coverage at once, 73.09 and 86.83 against column bests of 62.00 and 81.25 (Table 1). Temperature pays for its diversity in per-sample accuracy, which falls across the sweep and drags the vote down past $\tau = 2.5$; the penalty's holds between 53.30 and 57.31. At matched disagreement the per-sample gap plateaus near +7 from $\lambda = 16$ (deployed +7.36), a property of the dial, not one tuned point. The peer term produces it: a control reading the sample's own distribution returns to the temperature curve (Section 3.4). In blocks the contrast is +0.52.

### 3.3 Accuracy at matched compute

At its best variant per column, SR tops all six plurality columns of Table 2, over the strongest baseline by 2.65, 0.72 and 0.28 in blocks; a different variant leads each benchmark. The deployed count variant itself trails self-consistency on MATH in blocks by 0.10; on GSM8K the strongest baseline is the peer-conditioned ODD, which the temperature-zero count variant alone exceeds (80.38 against 78.00, 128 steps). TruthfulQA is the negative case, nothing to convert at a 94.37 chance ceiling.

Table 2: Every method at 1280 network function evaluations; Table 5 repeats it at 2560. GSM8K $n$=600, MATH $n$=492, TruthfulQA $n$=684, two to five seeds; TruthfulQA coverage omitted (94.37 chance). Rows run at $\tau$=0.6 except the three $\tau$=0 rows: row 1 is the greedy decode, equal to self-consistency (Proposition 1), and the two SR, count rows use gate 8 ($\lambda = 72$). Pure-diffusion rows carry the end-of-sequence correction (Appendix D.9); TAPS (†) is uncorrected. Bold: best per column within a regime, disagreement and the $K$=1 greedy row excluded.

| | Plurality (↑) | | | Coverage (↑) | | Per-sample (↑) | | | Disagreement | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Method | GSM | MATH | TQA | GSM | MATH | GSM | MATH | TQA | GSM | MATH | TQA |
| *Semi-autoregressive blocks of 32* | | | | | | | | | | | |
| Self-consistency, $\tau$=0 (= greedy, $K$=1) | 70.17 | 28.25 | 55.26 | 70.17 | 28.25 | 70.17 | 28.25 | 55.26 | 0.00 | 0.00 | 0.00 |
| Self-consistency | 77.05 | 40.14 | 59.02 | 89.70 | 56.81 | **70.04** | **30.02** | **54.44** | 27.14 | 62.01 | 34.09 |
| TAPS | 77.69 | 39.47 | 58.27 | 91.00 | 57.32 | 66.77 | 29.26 | 53.53 | 35.51 | 64.47 | 36.63 |
| ODD ($\alpha$=64) | 77.24 | 37.03 | 60.32 | 91.75 | 54.27 | 64.94 | 24.19 | 52.35 | 38.72 | 70.98 | 40.97 |
| ODD ($\alpha$=256) | 78.00 | 37.75 | 59.99 | 92.17 | 54.37 | 66.57 | 22.77 | 49.85 | 36.47 | 68.12 | 42.88 |
| SR, count | 79.11 | 40.04 | 59.77 | 93.53 | **59.86** | 68.21 | 27.67 | 48.07 | 35.53 | 68.82 | 49.88 |
| SR, count, $\tau$=0 | 80.38 | **40.86** | 59.84 | 94.33 | 57.11 | 69.33 | 27.13 | 46.92 | 34.35 | 69.13 | 51.61 |
| SR, exp. count | **80.65** | 39.58 | 59.82 | **94.67** | 59.65 | 66.47 | 26.78 | 50.75 | 40.60 | 71.45 | 43.66 |
| SR, collision-wt. | 80.07 | 38.81 | **60.60** | 94.42 | 55.39 | 61.87 | 24.44 | 52.78 | 48.63 | 75.08 | 40.53 |
| *Pure diffusion (whole response as one block)* | | | | | | | | | | | |
| Self-consistency | 57.02 | 24.00 | 56.27 | 67.00 | 37.40 | 54.47 | 22.87 | 54.17 | 20.40 | 40.89 | 24.81 |
| TAPS† | 56.82 | 25.67 | 55.14 | 69.75 | 39.74 | 50.34 | 22.46 | **55.26** | 32.74 | 45.38 | 14.96 |
| SR, count | 71.53 | 29.79 | 58.54 | 87.75 | 50.51 | 53.85 | 22.62 | 53.79 | 49.73 | 70.26 | 36.91 |
| SR, count, $\tau$=0 | 70.98 | 31.27 | **59.70** | 87.17 | 50.41 | 53.82 | 23.17 | 52.85 | 49.50 | 68.40 | 37.57 |
| SR, exp. count | 72.01 | 29.40 | 55.65 | **90.33** | 48.78 | 53.87 | 20.47 | 50.78 | 53.23 | 74.97 | 37.01 |
| SR, collision-wt. | **76.27** | **35.37** | 56.99 | 90.00 | **55.79** | **59.81** | **25.10** | 50.50 | 45.93 | 72.40 | 41.57 |

On MATH the pool improves where no aggregator we built converts it, pass@$k$ reaching $+3.05$ at $k = 10$ (Figure 2); the gain is a verifier's, not the vote's (Appendices D.5 and D.6).

Under pure diffusion, where duplication is worst (Appendix C.6), self-consistency's GSM8K vote falls to 57.02 against the collision-weighted penalty's 76.27 (Table 2); TAPS sits at 56.82 and blocking (Arriola et al., 2025; Nie et al., 2025) recovers most of the gap, at 77.05.

### 3.4 Peer-versus-own-distribution control

Reading the peers, not the disagreement the penalty produces, is what keeps per-sample accuracy above the temperature curve. A matched control isolates the two: with everything else fixed, each path reads its own predictive distribution in place of the peers' commitments. At matched disagreement this own-distribution control sits $+0.20$ and $+1.26$ above the temperature sweep, against $+7.35$ and $+12.71$ for the peer statistic (Table 3).

Table 3: Temperature against Self-Repulsion under pure diffusion, on the 1200 problem-seed pairs common to all nine configurations. The varied parameter is temperature for the first family and strength $\lambda$ for the other two (gate $g = \lambda/(K-1)$). The last two columns interpolate the temperature sweep to each configuration's own disagreement and give the per-sample accuracy it reaches there and the gap to it.

| Conditions on | Varies | Disagr. | Per-sample | Plur. | Cov. | Temp. here | Gap |
|---|---|---|---|---|---|---|---|
| Nothing (temperature) | $\tau$=0.6 | 15.02 | 53.42 | 54.53 | 63.17 | – | – |
| Nothing (temperature) | $\tau$=1.0 | 25.06 | 53.76 | 57.44 | 68.83 | – | – |
| Nothing (temperature) | $\tau$=1.2 | 28.25 | 53.24 | 57.82 | 71.33 | – | – |
| Nothing (temperature) | $\tau$=1.5 | 34.86 | 51.95 | 57.83 | 73.08 | – | – |
| Nothing (temperature) | $\tau$=2.0 | 50.61 | 47.77 | 60.72 | 79.92 | – | – |
| Own distribution | $\lambda$=1 | 24.48 | 53.94 | 57.57 | 69.33 | 53.74 | +0.20 |
| Own distribution | $\lambda$=2 | 32.63 | 53.64 | 59.72 | 74.17 | 52.39 | +1.26 |
| Peer comm., count | $\lambda$=64 | 46.60 | 56.19 | 73.09 | 86.83 | 48.84 | +7.35 |
| Peer comm., coll.-wt. | $\lambda$=128 | 40.91 | 63.06 | 76.85 | 91.25 | 50.34 | +12.71 |

## 4 Related work

No common mechanism for diversifying a model's $K$ samples lets a sample read, mutually and within the step, its peers' committed tokens at the same position. One family reshapes each sample's own distribution, by temperature, truncation, prompt-embedding perturbation (Wu et al., 2026), tempered remasking (Olausson et al., 2026), or within-sequence kernel-entropy guidance (Zhang et al., 2026); the draws stay independent. A second couples the pool through shared randomness: arithmetic codes with exact marginals (Vilnis et al., 2023, Prop. 3) or quasi-Monte Carlo, under a coverage ceiling on marginal-preserving samplers (Li et al., 2026, Thm. 2), and sampling without replacement, which changes the sampling probabilities (Kool et al., 2019). A third votes over varied decoding configurations, ensembling block schedules to avoid the failure modes of any single one (Lee et al., 2025); a fourth reads cross-sample agreement with the opposite sign, keeping agreed positions and resampling the rest (Feng et al., 2026).

In beam search, Diverse Beam Search subtracts a Hamming penalty across ordered beam groups (Vijayakumar et al., 2016) and determinantal beam search maximizes a log-determinant over a similarity kernel (Meister et al., 2021); SR instead couples $K$ samplers through the tokens they commit at each position as they draw. CoT-decoding builds a deterministic autoregressive ensemble (Wang & Zhou, 2024), assigning the top-$K$ first-step tokens to paths and continuing each greedily, so no path reads another. Sequential variants repel each sample from its predecessors, by an entailment score or branch avoidance (Park & Cho, 2025; Park et al., 2026), none voting over the pool; SemDiD instead runs its groups concurrently, repelling each group after the first, greedy one from its most similar other group in embedding space (Shi et al., 2025). The closest prior work is ODD (Lamont et al., 2026), which orthogonalizes each sample against its predecessors in a position-pooled confidence space, one gradient step per denoising step, its feature carrying no position index and its target $\mathrm{pass@}k$. ODD too diversifies at temperature zero, where each sample diverges only from its predecessors; SR instead keeps one symmetric penalty across paths, drawing its asymmetry from the information order alone. Kernel-gradient repulsion between concurrent members is established outside text (Liu & Wang, 2016; D'Angelo & Fortuin, 2021), and particle guidance does so in continuous diffusion through a fixed or learned potential (Corso et al., 2023). Self-repellent random walks (Doshi et al., 2023) and Stein self-repulsive dynamics (Ye et al., 2020) carry the name with a different construction, repelling one process from its own past, as does contrastive search, one sequence penalized by its own context (Su et al., 2022).

Coupling across samples also appears with the opposite sign: dVoting inherits agreed tokens (Feng et al., 2026), ThinkMerge averages $K$ traces into one (Wang et al., 2025), and others resample toward the trajectory a scorer prefers (Dang et al., 2025; Luo et al., 2026); D5P4 instead repels by selection: per diffusion step it keeps mutually dissimilar candidates, drawn from the model's denoising logits, under a partitioned determinantal point process (Lys et al., 2026; Kulesza & Taskar, 2012). A separate line measures the limits of pooling answers without acting on the sampler: vote accuracy can rise then fall in the call count (Chen et al., 2024), one analysis names decorrelation a lever though every intervention it names is peer-blind (Bay & Yearick, 2026), and models that err tend to agree on the wrong answer (Kim et al., 2025). No label-independent diversity effect exists for the 0-1 loss under any combiner (Wood et al., 2023, Thm. 10), and the classical diversity split is additive with no coverage channel (Brown & Kuncheva, 2010).

## 5 Conclusion

Self-Repulsion uses peer commitments to diversify masked-diffusion ensembles without additional model evaluations. Updating these commitments within each denoising step lets paths diverge even at temperature zero. On GSM8K, the count penalty improves plurality over self-consistency by $2.06$ points in blocks of 32 at matched budgets; ten deterministic paths reach $80.38\%$, against $70.17\%$ for the unpenalized greedy decoder (128 steps). The analyses identify higher coverage as the main source of voting gains, while showing that better coverage does not always improve the vote. **Limitations:** The evidence is limited to the LLaDA family: the main experiments use LLaDA-8B-Instruct, and LLaDA-1.5 covers GSM8K only. The exact allocation result applies locally under shared logits, not to complete trajectories. At fixed penalty strength, plurality changes little beyond $K = 10$ even as coverage rises (Appendix C.3); selecting the correct answer from the larger pool remains a limitation.

### Reproducibility statement

The method and every parameter it takes are specified in Section 2, and the full sampler is stated as Algorithm 1. Complete proofs of the propositions and the corollary in Section 2, with all assumptions stated, are given in Appendix A. The experimental setup, including the model checkpoint, the frozen per-benchmark problem indices, the prompting and answer extraction, the decoding schedule, the seed counts, and every deviation from the released decoder, is described in Appendix B, together with the baseline configurations, which run from their authors' released code at pinned commits; the metrics and their per-problem scoring are in Appendix B.1. Robustness to ensemble size, penalty strength, step count, penalty statistic, and decoding regime is reported in the ablations of Appendix C, with recommended defaults in Appendix C.7. The code that implements the sampler is included with the submission as supplementary material.

### Ethics statement

This work studies an inference-time sampling method for masked diffusion language models. It introduces no new data collection and involves no human subjects; all experiments use publicly released model checkpoints and standard public benchmarks. The method changes how a fixed model's samples are drawn and aggregated at decoding time and adds no new training signal, so it inherits the capabilities and limitations of the underlying model, including the possibility of factually incorrect or biased generations; it does not create risks beyond those of the base model. We report where the method does not improve results alongside where it does. The authors declare no conflicts of interest.

### AI use statement

We used generative AI tools as an aid for understanding background material, searching for literature and relevant known results used in our analysis, formalizing and typesetting the mathematical claims and proofs, drafting parts of the text, implementing parts of the experimental code, and creating figures. We did not use generative AI tools to develop the theoretical framework, propose hypotheses, design or give feedback on the experiments, or interpret results; synthetic data generation, dataset cleaning, translation, and qualitative data analysis are not applicable to this work. All proofs were checked by the authors, and all AI-assisted code was reviewed and tested. The authors take full responsibility for the content of this paper, including text, claims, and code.

## References

Avinash Amballa, Aditya Parashar, Aditya Vikram Singh, Jinlin Lai, and Benjamin Rozonoyer. Quasi-random multi-sample inference for large language models. In *ICLR 2025 Workshop on Frontiers in Probabilistic Inference: Learning meets Sampling*, 2025.

Marianne Arriola, Aaron Kerem Gokaslan, Justin T. Chiu, Zhihan Yang, Zhixuan Qi, Jiaqi Han, Subham Sekhar Sahoo, and Volodymyr Kuleshov. Block diffusion: Interpolating between autoregressive and diffusion language models. In *International Conference on Learning Representations*, 2025.

Yong Yi Bay and Kathleen A. Yearick. When more sampling hurts: The modal ceiling and correlation ceiling of test-time scaling. *arXiv preprint arXiv:2606.28661*, 2026.

Gavin Brown and Ludmila I. Kuncheva. "Good" and "Bad" diversity in majority vote ensembles. In *Multiple Classifier Systems*, volume 5997 of *Lecture Notes in Computer Science*, pp. 124–133, 2010.

Lingjiao Chen, Jared Quincy Davis, Boris Hanin, Peter Bailis, Ion Stoica, Matei Zaharia, and James Zou. Are more LM calls all you need? Towards the scaling properties of compound AI systems. *arXiv preprint arXiv:2403.02419*, 2024.

Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, et al. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*, 2021.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Łukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

Gabriele Corso, Yilun Xu, Valentin de Bortoli, Regina Barzilay, and Tommi Jaakkola. Particle guidance: non-I.I.D. diverse sampling with diffusion models. *arXiv preprint arXiv:2310.13102*, 2023.

Imre Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1):146–158, 1975.

Meihua Dang, Jiaqi Han, Minkai Xu, Kai Xu, Akash Srivastava, and Stefano Ermon. Inference-time scaling of diffusion language models via trajectory refinement. *arXiv preprint arXiv:2507.08390*, 2025.

Francesco D'Angelo and Vincent Fortuin. Repulsive deep ensembles are Bayesian. In *Advances in Neural Information Processing Systems*, volume 34, pp. 3451–3465, 2021.

DiffusionGemma Team. DiffusionGemma technical report. *arXiv preprint arXiv:2608.00146*, 2026.

Vishwaraj Doshi, Jie Hu, and Do Young Eun. Self-repellent random walks on general graphs - Achieving minimal sampling variance via nonlinear Markov chains. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 8403–8423, 2023.

Sicheng Feng, Zigeng Chen, Xinyin Ma, Gongfan Fang, and Xinchao Wang. dVoting: Fast voting for dLLMs. *arXiv preprint arXiv:2602.12153*, 2026.

Bennett Fox. Discrete optimization via marginal analysis. *Management Science*, 13(3):210–216, 1966.

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 *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1, 2021.

Inception Labs, Samar Khanna, Siddhant Kharbanda, Shufan Li, Harshit Varma, Eric Wang, Sawyer Birnbaum, Ziyang Luo, Yanis Miraoui, Akash Palrecha, Stefano Ermon, Aditya Grover, and Volodymyr Kuleshov. Mercury: Ultra-fast language models based on diffusion. *arXiv preprint arXiv:2506.17298*, 2025.

Elliot Kim, Avi Garg, Kenny Peng, and Nikhil Garg. Correlated errors in large language models. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, pp. 30038–30066, 2025.

Wouter Kool, Herke van Hoof, and Max Welling. Stochastic beams and where to find them: The Gumbel-Top-$k$ trick for sampling sequences without replacement. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pp. 3499–3508, 2019.

Alex Kulesza and Ben Taskar. Determinantal point processes for machine learning. *arXiv preprint arXiv:1207.6083*, 2012.

Sean Lamont, Christian Walder, Paul Montague, Amir Dezfouli, and Michael Norrish. Free lunch for pass@$k$? Low cost diverse sampling for diffusion language models. *arXiv preprint arXiv:2603.04893*, 2026.

Jihoon Lee, Hoyeon Moon, Kevin Zhai, Arun Kumar Chithanar, Anit Kumar Sahu, Soummya Kar, Chul Lee, Souradip Chakraborty, and Amrit Singh Bedi. Test-time scaling in diffusion LLMs via hidden semi-autoregressive experts. *arXiv preprint arXiv:2510.05040*, 2025.

Michael Y. Li, Anthony Zhan, Kanishk Gandhi, Noah D. Goodman, and Emily B. Fox. QuasiMoTTo: Quasi-Monte Carlo test-time scaling. *arXiv preprint arXiv:2607.01179*, 2026.

Tianyi Li, Mingda Chen, Bowei Guo, and Zhiqiang Shen. A survey on diffusion language models. *arXiv preprint arXiv:2508.10875*, 2025.

Stephanie Lin, Jacob Hilton, and Owain Evans. TruthfulQA: Measuring how models mimic human falsehoods. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 3214–3252, 2022.

Qiang Liu and Dilin Wang. Stein variational gradient descent: A general purpose Bayesian inference algorithm. *arXiv preprint arXiv:1608.04471*, 2016.

Ziwei Luo, Ziqi Jin, Lei Wang, Lidong Bing, and Thomas B. Schön. Self-rewarding sequential Monte Carlo for masked diffusion language models. *arXiv preprint arXiv:2602.01849*, 2026.

Jonathan Lys, Vincent Gripon, Axel Marmoret, Lukas Mauch, Fabien Cardinaux, Ghouthi Boukli Hacene, and Bastien Pasdeloup. D5P4: Partition determinantal point process for diversity in parallel discrete diffusion decoding. *arXiv preprint arXiv:2603.19146*, 2026.

André F. T. Martins and Ramón F. Astudillo. From softmax to sparsemax: A sparse model of attention and multi-label classification. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48, pp. 1614–1623, 2016.

Clara Meister, Martina Forster, and Ryan Cotterell. Determinantal beam search. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 6551–6562, 2021.

Roger B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer, 2nd edition, 2006.

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. In *Advances in Neural Information Processing Systems*, volume 38, pp. 50608–50646, 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. *arXiv preprint arXiv:2604.09921*, 2026.

Ji Won Park and Kyunghyun Cho. Efficient semantic uncertainty quantification in language models via diversity-steered sampling. *arXiv preprint arXiv:2510.21310*, 2025.

Kyeongman Park, Minha Jhang, and Kyomin Jung. A universal avoidance method for diverse multi-branch generation. *arXiv preprint arXiv:2604.17323*, 2026.

Weijie Shi, Yue Cui, Yaguang Wu, Jingzhi Fang, Shibo Zhang, Mengze Li, Sirui Han, Jia Zhu, Jiajie Xu, and Xiaofang Zhou. Semantic-guided diverse decoding for large language model. In *Advances in Neural Information Processing Systems*, volume 38, pp. 148182–148229, 2025.

Yuxuan Song, Zheng Zhang, Cheng Luo, Pengyang Gao, Fan Xia, Hao Luo, Zheng Li, Yuehang Yang, Hongli Yu, Xingwei Qu, Yuwei Fu, Jing Su, Ge Zhang, Wenhao Huang, Mingxuan Wang, Lin Yan, Xiaoying Jia, Jingjing Liu, Wei-Ying Ma, Ya-Qin Zhang, Yonghui Wu, and Hao Zhou. Seed Diffusion: A large-scale diffusion language model with high-speed inference. *arXiv preprint arXiv:2508.02193*, 2025.

Yixuan Su, Tian Lan, Yan Wang, Dani Yogatama, Lingpeng Kong, and Nigel Collier. A contrastive framework for neural text generation. In *Advances in Neural Information Processing Systems*, volume 35, pp. 21548–21561, 2022.

Ashwin K. Vijayakumar, Michael Cogswell, Ramprasaath R. Selvaraju, Qing Sun, Stefan Lee, David Crandall, and Dhruv Batra. Diverse beam search: Decoding diverse solutions from neural sequence models. *arXiv preprint arXiv:1610.02424*, 2016.

Luke Vilnis, Yury Zemlyanskiy, Patrick Murray, Alexandre Passos, and Sumit Sanghai. Arithmetic sampling: Parallel diverse decoding for large language models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 35120–35136, 2023.

Haonan Wang, Chao Du, Kenji Kawaguchi, and Tianyu Pang. Think in parallel, answer as one: Logit averaging for open-ended reasoning. *arXiv preprint arXiv:2512.02874*, 2025.

Xuezhi Wang and Denny Zhou. Chain-of-thought reasoning without prompting. *arXiv preprint arXiv:2402.10200*, 2024.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. In *International Conference on Learning Representations*, 2023.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 24824–24837, 2022.

Danny Wood, Tingting Mu, Andrew M. Webb, Henry W. J. Reeve, Mikel Luján, and Gavin Brown. A unified theory of diversity in ensemble learning. *Journal of Machine Learning Research*, 24 (359):1–49, 2023.

Jingxuan Wu, Zhenglin Wan, Xingrui Yu, Yuzhe Yang, Yiqiao Huang, Ivor Tsang, and Yang You. Time-annealed perturbation sampling: Diverse generation for diffusion language models. *arXiv preprint arXiv:2601.22629*, 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.

Mao Ye, Tongzheng Ren, and Qiang Liu. Stein self-repulsive dynamics: Benefits from past samples. In *Advances in Neural Information Processing Systems*, volume 33, pp. 241–252, 2020.

Jingwei Zhang, Haoyu Lei, Zijin Feng, Jiacheng Sun, and Farzan Farnia. Exploring more to solve more: Boosting diversity in text diffusion models via entropy-based guidance. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306, 2026.

Fengqi Zhu, Rongzhen Wang, Shen Nie, Xiaolu Zhang, Chunwei Wu, Jun Hu, Jun Zhou, Jianfei Chen, Yankai Lin, Ji-Rong Wen, and Chongxuan Li. LLaDA 1.5: Variance-reduced preference optimization for large language diffusion models. *arXiv preprint arXiv:2505.19223*, 2025.

# A Proofs and Theoretical Details

## A.1 Ensembles without randomness

*Proof of Proposition 1.* Write the canvases at the start of a step as $u = (u_1, \dots, u_K)$ and the step itself as $u_k \mapsto G(u_k, u_{-k})$, one function $G$ shared by all paths, where $u_{-k}$ is the tuple of path $k$'s peers in a fixed indexing. A statistic that reads the peers as a multiset, or all $K$ canvases symmetrically, is the special case where $G$ is symmetric in $u_{-k}$.

*One step preserves equality.* Suppose the $u_k$ all equal one canvas $w$. Every path then evaluates $G$ at the same argument $(w, (w, \dots, w))$, and each part of the step returns the same result across paths. The model reads the canvas alone and returns the same logits. The penalty is a function of the same argument and is the same vector. The argmax under a path-index-free tie-break writes the same token. The commit rule, one path-index-free function of the canvas, its raw logits and the written tokens, selects the same positions. So the canvases stay equal after the step.

*Induction.* The paths start from a common initial canvas. One step preserves equality, so induction over steps keeps the canvases identical and the pool holds one distinct sample. □

**Remark A.1** (Order-invariance of the cascade)**.** At temperature zero, for a processing order held fixed across steps, $\mathcal{C}_\pi = \mathcal{C}_{\pi'}$ for all processing orders $\pi, \pi'$, where $\mathcal{C}_\pi$ is the pool the cascade produces under order $\pi$. By induction over the pairs (denoising step, cascade position), the path at each pair reads its own canvas and a symmetric count of canvases written at earlier pairs, so the trajectory realized at cascade position $i$ does not depend on the order, which only assigns members to indices; at positive temperature the same induction gives equality in law. The deployed cascade lets each path read every peer; a triangular variant reads only predecessors.

## A.2 Allocation at a shared position

*Proof of Proposition 2. Concave objective.* Under the hypotheses all paths committing the position face the same logits $\ell$. The $i$-th path to take token $v$ gains $\ell_v - g\,(i-1)$, which decreases in $i$. A count vector $c$ then realizes total gain

$$\sum_v c_v \ell_v - g \sum_v \binom{c_v}{2},$$

a sum over tokens of concave functions of the counts. The cascade takes the largest current gain $\ell_v - g\,c_v$ at each turn, so it greedily maximizes this objective.

*Greedy exactness.* A separable concave objective on the integer simplex is maximized greedily (Fox, 1966). The gains the cascade takes are non-increasing, so every gain it takes is at least $\theta = \max_u(\ell_u - g\,c_u)$, the largest gain still available at the end. Take any feasible $c'$ and move one path from a token with $c'_v > c_v$, whose last gain is at most $\theta$, to a token with $c'_u < c_u$, whose next gain is at least $\theta$. The move does not lower the objective. Repeating it until $c' = c$ shows $c$ optimal.

*Water level.* The common threshold is $\theta = \max_u(\ell_u - g\,c_u)$. For $g > 0$, every token receiving a path has

$$(\ell_v - \theta)/g \;\le\; c_v \;\le\; (\ell_v - \theta)/g + 1,$$

while every token receiving none has $\ell_v \le \theta$. The expected-count and collision-weighted statistics charge tokens no peer has committed, so this allocation does not describe them. □

## A.3 Allocation as a sparsemax

*Fractional maximizer.* On the constraint $\sum_v c_v = K$ we have $\sum_v \binom{c_v}{2} = \frac{1}{2}\|c\|^2 - \frac{K}{2}$, so the objective of Proposition 2 for the count statistic is

$$-\tfrac{g}{2}\,\|c - \ell/g\|^2$$

plus terms independent of $c$. Over fractional allocations its maximizer is the Euclidean projection of $\ell/g$ onto the scaled simplex. After the substitution $c = Kp$ this projection is $K\,\mathrm{sparsemax}(\ell/gK)$ (Martins & Astudillo, 2016), with $gK = \lambda K/(K-1)$ as the sparsemax temperature.

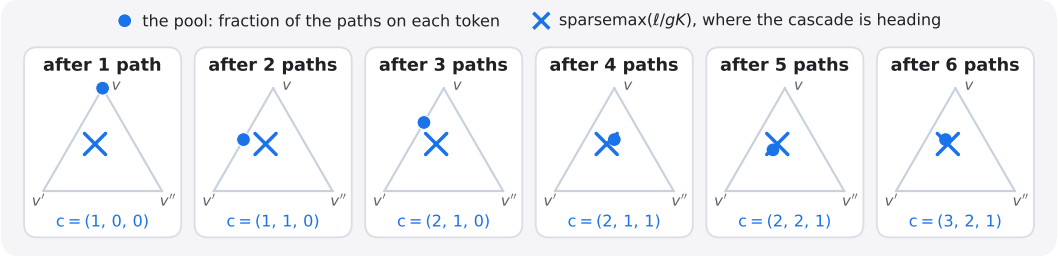


Figure 3: One position on the probability simplex, three of its tokens. Each frame shows the pool after $i$ paths, the fraction of the paths on each token, as a point of the triangle, against $\mathrm{sparsemax}(\ell/gK)$, marked by a cross. The cascade starts the pool at the mode's vertex and walks it, one path at a time, onto the cross: after six paths the pool is $(3, 2, 1)/6$, the nearest point of the $K$-grid to $\mathrm{sparsemax}(\ell/gK) = (0.46, 0.34, 0.21)$ (Proposition 2).

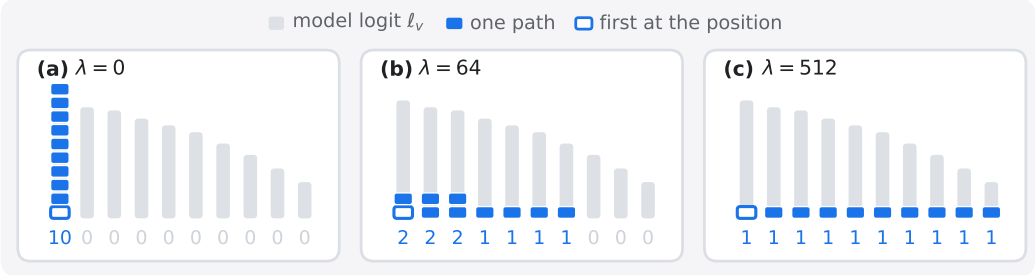


Figure 4: What the $K$ paths commit at one position at $\tau = 0$: with the penalty off **(a)**, at the deployed strength **(b)** and at a large one **(c)**. Each grey bar is a token's model logit $\ell_v$ and each blue cell is one path committed there, stacked in commit order; the outlined cell is the first path to commit, which faces no peer and pays no penalty. With the penalty off all $K$ paths tower on the mode; raising it spreads them, until at large strength each of the top $K$ tokens holds one path. Logits are illustrative and the allocation is computed by the update rule; the water level is the common threshold a token's logit must clear to receive a path.

*Integer counts.* Let $\theta = \max_u(\ell_u - gc_u)$ be the cascade's final level and set $y_v(t) = \max(0,\ (\ell_v - t)/g + \frac{1}{2})$. The end-state conditions of the proof above read $|c_v - y_v(\theta)| \le \frac{1}{2}$ for every $v$, while the projection is $y_v(\theta')$ at the level $\theta'$ where $\sum_v y_v(\theta') = K$. Every positive $y_v$ has slope $-1/g$ in $t$, so a level difference above $g/2$ would carry $\sum_v y_v$ strictly past $K$. Hence $|\theta - \theta'| \le g/2$, and

$$\max_v \left| c_v - K\,\mathrm{sparsemax}_v(\ell/gK) \right| \;<\; 1,$$

the equality excluded because $|\theta - \theta'| = g/2$ forces the two allocations to coincide. Figure 3 shows the six commits as a walk on the probability simplex toward the sparsemax point.

### A.4 Tilt

Subtracting $\tilde{\lambda}\Delta_v$ from the log-probabilities multiplies $p(v)$ by $e^{-\tilde{\lambda}\Delta_v}$ and renormalizes. Where the mass goes is immediate. The ratio is

$$\frac{q(v)}{p(v)} \;=\; \frac{e^{-\tilde{\lambda}\Delta_v}}{Z}, \qquad Z = \mathbb{E}_p\left[e^{-\tilde{\lambda}\Delta}\right],$$

an average of the factors over $\mathrm{supp}\,p$, so for $\tilde{\lambda} > 0$ a token gains mass, $q(v) \ge p(v)$, exactly when $\Delta_v \le -\log Z/\tilde{\lambda}$. For the count statistic under a full-support $p$, once some peer has committed and the support is larger than $K - 1$, $\min \Delta = 0 < \max \Delta$ and $Z$ lies strictly between the extreme

factors. So every untaken token gains mass and every token of the largest count loses it, the movement Figure 1 draws against temperature, which reshapes every token instead.

**Ratio preservation.** If $\Delta_u = \Delta_v$ the tilt factors cancel in the ratio $q(u)/q(v)$, which stays $p(u)/p(v)$ at every $\lambda$. Temperature instead raises every ratio to the power $\tau/\tau'$ and so distorts all but the unit ratio.

### A.5 Minimum-divergence characterization of the tilt

**Lemma A.2** (I-projection)**.** *For $\tilde{\lambda} \geq 0$ and $m = \mathbb{E}_q[\Delta]$, the law in equation (1) is the unique solution of* $\min\{\mathrm{KL}(q' \,\|\, p) : \mathbb{E}_{q'}[\Delta] \leq m\}$.

*Proof of Lemma A.2.* For any law $q'$ on $\mathcal{V}$, expand $q$ inside the divergence:

$$\mathrm{KL}(q' \,\|\, q) \;=\; \sum_v q'(v) \log \frac{q'(v)}{p(v)\, e^{-\tilde{\lambda}\Delta_v}/Z} \;=\; \mathrm{KL}(q' \,\|\, p) \;+\; \tilde{\lambda}\, \mathbb{E}_{q'}[\Delta] \;+\; \log Z.$$

Rearranged, this reads $\mathrm{KL}(q' \,\|\, p) + \tilde{\lambda}\, \mathbb{E}_{q'}[\Delta] \geq -\log Z$, with equality if and only if $q' = q$. The tilt $q$ is feasible, since $\mathbb{E}_q[\Delta] = m$. For any feasible $q'$, using $\tilde{\lambda} \geq 0$ at the second step,

$$\mathrm{KL}(q' \,\|\, p) \;\geq\; -\log Z - \tilde{\lambda}\, \mathbb{E}_{q'}[\Delta] \;\geq\; -\log Z - \tilde{\lambda} m \;=\; \mathrm{KL}(q \,\|\, p)\,,$$

the last equality being the line above evaluated at $q' = q$. Equality throughout forces $\mathrm{KL}(q' \,\|\, q) = 0$, hence $q' = q$. The minimizer therefore exists and is unique. □

The constraint is one-sided because $\tilde{\lambda} \geq 0$: the tilt is closest to $p$ among laws with expected agreement *at most* $m$, a negative strength bounding the agreement below instead. Fixing $\mathbb{E}[\Delta]$ exactly leaves the same minimizer, the I-projection onto a linear family.

### A.6 Coupling ceiling on coverage

*Proof of Proposition 4.* Fix the per-sample marginals $p_1, \ldots, p_K$ for the event that a sample produces the correct answer. Coverage over all couplings lies in $[\max_k p_k,\ \min(1, \sum_k p_k)]$, the Fréchet–Hoeffding bounds (Nelsen, 2006, Thms. 2.10.12–13) applied to the complement events, while independence gives $1 - \prod_k (1 - p_k)$. The excess of the upper bound over independence is

$$G(p) \;=\; \min\Big(1, \textstyle\sum_k p_k\Big) \;-\; \Big[1 - \prod_k (1 - p_k)\Big].$$

Write $s = \sum_k p_k$. For $s \leq 1$ the arithmetic-geometric mean inequality gives $G \leq s - 1 + (1 - s/K)^K$, whose derivative $1 - (1 - s/K)^{K-1}$ is non-negative, so the maximum is at $s = 1$. For $s > 1$, $G = \prod_k (1 - p_k) \leq (1 - s/K)^K < (1 - 1/K)^K$. Either way $G \leq (1 - 1/K)^K$, attained only at $p_k \equiv 1/K$. The value is below $e^{-1}$ for every finite $K$ and equals $0.3487$ at $K = 10$; the per-problem excess $(1 - p)^K$ is of order $10^{-3}$ at $p = 0.5$. □

## B Experimental details

### B.1 Metrics

Each metric is computed per problem, averaged over seeds then problems, from the answers $a_1, \ldots, a_K$ extracted from the $K$ samples, where extraction may fail. *Plurality accuracy* scores the vote under a uniform tie-break, $\mathbf{1}[a^\star \in M]/|M|$ for $M$ the top-count set among parsed answers, rather than an order-dependent break. *Coverage* is $\mathbf{1}[\exists k : a_k = a^\star]$, an upper bound on any aggregator over the pool; it saturates on small answer sets (four options give $1 - 0.75^{10} = 94.37\%$ by chance), so we omit it on TruthfulQA, where the count row's coverage sits at that ceiling (Table 15). *Per-sample accuracy* is the fraction of correct samples, an unparsed sample counted wrong; the unparsed rate reaches 17.4% on MATH and the two orderings disagree there under pure diffusion, so this counted-wrong denominator, not the parsed-only one, is the one the tables use. *Disagreement* is the rate at which two samples of a pool give different parsed answers, a diversity measure and not an objective.

### B.2 Model and computation

We run LLaDA-8B-Instruct (`GSAI-ML/LLaDA-8B-Instruct`) with frozen weights in BF16, $K = 10$ samples of generation length 256, so a problem costs $KT$ network function evaluations and no method uses more. Table 2 runs $T = 128$ (1280 evaluations); Tables 1, 5, 13 and 14 run $T = 256$ (2560); the sequential variant spends 1280. The temperature-zero rows of Tables 2 and 12 run at $T = 128$, those of Tables 13 and 14 at $T = 256$. Decoding is semi-autoregressive in blocks of 32 with the released low-confidence remasking, positions ranked by the untilted confidence of the sampled token; the pure-diffusion campaigns instead decode the whole generation as a single block. The pure-diffusion rows of Table 2 carry the released end-of-sequence confidence correction (Appendix D.9), while the 2560-evaluation campaigns of Tables 1 and 5 predate it and every TAPS row cannot take it, as their captions state. Our temperature 0.6 and length 256 deviate from the released greedy decoder, which uses length and step count 512.

### B.3 Selection, seeds and reporting

GSM8K (Cobbe et al., 2021) uses a frozen 600-problem index into the GSM8K test split (`test.jsonl`, 1319 items), MATH (Hendrycks et al., 2021) a 492-problem index into MATH-500 with the eight items whose reference answers do not parse removed, and TruthfulQA (Lin et al., 2022) a 684-problem index into the four-option items of the multiple-choice set (`truthful_qa_mc`), each item tagged by its SHA-256 prefix in every output. GSM8K is prompted zero-shot with a step-by-step instruction and the last number in the generation is taken as the answer; MATH and TruthfulQA place the final answer in a `\boxed{}` field inside a reasoning-then-answer template, extracted as the boxed expression under Hendrycks normalization (MATH) or the boxed option letter (TruthfulQA), each with a bare-answer fallback. Every within-campaign contrast is paired within one decoding run and not pooled across runs, which removes the half-precision kernel noise that compounds over the denoising steps; the few cross-campaign comparisons (the end-of-sequence correction of Appendix D.9, the block-versus-canvas baselines) are marked as such where they appear. Configurations carry two to five seeds as stated per table; temperature-zero configurations consume no randomness and have no seed axis, so a paired sampled configuration carries the seed variance. A reported result is a mean over at least 300 problems and two seeds; configurations at 200 problems or one seed are given in direction only and carry no table. The per-stratum diagnostics of Figure 6 report finer strata, some below this floor, as diagnostics and not results. A two-configuration comparison is a mean paired difference over the shared problems with seeds pooled; Appendix C.1 gives the single-configuration seed spread.

### B.4 Baselines

All baseline numbers are our reproductions under this protocol, each family swept over its own temperature, ties broken uniformly. *ODD* and *TAPS* run from their authors' released code at pinned commits.[1] ODD runs its own strategy and feature extractor inside our loop, at temperature 0.6 in the shared-schedule tables (Tables 2 and 5) and at 1.0, a temperature its authors report, in the own-temperature comparison (Table 12), with step sizes $\alpha \in \{64, 256\}$ from our sweep; TAPS runs its embedding-level conditioning perturbation, cosine-annealed Gaussian noise of scale 0.2 on the prompt embeddings, swept over $\{0.6, 1.0, 1.5\}$. Sampling hygiene is left as each author wrote it, so an unparsed sample is counted wrong rather than repaired.

## C Ablations

We vary the decoding parameters to assess the method's sensitivity. Ensemble size and decoding regime have the largest effects on plurality accuracy; the remaining parameters have smaller effects over the ranges tested. Unless stated otherwise, each sweep varies one parameter while holding the others at the reference configuration of Appendix B: ten samples, 128 steps, blocks of 32, $\tau = 0.6$, and the count statistic at strength 64, evaluated on GSM8K ($n = 600$) at 1280 model evaluations. Comparisons use the problems common to every row and seed, with metrics defined in Appendix B.1.

[1] ODD (`https://github.com/sean-lamont/odd`) and TAPS (`https://github.com/Johnny221B/TAPS`), each vendored at a pinned commit.

The ensemble-size sweep (Table 6) and the penalty-by-regime comparison (Table 8) use 256 steps instead. The ensemble-size sweep also includes a fixed-gate comparison at $g = 8.00$, alongside the fixed-strength comparison at $\lambda = 64$.

### C.1 Variation across decoding seeds

We measure variation across five decoding seeds while holding the evaluation problems and decoding configuration fixed. At 1280 evaluations, SR has mean plurality accuracy 79.11%, with a standard deviation of 1.08 percentage points and a range of 2.87 points. At 2560 evaluations, its standard deviation and range are 0.23 and 0.50 points, respectively. Self-consistency at 1280 evaluations has a standard deviation of 0.40 points and a range of 1.04 points. In these runs, reducing the step count from 256 to 128 increases SR's standard deviation by a factor of 4.7 and its range by a factor of 5.7. Per-sample accuracy varies less across seeds than plurality accuracy in every configuration.

For the comparison with self-consistency, we report the paired difference for each seed at the same evaluation budget. At 1280 evaluations, these differences are 3.51, 1.68, 1.74, 1.68, and 1.69 percentage points, giving a mean improvement of 2.06 points. The improvement is positive for all five seeds. These summaries describe sensitivity to sampling randomness on the fixed evaluation set; the seed standard deviations and ranges are not significance thresholds.

Table 4: Effect of the decode seed. Each configuration is held fixed, block 32, GSM8K, $n = 600$; only the seed varies. The reference configuration and self-consistency are at 1280 evaluations; the third configuration is the same method at 2560. SD and range are taken over the five seeds.

| Configuration | Metric | Seed 0 | Seed 1 | Seed 2 | Seed 3 | Seed 4 | Mean | SD | Range |
|---|---|---|---|---|---|---|---|---|---|
| SR, 1280 | Plurality | 80.93 | 78.89 | 78.79 | 78.06 | 78.89 | 79.11 | 1.08 | 2.87 |
| | Coverage | 94.33 | 92.17 | 94.00 | 93.50 | 93.67 | 93.53 | 0.83 | 2.17 |
| | Per-sample | 67.68 | 68.12 | 68.72 | 68.00 | 68.52 | 68.21 | 0.41 | 1.03 |
| Self-consistency, 1280 | Plurality | 77.42 | 77.21 | 77.06 | 76.38 | 77.20 | 77.05 | 0.40 | 1.04 |
| | Coverage | 89.50 | 89.67 | 90.67 | 89.50 | 89.17 | 89.70 | 0.57 | 1.50 |
| | Per-sample | 70.03 | 69.65 | 70.57 | 70.02 | 69.92 | 70.04 | 0.33 | 0.92 |
| SR, 2560 | Plurality | 78.82 | 79.22 | 78.76 | 78.72 | 79.14 | 78.93 | 0.23 | 0.50 |
| | Coverage | 94.33 | 93.33 | 94.00 | 92.00 | 93.00 | 93.33 | 0.91 | 2.33 |
| | Per-sample | 71.13 | 70.73 | 70.53 | 70.82 | 70.77 | 70.80 | 0.22 | 0.60 |

### C.2 Main comparison at 2560 evaluations

Table 5 repeats the temperature-0.6 rows of Table 2 at twice the computation, method by method, in both regimes; pure diffusion is the mechanism's measurement ground, not the operating point (Appendix C.6).

### C.3 Ensemble size

Ensemble size is the parameter with the largest effect. The sample count enters the budget directly and the penalty through its normalization, so a sweep over it is not budget-matched; the strength moves with it.

The vote rises to ten samples and is then flat at fixed $\lambda$: 78.93, 79.40, 79.38 at $K = 10, 20, 32$, while coverage keeps rising, 93.33, 95.50, 96.33, and per-sample accuracy holds, 70.80 to 71.04. At fixed gate the vote reads 80.29 and 81.18 at $K = 20, 32$, but per-sample accuracy falls 70.80 to 67.19 to 61.86 and disagreement climbs 30.85 to 48.25, the temperature signature of Table 1; that rise is more spreading and cannot be read as $K$-scaling. Self-consistency's vote is flat from ten onward while its coverage rises, so the vote is the pool's bottleneck, not the penalty's. Below ten the vote climbs, 71.47 to 78.93 from two samples to ten and coverage 79.72 to 93.33; five samples at 1280 evaluations read 76.88, above self-consistency at ten and twice the budget. We hold $\lambda$ across $K$.

Table 5: Both decoding regimes under the protocol of Section 3 ($K = 10$, 2560 NFEs, temperature 0.6, two to five seeds per configuration, five on the GSM8K SR, count row and two or three on the rest); TruthfulQA coverage is unreported, since guessing over four options already reaches 94.37. Bold: best per column within a regime, disagreement and the $K$=1 greedy reference row excluded. The pure-diffusion rows predate the end-of-sequence correction applied in Table 2 (Appendix D.9). Baselines are our reproductions (Appendix B).

| | Plurality (↑) | | | Coverage (↑) | | Per-sample (↑) | | | Disagreement | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Method | GSM | MATH | TQA | GSM | MATH | GSM | MATH | TQA | GSM | MATH | TQA |
| *Semi-autoregressive blocks of 32* | | | | | | | | | | | |
| Single greedy decode ($K$=1) | 73.00 | – | – | 73.00 | – | 73.00 | – | – | 0.00 | – | – |
| Self-consistency | 76.09 | 38.18 | 59.84 | 87.94 | 55.89 | **72.31** | **31.60** | 56.37 | 19.93 | 55.39 | 28.47 |
| TAPS | 75.37 | 37.54 | 58.46 | 90.17 | 55.39 | 68.88 | 30.26 | 55.42 | 29.36 | 57.89 | 30.80 |
| ODD ($\alpha$=64) | 77.75 | 37.49 | 59.31 | 91.58 | 56.20 | 69.14 | 26.88 | 54.01 | 31.00 | 66.43 | 37.55 |
| ODD ($\alpha$=256) | 78.24 | 39.37 | 59.93 | 92.33 | 54.78 | 69.52 | 27.71 | 54.20 | 31.10 | 63.11 | 38.05 |
| SR, count | 78.93 | 40.08 | 61.33 | 93.33 | **60.47** | 70.80 | 30.22 | **56.53** | 30.85 | 64.25 | 34.17 |
| SR, exp. count | **79.89** | 39.79 | 60.01 | 94.00 | 58.43 | 69.81 | 29.77 | 52.64 | 33.49 | 64.04 | 40.76 |
| SR, collision-wt. | 79.20 | **40.16** | **61.99** | **94.67** | 58.74 | 67.01 | 28.25 | 54.62 | 38.75 | 67.49 | 38.14 |
| *Pure diffusion (whole response as one block)* | | | | | | | | | | | |
| Self-consistency | 54.53 | 25.98 | 56.14 | 63.17 | 36.08 | 53.42 | 23.77 | 50.02 | 15.02 | 34.27 | 22.20 |
| TAPS | 56.77 | 27.13 | 56.20 | 69.92 | 40.96 | 52.91 | 23.69 | **55.87** | 28.69 | 41.91 | 13.02 |
| SR, count | 73.09 | 31.58 | 56.79 | 86.83 | 51.52 | 56.19 | 23.37 | 43.65 | 46.60 | 65.23 | 56.76 |
| SR, exp. count | **76.87** | 34.44 | 56.60 | 90.67 | 53.96 | 62.52 | 25.10 | 51.46 | 40.15 | 64.06 | 37.88 |
| SR, collision-wt. | 76.85 | **37.85** | **58.74** | **91.25** | **55.79** | **63.06** | **27.79** | 52.57 | 40.91 | 66.64 | 38.39 |

Table 6: Effect of ensemble size ($K$). Fixed: 256 steps, block 32, temperature 0.6, count statistic, GSM8K, $n = 600$. Not budget-matched: NFE $= K\times$ steps and is the price of each row. Rows below ten hold the gate at 8.00; the two conventions meet at $K = 9$ ($\lambda = 64$). Rows above ten hold $\lambda = 64$, the $K$-invariant scaling of Corollary 3, with the fixed-gate rows as the diagnostic; self-consistency rows above ten pool seeds, which their i.i.d. draws license. The $\lambda = 0$ rows are self-consistency, the same loop with the penalty off; a dash is not measured.

| $K$ | $\lambda$ | $g$ | NFE | $n$ | Seeds | Plurality | Coverage | Per-sample | Disagr. |
|---|---|---|---|---|---|---|---|---|---|
| *Fixed gate $g = 8.00$* | | | | | | | | | |
| 2 | 8 | 8.00 | 512 | 600 | 3 | 71.47 | 79.72 | 71.47 | 25.39 |
| 3 | 16 | 8.00 | 768 | 600 | 3 | 74.44 | 84.39 | 72.00 | 26.07 |
| 5 | 32 | 8.00 | 1280 | 600 | 3 | 76.88 | 88.78 | 71.39 | 28.03 |
| 7 | 48 | 8.00 | 1792 | 600 | 3 | 77.57 | 91.50 | 71.48 | 28.55 |
| *Fixed strength $\lambda = 64$, the $K$-invariant scaling of Corollary 3* | | | | | | | | | |
| 10 | 64 | 7.11 | 2560 | 600 | 5 | 78.93 | 93.33 | 70.80 | 30.85 |
| 20 | 64 | 3.37 | 5120 | 600 | 2 | 79.40 | 95.50 | 70.93 | 29.78 |
| 32 | 64 | 2.06 | 8192 | 600 | 2 | 79.38 | 96.33 | 71.04 | 29.50 |
| *Fixed gate $g = 7.11$ (diagnostic; per-sample falls, disagreement climbs)* | | | | | | | | | |
| 20 | 135 | 7.11 | 5120 | 600 | 2 | 80.29 | 96.58 | 67.19 | 38.48 |
| 32 | 220 | 7.11 | 8192 | 600 | 2 | 81.18 | 97.75 | 61.86 | 48.25 |
| *Self-consistency ($\lambda = 0$); rows above ten pool seeds* | | | | | | | | | |
| 10 | 0 | 0.00 | 2560 | 600 | 3 | 76.09 | 87.94 | 72.31 | 19.93 |
| 20 | 0 | 0.00 | 5120 | 600 | 5 | 76.37 | 89.81 | – | – |
| 32 | 0 | 0.00 | 8192 | 600 | 5 | 76.48 | 90.85 | – | – |
| 50 | 0 | 0.00 | 12800 | 600 | 5 | 76.83 | 91.67 | – | – |

### C.4 SMALL-EFFECT PARAMETERS

The remaining parameters move the vote least; Table 7 collects the two closest, the samples-versus-steps split of a fixed budget and the choice of statistic, as deviations from the shared reference. Sample count dominates step count at a fixed 1280-evaluation budget: ten samples of 128 steps

read 79.11 against 76.88 for five of 256, coverage 93.53 against 88.78, per-sample the other way. The two distribution-reading variants replace the count by the peers' predicted distributions $\bar{p}[j, \cdot]$: *expected-count* uses the expected number of peers on each token under $\bar{p}$, and *collision-weighted* scales that by the collision probability between the path's own law and the peers' mean. Among the statistics, expected-count reads highest at 80.65 and collision-weighted 80.07 against the count's 79.11; those gaps of 1.54 and 0.96 reverse across benchmarks, so we deploy the count everywhere.

Table 7: The small-effect parameters, as deviations from the reference configuration (GSM8K, $n = 600$, $K = 10$, 128 steps, block 32, $\tau = 0.6$, count statistic, 1280 network function evaluations). The reference and self-consistency rows are shared by both blocks; the statistic rows move only the statistic, while the split row re-allocates the fixed budget to five samples of 256 steps and so also lowers the strength to $\lambda = 32$ (gate 8.00). Seeds as stated; higher is better in the accuracy columns; disagreement is a rate, not an objective.

| Parameter | Setting | Seeds | Plurality | Coverage | Per-sample | Disagr. |
|---|---|---|---|---|---|---|
| Reference | count, $\lambda$=64, $g$=7.11 | 5 | 79.11 | 93.53 | 68.21 | 35.53 |
| Self-consistency | $\lambda$=0, $g$=0 | 5 | 77.05 | 89.70 | 70.04 | 27.14 |
| Split at 1280 NFE | 5 samples $\times$ 256 steps, $\lambda$=32 | 3 | 76.88 | 88.78 | 71.39 | 28.03 |
| Statistic | expected count | 2 | 80.65 | 94.67 | 66.47 | 40.60 |
| Statistic | collision-weighted | 2 | 80.07 | 94.42 | 61.87 | 48.63 |

The vote is flat in the denoising step count, which carries no table of its own: doubling the step budget from 1280 to 2560 evaluations moves the vote 0.18 over five seeds against the 2.06 the method gains over self-consistency.

### C.5 Penalty by decoding regime

On per-sample accuracy the ordering of the count and collision-weighted statistics reverses between the two regimes on all three benchmarks (Table 8), so neither distribution-aware statistic transfers between regimes.

Table 8: Penalty by decoding regime on per-sample accuracy, paired per problem and seed-averaged over seeds $\{0, 1\}$, at 2560 network function evaluations. Each entry is the named penalty minus the count statistic in that regime, positive favouring the named penalty; the interaction is the pure-diffusion effect minus the block-32 effect, taken within problem. These within-problem paired contrasts on two seeds do not reproduce the seed-mean differences of Table 5.

| | Expected count − count | | | Collision-weighted − count | | |
|---|---|---|---|---|---|---|
| Benchmark | Diffusion | Block 32 | Interaction | Diffusion | Block 32 | Interaction |
| GSM8K ($n = 600$) | $+6.33$ | $-1.53$ | $+7.86$ | $+6.87$ | $-4.23$ | $+11.10$ |
| MATH ($n = 492$) | $+1.73$ | $-0.46$ | $+2.18$ | $+4.42$ | $-1.97$ | $+6.39$ |
| TruthfulQA ($n = 684$) | $+7.81$ | $-3.70$ | $+11.51$ | $+8.91$ | $-1.45$ | $+10.36$ |

### C.6 Decoding regime

Pure diffusion decodes the whole response as one block and is LLaDA's default decoding; for the Instruct model, LLaDA reports blocks of 32 as stronger on GSM8K and MATH (Nie et al., 2025). The penalty reads $+2.06$ over self-consistency in blocks and $+15.91$ under pure diffusion, as differences of the marginals below (these pure-diffusion rows predate the end-of-sequence correction; with it the gain is $+14.50$, Table 2); the second regime is where the pool is most redundant, so it is where the mechanism has the most to convert.

Table 9: Effect of the decoding regime. Fixed: $K = 10$, 128 steps, temperature 0.6, count statistic, GSM8K, $n = 600$. Both pure-diffusion rows predate the released end-of-sequence correction; Table 2 carries the corrected rows and Appendix D.9 reports what the correction is worth. Per-sample accuracy under the parsed-answer denominator is 0.41 higher on both pure-diffusion rows.

| Regime | $\lambda$ | $g$ | $n$ | Seeds | Plurality | Coverage | Per-sample | Disagr. |
|---|---|---|---|---|---|---|---|---|
| Blocks of 32 | 64 | 7.11 | 600 | 5 | 79.11 | 93.53 | 68.21 | 35.53 |
| Blocks of 32 | 0 | 0.00 | 600 | 5 | 77.05 | 89.70 | 70.04 | 27.14 |
| Pure diffusion | 64 | 7.11 | 600 | 2 | 70.30 | 87.25 | 52.58 | 51.18 |
| Pure diffusion | 0 | 0.00 | 600 | 2 | 54.39 | 65.83 | 52.02 | 21.08 |

### C.7 Recommended defaults

Table 10: Recommended settings from the sweeps above. The range column states where the setting is supported by a measurement; a dash means we have no range.

| Parameter | Recommended | Range | Note |
|---|---|---|---|
| Ensemble size $K$ | 10 | 2 to 32 | Vote flat past ten at fixed $\lambda$; coverage still rising |
| Strength $\lambda$ | 64 | – | Selected on 200 problems at one seed |
| Strength across $K$ | $\lambda = 64$ | 10 to 32 | Hold $\lambda$, not $g$ (Corollary 3); fixed $g$ at $K \geq 20$ costs per-sample accuracy |
| Denoising steps | 128 | 128 to 256 | Vote flat in the step count; 128 is cheaper |
| Temperature | 0.6 or 0 | 0 to 0.6 | Zero (gate 8, $\lambda$=72) is 1.27 above the sampled configuration ($\lambda$=64) at 128 steps and 0.33 below it at 256; uses no randomness |
| Statistic | Count | – | Ordering is benchmark-dependent |
| Decoding regime | Blocks of 32 | – | Stronger than pure diffusion on GSM8K and MATH (Nie et al., 2025) |

### C.8 Second checkpoint: LLaDA-1.5

We repeat the deployed configuration on LLaDA-1.5 (Zhu et al., 2025), the preference-optimized checkpoint of LLaDA-8B-Instruct, which the released decoder runs unchanged; the paired gains over self-consistency are in Table 11, and the evidence is GSM8K only.

Table 11: Second checkpoint: LLaDA-1.5 on GSM8K, $K$=10, semi-autoregressive blocks of 32, 2560 evaluations, $n$=600 over two seeds (1200 paired observations). Bare means; higher is better for plurality, coverage (pass@10) and per-sample; distinct answers per pool and disagreement are descriptive. Bold marks the better method on each accuracy column. Paired SR minus self-consistency: coverage $+6.75$, plurality $+2.46$, per-sample $-1.82$. The coverage gain differs from the LLaDA-8B-Instruct gain on the same problems by $+0.58$.

| Method | Plur. (↑) | Cover. (↑) | Per-s. (↑) | Distinct | Disagr. |
|---|---|---|---|---|---|
| Self-consistency | 76.85 | 87.25 | **74.15** | 1.69 | 17.82 |
| SR, count | **79.31** | **94.00** | 72.33 | 2.26 | 28.22 |

## D Further results

### D.1 Position by position decoding

Figure 5 shows one problem decoded twice at temperature zero, once with the penalty off and once with it on, on the canvas the samples share. A row is one position and a column is one of the ten samples. Temperature zero makes the comparison exact, since the decode consumes no randomness. With the penalty off, the ten samples are one string repeated ten times, which is Proposition 1 and is the upper lattice. Every difference in the lower lattice is therefore the penalty and nothing else. A cell

is blue where the sample wrote a token no peer had taken, 57.5% of the written cells. A cell is grey where the schedule releases the penalty in the last quarter of a block, the steps left outside the penalty scope $\rho$ (Section 2.1), 24.1%. The rest are a path arriving first at a position, 8.9%, or a path writing a token a peer held, 9.5%. A path writes a token a peer held when its margin beats the gate of eight raw logits per peer. The top rows are mostly grey, which the commit order explains. The decoder commits in order of confidence. The opening word of an answer is among the least certain positions of its block, so most samples write it last. That is exactly when the penalty has been released. All twenty samples answer correctly. The pair therefore shows ten distinct decodes through the same slots against one decode repeated ten times. Sample 9 is shifted against its peers: the canvas aligns indices and not content, so part of the pool's disagreement is a phrasing shift of that kind.

### D.2 Each family at its own best temperature.

On GSM8K, self-consistency improves when its temperature is tuned. We therefore re-sweep every family and compare it at its own best setting as well as at the shared 0.6.

Table 12: Every family at its own temperature, 1280 network function evaluations, $K = 10$, blocks of 32. Left: plurality accuracy (↑) with each family at a temperature its authors report, self-consistency at both 0.6 and 1.0 since it has no published setting for this model. Right: the mean paired difference of SR with the count statistic at $\tau = 0.6$ over that row, on the problems the two methods share. GSM8K $n$=600, MATH $n$=492, TruthfulQA $n$=684; two seeds throughout, five for self-consistency at 0.6 and for the count statistic on GSM8K; the $\tau = 0$ row is one deterministic decode. Baselines are our reproductions.

| | Plurality at own $\tau$ (↑) | | | Difference from SR | | |
|---|---|---|---|---|---|---|
| Method | GSM | MATH | TQA | GSM | MATH | TQA |
| Self-consistency, $\tau = 0.6$ | 77.05 | 40.14 | 59.02 | +2.06 | −0.10 | +0.76 |
| Self-consistency, $\tau = 1.0$ | 78.28 | 39.47 | 58.65 | +0.83 | +0.57 | +1.12 |
| TAPS, $\tau = 1.0$ | 77.87 | 39.31 | 57.74 | +1.24 | +0.73 | +2.03 |
| ODD ($\alpha$=64), $\tau = 1.0$ | 78.23 | 36.66 | 60.86 | +0.88 | +3.38 | −1.08 |
| ODD ($\alpha$=256), $\tau = 1.0$ | 78.61 | 37.52 | 60.12 | +0.50 | +2.52 | −0.35 |
| SR, count, $\tau = 0.6$ | 79.11 | 40.04 | 59.77 | — | — | — |
| SR, count, $\tau = 0$ | 80.38 | 40.86 | 59.84 | −1.27 | −0.81 | −0.07 |
| SR, exp. count, $\tau = 0.6$ | 80.65 | 39.58 | 59.82 | −1.54 | +0.46 | −0.05 |
| SR, coll.-wt., $\tau = 0.6$ | 80.07 | 38.81 | 60.60 | −0.95 | +1.23 | −0.83 |

Tuning moves the baselines by about a point and in both directions: ODD ($\alpha$=256) gains 0.61 on GSM8K between 0.6 and 1.0 while losing 0.23 on MATH, so the shared temperature of Table 2 does not systematically favour us. The differences are largest on MATH against ODD and smallest on TruthfulQA, where ODD at $\alpha = 64$ leads. At 2560 evaluations the GSM8K picture is the same: $+0.68$ and $+0.63$ over the two tuned ODD settings; $+14.40$ and $+18.17$ over tuned TAPS under pure diffusion.

### D.3 Repulsion alone, at temperature zero

At temperature zero the penalty is the only source of a pool. Argmax paths with the penalty off are identical copies, so self-consistency collapses. We sweep the per-peer statistic $\lambda/(K-1)$, the margin in raw logits a peer's commitment must overcome. We repeat the design as a sequential chain of five drafts, each repelled from those completed. The decode uses no randomness, so each cell is one deterministic decode.

With the penalty off, both designs are exactly degenerate: they return 100.00% agreement, 1.00 distinct answers; per-sample, coverage and plurality all equal to the single greedy decode's 73.00. A pool of ten therefore adds nothing over a pool of one. Every number above it comes from the peer term. Plurality is monotone in the gate in both designs. At gate 8, nearest the deployed strength, the parallel design reaches 78.60 against self-consistency's 76.33. It comes within 0.33 of the deployed sampled configuration's five-seed plurality of 78.93 (Appendix C.1). Coverage still favours the sampled configuration, 93.33 against 92.33, so the penalty closes the plurality gap and not the coverage gap. The first

Penalty off, $\lambda = 0$ — 1 distinct sample of ten

| sample | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | To | To | To | To | To | To | To | To | To | To |
| 1 | determine | determine | determine | determine | determine | determine | determine | determine | determine | determine |
| 2 | how | how | how | how | how | how | how | how | how | how |
| 3 | much | much | much | much | much | much | much | much | much | much |
| 4 | the | the | the | the | the | the | the | the | the | the |
| 5 | chef | chef | chef | chef | chef | chef | chef | chef | chef | chef |
| 6 | spent | spent | spent | spent | spent | spent | spent | spent | spent | spent |
| 7 | on | on | on | on | on | on | on | on | on | on |
| 8 | the | the | the | the | the | the | the | the | the | the |
| 9 | onions | onions | onions | onions | onions | onions | onions | onions | onions | onions |
| 10 | , | , | , | , | , | , | , | , | , | , |
| 11 | we | we | we | we | we | we | we | we | we | we |
| 12 | need | need | need | need | need | need | need | need | need | need |
| 13 | to | to | to | to | to | to | to | to | to | to |
| 14 | follow | follow | follow | follow | follow | follow | follow | follow | follow | follow |
| 15 | these | these | these | these | these | these | these | these | these | these |
| 16 | steps | steps | steps | steps | steps | steps | steps | steps | steps | steps |
| 17 | : | : | : | : | : | : | : | : | : | : |
| 18 | ⏎ | ⏎ | ⏎ | ⏎ | ⏎ | ⏎ | ⏎ | ⏎ | ⏎ | ⏎ |
| 19 | ⏎ | ⏎ | ⏎ | ⏎ | ⏎ | ⏎ | ⏎ | ⏎ | ⏎ | ⏎ |
| 20 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 21 | . | . | . | . | . | . | . | . | . | . |
| 22 | Calculate | Calculate | Calculate | Calculate | Calculate | Calculate | Calculate | Calculate | Calculate | Calculate |
| 23 | the | the | the | the | the | the | the | the | the | the |

Penalty on, $\lambda = 72$, eight raw logits per peer — 10 distinct samples of ten

| sample | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | To | To | To | To | To | To | To | To | To | Sure |
| 1 | determine | determine | determine | determine | calculate | determine | find | determine | determine | , |
| 2 | how | how | the | the | the | the | how | how | how | to |
| 3 | much | much | total | total | amount | total | much | much | much | determine |
| 4 | the | chef | amount | cost | a | amount | total | did | the | how |
| 5 | chef | spent | the | of | chef | the | onions | chef | chef | much |
| 6 | spent | in | chef | spending | spent | chef | the | spend | spent | the |
| 7 | , | total | spent | on | on | spent | chef | , | on | chef |
| 8 | let | on | on | onions | the | the | spent | we | onions | spent |
| 9 | 's | onions | the | for | onions | onions | , | need | , | , |
| 10 | break | , | onions | the | , | , | we | to | we | we |
| 11 | it | we | , | chef | you | we | need | follow | can | need |
| 12 | down | need | we | , | can | need | to | a | break | to |
| 13 | step | to | can | follow | follow | to | consider | series | down | follow |
| 14 | by | follow | follow | these | these | follow | the | of | the | these |
| 15 | step | these | these | steps | steps | these | following | calculati | problem | steps |
| 16 | : | steps | steps | : | . | steps | : | . | into | : |
| 17 | ⏎ | : | : | ⏎ | ⏎ | : | ␣ | ⏎ | two | ⏎ |
| 18 | ⏎ | ⏎ | ⏎ | ⏎ | ⏎ | ⏎ | ␣ | ⏎ | main | ⏎ |
| 19 | 1 | ⏎ | ⏎ | 1 | 1 | ⏎ | 1 | 1 | steps | 1 |
| 20 | . | 1 | 1 | . | . | 1 | . | . | : | . |
| 21 | ** | . | . | Find | ** | : | Identify | ** | Calculate | ** |
| 22 | Identify | ** | ** | out | Determine | Calculate | the | Find | the | Calculate |
| 23 | the | Calculate | Calculate | the | The | the | total | the | total | the |

written under repulsion  written where the block releases the penalty

Figure 5: One problem decoded twice at temperature zero, the penalty off above and on below. A row is one position of the canvas, a column is one of the ten samples; in the lower half blue marks a token written under repulsion, grey one written where the block releases the penalty. The section text gives the rest.

path is unpenalised only in the sequential design, where draft one matches the greedy anchor exactly at gates 1 and 2. In the parallel design path zero commits inside the step it is read for, so it does not.

A sharper control fixes disagreement and removes the peer read while keeping the allocation law. Each sample draws its tokens in proportion to the water-filling counts of its own logits (Proposition 2), peer-blind, at the gate matching the penalty's disagreement (32.56 against 34.35; GSM8K in blocks, 1280 evaluations, three seeds). This control gains $0.69$ in plurality over self-consistency at $\tau = 0.6$,

Table 13: The penalty at temperature zero, GSM8K, $n = 600$, blocks of 32, 256 steps. One decode per cell. Agreement is the mean fraction of sample pairs returning the same answer; distinct is the mean number of different answers in a pool. The last rows of each block are the sampled references at $\tau = 0.6$, seed 0, and the single greedy decode.

| Configuration | Gate $g$ | $\lambda$ | Agreement | Distinct | Per-sample (↑) | Coverage (↑) | Plurality (↑) |
|---|---|---|---|---|---|---|---|
| *Parallel, $K = 10$, the deployed design* | | | | | | | |
| Penalty off | 0 | 0 | 100.00 | 1.00 | 73.00 | 73.00 | 73.00 |
| Count penalty | 1 | 9 | 77.54 | 1.91 | 72.17 | 89.50 | 76.97 |
| Count penalty | 2 | 18 | 75.62 | 2.04 | 71.85 | 90.83 | 77.31 |
| Count penalty | 4 | 36 | 73.04 | 2.18 | 71.90 | 92.00 | 77.97 |
| Count penalty | 8 | 72 | 68.27 | 2.43 | 69.65 | 92.33 | 78.60 |
| SR, $\tau = 0.6$ | 7.1 | 64 | 69.36 | 2.40 | 71.13 | 94.33 | 78.82 |
| Self-consistency, $\tau = 0.6$ | — | — | 79.85 | 1.83 | 72.53 | 88.17 | 76.33 |
| *Sequential, five drafts, each repelled from those completed* | | | | | | | |
| Penalty off | 0 | 0 | 100.00 | 1.00 | 73.00 | 73.00 | 73.00 |
| Count penalty | 1 | 4 | 76.32 | 1.61 | 73.00 | 88.83 | 76.50 |
| Count penalty | 2 | 8 | 76.25 | 1.63 | 73.00 | 88.50 | 76.67 |
| Count penalty | 8 | 32 | 70.93 | 1.77 | 71.90 | 90.33 | 78.03 |
| Count penalty | 16 | 64 | 59.93 | 2.10 | 66.33 | 90.17 | 78.12 |
| Single greedy decode | — | — | — | — | 73.00 | 73.00 | 73.00 |

a gain we do not establish, while SR at temperature zero gains $2.46$ over the control, which we do. The allocation law alone does not carry the gain; the peer read does.

### D.4 Path-index symmetry

Remark A.1 settles the processing order in structure. Permuting an order held fixed across steps returns the same pool relabeled, so no ablation over such orders exists; an order that varies across steps is a change of information structure, which the remark leaves free. The remark leaves open whether the position a path occupies predicts whether it is correct, a statement about marginals. With the penalty on, per-sample accuracy across the ten parallel indices spreads 3.17 points at gate 8 and shows no ordering. The sequential chain, where draft $i$ is repelled from $i - 1$ predecessors and a gradient exists by construction, runs 73.00, 69.67, 66.50, 63.83, 58.67 down the draft order at gate 16.

The information structure is the one order choice the remark leaves free. Table 14 quantifies it: a triangular filtration in which path $i$ reads only paths $0 \dots i - 1$, against the deployed mutual cascade at matched gate.

Table 14: The cost of the information structure. GSM8K, $n = 600$, temperature zero, one decode per cell. Path 0 of the filtration is the released greedy decode, its answers identical across the three configurations on all 600 problems and reading the anchor's 73.00; the symmetric control is the deployed mutual cascade at the same gate, decoded in the same campaign, reproducing the banked symmetric cell on all 600.

| Coupling | Gate $g$ | $\lambda$ | Agreement | Per-sample (↑) | Coverage (↑) | Plurality (↑) |
|---|---|---|---|---|---|---|
| Triangular filtration | 4 | 36 | 76.53 | 71.75 | 90.00 | 76.12 |
| Triangular filtration | 8 | 72 | 73.12 | 71.30 | 92.00 | 77.41 |
| Triangular filtration | 16 | 144 | 67.72 | 68.67 | 91.50 | 77.19 |
| Mutual cascade, symmetric | 8 | 72 | 68.27 | 69.65 | 92.33 | 78.60 |

At matched gate the filtration votes 1.19 below the symmetric cascade and is ahead on per-sample accuracy, 71.30 against 69.65. Its agreement at gate 16, 67.72, lands on the symmetric cascade's gate-8 value of 68.27, the halving that $K(K-1)/2$ ordered pairs against $K(K-1)$ predicts. At that matched agreement the symmetric cascade still converts better, 78.60 against 77.19. We keep the symmetric cascade deployed.

### D.5 Decomposition of the gain into coverage and selectivity

**Decomposition of the coverage gain into the per-sample term and dependence.** SR's coverage gain is the per-sample term, not dependence between samples: the Fréchet cap of Proposition 4 at the penalty's measured accuracies on GSM8K with blocks averages $1.94$ points, against a coverage gain of $+5.94$ over three seeds. Measured directly by a within-pool against cross-pool contrast that is exactly zero for a peer-blind sampler, the pool dependence is $+0.23$ for the penalty against $+0.02$ for self-consistency; the contrast draws three samples index-blind without replacement from one pool against one from each of three independent pools, which carry the same marginals.

One result bounds this measurement in advance. At fixed per-path marginals, re-coupling alone raises coverage above the i.i.d. value by at most $(1 - 1/K)^K < e^{-1}$, and by order $10^{-3}$ per problem at a per-path rate of one half. At the per-path rates measured on GSM8K a gain of practical size must therefore come from the induced marginals; at rates near $0.2$, as on MATH, the per-problem ceiling reaches ten points and the split must be read at the measured marginals. The net effect on the vote is therefore empirical. Plurality factors exactly as $P(\text{plur}) = P(\text{pass}) \cdot P(\text{plur} \mid \text{pass})$, coverage times the vote's accuracy given coverage. Writing $(a, x)$ for SR and $(b, y)$ for self-consistency, the gain $\Delta = ax - by$ splits exactly into a coverage term $(a - b)(x + y)/2$ and a selectivity term $(x - y)(a + b)/2$, each factor's change read at the mean of the other; this symmetric attribution averages the two orderings, which differ by the interaction $(a - b)(x - y)$. This attributes every change to a wider pool or a more selective vote.

Table 15: Decomposition of the gain. Each row is SR against self-consistency in the same regime at 2560 network function evaluations, $K = 10$, temperature 0.6, two seeds, paired on the problems both methods share. Coverage and selectivity are the two terms of the split defined in the text; they sum to $\Delta$ in every row. Fixes and breaks are the fractions of problem-seed units whose vote changes in each direction. On TruthfulQA ten uniform guesses over four options already cover the correct option with probability 94.37, which the pure-diffusion count row's coverage matches, so its coverage column is uninformative. Note: the pure-diffusion rows predate the end-of-sequence correction (Appendix D.9).

| Benchmark | Penalty | $\Delta$ plurality | Coverage | Selectivity | Fixes | Breaks |
|---|---|---|---|---|---|---|
| *Semi-autoregressive blocks of 32* | | | | | | |
| GSM8K | count | $+3.92$ | $+5.27$ | $-1.35$ | 7.17 | 3.25 |
| GSM8K | collision-wt. | $+2.92$ | $+5.98$ | $-3.06$ | 6.83 | 3.92 |
| MATH | count | $+2.34$ | $+3.09$ | $-0.75$ | 6.00 | 3.66 |
| MATH | collision-wt. | $+2.13$ | $+1.94$ | $+0.19$ | 5.18 | 3.05 |
| TruthfulQA | count | $+0.88$ | $+2.36$ | $-1.48$ | 6.65 | 5.77 |
| TruthfulQA | collision-wt. | $+3.07$ | $+3.43$ | $-0.36$ | 8.33 | 5.26 |
| *Pure diffusion* | | | | | | |
| GSM8K | count | $+19.17$ | $+20.19$ | $-1.03$ | 21.50 | 2.33 |
| GSM8K | collision-wt. | $+22.58$ | $+23.91$ | $-1.33$ | 24.92 | 2.33 |
| MATH | count | $+5.69$ | $+10.25$ | $-4.56$ | 9.65 | 3.96 |
| MATH | collision-wt. | $+11.89$ | $+13.71$ | $-1.82$ | 15.04 | 3.15 |
| TruthfulQA | count | $+0.88$ | $+16.36$ | $-15.48$ | 5.99 | 5.12 |
| TruthfulQA | collision-wt. | $+2.27$ | $+10.22$ | $-7.95$ | 8.99 | 6.73 |

Coverage accounts for between 51 and 95 percent of the moved mass in every row. The selectivity term is negative in eleven of the twelve. The vote is therefore never the source of the gain and usually offsets part of it. Churn is asymmetric and the asymmetry is the whole net effect. On GSM8K with blocks the penalty fixes 7.17% of problem-seed units and breaks 3.25%. Under pure diffusion it fixes 21.50% and breaks 2.33%. A rescue analysis settles a mechanism question. If the penalty helped by breaking a concentrated wrong consensus, then among the problems self-consistency gets wrong the rescue rate would rise with the concentration of its error. It falls instead, with a negative correlation in ten of the twelve rows; $+0.00$ and $+0.07$ in the other two. The problems rescued are the ones on which the baseline was scattered, not the ones on which it was confidently wrong.

Better aggregation does not help: across two waves of selectors, learned and heuristic, none beats plain plurality on the deployed configuration. The coverage-to-plurality gap is real, 79.38 against 96.33 at $K = 32$ on GSM8K (Appendix C.3); nothing we built closes it.

### D.6 COVERAGE PREDICTION

A wider pool is priced in per-sample accuracy: with $K$ near-independent draws, coverage on a problem whose per-sample accuracy is $a$ is $1-(1-a)^K$, so an extra unit of per-sample accuracy is worth $K(1-a)^{K-1}$ of coverage, the probability that no other sample would have found the answer. At $K=10$ that rate is 10 at $a=0$ and $10^{-8}$ at $a=0.9$. It is a falsifiable curve and the decodes test it.

A cross-benchmark reading appears to refute it. MATH has far lower per-sample accuracy than GSM8K, 31.60 against 72.31, and the smaller coverage term, $+3.09$ against $+5.27$ in blocks, which is the opposite of "largest where the model rarely succeeds". That reading confounds two things. The rate multiplies the per-sample shift the penalty produces, and that shift is not constant across benchmarks. The test has to hold the benchmark fixed and vary $a$, which is what Figure 6 does: problems are stratified by the baseline's per-sample accuracy, and each stratum's measured coverage change is set against the change independence predicts at that stratum's measured per-sample shift, $(1-a_{\mathrm{SC}})^K-(1-a_{SR})^K$.

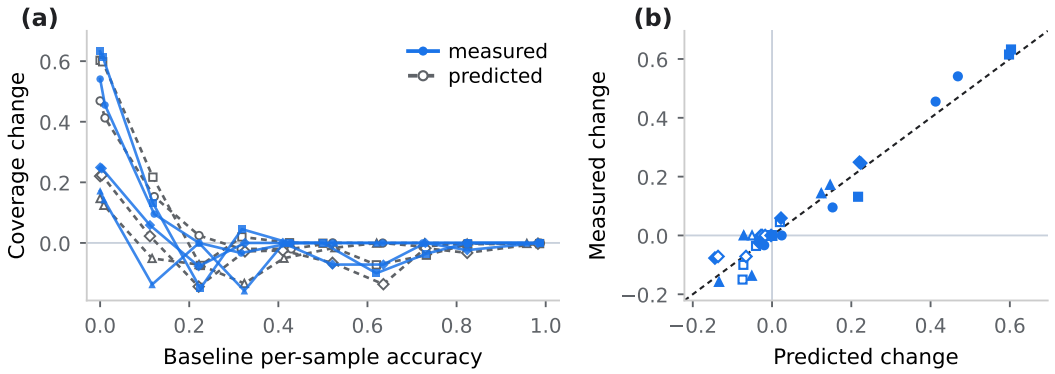


Figure 6: The exchange rate against the decodes. Problems are stratified by the baseline's per-sample accuracy; each point is one stratum of one benchmark and regime, at 2560 network function evaluations, $K=10$, temperature 0.6, two seeds, paired per problem. Predicted is $(1-a_{\mathrm{SC}})^K-(1-a_{SR})^K$, the coverage the stratum's measured per-sample shift buys if the samples were independent. **(a)** coverage change against baseline accuracy, measured and predicted. **(b)** the same points, measured against predicted, with the identity line; open markers are strata of fewer than 15 problems. Circles and squares are GSM8K in blocks and under pure diffusion, triangles and diamonds are MATH.

The account survives, and it survives quantitatively rather than in direction only. Summed over strata and weighted by their size, the ratio of measured to predicted coverage change is 1.07 and 1.02 on GSM8K in the two regimes and 1.11 and 1.16 on MATH: between two and sixteen percent more coverage than the per-sample shift buys under independence, which is the residual dependence measured directly in Section D.5 and is small beside the shift itself. The concentration the rate predicts is pronounced. On GSM8K in blocks the stratum the baseline never solves holds 61 problems and gains $0.541$ of coverage; every stratum above $a=0.35$ gains $0.000$ to three decimals. Under pure diffusion the same stratum holds 204 problems and gains $0.632$.

This also settles the MATH question without appeal to the vote. At $a=0$ the rate is $K$ on both benchmarks, so the rate is not what differs; the per-sample shift is. The penalty moves that stratum by $+0.113$ on GSM8K and by $+0.027$ on MATH, a factor of four, and the coverage gains follow at $0.541$ and $0.172$. MATH gains less coverage because the penalty finds fewer new answers there, not because the pool prices them differently.

The same fact is visible without any model, in how many of the $K$ traces are correct. The penalty empties both ends of that distribution and fills the middle. On GSM8K in blocks the share of problems no trace solves falls from 0.130 to 0.067 and the share every trace solves falls from 0.555 to 0.413; under pure diffusion the two fall from 0.383 to 0.137 and from 0.435 to 0.157. The first movement is the coverage gain, since coverage is one minus the share with no correct trace, and the second is what the vote pays for it, since a problem that leaves unanimity can lose its plurality. On MATH in blocks the share at zero barely moves, 0.461 to 0.435, which is the same shortfall the strata show and the reason its coverage gain is small.

### D.7 Reading the two sweeps

Three readings of Table 1 are not available, and the table marks none of them. Panel B's plurality column has no readable maximum: from $\lambda = 32$ to 384 it spans 1.4 points, which is why bold marks temperature's best rungs only. Its per-sample marginal exceeding its own $\lambda = 0$ value is a mixture effect of self-consistency's bimodality in this regime and supports no claim on its own; the per-sample edge is the one read at matched disagreement, which Section 3.2 states. And the penalty's disagreement saturates rather than ending where the sweep does, moving 52.01 to 53.46 as $\lambda$ doubles from 192 to 384, which is what a gate already past hard exclusion predicts; temperature has no such ceiling and reaches 88.87 at $\tau = 3.0$, by destroying the pool.

### D.8 Pass@$k$

Coverage in the tables above is pass@10; Table 16 and Figure 7 report the whole curve for three samplers at 2560 evaluations, and Table 17 the four-point curve for every method of Table 2 at its own campaigns; both use the estimator $1 - \binom{n-c}{k}/\binom{n}{k}$ over the $n = 10$ samples per problem (Chen et al., 2021), which is unbiased for independent samples. The shape is the same in all four cells: SR trails self-consistency at $k = 1$ in three of them, which is per-sample accuracy and the price already discussed, crosses by $k = 2$, and the gap widens monotonically in $k$ from there. Under pure diffusion the collision-weighted statistic is ahead at every $k$, including $k = 1$. MATH, where the vote converts least, is $+4.57$ at $k = 10$ in the deployed regime: the pool improves even where plurality cannot convert it, and the curve is the ceiling any selector over the pool could reach. TruthfulQA is excluded, since four options give a chance curve of $1 - 0.75^k$, already 94.37 at $k = 10$.

Table 16: Pass@$k$ at 2560 network function evaluations, $K = 10$, seeds 0 and 1 pooled; GSM8K $n$=600, MATH $n$=492. Left: that estimator at four $k$; bold is the best per column within a regime. Right: the count statistic's pass@$k$ gap over self-consistency, labeled per cell. The pure-diffusion configurations predate the end-of-sequence correction (Appendix D.9), which moves the baseline about two points and neither the crossover nor the ordering for $k \geq 2$.

| | GSM8K | | | | MATH | | | |
|---|---|---|---|---|---|---|---|---|
| Sampler | $k$=1 | 2 | 5 | 10 | $k$=1 | 2 | 5 | 10 |
| *Blocks of 32* | | | | | | | | |
| Self-consistency | **72.18** | 78.58 | 84.54 | 87.67 | **31.60** | 39.60 | 48.78 | 55.89 |
| SR, count | 70.93 | **81.52** | 90.07 | 93.83 | 30.22 | **40.51** | **52.12** | **60.47** |
| SR, coll.-wt. | 66.70 | 80.14 | **90.20** | **94.75** | 28.25 | 38.57 | 50.75 | 58.74 |
| *Pure diffusion* | | | | | | | | |
| Self-consistency | 53.42 | 56.96 | 60.53 | 63.17 | 23.77 | 27.99 | 32.87 | 36.08 |
| SR, count | 56.19 | 70.82 | 82.40 | 86.83 | 23.37 | 32.68 | 44.08 | 51.52 |
| SR, coll.-wt. | **63.06** | **76.23** | **86.41** | **91.25** | **27.79** | **37.30** | **48.46** | **55.79** |

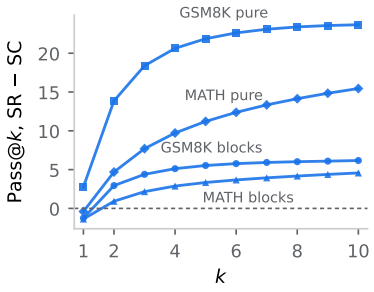


**Figure 7:** the pass@$k$ curves of Table 16; SR crosses self-consistency by $k$=2 and the gap widens thereafter.

### D.9 End-of-sequence correction

The pure-diffusion rows of Table 2 apply a remedy the model's authors attach to this regime: LLaDA sets the confidence of the end-of-sequence token to zero during pure-diffusion sampling on GSM8K, MATH and three other benchmarks, because the padding of its supervised data otherwise ends a decode early (Nie et al., 2025). Two questions attach to the correction. What is it worth, since our earlier campaigns omit it; and does the penalty's margin merely substitute for it, since with the content mask off peers' committed end-of-sequence tokens enter the count and the penalty is pushed off that token. Table 18 decides both: every configuration applies the correction, and one configuration additionally turns the content mask on, which removes the token from the count entirely so the penalty cannot substitute for anything.

The correction is material, and it does not explain the margin. It lifts the baseline from the 54.39 of Table 9 to 57.02 and removes the truncation the omission caused: no sample here fails to parse, against

Table 17: Pass@$k$ for every method of Table 2, same campaigns and settings; pass@1 and pass@10 are Table 2's per-sample and coverage columns. TruthfulQA reports $k$=1 only, chance alone reaching 94.37 at $k$=10; TAPS (†) runs the uncorrected protocol. Bold: best per column within a regime.

| | GSM8K | | | | MATH | | | | TQA |
|---|---|---|---|---|---|---|---|---|---|
| Method | $k$=1 | 3 | 5 | 10 | $k$=1 | 3 | 5 | 10 | $k$=1 |
| *Semi-autoregressive blocks of 32* | | | | | | | | | |
| Self-consistency | **70.04** | 82.45 | 86.03 | 89.70 | **30.02** | 44.14 | 49.78 | 56.81 | **54.44** |
| TAPS | 66.77 | 83.28 | 87.53 | 91.00 | 29.26 | 44.03 | 49.84 | 57.32 | 53.53 |
| ODD ($\alpha$=64) | 64.94 | 82.60 | 87.21 | 91.75 | 24.19 | 40.12 | 46.45 | 54.27 | 52.35 |
| ODD ($\alpha$=256) | 66.57 | 83.32 | 87.73 | 92.17 | 22.77 | 39.22 | 45.97 | 54.37 | 49.85 |
| SR, count | 68.21 | 85.06 | 89.50 | 93.53 | 27.67 | **44.22** | **50.98** | **59.86** | 48.07 |
| SR, count, $\tau$=0 | 69.33 | 85.67 | 90.02 | 94.33 | 27.13 | 43.56 | 49.80 | 57.11 | 46.92 |
| SR, exp. count | 66.47 | **85.74** | **90.61** | **94.67** | 26.78 | 43.24 | 50.28 | 59.65 | 50.75 |
| SR, collision-wt. | 61.87 | 84.72 | 90.15 | 94.42 | 24.44 | 40.75 | 47.46 | 55.39 | 52.78 |
| *Pure diffusion (whole response as one block)* | | | | | | | | | |
| Self-consistency | 54.47 | 61.55 | 63.95 | 67.00 | 22.87 | 29.06 | 32.37 | 37.40 | 54.17 |
| TAPS† | 50.34 | 61.75 | 65.58 | 69.75 | 22.46 | 30.81 | 34.57 | 39.74 | **55.26** |
| SR, count | 53.85 | 75.65 | 81.87 | 87.75 | 22.62 | 36.53 | 42.66 | 50.51 | 53.79 |
| SR, count, $\tau$=0 | 53.82 | 75.52 | 81.68 | 87.17 | 23.17 | 36.71 | 42.64 | 50.41 | 52.85 |
| SR, exp. count | 53.87 | 77.53 | 84.31 | **90.33** | 20.47 | 34.43 | 40.73 | 48.78 | 50.78 |
| SR, collision-wt. | **59.81** | **80.14** | **85.26** | 90.00 | **25.10** | **40.38** | **47.11** | **55.79** | 50.50 |

Table 18: The regime re-decoded with the released correction applied to every configuration. GSM8K under pure diffusion, $K = 10$, 128 steps, temperature 0.6, $n = 600$, two seeds. The third row additionally excludes the end-of-sequence token from the count. The rows with the mask off are the same campaign as Table 2's pure-diffusion block; the first row is the baseline. The last column is the difference in plurality from that baseline.

| Sampler | Mask | Plurality (↑) | Coverage (↑) | Per-sample (↑) | Disagr. | Δ plur. |
|---|---|---|---|---|---|---|
| Self-consistency | off | 57.02 | 67.00 | 54.47 | 20.40 | — |
| SR, count | off | 71.53 | 87.75 | 53.85 | 49.73 | +14.50 |
| SR, count | **on** | 70.44 | 88.00 | 53.07 | 50.45 | +13.42 |
| SR, coll.-wt. | off | 76.27 | 90.00 | 59.81 | 45.93 | +19.25 |

0.64% and 0.81% of samples in the uncorrected pair. The count penalty's margin over the baseline moves by only 1.41 with the correction, from 15.91 to 14.50, so about a point of the uncorrected pure-diffusion margin was ours rather than the sampler's; the 2.63 is the baseline's own lift. What remains is not two points. With the correction applied to both sides the collision-weighted penalty leads by 19.25, and with the end-of-sequence token removed from the count altogether, so that the penalty has no access to the channel in question, the count statistic still leads by 13.42 on the vote and by 21.00 on coverage. Turning the mask on costs 1.08 points of plurality, so the channel carries none of the effect.

The same campaign covers all three benchmarks and populates Table 2's pure-diffusion block. Against the corrected baseline the margins are +13.96 to +19.25 on GSM8K and +5.40 to +11.37 on MATH; on TruthfulQA the temperature-zero configuration leads by 3.42. Two protocols remain uncorrected: TAPS, whose released loop exposes no hook for the correction, and the 2560-evaluation campaigns of Tables 1 and 5.