# When Exact Loops Beat Shortcuts: Deep-Target Approximation Is Not Recurrent-Depth Acceleration

Anonymous authors
Paper under double-blind review

## Abstract

Recurrent-depth Transformers reuse a shared block for many sequential updates, making depth a candidate for parallel-in-time acceleration. A shortcut can look successful against the deepest endpoint while remaining worse than the model's native state after the same sequential depth. We formalize this comparison as the Native-Trajectory Criterion (NTC). On a prespecified 1,000-example Parcae-140M screen, identity-coarse Parareal yields pooled median forward KL $1.07 \to 29.1 \to 0.292$ across one to three corrections, while matched exact recurrence is better at every evaluated path budget. Anderson-style mixing at budget 8 reaches $8.94 \times 10^{-4}$ KL to R16, yet exact R8 reaches $6.31 \times 10^{-4}$. Paired evaluation confirms the difference. Boundary diagnostics recover the expected Parareal exactness front but reveal a near-antipodal unresolved final boundary after the second correction. The same ordering and geometry persist in a 370M cross-scale robustness check. Thus for post-hoc recovery of fixed-depth behavior, deep-target proximity is insufficient evidence of recurrent-depth acceleration.

## 1 Introduction

Looped language models repeatedly apply shared Transformer parameters and scale test-time computation through recurrence (Prairie et al., 2026). Their depth is elastic, but the native computation remains sequential. A shortcut may appear successful when judged only against the full-depth endpoint, even though the model's native intermediate state at the same sequential depth is closer to that endpoint. We show that this occurs in practice: Anderson-style mixing reaches very small R16 KL at budget 8 yet is worse than native R8, while identity-coarse Parareal exhibits a reproducible non-monotone boundary failure across two Parcae scales.

Classical Parareal alternates cheap coarse propagation with parallel fine solves over evolution intervals (Lions et al., 2001). Mapping "time" to recurrent depth preserves the released model as the fine operator, but a method that approximates the recurrence-16 target can still be inferior to simply stopping the same model after 4, 8, or 12 exact updates. This distinction is easy to miss when only target error and nominal path reduction are reported.

Contributions. We provide (1) the NTC, which compares a post-hoc recovery method with the native trajectory at equal fine-transition dependency depth; (2) a controlled negative result for identity-coarse Parareal and Anderson-style mixing; (3) per-example boundary evidence that provides an explanation for the non-monotone $K = 2$ output failure; and (4) a 370M cross-scale robustness check. We claim neither measured wall-clock acceleration nor general impossibility.

## 2 The native trajectory is the baseline

Let the released recurrent model define

$$h_{r+1} = F_r(h_r), \qquad r = 0, \ldots, R-1, \tag{1}$$



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.

and define

$$\Phi_{a:b} = F_{b-1} \circ \cdots \circ F_a, \qquad h^\star_B = \Phi_{0:B}(h_0). \tag{2}$$

We retain indexed notation for generality, but the pinned Parcae release uses one shared transition, $F_r \equiv F$ (Prairie et al., 2026): recurrence index and target depth are not inputs to its recurrent core or readout. Let $q$ denote that unchanged common readout. A post-hoc algorithm $A_B$ may be close to $h^\star_R$, but $h^\star_B$ is already available after the same fine-transition dependency depth. For a nontrivial budget $0 < B < R$, the Native-Trajectory Criterion requires

$$\mathrm{KL}(q(h^\star_R) \| q(A_B(h_0))) < \mathrm{KL}(q(h^\star_R) \| q(h^\star_B)) \,. \tag{3}$$

Here $B$ is the path budget: fine-transition dependency depth, not FLOPs, total work, or measured latency. For post-hoc methods whose objective is to recover the $R$-step computation at lower dependency depth, proximity to the deep target is necessary but not sufficient evidence of acceleration. NTC evaluates recovery of a specified full-depth computation; it does not rule out shortcuts that intentionally follow a different trajectory while preserving or improving downstream task correctness. Hidden-state distances below are diagnostics rather than the headline NTC discrepancy.

At budget 8, the distinction is concrete: Anderson attains pooled median token KL 0.00089422 to $h^\star_{16}$, an apparently strong approximation, but exact $h^\star_8$ attains 0.000631072. Anderson therefore succeeds at deep-target approximation while failing as an accelerator under Eq. 3.

## 3 Method and cost accounting

We partition $R$ updates into $M$ contiguous segments. Let $\Phi_m$ denote exact fine propagation through segment $m$ and $G_m$ a cheap coarse map. Additive Parareal initializes boundaries with $G_m$ and updates

$$h^{k+1}_{m+1} = G_m(h^{k+1}_m) + \Phi_m(h^k_m) - G_m(h^k_m). \tag{4}$$

Fine solves are independent within correction round $k$ under true segment parallelism. We use identity coarse propagation, $G_m(h) = h$, and report the ideal fine-transition dependency depth $K(R/M)$. Total fine work is $KR$. The path metric assumes ideal concurrent fine solves, ignores communication, and treats identity propagation as free; aggregate fine work is reported separately.

Identity coarse propagation also explains the first-round geometry algebraically. Equal-length segments in this shared-transition setting have the same propagator $\Phi$. With $h^0_m = h_0$, Eq. 4 gives

$$h^1_m = h_0 + m\big(\Phi(h_0) - h_0\big).$$

Thus the unresolved first-round boundaries form an affine extrapolation sequence whose displacement grows linearly with segment index, consistent with the observed norm ladder. This identity alone does not predict the second-round near-antipodal reversal.

| Method | Dependency depth | Total fine work |
|---|---|---|
| Exact R4 | 4 | 4 |
| DepthPar K1 | 4 | 16 |
| Exact R8 | 8 | 8 |
| DepthPar K2 | 8 | 32 |
| Exact R12 | 12 | 12 |
| DepthPar K3 | 12 | 48 |

Table 1: Cost accounting. DepthPar receives ideal four-way segment concurrency and performs four times the aggregate fine work of matched exact early exit.

Anderson-style mixing. Initialize $x_0 = h_0$. At transition $i$, let $y_i = F_i(x_i)$, $r_i = y_i - x_i$, $\ell_i = \min(5, i+1)$, and $J_i = \{i - \ell_i + 1, \ldots, i\}$. The residual-history matrix is $R_i = [r_j]_{j \in J_i}$.

We solve

$$\alpha_i = \arg\min_{\mathbf{1}^\top \alpha = 1} \|R_i \alpha\|_2^2 + \lambda \|\alpha\|_2^2, \qquad x_{i+1} = \sum_{j \in J_i} \alpha_{i,j} \big((1-\beta) x_j + \beta y_j\big). \tag{5}$$

We use history five, $\beta = 1$, and $\lambda = 10^{-4}$. Each constrained KKT system is at most $6 \times 6$ and solved in CPU float64. The favorable path metric omits these dense solves. The implementation retains the indexed adapter interface, while the pinned release has $F_i \equiv F$. We call this Anderson-style mixing because it is ridge-regularized and executed for a fixed budget; we invoke no standard convergence guarantee (Walker & Ni, 2011; Bai et al., 2019). An Anderson budget $B$ performs exactly $B$ sequential fine-transition evaluations, so its fine-transition dependency depth and aggregate fine work are both $B$, excluding the small dense KKT solves above.

Lemma 1 (Classical progressive exactness). In exact arithmetic, after $k$ Parareal corrections, boundaries $m \le k$ equal their sequential references.

Proof. Boundary zero is fixed. Assume $h_m^{k+1} = h_m^\star$ and $h_m^k = h_m^\star$. Substitution into Eq. 4 cancels the two coarse terms and gives $h_{m+1}^{k+1} = \Phi_m(h_m^\star) = h_{m+1}^\star$. Induction over corrections and boundaries completes the result. □

This classical property predicts the exactness front; our empirical observation is that the first unresolved boundary can worsen sharply immediately beyond it.

## 4 Experimental design

Data and pairing. We evaluate the validation shard data/val/train/shard_000000.parquet of SandyResearch/fineweb-edu-shuffled, revision 4d1138c0e4d5, using rows 0–999 for the prespecified screen and rows 0–255 for post-hoc diagnostics. Text is tokenized with SandyResearch/parcae-tokenizer at revision 6247b5d05928 and truncated to at most 129 tokens, yielding at most 128 teacher-forced next-token positions. Following the pinned Parcae forward procedure, the initial recurrent state $h_0$ is sampled stochastically. We seed this draw deterministically by example as 42+ row index and reuse the identical sampled $h_0$ across all paired candidates for that example. Exact R4/R8/R12 truncate the same R16 reference trajectory after $B$ fine transitions; no target-depth-dependent model input changes between these baselines. The 64-row smoke slice is rows 0–63 and therefore overlaps the screen by all 64 examples; it was used only for coarse-operator promotion under the prespecified rule, not for tuning the reported $R = 16, M = 4$ configurations. Because the overlap selected the promoted DepthPar operator rather than the native baseline, it cannot explain the negative finding by artificially favoring exact recurrence; the 1,000-example screen is not treated as held-out evidence for operator selection.

Protocol. The protocol precommitted the 140M checkpoint, $R = 16$, $M = 4$, MPS float32 execution, metrics, candidate selection, and Gate A before the screen outcomes were observed. Because the record has no independently timestamped external deposit, we describe it as prespecified. The absolute median/p95 KL limits (0.01/0.05) were operational promotion thresholds, not externally validated quality boundaries; the scientific comparison is matched native recurrence under NTC.

Gate A. A DepthPar candidate passes only if it simultaneously (i) satisfies the prespecified pooled median KL $\le 0.01$ and p95 KL $\le 0.05$ thresholds and (ii) strictly outperforms exact early exit at the same fine-transition dependency depth under NTC. Failure of Gate A stops the branch before learned coarse models, broader sweeps, or paid systems experiments.

Metrics. For each evaluated position $t$, we compute forward $\mathrm{KL}(p_{16,t} \| p_t)$, so probability mass important to the full-depth reference determines the penalty. Frozen Gate A medians and 95th percentiles pool all evaluated token positions. Top-1 agreement is the global fraction of those positions whose argmax matches R16. Separately, paired analysis first averages KL within each example. The pooled-token summary was fixed by the prespecified

protocol; per-example aggregation is used separately for paired uncertainty analysis. All evaluation is teacher forced. For boundary $m$ after correction $k$, we record

$$e_m^k = \frac{\|h_m^k - h_m^\star\|_2}{\|h_m^\star\|_2}, \tag{6}$$

cosine, norm ratio, correction magnitude, and solver residual.

## 5 Prespecified result

| Method | $B$ | Median KL | P95 KL | Top-1 |
|---|---|---|---|---|
| Exact R4 | 4 | 0.0139444 | 0.096674 | 0.902077 |
| DepthPar K1 | 4 | 1.06579 | 2.88891 | 0.61818 |
| Exact R8 | 8 | 0.000631072 | 0.00858599 | 0.974694 |
| DepthPar K2 | 8 | 29.1421 | 44.3714 | 7.83945e-06 |
| Exact R12 | 12 | 2.94723e-05 | 0.000690942 | 0.99347 |
| DepthPar K3 | 12 | 0.292168 | 0.87341 | 0.735724 |

Table 2: Prespecified screen. KL summaries pool evaluated token positions; Top-1 is global across those positions. Lower KL is better.

At paths 4, 8, and 12, DepthPar median KL is 1.06579, 29.1421, and 0.292168, versus 0.0139444, 0.000631072, and 2.94723e-05 for exact early exit. The $K = 2$ correction increases sharply from 1.06579 to 29.1421, while $K = 3$ partially recovers without approaching exact R12. Exact recurrence occupies the Pareto frontier of pooled median KL versus fine-transition dependency depth among evaluated configurations. Gate A fails; the narrow empirical conclusion is that identity-coarse Parareal fails on these Parcae configurations.

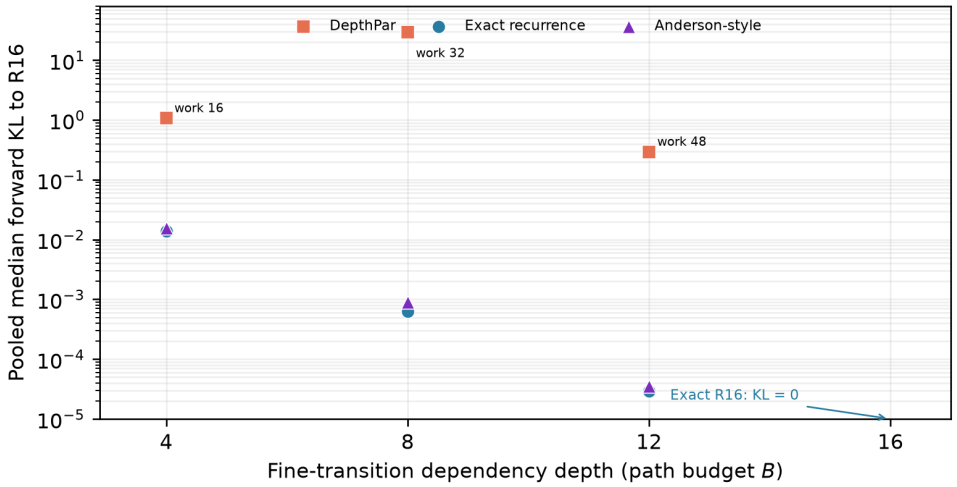



Figure 1: Pooled median forward KL versus fine-transition dependency depth. DepthPar labels report aggregate fine work. Exact R16 (zero KL) is annotated outside the log axis rather than clipped to an epsilon. This figure does not report measured latency.

### 5.1 Anderson shows why target proximity is insufficient

Anderson progressively approaches R16 and passes the operational absolute KL threshold at budget 8, but exact recurrence is better at every matched nontrivial budget. B16 has $B = R$ and is retained only as an endpoint sanity reference; Eq. 3 does not apply there. For paired per-example means, define $\Delta_i = \overline{\mathrm{KL}}_{i,\text{Anderson B8}} - \overline{\mathrm{KL}}_{i,\text{Exact R8}}$; positive values favor R8.

Across 1000 paired examples, median $\Delta$ is 0.000696693 with deterministic 10,000-resample percentile-bootstrap 95% CI [0.000660049, 0.000727153] (seed 42), and 99.8% of examples favor R8. (The paired median is not the difference of the pooled marginal medians because the per-example difference distribution is skewed.) This demonstrates the distinction formalized by Eq. 3: strong approximation to the full-depth target does not imply improvement over the native trajectory at the same dependency depth.

| Method | $B$ | Median KL | P95 KL | Top-1 |
|---|---|---|---|---|
| Exact R4 | 4 | 0.0139444 | 0.096674 | 0.902077 |
| Anderson B4 | 4 | 0.0156289 | 0.107947 | 0.897123 |
| Exact R8 | 8 | 0.000631072 | 0.00858599 | 0.974694 |
| Anderson B8 | 8 | 0.00089422 | 0.0117127 | 0.969818 |
| Exact R12 | 12 | 2.94723e-05 | 0.000690942 | 0.99347 |
| Anderson B12 | 12 | 3.50876e-05 | 0.000790685 | 0.992835 |
| Exact R16 | 16 | 0 | 0 | 1 |
| Anderson B16 | 16 | 8.19901e-07 | 4.92202e-05 | 0.998315 |

Table 3: Prespecified-screen Anderson comparison. The sequential numerical method approaches the R16 target, but exact recurrence is better at every matched nontrivial budget. B16 is an endpoint sanity reference, not an NTC acceleration test.

## 6 Why the second correction fails

Figure 2 recovers the expected exactness front, and the first-round states follow the affine norm ladder derived above. At $k = 2$, Lemma 1 makes $h_1^2$ and $h_2^2$ exact. The first unresolved state, $h_3^2$, is already strongly degraded: median relative error 1.01272, cosine -0.0746571, and norm ratio 0.100833. The following and final unresolved state, $h_4^2$, is near-antipodal to its reference: median cosine -0.991867 and 100% of examples have negative cosine, while median norm ratio is 2.05548. On this 256-example diagnostic subset, the broad reversal coincides with pooled-median KL 29.1768 (versus 29.1421 on the full 1,000-example prespecified screen), providing a boundary-state explanation for the observed spike. At $k = 3$, the exactness front advances through $h_3^3$, while $h_4^3$ recovers its direction but remains norm-biased, explaining partial rather than full output recovery.

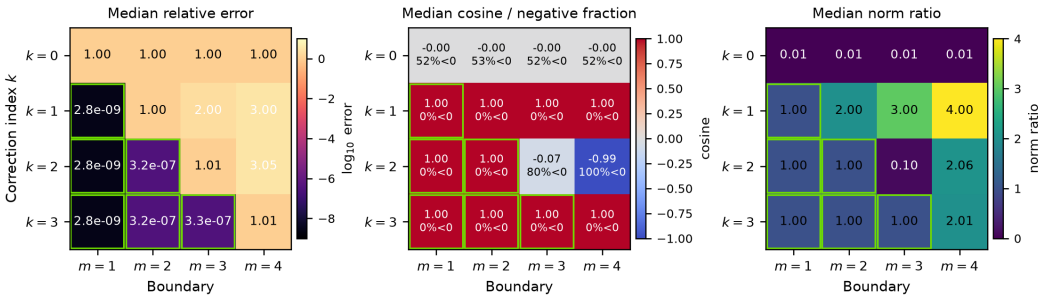



Figure 2: Post-hoc 140M boundary distributions on 256 fixed examples. Green boxes mark the classical exact region. Cells show medians; the cosine panel also shows the fraction below zero. At $h_4^2$, 10–90% ranges are error [3.02186, 3.07297], cosine [-0.993099, -0.989983], and norm ratio [2.02695, 2.07921].

Why this is not a generic instability claim. Every reported state and metric is finite, the implementation recovers the exact solution when the correction count reaches the number of segments in the earlier smoke test, and the first $k$ boundaries after correction $k$ obey causal exactness in the post-hoc run. The evidence therefore localizes the failure to useful early correction with this coarse map; it does not show that Parareal is algebraically incorrect or that every recurrent transition must amplify boundary errors.

## 7 Cross-scale robustness check

Parcae-370M reproduces the entire qualitative ordering: median KL is 1.12384→ 25.7537→ 0.33141, whereas exact early exit reaches 0.00903457, 0.00034036, and 1.41519e-05. Its boundary geometry also shows the same near-antipodal $h_4^2$ behavior: median cosine -0.997858, with negative cosine on 100% of examples. This shows that the pattern is not unique to the 140M checkpoint within the Parcae family, but both scales share an architecture family and the same post-hoc rows.

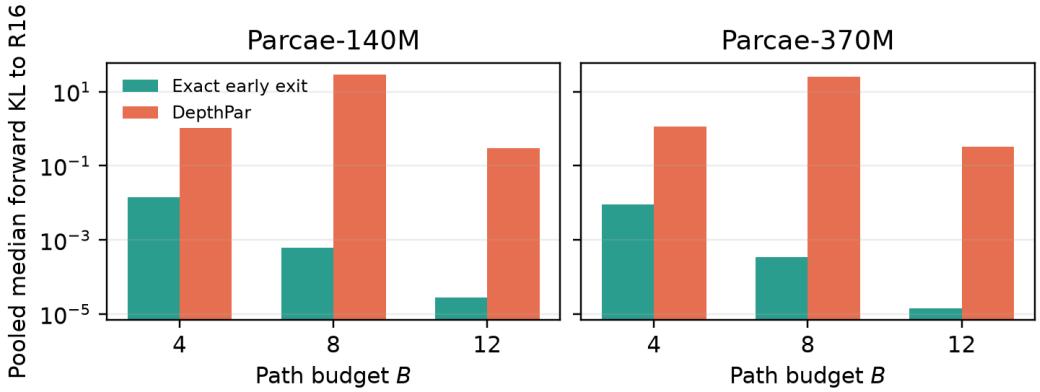



Figure 3: Same 256-example post-hoc comparison at two scales. Exact recurrence has lower pooled median KL at each matched path budget.

## 8 Related work and implications

Parareal and MGRIT expose concurrency across evolution intervals (Lions et al., 2001; Falgout et al., 2014). Lee et al. construct and train a coarse parallel neural network (Lee et al., 2024); we instead test post-hoc trajectory recovery in an unchanged released recurrent Transformer. Deep equilibrium models formulate inference as solving an implicit fixed point and commonly use root-finding or Anderson-style acceleration (Bai et al., 2019). Adaptive-depth and early-exit work makes intermediate computation operationally meaningful (Graves, 2016; Dehghani et al., 2018; Banino et al., 2021; Xin et al., 2020; Schuster et al., 2022; Popescu et al., 2026). Recent looped-Transformer work also studies learned trajectory shortcuts and stopping (Logan, 2026; Jeddi et al., 2026). Parallel Loop Transformer introduces trained architectural cross-loop parallelism to remove sequential loop dependencies (Wu et al., 2025); our setting instead asks whether an already-trained recurrent trajectory can be parallelized post hoc without altering its transition.

Our result suggests a stricter evaluation pattern for all of these shortcut families when applied post hoc to elastic recurrent depth: plot quality against fine-transition dependency depth, include $h_B^\star$ at every budget, and diagnose hidden-state boundaries rather than relying only on terminal logits. A method may approach $h_R^\star$ yet fail NTC throughout the useful budget range.

Practical evaluation checklist. Based on these results, we recommend reporting (i) the native exact trajectory at every claimed path budget; (ii) total fine work separately from dependency depth; (iii) whether segment solves actually execute in parallel; (iv) boundary-state error, cosine, and norm—not only terminal output error; and (v) a clear separation between prespecified decisions and post-hoc diagnosis; and (vi) peak activation/boundary-state memory and KV-cache overhead where applicable. These requirements are inexpensive compared with training or systems scaling and would have prevented the Anderson B8 result from being misread as acceleration.

## 9 Limitations

The prespecified conclusion is limited to Parcae-140M, one $R = 16, M = 4$ screen, and the identity coarse map selected by smoke tests. The 370M study is post-hoc, uses 256 examples from the same data source, and is a model-scale robustness check rather than evidence outside the Parcae architecture family. We compare next-token distributions on fixed contexts and do not claim equivalent behavior under autoregressive free-running generation, where small differences may compound. We did not train a coarse model, implement true multi-device segment parallelism, or measure wall-clock acceleration. The path metric assumes ideal concurrent fine solves, ignores communication and dense-solve overhead, and treats identity propagation as free; DepthPar's aggregate fine work is four times that of matched exact early exit.

## 10 Conclusion

For elastic recurrent models, the relevant question is not only "How close are we to the deepest trajectory?" but "Are we closer than the model would already be after the same sequential depth?" Here, identity-coarse Parareal fails that test across two Parcae scales, while Anderson shows that even a close deep-target approximation need not satisfy NTC.

## Reproducibility Statement

Section 4 specifies checkpoints, data, metrics, seeds, and protocol. Appendices A–C document prespecification separation, boundary diagnostics, and the corrected Anderson implementation. Exact run identifiers are withheld from the review copy. The anonymous supplement contains executable configurations, analysis code, tests, and regeneration scripts.

## AI Use Statement

Generative AI tools were used only for language editing and for creating or modifying figures and plots. The author reviewed and approved the resulting text and visual materials and takes full responsibility for the paper.

## References

Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. Deep equilibrium models. In Advances in Neural Information Processing Systems, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/01386bd6d8e091c2ab4c7c7de644d37b-Abstract.html.

Andrea Banino, Jan Balaguer, and Charles Blundell. Pondernet: Learning to ponder, 2021. URL https://arxiv.org/abs/2107.05407.

Mostafa Dehghani, Stephan Gouws, Oriol Vinyals, Jakob Uszkoreit, and Lukasz Kaiser. Universal transformers, 2018. URL https://arxiv.org/abs/1807.03819.

Robert D. Falgout, Stephanie Friedhoff, Tzanio V. Kolev, Scott P. MacLachlan, and Jacob B. Schroder. Parallel time integration with multigrid. SIAM Journal on Scientific Computing, 36(6):C635–C661, 2014. doi: 10.1137/130944230.

Alex Graves. Adaptive computation time for recurrent neural networks, 2016. URL https://arxiv.org/abs/1603.08983.

Ahmadreza Jeddi, Marco Ciccone, and Babak Taati. Loopformer: Elastic-depth looped transformers for latent reasoning via shortcut modulation. In International Conference on Learning Representations, 2026. URL https://openreview.net/forum?id=RzYXb5YWBs.

Chang-Ock Lee, Youngkyu Lee, and Jongho Park. Parareal neural networks emulating a parallel-in-time algorithm. IEEE Transactions on Neural Networks and Learning Systems, 35:6353–6364, 2024. doi: 10.1109/TNNLS.2022.3206797.

Jacques-Louis Lions, Yvon Maday, and Gabriel Turinici. A "parareal" in time discretization of PDEs. Comptes Rendus de l'Académie des Sciences - Series I - Mathematics, 332: 661–668, 2001. doi: 10.1016/S0764-4442(00)01793-6.

Joe Logan. Per-token fixed-point convergence in depth-recurrent transformers, 2026. URL https://arxiv.org/abs/2607.14427.

Andrei Cristian Popescu, Haitz Sáez de Ocáriz Borde, and Pietro Liò. Adaptive depth in looped transformers: Diagnosing learned halting gates and trajectory readouts, 2026. URL https://arxiv.org/abs/2607.20519.

Hayden Prairie, Zachary Novack, Taylor Berg-Kirkpatrick, and Daniel Y. Fu. Parcae: Scaling laws for stable looped language models, 2026. URL https://arxiv.org/abs/2604.12946.

Tal Schuster, Adam Fisch, Jai Gupta, Mostafa Dehghani, Dara Bahri, Vinh Tran, Yi Tay, and Donald Metzler. Confident adaptive language modeling. In Advances in Neural Information Processing Systems, volume 35, 2022. doi: 10.52202/068431-1269. URL https://papers.nips.cc/paper_files/paper/2022/hash/6fac9e316a4ae75ea244ddcef1982c71-Abstract-Conference.html.

Homer F. Walker and Peng Ni. Anderson acceleration for fixed-point iterations. SIAM Journal on Numerical Analysis, 49(4):1715–1735, 2011. doi: 10.1137/10078356X.

Bohong Wu, Mengzhao Chen, Xiang Luo, Shen Yan, Qifan Yu, Fan Xia, Tianqi Zhang, Hongrui Zhan, Zheng Zhong, Xun Zhou, Siyuan Qiao, and Xingyan Bin. Parallel loop transformer for efficient test-time computation scaling, 2025. URL https://arxiv.org/abs/2510.24824.

Ji Xin, Raphael Tang, Jaejun Lee, Yaoliang Yu, and Jimmy Lin. DeeBERT: Dynamic early exiting for accelerating BERT inference. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 2246–2251. Association for Computational Linguistics, 2020. doi: 10.18653/v1/2020.acl-main.204. URL https://aclanthology.org/2020.acl-main.204/.

## A Prespecification and smoke selection

The protocol preceded the 1,000-row screen but lacks an independently timestamped deposit. The 64-row smoke slice comprises screen rows 0–63, so it overlaps the headline screen by all 64 examples; it was used only for prespecified coarse-operator promotion, not configuration tuning. At $R = 8, M = 2$, identity scored 0.274647 while one-probe residual last and mean—the same operator with one probe—both scored 2.15199. The operational rule rejected K1 pooled median KL above 1.0, so only identity advanced.

## B Boundary distributions

Post-hoc boundary runs use the same 256 rows at both scales. Figure 2 and its caption report medians, negative-cosine fractions, and $h_4^2$ 10–90% ranges; machine-readable tables add correction magnitude and solver residual. These post-hoc diagnostics do not alter or reopen the prespecified Gate A decision.

## C Anderson validation and implementation incident

The original Anderson run failed because a fused MPS-float32 to CPU-float64 transfer corrupted a finite residual before the Gram solve. A two-stage source-dtype copy followed by CPU widening matched the native residual without changing solver hyperparameters. An independent NumPy float64 KKT reference agrees with the CPU solve below $10^{-12}$ maximum coefficient error; the ill-conditioned opt-in MPS fixture differs by at most $1.34 \times 10^{-5}$ (below $2 \times 10^{-7}$ relative error). Failed attempts remain preserved.