CONFLICT-AWARE MODEL MERGING FOR DATA-FREE LONG-TO-SHORT REASONING

Anonymous authors
Paper under double-blind review

ABSTRACT

Chain-of-thought (CoT) reasoning improves performance on challenging problems by generating structured reasoning steps, but it often produces unnecessarily long responses to easy ones. While typical long-to-short methods require additional training, calibration data, or input-dependent inference, model merging provides a promising way by combining long-CoT model with short model in parameter space. However, existing merging approaches either apply uniform coefficients across layers or rely on auxiliary data or activation, lacking systematic analysis about conflicts between long-CoT and short models. Importantly, strong CoT update is not necessarily beneficial when it conflicts with corresponding short-model update, motivating a explicit conflict-aware allocation of reasoning strength across layers. In this paper, we identify two patterns that guide long-to-short merging: (1) ***strength*** of CoT update, which reflects the potential reasoning benefit, and (2) ***conflict*** with short model, which indicates the risk of disrupting short-answer behavior at certain layer. Based on this observation, we propose BCL-Merge, a data-free method that design corresponding signals to select layers and assign proper coefficients into short model under given reasoning strength budget. Under explicit assumptions, we show that the resulting layer selection and strength allocation optimize an entropy-regularized proxy objective that balances reasoning benefit against short-answer disruption. Extensive experiments across reasoning benchmarks show that our method reaches 46.3% accuracy on Qwen2.5-1.5B, outperforming existing merging method by 2.9% while reducing response length by 81.5%. Notably, it also exceeds performance of long-CoT model by 5.3%. Further analysis reveals that our method shortens responses on easy examples and improves accuracy on hard examples, validating its utility.

1 INTRODUCTION

Chain-of-thought (CoT) reasoning improves performance on complex problems by generating explicit intermediate steps before producing a final answer (Wei et al., 2022; Guo et al., 2025). By externalizing intermediate computations, a model can decompose a problem, preserve partial results, and condition later steps on earlier deductions. This additional reasoning, however, comes with a substantial output cost. A long-CoT reasoning model can produce hundreds or thousands of tokens even for an easy question that a short model already solves correctly (Chen et al., 2024) with few tokens. Consequently, using one CoT model for workloads that mix routine and difficult queries increases decoding cost and latency, reduces serving throughput, and wastes computation on examples that do not require extended reasoning.

To reduce unnecessary response length while preserving the benefits of CoT reasoning, long-to-short reasoning aims to retain useful reasoning while shortening model responses. Existing methods prescribe or adapt a token budget (Han et al., 2025; Lou et al., 2025), compress or prune reasoning traces (Luo et al., 2026; Xia et al., 2025), or train models to produce concise reasoning (Ma et al., 2025; Munkhbat et al., 2025). Depending on their design, these methods require additional training data and optimization, a runtime controller, or input-dependent decoding. Such requirements complicate deployment by introducing additional training, calibration, or inference-time decision.

Recently, model merging provides a feasible approach to this question as it integrates multiple models into one unified model without an additional training (Ilharco et al., 2022; Yadav et al., 2023). In

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

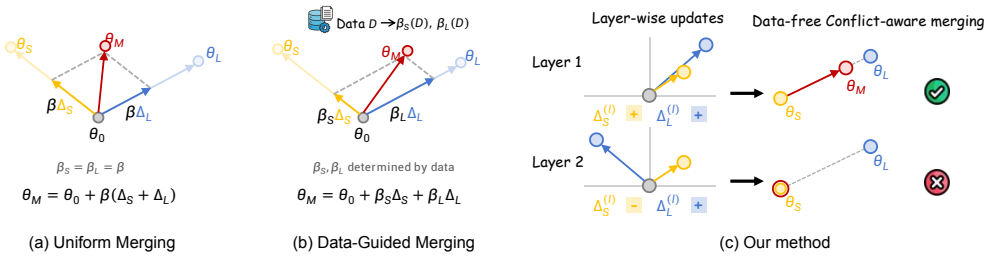

Figure 1: Paradigm comparison: prior long-to-short merging methods apply uniform coefficients or rely on auxiliary data or activation statistics. In contrast, our method is data-free, selecting layers and allocating coefficients via parameter-space scores.

the field of long-to-short reasoning, model merging combines a short model with a long-CoT model in parameter space to enhance effective reasoning while reducing unnecessary thinking procedure. Existing works develop activation-guided or calibration-guided layer-wise coefficients (Wu et al., 2025; Yao et al., 2025; Zhong et al., 2026). However, these methods either apply uniform coefficients across layers or remain dependent on auxiliary data or activation statistics to derive adaptive coefficients. Consequently, they do not explicitly characterize the conflict between long-CoT and short model: a layer may contain a strong long-CoT update while simultaneously posing a high risk of disrupting short-model behavior. It therefore remains unclear whether the model parameters contain structural patterns that can distinguish beneficial long-CoT updates from those that conflict with short-model behavior. This raises a more fundamental question: *What parameter-space patterns reveal the potential reasoning benefit and conflict of long-CoT updates, and how can these patterns enable principled long-to-short merging without data or activation signals?*

To address this problem, we introduce BCL-Merge, a data-free long-to-short merging framework. Our key observation is that model parameters exhibit two layer-wise patterns: (1) ***strength*** of CoT update, which serves as a proxy for potential reasoning benefit, and (2) ***conflict*** with short model, which reflects the risk of disrupting short-model behavior. Intuitively, a strong long-CoT update may also conflict substantially with corresponding short-model update, and the magnitude alone is insufficient to determine whether a layer is suitable for transfer. BCL-Merge therefore jointly considers both properties by defining strength and conflict scores, combining them to select a subset of layers under a given reasoning budget, and assigning proper coefficients for different layers that resolve conflict to the minimal. Under assumptions on short-model stability and reasoning-signal alignment, we show that the resulting selection and coefficient allocation optimize an entropy-regularized proxy objective that balances potential reasoning benefit against conflict disruption.

We evaluate our method across multiple reasoning benchmarks. Specifically, our method achieves 46.3% average accuracy on Qwen2.5-1.5B, outperforming the strongest baseline by 2.9% while reducing average response length by 81.5% relative to long-CoT model. Notably, it also surpasses performance of long-CoT model by 5.3%. The improvements remain consistent and substantial across different budgets, model scales and backbones. Further analysis shows that it preserves concise responses on easy examples while improving accuracy on hard ones, validating the roles of strength and conflict in identifying layers and assigning coefficients. Our contributions are as follows:

- We characterize long-to-short model merging from a parameter-space perspective and identify strength and conflict as crucial layer-wise patterns in long-to-short merging.
- We propose a data-free merging algorithm that translates the identified parameter-space patterns into scores that select layers under a given reasoning budget, and assigns coefficients to resolve conflict to the minimal, which we derive as the solution to an entropy-regularized proxy objective.
- We conduct comprehensive experiments across various benchmarks, budgets, model scales and backbones with substantial and consistent promotion, and further analysis validates the roles of patterns in allocating coefficients, showing its effectiveness and generalizability.

2 Related Work

Model merging. Model merging aims to equip one model with the ability to handle multiple tasks by integrating models in parameter space. Model soups (Wortsman et al., 2022) establishes that

compatible fine-tuned models can be averaged, while Fisher-weighted averaging (Matena & Raffel, 2022) introduces parameter importance into merging. Task arithmetic (Ilharco et al., 2022) represents adaptation as a displacement from base model, which makes addition and negation of task vectors possible. Ties-Merging (Yadav et al., 2023) and DARE (Yu et al., 2024) address distinct forms of interference through sign resolution and sparsification. Subsequent work uses activation statistics and sensitivity to refine the merge (Nobari et al., 2025; Liu et al., 2025). Our method builds on task-vector composition but addresses a different objective that how to transfer reasoning capability of a long-CoT model to short model through model merging.

Efficient reasoning. Chain-of-thought prompting, self-consistency, search, tool use, and verbal feedback improve multi-step inference by increasing or structuring intermediate reasoning (Wei et al., 2022; Wang et al., 2022; Yao et al., 2023; 2022; Schick et al., 2023; Shinn et al., 2023; Madaan et al., 2023). Long-CoT reasoning models make extended intermediate traces especially useful for difficult tasks (Guo et al., 2025). A complementary line controls reasoning cost through budget-aware decoding, pruning, length-aware training, and adaptive triggering (Han et al., 2025; Chen et al., 2024; Luo et al., 2026; Ma et al., 2025; Aggarwal & Welleck, 2025; Lou et al., 2025; Cao et al., 2026). These approaches adjust computation during training or inference. Our method achieves the same efficiency goal by integrating abilities of two models into a static model, without an input-time controller.

Long-to-short reasoning through model merging. Within efficient reasoning, model merging offers a feasible route for transferring useful long-CoT behavior into a short model without additional training. Wu *et al.* (Wu et al., 2025) evaluate task-vector, low-rank, and activation-informed merging for this setting. ACM (Yao et al., 2025) derives a plug-in layer coefficient from activation consensus, whereas RPAM (Zhong et al., 2026) uses a pattern-labeled calibration set and representation alignment to learn layer-wise coefficients. However, existing methods mainly rely on activation statistics to obtain better merging coefficients. By contrast, our method derives its layer ranking from parameter space without calibration examples, activations, gradients, or learned coefficients.

3 METHOD

3.1 PRELIMINARIES AND MOTIVATION

Preliminaries. Denote the base model, short model, and long-CoT model as θ_0, θ_S and θ_L. respectively. The short and long task vectors are:

$$\Delta_S = \theta_S - \theta_0, \qquad \Delta_L = \theta_L - \theta_0. \tag{1}$$

Generally, the merging process for layer l under a given budget is written as:

$$\theta_M^{(l)} = \theta_0^{(l)} + (1 - \beta_l)\Delta_S^{(l)} + \beta_l \Delta_L^{(l)}, \qquad 0 \le \beta_l \le M_l, \qquad M_l \in \{0, 1\}, \tag{2}$$

where M_l is the budget mask controlling the selection of layer l and β_l is the layer-wise merging coefficient. The mask and coefficient therefore separate layer selection from reasoning strength allocation. $\beta_l = 0$ and $\beta_l = 1$ represent the layers of short and long-CoT models, respectively.

Motivation. The central challenge in data-free long-to-short merging is to determine where long-CoT behavior can be introduced without substantially disrupting the behavior of the short model using statistics of model parameters alone. Uniform interpolation cannot make this distinction because it assigns the same coefficient to every layer, implicitly treating all long-CoT updates as equally beneficial without conflict. In practice, however, layers exhibit substantial heterogeneity: a long-CoT update may contain a strong transferable signal at certain layers, but it may also oppose the corresponding short-model update. Therefore, the strength of update alone is insufficient for deciding which layers should receive long-CoT influence. A data-free merging criterion must characterize both potential strength of a long-CoT update and its risk of conflict with short-model update.

Two layer-wise patterns in parameter space. Motivated by this, we aim to characterize each layer using two complementary patterns in parameter space. (1) The first is *strength of CoT update*, which serves as a proxy for the amount of long CoT model reasoning ability available for transfer. The second is *conflict with short model*, which measures the risk of disrupting short-model behavior when merging CoT model. Both patterns are crucial, as a strong long-CoT update may be highly conflicting, whereas a low-conflict update may contain little transferable reasoning signal. Therefore,

Table 1: Four types of merging policies categorized by conflict and strength.

Category	Conflict Score	Strength Score	Merge policy
Reasoning compatible	low	high	assign more long update
Conflict reasoning	high	high	restrict long update
Short dominant	high	low	retain short model
Low-signal	low	low	retain short model

useful merging CoT update requires jointly considering long-update strength and long-short conflict. As summarized in Table 1, layers with strong yet low-conflict long-CoT updates should receive the highest priority, while conflict or weak long updates should reduce their weight. Section 3.2 below formalizes these patterns as the strength score of CoT update and conflict score of short model and Figure 2 illustrates the framework.

3.2 Parameter-Space Strength–Conflict Scoring

Conflict between CoT and short model. The task vectors of long-CoT and short model may contain coordinates that move away from the base model in opposite directions. Interpolating such coordinates toward long-CoT model can overwrite parameter changes of the short model, producing interference. We first define a thresholded sign function to avoid potential noise:

$$\mathrm{sign}_\tau(x) = \begin{cases} \mathrm{sign}(x), & |x| > \tau \\ 0, & |x| \le \tau \end{cases}, \tag{3}$$

where τ is the threshold. For layer l, we compare only the parameters at which both task vectors contain non-negligible updates:

$$\mathcal{I}_l(\tau) = \left\{ i \in l : \mathrm{sign}_\tau(\Delta_S^i) \neq 0 \text{ and } \mathrm{sign}_\tau(\Delta_L^i) \neq 0 \right\}. \tag{4}$$

where i represents individual parameter within layer l. Restricting the comparison to $\mathcal{I}_l(\tau)$ prevents the merging model from being interfered by noise. Then the layer-wise conflict score is defined as:

$$C_l = \frac{1}{|\mathcal{I}_l(\tau)|} \sum_{i \in \mathcal{I}_l(\tau)} \mathbb{I}\left[\mathrm{sign}_\tau(\Delta_S^i) \neq \mathrm{sign}_\tau(\Delta_L^i) \right]. \tag{5}$$

Intuitively, the score measures the proportion of effective update for which the short and long task vetcor point in opposite directions. Therefore, a large conflict score indicates that moving this layer toward its long-CoT counterpart is more likely to interfere with the short-model update.

Strength of long-CoT update. Low conflict alone does not imply that a layer contains useful reasoning information. For example, the two task vectors may exhibit low conflict score simply because both updates are weak. We therefore additionally measure the strength score by the relative magnitude of long-CoT task vector:

$$R_l = \frac{||\Delta_L^{(l)}||_2}{||\Delta_S^{(l)}|_2 + |\Delta_L^{(l)}||_2}. \tag{6}$$

Intuitively, it compares the long update against the short update within the same layer and reduces sensitivity to scale differences across layers. A large strength score means that the long-CoT model introduces a relatively strong parameter change at layer l and can be interpreted as a proxy for the amount of long CoT model reasoning ability available for transfer.

Strength–conflict score. The two patterns describe different but important requirements for effective transfer of reasoning ability. Concretely, a layer with large strength score may contain substantial long-CoT signal but remain interfered if its conflict score is also large. Conversely, a low-conflict layer may be safe to modify but provide little reasoning capability when strength score is small. We therefore combine the scores as:

$$s_l = R_l(1 - C_l). \tag{7}$$

This construction ensures that a high score is assigned only when a layer simultaneously exhibits a strong long-CoT update and limited conflict with the short update. The rationality of employing the two scores as criterion are provided in Appendix A.

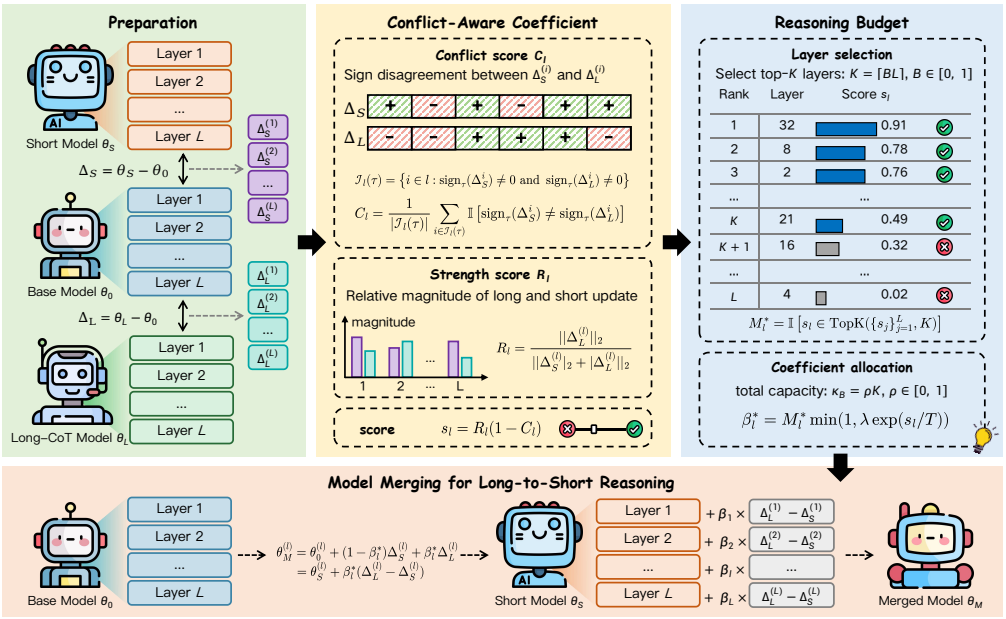

Figure 2: Overall framework. For each layer, s_l is computed by combining strength and conflict score. Under given budget, a hard mask M_l selects the top-K layers, followed by a capacity allocation that determines the merging coefficient for merging process.

3.3 Budgeted Layer Selection and Coefficient Allocation

The score determines the relative priority of each layer, but it does not directly control how the way reasoning behavior introduced. Specifically, moving every layer toward the long-CoT model can introduce unnecessary verbosity and overwriting short behavior. We therefore formulate merging as a two-stage budgeted allocation problem: first selecting which layers long-CoT influence is introduced, and then determining how much influence each selected layer receives.

Reasoning budget. Let L denote the total number of layers and $B \in [0, 1]$ be a prescribed reasoning budget, we convert B into a selected size: $K = \lceil BL \rceil$, which specifies the number of layers allowed to move toward their long-CoT counterparts. Moreover, selecting K layers does not yet determine how far they should move toward the long-CoT model. We therefore define the total capacity $\kappa_B = \rho K$, where ρ controls the average coefficient mass available for selected layers.

layer selection and coefficient allocation. We formulate the merging layer and coefficient of CoT models as the entropy-regularized coefficient allocation optimization:

$$\max_{M,\beta} \quad J_{\text{proxy}}(M, \beta) = \sum_{l=1}^{L} \beta_l s_l + \kappa_B T \sum_{l=1}^{L} \frac{\beta_l}{\kappa_B} \log \frac{\kappa_B}{\beta_l}, \tag{8}$$

subject to **(1)** layer constraint: $\sum_{l=1}^{L} M_l = K,\ M_l \in \{0, 1\}$, where binary mask $M_l = 1$ indicates that layer l is selected; **(2)** capacity constraint: $\sum_{l=1}^{L} \beta_l = \kappa_B,\ 0 \leq \beta_l \leq M_l$, which distributes the fixed capacity across the selected layers.

where T is the temperature. The first term allocates more long-CoT influence to layers with larger scores. The second term is the Shannon entropy of normalized coefficients, which prevents the entire capacity from collapsing onto a single layer and yields a smooth allocation over selected layers.

Proposition 1 (Optimality of Budgeted Coefficient Allocation)**.** *With the score combining strength and conflict proxies, the optimal mask selects the K layers with the largest scores:*

$$M_l^* = \mathbb{I}\left[s_l \in \mathrm{TopK}(\{s_j\}_{j=1}^L, K)\right], \tag{9}$$

where the constraint $\beta_l \leq 1$, the optimal coefficient has the capped softmax form:

$$\beta_l^* = M_l^* \min(1, \lambda \exp(s_l/T)), \tag{10}$$

where the scalar λ satisfies the capacity constraint.

Table 2: Evaluation of Qwen-1.5B on mathematical reasoning benchmarks. Each method spans two rows: accuracy is reported first, followed by average response length.

Method	GSM8K	MATH500	Minerva Math	Olympiad Bench	College Math	AIME24	**Avg.**
Baselines							
Qwen2.5-Math-1.5B	75.1	35.5	11.0	22.0	5.0	0.0	24.8
	(115)	(408)	(1030)	(600)	(860)	(860)	
DeepSeek-R1-1.5B	76.6	69.6	15.1	30.4	34.2	20.0	41.0
	(2743)	(4508)	(6374)	(7389)	(4058)	(8952)	
Merging methods							
Task Arithmetic	74.5	62.6	21.0	28.6	28.9	10.0	37.6
	(552)	(1622)	(1673)	(2590)	(1324)	(3398)	(-67.2%)
TIES-Merging	75.7	66.2	22.1	30.8	29.6	10.0	39.1
	(417)	(1041)	(1069)	(1966)	(879)	(2904)	(-75.7%)
Sens-Merging	71.3	63.8	24.8	28.9	30.5	13.3	38.8
	(723)	(1614)	(1776)	(2686)	(1397)	(3550)	(-65.5%)
ACM-Ties	78.4	71.4	28.7	33.8	37.9	10.0	43.4
	(965)	(1238)	(1353)	(1852)	(851)	(2692)	(-73.7%)
Ours	**81.8**	**72.2**	**30.0**	**35.0**	**39.0**	**20.0**	**46.3**
	(336)	(668)	(1145)	(1920)	(790)	(1440)	**(-81.5%)**

Remark. Proposition 1 provides a closed-form solution to the proxy optimization problem defined in Eq. (8), thereby turning it into a practical merging procedure. By formulating the as a entropy-regularized allocation optimization problem, the conflict and strength scores provide parameter-space proxies for risk and benefit, thereby determining the top-K mask selection and capped softmax coefficient allocation under given reasoning budget. Intuitively, if $M_l^* = 0$, the constraint gives $\beta_l^* = 0$, so the layer is derived directly from the short model. If $M_l^* = 1$, the capped-softmax coefficient determines how far the layer moves from short model toward CoT model. Full proofs are provided in Appendix A.

After obtaining the allocated coefficients, the merging process directly incorporates them as merging coefficients for layer l as follows:

$$\begin{aligned}\theta_M^{(l)} &= \theta_0^{(l)} + (1-\beta_l^*)\Delta_S^{(l)} + \beta_l^*\Delta_L^{(l)} = (1-\beta_l^*)\theta_S^{(l)} + \beta_l^*\theta_L^{(l)} \\ &= \theta_S^{(l)} + \beta_l^*(\Delta_L^{(l)} - \Delta_S^{(l)}).\end{aligned} \tag{11}$$

The layer selection and coefficient allocation are performed on each parameter of the model across layers, and obtain the final merged model. It is notable that although Eq. (8) is written as an optimization problem, its top-K and capped-softmax solution is available in closed form. Consequently, our method requires neither training nor activation statistics, showcasing its effectiveness.

4 Experiments

4.1 Setup

Models and implementation details. The main setting uses Qwen2.5-Math-1.5B, Qwen2.5-Math-1.5B-Instruct (Yang et al., 2024), and DeepSeek-R1-1.5B (Guo et al., 2025) as base, short and long-CoT model, respectively. Qwen2.5-Math-7B, LLaMA-8B (Grattafiori et al., 2024) and Qwen3-4B (Yang et al., 2025) are incorporated. The results in main experiment uses B=0.5, and T and ρ are set to be 0.1 and 0.5 by default. More implementations details are provided in Appendix B.

Datasets. We evaluate BCL-Merge on six mathematical reasoning benchmarks: GSM8K (Cobbe et al., 2021), MATH500 (Hendrycks et al., 2021; Lightman et al., 2024), Minerva Math (Lewkowycz et al., 2022), OlympiadBench (He et al., 2024), College Math (Tang et al., 2024), and AIME 2024 (Math-AI, 2024). For difficulty-aware evaluation, easy examples are those answered correctly by the short model, whereas hard examples are those answered incorrectly by the short model.

Baselines and evaluation metrics. We compare our method with short model, long-CoT model, Task Arithmetic (Ilharco et al., 2022), TIES-Merging (Yadav et al., 2023), DARE (Yu et al., 2024), Sens-Merging (Liu et al., 2025) and ACM (Yao et al., 2025). We report accuracy and average output

Table 3: Accuracy and average output length on the 7B setting. Each method spans two rows: accuracy is reported first, followed by average response length in parentheses.

Method	GSM8K	MATH500	Minerva Math	Olympiad Bench	College Math	AIME24	**Avg.**
Baselines							
Qwen2.5-Math-7B	88.5	51.5	12.0	17.0	22.0	3.3	32.4
	(128)	(520)	(950)	(1250)	(790)	(1520)	
DeepSeek-R1-7B	89.3	87.4	36.4	51.0	39.8	23.3	54.5
	(1062)	(2826)	(3056)	(5794)	(2462)	(8676)	
Merging methods							
Task Arithmetic	90.5	83.4	**41.9**	45.0	40.3	20.0	53.5
	(620)	(1620)	(1418)	(2653)	(1446)	(3598)	(-52.4%)
TIES-Merging	90.6	81.8	38.2	43.0	41.9	33.3	54.8
	(555)	(1495)	(1352)	(2476)	(1290)	(3305)	(-56.1%)
DARE-Ties	89.6	82.4	37.5	41.3	40.2	23.3	52.4
	(584)	(1589)	(1307)	(2655)	(1378)	(3579)	(-50.4%)
Sens-Merging	91.2	83.4	41.5	43.9	40.2	30.0	55.0
	(607)	(1616)	(1405)	(2676)	(1457)	(3360)	(-53.4%)
ACM-Ties	92.2	84.0	38.6	46.4	40.3	33.3	55.8
	(541)	(1453)	(1156)	(2360)	(1288)	(3080)	(-58.6%)
Ours	**92.5**	**85.0**	41.8	**48.2**	**42.6**	**36.7**	**57.8**
	(530)	(1420)	(1115)	(2290)	(1255)	(3010)	**(-59.7%)**

length for each benchmark, measure the length reduction relative to long-CoT model, and compare methods on the accuracy-length trade-off.

4.2 Main Results

Results on Qwen2.5-1.5B. We evaluate the effectiveness of our method on 1.5B model across six reasoning benchmarks. As shown in Table 2, BCL-Merge achieves 46.3% average accuracy across benchmarks, substantially and consistently outperforming existing merging baseline by 2.9%. Notably, it also surpasses the long-CoT model by 5.3% points, while reducing the average response length by 81.5%. These results show that the proposed conflict-aware layer selection and allocation can discriminate and transfer useful reasoning behavior with parameter-space criterion, while avoiding the excessive verbosity of the long-CoT model, demonstrating the effectiveness.

Scaling to 7B setting. To verify scalability to larger models, we extend the evaluation to 7B setting. It can be concluded from Table 3 that in the Qwen2.5-7B setting, our method reaches 57.8% average accuracy, exceeding the strongest merging baseline by 2.0% and the long-CoT model by 3.3%. At the same time, it reduces the average response length by 59.7% relative to the long-CoT model, exhibiting the efficiency. The gain is consistent across various benchmarks, validating that the conflict-aware criterion remains effective at a larger scale.

Budgeted accuracy–length trade-off. Figure 3 illustrates how the reasoning budget B controls the trade-off between reasoning accuracy and response length. Increasing B expands both the number of selected layers and the total available capacity, progressively introducing more long-CoT ability and moving the model towards higher accuracy with longer responses. Meanwhile, intermediate budgets concentrate the available capacity on strong, low-conflict layers, outperforming the full-budget setting while producing shorter responses. Across different values of B, our approach exhibits a more favorable accuracy–length frontier than existing merging baselines, demonstrating the effectiveness that the reasoning budget provides an explicit control over the merged model.

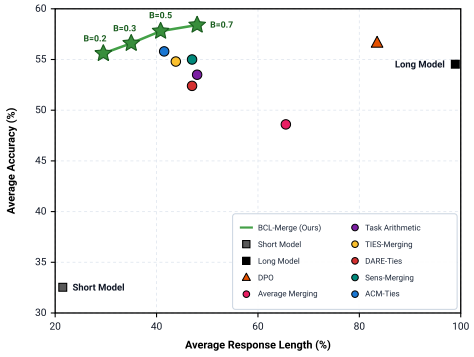

Figure 3: Accuracy-length trade-off.

Table 4: Generalization to the LLaMA-8B backbone.

Method	GSM8K	MATH500	AIME24	**Avg.**	**Avg. Len**
Baselines					
Llama-3.1-8B	83.6	58.4	16.7	52.9	540
DeepSeek-R1-8B	88.9	73.8	26.7	63.1	3540
Merging methods					
Linear interpolation	85.8	66.5	20.0	57.4	1325
Task Arithmetic	86.7	68.2	20.0	58.3	1480
Ties-Merging	87.1	69.0	23.3	59.8	1395
DARE-Ties	86.4	68.6	23.3	59.4	1518
Sens-Merging	88.0	71.2	26.7	62.0	1605
Ours	**89.2** (+1.2)	**73.0** (+1.8)	**30.0** (+3.3)	**64.1** (+2.1)	**1315**

Table 5: Effectiveness of each component. Easy subset is defined as questions correctly answered by short model.

Variant	GSM8K	MATH500	Easy Len	Hard Acc
Short model	88.5	51.5	118	0.0
Random selection	90.0	81.2	1468	66.0
High-conflict	88.3	79.5	2015	65.0
w/o conflict score	91.8	83.7	1585	71.0
w/o strength score	91.0	82.5	1040	68.5
Ours	**92.5**	**85.0**	**432**	**74.0**

Table 6: Difficulty-aware evaluation. Easy and hard subsets are split by short model.

Method	Easy		Hard	
	Acc	Len	Acc	Len
DeepSeek-R1-7B	97.2	1880	78.6	6520
Ties-Merging	96.6	690	70.2	2860
Sens-Merging	96.8	742	72.8	2915
ACM-Ties	96.9	718	73.0	2890
Ours	**97.3**	**432**	**74.0**	**2525**

Generalization across different backbones. We furthermore conduct experiments on LLaMA-8B (Grattafiori et al., 2024) to investigate whether the proposed parameter-space patterns transfer beyond the Qwen models. It can be found in Table 4 that while existing methods get performance inferior to long-CoT models, our method achieves 64.1% average accuracy, outperforming the strongest merging baseline by 2.1% and long-CoT model by 1.0%. Its average response length is 1,315 tokens, which is lower than that of compared merging method and 62.9% shorter than long-CoT model. This results strongly demonstrate that the effectiveness of our method does not depend on specific models, but follows layer-wise patterns that can generalize well to different backbones.

4.3 Ablation Study and Further Analysis

Effectiveness of each component. We conduct ablation study in Table 5 to analyze roles of different component. It can be seen that the inferior performance of high-conflict and random selection variants show the necessity of conflict resolution and layer selection, respectively. Furthermore, removing conflict score or strength score improves accuracy on hard questions, but substantially increases the output length on easy questions, showing that either of them alone cannot capture effective merging signal. In contrast, our method achieves the highest hard-task accuracy with shortest easy-subset response length by incorporating all components, which confirms better accuracy–length trade-off of the proposed method.

Difficulty-aware analysis. To better understand the influence of the patterns, we partition the questions into easy and hard subset by whether they can be answered by short model, and separately evaluate the behavior of different methods. Table 6 show that existing merging methods moderately reduce the response length of long-CoT model accompanied by a substantial accuracy degradation. By contrast, our approach achieves 97.3% accuracy with response length of 432 tokens on the easy subset, compared with 97.2% accuracy and 1880 tokens for the long-CoT model. Moreover, on the hard subset, it reaches 74.0% accuracy, outperforming all merging baselines while reducing the average response length to 2525 tokens, substantially below long-CoT model. It is consistent with the proposed selection and allocation strategy that low-conflict layers preserve behavior on easy questions, whereas strong long-CoT provides additional reasoning capacity for hard questions, greatly showing the effectiveness.

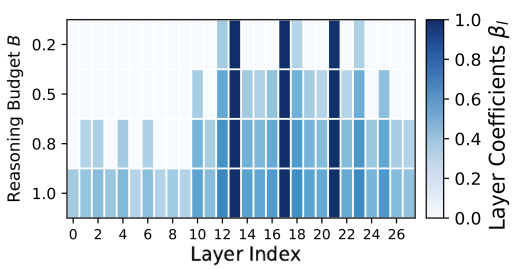

Figure 4: Heatmap of layer-wise merging coefficients across different reasoning budgets.



Figure 5: Distribution of strength and conflict scores. Orange dashed boundaries divide four regions, and green represents selected layers.

Hyperparameter robustness. Table 7 studies the sensitivity to temperature T, and capacity ρ. Across the values, the accuracy remains stable, indicating that the method does not depend on specific hyperparameters. Specifically, lower temperature concentrates the coefficient mass too aggressively, whereas larger values dilute it across more layers, and 0.1 provides a balanced allocation; The capacity exposes the expected accuracy–length trade-off. Increasing ρ slightly raises the mean accuracy, but also increases the average response length. We therefore use 0.5 in main experiments.

Table 7: Sensitivity analysis of hyperparameters.

Setting	Value	GSM8K	MATH500	Avg. Len	**Avg.**
T	0.05	92.2	84.6	1405	57.5
	0.1	**92.5**	**85.0**	**1420**	**57.8**
	0.2	92.4	84.8	1435	57.6
	0.5	92.0	84.2	1460	57.2
ρ	0.3	91.8	83.9	1210	56.9
	0.5	**92.5**	**85.0**	**1420**	**57.8**
	0.7	92.6	85.2	1680	58.0
	0.9	92.7	85.3	1950	58.1

Visualization of strength–conflict scores. Figure 4 qualitatively visualizes the layer-wise coefficients across different reasoning budgets to reveal the influence. It can be concluded that at a small budget, the coefficients are concentrated on a small number of layers with the most favorable scores. As the budget increases, additional layers are introduced progressively, while the coefficients remain non-uniform across layers focusing on deeper layers, indicating that reasoning ability mainly lies in deeper layers. It strongly shows that our method expands the selected layers and reallocates reasoning capacity according to the scores.

Figure 5 furthermore provides a direct distribution of the two underlying parameter-space signals. The two scores exhibit distinct and non-uniform patterns: long-update strength is generally higher in the middle and deeper layers. At the same time, conflict fluctuates across depth and can become large even for layers with strong long-CoT updates, showing why update strength alone is insufficient for layer selection. The orange dashed boundaries divide the layers into four regions corresponding to Table 1. By jointly considering the two scores, our method selects the reasoning-compatible layers highlighted in green, which retain relatively strong long-CoT updates while avoiding the conflict peaks observed in neighboring layers. This visualization provides an intuitive validation of the two scores: strength identifies where reasoning benefit is available, whereas conflict determines whether that benefit can be transferred without disruption, thereby confirming the rationality of our method.

5 Conclusion

In this paper, we investigate data-free model merging for long-to-short reasoning. Our analysis identifies two parameter-space patterns: long-update strength reflects potential reasoning benefit, while conflict with short model indicates the risk of disrupting behavior. Based on these observations, we propose BCL-Merge, which prioritizes strong yet low-conflict long-CoT updates and performs budgeted layer selection and coefficient allocation using only parameter-space criterion. Under explicit assumptions, we connect this procedure to an entropy-regularized proxy objective that balances reasoning capabilities against conflict disruption. Experiments across multiple reasoning benchmarks and multiple backbones demonstrate substantial accuracy improvements with reduced response length. Further analyses confirm that our method preserves concise behavior on easy examples while improving reasoning on hard ones, highlighting parameter-only conflict-aware merging as an effective methods for long-to-short reasoning.

AI USE STATEMENT

In this work, we used generative AI tools to improve the clarity, grammar, and presentation of the manuscript. All AI-assisted content was carefully reviewed and verified by the authors. All scientific ideas, methodological designs, experimental results, and conclusions were developed and validated by the authors. We take full responsibility for the final content of this work.

REFERENCES

Pranjal Aggarwal and Sean Welleck. L1: Controlling how long a reasoning model thinks with reinforcement learning. *arXiv preprint arXiv:2503.04697*, 2025.

Jie Cao, Tianwei Lin, Zhenxuan Fan, Bo Yuan, Ziyuan Zhao, Rolan Yan, Wenqiao Zhang, and Siliang Tang. Draft-thinking: Learning efficient reasoning in long chain-of-thought llms. *arXiv preprint arXiv:2603.00578*, 2026.

Xingyu Chen, Jiahao Xu, Tian Liang, Zhiwei He, Jianhui Pang, Dian Yu, Linfeng Song, Qiuzhi Liu, Mengfei Zhou, Zhuosheng Zhang, et al. Do not think that much for 2+ 3=? on the overthinking of o1-like llms. *arXiv preprint arXiv:2412.21187*, 2024.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, et al. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

Tingxu Han, Zhenting Wang, Chunrong Fang, Shiyu Zhao, Shiqing Ma, and Zhenyu Chen. Token-budget-aware llm reasoning. In *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 24842–24855, 2025.

Chaoqun He, Renjie Luo, Yuzhuo Bai, Shengding Hu, Zhen Thai, Junhao Shen, Jinyi Hu, Han Xu, Yujie Huang, Yuxiang Zhang, et al. Olympiadbench: A challenging benchmark for promoting agi with olympiad-level bilingual multimodal scientific problems. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, pp. 3828–3850, 2024.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the math dataset. *arXiv preprint arXiv:2103.03874*, 2021.

Gabriel Ilharco, Marco Tulio Ribeiro, Mitchell Wortsman, Suchin Gururangan, Ludwig Schmidt, Hannaneh Hajishirzi, and Ali Farhadi. Editing models with task arithmetic. *arXiv preprint arXiv:2212.04089*, 2022.

Aitor Lewkowycz, Anders Andreassen, David Dohan, Ethan Dyer, Henryk Michalewski, Vinay Ramasesh, Slone Ambrose, Imanol Schlag, Tehseen Gutman-Solo, Yuhuai Wu, et al. Solving quantitative reasoning problems with language models. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Hunter Lightman, Vineet Kosaraju, Yuri Burda, Harrison Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. In *International Conference on Learning Representations*, volume 2024, pp. 39578–39601, 2024.

Shuqi Liu, Han Wu, Bowei He, Xiongwei Han, Mingxuan Yuan, and Linqi Song. Sens-merging: Sensitivity-guided parameter balancing for merging large language models. In *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 19243–19255, 2025.

Chenwei Lou, Zewei Sun, Xinnian Liang, Meng Qu, Wei Shen, Wenqi Wang, Yuntao Li, Qingping Yang, and Shuangzhi Wu. Adacot: Pareto-optimal adaptive chain-of-thought triggering via reinforcement learning. *arXiv preprint arXiv:2505.11896*, 2025.

Haotian Luo, Haiying He, Yibo Wang, Shiwei Liu, Wei Li, Xiaochun Cao, Dacheng Tao, Naiqiang Tan, and Li Shen. O1-pruner: Length-harmonizing fine-tuning for o1-like reasoning pruning. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 14242–14257, 2026.

Xinyin Ma, Guangnian Wan, Runpeng Yu, Gongfan Fang, and Xinchao Wang. Cot-valve: Length-compressible chain-of-thought tuning. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 6025–6035, 2025.

Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, et al. Self-refine: Iterative refinement with self-feedback. *Advances in neural information processing systems*, 36:46534–46594, 2023.

Michael S Matena and Colin Raffel. Merging models with fisher-weighted averaging. *Advances in Neural Information Processing Systems*, 35:17703–17716, 2022.

Math-AI. AIME 2024. `https://huggingface.co/datasets/Maxwell-Jia/AIME_2024`, 2024.

Tergel Munkhbat, Namgyu Ho, Seo Hyun Kim, Yongjin Yang, Yujin Kim, and Se-Young Yun. Self-training elicits concise reasoning in large language models. In *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 25127–25152, 2025.

Amin Heyrani Nobari, Kaveh Alim, Ali ArjomandBigdeli, Akash Srivastava, Faez Ahmed, and Navid Azizan. Activation-informed merging of large language models. *arXiv preprint arXiv:2502.02421*, 2025.

Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language models can teach themselves to use tools. *Advances in neural information processing systems*, 36:68539–68551, 2023.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning. *Advances in neural information processing systems*, 36:8634–8652, 2023.

Zhengyang Tang, Xingxing Zhang, Benyou Wang, and Furu Wei. Scaling instruction tuning for mathematical reasoning. In *Proceedings of the 41st International Conference on Machine Learning*, 2024.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*, 2022.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

Mitchell Wortsman, Gabriel Ilharco, Samir Ya Gadre, Rebecca Roelofs, Raphael Gontijo-Lopes, Ari S Morcos, Hongseok Namkoong, Ali Farhadi, Yair Carmon, Simon Kornblith, et al. Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. In *International conference on machine learning*, pp. 23965–23998. PMLR, 2022.

Han Wu, Yuxuan Yao, Shuqi Liu, Zehua Liu, Xiaojin Fu, Xiongwei Han, Xing Li, Hui-Ling Zhen, Tao Zhong, and Mingxuan Yuan. Unlocking efficient long-to-short llm reasoning with model merging. *arXiv preprint arXiv:2503.20641*, 2025.

Heming Xia, Chak Tou Leong, Wenjie Wang, Yongqi Li, and Wenjie Li. Tokenskip: Controllable chain-of-thought compression in llms. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 3351–3363, 2025.

Prateek Yadav, Derek Tam, Leshem Choshen, Colin A Raffel, and Mohit Bansal. Ties-merging: Resolving interference when merging models. *Advances in neural information processing systems*, 36:7093–7115, 2023.

An Yang, Beichen Zhang, Binyuan Hui, Bofei Gao, Bowen Yu, Chengpeng Li, Dayiheng Liu, Jianhong Tu, Jingren Zhou, Junyang Lin, et al. Qwen2. 5-math technical report: Toward mathematical expert model via self-improvement. *arXiv preprint arXiv:2409.12122*, 2024.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2022.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. *Advances in neural information processing systems*, 36:11809–11822, 2023.

Yuxuan Yao, Shuqi Liu, Zehua Liu, Qintong Li, Mingyang Liu, Xiongwei Han, Zhijiang Guo, Han Wu, and Linqi Song. Activation-guided consensus merging for large language models. *Advances in Neural Information Processing Systems*, 38:124041–124069, 2025.

Le Yu, Bowen Yu, Haiyang Yu, Fei Huang, and Yongbin Li. Language models are super mario: Absorbing abilities from homologous models as a free lunch. In *Forty-first International Conference on Machine Learning*, 2024.

Zhaofeng Zhong, Wei Yuan, Tong Chen, Xiangyu Zhao, Quoc Viet Hung Nguyen, and Hongzhi Yin. Reasoning pattern alignment merging for adaptive reasoning. *arXiv preprint arXiv:2601.03506*, 2026.

A Full Proofs

A.1 Assumptions

Assumption A.1 (Local short-model optimality and smoothness)**.** *The short-answer loss $\mathcal{L}_S(\theta)$ is twice differentiable near θ_S, $\nabla\mathcal{L}_S(\theta_S) \approx 0$, and for $d = \theta - \theta_S$:*

$$\mathcal{L}_S(\theta_S + d) = \mathcal{L}_S(\theta_S) + \frac{1}{2} d^T H_S d + o(\|d\|_2^2), \tag{12}$$

with layer-wise spectral norm $\|H_S^{(l)}\|_2 \le \Lambda_l$.

Assumption A.2 (Layer-wise curvature approximation)**.** *The curvature of the short-answer loss admits a layer-wise block approximation, so cross-layer Hessian terms are treated as approximation error.*

Assumption A.3 (Long-update decomposition)**.** *Each long update can be decomposed as:*

$$\Delta_L^{(l)} = \Delta_R^{(l)} + \Delta_V^{(l)} + \Delta_N^{(l)}, \tag{13}$$

where the three components correspond to reasoning-critical, verbosity-inducing, and noise directions. The contamination condition is:

$$\mathbb{E}\left[\left\|\Delta_V^{(l)} + \Delta_N^{(l)}\right\|_2\right] \le \eta_l \mathbb{E}\left[\left\|\Delta_R^{(l)}\right\|_2\right]. \tag{14}$$

The conditional expectation $\mathbb{E}[\|\Delta_R^{(l)}\|_2 \mid \|\Delta_L^{(l)}\|_2 = t]$ is monotone nondecreasing in t.

Assumption A.4 (Hard-utility alignment)**.** *The reasoning component has positive expected alignment with the hard-task utility gradient:*

$$\mathbb{E}\left[\left\langle \nabla_{\theta^{(l)}} U_H(\theta_S), \Delta_R^{(l)} \right\rangle\right] \ge c_R \mathbb{E}\left[\left\|\Delta_R^{(l)}\right\|_2\right], \qquad c_R > 0, \tag{15}$$

with bounded layer-wise Hessian $\|H_U^{(l)}(\theta)\|_2 \le L_H$ near θ_S.

A.2 Lemmas

Lemma 1 (Merge-to-Short Displacement)**.** *For the BCL-Merge family,*

$$\theta_M^{(l)} - \theta_S^{(l)} = \beta_l(\Delta_L^{(l)} - \Delta_S^{(l)}). \tag{16}$$

Proof. Since $\theta_S^{(l)} = \theta_0^{(l)} + \Delta_S^{(l)}$, we have:

$$\begin{aligned} \theta_M^{(l)} - \theta_S^{(l)} &= \theta_0^{(l)} + (1-\beta_l)\Delta_S^{(l)} + \beta_l \Delta_L^{(l)} - \theta_0^{(l)} - \Delta_S^{(l)} \\ &= \beta_l(\Delta_L^{(l)} - \Delta_S^{(l)}). \end{aligned} \tag{17}$$

□

Lemma 2 (Short-Behavior Influence Bound)**.** *Under Assumptions A.1 and A.2, the bound for the influence of short model is:*

$$\mathcal{L}_S(\theta_M) - \mathcal{L}_S(\theta_S) \le \frac{1}{2}\sum_{l=1}^{L} \beta_l^2 \Lambda_l \|\Delta_L^{(l)} - \Delta_S^{(l)}\|_2^2. \tag{18}$$

For a single layer,

$$\mathcal{L}_S^{(l)}(\theta_M) - \mathcal{L}_S^{(l)}(\theta_S) \le \frac{1}{2}\beta_l^2 \Lambda_l \|\Delta_L^{(l)} - \Delta_S^{(l)}\|_2^2. \tag{19}$$

Proof. Let $d_M = \theta_M - \theta_S$ and $r_l = \Delta_L^{(l)} - \Delta_S^{(l)}$. Lemma 1 gives $d_M^{(l)} = \beta_l r_l$. Assumption A.1 gives:

$$\mathcal{L}_S(\theta_M) - \mathcal{L}_S(\theta_S) \approx \frac{1}{2} d_M^T H_S d_M. \tag{20}$$

Using Assumption A.2,

$$d_M^T H_S d_M \approx \sum_l (d_M^{(l)})^T H_S^{(l)} d_M^{(l)} = \sum_l \beta_l^2 r_l^T H_S^{(l)} r_l. \tag{21}$$

The spectral norm bound gives $r_l^T H_S^{(l)} r_l \le \Lambda_l \|r_l\|_2^2$, proving Eq. (18). Equation (19) follows by considering only layer l. □

Lemma 3 (Sign Conflict as a Conflict Score)**.** *For a single layer,*

$$\|\Delta_L^{(l)} - \Delta_S^{(l)}\|_2^2 = \|\Delta_L^{(l)}\|_2^2 + \|\Delta_S^{(l)}\|_2^2 - 2\langle \Delta_L^{(l)}, \Delta_S^{(l)} \rangle. \tag{22}$$

Define:

$$N_l^+ = \sum_{i:\Delta_S^i \Delta_L^i > 0} \Delta_S^i \Delta_L^i, \qquad N_l^- = \sum_{i:\Delta_S^i \Delta_L^i < 0} |\Delta_S^i \Delta_L^i|. \tag{23}$$

Then we have:

$$\|\Delta_L^{(l)} - \Delta_S^{(l)}\|_2^2 = \|\Delta_L^{(l)}\|_2^2 + \|\Delta_S^{(l)}\|_2^2 - 2N_l^+ + 2N_l^-. \tag{24}$$

Proof. The proof expands the square of task vectors:

$$(\Delta_L^i - \Delta_S^i)^2 = (\Delta_L^i)^2 + (\Delta_S^i)^2 - 2\Delta_L^i \Delta_S^i. \tag{25}$$

Summing over layer l yields Eq. (22). The inner product decomposes into aligned and non-aligned coordinates, $\langle \Delta_L^{(l)}, \Delta_S^{(l)} \rangle = N_l^+ - N_l^-$. Substitution gives Eq. (24). The term $2N_l^-$ is additive and positive, so non-aligned coordinates enlarge the mismatch norm that appears in Lemma 2. The conflict score C_l is an approximation when magnitudes are not extremely imbalanced. □

Lemma 4 (Long Update as a Strength Score)**.** *Holding $\|\Delta_S^{(l)}\|_2$ fixed, R_l is monotone increasing in $\|\Delta_L^{(l)}\|_2$. Under Assumptions A.3 and A.4, R_l is a data-free proxy for relative reasoning benefit.*

Proof. Let $x = \|\Delta_L^{(l)}\|_2$ and $a = \|\Delta_S^{(l)}\|_2 + \epsilon$. Then $R_l = x/(a+x)$ and $\frac{dR_l}{dx} = \frac{a}{(a+x)^2} \ge 0$. When layer norms are comparable or normalized, higher R_l indicates stronger relative long-update dominance. Assumption A.3 links larger long-update norm to a not smaller expected reasoning component. For the hard-utility part, define $\theta_\alpha = \theta_S + \alpha P_l(\Delta_R^{(l)})$. Taylor expansion gives:

$$U_H(\theta_\alpha) - U_H(\theta_S) = \alpha \left\langle \nabla_{\theta^{(l)}} U_H(\theta_S), \Delta_R^{(l)} \right\rangle + \frac{1}{2}\alpha^2 (\Delta_R^{(l)})^T H_U^{(l)}(\xi) \Delta_R^{(l)}. \tag{26}$$

Where U_H is the hard-utility subspace projection. Using Assumption A.4 and the Hessian bound:

$$\mathbb{E}[U_H(\theta_\alpha) - U_H(\theta_S)] \ge \alpha c_R \mathbb{E}\|\Delta_R^{(l)}\|_2 - \frac{L_H}{2}\alpha^2 \mathbb{E}\|\Delta_R^{(l)}\|_2^2. \tag{27}$$

For sufficiently small positive α, the right side is positive. Thus, under the assumptions, a layer with stronger relative long signal has higher expected hard-reasoning potential in the proxy sense. □

A.3 Proof of Proposition 1.

Proof. The score $s_l = R_l(1 - C_l)$ is monotone increasing in R_l and monotone decreasing in C_l:

$$\frac{\partial s_l}{\partial R_l} = 1 - C_l \ge 0, \qquad \frac{\partial s_l}{\partial C_l} = -R_l \le 0. \tag{28}$$

Consider the following entropy-regularized objective:

$$\max_{M,\beta} \quad J_{\text{proxy}}(M, \beta) = \sum_{l=1}^{L} \beta_l s_l + \kappa_B T \sum_{l=1}^{L} \frac{\beta_l}{\kappa_B} \log \frac{\kappa_B}{\beta_l}, \tag{29}$$

subject to (1) $\sum_l M_l = K$, $M_l \in \{0, 1\}$ and (2) $\sum_l \beta_l = \kappa_B$, $0 \le \beta_l \le M_l$.

If a feasible solution selects layer j but not layer i with $s_i > s_j$, moving the coefficient mass from j to i preserves feasibility and increases the linear term by $\beta_j(s_i - s_j)$. Thus the optimal selected set is the top-K set by s_l.

Table 8: Generalization to Qwen3-4B backbone.

Method	GSM8K	MATH500	Minerva Math	Olympiad Bench	AIME24	AIME25	GPQA	**Avg.**
Baselines								
Qwen3-4B	93.3	94.4	42.8	63.0	66.7	50.0	54.5	66.4
	(374)	(1670)	(1446)	(4199)	(7046)	(7368)	(5086)	
Qwen3-4B-Thinking	95.0	96.2	60.0	73.2	83.3	80.0	67.7	79.3
	(1521)	(6125)	(5501)	(14405)	(20469)	(23912)	(9032)	
Merging methods								
Average Merging	95.2	95.5	57.5	70.7	73.3	57.8	64.7	73.5
	(1090)	(3425)	(4107)	(7296)	(9870)	(10102)	(7683)	(46.2%)
Task Arithmetic	95.3	96.1	57.7	71.2	72.2	67.8	62.1	74.6
	(1184)	(3723)	(3819)	(7844)	(10938)	(11398)	(6194)	(44.3%)
TIES-Merging	94.8	95.3	58.0	70.4	75.6	60.0	63.6	74.0
	(1173)	(3414)	(3785)	(7171)	(9738)	(10894)	(6601)	(47.2%)
DARE-Ties	94.7	94.4	56.5	68.3	64.4	56.7	58.6	70.5
	(1645)	(3861)	(4562)	(7586)	(11680)	(12250)	(6407)	(40.7%)
AIM	95.0	95.0	55.6	70.7	64.4	60.0	70.7	73.1
	(1063)	(2926)	(3877)	(6569)	(10357)	(9937)	(8010)	(47.2%)
ACM	94.8	93.8	56.4	68.1	70.0	54.4	61.6	71.3
	(928)	(3211)	(3200)	(6894)	(10991)	(11083)	(6144)	(47.6%)
Sens-Merging	95.3	95.6	57.0	70.8	74.0	58.0	64.0	73.5
	(1158)	(3520)	(3910)	(7180)	(9980)	(10500)	(6580)	(47.1%)
RPAM	95.5	96.2	57.3	72.3	75.6	67.8	66.7	75.9
	(1088)	(3194)	(4172)	(7228)	(8397)	(10160)	(7615)	**(48.3%)**
BCL-Merge	**96.0**	**96.3**	**58.5**	**73.5**	**76.7**	**69.0**	**71.0**	**77.3**
	(1280)	(3180)	(4190)	(7340)	(8510)	(10300)	(7720)	(47.5%)

Given selected layer set and ignoring caps, define $\pi_l = \beta_l/\kappa_B$. The objective becomes:

$$\max_{\pi} \quad \sum_{l\in\mathcal{S}} \pi_l s_l + T \sum_{l\in\mathcal{S}} \pi_l \log \frac{1}{\pi_l}, \qquad \sum_{l\in\mathcal{S}} \pi_l = 1. \tag{30}$$

With KKT conditions $\beta_l \leq 1$, the Lagrangian first-order condition gives:

$$\beta_l^* = M_l^* \min(1, \lambda \exp(s_l/T)), \qquad \sum_l \beta_l^* = \kappa_B. \tag{31}$$

□

B More Implementation Details

The threshold τ is set to be 1e-6 and the reasoning budget varies $B \in \{0.2, 0.3, 0.5, 0.7\}$. The decoding temperature are 0 with maximum generation tokens 8192. The checkpoints are obtained from publicly available repositories. Loading all models simultaneously takes about 42GB of GPU memory.

C More Experimental Results

Generalization to Qwen3-4B. To validate generalization beyond the Qwen2.5 series, we conduct experiments on the Qwen3-4B (Yang et al., 2025) setting. Table 8 shows that our method achieves 77.3% accuracy with a 47.5% length reduction, substantially outperforming existing methods by 1.4%. This result shows that the parameter-space criterion transfers across model families, further showing the generalization ability.

More sensitivity analysis. the threshold is to suppress noise in task vectors. We alter the value of threshold and report the results in Tab 9. The performance is stable for over a wide range. Too small threshold may be unable to filter noise. When τ increases to 1e-4, the average accuracy decreases to 57.3%. This suggests that an excessively large threshold may treat informative conflicts as noise and distort the layer selection. We therefore use τ=1e-6 by default.

Table 9: More Sensitivity analysis of hyperparameters.

Setting	Value	GSM8K	MATH500	Avg. Len	**Avg.**
τ	1e-7	92.3	84.8	1428	57.6
	1e-6	**92.5**	**85.0**	**1420**	**57.8**
	1e-5	92.4	84.9	1415	57.7
	1e-4	92.1	84.4	1395	57.3

D LIMITATIONS

Despite great promotion, our method assumes that task-vector norms are sufficiently comparable across layers to provide a useful ranking. Layer-wise normalization can reduce scale sensitivity, but extreme variation may still affect the ranking. Also, a direct implementation of the 7B merge can require about 42GB of GPU memory when all three models are loaded simultaneously. Sequential model loading and layer-wise streaming reduce the peak memory requirement without changing the merging process.