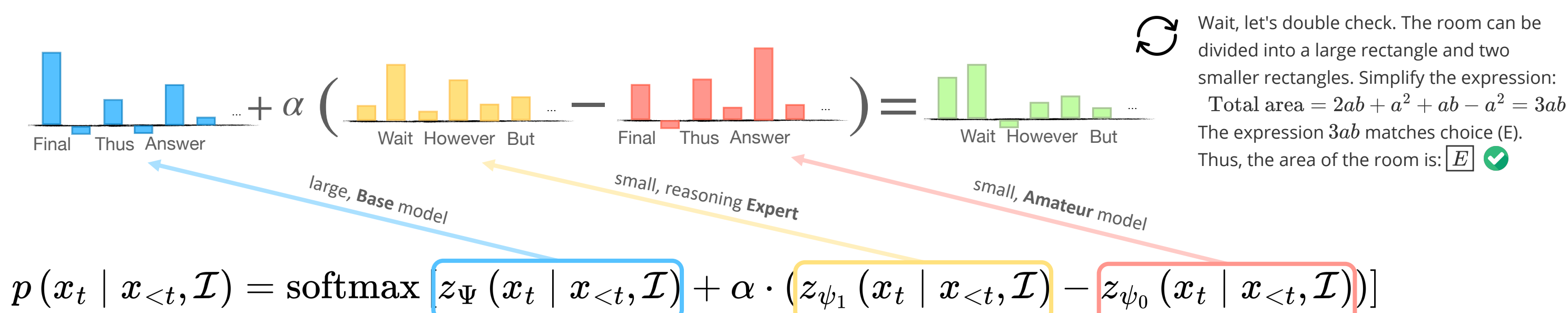
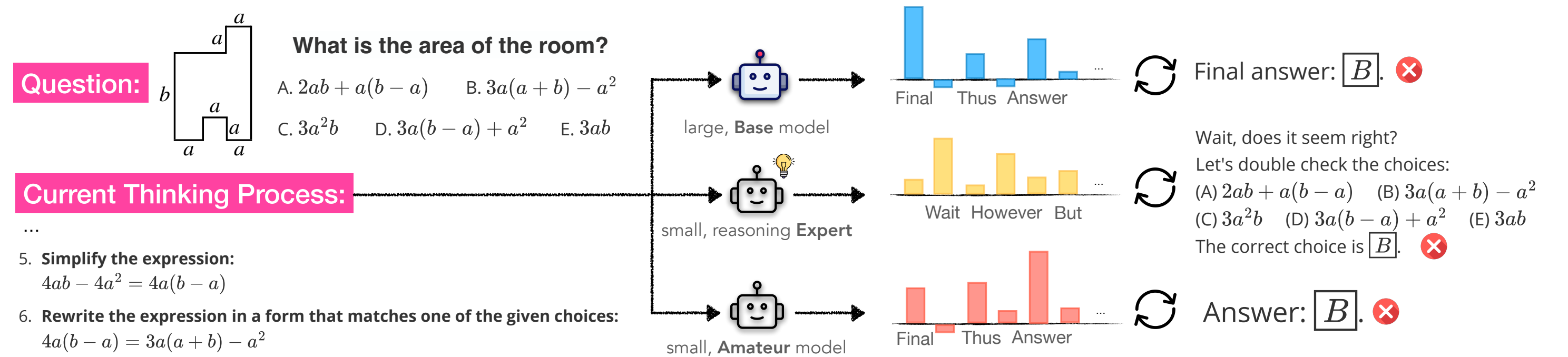


## ProxyThinker Overview



- Reinforcement fine-tuning (RFT) on **large models** is known to incur high training costs.

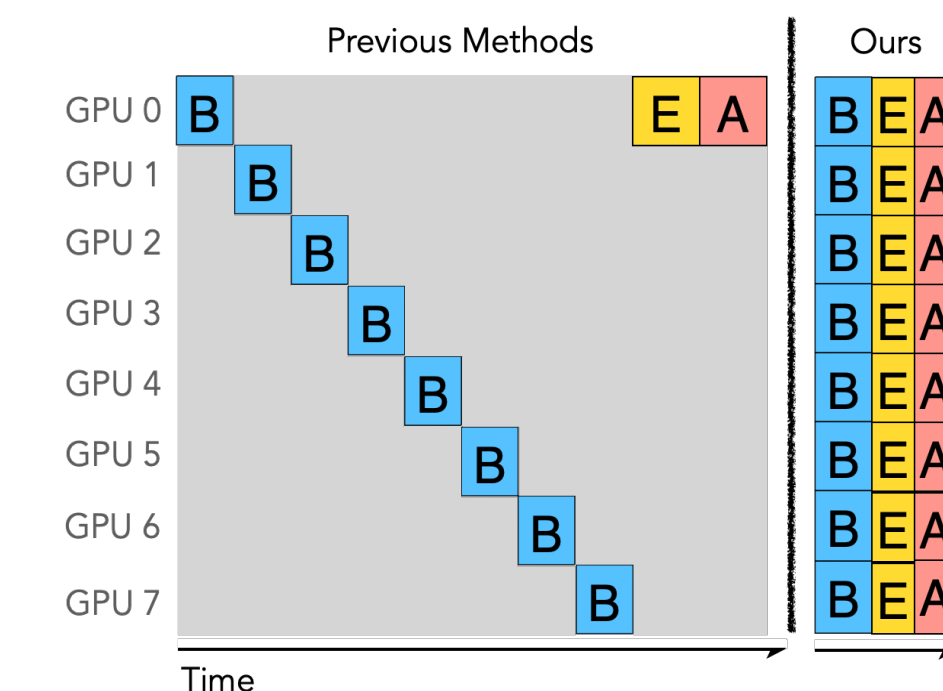
Related to ProxyTuning<sup>[1]</sup> and Contrastive Decoding<sup>[2]</sup>, ProxyThinker is designed to transfer visual reasoning from small RFT experts to large untuned VLMs.

## Efficient Implementation

- ProxyThinker needs to maintain and forward 3 models on the GPU.
- We parallelize model operations using vLLM and incur only 11% latency overhead.

Method	Duration (s)	Acc (%)
PROXYTHINKER		
Huggingface	19133	40.78
Ours Implementation	578	41.44
Optimized TP	501	41.44
Full-scale RFT expert		
VL-Rethinker-32B	451	41.77

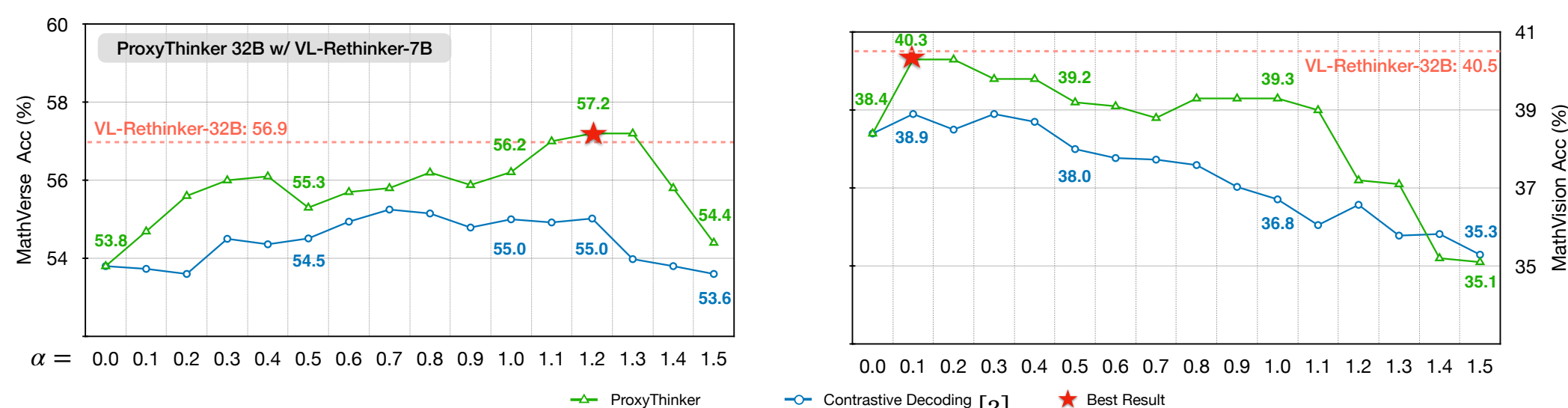
**+11.0% Latency**



## Strong Empirical Results on Visual Reasoning Benchmarks!

Model	Expert	Math Vista	Math Verse	Math Vision	MMMU Pro	R1-Bench	$\Delta$
Qwen2.5-VL-7B	–	68.2	46.3	25.1	36.9 <sup>†</sup>	32.1	–
💡 OpenVLThinker-7B	–	70.2	47.9	25.3	39.4*	32.9*	–
💡 ThinkLite-VL-7B	–	75.1	50.7	32.9	41.1*	39.0*	–
💡 VL-Rethinker-7B	–	74.9	54.2	32.3	41.7	47.2*	–
Qwen2.5-VL-32B	–	74.7	53.8*	38.4	49.5 <sup>†</sup>	49.4*	0.0
🤖 Qwen2.5-VL-32B	OpenVLThinker-7B	77.4 (+2.7)	53.8 (0.0)	<b>40.8</b> (+2.4)	51.8 (+2.3)	<b>53.0</b> (+3.6)	+2.2
🤖 Qwen2.5-VL-32B	ThinkLite-VL-7B	77.6 (+2.9)	<b>56.0</b> (+2.2)	38.8 (+0.4)	51.7 (+2.2)	49.7 (+0.3)	+1.6
🤖 Qwen2.5-VL-32B	VL-Rethinker-7B	<b>78.1</b> (+3.4)	55.3 (+1.5)	39.2 (+0.8)	<b>52.8</b> (+3.3)	52.5 (+3.1)	+2.4
💡 VL-Rethinker-32B	–	78.8	56.9	40.5	50.6	50.8*	+2.4
Qwen2.5-VL-72B	–	74.8	55.1*	38.1	51.6 <sup>†</sup>	50.4	0.0
🤖 Qwen2.5-VL-72B	OpenVLThinker-7B	77.8 (+3.0)	56.4 (+1.3)	36.2 (-1.9)	52.4 (+0.8)	50.4 (0.0)	+0.6
🤖 Qwen2.5-VL-72B	ThinkLite-VL-7B	<b>78.7</b> (+3.9)	57.2 (+2.1)	<b>40.4</b> (+2.3)	51.7 (+0.1)	50.2 (-0.2)	+1.6
🤖 Qwen2.5-VL-72B	VL-Rethinker-7B	78.1 (+3.3)	<b>58.6</b> (+3.5)	39.5 (+1.4)	<b>53.1</b> (+1.5)	<b>54.4</b> (+4.0)	+2.7
💡 VL-Rethinker-72B	–	80.3	61.7	43.9	55.9	57.9*	+5.9

## Strong Empirical Results on Visual Reasoning Benchmarks!



- Consistent improvement on visual reasoning benchmarks such as MathVerse and MathVision.
- ProxyThinker-32B even matches the performance of a full RFT model!

## Key References

[1] Liu et al. Tuning Language Models by Proxy. COLM 2024.  
 [2] Li et al. Contrastive Decoding: Open-ended Text Generation as Optimization. ACL 2023.

## Source Code / Implementation

<https://github.com/MrZilinXiao/ProxyThinker>