

Zhizheng's experience

Table of contents

[Development & Data Analysis of a Resistive Plate Counter \(RPC\)](#)

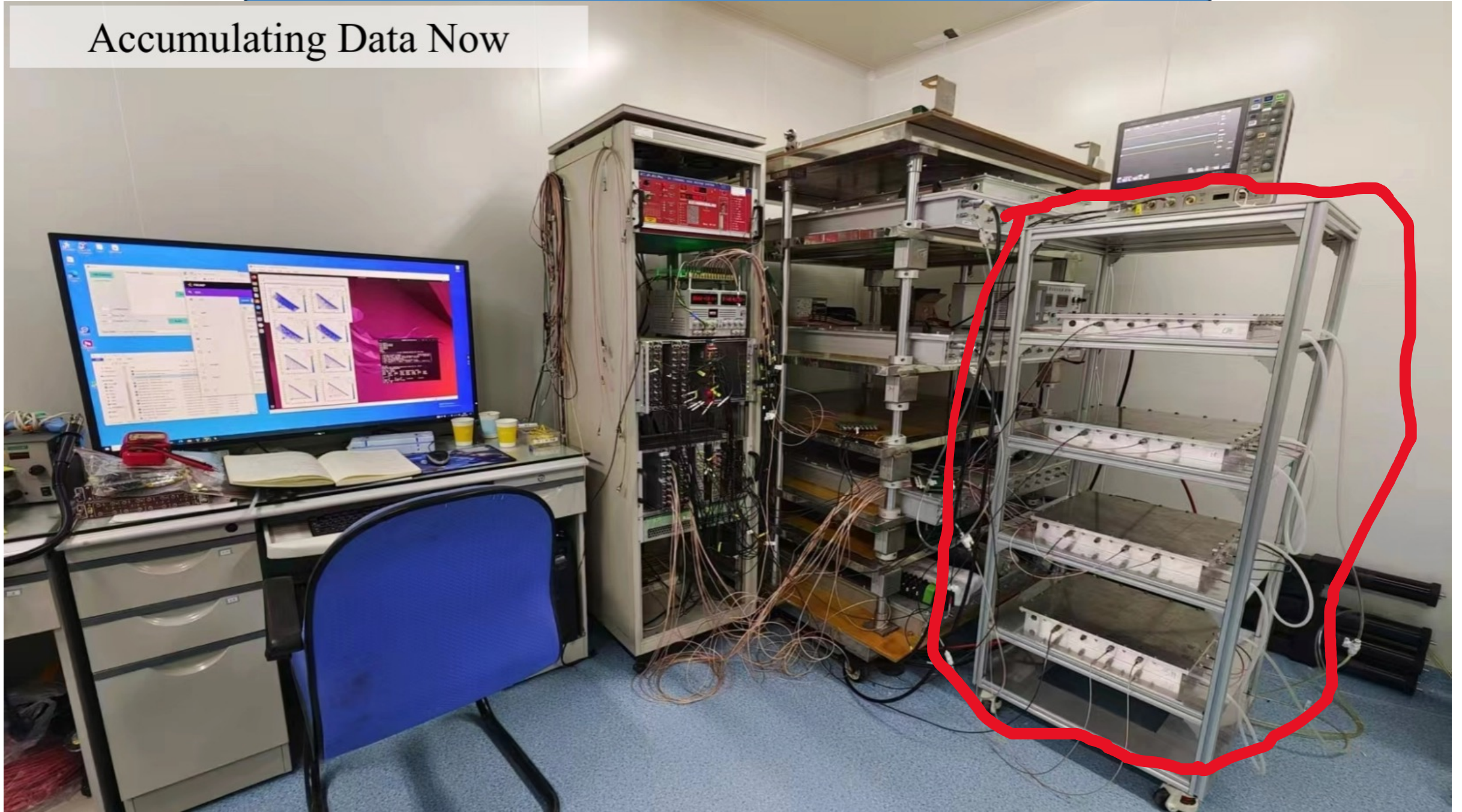
[Chain-of-Thought Reasoning for Advanced Image Generation](#)

[GRPO Optimization under Sampling–Reward Distribution Mismatch](#)

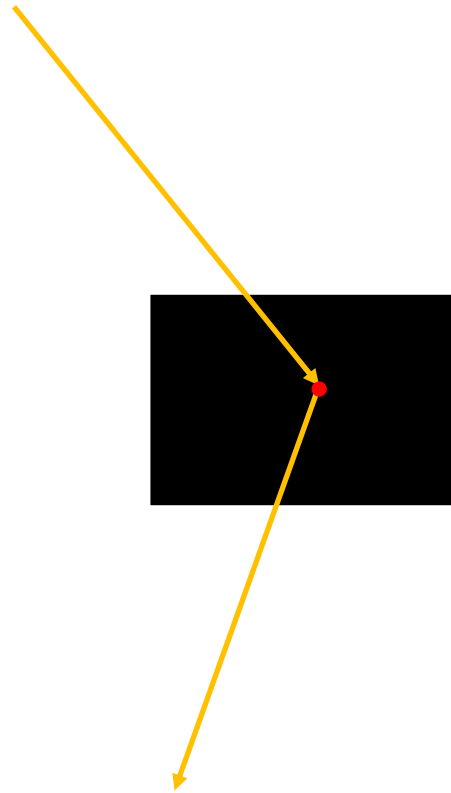
[Improving Crystal Structure Prediction via Niggli Reduction](#)

Development & Data Analysis of a Resistive Plate Counter (RPC)

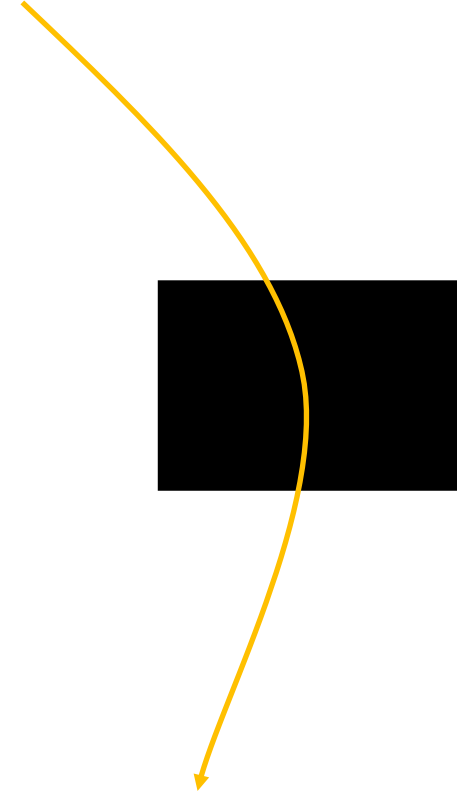
Accumulating Data Now



Development & Data Analysis of a Resistive Plate Counter (RPC)

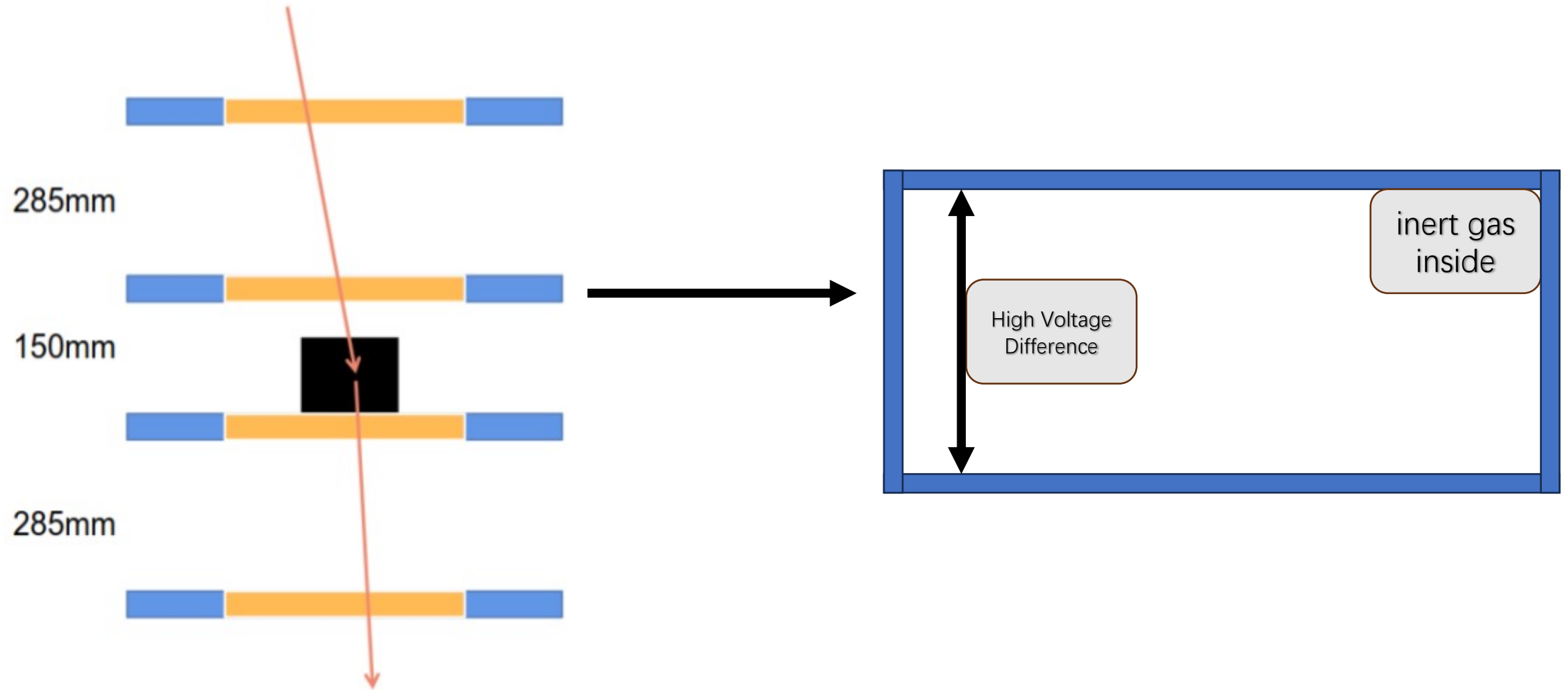


Algorithm



Real world

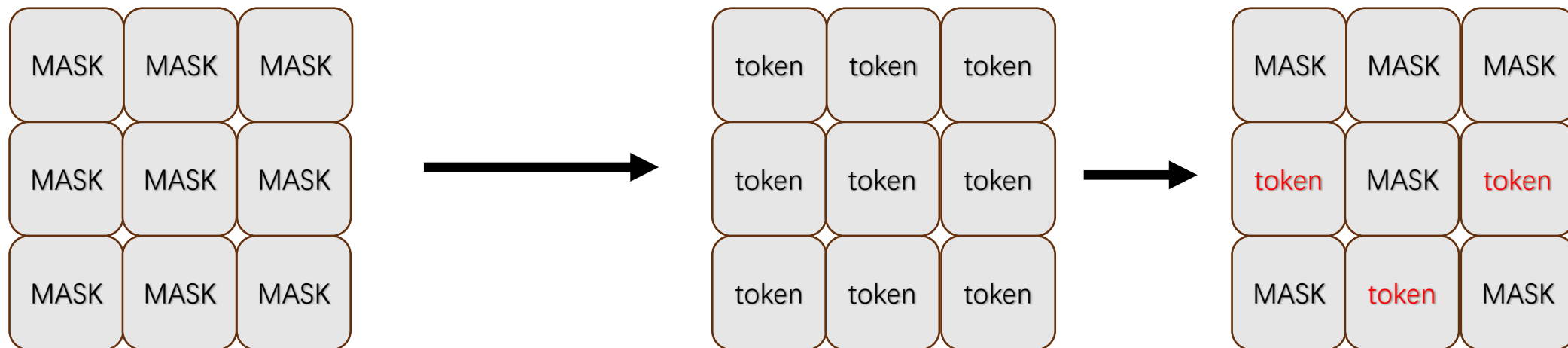
Development & Data Analysis of a Resistive Plate Counter (RPC)



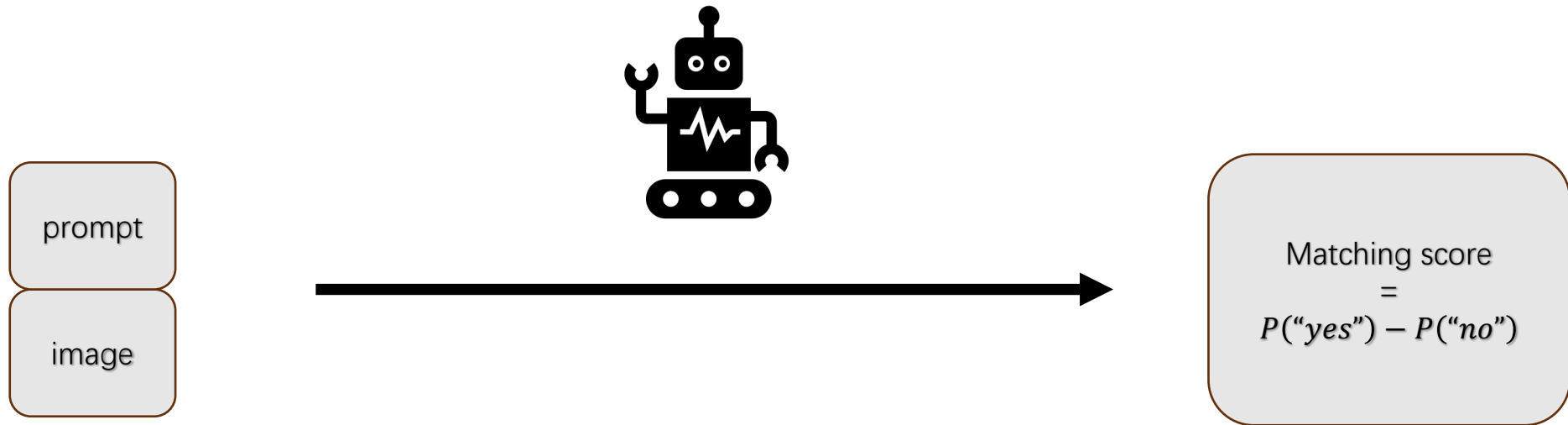
Chain-of-Thought Reasoning for Advanced Image Generation

Chain-of-Thought Reasoning for Advanced Image Generation

$$P(token_{ij} | prompt, [MASK], position\ embedding(i, j))$$

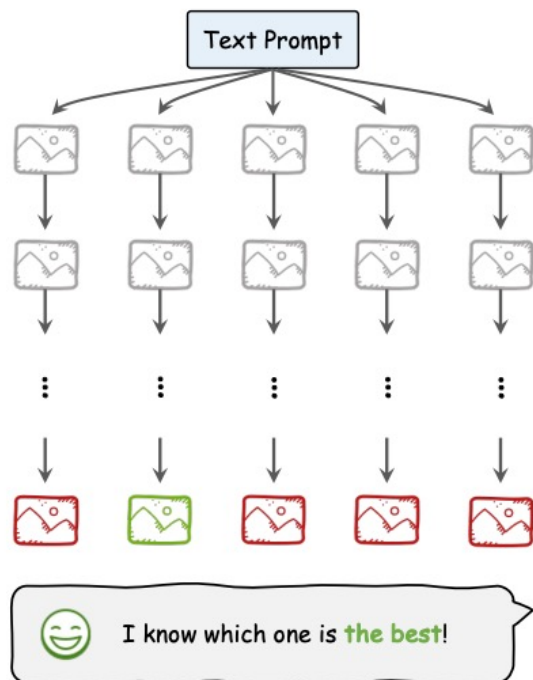


Chain-of-Thought Reasoning for Advanced Image Generation



Chain-of-Thought Reasoning for Advanced Image Generation

Outcome Reward Model (ORM)



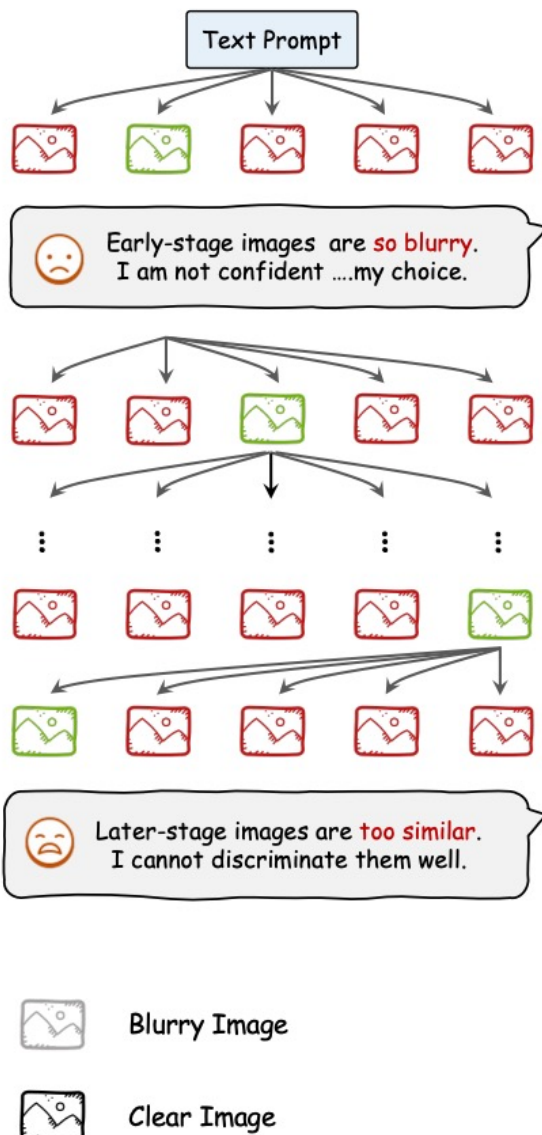
 Selected Bad Image

 Selected Good Image

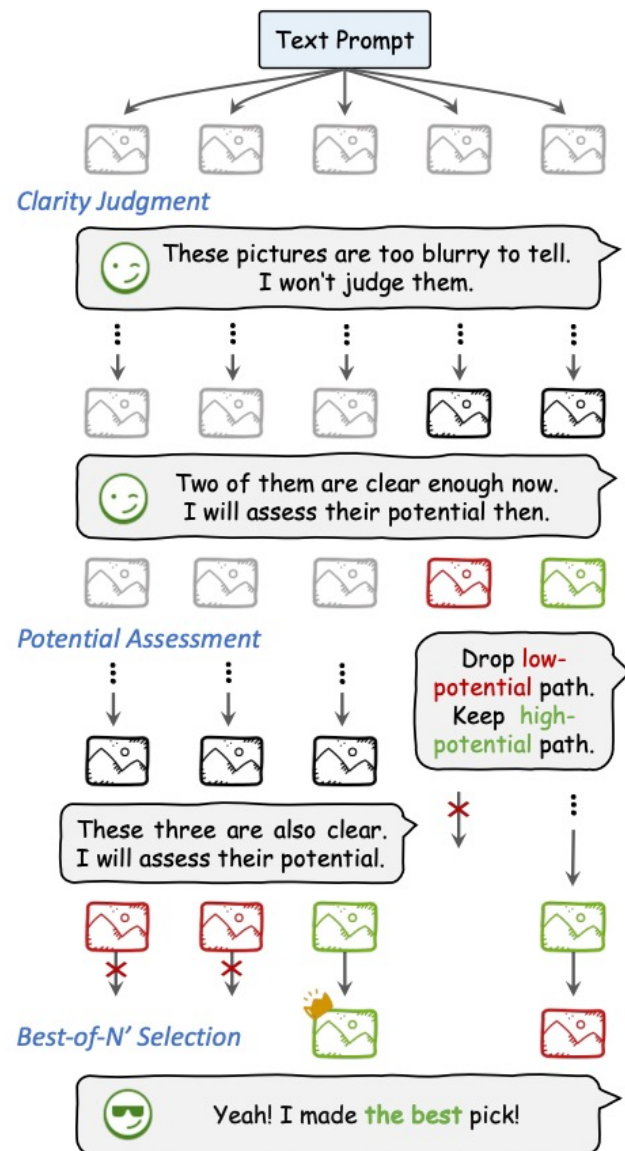
 Final Selected Image

 Verifier's Monologue

Process Reward Model (PRM)



Potential Assessment Reward Model (PARM)



Clarity Judgment

Potential Assessment

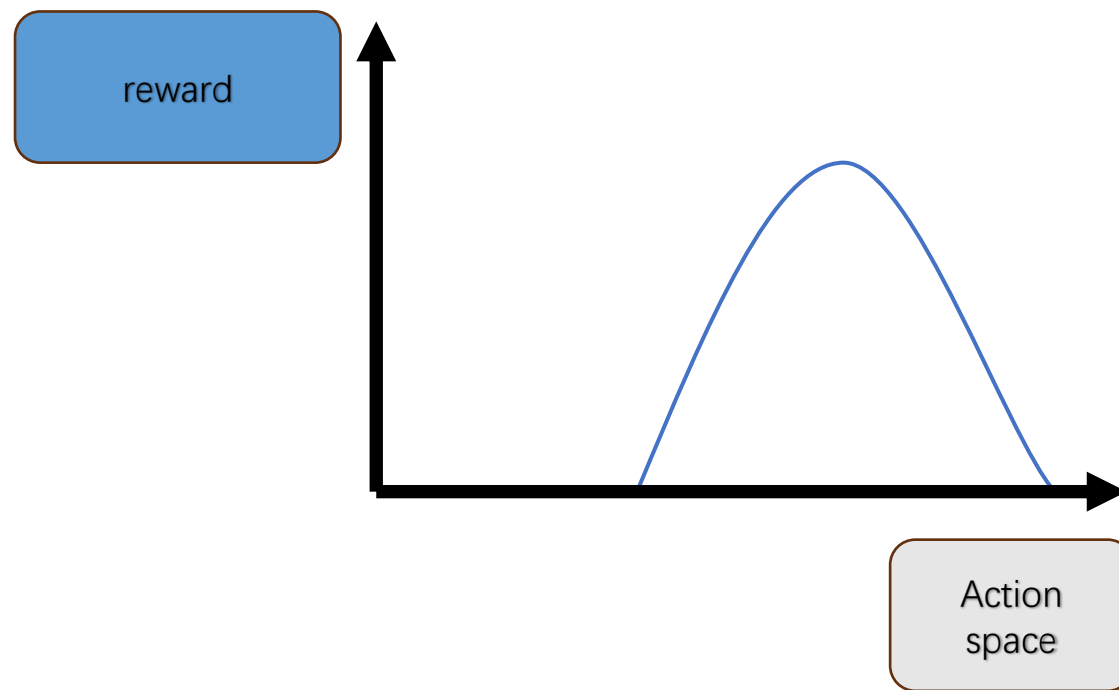
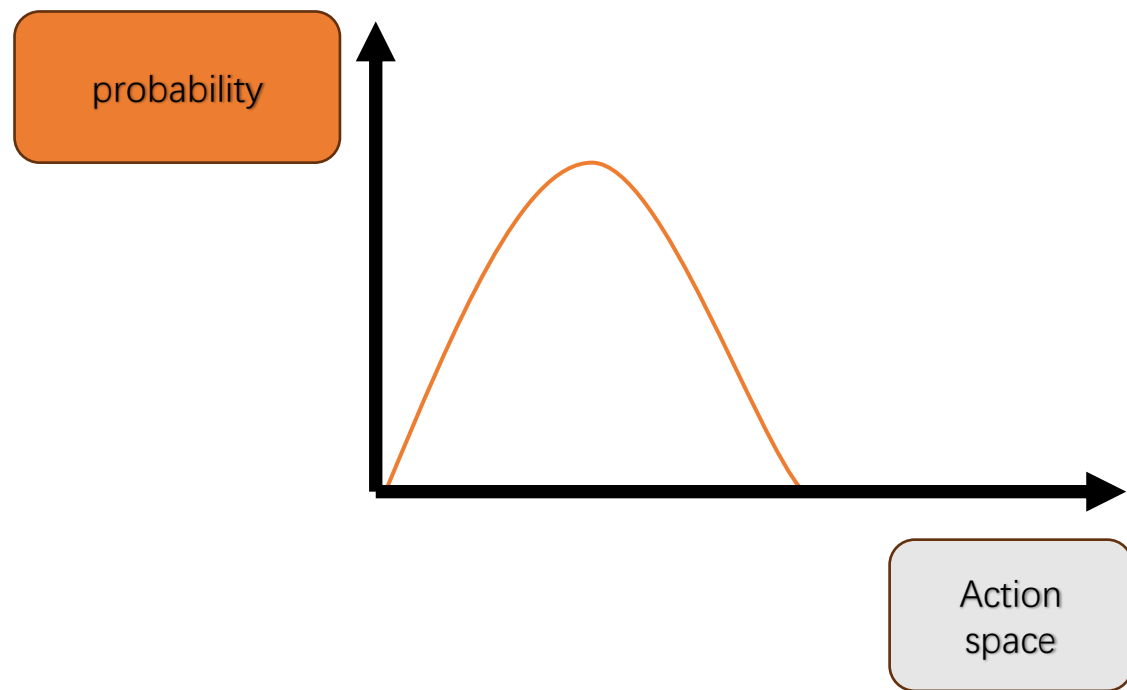
Best-of-N' Selection

Chain-of-Thought Reasoning for Advanced Image Generation

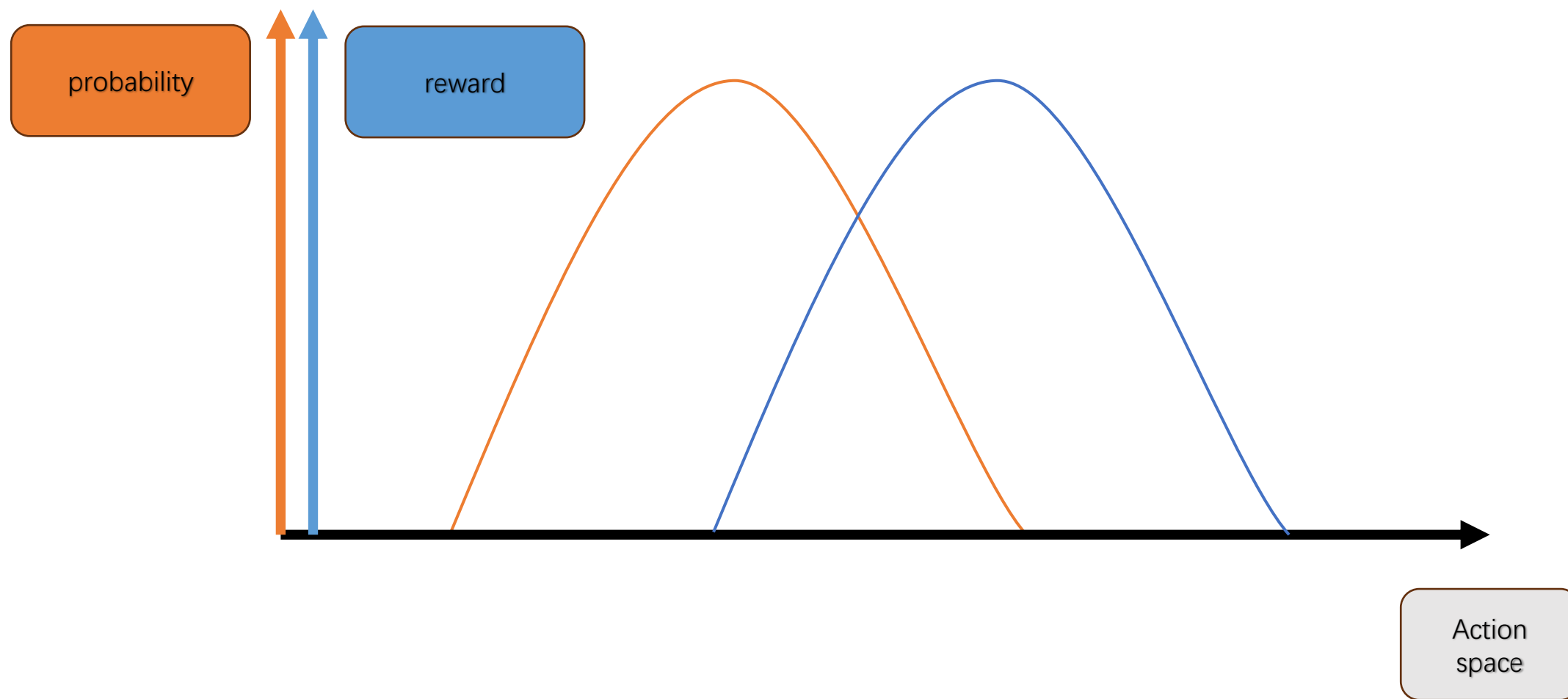
Model	Test-time Verifier	Preference Alignment	Reward Guidance	Single object	Two object	Counting	Colors	Position	Attribute binding	Overall
PixArt- α [4]	-	-	-	0.98	0.50	0.44	0.80	0.08	0.07	0.48
SD v2.1 [41]	-	-	-	0.98	0.51	0.44	0.85	0.07	0.17	0.50
DALL-E 2 [39]	-	-	-	0.94	0.66	0.49	0.77	0.10	0.19	0.52
SDXL [37]	-	-	-	0.98	0.74	0.39	0.85	0.15	0.23	0.55
SD 3 (d=24) [7]	-	-	-	0.98	0.74	0.63	0.67	0.34	0.36	0.62
LlamaGen [45]	-	-	-	0.71	0.34	0.21	0.58	0.07	0.04	0.32
Chameleon [46]	-	-	-	-	-	-	-	-	-	0.39
LWM [25]	-	-	-	0.93	0.41	0.46	0.79	0.09	0.15	0.47
SEED-X [11]	-	-	-	0.97	0.58	0.26	0.80	0.19	0.14	0.49
	-	-	-	0.95	0.52	0.49	0.82	0.11	0.28	0.53
Show-o [53]	Ft. ORM	-	-	0.99	0.72	0.65	0.84	0.25	0.33	0.63
	-	It. DPO	-	0.98	0.72	0.53	0.84	0.40	0.46	0.65
	Ft. ORM	It. DPO	Ft. ORM	0.98	0.84	0.64	0.85	0.66	0.52	0.75
	PARM	-	-	0.99	0.77	0.68	0.86	0.29	0.45	0.67
	-	It. DPO	PARM	0.97	0.75	0.60	0.83	0.54	0.53	0.69
	PARM	It. DPO	-	0.98	0.83	0.64	0.84	0.59	0.62	0.74
	PARM	It. DPO	PARM	0.99	0.86	0.67	0.84	0.66	0.64	0.77

GRPO Optimization under Sampling–Reward Distribution Mismatch

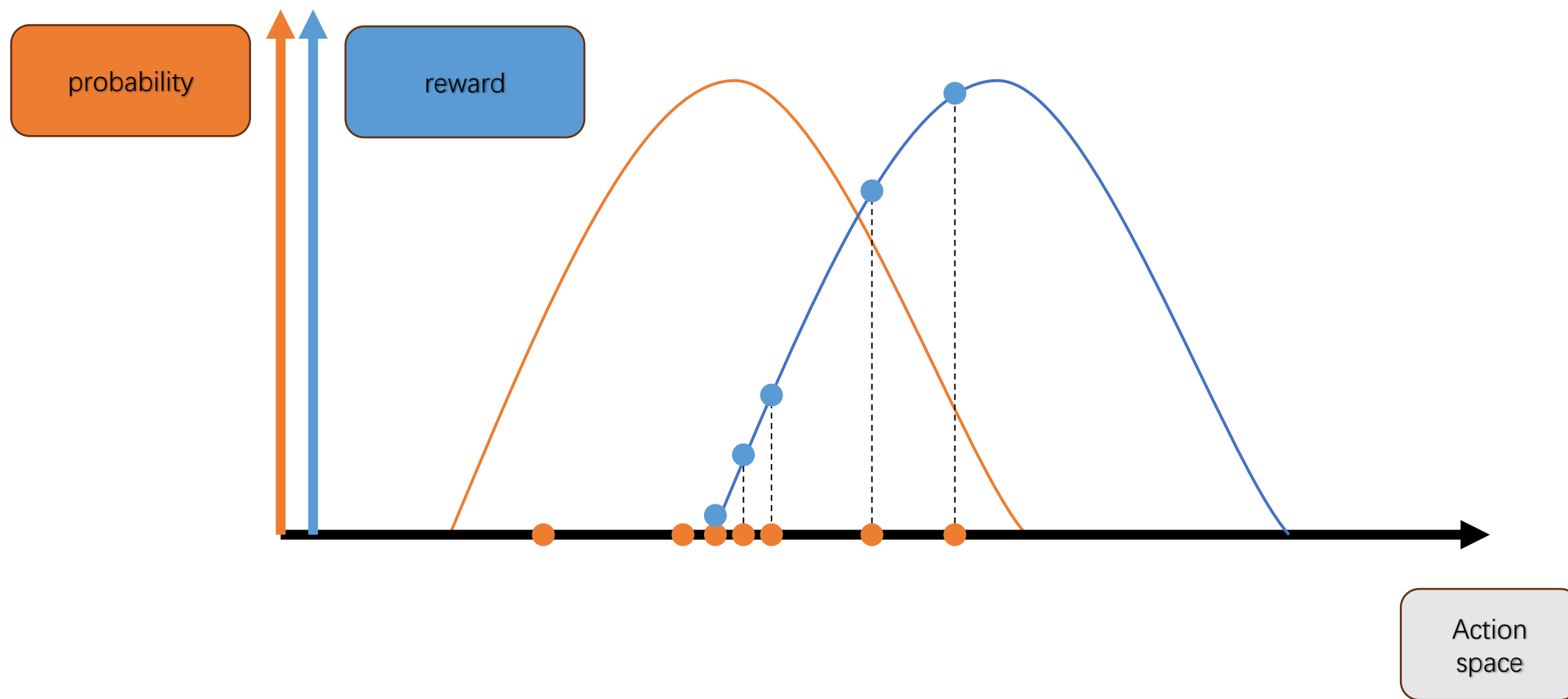
GRPO Optimization under Sampling–Reward Distribution Mismatch



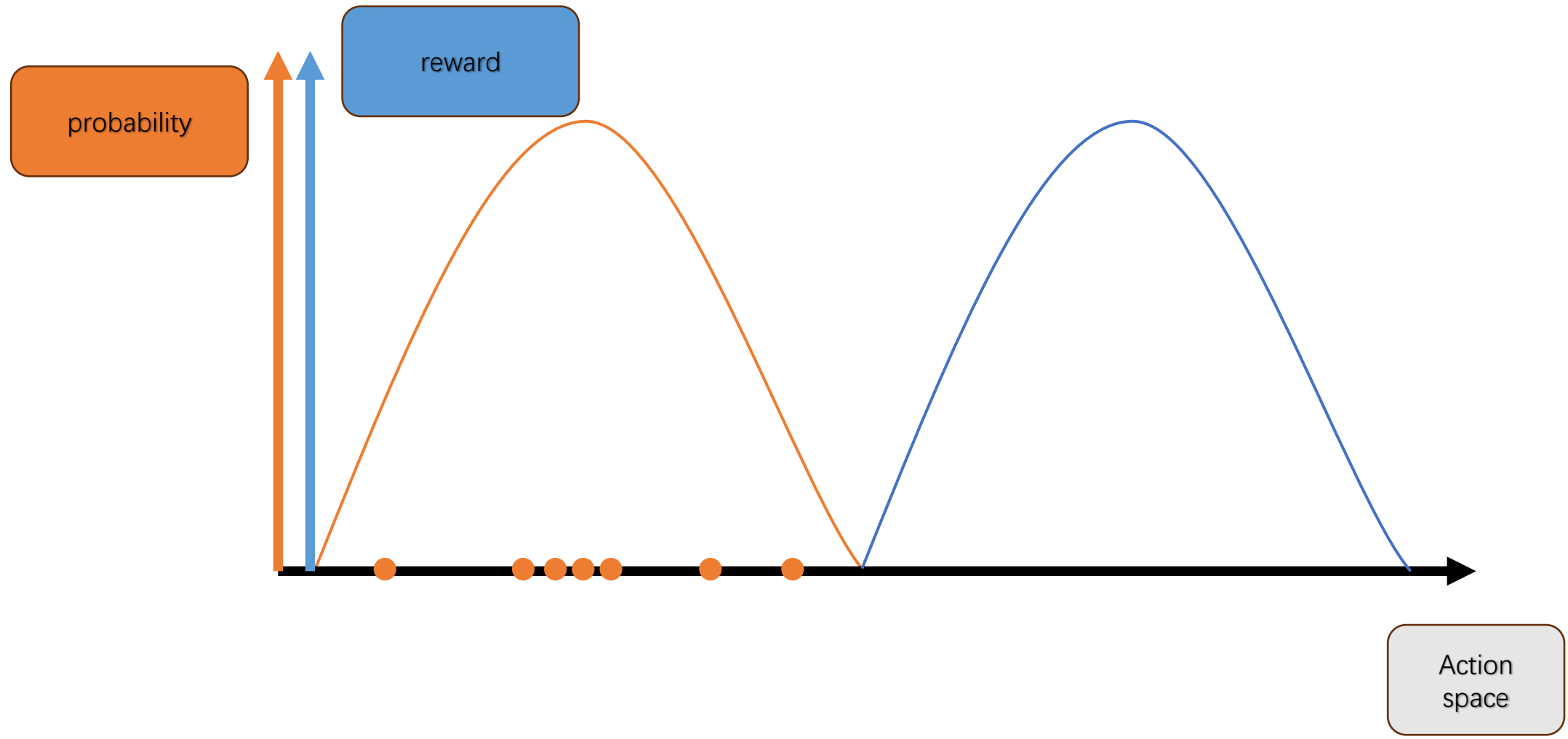
GRPO Optimization under Sampling–Reward Distribution Mismatch



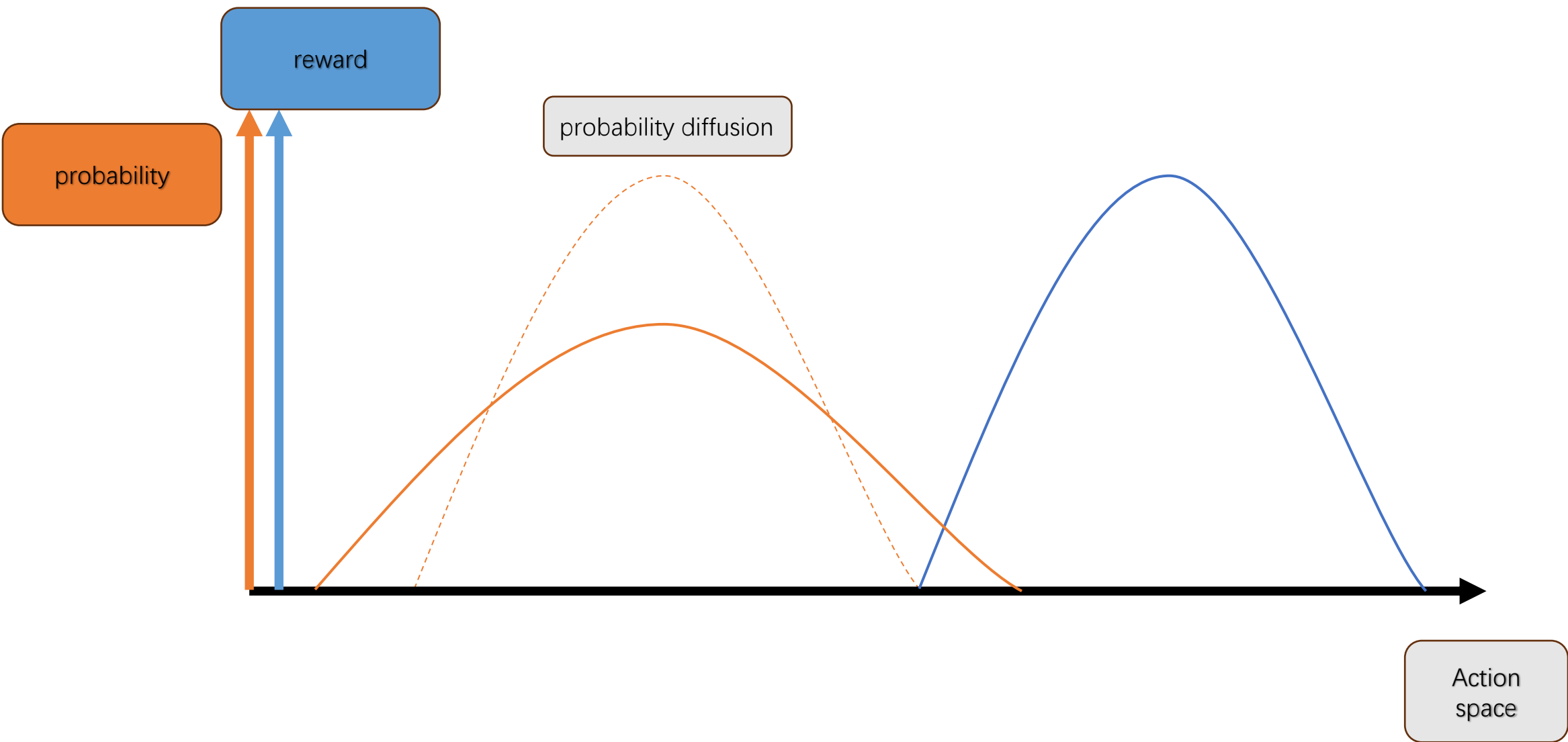
GRPO Optimization under Sampling–Reward Distribution Mismatch



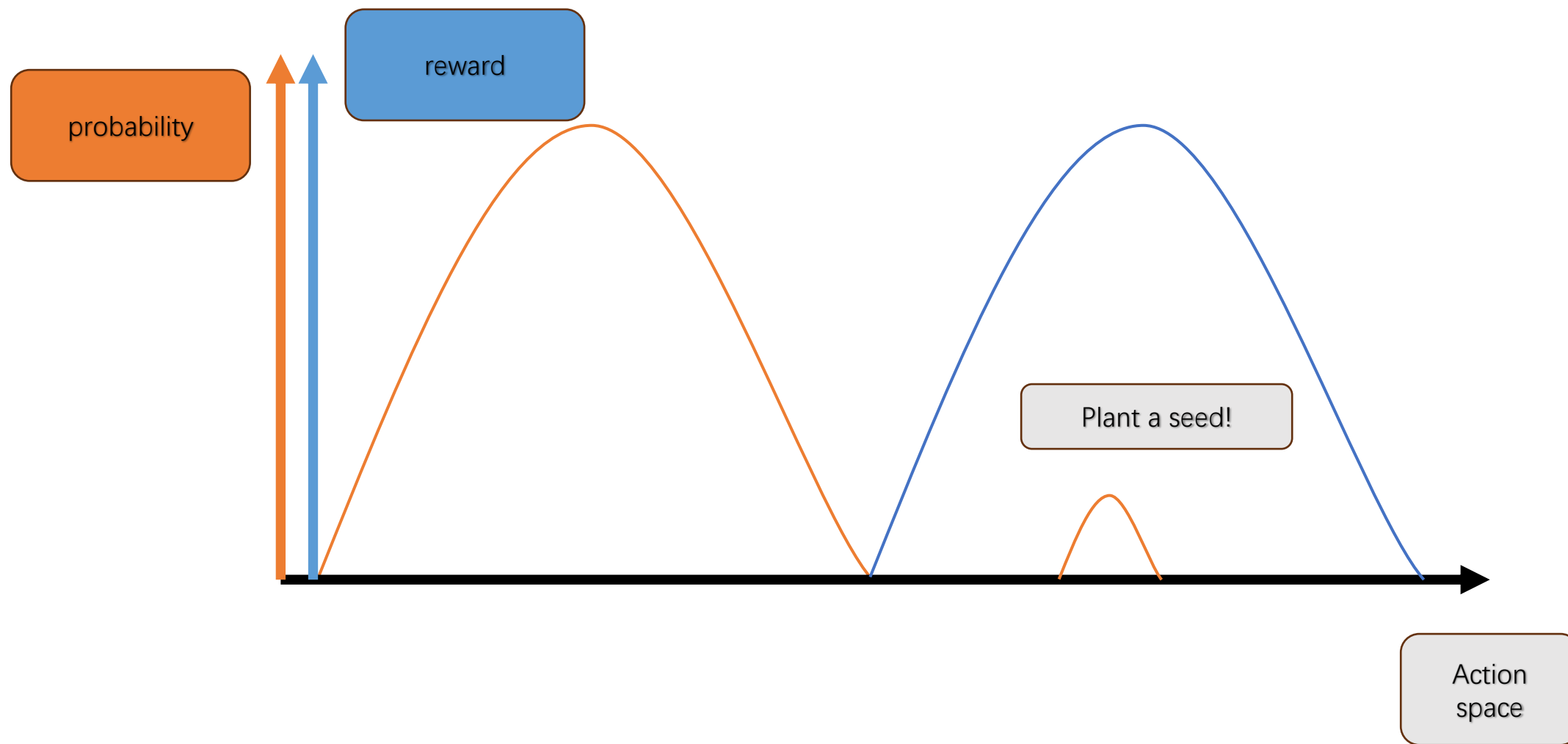
GRPO Optimization under Sampling–Reward Distribution Mismatch



GRPO Optimization under Sampling–Reward Distribution Mismatch

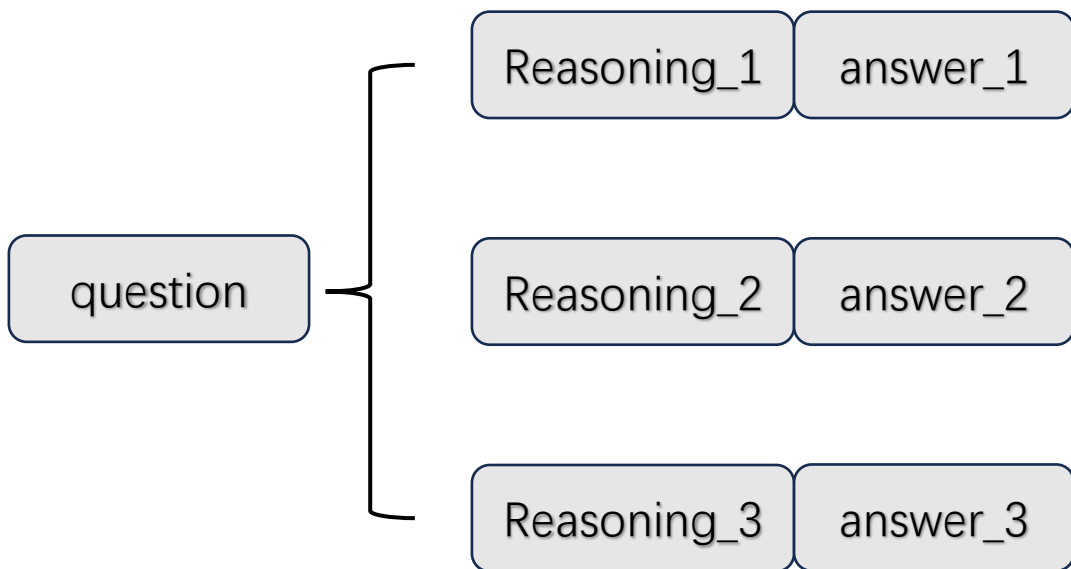


GRPO Optimization under Sampling–Reward Distribution Mismatch



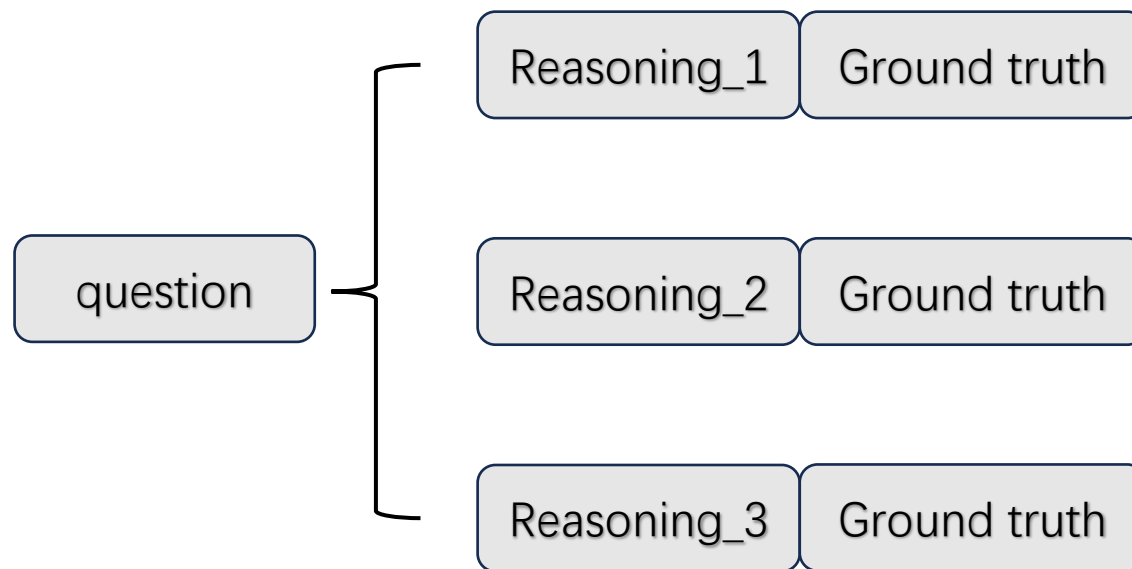
GRPO Optimization under Sampling–Reward Distribution Mismatch

$$reward = \delta(answer - correct\ answer)$$



$$Loss_1 = -A_{correctness} \times \sum_{t \in \{reasoning \cup answer\}} \log p_t$$

$$reward = P(ground\ truth \mid reasoning)$$



$$Loss_2 = -A_{GTP} \times \sum_{t \in reasoning} \log p_t - 1 \times \sum_{t \in ground\ truth} \log p_t$$

GRPO Optimization under Sampling–Reward Distribution Mismatch

All reward=0

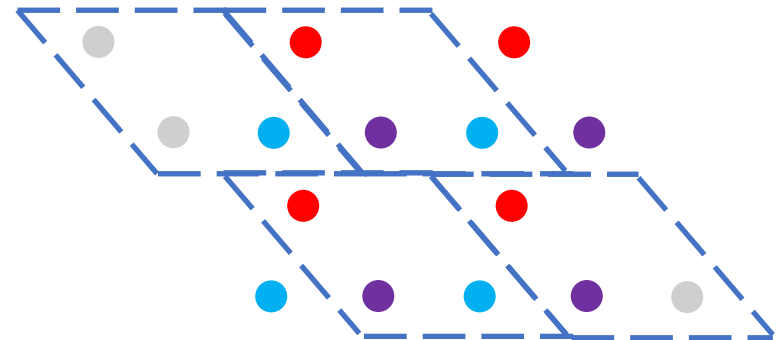
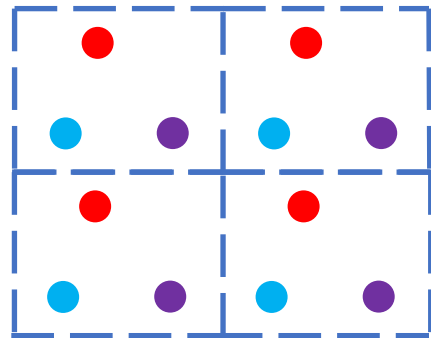
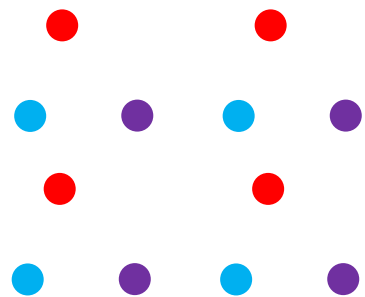
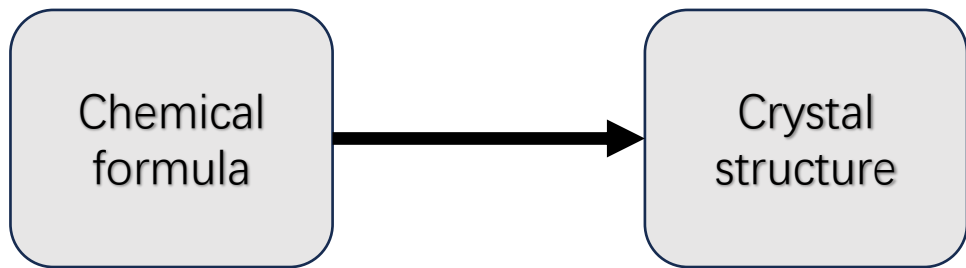


$$\text{Loss}_1 = -A_{\text{correctness}} \times \sum_{t \in \{\text{reasoning} \cup \text{answer}\}} \log p_t$$

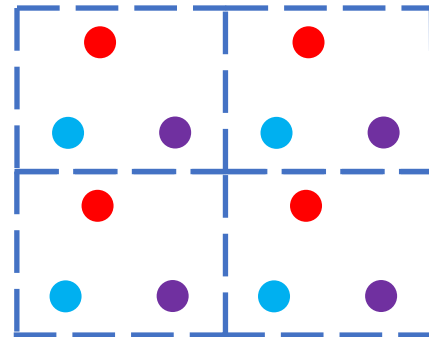
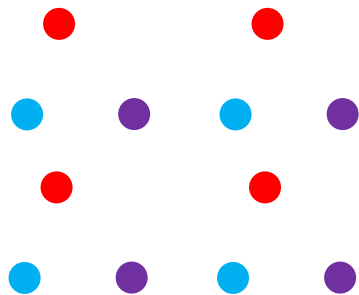
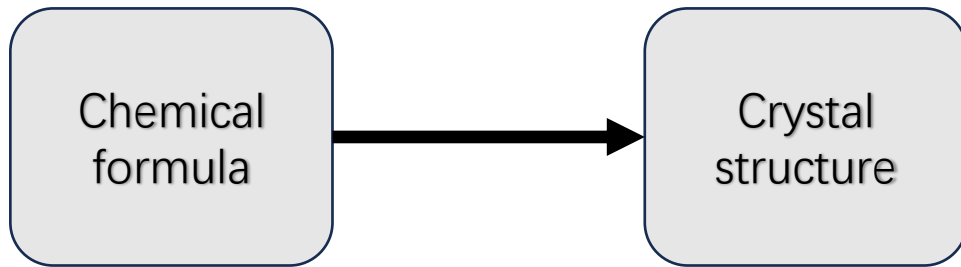
$$\text{Loss}_2 = -A_{\text{GTP}} \times \sum_{t \in \text{reasoning}} \log p_t - 1 \times \sum_{t \in \text{ground truth}} \log p_t$$

Evaluating Reasoning-Path Hidden Space for Crystals (Encode→Decode)

Evaluating Reasoning-Path Hidden Space for Crystals (Encode→Decode)

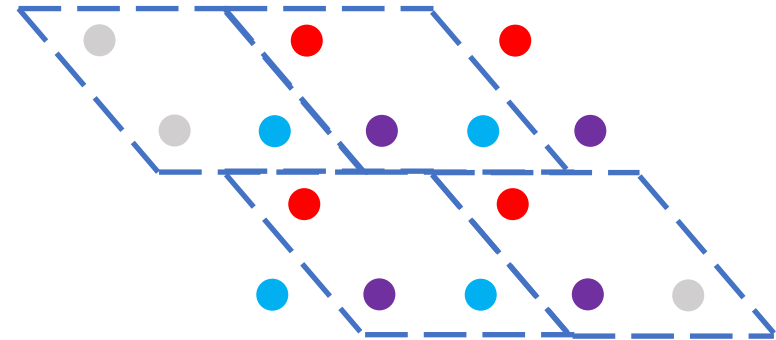


Evaluating Reasoning-Path Hidden Space for Crystals (Encode→Decode)



$$\vec{a} = (3, 0)$$
$$\vec{b} = (2, 0)$$

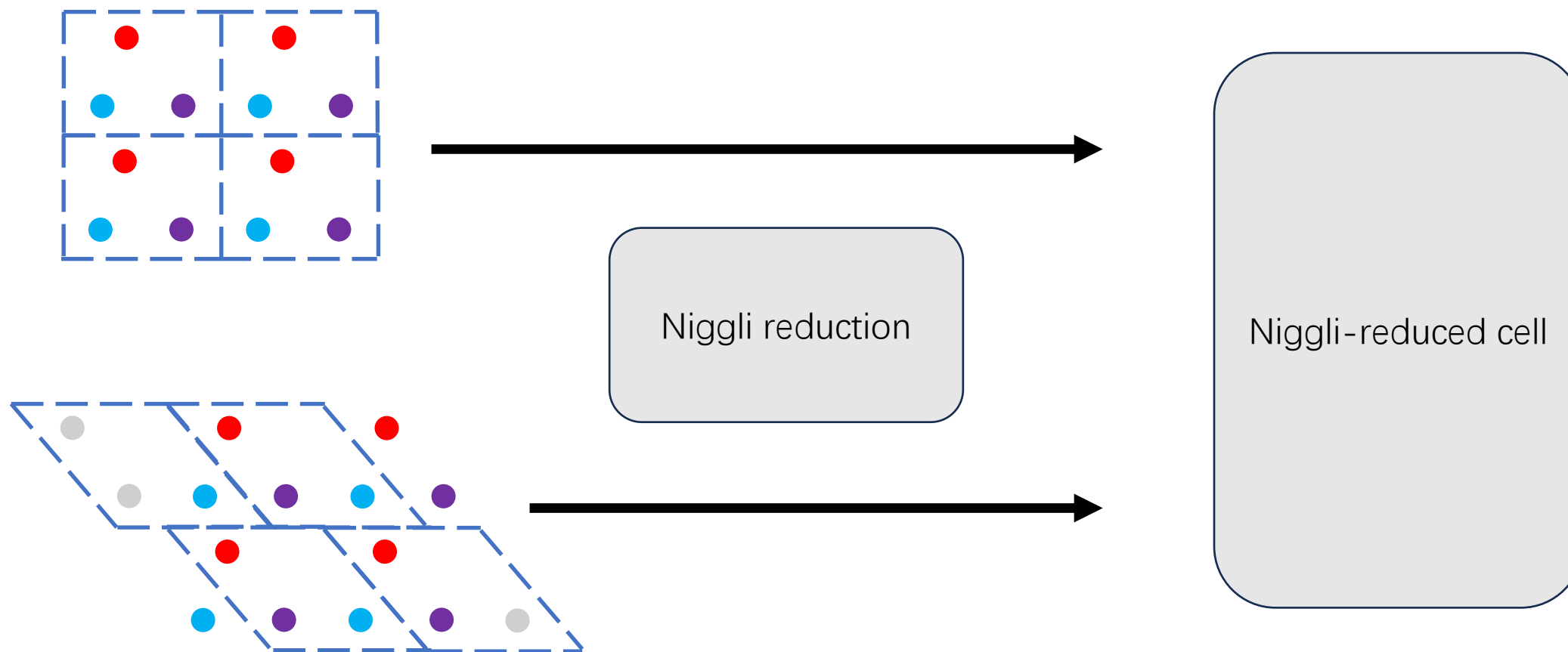
$$\text{red} = 0.3\vec{a} + 0.9\vec{b}$$
$$\text{blue} = 0.2\vec{a} + 0.15\vec{b}$$
$$\text{purple} = 0.8\vec{a} + 0.15\vec{b}$$



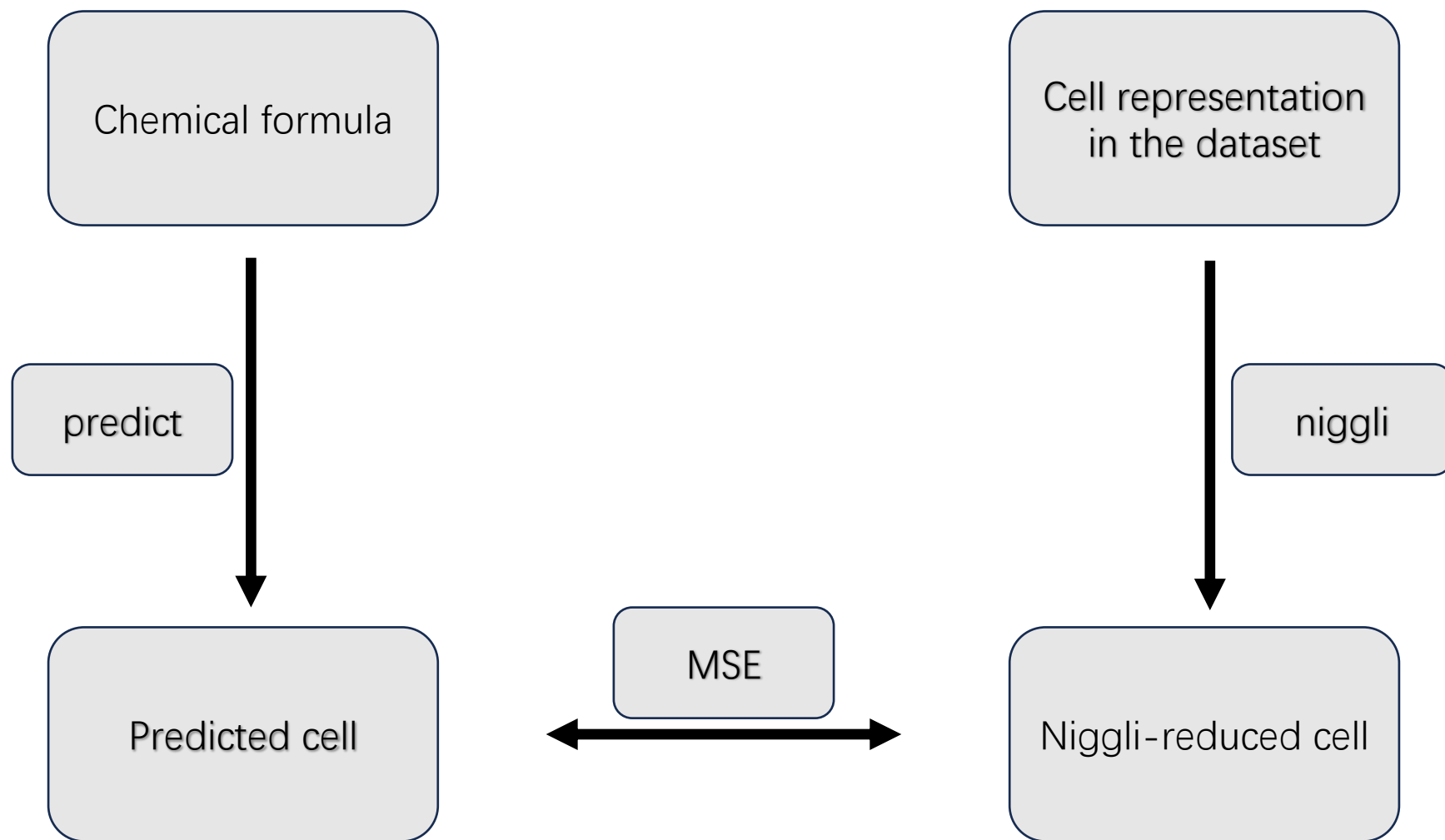
$$\vec{a} = (3, 0)$$
$$\vec{b} = (2, -1.5)$$

$$\text{red} = 0.75\vec{a} + 0.9\vec{b}$$
$$\text{blue} = 0.275\vec{a} + 0.15\vec{b}$$
$$\text{purple} = 0.875\vec{a} + 0.15\vec{b}$$

Evaluating Reasoning-Path Hidden Space for Crystals (Encode→Decode)



Evaluating Reasoning-Path Hidden Space for Crystals (Encode→Decode)



Thanks!