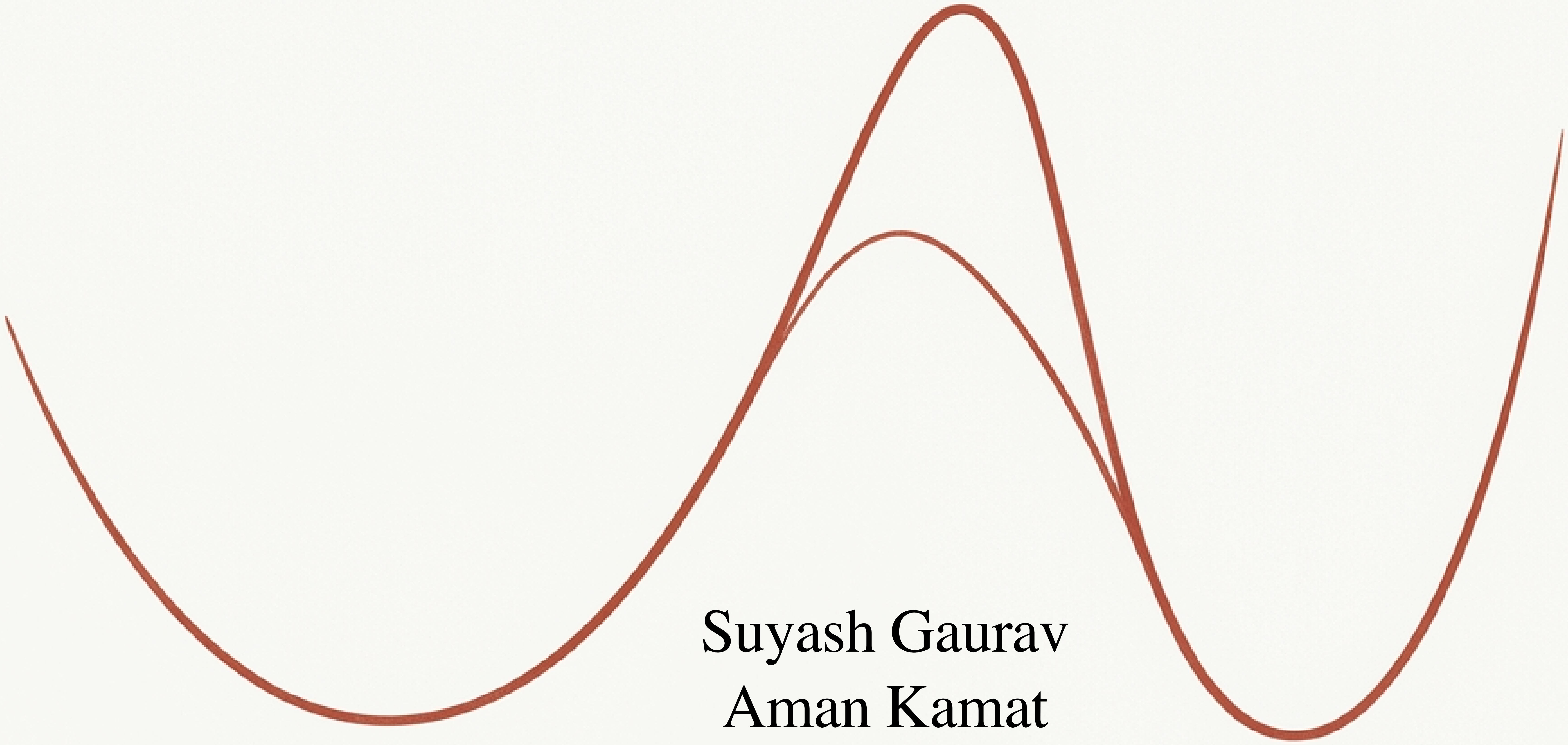
Interpolation-Aware PEFT

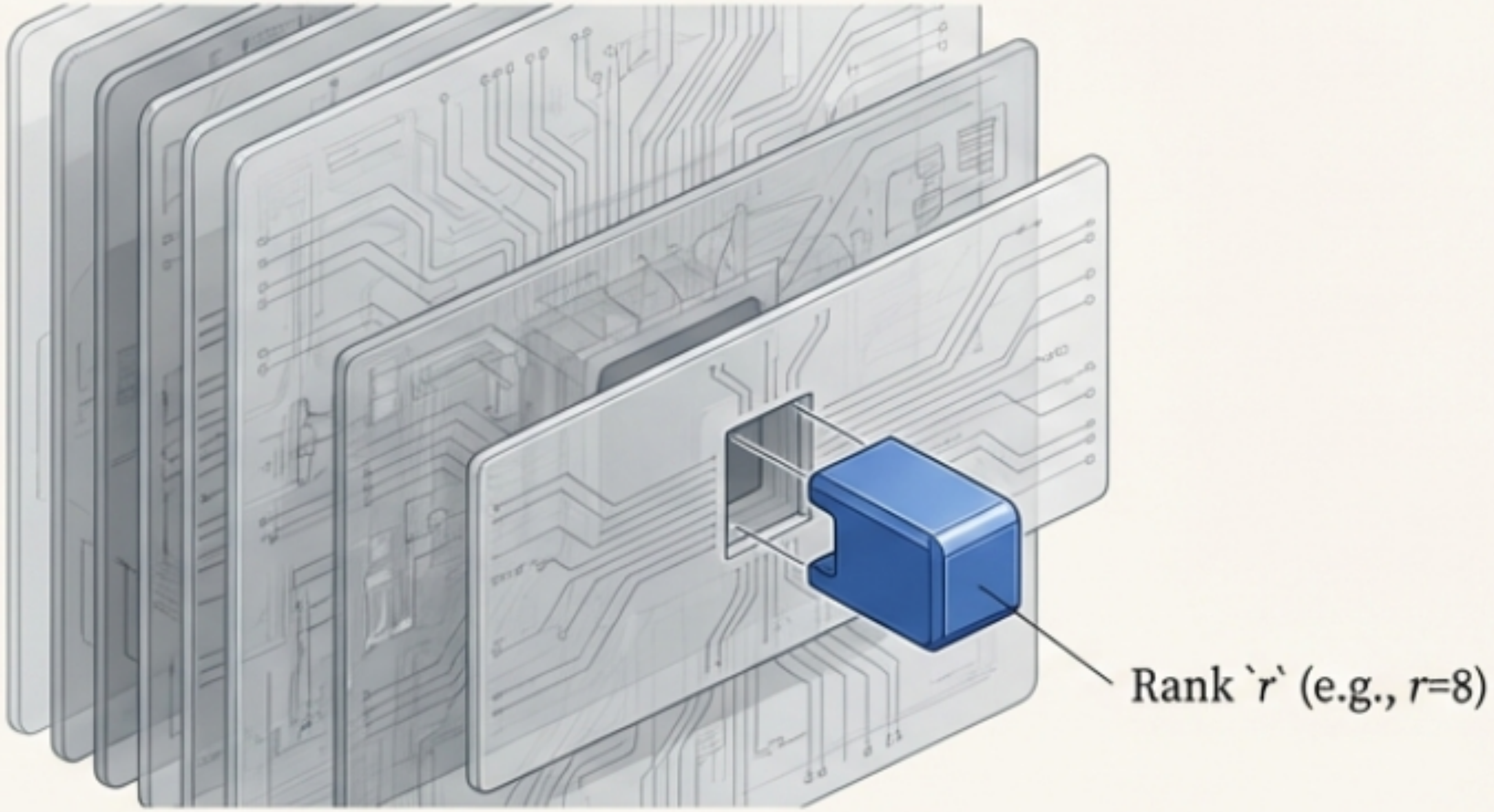
A New Paradigm for Fine-Tuning, Guided by the Theory of Double Descent



Suyash Gaurav
Aman Kamat

This work fundamentally repositions PEFT from a parameter-counting exercise to a principled approach grounded in high-dimensional statistics and implicit regularization theory.

The PEFT Dogma: Less is More

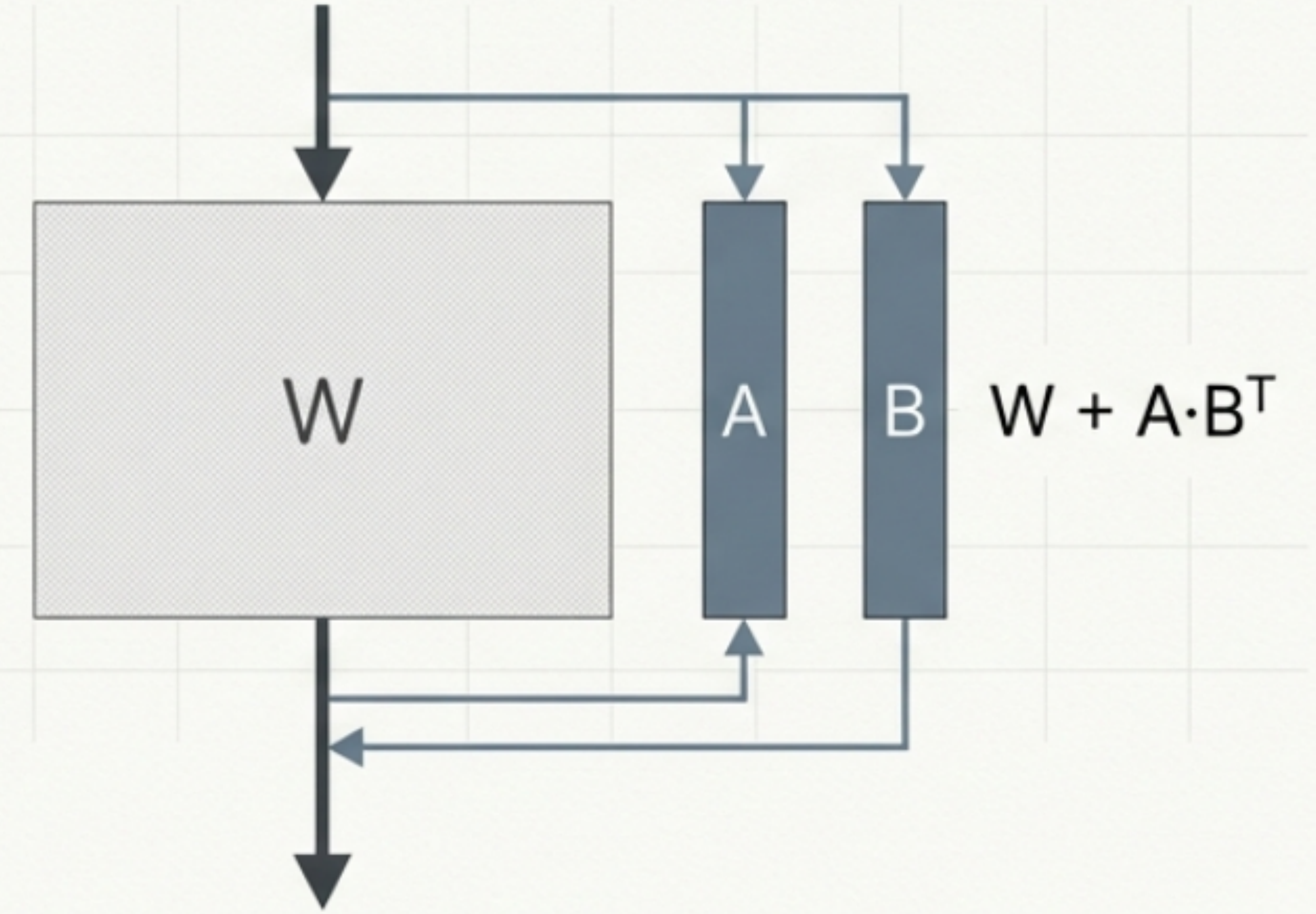


For years, the core principle of PEFT has been simple:
fewer parameters lead to better generalization and efficiency.

- **Dominant Method:** Low-Rank Adaptation (LoRA)
- **Core Assumption:** Smaller rank (`r`) = Better performance.
- **Goal:** Minimize trainable parameters.

The Dominant Method - LoRA

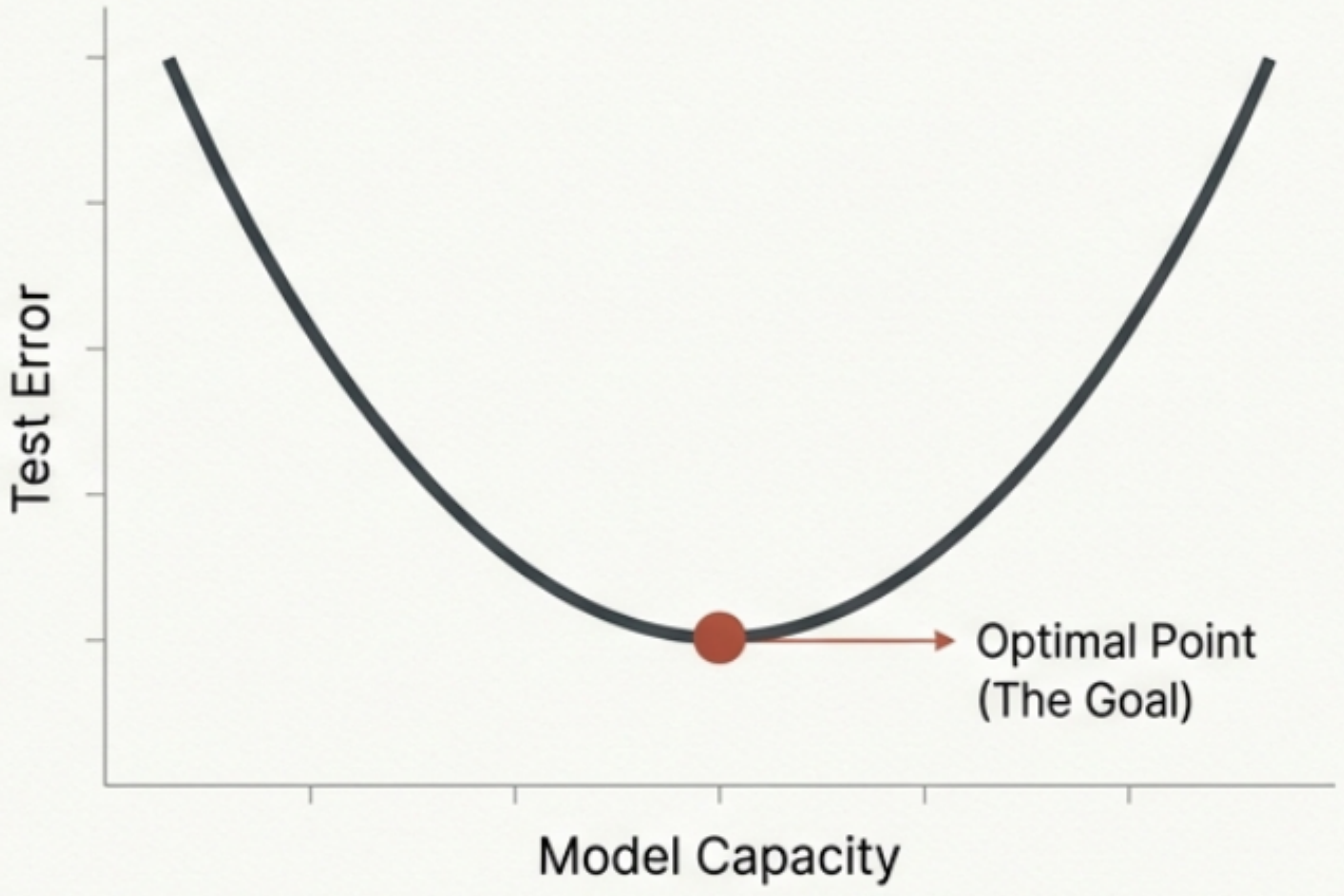
LoRA adapts models by training a small number of parameters in low-rank matrices (A and B), keeping the massive pretrained model frozen.



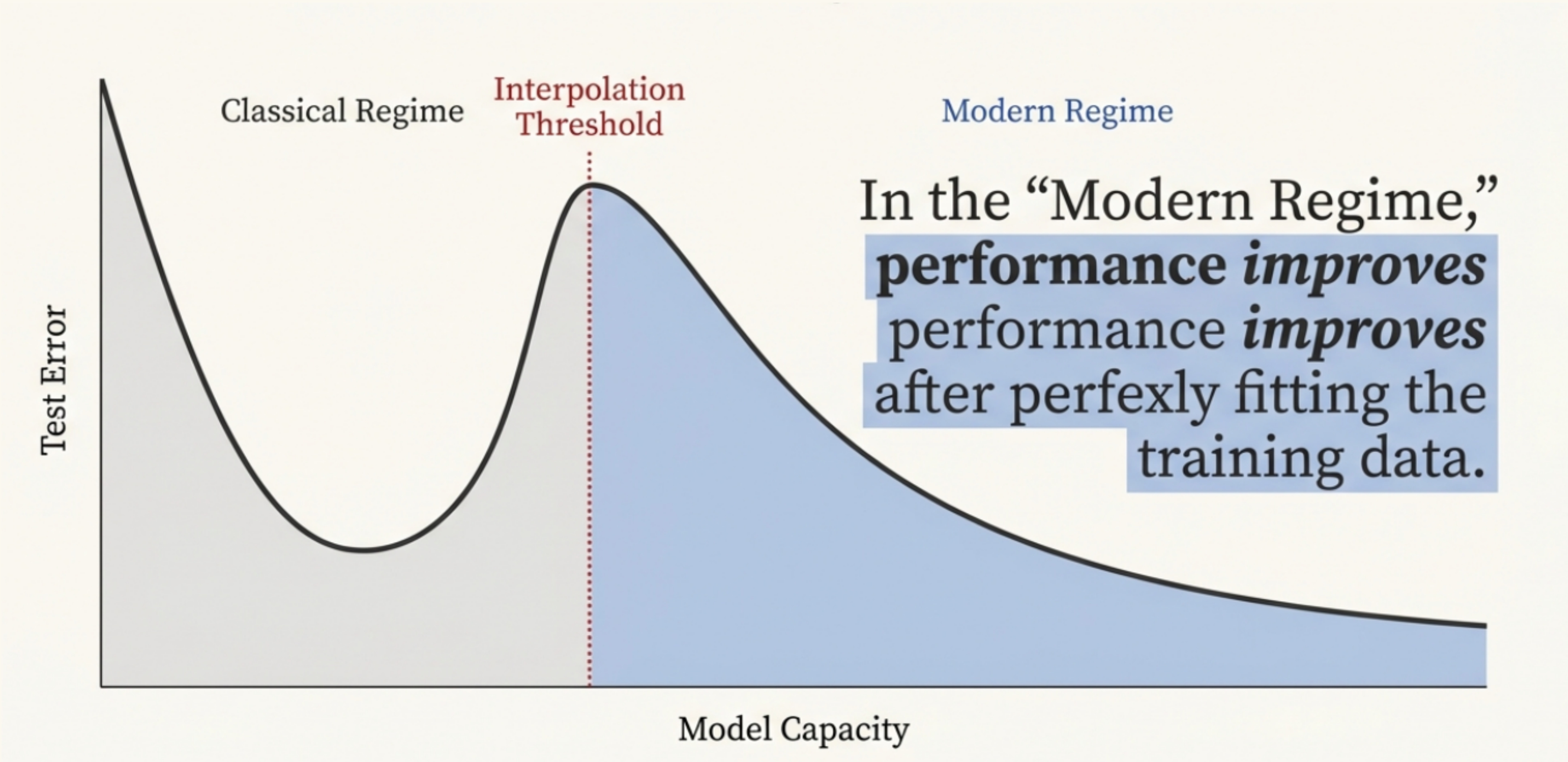
Core Idea behind ML methods

The Guiding Philosophy

The implicit assumption is that smaller rank r means fewer parameters, better efficiency, and protection against overfitting. This follows the classic U-shaped bias-variance tradeoff.

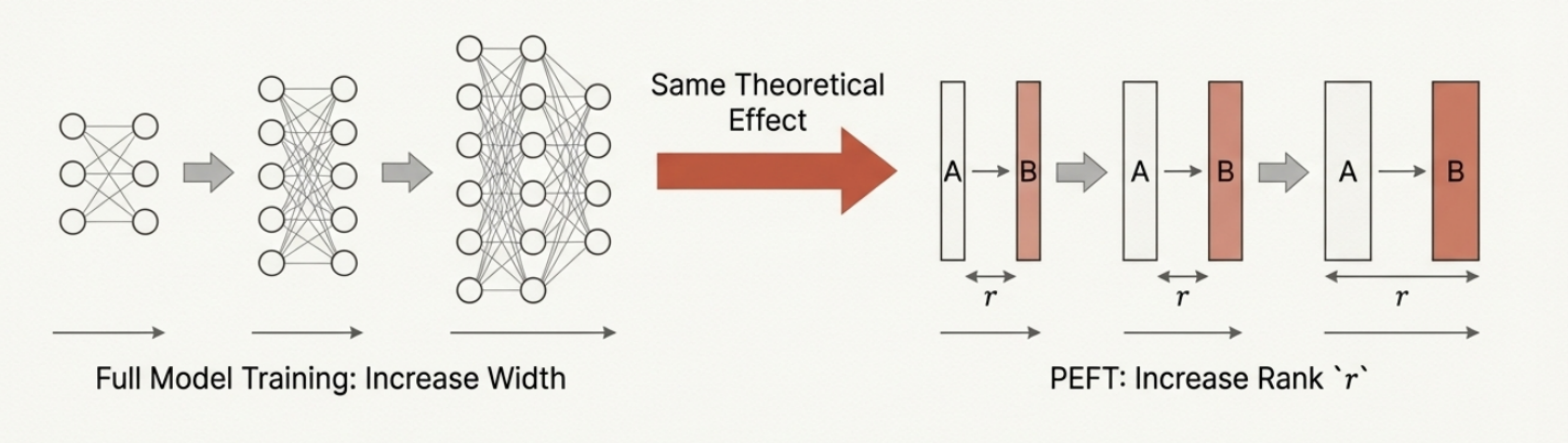


But Modern ML obeys a Differrent Law
“Double Decent Phenomena”



A New Model: LoRA Rank is the Effective Width of the Model

Our analysis reveals a novel perspective with profound implications for PEFT design: Rank in LoRA plays the role of effective model width.



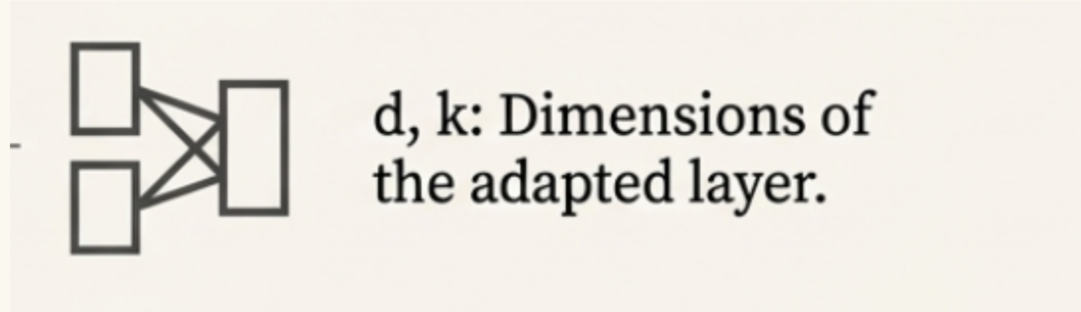
Just as increasing the width of a full model can induce double descent, increasing the rank of a LoRA adapter serves the same function. This reframes PEFT design. We are not just counting parameters; we are controlling the effective capacity of the fine-tuning update to strategically navigate the double descent curve

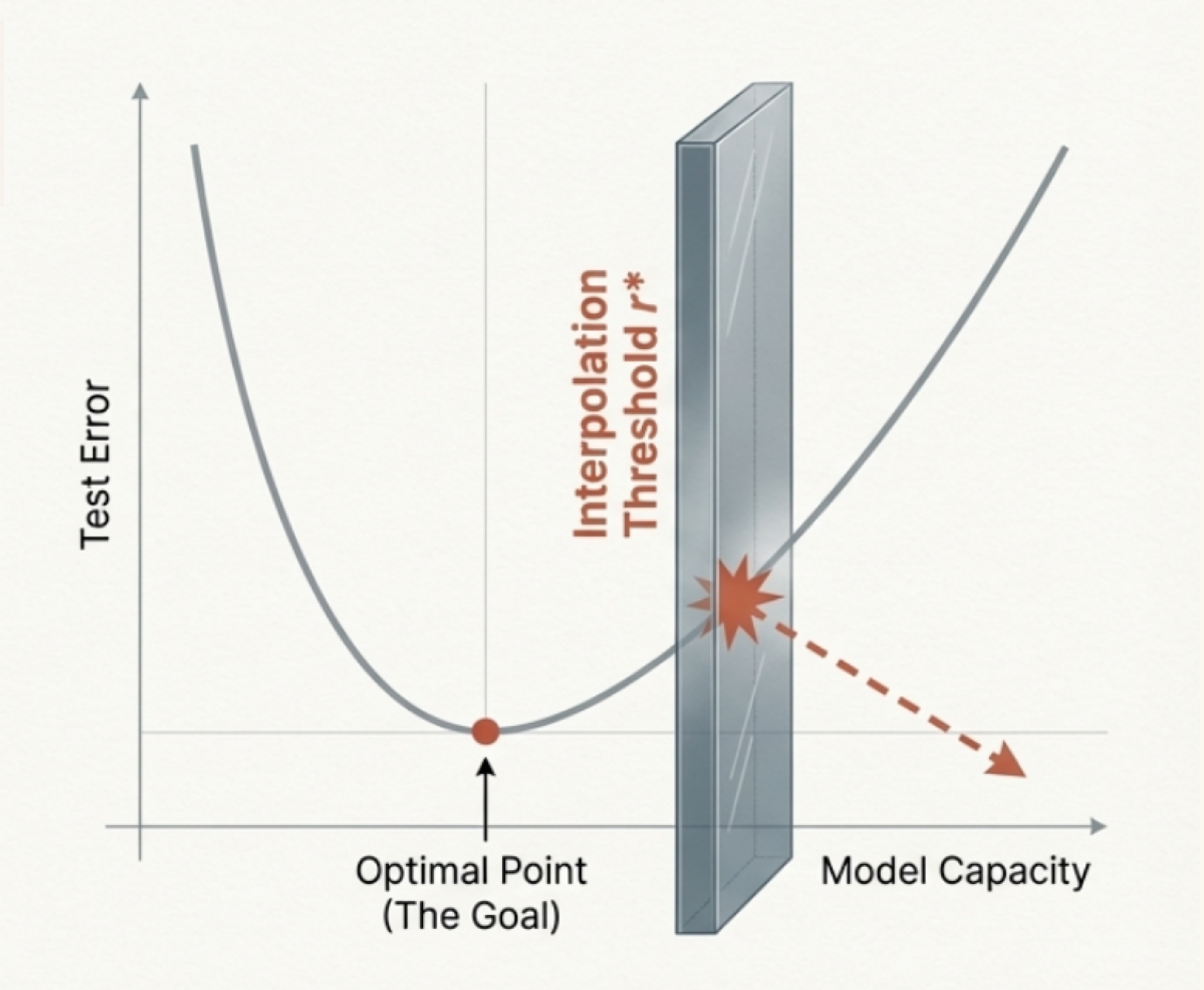
Hidden Flaw: Fixed Rank has an “Interpolation Ceiling”

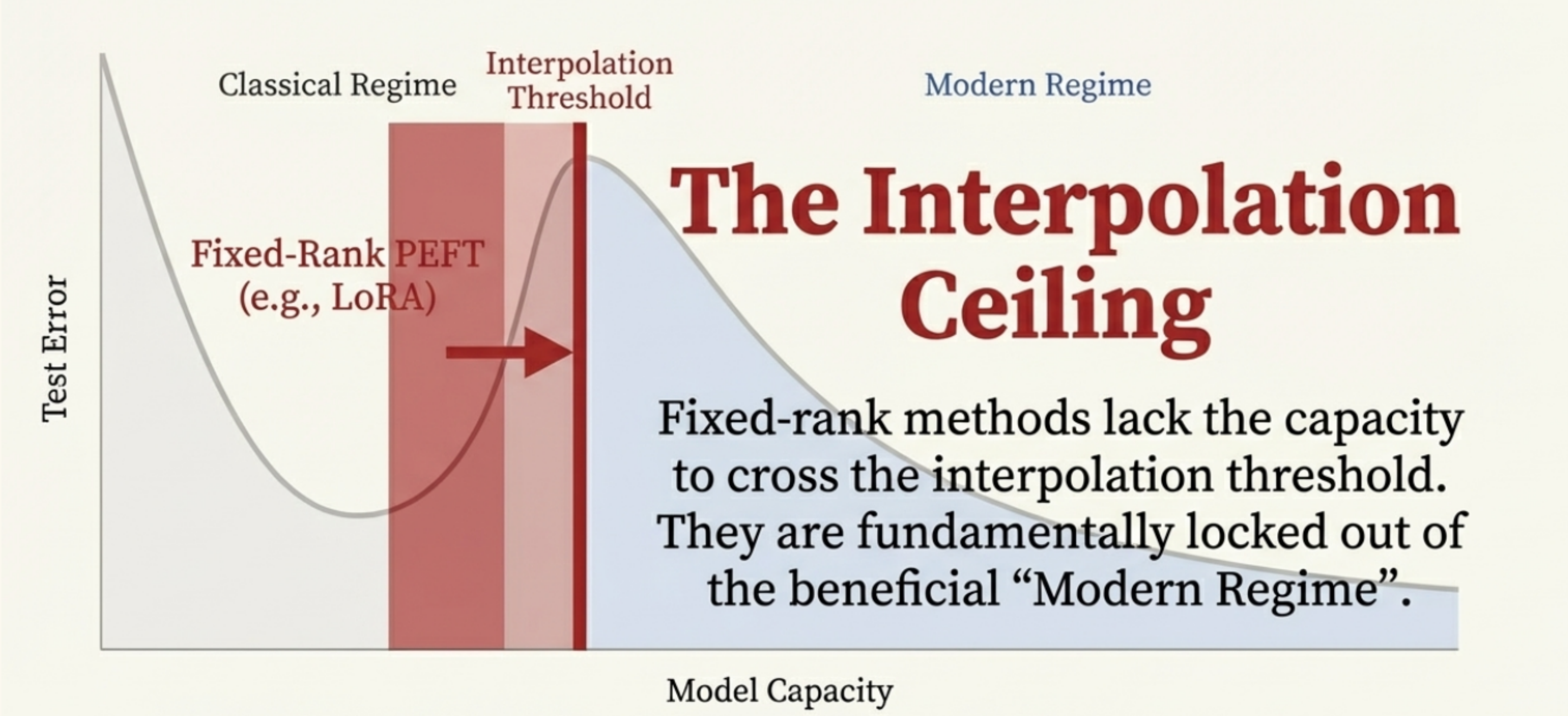
A LoRA adapter with fixed rank $\boldsymbol{r}$ has a maximum parameter count max $(P\_max) = r\,(d + k)$.

If P_max < n (number of training samples), the model cannot achieve zero training error (interpolation).

This "Interpolation Ceiling" prevents fixed-rank methods from accessing the beneficial overparameteriized regime where modern generalization theory applies.

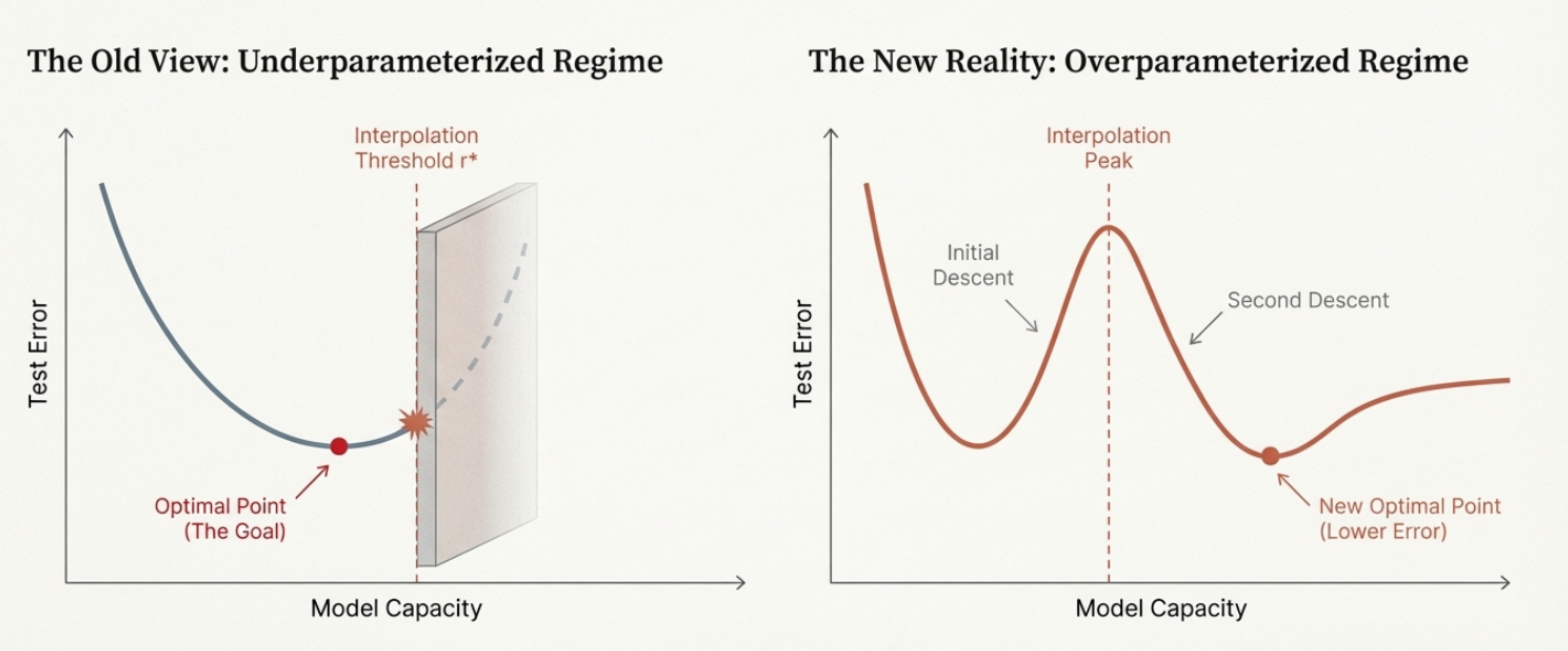






Beyond the Ceiling: Double Descent is an Opportunity, Not an Artifact

Contrary to the view that treats double descent as a harmful phenomenon to be mitigated, we establish that it represents an untapped opportunity for improved generalization in PEFT.



The Opportunity Theorem: How Overparameterization Unlocks Performance

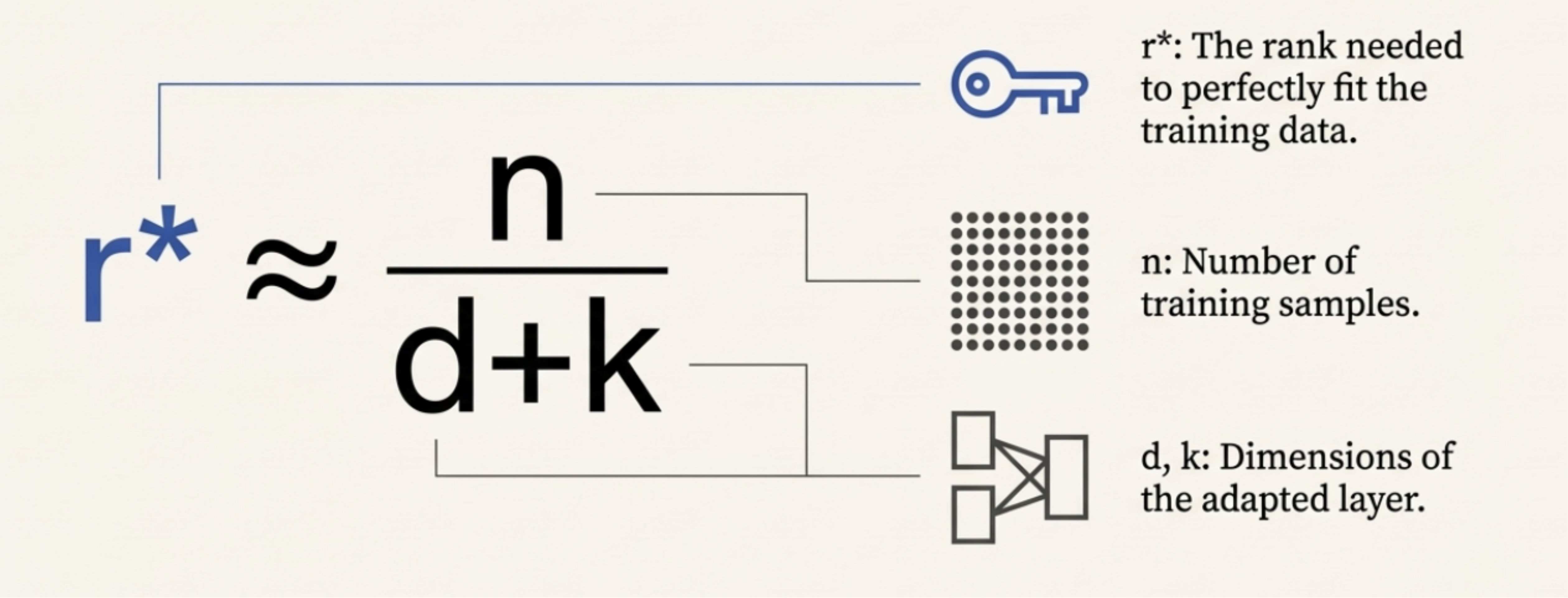
For a LoRA adapter, there exists a critical interpolation threshold $r^* = \lceil n/(d+k) \rceil$.

Underparameterized ($r < r^*$):
Test error follows the classic U-shaped curve.

$$\mathbb{E}[L_{\text{test}}] = O\left(\frac{1}{r} + \frac{r}{n}\right)$$

Overparameterized ($r > r^*$):
Test error exhibits double descent. The key term shows error *decreases* as rank `$r$ moves away from `r^*`.

$$\mathbb{E}[L_{\text{test}}] = O\left(\frac{1}{r} + \frac{\sigma^2}{r - r^*}\right)$$



The threshold ***(r*)*** isn't a fixed constant. It depends on the data (n) and the specific layer (***d***, ***k***). A one-size-fits-all rank is theoretically suboptimal.

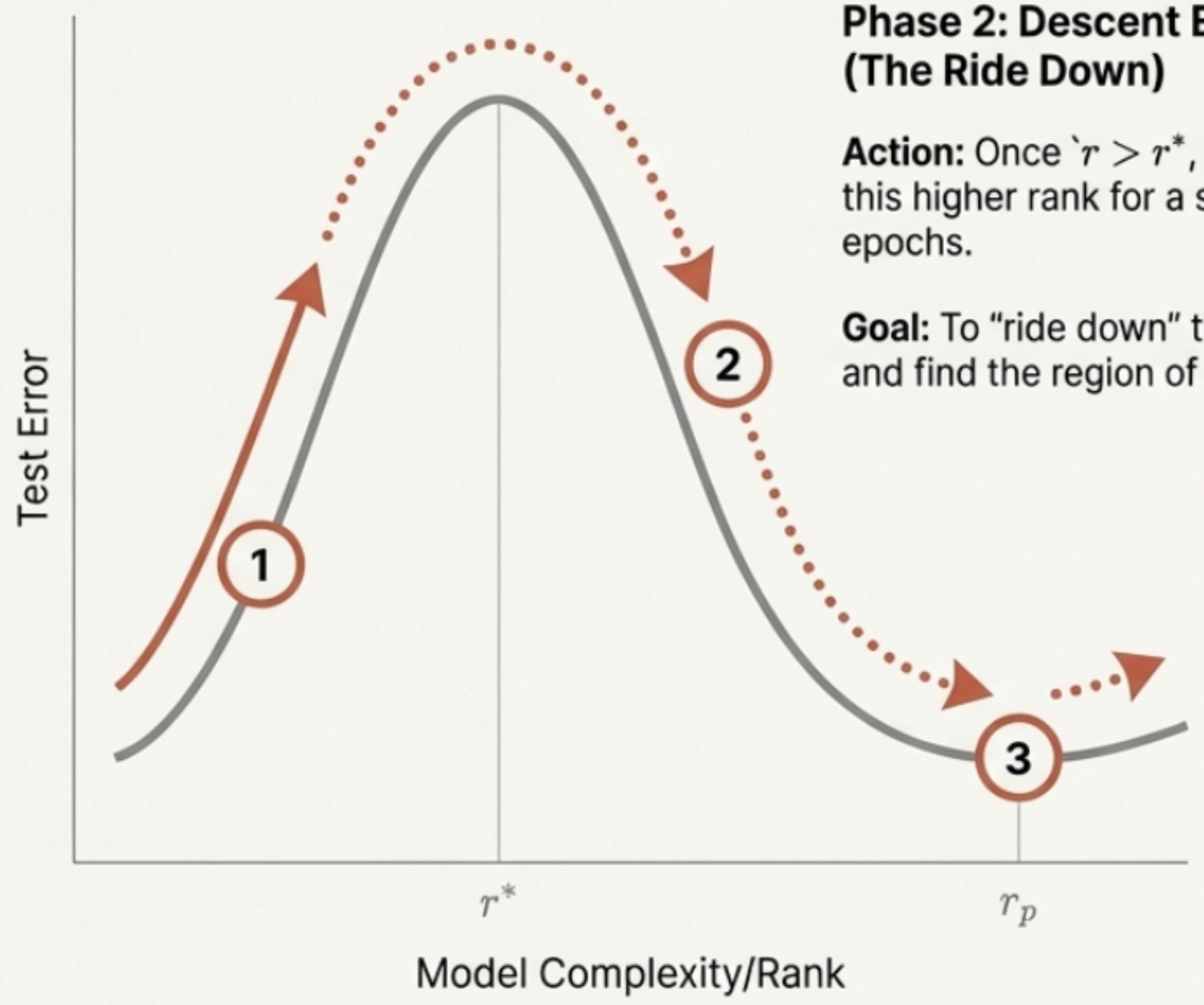
Takeaway : once r a n k is pushed past the critical threshold ' r* , test error begins to fall again. Fixed-rank methods, by design, cannot systematically explore this region.

AdaRank: The First PEFT Method to Exploit Double Descent

Phase 1: Rank Inflation (The Ascent)

Action: Start with a low rank `$r$` and incrementally increase it (`$r \leftarrow r + \alpha$`) until training loss approaches zero (`$\mathcal{L}_{train} < \epsilon$`).

Goal: To provably cross the interpolation threshold `$r^*$`.



Phase 2: Descent Exploitation (The Ride Down)

Action: Once `$r > r^*$`, continue training at this higher rank for a set number of epochs.

Goal: To “ride down” the second descent and find the region of lower test error.

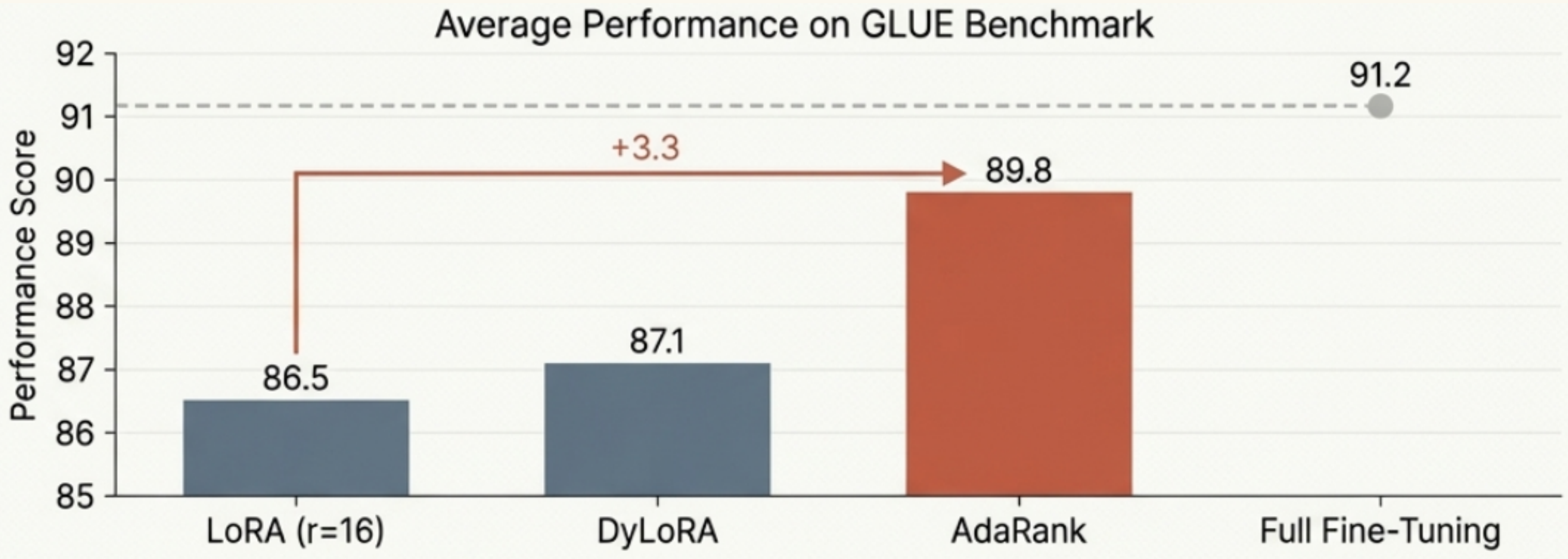
Phase 3: Pruning (Consolidation)

Action: Apply SVD-based truncation to reduce the final rank to `$r_p$`.

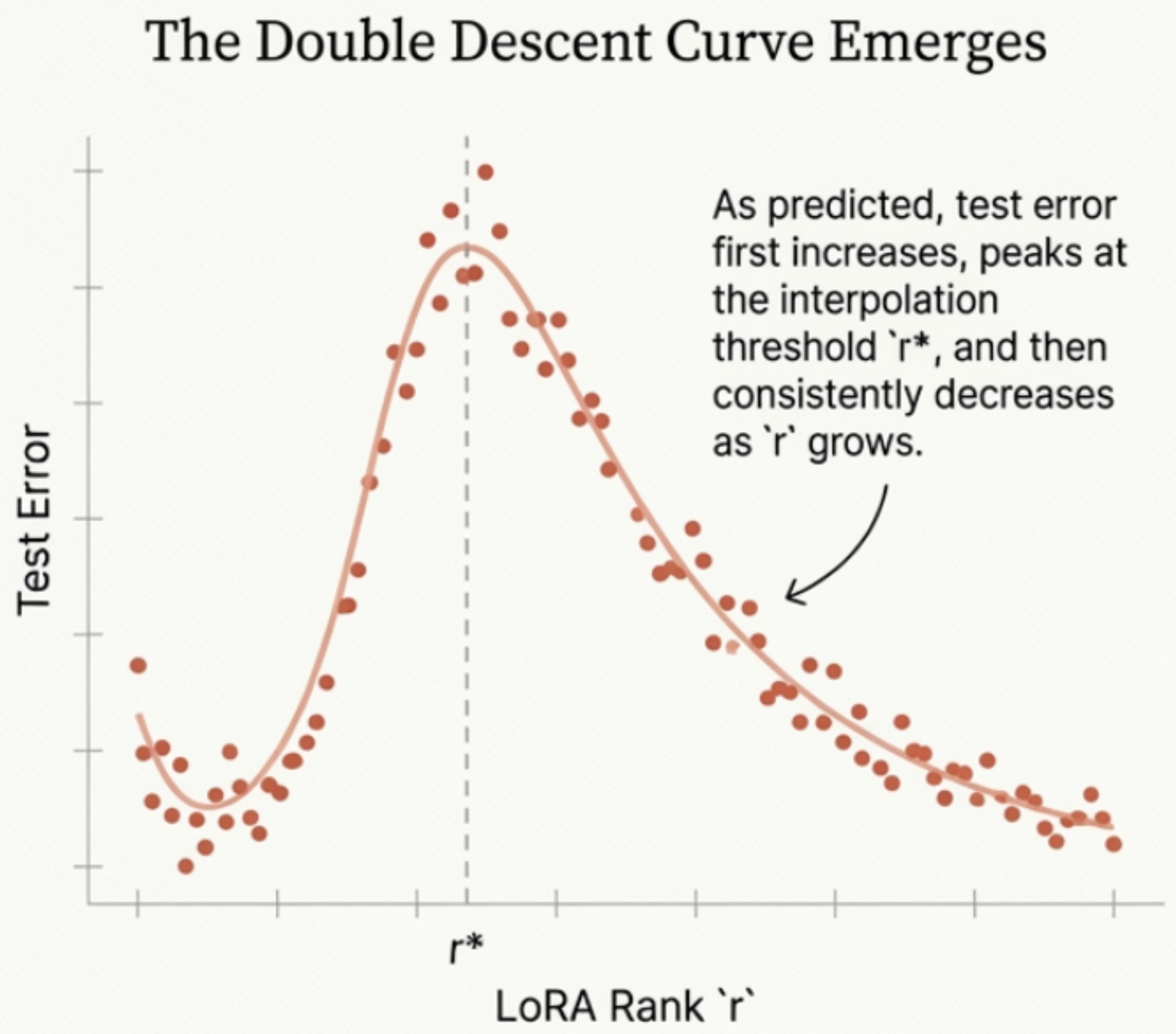
Goal: To preserve generalization gains while reducing final parameter count for efficient inference.

Experimental Validation:

AdaRank Consistently Outperforms State-of-the-Art PEFT Methods. Across NLP benchmarks (GLUE), AdaRank delivers +3.3 improvements over the strongest baselines.
Model used for finetunning: T4-base
Dataset: GLUE (MPRC).

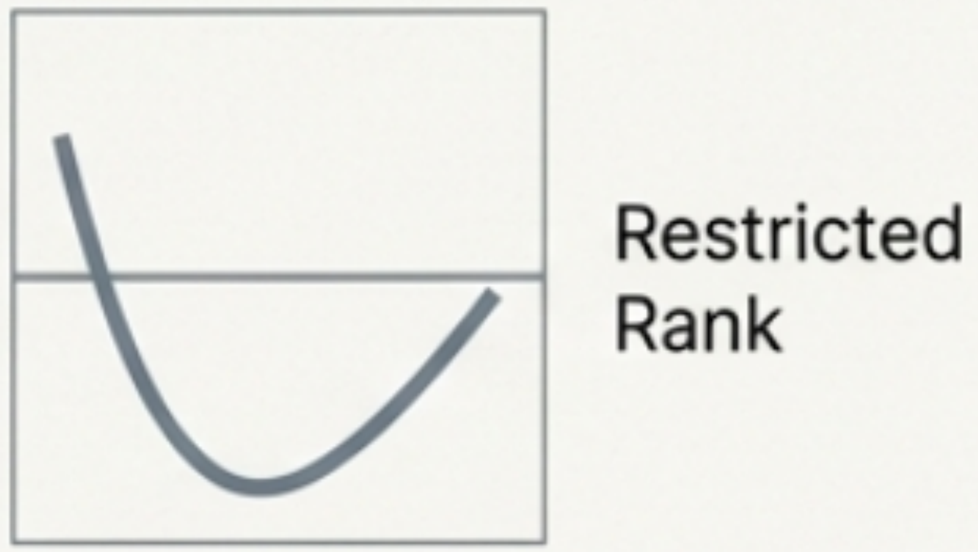


Method	Category	Avg. GLUE Score	Parameter Count
Full Fine-Tuning	Baseline	91.2	100%
LoRA (r=16)	Fixed-rank PEFT	86.5	0.01%
DyLoRA	Adaptive-rank PEFT	87.1	~0.01%
AdaRank	**Interpolation-Aware**	**89.8 (+3.3 vs LoRA)**	**~0.02% (final)**



The Paradigm Shift: From Parameter Counting to Principled Adaptation (Summary)

The Old Way (Fixed-Rank PEFT)



- Mindset: Less is more.
- Goal: Minimise trainable parameters.
- Core Belief: Overfitting is always harmful.
- Strategy: Choose a small, fixed rank `r` and stay below the interpolation threshold.

The New Way (Interpolation-Aware PEFT)

- Mindset: Harness controlled overparameterization.
- Goal: Find the optimal operating point on the double descent curve.
- Core Belief: Benign overfitting is a powerful tool for generalization.
- Strategy: Dynamically adapt rank `r` to cross the interpolation threshold and exploit the second descent.

Conclusion

From Counting Parameters to Crossing Thresholds

1. We identified a fundamental **'Interpolation Ceiling'** that limits all fixed-rank PEFT methods.
2. We proved this ceiling prevents access to the **beneficial 'double descent' regime**, where more parameters can lead to better generalization.
3. We designed **AdaRank**, a theory-driven method **that strategically crosses the threshold** to achieve new state-of-the-art performance.

Thankyou

Questions?