

Foresight: Adaptive Layer Reuse for Accelerated and High-Quality Text-to-Video Generation

Muhammad Adnan, Nithesh Kurella, Akhil Arunkumar, Prashant J. Nair

NeurIPS, 2025
San Diego, USA



THE UNIVERSITY
OF BRITISH COLUMBIA



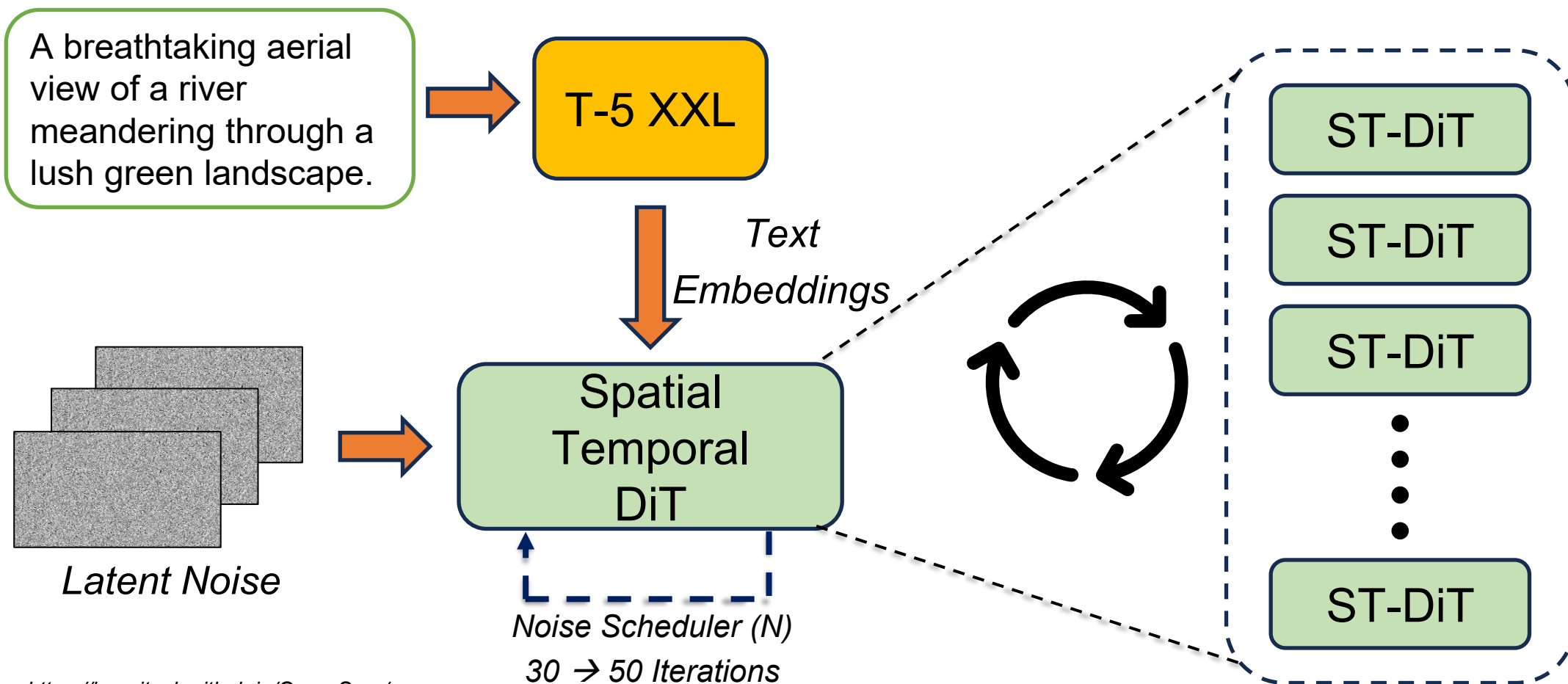
Background

Forward Process: $q(X_t|X_{t-1}) = \mathcal{N}(X_t|\sqrt{\alpha_t}X_{t-1}, (1 - \alpha_t)I)$

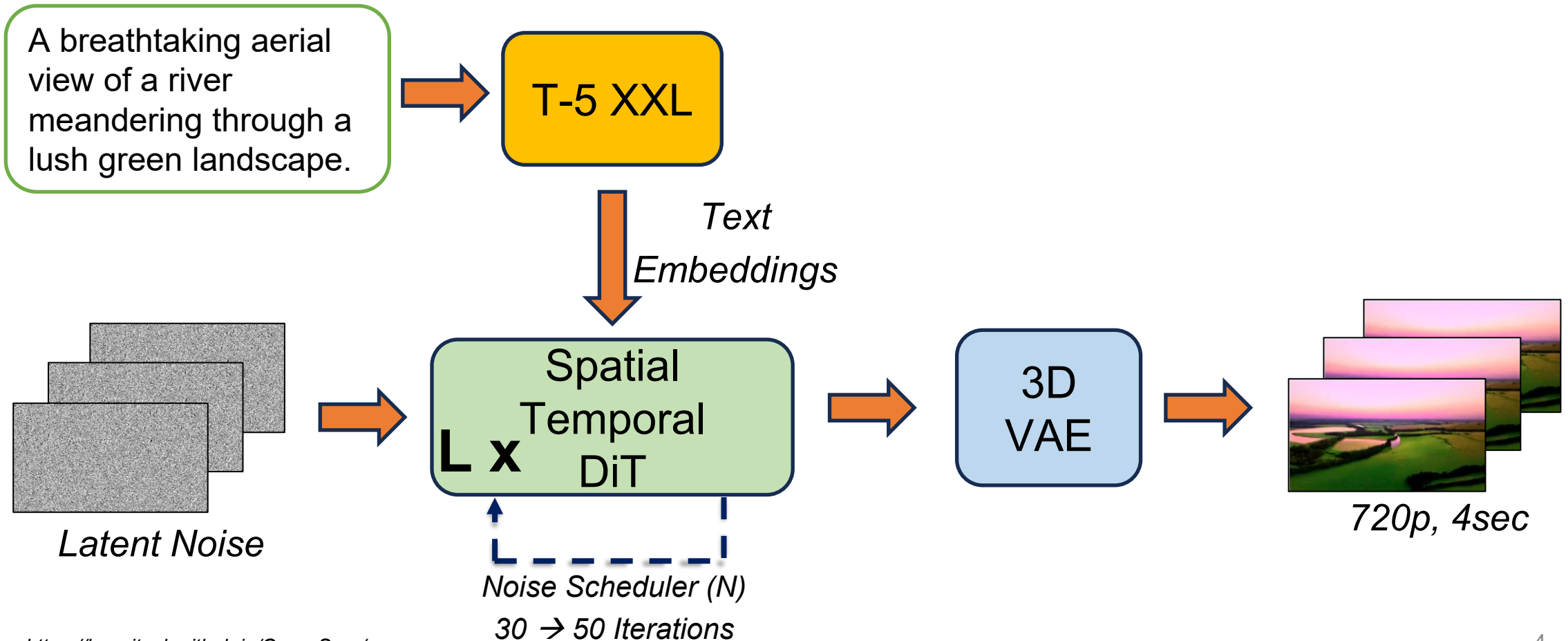


Reverse Process: $p(X_{t-1}|X_t) \approx \mathcal{N}(X_{t-1}|\mu_\theta(\cdot, t), \sigma_t)$

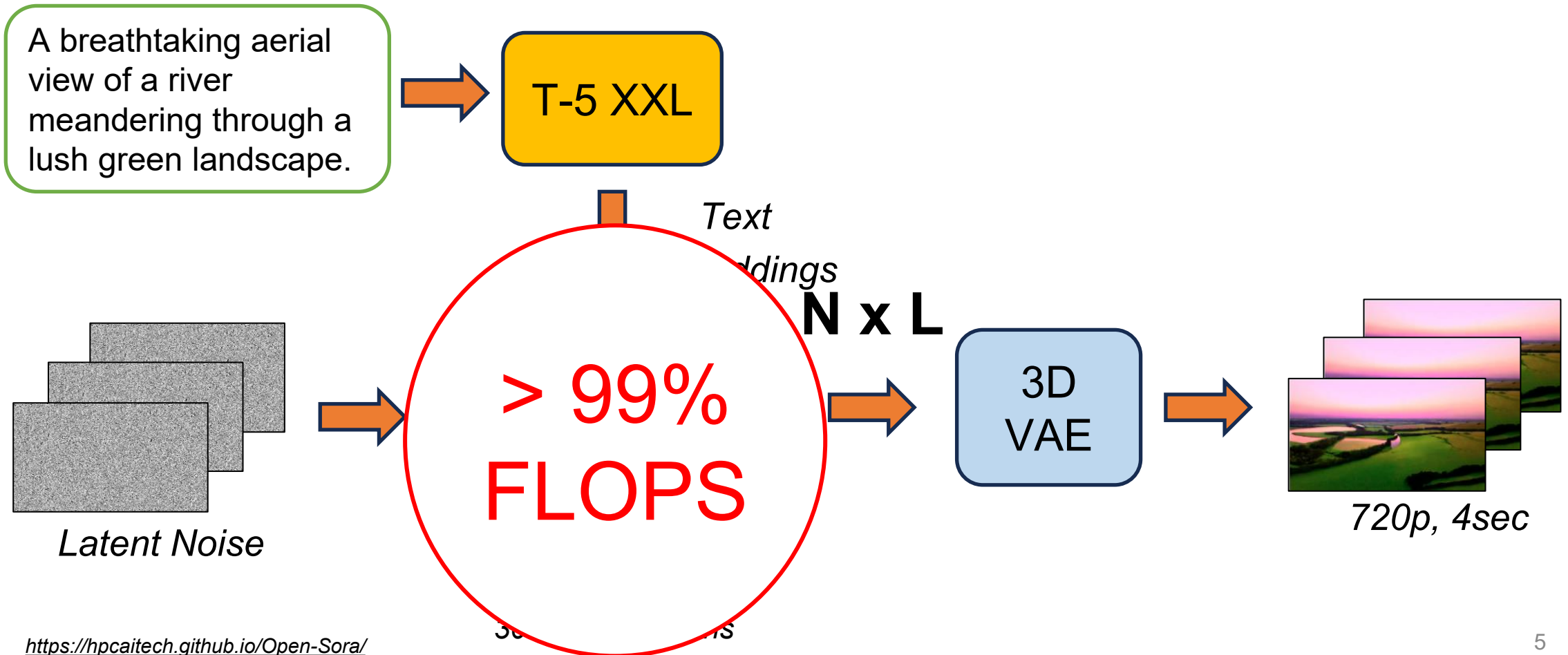
Text-to-Video Generation



Text-to-Video Generation

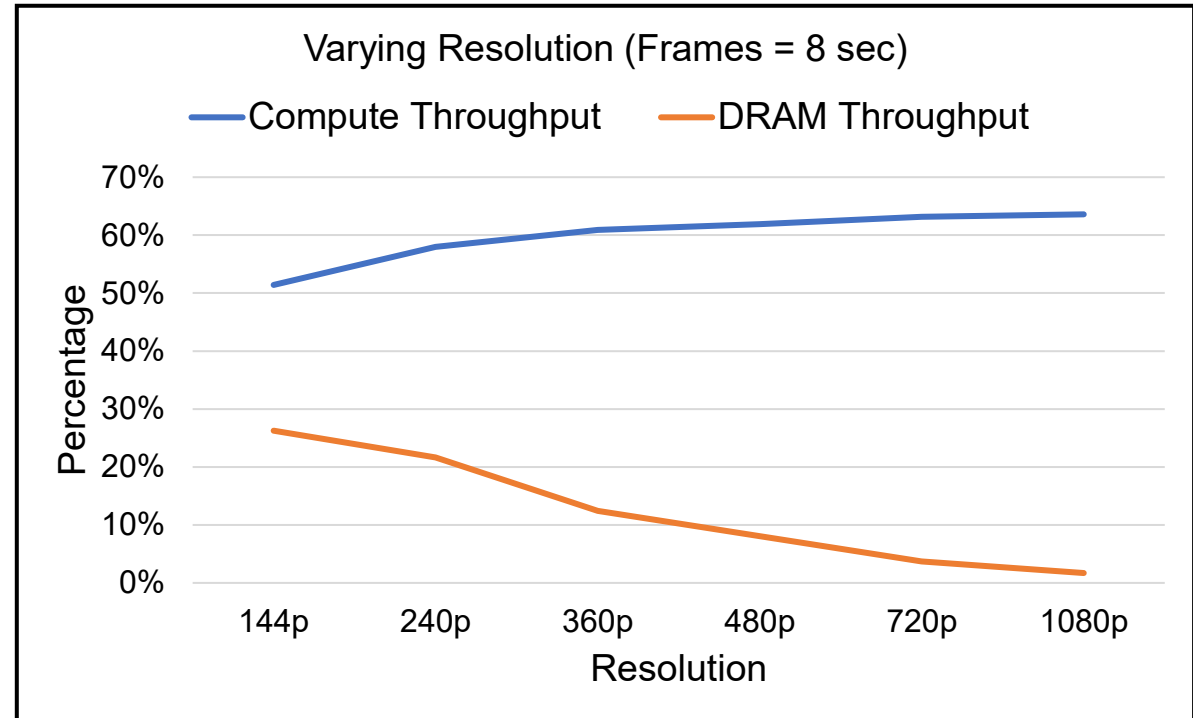


Text-to-Video Generation



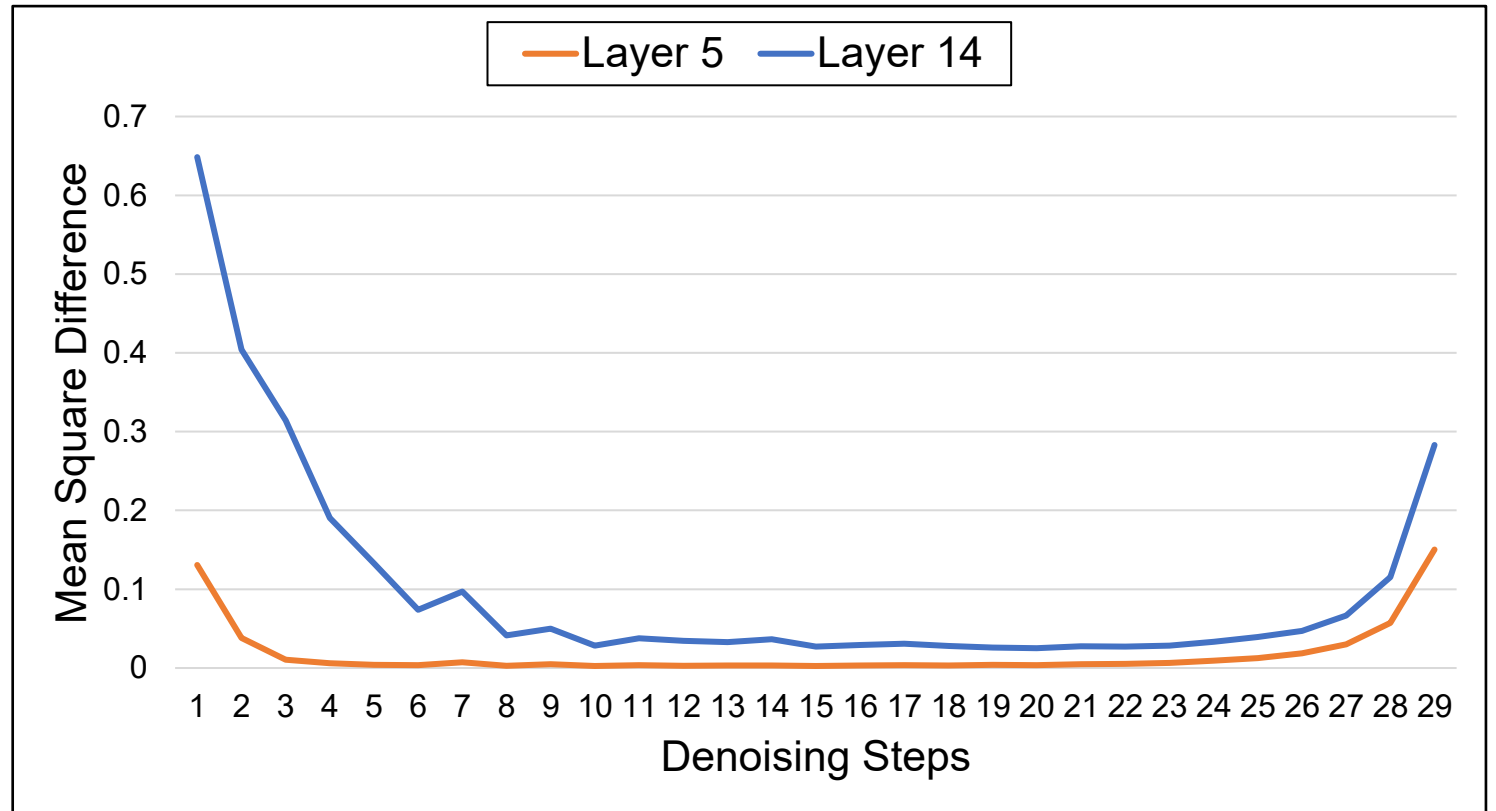
Compute vs. Memory

- Compute bound



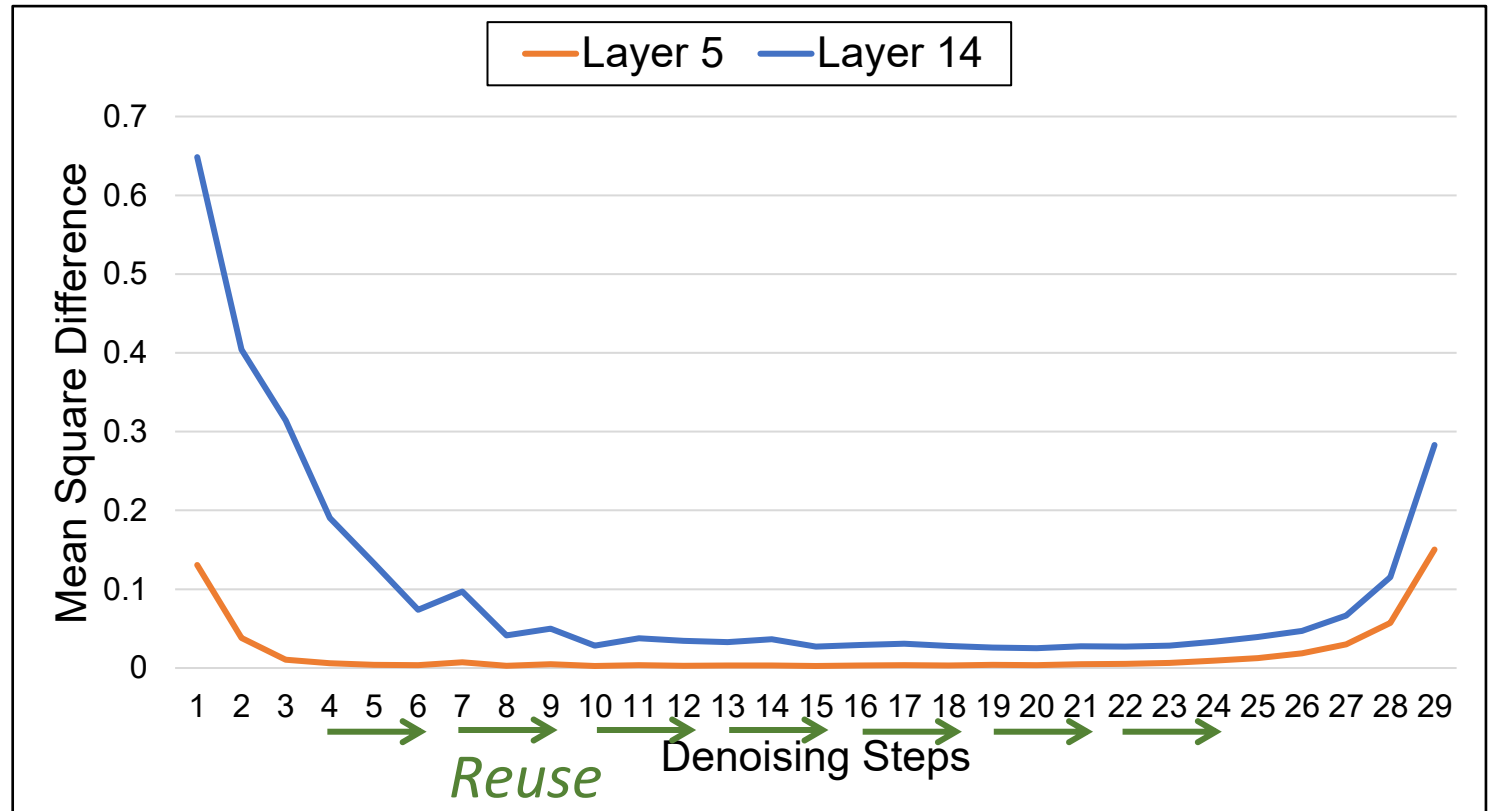
Motivation

- Intermediate Features reuse across Denoising Steps



Static Reuse

- Intermediate Features can be reused after fixed no of denoising steps



Qualitative Comparison

Prompt: A white and orange tabby cat is seen happily darting through a dense garden, as if chasing something. Its eyes are wide and happy as it jogs forward, scanning the branches, flowers, and leaves as it walks.



Baseline



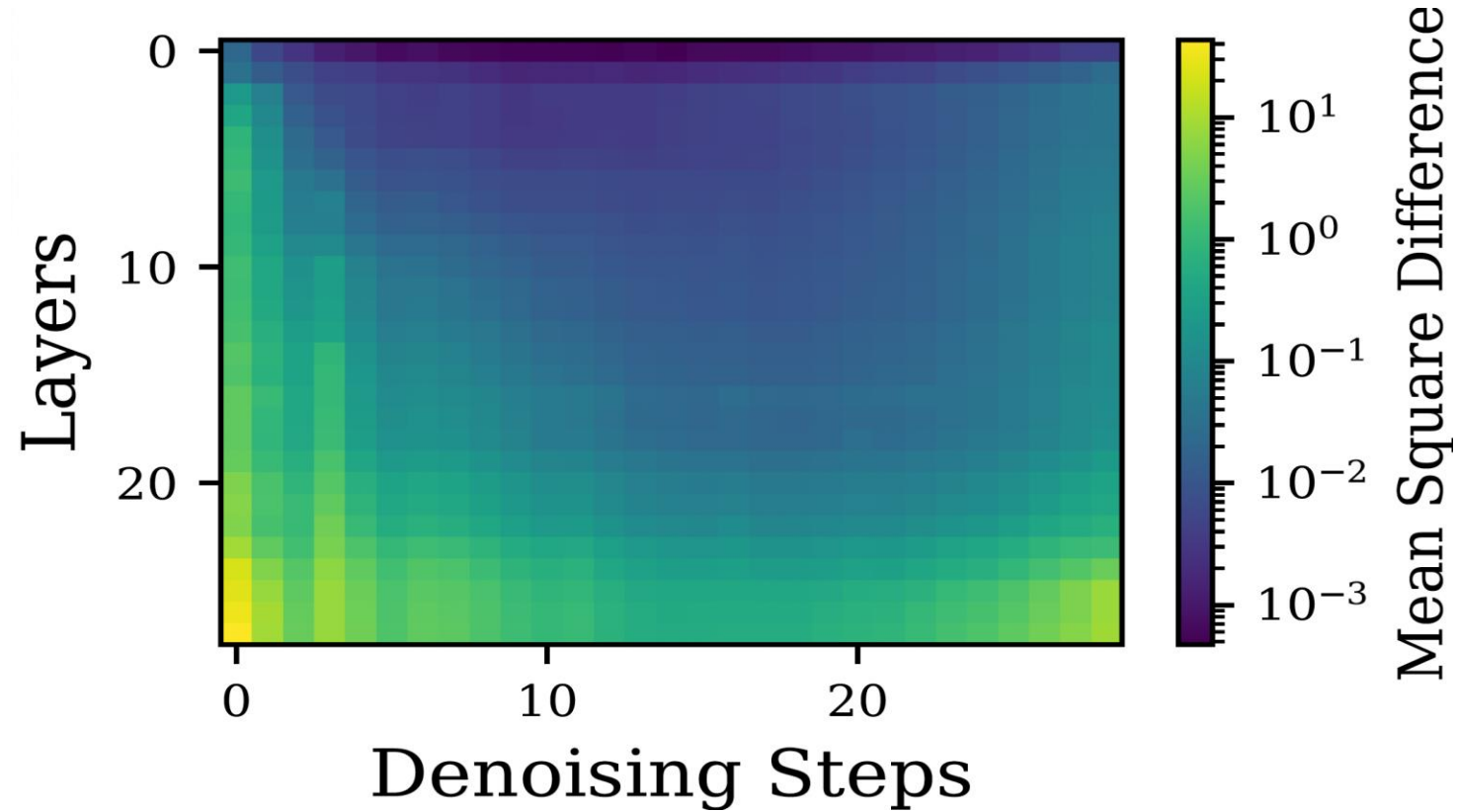
Static



PAB (Static)

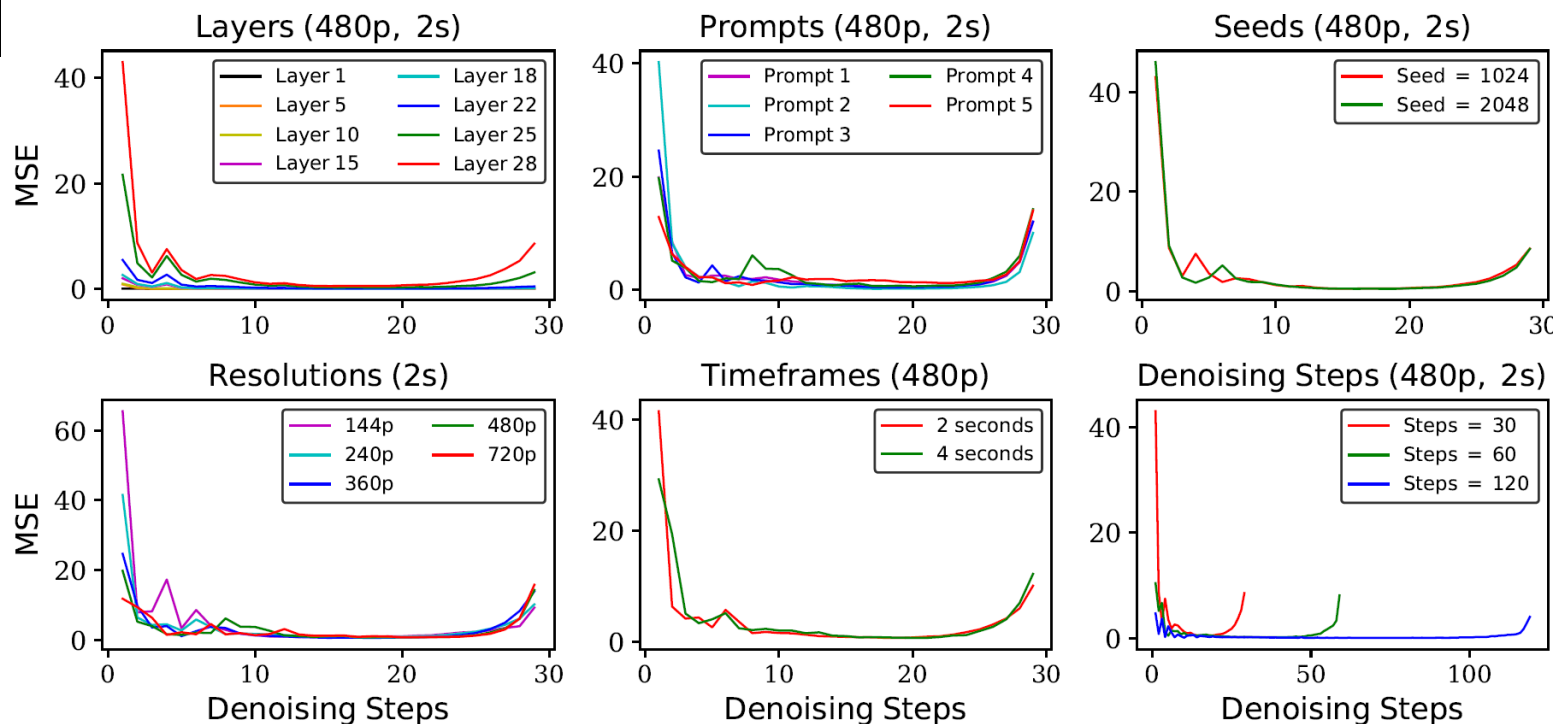
Key Insight - I

- Each layers has different reuse pattern



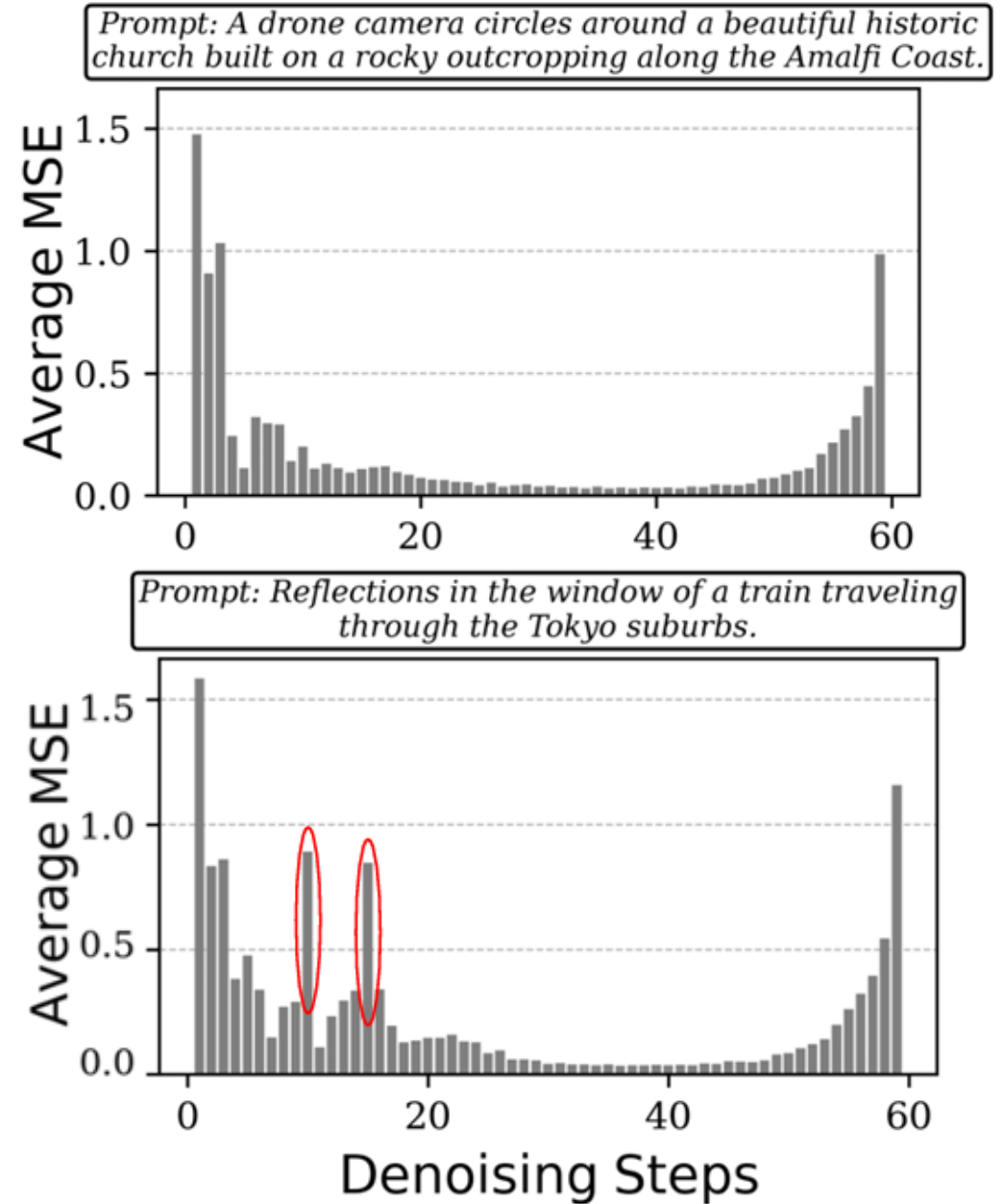
Key Insight - II

- Reuse pattern varies across different video configurations



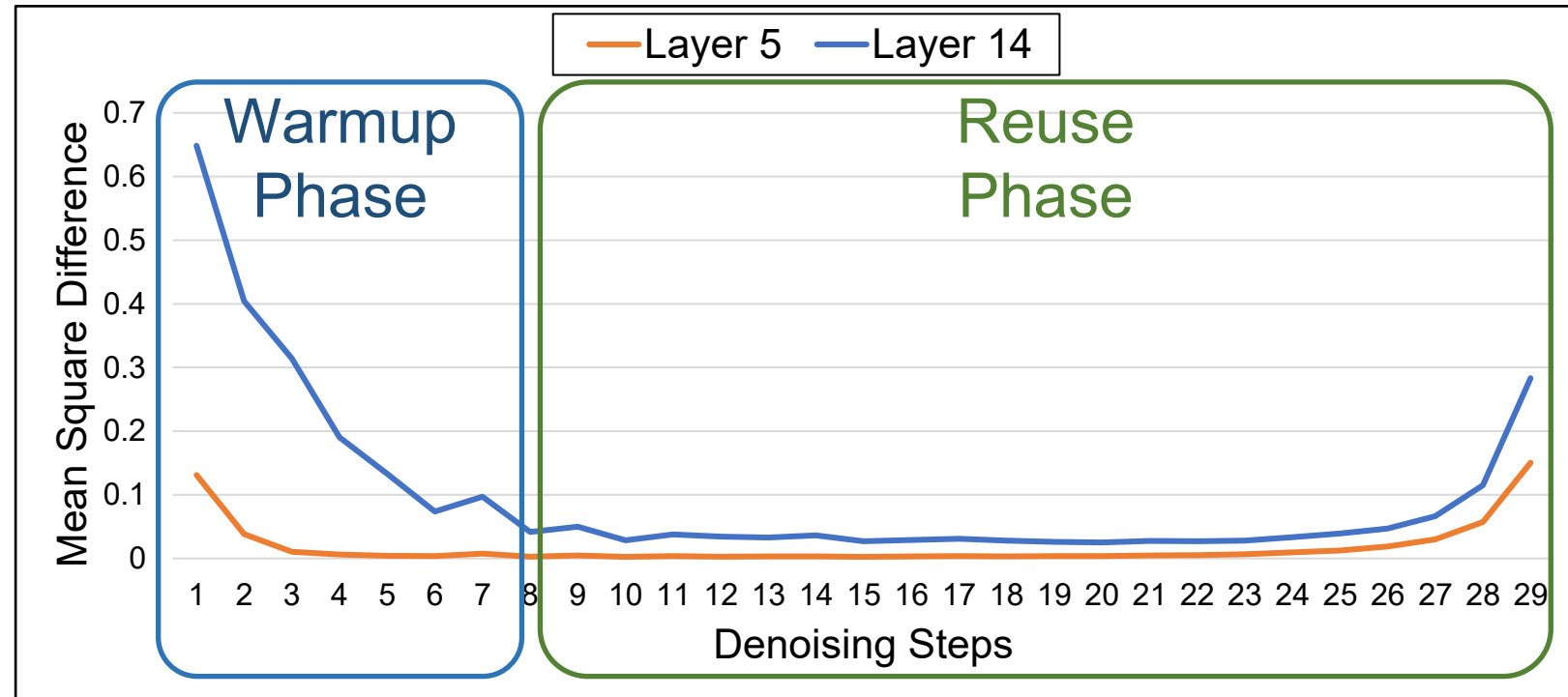
Key Insight - III

- Prompt-dependent dynamics of intermediate features

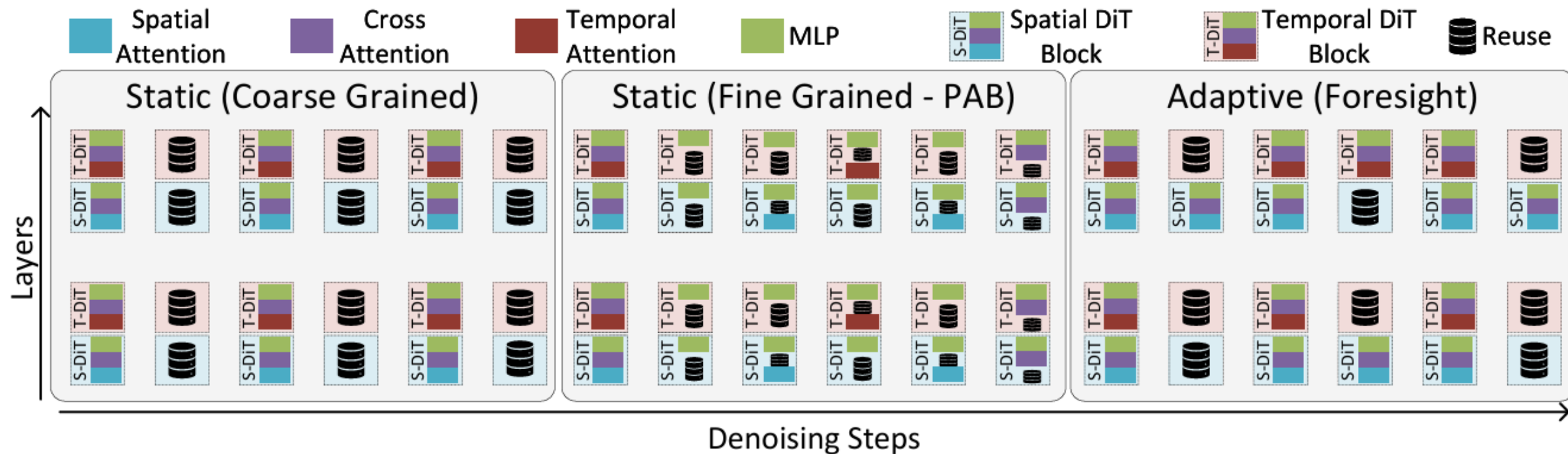


Foresight

- Adaptive/Dynamics Layer Reuse



Foresight



Foresight

- Adaptive/Dynamics Layer Reuse



Baseline



Foresight (Adaptive)

Foresight

- Adaptive/Dynamics Layer Reuse

