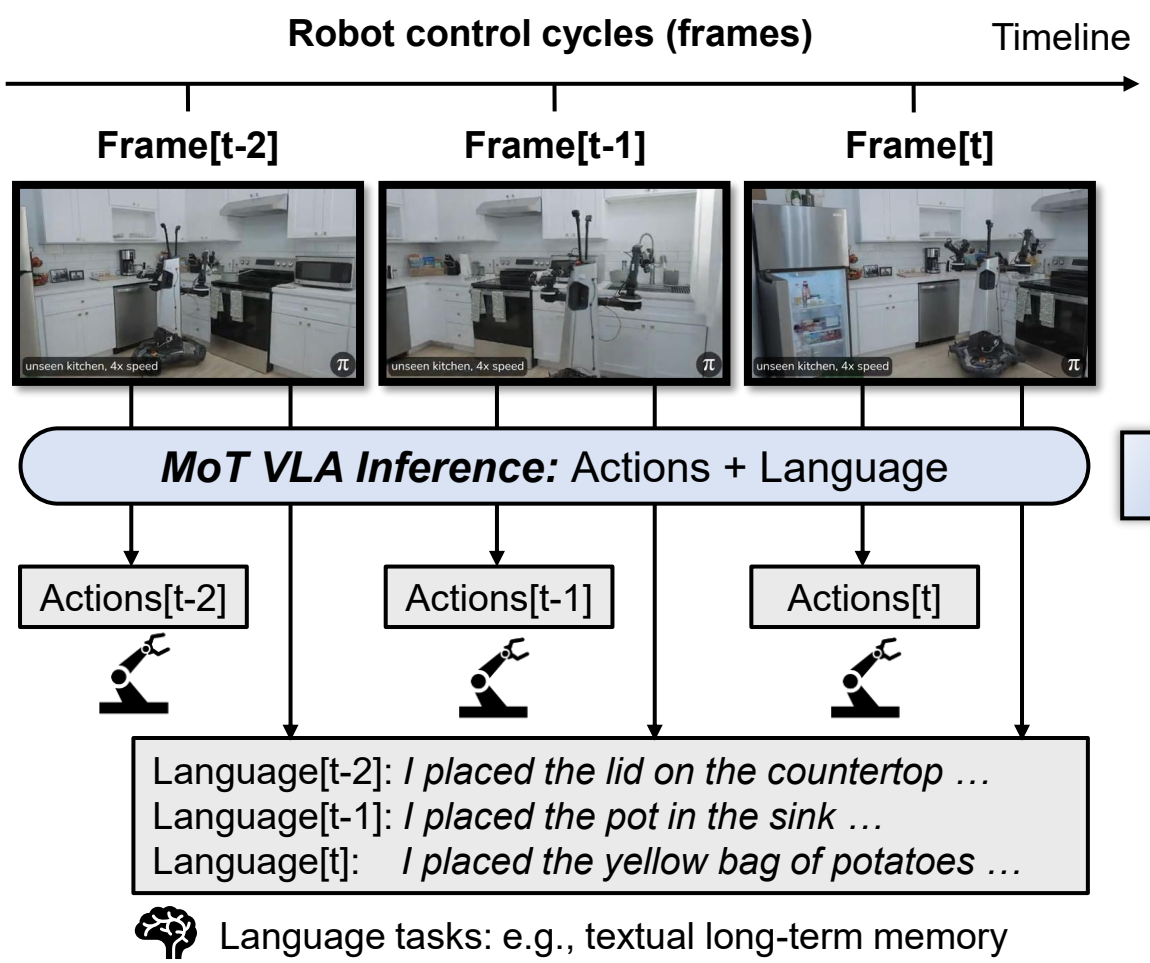
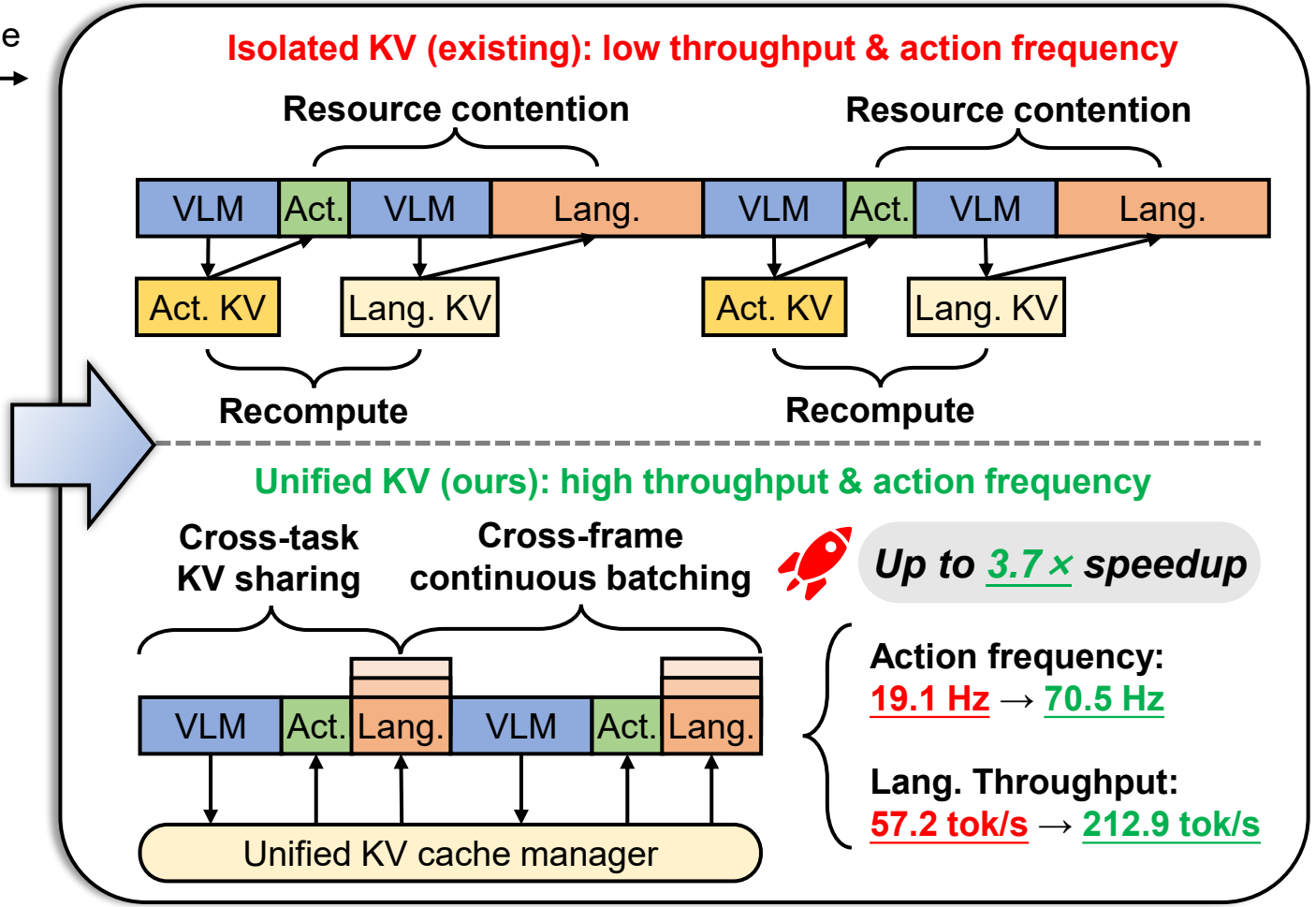

Robot control cycles (frames)
Timeline
Frame[t-2]
Frame[t-1]
Frame[t]
unseen kitchen, 4x speed
MoT VLA Inference: Actions + Language
Actions[t-2]
Actions[t-1]
Actions[t]
Language[t-2]: I placed the lid on the countertop …
Language[t-1]: I placed the pot in the sink …
Language[t]: I placed the yellow bag of potatoes …
Language tasks: e.g., textual long-term memory


Isolated KV (existing): low throughput & action frequency
Resource contention
Resource contention
VLM
Act.
VLM
Lang.
VLM
Act.
VLM
Lang.
Act. KV
Lang. KV
Act. KV
Lang. KV
Recompute
Recompute
Unified KV (ours): high throughput & action frequency
Cross-task KV sharing
Cross-frame continuous batching
Up to 3.7 × speedup
VLM
Act.
Lang.
VLM
Act.
Lang.
Unified KV cache manager
Action frequency:
19.1 Hz → 70.5 Hz
Lang. Throughput:
57.2 tok/s → 212.9 tok/s