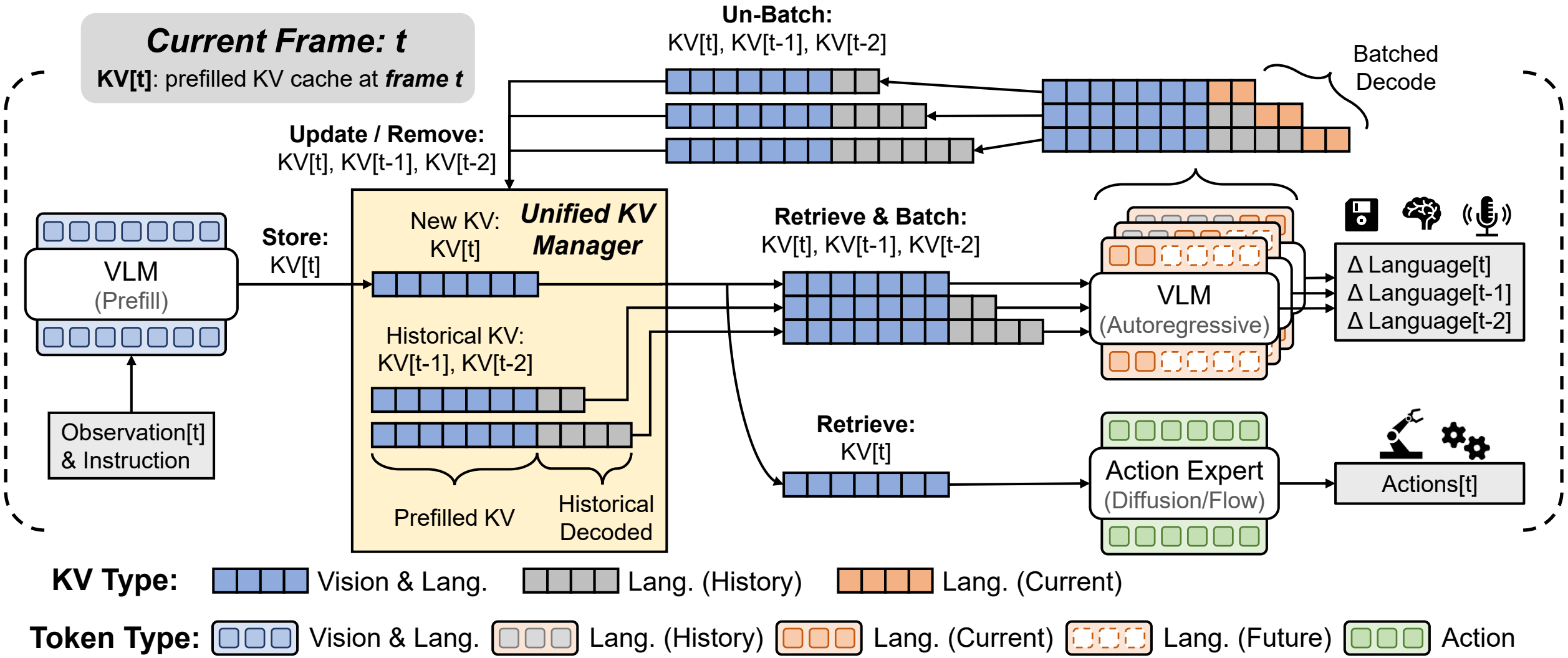
Current Frame: t
KV[t]: prefilled KV cache at frame t
Un-Batch:
KV[t], KV[t-1], KV[t-2]
Batched Decode
Update / Remove:
KV[t], KV[t-1], KV[t-2]
VLM
(Prefill)
Store:
KV[t]
New KV:
KV[t]
Unified KV Manager
Historical KV:
KV[t-1], KV[t-2]
Prefilled KV
Historical Decoded
Observation[t] & Instruction
Retrieve & Batch:
KV[t], KV[t-1], KV[t-2]
VLM
(Autoregressive)
Δ Language[t]
Δ Language[t-1]
Δ Language[t-2]
Retrieve:
KV[t]
Action Expert
(Diffusion/Flow)
Actions[t]
KV Type:
Vision & Lang.
Lang. (History)
Lang. (Current)
Token Type:
Vision & Lang.
Lang. (History)
Lang. (Current)
Lang. (Future)
Action