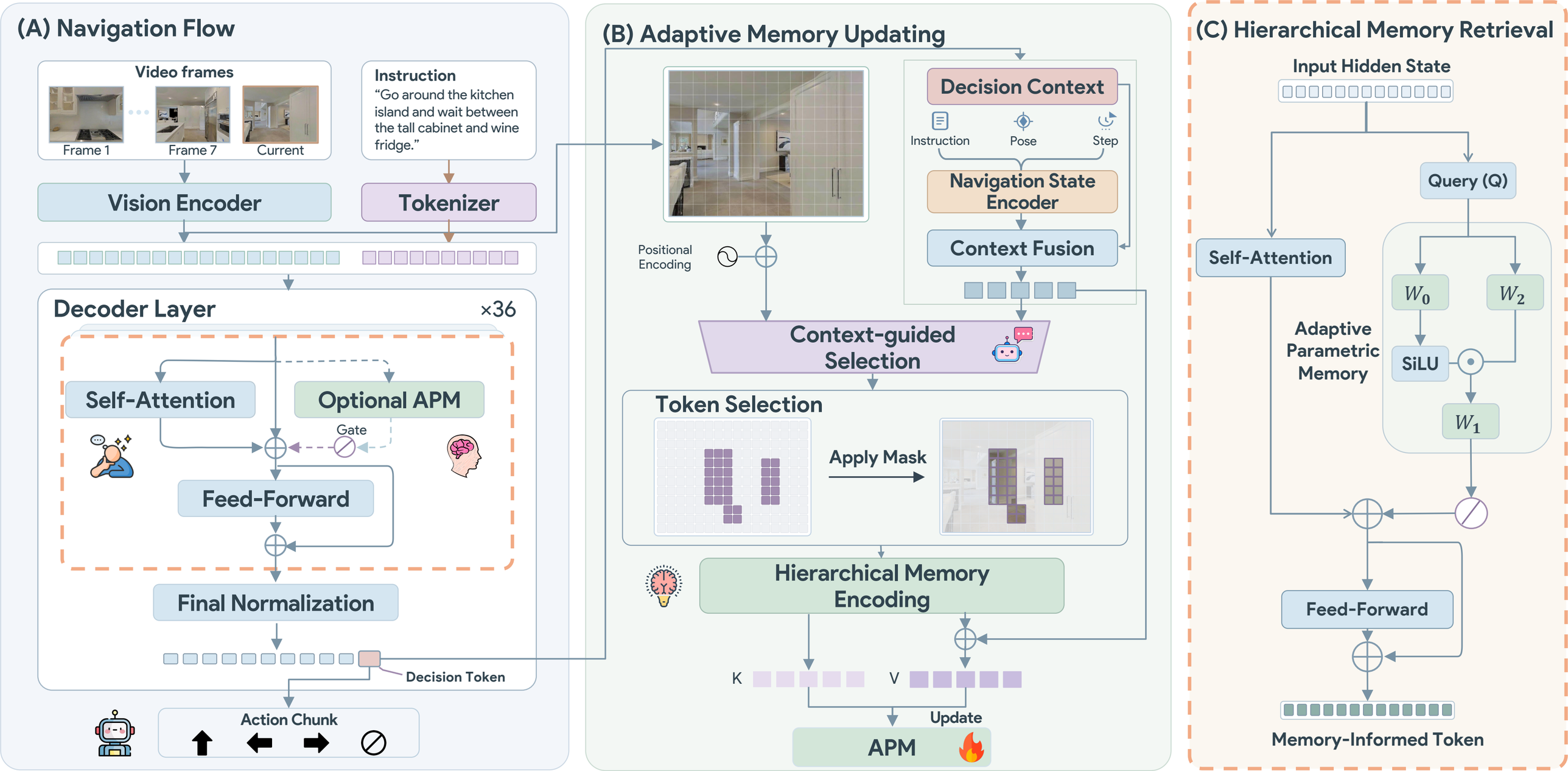
(A) Navigation Flow
Video frames
Frame 1
Frame 7
Current
Instruction
"Go around the kitchen island and wait between the tall cabinet and wine fridge."
Vision Encoder
Tokenizer
Decoder Layer
×36
Self-Attention
Optional APM
Gate
Feed-Forward
Final Normalization
Decision Token
Action Chunk
(B) Adaptive Memory Updating
Decision Context
Instruction
Pose
Step
Navigation State Encoder
Context Fusion
Positional Encoding
Context-guided Selection
Token Selection
Apply Mask
Hierarchical Memory Encoding
K
V
Update
APM
(C) Hierarchical Memory Retrieval
Input Hidden State
Query (Q)
Self-Attention
Adaptive Parametric Memory
W_0
W_2
SiLU
W_1
Feed-Forward
Memory-Informed Token