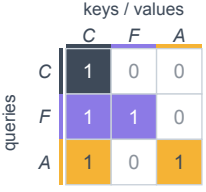
Training



filled: allowed; empty: blocked
bidirectional within a group

Shared Wan2.2-5B video DiT

Current *C*
clean RGB/XYZ, $\tau_C = 0$

Future *F*
noisy future RGB/XYZ

C reads *C* only
no future relay

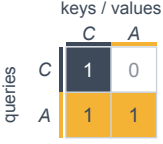
Action expert *A*
noisy action chunk

Separate T5 text conditioning to all groups

No *F* ↔ *A* attention

Inference

Future *F* and auxiliary CNN absent



current time stays zero
$\tau_C = 0$

Current RGB/XYZ
VAE + width fusion

Wan once
current *C* only

Cache current K/V
fixed, $\tau_C = 0$

Action expert
10 action-flow updates

noise → clean
τ_a : 1 → 0

clean actions

Inverse action normalization

Execute prefix

Replan
new observation