Time Axis

| Frame p | Frame 2p | Frame 3p | ...... | Frame Tp | Frame (T+1)p |

Video Tokens: (N, T, H, W, C)

N-th Denoise

DiT w/ Interactive Module denoise <u>Step x k</u>

Video Tokens: (N, T, H, W, C)

(N+1)-th Denoise

DiT w/ Interactive Module denoise <u>Step x k</u>

......