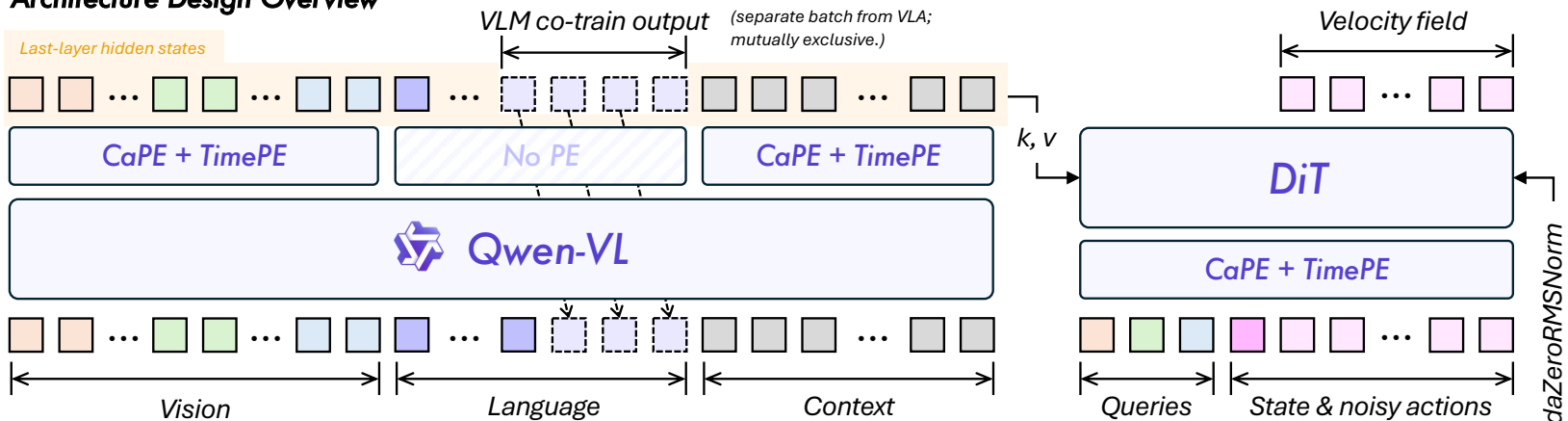
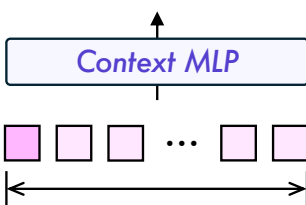


# Architecture Design Overview



Structured Embodiment Prompt	Embodied Chain of Thought
embodiment: robot_aloha instruction: fold the clothes speed: 1000 ...	scene: shirt flat, pinch sleeve progress: not complete next: grasp right sleeve ...

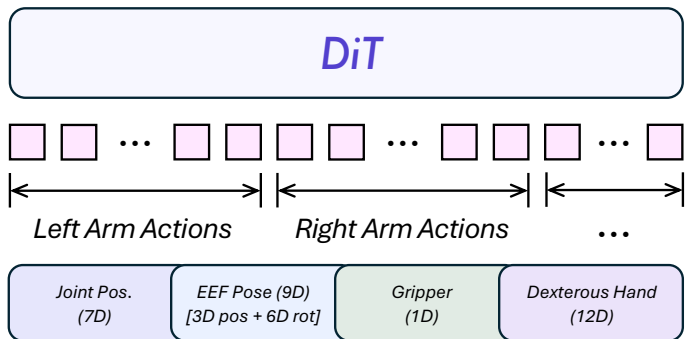


All cameras, sampled timestamps    Instruction & prompt

History state & actions

- Conditions:
- denoise timestep
  - end-effector type
  - has camera parameters

## Unified State & Action Representation



## Unified End-Effector Motion Prediction

