a) Codebook Learning b) Behavior Cloning Image Observations **Action Chunk Robot State Sequence** (Current + Goal) Task ID $\mathbf{S}_{t-h:t}$ $\mathbf{a}_{t:t+k}$ State ϕ_{state} **Action Tokenizer** Encoder Task ϕ_{task} Encoder **Transformer Encoder** \mathbf{Z}_{state} ϕ_{enc}^{act} \mathbf{Z}_{task} Discrete action tokens Codebook **Latent Encoder** ϕ_{enc}^{lat} Codebook **Transformer Decoder** Predicted ψ_{dec} action tokens **Predicted Action**