Project Website

Discrete Codebook World Models for Continuous Control

Aidan Scannell (Amos Storkey), Mohammadreza Nakhaei, Kalle Kujanpää, Yi Zhao, Kevin Luck, Arno Solin, Joni Pajarinen

University of Edinburgh
Finnish Center for Artificial Intelligence (FCAI)
Aalto University

World Models

$$p(s_{t+1}, r_t \mid s_t, a_t)$$

World Models

$$p(\boxed{s_{t+1}}, r_t \mid s_t, a_t)$$

World Models

$$p(\boxed{s_{t+1}}, \boxed{r_t} \mid s_t, a_t)$$

World Models

$$p(s_{t+1}, r_t \mid s_t, \boxed{a_t})$$

World Models

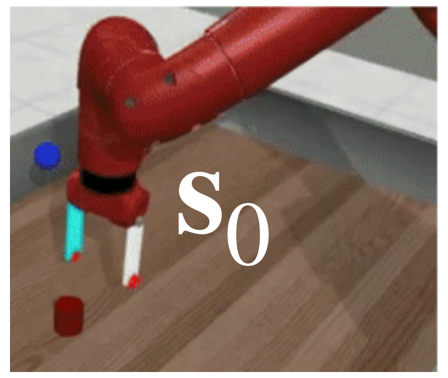
$$p(s_{t+1}, r_t \mid \boxed{s_t}, a_t)$$

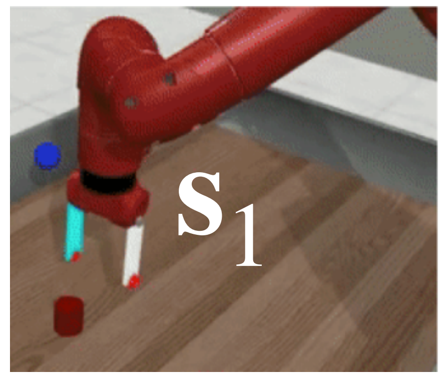
World Models

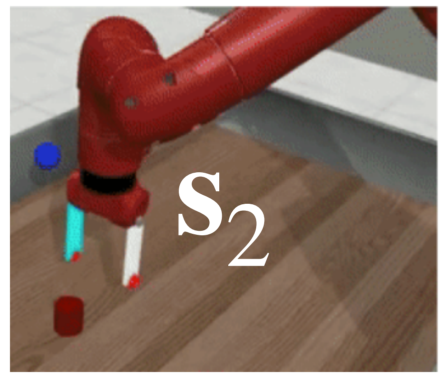
$$p(s_{t+1}, r_t \mid s_t, a_t)$$

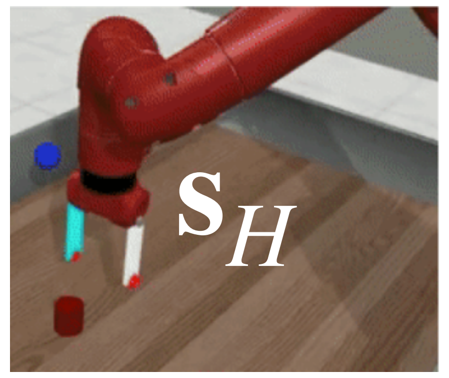
World Models

World Models

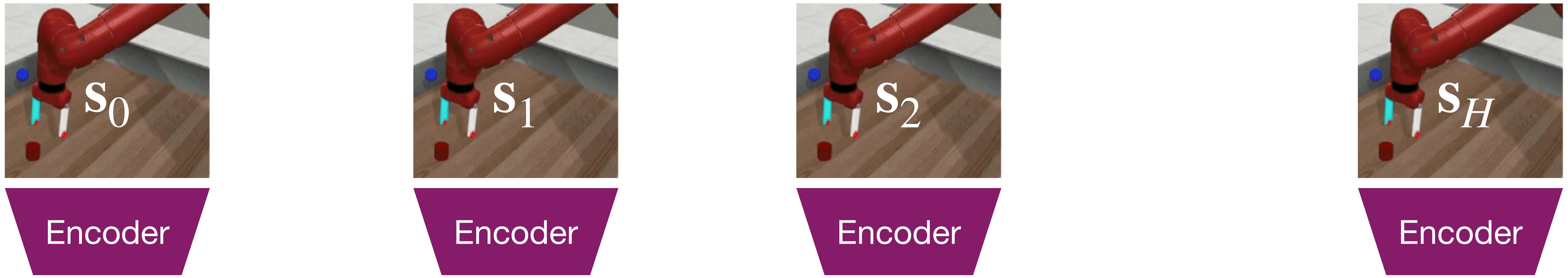




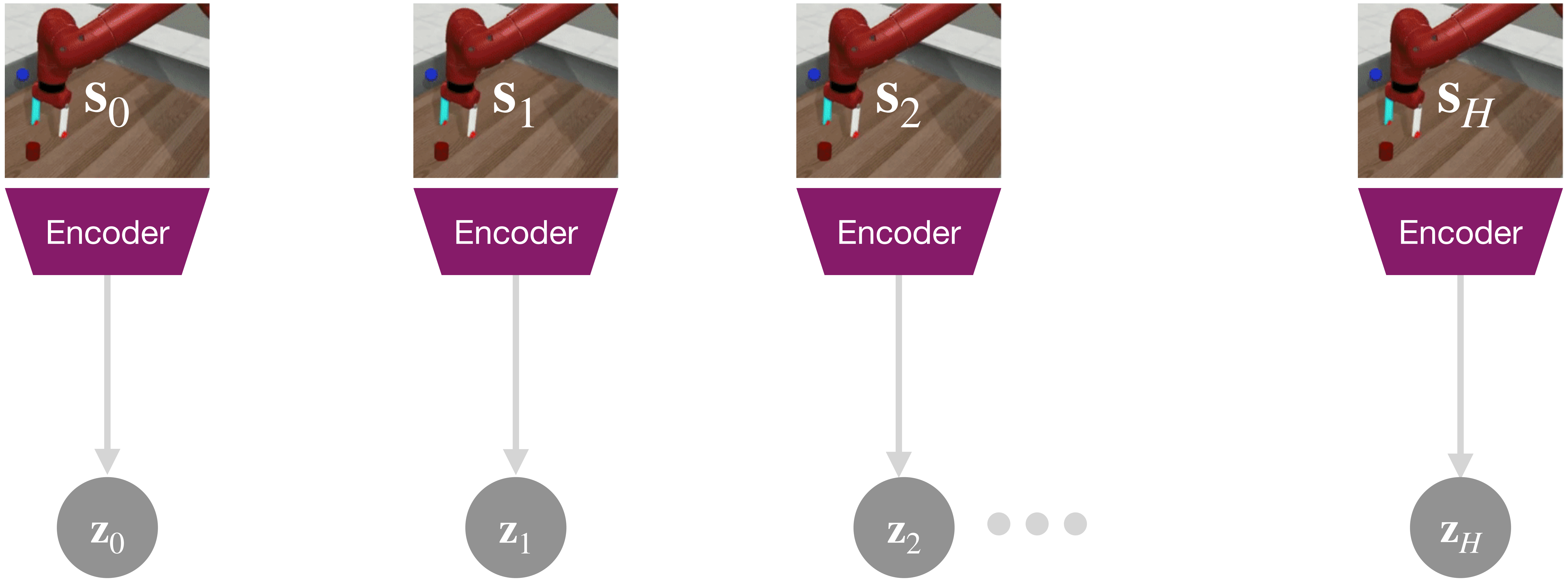




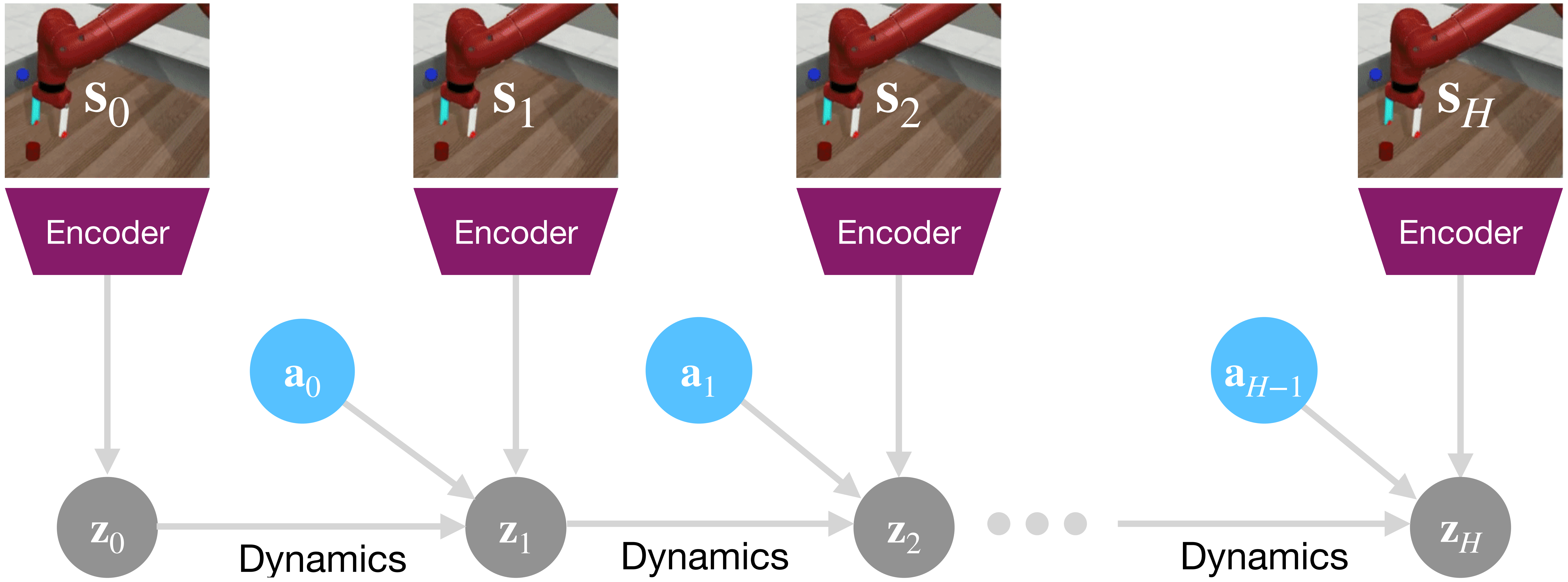
World Models



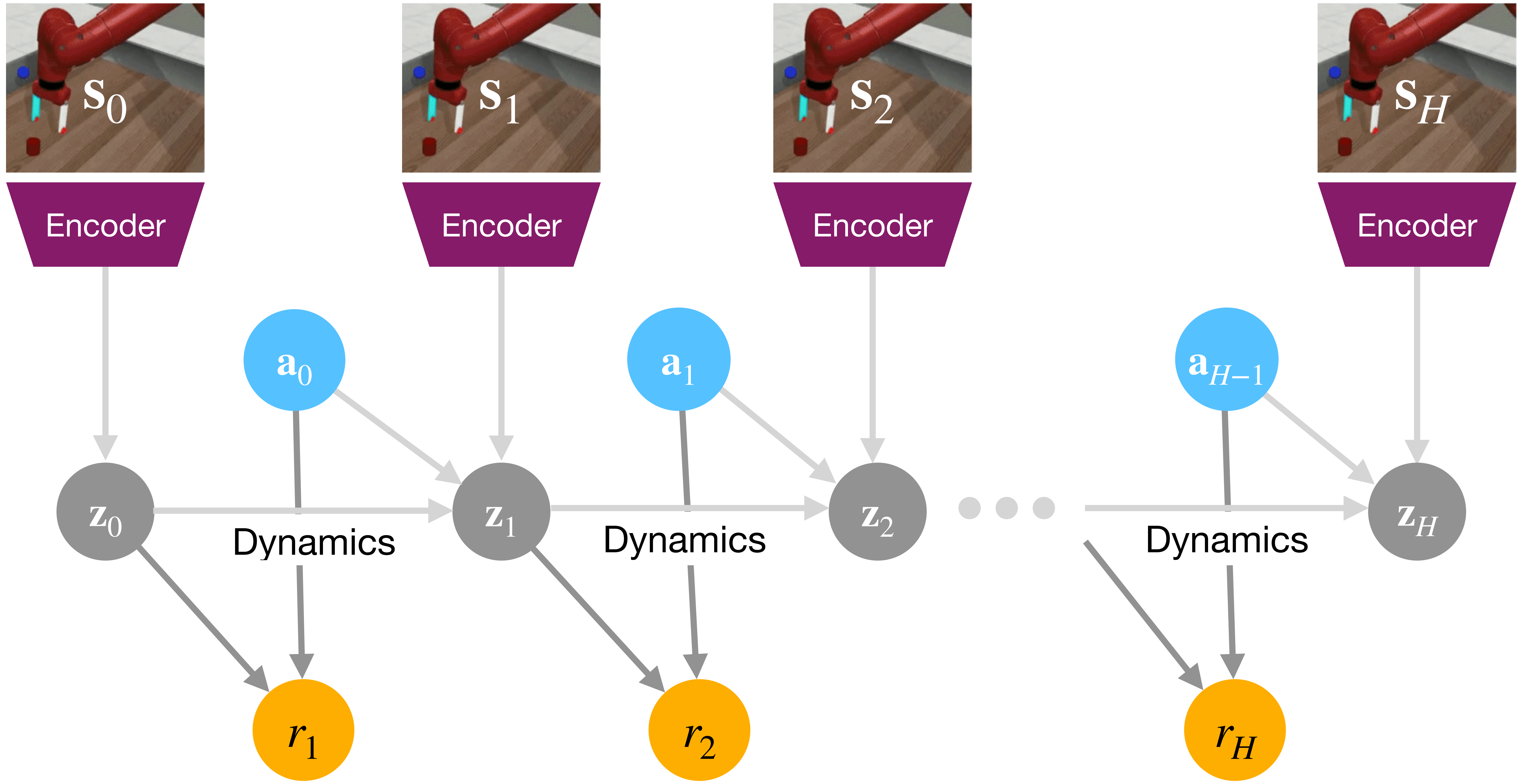
World Models



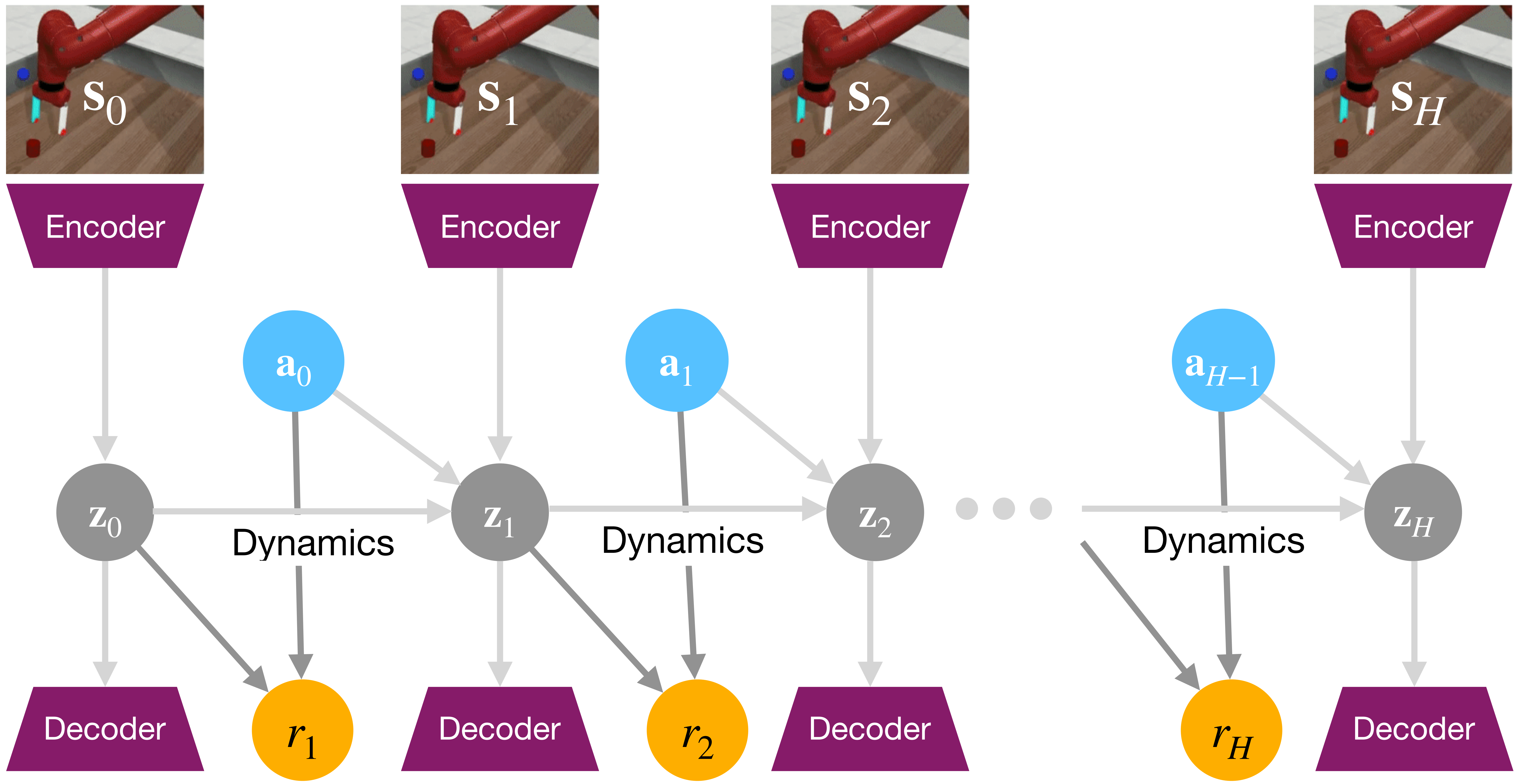
World Models



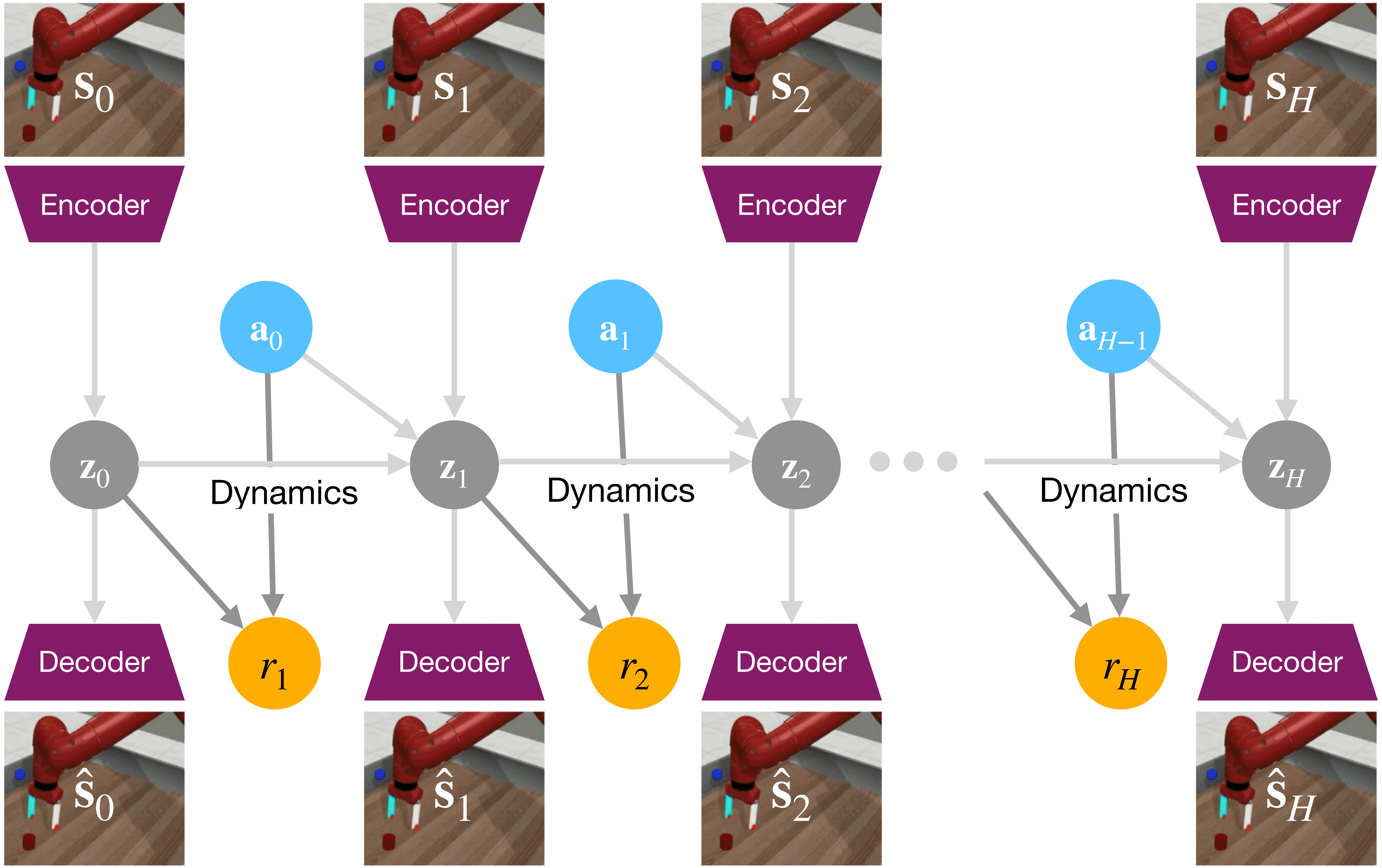
World Models



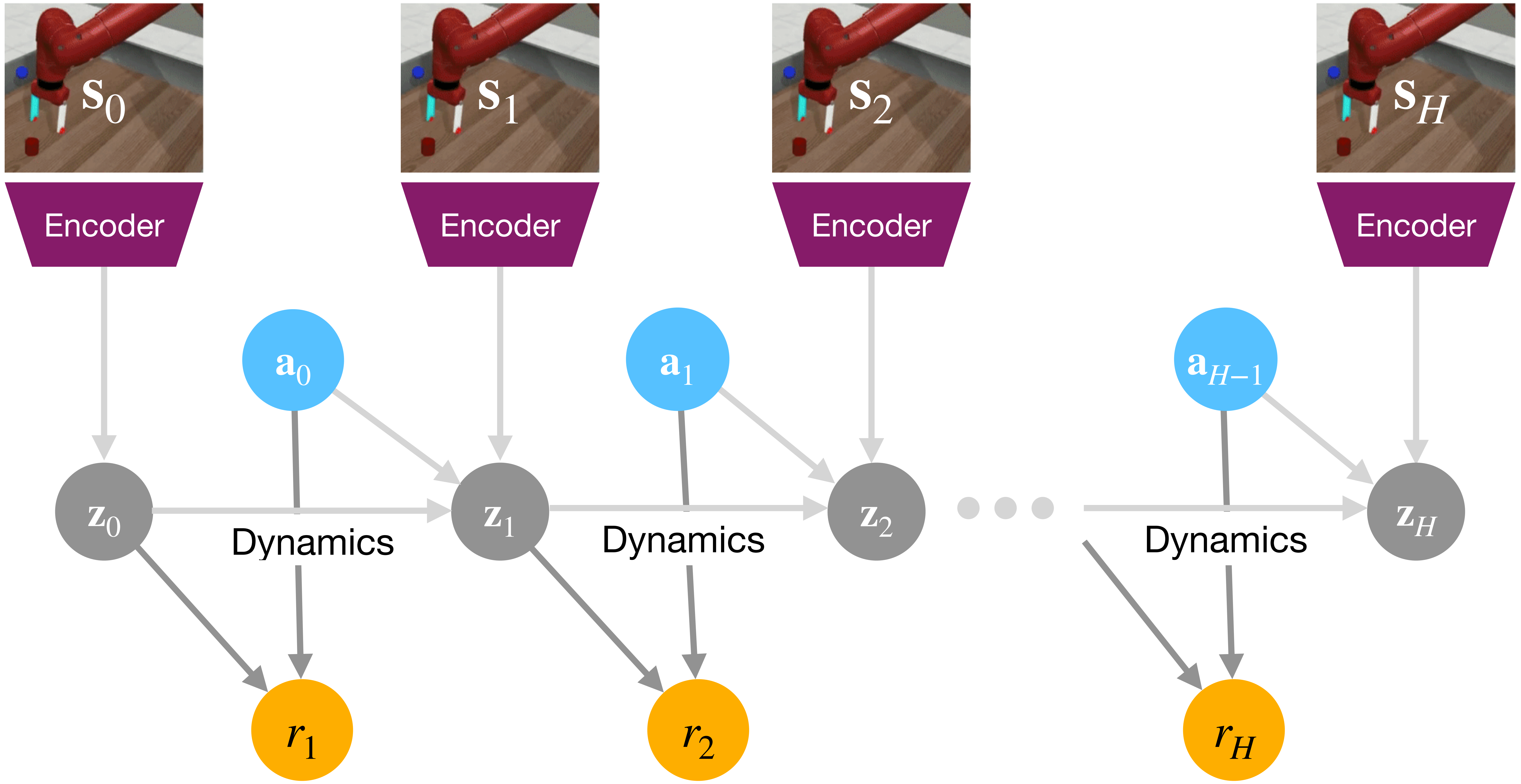
World Models



World Models



World Models



Latent Space Design Choices

Latent Space Design Choices

	Discrete Latent States?	Discrete Encoding Type	Stochastic Dynamics?	Reconstruction?
DreamerV3	✅	One-hot	✅	✅
TD-MPC2	❌	N/A	❌	❌

Danijar Hafner, et al. Mastering diverse domains through world models. arXiv preprint arXiv:2301.04104, 2023.

Nicklas Hansen, et al. TD-MPC2: Scalable, Robust World Models for Continuous Control. In The Twelfth International Conference on Learning Representations, October 2023.

Latent Space Design Choices

1. Do discrete latent spaces offer benefits over continuous ones?

Latent Space Design Choices

1. Do discrete latent spaces offer benefits over continuous ones?
2. How does the choice of discrete encoding (e.g., one-hot, label, or codebook encodings) affect performance?

Latent Space Design Choices

1. Do discrete latent spaces offer benefits over continuous ones?
2. How does the choice of discrete encoding (e.g., one-hot, label, or codebook encodings) affect performance?
3. Are there advantages to modelling the latent dynamics stochastically rather than deterministically?

Latent Space Design Choices

	Discrete Latent States?	Discrete Encoding Type	Stochastic Dynamics?	Reconstruction?
DreamerV3	✅	One-hot	✅	✅
TD-MPC2	❌	N/A	❌	❌

Danijar Hafner, et al. Mastering diverse domains through world models. arXiv preprint arXiv:2301.04104, 2023.

Nicklas Hansen, et al. TD-MPC2: Scalable, Robust World Models for Continuous Control. In The Twelfth International Conference on Learning Representations, October 2023.

Latent Space Design Choices

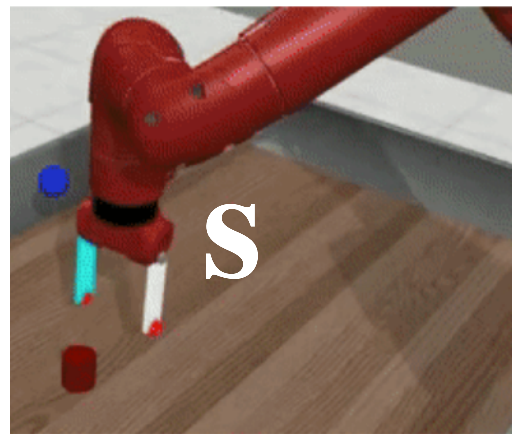
	Discrete Latent States?	Discrete Encoding Type	Stochastic Dynamics?	Reconstruction?
DreamerV3	✅	One-hot	✅	✅
TD-MPC2	❌	N/A	❌	❌
DC-MPC (ours)	✅	Codebook	✅	❌

Danijar Hafner, et al. Mastering diverse domains through world models. arXiv preprint arXiv:2301.04104, 2023.

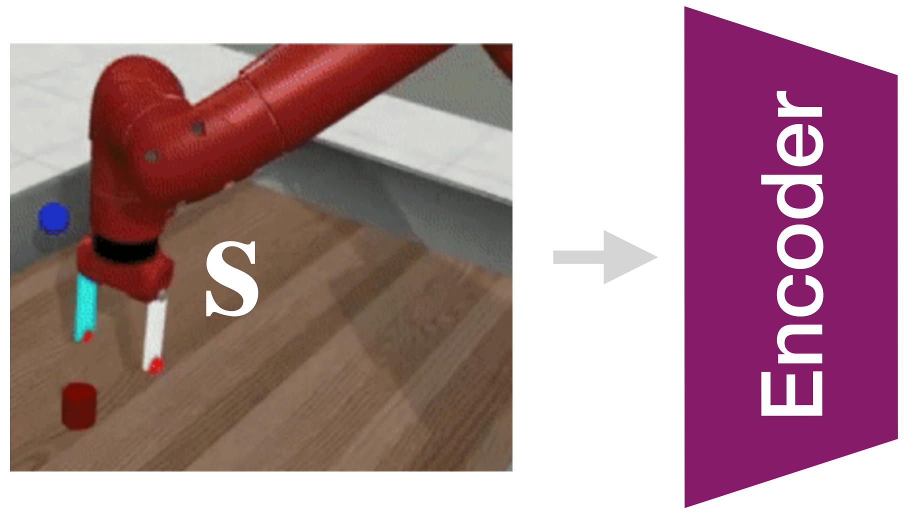
Nicklas Hansen, et al. TD-MPC2: Scalable, Robust World Models for Continuous Control. In The Twelfth International Conference on Learning Representations, October 2023.

DC-MPC: Discrete Codebook

DC-MPC: Discrete Codebook

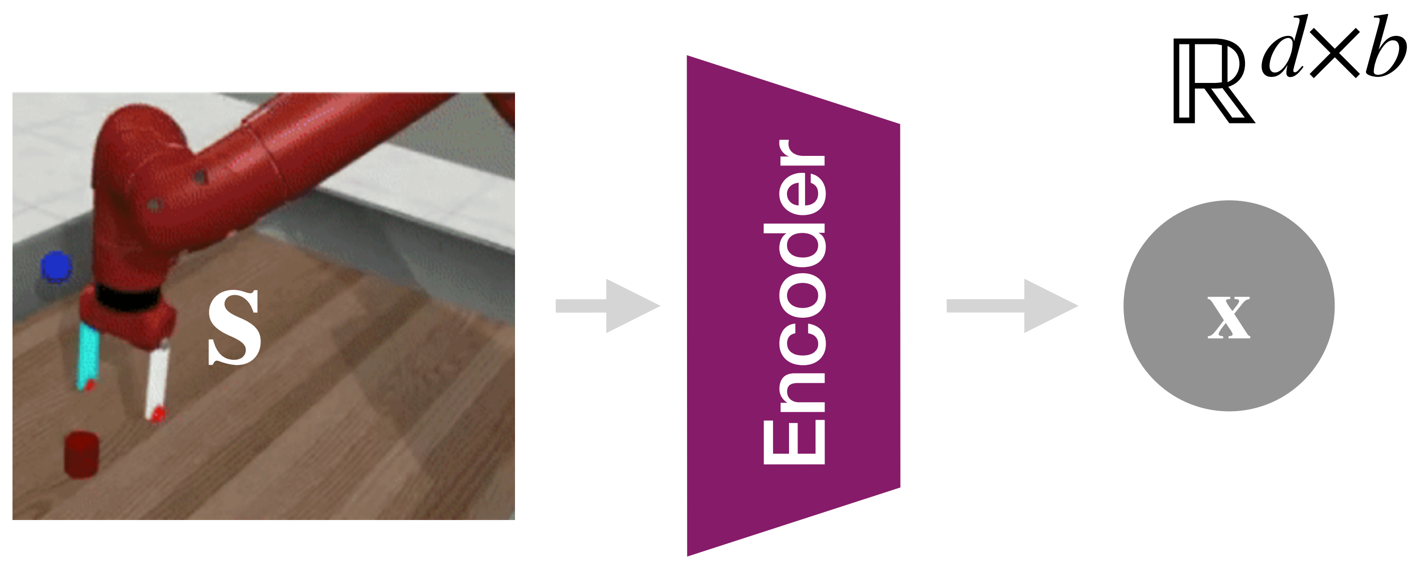


DC-MPC: Discrete Codebook

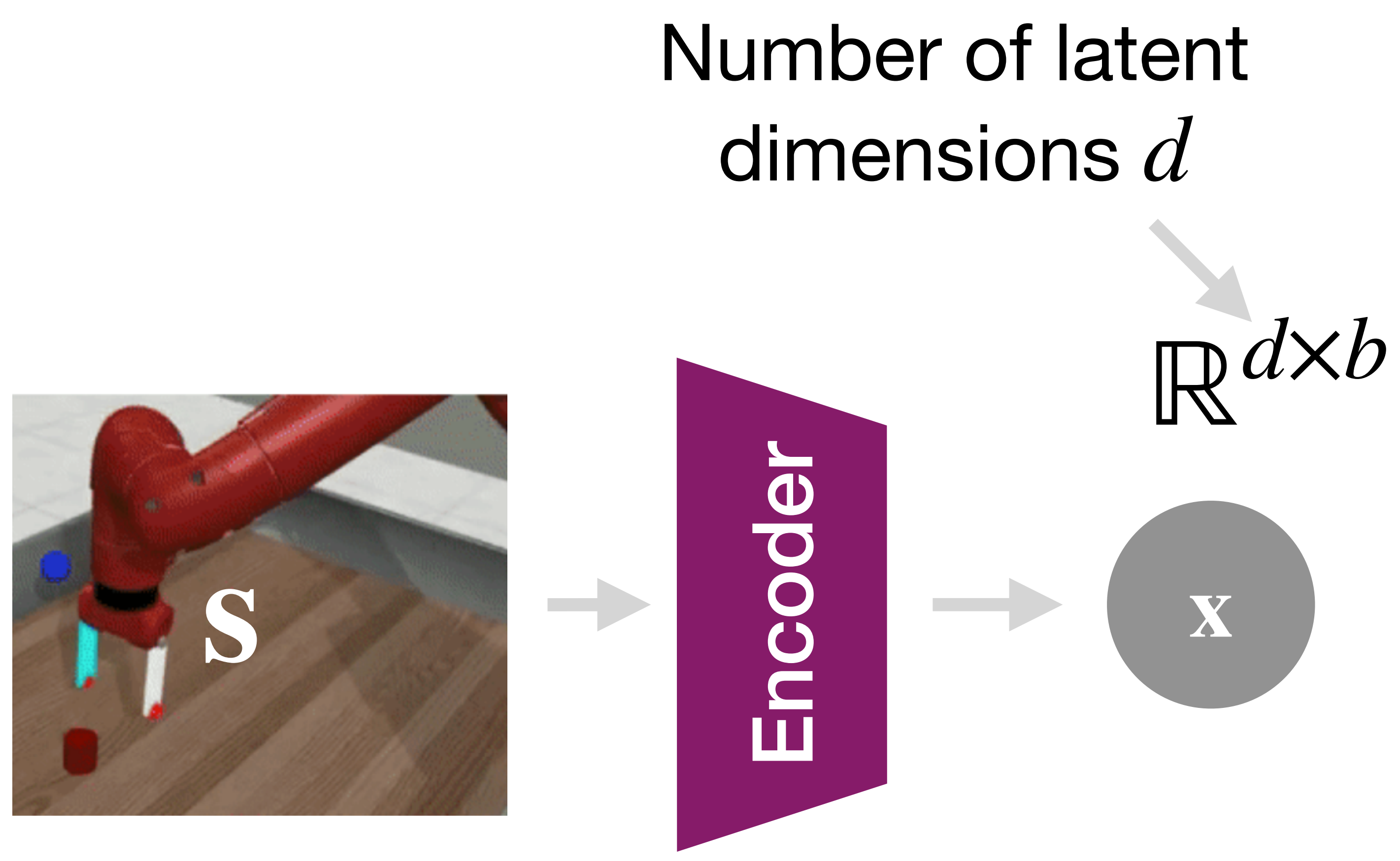


DC-MPC: Discrete Codebook

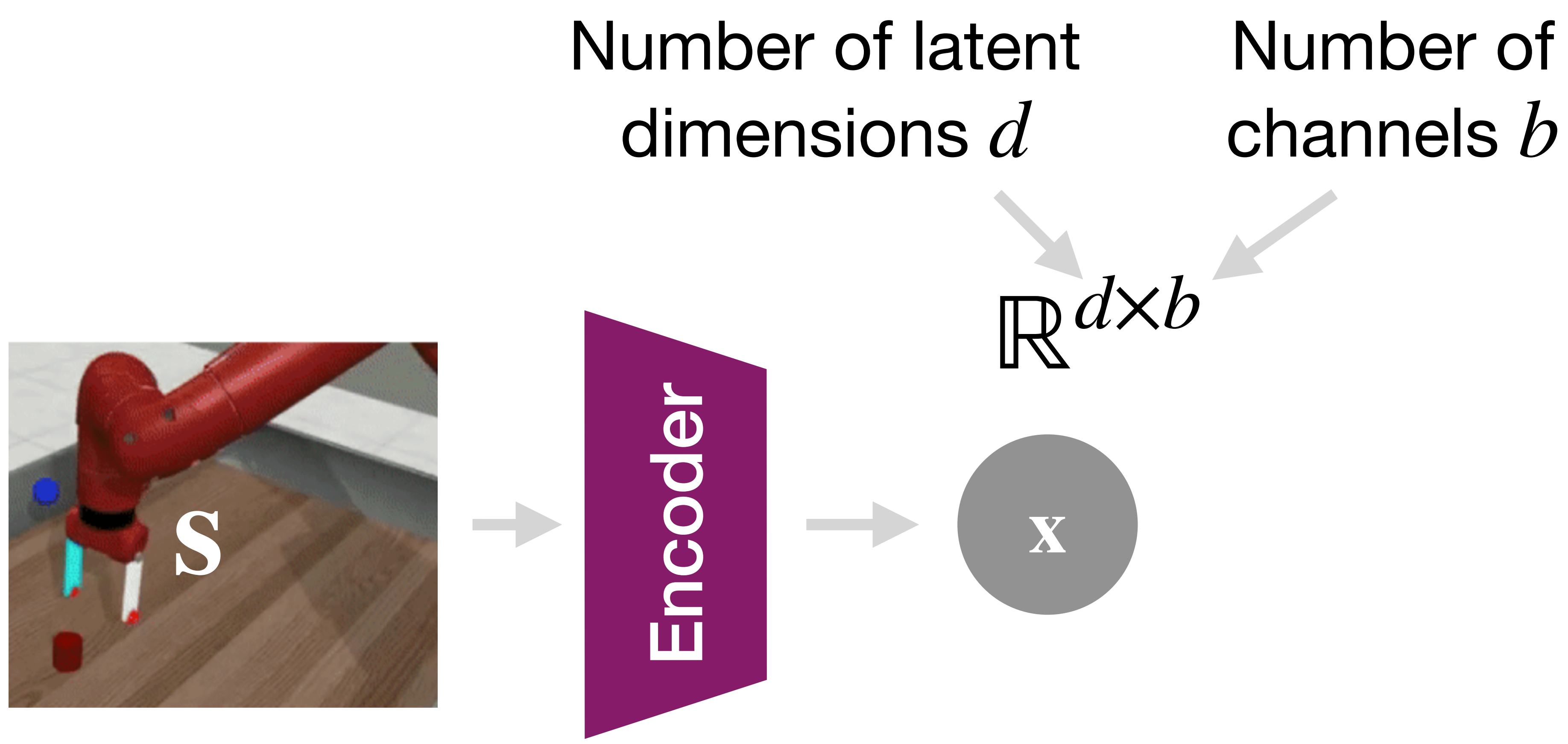
DC-MPC: Discrete Codebook



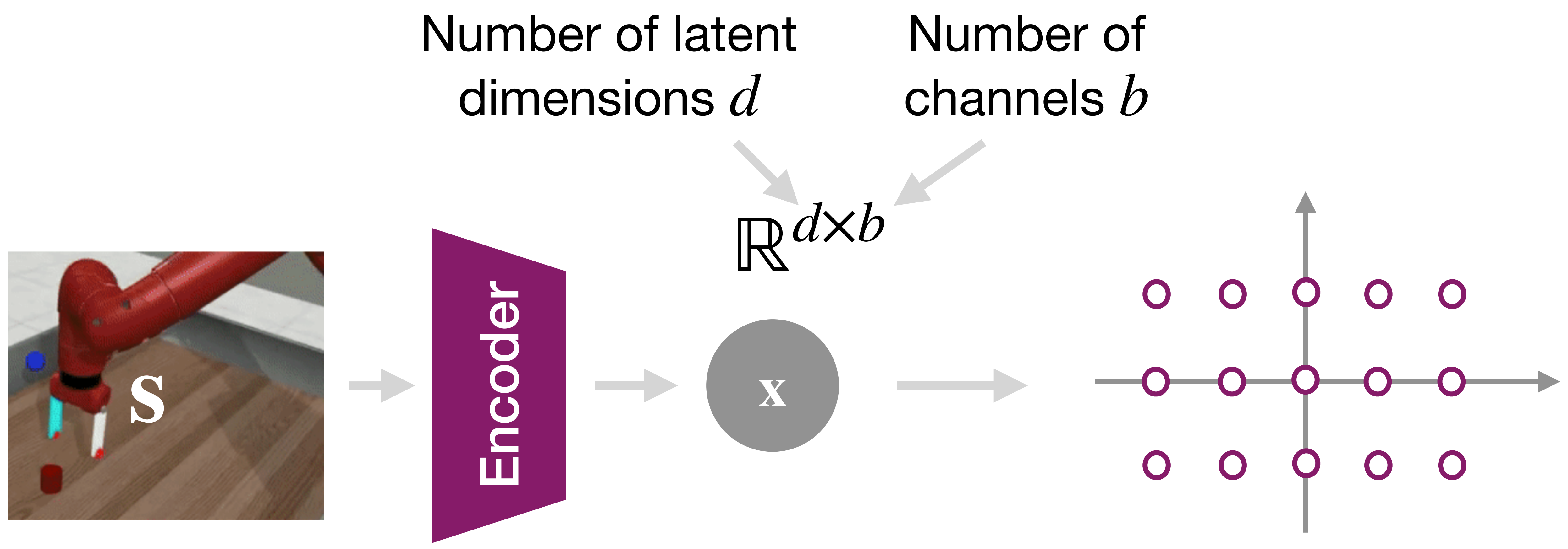
DC-MPC: Discrete Codebook



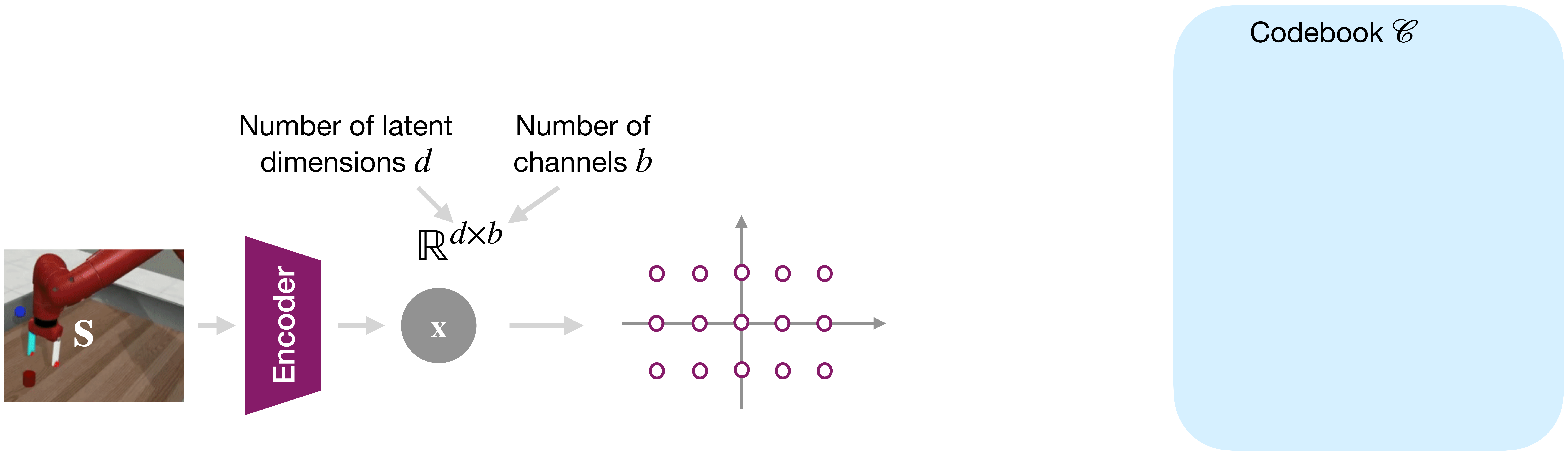
DC-MPC: Discrete Codebook



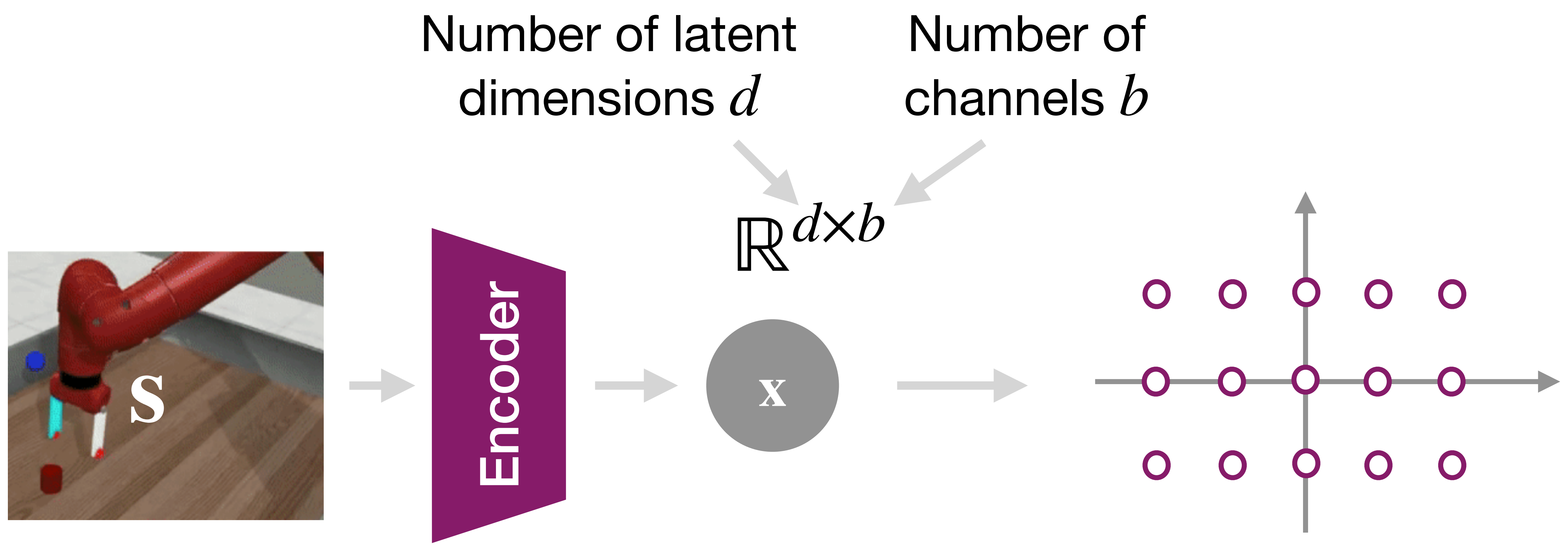
DC-MPC: Discrete Codebook

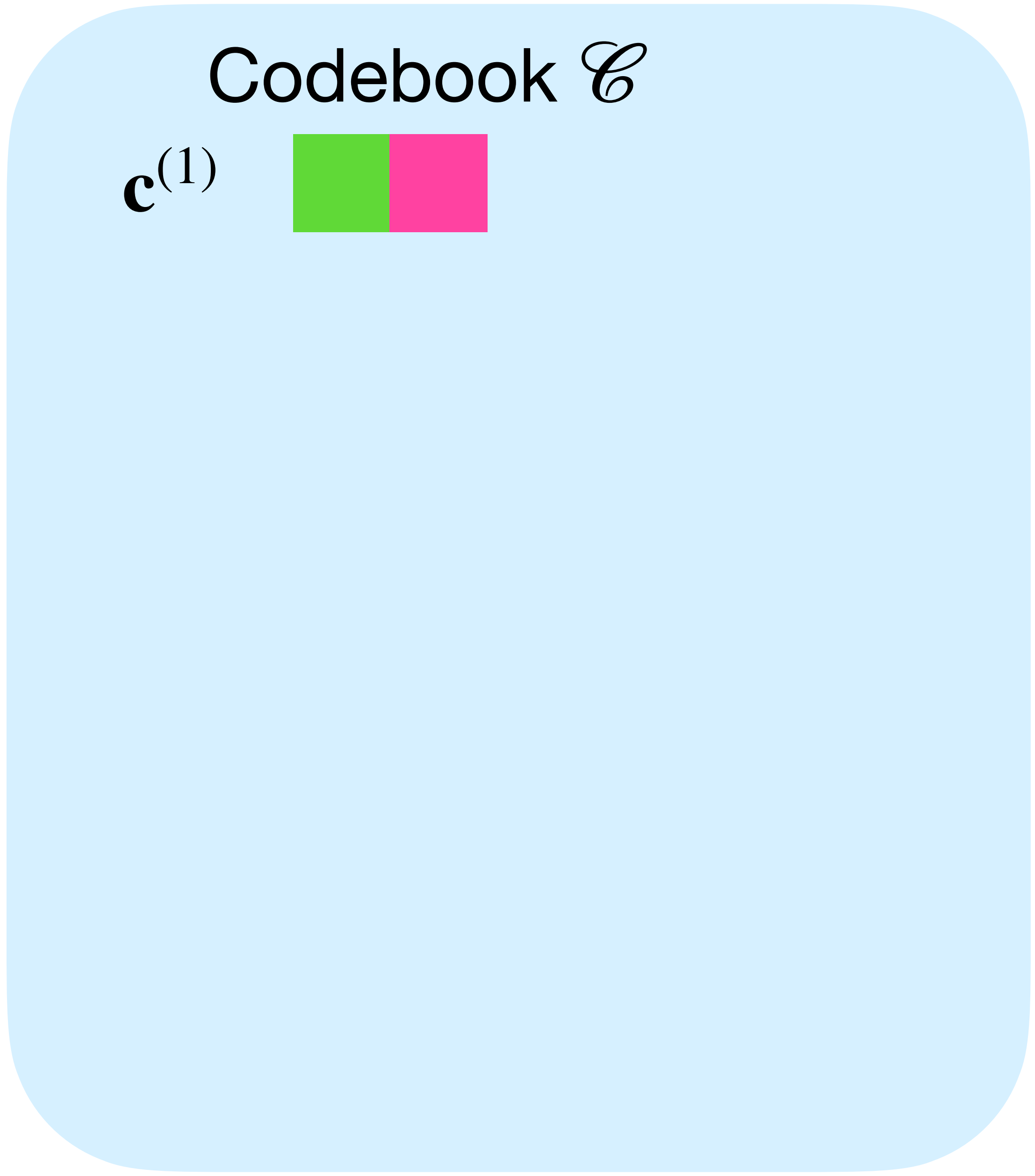


DC-MPC: Discrete Codebook

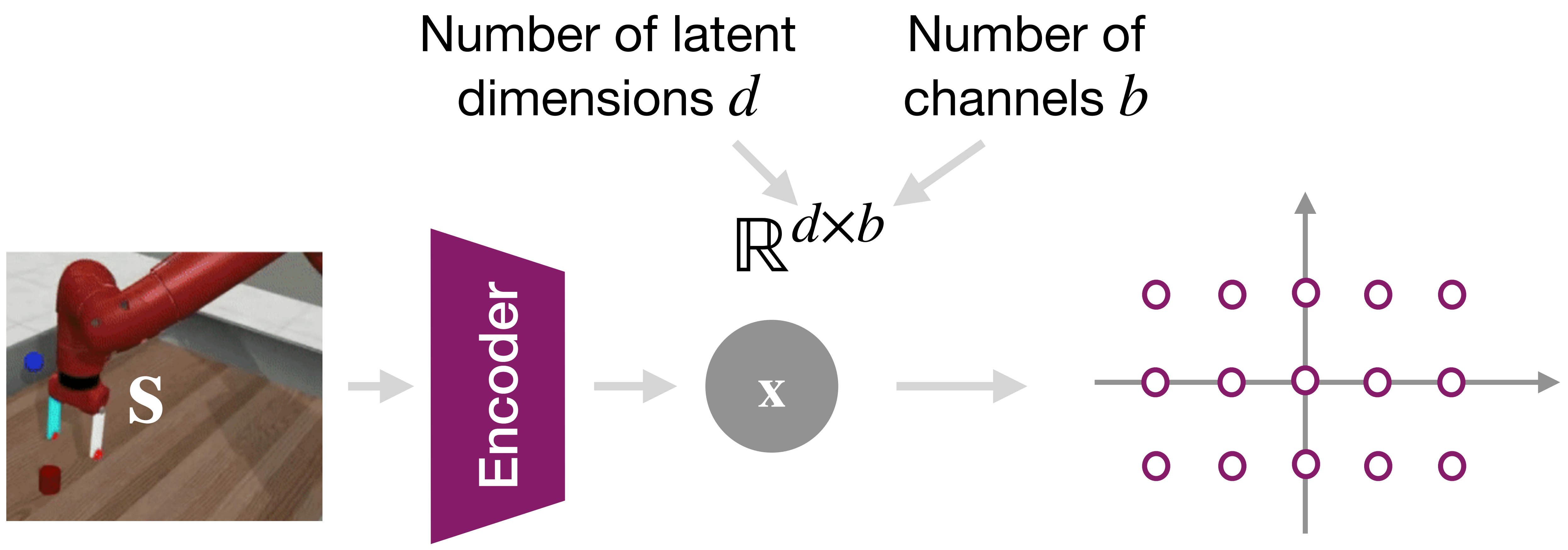


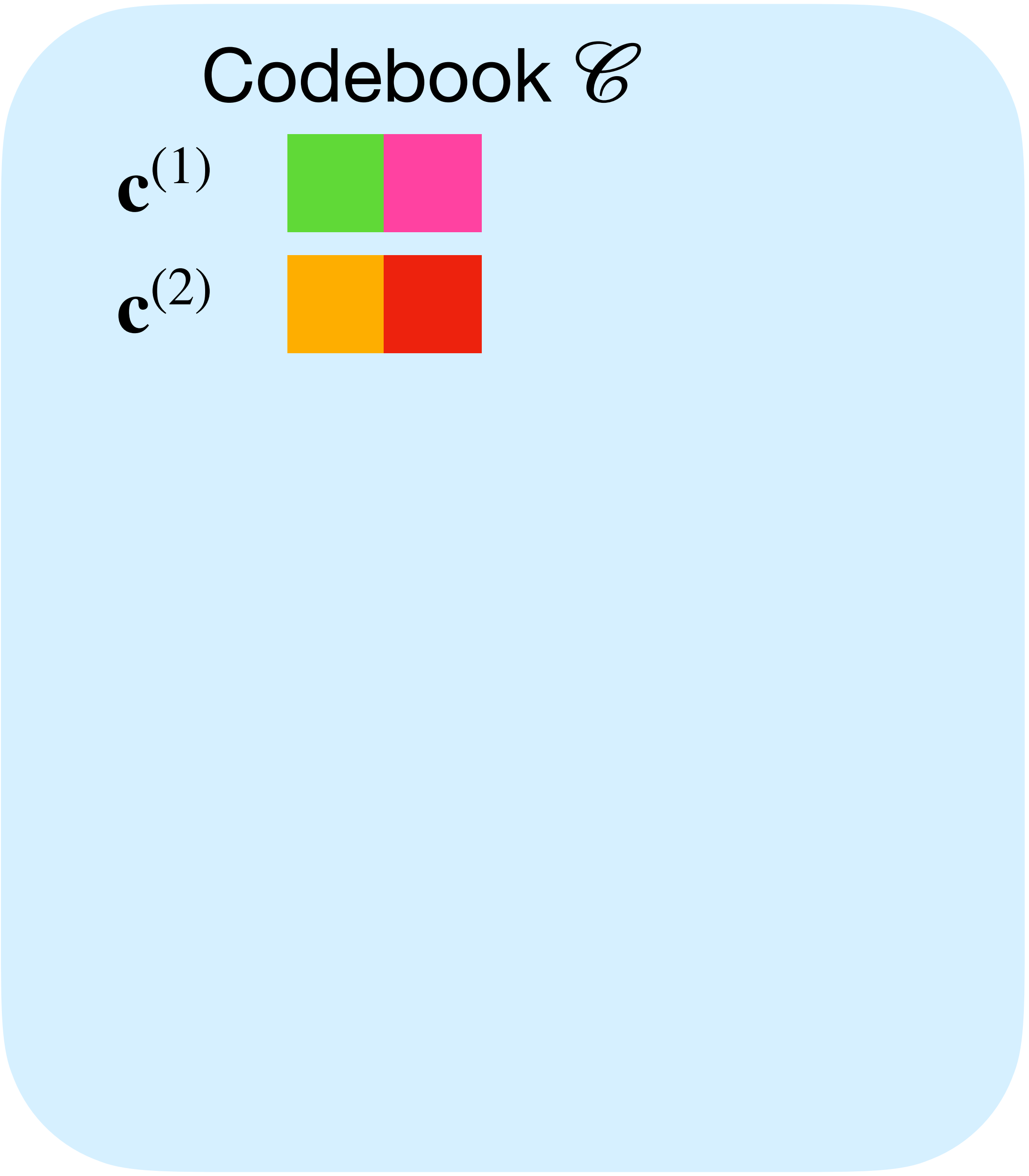
DC-MPC: Discrete Codebook



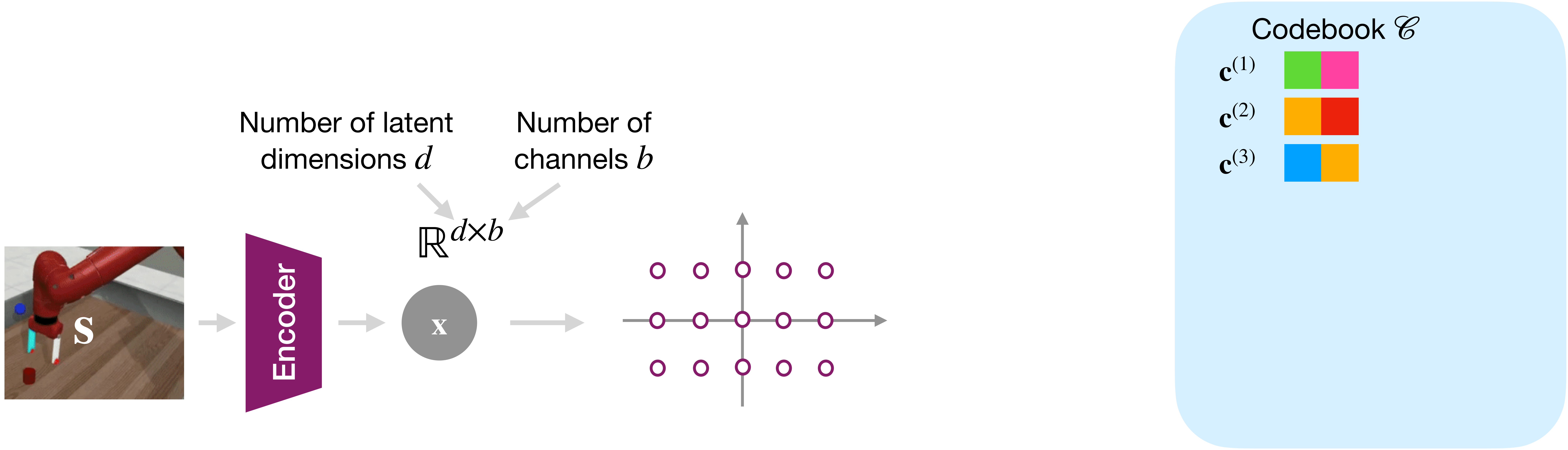


DC-MPC: Discrete Codebook

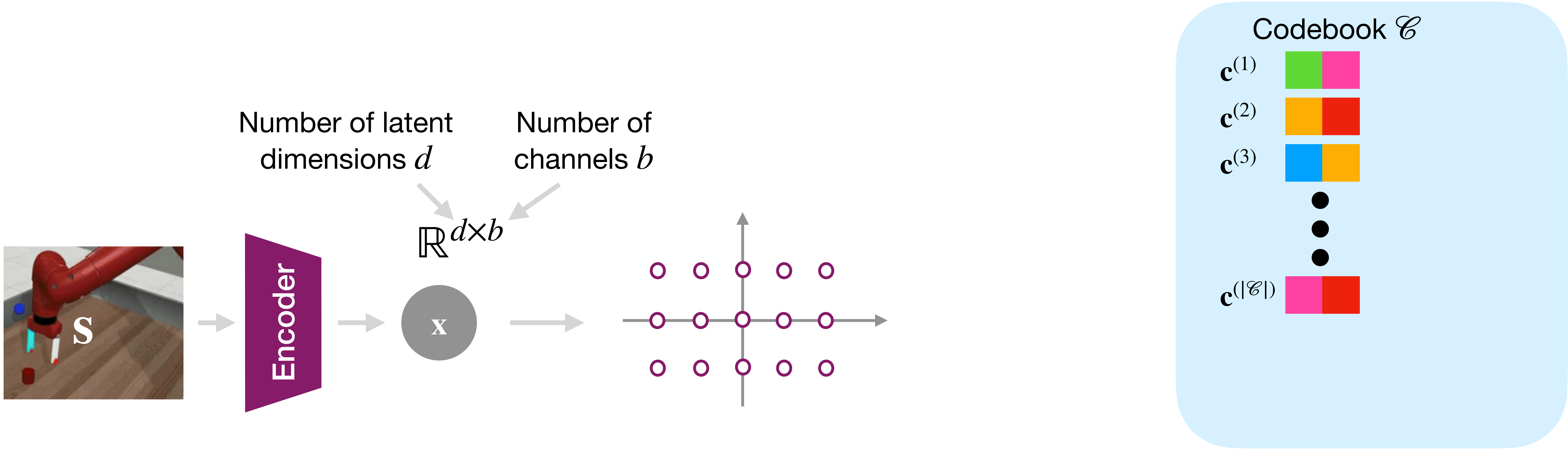




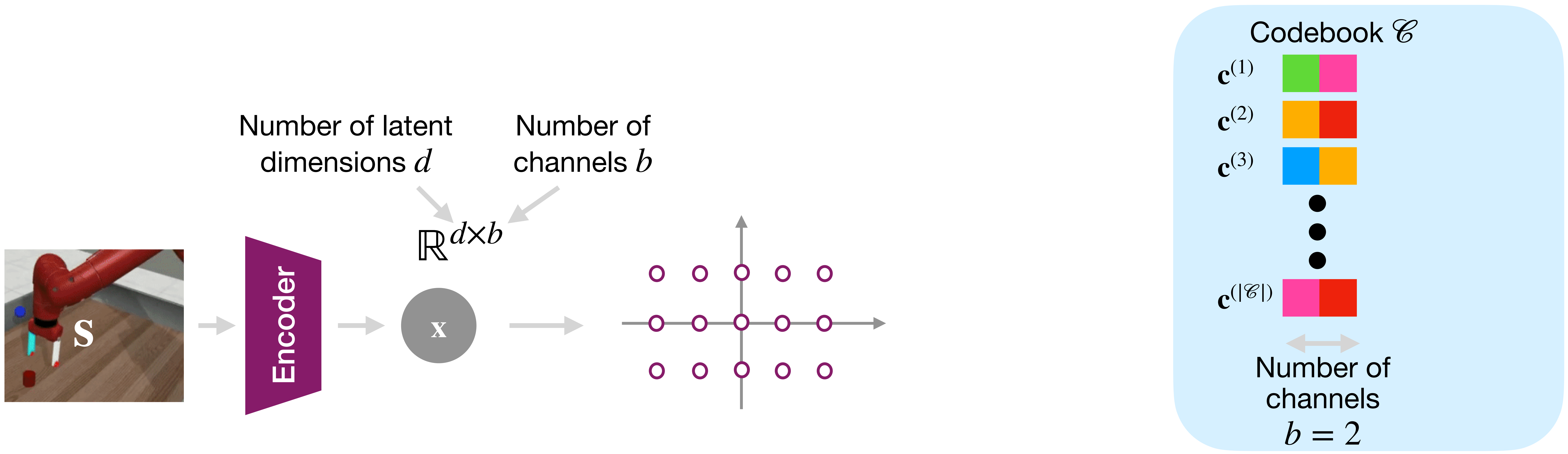
DC-MPC: Discrete Codebook



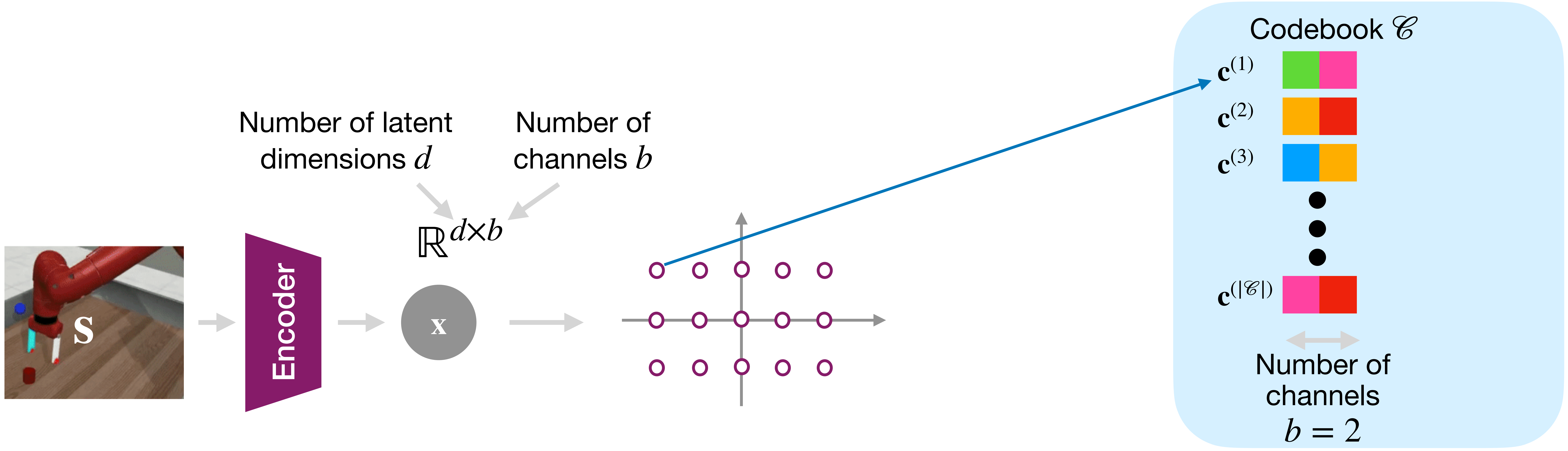
DC-MPC: Discrete Codebook



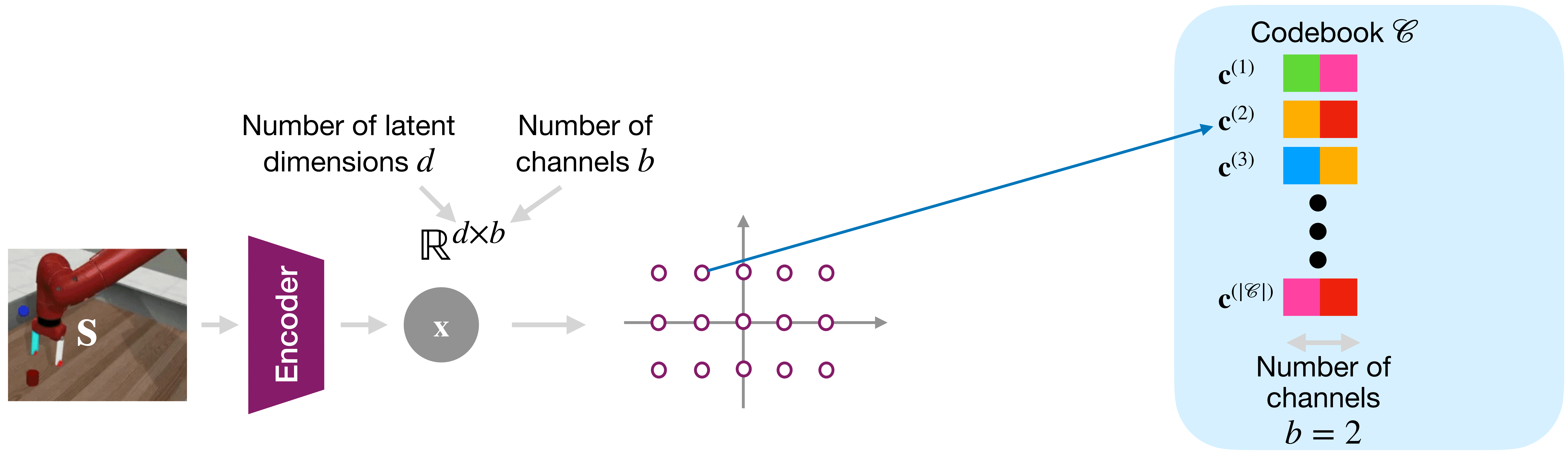
DC-MPC: Discrete Codebook



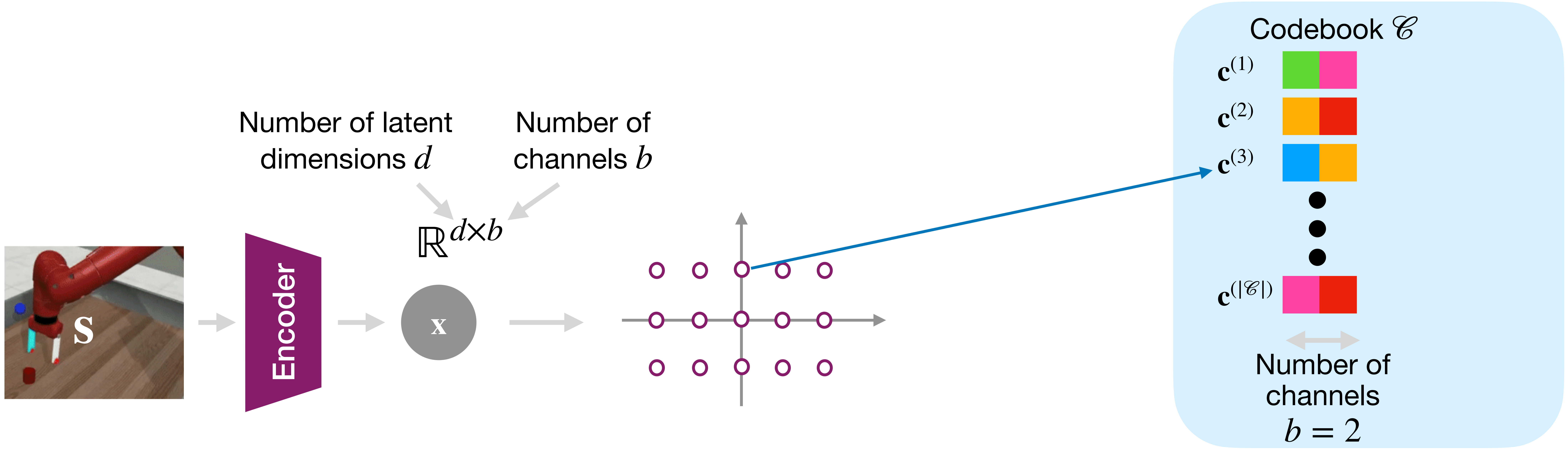
DC-MPC: Discrete Codebook



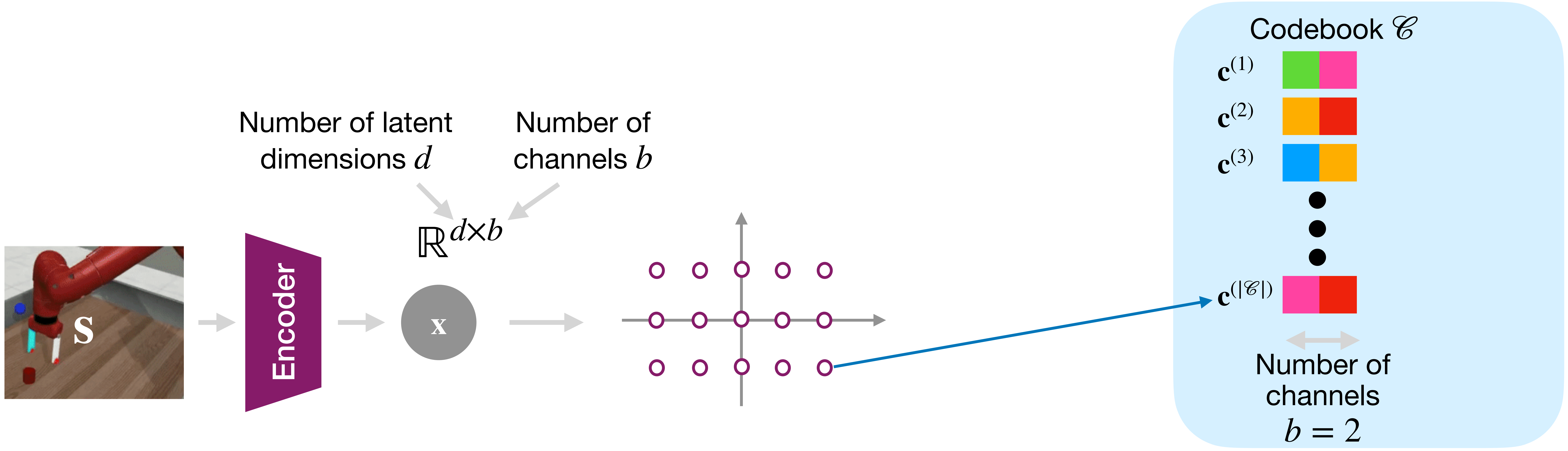
DC-MPC: Discrete Codebook



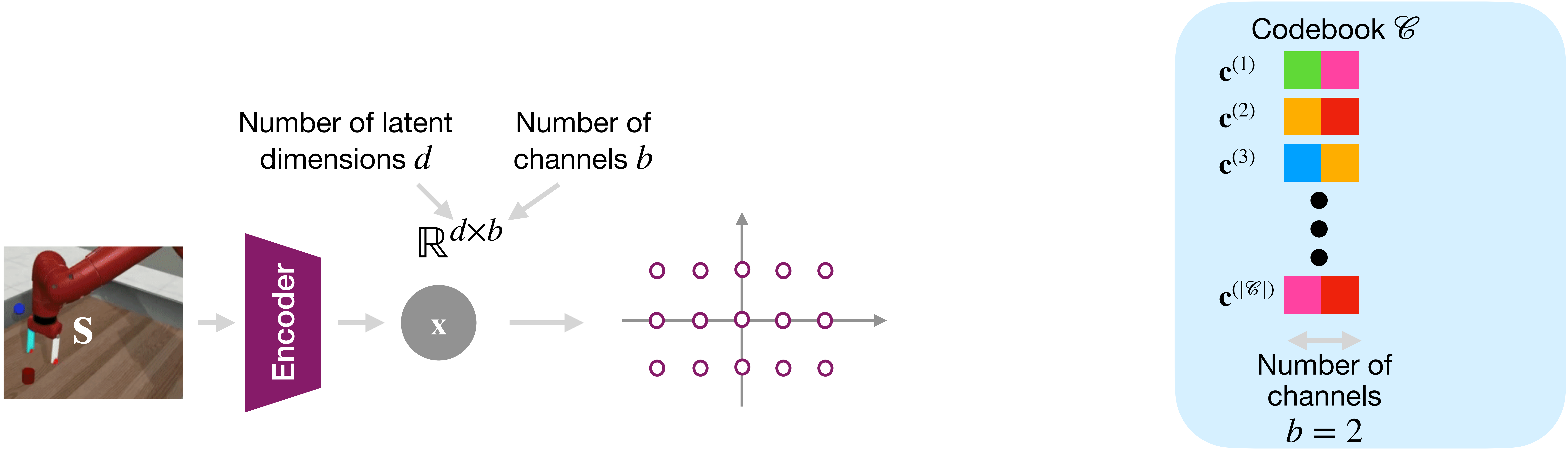
DC-MPC: Discrete Codebook



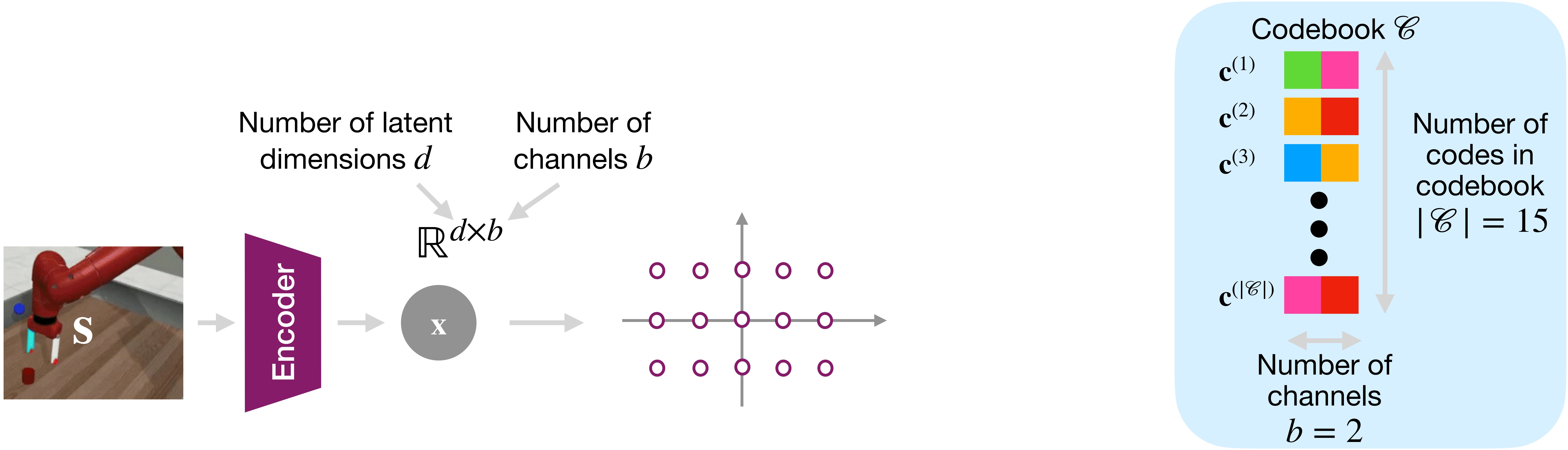
DC-MPC: Discrete Codebook



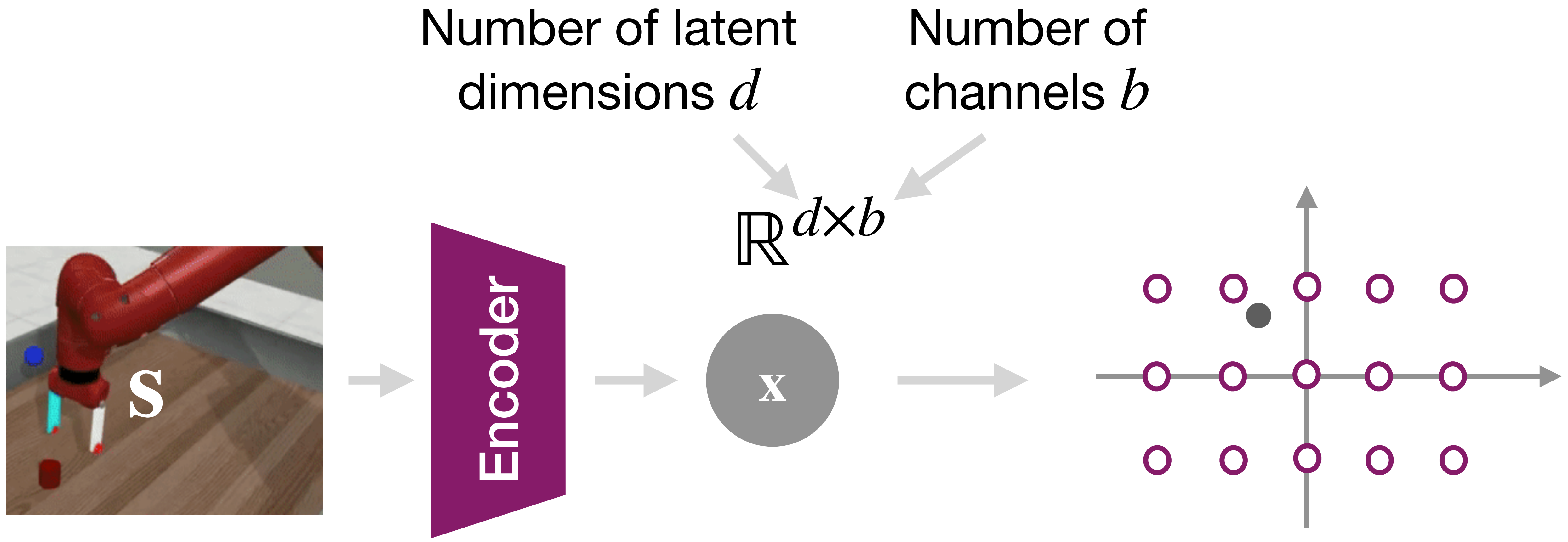
DC-MPC: Discrete Codebook

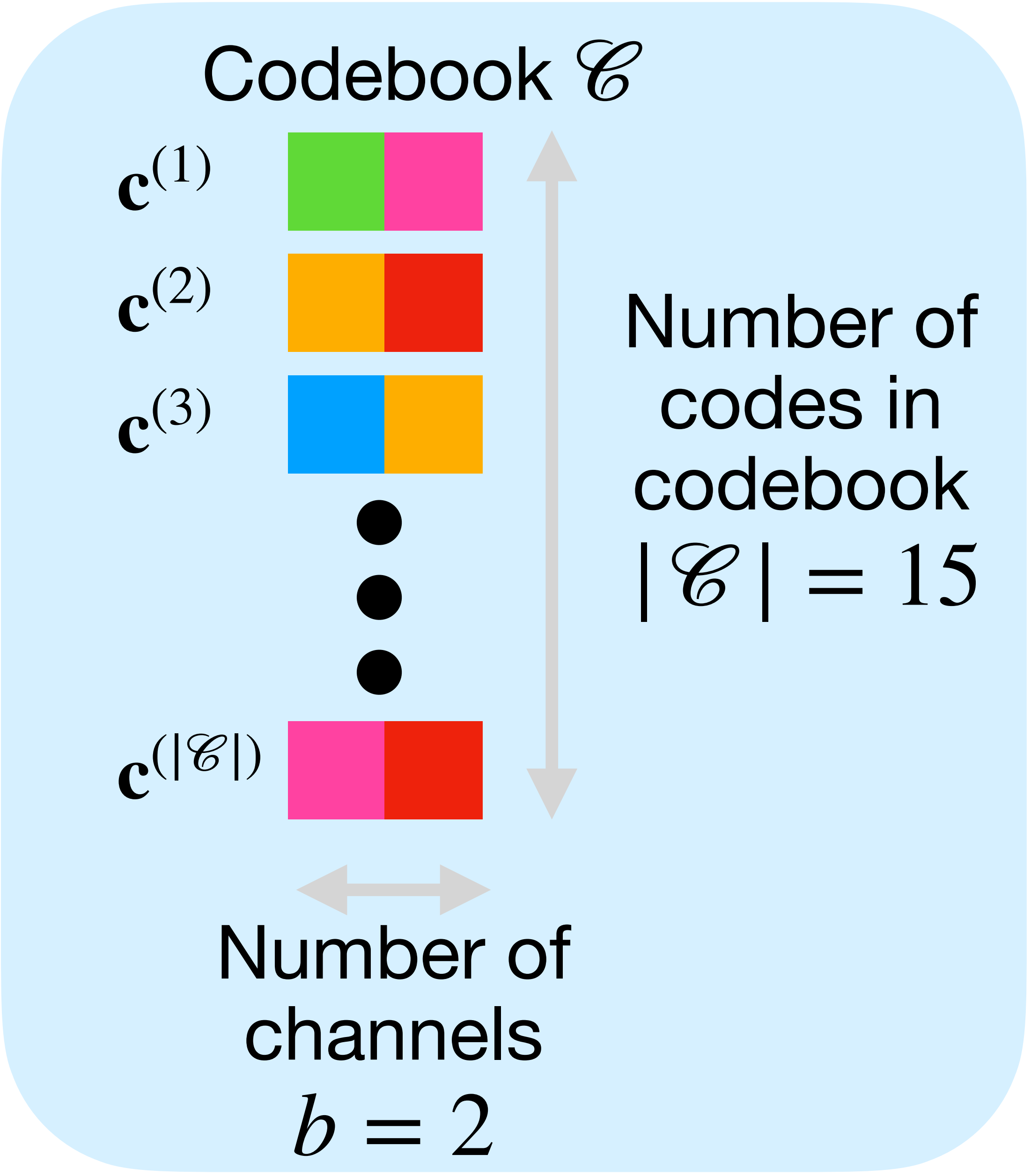


DC-MPC: Discrete Codebook

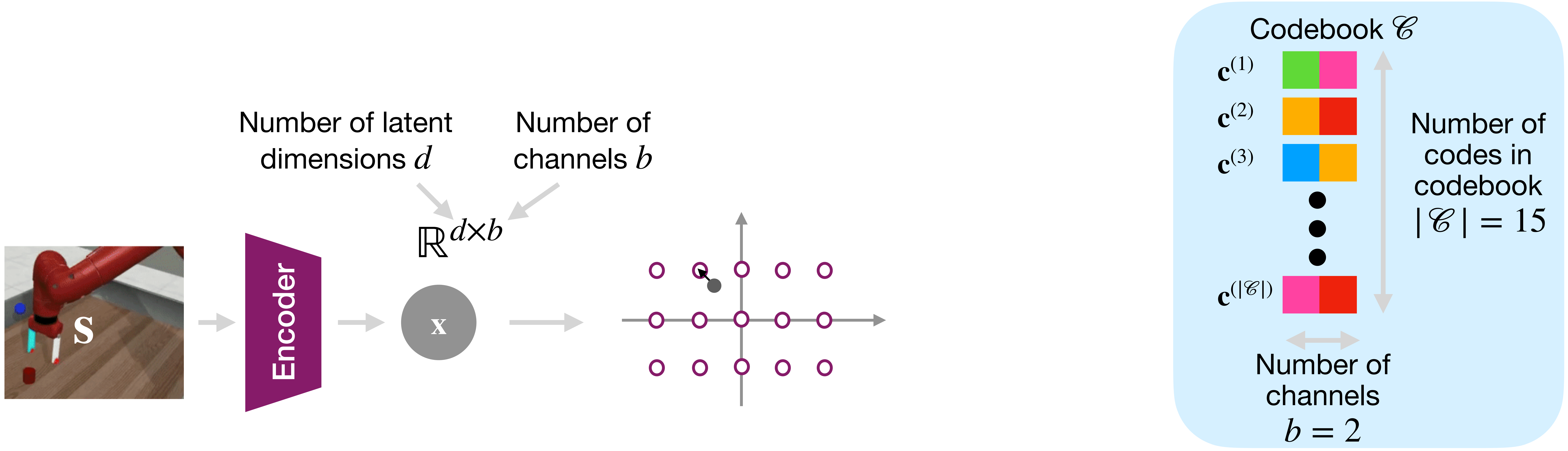


DC-MPC: Discrete Codebook

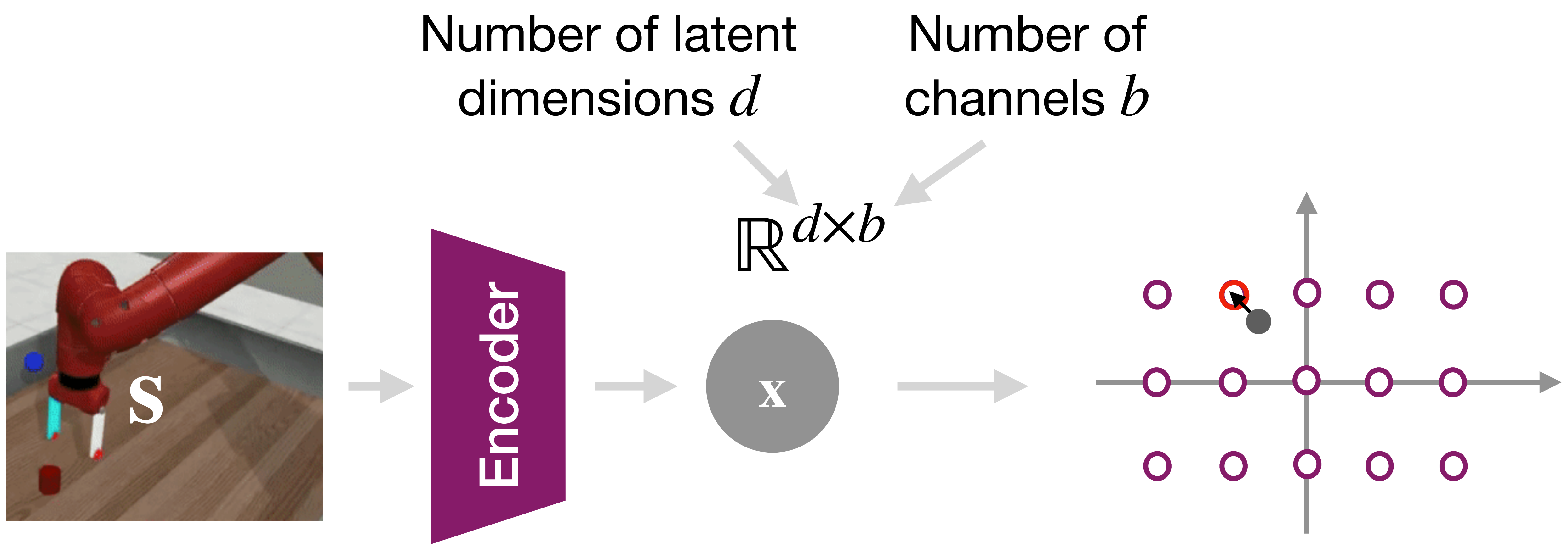


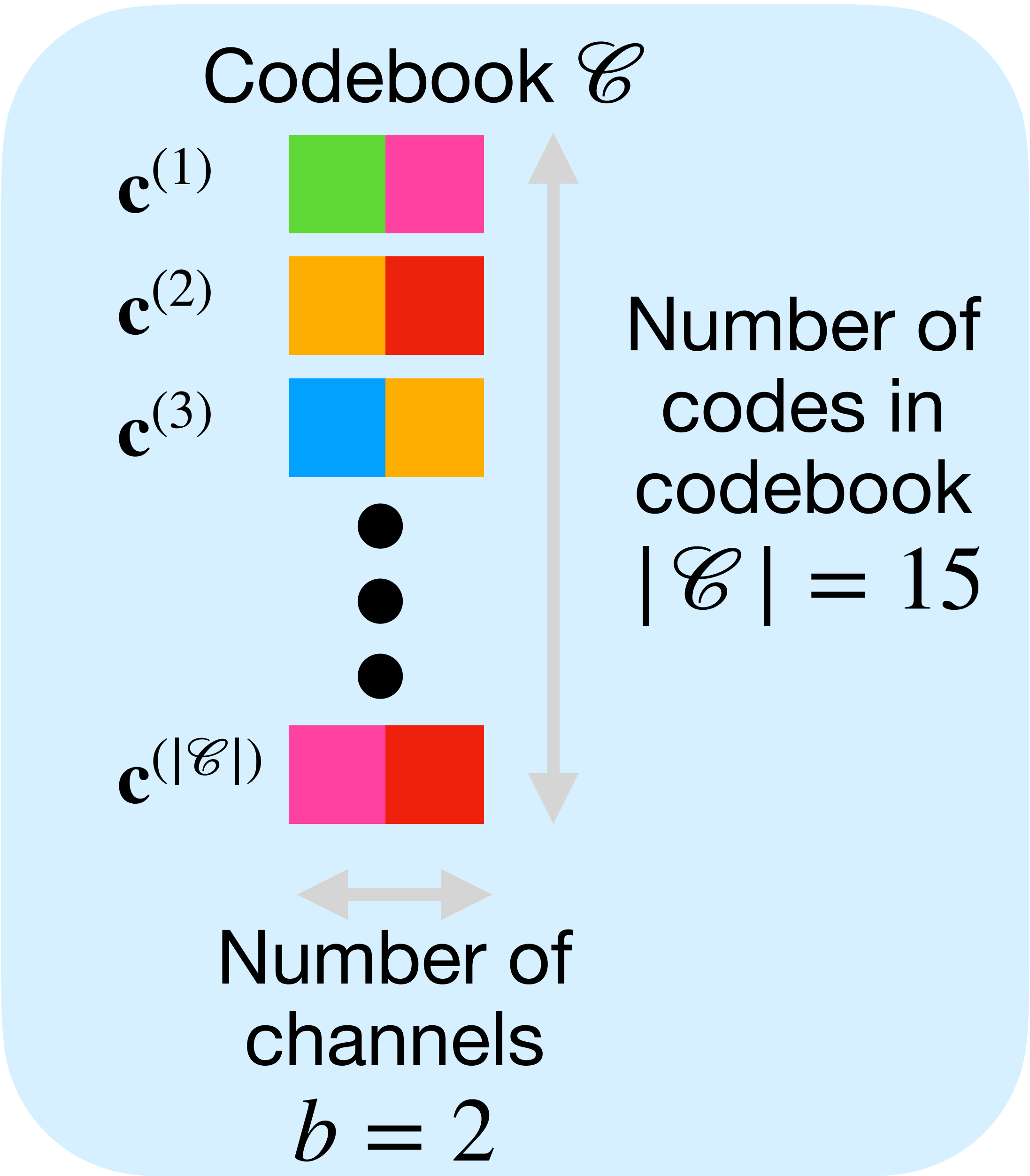


DC-MPC: Discrete Codebook

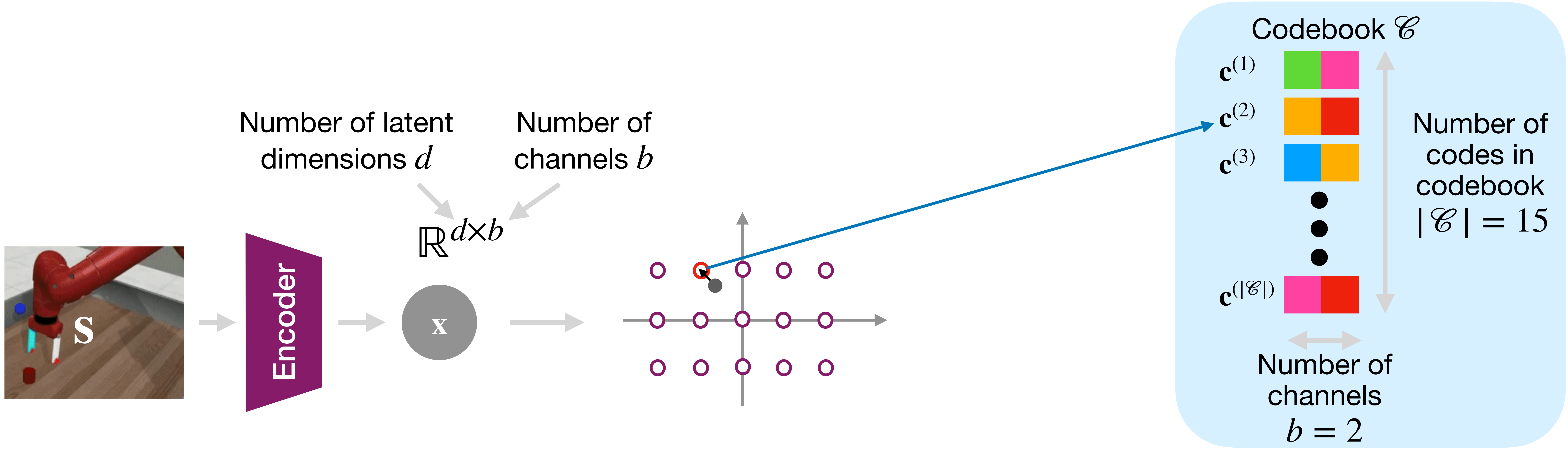


DC-MPC: Discrete Codebook

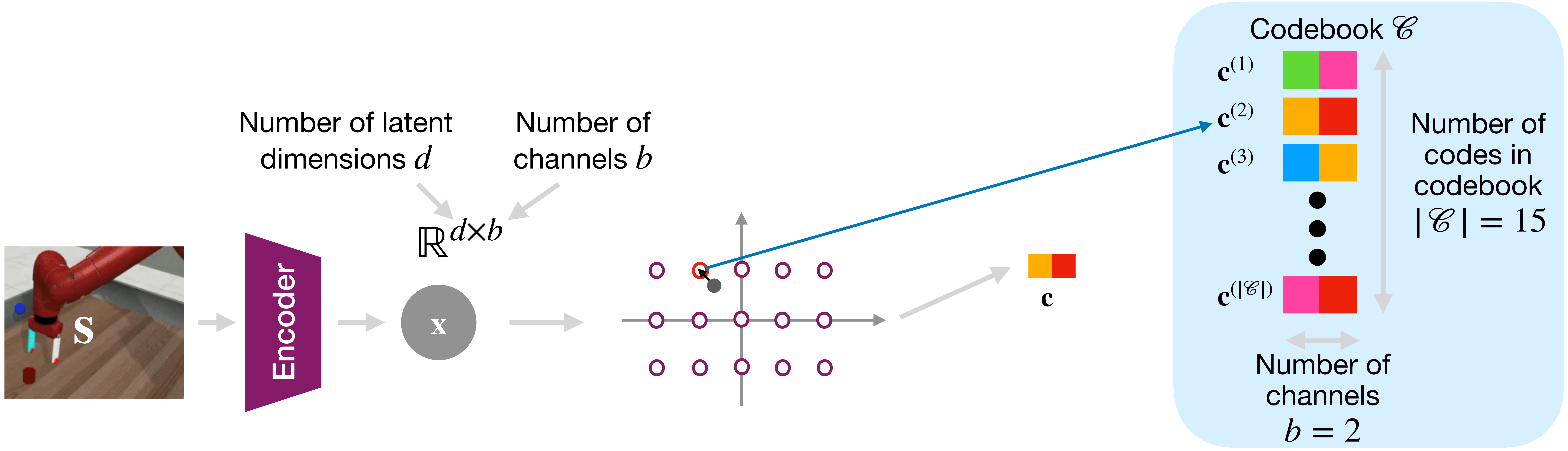




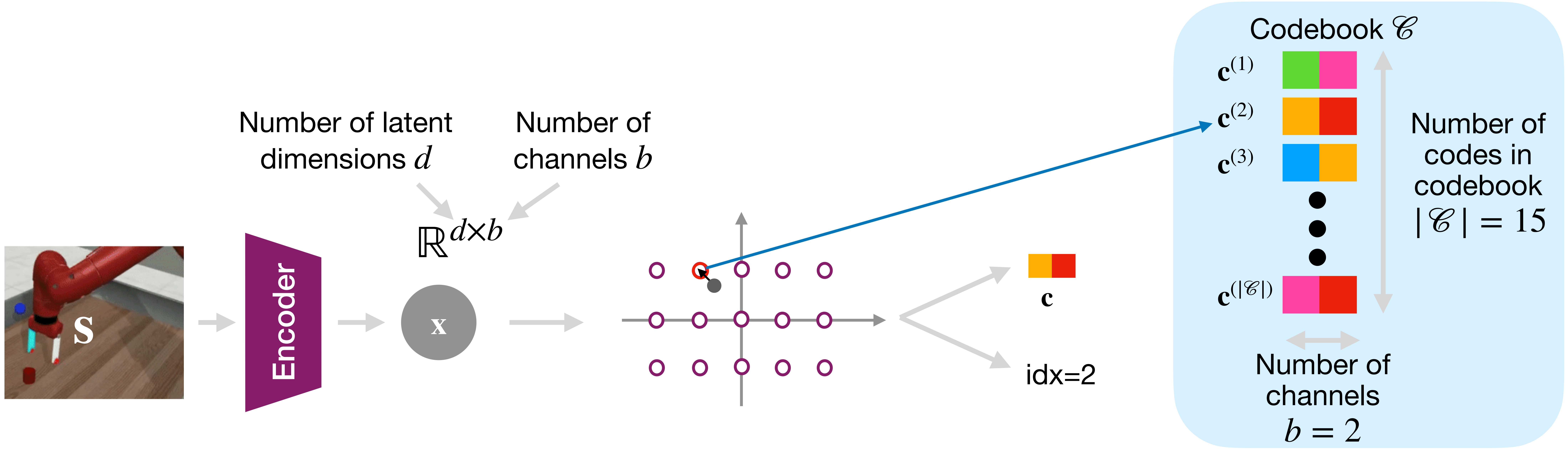
DC-MPC: Discrete Codebook



DC-MPC: Discrete Codebook

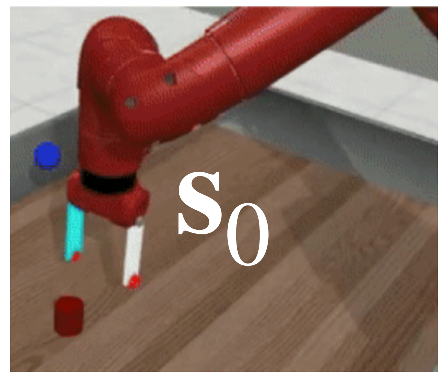


DC-MPC: Discrete Codebook

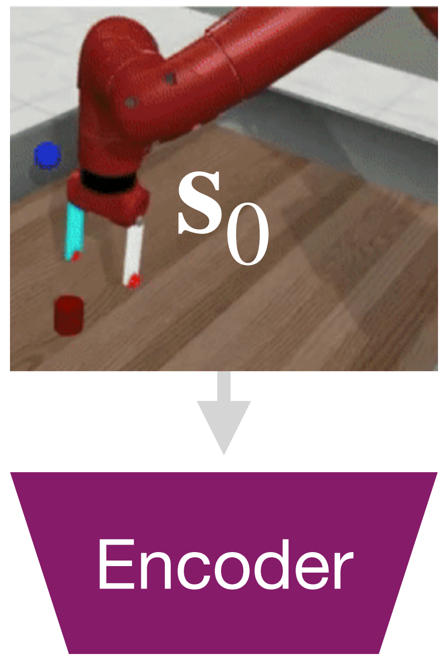


DC-MPC: World Model Training

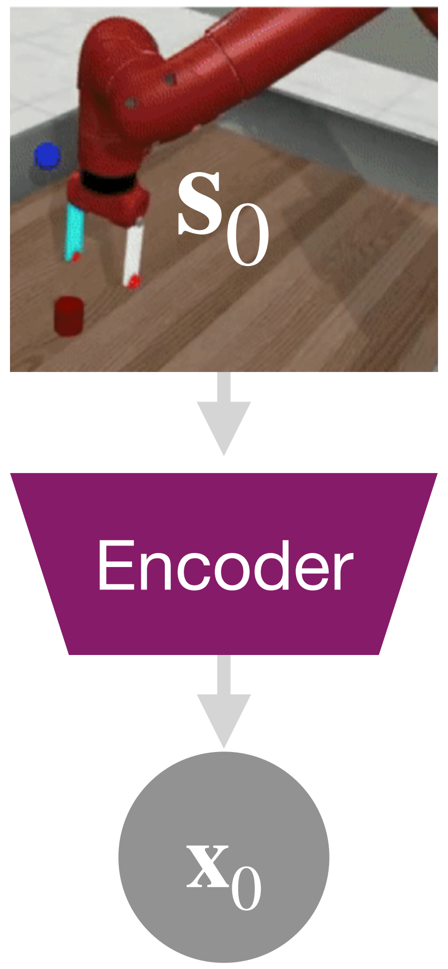
DC-MPC: World Model Training



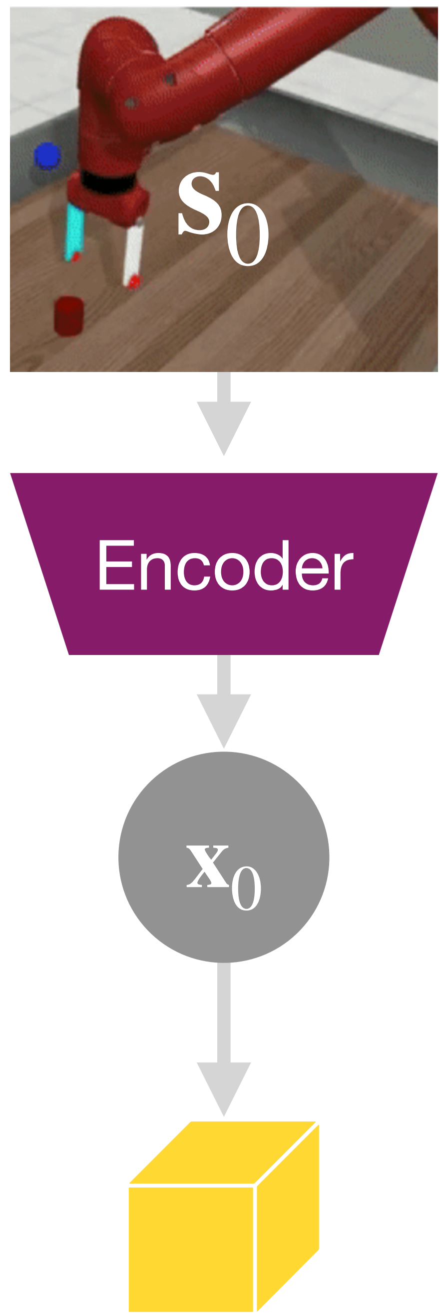
DC-MPC: World Model Training



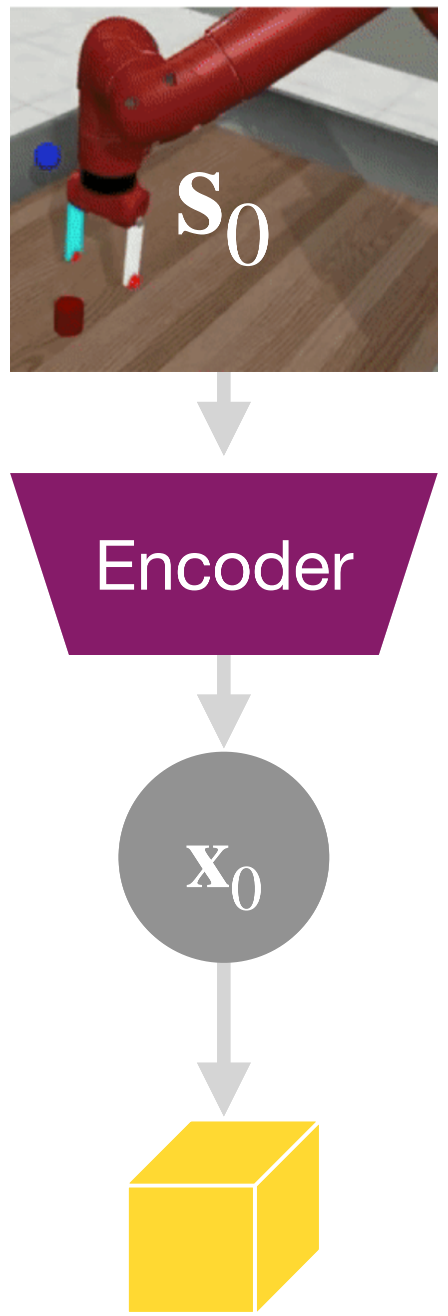
DC-MPC: World Model Training

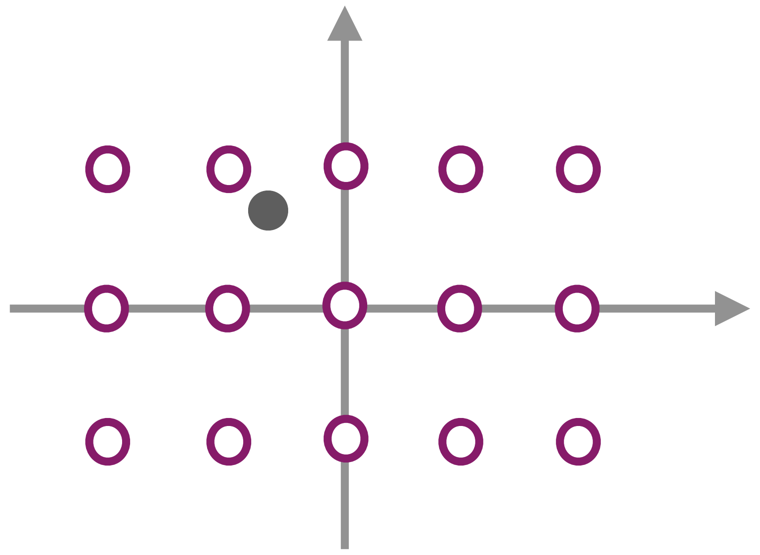


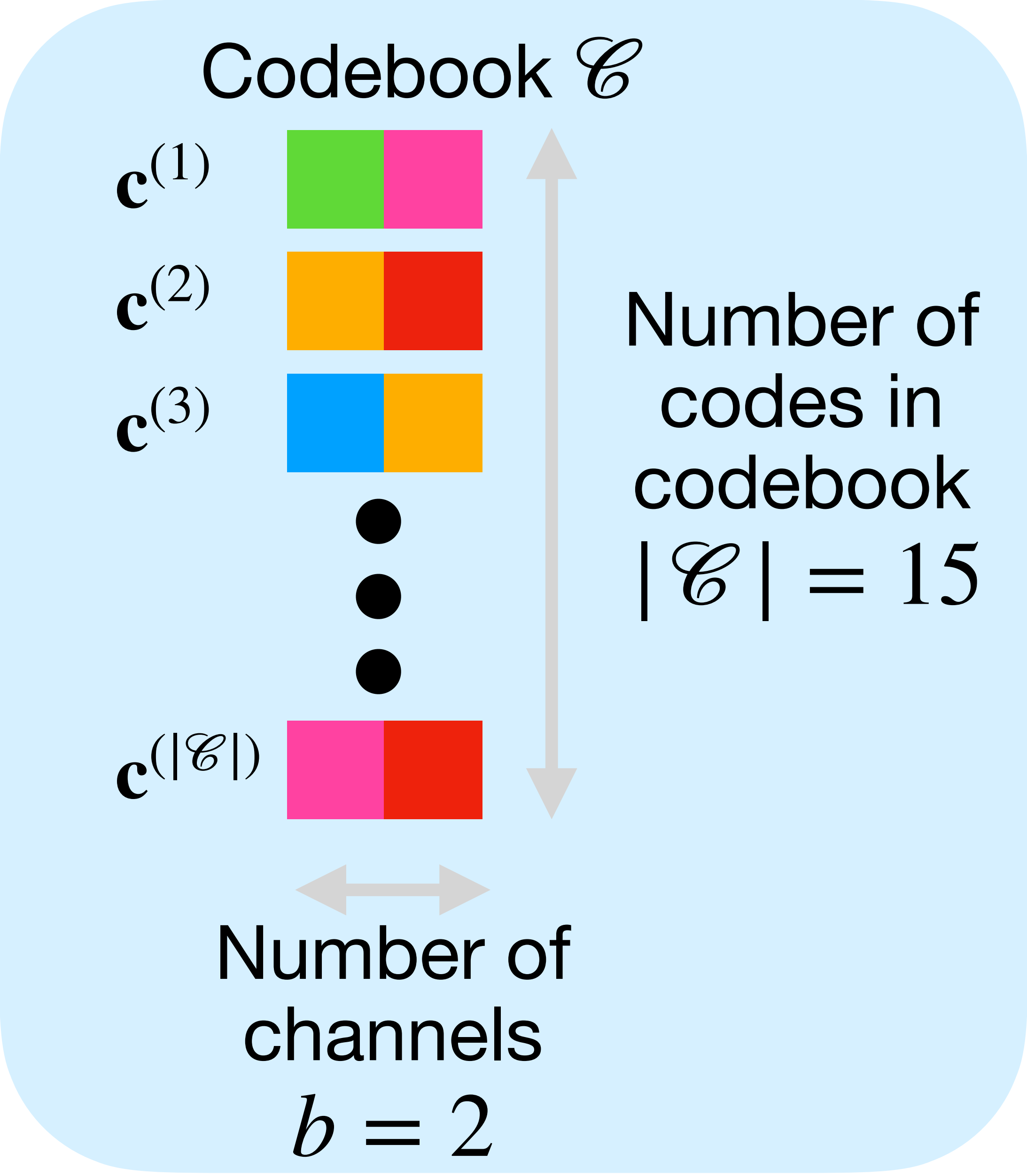
DC-MPC: World Model Training



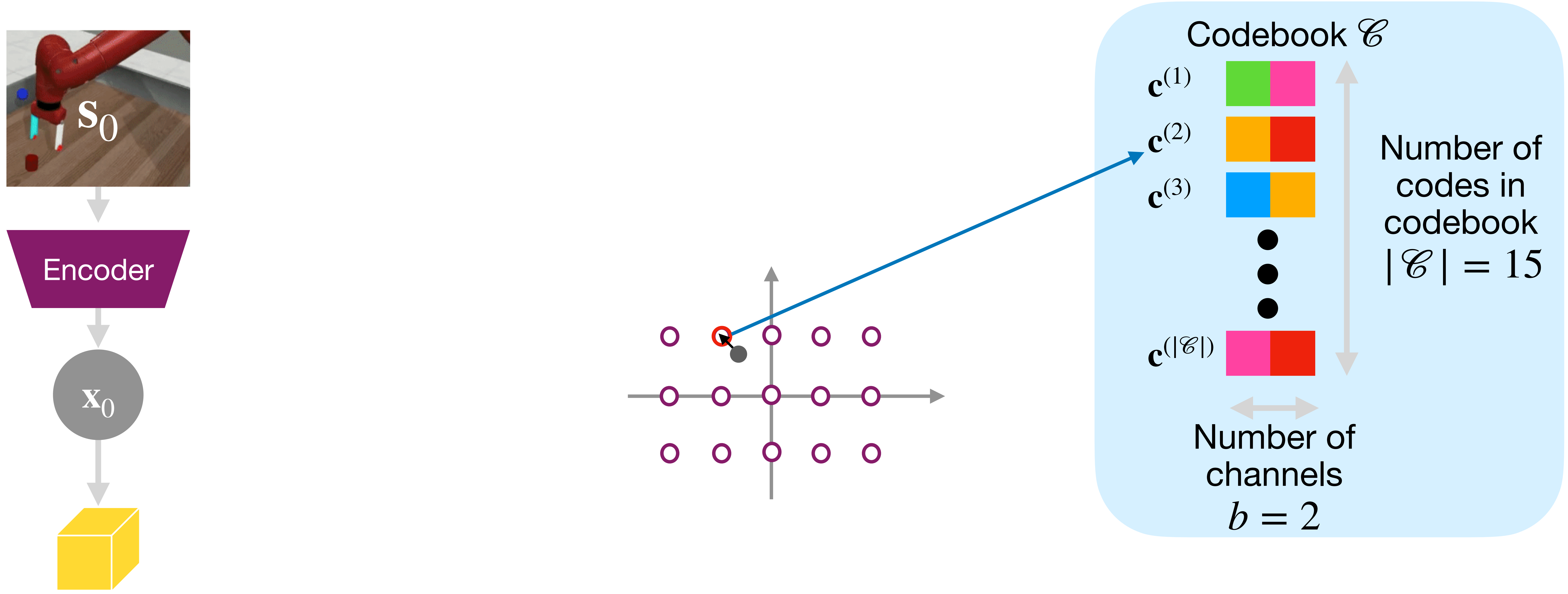
DC-MPC: World Model Training



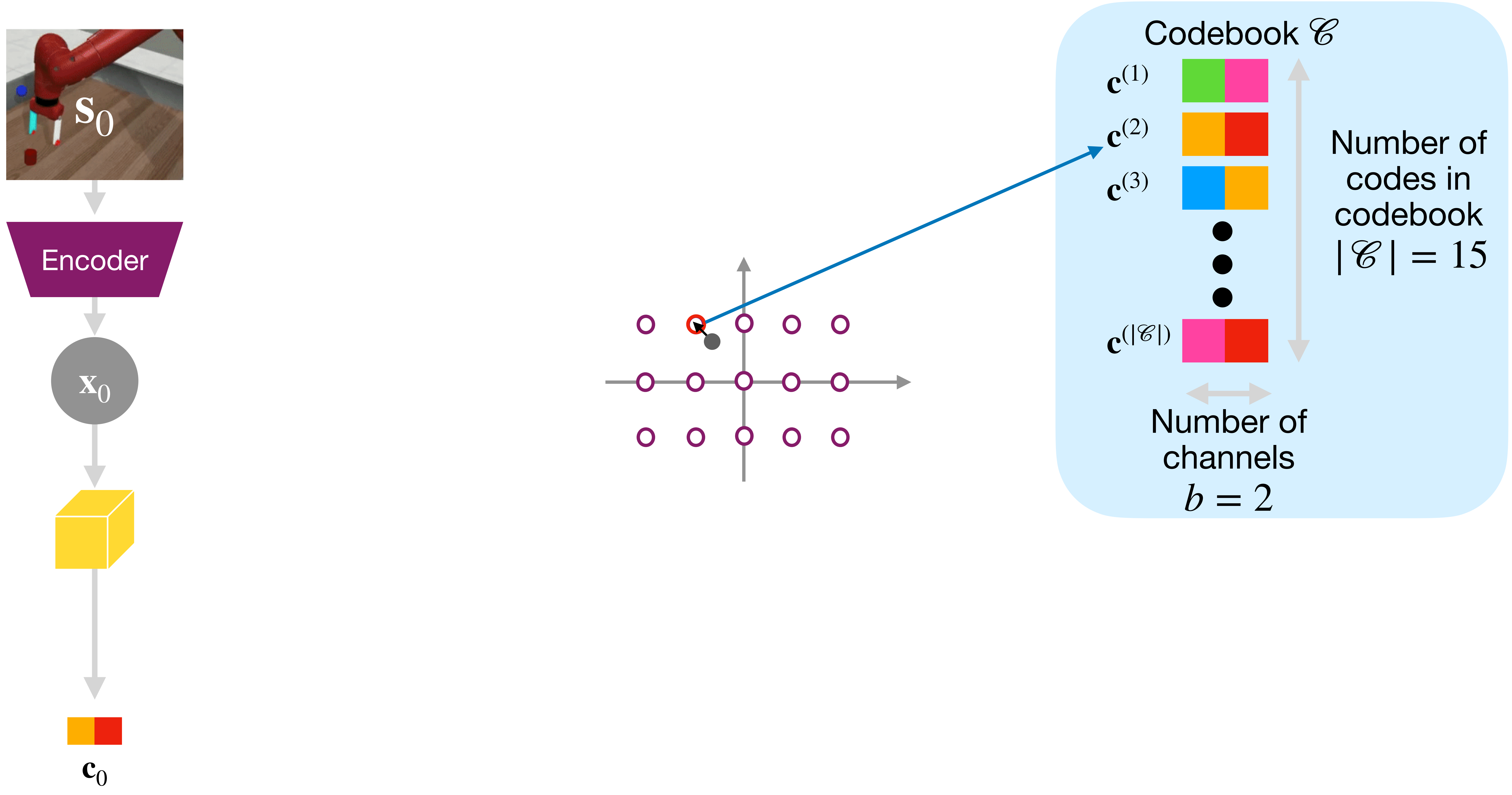




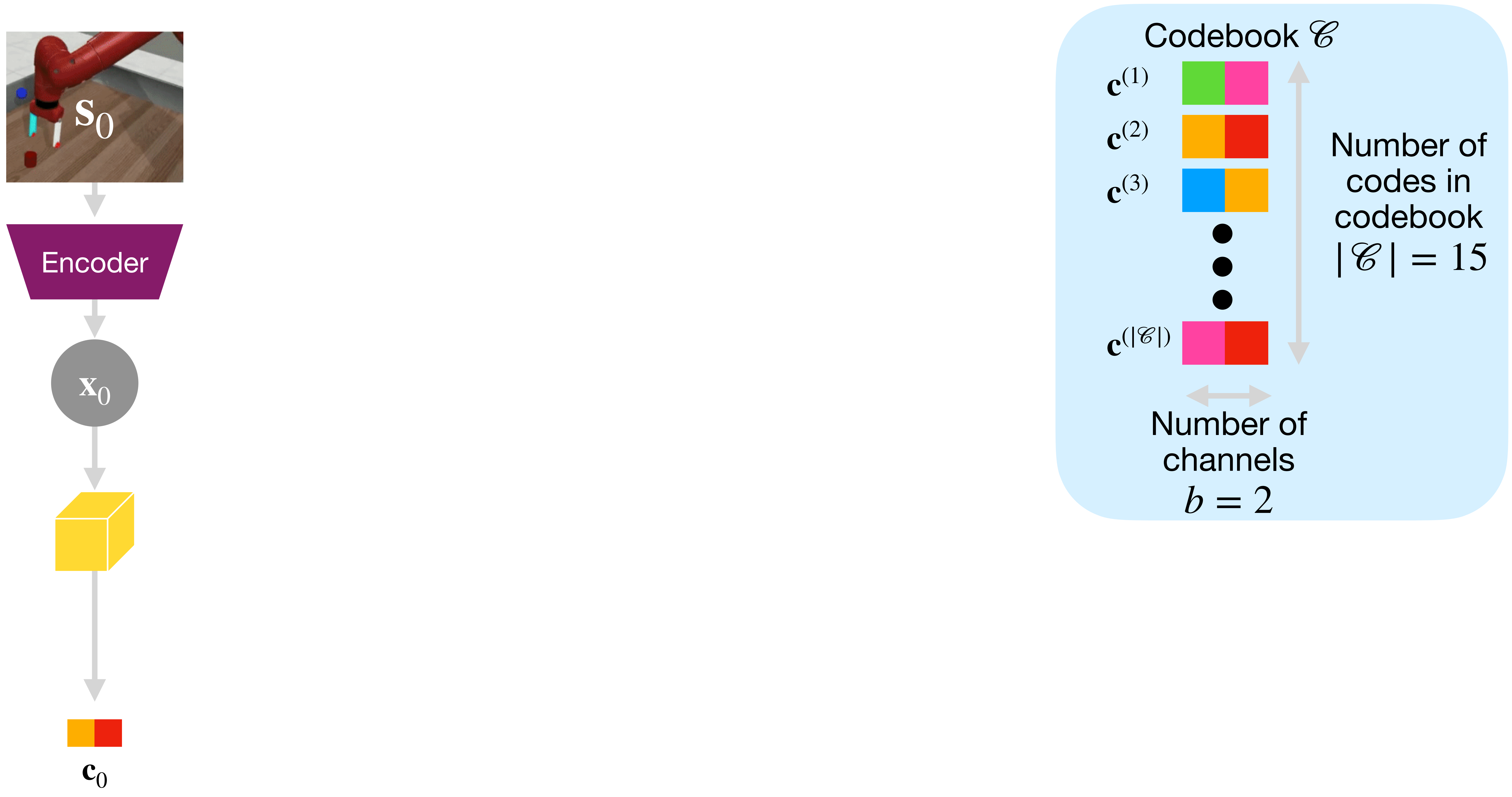
DC-MPC: World Model Training



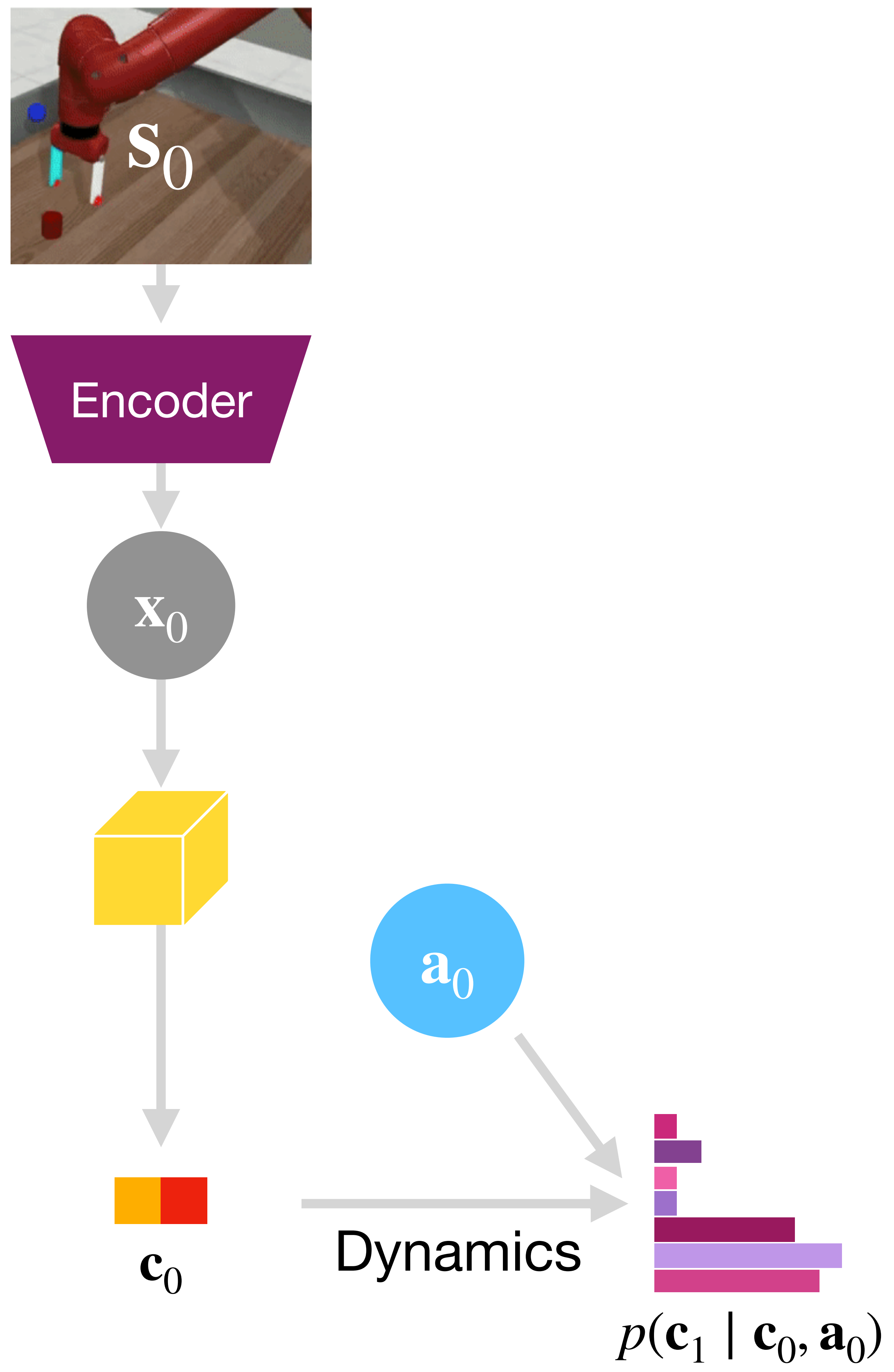
DC-MPC: World Model Training

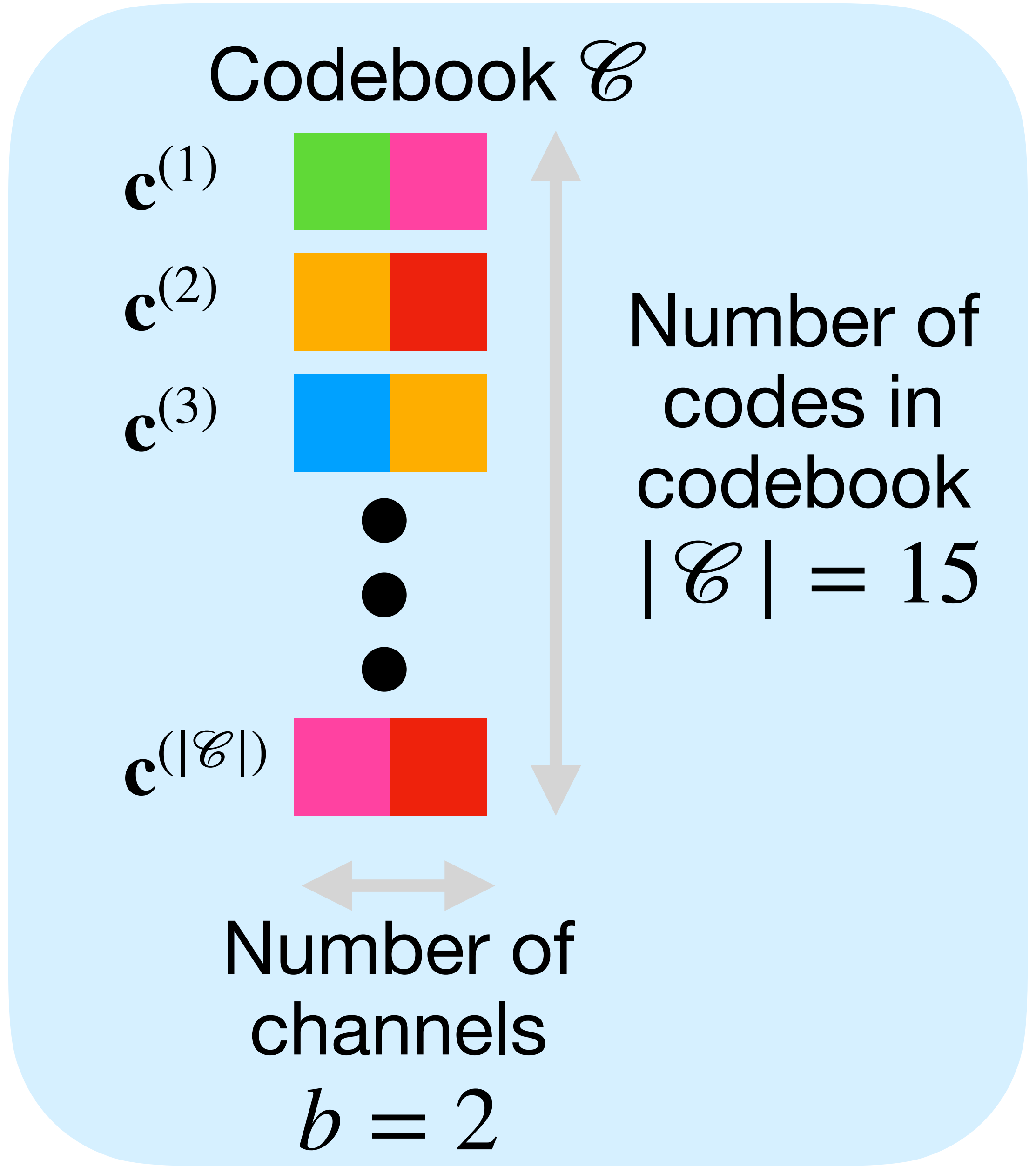


DC-MPC: World Model Training

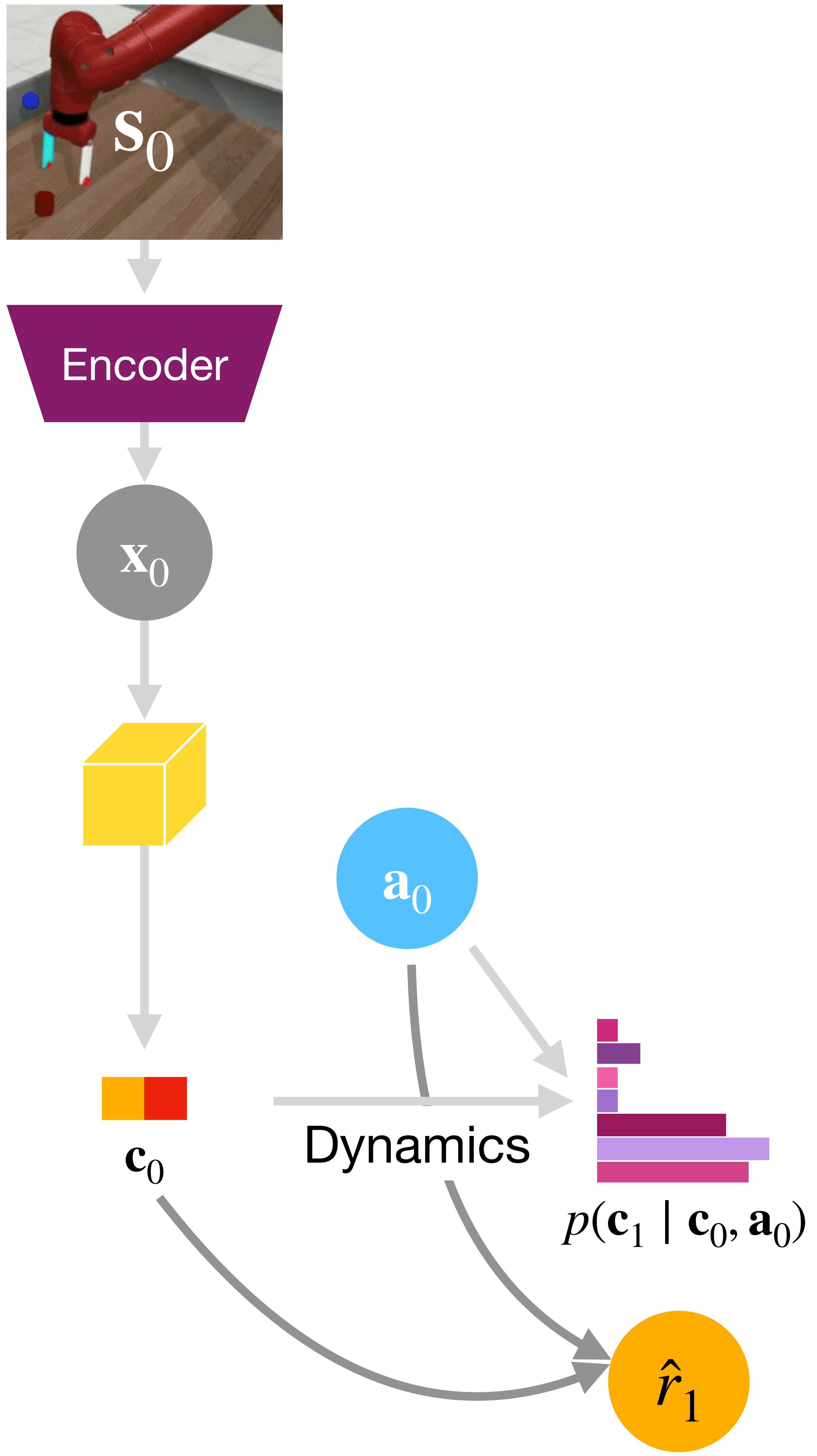


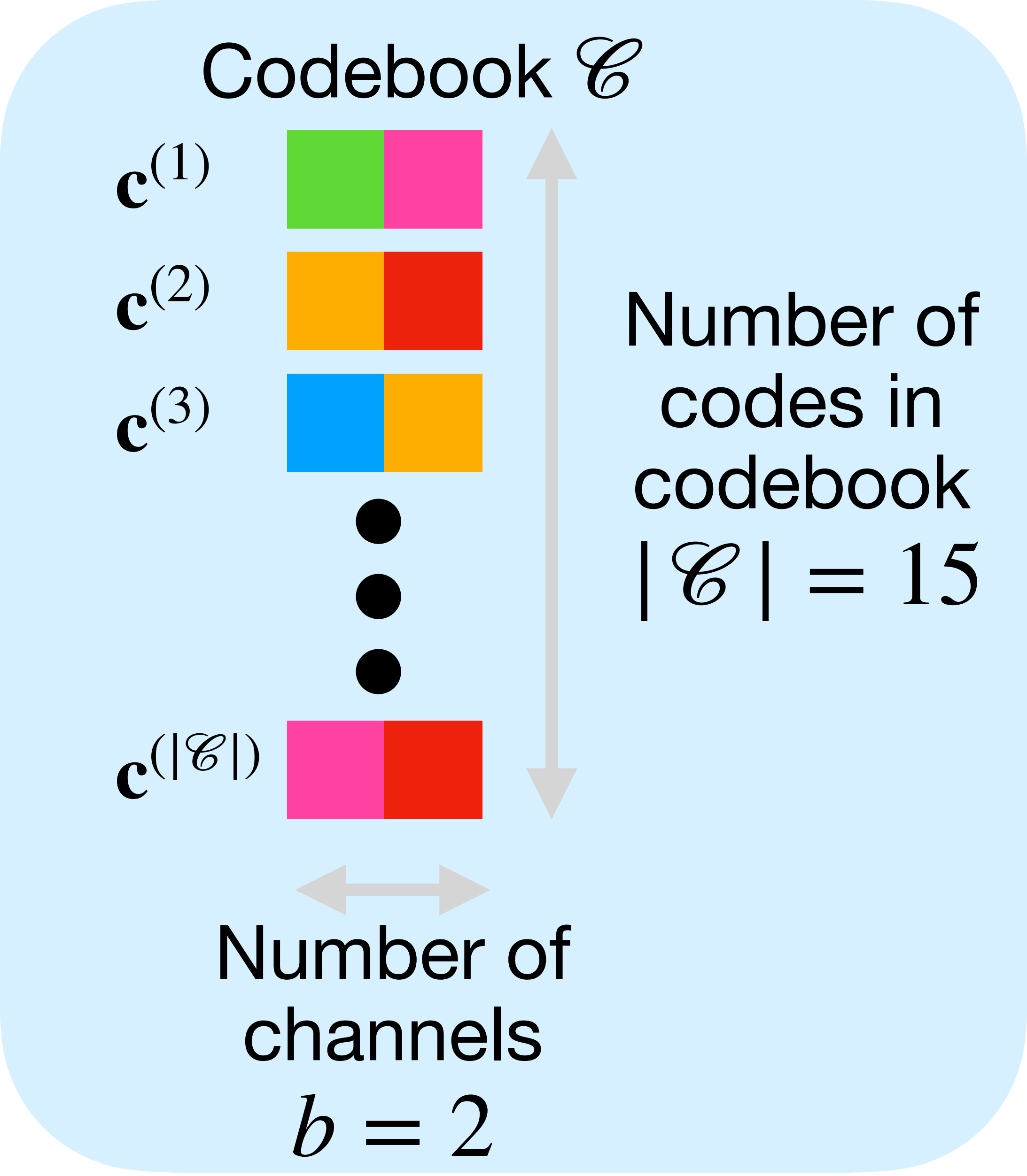
DC-MPC: World Model Training



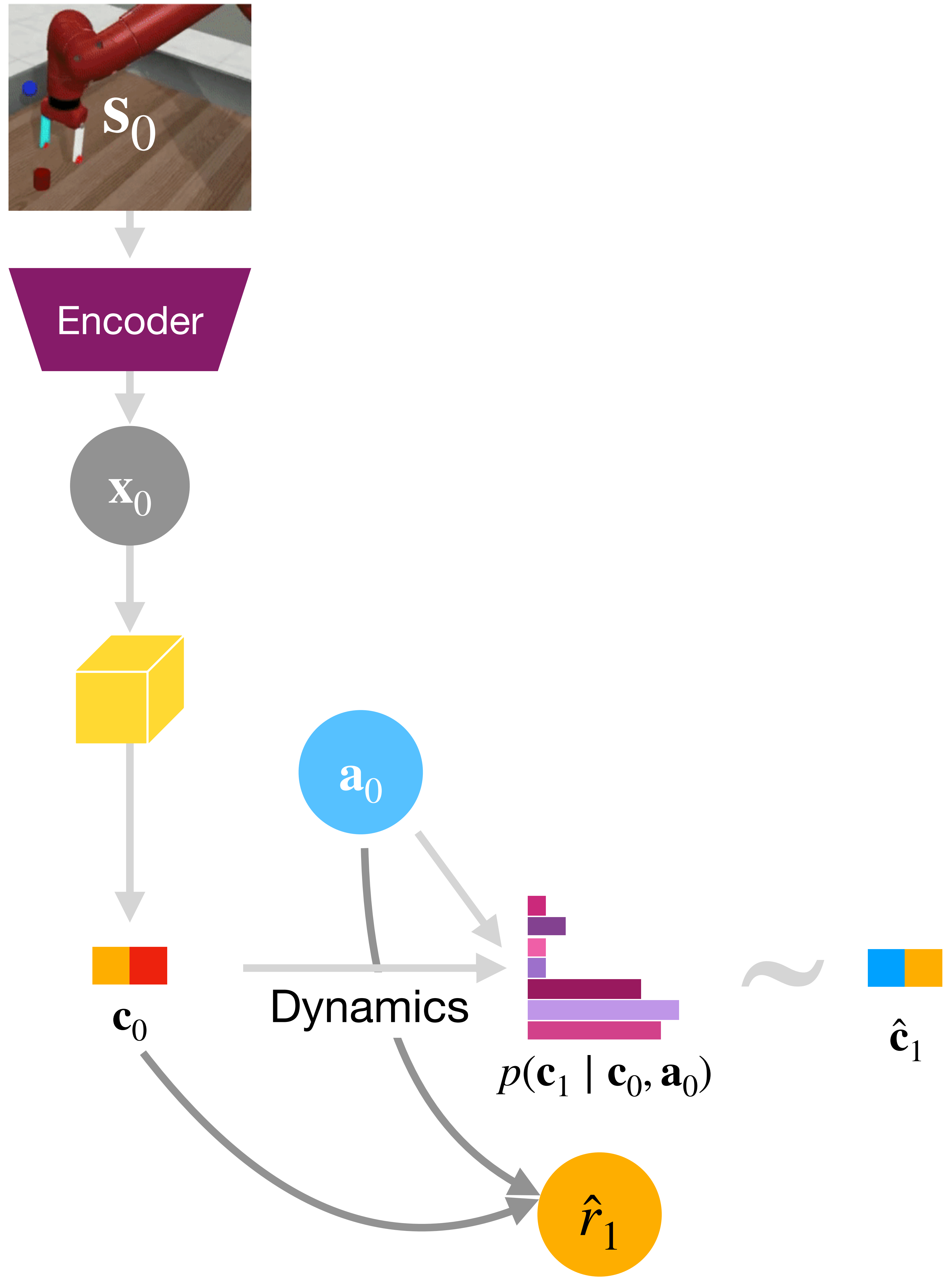


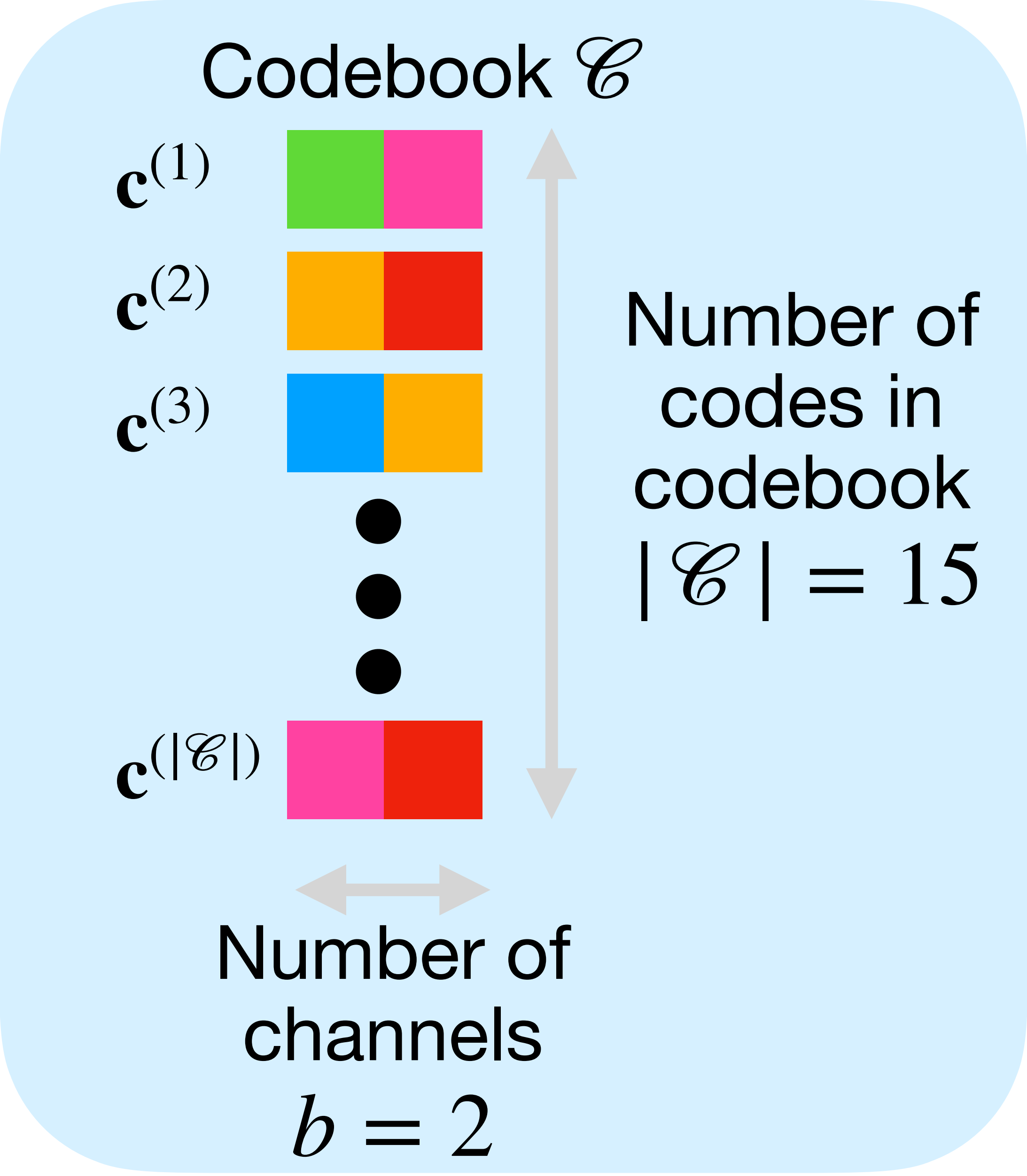
DC-MPC: World Model Training



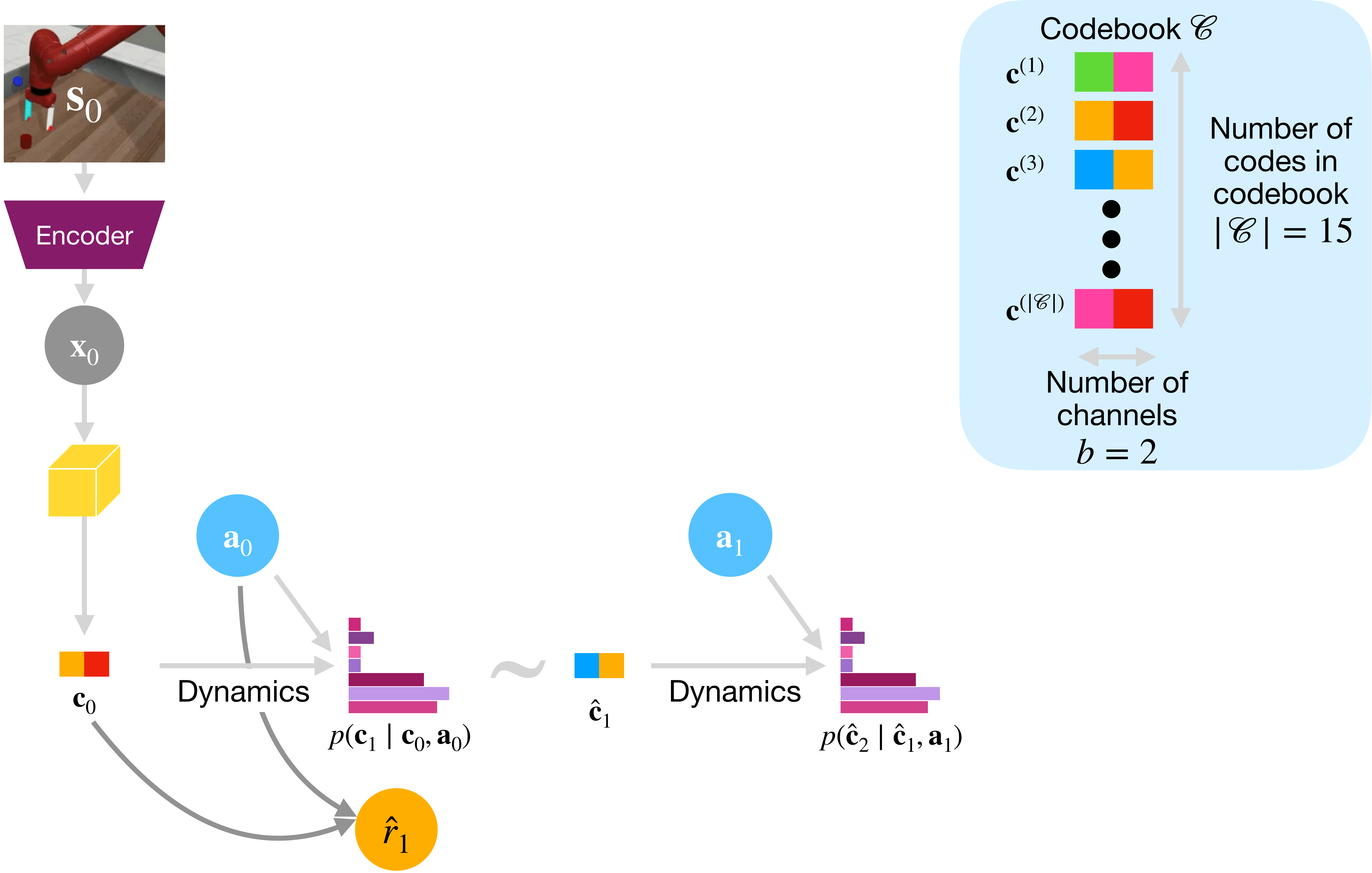


DC-MPC: World Model Training

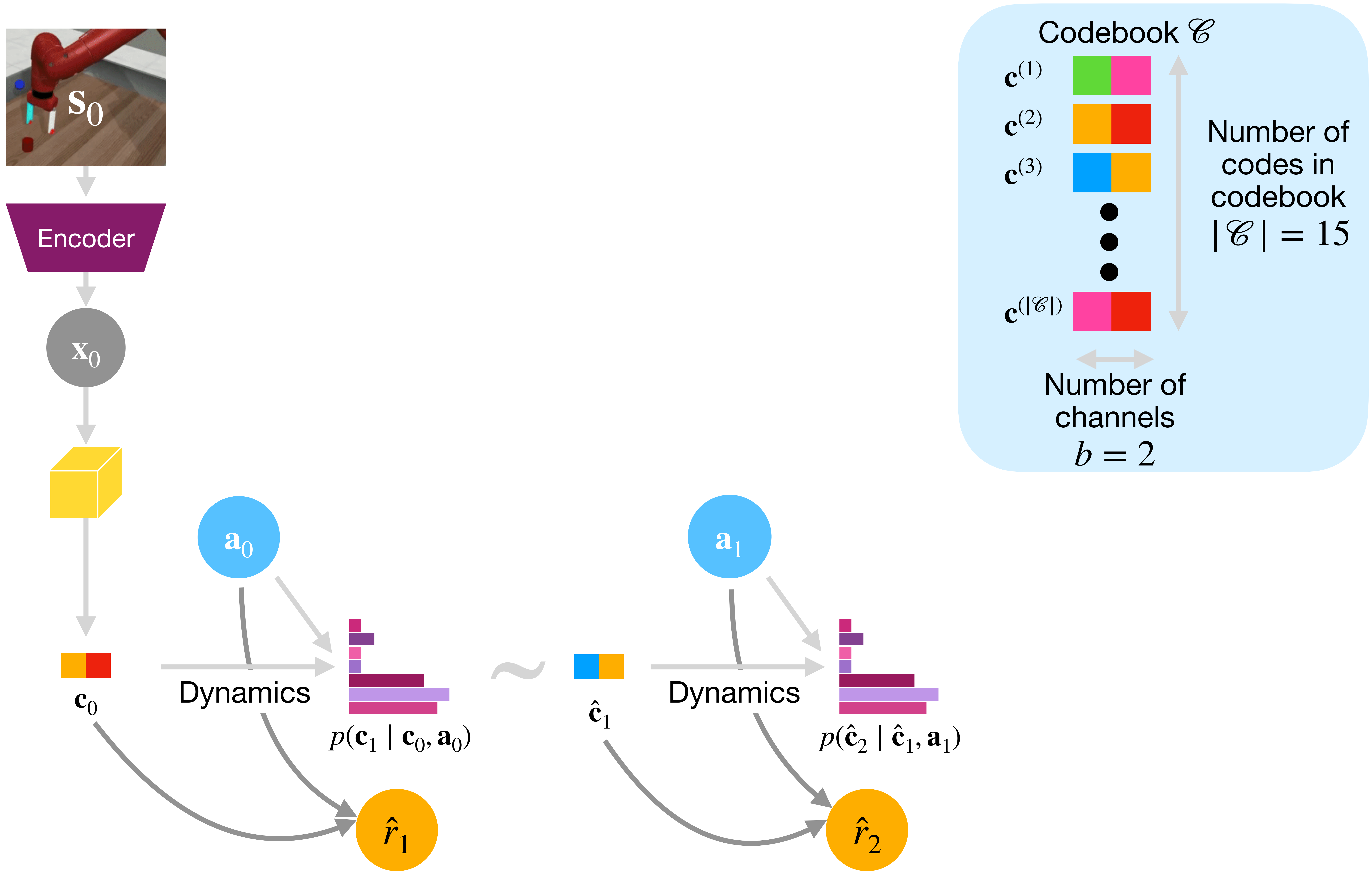




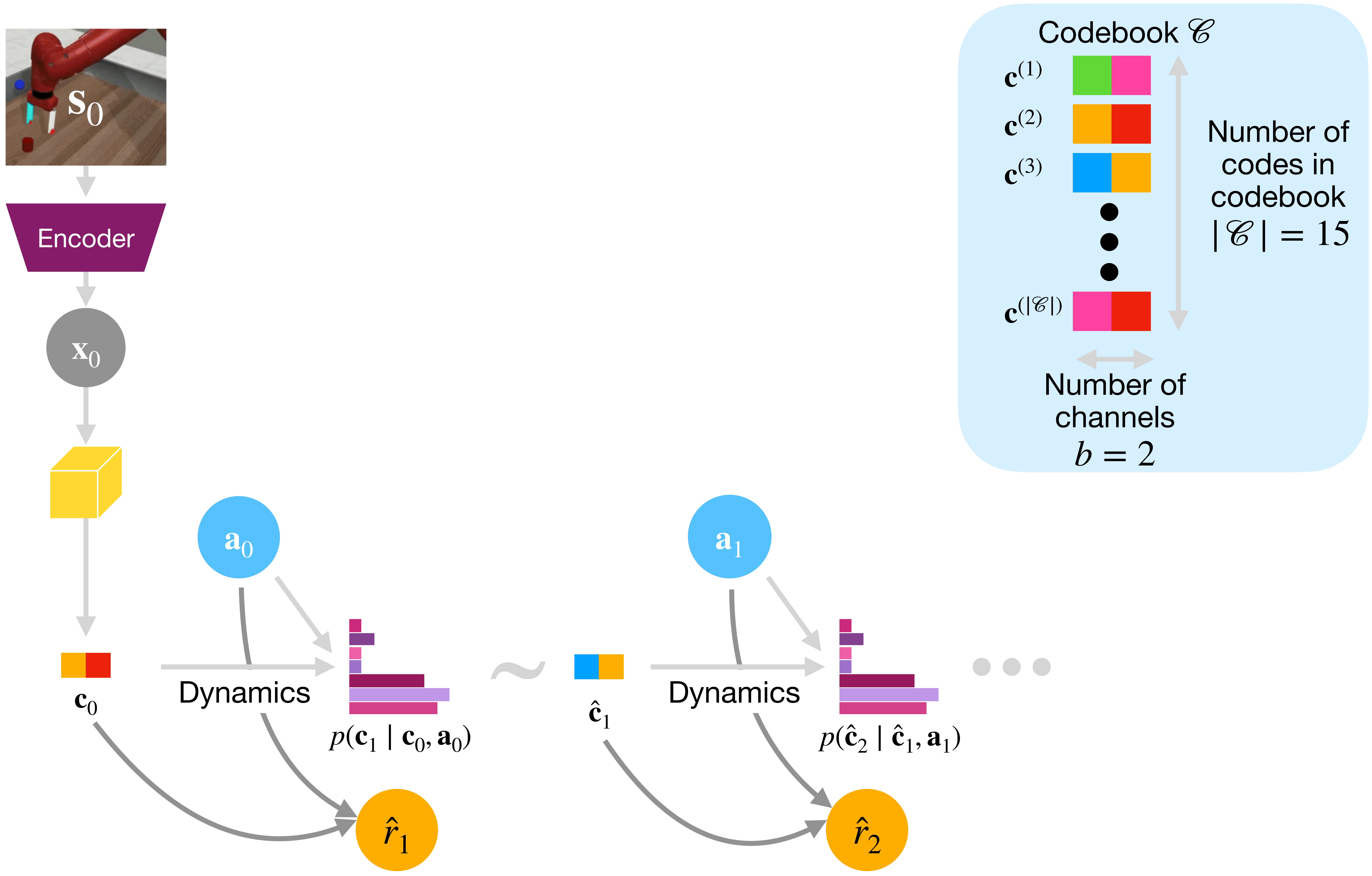
DC-MPC: World Model Training



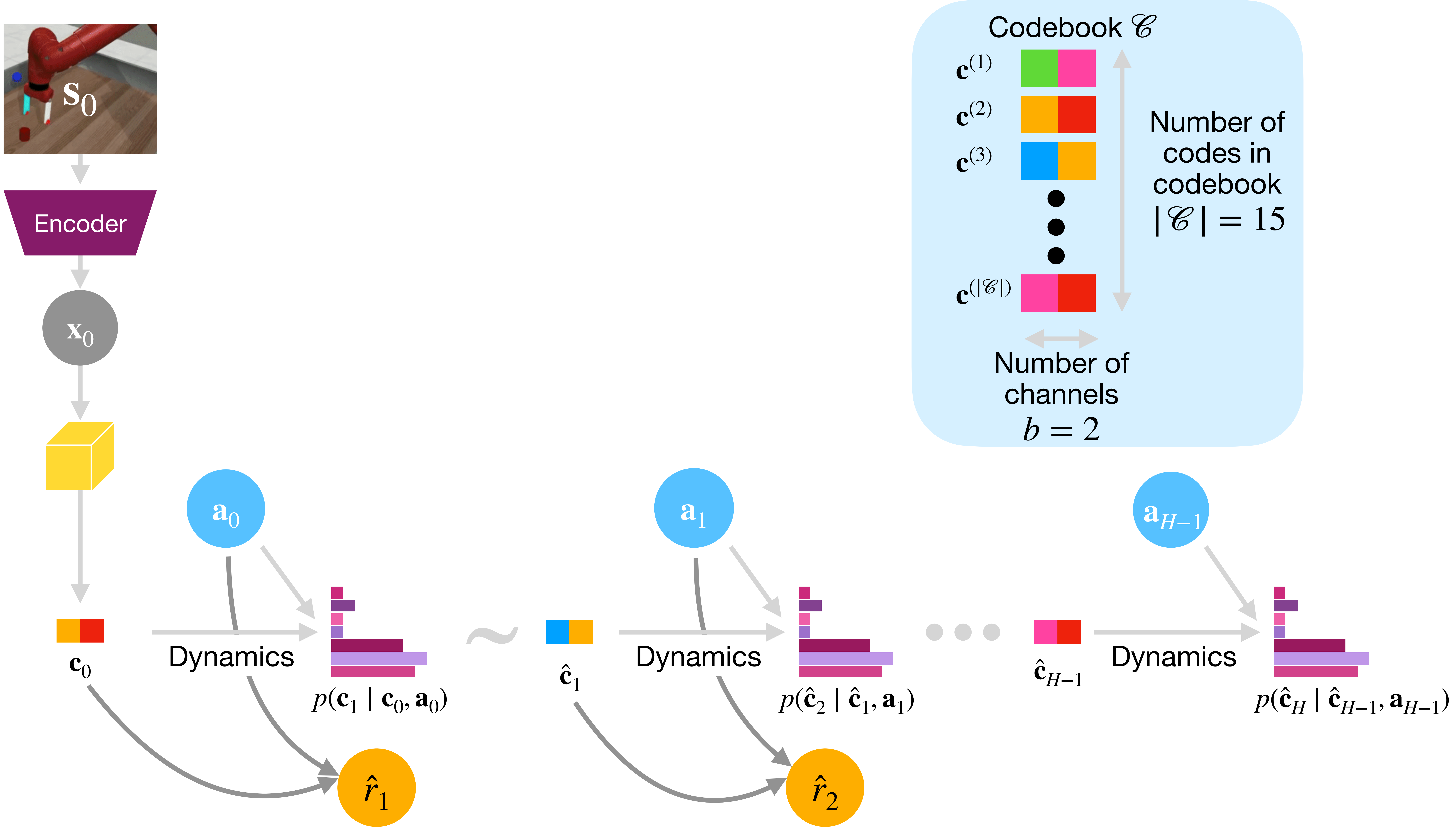
DC-MPC: World Model Training



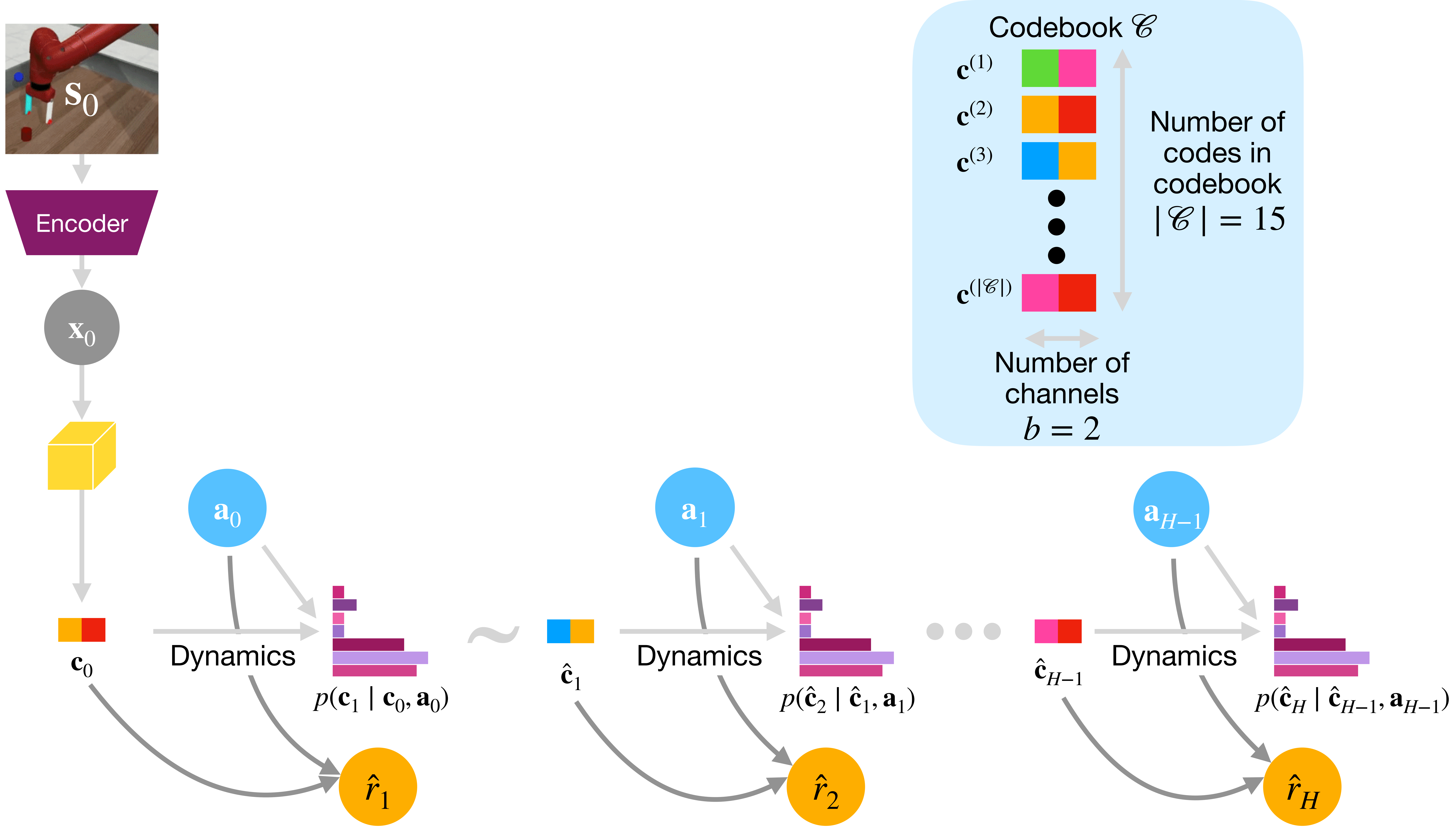
DC-MPC: World Model Training



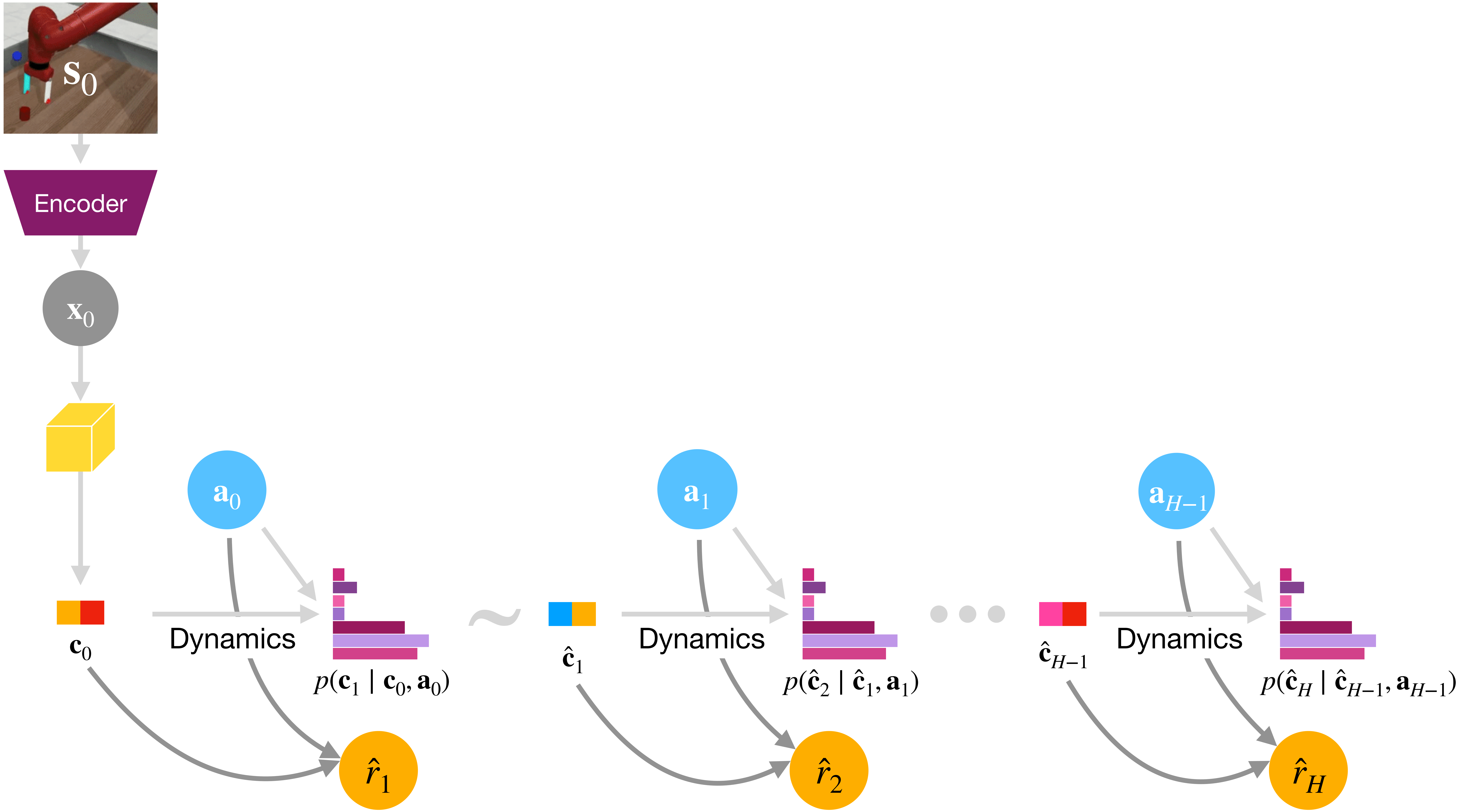
DC-MPC: World Model Training



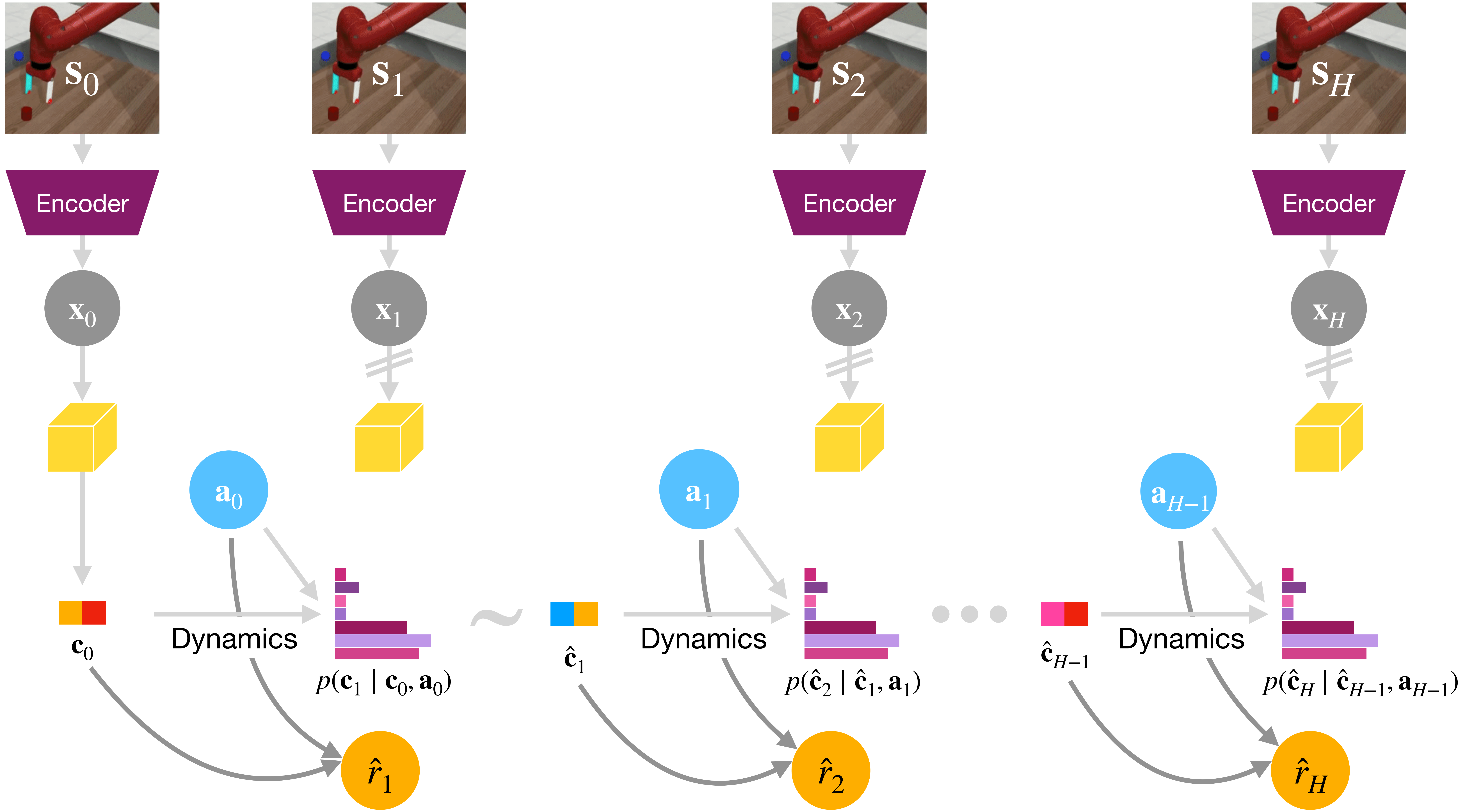
DC-MPC: World Model Training



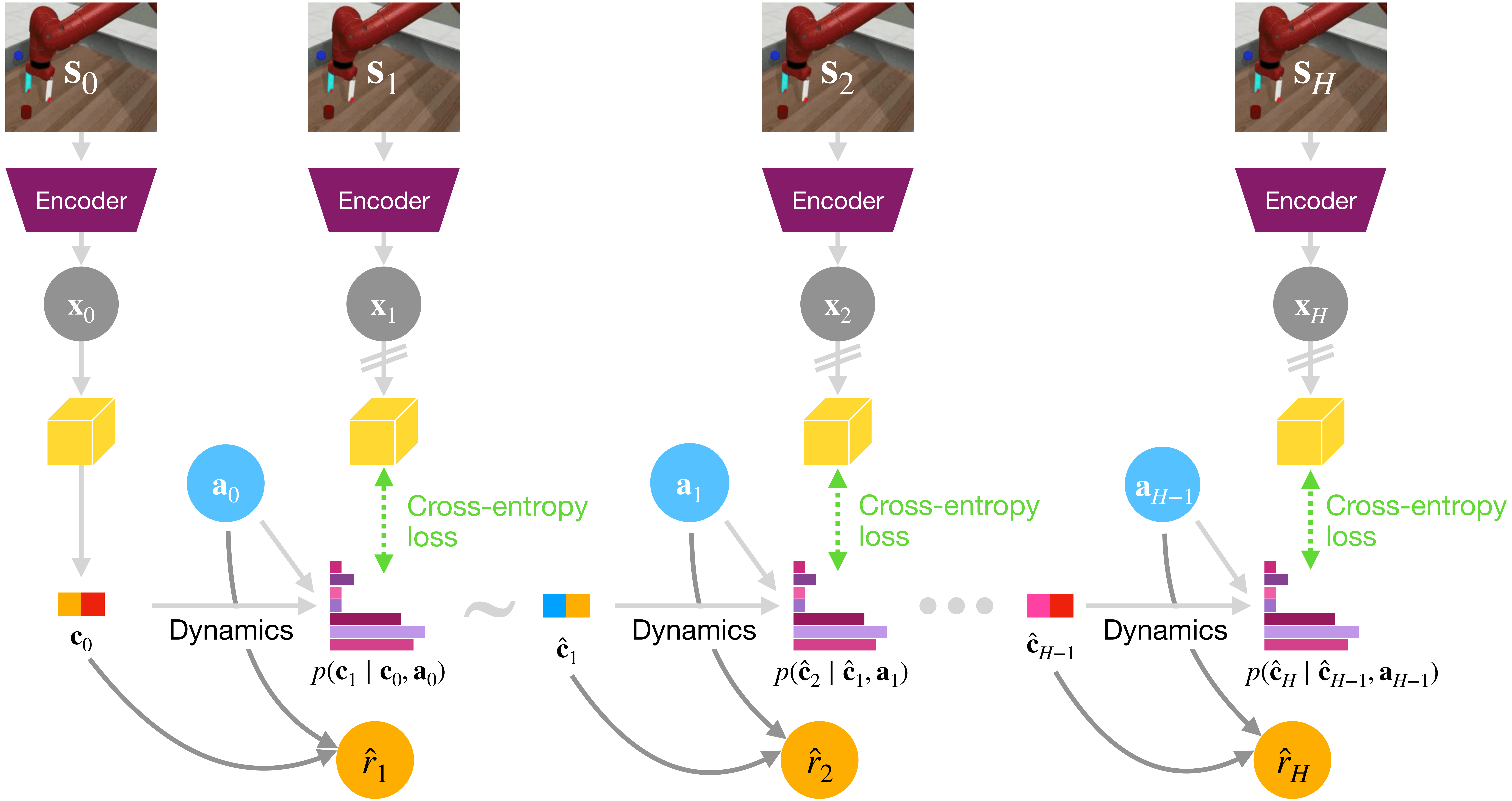
DC-MPC: World Model Training



DC-MPC: World Model Training

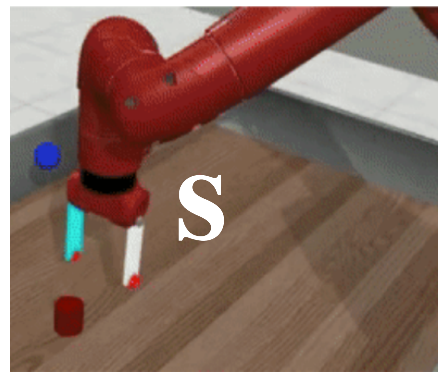


DC-MPC: World Model Training

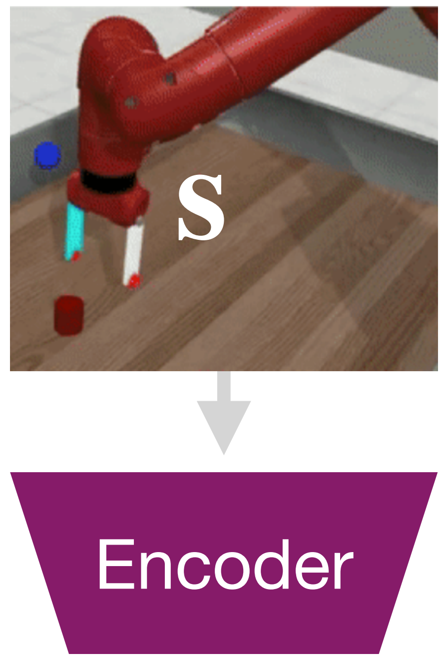


DC-MPC: Decision-time Planning

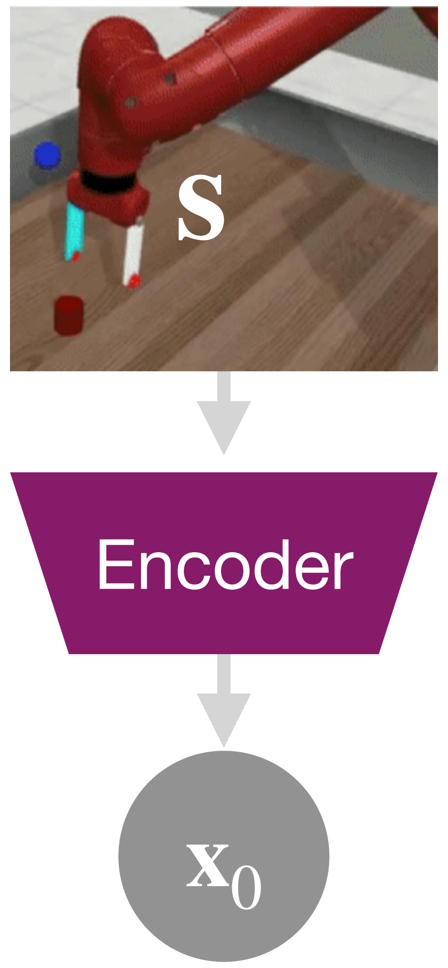
DC-MPC: Decision-time Planning



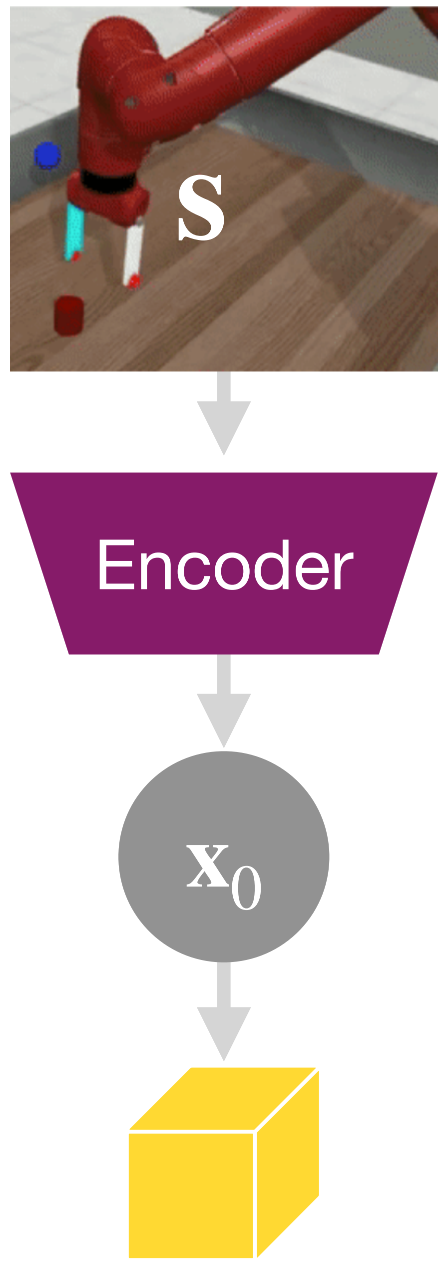
DC-MPC: Decision-time Planning



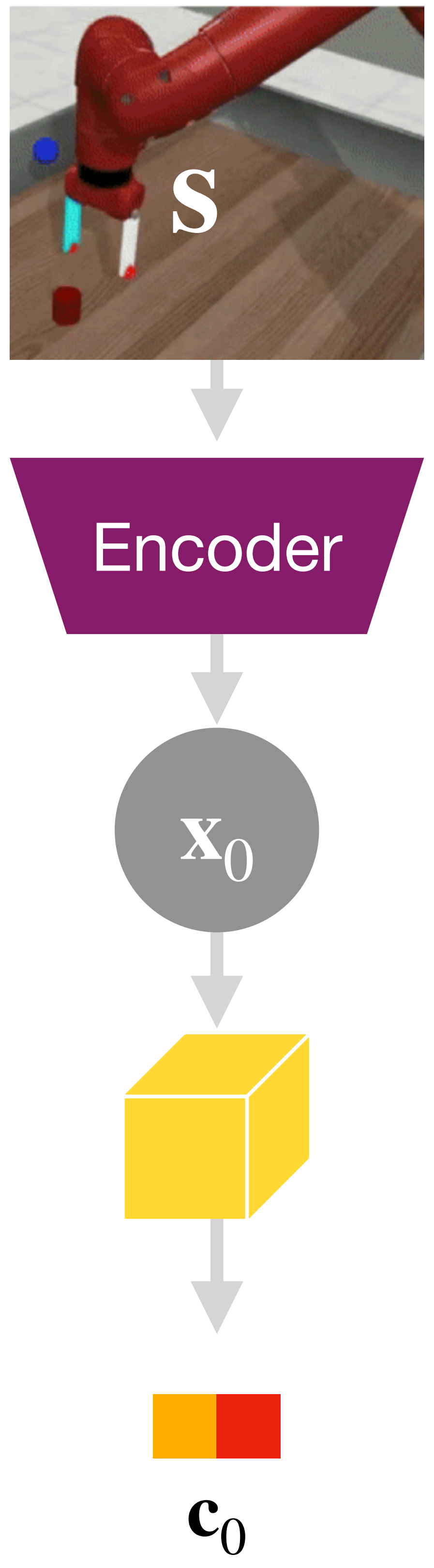
DC-MPC: Decision-time Planning



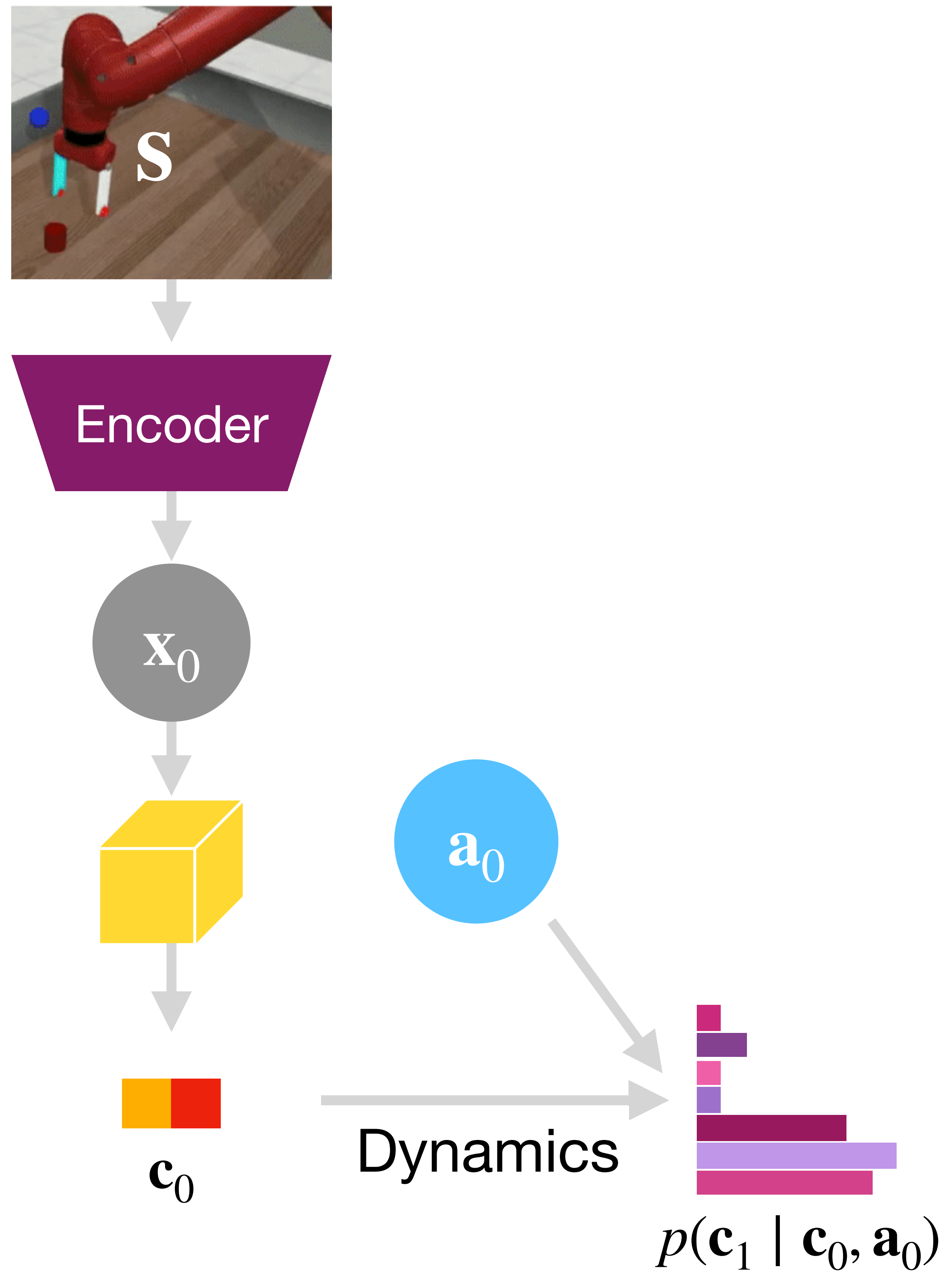
DC-MPC: Decision-time Planning



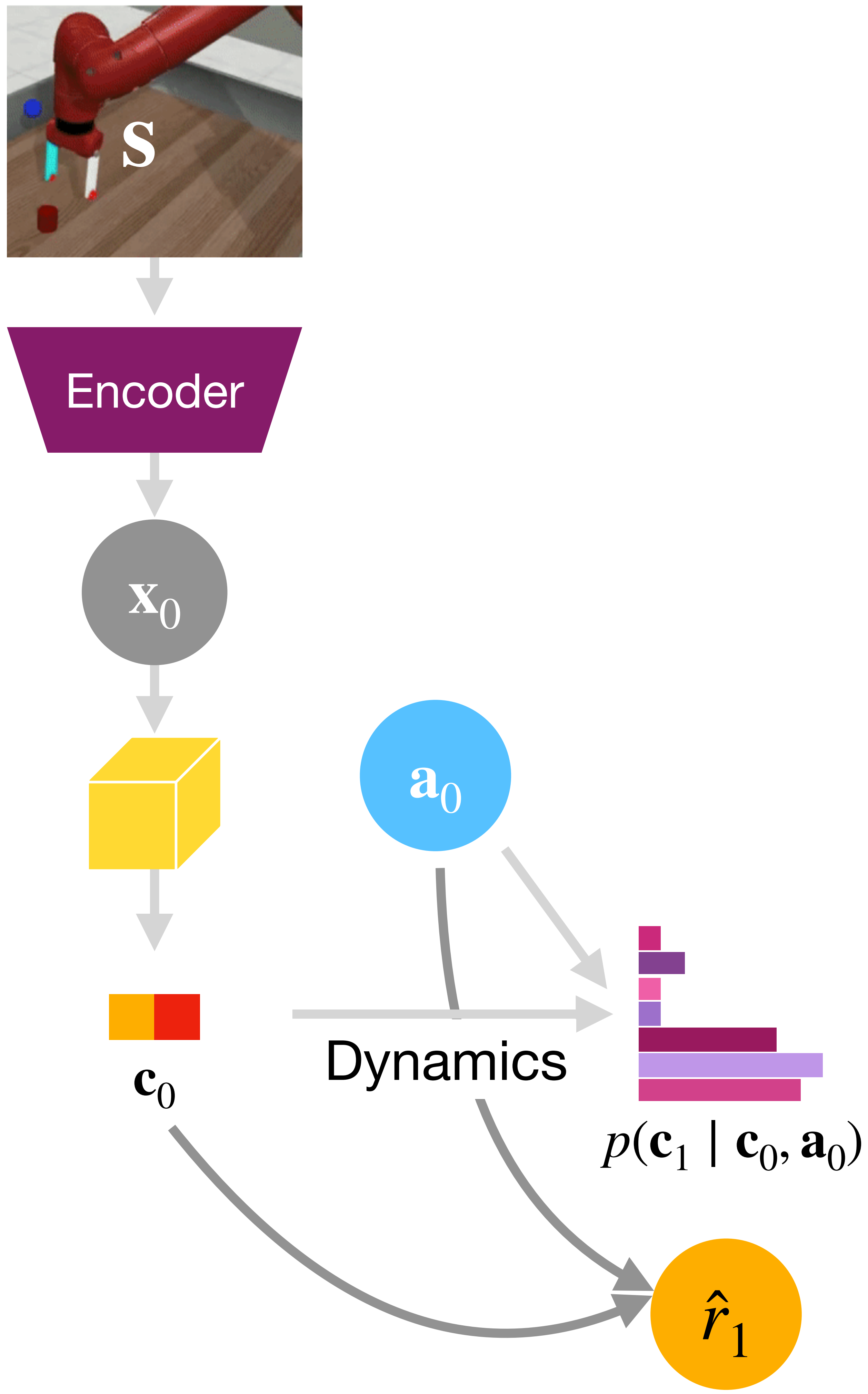
DC-MPC: Decision-time Planning



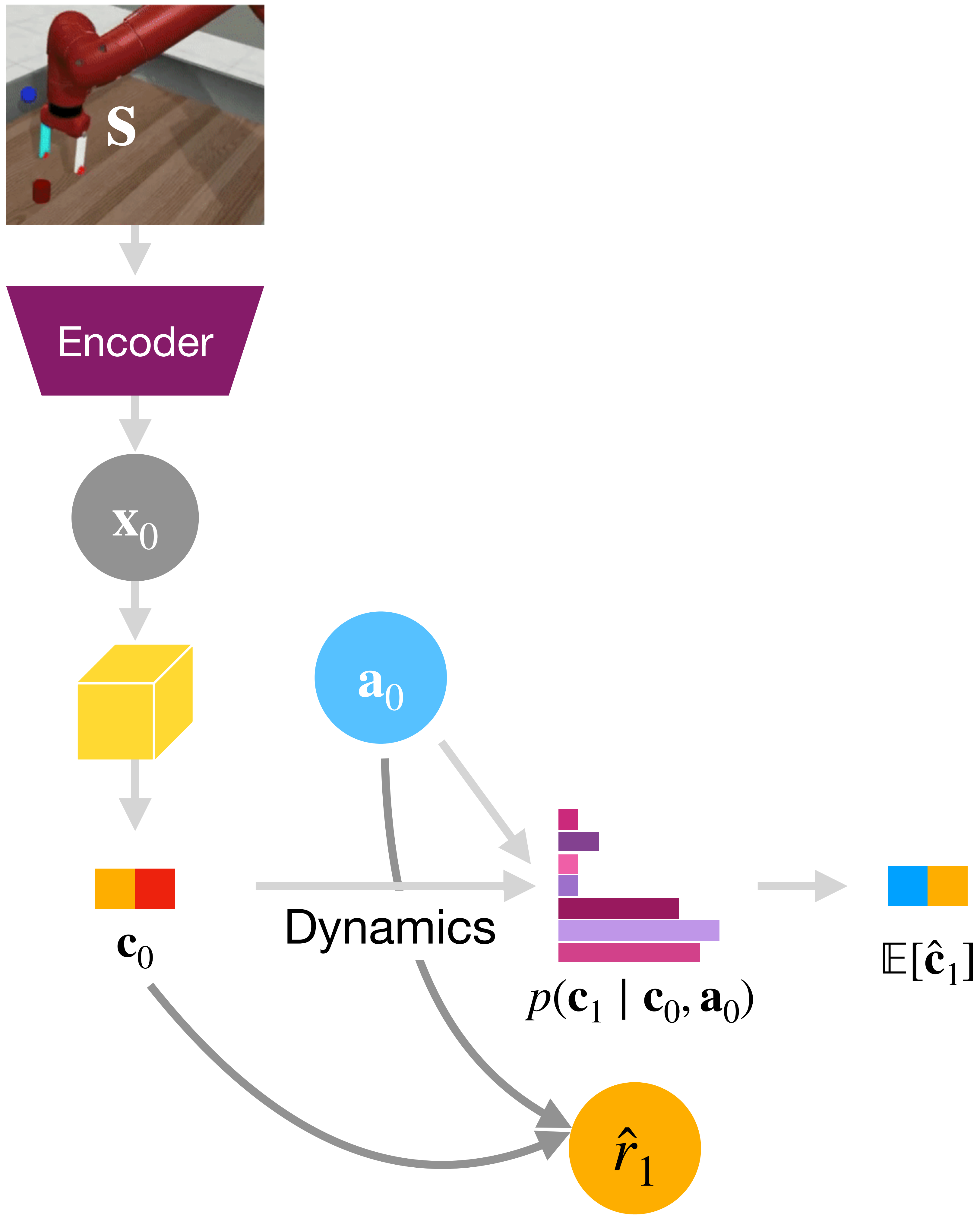
DC-MPC: Decision-time Planning



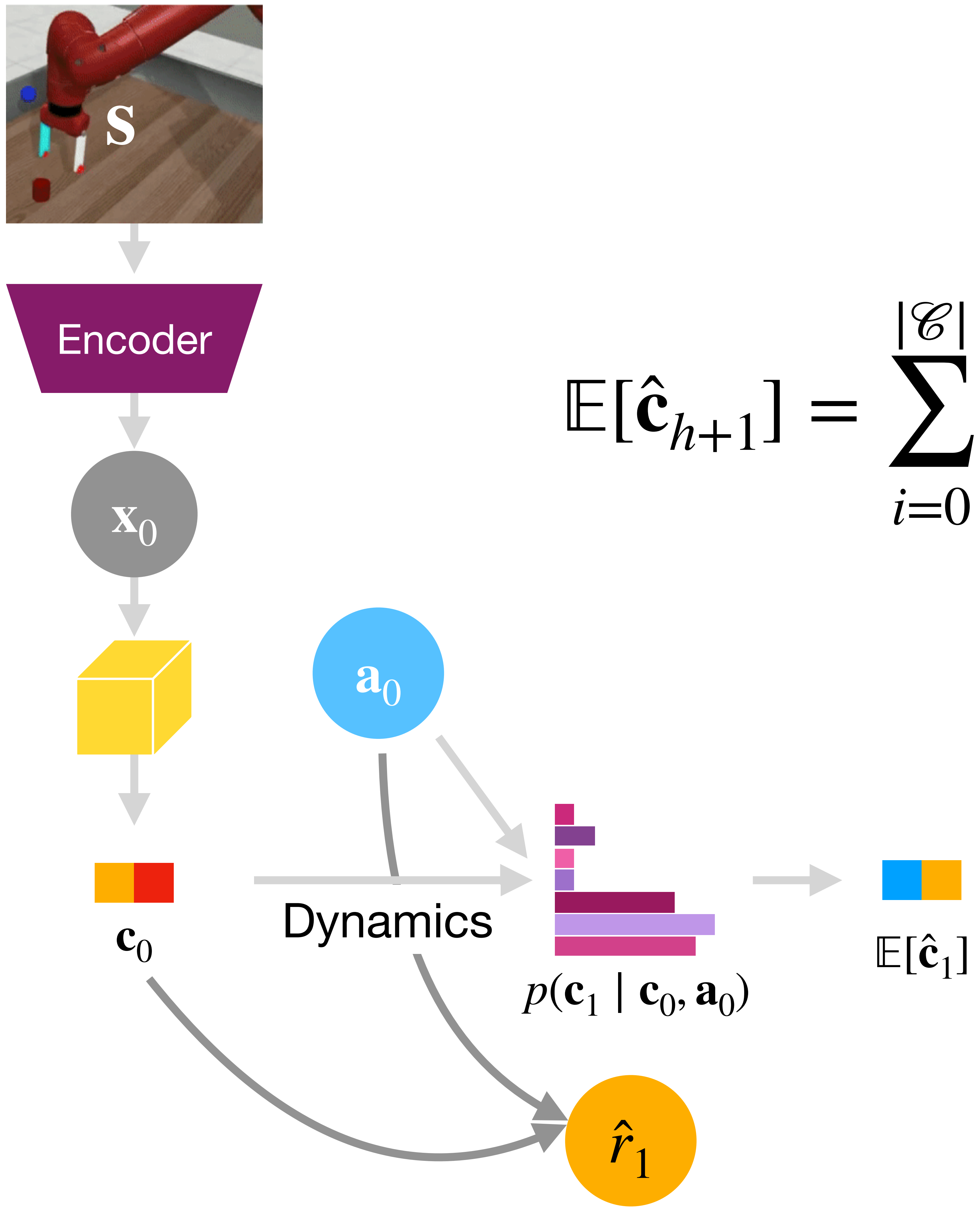
DC-MPC: Decision-time Planning



DC-MPC: Decision-time Planning

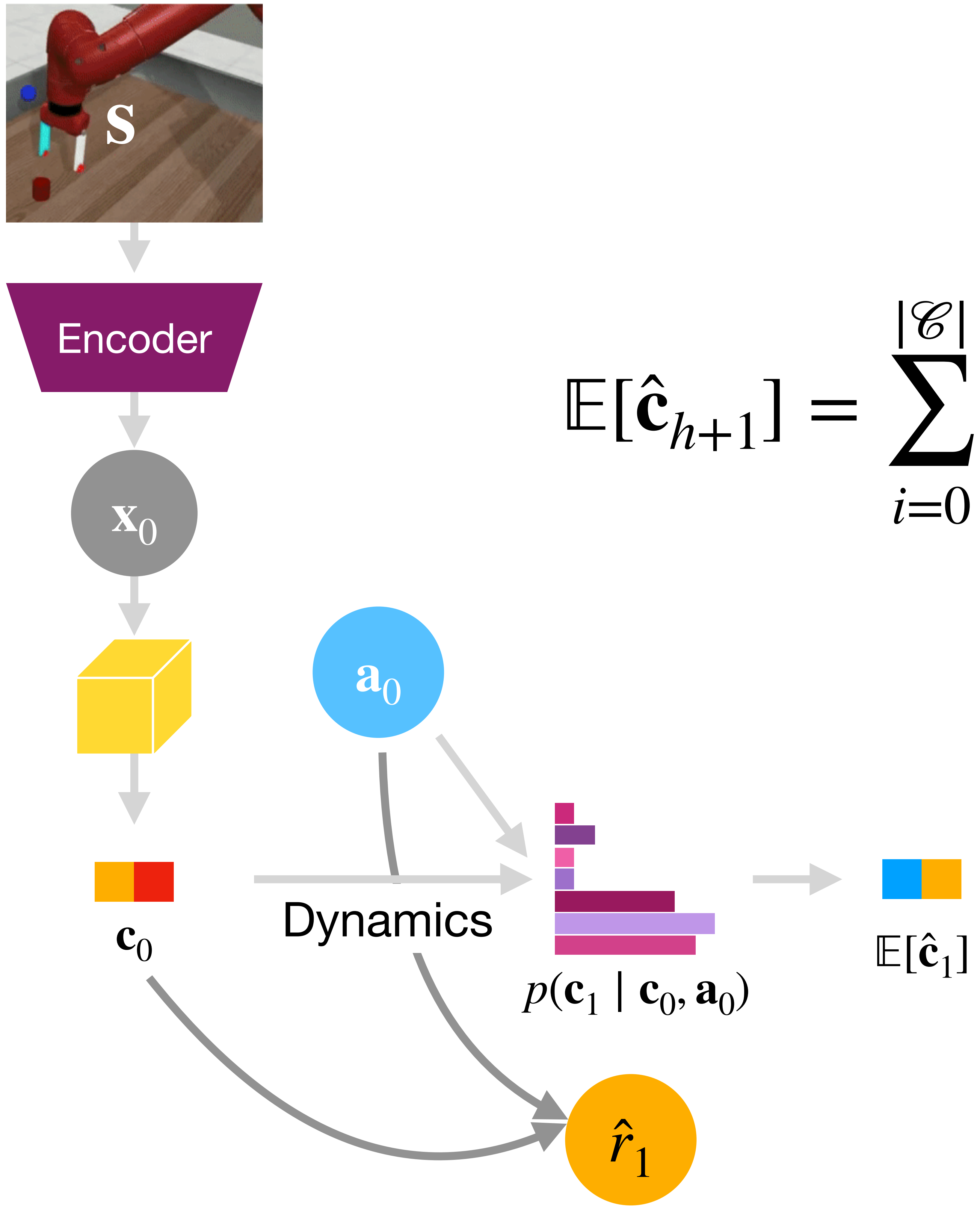


DC-MPC: Decision-time Planning



$$\mathbb{E}[\hat{\mathbf{c}}_{h+1}] = \sum_{i=0}^{|\mathscr{C}|} \underbrace{\Pr(\hat{\mathbf{c}}_{h+1} = \mathbf{c}^{(i)} \mid \hat{\mathbf{c}}_h, \mathbf{a}_h)}_{\text{prob. of code } i} \underbrace{\mathbf{c}^{(i)}}_{\text{code}}$$

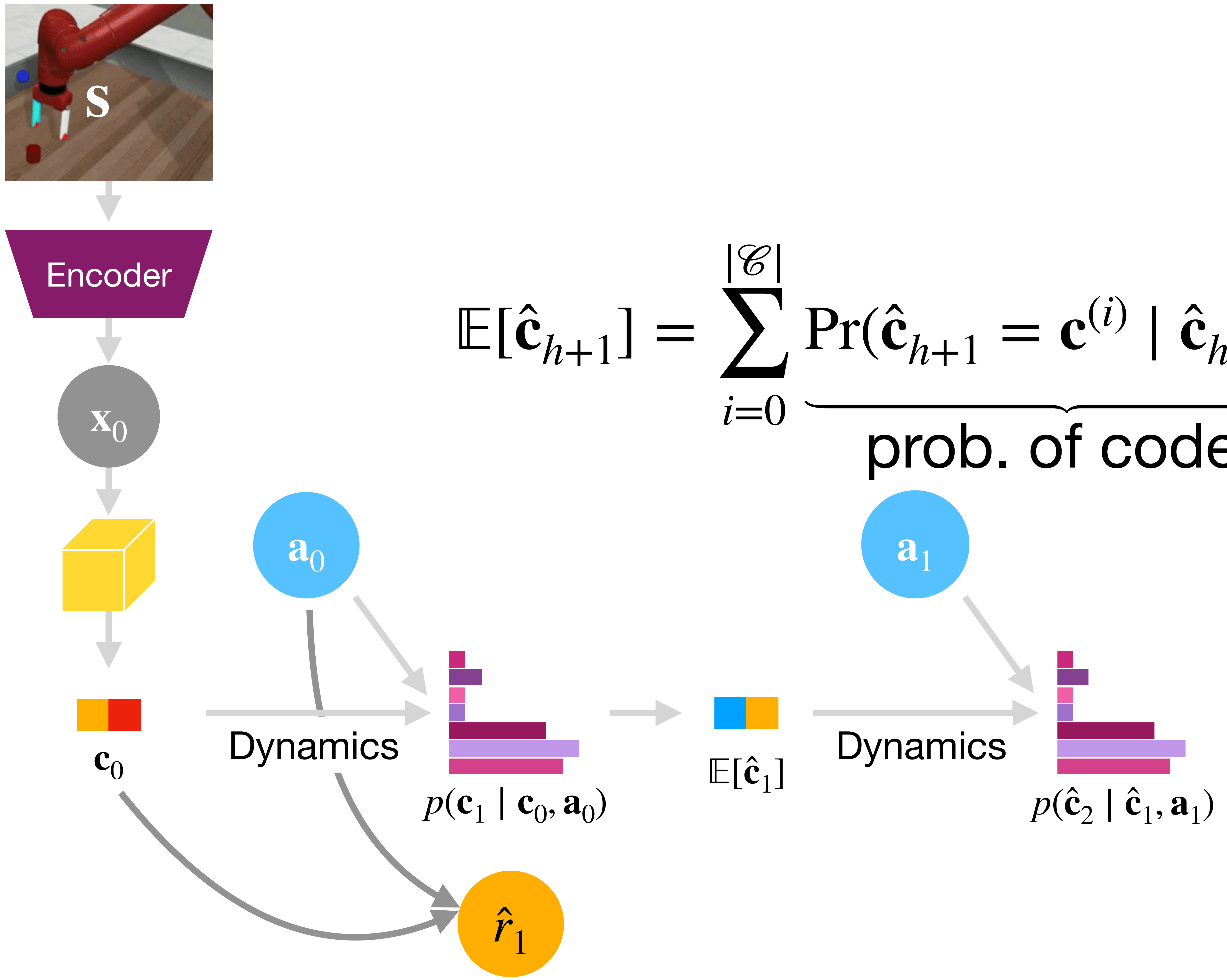
DC-MPC: Decision-time Planning



$$\mathbb{E}[\hat{\mathbf{c}}_{h+1}] = \sum_{i=0}^{|\mathscr{C}|} \underbrace{\Pr(\hat{\mathbf{c}}_{h+1} = \mathbf{c}^{(i)} \mid \hat{\mathbf{c}}_h, \mathbf{a}_h)}_{\text{prob. of code } i} \underbrace{\mathbf{c}^{(i)}}_{\text{code}}$$

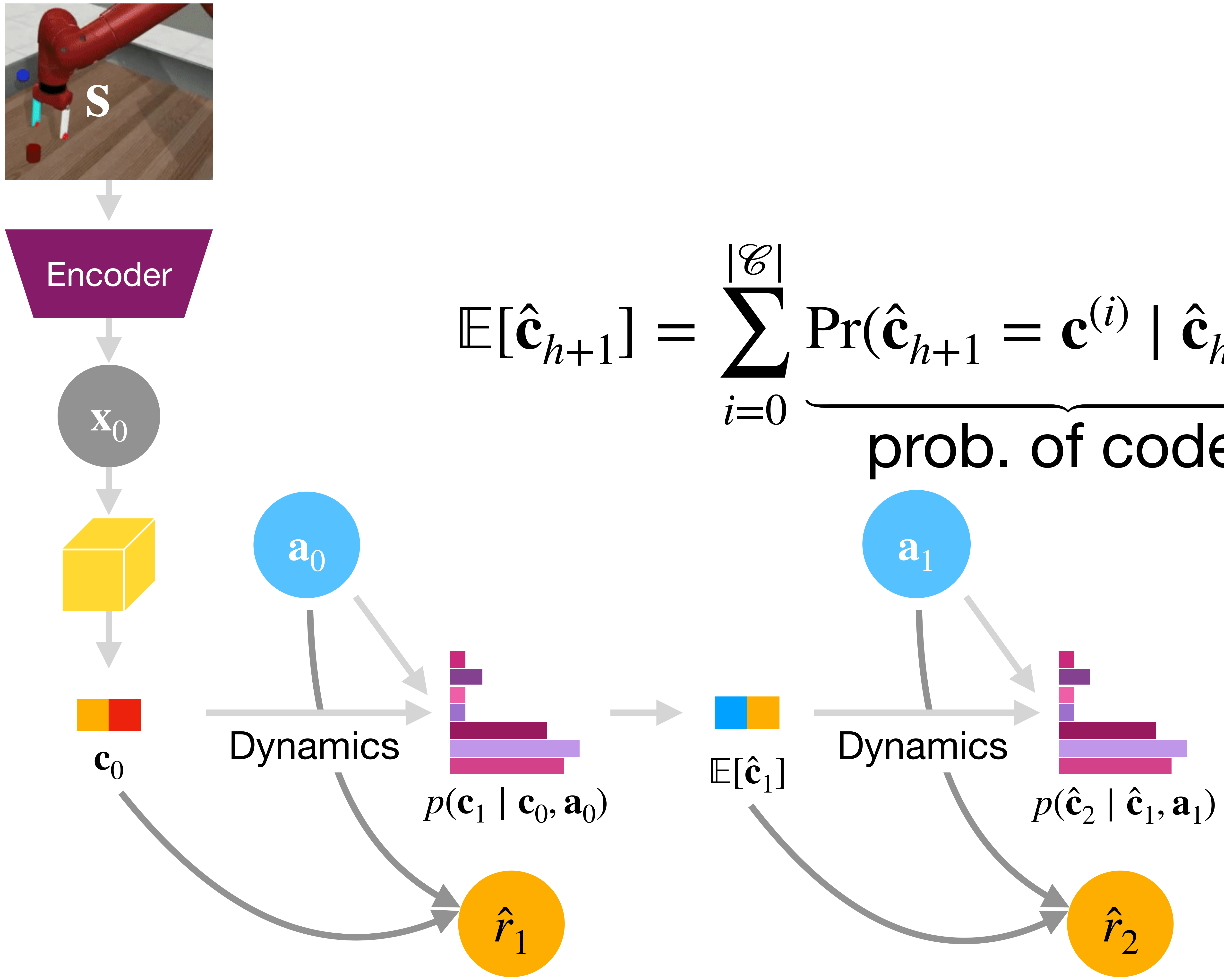
DC-MPC: Decision-time Planning

$$\mathbb{E}[\hat{\mathbf{c}}_{h+1}] = \sum_{i=0}^{|\mathscr{C}|} \underbrace{\Pr(\hat{\mathbf{c}}_{h+1} = \mathbf{c}^{(i)} \mid \hat{\mathbf{c}}_h, \mathbf{a}_h)}_{\text{prob. of code } i} \; \underbrace{\mathbf{c}^{(i)}}_{\text{code}}$$



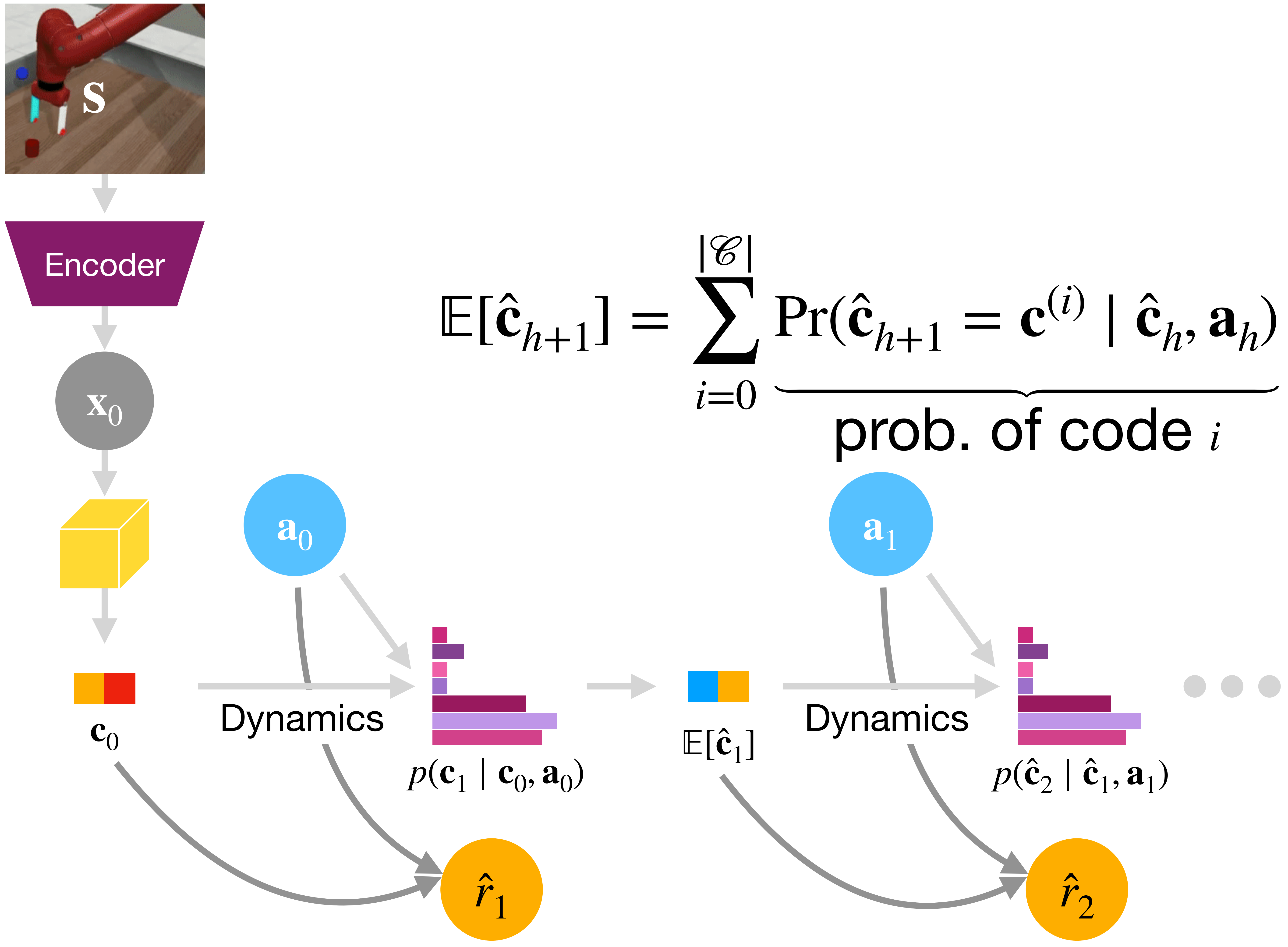
DC-MPC: Decision-time Planning

$$\mathbb{E}[\hat{\mathbf{c}}_{h+1}] = \sum_{i=0}^{|\mathscr{C}|} \underbrace{\Pr(\hat{\mathbf{c}}_{h+1} = \mathbf{c}^{(i)} \mid \hat{\mathbf{c}}_h, \mathbf{a}_h)}_{\text{prob. of code } i} \underbrace{\mathbf{c}^{(i)}}_{\text{code}}$$

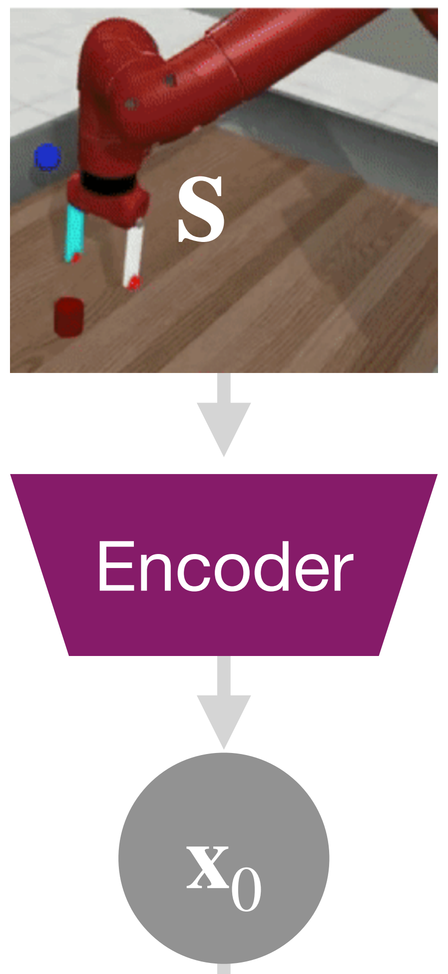


DC-MPC: Decision-time Planning

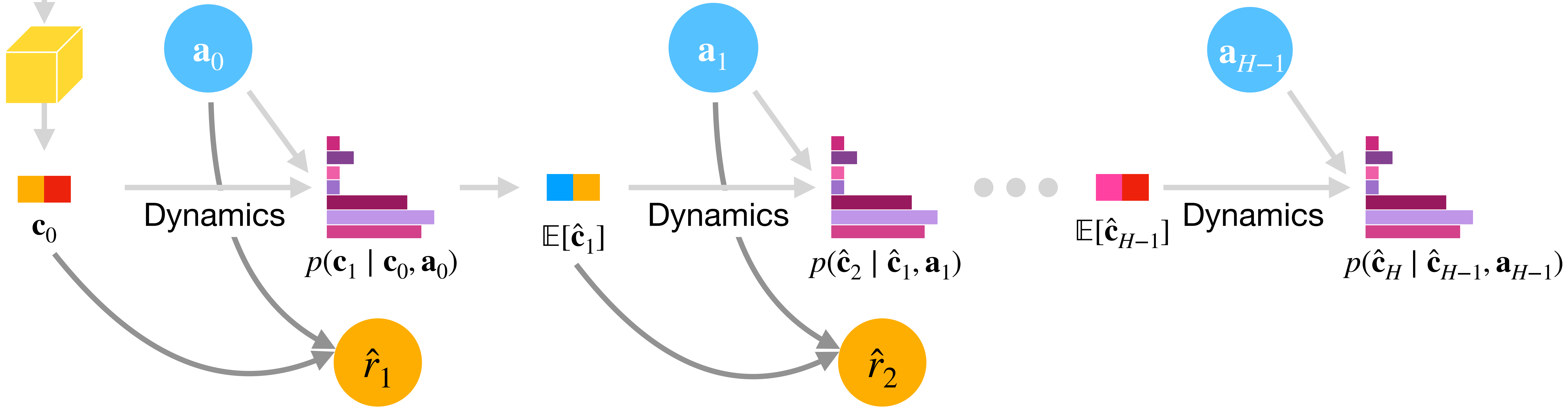
$$\mathbb{E}[\hat{\mathbf{c}}_{h+1}] = \sum_{i=0}^{|\mathscr{C}|} \underbrace{\Pr(\hat{\mathbf{c}}_{h+1} = \mathbf{c}^{(i)} \mid \hat{\mathbf{c}}_h, \mathbf{a}_h)}_{\text{prob. of code } i} \underbrace{\mathbf{c}^{(i)}}_{\text{code}}$$



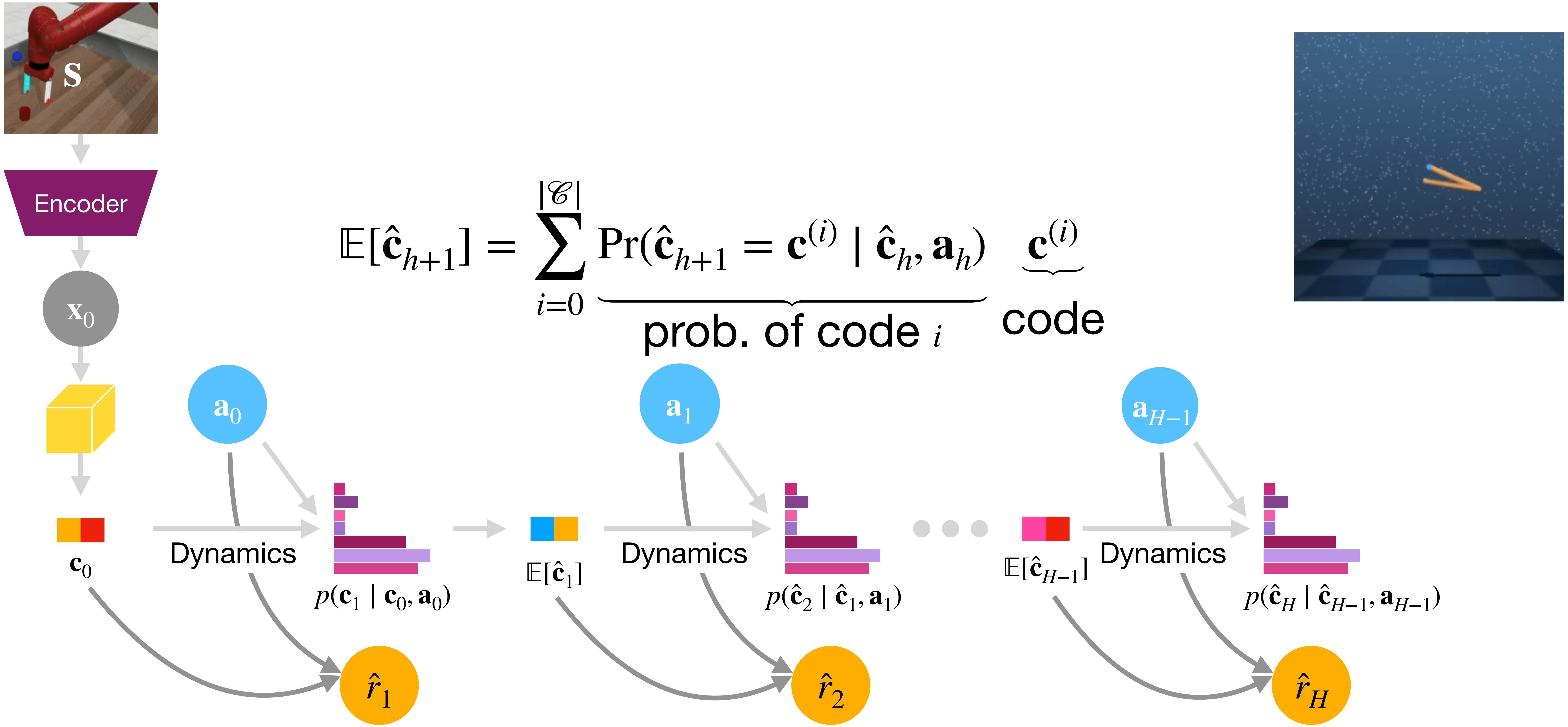
DC-MPC: Decision-time Planning



$$\mathbb{E}[\hat{\mathbf{c}}_{h+1}] = \sum_{i=0}^{|\mathscr{C}|} \underbrace{\Pr(\hat{\mathbf{c}}_{h+1} = \mathbf{c}^{(i)} \mid \hat{\mathbf{c}}_h, \mathbf{a}_h)}_{\text{prob. of code } i} \underbrace{\mathbf{c}^{(i)}}_{\text{code}}$$

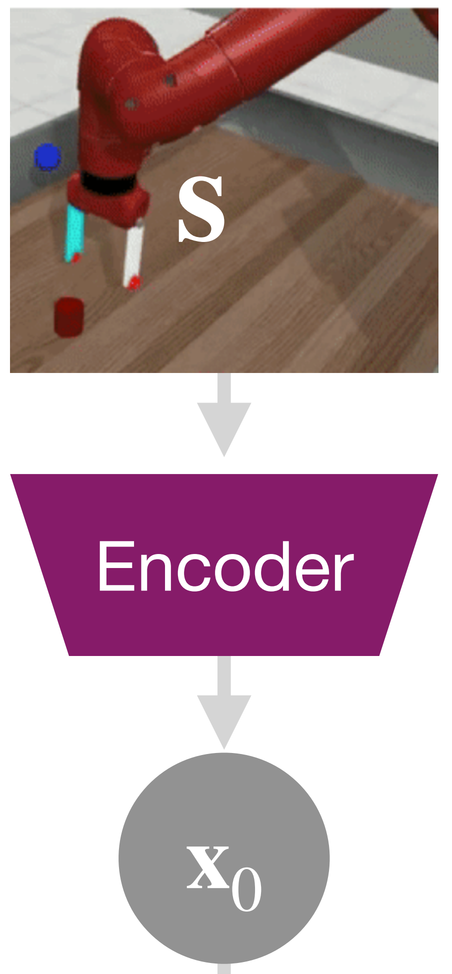


DC-MPC: Decision-time Planning



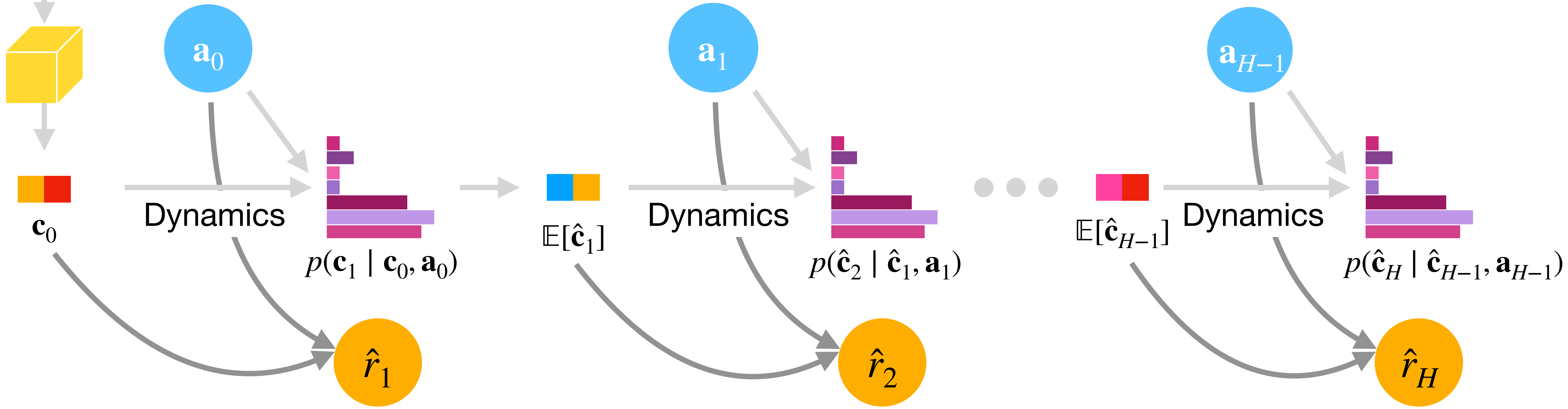
$$\mathbb{E}[\hat{\mathbf{c}}_{h+1}] = \sum_{i=0}^{|\mathscr{C}|} \underbrace{\Pr(\hat{\mathbf{c}}_{h+1} = \mathbf{c}^{(i)} \mid \hat{\mathbf{c}}_h, \mathbf{a}_h)}_{\text{prob. of code } i} \underbrace{\mathbf{c}^{(i)}}_{\text{code}}$$

DC-MPC: Decision-time Planning



$$J(\mathbf{a}_{0:H}, \mathbf{s}) = \gamma^H Q_\psi(\hat{\mathbf{c}}_H, \mathbf{a}_H) + \sum_{h=0}^{H-1} \gamma^h R_\xi(\hat{\mathbf{c}}_h, \mathbf{a}_h)$$

$$\mathbb{E}[\hat{\mathbf{c}}_{h+1}] = \sum_{i=0}^{|\mathscr{C}|} \underbrace{\Pr(\hat{\mathbf{c}}_{h+1} = \mathbf{c}^{(i)} \mid \hat{\mathbf{c}}_h, \mathbf{a}_h)}_{\text{prob. of code } i} \underbrace{\mathbf{c}^{(i)}}_{\text{code}}$$

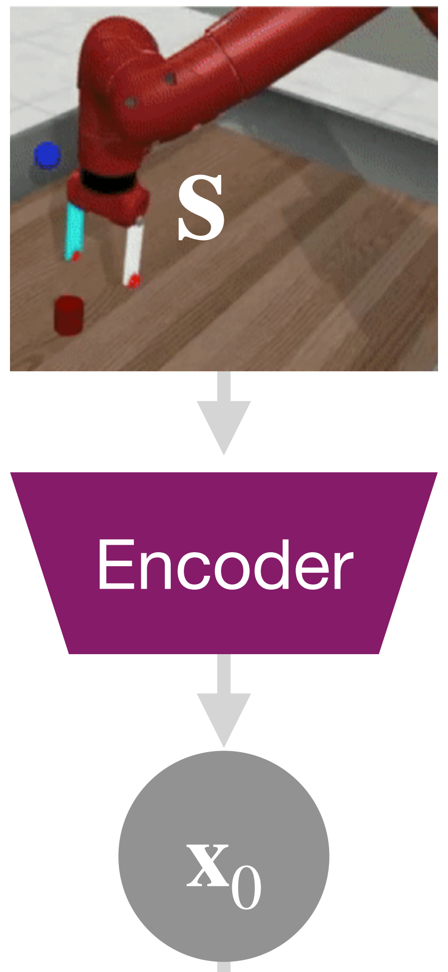


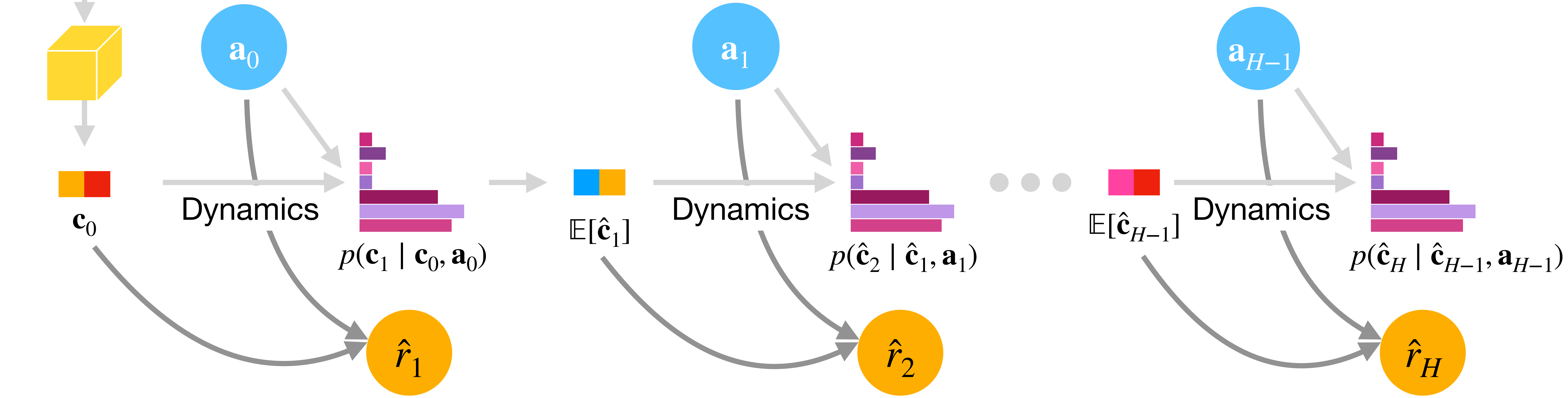
DC-MPC: Decision-time Planning

Reward func.

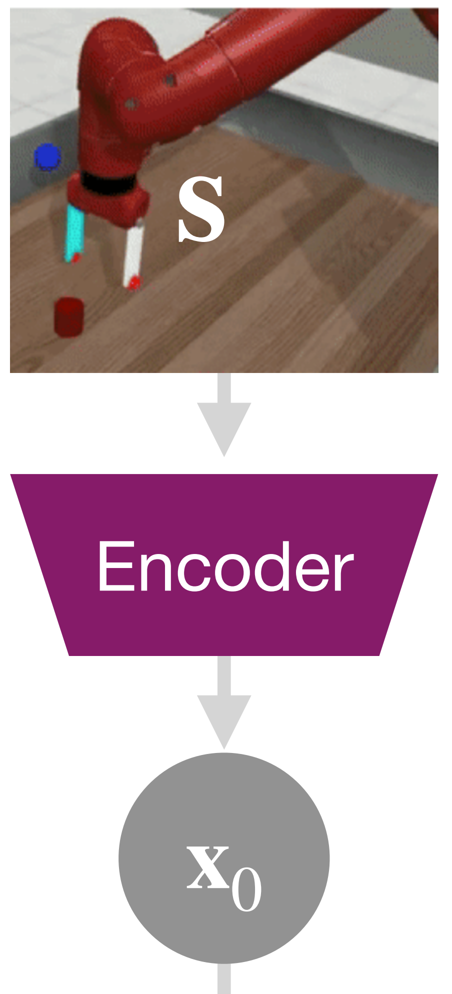
$$J(\mathbf{a}_{0:H}, \mathbf{s}) = \gamma^H Q_\psi(\hat{\mathbf{c}}_H, \mathbf{a}_H) + \boxed{\sum_{h=0}^{H-1} \gamma^h R_\xi(\hat{\mathbf{c}}_h, \mathbf{a}_h)}$$

$$\mathbb{E}[\hat{\mathbf{c}}_{h+1}] = \sum_{i=0}^{|\mathscr{C}|} \underbrace{\Pr(\hat{\mathbf{c}}_{h+1} = \mathbf{c}^{(i)} \mid \hat{\mathbf{c}}_h, \mathbf{a}_h)}_{\text{prob. of code } i} \; \underbrace{\mathbf{c}^{(i)}}_{\text{code}}$$



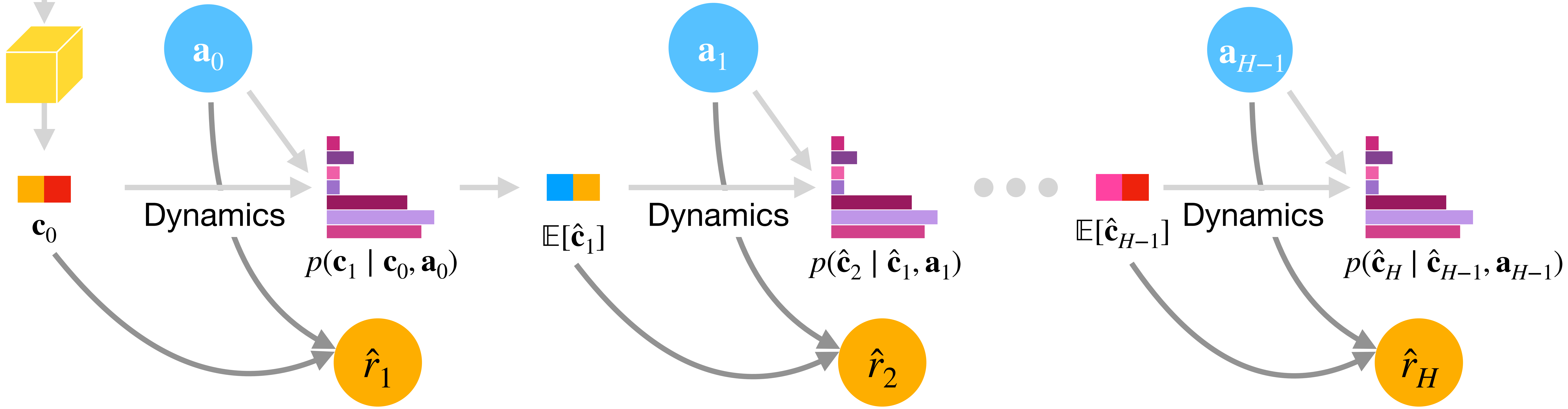


DC-MPC: Decision-time Planning



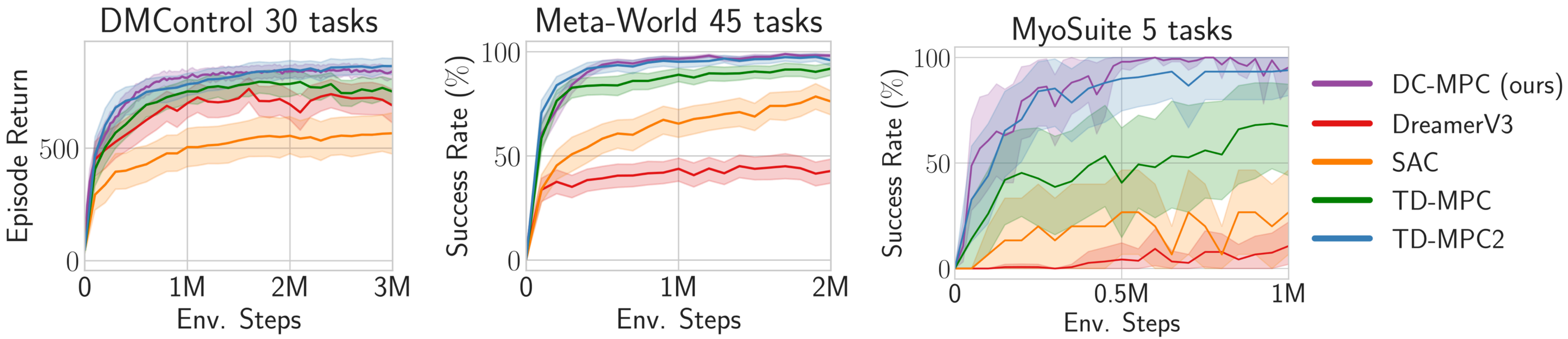
$$J(\mathbf{a}_{0:H}, \mathbf{s}) = \overbrace{\boxed{\gamma^H Q_\psi(\hat{\mathbf{c}}_H, \mathbf{a}_H)}}^{\text{Bootstrap with action-value}} + \overbrace{\boxed{\sum_{h=0}^{H-1} \gamma^h R_\xi(\hat{\mathbf{c}}_h, \mathbf{a}_h)}}^{\text{Reward func.}}$$

$$\mathbb{E}[\hat{\mathbf{c}}_{h+1}] = \sum_{i=0}^{|\mathscr{C}|} \underbrace{\Pr(\hat{\mathbf{c}}_{h+1} = \mathbf{c}^{(i)} \mid \hat{\mathbf{c}}_h, \mathbf{a}_h)}_{\text{prob. of code } i} \underbrace{\mathbf{c}^{(i)}}_{\text{code}}$$



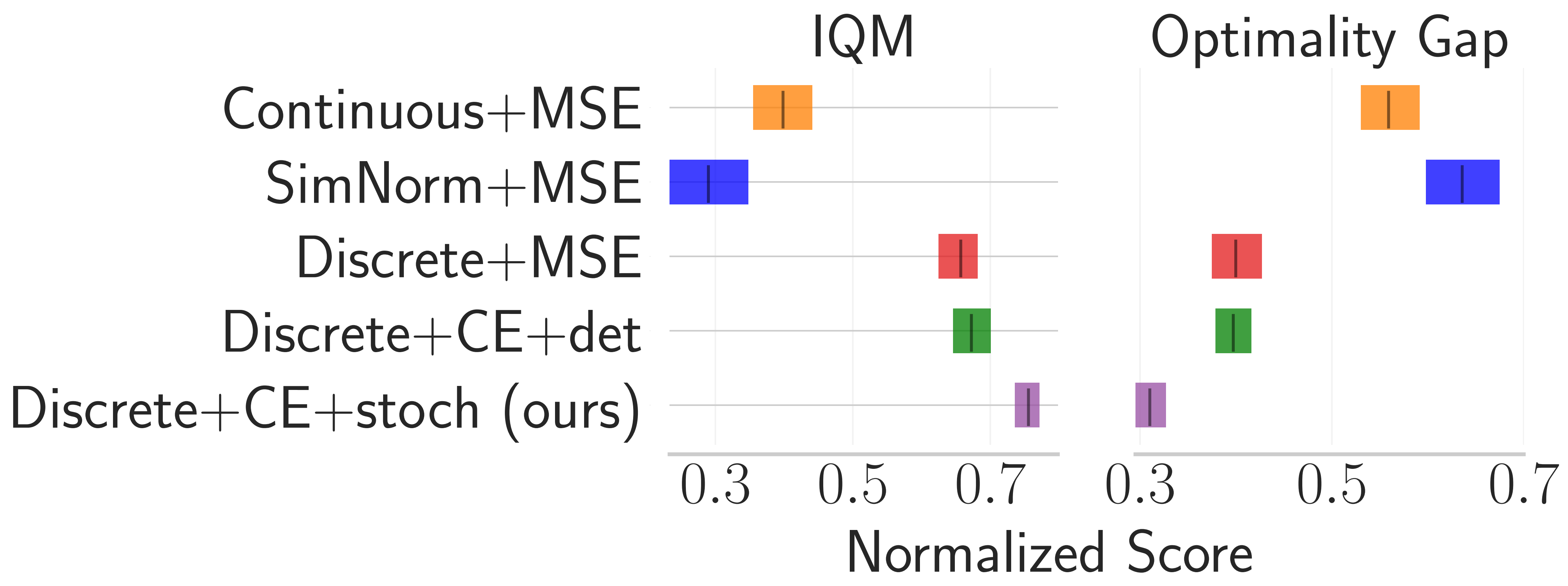
Results: Overview

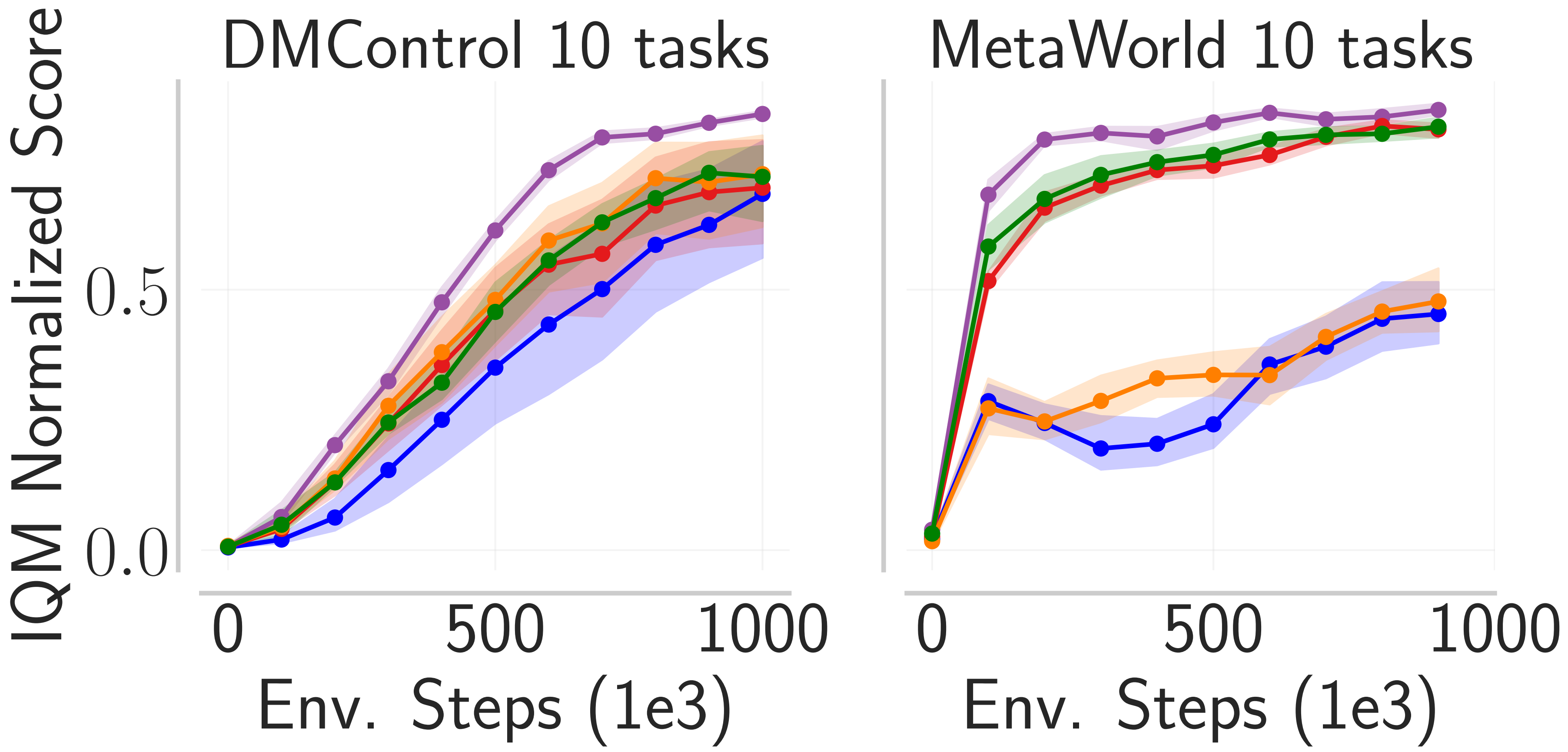
Strong Performance in DMControl, MetaWorld and MyoSuite Tasks



Why Does DC-MPC Work So Well?

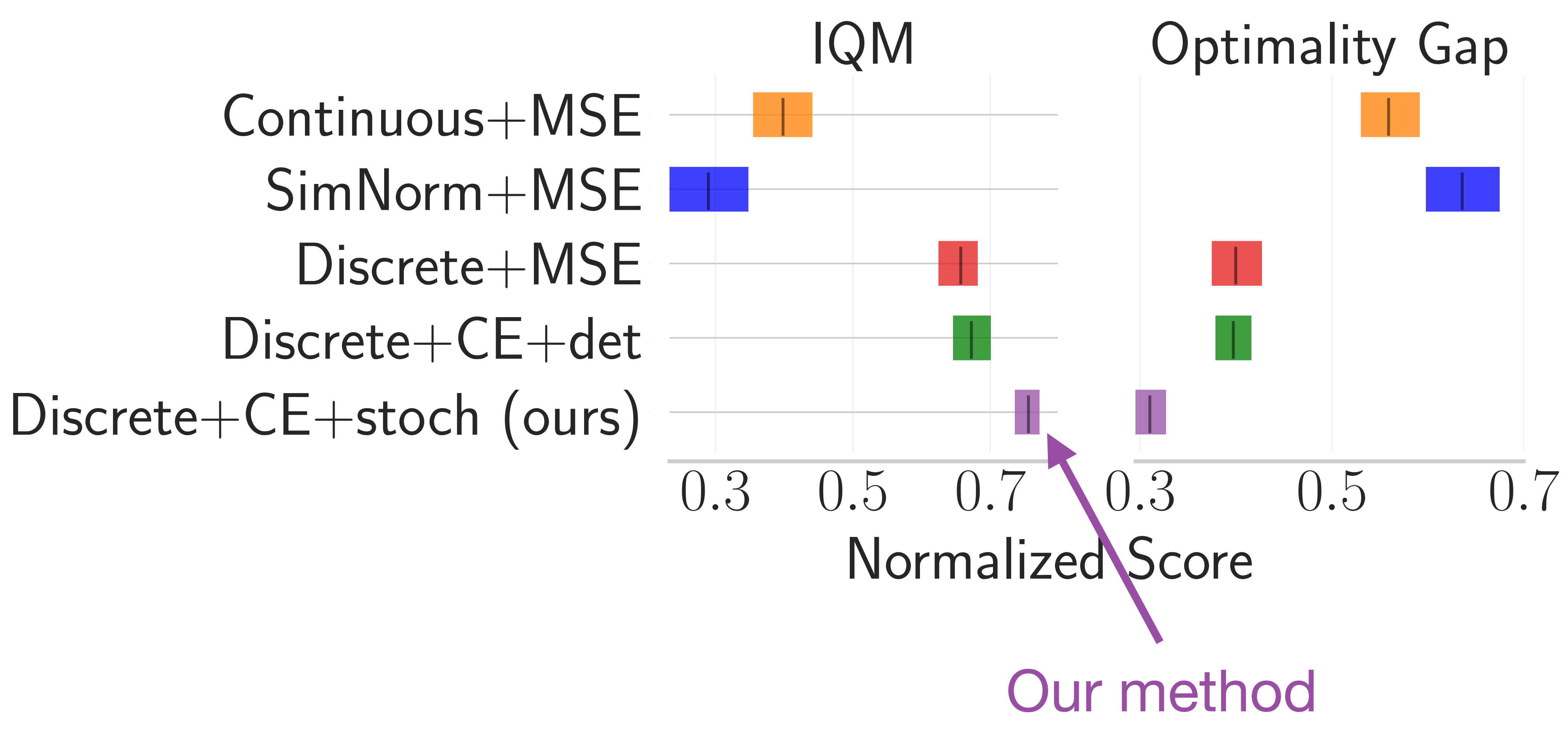
Combination of Discrete Codebook and Stochastic Dynamics

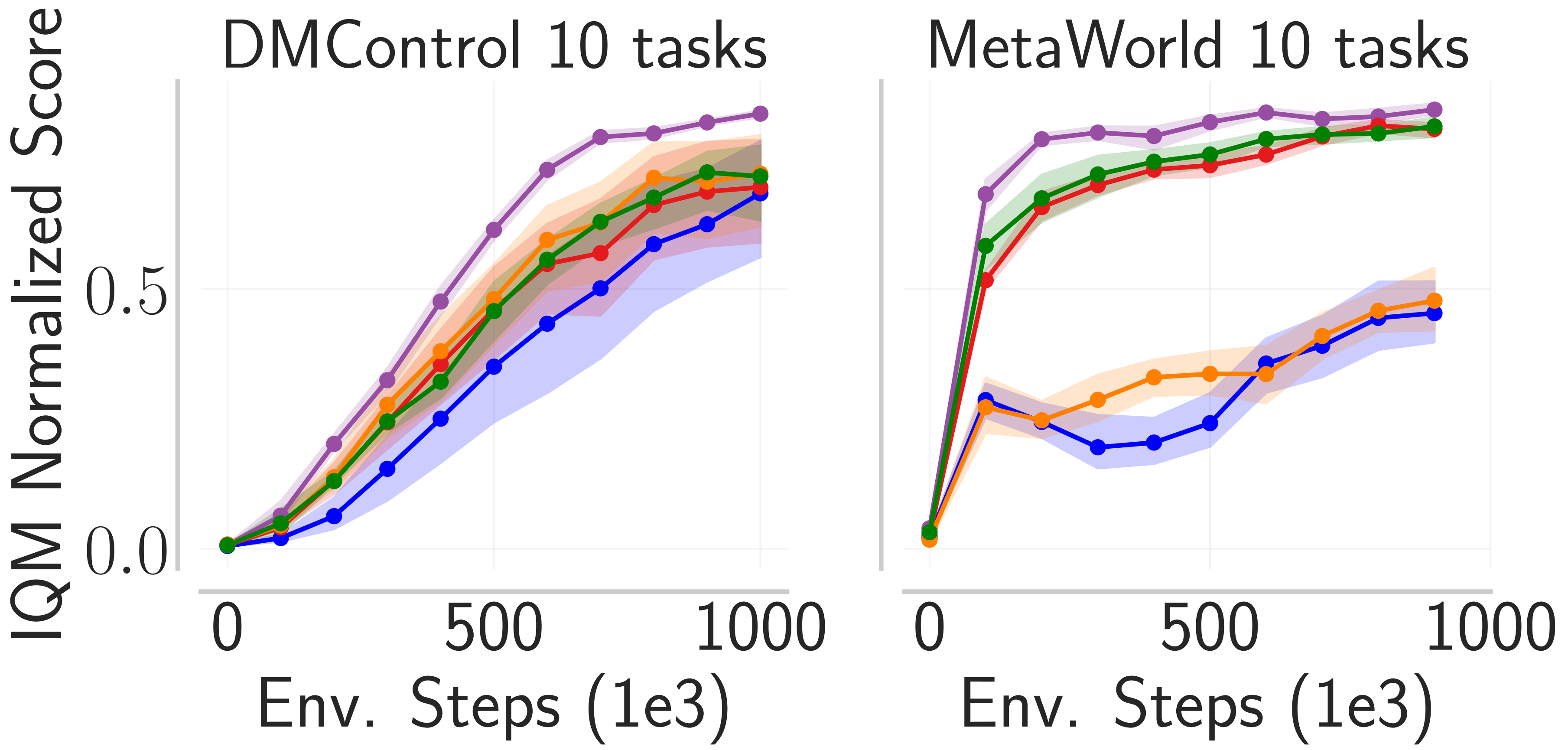




Why Does DC-MPC Work So Well?

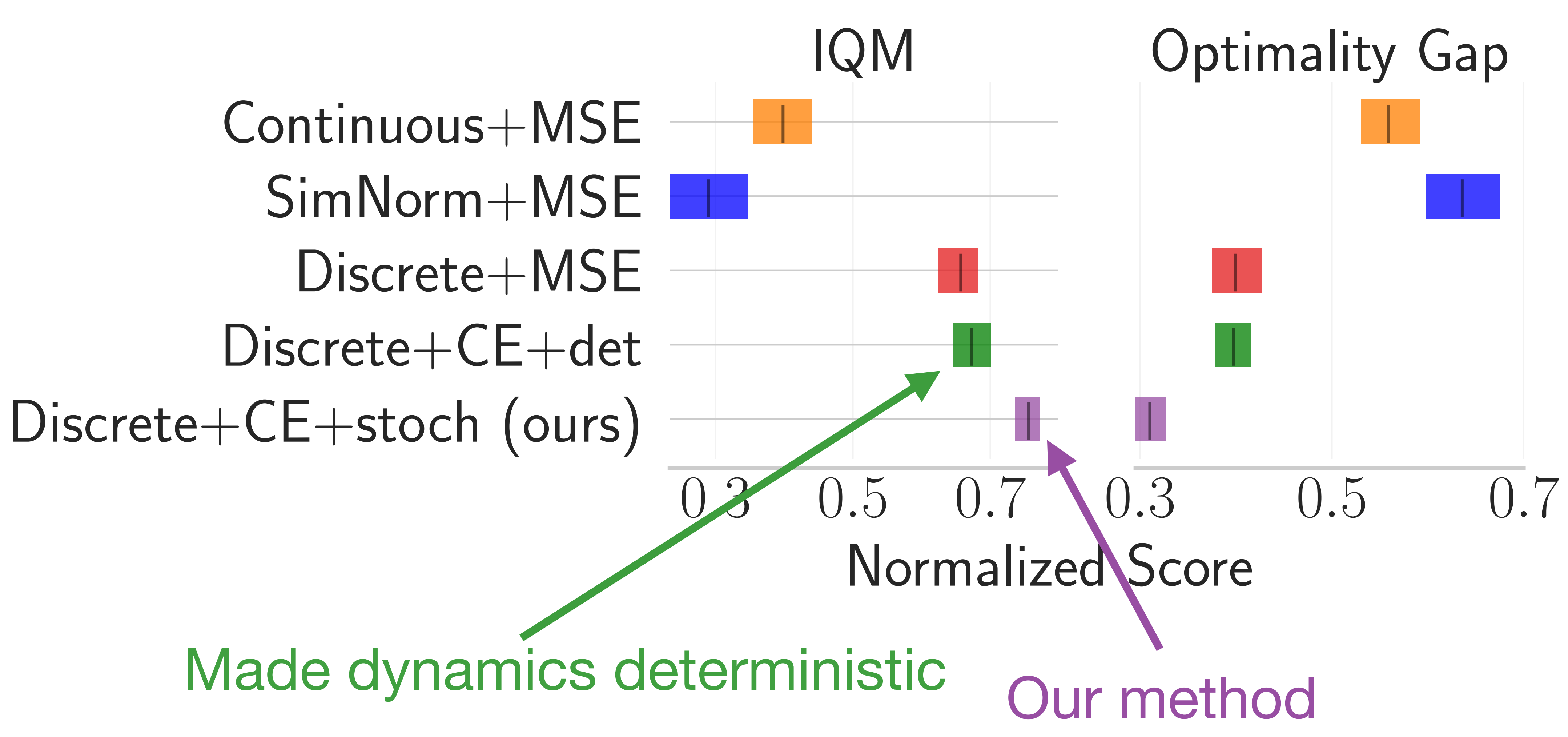
Combination of Discrete Codebook and Stochastic Dynamics

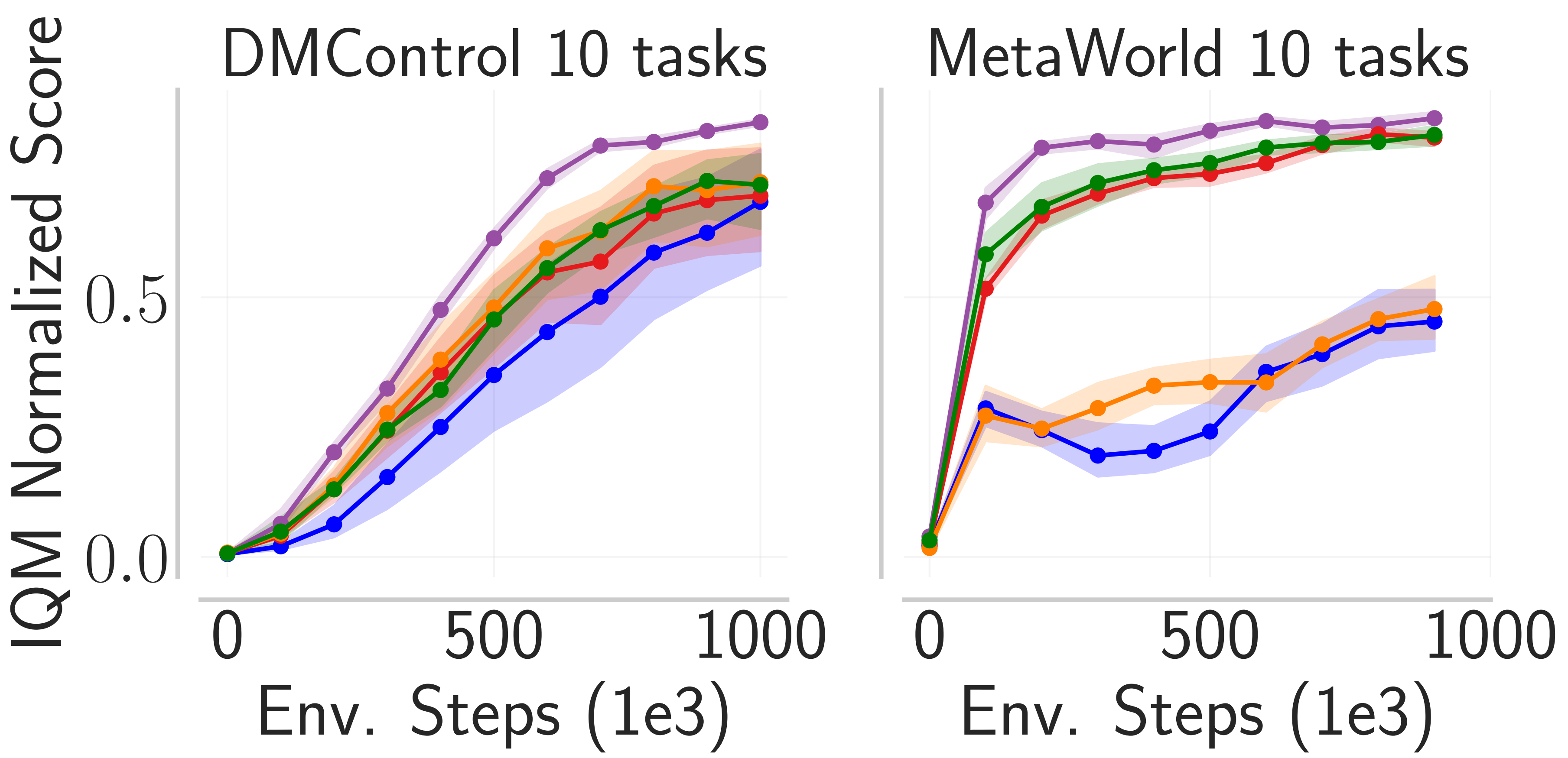




Why Does DC-MPC Work So Well?

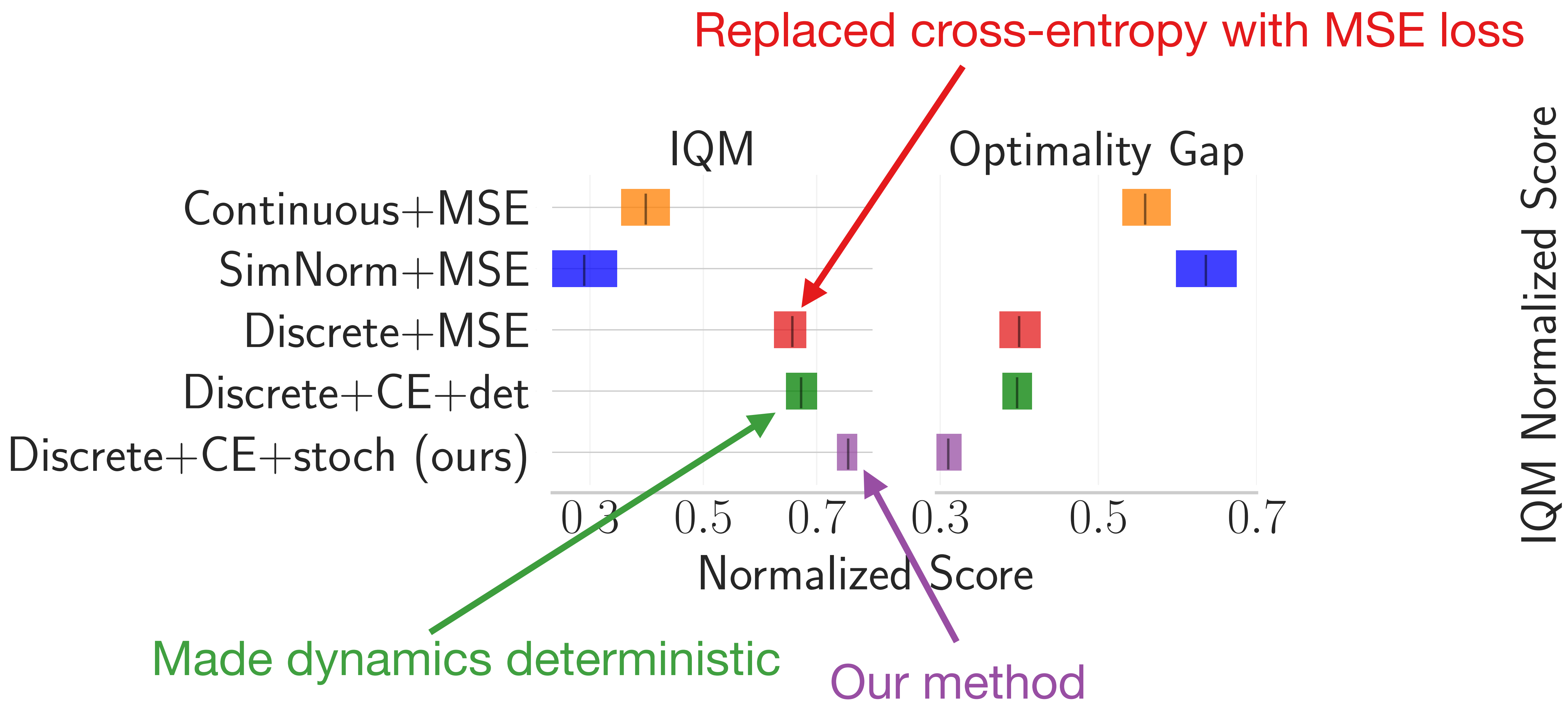
Combination of Discrete Codebook and Stochastic Dynamics

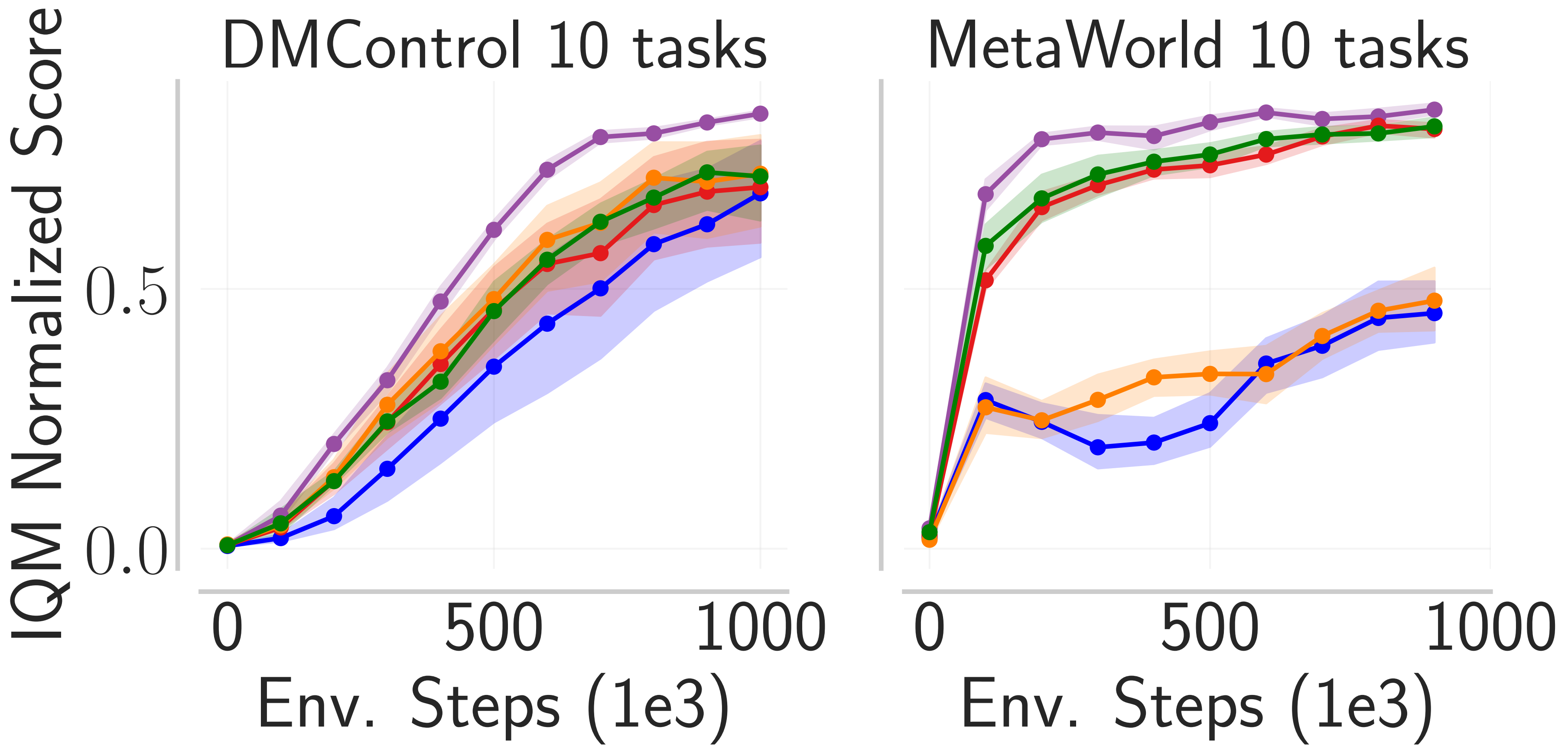




Why Does DC-MPC Work So Well?

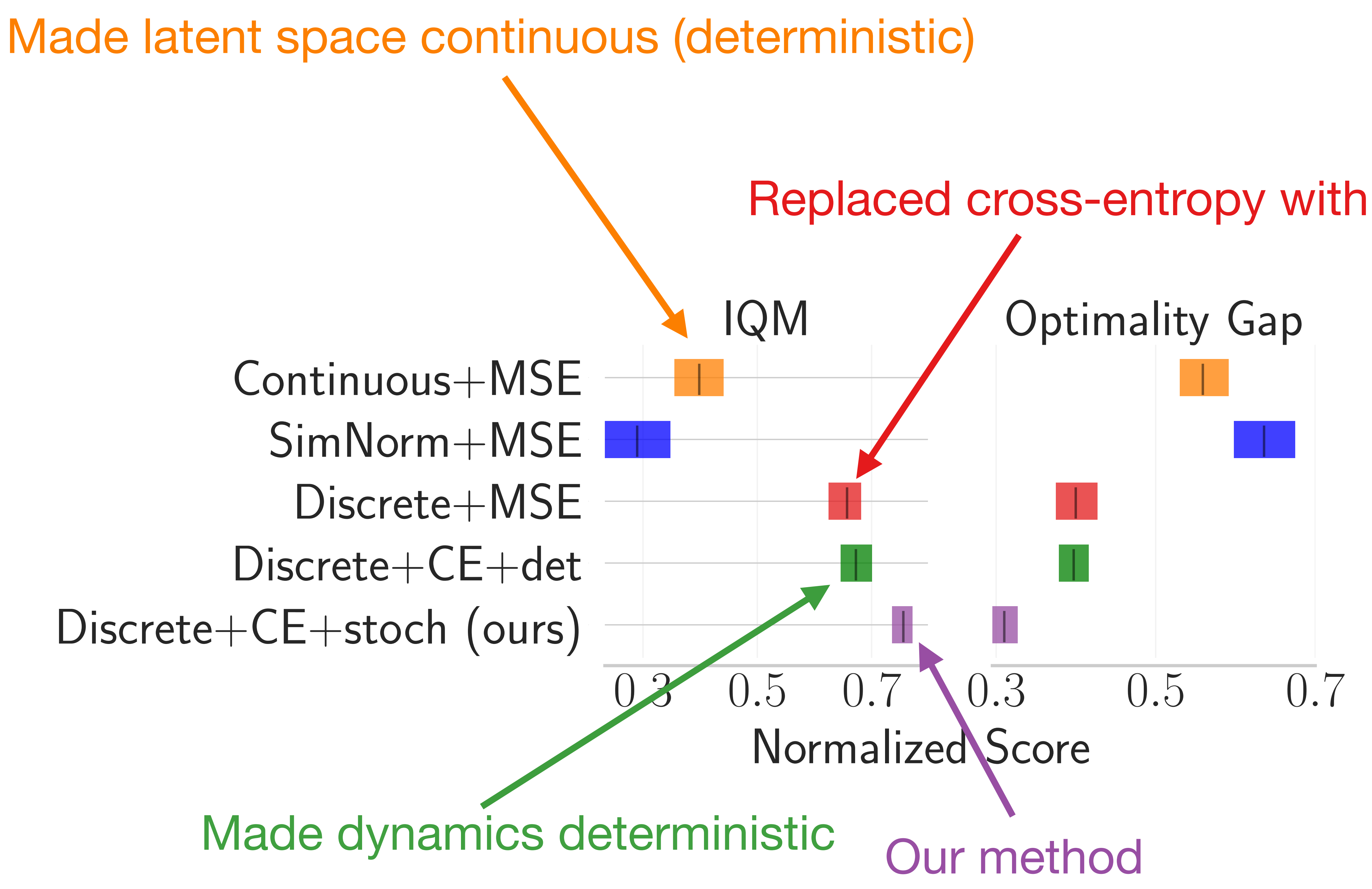
Combination of Discrete Codebook and Stochastic Dynamics

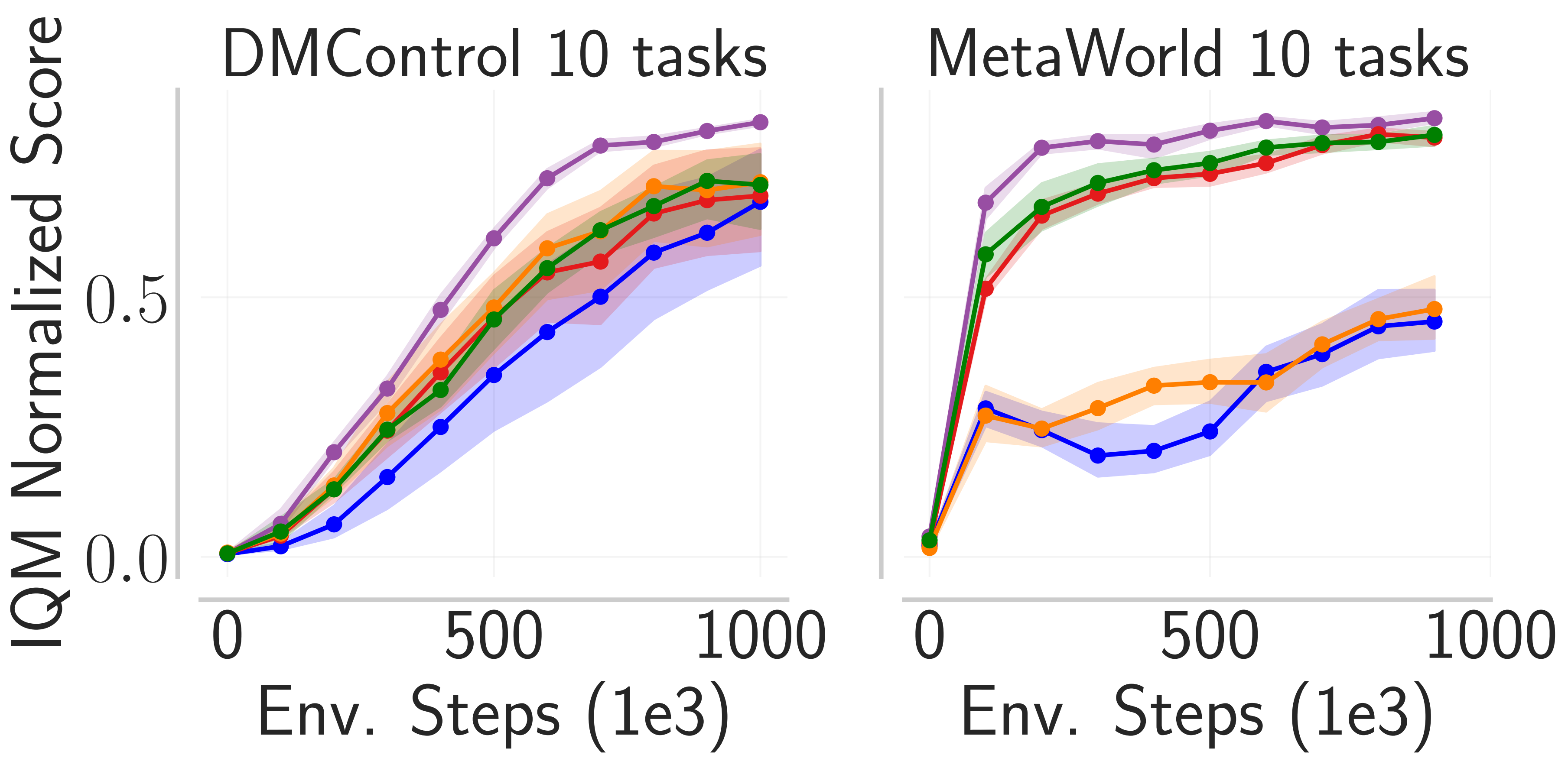




Why Does DC-MPC Work So Well?

Combination of Discrete Codebook and Stochastic Dynamics





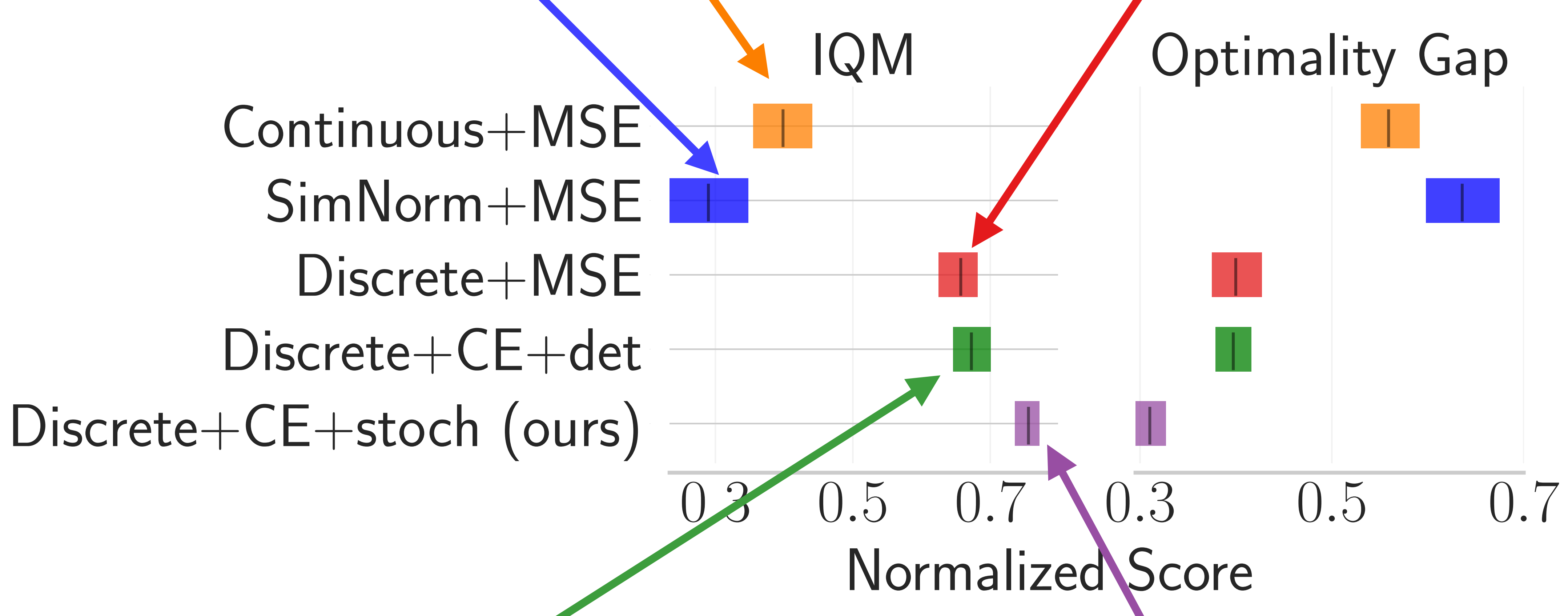
Why Does DC-MPC Work So Well?

Combination of Discrete Codebook and Stochastic Dynamics

Made latent space continuous (deterministic)

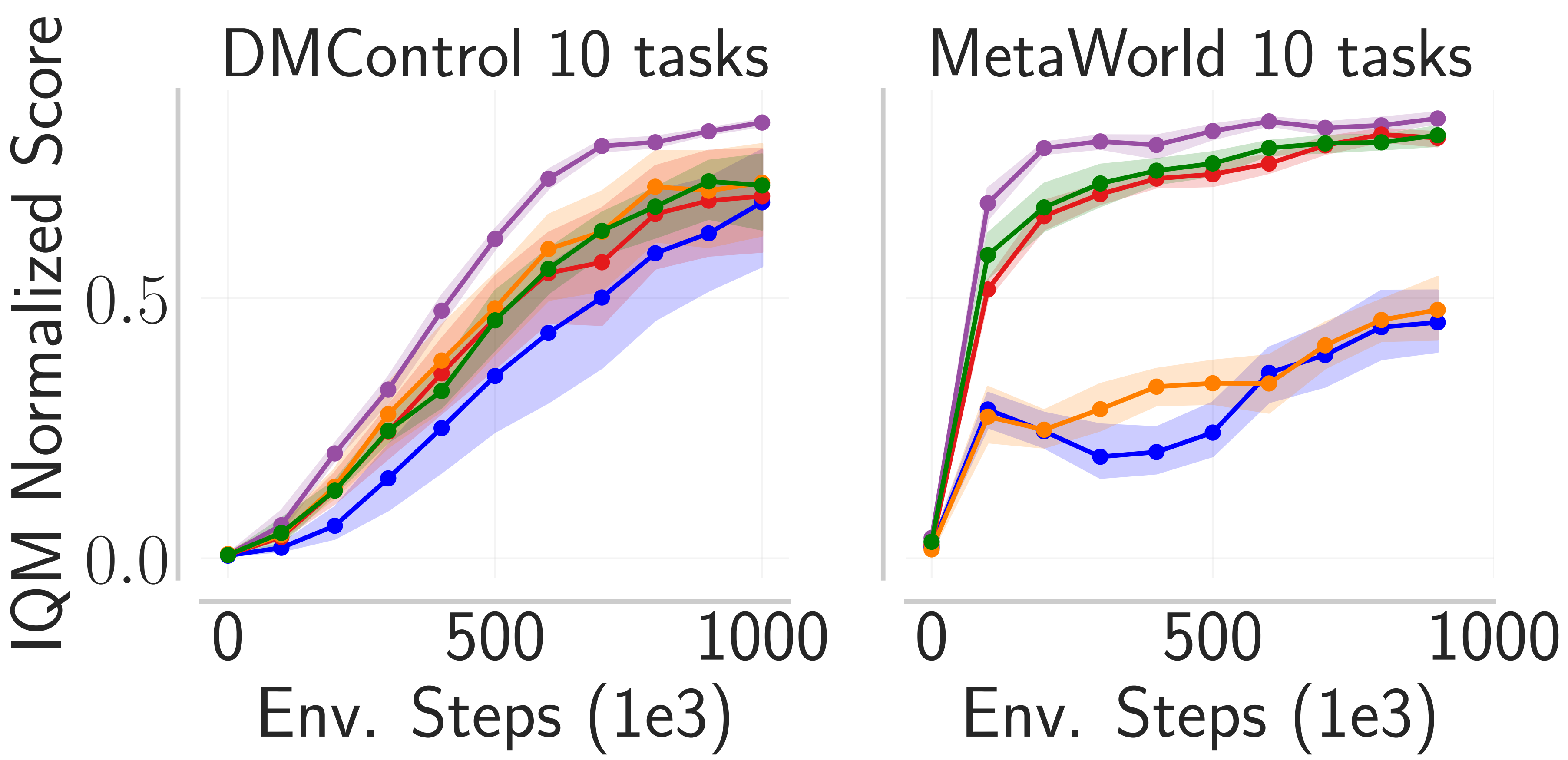
Bounded & continuous

Replaced cross-entropy with MSE loss



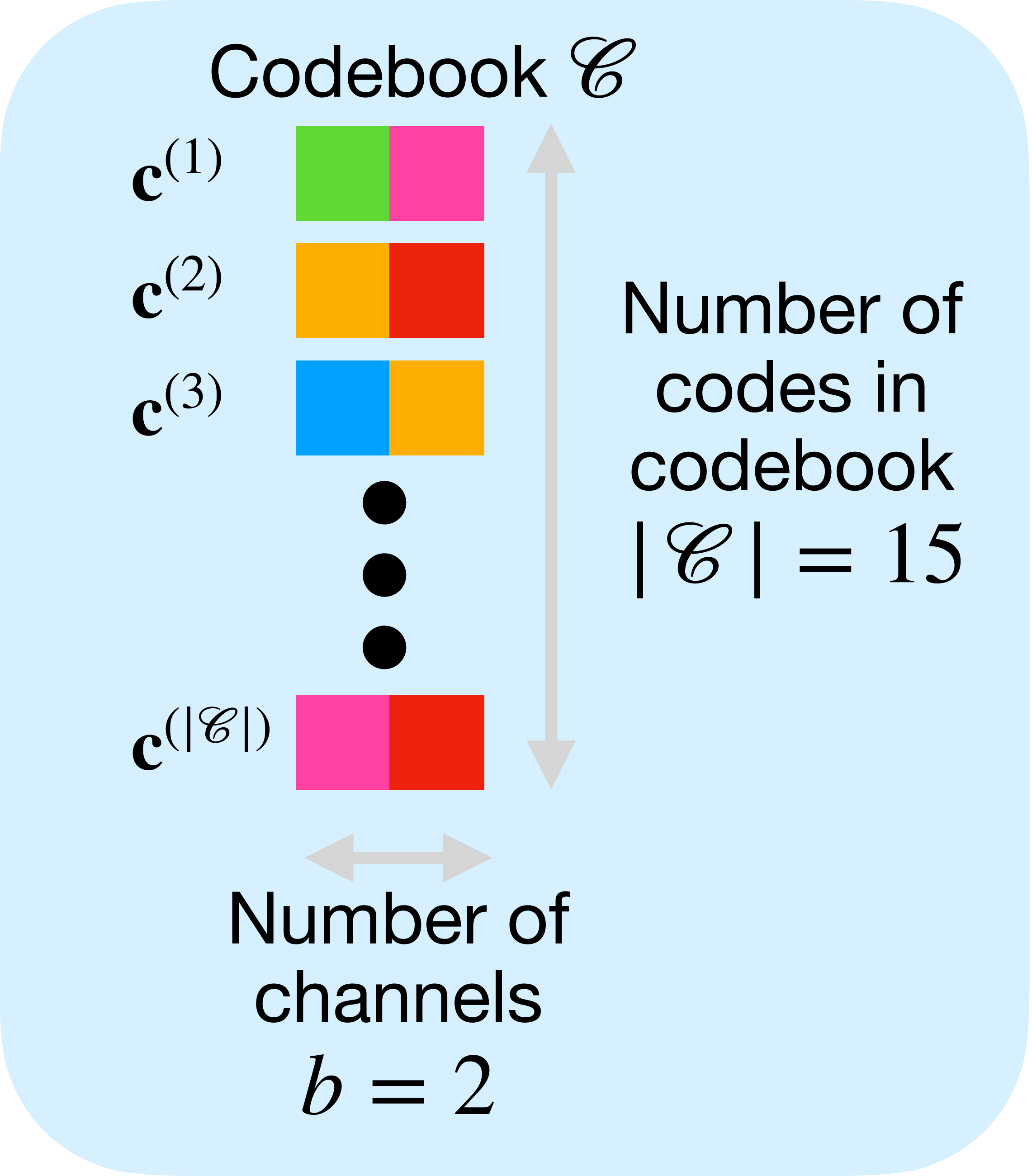
Made dynamics deterministic

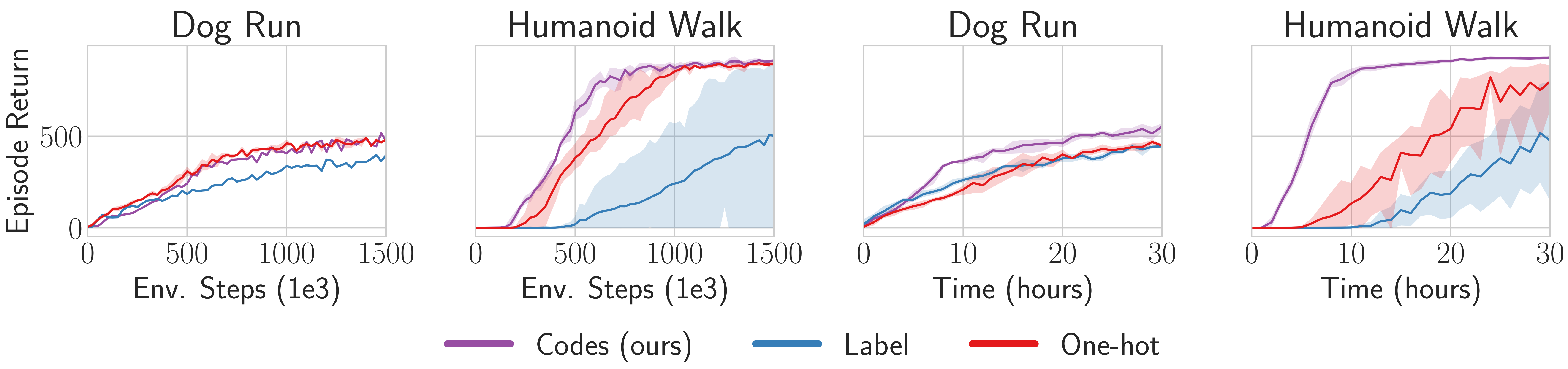
Our method



Comparison of Different Discrete Encodings

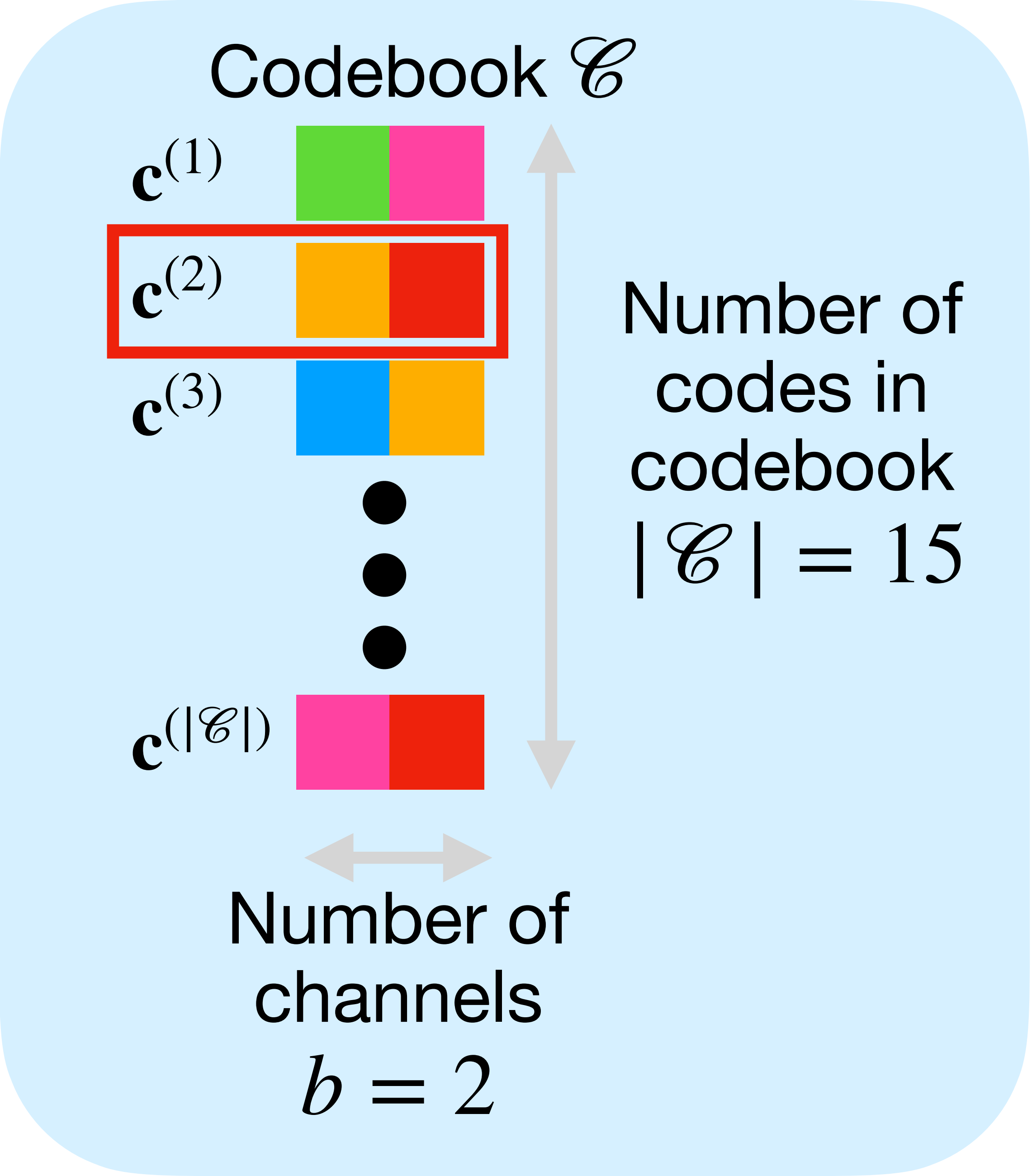
Codebook > One-hot > Label

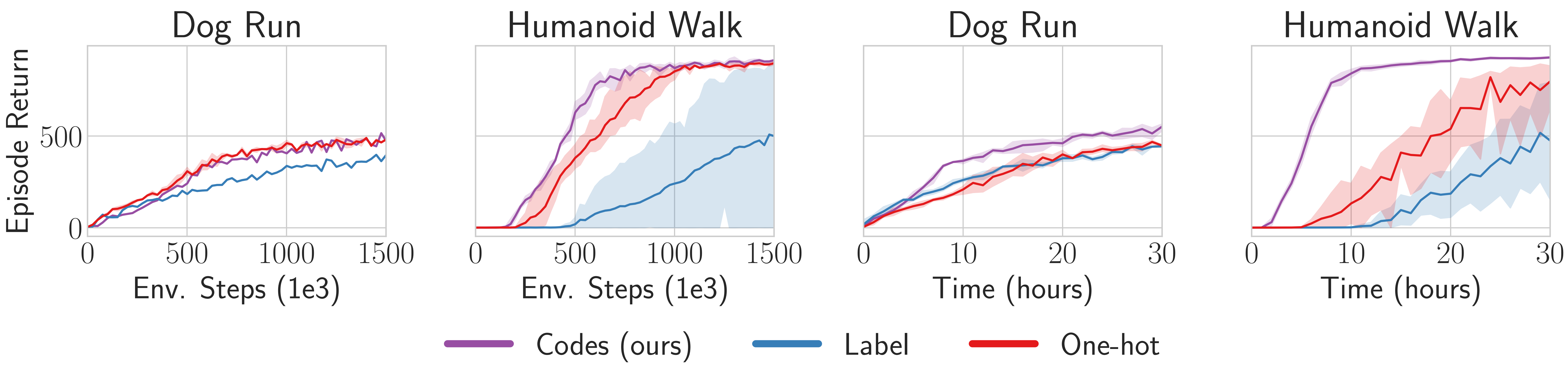




Comparison of Different Discrete Encodings

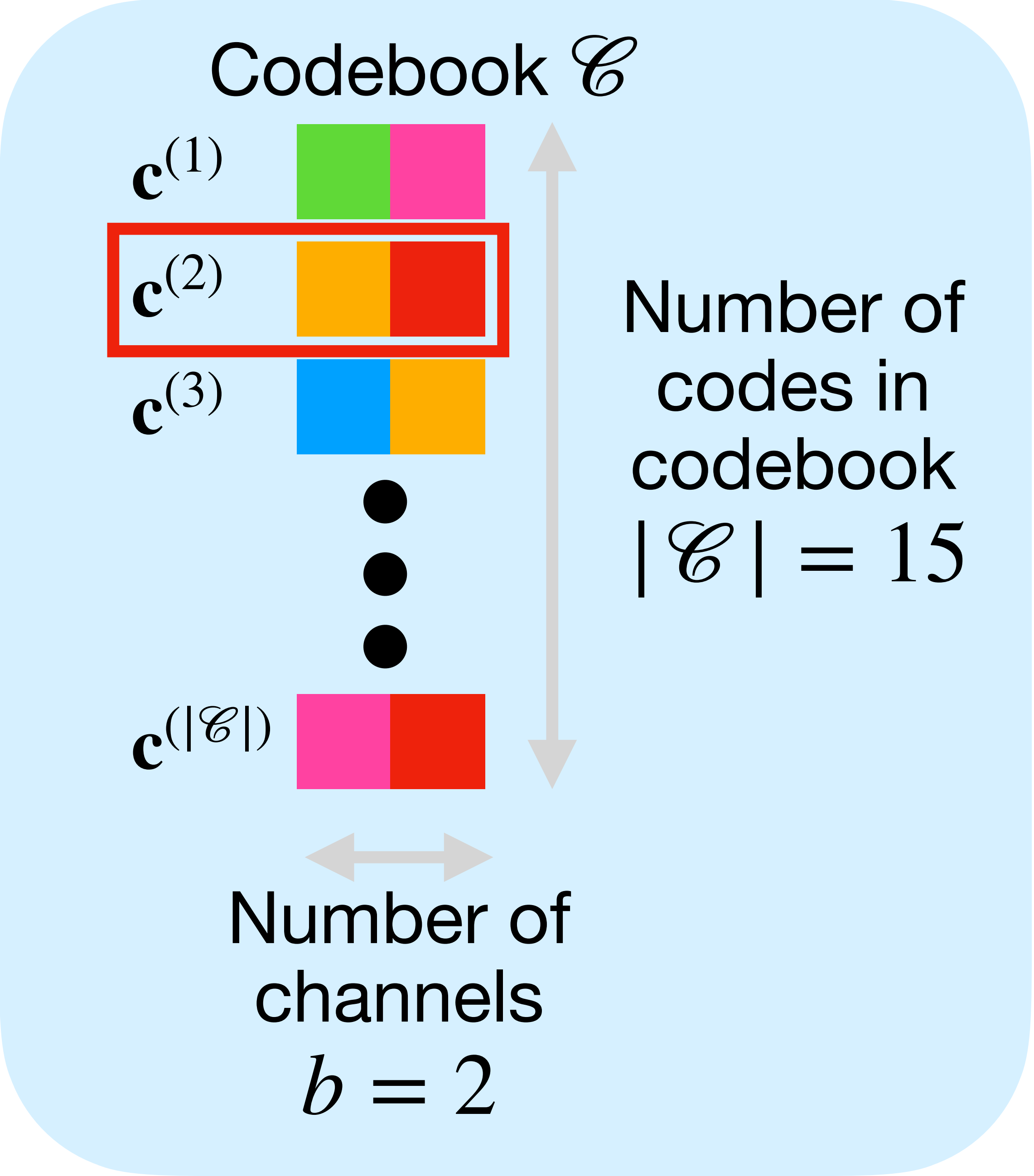
Codebook > One-hot > Label



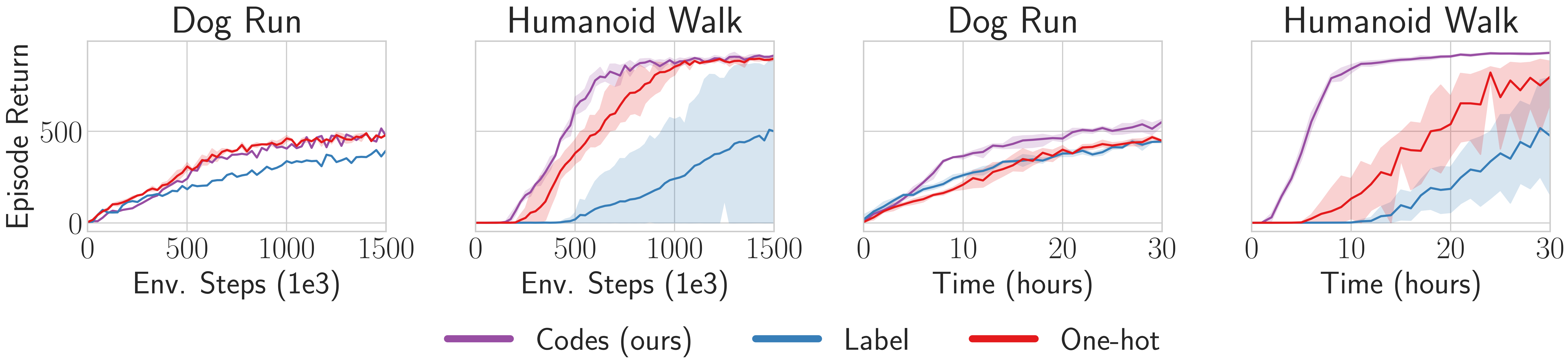


Comparison of Different Discrete Encodings

Codebook > One-hot > Label

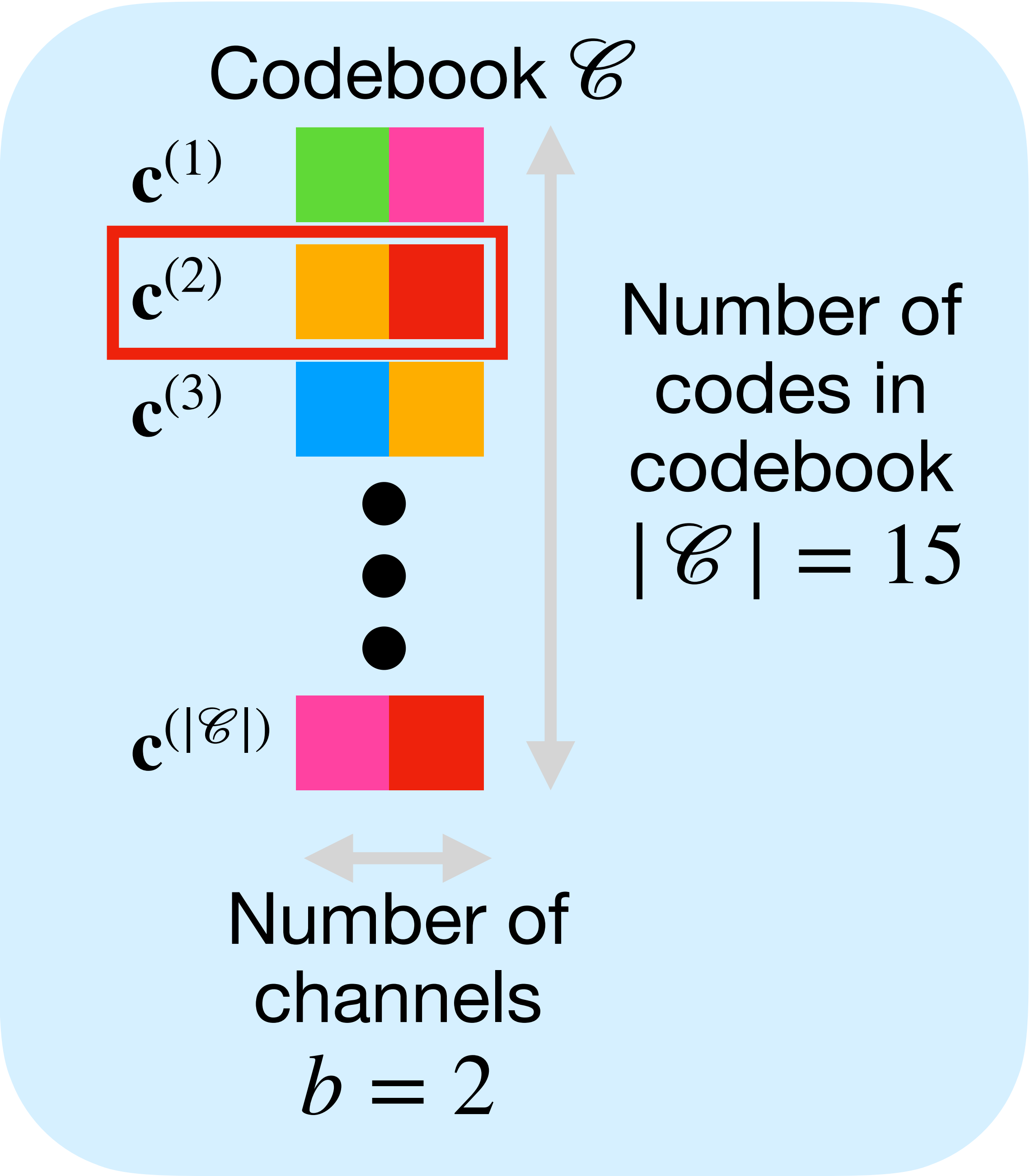


$\mathbf{e}_{\text{code}} = \mathbf{c}^{(2)} = \{-0.5, 1\}$



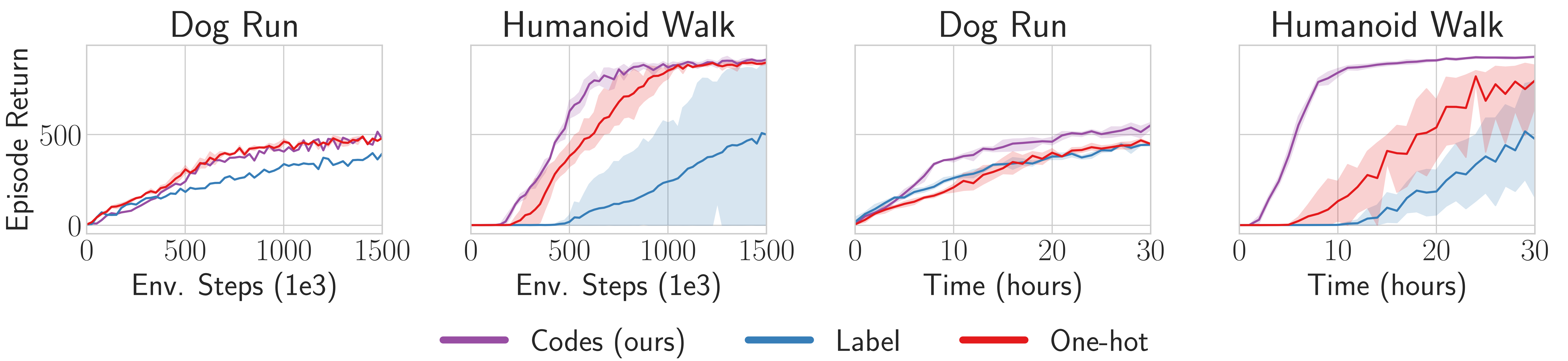
Comparison of Different Discrete Encodings

Codebook > One-hot > Label



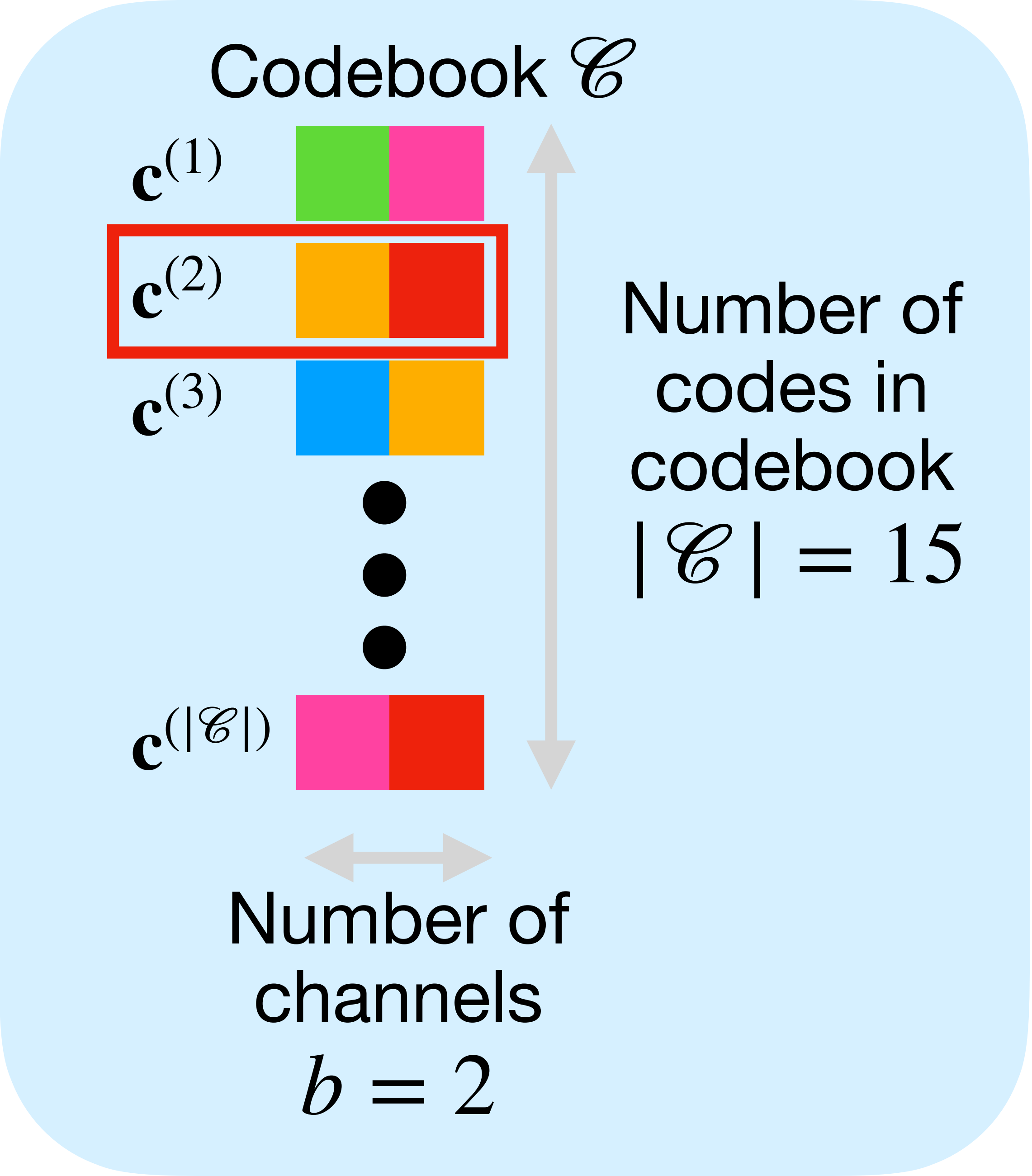
$\mathbf{e}_{\text{code}} = \mathbf{c}^{(2)} = \{-0.5, 1\}$

$\mathbf{e}_{\text{label}} = 2$



Comparison of Different Discrete Encodings

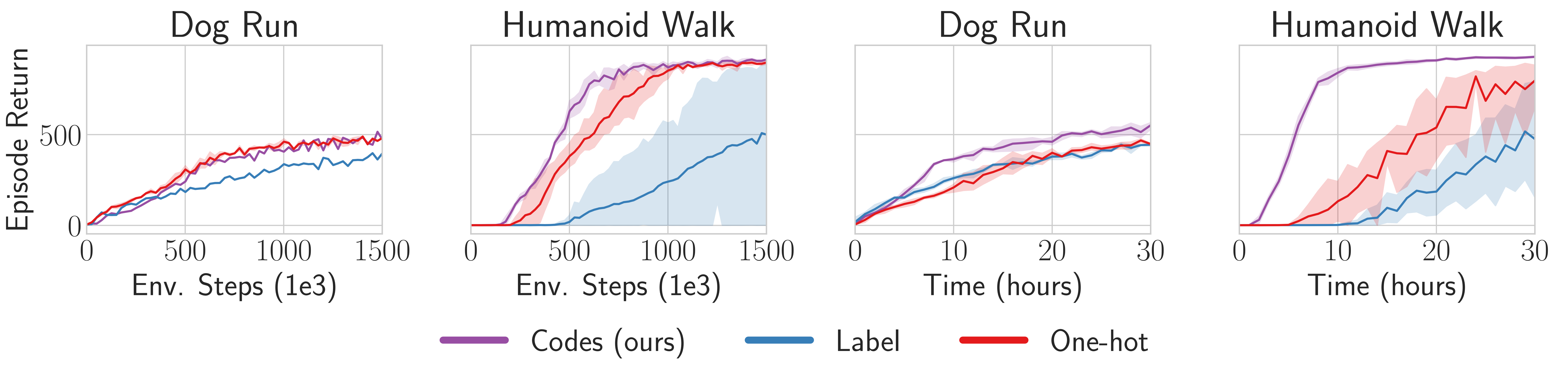
Codebook > One-hot > Label



$$\mathbf{e}_{\text{code}} = \mathbf{c}^{(2)} = \{-0.5, 1\}$$

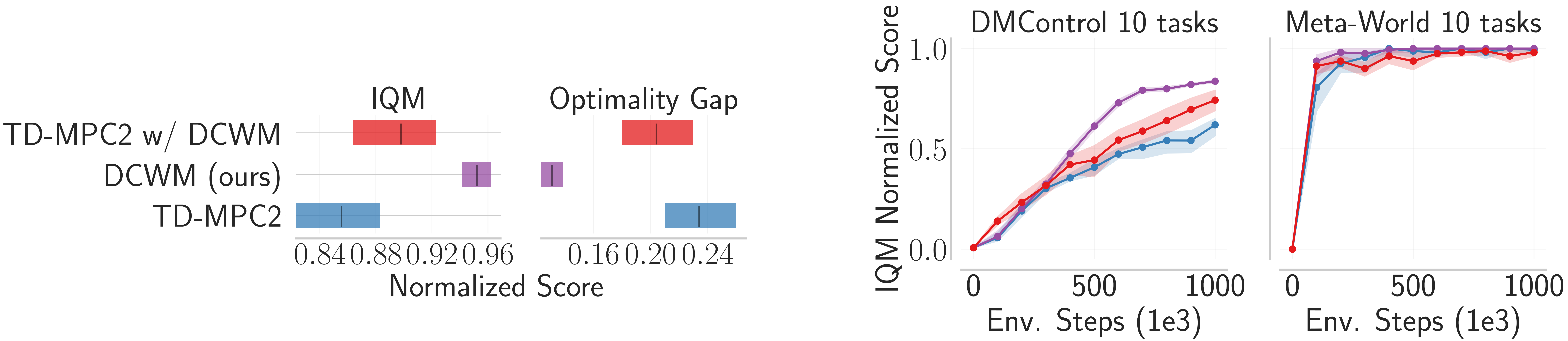
$$\mathbf{e}_{\text{label}} = 2$$

$$\mathbf{e}_{\text{one-hot}} = \{0,1,0,0,0,0,0,0,0,0,0,0,0,0,0\}$$



Discrete Codebooks Improve TD-MPC2

Combination of Discrete Codebook and Stochastic Dynamics



Email: ascannel@ed.ac.uk

Website: www.aidanscannell.com

Email: ascannel@ed.ac.uk

Website: www.aidanscannell.com

Main Takeaways:

Email: ascannel@ed.ac.uk

Website: www.aidanscannell.com

Main Takeaways:

Email: ascannel@ed.ac.uk

Website: www.aidanscannell.com

Main Takeaways:

- ***Discrete latent space better than continuous***

Email: ascannel@ed.ac.uk

Website: www.aidanscannell.com

Main Takeaways:

- *Discrete latent space better than continuous*
- *Codebook encodings > one-hot encodings > label encodings*

Email: ascannel@ed.ac.uk

Website: www.aidanscannell.com

Main Takeaways:

- *Discrete latent space better than continuous*
- *Codebook encodings > one-hot encodings > label encodings*
- *Stochastic dynamics better than deterministic*