

Appendix

Simulation for Evaluating Robot Policies Should Be *Asynchronous* and *Real-Time*

Website: <https://anonymous.4open.science/w/async-robosim>

397 A Robomimic full results

398 We provide detailed per-policy results on the Robomimic benchmark [3] across all five tasks: **Lift**,
 399 **Can**, **Square**, **Transport**, and **ToolHang** in Table 3. All values are success rates (%) evaluated
 400 over 400 episodes under synchronous and asynchronous settings. 95% Wilson score confidence
 401 intervals [35] are reported in brackets.

		Task Success Rate (%)						Perf. drop	Rel. drop (%)	Latency
		Sync / Async								
		Lift	Can	Square	Trans.	T. Hang	Avg.			↓
1	BC	86 / 80	31 / 28	19 / 14	0 / 0	0 / 0	27.2 / 24.4	-2.8	10.29	11.5 ms
2	BC-VAE	82 / 77	87 / 72	28 / 24	0 / 0	0 / 0	39.4 / 34.6	-4.8	12.18	11.7 ms
3	BC-Gaussian	60 / 50	38 / 54	23 / 19	0 / 0	0 / 0	24.2 / 24.6	+0.4	-1.65	12.4 ms
4	BC-GMM	100 / 100	94 / 93	54 / 46	52 / 4	0 / 0	60.0 / 48.6	-11.4	19.00	13.0 ms
5	BC-TF	100 / 100	98 / 88	90 / 72	33 / 2	0 / 0	64.2 / 52.4	-11.8	18.38	21.9 ms
6	BC-TF-GMM	100 / 100	98 / 90	83 / 62	78 / 3	0 / 0	71.8 / 51.0	-20.8	28.97	22.1 ms
7	DP (chunk=8)	100 / 100	96 / 61	86 / 52	81 / 6	30 / 29	78.6 / 49.6	-29.0	36.90	73.36 ms
8	DP (chunk=1)	100 / 45	95 / 32	81 / 33	76 / 0	31 / 0	76.6 / 22.0	-54.6	71.28	597.80 ms
9	DP (chunk=4)	100 / 98	97 / 69	78 / 68	82 / 2	35 / 26	78.4 / 52.6	-25.8	32.91	150.05 ms
10	DP (chunk=16)	100 / 100	98 / 42	84 / 54	83 / 5	46 / 32	82.2 / 46.6	-35.6	43.31	37.73 ms
11	DP (chunk=2)	100 / 72	97 / 10	83 / 30	82 / 0	33 / 2	79.0 / 22.8	-56.2	71.14	302.40 ms

Table 2: Synchronous and asynchronous simulation performance on Robomimic [3].

Policy	Sync	Async	Δ	Inference Latency
Can				
BC-TF	98 [96.1, 99.0]	88 [84.4, 90.8]	10	55.76 ms
BC-TF-GMM	98 [96.1, 99.0]	90 [86.7, 92.6]	8	56.01 ms
BC-Gaussian	38 [33.4, 42.8]	54 [49.1, 58.8]	-16	26.07 ms
BC	31 [26.7, 35.7]	28 [23.8, 32.6]	3	22.76 ms
BC-GMM	94 [91.2, 95.9]	93 [90.1, 95.1]	1	25.94 ms
BC-VAE	87 [83.3, 89.9]	72 [67.4, 76.2]	15	23.27 ms
DP (chunk=16)	98 [96.1, 99.0]	42 [37.3, 46.9]	56	574.39 ms
DP (chunk=8)	96 [93.6, 97.5]	61 [56.1, 65.7]	35	534.97 ms
DP (chunk=4)	97 [94.8, 98.3]	69 [64.3, 73.3]	28	602.67 ms
DP (chunk=2)	97 [94.8, 98.3]	10 [7.4, 13.3]	87	623.06 ms
DP (chunk=1)	95 [92.4, 96.7]	32 [27.6, 36.7]	63	587.53 ms
Square				
BC-TF	90 [86.7, 92.6]	72 [67.4, 76.2]	18	12.34 ms
BC-TF-GMM	83 [79.0, 86.4]	62 [57.2, 66.6]	21	12.38 ms
BC-Gaussian	23 [19.1, 27.4]	19 [15.5, 23.1]	4	7.13 ms
BC	19 [15.5, 23.1]	14 [10.9, 17.7]	5	6.72 ms
BC-GMM	54 [49.1, 58.8]	46 [41.2, 50.9]	8	7.75 ms

Continued on next page

Policy	Sync	Async	Δ	Inference Latency
BC-VAE	28 [23.8, 32.6]	24 [20.1, 28.4]	4	6.79 ms
DP (chunk=16)	84 [80.1, 87.3]	54 [49.1, 58.8]	30	595.73 ms
DP (chunk=8)	86 [82.3, 89.1]	52 [47.1, 56.9]	34	587.12 ms
DP (chunk=4)	78 [73.7, 81.8]	68 [63.3, 72.4]	10	602.23 ms
DP (chunk=2)	83 [79.0, 86.4]	30 [25.7, 34.7]	53	589.04 ms
DP (chunk=1)	81 [76.9, 84.5]	33 [28.6, 37.8]	48	594.50 ms
Lift				
BC-TF	100 [99.0, 100.0]	100 [99.0, 100.0]	0	11.82 ms
BC-TF-GMM	100 [99.0, 100.0]	100 [99.0, 100.0]	0	11.76 ms
BC-Gaussian	60 [55.1, 64.7]	50 [45.1, 54.9]	10	7.21 ms
BC	86 [82.3, 89.1]	80 [75.8, 83.6]	6	6.85 ms
BC-GMM	100 [99.0, 100.0]	100 [99.0, 100.0]	0	7.86 ms
BC-VAE	82 [77.9, 85.5]	77 [72.6, 80.9]	5	6.96 ms
DP (chunk=16)	100 [99.0, 100.0]	100 [99.0, 100.0]	0	614.90 ms
DP (chunk=8)	100 [99.0, 100.0]	100 [99.0, 100.0]	0	604.83 ms
DP (chunk=4)	100 [99.0, 100.0]	98 [96.1, 99.0]	2	594.14 ms
DP (chunk=2)	100 [99.0, 100.0]	72 [67.4, 76.2]	28	602.35 ms
DP (chunk=1)	100 [99.0, 100.0]	45 [40.2, 49.9]	55	596.97 ms
Transport				
BC-TF	33 [28.6, 37.8]	2 [1.0, 3.9]	31	16.82 ms
BC-TF-GMM	78 [73.7, 81.8]	3 [1.7, 5.2]	75	17.78 ms
BC-Gaussian	0 [0.0, 1.0]	0 [0.0, 1.0]	0	13.48 ms
BC	0 [0.0, 1.0]	0 [0.0, 1.0]	0	13.12 ms
BC-GMM	52 [47.1, 56.9]	4 [2.5, 6.4]	48	14.19 ms
BC-VAE	0 [0.0, 1.0]	0 [0.0, 1.0]	0	13.33 ms
DP (chunk=16)	83 [79.0, 86.4]	5 [3.3, 7.6]	78	660.48 ms
DP (chunk=8)	81 [76.9, 84.5]	6 [4.1, 8.8]	75	619.36 ms
DP (chunk=4)	82 [77.9, 85.5]	2 [1.0, 3.9]	80	630.00 ms
DP (chunk=2)	82 [77.9, 85.5]	0 [0.0, 1.0]	82	636.46 ms
DP (chunk=1)	76 [71.6, 79.9]	0 [0.0, 1.0]	76	643.10 ms
ToolHang				
BC-TF	0 [0.0, 1.0]	0 [0.0, 1.0]	0	12.97 ms
BC-TF-GMM	0 [0.0, 1.0]	0 [0.0, 1.0]	0	12.65 ms
BC-Gaussian	0 [0.0, 1.0]	0 [0.0, 1.0]	0	8.34 ms
BC	0 [0.0, 1.0]	0 [0.0, 1.0]	0	8.13 ms
BC-GMM	0 [0.0, 1.0]	0 [0.0, 1.0]	0	9.29 ms
BC-VAE	0 [0.0, 1.0]	0 [0.0, 1.0]	0	8.33 ms
DP (chunk=16)	46 [41.2, 50.9]	32 [27.6, 36.7]	14	572.72 ms
DP (chunk=8)	30 [25.7, 34.7]	29 [24.8, 33.6]	1	588.40 ms
DP (chunk=4)	35 [30.5, 39.8]	26 [21.9, 30.5]	9	572.12 ms
DP (chunk=2)	33 [28.6, 37.8]	2 [1.0, 3.9]	31	573.30 ms
DP (chunk=1)	31 [26.7, 35.7]	0 [0.0, 1.0]	31	566.81 ms

Table 3: Detailed performance on Robomimic tasks under synchronous and asynchronous simulation, evaluated over 400 episodes. Success rates (%) are shown with 95% Wilson score confidence intervals in brackets.

402 **B LIBERO full results**

403 We provide detailed per-policy results on the LIBERO benchmark [4] across four tasks: **Spatial**,
 404 **Object**, **Goal**, and **Long**, as well as their average.

		Task Success Rate (%)					Performance drop (sync → async)	Rel. drop (%)	IAR ↑	Latency ↓
		Sync / Async								
		Spatial	Object	Goal	Long	Avg.				
SmolVLA	450M	90 / 56	96 / 52	92 / 67	71 / 50	87.2 / 56.2	-31.0	35.6	64.5	9.31 ms
X-VLA	0.9B	93 / 83	98 / 95	99 / 94	98 / 92	97.0 / 91.0	-6.0	6.2	93.8	14.32 ms
MolmoAct2	5B	98 / 96	100 / 98	98 / 96	97 / 93	98.2 / 95.8	-2.5	2.4	97.6	16.14 ms
π_0	4B	90 / 52	86 / 68	95 / 72	73 / 44	86.0 / 59.0	-27.0	31.4	68.6	44.46 ms
$\pi_{0.5}$	4B	97 / 68	99 / 65	98 / 74	96 / 77	97.5 / 71.0	-26.5	27.2	72.8	66.73 ms

Table 4: Synchronous and asynchronous simulation performance on the LIBERO benchmark [4].

405 **C Real-World Evaluation Details**

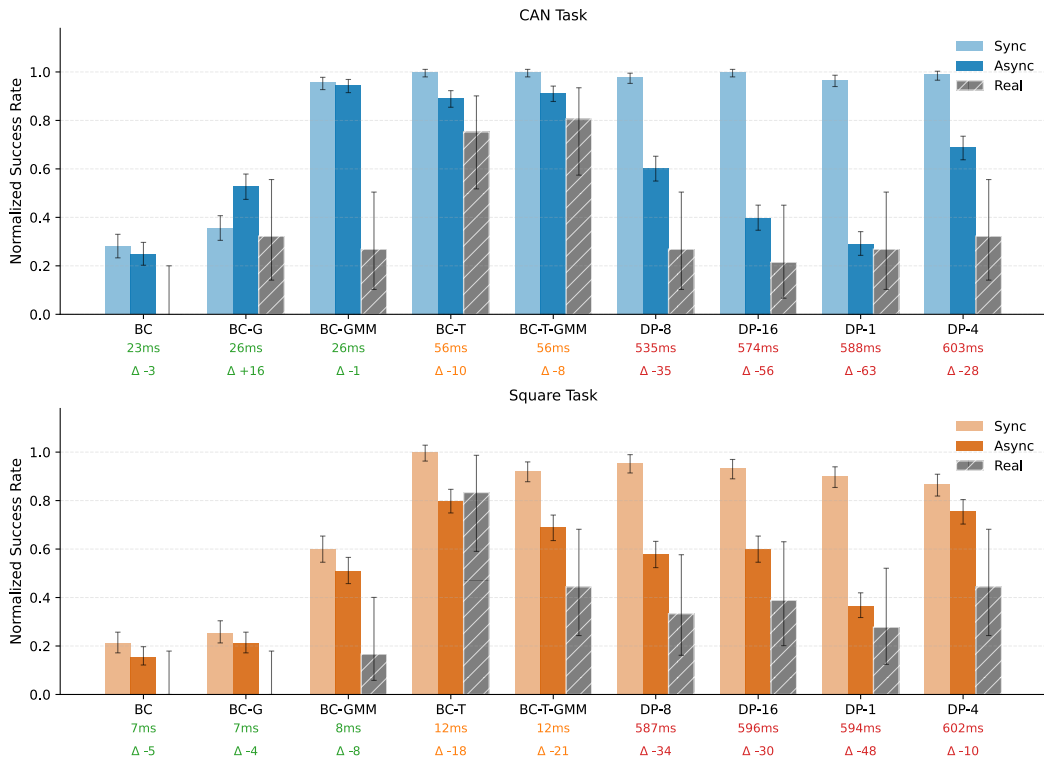


Figure 6: Per-policy success rates under synchronous, asynchronous, and real-world evaluation on the **Can** and **Square** tasks. Values are min-max normalized per task. Error bars indicate 95% Wilson score confidence intervals (400 trials for simulation, 20 for real world).

Policy	Sync	Async	Δ	Real	Inference Latency
BC-TF	98 [96.1, 99.0]	88 [84.4, 90.8]	10	75 [53.1, 88.8]	55.76 ms
BC-TF-GMM	98 [96.1, 99.0]	90 [86.7, 92.6]	8	80 [58.4, 91.9]	56.01 ms
BC-Gaussian	38 [33.4, 42.8]	54 [49.1, 58.8]	-16	35 [18.1, 56.7]	26.07 ms
BC	31 [26.7, 35.7]	28 [23.8, 32.6]	3	5 [0.9, 23.6]	22.76 ms
BC-GMM	94 [91.2, 95.9]	93 [90.1, 95.1]	1	30 [14.5, 51.9]	25.94 ms
BC-VAE	87 [83.3, 89.9]	72 [67.4, 76.2]	15	-	23.27 ms
DP (chunk=16)	98 [96.1, 99.0]	42 [37.3, 46.9]	56	25 [11.2, 46.9]	574.39 ms
DP (chunk=8)	96 [93.6, 97.5]	61 [56.1, 65.7]	35	30 [14.5, 51.9]	534.97 ms
DP (chunk=4)	97 [94.8, 98.3]	69 [64.3, 73.3]	28	35 [18.1, 56.7]	602.67 ms
DP (chunk=2)	97 [94.8, 98.3]	10 [7.4, 13.3]	87	-	623.06 ms
DP (chunk=1)	95 [92.4, 96.7]	32 [27.6, 36.7]	63	30 [14.5, 51.9]	587.53 ms

Table 5: Performance on the Robomimic **Can** task under synchronous simulation, asynchronous simulation, and real-world evaluation. Success rates (%) are shown with 95% Wilson score confidence intervals in brackets.

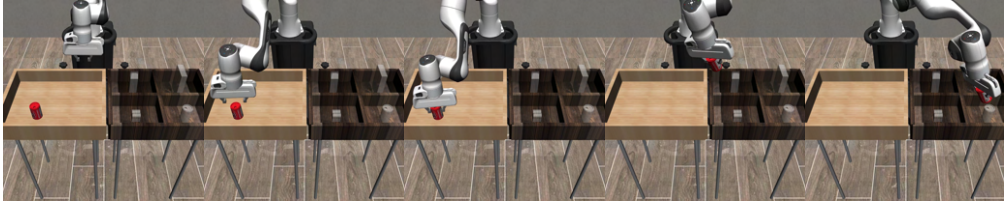
Policy	Sync	Async	Δ	Real	Inference Latency
BC-TF	90 [86.7, 92.6]	72 [67.4, 76.2]	18	75 [53.1, 88.8]	12.34 ms
BC-TF-GMM	83 [79.0, 86.4]	62 [57.2, 66.6]	21	40 [21.9, 61.3]	12.38 ms
BC-Gaussian	23 [19.1, 27.4]	19 [15.5, 23.1]	4	0 [0.0, 16.1]	7.13 ms
BC	19 [15.5, 23.1]	14 [10.9, 17.7]	5	0 [0.0, 16.1]	6.72 ms
BC-GMM	54 [49.1, 58.8]	46 [41.2, 50.9]	8	15 [5.2, 36.0]	7.75 ms
BC-VAE	28 [23.8, 32.6]	24 [20.1, 28.4]	4	-	6.79 ms
DP (chunk=16)	84 [80.1, 87.3]	54 [49.1, 58.8]	30	35 [18.1, 56.7]	595.73 ms
DP (chunk=8)	86 [82.3, 89.1]	52 [47.1, 56.9]	34	30 [14.5, 51.9]	587.12 ms
DP (chunk=4)	78 [73.7, 81.8]	68 [63.3, 72.4]	10	40 [21.9, 61.3]	602.23 ms
DP (chunk=2)	83 [79.0, 86.4]	30 [25.7, 34.7]	53	-	589.04 ms
DP (chunk=1)	81 [76.9, 84.5]	33 [28.6, 37.8]	48	25 [11.2, 46.9]	594.50 ms

Table 6: Performance on the Robomimic **Square** task under synchronous simulation, asynchronous simulation, and real-world evaluation. Success rates (%) are shown with 95% Wilson score confidence intervals in brackets.

406 D Simulation Demos

407 D.1 RoboMimic Simulation Demos

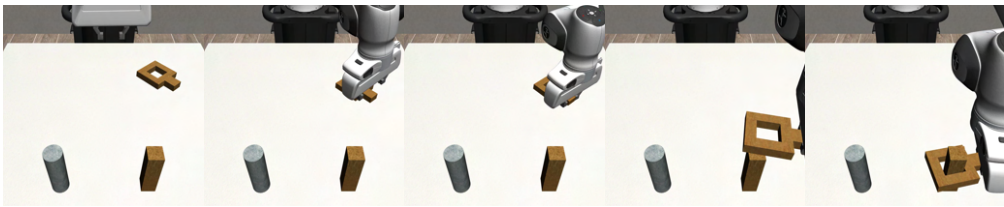
408 **Tasks.** We provide RoboMimic simulation demos on five representative manipulation tasks: **Can**,
409 **Square**, **Lift**, **Tool Hang**, and **Transport**.



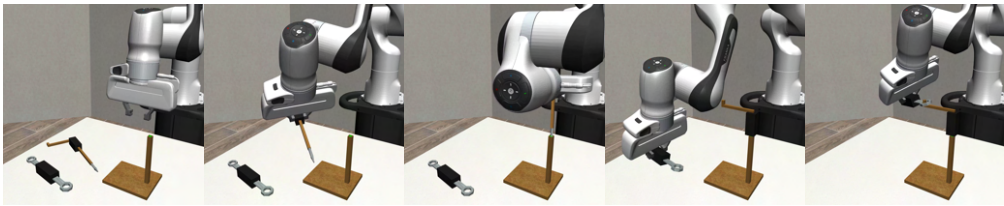
(a) **Can**: pick up a soda can from one bin and place it into the target bin.



(b) **Lift**: grasp a small cube from the table and lift it above a threshold height.



(c) **Square**: pick up a square nut and place it onto a vertical peg, requiring precise alignment.



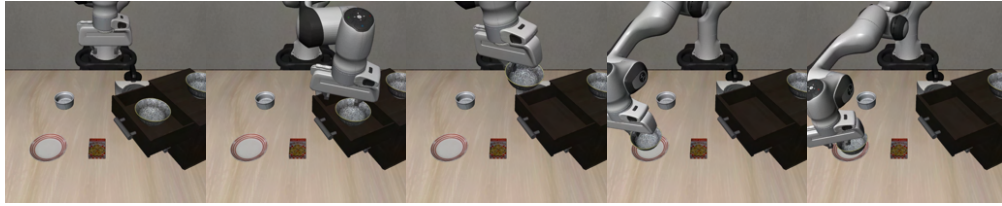
(d) **Tool Hang**: assemble a frame by inserting a hook into the base, then hang a wrench on the hook: a long-horizon, high-precision task.



(e) **Transport**: a bimanual task where two arms coordinate to transfer a hammer from a closed container to a target bin.

Figure 7: Rollout visualizations on the five RoboMimic tasks.

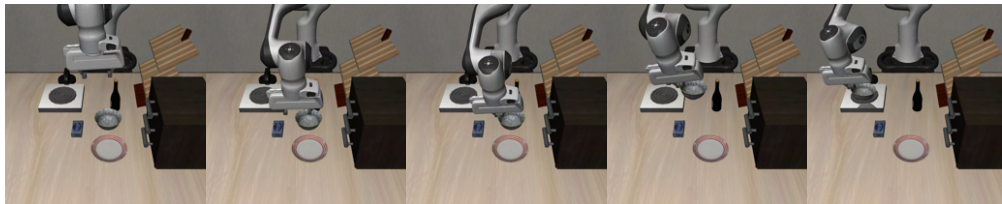
410 **D.2 LIBERO Simulation Demos**



(a) **LIBERO-Spatial**: pick-and-place tasks that require grounding spatial relations between objects in a tabletop scene.



(b) **LIBERO-Object**: pick-and-place tasks that require identifying and manipulating the correct household object in a kitchen scene.



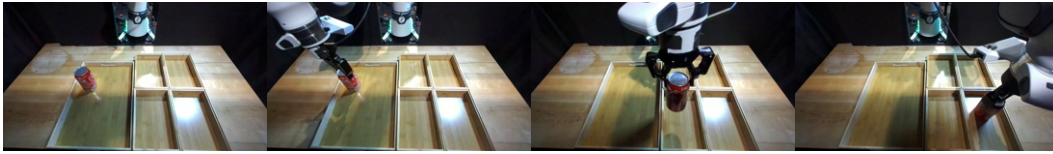
(c) **LIBERO-Goal**: short-horizon manipulation tasks where the policy must execute the goal specified by the language instruction.



(d) **LIBERO-Long**: long-horizon kitchen manipulation tasks that require chaining multiple sub-skills to completion.

Figure 8: Rollout visualizations on the four LIBERO task suites.

411 **E Real-world Demos**



(a) **Can**: pick up a soda can from one bin and place it into the target bin. (front camera view)



(b) **Square**: pick up a square nut and place it onto a vertical peg, requiring precise alignment. (external camera view; not used for training)

Figure 9: Rollout visualizations on the two real-world Robomimic [3] **Can** and **Square** tasks.