CONDENSED ENGINEERING REPORT

UR5e Cartesian Payload Transportation

Model-Based Control and Residual Reinforcement Learning

Sooyong Kim | October 2026

ROS 2 Humble • NVIDIA Isaac Sim 6.1 • Six-DOF UR5e

1 Research Question and Experimental Strategy

Industrial manipulators are deployed to perform physical tasks rather than to minimize joint-coordinate error as an isolated mathematical objective. During the transport of an attached object, the relevant outcome is the motion of the object in Cartesian space, together with the effort and mechanical behavior required to produce that motion. A small joint error may translate into a significant tool displacement in one configuration but not another because the relationship is mediated by the configuration-dependent manipulator Jacobian. Conversely, apparently accurate task-space motion can conceal different patterns of joint effort. These observations motivate a control investigation organized around the transport task, not around controller sophistication alone.

The central question is whether an established dynamics-based controller can remain useful when the actual payload differs from its assumed payload model, and whether *bounded residual reinforcement learning* can compensate the resulting loss of Cartesian tracking without replacing the nominal control law. The question has an equally important converse: when the nominal model is already appropriate, or when the mass discrepancy reverses sign, does the learned compensation remain useful or become detrimental? The research addresses both outcomes rather than presuming that a learned action must improve every condition.

The experimental design proceeds through distinct levels of evidence. An analytical UR5e model is constructed and cross-checked against Isaac Sim geometry; a deliberately demanding three-second transport is developed using the robot's workspace and dynamic limitations; computed-torque PD and local model predictive control (MPC) are screened as nominal candidates; and the chosen controller is commissioned through a ROS 2–PhysX effort interface. Residual policies are then trained in a numerical plant and transferred to PhysX for matched, heavier-than-modeled, lighter-than-modeled, reversed, and intermediate payload evaluations. These stages are not interchangeable. Analytical inverse dynamics predict effort, offline simulations test a mathematical plant, and PhysX experiments test the closed-loop response of a separately executed articulation.

The work deliberately excludes grasp acquisition, grasp stability, contact-rich insertion, payload release, and physical hardware trials. The transported 50-mm cube is rigidly fixed to the tool, allowing the study to isolate transportation dynamics and changes in payload parameters. A nominal simulation torque limit is not treated as a manufacturer's continuous actuator rating, and good Cartesian tracking is not interpreted as a collision-clearance or industrial safety guarantee. The aim is to produce an interpretable pre-hardware engineering assessment with clear evidence boundaries.

2 Manipulator Model and Payload Dynamics

2.1 Frames, kinematics, and model consistency

The UR5e is represented as a fixed-base, six-revolute-joint manipulator with joint vector $\mathbf{q} \in \mathbb{R}^6$, ordered shoulder pan, shoulder lift, elbow, wrist 1, wrist 2, and wrist 3. Homogeneous transformations map the base frame to Tool0 and provide the forward-kinematics function $\mathbf{p}_T = f_p(\mathbf{q})$. The accepted analytical-to-Isaac world-frame convention includes the rotation ${}^{W}R_B = \mathrm{diag}(-1, -1, 1)$; the tool's local positive Z axis points downward during the transport. These definitions determine both Cartesian tracking coordinates and the direction of payload-induced joint moments.

The differential kinematics are

$$\begin{bmatrix} \dot{\mathbf{p}}_T \\ \boldsymbol{\omega}_T \end{bmatrix} = J(\mathbf{q})\dot{\mathbf{q}}, \qquad \ddot{\mathbf{p}}_T = J_v(\mathbf{q})\ddot{\mathbf{q}} + \dot{J}_v(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}}. \tag{1}$$

The same Jacobian also connects a consistently expressed external wrench to generalized joint effort through $\boldsymbol{\tau}_{\text{ext}} = J^T\mathbf{w}$. This dual kinematic and static role makes the selected manipulator configuration relevant to

both position tracking and loading. Sequential damped-least-squares inverse kinematics is initialized from the previously accepted joint solution to maintain a continuous branch along the task reference.

A representative forward-kinematics comparison between the analytical model and the Isaac articulation gave a Tool0 position discrepancy of approximately 0.000206 mm after applying the adopted frame mapping. This supports numerical geometric consistency for the checked pose; it is not a measurement of hardware accuracy. Joint identities and conventions, the fixed payload attachment, and the Tool0-to-payload center-of-mass offset were separately inspected because matching Cartesian positions do not establish matching inertial properties or actuation behavior.

2.2 Payload-dependent dynamics

The rigid-body dynamics of the loaded manipulator are

$$M(\mathbf{q}; m_p)\ddot{\mathbf{q}} + \mathbf{c}(\mathbf{q}, \dot{\mathbf{q}}; m_p) + \mathbf{g}(\mathbf{q}; m_p) = \boldsymbol{\tau}, \tag{2}$$

where m_p is the rigidly attached payload mass. The cube has side length $a = 0.05$ m and Tool0-relative center of mass $\mathbf{r}_p = [0, 0, 0.025]^T$ m. For a homogeneous cube, its central rotational inertia is $I_{p,c} = (m_p a^2/6)I_3$. The payload inertia projected into joint coordinates includes translational and rotational contributions; in a consistent representation,

$$M_p = m_p J_{v,p}^T J_{v,p} + J_{\omega,p}^T R_p I_{p,c} R_p^T J_{\omega,p}. \tag{3}$$

Because the payload Jacobian depends on configuration, adding a constant mass does not add a constant generalized joint load along the trajectory. It changes the inertia matrix, gravity vector, and velocity-dependent terms throughout the motion. The analytical routines return the complete loaded robot model; payload contributions must not be added twice.

For the prescribed reference $\mathbf{q}_r(t)$, the mass-dependent inverse-dynamics requirement is

$$\boldsymbol{\tau}_{\mathrm{ID}}(t; m) = M(\mathbf{q}_r; m)\ddot{\mathbf{q}}_r + \mathbf{c}(\mathbf{q}_r, \dot{\mathbf{q}}_r; m) + \mathbf{g}(\mathbf{q}_r; m). \tag{4}$$

The nominal-model error has the physically interpretable form

$$\Delta\boldsymbol{\tau}_{\mathrm{ID}} = [M(\mathbf{q}_r; m_a) - M(\mathbf{q}_r; m_m)]\,\ddot{\mathbf{q}}_r + \Delta\mathbf{c} + \Delta\mathbf{g}, \tag{5}$$

where m_a is the actual payload and m_m is the mass assumed in control. This is the *reference inverse-dynamics discrepancy*, not necessarily the optimal residual command of an already feedback-controlled robot. It varies in time and depends on joint configuration, velocity, and acceleration.

3 Task-Oriented Cartesian Transportation Benchmark

3.1 Workspace and trajectory construction

The task was constructed from the UR5e's physical capabilities rather than from arbitrary Cartesian endpoints. The cited manufacturer specification lists nominal reach of 850 mm, maximum rated payload of 5 kg subject to center-of-gravity conditions, maximum joint speeds of $\pm 180°/\mathrm{s}$, and repeatability of ± 0.03 mm [5]. These are specifications for different physical quantities: nominal reach does not imply uniform effort capacity, and repeatability is not dynamic tracking RMSE. The adopted USD articulation effort limits were $[150, 150, 150, 28, 28, 28]^T$ Nm; these are research simulation settings, not verified UR5e hardware motor ratings.

The accepted target coordinates were

$$\mathbf{p}_0 = \begin{bmatrix} 0.4648 & -0.4959 & 0.005 \end{bmatrix}^T \mathrm{m}, \qquad \mathbf{p}_f = \begin{bmatrix} 0.5520 & 0.5889 & 0.355 \end{bmatrix}^T \mathrm{m}. \tag{6}$$

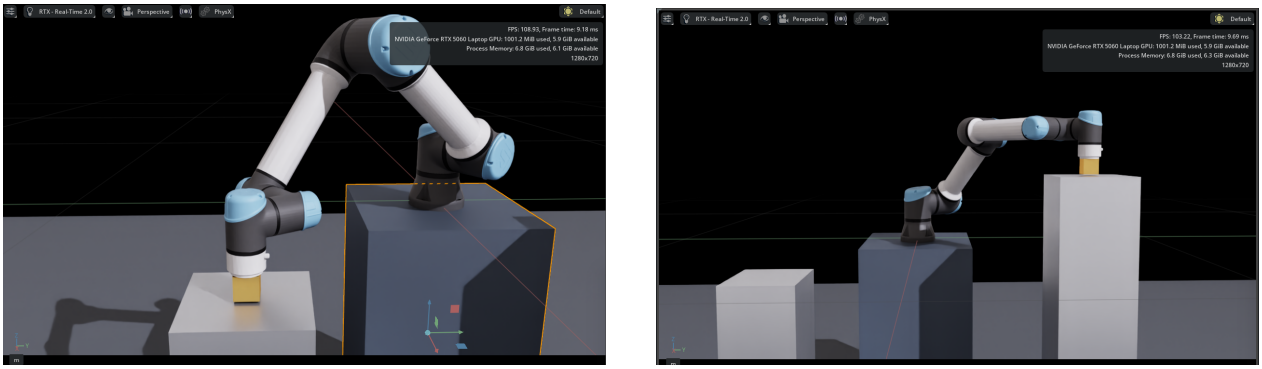
Their horizontal distances from the base are approximately 0.680 m (80% of nominal reach) and 0.807 m (95%), respectively. The elevated destination deliberately exposes configuration-dependent loading while the sequential inverse-kinematics solutions remain numerically feasible. All 121 reference-waypoint IK solutions converged; the recorded maximum Jacobian condition number was approximately 15.53 under the study's numerical convention. Sampled broad-phase and mesh-intersection checks found no collisions among the evaluated configurations, but these checks are not a continuous swept-volume safety certificate.

The normalized path parameter $s \in [0, 1]$ defines linearly changing horizontal coordinates and a two-parabola vertical path:

$$x(s) = 0.4648 + 0.0872s, \qquad y(s) = -0.4959 + 1.0848s, \tag{7}$$

$$z(s) = \begin{cases} 0.500 - 0.495(1 - s/s_j)^2, & 0 \le s \le s_j, \\ 0.500 - 0.145[(s - s_j)/(1 - s_j)]^2, & s_j < s \le 1, \end{cases} \qquad s_j = \frac{2}{3}. \tag{8}$$

The reference ascends from the pick target to a 0.500-m apex and descends to the elevated place target. It is continuous in position and first derivative, but its second derivative changes across the apex. The geometric curvature change is a deliberate test feature; it is not recommended without further smoothing for arbitrary production tasks. The time law slows path progress near the join while preserving a three-second duration, yielding a transition at approximately $t_j = 1.7944$ s. The prescribed reference achieved a recorded peak joint-speed demand of about 71.78°/s, below the cited speed specification. The original 1-kg reference analysis gave a peak shoulder-lift torque of 62.435 Nm, illustrating that gravity and inertia must be considered beyond kinematic reach.



(a) Pick configuration. **(b)** Extended place configuration.

Figure 1: Accepted Isaac Sim transportation workspace. The model uses a rigid, post-grasp payload; no physical clearance guarantee is inferred from this visualization.

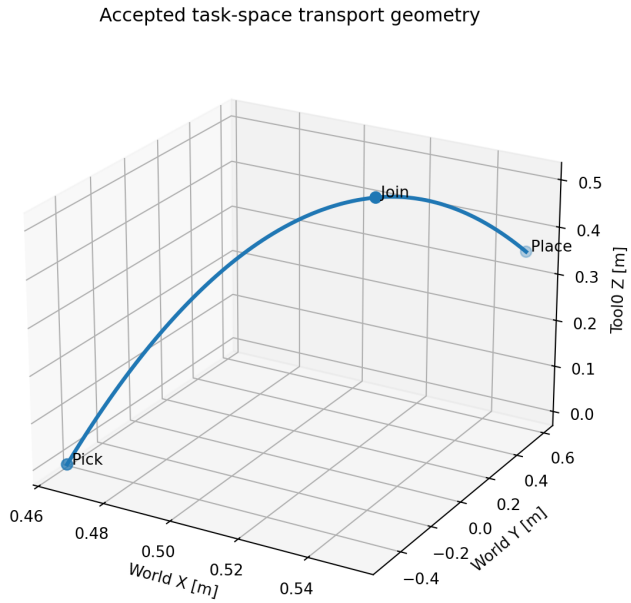


Figure 2: Accepted three-dimensional transportation path with the distinct ascending and descending segments.

The original 5-mm bottom-of-cube tabletop clearance was a geometric design assumption. The actual closed-loop payload-to-table minimum separation was not recorded, and the initial joint perturbation changed the starting Tool0 height relative to the unperturbed reference. Thus, trajectory tracking error must not be equated with verified clearance.

4 Nominal Control: PD/MPC Screening and PhysX Validation

4.1 Computed-torque PD and candidate selection

The nominal controller uses reference-state inverse-dynamics feedforward combined with feedback scaled by the inertia matrix at the measured configuration:

$$\boldsymbol{\tau}_{\mathrm{ff}}(t;m) = M(\mathbf{q}_r;m)\ddot{\mathbf{q}}_r + \mathbf{c}(\mathbf{q}_r,\dot{\mathbf{q}}_r;m) + \mathbf{g}(\mathbf{q}_r;m), \tag{9}$$

$$\boldsymbol{\tau}_{\mathrm{nom}} = \mathrm{sat}_{\boldsymbol{\tau}_{\max}}\left\{\boldsymbol{\tau}_{\mathrm{ff}} + M(\mathbf{q};m)\left[K_p(\mathbf{q}_r-\mathbf{q}) + K_d(\dot{\mathbf{q}}_r-\dot{\mathbf{q}})\right]\right\}. \tag{10}$$

This is not claimed to be identical to exact measured-state inverse-dynamics cancellation away from the reference. The gain search uses the ideal second-order interpretation $K_p = \omega_n^2$ and $K_d = 2\zeta\omega_n$ as a tuning guide, not as proof of exact closed-loop linear error dynamics in sampled PhysX execution.

Twenty-five offline PD configurations were tested using the same three-second reference and an initial perturbation of $+0.5°$ at shoulder lift and $-0.5°$ at the elbow. Three candidates were carried into PhysX validation. G022 achieved the lowest offline Cartesian RMSE, but G025 offered stronger damping and a more favorable selected combination of tracking and terminal behavior. Table 1 distinguishes offline from independently executed PhysX results.

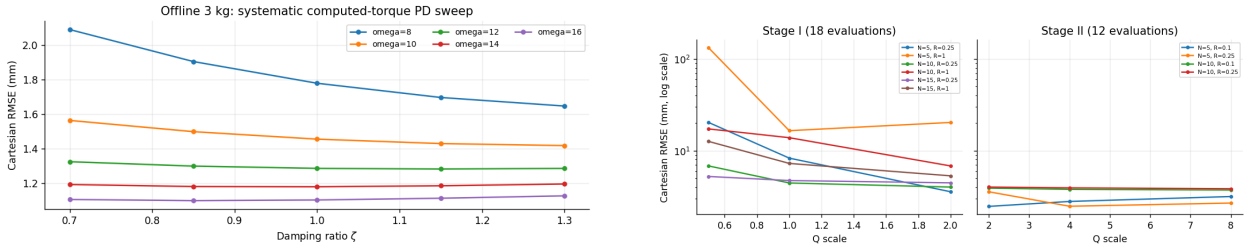
Table 1: Selected computed-torque PD gains and Cartesian RMSE in distinct evaluation environments.

Candidate	K_p	K_d	Offline (mm)	PhysX (mm)
G013	144	24.0	1.286	1.454
G022	256	27.2	**1.099**	1.233
G025	256	41.6	1.127	**1.195**

4.2 MPC as an alternative nominal controller

A local linearized predictive model was used for the state $\mathbf{x}_k = [(\mathbf{q}_k - \mathbf{q}_{rk})^T, (\dot{\mathbf{q}}_k - \dot{\mathbf{q}}_{rk})^T]^T$ and correction input $\mathbf{u}_k = \Delta\boldsymbol{\tau}_k$, with prediction $\mathbf{x}_{k+1} \approx A_k\mathbf{x}_k + B_k\mathbf{u}_k + \mathbf{d}_k$. The finite-horizon objective penalized state tracking, correction effort, and terminal state while constraining the total command $\boldsymbol{\tau}_{\mathrm{ff}} + \mathbf{u}_k$ within the adopted simulation effort limits. This was local linearized MPC, not full nonlinear MPC.

The initial 18-run sweep varied horizon $N \in \{5, 10, 15\}$ and two sets of state/effort weights. Its best Cartesian RMSE was 3.566 mm. A 12-run refinement with $N \in \{5, 10\}$ and expanded weights improved the best RMSE to 2.470 mm. The gain was not monotonic with increased state penalty. The representative QP solve time was approximately 0.779 ms on average and 0.986 ms at maximum for one tested configuration; hence computational infeasibility was not demonstrated. Nevertheless, the tested MPC did not provide enough task-level advantage over computed-torque PD to justify selecting it as the residual-learning baseline. This is a task- and implementation-specific decision, not a general ranking of the controller families.



(a) PD configuration screening.

(b) Two-stage MPC screening.

Figure 3: Controller search results from the original experimental records. The final G025 choice was confirmed independently in PhysX rather than selected by offline Cartesian RMSE alone.

4.3 ROS 2–PhysX effort interface and startup validation

The accepted execution system used Isaac Sim 6.1 and ROS 2 Humble. Joint states were published on `/joint_states`; joint-effort commands were received on `/joint_command`; both messages used `sensor_msgs/JointState`. Incoming commands included six ordered joint names and a six-element effort vector, while position and velocity command arrays were empty. The articulation-controller effort input applied the commanded generalized efforts. Angular-drive stiffness and damping were set to zero to avoid an unaccounted

internal position servo. Topic availability alone was not considered proof of successful torque actuation: the closed-loop holding response was observed.

PhysX was configured with a nominal 2-ms integration step (500 Hz) and the ROS controller with a requested 20-ms timer (50 Hz). Joint-state feedback arrived approximately 200–300 Hz, and the three-second logs contained roughly 286–302 records depending on the run. These rates describe different operations; a ROS wall-clock timer does not guarantee exactly ten PhysX steps between all controller callbacks. Position error was reconstructed from ROS joint-state forward kinematics and compared to a reference interpolated at the actual logged timestamps. The time-weighted position RMSE was

$$\mathrm{RMSE}_p = \sqrt{\frac{1}{T}\int_0^T \|f_p(\mathbf{q}_{\mathrm{ROS}}(t)) - \mathbf{p}_r(t)\|_2^2 \, dt}. \tag{11}$$

An essential commissioning finding was gravitational sag when the simulation was started without useful compensating effort. One initial trial departed by more than 100° from the intended configuration. A low-speed condition alone could not establish that the robot had returned to its starting pose. The corrected sequence was `PRELOAD` → `RETURN` → `STABILIZE` → `TRACKING` → `FINAL_HOLD`. A gravity-preload command was issued before the first valid joint-state reading; feedback then returned the robot to the deliberately offset starting pose. An explicit start request and two continuous seconds satisfying the stability gate preceded tracking.

Table 2: Startup gate thresholds for the accepted 3-kg ROS 2–PhysX trials.

Criterion	Threshold
Maximum joint-position error	0.05°
Maximum joint speed	0.30°/s
Tool0 relative height error	1.0 mm
Tool0 relative horizontal error	2.0 mm
Continuous accepted dwell	2.0 s, simulator time
Joint-state freshness	0.15 s, host time

The stabilized reported joint error was approximately 0.006–0.009° and joint speed 0.074–0.080°/s. The gate checked recovery to the intentionally offset initial pose, not the original no-offset pick pose. The accepted initial tracking position was approximately 2.9 mm below the original reference Tool0 height; no continuous independent payload-clearance measurement was available. Startup and terminal-hold samples were excluded from the scored three-second tracking interval.

Table 3: Nominal matched 3-kg PhysX performance of the PD shortlist.

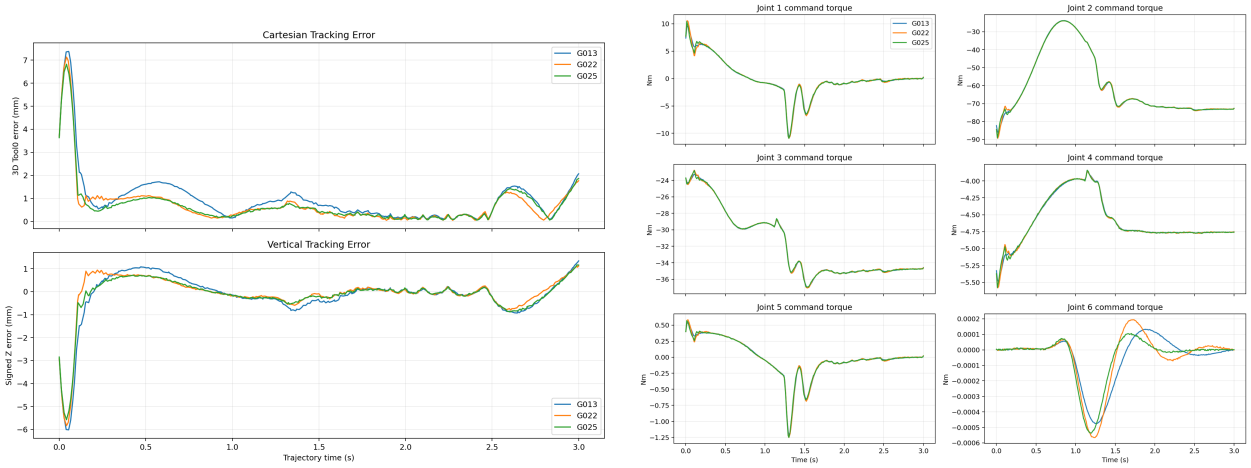
Measure	G013	G022	G025
Cartesian RMSE (mm)	1.454	1.233	**1.195**
Joint-position RMSE (°)	0.170	0.125	**0.115**
Peak Cartesian error (mm)	7.380	7.134	6.823
Terminal Cartesian error (mm)	2.066	**1.767**	1.862
Terminal joint speed (°/s)	**1.273**	2.181	1.513
Peak commanded effort (Nm)	85.966	89.397	89.095
Command saturation count	0	0	0

The largest Cartesian transients occurred near the deliberate initial-offset-to-reference transition, rather than characterizing the entire steady transportation segment. G025 achieved the lowest recorded Cartesian and joint-position RMSE among the three PhysX runs, although G022 retained a smaller terminal position error and G013 a smaller terminal speed norm. The difference between G025 and G022 is a one-run observation, not evidence of statistical dominance. The selected nominal G025 settings were $K_p = 256$ and $K_d = 41.6$.

5 Residual Reinforcement Learning: Formulation and Training

5.1 Residual architecture and reward

The nominal G025 controller is retained in full. A learned policy π_ϕ receives a 49-dimensional observation comprising eight six-dimensional vectors—actual joint position and velocity, reference joint position and velocity, position and velocity tracking errors, previous normalized residual action, and previous normalized



(a) Cartesian tracking error. **(b)** Joint effort history.

Figure 4: Representative nominal ROS 2–PhysX validation plots. The torque plot represents recorded simulated command quantities, not independent hardware motor-torque measurements.

nominal torque—plus normalized task time. It outputs a six-dimensional normalized continuous action $\mathbf{a}_t \in [-1, 1]^6$. The residual torque is

$$\Delta\boldsymbol{\tau}_{\pi,t} = \boldsymbol{\tau}_{\text{res,max}} \odot \mathbf{a}_t, \qquad \boldsymbol{\tau}_{\text{res,max}} = [5, 5, 5, 0.5, 0.5, 0.01]^T \text{ Nm}. \tag{12}$$

The final command is the nominal torque plus this bounded correction, clipped to the simulation's total-effort limits:

$$\boldsymbol{\tau}_{\text{cmd},t} = \text{clip}_{\boldsymbol{\tau}_{\max}} \left(\boldsymbol{\tau}_{\text{G025},t} + \Delta\boldsymbol{\tau}_{\pi,t}\right). \tag{13}$$

The joint-dependent residual bounds were reduced after unsuccessful exploratory rollouts and are distinct from final motor-command saturation. Persistent operation at a residual bound does not imply that the simulated complete joint-effort limit was reached. The policy receives no direct measurement of the actual payload mass and is not trained to estimate the mass as a supervised output.

The implemented reward gives primary importance to Cartesian task-space error, with smaller penalties for joint-position error, residual effort, and action change:

$$r_t = -\left[\left(\frac{\|f_p(\mathbf{q}_t) - \mathbf{p}_r(t)\|}{0.01\text{ m}}\right)^2 + 0.1\,\text{mean}(\mathbf{e}_{q,t}^2) + 0.01\,\text{mean}(\mathbf{a}_t^2) + 0.01\,\text{mean}\left((\mathbf{a}_t - \mathbf{a}_{t-1})^2\right)\right]. \tag{14}$$

These regularizers are soft preferences, not stability certificates, torque-slew constraints, or collision-safety guarantees. The relative scale of each term depends on the implementation's state and action normalization.

5.2 Training protocol and algorithm comparison

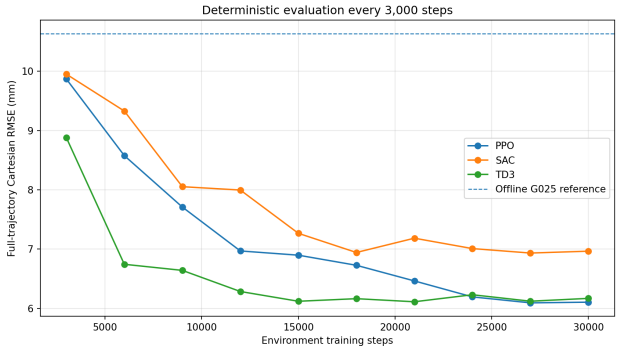
The numerical environment uses a 2-ms plant integration step and 20-ms action updates. A complete three-second evaluation contains 150 policy decisions; each commanded action is held across ten numerical substeps. The initial training condition fixes a nominal 3-kg model while the numerical plant carries 5 kg. PPO, SAC, and TD3 were each allocated 30,000 environment interactions with seed 42; deterministic full-episode evaluations occurred at 3,000-step intervals. Equal interaction counts do not imply equal gradient-update budgets or wall time. The saved checkpoints were chosen using the same task trajectory as training, so their offline scores are not independent task-generalization measurements.

Table 4: Selected offline best-checkpoint performance for the 3-kg model / 5-kg plant.

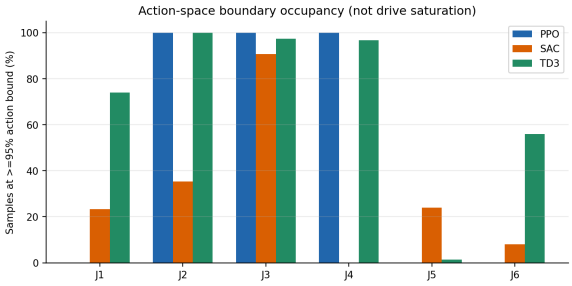
Policy	Best step	Cartesian RMSE (mm)	Terminal z error (mm)	Residual RMS (Nm)
G025, zero residual	–	10.636	–	0
PPO	27,000	**6.094**	−5.124	2.900
SAC	27,000	6.934	−6.254	2.916
TD3	21,000	6.112	−5.275	3.450

TD3 reached 6.744-mm Cartesian RMSE by 6,000 interactions and approximately 6.285 mm by 12,000, while PPO reached its lowest checkpoint at 27,000 interactions. The minimum PPO–TD3 RMSE difference

was only approximately 0.018 mm, too small and insufficiently replicated to justify a general algorithm ranking. TD3 was prioritized initially for PhysX transfer because of its earlier observed progress, while PPO remained a comparator. Following the independent physics results, PPO was chosen for the later bidirectional evaluation. This sequence records an evidence-based change in experimental preference, not a contradiction.



(a) Deterministic RL evaluation curves.



(b) Residual-bound occupancy.

Figure 5: Training progression and joint-level action constraints from the original best-policy evaluations. The joint residual bounds are distinct from full simulated actuator-effort limits.

5.3 Joint torque and the analytical mass discrepancy

The policies were not equivalent merely because their Cartesian RMSE values were similar. PPO requested an RMS residual of only 0.397 Nm at J1 compared with 4.632 Nm for TD3; PPO and TD3 both frequently applied −5 Nm to shoulder lift and elbow, and approximately −0.5 Nm to wrist 1. In PPO evaluation, J2, J3, and J4 occupied at least 95% of their negative residual bounds throughout all 150 steps. This makes the torque interpretation as important as the aggregate Cartesian score.

The independent analytical reference inverse-dynamics difference for a 5-kg plant minus a 3-kg model had means of about −13.384 Nm at J2, −8.860 Nm at J3, and −1.864 Nm at J4; these magnitudes exceeded their respective allowed residual corrections. This is consistent with the strong negative policy actions at those joints, but the residual is not expected to reproduce the complete theoretical difference because G025 feedback also reacts to error. At J1 the theoretical mean was nearly zero (−0.007 Nm) with a sign-changing range, making the large TD3 J1 compensation difficult to explain as a persistent missing gravitational feedforward. This motivates caution rather than a categorical judgment that one policy's torque distribution is wrong.

6 PhysX Payload-Mismatch Experiments

6.1 Matched, heavier, and lighter actual payloads

The independently executed ROS 2–PhysX transfer experiments retained G025, the three-second reference, and the same policy architecture. The nominal controller mass was fixed at 3 kg while the actual rigid-payload mass was changed among 1, 3, and 5 kg. Each controller's PhysX Cartesian RMSE is reported in Table 5. This comparison distinguishes policy performance under its intended heavier-than-modeled training condition from behavior when the model is accurate or the mass mismatch reverses sign.

Table 5: PhysX Cartesian tracking RMSE (mm) for a fixed nominal 3-kg model. Lower is better.

Controller	Actual 1 kg	Actual 3 kg	Actual 5 kg
G025, no residual	**10.920**	**1.195**	10.560
PPO	15.508	5.009	**6.061**
TD3	15.270	5.048	6.103
SAC	14.818	4.884	6.935

With a heavier actual payload, PPO reduced RMSE from 10.560 to 6.061 mm, approximately a 42.6% improvement. TD3 reduced it to 6.103 mm, and SAC to 6.935 mm. These observations support useful learned compensation for the trained mass-discrepancy direction, while also showing that the matched-mass accuracy of G025 was not recovered. The remaining error cannot be assigned uniquely to action bounds, simulator differences, or incomplete learning without additional controlled experiments.

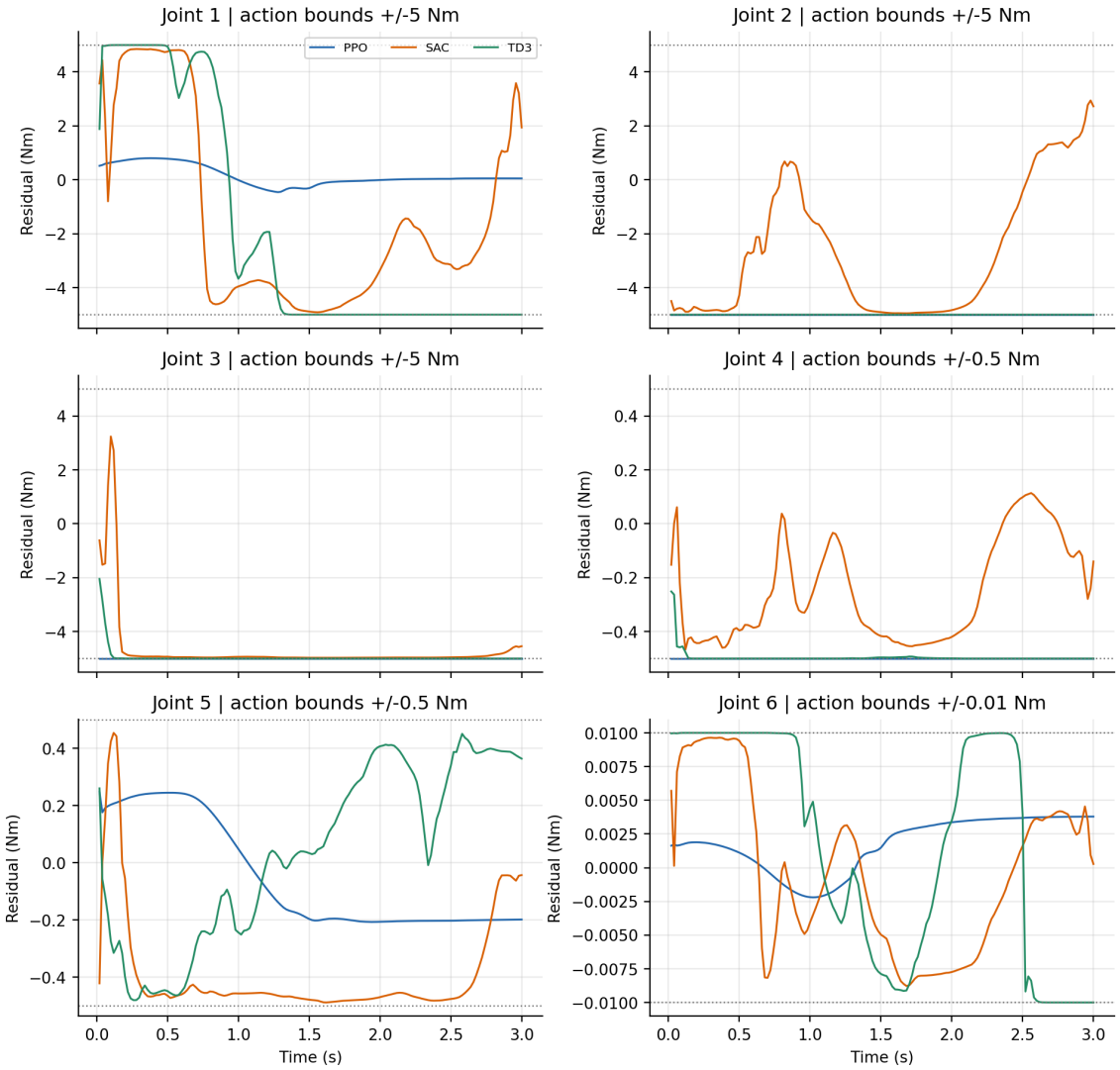
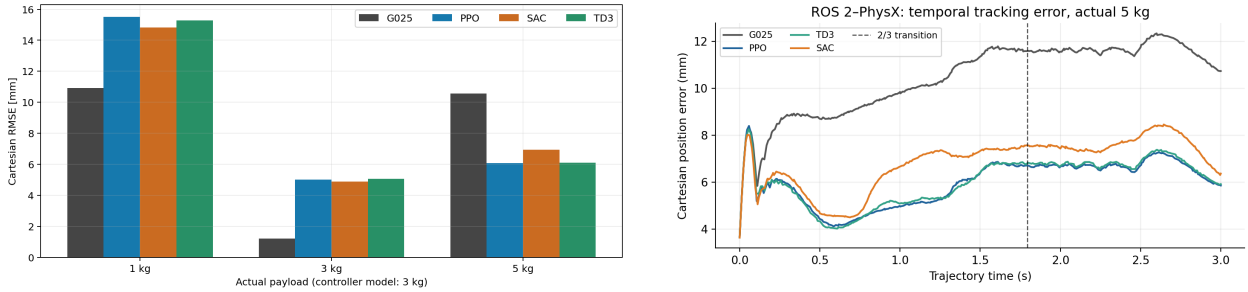


Figure 6: Saved best-policy joint-residual histories reveal different distributions of correction among PPO, SAC, and TD3.

With matched 3-kg model and plant, the original G025 already achieved 1.195 mm; every learned policy worsened that result. At 1 kg actual load, where the controller model assumed 2 kg too much, the original heavier-mismatch policies also increased tracking error above the zero-residual baseline. The two opposite ±2-kg conditions are not interchangeable even though they have the same absolute mismatch. Learned compensation is conditional on the sign of the dynamic discrepancy and must not be applied universally without a policy-validity mechanism.



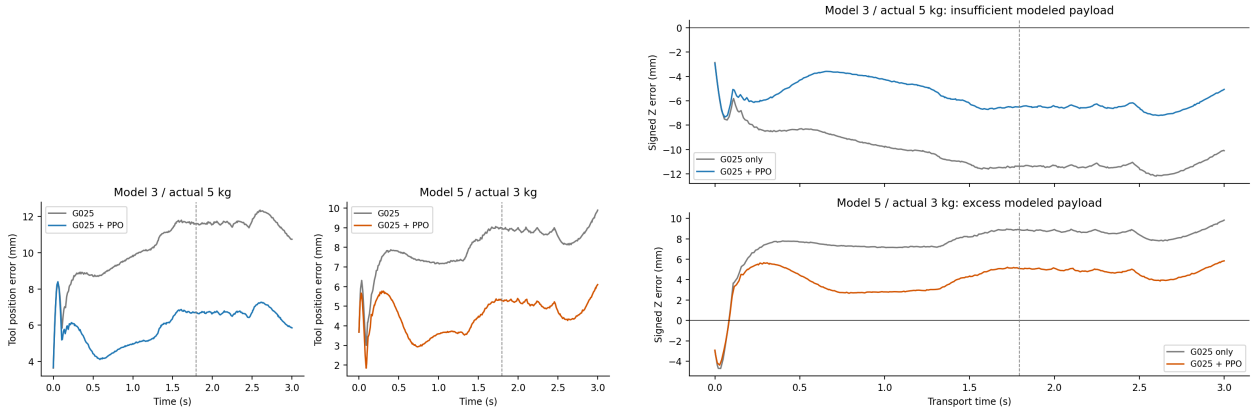
(a) Payload-by-controller comparison. **(b)** Error history at 5 kg actual mass.

Figure 7: Independent PhysX evaluation makes the conditional benefit of residual learning visible across loading conditions.

6.2 Reverse mismatch: model 5 kg, actual 3 kg

The reverse experiment retained the G025 gains and residual architecture but recomputed the nominal dynamics for 5 kg and trained a *new* PPO policy with a 3-kg numerical plant. The signed mass discrepancy was $m_a - m_m = -2$ kg, opposite the original training condition. In PhysX, the reverse PPO policy reduced Cartesian RMSE from approximately 8.039 mm for G025 alone to 4.585 mm, an improvement of about 43.0%. The dominant residual correction directions at J2, J3, and J4 reversed relative to the heavier-than-modeled PPO, consistent with $\Delta\boldsymbol{\tau}_{\mathrm{ID}}^{5\to3} = -\Delta\boldsymbol{\tau}_{\mathrm{ID}}^{3\to5}$ evaluated along the identical reference.

This does not establish that the two policies output pointwise exact negatives, nor that they perform physical parameter identification. They are independently optimized closed-loop policies operating with different nominal models. The result supports the applicability of the residual *architecture* to both directions through separate training; it does not establish one universally valid residual policy.



(a) Reverse-mismatch tracking response. **(b)** Signed vertical-error behavior.

Figure 8: Reverse-mismatch and bidirectional compensation results from the recorded PhysX experiments.

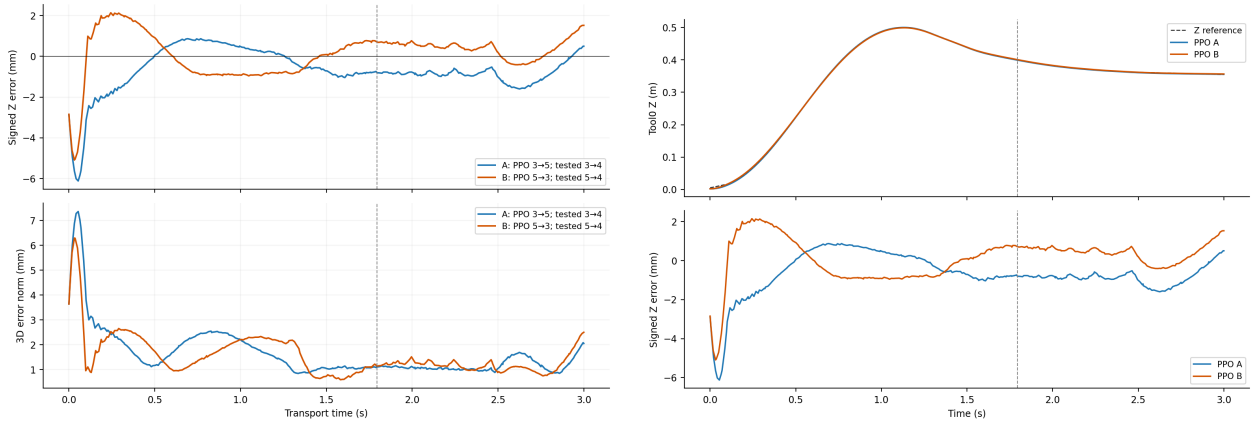
7 Intermediate 4-kg Cross-Evaluation

Two PPO policies were compared with the same *actual* 4-kg payload while retaining their respective nominal models. Policy A was trained at model 3 kg / plant 5 kg and evaluated at model 3 kg / plant 4 kg ($\Delta m = +1$ kg). Policy B was trained at model 5 kg / plant 3 kg and evaluated at model 5 kg / plant 4 kg ($\Delta m = -1$ kg). The policies thus faced reduced discrepancies of opposite sign, but the complete controllers differed both in their nominal model-based effort and learned residual. This is a comparison of *whole controllers*, not a controlled isolation of neural-network action alone.

Table 6: Cross-evaluation metrics with a common actual payload of 4 kg.

Metric	Policy A	Policy B
Nominal controller payload (kg)	3	5
Overall Cartesian RMSE (mm)	1.931	**1.763**
Ascent Cartesian metric (mm)	1.806	**1.740**
Descent Cartesian metric (mm)	**1.189**	1.236
Ascent vertical-error metric (mm)	**0.923**	1.023
Descent vertical-error metric (mm)	0.946	**0.588**

Policy B achieved the lower aggregate Cartesian RMSE by 0.168 mm, approximately 8.7% relative to Policy A. However, it did not improve every phase-specific metric. It had a smaller ascent Cartesian metric and descent vertical-error metric, while Policy A had a smaller descent Cartesian metric and ascent vertical-error metric. Thus, even when aggregate differences are small, the chosen measure and the physical phase of the task can affect the assessment. Because the reported Z-axis metrics are error magnitudes, they do not by themselves reveal signed bias; that requires the underlying $e_z(t)$ history. Nor can phase metrics be averaged arbitrarily to reconstruct the whole-trajectory RMSE without matching the original time weights and partitions.



(a) Actual 4-kg time histories.

(b) Ascent/descent metrics.

Figure 9: Intermediate-mass cross-evaluation of the two independently trained PPO configurations.

The 4-kg results provide limited evidence that the complete trained controllers can execute the fixed reference at an intermediate mass not used as the exact actual training load. They do not establish arbitrary payload generalization or that the residual itself improves on zero-residual G025 at 4 kg. The corresponding matched-protocol, zero-residual 3→4-kg and 5→4-kg evaluations are required to isolate incremental policy benefit. Changing the trajectory, timing, grasp offset, or initial posture would also create further evaluation conditions not addressed by this experiment.

8 Engineering Discussion and Deployment Boundaries

Three physical conclusions are supported by the combined experiments. First, accurate nominal compensation remains valuable: G025 tracks the matched 3-kg task with approximately 1.195-mm PhysX RMSE, and learning is not needed simply because learning is available. Second, fixed payload-model errors generate configuration- and motion-dependent effort discrepancies. A learned residual can improve the combined response when its training condition is relevant, but its torque must be interpreted jointly with the existing feedback action, not as a direct estimate of the missing inverse dynamics. Third, policy applicability is itself part of controller design: the original heavier-mismatch policy worsened matched- and lighter-mass behavior, while a separately trained reverse policy improved its intended condition.

The PhysX implementation further demonstrates why equations and executed motion require separate validation. With external torque control active but no useful compensation applied at startup, the robot sagged under gravity. The corrected controller required preload, return, stabilization, and an explicit tracking trigger. In another part of the preceding control-analysis workflow, the relative timing of effort application and physics integration affected tracking despite an unchanged nominal law. Such observations are a warning against blaming every deviation on feedback gains or reinforcement learning: frame conventions, load representation, effective actuation, state staleness, initialization, and simulator settings must first be checked. A controller

should be judged by its *physical response under applied commands*, not by the apparent correctness of calculated torque alone.

Simulation is an engineering risk-reduction stage, not a hardware-safety certificate. The loaded manipulator was modeled with a fixed rigid cube; actual grasp compliance, slip, structural elasticity, friction, proprietary drive behavior, and actuator thermal limits were not independently validated. The manufacturer repeatability figure does not quantify the simulated dynamic tracking error, and the authored USD effort limits are not independent proof of permissible hardware torque. Forward-kinematics-derived Tool0 positions also do not measure the minimum clearance of the moving payload or characterize physical deformation. The experiments do not establish repeated-trial confidence intervals, an unrestricted permissible payload interval, formal closed-loop stability, or safe operation around people or equipment.

An eventual hardware implementation would require verification of actual robot and tool loading, the manufacturer's supported control interface, startup and brake-release behavior, communication and drive latencies, protective stop behavior, collision geometry and swept volumes, and application-specific risk controls. The learned residual should not bypass the nominal controller's operational constraints or be activated outside a condition for which its behavior has been validated. Future development should prioritize payload-aware policy selection or adaptation, explicit suppression of unnecessary residual effort, wider trajectory and mass sampling, joint-resolved counterfactual experiments, and robust supervisory handling of policy failure. Such work would move the architecture from a fixed-task demonstration toward dependable variable-payload operation.

9 Conclusions

This investigation combined analytical UR5e modeling, a physically motivated Cartesian transportation task, nominal-controller screening, independently executed ROS 2–PhysX validation, and residual reinforcement learning under systematically varied payload conditions. The selected computed-torque PD configuration achieved low matched-mass error and supplied an interpretable feedforward–feedback foundation. Under a heavier-than-modeled 3→5-kg condition, residual PPO reduced PhysX Cartesian RMSE from 10.560 to 6.061 mm. The same learned policies worsened tracking under matched and lighter actual loads, showing that additional learned torque is not universally beneficial. A separately trained reverse PPO reduced error from 8.039 to 4.585 mm under 5→3-kg mismatch, and the intermediate 4-kg comparison revealed distinct phase-dependent behavior between the two complete trained controllers.

The principal contribution is therefore not that reinforcement learning replaces conventional manipulation control. It is that a structured nominal controller can be supplemented by bounded learned effort under selected dynamic uncertainty, and that the benefits and limitations of this correction can be evaluated through physical torque interpretation, controlled payload variation, and separate execution-level validation. The practical direction is to preserve accurate nominal behavior and enable additional compensation only when the operating condition supports it. Further safety, robustness, and hardware validation remain necessary before industrial deployment.

References

References

[1] International Organization for Standardization, *ISO 10218-1:2025: Industrial Robots* and *ISO 10218-2:2025: Industrial Robot Applications and Robot Cells*, 2025.

[2] M. W. Spong, "An Historical Perspective on the Control of Robotic Manipulators," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, 2022. doi:10.1146/annurev-control-042920-094829.

[3] J.-J. E. Slotine and W. Li, "On the Adaptive Control of Robot Manipulators," *International Journal of Robotics Research*, vol. 6, no. 3, pp. 49–59, 1987.

[4] T. Johannink et al., "Residual Reinforcement Learning for Robot Control," *IEEE International Conference on Robotics and Automation*, pp. 6023–6029, 2019.

[5] Universal Robots A/S, *UR5e Technical Specification*, November 2023.

[6] P. K. Khosla and T. Kanade, "Experimental Evaluation of Nonlinear Feedback and Feedforward Control Schemes for Manipulators," *International Journal of Robotics Research*, vol. 7, no. 1, 1988.

Code and experimental records: `https://github.com/greatsyong/ur5e-control-analysis`