Multi-Modal Non-Prehensile Estimation of Physical Parameters via Press-and-Pull Tipping

Steven M. Hyland, Jing Xiao, and Cagdas D. Onal

Worcester Polytechnic Institute
smhyland@wpi.edu

Abstract. Recovering an object's inertial and frictional parameters without a grasp is a long-standing problem in non-prehensile manipulation. Pushing alone cannot disentangle mass from friction, while pure tipping is friction-agnostic and breaks down for low-friction objects that slide before they tip. We present a multi-modal non-prehensile estimation framework that recovers mass m, center of mass height z_c, and surface friction μ_t by sequencing two complementary interaction modes that factor the parameter set into observable pieces. The torque-balance equations at the pivot recover (m, z_c) cleanly, with friction structurally absent from the model. The known m is then divided out of the cached sliding-phase product to recover μ_t. We validate the framework on an ABB IRB120 with a wrist-mounted F/T sensor across four objects spanning rigid prisms, irregular consumer items, and low-friction contacts. The framework reliably recovers all three parameters without prior knowledge of geometry, mass, or friction.

1 Introduction

Robots increasingly operate in environments where objects are too large, fragile, or unwieldy to be grasped: lifting a monitor by an edge, nudging a toolbox into position, or rebalancing a teetering package on a shelf. In all of these tasks, safe manipulation hinges on the robot knowing a small set of physical parameters: mass, center of mass (CoM), and the object-surface friction coefficient μ_t. Together, these parameters determine when an object will tip, slide, or remain stable. None of these quantities can be reliably read from vision alone.

A natural alternative is to estimate these parameters *through interaction*. Non-prehensile primitives such as pushing and tipping require no grasp planning, no mechanical fixturing, and no prior geometric model, and they apply to objects that are not easily graspable. No single primitive, however, recovers all three parameters: pushing and tipping are each informative about one part of the parameter set and uninformative about the other, and tipping in particular breaks down for low-friction objects that slide before they tip. We review these limitations in detail in Section 1.1.

We exploit the fact that these limitations are complementary: the parameter combination one mode confounds, the other resolves. We also implement a means of precise tipping through the *press-and-pull* interaction.

1.1 Related Work

Inertial Parameter Estimation Through Interaction Inertial parameter identification has a long history in robotics, from rigid-body dynamic identification of manipulators to in-hand estimation of grasped payloads [9,18]. Mavrakis and Stolkin [17] provide a comprehensive survey, organizing methods along three axes: the assumed contact model (rigid grasp, pinch grasp, point contact), the sensor modality (joint torque, F/T, tactile, vision), and whether estimation is passive or active. Most existing techniques assume a stable grasp [9,18], which limits their applicability to objects that cannot be securely held. Our framework operates in the non-prehensile regime, treating the object–table contact as a passive contributor to the wrench balance rather than something to be eliminated.

Pushing-Based Estimation Quasistatic planar pushing has been studied extensively as an active-perception primitive [1,14,16,20]. Some approaches learn pushing dynamics for control but treat the underlying parameters as latent [20], while others explicitly characterize the identifiability problem: in the quasistatic regime, a small number of pushes provides mass–friction pairs, retaining a probability distribution over both [1]. The Force Push controller sidesteps the identifiability issue by closing the loop on force feedback alone, never explicitly recovering parameters [3]. Others fuse vision and tactile sensing within a differentiable filter to actively select informative pushes, achieving joint inference of friction, mass, CoM, and inertia, but require learned object–robot interaction models [2]. Recent research has explored mobile robot pushing parameter estimation [5], which is later extended to onboard sensing without relying on motion capture [6]. Our approach is complementary: rather than try to disentangle m and μ_t within a single quasistatic interaction, we activate modes sequentially to reveal the parameters each mode is suited for.

Friction Estimation Friction estimation has historically been pursued through sliding interactions [1, 13], lateral pressing with tactile sensing [10], or slip-detection control loops on grasped objects [18]. Similar works actively command motion until tactile slip is detected, then identify μ at the slip onset [18]. Slippery-terrain methods for legged systems use proprioceptive ground-reaction estimates to infer terrain friction [8]. Our approach borrows the slip-onset principle but transposes it to a non-prehensile setting.

Tipping and Toppling Tipping has been studied as both a manipulation goal in its own right, e.g., for object reorientation [14], and as an estimation tool. Early tipping-based estimators rely on explicit shape models or vision-based pose tracking through the toppling threshold [15], both of which fail on transparent or texture-less objects. One similar line of work is estimating CoM using Gravity Equi-Effect Planes [19] for basic shape primitives through multiple tip interactions across unique axes. This work was later extended for round-base objects [21], though for both, object shape is known a priori, which our approach is not limited by.

Multi-Modal and Multi-Contact Estimation A growing body of work fuses multiple modalities or interaction modes for estimation. Visuo-tactile pipelines

[2,10] combine RGB-D shape priors with tactile force readings. Robust transport methods [4] explicitly account for inertial uncertainty rather than estimating it. Differentiable physics-informed world models [11] jointly identify multiple physical parameters through few-shot pushing trajectories. In contrast, our approach is deliberately minimalist: a single coordinated primitive, an analytical wrench-balance model, and decoupled sub-problems with closed-form structure. We do not require a learned model, shape priors, or multiple repeated interactions.

1.2 Contributions

This paper makes the following contributions:

- **A complementary multi-modal estimation framework.** We show that planar sliding and pivot tipping are individually insufficient but jointly sufficient for recovering (m, z_c, μ_t): sliding observes only the coupled product $\mu_t \cdot m$, while tipping recovers (m, z_c) with friction structurally absent from the torque balance. We recover (m, z_c) by propagating the full measured F/T wrench to the pivot frame via the adjoint transform, using proprioception alone and bypassing error-prone contact-point localization. Sequencing the two modes and dividing the cached slip product by the tipping-recovered mass yields all three parameters in closed form, with no second sliding interaction, learned model, shape prior, or repeated probing.
- **The press-and-pull primitive.** A coordinated contact primitive that raises the object–table sliding threshold through a downward press while a horizontal pull induces tipping, mechanically pinning the pivot through frictional engagement. This treats the object's own support surface as a *passive grasp*, enabling reliable tipping of low-friction and curved-base objects that standard forward tipping can only cause to slide or swirl. An accompanying adaptive press-selection loop converges to a valid press force without any prior knowledge of the surface friction.
- **Demonstration across a challenging object range.** We show the framework recovers all three parameters, without per-object tuning, on objects that defeat prior single-mode methods: spanning 0.24–5.0 kg, wide and low-friction bases, and irregular every day geometry. Notably, it resolves objects with a curved base.

2 Methodology

This section provides background on standard tipping, sliding, and the multi-modal framework proposed by this research.

2.1 Background: Forward Tipping

Tipping-based estimation recovers (m, z_c) from the torque balance at a fixed pivot edge $\{O\}$ [7,19]:

$$\left({}^{O}\mathbf{p}_{push} \times {}^{O}\mathbf{F}_{push}\right) + \left({}^{O}\mathbf{p}_{c} \times {}^{O}\mathbf{F}_{grav}\right) = \mathbf{0}, \tag{1}$$

from which (m, z_c) follow while μ_t drops out entirely: tipping is friction-agnostic by construction. Two limitations remain unresolved. First, μ_t is unrecoverable from any pure tipping interaction. Second, for smooth or low-friction objects, sliding precedes tipping for all admissible push geometries, making (m, z_c) unrecoverable as well. This paper resolves both.

Why Pure Tipping Fails for Smooth Objects
A limitation not addressed by [7] and [19] is *tipping feasibility*: whether the object is guaranteed to tip without slipping, often maximized by pushing at the top of the object.

Recall the sliding and tipping constraints [12]:

$$F_{slide} = \mu_t N \qquad F_{tip} = \frac{\|\mathbf{p}_c \times \mathbf{F}_{grav}\|}{(\|\mathbf{p}_{push}\| \sin \alpha)}, \quad \alpha : \angle(\mathbf{p}_{push}, \mathbf{F}_{tip}) \tag{2}$$

The object begins at rest and force is applied gradually, and the relative ordering $F_{tip} \gtrless F_{slide}$ of these two thresholds determines the resulting motion.

For smooth-bottomed objects, $F_{slide} < F_{tip}$ may hold for *all* admissible push heights, so sliding always precedes tipping. In this regime no choice of push geometry can recover (m, z_c).

2.2 Mode 1: Sliding for Friction Estimation

To estimate the object-surface friction coefficient μ_t, the object must be excited so that this parameter becomes identifiable. The sliding mode probes the force balance at slip onset and observes the coupled product $\mu_t \cdot m$.

The fingertip approaches the object at a low height, with a line of action through the a-priori known (x_c, y_c), which avoids tipping out of the support surface and avoids swirling about the support surface normal (pure translation). The end-effector moves at low constant speed while F/T measurements are taken. Guaranteed sliding is non-trivial, too: some objects have high enough friction that sliding without tipping is infeasible. For these "unslideable" objects, friction can be treated as *maximal* and is of little concern for downstream manipulation tasks. This basic manipulation primitive is well-documented in the literature [13, 16, 20, 22].

2.3 Mode 2: Press-and-Pull for CoM and Mass Estimation

To enforce tipping over sliding, we propose a contact primitive called *press-and-pull*, illustrated in Figure 1. Driving the constraint $F_{tip} < F_{slide}$ required for tipping admits two levers: *decrease* F_{tip} or *increase* F_{slide} as highlighted by Eq. (2). In practice, the robot has affordance over the finger contact point, $\mathbf{p}_{app}$, and, to some extent the applied force $\mathbf{F}_{app}$. However, decreasing F_{tip} means raising $\mathbf{p}_{push}$, which is bounded by the object height. Conversely, increasing F_{slide} instead demands either a larger friction μ_t (such as changing the support surface) or raising normal force N. We favor the latter, since the joint selection

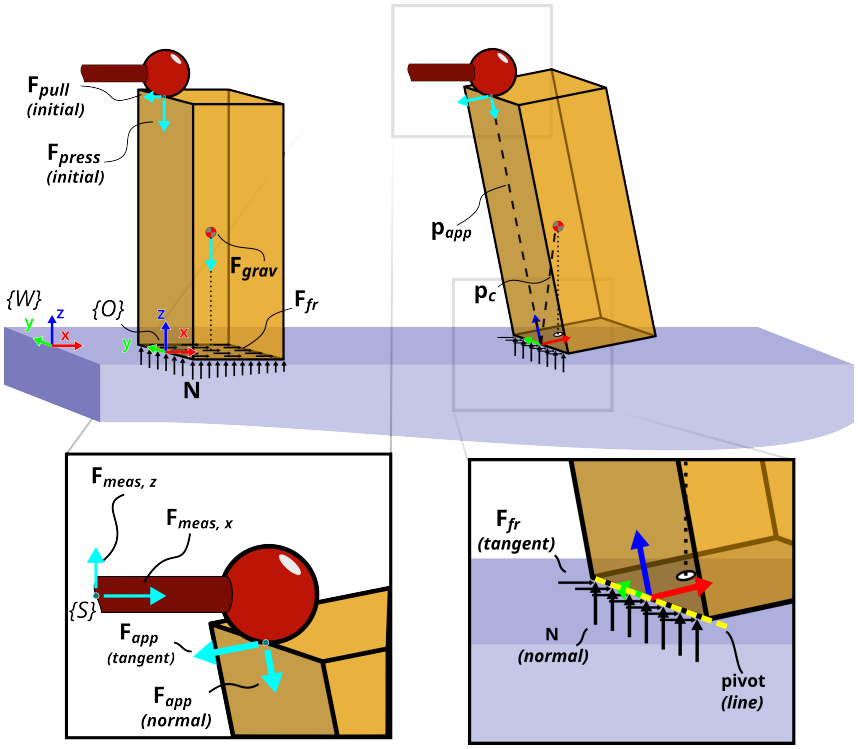

Fig. 1: The press-and-pull primitive. A downward press raises the sliding threshold at the object–table interface, while a horizontal pull induces tipping about the near-robot pivot edge.

of $\mathbf{p}_{push}$ and applied force modulates the support normal force directly without having to physically move the object to a high-friction support surface:

Press: a downward force F_{press} on the top face of the object raises normal force at object-table contact $N{=}mg{+}F_{press}$, and consequently elevates the sliding threshold to $F_{slide} = \mu_t(mg + F_{press})$.

Pull: a horizontal force F_{pull} is the primary driver of tipping about the near pivot edge of the object-surface contact.

Assuming $\mathbf{F}_{press}$ yields no-slip at all contacts, an arc trajectory is generated, centered at the pivot $\{O\}$ and in the pivot outwards-normal direction. This way, the press force can be maintained throughout tipping, no-slip can be satisfied, and the pivot can be stabilized, allowing for tipping of curved-base objects, unlike [19].

This primitive has the added benefit of causing the pivot to occur about the near-robot side of the object, which is often more observable than the aft-side. The press component fixes the pivot through frictional engagement and selecting

F_{press} such that $F_{tip} < F_{slide}$ guarantees tipping regardless of the (unknown) surface friction.

Adaptive Press Selection
Selection of F_{press} should satisfy three constraints:

1. $F_{tip} < F_{slide}$ with a margin sufficient to absorb model uncertainty.
2. $|\mathbf{F}_{press}|$ should be sufficient to maintain *no-slip* between finger and object.
3. $|\mathbf{F}_{press}|$ should not be too large such that deformation of the top face occurs. It should also not be too large such that a large *resistive, anti-tipping* moment is produced by the press about the pivot, which occurs under certain values of $\mathbf{p}_{push}$.

Since the bounds of these constraints are not identifiable until after the tipping procedure, selecting an appropriate press force is non-trivial.

Illustrated in Figure 2, we propose an *adaptive press selection* that converges to a press force that satisfies all three constraints. From a small nominal value for $|\mathbf{F}_{press}|$, force/torque measurements and object pose estimate are observed. If the initial press-and-pull fails to produce pure tipping or yields slip at the finger, press force is increased by some margin δ_{margin}, in this case multiplicative. The procedure repeats until clean tipping is observed. The upper bound for $|\mathbf{F}_{press}|$ is set by the shear strength of the rigid finger, though other bounds may be substituted, such as robot joint limits.

In practice, we initialize $F_{press} = 5\,\mathrm{N}$ and augment by $\delta_{margin} = 1.25$ per attempt.

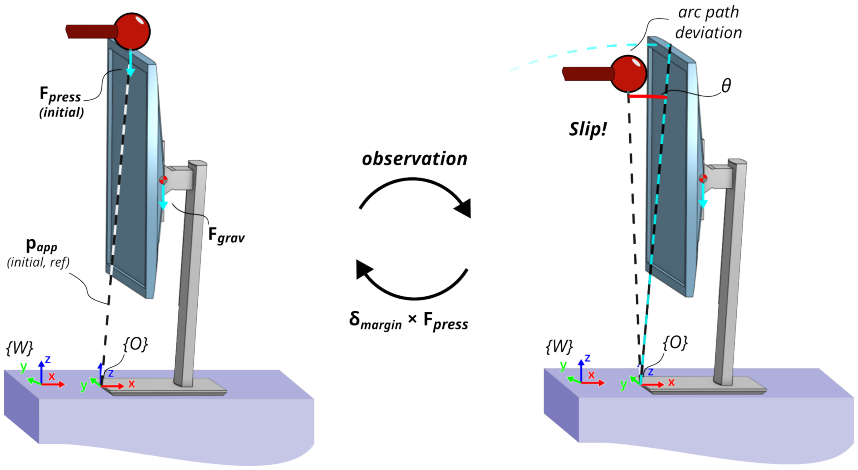

Fig. 2: Adaptive press procedure. (right) Slip is detected by deviation from the arc motion and observed vs. expected object pose. Press force that is ultimately found remains constant through the arc for all object angles θ (from proprioception).

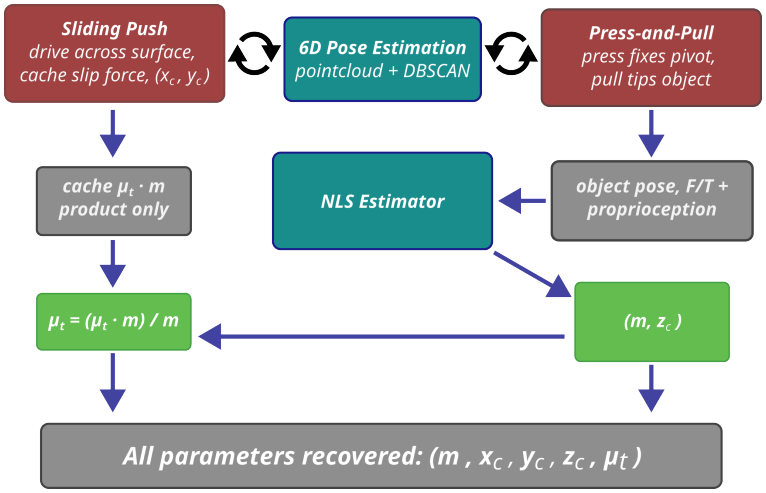

Fig. 3: Multi-modal pipeline. Pose estimation used to detect finger slip and helps with contact point selection. Friction estimation relies on output from NLS (non-linear least squares) estimation.

Finger orientation
Two end-effector strategies were considered during tipping: a fixed world-aligned orientation, and one that tracks the arc tangent throughout. The world-aligned finger produces a traveling contact patch on the spherical fingertip, introducing model error that is difficult to characterize due to fingertip compliance. The arc-tangent strategy fixes the contact patch in the object frame, eliminating this source of error at the cost of reduced workspace for short objects (the wrist approaches the table as tipping progresses). We adopt the arc-tangent strategy throughout, and the traveling contact point is solved by the wrench adjoint of Eq. (3).

Hysteresis Cancellation
The push–retract friction cancellation strategy of [7] carries over: during press-and-pull, the return-to-rest reverses the friction direction at the finger–object contact, producing a hysteresis structure in the measured torque–angle trajectory. Combining both segments in the regression cancels frictional bias in the torque fit.

2.4 Inertial Parameter Estimation Model

Because the pivot is fixed by frictional engagement, a quasistatic wrench-balance analysis is employed. The applied force now carries a large vertical (press) component, so we work directly with the full wrench. Reading the F/T sensor wrench as ${}^{S}\mathbf{w}_{\text{meas}} = [{}^{S}\mathbf{f}_{\text{meas}},\ {}^{S}\boldsymbol{\tau}_{\text{meas}}]^{T}$, we map it to the object frame $\{O\}$ via the adjoint

transform to yield the *applied wrench at {O}*:

$$ {}^{O}\mathbf{w}_{app} = -\mathrm{Ad}_{T_{SO}}\, {}^{S}\mathbf{w}_{\mathrm{meas}} = -\begin{bmatrix} {}^{O}\mathbf{R}_S & \mathbf{0} \\ [{}^{O}\mathbf{p}_{S/O}]\,{}^{O}\mathbf{R}_S & {}^{O}\mathbf{R}_S \end{bmatrix} \begin{bmatrix} {}^{S}\mathbf{f}_{\mathrm{meas}} \\ {}^{S}\boldsymbol{\tau}_{\mathrm{meas}} \end{bmatrix} \tag{3} $$

where ${}^{O}\mathbf{R}_S$ and ${}^{O}\mathbf{p}_{S/O}$ are the orientation (matrix) and position of the sensor frame in $\{O\}$, and $[\,\cdot\,]$ is the skew-symmetric (cross-product) matrix. The rotation and position of the sensor (encoded in transform matrix T_{SO}), follow from robot proprioception together with pivot location, known from the intersection of the object convex hull and table plane.

A subtle but important consequence of Eq. (3) is that the adjoint transform propagates the full measured wrench, directly to $\{O\}$ using only the sensor pose T_{SO}. This bypasses the need for explicit contact point calculation (i.e., $\mathbf{p}_{app}$ in $\{O\}$), which would otherwise require extrapolation from the robot kinematic chain and object pose estimate. Since T_{SO} is derived directly from proprioception, this formulation is more accurate than relying on contact point localization.

The complete quasistatic equilibrium at the pivot gives:

$$ {}^{O}\mathbf{w}_{app} + {}^{O}\mathbf{w}_{grav} + {}^{O}\mathbf{w}_{gnd} = \mathbf{0}, \tag{4} $$

with

$$ {}^{O}\mathbf{w}_{grav} = \begin{bmatrix} m\,{}^{O}\mathbf{g}(\theta) \\ {}^{O}\mathbf{p}_c \times \left(m\,{}^{O}\mathbf{g}(\theta)\right) \end{bmatrix}_{6\times 1}, \qquad {}^{O}\mathbf{w}_{gnd} = \begin{bmatrix} [\mathbf{f}_t]_{2\times 1} \\ N_t \\ [0]_{3\times 1} \end{bmatrix}_{6\times 1} \tag{5} $$

where θ is the object out-of-plane rotation (not about table $\hat{\mathbf{z}}$), and $[\mathbf{f}_t]_{2\times 1}$ and N_t are the friction and normal force at table contact, respectively. The ground wrench has zero moment since all ground reaction forces act through the pivot edge $\{O\}$. The moment rows of Eqs. (4) and (5) are therefore independent of μ_t and contain only the unknowns m and z_c, which we recover via the tipping and nonlinear least-squares pipeline.

Friction Estimation

The task is to determine the friction coefficient, μ_t. By the force balance (first three rows) of Eq. (4), the table reaction balances the applied and gravitational load; resolving it onto the table-tangent and table-normal directions gives the friction and normal contact forces

$$ f_t = -\left|\left({}^{O}\mathbf{f}_{app} + m\,{}^{O}\mathbf{g}(\theta)\right)\cdot\hat{\mathbf{v}}_t\right|, \qquad N_t = -\left|\left({}^{O}\mathbf{f}_{app} + m\,{}^{O}\mathbf{g}(\theta)\right)\cdot\hat{\mathbf{n}}_t\right|, \tag{6} $$

where $\hat{\mathbf{n}}_t$ is the table normal and $\hat{\mathbf{v}}_t$ is the in-plane unit vector along the direction of incipient slide. Since mass m is known from the tipping mode, the normal force N can be calculated leaving tangential friction force as unknown. Sliding (mode 1) induces a measurable tangential reaction, bringing friction to its limit-of-stability or its *Coulomb limit* [12] and making μ_t observable. At slip onset, friction reaches this limit (Eq. (2)), so $\mu_t = f_t/N$ follows directly from the measured contact forces.

Note that the slip measurement only observes the product of mass and friction, which cannot separate μ_t from m. We resolve it across modes: the press-and-pull tipping mode recovers m independently from the (friction-free) torque balance, and the cached slip measurement then yields μ_t with no second sliding interaction.

2.5 Estimation Process

As illustrated in Figure 3, parameter recovery proceeds through two sequential interactions:

1. **Planar push**: drive the object across the surface at low height. Cache the slip-onset tangential force, which observes the product $\mu_t \cdot m$. The same interaction also yields planar CoM components (x_c, y_c) via planar pushing techniques [5].
2. **Press-and-pull tipping**: press fixes the pivot; a slow pull tips the object sub-critically. Torque balance recovers (m, z_c) and the cached product then yields $\mu_t = f_t \,/\, mg$.

Both interactions share the following execution protocol:

1. Segment the pointcloud and select a contact point appropriate for the interaction mode.
2. Translate the finger at constant velocity; record the full wrench trajectory.
3. Detect slip or tip onset; verify no finger slip occurred. Return object to rest.
4. Apply a median plus Savitzky–Golay filter cascade to suppress sensor noise.
5. Estimate parameters according to interaction mode using filtered data.

3 Experiments

Physical experiments use an ABB IRB120 industrial manipulator with a rigid 3D-printed finger providing point contact. Figure 4 shows the heart object under the press-and-pull primitive. Shown in Figure 5, the fingertip is a high-friction rubber sphere to aid the press-and-pull mode. A wrist-mounted ATI Gamma six-axis Net F/T sensor measures the full interaction wrench at 500 Hz. Object pointcloud is recovered using an Intel RealSense D435i depth camera. All computation is performed on an Intel Core i7-13700H processor.

3.1 Object Pose Estimation

Two complementary techniques are implemented for object pose estimation.

Pointcloud pose estimation A fixed RealSense depth camera streams the pointcloud in $\{W\}$ via an extrinsic calibration, which is cropped to the workspace and voxel-downsampled for uniform density. Object clusters are extracted with DBSCAN and the corresponding 3D convex hulls are calculated, which provides coordinates for contact point selection during both interaction modes. The convex hull pose is also used for finger slip detection during adaptive press selection.

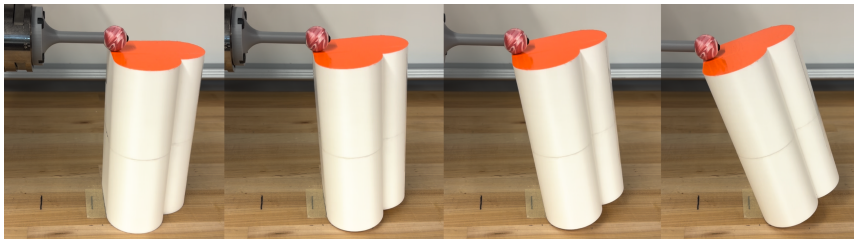

Fig. 4: Heart object during the press-and-pull mode. The finger first applies a downwards press force. Upon reaching the press force, the finger retracts and applies a pull force, which continues while the finger maintains the press force throughout tipping. This continues until the measured tangential force falls to the defined safety margin η_{safety}.

Fig. 5: (left) Robot finger used for both interaction modes. High-friction spherical tip provides better traction during press-and-pull and reduces required press force magnitude. (right) four objects tested, varying in size, shape, weight, friction, and CoM.

Proprioceptive pose estimation The object angle and pivot axis are estimated from robot proprioception during tipping. Provided no finger slip, the rigid object follows the trajectory enforced by the finger. During the arc motion, an accurate and smooth estimation of the object's pose is measured, as well as the primary tipping axis.

3.2 Push Point and Press Point Selection

Figure 6 illustrates the optimal and realistic selection of these contact points and the corresponding pointcloud.

Push point The push point selection is relatively simple: choose the lowest-height point of the object on any of its faces that allows free motion of the end effector.

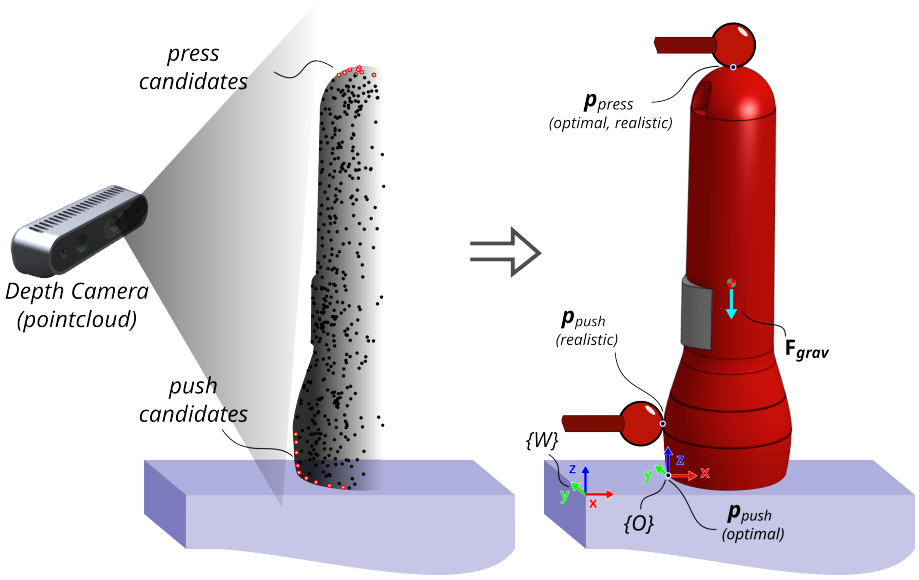

Fig. 6: Pointcloud and contact point selection. Candidate points taken from the pointcloud near the extremes of object height, and selected according to feasibility and the heuristics defined in Sec. 3.2. An optimal choice for either interaction mode may not be possible, in which case the next best feasible candidate is selected.

Press point The press point selection is crucial for stable tipping: select a point on the object top whose face normal approximates world $+\hat{z}$ (upwards) and minimizes the press force lever-arm ($\min \mathbf{p}_{app,\,x}$), often directly above $\{O\}$. These combined constraints are crucial towards a press force with minimally-induced torque about $\{O\}$. The pull is then directed towards the pivot edge outward normal.

3.3 Interaction Trajectory Generation

Trajectories follow the protocol of Section 2, with interaction velocities of 5 mm/s throughout. The arc-motion of press-and-pull continues until the measured pull force reaches the sub-critical safety threshold $\eta_{safety} = 0.1$ of [7] or in other words, until measured tangential force reaches 10% of its peak value at pull onset. The trajectory is then reversed to execute the retract phase symmetrically.

3.4 Physical Experiments

We test our approach on four unique objects shown in Figure 5: a hollow acrylic box, a 3D-printed heart-shaped prism, a handheld flashlight, and a computer monitor. Ground-truth (m, z_c) values are taken from independent experiments,

Table 1: Estimation results on physical objects.

	Ground Truth			Estimated			Relative Error		
Object	m (kg)	z_c (cm)	μ_t	$\hat{m}$ (kg)	$\hat{z}_c$ (cm)	$\hat{\mu}_t$	$\frac{\lvert\Delta m\rvert}{m}$	$\frac{\lvert\Delta z_c\rvert}{z_c}$	$\frac{\lvert\Delta \mu_t\rvert}{\mu_t}$
Box	0.676	15.00	0.185	0.709	15.38	0.176	4.9%	2.5%	4.9%
Heart	0.239	10.00	0.254	0.267	11.29	0.225	11.7%	12.9%	11.4%
Flashlight	0.387	9.38	0.293	0.414	10.03	0.276	7.0%	6.9%	5.8%
Monitor	5.040	23.20	0.638	5.325	23.76	0.607	5.6%	2.4%	4.9%

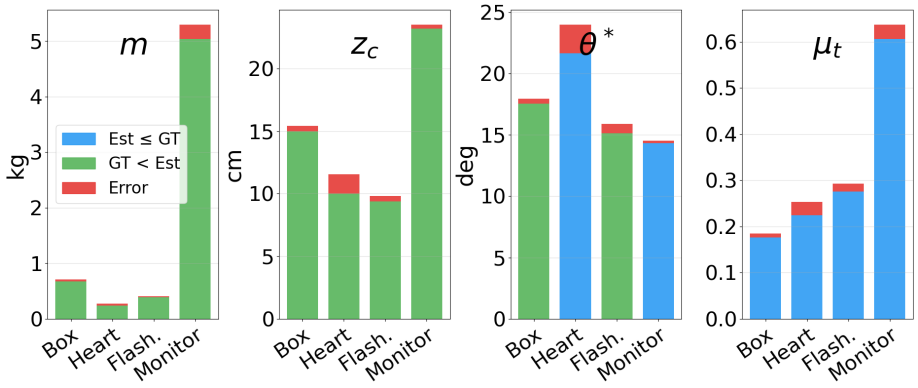

Fig. 7: Absolute error plots: ground truth versus estimated values for each object's parameters. Each individual bar is colored blue if that parameter was *underestimated*, and colored green if *overestimated*, with the error in red.

while ground-truth (μ_t, θ^*) are calculated from mass and CoM values for each object and validated independently.

Planar CoM components (x_c, y_c) were provided by our prior estimation pipeline [5] rather than re-estimated here, as validating that method is outside the scope of this work.

4 Results and Discussion

Table 1 summarizes estimation results across four objects. For the four rigid objects (box, heart, flashlight, monitor), all three parameters are recovered with relative errors under 13%, and three of the four objects achieve errors below 8% across all parameters. Figure 7 shows the absolute errors per object for each parameter and Figure 8 shows the corresponding relative errors. We refer the reader to the supplemental video for further information.

Estimator consistency. The push-retract structure follows the friction cancellation strategy of [7]: reversing the arc trajectory flips the friction direction at the

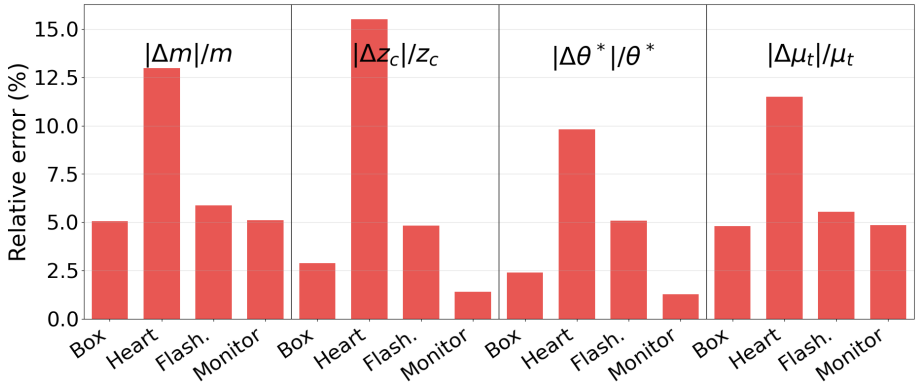

Fig. 8: Relative error plots: Per-object, per-parameter absolute error over ground truth. This metric provides a geometry and parameter-agnostic measure of the quality of the proposed multi-modal estimation framework.

finger-object contact, so arc and retract segments carry frictional bias with opposite signs. Arc and retract phases are estimated independently and averaged; the two agree to within $\Delta m < 0.26\,\text{kg}$, $\Delta z_c < 11\,\text{mm}$, and $\Delta\theta^* < 0.63°$ across all rigid objects, confirming that the wrench-balance model is well-conditioned across the tipping range and insensitive to phase direction under the rigid-body assumption.

Friction estimation. Friction errors for all rigid objects are below 11.5% (below 5.9% besides the heart), and the monitor correctly reports the highest $\hat{\mu}_t = 0.607$, consistent with its rubber feet. Friction error scales with mass error through $\hat{\mu}_t = f_t^{\text{slip}}/(\hat{m}\,g)$, though mass errors are small enough that this amplification remains benign.

Heart object. The heart prism shows the largest error among rigid objects ($\sim 12\%$ in $m, z_c,$ and μ_t). Two compounding factors degrade estimation quality: its low mass (239 g) produces a small absolute torque signal, and its wide, smooth base resists tipping, driving the press:pull force ratio much higher than for other objects. The arc/retract estimates agree closely ($\Delta m = 9\,\text{g}$, $\Delta z_c = 0\,\text{mm}$), confirming the error is systematic rather than noisy. Lightweight objects with wide, smooth bases represent the hardest case for this estimator; multiple press-and-pull trials would be the most direct path to improvement.

Flashlight: resolving a known failure mode. The flashlight is a documented failure case for pure forward tipping [7]: its low-friction curved base causes unwanted swirling about the world $+z$ axis. The press-and-pull primitive resolves this directly by mechanically pinning the pivot through the press force, yielding 7.0%, 6.9%, and 5.8% relative errors in m, z_c, and μ_t respectively, comparable to the best-performing rigid objects.

Monitor: large-object robustness. At 5.04 kg and a total height of 0.53 m, the monitor tests estimator robustness at the upper end of the payload range, yielding errors of 5.6% in mass, 2.4% in z_c, and 4.9% in μ_t, among the best performers across all objects and parameters. This accuracy is attributed to the heavier monitor weight, yielding a large, clean, and low-noise torque signal during tipping.

5 Conclusions

We presented a multi-modal non-prehensile framework that recovers an object's mass m, center-of-mass height z_c, and surface friction μ_t without a grasp, a shape prior, or a learned interaction model. Sequencing a planar sliding push with a press-and-pull tipping interaction yields all three parameters in closed form. The press-and-pull primitive is the enabling component: pinning the pivot through a controlled press lets the robot tip smooth- and curved-base objects that standard "forward" tipping can only cause to slide or swirl. Across four objects spanning rigid prisms, irregular consumer items, and low-friction contacts, the framework recovered all three parameters with relative errors under 13% (under 8% for three of four) and resolved the curved-base flashlight, which is infeasible under standard tipping.

One limitation is that the press action assumes a well-defined, visible, and reachable contact on the top face. Standard forward tipping does not share this requirement, since it may tip the object at any reachable, visible height; for such objects, forward tipping is preferable when feasible.

Future Work. Three natural extensions follow:

1. Relaxing the rigid-body assumption to model quasistatic fluid redistribution in partially filled containers, which may additionally enable fluid-height estimation.
2. Repeating the press-and-pull interaction over multiple trials to improve conditioning on lightweight, wide-based objects.
3. Incorporating an explicit, criticality-aware safety margin into the tipping trajectory, specifically to detect object-table slippage which may occur due to the press force. This would improve safety across all object classes.

References

1. Bauza, M., Rodriguez, A.: A probabilistic data-driven model for planar pushing. In: 2017 IEEE International Conference on Robotics and Automation (ICRA). pp. 3008–3015. IEEE, Singapore, Singapore (May 2017). https://doi.org/10.1109/ICRA.2017.7989345, http://ieeexplore.ieee.org/document/7989345/
2. Dutta, A., Burdet, E., Kaboli, M.: Push to Know! - Visuo-Tactile Based Active Object Parameter Inference with Dual Differentiable Filtering. In: 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 3137–3144. IEEE, Detroit, MI, USA (Oct 2023). https://doi.org/10.1109/IROS55552.2023.10341832, https://ieeexplore.ieee.org/document/10341832/

3. Heins, A., Schoellig, A.P.: Force Push: Robust Single-Point Pushing With Force Feedback. IEEE Robotics and Automation Letters **9**(8), 6856–6863 (Aug 2024). https://doi.org/10.1109/LRA.2024.3414180, https://ieeexplore.ieee.org/document/10556602/
4. Heins, A., Schoellig, A.P.: Robust Nonprehensile Object Transportation With Uncertain Inertial Parameters. IEEE Robotics and Automation Letters **10**(5), 4492–4499 (May 2025). https://doi.org/10.1109/LRA.2025.3551067, https://ieeexplore.ieee.org/document/10925467/
5. Hyland, S.M., Xiao, J., Onal, C.D.: Predicting Center of Mass by Iterative Pushing for Object Transportation and Manipulation. In: 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 1615–1620. IEEE, Detroit, MI, USA (Oct 2023). https://doi.org/10.1109/IROS55552.2023.10341534, https://ieeexplore.ieee.org/document/10341534/
6. Hyland, S.M., Xiao, J., Onal, C.D.: Onboard Sensing and Pushing of Unknown Payload for CoM Estimation with a Holonomic Mobile Robot. In: 2025 IEEE 21st International Conference on Automation Science and Engineering (CASE). pp. 2555–2561. IEEE, Los Angeles, CA, USA (Aug 2025). https://doi.org/10.1109/CASE58245.2025.11163849, https://ieeexplore.ieee.org/document/11163849/
7. Hyland, S.M., Xiao, J., Onal, C.D.: Before the Tipping Point: Force-Guided Active Perception for Shape-Agnostic Estimation of 3D Centers of Mass. In: 2026 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Under Review. (Mar 2026)
8. Kim, H., Kang, D., Kim, M.G., Kim, G., Park, H.W.: Online Friction Coefficient Identification for Legged Robots on Slippery Terrain Using Smoothed Contact Gradients. IEEE Robotics and Automation Letters **10**(4), 3150–3157 (Apr 2025). https://doi.org/10.1109/LRA.2025.3541428, https://ieeexplore.ieee.org/document/10884016/
9. Kubus, D., Kroger, T., Wahl, F.M.: On-line rigid object recognition and pose estimation based on inertial parameters. In: 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 1402–1408. IEEE, San Diego, CA, USA (Oct 2007). https://doi.org/10.1109/IROS.2007.4399184, http://ieeexplore.ieee.org/document/4399184/
10. Le, T.N., Verdoja, F., Abu-Dakka, F.J., Kyrki, V.: Probabilistic Surface Friction Estimation Based on Visual and Haptic Measurements. IEEE Robotics and Automation Letters **6**(2), 2838–2845 (Apr 2021). https://doi.org/10.1109/LRA.2021.3062585, https://ieeexplore.ieee.org/document/9364673/
11. Li, W., Zhao, H., Yu, Z., Du, Y., Zou, Q., Hu, R., Xu, K.: PIN-WM: Learning Physics-INformed World Models for Non-Prehensile Manipulation. arXiv (2025). https://doi.org/10.48550/ARXIV.2504.16693, https://arxiv.org/abs/2504.16693, version Number: 2
12. Lynch, K.M., Park, F.C.: Modern Robotics: Mechanics, Planning, and Control. Cambridge University Press, 1 edn. (May 2017). https://doi.org/10.1017/9781316661239, https://www.cambridge.org/core/product/identifier/9781316661239/type/book
13. Lynch, K., Maekawa, H., Tanie, K.: Manipulation And Active Sensing By Pushing Using Tactile Feedback. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. vol. 1, pp. 416–421. IEEE, Raleigh, NC (1992). https://doi.org/10.1109/IROS.1992.587370, http://ieeexplore.ieee.org/document/587370/

14. Lynch, K., Mason, M.: Dynamic underactuated nonprehensile manipulation. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS '96. vol. 2, pp. 889–896. IEEE, Osaka, Japan (1996). https://doi.org/10.1109/IROS.1996.571070, http://ieeexplore.ieee.org/document/571070/
15. Maeda, Y., Arai, T.: Planning of graspless manipulation by a multifingered robot hand. Advanced Robotics **19**(5), 501–521 (Jan 2005). https://doi.org/10.1163/156855305323383776, https://www.tandfonline.com/doi/full/10.1163/156855305323383776
16. Mason, M.T.: Mechanics and Planning of Manipulator Pushing Operations. The International Journal of Robotics Research **5**(3), 53–71 (Sep 1986). https://doi.org/10.1177/027836498600500303, https://journals.sagepub.com/doi/10.1177/027836498600500303
17. Mavrakis, N., Stolkin, R.: Estimation and exploitation of objects ' inertial parameters in robotic grasping and manipulation: A survey. Robotics and Autonomous Systems **124**, 103374 (Feb 2020). https://doi.org/10.1016/j.robot.2019.103374, https://linkinghub.elsevier.com/retrieve/pii/S0921889019302313
18. Sundaralingam, B., Hermans, T.: In-Hand Object-Dynamics Inference Using Tactile Fingertips. IEEE Transactions on Robotics **37**(4), 1115–1126 (Aug 2021). https://doi.org/10.1109/TRO.2020.3043675, https://ieeexplore.ieee.org/document/9325054/
19. Yong Yu, Fukuda, K., Tsujio, S.: Estimation of mass and center of mass of graspless and shape-unknown object. In: Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C). vol. 4, pp. 2893–2898. IEEE, Detroit, MI, USA (1999). https://doi.org/10.1109/ROBOT.1999.774036, http://ieeexplore.ieee.org/document/774036/
20. Yu, K.T., Bauza, M., Fazeli, N., Rodriguez, A.: More than a million ways to be pushed. A high-fidelity experimental dataset of planar pushing. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 30–37. IEEE, Daejeon, South Korea (Oct 2016). https://doi.org/10.1109/IROS.2016.7758091, http://ieeexplore.ieee.org/document/7758091/
21. Yu, Y., Kiyokawa, T., Tsujio, S.: Estimation of mass and center of mass of unknown cylinder-like object using passing-CM Lines. In: Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No.01CH37180). vol. 3, pp. 1788–1793. IEEE, Maui, HI, USA (2001). https://doi.org/10.1109/IROS.2001.977237, http://ieeexplore.ieee.org/document/977237/
22. Zhong, Z., Golestaneh, S., Chamzas, C.: ActivePusher: Active Learning and Planning with Residual Physics for Nonprehensile Manipulation (2025). https://doi.org/10.48550/ARXIV.2506.04646, https://arxiv.org/abs/2506.04646, version Number: 4