SourcePhys: Source-Conditional Physical Completion for Simulation-Ready 3D Assets

Anonymous authors
Paper under double-blind review

Abstract

Large 3D collections contain assets in very different states of physical readiness. Generated meshes may lack physical structure, while native simulation assets may already provide bodies, joints, and valid parameters. We present SourcePhys, a framework for source-conditional physical completion that preserves available evidence and infers unresolved attributes. It selects a part representation from the source, combines constrained vision-language inference with material priors and component geometry, and arbitrates direct and geometry-based mass estimates when they disagree. The completed physical records retain the origin of inferred values and can be exported to OpenUSD for simulation. We evaluate 775 assets from four source cohorts. On 150 randomly sampled Amazon Berkeley Objects, the full method reaches 43.6% median mass error; 743 assets have an original or repaired variant that passes cohort-specific OpenUSD and Isaac Sim checks. Controlled behavior probes and robot interactions further illustrate how the completed assets enter interactive workflows. These results offer a practical route from heterogeneous 3D content to physically grounded assets for embodied AI.

1 Introduction

Robot-learning environments need objects that respond to contact and manipulation, not only objects that look plausible. Generation and retrieval systems supply abundant 3D geometry, but an asset's physical state depends on its source. A generated mesh may contain only appearance; a grouped model may expose useful parts but no dynamics; an articulated OpenUSD file may already define bodies, joints, collisions, and some numerical parameters. These are not interchangeable inputs to a single "add physics" operation. Replacing all source information loses valid structure, whereas preserving everything leaves incomplete assets incomplete.

Recent systems connect generated assets to robot simulation and learning (Katara et al., 2024; Cao et al., 2025a;b; Yang et al., 2026), while scalable frameworks use generated or reconstructed environments for policy evaluation (Ranawaka et al., 2026; Wang et al., 2026; Jangir et al., 2026). Yet physical completion remains sensitive to what the input already contains. Source hierarchies may be more reliable than a new segmentation, while the appearance of a hollow or composite object may not reveal the material volume needed for mass estimation. Vision-language models (VLMs) provide semantic and mass cues; geometry provides measurable shape. Their errors differ, and an asset is useful in simulation only when the inferred values remain tied to coherent bodies, collisions, and interactions.

We formulate this problem as source-conditional physical completion. The method retains usable source physics, recovers parts from groups or visual evidence when needed, and estimates missing attributes from semantic, material, and geometric cues. A disagreement guard chooses between direct VLM and geometry-based mass estimates when the geometry is unreliable. The resulting physical records are exported to OpenUSD and tested in Isaac Sim. We evaluate numerical mass agreement, preservation of source structure, and execution in controlled and robot-interaction settings.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Our contributions are:

- Source-conditional completion: an evidence precedence rule preserves valid source-authored physics and selects part evidence for estimating only missing fields across heterogeneous inputs.
- Disagreement-aware mass derivation: structure-conditioned geometry and material priors provide one estimate; a deterministic guard arbitrates it with a direct VLM estimate. On 150 ABO assets, the full method reaches 43.6% median absolute percentage error.
- Traceable executable output: OpenUSD authoring records the origin and fallback of inferred values, followed by structural checks, controlled behavior probes, and robot interaction cases.

2 Related Work

Simulation-ready assets. Methods such as Gen2Sim, PhysX-3D, PhysX-Anything, and PhysForge connect 3D content with physical properties and robot interaction (Katara et al., 2024; Cao et al., 2025a;b; Yang et al., 2026). Real2Code reconstructs articulated parts and joint programs from images (Zhao et al., 2025); PhysPart completes geometry for interactable parts under motion constraints (Luo et al., 2025). UniPhysGen jointly learns articulation and intrinsic physical properties from heterogeneous assets (Li et al., 2026), while EmbodiedGen packages generated content with collision geometry and affordances (Wang et al., 2025a; 2026). Work on dynamic representations also infers material behavior from visual evidence (Lin et al., 2025; 2026). Our focus is complementary: the source may already contain a physical hierarchy, so completion must distinguish valid authored evidence from properties that still require inference.

Robot simulation and evaluation. SAPIEN provides part-aware interactive environments (Xiang et al., 2020), and Isaac Gym and Isaac Sim enable scalable GPU physics and robot-learning workflows (Makoviychuk et al., 2021; NVIDIA, 2024). AdaManip studies articulated-object manipulation environments (Wang et al., 2025b), while DreMa combines explicit scene representations with physics simulation for robot learning (Barcellona et al., 2025). At scene scale, UrbanVerse and SAGE-3D build physics-aware environments for embodied navigation (Liu et al., 2026; Miao et al., 2026). SimFoundry studies modular scene generation for policy learning and evaluation (Ranawaka et al., 2026); RobotArena Infinity uses real-to-sim translation for policy benchmarking (Jangir et al., 2026). These systems motivate the executable-asset checks used here, although our method operates at the level of individual assets.

Part and physical representations. Part-level geometry provides a carrier for material, inertia, contact, and affordance annotations. GAPartNet connects actionable part perception with manipulation (Geng et al., 2023). Segment Anything supplies image masks (Kirillov et al., 2023), recent 3D methods lift visual cues to meshes (SAM 3D Team et al., 2025), and GraspGen predicts six-degree-of-freedom grasp candidates (Murali et al., 2025). Our part representation follows the available source evidence, from authored rigid bodies to mesh groups and, when necessary, multi-view segmentation. Physical properties remain ambiguous even with a part decomposition: visually similar objects can differ in internal structure and density. We therefore retain material uncertainty and the evidence behind decisions between direct and geometry-based mass estimates.

3 Problem Formulation

The input is an asset $\mathcal{A}$ with geometry and views, and possibly mesh groups, a part hierarchy, or source-authored physical fields. The task is to choose a physical-part representation and complete missing fields while retaining valid source evidence. The completed record contains body and collision structure, scale, mass, contact parameters, applicable joints,

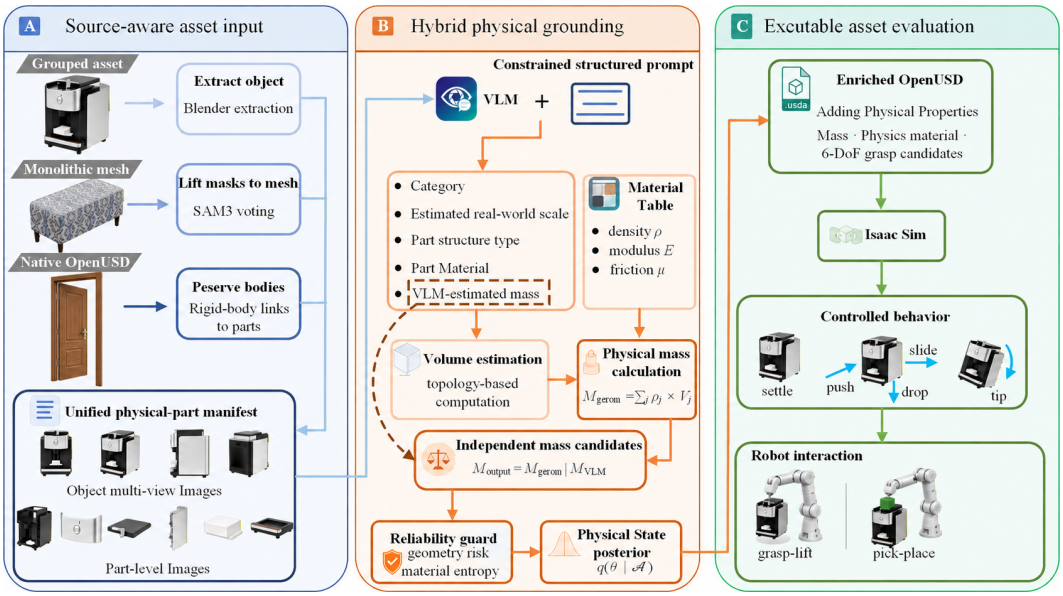

Figure 1: Source-conditional physical completion, from heterogeneous inputs to physical inference and downstream robot use.

and the evidence behind each inferred value. We export these records to OpenUSD for simulation. Center of mass and inertia are estimated in the audit; on some routes, the simulator derives executable inertials from collision geometry rather than explicit authoring. Evaluation examines parameter error, structural validity, controlled behavior, and robot-task integration.

4 Assets and Evaluation Protocol

4.1 Asset Sources and Evaluation Splits

The evaluation covers three source regimes. Retrieved visual assets contain textured geometry but little or no physics. Generated or monolithic assets require visual part discovery. Native simulation assets provide rigid-body, joint, collision, or mass annotations that can be retained or used as source-authored references in evaluation copies.

The current processed corpus contains 775 assets across four cohorts. The retrieved-mesh cohort contains 350 Amazon Berkeley Objects (ABO) (Collins et al., 2022). We randomly sample 150 assets with available reference mass and evaluate all methods on the same assets and shared part preprocessing. For metric alignment, the VLM estimates a real-world characteristic length; the ratio to the corresponding mesh extent gives a uniform asset scale. ABO listing dimensions are not used during inference, and listing mass is withheld until evaluation. The generated cohort contains 16 externally generated 3D assets. Each has a physical audit, OpenUSD export, and grasp annotation. The purely visual OpenUSD cohort contains 157 assets from object_sort; these sources lack rigid-body and joint definitions and are processed through the Blender part frontend. Finally, the native-physics cohort contains 7 single-rigid-body assets from a SimReady validation set and 245 articulated assets (243 from object_sort and 2 from SimReady). The 243 object_sort assets follow routine source-preserving completion. The nine SimReady assets instead use evaluation-only copies: source physical values are retained as references, while predicted mass and density are written to the copies. Original source files remain unchanged.

All 775 assets have an audit record, an original OpenUSD export, and an Isaac Sim validation report. Table 5 distinguishes those original exports from separately verified repair copies. We provide a browsable record of processed assets, visual previews, and inferred physical

properties at https://meshsimlab.xyz/. The catalog is maintained as a growing interface for inspecting these outputs.

Figure 2 shows the input sources, inferred labels, and output-field coverage. Reported mass coverage includes both predictions and preserved source values; accuracy is evaluated separately below.

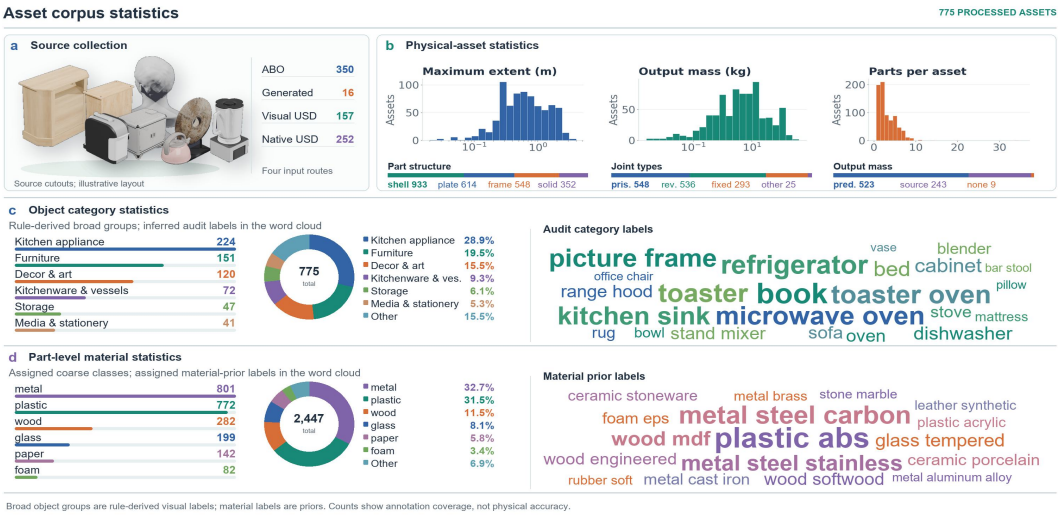

Figure 2: Processed asset corpus. (a) Input sources; (b) physical-asset statistics; (c) inferred object labels; (d) assigned material priors. Word size indicates frequency; counts summarize annotation coverage.

4.2 References and Downstream Protocols

Reference annotations include ABO listing mass, weak listing-level material metadata, and source-authored physics in native OpenUSD. Of 252 native assets, 249 have at least one explicit mass, but none has an explicit center of mass or inertia (Appendix E). These source-provided values support consistency and preservation checks; ABO material metadata provide a separate diagnostic comparison.

Controlled behavior is examined through settle, drop, slide, and push/tip probes on ten native OpenUSD assets. Source and enriched variants share geometry and initial conditions, with one execution per condition at a 1/60 s physics step. The probes measure responses under specified interventions, with settings given in Appendix B.

Robot integration is illustrated with ten randomly sampled pick-and-place cases from AutoMate, AssetTotal, ObjectSort, and SimReady. Four use Franka Panda with RMPFlow and six use UR10e with cuRobo. Candidate-level records cover planning, contact, lift, transport, and placement for movable objects. Each case has one recorded run (Appendix B).

5 Source-Conditional Physical Completion

Figure 1 shows how source evidence guides physical completion. The source hierarchy determines the part representation; visual inference and material priors then supply hypotheses for unresolved attributes. Component geometry provides a mass estimate independent of the VLM's direct prediction. The method resolves disagreements before export and retains both estimates, the selected value, and any fallback in the audit.

5.1 Evidence-Preserving Part Selection

We add physical properties only where source values are missing or invalid. Let $\boldsymbol{\theta}_s$ denote source-authored physical fields and $\mathbf{m}_s$ their validity mask. For scalar and tensor-valued fields, the authored output follows

$$\boldsymbol{\theta}_{\text{out}} = \mathbf{m}_s \odot \boldsymbol{\theta}_s + (\mathbf{1} - \mathbf{m}_s) \odot \hat{\boldsymbol{\theta}}, \tag{1}$$

where $\hat{\boldsymbol{\theta}}$ is inferred by the shared backend. For discrete structures such as bodies, joints, and collision hierarchies, the same rule is implemented as precedence rather than numerical interpolation: valid source structure is retained, and inferred structure is introduced only when no usable physical hierarchy exists. The SimReady estimation benchmark explicitly masks source mass and density in a separate prediction copy while retaining their original values as references.

The frontend route $r(\mathcal{A}) \in \{\text{USD}, \text{group}, \text{visual}\}$ is selected by an ordered evidence test:

$$r(\mathcal{A}) = \begin{cases} \text{USD}, & \text{valid}(\mathcal{B}_s, \mathcal{J}_s, \mathcal{C}_s), \\ \text{group}, & \text{usable}(\mathcal{H}, \mathcal{M}), \\ \text{visual}, & \text{otherwise}, \end{cases} \tag{2}$$

where $\mathcal{B}_s$, $\mathcal{J}_s$, and $\mathcal{C}_s$ are source bodies, joints, and collisions, while $\mathcal{H}$ and $\mathcal{M}$ denote object and material groups. This ordering avoids replacing strong physical evidence with a weaker appearance-derived segmentation.

In a native OpenUSD asset with usable PhysicsRigidBodyAPI, joints, and collision shapes, each rigid body defines a physical part and its descendant meshes are aggregated. Where the physical hierarchy is absent, meaningful object or material groups are extracted through Blender. For monolithic meshes, multi-view masks are projected to faces and combined by confidence-weighted voting, followed by neighborhood smoothing and connected-component cleanup.

All routes emit a common manifest containing source route, part mesh, aligned views, transforms, and source metadata. Fine visual fragments may be coalesced into non-destructive physics groups for physical estimation, while the original segmentation remains available for inspection.

5.2 Constrained Semantic Inference

Aligned global and local views provide the VLM with context for each object and part. Constrained outputs describe category, textual evidence, structural class $z \in \{\text{solid}, \text{shell}, \text{thin plate}, \text{frame}\}$, material hypotheses, and confidence. The VLM also predicts object-level mass and maximum real-world extent $\hat{\ell}$. The former serves as an independent candidate for mass arbitration, not as an input to geometric volume. When source scale is unavailable, the ratio of $\hat{\ell}$ to the maximum unscaled mesh extent ℓ_{mesh} gives a uniform scale $s = \hat{\ell}/\ell_{\text{mesh}}$ for all parts. Material-dependent physical constants come from an engineering prior table rather than unconstrained VLM text. The table provides distributions for density, static and dynamic friction, and restitution; structural and material classes provide wall-thickness priors.

For an uncertain material distribution with probabilities p_k, the density moments are

$$\bar{\rho} = \sum_k p_k \mu_k, \tag{3}$$

$$\sigma_\rho^2 = \sum_k p_k (\sigma_k^2 + \mu_k^2) - \bar{\rho}^2, \tag{4}$$

where (μ_k, σ_k) are material-prior parameters. Effective composites use normalized fractions and retain the assumed mixing rule. Contact-property moments are aggregated analogously and clamped to physically valid ranges.

5.3 Geometry-Conditioned Physical Derivation

Disconnected components are filtered and estimated separately rather than enclosed in one hull (Appendix A). For component j, material volume is selected according to its structural class:

$$V_j = \begin{cases} V_{\text{hull},j} r_{\text{fill},j}, & z = \text{solid}, \\ \frac{1}{2} A_j t_j r_{\text{cov},j}, & z \in \{\text{shell}, \text{plate}\}, \\ \min(\frac{1}{2} A_j r_j, \eta V_{\text{bbox},j}), & z = \text{frame}, \end{cases} \tag{5}$$

where A_j is mesh surface area, t_j is a wall-thickness prior, $r_{\text{fill},j}$ is occupancy, and $r_{\text{cov},j}$ discounts openings. Invalid or costly geometry triggers a bounded box or surface fallback; the selected estimator is recorded as provenance.

Mass and first-order uncertainty are $m_j = V_j \bar{\rho}_j$ and $\sigma_{m_j} = V_j \sigma_{\rho_j}$. For retained components, object mass, center of mass, and inertia are

$$M = \sum_j m_j, \qquad \mathbf{c} = M^{-1} \sum_j m_j \mathbf{c}_j, \tag{6}$$

$$\mathbf{I}_c = \sum_j \left[\mathbf{I}_j + m_j \left(\|\mathbf{d}_j\|^2 \mathbf{1} - \mathbf{d}_j \mathbf{d}_j^\top \right) \right], \tag{7}$$

where $\mathbf{d}_j = \mathbf{c}_j - \mathbf{c}$. Invalid tensors fall back to a positive box inertia and are marked as such.

5.4 Disagreement Guard and Mass Arbitration

Let G and L denote geometry-derived and direct-VLM mass. Large disagreement alone need not favor either estimate: a hollow or fragmented mesh can inflate G, while an appearance-based estimate may miss dense internal components. Our ordered guard selects L when disagreement is accompanied by geometry-risk or thin/flexible-part evidence, or exceeds a larger unconditional threshold; otherwise it retains G. Its exact thresholds and precedence are given in Appendix A. The guard is a deterministic decision rule based on the two estimates and mesh-risk indicators.

5.5 Provenance-Aware OpenUSD Authoring

Authoring proceeds on a copy of the source asset. Routine native completion preserves valid source mass and density, although documented structural repairs may change the copy. SimReady estimation uses separate prediction copies in which mass and density are intentionally replaced while the source files remain unchanged. Inferred values are assigned to the corresponding source rigid body; visual-only assets receive a rigid-body root, collision approximation, mass properties, and physics materials. Standard OpenUSD schemas store executable values (Alliance for OpenUSD, 2026), while namespaced fields retain material distributions, confidence, estimation source, fallback reason, and audit identifiers. GraspGen candidates (Murali et al., 2025) are transformed into the source coordinate frame and associated with a body. Each export is accompanied by its geometry, textures, and a machine-readable audit report.

Audit inertials are not always authored: the default native-USD route lets the simulator derive them from collisions (Appendix A).

6 Experiments

6.1 Evaluation Design

We evaluate the pipeline from physical inference to executable use. RQ1 (mass estimation) compares direct VLM, geometry-and-material, and full-method mass estimates on the same assets. RQ2 (mass arbitration) isolates the contribution of the reference-free disagreement guard beyond the geometry-and-material estimate. RQ3 (asset execution) examines whether

completed physical records preserve source structure and support OpenUSD validation and simulation in Isaac Sim, with controlled behavior and robot interaction assessed separately.

We compare five mass estimates on the same assets:

- Direct-VLM: object-level numerical prediction from global and part views, without part-wise volume computation;
- Fixed-density geometry: every retained component is treated as solid with density fixed at $1000\,\mathrm{kg/m^3}$;
- No-structure volume: predicted material density is retained, but every component is forced to be solid;
- Geometry+material: structure-conditioned component volume combined with material-density priors, without the disagreement guard;
- Full method: the source-aware hybrid pipeline with mass arbitration;

OpenUSD validity and source preservation are evaluated separately from mass error.

Mass metrics are median absolute percentage error (APE), median absolute log-ratio, and the proportions within 20%, 50%, and a factor of two of the ABO listing reference. The paired factor-of-two comparison uses an exact test on discordant outcomes. Median-APE differences use 10,000 paired bootstrap resamples of asset IDs with replacement for 95% confidence intervals. Behavioral probes report execution and response detection per probe. Robot interaction cases report first-candidate success, completion within the recorded Top-K budget, attempt counts, and failure stage. The two robot configurations are reported separately.

6.2 Mass Estimation on Randomly Sampled ABO Assets

The paired ABO results are reported in Table 1, with shared preprocessing and inference settings in Appendix A.

Table 1: Mass estimation on 150 ABO assets. Geometry variants use VLM-estimated characteristic length for metric alignment.

Method	N	Med. APE ↓	Med. $\lvert\log_{10} r\rvert$ ↓	20% ↑	50% ↑	2× ↑
Direct-VLM	150	49.0%	0.233	17.3%	52.0%	62.7%
Fixed-density	150	158.9%	0.764	4.0%	12.7%	18.7%
Material, all-solid	150	292.2%	0.855	8.7%	16.0%	20.0%
Geometry+material	150	46.5%	0.253	22.7%	54.7%	60.0%
Full method	150	43.6%	0.221	24.7%	61.3%	66.7%

Structure-conditioned volume is the main source of improvement: it reduces median APE from 292.2% for the all-solid material baseline to 46.5% (paired Wilcoxon $p < 10^{-17}$). The guard keeps the geometry estimate for 129 assets and selects Direct-VLM for 21. Relative to Geometry+material, these selections improve 19 cases and worsen two, reducing median APE by 2.9 percentage points (paired-bootstrap 95% CI $[-7.7, -0.3]$).

The full method reaches 43.6% median APE, 5.4 points below image-only Direct-VLM (95% CI $[-12.0, -0.1]$; paired Wilcoxon $p = 0.022$). Geometry variants obtain metric scale from the VLM-estimated characteristic length. Factor-of-two accuracy is 66.7% versus 62.7% for Direct-VLM (paired discordant-outcome $p = 0.47$). The 95th-percentile APE is 96.0% for the full method and 298.6% for Direct-VLM (linear interpolation; Appendix Fig. 5).

6.3 OpenUSD Structure and Isaac Sim Executability

We next test whether the completed records form executable simulator assets. All 775 original stages opened and reached a physics step in Isaac Sim. Physics-schema checks passed for 770, and 718 also had complete local dependencies. The latter check includes visual files, so it is stricter than dynamic execution. Of the original assets, 762 settled under

the prescribed test. Cohort-specific checks passed for 734 exports: unarticulated assets were tested for settling, while articulated assets were also actuated. Nine separately verified repair copies passed the same protocol, bringing the count to 743 distinct asset IDs with a verified variant (Table 2).

Table 2: OpenUSD and Isaac Sim checks. Complete includes local dependencies; verified IDs count a passing original or repair copy.

Cohort	N	Schema	Complete	Original pass	Repair pass	Verified IDs
ABO	350	350	350	342	5	347
3D-generated	16	16	16	15	0	15
Visual OpenUSD	157	157	108	155	0	155
Native physics	252	247	244	222	4	226
Total	775	770	718	734	9	743

Among 245 articulated native assets, 217 pass the joint protocol; 244 have a controllable joint tested. The 26 original ObjectSort joint failures coincide with failures in their source USDs. In the native source comparison, collision paths and existing joint definitions are retained, whereas auxiliary-body removal and SimReady reference-value replacement are explicitly recorded (Table 6).

Figure 3 illustrates the physical attributes produced for one native microwave.

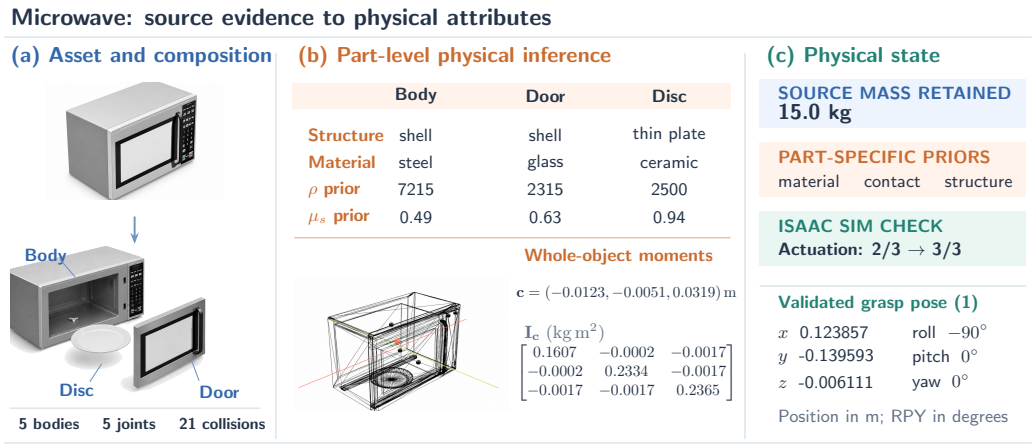

Figure 3: Native microwave case. (a) Asset and decomposed parts; (b) part priors and audit-derived whole-object moments; (c) retained source mass, Isaac Sim actuation, and one validated grasp pose. Moments are calculated from audited part masses in the extracted-mesh frame.

6.4 Controlled Behavior in Isaac Sim

All 40 probes on the ten enriched native assets completed with finite state and without escape (Table 3). Each asset settled and showed a detectable drop and slide response. Five responded to the weight-scaled push, three tilted beyond 45°, and median slide distance was 0.0227 m. Source and enriched runs exist for all ten assets, but the source CoffeeMachine fails the structural-comparison check. Paired comparison therefore comprises nine assets.

6.5 Robot Interaction Case Studies

The ten pick-and-place cases span four source groups (Table 4). Six succeed with the first grasp candidate, and all ten complete contact, lift, transport, and placement within four candidates. Across 18 attempts, failures comprise four trajectory-execution failures, two

pregrasp-reach failures, and two slips during transport. Four AutoMate plugs use Franka Panda with RMPFlow; the others use UR10e with cuRobo. Figure 4 shows two recorded sequences. Further case details are in Appendix B.

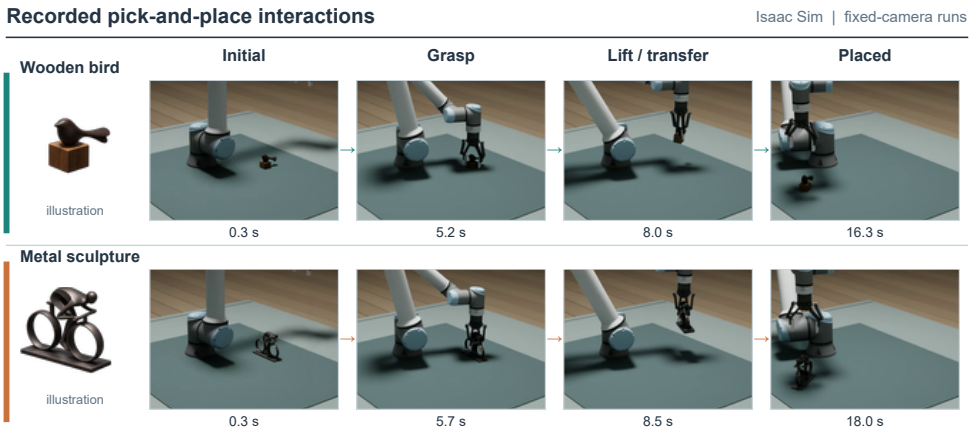

Figure 4: Two fixed-camera Isaac Sim pick-and-place runs. Object illustrations appear at left; recorded simulation frames show the task sequence.

7 Discussion

The experiments show the value of combining visual and geometric evidence: component structure and material priors improve aggregate mass estimation, while source-aware authoring carries the resulting records into simulation. Hollow interiors, wall thickness, and hidden materials remain important sources of uncertainty; contact priors also vary with the contacting pair and solver.

The disagreement guard reduces aggregate error and records its selection for each asset. Direct-VLM has lower error on 58 of the 150 evaluated assets, indicating where richer geometric evidence or calibrated priors may help.

OpenUSD and PhysX checks in Isaac Sim establish an executable asset path across the four cohorts. Measured dynamic trajectories and cross-simulator tests are natural extensions of this evaluation, as is articulation inference for assets without source joints.

8 Conclusion

We presented source-conditional physical completion, a framework for turning generated, retrieved, and native 3D assets into physically grounded assets for simulation. The framework uses source evidence to preserve authored structure and parameters where available, recovers physical parts where they are absent, and estimates unresolved attributes through visual, material, and geometric reasoning. An explicit mass decision combines complementary estimates while recording the origin of each selected value. Evaluation on heterogeneous assets links improved mass estimation with structural preservation and executable simulation checks; controlled probes and robot interactions further illustrate how the resulting assets can be used. Together, these results establish a practical route from heterogeneous 3D content to traceable physical asset collections for embodied AI.

AI use statement

Generative AI tools assisted with manuscript organization and language editing and analysis-code review. The research pipeline itself uses a vision-language model as an explicitly described component for material and structural inference. The authors are responsible

for verifying all citations, equations, experimental configurations, numerical results, figure labels, and claims before submission; numerical results are drawn from recorded experiments.

Ethics statement

This work does not involve human subjects in its current form. Dataset and asset licenses must be respected when releasing benchmark examples, derived OpenUSD files, and textures. Physical-property estimates should not be treated as certified engineering measurements or used for safety-critical deployment without validation. Any future human annotation or preference study will follow the relevant institutional and informed-consent requirements.

Reproducibility statement

The intended release includes the frozen evaluation roster, source-processing manifests, material-prior versions, constrained VLM prompts and model identifiers, OpenUSD authoring code, Isaac Sim configurations, probe definitions, evaluation scripts, and per-asset audit reports where available. Appendix B specifies the minimum information required to reproduce each level of evaluation.

References

Alliance for OpenUSD. Openusd documentation and physics schemas. https://openusd.org/release/, 2026.

Leonardo Barcellona, Andrii Zadaianchuk, Davide Allegro, Samuele Papa, Stefano Ghidoni, and Efstratios Gavves. Dream to manipulate: Compositional world models empowering robot imitation learning with imagination. In International Conference on Learning Representations, 2025.

Ziang Cao, Zhaoxi Chen, Liang Pan, and Ziwei Liu. Physx-3d: Physical-grounded 3d asset generation. In Advances in Neural Information Processing Systems, 2025a.

Ziang Cao, Fangzhou Hong, Zhaoxi Chen, Liang Pan, and Ziwei Liu. Physx-anything: Simulation-ready physical 3d assets from a single image. arXiv preprint arXiv:2511.13648, 2025b.

Jasmine Collins, Shubham Goel, Kenan Deng, et al. Abo: Dataset and benchmarks for real-world 3d object understanding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.

Haoran Geng, Helin Xu, Chengyang Zhao, Chao Xu, Li Yi, Siyuan Huang, and He Wang. Gapartnet: Cross-category domain-generalizable object perception and manipulation via generalizable and actionable parts. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7081–7091, 2023.

Yash Jangir, Yidi Zhang, Kashu Yamazaki, Chenyu Zhang, Kuan-Hsun Tu, Tsung-Wei Ke, Lei Ke, Yonatan Bisk, and Katerina Fragkiadaki. Robotarena ∞: Scalable robot benchmarking via real-to-sim translation. In The Fourteenth International Conference on Learning Representations, 2026. URL https://arxiv.org/abs/2510.23571.

Pushkal Katara, Zhou Xian, and Katerina Fragkiadaki. Gen2sim: Scaling up robot learning in simulation with generative models. In IEEE International Conference on Robotics and Automation, 2024.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, et al. Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015–4026, 2023.

Xian Li, Rong Wei, Lujie Yang, Haolin Huang, Junyuan Fang, Siliang Tang, Jun Xiao, Rui Tang, and Juncheng Li. Uniphysgen: Unified physical grounding for simulation-ready 3d assets. arXiv preprint arXiv:2607.13586, 2026.

Jiajing Lin, Shu Jiang, Qingyuan Zeng, Zhenzhong Wang, and Min Jiang. Visionlaw: Inferring interpretable intrinsic dynamics from visual observations via bilevel optimization. In International Conference on Learning Representations, 2026.

Yuchen Lin, Chenguo Lin, Jianjin Xu, and Yadong Mu. Omniphysgs: 3d constitutive gaussians for general physics-based dynamics generation. In International Conference on Learning Representations, 2025.

Mingxuan Liu, Honglin He, Elisa Ricci, Wayne Wu, and Bolei Zhou. Urbanverse: Scaling urban simulation by watching city-tour videos. In International Conference on Learning Representations, 2026.

Rundong Luo, Haoran Geng, Congyue Deng, Puhao Li, Zan Wang, Baoxiong Jia, Leonidas Guibas, and Siyuan Huang. Physpart: Physically plausible part completion for interactable objects. In IEEE International Conference on Robotics and Automation, 2025. URL https://arxiv.org/abs/2408.13724.

Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, et al. Isaac gym: High performance gpu-based physics simulation for robot learning. arXiv preprint arXiv:2108.10470, 2021.

Bingchen Miao, Rong Wei, Zhiqi Ge, Xiaoquan Sun, Shiqi Gao, Jingzhe Zhu, Renhan Wang, Siliang Tang, Jun Xiao, Rui Tang, and Juncheng Li. Towards physically executable 3d gaussian for embodied navigation. In International Conference on Learning Representations, 2026.

Adithyavairavan Murali, Balakumar Sundaralingam, Yu-Wei Chao, Wentao Yuan, Jun Yamada, Mark Carlson, Fabio Ramos, Stan Birchfield, Dieter Fox, and Clemens Eppner. Graspgen: A diffusion-based framework for 6-dof grasping with on-generator training. arXiv preprint arXiv:2507.13097, 2025.

NVIDIA. Isaac sim: Robotics simulation and synthetic data generation. https://github.com/isaac-sim/IsaacSim, 2024.

Nadun Ranawaka, Josiah Wong, Wei-Lin Pai, et al. Simfoundry: Modular and automated scene generation for policy learning and evaluation. arXiv preprint arXiv:2606.28276, 2026.

SAM 3D Team, Xingyu Chen, Fu-Jen Chu, et al. Sam 3d: 3dfy anything in images. arXiv preprint arXiv:2511.16624, 2025.

Xinjie Wang, Liu Liu, Yu Cao, Ruiqi Wu, Wenkang Qin, Dehui Wang, Wei Sui, and Zhizhong Su. Embodiedgen: Towards a generative 3d world engine for embodied intelligence. arXiv preprint arXiv:2506.10600, 2025a.

Xinjie Wang, Liu Liu, Taojun Ding, Andrew Choi, Chaodong Huang, Mengao Zhao, Ziang Li, Jackson Jiang, Chunlei Yu, Shengxiang Liu, Wei Xu, and Zhizhong Su. Embodiedgen v2: An agentic, simulation-ready 3d world engine for embodied ai. arXiv preprint arXiv:2607.07459, 2026.

Yuanfei Wang, Xiaojie Zhang, Ruihai Wu, Yu Li, Yan Shen, Mingdong Wu, Zhaofeng He, Yizhou Wang, and Hao Dong. Adamanip: Adaptive articulated object manipulation environments and policy learning. In International Conference on Learning Representations, 2025b.

Fanbo Xiang, Yuzhe Qin, Kaichun Mo, et al. Sapien: A simulated part-based interactive environment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.

Yunhan Yang, Chunshi Wang, Junliang Ye, Yang Li, Zanxin Chen, Zehuan Huang, Yao Mu, Zhuo Chen, Chunchao Guo, and Xihui Liu. Physforge: Generating physics-grounded 3d assets for interactive virtual worlds. arXiv preprint arXiv:2605.05163, 2026.

Mandi Zhao, Yijia Weng, Dominik Bauer, and Shuran Song. Real2code: Reconstruct articulated objects via code generation. In International Conference on Learning Representations, 2025.

Supplementary Material

The supplement records the estimation settings, evaluation protocols, and asset-level accounting behind the main results.

Section	Contents
A	Mass-evaluation provenance, shared geometry, guard precedence, and authoring scope
B	Asset inclusion, mass metrics, OpenUSD checks, behavior probes, and robot-case settings
C	Asset flow, repair-copy accounting, and native-source preservation
D	Common manifest, audit fields, and OpenUSD mapping
E	Coverage of source-authored physical fields in native OpenUSD

A Implementation Details for Mass Evaluation

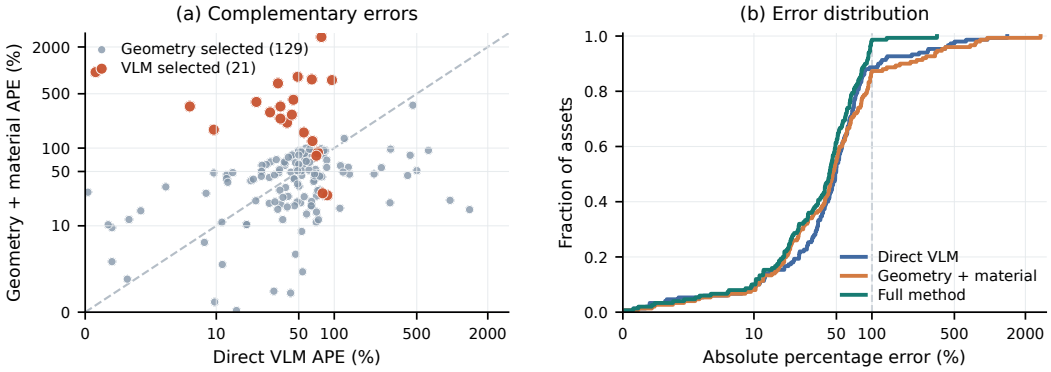

Figure 5: Paired mass-error analysis on 150 ABO assets. Left: Direct-VLM versus geometry-and-material APE; red points mark 21 guard selections. Right: empirical error CDF. Geometry-based methods use VLM-derived metric scale.

A.1 Mass Inference and Shared Geometry

The 150-asset analysis uses Ollama qwen3.5:122b (Q4_K_M), a 360-s timeout, 1,800 output tokens, structured JSON, and six parts per VLM batch. The primary VLM call uses temperature 0.1; JSON repair uses 0.0. The same prompts, material priors, structure-volume rules, component filtering, and mass guard are used throughout the evaluation. The VLM estimates the real-world characteristic length; its ratio to the unscaled mesh extent aligns the geometry estimates. ABO listing dimensions are not supplied to the VLM or used for alignment. Listing weight is withheld from inference and used only for scoring.

The five ABO mass variants are evaluated on the same 150 asset IDs. Geometry variants share VLM-derived metric alignment, part meshes, and retained connected components; the direct VLM candidate uses global and part images without part-wise volume computation. The final asset roster and per-asset result file each contain 150 unique IDs with an exact one-to-one match; all five mass outputs are present for every ID. The ablation evaluator keeps retained components separate. A component is major when its surface-area fraction is at least 0.01 or its face-count fraction is at least 0.005, and its bounding-box diagonal is at least 0.015 of the part diagonal. Non-major fragments within 0.025 of the part diagonal from a major component are retained separately; more distant fragments are discarded. If none is major, the original part mesh is the fallback. These values are the auditor defaults used in this evaluation.

For the solid-volume counterfactuals, each retained component receives its own convex hull. At most 10,000 vertices per component are selected at evenly spaced vertex indices before hull construction; an invalid hull falls back to the component bounding-box volume. The fixed-density baseline multiplies the sum of those volumes by $1000\,\mathrm{kg/m^3}$. The all-solid material baseline uses the same volumes with each part's inferred density. The geometry-and-material variant instead applies Eq. 5 using the inferred structural class, thickness, coverage, fill factor, and density. Shell and plate estimates use half the mesh area to account for paired surface sides. Frame radius is clipped to $[0.003, 0.025]$ m after multiplying the largest component extent by 0.008; frame volume is capped at 0.03 of the component bounding box. These are estimation assumptions, not measured material thicknesses or occupancies.

A.2 Ordered Mass Guard

Let $G > 0$ be the componentwise geometry-and-material estimate and $L > 0$ the independent direct-VLM estimate. Let r_g be the fraction of G assigned to parts with a geometry-failure flag, a risky convex-hull or safe-box volume source, or at least 30 raw connected components while labeled solid. Let r_t be the fraction assigned to parts whose identifier matches the thin/flexible semantic hints. The guard selects L at the first satisfied condition:

1. $G/L \geq 2$ and $r_g \geq 0.5$ (geometry risk);
2. $G/L \geq 3$ and $r_t \geq 0.5$ (thin/flexible part);
3. $G/L \geq 5$ (large overestimation);
4. $L/G \geq 3$ and at least one part has a recorded geometry failure.

Otherwise the guard retains G. The order matters because the conditions overlap. It never queries ABO reference mass. In the 150-asset evaluation, these branches account for 2, 2, 17, and 0 direct-VLM selections, respectively; the remaining 129 use G. The reported D_m is a diagnostic and is not itself the switching rule.

A.3 Mass and Inertial Authoring Scope

Mass arbitration selects an object-level scalar; it does not refit the part-wise center-of-mass or inertia estimates from observed motion. When explicit native-USD inertial authoring is enabled, valid part tensors are transformed to the source body frame, aggregated by the parallel-axis theorem, and uniformly scaled by $M_{\text{target}}/M_{\text{shape}}$. Uniform mass scaling leaves the aggregate center of mass unchanged. The tensor is symmetrized and diagonalized before OpenUSD principal axes and positive diagonal inertia are authored. This option is off by default in the available native-USD writer. Its ordinary completion writes a missing mass only when both source mass and density are absent; otherwise source values take precedence and the simulator may derive inertials from collision geometry. The available ABO single-rigid-body exporter writes the audit's recommended mass and convex-hull colliders without explicitly authoring center of mass or inertia. The separate ObjectSort visual-only USD writer has an additional 4× direct-VLM disagreement override; this is not the ABO guard evaluated above. The nine SimReady prediction copies follow their separately documented replacement protocol.

B Detailed Evaluation Protocol

B.1 Asset Inclusion and Reference Handling

The primary cohort is determined before physical inference. Assets are excluded only for documented source corruption, missing geometry, license restrictions, or inability to obtain the required reference annotation. Prediction error, VLM disagreement, and task failure cannot remove an asset from the evaluation cohort.

For ABO, listing dimensions are not used for scale alignment, and listing weight is reserved for evaluation. The inference configuration is given in Appendix A.

B.2 Mass Metrics

For asset i, let $\hat{m}_i > 0$ be the estimated mass and $m_i > 0$ the ABO listing reference. The absolute percentage and log-ratio errors are

$$e_i = 100\frac{|\hat{m}_i - m_i|}{m_i}, \qquad \ell_i = \left|\log_{10}\frac{\hat{m}_i}{m_i}\right|. \tag{8}$$

Table 1 reports the median of each error over the same 150 assets. The 2× rate is the fraction with $\frac{1}{2} \le \hat{m}_i/m_i \le 2$; the 20% and 50% rates use $e_i \le 20$ and $e_i \le 50$. For a paired comparison, asset IDs are resampled with replacement 10,000 times before taking the difference of median APEs. The reported interval is the 2.5th to 97.5th percentile of those differences.

B.3 OpenUSD Validity Checks

The RQ3 static scan checks stage opening, local file dependencies, rigid-body and collision presence, physics-material bindings, authored mass or density, and, where applicable, joint targets, types, and limits. Physics-schema completeness and the stricter conjunction with local dependencies are reported separately. Path-matched Isaac Sim reports establish load and physics stepping, settling, and joint actuation under the cohort-specific protocols.

B.4 Behavior-Probe Settings

The pilot reports a 1/60 s physics step, intervention magnitudes, and response measures. Reproducing each condition also depends on simulator and solver settings, contact and material settings, initial poses, run lengths, and response thresholds.

The pilot settles each asset from its initial pose and records time to rest and final motion. The drop probe releases it from 0.20 m and records contact, rebound, and re-settling. The slide probe applies an initial horizontal velocity of 0.50 m/s and records travel and stopping. The push/tip probe applies a force equal to 0.35 times the asset weight for 12 simulation steps and records displacement and tilt under a force proportional to asset weight. The paired analysis uses the nine structurally comparable source/enriched runs.

Table 3: Isaac Sim behavior pilot on ten enriched native assets, with one run per asset and probe.

Probe	Executed	Response	Additional observation
Settle	10/10	10/10 settled	Median time: 0.267 s
Drop	10/10	10/10 re-settled	0.20 m release
Slide	10/10	10/10 moved	Median distance: 0.0227 m
Push/tip	10/10	5/10 responded	3/10 tipped $> 45°$

B.5 Robot Interaction Configuration Sheet

Robot case records include the embodiment, gripper, controller, candidate attempts, task run identifier, and stage-specific failure reason where available. The ten cases come from four source groups: AutoMate (100 assets), AssetTotal (13 assets), ObjectSort (400 listed assets), and a separate SimReady pool. The pilot records one run per case under a fixed physical-parameter configuration. The recorded case IDs define this pilot; the complete source-pool draw lists and SimReady pool size were not archived.

C Asset Flow and Source Preservation

Table 5 counts source IDs, nonempty physical-audit records, original USD exports and tests, original Isaac passes, separately verified repair copies, and unique IDs with a verified variant. Repair copies are not reruns of the original production outputs. The online catalog provides access to the processed asset records.

Table 4: Ten sampled pick-and-place cases. First denotes first-candidate success; Top-K denotes completion within four candidates. One run per case.

Source group	Cases	First	Top-K	Attempts
AutoMate	4	1/4	4/4	9
SimReady	2	1/2	2/2	5
ObjectSort	1	1/1	1/1	1
Generated cubes	3	3/3	3/3	3
Total	10	6/10	10/10	18

Table 5: Asset flow by cohort. Repair passes are separately validated copies.

Cohort	Source	Audit	USD	Tested	Orig. pass	Repair	Verified IDs
ABO	350	350	350	350	342	5	347
3D-generated	16	16	16	16	15	0	15
Visual OpenUSD	157	157	157	157	155	0	155
Native physics	252	252	252	252	222	4	226
Total	775	775	775	775	734	9	743

Native-source preservation is evaluated on the 252 inputs that already have physics. Table 6 separates unchanged topology from documented repairs and evaluation-specific value replacement. The counts overlap and must not be summed; seven single-body assets have no joints.

Table 6: Source-to-output comparison for native physics OpenUSD; joint checks use the 245 articulated assets.

Check	Assets
Collision paths retained	252/252
Joint definitions retained	245/245
Full rigid/collision/joint topology unchanged	227/252
Documented auxiliary-body removals	25/252
Documented joint-drive normalization	209/252
Documented physics-material rebinding	243/252
Existing physical values unchanged	243/252
SimReady experimental value replacement	9/252
Unexpected changes	0/252

The nine SimReady replacements are evaluation-only copies: all nine replace source density and six replace an explicit source mass. The original files are unchanged. Twenty-six original ObjectSort joint failures were also observed on their source USDs; this explains neither the two other native joint failures nor the validity of any enriched output.

D Unified Asset and Audit Schema

Table 7 summarizes the common records emitted across frontends. Optional fields are retained when their producing stage runs.

The OpenUSD exporter maps these records to standard rigid-body, collision, mass, joint, and physics-material schemas. Auxiliary namespaced attributes preserve the audit information without changing the semantics of standard OpenUSD fields. Existing valid source-authored fields take precedence over inferred values in routine completion. The nine SimReady estimation copies are an explicit exception: density is replaced in all nine, and authored mass in six.

Table 7: Common manifest and audit fields; stored fields are not necessarily measured or explicitly authored into USD.

Record	Fields
Source manifest	Asset ID, frontend route, source metadata, metric alignment, part-mesh paths, views, and transforms
Object audit	Category and rationale, scale provenance, mass candidates, selected source, consistency ratio, geometry-risk fraction, and fallback
Part audit	Structure class, material distribution or composition, density and contact priors, volume and method, mass, center of mass, inertia and methods, component evidence, and confidence
Interaction	Optional grasp poses associated with the corresponding body

E Native OpenUSD Reference Coverage

Table 8 reports the reproducible scan of the 252 native physical OpenUSD assets. These values quantify source coverage and are not claimed to originate from physical measurement.

Table 8: Coverage of source-authored OpenUSD physics.

Source property	Assets	Evaluation use
At least one explicit mass	249/252	Reference consistency
Mass on all collision bodies	247/252	Strict mass subset
Mass or density on all bodies	250/252	Derived-mass subset
Explicit center of mass	0/252	Behavior/geometric checks
Explicit inertia	0/252	Behavior/geometric checks
Static and dynamic friction	252/252	Source preservation
Restitution	9/252	Small reference subset
Collision geometry	252/252	10,354 collision prims
Any bounded joint	244/245	Joint preservation
All movable joints bounded	212/245	Strict joint subset