Massachusetts Institute of Technology

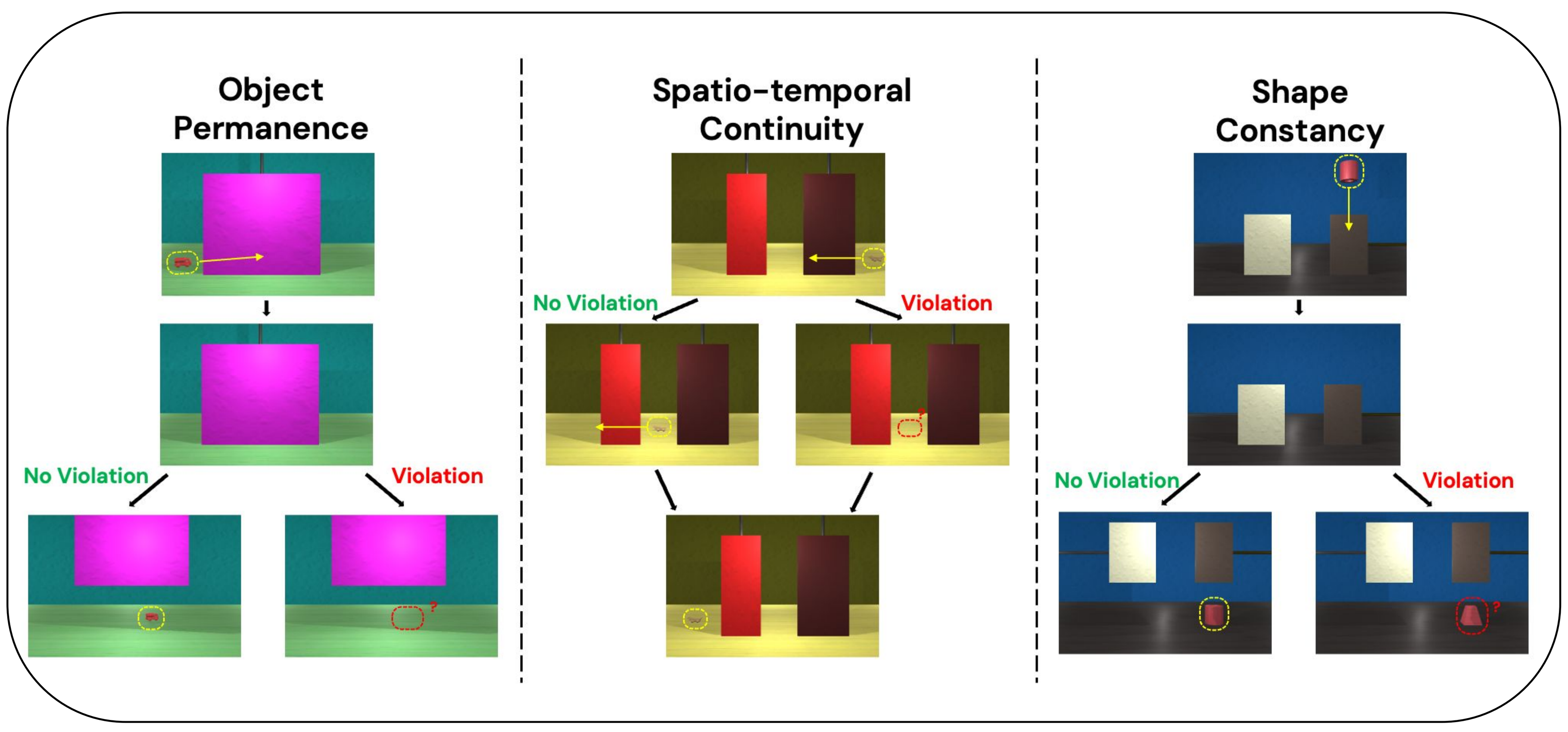
CoCoSys
CENTER FOR THE CO-DESIGN OF COGNITIVE SYSTEMS

Intuitive Physical Reasoning with Probabilistic Programs

Arijit Dasgupta, Nishad Gothoskar, Joshua Tenenbaum, Vikash Mansinghka

Intuitive Physics

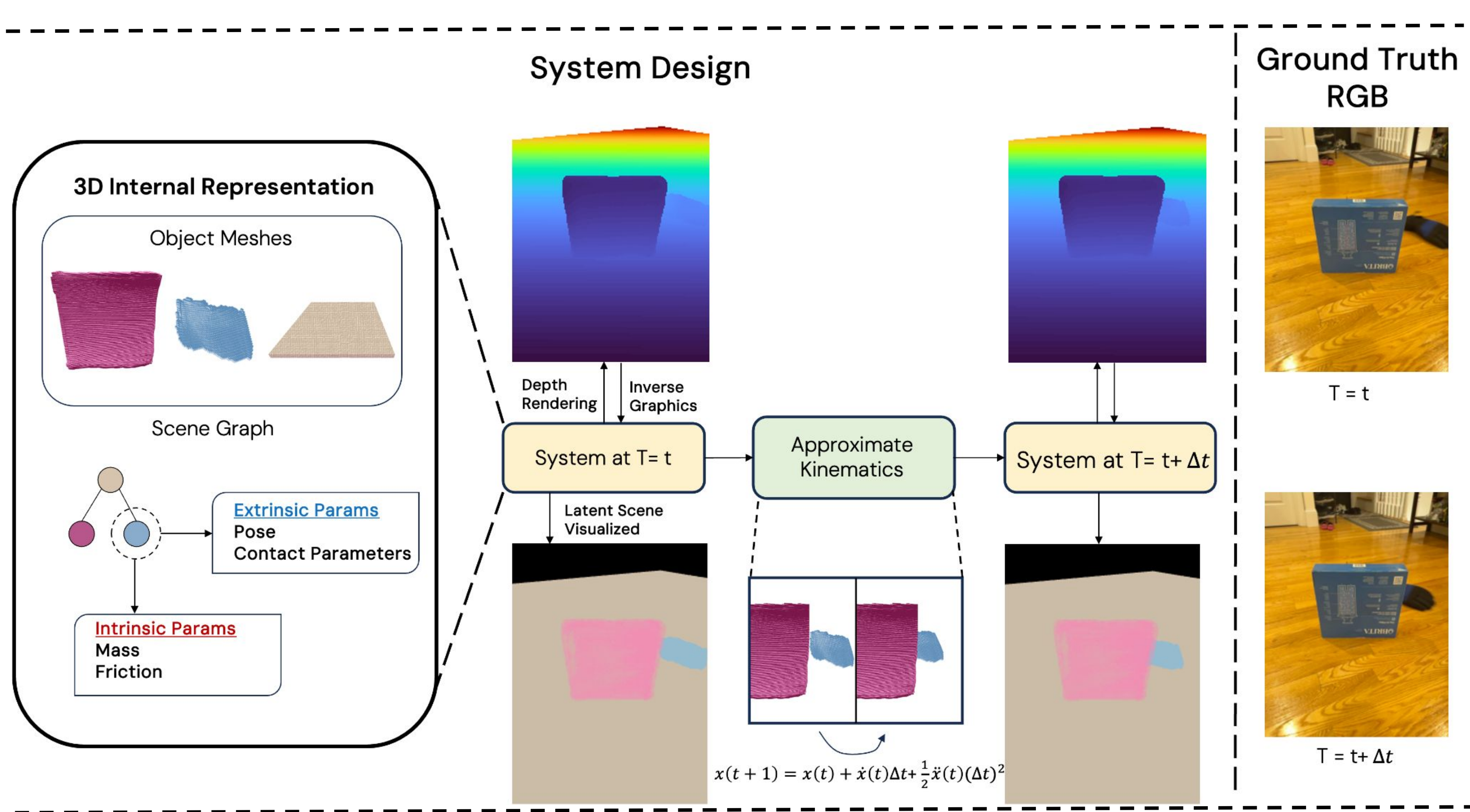
- Human infants recognize **violations** in physical reasoning from the age of 4 months (Violation-of-Expectation).



- If a vision system claims to **incorporate physical reasoning**, we should be able to **probe** it for such violations
- **Task:** Distinguish which RGB-D videos of rigid body interactions violate commonsense physical intuitions.

Inverse Graphics and Approximate Kinematics System

- We concretize the Intuitive Physics Engine (IPE) theory into a system (probabilistic program) with two core components:
- **Inverse Graphics**: Constructing 2D depth and RGB images from an interpretable 3D internal representation.
- **Approximate Kinematics**: Forward time-stepping of the 3D internal representation with approximately correct kinematics.

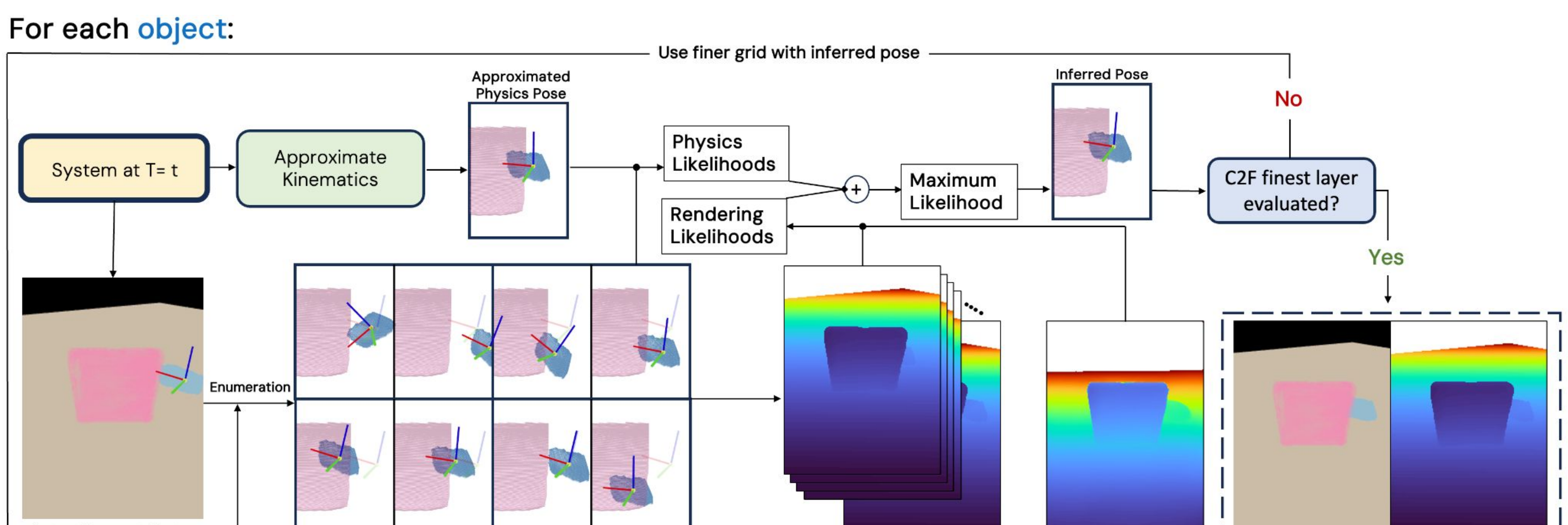


- Approximate dynamics lets us keep track of hidden objects that were previously visible.

Coarse-to-Fine Enumeration

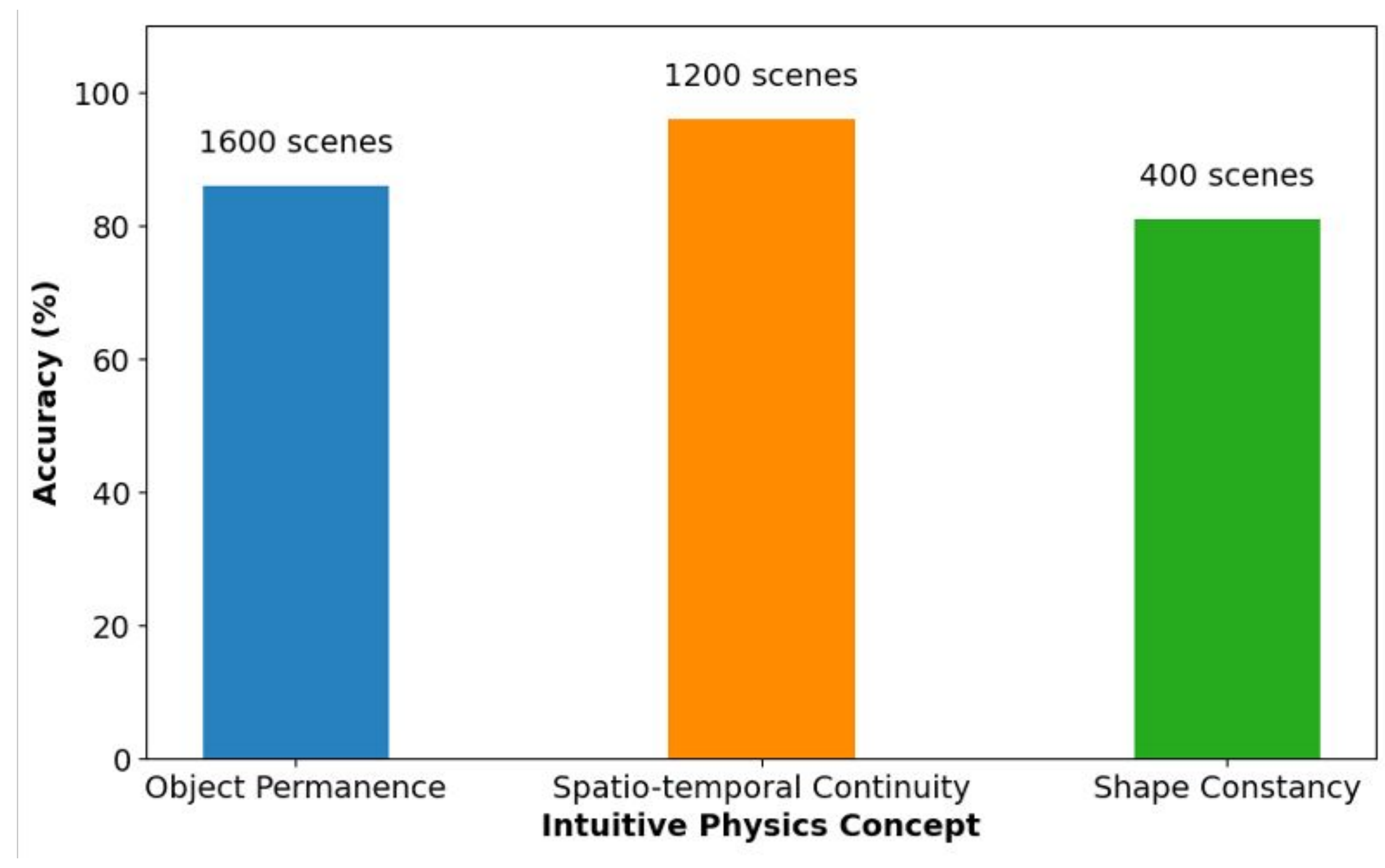
Inference: Online GPU-accelerated Sequential Monte Carlo with enumerative coarse-to-fine (C2F) pose proposals.

For each object:



Results

- By **evaluating** on a synthetic dataset, our **unified** approach works **robustly** across multiple physical concepts to **accurately** detect violations.



Future Steps

- Shift from a system-based approach to a **robust probabilistic model** of 3D objects and interactions.

Proposed Hierarchical IPE Model

procedure IPE-Model(N, T)
  $S \sim$ IPE-Initialize(N)
  for $t \leftarrow 2$ to T **do**
    $S \sim$ IPE-Step(S, N, t)
  end for
  return S
end procedure

procedure IPE-Initialize(N)
  $S \leftarrow$ EmptyDict
  $G(V, E) \sim$ Scene-Graph-Prior(N)
  for all $v \in$ Topo-Sort$(V, E) \wedge ((v = root) \vee (u \in V \text{ s.t. } (u, v) \in E))$ **do**
    $O^{(v)} \sim$ Uniform-Shape-Prior
    $M^{(v)} \sim$ Mass-Prior$(O^{(v)})$ ▷ Relative Mass
    $f^{(v)} \sim$ Friction-Prior() ▷ Kinematic Friction
    if $v = root$ **then**
      $X^{(v)} \sim$ Uniform-6DoF-Pose-Prior() ▷ World Frame Pose
    else
      $X^{(u,v)} \sim$ Contact-Prior$(G, O^{(v)})$ ▷ Parent Frame Pose
      $X^{(v)} \leftarrow X^{(u)} \times X^{(u,v)}$ ▷ World Frame Pose
    end if
    $\Phi^{(v)} = [O^{(v)}, M^{(v)}, f^{(v)}, X^{(v)}]$
  end for
  $S(1) \leftarrow [G, \Phi^{(1)}, \Phi^{(2)}, \ldots, \Phi^{(N)}]$
  $\hat{Y} \leftarrow$ Unproject(Depth-Render$(X^{(1:N)}, O^{(1:N)}, X^C, c)$)
  $Y \sim$ Noise-Model$(\hat{Y})$
  return S
end procedure

procedure IPE-Step(S, N, t)
  for all $v \in V$ **do**
    $X^{(v)}_{\text{step}} \sim$ Physics-Model-Finite-Diff(S, v, t)
    $\Phi^{(v)} = [O^{(v)}, M^{(v)}, f^{(v)}, X^{(v)}_{\text{step}}]$
  end for
  $S(t) \leftarrow [G, \Phi^{(1)}, \Phi^{(2)}, \ldots, \Phi^{(N)}]$
  $\hat{Y} \leftarrow$ Unproject(Depth-Render$(X^{(1:N)}_{\text{step}}, O^{(1:N)}, X^C, c)$)
  $Y \sim$ Noise-Model$(\hat{Y}, \theta)$
  return S
end procedure

procedure Physics-Model-Finite-Diff(S, v, t)
  $X^{(v)}(t-3), X^{(v)}(t-2), X^{(v)}(t-1) \leftarrow$ Get-Poses(S, v, t)
  $A^{(v)} \leftarrow \frac{X^{(v)}(t-1) - 2X^{(v)}(t-2) + X^{(v)}(t-3)}{(\Delta T)^2}$
  $V^{(v)} \leftarrow \frac{X^{(v)}(t-1) - X^{(v)}(t-2)}{\Delta T}$
  $\tilde{X^{(v)}}(t) \leftarrow X^{(v)}(t-1) + V^{(v)} \cdot \Delta T + \frac{1}{2}A^{(v)}(\Delta T)^2$
  $X^{(v)}(t) \sim$ Gaussian-Mixture-Model$(\tilde{X^{(v)}}(t), \kappa)$
  return $X^{(v)}(t)$
end procedure

Notation	Meaning
$O^{(m)}$	Object shape
N	Number of objects
c	Camera Intrinsics
X^c	Camera pose in world frame
$G = (V, E)$	Scene graph structure
$X^{(v)} \in SE(3)$	6DoF pose of v w.r.t. world frame
$X^{(u,v)} \in SE(3)$	6DoF pose of v w.r.t. parent (u) frame
$\hat{Y}$	Rendered point cloud
Y	Point cloud with noise
S	State dictionary over time
Φ	Object State
T Number of discrete time-steps	
ΔT	Time-step
θ	Noise model parameters
κ	GMM covariance parameters

Currently has

- **Scene graph** representation.
- **State-space** stepper model.
- Forward-euler finite difference kinematic approximation with **GMM** noise.
- **Mass** and **Friction** as random choices.

Will eventually require

- **Collision** detection and probabilistic collision resolution.
- Categorization of **static** and **moving** objects.
- Energy-based dynamics (effect of **friction** and **elasticity**).

Acknowledgments

This work was supported in part by CoCoSys, one of seven centers in JUMP 2.0, a Semiconductor Research Corporation (SRC) program sponsored by DARPA.

Semiconductor Research Corporation