

## 1 Appendix

### 2 A1 Optimization

3 Optimizing the force  $\mathbf{f}(\mathbf{x}, t)$  through differentiable physics is non-trivial. In the forward physical  
4 process, we repeatedly compute Equations 5-8 for hundreds of sub-steps  $\delta t$  between two consecutive  
5 video frames  $t$  and  $t + 1$ . During backpropagation, we have to roll out the same amount of sub-steps  
6 and accumulate the backward gradients, which is highly unstable and prone to gradient explosion.

7 To address the optimization challenge, we introduce a novel optimization scheme, where we discretize  
8 the temporally continuous forces into pulsed ones. Specifically, in the forward process (Equation 4),  
9 for each  $t \rightarrow t + 1$  frame, we only apply the same force  $\mathbf{f}(\mathbf{x}, t)$  in the first  $k$  sub-steps  $t \rightarrow t + k\delta t$   
10 and remove  $\mathbf{f}$  in the remaining sub-steps:

$$\mathbf{f}(\mathbf{x}, t') = \begin{cases} \mathbf{f}(\mathbf{x}, t) & \text{if } t' \in [t, t + k\delta t] \\ \mathbf{0} & \text{if } t' \in (t + k\delta t, t + 1]. \end{cases} \quad (1)$$

11 The rationale behind this is (1) reducing the temporal granularity of force estimation: we estimate the  
12 force  $\mathbf{f}(\mathbf{x}, t)$  per frame, instead of per sub-step; (2) reducing the number of sub-steps for applying  
13 force, which stabilizes optimization.

14 Similarly, in backward, we also reduce the number of sub-steps for accumulating force gradients.  
15 Specifically, we only accumulate the gradients of the first  $k = 1$  sub-steps:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{f}(\mathbf{x}, t)} = \sum_{t'=t+k\delta t}^t \frac{\partial \mathcal{L}}{\partial G^{t'}} \frac{\partial G^{t'}}{\partial \mathbf{f}(\mathbf{x}, t)}. \quad (2)$$

16 Since the time slots between frames are quite short (in milliseconds), the discretization effectively  
17 approximates the continuous forces. By reducing the gradient accumulation sub-steps, we mitigate  
18 the gradient explosion problem and ensure stable optimization of forces.

19 To validate the effectiveness of our proposed optimization scheme, we conducted an experiment  
20 comparing three cases: applying the force in the first sub-step, the last sub-step, and all sub-steps. The  
21 results in Table 1 show that applying forces in the first sub-step significantly outperforms applying  
22 them in the last sub-step, indicating that delayed force application fails to effectively influence object  
23 motion, leading to suboptimal supervision and unreliable optimization. Moreover, applying forces in  
24 all sub-steps hinders optimization convergence, whereas our proposed scheme successfully addresses  
25 this optimization challenge.

Table 1: Quantitative comparison of optimization schemes. N.C.: did not converge in optimization.

Force sub-steps	PSNR	SSIM	LPIPS	Mag. Error (%)	Dir. Error (°)
First sub-step	<b>37.02</b>	<b>0.9945</b>	<b>0.0022</b>	<b>4.81</b>	<b>2.94</b>
Last sub-step	21.28	0.9449	0.0411	100.45	22.90
All sub-steps	N.C.	N.C.	N.C.	N.C.	N.C.

### 26 A2 Implementation Details

27 For implementation, we set both the simulation grid resolution and the triplane grid resolution to 50.  
28 The MPM simulation employs 400 sub-steps within each temporal interval between successive video  
29 frames to ensure accurate physics simulation. The captured videos were rendered at a resolution of  
30 1920×1080 with a frame rate of 200 fps, simulating 50 frames of motion. On a single NVIDIA RTX  
31 4090 GPU, the optimization process takes approximately 5 minutes per frame.

### 32 A3 Synthetic Dataset Quality Results

33 To ensure the materials produce appropriate deformations that can be tracked, their properties are  
34 manually specified. Figure 1 illustrates the simulation results for objects made from different materials.  
35 For example, the Hotdog belongs to viscoplastic materials, Lego and Ficus represent elastic materials,

36 while the Toy is an example of elastoplastic materials. The visual comparison between the ground  
 37 truth and our re-simulated results highlights the accuracy of our method in recovering external forces  
 38 and reproducing motion details specific to each material type. These results demonstrate that our  
 39 approach effectively generalizes across a wide range of materials, accurately capturing the distinct  
 40 deformation and motion characteristics under various external force conditions.

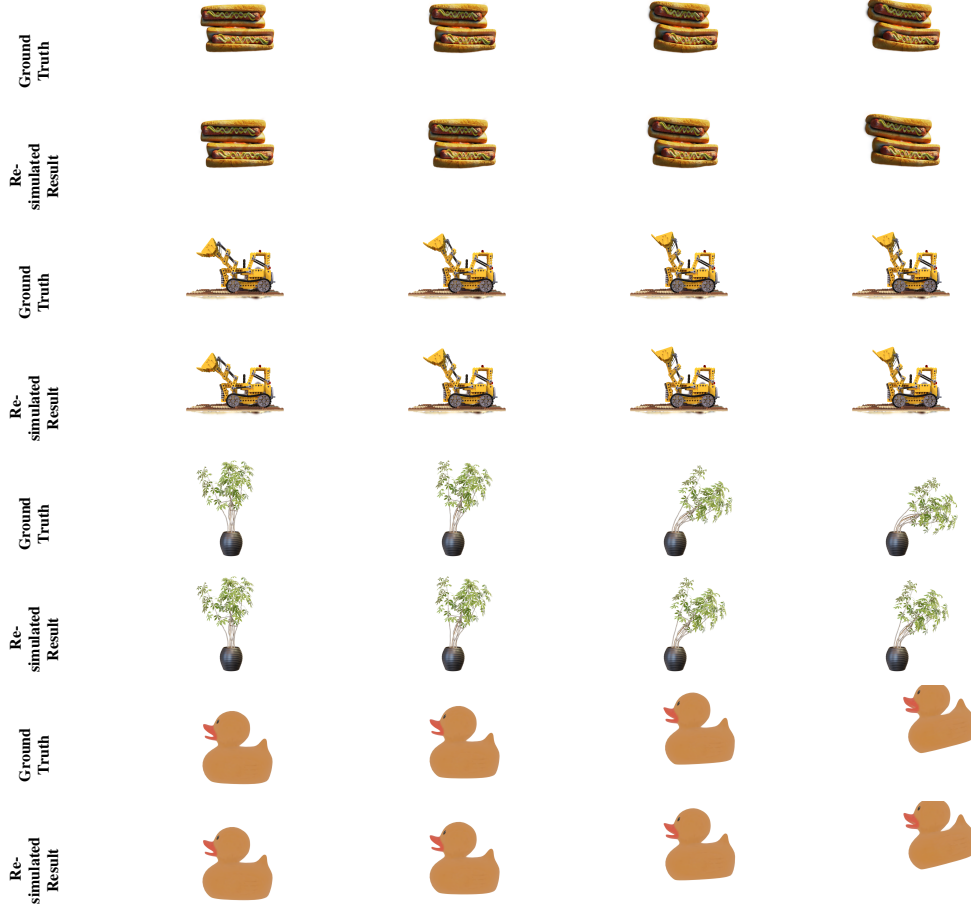


Figure 1: Visual comparison of simulation results for various objects. Each two-row group corresponds to a specific object: Hotdog, Lego, Ficus, and Toy. The first row in each group shows the ground truth from original simulations, while the second row presents our re-simulated predictions. These results illustrate the capability of our method to accurately replicate external forces and maintain motion details across different materials and environments.

Material type	Object	PSNR	SSIM	LPIPS	Mag. Error (%)	Dir. Error (°)
Elastic	Lego	$33.70 \pm 9.49$	$0.98 \pm 0.03$	$0.01 \pm 0.02$	$19.53 \pm 41.50$	$7.02 \pm 14.07$
	Ficus	$25.92 \pm 6.64$	$0.94 \pm 0.04$	$0.03 \pm 0.04$	$23.97 \pm 16.08$	$11.55 \pm 7.58$
	Sunflower	$34.08 \pm 4.24$	$0.99 \pm 0.01$	$0.01 \pm 0.00$	$14.38 \pm 8.69$	$7.85 \pm 12.93$
Elastoplastic	Toy	$41.35 \pm 10.47$	$0.99 \pm 0.01$	$0.00 \pm 0.00$	$29.19 \pm 10.08$	$8.11 \pm 4.60$
	Chair	$40.10 \pm 6.33$	$0.99 \pm 0.01$	$0.00 \pm 0.00$	$33.31 \pm 30.10$	$23.40 \pm 21.28$
Viscoplastic	Hotdog	$30.63 \pm 6.52$	$0.96 \pm 0.03$	$0.02 \pm 0.02$	$15.09 \pm 10.16$	$11.63 \pm 17.35$

Table 2: Force recovery is applied in synthetic scenarios, with statistical significance evaluated