Energetically Consistent Inelasticity for Optimization Time Integration

XUAN LI, University of California, Los Angeles, USA
MINCHEN LI, University of California, Los Angeles & TimeStep Inc., USA
CHENFANFU JIANG, University of California, Los Angeles & TimeStep Inc., USA

In this paper, we propose Energetically Consistent Inelasticity (ECI), a new formulation for modeling and discretizing finite strain elastoplasticity/viscoelasticity in a way that is compatible with optimization-based time integrators. We provide an in-depth analysis for allowing plasticity to be implicitly integrated through an augmented strain energy density function. We develop ECI on the associative von-Mises J2 plasticity, the non-associative Drucker-Prager plasticity, and the finite strain viscoelasticity. We demonstrate the resulting scheme on both the Finite Element Method (FEM) and the Material Point Method (MPM). Combined with a custom Newton-type optimization integration scheme, our method enables simulating stiff and large-deformation inelastic dynamics of metal, sand, snow, and foam with larger time steps, improved stability, higher efficiency, and better accuracy than existing approaches.


Additional Key Words and Phrases: MPM, Elastoplasticity, Inelasticity, Implicit Integration, Optimization Time Integration

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.
0730-0301/2022/7-ART52 $15.00
https://doi.org/10.1145/3528223.3530072

1 INTRODUCTION

Since the pioneering work of Terzopoulos and Fleischer [1988], the computer graphics community has observed increasing interests in modeling inelastic deformations governed by elastoplasticity, viscoelasticity, and viscoplasticity. These inelastic mechanical properties govern the behaviors of a wide range of everyday objects. Drawing inspirations from continuum mechanics, computer graphics researchers have successfully modeled and simulated many inelastic materials, ranging from metal, sand, snow and mud to foam, paint and organic tissues.

Inelasticity (mainly elastoplasticity and viscoplasticity) has been widely explored using mesh-based Finite Elements. During inelastic deformation, extreme element distortion and fracture commonly co-exist. Thus, remeshing [O’Brien et al. 2002] and virtual node [Hegemann et al. 2013; Molino et al. 2004] techniques are often applied. More recently, the Material Point Method (MPM) has emerged as a popular alternative for inelastic materials [Jiang et al. 2016] due to its natural support of topologically changing continuum materials.

Despite a large amount of work in modeling inelasticity, a loss of accuracy occurs in almost all existing work. In particular, when implicit time integration schemes are performed, the plastic strain...
is often treated as a constant, and the real plastic deformation is imagined to happen instantaneously at the beginning or the end of a time step. Such a semi-implicit lagged treatment of inelasticity results in unnoticeable visual artifacts for certain material models such as the heuristic snow plasticity in Stomakhin et al. [2013] but significant errors such as excessive artificial cohesion for others [Gao et al. 2018; Tampubolon et al. 2017].

The choice of semi-implicit is largely due to the prominent challenge in modeling implicit inelasticity. Klár et al. [2016] was the first to explore differentiating the plastic flow for Drucker-Prager soil plasticity and incorporating it into the implicit momentum balance. The authors proposed an implicit force formulation that resembles a similar format to semi-implicit formulations [Stomakhin et al. 2013]. Unfortunately, their formulation cannot be expressed as the negative gradient of analytical energy. Resultingly, the stiffness matrix is asymmetric, and GMRES became necessary for the associated nonlinear root-finding problem—a problem that by itself has no stability or convergence guarantees when solved with Newton’s method. Fang et al. [2019] used alternating direction method of multipliers (ADMM) to shift the asymmetry to local small linear systems, however without an energy, they could not perform global convergence techniques such as line search.

This paper tackles the challenge by revisiting the derivation of implicit plasticity. Our objective is to construct an analytical, augmented potential energy function whose derivative exactly reproduces the implicit force. Related work in classic engineering literature [Ortiz and Stainier 1999; Radovitzky and Ortiz 1999] formulated variational constitutive model updates based on the principle of maximum plastic dissipation and minimizing over the so-called dual inelastic potential. Taking a different path, we derive our method based on constructing a smooth energy that is consistent with existing return mapping-based plasticity treatments [Simo and Hughes 1998] in explicitly integrated inelasticity simulation systems. As a result, our implicit inelasticity formulation can be directly incorporated into recently advanced optimization time integrators [Gast et al. 2015; Li et al. 2020; Wang et al. 2020] to enable large step integration with guaranteed stability, theoretical consistency with return mapping, and a symmetric energy Hessian. Our contributions include:

- An implicit internal force formulation for fully implicit finite strain elastoplasticity;
- A strain energy augmentation method that yields analytically integrable elastoplastic forces and symmetric force derivatives for von Mises J2 plasticity;
- An extension of our model to support strain hardening, pressure-dependent soil plasticity, and rate-dependent viscoelasticity;
- Algorithms for incorporating our model in optimization-based time integrators with the Material Point Method and the Finite Element Method.

We demonstrate our results by simulating a wide range of inelastic materials, including metal, sand, snow, and foam. Our method allows the simulations of inelasticity to enjoy the advantages of guaranteed stability, global convergence, and large time step sizes brought by optimization-based time integrators without suffering from inaccuracy and numerical artifacts from prior work.

2 RELATED WORK

Inelasticity with FEM. Elastoplastic simulation with FEM has been extensively explored by the computer graphics community. O’Brien et al. [2002] used the additive decomposition of strain to separate elastic deformations and plastic deformations and used the von-Mises yield criterion. However, as Irving et al. [2004] pointed out, this decomposition does not support incompressibility for finite strain. Instead, Irving et al. [2004] used the multiplicative decomposition of deformation gradient with the volume-preserving return mapping algorithm. Our model is based on this decomposition as well. Under this framework, large plastic deformations may make the dynamic system ill-conditioned. To solve this problem, Molino et al. [2004] proposed the virtual node algorithm to allow topology changes when the simulated mesh is severely distorted, and Bargteil et al. [2007] used remeshing technique to maintain a high-quality mesh throughout the simulation. For high-performance simulation, Wojtan and Turk [2008] used frequently remeshed high-resolution surfaces combined with low-resolution interior tetrahedral mesh to resolve thin features near the boundaries. Wojtan et al. [2009] further improved the framework to allow topology changes in inelastic simulations. These methods introduced extra computational costs or complexities. Instead, we use optimization time integrators to maintain long-time stability and global convergence. Furthermore, Bargteil et al. [2007] proposed a volume-preserving plasticity model incorporating creep and work hardening/softening, which is also followed by Wojtan and Turk [2008]. These are important requirements for obtaining physical accuracy, which are all supported by our model as well. Jones et al. [2016b] proposed an examples-based approach for the mesh-based discretization, which search rest shapes on a predefined example manifold. This method is efficient for animation purposes but are less physically accurate.

Inelasticity with MPM. Extending the work of Harlow [1964] and Brackbill and Ruppel [1986] on PIC/FLIP, MPM was proposed as a hybrid Lagrangian/Eulerian method for solid mechanics by Sulsky et al. [1994]. Since its appearance in the graphics community [Hegemann et al. 2013; Stomakhin et al. 2013], it has attracted a lot of attentions. The most prominent advantage of MPM on modeling inelastic materials is its flexibility in handling extreme deformation and topological changes, which pose significant challenges to Lagrangian mesh-based approaches. Snow plasticity was first simulated by Stomakhin et al. [2013] in a semi-implicit fashion, enforcing thresholds on principal stretches with post-projections. Yue et al. [2015] used the Herschel-Bulkley model of non-Newtonian viscoplastic flow to approximate foam behaviors. Fei et al. [2019] derived an analytic plastic flow approach for Herschel-Bulkley fluid to simulate compressible, shear-dependent liquids. Daviet and Bertails-Descoubes [2016] modeled the granular materials as compressible viscoplastic fluids combined with the Drucker-Prager yield criterion. Their method suits the granular material simulations well, but follows a different perspective from ours. From the perspective of large strain solid mechanics, Klár et al. [2016] simulated granular continuum using the return mapping algorithm for the Drucker-Prager plasticity. Following Klár et al. [2016], Yue et al. [2018] proposed a hybrid method combining both discrete and continuum treatments to achieve a high level of details with less computational costs. Fang
et al. [2019] applied the return mapping approach to handle elastoplasticity and viscoelasticity in an ADMM framework. Except for Klár et al. [2016], these methods all temporally discretize inelasticity in an explicit or semi-implicit way, where the plastic correction was performed as an extra step at the end of each time step, fully decoupled from elasticity. Decoupled treatment in an explicit integration can be justified via operator splitting; however, it will cause artifacts for a (semi-)implicit integration. We use the return mapping framework as well for our fully implicit elastoplasticity and viscoelasticity, and we will show that ours is more temporally consistent compared to Klár et al. [2016].

Inelasticity with Other Discretizations. Inelasticity simulations are also explored with other types of spatial discretizations, e.g., Smoothed Particle Hydrodynamics (SPH), Position Based Dynamics (PBD), peridynamics, etc.

SPH is a mesh-free Lagrangian method originally invented for fluid simulations. Inspired by SPH, Jones et al. [2014]; Müller et al. [2004] applied the plasticity model in O’Brien et al. [2002] to moving least square particles for elastoplastic objects. Clavet et al. [2005] used springs between particles to mimic elasticity and achieved plasticity by modifying rest lengths during the simulation. These two plasticity models are not derived from the finite strain framework. Alduán and Otaduy [2011] simulated granular materials using an incompressible SPH framework combined with the Drucker-Prager yield criterion. Their plastic correction was performed in a Jacobi-like manner until convergence, while ours is performed with fixed-point iterations. Yang et al. [2017] proposed an elastoplastic model based on the Drucker-Prager yield criterion as well within an SPH framework. Gerszewski et al. [2009] introduced deformation gradients to the SPH framework so that plasticity models based on the multiplicative decomposition of deformation gradient can be applied. They used explicit time integrators combined with the plasticity model in [Irving et al. 2004]. Gissler et al. [2020] used an implicit compressible SPH solver to simulate the compression of snow. The plasticity is handled by an extra correction step on the deformation gradient following Stomakhin et al. [2013], which is still a semi-implicit method.

PBD was proposed by [Müller et al. 2007] for real-time simulations. This method replaced internal forces in force-based methods with constraints on positions. Plastic deformations can be introduced by the shape matching framework [Bender et al. 2017; Falkenstein et al. 2017; Jones et al. 2016a; Müller et al. 2005]. However, this simulation framework sacrifices physical accuracy for better efficiency.

The peridynamic theory is an emerging field in simulations, which was proposed by Silling [2000] to handle discontinuities caused by deformations, such as cracks. It defines pairwise force functions between particles and uses the integration over the interactions from neighboring particles to describe dynamics. He et al. [2017] used the peridynamics framework to simulate elastoplastic materials in a projective dynamics way. They adopted the Drucker-Prager criterion for plasticity. Their solver can also be extended to simulate viscoelasticity. Chen et al. [2018] derived a form of force functions based on the isotropic linear elasticity model to simulate elastoplastic materials. They used explicit time integrators and an additive plasticity model.

Optimization Time Integration. Large-scale implicit simulation methods usually require solving large systems of nonlinear equations. To solve these systems, the Newton method for root-finding problems is usually adopted, which needs careful tuning of the time step size to ensure convergence. In fact, many of these implicit equations can be integrated to get variational forms, where the equivalent minimization problem can be solved by applying robust optimization techniques. The optimization time integrators have advantages in terms of long-time stability even when simulating severe deformation with large time step sizes.

Bouaziz et al. [2014] proposed Projective Dynamics (PD), which reformulated the backward Euler time integration for a specific type of material into a local-global alternating solver. Both the local and global steps have simple variational forms that can be solved in a robust and efficient way. This framework was later extended to simulate hyperelastic materials [Liu et al. 2017], support Laplacian damping [Li et al. 2018], and utilize other time integration schemes [Dinev et al. 2018]. Narain et al. [2016] then extended PD to a more general form within the ADMM framework. Brown and Narain [2021] improved the ADMM framework to resolve large rotations. Gast et al. [2015] recast the backward Euler time integration with hyperelastic materials, Rayleigh dampings, and collision penalties as a minimization problem. Li et al. [2019] and Wang et al. [2020] explored domain decomposed and hierarchical preconditioning strategies respectively within a quasi-Newton optimization framework for robust and efficient time integration. Wang and Yang [2016] proposed a gradient descent solver for GPUs to accelerate optimization time integrations. Li et al. [2020] proposed Incremental Potential
Contact (IPC), a variational form for frictional contacts. Their friction bases are iterated in a similar manner to our iterative yield stresses. IPC is later proven effective for simulating codimensional objects [Li et al. 2021b], rigid bodies [Ferguson et al. 2021], reduced elastic solids [Lan et al. 2021], and FEM-MPM coupled domains [Li et al. 2021a], all within the optimization time integration framework. In this paper, we follow Gast et al. [2015] and Li et al. [2020] for the optimization time integration of MPM and FEM respectively. The hessian matrices are enforced to be positive definite following the per-stencil projection technique in Tan et al. [2005].


3 FOUNDATIONS

In this section we start with reviewing finite strain elastoplasticity (Section 3.1), MPM spatial discretization (Section 3.2), optimization-based time integration (Section 3.3), and discretized plastic flow rule (Section 3.4). Our review is by no means complete, and they are provided as necessary background knowledge for our new model.

In Section 3.5, we present a new implicit force formulation that is consistent with the variational weak form. It has a remarkable advantage – integrability, and thus lays important theoretical foundations for our method.

3.1 Finite Strain Elastoplasticity

Our variational inelasticity model is derived under the finite strain elastoplasticity framework. Here we review some basic concepts and refer to Simo [1992], Simo and Hughes [1998] for more details.

Let \( Q^0 \subset R^3 \) be the reference configuration of the continuum body and denote \( x := \Phi(X, t) \) the deformation map from \( Q^0 \) (with coordinate \( X \)) to the world space \( \Omega^0 \) (with coordinate \( x \)). The deformation gradient \( F = \frac{\partial \Phi}{\partial X}(X, t) \) measures the local deformation of the infinitesimal region around \( X \). With finite strain elastoplasticity, \( F \) is multiplicatively decomposed into \( F = FEF^P \), where \( F^P \) denotes the permanent plastic deformation, and \( FE \) denotes the elastic deformation which results in elastic forces (Figure 3). Plasticity requires that the Kirchhoff stress \( \tau \) associated with \( FE \) is inside the admissible area defined by a yield condition \( y(\tau) \leq 0 \). The surface characterized by \( y(\tau) = 0 \) is often referred to as the yield surface. When \( F \) changes, \( FE \) will follow some plastic flow to evolve so that it lies within the yield surface. In this paper, we follow the volume preserving plastic flow from [Klar et al. 2016].

From the Lagrangian view point, the state of dynamics of an elastoplastic continuum can be described by a Lagrangian density field \( R(X, t) \) on \( \Omega^0 \) and a Lagrangian velocity field \( V(X, t) := \frac{\partial \Phi(X, t)}{\partial t} \) on \( \Omega^0 \). The two fields are governed by the conservation of

\[
\frac{1}{\Delta t} \int_{\Omega^0} \rho(x, t^n)(\tilde{\psi}^{n+1} - \tilde{\psi}^n) q_{a} dX = - \int_{\Omega^0} \rho \tilde{\psi}^{n+1} \tau_{a} dX + \int_{\Omega^0} R(X, 0) \partial_t \tilde{\psi}^{n+1} q_{a} dX,
\]

where \( \rho, \tilde{\psi}^n, \tilde{\psi}^{n+1}, \) and \( q_{a} \) are Eulerian counterpart of \( R, V^n, \nabla \psi, \) and \( Q_{a} \), obtained by pushing forward from \( \Omega^0 \) onto \( \Omega^n \).

In MPM, B-Spline-based interpolations are often applied to define \( R^n \), and material particles serve as quadratures to approximate the volume integration. Let \( X^n_p \), \( \hat{X}_p \), be the coordinate of particle \( p \) in \( \Omega^n \) and \( \hat{Q}^a_p \), and \( \nabla \hat{w}_p^n \) be the weight and weight gradient between particle \( p \) and grid \( i \). With the mass lumping technique, the force equilibrium of grid \( i \) can be discretized as

\[
\frac{1}{\Delta t} m^i_p (\tilde{\psi}^n_p - \tilde{\psi}^i) = - \sum_{p} P_p F_p^T \nabla w^n_p \tilde{\psi}^n_p,
\]

where \( V^n_p \) is the initial volume of particle \( p \), \( m^i_p = \sum_p m_p w^n_p \) is the lumped mass on grid \( i \) and \( m_p \) is approximated by \( R(X_p, 0) V^n_p \). The right hand side of Equation 5 is the internal elastic force on grid \( i \).

At each time step, the velocity field \( \tilde{\psi}^n_p \) is transferred from material particles to grid nodes, and the new velocity field \( \tilde{\psi}^{n+1} \) is solved and transferred back to material particles for advection. In this paper, we
use the quadratic MLS kernel [Hu et al. 2018] as the weight function, and APIC [Jiang et al. 2015] as the particle-grid transfer scheme.

3.3 Optimization Time Integration
Assuming implicit integration with BDF1 (backward Euler), the first Piola-Kirchhoff stress $P$ in Equation 5 is associated with the deformation gradient $F^{n+1}$ at time step $t^{n+1}$. The deformation gradients $F^n$ and $F^{n+1}$ are related by

$$F^{n+1} = (I + \Delta t \nabla \Psi^{n+1}) F^n,$$  

where $\nabla \Psi^{n+1} = \sum_i \hat{\Psi}^{n+1} \nabla \hat{\Psi}^i$.

Existing optimization-based time integrators in computer graphics often assume hyperelastic materials. Without plasticity, the first Piola-Kirchhoff stress is simply the derivative of the corresponding elastic strain energy density function: $P(F) = \frac{\partial \Psi}{\partial F}$. Equation 5 is then equivalent to the following optimization problem:

$$\Delta \Psi = \arg\min_{\Delta \Psi} E(\Delta \Psi) = \sum_i m_i \| \Delta \Psi^i \|^2 + \sum_p \Psi(F^{n+1}_p (\hat{\Psi} + \Delta \Psi)) \nabla \psi^n_p,$$  

where $F^{n+1}_p (\hat{\Psi} + \Delta \Psi)$ is the elastic predictor. The optimization problem can be robustly solved by projected Newton’s method with backtracking line search [Wang et al. 2020].

**Gravity.** For the effect of gravity, we add the term $m_i g$ on the right-hand side in Equation 5, which corresponds to the extra term $- \sum_i m_i g^T \Psi$ in Equation 7.

3.4 Discretization of Plastic Flow
In the discrete setting, plasticity is most commonly achieved by the return mapping algorithm [Klár et al. 2016; Simo and Hughes 1998], which is equivalent to solving for a strain that satisfies the plastic flow rule. Geometrically, the return mapping defines how elastic predictors outside the yield surface should be corrected so that the effective stresses lie inside the yield surface. We follow the notations in Klár et al. [2016] to describe discrete plastic flows in this paper.

For elasticity, we adopt the St. Venant-Kirchhoff (StVK) model with Hencky strains. The elastoplasticity of isotropic materials can be characterized in the principal stretch space using the singular value decomposition (SVD) [Stomakhin et al. 2012]. Let $\Sigma^{ir}$ be the SVD of an elastic predictor $F^{ir}$. The Hencky strain is defined as $\epsilon = \log \Sigma^{ir}$, and the Kirchhoff stress for the StVK model is $\tau = 2\mu \epsilon + \lambda \text{tr}(\epsilon) I$, where $\mu, \lambda$ are Lamé parameters. For a discrete plastic flow, if the stress associated with an elastic predictor is outside the yield surface, then the stress is projected back onto the yield surface. The projection procedure in the principal stress space is illustrated in Figure 5. Note that along the perpendicular direction to the diagonal, we would have $\det F^{ir} = 1$, which corresponds to a volume-preserving plastic deformation. We denote the endpoint of the return mapping (also known as the corrector or the effective stress) as $F^{E} = Z(F^{ir})$ where $Z(\cdot)$ is the return mapping.

3.5 Force Balance with Implicit Plasticity
For elastoplastic materials, the first Piola-Kirchhoff stress $P$ in Equation 5 should be rewritten as [Bonet and Wood 1997]

$$P = \frac{\partial \Psi^E}{\partial F^E} F^{E-\tau}.$$  

Here $\Psi^E$ is the elastic strain energy density function. We add a superscript to emphasize the elastic energy is only associated with the elastic deformation gradient $F^{E}$.

Through a weak form derivation of the updated Lagrangian dynamics (see the supplemental document for details), we show that

\begin{align*}
\end{align*}
the implicit internal force on grid node $i$ is:

$$f_{n+1}^i = -\sum_p v_0 \frac{\partial \Psi^E}{\partial F^E}(F_{p,n+1}^E) F_{p,n+1}^{E,n+1} \nabla w_{ip}^n.$$  

(9)

In practice one does not need to track the plastic deformation gradients $F^P$ on the particles. The nodal force can be expressed in terms of $F^{E,tr}$ (defined as $F^{E,tr}(v) = (I + \Delta t \nabla \psi_p) F_{p,n}^E$):

$$f_{n+1}^i = -\sum_p v_0 \frac{\partial \Psi^E}{\partial F^E}(F_{p,n}^E) F_{p,n+1}^{E,tr} \nabla w_{ip}^n.$$  

(10)

Note that when doing explicit time integration, we can directly replace $F_{p,n+1}^{E,tr}$ with $F_{p,n}^E$, which gives the common force expression for explicit MP:

$$f_{n}^i = -\sum_p v_0 \frac{\partial \Psi^E}{\partial F^E}(F_{p,n}^E) F_{p,n}^{E,n} \nabla w_{ip}^n.$$  

(11)

For implicit plasticity, Klár et al. [2016] directly replace $\frac{\partial \Psi^E}{\partial F^E}(F_{p,n}^E)$ in Equation 11 with $\frac{\partial \Phi^E}{\partial F^E}(Z(F_{p,n}^{E,tr}))$ and define the resulting expression as the implicit force. We can clearly observe that

$$f_{n+1}^i = -\sum_p v_0 \frac{\partial \Psi^E}{\partial F^E}(F_{p,n}^E) F_{p,n+1}^{E,tr} \nabla w_{ip}^n,$$  

(12)

i.e., the implicit force in [Klár et al. 2016] is not equivalent to our formulation (Equation 10). As we discuss in the supplemental document, the choice of Klár et al. [2016] is only semi-implicit. Furthermore, their formulation is not integrable because their force derivative is not symmetric. Therefore in [Klár et al. 2016] it is necessary to adopt Newton-Ralphson root finding with GMRES for the asymmetric linear system solve. In the next section, we elaborate on our new model which enables the existence of an analytical energy.

4 ENERGETICALLY CONSISTENT INELASTICITY (ECI)

4.1 One-Dimensional Investigation

To motivate ECI, let’s start with applying a discrete plasticity model to a one-dimensional spring with a constant yield stress.

Consider a one-dimensional elastoplastic spring with rest length $V_0 = 1$. We fix its one end at $x = 0$, and place the other end at $1$ initially. With the initial state being the reference configuration, we model the spring with finite strain elastoplasticity, where the deformation gradient can be conveniently calculated as $F(x) = x$ with $x$ being the coordinate of its free end.

Discretizing time into steps $0, 1, \ldots, n$ with equal time step size $\Delta t$, for time step $n$, the elastic predictor $F^{E,tr}$ by assuming a purely elastic deformation is given by

$$F^{E,tr}(x) = \frac{1}{F_{n,n}^E} x.$$  

(13)

We assign the following strain energy density function:

$$\Psi^E(F^E) = \frac{k}{2} (\log F^E)^2,$$  

(14)

where $k$ is the stiffness. Viewing the spring as a single-element FEM discretization, since $V_0 = 1$, $\Psi^E$ equals the total elastic potential.

Fig. 6. The strain-energy (left) and the strain-stress (right) plot of an elastoplastic spring. Here $\epsilon = \log(F)$ and $\tau = \frac{\partial \Psi}{\partial F}.$

Assume $k = 1$ for brevity, the Kirchhoff stress is then given by

$$\tau(F^E) := \frac{\partial \Psi^E}{\partial F^E} F^E = \log(F^E).$$  

(15)

Let the constant yield stress be $\tau_0 = \log(F_Y)$, where $F_Y \in [1, \infty)$ is a critical strain, and define the yield function to be $\tau - \tau_0 \leq 0$, we can then follow standard plasticity treatment [Simo and Hughes 1998] to define a simple return mapping procedure with the form

$$F^{E,n+1} = \begin{cases} F^{E,tr} & F^{E,tr} \leq F_Y \\ F_Y & \text{otherwise}. \end{cases}$$  

(16)

In terms of the Hencky strain $\epsilon^E = \log(F^E)$, the yield condition is equivalent to

$$\epsilon^{E,tr} - \epsilon^{E,n+1} = \delta \gamma > 0.$$  

(17)

Geometrically, the quantity $\delta \gamma$ measures how far away the elastic strain predictor is from the yield surface in the principal strain space. This quantity plays an important role in our variational modeling of plasticity. Specifically, we have the following theorem for the elastoplastic springs:

**Theorem 4.1 (Augmented energy density for springs).** In the problem setting described above ($V^0 = 1$, $k = 1$), using the following energy density function

$$\Psi(x) = \begin{cases} \Psi^E(Z(F^{E,tr}(x))) + \gamma \delta \gamma (F^{E,tr}(x)) & F^{E,tr} > F_Y \\ \Psi^E(F_Y(x)) & \text{otherwise} \end{cases}$$  

(18)

reveals a force that is equivalent to what one would get if one performed the force-based implicit plasticity.

We include in the supplemental document the proof for this theorem as well as details showing that $\Psi(x)$ is piecewise $C^0$ and everywhere $C^1$. A comparison between the augmented energy and the pure-elastic energy is shown in Figure 6.

Taking the inertia into consideration, we test the model on a small dynamic mass-spring system. At the end of each time step, $F_p$ is updated from the following relation:

$$F_{n+1} = \frac{x_{n+1}}{x^n} = F^{E,tr} F_{P,n+1} = F^{E,tr} F_{P,n}.$$  

(19)

The ECI simulation results quantitatively match the results using explicit integration (see Figure 8).
4.2 Extending to Von-Mises Plasticity

A natural analogy of elastoplastic spring with constant yield stress for the plasticity of isotropic hyperelastic materials is the von-Mises plasticity model, which also associates all stress predictors with a constant yield stress \( \tau_Y \) (the norm of the deviatoric Kirchhoff stress on the yield surface). We study von-Mises plasticity under the St. Venant-Kirchhoff constitutive model with Hencky strains. Following the notations from Section 3.4, the yield surface is defined as

\[
y(\tau) = \|\hat{\tau}\|_F - \tau_Y = 0,
\]

where \( \hat{\tau} = \frac{1}{2} \text{tr}(\tau)I \) is the deviatoric part of the Kirchhoff stress.

The equivalent yield condition is

\[
\delta_Y = \|\hat{\epsilon}\| - \frac{\tau_Y}{2\mu} > 0,
\]

where \( \epsilon = \log(\Sigma^{tr}) \) is the trial Hencky strain and \( \hat{\epsilon} = \epsilon - \frac{1}{4} \text{tr}(\epsilon)I \) is the deviatoric part of the Hencky strain. The corresponding return mapping (Figure 9) is

\[
Z(F^{E,\tau}) = \begin{cases} 
F^{E,\tau} & \delta_Y \leq 0 \\
U \exp(\epsilon - \delta_Y \hat{\epsilon} \|\hat{\epsilon}\|_F) V^T & \text{otherwise}
\end{cases}.
\]

We have the following key lemma for the von-Mises plasticity:

**Lemma 4.2.** Define the augmented elastoplastic energy density function as:

\[
\Psi(F) = \begin{cases} 
\Psi^E(F), & \delta_Y(F) \leq 0 \\
\Psi^E(Z(F)) + \tau_Y \delta_Y(F), & \text{otherwise}
\end{cases}
\]

This energy density function satisfies the following identity for any \( F \):

\[
\frac{\partial \Psi(F)}{\partial F} \equiv \frac{\partial \Psi^E}{\partial F}(Z(F))Z(F)^T F^{-T}.
\]

The proof of this lemma is provided in the supplementary document. With this lemma, it is easy to prove the following theorem:

**Theorem 4.3 (Augmented energy theorem for von-Mises plasticity).** The augmented elastoplastic energy density function (Equation 23) viewed as a hyperelastic strain energy density function...
The corresponding return mapping (Figure 10) is

\[
f_f = -\sum_p v_p \left[ \frac{\partial \Phi}{\partial \mathbf{F}} (\mathbf{F}_p^{E,ir}) \right] \mathbf{F}_p^{E,ir} \nabla \mathbf{w}_p^\alpha
\]

\[
= -\sum_p v_p \left[ \frac{\partial \Psi}{\partial \mathbf{F}} (Z(\mathbf{F}_p^{E,ir})) Z(\mathbf{F}_p^{E,ir})^\top \mathbf{F}_p^{E,ir} \nabla \mathbf{w}_p^\alpha \right].
\]

When performing optimization time integration, we can simply view the elastoplastic free energy density as a new strain energy density function. At the end of each time step, we update \( \mathbf{F}^E \) with \( Z(\mathbf{F}_p^{E,ir}) \). The detailed pipeline is elaborated in Section 5.

4.3 Extending to Pressure Dependent Soil Plasticity

Drucker-Prager plasticity is widely applicable to the simulations of granular materials such as sand. The yield surface under the St. Venant-Kirchhoff constitutive model with Hencky strains is defined as

\[
y(\mathbf{r}) = \|\mathbf{r}\|_F + \alpha \operatorname{tr}(\mathbf{r}) = 0,
\]

where \( \alpha = \sqrt{\frac{2\sin \phi_f}{3(1-\sin \phi_f)}} \) and \( \phi_f \) is the friction angle.

The equivalent yield condition is then

\[
\operatorname{tr}(\mathbf{e}) > 0, \quad \text{or} \quad \delta_Y = \|\mathbf{e}\|_F + \alpha \left( \frac{d\lambda + 2\mu}{2\mu} \right) \operatorname{tr}(\mathbf{e}) > 0.
\]

The corresponding return mapping (Figure 10) is

\[
Z(\mathbf{F}_p^{E,ir}) = \begin{cases} \mathbf{U} \mathbf{V}^T & \operatorname{tr}(\mathbf{e}) > 0 \\ \mathbf{U} \exp(\mathbf{e} - \delta_Y \frac{\mathbf{e}}{\|\mathbf{e}\|}) \mathbf{V}^T & \delta_Y \leq 0, \operatorname{tr}(\mathbf{e}) \leq 0 \end{cases}.
\]

4.3.1 Extrapolating St. Venant-Kirchhoff. To solve the issue of \( \delta_Y \) for the area defined by \( \operatorname{tr}(\mathbf{e}) > 0 \), we extrapolate the St. Venant-Kirchhoff constitutive model in this area as:

\[
\Psi^E(\Sigma) = \begin{cases} \rho \|\mathbf{e}\|_F^2 & \operatorname{tr}(\mathbf{e}) \geq 0 \\ \rho \|\mathbf{e}\|_F^2 + \left( \frac{\mu}{2} + \frac{\mu}{4} \right) (\operatorname{tr}(\mathbf{e}))^2 & \operatorname{tr}(\mathbf{e}) < 0 \end{cases}.
\]

When \( \operatorname{tr}(\mathbf{e}) < 0 \), \( \Psi^E \) is just equivalent to the St. Venant-Kirchhoff strain energy density, which separates the deviatoric term and the pressure term. When \( \operatorname{tr}(\mathbf{e}) \geq 0 \), we extrapolate the energy only with the deviatoric term and define the yield stress to be zero. This extrapolation does not change the yield surface in the principal stress space. Instead, the yield surface in the principal strain space is extended to include the diagonal line of the first quadrant, and all the points on this ray correspond to the tip of the yield surface in the principal stress space (see Figure 11). With this extrapolated model, the volume-preserving projection can be done as well in the area of \( \operatorname{tr}(\mathbf{e}) \geq 0 \), and \( \delta_Y \) is well-defined.

In summary, with our extrapolation, the return mapping is simplified as

\[
Z(\mathbf{F}_p^{E,ir}) = \begin{cases} \mathbf{F}_p^{E,ir} & \operatorname{tr}(\mathbf{e}) > 0 \\ \mathbf{U} \exp(\mathbf{e} - \delta_Y \frac{\mathbf{e}}{\|\mathbf{e}\|}) \mathbf{V}^T & \delta_Y \leq 0, \operatorname{tr}(\mathbf{e}) \leq 0 \end{cases}.
\]

where

\[
\delta_Y = \begin{cases} \|\mathbf{e}\|_F & \operatorname{tr}(\mathbf{e}) > 0 \\ \|\mathbf{e}\|_F + \frac{d\lambda + 2\mu}{2\mu} \operatorname{tr}(\mathbf{e}) & \operatorname{tr}(\mathbf{e}) < 0 \end{cases}.
\]

4.3.2 Recover Integrability. To resolve the non-integrability, we update the yield stress iteratively during integration (Figure 13). At each time step, we solve a series of optimization problems with constant yield stresses. The yield stress \( r_Y \) for each particle \( p \) is computed from its elastic predictor \( \mathbf{F}_p^{E,ir} \) at the beginning of the optimization:

\[
r_Y = \begin{cases} 0, & \operatorname{tr}(\mathbf{e}) > 0 \\ -\alpha (d\lambda + 2\mu) \operatorname{tr}(\mathbf{e}), & \text{otherwise} \end{cases}.
\]

and the corresponding \( \delta_Y \) is defined with a fixed yield stress:

\[
\delta_Y = \begin{cases} \|\mathbf{e}\|_F, & \operatorname{tr}(\mathbf{e}) > 0 \\ \|\mathbf{e}\|_F - \frac{r_Y}{\mu}, & \text{otherwise} \end{cases}.
\]
The hardening mechanism plays an important role in simulations of materials like metal [Chakrabarty and Drugan 1988] and snow [Gaume et al. 2018; Stomakhin et al. 2013]. In general, the hardening mechanism is associated with some hardening state set \( q_n \) and some hardening parameter set \( \xi \). Theoretically, hardening controls how the yield surface evolves according to the hardening state.

A linear hardening rule for the von-Mises plasticity can be defined as

\[
q^{n+1} = q^n + 2\mu_\xi \delta \gamma (F^{E, tr}),
\]

\[
\tau^{n+1} = q^{n+1},
\]

\[
q^0 = \tau^{tr, \text{init}}.
\]

This effectively makes the yield stress \( \tau^{n+1} \) in the equilibrium state at time step \( t^n \) depend on \( F^{E, tr} \), which is not a constant anymore for different \( F^{E, tr} \). Similarly to the iterative stress update for Drucker-Prager, we can also iterate on the hardening state. At the beginning of each optimization, the trial hardening state and the trial yield stress are updated as

\[
\tau^{tr, i+1} = \tau^{tr, i},
\]

\[
\tau_n^{tr, i} = q_n^{tr, i} + \xi \delta \gamma (F^{E, tr}).
\]

At the end of the time step, the hardening state \( q^{n+1} \) is updated to be the last trial hardening state \( q^{tr, n} \).

### 4.5 Viscoelasticity

In addition to rate-independent elastoplasticity, ECI can also be applied to rate-dependent viscoelasticity. Here we model viscoelasticity based on a decomposition of the deformation gradient, which is independent of the elastoplastic decompostion. At each time step, the deformation gradient \( F \) can be decomposed into two different ways (Figure 14)

\[
F = F^E F^P = F^N F^V,
\]

where \( F^N \) is the non-equilibrated elastic deformation gradient, and \( F^V \) is the viscous deformation gradient. \( F^N \) and \( F^E \) provide elastic responses additively. The evolution of \( F^N \) follows a similar principle as \( p^E \), which is characterized by a return mapping-like projection in the discrete setting. We follow the formulation of Fang et al. [2019]:

\[
Z(F^{N, tr}) = U(A(e - B \tau(e)) I)^T V^T,
\]

where

\[
A = \frac{1}{\alpha + \Delta \tau \alpha}, \quad B = \frac{\lambda \Delta \tau}{\alpha + \Delta \tau \alpha}, \quad \alpha = \frac{2\mu_N}{\omega_d}, \quad \beta = \frac{2(2\mu_N + \lambda_N \alpha_d)}{9\omega_v} - \frac{2\mu_N}{\alpha_d \omega_d},
\]

and \( F^{N, tr} \) is the elastic predictor assuming no viscosity:

\[
F^{N, tr}_p = (I + \Delta \tau \nabla \Phi_p^{n+1}) F^{N, n}. \tag{41}
\]

Here \( \nu_p \) and \( \gamma_p \) are viscosity parameters, and \( \mu_N \) and \( \lambda_N \) are independent Lamé parameters for viscoelasticity to the ones for elastoplasticity. For simplicity, we use \( \nu_p = \gamma_p = 2\mu_N \omega_p \) for some \( \omega_p \).
Although the return mapping for viscoelasticity is totally different from the one for elastoplasticity, the vector field
\[
\frac{\partial \Psi^N}{\partial F^\text{E}} (Z(F) F^\text{F})^T F^{-T}
\]
turns out to be integrable if \( \Psi^N \) is from the St. Venant-Kirchhoff constitutive model, and the augmented ECI energy for this vector field is
\[
\Psi^{\text{Visco}}(\Sigma) = \dot{\mu} \text{tr}((\log \Sigma)^2) + \frac{\hat{A}}{2} \text{tr}((\log \Sigma)^2),
\]
where \( \dot{\mu} = A \mu_N \) and \( \hat{A} = A \lambda_N - AB(z \mu_N + d \lambda_N) \).

Without plasticity, \( F^\text{F} \equiv I \), and then the strain energy density for a viscoelastic material is simply
\[
\Psi(F) = \Psi^E(F) + \Psi^{\text{Visco}}(F).
\]

With MPM discretization, each particle \( p \) independently tracks the evolutions of \( F^N \) and \( F^E \) and independently updates them accordingly at the end of each time step.

5 SPATIAL-TEMPORAL INTEGRATION

In this section, we present the detailed pipeline of ECI applied to MPM. The algorithm stages from \( t^n \) to \( t^{n+1} \) are listed as follows:

1. **Particles-to-grid transfer.** Grid mass \( m_p^F \) and velocity \( v_p^F \) are transferred from particle mass \( m_p \), velocity \( v_p^N \), and angular velocity information \( C_{p}^N \) from APIC [Jiang et al. 2015].

2. **Optimize new grid velocity.** A series of optimization problems in the form of Equation 7 using ECI augmented energies are solved until the fixed-point iteration converges or the maximal number of iterations is reached. See Section 5.1.

3. **Grid-to-particles transfer.** The grid velocity \( v_{p}^{t+1} \) from the time integration are transferred back to particles to update particle velocity \( v_p^{n+1} \) and angular velocity information \( C_{p}^{n+1} \).

4. **Particle strain update.** The elastic strain \( F^E \) or \( F^N \) are updated according to return mappings.

5. **Particle advection.** Particles are advected via particle velocity:
\[
X_p^{n+1} = X_p^n + v_p^{n+1} \Delta t.
\]

We only elaborate on the second stage in the following section. The other stages are the same as the standard explicit MPM simulation pipeline [Jiang et al. 2016].

5.1 Iterative Stress Optimization Time Integration

To make the internal force of implicit plasticity integrable, the yield stress is viewed as constant in the force formulation, i.e., each particle sees a local cylinder-like yield surface with a different yield stress. Multiple optimizations with updated yield stresses are needed to make the final computed stresses consistent with the true yield surface. Each optimization problem is solved robustly using the projected Newton method with backtracking line search [Wang et al. 2020], where the Hessian matrix is projected to a nearby positive definite form [Teran et al. 2005]. See Algorithm 1 for the pseudocode.

The update procedure of \( \tau_Y^{t+1} \) can be viewed as a fixed point iteration:
\[
\tau_Y^{t+1} = \tau_Y^{t} (F^{E,tr}(\Delta v(\tau_Y^{t}))).
\]

Here \( j \) is the index of stress iteration, \( \Delta v(\tau_Y^{t}) \) is the equilibrated grid velocity field returned by a single optimization based on the yield stress vector \( \tau_Y^{t} \), and the bold symbol represents the stacked stress vector from all particles or all grid nodes. Since the Jacobian of this iteration has a scalar \( \Delta \tau^2 \) (see the supplemental document for details), the convergence of this fixed-point iteration is guaranteed if \( \Delta \tau \) and the residual for the equilibrium are both small enough.

In practice, we find that even with large time steps, only several fixed-point iterations are required to produce visually high-quality results.

5.1.1 Boundary Conditions. The boundary conditions in our simulations are all from rigid collision objects. At the beginning of each time step, we detect the set of grid nodes colliding with the collision objects and directly enforce the velocity continuity condition across

![Diagram](image-url)
we use an inexact Newton-Krylov method. The tolerance for the 
fully converge, but we have not observed non-converging cases.

5.1.5 Timestep Size Restriction. For those with stress-iterations, theoretically, there is indeed a timestep size restriction for the stress iteration to fully converge, but we have not observed non-converging cases.

5.1.4 Stopping Criteria. To terminate the Newton iterations early 
for the collision interface. In each Newton iteration, the linear solver is 
projected so that the solved search direction remains tangent to the 
constraint manifold.

5.1.3 Inversion-free Line Search. The Hencky strain requires that 
the deformation gradient is not inverted, i.e., \( \det(F^E,t^r) > 0 \). Following [Li et al. 2020, 2021c; Smith and Schaefer 2015], before the line search, we first compute a large admissible step size \( \alpha \) for the search direction such that the energy is well-defined for any step size \( t \in [0, \alpha] \), and then the backtracking procedure starts with the 
filtered step size \( \alpha \).

5.1.2 Inexact Newton-Krylov Methods. Following Wang et al. [2020], we use an inexact Newton-Krylov method. The tolerance for the 
linear systems is set to relatively large in an adaptive way. Although more 
Newton iterations are needed, the reduced linear solve cost can 
still improve the world-clock timing of Newton convergence. 
Specifically, we use matrix-free Minimal Residual Method (MINRES) 
to solve the linear systems and the relative tolerance of each MINRES 
solve is set to \( \min(0.5, 0.1, \sqrt{t/\tau_i}) \), where \( \tau \) is the right-hand 
side vector and \( P \) is the preconditioning matrix.

5.1.1 Stopping Criteria. To terminate the Newton iterations early 
while ensuring visually high-quality simulation results, we 
normalize the grid residual vector \( r \), the gradient of the system energy, by 
the grid mass vector. This gives a residual in the unit of velocity 
\( (m/s) \), which is associated with a physical meaning. However, due to 
numerical rounding errors, small-mass nodes sometimes can have 
large residuals but contribute little to the particle advection. 
Therefore, we use grid-to-particle transfer to transfer the grid residual 
vector onto the particles to get the final residual vector. All our 
examples are running with tolerance \( 10^{-2}m/s \) based on the infinity 
\( E_{\text{ramp}} \) of the velocity-unit residual vector on particles.

6 DISCRETIZATION WITH FEM 
ECI is independent of spatial discretization choices. Hence it can 
also be conveniently applied in Finite Element Methods (FEM). 
In FEM, the conservation-of-momentum equation (Equation 3) is 
directly discretized and solved in the material space. For FEM with 
linear tetrahedral elements, the discretized nodal internal force is 
\[ f_i = -\sum_e V_e^0 P_e \nabla N_{ie}, \] 

where \( e \) indices all tetrahedral elements, \( V_e^0 \) is the rest volume of 
element \( e \), and \( \nabla N_{ie} \) is the gradient of the shape function on node \( i \) 
evaluated at the barycenter of element \( e \) [Irving et al. 2006]. 
Considering implicit plasticity, the internal force can be written 
as (see the supplemental document for details) 
\[ f_i^{n+1} = -\sum_e V_e^0 \frac{\partial \Psi}{\partial F} (Z(F_e^{E,tr})Z(F_e^{E,tr})^T) F_e^{E,tr} \nabla N_{ie}. \] 
The integrability of the vector field \( \frac{\partial \Psi}{\partial F} \) 
leads us to the integrable internal force from the augmented elasto-
plastic energy density \( \Psi \): 
\[ f_i^{n+1} = -\frac{\partial \Psi}{\partial F} \nabla N_{ie}. \] 
where \( x_i \) is the world space coordinate of node \( i \). 
At the end of each time step, we need to track and update \( F^P \) on 
each element with 
\[ Z(F_e^{E,tr})F_e^{P,n+1} = F_e^{E,tr} F_e^{P,n}. \]
ECI combined with Incremental Potential Contact (IPC) [Li et al. 
2020] allows us to simulate various scenarios where both accurate 
frictional contacts and inelastic responses are essential.

Timestep Size Restriction. Similar to MPM, there is also a timestep 
size restriction for the stress iterations to fully converge in FEM. 
Other than that, no further restrictions are needed. However, there 
is certainly a tradeoff between the number of timesteps and the 
accuracy and overall efficiency of the simulation [Li et al. 2020], 
which holds for all time discretized numerical schemes.
Fig. 15. Sand Column Collapse. (a) The explicit method explodes with $\Delta t = 10^{-4}$s. The implicit method [Klár et al. 2016] with vanilla Newton fails at a time step where the scene almost becomes static $\Delta t = 10^{-5}$s, and produces consistent results. The semi-implicit method produces artificial elastic behaviors even with a small time step size. (b) With the explicit method as the ground truth, our method has a smaller error (larger IoU score) than the semi-implicit method.

Table 2. Iteration statistics of 2D sand column collapse.

<table>
<thead>
<tr>
<th>$\Delta t$</th>
<th># Stress Iter.</th>
<th># Newton Iter.</th>
<th># Line search</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Avg. / Max)</td>
<td>(Avg. / Max)</td>
<td>(Avg. / Max)</td>
</tr>
<tr>
<td>0.01</td>
<td>8.8 / 13</td>
<td>112.3 / 186</td>
<td>202.0 / 476</td>
</tr>
<tr>
<td>0.005</td>
<td>6.8 / 9</td>
<td>45.2 / 75</td>
<td>48.8 / 201</td>
</tr>
<tr>
<td>0.0025</td>
<td>5.1 / 7</td>
<td>18.4 / 34</td>
<td>9.6 / 60</td>
</tr>
<tr>
<td>0.00125</td>
<td>3.7 / 6</td>
<td>9.2 / 14</td>
<td>1.3 / 9</td>
</tr>
<tr>
<td>0.000625</td>
<td>2.6 / 4</td>
<td>4.7 / 8</td>
<td>0.0 / 0</td>
</tr>
</tbody>
</table>

7 EVALUATION
We demonstrate the versatility of ECI with both MPM and FEM simulations. Among these examples, the ones that do not contain topological changes are simulated with FEM, and the frictional contact is modeled with IPC [Li et al. 2020]. For our MPM simulations, we use a CFL number of 0.6 [Gast et al. 2015]. The world-clock timing and the simulation setup are reported in Table 1. The statistics are based on Intel Core i9-10920X 3.5-GHz CPU with 12 cores.

7.1 Unit Tests

Convergence of Stress Iteration.
We test the convergence of the stress iteration on a 2D sand column collapse experiment. We use a direct solver to solve linear systems in the optimization time integrator, to avoid complicating the experiments with possibly inexact Krylov solves. The convergence criteria of the stress iteration is $\| (\tau_{j+1} - \tau_j) \|_2 < 10^{-9} (2\mu \sqrt{N})$, where $N$ is the number of particles, and the Newton tolerance is $10^{-5}$. Note that these tolerances are much tighter than needed so that we can verify that our method can converge with high accuracy. We consecutively halve the time step size from $\Delta t = 0.01$s. All these tests successfully converge with the given convergence criteria and have consistent results (see the inset figure). The iteration statistics are listed in Table 2, which shows that as the time step size decreases, the required number of stress iterations, Newton iterations, and line searches all decrease as expected.

Long-Time Stability. To test the long-time stability of our method, we simulate a soda can being periodically compressed and stretched 10 cycles with $\Delta t = 10^{-2}$s (Figure 16). The Young’s modulus of the soda can is 7 GPa. The stored elastic energy over time is always bounded and it oscillates along with the compressing–stretching cycles, demonstrating the strong long-time stability property of our method.

7.2 Comparisons to Explicit and (Semi-)Implicit Plasticity
We compare our variational method with both explicit and implicit methods proposed in Klár et al. [2016] on a 3D sand column collapse
Increasing cohesion

where we compute the Intersection over Union (IoU) metrics be-
tween MPM grid mass distributions (computed as the ratio of the
number of common grid nodes to the number of union grid nodes).

that our method has a smaller error than the semi-implicit method,
material’s resistance to tensile deformation. Our method, on the
other hand, fully resolves plasticity in the implicit solve and does
not suffer from any such artifacts. We use the explicit method as
an integrable implicit force (Equation 12, with asymmetric force
Jacobian), the convergence of time integration can only be reached
if the initial guess is sufficiently close to the local optimum.
Fur-
thermore, the search performed by the Newton-Raphson iterations
(we refer it as the vanilla Newton method) can result in deforma-
tion gradients with non-positive determinants that cause simulation
failure.

We experiment under three different time step sizes \( \Delta t = 10^{-3}s, 10^{-4}s \) and \( 10^{-5}s \). Explicit MPM can run with \( \Delta t = 10^{-5}s \), but it
explodes with \( \Delta t = 10^{-4}s \) (Figure 15a top left). The vanilla Newton
method can run with \( \Delta t = 10^{-3}s \), but fails at a step when the
simulation almost becomes static with \( \Delta t = 10^{-4}s \) and at the first
step with \( \Delta t = 10^{-3} \) (Figure 15a top middle). Our method, on
the other hand, works well with all these three time step sizes and
produces consistent results (Figure 15a right).

A common heuristic treatment is to directly replace the grid
update step in the explicit MPM simulation with implicit time inte-
gration without plasticity and only conduct return mappings at the
end of the time steps. We refer to this elasticity-plasticity-decoupled
scheme as the semi-implicit method in this paper. Although its sta-

tility and convergence can be guaranteed by the optimization time
integration, the semi-implicit method can lead to severe artifacts
as shown in Figure 15a bottom left, where the forces provided by
the stresses outside the yield surface make the continuum behave
more like a purely elastic body. This is due to the ignorance of the
plasticity by the implicit solve, which in turn overestimates the
lability of our method on simulating different levels of chunkiness
(Figure 20).

Increasing cohesion

In practice, we can limit the number of Newton iterations and
Krylov iterations. On a sand column collapse experiment with the
same physical parameters and initial setup as above, our method
with \( \Delta t = 2 \times 10^{-5}s \) achieves \( 2\times \) speedup compared to the explicit
method with \( \Delta t = 10^{-3}s \), as shown in Figure 17. With 2 stress
iterations per time step, 1 Newton iteration per stress iteration, and
5 MINRES iterations per Newton iteration, our method can still
generate physically plausible results. To make it a fair comparison,
the maximal numbers of Newton iterations and GMRES iterations
are set to 2 and 10 respectively for Klár’s implicit method with the
same \( \Delta t = 2 \times 10^{-5}s \) as ours. The simulation using Klár’s method

does not go unstable in this setting, and its computational cost is
similar to ours, as expected.

7.3 Drucker-Prager Plasticity with Cohesion

Snow Castle. To further demonstrate the artifacts caused by fully
decoupling elasticity and plasticity, we simulate a snow castle hit
by a high-speed elastic fish. The snow is modeled with wet soil by
Druker-Prager plasticity with cohesion. With our variational model,
the fish smashes the snow castle into pieces completely. However,
with the semi-implicit method, the castle behaves like an elastic
body and ends up holding the fish in an unrealistic way.

Our extrapolated StVK constitutive model combined with the
volume-preserving return mapping plays a vital role in generating
fractures in this example. Intuitively, our scheme mimics the cohe-
sion behavior better because it allows particles to be compressed a
little before exerting resisting force. Under the same time step size
(\( \Delta t = 5 \times 10^{-5} \)), we use the result from the explicit method with the
extrapolated StVK model as the ground truth to compare the accu-
cacy between our method and semi-implicit methods (with/without
the extrapolated StVK model). The visual and quantitative compar-
sions in Figure 20 both show that our method is more accurate.

Snow Ball. We use the Druker-Prager plasticity model with cohe-
sion to simulate a snow ball hitting a static dragon (Figure 2). We
also simulate with different cohesion strengths to show the control-

ability of our method on simulating different levels of chunkiness
(Figure 18).

7.4 Von-Mises Plasticity (with Hardening)

Play-Doh Noodle. MPM can automatically handle topology changes.
By leveraging this feature, we simulate a Play-Doh modeled by the
von-Mises plasticity pressed through a cylindrical noodle mold (Fig-

ure 1 (top row)).

Hydraulic Tests on Metals. Hardening is widely observed in metals.
We simulate hydraulic tests on soda cans with different hardening

Fig. 19. **Snow Castle.** With our variational inelasticity model, the castle can be smashed into pieces after hitting by the fish, while with the semi-implicit method the castle behaves like an elastic body, holding the fish in an unrealistic way.

Fig. 20. Our method is more accurate visually and quantitatively than the semi-implicit methods with/without the extrapolated SIVK constitutive model.

Fig. 21. Different hardening coefficients lead to varying restorations towards the rest shape and generate different crushing patterns. From left to right, the hardening coefficient $\xi = 0.5, 0.3, 0$.

coefficients and compare with the simulation without hardening (Figure 1 bottom row). The hardening mechanism makes plastic deformations harder to happen as the yield surface expands. This lets the object restore its original rest shape partially when all boundary conditions are released. As shown in the last frame when the upper press withdraws (Figure 21), the red can with the largest hardening coefficient restores the most, and the orange can with no hardening almost does not restore at all. Furthermore, different hardening coefficients generate different crushing patterns. As shown in Figure 22, the deformation patterns in one of our compressed can match that from a real experiment. Modeling hardening also allows us to successfully capture the snap-through instability of metal, which can be observed in real experiments (see our video demonstration). When we swap in long steel pipes for the hydraulic tests (one cylindrical, one square), the crushing patterns also match real experiments well (Figure 23, 24).

Fig. 22. One of our hydraulic test simulations on metal cans generate a crushing pattern well matching that in a real video footage [Youtube 2021].

Fig. 23. **Hydraulic Test on a Cylinder Pipe.** The crushing pattern matches the result of a real-world experiment [Youtube 2021].

**Car Crash and Crushed Armadillo.** To further demonstrate the hardening behaviors of metals, we simulate a high-speed car crashing into another stationary car (Figure 4) and an armadillo rolling
Our method is most naturally “plug and play” when applied to J2 von Mises materials and finite strain viscoelastic materials. However, when extended to pressure-dependent plasticity or strain hardening mechanisms, additional iterations on the stress are necessary to achieve final convergence. In our examples, usually 1–2 stress iterations are sufficient to generate convergent or visually high-quality results. It is promising future work to devise theoretical and algorithmic improvements to guarantee and accelerate the convergence, particularly for accuracy-demanding applications.

The integrability of the implicit elastoplastic force depends on both the elastic model and the plastic model. For instance, although the combination of St. Venant-Kirchhoff elasticity with von-Mises plasticity adopted by ECI leads to a symmetric force Jacobian, neo-Hookean elasticity with von-Mises plasticity does not. It is an interesting future work to explore integrable approximations to other combinations.

ECI assumes the full-dimensional volumetric deformation gradient. Accordingly, our metal cans and pipes are all simulated with thin single-layer linear tetrahedral elements, which could potentially suffer from shear locking. It would be interesting to extend ECI to codimensional geometries like shells and rods [Narain et al. 2013].

Finally, our augmentation to the strain energy density function changes the conditioning of the global stiffness matrix. It is interesting future work to study its effect on the linear solve, and strategies to precondition the ECI-augmented system.

ACKNOWLEDGMENTS
We would like to thank Pingying Chen for narrating the video, assisting with the video editing, and her sublime finishing touches assisting with the video editing, and her sublime finishing touches on renderings, especially the teaser figure. We would also like to thank the anonymous reviewers for their valuable comments. This work has been supported in part by NSF CAREER 2153851, ECCS-2153863, ECCS-2023780, DOE ORNL contract 4000171342.

REFERENCES

Energetically Consistent Inelasticity for Optimization Time Integration • 52:15

8 DISCUSSION
In summary, we developed ECI, a new formulation that augments hyperelastic energy density functions to enable variational forms for a wide range of elastoplastic and viscoelastic materials. Our algorithm enables the fully implicit simulation of inelasticity in recently advanced optimization-based time integrators, embracing advantages of long-time stability, global convergence, large time step sizes, and high accuracy.

Fig. 24. Hydraulic Test on a Square Pipe. The crushing pattern matches the result of a real-world experiment [Youtube 2018].

Fig. 25. The semi-implicit von-Mises plasticity (right) overestimates the resistance response and results in a large error compared to the ground truth (left). Ours (middle) is much more accurate.

through a metal crushe driven by frictions (Figure 7). Both examples show realistic denting effects with sufficient restoration towards the rest shape enabled by hardening.

Comparison to Semi-Implicit Plasticity. We simulate a stiff elastic ball hitting a wall modeled by the von-Mises plasticity to compare our method with the semi-implicit method. As shown in Figure 25, the permanent deformations of the wall clearly show that the semi-implicit plasticity overestimates the material’s resistance. We use the explicit method as the ground truth to compare the position error of the wall (computed as the average squared norm of vertex position differences), which shows that our method is more accurate than the semi-implicit method.

7.5 Viscoelasticity
Memory foam is a typical material demonstrating the viscoelastic behaviors in the real world. We simulate a pillow made by memory foam pressed down by a hand for a while, and then we lift the hand suddenly. We successfully capture the intricate process where the pillow slowly recovers its rest shape, completely removing the imprint of the hand (Figure 12).