Reflections on Simultaneous Impact

Breannan Smith
Columbia University

Danny M. Kaufman
Columbia University

Etienne Vouga
Columbia University

Rasmus Tamstorf
Walt Disney Animation Studios

Eitan Grinspun
Columbia University

Abstract

Resolving simultaneous impacts is an open and significant problem in collision response modeling. Existing algorithms in this domain fail to fulfill at least one of five physical desiderata. To address this we present a simple generalized impact model motivated by both the successes and pitfalls of two popular approaches: pair-wise propagation and linear complementarity models. Our algorithm is the first to satisfy all identified desiderata, including simultaneously guaranteeing symmetry preservation, kinetic energy conservation, and allowing break-away. Furthermore, we address the associated problem of inelastic collapse, proposing a complementary generalized restitution model that eliminates this source of nontermination. We then consider the application of our models to the *synchronous* time-integration of large-scale assemblies of impacting rigid bodies. To enable such simulations we formulate a consistent frictional impact model that continues to satisfy the desiderata. Finally, we validate our proposed algorithm by correctly capturing the observed characteristics of physical experiments including the phenomenon of extended patterns in vertically oscillated granular materials.

CR Categories: I.6.8 [Simulation and Modeling]: Types of Simulation—Animation

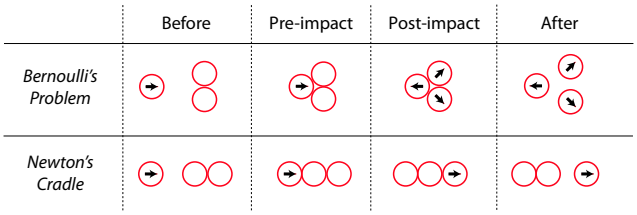
Keywords: physics, simulation, impact, mechanics, rigid bodies

Links: DL PDF

1 Introduction

Modeling the dynamics of sustained *contact* (resting, sliding) and of instantaneous *impact* (transient collisions, bouncing) is a fundamental thread of research in graphical simulation [Hahn 1988; Baraff 1989; Mirtich and Canny 1995; Witkin and Baraff 2001]. We focus on the open problem of modeling *multi-impact*, where either as a result of time-discretization or by consequence of formulation, we are asked to *simultaneously resolve multiple collisions occurring at an instant.* Models (e.g., sustained contact) and methods (e.g., penalty) that require finite time intervals to act are not considered here.

Consider arrangements of three balls at the instant they all collide:



In each case, leaving pre-impact velocities unchanged leads to penetration. These velocities must therefore be altered via instantaneous *impulses* to avoid penetration, i.e., to become ***feasible***. What does physics tell us about the requisite impulses and the attendant post-impact velocities?

(**BRK**) **Break away.** Bodies that were previously in contact may break away from each other as a result of impact. This might occur as an immediate consequence of the impact, as in *Bernoulli's Problem*, or it may be the result of *shock propagation*—a sequence of ordered events occurring at an instant—as in *Newton's Cradle.*

(**SYM**) **Symmetry preserved.** Spatial symmetries (e.g., about a reflection line) that exist in pre-impact configurations should also exist in post-impact configurations. After all, in an ideal system, what factor breaks such a symmetry [Bernoulli 1742]? As depicted above, both *Bernoulli's Problem* and *Newton's Cradle* are symmetric about the horizontal bisector.

(**KIN**) **Energy bounded.** *Elastic* impact ($\mathsf{c}_r = 1$) conserves kinetic energy. *Inelastic* impact under a ***coefficient of restitution*** ($0 \leq \mathsf{c}_r < 1$) reduces kinetic energy. Barring the esoteric case $\mathsf{c}_r > 1$, kinetic energy does not increase.

(**MOM**) **Momentum conserved.** Because impacts are internal to the (closed) system of bodies, total momentum is conserved.

(**ONE**) **One-sided impulses.** Impulses may push bodies apart but not pull them together—the so-called "no-velcro" condition.

These five physical desiderata define our notion of a correct algorithm for instantaneous impact. Given these goals, how well do existing families of models perform?

MODEL	(BRK)	(SYM)	(KIN)	(MOM)	(ONE)
Our model	•	•	•	•	•
LCP	X	•	•	•	•
Gauss-Seidel	•	X	•	•	•
Jacobi	•	•	X	•	•

Table 1: ***Multi-impact feature chart:*** *Physics demands a great deal from a multi-impact solution, and previous models punt on one goal or another.*

We summarize the answer in Table 1. Unfortunately, existing models, taxonomized by their use of a *linear complementary problem* (LCP), *Gauss-Seidel*, or *Jacobi* structure (§3), fail to satisfy either *break-away*, *symmetry preservation*, or *kinetic energy conservation*.

Contributions We propose a *generalized reflections* multi-impact operator that satisfies all five desiderata. In its simplest inter-

pretation the approach we will present amounts to a careful combination of the LCP and Gauss-Seidel formulations. And while forming hybrids of two methods is often a recipe for aggregating flaws, in this case, the resulting ***generalized reflection operator*** provably fulfills all desiderata.

Practical implications Beyond the satisfaction of capturing the inherent beauty of the physical laws, the fulfillment of the desiderata offers important practical benefits:

Breaking contact is an experimentally validated and expected behavior in stiff impact, as in the propagation of a shock in Newton's Cradle: impact from the leftmost ball propagates, traversing the chain, until the rightmost ball breaks away. On high-speed film it is observed that the propagation is *effectively instantaneous*, in that it traverses the domain at time scales separated by orders of magnitude from the gross dynamics.

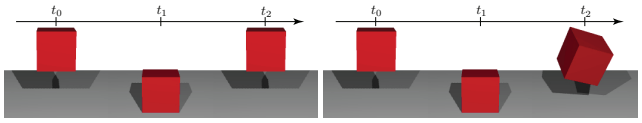


Figure 1: ***Symmetry preservation*** *requires a box dropped face-first onto a floor to bounce straight back up (left); symmetry breaking can manifest in unexpected ways, as with Gauss-Seidel (right).*

Shock waves are observed to propagate through contacting media in a time-ordered [Hascoët et al. 1999; Pudasaini and Kröner 2008], ***symmetry-preserving*** [Bernoulli 1742; Brogliato 1999] manner. Artificial symmetry breaking can yield unexpected results. Consider the simple canonical case of a rigid box dropped face-first onto a horizontal floor. At impact, standard box/box collision routines detect all four face vertices as colliding. Constraints are then generated by assigning these corner vertices the floor normal. Since the projection of the box's pre-impact velocity (i.e., its rigid-body twist) onto all four constraint normals is identical, the box's post-impact velocity should retain this symmetry and thus bounce straight back up (Fig. 1, left). Post-impact velocities that violate this symmetry, but otherwise satisfy all four remaining desiderata, contain non-zero angular velocity components that cause the box to incorrectly fly away along an oblique trajectory (Fig. 1, right). More generally, *all* symmetries between pre-impact velocities and constraints should be preserved by post-impact velocities.

Bounded energy, in both the conservative ($\mathsf{c}_r = 1$) and dissipative ($\mathsf{c}_r < 1$) regimes, leads to *stable* simulations. Energy behavior (whether conservative or dissipative) that remains independent of evaluation order and permutation of degrees of freedom (DoFs) provides more *consistent* simulations. Here we develop a model that guarantees bounded energy for the instantaneous resolution of simultaneous impacts.

Overview Our story begins by considering a single instant in time. We first explore the uniqueness of solutions in the case of multiple impacts (§2). Through a careful analysis of existing models (§3) we then arrive at our new *generalized reflections operator* (§4). In the presence of restitution, *inelastic collapse* challenges termination of this and other impact models. However, we show how such collapse can be entirely avoided with the application of a simple energetic restitution model (§5). To conclude our instantaneous story we present a compatible friction formulation (§6) and a scalable numerical implementation (§7).

While the instantaneous story reveals the inherent beauty of the underlying physics, any practical application must consider integration over finite amounts of time. The extension to handle this poses real challenges (§8). We address some of these, which leads us to propose a collision time-integration algorithm (§8.2) that is applicable to rigid body simulation and preserves all of the desiderata presented earlier. To validate the method we consider a wide range of benchmark examples (§9, §10, and §11). Finally, looking forward, we conclude with a discussion of limitations and open questions that this work raises (§12).

2 Impact

Impacts and contact occur whenever we impose one-sided constraints between objects or DoFs. A constraint is expressed by an inequality $g_i(\mathsf{q}) \geq 0$, where the constraint function $g_i : \mathsf{Q} \to \mathbb{R}$ maps each configuration $\mathsf{q} \in \mathsf{Q}$ to a non-negative number iff q is admissible. Trajectories, $\mathsf{q}(t) \in \mathsf{Q}$, are then required to be non-negative, $g_i(\mathsf{q}(t)) \geq 0$, for all time $t \in [0, T]$.

Our story takes place at instants where objects touch, i.e., $g_i(\mathsf{q}) = 0$. Differentiating the constraint with respect to time then gives

$$\nabla g_i(\mathsf{q})^T \dot{\mathsf{q}} \geq 0. \tag{1}$$

We cannot *sustain* a velocity opposing the normal $\nabla g_i(\mathsf{q})$.

Contact treats the case $\nabla g_i(\mathsf{q})^T \dot{\mathsf{q}} \geq 0$ when $g_i(\mathsf{q}) = 0$.

Impact treats the case $\nabla g_i(\mathsf{q})^T \dot{\mathsf{q}}^- < 0$ when $g_i(\mathsf{q}) = 0$. Here the pre-impact velocity $\dot{\mathsf{q}}^-$ opposes a constraint normal, necessitating an *impulsive* change to yield a post-impact velocity $\dot{\mathsf{q}}^+$ satisfying (1).

Isolated impact(s) At an instant when exactly one constraint experiences impact, $\dot{\mathsf{q}}^+$ is uniquely determined from conservation of momentum by the (c_r scaled) reflection

$$\nabla g(\mathsf{q})^T \dot{\mathsf{q}}^+ = -\mathsf{c}_r \, \nabla g(\mathsf{q})^T \dot{\mathsf{q}}^-. \tag{2}$$

Two cases have special properties: *elastic* impact ($\mathsf{c}_r = 1$) conserves energy; *purely inelastic* impact ($\mathsf{c}_r = 0$) dissipates more energy than any other momentum-respecting response.

At an instant when *two* constraints experience impact, we might get lucky with an easy case: if the normals $\nabla g_1(\mathsf{q})$ and $\nabla g_2(\mathsf{q})$ are orthogonal then each constraint is safely isolated as above. Two independent collisions across the room from each other, for example, possess orthogonal normals. This lucky strike generalizes to n simultaneous impacts, when all n normals are mutually orthogonal.

Multi-impact A more interesting case occurs at the instant where n constraints experience impact, with general (not necessarily orthogonal) constraint normals $\nabla g_1(\mathsf{q}), \ldots, \nabla g_n(\mathsf{q})$. This is the typical situation when multiple bodies collide at once, when a body collides against another with multiple points of contact, or when a particle collides against a kink of an enclosing boundary.

In the elastic case ***the core properties of elastic impact appear to delimit, but not uniquely pin down, the impulse***. This issue is raised in multiple works [Moreau 1988; Ivanov 1995; Brogliato 1999; Chatterjee and Ruina 1998; Glocker 2004] that analyze the time-continuous setting. In essence, we understand how a particle bounces off of a wall, but not how it bounces off of an arbitrarily-shaped corner. Since core conservative properties do not uniquely prescribe impact at a kink, we seek a principled way of choosing one such canonical outbound trajectory out of the myriad options.

The resolution of multi-impact is thus tricky. Yet, it would be dangerous to treat multi-impact as a degenerate case, since in practice the resolution of multiple impacts is the *typical* case. For

example, (a) when *synchronous* time stepping methods advance, *all* constraints violated en route are treated as simultaneous collisions [Bridson et al. 2002]; (b) asynchronous *time of impact* (TOI) methods are used to compute the motion of huge, multi-body systems [Lubachevsky 1991], here the probability of multiple simultaneous impacts increases with problem size; (c) for dissipative physical systems, such as those with $\mathsf{c}_r < 1$, the average time between collisions gradually decreases, consequently the probability of multi-impact increases (until ultimately a stable sustained contact forms).

3 Simultaneity vs. Propagation

Existing multi-impact models fall into two categories [Baraff 1989]: they focus on either *simultaneity* or *propagation*, as exemplified by Bernoulli's and Newton's problems, respectively.

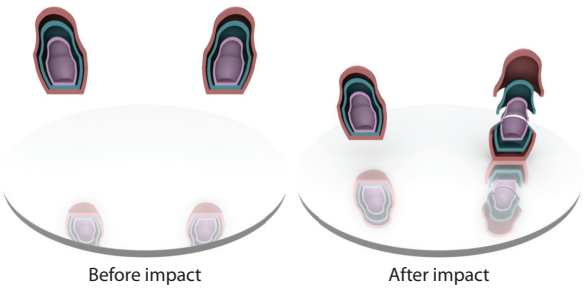


Figure 2: ***Matryoshka Dolls****: A set of dolls dropped on a table rebounds as a solid object with LCP due to lack of break-away (left doll) but separates with Generalized Reflections (right doll).*

3.1 Simultaneity

The Linear Complementarity Program (LCP) approach to multi-impact begins by determining the *active constraint set* $\mathbb{A}(\mathsf{q}) = \{i : g_i(\mathsf{q}) = 0\}$, and then requires the impact-level *Signorini-Fischera condition* [Moreau 1983; Baraff 1989; Stewart 2000]

$$0 \;\leq\; \lambda \;\perp\; \mathsf{G}_{\mathbb{A}}^T \dot{\mathsf{q}}^+ \;\geq\; -\mathsf{c}_r\, \mathsf{G}_{\mathbb{A}}^T \dot{\mathsf{q}}^-, \tag{3}$$

where $\mathsf{G}_{\mathbb{A}}$ is the matrix with columns $\{\nabla g_i(\mathsf{q}) : i \in \mathbb{A}\}$, $\lambda \in \mathbb{R}^{|\mathbb{A}|}$ is the vector of impulse coefficients, and $\mathsf{x} \perp \mathsf{y}$ is the *complementarity condition* $\mathsf{x}_i \mathsf{y}_i = 0,\ \forall i$. Because $\mathsf{G}_{\mathbb{A}}$ is not generally full-rank, the λ satisfying the Signorini-Fischera condition is not necessarily unique; nevertheless, when scaled by the inverse mass matrix M^{-1}, this impulse leads to a unique post-impact velocity

$$\dot{\mathsf{q}}^+ = \dot{\mathsf{q}}^- + \mathsf{M}^{-1}\mathsf{G}_{\mathbb{A}}\lambda.$$

Benefits and failures of LCP As the LCP considers only spanning spaces ($\mathrm{span}\ \mathsf{G}_{\mathbb{A}}$), ignoring the choice of basis vectors (∇g_i), it is *geometric* in the sense of being *basis- (or coordinate) independent*. Concretely, it (a) produces a result independent of DoF or sampling permutations and (b) *preserves symmetry* by construction.

Furthermore, for purely inelastic impact ($\mathsf{c}_r = 0$), the LCP formulation produces the *unique* solution that maximally dissipates normal velocities [Moreau 1983], *satisfying all desiderata*.

Indeed, LCPs satisfy one-sidedness (ONE) for any c_r, by construction. This, then, serves to highlight the distinction between one-sidedness (ONE) and break-away (BRK). For in the elastic case ($\mathsf{c}_r = 1$), LCPs do not satisfy (BRK) [Chatterjee and Ruina 1998; Glocker 2004]: the LCP solution of the elastic Newton example exhibits *sticking* where we expect break-away (see Fig. 3). Intuitively, the LCP solution reflects each relative velocity; when the relative velocity is zero (sustained contact), the reflected relative velocity remains zero, so that LCP has pinned the constraint, instead of allowing breaking contact. Put precisely, LCPs "stick" by producing zero (rather than positive) post-impact relative velocities, $\nabla g_i(\mathsf{q})^T \dot{\mathsf{q}}^+ = 0$, at contact points where corrective impulses, $\lambda_i > 0$, have been applied. As a consequence, LCP solutions do not capture shock-propagation effects.

Thus, while LCPs might be *the ideal solution* for purely inelastic multi-impact, as given below[1] in Alg. 1, LCPs *do not and can not* correctly treat multi-impact in general.

Algorithm 1 `Inelastic_Impact`$(\mathsf{q}, \mathsf{p}, \mathbb{A})$

1: $\mathsf{G} \leftarrow \mathsf{G}_{\mathbb{A}}(\mathsf{q})$
2: $\lambda \leftarrow \mathrm{argmin}_{\mathsf{y}} \left(\frac{1}{2}(\mathsf{G}\mathsf{y} + \mathsf{p})^T \mathsf{M}^{-1} (\mathsf{G}\mathsf{y} + \mathsf{p}) : \mathsf{y} \geq 0 \right)$
3: **return** λ

LCP's behavior for our two model problems summarizes its strengths and weaknesses as a method for elastic impact. LCP produces the correct behavior for Bernoulli's problem because it preserves symmetry. However, LCP produces incorrect *sticking* for Newton's Cradle (see Fig. 3).

3.2 Propagation

Pairwise propagation models leverage the well-posed behavior and computational ease of resolving a single-point impact. Dating back to the foundations of impact mechanics [Maclaurin 1742; D'Alembert 1743], these methods sequentially resolve each collision in isolation. Because each collision is treated separately, the communication between collisions occurs *explicitly* when a previously-treated collision is revisited; in general, propagation models iterate multiple times over the full set of active collisions.

Some variants use random ordering for the sequence [Ivanov 1995], others invoke physical [Chatterjee and Ruina 1998] or geometric [Johnson 1976; Ivanov 1995] considerations. Two popular variants [Bridson et al. 2002; Guendelman et al. 2003] mimic the styles of the simplest iterative linear solvers: the *Jacobi* variant first computes every pairwise impulse, and then applies them all [Maclaurin 1742], whereas the *Gauss-Seidel* variant computes and applies each impulse in turn [D'Alembert 1743; Johnson 1976; Ivanov 1995; Chatterjee and Ruina 1998]. All of these variants are not to be confused with the similar-sounding names of iterative splitting methods used to solve LCPs [Cottle et al. 1992; Erleben 2007].

Benefits and failures of pairwise propagation Pairwise propagation models enable both breaking contact and shock-propagation. However, all such methods *either* violate energy conservation, e.g., Jacobi (see Sec. 9), leading to large instabilities or excess dissipation, *or* destroy symmetry, e.g., Gauss-Seidel (see Fig. 1, right), and thus produce non-physical and inconsistent trajectories (see also Figs. 3, 7, and 5).

Recap Thus far, we have observed that the standard *simultaneous* LCP impact model is only well-behaved for $\mathsf{c}_r = 0$, but otherwise is prone to *sticking*. On the other hand, while *pairwise propagation* models recover break-away behavior, they violate either energy conservation or symmetry preservation.

[1] The inelastic LCP given by (3) with $\mathsf{c}_r = 0$ is the KKT optimality condition [Boyd and Vandenberghe 2004] for the equivalent minimization employed in Alg. 1, line 2. We denote the system's momentum as $\mathsf{p} = \mathsf{M}\dot{\mathsf{q}}$.

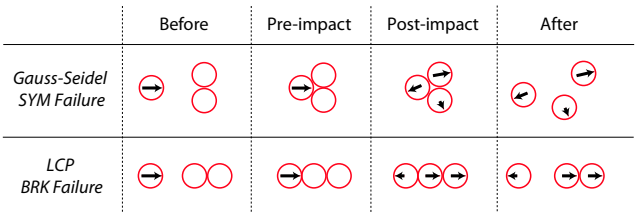


Figure 3: ***Simultaneous vs. propagation perspective:*** *The* simultaneous *approach of LCP fails to capture the break-away of Newton's Cradle, while the* Gauss-Seidel variant of pairwise propagation *fails to capture the symmetry of Bernoulli's problem.*

4 Simultaneity *and* Propagation

Generalized reflection We begin our development with a simple, yet critical, observation:

> **Lemma.** In the special case where the approaching velocity opposes *all* constraint-normal directions, the LCP model generates a unique solution that is *always free of both sticking artifacts and feasibility violations*, for any $0 \leq \mathsf{c}_r \leq 1$.

Proof. The LCP in (3) generates a *unique* post-impact velocity for all feasible problems [Cottle et al. 1992]. By assumption, all normal velocities are initially negative. Then, by the Signorini-Fischera condition,

$$\mathsf{G}_{\mathbb{A}}^T \dot{\mathsf{q}}^+ \;\geq\; -\mathsf{c}_r\, \mathsf{G}_{\mathbb{A}}^T \dot{\mathsf{q}}^- \;> 0,$$

for all $\mathsf{c}_r > 0$. □

In particular, for this special case, the LCP solution is a *generalized reflection*, in the sense that the component of the velocity in the subspace spanned by all pushing normal directions (i.e, directions along which an impulse is applied) is negated (and scaled by c_r), while the component in tangential directions remains untouched. The solution also inherits the energy and symmetry preservation properties inherent to the LCP formulation.

Generalized reflections impact operator During multi-impact, the incoming velocity will generally oppose *some, but not all* normal directions. Let these opposing normals define a *violator subset* of the *active set* of constraints, $\mathbb{V} \subset \mathbb{A}$. If we temporarily ignore the remaining active constraints and apply LCP *just* to the violator subset, our special-case *generalized reflection*, described above, yields the *unique* impulse that recovers all core properties. This feasible "post-impact" velocity no longer opposes the violator normals.

However, having ignored the remaining active constraints, this velocity may not be feasible with respect to the full set of original constraints. Thus, again, we gather the set of opposing normals, apply the generalized reflection, and repeat (Alg. 2).

This resulting *Generalized Reflections (GR) impact operator* consisting of back-to-back special-case impulses, *directly corresponds to the view of a shock wave as an advancing front of constraint violations.* A shock wave can be viewed as an *instant* in time—positions are fixed—during which a *moving front* of velocity modifications sweeps over the material. In particular, feasibility is satisfied everywhere except on the moving front, where constraint-restoring impulses induce new violations ahead of the front and thus advance the front forwards.

Algorithm 2 `Generalized_Reflection`$(\mathsf{q}, \mathsf{p}, \mathbb{A}, \mathsf{c}_r)$

$\epsilon \leftarrow 1 + \mathsf{c}_r$
$\lambda \leftarrow 0$
$\tilde{\mathsf{p}} \leftarrow \mathsf{p}$
while true **do**
  $\mathbb{V} \leftarrow \emptyset$
  for k in $\mathbb{A}$ **do**
    if $\nabla g_k(\mathsf{q})^T \mathsf{M}^{-1} \tilde{\mathsf{p}} < 0$ **then**
      $\mathbb{V} \leftarrow k$
    end if
  end for
  if $\mathbb{V} \neq \emptyset$ **then**
    $\mathsf{G} \leftarrow \mathsf{G}_{\mathbb{V}}(\mathsf{q})$
    $\tilde{\lambda} \leftarrow \mathrm{argmin}_{\mathsf{y}} \left(\frac{1}{2} (\mathsf{G}\mathsf{y} + \epsilon\tilde{\mathsf{p}})^T \mathsf{M}^{-1} (\mathsf{G}\mathsf{y} + \epsilon\tilde{\mathsf{p}}) : \mathsf{y} \geq 0 \right)$
    $\tilde{\mathsf{p}} \leftarrow \tilde{\mathsf{p}} + \mathsf{G}\tilde{\lambda}$
    $\lambda \leftarrow \lambda + \tilde{\lambda}$
  else
    return λ
  end if
end while

5 Restitution and Inelastic Collapse

Pairwise propagation methods are well known to suffer from poor convergence whenever $\mathsf{c}_r < 1$ is applied; in the extreme case, they cannot converge in finite iterations due to *inelastic collapse* [Baraff 1989; McNamara and Young 1994]. A simple exercise is to observe that for Newton's Cradle with $\mathsf{c}_r = 0$, each iteration halves the negative relative velocity, thus by Zeno's Paradox ("Achilles and the tortoise") finite iterations cannot reach a feasible velocity.

More generally, inelastic collapse extends well above the fully inelastic case of $\mathsf{c}_r = 0$ (see Fig. 4, (a) and (b)). Indeed, the range of c_r for which inelastic collapse can occur increases as the size of an impacting system grows (see Fig. 4, (c)), and quickly approaches unity as the number of colliding bodies become sufficiently large [Bernu and Mazighi 1990; McNamara and Young 1994]. Thus inelastic collapse is effectively unavoidable for the large-scale colliding systems that we consider.

In the numerical setting, round-off somewhat ameliorates this issue and the iterative process generally terminates [Chatterjee and Ruina 1998]. Relying on round-off is not acceptable, however. Moreover, we observe that convergence behavior consistently worsens in proportion to the decrease in c_r (see Fig. 4, (d)). While iterating the generalized-reflection operator with a $\mathsf{c}_r < 1$ would require fewer iterations than pairwise propagation, GR is nevertheless a propagation approach, and so it would not escape inelastic collapse.

An energetic restitution model safe from inelastic collapse Instead of iteratively resolving impacts at $\mathsf{c}_r < 1$ and suffering the consequences of inelastic collapse, we propose a simple and, to our knowledge, novel energetic restitution model that obtains consistent convergence and dissipation behaviors across all c_r values. We first observe that purely inelastic multi-impact is well-posed and soluble using the standard LCP formulation [Moreau 1983; Anitescu and Potra 1997; Stewart 2000], i.e., (3) with $\mathsf{c}_r = 0$, yielding $\dot{\mathsf{q}}_0^+$. Similarly, elastic multi-impact is unaffected by collapse; applying the generalized reflection operator to the elastic case we obtain $\dot{\mathsf{q}}_1^+$. We view c_r as the *interpolant* between the two and thus obtain

$$\dot{\mathsf{q}}^+ = (1 - \mathsf{c}_r)\dot{\mathsf{q}}_0^+ + \mathsf{c}_r \dot{\mathsf{q}}_1^+.$$

Notice that this definition of c_r now gives the exact interpolation between the maximum (physically) allowable dissipation and total

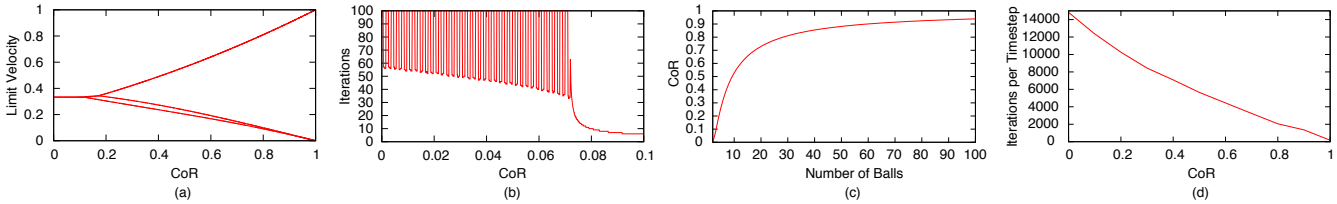


Figure 4: ***The danger of propagation with*** $\mathsf{c}_r < 1$***:*** *Consider the three-ball Newton's Cradle example with* $\mathsf{c}_r = 0$*; a single pairwise iteration halves the negative relative velocity at a constraint, implying that pairwise iterations only terminate in the limit with a fixed outgoing velocity. In (a) we plot the post-response velocities of the three balls as a function of* c_r*. Note that a common limiting velocity occurs across a significant range of* c_r *values, suggesting that within this span, pairwise iterations will require an infinite number of iterations. This is indeed the case as illustrated in (b), where we plot the number of iterations required to reach a collision free state as a function of* c_r*. For* $\mathsf{c}_r < 0.0717$ *the graph oscillates at saturation of maximum iterations – if we performed computations with unlimited precision, an infinite number of iterations would be required. This is known as "inelastic collapse". Even worse, the range of dangerous* c_r *values increases with the number of impacting bodies. Consider (c) where, following the analysis of Bernu et al. [1990], we plot the value of* c_r *at which inelastic collapse occurs as a function of the number of balls in a Newton's cradle. Observe that the range quickly approaches* $\mathsf{c}_r = 1$ *(the same effect is observed more generally). Finally in (d) we show that in actual numerics, the cost of pairwise propagation (here with pairwise Gauss-Seidel for a box of 10,000 balls) correspondingly blows up as* c_r *decreases.*

conservation of energy. In the special case of an isolated impact, we recover the classical isolated restitution model (2).

By construction this interpolation satisfies (KIN), (MOM), (ONE), (SYM), and (BRK). Feasibility follows from the linearity of the impact constraints.

6 Friction

As in contact, friction is a critical and often dominant aspect of transient collision behavior [Brogliato 1999]. We incorporate impulsive friction (Alg. 3) while preserving the physical desiderata. The short time-scales of impact simplify the computation.

The impact solution, λ, gives the magnitude of the normal impulse and thus, in analogy to frictional *contact*, defines local, pointwise Coulomb inequalities, $\|\boldsymbol{f}_k\| \leq \mu_k \lambda_k$, for friction impulse feasibility at each collision point k. In the instantaneous setting, as with normal directions, the directions of tangential dissipation are effectively fixed and thus the maximal dissipation of friction [Stewart 2000] is reduced to impulses applied along a single *fixed* tangential direction per collision point.

Instantaneously we have

$$\dot{\mathsf{q}}^+ = \dot{\mathsf{q}}^- + \mathsf{M}^{-1}\mathsf{G}\lambda + \mathsf{M}^{-1}\mathsf{D}\beta, \tag{4}$$

where λ is the impact solution, D is the generalized basis of instantaneous friction directions that oppose velocity at each collision point, and β is the vector of corresponding frictional impulse magnitudes. Maximal dissipation then gives the frictional impulses as

$$\begin{aligned} \beta = \underset{\mathsf{z}}{\operatorname{argmin}} \;\; & \mathsf{z}^T\mathsf{D}^T\left(\dot{\mathsf{q}}^- + \mathsf{M}^{-1}\mathsf{G}\lambda + \mathsf{M}^{-1}\mathsf{D}\mathsf{z}\right) \\ \text{s.t.} \quad & 0 \leq \mathsf{z} \leq \operatorname{diag}(\mu)\lambda. \end{aligned} \tag{5}$$

To compute the friction basis D we let Γ_k denote the relative velocity Jacobian [Kaufman et al. 2008] so that $\boldsymbol{v}_k = \Gamma_k\dot{\mathsf{q}}^- \in \mathbb{R}^3$ gives the relative velocity at collision point k. Extracting the relative velocity local tangent component, $\boldsymbol{v}_{t,k}$, we then generate a *single* friction basis contribution per collision point, $\mathsf{d}_k = \Gamma_k^T\boldsymbol{v}_{t,k}/\|\boldsymbol{v}_{t,k}\|$. The generalized friction basis follows as $\mathsf{D} = (\mathsf{d}_1, ..., \mathsf{d}_m)$.

Solving maximal dissipation globally guarantees monotone energy decay at all active contacts thus ensuring (KIN), our tangent basis choice preserves (MOM), while *basis independence* continues to maintain (SYM). We then inherit (BRK) and (ONE) from GR.

Algorithm 3 $\texttt{Friction}(\mathsf{q}, \mathsf{p}, \mathbb{A}, \lambda)$

$\mathsf{G} \leftarrow \mathsf{G}_{\mathbb{A}}(\mathsf{q})$
$\mathsf{D} \leftarrow \mathsf{D}_{\mathbb{A}}(\mathsf{q}, \mathsf{p})$
$\beta \leftarrow \operatorname{argmin}_{\mathsf{z}}(\mathsf{z}^T\mathsf{D}^T\mathsf{M}^{-1}(\mathsf{p}+\mathsf{G}\lambda+\mathsf{D}\mathsf{z}) : 0 \leq \mathsf{z} \leq \operatorname{diag}(\mu)\lambda)$
return β

7 Numerical Implementation

Our method requires the solution of numerous LCPs. Resolving even a single, large-scale LCP has long been considered computationally burdensome and has thus motivated many approximations and failsafes [Erleben 2007; Harmon et al. 2008]. We follow the observation that each impact LCP problem corresponds to the optimality conditions of a convex Quadratic Program (QP) [Boyd and Vandenberghe 2004]. In particular, we solve the large-scale, sparse impulse QP, dual to the LCP, employing an interior-point solver with a swappable linear backend. As we will see in §11 this leads to practical timings for exact LCP solves, even for very large ($> 2M$ DoFs and $> 2M$ constraints) impact systems.

QP solution The constraint gradients forming the linear constraints at each such solve are highly sparse[2]. To exploit the sparse problem-structure of these QPs we employ an interior-point method. Here the computational crux is to robustly and efficiently solve the many repeated inner-loop linear, KKT systems [Boyd and Vandenberghe 2004]. We use the Ipopt solver [Wächter and Biegler 2006] where the ability to customize our choice of an adaptive, swappable suite of linear-solver backends far outweighs the overhead of employing a general-purpose code. In particular, we employ MA27 [HSL 2001] as our first-line, linear solver with MUMPS [Amestoy et al. 2001; Amestoy et al. 2006] invoked in the rare event that MA27 fails to solve a linear system.

For the impact QP, we observe that imposing non-negativity as a bound constraint yields a significant performance increase over imposing non-negativity as an affine constraint. In the case of a single friction disk sample, the Coulomb constraint similarly reduces to a bound constraint, which therefore admits the same optimization.

QP solvers and scalability While scaling our method to larger simulations, we tested a number of algorithms for solving the im-

[2]Non-zeros in each column are generally restricted to the DoF stencil of any two in-contact surface patches.

pact and friction QPs. For QPs with order 100 constraints, we find that the QL [Schittkowski 2005] implementation of the Goldfarb and Idnani [Goldfarb and Idnani 1983] dual active set method performs admirably. In fact this is what we have used for the small simulations in section 9. QL is a dense method, however, and as our simulations approach order 1,000 constraints, storing the normal equations in a dense manner grows computationally prohibitive.

For QPs of order 1,000 to 10,000 constraints, we tested an operator-based non-negative least-squares (NNLS) [Lawson and Hanson 1974] approach that avoids the normal equations altogether. We employ a two-metric, projected-descent solver [Friedlander 2007] that enables warm starting and requires only the evaluation of operator-based callbacks for multiplications of impulse subvectors by submatrices of the constraint gradient. As we scale to larger contact problems, however, the NNLS approach requires increasingly aggressive applications of Tikhonov regularization, degrading overall performance compared to the interior point approach.

8 From One Instant to Finite Time

Up until this point, our intentionally myopic discussion has focused exclusively on a *single instant in time.* We have developed an instantaneous multi-impact operator that, for the first time, is able to fulfill five important physical principles that apply to instantaneous impact. A natural question is what happens in the surrounding *interval* of time, and in particular whether an algorithm that satisfies the five principles of instantaneous impact fares better than others when employed as one component of a *finite* time integrator. In the remainder of this paper we explore this question.

As in the instantaneous picture, finite time behavior can be studied in terms of break-away, symmetry, momentum, the no-velcro condition, and energy conservation/boundedness. The last criterion however raises a subtle point.

8.1 Finite Time Energy Conservation

To simplify the discussion consider purely elastic restitution ($\mathsf{c}_r = 1$). In the instantaneous picture, positions were fixed, only momenta varied, and we sought exact conservation of *kinetic* energy. In the finite time picture, both configurations and momenta evolve, and we consider instead the requirement of exact conservation of the *Hamiltonian*, or *total* energy, over extended durations of simulation, which we denote (HAM). For inelastic impact ($\mathsf{c}_r < 1$) we relax (HAM) to allow for bounded rather than conserved total energy.

Do impact operators satisfying (KIN) lead to integration algorithms satisfying (HAM)? When we combine instantaneous impact response with finite time integration, we must distinguish between three related notions of "energy behavior:"

- where the long-term Hamiltonian conservation of the discrete time integrator alone (without impact response) satisfies (HAM);
- whether the impact response *instantaneously* obeys (KIN);
- where the long-term Hamiltonian conservation of both pieces combined satisfies (HAM).

If either the integrator has poor energy behavior, or the impact response violates (KIN), it is exceedingly unlikely that their aggregate will have good behavior. On the other hand, if both exactly conserve energy, their combination will as well. In the remainder of this paper, we explore the latter case, identifying two classes of physical systems where it is possible to combine GR with energy-momentum preserving integrators. We then discuss inherent limitations and pose further avenues for exploration.

8.2 Synchronous Time Integration

In our finite time exploration we consider the typical case of *synchronous* or fixed-time step collision integration. At the beginning of each time-step we gather all constraints expected to be violated along the course of a fixed-size time-step and then resolve *all* of them instantaneously and simultaneously. We do this by applying our chosen method (GR, Gauss-Seidel, Jacobi, or LCP) for `Elastic_Impact` interpolated, by our restitution model, with `Inelastic_Impact`. Friction is then applied, followed by a fixed time-step with an unconstrained, free-flight integrator, denoted `FF_Integrator`. The resulting collision-integration method is given below in Alg. 4.

While the solution of the maximal dissipation problem in impact remains (as with contact) coupled to constraint resolution forces [Kaufman et al. 2005], we will assume that most of the impacts are sufficiently high-speed that it is not critical to find the exact equilibrium between friction and contact forces. As such we discretize in time by applying a single pass sequence of GR followed by a maximal dissipation solve of the resulting QP in equation (5) above with λ given. If a more accurate friction solution is needed then additional staggered iterations can be made [Kaufman et al. 2008].

Algorithm 4 `Collision_Integrator`$(\mathsf{q}^t, \mathsf{p}^t, h, \mathsf{c}_r)$

$(\tilde{\mathsf{q}}, \tilde{\mathsf{p}}) \leftarrow$ `FF_Integrator`$(\mathsf{q}^t, \mathsf{p}^t, h)$
$\mathbb{A} \leftarrow$ `Get_Active_Set`$(\tilde{\mathsf{q}})$
$\lambda_1 \leftarrow$ `Elastic_Impact`$(\mathsf{q}^t, \mathsf{p}^t, \mathbb{A})$
$\lambda_0 \leftarrow$ `Inelastic_Impact`$(\mathsf{q}^t, \mathsf{p}^t, \mathbb{A})$
$\lambda \leftarrow \mathsf{c}_r\, \lambda_1 + (1 - \mathsf{c}_r)\, \lambda_0$
$\beta \leftarrow$ `Friction`$(\mathsf{q}^t, \mathsf{p}^t, \mathbb{A}, \lambda)$
$\mathsf{p}^+ \leftarrow \mathsf{p}^t + \mathsf{G}_{\mathbb{A}}(\mathsf{q}^t)\lambda + \mathsf{D}_{\mathbb{A}}(\mathsf{q}^t, \mathsf{p}^t)\beta$
$(\mathsf{q}^{t+1}, \mathsf{p}^{t+1}) \leftarrow$ `FF_Integrator`$(\mathsf{q}^t, \mathsf{p}^+, h)$
return $(\mathsf{q}^{t+1}, \mathsf{p}^{t+1})$

9 Case Study: 2D Billiard Balls

As a didactic starting point we consider frictionless billiard balls (two translational DoFs) in two dimensions under linear potentials. We employ Verlet [Hairer et al. 2002a] time integration, which for this physical system exactly conserves both momentum and energy. We consider friction in section 10. The ball-ball non-overlap constraint between 2D balls is easily formulated analytically. We compare Jacobi, Gauss-Seidel, Generalized Reflections, and LCP for elastic impact problems.

Symmetry in multi-impacts During a clean pool break, racked balls disperse across the pool table in a nearly symmetric pattern. We consider the "cleanest" possible pool-break, i.e., the outcome of an impact subject to *perfectly* symmetric initial conditions. The initial setup for this simulation is an exactly aligned rack of 55 balls with an additional cue ball fired along the axis of symmetry. In Fig. 5 we compare the results after ten seconds of simulation time between GR and a Gauss-Seidel ordering.

Breaking contact in multi-impacts Consider a uniform grid of balls colliding with a circular boundary. Unless the impact is perfectly inelastic, the grid is expected to scatter with a fountain-like

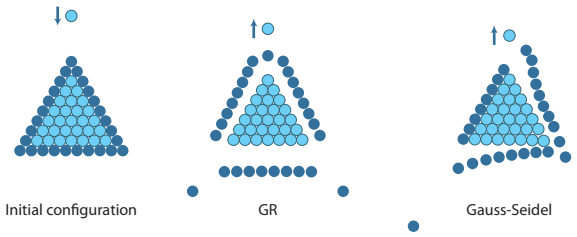


Figure 5: ***The "cleanest" pool break:*** *A perfectly symmetric, elastic pool break is simulated for 56 balls, starting from symmetric initial conditions. After ten seconds of simulation time GR retains the initial symmetries while Gauss-Seidel orderings all generate unexpected, order-dependent results (the right-hand figure shows one representative Gauss-Seidel simulation).*

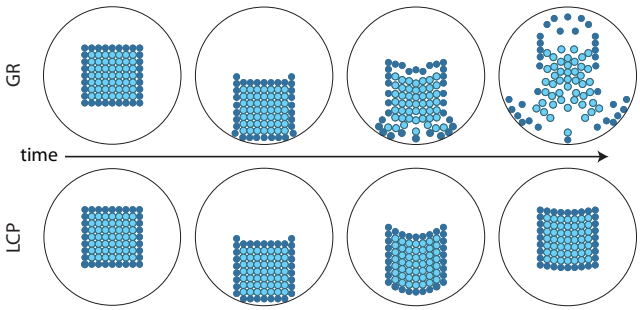
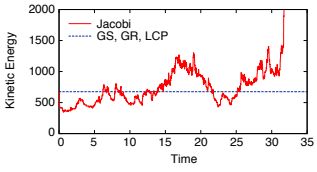


Figure 6: ***Sticky impacts:*** *As in the 1D case, LCP's inability to resolve breaking contact introduces noticeable sticking artifacts in simulation (bottom). Compare to the breaking splash generated by GR (top).*

trajectory (see Fig. 6, top). Resolution of breaking contact is essential to produce this expected behavior. Indeed, simulating this system with LCP, and thus losing breaking contact, produces the unexpected behavior of the entire mass of balls bouncing back up, cohering together as a uniform, sticky material (see Fig. 6, bottom).

Long-term kinetic energy conservation In the plot to the right we confirm the *exact* kinetic energy conservation of our collision integrator using LCP, Gauss-Seidel, and GR as impact operators and note the characteristic poor energy behavior of Jacobi. Here we plot the energy of a 9x9 grid of balls, constrained in a drum, given initial random velocities, and stepped at $h = 10^{-2}$ in zero gravity.



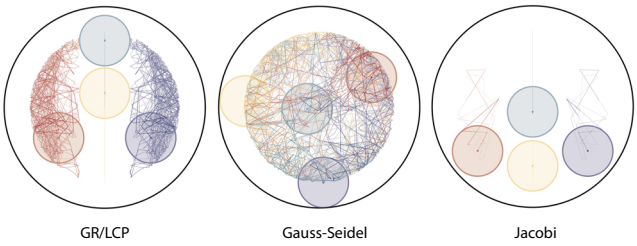


Figure 7: ***Long-term symmetry****: Long-term trace of particle trajectories for GR/LCP, Gauss-Seidel and Jacobi. GR, LCP, and Jacobi remain symmetric, while Gauss-Seidel breaks symmetry. Note that LCP does not suffer from 'sticking' for this example, and so its solution does not differ from GR's. Jacobi's trajectory quickly damps out due to parasitic dissipation.*

Long-term symmetry conservation We place four balls in a circular enclosure with initial state symmetric about the vertical bisector, and plot trajectories generated with the GR, LCP, Jacobi, and Gauss-Seidel impact operators in the collision integrator (see Fig. 7). In this example LCP does not exhibit sticking, therefore the trajectories of GR and LCP coincide. Observe GR/LCP's long-term symmetry preservation. The GS trajectory breaks symmetry upon the first impact. Jacobi's trajectory preserves symmetry, but rapidly dissipates to a crawl, evidenced by the short trace length.

10 Case Study: 3D Rigid Bodies

We now shift our focus to the simulation of rigid bodies with friction in three dimensions. For free-flight integration in SE(3) with a linear gravitational potential we apply the Discrete Moser-Veselov (DMV) integrator [Moser and Veselov 1991], which is exactly energy, momentum, and angular-momentum conserving for this special case. In our implementation we employ a freely available implementation of DMV [Hairer and Vilmart 2006]. Constraints are obtained for boxes using the standard box-box routine [Smith 2006], while constraints for potentially non-convex mesh-mesh impacts are obtained with the signed distance field based implementation in the freely available OpenTissue package [Erleben and Dohlmann 2007]. As an acceleration broad-phase culling is performed using a uniform spatial partition [Ericson 2004].

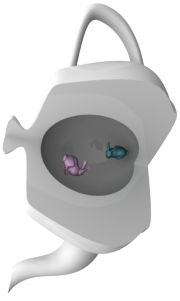
10.1 Long-Term Energy Conservation

When married to a discrete time integration method, the choice of an instantaneous impact technique can lead to a harmonious partnership or a short-lived and explosive affair. In Fig. 8 we plot the long-term total energy of DMV when paired with Jacobi, Gauss-Seidel, and Generalized Reflections for three select simulations. Jacobi's failure to respect (KIN) quickly destabilizes the entire simulation, leading to a catastrophic failure. Gauss-Seidel and Generalized Reflections, in contrast, act in concert with DMV to yield constant Hamiltonians and stable simulations.

10.2 Long-term Symmetry Preservation

Similarly, pairing DMV with the wrong instantaneous impact technique can send a simulation on an asymmetric and meandering trajectory. In Fig. 8 we plot traces of points on rigid bodies for the same three simulations. The trajectories produced by DMV with Jacobi and with Generalized Reflections mirror all spatial symmetries in the initial conditions, while the trajectories due to Gauss-Seidel quickly wander into disorder.

10.3 Long-term Angular Momentum Conservation



To verify conservation of (MOM), we place two bunnies in a simulated teapot with a coefficient of restitution of $\mathsf{c}_r = 0.8$ and a coefficient of friction of $\mu = 0.5$ and initialize the system with zero net linear momentum and nonzero net angular momentum. As this system is closed, we expect both linear and angular momentum to remain constant. Indeed, evolving the system with GR and DMV using a timestep of $h = 0.01$ we observe a maximum drift of 4.37×10^{-11} percent in each component of angular momentum and an absolute drift of 1.36×10^{-11} in each component of the linear momentum for 1000 seconds of simulation.

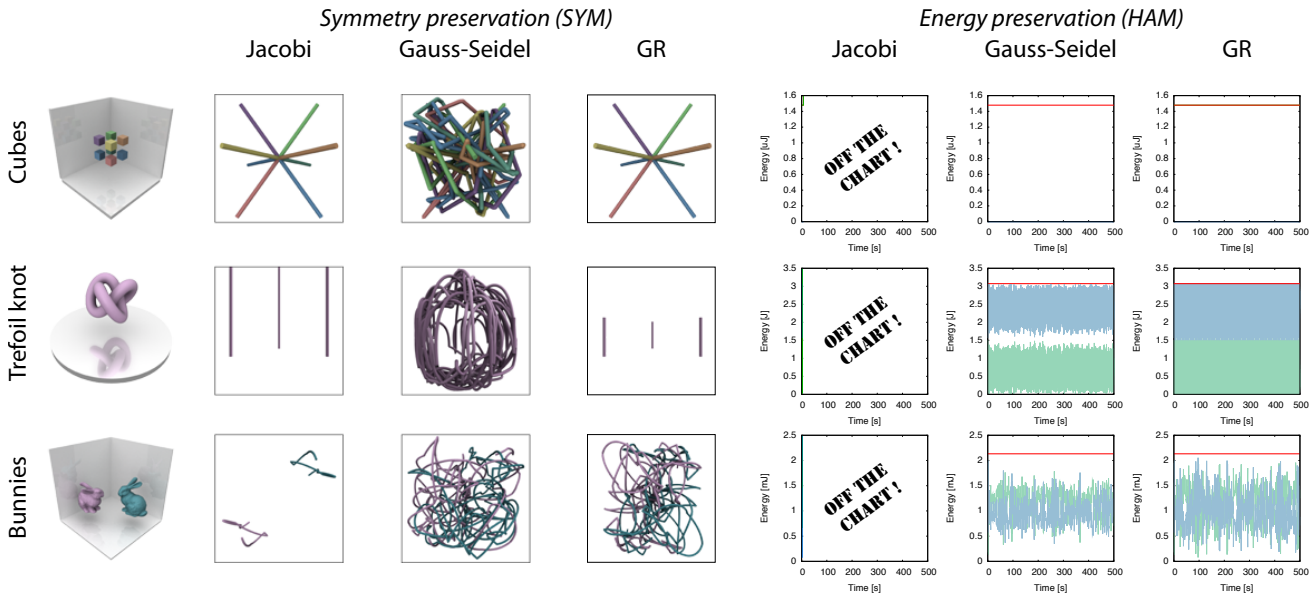


Figure 8: (***HAM***) ***and*** (***SYM***) ***with DMV****: Here we show the results of three simulations.* **Top:** *Eight cubes, initially touching, with symmetric outward velocities constrained to lie in a large cube. There is no gravity in this simulation.* **Middle:** *A trefoil knot with threefold rotational symmetry is dropped on a plane under gravity.* **Bottom:** *Two bunnies with 180 degree rotational symmetry are released under gravity in a box.* **Left:** *For each simulation, we plot traces of points fixed on each body.* **Right:** *For each simulation, we plot potential (blue), kinetic (green), and total energy (red).* **Long term desiderata:** *Coupled with a time integrator, Jacobi, to the left, respects* (*SYM*) *but drops* (*HAM*) *while Gauss-Seidel, in the middle, conserves* (*HAM*) *but destroys* (*SYM*)*. Only GR, on the right, satisfies both* (*HAM*) *and* (*SYM*) *when paired with DMV. To avoid ambiguity in selecting a vertical scale to display the Hamiltonian with, we employ the natural scale defined by the (always positive) kinetic energy.*

10.4 Controlled Dissipation

Our generalized restitution model, when paired with a discrete time integration method, yields a predictable and controlled rate of dissipation. To demonstrate this, we simulate 1,000 bunnies in a fixed container with gravity using Generalized Restitution and DMV. As we decrease the coefficient of restitution, we observe an attendant controlled decrease in the long term energy dissipation (Fig. 9).

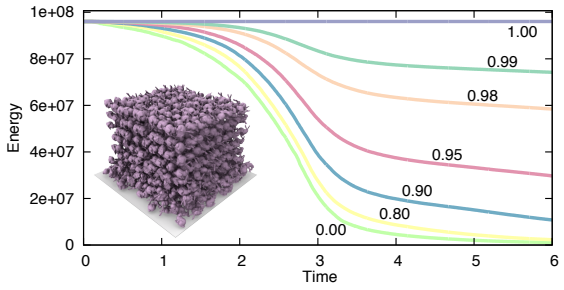


Figure 9: ***Controlled Dissipation****: As we decrease* c_r*, the overall dissipation decreases in a controlled manner.*

10.5 Break-Away and Restitution

Neglecting (BRK) can produce unexpectedly uneventful simulations, even for systems subject to restitution effects. Dropping a set of matryoshka dolls with $c_r = 0.8$, the LCP model causes the dolls to rebound as a single, solid object (Fig. 2, left doll). With the Generalized Restitution model, in contrast, the dolls separate upon first impact, revealing the interior matryoshkas (Fig. 2, right doll).

11 Scaling and Experimental Benchmarks

We now evaluate the scaling properties of our method in 2D and 3D and benchmark against classical problems in granular media.

11.1 Related Methods in Granular Simulation

While our proposed algorithms enable the simulation of arbitrary rigid body models, many of the following examples particularly focus on granular systems. Such examples enable us to consider correctness, efficiency, scaling, and emergent behavior in assemblies (treated here as collections of rigid bodies) while allowing us to validate large-scale simulations against experimental observations.

The simulation of granular materials has been broadly treated in engineering and mechanics. Pöschel and Schwager [2005] provide a comprehensive survey of current methods in the literature.

Within graphics research on granular simulation has largely focused upon continuum-based models [Zhu and Bridson 2005; Lenaerts and Dutré 2009; Narain et al. 2010; Alduán and Otaduy 2011] that efficiently and convincingly capture the complex behaviors of granular flow. However, as noted by Narain et al. [2010], these methods are unable to resolve the collisional behaviors of granular systems.

Alternately, a range of damped-spring-based interaction methods [Miller and Pearce 1989; Luciani et al. 1995; Bell et al. 2005; Alduán et al. 2009], originating in the discrete element and molecular dynamic models of mechanics, have also been considered. While effective for resolving slower contact modes, under impact these methods must often deal with many of the same stability issues commonly encountered in comparable, and more familiar, penalty-based methods [Pöschel and Schwager 2005].

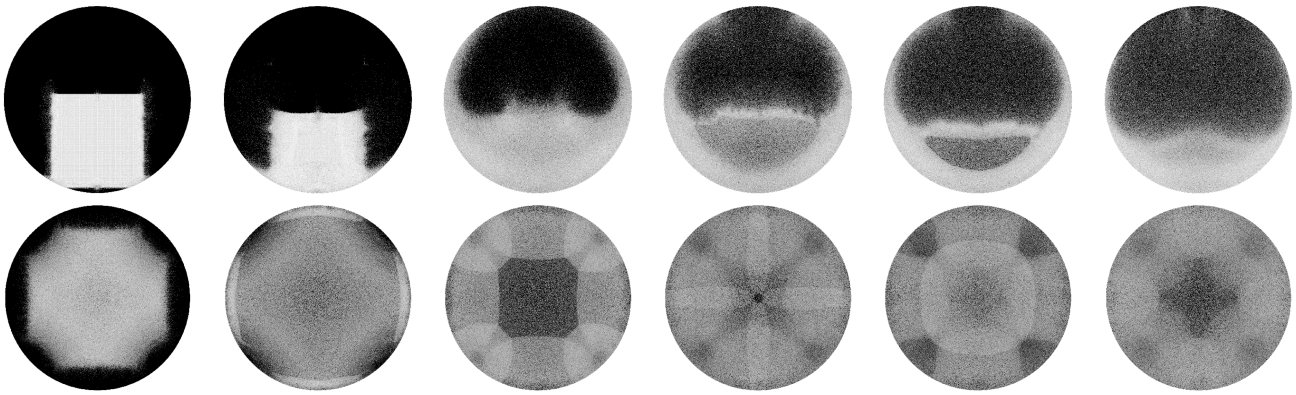


Figure 10: ***Scaling of our system****: Here we show two sequences of configurations from large-scale, 1M-ball simulations in 2D. Impacts are frictionless and elastic (i.e.,* $\mathsf{c}_r = 1$*).* **Top (Drop)***: Spheres are initialized to a uniform square grid configuration, dropped from rest, under a gravitational potential of magnitude* 10*.* **Bottom (Gas):** *Spheres are initialized to a uniform square grid configuration. Each sphere is assigned an initial, random, unit length velocity. No external forces are exerted on the system.*

Simulation	DoFs	Integration	Detection			Impact			GR Iterations			Overall Constraints		
		avg	avg	max	min	avg	max	min	avg	max	min	avg	max	min
Drop 10K	20,402	0.0005	0.01	0.03	0.01	0.01	3.08	0.00	1	291	0	6,914	20,238	0
Drop 100K	200,978	0.0066	0.23	0.33	0.19	1.30	279.90	0.00	7	1,629	0	85,381	200,346	10,253
Drop 1M	2,004,002	0.0930	3.08	4.33	2.56	137.62	7,434.73	0.04	76	11,181	0	1,242,689	2,271,061	373,610
Random 10K	20,402	0.0005	0.01	0.07	0.01	0.00	0.09	0.00	1	5	1	147	9,720	0
Random 100K	200,978	0.0060	0.20	0.27	0.15	0.05	2.04	0.01	1	5	1	4,452	96,694	50
Random 1M	2,004,002	0.0854	2.61	3.19	2.32	1.67	22.76	0.10	2	6	1	135,417	968,392	6,933

Table 2: ***Timings****: Performance statistics for balls constrained to a drum. For each simulation we report the wall-clock time (seconds) spent in unconstrained integration, collision detection, and impact response per timestep. We report the min, max, and average time across all timesteps. We also report the number of iterations spent in each GR solve, as well as the overall active set size. The system was integrated with a fixed timestep of* 0.01s *and employed the nonlinear optimization package Ipopt to solve the impact QP. Timings were recorded with a single thread on a 3.33GHz Intel Core i7-975.*

11.2 Scaling

Formulating our multi-impact problem as a QP and solving it in a sparsity preserving manner enables our method to scale to large problems composed of many DoFs in impact with one another. We instrumented two example configurations, consisting of increasingly larger numbers of frictionless spheres, all with unit radius, subject to $\mathsf{c}_r = 1$, and constrained to stay inside a circular drum:

Drop examples are initialized to a uniform square grid configuration and dropped from rest, under gravity.

Gas examples are likewise initialized to a uniform square grid configuration. Each sphere is assigned an initial, random, unit length velocity. No external forces are exerted on the system.

In both examples, note the preservation of large scale symmetries. Timings and statistics for these examples are reported in Table 2. A video of these simulations can be found in the supplementary materials while configuration snapshots are shown in Fig. 10. Note that the dominant cost, the solution of the impact QP, scales with the number of constraints.

11.3 Granular Maxwell's Demon

Dissipative impact plays an intriguing role in spontaneous symmetry breaking. An interesting example is the so-called Granular Maxwell's Demon. In analogy to Maxwell's Demon, two identical grids of balls are placed in a box separated from one another by a partial-height partition. Driving the box's floor at constant frequency and amplitude all balls eventually "choose" one of the two box partitions *without* the assistance of a demonic agent.

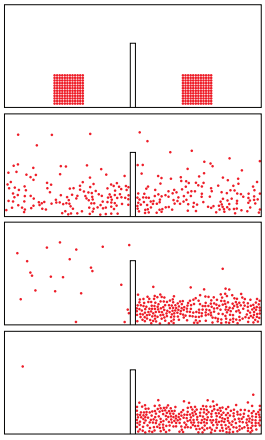


Figure 11: ***Maxwell's Demon*** *demonstrates a dramatic consequence of restitution behavior* (top to bottom)*. Balls placed evenly in two sides of a vibrating box, split by a solid fence, aggregate on one side due to dissipative collisions. The setup here simulates 288 unit mass steel balls with* $\mathsf{c}_r = 0.6$*.*

The standard explanation is that multiple dissipative collisions "trap" particles preferentially on the side with more particles [van der Weele et al. 2001]. More particles colliding means more dissipation due to the coefficient of restitution, keeping the kinetic energy of the balls on the more populous side low. Similarly, balls on the sparser side experience a lower collision rate and thus their energy is not damped as quickly. Somewhat paradoxically, balls on the sparser side are thus more likely to cross over to the trapping side than vice-versa; as the trapping side becomes more loaded, this process accelerates via positive feedback.

Duplicating the experimental setup's initial configuration we simulate the same effect using Generalized Reflections. Fig. 11 shows simulation snapshots of the granular Maxwell's Demon phenomenon. With initially symmetric positions and random initial velocities, progressively more balls migrate to one side (the right in this example) as expected. This provides one validation of our proposed definition for the multi-impact coefficient of restitution.

11.4 Extended Patterns

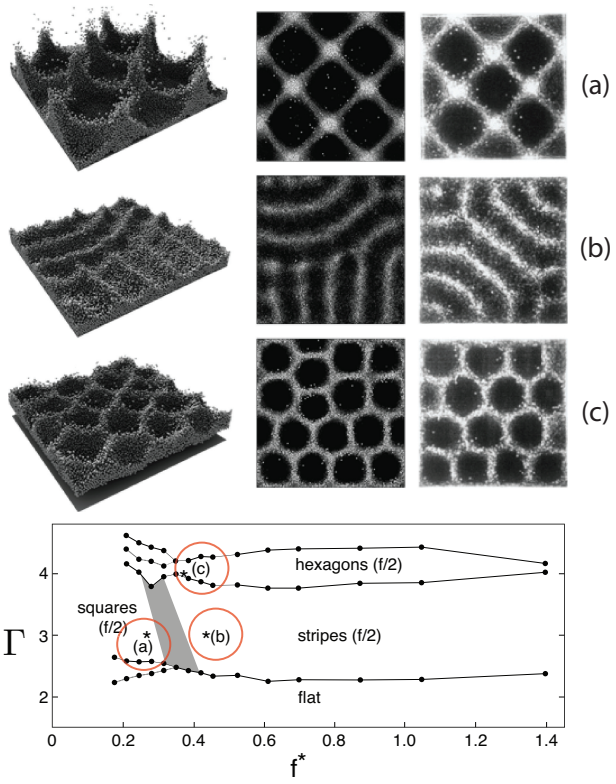


Figure 12: ***Extended Patterns experiment and GR Simulation****: We simulated the experiments of Melo et al. [1994] across a range of parameters.* **Top***: In the left and middle columns we show side and top view simulation snapshots and note a match with the experimentally obtained square, stripe, and hexagonal patterns in the right column.* **Bottom***: We also reproduce the phase diagram from Bizon et al. [1998] showing the points that generated both the experimental results and the matching simulations. The phase diagram is Copyright 1998 by The American Physical Society.*

Molecular and crystalline assemblies play a critical role in systems of interest in biology, chemistry and physics but are often difficult and costly to study. To understand these phenomena scientists have turned to simpler systems such as vibrated bins of granules [Umbanhowar et al. 1996]. In this setting spontaneous, stable patterns are observed [Melo et al. 1994]. The hope is that these systems can be studied in analogy to crystalline and molecular structures and thus elucidate otherwise complex behaviors.

The so-called extended patterns are emergent behaviors known to be driven solely by floor vibrations, high-speed inelastic impacts between granules, and corresponding frictional forces [Melo et al. 1994]. As such they are ideal computational benchmarks to examine both validation and efficiency. The challenges to simulate them are directly in line with our goals: the generation of extended patterns requires the accurate resolution of very-large collections of rigid-bodies under high speed impact (and are thus not amenable to continuum models), they require the accurate, scalable modeling of multi-impact (simultaneous collisions are generic), restitution (inelasticity is required), and friction (without friction the patterns do not emerge [Moon et al. 2004]) at large scales.

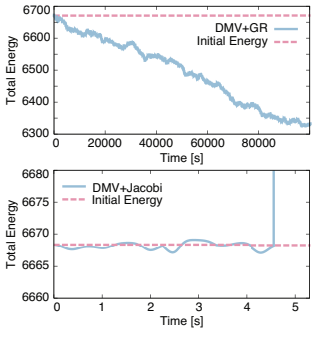
We first confirm that our algorithm captures the stable stripe, square and hexagonal patterns experimentally observed in stability regions of a phase diagram plotting driving amplitude vs. frequency. Duplicating the reported [Bizon et al. 1998] material restitution and friction, packing ratios, amplitudes, and frequencies we obtain a one-to-one correspondence with experimental results [Bizon et al. 1998] for all three patterns: hexagons at $f^* = 0.38$ and $\Gamma = 4.00$, squares at $f^* = 0.27$ and $\Gamma = 3.00$, and stripes at $f^* = 0.44$ and $\Gamma = 3.00$ (see Fig. 12). These parameters correspond to the non-dimensionalized frequency and amplitude of the driving plate, which are employed to reduce the dimensionality of the problem search space. Note, as well, as seen in our supplemental video, that these extended patterns emerge, as expected, dynamically over time and remain stable throughout the remainder of all simulation runs in correspondence with the experimentally observed stability.

Next we note that these patterns are simulated "out of the box" in a parameter-free manner with no tuning. However, since we employ a synchronous framework for simulation (see §8.2) in these examples, there is a clear trade-off between the accuracy of our results, e.g., how well we capture the extended patterns, and the size of the time-step employed, i.e., the efficiency of the method. In Table 3 we consider these trade-offs in the simulation of the hexagonal pattern holding all experimental constants fixed and varying only timestep size in increments down from $h = 10^{-3}$ (above which only a flat "noisy" pattern is obtained) decrementing down to $h = 10^{-4}$. At $h = 10^{-3}$ the expected hexagonal pattern emerges; however, the wavelength is smaller than reported. As we decrease timestep the wavelength of the generated hexagonal pattern correspondingly increases until, starting at stepsizes of $h = 3 \times 10^{-4}$ and below, simulations converge to both the correct pattern and wavelength.

12 Limitations and Future Work

Focusing on impact, we have not addressed a number of questions endemic to contact, chiefly the problem of drift-free satisfaction of sustained contact constraints. Drift-free trajectories are typically obtained with a shock propagation or position-based stabilization technique [Guendelman et al. 2003; Erleben 2007], sacrificing (HAM) and/or (ONE). Can one avoid this sacrifice?

While the iterative GR model terminates for all attempted examples, the termination of GR, and, to our knowledge, even that of the Gauss-Seidel pairwise model, remains an open question. Likewise, appropriately extending the generalized restitution model to multiple coefficients of restitution is an interesting problem.



Alg. 9 obtains (HAM) by interleaving an energy-preserving impact operator and a synchronous energy-preserving time integrator. Unfortunately it is known [Zhong and Marsden 1988] that the class of exact energy-momentum integrators is limited, and that with fixed-size time steps these integrators do not, in general, conserve all of a system's physical invariants. Thus, we are also interested in extending our (HAM) guarantees to a wider class of simulations.

Timestep	Integration	Detection			Impact			Friction			GR Iterations			Overall Constraints		
	avg	avg	max	min	avg	max	min	avg	max	min	avg	max	min	avg	max	min
0.00100	0.1025	0.48	0.55	0.30	38.22	404.99	0.00	11.33	326.09	0.00	14	32	1	89,126	181,261	141
0.00050	0.1003	0.46	0.54	0.30	26.39	446.20	0.00	12.63	334.63	0.00	15	42	1	61,469	168,637	2
0.00045	0.1005	0.41	0.48	0.25	23.66	388.16	0.00	11.80	348.47	0.00	15	39	1	57,355	165,425	79
0.00035	0.0996	0.40	0.47	0.25	13.97	331.79	0.00	7.63	311.10	0.00	14	45	1	44,773	157,099	14
0.00030	0.1000	0.45	0.52	0.29	9.52	271.89	0.00	5.31	262.58	0.00	13	43	1	38,177	151,024	33
0.00020	0.0989	0.45	0.54	0.29	3.88	147.24	0.00	2.22	157.54	0.00	12	39	1	27,521	135,653	14
0.00010	0.0982	0.41	0.48	0.26	1.48	64.45	0.00	0.81	54.15	0.00	11	28	1	18,491	109,525	2

Table 3: ***Hexagonal Extended Pattern Statistics****: Performance for decreasing timestep. Timings (reported above in seconds) were recorded with a single thread on a 2.67GHz Intel Xeon 5650 for a system of 360,000 DoFs.*

For example, when we apply GR-DMV (Alg. 9) to a rigid body under a *nonlinear* potential, we observe a long-term energetic walk (adjacent, top). Compare to the behavior of Jacobi (adjacent, bottom), noting the different plot-axis scales. While Jacobi immediately blows up, GR-DMV has a relatively mild energy drift instead of exact conservation. This is because DMV exactly preserves energy for linear, but not for *non*linear, potentials.

DMV belongs to the class of symplectic integrators [Hairer et al. 2002b], which strike a compromise: for general systems (e.g., nonlinear potentials), they *approximately* conserve all physical invariants, including energy. The measured energy using such an integrator oscillates about the true constant value, but the amplitude is guaranteed to be bounded for an amount of time exponential in the time step. The interaction of a symplectic integrator with (KIN)-obeying impact response is more subtle—the resulting time integrator, which can be viewed as the composition of a symplectic-momentum and an energy-momentum integrator, is typically neither.

Can one gracefully combine impact response with symplectic integration to yield a new integrator that is still symplectic? Is there an alternative desideratum to (KIN) that trades instantaneous energy conservation for long-term, approximate energy conservation?

Finally, we are interested in the long-term preservation of (SYM) under floating point arithmetic (FPA). During a simulation FPA errors can accumulate and eventually break symmetry (see, e.g., the top row, fourth image of Fig. 10). Even so, we observe that operators like GR that avoid the *systematic* destruction of symmetry are better able to capture approximate symmetries over long run times.

Acknowledgments

We thank Fang Da for his assistance with 2D rendering and Stephen Morris for his valuable discussions on granular media. This research is supported in part by the Sloan Foundation, the NSF (CAREER Award CCF-06-43268 and grants IIS-09-16129, IIS-10-48948, IIS-11-17257, CMMI-11-29917), and generous gifts from Adobe, Autodesk, Intel, mental images, NVIDIA, Side Effects Software, and The Walt Disney Company.

References

ALDUÁN, I., AND OTADUY, M. A. 2011. SPH granular flow with friction and cohesion. In *Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, ACM, New York, NY, USA, SCA '11, 25–32.

ALDUÁN, I., TENA, A., AND OTADUY, M. A. 2009. Simulation of High-Resolution Granular Media. In *Proc. of Congreso Español de Informática Gráfica*.

AMESTOY, P. R., DUFF, I. S., KOSTER, J., AND L'EXCELLENT, J.-Y. 2001. A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM Journal on Matrix Analysis and Applications 23*, 1, 15–41.

AMESTOY, P. R., GUERMOUCHE, A., L'EXCELLENT, J.-Y., AND PRALET, S. 2006. Hybrid scheduling for the parallel solution of linear systems. *Parallel Computing 32*, 2, 136–156.

ANITESCU, M., AND POTRA, F. R. 1997. Formulating Dynamic Multirigid-Body Contact Problems with Friction as Solvable Linear Complementarity Problems. *ASME Nonlinear Dynamics 14*, 231–247.

BARAFF, D. 1989. Analytical methods for dynamic simulation of non-penetrating rigid bodies. In *Computer Graphics (SIGGRAPH 89)*, 223–232.

BELL, N., YU, Y., AND MUCHA, P. J. 2005. Particle-based simulation of granular materials. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, ACM, New York, NY, USA, SCA '05, 77–86.

BERNOULLI, J. 1742. Op. CLXXVII, Propositiones variæ Mechanico-dynamicæ. In *Opera Omnia*. 253–313.

BERNU, B., AND MAZIGHI, R. 1990. One-Dimensional Bounce of Inelastically Colliding Marbles on a Wall. *Journal of Physics A: Mathematical and General 23*, 24, 5745–5754.

BIZON, C., SHATTUCK, M. D., SWIFT, J. B., MCCORMICK, W. D., AND SWINNEY, H. L. 1998. Patterns in 3d vertically oscillated granular layers: Simulation and experiment. *Phys. Rev. Lett. 80*, 1 (Jan), 57–60.

BOYD, S., AND VANDENBERGHE, L. 2004. *Convex Optimization*. Cambridge University Press.

BRIDSON, R., FEDKIW, R. P., AND ANDERSON, J. 2002. Robust Treatment of Collisions, Contact, and Friction for Cloth Animation. *ACM Trans. Graph. (SIGGRAPH 02) 21*, 3 (July), 594–603.

BROGLIATO, B. 1999. *Nonsmooth Mechanics: models, dynamics, and control*, 2nd ed. Springer-Verlag.

CHATTERJEE, A., AND RUINA, A. L. 1998. A New Algebraic Rigid-Body Collision Law Based on Impulse Space Considerations. *Journal of Applied Mechanics 65*, 4, 939–951.

COTTLE, R. W., PANG, J. S., AND STONE, R. E. 1992. *The Linear Complementarity Problem*. Academic Press.

D'ALEMBERT, J. 1743. *Traite de Dynamique*.

ERICSON, C. 2004. *Real-Time Collision Detection*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

ERLEBEN, K., AND DOHLMANN, H. 2007. Signed Distance Fields Using Single-Pass GPU Scan Conversion of Tetrahedra. In *GPU Gems 3*, 741–762.

ERLEBEN, K. 2007. Velocity-based shock propagation for multibody dynamics animation. *ACM Trans. Graph. 26*, 2.

FRIEDLANDER, M. P., 2007. BCLS: Bound Constrained Least Squares.

GLOCKER, C. 2004. Concepts for Modeling Impacts without Friction. *Acta Mechanica 168*, 1–19.

GOLDFARB, D., AND IDNANI, G. 1983. A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical Programming 27*, 1–33.

GUENDELMAN, E., BRIDSON, R., AND FEDKIW, R. 2003. Nonconvex Rigid Bodies with Stacking. *ACM Trans. Graph. (SIGGRAPH 03) 22*, 3, 871–878.

HAHN, J. K. 1988. Realistic animation of rigid bodies. In *Computer Graphics (SIGGRAPH 88)*, 299–308.

HAIRER, E., AND VILMART, G. 2006. Preprocessed discrete Moser–Veselov algorithm for the full dynamics of a rigid body. *Journal of Physics A: Mathematical and General 39*, 42, 13225.

HAIRER, E., LUBICH, C., AND WANNER, G. 2002. *Geometric numerical integration: Structure-Preserving Algorithms for Odinary Differential Equations*. Springer.

HAIRER, E., LUBICH, C., AND WANNER, G. 2002. *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer-Verlag.

HARMON, D., VOUGA, E., TAMSTORF, R., AND GRINSPUN, E. 2008. Robust Treatment of Simultaneous Collisions. *SIGGRAPH 08, ACM TOG*.

HASCOËT, E., HERRMANN, H. J., AND LORETO, V. 1999. Shock Propagation in a Granular Chain. *Phys. Rev. E 59*.

HSL. 2001. A collection of Fortran codes for large scale scientific computation. *http://www.hsl.rl.ac.uk*.

IVANOV, A. P. 1995. On Multiple Impact. *Journal Applied Mathematics and Mechanics 59*, 6, 887–902.

JOHNSON, W. 1976. Simple Linear Impact. *Int. J. Mech. Eng. Educ. 4*, 167–181.

KAUFMAN, D. M., EDMUNDS, T., AND PAI, D. K. 2005. Fast frictional dynamics for rigid bodies. *ACM TOG (SIGGRAPH 05) 24*, 3, 946–956.

KAUFMAN, D. M., SUEDA, S., JAMES, D. L., AND PAI, D. K. 2008. Staggered Projections for Frictional Contact in Multibody Systems. *ACM TOG (SIGGRAPH Asia 08) 27*, 5, 1–11.

LAWSON, C. L., AND HANSON, R. J. 1974. *Solving least squares problems*. Prentice-Hall.

LENAERTS, T., AND DUTRÉ, P. 2009. Mixing fluids and granular materials. *Computer Graphics Forum 28*, 2, 213–218.

LUBACHEVSKY, B. 1991. How to Simulate Billiards and Similar Systems. *Journal of Computational Physics 94*, 255–283.

LUCIANI, A., HABIBI, A., AND MANZOTTI, E. 1995. A multiscale physical model of granular materials. In *Graph. Interf.*

MACLAURIN, C. 1742. *A Treatise on Fluxions*.

MCNAMARA, S., AND YOUNG, W. R. 1994. Inelastic collapse in two dimensions. *Phys. Rev. E 50*, 1 (Jul), R28–R31.

MELO, F., UMBANHOWAR, P., AND SWINNEY, H. L. 1994. Transition to parametric wave patterns in a vertically oscillated granular layer. *Phys. Rev. Lett. 72*, 1 (Jan), 172–175.

MILLER, G., AND PEARCE, A. 1989. Globular dynamics: A connected particle system for animating viscous fluids. *Computers and Graphics 13*, 3, 305 – 309.

MIRTICH, B., AND CANNY, J. F. 1995. Impulse-based dynamic simulation of rigid bodies. In *Symp. on Inter. 3D Graph.*

MOON, S. J., SWIFT, J. B., AND SWINNEY, H. L. 2004. Role of friction in pattern formation in oscillated granular layers. *Phys. Rev. E 69*, 3 (Mar), 031301.

MOREAU, J. J. 1983. *Unilateral Problems in Structural Analysis. International Centre for Mechanical Sciences, Courses and Lectures - No. 288*. ch. Standard Inelastic Shocks and the Dynamics of Unilateral Constraints., 173–221.

MOREAU, J. J. 1988. Unilateral Contact and Dry Friction in Finite Freedom Dynamics. *Nonsmooth Mechanics and Applications, CISM Courses and Lectures*, 302, 1–82.

MOSER, AND VESELOV. 1991. Discrete Versions of Some Classical Integrable Systems and Factorization of Matrix Polynomials. *Communications in Mathematical Physics 139*, 2, 217–243.

NARAIN, R., GOLAS, A., AND LIN, M. C. 2010. Free-Flowing Granular Materials with Two-Way Solid Coupling. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2010)*.

PÖSCHEL, T., AND SCHWAGER, T. 2005. *Computational granular dynamics: models and algorithms*. Springer-Verlag.

PUDASAINI, S. P., AND KRÖNER, C. 2008. Shock waves in rapid flows of dense granular materials: Theoretical predictions and experimental results. *Phys. Rev. E 78* (Oct), 041308.

SCHITTKOWSKI, K. 2005. QL: A Fortran code for convex quadratic programming - User's guide, Version 2.11. *Report, Department of Mathematics, University of Bayreuth*.

SMITH, R. 2006. Open Dynamics Engine, V0.5, User Guide.

STEWART, D. E. 2000. Rigid-Body Dynamics with Friction and Impact. *SIAM Rev. 42*, 1, 3–39.

UMBANHOWAR, P. B., MELO, F., AND SWINNEY, H. L. 1996. Localized excitations in a vertically vibrated granular layer. *Nature 382* (8/1996), 793 – 796.

VAN DER WEELE, K., VAN DER MEER, D., VERSLUIS, M., AND LOHSE, D. 2001. Hysteretic clustering in granular gas. *EPL (Europhysics Letters) 53*, 3, 328.

WÄCHTER, A., AND BIEGLER, L. T. 2006. On the Implementation of an Interior-Point Filter Line-Search Algorithm for Large-Scale Nonlinear Programming. *Mathematical Programming 106*, 25–57.

WITKIN, A., AND BARAFF, D. 2001. Physically Based Modeling. In *SIGGRAPH 2001 COURSE NOTES*.

ZHONG, G., AND MARSDEN, J. E. 1988. Lie-Poisson Hamilton-Jacobi theory and Lie-Poisson integrators. *Physics Letters A 133*, 3 (Nov), 134–139.

ZHU, Y., AND BRIDSON, R. 2005. Animating sand as a fluid. *ACM Trans. Graph. (SIGGRAPH 05) 24* (July), 965–972.