Multiple impacts: A state transition diagram approach

Yan-Bin Jia, Matthew T Mason and Michael A Erdmann

The International Journal of Robotics Research 2013 32: 84 originally published online 1 November 2012
DOI: 10.1177/0278364912461539

The International Journal of
Robotics Research
32(1) 84–114

DOI: 10.1177/0278364912461539
ijr.sagepub.com

Multiple impacts: A state transition diagram approach

Yan-Bin Jia[1], Matthew T Mason[2] and Michael A Erdmann[2]

Abstract

Impact happens when two or more bodies collide, generating very large impulsive forces in a very short period of time during which kinetic energy is first absorbed and then released after some loss. This paper introduces a state transition diagram to model a frictionless multibody collision. Each state describes a different topology of the collision characterized by the set of instantaneously active contacts. A change of state happens when a contact disappears at the end of restitution, or when a disappeared contact reappears as the relative motion of two bodies goes from separation into penetration. Within a state, (normal) impulses are coupled differentially subject to relative stiffnesses at the active contact points and the strain energies stored there. Such coupling may cause restart of compression from restitution during a single impact. Impulses grow along a bounded curve with first-order continuity, and converge during the state transitions. To solve a multibody collision problem with friction and tangential compliance, the above impact model is integrated with a compliant impact model. The paper compares model predictions to a physical experiment for the massé shot, which is a difficult trick in billiards, with a good result.

Keywords

impact, multiple impacts, energy-based restitution, state transition diagram, stiffness ratio, impulse curve, compliance, billiard shooting

1. Introduction

Impact is a phenomenon that takes place and is made use of in our daily life. It occurs when two or more bodies collide, establishing a short period of contact. During this period, impact generates an impulsive contact force with magnitude tens or hundreds of times higher than that of a static contact force. We apply impulsive forces to accomplish tasks that would be inefficient or difficult to achieve otherwise. Examples include landing after a jump, putting a cup of coffee on the table, punching a staple through a thick stack of documents, kicking a soccer ball, shooting a billiard, a bullet hitting a target, etc. On other occasions we want to extend impact periods to minimize impulsive forces that could do harm to ourselves. Devices designed for such purpose include air bags in an automobile, baseball and boxing gloves, football helmets, gymnastic mats, etc.

Impact has applications in many manufacturing and robotic manipulation tasks that involve high-speed operations. A bowl feeder (Boothroyd and Redford, 1968) exerts vibrations to channel parts along a spiral inclined track designed to sort them in a small number of fixed orientations. Every part will go through multiple collisions with gates along the track. The earliest work on impulsive manipulation can be traced back to Higuchi (1985), who demonstrated linear positioning of tools using electromagnetic impulsive force. Parts with simple geometry floating on an air table can be sorted by hitting them with a rotating stick (Hirai et al., 1999). A part can be struck only once (Han and Park, 2001) or tapped repetitively (Huang and Mason, 2000) to reach a desired configuration, by imparting initial velocities calculated through analyses of impacts and the part's post-impact sliding motions. For space robots, collisions are inevitable either with other free-floating objects (Yoshida and Nenchev, 1995) or from landing on other planets, either inside spacecraft or by themselves.

Meanwhile, work has been carried out on making a robot with low degrees of freedom juggle an object via impact. A common objective here is to turn the object's original

[1]Department of Computer Science, Iowa State University, Ames, IA, USA
[2]School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA

Corresponding author:
Yan-Bin Jia, Department of Computer Science, Iowa State University, 226 Atanasoff Hall, Ames, IA 50010, USA.
Email: jia@iastate.edu

free flying motion into a vertical periodic one. Rizzi and Koditschek (1992) implemented a robotic paddle that was capable of batting a ping-pong ball into a steady periodic motion via controlled impacts. Zavalo-Rio and Brogliato (1999) studied the feedback control of a one-degree-of-freedom juggling robot by treating it as a mechanical system under some unilateral constraint. Their control design was later extended by Brogliato and Zavalo-Rio (2000) to a class of complementary-slackness juggling systems through integration of such constraint with an impact law. Lynch and Black (2001) used the idea of force recurrence to demonstrate that a planar juggler, under controls found via nonlinear optimization, could converge an initial free flight trajectory of a puck to a vertical limit cycle.

Despite accomplishments in the design of a practical walking machine (Raibert, 1986), most existing walking robots are still lab prototypes that perform statically stable gaits at low and constant speeds to avoid inertial effects. Understanding the foot–ground collision can be important for resolving instability due to large inertial forces generated by fast walking and running (García and de Santos, 2005). Research efforts on robot dynamics with impact include modeling of collisions between a robot and its environment (Zheng and Hemami, 1985), design of control schemes for stable contact during such collisions (Volpe and Khosla, 1993), evaluation of collision effects on a robot (Walker, 1994), and generation of a large impulsive force by a humanoid robot without losing balance (Konno et al., 2011).

Impact also has a large potential application in sports robotics. An air-hockey-playing robot (Partridge and Spong, 1999) applies impact to control the trajectory of a puck sliding on an air table with low friction. Skills in ball sports are essentially about understanding how to strike balls to make them follow desired trajectories while moving and spinning fast enough to evade the opponents. A banana kick in soccer, performed by striking the bottom left or right side of the ball, results in a curved trajectory in the air along which the ball could travel past an array of defenders, the goalie, or a post into a net. This is a perfect example of impact with tangential compliance. In pocket billiards (Shamos, 1993), a shot often results in two concurrent impacts: one between the cue stick and the cue ball and the other between the ball and the table. Tangential compliance between the cue tip and the cue ball is critical for generating desired velocity and spinning of the cue ball with two objectives: to pocket an object ball via collision, and to make the cue ball go to a location where the next shot can continue comfortably.

Multiple impacts occur when several bodies collide at more than one contact, or more common in reality, when one body hits another body that is at rest and in contact with several other motionless bodies. Multiple impacts are unavoidable in the simulation of dynamical systems that involve rigid-body collisions. They are also common in sports actions such as a billiard shot mentioned earlier (and also a break shot) and a bowling throw to knock down pins.

An impulsive force has very short execution time, and therefore good potential for improving task efficiency. Its use could considerably simplify the robotic mechanism performing a manipulation task, while avoiding uncertainties accumulated over repeated complex operations otherwise. Despite many potential applications of impact, impulsive manipulation has remained an area in robotics where relatively little work is known,[1] likely because the foundation for modeling rigid-body impact is not fully developed and existing theories often seem either too simple to be realistic or too complex to be applicable, especially in the presence of friction and tangential compliance, not to mention nonlinear viscoelastic effects. Discrepancies often exist between an introduced theory and the findings from an experiment intended for its validation.

Robotic manipulation is often treated as a multibody system with point contacts. Its state is described by generalized coordinates $\boldsymbol{q}$ that gather the positions and orientations of the rigid bodies (including the links of the manipulator) in the system. It is well known that the system dynamics, combined with the contact kinematics, can be described as

$$M(\boldsymbol{q})\ddot{\boldsymbol{q}} + C(\boldsymbol{q},\dot{\boldsymbol{q}}) + G(\boldsymbol{q}) = F(\boldsymbol{q},\dot{\boldsymbol{q}}), \tag{1}$$

where M is the symmetric and positive-definite mass matrix, C gathers all centrifugal and Coriolis terms, G includes gravity terms, and F represents the generalized force, which includes the applied forces, torques, as well as the effects of friction and compliance. Such effects typically depend on the generalized velocity $\dot{\boldsymbol{q}}$.

Suppose that the rigid bodies in the system are simultaneously engaged in a collision. It is reasonable to consider that the collision has a very small duration τ. We integrate the dynamics equation (1) over τ. Since the terms $C(\boldsymbol{q},\dot{\boldsymbol{q}})$ and $G(\boldsymbol{q})$ are bounded, their integrals vanish as $\tau \to 0$. Also, when $\tau \to 0$, the configuration $\boldsymbol{q}$ of the system remains unchanged. Hence, we have

$$M(\boldsymbol{q})\dot{\boldsymbol{q}} = \int_0^{\tau} F(\boldsymbol{q},\dot{\boldsymbol{q}})\, dt = A(\boldsymbol{q})\boldsymbol{I}_n(\boldsymbol{q}) + B(\boldsymbol{q})\boldsymbol{I}_{\perp}(\boldsymbol{q},\dot{\boldsymbol{q}}), \tag{2}$$

where $\boldsymbol{I}_n$ gathers all of the impulses due to the normal contact forces between the rigid bodies, and $\boldsymbol{I}_{\perp}$ gathers all of the impulses due to their tangential contact forces. Typically, the impact configuration $\boldsymbol{q}$ can be computed. This means that $M(\boldsymbol{q})$ is a constant matrix. Also, since the torques are linear in the contact forces, A and B are constant matrices as well.

Solution of the above multibody collision problem requires determination of the post-collision velocities of the rigid bodies from their configuration $\boldsymbol{q}$ and pre-collision velocities $\dot{\boldsymbol{q}}_0$. Essentially, this requires computation of the normal and tangential impulse vectors $\boldsymbol{I}_n$ and $\boldsymbol{I}_{\perp}$. We solve the problem via a decomposition as follows:

(i) At each contact, obtain the tangential impulse from the normal impulse using their differential relationship via an energy-based contact mode analysis.

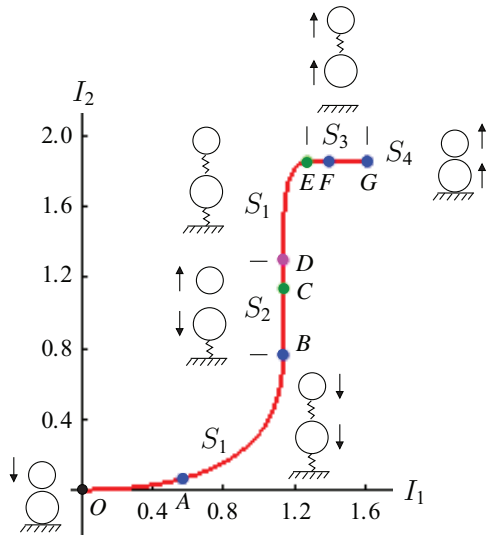


Fig. 1. State sequence $\langle S_1, S_2, S_1, S_3, S_4 \rangle$ and growth of impulses along a curve that results from a collision. Virtual springs represent active contacts.

(ii) Analyze the differential relationships among the normal impulses at different contacts so we can choose one normal impulse at a time as the sole variable for the collision process.

Part (i) is a problem of impact with friction and tangential compliance. It is treated in Jia (2012) via the introduction of a compliant impact model. Part (ii) is a problem of multiple impacts to be investigated in this paper. More specifically, we will study multiple frictionless impacts with the introduction of a state transition model that tracks impulse accumulations at different contacts. Near the end of the paper, we will combine the two impact models to solve the multibody impact problem. With this work we hope to enhance the foundation for impulsive manipulation, especially in a situation where multiple impacts happen at the same time between a robot and the objects it is manipulating.

Objects engaged in impacts simultaneously may break and re-establish contact multiple times. In our model, a state represents a distinct subset of the contact points that are active within an interval of the entire duration.

Figure 1 illustrates a collision that results from a ball with downward velocity striking another ball that is resting on the table. The initial configuration is drawn next to the origin with the centers of the two balls aligned vertically. The collision process is described in terms of accumulations of the impulses I_1 and I_2 at the ball–ball and ball–table contacts, respectively.[2] It is decomposed into a sequence of five states $\langle S_1, S_2, S_1, S_3, S_4 \rangle$, which correspond to five segments of growths of (I_1, I_2): $[O, B)$, $[B, D)$, $[D, E)$, $[E, G)$, and G. In the figure, every state is augmented with a configuration during the state, where active contacts are represented by (virtual) springs.

The collision starts out with the upper ball hitting the lower ball and the lower ball in turn hitting the table. In the first state S_1, both impacts are active. The upper ball's velocity increases (i.e. its downward speed reduces), while the lower ball's decreases. Strain energies stored at the two contacts are increasing until at A, where the ball–ball contact begins to restitute, releasing part of its stored energy while losing the other part to irreversible deformation at contact. Restitution ends when the two impulses reach B with the two balls separating, one going upward and the other downward. The second state S_2, with only active ball–table contact, immediately follows. During S_2, the lower ball's velocity reaches zero at point C and continues growing afterward. It eventually catches up with the upper ball's velocity at D, reestablishing the ball–ball contact. Since both contacts are now active, S_2 transitions back to S_1 at D. During S_1, the ball–table impact finishes at E before the ball–ball impact. The ball–table contact breaks, and state S_3, with only the ball–ball contact active, starts. The ball–ball impact ends compression at F. After some loss due to the contact's plasticity, the remaining strain energy is converted into the upper ball's kinetic energy. As restitution ends at G, the terminal state S_4 is reached. Collision ends with both balls going upward and the upper ball at a faster speed.

1.1. Paper outline

This paper presents a formal treatment of the model for multiple impacts which was proposed by Jia et al. (2010) with new theoretical developments, and combines the model with a recently developed model (Jia, 2010)[3] for three-dimensional impact with friction and compliance. To focus on presenting the model for multiple impacts, we consider the absence of friction and tangential compliance. The dynamics equation (2) reduces to

$$M(\boldsymbol{q})\,\dot{\boldsymbol{q}} = A(\boldsymbol{q})\,I_n(\boldsymbol{q})\,. \tag{3}$$

As mentioned earlier, both matrices $M(\boldsymbol{q})$ and $A(\boldsymbol{q})$ depend on the geometry and configuration of the system, and can be set up for individual problems from dynamics and contact kinematics.

We focus on analyzing the ball–ball–table collision presented in Figure 1. There are several reasons for such a choice. Turning out to be two-dimensional, the collision is likely the simplest multiple impact problem. Yet, its solution can be easily generalized to solve a three-dimensional problem, like modeling of a massé billiard shot (to be examined later), in which one of the two impacts is eccentric.[4] A collision involving three rigid bodies with arbitrary geometry will be handled easily once the dynamics equation (3) is set up.

In Section 2, we present a quick review of single impact with frictionless contact, offering the derivative of contact energy with respect to impulse.

In Sections 3–7, we give a formal treatment of the multiple impact model for the ball–ball–table collision. Section 3 presents a transition diagram involving four states,

each representing a different combination of active contacts during the collision, and introduces a set of energy-based impact assumptions. Section 4 describes impact evolutions within the states. Section 5 reveals the effects of the stiffness ratio and the mass ratio, and establishes the scalability of the post-impact ball velocities over the pre-impact velocity of the upper ball. Section 6 introduces the impulse curve as a trajectory of growth in the ball–ball and ball–table impulses, and shows that during the collision the curve lies inside an ellipse derived based on non-negative contact strain energy. Instances of elastic and plastic collisions are illustrated. Section 7 proves the convergence of state transitions.

Section 8 integrates the multiple impact model with a model for three-dimensional impact with tangential compliance[5] developed by Jia (2012), to simulate the task of shooting a billiard ball. Impact dynamics and contact kinematics are combined with frictional contact modes analyzed according to the second model. Section 9 describes two experiments: one of dropping a ping-pong ball onto another and the other of a massé shot. In Section 10, we give a summary and discuss an extension to simultaneous collisions involving three or more bodies, along with other future research directions.

1.2. Related work on multiple impacts

All collisions conserve momentum under Newton's third law. Those that also conserve kinetic energy are elastic, while those that do not are plastic. Molecular collisions can be regarded as elastic. Collisions that we come across in daily life are generally plastic. The problem of determining post-collision velocities is under-constrained by momentum conservation alone. One or more impact laws need to be imposed accordingly. There are three commonly used laws: Newton's law, Poisson's hypothesis, and the energy hypothesis, which respectively specify the ratios between the velocities before and after an impact, the impulse growths during two different impact phases, and the strain energies stored and released during those phases. Jia (2012) presents a progressive overview of research in impact mechanics centered around these laws. Here we primarily survey work on multiple impacts, which happen simultaneously but are sometimes not treated exactly so in the analysis.

Newton's law of impact (Mason, 2001) asserts that the speed of an object after an impact is a constant fraction of that before the impact. It was applied very early on in the study of multiple impacts by (MacLaurin, 1742), and in particular, in the treatment of elastic collision in a system of spheres by Bernoulli (1969–1993). More recently, Ivanov (1995) examined four approaches that apply Newton's law to solve multiple impact problems with no friction. He reported that two of the approaches generated unrealistic results or no unique solution, one was highly sensitive to the initial conditions, and one had to consider all possible sequences of pairwise collisions (thus, increasing the complexity significantly).

Newton's law implies that still objects before an impact should remain still. This is nevertheless incorrect since energy often gets transferred to these objects in the impact duration. Consider a simple version of Newton's cradle (Brogliato, 1999), where three balls are aligned horizontally with the leftmost ball having an initial velocity to start the collisions. All three balls may get velocities out of the process (Liu et al., 2009). Another counterexample is dropping a ping-pong ball onto another one resting on the table. Both balls will bounce up as a result. Newton's law is generally inadequate for modeling multiple impacts and impacts with contact friction and tangential compliance.

Poisson's hypothesis (Routh, 1905) states that an impact between two bodies begins with a compression phase until their approaching velocity decreases to zero, and follows with a restitution phase until the two bodies fully separate. The hypothesis also asserts that the impulse accumulated during restitution is a fraction of that accumulated during compression. Aside from the possibility of predicting an increase in kinetic energy (Wang et al., 1992; Stewart and Trinkle, 1996), Poisson's hypothesis applies to the impact at each contact during simultaneous collisions in an isolated way, whereas the phases of the impacts at different contacts often depend on each other. There may not be enough energy stored at the contact to provide the amount of impulse growth during restitution as prescribed by Poisson's hypothesis (Jia et al., 2010). Impact analysis based on Poisson's hypothesis has been a subject of controversy over its consistency with Coulomb's law of friction and the law of energy conservation.

Glocker and Pfeiffer (1995) formulated two-dimensional impact under Coulomb friction and Poisson's hypothesis as a linear complementarity problem (LCP). They assumed that impacts at different contacts end compression and restitution simultaneously, and used a polyhedral approximation of the Coulomb friction cone to realize complementarity in the tangential direction (in addition to the non-negativeness of impulse in the normal direction). Impulses at the ends of compression and restitution were obtained under different LCP formulations, by employing Lemke's algorithm (Anitescu et al., 1996) which pivoted like the simplex algorithm in linear programming. A similar formulation of multi-rigid-body impact problems with friction was proposed by Stewart and Trinkle (1996), who applied a time-stepping (integration) scheme that extended to a nonlinear complementarity (NCP) formulation. Their method was modified by Anitescu and Porta (1997) in order to guarantee a solution (although multiple solutions may exist). We refer to Glocker (2001) for a survey of LCP-based methods for frictionless collision problems.

A LCP formulation of multiple impacts, however, does not correspond to fundamental physical properties (Chatterjee and Ruina, 1988). It is based on an unrealistic assumption that all impacts synchronize in compression and restitution. Lemke's algorithm only generates the impulse values after compression and restitution, but does not describe how the impulse accumulates, which is important

for contact mode analysis during the impact. Also, ambiguities exist because often the LCP solution is not unique.

It is worthy to note that none of the frictional impact models mentioned so far under Poisson's hypothesis handles tangential compliance, which would return a portion of the work done by the tangential interaction force during an impact rather than let it dissipate completely under friction.

Concurrent multiple collisions are sometimes sequenced into two-body collisions either by order of position (Ivanov, 1995), by order of normal approach velocity (Chatterjee and Ruina, 1988), or by a variational approach (Seghete and Murphey, 2010) that extremizes action over nearby trajectories of successive single impacts. Most of this type of works deal with frictionless impact. High-speed photographs of such collisions show that multiple objects are simultaneously in contact rather than only two at a time (Stewart, 2000). Observations also seem to suggest that, during the physical process, the involved objects may have broken and re-established contact multiple times. These coupling effects among impacts at various contacts were approximated by some impulse correlation ratios (Ceanga and Hurmuzlu, 2001). Similar studies of a 3-ball chain were conducted by Acary and Brogliato (2003) and by Acary and Taha (2005) with the introduction of a ratio between the total kinetic energies after and before the collision. One problem with the impulse correlation ratio is its lack of precise physical description.

Stronge (1990) developed an energy-based model that defines the energetic coefficient of restitution as the square root of the portion of the strain energy absorbed during compression that is to be released during restitution. This coefficient measures energy dissipation directly, and is consistent with the law of energy conservation, unlike Newton's impact law and Poisson's hypothesis.

Liu et al. (2008) described a framework for frictionless multiple impacts in a multibody system. Impulses are related to each other differentially, and determined by the relative contact stiffness as well as the ratios of corresponding contact potential energies. Numerical integration is always carried out over the impulse at the contact currently storing the maximum potential energy. Energetic coefficients of restitution are applied to individual impacts, each of which may go through multiple compression-restitution phases. Referred to as the 'LZB multiple impact model', this work deals with general contact stiffness that subsumes those for Hertz and linear contacts. Their sequel paper (Liu et al., 2009) presented a numerical algorithm and included simulation results for several benchmark problems including Newton's cradle, the Bernoulli problem, etc. As nicely as the theory was presented, it did not analyze the collision behavior in the course of the accumulation of individual impulses, nor did it provide a proof for termination or convergence.

Based on the LZB model, Zhao et al. (2008) studied energy dissipation and transfer during multiple frictionless impacts. They obtained numerical results for a column of particles (a granular chain) that were in good agreement with experimental findings (Falcon et al., 1998). Nguyen and Brogliato (2012) recently used the same model to describe wave propagation during impacts of multiple granular chains, matching some main observations by Nakagawa et al. (2003) from their experiment.

Jia et al. (2010) independently proposed a multiple impact model that observes the same differential relationship between impulses at various contacts, keeps track of contact strain energy for restitution, and performs numerical integration over the currently dominant impulse. The main difference from the LZB model is that Jia et al.'s model formulates the physical process as a state transition diagram, where each state represents a different combination of contacts that are instantaneously active. A transition from one state to another happens when either an active impact finishes restitution or an inactive impact gets reactivated. Every collision instance yields a sequence of states with proven termination/convergence. Jia et al. (2010) also showed that, for a linear arrangement of three objects (the third of which having infinite mass), the impulses grow along a planar curve bounded by an ellipse derived from non-negative strain energy during the collision. As mentioned earlier, this paper extends the work of Jia et al. (2010) in formal analysis, and integrates the model with another model (Jia, 2012) to handle friction and tangential compliance in multibody collisions in three dimensions.

2. Single impact with frictionless contact

This section presents a very brief review of single impact with frictionless contact. Suppose that a particle of mass m with downward velocity $v_0 < 0$ impacts on a horizontal plane. The particle receives a contact force $F \geq 0$ that varies in the duration τ of the impact. The force is physically continuous, and thus integrable over $[0, t]$, $0 \leq t \leq \tau$, yielding the impulse $I = \int_0^t F\, dt$. Conversely, we have $\dot{I} = F$ during the impact, where the dot denotes differentiation with respect to time. Furthermore, a one-to-one correspondence exists between t and I given that $F > 0$ except at the beginning and end of the physical process.

On the particle, the gravitational force is significantly less than the impulsive force F and thus ignored. It follows from Newton's second law that during the impact $F = m\dot{v}$, where v is the particle's velocity. Integration of the equivalent equation $\dot{I} = m\dot{v}$ establishes that the total impulse is equal to the particle's change in momentum. This yields the velocity equation during the impact:

$$v = v_0 + \frac{I}{m}. \tag{4}$$

Under energy conservation, the particle's post-impact velocity cannot exceed $-v_0$. The total impulse is thus finite. As the impact duration $\tau \to 0$, $F \to \infty$ in order to keep the integral I finite.

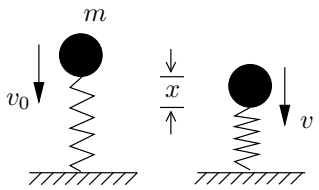


Fig. 2. Particle impacting a table.

It is reasonable to assume that the impact happens in infinitesimal time. The physical process is thus best analyzed in the impulse space. To better understand this process, we connect the particle and the plane with a virtual spring with stiffness k. See Figure 2. Let x measure the change in the virtual spring's length from its rest length. The value of x is negative when the spring is compressed. We have the ball velocity $v = \dot{x}$, the contact force $F = -kx$, and the potential energy stored by the spring $E = \frac{1}{2}kx^2$. Note that $E = 0$ only when the impact begins and finishes.

The impact can be divided into two stages (Mason, 2001, p. 212): compression and restitution. Compression transforms the particle's kinetic energy into the potential energy E of the spring, and ends at $\dot{x} = 0$ when the energy reaches its maximum value $E_{\max}$. The impulse exerted up to this point is $I = -mv_0$ by (4). Restitution then releases the elastic portion of the stored energy, of the amount $e^2 E_{\max}$. Here e, $0 \le e \le 1$, is referred to as the energetic coefficient of restitution,[6] which was analyzed by Stronge (1990).[7] The remaining portion $(1 - e^2)E_{\max}$ is simply dissipated due to some irreversible internal deformations. With the energy release, the impulse grows by an additional amount of $-emv_0$. The impact ends with the particle velocity $-ev_0$.

We adopt the explanation by Stronge (2000, p. 98) for the energy loss:[8] when compression ends the spring stiffness suddenly increases from k to k/e^2. Such increase can be attributed to the hardening of the material due to some irreversible deformation.[9] Continuity of F then ensures a simultaneous reduction of the change x in length of the spring to e^2x.[10]

The one-to-one correspondence between I and t establishes E as a function of I, even though the loss of energy after compression is directly due to deformation, not I. Discontinuous only at the end of compression, the function E is differentiable during each impact phase:

$$\begin{aligned} \frac{dE}{dI} = \frac{\dot{E}}{\dot{I}} &= \frac{d(\frac{1}{2}kx^2)/dt}{-kx} \\ &= -\dot{x} \qquad (5) \\ &= -v \\ &= -\left(v_0 + \frac{I}{m}\right), \qquad \text{by (4).} \qquad (6) \end{aligned}$$

The third equation above holds because the stiffness k does not change its value during compression or restitution. Since I is continuous, at the impact phase switch the left and right derivatives of E, both given by (6), are equal.

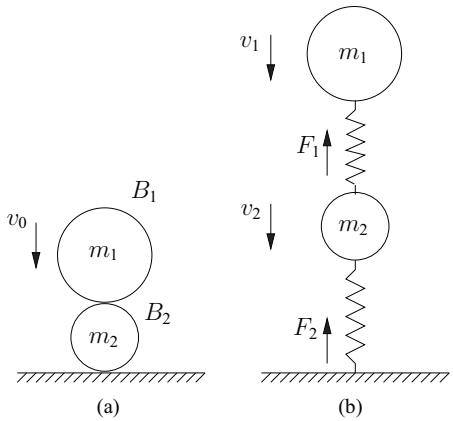


Fig. 3. (a) Ball-ball–table collision and (b) virtual springs for modeling.

3. Multiple impacts

From now on until Section 7, we study the following problem. A rigid ball $\mathcal{B}_1$ with mass m_1 and initial velocity $v_0 < 0$ strikes downward onto another rigid ball $\mathcal{B}_2$ with mass m_2 resting on a horizontal table. The two balls have their centers vertically aligned, and frictionless contacts between them and between $\mathcal{B}_2$ and the table. The scenario is depicted in Figure 3(a). As $\mathcal{B}_1$ collides with $\mathcal{B}_2$, the latter ball in turn impacts the table. The goal is to determine the rebound velocities of the two balls. This seemingly simple problem demands non-trivial analysis (which can be generalized to solve more complex problems).

Let v_1 and v_2 be the respective velocities of $\mathcal{B}_1$ and $\mathcal{B}_2$ during the collision. We attach a virtual spring between $\mathcal{B}_1$ and $\mathcal{B}_2$ with stiffness k_1, and another one between $\mathcal{B}_2$ and the table with stiffness k_2, as shown in Figure 3(b). The stiffness k_i, $i = 1, 2$, has original value $\bar{k}_i$. Following the discussion over single impact in Section 2, k_i will increase by a factor of $1/e_i^2$ whenever compression ends, where $e_i \in [0, 1]$ is the impact's coefficient of energy restitution.

Let x_1 and x_2 be the changes in length of the two virtual springs. The gravitational forces are negligible compared with the large impulsive forces F_1 and F_2 exerted on $\mathcal{B}_1$ and $\mathcal{B}_2$, respectively. The kinematic and dynamic equations are

$$\dot{x}_1 = v_1 - v_2, \qquad (7)$$

$$\dot{x}_2 = v_2, \qquad (8)$$

$$m_1\dot{v}_1 = F_1 = -k_1x_1, \qquad (9)$$

$$m_2\dot{v}_2 = F_2 - F_1 = k_1x_1 - k_2x_2. \qquad (10)$$

In the duration τ of the ball–ball–table collision, there are two impulses: $I_1 = \int_0^t F_1\,dt$ and $I_2 = \int_0^t F_2\,dt$, $0 \le t \le \tau$. Since $F_1, F_2 \ge 0$, they never decrease. Integrate (9) and (10):

$$v_1 = v_0 + \frac{1}{m_1}I_1, \qquad (11)$$

$$v_2 = \frac{1}{m_2}(I_2 - I_1). \qquad (12)$$

The two virtual springs store strain energies $E_1 = \frac{1}{2}k_1x_1^2$ and $E_2 = \frac{1}{2}k_2x_2^2$, respectively. Their derivatives[11] assume the form (5), into which we substitute (7), (8), (11), and (12):

$$\begin{aligned}\frac{dE_1}{dI_1} &= -\dot{x}_1 = v_2 - v_1 \\ &= -\left(v_0 + \left(\frac{1}{m_1} + \frac{1}{m_2}\right)I_1 - \frac{1}{m_2}I_2\right), \qquad (13)\\ \frac{dE_2}{dI_2} &= -\dot{x}_2 = -v_2 \\ &= \frac{1}{m_2}(I_1 - I_2). \qquad (14)\end{aligned}$$

In the meantime, a differential relationship exists between the two impulses I_1 and I_2 that depends on the two energies:[12]

$$\frac{dI_2}{dI_1} = \frac{\dot{I}_2}{\dot{I}_1} = \frac{k_2x_2}{k_1x_1} = \frac{\sqrt{k_2E_2}}{\sqrt{k_1E_1}}. \qquad (15)$$

The ordinary differential equations (ODEs) (13)–(15) can be viewed as a system in one variable I_1 (or I_2) that governs how the two impacts evolve.

3.1. State transition diagram

Although the ball–ball and ball–table impacts begin simultaneously, they will hardly end compression or restitution at the same time. For instance, when the ball–ball impact finishes restitution, the ball–table impact may be still undergoing restitution. As collision continues, the two balls may start moving toward each other again at some later point, reactivating the ball–ball impact.

This suggests that we sequence the collision process into (repeats of) three states: S_1, when both impacts are active; S_2, when only the ball–table impact is active; and S_3, when only the ball–ball impact is active. The three are referred to as the impact states. We also introduce a fourth state S_4, which is terminal and happens only if all active impacts end restitution simultaneously in S_1, S_2, or S_3.

With two active impacts the system is in S_1. Both impacts could end simultaneously, placing the system in state S_4. Otherwise, one impact ends before the other, placing the system in either state S_2 or state S_3. In S_2, there is no ball–ball impact, just a ball–table impact. If the ball–table impact causes the lower ball to catch up with the upper ball in velocity, the system then re-enters state S_1. Similarly, in state S_3, there is only a ball–ball impact. Again, the system re-enters state S_1 if the lower ball starts moving downward.

The above idea leads to a state transition diagram shown in Figure 4. The collision always starts with state S_1. A transition $S_1 \to S_2$ happens when the ball–ball impact finishes restitution before the ball–table impact. So the two balls are 'breaking' contact momentarily. Since the ball–ball impact was in restitution just before the transition, $\dot{x}_1 \geq 0$, which by (7) implies $v_1 \geq v_2$ when S_2 begins. Because gravity is neglected during the collision, v_1 will not vary during S_2. The state will transition back to S_1 if v_2 increases to become equal to v_1 before restitution of the ball–table impact ends. Otherwise, state S_4 will be reached with $v_1 \geq v_2$ to end the collision.

Note positiveness of the coefficient of restitution e_1 as part of the condition for the state transition $S_2 \to S_1$. When $e_1 = 0$, the state S_1 just before S_2 must have ended with compression of the ball–ball impact. There was no restitution phase in S_1 because the strain energy E_1 was lost under $e_1 = 0$. The two balls collapse into one object, reducing the remaining process to that of a single impact. This case is analyzed in the supplement Jia et al. (2011a, Section E.3).

Similarly, a transition $S_1 \to S_3$ happens when the ball–table impact finishes restitution before the ball–ball impact. At the moment, $v_2 \geq 0$. State S_3 will transition to S_1 when $v_2 = 0$, that is, when the lower ball re-establishes contact with the table. An extra condition $e_2 > 0$ is imposed over the transition, since otherwise the lower ball and the table collapse into one body, leading to a single impact problem (treated by Jia et al. (2011a)). If the ball–table impact finishes restitution in S_3, S_4 follows to end the collision.

State S_4 is the terminal state. The only computation within the state happens with plastic ball–ball collision ($e_1 = 0$) when this impact ends compression in state S_3. We again refer to Jia et al. (2011a, Section E.3) for a treatment of this special case.

An impact may start with one state, end compression in another, and finish restitution to end a third one. By induction on the number of states, we can easily establish that Equations (11) and (12) hold across any state sequence.

Proposition 1 *The velocities of the two balls are (11) and (12) during the collision.*

The collision described in Figure 1 is generated with $m_1 = 1$, $m_2 = 2/\sqrt{3}$, $k_2/k_1 = 1$,[13] $e_1 = 0.9$, and $e_2 = 0.7$. The initial velocity of the upper ball is $v_0 = -1$. The impulse trajectory takes the shape of a curve.[14] The dots A, B, F, G on the curve mark the ends of a ball–ball impact phase, while the dots C and E mark the ends of a ball–table impact phase. The dot D represents a transition $S_2 \to S_1$ due to $v_1 = v_2$ not the end of any impact phase. The coordinates of G are the final impulses exerted on the balls. Substituting them into Equations (11) and (12), we obtain the ball velocities after the collision. In Figure 1, we see that I_1 does not change in S_2 while I_2 does not change in S_3.

3.2. Restart of compression from restitution

From Section 2, a single impact starts with compression and ends with restitution. Once started, restitution will never switch back to compression. When two impacts are simultaneously active (i.e. in state S_1), however, this is no longer true. The two ball velocities, given in (11) and (12), are governed by the differential equations (13)–(15), the last of which is nonlinear. It may happen that during restitution of

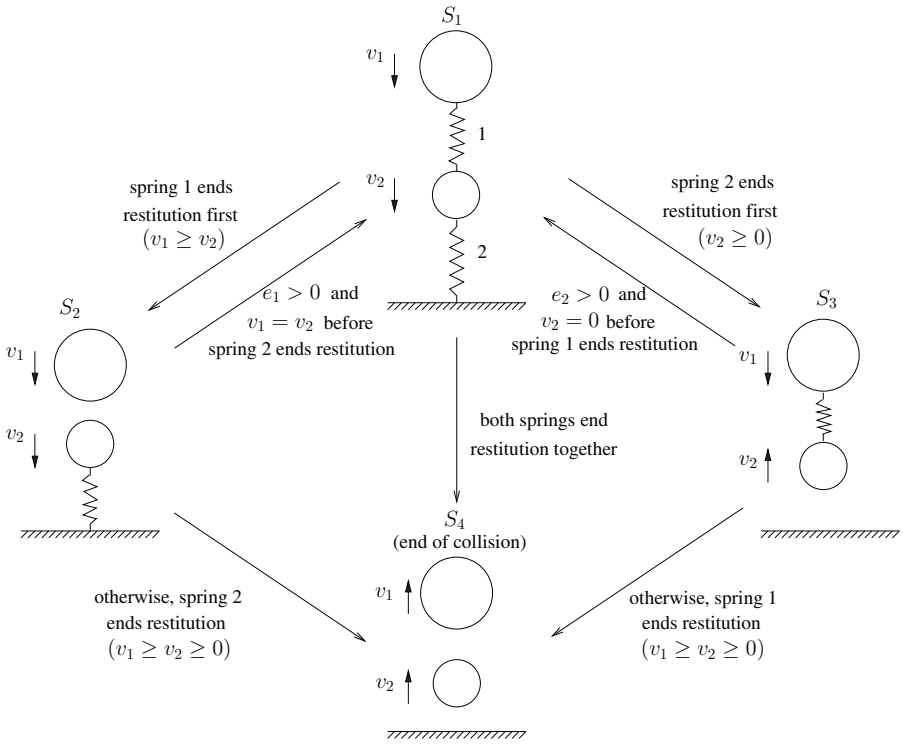


Fig. 4. State transition diagram for the ball–ball–table collision in Figure 3. During each state S_i, $1 \leq i \leq 3$, v_1 and v_2 can be either upward or downward, except v_2 is always upward in S_3. Both velocities are upward in S_4.

the ball–ball impact $v_1 - v_2$ decreases to zero and becomes negative. Then the impact will end restitution and restart compression at $v_1 = v_2$.[15] We defer an example to Section 6.3.3 after the notion of impulse curve is introduced.

3.3. Assumptions of multiple impacts

We extend the law of a single impact in Section 2 to the following set of three assumptions for multiple impacts (for $i = 1, 2$):

(i) Whenever the ith impact ends compression, the contact stiffness suddenly increases from k_i to k_i/e_i^2 while the strain energy suddenly drops from E_i to $e_i^2 E_i$. Or, more succinctly in the assignment forms, $k_i \leftarrow k_i/e_i^2$ and $E_i \leftarrow e_i^2 E_i$.

(ii) The value of k_i does not change whenever the ith impact switches from restitution back to compression.

(iii) Reactivation of the ith impact does not affect the value of k_i.

Under assumption (i), the contact force $\sqrt{2k_iE_i}$ is continuous when the impact ends compression. So is the impulse derivative dI_j/dI_i, $j \neq i$, which equals the ratio of the two contact forces.

When restitution switches back to compression, loading of the ith virtual spring restarts. Under assumption (ii), its stiffness does not change as the spring begins absorbing work into internal energy. Neither do the contact force F_i and the spring energy E_i. Hence, the impulse derivative is continuous at this transition.

Owing to possible restarts of compression, an impact may go through multiple compression and restitution phases. Suppose that the impact has ended compression j times, that is, restitution has switched back to compression $j-1$ times. Then the stiffness $k_i = \bar{k}_i/e_i^{2j}$ under assumptions (i) and (ii), where $\bar{k}_i$ is the original stiffness value.

When an impact gets reactivated, it should respect the effect of irreversible deformation that has happened before. This is reflected in zero change to the contact stiffness under assumption (iii). More specifically, at a transition $S_2 \rightarrow S_1$ or $S_3 \rightarrow S_1$, k_1 or k_2 retains its value at the end of the last state S_1.

3.4. Energy conservation unless end of compression

The total energy of the impact system consists of the kinetic energies of the two balls and the strain energies at the ball–ball and ball–table contacts. Energy loss happens only at an end of compression if the corresponding coefficient of restitution is less than one, as stated in the following theorem.

Theorem 2 *The total energy of the two-ball impact system decreases whenever impact i with $e_i < 1$, for some $i = 1, 2$, ends compression, but does not vary otherwise.*

Proof Under the impact assumptions in Section 3.3 when impact i with $e_i < 1$ ends compression, a loss of the strain energy E_i by a factor of $1 - e_i^2$ happens. We now prove that the total energy

$$U = \frac{1}{2}m_1 v_1^2 + \frac{1}{2}m_2 v_2^2 + E_1 + E_2$$

is conserved over a period of collision during which no compression of any impact ends.

From (13) and (14) we obtain

$$\begin{aligned}
-d(E_1 + E_2) =& v_0\, dI_1 + \left(\frac{1}{m_1} + \frac{1}{m_2}\right) I_1\, dI_1 \\
& - \frac{1}{m_2} I_2\, dI_1 + \frac{1}{m_2} I_2\, dI_2 - \frac{1}{m_2} I_1\, dI_2 \\
=& \left(v_0 + \frac{I_1}{m_1}\right) dI_1 + \frac{1}{m_2}(I_1\, dI_1 \\
& + I_2\, dI_2 - d(I_1 I_2)). \\
=& v_1\, dI_1 + \frac{1}{m_2} d\left(\frac{1}{2}(I_2 - I_1)^2\right) \quad \text{by (11)} \\
=& d\left(\frac{1}{2}m_1 v_1^2\right) + d\left(\frac{1}{2}m_2 v_2^2\right) \text{ by (11) and (12).}
\end{aligned}$$

The above equation implies energy conservation:

$$dU = d\left(\frac{1}{2}m_1 v_1^2 + \frac{1}{2}m_2 v_2^2 + E_1 + E_2\right) = 0.$$

Note that $dI_1 = 0$ in state S_2 and $dI_2 = 0$ in state S_3. □

3.5. Energy-based criteria for impact phases

For convenience, we sometimes refer to the ball–ball and ball–table impacts as the first and second impacts, respectively. Compression of the ith impact ends when $\dot{x}_i = 0$ and $\ddot{x}_i > 0$. Since $\dot{x}_i = -dE_i/dI_i$ by (5), $\ddot{x}_i = -(d^2E_i/dI_i^2)\cdot\dot{I}_i$, and $\dot{I}_i = F_i > 0$ at the moment, the two sufficient conditions are equivalent to $dE_i/dI_i = 0$ and $d^2E_i/dI_i^2 < 0$. Namely, the strain energy stored at the ith contact has reached a (local) maximum.

Under (13) and (14), the first-order necessary conditions for ends of the ball–ball and ball–table compression are, respectively,

$$v_0 + \left(\frac{1}{m_1} + \frac{1}{m_2}\right) I_1 - \frac{1}{m_2} I_2 = 0, \tag{16}$$

$$I_1 - I_2 = 0. \tag{17}$$

In state S_2, the ball–ball contact is inactive with $I_1 \equiv I_1^{(0)}$, its value at the start of the state. From (14) we obtain that $d^2E_2/dI_2 = -1/m_2 < 0$. If within the state I_2 reaches $I_1^{(0)}$ to make (17) hold, compression will end. Similarly, in state S_3, the ball–table contact is inactive with I_2 assuming its value $I_2^{(0)}$ at the state of the state. From (13) we obtain that $d^2E_1/dI_1^2 = -(1/m_1 + 1/m_2) < 0$. If I_1 grows to make (16) hold during the state, compression will end.

In state S_1, the end of compression will be conveniently determined using the impulse curve and compression lines introduced in Section 6.2.

4. Three impact states

To focus our investigation, from now through Section 6 we make several assumptions: (a) $k_2/k_1 > 0$ and is finite; (b) $e_1 \neq 0$; and (c) $e_2 \neq 0$. To save space, the degenerate cases $k_2/k_1 = 0$, $k_2/k_1 = \infty$, $e_1 = 0$, and $e_2 = 0$ are treated in detail in the supplement Jia et al. (2011a, Section E).[16] We only make some remarks about these cases when needed.

This section looks at how each of the states S_1, S_2 and S_3 evolves and how to decide the next state. State S_4 need only handle perfectly plastic ball–ball impact, which is also described in Jia et al. (2011a, Section E.3). Before examining the other three states one by one, we look at how to track the strain energies stored at the two contacts within a state.

4.1. Changes in strain energies within a state

Let us use the superscript '(0)' to refer to the value of a physical quantity at the start of a state, and the notation 'Δ' to refer to its amount of change so far in the state. Rewrite the velocities of the two balls (11) and (12) in terms of the impulse accumulations $\Delta I_1 = I_1 - I_1^{(0)}$ and $\Delta I_2 = I_2 - I_2^{(0)}$:

$$v_1 = v_1^{(0)} + \frac{1}{m_1}\Delta I_1, \tag{18}$$

$$v_2 = v_2^{(0)} + \frac{1}{m_2}(\Delta I_2 - \Delta I_1). \tag{19}$$

With (18) and (19), the energy derivative $dE_1/dI_1 = -\dot{x}_1$ from (13) is rewritten as

$$\frac{dE_1}{dI_1} = -\left(v_1^{(0)} - v_2^{(0)} + \left(\frac{1}{m_1} + \frac{1}{m_2}\right)\Delta I_1 - \frac{1}{m_2}\Delta I_2\right). \tag{20}$$

Integration of the above yields the change in E_1 from its value $E_1^{(0)}$ when the state began:

$$\begin{aligned}
\Delta E_1 = & \left(v_2^{(0)} - v_1^{(0)}\right)\Delta I_1 - \frac{1}{2}\left(\frac{1}{m_1} + \frac{1}{m_2}\right)\Delta I_1^2 \\
& + \frac{1}{m_2}\int_{I_1^{(0)}}^{I_1} \Delta I_2\, dI_1.
\end{aligned}$$

There is one bug with the above formula though. If compression has ended during the growth of I_1 from $I_1^{(0)}$ to its current value, we must include an energy loss by the amount of $(1 - e_1^2)E_{1\max}$, where $E_{1\max}$ was the value of E_1 at the end of compression. Introduce a two-value variable α_1 such that $\alpha_1 = 1$ if compression ends at some impulse value in

$(I_1^{(0)}, I_1]$, and $\alpha_1 = 0$ otherwise. Then the correct form of the energy change is

$$\Delta E_1 = \left(v_2^{(0)} - v_1^{(0)}\right)\Delta I_1 - \frac{1}{2}\left(\frac{1}{m_1} + \frac{1}{m_2}\right)\Delta I_1^2 + \frac{1}{m_2}\int_{I_1^{(0)}}^{I_1} \Delta I_2\, dI_1 + \alpha_1(e_1^2 - 1)E_{1\max}. \quad (21)$$

Similarly, from Equations (8) and (19) we rewrite (14) and then integrate it to obtain the energy change from $E_2^{(0)}$:

$$\frac{dE_2}{dI_2} = -\left(v_2^{(0)} + \frac{1}{m_2}(\Delta I_2 - \Delta I_1)\right), \quad (22)$$

$$\Delta E_2 = -v_2^{(0)}\Delta I_2 - \frac{1}{2m_2}\Delta I_2^2 + \frac{1}{m_2}\int_{I_2^{(0)}}^{I_2} \Delta I_1\, dI_2 + \alpha_2(e_2^2 - 1)E_{2\max}, \quad (23)$$

where α_2 takes on the value 1 or 0 depending on whether or not the ball–table impact ends compression at some impulse value in $(I_2^{(0)}, I_2]$, and $E_{2\max}$ is the strain energy at the end of compression.

4.2. State S_1: simultaneous ball–ball and ball–table impacts

Both impulses I_1 and I_2 are increasing in the state according to (15).

- If S_1 is the start of the collision or follows S_3, I_1 is the primary impulse (variable), and I_2 is the secondary impulse (function of I_1). We solve the system (13)–(15).
- If S_1 follows S_2, the two impulses switch their roles. We obtain I_1 as a function of I_2 defined by the equation

$$\frac{dI_1}{dI_2} = \sqrt{\frac{k_1}{k_2}} \cdot \sqrt{\frac{E_1}{E_2}}. \quad (24)$$

  This time we need to solve the system made up of (13), (14), and (24).

We first consider the case that S_1 is at the start of the collision, and characterize how I_2, E_1, and E_2 vary with I_1. From (13), we have $dE_1/dI_1(0) = -v_0$, and hence the Taylor series

$$E_1 = -v_0 I_1 + O(I_1^2). \quad (25)$$

In the above, we used the big-O notation such that $O(I_1^2) < cI_1^2$ for some constant c and some small enough value of I_1.

Next, we determine the Taylor series for I_2. Let $\rho = (dI_2/dI_1)(\delta)$, where δ is some small enough positive number. The idea is to solve for ρ from Equation (15) and then take the limit as $\delta \to 0$ to obtain the value of $dI_2/dI_1(0)$. With $I_1 = \delta$ and $I_2(\delta) \approx \rho\delta$, we substitute into (21) and (23) the following impulse integrals:

$$\int_0^{\delta} \Delta I_2\, dI_1 \approx \int_0^{\rho\delta} \frac{I_2}{\rho}\, dI_2 = \frac{1}{2}\rho\delta^2,$$
$$\int_0^{\rho\delta} \Delta I_1\, dI_2 \approx \int_0^{\delta} \Delta I_1 \cdot \rho\, dI_1 = \frac{1}{2}\rho\delta^2. \quad (26)$$

The two energies are then approximated as (setting $\alpha_1 = \alpha_2 = 0$ because no compression ends)

$$E_1 = \Delta E_1 \approx -v_0\delta - \frac{1}{2}\left(\frac{1}{m_1} + \frac{1}{m_2}\right)\delta^2 + \frac{1}{2m_2}\rho\delta^2,$$
$$E_2 = \Delta E_2 \approx \frac{1}{2m_2}\rho\delta^2(1 - \rho).$$

Plug the above equations into (15) and square both sides. A few more steps of symbolic manipulation lead to a quadratic equation:

$$\rho^2 + b_1\rho - \frac{k_2}{k_1} = 0,$$

where $b_1 = -2 \cdot \frac{m_2 v_0}{\delta} - 1 - \frac{m_2}{m_1} + \frac{k_2}{k_1} > 0$. Thus,

$$\left.\frac{dI_2}{dI_1}\right|_{\delta} = \rho \approx \frac{1}{2}\left(\sqrt{b_1^2 + 4k_2/k_1} - b_1\right) = \frac{2k_2/k_1}{b_1 + \sqrt{b_1^2 + 4k_2/k_1}}. \quad (27)$$

Since $4k_2/k_1 \ll b_1$ as $I_1 = \delta \to 0$, we have that

$$\frac{dI_2}{dI_1} = -\frac{k_2}{2m_2 v_0 k_1} I_1 + o(I_1), \quad (28)$$

where the small-o notation is used such that $\lim_{I_1 \to 0} o(I_1)/I_1 = 0$. From (28) we have

$$\left.\frac{dI_2}{dI_1}\right|_0 = 0 \quad \text{and} \quad \left.\frac{d^2 I_2}{dI_1^2}\right|_0 = -\frac{k_2}{2m_2 v_0 k_1}. \quad (29)$$

The ball–table impulse has a Taylor series about $I_1 = 0$:

$$I_2 = -\frac{k_2}{4m_2 v_0 k_1} I_1^2 + O(I_1^3). \quad (30)$$

Similarly, from (14) and applying (28) and (30), we obtain for small enough I_1:

$$\frac{dE_2}{dI_1} = \frac{dE_2}{dI_2} \cdot \frac{dI_2}{dI_1} = -\frac{k_2}{2m_2^2 v_0 k_1} I_1^2 + o(I_1^2).$$

This gives us the first three derivatives of E_2 at $I_1 = 0$, and the Taylor series about the point:

$$E_2 = -\frac{k_2}{6m_2^2 v_0 k_1} I_1^3 + O(I_1^4). \quad (31)$$

Proposition 3 *A unique solution to the system (13)–(15) exists in state S_1 that starts the collision.*

Proof Collision starts with $I_1 = I_2 = 0$ and $E_1 = E_2 = 0$. The initial changes in I_2, E_1, and E_2 as I_1 increases are unique according to (30), (25), and (31), respectively.

Once $I_1, I_2, E_1, E_2 > 0$, existence and uniqueness of the solution to the system follows from Picard's existence theorem (Wolfram Mathworld[17]) because the right-hand sides

of (13)–(15) are continuous and have continuous partial derivatives with respect to I_1, I_2, E_1, and E_2.

The right-hand side of Equation (15) stays continuous when the ith impact ends compression because the contact forces are continuous. Meanwhile, the discrete loss in E_i and increase in k_i have no effect over (13) and (14). □

Similarly, we can derive the Taylor series for the secondary impulse and strain energies in terms of the primary impulse at the start of S_1 when the state succeeds S_2 or S_3. Existence and uniqueness of solution within S_1 can be established. Here, $(dI_1/dI_2)(I_2^{(0)}) = 0$ if S_1 succeeds S_2 and $(dI_2/dI_1)(I_1^{(0)}) = 0$ if it succeeds S_3.

With no closed-form solution to (13)–(15) in general, the state evolution is simulated via numerical integration.[18] At each integration step, we check whether compression of either impact has just ended as described in Section 3.5, and if so, scale the corresponding strain energy E_i by e_i^2 accordingly to recognize the energy loss. The state transitions to S_2 when E_1 decreases to zero during restitution of the ball–ball impact, or to S_3 when E_2 decreases to zero during restitution of the ball–table impact, whichever occurs earlier.

In numerical integration, we use (27) instead of (29) so that I_2 increases if S_1 is at the start of the collision. For the other two cases, we initialize the impulse derivatives as follows:

$$\left.\frac{dI_1}{dI_2}\right|_{I_2^{(0)}+\delta} \approx \frac{2k_1/k_2}{b_2 + \sqrt{b_2^2 + 4k_1/k_2}},$$

if the preceding state was S_2,

$$\left.\frac{dI_2}{dI_1}\right|_{I_1^{(0)}+\delta} \approx \frac{2k_2/k_1}{b_3 + \sqrt{b_3^2 + 4k_2/k_1}},$$

if the preceding state was S_3,

where $b_2 = \frac{2m_2E_2^{(0)}}{\delta^2} - \frac{2m_2v_2^{(0)}}{\delta} - 1 + \frac{k_1}{k_2}\left(1 + \frac{m_2}{m_1}\right)$ and $b_3 = \frac{2m_2E_1^{(0)}}{\delta^2} - \frac{2m_2v_1^{(0)}}{\delta} - 1 - \frac{m_2}{m_1} + \frac{k_2}{k_1}$.

In the supplement Jia et al. (2011a, Sections E.1 and E.2), S_1 will be augmented to handle the special cases $k_2/k_1 = \infty$ and $k_1/k_2 = 0$. If $e_i = 0$ and impact i ends compression in S_1, another special treatment is needed. This will also be described in the same supplement (Jia et al., 2011a, Section E.3).

4.3. State S_2: ball–table impact only

The state S_2 starts with $\dot{x}_1 > 0$, or by (7), $v_2^{(0)} < v_1^{(0)}$ since the ball–ball impact has just finished restitution. During the state, $v_1 \equiv v_1^{(0)}$, $\Delta I_1 \equiv 0$, $\Delta E_1 \equiv 0$, and thus from (19),

$$v_2 = v_2^{(0)} + \frac{1}{m_2}\Delta I_2. \tag{32}$$

If $v_1^{(0)} < 0$, then $v_2^{(0)} < 0$. Thus, state S_2 starts during compression of the ball–table impact. Since compression would not end until $v_2 = 0$, v_2 will reach $v_1^{(0)}$ first, triggering a transition to S_1.

If $v_1^{(0)} > 0$, there are two cases depending on the phase of the ball–table impact when S_2 starts.

- Compression. Since $\dot{x}_2 = v_2$ by (8), we conclude that restitution will start during the state when $v_2 = 0$ at $\Delta I_2 = -m_2v_2^{(0)}$ by (32). Substituting this value and $\alpha_2 = 0$ into (23) to determine ΔE_2, we obtain $E_{2\max} = E_2^{(0)} + \frac{1}{2}m_2{v_2^{(0)}}^2$. The next state is $\mathcal{S}_1$ if there is enough strain energy after the loss to increase the velocity of the lower ball $\mathcal{B}_2$ to $v_1^{(0)}$. This leads to the condition $e_2^2E_{2\max} > \frac{1}{2}m_2{v_1^{(0)}}^2$. Otherwise, the final state S_4 is next to end the collision with $\mathcal{B}_2$ having the velocity

$$v_2 = \sqrt{\frac{2e_2^2E_{2\max}}{m_2}} = e_2\sqrt{{v_2^{(0)}}^2 + \frac{2E_2^{(0)}}{m_2}}.$$

- Restitution. The ball $\mathcal{B}_2$ will increase its velocity to $v_1^{(0)}$ to reactivate the ball–ball impact (and switches to the state S_1) if $\frac{1}{2}m_2{v_2^{(0)}}^2 + E_2^{(0)} > \frac{1}{2}m_2{v_1^{(0)}}^2$. Otherwise, a transition to state S_4 takes place to end the collision with

$$v_2 = \sqrt{{v_2^{(0)}}^2 + \frac{2E_2^{(0)}}{m_2}}.$$

The above analysis of state $\mathcal{S}_2$ can be augmented with the case where $e_2 = 0$ and the ball–table impact ends compression in the state. We also need to handle the case $k_2/k_1 = 0$ when S_2 is the second state and followed by S_4. These two cases are described in Jia et al. (2011a, Sections E.2 and E.3).

4.4. State S_3: ball–ball impact only

During state S_3, $\Delta I_2 \equiv 0$, and $E_2 \equiv 0$. The lower ball $\mathcal{B}_2$'s velocity has a reduced form from (19):

$$v_2 = v_2^{(0)} - \frac{1}{m_2}\Delta I_1, \tag{33}$$

while the upper ball $\mathcal{B}_1$'s velocity is still (18). Through an analysis based on the phase of the ball–ball impact during which S_3 starts, the maximum possible accumulation of I_1 within the state can be determined (Jia et al., 2011a, Section C) to be

$$\Delta I_{1\max} = \begin{cases} M\left(v_2^{(0)} - v_1^{(0)} + e_1\sqrt{\left(v_2^{(0)} - v_1^{(0)}\right)^2 + \frac{2}{M}E_1^{(0)}}\right), \\ \text{if } S_3 \text{ starts during compression,} \\ M\left(v_2^{(0)} - v_1^{(0)} + \sqrt{\left(v_2^{(0)} - v_1^{(0)}\right)^2 + \frac{2}{M}E_1^{(0)}}\right), \\ \text{if } S_3 \text{ starts during restitution,} \end{cases} \tag{34}$$

where $M = 1/(\frac{1}{m_1} + \frac{1}{m_2})$.

A transition $S_3 \to S_1$ happens if the lower ball's velocity reduces to zero before S_3 ends, that is, if

$$v_2^{(0)} - \frac{\Delta I_{1\max}}{m_2} < 0. \tag{35}$$

In this case, the impulse accumulation within S_2 is $\Delta I_1 = m_2 \left(v_2^{(0)} - 0\right) = m_2 v_2^{(0)}$ by (33).

If condition (35) does not hold, the state ends with impulse accumulation $\Delta I_1 = \Delta I_{1\max}$. The terminal state S_4 is next. The final ball velocities are obtained by plugging $\Delta I_1 = \Delta I_{1\max}$ and $\Delta I_2 = 0$ into (18) and (19).

A special treatment is given in Jia et al. (2011a, Section E.3) for the case where $e_1 = 0$ and the ball–ball impact ends compression in S_3.

5. Stiffness and mass ratios and input/output scalability

Sections 4.3 and 4.4 offered closed-form solutions respectively in states S_2 and S_3 based on the values of the impulses, velocities, and energies entering the states. Section 4.2 established existence of a unique solution in S_1 whether it starts the collision or succeeds S_2 or S_3.

Theorem 4 *A unique solution exists for the ball–ball–table collision problem based on the state transition diagram in Figure 4.*

Having described the states, we now take a high-level view of the ball–ball–table collision: as a system with input being the initial velocity v_0 of the upper ball $\mathcal{B}_1$, and output (or outcome) being the final (separation) velocities v_1 and v_2 of the two balls. The theorem below states how the collision outcome is influenced by the physical parameters.

Theorem 5 *The following statements hold about the collision outcome:*

(i) It depends on the stiffness ratio k_2/k_1 but not on individual values of k_1 and k_2. Scaling of k_1 and k_2 by the same factor does not change evolutions of the two impulses I_1 and I_2.

(ii) It depends on the mass ratio m_2/m_1 but not on individual values of m_1 and m_2. Scaling of m_1 and m_2 by a factor of s results in scaling of I_1 and I_2, and the strain energies E_1 and E_2, all by s, during the collision.

(iii) The output–input velocity ratios v_1/v_0 and v_2/v_0 are invariant to any change in v_0. Scaling of v_0 by a factor of s will result in scaling of I_1 and I_2 by s, and E_1 and E_2 by s^2 during the collision.

In fact, part (i) of Theorem 5 is a known result. Brogliato (1999, pp. 306–307) applied a time-based analysis to establish that the separation velocities of three balls colliding in a chain configuration depend on the relative stiffness at their two contact points.

Part (ii) of the theorem can be interpreted differently. Suppose that we impose two unilateral constraints: (a) the distance between the centers of the two balls is no less than the sum of their radii; and (b) the distance between the center of the lower ball and the table is no less than its radius. The kinetic angle, defined by Brogliato (1999) as that between the normals of the corresponding equality constraints, depends on the mass ratio but not on individual masses. The geometry of the collision is determined by the mass ratio (up to scaling), as we will soon examine in Section 6.

Following the theorem, we see that the ratios v_1/v_0 and v_2/v_0 depend on the stiffness ratio k_2/k_1, the mass ratio m_2/m_1, and the coefficients of restitution e_1 and e_2 only.

Input/output scalability over the initial upper ball velocity and dependence on the mass ratio are two desired properties for an impact model using linear stiffness. They also simplify our analysis of the collision process and computation of the final ball velocities. To solve a collision instance with initial velocity $v_0 < 0$, and ball masses m_1 and m_2, we instead consider another instance where the upper ball has unit mass and unit initial velocity -1, and the lower ball has mass m_2/m_1. Suppose the new instance yields final ball velocities $\bar{v}_1$ and $\bar{v}_2$. Then the original collision instance yields final velocities $-v_0\bar{v}_1$ and $-v_0\bar{v}_2$. The state sequences from the two instances are exactly the same under a one-to-one correspondence between their impulse growths.

6. The impulse curve

From now on, we need only consider a collision instance where

$$m_1 = 1 \quad \text{and} \quad v_0 = -1. \tag{36}$$

Under Theorem 5, analysis of impulse evolution will carry over for arbitrary $v_0 < 0$ and $m_1 > 0$ merely by scaling the impulses and velocities by a factor of $-m_1 v_0$ and the strain energies by a factor of $m_1 v_0^2$.

Accumulations of the ball–ball impulse I_1 and the ball–table impulse I_2 are best described in the impulse plane with I_1 and I_2 as coordinates. Every point in this I_1-I_2 plane represents a pair of impulse values that may or may not be attained during the collision. By Proposition 1, the point would yield intermediate ball velocities according to (11) and (12). As I_1 and I_2 accumulate, the moving point (I_1, I_2) traces out a set of points in the plane. Since the contact forces F_1 and F_2 are continuous and non-negative, the impulse I_1 and I_2 as their integrals are continuous and monotonically increasing.

The impulse curve consists of all pairs of intermediate impulse values (I_1, I_2). It grows upward and/or rightward. The next proposition establishes the first order continuity of the impulse curve.

Proposition 6 *The impulse curve is first-order continuous, namely, its tangent is defined everywhere as the curve extends with the growing impulses during the collision.*

Proof The impulse derivative $dI_j/dI_i = F_j/F_i$, where I_i and I_j are the primary and secondary impulses, respectively, is continuous within state S_1. Within state S_2 or S_3, the secondary impulse I_i is constant, and the derivative $dI_i/dI_j = 0$, $j \neq i$, is clearly continuous.

Now we show that the tangent is defined at every possible state transition by establishing the equality between the left and right derivatives of the secondary impulse with respect to the primary impulse. There are four types of transitions among states S_1, S_2, and S_3 as shown in Figure 4.

- $S_1 \to S_2$. At the end of S_1, the two balls break contact with $E_1 = 0$. Thus, the left derivative $dI_1^-/dI_2 = \sqrt{k_1/k_2} \cdot \sqrt{E_1/E_2} = 0$. Throughout state S_2, I_1 remains constant, which implies the right derivative $dI_1^+/dI_2 = 0$ at the transition point.
- $S_1 \to S_3$. Similar to the above case.
- $S_2 \to S_1$. The left derivative $dI_1^-/dI_2 = 0$ since I_1 does not change its zero value within S_2. The right derivative $dI_1^+/dI_2 = \sqrt{\frac{k_1 E_1}{k_2 E_2}} = 0$ because $E_1 = 0$ at the start of S_1.
- $S_3 \to S_1$. Similar to the above case.

A transition to state S_4 ends the impact right away.[19] □

6.1. Bounding ellipse

In this section, we show that the two growing impulses are bounded inside an ellipse determined by m_2. To be derived from non-negativeness of the total contact strain energy, this bounding ellipse will help us not only visualize impulse growth during each state but also prove convergence of state transitions later in Section 7. The next lemma bounds the total strain energy using the impulses.

Theorem 7 *The following inequality is satisfied during a collision where the upper ball has unit mass ($m_1 = 1$) and unit initial velocity ($v_0 = -1$):*

$$E_1 + E_2 \le I_1 - \frac{1}{2} I_1^2 - \frac{1}{2m_2}(I_1 - I_2)^2 . \tag{37}$$

Proof By induction. The energy equations (21) and (23), after substitution of (36), become inequalities with removal of the terms containing α_1 and α_2 from their right-hand sides:

$$\Delta E_1 \le \left(v_2^{(0)} - v_1^{(0)}\right) \Delta I_1 - \frac{1}{2}\left(1 + \frac{1}{m_2}\right)\Delta I_1^2 + \frac{1}{m_2}\int_{I_1^{(0)}}^{I_1} \Delta I_2 \, dI_1, \tag{38}$$

$$\Delta E_2 \le -v_2^{(0)} \Delta I_2 - \frac{1}{2m_2}\Delta I_2^2 + \frac{1}{m_2}\int_{I_2^{(0)}}^{I_2} \Delta I_1 \, dI_2. \tag{39}$$

From $d(I_1 I_2) = I_2 \, dI_1 + I_1 \, dI_2$ we have

$$\int_{I_1^{(0)}}^{I_1} \Delta I_2 \, dI_1 + \int_{I_2^{(0)}}^{I_2} \Delta I_1 \, dI_2 = \Delta I_1 \Delta I_2.$$

Add up the inequalities (38) and (39):

$$\begin{aligned}
\Delta E_1 + \Delta E_2 &\le \left(v_2^{(0)} - v_1^{(0)}\right) \Delta I_1 - v_2^{(0)} \Delta I_2 \\
&\quad - \frac{1}{2m_2}(\Delta I_1 - \Delta I_2)^2 - \frac{1}{2}\Delta I_1^2 \\
&= \left(\frac{1}{m_2}\left(I_2^{(0)} - I_1^{(0)}\right) + 1 - I_1^{(0)}\right) \Delta I_1 \\
&\quad - \frac{1}{m_2}\left(I_2^{(0)} - I_1^{(0)}\right)\Delta I_2 - \frac{1}{2}\Delta I_1^2 \\
&\quad - \frac{1}{2m_2}(\Delta I_1 - \Delta I_2)^2 ,
\end{aligned} \tag{40}$$

where in the last step we substituted $v_1^{(0)} = -1 + I_1^{(0)}$ and $v_2^{(0)} = (I_2^{(0)} - I_1^{(0)})/m_2$ by (11) and (12).

In the initial state S_1, $v_1^{(0)} = -1$, $v_2^{(0)} = 0$, $I_i^{(0)} = 0$, $\Delta I_i = I_i$, and $\Delta E_i = E_i$, for $i = 1, 2$. This simplifies (40) to

$$E_1 + E_2 \le I_1 - \frac{1}{2} I_1^2 - \frac{1}{2m_2}(I_1 - I_2)^2 .$$

Next, assume that (37) holds after the first k states; namely, at the start of the $(k+1)$th state,

$$E_1^{(0)} + E_2^{(0)} \le I_1^{(0)} - \frac{1}{2} I_1^{(0)2} - \frac{1}{2m_2}\left(I_1^{(0)} - I_2^{(0)}\right)^2 .$$

Add the above inequality to inequality (40), merging the terms on the right-hand side:

$$\begin{aligned}
E_1 + E_2 &= E_1^{(0)} + E_2^{(0)} + \Delta E_1 + \Delta E_2 \\
&\le \left(I_1^{(0)} + \Delta I_1\right) - \frac{1}{2}\left(I_1^{(0)} + \Delta I_1\right)^2 \\
&\quad - \frac{1}{2m_2}\left(\left(I_1^{(0)} + \Delta I_1\right) - \left(I_2^{(0)} + \Delta I_2\right)\right)^2 \\
&= I_1 - \frac{1}{2} I_1^2 - \frac{1}{2m_2}(I_1 - I_2)^2 .
\end{aligned}$$

□

From the proof we see that (37) becomes an equality if and only if there has been no energy loss. This is the case throughout the collision if $e_1 = e_2 = 1$. Otherwise, no energy loss occurs until the first impact, with coefficient of restitution less than one, ends compression.

Since $E_1 + E_2 \ge 0$, condition (37) imposes a constraint over the ball–ball and ball–table impulses.

Corollary 8 *Every point on the impulse curve satisfies*

$$\frac{1}{2} I_1^2 + \frac{1}{2m_2}(I_1 - I_2)^2 - I_1 \le 0. \tag{41}$$

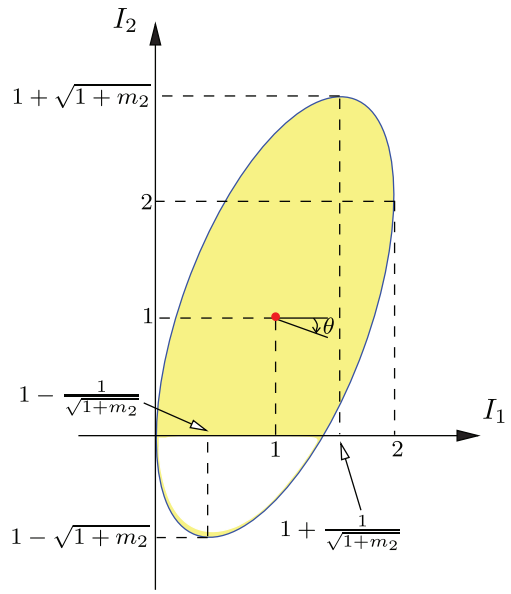


Fig. 5. The impulse curve must lie inside an ellipse and above the I_1-axis. Here $m_1 = 1$ and $v_0 = -1$.

Appendix A shows that the equation

$$\frac{1}{2}I_1^2 + \frac{1}{2m_2}(I_1 - I_2)^2 - I_1 = 0 \tag{42}$$

defines an ellipse (see Figure 5) centered at $(1, 1)$. The semi-minor axis rotates from the x-axis through an angle

$$\theta = -\frac{1}{2}\arctan\left(\frac{2}{m_2}\right). \tag{43}$$

The ellipse is tangent to two vertical lines: the I_2-axis at the origin and $I_1 = 2$ at the point $(2, 2)$. It is also tangent to two horizontal lines: $I_2 = 1 - \sqrt{1+m_2}$, and $I_2 = 1 + \sqrt{1+m_2}$.

During the collision, the growing impulses (I_1, I_2) must stay inside the region bounded by the ellipse (42) and above the I_1-axis. This region, shaded in Figure 5, is referred to as the impulse region. Rewrite the velocity equations (11) and (12) into the matrix form (note $m_1 = 1$ and $v_0 = -1$):

$$\begin{pmatrix} 1 & 0 \\ -\frac{1}{m_2} & \frac{1}{m_2} \end{pmatrix}\begin{pmatrix} I_1 \\ I_2 \end{pmatrix} = \begin{pmatrix} v_1 + 1 \\ v_2 \end{pmatrix}.$$

Since the coefficient matrix is non-singular, different points in the impulse region, if representing the impulses at the end of the collision, would generate different final ball velocities.

Given the same lower ball mass, every triple $(k_2/k_1, e_1, e_2)$ defines a different collision instance. Thus, the impulse region bounds a three-dimensional set of impulse curves resulting from all possible collisions. Every point in the region potentially lies on the impulse curve of some collision instance.

It is easy to see that the result on bounded impulse curve generalizes to a collision at $n > 2$ contacts by deriving an inequality for the total contact strain energy in terms of impulses.

6.2. Partitioning the impulse region

In the impulse plane, we add two lines ℓ_1 and ℓ_2 defined by Equations (16) and (17), respectively. Under (13) and (14), these two lines have equations $v_2 - v_1 = 0$ and $v_2 = 0$. They respectively represent the first-order necessary conditions for ends of compression of the ball–ball and ball–table impacts. Referred to as the compression lines by convention (Mason, 2001, p. 215), they partition the impulse region into four smaller regions I–IV as shown in Figure 6(a). Inside each region is a different combination of the signs of $v_1 - v_2$ and v_2 or, equivalently, according to (7) and (8), of the signs of the change rates $\dot{x}_1$ and $\dot{x}_2$ of the virtual contact spring lengths, and thus a different combination of the phases of the two impacts. The interior of region I includes all impulse values that could possibly be attained during compression of both the ball–ball and ball–table impacts. Accordingly, the region is labeled C1C2 in Figure 6(b). The interior of region II includes all impulse values that could possibly be attained during restitution of the ball–ball impact and compression of the ball–table impact. The region is thus labeled R1C2. And so on.

The impulse curve evolves from the origin at which it is tangent to the I_1-axis since dI_2/dI_1 has initial value 0 by (29). The curve enters region I as both impulses increase. The curve later crosses ℓ_1 (ℓ_2, respectively) where $dE_1/dI_1 = 0$ ($dE_2/dI_1 = (dE_2/dI_2)\cdot(dI_2/dI_1) = 0$, respectively). It is not difficult to picture that in state $\mathcal{S}_1$, both impacts could be in either compression or restitution.

In state S_2, $v_1 > v_2$ holds because the two balls are instantaneously separate. So the curve grows vertically upward to the right of ℓ_1. In S_3, the lower ball has velocity $v_2 > 0$ since it is separate from the table. The curve grows horizontally rightward to the left of ℓ_2.

When a state finishes with the value (I_1, I_2) inside region IV, $v_1 > v_2 > 0$, state S_4 follows to end the collision.

During the collision, $E_1 + E_2 > 0$. Thus, by (37) the inequality (41) is strict. It is also strict after the collision if there has been some energy loss. Therefore, after the start, the impulse curve will not reach the bounding ellipse (42) again unless the collision ends and $e_1 = e_2 = 1$.

6.3. Collision examples

This section presents several collision instances: an elastic one, a plastic one, and a third one in which compression restarts during restitution. In all instances, the upper ball has unit mass ($m_1 = 1$) and unit downward initial velocity ($v_0 = -1$).

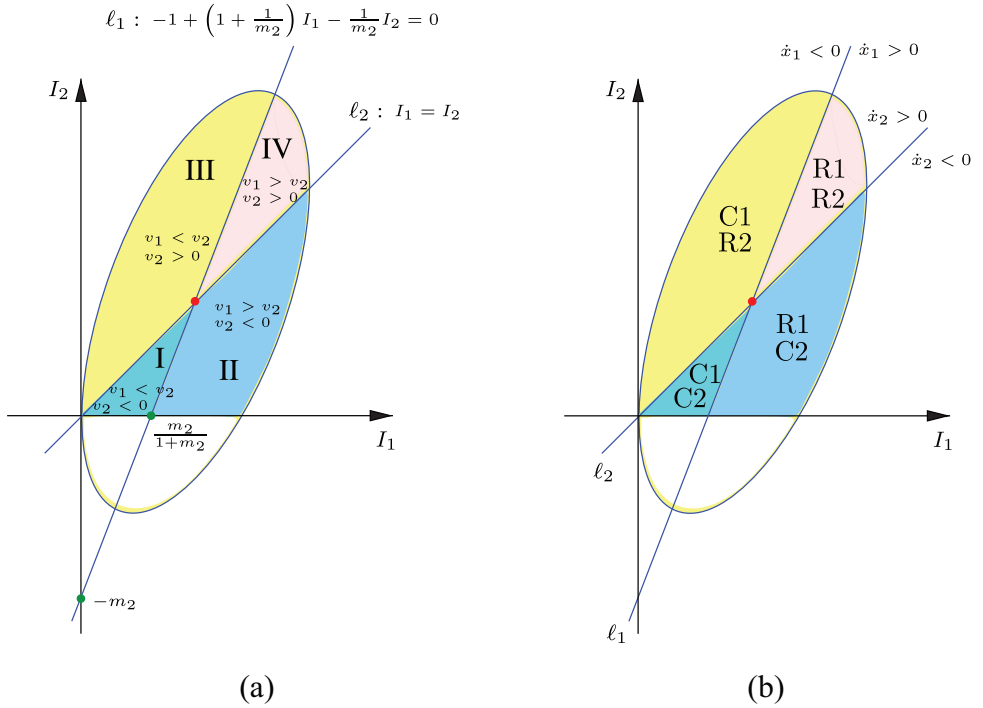


Fig. 6. Regions of four different combinations of impact phases.

6.3.1. Elastic collision Figure 7 illustrates the growth of the impulse curve during a collision with additional physical parameter values specified as follows:

$$m_2 = \frac{2}{\sqrt{3}}, \quad \frac{k_2}{k_1} = 1, \quad \text{and} \quad e_1 = e_2 = 1. \tag{44}$$

The bounding ellipse has its minor axis rotated clockwise from the x-axis through $\pi/6$. The equations of the ellipse and the two compression lines l_1 and l_2 are marked in the figure. The dots A, D, H, J on the impulse curve each marks the end of a ball–ball impact phase, while the dots E and G each marks the end of a ball–table impact phase. The dot F represents a transition from state S_2 to state S_1 that is triggered by a velocity condition ($v_1 = v_2$) not by the end of an impact phase.

Starting at the origin tangent to the I_1-axis and entering region I, the impulse curve crosses the compression line ℓ_1 at the point A. By (13), the two balls have the same velocity, i.e. $v_1 = v_2$, at the point. Since the curve is entering region II, the ball–ball impact has finished compression. The curve continues growing to reach the point B, at which its tangent is parallel to the compression line ℓ_2 with slope 1. At B we infer $dI_2/dI_1 = 1$ from the equation of ℓ_2, and $dv_2/dI_1 = 0$ from (12). Since the curve slope is increasing at B, so is dv_2/dI_1. Hence, $d^2v_2/dI_1^2 > 0$ holds at B, which implies that v_2 has reached a local minimum value. The value turns out to be an absolute minimum as the only local minimum below ℓ_2 (where $v_2 < 0$). In other words, the lower ball has the largest downward velocity when the two impulses are at B.

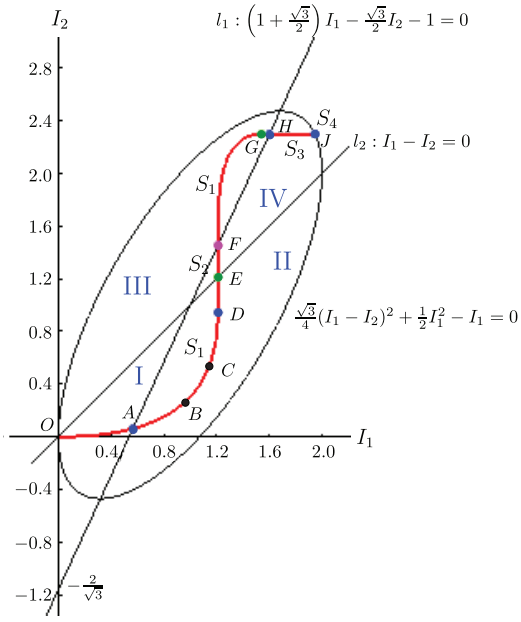


Fig. 7. Elastic collision. The values of the physical constants are given in (44). There are five states sequentially: S_1 over the impulse segment $[O, D)$, S_2 over $[D, F)$, S_1 over $[F, G)$, S_3 over $[G, J)$, and S_4 at J.

The next point of interest on the curve is C, at which the tangent is parallel to ℓ_1 with slope $1+2/\sqrt{3}$. Differentiating the last equation in (13), we obtain

$$\frac{d}{dI_1}(v_2 - v_1) = \frac{1}{m_2}\left(\frac{dI_2}{dI_1} - 1\right) - 1. \tag{45}$$

Since $dI_2/dI_1 = (1 + \sqrt{3}/2)/(\sqrt{3}/2)$ at C, the same as the slope of ℓ_1, the derivative of $v_2 - v_1$ vanishes. At the point, the curve slope is increasing, so $d^2 I_2/dI_1^2 > 0$. This suggests $d^2(v_2 - v_1)/dI_1^2 > 0$ according to (45). Thus, $v_2 - v_1$ has a local minimum at C. Since $v_1 > v_2$ to the right of ℓ_2, the lower ball's velocity is the most below the upper ball's.

The ball–ball impact ends restitution when the impulse curve reaches D. State S_2 follows immediately. During the state, the ball–table impact ends compression when it crosses the line l_2 at E. As I_2 continues to accumulate, the curve extends to F, where a transition $S_2 \to S_1$ happens. The ball–table impact ends restitution at G with a transition $S_1 \to S_3$.

In state S_3, the ball–ball impact as the only active one ends compression when the impulse curve crosses l_1 at H and ends restitution at $J = (1.94484, 2.29559)$, which lies on the bounding ellipse since there is no energy loss under $e_1 = e_2 = 1$. The velocities after the collision are $v_1 = 0.94484$ and $v_2 = 0.30376$, respectively.

6.3.2. Plastic collision We change the coefficients of restitution to $e_1 = 0.9$ and $e_2 = 0.7$ while keeping the values of the other collision parameters in (44). Now, neither impact is elastic. This is the collision instance shown in Figure 1. With the impulse growth already overviewed following that figure, we focus on how the ball velocities and the strain energies evolve as the impulses accumulate.

The collision goes through a sequence of five states: S_1, S_2, S_1, S_3, and S_4. They correspond to impulse curve segments $[O, B)$, $[B, D)$, $[D, E)$, $[E, G)$, and G in Figure 1. Owing to the energy loss, the ending point G is now inside a bounding ellipse, the same as that shown in Figure 7. Part (a) of Figure 8 plots the varying ball velocities against each other, while part (b) plots the varying strain energies stored at the two contacts. The velocity curve in Figure 8(a) is continuous, following the continuity of the impulse curve and the velocity formulae (11) and (12).

Points on the three curves in Figure 1, and Figures 8(b) and (c) that are labeled by the same letter correspond to the same instant during the collision. On the latter two curves, ⓘ labels all segments of impulse evolution within the ith state in the sequence.

The velocity curve starts at $(-1, 0)$ in Figure 8(a), and the energy curve at the origin in (b). The energy loss (of E_1) by a factor $1 - e_1^2 = 0.19$ happens at the point A when the ball–ball impact finishes compression. It is shown as a discontinuity of the energy curve at A in Figure 8(b). After the loss, the curve resumes at the point A' to the left of A.

In Table 1, each row lists the values of the impulses, the ball velocities, and the strain energies at the start of a state in the sequence. For example, the second to last entry in the row of S_2 gives the total energy $\frac{1}{2}m_1v_1^2 + \frac{1}{2}m_2v_2^2 + E_1 + E_2 = 0.44736$ of the system at the end of the first state: S_1, following the energy loss at A. The last entry in the previous row, C1–R1, marks that impact 1 (i.e. the ball–ball impact) has finished compression and restitution within the state S_1.

In the second state (S_2) starting at B with only the ball–table impact active, the velocity and energy curves both evolve vertically from B upward since I_1, v_1, and E_1 do not change their values. The energy curve, however, has a discontinuity at the point C when the ball–table impact ends compression. Here E_2 suffers a loss by the factor of $1 - e_2^2 = 0.51$. The curve resumes at C' below on the E_2-axis. When the velocity curve reaches D at which $v_1 = v_2$, the ball–ball impact gets reactivated immediately to end S_2. Note the small decrease in E_2 over $[C', D]$.

The third state S_1 finishes with restitution of the ball–table impact and no energy loss.[20] In the following state S_3, the velocity curve extends with slope $-1/m_2$ according to (33). The energy curve segment, labeled ④ in Figure 8(b), is on the E_1-axis. It first grows to the right to reach F where compression of the ball–ball impact ends. Restitution starts at F' following an energy loss, and releases the remaining strain energy. The energy curve comes back to the origin with the final ball velocities 0.61377 and 0.20947 at G in Figure 8(a) and in the last row of Table 1.

6.3.3. Restart of compression Figure 9 shows the impulse curve of a collision with only three states: S_1, S_3, and S_4. In S_1, the impulse curve crosses the compression line l_1 three times, at A, B, and D, respectively. The first and third crossings are from left to right, indicating ends of compression. Each of them results in increase of the contact stiffness k_1 by a factor of $1/e_1^2$. The second crossing indicates a switch from restitution back to compression with no change to the value of k_1. Note that in S_1 the impulse curve passes through all four regions.

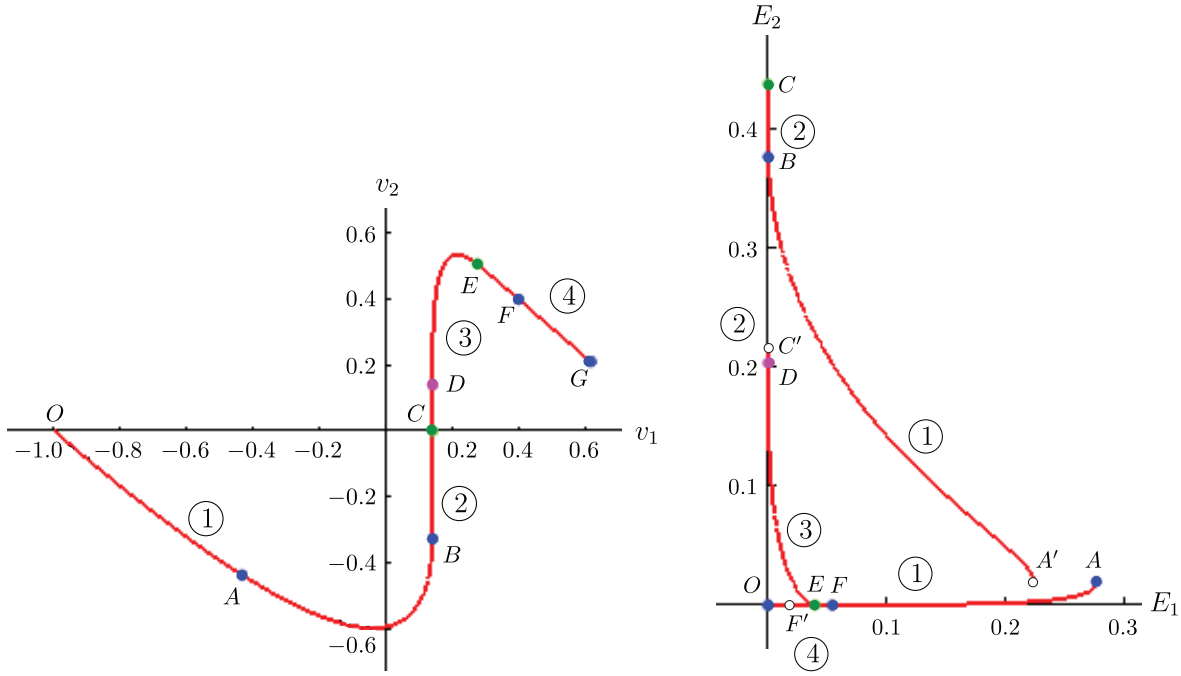
The impulse curve also crosses the compression line l_2 at C and E within state S_1. The ball–table contact becomes stiffer by $1/e_2^2$ after C. Similarly, the second crossing point E corresponds to a switch from restitution to compression of the ball–table impact.

7. Convergence of state transitions

Every state terminates. This is trivial for states S_2 and S_3 in which the impulse accumulations ΔI_2 and ΔI_1 are given in closed forms, respectively, in Sections 4.3 and 4.4. In state S_1, the primary impulse must stop increasing because the growth of (I_1, I_2) is bounded by the ellipse (42).

Special collisions (with $k_2/k_1 = 0$, $k_2/k_1 = \infty$, $e_1 = 0$, or $e_2 = 0$) all terminate after a few state transitions (Jia et al., 2011a, Section E). In this section we look at the general case where k_2/k_1 is finite and non-zero, and e_1 and e_2 are non-zero. Under Theorem 5, we need only consider a collision instance with $v_0 = -1$ and $m_1 = 1$.

Let $\langle s_1, s_2, \ldots \rangle$ be the sequence of states that results from the collision. Denote by I_{1i} and I_{2i} the respective values of the two impulses I_1 and I_2 at the end of the ith state s_i. The sequence $\{(I_{1i}, I_{2i})\}$ is monotonically non-decreasing. In the case that it is finite, from the diagram in Figure 4, state transitions will terminate with $v_1 \geq v_2 \geq 0$. In other



(a) Velocity curve (b) Energy curve

Fig. 8. Collision with energy loss. Diagrams (a) and (b) show the evolutions of the ball velocities and the strain energies at the two contacts. Here $m_1 = 1$, $m_2 = 2/\sqrt{3}$, $k_2/k_1 = 1$, $e_1 = 0.9$, $e_2 = 0.7$, and $v_0 = -1$.

Table 1. Values of I_1, I_2, v_1, v_2, E_1, and E_2 at the beginning of every state during the collision depicted in Figures 1 and 8.

State	Interval	I_1	I_2	v_1	v_2	E_1	E_2	Total energy	Phases
S_1	$[O, B)$	0	0	−1	0	0	0	0.5	C1-R1
S_2	$[B, C)$	1.13807	0.76239	0.13807	−0.32535	0	0.37671	0.44736	C2
S_1	$[C, E)$	1.13807	1.29750	0.13807	0.13807	0	0.20353	0.22407	R2
S_3	$[E, G)$	1.27281	1.85565	0.27281	0.50475	0.03966	0	0.22397	C1-R1
S_4	G	1.61377	1.85565	0.61377	0.20947	0	0	0.21369	

words, the impulse curve will stop its growth in region IV shown in Figure 6(a), with its boundary included.

Suppose that the monotonically increasing sequence is infinite. Bounded inside the ellipse (42), it must converge to some point (I_1^*, I_2^*) by a known result from calculus (Kaplan, 1991, p. 173).

Lemma 9 *The sequence* $\{(I_{1i}, I_{2i})\}$ *of pairs of impulses ending the states* $s_1, s_2, \ldots$ *will not converge to a point in the interior of region I, II, III, or IV drawn in Figure 6(a).*

Proof Suppose that the sequence is infinite and converges to some point q in the interior of one of the regions, say, IV. There exists some integer m such that, for all $i > m$, the points (I_{1i}, I_{2i}) are within some neighborhood N of q that is inside the region. However, because the sequence is infinite, state S_1 must transition to S_2 or S_3. State S_2 and S_3 each has single impact, and the impulse curve will always reach the compression line ℓ_1 or ℓ_2 at the end of such a state when it transitions to $\mathcal{S}_1$. The sequence must therefore exit the neighborhood N at the end of every other state at least. A contradiction. □

Following Lemma 9, an infinite sequence $\{(I_{1i}, I_{2i})\}$ must converge to a point (I_1^*, I_2^*) on the compression line ℓ_1 or ℓ_2. It only makes physical sense to have an outcome with both balls rebounding and the upper ball at a speed no less than that of the lower ball. The convergence point (I_1^*, I_2^*) should be at or above $(1, 1)$ to yield $v_1 \geq v_2 \geq 0$.

To show that (I_1^*, I_2^*) is above $(1, 1)$, we will use the following result regarding the minimum increase of the ball–ball impulse or the ball–table impulse within state S_1 during compression.

Lemma 10 *State* S_1 *will not end*

(i) if the accumulation of the ball–ball impulse within the state satisfies

$$\Delta I_1 \leq -\frac{v_1^{(0)}}{1 + \frac{1}{m_2}\left(1 + \frac{k_2}{k_1}\right)} \tag{46}$$

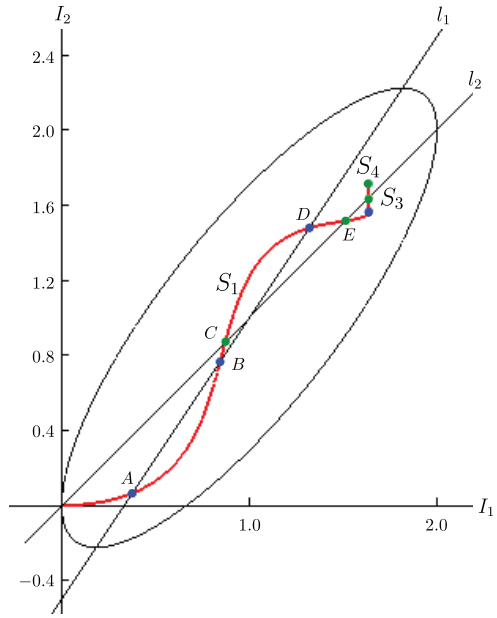


Fig. 9. Restarts of compression during restitutions of the ball–ball impact at B and the ball–table impact at D. The lower ball has mass $m_2 = 0.5$, while the other collision parameters have the same values as for the instance in Figure 8.

in the case that S_1 is the first state (with $v_1^{(0)} = -1$) or follows state S_3 with $v_1^{(0)} < 0$, or

(ii) if the accumulation of the ball–table impulse within the state satisfies

$$\Delta I_2 \le -\frac{m_2 v_2^{(0)}}{1 + \frac{k_1}{k_2}(1 + m_2)} \tag{47}$$

in the case that S_1 follows state S_2 with $v_2^{(0)} < 0$.

Note that $v_2^{(0)} = 0$ at the transition $S_3 \to S_1$ and $v_1^{(0)} = v_2^{(0)}$ at $S_2 \to S_1$. The proof of Lemma 10 is rather involved, and postponed to Appendix B, without interrupting our flow of presentation.

Proposition 11 *If the sequence $\{(I_{1i}, I_{2i})\}$ is infinite, it must converge to a point on ℓ_1 or ℓ_2 but not below $(1, 1)$.*

Proof By contradiction. Lemma 9 states that the point of convergence $q = (I_1^*, I_2^*)$ must be on either ℓ_1 or ℓ_2. Suppose that the point q lies below $p = (1, 1)$. Representing a pair of final impulse values, it would yield the final ball velocities v_1^* and v_2^*.

First, we suppose that q is on ℓ_2 below p. Then q lies to the left of ℓ_1, as shown in Figure 10(a). We have $v_1^* < v_2^* = 0$. Denote by $N(\epsilon)$ an open disk centered at q and with radius ϵ. This radius can be chosen small enough such that the disk lies completely to the left of ℓ_1. Every impulse point in N would yield $v_1 < v_2$ (cf. Figure 6(a)), and $v_1 < 0$ from (11) since $I_1 < 1$, $m_1 = 1$, and $v_0 = -1$.

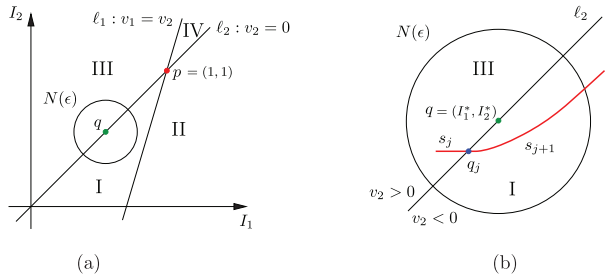


Fig. 10. Contradiction from the hypothesis that the state sequence converges to a point q on ℓ_2 below the intersection p of the compression lines ℓ_1 and ℓ_2: (a) the sequence is supposed to alternate between S_1 and S_3 inside a disk $N(\epsilon)$; meanwhile, (b) the impulse curve must exit the disk (enlarged for display).

It follows from the convergence of the sequence that there exists some integer $\sigma > 0$ such that (I_{1i}, I_{2i}) lies in N for $i > \sigma$. Because $v_1 < v_2$ for $i > \sigma$, the ball–ball contact remains in compression throughout the states $s_{\sigma+1}, s_{\sigma+2}, \ldots$. So we infer that $s_i \neq S_2$ (in which the ball–ball impact is inactive), for $i > \sigma$. From then on the state sequence must be alternating between S_1 and S_3.

Let us consider some state $s_{j+1} = S_1$, where $j > \sigma$. It begins with the impulse curve at $q_j = (I_{1j}, I_{2j})$ inside $N(\epsilon)$, as shown in Figure 10(b). The point q_j is on ℓ_2 because $v_2 = 0$ at the transition $S_3 \to S_1$. By part (i) of Lemma 10 the impulse I_1 will increase by an amount of $\Delta I_1 \ge -v_{1j} / \left(1 + \frac{1}{m_2}\left(1 + \frac{k_2}{k_1}\right)\right)$ in the state s_{j+1}, where v_{1j} is the value of v_1 as the state s_j ends at q_j. We can choose ϵ small enough such that $\Delta I_1 > 2\epsilon$ holds (this is described in the next paragraph). Consequently, $I_{1j} + \Delta I_1$ is greater than $I_1^* + \epsilon$, which bounds from above the I_1-coordinates of all points inside the disk $N(\epsilon)$. In other words, the impulse curve must exit the neighborhood within the state s_{j+1}. Hence, a contradiction.

Now we look at how to choose ϵ. The condition $\Delta I_1 > 2\epsilon$ is satisfied if

$$-\frac{v_{1j}}{1 + \frac{1}{m_2}\left(1 + \frac{k_2}{k_1}\right)} > 2\epsilon, \text{ i.e. } v_{1j} < -2\epsilon\left(1 + \frac{1}{m_2}\left(1 + \frac{k_2}{k_1}\right)\right).$$

Since $q_j \in N(\epsilon)$, $I_{1j} < I_1^* + \epsilon$, which is equivalent to $v_{1j} < v_1^* + \epsilon$ by (11) with $m_1 = 1$ and $v_0 = -1$. It suffices to make sure that $v_1^* + \epsilon < -2\epsilon\left(1 + \frac{1}{m_2}\left(1 + \frac{k_2}{k_1}\right)\right)$, which is satisfied when $\epsilon < -v_1^* / \left(3 + \frac{2}{m_2}\left(1 + \frac{k_2}{k_1}\right)\right)$.

Next, we show that the convergence point q cannot lie on ℓ_1 below p; again, by contradiction. Suppose otherwise. So q lies on ℓ_1 below p. The states will eventually be alternating between S_1 and S_2 with the impulses in some neighborhood $N(\epsilon)$ of q that is below ℓ_2. Inside $N(\epsilon)$, at a transition $S_2 \to S_1$, $v_1 = v_2 = u$ for some $u < 0$ because the impulse point is on ℓ_1 but below ℓ_2. The ball–table impact is in compression at the moment since $u < 0$. The ball–ball impact immediately goes into compression as v_2 increases to make $v_1 - v_2 < 0$. Hence, state S_1 starts with both impacts in

compression, and the impulse curve leaving the line ℓ_1 and going into region I.

By part (ii) of Lemma 10, within state S_1 the impulse I_2 will grow by an amount of at least $\Delta I_2 = -m_2 u / \left(1 + \frac{k_1}{k_2}(1 + m_2)\right)$. Again, we just need to choose ϵ small enough such that u is close enough to v_1^* to satisfy $\Delta I_2 > 2\epsilon$. Within the state the impulses (I_1, I_2) will then exit the neighborhood $N(q)$. A contradiction to the convergence at q. □

The point $(1, 1)$ represents the outcome of plastic ball–ball and ball–table impacts where both balls have zero velocities after collision. Proposition 11 therefore implies that an infinite sequence of states results in a collision outcome equivalent to that when one of the two impacts is plastic.

Finally, we combine Proposition 11 for a general collision with the results on the output velocities in special collisions treated in Jia et al. (2011a).

Theorem 12 *The state transitions will either terminate with $v_1 \geq v_2 \geq 0$ or the impulse sequence will converge to generate either $v_1 = v_2 \geq 0$ or $v_1 > v_2 = 0$.*

8. Shooting a billiard: integration of the models for multiple compliant impacts

The model described in Sections 3-7 characterizes impulses along the normals at two contacts where impacts take place simultaneously. In the presence of friction and tangential compliance, at each contact, a tangential impulse also exists and has a differential relationship to the normal impulse. Only part of the work done by the tangential force dissipates due to friction. The remaining part gets converted into strain energy to be released later. A model for 3-dimensional impact with tangential compliance has been developed by Jia (2010, 2012) to compute tangential impulse from normal impulse. Integration of the two models will equip us to solve a general impact problem involving multiple frictional and compliant contacts. In this section, we describe how such integration can be done via modeling a shot in the game of pool.

As shown in Figure 11, during a shot by the cue stick at the cue ball, the cue–ball impact and the ball–table impact happen simultaneously at the cue–ball contact point $\boldsymbol{p}$ and the ball–table contact point $\boldsymbol{q}$. Denote by $\hat{\boldsymbol{n}}$ the unit normal of the ball at $\boldsymbol{p}$, and $\hat{\boldsymbol{z}}$ that of the table at $\boldsymbol{q}$. At $\boldsymbol{p}$ two unit tangent vectors $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{w}}$ form a right-handed contact frame with $\hat{\boldsymbol{n}}$ such that $\hat{\boldsymbol{u}} \times \hat{\boldsymbol{w}} = \hat{\boldsymbol{n}}$. Similarly, at $\boldsymbol{q}$ the unit tangent vectors $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ form a right-handed contact frame with $\hat{\boldsymbol{z}}$.

Let the cue stick have mass m_c, and the cue ball have mass m and radius r, thus moment of inertia $\frac{2}{5}mr^2$. The cue stick shoots at a velocity $\boldsymbol{v}_{c0}$ in the unit direction $\hat{\boldsymbol{c}} = \boldsymbol{v}_{c0}/\|\boldsymbol{v}_{c0}\|$ that is aligned with the cue's axis of symmetry. The cue–ball impact is eccentric when $\hat{\boldsymbol{n}}$ and $\hat{\boldsymbol{c}}$ are not collinear. The condition $\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{c}} < 0$ must hold for the shot to happen.

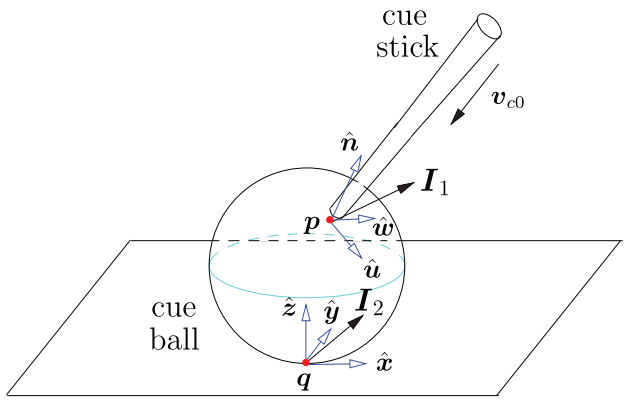


Fig. 11. Pool shot results in simultaneous cue-ball and ball–table impacts at $\boldsymbol{p}$ and $\boldsymbol{q}$.

To ensure the ball–table impact to be active during the shot, we first assume that the point $\boldsymbol{p}$ of hit is above the ball's equator (i.e. $\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{z}} > 0$) or on the equator with $\hat{\boldsymbol{c}} \cdot \hat{\boldsymbol{z}} < 0$. We also assume that during the shot, the cue stick is constrained to move along $\hat{\boldsymbol{c}}$ or $-\hat{\boldsymbol{c}}$. This way our analysis is not overcomplicated by taking into account rotation or bending of the cue stick or shearing of the cue tip during the shot, although these effects occur during a real pool shot. This second assumption is satisfied by our design of a mechanical cue stick to be described in Section 9.2.

During the shot, the cue stick has velocity $\boldsymbol{v}_c$, the ball has velocity $\boldsymbol{v}$ and angular velocity $\boldsymbol{\omega}$, all varying as the cue–ball and ball–table impacts develop. Denote by $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ the impulses exerted on the cue and the ball at $\boldsymbol{p}$ and $\boldsymbol{q}$, respectively. The velocity equations are as follows:

$$\boldsymbol{v}_c = \boldsymbol{v}_{c0} + \frac{1}{m_c}(\boldsymbol{I}_1 \cdot \hat{\boldsymbol{c}})\,\hat{\boldsymbol{c}}, \tag{48}$$

$$\boldsymbol{v} = \frac{1}{m}(\boldsymbol{I}_2 - \boldsymbol{I}_1), \tag{49}$$

$$\boldsymbol{\omega} = -\frac{5}{2mr}(\hat{\boldsymbol{n}} \times \boldsymbol{I}_1 + \hat{\boldsymbol{z}} \times \boldsymbol{I}_2). \tag{50}$$

Let $\boldsymbol{v}_1$ be the relative contact velocity of the cue stick to the ball, and $\boldsymbol{v}_2$ be that of the ball to the pool table. It follows from contact kinematics with substitutions of (48)–(50) that

$$\begin{aligned}\boldsymbol{v}_1 &= \boldsymbol{v}_c - \boldsymbol{v} - \boldsymbol{\omega} \times (r\hat{\boldsymbol{n}}) \\ &= \boldsymbol{v}_{c0} + \left(\frac{1}{m_c}\hat{\boldsymbol{c}}\hat{\boldsymbol{c}}^{\mathrm{T}} + \frac{7}{2m} - \frac{5}{2m}\hat{\boldsymbol{n}}\hat{\boldsymbol{n}}^{\mathrm{T}}\right)\boldsymbol{I}_1 \\ &\quad + \frac{1}{m}\left(-1 + \frac{5}{2}(\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{z}}) - \frac{5}{2}\hat{\boldsymbol{z}}\hat{\boldsymbol{n}}^{\mathrm{T}}\right)\boldsymbol{I}_2,\end{aligned} \tag{51}$$

$$\begin{aligned}\boldsymbol{v}_2 &= \boldsymbol{v} - \boldsymbol{\omega} \times (r\hat{\boldsymbol{z}}). \\ &= \frac{1}{m}\left(\left(-1 + \frac{5}{2}(\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{z}}) - \frac{5}{2}\hat{\boldsymbol{n}}\hat{\boldsymbol{z}}^{\mathrm{T}}\right)\boldsymbol{I}_1 + \left(\frac{7}{2} - \frac{5}{2}\hat{\boldsymbol{z}}\hat{\boldsymbol{z}}^{\mathrm{T}}\right)\boldsymbol{I}_2\right).\end{aligned} \tag{52}$$

In the above equations, scalars appearing in the sums with matrices are treated as multiples of the 3×3 identity matrix. Such a sum is always multiplied with a vector so it can be viewed as multiplications distributed into the sum.

In (48), an impulse due to the implicit constraint force, which enforces the linear motion constraint on the cue stick, cancels out the component of $\boldsymbol{I}_1$ orthogonal to $\hat{\boldsymbol{c}}$. Decompose the two impulses in their respective contact frames:

$$\begin{aligned} \boldsymbol{I}_1 &= I_u \hat{\boldsymbol{u}} + I_w \hat{\boldsymbol{w}} + I_n \hat{\boldsymbol{n}}, \\ \boldsymbol{I}_2 &= I_x \hat{\boldsymbol{x}} + I_y \hat{\boldsymbol{y}} + I_z \hat{\boldsymbol{z}}. \end{aligned} \tag{53}$$

We have $I_n > 0$ or $I_z > 0$ whenever the corresponding impact has been active, and $I_n = 0$ or $I_z = 0$ otherwise.

8.1. Embedding the compliant impact model

The tangential impulses $I_u \hat{\boldsymbol{u}} + I_w \hat{\boldsymbol{w}}$ and $I_x \hat{\boldsymbol{x}} + I_y \hat{\boldsymbol{y}}$ are due to compliance and friction and dependent on the normal impulses I_n and I_z, respectively. To model them, we use the three-dimensional contact structure from Jia (2012), which extends the planar structure introduced by Stronge (2000, pp. 95–96). As illustrated in Figure 12(a), we view the cue–ball contact $\boldsymbol{p}$ as a massless particle connected to the cue tip via three virtual springs respectively aligned with $\hat{\boldsymbol{u}}$, $\hat{\boldsymbol{w}}$, and $\hat{\boldsymbol{n}}$. For convenience, they are referred to as the u, w, and n springs, while u, v, and n also denote the changes of their lengths by a slight abuse of notation. The infinitesimal area on the cue tip attached to the springs may be thought of as 'concave'. Similarly, part (b) of the figure shows $\boldsymbol{q}$ as a massless particle connected to the cue ball via three springs in the directions of $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{z}}$. These springs are referred to as the x, y, and z springs with changes of length x, y, and z, respectively. The impulses $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ result from compressions and extensions of these six springs over the shot period.

Let k_n and k_z be the stiffnesses of the n and z springs with original values $\bar{k}_n$ and $\bar{k}_z$, respectively. The u and w springs at $\boldsymbol{p}$ have the same stiffness whose ratio to $\bar{k}_n$ is determined by the material properties of the cue tip and the ball. The x and y springs at $\boldsymbol{q}$ also have the same stiffness with a ratio to $\bar{k}_z$ determined by the material properties of the ball and the table. The two normal springs vary their lengths at the rates

$$\dot{n} = \hat{\boldsymbol{n}} \cdot \boldsymbol{v}_1 \qquad \text{and} \qquad \dot{z} = \hat{\boldsymbol{z}} \cdot \boldsymbol{v}_2. \tag{54}$$

The cue–ball and the ball–table impacts are each described by a copy of the compliant impact model (see Jia, 2012). Each copy is completely characterized by the coefficient of friction and the ratio of the normal stiffness to the tangential stiffness at the same contact. The sliding velocities of $\boldsymbol{p}$ and $\boldsymbol{q}$ in their respective contact tangent planes are (cf. Figure 12) $\boldsymbol{v}_1 - \dot{n}\hat{\boldsymbol{n}} - \dot{u}\hat{\boldsymbol{u}} - \dot{w}\hat{\boldsymbol{w}}$ and $\boldsymbol{v}_2 - \dot{z}\hat{\boldsymbol{z}} - \dot{x}\hat{\boldsymbol{x}} - \dot{y}\hat{\boldsymbol{y}}$. If one of the sliding velocities is zero, the corresponding contact sticks; otherwise, the contact slips. Mode analysis can be performed for each contact based on the energies stored by its normal and tangential springs. The rates $\dot{u}, \dot{w}, \dot{x}$, and $\dot{y}$ of changes in length of the tangential springs are updated based on the contact modes. The tangential impulse is updated as a function of the accumulating normal impulse. Details of the procedure for such update can be found in Jia (2012). The output from each compliance model includes the derivatives of the tangential impulses I_u and I_w (or, I_x and I_y) with respect to the normal impulse I_n (or I_z).

Denote by e_n and e_z the coefficients of restitution for the cue–ball and ball–table impacts, respectively. The normal stiffness k_n (or k_z) scales up by a factor of $1/e_n^2$ (or $1/e_z^2$) every time the corresponding normal impact finishes compression. Its ratio to the tangential stiffness at the contact needs to be scaled up by the same factor.

Figure 13 illustrates integration of the models for multiple impacts and compliant impact with dynamics and contact kinematics. We apply the multiple impact model to compute accumulations of the two normal impulses I_n and I_z. It is responsible for generating the state sequence. The shot by the cue stick has four states: S_1, during which both the cue–ball and the ball–table impacts are active; S_2, during which only the ball–table impact is active; S_3, during which only the cue–ball impact is active; and S_4, the final state. The shot starts with S_1 and ends with S_4. The state transition diagram has the same structure[21] as that depicted in Figure 4 except for respective replacements of $v_1 - v_2$ and v_2 with the normal velocity components $\dot{n}$ and $\dot{z}$. A cue–ball impact is in compression when $\dot{n} < 0$. A ball–table impact is in compression when $\dot{z} < 0$.

In a state, the primary impulse is the variable that drives the entire system. All tangential impulses are eventually functions of the primary impulse in the state.

8.2. Algorithm for simulating a billiard shot

Algorithm 1 simulates a billiard shot via numerical integration. On line 7, update of the strain energies E_n and E_z stored by the two normal springs proceeds according to the derivative (5) of the strain energy stored by a single impact with respect to impulse, or more specifically here,

$$\frac{dE_n}{dI_n} = -\dot{n} \quad \text{and} \quad \frac{dE_z}{dI_z} = -\dot{z}. \tag{55}$$

Line 13 tests whether the cue–ball impact is active. If so, lines 14–17 calls upon its compliant impact model to determine the contact mode and evaluate the derivatives of the two tangential impulses at the contact. Lines 21–24 handles the situation where the ball–table impact is active. Lines 7–12, 28–29, and 32 are carried out by the introduced multiple impact model.

Below we explain initialization of a state in line 3. The single-impact states S_2 and S_3 need not be initialized, because the impulse derivative dI_n/dI_z or dI_z/dI_n inherits its value from the previous state S_1, and so do the strain energies and the active contact modes.

State S_1 either is the first state or follows one of S_2 and S_3. If S_1 is the first state, dI_z/dI_n has the initial value

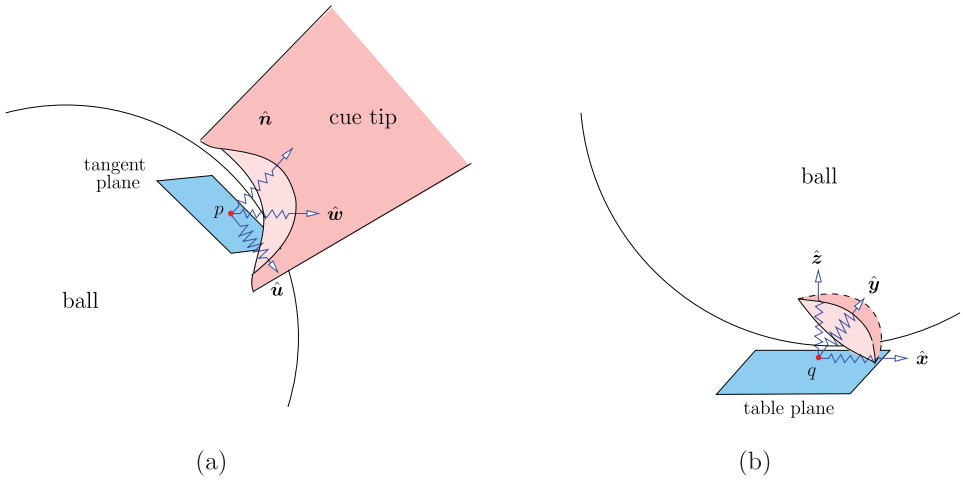


Fig. 12. Contact structures attaching massless points $\boldsymbol{p}$ and $\boldsymbol{q}$ via virtual springs to (a) the cue tip and (b) the ball, respectively.

zero by (29). Let δ be the step size of numerical integration (over either I_n or I_z) used in Algorithm 1. So we determine the value ρ of dI_z/dI_n at $I_n = \delta$. Use the compliant impact model to determine whether the initial cue–ball contact sticks or slips, and along with it, the derivative

$$\frac{d\boldsymbol{I}_1}{dI_n}(\delta) = \hat{\boldsymbol{n}} + \frac{dI_u}{dI_n}(\delta)\,\hat{\boldsymbol{u}} + \frac{dI_w}{dI_n}(\delta)\,\hat{\boldsymbol{w}}.$$

Initially, $\boldsymbol{I}_2$ is dominated by $\boldsymbol{I}_1$, and thus ignored in obtaining the ball–table contact velocity using (52). Apply the same initialization approach from Section 4.2:

$$\rho = \frac{dI_z}{dI_n}(\delta) = \sqrt{\frac{k_z}{k_n}} \cdot \sqrt{\frac{E_z(\delta)}{E_n(\delta)}},$$

where $E_n(\delta)$ and $E_z(\delta)$ are in terms of $\boldsymbol{I}_1(\delta)$ (with $\boldsymbol{I}_2(\delta)$ ignored in comparison) via one step numerical integration over I_n and I_z, respectively. This will set up a quadratic equation in ρ. Solve it for ρ. We approximate $I_z(\delta) \approx \rho I_n(\delta) = \rho\delta$ and use the compliant impact model to obtain $I_x(\delta)$ and $I_y(\delta)$, and thus $\boldsymbol{I}_2(\delta)$.

If S_1 succeeds S_3, that is, the ball–table impact is reactivated during the cue–ball impact, we can estimate dI_z/dI_n similarly as in the above case, except plugging in the values of E_n, $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, and $d\boldsymbol{I}_1/dI_n$ at the transition point

If S_1 succeeds S_2, that is, the cue–ball impact is reactivated during the ball–table impact, suppose that the transition $S_2 \to S_1$ happens at $I_z = I_z^{(0)}$. By ignoring the ball–ball impulse compared with the ball–table impulse just after S_1 starts, we can solve for dI_n/dI_z at $I_z^{(0)} + \delta$ from another quadratic equation.

9. Experiments

In this section, we first validate the ball–ball–table collision model described in Sections 3–7 using an experiment with two ping pong balls. Then we will compare measurements from a real massé shot with predictions by Algorithm 1 that integrates the multiple impact model and the compliant impact model.

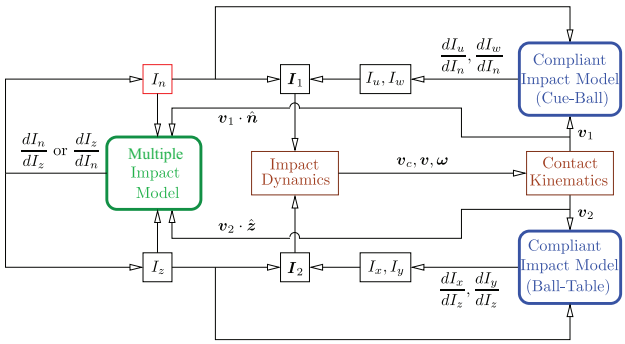


Fig. 13. Integrating the models for multiple and compliant impacts.

9.1. Collision of ping-pong balls against plexiglass surface

As shown in Figure 14(a), a ping pong ball $\mathcal{B}_1$ was dropped by hand onto another one $\mathcal{B}_2$ resting on a plexiglass block. The two balls were identical with mass 0.00023 kg and radius 0.019 m. The block was placed horizontally on the marker tray of a (vertical) office whiteboard, and against a vertical axis ℓ drawn on the board. Both balls were positioned almost in contact with the whiteboard so the line through their centers was as close to be parallel to ℓ as possible.

To measure the coefficient e_2 of restitution between a ball and the plexiglass surface, we dropped the ball from certain height h_1 onto the surface and recorded the rebounding height h_2 (on the axis by human vision). So $e_2 = \sqrt{h_2/h_1}$ after ignoring air resistance. A total of 16 measurements from different drop heights generated a mean estimate of 0.846529 with a standard deviation of 0.020827. To measure the coefficient e_1 of ball–ball restitution, both balls

Algorithm 1 Billiard shot.

```
state s ← S1
while not S4 do
    initialization
    while state s do
        update cue and ball velocities according to impact dynamics (48)–(50)
        update contact velocities according to contact kinematics (51)–(52)
        update strain energies using In and Iz according to (55)
        if s = S3 or s is the initial state or (s = S1 and previous state is S3) then
            dIz/dIn ← sqrt(kz Ez / kn En)
        else
            dIn/dIz ← sqrt(kn En / kz Ez)
        end if
        if s ≠ S2 then
            update strain energies stored by the u and w springs
            determine cue–ball contact mode
            evaluate dIu/dIn and dIw/dIn
            update Iu and Iw via one-step integration
            increment In
        end if
        if s ≠ S3 then
            update strain energies stored by the x and y springs
            determine ball–table contact mode
            evaluate dIx/dIz and dIy/dIz
            update Ix and Iy via one-step integration
            increment Iz
        end if
        update I1 and I2 according to (53)
        if any active impact ends compression or restitution then
            update impact phase
        end if
    end while
    s ← next state according to the transition diagram in Figure 4
end while
```

were hung by parallel threads at the same height (like two pendula) and a distance of twice the radius apart so the balls are in contact. Next, one ball was raised to a height of h_3 relative to the other one while its attached thread was kept straight.[22] Then it was released, colliding with the other ball, which reached a height of h_4. We obtained $e_1 = \sqrt{h_4/h_3}$. The mean value of e_1 calculated over 11 different dropping heights was 0.950043 with standard deviation 0.0125718.

We could not get a good measure of the stiffness k_1 between the two ping-pong balls or the stiffness k_2 between the ping-pong ball and the plexiglass. We know that the ratio $k_2/k_1 > 1$ since the plexiglass is considerably stiffer.

A collision trial involved dropping $\mathcal{B}_1$ from a fixed height onto $\mathcal{B}_2$ multiple times, and choosing the one with the highest rebounds of both balls.[23] The velocity v_0 of $\mathcal{B}_1$ before the collision and the ball velocities v_1 and v_2 after the collision were calculated from the heights reached by the balls.

Results from 10 trials, each with a different dropping height, are plotted in Figure 14(b) as pairs of points $(-v_{0,i}, v_{1,i})$ and $(-v_{0,i}, v_{2,i})$, $0 \le i \le 9$ for the guessed value $k_2/k_1 = 2.25$, where v_0 varies from $v_{0,0} = -2.02014$ m s^{-1} to $v_{0,9} = -3.59304$ m s^{-1}. Also shown are two lines of v_1 and v_2 predicted by the multiple impact model as v_0 varies from -2 to -3.6 m s^{-1}. (The ratios v_1/v_0 and v_2/v_0 are invariant to v_0 under Theorem 5.) In Figure 14(c), the same data points are plotted against model predictions with $k_2/k_1 = 2.7$. The model generated the same state sequence $\langle S_1, S_3, S_4 \rangle$ in all trials.

In Figure 14(b), the model predictions had the best match with the data over v_1; while in Figure 14(c), the model predictions did over v_2 (except for the two outliers $v_{2,0}$ and $v_{2,9}$). Overall, Figure 14(c) represents a better match. Despite the rather primitive setup and measurements, the results suggest a reasonably good match between our model and the physical collision process.

9.2. A massé billiard shot

To experimentally validate Algorithm 1, we designed a (manual) mechanical cue (see Figure 15(a)). It was constrained to linear motions by ball bearings embedded inside an aluminum block. A small flat aluminum piece attached to the other end of the cue stick served as a trigger which, connected to the aluminum block via two identical springs, could be pulled and then released to execute a shot on the cue ball. The block rested on an adjustable incline so the cue stick can hit the ball from a wide range of directions.

The cue together with the trigger weighed $m_c = 0.5018$ kg. The cue ball had radius $R = 0.0305$ m and weight $m = 0.01701$ kg. The two springs[24] had stiffness $k_s = 353.756$ N m^{-1} and initial stretches of $\Delta x_0 = 0.011$ m. A ruler was fastened to the back side of the aluminum block to measure their changes in length Δx after the pulling. The cue hit the ball at speed $\sqrt{2k_s(\Delta x^2 - \Delta x_0^2)/m_c}$.

The coefficient of restitution e_2 for a ball–table impact was measured by dropping a ball onto the table and recording its drop height and bounce height, similarly to the measurement method in Section 9.1 for a ping-pong ball. The average value calculated from drops at four different heights was $e_2 = 0.51625$. The coefficient of restitution e_1 between the cue and the ball was measured in a similar way by dropping the ball onto an identical cue tip resting bottom-up on a hard surface. Our estimate was 0.656532.

We taped four object balls (of the same material as the cue ball) from the top so they would slide together. Then

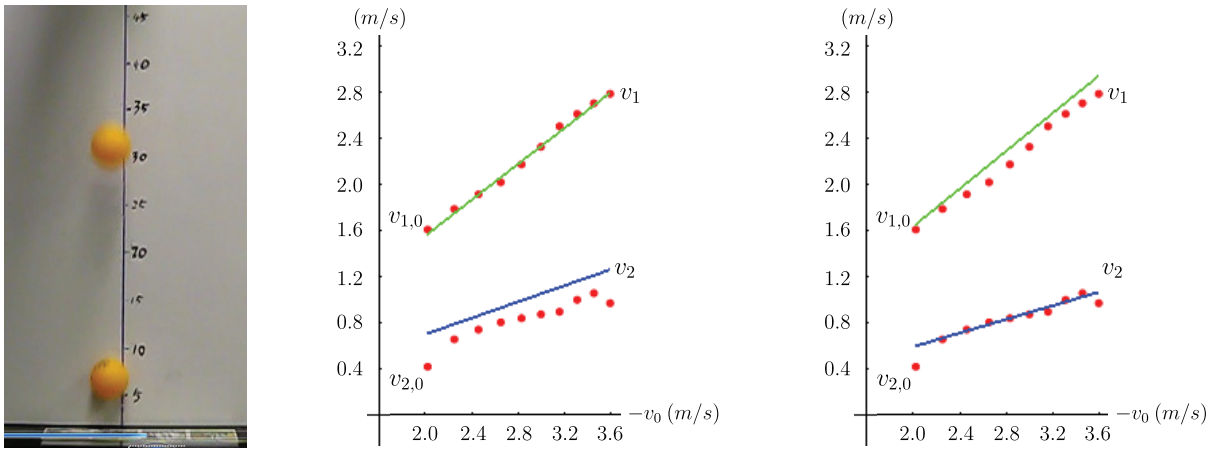


Fig. 14. (a) Collision between two ping-pong balls and a plexiglass surface (from a video capture); (b) results from 10 trials (dots) versus predictions (lines) by the impact model (with guessed value $k_2/k_1 = 2.25$); and (c) match with $k_2/k_1 = 2.7$.

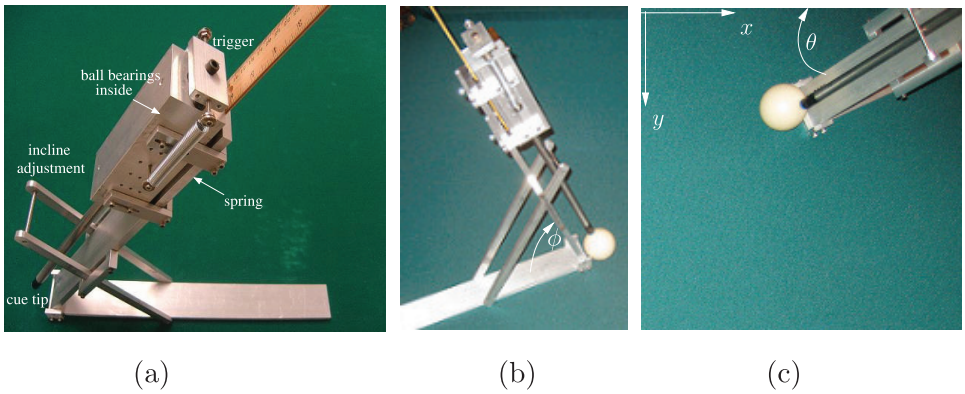


Fig. 15. (a) Adjustable billiard shooter using springs and ball bearings (embedded). Setup for a massé shot: (b) side view; and (c) top-down view.

tilted the pool table until the balls started sliding. The coefficient μ_s of sliding friction was calculated as the slope of the tilted table ($\mu_s = 0.152479$).

We let the cue ball roll down a slope of height h set up on the pool table, and measured how far it would continue on the table. The ball velocity right at the bottom of the slope was calculated from the height of the slope. Then the coefficient μ_r of rolling friction was estimated from the rolling distance d on the table due to the deceleration of $-\frac{5}{7}\mu_r g$ under friction, where $g > 0$ is the gravitational acceleration.[25] It is easy to derive that $\mu_r = 7h/(5d)$. The calculated value was 0.0209727.

The coefficient of friction between the cue and the ball can be as high as 0.7 depending on the amount of chalk (Cross, 2007). We used the value 0.7 in the experiment.

The relative stiffnesses were set as $k_2/k_1 = 0.8$, $\eta_1^2 = 7$, and $\eta_2^2 = 10$. Here, k_2/k_1 is the ratio between the normal stiffnesses at the ball–table and cue-ball contacts, η_1^2 and η_2^2 the ratios between the normal and tangential stiffnesses at the cue-ball and ball–table contacts, respectively. Unable to measure these three parameters, we had to guess their values. The felt was softer than the cue tip but not by too much because of the hard surface of the table underneath it. So we chose k_2/k_1 to be less than one but not by too much. The cue–ball contact likely had less tangential compliance than the ball–table contact, so we set $\eta_1 < \eta_2$.

A massé shot was performed. Figure 15(b) and (c) present two different views of the pre-shot configuration. The cue stick tilted at an angle of $\phi = \frac{2}{5}\pi$ with the pool table. Its projection onto the table formed an angle of $\theta = 0.465762$ with the negative x-axis which was aligned with a table edge. The cue tip hit at a location on the cue ball with measured outward normal $\boldsymbol{n} = (0.480151, -0.102889, 0.871131)$. Before the shot, the springs were stretched by $\Delta x = 0.079$ m due to the hand pull. Based on the mechanical model for the shooter,

we calculated that the cue tip hit the ball at the speed $\boldsymbol{v}_c = 3.089095\ \mathrm{m\ s^{-1}}$.

After the massé shot, the cue ball slid along a parabolic trajectory before rolling along a straight trajectory to halt. A video of the shot was taken by a Cannon PowerShot SD7901S digital camera at 30 frames/s. The video was sequenced at the same frame rate, and in each frame, the ball's center was measured manually. Figure 16(a) plots the positions (dots) of the ball's center every 1/30th of a second during the motion. Also shown are the trajectory fit over these positions along with the parabola, and the line on which the ball had moved. Note that the leftmost dot represents only the last position before the ball exited the camera's field of view. Jia et al. (2011b) presented an involved analysis for the trajectory of a billiard given its initial velocity and angular velocity, along with an algorithm to recover these two velocities given a billiard trajectory. Applying the algorithm, we estimated the velocities of the cue ball immediately after the shot to be

$$\begin{aligned} \bar{\boldsymbol{v}} &= (-1.67063, -0.274431, \_\_\_)^{\mathrm{T}}, \\ \bar{\boldsymbol{\omega}} &= (28.4584, 79.1561, \_\_\_)^{\mathrm{T}}. \end{aligned} \tag{56}$$

Their z-components, on the other hand, could not be recovered because they did not affect the trajectory.

Figure 16(b) shows the free ball trajectory (at a different scale) that would have resulted from the ball velocities (56). The actual ball trajectory ended earlier due to a collision with the table cushion.

Using the measured physical parameter values, our integrated impact model predicts the post-shot ball velocities to be

$$\begin{aligned} \tilde{\boldsymbol{v}} &= (-1.67629, -0.075349, 0.637937)^{\mathrm{T}}, \\ \tilde{\boldsymbol{\omega}} &= (40.3064, 83.6145, -15.4927)^{\mathrm{T}}. \end{aligned} \tag{57}$$

The state sequence is $\langle S_1, S_2, S_4 \rangle$. The ball trajectory that would have been generated by these velocities is shown in Figure 16(c). It looks quite similar to the reconstructed trajectory in Figure 16(b). Comparing (56) and (57), the (dominating) x-components of the two velocities are very close to each other and so are the (dominating) y-components of the two angular velocities.

Figure 16(d) displays a bundle of trajectories predicted by the combined impact model as the stiffness ratio η_1^2 at the cue–ball contact increases discretely from 2 to 7. Accordingly, the tangential stiffness decreases, or equivalently, the tangential compliance increases at the contact. With small compliance ($\eta_1^2 = 2$ or 3), the ball would slide with a right turn, and the resulting trajectory would differ significantly from the one captured by the camera shown in Figure 16(a). When η_1^2 is 4 or above, the ball would slide to the left, and its trajectory would look increasingly like the real one. This shows the role of tangential compliance in modeling the actual parabolic trajectory generated by the massé shot. The dependence of a massé shot on large compliance is consistent with our practice of putting chalk on the cue tip before this type of shot.

In order to generate a parabolic trajectory as shown in the video, there has to be an enough amount of tangential impulse to impart the spinning of the ball about its traveling direction. Such an amount could not be provided by contact friction alone or generated under Poisson's hypothesis.[26]

10. Discussion and future work

The introduced state transition diagram breaks down a three-body frictionless collision with two contact points into a state sequence generated from four states (with repeats). A transition from one state to another happens when either an impact finishes restitution or an inactive contact is reactivated because the two engaged objects are approaching each other again.

Every contact is associated with an energetic coefficient of restitution. The impact at this contact is related to that at another contact differentially according to their relative stiffness and the ratio between their stored strain energies. As a result, the impact may go through multiple rounds of compression and restitution. The above aspects are the same as those of the impact model introduced by Liu et al. (2008, 2009), although impulse growth was not characterized in that work. Such characterization becomes prominent for detecting contact modes (slip and stick) in the presence of friction and compliance so the whole physical process can be accurately described. Another distinction is that our model increases the contact stiffness by a factor of $1/e^2$ every time compression ends. In addition, we have shown that the growth of impulse is along a bounded curve that is first-order continuous, and proved that any state sequence either terminates or converges.

The same ball–ball–table collision problem was also treated by Glocker and Pfeiffer (1995) based on the Poisson restitution under a linear complementarity formulation. The result was, however, not quite satisfactory. The ball–ball impulse had to be enlarged in an ad hoc fashion to prevent ball penetration. Simulation yielded an unnatural outcome where the two balls would bounce up in contact if the value of the coefficient of restitution did not exceed 0.5.

Although the multiple impact model in Sections 3–7 considers central impacts, Section 8 demonstrates that it is readily applicable to an eccentric impact. Objects of arbitrary geometry complicate only the dynamics and contact kinematics that set up the impact equation (3), but not modeling of multiple concurrent impacts itself.

For $n \geq 3$ contact points, states can still be defined according to which impacts are instantaneously active and which are not. Although there could be up to 2^n states, many of them may not occur in the collision. We can generate the transition diagram on the fly, starting with the initial state where all contacts are active. A transition leads to either

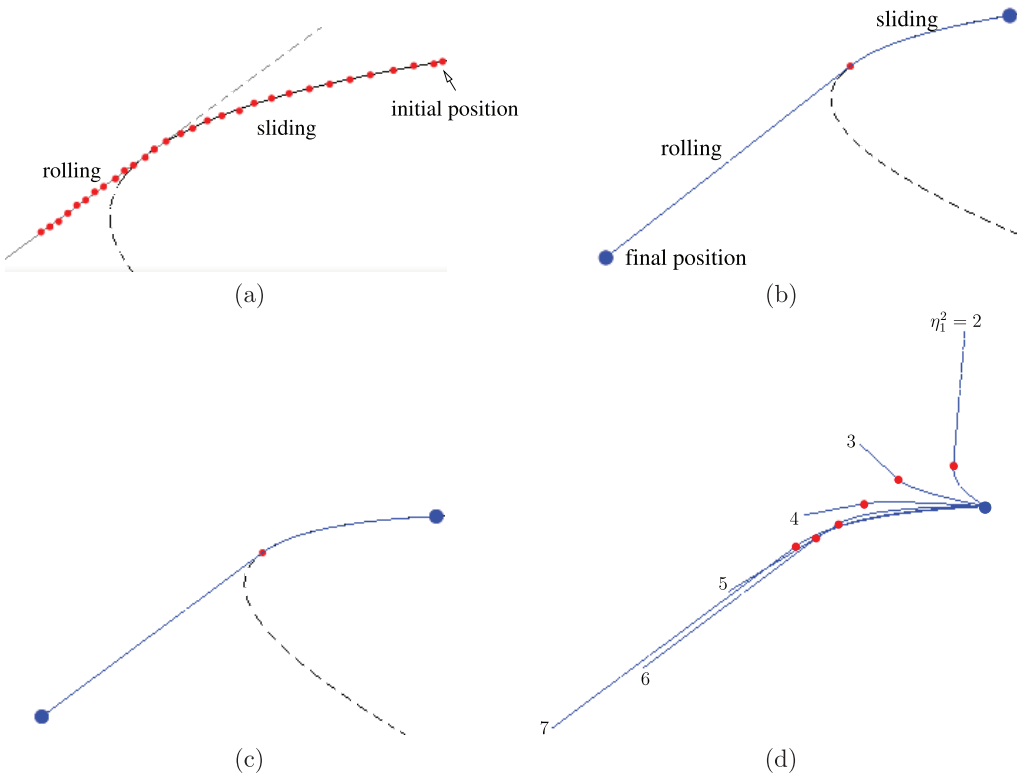


Fig. 16. Billiard trajectories related to the massé shot in Figure 15: (a) recovered trajectory via fitting over ball positions sampled at 30 Hz from the shot video; (b) trajectory extended from (a) were it not ended by the cushion on the pool table; (c) predicted trajectory by the integrated impact model ($\eta_1^2 = 7$); and (d) a bundle of trajectories as the cue–ball compliance increases with η_1.

a new state or to an already generated state. Expansion of the transition diagram stops as soon as the end state with no active contacts is reached. The impulse curve will be n-dimensional.

Algorithm 2 describes how the multiple-impact model, extended for more than two contacts, can be combined with multiple copies of the compliant impact model to solve a general simultaneous multibody collision problem. At a contact, the normal contact velocity is used for updating the strain energy stored by the normal spring, just like in (55).

Further investigation needs to be conducted to abstract out state evolution and transition templates, and to find the range for the number of contacts to which this model is applicable in practice. Measuring relative contact stiffness is important for real application of the model.[27] Upgrade and better design of hardware will improve accuracy in validation. Other issues worthy of exploration include area contact and gravity (when impact has a non-negligible duration).

If the time periods of multiple impacts have little overlaps so their correlations are weak, we may opt for more efficient methods such as impact sequencing (Chatterjee and Ruina, 1988). If the impacts are synchronized on compression and restitution, we may use LCP (Glocker and Pfeiffer, 1995; Stewart and Trinkle, 1996).

Since a frictionless contact force acts in the direction of the outward normal, this imposes a unilateral constraint during multiple impacts. In the control theory, it is known that a system of differential equations with unilateral constraints could be very sensitive to small perturbations in the input and round-off errors in numerical simulation, and have discontinuity in its solution.[28] Our impact system fits in this category. In the future, we hope to understand more about its sensitivities to physical parameters including mass, coefficients of friction and restitution, stiffness ratios, and the configuration of collision.

A longer-term objective is to design a robot able to play billiards with human-level skills. To the best of the authors' knowledge, none of the developed systems (Moore et al., 1995; Long et al., 2004; Ho et al., 2007) execute shots by exploiting the underlying mechanics, or exhibit human-like shooting skills.

The state transition diagram adds a flavor of computer science to impact problems. The model is applicable to manipulation and mobile robot tasks, in situations where multiple collisions occur. It also has potential impact in computer graphics on dynamic simulation of collisions over which little physics-based work (Baraff, 1993; Mirtich and Canny, 1995) is known.

Algorithm 2 Frictional multibody collision with tangential compliance.

```
state s ← S1
while not end of collision do
    initialization
    while state s do
        update object and contact velocities based on dynamics and contact kinematics
        for each contact do
            update the strain energy stored by the normal spring
            update the strain energies stored by the two tangential springs
            determine the contact mode using the above three strain energies
            update normal and tangential impulses
            if the impact ends compression or restitution then
                update the impact phase
            end if
        end for
        increment the primary normal impulse
    end while
    s ← next state according to the (extended) multiple impact model
end while
```

Notes

1. In addition to the aforementioned work, only Izumi and Kitaka (1993) and Tagawa et al. (2010) to our knowledge.
2. Values are properly set to physical parameters including the ball masses, the contact stiffnesses, and the coefficients of restitution of the ball–ball and ball–table impacts.
3. This model is investigated in more detail in Jia (2012).
4. A central impact between two bodies happens when their centers of mass lie on the common normal line passing through the contact. The impact is eccentric if one of the centers does not lie on the line.
5. Compliance allows part of the work done by the tangential reaction force to be recoverable rather than dissipate entirely due to friction.
6. Poisson's hypothesis states that I will accumulate $-\tilde{e}mv_0$ more during restitution to yield the final velocity $-\tilde{e}v_0$, where $\tilde{e} \in [0, 1]$ is called the kinetic coefficient of friction. When friction exists at the contact and the direction of contact slip varies during collision, the kinetic coefficient of restitution $\tilde{e}$ is not consistent with energy conservation (Stronge, 2000, p. 47). The two coefficients of restitution $\tilde{e}$ and e are otherwise equivalent for single impact unless the bodies in impact are rough and the impact configuration is eccentric (Stronge, 2000, p. 28).
7. This was initially due to Boulanger (1939).
8. An alternative is a mono-stiffness model (Goldsmith, 1960) that does not change the stiffness at the end of compression but rather ends restitution at the energy value $E = (1 - e^2)E_{\max}$. This approach, used by both Liu et al. (2008) and Jia et al. (2010), is able to produce cycles of compression and restitution that have been experimentally validated (Antonyuk et al., 2010). In our paper, the increase in stiffness after compression reflects the hardening of material better as well as makes it convenient to model transition of compression back to restitution, a phenomenon remarked on by Liu et al. (2008) and described later in Section 3.2.
9. Such material deformation is assumed here to happen much faster than impact.
10. A different bi-stiffness model (Zhao et al., 2008; Nguyen and Brogliato, 2012) subtracts from the change in length at the end of compression a portion of $1 - e^2$ due to plastic deformation.
11. We abuse the meaning of 'derivative' at the end of compression to refer to the equal left and right derivatives of the energy E_i. The same convention will apply to its second derivative (which is continuous according to Proposition 6).
12. The ratio was called by Liu et al. (2008) as the distributing law and presented for nonlinear contact as well.
13. Section 5 will show that the collision outcome is affected by the stiffness ratio k_2/k_1 but not individual stiffness values.
14. This will be formally introduced as the impulse curve in Section 6.
15. Such a phenomenon was first observed by Liu et al. (2008).
16. In such a degenerate case, the problem can be simplified with a closed-form solution. For example, if $k_2/k_1 = \infty$ (infinitely stiffer ball–table contact than ball–ball contact), the lower ball can be treated as an integral part of the table so the collision reduces to single impact.
17. See http://mathworld.wolfram.com/OrdinaryDifferentialEquation.html.
18. For numerical stability, it is better to choose I_1 as the variable whenever $k_1E_1 \geq k_2E_2$ and I_2 as the variable otherwise. In Liu et al. (2008), $I_j, j = 1, 2$, is chosen if E_j is the larger one of the two energies.
19. We are assuming $e_1, e_2 \neq 0$ through Section 6. Otherwise, state S_4 needs to handle the special case where the ball–ball impact is plastic ($e_1 = 0$) and ends in S_3 Jia et al. (cf. 2011a, Section E.3). The impulse curve has a continuous tangent during S_4.
20. Owing to a round-off error in numerical integration, a slight discrepancy of 0.0001 can be seen in the total energy column of Table 1 between the second S_1 row and the following S_3 row.
21. The diagram now is even simplified because the coefficients of restitution for both impacts are non-zero (and the contact stiffnesses are finite and non-zero).
22. For more accuracy, the horizontal distance of each ball at its maximum height to its resting position was measured, and the height was calculated from this distance.
23. Thus, the balls were the closest to be vertically aligned before the collision.
24. Model number 9564K376 from McMaster-Carr, Inc.
25. The deceleration of a rolling ball is derived as equation (16) in Jia et al. (2011b).
26. From the first author's own failed attempts between 2007 and 2011.
27. A formula for (nonlinear) Hertz contact is known (Zhao et al., 2008; Nguyen and Brogliato, 2012).
28. Brach (2007) used the design of experiments method to examine the sensitivity of planar impact (particularly in a two-vehicle collision) to the objects' physical properties, the collision configuration, and the coefficient of restitution.

Funding

The work was sponsored in part by Carnegie Mellon University and Iowa State University, and in part by DARPA (contract number HR0011-07-1-0002). This work does not necessarily reflect the position or the policy of the U.S. Government. No official endorsement should be inferred.

Acknowledgments

This work began during the first author's 6-month sabbatical visit at Carnegie Mellon University (CMU) in 2007, and has since continued primarily at Iowa State University and in China (during summer breaks of the first author under self-support). Earlier progress (Jia et al., 2010) was reported at the Eighth International Workshop on Algorithmic Foundations of Robotics (WAFR). The authors are grateful to Amir Degani (now at Technion) and Ben Brown (at CMU) for their generous help in the design of the billiard shooting mechanism. They very much appreciate the valuable and constructive comments from the anonymous reviewers of IJRR that have helped improve the paper's presentation significantly. Thanks also go to the anonymous WAFR reviewers for their earlier comments. The first author is also grateful for the help from his Ph.D. student Feng Guo in performing measurements that were used for calculating the coefficient of restitution for two colliding ping-pong balls.

References

Acary V and Brogliato B (2003) Concurrent multiple impacts modeling: Case study of a 3-ball chain. In Bathe KJ (ed.), *Proceedings of the MIT Conference on Computational Fluid and Solid Mechanics*. Amsterdam: Elsevier Science, pp. 1836–1894.

Acary V and Taha D-E (2005) Concurrent multiple impacts in rigid bodies: formulation and simulation. In Van Campen DH (ed.), *Proceedings of the Fifth Euromech Nonlinear Dynamics Conference*. Eindhoven: Springer-Verlag.

Anitescu M and Porta FA (1997) Formulating dynamic multi-rigid-body contact problems with friction as solvable linear complementarity problems. *Nonlinear Dynamics* 14: 231–247.

Anitescu M, Cremer JF and Potra FA (1996) Formulating three-dimensional contact dynamics problems. *Mechanics of Structures and Machines* 24: 405–437.

Antonyuk S, Heinrich S, Tomas J, Deen NG and van Buijtenen MS (2010) Energy absorption during compression and impact of dry elastic–plastic spherical granules. *Granular Matter* 12: 15–47.

Baraff D (1993) Issues in computing contact forces for non-penetrating rigid bodies. *Algorithmica* 8: 292–352.

Bernoulli J (1969–1993) *Die Werk von Jakob Bernoulli*. Basel: Birkhäuser.

Boothroyd G and Redford AH (1968) *Mechanized Assembly: Fundamentals of Parts Feeding, Orientation, and Mechanized Assembly*. London: McGraw-Hill Inc.

Boulanger G (1939) Sur le choc avec frottement des corps non parfaitement élastiques. *Revue Scientifique* 77: 325–327.

Brach RM (2007) Design of experiments and parametric sensitivity of planar impact mechanics. In: *Vereinigung fr Unfallforschung und Unfallanalyse (EVU) – Conference Uncertainty in Reconstruction of Road Accidents*, Krakow, Poland. http://www.brachengineering.com/publications/DOE_sensitivity_PIM.pdf.

Brogliato B (1999) *Nonsmooth Mechanics*, 2nd edition. New York: Springer.

Brogliato B and Zavalo-Rio A (2000) On the control of complementary slackness juggling mechanical systems. *IEEE Transactions on Automatic Control* 45: 235–246.

Ceanga V and Hurmuzlu Y (2001) A new look at an old problem: Newton's cradle. *Journal of Applied Mechanics* 68: 575–583

Chatterjee A and Ruina A (1988) A new algebraic rigid-body collision law based on impulse space considerations. *Journal of Applied Mechanics* 65: 939–951.

Cross R (2007) Billiards. http://physics.usyd.edu.au/∼cross/Billiards.htm.

Falcon E, Laroche C, Fauve S, and Coste S (1998) Collision of a 1-D column of beads with a wall. *European Physics Journal B* 5: 111–131.

García E and de Santos PG (2005) An improved energy stability margin for walking machines. *Robotica* 23: 13–20.

Goldsmith W (1960) *Impact: The Theory and Physical Behaviour of Colliding Solids*. London: Edward Arnold Ltd.

Glocker G (2001) On frictionless impact models in rigid-body systems. *Philosophical Transactions of the Royal Society A* 359: 2385–2404.

Glocker G and Pfeiffer F (1995) Multiple impacts with friction in rigid multibody systems. *Nonlinear Dynamics* 7: 471–497.

Han I and Park S-U (2001) Impulsive motion planning for positioning and orienting a polygonal part. *The International Journal of Robotics Research* 20: 249–262.

Harris JW and Stocker H (1998) *Handbook of Mathematics and Computational Science*. Berlin: Springer-Verlag.

Higuchi T (1985) Application of electromagnetic impulsive force to precise positioning tools in robot systems. In: Faugeras O and Giralt G (eds.), *Robotics Research: The Second International Symposium*. Cambridge, MA: The MIT Press, pp. 281–285.

Hirai S, Niwa M and Kawamura S (1999) Development of impulsive object sorting device with air floating. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 3065–3070.

Ho KHL, Martin T and Baldwin J (2007) Snooker robot player — 20 years on. In: *Proceedings of the IEEE Symposium on Computational Intelligence and Games*, pp. 1–8.

Huang WH and Mason MT (2000) Mechanics, planning, and control for tapping. *The International Journal of Robotics Research* 19: 883–894.

Ivanov AP (1995) On multiple impact. *Journal of Applied Mathematics and Mechanics* 59: 887–902.

Izumi T and Kitaka Y (1993) Control of a hitting velocity and direction for a hammering robot using a flexible link. *Journal of the RSJ* 11: 436-443 (in Japanese).

Jia Y-B (2010) Energy-based motion of tangential compliance in 3-dimensional impact. In: Hsu D et al. (eds.), *Algorithmic Foundations of Robotics IX*. Berlin: Springer-Verlag, pp. 267–284.

Jia Y-B (2012) Three-dimensional impact: energy-based modeling of tangential compliance. *International Journal of Robotics Research*, 32: 56–83.

Jia Y-B, Mason MT and Erdmann MA (2010) A state transition diagram for simultaneous collisions with application in billiard shooting. In: Chirikjian GS et al. (eds.), *Algorithmic Foundations of Robotics VIII*. Berlin: Springer-Verlag, pp. 135–150, 2010. (Also presented at the Eighth International Workshop on Algorithmic Foundations of Robotics, Guanajuato, Mexico, 2008.)

Jia Y-B, Mason MT and Erdmann MA (2011a) Supplementary material for *Multiple impacts: a state transition diagram approach*. http://www.cs.iastate.edu/∼jia/papers/multi-impacts-supplement.pdf.

Jia Y-B, Mason MT and Erdmann MA (2011b) Trajectory of a billiard ball and recovery of its initial velocities. http://www.cs.iastate.edu/∼jia/papers/billiard-analysis.pdf.

Kaplan W (1991) *Advanced Calculus*, 4th edition. Reading, MA: Addison-Wesley.

Konno A, Myojin T, Matsumoto T, Tsujita T and Uchiyama M (2011) An impact dynamics model and sequential optimization to generate impact motions for a humanoid robot. *The International Journal of Robotics Research* 30: 1596–1608.

Liu C, Zhao Z and Brogliato B (2008) Frictionless multiple impacts in multibody systems. I. Theoretical framework. *Proceedings of the Royal Society of London A* 464: 3193–3211.

Liu C, Zhao Z and Brogliato B (2009) Frictionless multiple impacts in multibody systems. II. Numerical algorithm and simulation results. *Proceedings of the Royal Society of London A* 465: 1–23.

Long F, Herland J, Tessier M-C, et al. (2004) Robotic pool: an experiment in automatic potting. In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2520–2525.

Lynch K and Black CK (2001) Recurrence, controllability, and stabilization of juggling. *IEEE Transactions on Robotics and Automation* 17: 113–124.

MacLaurin C (1742) *A Treatise on Fluxions*. Edinburgh: Ruddimans.

Mason MT (2001) *Mechanics of Robotic Manipulation*. Cambridge, MA: The MIT Press.

Mirtich B and Canny J (1995) Impulse-based simulation of rigid bodies. In: *Proceedings of the Symposium on Interactive 3D Graphics*, Monterrey, CA, pp. 181–188.

Moore AW, Hill DJ and Johnson MP (1995) An empirical investigation of brute force to choose features, smoothers and function approximators. In Hanson SJ et al. (eds.), *Computational Learning Theory and Natural Learning*. Cambridge, MA: The MIT Press, pp. 361–379.

Nakagawa M, Agui JH, Wu DT and Extramiana DV (2003) Impulse dispersion in a tapered granular chain. *Granular Matter* 4: 167–174.

Nguyen N-S and Brogliato B (2012) Shock dynamics in granular chains: numerical simulations and comparison with experimental tests. *Granular Matter* 14: 341–362.

Partridge CB and Spong MW (1999) Control of planar rigid body sliding with impacts and friction. *The International Journal of Robotics Research* 19: 336–348.

Raibert M (1986) *Legged Robots That Balance*. Cambridge, MA: The MIT Press.

Rizzi AA and Koditschek DE (1992) Progress in spatial robot juggling. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 775–780.

Routh EJ (1905) *Dynamics of a System of Rigid Bodies*. London: MacMillan and Co.

Seghete V and Murphey T (2010) Variational solutions to simultaneous collisions between multiple rigid bodies. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2731–2738.

Shamos MI (1993) *The Illustrated Encyclopedia of Billiards*. Lyons and Burford Publishers.

Stewart DE (2000) Rigid-body dynamics with friction and impact. *SIAM Review*, 42(1): 3–39.

Stewart DE and Trinkle JC (1996) An implicit time-stepping scheme for rigid body dynamics with inelastic collisions and Coulomb friction. *International Journal for Numerical Methods in Engineering* 39: 2673–2691.

Stronge WJ (1990) Rigid body collisions with friction. *Proceedings of the Royal Society of London A* 431: 168–181.

Stronge WJ (2000) *Impact Mechanics*. Cambridge: Cambridge University Press.

Tagawa K, Hirota K and Hirose M (2010) Manipulation of dynamically deformable object using impulse-based approach. In Zadeh MH (ed.), *Advances in Haptics*, pp. 16–33. New York, NY: InTech.

Volpe R and Khosla P (1993) A theoretical and experimental investigation of impact control for manipulators. *International Journal of Robotics Research* 12: 351–365.

Walker ID (1994) Impact configurations and measures for kinematically redundant and multiple armed robot system. *IEEE Transactions on Robotics and Automation* 10: 670–683.

Wang Y-T, Kumar V and A Jacob (1992) Dynamics of rigid bodies undergoing multiple frictional contacts. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2764–2769.

Yoshida K and Nenchev DN (1995) Space robot impact analysis and satellite-base impulse minimization using reaction nullspace. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1271–1277.

Zavalo-Rio A and Brogliato B (1999) On the control of a one degree-of-freedom juggling robot. *Dynamics and Control* 9: 67–90.

Zhao Z, Liu C and Brogliato B (2008) Energy dissipation and dispersion effects in granular media. *Physical Review E* 78: 031307.

Zheng YF and Hemami H (1985) Mathematical modeling of a robot collision with its environment. *Journal of Robotic Systems* 2: 289–307.

Appendix A: Geometry of the impulse bounding ellipse

We rewrite the ellipse equation (42) into the standard form of a conic:

$$aI_1^2 + 2bI_1I_2 + cI_2^2 + 2dI_1 + 2fI_2 + h = 0,$$

where

$$a = \frac{1}{2} + \frac{1}{2m_2}, \quad b = -\frac{1}{2m_2}, \quad c = \frac{1}{2m_2},$$
$$d = -\frac{1}{2}, \quad \text{and} \quad f = h = 0.$$

To determine the type of conic, let us evaluate the following determinant:

$$A = \begin{vmatrix} a & b & d \\ b & c & f \\ d & f & h \end{vmatrix} = -\frac{1}{8m_2}.$$

Since $A \neq 0$, the curve is of second order. Next, we obtain

$$B = ac - b^2 = \left(\frac{1}{2} + \frac{1}{2m_2}\right)\frac{1}{2m_2} - \left(\frac{1}{2m_2}\right)^2 = \frac{1}{4m_2} > 0.$$

Thus, the curve is an ellipse, which has the canonical form (Harris and Stocker, 1998, p. 394):

$$\frac{u^2}{\left(2 + m_2 - \sqrt{4 + m_2^2}\right)/2} + \frac{w^2}{\left(2 + m_2 + \sqrt{4 + m_2^2}\right)/2} = 1. \tag{58}$$

The positive square roots of the above two terms are respectively the semi-minor and semi-major axes of the ellipse.

To determine the orientation of the ellipse, suppose a rotation of the I_1–I_2 coordinate frame through θ would result in an x–y coordinate frame whose two axes are parallel to those of the ellipse. We have $I_1 = x\cos\theta - y\sin\theta$ and $I_2 = x\sin\theta + y\cos\theta$. Substitute the above expressions into the ellipse equation (42). The quadratic term xy must vanish, that is, its coefficient $\frac{1}{2}\sin(2\theta) + \frac{1}{m_2}\cos(2\theta) = 0$. This yields $\tan(2\theta) = -\frac{2}{m_2}$. Thus, the ellipse in the impulse plane has undergone a rotation given in (43).

Next, we determine the center (ξ, η) of the ellipse in the impulse plane. If we replace I_1 with $I_1 - \xi$ and I_2 with $I_2 - \eta$ in the ellipse equation (42), the coefficients of I_1 and I_2 in the new equation must vanish. This results in two conditions:

$$2\left(\frac{1}{2} + \frac{1}{2m_2}\right)\xi - \frac{1}{m_2}\eta - 1 = 0,$$
$$\frac{1}{m_2}(\eta - \xi) = 0.$$

The linear system has solution $\xi = \eta = 1$. Hence, the ellipse is centered at $(1, 1)$. See Figure 5.

In (42), $I_1 \geq 0$ following that the first two terms are non-negative. Since $(I_1 - I_2)^2 \geq 0$, we also infer from the same equation $\frac{1}{2}I_1^2 - I_1 \leq 0$, which implies that $I_1 \leq 2$. So the ellipse is defined over $[0, 2]$. The two points $(0, 0)$ and $(2, 2)$ are the points of tangency of the vertical lines $I_1 = 0$ and $I_1 = 2$ to the ellipse, respectively.

To find out the range of I_2 on the ellipse, we view I_2 as an implicit function of I_1 defined by (42). Differentiating the equation with respect to I_1 gives us

$$\frac{dI_2}{dI_1} = 1 + \frac{m_2(I_1 - 1)}{I_1 - I_2}.$$

At the highest and lowest points of the ellipse, the above derivative vanishes, yield $I_2 = I_1(1 + m_2) - m_2$. Substitute the above into (42), obtaining two roots: $I_1 = 1 \pm \frac{1}{\sqrt{1+m_2}}$. Hence, the corresponding values: $I_2 = 1 \pm \sqrt{1 + m_2}$. They locate two horizontal tangent lines of the ellipse.

Appendix B: Proof of Lemma 10 on minimum impulse growth in state S_1

This appendix proves the lower bounds for minimum accumulation of the ball–ball impulse or the ball–table impact in state S_1 when it starts with one ball at some negative velocity and the other at some non-positive velocity.

Proof There are three cases: (a) S_1 begins the collision; (b) S_1 starts after S_3 with upper ball velocity $v_1^{(0)} < 0$; and (c) S_1 starts after S_2 with lower ball velocity $v_2^{(0)} < 0$. We will establish that both impacts are in compression under condition (46) in cases (a) and (b), and under condition (47) in case (c). Consequently, S_1 cannot end because it must end with restitution according to the transition diagram in Figure 4.

Case (a) Note that $v_1^{(0)} = -1$, $\Delta I_1 = I_1$, and $\Delta I_2 = I_2$. So condition (46) becomes

$$I_1 \leq \frac{1}{1 + \frac{1}{m_2}\left(1 + \frac{k_2}{k_1}\right)}. \tag{59}$$

The first end of compression of any impact happens in the start state S_1 when the impulse curve crosses either of the two compression lines ℓ_1 and ℓ_2. Within the state, the ball–ball impulse I_1 is the variable while the ball–table impulse I_2 and the strain energies E_1 and E_2 stored at the two contacts are functions. The state begins with $I_1 = I_2 = 0$ and $E_1 = E_2 = 0$. Equations (13) and (14) give us the initial energy derivatives:

$$\left.\frac{dE_1}{dI_1}\right|_{I_1=0} = 1 \quad \text{and} \quad \left.\frac{dE_2}{dI_2}\right|_{I_1=0} = 0. \tag{60}$$

Thus,

$$\left.\frac{dE_2}{dI_1}\right|_{I_1=0} = \left.\frac{dE_2}{dI_2}\right|_{I_1=0} \cdot \left.\frac{dI_2}{dI_1}\right|_{I_1=0} = 0. \tag{61}$$

Since the compression line ℓ_1 defined by (16) has positive slope $1 + m_2$ (see Figure 6(a)), the point of crossing by the impulse curve must have its abscissa greater than $m_2/(1 + m_2)$, the abscissa of the intersection of ℓ_1 and the I_1-axis. Thus, the ball–ball impact is during compression if

$$I_1 \leq \frac{m_2}{1 + m_2}. \tag{62}$$

Compression of the ball–table impact ends (before a state transition) only if $dE_2/dI_2 = 0$ or, equivalently, $I_1 = I_2$ after collision starts. Since dI_2/dI_1 has initial value zero by (29), the condition $I_1 = I_2$ will not be satisfied until after $dI_2/dI_1 = 1$. The latter condition, according to (15), is equivalent to $k_1E_1 = k_2E_2$ or $E_1 = (k_2/k_1)E_2$. Since

$dE_1/dI_1 > (k_2/k_1)\, dE_2/dI_1$ at $I_1 = 0$ by (60) and (61), $E_1 > (k_2/k_1)\, E_2$ will remain true before

$$\frac{dE_1}{dI_1} = \frac{k_2}{k_1} \cdot \frac{dE_2}{dI_1} = \frac{k_2}{k_1} \cdot \frac{dE_2}{dI_2} \cdot \frac{dI_2}{dI_1}.$$

Substitutions of (13)–(15) along with $v_0 = -1$ and $m_1 = 1$ into the above yields

$$\begin{aligned} -\left(-1 + \left(1 + \frac{1}{m_2}\right) I_1 - \frac{1}{m_2} I_2\right) &= \frac{k_2}{k_1} \cdot \frac{I_1 - I_2}{m_2} \\ &\cdot \sqrt{\frac{k_2}{k_1}} \cdot \sqrt{\frac{E_2}{E_1}} \\ &= \frac{1}{m_2} \cdot \left(\frac{k_2}{k_1}\right)^{3/2} \cdot \sqrt{\frac{E_2}{E_1}} \cdot (I_1 - I_2). \end{aligned}$$

Solving for I_1 in terms of I_2 and then setting $I_2 = 0$, we infer that $\frac{dE_1}{dI_1} > \frac{k_2}{k_1} \cdot \frac{dE_2}{dI_1}$ holds when

$$I_1 \le 1 \Big/ \left(1 + \frac{1}{m_2} + \frac{(k_2/k_1)^{3/2}}{m_2} \sqrt{\frac{E_2}{E_1}}\right). \tag{63}$$

When condition (63) holds, $E_2/E_1 \le k_1/k_2$, which implies

$$1 \Big/ \left(1 + \frac{1}{m_2} + \frac{(k_2/k_1)^{3/2}}{m_2} \sqrt{\frac{E_2}{E_1}}\right) \ge 1 \Big/ \left(1 + \frac{1}{m_2} + \frac{k_2}{m_2 k_1}\right).$$

Thus, (59) is a stronger sufficient condition than (63).

Tracing back the reasoning steps, we have right after the collision starts:

$$\begin{aligned} (59) \Rightarrow (63) \;\Rightarrow\; & k_2 \frac{dE_2}{dI_1} < k_1 \frac{dE_1}{dI_1} \\ \Rightarrow\; & k_2 E_2 < k_1 E_1 \;\Rightarrow\; \frac{dI_2}{dI_1} < 1 \\ \Rightarrow\; & I_2 < I_1 \;\Rightarrow\; \frac{dE_2}{dI_2} > 0 \qquad \text{by (14)} \\ \Rightarrow\; & \text{the ball–table impact in compression.} \end{aligned}$$

Since the inequality (59) also implies inequality (62), we conclude that restitution of neither impact could start as long as (59) holds. The state does not end.

Case (b) State S_1 starts right after state S_3 ends. The ball–ball impact has not finished restitution while the ball–table impact is just starting compression. So $E_1 > 0$, $E_2 = 0$, and $\Delta I_1 = \Delta I_2 = 0$. Since $v_1^{(0)} < 0$ (as stated in the lemma) and $v_2^{(0)} = 0$, the rate of change in length of the ball–ball spring is $\dot{x}_1 < 0$ by (7). Hence, the ball–ball impact is in compression. We substitute the impulse and energy values into the energy derivatives (20) and (22) because of compression, obtaining the initial energy derivatives

$$\left.\frac{dE_1}{dI_1}\right|_{I_1^{(0)}} = -v_1^{(0)} \quad \text{and} \quad \left.\frac{dE_2}{dI_2}\right|_{I_2^{(0)}} = 0.$$

Since the current state starts with $E_1 > E_2 = 0$, $k_2(dE_2/dI_1) < k_1(dE_1/dI_1)$ still implies $k_2 E_2 < k_1 E_1$. The rest of the proof parallels that of case (a), where the inference steps near the end would carry over after replacement of condition (59) with condition (46) in the first step.

Case (c) State S_1 starts right after state S_2 ends. The condition $v_2^{(0)} < 0$ implies that the ball–table impact has not finished restitution. The ball–ball impact is starting compression. At the transition, $E_2 > 0, E_1 = 0$, and $\Delta I_1 = \Delta I_2 = 0$. Also, $v_1^{(0)} = v_2^{(0)} = h$ for some $h < 0$ at the transition $S_2 \to S_1$. Clearly, the ball–table impact is in compression. Since v_2 is increasing, the ball–ball impact goes into compression immediately. Substituting (18) and (19) into (7) and (8), we have the rates of length changes of the two virtual springs:

$$\begin{aligned} \dot{x}_1 &= \left(1 + \frac{1}{m_2}\right) \Delta I_1 - \frac{1}{m_2} \Delta I_2, \\ \dot{x}_2 &= h + \frac{1}{m_2} (\Delta I_2 - \Delta I_1). \end{aligned}$$

The two impacts are in compression as along as $\dot{x}_1 \le 0$ and $\dot{x}_2 \le 0$, which are true if the following two conditions together hold:

$$\Delta I_1 \le \frac{1}{1 + m_2} \Delta I_2, \tag{64}$$

$$\Delta I_2 \le -m_2 h. \tag{65}$$

Below we derive a condition on ΔI_2 that will satisfy (64).

Section 4.2 showed that $dI_1/dI_2 = 0$ at the transition $S_2 \to S_1$. Applying a sequence of reasoning steps similar to the one listed in the end of the proof for case (a), we have that, after S_1 starts:

$$\begin{aligned} \Delta I_1 \le \frac{1}{1 + m_2} \Delta I_2 &\Leftarrow \frac{dI_1}{dI_2} \le \frac{1}{1 + m_2} \\ &\Leftarrow k_1 E_1 \le \frac{1}{(1 + m_2)^2} \cdot k_2 E_2 \\ &\Leftarrow \frac{dE_1}{dI_2} \le \frac{k_2}{k_1 (1 + m_2)^2} \cdot \frac{dE_2}{dI_2}. \end{aligned} \tag{66}$$

In the state, the energy derivatives with respect to I_2 follow from (20), (22), and (24), with $v_1^{(0)} = v_2^{(0)} = h$:

$$\begin{aligned} \frac{dE_1}{dI_2} &= \frac{dE_1}{dI_1} \cdot \frac{dI_1}{dI_2} = \left(-\left(1 + \frac{1}{m_2}\right) \Delta I_1 + \frac{1}{m_2} \Delta I_2\right) \cdot \\ &\quad \sqrt{\frac{k_1 E_1}{k_2 E_2}}, \\ \frac{dE_2}{dI_2} &= -\left(h + \frac{1}{m_2} (\Delta I_2 - \Delta I_1)\right). \end{aligned}$$

We substitute the above derivatives into the last inequality in (66):

$$\left(-\left(1+\frac{1}{m_2}\right)\Delta I_1+\frac{1}{m_2}\Delta I_2\right)\cdot\sqrt{\frac{k_1E_1}{k_2E_2}}\leq -\frac{k_2}{k_1(\,1+m_2)^2}\left(h+\frac{1}{m_2}(\,\Delta I_2-\Delta I_1)\right). \quad (67)$$

Next, divide both sides by $\sqrt{k_1E_1/k_2E_2}$. Move all of the terms involving ΔI_2 to the left-hand side, while all of the terms involving ΔI_1 to the right-hand side. After merging, the coefficients of ΔI_1 and ΔI_2 are positive. Then we remove the term of ΔI_1, obtaining a condition over ΔI_2 that implies condition (67):

$$\begin{aligned}\Delta I_2 &\leq -\frac{h\left(\frac{k_2}{k_1}\right)^{3/2}\cdot\sqrt{\frac{E_2}{E_1}}\cdot\frac{1}{(1+m)^2}}{\frac{1}{m_2}+\frac{1}{m_2}\left(\frac{k_2}{k_1}\right)^{3/2}\cdot\sqrt{\frac{E_2}{E_1}}\cdot\frac{1}{(1+m)^2}}\\ &= -\frac{m_2h}{1+(\,1+m_2)^2\sqrt{\frac{E_1}{E_2}}\cdot\left(\frac{k_1}{k_2}\right)^{3/2}}. \quad (68)\end{aligned}$$

Because (68) implies $E_1/E_2 < k_2/(\,k_1(\,1+m_2)^2\,)$ from the steps of reasoning in (66), we can substitute $k_2/(\,k_1(\,1+m_2)^2\,)$ for E_1/E_2 into the right-hand side of (68) and get a condition that implies it:

$$\begin{aligned}\Delta I_2 &\leq -\frac{m_2h}{1+\frac{k_1}{k_2}(\,1+m_2)}\\ &= -\frac{m_2v_2^{(0)}}{1+\frac{k_1}{k_2}(\,1+m_2)}. \quad (69)\end{aligned}$$

The inequality (69) also implies (65). Thus, it guarantees both impacts to be in compression. □