Unified Particle Physics for Real-Time Applications

Miles Macklin∗ Matthias Müller† Nuttapong Chentanez‡ Tae-Yong Kim§

NVIDIA

Abstract

We present a unified dynamics framework for real-time visual effects. Using particles connected by constraints as our fundamental building block allows us to treat contact and collisions in a unified manner, and we show how this representation is flexible enough to model gases, liquids, deformable solids, rigid bodies and cloth with two-way interactions. We address some common problems with traditional particle-based methods and describe a parallel constraint solver based on position-based dynamics that is efficient enough for real-time applications.


Keywords: simulation, fluid simulation, unified solver, position-based dynamics, two-way fluid coupling, cloth simulation

1 Introduction

Unified solvers have become popular tools in offline visual effects in recent years. Notable examples include Maya’s Nucleus solver [Stam 2009] and Softimage’s Lagoa. We were inspired by the features of these packages, and our goal here is to create a unified solver capable of simulating gases, liquids, deformable solids, rigid bodies, and cloth, interacting with one another in real-time.

The concept of a unified dynamics solver is attractive. Having a single piece of software that can simulate multiple object types that were previously handled by specialized solvers means less code to write, optimize, and maintain. Additionally, objects that were previously simulated in isolation can now interact with each other in a fully coupled way. In this paper we present a unified approach that makes some compromises in realism to meet our goal of real-time performance, but enables a wider range of effects than was previously possible. Our framework is built on the position-based dynamics method (PBD) [Müller et al. 2007] which is popular in computer games and interactive applications for its simplicity and stability.

Our paper is organized as follows: first we discuss our particle-based representation (section 3), we then present our general purpose parallel constraint solver (section 4) followed by the constraint types used in our solver (sections 5-8). Our main contributions are summarized below:

• Extending constraint averaging to PBD to support parallel execution (section 4)
• Approximate shock propagation to increase convergence of rigid stacks (section 5.2)
• Static and dynamic friction in PBD (section 6.1)
• Two-way fluid solid coupling (section 7.1)
• Gas simulation with PBD (section 7.2)

2 Related Work

There has been increasing work towards unified simulation models recently, and point-based methods are well suited to the problem. Müller et al. [2004a] use a point based representation to model elastic and plastic solids that can topologically deform. Solenthaler et al. [2007] described a method for simulating fluids and solids based on smoothed particle hydrodynamics (SPH). Becker et al. [2009] improve the rotational invariance of their method using a co-rotated...
Our core particle state consists of the following properties: advanced algorithms that run serially. Consider many simple particle interactions in parallel, over more granularity at which to express parallelism. To take advantage to process, and avoid complex algorithms for generating contacts. Cles, we significantly reduce the number of collision types we need objects we wish to simulate. By constructing all objects from particles. Particles are attractive for their simplicity and ease of implementation. While being flexible enough to represent the range of types. Particle representation connected by shape matching and distance constraints to model elastic and plastic solids.

One aspect of unified simulation of particular interest is two-way interaction of fluids and solids. Carlson et al. [2004] presented a method to couple rigid bodies with a grid-based fluid simulation. Carlson [2004] also simulate rigid, melting and flowing materials in a unified way. Gueudet et al. [2004] show how to couple Eulerian fluid simulations to solid and thin-shell deformables. A full discussion of grid-based methods is, however, beyond the scope of our paper. Coupling particle fluids with rigid bodies was explored by Müller et al. [2004b] who couple a SPH fluid simulation to mesh based deformable objects. Clavet et al. [2005] couple fluids with rigid bodies by treating particles as hard spheres and applying impulses to a traditional rigid body simulator. Akinci et al. [2012] couple SPH fluids with rigid bodies using more accurate boundary force calculations, and extend their method to handle deformable solids in [Akinci et al. 2013b]. In contrast to these works, we simulate fluids and rigid bodies inside the same framework.

In addition to the work discussed here, we will refer to more specific related work in the respective sections.

3 Particle Representation

We choose particles as the fundamental building block for all object types. Particles are attractive for their simplicity and ease of implementation, while being flexible enough to represent the range of objects we wish to simulate. By constructing all objects from particles, we significantly reduce the number of collision types we need to process, and avoid complex algorithms for generating contacts between mesh based representations. Particles also provide a natural granularity at which to express parallelism. To take advantage of modern graphics processors (GPUs) we prefer algorithms that consider many simple particle interactions in parallel, over more advanced algorithms that run serially.

Our core particle state consists of the following properties:

```c
struct Particle {
    float x[3];
    float v[3];
    float invmass;
    int phase;
};
```

We extend the common dynamics quantities with a particle phase identifier. This integral value is used to organize particles into groups, and provides a convenient way to adjust their properties and control how they interact, for example, disabling collisions between groups of particles. We restrict ourselves to a fixed particle radius per scene in order to leverage efficient collision detection based on uniform grids.

4 Parallel SOR Solver

4.1 Background

Our system is built around a general purpose constraint solver that is derived from position-based dynamics. We give some background on position-based dynamics here, but refer the reader to Müller et al. [2007] and Bender et al. [2014] for more information. PBD solves a system of non-linear equality and inequality constraints such that

\[
C_i(x + \Delta x) = 0, \quad i = 1, \ldots, n \tag{1}
\]

\[
C_j(x + \Delta x) \geq 0, \quad j = 1, \ldots, n. \tag{2}
\]

where \(x = [x_1, x_2, \ldots, x_n]^T\) is the vector of particle positions. Constraints are typically solved through Gauss-Seidel iteration, where each constraint is solved sequentially using the linearization of \(C\) around \(x\).

\[
C_i(x + \Delta x) \approx C_i(x) + \nabla C_i(x) \Delta x = 0. \tag{3}
\]

The position change \(\Delta x\), is restricted to lie along the constraint gradient, and is weighted by the inverse of the mass matrix \(M = \text{diag}(m_1, \ldots, m_n)\).

\[
\Delta x = M^{-1} \nabla C_i(x)^T \lambda_i. \tag{4}
\]

Combining Eq. (3) and (4), \(\lambda_i\) is given by

\[
\lambda_i = - \frac{C_i(x)}{\nabla C_i(x) M^{-1} \nabla C_i(x)^T}. \tag{5}
\]

Typically positions are updated after each constraint is processed, and after a number of iterations a change in velocity is computed according to the total constraint delta

\[
\Delta v = \frac{\Delta x}{\Delta t}. \tag{6}
\]

4.2 Optimization Viewpoint

Here we present another perspective on position-based dynamics by viewing it as the solution to a constrained optimization problem. Considering only equality constraints, the problem to be solved can be stated as:

\[
\text{minimize} \quad \frac{1}{2} \Delta x^T M \Delta x \\
\text{subject to} \quad C_i(x + \Delta x) = 0, \quad i = 1, \ldots, n \tag{7}
\]

The solution variable \(\Delta x\) is the change in position such that the constraints are satisfied. According to Eq. (6) the resulting change in position is equivalent to applying an impulse at the beginning of the time-step. As such, the problem is equivalent to finding the minimum change in kinetic energy that satisfies the constraints, which is consistent with Gauss’s principle of least constraint.
Because the two constraints are identical, then (7) would be a constrained quadratic minimization problem (QP) with a closed form solution. In practice, the constraints are arbitrary non-linear, non-convex functions, for which fast and globally optimal solvers do not exist [Boyd and Vandenberghe 2004]. Position-based dynamics proceeds in the spirit of sequential quadratic programming (SQP) by linearizing the constraints and solving a sequence of local constrained quadratic minimizations:

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \Delta x^T M \Delta x \\
\text{subject to} & \quad J \Delta x = b
\end{align*}
\]

where \( J = \nabla C(x) \), and \( b = [-C_1, \ldots, -C_n]^T \). Problems of this type (minimizing a quadratic objective with linear constraints) can be transformed into the following optimality conditions:

\[
\begin{align*}
M \Delta x &= J^T \lambda \\
J \Delta x &= b
\end{align*}
\]

The first condition comes from the theory of Lagrange multipliers where the left-hand side is the gradient of the objective function, and the second condition comes from from the feasibility requirement. Equations (4) and (5) follow immediately from these conditions, and when considering the system as a whole, we can eliminate \( \Delta x \) and write

\[
JM^{-1}J^T \lambda = b.
\]

Eq. (11) is a matrix equation for the linearized problem. This is similar to the fast projection of Goldenthal et al. [2007]. However, in PBD, the matrix is never explicitly formed, and rather than solve for \( \lambda \) exactly, PBD applies a single Gauss-Seidel iteration over (11) before updating \( \Delta x \) and restarting the problem. Projected Gauss-Seidel is used in the presence of inequality constraints.

Gauss-Seidel iteration is inherently serial, so to increase parallelism we solve constraints in a projected Gauss-Jacobi fashion. Unfortunately, Jacobi iteration is not guaranteed to converge if the system matrix is not positive definite. In some cases this is not a problem and Macklin and Müller [2013] successfully used Jacobi iteration to simulate fluids in the position-based framework using constraint regularization. However, for many common constraint configurations it is clear that a solution will not be reached. We illustrate the problem by considering a 1-dimensional particle constrained to the origin by two identical distance constraints (Figure 2).

![Figure 2: A particle constrained to lie at the origin by two identical distance constraints.](image)

In this example, the system of constraint equations to solve is:

\[
\begin{align*}
C_1(x) &= x = 0 \\
C_2(x) &= x = 0
\end{align*}
\]

Because the two constraints are identical, \( J \) is rank deficient, and the system matrix is singular. Although Gauss-Seidel iteration would find a solution, Jacobi iteration will oscillate between two fixed solutions indefinitely (the positive and negative side of the origin). If the constraints are not identical but merely close, then the condition number of the matrix is large and convergence will be slow. Situations like this are common in practice.

To address this problem, we apply under-relaxation based on the concept of constraint averaging [Bridson et al. 2002], or mass-splitting [Tonge et al. 2012]. We first process each constraint in parallel and accumulate position deltas for each particle. At the end of the iteration, once all constraints are processed, we divide each particle’s total constraint delta by \( n_i \), the number of constraints affecting the particle, to get the averaged position update \( \Delta \tilde{x} \):

\[
\Delta \tilde{x} = \frac{1}{n_i} \Delta x_i.
\]

This form of local relaxation is not guaranteed to conserve momentum when neighboring particles have differing number of constraints, however visual errors are typically not noticeable.

### 4.3 Successive Over-Relaxation

Averaging constraint forces as described above ensures convergence, but in some cases this averaging is too aggressive and the number of iterations required to reach a solution increases. To address this we introduce a global user-parameter \( \omega \) which controls the rate of successive over-relaxation (SOR),

\[
\Delta \tilde{x}_i = \frac{\omega}{n_i} \Delta x_i.
\]

In all our simulations we have used \( 1 \leq \omega \leq 2 \), although higher values may be used depending on the scene being simulated. Additional under-relaxation (\( \omega < 1 \)) is not typically required as the constraint averaging is sufficient to avoid divergence.

In practice it can be desirable to have some constraint types take precedence over others. To achieve this, we process constraints in groups and immediately apply the accumulated position delta to the particle positions before the next constraint group is processed. Typically, all constraints of the same type are processed in the same group. For example, after processing all density constraints in parallel, the accumulated delta can be applied to our candidate position before processing contact constraints in parallel (steps 17-21 in Algorithm 1). This method also propagates constraint corrections faster which improves convergence speed.

### 4.4 Initial Conditions

A common situation is for constraints to be violated at the beginning of a time-step. This could occur, for example, due to convergence not being reached at the end of the last step, or a kinematic object being moved to an invalid state in response to user input. Due to the way position-based dynamics solves constraints, these invalid initial conditions may cause kinetic energy to be added to the system in an undesirable manner. To illustrate the problem, we consider a particle initially interpenetrating the ground (Figure 4). The position-based solver calculates the corrected position at the surface \( x^* \), projects the particle there and updates the velocity according Eq. (6).

This new velocity causes the particle to continue to travel upwards and ‘pop’ out of the ground. What’s worse, this incorrect velocity becomes larger as the time-step decreases! This is a serious problem for stable stacking and can cause obvious non-physical behavior such as cloth bouncing off the ground. One way to view this correction of initial error is as a Baumgarte stabilization [Ascher et al. 1995] with a stiffness factor of 1. The differential algebraic equations (DAE) community have long known about the
drawbacks of Baumgarte stabilization and have moved to pre- and post-stabilization techniques that modify position independent of velocity. For further reading we refer to [Weinstein et al. 2006], which provides a good overview of stabilization methods from a computer graphics perspective.

We address this issue by optionally performing a pre-stabilization pass on the initial positions in order to move particles to a valid state. The process involves first predicting new positions, generating contacts, and then solving the resulting contact constraints with the original, rather than predicted positions. Any delta applied to the original positions are also applied to the predicted positions before the main constraint solving loop begins. In practice we perform this pre-stabilization only for contacts, as this is the most visible source of error. If convergence is not fully reached during this pre-stabilization pass then some energy may still be added to the system, however 1-2 iterations is usually sufficient to resolve most visual artifacts.

4.5 Particle Sleeping

Positional drift may occur when constraints are not fully satisfied at the end of a time-step. We address this by freezing particles in place if their velocity has dropped below a user-defined threshold,

$$\mathbf{x}(t + \Delta t) = \begin{cases} \mathbf{x}^*, & |\mathbf{x}^* - \mathbf{x}_0| > \epsilon \\ \mathbf{x}_0, & \text{otherwise} \end{cases}$$

(14)

Figure 3: Showing the Stanford bunny at different sizes sampled by particles.

Algorithm 1 Simulation Loop

1: for all particles $i$ do
2: apply forces $\mathbf{v}_i \leftarrow \mathbf{v}_i + \Delta t \mathbf{f}_{ext}(\mathbf{x}_i)$
3: predict position $\mathbf{x}_i^* \leftarrow \mathbf{x}_i + \Delta t \mathbf{v}_i$
4: apply mass scaling $m_i^* = m_i e^{-kh} (\mathbf{x}_i^*)$
5: end for
6: for all particles $i$ do
7: find neighboring particles $N_i(\mathbf{x}_i^*)$
8: find solid contacts
9: end for
10: while iter < stabilizationIterations do
11: $\Delta \mathbf{x} \leftarrow 0$, $n \leftarrow 0$
12: solve contact constraints for $\Delta \mathbf{x}$, $n$
13: update $\mathbf{x}_i \leftarrow \mathbf{x}_i + \Delta \mathbf{x}/n$
14: update $\mathbf{x}^* \leftarrow \mathbf{x}^* + \Delta \mathbf{x}/n$
15: end while
16: while iter < solverIterations do
17: for each constraint group $G$ do
18: $\Delta \mathbf{x} \leftarrow 0$, $n \leftarrow 0$
19: solve all constraints in $G$ for $\Delta \mathbf{x}$, $n$
20: update $\mathbf{x}^* \leftarrow \mathbf{x}^* + \Delta \mathbf{x}/n$
21: end for
22: end while
23: for all particles $i$ do
24: update velocity $\mathbf{v}_i \leftarrow \frac{1}{\Delta t} (\mathbf{x}_i^* - \mathbf{x}_i)$
25: advect diffuse particles
26: apply internal forces $\mathbf{f}_{drag}, \mathbf{f}_{vort}$
27: update positions $\mathbf{x}_i \leftarrow \mathbf{x}_i^*$ or apply sleeping
28: end for

5 Rigid Bodies

Harada [2007] and Bell et al. [2005] used a particle representation to simulate rigid bodies with a penalty force model, while Tonge et al. [2010] used particles to detect contacts for a constraint based GPU rigid body solver. An alternative approach is to construct rigid bodies by connecting particles with a lattice of distance constraints, however this method requires many iterations to appear stiff.

We represent non-convex rigid bodies using particles, and use rigid-shape matching constraints [Müller et al. 2005] to maintain particle configurations. Shapes are created by first performing solid voxelization of a closed triangle mesh and placing particles at occupied cells interior to the shape. We then assign all particles in the shape the same phase-identifier, disable collision between them, and add a shape-matching constraint to the system.

During simulation we treat particles as if they were unconnected, subjecting them to displacements from the other constraint types. We then find the least squares best transform that maps the rest-
positions to the deformed state and transform particles accordingly (Figure 5). This deformation and re-targeting is similar to the approach taken by Bao et al. [2007] in the context of a finite element simulation.

The position delta due to a shape-matching constraint is given by

$$\Delta x_i = (Qr_i + c) - x_i^*$$

(15)

where $Q$ is a rotation matrix given by the polar-decomposition of the deformed shape’s covariance matrix $A$, calculated as:

$$A = \sum_{i}^{n} (x_i^* - c) \cdot r_i^T$$

(16)

where $r_i$ is the particle’s offset from the center of the mass in the rest configuration, and $c$ is the center of mass of the particles in the deformed configuration. The evaluation of Eq. (16) can be performed efficiently in parallel by assigning a thread per particle to calculate each outer product and then forming $A$ using a parallel reduction across all threads in the shape.

We note that using this method we never explicitly store the rigid body’s linear or angular velocity, as these are given implicitly by the particle state. The shape matching step automatically respects the inertial properties of the object, making this a very simple rigid body simulator.

One limitation of this method is that, when using particles with a fixed radius $r$, the number of particles required to represent a given shape grows proportional to $O(n^3)$ where $n = \frac{1}{r}$. This places a limit on the maximum size ratio that can be efficiently represented, so for real-time simulations we suggest a rule of thumb of 1:10. Placing particles in a thin layer around the surface can reduce the number between the smallest and largest feature (see Figure 3). To prevent tunneling.

To resolve collisions, Harada [2007] used discrete element forces between pairs of particles belonging to rigid bodies. The simplicity of this approach is attractive as only local collisions between particle pairs need be considered. However, if tunneling occurs then bodies can interlock and fail to separate (Figure 6). We address this problem using a new particle collision algorithm based on a sparse signed distance field (SDF) representation.

Guendelman et al. [2003] used signed distance fields to generate contacts between non-convex rigid bodies by point sampling triangle mesh features within each overlapping shape’s SDF. Collisions are then resolved using an sequential, iterative scheme where each feature is projected out of the shape in order of deepest penetration first. In contrast to Guendelman et al. [2003] who use grid-based SDFs, we sample our field function onto each particle belonging to a rigid shape. By storing the SDF on particles we are able to re-use all the machinery of our particle-particle collision detection to resolve deeper overlaps between shapes. We store both the magnitude and gradient we choose the SDF gradient corresponding to the minimum translation distance, $d = \min(|\phi_i|, |\phi_j|)$.

The position delta for each particle is then given by

$$\Delta x_i = -\frac{w_j}{w_i + w_j} (d \cdot n_{ij})$$

(18)

$$\Delta x_j = -\frac{w_i}{w_i + w_j} (d \cdot n_{ij})$$

(19)

where $w_i = 1/m_i$.

The signed distance field is often undersampled near the surface of a shape which can lead to discontinuous penetration depths and jittering. To address this we treat boundary particles (those with $|\phi| > r$) separately from the method for interior particles described above. When a collision between a boundary particle and another particle is detected, we modify the contact normal according to [Müller and Chentanez 2011a], and use $d = |x_i - x_j| - r$, effectively treating boundary particles as one-sided hard spheres.

To resolve collisions with this representation we first detect overlaps between particles with distance $|x_i - x_j| < r$. As the contact normal we choose the SDF gradient corresponding to the minimum translation distance, $d = \min(|\phi_i|, |\phi_j|)$.

$$n_{ij} = \begin{cases} \nabla \phi_i & \text{if } |\phi_i| < |\phi_j| \\ -\nabla \phi_j & \text{otherwise} \end{cases}$$

(17)

The position delta for each particle is then given by

$$\Delta x_i = -\frac{w_j}{w_i + w_j} (d \cdot n_{ij})$$

(18)

$$\Delta x_j = -\frac{w_i}{w_i + w_j} (d \cdot n_{ij})$$

(19)

where $w_i = 1/m_i$.

The signed distance field is often undersampled near the surface of a shape which can lead to discontinuous penetration depths and jittering. To address this we treat boundary particles (those with $|\phi| > r$) separately from the method for interior particles described above. When a collision between a boundary particle and another particle is detected, we modify the contact normal according to [Müller and Chentanez 2011a], and use $d = |x_i - x_j| - r$, effectively treating boundary particles as one-sided hard spheres.

This modified boundary normal (Figure 8) can be computed efficiently as the reflection of the collision direction around the SDF gradient (from Eq. 17) when a particle lies in the negative half-space of the particle.

$$n_{ij}^* = \begin{cases} x_{ij} - 2(x_{ij} \cdot n_{ij})n_{ij} & x_{ij} \cdot n_{ij} < 0 \\ x_{ij} & \text{otherwise} \end{cases}$$

(20)
A limitation of this approach is that particle surfaces are not entirely smooth. To reduce bumping and artificial sticking, we ensure some overlap between particles in the rest pose.

5.2 Stiff Stacks

Jacobi methods can only propagate information (collision deltas) over a distance of one particle per iteration. This can lead to large of piles rigid bodies requiring many iterations to appear stiff. One method for increasing the rate of information propagation is shock propagation [Guendelman et al. 2003]. Shock propagation works by iterating bodies from the ground upwards, fixing each layer of bodies in place after they have been processed. Unfortunately this process is inherently serial (layers are processed one at a time in a sequential fashion). Here we present a new, parallel-friendly method to improve the stability of rigid stacks that does not require changing iteration order.

First, we estimate the stack height of each particle, \( h \). This may be done through a contact graph search, or field propagation, from the solid boundary. The main requirement is that the height function increases in the direction opposite gravity. In our examples we have used a simple heuristic function measuring height from a fixed ground plane.

Once we have an estimate of stack height, we temporarily modify each particle’s mass such that lower particles have a larger mass than the ones above. In our experiments, we found the following exponential function works well as a scale factor for particle mass,

\[
s_i(x_i) = e^{-kh_i(x_i)}
\]

where \( h_i(x_i) \) is the stack height for a particle. In the case that \( h_i(x_i) \) is equal to height from the ground plane then this scaling function provides a constant mass ratio of \( s_i/s_j = e^{-kr} \) for two particles \( i \) and \( j \) stacked a distance \( r \) apart. The resulting mass ratio causes lower particles to feel less pressure and to resist compression. This causes stacks of bodies to converge to a solution much faster (Figure 10). In our rigid piling scenes we have typically used values of \( k \in [1 \ldots 5] \).

The scaled particle mass \( m_i^* = s_i m_i \) is used only during the contact processing phase and is reset post-constraint solve. Because each particle is processed independently, we can perform this mass adjustment in parallel (step 4 in Algorithm 1).

We note that our method is not specific to particle-based solvers and may be applied equally well in the context of a traditional rigid body solver. However a potential limitation of this method is that if the coefficient \( k \) is set too high then the lower particles will stop responding to interaction from above. One solution to this problem is to perform mass modification only in the final solver iteration, however we have not found this to be necessary in order to achieve good results.

5.3 Deformation

Each shape-matching constraint has an associated stiffness parameter that allows us to model small scale elastic deformation. In addition, we can deform objects plastically using the method described in [Müller and Chentanez 2011b]. Briefly, we detect when a particle belonging to a rigid shape has been deformed past a user threshold. We then mix this deformation back into the particle’s local-space rest position according to a plastic creep coefficient. Because this process invalidates our SDF field, we convert particles back to regular spherical particles if they deform past a certain threshold.

Currently we use a single shape-matching constraint per object, which limits the amount of deformation possible. Larger deformations can be supported by combining multiple shape-matching constraints, or using methods such as oriented particles or tetrahedral volume constraints, both of which can be accommodated in our constraint framework.
Granular Materials and Friction

We present our friction model in the context of granular materials whose behavior is heavily dependent on frictional effects. We note, however, that the same method presented here is used for all object types in our framework.

Previous work by Bell et al. [2005] modeled granular materials using the discrete element method and irregular groups of particles. Zhu and Bridson [2005] used the FLIP method to animate sand, while Alduán and Otaduy [2011], and Ihmsen et al. [2012b] used iterative Smoothed Particle Hydrodynamics (SPH) solvers to model granular materials with friction and cohesion. Position-based dynamics traditionally models friction by damping velocity after the position-based constraint solve has completed [Müller et al. 2007]. This method cannot model static friction because the frictional damping forces cannot correct for the position changes already made during the constraint solve. Consequently, particle positions drift, and piles quickly collapse. We address this problem using a novel formulation that applies friction at the position level.

6.1 Friction Model

During contact handling, we first resolve interpenetration by projecting particles a distance $d$ along the collision normal according to the following non-penetration constraint,

$$C(x_i, x_j) = |x_{ij}| - r \geq 0.$$  (22)

Once interpenetration has been resolved, we calculate a frictional position delta based on the relative tangential displacement of the particles during this time-step

$$\Delta x_{\perp} = [(x_i^* - x_i) - (x_j^* - x_j)] \perp n,$$  (23)

where $x_i^*$ and $x_j^*$ are the current candidate positions for the colliding particles including any previously applied constraint deltas, $x_i$ and $x_j$ are the positions of the particles at the start of the time-step, and $n = x_{ij}^*/|x_{ij}^*|$ is the contact normal. The frictional position delta for particle $i$ is then computed as

$$\Delta x_i = \frac{w_i}{w_i + w_j} \left( \frac{\Delta x_{\perp}}{w_i + w_j} \cdot \min\left(\frac{\mu_s d}{|\Delta x_{\perp}|}, 1\right), \frac{|\Delta x_{\perp}| < \mu_s d}{|\Delta x_{\perp}|}, \text{otherwise} \right)$$  (24)

where $\mu_k, \mu_s$ are the coefficients of kinetic and static friction respectively. The first case in Eq. (24) models static friction by removing all tangential movement when the particle’s relative velocity is below the traction threshold. The second case models kinetic Coulomb friction, limiting the frictional position delta based on the penetration depth of the particle. The position change on particle $j$ is given by

$$\Delta x_j = -\frac{w_j}{w_i + w_j} \Delta x_i.$$  (25)

Treating the frictional impulses this way allows us to generate large particle piles with high angles of repose (Figure 11). However, as with position-based dynamics generally, a limitation is that friction strength is somewhat dependent on iteration count. Müller et al. [2007] suggest a method to reduce this effect.

7 Fluids

We simulate fluids in our framework using the position-based fluids method [Macklin and Müller 2013], where the fluid density constraint (Eq. 26) is considered simply as another constraint in the system. In contrast to [Macklin and Müller 2013] we clamp the density constraint to be non-negative (unilateral), so that it only acts to separate particles,

$$C(x_1, \ldots, x_n) = \frac{\rho_i}{\rho_0} - 1 \leq 0,$$  (26)

we then use the model of [Akinci et al. 2013a] to handle cohesion and surface tension effects.

7.1 Fluid-Solid Coupling

Akinci et al. [2012] show how to accurately couple SPH fluids with rigid bodies by placing particles at solid boundaries and calculating the Rayleigh-Taylor instability.
fluid pressure forces which are then input to a traditional rigid body simulator. Our time-steps are typically large enough that we cannot rely on fluid pressure to prevent interpenetration and tunneling. Consequently, when a fluid particle comes into contact with a solid particle, we treat both as solid particles with the fluid rest distance used as the contact distance $r$.

We include solid particles in the fluid density estimation, so the density for a fluid particle is then given by

$$\rho_i = \sum_{\text{fluid}} W(x_i - x_j, h) + s \sum_{\text{solid}} W(x_i - x_j, h) \quad (27)$$

where $h$ is width of the smoothing kernel $W$. The parameter $s$ accounts for any difference in sampling density between solid particles and fluid particles. If the distance between fluid particles at the rest density is the same as solid particles then $s$ can be set to 1. If solids are sampled more densely than fluids then it should be set $< 1$. Because $s$ is a constant, we assume solid particles are sampled at relatively uniform density. Akinci et al. [2012] use a per-particle correction, however we have not found this to be necessary, partly because we typically use regular particle samplings, and also because we do not rely on fluid pressure forces to resolve fluid-solid interactions.

### 7.1.1 Buoyancy

The density constraint formulation of [Macklin and Müller 2013] treats fluid particles as having equal mass, but by using the mass weighted version of position-based dynamics (Eq. 4) we can simulate fluids and rigid bodies with differing densities. This simple adjustment automatically gives rise to buoyancy and sinking of objects with differing mass ratios (Figure 12).

For the fluid density constraints, we pre-compute the denominator in Eq. (5) for a reference filled particle configuration. During this pre-computation step we assume that neighboring particles have the same mass, so to ensure the subsequent position corrections from Eq. (4) are conservative, the smallest expected particle mass in the system should be used.

To simulate immiscible multi-phase liquids (Figure 13) we allow fluid particles of different phases to contribute to each other’s density estimation. We apply cohesive and surface tension forces only between particles of the same phase as in [Solenthaler and Pajarola 2008].

### 7.2 Gases

Grid-based methods have proved popular for animating gases and smoke, in particular [Stam 1999] and [Fedkiw et al. 2001]. More recently, Lagrangian discretizations for smoke have gained interest in computer graphics, notably vortex methods such as [Park and Kim 2005], and mesh-tracking methods [Pfaff et al. 2012][Brochu et al. 2012]. Stam and Fiume [1995] were the first to apply SPH to the simulation of gas and fire, and Gao et al. [2009] combined SPH with a grid-based fluid solver to simulate fast moving gases.

We present a fully Lagrangian method for simulating gases based on position-based fluids. Our method is sparse in the sense that it only computes fluid dynamics in areas of visual interest, and, in contrast to vortex particle methods which require special handling for solid boundaries, our method handles solid boundary conditions through simple position projection during the regular particle collision constraint pipeline.

Figure 14: A rising smoke plume simulated using position-based fluids, this example uses 16k fluid particles (32x32x16) and 65k smoke particles.

Figure 15: Slice from a smoke simulation with closed boundaries. The domain is seeded with fluid particles (blue) and smoke particles (white) are passively advected through the velocity field.

### 7.2.1 Closed Boundaries

To model smoke in closed environments, we begin by filling the computation domain with fluid particles and reducing gravity for gas particles according to a user parameter $\alpha$, although a temperature-based buoyancy model would work equally well if tracking temperature per particle. We treat the gas as incompressible and use the unilateral density constraint to maintain a fixed rest density.

We inject visual smoke particles into the simulation domain and passively advect them through the velocity field defined on the fluid particles (Figure 14). To calculate the velocity at a smoke particle’s location $x_s$ we use the following weighted averaging over fluid particles:

$$v(x_s) = \frac{\sum_j v_j W(x_s - x_s)}{\sum_j W(x_s - x_s)} \quad (28)$$

where $W(x)$ is any symmetric kernel function, e.g. the cubic Poly6 kernel given in [Müller et al. 2003], and $v_j$ is the velocity of a fluid particle which overlaps the sampling location. Note that we use the velocity calculated immediately after the constraint solve to ensure the velocity used for advection is consistent with the incompressibility constraints (step 24 in Algorithm 1). We note that these passive visual marker particles are useful in other contexts, for example we also use them to represent spray and foam particles in liquid simulation as in [Hindsen et al. 2012a].
7.2.2 Open Boundaries

To generate plumes of smoke, we inject fluid particles and smoke particles together, ensuring a 1-2 layer thick boundary of fluid particles surrounding the smoke particles at their emission point. Interior fluid particles are emitted with a higher velocity than the boundary particles, and to model the effect of fast moving air interacting with the surrounding environment we introduce a new drag force:

\[ f_{\text{drag}_i} = -k(v_i - v_{\text{env}})(1 - \frac{\rho_i}{\rho_0}) \]  

where \( v_{\text{env}} \) is the environmental velocity at the particle’s location. In our examples, \( v_{\text{env}} \) has been taken to be 0 to model still air, however any procedural velocity field may be used. The last factor on the right-hand side of Eq. (29) ensures that only particles near the free surface experience drag. If smoke particles become isolated from the fluid we revert them to ballistic motion, or advect them by the procedural background velocity field. Additionally, fluid particles may be removed once they have no smoke particles within their kernel radius, or after a pre-defined lifetime.

Because we do not simulate fluid where smoke does not exist, this technique is not well suited to simulating long range effects, e.g.: a gust of wind from a window disturbing smoke at the other end of the room. However these types of effects may be approximated by approximating each triangle as a thin airfoil and distributing lift and drag forces to the triangle’s particles [Keckeisen et al. 2004].

7.2.3 Baroclinic Turbulence

Our drag model creates a velocity gradient that gives rise to some interesting motion at the fluid boundary, however it is often desirable to introduce additional turbulent motion. Fedkiw et al. [2001] introduced vorticity confinement to computer graphics in the context of smoke simulation, and Selle et al. [2005] used vortex particles to add vorticity to grid-based smoke simulations. Pfaff et al. [2012] use a mesh-based representation to model the smoke-air interface where vortex sheets are generated for each triangle and edge including a baroclinity term. [Kim et al. 2012a] use vortex particles to model baroclinic turbulence in grid-based smoke simulations.

In this section, we present a turbulence model inspired by the Lagrangian vortex sheets method [Pfaff et al. 2012]. Unlike mesh-tracking methods, we do not have an explicit surface representation, instead we use the following unnormalized vector as an approximate surface normal

\[ n_i = \sum_j \frac{m_j}{\rho_j} \nabla W(x_i - x_j). \]  

The magnitude of this vector provides a measure of how close to the surface a particle is. Larger magnitudes indicate surface particles and smaller magnitudes indicate interior particles. We combine it with the Boussinesq pressure gradient approximation to evolve a per-particle vortex strength according to the following differential equation

\[ \frac{d\omega_i}{dt} = \omega_i \cdot \nabla v + \beta (n_i \times g), \]  

where the first term is due to vortex stretching, and the second term represents the baroclinity. \( \beta \) is an exposed parameter used to control the overall strength of the effect, and the velocity gradient \( \nabla v \) can be approximated using SPH gradient kernels. The final driving force on a particle due to its neighbors is given by

\[ f_{\text{vort}_i} = \sum_j (\omega_j \times x_{ij}) W(x_{ij}). \]  

Good results can be obtained by using a high particle drag force e.g.: \( k = 50 \), and allowing particle vorticity to drive the flow.

7.2.4 Rendering

Unlike grid-based solvers there is no built-in diffusion to Lagrangian schemes. To visually approximate diffusion of smoke density, we increase smoke particle size and decrease opacity over the lifetime of the particle, similar to [Brochu et al. 2012]. For our real-time renderer, we sort smoke particles according to depth from camera and render them as point sprites. Light transmission effects are approximated by rendering particles as opaque points into a shadow map and treating the light-space depth delta as input to an exponential scattering function.

8 Cloth and Ropes

We build cloth models using networks of distance constraints (Fig. 17) along triangle edges to model stretching, and across edges to model bending. In addition to the basic distance constraint we expose unilateral distance constraints (“tethers”) and use them as long-range attachments to reduce stretching [Kim et al. 2012b]. Cloth self-collision and inter-collision with other cloth pieces is handled automatically by the particle collision pipeline, but we note this requires sampling the cloth surface with enough particles to prevent tunneling. An aerodynamic model is applied to the cloth by approximating each triangle as a thin airfoil and distributing lift and drag forces to the triangle’s particles [Keckeisen et al. 2004].

To model ropes, we connect chains of particles with stretch and bending constraints, then extrude tessellated cylinders along their path. More advanced rope models such as [Rungjiratananon et al. 2011] may be integrated into our solver and would allow modeling of torsion. Inflatable objects can be modeled by combining a cloth mesh with a volume constraint as described in [Müller et al. 2007].

9 Implementation Details

For particle neighbor finding, we use an efficient parallel hash-grid [Green 2008]. Particle interactions are found using discrete overlap tests at the beginning of a time step, however if particles move significantly during constraint solving then subsequent new collisions may be missed. To mitigate this, we allow expanding the
collision radius by a fixed percentage during the overlap checks. This increases the number of potential colliders that are processed so should be set as a small fraction of the particle radius.

Constraints may be solved in a particle-centric or constraint-centric manner. In the particle-centric approach (Algorithm 2), we assign a GPU thread per particle and loop over all constraints affecting it, performing a single write operation per particle once all constraints are processed. Alternatively, we may solve constraints in a constraint-centric manner (Algorithm 3) where each thread processes a constraint and uses atomic operations to scatter position deltas to each affected particle. In practice we use both methods depending on the constraint type. For instance, fluid density constraints are solved in a particle-centric manner, and distance constraints are solved in a constraint-centric manner. Although scattered writes based on atomic memory transactions are generally less efficient, we have found this approach to be quite practical on current GPU hardware. Coherence in memory transactions is improved by re-ordering particle data according to the hash-grid cell index.

Particles are an inefficient choice of collision primitive for large shapes such as walls and floors. We represent these kinematic shapes using traditional representations such as convex hulls, signed distance fields, and triangle meshes.

10 Results

We have implemented our solver in CUDA and timings for various scenes were measured on an NVIDIA GTX680 GPU (Table 1). These times do not include rendering, however all our results were captured in real-time apart from the Rayleigh-Taylor scene which consists of 300k particles and simulates in around 150ms per frame. We render fluids using the ellipsoid splatting and screen-space surfacing described in [Macklin and Müller 2013].

Figure 1 shows fluids, rigid bodies, cloth and ropes together with two-way interactions, the drag model on our cloth slows the rigid bunnies descent via the connecting ropes. Figure 16 shows a frame from a simulation where a cloth sheet falls knocking over rigid bodies and then interacting with a smoke plume rising due to baroclinic turbulence. Figure 18 shows water balloons built from a cloth mesh filled with fluid particles. When the user pulls on the cloth mesh we dynamically remove constraints allowing them to tear the mesh and release the fluid. Because the fluid and cloth are two-way coupled, the escaping fluid forces the balloon backwards.

Table 1: Performance results for several examples. A frame time of 16ms is used in all cases.

<table>
<thead>
<tr>
<th>Scene</th>
<th>particles</th>
<th>diffuse</th>
<th>steps/frame</th>
<th>iters/step</th>
<th>ms/frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bunny Splash</td>
<td>50k</td>
<td>30k</td>
<td>2</td>
<td>4</td>
<td>10.1</td>
</tr>
<tr>
<td>Bunny Pile</td>
<td>44k</td>
<td></td>
<td>2</td>
<td>2</td>
<td>3.8</td>
</tr>
<tr>
<td>Smoke Cloth</td>
<td>10k</td>
<td>100k</td>
<td>2</td>
<td>6</td>
<td>5.6</td>
</tr>
<tr>
<td>Sandcastle</td>
<td>73k</td>
<td></td>
<td>2</td>
<td>12</td>
<td>10.2</td>
</tr>
<tr>
<td>Water Balloons</td>
<td>29k</td>
<td></td>
<td>3</td>
<td>12</td>
<td>12.1</td>
</tr>
</tbody>
</table>

11 Limitations and Future Work

In the future, we would like to remove the requirement of fixed particle sizes through the use of hierarchical acceleration structures. This would allow rigid bodies and fluids to be simulated more efficiently. Solving certain constraint types at different frequencies would also improve efficiency as in [Stam 2009].

Particles alone cannot accurately represent large flat surfaces, and our first order SDF is not accurate enough to provide useful collision normals near the surface of poorly sampled shapes. Using a quadratic basis function, as in [Corbett 2005], and reconstructing the signed distance value (instead of point sampling) could improve results for sparse samplings.

Shape-matching convergence is dependent on the total number of particles in the shape. This makes our method more suitable to smaller debris-like rigid bodies, and less suitable for large shapes. Buoyancy is also affected by the number of particles in the shape because the slower convergence makes the rigid body behave as if it were heavier. Consequently, our mass ratios do not correspond to real-world submersion depths.

12 Acknowledgments

The authors would like to thank NVIDIA and the PhysX team for supporting our work, especially Richard Tonge, Simon Schirm and Christian Sigg for many insightful discussions. We also thank the anonymous reviewers for their valuable feedback. The Bunny model is used courtesy of the Stanford Computer Graphics Laboratory.
References


Computer Animation and Virtual Worlds.


ACM Trans. Graph. 21, 3 (July), 594–603.


GREEN, S. 2008. CUDA particles. nVidia Whitepaper 2, 3.2, 1.


