Physically Based Deformable Models in Computer Graphics

Andrew Nealen\textsuperscript{1}, Matthias Müller\textsuperscript{2,3}, Richard Keiser\textsuperscript{3}, Eddy Boxerman\textsuperscript{4} and Mark Carlson\textsuperscript{5}

\textsuperscript{1} Discrete Geometric Modeling Group, TU Darmstadt
\textsuperscript{2} NovodeX / AGEIA
\textsuperscript{3} Computer Graphics Lab, ETH Zürich
\textsuperscript{4} Department of Computer Science, University of British Columbia
\textsuperscript{5} DNA Productions, Inc.

Abstract

Physically based deformable models have been widely embraced by the Computer Graphics community. Many problems outlined in a previous survey by Gibson and Mirtich [GM97] have been addressed, thereby making these models interesting and useful for both offline and real-time applications, such as motion pictures and video games. In this paper, we present the most significant contributions of the past decade, which produce such impressive and perceptually realistic animations and simulations: finite element/difference/volume methods, mass-spring systems, meshfree methods, coupled particle systems and reduced deformable models based on modal analysis. For completeness, we also make a connection to the simulation of other continua, such as fluids, gases and melting objects. Since time integration is inherent to all simulated phenomena, the general notion of time discretization is treated separately, while specifics are left to the respective models. Finally, we discuss areas of application, such as elastoplastic deformation and fracture, cloth and hair animation, virtual surgery simulation, interactive entertainment and fluid/smoke animation, and also suggest areas for future research.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Physically Based Modeling I.3.7 [Computer Graphics]: Animation and Virtual Reality

1. Introduction

Physically based deformable models have two decades of history in Computer Graphics: since Lasseter’s discussion of squash and stretch [Las87] and, concurrently, Terzopolos et. al’s seminal paper on elastically deformable models [TPBF87], numerous researchers have partaken in the quest for the visually and physically plausible animation of deformable objects and fluids. This inherently interdisciplinary field elegantly combines newtonian dynamics, continuum mechanics, numerical computation, differential geometry, vector calculus, approximation theory and Computer Graphics (to name a few) into a vast and powerful toolkit, which is being further explored and extended. The field is in constant flux and, thus, active and fruitful, with many visually stunning achievements to account for.

Since Gibson and Mirtich’s survey paper [GM97], the field of physically based deformable models in Computer Graphics has expanded tremendously. Significant contributions were made in many key areas, e.g. object model-
physical solution of PDEs, modal analysis, space-time adaptivity, multiresolution modeling and real-time simulation. Nonphysical models, such as parametric curves and surfaces and free-form deformations, are not discussed in this report. The inclined reader is therefore encouraged to browse more recent literature on t-splines [SZBN03] [SCF*04], space-warping [MBF02] [LKG*03] [BK05] and methods based on differential surface properties [SLCO*04] [YZX*04] [BK04] [LSLC05] [NSAC05] [IMH05]. For advances in character skinning see e.g. [WP02] [KIP02] [JT05]. Since we are not able to cover basic elasticity theory and continuum mechanics in this report, we would like to point out that a nice review of the history of elasticity theory, starting with the discovery of Hooke’s Law in 1660 (Fig. 1) and leading up to the general equations of Navier in 1821, is given in [Lov27]. Furthermore, great introductions to continuum mechanics and dynamics can be found in [WB97] and in general textbooks, such as [Chu96] [Coo95] [BW97] [Gdo93] [BLM00]. For application specific presentations, we refer the reader to a number of recent works. For cloth simulation, there is the text by House and Breen [HB00], as well as the recent, extensive tutorial by Thalmann et al. [MTCK*04]. For hair simulation, there is the (already slightly dated) overview by Thalmann et al. [MTM00]. For collision detection and haptic force-feedback rendering for deformable objects are other challenging and active areas of research. For a summary of recent work in these fields, we refer the reader to the report by Teschner et al. [TKH*05] and the course notes of Lin and Otaduy [LO05].

In this report we take a model based point of view, motivated by the fact that there are many readily available physical models for very similar applications, i.e. we can animate an elastically or plastically deforming solid with many different underlying models, such as mass-spring systems, finite elements or meshfree methods. We furthermore make a distinction between Lagrangian methods, where the model consists of a set of points with varying locations and properties, and Eulerian methods, where model properties are computed for a set of stationary points. To give a coarse overview, we describe recent developments for

- Lagrangian Mesh Based Methods
  - Continuum Mechanics Based Methods
  - Mass-Spring Systems
- Lagrangian Mesh Free Methods
  - Loosely Coupled Particle Systems
  - Smoothed Particle Hydrodynamics (SPH)
  - Mesh Free Methods for the solution of PDEs
- Reduced Deformation Models and Modal Analysis
- Eulerian and Semi-Lagrangian Methods
  - Fluids and Gases
  - Melting Objects

In each section we present the basic model formulation, recent contributions, benefits and drawbacks, and various areas of application. The section on fluids, gases and melting objects contains an overview of recent work and establishes the connection to the field of physically based deformable models. A complete survey on the animation of fluids and gases would easily fill its own report and is therefore beyond our scope.

Our goal is to provide an up-to-date report to the Computer Graphics community, as an entry point for researchers and developers who are new to the field, thereby complementing the existing survey paper [GM97].

2. Background

2.1. Continuum Elasticity

A deformable object is typically defined by its undeformed shape (also called equilibrium configuration, rest or initial shape) and by a set of material parameters that define how it deforms under applied forces. If we think of the rest shape as a continuous connected subset $M$ of $\mathbb{R}^3$, then the coordinates $m \in M$ of a point in the object are called material coordinates of that point. In the discrete case $M$ is a discrete set of points that sample the rest shape of the object.

When forces are applied, the object deforms and a point originally at location $m$ (i.e. with material coordinates $m$) moves to a new location $x(m)$, the spatial or world coordinates of that point. Since new locations are defined for all material coordinates $m$, $x$ is a vector field defined on $M$. Alternatively, the deformation can also be specified by the displacement vector field $u(m) = x(m) - m$ defined on $M$ (see Fig. 2). From $u(m)$ the elastic strain $\varepsilon$ is computed ($\varepsilon$ is a dimensionless quantity which, in the (linear) 1D case, is simply $\Delta l/l$). A spatially constant displacement field represents a translation of the object with no strain. Therefore, it becomes clear that strain must be measured in terms of spatial variations of the displacement field $u = u(m) = (u, v, w)^T$. Popular choices in Computer Graphics are

$$\varepsilon_G = \frac{1}{2}(\nabla u + [\nabla u]^T + [\nabla u]^T \nabla u) \quad (1)$$

$$\varepsilon_C = \frac{1}{2}(\nabla u + [\nabla u]^T), \quad (2)$$

where the symmetric tensor $\varepsilon_G \in \mathbb{R}^{3\times3}$ is Green’s nonlinear strain tensor and $\varepsilon_C \in \mathbb{R}^{3\times3}$ its linearization, Cauchy’s linear strain tensor. The gradient of the displacement field is a 3 by 3 matrix

$$\nabla u = \begin{bmatrix} u_x & u_y & u_z \\ v_x & v_y & v_z \\ w_x & w_y & w_z \end{bmatrix}, \quad (3)$$

where the index after the comma represents a spatial derivative.

The material law (or constitutive law) is used for the computation of the symmetric internal stress tensor $\sigma \in \mathbb{R}^{3\times3}$ for
each material point $m$ based on the strain $\epsilon$ at that point ($\sigma$ is measured as force per unit area, where $1\text{ Pascal} = 1\text{ Pa} = 1\text{ N}/\text{m}^2$). Most Computer Graphics papers use Hooke’s linear material law

$$\sigma = E \cdot \epsilon,$$  

where $E$ is a rank four tensor which relates the coefficients of the stress tensor linearly to the coefficients of the strain tensor. For isotropic materials, the coefficients of $E$ only depend on Young’s modulus and Poisson’s ratio.

2.2. Time Integration

In order to simulate dynamic deformable solids, we need to know the time dependent world coordinates $x(m, t)$ of all points in $M$. Given $x(2m, t)$, we can subsequently display the configurations $x(0), x(\Delta t), x(2\Delta t), \ldots$ resulting in an animation of the object. Here $\Delta t$ is a fixed time step of the simulation and $x(t)$ represents the entire vector field at time $t$.

The unknown vector fields $x(t)$ are not given directly but implicitly as the solution of a differential equation, namely Newton’s second law of motion of the form

$$x = F(x, x, t),$$  

where $x$ and $\dot{x}$ are the second and first time derivatives of $x$, respectively and $F(\cdot)$ a general function given by the physical model of the deformable object. In order to find the solution $x(t)$, this second order differential equation is often rewritten as a coupled set of two first order equations

$$\dot{x} = v$$  

$$v = F(x, x, t),$$  

where the new quantity $v$ represents $\dot{x}$. A discrete set of values $x(0), x(\Delta t), x(2\Delta t), \ldots$ of the unknown vector field $x$ which is needed for the animation can now be obtained by numerically solving (i.e., integrating) this system of equations. Numerical integration of ordinary differential equations is the subject of many textbooks (e.g., [PTVF92, AP98]). See [HES02] for an excellent overview in the context of deformable modeling in computer graphics. We give a few examples here which appear in subsequent sections.

The simplest scheme is explicit (or forward) Euler integration, where the time derivatives are replaced by finite differences $\dot{x}(t) = [x(t + \Delta t) - x(t)]/\Delta t$ and $\dot{v}(t) = [v(t + \Delta t) - v(t)]/\Delta t$. Substituting these into the above equations and solving for the quantities at the next time step $t + \Delta t$ yields

$$x(t + \Delta t) = x(t) + \Delta t \cdot v(t)$$  

$$v(t + \Delta t) = v(t) + \Delta t \cdot F(v(t), x(t), t).$$  

Time integration schemes are evaluated by two main criteria, their stability and their accuracy. Their accuracy is measured by their convergence with respect to the time step size $\Delta t$, i.e. first order $O(\Delta t)$, second order $O(\Delta t^2)$, etc. In the field of physically based animation in Computer Graphics, stability is often much more important than accuracy.

The integration scheme presented above is called explicit because it provides explicit formulas for the quantities at the next time step. Explicit methods are easy to implement but they are only conditionally stable, i.e. stable only if $\Delta t$ is smaller than a stability threshold (see [MHTG05] for a formalization). For stiff objects this threshold can be very small. The instability is due to the fact that explicit methods extrapolate a constant right hand side blindly into the future as the above equations show. For a simple spring and a large $\Delta t$, the scheme can overshoot the equilibrium position arbitrarily. At the next time step the restoring forces get even larger resulting in an exponential gain of energy and finally an explosion. This problem can be solved by using an implicit scheme that uses quantities at the next time step $t + \Delta t$ on both sides of the equation

$$x(t + \Delta t) = x(t) + \Delta t \cdot v(t + \Delta t)$$  

$$v(t + \Delta t) = v(t) + \Delta t \cdot F(v(t + \Delta t), x(t + \Delta t), t).$$  

The scheme is now called implicit because the unknown quantities are implicitly given as the solution of a system of equations. Now, instead of extrapolating a constant right hand side blindly into the future, the right hand side is part of the solution process. Remarkably, the implicit (or backward) Euler scheme is stable for arbitrarily large time steps $\Delta t$ (There is, however, a lower time step limit which, for practical purposes, poses no problem). This gain comes with the price of having to solve an algebraic system of equations at each time step (linear if $F(\cdot)$ is linear, non-linear otherwise).

A simple improvement to the forward Euler scheme is to swap the order of the equations and use a forward-backward scheme

$$v(t + \Delta t) = v(t) + \Delta t \cdot F(v(t), x(t), t)$$  

$$x(t + \Delta t) = x(t) + \Delta t \cdot v(t + \Delta t).$$  

The update to $v$ uses forward Euler, while the update to $x$ uses backward Euler. Note that the method is still explicit; $v(t + \Delta t)$ is simply evaluated first. For non-dissipative systems (i.e. when forces are independent of velocities), this reduces to the second order accurate Störmer–Verlet scheme. The forward-backward Euler scheme is more stable than standard forward Euler integration, without any additional computational overhead.

3. Lagrangian Mesh Based Methods

3.1. The Finite Element Method

The Finite Element Method (FEM) is one of the most popular methods in Computational Sciences to solve Partial Differential Equations (PDE’s) on irregular grids. In order to use the method for the simulation of deformable objects, the object is viewed as a continuous connected volume as
in Section 2.1 which is discretized using an irregular mesh. Continuum mechanics, then, provides the PDE to be solved. The PDE governing dynamic elastic materials is given by

$$\rho \ddot{x} = \nabla \cdot \sigma + f,$$  \hspace{1cm} (14)

where $\rho$ is the density of the material and $f$ externally applied forces such as gravity or collision forces. The divergence operator turns the 3 by 3 stress tensor back into a 3 vector

$$\nabla \cdot \sigma = \begin{bmatrix} \sigma_{xx,x} + \sigma_{yy,y} + \sigma_{zz,z} \\ \sigma_{yx,y} + \sigma_{xy,x} + \sigma_{zz,z} \\ \sigma_{zx,z} + \sigma_{zy,y} + \sigma_{zz,z} \end{bmatrix},$$  \hspace{1cm} (15)

representing the internal force resulting from a deformed infinitesimal volume. Eq. 14 shows the equation of motion in differential form in contrast to the integral form which is used in the Finite Volume method.

The Finite Element Method is used to turn a PDE into a set of algebraic equations which are then solved numerically. To this end, the domain $M$ is discretized into a finite number of disjoint elements (i.e. a mesh). Instead of solving for the spatially continuous function $x(m,t)$, one only solves for the discrete set of unknown positions $x_i(t)$ of the nodes of the mesh. First, the function $x(m,t)$ is approximated using the nodal values by

$$\tilde{x}(m,t) = \sum_i x_i(t)b_i(m),$$  \hspace{1cm} (16)

where $b_i(t)$ are fixed nodal basis functions which are 1 at node $i$ and zero at all other nodes, also known as the Kronecker Delta property (see Fig. 2). In the most general case of the Finite Element Method, the basis functions do not have this property. In that case, the unknowns are general parameters which can not be interpreted as nodal values. Substituting $\tilde{x}(m,t)$ into Eq. 14 results in algebraic equations for the $x_i(t)$. In the Galerkin approach [Hun05], finding the unknowns $x_i(t)$ is viewed as an optimization process. When substituting $x(m,t)$ by the approximation $\tilde{x}(m,t)$, the infinitely dimensional search space of possible solutions is reduced to a finite dimensional subspace. In general, no function in that subspace can solve the original PDE. The approximation will generate a deviation or residue when substituted into the PDE. In the Galerkin method, the approximation which minimizes the residue is sought. In other words, we look for an approximation whose residue is perpendicular to the subspace of functions defined by Eq. 16.

Many papers in Computer Graphics use a simple form of the Finite Element method for the simulation of deformable objects, sometimes called the explicit Finite Element Method, which is quite easy to understand and to implement (e.g. [OH99], [DDCB01], [MDM*02]). The explicit Finite Element Method is not to be confused with the standard Finite Element Method being integrated explicitly. The explicit Finite Element Method can be integrated either explicitly or implicitly.

In the explicit Finite Element approach, both, the masses and the internal and external forces are lumped to the vertices. The nodes in the mesh are treated like mass points in a mass-spring system while each element acts like a generalized spring connecting all adjacent mass points. The forces acting on the nodes of an element due to its deformation are computed as follows (see for instance [OH99]): given the positions of the vertices of an element and the fixed basis functions, the continuous deformation field $u(m)$ inside the element can be computed using Eq. 16. From $u(m)$, the strain field $\varepsilon(m)$ and stress field $\sigma(m)$ are computed. The deformation energy of the element is then given by

$$E = \int_V \varepsilon(m) \cdot \sigma(m) \, dm,$$  \hspace{1cm} (17)

where the dot represents the componentwise scalar product of the two tensors. The forces can then be computed as the derivatives of the energy with respect to the nodal positions. In general, the relationship between nodal forces and nodal positions is nonlinear. When linearized, the relationship for an element $e$ connecting $n_e$ nodes can simply be expressed as

$$f_e = K_e u_e,$$  \hspace{1cm} (18)

where $f_e \in \mathbb{R}^{3n_e}$ contains the $n_e$ nodal forces and $u_e \in \mathbb{R}^{3n_e}$ the $n_e$ nodal displacements of an element. The matrix $K_e \in \mathbb{R}^{3n_e \times 3n_e}$ is called the stiffness matrix of the element. Because elastic forces coming from adjacent elements add up at a node, a stiffness matrix $K \in \mathbb{R}^{3nx3n}$ for an entire mesh with $n$ nodes can be formed by assembling the element’s stiffness matrices

$$K = \sum_e K_e.$$  \hspace{1cm} (19)

In this sum, the element’s stiffness matrices are expanded to the dimension of $K$ by filling in zeros at positions related to nodes not adjacent to the element. Using the linearized elastic forces, the linear algebraic equation of motion for an entire mesh becomes ($u = x - x_0$)

$$Mu + Du + Ku = f_{\text{ext}},$$  \hspace{1cm} (20)

where $M \in \mathbb{R}^{3nx3n}$ is the mass matrix, $D \in \mathbb{R}^{3nx3n}$ the damping matrix.
matrix and \( f_{ext} \in \mathbb{R}^n \) externally applied forces. Often, diagonal matrices are used for \( M \) and \( D \), a technique called mass lumping. In this case, \( M \) just contains the point masses of the nodes of the mesh on its diagonal. The vectors \( x \) and \( x_0 \) contain, respectively, the actual and the rest positions of the nodes.

The Finite Element method only produces a linear system of algebraic equations if applied to a linear PDE. If a linear strain measure is used and Hooke’s law for isotropic materials is substituted into \( 14 \), Lamé’s linear PDE results:

\[
\rho x = \mu \Delta u + (\lambda + \mu) \nabla (\nabla \cdot u),
\]

(21)

where the material constants \( \lambda \) and \( \mu \) can be computed directly from Young’s modulus and Poisson’s ratio. This equation is solved in [DDBC99] in a multiresolution fashion. Using discretized versions of the Laplacian (\( \Delta = \nabla^2 \)) and gradient-of-divergence (\( \nabla (\nabla \cdot \cdot) \)) operators, they solve the Lamé equation on an irregular, multiresolution grid. The system is optimized for limited deformations (linearized strain) and does not support topological changes. Based on Gauss’ Divergence Theorem, the discrete operators are further evolved in [DDCB00], which leads to greater accuracy through defined error bounds. Furthermore, the cubic octree hierarchy employed in [DDBC99] is succeeded by a non-nested hierarchy of meshes, in conformance with the redefined operators, which leads to improved shape sampling. In [DDCB01] the previous linearized physical model is replaced by local explicit finite elements and Green’s nonlinear strain tensor. To increase stability the simulation is integrated semi-implicitly in time [DSB99].

O’Brien et al. [OH99] and [OBH02] present a FEM based technique for simulating brittle and ductile fracture in connection with elastoplastic materials. They use tetrahedral meshes in connection with linear basis functions \( b_i() \) and Green’s strain tensor. The resulting nonlinear equations are solved explicitly and integrated explicitly. The method produces realistic and visually convincing results, but it is not designed for interactive or real-time use. In addition to the strain tensor, they use the so-called strain rate tensor (the time derivative of the strain tensor), to compute damping forces. Other studies on the visual simulation of brittle fracture are [SWB00] and [MMDJ01].

Bro-Nielsen and Cotin [BNC96] use linearized finite elements for surgery simulation. They achieve significant speedup by simulating only the visible surface nodes (condensation), similar to the BEM. Many other studies successfully apply the FEM to surgery simulation, such as (but surely not limited to) [GT'T89], [CEO*93], [SBMH94], [KGC+96], [CDA99], [CDA00], [PDA00] and [PDA01].

As long as the equation of motion is integrated explicitly in time, nonlinear elastic forces resulting from Green’s strain tensor pose no computational problems. The nonlinear formulas for the forces are simply evaluated and used directly to integrate velocities and positions as in [OH99]. As mentioned earlier (Section 2.2), explicit integration schemes are stable only for small time steps while implicit integration schemes allow arbitrarily large time steps. However, in the latter case, a system of algebraic equations needs to be solved at every time step. Linear PDE’s yield linear algebraic systems which can be solved more efficiently and more stably than nonlinear ones. Unfortunately, linearized elastic forces are only valid for small deformations. Large rotational deformations yield highly inaccurate restoring forces (see Fig. 3).

To eliminate these artifacts, Müller et al. extract the rotational part of the deformation for each finite element and compute the forces with respect to the non-rotated reference frame [MG04]. The linear equation 18 for the elastic forces of an element (in this case a tetrahedron) is replaced by

\[
f = RK (R^T x - x_0),
\]

(22)

where \( R \in \mathbb{R}^{12 \times 12} \) is a matrix that contains four by 3 by 3 identical rotation matrices along its diagonal. The vector \( x \) contains the actual positions of the four adjacent nodes of the tetrahedron while \( x_0 \) contains their rest positions. The rotation of the element used in \( R \) is computed by performing a polar decomposition of the matrix that describes the transformation of the tetrahedron from the configuration \( x_0 \) to the configuration \( x \). This yields stable, fast and visually pleasing results. In an earlier approach, they extract the rotational part not per element but per node [MDM*02]. In this case, the global stiffness matrix does not need to be reassembled at each time step but ghost forces are introduced.

Another solution to this problem is proposed in [CGC’02]: each region of the finite element mesh is associated with the bone of a simple skeleton and then locally linearized. The regions are blended in each time step, leading to results which are visually indistinguishable from the nonlinear solution, yet much faster.

An adaptive nonlinear FEM simulation is described by Wu et al. [WDGT01]. Distance, surface and volume preservation for triangular and tetrahedral meshes is outlined in [THMG04]. Grinspun et al. introduce conforming, hierarchical, adaptive refinement methods (CHARMS) for general finite elements [GKS02], refining basis functions instead of elements. Irving et al. [ITF04] present a method which robustly handles large deformation and element inversion by

Figure 3: The pitbull with its inflated head (left) shows the artifact of linear FEM under large rotational deformations. The correct deformation is shown on the right.
computing a problem-specific diagonalization of the deformation gradient, from which the forces are derived. Müller et al. [MTG04] propose simulating and fracturing objects represented by surface meshes, by embedding the surface in a cuboid element voxelization (Fig. 4). This domain embedding strategy is also used by James et al. in their squashing cubes simulator [JBT04]. To support topological changes while maintaining well-shaped elements, Molino et al. create duplicates of the original elements in their virtual node algorithm [MBF04].

There is a more direct way to get to the nodal forces of an element, using the Method of Finite Volumes. The stress tensor \( \sigma \) can be used to compute the internal force \( f \) per unit area with respect to a certain plane orientation as

\[
f = \sigma \mathbf{n},
\]

where the normalized vector \( \mathbf{n} \) is the normal on that plane. Thus, the total force acting on the face \( A \) of a finite element is given by the surface integral

\[
f_A = \int_A \sigma \mathbf{n} dA.
\]

When linear basis functions are used, the stress tensor is constant within an element. Then, for planar element faces, the surface integral reduces to the simple product

\[
f_A = A \sigma \mathbf{n},
\]

where the scalar \( A \) is the area of the face and \( \mathbf{n} \) its normal. To get the nodal forces, one loops over all the faces \( A_f \) of the element and distributes the force \( f_A \) evenly among the nodes adjacent to that face. Teran et al. [TBHF03] use this method to simulate skeletal muscle. They also use a geometrically motivated way to compute strain which leads to an intuitive way of integrating the equations of motion.

### 3.4. The Boundary Element Method

The Boundary Element Method (BEM) is an interesting alternative to the standard Finite Element approach because all computations are done on the surface (boundary) of the elastic body instead of on its volume (interior). For a very good introduction to the subject see [Hunt05]. Roughly speaking, the integral form of the equation of motion is transformed into a surface integral by applying the Green-Gauss theorem. The method achieves substantial speedup because the three dimensional problem is reduced to two dimensions. However, the approach only works for objects whose interior is composed of a homogenous material. Also, topological changes (e.g. fractures) are more difficult to handle than in the explicit FEM approach where only local changes to the stiffness matrix or the connectivity of the elements need to be done.

In the ArtDefo System [JP99] surface nodes are simulated using the Boundary Element Method and a database of precomputed reference boundary value problems (RBVPs). By employing a fast update process, which exploits coherence, accurate real-time deformation simulation is achieved. In [JP02b] the RBVPs are expressed in terms of linear elastostatic Green’s functions (LEGFMs) and multiple RBVP’s are linked via interface boundary conditions, resulting in a multizone elastokinematic model, which properly simulates large nonlinear relative strains. The system is further augmented by multiresolution Green’s functions (wavelet Green’s functions) in [JP03], with which large-scale objects can be simulated.

© The Eurographics Association 2005.
3.5. Mass-Spring Systems

Mass-spring systems are arguably the simplest and most intuitive of all deformable models. Instead of beginning with a PDE such as Eq. 14 and subsequently discretizing in space, we begin directly with a discrete model. As the name implies, these models simply consist of point masses connected together by a network of massless springs.

The state of the system at a given time $t$ is defined by the positions $x_i$ and velocities $v_i$ of the masses $i = 1..n$. The force $f_i$ on each mass is computed due to its spring connections with its neighbors, along with external forces such as gravity, friction, etc. The motion of each particle is then governed by Newton’s second law $f_i = m_i\ddot{x}_i$, which for the entire particle system can be expressed as

$$M\ddot{x} = f(x, v)$$

(26)

where $M$ is a $3n \times 3n$ diagonal mass matrix. Thus, mass-spring systems "only" require the solution of a system of coupled ordinary differential equations (ODEs). This is done via a numerical integration scheme as described in Section 2.2.

![Figure 5: A mass-spring system.](image)

Depicted in Fig. 5 is a simple example of a volumetric mass-spring system (similar to that in [CZKM98]). The masses are regularly spaced in a lattice. They are connected together along the twelve edges of each hexahedron by "structural" springs. Masses on opposite corners of each hexahedron are connected together by "shear" springs. These springs cause the solid to resist longitudinal and shear deformations respectively, and their rest lengths define the undeformed state of the body. Typically, the spring types have different properties; and for anisotropic materials each springs’ properties also depend on its orientation.

Springs are commonly modeled as being linearly elastic; the force acting on mass $i$, generated by a spring connecting $i$ and $j$ together is

$$f_i = k_s(|x_{ij} - l_{ij}|)\frac{x_{ij}}{[x_{ij}]}$$

(27)

where $x_{ij}$ is the difference between the two masses’ position vectors $(x_j - x_i)$, $k_s$ is the spring’s stiffness and $l_{ij}$ is its rest length.

Physical bodies are not perfectly elastic; they dissipate energy during deformation. To account for this, viscoelastic springs are used to damp out relative motion. Thus, in addition to (27), each spring exerts a viscous force. It is common to model this as

$$f_i = k_d(v_j - v_i)$$

(28)

where $k_d$ is the spring’s damping constant. This is unfortunate, as it damps rigid body rotations. Worse still, when modeling cloth it damps bending and wrinkling — a key feature of these objects. It is preferable to use

$$f_i = k_d(\frac{v_j^T x_{ij}}{x_{ij}^T x_{ij}})x_{ij}$$

(29)

where $v_{ij} = v_j - v_i$. This projects the velocity difference onto the vector separating the masses, and only allows a force along that direction.

In the literature, the concept of a mass-spring system is often more general than the canonical example given above. These models are still represented by point masses and a fixed connectivity, but the notion of a spring is generalized. Often the name "particle system" is used instead, which is perhaps a more fitting name. (Though they should be distinguished from the models found in Section 4.1, which have no fixed connectivity.) In any case, the term "mass-spring system" is very suggestive, and will most likely persist.

In more general particle systems, such as those found in [BHW94, BW98, THMG04], deformation energies are defined. These are often derived from soft constraints that are to be maintained in the model; given some constraint of the form $C(x) = 0$, an associated energy is defined as

$$\frac{1}{2}C^T(x)C(x).$$

These energies are minimized at the model’s rest state, and are used to enforce the preservation of mesh distances, angles, areas, volumes, etc. Particle forces are then computed as the derivatives of the energies with respect to the particle positions

$$f_i = \frac{\partial E}{\partial x_i}.$$  

(30)

Each energy term typically involves only a few particles. In the case of a distance constraint, we are back to our simple, linear spring model described above. However, for other constraints and energies, we must imagine more general spring types: angular, bending, areal, volumetric, etc.

3.5.1. Early Work

Mass-spring networks first saw use in Computer Graphics for facial modeling [PBS81, Wai87]. These early works solve static problems of the form (18). Soon after, dynamic models were introduced to model skin, fat and muscle [CHP89, TW90, WT91].

The locomotion of simple creatures such as snakes,
worms and fish [Mil88, TT94] was simulated using mass-spring systems. In these systems, spring rest-lengths vary over time to simulate muscle actuation.

In work by Terzopoulos et al. [TPF89] the equations of thermal conductivity are used to simulate heat transfer in a volumetric mass-spring system. Each spring’s stiffness is determined by its temperature, which is set to zero once the melting threshold is exceeded. A discrete fluid model (similar to those presented in Section 4.1) is applied to particles for which all connecting springs have melted.

Breen et al. [BHW94] presented a particle system to model cloth. They posited that cloth is a mechanism of warp and weft fibres — not a continuum — and that mass-spring systems are thus more appropriate than Finite Element techniques for modeling cloth. Using measured data from the Kawabata Evaluation System [Kaw80], they took care in formulating their energy functions for stretching, bending and trellising (shearing), and predicted the static shape of real materials quite accurately. Since [BHW94], particle systems have dominated the cloth simulation literature — although new FEM formulations (such as the one described in [EKS03]) continue to be proposed.

3.5.2. Drawbacks and Improvements

Particle systems tend to be intuitive and simple to implement. Moreover, they are computationally efficient and handle large deformations with ease. However, unlike the Finite Element and finite difference methods, which are built upon elasticity theory, mass-spring systems are not necessarily accurate. Primarily, most such systems are not convergent; that is, as the mesh is refined, the simulation does not converge on the true solution. Instead, the behavior of the model is dependent on the mesh resolution and topology. In practice, spring constants $k_s$ are often chosen arbitrarily, and one can say little quantitatively about the material being modeled. In addition, there is often coupling between the various spring types. For instance, the "shear" springs of the model in Fig. 5 also resist stretching. For medical applications, as well as for virtual garment simulation in the textile industry, greater accuracy is required. For applications such as film and games, this lack of accuracy may be acceptable; convincing animations have been produced using these systems. However, a judicious choice of model is still advised, as the behavior of the material being modeled can be highly mesh dependent.

Several researchers have explored and have tried to mitigate the accuracy issues in mass-spring systems. Kass [Kas95] presents a simple equivalence in one dimension between a mass-spring system and a corresponding finite difference spatial discretization. Eischen and Biglami [HB00] perform a comparison between a Finite Element model and a carefully tuned particle system, which gave similar experimental results for small deformations. Etzmuss et al. [EGS03] carefully derive a cloth particle system from a continuous formulation by a finite difference discretization. Their model is convergent, and they show how to choose spring constants based on continuous material parameters.

These more accurate particle systems, however, only apply to rectangular meshes. Like finite difference techniques, they do not generalize easily to triangular (or tetrahedral) meshes. Volino and Thalmann [VMT97] present a triangular mesh which attempts to model the physical properties of cloth regardless of edge orientations. However, the accuracy of their method is unclear. Baraff and Witkin [BW98] model cloth using a triangular mesh, deriving in-plane particle forces from a continuum formulation. Their approach supports anisotropic behavior, but is not convergent. In [BC00], the authors present a novel mass-spring system which gives consistent results (though it is not convergent) for tetrahedral and hexahedral meshes, independent of the orientation of the elements. In their system, springs emanate from the barycenter of each element along principle coordinate axes and are attached to element faces; forces acting on each particle are then computed via interpolation. Their approach also allows the user to control anisotropic material behavior. Teschner et al. [THMG04] model volumes and surfaces using tetrahedral and triangular meshes respectively. They employ generalized springs which preserve distances, areas and volumes. Their model is efficient, convergent (though it does not incorporate continuous material properties), and supports plastic deformation.

Bending models for surfaces tend to be particularly ad-hoc; Bridson et al. [BMF03] present a physically correct bending model for triangle meshes by isolating the bending mode from all other modes of deformation (Fig. 7). Grinspun et al. present a similar bending model for the simulation of discrete shells [GHDS03].

An interesting approach is to use learning algorithms to search for mesh parameters. (See [BTH’03, BSSH04] and references therein.) Bhat et al. use simulated annealing to estimate the spring constants in a cloth mesh. They employ Baraff and Witkin’s model [BW98], and use video of real cloth as their reference for comparison. Bianchi et al. use a genetic algorithm to identify spring constants as well as mesh topology in a volumetric mass-spring system. They
compare their results to FEM reference solutions, and obtain close correspondence. While such methods are general, they are computationally expensive and require reference solutions; moreover, the results of a specific search may not hold for different mesh resolutions.

Adaptive meshing techniques are particularly complicated by the non-convergent behavior of many mass-spring systems, as they require modeling at multiple resolutions. Another difficulty is the handling of so-called T-junctions between regions of differing resolutions, though this problem must also be addressed in FEM and other methods.) A number of authors have addressed this issue for cloth simulation. In these systems, splitting/merging criterion are often based on bending angles between triangles. Villard and Borouchaki [VB02] apply adaptive refinement to a simple, convergent model based on quadrilateral meshes. They show consistent results between meshes of multiple resolutions, maintaining rich detail at reduced computation times. Earlier work by Hutchinson et al. [HPI96] takes a similar approach, though the convergence of their model is not clear as they couple their spatial and time discretizations together in an ad-hoc manner. Li and Volkov [LV05] present an adaptive framework for triangular cloth meshes, avoiding the problematic T-junctions. They present attractive results, though their physical model is not convergent.

Finally, a number of researchers have sought to improve the realism of particle systems by modeling nonlinear material properties. For instance, Eberhardt et al. [EWS96] base their spring model on measured cloth data to model hysteresis effects; Choi and Ko [CK02] approximate cloth’s buckling response using a fifth order polynomial.

### 3.5.3. Time Integration

Once a force model is chosen, the particle system is stepped forward in time via a numerical integration scheme as in Section 2.2. Time integration schemes have received particular attention in the cloth simulation literature (see [HES02]). As cloth is often modeled by mass-spring systems, we further discuss the subject of time integration here.

Cloth is a highly anisotropic material due to its structure: it resists bending weakly, but has a relatively strong resistance to stretching. When simulating a highly stiff volume of material, one can employ a reduced deformation model, such as modal analysis (see Section 5) — and in the limit, use a rigid body model. Clearly, this cannot be done with cloth. We are interested in visualizing the large scale, out-of-plane folding and wrinkling of cloth, however we must still deal with these planar energies in our model. Numerically speaking, the resulting ODE system (26) is stiff; that is, it possesses a wide range of eigenvalues. This forces us to take excessively small time steps when using an explicit scheme.

Provot [Pro95] tackles the issue of stiffness by using much weaker stretching energies and then post-processing the cloth mesh at each time step, iteratively enforcing constraints. Springs that are stretched by more than 10% are relaxed (shortened). This in turn stretches nearby neighboring springs which are then relaxed, and so on until convergence is obtained. This, in effect, also models the biphasic nature of cloth — small deformations are resisted weakly until a threshold is reached, whereupon stiffness dramatically increases. In practice, this method gives efficient, attractive results, though convergence is not guaranteed.

Baraff and Witkin [BW98] present a linear implicit scheme which allows for large time steps while maintaining stability. This has proven to be a robust and efficient solution to the stiffness problem, and has become a popular technique. Specifically, they apply a backward Euler scheme to Eq. 26, giving:

$$\begin{bmatrix} \Delta x \\ \Delta v \end{bmatrix} = \Delta t \begin{bmatrix} v_n + \Delta v \\ M^{-1} f(x_n + \Delta x, v_n + \Delta v) \end{bmatrix}$$  \hspace{1cm} (31)

where \(x_n = x(t)\) and \(v_n = v(t)\). This is a nonlinear equation in \(\Delta x\) and \(\Delta v\). A linear implicit version of (31) is obtained by using a first order Taylor series expansion of \(f\):

$$f(x_n + \Delta x, v_n + \Delta v) = f_n + \frac{\partial f}{\partial x} \Delta x + \frac{\partial f}{\partial v} \Delta v$$  \hspace{1cm} (32)

where \(\frac{\partial f}{\partial x}\) and \(\frac{\partial f}{\partial v}\) are the Jacobian matrices of the particle forces with respect to position and velocity respectively. In their paper, Baraff and Witkin derive expressions for the Jacobians of the various internal forces in their model. Due to the local connectivity structure of the mesh, these are sparse matrices. They then solve the resulting linear system at each time step

$$A \Delta v = b,$$  \hspace{1cm} (33)

where

$$A \equiv (I - \Delta t M^{-1} \frac{\partial f}{\partial v} - \Delta t^2 M^{-1} \frac{\partial^2 f}{\partial v^2})$$  \hspace{1cm} (34)

and  \(b \equiv \Delta t M^{-1} f_n + \Delta t \frac{\partial f}{\partial v} v_n\).  \hspace{1cm} (35)

They solve this system, in the presence of constraints, using a modified preconditioned conjugate gradient solver (MPCG). This is equivalent to applying one Newton iteration to Eq. 31. Although solving this system increases the
computational cost at each time step, this is more than offset by the ability to take large steps when desired. However, as illustrated by Desbrun et al. [DSB99], the larger the time step size, the more artificial damping is added to the system.

Others have since attempted to improve upon Baraff and Witkin’s approach. Desbrun et al. [DSB99] precompute the linear part of A, achieving an $O(n)$, unconditionally stable scheme. Their technique, however, is inaccurate and does not generalize well to large systems. Kang et al. [KCC’00] improve upon this approximation, but ultimately are only replacing the CG iterations of [BW98] with a single, Jacobi-like iteration. Volino and Magenat-Thalmann [VMT00] use a weighted implicit-midpoint method that appeared to give attractive results, but which is less stable and may be difficult to tune in practice. Choi and Ko [CK02] use the more accurate second-order backward difference formula (BDF2); moreover, their coupled compression/bending model improves the stability of the system, eliminating the need for fictitious spring damping.

Instead of applying an implicit scheme to the entire system, a number of researchers have employed implicit-explicit (IMEX) schemes [ARW95]. The essential idea is to separately treat the stiff and non-stiff parts of the ODE, handling the stiff parts with an implicit method and the non-stiff parts with an explicit method. This combines the stability of an implicit scheme where needed with the simplicity and efficiency of an explicit scheme where possible. Hauth et al. [HES02] base their IMEX splitting on connection type: stretch springs are handled implicitly, whereas shear and bend "springs" are handled explicitly. This simplifies implementation and improves the sparsity of A — and thus the cost of solving Eq. 33. The authors also use BDF2, and embed their linear solver within a Newton solver, making theirs a fully (as opposed to linear) implicit technique. Boxerman and Ascher [BA04] take a similar approach, optimizing their IMEX splitting adaptively based on a local stability criterion. Each spring is handled either explicitly or implicitly, based on time step, local mesh resolution and material properties. They then take further advantage of the improved system sparsity to decompose the mesh into components which can be solved more efficiently (Fig. 8).

Bridson et al. [BFM03] apply a similar IMEX splitting to cloth as that used for advection-diffusion equations [ARW95], applying an implicit method to the damping term and an explicit method to the elastic term. This is appropriate when damping plays a significant role, as damping imposes quadratic time step restrictions with respect to mesh size (as opposed to a linear one for elastic terms). They incorporate this within a second order accurate integration scheme, and include a strain limiting step. Coupled with their careful handling of bending and collisions, they produced visually stunning cloth animations.

4. Lagrangian Mesh Free Methods

4.1. Loosely Coupled Particle Systems

Particle systems were developed by Reeves [Ree83] for explosion and subsequent expanding fire simulation in the movie "Star Trek II: The Wrath of Khan". The same technique can also be used for modeling other fuzzy objects such as clouds and water, i.e. objects that do not have a well-defined surface. Particles are usually graphical primitives such as points or spheres, however, they might also represent complex group dynamics such as a herd of animals [Rey87]. Each particle stores a set of attributes, e.g. position, velocity, temperature, shape, age and lifetime. These attributes define the dynamical behavior of the particles over time and are subject to change due to procedural stochastic processes. Particles pass through three different phases during their lifetime: generation, dynamics and death.

In [Ree83], particles are points in 3D space which represent the volume of an object. A stochastic process generates particles in a predetermined generation shape which defines a region about its origin into which the new particles are randomly placed. Attribute values are either fixed or may be determined stochastically. Initially, particles move outward from the origin with a random speed. During the dynamics phase, particle attributes might change as functions of both time and attributes of other particles. Position is updated by simply adding the velocity. Finally, a particle dies if its lifetime reaches zero or if it does not contribute to the animation anymore, e.g. if it is outside of a region of interest. A particle is rendered as a point light source which adds an amount of light depending on its color and transparency attribute. Motion blurring is very easy to achieve by rendering a particle as a streak using its position and velocity. An advantage of particles is their simplicity, which enables the animation of a huge number of particles for complex scenes. The procedural definition of the model and its stochastic control simplifies the human design of the system. Furthermore, with particle hierarchies, complicated fuzzy objects such as clouds can be assembled and controlled. Although the particles are simulated omitting inter-particle forces, the resulting animations are convincing and fast for inelastic phenomena. Therefore, the technique has been widely employed in movies and video games. Examples of modeling waterfalls, ship wakes, break-

© The Eurographics Association 2005.
ing waves and splashes using particle systems can be found in [FR86, Pea86, Gos90, Sim90, OH95].

Particles which interact with each other depending on their spatial relationship are referred to as spatially coupled particle systems [Ton92]. The interactions between particles evolve dynamically over time, thus, complex geometry and topological changes can be easily modeled using this approach. Tonnesen presented a framework for physically based animation of solids and liquids based on dynamically coupled particles which represent the volume of an object [Ton91, Ton98]. Each particle \( p_i \) has a potential energy \( \phi_i \) which is the sum of the pairwise potential energies \( \phi_{ij} \) between \( p_i \) and all other other particles \( p_j \), i.e.

\[
\phi_i = \sum_{j \neq i} \phi_{ij}.
\]

The force \( f_i \) exerted on \( p_i \) with position \( x_i \) is then

\[
f_i = -\nabla_{x_i} \phi_i = -\sum_{j \neq i} \nabla_{x_i} \phi_{ij}.
\]

So far, all particles interact with each other, resulting in \( O(n^2) \) complexity where \( n \) is the number of particles. The computational costs for computing the force can be reduced to \( O(n) \) when restricting the interaction to a neighborhood within a certain distance, and \( O(n \log n) \) for neighbor searching. To avoid discontinuities at the neighborhood boundary, the potentials are weighted with a continuous, smooth and monotonically decreasing weighting function which depends on the distance to the particles and ranges from one at the particle position to zero at the neighborhood boundary.

For deriving inter-particle forces, the Lennard-Jones potential \( \phi_{LJ} \) is used:

\[
\phi_{LJ}(d) = \frac{B}{d^n} - \frac{A}{d^m},
\]

where \( d \) is the distance between two particles, and \( n, m, B \) and \( A \) are constants. The Lennard-Jones potential is well known in molecular dynamics for modeling the interaction potential between pairs of atoms. It creates long-range attractive and short-range repulsive forces, yielding particles arranged into hexagonally ordered 2D layers in absence of external forces. A more convenient formulation, called the Lennard-Jones bi-reciprocal function, is written as

\[
\phi(d) = -\frac{e_0}{m-n}(m\frac{d_0}{d})^n - n\frac{d_0}{d} (m),
\]

where \( d_0 \) is the equilibrium separation distance, and \(-e_0\) is the minimal potential (the magnitude is called the dissociation energy). Increasing the dissociation energy increases the stiffness of the model, while with the exponents \( n \) and \( m \) the width of the potential well can be varied. Therefore, large dissociation energy and high exponents yield rigid and brittle material, while low dissociation energy and small exponents result in soft and elastic behavior of the object. This allows the modeling of a wide variety of physical properties ranging from stiff to fluid-like behavior. By coupling the dissociation energy with thermal energy such that the total system energy is conserved, objects can be melted and frozen. Furthermore, thermal expansion and contraction can be simulated by adapting the equilibrium separation distance \( d_0 \) to the temperature.

One problem of particle systems is that the surface is not explicitly defined. Blinn [Bli82] proposed using an algorithm which was developed to model electron density maps of molecular structures. A Gaussian potential \( \phi_i(x) = b e^{-a d^2} \) (which is often called a blob) is assigned to each particle, where \( a \) and \( b \) are constants and \( d = ||x - x_i|| \) is the distance from an arbitrary point \( x \) in space to the particle’s position \( x_i \). A continuously defined potential field \( \psi(x) \) in space is obtained by summing the contribution from each particle

\[
\psi(x) = \sum_i \psi_i(x).
\]

The surface is then defined as an iso-value \( S \) of \( \psi \). This yields an implicit coating of the particle which handles topological changes such as splitting and merging by construction. For a more intuitive control of the surface, the constants \( a \) and \( b \) can be computed as \( a = -B/R^2 \) and \( b = Sc^{-B} \), where \( R \) is the radius in isolation and \( B \) the blobiness.

The implicit coating of particles for soft inelastic substances undergoing topological changes pose challenging problems such as volume preservation, avoiding unwanted blending and contact modeling. These were addressed by Desbrun and Cani in a series of papers [Can93, DC94, DC95]. A hybrid model is used which is composed of two layers: Particles are used to simulate soft inelastic substances as described above, while an elastic implicit layer defines the surface of an object and is locally deformed during collisions. A problem of the implicit coating is that the volume may change significantly during deformation, especially for splitting and merging. However, efficiently computing the volume of a soft substance is not trivial. A territory of a particle \( p_i \) is defined as the (volumetric) part \( V_i \) of the object where the field contribution of \( p_i \) is the highest. Note that territories form a partition of the implicit volume of an object. Each particle samples its territory boundary by sending a fixed number of points called seeds in a set of distributed directions until they reach the boundary. The volume of a particle is approximated by simply summing up the distances from the particle to the seeds. The local volume variation can then be easily approximated for each particle, and the field function is changed accordingly such that the volume is preserved. Another problem is that split object parts might blend with each other when they come close. To avoid this, an influence graph is built at each animation step by recursively adding the neighbors of a particle which are in its sphere of influence to the same influence connected component. Only the particles of the same component can interact and their field functions are...
produces both the programming and computational complexity is trivially conserved and convection is dispensable. This directly from the state equations. Furthermore, as a particle-compute approximate values of needed physical quantities technique where discrete, smoothed particles are used to Smoothed Particle Hydrodynamics (SPH) is a Lagrangian collision of granular materials with rigid bodies, or even be-

Molecular dynamics (MD) is used to compute contact forces granular materials, such as sand and grains, using a parti-

tation yields forces acting on the positions and torques of the particle energy with respect to its position and ori-

control the bending and stiffness of the surface. Variation of the particle energy with respect to its position and ori-

tentials yield the energy of a particle, where the weights control the bending and stiffness of the surface. Variation of the particle energy with respect to its position and ori-
aton yields forces acting on the positions and torques acting on the orientations, respectively. Using these forces and torques, the Newtonian equations for linear and angular motion are solved using explicit integration as described in Section 2.2.

Recently, Bell et al. presented a method for simulating granular materials, such as sand and grains, using a particle system [BYM05]. A (non-spherical) grain is composed of several round particles, which move together as a single rigid body. Therefore, stick-slip behavior naturally occurs. Molecular dynamics (MD) is used to compute contact forces for overlapping particles. The same contact model is used for collision of granular materials with rigid bodies, or even between rigid-bodies, by simply sampling the rigid body surface with particles.

4.2. Smoothed Particle Hydrodynamics

Smoothed Particle Hydrodynamics (SPH) is a Lagrangian technique where discrete, smoothed particles are used to compute approximate values of needed physical quantities and their spatial derivatives. Forces can be easily derived directly from the state equations. Furthermore, as a particle-based Lagrangian approach it has the advantage that mass is trivially conserved and convection is dispensable. This reduces both the programming and computational complexity and is therefore suitable for interactive applications. A drawback of SPH is that it is not possible so far to exactly maintain the incompressibility of material.

Following, we will give a short introduction of the SPH method and discuss applications in Computer Graphics. For a more detailed introduction we refer the reader to the excellent paper of Monaghan [Mon92].

A function A is interpolated at a position x from its neighboring particles using a summation interpolant:

\[ A(x) = \sum_j m_j \frac{A_j}{\rho_j} W(r, h), \]

where \( A_j \) denotes the function value of a particle \( p_j \) at \( x_j \) (the interpolation point), \( m_j \) and \( \rho_j \) are the mass and density of a particle \( p_j \), respectively, and \( r = x - x_j \). \( W(r, h) \) is a smoothing kernel with the properties \( \int W(r, h) dr = 1 \) and \( \lim_{h \to 0} W(x, h) = \delta(x) \), where \( h \) is the support radius and \( \delta \) the Dirac-function. For efficiency reasons, \( h \) is usually chosen such that the kernel \( W(r, h) \) has compact support. The choice of the smoothing kernel \( W(r, h) \) is very important. Often spline Gaussian kernels are used, see e.g. [Mon92] for a discussion of kernels. If \( W(r, h) \) is differentiable, a differentiable interpolant of a function can be derived by simply computing the gradient of \( W(r, h) \), i.e.,

\[ \nabla A(x) = \sum_j m_j \frac{A_j}{\rho_j} \nabla W(r, h) \]

Therefore, there is no need for finite differences or a grid. To obtain higher accuracy the interpolant can be computed as

\[ \rho \nabla A(x) = \nabla (\rho A(x)) - A(x) \nabla \rho. \]

The density is estimated at an arbitrary point as

\[ \rho(x) = \sum_j m_j W(r, h). \]

The particle densities could be computed for each time step. In practice, this is often not appropriate because the density drops close to the object boundary. Furthermore, difficulties arise when adapting the spatial resolution of the particles [DC99]. Instead, we can assign an initial density to each particle. The continuity equation (conservation of mass) is then used for computing the variation of density over time

\[ \dot{\rho}_i = -\rho_i \nabla v_i, \]
where the divergence of velocity $\nabla v_i$ is approximated by

$$\nabla v_i = \frac{1}{\rho_i} \sum_{j \neq i} m_j (v_j - v_i) \cdot \nabla W(r, h). \quad (46)$$

The disadvantage of this update of the density relative to the motion of the neighboring particles is that exact mass conservation is not guaranteed.

The equations of motion are solved by deriving forces. E.g. the pressure force can be estimated using Eq. 43

$$f^\text{pressure}_i = -m_i \sum_j \frac{m_j}{\rho_j^2} \left( \frac{p_j}{\rho_j^2} + \frac{P_i}{\rho_i^2} \right) \nabla W(r, h), \quad (47)$$

where the force is symmetrized to fulfill the action-reaction principle of Newton’s third law. For keeping a constant rest where the force is symmetrized to fulfill the action-reaction.

E.g. the pressure force can be estimated using Eq. 43

$$f^\text{pressure}_i = -m_i \sum_j \frac{m_j}{\rho_j^2} \left( \frac{p_j}{\rho_j^2} + \frac{P_i}{\rho_i^2} \right) \nabla W(r, h), \quad (47)$$

where the force is symmetrized to fulfill the action-reaction principle of Newton’s third law. For keeping a constant rest density $\rho_0$, the pressure $P_i$ is computed by a variation of the ideal gas state equation [DC96]

$$P_i = k(\rho - \rho_0), \quad (48)$$

where $k$ is a gas constant.

For smoothed particles, local stability criteria have been found which, in conjunction with time-adaptive integration, yield both stable and efficient animations. The Courant-Friedrichs-Lewy (CFL) criterion requires that each particle $i$ must not be passed by, giving a limit for the time step $\Delta t \leq \lambda h/c$, where $\lambda$ is the Courant number and $c$ is the speed of sound of the material, which is the maximum velocity of a deformation wave inside the material. Pressure waves propagate at speed $c = \sqrt{\partial P/\partial \rho}$, therefore using Eq. 48 induces $c = \sqrt{k}$. Considering viscosity further decreases the maximum time step, see [Mon92] and [DC96] for details.

Generally, the animation quality and accuracy increases with a higher number of particles. In [DC99], an adaptive framework is proposed where space and time resolutions are chosen automatically. Individual particles are refined where pressure differences are above a user-defined threshold. A refined particle $p_i$ with mass $m_i$ and volume $m_i/\rho_0$ is replaced by $n$ particles with smaller volume and mass $m_i/n$ such that they occupy the same volume as $p_i$. Considering the particles as spheres with radius proportional to their support radius, an individual support radius $h_i$ can be computed as $h_i = \bar{\rho} \sqrt{m_i/\rho_0}$, where the constant $\bar{\rho}$ is chosen according to the desired average number of neighboring particles. Neighboring particles can be merged in stable areas where pressure is almost uniform. However, since a particle is isotropic, the neighboring particles should approximately fill a sphere. The new particle is positioned at the center of gravity of the replaced ones. Its mass and velocity are computed such that mass and linear momentum are conserved.

Considering the object as a set of smeared-out masses, the surface can be defined as an iso-surface of the mass density function [DC96]. This yields an implicit representation coherent with the physical model. Desbrun and Cani [DC98] improve this implicit coating of particles by using an active surface which evolves depending on a velocity field, similar to snakes [KWT88]. Therefore, surface tension and other characteristics such as constant volume can be added to the surface model.

### 4.2.1. Applications

SPH was introduced independently by Gingold and Monaghan [GM77] and Lucy [Luc77], for the simulation of astrophysical problems such as fission of stars. Stam and Fiume introduced SPH to the Computer Graphics community to depict fire and other gaseous phenomena [SF95]. They solve the advection-diffusion equation for densities composed of “warped blobs”. Desbrun and Cani solve the state equations for the animation of highly deformable bodies using SPH [DC96]. They achieve the animation of inelastic bodies with a wide range of stiffness and viscosity. Lava flows are animated by coupling viscosity with a temperature field and simulated heat transfer between the particles [SAC99]. By considering hair as a fluid-like continuum, Hadap and Magnenat-Thalmann [HMT01] used a modified formulation of SPH to simulate hair-hair interactions. Premož et al. introduced the use of the moving particle semi-implicit method (MPS) for simulating incompressible multiphase fluids [PTB03]. Impressive visual results were produced by coupling the physical particles with level sets for surface reconstruction. Müller et al. presented a method based on SPH and new smoothing kernels, with which fluids with free surfaces can be simulated at interactive rates with up to 5000 particles [MCG03]. Furthermore, they proposed a term to model surface tension effects. In [MST04], the interaction of Lagrangian fluids and mesh-based deformable solids are modeled by placing virtual boundary particles, so-called ghost particles [Mon94], on the surface of the solid objects according to Gaussian quadrature. Recently, the SPH method was extended in [MSKG05] so that the simulation of phenomena such as boiling water, trapped air and the dynamics of a lava lamp are possible (Fig. 9). Liquids with different polarities are simulated by
computing a body force which acts perpendicular to the interface of two liquids. The force is proportional to the curvature of the interface and the interface tension. Additionally, air particles are generated and deleted dynamically where air pockets are likely to be formed, making it possible to simulate trapped air.

4.3. Mesh Free Methods

Interesting and fairly new approaches to deformable modeling are the so-called meshfree methods for the solution of partial differential equations, which originated in the FEM community approximately a decade ago ([NTVR92], [Suk03]). For an extensive and up-to-date classification and overview of mesh-free methods, see [FM03], [Liu02] and also the excellent review paper [BKO96]. A nice introduction to the element-free Galerkin method is given in [Ask97].

4.3.1. Point-based Animations

Recently, the combination of mesh-free physics with point-sampled surfaces [PZvBG00] in so-called point-based animations [pbaf04] have become popular. Müller et al. introduced to the graphics community a mesh-free continuum-mechanics-based framework for the animation of elastic, plastic and melting objects [MKN04]. Elastic body forces are derived via the strain energy density

$$ U = \frac{1}{2} \varepsilon \cdot \sigma, $$

where $\varepsilon$ is the strain and $\sigma$ the stress tensor as described in Section 3.1. The elastic force per unit volume at a particle $p_i$ with material coordinates $\mathbf{m}_i$ is computed using the directional derivative $\nabla_{\mathbf{m}_i}$. For a Hookean material, i.e. $\sigma = E \cdot \varepsilon$ (see Eq. 4), this yields

$$ f_i = -\nabla_{\mathbf{m}_i} U = -\frac{1}{2} \nabla_{\mathbf{m}_i} (\varepsilon_i \cdot E \cdot \varepsilon_i) = -\sigma \cdot \nabla_{\mathbf{m}_i} \varepsilon_i, $$

where Green’s nonlinear strain tensor (see Eq. 1) is used for $\varepsilon_i$. For computing $\varepsilon_i$, the spatial derivatives of the displacement field $\nabla \mathbf{u}$ at $\mathbf{m}_i$ is needed. To guarantee zero elastic forces for rigid body modes, the approximation of $\nabla \mathbf{u}$ from the displacement vectors $\mathbf{u}_j$ of the neighboring particles must be at least first order accurate. Therefore, the Moving Least Squares method [LS81] with a linear basis is used which yields first order accurate interpolation of point-sampled functions (contrary to the SPH approximation described in Section 4.2 which is not even zeroth-order accurate). For the following, we will only consider the x-component $u$ of the displacement field $\mathbf{u} = (u, v, w)^T$. The basic idea is to approximate a continuous scalar field $u(\mathbf{m})$ using a Taylor approximation. For particles $p_j$ close to a particle $p_i$, we get a first order approximation $\tilde{u}_j$ of the values $u_j$ as

$$ \tilde{u}_j = u_i + \nabla u(\mathbf{m}_i) \cdot \mathbf{m}_{ij}, $$

where $\mathbf{m}_{ij} = \mathbf{m}_j - \mathbf{m}_i$ and $\nabla u(\mathbf{m}_i) = (u_x, u_y, u_z)^T$ at $\mathbf{m}_i$. The index after the coma denotes a spatial derivative. The weighted least-squares error $e_i$ of the approximation $\tilde{u}_j$ is given by

$$ e_i = \sum_j (\tilde{u}_j - u_j) w_{ij}, $$

where $w_{ij}$ is a normalized, continuously defined and monotonically decreasing weighting function. We want to find the unknowns $u_x, u_y$ and $u_z$ such that the error $e$ is minimized. Therefore, we set the derivatives of $e$ with respect to $u_x, u_y$ and $u_z$ to zero, yielding three equations for the three unknowns

$$ \left( \sum_j \mathbf{m}_j \mathbf{m}_j^T w_{ij} \right) \nabla u|_{\mathbf{m}_i} = \sum_j (u_j - u_i) \mathbf{m}_j w_{ij}. $$

Finally, we obtain the spatial derivatives of $u(\mathbf{m})$ at $\mathbf{m}_i$ as

$$ \nabla u|_{\mathbf{m}_i} = A^{-1} \left( \sum_j (u_j - u_i) \mathbf{m}_j w_{ij} \right), $$

where the inverse of $A = \sum_j \mathbf{m}_j \mathbf{m}_j^T w_{ij}$ needs to be computed only once per particle and can be used for computing the derivatives of $v$ and $w$ as well.

With this approach, material properties ranging from stiff elastic to highly plastic can be simulated. In [KAG05], the authors describe how to extend this range to viscous materials such as fluids by merging the solid mechanics equation with the Navier-Stokes equations. Viscosity, pressure and surface tension forces are computed using SPH as described in Section 4.2 and added to the elastic force. Note that this allows the simulation of non-realistic materials such as elastic fluids, similar to [GBO04]. The material properties such as stiffness, viscosity, plasticity and cohesion can be defined per particle. By coupling these properties to the temperature of a particle, local melting and freezing of objects can be achieved (Fig. 10).

In the case of elastic materials which do not undergo

Figure 11: Highly plastic deformations and ductile fracture [PKA05].
any topological changes, the surface can be simply dragged along with the particles [MKN’04]. This is done by computing the displacement for each surface element (surfel) from the displacements of neighboring particles \( p_j \). The first order approximation of \( \nabla u \) at \( m_j \) can be used to obtain a smooth displacement vector field which is invariant under linear transformations:

\[
\mathbf{x}(\mathbf{m}_i) = \mathbf{m}_i + \sum_j \omega(\mathbf{m}_i, \mathbf{m}_j)(\mathbf{u}_j + \nabla \mathbf{u}|_{\mathbf{m}_i}(\mathbf{m}_j - \mathbf{m}_i)),
\]

where \( \mathbf{m}_i \) and \( \mathbf{x}(\mathbf{m}_i) \) are the material and world coordinates of a surfel \( s_i \), respectively, and \( \omega \) is a normalized weighting kernel. The deformation is computed and applied to both the center and the axes of a surfel. To adapt the sampling of the surface in case of stretching or compression, surfels can be split or merged similar to [PKKG03]. The animation of the physics particles combined with this surface skin allows the simulation of elastic and plastic materials of detailed models in real-time. Furthermore, Adams et al. [AKP’05] showed that the surface can be efficiently raytraced by exploiting the deformation field for bounding hierarchy updates.

However, dragging the surface along with the particles fails if topological changes occur. In this case, an implicit representation can be used as discussed in the Section 4.2. In [MKN’04], this implicit representation is sampled with surfels. After an animation step the new position of the surfels is estimated using the displacement computation in Eq. 55. The deformed surfels are then projected onto the implicit surface. Finally, a resampling and relaxation scheme ensures that the surface is completely covered and uniformly sampled with surfels. In [KAG’05], the implicit representation is improved by applying energy potentials and geometric constraints which give better control over the surface. However, as the particle representation of the volume is usually much coarser than the surface representation, the implicit coating of the particles cannot represent fine detail of the surface. A possible solution is to use the explicit deformation approximation of (55) for solids (with low temperature), the implicit representation for fluids (with high temperature), and blend between the explicit and the implicit representation depending on the temperature for melting object parts. This way, surface detail smoothly fades out and the melting surface approaches the implicit representation. Freezing fluids are handled analogously.

The framework is extended in [PKA’05] to cope with fracturing (Fig. 11). Cracks are created at surfels where the main principal stress exceeds a threshold. The crack propagates in a plane perpendicular to the main principal stress, where the crack surface is dynamically sampled with surfels. Visibility tests detect neighboring particles, which are separated by a crack surface. A splitting, merging and termination operator handles the topological events which occur when cracks merge or branch. Furthermore, a resampling scheme adapts the particles resolution to handle small fractured object pieces. With the suggested methods a wide range of materials can be fractured, from stiff elastic to highly plastic objects that exhibit brittle and/or ductile fracture.

Collision detection of point-based models has been discussed for rigid [KZ04], quasi-rigid [PPG04] and deformable objects [KMH’04]. For rigid collision detection, a bounding sphere hierarchy is built with the surfels as leaves, which is suitable for time-critical collision detection [KZ04]. Pauly et. al [PPG04] compute a consistent contact surface and the traction distribution by solving a Linear Complementarity Problem (LCP). Collision response forces are computed from the local tractions and integration yields the total wrench that acts on the quasi-rigid bodies. Keiser et al. propose to decouple the collision handling and deformation to achieve stable and efficient contact handling for deformable objects [KMH’04]. Collisions are detected using a high resolution surface sampled with surfels and a simple scheme is used to compute a consistent contact surface. A penalty force depending on its displacement onto the contact surface is computed for each surfel. These forces are then distributed to the neighboring particles where they are applied as external collision response forces in the next animation step.

Wicke et al. propose a method for animating point-sampled thin shells [WSG05]. The geometric surface properties are approximated using splines (so-called fibres) embedded in the surface. Physical effects such as plasticity and fracturing can be easily modeled. The authors also describe a skinning technique for point-sampled surfaces, with which a high-resolution surface can be animated using a sparse set of samples.

### 4.3.2. Meshless Deformations Based on Shape Matching

Müller et al. [MHTG05] propose a meshless method for animating deformable objects which does not fit into any of the preceding categories. The nodes of a volumetric mesh are treated as point masses and animated as a simple particle system without connectivity. Then, at every time step, the original configuration of the points (the rest state mesh) is fitted to the actual point cloud in the least squares sense, using shape matching techniques for point clouds with correspondence. The fitted rest shape yields goal positions for all point masses. Each point mass is then pulled towards its goal position while the displacement divided by the time step is added to the velocity of the point. This geometric integration scheme, although explicit, is shown to be unconditionally stable (Fig. 12).

Figure 12: Stability and the ability to recover from highly deformed or inverted configurations, shown in [MHTG05].
5. Reduced Deformation Models and Modal Analysis

5.1. Basic Formulation

As described by James and Pai [JP04], given \( N \) undeformed point locations \( p = [p_1, \ldots, p_N] \), a reduced deformation model approximates deformed point locations \( p' \) by a linear superposition of \( M \) displacement fields (the columns \( U_i \) of \( U \) in Eq. (56)). The amplitude of each displacement field is given by the reduced coordinates \( q \) such that
\[
p' = p + Uq. \tag{56}
\]
as shown in Fig. 13. In [JP04] it is stated that the columns of \( U \) could arise from any possibly nonlinear black box process, such as an interpolation process, multiresolution modeling or nonlinear/linear modal analysis. The latter will be the focus of this section.

\begin{figure}[h]
\centering
(a) \hspace{1cm} (b) \hspace{1cm} (c) \hspace{1cm} (d)
\caption{Reduced deformation models: (a) reference shape \( p \), (b) displacement field \( U_1 \), (c) displacement field \( U_2 \), and (d) one possible deformed shape \( p' = p + U_1 + 0.5U_2 \).
\end{figure}

5.2. Linear Modal Analysis

Pentland and Williams [PW89] pioneered the use of reduced deformable models in Computer Graphics, using modal analysis. Given the mass, damping and stiffness matrices \( M, C \) and \( K \), Eq. (20) can be decoupled into \( N = 3n \) linearly independent ordinary differential equations (ODEs) by solving the generalized eigenvalue problem
\[
M\Phi\Lambda = K\Phi, \tag{57}
\]
such that \( \Phi^T M\Phi = I \) and \( \Phi^T K\Phi = \Lambda \) are diagonal matrices. The entries of \( \Lambda \) contain the eigenvalues, and the columns of \( \Phi = [\Phi_1, \Phi_2, \ldots, \Phi_N] \) contain the eigenvectors of \( M^{-1}K \). \( \Phi \) is often termed the modal matrix or modal displacement matrix, where the \( i \)-th column \( \Phi_i \) represents the \( i \)-th mode shape, and the \( i \)-th entry of \( \Lambda \) is proportional to the resonant frequency of the \( \Phi_i \). The columns of \( \Phi \) form a basis (the modal basis, or eigenbasis) of \( 3n \) dimensional space, so any displacement \( u(t) = x(t) - x_0 \) can be written as a linear combination of the columns
\[
u(t) = \Phi q(t). \tag{58}
\]
The vector \( q(t) \) contains the modal coordinates (or modal amplitudes). Substituting Eq. (58) into Eq. (20) and premultiplying by \( \Phi^T \) yields
\[
\Phi^T M\Phi q + \Phi^T C\Phi q + \Phi^T K\Phi q = \Phi^T f_{ext}. \tag{59}
\]
Note that for general damping, \( \Phi^T C\Phi \) is dense, but if we assume proportional (Raleigh) damping \( C = \alpha M + \beta K \), then \( \Phi^T C\Phi = \alpha I + \beta \Lambda \) is also diagonal (taking into consideration that \( M, C \) and \( K \) are normally symmetric positive definite).

With \( g = \Phi^T f_{ext} \) we can now express Eq. (59) as \( 3n \) independent scalar 2nd order differential equations
\[
\ddot{q}_i + (\alpha + \beta \lambda_i)q_i + \lambda_i q_i = g_i, \tag{60}
\]
where \( \lambda_i \) is the \( i \)-th diagonal element of \( \Lambda \).

With this decoupling, one can examine the eigenvalues, discard high frequency mode shapes, and thereby only use dominant modes (i.e. dominant columns \( \Phi_i \)). This can significantly reduce computational cost. Furthermore, the motion components \( q_i \) of the individual modes can now be computed independently and combined by linear superposition.

In detail, each of these equations (60) has an analytical solution of the form
\[
q_i = c_1 e^{\alpha t} + c_2 e^{\beta t}, \tag{61}
\]
where the constants \( c_1 \) and \( c_2 \) depend on the initial conditions \( (q_i \) and \( \dot{q}_i \) at time \( t = t_0 \)), and the roots of the characteristic equation \( r^2 + (\alpha + \beta \lambda) r + \lambda = 0 \) are
\[
r_{1,2} = -\frac{(\alpha + \beta \lambda_i) \pm \sqrt{\alpha + \beta \lambda_i}^2 - 4\lambda_i}{2}, \tag{62}
\]
The solution depends on the sign of \( \nu = (\alpha + \beta \lambda_i)^2 - 4\lambda_i \): \( \nu > 0 \), \( \nu = 0 \), \( \nu < 0 \) produces the overdamped (\( r_1, r_2 \) real and different), critically damped (\( r_1, r_2 \) real but repeated) and underdamped case (\( r_1, r_2 \) are complex conjugates) respectively. If \( r_1, r_2 \) are real then \( c_1, c_2 \) are also real, if \( r_1, r_2 \) are complex conjugate pairs then so will be \( c_1, c_2 \). In any case, \( q_i \) in Eq. (61) will be a real value.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure14.pdf}
\caption{Underdamped (solid, blue) and critically damped (dashed, red) vibration. The black (dash-point) line is the plot of \( \pm e^{-\nu t}/2 \).
\end{figure}

Most interest is given to the underdamped case, where damped oscillations occur, and for which Eq. (61) can be written as
\[
q_i = e^{-(\alpha + \beta \lambda_i)/2}(c_1 \cos \mu t + c_2 \sin \mu t), \tag{63}
\]
\[
\mu = \frac{\sqrt{\nu} - (\alpha + \beta \lambda_i)^2}{2}, \tag{64}
\]
see Fig. 14. This can be used to compute the response of a mode to an external force at some time \( t = t_0 \) (see [HSO03]): first, the generalized force \( f_{ext} \) must be transformed to modal...
coordinates by $\Delta \mathbf{g} = \Delta \mathbf{r}^T \mathbf{f}_{\text{ext}}$, and then $c_1, c_2$ can be computed by substituting Eq. 63 into Eq. 60 and setting $q_0 = 0$ and $\dot{q}_0 = \Delta \mathbf{g}$ [HSO03]. Naturally, the system of decoupled ODEs (Eq. 60) can also be numerically integrated through time.

5.3. Applications and Other Reduced Models

Interestingly, modal analysis was introduced to Computer Graphics in 1989, but only relatively few significant contributions have been presented since then. The first work of note is Stam’s application of modal analysis to the simulation of tree branches subjected to turbulence [Sta97].

James and Pai [JP02a] map runtime dynamics to graphics hardware: in a precomputation stage they build a so-called dynamic response texture (DyRT), where mode shapes $\Phi_i$ and other quantities are stored (see Fig. 15). At runtime, the modal coordinates $\mathbf{q}$ are computed from rigid bone transforms or external excitations. Finally, displacements are applied to the undeformed mesh by computing $\mathbf{u} = \sum \Phi_i \dot{q}_i$ for a few dominant low frequency modes using a vertex program running on graphics hardware.

Whereas enforcing direct manipulation and collision constraints is relatively straightforward with node positions in Euclidean space, applying these in a modal basis can be somewhat unintuitive. Hauser et al. [HSO03] provide a solution, where generalized forces are computed for constrained nodes using the modal basis. Since the force computation involves evaluating a pseudoinverse using the singular value decomposition (SVD), only few constraints (up to ten in their examples) can be applied in a real-time simulation environment. Furthermore, they couple the deformable model with a rigid body reference frame and simulate dynamics, collisions and friction.

In the above methods, a linear Cauchy strain model is employed to obtain the stiffness matrix $K$, resulting in well known artifacts for large rotational deformations away from the rest shape (see Section 5.1 and Fig. 15). To suppress these artifacts, Choi and Ko [CK05] identify per-node rotations and extend the basic modal analysis formulation (Section 5.2) to accommodate these rotations, similar in spirit to the warped stiffness approach [MDM’02]. Although their modal warping procedure is not guaranteed to perform well for large deformations, their results show visually convincing improvements over the standard linear model (Fig. 16).

The decoupled ODEs are solved using semi-implicit numerical integration, instead of applying the analytical solution of Eq. 61.

Barbić and James [BJ05] take a different approach: instead of employing Cauchy’s linearized strain (Eq. 2) they use full quadratic Green strain (Eq. 1) throughout the entire computation, thereby necessitating the solution of a nonlinear version of Eq. 20 per time step. The otherwise computationally burdening implicit Newmark integration is greatly accelerated by carrying it out in reduced coordinates (also known as subspace integration), and thereafter performing the matrix multiplication of Eq. 56 in the fragment shader of current graphics hardware. For the generation of an appropriate reduced deformation basis $\mathbf{U}$, which contains sufficient nonlinear deformation, they provide a fully automatic method based on modal derivatives (Fig. 17), and also show how user-defined (force) sketches can assist a full unreduced offline static solver in precomputing a suitable $\mathbf{U}$. Note that due to the nonlinear coupling of modes, using independent harmonic oscillators as previously described (Eq. 61) does not suffice.

Figure 16: Evolution of one mode shape in linear modal analysis (top row) and modal warping (bottom row). Image courtesy of Min Gyu Choi, Kwangwoon University.

Figure 17: Using subspaces based on the first few dominant vibration modes and their derivatives is sufficient to describe large nonlinear deformations [BJ05].
A somewhat different model reduction strategy is devised by James and Fatahalian [JF03]: they entirely precompute collision, contact and dynamics (as well as co-parameterized global illumination) of deformable objects by limiting the range of interactions to a finite set of user impulses. The dynamic responses, termed impulse response functions (IRFs), each a chronological set of generalized deformation vectors, are dimensionality reduced using principal component analysis (PCA). At runtime, IRFs are blended based on user interaction.

6. Eulerian and Semi-Lagrangian Methods

There are two general points of view we can take when describing an object that we wish to physically simulate. The main distinctions of the two view points is how they discretize an object in order to work with it numerically, and how they define the boundary of the object itself. So far, this report has covered the Lagrangian point of view. As Section 2 explains, the Lagrangian point of view describes an object as a set of moving points (material coordinates) that travel around and change position over time; these points carry their material properties with them as they move through the world. Lagrangian techniques are a convenient way to define an object as a connected mesh of points (Sec-

A somewhat different model reduction strategy is devised by James and Fatahalian [JF03]: they entirely precompute collision, contact and dynamics (as well as co-parameterized global illumination) of deformable objects by limiting the range of interactions to a finite set of user impulses. The dynamic responses, termed impulse response functions (IRFs), each a chronological set of generalized deformation vectors, are dimensionality reduced using principal component analysis (PCA). At runtime, IRFs are blended based on user interaction.

Eulerian and Semi-Lagrangian Methods

There are two general points of view we can take when describing an object that we wish to physically simulate. The main distinctions of the two view points is how they discretize an object in order to work with it numerically, and how they define the boundary of the object itself. So far, this report has covered the Lagrangian point of view. As Section 2 explains, the Lagrangian point of view describes an object as a set of moving points (material coordinates) that travel around and change position over time; these points carry their material properties with them as they move through the world. Lagrangian techniques are a convenient way to define an object as a connected mesh of points (sec-

These two equations represent the conservation of mass and momentum for an incompressible fluid. The vector field $u_t$ is the time derivative of the fluid velocity. The scalar pressure field is $p$, and $v$ is the kinematic viscosity. The vector field $f$ represents the body force per unit mass (usually just gravity). In this Eulerian formulation the quantities of the fluid are stored in a grid of cells (a regular grid of cubes or voxels if you like). The velocity is stored on the cell faces and the pressure is stored at the center of the cells. This is commonly called a staggered (or MAC) grid. Notice that the position, $x$, of the fluid is not defined; in this Eulerian frame work, the grid positions remain fixed.

Solving equations 65 and 66 is usually done in several steps (breaking up a PDE into simpler pieces and solving the pieces separately is known as operator splitting). The $f$ term is simply scaled by the time step and added to the current velocity. The advection term, $-(u \cdot \nabla)u$, is then solved. One popular way to solve the advection term is with the semi-

The Eulerian approach to fluids was popularized by a series of papers by Foster and Metaxas [FM96], [FM97b], and [FM97a]. Their formulation solves the fluid (Navier-Stokes) equations on a regular voxel grid and uses a finite difference formulation for all the differential equations. To understand the Eulerian framework, and how it differs from a Lagrangian one, we will look briefly at how the Navier-Stokes equations (equations of motion for a fluid) are formulated and solved for a single time step. We refer the reader to [Car04] for an in-depth discussion of liquids, and to [Sta03] for a practical implementation of a smoke simulator.

There are two parts to the Navier-Stokes equations:

$$\nabla \cdot u = 0 \tag{65}$$

$$u_t = -(u \cdot \nabla)u + \nabla \cdot (v \nabla u) - \nabla p + f \tag{66}$$

† In this formulation we assume a constant density and absorb it into the pressure term for simplicity.
The quantity that we desire to update is then interpolated from the new location and that interpolated value is used as the updated quantity. It is interesting to note that if the time step is scaled so that the path traverses only a single cell, then the semi-Lagrangian technique is a first order upwind technique. After advection, the viscosity term is solved. If the viscosity is expected to be large then an implicit technique should be used (see Section 2.2). At this point we have what can be called a best guess velocity:

$$\tilde{u} = u + \Delta t [-(u \cdot \nabla)u + \nabla \cdot (\nu \nabla u)]$$

which does not take into account pressure or mass conservation (Eq. 65). The final step, commonly called a pressure-projection, is in fact one of the reasons that Eulerian techniques are popular for fluids. The trick of the pressure-projection step is to solve for a pressure such that subtracting that pressure’s gradient from the best guess velocity will give a final velocity that conserves mass. So, we know two things: first, the final velocity, \(u^{new}\), must contain the pressure term which is missing from Eq. 67, so we add it back in to get

$$u^{new} = \tilde{u} - \Delta t \nabla p$$

Second, we know the final velocity has to conserve mass, so we plug Eq. 68 into Eq. 65 to get

$$\nabla \cdot u^{new} = \nabla \cdot \tilde{u} - \Delta t \nabla \cdot (\nabla p) = 0$$

Rearranging Eq. 69 gives us the following Poisson equation

$$\Delta t \nabla^2 p = \nabla \cdot \tilde{u}$$

with which we must solve for \(p\). We then plug \(p\) back into Eq. 68 to complete the pressure projection and find our final velocity. The final velocity is incompressible (divergence free), and for a constant density fluid this ensures mass conservation. Compressible fluids can also conserve mass, but their density must change to do so (see [YOH00]).

The system of equations formed by Eq. 70 is symmetric and positive definite (as long as there is at least one known Dirichlet boundary condition), and can be solved with a conjugate gradient method. Although the above explanation assumes a regular grid, it can also be carried out on an adaptively refined grid, such as an octree [SYO02, LGF04].

In the rest of this section we will explore the numerous types of deformable objects that are modeled within an Eulerian framework. As the techniques are described we will also describe how the standard Navier-Stokes equations are changed in each case, and how the objects are represented.

In the early work of Foster and Metaxas [FM96] liquid is displayed either as a height field or a collection of massless particles. These massless particles are different from the material coordinate particles used in a Lagrangian simulation because they are passively moved in the velocity field defined by the fluid, and the particles’ velocity is interpolated from the grid. Foster and Metaxas render liquid particles as smoothed ellipsoids [FM97a] with an orientation based on the velocity of the particle and a normal computed from the position of the other nearby particles. Foster and Metaxas also define smoke as massless particles [FM97b]; they advect and diffuse an added temperature field on the grid, and use that temperature to define thermal buoyancy in the smoke to create turbulent motion.

Stam [Sta99] uses a scalar density field to define quantities of smoke. These smoke densities are diffused on the grid and advected with the semi-Lagrangian technique (which Stam did not know about at the time, so when he published his seminal work he called the semi-Lagrangian technique the method of characteristics). Stam also adds buoyancy forces based on the local smoke density, and Fedkiw et al. [FSJ01] add a vorticity confinement term to artificially increase small scale swirling motion which is lost when using semi-Lagrangian advection on coarse grids. Kim et al. [KLLR05] used the semi-Lagrangian technique with back and fourth error compensation and correction to keep both small and large scale swirling motion in the fluid while maintaining stability. Their technique greatly reduces the dissipation present in the semi-Lagrangian technique, while maintaining its stability.

Carlson et al. [CMVT02] add a temperature and variable viscosity field to model solid objects that can melt into liquid (Fig. 19). In their model, a solid object is simply a fluid with very high viscosity, and as the temperature rises the viscosity decreases, and the object melts into a liquid state. Like [FM96], the object in their simulation is defined by a set of massless particles, but instead of just rendering the particles they use a splatting technique to extract a mesh from the particles for rendering. Their method also animates the building of a sand castle, but only high viscosity is used to model the mud.

More recently, Zhu and Bridson [ZB05] model sand as a
continuum (fluid) on an Eulerian grid (Fig. 20). They use a Lagrangian technique to move particles, then interpolate the velocity from the nearby particles onto to the Eulerian grid. They then use the grid velocity to solve the pressure projection step. Thus, their technique combines the strengths of both the Lagrangian and Eulerian techniques. After solving the Navier-Stokes equations, they add in a model for sand. At each grid cell they identify the strain rate (Eq. 2) and use it to calculate stress forces in that cell. All connected groups of cells which can resist the inertia trying to make them slide have their velocity projected to rigid body motion. The velocity in all other cells get forces created by sliding frictional stress added to them. Their method also describes a way to build an implicit function from the particles for rendering.

Carlson et al. [CMT04] project rigid bodies on an Eulerian grid to model their interaction with a fluid (Fig. 18). A rigid body solver is run in tandem with a fluid solver: at each frame, the rigid bodies that are in contact with fluid are projected onto the fluid, where their equations of motion are solved with a Navier-Stokes equation that is modified with buoyancy and collision forces. After solving the fluid equation, they project the velocity in the rigid body cells to rigid body motion (effectively setting the strain rate to zero inside the cells that contain rigid bodies) and use that velocity for the next step of the rigid body solver.

Foster and Fedkiw [FF01] were the first in graphics to model the boundary of a fluid as a level set. A level set, in this context, is an extra scalar field, \( \phi \), that stores the signed distance to the fluid surface at each grid point. The change of \( \phi \) is computed each step with the level set equation,

\[
\phi_t + u \cdot \nabla \phi = 0, \tag{71}
\]

which effectively updates the position of the iso-contour that delineates the fluid's surface. The level set method has become one of the most popular ways to define a fluid's surface. It has the drawback of defining the surface only at Eulerian grid points, and thus rounding off high resolution details, but this drawback can be reduced by using the particle level set technique [EMF02], which patches the level set values with a group of particles that is passively advected with the surface. The level set can be directly ray traced with a root finder, or a mesh can be extracted from it with marching cubes. We would like to take note of a common confusion when discussing level sets and fluids. Level sets are not used to solve the fluid equations; they are used to define the boundary of the fluid (often called surface tracking), so that we know where to solve the fluid equations.

Another way to define a fluid surface is with an explicit polygonal mesh. A mesh representation not only allows a direct way to render the surface, but can also keep track of surface properties like texture coordinates. It can be difficult to deal with topology changes, warping, and self intersection in a mesh if a single mesh is used. The contouring method described by Bargteil et al. [BGOS05] defeats these difficulties by re-defining a mesh each frame of the fluid simulation. They also maintain a signed distance field and show volume conservation on par with the particle level set method with octree refinement.

Goktekin et al. [GB004] animate viscoelastic substances (that display characteristics of both a liquid and a solid) by adding elastic terms to the basic Navier-Stokes equations (Fig. 21). The elastic terms are controlled by von Mises's yield condition and a quasi-linear plasticity model. They keep track of a strain tensor field that is advected throughout the fluid grid, and its components are stored at edge centers (off-diagonal components) and cell centers (diagonal components). The surface of their fluid is defined with a particle level set.

Feldman et al. [FOK05] use an Eulerian grid based technique to model smoke, but they use an unstructured mesh, tetrahedra instead of cubes, near the boundary of complicated objects. Their hybrid mesh formulation allows them to model the boundary with curved or highly detailed objects in a more accurate fashion. In their paper they render smoke as a set of massless particles.

7. Conclusion

Physically based deformable models have seen wide application in many fields of Computer Graphics, and research and development efforts are as active and fruitful as ever. As pointed out throughout this paper, no one model is suited best for any given application. Instead, many parameters and considerations need to be taken into account, such as model representations, the range of physical parameters, topological changes, real-time or interactive simulation, and so on.
And even though such a seemingly exhaustive toolkit exists, with which very impressive results have been achieved, we still have quite a way to go until we can plausibly and interactively simulate the natural phenomena in our everyday lives. To quote Richard Feynman [FLS63]

"The things with which we concern ourselves in science appear in myriad forms, and with a multitude of attributes. For example, if we stand on the shore and look at the sea, we see the water, the waves breaking, the foam, the sloshing motion of the water, the sound, the air, the winds and the clouds, the sun and the blue sky, and light; there is sand and there are rocks of various hardness and permanence, color and texture. There are animals and seaweed, hunger and disease, and the observer on the beach; there may be even happiness and thought. Any other spot in nature has a similar variety of things and influences. It is always as complicated as that, no matter where it is. Curiosity demands that we ask questions, that we try to put things together and try to understand this multitude of aspects as perhaps resulting from the action of a relatively small number of elemental things and forces acting in an infinite variety of combinations."  

We think it is safe to say that this statement still holds, and probably will hold for years to come, a burden and a blessing alike to researchers in the field.

Yet even with the current methodology, the algorithms and models have seen somewhat limited application in production environments and video games. One reason for this is the lack of computational power: many of the production environments and video games. One reason for this is the lack of computational power; many of the current techniques are inherently offline, and can take hours or days to produce results. In the field of interactive entertainment, small physically based deformable models are already being implemented, and thanks to developments such as the upcoming physics processing unit (PPU), this trend is likely to carry on. A second reason we see for a reluctance to adopt these new methods is the limited control the users have over the resulting animations. Although this has been previously addressed, we see great potential in coupling more intuitive user interfaces with physically based simulations.

Acknowledgements. We would like to thank the following people for their contributions, suggestions and general help in shaping and improving this paper: Bart Adams, Adam Bargteil, Robert Bridson, Min Gyu Choi, Nico Galoppo and Doug James.

References


© The Eurographics Association 2005.
[CKO2] CHOI K. J., KO H. S.: Stable but responsive cloth. In SIGGRAPH '92 (pp. 604–611). 9, 10
[HES02] HAUPT M., ETZMUS O., STRASSER W.: Analysis of numerical methods for the simulation of deformable models. The Visual Computer (2002). Accepted for publication. 3, 9, 10


Las87 Lasseter J.: Principles of traditional animation applied to 3D computer animation. In SIGGRAPH ’87 (1987), pp. 55–44. 15


L05 Lin M., Otaduy M.: Recent advances in haptic rendering and applications. SIGgraph Course Notes (2005). 2


Lac77 Lucy L. B.: A numerical approach to the testing of the fission hypothesis. The Astronomical Journal 82, 12 (1977), 1013–1024. 13

LV05 Li L., Volkov V.: Cloth animation with adaptively refined meshes. In ACSC (2005), pp. 107–114. 9


MSK05 Müller M., Solenthaler B., Keiser R., Gross M.: Particle-based fluid-fluid interaction. ACM. to appear. 12, 13


Nealen, Müller, Keiser, Boxerman and Carlson / Physically Based Deformable Models in Computer Graphics


[Sis06] SISI K.: Particle animation and rendering using data parallel computation. In SIGGRAPH ’96 (1996), pp. 405–413. 11


[Ton92] TONNESEN D.: Spatially coupled particle systems. 4.1 – 4.21. 11


© The Eurographics Association 2005.