Eurographics Symposium on Geometry Processing (2005), pp. 1–10
M. Desbrun, H. Pottmann (Editors)

Streaming Simplification of Tetrahedral Meshes

Submission ID: 114

Abstract

Unstructured tetrahedral meshes are commonly used in scientific computing to represent scalar, vector, and tensor fields in three dimensions. Visualization of these meshes can be difficult to perform interactively due to their size and complexity. By reducing the size of the data, we can accomplish real-time visualization necessary for scientific analysis. We propose a two-step approach for streaming simplification of large tetrahedral meshes. Our algorithm arranges the data on disk in a streaming, I/O-efficient format that allows coherent access to the tetrahedral cells. A quadric-based simplification is sequentially performed on small portions of the mesh in-core. Our output is a coherent streaming mesh, which facilitates future processing. Our technique is fast, produces high quality approximations, and operates out-of-core to process meshes too large for main memory.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling

Keywords: Out-of-core algorithms, mesh simplification, large meshes, tetrahedral meshes

1. Introduction

Simplification techniques have been a major focus of research for the past decade due to the increasing size and complexity of geometric data. Scientific simulations and measurements from fluid dynamics and partial differential equation solvers have produced data sets that are too large to visualize with current hardware. Thus, approximations are necessary to achieve a level of interactivity that is necessary for proper analysis.

Although significant work has been done in simplifying triangle meshes, relatively little has been done with tetrahedral meshes. Most of the work in tetrahedral simplification falls into two categories: Edge-collapse methods and point sampling methods. These algorithms assume that the entire mesh can be loaded into main memory. However, due to the high memory overhead of storing the mesh connectivity in addition to the geometry, there are limitations on the size of the data set that can be simplified in this manner.

We present an algorithm that *streams* the data from disk through memory and performs the simplification on a localized portion of the entire mesh. Our approach consists of two steps. First, the tetrahedral mesh is arranged in a streaming format that supports coherent sequential access. Then, this streaming mesh is sequentially simplified using a quadric error-based scheme that respects boundaries and fields in the mesh. The resulting mesh is output in the same streaming format and can be used directly in subsequent processing.

Our streaming algorithm requires only one pass to simplify the entire mesh. Thus, the layout of the mesh is of great importance to produce high quality results. We perform an out-of-core ordering of the tetrahedral cells and store them on disk using a streaming tetrahedral mesh format. This format provides concurrent access to coherently ordered vertices and tetrahedra. It also minimizes the duration that a vertex remains in-core, which limits the memory footprint of the simplification.

Our tetrahedral simplification incrementally works on overlapping portions of the mesh in-core. We use the quadric-error metric to perform a series of edge collapses until a target decimation is reached. By weighting the boundaries and incorporating the field data in our error metric, we can keep the error in the simplified approximation low. This results in a simplification algorithm that can efficiently simplify extremely large data sets. In addition, through the use of carefully optimized algorithms, linear solvers, and data structures we show that significant improvements in speed and stability can be achieved over previous techniques.

The main contributions of this paper include:

- We provide a streaming representation of tetrahedral

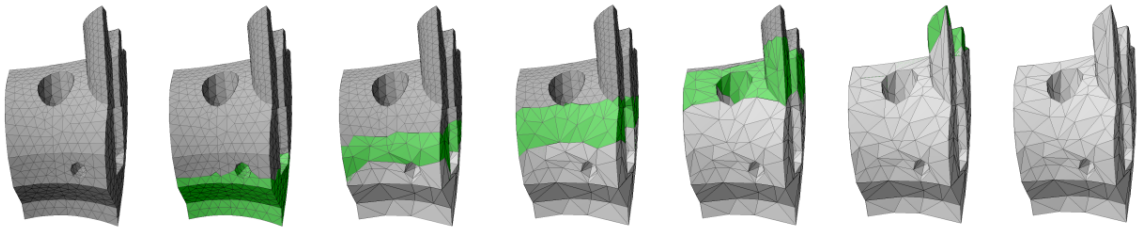

Figure 1: *Streaming simplification performed on a tetrahedral mesh ordered from bottom to top. The portion of the mesh that is in-core at each step is shown in green.*

meshes that supports coherent sequential access to the mesh elements.

- We describe a quadric-based edge-collapse simplification algorithm that operates on portions of the streaming tetrahedral mesh. This operation occurs in a single pass, runs quickly, handles data of arbitrary size, respects field data, and works out-of-core.
- We improve upon previous stream simplification methods by ensuring that the output stream is coherent in order to accommodate further downstream processing. We also introduce optimizations in the data structures and simplification algorithm which dramatically improve the speed and efficiency of tetrahedral simplification.
- We provide a new solver for quadric-based simplification that improves stability and speed of existing algorithms. We also provide both stability and error analysis of the results generated using this technique.

The remainder of this paper is organized as follows. We summarize related work in Section 1.1. In Section 2, we describe our algorithm for arranging the data in a coherent, streaming mesh. Section 3 provides details on our out-of-core simplification, Section 4 contains our stability and error analysis followed by performance measures, Section 5 discusses the benefits of our approach over previous algorithms, and Section 6 provides final remarks and directions for future work.

1.1. Related Work

A common result from scientific computations is a scalar field f in $\mathbb{R}^3$. This scalar field f can be represented over a domain D as a tetrahedral mesh. When it is not possible to achieve interactive visualization of f, it is common to find a tetrahedral mesh with fewer elements and an associated scalar field f' such that the approximation error $\|f' - f\|$ is minimized. Many algorithms have been proposed in an attempt to compute f' quickly and with little error.

Trotts *et al.* [THJW98, THJ99] developed a technique that collapses one edge at a time, deciding which edge to collapse next based on an error bound calculated at each step. They provide a bound on the maximum deviation of the field data in the simplified mesh from the original.

Several techniques for simplification have recently been proposed that act on the vertices. Van Gelder *et al.* [GVW99] remove vertices based on mass and data error metrics. Uesu *et al.* [UBFS04] provide a fast point-based method which works directly on the underlying scalar field. These techniques are more memory efficient than edge collapse methods, but require the addition of Steiner points to handle non-convex meshes. This requirement makes them difficult to modify for streaming algorithms.

The idea of a progressive mesh for surface level of detail control was proposed by Hoppe [Hop96] and later extended to simplicial complexes by Popović and Hoppe [PH97]. Staadt and Gross [SG98] define appropriate cost functions to account for volume preservation, gradient estimation, and scalar data with progressive tetrahedral meshes. Chopra and Meyer [CM02] propose a fast progressive mesh decimation scheme that is based on the scalar field of the mesh.

Many algorithms have been developed that use different error metrics to perform the simplification via edge collapses. Cignoni *et al.* [CCM*00] use domain and field (*i.e.* range) error metrics to approximate the original mesh. The use of a quadric error metric for surface simplification was introduced by Garland and Heckbert [GH97]. Their method uses iterative contractions on vertex pairs and calculates the error approximations using quadric matrices. Natarajan and Edelsbrunner [NE04] extend the quadric error metric to preserve topological features. Garland and Zhou [GZ05] recently generalized the quadric error metric for simplifying simplicial elements in any dimension.

As model size has continued to increase faster than main memory size in commodity PCs, techniques have been developed to simplify these data sets out-of-core. Lindstrom [Lin00] proposed an algorithm that simplifies triangle meshes of arbitrary size. This algorithm improves upon Rossignac and Borel's [RB93] vertex-clustering method by using the quadric error metric. The mesh is stored as a redundant list of three vertex positions per triangle. This "triangle soup" is read one triangle at a time and a simplified mesh is constructed incrementally and kept in-core. Lindstrom and Silva [LS01] improve upon the quality of this algorithm while making the method more memory efficient by storing the simplified mesh out-of-core during process-

ing. They handle boundaries separately to preserve the overall shape of the mesh. Wu and Kobbelt [WK03] propose an edge collapse-based streaming method for large triangle meshes that requires only one pass to decimate the entire mesh. All three of these methods make use of a redundant on-disk mesh representation that is two times larger than an indexed mesh. However, they are fast because no global indexing is required, which results in better cache coherency. Isenburg *et al.* [ILGS03] present a triangle mesh simplification algorithm based on the processing sequence paradigm. A processing sequence is a mesh representation that interleaves the ordering of the indexed triangles and vertices. This representation provides fast out-of-core access because it arranges the data in a memory-efficient order while requiring no additional storage. We extend the method of Isenburg *et al.* to tetrahedral meshes and improve upon it by recognizing the importance of preserving the streamability of the mesh after it undergoes simplification, *e.g.* for subsequent pipelined stream processing. We furthermore place fewer restrictions on the input mesh, making the algorithm directly applicable to a larger class of mesh producing applications.

With an increase in streaming algorithms, the need for a streamable format that efficiently codes both geometry and connectivity becomes necessary. Isenburg and Lindstrom [IL05] provide the underlying work for streaming representations of polygonal meshes. They provide metrics and diagrams for measuring the streamability of a mesh and discuss methods for improving its layout so as to reduce its memory footprint and the time each mesh element remains in-core. Mascarenhas *et al.* [MIPS04] extend this format to handle volumetric grids, which allows for streaming out-of-core isosurface extraction.

2. Mesh Layout

Traditional object file formats consist of a list of vertices followed by a list of polygonal or polyhedral elements that are defined by indexing into the vertex list. Dereferencing such a mesh, *i.e.* accessing vertices via their indices, requires the whole vertex list to be in memory since elements are generally not assumed to reference vertices in any particular order; an element can arbitrarily reference *any* vertex in the list. Furthermore, streaming such a mesh implies buffering all vertices before the first element is encountered in the stream. A logical progression for large meshes is to store them in a streaming mesh representation that interleaves the vertices and elements and stores them in a "compatible" order. This representation allows a vertex to be introduced (added to an in-core active set) when needed and finalized (removed from the active set) when no longer used.

Isenburg and Lindstrom [IL05] provide a streaming mesh format for triangle meshes that extends the popular OBJ file format. We use their format with additions to handle tetrahedral meshes. This includes extending the vertices to four values $\langle x, y, z, f \rangle$ that represent the position in 3D space and

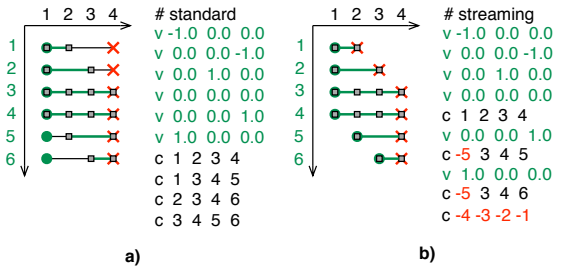

Figure 2: *Layout diagrams with cell indices on the vertical and vertex indices on the horizontal axis. A vertex is active from the time it is introduced until it is finalized, as indicated by the horizontal lines. (a) A standard format for tetrahedral meshes based on the OBJ format. (b) A streaming format that interleaves vertices with cells. Vertex finalization is provided with a negative index (i.e. relative to the most recent vertex).*

a scalar value. In addition, we provide a new element type for a tetrahedral cell that indexes four vertices. This format allows us to finalize a vertex when it is no longer in use by using a negative index into the vertex list. Figure 2 shows an ASCII example of the streaming tetrahedral format.

An important consideration with streaming meshes is the ability to analyze the efficiency of a mesh layout. Isenburg and Lindstrom developed several techniques for visualizing properties of the mesh to determine the effectiveness of the layout. We use similar tools to measure tetrahedral mesh quality. An important property is the *front width*, or the maximum number of concurrently active vertices. An *active* vertex is one that has been introduced, but not yet finalized. The width of a streaming mesh gives a lower bound on the amount of memory required for dereferencing (and thus processing) the mesh. Another important property of the mesh is the *front span*, which measures the maximum index difference (plus one) of the concurrently active vertices. A low span allows a faster implementation because optimizations can be performed that achieve the best performance when the span is similar to the width. The efficiency of our layout depends on quantifying these properties, thus we provide analysis on different layout techniques so that we can choose the most streamable layout for a given data set.

One simple mesh layout is to sort the vertices on a *spatial* direction, in particular one that crosses the most tetrahedra. Wu and Kobbelt [WK03] use this technique for triangle mesh simplification. This can be accomplished for large meshes by performing an out-of-core sort on the vertices [LS01] and writing them into a new file. An additional file is created to contain a mapping of the old ordering to the new one. Next, the tetrahedral cells are written to a new file and re-indexed according to the mapping file. A sort is then performed on the file containing the tetrahedral cells based on the largest index of each cell. Finally, the vertex file and the cell file are read simultaneously and interleaved

Data Set	Vertices	Tetrahedra	Spatial Sort		Z-Order		Spectral		Breadth-First	
			Width	Span	Width	Span	Width	Span	Width	Span
Torso	168,930	1,082,723	3,118	20,784	7,256	122,174	**2,894**	13,890	5,528	**6,370**
Fighter	256,614	1,403,504	**3,894**	110,881	9,382	215,697	3,916	28,638	16,629	**19,523**
Rbl	730,273	3,886,728	2,814	5,270	10,232	371,269	**2,291**	21,764	3,206	**3,495**
Mito	972,455	5,537,168	19,876	33,524	10,202	642,550	**6,745**	44,190	10,552	**11,498**
Sf1	2,461,694	13,980,162	16,898	65,921	48,532	1,958,212	**12,851**	131,152	30,258	**33,378**

Table 1: *Analysis of laying out the mesh spatially, by z-order, by spectral sequencing, and breadth-first.*

into a new file by writing each cell immediately after the vertex corresponding to the cell's largest index has been written. Spatial layouts work especially well when considering meshes that have a dominant principal direction.

Other techniques may be desirable if the mesh does not have a dominant principal direction, such as a sphere. An approach to handle this type of data is to use a *bricking* method similar to the one proposed by Cox and Ellsworth [CE97] in which the vertices are ordered into a fixed number of small cubes for better sequential access. A similar approach is to arrange the vertices using a Lebesgue space-filling curve (*i.e. z-order*), which provides better sequential access in the average case. This arrangement can be generated by creating an out-of-core octree [CMRS03] of the vertices and traversing them in-order. The interleaved mesh is then written to a file in the same manner described above for spatial sorting. The results of the layout produced by bricking and *z*-order traversal are similar. When streamed they provide a more contiguous portion of the mesh on average, but the front width and span are typically much larger than sorting spatially.

Another approach used for laying out the mesh on disk is *spectral sequencing*. This heuristic finds the first non-trivial eigenvector (the *Fiedler vector*) of the mesh's Laplacian matrix and was shown by Isenburg and Lindstrom [IL05] to be very effective at producing low-width layouts. They provide an out-of-core algorithm for generating this ordering for streaming triangle meshes, which we have extended to handle tetrahedra. This method works particularly well for curvy triangle meshes, but tetrahedral meshes are generally less curvy and more compact. Still, in most cases this ordering results in the lowest width, which is ideal for minimizing memory consumption.

A final approach is to create a *topological* layout, which starts at a vertex on the boundary and grows out to neighboring vertices. To grow in a contiguous manner, we use a breadth-first traversal with optimizations to improve coherence. Instead of a traditional FIFO priority queue, we assign priority using three keys. First, the oldest vertex on the queue is used in the same way that it would be in standard breadth-first algorithms. However, if multiple vertices were added to the queue at the same time, a second and third key are used to achieve a more coherent order. The second key is boolean, and gives preference to a vertex if it is the final one in a cell that has not been processed. Finally, the third key is to use the vertex that was most recently put on the queue, which is more likely to be adjacent to the last vertex. These sort keys guarantee a layout that is *compact* [IL05], such that runs of vertices are referenced by the next cell, and runs of cells reference the previous vertex (*e.g.* as in Figure 2b). Using this approach, we were able to minimize the front span of the data sets in all of our experimental cases. This is ideal because having a span and width that are similar allows us to exploit optimization techniques described in the simplification algorithm.

Table 1 shows the front width and span for four different data sets produced by the layout techniques described above. Spectral sequencing proves to be the superior choice when low width and thus memory efficiency is required. Breadth-first layouts are not as memory-efficient, but as we will see can be processed fast. Note that unlike [ILGS03] we do not impose the constraint that the sequence of tetrahedra be grown in a face-adjacent manner (equivalently edge-adjacent for triangle meshes). Therefore we can accept a streaming mesh in whatever order it arrives without having to locally reorder elements.

3. Tetrahedral Simplification

3.1. Quadric-based Simplification

To achieve high-quality approximations, we use the quadric error metric proposed by Garland and Zhou [GZ05]. This metric measures the squared (geometric and field) distances from points to hyperplanes spanned by tetrahedra. The volume boundaries are preserved using a similar metric on the boundary faces and by weighting boundary and interior errors appropriately. The generalized quadric error allows the flexibility of representing field data by extending the codimension of the manifold. Given a scalar function $f : D \subseteq \mathbb{R}^3 \rightarrow \mathbb{R}$ defined over a domain D represented by a tetrahedral mesh, we can represent the vertices at each point $\mathbf{p}$ as $\langle x_p, y_p, z_p, f_p \rangle$, which can be considered a point on a 3D manifold embedded in $\mathbb{R}^4$. Thus, by extending the quadric to handle field data, the algorithm intrinsically optimizes the field approximation along with the geometric position.

The quadric error of collapsing an edge to a single point

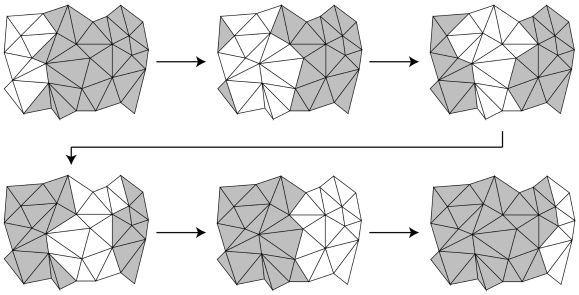

Figure 3: *Example of a buffer moving across the surface of a tetrahedral mesh sorted from left to right. As new tetrahedra are introduced, the tetrahedra that have been in-core the longest are removed from memory.*

is expressed as the sum of squared distances to all accumulated incident hyperplanes, and can in n dimensions be encoded efficiently as a symmetric $n \times n$ matrix $\mathbf{A}$, an n-vector $\mathbf{b}$, and a scalar c. It is sufficient to component-wise add these terms to combine the quadric error of two collapsed vertices. Finding the point $\mathbf{x}$ that minimizes this measure amounts to solving a linear system of equations $\mathbf{Ax} = \mathbf{b}$. Once $\mathbf{x}$ is computed, we test whether collapsing to this point causes any tetrahedra to flip, *i.e.* changes the sign of their volume, in which case we disallow the edge collapse. Because $\mathbf{A}$ is not necessarily invertible, it is important to choose a linear solver that is numerically stable. Since quadric metrics are covered in great detail elsewhere (see, *e.g.* , [GH97, GZ05]), we here focus only on the numerical issues of robustly minimizing quadric errors (see Section 3.4).

3.2. Streaming Simplification

Combining streaming meshes with quadric-based simplification, we introduce a technique for simplifying large tetrahedral meshes out-of-core. We base our streaming algorithm on [ILGS03], but make several general improvements and provide a list of optimizations that compared to a less carefully engineered implementation results in dramatic speed improvements.

First, unless already provided with streaming input, we convert standard indexed meshes and optionally reorder them for improved streamability. Then, portions of the streaming mesh are loaded incrementally into a fixed-size main memory buffer and are simplified using the quadric-based method. Once the in-core portion of the mesh reaches the user-prescribed resolution, simplified elements are output, *e.g.* to disk or to a downstream processing module. Thus input and output happen virtually simultaneously as the mesh streams through the memory buffer (see Figure 3).

To ensure that the final approximation is the desired size, two control parameters have been added: *target reduction* and *boundary weight*. Target reduction is the ratio between the number of tetrahedra in the output mesh and the number of tetrahedra in the original mesh. Alternatively, this parameter can be expressed as a target tetrahedral count of the resulting mesh. The boundary weight prevents the shape of the mesh from changing throughout the simplification. We use a fixed value of 100 times the maximum field value in the data for the weight in our experiments.

Because we only keep a small set of tetrahedra in memory, we do not know the entire mesh connectivity. Thus, we keep the boundary between the tetrahedra currently in memory and all remaining elements—tetrahedra that have not yet been read or that have been output—fixed to ensure that the simplified mesh is crack-free. We call this boundary the *stream boundary*, which consists entirely of faces from the interior of the mesh. We can identify the stream boundary faces as they are read in by utilizing the finalization information stored in the streaming mesh. A face of the current in-core mesh is part of the stream boundary if none of its three vertices are finalized. We disallow collapsing any edge that has one or both vertices on the stream boundary.

Due to the stream boundary constraint, if we read in one portion of the mesh, simplify it, and write it out to disk in one phase, our output mesh will have unsimplified areas along the stream boundaries. This results in an approximation that is oversimplified in areas and undersimplified in others. To avoid this problem, we follow the algorithm proposed by Wu and Kobbelt [WK03]. Their algorithm consists of a main loop in which READ, DECIMATE, and WRITE operations are performed in each iteration. The READ operation introduces new elements until it fills the buffer. Next, DECIMATE simplifies the elements in the buffer until either the target ratio is reached or the buffer size is halved. Finally, in their method the WRITE operation outputs the elements with the largest error to file.

We improve upon [WK03, ILGS03] by ensuring, to the extent possible, that the relative order among input elements is preserved in the output stream, with the caveat the tetrahedra whose vertices have not yet been finalized (*i.e.* are on the input stream boundary) must be delayed. Therefore, the output typically retains the coherence of the input. An error-driven output criterion, on the other hand, can considerably fragment the buffer and split off small “islands” that remain in the buffer for a long time without being eligible for simplification, and thus unnecessarily clog the stream buffer. Furthermore, such an output stream generally has poor stream qualities, which affects downstream processing. The front width (*i.e.* number of active vertices), for example, is particularly important for tetrahedral meshes, for which each active vertex affects on average four times as many elements as in a triangle mesh, and therefore more adversely affects memory requirements and processing delay. Furthermore, we relax the requirement that the stream of tetrahedra (triangles) advance in a face (edge) adjacent manner [ILGS03], as this is of no particular value to us, and we allow any coherent ordering of mesh elements. Finally, using the more

streamable layouts and simpler streaming mesh formats and API from [IL05], we gain considerably in performance and memory usage over [ILGS03].

3.3. Implementation Details

Since our method processes different mesh portions of bounded size sequentially, a statically allocated data structure is more efficient than dynamic allocations, which collectively increase the memory footprint. In our implementation, we extended Rossignac's corner table [Ros01] for triangle meshes to tetrahedral meshes. The original corner table requires two fixed-size arrays V and O indexed by corners (vertex-cell associations) c, where $V[c]$ references the vertex of c and $O[c]$ references the "opposite" corner of c.

In the case of tetrahedral simplification, the most common query is to find all tetrahedra around a vertex. Therefore, we replace the O array with a link table L of equal size, which joins together all corners of a given vertex in a circular linked list. We additionally store with each vertex an index to one of its corners.

We store the mesh internally as three fixed-size arrays of vertices, tetrahedra, and corners (*i.e.* the links L). Each vertex contains a pointer to one corner and the quadric error information $(\mathbf{A}, \mathbf{p}, \varepsilon)$ using a parameterization that explicitly represents the vertex geometry and scalar data in $\mathbf{p}$ (see 3.4 and Figure 4). The quadrics for the tetrahedra are calculated when we read in a new set of tetrahedra, and are then distributed to the vertices. For each finalized vertex, we compute boundary quadrics for all incident boundary faces (if any) that have no other finalized vertices, and distribute these quadrics to the boundary vertices.

Garland and Zhou [GZ05] use a greedy edge collapse method and maintain a priority queue for the edges ordered by quadric error. Forsaking greediness, we obtain comparable mesh quality by using a multiple choice randomized approach [WK03] with eight candidates per collapse. There are several advantages of using randomized selection. One is that we no longer need a priority queue or explicit representation of edges. Instead an edge can be found by randomly picking a tetrahedron and then randomly selecting two of its vertices. Another advantage is that the randomized technique can be further accelerated by exploiting information readily available through our quadric representation (see 3.4).

Before we output a tetrahedron, we must ensure that its four vertices are output first. Once a vertex is output, we mark it prevent it from being decimated in future iterations. To enhance the performance, we use a lazy deletion scheme, where all vertices and tetrahedra to be deleted are initially marked. At the end of each WRITE phase, we make a linear pass through all vertices and tetrahedra to remove marked elements and compact the arrays. Since we do not allocate additional memory during simplification, keeping deleted vertices and tetrahedra does not increase the memory footprint.

```
VERTEX
    matrix A          10 floats
    vector p          4 floats
    float e
    int idx           global index
    int corner        vertex-to-corner index
    uchar flags       deleted, written
TETRAHEDRON
    int vidx[4]       vertex indices
    bool deleted
    int order         for ordering tetrahedra
    int L[4]          corner links
```

Figure 4: *Data structures used for our quadric-based simplification.*

Improvements	Time (sec)	Memory (MB)
Initial Implementation	212.95	310
QEF + CG	132.68	282
Multiple Choice + Corner Table	38.35	130
4D Normal, Floats	35.99	78
Final / Binary I/O	22.10	78

Table 2: *Improvements over straightforward implementation due to our optimization techniques. The results are for simplifying the Fighter data set (1.4 M tetrahedra) completely in-core on a P4 2.2GHz with 1GB of RAM.*

Storing large data sets on disk in ASCII format can adversely affect performance because converting ASCII numbers to an internal binary representation can be surprisingly slow. We have extended the ASCII stream format in Figure 2 to a binary representation. Because our program spends over 30% of the time on disk I/O, this optimization results in a nonnegligible speedup. For example, on the SF1 data set it improves overall performance by 17%.

Since we only maintain a small portion of the mesh in-core, we require a way of mapping global vertex indices to in-core buffer indices. Usually a hash map is used, but with our low-span breadth-first mesh layout, this hash map can be replaced by a fixed-size array indexed using modular arithmetic. We move occasional high-span vertices that cause "collisions" in this circular array to an auxiliary hash.

With all of the optimizations described above, our simplifer can run at high speed without any dynamic memory allocation at run time. The performance and memory summary can be found in Table 3.3. Further efficiency improvements relating to quadric error metrics will be discussed in the following section.

3.4. Numerical Issues

Great care has to be taken when working with quadric metrics to ensure numerical stability while retaining efficiency. To minimize quadric errors, a positive semidefinite system of linear equations must be solved, for which numerically accurate but heavy-duty techniques such as singular value decomposition (SVD) [Lin00, KBSS01] and QR factorization [JLSW02] have been proposed. However, even constructing, representing, and evaluating quadric errors require that special care be taken. We here outline an efficient representation of quadric error functions that leads to numerically stable operations, improved speed, and less storage.

The standard representation [GZ05] of quadric errors is parameterized by $(\mathbf{A}, \mathbf{b}, c)$, and is evaluated as

$$Q(\mathbf{x}) = \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} - 2\mathbf{b}^\mathsf{T}\mathbf{x} + c \qquad (1)$$

Typically the three terms in this equation are "large," but sum to a "small" value, resulting in a loss of precision. One can show that the roundoff error is proportional to $\|\mathbf{A}\|\|\mathbf{x}\|^2$. Furthermore, in addition to this quadric information, it is common to store the vertex position (and field value) $\mathbf{p}$ that minimizes Q separately. Lindstrom [Lin03] suggested an alternative representation that removes this redundancy:

$$Q(\mathbf{x}) = (\mathbf{x}-\mathbf{p})^\mathsf{T}\mathbf{A}(\mathbf{x}-\mathbf{p}) + \varepsilon \qquad (2)$$

where $\mathbf{A}$ is the same as in the standard representation and

$$\begin{aligned} \mathbf{A}\mathbf{p} &= \mathbf{b} \\ \mathbf{p}^\mathsf{T}\mathbf{A}\mathbf{p} + \varepsilon &= c \end{aligned}$$

This parameterization $(\mathbf{A}, \mathbf{b}, \varepsilon)$ provides direct access to the minimum quadric error ε and the minimizer $\mathbf{p}$. This not only saves memory but also results in a more stable evaluation of Q, as the roundoff error is now proportional to $\|\mathbf{A}\|\|\mathbf{x}-\mathbf{p}\|^2$, and we are generally interested in evaluating $Q(\mathbf{x})$ near its minimum $\mathbf{p}$ as opposed to near the origin. Another significant benefit of this representation is that it provides a lower bound $\varepsilon_i + \varepsilon_j$ on $Q_i + Q_j$ when collapsing two vertices v_i and v_j. Using randomized edge collapse [WK03], we can thus often avoid minimizing $Q_i + Q_j$ if the lower bound already exceeds the smallest quadric error found so far.

Our quadric representation also lends itself to an efficient and numerically stable iterative linear solver. To handle ill-conditioned matrices $\mathbf{A}$, we have adapted the well-known conjugate gradient (CG) method [GV96] to work on semidefinite matrices. As in SVD, we provide a tolerance κ_{max} on the condition number $\kappa(\mathbf{A})$, and pre-empt the iterative solver when all remaining conjugate directions are deemed "insignificant" for reducing Q. This is equivalent to zeroing small singular values in SVD. Using our quadric representation, we conveniently initialize the CG solver with the guess $\mathbf{x} = (\mathbf{p}_i + \mathbf{p}_j)/2$.

A final word of caution: The computation of generalized quadrics presented in [GZ05] computes $\mathbf{A}$ = $\mathbf{I}$ - $\mathbf{N}$, whose null space $\text{null}(\mathbf{A}) = \text{range}(\mathbf{N})$ is spanned by the tetrahedron, via subtraction, which due to roundoff error can leave $\mathbf{A}$ indefinite, *i.e.* with one or more negative eigenvalues. This causes Q to have a "saddle" shape with no defined minimum, and can cause numerical instability. Instead, we compute a 4D "volume normal" using a generalization of the 3D cross product to 4D. The outer product of this normal with itself gives a positive semidefinite $\mathbf{A}$ for a tetrahedron.

```
SOLVE(A, x, b, n, κmax)
 1  r = b − Ax                          negative gradient of Q
 2  p = 0
 3  for k = 1,...,n                     iterate up to n times
 4     s = r^T r
 5     if s = 0 then exit               solution found?
 6     p = p + r/s                      update search direction
 7     q = Ap
 8     t = p^T q
 9     if st ≤ tr(A)/(nκmax) then exit  insignificant direction?
10     r = r − q/t                      update gradient
11     x = x + p/t                      update solution
```

Figure 5: *Conjugate gradient solver for positive semidefinite systems* $\mathbf{A}\mathbf{x} = \mathbf{b}$. *On input* $\mathbf{x}$ *is an estimate of the solution,* $n = 4$ *is the number of linear equations, and* κ_{max} *is a tolerance on the condition number.* $\text{tr}(\mathbf{A})$ *is the trace of* $\mathbf{A}$.

Because of our attention to numerical stability, with $\kappa_{max} = 10^3$ we are able to use single precision floating point throughout our simplifier, even for the largest meshes. Since the 4D quadric information requires 15 scalars per vertex, this saves considerable memory and improves the speed.

4. Results

4.1. Stability and Error Analysis

We have described a CG method for solving the linear equations that arise when minimizing the quadric error. The choice of solver is important because degenerate tetrahedra and regions of near-constant field value can cause singularity. For testing purposes, we constructed a data set by subdividing a tetrahedron into hundreds of smaller tetrahedra by interpolating the vertices and field data. Obviously, these small tetrahedra all lie on the hyperplane spanned by the original tetrahedron, thus they are solutions to the linear equations. We then picked a solution as a target for each collapsed edge. We experimented with several linear solvers as shown in Table 4.1. We experimented with Cholesky factorization, SVD, SVD with Cramer's Rule, and CG. Our hybrid SVD/Cramer's Rule method uses a conservative condition estimator and uses Cramer's Rule when the matrix is well-conditioned. Using all of our solvers, we were able to simplify our subdivided tetrahedron to its original shape with small error in both field and geometry. In our experiments, CG provides the fastest, most stable solution.

To estimate the error in the simplified mesh we use two

Linear System Solver	Performance	Stability
Cholesky	0.94	Unstable
SVD	2.86	Stable
SVD with Cramer's Rule	1.16	Stable
Conjugate Gradient	1.0	Stable

Table 3: *Performance of linear solvers. The performance is the time relative to CG for simplifying the Fighter data set (1.4 M tetrahedra) completely in-core.*

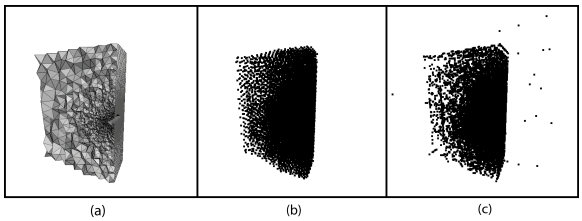

Figure 6: *Demonstration of the stability of linear system solvers on the Fighter data set. A portion of the original mesh is shown in (a). Solutions to linear equations are shown as black dots using SVD with Cramer's Rule (b) and Cholesky factorization (c).*

different methods. The first method is to measure the error on the surface boundary of the mesh using the tool *Metro* [CRS98]. The second method is to measure the error in the field data using a similar approach to Cignoni *et al.* [CCM*00]. Our implementation differs because it ignores points outside the domain of the simplified mesh since these points become part of the surface boundary error. Table 4 shows these error estimations.

4.2. Performance

All timing results were generated on a 3.2 GHz Pentium 4 machine with 2.0 GB RAM. For the streaming experiments, we limit the operating system to only 64 MB RAM by using the Linux bootloader. Using our algorithm, we simplified all of our data sets to 10% of original size. Table 4 shows the results of simplifying a collection of data sets to 10% of their original size using our streaming algorithm and the same implementation optimized for in-core execution.

We were able to achieve streaming simplification with only a slight increase in time and error compared to an in-core implementation. The streaming technique has the advantage of a smaller memory footprint. With our algorithm, we were able to simplify 14 million tetrahedra while only using 20 MB RAM. Due to the large size of the Sf1 data set, certain parts of the stream were not able to be simplified accurately, resulting in a larger error. By increasing the memory slightly, the quality of the simplification is greatly improved and approaches the in-core quality.

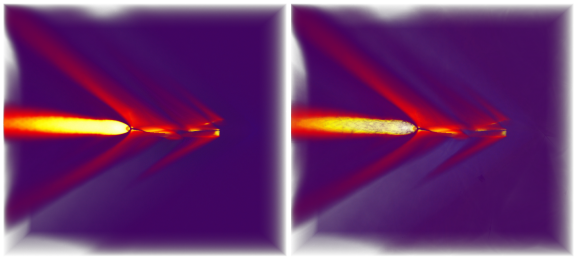

Figure 7: *Volume Rendered images of the Fighter data set show the preservation of scalar values. The original data set is shown on the left (1,403,504 tetrahedra) and the simplified version is shown on the right (140,348 tetrahedra).*

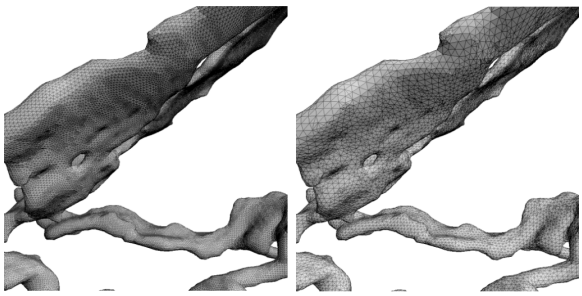

Figure 8: *Views of the mesh quality on the surface of the Rbl data set. The original data set is shown on the left (3,886,728 tetrahedra) and the simplified version is shown on the right (388,637 tetrahedra).*

5. Discussion

The use of streaming meshes for simplification reduces the memory footprint of a large mesh considerably. We improved on the algorithm of Wu and Kobbelt [WK03] by preserving the stream order of the mesh between input and output. A direct comparison with their algorithm shows that our method consistently achieves a lower width, *e.g.* 9% versus 59% on the Fighter data set, and span, *e.g.* 45% versus 98%, without reducing the approximation quality. In addition, with the optimizations that we employ to our data structures, we have been able to simplify up to 14 million tetrahedra while using only 20MB RAM. Only the smallest data sets (Torso and Fighter) could be simplified using our implementation of Wu and Kobbelt's algorithm.

Apart from providing a streaming algorithm that operates on meshes of arbitrary size, we also described speed and stability optimizations that improve the performance of tetrahedral simplification. Our quadric representation improves linear solver performance. In addition, our adapted conjugate gradient method and the use of "volume normals" for tetrahedra reduce the numerical errors and allow the use of single precision floating-point numbers. By using a binary format, we improve on storage and speed up I/O. Finally, through the use of a breadth-first mesh layout, we have improved the width and the span, which enables the use of a fixed-size cir-

Data Set	Number of Tets Input	Number of Tets Output	Time (sec)	Max RAM (MB)	Max Field Error (%)	RMS Field Error (%)	Max Surface Error (%)	RMS Surface Error (%)
In-core								
Torso	1,082,723	108,271	14.88	57	0.012	0.000884	0.120	0.013360
Fighter	1,403,504	140,348	15.46	78	4.845	0.280266	0.038	0.000352
Rbl	3,886,728	388,668	59.10	212	0.020	0.002574	0.025	0.000055
Mito	5,537,168	553,711	47.13	285	0.045	0.007355	0.001	0.000008
Sf1	13,980,162	1,398,013	191.69	709	5.626	0.262335	0.036	0.000811
Streaming								
Torso	1,082,723	108,270	19.07	20	0.019	0.000879	0.161	0.001226
Fighter	1,403,504	140,345	20.87	20	4.549	0.299081	0.102	0.000470
Rbl	3,886,728	388,671	95.54	20	0.025	0.002833	0.036	0.000089
Mito	5,537,168	553,716	73.58	20	0.045	0.007614	0.044	0.000009
Sf1	13,980,162	1,398,012	246.15	20	23.287	0.472869	0.169	0.004150
Sf1	13,980,162	1,398,017	244.42	50	5.834	0.315583	0.050	0.001394

Table 4: *Summary of the results using both in-core and streaming simplification. Field error percentages are in relation to the range of the field and surface error percentages are in relation to the bounding box diagonal.*

cular array instead of a hash table. With these improvements, we have improved the speed of our simplification method by an order of magnitude over our initial implementation of Garland's quadric-based simplification [GZ05].

Due to the efficiency of our algorithm, we easily handle the largest data sets we obtained. In our experimental results, our streaming algorithm simplifies about 60K tetrahedra per second and achieves high quality results. Given a machine with 2.0 GB of RAM, the in-core implementation of our algorithm could handle approxmately 35 million tetrahedra. Using our streaming algorithm, this maximum bound is much larger and is only constrained by the front width of the mesh, which we have shown to be very small when using a good mesh layout.

6. Conclusions and Future Work

We have presented a streaming technique for simplifying tetrahedral meshes of arbitrary size. We describe several methods for laying out a tetrahedral mesh on disk in a coherent, I/O-efficient format. We also show an analysis of these layouts and the effects they have on the final simplified mesh. We provide a technique for simplifying small portions of the mesh in memory while obtaining a smooth simplification over the entire mesh. The simplification occurs in one pass, preserves mesh topology and scalar information, requires little memory, and runs quickly. We also provide optimizations to traditional simplification data structures that improve speed and efficiency. We present a linear solver that improves stability and speed of quadric-based simplification. Finally, we provide stability and error analysis of our algorithm with results for specific examples that show the time and memory required for processing.

Because the generalized quadric error works on a variety of data types, an interesting extension would be to attempt to handle meshes consisting of other types, *e.g.* hexahedra. Finally, it would be interesting to take advantage of the coherent streaming tetrahedral format to perform fast, out-of-core isosurface extraction.

References

[CCM*00] CIGNONI P., CONSTANZA D., MONTANI C., ROCCHINI C., SCOPIGNO R.: Simplification of tetrahedral meshes with accurate error evaluation. In *IEEE Visualization '00* (2000), IEEE Computer Society Press, pp. 85–92. 2, 8

[CE97] COX M., ELLSWORTH D.: Application-controlled demand paging for out-of-core visualization. In *IEEE Visualization '97* (1997), pp. 235–244. 4

[CM02] CHOPRA P., MEYER J.: Tetfusion: an algorithm for rapid tetrahedral mesh simplification. In *IEEE Visualization '02* (2002), IEEE Computer Society, pp. 133–140. 2

[CMRS03] CIGNONI P., MONTANI C., ROCCHINI C., SCOPIGNO R.: External memory management and simplification of huge meshes. *IEEE Transactions on Visualization and Computer Graphics 9*, 4 (2003), 525–537. 4

[CRS98] CIGNONI P., ROCCHINI C., SCOPIGNO R.: Metro: measuring error on simplified surfaces. *Computer Graphics Forum 17*, 2 (1998), 167–174. 8

[GH97] GARLAND M., HECKBERT P. S.: Surface simplification using quadric error metrics. In *Proceedings of SIGGRAPH 97* (Aug. 1997), Computer Graphics Proceedings, Annual Conference Series, pp. 209–216. 2, 5

[GV96] GOLUB G. H., VAN LOAN C. F.: *Matrix Computations*, third ed. Johns Hopkins University Press, 1996. 7

[GVW99] GELDER A. V., VERNA V., WILHELMS J.: Volume decimation of irregular tetrahedral grids. *Computer Graphics International* (1999), 222–230. 2

[GZ05] GARLAND M., ZHOU Y.: Quadric-based simplification in any dimension. *ACM Transactions on Graphics 24*, 2 (Apr. 2005). 2, 4, 5, 6, 7, 9

[Hop96] HOPPE H.: Progressive meshes. In *Proceedings of SIGGRAPH 96* (Aug. 1996), Computer Graphics Proceedings, Annual Conference Series, pp. 99–108. 2

[IL05] ISENBURG M., LINDSTROM P.: *Streaming Meshes*. Tech. rep., Lawrence Livermore National Laboratory, 2005. UCRL-TR-211608-DRAFT. 3, 4, 6

[ILGS03] ISENBURG M., LINDSTROM P., GUMHOLD S., SNOEYINK J.: Large mesh simplification using processing sequences. In *Visualization'03 Conference Proceedings* (2003), pp. 465–472. 3, 4, 5, 6

[JLSW02] JU T., LOSASSO F., SCHAEFER S., WARREN J.: Dual contouring of hermite data. *ACM Transactions on Graphics 21*, 3 (July 2002), 339–346. 7

[KBSS01] KOBBELT L. P., BOTSCH M., SCHWANECKE U., SEIDEL H.-P.: Feature-sensitive surface extraction from volume data. In *Proceedings of ACM SIGGRAPH 2001* (Aug. 2001), Computer Graphics Proceedings, Annual Conference Series, pp. 57–66. 7

[Lin00] LINDSTROM P.: Out-of-core simplification of large polygonal models. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (2000), ACM Press/Addison-Wesley Publishing Co., pp. 259–262. 2, 7

[Lin03] LINDSTROM P.: Out-of-core construction and visualization of multiresolution surfaces. In *2003 ACM Symposium on Interactive 3D Graphics* (Apr. 2003), pp. 93–102. 7

[LS01] LINDSTROM P., SILVA C. T.: A memory insensitive technique for large model simplification. In *Proceedings of the conference on Visualization '01* (2001), IEEE Computer Society, pp. 121–126. 2, 3

[MIPS04] MASCARENHAS A., ISENBURG M., PASCUCCI V., SNOEYINK J.: Encoding volumetric grids for streaming isosurface extraction. In *International Symposium on 3D Data Processing, Visualization, and Transmission '04 Proceedings* (2004). 3

[NE04] NATARAJAN V., EDELSBRUNNER H.: Simplification of three-dimensional density maps. *IEEE Transactions on Visualization and Computer Graphics* (2004). To appear. 2

[PH97] POPOVIĆ; J., HOPPE H.: Progressive simplicial complexes. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques* (1997), ACM Press/Addison-Wesley Publishing Co., pp. 217–224. 2

[RB93] ROSSIGNAC J., BORREL P.: *Geometric Modeling in Computer Graphics*. Springer-Verlag, 1993, ch. Multi-resolution 3D approximations for rendering complex scenes, pp. 455–465. 2

[Ros01] ROSSIGNAC J.: 3d compression made simple: Edgebreaker with zip&wrap on a corner-table. In *2001 Proceedings of the International Conference on Shape Modeling & Applications* (Washington, DC, USA, 2001), IEEE Computer Society, p. 278. 6

[SG98] STAADT O. G., GROSS M. H.: Progressive tetrahedralizations. In *IEEE Visualization '98* (1998), pp. 397–402. 2

[THJ99] TROTTS I. J., HAMANN B., JOY K. I.: Simplification of tetrahedral meshes with error bounds. In *IEEE Transactions on Visualization and Computer Graphics*, vol. 5 (3). IEEE Computer Society, 1999, pp. 224–237. 2

[THJW98] TROTTS I. J., HAMANN B., JOY K. I., WILEY D. F.: Simplification of tetrahedral meshes. In *IEEE Visualization '98 (VIS '98)* (Washington - Brussels - Tokyo, Oct. 1998), IEEE, pp. 287–295. 2

[UBFS04] UESU D., BAVOIL L., FLEISHMAN S., SILVA C. T.: *Simplification of Unstructured Tetrahedral Meshes by Point-Sampling*. Tech. rep., Scientific Computing and Imaging Institute, University of Utah, 2004. UUSCI-2004-005. 2

[WK03] WU J., KOBBELT L.: A stream algorithm for the decimation of massive meshes. In *Graphics Interface '03 Conference Proceedings* (2003), pp. 185–192. 3, 5, 6, 7, 8