Ray-traced Shell Traversal of Tetrahedral Meshes for Direct Volume Visualization

Alper Sahistan*  
Bilkent University

Serkan Demirci  
Bilkent University

Nathan Morrical  
University of Utah

Stefan Zellmann  
University of Cologne

Aytek Aman  
Bilkent University

Ingo Wald  
NVIDIA Corporation

Ügur Güdükbay  
Bilkent University

Figure 1: Renderings with secondary effects obtained with the proposed approach on an NVIDIA RTX 8000 GPU: (a) Plasma64 dataset rendered at 46.6 frames per second (fps). (b) Jets dataset rendered at 48.5 fps. (c) Fusion dataset rendered at 13.9 fps. (d) Agulhas dataset rendered at 5.6 fps.

ABSTRACT

A well-known method for rendering unstructured volumetric data is tetrahedral marching (tet marching), where rays are marched through a series of tetrahedral elements. However, existing tet marching techniques do not easily generalize to rays with arbitrary origin and direction required for advanced shading effects or non-convex meshes. Additionally, the memory footprint of these methods may exceed GPU memory limits. Interactive performance and high image quality are opposing goals. Our approach significantly lowers the burden to render unstructured datasets with high image fidelity while maintaining real-time and interactive performance even for large datasets. To this end, we leverage hardware-accelerated ray tracing to find entry and exit faces for a given ray into a volume and utilize a compact mesh representation to enable the efficient marching of arbitrary rays, thus allowing for advanced shading effects that ultimately yields more convincing and grounded images. Our approach is also robust, supporting both convex and non-convex unstructured meshes. We show that our method achieves interactive rates even with moderately-sized datasets while secondary effects are applied.

Index Terms: Human-centered computing—Visualization—Visualization application domains—Scientific visualization; Computing methodologies—Computer Graphics—Rendering—Ray Tracing

1 INTRODUCTION

One of the most common techniques to render scientific datasets is direct volume rendering (DVR). Direct volume rendering is usually performed using ray traversal, which allows for evaluating advanced shading effects such as ambient occlusion or single or multiple scattering. The widely utilized methods for unstructured meshes; however, often use a rasterization framework to initiate and perform ray marching and thus do not allow for simple integration of these effects. A widely utilized method for unstructured meshes is ray-marching via connectivity information, where each ray samples and integrates the opacity and color information for each cell they pass through. DVR techniques that can render unstructured volumes come with their own challenges. We aim to tackle these challenges while avoiding the shortcomings of raster-based approaches.

Unstructured volume traversal via face connectivity information is often met with challenging limitations like high memory usage and robustness issues. These issues usually hinder performance or exceed the memory limitations of modern GPUs. We address these challenges for non-raster frameworks using compact memory layouts and robust ray-tetrahedron intersection methods.

To further improve visualization quality and perception, we trace arbitrary rays to apply visual effects such as ambient occlusion, shadows, shading, and border contours (see Fig. 1). These effects usually require tracing secondary rays, cast after primary rays interact with the scene geometry. Unlike camera rays, these secondary rays are arbitrary; achieving these effects in raster-based pipelines is challenging [10]. In addition, existing solutions usually include extra raster passes, hindering interactive performance.

We propose a novel approach to ray-marching unstructured meshes that combine three main improvements:

- a compact, cache- and GPU-friendly memory layout that facilitates fast ray-tetrahedra intersection and efficient tetrahedra-to-tetrahedra traversal,
- the ability to handle arbitrary, non-common-origin rays as required for secondary effects like ray-traced reflections, shadows, or ambient occlusion, and
- the ability to efficiently handle convex and non-convex datasets, datasets with holes, curves and discontinuities.

2 RELATED WORK

Several research works have recently proposed to exploit RTX hardware for DVR [14, 20, 23, 24]. Wald et al. [23] use RTX BVHs for point location in unstructured tetrahedral meshes, at the cost of considerable memory consumption. The authors used these point
location kernels to accelerate a sampling-based ray marcher for tet meshes. This work was later extended by Morrical et al. [15] to support other cell types than tetrahedra. In contrast to these approaches, we use BVHs for just the exterior (shell) faces of tetrahedral volumes. Our work is orthogonal to that by Muigg et al. [16], which is tailored to the rasterization pipeline. They first subdivide the volume into bricks using a kd-tree and then render these in front-to-back order. Then they perform a depth peeling step where each brick’s entry and exit faces are determined. In contrast to theirs, our method allows us to trace secondary rays whose origin and direction are different from those of viewing rays.

When ray marching through tetrahedral volumes, we need to determine the next tetrahedron at every step. Several methods have been proposed to accomplish that [1, 7–9]. Our work is based on that by Aman et al. [1] who proposed highly optimized memory layouts to improve marching performance and reduce memory consumption. Their method is based on projecting tetrahedra vertices into a 2-D ray-centric coordinate system to reduce instruction count. We extend their ray marching algorithm to support DVR.

3 Method Overview
In order to support high-quality DVR, we need to be able to a) find entry faces and their associated tetrahedra, and b) to efficiently march from one tetrahedron to the next, a method we call tetrahedron marching, or tet marching. The process of finding entry and exit faces on what we call the shell (the triangle mesh induced by the generally non-convex hull of the tet mesh) we accelerate using an OptiX bounding volume hierarchy (BVH) [11] (the shell-BVH). Traditional raster-based approaches like the one by Muigg et al. [16] use multiple passes for finding entry faces. Here, each pixel is assigned the tetrahedron ID where marching starts. While powerful, these approaches do not easily extend to non-convex meshes and only allow for common-origin projection. This work demonstrates how we can address these shortcomings by using a dedicated ray-tracing framework like OptiX. These adjustments allow for arbitrary ray / tet-mesh intersections and thus for high-quality ray-traced shading effects that, to our knowledge, have not been demonstrated in a tetrahedra marching framework before.

Efficient tetrahedra marching relies on the principle of ray-connectivity between elements. In this work, all primitives are tetrahedra, which are occasionally derived by tetrahedralizing higher-dimensional elements. Ray-connectivity implies that each tetrahedron neighbors another tetrahedron if they share one of their four faces, which makes up a continuous path of tetrahedra on a ray segment. If a face does not connect two tetrahedra, this makes that face a shell-face by definition. The tetrahedra marching method operates between these shared faces where rays go face to face until they reach an opaque region or a termination condition is met.

Our marchers are optimized to reduce memory accesses, cache misses, and arithmetic complexity. We exploit the fact that each tetrahedron shares three vertices with a neighbor to minimize memory access. Furthermore, we use a specialized 2-D projection to reduce arithmetic complexity. Furthermore, our memory scheme stores, sorts, and compresses neighborhood and vertex information.

4 Shell-to-Shell Traversal
The shell-BVH is just an ordinary BVH of triangles that we realize using the OptiX framework so we can leverage hardware acceleration for triangle geometry. The OptiX API requires us to specify the shell triangles in the form of one list of triangle vertices and another list of triangle indices. Memory-wise those lists come on top of the already stored tetrahedron data structures as proposed in Sect. 5.1. For the indices, we fact in use int4’s so that we can store, in addition to the triangle indices, the index of the tetrahedron that this triangle is associated with and that we use as an entry point for tet marching (see Sect. 5).

Figure 2: Shell-to-shell traversal in 2-D. Left: a non-convex tetrahedral volume with the shells in purple. Right: Example path where I) is the viewing ray, II) is a ray (blue) we cast backward to find the entry shell-face of the current volume segment. III) takes us to the next segment after tet marching the first segment completed. Finally, IV) is again cast to find the entry face and V) is cast to find the exit face for the second volume segment.

To initiate tet marching, we first need to find an entry face on the shell. This is simple if the entry face is in front of the ray origin, but more involved when we start marching inside the volume and the entry face is behind the ray origin. We therefore generally first find the exit face, by tracing a ray against the shell BVH, but with front face culling activated. We then trace another ray in the opposite direction, starting at the point of intersection with the exit face, to find the entry face. We march through the segment—potentially using early ray termination if the ray is used to compute radiance and then exit altogether. We then extend the ray and repeat that until all segments were processed. The process is illustrated in Fig. 2 and in Alg. 1, which we realized using an OptiX RayGen program [12].

Algorithm 1 Shell-to-shell traversal.
1: procedure SHELL2SHELLTRAVEL(ray, shells, tetMesh)
2: \( C_{dit} \leftarrow (0,0,0) \quad \triangleright \) final color
3: \( \alpha \leftarrow 0 \quad \triangleright \) opacity
4: \( \text{while } \alpha < 1 \text{ do} \)
5: \( \text{payload } \leftarrow \text{traceRay}(ray, shells, \text{CULLFRONT}) \)
6: \( \text{if payload.hit then} \)
7: \( bRay.origin \leftarrow \text{payload.hitPoint} \quad \triangleright \) backwards ray
8: \( bRay.direction \leftarrow -ray.direction \)
9: \( bRay \leftarrow \text{traceRay}(bRay, shells, \text{CULLFRONT}) \)
10: \( v_{did},bPayloa\_face \leftarrow v_{Payloa\_face}bRay.face \quad \triangleright \) face vert. ids
11: \( Id_{set} \leftarrow \text{Payload.face}_{3} \quad \triangleright \) entry tet. id
12: \( C_{voi},\alpha_{voi} \leftarrow \)
13: \( \text{marchVolume(ray, v_{did}, Id_{set}, tetMesh)} \)
14: \( C_{dit} + = C_{vol} \times \alpha_{vol} \times (1 - \alpha) \)
15: \( \alpha + = (1 - \alpha) \times \alpha_{vol} \)
16: \( \text{ray.origin } \leftarrow bRay.origin \)
17: \( \text{else} \quad \triangleright \) Nothing left to hit
18: \( C_{dit} + = C_{mix} \times (1 - \alpha) \quad \text{break} \)

5 Efficient Tetrahedra Marching
We propose volume rendering variants of three memory layouts, Tet32, Tet20, and Tet16, with their marching algorithms, originally proposed by [1]. In the proposed structures, we store vertices in a separate list along with scalars. These layouts aim to reduce the memory footprint of the method while sustaining fast traversal times. We modify their traversal algorithm to approximate the volume rendering equation [5] and to work as a RayGen [12]. Furthermore, we extend the algorithm to trace arbitrary rays inside the tetrahedral mesh, allowing us to achieve advanced effects.

5.1 Memory Layouts
The tetrahedron representations that we use [1] have an exclusive-or-sum (xor-sum) field in common. This field, called vx in Fig. 3, allows us to reduce memory requirements, similar to [18] and [13]. We calculate the xor-sum as \( vx^{i} = v_{v0}^{i} \oplus v_{v1}^{i} \oplus v_{v2}^{i} \oplus v_{v3}^{i} \). Since \( a \oplus a = 0 \) we can obtain a vertex index from \( vx^{i} \) given we know the other three.
In addition to the xor field, the structures contain up to three vertex
and four neighbor indices, denoted as \( v_i \) and \( n_j \), respectively.

\[
\text{struct \ Tet32} \{ \begin{align*}
\text{int3 \ v; } & \quad \text{struct \ Tet20} \{ \begin{align*}
\text{int \ v; } & \quad \text{struct \ Tet16} \{ \begin{align*}
\text{int \ v; } & \quad \text{int \ n; } \} ; \\
\text{int4 \ n; } & \} ; \\
\text{int3 \ nx; } & \} ; \\
\end{align*} \end{align*} \end{align*}
\]

Figure 3: Tetrahedra memory layouts: each integer is four bytes. Tet32, Tet20, and Tet16 occupy 32, 20, and 16 bytes, respectively.

Tet32 layout stores all connectivity information, i.e., we use four neighbors of the tetrahedron along with three vertices and the xor field to obtain the 4\(^{th}\) vertex. Tet20 layout stores all connectivity information for a tetrahedron, i.e., four of its neighbors along with the xor field. Because neighboring tetrahedra share three vertices during tetrahedra marching, we obtain the unshared vertex of the adjacent tetrahedron using the xor field. On top of that, in Tet16 representation, instead of storing the indices of the neighboring tetrahedra, we keep values that can reconstruct indices along with the xor field. We pre-compute these fields as \( n_j = n_j' \oplus n_3' \) where \( j \in \{0, 1, 2\} \) and \( i \) is tetrahedron index.

5.2 Marching Algorithms

Aman et al. [1] assume that the camera is always inside the volume, they use a “source tetrahedron” to start traversal. While this assumption is valid in their case, this is, in general, only true for DVR if the camera is inside the mesh. Instead, we utilize the tetrahedron index information explained in Sect. 4 to compute entry and exit faces thus initializing marching. Due to our compaction, without a tetrahedron index or a source tetrahedron, tet marching cannot be initiated.

Another critical component is the intersection tests that determine the tetrahedron to traverse next. We adopt the point projection on a specialized basis method proposed by Aman et al. [1] where points are projected to a 2-D ray centric space whose origin coincides with the ray origin and whose z-axis is the ray’s direction vector. We start marching by obtaining the missing vertex index using the xor (⊕) field of the current tetrahedron. At each step, we get new vertices using the xor field and perform front-to-back color compositing. The accumulated opacity is obtained through Beer’s Law [5, 6, 19]. We use early ray termination with a threshold of 98%.

Each memory layout given in Sect. 5.1 comes with its marching algorithm which share the same skeleton. The differences come from unwrapping the applied compaction schemes. Alg. 2 summarizes the marching procedure for volume rendering. Algs. 1 and 2 can be generalized by changing the return types and sampling function.

Algorithm 2: Tetrahedral marching. \( v_{id} \) is the vertex index list for the current tetrahedron, \( index \) is the tetrahedron index and \( tetMesh \) is a list of tetrahedra, represented with one of Tet32, Tet20, or Tet16.

1: procedure MARCHVOLUME\(ray, v_{id0}, \ldots, 2, index, tetMesh) \n2: \( C_{vor}, \alpha \leftarrow (0,0,0), 0 \n3: \) \( v_{id3} \leftarrow tetMesh[index].VX \oplus v_{id0} \oplus v_{id1} \oplus v_{id2} \) \( \triangleright \) 3rd \( v_{id} \n4: V_0^' \ldots, 3 \leftarrow \text{projectToBasis}(V_{0,3}) \triangleright \) points in ray-space \n5: exitFaceId \leftarrow \text{GetExitFace}(V_0^', 3) \n6: while \( index \neq -1 \) AND \( \alpha < 1 \) do \n7: \( C_{vor}, \alpha \leftarrow \text{sample}(V_0^', \ldots, 3, v_{id0}, \ldots) \n8: index \leftarrow \text{matchToNextTet}(V_0^', \ldots, 3, exitFaceId, tetMesh) \n9: \) \( v_{idexitFaceId} \leftarrow v_{id3} \n10: \) \( exitFaceId \leftarrow v_{id3} \n11: \) \( v_{id} \leftarrow tetMesh[index].VX \oplus \text{projV}_{0,3} \oplus v_{id1} \oplus v_{id2} \n12: \) \( V_3' \leftarrow \text{projectToBasis}(V_{0,3}) \) \( \triangleright \) new point \n13: exitFaceId \leftarrow \text{GetExitFace}(V_0^', 3) \n14: return \( C_{vor}, \alpha \)

In order to connect to the next tetrahedron with Tet32, we determine which neighbor is behind the exit face based on the rank of the exit face’s index in the current tetrahedron’s vertex index list. For instance, let the exit face index be 42 and let the current tetrahedron’s vertex indices be \{32, 20, 42, 10\}. Then, the algorithm will pick the third neighbor index because 42 is the third in the list.

The Tet20 representation does not explicitly store vertex indices. Initially, we sort each neighbor index using its corresponding vertex index, i.e., a vertex that does not share an edge with that tetrahedron. During marching, to get the next tetrahedron, we pick \( n_{th} \) neighbor stored at the current tetrahedron where \( n \) is the sorted order of the last vertex index. For instance, let the current tetrahedron’s vertex indices be \{20, 10, 42, 32\}. Then algorithm picks the rank of the last vertex index as the next neighbour to be visited. When sorted, 32 falls into third place; hence, the third neighbor’s index is picked.

Marching to the next tetrahedron for the Tet16 representation is a bit more complicated since it inherits Tet20’s compaction steps and also reduces neighborhood information. During traversal, unlike Tet20, we also keep the index of the previous tetrahedron. In this way, we can extract the next tetrahedron index from the \( nx \) fields solving \( n_j = nx_j \oplus n_3 \) where \( j \) is the exit face index.

6 Increasing Rendering Quality

We describe how to implement several effects to achieve high-quality shading by tracing secondary rays. Using our marching algorithms that support traversal from arbitrary locations and in random directions, secondary rays naturally integrate with our framework.

6.1 Gradient Calculation

Gradients are commonly used as surface normals for local shading. As our method supports traversal starting at arbitrary origins, central-difference gradients [3, 5, 21] can be computed by marching six rays in orthogonal directions starting at the sample position, giving us correct, high-quality gradients even if the gradient sample positions \( \Delta x \) fall outside the tetrahedron that the current sample is inside. As this exact method is relatively costly, we restrict it to high-density regions (above 80%). Our marchers can, however, only stop at tet faces and not at arbitrary positions, whereas the gradient sample positions will generally fall somewhere in-between. While we could imagine using a more exact scheme or an interpolation method as proposed by Shirley and Tuchman [17], we found it acceptable in practice to evaluate the gradients at the position \( \Delta x' \) where the marcher stops and then divides the sample value by \( \Delta x' \). Gradient shading with the Phong model is demonstrated in Fig. 4 and Fig. 5.

Figure 4: Gradient-shaded depth cues: (a) Jets dataset with shading off (left) and on (right) (113.3 fps vs. 70.45 fps). (b) Fusion dataset with shading off (left) and on (right) (87.3 fps vs. 12.2 fps).

6.2 Volumetric Shadows

Another way to improve depth perception is by rendering shadows that involve tracing arbitrary rays, thus allowing volumes to cast shadows on surfaces. To calculate the radiance that reaches a certain point, we cast a shadow ray from that point using shell-BVH. If the shadow ray hits a volume shell, we start tetrahedra marching to
accumulate transmittance. Termination conditions and traversal are
the same as Alg. 2.

Marching needs to be initiated from a shell-face. If a volume and
surface mesh intersect, we cannot simply start from the tetrahedron
containing the incident point. Because we are only interested in the
transmittance and not in the radiance emitted from the volume,
the order of tetrahedra on the shadow ray’s path does not matter.
Instead of marching in the direction of the shadow ray, we go backward,
starting from the closest back-facing shell. If the light source lies
inside the volume, we start accumulating transmittance after passing
through the light’s position. Fig. 6 (c) displays this effect.

6.3 Ambient Occlusion

Ambient occlusion (AO) can help even more with depth perception
and overall rendering quality. We use the standard ray traced AO
method, for example, proposed in [2,4]. Tracing the required shadow
rays is technically very similar to tracing rays towards point light
sources located at a distance of \( r \). We compute AO by averaging \( N \)
hemisphere samples, the effect of which can be seen in Fig. 6 (b).

7 IMPLEMENTATION AND EVALUATION

Our implementation is based on OptiX 7 [11] and the OptiX Wrappers Library (OWL) [22], which allows us to make use of NVIDIA’s hardware ray tracing extensions. We use a ray generation program to initiate traversal and use optixTrace to trace rays against the shell’s triangle mesh. Our tetrahedra structures (see Sect. 5.1) reside in arrays in GPU memory.

Experiments are performed on a workstation with an Nvidia Quadro RTX8000 GPU and Ubuntu 18.04. We evaluate our approach on various tetrahedral volume meshes with varying memory sizes, some of them coming with surface meshes always at 1024×1024 resolution. We test each scene with all of our memory layouts with their respective marchers (see Sect. 5.1 and Sect. 5.2).

8 SUMMARY AND DISCUSSION

We propose shell traversal-based tetrahedra marching algorithms for
direct volume rendering. Our method offers high cache coherence
while tracing arbitrary rays, generating secondary effects such as
gradient shading, shadows, and ambient occlusion. It also benefits
from NVIDIA’s ray-tracing cores to achieve hardware acceleration.

Table 1 provides the computational cost of our implementation
for the tested scenes with various secondary effects using different
tetrahedral representations. In all cases, the Tet20 and Tet16 re-
presentations outperform the Tet32 layout. When we compare the
Tet20 and Tet16 representations, there is no significant difference in
performance. Using Tet20 or Tet16 has some practical advantages,
and we can select one depending on the memory constraints.

We also evaluate ray tracing overhead (see Table 1). We observe
that the memory overhead of OptiX scales reasonably well with
the number of shell faces. We see that the shell-face count is consider-
ably less than the total vertex count for the given scenes. Besides,
the shell traversal cost constitutes \( \approx 25\% \) of the total rendering time
for our test (the \( \approx 75\% \) is tet-marching plus effects). We show that,
unlike raster-based methods, applying any secondary effect that
requires tracing arbitrary rays is cheap and easy with our method.

Although our approach reduces memory consumption while al-
lowing efficient DVR, it has some limitations. Our method only
works with pure tetrahedral-meshes. Many large volumetric datasets
fail to comply with this constraint. Other types of meshes can be
tetrahedralized, but the tetrahedralization of different primitives in-
creases the memory cost. Some large datasets cannot be fit into the
Video Random Access Memory on the GPU when tetrahedralized.
Additionally, our marching procedure tends to go to the wrong tetra-
hedron and circle around when it encounters degenerate faces or
tetrahedra. We break these loops after a few iterations; they do not
produce any noticeable artifact to our observation.

ACKNOWLEDGMENTS

This research is supported by The Scientific and Technological
Research Council of Turkey (TÜBİTAK) under No. 117E881. The
Agulhas dataset is courtesy of Dr. Niklas Röber (DKRZ); the Deep
Water Asteroid Impact is courtesy of John Patchett and Galen Gisler
of LANL. Plasma64 dataset is courtesy of AIM@SHAPE. Jets and
Fusion datasets are courtesy of the SCI Institute of the University
of Utah. Hardware for development and testing was graciously
provided by NVIDIA Corp.
REFERENCES


