

LA-UR- 95 - 1169

Conf-9510137--1

Title: PARALLEL SPHERE RENDERING

Author(s): Michael Krogh  
Charles Hansen  
James Painter  
Guillaume Colin de Verdiere

Submitted to: IEEE Parallel Rendering Symposium 95

RECEIVED  
MAY 08 1995  
OSTI

**Los Alamos**  
NATIONAL LABORATORY



Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy.

Form No. 836 R5  
ST 2629 10/91

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

## **DISCLAIMER**

**Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.**

# Parallel Sphere Rendering

Michael Krogh, Charles Hansen, James Painter  
Advanced Computing Laboratory  
Los Alamos National Laboratory  
Los Alamos, New Mexico 87545

Guillaume Colin de Verdière\*  
Centre d'Etudes de Limeil-Valenton  
CEL-V/DMA/AIM  
94195 Villeneuve-Saint-Georges, France

## Abstract

Sphere rendering is an important method for visualizing molecular dynamics data. This paper presents a parallel divide-and-conquer algorithm that is almost 90 times faster than current graphics workstations. To render extremely large data sets and large images, the algorithm uses the MIMD features of the supercomputers to divide up the data, render independent partial images, and then finally composite the multiple partial images using an optimal method. The algorithm and performance results are presented for the CM-5 and the T3D.

**Key Words:** Parallel rendering, sphere rendering, compositing.

## 1 Introduction

In recent years, massively parallel processors (MPPs) have proven to be a valuable tool for performing scientific computation. Available memory on these types of computers is greater than that which is found on most traditional vector supercomputers. For example, a fully populated 256 node T3D has 16 gigabytes of physical memory. A 1024 node CM-5 contains 32 gigabytes of physical memory. As a result, scientists who utilize these MPPs can execute their three dimensional simulation models with much greater detail than previously possible. While current simulations don't typically utilize the entire memory systems of these machines, it is not uncommon for a data set from a single time-step in a dynamic simulation to be in excess of several gigabytes. For example, molecular dynamics simulations of structural materials have reached 600 million atoms [1, 2, 12]. While researchers don't usually perform simulations with 100 million atoms, 10 million to 40 million atom simulations are becoming routine. Figure 1 and Figure 2 show images of such data.

With such large data sets, visualization is an essential tool for analyzing the simulation output. Researchers wish to gain insight into both macroscopic phenomenon as well as microscopic phenomenon. Traditional methods, such as statistical analysis and browsing through the raw simulation data, aren't adequate by themselves for analyzing data sets ranging in size from gigabytes to terabytes. Visualization, because of its high bandwidth, enables a researcher to explore his or her data sets in a more timely and fruitful manner [14].

---

\* Author currently at the ACL through a grant from DGA/DRET

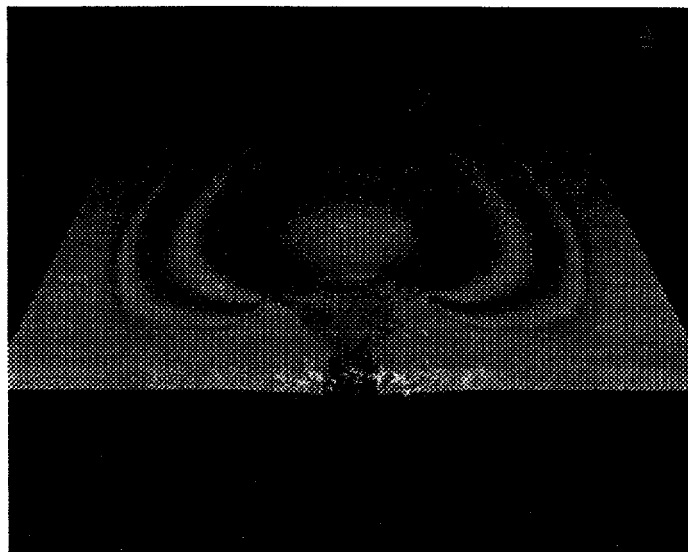


Figure 1: 104M Atoms

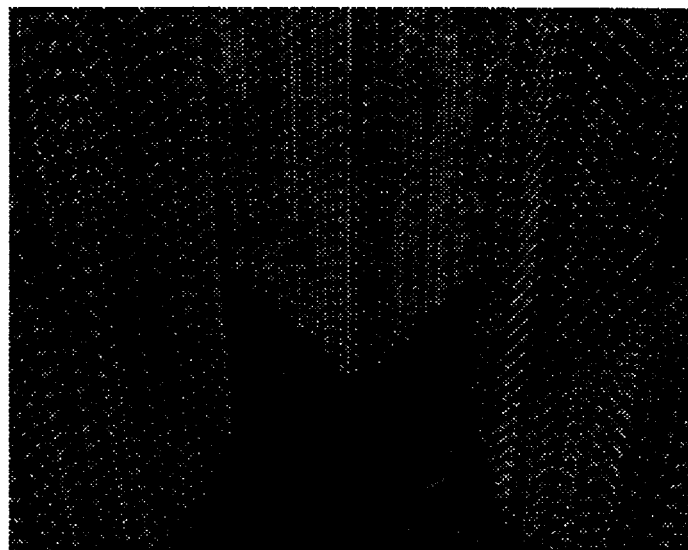


Figure 2: 104M Atoms (close up)

In structural molecular dynamics, researchers wish to see both the large scale structures in the data as well as microscopic phenomenon, such as the onset of dislocations. This requires portrayal of the entire data set in a single image, where individual atoms may not be distinguishable, and the ability to zoom in on minute details, where only a fraction of the data can be seen. Of course, researchers also wish to see these features as they vary with time.

For a molecular dynamics simulation, the atoms are usually represented by spheres. One could argue that with several tens of millions of atoms, most would be subpixel in size and that a pixel representation would be adequate. This argument is certainly true when the viewpoint is sufficiently far away from all of the data. In this situation the atoms shrink to pixel or subpixel sizes. However, we tend to move the viewpoint in and around the data set thus causing the atoms to cover tens of pixels in diameter. This technique is very useful for showing minute details within the data sets as can be seen in Figure 2.

Empirically, we have found that a sphere tends to be the best representational form for an atom in our simulations. Spheres can vary cleanly from subpixel sizes to covering the entire image in extreme cases; they seem to introduce the fewest artifacts. Also, it is a long standing and familiar representational form for atomic data.

The problem with spheres is that they can be expensive to render. Most graphics workstations tend to approximate spheres with a collection of polygons which causes a data explosion. Every sphere is tessellated into many tens of polygons. Instead of rendering 40 million spheres for a single time step, one may need to render upwards of 400 million polygons. Clearly, a different approach to rendering spheres is needed or vast improvements in hardware rendering.

Another question that can be asked is where to do the rendering. Given the availability of a MPP and a graphics workstation, one could render on either the MPP, the graphics workstation, or distributed across both. The graphics workstation is probably not the right choice for two reasons. First, 40 million atoms requires approximately 800 megabytes of storage and most workstations don't have sufficient memory to store many time steps of this size. Second, from our own experiments, we know it takes approximately 34 minutes to render 40 million atoms on a Silicon Graphics Inc. Onyx workstation equipped with RealityEngine2 graphics hardware. Additionally, it requires on the order of 30 minutes to move the data set from the supercomputer to the workstation. Two thousand time steps would take over 83 days to render! The distributed approach doesn't provide a solution either. The primary bottleneck is still the graphics hardware, and furthermore, moving hundreds of megabytes per time step can take a substantial amount of time even over high speed networks. Rendering on the MPP is an attractive solution given that the data resides on the MPP's disk farm and that the MPP has sufficient memory. The challenge becomes developing an algorithm that can exploit the multitudes of processors while minimizing overhead, typically communication costs.

## 2 Related Work

Spheres can be rendered using several approaches. One approach, which is used by Silicon Graphics Inc., is to tessellate a sphere into a set of polygons which can then be rendered as any other polygons using the custom hardware [10]. Several tessellation methods exist, such as a bilinear, octahedral, icosahedral, cubic, and others. The problem with this approach is that each sphere is transformed into many polygons. The number of polygons varies depending upon the degree of smoothness desired. Spheres with lesser degrees of tessellation appear less spherical than those with higher degrees of tessellation.

Others have used similar approaches. Staudhammer describes a method where spheres are tessellated but the shading is based on the sphere not the polygons, thus the rendered sphere has a more spherical appearance while perhaps using fewer polygons[18].

Porter's approach treats a sphere as a primitive in his scan line based rendering algorithm [17]. For any spheres that intersect the current scan line, the intersection of the sphere silhouette and the image plane is determined and then a Bresenham algorithm is used to calculate  $z$  values for hidden surface removal. Shading is approximated with a cosine function. Additional features of his algorithm include sorting for transparency and pixel averaging at sphere intersections and border intersections. This is done for antialiasing.

Starting with the Pixel-Planes machines through the PixelFlow machine, UNC has explored several parallel rendering methods [5, 6, 16]. The PixelFlow machine is different from its predecessors in that it switched from a screen subdivision process to an image composition process. Conceptually, given  $N$  primitives and  $P$  processors, each processor gets  $N / P$  primitives to render into its own frame buffer. After all processors have finished rendering, the frame buffers are composited together in a final image. With an image composition approach, it can be easier to avoid the problem of poor load balancing in the rasterization stage since all processors render and rasterize their own primitives with no communication. The only communication is the composition of the partial images. As described by Molnar et. al., the actual implementation of the PixelFlow machine is somewhat more complicated being comprised of several custom circuit boards: renderers, shaders, and a frame buffer [16]. Renderers are comprised of an Intel i860XP for 3D operations and a custom 128x128 SIMD pixel processor that performs rasterization of transformed primitives. Since the SIMD array is only 128x128, a 1280x1024 image must be computed in chunks. The partial images computed by the multiple renderer boards are composited together in 128x128 chunks and then sent to shading boards for operations that can be deferred, such as shading and texture mapping. The final image is sent from the shading boards to a frame buffer board for display.

### 3 Massively Parallel Processors at the ACL

The Thinking Machines Corp. CM-5 is a production quality massively parallel supercomputer [4]. The CM-5 can consist of 32 to 16384 Sparc processors. The CM-5 at the Advanced Computing Lab (ACL) at Los Alamos National Laboratory consists of 1024 processors, each with 32MB of local RAM for a total of 32GB. Each Sparc processors also has four 64-bit wide vector units which can assist in floating math operations. With four vector units up to 128 MFlops per node can be performed. This yields a theoretical speed of 128 GFlops for the ACL CM-5. In practice roughly 60 GFlops has been sustained. It should also be noted that not all vendor supplied languages use the vector units and may, in fact, rely on the Sparc processors to do all floating point math. In this case a user can expect to see only 5 MFlops per processor.

The 1024 node processors are hosted by 8 control processors (also Sparc processors). The control processors are responsible for system administration tasks and executing non-parallel code. A control processor and some subset of the node processors form a partition. A user's program is constrained to operating within a partition.

The CM-5 has three internal high-speed networks: the control network, the data network, and the diagnostic network. The control network is used for data operations, such as broadcasts, global opera-

tions, and combining operations, that involve all node processors in a given partition. It is also used for synchronization and error handling. The data network is used for routing data between nodes. Each node has two data network connections which are rated at 20 MB/s each, and the control network links are also rated at 20 MB/s. The diagnostic network is not available to user programs. High-speed I/O devices, such as parallel disk arrays, HIPPI network interfaces, and frame buffers are also attached to the CM-5 data network.

The T3D from Cray Research Inc. is a MPP consisting of 32 to 2048 DEC Alpha processors [9]. The ACL T3D has 256 processors, each with 64MB of RAM for a total of 16GB. Each Alpha chip is rated at 150 MFlops, thus the machine is theoretically capable of 38.4 GFlops total performance. However, we have seen speeds in the 2 MFlop to 20 MFlop per processor range on average. Like the CM-5 the nodes are arranged into partitions which must contain a power of 2 processors. The nodes are hosted by one of three CPUs of the YMP front-end. Unlike the CM-5, the partitions are dynamically determined and are not tied to a specific front-end host processor.

The T3D nodes (2 Alpha processors comprise a node) are organized into a torus network via 300 MB/s links. The torus is then connected to the YMP through two 200 MB/s high speed links. All I/O (including HIPPI, FDDI, and disk) from the Alpha processors is done through the YMP.

### 3.1 Programming Models

Both the CM-5 and the T3D support SIMD and MIMD programming models. The SIMD (Single Instruction, Multiple Data) model performs the same operation on all selected data elements. For example, given an array of numbers, a constant could be added to each number. When using the SIMD model, this operation would logically occur simultaneously on each element of the array. Actual hardware may or may not perform this operation simultaneously on all selected data elements. This would depend on whether or not enough physical processors exist for each element in the array.

The MIMD (Multiple Instruction/Multiple Data) model, divides a task up into a number of subtasks that can run concurrently and independently. Some subtasks can occur in parallel while others might occur serially. For example a set of processors might be used to render frames for an animation. Each processor would have a data set to render along with the viewing and/or motion transformations for its given frame within the animation. Each renderer can execute in parallel and independently from the other processors to produce a finished image. In this example, no synchronization is needed between individual processors.

MIMD programs can be either host/node programs or hostless programs. A host/node program consists of a host program, frequently referred to as a master program, and some number of node programs, also known as slaves. When a host/node program commences execution, the host contains the thread of execution and is responsible for invoking the node programs. The host typically communicates some initialization parameters to the nodes and then instructs them to perform some task. The node programs do not need to stay in synchronization after this point, although programs typically involve some synchronization points, usually where communication between nodes or between nodes and the host occurs. On the CM-5 communication is performed via the CMMD library whereas the T3D uses the PVM message passing library [7]. Hostless programs differ from host/node programs in that they do not have a user supplied host program as the name implies, although CMMD supplies a standard host program to provide initialization and I/O serving. On the T3D hostless programs can use either PVM, Cray's explicit shared memory library, or some other library such as FM [11].

## 4 Parallel Rendering Classification

Molnar et. al., provide a useful taxonomy for parallel rendering which classifies parallel rendering methods as sort-first, sort-middle, or sort-last [15]. This proves useful when deciding on how to structure a rendering algorithm. Basically, the rendering process requires a mapping from 3D object space to 2D image space which typically requires sorting the data. This sorting can occur at three different locations: at the beginning of the rendering process, during the middle, or at the end. Each method has its advantages and disadvantages.

Sort-first assigns each renderer to specific screen region. The renderers are responsible for the complete rendering process, both geometry processing and rasterization, for all primitives that map into their region of responsibility. The mapping of primitives from object space to image space is decided initially by a relatively simple pre-transformation operation. This approach is advantageous if a primitive has a high degree of tessellation, because less data is communicated, or when there is high frame-to-frame coherence, because fewer primitives need to be communicated on successive frames. One disadvantage of this approach occurs when there is a high degree of depth complexity. In this case load balancing becomes a problem.

With sort-middle, primitives are evenly divided among the processors where tessellation and geometry processing are performed. Then the transformed primitives are sent to the rasterizer processors. These processors, assigned to disjoint regions of the image plane, are responsible for rasterizing primitives that lie within their region of the image. The advantage of this approach is that geometry processing tends to be fairly well load balanced, however disadvantages include potentially poor load balancing for the rasterizers (for the same reasons as with sort-first) and high communication costs if tessellation is high.

With the sort-last method, also known as image compositing, each processor is responsible for the complete rendering process, as with the sort-first approach. Instead of mapping primitives up front to their correct image space processor, primitives are equally distributed among the processors at the beginning, rendered into a complete image, and then the images are composited. Two advantages of this method are: potentially easier load balancing, and image compositing times depend only on image size and not model complexity. The disadvantage is that each processor must have enough memory for the image buffer and that communication involves transferring the image buffer during compositing. This can be rather significant for very large images.

## 5 Parallel Sphere Algorithm

Our sphere rendering algorithm is based on a sort-last approach. This approach has two essential advantages for our problem: better load balancing and reduced communications. Better load balancing is achieved because the spheres are evenly distributed across all processors for both geometry processing and rasterizing. This is essential because our data exhibits an extremely high degree of depth complexity. If a sort-first or sort-middle approach was employed, we would have to either resort to virtual memory techniques due to insufficient memory per processor, or we would have to block communications between geometry processors and rasterizers as the rasterizers became inundated. Reduced communication cost can easily be seen. Instead of having to transfer 40 million atoms, each of which consists of 20 bytes of data, between geometry processors and rasterizers, we only need to transfer images or partial images. The cost of transferring spheres can even be greater if the spheres are tessellated. An additional cost



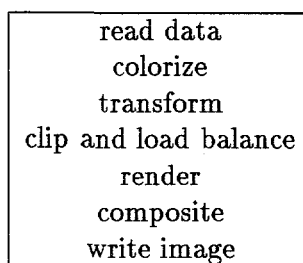


Figure 3: Steps in the Algorithm

saving feature is the use of a combining operation for image compositing which executes in  $O(N \log P)$  time on the CM-5, where  $N$  is the number of pixels per processor and  $P$  is the number of processors, and  $O(N)$  time on the T3D. This is discussed later in the paper.

The sphere renderer was originally designed as a MIMD program using a host/node model, written in C with CMMD, for the CM-5. The choice of MIMD over SIMD was made because of the desire to allow processors to compute asynchronously as much as possible. Even though each processor has the same number of spheres to render (plus or minus one), each sphere can be a different diameter in pixels. Additionally, spheres can be clipped if they fall outside of the image. This dictates varying amounts of computation. The host/node model was chosen over the hostless model because of the desire to display the resultant images through an X11 window. At this time the hostless model on the CM-5 does not support X11. Otherwise, the program can be easily structured using either model.

Molecular dynamics simulations are frequently time dependent and typically, the researcher wishes to see an animation of the time series. Thus, we included support for batch processing of sequences of data sets within the renderer.

Two additional features we required were the capability to render any number of atoms and images of any size while working within the memory constraints of any given partition size. The amount of memory available to the renderer is at most 32MB per node on the CM-5 (which might be shared with other running jobs) and 64MB per processor on the T3D (which is not shared); virtual memory does not exist. Because of this, the renderer must determine how much of the data set it can fit into memory as well as how much of the image will fit. If sufficient memory is not available, then either the data set, image, or both may be processed in chunks. Figure 3 shows the algorithmic steps for the main part of the renderer.

The only function of the host is to interact with the user and to initialize the nodes. The host will also be responsible for image display, via X11 or HIPPI frame buffers.

The main body of the renderer is a loop that iterates over the sequence of data sets. A data set, or some portion of it, is read in from disk: either an NFS (Network File System) mounted disk or from parallel disk (SDA on the CM-5 or RAID-3 on the T3D). The read operation on the CM-5, which looks like a normal UNIX read, has the option of automatically dividing up the data set among the processors in the current partition. For example, the first processor can read the first  $n$ 'th of the data, the second processor can read the second  $n$ 'th of the data, and so on.

Next each processor colorizes its atoms based on a scalar quantity associated with each atom. Atoms

are then transformed from object space to screen space. Part of the transformation operation includes transforming the atom radius into an image space radius with a perspective transformation (i.e. spheres closer to the viewpoint will appear to be larger than those further away). This is accomplished with a perspective scaling factor and linear interpolation along the  $z$  depth. While this is not entirely accurate, the results are more than adequate and faster than the “real” perspective transformation.

## 5.1 Load Balancing

After transformation certain spheres may fall outside of the image, these spheres can be clipped. This has lead to significant load balance problems. To compensate for this, we employ a simple heuristic. A list consisting of the number of visible atoms per processor is sorted. From this list the processors are paired up. The processor with the most atoms is paired with the processor with the least atoms; the processor with the second most atoms is paired with the second least, and so on. For a given pair of processors, the processor with the most atoms sends a portion of its atoms to the other so that both end up with the same number atoms plus or minus one.

Empirically, we’ve found that performing this simple load balancing step twice can yield good results while incurring only a small amount overhead (the third and fourth times cost more than they save). The choice of this simple method over more complex, but potentially more optimal, methods is due to the desire to avoid more complicated communication between nodes which can be very slow.

## 5.2 Scan Conversion

After all spheres have been transformed, each is scan converted, one at a time, into the image and  $z$  buffers. The image buffer consists of 4 bytes per pixel: red, green, blue, and alpha (which is not used at this time). The  $z$  buffer is a single floating point value per pixel. The scan conversion is done by evaluating a distance equation for each pixel within the bounding box for the sphere’s silhouette. If the current pixel is within the sphere, then a  $z$  buffer comparison is made to determine visibility. If this checks, then the color is determined. Instead of computing the true perimeter of the sphere, which involves a square root operation, we simply leave off the square root operation for both the distance check and the  $z$  buffer comparison. This yields significant savings.

We have additional savings in determining pixel colors. Instead of using photorealistic lighting calculations (e.g. Phong), we use a simple formula which approximates diffuse and ambient shading with a single light source with very good results. We compute the distance from the center of the sphere’s silhouette to the pixel. This value is used to linear interpolated an intensity that varies from “white” at the center to “almost black” at the perimeter. This value is multiplied by the atom’s color to achieve a “shaded” RGB color. This is similar to precomputing the intensity term in diffuse shading. To simulate a light source located above and to the right of the viewpoint, we add in an offset in the distance calculation.

## 5.3 Compositing

After all  $P$  processors have rendered their spheres, they synchronize and the  $P$  images are composited. Since transparency is not supported, we do not have to worry about compositing order. The CM-5 has a

CMMD function, `CMMD_reduce_v`, that implements a global reduction on an array using a specified operator. We use `CMMD_reduce_v` twice, once to composite the z buffer and then a second time to composite the image buffer. When compositing the z buffer, we specify a minimum operator and our z buffer as the array to reduce. After this call, each processor has a new z buffer array that contains the minimal z value across all processors for each pixel location. This will select the values for each pixel which are closest to the viewpoint. Next the processors loop over the composite z buffer and compare its value at each pixel location with that of its original z buffer. If the values match, then it knows that its pixel is closest to the viewpoint, otherwise it is obscured by some other pixel on another processor. In this case, the processor would set its pixel at that location to black, a minimal value. Following this, the processors would call `CMMD_reduce_v` again but this time specifying the image buffer along with the maximum operator. Only pixels still having a color will contribute to the resultant image. After this step each processor contains an identical image for the current image chunk. On the CM-5 this operation takes  $O(N \log(P))$  where  $P$  is the number of processors and  $N$  is the number of pixels in the image.

If the renderer was not able to read in the entire data set, it then reads in the next portion of the data set and renders those atoms for the same partial image. But instead of starting with an empty image and z buffer, it uses the ones which were previously calculated. After the entire data set has been processed, the first processor writes out the final image, or current partial image, to the image file. If the image is being computed in tiles, then the next tile is rendered. This continues until a full image is rendered.

## 6 T3D Issues

ACL acquired its T3D after we completed the CM-5 version of the renderer. We felt that an implementation for the T3D would be trivial since Cray had implemented PVM on the MPP. This proved to be false for several reasons. First, debugging was tedious due to the frequency of TotalView (the debugger) crashes. This prompted us to first implement the renderer on a cluster of SGI workstations running PVM and then to move this version to the T3D. Next we found that the Cray version of PVM has several "magic" environment variables which control memory utilization by PVM. Picking correct values was difficult at best; whereas picking wrong values resulted in a core file. We were unable to pick correct values to get our renderer to run on 128 or 256 processors. Finally, PVM routines were poorly implemented. Global operations appear to be implemented using linear techniques rather than more efficient logarithmic methods. Additionally, any PVM operation that involved the YMP, explicitly or implicitly, had very poor performance.

The routines for doing I/O had to be rewritten for the T3D. If all processors try to read their own portion of the data file, this results in overloading the YMP host since it performs the actual I/O and would spawn a task for each Alpha processor. To avoid this, we had the host read the data file and parcel out the data to the Alpha processors. This did not yield good performance due to PVM. Thus, the YMP was removed as the host and processor 0 inherited the host's tasks and all file I/O.

Ultimately, we abandoned PVM on the T3D due to poor performance and the inability to scale up to partitions larger than 64 nodes. It should be noted that after our experiences with PVM, we found an informative paper which describes many of the flaws with Cray's PVM implementation [3].

We were left with the choice of using either Cray's explicit shared memory library or some other

message passing library such as FM [11]. We did not want to use FM for this particular program for two reasons: it uses an active message model which would require extensive rewriting; and it does not support messages lengths as large as were required. The shared memory library places a fairly large burden on the user. The user must manage all synchronization between processors and cache coherency. Our decision was to develop a lightweight library, which looks quite similar to CMMD, on top of the explicit shared memory library. Our timings indicate that our additional layer incurs very little overhead beyond direct shared memory calls.

One improvement the T3D version has over the CM-5 version is in compositing. By writing our own global reduction operations, we were able to merge the two separate global reduction operations and intervening comparison into a single operation. This new z buffer reduction uses our Binary Swap compositing scheme which runs in  $O(N/2 + N/4 + N/8 + \dots + N/P)$  time where  $N$  is the number of pixels per processor and  $P$  is the number of processors<sup>1</sup> [13]. In the limit as  $P$  goes to infinity, this reduces to  $O(N)$  time. Since the total problem size grows by the product  $N \times P$ , an  $O(N)$  algorithm on  $P$  processors is optimal.

## 7 Results

To gauge our algorithm's performance we timed several experiments on various CM-5 partitions and T3D partitions. The test case is a data set consisting of 5 million atoms and the image size is 1024 x 1024. For comparison we also ran a test case on a Silicon Graphics Inc. Onyx with a RealityEngine2 graphics board to get a sense of how long a typical high-end graphics workstation would take to complete the task.

The SGI workstation uses a simple, but optimized, program that invokes the SGI sphere drawing routine. Their sphere drawing routine tessellates a sphere into a set of triangle or quadrangle meshes (depending on user selected tessellation method). The number of polygons generated depends upon the user selected tessellation factor. Using a tessellation factor of 3 and the octohedral tessellation method, the SGI renders our test data set in 261 seconds, including 5 seconds for reading the data.

One factor not included in the SGI times is the time it takes to transfer the data from the SDA to a local disk on the SGI. Over ethernet it takes several minutes for a 5 million atom data set. This is a can be a significant cost that we would incur while generating a time series animation.

Figure 4 and Table 1 show rendering times for various partition sizes for the CM-5. Our initial results are quite promising. **Total** is the time for the program to read in the data, render, and write out the image. **Read** includes the time to read in the data. **Write** is the time to write out the image. Both will vary from run to run due to other jobs using the SDA. **Composite** is the time the nodes spend doing compositing. **Computation** is the difference between the **Total** less the **Read**, **Write**, and **Composite**. As can be seen, the algorithm scales quite well across all partition sizes. The knee between 64 and 128 processors for the **Read** time corresponds to there being enough total memory to read the data set in a single pass. For partition sizes 32 and 64 two reads are required to process all of the data. The limiting factor for the CM-5 version of the renderer is the compositing operation.

---

<sup>1</sup>This analysis is correct for network topologies such as the hypercube where each path of communication is a single hop for this particular operation. Other topologies will have greater costs for some hops.

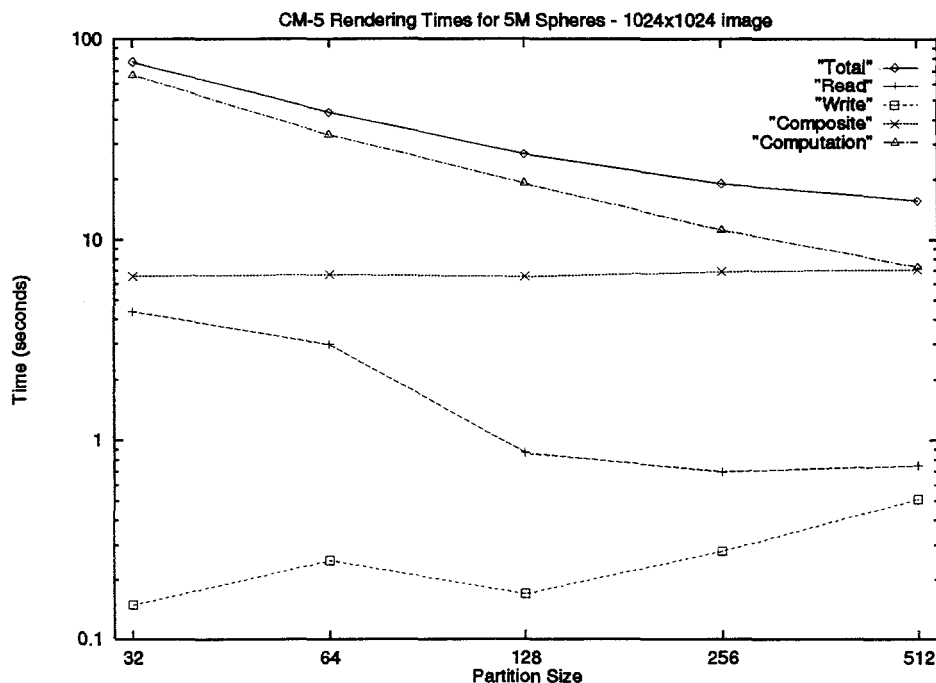


Figure 4: CM-5 Rendering Times

Partition	Read	Write	Composite	Computation	Total
32	4.37	0.15	6.57	66.57	77.66
64	2.99	0.25	6.68	33.45	43.37
128	0.87	0.17	6.54	19.20	26.78
256	0.70	0.28	6.93	11.21	19.12
512	0.75	0.51	7.11	7.34	15.71

Table 1: CM-5 Rendering Times

Figure 5 and Table 2 show rendering times for various partition sizes for the T3D. **Read** also includes the time spent moving the data from processor 0 to the other processors during the read operation. Again, **Read** and **Write** will vary due to other processes and YMP load. The knee between 16 and 32 nodes corresponds to there being enough total memory for the entire data set to be processed in one pass. Here the algorithm exhibits better scaling than the CM-5 version as can be seen with the compositing and computation times. Although, the limiting factor with the T3D version is reading and distributing the data.

When comparing the T3D version to the CM-5 version, it is easy to see that the processors in the T3D are significantly faster. However, the T3D is hindered by not having parallel disk I/O available to the Alpha processors directly, thus requiring one processor to perform the read and data distribution.

Computational times with load balancing turned on and turned off are presented in Figure 6 and Table 3. Here **Computation Time** also includes any time spend load balancing with associated communications. It should be noted that the CM-5 does not support partition sizes smaller than 32 nodes and that time on the 1024 node configuration was not available for these experiments. The numbers differ from the

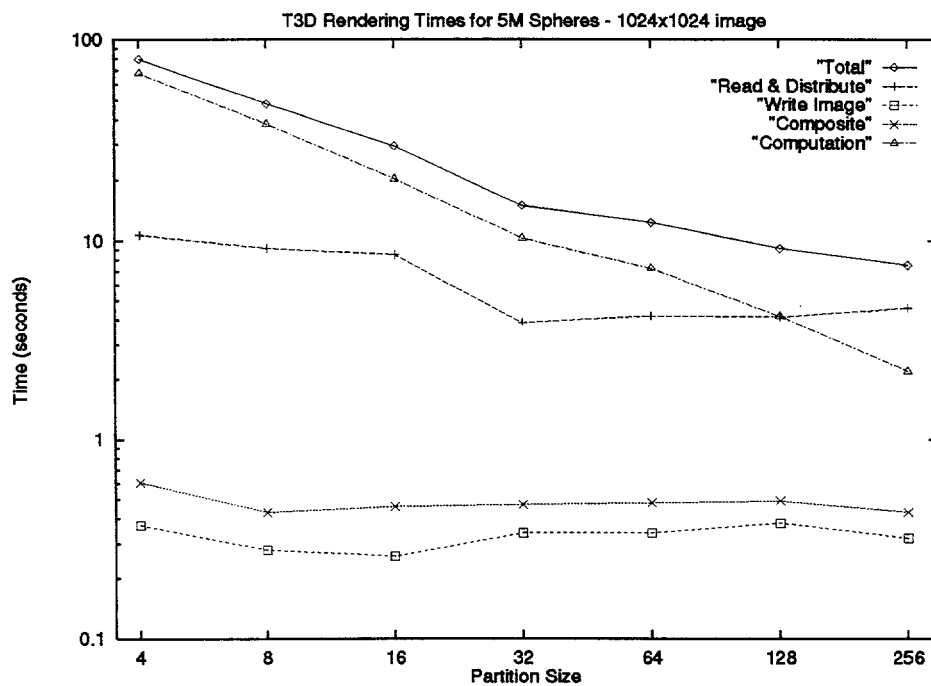


Figure 5: T3D Rendering Times

Partition	Read	Write	Composite	Computation	Total
4	10.71	.37	.61	68.25	79.94
8	9.20	.28	.43	38.25	48.16
16	8.52	.26	.46	20.39	29.63
32	3.87	.34	.47	10.30	14.98
64	4.20	.34	.48	7.28	12.30
128	4.14	.38	.48	4.16	9.17
256	4.61	.32	.43	2.21	7.57

Table 2: T3D Rendering Times

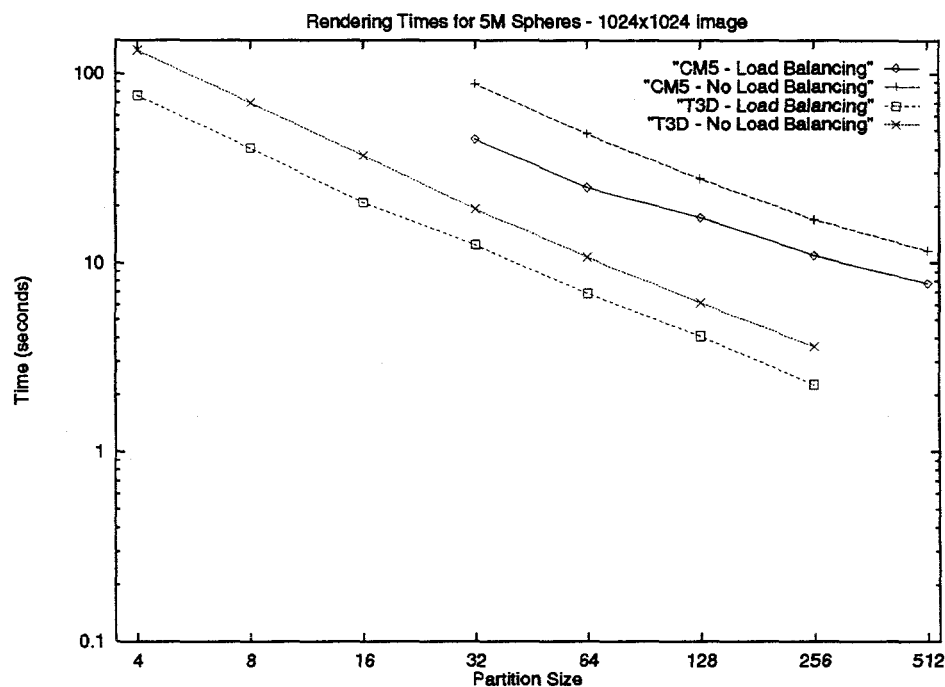


Figure 6: Computation Times

previous results due to a different viewpoint which results in roughly 1/3 of the atoms being clipped. It is easy to see that load balancing is beneficial for situations in which it can be applied.

## 8 Conclusions

We have shown that parallel sphere rendering can be performed at very high rates on massively parallel supercomputers and that MPPs can be significantly faster than a graphics workstation. Rendering rates on the three platforms are in the range of: 660K spheres/second for the T3D, 318K spheres/second for the CM-5, and 19K spheres/second for the SGI. If reading the data was not included in these rates, such as for an interactive viewer, then the rates would be: 1689K spheres/second for the T3D,

Partition	T3D LB	T3D no LB	CM-5 LB	CM-5 no LB
4	76.96	133.30	NA	NA
8	40.66	70.37	NA	NA
16	20.95	37.10	NA	NA
32	12.48	19.40	45.22	88.57
64	6.92	10.76	25.20	48.13
128	4.12	6.18	17.53	28.06
256	2.27	3.62	10.96	17.04
512	NA	NA	7.82	11.63

Table 3: Computation Times

334K spheres/second for the CM-5, and 19K spheres/second for the SGI. The algorithm makes use of several optimizations to enhance performance: direct sphere rendering, simplified rendering and lighting equations, trivial but efficient load balancing, and optimal compositing.

Our algorithm should be portable to other supercomputers or to workstation clusters due to its straight forward design. The algorithm only requires basic message passing primitives for communications, although an optimized global reduction operator can be of value.

Our algorithm has the advantage of not relying on a hardware specific rendering engine. This gives us the capability of integrating it in with other types of renderers, such as polygonal and volumetric. Writing a renderer that combines our polygon renderer, sphere renderer, and volume renderer is the focus of our next effort. Also, we plan to look at additional load balancing techniques that deal with spheres of varying radii [8].

## References

- [1] D. M. Beazley and P. S. Lomdahl. A practical approach to portability and performance problems on massively parallel supercomputers. In *To appear in Proceedings of Debugging and Performance Tuning for Parallel Computing Systems: Toward a Unified Environment*. IEEE Computer Society Press, October 1994.
- [2] D. M. Beazley, P. S. Lomdahl, N. Gronbech-Jensen, and P. Tamayo. High performance communication and memory caching scheme for molecular dynamics on the CM-5. In *Proceedings of the 8th International Parallel Processing Symposium*, pages 800–809. IEEE Computer Society Press, 1994.
- [3] S. Booth, A. Simpson, and C. Brough. Notes on porting PVM codes to the T3D. Technical report, Edinburgh Parallel Computing Centre, The University of Edinburgh, August 1994.
- [4] Thinking Machines Corporation. The Connection Machine CM-5 technical summary, 1991.
- [5] H. Fuchs et al. Fast spheres, shadows, textures, transparencies, and image enhancements in Pixel-Planes. In *SIGGRAPH 85 Conference Proceedings*, pages 111–120. Association for Computing Machinery, Inc., July 1985.
- [6] H. Fuchs et al. Pixel-Planes 5: A heterogeneous multiprocessor graphics system using processor-enhanced memories. In *SIGGRAPH 89 Conference Proceedings*, pages 79–88. Association for Computing Machinery, Inc., July 1989.
- [7] A. Geist et al. PVM3 user's guide and reference manual. Technical Report ORNL/TM-12187, Oak Ridge National Laboratory, Oak Ridge, TN, 1993.
- [8] C. Hansen, M. Krogh, J. Painter, and G. Colin de Verdière. Binary-swap volumetric rendering on the T3D. In *to appear in the Cray User Group 95 proceedings*, 1995.
- [9] Cray Research Inc. CRAY T3D hardware reference manual. Technical Report CMM-0602-000, Cray Research Inc., Chippewa Falls, WI, October 1993.
- [10] Silicon Graphics Inc. libsphere. online man page.
- [11] V. Karamcheti and A. Chien. A comparison of architectural support for the TMC CM-5 and the Cray T3D. In *Proceedings of ISCA '95*, 1995.



- [12] P. S. Lomdahl, P. Tamayo, N. Gronbech-Jensen, and D. M. Beazley. 50 GFlops molecular dynamics on the Connection Machine 5. In *Proceedings of Supercomputing 93*, pages 520–527. IEEE Computer Society Press, 1993.
- [13] K.L. Ma, J.S. Painter, C.D. Hansen, and M.F. Krogh. Parallel volume rendering using binary-swap compositing. *IEEE Computer Graphics and Applications*, 14(4):59–68, July 1994.
- [14] B.H. McCormick, T.A. DeFanti, and M.D. Brown. Visualization in scientific computing. *Computer Graphics*, 21(6), November 1987.
- [15] S. Molnar, M. Cox, D. Ellsworth, and H. Fuchs. A sorting classification of parallel rendering. *IEEE Computer Graphics and Applications*, 14(4):23–32, July 1994.
- [16] S. Molnar, J. Eyles, and J. Poulton. PixelFlow: High-speed rendering using image composition. In *SIGGRAPH 92 Conference Proceedings*, pages 231–240. Association for Computing Machinery, Inc., July 1992.
- [17] T. Porter. Spherical shading. In *SIGGRAPH 78 Conference Proceedings*, pages 282–285. Association for Computing Machinery, Inc., August 1978.
- [18] J. Staudhammer. On display of space filling atomic models in real-time. In *SIGGRAPH 78 Conference Proceedings*, pages 167–172. Association for Computing Machinery, Inc., August 1978.

## DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

---