A REALTIME SIMULATION BENCHMARK SUITE
FOR TESTING LOW COST VISUALIZATION SYSTEMS

Rodney Rogers, Gary Green, Michelle Sartor
Institute for Simulation and Training
3280 Progress Drive
Orlando, FL 32826

Pamela Woodard
STRICOM, AMSTI-ET
12350 Research Parkway
Orlando, FL 32826

ABSTRACT

The *Low Cost Visualization (LCV) Project* team at the Institute for Simulation has developed a prototype *Modeling and Simulation (M&S) Benchmark Suite* for testing 3D image rendering and realtime simulation capabilities of *Low Cost Visualization Systems*, i.e. PC-based computer systems capable of rendering 3D images and costing several thousand dollars. Previously reported research led to a focus on four OpenGL-based public domain benchmarks running under Windows NT. From these four, we selected *scenario tests* most resembling real-time simulation applications in scene complexity (terrain, culture, moving models and viewpoint, level of detail control, special effects, &c) and graphics rendering parameters (smooth shading, antialiasing, texture mapping, double buffering, transparency, hidden surface removal, &c). In addition, we included *primitive tests* designed to measure the maximum performance capabilities of an LCV system. Finally, we selected tests to evaluate the *rendering quality* of LCV systems, which turns out to be a matter of fundamental importance. We describe the M&S Benchmark Suite and analyze data produced by running it on nine different LCV systems. While the primary goal in testing was to verify the consistency and usefulness of the suite, our results reveal insights into performance capabilities of LCV Systems and software that runs on them. We also show how we revised the prototype M&S Benchmark Suite as a result of what we learned, and indicate future research directions in the LCV Project.

ABOUT THE AUTHORS

Rodney Rogers is a member of the Computer Science Department at Embry-Riddle Aeronautical University, Daytona Beach, Florida, and a Research Associate at the Institute for Simulation and Training. A former U.S. Navy Jet Carrier Pilot, he received Ph.D. degrees in Computer Science from the University of Central Florida and in English and American Literature from the University of Virginia.

Gary Green is a Research Associate and Operations Research Analyst at the Institute for Simulation and Training. He has extensive experience applying rigorous operations research techniques to planning, management, resource allocation, and other issues. A retired Colonel in the U.S. Army, he received an M.S. in Operations Research and System Analysis from the US Naval Postgraduate School.

Michelle Sartor is a Research Associate in the Visual Systems Laboratory at the Institute for Simulation and Training, where she helped develop physically based models for networked realtime simulations on both the Dynamic Terrain and Dynamic Virtual Environments projects. Ms. Sartor previously was employed in the Radar Department at Martin Marietta (now Lockheed-Martin). She received a B.S. in Biology from Florida State University, and a B.S. and M.S. in Electrical Engineering from the University of Central Florida.

Pamela Woodard is an Electronics Engineer in the Engineering Directorate at the U.S. Army's Simulation, Training, and Instrumentation Command, where she is a visual systems engineer on engineering and acquisition programs. She has over twenty years experience in visual simulation and synthetic natural environments. Ms. Woodard received a B.S. in Mathematics from George Mason University and an M.S. in Computer Engineering from the University of Central Florida.

A REALTIME SIMULATION BENCHMARK SUITE FOR TESTING LOW COST VISUALIZATION SYSTEMS

Rodney Rogers, Gary Green, Michelle Sartor
Institute for Simulation and Training
3280 Progress Drive
Orlando, FL 32826

Pamela Woodard
STRICOM, AMSTI-ET
12350 Research Parkway
Orlando, FL 32826

INTRODUCTION

An *LCV (low cost visualization) system* is a commercially available PC-based computer system capable of rendering 3D images and costing several thousand dollars. Under contract to STRICOM, the LCV Project Team at the Institute for Simulation and Training is constructing a *Modeling and Simulation (M&S) Benchmark Suite* to test the suitability of LCV systems for realtime simulation requiring 3D image generation. As explained in [R98, R98a], the current focus is on single processor Intel-based LCV systems running Windows NT, and using the OpenGL graphics API. The benchmark suite is comprised of tests drawn selectively from *public domain* benchmarks, i.e., benchmarks available free of charge via the Internet. We believe the suite is more objective, less complicated, and more focused on M&S needs than the public domain benchmarks from which it is derived.

Here we report on tests we conducted on LCV systems using the M&S Benchmark Suite. Our goal was to verify that the suite gives consistent results, and to perfect the suite by modifying its composition as necessary. What follows describes the tests in the Benchmark Suite and the computer systems used to execute the suite; presents and analyzes data gathered from running the suite; explains how we modified the M&S Benchmark Suite as a result of what we learned from the testing; and describes future work.

THE M&S BENCHMARK SUITE

The M&S Benchmark Suite is comprised of tests drawn from four public domain OpenGL benchmarks: *Viewperf* [VP], *GLPerf* [GLP], *Gemini Real-World* [GRW], and *Indy3D* [I3D]. For a description of these benchmarks, see [R98, R98a]. We categorize each test as a *scenario test* or a *primitive test*. A primitive test renders *graphics primitives* (lines, triangles, quadrangles, &c.) of a fixed size, typically using a fixed viewpoint and view volume. Ordinarily, projected primitives are all the same size and are rendered in a single frame. A primitive test can reveal a graphics system's maximum throughput. The output statistic is primitives drawn per second (e.g., triangles/ second) or pixels filled per second (pixels/second).

By contrast, a scenario test renders *complex graphics objects* (often called *meshes*) constructed from primitives as components of a graphics scenario depicting object motion. The output statistic is frames rendered per second (frames/second). Like a realtime simulation, a scenario test is characterized by object transformations (rotation, scaling, translation) and by a changing viewpoint, viewplane, and perhaps view volume. Objects are textured or shaded, and successive frames are rendered using double-buffering and hidden surface removal techniques. Realtime simulations, unlike scenario tests, are usually interactive and do not produce output statistics. Also, the former, unlike the latter, often must handle database paging and level of detail control, and must respond promptly to user input.

Table 1 presents database statistics for the nine scenario tests in the M&S Benchmark Suite. A missing entry indicates we were unable to obtain the relevant information. Italicized entries are explained later. *Gvr* and *gvf* are tests from the Gemini Real-World Benchmark. Gvr depicts racecars traversing a racetrack, while in gvf a fighter aircraft is flying at low to medium altitude over mountainous terrain. *Simulation* and *Animation* are Indy3D tests. Animation depicts a man walking on a city street, while Simulation shows a boat sailing in a bay. The remaining five scenario tests are drawn from various Viewperf *viewsets*. *Advanced Visualizer* (*AdvViz*) shows a human figure rotating in space. Unlike the other scenario tests in the M&S Suite, this test has a fixed viewpoint and viewplane. *Design Review* (*DesRev*) *Flat* and *Round* both simulate a tour through a model of the North Sea GYDA offshore oil production platform. In *Lightscape* (*Lt*) *Box* and *Parliament*, the viewpoint moves through rooms lighted using a proprietary *radiosity* algorithm. The Box test depicts a room with flat surfaces, while the Parliament test rooms have fluted columns, ornate woodwork and lights, and elaborate staircases.

The M&S Benchmark Suite also contains a single primitive test, the GLPerf TriangleStrip test, which renders strips of triangles of a fixed pixel size. Specifiable rendering parameters include triangle size, fill type (flat, Gouraud, texturing), and texture filtering method. As explained in Section 4, specifying small triangles stresses the CPU, while large size triangles stress the graphics card.

Benchmark Test	Ave Polys Rendered per Frame	Ave Vertices Transformed per Frame	Average Depth Complexity	Ave Polygon Size in Pixels	Total Polygons in Database	Total Vertices in Database	Number or Size of TexMaps
gvr	2,844	*9,746*	2.0-4.0		7,378	25,628	38 Maps
gvf	3,667	8,251	1.5-3.0		6,400	14,121	13 Maps
Simulation	7,257		2.15	574	7,710		3.64 MB
Animation	22,865		1.35	114	133,004		2.72 MB
AdvViz	*18,437*	65,576					
DesRev Flat	28,013	119,634					
DesRev Round	*21,526*	151,379					
Lt Box	281,182	1,124,728					
Lt Parliament	450,029	1,800,116					

Table 1. Database Statistics for Scenario Tests in the M&S Simulation Benchmark Suite.

Table 2 shows the 19 test instances in the M&S Benchmark Suite together with their rendering parameters. There are 15 scenario test instances in the suite. In addition, the primitive test TriangleStrip runs in four rendering modes, producing for each run, as seen later, a single plotted curve from the ten data points it generates. Thus the suite produces 19 *data objects* for analysis. We also use two additional public domain tests to check an LCV system's *rendering quality*. These tests are the *Glaze Benchmark* [GL] from Evans and Sutherland, and the Indy3D *Image Quality* test. See [R98, R98a] for details.

In Table 2, *Gouraud* alone indicates smooth shading without texturing. Texture mapping filtering modes are *Nearest* (one texel on a single texture mapped to one pixel); *Linear* (four texels on a single texture mapped to one pixel); *MMNearest* (one texel on a single mipmap mapped to one pixel); *MMLinear* (four texels on a single mipmap mapped to one pixel); *BiLinear* (one texel on each of two mipmaps mapped to one pixel); and *Trilinear* (four texels on each of two mipmaps mapped to one pixel). *Decal* indicates the texture is pasted on a facet without regard to lighting effects, while *Modulate* means a texture is combined with lighting to produce the final appearance. A "?" means we surmised an entry by observing the executing test, as opposed to deriving it from documentation or output generated by a benchmark.

Test	Lighting	Fill /Texture Filter	Tex Mode	Other	Window Size
gvr	Infinite (1)	Gouraud? Trilinear	Modulate?	Fog	1024x768
gvf	Infinite (1)	Gouraud? Trilinear	Modulate?	Fog	1024x768
Simulation 1	Ambient+Directed(1)	Gouraud		Fog, Trans	1024x768
Simulation 2	Ambient+Directed(1)	Gouraud Bilinear	Decal?	Fog, Trans	1024x768
Simulation 3	Ambient+Directed(1)	Gouraud Trilinear	Decal?	Fog, Trans	1024x768
Animation 1	Ambient+Directed(1)	Gouraud		Fog	1024x768
Animation 2	Ambient+Directed(1)	Gouraud Bilinear	Decal?	Fog	1024x768
Animation 3	Ambient+Directed(1)	Gouraud Trilinear	Decal?	Fog	1024x768
AdvViz 1	Local (2)	Gouraud			1024x768
AdvViz 2	Local (2)	Gouraud Nearest	Modulate	Texture 1	1024x768
AdvViz 3	Local (2)	Gouraud Trilinear	Modulate	Texture 2	1024x768
DesRev Flat	Infinite (1)	Gouraud Trilinear	Modulate		1024x768
DesRev Rnd	Infinite (1)	Gouraud Trilinear	Modulate		1024x768
Lt Box	Proprietary	Gouraud			1024x768
Lt Parliament	Proprietary	Gouraud			1024x768
TriangleStrip 1	Infinite (1)	Gouraud		10 Tri Sizes	768x768
TriangleStrip 2	Infinite (1)	Gouraud Nearest	Modulate	10 Tri Sizes	768x768
TriangleStrip 3	Infinite (1)	Gouraud Linear	Modulate	10 Tri Sizes	768x768
TriangleStrip 4	Infinite (1)	Gouraud Trilinear	Modulate	10 Tri Sizes	768x768

Table 2. Rendering Parameters for M&S Benchmark Suite Scenario and Primitive Test Instances.

LAB ENVIRONMENT AND TEST PROCEDURES

We executed the M&S Benchmark Suite on each of three Dell Dimension Intel-based computer systems configured to run Windows NT. Our Pentium machine has a 166 MHz processor with 16 KB of Level I cache, and 64 MB of main memory. Our Pentium II machines have 266 MHz and 300 MHz processors, both with 32 KB of Level I cache and 128 MB of main memory. All three machines have 512 KB of Level II cache, MMX capability, and 66 MHz PCI bus architectures. In addition, the 300 MHz machine has a 66 MHz Accelerated Graphics Port (AGP).

We alternated three 3D graphics controller boards in the 300 MHz, 266 MHz, and 166 MHz machines: a Leadtek Winfast 3D L2200, an Elsa GLoria Synergy, and a reference board from an OEM whose name we are not authorized to mention. For convenience, we refer to these boards as the *L2200*, the *Elsa*, and the *ABC* respectively. The L2200 has a 3D Labs Permedia NT graphics processor, while the Elsa has a Permedia 2 processor; both use Delta geometry and have 8 MB of on-card memory. The ABC uses a different chipset and has 15 MB of memory. We are aware that the Permedia NT chip is an earlier and now obsolete version of the Permedia 2 chip. However, since our aim was to validate the M&S Benchmark Suite, rather than evaluate current LCV systems, we believe use of the L2200 card is not a questionable procedure.

By combining three machines with three graphics controllers, we produced a total of 3 x 3 = 9 distinct LCV systems, then ran the M&S Benchmark Suite on each system. All machines were disconnected from the network, and only essential processes were memory resident. To assure data integrity, we used batch files to execute the tests, and independently ran the suite twice on each system. Results achieved are repeatable to within 1% accuracy or better, usually much better.

ANALYSIS OF DATA

The *speedup* one LCV system achieves over another system for a given test is the ratio of the two rendering rates. For example, if System A runs a test at an average frame rate of 15.0 frames/second, and System B runs the same test at 10.0 frames/second, then the speedup System A achieves over system B is 15.0 / 10.0 = 1.50, and the speedup (sometimes called *slowdown*) B achieves with respect to A is 0.67.

In most LCV *graphics pipelines*, the CPU does 3D vertex operations (rotation, scaling, translation, projection, lighting), while 2D operations (triangle setup, rasterization) are accomplished on the graphics card. The CPU puts the results of its operations in a vertex queue where the card can access them in order. The *bottleneck* in the pipeline is the point where processing proceeds most slowly. Since hardware buses connect the CPU, main memory, and the graphics card, the bottleneck at any given point can exist in the CPU, on the graphics card, or on the buses. As seen below, in scenario tests the bottleneck migrates back and forth between the CPU and the graphics card, while primitive tests can be made to bottleneck in either place by adjusting primitive size.

Scenario Tests

Table 3 gives average frame rates for scenario test instances for the nine LCV machines we tested. We have arranged the results from top to bottom in descending order of rendering speed on the system with the most powerful graphics card (ABC) and the most powerful CPU (300 MHz). Frame rates for individual tests on the fastest system vary from a low of 0.74 to a high of 23.33, with intermediate results distributed evenly. This reveals that Benchmark Suite tests are adequate to challenge relatively powerful LCV systems while still accommodating relatively weak ones.

Test	300 MHz			266 MHz			166 MHz		
	ABC	Elsa	L2200	ABC	Elsa	L2200	ABC	Elsa	L2200
gvr	23.33	12.01	6.87	21.89	11.67	7.29	15.65	11.30	6.40
gvf	21.62	11.87	6.02	20.49	*11.90*	6.47	15.14	11.31	5.71
Simulation 1	19.98	*14.33*	*8.32*	18.02	*14.90*	*8.72*	11.32	*11.34*	*6.81*
Simulation 2	16.16	11.11	5.13	14.89	11.02	5.93	10.18	9.09	4.58
Simulation 3	15.97	10.83	4.86	14.59	10.69	5.10	9.28	8.63	4.19
Animation 1	12.94	10.60	*6.66*	11.33	9.87	*6.60*	6.36	6.42	*4.57*
Animation 2	12.54	9.31	4.93	11.04	8.76	4.95	6.34	5.97	3.69
Animation 3	12.23	9.32	4.91	10.79	8.66	4.95	6.22	5.95	3.69
AdvViz 1	9.21	6.76	5.06	8.48	6.22	4.74	3.66	3.08	2.58
AdvViz 2	7.45	6.45	4.58	6.59	5.85	4.19	3.04	2.83	2.35
AdvViz 3	6.69	6.43	4.56	5.93	5.84	4.20	2.70	2.82	2.34
DesRev Flat	2.89	3.08	2.03	2.44	2.76	1.98	1.26	1.61	1.18
DesRev Rnd	2.34	2.53	1.62	1.98	2.33	1.58	1.00	1.38	0.97
Lt Box	1.22	1.14	1.09	1.08	1.03	0.99	0.47	0.45	0.49
Lt Parliament	0.74	0.70	0.69	0.67	0.64	0.62	0.30	0.29	0.32

Table 3. Frame Rates for Scenario Tests Sorted Descending on the 300 MHz/ABC Rendering Speed

Notice that the columns of Table 3 are nearly sorted (the ten "misordered" entries are italicized). This suggests that the M&S Benchmark Suite will give consistent results across different LCV Systems. Now consider that the tests in Tables 1 and 3 are in the same order, and that in Table 1 the tests (with three italicized exceptions) are sorted ascending on increasing number of polygons rendered per frame in Column 2, and on average vertices transformed per frame in Column 3. Thus these two naïve metrics, independent of polygon size, depth complexity, and related properties, seem reasonable predictors of rendering speed on LCV systems. Also, nine of the ten misordered entries in the columns of Table 3 are for the Elsa and L2200 cards rendering Simulation and Animation in smooth shading mode, implying that these cards perform smooth shading much more efficiently than texturing.

Test	ABC			Elsa			L2200		
	300/266	300/166	266/166	300/266	300/166	266/166	300/266	300/166	266/166
gvr	1.07	1.49	1.40	1.03	1.06	1.03	0.94	1.07	1.14
gvf	1.06	1.43	1.35	1.00	1.05	1.05	0.93	1.05	1.13
Simulation 1	1.11	1.77	1.59	0.96	1.26	1.31	0.95	1.22	1.28
Simulation 2	1.09	1.59	1.47	1.01	1.22	1.21	0.87	1.12	1.29
Simulation 3	1.09	1.72	1.57	1.01	1.25	1.24	0.95	1.16	1.22
Animation 1	1.14	2.04	1.78	1.07	1.65	1.54	1.01	1.46	1.44
Animation 2	1.14	1.98	1.74	1.06	1.56	1.47	1.00	1.34	1.34
Animation 3	1.13	1.97	1.76	1.08	1.57	1.46	0.99	1.33	1.34
AdvViz 1	1.09	2.52	2.32	1.08	2.19	2.02	1.07	1.96	1.84
AdvViz 2	1.13	2.45	2.17	1.10	2.28	2.07	1.09	1.95	1.78
AdvViz 3	1.13	2.48	2.20	1.10	2.28	2.09	1.09	1.95	1.79
DesRev Flat	1.18	2.29	1.94	1.12	1.91	1.71	1.03	1.72	1.68
DesRev Rnd	1.18	2.34	1.98	1.09	1.83	1.69	1.03	1.67	1.63
Lt Box	1.13	2.60	2.30	1.11	2.53	2.29	1.10	2.22	2.02
Lt Parliament	1.10	2.47	2.23	1.09	2.41	2.21	1.11	2.16	1.94

Table 4. CPU Speedups by Graphics Card, Tests Sorted Descending on 300 MHz/ABC Rendering Speed.

Test	300 MHz			266 MHz			166 MHz		
	ABC / Elsa	ABC / L2200	Elsa / L2200	ABC / Elsa	ABC / L2200	Elsa / L2200	ABC / Elsa	ABC / L2200	Elsa / L2200
gvr	1.94	3.40	1.75	1.88	3.00	1.60	1.38	2.45	1.77
gvf	1.82	3.59	1.97	1.72	3.17	1.84	1.34	2.65	1.98
Simulation 1	1.39	2.40	1.72	1.21	2.06	1.71	1.00	1.66	1.67
Simulation 2	1.45	3.15	2.17	1.35	2.51	1.86	1.12	2.22	1.98
Simulation 3	1.47	3.29	2.23	1.36	2.86	2.10	1.08	2.21	2.06
Animation 1	1.22	1.94	1.59	1.15	1.72	1.50	0.98	1.39	1.40
Animation 2	1.35	2.54	1.89	1.26	2.23	1.77	1.06	1.72	1.62
Animation 3	1.31	2.49	1.90	1.25	2.18	1.75	1.05	1.69	1.61
AdvViz 1	1.36	1.82	1.34	1.36	1.79	1.31	1.19	1.42	1.19
AdvViz 2	1.16	1.63	1.41	1.13	1.57	1.40	1.07	1.29	1.20
AdvViz 3	1.04	1.47	1.41	1.02	1.41	1.39	0.96	1.15	1.21
DesRev Flat	0.94	1.42	1.52	0.88	1.23	1.39	0.78	1.07	1.36
DesRev Rnd	0.92	1.44	1.56	0.85	1.25	1.47	0.72	1.03	1.42
Lt Box	1.07	1.12	1.05	1.05	1.09	1.04	1.04	0.96	0.92
Lt Parliament	1.06	1.07	1.01	1.05	1.08	1.03	1.03	0.94	0.91

Table 5. Graphics Card Speedups by CPU, Tests Sorted Descending on 300 MHz/ABC Rendering Speed

Table 4 shows speedups attributable to increased CPU power, i.e. speedups achieved by the 300 MHz over the 266 MHz, the 300 MHz over the 166 MHz, and the 266 MHz over the 166 MHz. Table 5 shows speedups resulting for increased graphics card power, i.e. speedups achieved by ABC over Elsa, ABC over L2200, and Elsa over L2200. In both tables, the tests are in the same order as in Table 3. It may also be helpful to recall that the 300 MHz and 266 MHz machines have Pentium II CPUs, while the 166 MHz machine has a Pentium; and that in general the ABC card renders faster than the Elsa, and the Elsa faster than the L2200.

In Column 2 of Table 4, the 300 MHz with the ABC card achieves speedups over the 266 MHz ranging from 1.07 to 1.18, with the average being 1.12. This is commensurate with the relative clock speeds of the two machines, 300/266 = 1.13. From Column 3, the 266 MHz is an average of 2.08 times faster than the 166 MHz, with the clock speed ratio being 300/166 = 1.81. We attribute the additional speedup to the fact that the 266 MHz has a Pentium II, and the 166 MHz a Pentium. In Column 4, the speedup averages 1.85, and the corresponding clock speed ratio is 266/166 = 1.60.

We conclude that average rendering speed of scenario tests in the M&S Suite scales with processor power. However, larger speedups tend to occur in tests which render at low frame rates, implying that these tests bottleneck more often in the CPU than on the graphics card. Thus tests toward the bottom of the table will be increasingly useful as CPU power increases in future LCV systems, while tests near the top are currently the most suitable for comparing relative graphics card power. Similar observations apply to data in Table 4 for the Elsa and L2200 cards, but since these cards are less powerful than the ABC card, the CPU is less often a bottleneck, which accounts for the fact that speedups due to increased CPU power are more modest. A counter-intuitive result is that not infrequently the 300 MHz is slightly slower than the 266 MHz with the L2200 card.

In Columns 2, 3, and 4 of Table 5, the average speedups in the 300 MHz for the ABC Card over the Elsa card, the ABC card over the L2200 card, and the Elsa card over the L2200 card are respectively 1.12, 2.18, and 1.63. However, the maximum speedups are 1.94 (gvr), 3.59 (gvf), and 2.23 (Simulation), and these three tests, as revealed above, are among those where the graphics card is the primary bottleneck. That is, when the three cards receive data from a CPU in a timely fashion, the ABC card is nearly twice as fast as the Elsa card and three and one-half times faster than the L2200, while the Elsa card is more than twice fast as the L2200. The data in Table 5 for the 266 MHz and the 166 MHz also support this assertion. A related idea is that the LCV systems we tested tend to be limited more often by graphics card power than by CPU power. However, this might not have been the case had we used the most powerful graphics cards available. As an example, we have an Obsidian 100SB-4440 card from Quantum3D which in the 166 MHz, the slowest CPU, at 800 x 600 resolution and using OpenGL drivers, renders gvf at more than 51 frames/second, and gvr in excess of 38 frames/second. We did not include this card in our tests, however, since it is based on the 3Dfx Voodoo chip, which runs only these two tests in the M&S Suite.

Primitive Tests

Table 6 gives data yielded by running various instances of TriangleStrip test. In tabular form, these data are not easy to interpret. The ABC card emerges as the best triangle renderer, followed by the Elsa and L2200 in order. Smooth shading is faster than texturing on all three cards. The Elsa and L2200 perform linear and trilinear texturing at the same speed, which could mean that there is no penalty for a better image, but is actually explained by the fact that these two cards cannot do trilinear mipmapping, as determined independently. Finally, systems using the ABC card can render more 25 pixel than 10 pixel triangles per second, which is counterintuitive.

Displayed in graphical form, the same data are more easily comprehended. Suppose a graphics card can rasterize k pixels/second in a given rendering mode, and that triangles contain p pixels. Then if the card is the bottleneck, the system should rasterize $t = k / p$ triangles/second, i.e., $t \cdot p = k$, which is a polynomial of degree two, specifically a hyperbola symmetric to $t = p$. If however the CPU is the bottleneck, it cannot process 3D vertices and normals fast enough to provide the card an uninterrupted supply of 2D triangles to rasterize. In that case, $t = k_t$, where k_t is the number of triangles (i.e., vertices and normals) the CPU can process per second. This plots as a straight line parallel to the p-axis. The point where the two curves intersect gives the triangle size (k / k_t) where the CPU and graphics card are *balanced*. If, however, we plot the curves on axes with logarithmic scales, then *two* straight lines appear, since $\log(t \cdot p) = \log t + \log p = \log k = k'$, which with triangle size p on the horizontal axis and triangles/second t on the vertical, gives a straight line with a slope of negative one.

That is, if the data in any row of Table 6 is graphed using logarithmic scales, the result should approximate the ideal curve shown in Figure 1. Increased CPU power moves the horizontal portion of the curve upwards, while more rasterization power moves the diagonal portion to the right. Increased area beneath the curve indicates increased rendering power. Minor deviations from the predicted shape in plots of actual data are probably due mainly to efficiency considerations in accessing the vertex buffer between the CPU and graphics card, and to interaction between the circuits on the card that do triangle setup and those that compute pixel values for rasterization. In Figure 2, we have plotted 24 rows of data from Table 6. There are many approaches to interpreting the data in the table. What follows is meant to be suggestive rather than exhaustive.

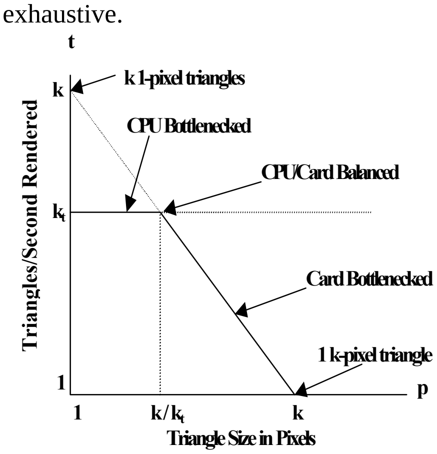

Figure 1. Idealized Plot of Triangle Throughput Using Logarithmic Scaled Axes.

CPU	Card	Fill	Triangle Size in Pixels 10	25	50	100	200	300	400	500	750	1000
300 MHz	ABC	smooth	764	965	688	479	257	174	133	108	73	55.5
		MMnear	593	712	538	421	241	165	127	103	69.9	53.1
		bilinear	596	717	536	358	195	132	100	80.6	54.3	41.1
		trilinear	413	464	385	335	190	129	97.8	78.9	52.3	40.2
	Elsa	smooth	780	519	358	237	151	116	95.1	81.9	61.9	50.4
		MMnear	569	443	300	187	117	89.3	74.4	64.4	50.1	41.6
		bilinear	562	329	236	143	88.9	67.1	56	48.6	38.8	33.2
		trilinear	563	359	237	143	88.9	67	56.1	48.7	38.8	33.2
	L2200	smooth	556	324	215	135	80.9	60.5	48.9	41.6	30.4	24.2
		MMnear	455	278	181	113	67.3	49.5	40.3	34.2	25.5	20.7
		bilinear	313	145	85.0	47.1	26.4	19.1	15.7	13.4	10.7	9.2
		trilinear	313	146	85.0	47.2	26.5	19.1	15.7	13.4	10.7	9.2
266 MHz	ABC	smooth	623	777	582	473	256	175	134	108	73	55.5
		MMnear	465	527	423	385	241	165	127	103	69.7	53.1
		bilinear	464	527	426	350	195	132	100	80.8	54.3	41.1
		trilinear	348	383	325	299	190	129	97.8	78.9	53.1	40.3
	Elsa	smooth	795	566	393	260	166	127	104	89.4	66.6	53.7
		MMnear	563	481	328	204	128	98.2	81.4	70.3	54.2	44.7
		bilinear	567	390	258	157	97.6	73.6	61.4	53.2	42.0	35.9
		trilinear	567	391	258	157	97.5	73.4	61.3	53.1	42.0	35.9
	L2200	smooth	598	377	250	158	95.3	70.6	56.4	47.8	33.7	26.3
		MMnear	348	167	98.4	54.5	30.6	22.1	18.1	15.4	12.3	10.5
		bilinear	348	167	98.4	54.5	30.6	22.0	18.1	15.5	12.2	10.5
		trilinear	348	167	98.4	54.5	30.6	22.1	18.1	15.4	12.3	10.5
166 MHz	ABC	smooth	255	302	231	207	256	175	133	108	73.0	55.5
		MMnear	199	231	184	166	216	165	127	102	69.7	53.1
		bilinear	200	230	182	166	194	132	99.9	80.6	54.3	41.0
		trilinear	148	158	134	126	156	128	97.6	78.8	53.2	40.2
	Elsa	smooth	328	318	321	261	166	127	104	89.5	66.7	53.8
		MMnear	238	234	235	204	128	98.2	81.4	70.4	54.2	44.8
		bilinear	238	231	234	157	97.8	73.7	61.5	53.4	42.2	35.9
		trilinear	238	231	235	157	97.8	73.7	61.4	53.4	42.1	35.9
	L2200	smooth	311	304	215	135	80.9	60.5	48.9	41.7	30.4	24.1
		MMnear	198	194	181	113	67.4	49.5	40.3	34.2	25.5	20.7
		bilinear	198	145	85.0	47.1	26.4	19.1	15.7	13.4	10.7	9.17
		trilinear	197	145	85.0	47.2	26.4	19.1	15.7	13.4	10.7	9.20

Table 6. Results of TriangleStrip Tests in Thousands of Triangles Rendered per Second.

The 12 curves on the left in Figure 2 reflect data produced by the ABC card installed in the 300 MHz, the 266 MHz, and the 166 MHz. They suggest that the ABC card favors triangles of size approximately 25 and 200 pixels, for all four rendering modes. The diagonal curve portions shows that, for triangles of 200 pixels or larger, the ABC card performs nearest texturing with about the same efficiency as smooth shading, that there is no added expense for trilinear as opposed to linear texturing, and that the latter two rasterization modes are somewhat more expensive than the former two. The distribution of the 12 “horizontal” curve portions suggests that CPU power is extremely important in an application that uses small triangles. In addition, we see that different rasterization modes impose different costs even when small triangle sizes ensure the CPU is the main system bottleneck. The available information on proprietary system architectures does not allow explaining this counterintuitive result.

The curves on the right of Figure 2 result from data produced when the three cards we tested were installed in the 300 MHz system. The horizontal curve portions for the L2200 and Elsa cards are absent, indicating that these two cards in the 300 MHz bottleneck even with small triangle sizes and is a waste of CPU power. If we were to show the Elsa and L2200 curves for the 266 MHz and the 166 MHz, we would see this is true for the 266 MHz as well, but that horizontal curve portions appear for both cards on the 166 MHz, implying that power of these two cards is better matched with the 166 MHz than with the faster machines. This reinforces our earlier observation that graphics card power limits rendering speed more than CPU power in the LCV systems we used to evaluate the M&S Benchmark Suite.

Evaluation of the M&S Benchmark Suite

We believe the M&S Benchmark Suite produces data which allow a consistent, reasonably thorough evaluation of an LCV system. The scenario tests in general have image generating characteristics typical of realtime simulations, e.g., moving objects, a changing viewpoint, lighting, and smooth shading or texturing. Moreover, these tests have a good distribution of image complexity measured in average polygons rendered and vertices transformed per frame. A drawback is that they lack certain important properties of realtime simulations, e.g., level of detail control, collision detection, page management for large databases, and special effects such as smoke, fire, and explosion debris. We return to this subject later. Finally, the sole primitive test, TriangleStrip test, allows independent examination of rendering power in either the CPU or graphics card, and determination of triangle sizes for which the CPU and card are approximately balanced.

One drawback of the Benchmark Suite is that the scenario test instances do not use consistent rendering parameters, making it difficult to compare rendering efficiency between tests. For example, AdvViz uses nearest and trilinear filtering, Simulation and Animation use bilinear and trilinear, and gvr, gvf, and Design Review use only trilinear.

A second drawback is that 3Dfx Voodoo based cards, such as the Obsidian card from Quantum3D, can execute only gvf and gvr among the tests in the suite, because the Voodoo chip operates in full-screen mode only. This difficulty is not insignificant, since it is widely accepted that 3Dfx chips are among the most powerful rasterizers available for LCV Systems.

A third drawback is that our TriangleStrip test instances do not produce sufficient data to plot horizontal curve portions accurately, as evidenced in Figure 2. Since the break point between horizontal and diagonal curve portions occurs at around 100-200 pixel triangles on current LCV systems, we need to collect additional data in the small triangle region.

THE REFINED BENCHMARK SUITE

To address shortcomings identified above, we have refined the M&S Benchmark Suite as shown in Table 7, first by adding two scenario tests which run in full screen mode. Gvb is a Gemini Real-World test which we believe will soon be available on the web. The scenario involves flying through a battle zone with helos, tanks, missiles, ground-to-air fire, burning trains, and explosions, smoke, and similar effects. Quake II Demo [Q2] is a public domain version of the famous computer game of the same name. Thus four tests now exist which allow

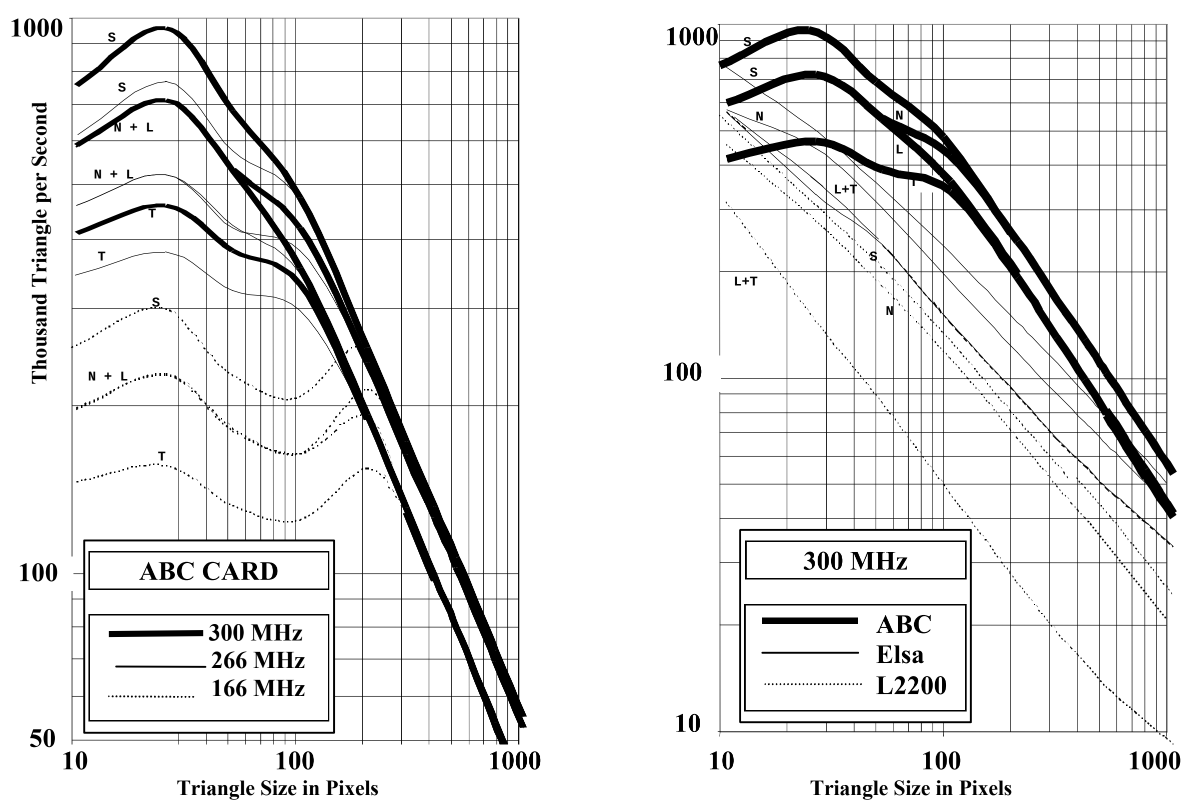

Figure 2. TriangleStrip Test Results for Graphics Cards on 300 MHz and for 3 CPUs with ABC Card (S = Smooth; N = Nearest; L = Linear; T = Trilinear)

using graphics cards which render in full screen mode only. We include instances of all four tests at 800x600 and 1024x768 resolution, since the Voodoo chip will not render at the higher mode. Then, to keep the total number of scenario tests in the suite low, we eliminated the Design Review Flat and Lightscape Box tests.

Second, we have added additional triangle sizes for TriangleStrip test instances, raising the total to 23 (varying from 10 to 3000 pixels), and have included corresponding instances of GLPerf QuadStrip test, which produces rectangle strips using the OpenGL quad strip data structure instead of triangle strips. Moreover, we have added GLPerf LineStrip tests to draw line strips in both aliased and antialiased mode, allowing comparisons of rendering speeds between the two. A total of 20 line sizes vary between 10 and 750 pixels.

Finally, where possible we have made polygon fill rendering modes consistent between tests. We believe the four rendering modes shown in Table 7 are most consistent with image quality required in realtime simulation. As reflected in the table, the revised M&S Benchmark Suite has 28 scenario test instances and ten primitive test instances. As the data points generated by the TriangleStrip, QuadStrip, and LineStrip tests are combined into ten distinct plotted curves, the Suite produces 38 data objects, a manageable number.

Test	Lighting	Fill /Texture Filter	Tex Mode	Other	Window Size
gvr 1	Infinite (1)	Gouraud? Trilinear	Modulate?	Fog	1024x768
gvr 2	Infinite (1)	Gouraud? Trilinear	Modulate?	Fog	800x600
gvf 1	Infinite (1)	Gouraud? Trilinear	Modulate?	Fog	1024x768
gvf 2	Infinite (1)	Gouraud? Trilinear	Modulate?	Fog	800x600
gvb 1	Infinite (1)	Gouraud? Trilinear	Modulate?	Fog	1024x768
gvb 2	Infinite (1)	Gouraud? Trilinear	Modulate?	Fog	800x600
Simulation 1	Ambient+Directed(1)	Gouraud	Decal?	Fog, Trans	1024x768
Simulation 2	Ambient+Directed(1)	Gouraud MMNearest	Decal?	Fog, Trans	1024x768
Simulation 3	Ambient+Directed(1)	Gouraud Bilinear	Decal?	Fog, Trans	1024x768
Simulation 4	Ambient+Directed(1)	Gouraud Trilinear	Decal?	Fog, Trans	1024x768
Animation 1	Ambient+Directed(1)	Gouraud	Decal?	Fog	1024x768
Animation 2	Ambient+Directed(1)	Gouraud MMNearest	Decal?	Fog	1024x768
Animation 3	Ambient+Directed(1)	Gouraud Bilinear	Decal?	Fog	1024x768
Animation 4	Ambient+Directed(1)	Gouraud Trilinear	Decal?	Fog	1024x768
Quake II Demo 1	Ambient?+Directed?	Gouraud Tex Filter Unk	Modulate?	Sound	1024x768
Quake II Demo 2	Ambient?+Directed?	Gouraud Tex Filter Unk	Modulate?	Sound	800x600
AdvViz 1	Local (2)	Gouraud	Modulate		1024x768
AdvViz 2	Local (2)	Gouraud MMNearest	Modulate		1024x768
AdvViz 3	Local (2)	Gouraud Bilinear	Modulate		1024x768
AdvViz 4	Local (2)	Gouraud Trilinear	Modulate		1024x768
DesRev 1	Infinite (1)	Gouraud	Modulate		1024x768
DesRev 2	Infinite (1)	Gouraud MMNearest	Modulate		1024x768
DesRev 3	Infinite (1)	Gouraud Bilinear	Modulate		1024x768
DesRev 4	Infinite (1)	Gouraud Trilinear	Modulate		1024x768
Parliament 1	Proprietary Algorithm	Gouraud	Modulate		1024x768
Parliament 2	Proprietary Algorithm	Gouraud MMNearest	Modulate		1024x768
Parliament 3	Proprietary Algorithm	Gouraud Bilinear	Modulate		1024x768
Parliament 4	Proprietary Algorithm	Gouraud Trilinear	Modulate		1024x768
LineStrip	Infinite (1)	Gouraud	Modulate	20 Line Sizes	1024x768
LineStrip AA	Infinite (1)	Gouraud	Modulate	20 Line Sizes	1024x768
TriangleStrip 1	Infinite (1)	Gouraud	Modulate	23 Tri Sizes	1024x768
TriangleStrip 2	Infinite (1)	Gouraud MMNearest	Modulate	23 Tri Sizes	1024x768
TriangleStrip 3	Infinite (1)	Gouraud Bilinear	Modulate	23 Tri Sizes	1024x768
TriangleStrip 4	Infinite (1)	Gouraud Trilinear	Modulate	23 Tri Sizes	1024x768
QuadStrip 1	Infinite (1)	Gouraud	Modulate	23 Quad Sizes	1024x768
QuadStrip 2	Infinite (1)	Gouraud MMNearest	Modulate	23 Quad Sizes	1024x768
QuadStrip 3	Infinite (1)	Gouraud Bilinear	Modulate	23 Quad Sizes	1024x768
QuadStrip 4	Infinite (1)	Gouraud Trilinear	Modulate	23 Quad Sizes	1024x768

Table 7. Rendering Parameters for Test Instances in Revised M&S Benchmark Suite.

CONCLUSION AND FUTURE WORK

The quality of the revised M&S Benchmark Suite is high, given available public domain benchmarks and our decision to use OpenGL as the rendering API. However, the suite will need to be continually updated as new benchmarks appear. As a next step in utilizing the suite, we will obtain high-end graphics boards and test them in 300 MHz, 266 MHz, and a recently acquired Dell Dimension 400 MHz

Pentium II machine with AGP and an 100 MHz internal bus. The 166 MHz machine represents outdated technology and will no longer be used for testing. We are also in the process of porting the M&S Benchmark Suite (except the Quake II Demo test) to SGI machines running Unix, where we expect results of running the suite on these machines to substantiate our claim that LCV Systems are already cost-effective alternatives to Unix Workstations for realtime simulation. Also, we currently are investigating ways to cause 3Dfx Voodoo based graphics cards to run Viewperf, GLPerf, and Indy3D benchmark tests.

We believe the M&S Benchmark Suite is a good instrument to measure image generation power for realtime simulation. However, it is clear that individual tests in the suite do not tax an LCV system to the same extent an actual simulation would. This is because no tests undertakes high-overhead simulation tasks such as collision detection, paging management for large databases, level of detail control, special effects, etc., with the exception of the Quake II demo, which does collision detection and services the game logic overhead, and gvb, which contains special battlefield effects. Moreover, no test in the suite has to deal with the problem of *transport delay*, i.e., of servicing user input in a timely fashion. Thus, although the M&S Benchmark Suite may reveal adequate image rendering power in LCV systems to support realtime simulation, it remains to be determined if sufficient additional processing power exists to sustain the overhead of these and similar computational tasks.

To address this question, we are considering the possibility of implementing a realtime simulation using LCV systems. As long as the CPU does 3D graphics processing in LCV systems, we anticipate that the excess rendering power in single processor machines may not be adequate for our purposes. Even when 3D computations move to the graphics card, as seems already to be happening, it may be necessary to use multiprocessor NT machines or to distribute different simulation tasks to various loosely coupled parallel LCV machines. We hypothesize that the former approach will be cheaper and far more efficient than the latter. Also, it will be necessary to examine high-level graphics APIs which run under NT on top of OpenGL to determine which is most suitable for realtime simulation. Three with which we are familiar are *OpenGVS* from Gemini Technology, *World Toolkit* from Sense8, and *SortVR* from Soft Reality; many additional APIs exist. We also wonder what the recent SGI/Microsoft partnership will produce by way of combining OpenGL and Direct3D into a new API.

Finally, we are currently analyzing M&S terrain databases in Open Flight format (*Quantico*, *Hunter-Liggett*, *Monterey*) using an SGI UNIX tool, *Perfly*. Our goal is to define database complexity in a way that allows relating it to rendering speed. Then, we will compare the simulation database complexities to the complexity of databases used by the M&S Benchmark Suite scenario tests. Finally, using data we collect from running the M&S Benchmark Suite on Unix machines together with our relative database complexity information, we will formulate an estimation of how fast the simulation databases could be rendered on the LCV systems in our lab. This provides us an approach to estimating how much additional processing power we might need in a multi-processor LCV machine to implement a realtime simulation on these systems using the simulation databases we are analyzing.

ACKNOWLEDGEMENT

The U.S. Army under STRICOM contract N61338-97-K-0010 funded research reported in this paper. We also wish to thank Jeff Potter of Real3D, who suggested the method we used to graph the data produced by TriangleStrip test.

REFERENCES

[GL] Evans and Sutherland Corporation, *Glaze Benchmark*, available from http://www.es.com.

[GLP] OpenGL Performance Committee, *GLPerf Benchmark* (author: John F. Spitzer: spit@sgi.com), available from http://www.specbench.org.

[GRW] Gemini Technology Corporation, *Gemini Real-World 3-D Benchmarks*, available from http://www.gemtech.com.

[I3D] Sense8 Corporation, *Indy3D Benchmark*, available from http://www.sense8.com.

[Q2] id Software, *Quake II Demo*, available from http://www.idsoftware.com.

[R98] Rogers, Rodney, Gary Green, and Brian Goldiez, *Preliminary Testing of Low Cost Visualization Systems Using Public Domain Benchmarks*, Institute for Simulation and Training Report, February 1998, available from http://www.vsl.ist.ucf.edu/groups/lcv/.

[R98a] Rogers, Rodney, Brian Goldiez, Pamela Woodard, and Benito Graniela, Testing Low Cost Visualization Systems Using Public Domain Benchmarks, *Proceedings of the 1998 Image Conference*, Scottsdale, Arizona, 2-7 August 1998.

[VP] OpenGL Performance Committee, *Viewperf Benchmark*, available from http://www.specbench.org.