Graphical and Arithmetical Benchmarking of Workstations*

Gerhard Kurka
e-mail: kurka@gup.uni-linz.ac.at
SFB F013, Project 1301

Abstract

This paper summarizes the results of a series of performance tests carried out as part of an evaluation of different graphics workstations. The graphics performance was determined by Viewperf 5.1. In addition, we present some results of a relatively simple graphics benchmark, which measures the raw performance with regard to Gouraud-shading. In order to get a good estimation of the processing power under real world conditions, a multi-grid software package was executed on all machines.

1 Systems and their configuration

In this report we investigate the graphics and processor performance of three contemporary commercial graphics workstations. In addition, we will also compare these machines with an ordinary mainstream PC.

1.1 Octane/MXE by Silicon-Graphics, inc.

1.1.1 Configuration

processor	MIPS R10000
clockrate	250Mhz
L1 cache size	32KB for instructions, 32KB for data
L2 cache size	1MB for instructions and data
spec95 integer / floating point	13.6 / 24
execution units / max. simultaneaus issues	5 / 5
pipe-stages (integer, load/store, fp)	5 / 6 / 7
main memory	640MB SDRAM
number of processors	2
width of the data-bus	—
width of the address-bus	—
system interconnection system	7x7 crossbar
graphics subsystem	MXE with 4MB texture cache
operating system	Irix 6.4

1.1.2 Remarks

Silicon Graphics, inc. claims that their Octane/MXE offers the most powerful desktop graphics available, providing unmatched realism and polygon performance. However, a direct head-to-head comparison with the two competitors Sun and HP has shown that SGI's Octane/MXE offers a relatively low graphics performance in the case of large Gouraud-shaded surfaces, only textured surfaces are drawn at high speed.

The MXE graphics subsystem consists of two geometry-accelerators providing a peak performance of 1.344 Gflops each, and two rasterizers that deliver together a peak fill rate of approximately 240 million Gouraud-shaded pixels per second. In spite of its texturing capabilities it provides only a 4MB texture-cache on the graphics board. This leads to a notable loss in performance if the desired textures do not fit into the small cache.

*Acknowledgments: This work is supported by the Austrian Science Fund – 'Fonds zur Förderung der wissenschaftlichen Forschung' – within the SFB F013 "Numerical and Symbolic Scientific Computing".

1.2 C240 by Hewlett-Packard

1.2.1 Configuration

processor	HP PA-RISC 8200
clockrate	236Mhz
L1 cache size (off chip)	2MB for instructions, 2MB for data
L2 cache size	none
spec95 integer / floating point	15.5 / 25
execution units / max. simultaneaus issues	10 / 4
pipe-stages (integer, load/store, fp)	7 / 9 / 9
main memory	1024MB ECC EDO-DRAM
number of processors	1
width of the data-bus	—
width of the address-bus	—
system interconnection system	crossbar
graphics subsystem	VisualizeFX6 with texture extension
operating system	HP UX-10.20

1.2.2 Remarks

The C240 is the successor of the C180/Visualize48. The new Visualize/FX6 subsystem offers twice the performance of the Visualize48.

An interesting detail of this machine is that the CPU doesn't contain an on-chip L1 cache. Instead of including the L1 cache on the chip-area, HP has attached a large external L1 cache (2MB for instructions, 2MB for data) with the ability to hide memory latencies in hardware [2]. Although this approach entails some physical limits on the operating frequency of the system because of the tight coupling between the CPU and the SRAMs, it makes the PA-8200 extremely fast when accessing large memory blocks in main memory. However, the new PA-8500, which is the successor of the PA-8200, has broken with the HP-tradition of using an external L1 cache. The PA-8500 includes a huge 1.5MB L1-Cache on the chip-area, this approach enables higher operating frequencies than with the old design.

With respect to our tests, the C240 has been proven to provide the highest graphics and processing performance of the three workstations.

1.3 Ultra60 by Sun Microsystems

1.3.1 Configuration

processor	SUN UltraSPARC II
clockrate	296 Mhz
L1 cache size	16KB for instructions, 16KB for data
L2 cache size	2MB for instructions and data
spec95 integer / floating point	13.2 / 18.4
execution units / max. simultaneaus issues	9 / 4
pipe-stages (integer, load/store, fp)	9 / 9 / 9
main memory	512MB
number of processors	2
width of the data-bus	128
width of the address-bus	36
system interconnection system	SUNs Ultra Port Architecture
graphics subsystem	Elite3D m6
operating system	Solaris 2.6.0

1.3.2 Remarks

The recently available Elite3D m6 is the first high-end 3D-accelerator produced by Sun Microsystems, which delivers an outstanding graphics performance for scientific and engineering applications. Contrary to HP's Visualize/FX6 and SGI's MXE the Elite3D m6 doesn't implement any hardware-acceleration for textured surfaces. Instead, its architecture has been carefully tuned for applications like Computer Aided Design (CAD), where a huge set of vertices has to be transformed by the geometry-accelerator on the graphics board. In order to provide high performance for such applications, the Elite3D m6 employs six independent geometry-accelerators for vertex-processing, triangle setup, and lightning calculations. Rendering is done by two fast rasterizers, which are directly connected to the so called 3D-RAM memory.

This new kind of graphics memory was specially developed to speed up load/modify/store pixel-operations often invoked on the depth-buffer. For this purpose 3D-RAM includes a small SRAM cache and an ALU which can do a depth comparison and selective write. Once the pixel color and depth value are delivered by the rasterizer the 3D-RAM handles everything from there on [4]. Thereby, 3DRAM technology reduces the time required by depth buffer operations from 160ns to only 10ns [3].

1.4 Intel PC running at 350Mhz (100Mhz memory clock)

processor	Intel Deschutes
clockrate	350Mhz
L1 cache size	16KB for instructions, 16KB for data
L2 cache size	0.5MB for instructions and data
spec95 integer / floating point	13.9 / 10.9
execution units / max. simultaneaus issues	5 / 5
pipe-stages (integer, load/store, fp)	—
main memory	512MB
number of processors	2
width of the data-bus	64
width of the address-bus	32
system interconnection system	Intel BX AGP-Chipset
graphics subsystem	Diamond FireGL 1000 pro
operating system	Linux-2.1.113

1.4.1 Remarks

With the availability of the advanced Intel-BX chipset [5], which can be found inside many of today's PCs, Intel systems have raised their performance up to workstation level. The improved SD-RAM interface of Intel's BX chipset and its successors makes it possible to use mainstream PCs for solving memory demanding numerical problems. Various analysis of the memory throughput have shown that a modern PC is able to outperform many of its competitors in random memory accesses. In addition, the PC provides a remarkably high memory bandwidth in respect of linear read and write operations of about 450MB to 600MB per second.

Because of the lack of any 3D-hardware support at the time of creating this report, the machine wasn't benchmarked for its graphics performance.

2 Graphics-Benchmarks

2.1 Viewperf 5.1 results

Viewperf was originally developed by IBM employes to provide a test bed where 3D-graphics hardware can be measured under real-world conditions. Viewperf reports graphics performance in frames per second. Other information about the system under test – all the rendering states, the time to build display lists, and the data set used – are also output in a standardized report. Viewperf uses a series of real-world OpenGL [1] application datasets, which in common are called viewsets. Each viewset simulates the graphics load of a certain application area. Currently, there are five standard viewsets. Because our application field focuses on pre- and postprocessing with respect to finite element applications, only the two viewsets DX-03 and CDRS-03 were considered for benchmarking.

2.1.1 Results of the viewset DX-03

The viewset DX-03 was generated by IBM's Data Explorer (DX). DX is a general-purpose software package for scientific data visualization and analysis. The tests visualize a set of particle traces through a vector flow field. Data such as that might result from simulations of fluid flow through a construction. Triangle meshes are the primitives for this viewset. The object represented contains about 1000 triangle meshes, containing approximately 100 vertices each. DX-03 does not include any texture maps.

HP C240/VisualizeFX6	Sun Ultra60/Elite3D m6	Octane/MXE
28.98	19.79	18.37

Due to the small size and the huge amount of visible triangles the frame rate – the number of images presented to the user per second – depends to a high degree on the performance of the geometry-accelerators and on the bandwidth between main memory and graphics board. We observe that SGI and Sun deliver similar frame rates while that delivered by HP's C240 is significant higher. The obvious performance gain of HP's C240 over SGI and Sun might be a result of HP's powerful geometry-accelerators. More precisely, the Visualize/FX6 [6] graphics-board includes three geometry accelerator-chips, each chip contains two floating-point processors based on HP's own PA-RISC technology. Together all floating-point processors of the graphics-board deliver a peak performance of approximately 3.6 GFlops. In addition, the Visualize/FX6 is equipped with four fast rasterizer chips.

2.1.2 Results of the viewset CDRS-03

This viewset simulates CDRS a CAD-system for industrial-design. CDRS was originally created by Evans and Sutherland, but it is now owned by Parametric Technologies, Inc. It is used to create concept models of interiors and exteriors in the automotive industry and for appliances. The viewset simulates a coarse surface-model of a lawn-mover. CDRS does, however, include two tests that use a small (860KB) texture map. Graphics subsystems that provide hardware acceleration of texture processing usually show higher frame rates than such that do not accelerate textures.

The model is depicted in three different ways: as wireframe, as Gouraud-shaded surface model, and as textured surface model. The resulting frame rate depends to a high degree on the fill rate of the rasterizers, especially on the fill rate for lines.

HP C240/VisualizeFX6	Sun Ultra60/Elite3D m6	Octane/MXE
200.00	130.18	55.40

Contrary to viewset DX-03 the performance of the geometry-accelerators doesn't play a dominant role here, because only a small number of vertices are processed. The results rather reflect the performance of the rasterizers with respect to lines and Gouraud-shaded surfaces. We recognize that Sun and HP deliver a significant higher frame rate than SGI's Octane/MXE, which suffers from its slower rasterizers that are not well optimized for rendering of lines and Gouraud-shaded surfaces.

The lack of any hardware-texturing capabilities on the Elite3D m6 results in a much lower frame rate as compared to HP. But it should be mentioned that in spite of the fact that Sun's Elite3D m6 doesn't provide texture-acceleration in hardware it outperforms SGI's fully equipped Octane/MXE.

2.2 Depiction of Gouraud-shaded surfaces

The depiction of Gouraud-shaded surfaces is a crucial operation in scientific visualization, especially within the field of finite-element/finite-volume methods. For this reason an additional benchmark originally created by Mark J. Kilgard as part of SGI's Glut graphics library was applied. The benchmark generates an OpenGL display-list, which contains quad-strips that form a Gouraud-shaded torus. In all, 10 000 Gouraud-shaded, depth buffered and lit quads are drawn. The following table shows the resulting frame rates at different window resultions.

Window resolution	C240/VisualizeFX6	Ultra60/Elite3D m6	Octane/MXE
256x256	226.5	273.0	108.0
1024x768	189.0	226.0	73.0
1280x1024	173.0	181.5	55.5

As expected, Sun and HP deliver higher frame rates than SGI's Octane/MXE. A closer look at the frame rates of the Octane/MXE shows that there exists a noticeable correlation between the window resolution and the frame rate. From that we can conclude that the Octane/MXE mainly suffers from its low fill rate concerning Gouraud-shading. Moreover, there are two other reasons why the Octane/MXE couldn't compete with Sun and HP in this test:

- In contrast to Sun and HP, SGI's Octane wasn't well optimized to accelerate OpenGL display-lists, in place of that it was tuned for the OpenGL immediate mode.

  On Sun and HP the usage of display-lists for static models leads to a performance gain of up to 300% compared to the immediate mode.

- An independent review of the reported benchmark results by SGI has shown that by rewriting the benchmark so that it uses triangle-strips instead of quad-strips, a performance gain of about 100% is possible. This is because the Octane/MXE requires quads to be tested whether they are convex or not. This test isn't neccessary if quads are replaced by triangles.

  It seems that HP and Sun automatically split quads into triangle-pairs, which don't need such time consuming tests before rendering actually occurs.

2.3 Experiences with real-world applications

Beside the results of the different benchmarks within this report, we have applied the application GMV [8] (General Mesh Viewer, developed at Los Alamos National Laboratories) on all three platforms in order to get a more realistic impression of the graphics performance under real-world conditions. To our surprise the achived frame rates were reflecting the results of the viewset DX-03. While there was no noticeable difference between the Ultra60/Elite3D m6 and the Octane/MXE, HP's C240/VisualizeFX6 was delivering frame rates 25% to 50% higher than both of its competitors.

2.4 CPU Performance

To test the performance of the CPU and the throughput to main memory we used cds.

cds is a relatively simple multi-grid solver, it calculates the solution of a finite-difference problem by using multi-grid methods in order to reduce the calculation time. In contrast to many other CPU benchmarks cds accesses large arrays within main memory, hence the calculation time depends on the bandwidth between CPU and memory rather than on the bandwidth between CPU and cache. This makes the benchmark's results a good indicator for software applied to solve large scale scientific problems.

Due to the complicated address generation code contained in implementations of multi-grid techniques, the resulting codes contain much integer blended code. Therefore, cds measures the overall performance of the CPU rather than its raw floating-point performance.

To compare the performance of the tested graphics workstations with contemporary PC technology, we have also added the results determined on a Linux-PC equipped with a 350Mhz-Deschutes (PentiumII) processor, a GigaByte BX mainboard, and 128MB-SDRAM (100Mhz/8ns).

2.4.1 Results of the cds-benchmark

The results represent the time (seconds) required for solving the equation system. Note, the numbers in the headline of the table indicate the memory size (MB) of the used equation system.

Machine	CPU	0.125	0.49	1.9	7.4	29
SUN Ultra60	UltraSPARC II 300Mhz	0.01	0.06	0.28	1.49	6.80
SGI Octane/MXE	MIPS R10000 250Mhz	0.01	0.04	0.23	1.18	5.30
HP C240	HP PA-RISC 8200 236Mhz	0.01	0.04	0.15	1.00	4.40
PC	IntelBX 350Mhz 128MB	0.01	0.05	0.27	1.21	5.09

2.4.2 Annotation to the performance of the IntelPC

The reader may be surprised that Intel has raised the power of its mainstream processors up to high-end RISC level. Despite the fact that Sun, SGI, and HP are using CPUs that offer approximately twice the floating point performance of the PentiumII, they were not able to outperform the Intel-CPU in our tests. The reasons for that are summarized in following points:

- The Intel PC provides a very high memory bandwidth between the CPU and the SDRAM-memory. Data can be fetched with about 500MB per second from the main-memory into the L2-cache of the CPU. One disadvantage of using low cost memory is that it doesn't ensure data integrity by ECC-protection. This means, the user gets high speed at the risc of undiscovered memory failures, which are, however, often unacceptable to so called mission critical applications.

- As stated above, the code contains a lot of integer blended parts. This enables the out-of-order RISC-core of the PentiumII [7] to hide floating-point latencies by executing independent integer code – for instance for address generation – simultaneously. It should be mentioned that the integer performance of the PentiumII is nearly as high as that of its competitors, but its floating point performance is much lower because Intel has integrated only one fp-unit – that suffers from its 32bit multiplier – compared with two 64bit fp-units available on the majority of high-end RISC-processors. In addition, Intel has organized the floating-point register file as stack, so compilers have to work hard in optimizing floating-point register usage.

In spite of some drawbacks of the Intel CPU, the benchmark results show that even a low cost solution can provide sufficient computing power for a wide range of performance demanding tasks. And, if it is really neccessary, one can also use more expensive ECC-RAMs to ensure data integrity for mission critical applications.

3 BYTEmark CPU-Test Release 2.0

BYTEmark [9] is a benchmark used to compare the raw CPU-performance of disparate machines. This benchmark is meant to expose the theoretical upper limit of the CPU, FPU, and memory architecture of a system. But the reader should keep in mind, that this benchmark is synthetic. This quite reasonable argument is based on the fact that people don't run benchmarks for a living, they run applications.

The compiler switches were tuned on each machine to get the best results. It should also be noted that this benchmark does not fully utilize the memory interface because most of its temporary datasets fit into the L1 or L2 cache respectively. The following table shows the results of the benchmark for all tested machines (high values indicate high speed):

Test	C240	Octane/MXE	Ultra60	PentiumII 350Mhz
Numeric Sort	4.18	3.91	3.95	4.09
String Sort	2.11	11.38	2.12	5.46
Bitfield	6.06	5.52	5.13	9.29
FP Emulation	8.72	8.01	4.81	4.58
Fourier	7.00	7.18	2.95	3.75
Assignment	10.36	7.81	6.18	5.52
Idea	6.49	7.28	3.96	5.89
Huffman	6.61	6.42	6.03	6.29
Neural Net	9.94	6.32	7.18	5.86
LU Decomposition	8.38	11.22	8.65	6.11
Integer Index	5.744	6.864	4.356	5.697
FP Index	8.354	7.983	5.629	5.119

Short description of the benchmark results

Numeric sort: Generic integer performance. Should exercise non-sequential performance of cache (or memory if cache is less than 8K). Moves 32-bit longs at a time, so 16-bit processors will be at a disadvantage.

String sort: Tests memory-move performance. Should exercise non-sequential performance of cache, with added burden that moves are byte-wide and can occur on odd address boundaries. May tax the performance of cell-based processors that must perform additional shift operations to deal with bytes.

Bitfield: Exercises "bit twiddling" performance. Travels through memory in a somewhat sequential fashion; different from sorts in that data is merely altered in place. If properly compiled, takes into account 64-bit processors, which should see a boost.

Emulated F.P.: Past experience has shown this test to be a good measurement of overall performance.

Fourier: Good measure of transcendental and trigonometric performance of FPU. Little array activity, so this test should not be dependent of cache or memory architecture.

Assignment: The test moves through large integer arrays in both row-wise and column-wise fashion. Cache/memory with good sequential performance should see a boost (memory is altered in place – no moving as in a sort operation). Processing is done in 32-bit chunks – no advantage given to 64-bit processors.

Huffman: A combination of byte operations, bit twiddling, and overall integer manipulation. Should be a good general measurement.

IDEA: Moves through data sequentally in 16-bit chunks. Should provide a good indication of raw speed.

Neural Net: Small-array floating-point test heavily dependent on the exponential function; less dependent on overall FPU performance. Small arrays, so cache/memory architecture should not come into play.

LU decomp.: A floating-point test that moves through arrays in both row-wise and column-wise fashion. Exercises only fundamental math operations (+, -, *, /).

4 Conclusions

HP's C240 seems to be the fastest machine regarding real-world applications, it combines highest graphics performance with a very fast CPU and a high memory bandwidth, hence the C240 may be the best choice for technical computing.

Beside of that the Ultra60 is an excellent choice for somebody who do not require texture-mapping. The Ultra60 comes with a very stable operating system and offers high CPU performance at a favourable price.

The Octane/MXE might be an optimal solution for those users that are mainly interested in virtual reality and multimedia applications. This is because the graphics board provides superior performance for textured-surfaces and includes powerful compression/decompression video-hardware indispensable for realtime video- and audio-processing. On the contrary, the machine is tuned for direct memory access, which allows for very high data transfer rates, also for image processing. Another essential advantage of the Octane/MXE is the availability of a wide range of high-quality graphics software, especially for virtual-reality applications, for instance IRIS-Performer and Inventor.

As we have seen also low-cost Intel-PCs provide enough processing power for many computationally demanding applications. The low price and their high performance makes them an interesting alternative to expensive high-end workstations. In our opinion, the vendors of high-end workstations will have to lower the price of their machines, if they want to stay competitive against low cost solutions in the future.

Workstations from SGI, like the Octane/MXE, are in general known to have a relatively bad price/performance ratio (see www.spec.org/gpc/opc.static/viewin%7E1.html for a more detailed comparison of SGI's workstations with those of other companies), hence these machines can not be recommended for many scientific applications, even if high graphics performance is required. Despite the fact that the Octane/MXE was the most expensive machine of our test candidates, it wasn't able to outperform its competitors in graphics performance.

References

[1] Silicon Graphics, "OpenGL Reference Manual", Addison-Wesley Publishing Company, ISBN-0-201-63276-4, 1992.

[2] Anne P. Scott, Kevin R. Brukhart, Ashok Kumar, Richard M. Blumberg, and Gregory L. Ranson, "Four-Way Superscalar PA-RISC Processors", Article 1, Hewlett-Packard Journal, August, 1997.

[3] Sun Microsystems, http://www.sun.com/desktop/products/Graphics/elite3darch.html

[4] Kwan-Liu Ma (Organizer), Frank Crow, Mark Grossman, Gordon Stoll, "Parallel Graphics and Visualization Techology", Course Notes c42 for SIGGRAPH 98.

[5] Intel, "Intel 440BX AGPset: 82443BX Host Bridge/Controller", Datasheet, Order Number:290633-001, April, 1998.

[6] Noel D. Scott, Daniel M. Olsen, Ethan W. Gannett, "An Overview of the VISUALIZE fx Graphics Accelerator Hardware", Hewlett-Packard Journal, May, 1998.

[7] Intel Architecture Optimization Manual, Order Number 242816-003, 1997.

[8] GMV General Mesh Viewer, User's Manual LAUR 95-2986, Los Alamos National Laboratory, 1998.

[9] BYTE Benchmarks, http://www.byte.com/bmark/bmark.htm