Computer System Performance Analysis and Benchmarking

Winter Semester 1997-1998

by Marko Aho
Christopher Vinckier

Part I

I/O Subsystem Performance Analysis

by Christopher Vinckier

Table of Contents

Introduction

When we consider the current evolution in the computing industry, there are a few aspects that immediately draw one's attention. Computing power is increasing at a tremendous rate. This has been nicely captured by Gordon Moore : 'Expect a doubling in processor power every 18 months.', an expression now referred to as Moore's Law. One can also easily observe that no other component follows this growth rate. Memory, RAM and storage, are becoming ever cheaper. However, what is not occurring is that access time and throughput are decreasing resp. increasing at exponential rates. Considering magnetic disk technology specifically we see that disk density has been improving by about 50% per year, almost quadrupling in three years. Access time has only improved by one-third in 10 years. Taking this into account, we have decided to concentrate our attention on the performance analysis of currently available magnetic disk storage. If current trends continue, magnetic disk storage is bound to become the bottleneck of most systems. Super fast processors and huge memories have to be 'fed' and a system is only as fast as it's slowest component, currently the disk.

In our analysis we shall consider the advantages and disadvantages of currently available technologies and their impact on system performance and effectiveness. To narrow down this still very broad subject even more, we shall focus our attention on standard Personal Computers. The reason being that 90% of computers used world wide are of this type. Along the way we shall try to demystify the difference between the currently used interfaces, EIDE (Enhanced Integrated Drive Electronics) and SCSI (Small Computer System Interface) as well as their cost-performance correlation.

We shall try to analyze the performance of different drives by means of several benchmarks. The reason for using multiple benchmark programs being that performance bottlenecks may be easier to detect by comparing the results from different benchmarks. Each benchmark tells part of the story, and together a more complete picture can be drawn.

The Art of Benchmarking

Without a doubt, benchmarking is one of the most controversial subjects in the computer industry. This is understandable, given the number of exaggerated performance claims that consumers are exposed to on a daily basis. As a result, the process of benchmarking is often viewed with doubt and suspicion.
So how can we make a distinction between questionable and valid benchmarks ? Benchmarks can be fair, but most of all useful, if the test are understood and a few simple strategies are followed. We shall examine common issues affecting disk benchmarks as well as discuss reasons why results may not always be what one expects.

Two basic types of benchmarks exist in the computer industry : those that run a real-world workload (ex. Winstone 97® and Winbench 97®), and those that simulate a workload (ex. ThreadMark). Each have their strength and weaknesses, and it often works well to run both types.
Real-world benchmarks may be as simple as timing how long it takes to copy a file, or they may involve executing complex scripts that perform functions in many different applications. The main advantage of these benchmarks is they measure performance using real applications. In doing so, they often exercise all the components in a system, so the effect that any one component has on overall system performance can be measured. However, this same advantage can also be a limitation, since the performance of one component may be artificially limited by another. For example, we may be interested in disk performance, but a slow video accelerator will slow down a benchmark that has a high proportion of video content. This in turn could affect the workload presented to the disk drives.
Simulated benchmarks usually target a specific system component, and are not as sensitive to the effects of other components. The function of simulated benchmarks is to measure how well the targeted component handles periods of high activity. Performance in periods of high activity is critical, since the user will feel his machine is sluggish during these periods. The better these ‘hot spots’ are handled, the more responsive the system will be.
Care should be taken in choosing simulated benchmarks, since some may be designed to present a workload that does not relate to the intended system application. Likewise, certain benchmarks may bypass critical components of the operating system or measure aspects of performance that don’t relate to a given user. The documentation included with the benchmark should give a detailed description of its operation. Otherwise, experimentation will be necessary to find out how the benchmark makes measurements.

Considering the myriad of benchmarks available on the market today, it is not uncommon to obtain entirely different or even conflicting results depending on which benchmark you run. That can make it very difficult to draw any concrete conclusions on performance. The solution to this problem consists of a few basic strategies.
The first is to understand what each benchmark you run measures, and how it makes those measurements. Armed with this knowledge, the discrepancies and anomalies that invariably arise can often be explained.
Take the case of the disk benchmark that repeatedly reads the content of one sector. The goal may be to measure disk performance, but if the drive or operating system uses cache, the result is a measurement of how fast the cache is instead of how fast the drive is. Other disk benchmarks may bypass he operating system cache all together, or a drive may have caching disabled. All of these situations may lead to unexpected results, but can be explained by how the benchmark operates and how the system is configured.
The next step is closely related to the first, and that is to obtain as many reference points as possible by using several different benchmarks and comparing results.
When running benchmarks, as with any scientific experiment, it is best to change only one variable at a time. In reality, this often proves to be difficult. For example, when comparing two host adapters, both host adapters and drive software will need to change. In situations like this it is important to document the differences so that those who examine the results can understand the factors involved.
Thorough documentation of the test environment and a detailed explanation of how the benchmarks were run will help avoid confusion and misinterpreted results.

Why Benchmark results may not be what we expect

Even if everything is done correctly, there will still be some odd results. Sometimes the reason is straightforward, while other times there might be no apparent reason. In this section we go into some detail as to why the disk drive benchmarks may turn out differently than originally expected.

- **Mechanical Latencies**

  The single largest contributor to low performance is the disk drive mechanical overhead. The mechanical overhead consists of drive head seeks and rotational latencies. Rotational latency is the time it takes for a sector on the disk media to rotate to the point under the read/write head. It is inversely proportional to the rotational speed of the disk : the faster the disk spins, the shorter the rotational latency. The mechanical overhead is significant because it is measured in milliseconds, while most of the other overheads are measured 3 orders of magnitude smaller.

- **Software and Firmware Overheads**

  Software overhead is the time it takes for commands to be passed through the operating system and software drivers. Firmware overhead relates to the time it takes for the host adapter and disk drive to process the commands. Both are substantially smaller than the mechanical overhead, and usually fall under 100 microseconds.

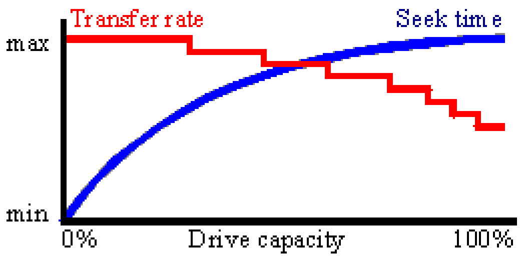
- **Data Transfer Rates**

  Compared to the mechanical overheads, the data transfer periods are also very short. For example, it only takes approximately 200 microseconds to transfer a typical 4 KB block of data at the 20 MB/s Ultra SCSI rate.
  To illustrate how the different overheads affect performance, consider this example. A good 7200 rpm hard drive will have an average seek time of 8 milliseconds, and a rotational latency of 4.2 milliseconds. The combined soft- and firmware overheads will only add 100 microseconds, and the data transfer time for 4 KB of data will be about 200 microseconds. All the factors together amount to 13.5 milliseconds of overhead. The soft- and firmware overheads together with the data transfer time only account for 2.4 % of the total overhead. From this it would appear that it doesn’t really matter how low the soft- and firmware overheads are, or how fast the data transfers are, since the mechanical overheads are much larger. In reality, these factors do matter, depending on what type of I/O occurs. With random I/O, where the data is not on contiguous areas of the disk, seeks and rotational latencies are encountered with almost every command. Thus, here the mechanical overheads become the most important component of disk performance, and data transfer rates are minor contributors. When considering sequential I/O however, the requested data is laid out contiguously. A consequence is that multiple commands can often be performed after the initial seek to the first sector of the requested data. Eliminating long seeks and rotational latencies increases the percentage of time spent transferring data, hence the data rates become more important.
  Data rates are also important in environments where there are multiple SCSI disk drives processing multiple commands simultaneously. In these situations, two or more drives may be attempting to transfer data across the bus at the same time. Since only one drive can transfer data across the SCSI bus at a time, the faster a drive can transfer its data and free the bus for other drives, the faster the overall data transfer rate will be.

- **Hard Disk Organization**

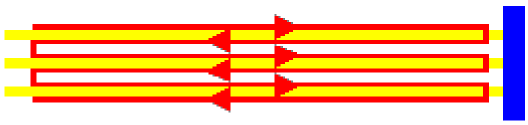
  In the pictures we can see the several ways how data can be stored physically on the harddisk. With a benchmark program that calculates the transfer rate or seek time of the whole harddisk we can see if the drive is using a 'vertical' or a 'horizontal' mapping. Depending on what kind of read/write heads and servo-motors (for positioning the actuator arm) are used it is faster to switch heads or to change tracks.

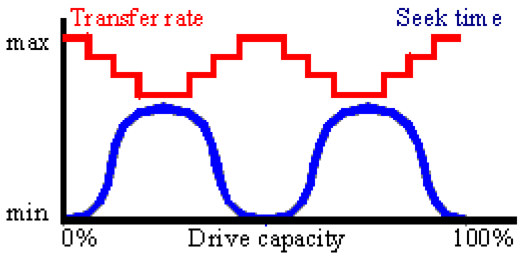
- **Vertical mapping**



Traditional harddisks orders their capacity in ‘vertical’ mapping. The data is read/written from one cylinder first, starting at the top track down to the bottom, before the heads are moved to the next cylinder.

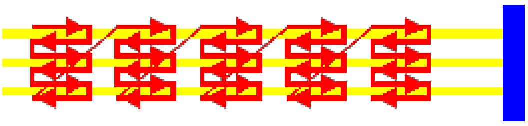
- **Horizontal mapping**

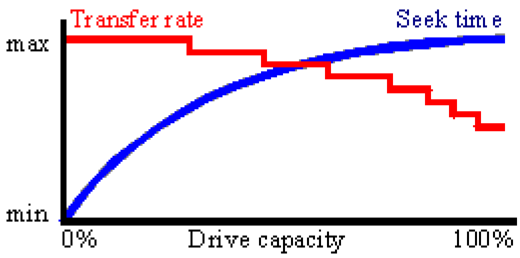




The ‘horizontal’ mapping. The data is read/written starting from the outer cylinder to the inner cylinder, before switching the heads to the next track.

- **Vertical and horizontal mapping**





Some harddisks use a combination of ‘vertical’ and ‘horizontal’ mapping.

As we can see in the above pictures, transfer rate is higher when data is read or written to the outer parts of a disk. The reason is that there is more space for sectors. The number of sectors varies in steps. Usually on a disk there are 10 to 20 zones (called 'notches') with a constant sector number. That is the reason why we see the steps in the transfer rates. Some harddisk use the combination of 'vertical' and 'horizontal' mapping. The 'horizontal' mapping is used inside the zones, the 'vertical' mapping between the zones. However, transfer rate and seek time look the same to 'vertical' mapping. Modern harddisks use different track sizes. The outer parts of a disk have more space for sectors than the inner parts. Usually, HD’s begin to write from the outside to the inside of a disk. Hence, data written or read at the beginning of a HD are accessed faster and transferred at a higher rate.

If we are going to buy a HD we have may need to know what kind of mapping the drive does. If we need constant transfer rates (for video, audio) we should get a drive which doesn't do the horizontal mapping. However, drives with horizontal mapping are not very common.
What this section has as goal is to show to the reader that the transfer rate and seek time can be influenced not only by the organization of the data on the hard disk but also by the presence of 'other' data on the disk. This means that a disk benchmark measuring the two aforementioned quantities will return different results depending on how much data is already present on the drive prior to running the benchmark. If one would push the performance measuring to the extreme, we would have to run the benchmark several times with the drive filled to different degrees. Having the results of the tests, it would suffice to take the arithmetic mean as a good approximation of the average disk performance. As we did not have access to virgin disks, we we're not able to perform these extensive measurements.

- **Other Factors**

Drive media rates, spindle speeds, buffer sizes, system bus speeds and bus chipsets are important items that can have a major effect on the overall data transfer rate. The higher the drive's media rate, the faster the data can be read into or written from the drive's buffer. The drive's media rate is closely tied to its spindle speed, since the faster the disk platters spin, the more data passes the drive heads in a given time. Today's high-end SCSI drives spin at rates up to 10000 rpm while EIDE drives seem to top out at 5400 rpm.
The size of a drive's cache buffer is dictated by a number of factors, including the drive's media rate, data transfer rate and the system bus speed. The function of the buffer is to allow data transfers to proceed smoothly and minimize costly mechanical latencies. For example, I/O and system bus speeds are generally faster than drive media rates. On write commands, the buffer serves to store incoming data from the host until the drive can write the data onto the media. If the cache buffer fills up, the host must suspend data transfers while the drive writes the data in the buffer to the disk. On read commands, the drive can read additional sectors of data into its buffer in anticipation that they will be requested later. If a command requests data that already exists in the drive's buffer, it can be sent directly, eliminating mechanical latencies and boosting performance. SCSI drives tend to have larger buffers than EIDE drives, since many SCSI drives can handle multiple commands at the same time and the SCSI spec higher data transfer rates.

The type of I/O request determines which of the above factors come into effect. As discussed above, with the sequential I/O a high proportion of time is spent transferring data and relatively little time is spent on mechanical overhead. Because of this, the drive's media rate is ultimately the determining factor in overall performance. The I/O bus speed is also an important factor if high performance drives are used. A large drive cache coupled with a good caching algorithm will allow the drive to stay on the bus longer, transferring data at the I/O bus burst rate. On the other hand, random I/O performance is dominated by drive mechanical latencies. In this case the bus transfer rate have little effect and seek times and rotational speeds are the most important specs to look at. The rotational speed is important because it determines the rotational latency.

From our discussion above, it becomes clear that just because a drive spec sheet states the drive has an x MB/s burst rate, it doesn't mean a benchmark will show the same rate. Benchmark results are a complex combination of hardware and software performance characteristics and I/O types. Often these components can interact I strange ways, producing numbers that don't seem to make sense. Sometimes, when there are no specific hardware or software limitations that account for odd results, they may be the result of timing situations that cause additional mechanical latencies. Bus analyzers could be used to show where the delays are occurring and to track down problems.

A final note on SCSI drives. Targeting the high-end drive market, SCSI drives have been designed to overcome mechanical overheads with two unique features. The first is called disconnect/reconnect. While one drive is busy seeking or reading data into its buffer, it can electrically disconnect from the bus. While it is disconnected, another drive can connect and receive commands or transfer data. The end result is better bus utilization and higher overall throughput because the time spent waiting for mechanical latencies can be put to use by other peripherals. The second feature is called tagged queuing. This feature allows a drive to accept and queue up multiple commands while processing an active command. The drive can then sort the commands and process them in the order that minimizes head seeks and rotational latencies. These added features to SCSI drives as well as the higher rotational speeds are just some of the reasons why SCSI drives outperform their EIDE counterparts and is naturally reflected in their higher cost.

The Benchmarks

In this section we shall briefly discuss the benchmarks we used for doing the performance analysis. In addition to analyzing just what these benchmarks do, we shall also mention how they calculate their respective scores. This is done to create a clear and concise picture of the whole process of the performance analysis. Two out of the three benchmarks used were provided to us courtesy of Mr. Kurmann. They are Winstone 97® and Winbench 97®. The third benchmark has been downloaded from the Web, is called Threadmark and was developed by Adaptec. After having looked at Threadmark's technical specifications, we decided to use it. The reason for this being that we could check if the results of benchmarks from different vendors are consistent with each other.

- **Winstone 97**®

Winstone 97® is a system-level, application-based benchmark that measure's a PC's overall performance when running today's top-selling Windows-based applications. The Business WinStone Benchmark Suite contains the following applications :

- Word-processing/Spreadsheet
  - Excel 7.0
  - Word 7.0
  - Word Pro 7.0
- Database
  - Paradox 7.0
  - Access 7.0
- Businessgraphics/ Desktop Publishing
  - PageMaker 6.0
  - CorelDRAW! 6.0
  - PowerPoint 7.0

The HighEnd WinStone Benchmark Suite contains the following applications :

- Imageprocessing
  - Photoshop 3.0.5
  - Picture Publisher 6.0
- CAD / 3-D
  - Microstation 95
  - PV-WAVE 6.0
  - AVS 3.0
- Application Development
  - Visual C++ 4.1

Winstone 97® uses scripts to execute commands within an application and stores the time it took the PC to execute them. After all the applications have run, Winstone 97® uses each application's elapsed run time to calculate a score. They way this score is calculated is as follows :

1. *Winstone 97® computes the application's execution speed on the test PC relative to the application's actual execution speed on the base machine.* This is done to remove an undue influence an application may have the script for that application took longer to execute than the scripts of the another applications.

The basemachine used for the normalization process in Winstone 97® as well as for Winbench 97® has the following configuration :

	Gateway 2000 4Dx2-66V
Processor	Intel 486DX2, 66MHz
Operating System	Windows 95, Enhanced Mode with Paging
Memory	16 MB
Level-1 Cache	8 KB
Level-2 Cache	64 KB
Virtual memory	64 to 96 MB
Graphics- accelerator	ATI Graphix Ultra Pro, 1024x768x16, 1MB VRAM
IDE-controller	On Board
Hard Disk	Western Digital Caviar 2250

2. *Winstone 97® combines the relative speed for applications in a category suite.* This is done with use of a weighted harmonic mean, producing a relative speed for the entire category. The weight assigned to each applications is its unit market share in the category and is kept secret. To compute the weighted harmonic mean, Winstone 97® divides the weight of each application by the application's relative speed to produce a workload execution time for the application. The benchmark then uses a weighted harmonic mean to combine those application numbers to get a number for each category.
3. *The same formula is used to compute an overall score, but the final number is multiplied by 10.*

Interpreting the results is straightforward : Higher numbers mean better performance.

- **Winbench 97**®

Winbench 97® is a subsystem-level benchmark that measures the performance of the following subsystems in a Windows-based environment:

- Graphics
- Disk
- Processor
- CD-ROM
- Video
- DirectDraw

Of these, only the first two shall be used. Concerning the disk benchmarks, Winbench 97® returns the following results that provide an overview of the disk performance:

- High End Disk Winmark 97
- Business Disk Winmark 97
- Disk Winmark 97

The synthetic disk tests consist of the following accesses to a 90 MB file:

- Disk/Read, CPU Utilization (%)
- Disk/Read, Sequential 200, 512, 2048, 4096 Bytes blocksize
- Disk/Read, Random 200, 512, 2048, 4096 Bytes blocksize
- Disk/Write, CPU Utilization (%)
- Disk/Write, Sequential 200, 512, 2048, 4096 Bytes block size
- Disk/Write, Random 200, 512, 2048, 4096 Bytes block size

The three elements : mode of operation (read/write), the type of access (sequential/random) and the different block sizes, make a Full Factorial Design out of these synthetic tests. One can easily calculate that the total number of combinations of the three factors is 16 (2 x 2 x 4) giving us some more insight into the performance characteristics of the system. For these tests, bigger numbers mean better performance.

The Disk Winmark 97 tests can also produce CPU utilization results, a measure of the percentage of the processor that disk usage consumes. For this test, the lower the result the better, because more processor time is then available for other activities. The Disk Winmark 97 test executes the same series of disk operations Winstone 97®'s business and high-end applications perform. The results are computed in the following manner :

1. During the Disk WinMark test, Winbench 97® plays back the logs we recorded for each Winstone 97® Business or High-End application.
2. The benchmark times how long the test machine plays back each application log.
3. Winbench 97® calculates the transfer rate (how many thousands, not 1024 but 1000, of bytes per second the system read or wrote to the disk) for each application.
4. Winbench 97® uses a weighted harmonic mean of transfer rates (using the Winstone 97® application weights) to produce the Disk WinMark 97 result. The result is the number of thousands of bytes (1000, not 1024) per second transferred.

Note that a transfer rate test actually stresses the whole disk subsystem, which includes:

- the disk drive
- the pathway between the hard disk and the controller
- the pathway between the disk controller and the main memory
- the disk device drivers
- the interaction of the disk subsystem with the OS

Because the test reproduces the kind of disk activities applications carry out, we can use the results as a guide to the kind of throughput we can expect to see when working with the PC's disk subsystem.

• Threadmark

ThreadMark is a standard 32-bit application that measures two key disk performance indicators : data transfer rate in MB/s and percent CPU utilization. It makes a comprehensive series of measurements using a mixture of single and multithreaded requests across a range of request block sizes. The measurements results in 128 individual data points, which are weighted according to a profile that represents typical I/O behaviour in a Windows dektop system. After the measurements are weighted, they are averaged together to arrive to a single-number data transfer rate and CPU utilization scires. This is done in the following way. When ThreadMark is executed for the first time, a CPU baseline number is established for use in calculating the amount of CPU utilization required to perform the I/O tasks. The CPU baseline is established by calculating prime numbers while there is no other system activity. The number of calculations completed over a pre-defined period of time is recorded, with the result representing 100% CPU utilization.

Once the baseline is established the benchmark will begin taking I/O measurements. During this time, the benchmark will continue to calculate prime numbers in the background using a thread with idle priority. Because the thread has idle priority, it will only be active if there is no other system activity, and it will not affect I/O measurements. When the measurements are finished, the number of completed prime number calculations is scaled to match the time frame of the original CPU baseline. The scaled number is then compared with the original according to the formula:

CPU utilization = 100 - scaled number of calculations during I/O measurements x 100 baseline number

Test files for read measurements are created before taking each measurement so the files are not accessed out of operating system cache. Each file is twice the size of the system RAM to further minimize caching effects. Test files are created one at a time to minimize fragmentation. Once the test files are created, the benchmark starts taking I/O measurements. Each file is accessed using a single thread. Threads are permitted to begin I/O as soon as they are created, and all thread management is handled through the operating system scheduler. The measurements stop at the time when any thread reaches the end of its file. The benchmark calculates the data transfer rate by measuring the number of bytes transferred and the transfer time. It also performs the calculations used to determine CPU utilization. The results are stored in system memory to avoid influencing disk performance.

ThreadMark runs a series of measurements using one to four threads. Both read and write requests are generated using 512 Bytes, 1 KB, 2 KB, 4 KB, 8, KB, 16 KB, 32 KB and 64 KB block sizes. The result is 64 individual test passes, with each pass recording the data transfer rate and the percent of CPU utilization. When all measurements are complete, the results are weighted using the profile below.

Overall	Thread Type	Block Size							
%	% Read or Write	512	1K	2K	4K	8K	16K	32K	64K
50 %	1 Thread Reads - 80 %	10%	2%	2%	40%	2%	4%	10%	30%
	1 Thread Writes - 20 %	10%	5%	5%	60%	2%	4%	4%	10%
35 %	2 Thread Reads - 80 %	10%	2%	2%	40%	2%	4%	10%	30%
	2 Thread Writes - 20 %	10%	5%	5%	60%	2%	4%	4%	10%
10 %	3 Thread Reads - 80 %	10%	2%	2%	40%	2%	4%	10%	30%
	3 Thread Writes - 20 %	10%	5%	5%	60%	2%	4%	4%	10%
5 %	4 Thread Reads - 80 %	10%	2%	2%	40%	2%	4%	10%	30%
	4 Thread Writes - 20 %	10%	5%	5%	60%	2%	4%	4%	10%

Each transfer rate and CPU measurement is weighted according to the block size and the results are summed according to thread type. The eight resulting numbers are then multiplied by the percentage in the ‘Thread Type’ column and the read and write results for each thread count are added together. The resulting four numbers are then multiplied by the percentages in the ‘Overall’ column and added together to arrive at the single-number result.

The Systems

A specification of the machines we have used for our performance analysis can be found below. The first of these machines, a Dell computer, can be found in the lab for Computer Systems (RZ J-Stock, ETH). The other machines tested have no brandname as they have been assembled by Christopher. The latter of the two is still in his possesion.

- **Dell System OptiPlex GXpro 200USD Enhanced**

	Dell System OptiPlex GXpro 200USD Enhanced
Processor	Intel Pentium Pro 200 MHz, Step 9 Features fbffh
Operating System	Windows NT 4.0 Service Pack 3, Build 1381, Uniprocessor Free
Memory	64 MB EDO 60ns
Level-1 Cache	16 KB
Level-2 Cache	256 KB
Virtual memory	64 to 96 MB
Graphics-accelerator	Matrox Power Graphics Accelerator, MGA-2064W B4 R2, DAC: TI TVP3026 (220MHz), 4 MB DRAM
EIDE-Kontroller	On Board, bis PIO-Mode 4, inkl. DMA
SCSI-Kontroller	Adaptec AHA-2940 Ultra Wide SCSI, Bios V. 1.03

The disks that were benchmarked using this machine are the following :

	Cheetah	Hawk 4	Caviar
Capacity	4.55 GB	4.294 GB	3.166 GB
Rounds Per Minute (RPM)	10033 RPM	5411 RPM	5200 RPM
Cache	512 KB	512 KB	128 KB
Avg. Access Time (rd / wr)	7.7 / 8.7 ms	10.4 / 11.4 ms	12 ms
Interface	Ultra Wide SCSI	SCSI-2 Fast	EIDE
Filesystem	NTFS	NTFS	FAT
Partition-size	2 GB	2 GB	1 GB

- **Intel Pentium 200 MMX**

	Intel Pentium 200 MMX
Processor	Intel Pentium MMX 200 MHz, Step 9 Features fbffh
Operating System	Windows NT 4.0 Service Pack 3, Build 1381, Uniprocessor Free
Memory	64 MB EDO 60ns
Level-1 Cache	16 KB
Level-2 Cache	512 KB
Virtual memory	75 to 120 MB
Graphics-accelerator	Matrox Millenium II, MGA-2064W B4 R2, DAC: TI TVP3026 (220MHz), 4 MB WRAM
EIDE-Kontroller	On Board, up to PIO-Mode 4, inkl. DMA and Busmastering
SCSI-Kontroller	None

	Quantum Fireball ST 3220
Capacity	3.22 GB
Rounds Per Minute (RPM)	5400 RPM
Cache	128 KB
Avg. Access Time (rd / wr)	11 ms
Interface	Ultra ATA
Filesystem	NTFS
Partition-size	1 GB

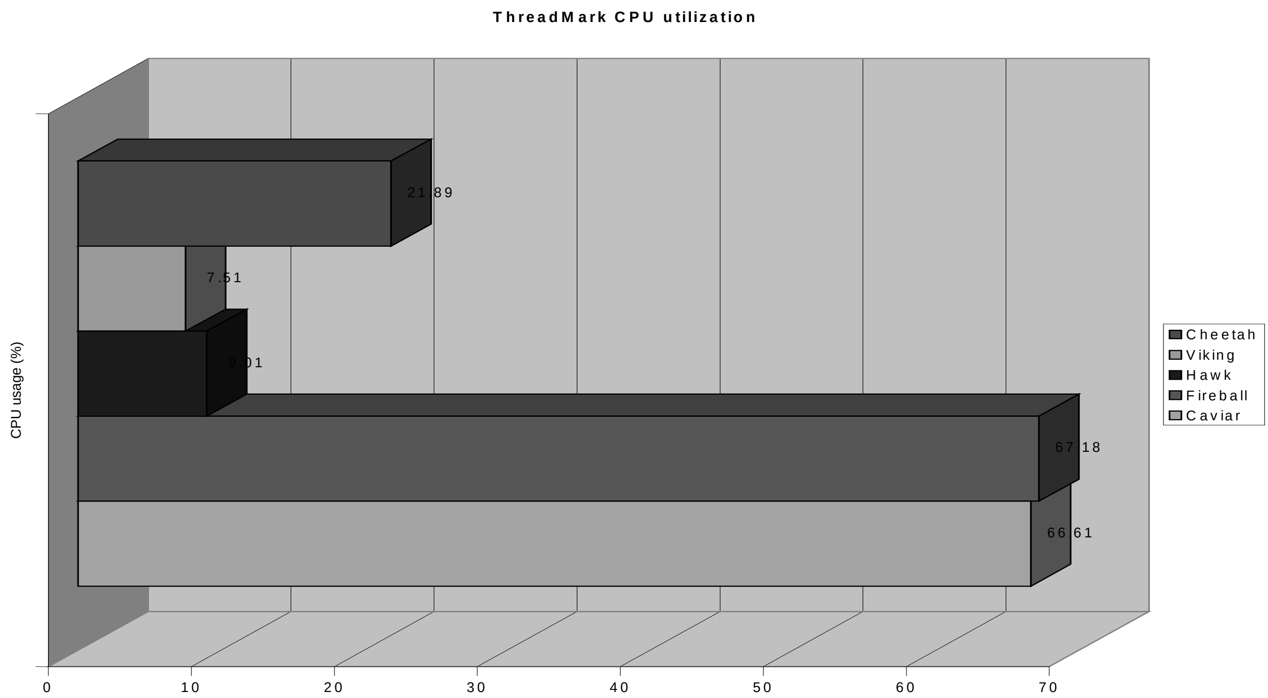
- **Intel Klamath (P6) 266**

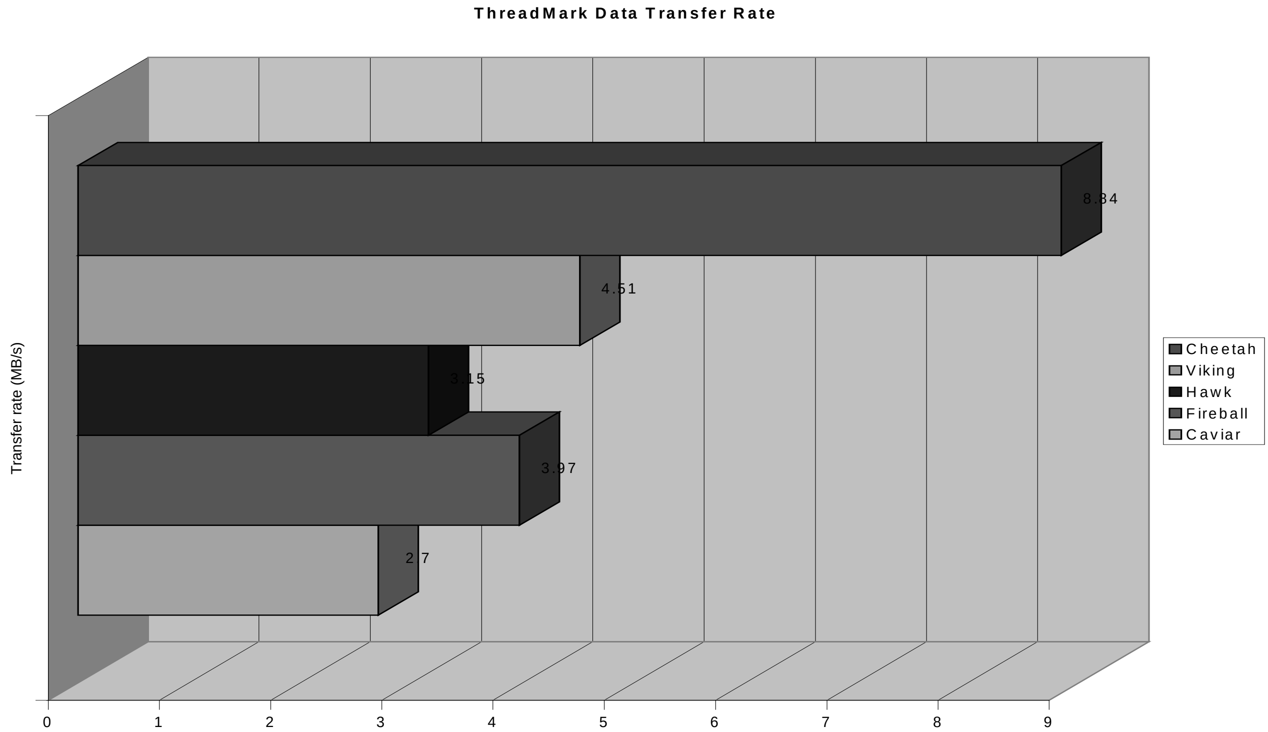
	Intel Klamath (P6) 266
Processor	Intel Pentium(R) II; Model 3 with MMX(tm), Technology Step 3 Features 80f9ffh
Operating System	Windows NT 4.0 Service Pack 3, Build 1381, Uniprocessor Free
Memory	64 MB SDRAM 10ns
Level-1 Cache	16 KB
Level-2 Cache	512 KB
Virtual memory	75 to120 MB
Graphics-accelerator	Asus V3000 AGP, SGS-THOMSON RIVA 128™ 128-bit 2D/3D graphics engine, 4MB 128-bit 100MHz SGRAM frame buffer
EIDE-Kontroller	On Board, up to PIO-Mode 4, inkl. DMA and Busmastering
SCSI-Kontroller	Adaptec AIC-7880 UltraSCSI, Bios V. 1.03

	Quantum Viking 4.5
Capacity	4.55 GB
Rounds Per Minute (RPM)	7200 RPM
Cache	512 KB
Avg. Access Time (rd)	8 ms
Interface	Ultra Wide SCSI
Filesystem	NTFS
Partition-size	1.5 GB

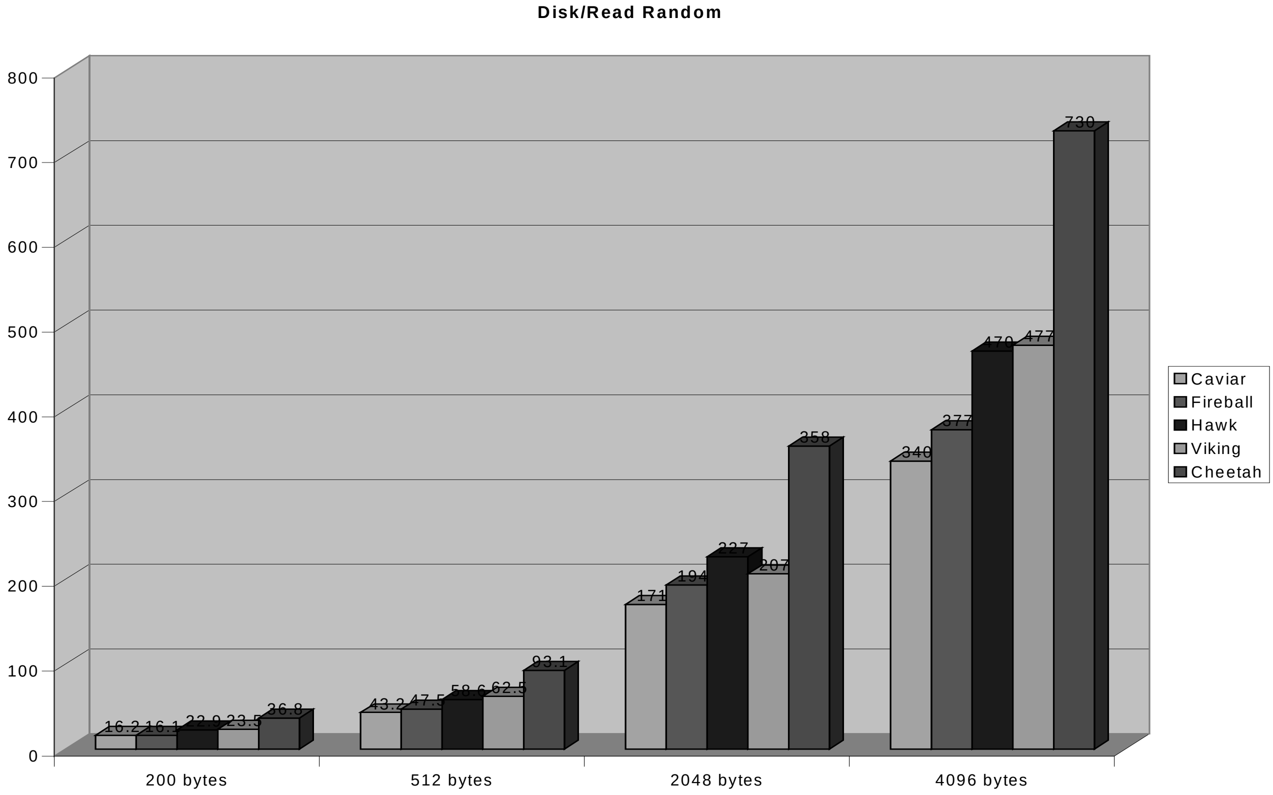
The Benchmark Results

- **ThreadMark**

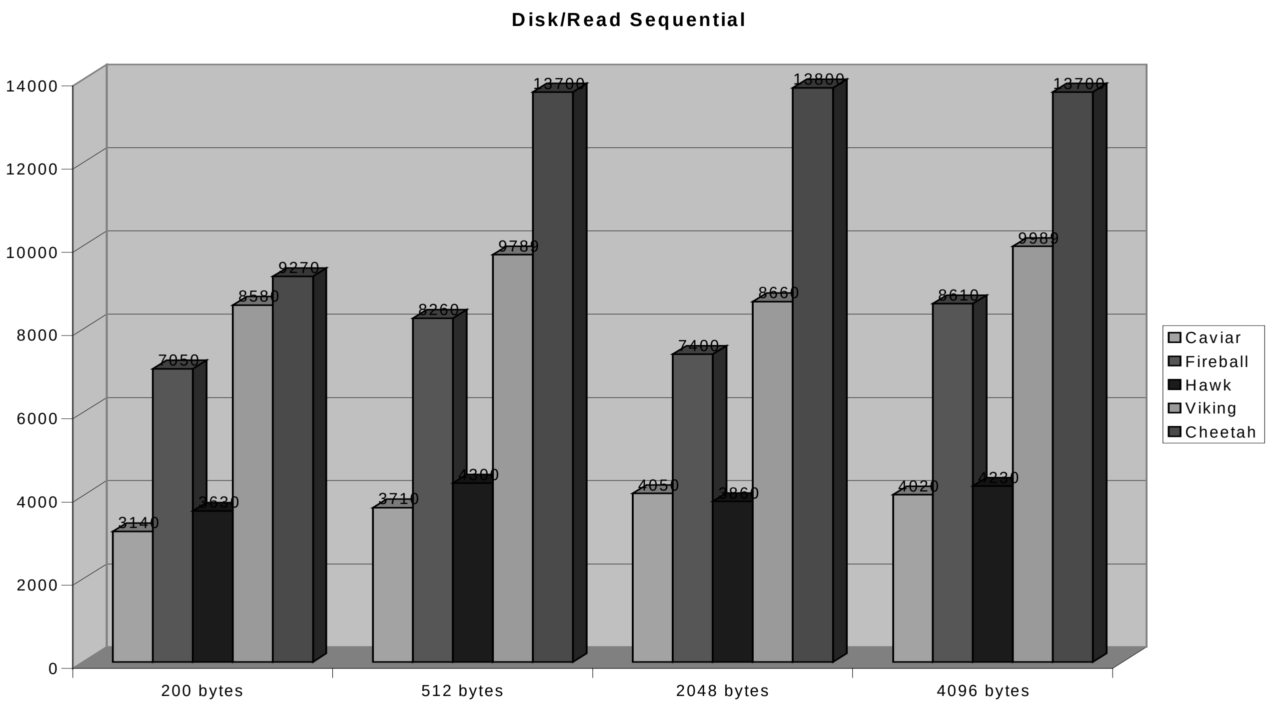




- **Winbench 97®**

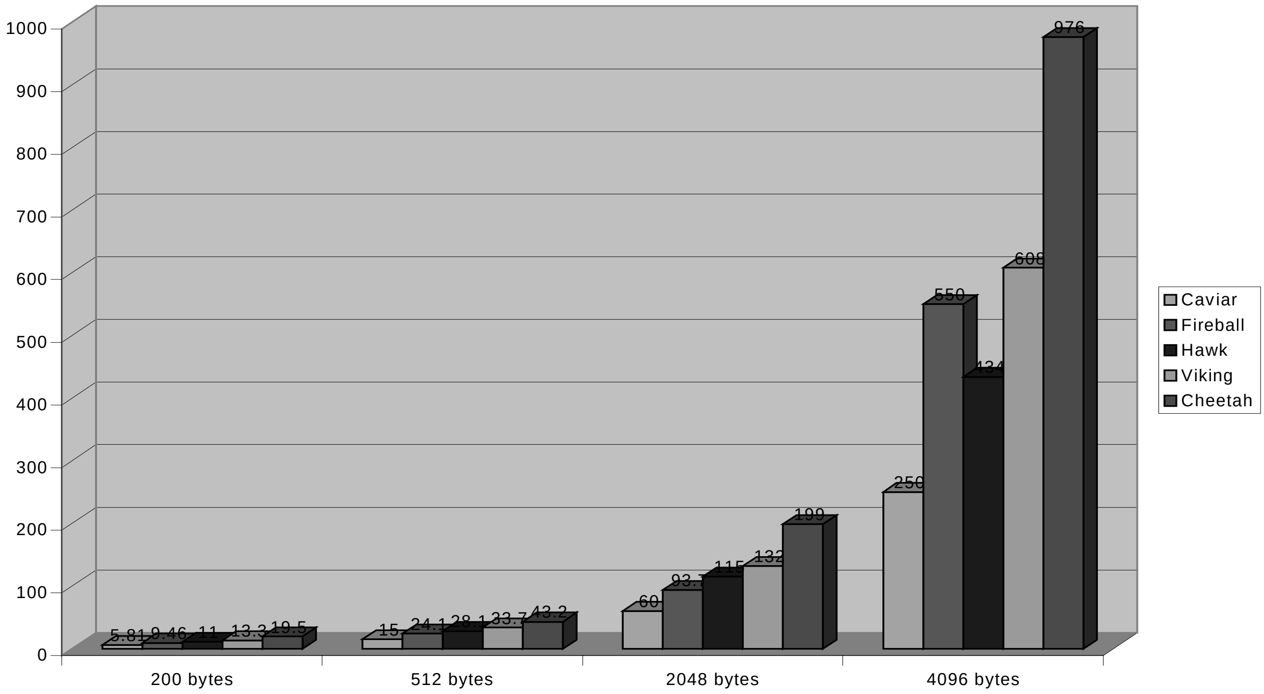


Disk/Read Random
800
700
600
500
400
300
200
100
0
16.2
16.1
22.9
23.5
36.8
43.2
47.5
58.6
62.5
93.1
171
194
227
207
358
340
377
470
477
730
200 bytes
512 bytes
2048 bytes
4096 bytes
Caviar
Fireball
Hawk
Viking
Cheetah

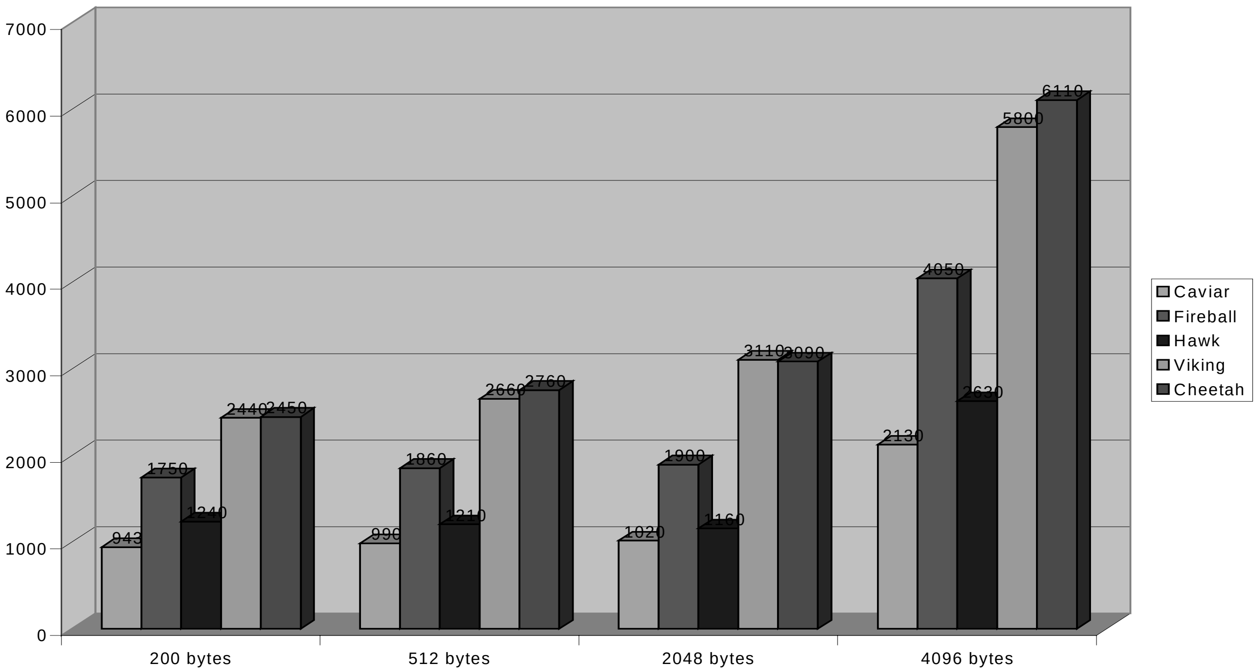


Disk/Read Sequential
14000
12000
10000
8000
6000
4000
2000
0
3140
7050
3630
8580
9270
3710
8260
4300
9789
13700
4050
7400
3860
8660
13800
4020
8610
4230
9989
13700
200 bytes
512 bytes
2048 bytes
4096 bytes
Caviar
Fireball
Hawk
Viking
Cheetah

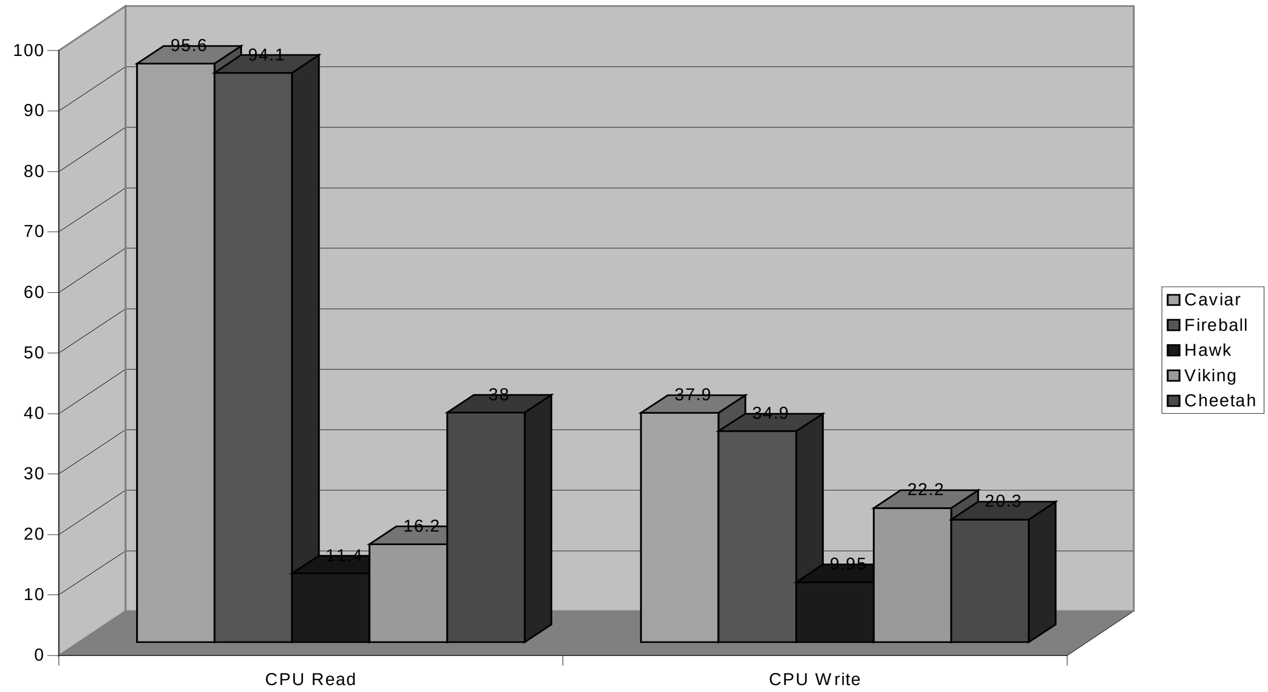
Disk/Write Random



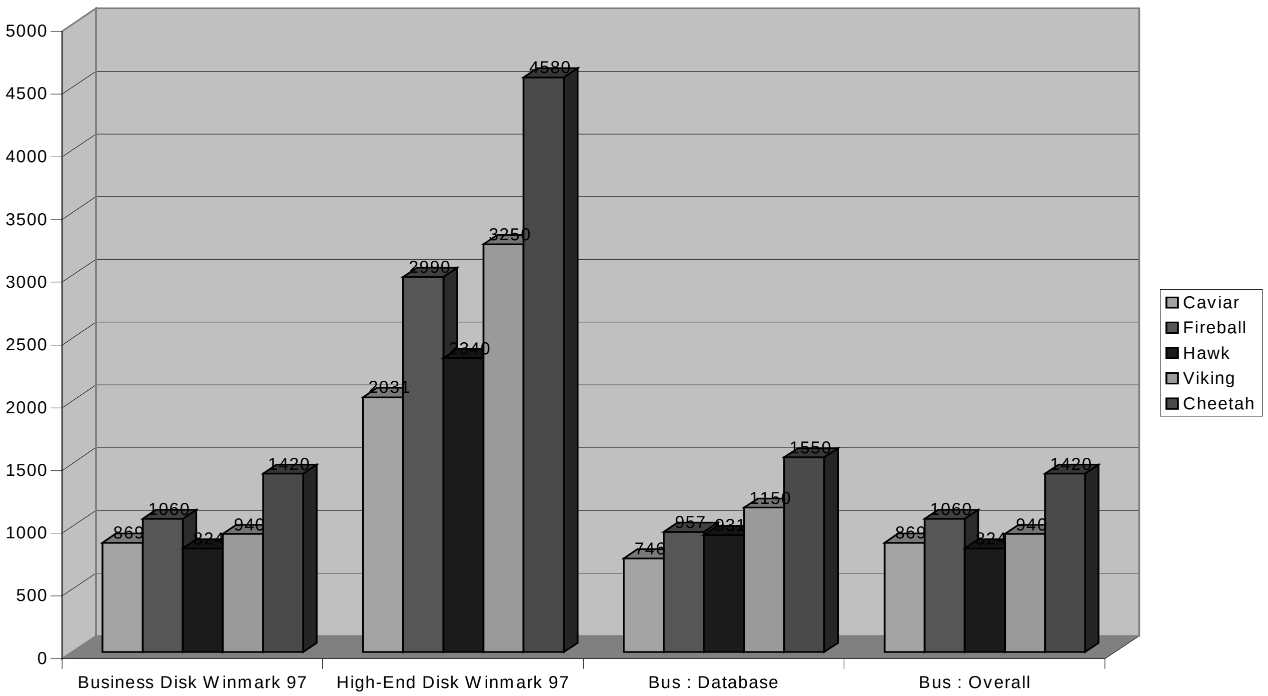
Disk/Write Sequential

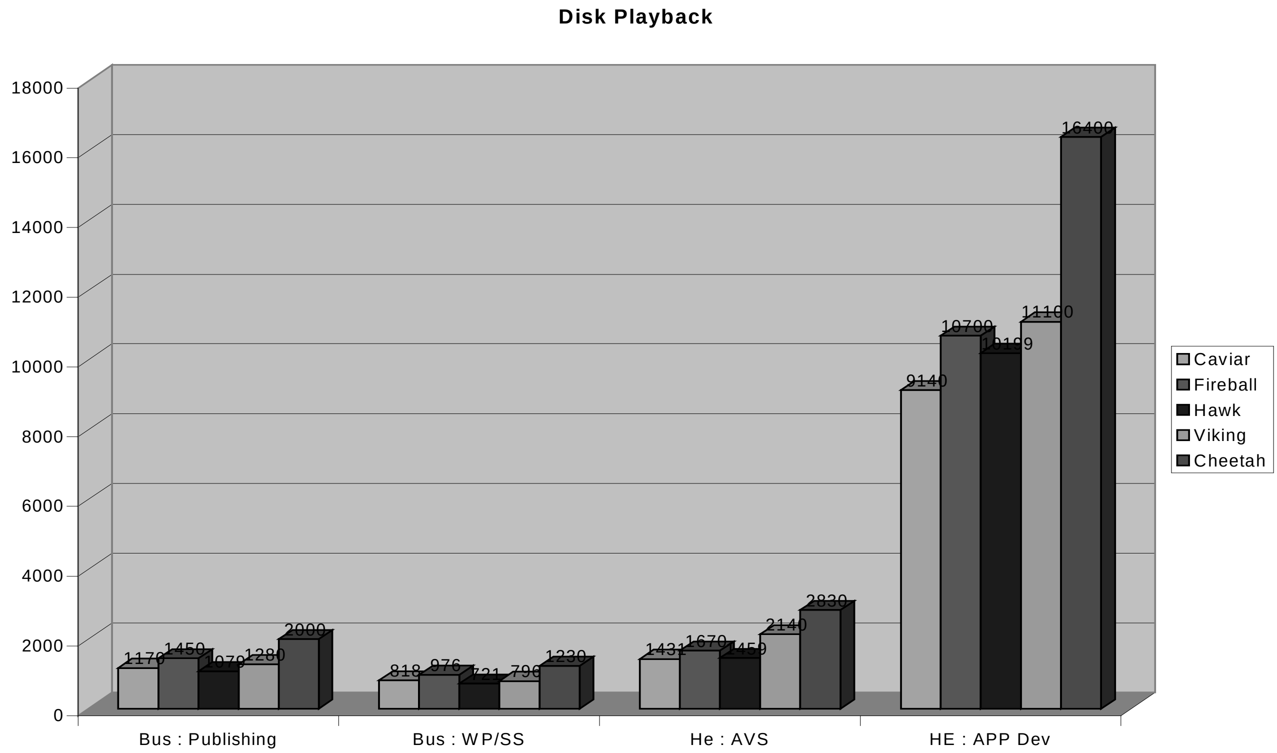


CPU Usage

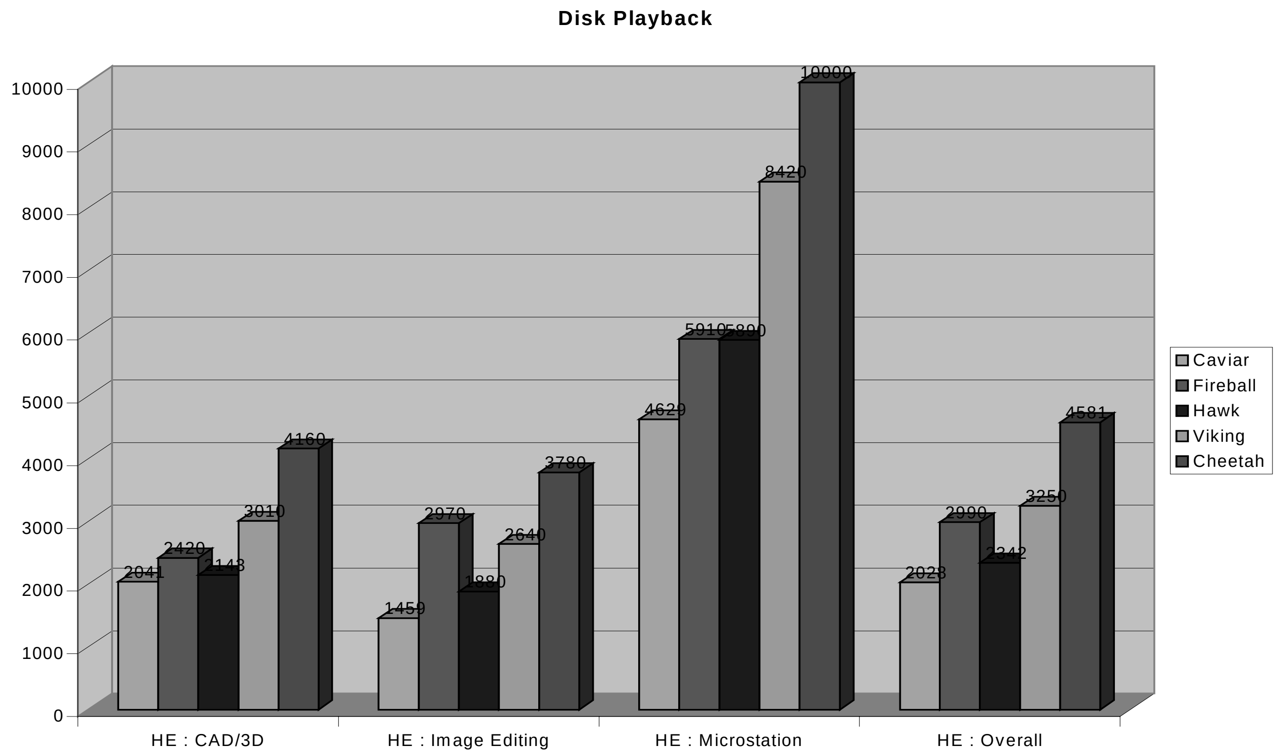


Disk Playback

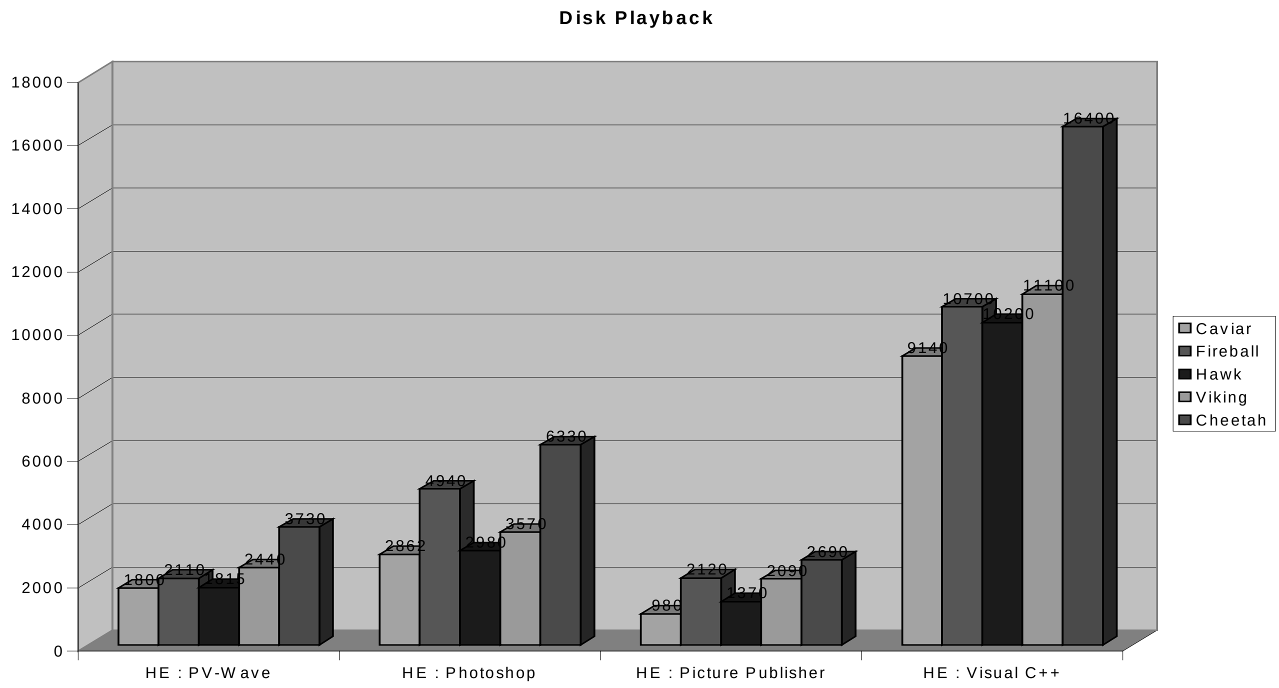




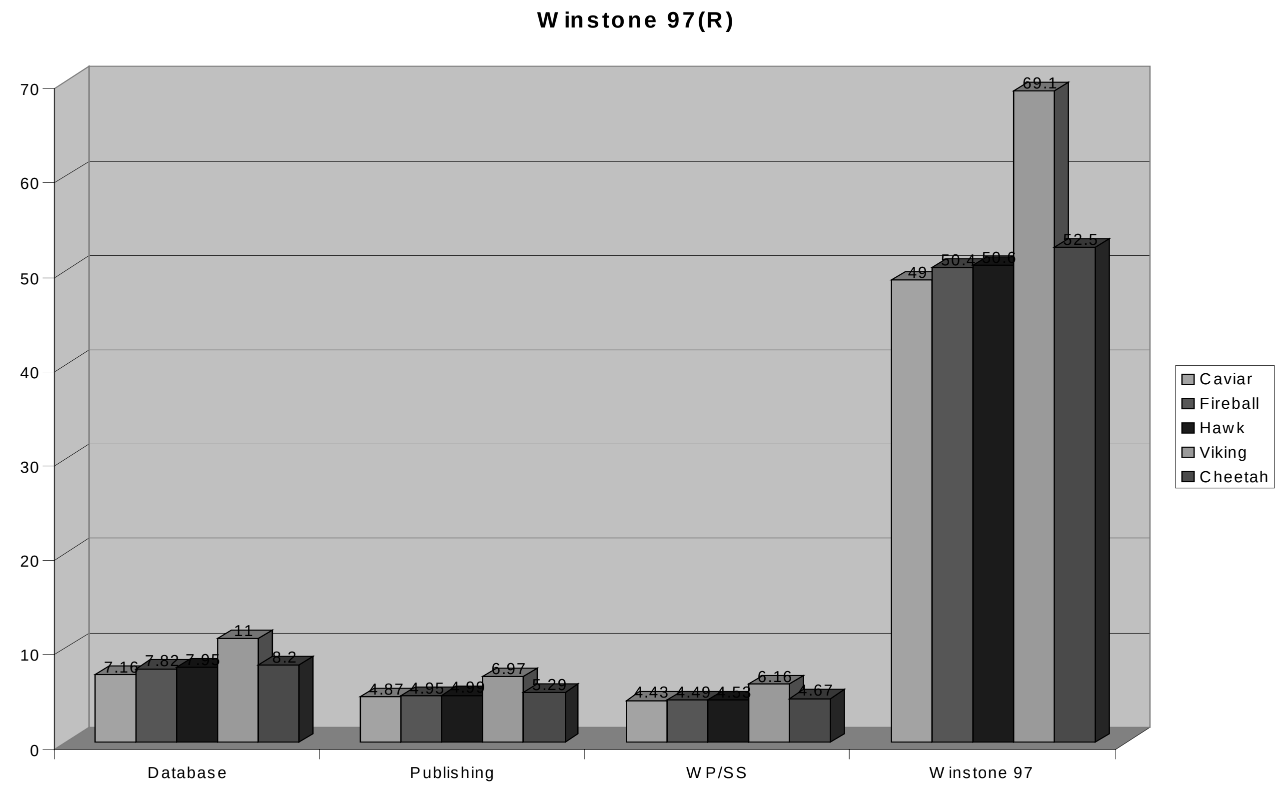
Disk Playback
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
1170
1450
1070
1280
2000
818
976
721
799
1230
1431
1670
1450
2140
2830
9140
10700
10199
11100
16400
Bus : Publishing
Bus : WP/SS
He : AVS
HE : APP Dev
Caviar
Fireball
Hawk
Viking
Cheetah



Disk Playback
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
2041
2420
2143
3010
4160
1459
2970
1880
2640
3780
4620
5910
5890
8420
10000
2028
2990
2342
3250
4581
HE : CAD/3D
HE : Image Editing
HE : Microstation
HE : Overall
Caviar
Fireball
Hawk
Viking
Cheetah


Disk Playback
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
1800
2110
1815
2440
3730
2862
4940
2980
3570
6330
980
2120
1370
2090
2690
9140
10700
10200
11100
16400
HE : PV-Wave
HE : Photoshop
HE : Picture Publisher
HE : Visual C++
Caviar
Fireball
Hawk
Viking
Cheetah

- **Winstone 97®**


Winstone 97(R)
70
60
50
40
30
20
10
0
7.16
7.82
7.95
11
8.2
4.87
4.95
4.95
6.97
5.29
4.43
4.49
4.53
6.16
4.67
49
50.4
50.6
69.1
52.5
Database
Publishing
WP/SS
Winstone 97
Caviar
Fireball
Hawk
Viking
Cheetah

After having compared the ThreadMark results to the ones returned by Winbench®, we can draw the conclusion that EIDE-drives will become the bottlenecks in multitasking, multi-user or server systems. Both benchmarks return similar results. The EIDE-I/O consumes more than 68% of processor resources.
A detail that should be mentioned is the impact of processor speed when using EIDE–drives. Taking into account the amount of CPU resources that these drives consume, it is clear that a faster processor will be able to handle I/O traffic more efficiently than a slower one. On the machines using EIDE-drives that we tested, the processor performance difference is not very big. However, if we were to test the same drive on for example a Pentium II 266 and a Pentium Classic 133, we would expect to see a difference in transfer rate although we are testing exactly the same drive. How big this difference would be is unclear. Concerning SCSI drives, CPU speed should not have a very great impact because the CPU load of the I/O operations remains under 10%. This is one of the characteristics that make SCSI-drives great for disk-intensive environments such as multi-tasking systems or disk intensive applications such as image and audio editing. The read or write operation is submitted and the SCSI controller handles most of the overhead making the CPU available for other operations. The data transfer rates returned by ThreadMark are difficult to compare to the tranfer rates returned by Winbench as both programs use totally different blocksizes. What’s more, ThreadMark returns only a single value having been calculated in the way mentioned above. The global trend is however comparable.

The Winbench 97® results give us some more insight into the behaviours of the different drives when reading and writing varying amounts of data. What we are forced to mention is that contrary to Threadmark, Winbench 97® uses rather small blocksizes for its tests. Knowing that Windows NT™ operates on blocksizes of 4096 bytes, we can expect some considerable caching effects when performing sequential reads on smaller data units. At this point, we are not evaluating drive performance anymore but rather cache performance. Personally, we find the approach of using much larger data units, as implemented by ThreadMark, a better approach since it tries to minimize the influence of caching. What can be deduced from the Winbench 97® results is that for small data units (200 bytes), the data transfer rates we measure are extremely low. The mechanical overhead when performing random reads and writes becomes dominant. Looking at the sequential operations for the same block size, we find data transfer rates that are up to 300 times higher. This can probably be attributed to caching. As the block sizes increase, the influence of mechanical overhead decreases and we notice a corresponding increase in data transfer rate. When looking at block sizes of 4096, NT’s operating mode, the difference between random and sequential writes falls back to a mere factor of 6. Another striking feature is the near lacking in difference between sequential read performance of blocks of size 512 KB, 2048 KB and 4096 KB. Once again the probable reason being caching of the data.
There are also however some results that we don’t fully understand such as the Quantum Fireball scoring a higher Business Disk Winmark than the Quantum Viking, the more performing drive when we look at both their specifications. Technicaly and theoretically this should not be such. If it it due to the testing being done on two very different machines or if it simply due to measurement errors remains a mistery. More extensive testing would be required to solve this dilemma.

The Winstone 97® results are somewhat peculiar. We see that drive performance does not have a great impact on global system performance. We do not get a factor two speed up by using a super fast Ultra Wide SCSI drive in stead of a EIDE-drive. The slightly higher performance of the Dell machine using SCSI drives can be explained by the aforementioned reduced CPU load of the I/O operations. The difference between the different SCSI-drive configurations can in its turn be explained by the higher troughput that the Cheetah drive has, thus pumping the data into memory at a higher rate. The Winstone 97® result for the Pentium II 266 machine has been added to show that a slightly lower performing drive, i.e. the Viking compared to Cheetah, can be more than compensated by building an overall well balanced system. It is very hard to determine the exact reasons why the Klamath configuration shows such an advantage over the Pentium Pro machine, but this is certainlz not because of the CPU. Other factors that are very likely to play a role are different motherboard chipsets, different SCSI-controller chips and drivers (AHA-2940 UW vs. AIC-7880), different types of RAM (EDO DRAM vs. SDRAM), different graphics subsystems, etc.

Conclusion

As mentioned in the introduction, the goal of this paper was to compare the performance of several disk drives using benchmarks. Although we have focused our attention on subsystem level performance of the different drives, we have included the results for the Winstone 97® benchmark suite. This has been done to show the impact of the disk drives performance on global system performance. The reader should however take care when reading the Winstone 97® results. These results only indicate the influence of the drive if the benchmark has been run on the same or identical systems, except for the disk drive. This is due to the fact that system-level performance is determined by many more factors than merely the disk drive. Processor type, amount of memory, operating system, system bus chipset, graphics accelerator and many others contribute to the systems performance. Knowing this, the reader can use the Winstone 97® results for the Dell machine to compare disk drive performance impact on global performance.

Taking into account the arguments put forward in the previous chapters, it becomes clear that measuring real drive-performance is not a trivial task just consisting of installing a program, running it and publishing the results. In this paper we have tried to tackle the problem of comparing different drives and drive-types in an intelligent and organized fashion, pointing out some of the factors one should consider… and there are a lot.

Part II

Graphics Subsystem Performance Analysis

by Marko Aho

Part II

Graphics Subsystem Performance Analysis

by Marko Aho

Table of Contents

I. Introduction

The monitor is the primary output device of a modern day computer system. In the last decade the processing power has increased rapidly and ever increasing resolutions and color depths have become common on the mainstream PC or workstation. This imposes growing demands on the performance and throughput of the graphics subsystem. As the quality of visual output is important for both professional and home users, this sector of technology has grown into a major industry with considerable research and development.

The components that build up the graphics subsystem, or have a direct impact on it, are:

- the internal bus
- graphics adapter
- graphics adapter driver
- the monitor

This paper will discuss the comparison of four conventional graphics subsystems and give an indication of their effect on the overall system performance. Specialized 3D capabilities are not benchmarked. A modern graphics subsystem based on the Advanced Graphics Port has been included as reference point to the newest technologies used on the desktop market.

II. Tested subsystems and test configurations

The following table summarizes the properties of the tested graphics subsystems:

Adapter	Graphics chip	Driver / Date	Memory	RAMDAC	Bus
Matrox Millenium	MGA-2064W	mga_mil.sys, 14.10.1996, 2.20.1371.01 mga.dll, 30.4.1997, 4.00.1381.04	4 MB WRAM	220 Mhz	PCI
STB PowerGraph	S3 Trio64	s3.sys, 30.4.1997, 4.00.1381.04 s3.dll, 30.4.1997 4.00.1381.04	2 MB DRAM	N/A	PCI
S3 ViRGE/DX	S3 3D	s3virge.sys, 21.11.1996, 3.51.1057.01 S3Virge.dll, 2.11.1996	2 MB EDO DRAM	170Mhz	PCI
ASUS V-3000 AGP	Riva 128	nv_disp.dll, 05.9.1997, 4.00.1381.01	4 MB VRAM	230 Mhz	AGP

The following machines were used to perform the tests:

- 2 Dell OptiPlex GXpro / Pentium Pro 200MHz (Matrox Millenium I, STB PowerGraph)
- AMD 200MMX (S3 ViRGE/DX)

Exact configuration descriptions of all systems can be found from the tables on page 14 of Part I. To ensure comparability of the graphics subsystem results, all machines were equipped with:

- 64 MB of RAM
- a 17 inch monitor
- Windows NT 4.0
- PCI bus running at 33MHz

For the reference AGP system, the following configuration was used :

- Pentium II 266MHz AGP (ASUS V-3000 AGP)

III. Discussion of parameters

Numerous characteristics of the graphic subsystem contribute towards the perceived speed and versatility of use. The following section explains the effects of the various factors and technical implementations.

Technical factors:

- *Graphics chip* — Several manufacturers develop special chip sets to accelerate graphics rendition. Modern chips implement many functions in hardware making them faster while lifting the load off the CPU. The number of hardware supported functions plays a decisive role especially in the 3D world, and most of the development work is currently being conducted in this field.

  Manufacturers are developing new accelerated chip sets at product cycles as short as a half a year, while more complex functions are developed and taken into use by updated Application Programming Interfaces (API).

- *System bus* — The system bus type and speed control the available bandwidth between the CPU, main memory and the graphics adapter. Current desktop configurations use almost exclusively a PCI bus, as older and slower ISA and VESA cards are disappearing from the market. Older Pentium class configurations (P75 – P150) run the PCI bus at 25 or 30MHz, whereas all the current marketed systems are able to run the PCI bus at 33MHz. The new Advanced Graphics Port (AGP) available e.g. on the Intel 440LX chip set can quadruple the bandwidth as it operates at 66Mhz and can transfer data at the rising and falling slope of the trigger signal. The down side of AGP is that is not an actual bus architecture but a port, and can run only one device on it.

  The motherboard chip set that controls the bus can also affect its performance, and better performance can be achieved using the more current chip sets available on the market.

- *Adapter internal bus* — The internal bus needs to transfer data between the graphics chip, video memory and the RAMDAC. Since both the chip and the RAMDAC need to access the video memory virtually simultaneously the bandwidth of the internal bus can easily become a bottleneck. One way of improving the performance is to make the bus wider. Only a few years ago 32 bit buses were the top of the technology being able to transfer 4 bytes per cycle. Nowadays current adapters feature 64 and 128 bit wide buses, being able to simultaneously transfer blocks of 8 and 16 bytes, respectively.

  It is important to notice though, that the data bus of the most commonly used 8x1Mbit memory chip is only 32 bits wide. This means that to get the benefit of a 64 or 128 bit wide bus, 2 or 4 such chips need to be used parallel. Hence an adapter with a 128 bit bus should not be equipped with less than 4 MB of video memory.

- *RAMDAC* — RAMDAC is the component responsible for translating digital information to an analog form that can be sent to the monitor. A screen includes a certain amount of pixels depending on the resolution, and this amount of data has to be transferred to the monitor several times a second depending on the used refresh rate. From an ergonomic point of view, the refresh rate should be over 75Hz, which means that the RAMDAC needs to operate at a very high frequency. A few years old RAMDAC operate at 120-140MHz, while the newer ones achieve rate in excess of 200MHz.

The newest chips have an integrated RAMDAC in an effort to lower costs, but external ones are usually of higher quality.

The speed of the RAMDAC constricts the available refresh rates, which explains why at higher resolutions and color depths, sometimes only mediocre refresh rates can be achieved. It must be carefully checked that the adapter is able to provide sufficiently high non-interlaced refresh rates at the screen modes that are planned to be used.

- *Refresh rate* — The higher the screen redraw rate is, the higher throughput is needed, which stresses the RAMDAC. To cancel out this factor, these tests have been run using an ergonomically recommendable refresh rate of 75Hz.

Software related factors:

- *Driver* — The driver is responsible for translating graphical API calls to the corresponding functions of the chip. The drivers can also emulate (especially 3D) functions that are not supported at hardware level by the chip. For CAD and high end rendering, conformity of the driver to the HEIDI and OpenGL standards can be a decisive factor, whereas for gaming the compatibility with Direct3D is fairly crucial. A driver not implementing emulation or whose API translation is done inefficiently will reduce the overall efficiency, cause incompatibility problems and will be unable to display functions not implemented at hardware level.

Memory related factors:

- *Amount of memory* — The amount of memory determines the maximum resolutions and color depths that can be used. In 2D use the needed amount of memory is easy to calculate by multiplying the screen width, height and color depth in bytes. In 3D a lot more video memory is needed to store the front, back and z-buffer color values. In some modern 3D cards there is also additional memory for texture maps.

  Following calculations show the memory requirements of some commonly used video modes:

800 pixels x 600 pixels x 256 colors (8bit)	=	*0.46M*
800 pixels x 600 pixels x 65K colors (16bit)	=	*0.92M*
1024 pixels x 768 pixels x 65K colors (16bit)	=	*1.5M*
1024 pixels x 768 pixels x 16.7M colors (24bit)	=	*2.25M*
1024 pixels x 768 pixels x 65K + Z-buffer (32bit)	=	*3.0M*
1280 pixels x 1024 pixels x 65K (16bit)	=	*2.5M*

- *Memory type* — The video memory is situated between the graphics chip and the RAMDAC. The chip reads and writes data from the memory, and the RAMDAC would need to simultaneously read the data to send it to the monitor. With conventional DRAM only one of these two components can access the memory at a time, which causes a lot of latency. Due to the extremely high frequency demands the high resolutions and refresh rates impose on the RAMDAC, dual ported solutions have been developed that enable both components to simultaneously and independently access the video memory. Two of the most used dual ported memory types are VRAM and WRAM, of which WRAM has an improved internal structure making it slightly faster than VRAM. Because of their more complex architecture they are more expensive than single ported memory but offer higher performance.

- *Memory speed* Video memory performance can be improved also by higher frequencies as is the case with synchronous memory chips SDRAM and SGRAM that can run at frequencies of 100MHz or higher.

IV. The applied tests

The graphics subsystems have nowadays several tasks to handle, including business graphics, high end picture processing and 3D rendition. This test will focus itself on 2D performance for several reasons. First of all none of these cards, except for the ASUS AGP card, have significant built-in 3D capabilities. 3D rendition is also still limited to a rather narrow application area consisting mainly of games, CAD and professional image editing. Current 3D engines are also reliant on specialized APIs or drivers like DirectX 5, which are not currently available for Windows NT 4.0.

For testing the cards, Ziff-Davies' WinBench 97 was used. WinBench 97 is a complete suite of system benchmarks including benchmarks to measure the graphics subsystem performance with workloads emulating some of the most widely used business and image editing packages. For this comparison, both Business Graphics WinMark and High End Graphics WinMark were run to widely simulate the various real life workloads of professional and home users.

The following is a list of the components of the two benchmarks:

Business Graphics WinMark	High End Graphics WinMark
Access	AVS
CorelDRAW!	MicroStation
Excel	Photoshop
Paradox	Picture Publisher
PowerPoint	PV-WAVE
Word	Visual C++
Word Pro	

Considering the most often used resolutions and the limitations imposed by some cards being equipped only with 2 MB of video memory, every card was subjected to the following tests:

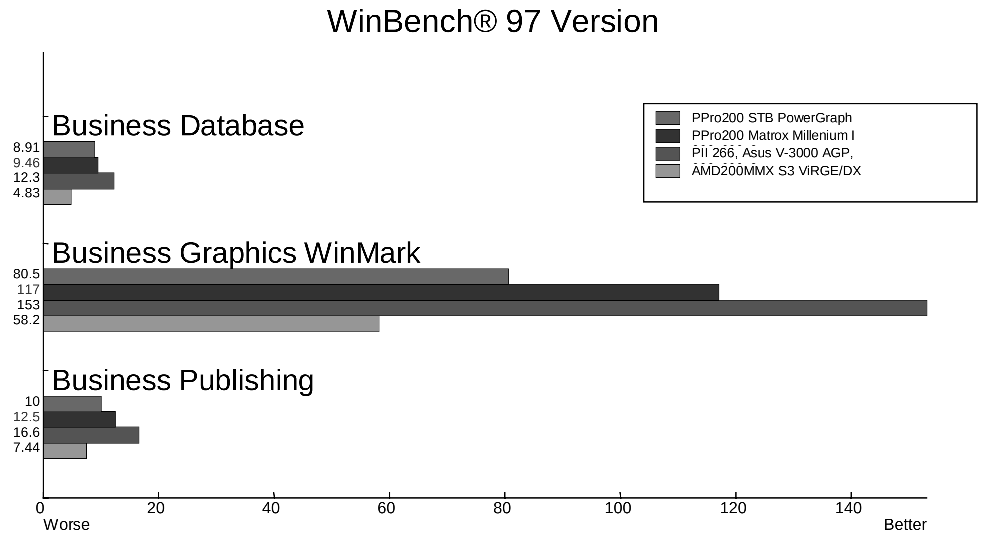
- Business WinMark 800 x 600 x 8 bit color
- Business WinMark 1024 x 768 x 16 bit color
- High End WinMark 1024 x 768 x 16 bit color
- High End WinMark 1280 x 1024 x 8 bit color

All tests were carried out using a refresh rate of 75 Hz.

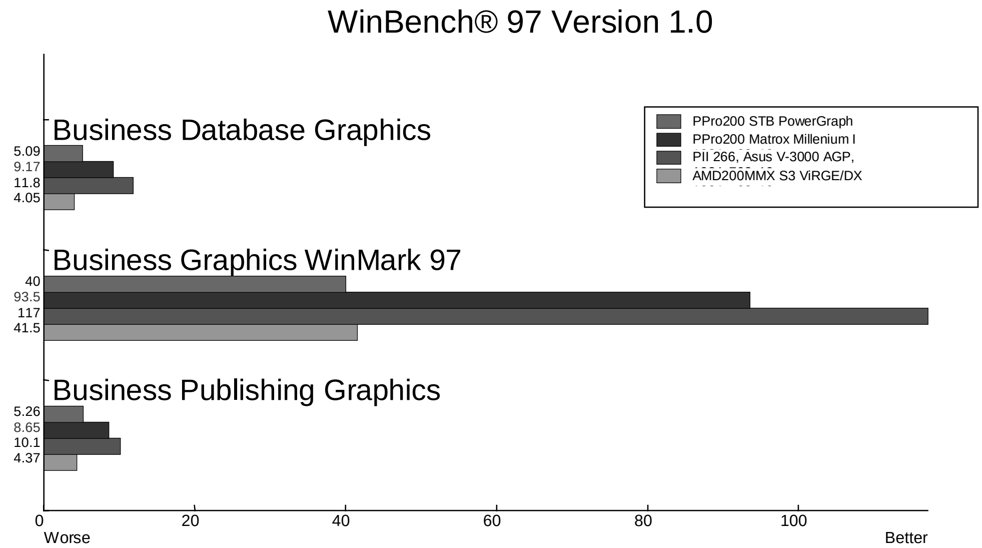
To investigate the impact of the graphics subsystem performance on the overall system performance, the two otherwise identically equipped Pentium Pro 200MHz systems were benchmarked with WinStone 97 suite. This benchmark is described in detail on page 10 of Part I.

The WinStone 97 suite was run using resolution 800x600x8bits and 1024x768x16bits, which correspond to the resolutions used for the WinBench 97 Business Graphics WinMark.

V. Summary of results

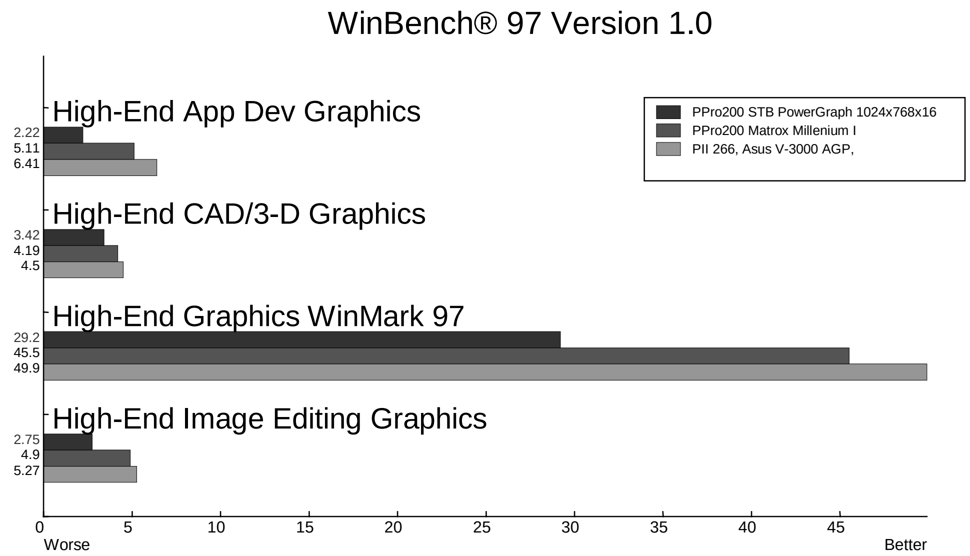


Business Graphics WinMark at 800x600x8bits

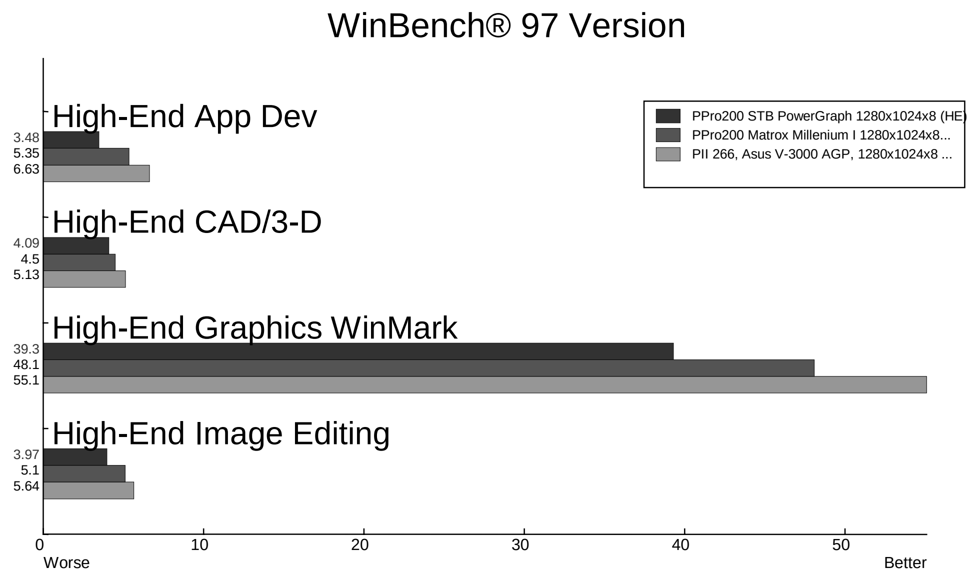


Business Graphics WinMark at 1024x768x16bits

From these results we can see how the Millenium chip clearly outperforms the S3 chips. Interesting to note though is the fact how the STB PowerGraph chip has trouble with the higher resolution, and ends up the slowest of the tested adapters even though it had a healthy marginal over the S3 ViRGE at the lower resolution. On the contrary Matrox Millenium I adapter actually closes the gap to the reference ASUS card a bit at the higher resolution.

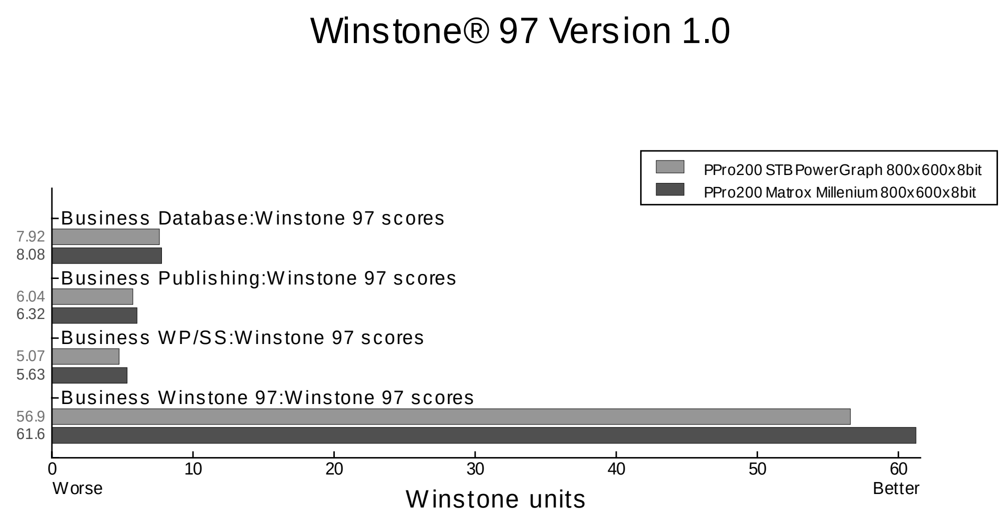


High End Graphics WinMark at 1024x768x16bits

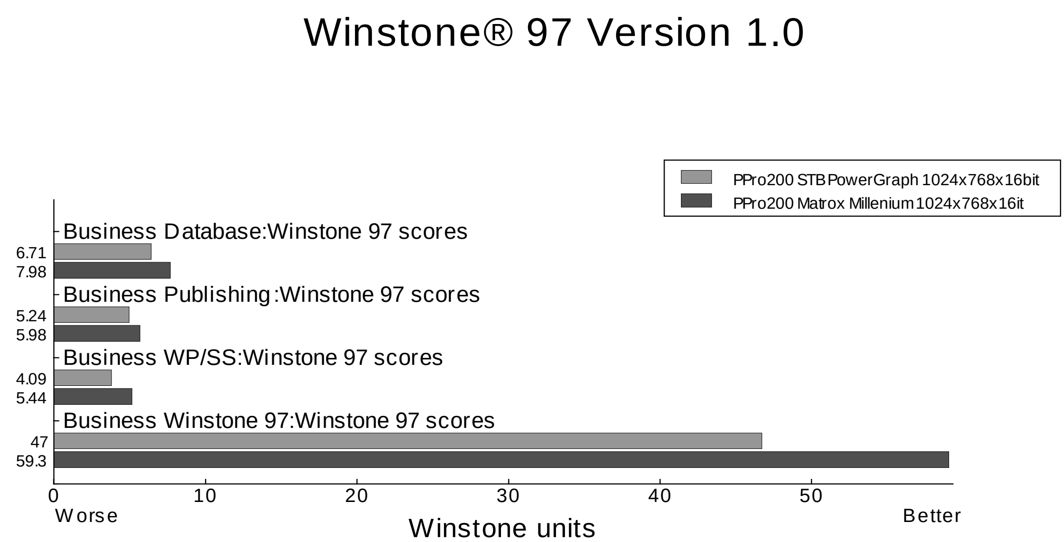


High End Graphics WinMark at 1280x1024x8bits

The trend remains the same during the High End benchmark also. The test could not be run on the S3 ViRGE adapter due to persistent crashes caused most probably by the unstable NT driver.
The performance difference between the Matrox Millenium I adapter and the reference ASUS adapter are only in the region of 10 – 15 percent. Even though the MGA-2064W isn't the current generation of MGA chips anymore, it is known for its good 2D performance. The general improvement curve of 2D performance has already flattened out considerably lately, which explains the modest performance difference. Great improvements in this field of graphics acceleration are not expected anymore.



WinStone 97 at 800x600x8bits



WinStone 97 at 1024x768x16bits

VI. Discussion of overall performance effects

The graphics subsystem is only one component of the complete system. The CPU and the disk system also have a decisive role in determining the overall performance of a computer system, so a twofold performance edge of the graphics subsystem will never translate into a twice faster overall system.

The effect of the graphics subsystem depends also on how screen intensive the application is. In the case of purely mathematical applications or database functions, the performance of the graphical subsystem doesn't really make a difference. The most common desktop applications are fairly screen intensive with frequent redraws due to screen scrolling and windowing. When working with image editing tools, the graphics subsystem is heavily loaded and will have a noticeable effect on the overall performance.

The results of the WinStone benchmarks illustrate both of these points. The machines tested are identical systems apart from the graphics subsystem. Still the results differ by margins of 8 and 25 percent, depending on the used resolution, to the favor of the system with the faster graphics subsystem. A performance boost of one quarter using a common 1024x768x16bit resolution shows clearly the importance of the graphics subsystem on the overall performance.

On the other hand we can refer to WinBench Business WinMark results. From there we can easily calculate that the Matrox Millenium I card performed 45 and 133 percent better when running the tests at the same resolution as the WinStone tests. Still the WinStone figures showed the aforementioned margins of 8 and 25 percent. This clearly shows that the other components of the system have a major influence on the overall performance and performance improvements of single subsystems do not directly translate into correspondingly big overall performance improvements.

CPU Utilization

The CPU utilization was measured for every benchmark that was performed. The tested adapters did not show significant differences in CPU utilization, as all result values were within 96 and 100 percent.

VII. Conclusions about the results

Both in business and high end performance the cards equipped with Matrox' chips clearly outperform the cards equipped with chips manufactured by S3. Since the resolutions and the refresh rates were selected to be such that all cards were capable of handling them, the amount of video memory did not play a decisive role within these results. All graphics subsystems were also running on a 33MHz PCI system bus, and had a 64 bit wide internal video bus. Any 3D functionality built in to the cards was also ignored by the performed tests. Nevertheless several other factors can be pointed out when to explain the measured results.

- Accelerator chip — The MGA chip is a few generations newer than the S3 chips, and deliver considerably higher 2D acceleration.

- Memory type — The Matrox card is equipped with dual ported WRAM, which is faster than the conventional DRAM. The effects of the WRAM architecture, which allows simultaneous video memory accesses for the chip and RAMDAC, and the speed of the RAMDAC, can be seen when inspecting how big of an impact the increase of resolution or color depth has on performance. In the case of a DRAM memory system, with larger amounts of data to be transferred per screen redraw, the RAMDAC reserves more and more video access time, decreasing the overall performance. In WRAM systems the performance is slowed down only due to the increased amount data that needs to be transferred.

  This point is clearly demonstrated by the Business Graphics WinMark results for the different resolutions. The Millenium I card is about 1.4 times faster than the STB card when using the resolution 800x600x8bits, but the advantage is 2.3 fold, when the same test is run at a resolution of 1024x768x16bits. The machines are otherwise identically configured.

- Driver — The newer drivers also play a significant role. The importance of fine tuned drivers became painfully evident through the persistent crashes of the S3 ViRGE during the High End Graphics WinMark.

Matrox Millenium I

The Millenium adapter was the clear performance winner of this comparison. The card is equipped with the MGA-2064W chip, which has already been replaced by the newer generation chip MGA 2164W. In addition to good results in this comparison, the card offers higher resolutions, larger color palettes and higher refresh rates than the ones used in these tests, being able to display video modes up to 1280x1024x24bit and 1600x1200x16bit.

The performance figures of the High End Graphics WinMark are quite impressive as they lag only 10-15 percent behind the reference ASUS adapter with a brand new accelerator chip and that was connected to an AGP port of a powerful Pentium II 266MHz machine. This shows how the level of 2D performance is nearing its maximum, which means that for a business user a Matrox Millenium adapter will be suitable for quite a long time even though the marketplace generally speaks of very short product lifecycles in this field.

STB PowerGraph

The STB PowerGraph is already a two years old adapter equipped with the 64 bit Trio64V+ accelerator chip from S3. The older generation chip cannot keep up with the competition from current chips. The PowerGraph did clearly outperform the S3 ViRGE when using a low resolution, only to be eclipsed by it on the Business WinMark at the 1024x768x16bits resolution. The PowerGraphs 2 MB of DRAM memory does not allow higher resolutions or color depths than the ones used in these tests and higher refresh rates than 75Hz are not available. There are also no builtin 3D capabilities.

S3 ViRGE /DX

The ViRGE /DX is another adapter equipped by a S3 accelerator, the S3D in this case. The performance in 800x600x8bit Business Graphics WinMark was the lowest of the bunch, but the low impact of increased resolution allowed it to win the STB PowerGraph by a fraction in the 1024x768x16bit Business Graphic WinMark. The performance of the drivers let the card down badly, causing persistent crashes when trying to run the high end tests. As with the STB PowerGraph, the 2 MB of DRAM memory does not allow higher resolutions than the ones used in this test.

ASUS V-3000 AGP

The brand new ASUS card used as a reference point was expectedly the winner of all performance tests. The performance margin against the winning Matrox Millenium card lies between 10-30% depending on the test. But a slight improvement in speed is not at all the only advantage of the ASUS card. Containing one of newest 3D capable accelerator chip, this adapter can offer numerous hardware accelerated special graphics functions commonly used in the 3D field. This combined with the ability to benefit from the AGP connection, enables this card to cover by far the broadest spectrum of usage needs.

VIII. Literature:

- Patterson and Hennessy, Computer Organization & Design - ISBN 1-55860-281-X
- WinBench 97 User documentation
- WinStone 97 User documentation
- ThreadMark User documentation
- Excerpts from manufacturers homepages,
  e.g. www.matrox.com, www.adaptec.com, www.stb.com, www.zdnet.com