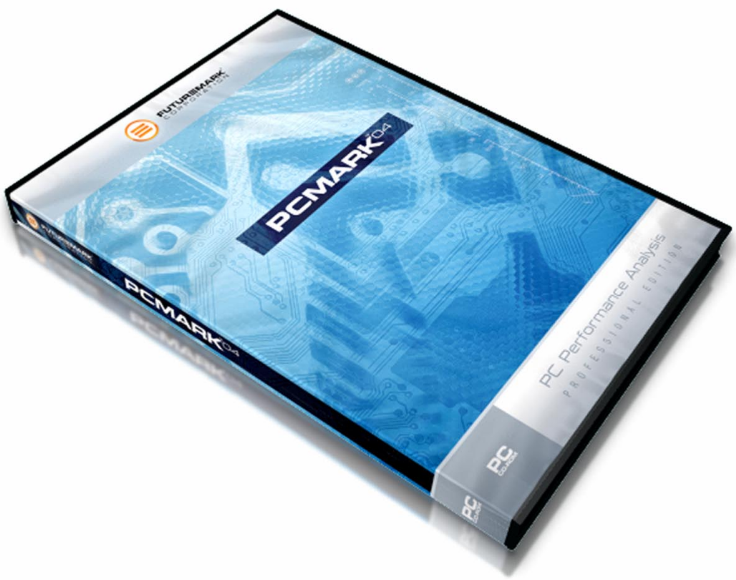
FUTUREMARK CORPORATION

WHITE PAPER

PCMARK04

PC Performance Analysis



Maneesh Dhagat
Futuremark Corporation

12930 Saratoga Avenue,
Suite B-2, Saratoga, CA 95070 USA

November 11 2003

Table of Contents

Overview

This paper introduces PCMark™04, the latest version in the PCMark benchmark series by Futuremark® Corporation. PCMark04 is a state-of-the-art benchmarking tool designed to be easy to use and widely accessible. It produces highly reliable and detailed benchmarking results with a simple, intuitive user interface.

PCMark™2002, the previous version, became a popular tool for benchmarking home and office PC performance. In a short time it grew to be quoted by 73%[*] of on-line and paper publications. PCMark has proved to be a highly accurate tool for measuring mainstream PC usage, much like its sister product 3DMark® has for 3D graphics usage. PCMark has developed a very large following worldwide in its own right. Users of PCMark2002 worldwide have submitted more than 1.5 million benchmark results to Futuremark's Online ResultBrowser database.

PCMark04 builds on Futuremark's strong benchmark development experience by providing a sophisticated tool for measuring PC performance for home usage. PCMark04 supports the complete benchmark cycle – allowing you to benchmark your PC, view the resulting benchmark details, compare your results to those of others, and finally analyze how to improve your PC performance.

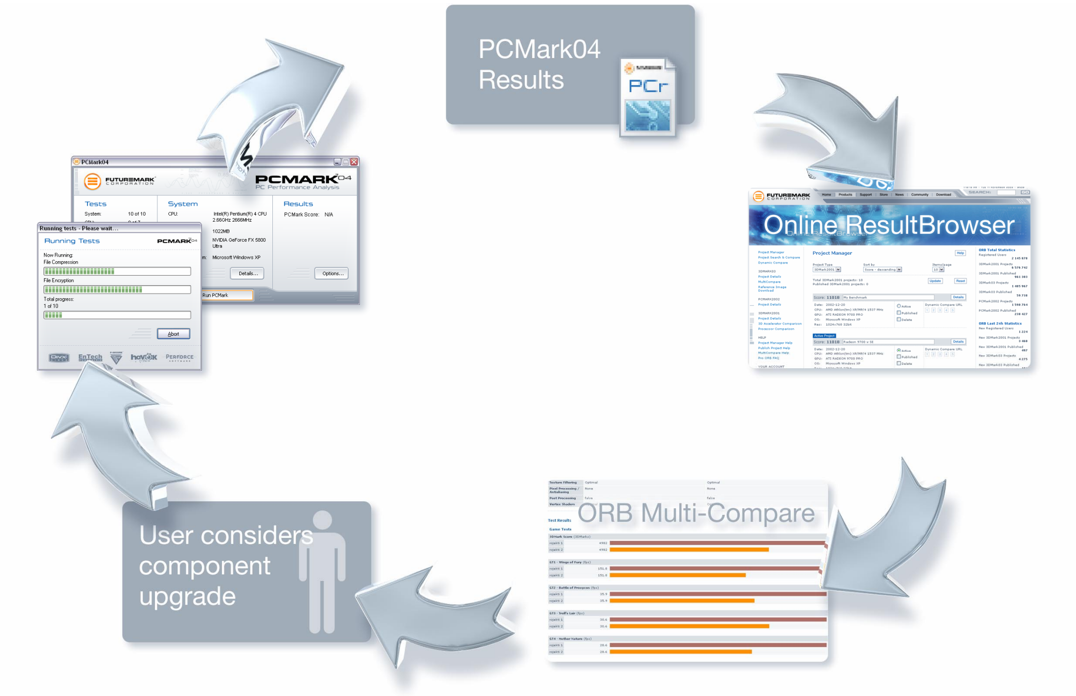


Figure 1: PCMark04 Supports the Complete Benchmarking Cycle.

PCs continue to improve at a dizzying pace. As each PC component – CPU, system memory, graphics card, hard disk, etc. – increases in speed or capacity, it becomes increasingly difficult for the home user to evaluate benefits of these changes. Software applications that may or may not take advantage of the latest hardware add further to the confusion. The goal of PCMark is to solve this problem by providing an easy-to-use tool that outputs easy-to-understand performance figures.

In this paper we discuss our approach to PC benchmarking and the development of PCMark04. We first explore the various facets of successful system and component benchmarking. We next cover our multi-step development process that is central to developing a dependable benchmark. We describe the individual test suites and technologies used in each in detail[†]. Our score calculation approach and the formulae used to

[*] Of 45 reviews covered, 33 reviews used PCMark. *MadOnion Benchmark Media Usage Report – Q1 2002.*
[†] Note that the free version of PCMark04 does not include all tests.

generate the results are also presented. Lastly, the Online ResultBrowser, our integrated and highly popular online comparison tool, is described.

System and Component Benchmarking

System benchmarking allows users to accurately evaluate the overall performance of their PC hardware for a defined usage type. *Component benchmarking*, by contrast, measures the performance of individual PC components, such as CPU, memory, graphics card, etc. Benchmarking PCs can be a complex process as hardware comes in a tremendous variety. Each PC component – motherboard, CPU, system memory, graphics card, etc. – has multiple possible manufacturers and variations. Furthermore, the rapid pace with which manufacturers introduce improved versions of these components makes system benchmarking a very challenging task.

PCMark04 supports both system and component benchmarking. System benchmarking produces a measure of the PC's overall performance for a specific type of usage, in our case, home usage. It is a single number that is easily comparable to that of other PCs. However, advanced users and professional testers may want to further explore the performance of individual components such as the CPU, the memory, the graphics subsystem, the hard disk, etc. Component benchmarking helps them to isolate the performance of individual components, exposing their strengths and weaknesses. PCMark04 even includes the ability to create custom benchmarks by putting together tests tailored to specific needs.

Usage Profile: Identifying the typical usage being measured – the usage profile – is central to building a benchmark. Average home PCs may predominantly be used to browse the internet, compose emails, and view digital photographs. In contrast, gaming PCs will likely have heavy 3D graphics usage for playing the latest games. The set of tasks for the benchmark must closely emulate the chosen usage profile. PCMark04 targets typical home usage.

Benchmarking Approaches: A well-defined usage profile allows us to design the workload – the set of tasks that are run and measured in the benchmark. The workload must have the characteristics of the chosen usage profile. For PCMark04 the usage profile is typical home PC usage. Thus, PCMark04 stresses the PC in a similar manner as PCs are stressed in real life-home usage. There are two common approaches for building the workload: *Application* and *Synthetic* benchmarking.

Application benchmarks include complete applications the user may use. Such benchmarks would include the complete word processor, email client, graphics authoring software, etc. The primary benefit of this approach is that the performance results are likely to correlate closely with the actual performance of using the included applications. Application benchmarks have some drawbacks in their usability and accessibility. They typically have a very large installation footprint – they require a large amount of disk space making distribution and usage more challenging. Moreover, they may take a long time to run due to the inherent size and complexity of the benchmark.

Synthetic benchmarks tend to have the opposite strengths and weaknesses as application benchmarks. They include code fragments written specifically to mimic the work performed by real-world applications. This allows them to be small in file size and quick to run. However, synthetic benchmarks often face the criticism of not being the "real thing" as actual applications are not being used. Despite this, synthetic benchmarks can be powerful tools for isolating performance of certain parts of the hardware.

PCMark04 draws the best parts of both of these approaches by building the workload using an *Application-based* approach. The workload consists of actual applications with freely available public code as much as possible[‡]. However to keep the benchmark small and easy to run, we use only relevant parts of these applications instead of complete applications.

Complete Benchmarking Cycle: A good benchmark provides an accurate performance measurement as well as assists the user through the full benchmarking cycle of analyzing the results; comparing results against other PCs; exploring PC improvements; and leading finally to higher performing PCs. PCMark04 and Futuremark's online tools were designed to support this process.

[‡] Note that certain tests in PCMark04 have been written entirely by Futuremark. These tests, e.g. the memory test suite, require us to isolate the performance of a particular subsystem. Such code is not easily extractable from existing applications and requires us to implement the tests ourselves.

PCMark04's system score is a single, globally recognized number that represents the overall PC performance for home usage. It also produces several component scores. The complete benchmark data available to the user consists of results from 70 tests and over 300 pieces of detailed PC configuration information. As benchmark results are useless in isolation, the user can take the next step and submit the results to Futuremark's Online ResultBrowser (ORB). The ORB allows the user to compare the results to millions of other benchmark results in Futuremark's benchmark databases. With this comparison, the strengths and weakness of the user's PC are revealed. The user can then explore the effect of upgrading some components on overall PC performance using the ORB's multi-compare functionality. This provides the user unique insights into the how much performance improvement each component upgrade may yield. The user, as a result, is more educated and can make more informed purchasing decisions.

Other Benchmarks: There have been several efforts in PC benchmarking. Some prominent ones are listed below:

- SPEC (www.spec.org), a not-for-profit consortium of companies, produces CPU integer-workload and floating point-workload component benchmarks.
- SYSmark® and MobileMark™ series from the BAPCo® Corporation (www.bapco.com) are application benchmarks for internet content creation and office productivity usage. They include a large set of complete applications such as the Microsoft® Office suite and Adobe® Photoshop®.
- SiSoftware™'s Sandra™ (www.sisoftware.co.uk) consists of a large set of different diagnostic tools. It also includes several components-level benchmarks.

With PCMark04 it has been our goal to create a benchmark that is accessible to all users. We want to help users understand their PC performance. For this reason, we have made the benchmark very intuitive to run. With a single click, the user is able to get performance results in minutes. We also have tried to keep the installation size small so that users with limited disk sizes will also be able to use it. We have done this by only including the relevant parts of real applications, and not the complete application. Further, we have worked hard to make the benchmark highly accurate. We have worked in cooperation with all key players in the industry to ensure that the workload accurately reflects real-world usage. Lastly, we provide online tools that allow users put their PC performance results in context and make intelligent purchasing choices.

Our Development Methodology

Futuremark approaches all of the benchmarks it creates with a standard development methodology. We believe that the process we follow is central to the development of a successful and dependable benchmark.

The key part of the development process is cooperation with the key PC technology developers. This cooperation ensures that our benchmark is designed to measure the right things correctly. It ensure that the benchmark will contain tests that are relevant and that the measurements are meaningful. PCMark04 was developed in cooperation with the following members of Futuremark's Benchmark Development Program (BDP): AMD, ATI, Creative Labs, Dell, Gateway, Imagination Technologies, Innovision, Intel, Microsoft, NVIDIA, S3 Graphics, Transmeta, and XGI. Additionally, the following members of media have also been part of the same process: Beyond3D, CNET, and Extremetech.

These companies are willing to cooperate with us because they share the vision that strong, objective benchmarks are in everyone's interest. The Benchmark Development Program allows these companies to participate in designing leading benchmarking standards in the PC industry. The cornerstones of our design process are *transparency* and *neutrality*. We make a strong effort to document all processes that make up the benchmark; we continuously strive to make these documents better. Also, we always maintain the highest standards of neutrality, neither favoring nor ignoring any party.

The figure below depicts a high-level view of our benchmark development methodology.

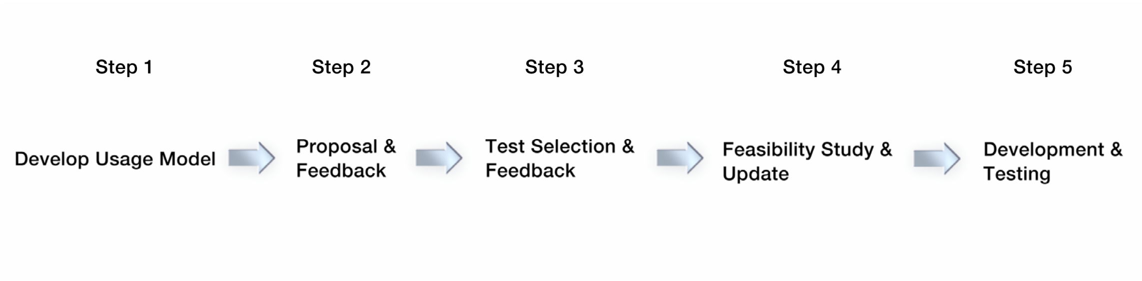


Figure 2: Benchmark Development Methodology

Step One. The benchmark development process begins with identification of target usage profile – the type of PC usage we want to benchmark. For PCMark04, we chose Home PC Usage. Using a variety of resources, we developed a usage model consisting of several categories of tasks. The usage model for home PC is shown in the table below.

Table 1: Home PC Usage Model

Task Category	Task	Weight		
		Low	Medium	High
Productivity	Text Editing		×	
	Spreadsheet Calculations	×		
	Presentations	×		
	Picture Viewing		×	
	High-End Document Editing	×		
Internet	Electronic Mail		×	
	Internet Browsing			×
Entertainment	Music Playing & Recording			×
	Video Playing & Recording			×
	Game Play			×
Other	Desktop Usage			×
	Compression			×
	Encryption			×
	Virus Checking			×

Within each category we list a set of representative tasks. Each task is further assigned a weight reflecting its importance to the chosen usage.

To develop this model, we drew upon our own experience with previous benchmarks and the feedback we have received from the millions of users of these products. Communications with our benchmarking community also provided valuable input. BDP member companies were another source of insights. At this early stage, the specific tests to use and implementation options were intentionally kept open.

Step Two. The document produced in the previous step is a proposal. It is designed to present features and implementation options in a format amenable for getting constructive feedback. This proposal was circulated to our BDP members. The feedback received aided us in choosing tests and implementation methods.

Step Three. In this step, we incorporated the feedback and modified the task list as was appropriate. At this point some lower priority tasks were dropped. We next selected the application fragments to use for each task in the usage model. The applications chosen depended on variety of factors: licensing, size, popularity, etc.

We then created a written benchmark specification. Each of the workload tests is specified in detail with exact versions of the technologies used. For PCMark04, this is summarized below:

Table 2: PCMark04 Tests Used

Task Category	Task	Test Used	Applications Used
Productivity	Text Editing	Grammar check	Link Parser
	Picture Viewing	Image processing	Jpeg image decompression
		Graphics memory	DirectX®
Internet	Internet Browsing	Web page rendering	Microsoft® Internet Explorer 6
Entertainment	Music Playing and Recording	Audio conversion	OGG Vorbis
	Video Playing and Recording	Video en/decoding	WMV and DivX video encoding and decoding
	Game Play	Physics Calculation	Havok
		3D Graphics	DirectX®
General	Desktop Usage	Transparency, moving windows	Windows® 2D graphics API
	Compression	File compression and decompression	Zlib
	Encryption	File encryption and decryption	Blowfish
	Virus Checking	Virus scanning	F-Secure®

The specification was circulated to the BDP members. Their feedback was analyzed and incorporated at our discretion.

Step Four. In the next step we implemented prototype code to see if the available technology would support our plans. This is the stage where we may discover that certain tests are not possible, or may be surprised to discover that more can be achieved. The results were then incorporated into the specification and again circulated to the BDP members.

Step Five. The fifth step consists of implementing the workloads or tests. Periodic releases were made to the BDP members and after each step their feedback was taken into account.

PCMark04 Overview

PCMark04 is an application-based benchmark. It uses portions of real applications instead of including very large applications or using specifically created code. This allows PCMark04 to be a smaller installation as well as to report very accurate results. As far as possible, PCMark04 uses public domain applications whose source code can be freely examined by any user.

PCMark04 measures home PC usage. The workload is designed to stress the PC in the same manner as typical home usage does. The workload includes a test suite that gives an overall PCMark score as well as several component test suites that give individual scores measuring the CPU, memory, graphics, and hard disk drive. Some of the tests run concurrently in separate threads. Multithreading is commonly used in software design to maximize performance and resource utilization. Lastly, PCMark04 includes the ability to define custom test suites that are tailored to your own specific needs. The user can choose various tests and compose them in multithreaded sets. This allows the user to model real-life usage where several applications running in a multithreaded fashion compete for PC resources.

PC component manufacturers often provide tools to optimize code for their hardware. Typical examples are compiler optimization flags. Software developers use these to achieve the best possible performance for their software. As these are generic optimizations, they are also acceptable for use in PCMark04. We allowed CPU manufacturers to propose a compiler and a set of optimizations flags for any code Futuremark compiles. Guidelines for these optimizations were:

- Any compiler and flags proposed must produce correct code, should be documented, must be supported by the manufacturer and should be encouraged for public use
- Futuremark reserves the final right to accept or reject any compiler and flags proposals.
- If no compiler and flags are proposed, generic Microsoft® Visual Studio® .net compilation is used

Optimized compilation is used by default and can be disabled from the user interface.

We also allowed vendors to examine any code that was written by us, for example the memory tests. The vendors suggested optimized code paths, which we reviewed for accuracy and implemented at our discretion.

Note that vendors are not allowed to detect a running PCMark04 instance in any of their drivers or software and use that to reduce the workload or alter the behavior. Our policy for generic optimizations was published earlier this year.

Next, we will describe the contents of the various test suites in PCMark04.

System Test Suite

The system test suite is a collection of tests that are run to generate an overall PCMark score. There are thirteen tests in all in this suite. Three pairs of tests are run multithreaded – each test in the pair is run in its own thread. The remaining seven tests are run single threaded. The setup for the default run is shown in the figure below.

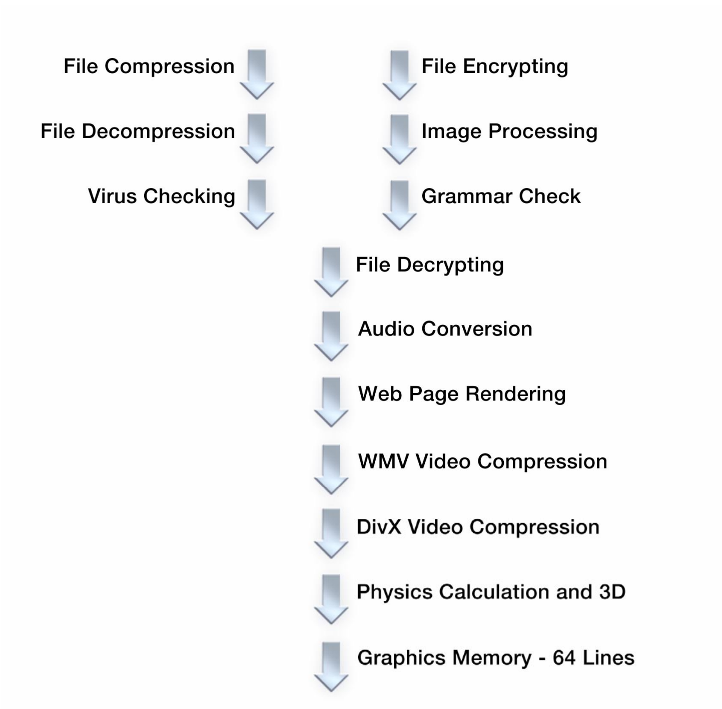


Figure 3: System Test Suite Default Setup

System tests stress different parts of the PC. These tests include some tests which stress mainly the CPU, some tests stressing both the CPU and memory, as well as some tests stressing the CPU, memory and graphics subsystem. We eliminate hard disk activity during system tests as far as possible to avoid variation in the results. If not stated otherwise, all the test data is loaded into memory before running the test.

File Compression / De-compression

With increasingly powerful PCs and the ability of applications to handle larger, more sophisticated documents, file compression has become a very common task. Typical Microsoft® Office documents can now include – besides text – images, sounds, and even video clips. They can easily grow to be several megabytes. It is common to store and exchange them in compressed format. Compression algorithms are also embedded in many applications; very often the home user may use compression and decompression without realizing it.

For file compression and decompression tasks we use the zlib library. Zlib is a free, general-purpose data compression library available on a large number of platforms. It is used in over 500 applications (listed on http://www.gzip.org/zlib/apps.html) including several Microsoft® Office products, Microsoft® Internet Explorer, Symantec™ Norton AntiVirus™, Adobe® Photoshop®, and Macromedia® Shockwave™.

Compression and decompression is integer computation and primarily stresses the CPU. Zlib library version 1.1.4 is used. For this test, the input data files are:

- A 2 Megabyte executable
- A 2 Megabyte Microsoft® Word document
- A 7.4 Megabyte AVI video file
- A 2.7 Megabyte DDS texture file

The test is run continuously for 20 seconds. The result is in Megabytes processed per second.

Note that in the default setup the File Compression test is run concurrently with the File Encryption test. Also, the File Decompression test is run concurrently with the Image Processing test.

Grammar Check

Grammar checking is the automated analysis of text for grammatical correctness. Users of Microsoft® Word are getting increasingly accustomed to getting real-time feedback on mistakes in grammar and style. This functionality makes us more efficient by cutting down the amount of time spent on proofreading. Typical approaches for grammar checking comprise of parsing of sentences followed by applying a rules-based system. The rules are designed to detect issues with capitalization, punctuation, hyphenation, misused words, fragments, use of idioms, etc.

For the grammar checking, task we use the Link Grammar Parsing Library. It was developed at the School of Computer Science at Carnegie Mellon University. The source code and binaries are freely available at http://www.link.cs.cmu.edu/link/. Our analysis using tools such as the Intel® VTune™ Performance Analyzer has shown that this library has a similar runtime profile as grammar checking in commercial applications, such as Microsoft® Word.

Grammar Checking is integer computation and primarily stresses the CPU. Link Grammar Parsing Library version 4.2 is used. The input for this test is a 130 Kilobyte text document. The test is run continuously for 20 seconds. The result is in Kilobytes processed per second.

Note that in the default setup the Grammar Check test is run concurrently with the Virus Scanning test.

Web Page Rendering

Web browsing is becoming the primary use of many home PCs. The Internet has proven a successful way to make money both for internet-only businesses and well as traditional stores. The highest traffic sites have moved cautiously towards making their web pages increasingly complex. Internet connection speeds still vary among users and these sites must balance providing rich content and acceptable response times.

For the Web Page Rendering task we use Microsoft® Internet Explorer 6. This is clearly the predominant web browser. We require that IE 6 be previously installed on the PC. Our goal for this test is to measure the performance of rendering of web pages after they have already been downloaded. We do not measure the speed of downloading web pages. Internet connection speeds can vary due to several factors – type of connectivity (dialup, DSL, cable, etc.), proximity to provider, proximity to web site, etc. – and may not accurately reflect the performance of the PC.

Web page rendering is integer computation and stresses both the CPU and the memory subsystem. The input data for this test are three html pages and related images. These are:

- A 200 Kilobyte web page representing a typical company main page
- A web page containing 4 images (total size approximately 1.4 Megabyte) representing a page containing photos
- A 160 Kilobyte web page containing text representing rendering of documentation

The test is run continuously for 20 seconds. The result is in Pages processed per second.

Image Processing

Image processing for home PC usage primarily includes viewing photographs taken with a digital camera or shared on a web site. It may also include viewing larger images embedded in a web page or document. All of these involve the process of decoding a large image. The most common format found for large images is JPEG. There are two reasons for this. First, JPEG is particularly well suited for rich color photographs and artwork as it uses 24 bits per pixel to store color data. Second, JPEG compresses well allowing images files to be quite small and hence more suitable for storing and transmitting over networks.

For the Image Processing task we use the Standard JPEG library from the Independent JPEG Group. It is production quality software and is freely available for both commercial and non-commercial use at http://www.ijg.org.

Image processing is integer computation and primarily stresses the CPU. The JPEG decoding pipeline uses fixed-point IDCT and RGB-24 output pixel format. The input image is decoded one scan line at the time to a buffer to provide cache coherent memory usage. The input data files are:

- A 130 Kilobyte file
- Two 900 Kilobyte files
- A 1.1 Megabyte file

The image compression ratios for these files vary from 3 to 18 with the content varying from photographs to screenshots. The test is run continuously for 20 seconds. The result is in millions of pixels (MPixels) per second.

Note that in the default setup the Image Processing test is run concurrently with the File Decompression test.

Audio Conversion

The use of audio in PCs is growing beyond simply replying your favorite songs. Audio is becoming an important part of games, web sites and speech enabled applications. The process of recording music involves encoding audio into a music file. Like images, music files come in various formats. Ogg Vorbis and MP3 are some examples. Both of these are similar in that they are considered “lossy” – they discard data in order to compress better than would be otherwise possible. This makes them small and more amenable to transmission over networks.

For the Audio Conversion task we use Ogg Vorbis libraries developed by xiph.org. Ogg Vorbis uses a comparable approach to MP3. However, unlike MP3, it is a non-proprietary format and comes with free, open-source libraries. Vorbis has been used in commercial games like Epic Games Unreal Tournament 2003 and Electronic Arts™ Harry Potter and the Chamber of Secrets™.

Audio Conversion uses floating-point operations and stresses mostly the CPU. Ogg Vorbis libraries libogg 1.0 and libvorbis 1.0 are used. This test encodes a 1.8 Megabyte uncompressed WAV audio file into Ogg Vorbis format. The result is in Kilobytes processed per second.

Video Compression

With digital video cameras gaining increasing popularity, video editing is becoming a popular task on the home PC. For this test we concentrate on the process of compressing movies into formats more amenable to use on

PCs. As with images and audio, various file formats exist for storing video. These formats similarly vary in how much they compress.

For the Video Compression task we have chosen WMV and DivX as the target formats. These are two popular formats that compress well and hence are convenient for use on PCs with limited disk space and for sharing over networks. The process for Video Compression involves encoding video data from one format to a more compact format. Video compression primarily stresses the CPU with floating point operations. For both WMV and DivX compression results are in Frames processed per second.

WMV Compression: For the WMV Compression task we use Windows® Media encoder 9 which must be pre-installed on the PC. For this test the input is lower resolution (320x240) video representing a downloadable movie clip. It is a 1.8 Megabyte MPEG file that is encoded to the WMV data format. The encoding uses a constant bit-rate of 1000 Kilobytes per second for video and 160 Kilobytes per second for audio.

DivX Compression: For the DivX Compression task we use DivX video encoder 5.0.5 included with PCMark04. In this test we use larger resolution (720x480) video. It is a 7.4 Megabyte DV format file that is first decoded internally to a buffer from and then encoded to the DivX format. The decoding and encoding are performed in separate threads. This setup is typical of commercial applications. The bit-rate for encoding is 4000 Kilobytes per second. Note that as DivX Compression reads and writes to the hard disk, hard disk performance may have a minor effect on the results.

Graphics Memory

The Graphics Memory test is designed to stress the video memory subsystem in same manner as typical Windows desktop applications. In home usage, common activities that affect the video memory are moving windows, resizing windows, and scrolling through documents.

The Graphics Memory test stresses CPU, memory, AGP graphics bus and graphics memory subsystem. These tests are written by Futuremark and do not include any public domain or commercial code. They use the Microsoft® DirectX® APIs and are designed specifically to isolate Video Memory Performance.

The test creates a back-buffered primary surface with 1024 x 768 (16 bit) resolution in DX-exclusive mode. A second work surface is created off-screen that is twice as high. For each frame, the work surface is updated by transforming data through the AGP bus with a scrolling speed of 64 scan lines per second. The work surface is then copied every frame to the displayed primary surface to stress the internal memory bandwidth.

The test is run continuously for 10 seconds. The result is in Frame updates per second.

Physics Calculation and 3D Graphics

3D games like Electronic Arts™ Madden NFL™ have been tremendously popular and game developers have been very successful in creating and selling new games. With continual improvement in 3D graphics hardware and software, 3D games are becoming very sophisticated. Besides 3D graphics, games makers are employing complementary technologies like real-time physics, artificial intelligence, and 3D sound. We include this test in PCMark04 to represent 3D game usage on home PCs. Note that Futuremark also produces another benchmark, 3DMark03, which focuses solely on 3D graphics performance.

For this test, we use Microsoft® DirectX® 9 and Havok Physics engine 2.1. DirectX® is the dominant 3D platform on PCs. Havok is a highly popular real-time physics engine. This test is a real-time rendering of several falling blocks. As they fall they collide with each other and their surroundings. Note that this is not an animation or playback of recorded events. All 3D graphics and physics computations are being performed in real-time – exactly as in 3D games.

For this test, the 3D graphics part stresses the 3D graphics subsystem with integer operations and the physics part stresses the CPU with floating point operations. For the 3D graphics workload we use multi-texturing and high-polygon throughput. The physics workload consists of rigid body calculations and collisions. There are 600 blocks and 4 light sources in the scene. We used frame-based rendering (see the 3DMark03 whitepaper for an explanation) – an approach that forces the same number of frames to be generated on all PCs. This ensures the same workload regardless of the hardware used. The result of this test is in Frames rendered per second.

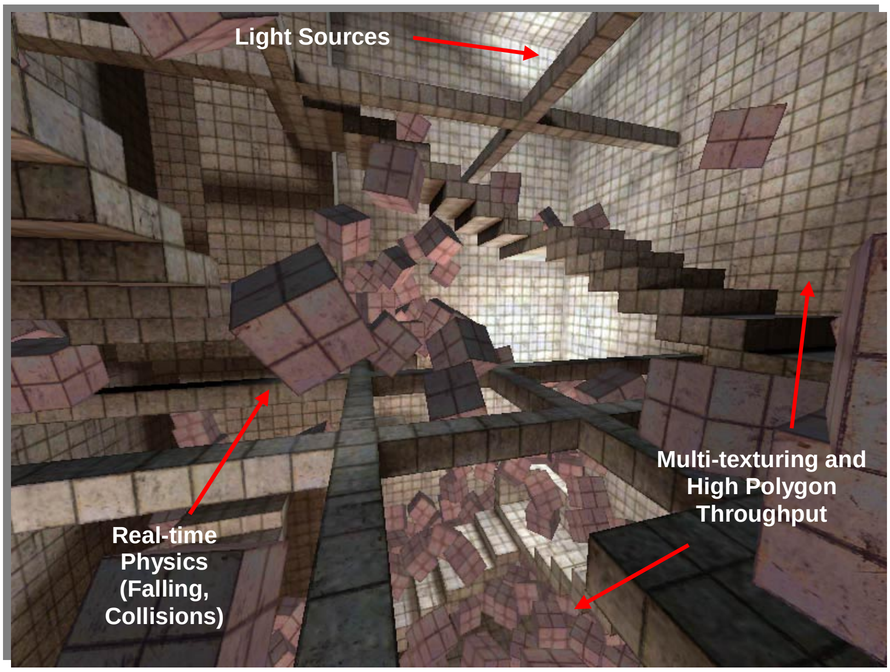


Figure 4: Physics Calculation and 3D Graphics Test

Virus Scan

Virus checking is a critical task. It is becoming common to hear about new viruses disabling computers worldwide. Minor deficiencies in common applications can give the malicious user a way to introduce a new virus. Email and web browsing have become effective mechanisms for unwittingly spreading viruses. Fortunately virus-checking programs have kept pace and are now a required preventative measure against viruses. The virus checker is ubiquitous in home PCs and is typical run in the background, scanning new emails and downloads.

For the Virus Scan task, we use the F-Secure® Anti-Virus™ (www.f-secure.com), a leading commercial virus checking application. It has a large customer base and has received numerous awards. The software has been particularly recognized by reviewers for its ease of use. Other equivalent products are McAfee® VirusScan™ and Symantec™ Norton AntiVirus™.

Virus scanning stresses the CPU and memory with integer operations. In this test the scanner engine consists of two DLLs and a small virus database file. Twenty-one files of different types totaling approximately 23 Megabytes are scanned. The files are scanned consecutively in a loop for 20 seconds. The result is in Megabytes scanned per second. The hard disk drive may have a minor effect on the result.

Note that as virus scanning is typically performed in the background, the Virus Scan test is run concurrently (in a separate thread) with the Grammar Check test.

File Encryption / Decryption

Encryption is the conversion of data into a form (called a ciphertext) that cannot be easily understood by unauthorized people. Decryption is the process of converting encrypted data back into its original form, so it can be understood. Although the use of encryption/decryption is as old as the art of communication, it is particularly important for current day networks, which may be easier to tap into and can carry highly sensitive information. Many companies already enforce policies where all emails must be encrypted.

For the File Encryption and Decryption task we use the Blowfish Algorithm (www.schneier.com/blowfish.html). Blowfish is a symmetric block cipher that can be used as a drop-in replacement for DES or IDEA. It takes a variable-length key, from 32 bits to 448 bits, making it ideal for both domestic and exportable use. Blowfish was designed in 1993 by Bruce Schneier as a fast, free alternative to existing encryption algorithms. Since then it has been analyzed considerably, and is gaining acceptance as a strong encryption algorithm. Blowfish is unpatented and license-free, and is available free for all uses.

Encryption and Decryption stress the CPU with integer operations. For both we use the Crypto++ library version 5.0 (www.cryptopp.com). This is a free C++ class library containing implementations of various cryptographic algorithms including Blowfish. The input files are:

- A 2 Megabyte executable
- A 2 Megabyte Microsoft® Word document
- A 1.1 Megabyte JPEG image
- A 1.8 Megabyte WAV audio file

All input files are encrypted and decrypted in 8 byte padded blocks and the key size is 56 bits. The test is run continuously for 20 seconds. The result is in Megabytes processed per second.

CPU Test Suite

The CPU test suite is a collection of tests that are run to isolate the performance of the CPU. These tests are a subset of the tests included in the system test suite. Excluded are tests that also stress the memory and graphics subsystems. There are nine tests in all. Two pairs of tests are run multithreaded – each test in the pair is run in its own thread. The remaining five tests are run single threaded. The setup for the default run is shown in the figure below.

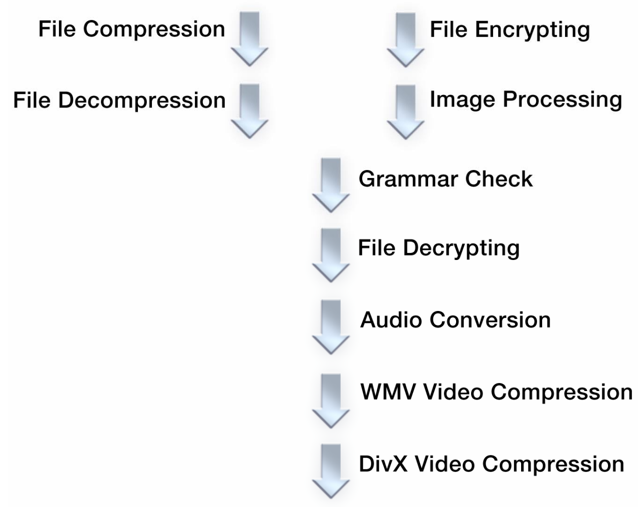


Figure 5: CPU Test Suite Default Setup

Memory Test Suite

The memory test suite is a collection of tests that isolate the performance of the memory subsystem. The memory subsystem consists of various devices on the PC. This includes the main memory, the CPU internal cache (known as the L1 cache) and the external cache (known as the L2 cache). As it is difficult to find applications that only stress the memory, we explicitly developed a set of tests geared for this purpose. The tests are written in C++ and assembly. They include:

- Reading data blocks from memory

- Writing data blocks to memory
- Performing copy operations on data blocks
- Random access to data items
- Latency

Depending on the size of the data blocks, the tests use the main memory, L2 cache or L1 cache. Before each test, cache cleanup procedures are performed to prevent the cache contents from affecting the results.

The read, write, and copy tests in the default setup use data blocks of 8 Megabytes and 4 Megabytes for the main memory and 192 Kilobytes for the L2 cache and 4 Kilobytes for the L1 cache. The test is run continuously for 5 seconds and Megabytes of data processed per second is given as the result. These memory tests are optimized differently depending on the platforms to report the best result.

The random access test uses a large list of data items. The size of each data item in the list is 64 bytes. The list is unsorted (or sorted using a pseudo-random key). Because of this setup, accessing each consecutive item in the list represents a random access operation. The list is iterated through as many times as possible in 5 seconds and Megabytes of items accessed per second is given as the result. Again, 8 Megabytes and 4 Megabytes lists used for the main memory and 192 Kilobytes for the L2 cache and 4 Kilobytes for the L1 cache in the default setup. This test suite, besides the block sizes mentioned, also includes additional block sizes for more testing.

The latency tests are theoretical tests and are meant to isolate the performance a particular part of the memory subsystem. They measure the time it takes to access data memory. These tests use data blocks of 16 Megabytes for main memory, 192 Kilobytes for the L2 cache and 4 Kilobytes for the L1 cache. The results are in Nanoseconds needed to access the data. Note that latency results are not included in the overall memory score.

Graphics Test Suite

The graphics test suite is a collection of tests that isolate the performance of the graphics subsystem of the PC. With each revision of Microsoft® Windows® and related productivity applications, the 2D graphics capabilities of the home PC are pushed further. PCs are being used to look at increasingly sophisticated content such as high-resolution images and video. 3D game play also continues to get more complex fueled by improvements in 3D software and hardware platforms. With this in mind, we created a set of test to expose the graphics performance of your PC.

2D Tests

These tests measure the 2D graphics capabilities of the PC.

Transparent Windows: This test measures the performance of typical windowing operations of your PC. The test uses a workspace of 1024 x 768 with 32-bit color precision. Ten windows are created, each with its own moving and fading speed. Each of the windows is created fully visible and its alpha-blending value is first reduced and then increased to make a fading effect. Fading is a newer Windows® feature, which enhances the visual content of programs. User can set their own preferences for applications for highlighting certain dialogs. The result for this test is the Average Number of windows drawn per second.

Graphics Memory: These two tests are the same as the Graphics Memory test in the system test suite. They only differ in the number of scan lines drawn per second during scrolling. These tests draw 16 and 32 scan lines per second.

Video Playback Tests: This test measures the performance of playing back video files. There are two tests: one plays back WMV content and the other DivX content.

The WMV file has a resolution of 1024 x 768. In this test it is decoded through Windows Media Player. The file data is first passed through a MPEG1 splitter, followed by a MPEG Decoder, and the finally the Video Renderer. The DivX file has a resolution of 640 x 480. During the test it is decoded through DivX codec. The file data first goes through an AVI Splitter, followed by a DivX Decoder, and finally the Video Renderer.

The videos are played back at the highest rate possible. The result is in frames per second.

Note that the score of these tests are not included in the overall graphics test suite score. This is because the maximum frame rate that can be achieved depends on the value the user has set for the screen refresh rate. The score cannot go any higher than this value, even though the graphics hardware may be capable of providing higher frame rates. If the result of this test is equal to the screen refresh rate, the graphics system is capable of playing a variety of video files.

3D Tests

Modern PCs typically have a graphics adapter capable of 3D operations. We have specifically created the following 3D tests to measure the performance of these operations. These tests require DirectX® 7 capable hardware and the DirectX® 9 runtime system pre-installed. The two sets of tests measure fill rate and polygon throughput performance. For more comprehensive 3D graphics benchmarking we suggest using 3DMark03.

Fill Rate Tests: The fill rate is the speed at which your graphics hardware is capable of drawing textures onto 3D objects. These tests report result in million *texels* drawn per second (MTexels/s). Texels or texture elements are the pixels in the source texture. We draw a number of large surfaces covering the entire screen and apply textures to these surfaces using fixed function pixel processing. Alpha blending is also used for transparency. There are two fill rate tests.

Single-Texturing: This test measures single-texturing performance – how fast the graphics card is capable of drawing single textures onto 3D objects. A single texture is applied to each of 64 objects. This means that the graphics hardware must fill each of these objects separately, irrespective of how many texture layers the card is capable of drawing in a single pass.

Multi-Texturing: This test measures multi-texturing performance – how fast the graphics card is capable of drawing multiple textures onto 3D objects. Multiple textures are applied to an object in a single pass until a total 64 textures have been applied. This test benefits from the use of modern graphics cards, which are capable of applying several textures in a single pass. For example, if a graphics card is capable of applying 6 texture layers in a single pass, 6 textures will be applied to an object in 10 passes and the remaining 4 textures will be applied in the 11th pass. Note that fixed pixel texturing has an upper limit of 8 textures per pass. So, even modern DirectX® 9 hardware, capable of up to applying 16 textures per pass, will not be able to apply more than 8.

Polygon Throughput Tests: Polygon throughput is a measure of the ability of your graphic hardware to process 3D graphics primitive objects. The tests report results in millions of triangles drawn per second (MTriangles/s). The scene is a real-time rendering containing a large number of polygons and no texturing. Specular lighting is also used. There are two polygon throughput tests representing different kind of 3D content.

Single Light: For this test we use only a single directional light source.

Multiple Lights: For this test we use 8 directional light sources.

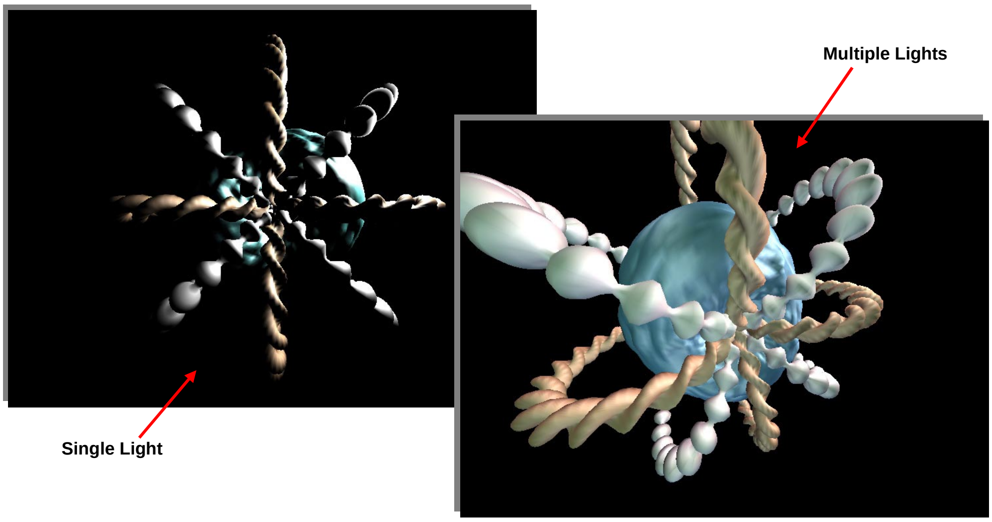


Figure 6: Polygon Throughput 3D Tests

Hard Disk Drive Test Suite

The hard disk drive test suite is a collection of four tests that isolate the performance of the hard disk. The hard disk is an important part of the PC providing quick access to large amounts of data. Today's computers come with hard disks that can store several billion bytes (gigabytes) of data.

For these tests we use RankDisk, an application developed and copyrighted by Intel®. RankDisk is used to record a trace of disk activity during usage of typical applications. These traces can then be replayed to measure to performance of disk operations for that usage.

RankDisk records disk access events using the device drivers and bypasses the file system and the operating system's cache. This makes the measurement independent of the file system overhead or the current state of the operating system. In replaying traces, RankDisk always creates and operates on a new "dummy" file. This file is created in the same (or closest possible) physical location of the target hard disk. This allows the replaying of traces to be safe (does not destroy any existing files) and comparable across different systems. Due to the natural fragmentation of hard disks over time, they should be defragmented before running these tests.

The traces used for each test were created from real usage. The following four input traces are used:

Windows XP Startup: This is the Windows® XP start trace, which contains disk activities occurring at operating system start-up. This trace contains no user activity.

Application Loading: This is a trace containing disk activities from loading various applications. It includes opening and closing of the following applications:

- Microsoft® Word
- Adobe® Acrobat® Reader 5
- Windows® Media Player
- 3DMark 2001SE
- Leadtek® Winfast® DVD
- Mozilla Internet Browser

File Copying: This trace simply contains disk activities from copying approximately 400 Megabytes of files.

General Hard Disk Drive Usage: This trace contains disk activities from using several common applications. These are:

- Opening a Microsoft® Word document, performing grammar check, saving and closing
- Compression and decompression using Winzip
- Encrypting and decrypting a file using PowerCrypt
- Scanning files for viruses using F-Secure® Antivirus™.
- Playing an MP3 file with Winamp
- Playing a WAV file with Winamp
- Playing a DivX video using DivX codec and Windows® Media Player
- Playing a WMV video file using Windows® Media Player
- Viewing pictures using Windows® Picture Viewer
- Browsing the internet using Microsoft® Internet Explorer
- Loading, playing and exiting a game using Ubisoft™ Tom Clancy's Ghost Recon

Disk idle times were compressed to 50 milliseconds to speed up the playback time. Our studies showed that 50 milliseconds was the smallest idle time interval that didn't affect the test results. The results of these tests are in Megabytes processed per second.

Score Calculation

PCMark04 provides an overall PCMark score and four component scores. The PCMark score is a measure of the overall performance of the PC and is obtained by running the system test suite. The components scores are obtained by running the corresponding component test suites: CPU, Memory, Graphics, and Hard Disk Drive test suites. Users can also create custom test suites to meet their needs by putting together various tests. Note that no overall score will be presented for custom test suites. However using the exporting to Microsoft® Excel feature the user can implement this functionality as an Excel formula[§].

PCMark Score

PCMark04 scores will initially range between a 1100 and 5000 PCMarks. They are scaled such that an entry-level system will score approximately 1100 and a high-end system, at the time of product release, will score approximately 5000. An entry-level system is approximately one with a CPU corresponding to 800 MHz clock speed. A high-end system roughly has a CPU corresponding to 3.2 GHz clock speed. Of course, the high scores are expected to keep increasing over time.

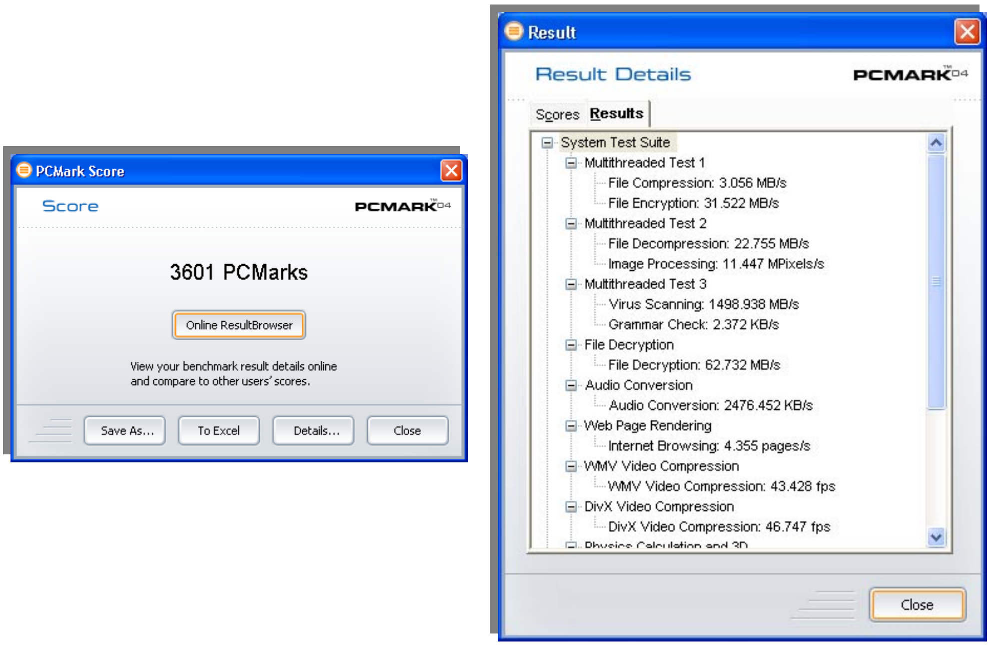


Figure 7: Example PCMark04 Scores

The PCMark score is calculated by using the results of the various tests in the systems test suite. The individual tests are combined using a geometric mean. The geometric mean provides a fair mechanism to combine a large number of test results as compared to assigning arbitrary weights to individual scores. Geometric means are used in other successful benchmarks such as those from SPEC. The geometric mean is scaled using results from reference systems to produce the appropriate range of scores. The formula for the overall PCMark score is:

$$\text{PCMark Score} = 66 \times (\text{File Compression} \times \text{File Encryption} \times \text{File Decompression} \times \text{Image Processing} \times \text{File Decryption} \times \text{Virus Scanning} \times \text{Grammar Check} \times \text{Audio Conversion} \times \text{Web Page Rendering} \times \text{WMV Video Compression} \times \text{DivX Video Compression} \times \text{Physics and 3D} \times \text{Graphics Memory})^{1/13}$$

[§] Note that such custom scores are not comparable to official PCMark scores.

CPU Score

The CPU score is also the geometric mean calculated from the CPU suite test results. The scaling factor is created in a same manner as for the PCMark score. The CPU score formula is:

$$\text{CPU Score} = 110 \times (\text{File Compression} \times \text{File Encryption} \times \text{File Decompression} \times \text{Image Processing} \times \text{File Decryption} \times \text{Grammar Check} \times \text{Audio Conversion} \times \text{WMV Video Compression} \times \text{DivX Video Compression})^{1/9}$$

Memory Score

The memory score is also calculated from the memory suite test results using a geometric mean. The scaling factor is created in a same way as for the PCMark score. The Memory score formula is:

$$\text{Memory Score} = 0.9 \times \{\text{Read 8M} \times \text{Read 4M} \times ((\text{Read 192k} + \text{Read 4k})/2) \times \text{Write 8M} \times \text{Write 4M} \times ((\text{Write 192k} + \text{Write 4k})/2) \times \text{Copy 8M} \times \text{Copy 4M} \times ((\text{Copy 192k} + \text{Copy 4k})/2)) \times \text{Random Access 8M} \times \text{Random Access 4M} \times ((\text{Random access 192k} + \text{Random access 4k})/2)\}^{1/12}$$

Note that, for cache performance, averages of L1 and L2 results were used in the score formula to give more emphasis to system memory performance.

Graphics Score

As the graphics test suite contains fewer tests we do not use a geometric mean to combine the scores. We are able to assign weights to the individual tests based on typical workloads. The weights are obtained by running the tests on reference systems and assigning factors so that approximately: Transparent Windows test accounts for 20%; average of the Graphics Memory tests accounts for 30%; average of the Fill Rate tests accounts for 25%; and average of Polygon Throughput tests accounts for 25%. Note that in the overall graphics score the 2D and 3D test sets are weighted equally. The formula used is:

$$\text{Graphics score} = 0.5 \times \text{transparent windows} + 0.4 \times ((\text{video memory 16 lines} + \text{video memory 32 lines})/2) + 0.6 \times ((\text{fill rate single text.} + \text{fill rate multitext.})/2) + 50 \times ((\text{polygon throughput single light} + \text{polygon throughput multiple lights})/2)$$

HDD Score

Again, as the hard disk drive test suite contains fewer tests we do not use a geometric mean to combine the scores. We are able to assign weights to the individual tests based on actual workloads. The weights are obtained by running the tests on reference systems and assigning factors so that approximately: XP Startup test accounts for 25%; Application Loading test accounts for 28%; File Copying test accounts for 12%; and the General HDD Usage test accounts for 35%. The formula used is:

$$\text{HDD Score} = (\text{XP Startup Trace} \times 120) + (\text{Application Load trace} \times 180) + (\text{File Copy Trace} \times 28) + (\text{General Usage} \times 265)$$

Example Scores

The table below shows sample scores on three PC configurations. Two systems have the same graphics card, and two have the same CPU. Note that the overall PCMark score rises significantly for the more powerful CPU. The overall score is less affected by the other PC components.

Table 3: Example of PCMark and CPU Scores Scaling

System	CPU	AMD® Athlon™ XP 1700+	AMD® Athlon™ XP 2700+	AMD® Athlon™ XP 2700+
	Internal Clock	1469 MHz	2170 MHz	2170 MHz
	Memory	512 MB	512 MB	512 MB
	Graphics Card	ATI® Radeon™ 9700 Pro	ATI® Radeon™ 9700 Pro	ATI® Radeon™ 9000
PCMark Score		**2663**	**3549**	**3309**
CPU Score		**2470**	**3434**	**3456**

The next table shows scaling of the memory score again using three PC configurations. All the systems have the same graphics card, two have the same CPU, and another two have the same memory type. Note that the overall PCMark score increases both with the CPU speed as well as faster memory type. The memory score also rises with both faster CPUs and faster memory. This expected as more powerful CPUs can handle memory more efficiently.

Table 4: Example of PCMark and Memory Scores Scaling

System	CPU	Intel® Celeron® 2.6GHz	Intel® Pentium® 4	Intel® Pentium® 4
	Internal Clock	2593 MHz	3192 MHz	3192 MHz
	Memory	256 MB DDR 266 MHz	256 MB DDR 266 MHz	256 MB dual channel DDR 400 MHz
	Graphics Card	NVidia® GeForce™ Ti4600	NVidia® GeForce™ Ti4600	NVidia® GeForce™ Ti4600
PCMark Score		**2894**	**4486**	**4753**
Memory Score		**2317**	**2993**	**4679**

Online ResultBrowser

The utility of a benchmark result can be rather limited in isolation. Without a mechanism to compare your system to others, it is difficult to say if your system is a high-end state-of-art PC, a mid-end system, or even a low-end beige box. To allow benchmark users to come together to compare and analyze results, we provide a web service called the Online ResultBrowser or ORB.

The ORB has become Futuremark's most popular online service. It provides the users a web application to manage and compare their benchmark results. The ORB database contains over 10 million results. After running the benchmark, the user can choose to *upload* the results and system information to the ORB. We enforce data privacy, so no one except the user will be able to see the individual results. Futuremark also verifies the uploaded results for accuracy. The ORB helps the user by adding meaning to the benchmark data; it allows the user to compare the results with those from other PCs. The user is able evaluate the PC's relative performance - determine the PC's weaknesses and its strengths.

Of course, users can decide to share their results by explicitly *publishing* them. This allows them to show their results to the rest of the world. For many users, their performance results are a point of pride. For some, their position in our top performing PC rankings is the source of a bit of fame. Many proud PC owners include ORB URL links to their published benchmark result in their email signatures.

The ORB is the user's virtual benchmark laboratory. Users can experiment with different system configurations such as increasing the RAM or upgrading the CPU. They can get the new PC performance by searching through results published by others. Users can also maintain their own performance track record by submitting multiple projects to the ORB over time. This is especially useful for keeping track of what an effect do new driver versions have on performance. They can assemble custom multi-compare sets to compare their PC to multiple other configurations.

As the ORB is an online service, Futuremark continues to improve it by adding new functionality. The benchmark data collected by Futuremark is used for generating statistics and recommendations for the user community. This means that every result submitted helps all users to select reliable hardware upgrades. The ORB aids in making buying decisions; before spending money, the user can validate expectations of different hardware options.

System Information

Associated with each benchmark result is the complete profile of the target PC; we call this the *System Information*. The system information not only provides detailed configuration information (CPU speed, RAM, graphics chipset, etc.) to go along with the performance data, but also gives the system state (open applications, AGP mode, free system memory, etc.).

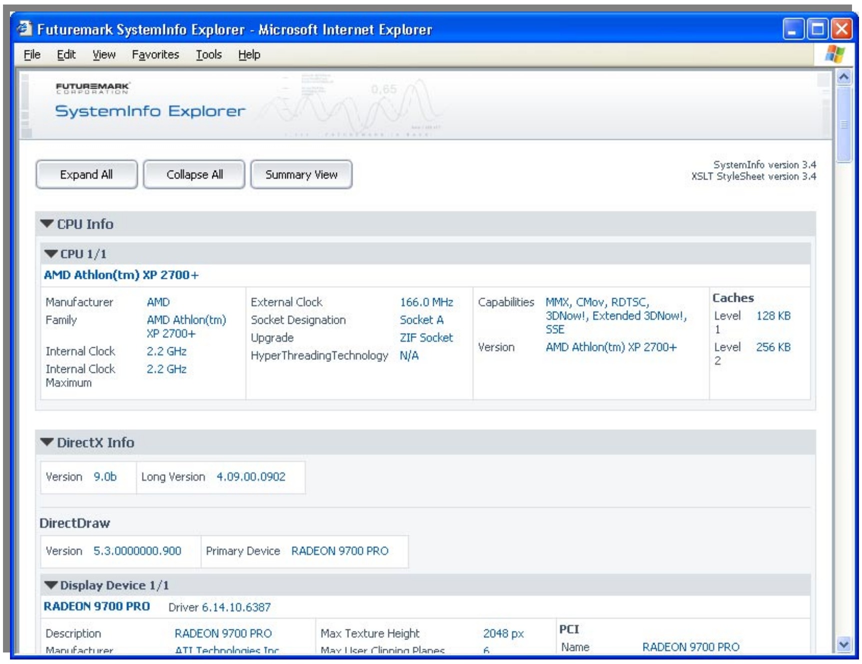


Figure 8: Example PCMark04 System Information

The complete system information consists of over 300 fields**. These include:

- **CPU** Information – clock speeds, internal and external caches, amount of physical and logical CPUs
- **DirectX**® Information – display drivers, direct draw/show attributes, texture formats, 3D capabilities
- **Memory** Information – memory arrays and modules
- **Motherboard** Information – bios, card slots, system devices
- **Monitor** Information – monitor attributes
- **Power Supply** Information – batteries
- **Operating System** Information – version, installed service packs
- **Open Processes** Information – applications, processes
- **Logical Drives** Information – local and network logical drives
- **Hard Disk** Information – disk drive attributes

The ORB uses the system information to enable the search and compare functionality. System information also allows Futuremark to verify the accuracy of published benchmark results. These increasing numbers of system information records are used to provide information back to the users in the form of lists of most popular and powerful hardware components shown on our web site. We also use system information data to build other tools such as the Futuremark Performance Analyzer – an online tool used for product comparisons.

** Note that no private information is ever collected.

Limitations

We have presented the features and benefits of PCMark04. We have detailed where PCMark04 is particularly useful. It is equally important to consider what PCMark04 is not. Below we list what we consider inappropriate uses and limitations of this benchmark tool.

- PCMark04 is not a tool to evaluate the performance of your PC inone particular application. For example, it is not appropriate for comparing the performance of only Microsoft® Word on two PCs. It measures the overall performance of your PC for home usage and performance of individual PC components.
- The system test suite represents average home usage. It may not correlate perfectly with other types of usage. For this, we provide the ability to run custom test suites and to export results to Excel to compose custom scores.
- As some test suites include multithreaded tests, scores will naturally be higher on multiprocessor and hyperthreading systems. So the scores may not be a measure of pure “number-crunching” performance of the PC. For this one should use only the single threaded tests.
- Although component test suites isolate the performance of a particular component such as the CPU or memory, they may be nominally affected by the performance other subsystems. For example, the virus scan test must read and write to the file system and hence is affected to a minor degree by the hard disk drive.
- The graphics tests only measure the performance of the graphics subsystem and not the quality of images rendered. Futuremark’s graphics oriented benchmarks, like 3DMark03, include image quality tools that can help with some of this.

Conclusion

PCMark04 is an application based benchmark with both system and component level tests that has been designed to be easy to install and run. We have taken the application-based benchmark approach by including only the relevant parts of real applications. This has allowed PCMark04 to represent real home usage while keeping it small and quick to run. PCMark04 needs neither gigabytes of disk space nor days to run.

Futuremark cooperates with all major PC industry players, following a neutral and transparent development process, to ensure that the benchmark results are highly accurate.

PCMark04 is unique in enabling the complete cycle for benchmarking PCs – running the benchmark, analyzing the results, comparing results against other PCs, exploring PC improvements, leading finally to higher benchmark results and a more powerful PC.

As PCs and PC components continue to improve at a rapid pace, we believe PCMark04 will serve as a highly dependable tool for the benchmarking professional.