1.0 Introduction

SYSmark® 2004 SE is BAPCo's latest revision of the mainstream office productivity and Internet content creation benchmark used to characterize the performance of the business client. It is based on a foundation of extensive research into emerging usage models and computing trends and incorporates the latest benchmarking methodologies to evaluate platform technologies. It uses application experts (with at least five-years of professional experience) to completely develop the content and operations used throughout the benchmark. As a result, SYSmark 2004 SE contains scientifically designed workloads that represent a range of activities that an office productivity or Internet content creation worker may encounter. In addition with extensive validation across multiple platforms, users can be sure of a consistent and reliable performance comparison.
SYSmark 2004 SE supports Windows XP (Home and Professional) and Windows XP Professional x64 Edition. The scores of SYSmark 2004 and SYSmark 2004 SE can not be compared.

Benchmarks designed by BAPCo are the result of a cooperative agreement between industry leading companies from a variety of disciplines ranging from publications, testing labs, PC manufacturers and semiconductor manufacturing. BAPCo's depth of membership represents the breadth of the computing industry and harnesses a consortium of knowledge to better reflect today's and tomorrow's emerging business trends.

This document describes the methodologies deployed in the development of SYSmark 2004 SE. For detailed instructions on how to install and run SYSmark 2004 SE, please refer to the documentation provided on the DVD-ROM and the BAPCo web site (www.bapco.com)

2.0 Development Methodology

BAPCo follows a development cycle in which it applies its standard development methodologies to a set of parameters driven by usage categories, usage models within the categories, software market segment indicators and technical feasibility. These parameters, along with the application of logical and physical workload characteristics, determine the composition of workloads in SYSmark.

Over fifteen company representatives from BAPCo's membership in addition to numerous experts have collaborated over an eighteen month period so that SYSmark 2004 SE can build upon its predecessors through the utilization of the strongest and broadest development team in PC benchmark history.

The benchmark development process begins with identification of business usage categories of Personal Computers and is followed by determination of the types and characteristics of the output created by users in those categories. By determining the output characteristics, BAPCo can reconstruct the user interactions required to create such outputs. These interactions are converted into instructions (or "scripts") and integrated into BAPCo's automated benchmarking environment resulting in candidate workloads for final placement in the benchmark suite. Candidate workloads are then judged based on their logical and physical workload characteristics and minor adjustments are made in order to arrive at a balanced workload. A key participant in the development process is the application expert provided by member companies. These

application experts have at least five years of professional experience working with the applications.
A graphical representation of BAPCo's output-driven benchmark development methodology for creating workload candidates is shown in Figure 1.

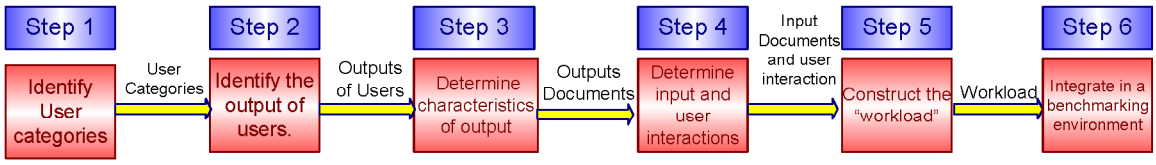

Figure 1: BAPCo's output-driven benchmark development methodology

The following sections will expand on these steps.

The development process is an output-driven methodology as it works backward from a representative output to determine a representative workload. The development process consists of the following basic steps:

- Determination of characteristics of representative output
- Creation of application workloads to generate representative output
- Conformance of the workloads to benchmarking constraints
- Creation of the workflow in each scenario (Internet Content Creation or Office Productivity)

BAPCo members discuss and vote upon the outcome.

The SYSmark 2004 SE workload development process aims to produce representative application workloads for the benchmark. Each application workload consists of three elements: the input data set, operations performed on the input data set and the generated output. An example of generated output would be a PowerPoint presentation that marketing professionals create. These three elements of the workload are used to represent the workflow of a typical user skilled in each given application.

The workload design process begins with BAPCo research to determine the characteristics of the typical finished output created by skilled users. Next, experts create representative input data sets and perform operations to transform the input to the finished output. For each application a number of skilled experts are used. An expert is a skilled user who has at least five years experience in using the given application as part of their work duties. The BAPCo committee members and experts ensure that the operations are in accordance with the storyboard (a narrative of the activities of the professional being modeled in a scenario).

BAPCo experts and members join together for multiple development sessions to collaborate and develop the workload.

BAPCo developers then implement the workload while satisfying benchmarking constraints to ensure correct implementation, relevant and objective workloads and feasible distribution of the benchmark.

Step 6

2.6 Automation and Workload Integration

The final step in creating SYSmark 2004 SE is translating the workload into an automated benchmark. SYSmark 2004 SE emulates a human that is using a PC and mouse. These commands are converted into scripts that will be used by the benchmark to control each of the individual applications included in SYSmark 2004 SE. In order to make sure the benchmark is deterministic, BAPCo uses a framework to install applications, run the scripts, record and display the performance results. For screen shots of the applications running through the framework see Appendix C. The measurement methodology used by the framework is shown below.

2.6.1 Measurement Methodology

The fundamental performance unit in SYSmark 2004 SE is "Response Time". Response time, in the context of SYSmark 2004 SE, is defined as the time it takes the computer to complete a task that has been initiated by the automated script. The task can be initiated by a mouse click or a keystroke from the keyboard. For example, the response time for a Replace All command in Word 2002 is the time between clicking the Replace All button in the Edit/Replace window and the time that Word 2002 brings up the completion window. In SYSmark 2004 SE, only the response time of individual operations is included in the performance metric. SYSmark 2004 SE does not measure the time to send keystrokes and mouse clicks to the application. In the real world, these tasks are initiated and controlled by a human and therefore do not measure computer performance. Since these tasks don't measure computer performance they are not timed. This allows think time to be included in the workloads to more accurately reflect how humans interact with the PC.

SYSmark 2004 SE is designed to emulate human interaction with the computer by sending keystrokes at a rate that approximates how fast a human will type (15 to 50 words per minute). SYSmark 2004 SE departs from the common practice of using an automation tool (harness) that sends keystrokes as fast as the computer can generate them because today's computers are able to generate keystrokes much, much faster than a human can type. In SYSmark 2004 SE, Think Time (of the order of a second) is added between operations to emulate human input to the computer but Think Time is not included in the performance measurement.

Operating system and application behavior is more realistic in SYSmark 2004 SE because Think Time enables the operating system and applications to perform normal housekeeping activities. For example, applications can properly schedule spell checking and grammar checking, the operating system can execute idle time optimizations. Think Time avoids unnatural compression of page use information which can make memory management algorithms in the operating system less effective. The addition of Think Time also enhances cross-platform stability.

In addition to the measurement methodology, BAPCo also needs to make sure that all operations recommended by experts are scriptable by the scripting language. In case the operation is not scriptable, developers and experts consider alternative operations. In order to ensure correct measurement of operations it is also necessary to accurately detect the end of the operation. Similarly, if this is not possible the experts and the developers look for alternatives.

3.0 SYSmark 2004 SE scoring methodology

SYSmark 2004 SE adds the individual response times of all operations within a group (e.g. 2D creation. See section 3.1) and uses the total response time to compare the respective groups on 2 systems (using the calibration system as the base). The scoring methodology reflects BAPCo's decision to reward the responsiveness of a system. This can be illustrated by the following example.

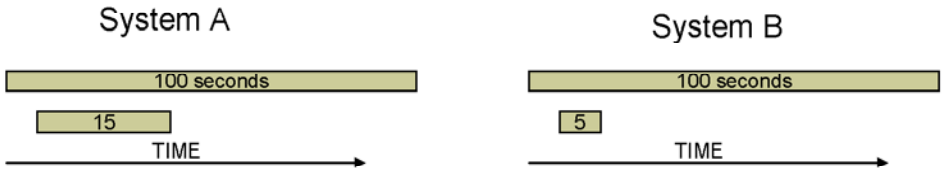

Figure 3: Comparison of 2 tasks on two systems

Two systems are shown running the same 2 concurrent tasks. System B completes the shorter task sooner than System A. By adding the response times of the two tasks, the SYSmark 2004 SE scoring methodology rates system B higher than system A since it is more responsive to the user. The fundamental premise of the SYSmark 2004 SE scoring methodology is that users care about the interactive responsiveness of systems. A throughput-based scoring methodology, which rates a system based on its ability to complete as many tasks as possible per unit-time, would conclude that these systems are rated equally as both complete the 2 tasks in 100 seconds. This ignores the fact that system B is more responsive to user interaction.

The scores of SYSmark 2004 and SYSmark 2004 SE can not be compared.

3.1 SYSmark 2004 SE Performance Rating

Table 1 and Table 2 show the applications that constitute the Internet Content Creation and Office Productivity scenarios. Each of these scenarios is categorized into 3 groups each.

Internet Content Creation Groups:

3D Creation: The user renders a 3D model to a bitmap using 3ds max 5.1, while preparing web pages in Dreamweaver MX. Then the user renders a 3D animation in a vector graphics format.

2D Creation: The user uses Premiere 6.5 to create a movie from several raw input movie cuts and sound cuts and starts exporting it. While waiting on this operation, the user imports the rendered image into Photoshop 7.01, modifies it and saves the results. Once the movie is assembled, the user edits it and creates special effects using After Effects 5.5.

Web Publication: The user extracts content from an archive using WinZip 8.1. Meanwhile, he uses Flash MX to open the exported 3D vector graphics file. He modifies it by including other pictures and optimizes it for faster animation. The final movie with the special effects is then compressed using Windows Media Encoder 9 series in a format

that can be broadcast over broadband Internet. The web site is given the final touches in Dreamweaver MX and the system is scanned by VirusScan 7.0.

Office Productivity Groups

Communication: The user receives an email in Outlook 2002 that contains a collection of documents in a zip file. The user reviews his email and updates his calendar while VirusScan 7.0 scans the system. The corporate web site is viewed in Internet Explorer 6.0. Finally, Internet Explorer is used to look at samples of the web pages and documents created during the scenario.

Document Creation: The user edits the document using Word 2002. He transcribes an audio file into a document using Dragon NaturallySpeaking 6. The user creates a marketing presentation in PowerPoint 2002 and adds elements to a slide show template.

Data Analysis: The user opens a database using Access 2002 and runs some queries. A collection of documents are archived using WinZip 8.1. The queries' results are imported into a spreadsheet using Excel 2002 and used to generate graphical charts.

Group Rating

Each of the groups above have a rating that is calculated by taking the sum of duration of operations that constitute that group and then comparing it with the equivalent sum of the duration of operations on the Calibration System (see section 3.2). The rating is calculated by dividing the sum of the duration of operations in the group on the calibration system by the sum of the duration of operations on the test system and multiplying by 100. The result is then rounded to the nearest integer. At the end of a benchmark run, the Group rating for each of the 6 groups is calculated.

Scenario Rating

The Scenario rating (for each of the 2 scenarios) is calculated by taking the Geometric mean of the Group ratings of all the three groups within each scenario. The scenario rating is rounded to the nearest integer.

SYSmark 2004 SE Overall Rating

The SYSmark 2004 SE overall rating is calculated by taking the geometric mean of the two Scenario ratings. The SYSmark 2004 SE rating is rounded to the nearest integer.

3.2 Calibration System

Each time SYSmark 2004 SE is run on a PC, it calculates a performance rating for each Group (see section 3.1). The group ratings are used to calculate the scenario and overall rating. The group rating is calculated by dividing the sum of the duration of operations in the group on the calibration system by the sum of the duration of operations in the corresponding group on the test system and multiplying by 100. For example, a SYSmark 2004 SE rating of 100 indicates the test system has a SYSmark 2004 SE performance equal to that of the calibration platform when using SYSmark 2004 SE. A rating of 200 indicates the test system has twice the SYSmark 2004 SE performance of the calibration platform. The SYSmark 2004 SE calibration platform has the following configuration:

OS	Microsoft® Windows® XP Professional (Build 2600) plus Service Pack 1
Motherboard	Intel® D845GBV
BIOS	P15
CPU	Intel Pentium® 4 processor 2.0 GHz
Memory	512 MB DDR 266, CL2.5
Sound Card	Creative Labs SoundBlaster Audigy PCI (Model #SB0090)
Sound Card Driver	v5.12.0129 - 1.00.0010
Video	ATI® Radeon® 9700 Pro with 64MB DDR RAM onboard
Resolution	1024x768x32bpp, 75Hz refresh
Video Driver	ATI® reference
File system	NTFS
Disk	IBM® 80GB (2MB cache, 7200 RPM)
Disk Adapter	IDE ATA-100
NIC	Intel® Pro 100+

4.2 System Characteristics of SYSmark 2004

System characteristics of SYSmark 2004 SE are close to SYSmark 2004. The characteristics of SYSmark 2004 are shown in this white paper for reference.
The characteristics of SYSmark 2004 are examined with respect to the different system components. The components that will be looked at are the CPU frequency, main memory, disk I/O and video card. For each of these studies only one system component (e.g. CPU frequency) is varied. All the other system components are held constant. The impact of system changes on concurrency and measured performance is also studied.

4.2.1 CPU Frequency

The CPU frequency of a system is changed from 2.6 to 3.0 GHz. The data in Table 3 shows that the 3D Creation and 2D Creation groups scale close to CPU frequency. On the other hand, the Office Productivity groups are less affected by changes in CPU frequency.

CPU Frequency(GHz)	2.6	2.8	3.0
Frequency	baseline	**8%**	**15%**
SYSmark® 2004 Rating	baseline	**4%**	**9%**
Internet Content Creation	baseline	7%	14%
3D Creation	baseline	7%	13%
2D Creation	baseline	7%	15%
Web Publication	baseline	6%	13%
Office Productivity	baseline	1%	4%
Communication	baseline	0%	3%
Document Creation	baseline	5%	9%
Data Analysis	baseline	-1%	3%

Table 3. SYSmark 2004 CPU Frequency Sensitivity. (Intel * Pentium*4 processor, i865G motherboard, 512 MB memory, Maxtor 6E040L0 37GB IDE, Windows* XP Professional)

4.2.2 Main Memory

Main memory of a system is changed from 256 MB to 2 GB. The results are shown in Table 4. Both Internet Content Creation and Office Productivity benefit from increasing memory size. Particularly, the 2D Creation and Communication groups are most sensitive to memory increase. The 2D Creation group rating benefits by almost 40% when the memory size increases from 256 MB to 2 GB. Similarly, in Office Productivity, the Communication group rating goes up by 27% when the memory size goes from 256 MB to 512 MB and remains constant from then on.

Main Memory (MB)	256	512	1024	2048
SYSmark® 2004 Rating	**baseline**	**15.8%**	**18.0%**	**18.7%**
Internet Content Creation	**baseline**	**15.5%**	**16.8%**	**18.0%**
3D Creation	baseline	2.9%	4.6%	6.9%
2D Creation	baseline	36.3%	38.8%	39.4%
Web Publication	baseline	8.6%	8.6%	9.3%
Office Productivity	**baseline**	**16.7%**	**19.2%**	**19.2%**
Communication	baseline	27.3%	29.5%	28.4%
Document Creation	baseline	10.4%	12.5%	13.2%
Data Analysis	baseline	14.1%	16.3%	17.8%

Table 4 – SYSmark 2004 Main Memory Sensitivity (Intel* Pentium*4 processor at 3.0 GHz, i865G motherboard, Windows* XP Professional)

The 3D Creation group rating increases between 2.9% to 8.9% across all four memory sizes. Over the entire spectrum, there is a big jump in the Internet Content Creation and Office Productivity ratings when system memory increases from 256 MB to 512 MB. However, the ratings are almost flat when system memory increases above 512 MB.

APPENDIX B: Proportional contribution to the score as measured on the calibration system

Due to measurement constraints imposed by the scripting language many tiny operations cannot be specifically measured and hence directly attributed to any group rating. This does not mean that these operations are not performed rather they are retained as so to remain reflective of the user experience. These tiny operations reduce system resources available to those concurrent operations that are specifically measured (time taken is inherently longer than if these operations were not carried out).

In considering the distribution of measurable operations it is therefore important to remember that the proportional contribution (in the table 1) for each application are impacted by operations that cannot be specifically measured from all of the applications within a given group. Furthermore, due to variances in concurrency between configurations the figures shown below are true only for the given calibration system and will vary from system to system.

Application contributions to each group are determined not by the physical time that is spent in the application but by response time of each application. Response time, in the context of SYSmark 2004 SE, is defined as the time it takes the system to complete a task that has been initiated by the user.

In the example below, even though Outlook is 17.2% of the OP-Communications group, this figure shows the time that the system is taking to complete the Outlook task rather than the physical time that a user would have spent in Outlook.

Group	Application	Proportional Contribution
ICC-3D Creation	Dreamweaver MX - Part 1	4.5%
	3ds max 5.1	95.5%
ICC-2D Creation	AfterEffects 5.5	17.9%
	Adobe Photoshop 7.0.1	43.9%
	Premiere 6.5	38.2%
ICC-Web Publication	WinZip 8.1	0.3%
	Dreamweaver MX - Part 2	3.7%
	Window Media Encoder 9 Series	55.6%
	Flash MX	10.1%
	VirusScan 7.0	30.3%
OP-Communications	Outlook 2002	17.2%
	VirusScan 7.0	79.5%
	Internet Explorer 6.0	3.3%
OP-Document Creation	Word 2002	14.8%
	PowerPoint 2002	23.6%
	Dragon NaturallySpeaking 6.0	50.6%
	Acrobat 5.0.5	11.1%
OP-Data Analysis	Excel 2002	76.8%
	WinZip 8.1	3.5%
	Access 2002	19.7%