PC Performance - How to Upgrade Existing Equipment, What Kind of Machine to Buy

By MARK W. PABST

A user wants to run the SAS® System on a PC. Usually within a short time, performance becomes a concern. This paper will discuss in simple terms how the operating system, hardware, and the SAS System work together to get the job done. Knowing this, recommendations are made on what to do with an existing PC and what to look for in future PC purchases. Costs associated with the recommendations are also discussed.

How a PC Works

Before we begin to find the best configuration for a PC, we must first come to understand a little bit of how they work. Figure 1 shows, in general, what a PC looks like if you take off the cabinet. For this example, I have chosen a 286 with a math coprocessor, mouse, and add-in memory. Figure 2 is the schematic for Figure 1. Some selected components are; the RAM (Random Access Memory) or memory, the CPU (Central Processing Unit) or brain of the PC, the NDP (Numeric Data Processor) or math coprocessor which helps the CPU do numeric calculations, the hard disk which stores your applications and data, the mouse or pointing device, the video or screen, and add-in memory which increases the amount of RAM.

Mark Pabst is a Civil Engineer in the Denver Office of the Bureau of Reclamation and uses the SAS® System to analyze field data on existing dams.

Figure 1 - PC without a cabinet

Figure 2 - Schematic of PC

Notice in Figure 1 the RAM, CPU, and NDP are on the motherboard and the other items are on cards that fit into slots located on the motherboard. These card slots are known as the bus. In general, items that are directly on the motherboard will be able to operate more quickly than items on the bus.

The lines between the boxes on Figure 2 are pathways that carry information. With this in mind, think of Figure 2 as a city with several post offices (RAM, CPU, etc.) connected by streets and sidewalks. Notice that the width of the paths between offices is important. The path between RAM and CPU is very wide, perhaps a four-lane street. The path between the bus and the CPU is not quite as wide, perhaps a two-lane street. And finally the path between the video and the bus is quite narrow, a sidewalk. The speed of the offices is also important. If one office is faster than another or faster than it needs to be for the path it is going to use, the operation as a whole will become inefficient. By using our post office analogy it is easy to see how a PC (or our postal operation) could get in trouble. Suppose the hard disk office is very fast as is the CPU office but delivery is slow. The problem is in the path that is taken. The workers at the CPU office have to wait for the delivery person to walk across town on a sidewalk. The converse is true for the workers in the hard disk office on the return route. In a computer, this condition is known as I/O bound. This just means that the deliveries and returns (Input/Output) have become bottlenecked.

Speed of CPU

Processor speed has been increasing greatly in the past few years. At this time, the fastest processors for PC's run at about 33 MHz (larger values are better). Manufacturers are currently working on processors in the 50 to 70 MHz range which should be available in a year or so. Processor speed is not as important as it is made out to be by PC magazines and dealers [1]. On the average, CPU's are doing something (CPU busy) less than half of the time. The rest of the time they are waiting for other parts of the PC to do its work.

Speed of Hard Disk

Not promoted to nearly the extent as CPU speed, but probably more significant are the advancements in hard disk speed. In older machines, hard disk seek times were around 60 ms (smaller
values are better). Today the average hard disk seek time is 28 ms with top-of-the-line models getting 14 ms.

Speed of RAM
Early PCs had RAM speeds of 150 ns (smaller values are better) or slower. Today most PCs come with 100-ns chips with some as fast as 80 ns. Figure 3 illustrates the memory throughput (larger is better) for various machines. Notice that the throughput varies between PCs and even within the PC itself.

Types of CPU
The earliest models of PC's contained 8086 CPUs, followed shortly afterward by the 8088. Then came the 80x86 family of processors. First was the 80286, then the 80386 and now the 80486. In development are the 80586 and 80686. As processors are developed, the two major areas of improvement have been the amount of addressable memory and processor speed. Other features have also been added along the way, such as multitasking and data path width. Unfortunately, processor development has progressed much quicker than any other part of the PC. The 80286 processors can perform multitasking but this has been bypassed because it is an 'old' processor. 80386 processors have a 32-bit data path which isn't used because the software is not available. It appears that it will take about 4 years for software developers to take full advantage of this feature.

Types of Hard Disk
Hard disks must be selected to be compatible with their controller. Typical controllers today use the following interfaces: MFM, ESDI (ezzy), SCSI (souzy) and RLL. There is a lot of debate today over what is the best interface method. Of those listed, the oldest and slowest is MFM. It is also the most tried and true. ESDI and SCSI were then developed to increase performance. RLL was introduced to increase performance over MFM. Now there are hardware disk caching ESDI controllers for even greater speed. Although performance has increased with these methods, sometimes compatibility is lost.

Types of RAM
Earlier PC's recognized 640 Kb of RAM on the motherboard (conventional memory). Later, PC's became available that were able to address larger amounts of RAM. Memory added past 640 Kb became known as expanded memory (EMS) which was developed by Lotus-Intel-Microsoft (LIM). This type of memory requires special circuitry and a software driver. Afterward, extended memory was developed, which requires no special circuitry.

Extended Memory
Adding more chips to the motherboard is one way of adding extended memory. It was found that most motherboards couldn’t hold enough memory so manufacturers produced add-in cards that allowed extended memory additions. The problem with extended memory is that few applications, including PC-SAS®, are able to utilize it. Today you will see PC's with 1 Mb RAM (640 conventional, 384 extended) with the extended memory laying fallow.

Expanded Memory
This memory is also known as EMS (Expanded Memory System). For 80286 class PC's, EMS is usually supplied on its own add-in board because of the need to supply the EMS circuit. For 80386 class PC's, additional memory on the motherboard can be used as EMS because the circuit is part of the 80386 CPU. Some of the newest 80286 class PC's will have the circuit on their motherboard (the NEAT chipset) and don't require a special add-in board. In this way, newer 80286 class PC's can use additional memory on the motherboard as EMS memory. Be aware that some vendors are trying to sell 80286 class PC's which will use extended memory as expanded. The circuit is not in these machines (no NEAT chipset) and this is not the true LIM standard. Performance is very poor with this nonstandard method and it should not be used.

Most add-in memory boards today are expanded memory boards and can be switched off (bypass the circuit) to make them extended memory boards. Better boards will even allow you to mix the amount of extended and expanded memory on the same board. To cause even more confusion, some vendors will use the word 'extended' in the title of their expanded memory boards.
Memory boards also come in varieties of 8, 16, and 8/16 bit bus widths. Be advised that even if you have purchased a 16-bit board, you may have to set it at 8 bit if you have a faster CPU. The quality of the memory board dictates the speed at which it can run.

Shadow RAM
Newer machines will have a type of memory known as Shadow RAM. As the name implies, this memory holds the shadow, or duplicate, of the BIOS. BIOS is even more rudimentary than the operating system (DOS). It tells DOS what hardware is available on the PC. The shadow is needed to move BIOS off of its slow ROM chip and into faster RAM. You will not be able to use shadow RAM for your own purposes. It is dedicated and can’t be altered. Keeping this in mind, you will realize that buying a 1-Mb PC with Shadow RAM really gives you a 640-Kb PC.

How SAS and the PC Work Together
Since the SAS System is so large, being originally developed for the mainframe, some tricks had to be incorporated to have it fit on a PC.

RAM and Overlays
Figure 5 shows an idealization of how the SAS System would look when using PROC GLM. Notice that the space required to run on a mainframe is not available on a PC running DOS. That is, a PROC step requiring 7 Mb will not be able to run on a PC using DOS because only 640 Kb are available. This is known as the 640-Kb barrier.

Conventional

<table>
<thead>
<tr>
<th>640 Kb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proc GLM</td>
</tr>
</tbody>
</table>

Figure 5 - SAS and Proc GLM

To overcome this, a technique known as overlays is used. This is illustrated in Figure 6. In this method, the code is broken up into smaller blocks which will individually fit into the 640-Kb area. Because of this, code and data are constantly being sent back and forth between RAM and the hard disk. This is one reason you see the hard disk light flashing so often when you run your SAS System jobs.

Conventional

<table>
<thead>
<tr>
<th>640 Kb</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM 1/3</td>
</tr>
</tbody>
</table>

Figure 6 - Overlays

Another problem arises when considering the amount of data the SAS System will allow you to analyze. It allows you as much as your hard disk can hold. That is because the SAS System stores your data (and its temporary data files) on the hard disk during execution. This permits you to analyze very large datasets but can cause poor performance. This is another reason the hard disk comes on often.

Now let’s look again at how the SAS System is going to fit into the 640-Kb memory space. Figure 7 shows the SAS System code, broken up into its overlay modules, the data, and DOS files required for supervision of the PC. We will call this group of files the DOS workspace and it needs to fit into the 640 Kb of RAM. DOS Workspace isn’t a physical entity, but rather an idea I am using to help explain this topic.

Conventional

<table>
<thead>
<tr>
<th>640 Kb</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM.EMS 1/3</td>
</tr>
<tr>
<td>GLM.EMS 2/3</td>
</tr>
<tr>
<td>GLM.EMS 3/3</td>
</tr>
<tr>
<td>Data</td>
</tr>
<tr>
<td>DOS and stuff</td>
</tr>
</tbody>
</table>

DOS Workspace

Figure 7 - Getting Into Memory I

As if this weren’t enough, there is the scratch area that is used by the SAS System. In order for the SAS System to run, it requires a scratch area to store temporary files. You have probably noticed this as the SASWORK subdirectory on your hard drive. The SAS System uses this area to store dataset images, matrices and windows information when using Display Manager. Once again, updating our figure we see in Figure 8 the information that needs to get into the 640 Kb of memory.

Conventional

<table>
<thead>
<tr>
<th>640 Kb</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM.EMS 1/3</td>
</tr>
<tr>
<td>GLM.EMS 2/3</td>
</tr>
<tr>
<td>GLM.EMS 3/3</td>
</tr>
<tr>
<td>Data</td>
</tr>
<tr>
<td>DOS (Scratch)</td>
</tr>
<tr>
<td>DOS and stuff</td>
</tr>
</tbody>
</table>

DOS Workspace

Figure 8 - Getting Into Memory II

EMS
As you may have guessed, this is a lot of information to squeeze into 640 Kb of RAM. Even with using overlay structures, sometimes there is not enough room for parts of the SAS System to execute. SAS/Graph procedures are a good example. To overcome this, the SAS System will use EMS memory for execution space. Now the SAS System has 640 Kb plus the amount of EMS on your PC. So instead of 640 Kb the SAS System would have as much as 1024 or 2048 Kb to execute in.
Most software developers will use EMS memory for data storage so you can run larger problems. Lotus 1-2-3 prior to version 3.0 is a good example. The more EMS you have, the larger the spreadsheet you can create. The SAS System doesn't use EMS for data, it uses it for code. Also, if SAS finds EMS in your PC, it will load a different set of code (images). So when the SAS System had to have three images to run PROC GLM in conventional memory, it will only need two when EMS is available. Because of this, the SAS System will appear to run faster because the code used without EMS is different than when you have EMS. I say that it will appear to run faster, because EMS memory is typically physically slower than other types of memory. EMS by itself not only does not enhance performance, it degrades it. The spreadsheet that took so long to calculate without EMS will take even longer with EMS. The SAS System only runs faster because there are fewer images to swap.

Now let's say we have a PC which has 640 Kb conventional memory and 384 Kb EMS. Figure 9 shows how the SAS System will use EMS. It is less complicated than the previous figures.

<table>
<thead>
<tr>
<th>Conventional</th>
<th>EMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>640 Kb</td>
<td>384 Kb</td>
</tr>
</tbody>
</table>

GLM.EMS 1/2
GLM.EMS 2/2
SASWORK (Scratch)
DOS and stuff

DOS Workspace

Figure 9 - Getting Into Memory III

Disk Caching

There are still a lot of exchanges going on between the hard disk and RAM. There is one more thing that can be done to reduce this. Disk caching, as shown in Figure 10, is a technique which is used to store the most recently used hard disk information in RAM.

<table>
<thead>
<tr>
<th>Conventional</th>
<th>EMS</th>
<th>Disk Cache</th>
</tr>
</thead>
<tbody>
<tr>
<td>640 Kb</td>
<td>384 Kb</td>
<td>1024 Kb</td>
</tr>
</tbody>
</table>

GLM.EMS 1/2
GLM.EMS 2/2
SASWORK (Scratch)
DOS and stuff

DOS Workspace

Figure 10 - Getting Into Memory IV

You will need more memory in your PC so the disk cache can use it. Let's say we add another 1024 Kb of memory to our PC. The disk cache will now temporarily store 1024 Kb worth of information. This will be available to all of your applications also, not just the SAS System. As far as the SAS System goes, the scratch area (SASWORK) information will now be stored in RAM. Comparing Figures 8 and 10, we see the number of disk accesses has been greatly reduced. Most work is done between the CPU and RAM. The hard disk light will seldom come on.

PC Performance

Now that it is understood how the PC and the SAS System work, we can make some observations about what to expect in the way of performance. Using a benchmark application [3], different systems can be compared.

Plain PC's

Figure 11 shows that PCs without math coprocessors and extra memory perform about the same regardless of CPU speed. Shown on the left axis are a variety of PC's with different processor types and speeds. The only significant difference in these machines is the 486-25. For reasons discussed later, this is of little consequence. This dramatically illustrates that as far as the SAS System is concerned, PC's are I/O bound.

Enhanced PC's

Figure 12 shows PC's which have had math coprocessors and extra memory added. Notice that the 286-12 outperforms the 386-16. This could be explained by the vintage of the machines. The 286-12 is an early 1990 machine whereas the 386-16 is a 1987. The 386-16 had just come onto the market, as did the 80386 processor, and possibly did not have its design fine tuned. The other 386's and the 486 are all 1990 machines, which include the Shadow RAM feature. You can see that by adding options, there is no appreciable difference between the 486-25 and the faster 386's. Notice that adding options to these machines can improve performance by as much as six times.
As far as performance is concerned, disk caching can give improvements of 200% or more, EMS and NDP about 10% each.

**Costs**

One Mb of RAM costs $60 to $100. Extended/Expanded memory cards that hold up to 3 Mb of RAM cost about $100 each. Disk caching software is about $50. Math coprocessors range from $100 to $600.

**Example 1** - 80386 PC with 640 Kb. The motherboard can hold 4 Mb (4096 Kb).

Adding
- Memory (3456 Kb) $340
- Cache software $50
- NDP $300

**Example 2** - 80286 PC with 640 Kb. The motherboard can hold 640 Kb.

Adding
- 3 Mb expanded memory board $90
- Memory (3072 Kb) $300
- Cache software $50
- NDP $440

**Questions and Answers**

**Q:** Should I replace or add a hard disk?

**A:** This depends on your comfort in going into your PC and tearing it apart. The newer hard disk usually won't work with an older controller. The BIOS is also probably incompatible. A driver may be needed which will take some conventional memory. With these things in mind, it's probably not worth the trouble.

**Q:** Should I add a speedup board (new processor)?

**A:** As far as the SAS System is concerned, you will not get much benefit from this for the cost. Be on the lookout for heat and power problems. This should only be done after memory is added.

**Q:** Should I change the interleave on my hard disk?

**A:** Yes. This is an inexpensive way to improve hard disk performance. Not all PC's will need to have this done. Disk maintenance software, such as Spinrite II, costs about $90.

**Q:** Should I compress my hard disk?

**A:** Yes. Here I am talking about hard disk defragmentation, not file compression. This is another economical way of improving hard disk performance. Compression software costs about $80.

There is one additional question to keep in mind when considering whether to upgrade a PC. Is the machine worthy of an upgrade? Recall that life expectancy of older hard disks may not be much more than four years. Other components may also be near the end of their useful life. I upgraded one machine, which did give it a boost in performance. It also taxed the rest of the system and promptly fried the memory and several chips. The machine is now used for spare parts.

**What to Buy**

When considering which machine to buy, you want the best performance for the least cost (most bang for the buck). In last year's paper [3], I presented a measure of this (performance product) for several machines. Since then, I have benchmarked and computed the performance product for additional machines. The results are shown in Figure 13.
Notice that with the PCD machine the addition of memory more than pays for itself. Also notice that the DTK and Dell have the same performance product. The Dell ran the benchmark in half the time, but costs twice as much as the DTK. What this figure doesn't take into account is age of the machine and decreasing market prices. The top four machines and the DTK were all purchased in the last year. The PCD is about four years old.

The machine that will best suit your needs is determined by the operating system you will be using in the future. The following table gives the order of importance for PC options.

<table>
<thead>
<tr>
<th>DOS/Windows</th>
<th>1. Amount of RAM: 2-8 Mb</th>
<th>OS/2</th>
<th>1. Type of processor: 386, 486, 586, etc.</th>
</tr>
</thead>
</table>

Another performance issue which is promoted is the number of wait-states the processor uses. In previous testing [3] this was found not to be important. In fact, some compatibility problems do arise with this technique. Do not place much importance on this feature.

Some disk controllers now have their own cache - known as hardware disk cache. With this product, memory is added to the controller board, true eliminating the need for additional memory on the motherboard and disk cache software. Some compatibility problems arise with hardware disk caches and their performance is about the same as software disk cache. Some flexibility is also lost in that you will not be able to utilize the memory for EMS.

Shadow RAM will be beneficial for increased performance of video, keyboard and hard disk handling. Do not avoid getting this feature. Recall that you do lose 384 Kb by this method. Some PC's will allow you to disable this feature and thereby permit you to utilize it's RAM. Of course you then lose the performance increase.

Additional Thoughts

It is possible that you have upgraded your machine or bought a new PC using the recommendations in this paper and your performance is still unacceptable. In this case there are several steps you can investigate.

1. Is your code efficient? Hints on efficient coding can be found in several sources [4, 5].

2. Multi-tasking can be used to allow a large SAS application to run in the background while doing other work in the foreground.

Hint

1. Don't bother adjusting the FILEBUFFERS statement in CONFIG.SAS. When using disk caching, the number of SAS FILEBUFFERS is inconsequential. Keep them at their defaults. It is possible to increase the number of FILEBUFFERS, which will take up RAM and not improve performance.

2. Not all disk caches are the same. We have found PC-Kwik to give the greatest improvement in performance. Many disk caches are practically ineffective.

3. The math coprocessor in your PC may have compatibility problems with the rest of the machine. If you find you are getting strange error messages from the SAS System, try disabling the SAS code that uses the math coprocessor with the $NONOP statement in CONFIG.SAS.

4. Zero-wait state techniques used by some PC's may also cause some compatibility problems. Once again, if you are getting strange SAS System errors, disable the 0-wait state option on your PC (this will be different from machine to machine).

5. RAMdrives are a possible way to increase performance. Since it is relatively difficult to utilize this method, I recommend against it.

What's Ahead?

There is quite a bit of discussion about what the operating environment will be for the PC in the future. Currently most all PC's are running in DOS along with it's 640-Kb barrier. This paper has pointed out what a problem this can be for programs such as the SAS System.

Two alternatives are available at this time - OS/2 and Windows. It would appear that DOS by itself is on it's last legs. The decision is then which way to go. OS/2 is nearly impossible to install on an existing machine, and the PC would probably need a hardy upgrade to boot (pun intended). Windows on the other hand is easily installed and will probably only need an increase in the amount of RAM.

References


The following are registered trademarks of SAS Institute, Cary, NC, USA.: SAS, SAS/Graph and associated procedures. AT is a trademark of IBM Corporation. The following are trademarks of Microsoft Corporation: MS DOS, Windows and OS/2. PC-Kwik is a trademark of Multisoft Corporation. Lotus 1-2-3 is registered trademark of Lotus Development Corp. Dell 325 is a trademark of Dell Computer Products.

Disclaimer - The mention of product names, trademarks, or discussion of application by any vendor is for information purposes and does not constitute endorsement or recommendation by the U.S. Government.