Answers to Exercises

♦1. Table 11.2 shows an execution mix and run times for two computers, System A and System C. In this example System C is 83% faster than System A. Table 11.3 shows run times for System A with a different execution mix. Using the execution mix in Table 11.3, calculate the percentage by which System C would be faster than System A. Using the original statistics from Table 11.2, by how much has the performance of System A degraded under the new execution mix?

System performance is one of the key aspects of a processor. It's what determines how fast a given problem can be solved. It can also determine other factors such as the number of problems that can be supported at a given time as well as the range of problems that can be tackled by the processors. The most common way to measure the relative performance between two systems is by the expected running time of a program on each.

We are given the following two tables having information about systems A and C. They show distribution of the types of programs that are run on each system.

Program	Execution Frequency	System A execution time (seconds)	System C execution time (seconds)
V	50%	50	500
W	30%	200	600
X	10%	250	500
Y	5%	400	800
Z	5%	5000	3500

We are given yet another table showing a different distribution of programs that are run on system A:

Program	Execution Frequency	System A execution time (seconds)
V	25%	50
W	5%	200
X	10%	250
Y	5%	400
Z	55%	5000

In the first table, we see that the weighted average time for System A is given by:

$$0.5 \times 50 + 0.3 \times 200 + 0.1 \times 250 + 0.05 \times 400 + 0.05 \times 5000 = 380$$

From the second table, we see that the weighted average time for System A is given by:

$$0.25 \times 50 + 0.05 \times 200 + 0.1 \times 250 + 0.05 \times 400 + 0.55 \times 5000 = 2817.5$$

The weighted average time for System C is given by:

In the first table, we see that the weighted average time for System A is given by:

$$0.5 \times 500 + 0.3 \times 600 + 0.1 \times 500 + 0.05 \times 800 + 0.05 \times 3500 = 695$$

In the first type of distribution for System A, we see that:

$$\frac{Average\_Exceution\_time\_of\_C}{Average\_Exceution\_time\_of\_A} = \frac{695}{380} = 1.82895$$

Hence, System A is 1.83 times faster or 83% faster than System C.

However, using the execution mix for A in the second table, we have:

$$\frac{Average\_Exceution\_time\_of\_A}{Average\_Exceution\_time\_of\_C} = \frac{2817.5}{695} = 4.0539$$

Hence, System C is 4.053 times faster or 305% faster than System A.

Comparing the execution times of System A with the two execution mixes, we have:

$$\frac{Exceution\_Mix1\_of\_A}{Exceution\_Mix2\_of\_A} = \frac{380}{2817.5} = 0.1348$$

Hence, we see that when the Execution Mix2 is used, the performance is just 0.135 or 13.5% of its original. Hence, there is a degradation in performance by **86.5%**

2. With regard to the performance data cited for programs v, w, x, y, and z in Section 11.3, find the geometric means of the run times of the programs for System B and System C using System A as the reference system. Verify that the ratios of the means is consistent with the results obtained using the other two systems as reference systems.

System performance is one of the key aspects of a processor. It's what determines how fast a given problem can be solved. It can also determine other factors such as the number of problems that can be supported at a given time as well as the range of problems that can be tackled by the processors. The most common way to measure the relative performance between two systems is by the expected running time of a program on each.

We are given the following table showing the run times of various types of programs that are run on systems A, B and C as follows:

Program	**System A execution time (seconds)**	**System B execution time (seconds)**	**System C execution time (seconds)**
V	50	100	500
W	200	400	600
X	250	500	500
Y	400	800	800
Z	5000	4100	3500

If there are *n* programs, each having their respective run times on each system, the geometric mean of the run times of one system normalized to another is calculated by taking the product of the ratios of each of the run times and then taking the n^{th} root of the product. Thus, if both System B and System C are to be normalized to System A, then we have the following table:

Program	System A execution time (seconds)	Normalized to A	System B execution ti (seconds)
V	50	1	100
W	200	1	400
X	250	1	500
Y	400	1	800
Z	5000	1	4100
Geometric Mean	$=(1\times1\times1\times1\times1)^{1/5}$ $=1$	$=(0.5\times0.5\times0.5\times0.5\times1.22)^{1/5}$ $=(0.07625)^{1/5}\ =0.5976$	$=(0.1\times0.33\times0.5\times0.$ $=(0.0118)^{1/5}\ =0.41$

The various ratios are as follows:

$$\frac{Geometric\_mean\_A}{Geometric\_mean\_B}=\frac{1}{0.5976}=1.67$$

$$\frac{Geometric\_mean\_B}{Geometric\_mean\_C}=\frac{0.5976}{0.4114}=1.45$$

$$\frac{Geometric\_mean\_A}{Geometric\_mean\_C}=\frac{1}{0.4114}=2.43$$

Thus we see that the ratios of the Geometric means are same irrespective of which one is used as a baseline against which the others are normalized.

3. The execution times for three systems running five benchmarks are shown in the table below. Compare the relative performance of each of these systems (i.e., A to B, B to C, and A to C) using the arithmetic and geometric means. Are there any surprises? Explain.

Program	System A Execution Time	System B Execution Time	System C Execution Time
v	150	200	75
w	200	250	150
x	250	175	200
y	400	800	500
z	1000	1200	1100

Arithmetic mean
formula= (t1+t2+t3+t4+t5)/5
average execution time
A 400
B 525
C 405

Geometric Mean
formula= (t1xt2xt3xt4xt5)^(1/5)
A 13.96
B 165.29
C 1.643

4. The execution times for three systems running five benchmarks are shown in the table below. Compare the relative performance of each of these systems (i.e., A to B, B to C, and A to C) using the arithmetic and geometric means. Are there any surprises? Explain.

Program	System A Execution Time	System B Execution Time	System C Execution Time
v	45	125	75
w	300	275	350
x	250	100	200
y	400	300	500
z	800	1200	700

---

5. A company that is selling database management optimization software contacts you to pitch its product. The representative claims that the memory management software will reduce page fault rates for your system. She offers you a 30-day free trial of this software. Before you install it, however, you decide to first determine a baseline for your system. At specific times of the day, you sample and record the page fault rate of your system (using the system's diagnostic software). You do the same after the software has been installed. How much of an average performance improvement has the new software provided? (Hint: Use the harmonic mean.)

The fault rates and times of day are shown in the table below.

Time	Fault Rate Before	Fault Rate After
02:00-03:00	35%	45%
10:00-11:00	42%	38%
13:00-14:00	12%	10%
18:00-19:00	20%	22%

Here, we are given the information showing the cache fault rates of a system. We must note that the greater the fault rate, the lesser is the performance. After installation of a new product the fault rate is thus expected to decrease. We are given four readings each of the old system and the improved system at different times. Since the readings are rates, we calculate the harmonic mean of each to compare performance.

If there are *n* readings, each expressed as a rate, then the **harmonic mean** is calculated by taking the sum of the reciprocals dividing the number *n* by this sum. Thus, we have the following table:

Serial No.	**Fault Rate Before**	**Fault Rate After**
1	35%	45%
2	42%	38%
3	12%	10%
4	20%	22%
Harmonic Mean	$= \frac{4}{(1/0.35 + 1/0.42 + 1/0.12 + 1/0.20}$ $= 0.215$	$= \frac{4}{(1/0.45 + 1/0.38 + 1/0.1 + 1/0.22}$ $= 0.206$

Thus, we can see that before the installation of the software, the mean cache default rate was at 21.5%. This has now dropped to 20.6%. Hence we can say that the system has improved.

6. What are the limitations of synthetic benchmarks such as Whetstone and Dhrystone? Do you think that the concept of a synthetic benchmark could be extended do overcome these limitations? Explain your answer.

System performance is one of the key aspects of a processor. It's what determines how fast a given problem can be solved. It can also determine other factors such as the number of problems that can be supported at a given time as well as the range of problems that can be tackled by the processors.

The CPU speed is often a key deciding factor about the performance of a system. This factor would be misleading in the sense that it only applies to identical architectures. For different architectures, however, a deeper analysis would be required to arrive at a decision. In order to set a standard that would aspire to guarantee reliability in the eyes of a user, computer researchers derive benchmarks against which users can easily compare metrics to get a fair idea of the performance. Examples of such benchmarks are the Whetstone, Linpack and Dhrystone systems.

The problem with the above contrived systems is that they are basically quite simple. This makes it easy for manufacturers to optimize their products to meet these benchmarks. What results is hardly any differentiation between products to the user. In order to make the process of forming a benchmark more meaningful and comprehensive, a consortium of computer manufacturers, backed by a popular technology magazine came together to form the Standard Performance Evaluation Corporation (SPEC) in 1988.

Establishing reasonable and practical means for measuring performance of computers were the main objectives of this group. The group is well structured and split into three committees each of them focused on a particular aspect of the systems functioning: Open Systems Group focusing on desktops, High-Performance Group targeting multi-processors and supercomputers and Graphics-centric group focusing on visual effects.

Each of these groups identify representative samples of a particular type of work or application. Those aspects of the program which are not related to computing such as input/output etc. are stripped away and the core functionality that remains is referred to as the kernel. The SPEC carefully selects such kernels submitted by various individuals/organizations aimed at a particular area of the three mentioned earlier. The most effective kernels which, in combination are capable of exhaustively testing the system performance are collectively known as the **benchmark suite**.

One major advantage of the meticulousness in this method employed by the SPEC is that while in the earlier benchmarks where manufacturers would easily find loopholes with running the test programs, with the **benchmark suite** it wasn't as simple any longer for the manufacturers to determine the right match of compiler configurations that would help pass the test.

Another avenue the SPEC aimed to offer benefits to users was regarding the "benchmark special" options that several manufacturers often used to market their product by focusing solely on the particular areas of optimization of their products. This was accomplished by adjusting the compiler settings in a particular way to highlight the strengths of the system prior to running the program. The SPEC mandated that the programs were to be run using a prescribed settings so the users could get a circumspect view of the system. While "benchmark special" options were allowed to be published, it wasn't to be without the accompanying prescribed settings results.

Hence, we can see that a more comprehensive set of aspects is required to form a benchmark. Not just the metrics, but also measures to ensure they are not cleverly circumvented need to be focused on.

7. What would you say to a vendor that tells you that his system runs 50% of the SPEC benchmark kernel programs twice as fast as the leading competitive system? Which statistical fallacy is at work here?

System performance is one of the key aspects of a processor. It's what determines how fast a given problem can be solved. It can also determine other factors such as the number of problems that can be supported at a given time as well as the range of problems that can be tackled by the processors.

The CPU speed is often a key deciding factor about the performance of a system. This factor would be misleading in the sense that it only applies to identical architectures. For different architectures, however, a deeper analysis would be required to arrive at a decision. In order to set a standard that would aspire to guarantee reliability in the eyes of a user a consortium of computer manufacturers, backed by a popular technology magazine came together to form the Standard Performance Evaluation Corporation (SPEC) in 1988.

Establishing reasonable and practical means for measuring performance of computers were the main objectives of this group. The group is well structured and split into three committees each of them focused on a particular aspect of the systems functioning: Open Systems Group focusing on desktops, High-Performance Group targeting multi-processors and supercomputers and Graphics-centric group focusing on visual effects.

Each of these groups identify representative samples of a particular type of work or application. Those aspects of the program which are not related to computing such as input/output etc. are stripped away and the core functionality that remains is referred to as the kernel. The SPEC carefully selects such kernels submitted by various individuals/organizations aimed at a particular area of the three mentioned earlier. The most effective kernels which, in combination are capable of exhaustively testing the system performance are collectively known as the **benchmark suite**.

It is recommended that every program in the suite is tested on the target system. The weighted average time of the composite of all the programs is then measured. This would give a clearer picture since the frequency with which each type of program generally occurs varies. Hence, suppose a vendor claims that his system can run half of the benchmark programs faster than the leading provider, it might still not be enough to convince the buyer of its superiority.

For one, these might be lesser frequently occurring programs, so rare that they could be irrelevant. Another possibility is that even if one half of the programs do run a lot faster than the competitor's, the other half might be slower enough to tilt the scale against it.

8. Suppose that you are looking into purchasing a new computer system. You have suitable benchmark results for all of the systems that you are considering except for System X Model Q. The benchmark results have been reported for System X Model S, and they are not quite as good as several competing brands. In order to complete your research, you call System X computer company and ask when they plan to publish benchmark results for the Model Q. The company tells you that they will not be publishing these results anytime soon, but because the disk drives of Model Q give an average access time of 12ms, while Model S had 15ms drives, Model Q will perform better than Model S by 25%. How would you record the performance metrics for System X model Q?

System performance is one of the key aspects of a processor. It's what determines how fast a given problem can be solved. It can also determine other factors such as the number of problems that can be supported at a given time as well as the range of problems that can be tackled by the processors.

The CPU speed is often a key deciding factor about the performance of a system. This factor would be misleading in the sense that it only applies to identical architectures. For different architectures, however, a deeper analysis would be required to arrive at a decision. Disk access speed is another major factor that determines performance, since the access time is significantly higher than RAM- almost in the order of millions.

We are given that the disk access time of Model Q= 12ms, while that of Model S=15ms. Taking the ratios of the access times, we have:

$$\frac{Model\_S}{Model\_Q} = \frac{15}{12} = 1.25$$

Hence, by this measure, we can say Model Q is better than Model S by 25%.

The performance metrics would need to be recorded as a **Weighted Average** since the frequency aspect needs to be taken into account while measuring performance. This is due to the fact that disk accesses occur much rarer than memory accesses. Hence, the memory access times would also need to be considered before determining which system is to be chosen.

9. What value do you think there would be in comparing the results of two different SPEC CPU releases, say, SPEC95 with SPEC2000?

In order to set a standard that would aspire to guarantee reliability in the eyes of a user a consortium of computer manufacturers, backed by a popular technology magazine came together to form the Standard Performance Evaluation Corporation (SPEC) in 1988.

Establishing reasonable and practical means for measuring performance of computers were the main objectives of this group. The group is well structured and split into three committees each of them focused on a particular aspect of the systems functioning: Open Systems Group focusing on desktops, High-Performance Group targeting multi-processors and supercomputers and Graphics-centric group focusing on visual effects.

Each of these groups identify representative samples of a particular type of work or application. Those aspects of the program which are not related to computing such as input/output etc. are stripped away and the core functionality that remains is referred to as the kernel. The SPEC carefully selects such kernels submitted by various individuals/organizations aimed at a particular area of the three mentioned earlier. The most effective kernels which, in combination are capable of exhaustively testing the system performance are collectively known as the **benchmark suite**.

These benchmark suites are constantly updated since they are based on empirical analysis. As more past data is accumulated about the kind of deliverables expected out of a workload as well as the advancing technology. Hence, a later version is conceptually more reliable. However, comparing benchmarks could normally be done by taking surveys from users after the purchase of and use of a system and which benchmark they used to select. A weighted score can be taken for each benchmark depending on how the users graded it.

10. Besides the retail business sector, what other organizations would need good performance from a transaction-processing system. Justify your answer.

The end-users of CPU's, to whom performance is the major criteria, benefit significantly from the benchmarks that are constructed by the SPEC. However, there are also another class of users: those interested in software with a much wider application such as servers which processes multiple transactions from several users. Systems capable of processing numerous simultaneous short tasks would be of more interest to these users.

Standardization of analysis of the performance of such advanced systems troubled manufacturers for a long time. Unlike the benchmarking exercises where results were instantly flashed and bragged about to potential customers, the outputs of such analysis of the enterprise systems are not for public consumption but instead used by the manufacturers to tweak their product. IBM formed such a benchmarking system called the TP1 (Transaction processing). Slowly the system became popular. Even competitors began using the TP1 to check various aspects of performance of the system.

While transaction processing systems is definitely of a major benefit in the retail sector where activities such as shopping cart inventory, order processing and bill generation are handled by the system, these systems also have application in other areas which involve a large number of customers.

The **banking sector** is one such area. Aspects of banking such as cash deposits, passbook updating and cheque transfers are all capable of being efficiently supported with the help of such transaction processing systems. The **hospitality sector** is another such area where, aspects such as suite selection, room allocation, check-in and check-out details, maintaining the bills for each guest can be supported by the transaction processing system. Yet another area of application is at various **fast-food eating chains** such as Mc Donald's or Burger King, where the order of each customer is punched into a machine and the bill is after calculating the total with the help of item prices pre-fed into the machine.

11. Which of the benchmarks discussed in this chapter would be most helpful to you if you were about to purchase a system to be used in DNA research? Why would you choose this one? Would any of the other benchmarks be of interest to you? Why or why not?

In order to make the process of forming a benchmark more meaningful and comprehensive, a consortium of computer manufacturers, backed by a popular technology magazine came together to form the Standard Performance Evaluation Corporation (SPEC) in 1988.

Establishing reasonable and practical means for measuring performance of computers were the main objectives of this group. The group is well structured and split into three committees each of them focused on a particular aspect of the systems functioning: Open Systems Group focusing on desktops, High-Performance Group targeting multi-processors and supercomputers and Graphics-centric group focusing on visual effects.

Each of these groups identify representative samples of a particular type of work or application. Those aspects of the program which are not related to computing such as input/output etc. are stripped away and the core functionality that remains is referred to as the kernel. The SPEC carefully selects such kernels submitted by various individuals/organizations aimed at a particular area of the three mentioned earlier. The most effective kernels which, in combination are capable of exhaustively testing the system performance are collectively known as the **benchmark suite**.

The benchmark suite contains kernels specific to various functions in a variety of areas of use and research in many fields such as mathematics, physics, chemistry etc. DNA research involves the analysis of the basic human building blocks and the biochemical implications of it. From the set of benchmarks listed in the chapter, we see two that would be of use for this purpose:

Benchmark Suite	**Benchmark**	**Area of Application**	**Description**
SPEC CINT2006	456.hmmer	Search Gene sequence	Analysis of protein sequences
SPEC Cfp2006	444.namd	Biology/Molecular Dynamics	Used for simulation of large biomolecular systems

12. Suppose a friend has asked you to help him to make a choice as to what kind of computer he should buy for his personal use at home. What would you look for in comparing various makes and models? How does your line of thinking differ in this situation than if you were to help your employer purchase a Web server to accept customers' orders over the Internet?

In order to make the process of forming a benchmark more meaningful and comprehensive, a consortium of computer manufacturers, backed by a popular technology magazine came together to form the Standard Performance Evaluation Corporation (SPEC) in 1988. Establishing reasonable and practical means for measuring performance of computers were the main objectives of this group.

The end-users of CPU's, to whom performance is the major criteria, benefit significantly from the benchmarks that are constructed by the SPEC. However, there are also another class of users: those interested in software with a much wider application such as servers which processes multiple transactions from several users. Systems capable of processing numerous simultaneous short tasks would be of more interest to these users.

Standardization of analysis of the performance of such advanced systems troubled manufacturers for a long time. Unlike the benchmarking exercises where results were instantly flashed and bragged about to potential customers, the outputs of such analysis of the enterprise systems are not for public consumption but instead used by the manufacturers to tweak their product. IBM formed such a benchmarking system called the TP1 (Transaction processing). Slowly the system became popular. Even competitors began using the TP1 to check various aspects of performance of the system. One disadvantage of the system, though, is that it only reported the best performance achievable under ideal conditions.

Hence, to search for a personal computer, like that requested for by a friend, the conformance to the SPEC benchmarks would be of primary importance while, for a system such as a web server capable of servicing a large number of requests, the TP1 (transaction processing) benchmarks would need to be fulfilled.

13. Suppose you have just been assigned to a committee that has been tasked with purchasing a new enterprise file server that will support customer account activity as well as many administrative functions, such as producing a weekly payroll. (Yes, a committee frequently makes these decisions!) One of your committee members has just learned that a particular system has blown out the competition in the SPEC CPU2000 benchmarks. He is now insisting that the committee buy one of these systems. What would be your reaction to this?

In order to make the process of forming a benchmark more meaningful and comprehensive, a consortium of computer manufacturers, backed by a popular technology magazine came together to form the Standard Performance Evaluation Corporation (SPEC) in 1988. Establishing reasonable and practical means for measuring performance of computers were the main objectives of this group.

The end-users of CPU's, to whom performance is the major criteria, benefit significantly from the benchmarks that are constructed by the SPEC. However, there are also another class of users: those interested in software with a much wider application such as servers which processes multiple transactions from several users. Systems capable of processing numerous simultaneous short tasks would be of more interest to these users.

Standardization of analysis of the performance of such advanced systems troubled manufacturers for a long time. Unlike the benchmarking exercises where results were instantly flashed and bragged about to potential customers, the outputs of such analysis of the enterprise systems are not for public consumption but instead used by the manufacturers to tweak their product. IBM formed such a benchmarking system called the TP1 (Transaction processing). Slowly the system became popular. Even competitors began using the TP1 to check various aspects of performance of the system. One disadvantage of the system, though, is that it only reported the best performance achievable under ideal conditions.

Hence, to search for a personal computer. the conformance to the SPEC benchmarks would be of primary importance whereas, for a system such as a web server capable of servicing a large number of requests, the TP1 (transaction processing) benchmarks would need to be fulfilled. Even though the new system is better than its competition, what is important to understand is that the benchmarks it does well running is of the SPEC type. These are used for mostly personal use. Even though many types of complex applications could be run on the system, it still wouldn't be able cope with the sheer volume of requests that an industrial transaction processing system is designed to support.

Hence, it would not be advisable to employ the use of the system meant for personal use for the purpose of a transaction processing system no matter how efficient it is since each type of system has a specific purpose.

system has a specific purpose.

* 14. We discussed the limitations of the harmonic mean in its application to computer performance assessment. A number of critics have suggested that the SPEC should use the harmonic mean instead. Suggest a unit of “work” that would be appropriate for reformulating the SPEC benchmarks as rate metrics. Test your theory using results from SPEC’s Web site, www.spec.org.

In order to make the process of forming a benchmark more meaningful and comprehensive, a consortium of computer manufacturers, backed by a popular technology magazine came together to form the Standard Performance Evaluation Corporation (SPEC) in 1988. Establishing reasonable and practical means for measuring performance of computers were the main objectives of this group.

The end-users of CPU’s, to whom performance is the major criteria, benefit significantly from the benchmarks that are constructed by the SPEC. However, there are also another class of users: those interested in software with a much wider application such as servers which processes multiple transactions from several users. Systems capable of processing numerous simultaneous short tasks would be of more interest to these users.

The SPEC publishes benchmarks that test the systems for competence in various aspects. The performance of these benchmark is measured in terms of the time taken up by each benchmark to run on the system. In order to assess the performance of the system in terms of harmonic mean, we would need to express the benchmarks as rates.

Since, during the process of a program running on a system, several checkpoints or function points are passed along the way, we could measure the number of such points passed per unit time. We could also arrange for a number of flags to be raised as each point is passed. This rate could then be used to measure the harmonic means.

* 15. SPEC and the TPC both publish benchmarks for Web server systems. Visit the respective Web sites of these organizations (www.spec.org and www.tpc.org) to try to find identical (or comparable) systems that have results posted on both sites. Discuss your findings.

System performance is one of the key aspects of a processor. It's what determines how fast a given problem can be solved. It can also determine other factors such as the number of problems that can be supported at a given time as well as the range of problems that can be tackled by the processors.

In order to set a standard that would aspire to guarantee reliability in the eyes of a user,

a consortium of computer manufacturers, backed by a popular technology magazine came together to form the Standard Performance Evaluation Corporation (SPEC) in 1988. Establishing reasonable and practical means for measuring performance of computers were the main objectives of this group.

The end-users of CPU's, to whom performance is the major criteria, benefit significantly from the benchmarks that are constructed by the SPEC. However, there are also another class of users: those interested in software with a much wider application such as servers which processes multiple transactions from several users. Systems capable of processing numerous simultaneous short tasks would be of more interest to these users.

Standardization of analysis of the performance of such advanced systems troubled manufacturers for a long time. Unlike the benchmarking exercises where results were instantly flashed and bragged about to potential customers, the outputs of such analysis of the enterprise systems are not for public consumption but instead used by the manufacturers to tweak their product.

But even though SPEC was generally used to assess performance in the personal computing space while TPS was targeted at the enterprise space, there are some systems which meet benchmarks in both the SPEC and TPC formats. Some of those systems are as follows:

1) **WebSphere Application Server V7 on IBM System x3650 and DB2 9.7 on IBM System x3850:** This software is used for developing code in the IBM based technologies such as Lotus Notes, Document Management etc.

2) **Oracle Weblogic Server Standard Edition Release 10.3.4 on Cisco UCS B440 M1 Blade Server**: This software is used for developing code in the Oracle based technologies such as PL/SQL, Database queries etc..

3) **SAP NetWeaver CE 7.1, MaxDB 7.8 on HP DL380 G6 X5570/X5560, Novell SLES10:** This software is used to build the ERP architecture for large organizations which helps set up integration of data.