No. 54, May 1998

THE COMPUTER CORNER

The Numbers Game

- by Stan Kaplan, WB9RQR
105 Martin Drive
Port Washington, WI 53074-9654
(414) 284-9346
WB9RQR @ N9PBY.EN63BI.WI.USA.NA
skaplan@mcw.edu

Last month was a hardware topic - cleaning up your power supply. Sort of a practical-type topic. This month I will get more academic, but it should lead to some understanding where there was confusion before. At least it will show you why the confusion cannot really be resolved!

You just bought a new 1.2 GB (gigabyte) hard drive, installed it, and it works fine. However, you notice from time to time that your CMOS, or one program or another, reports that you have a 1222.6 MB (megabyte) disk. What gives? Where did the extra 22.6 MB come from? On the other hand, when you run the DOS program CHKDSK, it reports 1281.9 MB for the same drive. Why are they not all exactly the same?

Basically, the answer stems from two facts. First, the math used in the computer world uses somewhat different rules than those used in everyday math. Second, the hard drive manufacturers themselves don't know!

You and I expect that the prefix kilo means that there are 1,000 of something. The ship's captain who says his ship carries 14 kilotons of cargo is saying that the cargo weighs 14,000 tons. A kilovolt is 1,000 volts.

In the language and math of computers, however, kilo means 2^{10}, the 2 standing for the binary system. Try it on your calculator. Multiply 2 x 2 x 2 x 2 x 2 x 2 x 2 x 2 x 2 x 2, or two multiplied by itself ten times. The answer is exactly 1,024. In the computer world, then, a kilobyte is exactly 1,024 bytes. Half a kilobyte is 512 bytes; the smallest "chunk" of data that can be written to a hard drive at one time in the MS-DOS operating system. A quarter of a kilobyte is 256 bytes.

As an aside, the correct abbreviation for a kilobyte in the computer world is KB, the uppercase K meaning 1,024. In the rest of the world, a lowercase k is used, meaning 1,000. Therefore, one writes 3150 kHz to describe a frequency, but 40 KB to describe the size of a computer file. Confusing? You bet. I will try to always use the correct abbreviations in future columns.

Here is an accurate and fairly extensive definition of terms used in the world of computers. Notice that you can jump a power of ten as you go up by squaring the original number.

UNIT and ABBREVIATION	COMPUTER WORLD DEFINITION	
1 byte (B)	8 bits	(bit stands for binary digit)
1 kilobyte (KB, 2^{10} bytes)		1,024 bytes
1 megabyte (MB, 2^{20} bytes)	(1 KB^2)	1,048,576 bytes
1 gigabyte (GB, 2^{30} bytes)	(1 MB^2)	1,073,741,824 bytes
1 terabyte (TB, 2^{40} bytes)	(1 GB^2)	1,099,511,627,776 bytes
1 petabyte (PB, 2^{50} bytes)	(1 TB^2)	1,125,899,906,842,624 bytes

Well now, suppose you are a hard drive manufacturer and you make a drive that will hold exactly 40 million (40,000,000) bytes after it is formatted and ready for use. Can you call it a 40 MB drive? No, not in computer world math. To call it a 40 MB drive, it would have to hold 41,943,040

bytes, or about 5% more than it actually does. A few years back, some (but not all) hard drive manufacturers decided not to use computer-style math to describe their devices (even though those devices do go in computers). That gave them an edge. Their drives appeared to be about 5% larger than they really were by using kb instead of KB. This is a bit like advertising an item for $9.95 instead of $10. A small deception , but a deception. For a while there, the deception was even bigger. Before very recently, many manufacturers were describing their drives using unformatted capacity, which made them appear much larger still.

When I ordered my computer from Computer Research and Applications (see No. 25 in this series, late 1995), I specified a 1.2 GB drive. Here is what I got, as specified by the various sources of information in the table:

SOURCE	SIZE REPORTED BY THE SOURCE	①BYTES	ERROR (bytes)	ERROR (%)
MS-DOS 7's CHKDSK.EXE	1,394,786,304 bytes	1,394,786,304	②none	②none
Vendor who sold me the drive	1.2 GB	1,288,490,189	106,296,115	8%
③Drive manufacturer (Samsung)	1.2 GB	1,288,490,189	106,296,115	8%
③Drive manufacturer (Samsung)	1208.5 MB	1,267,204,096	127,582,208	9%
SYSCHK.EXE (utility)	1330.2 MB	1,392,815,795	29,491	> 1%
CMOS	1179.3 MB	1,236,585,677	158,200,628	11%
MS-DOS 7 (DIR)	1207.6 MB	1,266,260,378	128,525,926	9%
Norton Utilities V2 for Win95	1.5 GB	1,610,612,736	215,826,432	15%
Windows 95	1.5 GB	1,610,612,736	215,826,432	15%

①Calculated by WB9RQR (using the calculator that comes with Windows 95).
②The error values shown assume that CHKDSK's value (1,394,786,304 bytes) is correct.
③Both values (1.2 GB and 1208.5 MB) were reported in the manual that came with the drive.

What a mess! Assuming CHKDSK.EXE is correct, the other values range from 9 to 15% off! Only the SYSCHK.EXE utility comes close, with an error of much less than 1%.

Which value is really correct? I haven't got the foggiest idea. To tell the truth, I don't even know how to find out! Moreover, the Samsung Corporation itself doesn't know the true value. Their numbers are only approximations because of a number of factors, outlined below. However, it doesn't foster too much confidence when they report about a 21-MB difference for the same model drive within the pages of the same drive manual.

When I gave this article to KB9QQE, Jeff, to read for errors and to generally keep me honest, he rightly pointed out that hard drive manufacturers have a lot of unknown variables to work with. The only completely accurate number they can truthfully report is the unformatted size based upon the controller they used during manufacture. Formatted capacity depends on (as Jeff calls it) "techie stuff" such as block size and File Allocation Table (FAT) size. Another factor is landing on an even number of blocks at the end of the disk (anything left over after the last even-numbered block is ignored and lost to the user). This factor, in turn, depends on the number of cylinders and heads. Still another is the number of bad spots on the disk, usually locked out by the manufacturer when the drive is tested after assembly. Moreover, many of these factors vary from operating system to operating system even for the same drive, and so the manufacturer cannot really come up with the exact capacity that the end user will end up with.

I suppose the exact capacity of the drive really doesn't matter. After all, when working with numbers as large as my drive listed above, a 1% error is a mind-boggling 14 megabytes! Remember back when 14 MB was 70% of your hard drive space? It was not many years ago. We've come a long, long way since then in drive capacity. Furthermore, you can now purchase a 1.2 GB drive for around $100. That works out to a buck for 12 MB, or about 8.3¢ per meg!

So what this leads to, as correctly pointed out by Jeff, is that different applications can read the disk in different ways. Some will read the true capacity while ignoring the formatting, while others

will report what DOS reports to it. Even others may have their own way of calculating capacity based on what they see as the number of heads * cylinders * blocks.

Well, those are some of the details. In summary, keep in mind when trying to justify and conceptualize numbers in the world of computers, that there are two general pitfalls or potential sources of error. The first, already mentioned, is that the math of computers is somewhat different than everyday arithmetic. The second is that the numbers we encounter today are so large that we can no longer think in terms of exact figures. Rather, we are forced to generalize, and conceptualize using “ballpark” estimates. For some of us, this represents a new way of thinking about numbers. Those of us who are most comfortable with exact numbers feel a bit uneasy when we are forced to accept an approximation, but that is the nature of a world in which we must consider astronomical quantities!

My thanks to KB9QQE for his efforts at keeping me honest and on-track. Happy computing!