For Tandy and MS-DOS Compatibles

Taking data transmission to new speeds

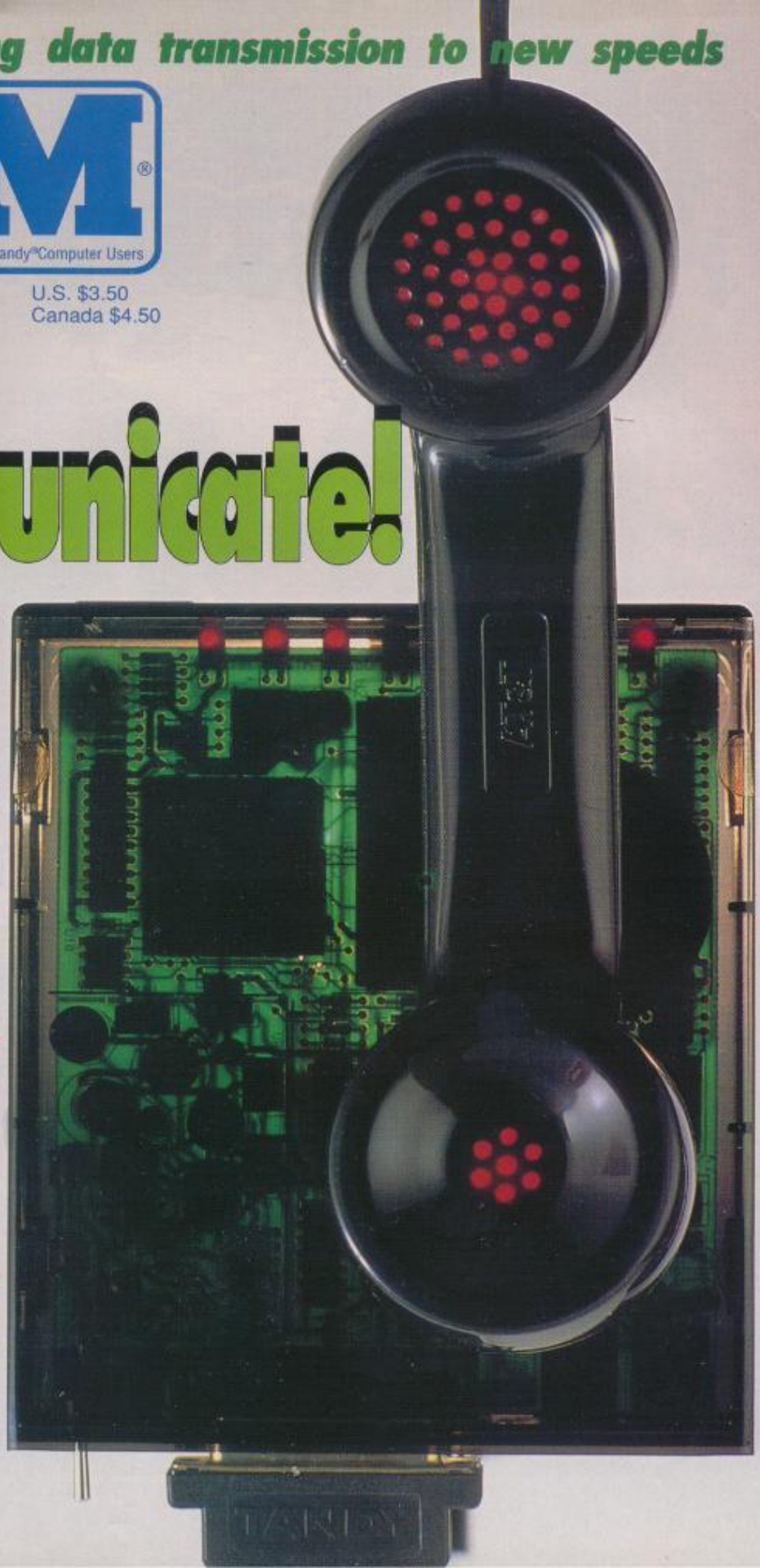
PCM®

The *Premier* Personal Computer Magazine for Tandy® Computer Users

Vol. IX No. 8
February 1992

U.S. $3.50
Canada $4.50

Communicate!



Fax machines and your PC

MLite: user interface for Delphi

Voltage converter increases laptop battery power

Software Shopper utilities for telecommunications

0 44254 00002 02

DOS 5.0: To Partition or Not to Partition?

by Gregory Shultz
Technical Editor

MS-DOS 5.0 is a great product, but the documentation can be a trifle misleading when it comes to partitioning hard disks in the 120- to 500-Meg range. The question is, when is partitioning wise rather than when is it required?

Most of us are familiar with the 32-Meg partition limit imposed by previous versions of DOS. We are told this limit has been eliminated in DOS 5.0, which can handle a hard drive with a single partition of up to 160 gigabytes or so. Meanwhile the documentation innocently states only that the size of an allocation unit or cluster is adjusted to match the size of the partition. No guidance is provided about the wisest partition size, so I did some experimentation.

When I partitioned my hard disk with partitions below 120 Meg, the allocation unit or cluster size was 2048 bytes. When I used partitions of over 150 Meg, the allocation unit was always 4096 bytes. The lower figure provides more efficient storage and access to more of the disk's capability. After some additional experiments, I have concluded that DOS 5.0 defaults to the smallest allocation unit that does not cause the total number of allocation units to exceed 65,535. Once you have partitioned the drive, DOS 5.0 uses an allocation-unit size based on the size of the partition rather than that of the hard drive.

Let's say you have a 100-Meg drive and are considering upgrading to a 200-Meg drive. If you create no more than one partition on either drive, conceivably you might get no extra storage at all from the larger drive. The reason: The default cluster size for a single-partition, 100-Meg drive is 2048 bytes; the default cluster for a single-partition, 200-Meg drive is 4096 bytes. If most of your files are small — files using only one cluster — you use just as many allocation units in both cases, but on the larger disk these are only half as full. If you have a large number of files in the multimegabyte range, on the other hand, partitions of 100 to 200 Meg or more make sense.

Being practical, very few of us use hard disks solely for thousands of very small, single-cluster files. The percentage of users truly injured as badly as my examples claim would be small. However, the DOS 5.0 manual and various other DOS 5.0 books should have considered this allocation-unit size difference and provided practical advice on when to partition in DOS 5.0.

After these experiments my system, running DOS 5.0, has partitions all set between 50 and 100 Meg. Unless you need to be outside that range, I'd suggest staying within these boundaries in order to get the most efficient use of your hard drive. What is your opinion on this subject?

H. Lawrence Elman
Port Jefferson, New York

Greg Shultz, like most married computer jocks, is in a constant struggle trying to divide his time equally between his spouse and his computer. Luckily he works with computers at the office because his wife seems to win the struggle on the home front. You can contact him on Delphi (username GREGSHULTZ) or on PC-Link (username PCM GREG). Of course you can contact him via U.S. mail at PCM.

This is an interesting and important subject many of our readers with hard drives of 200 Megs or more have perhaps never considered. I must admit, it is a subject I had given little thought until your letter provoked me into researching it more thoroughly. What I discovered confirms some of your basic assertions, but the subject is even more complex than you thought. To give you a better understanding of this intricate subject, let's look at the reasoning behind the use of various-sized allocation units. We start with how DOS stores and manages files on a disk.

The basic unit of storage on both floppy and hard disks is known as a *sector*. Each sector contains 512 bytes, and each type of disk contains numerous sectors. For example, a 360K floppy disk contains 708 sectors while a 212-Meg hard drive contains 414,618.

Since the sector is the basic storage unit, you might assume that DOS manages files on a disk by keeping track of each sector on the disk. However, the sheer volume of sectors found on a hard disk as large as 212 Meg makes this method of tracking files impractical. It is true that DOS manages files by keeping track of their locations on the disk; but rather than try to keep track of each of the sectors, DOS groups several sectors together and considers them a single unit, called a *cluster* or an *allocation unit*. The cluster then becomes the unit of storage for a file. Using this scheme, DOS reduces the amount of work it must do in order to manage all the files on a disk. This reduction in work means an increase in system performance.

The actual file-management operation is performed by the *file allocation table* (FAT). The FAT is one of three logical housekeeping areas established on a disk when it is formatted. During the format procedure each cluster on the disk is assigned a unique reference number. These reference numbers are then entered in the FAT. When a file is stored on the disk, its name is then linked to the reference number for the cluster or clusters it occupies. (This linkage is performed by what is known as the *directory entry*, also one of the housekeeping areas established during the format procedure but beyond the scope of this article.) As part of its responsibilities, the FAT also keeps track of which sectors are damaged as well as the amount of disk space available. Before every read or write operation, DOS refers to the FAT to determine where a file is for a read and what clusters are available for a write. The FAT is basically a detailed road map of a disk.

Considering that the FAT is extremely crucial to DOS, the format procedure actually places two copies of the FAT on a disk. If anything should happen to the first copy of the FAT, DOS can still locate files by referring to the second copy.

Since the FAT occupies double space on the disk, space is taken away from the total amount available for files. To compensate for this, DOS uses different-sized FATs for different-sized disk partitions.

On a large hard-disk partition, each entry in the FAT is 16 bits (2 bytes) long. On a small hard-disk partition, each entry is 12 bits (1½ bytes) long. This alteration in entry size effectively reduces the total FAT size by about 25 percent.

The size of the FAT has a direct relation to the size of the clusters used on a disk. A FAT with 12-bit entries can maintain a maximum of 4079 clusters, and a FAT with 16-bit entries can maintain a maximum of 65,519 clusters. In order to stay within these set limits, DOS adjusts the number of sectors per cluster. However, the number of sectors per cluster is always a power of two.

In order to confirm these findings, I set up a 212-Meg hard drive with a number of partition sizes. If the drive is set up with a 15-Meg partition, this partition uses a 12-bit FAT and has a total of 30,970 sectors. Dividing 30,970 by 4079 yields a result of 7.59. Implementing the power-of-two rule, we come up with eight sectors per cluster. If the drive is set up with a 127-Meg partition, this partition uses a 16-bit FAT and has a total of 260,186 sectors. Dividing 260,186 by 65,519 yields a result of 3.97 (four sectors per cluster with the power-of-two rule). If the drive is set up with a single partition, it uses a 16-bit FAT and has 414,618 sectors or eight sectors per cluster.

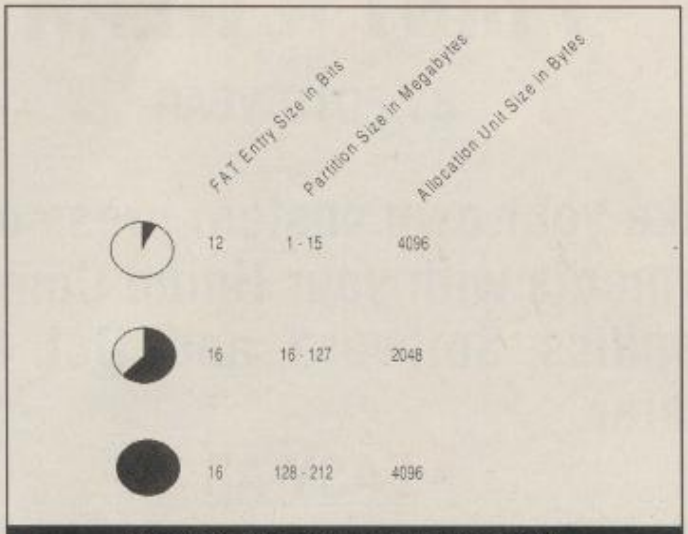


Figure 1: Allocation-Unit Size-Break Points for a 212-Meg Drive Partitioned and Formatted With DOS 5.0

The chart in Figure 1 shows the actual break points where the size of the allocation unit changes with respect to the size of the partition when the 212-Meg drive is partitioned and formatted with DOS 5.0.

Now that we understand *how* cluster sizes are determined, let's look at the *why*. Cluster size reflects a compromise made in the size of the FAT. If the 212-Meg hard drive were set up with a single partition using four-sector clusters rather than eight-, the size of the FAT would become excessively large. The larger the FAT, the more disk accesses are required and the more time computations take. In order to skirt these delays, DOS selects a FAT size based on a compromise between the amount of disk space wasted and the total disk space available and operation speed.

To verify these findings, a PCM colleague and I used a combination of the *Norton Utilities* Disk Edit and Safe Format utilities to set up a hard disk with the cluster size equal to that of the sector size — 512 bytes. While the result was a hard disk with the most efficient use of the available storage space possible, there was a price to pay. The hard disk-access time went far below acceptable standards, and we were forced to wait an excruciating length of time when running a program or copying files. The reason was that we had quadrupled the number of clusters DOS must manage and dramatically increased the size of the FAT.

With the results of that experiment in mind, we can see that the compromise between wasted disk space and operation speed is a valid one. Anyone who has used both an 80386 with a fast hard disk and an 8088 with only floppy drives can appreciate the difference in disk-access speed we experienced. For the sake of argument, though, let's take a graphic look at the amount of space wasted with large clusters and small files. Figure 2 shows the amount of space wasted when a 50-byte batch file is allocated to an entire cluster. The wasted space is paltry in the one-sector cluster and startling when the file is allocated to an eight-sector cluster — almost the entire 4K is wasted space.

Before you become overly concerned about this waste of space, remember that DOS is making a compromise between wasted space and operation speed. Yes, it is true that if you had hundreds of small files, the amount of wasted disk space could run

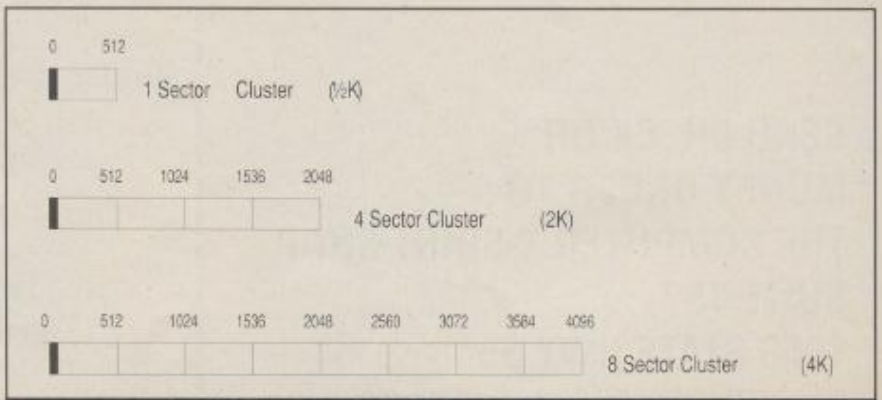


Figure 2: Amount of Space Wasted When a 50-Byte Batch File Is Allocated to an Entire Cluster

as high as 25 percent or more. But the opposite is also true: Large files benefit from large cluster sizes.

So how can you get the most from your large hard drive with the effective use of partitions? Deciding when and how large or small to partition your large hard disk can be determined by the types of programs you use and the ways you use them. A determining factor is the size and number of the data files your programs use. If you use accounting or database programs, you usually have only a couple of data files that range in size from 100K to 10 Meg or more. If you primarily use word-processing programs, more than likely you have many small files ranging in size from 2K to 10K or more.

Accounting and database programs with large data files effectively use storage space and operate marginally faster on a partition with large cluster sizes of 4K. On the other hand, word-processing programs are more efficient on a partition with 2K clusters. However, you might want to consider such other considerations as whether working with one or several partitions would be more convenient in your specific situation.

Another factor coming into play here is that you may have a varied combination of both large and small data files on your hard disk. If that is the case, the amount of space wasted with large cluster sizes is marginal. It should now be apparent that there really is no right or wrong partition strategy and that every strategy has both advantages and disadvantages. While keeping your partitions within the 50- to 100-Meg range, as you stated in your letter, works well for you, it may not be the best strategy for others. No matter what strategy you follow, remember that regular disk optimization reduces the amount of file fragmentation and increases the performance of your hard disk.

Disk Size	Bytes in each Allocation Unit	Number of Allocation Units on Disk	Total Disk Space in Bytes
360K	1024	354	362,496
720K	1024	713	730,112
1.2MB	512	2371	1,213,952
1.44MB	512	2847	1,457,664
2.88MB	1024	2863	2,931,712

Figure 3: Cluster Sizes Used on Floppy Disks

While working on this article, I began to wonder about the cluster sizes used on floppy disks. My findings are shown in Figure 3. While all floppy disks use 12-bit FATs, the size of the cluster in relation to the size of the disk was not what I had expected.

PCM