GIS HARDWARE

Hardware components

Computer hardware is one of the most difficult areas to maintain currency. It is a technology that is constantly changing. This change brings to the user more computing power, better access to data, superior output options, and lower costs. Many recall how we agonized over the first computer purchase. If I buy now I know it will be out of date in 6 months. Maybe I should wait until it becomes cheaper. But the cycle never ends. Finally, we broke down and purchased what we felt was the best system that met our needs at the time. But now we need more. More power, more storage, more auxiliary devices. Because computer hardware is changing so rapidly, it should not be the overriding asset that is used when evaluating what kind of GIS we should develop.

There are a myriad of different types of computers in use today. These include: supercomputers, mainframe computers, minicomputers, workstations, and personal computers. This latter type also has different kinds such as the desktop, notebook and laptop, palmtop, and personal data assistant (PDA). The important criterion in selecting which computer should be purchased, besides the obvious issue of cost, is its capability to handle the applications required by the user.

Within a GIS the main computer components can be identified as the processing unit and peripheral devices. The central processing unit is where the computer operations are performed. Peripheral devices are those tools we use to input data, store the data, and output the results of our analysis.

The central processing unit (CPU) is where the bulk of the work is done within the computer system. It controls access to the computer and helps monitor security, compiles the programs written in a high-level language into machine readable form, controls the communications between the peripheral devices, performs system accounting and diagnostics, and executes the program commands either in an interactive or batch processing mode.

CPUs can be classified as either centralized processors or distributed processing units. In a centralized processing unit there is one host processor that performs all of the processing functions within the computer system. Distributed processing systems have several processing units which perform some of the work instead of having the central processor unit handle all of these tasks. An example of a distributed processing unit is shown in figure 1.

Within the CPU lies the arithmetic and logic unit (ALU) and the control unit. These contain the electronic circuitry designed to process instructions. Information is stored in the ALU in registers. These are temporary storage areas which contain not only the computer instructions but also data. These items are stored in binary digits. A binary digit is a value stored in a base 2. As an example, our arithmetic is based on a base 10.

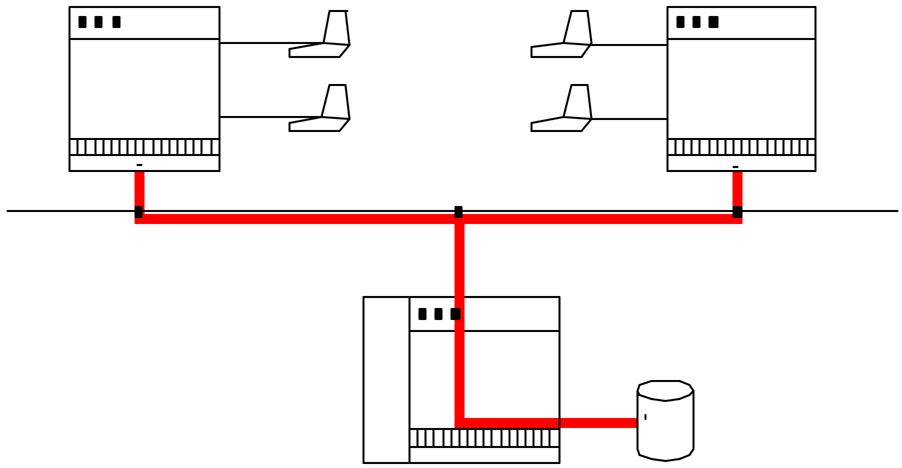

Figure 1. Distributed computer system

To convert a base 2 to a base 10 is not that difficult. Lets look at an example for base 10 as shown in figure 2. Lets look at the number 4,376. Working with a base involves taking that base to a power. For base 10, take 10 to the power 0, 1, 2, 3, etc. This is shown in the first line of the figure. The result of that arithmetic is shown in the second line. Thus, 10 to the power 3 is 1,000, 10 to the power 2 is 100, etc. Note that any number taken to the zero power is one. Now, simply count how many thousands there are in the number, then how many hundreds, tens and digits. Of course, with this simple example, the number is 4 thousands, 3 hundreds, 7 tens and 6 unit digits. The answer is, as we would expect, 4,376. Now, I know we do not count this way but it shows us a method of how we can convert from binary to our base 10 number and vice versa.

4376

10^3	10^2	10^1	10^0	
1,000	100	10	1	
4	3	7	6	= 4,376

Figure 2. Base 10 numbering scheme.

Figure 3 shows the number 4,396. This time we are going to represent this in a binary number (base 2). The computer stores all values in a binary form. As before, take the base to the power. Thus, 2 to the 12th power is 4096, 2 to the 11th power is 2048, 2 to the 10th power is 1024, etc. As before, the base 2 to the power is given in the first line while the arithmetic value is shown in the second line. The number we want to convert is 4,396. We can see that there is 1 4096 within 4396. Therefore, place a 1 in the third row under 2 to the 12th power. Subtract 4096 from 4396 and this leaves us with 300. Two to the 11th power is 2048. But we only have 300 left from our number. Thus, there are no 2

```
  4396        12
 -4096       - 8
 -----       ---
   300         4
 - 256       - 4
 -----       ---
    44         0
 -  32
 -----
    12
```

2^{12}	2^{11}	2^{10}	2^9	2^8	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
4096	2048	1024	512	256	128	64	32	16	8	4	2	1
1	0	0	0	1	0	0	1	0	1	1	0	0

Figure 3. Converting base 10 number to binary form.

to the 11th power numbers left in our original value so place a zero in the third row under the 11th power. As we can see, there are also no values for the power 10 or 9 left so zeros are placed in the appropriate column on the third row. Two to the power 8 is 256. This is less than our remainder value of 300 so a one is placed under that column and taking 256 from 300 gives us a remainder of 44. Looking at the values remaining in row 2, we see that 2 to the 7th and 6th powers are both larger than 44 thus place a zero in those columns in row 3. Continue the arithmetic. The result is that the number 4,396 can be represented in a binary form as 1000100101100.

Going from binary to base 10 is easier. Here, set up the table as shown in figure 3. This time, we can fill it in completely. Then simply add the values in row 2 that have a one in row 1. In our example,

```
 4096
+256
+  32
+   8
+   4
-----
 4396
```

The computer main memory unit is designed to provide high speed access to the data and program statements during execution of the program. This main memory is called core memory, or more commonly random access memory (RAM). When we use our computers we probably have experienced that the more RAM the faster the program executes. RAM holds the programs and data when the computer is in operation. The more memory, the more programs one can run at the same time. On the other hand, if we try to extract data from an auxiliary storage device, such as a disk, it takes longer to retrieve the required data thereby increasing the time necessary to execute the command. A lot of computer processors utilize cache memory. This is a temporary storage area where frequently accessed data and programs are stored. These units are designed to maintain a high bit rate between the CPU and the memory. Read only memory (ROM) stores the computer's control information. Therefore it will save this information at all times. One cannot write to ROM.

Data is moved between computer components on high-speed paths called buses. Data travels over buses in a synchronous manner. A sequence of bits, representing one computer word or part of a computer word, is transmitted simultaneously. The information-carrying capacity of a bus depends on its width, or the number of bits that can be transferred simultaneously. Buses vary in width depending on the design and size of the computer. Microcomputer busses typically are 8, 16, or 32 bits wide.

Auxiliary storage devices

The main memory discussed above is not sufficient to meet our computer processing needs. Thus, auxiliary storage devices or mass storage devices are used to store large volumes of data that are accessible by the processor. These are peripheral devices and can be categorizes as off-line or direct access. The most popular off-line storage medium is the tape drive, although this is quickly changing today. It is off-line because the tape must be mounted physically on the drive before the data can be read. Direct access storage devices allow users to store and retrieve data directly without loading it form another source. The most popular is the disk drive, either magnetic or optical.

Tape drives are usually available in reel-to-reel or cartridge format. The main usefulness of tape drives are for archival storage and transport of large volumes of data or software. In addition, they are often used to perform routine backups of data and software on disk drives to protect against loss in the case of disk failure. Tape drives are also a relatively low-cost method to store data off-line. The main problem with tape drives are that the data are stored in a sequential manner therefore it has to be read from the beginning. As an analogy, look at the tape cassettes that we may use for music. If we want to go to a particular song, the tape has to be advanced to that starting point. This is different than selecting songs from compact disks. Most tape formats do not permit updating of the data.

There are two classes of reel-to-reel tape drives: streaming or start/stop. These classes are based on how the data is written to the tape. In the streaming method, data is written to the tape without any physical gaps between files or records and without marks on the tape that delimit files or blocks. In other words, there is no directory structure. This is very efficient and fast for routine backups but is difficult to find where a particular data record exists. The start/stop method puts a physical delimitation between files and records. This is slower but more efficient if a particular record is to be retrieved for processing because individual records can be retrieved. Even with the emergence of newer technologies, research in applying tape storage continues. From a historical perspective, the adoption of a new technology must be 10 times cheaper per bit for the same performance or have 10 times the performance for the same cost.

While newer technologies exist, tapes will remain, for the foreseeable future, an important storage media. Part of this is because of the historical archival data that is already stored in tape format. Additionally, it will remain important because:

"1) Magnetic tape storage will remain significantly less expensive than hard disk storage. While lower cost per storage will be a continuing trend for all technologies, it is expected that the cost ratios will remain intact.

2) The volumetric density of tape storage relative to other storage technologies will, for fundamental reasons, always be a large ratio.

3) With each advance in technology, the demand for data storage increases. Improved storage devices enable new applications that previously were not economical and thus, in turn, leads to increased demand for additional storage.

4) Software-managed automated removable media storage libraries continue to evolve and will be common for all applications. With this in place, optimally-designed tape storage devices will provide a continuum of storage characteristics along with semiconductor memory, hard disk and optical disk storage.

5) Although significant advances in electronic data transfer communication networks can be expected in the next decade, because of band width limitations and telecommunication costs, data interchange via physical transport of removable media volumes will remain the most economical procedure for many applications"[1]

How efficient are tapes? The National Storage Industry Consortium (NSIC) arrived at a bit density of 6 Gb/in^2. By increasing the number of layers, storage of up to 12 TB can be realized in a 3480[2] form factor cartridge. Current limitations to track density include: substrate instability, tracking, mechanical misalignments, mechanical tolerances, media defects, and media intrinsic SNR[3]. If developments continue, storage costs of 1¢/GB are possible.

Magnetic disk are probably the most used system of storing data today. While it does not have the storage capacity of optical disks and it is slower than the solid state memory in the computer, it offers the best combination of data access speed and capacity. Magnetic disk stored data by use of magnetized fields on the disk surface. An electromagnetic head subsequently reads the data. Data are stored in tracks and sectors on a rotating disk. It is a random access medium in that the reader can jump form track to track to access the data. The downside is that this media is subject to corruption by stray magnetic fields that can affect the direction of the magnetized fields, and by head crashes, which involve the actual collapse of the read/write head onto the magnetic surface. In addition, the diskettes can be bent or lost. The typical disk holds up to 1.44 million bytes of data. Zip and jazz drives are higher capacity storage devices that are almost as fast as some hard disk drives. A single diskette can hold up to 1 GB of data.

There are three main disk drive interfaces.[4] The Enhanced Integrated Drive Electronic (EIDE) interface that is used with many PCs. This is a low cost approach costing about $7/GB. The Small Computer System Interface (SCSI) is primarily found in the workstation and server systems. The costs are about 2-3 times that of the EIDE interface. The Universal Serial Bus (USB) and the IEEE 1394 Firewall interfaces are being introduced for the low-end market. At the high end, for supercomputers and high-performance workstations, the fibre-channel drive is generally used.

Optical disk technology is advancing and offers more storage capability than is found in magnetic disks. WORM (Write Once Read Many) disks are created by burning small impressions onto the disk using high-powered lasers. The lasers create a pattern of pits or bubbles, representing the stored data, on a highly reflective surface sandwiched between two layers of plastic. A low-powered laser is used to read the binary data by measuring the reflection of the laser light beam off the reflective surface (see figure 4). Since lasers can be used to focus on very small areas of the disk, this gives this medium the capability of high storage. While the performance is not as good as magnetic disks, its storage capability is very high. It is good for storing data which is read frequently, but not updated, such as video images. This media is ideal in delivery of the final product from a large project.

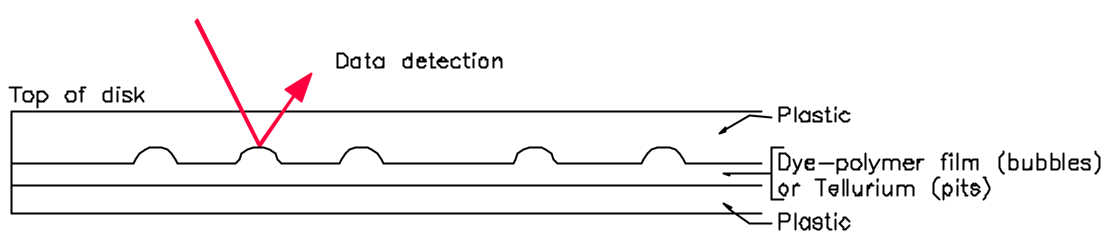

Figure 4. WORM Optical disk.

There is a new technology that is beginning to find more use in personal computer systems and this is re-writable optical disks, or sometimes called erasable optical disks. Two technologies are most often used: magneto-optical and phase change. The magneto-optical utilizes a cross between magnetic and optical storage. Magnetized fields are used to store the data while lasers are employed to read that data. It is beyond the scope of this class to delve into the theory of these technologies. Table 9-1 gives a comparison between different mass media. Although dated, it does give a measure of comparison from the time this paper was published.[5]

Media	Date Introduced	Primary Storage	Secondary Storage	Backup	Distribution	Maximum Capacity	Cost per Mb.
Hard Disk	1974	Excellent	Inappropriate	Good	Inappropriate	1.2 Gb	$4.50-10.00
CD-ROM	1985	Inappropriate	Excellent	Inappropriate	Excellent	600 Mb	<$0.01
WORM	1985	Poor	Good	Poor	Poor	2.4 Gb	$0.10-0.25
Mag-Opt	1988	Poor	Excellent	Good	Good	1 Gb	$0.13-0.34

Table 1. Evaluation of peripheral devices.

Newer technologies include the digital versatile disk (DVD). These are generally read-only devices that are becoming common in the storage of high-quality movies. There are writable DVD devices in development.

A very exciting computer development is the introduction of the personal digital assistant (PDA). These are basically portable computers. They employ very friendly interfaces with the programs. Many include other applications like address and appointment books, web browsers, and smaller versions of Microsoft Office applications like MS Word and Excel. They can be either directly connected to regular computers for data transfer or connected using wireless communications. There are many who feel that the trend will be in Java-based systems. These systems are having a significant impact in the GIS arena. Some of the current systems are described by L. Graham[6] and include OnSite by Autodesk, ArcPad 5.0 by ESRI,

Figure 6a. Example of the OnSite notebook.

Figure 5b. Example of ESRI's ArcPad.

FieldSmart by MapFrame Corp., iMap by Sokkia, and SoloCE and Ranger by Tripod Data Systems. An example of OnSite and ArcPad are shown in figures 5a and 5b.

One input device actually involved collecting the data in the field and this is surveying instrumentation. Also included in this category are global positioning system (GPS) receivers even though not all of them can be classified as survey grade receivers. The surveyor generally collects data using electronic instrumentation called total stations. These instruments will record the angles and distances measured in the field. This information is downloaded into the computer and using coordinate geometry, coordinates of the features can be determined. Usually the output is a map for the client showing all of the important features on the site. Total stations require setting up the instrument on a tripod over survey monuments. The process of setting the instrument up and measuring the angles and distances requires a specialized technician who understands something of the survey process. Using GPS receivers, features can also be recorded within the receiver and downloaded onto a computer. Since the receiver stores the coordinates, processing is minimized in the computer. In both instances, attribute information can also be collected in the field. The accuracy of survey data is probably the highest of all the data input methods, on the order of a couple of centimeters in relative positioning. But it is a labor intensive and time consuming process. GPS data is very economical and have been used to locate features in a number of GIS's. The accuracy of receivers depends on the instrument being used. The technology is capable of measuring very accurately, down to the millimeter range and lower. But, this instrumentation is only used for the most demanding type of work such as crustal motion. Most receivers used in GIS data collection are sub-meter instruments. They provide high productivity and maintain some semblance of accuracy. While these receivers report accuracies below a meter, these are based on statistical results at about the 68% level. This means that some values can be in error by more than a meter. On the other hand, practical results have shown that most of the time the receivers are much better than the reported accuracies made by the manufacturers.

Photogrammetric workstations are one of the workhorses in GIS data collection. We have reviewed the concepts of photogrammetry in the first two lessons and, hopefully, can see that this technology has a lot to offer those building a GIS. Right now, the mapping process is still a labor intensive operation. Yet, it is still very economical and offers a high degree of accuracy over a large area. There are promising developments on the horizon. There is a lot of research and development into extracting features from imagery. In some instances this is already in operational mode. But more needs to be done. Nonetheless, many geographic information systems have started with a good base map derived from photogrammetric methods. Also, there is an increasing utilization of digital orthophotos in these systems. With new satellites offering meter-level resolution and other types of developments, photogrammetric workstations will probably grow in importance as a primary data collector for GIS.

Digitizing an existing map is another very popular means of entering data into a GIS. There are two basic methods of digitizing: stream digitizing and point digitizing (figure 6). When digitizing, the map is placed on a digitizing tablet/table. The operator uses a cursor which consists of a cross-hair and a series of buttons. The cross-hair is placed over the point and the appropriate button is depressed. This records a set of coordinates based on the digitizer coordinate system. In a point mode, each important point is manually digitized by the operator. With stream digitizing, the operator begins by digitizing the first point and then follows the feature with the cursor. The digitizer will automatically record the coordinates of the feature either based on distance traveled or based on time. As an example of the latter, the digitizer may record a feature every 3 seconds. With distance as the criteria, the digitizer may automatically record every time the digitizer moves a half inch from the last point collected.

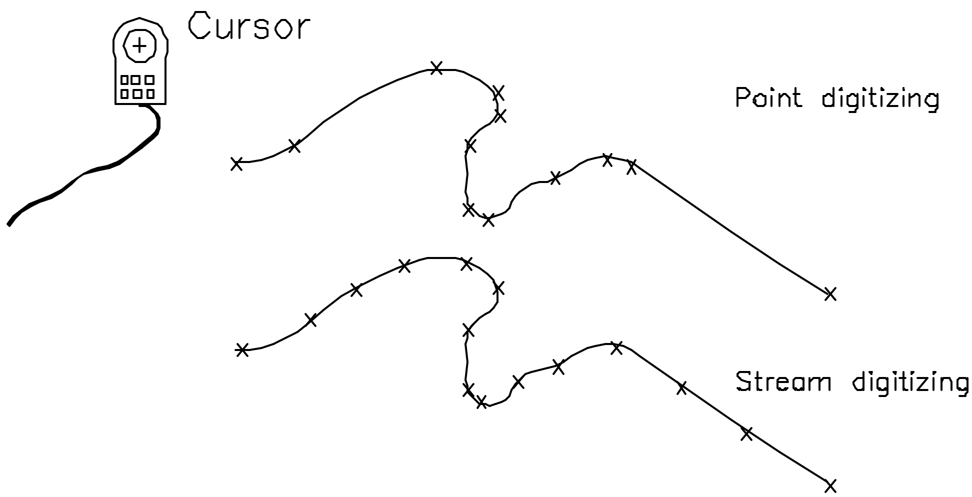

Figure 6. Point and stream digitizing.

The advantage of point digitizing is that all of the important points are measured whereas with stream digitizing, some might be missed since the operator is only passively collecting point data. For features that are pretty straight, only the beginning and ending points need to be collected, although it is good practice to collect another point in the middle. Therefore, less points are collected. The disadvantage is that the operator needs to make the decision as to whether a point is important or not. No two operators will identify the same points all of the time. The basic philosophy in favor of stream digitizing is that the operator will probably trace the feature faster when the feature is straight and will be required to slow down when they are complex features, thereby collecting more data and ensuring the feature is adequately identified. After the digitizing is completed, the coordinates need to be transformed to real-world coordinates using some type of transformation, such as an affine transformation.

Scanning also offers an economical means of collecting data for a GIS. Scanners can be classified as flat-bed scanners or drum scanners. In either case, the scanner senses variations in reflected light from the map surface. Any marks on the hard copy, including stains or wrinkles, that exceed a given threshold of contrast between the mark and the background media will be encoded. The result is a raster data set containing showing the features on the map. Sometimes this raster data is converted to vector form for a particular GIS application.

Flat-bed scanners utilize a table where the document is laid. Then a scanning head is moved along both the x and y directions over the document. In a drum scanner, the document is mounted on a cylinder that rotates while the scan head moves horizontally across the cylinder. The disadvantage of the flat-bed plotter is that it takes up a lot of space and it is generally slower to perform the scanning. The advantage is that it is more accurate. The drum scanner may slip as the drum rotates thereby creating an error. It is much faster though and the error rate is small with today's technology. It also occupies less space which may be an issue.

The last type of scanner is a line scanner. Here, the operator places the cursor on the feature and the scanner will automatically follow the line. This type of scanner requires a lot of operator interaction because there may be problems when lines intersect (the scanner may not know where to go). The line scanner is not often used today.

Conclusion

There are a lot of other issues dealing with computer hardware that are not included in this lesson. Again, it is important to realize that hardware is constantly changing and it is important for someone within the organization to keep track of these developments as the GIS matures. The one given is that the hardware that is used at the beginning of a GIS development process will not be the same in five years (or less). It is also important to recognize that with data input, there are numerous processing steps that may be needed. For example, transforming the data to the coordinate system used in the GIS. Maintenance of the graphic record within the GIS will require decisions as to what needs to be changed to fit new data to the old.

Additional Readings

White paper entitled "Mobile Computing" by MapFrame. Available at http://www.mapframe.com/wp.htm. There are two white papers listed. Read the first one on mobile computing.

"A Guide to the Central Processing Unit and Its Architecture" by Alex Pascoe, available at http://sis.bris.ac.uk/~ap7371/assign.html.

The PC Technology Guide – Scanners, available at http://www.pctechguide.com/18scanners.htm.

Distributed Computing, available at http://www.ncgia.ucsb.edu/~good/newwhitepaper.html.

"Field Note: Extending a GIS into the Field", by N. Ryan, J. Pascoe, and D. Morse, available at:

http://www.cs.ukc.ac.uk/research/infosys/mobicomp/Fieldwork/Papers/CAA98/MobileGIS.html

"ArcPad: Mobile Mapping and GIS", and ESRI White Paper, available at:
http://www.esri.com/library/whitepapers/arcpad_lit.html

Endnotes

[1] Gniewek, J. and Vogel, S. "Influence of Technology on Magnetic Tape Storage Device Characteristics", http://esdis-it.gsfc.nasa.gov/MSST/conf1995/B1_4.html, accessed 07/12/00.
[2] The IBM-3480 was the first cartridge tape drive recording system.
[3] SNR is signal to noise ratio.
[4] Harikaran, P. and Kobler, B., 2000. "Data Storage Choice Abound", GEOWorld, 13(7):36-40.
[5] Rivamonte, L., 1992. "A Multimedia Primer for GIS", URISA Proceedings
[6] Graham, L., 2000. "Life in the Fast Lane", GEOWorld, 13(7):30-35.