Computer Basics

Computers manipulate digital data. Every sort of information a computer deals with — text, sound, graphics, to name a few — is represented within the computer as a stream of numbers.

Computers represent these numbers using the binary system. While modern human number systems are based on 10 (we have 10 digits, 0 through 9) computers use a **base 2** system. In base 2, there are only 2 digits: 1 and 0. These binary digits are called **bits**. Everything a computer does involves moving, manipulating, calculating and storing numbers using the base 2 system. While computers may translate these base 2 numbers into other sorts of information that is easier for humans to work with, a computer's only true ability is the manipulation of these base 2 numbers.

Computers are built of a collection of components or subsystems, each performing a different task. Three critical ones are **processing, short term storage** and **long term storage.**

Most processing is handled by the computer's CPU, its Central Processing Unit. The CPU can be likened to the engine of the computer. This is where the data (remember, we're talking 1s and 0s here) is manipulated. A CPU can only perform a few basic arithmetic operations, such as adding, subtratacting and shifting groups of numbers around. However, since modern computers can perform billions of these basic operations per second, large and complex tasks can be performed on computers by breaking the tasks down into billions of very simple arithmetical operations, the kind of operation a computer's CPU is designed to do.

While the CPU manipulates data, the data also needs to be stored somewhere so that the CPU can access the raw data and store the results of its calculations. Most computer systems have two types of storage: long-term storage and short-term storage.

The CPU's ability to deal with billions of basic arithmetical operations per second means it needs to be fed data at a very high rate. The data that supplies the CPU's appetite is held in short-term storage. On most computer systems, this takes the form of Random Access Memory chips, or RAM. The data the CPU needs to access is held in RAM. It is then moved to the processor along a path called a **bus** (think of a bus as a 32 lane highway for binary digits). The CPU makes the necessary calculation on a small number of bits,

and the result of the calculation is shuttled along the bus back to the RAM memory, where it is held in storage. The cycle is then repeated.

RAM memory allows for very fast access to data, but it has two inherent disadvantages. One it that it is relatively expensive. A greater disadvantage is that most types of RAM memory can only store data as long as the memory chips are supplied with electrical power. Cut off the electricity, and the data vanishes instantly. These two drawbacks are the reasons we have a second sort of storage, long-term storage.

In most computer systems, long-term storage takes the form of disk drives. Disk storage is much slower than RAM, but it is also much cheaper. More importantly, and unlike RAM, information stored on a disk is preserved when the supply of electricity is shut off. While not fast enough to directly feed the CPU's ravenous appetite for data, disk storage is used to store information not immediately needed by the computer's CPU, and to store all information when the power is shut off.

Computers generally move information from the disk to RAM as the information is needed, shuttling it back and forth between RAM and CPU as the information is manipulated, and finally storing the information back in long-term storage (a hard drive, Zip disk or similar system) when the processed information needs to be saved. The rate at which data can be shuttled from disk to RAM to CPU and back is often a more important factor when gauging the overall speed of a computer then the speed of the CPU alone.

A simplified computer