February 5, 2011

Binary Tree Computer Systems

By Thomas O. Jones

A Binary Tree Computer System means a computer system of nodes connected in a binary tree configuration. A binary tree configuration means an arrangement of nodes where each node has a single parent and two children nodes, except the root node, which has no parent, and the leaf nodes, which have no children.

In order to visualize a Binary Tree Computer System, see Figure 2 below from US Patent 4,860,201, which is owned by Fifth Generation Computer Corporation.
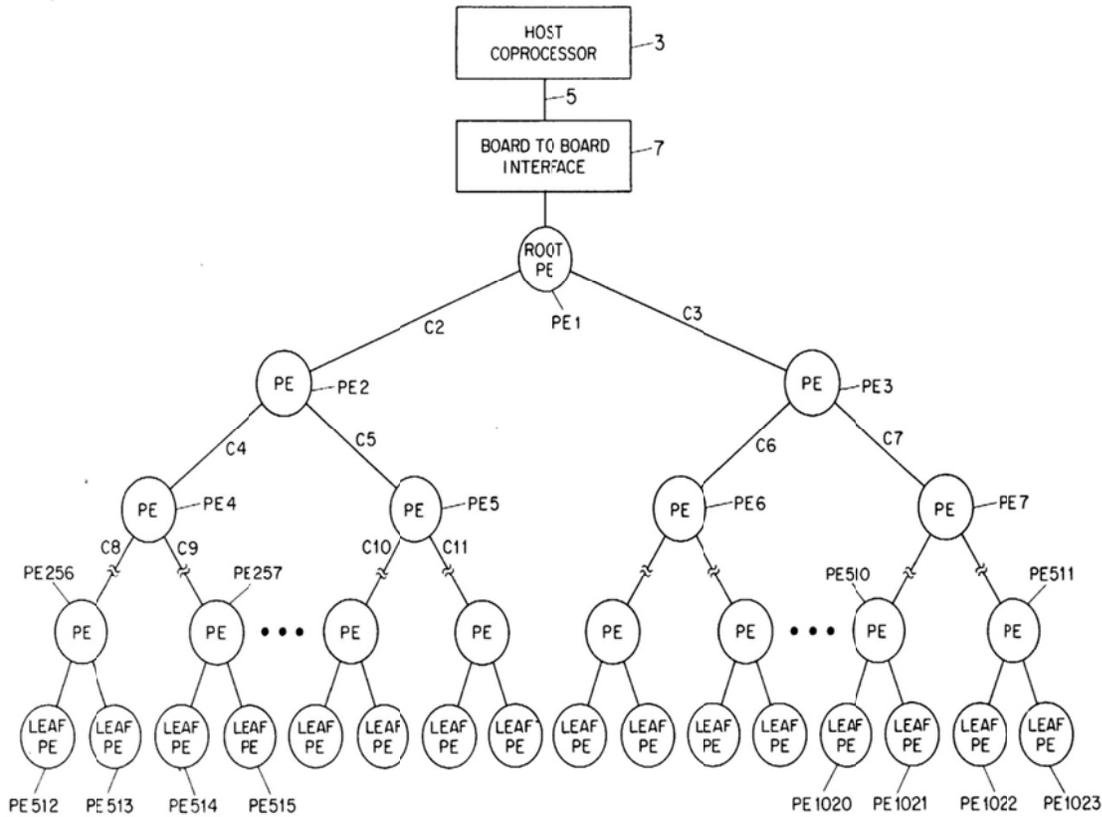

FIG. 2

If you look closely at the diagram, you will see that it illustrates a tree of 1023 nodes with some nodes in the middle of the tree eliminated for the purpose of simplicity. You can see that the binary tree computer system is scalable to any size that might be required for an application.

Many Improvements

Since the early days of binary tree computer systems many improvements have been implemented in order to eliminate the bottleneck at the root of the tree. One such improvement was the invention of virtual channels by William Dally, sponsored by DARPA. Processor and memory speeds have also been dramatically increased.

Examples of Commercial Binary Tree Computer Systems

In order to illustrate the power of binary tree computer systems, we have selected two examples of successful commercial systems whose architecture includes a form of a binary tree computer system.

AT&T BT-100 System

On October 13, 1986, AT&T and Fifth Generation Computer Corporation, AT&T's sole subcontractor on the project, were awarded a contract from the Defense Advanced Research Projects Agency (DARPA) "to develop prototypes of a computer that can recognize speech and images and do other complex pattern-matching tasks in a fraction of the time of today's best computers."

The AT&T Press Release included the following additional technical information:

> "To demonstrate the power of its parallel-processing architecture, the AT&T machines will run speech-processing algorithms currently under 'development at AT&T \Bell Laboratories.

> "Each processing node in the machines will consist of a microprocessor, a digital signal processor, and some memory and input-output hardware. The Dado architecture links processing nodes in a so-called binary tree architecture in which each node communicates only to a parent node" and two "descendants," thus forming a "family tree" of related processors."

The prototypes that the two companies developed were named the AT&T BT-100 Systems which is depicted in the Diagram below.

Figure 2: Parallel Processor for Speech Recognition
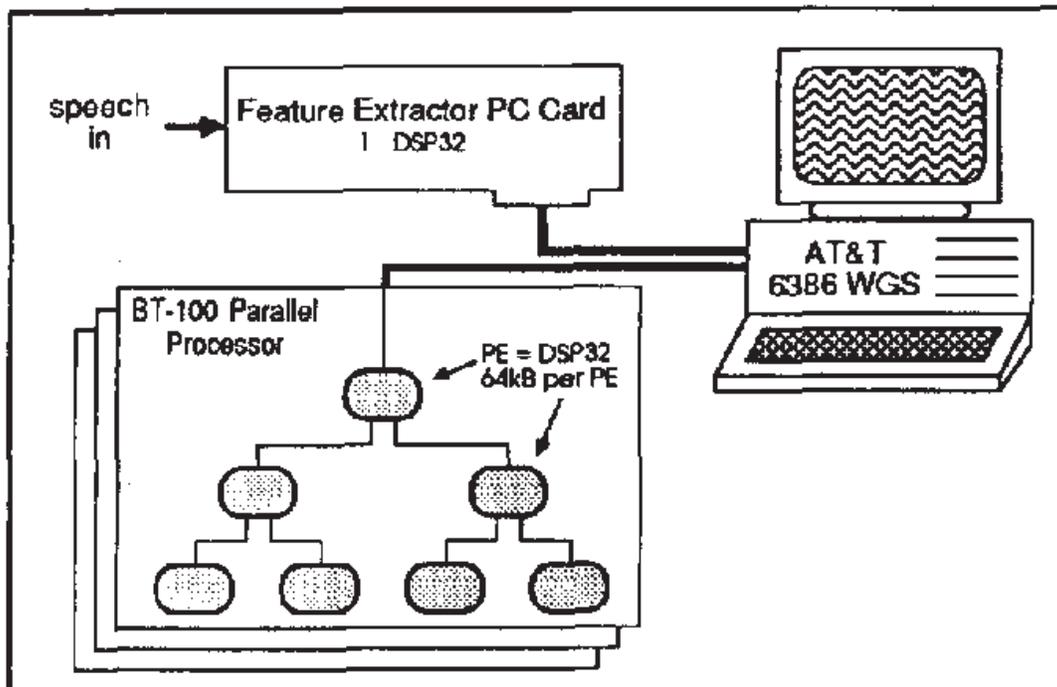
On Tuesday, March 3, 1992 announced "it will deploy voice-recognition technology nationwide to automate many long-distance calls now handled by operators."

As planning and deployment took place, AT&T later announced that "when completed, the phase-out (of jobs) is expected to save the company at least $900 million a year as new technology is deployed and management layers are reduced."

IBM Blue Gene Systems

Inside every IBM Blue Gene System is an embedded Binary Tree Computer System which IBM calls, the "Collective Network."
The IBM Blue Gene System was jointly developed with Lawrence Livermore National Laboratories (LLNL) under a government contract.

In the April 2005 article, entitled "Into the Wild Blue Yonder with BlueGene/L," Mark Seager, is quoted (at the bottom of page 2):

"Another difference between BlueGene/L and other platforms is that it has not one but three interconnects for applications: a 3D torus network, a binary-tree (combining and

broadcasting) network, and a barrier network."

"BlueGene/L's binary-tree network is useful for low-latency global operations that share data and synchronize programs. This interconnect determines how a highly parallel computer program "talks" to all the nodes quickly and efficiently. "Different ways exist to deliver a message to a large number of nodes," says Seager. "In a binary-tree network, one node talks to two neighbors, those two talk to two of their neighbors, and so on. Getting the message out to 65,536 nodes is a very efficient process, taking only 16 tree operations, or hops."

"The binary tree can operate in broadcast mode to replicate information across the machine or in combining mode to gather data distributed across the machine into a single location. Both broadcast and combining modes are used in operations performed millions of times in real scientific applications. In BlueGene/L, the binary-tree interconnect is implemented in the hardware rather than in the software, making those hops extremely fast. Performing those operations in the hardware, says Seager, is a huge leap forward in making BlueGene/L scalable and fast."

In an IBM research report, entitled "Overview of the Blue Gene/L system architecture," authored by Alan. Gara et al, in the IBM Journal of Research and Development (Vol. 49, No. 2/3, 2005 Special Issue on Blue Gene), the authors refer to the tree network as the "Collective Network."

Dr. Alan Gara and three other scientists had been recruited from Columbia University in 1998. He was later named as the Chief Architect of Blue Gene by IBM.

In describing the collective network, Alan Gara calls it a "tremendous improvement:"

"Arithmetic and logical hardware (ALU) is built into the collective network to support integer reduction operations including min, max, sum, bitwise logical OR, bitwise logical AND, and bitwise logical XOR…**The latency of the collective network is typically at least ten to 100 times less than the network latency of typical supercomputers,** allowing for efficient global operation, even at the scale of the largest BG/L machine.

"The collective network is also used for global broadcast of data, rather than transmitting it around on rings on the torus. For one to all communications, this is a **tremendous improvement** from a software point of view over the nearest-neighbor 3D torus network."