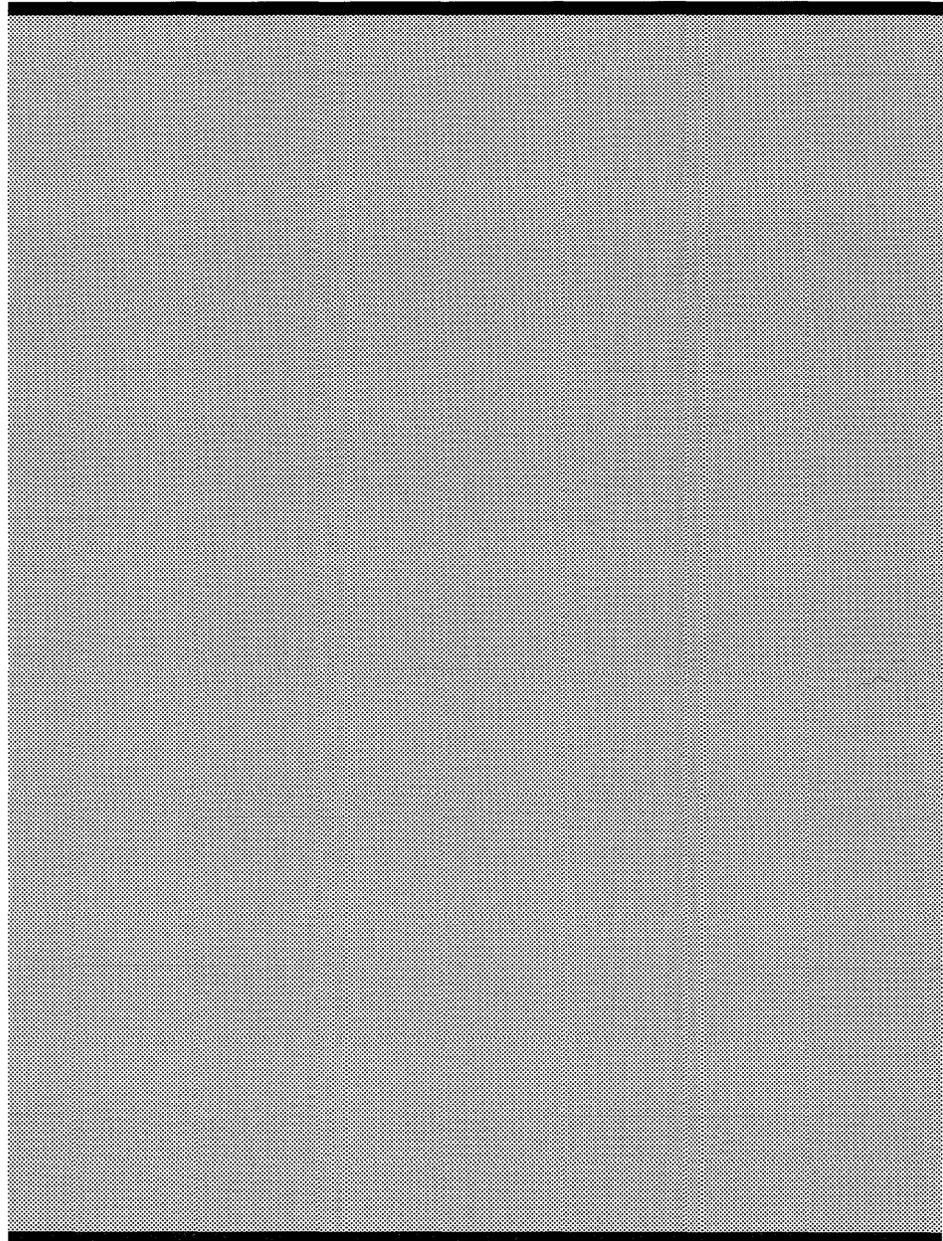


History and Evolution of 1100/2200 Mainframe Technology

UNISYS



This paper was
prepared in
recognition of the
35th anniversary of
USE inc. and was
presented at the
Fall 1990 Conference
in Seattle,
Washington.

History and Evolution of 1100/2200 Mainframe Technology

UNISYS

November 8, 1990
Richard J. Petschauer

HISTORY AND EVOLUTION OF 1100/2200 MAINFRAME TECHNOLOGY

INTRODUCTION

This paper covers the highlights of the history and evolution of the Unisys 1100/2200 Series processor and memory technologies. The period covered spans the last 35 years and includes the recent product announcements in 1990. A number of articles have already covered the earlier years including the computers that were developed then. Others reviewed the formative time of the Eckert Mauchly and Engineering Research Associates (ERA) companies and their joining Remington Rand which later merged into the Sperry Rand organization. Still others have traced the architectural aspects of the 1100 Series. This paper instead concentrates on the internal logic and memory technologies used in the 1100/2200 systems and some of the industry factors present during this time.

These early machines originated during a period when computer designers were searching for better ways to best do the basic functions of computing and memory. This computing, or logic technology as we will call it, had a time lead over that of memory since it had been used in early digital calculators and some control equipment. However these machines were special in nature since their inherent construction and "hook up" determined what function they would perform. The revolutionary concept of the computer was an idea where a general set of basic operations, which could be called out in any sequence, would operate on "data" provided, and return and store the results. The "data" could be numbers or codes representing letters and words.

This programmability of computers, which we all now take for granted, is really what makes them so powerful since it allows them to be

used to solve such a wide variety of problems. However, it also required a new class of technology that the earlier calculators did not need—a means to store the sequence of operations, or instructions, to be performed and also a place to hold the input data and the results. This function of "memory" presented a large problem in the early days of computers. The memory used a different technology than the logic for the processor, and the two developed on separate but parallel tracks, which to some extent still continue today. For quite a few years, until the mid 1970s when the semiconductor memory became so capable, memory technology greatly affected the performance, capability, and cost of the computer system.

Early in the memory technology development, a split occurred. One technology family was developed for the program and internal data needed by the computer—which required fast access; and another was developed for less frequently used data which needed to provide larger capacity storage at a lower cost. This "external" storage class, mostly represented today by disk and tape storage, has seen dramatic improvements over the years. However, it is beyond the scope of this paper which will concentrate on the central processor and its internal memory.

THE 1101 COMPUTER

The 1101 computer is noteworthy since it was the first "1100". It was a commercial offering of a version of the first stored program computer designed by Engineering Research Associates (ERA) under a Navy contract and delivered in 1950. While this was only a 24-bit machine with no actual commercial sales, it did provide an engineering learning base for

the very successful 1103 which followed it. And the old timers like to recall how the commercial model number was chosen: 1101 is the binary representation for the number 13, the task number the Navy assigned to the original development contract.

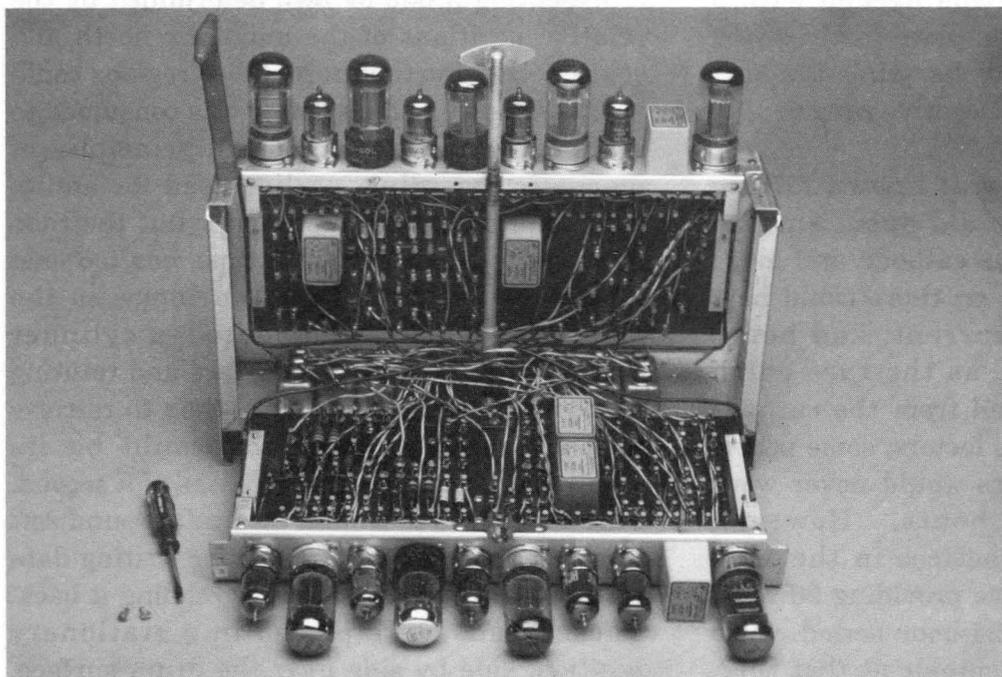
THE 1103 UNIVAC SCIENTIFIC COMPUTER

The 1103 was the first 1100 computer with significant commercial sales. It was introduced in 1953 as the Univac Scientific Computer. It set the standard for the 1100 36-bit word, although its internal organization was changed radically in later machines. Logic was done with vacuum tubes and crystal diodes, mounted on many "suitcase" chassis. A total of about 3900 tubes and 9000 diodes were required. Total power for an installation could be up to about 100 kilowatts including that for the air conditioner, chilled water supply and associated blowers. The recommended floor space was at least 58 by 30 feet for the 38,000 pound computer and its supporting equipment.

Vacuum Tubes

Most of the 1103 vacuum tubes were triodes. A triode contains a filament which is heated by current passing through it and is placed close to a "cathode" which in turn becomes hot causing electrons to be emitted from a rare earth coating on the cathode surface. The negatively charged electrons are attracted to a surrounding positively charged anode or "plate". When a fine mesh termed a "grid" is placed between the cathode and plate, it acts as a control element. A negative voltage on the grid can greatly reduce the current going to the plate. Usually about 100 to 200 volts is applied to the plate, and about -20 volts on the grid can cut off the tube current. Two triodes were contained in one tube envelope, and this pair could make one flip flop. A flip flop, storing one bit of information could be set to a "one", cleared to a "zero", or toggled, i.e., changed to its opposite state. The later function was quite handy in certain logic and arithmetic operations. Capacitors stored the state of the flip flop for a short time so it would not toggle twice with a single input pulse. Another triode was usually connected to each flip flop output as a "cathode follower" (with

its output taken from the cathode rather than the usual plate). This isolated the output wiring load from the flip flop allowing it to change its state faster. The 1103 used about 12 types of tubes; the most common was the 5963 dual triode, an industrial version of the 12AU7, a common part in home television sets at the time.



The 1103 "suitcase" chassis opened up easily for test and repair.

Another tube type that was available then was a pentode (so called since it had five rather than the three basic components of the triode). The pentode required less control grid voltage change for a corresponding plate current change and has two special grids. In a modified version, the 7AK7 dual control pentode, one of these grids was made to provide an extra control grid. Therefore, a fast two-input "AND" circuit could be made since plate current would flow only when both grids were positive. This provided the basis of the "gated pulse amplifier", widely used in the 1103 computer. One of the inputs was usually a narrow clock pulse, and the other a wider logic level from another output. Diodes were also used to form "AND" and "OR" circuits, but the pentode tube circuit was preferred for the narrower pulses. A pulse transformer containing one primary and two secondary windings on a linear ferrite core and potted in a special package was also a key feature of the gated pulse amplifier. The two secondary windings provided the option of either a positive or negative pulse for the output. The transformer also allowed the output to be at a different voltage level than the plate, thereby providing the level shifting needed with a vacuum tube circuit. (The plate would swing from about +100 to +200 volts, while the next grid needed values in the negative range).

Vacuum tubes had several known failure modes: the filament could open, surface particles could fall off the cathode and lodge between it and the grid; or, there could be a gradual loss of plate current, and hence reduced output signal, as the rare earth material became depleted from the cathode surface. Because of these factors, some people predicted large computers would never work for more than a few hours. However, conservative design practices in the 1103 proved to be quite effective providing 90 to 98 percent availability in a 24-hour period which was considered quite acceptable at that time. One thing that helped was testing periodically

with lowered filament voltages. This reduced the tube currents and simulated end of life conditions on the weaker tubes so they could be replaced before they failed. Thus "marginal checking" was born and is still being used even though most modern parts do not fail often by gradual deterioration.

One additional new problem was seen in some vacuum tubes. If they were held in a cutoff condition for a long time, something happened at the cathode surface so that when the grid became more positive, plate current would not reach its full, normal value. Tubes were originally developed for the radio industry and did not operate in this cutoff mode, so the problem had not been seen there. This "sleeping sickness" problem was eventually solved by the vacuum tube manufacturer changing the cathode coating process.

Magnetism for Memory

Magnetism played a strong role in early memory technology and still does for external storage such as disks and tapes. It seemed a natural for this since a magnetic surface could be magnetized in small areas with each region representing a one or zero determined by the relative positions of the magnetic north and south poles. Furthermore, these regions could be quite small, and no power was consumed to maintain their state. So the principles of magnetic recording, earlier used for audio, were applied to digital storage. But the time to retrieve the data from a tape was too long for the computer internal memory, so the magnetic drum was invented—a cylinder coated with an iron-oxide surface and rotating at high speed. The maximum time to retrieve any data from the drum would be its revolution time—a few hundredths of a second. Fine copper wires wound in a coil around soft iron cores formed the "heads" for writing data on the drum surface and later reading it back. These heads were placed in a stationary position, side by side near the drum surface. The direction of the current flow in the head

determined whether a one or zero was written. Reading was accomplished by sensing the direction of the voltage across the head winding caused by the small magnetic field around each region moving past the head coil as the drum rotated. ERA had some of the original drum patents. Early tests involved magnetic tape glued on a cylinder with one head used for each bit in the computer word. The track was one-quarter-inch wide, with a bit density of 50 per inch, and could provide "as much as 200,000 bits for a drum with a 34-inch diameter and 10 inches long" according to a 1948 patent. Later, an iron oxide material was sprayed directly on the drum surface and bit density increased dramatically.

The 1103 used such a drum memory for program storage and provided 16,384 words with a parallel word access in a maximum time of 17 milliseconds. Many program steps are executed in a certain constant sequence, so the time from one step to the next could be much less than that for a drum revolution if they were stored physically on the drum surface in positions corresponding to their sequential use. However, this made programming difficult, and many times a branch occurred, so that the next program step could not be predicted. Also the sequence and location of the data to be read from memory could usually not be determined very well. A faster memory was needed—one that did not use mechanical rotation. Early 1103 machines used a form of cathode ray tube for this, known as the Williams tube. An electron beam was scanned across a phosphor-coated screen. A large current at a spot would charge it so that in the next scan sequence a small charge would temporarily remain and be sensed. The data had to be continually read and refreshed on a periodic basis. Each tube contained a 32 by 32 matrix and could store 1024 bits. The first five 1103 systems used a 1024-word memory of this type. A total of 36 five-inch diameter tubes were needed which took a considerable amount of space. This

type of memory, while faster than the drum, proved to be unreliable and difficult to maintain. What was needed was a low-cost way to store and retrieve data in any random sequence with a short and uniform time delay for access. To answer this problem the ferrite core memory, using a new form of magnetics, was invented in the early 1950s.

The 1100 engineers were quick to seize on the benefits of the ferrite core memory and incorporated a 1024-word unit in the 1103 starting with serial number 6. Delivered in November, 1954, it is believed to be the first commercial unit delivered with a core memory. Beginning with serial number 10, the memory was expanded to 4096 words, and the computer was designated the 1103A. Most of the 1103s used this Random Access Memory, as it was termed, (spawning the acronym "RAM" still commonly used today to refer to this class of memory). This memory also needed vacuum tubes for its operation. The 4096-word unit fit into a separate cabinet and used 470 tubes and 2200 diodes and required about three kilowatts of power. At an engineering meeting in 1956, per bit costs of core memory with electronics were quoted at about \$1.25, with drum at three cents, and tape at 0.1 cent.

The 1103 computer system used clock pulses which came from a timing track on the drum and occurred every two microseconds. It is difficult to control the speed of a drum motor, so instead it provided the computer clock. The 1103A computer had an ADD time of about 30 microseconds and the memory had an access time of six microseconds and cycled at twelve microseconds. The 1103 sold for about one million dollars including the 16,384-word drum memory and a 4096-word core memory. An additional 4096-word core cabinet cost \$200,000. Later an enhanced version of the 1103 was designed and named the 1105. The major changes were a new buffered Input/Output (I/O) and an optional third 4K

word core cabinet. (In referring to memory capacity, we will follow the convention where K equals 1024 and M or Meg equals 1024x1024 or 1,048,576). The central computer was unchanged. Combined deliveries of the 1103 and 1105 were about 45 machines.

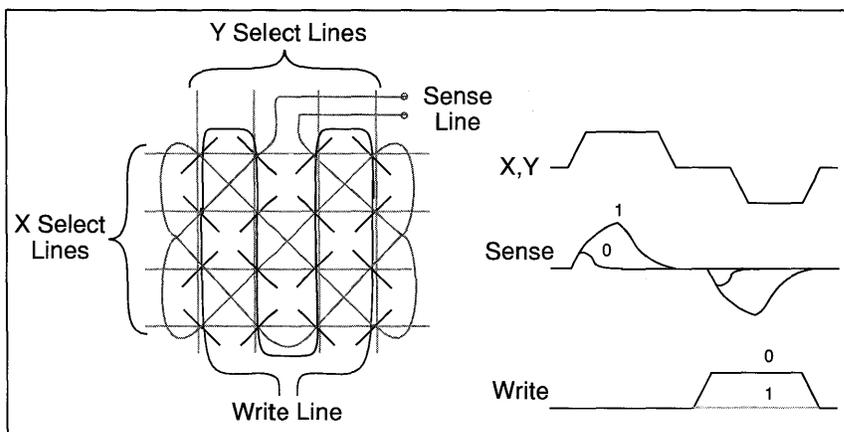
THE FERRITE CORE MEMORY

The core memory as first used in the 1103 was an ingenious device that was to last nearly 20 years in the industry and spawn entire organizations using special skills and techniques. The "cores" referred to tiny doughnut shaped rings (toroids) which were made from an iron oxide powder known as ferrite. These were fabricated using ceramic technology which consisted of grinding and powdering the material and mixing it with a binder which held the particles together. The material was then stamped into its final shape with modified pill presses such as those used to make aspirin tablets. Finally, they were fired at high temperature and individually tested for precise uniformity. The cores, used in the 1103A memory, measured 80

thousandths of an inch outside diameter and 50 thousandths of an inch inside diameter. Each core could store one bit of information, corresponding to whether the magnetization flowed in the clockwise or counter clockwise direction. Writing into the core was done by passing a sufficiently large current through a wire passing through the core, and this would create a clockwise or counter clockwise magnetic field in the core (depending on the direction of the current flow) and cause the magnetization, or lines of flux, to take the same direction. When the current was removed, most of the ferrite core, having a high degree of magnetic retention, would maintain its internal direction of magnetic flux.

Reading was accomplished by passing a current through the core forcing it to the zero state. Any change in the flux state of the core would be sensed by another wire passing through the core. A large flux change would mean that the opposite state, i.e., a one had been stored while a small change indicated that the zero state had already been in the core. The act of reading thus cleared the core to the zero state so that after reading, a stored one had to be rewritten back to its original condition.

Core Memory Operation



Coincident current core memory, four-wire arrangement and read/write current pulses.

The other phase of the core memory operation involved how it was organized into a complete memory and how a relatively small set of drive and sense electronics controlled a large number of individual cores. Each core had a threshold current that if not exceeded would not cause the core to

change its magnetic state. But if it were exceeded by a relatively small amount, the core would switch to the opposite state. (The time to switch was inversely proportional to the amount of excess drive current applied). So the core actually provided both the functions of storage as well as a form of a two input "AND" operation.

If one wanted to build a 4096-word, 36-bit per word memory, for example, the person would construct 36 separate planes, each made from an array of 64 by 64 cores. These would be 64 X and 64 Y selection lines on each plane. By placing a "half-current", a value about 10 percent less than the threshold value, down any X and any Y line, only the core at the intersection would receive a magnetic field of sufficient magnitude to switch. A single sense line could link all the 4096 cores in each plane and detect if the selected core had switched. The sense line was wound in a special diagonal way so that each row and column of cores would give an opposite polarity voltage. This was important since the 62 half-selected cores each would generate a small, but reversible, magnetic flux change, and the summation of all of them would be larger than that from a single core switching. With the diagonal sense line, these small voltages would cancel. Actually, a half-selected one would generate a little more voltage than a half-selected zero. Consequently the "worst case pattern", the then familiar arrangement of checkerboard ones and zeros, would sense a net difference between each pair—the so called delta noise—and sense these pairs of differences. This combined delta noise could approach a full amplitude one switching. Fortunately however, this noise would die out before the switching was complete, which took a few microseconds. Therefore the sense amplifier output was strobed shortly after this noise had subsided.

The X and Y lines would be connected in series for all 36 planes. To accomplish this physically the planes were placed on top of one another,

forming the "stack". This made a compact arrangement. The stack was difficult to repair, but this was quite an infrequent occurrence.

Each plane then had one sense line for reading. For writing another wire had to be included in each plane, so a total of four wires would thread each core. The write line linked all the cores by first flowing along one row then turning around and going back through the next row until all the cores in one plane were threaded. Writing was accomplished in the second half of the cycle. Either the data was rewritten back into the selected core, which had been cleared to zero when it was read out, or new data could be written in. This writing was accomplished by again applying currents down the same pair of X and Y select lines, but in a reverse direction. This would then switch the core to a one state. If a zero were to be written, a half-select current would be passed through the write line overlapping in time the X and Y currents but in an opposite direction. This would nullify the effect of one of the X-Y select currents, the core would not switch, and it would remain in the zero state. Since the write line acted by preventing switching, it was commonly called the "inhibit" line. Since the inhibit line drove all the cores in the plane, it coupled a great deal of noise into the sense amplifier—about 50 times an output signal—both when it was turned on and off. The sense amplifier had to recover from this large pulse before it could be ready for the next read operation. Drive currents required to switch the cores were several hundreds of milliamperes, and the core output voltage was about 50 millivolts.

In the case cited above, it can be seen that 64 X lines, 64 Y lines, 36 sense lines and 36 write lines—a total of only 200 lines—could control $64 \times 64 \times 36$ or 147,456 individual cores. Furthermore, the 64 X and Y lines were usually driven by an eight by eight transformer-diode matrix. (A single transformer-diode was selected by an eight-driver, eight-enable

combination. A separate diode and driver had to be used for each of the X and Y current polarities, but the transformer and enable driver could be shared). So one can see that a relatively few circuits could control a rather large number of cores. The larger the core memory was, the more efficient it became. This was particularly important when electronics costs were high as during the days of vacuum tubes and those of the early transistors.

EARLY TRANSISTOR LOGIC

The transistor was invented in 1947 and by the early '50s was a prime candidate for computers mostly because of its smaller size and greatly reduced power consumption compared to vacuum tubes. These early units were made from germanium, one of the natural elements that exhibit semiconductor properties. They operated in the "bipolar" mode where current is amplified between the input and output. A relatively small current forced into the "base" would cause a much larger current to flow in the "collector". The third terminal, the "emitter", was usually grounded and carried both the base and collector currents. This operation was quite different from a vacuum tube where grid voltage was the control element, and no appreciable grid current was involved. The transistor also had the advantage of operating with a low collector voltage and thus acted as a very good "switch" to turn on and off signals. Furthermore, the output voltage from one transistor stage could usually be directly coupled to a following stage without the large voltage shift as was required by vacuum tubes.

Still, it took a while for the engineers to figure out the best circuit arrangements to use with transistors. At first, some people tried to copy the vacuum tube circuits with their pulses and transformers. (Cartoons were seen on some of the lab walls which showed a small three

terminal tubular device under the sole of a large shoe; a caption read: "Stamp Out Transistors"). It later turned out that a simple direct-coupled circuit combined with diodes was a better approach.

One of the problems with transistors at this time was that their frequency response was less than tubes. To overcome this, more current than normally would be needed was forced into the base causing the transistor to turn on faster. However, when the current was later removed, the transistor would remain on for some time such as a large fraction of a microsecond. This delay was defined as the transistor "storage time". A larger "on" current would cause a larger storage time. This effect was minimized by providing a current in the reverse direction to turn the device off harder. The turn-off current sometimes had to be nearly as large as the turn-on value. Improved transistor construction finally reduced this effect, but this took a number of years.

Early transistors were quite expensive and delicate. There was also some concern about their long-term reliability and ruggedness. One competing and attractive technology at the time used low cost magnetic switch cores. These were cores made from a ribbon of permalloy, a magnetic iron-nickel alloy, wrapped on a small bobbin and then wound with several windings. The windings could be connected to inputs, an output, or a clock pulse. The power to operate a logic circuit came from the clock which would force the core to a cleared state. A switching output voltage would indicate that a one state had been stored and would cause a following core to switch. Since read out of a switch core cleared its state, a two-phase clock was needed. Data from cores clocked on the "A" phase would be captured by cores which would be clocked at the "B" phase and vice versa.

In order to evaluate these approaches, two

internal test vehicles were built in the mid '50s: the Magstec, using the magnetic technology, and the Transtec using transistors. Both were built and evaluated; the transistor technology version was used for most applications after this. Some notable non 1100 exceptions in the company were smaller computers developed for early business applications. One was the Solid State 80/90 designed in the Philadelphia location, and the other was the File Computer designed in St. Paul. These combined tubes, magnetic logic, and some transistors.

The first all-transistor computer built by Univac was the Athena, a computer design starting in 1956 for the Air Force which was used for missile guidance. Small diodes were also readily available then. The basic AND-OR logic was done with the diodes; and the transistors provided the amplification, or gain, between stages.

In those days, the companies that manufactured vacuum tubes were the first to market transistors. Philco made one of the faster transistors, using a process termed "surface barrier technology" or SBT. The SB100, as it was called, was the fastest transistor at the time, but it could not handle much current and had a low voltage rating. General Electric made an "alloy junction" processed device, the 2N123, which was slower, but huskier, and could handle more current. The Athena computer used both of these in each logic circuit, the first to provide the voltage gain, and the other to provide current gain and drive the logic loads and wiring capacitance on the output. The Athena computer exceeded all availability goals by a large margin, and proved the reliability of transistor technology. Transistors improved substantially over the next few years in performance, current and voltage handling capability, and cost. In a short time they could even accommodate the higher currents and voltages needed by core memories. Vacuum

tubes quickly faded from most computer designer's minds.

COMPANY SITUATION DURING THE LATE 1950s

Before we cover the technology for the next 1100, it would be helpful to discuss the situation in the St. Paul Development Labs at this time. (The Roseville Plant was not occupied until later in 1961). There were many and various computer and memory projects then. Most of them were funded by one of the government agencies who set the technical objectives for the work. The corporation was now Sperry Rand, and there was a sister organization in Philadelphia also developing computers but with less government work and with designs more geared to business needs than scientific. At about this time the Lawrence Radiation Laboratory at Livermore, California solicited proposals on the next generation of a high performance computer. The St. Paul and Philadelphia groups submitted competing proposals. St. Paul, being conditioned by the Navy which stressed reliable and maintainable equipment, submitted a more conservative design; the Philadelphia design was much more ambitious in both its technology goals and in the logic organization, and promised much more performance. The contract was fixed price, and the customer selected the higher performance machine. This computer, the LARC, proved quite a challenge to complete. It was never delivered to any users besides Livermore because of its high costs. It is believed that the existence of LARC probably delayed the starting of an alternate follow on to the 1103. Besides the Athena computer mentioned earlier, other non 1100 significant developments in this time period included the Naval Tactical Display System (NTDS), which was later commercialized as the 490 Real Time System, and which then spawned the 494 Product.

The Process Control Computer, developed for an industrial customer, formed the foundation for the 418 System which led to its own follow on family. Many of the engineers who designed the later 1100 products had earlier worked on these and other commercial and government computers of various types and gained by this experience.

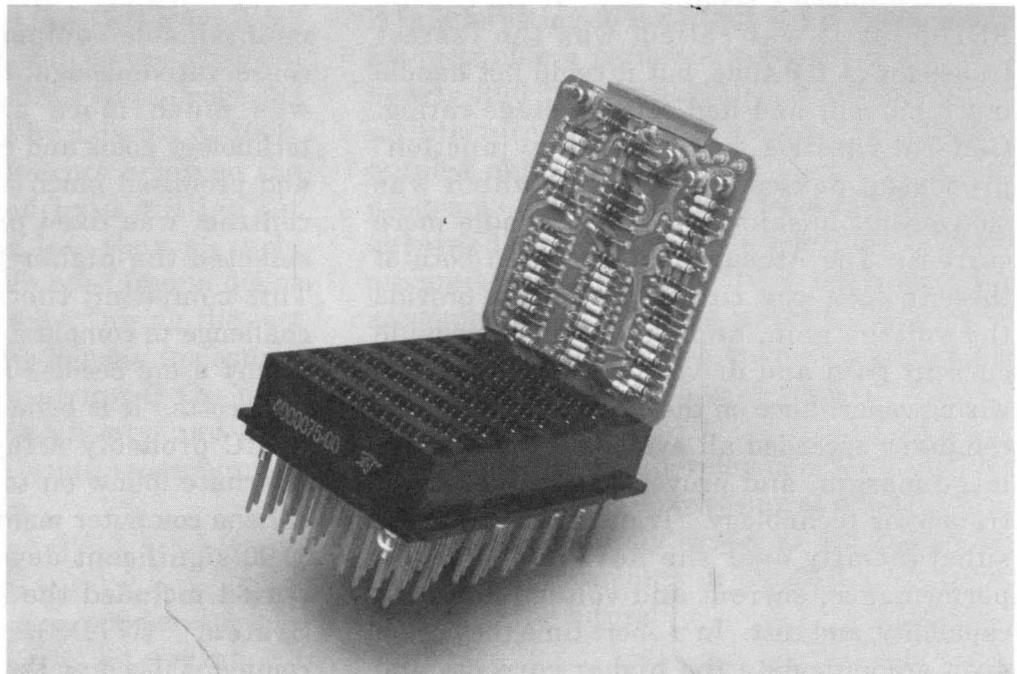
THE 1107 COMPUTER

The 1107 computer development started in the late 1950s and was announced in December, 1960. It maintained the 36 bit word length from the 1103, but otherwise was quite new in its organization and set the basic architecture for future 1100 machines. A key feature of this is the GRS or General Register Set of 128 words. In the 1107 this was called the Control Memory and was constructed from a new type of magnetic memory fabricated from deposited thin films which will be described in more detail later. A number of companies had been working on film memories including IBM and Burroughs, but this was the first commercial computer to employ them. The company played up this new technology, and in fact named the system the 1107 Thin Film Memory Computer, with this name appearing on the operator console. In a Business Week article dated December 17, 1960, it was stated that the news leaked out about this new computer, and the company stock became the most active on the New York Stock Exchange. Following this, the Exchange asked the

company to announce the product earlier than had been planned, the article said.

Logic Technology

The basic 1107 logic circuit used transistors, diodes, and resistors. They were mounted on small plug-in printed circuit cards measuring about three by four inches. About 20 variations of these were needed to perform all the computer and I/O logic. An edge connector with 22 contacts was used. These cards plugged into molded connector blocks, each holding eight cards on 0.5 inch centers. A five by five array of these blocks formed a deck holding up to 200 cards. A bay consisted of four decks arranged vertically. Four bays in a "U" shape then made up the computer complex. The film memory used two decks, and the remaining fourteen—with 2800 card locations—were available for the processor and the input/output function for the entire system. Wire-wrap technology, a very reliable technique that had been developed by the phone company, was used to interconnect the cards. It was interesting to note that in this type of design the complete personality of the computer logic was determined by what card



A 1107 logic card plugged into a section of the connector block.

types were plugged into which slots and the wire wrap interconnection on the backplane. Logic design changes could be easily incorporated by changing the backplane wiring and perhaps a card type change in a few slots. The logic delay per stage was about 80 nanoseconds, although it could vary greatly with the length of the wires driven by a given circuit. An innovative feature of the 1107 logic technology was a transformer-diode shift matrix which was important in accomplishing certain arithmetic functions.

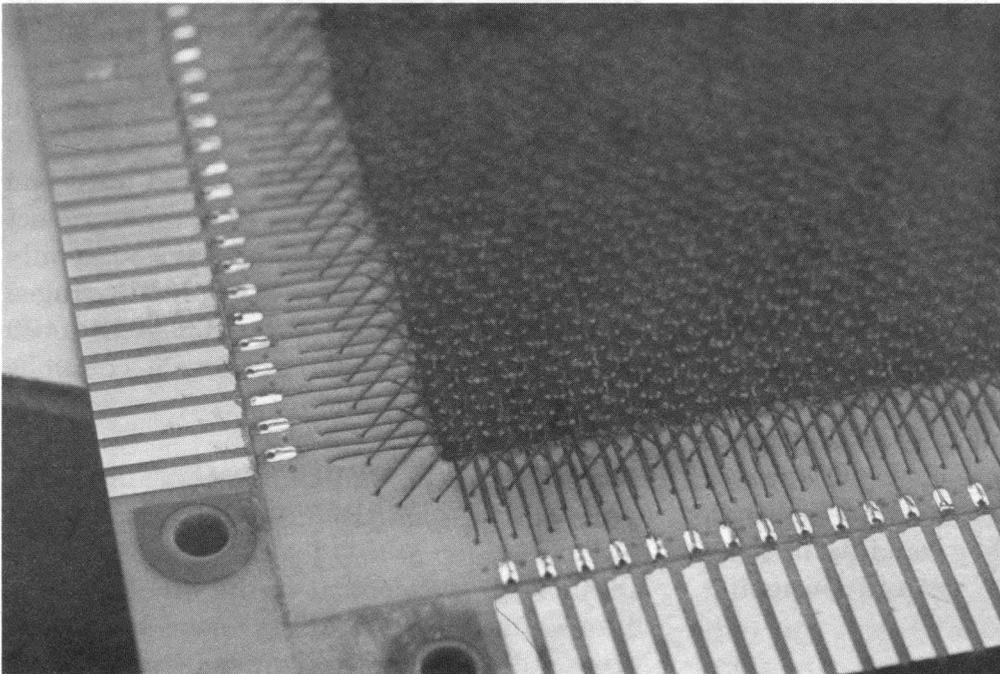
Core memory

Ferrite core memory technology progressed rapidly during the late 1950s, and better transistors and diodes became available for handling the larger currents. The 1107 core memory was large enough to handle both active programs and data so that a drum was not needed for this function. Cycle time was reduced to four microseconds. To further improve performance, it was organized into two separate banks. Every two microseconds, one bank containing the program, or the other containing the data, could be referenced. The entire memory of 65,536–36 bit words fit into a

36-inch wide, 84-inch high cabinet. Since the memory was still fairly expensive, smaller capacities in 16,384 word increments were also available. The core plane contained 4096 cores each measuring 0.050-inch diameter. Pulse transformer and diode arrays were mounted at the end of the core stack to provide the X and Y current selection. Since the memory cores were somewhat temperature sensitive, the drive currents had to be compensated to adjust for the computer room environment.

Thin film memory

As mentioned above, the 1107 also featured a new, fast type of memory made of small dots of thin magnetic film. It was a small memory of 128 words but cycled at about 0.67 microseconds—six times faster than the core memory. Its function was to store frequently used control words and local registers for the arithmetic and logic sections, so it was important that it operated at a speed close to that of the logic circuits. The access time was half of the cycle time—about 0.33 microseconds. The use of this small high speed memory had a lasting impact on the 1100/2200 architecture and certainly improved its performance.



A close up view of the 4096 bit 1107 core memory plane using 50 mil cores.

The 1107 was the first commercial computer to use a thin film memory. (MIT Lincoln Labs built a smaller one earlier). It represented payoff from about seven years of R and D when it was announced. Many other companies were also working on these types of memories. Film was a competitor to the core memory. It was made from a thin film of iron-nickel

composition deposited in a vacuum through a mask used as a stencil which formed an array of small spots on a thin smooth glass substrate. This was done in the presence of a magnetic field which produced an axis of preferred magnetization in the plane of the film. The magnetization, which represented the two stored states, could be switched to point either parallel or anti-parallel to this direction. All magnetic flux paths must close back on themselves. The path in the thin film had to close through the air and surrounding material which was non-magnetic. Since it is harder to push these flux lines through a non-magnetic material, a back magnetic force results which has a direction opposite to the original flux lines and attempts to demagnetize the magnetic material. Because of this, the film had to be kept thin which reduced the flux and associated demagnetizing force. (The ferrite core did not have this problem since the flux lines were circular and closed inside the core itself). Because of this limitation, thin film had an output signal about one-tenth that of a typical core.

Writing and reading were accomplished by placing the glass substrate between etched planes of copper conductors which had been laminated onto thin Mylar layers and bonded together. The thin film memory promised lower cost since many storage elements could be made in one deposition, and since the difficult job of threading several wires through a core could be eliminated. It turned out also that film switched from one stored state to the other much faster than core. So it appeared that film could become the next generation of computer memory and the 1107 version was just the beginning. However, the engineers that had to work with these elements knew of the new problems that films introduced that had to be solved before they could fulfill their promise of having all around superior features.

Earlier, the company had designed an experimental model of 1024-word, two-microsecond memory and reported on this in 1956. It used the same coincident current scheme as cores. However in the construction of the test vehicle it was learned that the uniformity of the magnetic properties of all the individual film elements varied too much for reliable operation. Consequently, the next film memories developed used a linear selection scheme. Instead of the desired word being selected by the coincidence of two pulses, it was directly selected by a full current which could be made large enough to switch all the elements on a given glass substrate. Sensing and writing were accomplished by using conductors flowing at right angles to the word lines. These linear select or "word organized" memories were more costly because of increased driver/selection circuitry, but in a small memory such as that on the 1107 this was not a significant item. The 1107 thin film elements were vacuum deposited onto two-by-two-inch, thin glass substrates. Each contained a 16 by 18 array of 0.05-inch diameter circular elements. The magnetic elements were only about four millionths of an inch thick. Four arrays were mounted on planes and four planes comprised the entire memory.

The 1107 was scheduled to be delivered in December of 1961, but was delayed until mid 1962. It was technically a successful product and operated well in the field although only about 38 of the units were sold. During a similar time period, about 60 of the 490 systems were delivered.

TECHNOLOGY DEVELOPMENT IN THE EARLY '60s

Technology development in the early '60s was diverse and rapid. Many new ideas were proposed and some were used in products. Transistors became faster and less expensive.

A major innovation was the invention of the integrated circuit which placed many transistors and resistors onto a single piece of silicon, or “chip” as it became known. Originally these devices were quite simple, perhaps one flip flop; their performance then was inferior to that which could be obtained with separate transistors and diodes. But progress was steady, including determining what functions should be performed by the circuits contained on the chips.

Printed circuit board technology was also key for a means to mount and interconnect the components. The company has had a strong printed circuit board development and manufacturing capability since the 1950s. The company also developed special techniques for manufacturing ferrite cores and assembling them. Work continued on improving the magnetic film memory, and a new large continuous production line for film deposition was installed. There were several applications by the company using film memories for military computers during the 1960s. One noteworthy example involved a concept termed “Mated Film”. This was a clever idea that produced a very compact and rugged stack with excellent temperature characteristics and operating margins. Both the films and a common write/sense line were deposited on the substrate. There was a pair of film spots for each bit—one on each side of the sense line. This helped close the flux path and improved the operating margins substantially. Holes were etched in the substrate to accommodate a diode lead that was used for word selection. The word lines were merely the diode leads that passed perpendicularly through the planes. This novel arrangement meant that no connections had to be made between planes so they could be placed very close together. Manufacture of this type of memory actually continued through the late 1980s for a number of military computers requiring relatively small, compact memories. However, magnetic film memories never really fulfilled their

promise in the commercial computer area even though many large companies devoted extended efforts to this technology. The principal reason was that core memory stacks costs dropped considerably more than anyone had anticipated, and that film memories needed more costly electronics with their requirement for word organization and the rather small output signal. In addition, even though the films switched faster, the delays in a stack and the electronics diluted this. As cores became faster, there remained less than a 2:1 difference in the speed of the two technologies, not enough to warrant the cost differences. And as customers demanded larger memories, the cost factor became more important.

Many other magnetic memory approaches were proposed during this time by various people and companies in the industry. Some of these were the ferrite sheet, transfluxor, the biax, magnetic rod, twistor, and waffle iron among others. None of these were used to any great extent in general purpose computers. One exception was the plated wire, but we will cover that more a little later. During this time the first semiconductor memories were proposed. The ones made then were quite small and quite expensive, although they were fast and simple to design with.

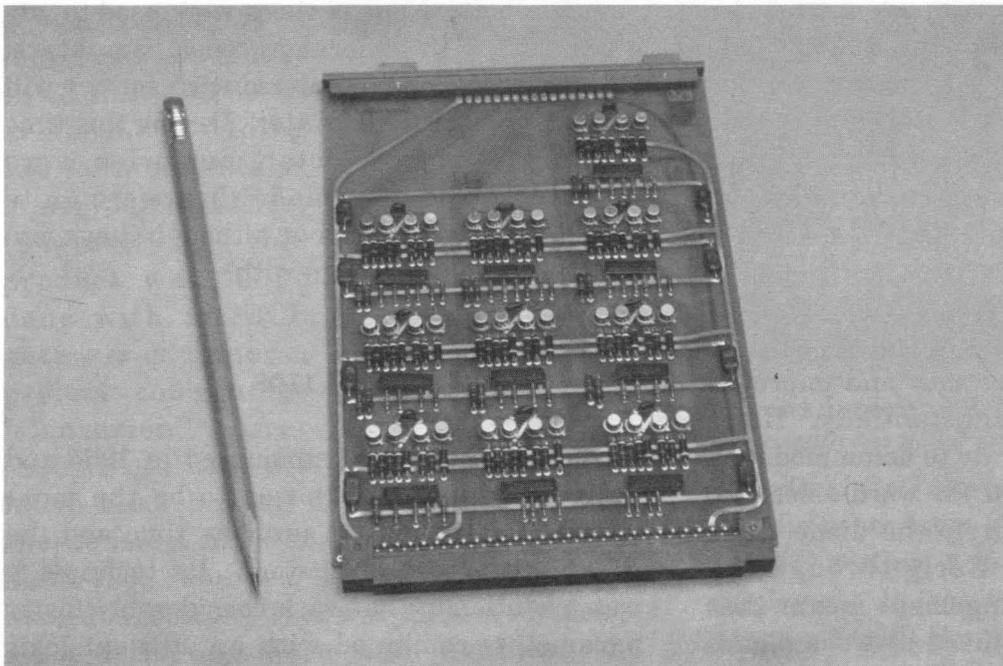
THE 1108

The 1108 system was announced in 1964 and delivered in 1965. It was to be the most successful 1100 system until its time and the first multiprocessor system. Its technology was evolutionary but with considerably faster parameters combined with an efficient logic organization. The 1108 performance exceeded that of the 1107 by about four to five times.

Logic Technology

At the time that the 1108 technology selection was made, integrated circuit logic still wasn't

fast enough, so transistor-diode logic using the latest devices was used. The germanium transistors were replaced with those made from silicon. "Two level logic" was chosen which meant that both an AND and an OR logic function using the diodes could be done before passing through the transistor amplifier which now consisted of two transistors in an arrangement similar to the earlier Athena computer. Logic card size was increased to five by seven inches with 55 I/O pins. Cards were on close 0.3 inch centers, and a total over 900 were used for the processor and the I/O section. The number of card types was increased, but the idea of general purpose cards with the real logic in the wire wrapped backplane—similar to the 1107—remained. A high fan out driver was added to the logic family. It was able to drive many internal loads and long backplane wires, and contributed to simplifying the machine design and aiding its performance. The typical logic stage delay was reduced to 15 nanoseconds, and the clock cycle time came in at 125 nanoseconds.



The 1108 logic card; It was in production for 15 years.

An integrated circuit containing one bit of storage was used to perform the GRS function that the film memory had done on the 1107. Its function was enhanced over the thin film

memory in that simultaneous reading and writing from two different addresses was possible. Its cycle time was reduced to 125 nanoseconds. Sixty four chips in circular cans were mounted on a standard logic card together with other supporting components. A total of 127 cards were needed to provide the complete GRS function which became the performance limiting section of the processor complex.

The ferrite core memory cycle time was reduced to 750 nanoseconds—over five times faster than the one in the 1107. This was brought about by the use of a smaller, faster 23 mil core and a new "2-1/2 D" organization. In this new organization the Y select drivers were duplicated for each bit so that if a zero were to be written, the corresponding Y current would merely be left off. This meant that the inhibit line could be eliminated, reducing the large noise generated after writing. It also meant that only three wires

were needed, allowing a smaller, faster core. The cost of doing this was to require 36 sets of Y lines, one per bit, rather than only one set. The reduced cost of transistor drivers minimized the impact of this. Furthermore, each "plane" was made "non square" decreasing the number of Y lines per bit and increasing the X lines but minimizing the total number of drivers. This also reduced the back voltage on both X and

Y lines since they drove fewer cores and allowed a planar type stack which reduced interconnections between planes as well as manufacturing costs.

One of the concerns at this time was the self heating of a core which would occur if it were switched repeatedly at this fast cycle rate. While repeating one address for many successive cycles would be quite unlikely, the core switching threshold varied with temperature, so if there were heating it might cause a problem. Logic was added to detect this and cause a slight slow down in the rare event such would occur. Another problem with some cores at this time was that they were magnetostrictive. This meant that mechanical stress applied to the core would cause a change in its magnetic properties. It also meant that the switching of a core would physically stress it. Consequently when a core was switched, a shock wave starting from its inner diameter would propagate to its outer diameter edge. Then it would be reflected back in a wave action and proceed to the inner edge where it would be again reflected back out. Normally the effect of this would be quite small. However, if the core were switched at the same frequency corresponding to the round trip delay of this shock wave, a large reinforcement action would occur which might cause a read out error. This problem was usually controlled by picking the right core material and covering the cores with a light spray coating which absorbed most of the shock wave. The coating also prevented the core from rubbing on the wires as a result of vibration caused by cooling fans or other sources.

The 1108 was quite large by today's standards. The processor required a 64-inch wide, 64-inch high cabinet with a separate 36-inch cabinet for power and maintenance. A 65,536 word memory, organized in two banks, took a 48-inch unit plus another 36 inches for its power supplies. The original 1108 was delivered in 1965. Enhancements were made soon after allowing multiprocessor capability with up to four memory cabinets for a total of 262,144 words of storage.

1108 Product Enhancements

In about 1969, a slower version of the 1108 with a 1.5 microsecond, 128K word, less costly core memory was introduced. It was named the 1106, and expanded the usage of the 1100s to users who did not need the full power of the 1108. It became a very popular system. Later on in the early 1970s when semiconductor memory technology became cost competitive, the core memories in both were replaced with semiconductor technology and the maximum system memory was doubled to 524,288 words. The GRS chip was also replaced with a more advanced version containing 256 bits. There were no 128 bit versions made, so only half the address locations were used in each chip. The number of printed cards needed for the GRS function dropped from 127 to only 15. The resulting lower cost machines were repriced and renamed the 1100/20 and the 1100/10 with active production continuing until the late 1970s. The 1108 including its enhanced derivatives was one of the most successful computer systems ever developed and greatly expanded the base of loyal 1100 users. Approximately 1000 of these computers were manufactured and delivered to many installations around the entire world.

THE 1110

Background

The 1110 system was the next 1100 machine. Its development started in the late 1960s, and it was announced in November, 1971. At its inception both it and the 1106 were planned as extensions to the 1108 family with the objective of extending the performance range to both higher and lower levels. The 1106 as mentioned above came out earlier, since it did not need a new processor but only a new memory. The 1110 objective was to get a machine out in a short schedule with low technology risk yet providing two or more times the performance of the 1108. By this time integrated logic circuits had progressed

sufficiently so that they were faster than separate transistors but not by a large amount—only about 25 percent. In addition to this, core memory technology was benefiting from minor cost improvements, but it was difficult to improve their speed very much unless quite small modules were used. But they would be more expensive since the supporting electronics would be prorated over fewer storage bits. Under these conditions how was the 1110 to meet objectives? Part of the answer was to improve the processor performance by increasing the “overlap”, or the number of instructions being concurrently executed. While this would increase the amount of logic in the computer by a large factor, the denser and less expensive integrated circuits should accommodate this. In addition, it was felt that more improvement could be obtained at the system level by a more efficient organization.

This was to be accomplished in three principal ways:

(1) The main memory would be broken into two sections. One would be of less capacity but faster, and the other would be larger and operate slower.

(2) A separate Input/Output or “I/O” section would be used to offload this function from the central computer which was now renamed the “Instruction Processor” or IP.

(3) The number of processors that could be connected in a multiprocessor system was also increased to four and later to six.

Logic Technology

Integrated circuit logic speed had improved significantly and the cost projections became favorable. A type of circuit termed TTL which stood for Transistor-Transistor Logic was becoming standard in the industry. The TTL circuit operated a lot like the diode-transistor circuit everyone was already familiar with. In integrated circuits, transistors are as easy to make as diodes and provided a little better operation. A new standard package adopted by the industry, the dual-inline package, or DIP as it became known, was used. A total of 14 leads at 0.1-inch spacing left two sides of a rectangular package. While the circuit delay only went down to about 12 nanoseconds (from 15 for the 1108), it was felt that the capability of the higher density, lower cost technology would allow the use of additional logic, and thus help performance by producing more concurrent operation in the processor.

The original plan called for a printed circuit card connector having about 100 pins. However since excellent experience was had with the 1108 printed circuit card and its connector, it was decided to continue its use in the 1110. Hindsight indicates this was



The 1110 was the last system to use separate wires on the backplane.

probably a mistake since the 55 pins were not sufficient to support the 30 integrated circuit logic chips, each with 14 pins, that fit on the card. Many cards were half filled, and about 1050 cards were needed for the entire processor. This integrated circuit also had less noise immunity than the 1108 circuit which forced the addition of considerable twisted pair wiring in order to limit signal crosstalk in the backplane wiring.

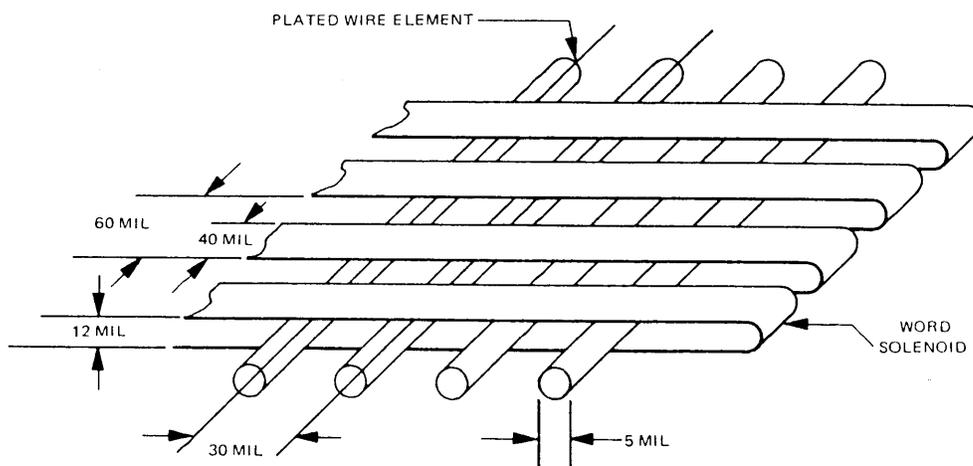
Another difficulty in developing the 1110 was the fact that much of the logic personality moved onto the printed circuit wiring where it had been in the easily changed backpanel on the 1108 with its simpler, general purpose cards. This caused the number of unique card types to increase to about 400. It consumed considerable time to layout the interconnect on these cards. Also, many of the design changes arising during computer test required changes to the printed circuit card layouts. All of these factors and the more complicated design with the extra logic caused the development schedule to be delayed from the plan.

Memory Technology
- The Plated Wire
Memory

The 1110 used two levels of main memory: A 1.5 micro-second "Extended" core memory—the same 128K word memory that the 1106 used—and a higher speed 64K word plated wire memory for the "Primary" storage which had a 300 nanosecond read cycle time. A maximum of eight cabinets of extended memory could be used providing one million

words, while up to four cabinets of primary memory were allowed.

The plated wire memory concept was originally described by Bell Labs at the 1958 Magnetics Conference, and the company became interested in it as having some of the performance benefits of magnetic film, but with lower cost potential and larger output signals. The idea was quite simple: A permalloy, or iron-nickel, alloy of the composition used in magnetic film was plated onto a small copper wire in a continuous fashion as the wire passed through an electroplating bath. Direct current was passed through the wire during plating to force a preferred direction of magnetization to be in the circumferential direction. This has the advantage of a closed flux path allowing the magnetic material to be about 10 times thicker than had been used in the thin film memory in the 1107 and resulted in a larger, easier to use output signal.



Plated wire memory organization.

To make a memory, one merely wrapped a one-turn word line strap around a group of wires. To write, a large current was passed through the word strap, and a small current was driven through each plated wire. A positive wire current wrote a one while a negative current wrote a zero. For reading, the same word current as writing was used, and it produced a positive or negative signal across the wire depending on whether a one or zero had been stored there. Since the magnetic field produced by the word strap was parallel to the wire and thus perpendicular to the remnant flux (which was circumferential around the wire), read out did not change the state of the stored data. This provided a "Nondestructive Readout" or NDRO operation. As a result, rewrite was not needed—as it was in a core memory—following a read operation. This meant that the read cycle time could be faster than a write cycle since it did not have to recover from write current noise. This was important since normally about 80 percent of memory operations involve reading.

Earlier the company introduced a small byte organized processor, the 9200, which contained an 8K byte plated wire memory module, the first to be used in a commercial product. This created a lot of interest in plated wire memory technology around the country, and many companies began a program to evaluate this technology. While the 1110 memory used the same plated wire, the memory stack was larger at 8K words. The plated wire was only 0.005 inch in diameter, placed 0.03 inches apart. The word straps, running at right angles, were on 0.06 inch centers. Each module of 8K words had its own stack and complete memory electronics. The read cycle time was about 300 nanoseconds while the write cycle was 500. Eight of these were mounted in a 64K word cabinet, and up to four of these could be used in a full system. All modules could be operated concurrently. This small module of 8K words helped performance in a multiprocessor system since it reduced the

chance that different processors or I/O units would request data from the same module and have to wait.

A multiple memory adapter (MMA) in each cabinet provided an interface between the requesters and the eight modules. It provided a crossbar switch function and also resolved the priority in the event that two requesters wanted simultaneous access to the same module.

Like many new technologies, plated-wire memory had some problems that the early developers had not envisioned. For one thing it was somewhat strain sensitive. Handling and soldering it to its connections could change its magnetic properties. A more serious problem, also seen in thin film memory, was that repeated writing in one bit location could cause a neighbor to partially lose some of its stored flux and give a smaller signal when later read out. This "disturb" problem was difficult to solve and took considerable experimentation to determine the sequence of reading, writing, etc., so that the wire could be properly tested in the worst case mode before it was assembled into a stack. The plated-wire memory proved more difficult to manufacture than was envisioned and did not reach its cost goals. But even if it had, both wire and core memory were being rapidly overtaken by an offshoot of the integrated logic technology - i.e., semiconductor memory. The 1110 was the last 1100 machine to use magnetic main memory. And in fact, both the wire and core memory on the 1110 system were later replaced with semiconductor memories at lower prices. The renamed system, the 1100/40, also offered twice as much primary memory for a maximum of 512K words.

1110 System

The 1110 hardware took more floor space than the 1108. This was principally because separate cabinets were needed for the I/O, and

the two types of memory each had their own. A "4 X 4" system (four IPs and four I/Os) required an area of nearly 900 square feet. The first 1110 systems were delivered in 1972, and over 400 processors were manufactured in the next few years.

SEMICONDUCTOR MEMORY

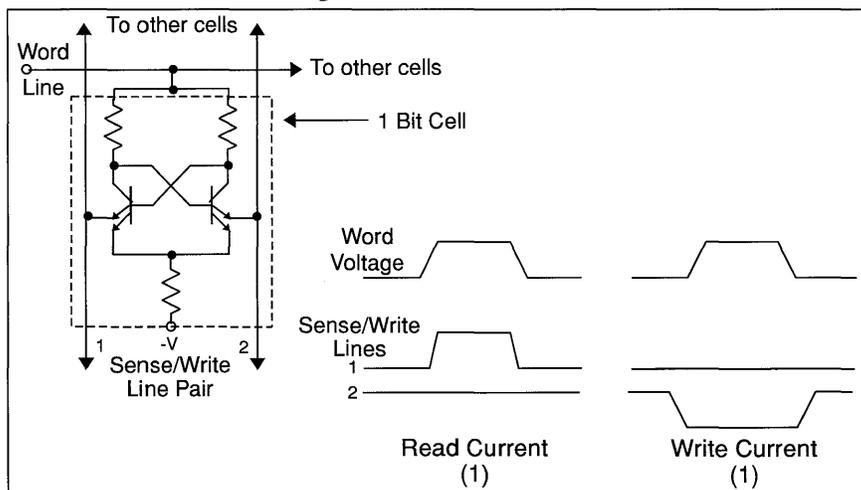
Using transistors to make memory elements began in the late sixties as the number of components fabricated on the integrated circuit chips increased sufficiently. At first these were limited to quite small memory applications such as register files or like the chip used in the GRS for the 1108 as mentioned earlier. They used similar bipolar transistors as were used for logic and were quite fast. While these were still much too expensive for main memory, they did allow a new memory organization concept to be adopted—the cache memory. The idea behind the cache memory was to have a small, fast memory store the contents of every word read from the larger main memory as well as its adjacent neighboring addresses that shared a

common block of perhaps 8 to 16 words. Studies of address sequences had shown that any address requested was likely to be in the same block as the preceding one requested. So if a number of such blocks were stored in the small high speed memory, a large portion of the time the processor could get the data from that memory and avoid going to a larger, slower unit. This concept was a significant development for computer organization and removed the demand for a large, fast, and low cost memory all in a single technology.

The second significant development was the MOS (Metal-Oxide-Semiconductor) transistor. It was also known as the Field Effect Transistor or FET. This type of device was smaller, denser and less expensive to fabricate than the bipolar transistor. Its lower speed was still sufficient for main memory especially if a cache memory were used. The MOS transistor operates by an electric voltage that is applied to the "gate" which controls current flow between the "source" and the "drain". A thin silicon oxide insulator separates the gate from the rest of the device, and no gate current is needed in operation. Interestingly enough,

the MOS electrical behavior was similar to a vacuum tube pentode, but with much reduced voltage and power levels. Early MOS memory chips stored 256 bits, but were not used widely for main memory. Industry interest perked up when the 1024-bit device came out early in the 1970s. It had two major innovations. First, address decoding circuits—those that

Bipolar Semiconductor Static Memory Cell



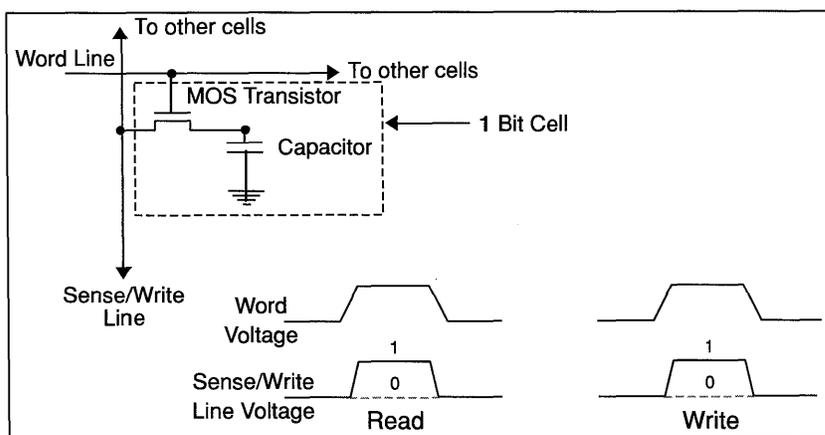
The bipolar semiconductor memory cell; its fast speed made cache memory feasible.

take the word address and select one each from the 32-word (row) lines and 32-bit (column) lines inside the chip—were duplicated on every chip. While this made the chip about 30 percent larger, it reduced the leads needed for addressing from 64 to only 10 which made the chip package much smaller. The second idea was to use one transistor and a small capacitor for each storage bit rather than the four to six transistors normally needed to make a flip flop. The capacitor stored one bit of data represented by the presence or absence of a voltage charge. The transistor acted as a switch to set or sense this voltage. This produced a very small memory cell. No current in the cell existed except very low leakage currents. Even a very small capacitor could maintain its voltage charge for many milliseconds. Refreshing to restore the charge was required periodically, but since 32 bits could be refreshed in parallel, only a small fraction of the time had to be used to do this. Designers quickly found ways to incorporate the refresh function into the computer timing with minimal impact.

One of the concerns with early semiconductor memory technology involved their reliability, since there were so many components and connections inside of each chip, and since quite a few chips were usually needed to make a complete memory. Since each chip read out only one bit for each read operation, single bit error correction was usually used especially in larger memories. In error correction extra bits are added to each word. For example a 36-bit word requires seven additional bits for error correction. During writing these are set based on the parity of various combinations of the 36-bit data word. Later when this word is read out, if any bit has changed, including the added bits, an error syndrome is formed from again parity checking the combinations. This error syndrome uniquely identifies which bit has changed, and the logic merely changes it to the opposite state. If two bits are in error, they cannot be corrected, but their presence can be detected. This scheme works well with semiconductor memory when each chip is only one bit wide. Then, any number of bits in a chip could fail, but no more the one bit for any address selected would be involved, and the

error correction circuitry will correct it.

MOS Semiconductor Dynamic Memory Cell



The one transistor MOS dynamic memory cell still provides high density and low cost.

Another early concern with semiconductor memory arose since their data was “volatile”. This means that when power is removed, the stored data is lost. Magnetic memory such as core and plated wire could retain their data with no power applied. However, most users did not rely on this feature

in these random access internal magnetic memories, and most of these memories were not designed to prevent incorrect writing during a nonscheduled power loss anyway. Backup to drum or disk was normally done, so this concern with volatility seemed to be short lived for most commercial applications. In very critical systems an uninterruptible power supply would probably be chosen.

The original 1K MOS memory devices were somewhat difficult to design with and required 20-volt power supplies and special drivers. Also, the core memory costs were still dropping. But when the 4K bit device came out about three years later, these design deficiencies were corrected. Standard five-volt drivers could be used, and the reduced cost per bit—due to the better than expected chip yields—put an end to core memory and other magnetic memory for almost all new commercial products. Second generation 4K devices added another feature—address multiplexing. Normally 12 signals would be needed to select an address for a 4096 by 1-bit chip. With address multiplexing, they were gated onto the chip, six lines at a time, cutting the address signals in half and reducing the package size. While multiplexing slowed the cycle time somewhat, the design community thought the tradeoff was well worth it since it allowed more chips to be packed on a printed circuit board.

The semiconductor memory technology continued to develop very rapidly after this with a new generation coming out about every three years and having four times the capacity of its predecessor. No longer was the main memory such a limiting part of the computer system. Let us return then to what was happening in the area of technology for the computer logic for the next generation of products.

THE 1100/80

Work began on the 1100/80 technology well before the 1110 system was completed. The objective was to get substantial performance improvements over the 1110. The solutions used were well conceived and set the basis for technology direction for the high end of the 1100 Series for sometime in the future. The major items involved an “ECL” type of logic circuit, controlled impedance printed circuit boards and backpanels that used automated design and layout to strict electrical rules, use of a cache memory organization, and a semiconductor main memory.

Controlled Impedance Printed Circuit Cards

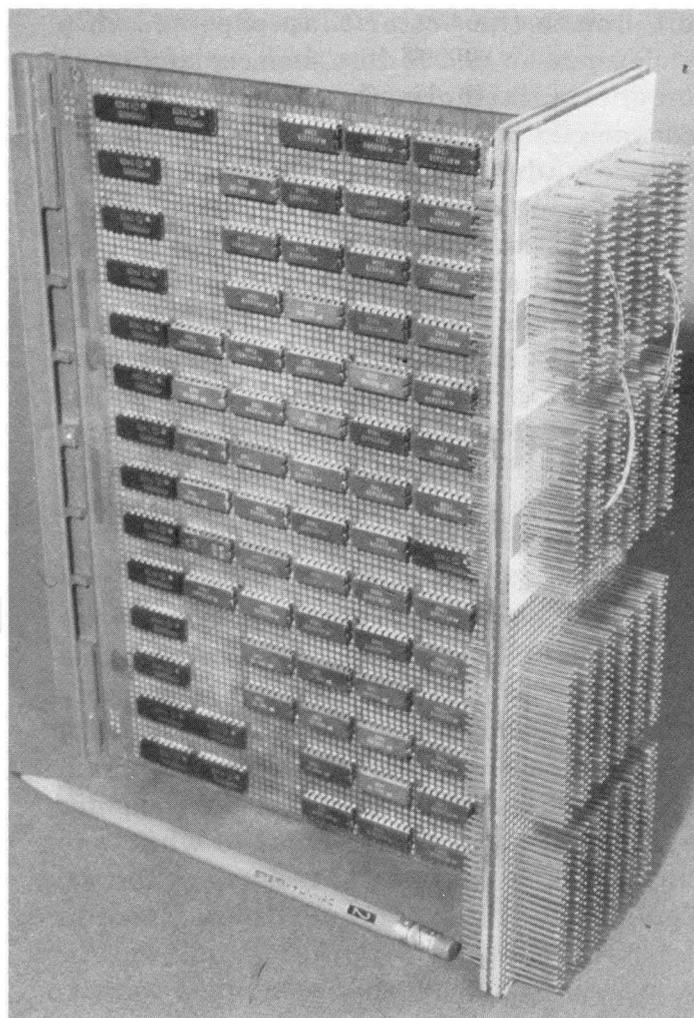
Experience on the 1110 had shown that if its circuits had been any faster, the interconnection and wiring between the chips would have become a limitation. The fastest electrical signals can travel is equal to the speed of light, or about one nanosecond per foot in free space. However in most wires and printed circuit boards the insulation material has a dielectric constant higher than free space, slowing the signal down to about two nanoseconds per foot. While most wires between chips were only a few inches, lengths of a couple of feet were not unusual. Faster chips were now switching in a few nanoseconds and getting close to the limits of interconnections. But the added delay at this time was not the most serious problem. A bigger problem was due to reflections and crosstalk which could produce noise and errors in the logic signals if not treated properly. The logic circuit and interconnection wires became part of a complex circuit that had to be treated as a single entity. Formerly the wiring was considered as a load on the logic circuit that looked like an electrical capacitor with a value that was proportional to its length. This was all right for the slower circuits. But as the transition time of the faster circuits approached the transit time of the interconnecting wires, new problems were encountered.

All wires have a characteristic impedance which is determined by their distance to other conductors in the region and the dielectric constant of the intervening space. If one applies a fast voltage pulse to one end of wire, a current will flow in the wire that is independent of what is connected to the other end of the wire—in fact it takes about two nanoseconds for each foot of wire before the far end “knows” that any voltage pulse was applied. The current flowing in the wire is proportional to the size of the voltage pulse, so the ratio between the voltage and the current is constant. This ratio is the “characteristic impedance” measured in ohms (since this is similar to ohms low for dc currents). The voltage pulse and its associated current pulse begin to propagate down the wire as a wave front. Now, if the far end is open with nothing connected to it, the current must be zero there even after the wave front arrives. This creates a discontinuity which then starts another wave that propagates back to the source, where it turns out another discontinuity can occur, and the signal will bounce back and forth many times only being reduced by the small losses in the wire and dielectric material. If the far end had been shorted, the voltage would be zero there, so another type of wave is propagated back to the source, and a similar thing occurs. Now if a resistor having a resistance equal to the characteristic impedance is connected at the end of the line, the current and voltage will maintain their proper relationship when the wave front arrives at the resistor. The wave front is thus completely absorbed, and there are no reflections rattling back and forth. We have a quiet line. (If the resistor is a little different than the characteristic impedance, the reflection is calculated by taking the difference between the two, divided by their sum. So a ten percent difference gives only about a five percent reflection which is usually quite tolerable).

So for fast moving logic signals traveling over

reasonable distances, the wires now had to be thought of as transmission lines. Most logic circuits outputs usually had low resistance so they could drive many loads well. Their inputs had high resistance so they would take less current and be easy to drive. This is the worst combination for a transmission line interconnection. Signals could bounce back and forth several times between two logic circuits before the signal settled to its proper level. So a wire with only a two-nanosecond one-way delay would take 16 nanoseconds for four 2-way trips! The engineers would say, “It isn’t the delay that kills you, it is the ringing”.

The other problem with the wiring was crosstalk between neighboring wires on a printed circuit board or in the backplane. If



The 1100/80 multilayer card and backpanel technology had many innovations.

one conductor was driven with a signal, it would induce a small voltage on its neighbor through capacitive (voltage) and inductive (current) coupling. This noise from several wires all passing signals at the same time could add up on a nearby single "quiet wire" and cause an error. Twisted pair wire, with one of the wires grounded, helped this quite a bit, but this was quite expensive and not completely effective, especially at the higher speeds.

What was needed was an interconnection wiring scheme that provided a controlled characteristic impedance and also limited the crosstalk to acceptable levels. The answer was a precision multilayer printed circuit board and backpanel technology. The latter would replace the wirewrap backplane that had been used. All interconnections between points had to be made by using a grid of X and Y lines running at a 90 degree angle between each other. (There is almost no crosstalk between perpendicular crossing wires). Each pair of X-Y lines was contained between a pair of ground (or voltage) planes which shielded them from the other pairs. The only crosstalk concern was from the parallel, adjacent conductors, and this could be easily controlled by varying the distance between the adjacent conductors relative to the distance to the ground planes. A smaller distance to the ground plane reduced the crosstalk, but it also reduced the characteristic impedance which would demand more current from the logic circuits and increase system power. The final design used 0.008 inch wide lines, with 0.012 inch separation, and a distance to the ground planes that gave a 50 ohm characteristic impedance and only a few percent crosstalk.

ECL Logic Circuit and Automated Card Layout

The selection of this new controlled impedance interconnect had a lot to do with the choice of the new logic circuit for this technology. TTL devices, such as those used by the 1100 and

quite common in the industry,, had a 3.5 volt logic swing. If they were to drive the 50 ohm interconnect, which had to be terminated at the end with a 50 ohm resistor, 70 milliamperes per output would be needed. The TTL parts couldn't drive this much current, and if they were redesigned to do this, the power consumed would have been quite excessive. Since this power varies with the square of the voltage, a lower logic signal would be quite advantageous. The Emitter Coupled Logic (ECL) circuit was selected for this. It is still used by almost all high speed computers built today. It has only a 0.8 volt logic swing and can readily drive 50 ohms. Both the output signal levels and input switch points are quite closely controlled, providing room for noise margins. The transistors in the circuit operate in the linear mode which keeps enough voltage across the transistor to get its full speed potential and avoid any "voltage saturation" problem. ECL circuits are also considerably faster than TTL circuits built with similar base technologies, so it became the obvious choice for larger machines.

The 1100/80 logic card size was 7 by 10 inches, and it held up to 84 logic chips. The chip used a 16-pin DIP package similar to that of the 1110. A new connector with 240 pins was developed of which 160 could be used for signals, a big increase over previous logic circuits. This helped support the amount of logic on the card. It took four signal layers (2X + 2Y) to interconnect the logic circuits, and with the ground and voltage planes, a total of ten layers were needed. Previously only two or sometimes four layers were sufficient. Sixty cards in two rows could plug into one 50 ohm multilayer printed circuit backpanel measuring about 18 by 21 inches. Four backpanels then fit into one cabinet.

One of the problems encountered was that the thickest backpanel that could be then manufactured could not contain sufficient wires to interconnect all the printed circuit

cards. This problem was solved by combining two backpanels. At every 0.10-inch location in both directions a plated through-hole was formed into each backpanel. One backpanel was placed on top of the other. Close tolerance square pins were forced through all the holes in the two backpanels and then were all soldered in one operation, interconnecting the two panels. The pins were made long enough to protrude on both sides. One side was used as the connector contact to mate with the printed circuit card, and the other was used for test probes and provisions for engineering wire changes. This concept was quite effective and has since been used extensively on products which require many complex backplane interconnections.

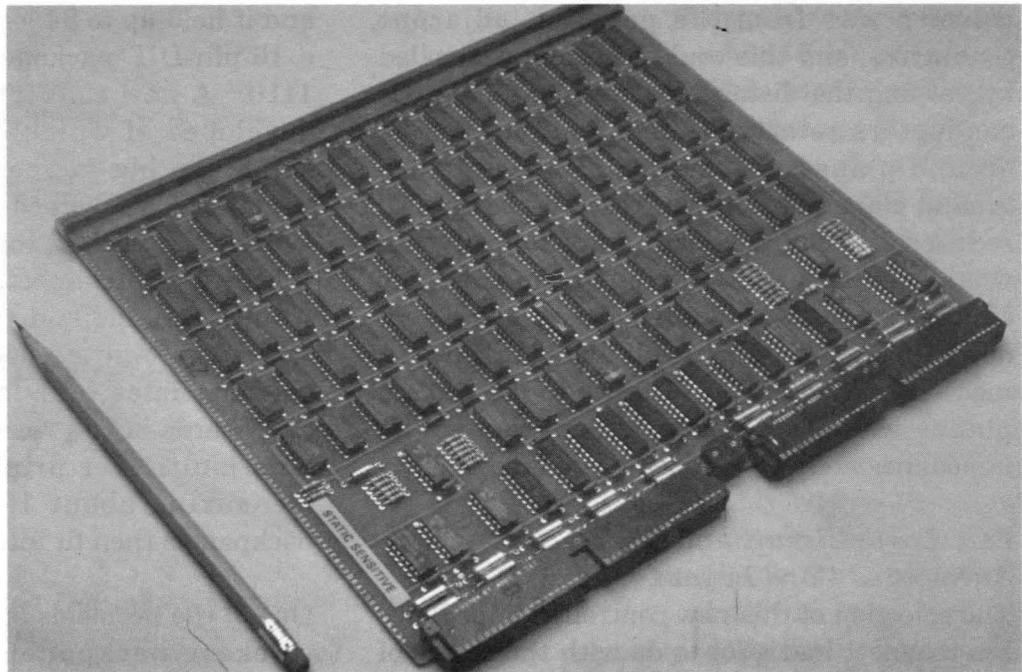
With so many chips on the printed circuit card, it was apparent that each could have its unique interconnection wiring. Furthermore, there were four layers of wiring, and strict rules had to be followed on how the various connections between one logic output and its various loads would be interconnected in order to avoid any transmission line problems. Since many of these boards had to be done in a short time period, it was obvious that placement of the chips on a board and the location and arrangement of the wires between them had to be done with the use of a computer. Software to do this was not available at that time, so it was developed internally. The output of the program drove a machine which exposed light on photographic glass plates which became the artwork masters

used to manufacture the board. Computer output also provided the data for board testing. The same program handled the larger backpanel as well. The existence of a design automation capability within the engineering department became a necessity, and its role was destined to become even more important in the future. However, it is beyond the scope of this paper to cover this important area in proper detail.

The printed circuit technology developed for the 1100/80 was named MLP (for Multi Layer Packaging) and was used on a number of other projects. It has also provided a technology basis for handling ECL logic that is still used today.

Memory Technology

As mentioned earlier, the 1180 was the first 1100 to use a cache memory. It was a shared cache memory in a separate cabinet; any processor in the system was able to access it. The actual term applied to this unit was the "Storage Interface Unit" or SIU. It contained up to 16,384 words, used a 256-bit semiconductor device with an ECL logic



Typical semiconductor memory card of mid 1970s; 16K chip provides 32K words/card for the 1100/80 and 1100/60 systems.

interface and a 45-nanosecond access time. The memory control used the same cards and ECL logic technology as the processor.

The main memory used the 4K bit storage device. Eighty-eight chips were mounted on a 10-by 10-inch printed circuit card. The memory chips had TTL type of signal levels and did not have to operate as fast as the cache memory, so a wirewrap backpanel was still used here. Although the main memory was slower than the cache, it would read out eight words at a time, so that the entire block could be loaded into cache at a faster rate. The eight words were transmitted to the cache memory two words at a time.

The 1100/80 System

The processor and cache memory each took their own cabinet with four backpanels of circuitry. The processor used 196 cards of 130 unique types with an averaged usage of 52 chips. The I/O section used the same TTL technology as the 1110, but had a new connector with more signal pins. The power system was supplied with 400 cycle power via a motor alternator, since that was felt to improve the immunity from any spikes on the incoming electric power lines. Another first on the 1100/80 was the use of a system support processor, the "SSP", a minicomputer used to monitor all the registers, or flip flops, in the system. This eliminated the need for the large control panels, containing hundreds of indicator lights, that had become common on large computers. The 1100/80 was about twice as fast as the 1110 and was delivered in 1976 as a dual processor system and later was upgraded to four processor capability. At that time the memory device was changed to a 16K chip, and two million words fit into a single cabinet - 16 times what had fit into one 1110 core memory unit and 500 times as much as the old 1103A! The space for a 4 X 4 system was reduced to 512 square feet compared to

864 for the earlier 1110 system. The 1180 was a very successful system. Over 1000 processors were delivered. It also created new technology techniques that set the foundation for many future products starting with the 1100/60-70 systems.

THE 1100/60-70

During the 1100/80 development, some computer engineers at the Sperry Research Center observed that microprocessor chips made with MOS technology were becoming quite capable and wondered what would happen if many of these were tied together in a single computer. After they completed a study project, their results showed then that the optimum cost-performance occurred with about four processors, partially because the memory became a bottleneck at much above that. It was also noted that it would be nice if something faster than MOS technology were used. ECL technology density was improving then, and a chip handling four bits of 16 primitive instructions had been developed. Nine of these together with other standard ECL logic chips would fit onto one 1100/80 type card to form a 36 bit ECL microprocessor. This formed the basis for a new product. Up until then the objective for a new processor design had always been to make a machine "bigger and better" than the existing ones. Here was a chance to make a smaller system, yet with reasonable performance, to broaden the usage base for the 1100. (At this time code names began to be used for major design projects with the final name—or model number—selected later at the product announcement date. The code name "Vanguard" was selected for this project; the 1100/60 nomenclature came later).

Four of the microprocessor cards were used in one processor together with microprogrammed instructions—a common practice in microprocessor chips—but the first time to be

used in an 1100 system. Microprogramming moves random control from logic circuits to memory technology, and with the progress in semiconductor technology a dedicated, small, fast memory for this was quite practical. Otherwise the basic technology and design automation tools already developed for the 1100/80 were a good match for the 1100/60, and this allowed a short development schedule.

The processor, with a private cache memory using a 1K bit chip, a main memory using the same 16K memory card as the 1100/80, and the I/O logic all fit into a single cabinet - another first for an 1100 machine. Microprogrammed computers are usually associated with rather low performance, but the four microprocessors running at ECL speeds provided an overall speed close to what the 1108 had, yet in a much smaller space and with a reduced price tag. A cost-reduced version without a cache memory was also made available.

were shipped—100 more than the original business plan. Later enhancements were made such as upgrading from a two to a four multiprocessor system and the addition of a denser main memory. At this time the system was renamed the 1100/70. The 1100/60-70 was a very successful family of products and found broad usage in the industry. Many 1100/60-70 purchases were from first time 1100 users. Nearly 4000 processors were delivered in all—a new high for an 1100 product.

THE 1100/90

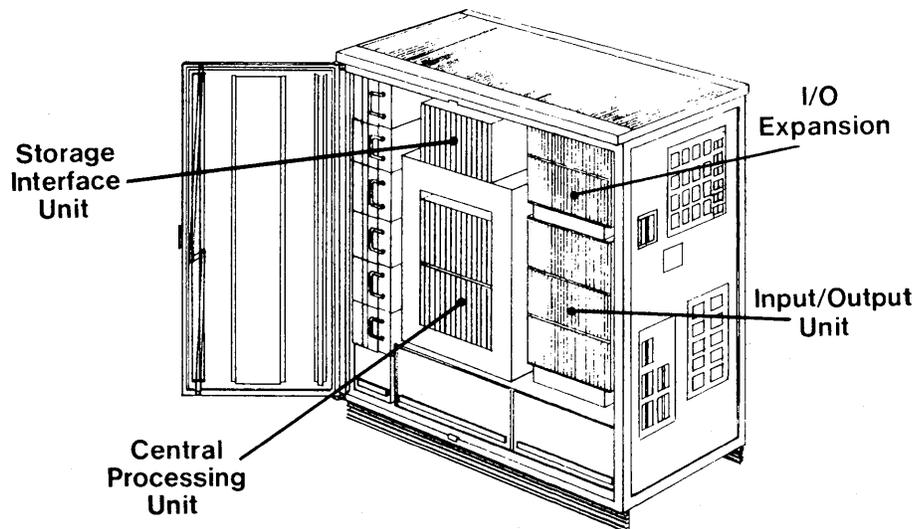
Technology Challenge

The development objective for the 1100/90 system was to improve performance over that of the 1100/80 by a factor of 3.5. At the time, this was considered quite a challenge. This represented the largest improvement ratio from one model over its predecessor since the 1108. The logic designers had some ideas on

how some significant benefits could come from logic organization—up to about 40 percent. This meant that the technology itself—the delay to do a simple function—still had to be reduced by about 2.5 to 1, or 40 percent of what it was in the 1100/80. An analysis of key 1100/80 logic paths showed that about one half of the delay was in the logic circuit itself, inside the chip, and the other half in the interconnecting wires

between them on the printed circuit boards and backpanel. Inside the chip, the basic logic

VANGUARD Central Complex Cabinet Back View



*The 1100/60 was the first 1100 central complex to fit into a single cabinet.
The memory module fits on the other side.*

The first 1100/60 was installed in 1979. In the first full year of production, 528 processors

gate delays were in the two to three nanosecond range. At about this time the newer improved ECL circuits were getting down to slightly below one nanosecond. But even if they could be reduced to near zero, the total delay could only be cut in half which did not meet the goal. Something had to be done to reduce the delay between the chips.

The LSI Question

The year was 1975 when these studies were going on—shortly before the first 1100/80 was delivered to customers. Back in the mid 1960s, when integrated circuits started to make their mark, visionaries quickly foresaw the future when many circuits could be made on a single chip. The term LSI, for Large Scale Integration, was coined. However, several years went by, and no one could define any significant LSI functions that could be made in sufficient volume except for memory chips, hand held calculators, and a little later the microprocessor chip. While the 1100/60 already exploited that, there was a lot of logic that still had to be done elsewhere. Furthermore, the 1100/60 organization was not optimized for the highest performance but rather lower cost.

Earlier the industry had discovered that a number of smaller functions such as parity checkers, decoders, and multiplexers could be universally applied to almost any computer architecture. Thus the term MSI, for Medium Scale Integration, was born as a practical back off from the highly touted—but hard to use—LSI. A big market for MSI followed especially in the slower but popular TTL circuit family. The original simple gate functions also needed a name, so SSI for Small Scale Integration was coined for them. The conventional agreement in the industry then was to refer to functions with 1 to 10 logic gates as SSI, from 11 to 100 as MSI, and from 101 to 1000 as LSI. For functions over 1000 gates, the term VLSI for Very Large Scale Integration was adopted.

The question quickly arose: Would the use of high speed LSI chips—containing more logic circuit interconnections within themselves and thus fewer left to be made on the relatively larger printed circuit boards—allow attainment of the 1100/90 performance goal. In order to address this problem of how various technology alternatives would measure up to meeting the performance goal, a mathematical model was constructed representing the basic delay elements in a typical logic path. A logic path performs one sequence of perhaps 10 to 20 logical decision steps that must be done within a basic computer cycle. The processor contains thousands of such paths. The inputs to the model were how much of the machine logic would be in LSI, the basic speed and number of gates in the chips, the distance between chips, the number of chips in a printed circuit board, etc. The output was a value representing a typical path delay for various number of logical steps in the path. The 1100/80 technology parameters were run through the model to establish a baseline reference value to compare with.

By then the performance target of the next generation internal gates speeds on the chip was being bracketed: below a nanosecond and possibly down to 0.5 nanosecond or 500 picoseconds. But there was the question of how large a chip should be used. All semiconductor wafers experience some unavoidable number of defects originating during the fabrication process. Larger chips are more likely to experience a defect, so if the chip is too big, yield is low and the cost of the good chips becomes excessive. Bipolar ECL wafers have more defects than those using the slower MOS technology since their process is more complicated. A chip 0.2 inch on each side was selected based on experience of the semiconductor manufacturer. This allowed the gate capacity to be established and the number of signal leads required from the package. The other part of the problem was

determining the number of unique type of LSI chips that could be designed during a reasonable schedule. The decision was made to have a mixture of LSI chips of about 30 unique types and do the rest of the processor logic with the general purpose SSI and MSI types. The model showed that if 70 percent of the logic could be made to fit into the LSI parts, the performance goal could be met based on some assumed printed circuit and other packaging parameters that had been established in a separate study. A logic design study was started. Within a short time the study group designed a number of new general purpose LSI logic functions, and it was estimated the goal could be met. Later some improved functions were designed, and the final LSI content in the 1100/90 approached 90 percent.

At this time the semiconductor manufacturers were not interested in designing custom or special purpose LSI chips for individual computer manufacturers. They had a limited number of chip designers and wanted to sell a hundred thousand chips or more per year of every type. The company decided it was time to develop the knowhow to design LSI chips for the 1100 architecture. It selected several circuit design engineers and layout technicians to learn this technology. At this time an internal company semiconductor manufacturing capability did not exist. So an agreement was reached with a semiconductor manufacturer—one which had the best ECL technology at the time—to obtain his mask design rules. This allowed the 1100 chips to be designed and then be made with the manufacturer's standard process.

The 1100 logic designers could define the functions they desired, circuit designers would convert these to electrical schematics, and the layout technicians would design the masks. The masks, with the aid of a lithographic process, are used to define the actual arrangements of all the components on the

chips as they are being manufactured. At that time there were many stories in the industry about the difficulties of designing LSI chips and getting them to work properly even after many cycles of testing, redesign, and refabrication, which could take several months per cycle. Some designs never did work, and were abandoned. Others worked under only limited conditions or were plagued with very low manufacturing yields. Most of these circuits were custom designed to do a particular function. Years earlier computer people had learned that the best way to build a computer was to design only a small set of simple circuit building blocks that did primitive logic functions, and then use logic designers to specify how these should be interconnected so as to perform the functions of the various sections on the computer. The logic designer dealt with Boolean logic—a world of ones and zeros and clock pulses—and didn't have to know transistor physics or electrical circuit operation. Computer simulation and verification of large sections at the logic design level is also more practical than at the circuit design level.

Other computer manufacturers came to the same conclusion about this time. The answer was the "gate array" chip—so called because it was built from a physical array of simple logic gate circuits. Each unique design would merely interconnect these gates together much like the backpanel wiring used to connect the early printed circuit cards containing their simple AND and OR circuits. Besides making the design easier, all the various types were manufactured identically until the last few steps where the metal was added. The way it finally worked out, actual gates were not laid out. Instead a common arrangement of transistors and resistors were used which gave more flexibility. These could then be interconnected to form a predetermined, limited set of gate type functions rather than just one type. So the metal patterns—only two layers then—would interconnect the transistors

and resistors to form the selected gate function, and also provide the interconnections among the gate functions as well as the I/O leads leaving the chip. Thus a unique chip design was formed, yet made from a common base component.

The training of the LSI designers was completed after they spent some time at the semiconductor manufacturer's facility becoming familiar with the design rules—they already were quite expert in ECL circuitry. It was mid 1977. The basic chip design was then completed, masks built, and the first test parts were received in early 1978. At this time the parts were made on two-inch diameter wafers and contained about 100 chips. Each chip had 168 basic circuits; 144 internal and 24 for outputs. The internal circuits were clustered in fours to make 36 “cells”. The total of these circuits could provide the equivalent of up to about 350 simple gates depending on the function performed.

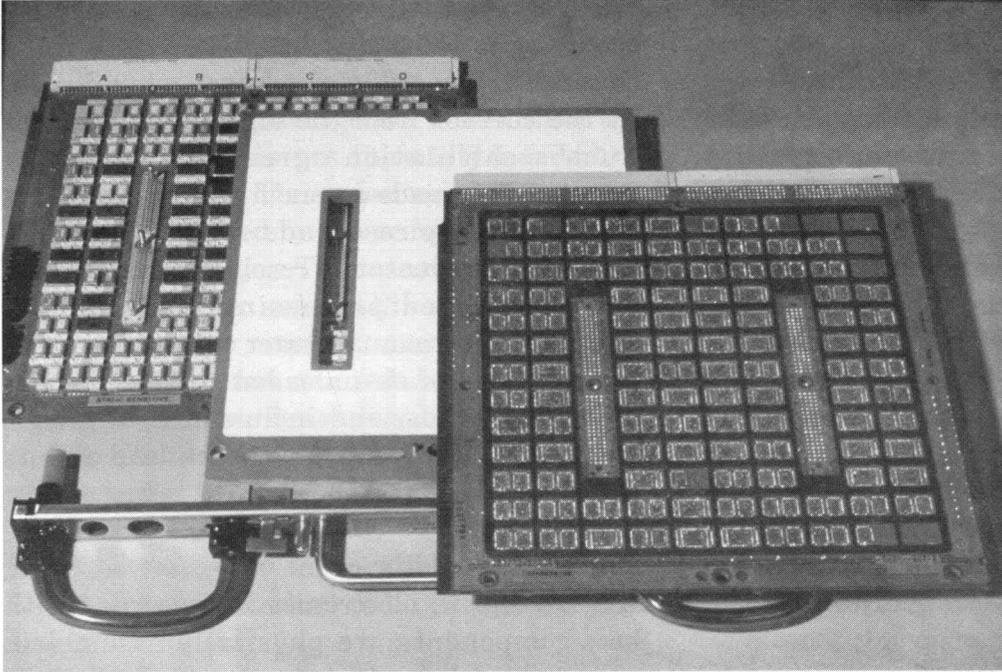
The performance target was 500 picoseconds for a circuit driving three internal loads. The tests showed that with only one nearby load the delay was 500 picoseconds, but with three loads at more remote locations on the chip the delay would rise to about 900 picoseconds. We quickly learned that even though the wiring is very short on a chip compared to printed circuit boards and that the capacitance of an internal transistor is very small, on chip slowdown was real. Quick calculations verified this. We had been led to believe by the semiconductor manufacturer to expect 500 picoseconds but had not pinned down the actual conditions for this. For circuits within a chip, one cannot use the two nanosecond per foot delay rule as on a printed circuit card. This would be only 33 picoseconds to travel a 0.2-inch distance on the chip. This only works when the driving current is large enough so that it can charge the characteristic impedance of the line a full signal swing. This would require about 20 milliamperes which is

much higher than practical for logic circuits on a chip. Even if one tried this, the resistance of the wiring on the chip—which is much higher than a printed circuit card because of its very small cross section—would prevent a large current flow. With the lower drive situation the delay is about equal to the capacitance of the load times half the voltage swing divided by the current from the logic circuit output. Such a calculation agreed with the test results. This made a person realize how fast a few hundred picoseconds really are. This problem was eventually resolved by the help of some improved processing made by the semiconductor manufacturer that reduced the typical delay of an unloaded circuit to about 350 picoseconds, and inclusion in the logic usage rules delay values for each load and its associated wiring on the chip.

Packaging

“Packaging” in electronics is concerned with how components are physically assembled, mounted, cooled, and interconnected together. Packaging was becoming a key factor in computer performance. In attempts to increase cycle time, the time for signals to propagate between the components became a major factor. To meet the 1100/90 performance goals it was felt important to fit the entire computer along with its cache memory into a single backpanel. The 1100/80 had used four backpanels contained in separate cabinets for each of these functions. The use of LSI circuits would reduce the physical size but not enough. This was because the technology then did not allow enough logic gates on each chip. In addition, in order to keep the number of unique LSI types to a manageable level, there would still be quite a few SSI/MSI low complexity parts used. A higher density packaging approach was needed. The result was the HPP (High Performance Packaging) technology. Up until then, connections and the feed-through holes on printed circuit boards were usually made with 100 mil (0.10 inch) spacing. The HPP technology reduced this to

50 mils resulting in a smaller board with chips closer together. This also required new 50 mil packages to be designed for the LSI and MSI parts.



The dense 1100/90 card pair used a 50 mil grid, LSI and MSI parts, and liquid cooling.

The new 54-pin LSI package was the same size as the older 16-pin DIP, as used in the 1100/80, and the 24-pin MSI package was half as big. Up to 114 LSI packages or 228 MSI packages, or any combination of the two, fit onto a 10-inch square printed circuit board. Because of the great deal of logic on a single card, 650 connections were provided from three sides of the card. The backpanel and two "sidepanels" remained at the 100 mil grid. Each LSI chip dissipated about five watts, and the MSI parts up to one half of a watt. High speed memory chips for the cache and GRS were also mounted in the MSI packages, and they dissipated about one watt. A single printed circuit board could involve up to about 500 watts. Since close spacing was needed in order to fit all the cards into one backpanel, liquid cooling was selected since it requires much less volume than air cooling. For a given temperature rise, a cubic foot of water

can absorb 3500 times more heat than air. Chip reliability is strongly affected by its operating temperature; a reduction of 10° C can reduce failure rates about 30 to 60 percent. The goal was to maintain chip temperatures at an equal or lower values than in the past, and this led to the liquid cooling decision.

In the final 1100/90 design, about 2000 LSI and 4000 MSI/memory parts were used. Fifty-two cards were needed to perform all the functions of the processor and cache memory. These were assembled in card pairs with each pair sharing a cold

plate cooling frame which contacted the back of the chip packages. The LSI chip was attached directly to a metal heat spreader inside the package for efficient heat conduction to the cold plate. The cold plate was made with a conforming thermally conductive material and ensured good contact with all the chip packages. Each card pair with its cold plate was mounted on one inch centers and plugged into a backpanel that measured approximately 12 inches by 27 inches. It was made from three individual panels pinned together, with one panel being used only for power and ground. The printed circuit board used up to 12 layers for interconnection and a total of 22 layers including those for power and ground.

I/O and Memory Technology

The I/O and memory sections were implemented in a modified 1100/80 MLP style

technology. The card size was increased to 10 by 10 inches which could then hold up to 100 MSI chips. The logic was done with MSI chips. The main memory chip provided 64K bits of storage and 88 of these, together with associated driver chips, fit on one printed circuit card yielding 128K words. Due to the increased performance of the 1100/90, additional channels and memory modules were needed in order to keep the processors busy. One memory cabinet contained four million words with a maximum of four cabinets, or 16 million words, per system.

1100/90 System

The first 1100/90s were completed in 1983, and met the ambitious performance goal of three to four times the 1100/80. The 1100/90, or the "Cirrus project" as it was called, was also the first system using the upgraded "C Series" 1100 architecture containing increased memory and "new mode" instructions. Because of this complexity and the challenging new dense packaging, the 1100/90 development schedule was longer than planned. However, it was a very successful program with about 20 percent more processors being manufactured than the 1100/80 even though the 1100/90 was in production for a shorter time period.

THE SYSTEM 11

The 1100/60 System was very successful in expanding the 1100 base of users. Many users liked the capability of the 1100 system and its software, but did not need the performance of the 1100/80. They chose the 1100/60, with its lower cost. The question arose: What could be done to further reduce the cost of the entry level 1100, yet maintain acceptable performance and capability?

New LSI Chip

At this time the cost of MSI logic circuits had pretty much bottomed out and was mostly dominated by their assembly and package

costs which weren't expected to drop much further. And the cost of a printed circuit board and its assembly exceeded the costs of the MSI components. We had seen LSI help the performance in the case of the 1100/90. LSI seemed to be the only way to also attack the cost problem but probably using something different than ECL which was more expensive and less dense. It was thought that the use of LSI would reduce cost because the number of printed circuits would drop greatly as would the power consumption and cabinet hardware required. At the same time a maximum delay per gate of five nanoseconds was felt needed to provide adequate performance. The metal oxide technology (MOS) was considered, but at the time in 1978 its delay was closer to ten nanoseconds. Here we are referring to the delay value the logic designers use, which must include deviations for manufacturing tolerances, temperature and power supply variations as well as additions for driving other loads and wiring on the chip. Quite often this value is two to three times larger than that quoted by technologists (known as the "advertising" value) which is simply a measurement of a typical gate driving one load very close by, under average conditions.

Shortly before this time, company management authorized the establishment of a prototype capability in semiconductor fabrication, and later its expansion, to include full manufacture of LSI components. Also at this time there were very few gate arrays available from outside suppliers or design automation tools to support chip design. While in the case of the 1100/90 the types of LSI chips were limited to about 30, it was felt that virtually all the logic had to move to LSI to obtain the full cost-reduction benefits. It was estimated that this would involve 100-200 different chip designs. Obviously, automated error-free design would be required to do this. A simple gate structure arrayed across the chip seemed to be the best way to accommodate the automation. Also a low

power circuit was desired which could be air cooled. A simple four-input NAND circuit was selected which was made from two transistors and three resistors and operated on only a two-volt power supply (compared to the standard five-volt supply). A NAND circuit performed the “NOT-AND” function which is an AND followed by a logical inversion. It turns out that any logical function including flip-flop registers can be constructed from a NAND—or its cousin the NOR circuit. In theory, this meant that all the chip design could be quickly automated once the logic design was completed. At the time there were no such tools available, so a special project was started inside the Computer Aided Design (CAD) department to develop such a capability.

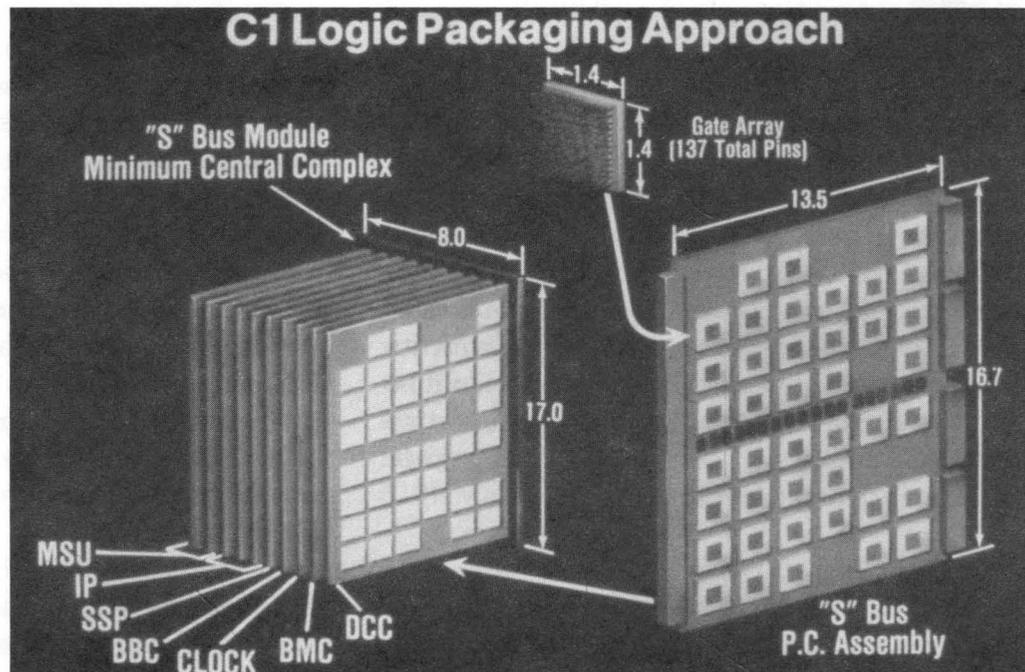
A major issue in the LSI technology definition was the selection of the chip size, the number of gates and I/O signals that would be compatible. It was also important to provide sufficient wiring on each chip to ensure that the CAD tool could complete all the layouts in a short time. Furthermore, the length of the wires on the chips could affect their delay to a considerable amount, but the logic designers needed delay values early in the design, not after the chips were laid out by the CAD system. The project, code named “Chaparral”, was on a short schedule, there was no time to run a lot of trial designs through evaluation, and the CAD tool wasn’t ready yet anyway.

The answer came from a combination of data presented in some research papers

combined with some mathematical models developed internally. This allowed a prediction to be made of the average wire length which then was used to calculate the number of wiring tracks on the chip, and also the “statistical” minimum and maximum wire lengths that were needed for logic delay estimations. The final result was a gate array of about 1000 gates, 120 signals and measuring 0.25 inch on a side. The package measured about 1.4-inch square, mounted on two-inch centers onto a 17 by 19-inch printed circuit board. Worst case delay was about two nanoseconds plus one for each load and its wiring. Typical values were about 70 percent of this.

System Bus

Another problem that had to be solved on the System 11 was limiting the number of signal leads leaving the printed circuit board which could contain 50,000 to 100,000 gates each. Normally in a computer system, each major section must communicate with all other sections. In high-performance systems this should take place concurrently so parallel and separate connections must be made between all units. These are called “radial” connections



The System 11 S bus; a dual processor system could fit into one backpanel.

since they appear to radiate out from each unit. Another method uses a "bus" organization wherein all units communicate via a set of common wires known as a bus. The bus contains the data to be transmitted and also additional lines which control the data flow. A principal advantage of the bus organization is that it greatly reduces the number of signal wires leaving each unit which now only have to connect to the bus, not all other units. The primary drawback is that the bus limits communication between units to only one transmission/reception for each bus cycle time. This drawback can be minimized with a higher speed bus.

The System 11 was the first 1100 to use a bus organization for its main system interconnections. It was named the "S" bus. Another lower speed bus, the "L" bus, was used for I/O sections. Cards to provide the maximum configuration of three processors, all the associated memory, and I/O circuitry fit into two backpanels. A section of the bus was in each backpanel with a special cable connecting the two for the maximum system. Most configurations, however, needed only one backpanel.

The bus was treated electrically as a transmission line. In order to maintain data integrity at the 100 nanosecond cycle time of the system, it was important to terminate the bus and avoid any reflections. Since a driver and a receiver can be at any location in the bus, both ends of the bus have to be terminated. The extra electrical capacitance caused by each card plugged into the backpanel bus has the effect of increasing the delay of the bus and also of lowering its characteristic impedance. With all the cards plugged in, the delay went up to about four nanoseconds per foot and the characteristic impedance dropped to about 25 ohms. A driver looking into the bus sees two sections in parallel, and thus must drive about 12.5 ohms. This low effective impedance meant the bus

driver had to deliver a large current. In order to minimize this, a special low signal swing driver was designed and a compatible receiver circuit. The S bus was bidirectional which means that any wire in the bus could send data in either direction. A driver-receiver pair, or "transceiver" circuit, was located at each printed circuit connection point. A special LSI chip was developed to implement these circuits. This was done by modifying the System 11 gate array and adding 24 transceiver circuits across one side of the chip and using the same package.

Memory Technology

The System 11 main memory used the 256K bit DRAM device. One million words of storage with all the required control logic fit onto a single board. There was no cache memory. The GRS function was done with a MOS ram chip. MOS technology also provided the circuits for the microprogram storage.

System 11 Firsts

The System 11 provided many firsts for an 1100 product: 100 percent LSI for logic in the processor, memory and I/O; system level bus organization; complete one million word memory on a single card; 1100 in a low cabinet capable of operating in an office environment; and also a complete dual processor system with memory and I/O in one backpanel. The System 11 was also marketed as the "MAPPER 10" product. Shipments began in 1984. While total System 11 sales were less than planned, the S bus organization and many of its I/O modules were carried over into the later 2200/200 and 2200/100 systems and are still in production today.

THE 2200/200

Background

The 2200/200 product had an unusual start and method of implementation. For a top management conference to be held in January

of 1983, several people were assigned a question which they were asked to study and report on: "What should the company do relative to the emerging (and successful) microprocessor technology". The answer came back: Apply the technology and design techniques to an 1100 machine. A rather optimistic report presented what could be done in a short development project. The President of the division approved the project on the spot. The effort, which became known internally as the "Micro 1100", started about two weeks later with a small dedicated team of logic and chip circuit designers, who already had been thinking about how the latest CMOS technology might be used in an 1100 machine, together with some CAD people.

CMOS Technology

As mentioned earlier, MOS technology had been around quite a while and was used for mostly microprocessor chips and memory devices. MOS transistors can be of either a "P" type, made with positively doped semiconductors, or "N" type, with negatively doped material. The earliest chips used the P type material which switched slower, but was easier to fabricate. However by the late 1970s, N type yields were good and became used almost exclusively. Several transistors connected in parallel could form a NAND circuit with another transistor serving as a resistor connecting between the power supply and the common output of the transistors. The transistors would discharge the output wiring capacitance and the resistor would charge it if all the transistors were turned off via their inputs. When any transistor was on, a constant current would flow from the power supply through the resistor and transistor to ground. To improve speed, this current would be increased, but this also increased the circuit power dissipation. The CMOS technology uses both P and N type transistors (the "C" stands for complimentary). One type charges the output net, and the other discharges it, but only one or the other is on at any time (except

for a brief period during which they interchange states). This means that after a logic network finishes switching to a new state, no additional power is dissipated. CMOS technology is also easier to design with since one need not keep track of the effect of these internal currents flowing in the chip. While CMOS required more transistors than NMOS and was somewhat slower, it appeared to be an attractive technology. This was because rapid progress being made in scaling down transistor sizes was improving both density and speed. Some of this could be sacrificed in order to keep power manageable and ease design complexity. CMOS was selected for the Micro 1100 and shortly thereafter became the preferred technology in microprocessors and other general applications in the industry that didn't require the highest speed.

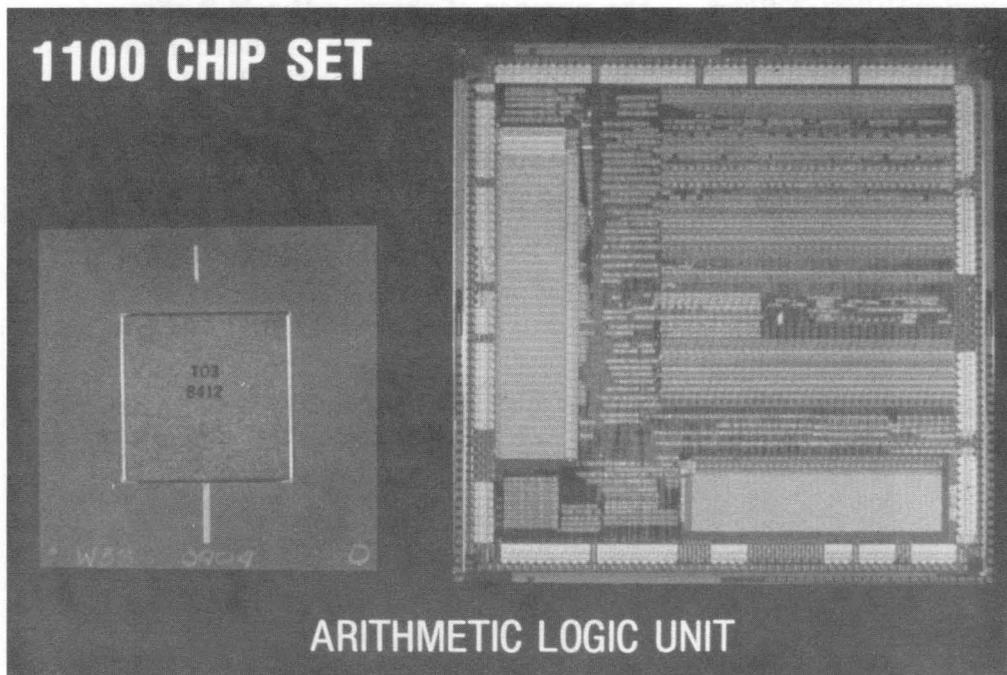
The Micro 1100 Chip Set

The basic Micro 1100 consisted of four chips which could perform all the 1100 functions. Two optional chips were added a little later: one for some enhanced instructions and the other for speeding up multiply and divide operations. These chips averaged over 100,000 transistors each; in order to pack this many transistors onto a 0.4-inch chip it was important to optimize the logic and circuit, transistor, and interconnection design. A custom layout for each chip, rather than the automatic gate array layout method, was needed. A logic designer—familiar with the 1100 and its internal operation—and a "silicon" designer—familiar with the CMOS circuitry needed to make various logic functions—formed a team for each of the chips. A new design "language" to describe the design details was developed as were a number of custom design automation tools necessary to verify design correctness and internal timing. Besides the logic functions, these chips contained a number of internal memories, as well as a read-only-memory (ROM) for the control store which contained the microprogram

instructions. ROM was used here since it occupies considerably less space than a read/write type of memory. This put a large demand on accurate logic simulation, since changes in the microcode could only be made by building new chips. The final chips measured about 0.375 inch on a side and were mounted in a package with 224 leads, measuring about two inches square.

Because of this we tend to think of both series as a continuum and call them the 1100/2200 series or sometimes "1100" is still used to refer to both). As implemented the 2200/200 continued use of the System 11 bus and I/O modules. A cache memory was added, and a new main memory using CMOS gate arrays, filled out the central complex. Performance was over double the System 11, putting it near the 1100/60 range. The chip set, like most

microprocessor designs, did not have the many checking features of mainframes. Consequently, this capability was obtained by providing two complete chip sets and a compare of their operations was performed. All this, and supporting cache memory fit onto a single printed circuit board—the first time for an 1100/2200 machine. It was interesting to note that the GRS function, which needed a thin film memory stack



The Arithmetic Logic chip from the Micro 1100 chip set; the first CPU use of CMOS.

The 1100 Desktop

One of the applications for the Micro 1100 chip set was in a desktop 1100 workstation. This design project, known internally as "Orion", was completed with an operating prototype, but was never brought to production.

The 2200/200 System

The other application of the chip set was a small mainframe. This project, code name "Swift", turned into the 2200/200. (At this time new versions of the 1100 Series were to be named the 2200 Series. While this was meant to indicate evolutionary enhancements over the older systems, the new systems were fully compatible with the older 1100's.

and numerous support drivers and sense amplifiers in the 1107, or 128 cards in the 1108, now fit neatly onto one part of the 2200/200 Arithmetic/Logic chip! (It also ran a little faster than the 1108.)

The first 2200/200 was delivered in 1986. It will be remembered for its pioneering use of CMOS technology. Details of the chip set were reported at the 1986 International Solid State Circuits Conference held at Anaheim, California. Today, nearly one-thousand 2200/200s are in use, and the Micro 1100 continues to be used in the 2200/100.

THE 2200/600

advantage of the “standard cell” approach is that the library design is not limited to fixed transistor sizes and fixed placement as in a gate array. The primary drawback is that each chip design involves a complete new mask set of over ten layers, not just two new metal layers.

New I/O and System Bus

The 2200/400 also employs a new “M” bus. Separate, small bipolar bus transceivers are mounted near the printed circuit card connector. A new I/O section using CMOS technology was also developed.

Memory Technology

The 2200/400 memory uses the one megabit chip. Further density improvement is obtained with small, dense “daughter” cards containing 16 small outline memory chips surface mounted on both sides. A single board contains four million words of storage and all the memory control logic.

System Comparisons

The 2200/400 offers striking improvements over the very successful 1100/60-70. The 1100/60-70 used over 90 cards for the processor and cache; the 2200/400 has only one. Similar reductions occurred in the memory and I/O section. Even with two to three times the performance of the 1100/60-70 and twice as much memory, the space required for a typical 2200/400 dual processor central complex dropped from about 72 to 22 square feet, and the power consumption fell to only about ten percent of the 1100/60-70 value. In addition a single, low cabinet can contain up to four processors with 16M words of memory.

Delivery of the 2200/400 began in early 1989, and active production continues at this time. As of September of 1990, nearly 1000 processors had already been delivered. Indications are that this system will continue the success story of the mid-range systems first started by the 1100/60.

Technology Selection

Planning for the next largest machine, the one following the 1100/90, commenced before the first 1100/90 was built. This is normal. Technology selection and development for a large scale machine can take about two years and must be done before detail design can begin. In the case of the 1100/90 follow-on, two aggressive technologies were evaluated and dropped, before a more conventional evolutionary one was adopted. Both of the technologies which were shelved involved attempts to get more from the silicon circuitry than otherwise could be obtained by packaging the logic circuits closer together. The alternative is to wait a little longer for denser semiconductor circuits which can accomplish a similar improvement, and let these drive any packaging design changes needed to accommodate the semiconductors.

The first approach that was evaluated for use involved placing a number of uncased chips close together in “multichip” packages, which also contained the interconnection wires between the chips. While versions of this technology still appear interesting for the future, it was not used for the 1100/90 follow-on because the time to get it ready for production and prove its reliability would have taken too long.

Another approach that was seriously evaluated involved a technology being developed by Trilogy, a new, but well funded and staffed organization that was located in Silicon Valley. Their approach used “Wafer Scale Integration” or WSI. The idea here is rather than using conventional means of cutting up a complete wafer into many small chips which are later tested and interconnected, leave them on the wafer where they are already quite close and interconnect them with the same metal lines that are used to make the chips. The challenge is to come

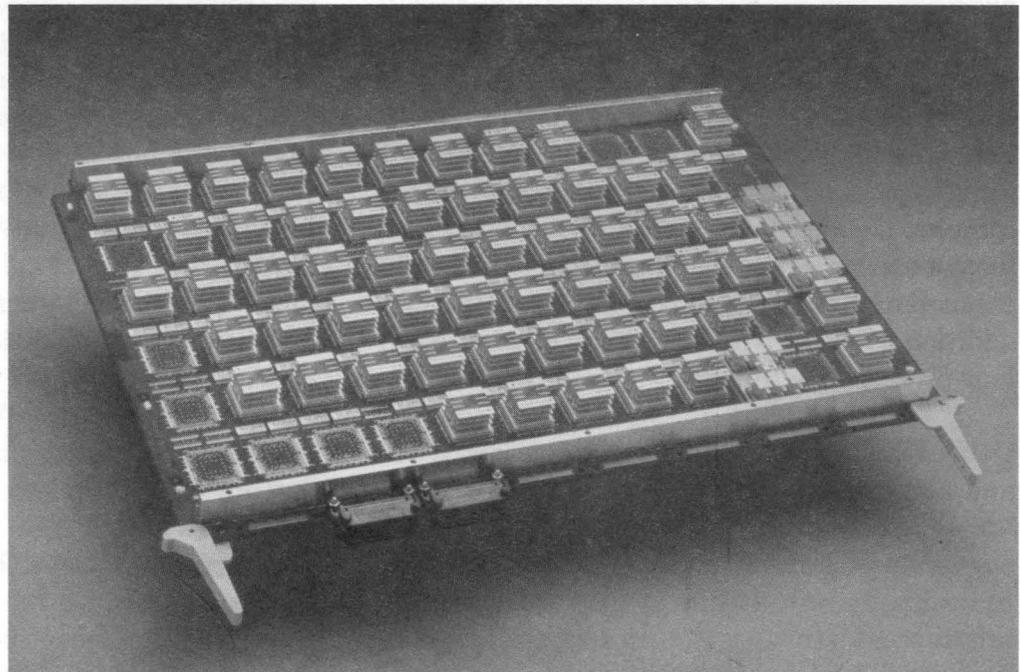
up with a means of circumventing the dozens of flaws and defects normally present on all wafers and which usually set the maximum practical chip size. Trilogy had developed a method to do this using redundant circuits. Sperry obtained an option to license this technology for both design and manufacture. The semiconductor facility at Eagan, Minnesota had most of the equipment needed for fabrication. Evaluation of the Trilogy technology began in June of 1983. This consisted of the design of a small evaluation test vehicle that Trilogy was to build, and the paper evaluation of a complete processor design. During the next few months, a considerable amount of detail was learned about this new technology. However most of this cannot be disclosed here due to agreements with Trilogy. Suffice it to say that within about eight months it became apparent that the cost/performance goals of the technology were not going to be obtained, and the manufacturing risks were fairly high. During this time period there was considerable interest in the technical press regarding this WSI technology and rumors surrounding progress of its development. Much of this was erroneous information. For example, one reporter wrote, "The technology is running so fast, they are having trouble cooling it". Actually, if one were to stop the clock, the power dissipation hardly changes in this type of technology. And while the wafers did dissipate a large amount of power, the Trilogy cooling system was very effective, and was not one of the problems.

The original goal for

the 1100/90 follow-on was a performance of three to four times the 1100/90. It was finally decided that such was not obtainable in the desired schedule. In the meantime, ECL LSI technology had progressed in both density and speed—as had printed circuit board capability. According to the performance model used for 1100/90 technology selection, these could now provide about two times performance improvement. Since this was still in keeping with what is common from one generation of large system to the next, it was decided to use this for the next high performance system. At this time two separate and fully staffed design teams were set up. The first team started the 1100/90 follow-on, named the 2200/600. The other started the next generation of machine, with a technology suitable to its timeframe, and it is now progressing well in its development. This dual project setup was possible since the 2200/400 was being done at the Blue Bell development facility, and other major projects had been completed at the Roseville location.

Processor Technology

The 2200/600 system involved a new processor and memory and continued the 1100/90 I/O



The 2200/600 ECL VLSI processor card.

subsystem. The processor uses an ECL LSI gate array with 1500 circuits—almost nine times that in the 1100/90 and equivalent to about 2500 gates of logic. The delay is about 60 percent of the 1100/90 LSI devices. Automated layout of the gate arrays was done allowing all the logic to be in the LSI parts. Over three hundred unique types were used to provide a total of approximately 850 LSI chips mounted on 18 printed circuit cards. These held the processor and the cache memory which was expanded by four times to further help performance by improving cache hits. A special package containing one gate array and eight ram chips was used for the memory functions. The GRS function was implemented with the gate array.

Memory Technology

The 2200/600 memory was completely redesigned to include 100 percent ECL LSI. A one-megabit memory chip was used. Sixteen of these were attached to both sides of a small daughter board. This dense packaging provided eight million words of storage, and related control logic on a single printed circuit card. The complete memory system of 16 million words—replacing four cabinets on the

1100/90—now easily fit into one cabinet. So an option with two separate 16 million word memory subsystems in the same cabinet was provided.

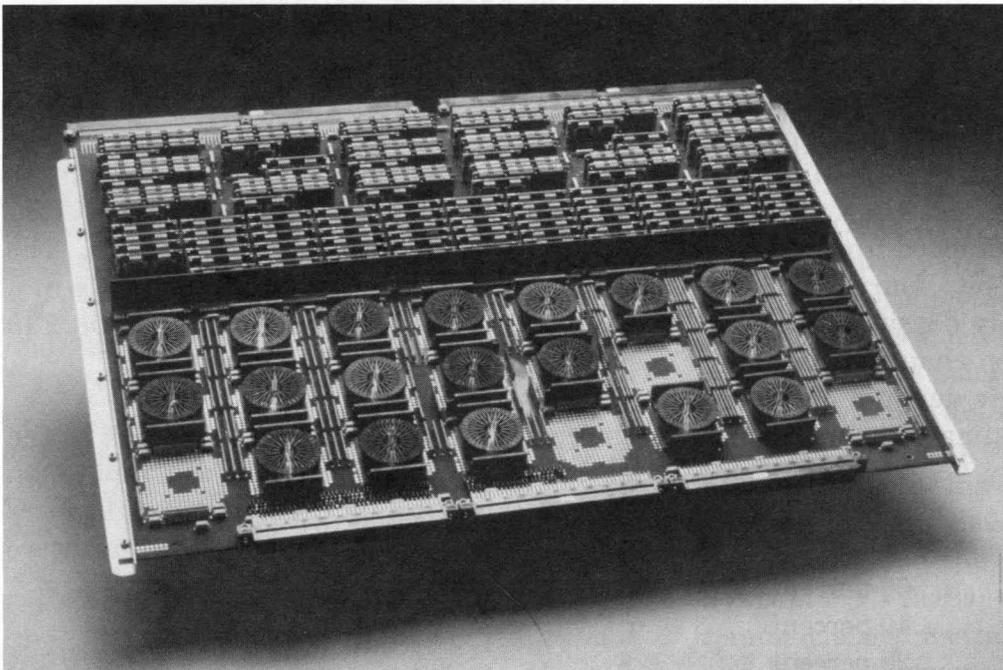
2200/600 System

The first 2200/600 systems were delivered in December, 1988, ahead of schedule. Very extensive computer simulation was performed, and this significantly reduced design and test time compared to the 1100/90. Performance met the two times goal over the 1100/90. Field reliability of the 2200/600 system has been impressive and has set new standards for large mainframes.

THE 2200/600ES

The 2200/600ES was announced in 1990 and is the current large 2200 system in production. It consists of evolutionary enhancements over the successful 2200/600. The CMOS technology of the 2200/400 was adopted for the I/O. This resulted in a reduction in the printed circuit card count per I/O subsystem from over 200 to 25. New power supplies and cabinet technology were also used. The

processor and main memory now fit into a single cabinet. The 2200/600ES has substantially reduced floor space and power requirements. For example, compared to a dual processor system 1100/90, floor space is reduced by nearly two to one and power consumption over three to one.



The 2200/600 eight million word memory card.



The high performance 2200/600ES system cuts required space and power.

A SUMMARY LOOK AT THE TRENDS

It would be difficult to find in the history of mankind a technology that has had the rapid progress and growth as that of computer hardware. A large part of this has been the result of improvements of electronics, especially semiconductors. We have gone from a vacuum tube chassis on the 1103 containing about 25 gates to a VLSI chip on the 2200/400 with over 25 thousand gates. Gate delay has dropped on the 2200/600 by a factor of 1000. The five-inch square 4096 core array on the 1107 has been replaced by a million bit memory chip measuring less than an inch. And memory costs have dropped from a dollar per bit to millicents per bit. During the last twenty years memory chip density (measured in bits per chip) has been changing the most rapidly, followed by logic density, and then logic speed. The approximate compounded annual improvements are:

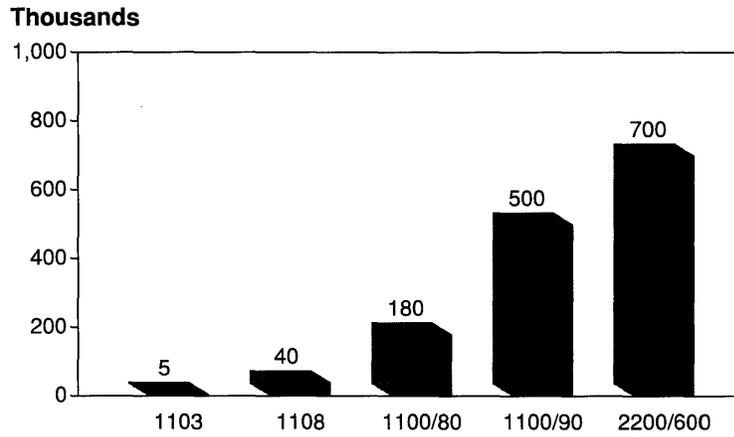
Memory capacity60 percent
 Logic capacity40 percent
 Logic speed.....20 percent

At the same time the number of logic gates used for the various functions has also been rising. There are a number of reasons for this including additional features, more checking hardware, and sections to help performance.

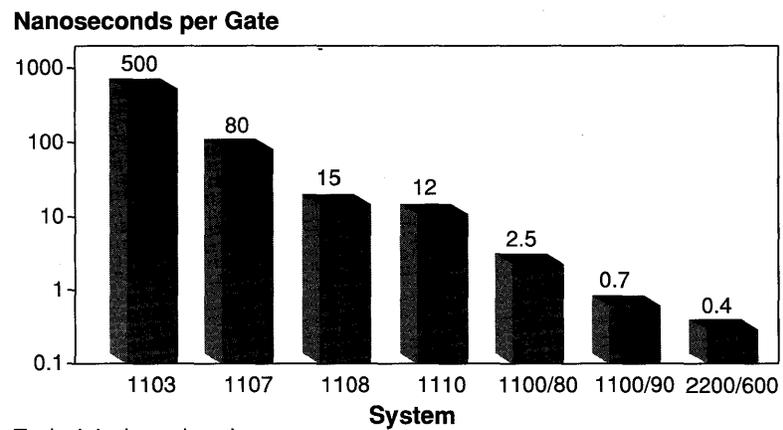
It is interesting to note that the current 2200 products offer a performance range of three to nearly 300 times the 1107, the first 1100 architecture machine. These are for tightly coupled multiprocessors, all running similar programs. Additional performance can be obtained with the use of the XTPA architecture and other software products. The accompanying bar charts present trends in a number of metrics for various 1100/2200 systems over the last 25 to 35 years.

The technology and development work in the laboratories is still progressing at a rapid rate. And we can certainly expect to see still more exciting improvements in the 2200 hardware designs in the future!

Processor Logic Gates

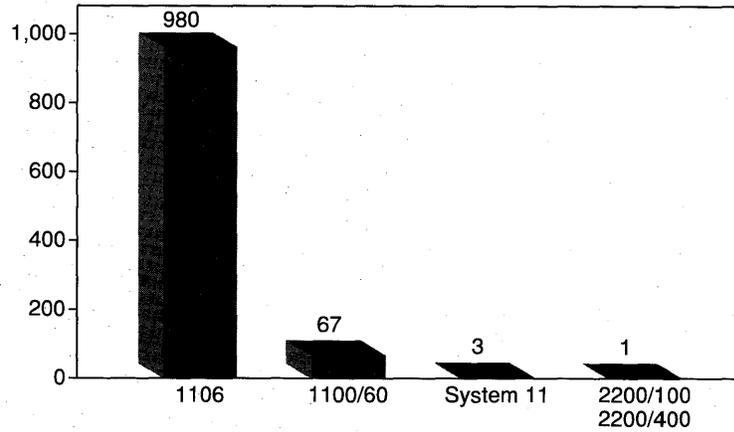


Logic Circuit Delay Trend

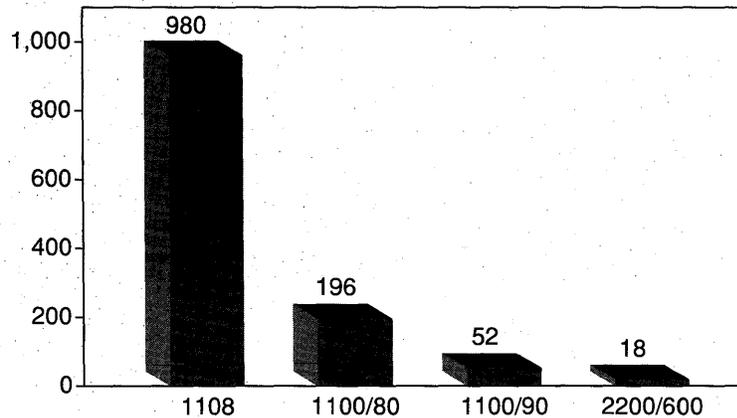


Typical design value shown
Measured unloaded value about half design value

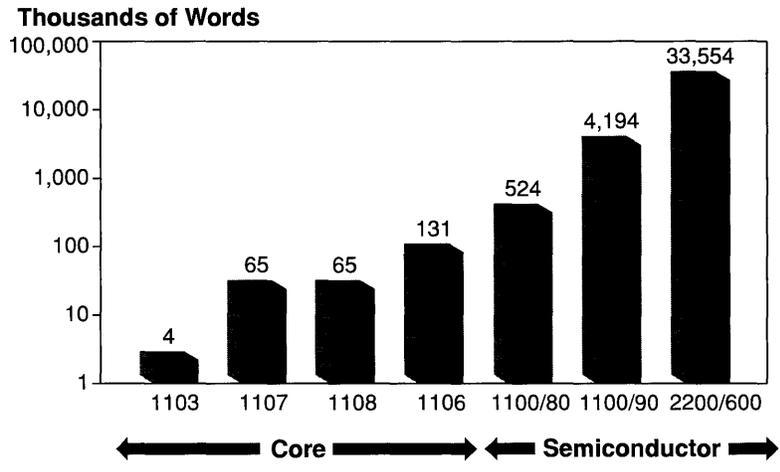
Processor Printed Circuit Cards Low – Medium Performance Systems



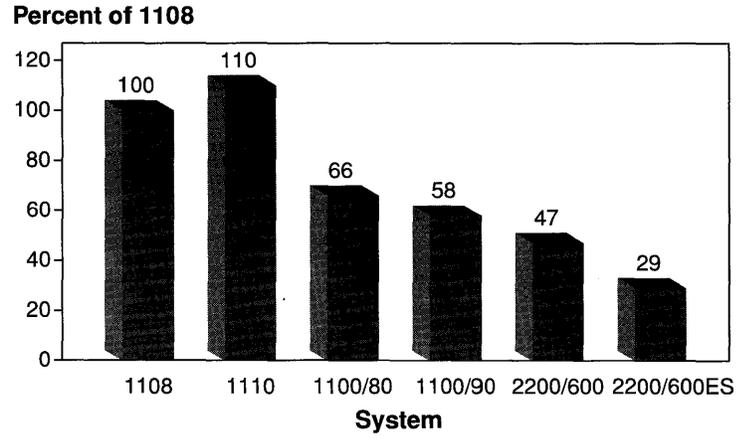
Processor Printed Circuit Cards High Performance Systems



Memory Capacity per Cabinet



Relative Floor Space Trend High Performance Systems



1100/2200 SYSTEM FIRSTS AND NOTABLE FEATURES

1101 (1950)

First "1100", Drum Memory

1103 (1953)

Commercial Sales, 36-Bit Word, Ferrite Core Memory

1107 (1962)

Transistor Logic, 1100 Architecture, Film Memory (GRS)

1108 (1965)

Multiprocessor Organization, Created Large User Base

1110 (1972)

Integrated Circuits, Separate I/O Processor, Plated Wire Memory

1180 (1976)

ECL Logic, Precision PC Boards, Cache and Semiconductor Main Memory

1100/60-70 (1979)

Single Cabinet System, Microgrammed Logic, Most 1100s Made

1100/90 (1983)

LSI Logic, Dense (50 mil) Packaging, Liquid Cooling

System 11 (1984)

100% LSI, System Bus, Office Environment

2200/200 (1986)

CMOS VLSI Processor, Single Board Processor

2200/600 (1988)

100% VLSI ECL Processor, Very High Availability/Performance

2200/400 (1989)

Six Processors/16 Megawords in One Cabinet, CMOS I/O, Very Low Power

2200/100 (1990)

Upright Cabinet, Very Small Footprint

2200/600ES (1990)

High Performance Combined with Low Power/Space, Evolutionary I/O.

ACKNOWLEDGEMENTS

Much of the information in this paper came from the author's own experience, and from various internal documents. He also appreciates the additional information and help attained from many Unisys employees, past and present, including the following: Bob Bergman, Tom Currie, Elwin Crandall, Curt Christensen, Les Cochran, Ken Englebrecht, Gerry Faggerness, Bill Heer, Duane Hjermstad, George Johnson, Gary Kemmetmueller, Glenn Kregness, Keith Miller, Walt Quinton, Gene Rodi, Jim Rogneby, Wes Swanson, Jim Scheuneman, Bill Swenson, Jorge Slater, Merlyn Schubloom, Gene Thomas, Ernie Unruh, Bob Wendt, Wayne Ward, and Dick Zoya.

And thanks to Helen Nelson who typed the initial draft material and Barb Harvey for the final composition.

BIBLIOGRAPHY

Borgerson, B. R.; Hanson, M. L.; and P. A. Hartley. "The Evolution of the Sperry Univac 1100 Series: A History, Analysis, and Projection"; *Communications of the ACM*, Vol. 21, No. 1, Association for Computing Machinery, January 1978.

Sperry, *Engineering Research Associates. The Wellspring of Minnesota's Computer Industry*, Sperry Corporation, 1986.

Tomash, E., and Arnold A. Cohen. "The Birth of an ERA: Engineering Research Associates, Inc. 1946-1955"; *Annals of the History of Computing*, Vol. 1, No. 2, American Federation of Information Processing Societies, Inc., October 1979.

Gray, George. "ENIAC, UNIVAC and Other Early ACs", 1990 Fall USE Conference.

Smith, R. Q. "Why It Is The Way It Is", 1990 Fall USE Conference.

Electronic Design, August 30, 1961 (Several articles on technology status at this time).

Pugh, Emerson. *Memories That Shaped An Industry*, The MIT Press.