Information Technology Fundamentals

Memory Technologies

School of Information Technology and Mathematical Sciences

Just a quick note: the success of a given technology is as much based on market forces as it is technical specifications.

Early Memory Technologies

Memory cards served as immutable nonvolatile storage in computers up until the 1970’s.

Short-term memory was, originally, handled by two main technologies.

Acoustic Delay Line

Relied on converting data into sound.

Sound is non-volatile, and slower than electricity.

Accordingly, while the data exists as sound,
it is stored in memory.

Acoustic Delay Memory needs four parts.

A radar **amplifier** to receive the data and amplify it for the ....

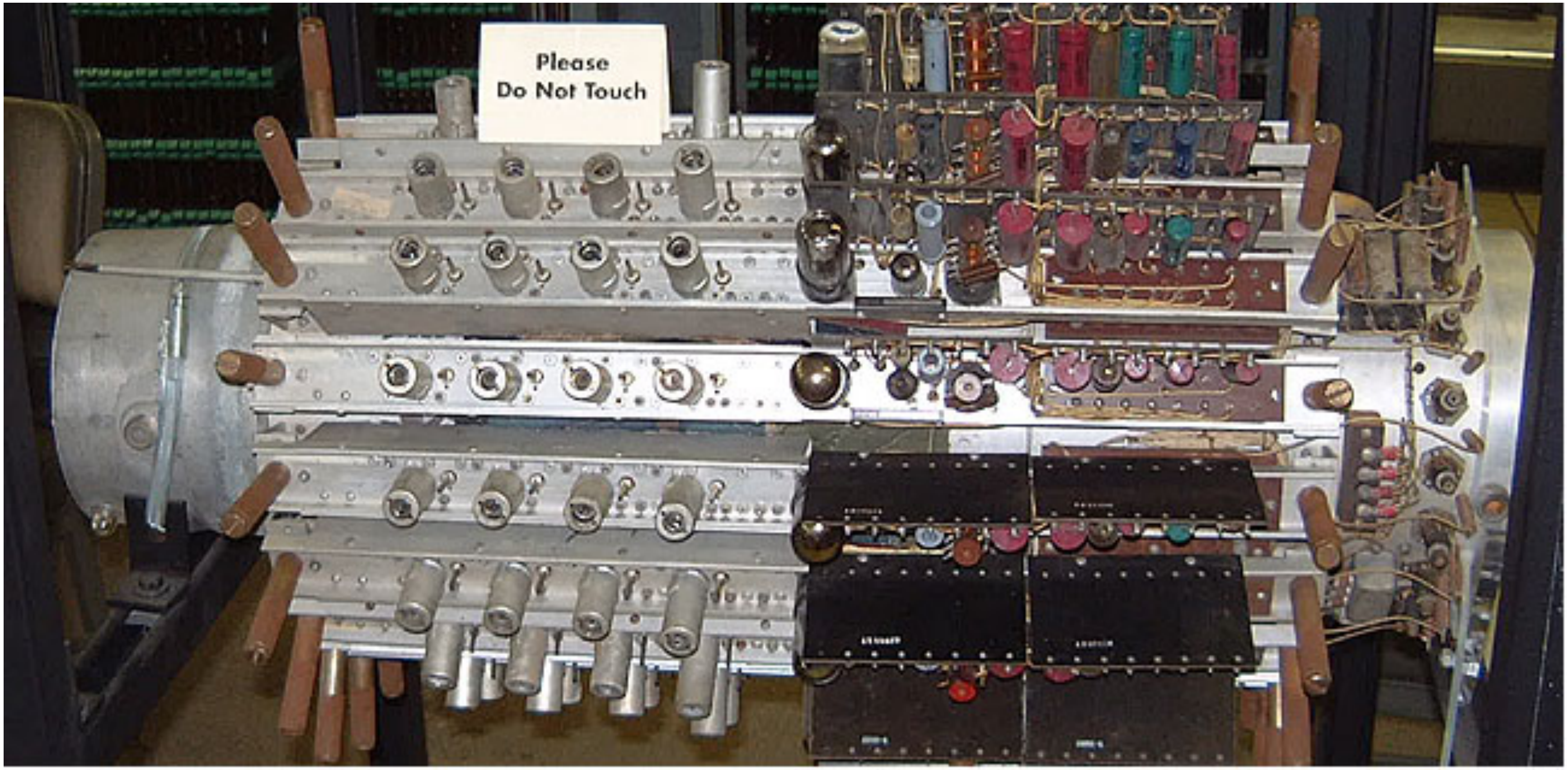
… **transducer**, (or speaker) which converted the data into sound and …

… transmitted the sound through a **medium** (typically mercury) until …

… the sound/data is picked up by another **transducer** (a microphone) and sent back to the amplifier.

While the data was sound in the medium,
it was stored in memory.

It had to be read back in the order it was sent, so it was not RAM.



Acoustic memory from the UNIVAC

Williams-Kilburn Tubes

Think of an old CRT television set.

An electron beam causes pixels on a phosphorescent screen to glow.

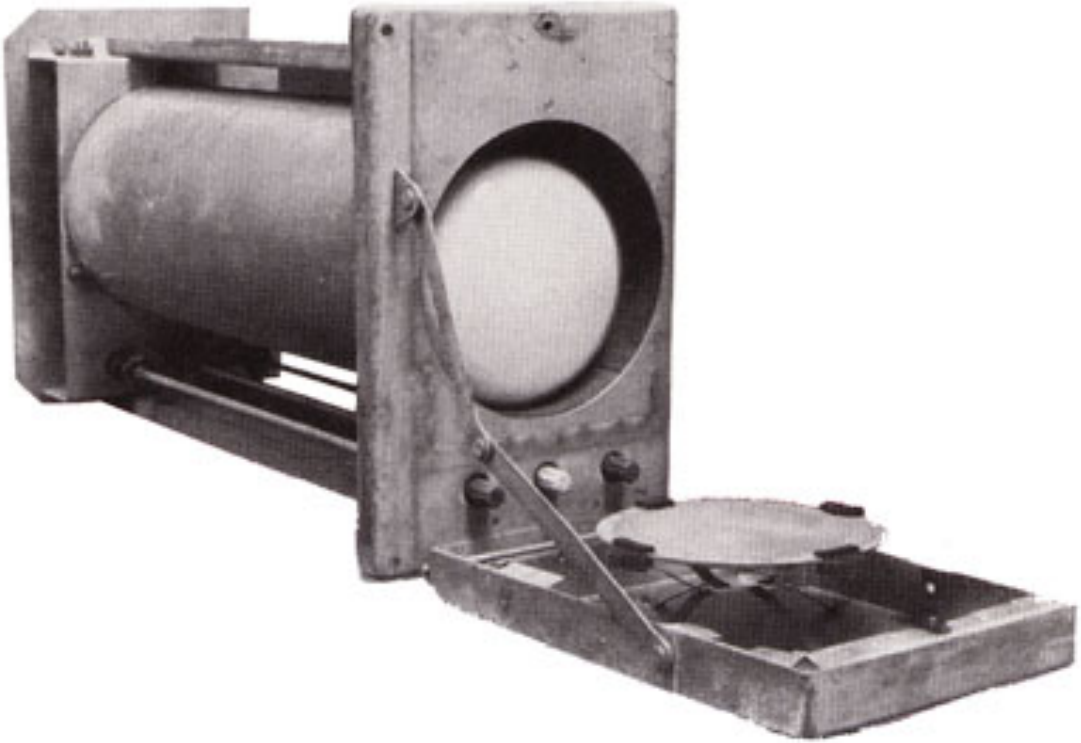
The phosphor continues to glow
after the beam moves on.

Williams-Kilburn Tubes used the same idea.

Each pixel is one bit of memory.

While it glowed it was storing data.
On if it glows (1) or off it doesn’t (0)

A thin metal plate detected which
pixels were glowing, so it could read them back.



Williams-Kilburn Tube

Williams-Kilburn tubes were fast but unreliable.

Therefore acoustic delay lines were more popular.

Drum Memory

Drum memory works like a hard drive – data is stored magnetically on the surface of an object.

But hard drives use disks, and drum memory uses … drums.

Drum memory.

Technology	*Example*	*Capacity*	*Access Time*
Acoustic Delay Line	EDVAC (1951)	1024 words x 44 bits	48 μs (min) 240 μs (ave)
Williams-Kilburn Tube	Manchester Mark I (1949)	128 words x 40 bits (two tubes)	400 μs
Drum	Manchester Mark I (1949)	4096 words x 40 bits (two drums)	15 ms (ave)

μs = microsecond = 10^{-6} second

ms = millisecond = 10^{-3} second

Comparison of early memory technologies.

The next phase

Magnetic Core Memory

Emerged in the 1950’s at Harvard, remained dominant until the 1970’s.

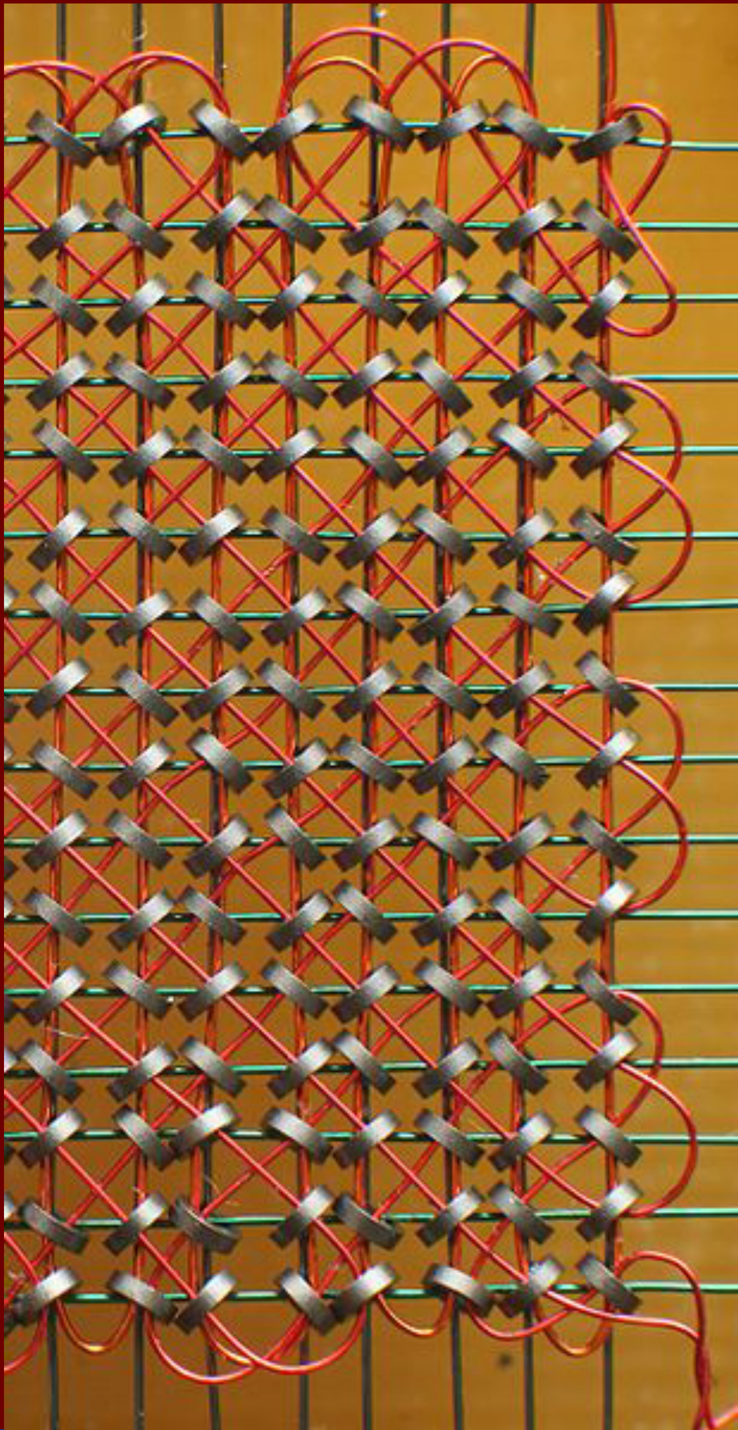
Also known as “Core memory”,
hence “core dump” for Unix developers.

Consists of a large number of small, donut-shaped magnets.

Each magnet could store a clockwise or anti-clockwise charge. (One bit, interpreted as 0 or 1).

By changing the direction, the stored data is changed.

- Each magnet has three (or four) wires running through it:
  - X and Y wires to change its state
  - A sense/inhibit line.



For writing, the X and Y wire contains half the charge needed to flip the direction. Where they cross is where the data is stored.

For reading, the computer attempts to “flip” (change the state) of the magnet.

The sense line determines if a change occurred.

If it changes from 0 to 1, it used to be 0.

If it doesn't change, it was already 1.

Core memory was non-volatile and mutable.

It was also sensitive, especially to magnetic fields – including itself.

Thin Film Memory

Thin Film Memory used the same idea, but instead of magnetic donuts used dots of magnetic material on glass.

Rather than passing wires through the magnets, they sat on top.

They proved to be much faster, but very expensive to make.

Twistor Memory

Like magnetic core memory, but magnetic tape was wrapped around the wire, rather than the wire threaded through magnets.

Had a short lifespan, because it was produced just before semiconductor memory came onto the scene.

Bubble Memory

Works by magnetizing points in a crystal.

Each “bubble” could be created along one edge of the crystal.

A charge along that side would then push the bubbles across the crystal, making room for new bubbles.

When the bubbles reach the other side of the crystal they could be read.

Provided non-volatile, mutable memory, with HD-like density and core memory like performance.

Killed by Flash RAM, semi-conductor memory, and HD capacity.

Integrated Circuits

Emerged in the 1970’s, killed all other technologies.

Volatile and either mutable or immutable, the main advantages were speed, size and cost.

Use integrated circuits, which can be categorized based on complexity:

SSI: Small Scale Integration.

Only a few components.

Used in the Apollo Space Mission’s guidance computer.

MSI: Medium Scale Integration.

Hundreds of components on the one chip.

Could build much more complex systems for little increase in cost.

LSI: Large Scale Integration.

Tens of thousands of components.

Same economic benefits as MSI, only greater.

VLSI: Very Large Scale Integration.

The standardisation of the design of CMOS chips made a significant contribution to VLSI emerging.

CMOS: Complimentary metal-oxide-semiconductor.

Hundreds of thousands of components.

Enabled the first single chip microprocessors.

ULSI: Ultra Large Scale Integration.

Qualitatively no different from VLSI.

Usually when people say VLSI, they are also talking about ULSI.

Millions of components on one chip.

WSI: Wafer Scale Integration

1980s

Entire computers on one silicon wafer

Microprocessor and memory

Difficult to manufacture without defects, so failed in the market

SOC: System-on-chip

All devices on one chip

Clock, microprocessor, memory, I/O, communications, etc

Used widely in embedded systems

Each system's chip has to be specially designed

IC memory comes in small packages.

Dual Inline Package (DIP).

Two parallel rows of connecting pins for conducting electricity.

Single Inline Memory Module (SIMM).

Much like DIP, but with contact plates for conducting electricity on one side.

Dual Inline Memory Module (DIMM).

If a SIMM handles 32 bits, a DIMM handles 64 bits.

Synchronous Dynamic RAM (SDRAM) vs DRAM

DRAM is asynchronous, which makes it quicker.

SDRAM is synchronous: it waits for the clock signal.

This means that it can be synchronised with the bus and the microprocessor, which enables pipelining.

Double Data Rate SDRAM (DDR SDRAM)

As the name suggests: wider bandwidth, 2 words per clock cycle

DDR2 SDRAM

Latency is the same as DDR SDRAM, but twice the bandwidth, 4 words per cycle

DDR3 SDRAM

8 words per cycle

Lower power consumption than DDR or DDR2

DDR4 SDRAM

Lower power consumption, higher density, higher data transfer.

Flash Memory

IC which can maintain its charge for years without using any power.

Allows random access for reading but not for writing.

Emerging Technology

Programmable ICs.

Can be programmed by the user instead of functionality being fixed at manufacture.

Used in *reconfigurable computing*.

IC Nanotechnology (NEMs)

In 2001 IBM announced that it had developed the first single-molecule integrated circuit

Predicted to overtake silicon at some point in the future.

Manufacturing debate: do we go with top down or bottom up?

Memristors

Proposed theoretically in 1971.

Implemented in April 2008.

Operates at about 1Hz (slow) on 1V (not much power).

Solid state

High density

100gb of storage in a square centimetre

Flash memories of the same size have lower capacity

Could eventually replace both RAM and hard-drives.