Week 3: System design

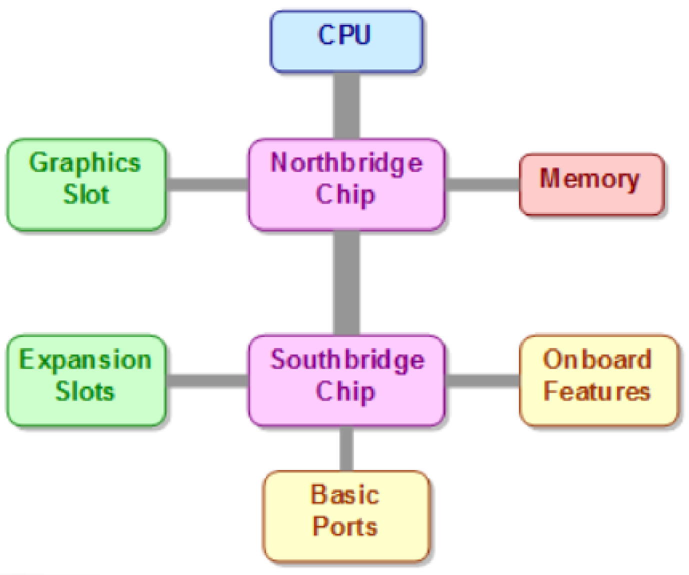
Legacy

Legacy systems (as depicted on the right side of the diagram above) often broke this task up into two discrete chips, known as a ***northbridge*** **(http://en.wikipedia.org/wiki/Northbridge_(computing))** and ***southbridge*** **(http://en.wikipedia.org/wiki/Southbridge_(computing))**. The northbridge connected directly to the CPU, and provides it with high-speed access to the computer memory and graphics card, two devices that the CPU needs the fastest communication with.

> The reason why the chipset was split in this way is to improve performance: by placing the most speed-critical components "closest" to the CPU, they can be accessed more quickly (a metric known as *latency* -- how long it takes to access a certain peripheral device).

The southbridge connected to the northbridge, and provides all other functionality that is not so time critical (relatively speaking; we're talking in terms of nanosecond delays here). The southbridge chip provided connections to other peripherals both inside and outside the computer case, including:

- **disk drives:** these are used for permanent storage and retrieval of data.
- **expansion slots:** these are connector slots in the computer that allow direct connection to the system's internal busses. They are used to add internal
- **basic ports:** this includes necessities like keyboard and mouse ports.
- **onboard features:** some computers have features such as audio-generating devices and network adapters integrated right onto the southbridge. This simplifies the overall design of the computer system (less chips and/or expansion boards are required), and helps reduce costs.



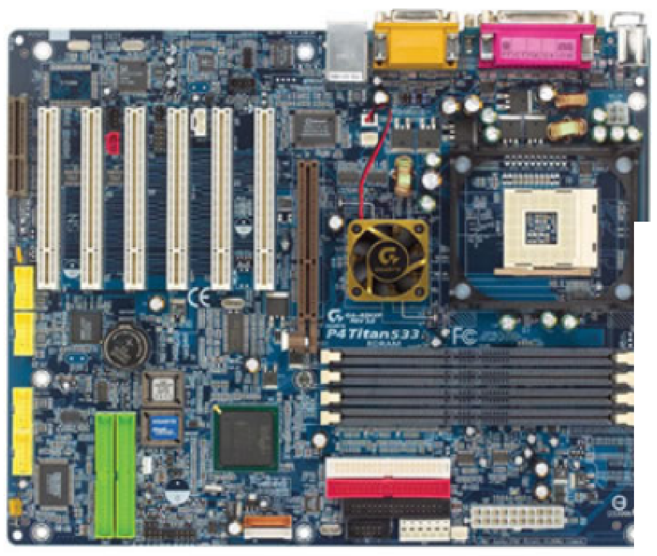
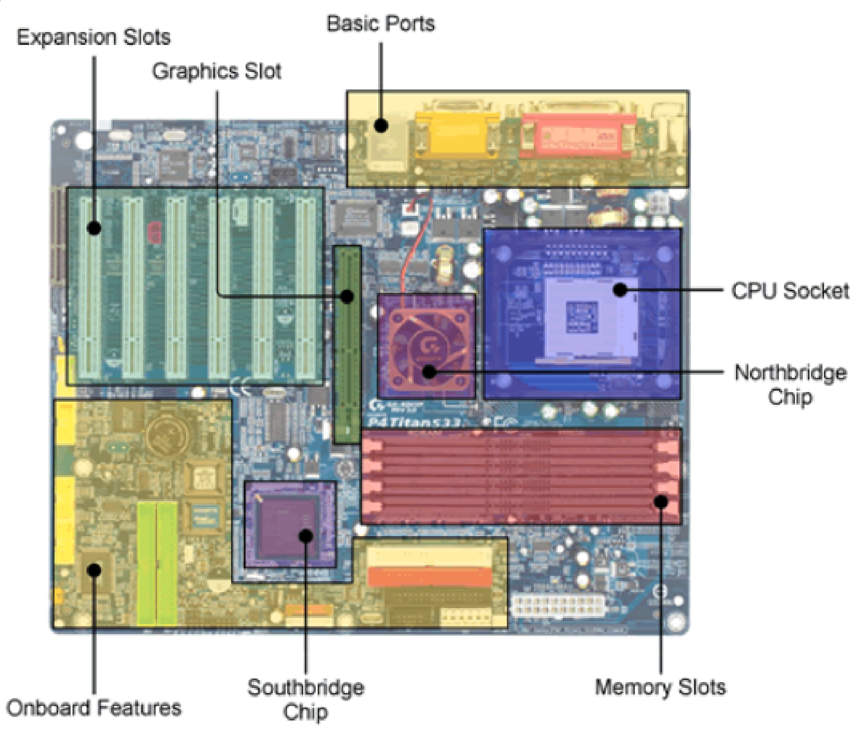


Photo: Gigabyte Co



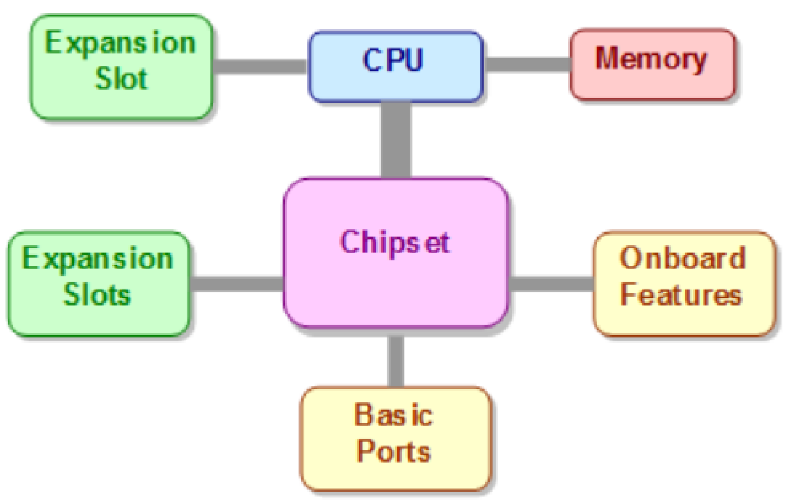
Modern

In electronic systems, it makes good sense to integrate devices (i.e. put all the functions onto a single chip) in order to boost performance and decrease latency. When the northbridge-southbridge model was used, it was not realistically possible to host so many features on a single silicon die without serious manufacturing and heat dissipation issues.

As IC manufacturing processes improved, it gradually became possible to integrate more features onto fewer chips.

- The first such integration, introduced with AMD's Athlon 64, was to put the **memory controller directly on the CPU die**, so that CPU-memory data transfers were as fast as possible.
- Another integration was to **combine the northbridge and southbridge** functions onto a single chip chipset. (Despite being only a single chip, they are still referred to as "chipsets".)
- As transistor density afforded even further integration, **high speed expansion interfaces were placed directly onto the CPU** - typically used for the graphics system, the most performance-dependent type of system expansion.

Some systems, notably those for embedded systems and small mobile computers such as tablets and smartphones, have the entire computer system - CPU, memory controller, graphics controller, I/O, and even system memory - on a single physical chip. These are called SoC (**System on a Chip** **(http://en.wikipedia.org/wiki/System_on_a_chip)**) integrated circuits.



Actual system designs

The system diagrams below are of actual computer systems. Note the high level of integration on the newer Intel Intel Z87 chipset (pictured on the left), in comparison to the older PowerPC G5-based system diagram on the right.

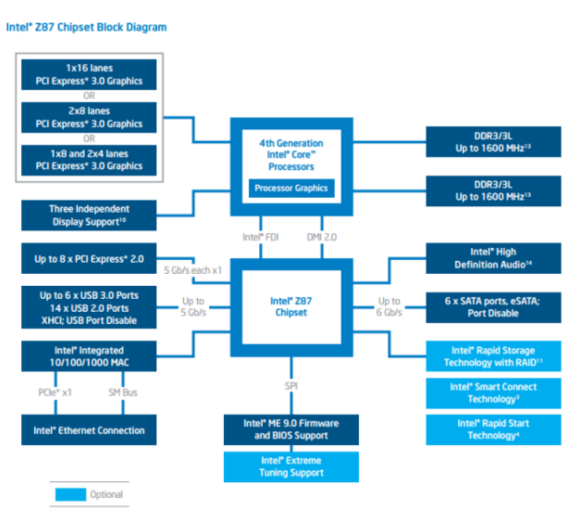
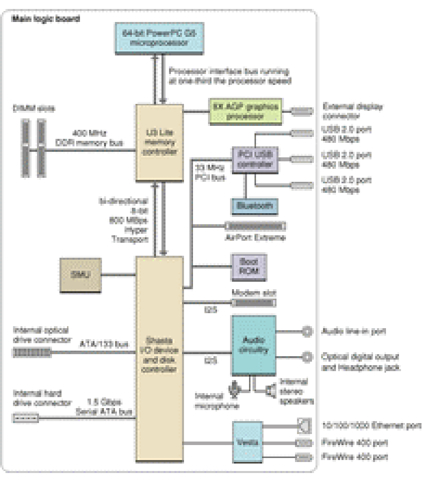


diagram: Intel Corporation



- Intel Z87 chipset – high level of integration

Data links

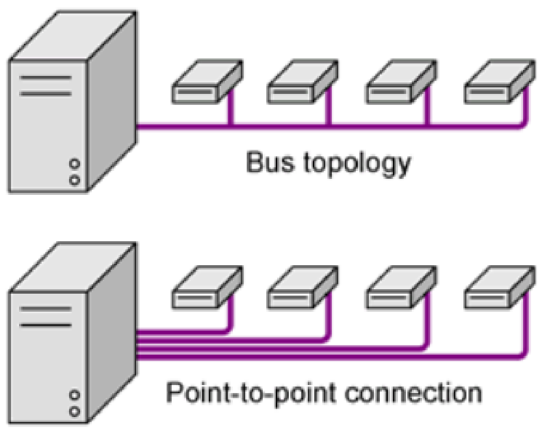
Between every internal component or subsystem in a computer (processor, memory, expansion cards, etc.) lies a data link. There are a variety of link types, each designed for a particular purpose in mind (or, in the case of expansion links, designed on purpose to support a wide variety of potential use cases).

A computer's functionality can also be expanded using external devices known as *peripherals*. These include common computing devices such as keyboard, mice, printers and modems. Perhiperals are connected through an interface port on the computer: there are a number of interface types, each best suited to a particular task: low or high speed, *hot pluggable* (can be connected/disconnected while the computer is on), cheap or expensive to implement, etc.

Although these connection schemes are all different, there are a few common attributes that are shared by all.

Topology - bus vs. point-to-point

A *topology* refers to how the signal ends are connected to one another. Most all modern signal interfaces use either a bus (shared data path), or point-to-point (dedicated data path) topology. They can be summarised in the following diagram:



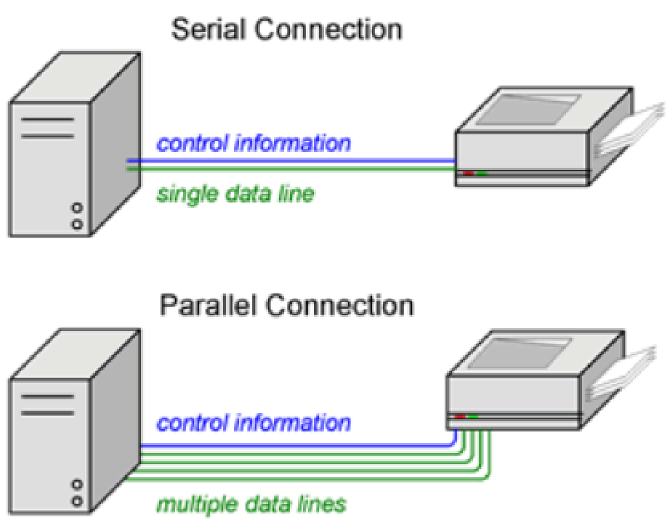
As we mentioned earlier, a "bus" is an electrical path onto which devices can be connected. Unlike a "point-to- point" system, where each device is guaranteed its own set of wires, a "bus" shares the wires across all devices connected to it.

This brings about the issue of the bus being congested - where so many devices try to communicate on the bus, it simply can't handle the volume of information. The bus interfaces are typically designed such that they are sufficiently fast that the data demands do not normally overwhelm the bus in this way. Issues arise, however, in time when increasing computing and data demands mean that buses are pushed further than originally intended or designed for.

Data transfer - parallel vs. serial

A "serial" interface means that data is transferred along a wire one bit at a time. Along with this "data" wire, there are a few other connections, such as an electrical ground and some signal wires to control the flow of information.

A "parallel" interface means that instead of data being transmitted one bit at a time, there are a number of signals travelling along many wires at the same time.



Due to the decreased number of wires and the unique way in which the electrical signals are transmitted, serial cables can span a far longer distance before the electrical signals degrade.

In the past, when computer clock (and respectively, the interface) speeds were relatively slow, parallel interfaces were far faster than their serial counterparts.

Virtually all new-generation interface connections are serial in nature: with modern technology, the interface can be run incredibly quickly, while still allowing the long cable length typical of a serial data transfer scheme.

Interface connection - hot vs. cold swap

A *hot- swappable* interface connector means that the device can be safely plugged and unplugged while the computer is switched on. This is usually achieved through two means:

- the physical connector is designed so that the device is properly grounded before a connection is made to any electrical circuitry
- the way in which the computer interacts with the device is such that it can be removed at any time without causing harm or loss of data

A *cold- swappable* device should only be plugged and unplugged after switching off the computer. If this is done while the computer is on, there is a risk of the electrical connections shorting out, causing potential damage to the device and/or the computer port.

> Most external cold-swappable connectors and computer interfaces are usually equipped with some form of short-circuit protection, which vastly reduces the risk of disaster if a cold- swappable device is hot-swapped.
>
> This protection shouldn't be relied upon, though.

There is also the issue in that a computer might rely on a cold-swappable device being present at all times, if it was available at startup. If it is disconnected while the computer is on, it may try to communicate with it later on (on the assumption it's still there) causing a system crash or other undesirable behaviour.

> Almost all internal connections in a computer are cold-swappable, and offer little or no short circuit protection.
>
> **Never** work on the insides of a computer that's switched on.

There is also the concept of a *warm - pluggable* device, usually in portable computers. This means that a component is unsafe to plug in or remove when switched on, but okay when the computer is in sleep mode. The most common such device are laptop batteries: a good proportion of laptops can remain in sleep mode without a battery installed, just long enough to remove and replace the battery. This allows the user to replace the battery without having to shut down and restart the entire system.

Legacy interfaces

A "legacy" interface is one that is a hangover from a previous generation of computing, which has since been replaced by a newer equivalent. In some cases, however, the legacy interfaces persist in the marketplace for a number of reasons:

- they are very cheap and easy to implement;
- there are so many devices that still use them;
- they're "good enough" for the task at hand.

Many legacy interfaces are electrically (and mechanically) simple, and this has promoted their longevity: they are not only cheap and easy to implement, but also very easy to build peripherals for.

The three main legacy interfaces that still exist today are those for keyboards and mice, parallel and serial communications.

Keyboard and mouse connections

Keyboards and mice are very undemanding from an interface point of view: they don't require much data to be transmitted (low bandwidth), and they don't require a connection with low latency. There are two legacy connections for keyboards; the AT and PS/2 standards.

- the **AT standard** (http://en.wikipedia.org/wiki/AT_keyboard) , in common use from the IBM PC AT in 1984, up until the early 1990s, uses a large 5-pin circular connector known as "DIN-5".
- this was replaced by the **PS/2 standard** (http://en.wikipedia.org/wiki/PS/2_keyboard#Keyboard.2Fmouse_interface) in the early 1990s (and earlier still, in IBM's **Personal System/2** (http://en.wikipedia.org/wiki/PS/2) range of desktop computers). This used a smaller "mini-DIN" connector, and included mouse support.
- Despite their identical appearance, PS/2 mouse and keyboard ports are not the same.
- Apple Macintosh systems from the late 1980s to mid 1990s used the **Apple Desktop Bus** (http://en.wikipedia.org/wiki/Apple_desktop_bus) (ADB); this used a mini-DIN connector similar to the PS/2 standard, but is very different electrically. Its major advantage was that it allowed keyboards, mice, and other input devices to be connected to each other, in the end requiring only one connection to the computer.

> There was an even older keyboard connector, the "XT standard", used on the first **IBM PC** (http://en.wikipedia.org/wiki/IBM_PC) computers. It uses the same connector as the AT keyboards, but is electrically incompatible. "XT standard" keyboards have long since disappeared.
>
> During the transition from XT to AT systems in the mid-80s, keyboards often had a switch that let them work with both types of computer system.

All these legacy interfaces are cold swappable, and particularly susceptible to electrical damage from hot-swapping. Before the advent of ADB (on Macintosh systems) and the PS/2 mouse port (on IBM PC systems), mice were connected using a standard serial interface, described in detail below.

These interfaces have since been replaced by the USB (Universal Serial Bus) standard.