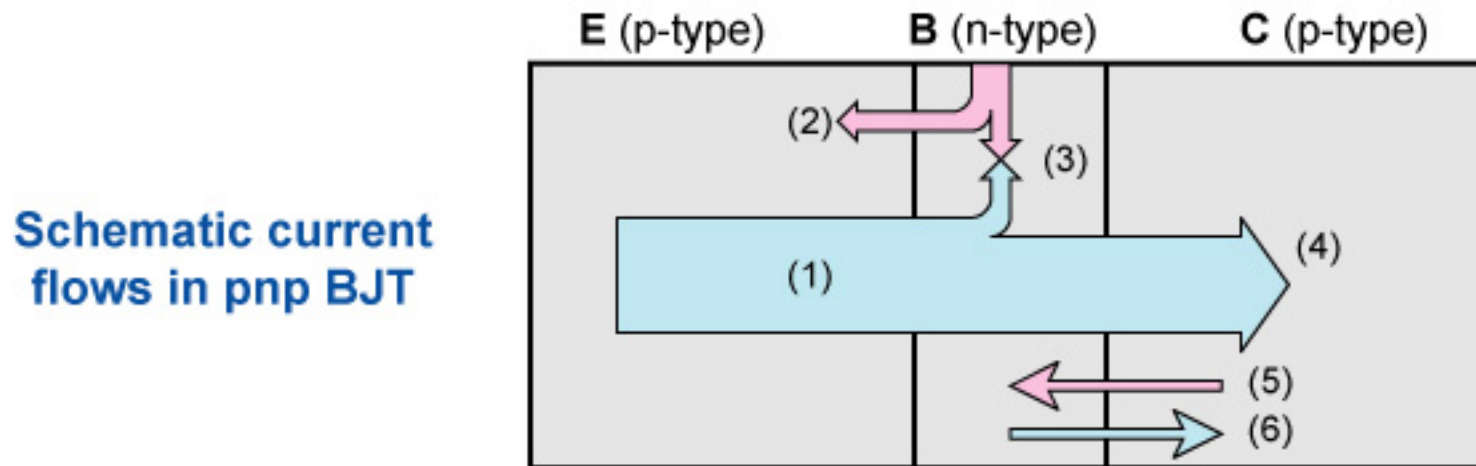


Bipolar junction transistor - History

Basic operation of BJT



- (1) Hole particle current from E to B
 - (2) Electron particle current from B to E
 - (3) Recombination current in B
 - (4) Hole particle current originating in E and reaching C
 - (5) Reverse electron particle current from C to B
 - (6) Reverse hole particle current from B to C
- (What is difference between “current” and “particle current”?)

Qualitative basic operation of BJTs

What is a BJT?

A BJT consists of two back-to-back p-n junctions.

The middle region, the base, is very thin.

The three regions are the emitter, base, and collector.

Carriers are injected (“emitted”) into the base from the emitter.

Since the base is thin, most carriers injected into base diffuse into collector.

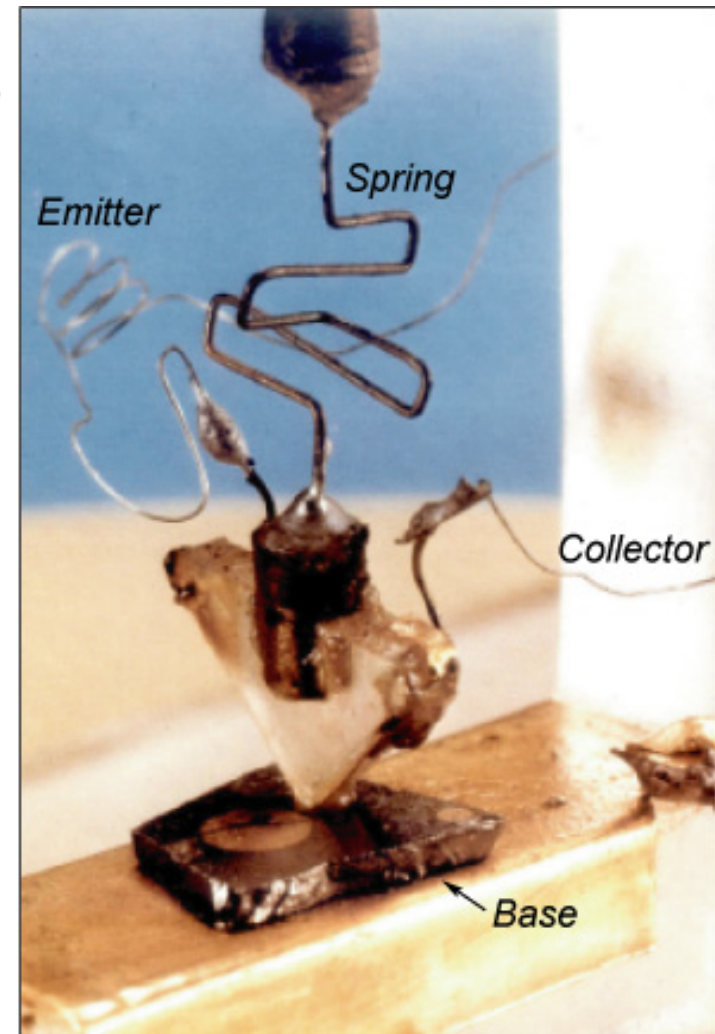
What does a “thin base thickness” mean? → Base thickness is much thinner than the diffusion length of carriers injected from the emitter.

History of point-contact transistor

The first transistor was a point-contact transistor

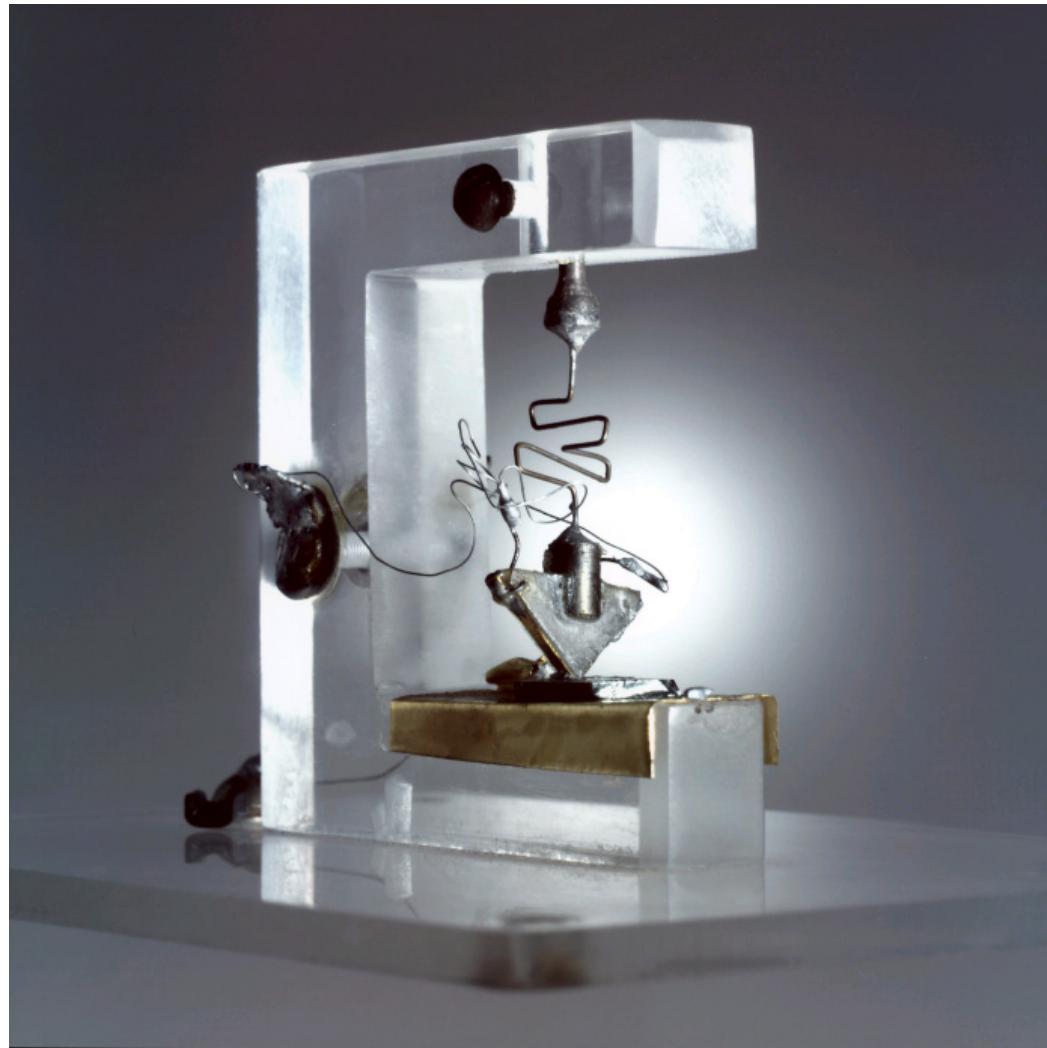
The first point-contact transistor

*John Bardeen, Walter Brattain, and William Shockley
Bell Laboratories, Murray Hill, New Jersey (1947)*



Model of transistor in Bell Labs museum

Museum is located at 600 Mountain Ave., Murray Hill, New Jersey



History of the first transistor

“The first transistor was about half an inch high. Before Brattain started, John Bardeen told him that they would need two metal contacts within **0.002 inches** (0.002 inches = 50 μm) of each other - about the thickness of a sheet of paper. But the finest wires then were almost three times that width and couldn't provide the kind of precision they needed. Instead of bothering with tiny wires, Brattain attached a single strip of gold foil over the point of a plastic triangle. With a razor blade, he sliced through the gold right at the tip of the triangle. Voila: two gold contacts just a hair-width apart.

The whole triangle was then held over a crystal of **germanium** on a spring, so that the contacts lightly touched the surface. The germanium itself sat on a metal plate attached to a voltage source. This contraption was the very first semiconductor amplifier, because when a bit of current came through one of the gold contacts, another even stronger current came out the other contact.

Here's why it worked: Germanium is a semiconductor and, if properly treated, can either let lots of current through or let none through. **This germanium had an excess of electrons**, but when an electric signal traveled in through the gold foil, it injected holes (the opposite of electrons) into the surface. This created a thin layer along the top of the germanium with too few electrons.

Semiconductors with too many electrons are known as N-type and semiconductors with too few electrons are known as P-type. The boundary between these two kinds of semiconductors is known as a P-N junction, and it's a crucial part of a transistor. In the presence of this junction, current can start to flow from one side to the other. In the case of Brattain's transistor, current flowed towards the second gold contact.

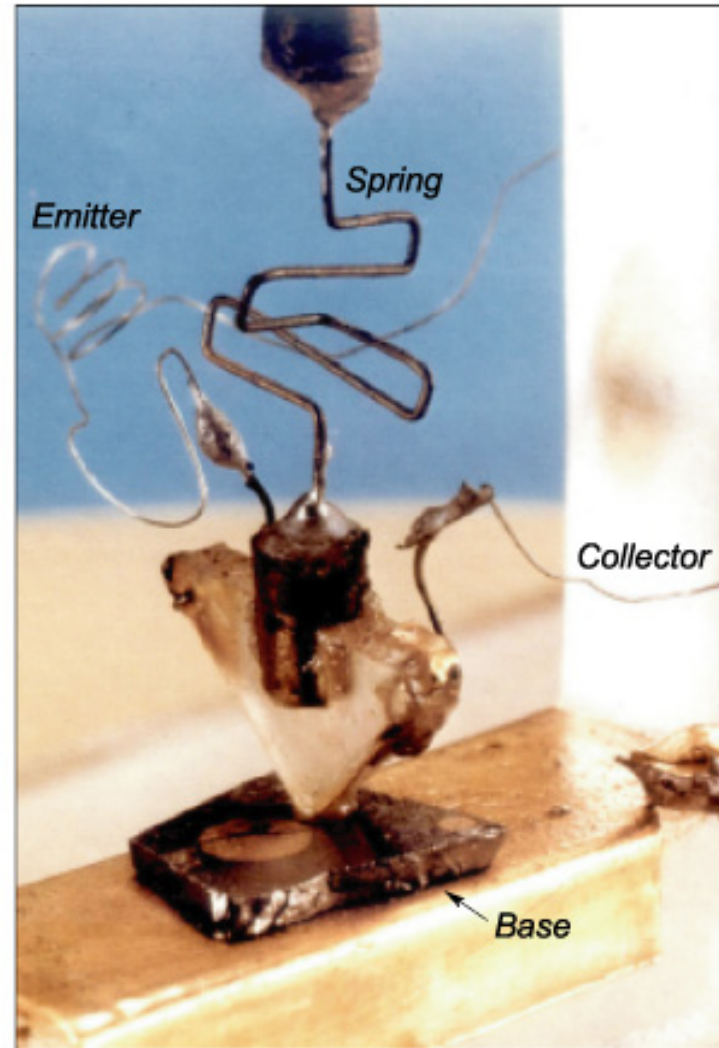
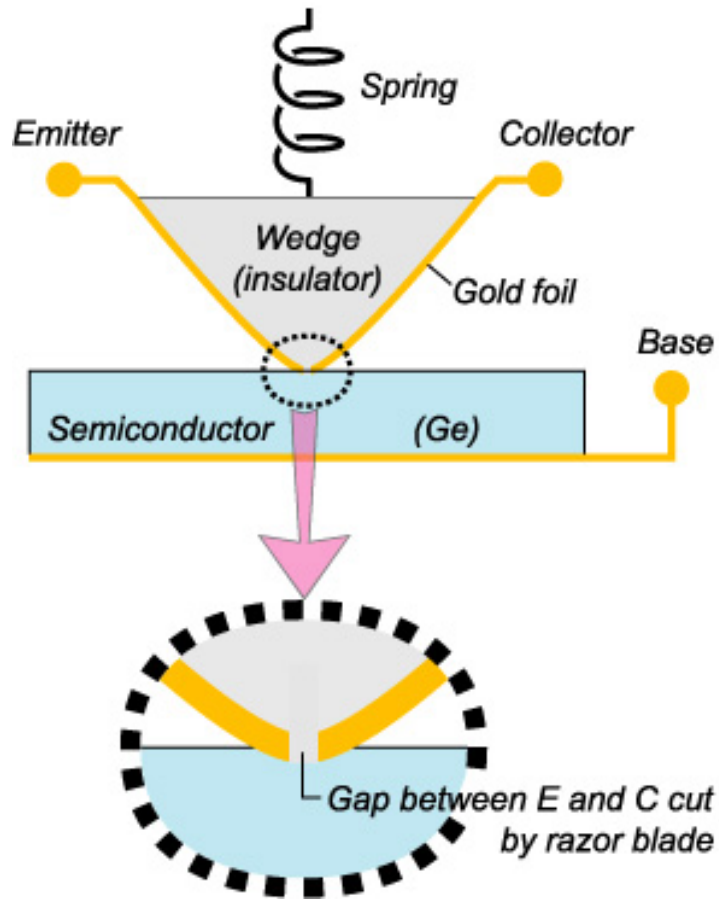
Think about what that means. A small current in through one contact changes the nature of the semiconductor so that a larger, separate current starts flowing across the germanium and out the second contact. **A little current can alter the flow of a much bigger one**, effectively amplifying it. Of course, a transistor in a telephone or in a radio has to handle complex signals. The output contact can't just amplify a steady

hum of current, it has to dutifully replicate a person's voice, or an entire symphony. Luckily, a semiconductor is perfectly suited to this job. It is exquisitely sensitive to how many extra or missing electrons are inside. Each time the input signal shoves more holes into the germanium, it changes the way current flows across the crystal - the output current instantly gets larger and smaller, perfectly mimicking the input."

(after Public Broadcasting System (PBS) television series, 1999)

How did first point-contact transistor work?

Schematic of the first point-contact transistor



A gold foil was glued to a triangular insulating wedge.

A narrow gap was cut with a razor blade to form the E and C.

The gap was approximately 50 μm wide.

Under forward bias of the EB junction, minority carriers are injected into base (In case of point-contact transistor, strong forward bias is required).

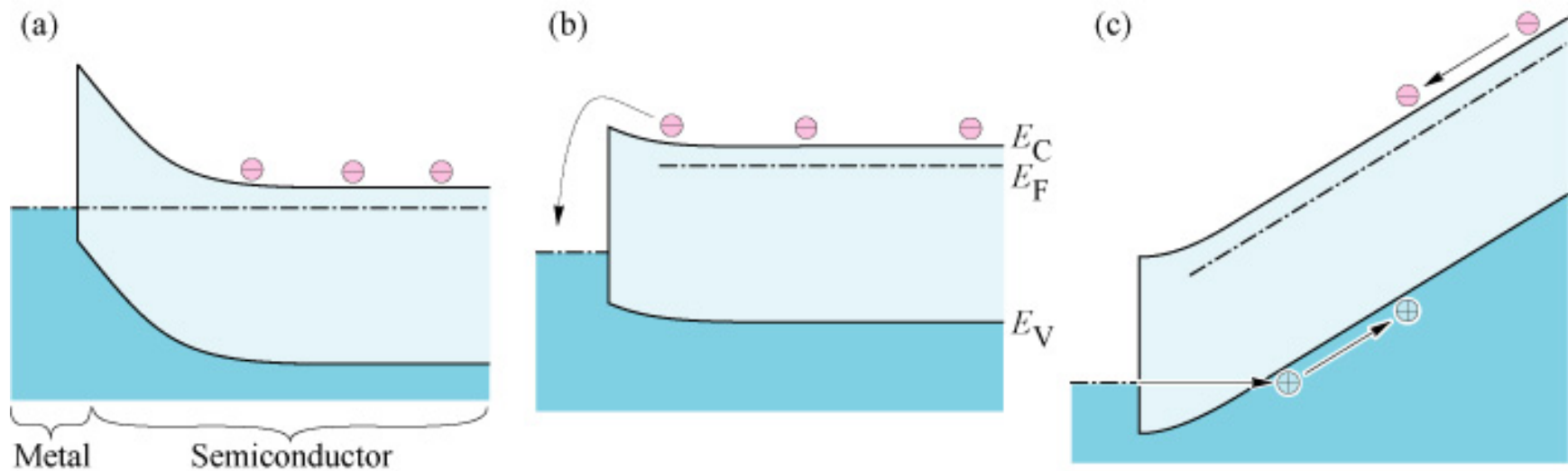
How are minority carriers injected without a p-n junction?

Most minority carriers are collected by the reverse-biased BC junction.

Thus the base controls the current flow between E and C.

How are minority carriers injected without a p-n junction?

Answer:



Band diagram of Schottky diode under (a) equilibrium conditions, (b) forward bias, and (c) strong forward bias. Under strong forward bias, minority carrier injection occurs.

... strong forward-bias is needed for minority carrier injection.

Can Schottky contacts emit light? If yes, under which circumstances?

Are there other circumstances under which Schottky contacts emit light?

Problems of the point-contact transistor:

The point-contact transistor is a **surface-effect device**, *i. e.* important effects occur at the semiconductor surface.

Surfaces are easily contaminated.

Surface-effect devices are unstable (*e. g.* mechanical vibrations).

Remedy:

William Shockley invented the pn-junction transistor.

The physically relevant region is **moved to the bulk** of the material.

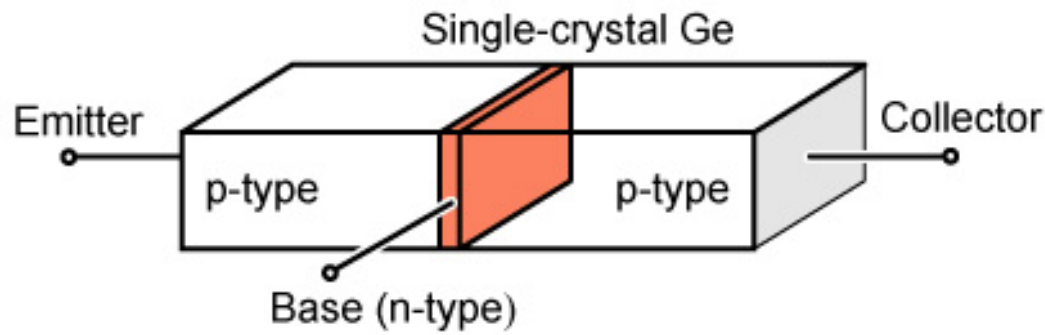
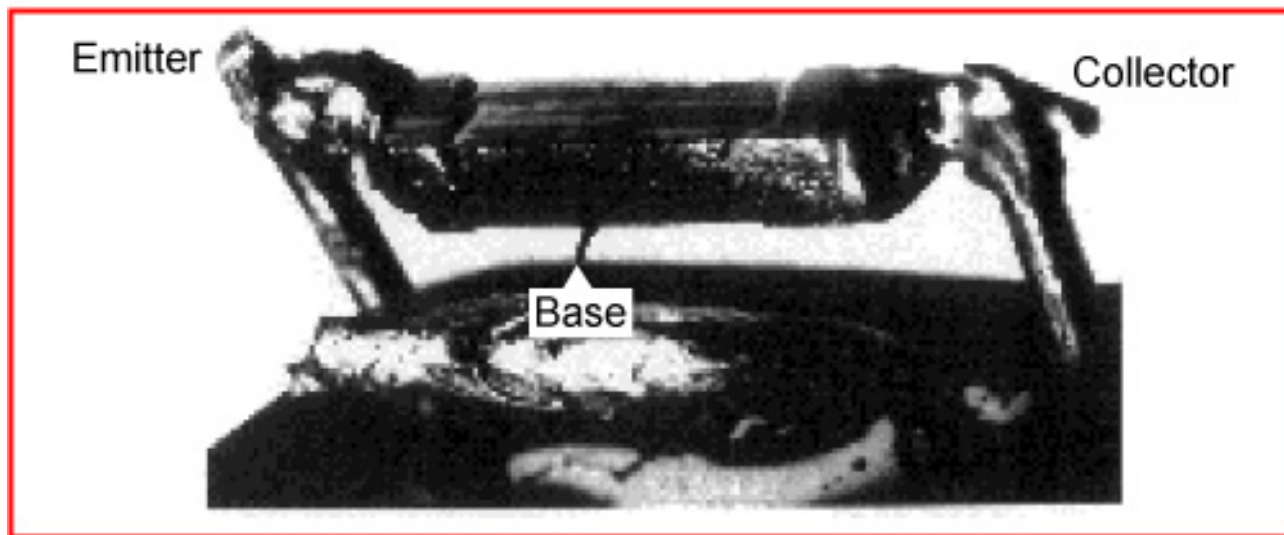
The p-n junction transistor is a bulk device.

Shockley, Bardeen, and Brattain were awarded the Nobel Prize in 1956.

History of BJTs

The First Junction Transistor

First transistor with diffused pn junctions by William Shockley
Bell Laboratories, Murray Hill, New Jersey (1949)



Questions

Why is the E called E?

Why is the C called C?

Why is the B called B?

What is meant by “thin” in context of “thin base thickness”?

The EB junction is forward biased. Many more carriers flow from E → B than from B → E. Why is it supposed to be that way? How is this achieved?

What are differences between point-contact transistor and junction transistor?

Which of these two transistors is no longer used and why?

References

Historical materials are courtesy of AT&T Bell Laboratories

B. G. Streetman and S. Banerjee "Solid State Electronic Devices"
(Prentice Hall, Englewood Cliffs NJ, 1999)

PBS television series on transistor (1999)