Current — which way does it flow?

From the Museum of New York, this Currier and Ives print depicts the 1752 experiment by Benjamin Franklin which established a relationship between lighting and static electricity — altogether a rather hazardous experiment.

Which way does electric current flow in a circuit? From positive to negative, as originally proposed by Benjamin Franklin, or in the reverse direction, as maintained by many adherents of electronics? This somewhat historical survey may help you to get the matter in perspective.

by NEVILLE WILLIAMS

Electrical phenomena have been observed and documented over hundreds — even thousands — of years, although not understood at the time. The ancient Greeks, for example, were aware of static electricity, as manifest by the way in which a piece of vigorously rubbed amber would attract particles of dust and lint. In fact, the Greek word "elektron", meaning amber, is the root of our modern term "electronics".

Again, the Romans were aware of the peculiar behaviour of "loadstone", or magnetite, an oxide or iron which provided the key to the early mariner's compass. Even in those far off days, Roman writers were unsure whether their word for the effect had been derived from the island of "Magnesia", the source of the loadstone or "Magnus", the shepherd who allegedly discovered its strange properties. Sufficient to say that, between them, the words provided the root for our modern term "Magnetism".

Much later, from about the 18th century onward, European researchers began to study electrostatic and magnetic effects more deliberately and the conviction grew that there was a link, as yet unresolved, between the two. In 1752, Benjamin Franklin extended the speculation by demonstrating a further link between static electricity and atmospheric lighting — an experiment that, in retrospect, has been classified as one of the most foolhardy of all time!

Indeed, Franklin postulated that electricity was some kind of intangible fluid. When an object had more than its share of this fluid, it would attract (or electricity would flow into) another body which had less than its share. He also suggested that a body with surplus fluid should be regarded as being electrically positive, and bodies with less than normal fluid as being electrically negative.

A concentration of the "fluid" was seen as "static" electricity. Electrical fluid in motion came to be known as "current" electricity and, in line with Franklin's proposition, current was deemed to flow from positive to negative.

At that point in time, there was no special reason to adopt a contrary view and it was accepted as fact.

So the convention was established.

In subsequent years, Luigi Galvani (1780) and Alessandro Volta (1880) contributed to research which produced the first electrical cells (or "batteries" of cells). In due course, one connection was identified as the source of electrical "fluid" and variously branded "anode", "positive", "plus" or "+". The other connection was called "cathode", "negative", "minus" or "-".

Thus equipped, Hans Christian Oerstadt began a long series of experiments in 1807, which established a firm relationship between electrical current flow and magnetism:

When current from a battery flowed through a wire (from anode or positive to cathode or negative, of course) it produced a magnetic field around the wire. This could be sensed by a compass and shown to have a north and south polarity comparable with that of the Earth.

Thus, in the early part of the last century, the conventions had become firmly established and interlocked.

In successive decades, many famous physicists and engineers expanded the basic concepts and developed practical equipment, ultimately giving birth to a fledgling electrical industry, with batteries, generators, motors, lighting systems, communications systems and so on.

Associated with it was a growing mountain of textbooks, literature and academia — all conforming to the original assumptions about voltage polarity and the direction of current flow and magnetic flux. These relationships were epitomised in Fleming's classic right-hand and left-hand rules, which have graced electrical textbooks for more years than most can remember.

But, around the turn of the century, Franklin's long-standing electric "fluid" concept began to be questioned by Rutherford, Bohr, Thompson and others, who put forward the then quite revolutionary electron theory. It had enormous implications.

They maintained that atoms were not, after all, the basic building blocks of matter. That atoms, in turn, were made up of still smaller particles (the real building blocks) notably "protons" with a positive electrical charge and "electrons" with an equivalent negative charge.

Each atom was like a miniature solar system with a nucleus of positively charged protons and (usually) an equal number of electrons spinning around it in specific orbits. The number and arrangement of protons and electrons determined the nature of the atom.

In certain substances, electrons in the outermost orbit could migrate from atom to atom. If, for any reason, there was an overall surplus of electrons in a particular area, it was equivalent to a negative electrostatic charge; a deficit of electrons was equivalent to a positive electrostatic charge. A discernable migration of electrons from one region to another was an electric current.

As a logical extension of this, a cell (or battery) could be regarded as a device in which internal galvanic (electro-chemical) action produced a surplus of electrons at one pole (negative) and a lack of such electrons at the other (positive) pole. In a rotary generator, a similar result was produced by magnetic effects. Similarly for an electrostatic generator or any other source of electrical energy.

The current which flowed when the two poles were bridged by an external wire was not some intangible "fluid", but simply a migration of free electrons through the wire.

NEGATIVE TO POSITIVE!

More to the point, this movement was from negative to positive!

The electron theory was not well received at first, partly because it challenged existing physical, chemical and electrical concepts, and partly because it appeared to contradict the "sacred" laws about polarity, current flow and magnetic flux. Indeed, it remained a "theory" for many years, until progressive developments rendered the traditional view untenable.

What gave particular point to the argument was the development of the thermionic valve, better known to the Americans as a vacuum tube or radio tube. As it happened, the operation of this vital new device was relatively easy to depict and comprehend and, of course, it depended on electrons migrating from atom to atom and even across free space.

In a thermionic diode valve, electrons emitted by a heated filament (or cathode) were attracted to an adjacent

THE TEACHERS' DILEMMA . . .

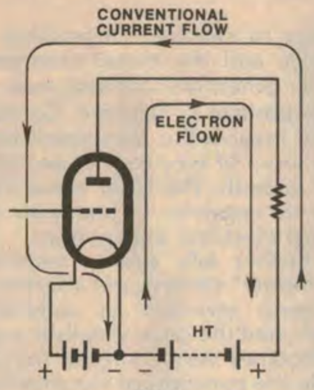


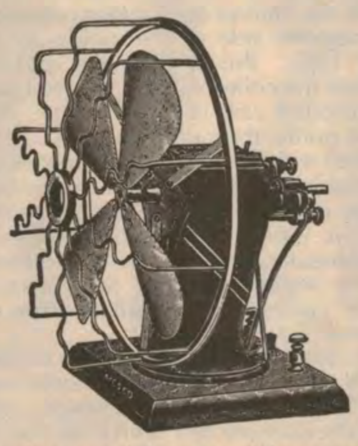
FIG. 1

In a simple triode valve circuit, electrons flow from the negative side of the (HT) high tension supply to the filament or cathode, across the evacuated space to the valve anode, then through the external anode circuit to HT+. This conflicts with the conventional current flow concept from plus to minus.

plate (or anode) within the evacuated glass envelope, whenever the plate was at a positive potential with respect to filament. On this property depended a diode's ability to function as a rectifier or detector.

In a triode valve, electrons were emitted from filament to plate in the same manner, except that the electron stream was subject to further control by an additional electrode (a "grid") mounted between the filament and plate. This made it possible for a triode to act as an amplifier or oscillator.

To understand the operation of wireless circuits using valves, it seemed quite natural to think in terms of electron flow: from the negative side of the supply to the valve filament (or cathode), thence across to the valve plate (or anode), and on through components in the external circuit back to the positive side of the supply. (Fig. 1.)



On sale around the turn of the century, this electric fan was a product of the "fledgling electrical industry". One thing it couldn't cool was the then current argument about the nature of electricity: was it a "fluid" flowing in the accepted direction or a massive migration of electrons moving the other way?

In wireless magazines and textbooks, arrows were commonly drawn on circuit diagrams to emphasise this quite fundamental concept.

Over the following decades, as the science and application of wireless expanded (call it "radio" if you like), an inevitable confrontation developed.

The wireless/radio fraternity had no real option but to develop their understanding on the basis of electrons and electron flow. But, to them, electron flow was current flow; how could it be any other way? Indeed, such was the logic of their position that they were confident that, one day, electrical types would have to see it that way!

But, in this, they were disappointed. Possibly because they were less directly concerned with electron physics, the electrical fraternity chose to retain their long established conventions governing current flow and magnetic polarity. What's more, they have never budged from that position.

BASIC CONFLICT

In the centre of things, teachers and lecturers were commonly obliged to explain electron (therefore current) flow in one direction, despite the fact that other members of the staff were talking about current flow in the opposite direction!

Publications like our own were also caught in a bind, particularly as electrical and radio interests began to overlap onto the common ground which is now all part of "electronics". As surely as we gave expression to the electron/current concept, just as surely would readers challenge us with chapter and verse to "prove" that we had our ideas and arrows the wrong way round!

More or less of necessity, we adopted the following practices:

- The word current was used only in a generalised sense; eg, voltage and current.
- If the context required a reference to the direction of current flow, we added the word "conventional", thereby indicating that the statement was in accordance with the established electrical convention.
- In the context of radio (particularly valve) circuits, and where possible, we wrote in terms of electron flow.

This practice, also followed by many other electronics publications, minimised confusion and kept most people happy. Or, perhaps we should say: made them less miserable!

In general, we still follow the practice, but the emphasis has changed again since the advent of solid-state devices: diodes, transistors, integrated circuits, etc.

Unlike valves, their operation is not

CURRENT — which way does it flow?

at all easy to visualise. Depending on the device and the circuit configuration, the potentials applied may be variously positive or negative. Conduction may involve the movement from atom to atom of electrons or so called "holes" or both. The latter move from positive to negative, in line with the traditional electrical convention.

Add further talk about "majority" and "minority" carriers and a variety of other terms invented by solid-state physicists, and the once simplistic argument becomes very confused indeed. So, while the concepts of electron flow and movement remain as valid as ever, the urge to make an issue out of them has greatly diminished.

It may even be said that, if the original positive-to-negative convention was a guess, the view of electronics adherents in the valve era was, itself, an over-simplification of the mechanism of conduction!

No less to the point, solid-state devices are represented, in many cases, by circuit symbols which perpetuate the concept of conventional current flow. Whether he likes it or not, the modern electronics devotee is stuck with them.

It seems almost certain that this came about because the very early rectifiers were allotted circuit symbols which involved an arrowhead for the anode pointing in the direction of (conventional) current flow to a line representing the cathode. For the sake of consistency, this symbol has been maintained for all rectifiers, right through to the modern (?) germanium and silicon diode.

When the transistor arrived, virtually a derivative of the diode, its symbol extended the diode convention. As a result, the emitter lead always points to the negative side of the supply. In an NPN type, the arrowhead points towards the collector, which is fed from the negative rail. In an NPN type, the arrow points the other way.

Faced with this situation, followers of electronics are tending not to argue but simply to adapt to the traditional convention. Where there is the likelihood of ambiguity, they can do as we have already suggested: refer to "conventional" current flow, or talk in terms of electron flow, or both.

It is possible to come up with similes which purport to bridge the remaining gap. Some time ago, for example, we used the example of a nearly-full row of seats in a theatre; by moving the patrons (electrons) in one direction, vacant seats (holes) could be concentrated at the opposite end. A Tandy publication invokes an alternative picture of a tube nearly filled with liquid; as the tube is tilted to and fro, the liquid (electrons) flow in one direction while bubbles (holes) move in the other.

Whether such mental pictures are necessary, or even helpful, at this stage seems debatable. Having come to appreciate the reasons for the apparent confusion, it is probably best to simply accept the situation and conform to it.

Which just about wraps it up, except for one point: why is the cathode of a semiconductor diode commonly thought of or marked as the "plus" end? Traditionally, and in just about every other device, the cathode is regarded as the negative electrode. Why the difference?

In fact, the diode is not an exception. In terms of conventional current flow, its conduction is also from anode to cathode. The markings on a diode follow from the fact that they are most commonly used in rectifier situations, with AC input and DC output.

Consider the basic circuit of Fig. 2, which could represent a diode functioning as either a rectifier in a mains power supply, or as a signal detector, (ie, a signal rectifier) in a radio receiver.

An alternating voltage is typically fed to the diode from a winding, which may be part of a mains transformer or a tuning coil. During the half-cycles when the upper end of the coil is positive, the diode conducts and feeds that positive potential to the filter capacitor and to the upper output terminal. On the reverse half-cycle, the diode does not conduct, since a negative voltage is being applied to the anode.

In short, only positive pulses are applied to the upper output terminal. The pulses are stored and smoothed by the filter capacitor.

Without trying to go any further, the point we want to make becomes apparent. In a rectifier situation, as in Fig. 2, the DC output is derived from the cathode end; hence the marking convention.

If the diode in Fig. 2 was simply reversed, the diode would conduct on the alternative half-cycle and the upper output terminal would become minus — derived from the anode!

THE DIODE PROBLEM

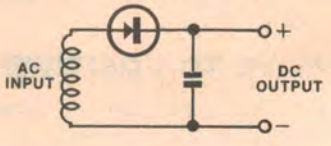


FIG. 2

In a diode, the anode is represented as an arrow pointing, in the direction of conventional current flow, to the cathode. The latter is sometimes marked "plus" because in a rectifier circuit, as shown, it is at the positive side of the DC output.