# The diode as an energy-controlled, not a charge-controlled device

**The conventional notion of electricity as electrons jostling their way down a wire cannot, according to Catt's Anomaly, be the thing that arrives at and controls the operation of a diode.**

IVOR CATT

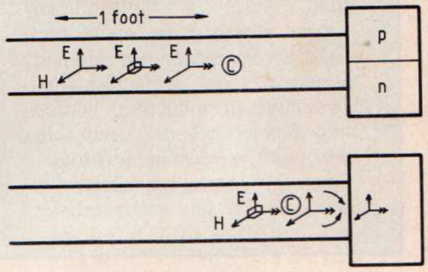
The traditional theory of operation of the diode is, for me, one of the many casualties of advances in electromagnetic theory during the last 25 years.

Whilst at Motorola, Phoenix, in 1964, work on the problem on how to interconnect high-speed (one nanosecond) logic gates led me to the same general conclusion as had been reached (unknown by me until 1972) by Oliver Heaviside a century before when he tackled the problem of how to improve undersea telegraphy from Newcastle to Denmark.

> "... (The electric and magnetic fields) are supposed to be set up by the current in the wire. We reverse this; the current in the wire is set up by the energy transmitted through the medium around it. The sum of the electric and magnetic energies is the energy...
>
> "... A line of energy-current is perpendicular to the electric and the magnetic force,..."
>
> – O. Heaviside, Electrical Papers vol.1, 1892, p.438.

Our conclusion was that what he called the "energy current" travelling down between the two conductors (i.e. the Poynting vector) guided by them as a train is guided by two rails, was the important feature of signalling, and not the electric charge and electric current in or on the wires. Twenty years later my view hardened when I came across an anomaly, explained in panel 1.

Let us deliver a 1ns-wide pulse down a long transmission line terminated by a diode. When the pulse reaches the diode, *it does not carry any charge with it;* Catt's Anomaly shows that charge could not have travelled fast enough to keep up with the pulse, which travels at the speed of light. If we are agreed that the diode will respond (for instance 'start to conduct') after a delay which is small (say 100ps) compared with the time delay down the transmission line delivering the pulse, then it must be responding to the *energy current,* that is the t.e.m. wave or pulse approaching it in between the two conductors. This t.e.m. pulse enters directly into the side of the crucial interface or surface between the p-region and the n-region which together make up the diode.




Note the phrase in panel 2: "Nothing ever traverses a capacitor from one plate to the other." Applied to the diode, this seems to say that nothing travels across the junction from the p-region to the n-region, or vice versa. The only travel is along the surface between the two regions, in a direction *at right angles* to the generally supposed direction of movement.

When the leading edge of the pulse reaches the near edge of the diode, it finds a change in characteristic impedance. As a result, most of it is reflected, but a very small portion continues forwards to the right, down the very narrow transmission line made by the surface between the p and n regions. It is possible that the effective dielectric constant $\epsilon$ is large so that the velocity of propagation, $(\mu.\epsilon)^{-1/2}$, from left to right along the p-n interface is very slow. At the speed of light in a vacuum, the round trip across the p-n interface of a diode a tenth of an inch wide would be 20 picoseconds but since the effective $\epsilon$ will be bigger than that for a vacuum, the delay will be greater.

When the step reaches the right-hand edge of the diode, it sees an open circuit and reflects back toward the left, so that the total voltage across the junction doubles. When it gets back to the front (left-hand) end, it reflects toward the right again (except for the very small portion which escapes across the $Z_0$ mismatch back into the transmission line leading away to the left). By this mechanism of zig-zag repeated reflections across the diode, the amount of energy

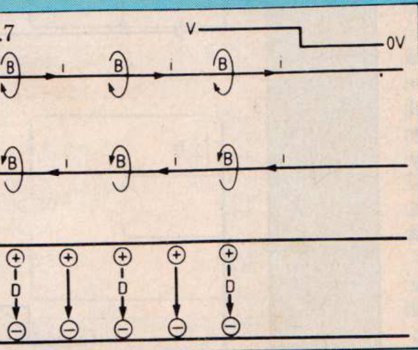
## 1: CATT'S ANOMALY

Traditionally, when a TEM step (i.e. a logic transition from low to high) travels through a vacuum from left to right (Fig.7), guided by two conductors (the signal line and the 0V lines), there are four factors which make up the wave:

— electric current in the conductors
— magnetic field, or flux, surrounding the conductors
— electric charge on the surface of the conductors
— electric field, or flux, in the vacuum terminating on the charge.

The key to grasping the anomaly is to concentrate on the electric charge on the bottom conductor. During the next 1 nanosecond, the step advances one foot

Fig.7




to the right. During this time, extra negative change appears on the surface of the bottom conductor in the next one foot length, to terminate the lines (tubes) of electric flux which now exist between the top (signal) conductor and the bottom conductor.

Where does this new charge come from? Not from the upper conductor, because by definition, displacement current is not the flow of real charge. Not from somewhere to the left, because such charge would have to travel at the speed of light in a vacuum. (This last sentence is what those "disciplined in the art" cannot grasp, although paradoxically it is obvious to the untutored mind.) A central feature of conventional theory is that the drift velocity of electric current is slower than the speed of light.

For further information on the Catt Anomaly, see letters in the following issues of *WW* 1981: Aug, 1982: Aug. Oct. Dec. 1983: Jan.

For what happens next, refer to my new model for the capacitor, panel 2.

(current) in the p-n surface increases 'exponentially' as indicated by the graph. When the energy density builds up beyond some critical level (0.7V), there is a 'snap', and the later advancing energy current sees a short circuit, and reflects with inversion.

Since no charge has been introduced into the p-n interface, it is totally inappropriate to explain the mechanism of the diode in terms of extra electrons. The explanation must be novel, in terms of the amount of electromagnetic energy present; that a level in excess of some critical value (0.7V) causes the t.e.m. wave travelling down the p-n interface to see a change in what is ahead of it, from open circuit to short circuit. That is, beyond that critical amplitude the p-n interface cannot accept more energy and rejects it.

This developing analysis of the behaviour of a diode is totally at odds with the traditional view, based on electrons, holes, energy barriers across the p-n interface that charges are trying to climb up. Why does this earlier theory succeed in correlating *at all* with experimental results?

"... if in conversation you insisted that your elder daughter was *identical* to your younger daughter, whereas in fact their "equality" only related to their parentage, every conclusion that followed this absurd assertion would not necessarily be absurd. For instance, if you knew the address of one daughter you might therefore know the address of the other. In the same way, it is possible for "valid" results to come from absurd postulates (like the absurd postulate that a diode is full of particles called electrons buzzing around trying to climb hills). "... the two non-identical daughters might have the same address. It is these 'echoes of truth' which masquerade as scientific truth today."
– I. Catt, Electromagnetic Theory vol.1, CAM Publishing, 1979, p15.

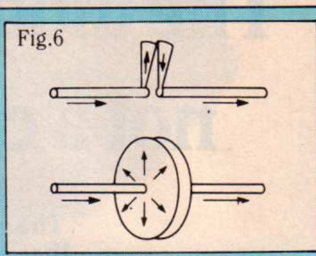
False theories, like the theory that the diode is a device controlled by charge, exist in the real world, and so are influenced, or somewhat directed, by the imperatives of the real world in which they find themselves, at least when it comes to the moment of truth: the checking of theory against experimental result.

## 2: THE CAPACITOR

In the early 1960's I pioneered the inter-connection of high speed (1 ns) logic gates at Motorola, Phoenix, Arizona[12]. One of the problems to be solved was the nature of the voltage decoupling at a point given by two parallel voltage planes. I asked Bill Herndon about this problem, and he gave me the answer: "It's a transmission line"[13]. Bill learnt this from Stopper, whom I never met, who now works for Borroughs in Detroit.

The fact that parallel voltage planes, when entered at a point, present a resistive, not a reactive, impedance, was for me an important breakthrough. (It meant that as logic signal speeds increased, there would be no limitation presented by the problem of supplying +5V.) The reader should be able to grasp the reason why voltage plane decoupling is resistive by studying Fig. 6, which shows the effect of a segment only of two planes as they are seen from a point[14].

During the next ten years, with the help of Dr D.S. Walton, I gradually came to appreciate that, since a conventional capacitor was made up of two parallel voltage planes it also had a resistive, not a reactive (i.e. capacitive or inductive) source impedance when used to decouple the +5V supply to logic. Since the source impedance (= transmission line characteristic resistance) is well below one ohm, the transient current demand of logic gates approaching infinite speed can still be successfully satisfied with +5V from a standard capacitor of any type[15]. The reason why the myth has developed that the worst (low capacitance, 'r.f.') capacitors are the best in this role is discussed elsewhere[16].




The capacitor is an energy store, and when energy is injected, it enters the capacitor sideways at the point where the two leads are joined to the capacitor. Nothing ever traverses a capacitor from one plate to the other. This is clearly understood in the case of a transmission line. By definition, when a TEM wave travels down a transmission line, nothing travels sideways across the transmission line, or we would not have a transverse electromagetic wave.

**For further information, see the *Wireless World* articles dated December 1978 and March 1979.**

**Excerpts 1 & 2 are from Fundamentals of electro-magnetic energy transfer, by Ivor Catt, *Wireless World* September 1984.**

WIRELESS WORLD, DECEMBER 1978

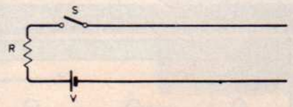
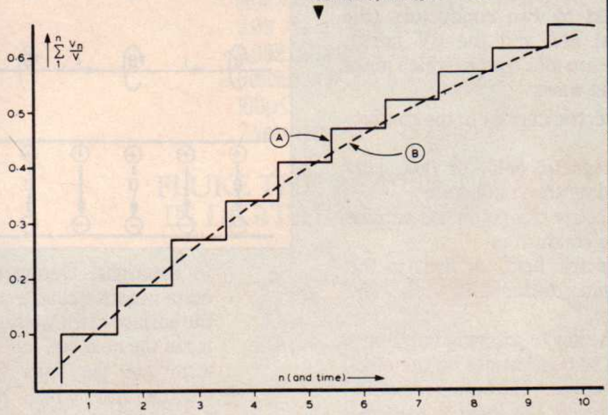



**Fig. 2** *An open-ended transmission line.*

**Fig. 3** *Comparison of the transmission line model $1-(1-2Z_0/R)^n$ in the curve A with the lumped model $1-e^{-2Z_0 n/R}$ in curve B, for $2Z_0/R = 0.1$.*




## Semiconductor Centre

The Semiconductor and Microelectronics Centre, a combined research and development centre created by the University of Wales Institute of Science and Technology and University College, Cardiff, has recently opened in Cardiff. The centre will be especially looking at new semiconductor materials of the III –V group, possibly combined with material from the II – VI group.

UCC physicists have developed a wide range of techniques to build and examine semiconductors, while UWIST has a clean room facility to use construct transistors from the materials. Both groups work closely with industry to find applications for the new chips. Projects have already included the development of a thermoelectric generator for the US space programme, lasers for communications systems and high-speed devices.

SEPT 1987