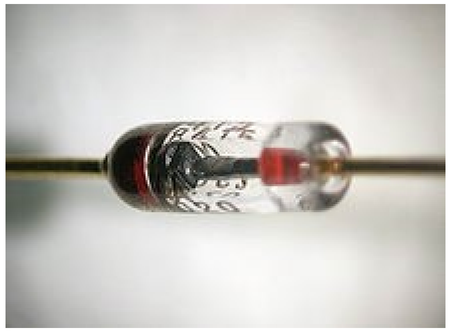
return to updates

P-N Junctions without Holes



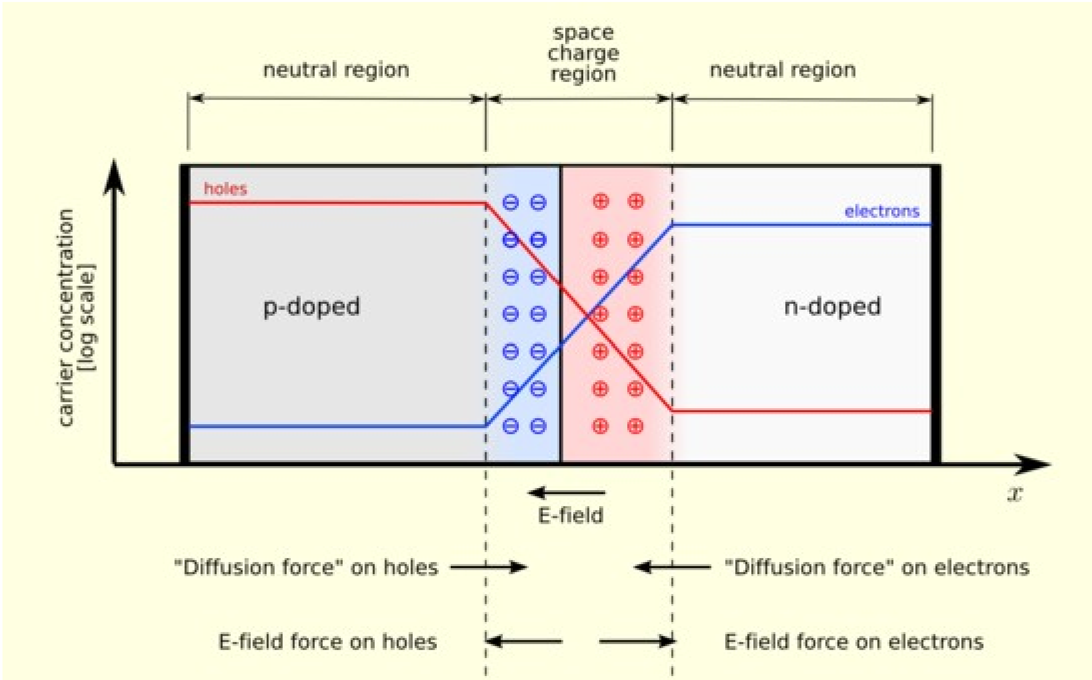
by Miles Mathis

First published June 24, 2014

A P-N junction is a term used in semiconductors. It is the boundary between two types of conducting substances, a **P**ositive type and a **N**egative type. Although current physicists and engineers are quite adept at getting the results they want with semiconductors, the theory beneath these fields is pathetic. I will correct it and it flesh it out here using my charge field, showing you that the correct theory is completely mechanical and completely visualizable, without any cheats or fudges.

Currently, P-N diodes are explained with an embarrassing series of cheats, including electron holes, quantum tunneling, and electron donors. I have been told that my critique of quantum tunneling cannot be right because these diodes are experimental proof of it. But no experiment can be proof of such a lame, non-physical theory. The experiments are only proof of the results, which we then must explain with theory. But what we want is a sensible physical theory, not a series of illogical and insupportable pushes. And we want clear, mechanical diagrams, not fudged diagrams of manufactured non-particles and impossible motions. The theory of diodes is actually one of the worst theories ever dreamed up by the human mind, and it has had a lot of competition recently.

Before I show you the right answer, I will show you the current theory. I don't have to beat it up or catch it on a bad day, since all the mainstream sites have published it in its full ugliness. All I have to do is circle the things they already admit. To start with, I will republish their diagram of the reaction:



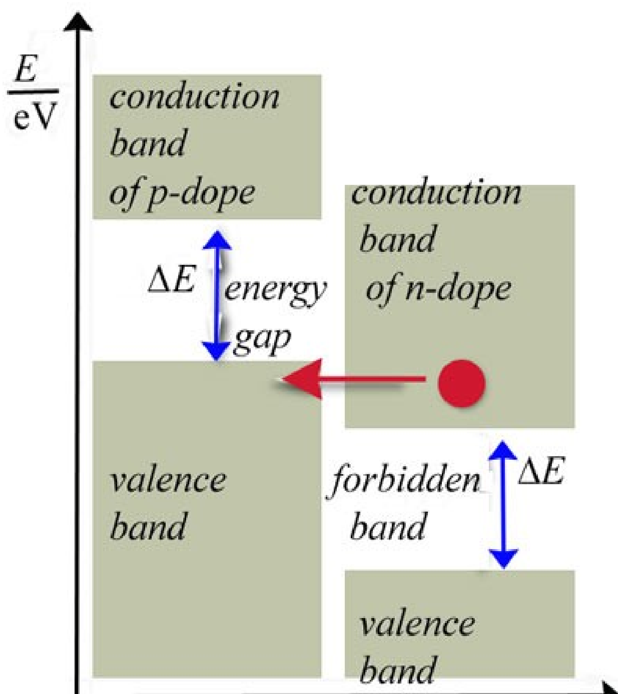
It helps if we apply that diagram to some real elements, so that you understand what is going on here. The most common semiconductor is Silicon, of course, and it is often P-doped with Boron and N-doped with Phosphorus. Silicon is element number 14, while Boron is 5 and Phosphorus is 15. Doping just means those elements are added to Silicon to make it more conductive. But since the conductivity of the two doped areas are still different, a voltage is created across the junction. This built-in voltage can be later augmented by attaching the whole diode to a battery of some sort.

The biggest fudge of many there is the electron hole, represented by the blue circles. We will get to those in moment. But first, study the red circles. The red circles are supposed to be the electrons. So they are actually giving the electron a plus sign here. That is to stir your brain. The only way you could possibly accept any of this is if your critical faculties have been utterly destroyed, and by giving the electron a red plus sign the theorists hope to hypnotize you into a state of faux-rapture, in which their every suggestion becomes your perfect bliss.

That's bad enough, but the electron hole is an even greater piece of mystification than the plus-red electron. You see, that blue circle doesn't represent any known particle. No one has ever captured or witnessed an electron hole, and the only "proof" of it is found in these diagrams. The theorists come up with it, they draw it, and then the fact that they drew it is supposed to stand as evidence of its existence. They assign their diagram to some experiment, and so the entities in the diagram become "experimental." But if you can still remember a time before the brainwashing, you will remember that physics didn't work that way. Back when people still had some critical faculties, a diagram or theory was only accepted as a likely explanation if it fit the data in a lovely, natural, and reasonable way. Making up particle holes is not any of those things. Back in the days of real science, anyone who proposed or accepted something like a particle hole would have been immediately drummed out of the field as a nuisance, if not worse.

You may also ask why they have labeled the two doped regions as "neutral," although they have already told us both are conductive. If they are conducting, they can't be neutral, either absolutely or relative to one another. Again, your mind is being squeezed. They are weaning you off charge differentials or potential differences—which are charge characteristics. They want you looking mostly away from charge and toward the electrons and electrons holes. Using textbook theory, they have no way to explain this with real charge field characteristics, since in the standard model the charge field doesn't have any. So they have to explain everything with ions. As with A/C and D/C current, they have to keep teaching you that electrons are causing everything, even though they know that isn't true. They have known for a century that electricity isn't the motion of electrons, but since they have nothing else to sell you, they keep selling you electrons. We saw that in my paper on the battery circuit, where they admit the motion of electrons isn't the cause of anything. It can't be because the electrons are moving too slowly to cause the effect. The same is known to apply to A/C and the properties of transition metals, but they go out of their way to pretend they don't know it. We saw it with my critique of the Drude-Sommerfeld model, which is a lot like this Diode model. They know it isn't the motion of electrons causing it, and they can't explain it with electron motion anyway; but rather than try something new, they just hammer on the electron model until it yields the ugliest and most irrational theory possible, then rush it into permanent print. They then pass out some ugly prizes for these hideous theories to cement them in, and hire a few thousand people to force them down the throats of anyone who blunders by. That is new physics in a nutshell.

Here's another diagram from the mainstream:



As you see, we have energy per electron volt on the vertical axis, but nothing on the horizontal axis. So why is the n-doped band to the right of the p-doped band? Why is it above the valence bands and the forbidden band? Why is one gap called an energy gap and the other a forbidden band? That's just a terrible diagram no matter how you look at it, but to see how bad it really is, we have to look more closely at the underlying theory. Returning to Wikipedia, we find this:

After joining p-type and n-type semiconductors, electrons from the n region near the p–n interface tend to diffuse into the p region. As electrons diffuse, they leave positively charged ions (donors) in the n region. Likewise, holes from the p-type region near the p–n interface begin to diffuse into the n-type region, leaving fixed ions (acceptors) with negative charge. The regions nearby the p–n interfaces lose their neutrality and become charged, forming the space charge region or depletion layer.

Although this is meant as an explanation of the physics involved, it contains no physics. To start with, we are told the electrons diffuse across the gap into the p-region. Likewise, electron holes (which are positive) diffuse into the n-region. We are told this leaves areas of higher charge in the boundaries, but that is illogical. Study the stated mechanism closely and you see that all we have is a charge swap. If electrons move into the p-region, then the p-region becomes *more neutral*. And if electron holes move into the n-region, that region also becomes more neutral. All the theorists have done is perform a poor magic trick, telling you to keep your eyes on the donors and acceptors. Since they have become ionized, you are told we have more charge. But what about the electrons and holes, which are also ions? Have they just evaporated?

Just think about it: they just told you the p-region has gained a lot of electrons but become *more* charged? How's that? That doesn't work, does it? The p-region can only attract the electrons if it is positively charged to start with, hence its name. If we then add negative charge to that positive region and subtract positive charge (the holes), the region becomes *more neutral*. So why are they telling you it is becoming more charged?

The same applies to the n-region, in reverse. A negative region has lost negative charge and gained positive charge, therefore it must be more neutral. Why are they telling you it has gained charge? The only thing that has changed in the swap is mass, not charge. Since the electron hole has a negative mass, the p-region has gained mass and the n-region has lost it.

And what exactly has become "depleted" by this fake physics? They need a depletion layer in between the regions, so they just tell you one has been created. But I don't see how any depletion layer has been created, do you? We have a double increase in ionization in the layer, so what has been depleted? The electrons now in the p-region can't fill the places the holes left, since those holes are *negative*. The acceptors there have negative charge, as they admit. Electrons aren't attracted to negative charge, are they? So even if the electrons join to positive ions in the p-region (which were attracting them across the boundary), the p-region has still gained negative charge, which has made it more neutral.

You will say it is this neutrality they are calling depletion. But it wouldn't work that way, and they should know it. This neutrality wouldn't create a depletion layer: neutrality in a charge field disperses, by all the old definitions. Both regions as a whole should lose potential relative to one another, and the initial voltage would just drain out. In other words, the two substances would stabilize relative to one another. So the current theory is upside down to all sense in several ways.

To try to counter my obvious analysis here, Wikipedia says this:

The electric field created by the space charge region opposes the diffusion process for both electrons and holes.

No it doesn't. How and why would it do that? There was already an electric field there before the electrons and holes started moving: what else would be causing them to move in the first place? As you see, they are telling you a field was there that caused the particles to move, and that the movement of the particles then set up an opposite field. Nonsense from the first word. It is like saying the west wind blew my barn over, and my barn falling to the east caused an opposing wind to the west.

They know that neutrality should disperse, which is precisely why they had to come up with this idiotic theory of electron holes. The electron holes prevent dispersal of neutrality by recombining with electrons. However, the theory of electron holes is another non-starter, since the idea of holes traveling through space is a blatant fudge. We see that immediately on the Wiki page for electron hole, where the second sentence is this:

The concept describes the lack of an electron at a position where one could exist in an atom or atomic lattice.

The electron hole *has a position*. Even in this fudged math, it can only exist where the electron could exist: in an atom or atomic lattice. Having holes travel across boundaries is non-physical. It is a double cheat. To answer this, Wiki tells us this charming story:

Hole conduction in a valence band can be explained by the following analogy. Imagine a row of people seated in an auditorium, where there are no spare chairs. Someone in the middle of the row wants to leave, so he jumps over the back of the seat into an empty row, and walks out. The empty row is analogous to the conduction band, and the person walking out is analogous to a free electron.

Now imagine someone else comes along and wants to sit down. The empty row has a poor view; so he does not want to sit there. Instead, a person in the crowded row moves into the empty seat the first person left behind. The empty seat moves one spot closer to the edge and the person waiting to sit down. The next person follows, and the next, et cetera. One could say that the empty seat moves towards the edge of the row. Once the empty seat reaches the edge, the new person can sit down.

What they don't tell you is that could only work in a row of electrons. That is, it *might* work in orbitals [if we had orbitals], where one electron could jump down to a lower level, and so on. But it doesn't begin to explain motion of the hole through semiconducting material. In the p-n junction problem, we have many electrons not only coming into the auditorium, but being drawn to open seats there. If there were no open seats, why were they drawn there? So the electrons *aren't* waiting for the seats to open up (that is, they aren't waiting to meet an electron hole). If the electrons are moving to the p-region, it must be because open seats are *already* there, before the creation of electron holes. It *can't* be the electron holes attracting the electrons to the p-region, since that process would obviously be circular. The electron can only be attracted by a positively charged ion, since only the ion has a *position* waiting for it. Given current theory, the electron can only be attracted to an orbital position, since only that position has the right energy for the electron. You can't detach an empty orbital from a nucleus and let it travel across boundaries. But that is exactly what these theorists are doing. They are detaching the orbital position from the atom, calling an empty orbital a hole, and then moving it into the "charge space region." They have to do that to create this depletion layer. If they didn't do that, they could only create charge motion in one direction. Without moving electron holes, all you have is electrons moving to the p-region, but no opposite motion. And if you have that, you have no opposing charges. You just have a normal old current.

Or do you? Let's leave this embarrassing mainstream attempt at physics and move into some real physics. Since it is clear that the current theory explains nothing, let us use my charge field to show why the diode has the characteristics is has. To do that, we will once again ditch the electrons altogether. It is not electrons but charge photons that move from one region to the other.

The main characteristic of the diode is that it has low resistance in one direction and high resistance in the other. Another clue is that at high reverse bias, we find reverse breakdown that causes a large increase in current. Another is that a Zener diode can be created which prevents breakdown—or which forces it to happen at a lower voltage that causes much less damage. We will stick with those for now.

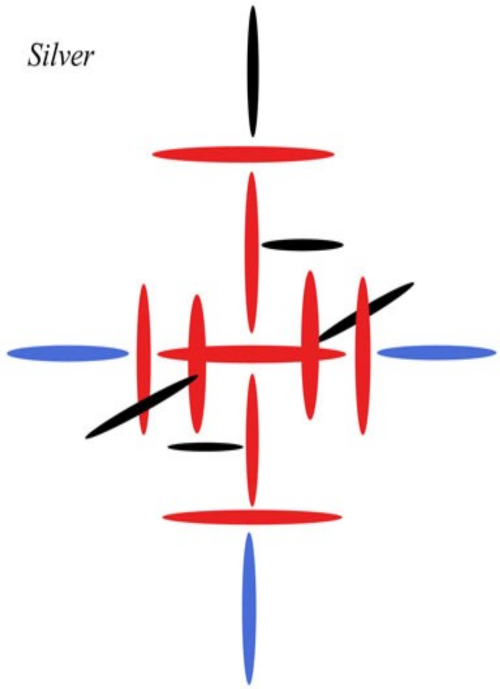
The biggest problem with mainstream theory is that it tries to solve the problem using the electric field only, ignoring the magnetic field. As you have seen, current theory doesn't mention the magnetic field once when talking about diodes. But as we have been reminded by my recent paper on A/C current, all electric fields have chirality (handedness). Even if they aren't created by spinning magnets, they have this chirality. In some instances, we can ignore this chirality, since it may not raise its head above the covers. But in the p-n junction problem, it has raised its head high.

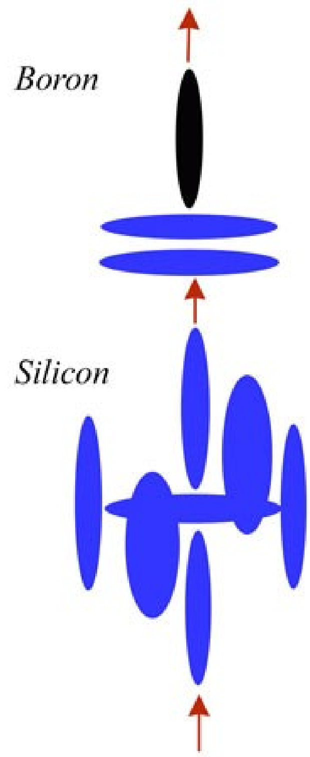
All electric fields have chirality simply because charge photons always have chirality. All current is caused by charge fields, charge fields are made up of photons, and these photons are always spinning. The summed spin of these photons creates chirality, and this chirality is what we call the magnetic field. As you will now see, this chirality explains the forward and reverse bias, the direction of resistance, and the cause of breakdown. And once we have this chirality, we can ditch mainstream theory altogether. Given my simple spin mechanics, you no longer need all the cheats of historical electrical theory.

We will start with the built-in voltage. This just means we have doped one side of the diode differently than the other, setting up a potential across the boundary. In some cases, this will cause a built-in current; in others it won't. Since the Earth has an ambient E/M field of some strength, that ambient field may be able to overcome the built-in resistance of the diode*. In that case, the diode will have current even with no voltage applied. But normally the resistance of the diode will prevent current up to some applied voltage.

OK, but what causes resistance in my theory? Resistance is simply caused by longer charge paths through the material. In a conductor, the atoms align pole-to-pole, allowing charge to channel through the nuclei in the most efficient manner. In an insulator or resistor, the channels through the nucleus are orthogonal, or perpendicular, creating longer paths. See my paper on the dielectric for diagrams of Iron and Sulfur, which show how this happens. In some cases, an applied voltage can turn nuclei, increasing conductivity. And in more extreme cases, *very* high voltages can break nuclei, actually re-arranging the nucleons to increase conduction even more. But nothing like that is happening here.

The reason doped Silicon has such a bias in one direction is due to the structure of the Silicon nucleus. You cannot understand the properties of semiconductors or diodes without knowing how the nucleus channels charge. This is because without that knowledge, it is impossible to explain why doped Silicon acts differently than metals like Silver. Doped Silicon is being made more conductive, so why doesn't it simply act more like Silver?



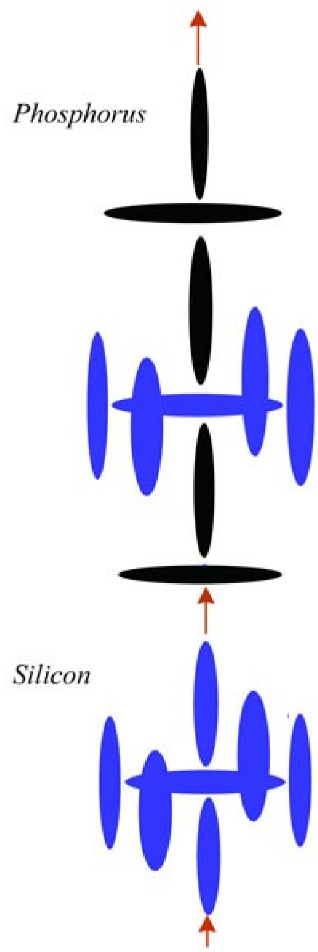


[Pay no attention to the skinny disks in Silver: I drew that diagram several years ago, and I tend to draw the disks skinnier in the bigger elements so I can fit more in. Just study the architecture of the nuclei. To see what the disks represent, consult my long paper on nuclear diagramming.]

Both these configurations are conductors because both have a differential bottom to top, along the pole. Both have blue disks bottom and black disks top. That is, two protons bottom and one top. The disks acts like little fans, so what this configuration means is that the charge streams know which way to go. There is a strong potential here for through charge going north. See my paper on Period Four for more on this.

So both configurations are conductors. But they aren't the *same sort* of conductors because we have many important differences in the architecture, as you see. To start with, while Silver has a strong carousel level, Silicon and Boron don't. Boron has no real carousel level at all (or only the hub), while Silicon only has the vertical alphas. It doesn't have the multiple horizontal protons in the 4^{th} carousel level that Silver has, pulling charge out the nuclear equator. This means that although the conductivity of Si-B isn't as strong as Silver (because Silver has a bigger core and more total channeling), it is more linear. Silver has a transverse or equatorial field that Si-B doesn't. This is why Si-B has a low resistance in the forward direction: once you align your Si-B to the field, the through charge moves through very easily, with little loss by the carousel levels.

To see why it doesn't like the reverse voltage, we have to look at Silicon doped with Phosphorus.



As you see, Si-P creates conductivity in the same way as Si-B, and it also has a weak carousel level. But here we have 29 protons channeling instead of the 19 of Si-B. So Si-P will both channel and conduct more. This means the charge field will move from Si-P to Si-B. In fact, that is what is known to happen. Consult the first mainstream diagram above, where the E field is moving right to left from Si-P to Si-B. In my theory, the E field simply follows the summed motion of the charge photons. Electrons then drift with that field. I do not allow, or need, electron holes.

Now, if we try to reverse the direction of this charge motion, we have a problem we don't encounter if we try to reverse the charge motion through Silver, say. If we reverse the voltage through Silver, we are just trying to reverse the channel through the nucleus. We can't reverse the channel through the nucleus, but under the right conditions, Silver can turn to accommodate the reverse bias. But we can't do the same with our diode. Even if all the nuclei flipped for some reason, that wouldn't solve the problem. We would still be trying to channel from Si-B to Si-P, and flipping all the nuclei doesn't help us do that. We would have to move all the Si-B right and all the Si-P left, you see, and that isn't what is happening.

This is why the diode has high resistance in the reverse direction. Charge simply doesn't want to move from Si-B to Si-P, no matter how the nuclei are aligned.

To understand what happens in Zener diodes, we have to understand that these diodes are more highly doped. The more doping you have, the less likely it is that all Boron or Phosphorus will be linked to Silicon. In that case, you can get reverse current running from Si-B to free P, or from B to P. Although Si-P is more conductive than Si-B, free Boron is more conductive than free Phosphorus.

How can that be, you may ask. Again, you only have to study the nuclear architecture. Taken alone, Phosphorus doesn't have the through charge Boron does, simply because Phosphorus has more alphas in the carousel level. These alphas pull charge out equatorially, and if charge is pulled out equatorially, it can't also be moving pole to pole. Boron has two alphas in the carousel hub and Phosphorus has four alphas pulling charge out the carousel level. So Boron has a ratio of one proton pulling in charge at the pole to four in the carousel, while Phosphorus has one to eight. This is why Phosphorus by itself is about half as conductive** as Boron.

You will say, "Doesn't Phosphorus have three protons on the axis? Shouldn't all three count in that ratio?" No, because those other two can't pull in more than the one at the pole. Since Phosphorus has a blue core, it *could* channel more than one unit of charge if we put more protons on the poles. But then of course it wouldn't be Phosphorus anymore. It would be Chlorine.

But once we plug Phosphorus into Silicon, everything changes. For a start, Silicon linked to Phosphorus won't act the same as Silicon linked to Boron. Simply because P is a larger element than B, it will pull more charge through Silicon. Remember, we are linking along the poles, so P on the pole of Si will draw more charge through than B on the pole. Although P doesn't conduct as well as B, it can draw more charge. Even before we look at conductivity, we know Si-P must be channeling more *total* charge than Si-B—about 53% more in the first analysis [29/19]. Although Phosphorus does blow out more charge equatorially than Boron, we find that Si-P still conducts better than Si-B. We can do the quick math. Let us say that Boron conducts 25% of its total charge intake, as I showed above (1 to 4). That would mean that Phosphorus conducts about 11% (see thermal conductivity numbers for both elements or the second footnote below). If Phosphorus channels 300% more total charge than Boron (15 to 5), then we multiply 25% times 100 and 11% times 300, giving us 25 for Boron and 33 for Phosphorus. Therefore, if we plug the same element into Boron and Phosphorus, Phosphorus will increase the conductivity of that element 32% more than Boron will.

You will say, "By that analysis, Phosphorus should always channel 32% better than Boron, in which case you can't explain why P has a thermal conductivity 2.26 times less than Boron." That seems like a good argument, but to make it you have to ignore how conductivity is normally measured. It isn't measured *relative to* a given or standard element, as we are relating B and P to Si above; it is measured relative to an ambient or applied field. Heat or charge is applied to the element in question, that is, and the transference is measured on the far side. In that case, P *won't* take in 300% more charge than B, because there isn't 300% more charge to take in. The ambient or applied field is a constant, so P takes in what B takes in, giving us more conductivity with B.

But if we plug both B and P into Si, Si's channeling ability is variable. It has a maximum charge potential across it, but it is never anywhere near that potential on Earth. Therefore P is free to pull more charge through Si than B does. This confirms my analysis and math above. And it shows us why a diode that is doped correctly can create a reverse current. Initially it doesn't want to, because the built-in potential has aligned the atoms in the forward bias positions. But at a given voltage, the atoms and faux-molecules can reverse. This still won't create current from Si-B to Si-P, but it can create current from Si-B to P and from B to P.

That is what is happening in the Zener effect, not any quantum tunneling. But what is happening with the avalanche, when we have lower doping? This avalanche requires a higher voltage than the Zener reversal. I would assume that the Si-B is harder to turn than the free P or B. In the Zener effect, we may be seeing only the P and B turn, creating a path from the second to the first. In the avalanche, we are probably seeing the Si-B turn, allowing it to channel to P. At even higher voltages, we could probably force Si-P to turn as well, and force a current from Si-B to Si-P, but I doubt that is happening here. It would take more study to figure out which voltage applies to which mechanics. But as you can see, we have lots of possibilities here, all of them better than electron holes.

To complete this paper, let us look at the so-called tunnel diode. This diode is supposed to be proof of quantum tunneling, and two physicists—Brian Josephson and Leo Esaki—won the 1973 Nobel Prize for the theory of tunneling in regard to this diode. In fact, this is why the theory still stands, despite being another awful push. These big prizes cement the theories into place, and those who win the prizes (and all their students) block any better theories until they are all in the grave.

In a tunnel diode, both sides are heavily doped, creating no reverse resistance. The breakdown voltage goes to near zero, so this diode can conduct in both directions. The weird thing is that this diode can create a decrease in forward current even with an increase in forward voltage. Since this was unexplainable using the given electrical theory, these old guys needed a trick, and the trick they chose to apply was quantum tunneling. Quantum theory is a big bag of tricks, and they could have used any number of them, but they chose quantum tunneling for some reason. If I had been around at the time, I would have recommended they use borrowing from the vacuum or symmetry breaking of some sort, but it is too late for that. These cheats would have been preferable since they aren't quite as transparent. They aren't so easy to shoot down.

Quantum tunneling is a poor fudge here because it is difficult to apply to the problem at hand. This is because quantum tunneling was originally proposed decades earlier to fudge solutions to wavefunction problems. In the diode, the problem is not a wavefunction problem. It is simply a field problem.

I will tell you what I mean. I have shown that the current wavefunction is short a degree of freedom. I did this while shooting down non-locality. The Bohr and Schrodinger equations are also flawed in major ways, and I have gone through both line-by-line, showing the obvious flaws. Because these early equations were wrong, physicists in the 20^{th} century quickly came upon data they couldn't explain. The particles had too little energy or too few degrees of freedom to do the things they were caught doing in experiments. Since these physicists didn't know how to correct the equations, they fudged them instead. Quantum tunneling is an obvious fudge because it basically says that although the equations say the phenomenon shouldn't happen, it happens anyway because probability is never zero. We are told that in quantum mechanics, nothing is impossible. This is convenient for theorists, since it means that nothing will ever disprove one of their theories. All their equations then take on the status of "suggestions," but the field can ignore them at will. If data comes along that does not confirm the theories or equations, that data is not negative data. It is just data that indicates quantum tunneling or one of the other dozens of cheats.

But tunneling itself was originally a fudge that only applied to wavefunction problems, since it was proposed to explain how quanta got through barriers they shouldn't have gotten through. It was wave equations that were being fudged. But with the diode, we aren't really seeing that. There is no barrier here. We are dealing with electrical field equations, not wavefunctions. More than that, it is a field *reversal* we are trying to explain. Something is lowering current while voltage is rising. Even at a glance, you can see that tunneling doesn't address that. Given current theory, we would need to explain why electrons were turning around or getting blocked, not why they were tunneling.

Once we fill in the old theory with some real mechanics, we don't need any of these old cheats. I have expanded the wavefunction and corrected the Bohr and Schrodinger equations, so quantum tunneling was destroyed from its foundations. But here, we don't even need to bring in the wavefunction. We can solve the problem with field mechanics. We don't have to explain why or how the electron does anything, since I have shown the electron is just a buoy in the field. Its motion doesn't matter. The electron isn't the field particle, so we don't follow the electron to explain the field. We follow the charge field, not the electron. Since the charge field follows current created by charge channeling by the nucleus, that is the field we study.

The important fact of the tunneling diode is that it has zero or low resistance in both directions. From my analysis above, this just means that it has charge paths set up in both directions. There is so much dopant in the diode that much of it isn't even used in the forward bias position. When we apply the forward voltage, enough current is created by the doped Silicon that the free dope doesn't even have to turn. In other words, if we stick with B and P, we have enough free B and P that much of it doesn't have to align with the Silicon to support the forward voltage. It therefore remains aligned in the reverse (or random) position. If we reverse the voltage, that free B and P can align to create the reverse current. And once you apply the reverse voltage, the free B and P is then aligned in the reverse position. Both paths exist simultaneously, and the paths do not interfere.

Now, what happens when the forward voltage begins causing a decrease in forward current? Obviously, it has nothing to do with tunneling. Something is causing the forward paths to close or lengthen. It is simply the voltage at which the free B and P is no longer left alone by the forward current. The forward charge stream is no longer contained by the forward path of the Si-P to Si-B, and it overflows the banks. It begins spilling out into the charge path of the B to P, and as it washes past those atoms, it ruins their reverse alignment. Some are washed into the Si-P to Si-B path, diminishing it. This diminishment is what we read a new resistance. The forward current drops.

I told you I could make all this mechanical and easy to visualize, and I have. I haven't answered every question in this short paper, but I have shown that all the questions have a straightforward answer, one that does not require any sort of mainstream fudging or hedging. In newer papers, I will continue to lead you through the solutions. But you can already see that having a real charge field and real channeling is the key to this and just about every other similar problem.

*Especially if you position the diode vertically.
**To find the actual number 2.26 would require we look at the secondary charge channels (like of the cap protons), and I won't do that here.