4.4 Operation in the Reverse Breakdown Region — Zener Diodes

Reading Assignment: pp. 190-191

A Zener Diode is simply a p-n junction diode that is meant to be operated in breakdown.

Understanding the differences between a Zener diode and a junction diode is crucial.

**HO: THE ZENER DIODE**

Since a Zener diode is usually operating in the breakdown or reverse bias regions, we alter the current/voltage notation (but this does not alter the device) for Zener diodes.

**HO: ZENER DIODE NOTATION**

Our junction diode models are not accurate if a diode is operating in breakdown. Since a Zener diode is often operating in the breakdown region, we require some new diode models.

**HO: ZENER DIODE MODELS**

Let's use these new models to analyze a Zener diode circuit!
**Example: Fun With Zener Diodes**

**Q:** I'm tired of all these *pointless academic exercises*!

*What's the point of diodes and diode circuits if they don't actually do something useful? Can't we get to the “practical” and “hands-on” portion of this course?*

**A:** That is exactly the place at which we have arrived in this course (or at least, with respect to diodes)! Note that you now:

- understand the operation of junction and Zener diodes
- understand how to approximate these devices using ideal diode models (e.g., the CVD model).
- understand how to analyze junction and Zener diode circuits, implementing various ideal diode models.

Our tool box is now full—it's time to go build something useful!

The first useful application we will consider is voltage regulation. A Zener diode in breakdown acts somewhat like a voltage source—the voltage across it is nearly independent of the (positive) current through it. This makes a Zener diode in breakdown very useful for voltage regulation!
**HO: THE SHUNT REGULATOR**

Say your boss wants you to design a shunt regulator; be careful, engineering is not as easy as it looks!

**EXAMPLE: THE SHUNT REGULATOR**

In addition to good load and line regulation, another important parameter of voltage regulators is efficiency—we do not wish to waste energy by creating heat.

**HO: REGULATOR POWER AND EFFICIENCY**

Finally, we should note that the shunt regulator is far from the only way to provide voltage regulation. Not only can we use more efficient switching regulators, better linear regulator designs (i.e., better than the shunt regulator) are available in the form of integrated circuit linear regulators.

**HO: LINEAR VOLTAGE REGULATORS**
The Zener Diode

➔ A Zener diode is simply a p-n junction diode that is meant to be placed in breakdown!

Q: Yikes!

A whole new electronic device to learn about.

I'm sure it's completely different than a junction diode; go ahead and start listing all the differences between a junction diode and a zener diode!

A: Um, OK.

Let's contrast the differences between junction diodes and Zener diodes!
The first difference

On the one hand, a junction diode is:

* an asymmetric device
* non-linear device
* a device with two terminals, called the anode and the cathode.

On the other hand, a Zener diode is:

* an asymmetric device
* non-linear device
* a device with two terminals, called the anode and the cathode.
The second difference

Three device parameters characterize a junction diode:

* its ideality factor,
* its scale current,
* and its zener breakdown voltage.

However, three device parameters characterize a Zener diode:

* its ideality factor,
* its scale current,
* and its zener breakdown voltage.
Now the third difference

Likewise, a junction diode has three operating “regions”, called:

* forward bias,
* reverse bias,
* and breakdown.

In contrast, a Zener diode has three operating “regions”, called:

* forward bias,
* reverse bias,
* and breakdown.
Yet another difference—
I’ve lost track of which one

We know that a junction diode operating in the forward bias region will exhibit:

* significant current will flow from anode to cathode,

* and the anode voltage will be around 700 mV higher than the cathode voltage.

But, we find instead that a Zener diode operating in the forward bias region will exhibit:

* significant current will flow from anode to cathode,

* and the anode voltage will be around 700 mV higher than the cathode voltage.
So completely different!

And, we know that a junction diode operating in the reverse bias region will:

* exhibit a tiny current flowing from cathode to anode,

* and a higher voltage at the anode than at the cathode.

Whereas, we know that a Zener diode operating in the reverse bias region will:

* exhibit a tiny current flowing from cathode to anode,

* and a higher voltage at the anode than at the cathode.
The last difference—a really big one!

Finally, a junction diode in the breakdown region will show:

* significant current flowing from cathode to anode,

* and a cathode voltage that is a value $V_{ZK}$ higher than the anode voltage.

Alternatively—in a stunning reversal—a Zener diode in the breakdown region will show:

* significant current flowing from cathode to anode,

* and a cathode voltage that is a value $V_{ZK}$ higher than the anode voltage.
**Sarcasm: it’s so darn effective**

Q: *What?*

*This Zener diode sounds **exactly the same** as a junction diode!*

A: That’s correct!

I’ll say it again:

“A **Zener diode is simply a p-n junction diode that is meant to be placed in breakdown!**”

Q: *But if a Zener diode is a junction diode, then **why** give it a different name?*

A: Because a Zener diode is a junction diode that is **meant to be placed in breakdown** (I hope you’re finally getting this)!
Speaking of annoying

Q: You seem to be suggesting that a junction diode is not meant to be placed in breakdown?

A: That’s exactly correct!

For many diode applications, breakdown is an annoying and useless operating mode that is not at all like an IDEAL diode.

For these applications, breakdown is to be avoided—we seek a junction diode with a very large Zener breakdown voltage $V_{ZK}$.

Thus, for these diode applications, the ideal value of $V_{ZK}$ is infinity (i.e., making breakdown impossible)!
Thankfully, a badly receding hairline is no longer required for a job in engineering

Q: Is that why our junction diode models did not consider breakdown?

A: That’s right—it is implicitly assumed in those models that $V_{z_k}$ is sufficiently large, such that breakdown cannot happen.

The forward and reverse bias regions only are modeled.

But—and here’s the exciting part—industrious and creative electrical engineers (as they are wont to do) discovered that the “useless” and “annoying” breakdown region actually has some especially useful applications!

So, there are some important differences between “normal” junction diodes, and the $p$-$n$ junction diodes that we call Zener diodes.
The first real difference

On the one hand, the Zener breakdown voltage $V_{ZK}$ of a junction diode:

* is typically large (e.g., 50 V or greater).

* The manufacturing tolerance associated with the breakdown voltage is also usually large (e.g., 20%).

* so that $V_{ZK}$ is often specified as an inequality (e.g., $V_{ZK} > 65 V$).

On the other hand, the Zener breakdown voltage $V_{ZK}$ of a Zener diode:

* varies over a wide range of values (e.g., 2 to 200 V), with some values quite small.

* The manufacturing tolerance associated with the breakdown voltage is also small (e.g., < 5%).

* so that $V_{ZK}$ is precisely specified as an equality (e.g., $V_{ZK} = 12 V$).
**The second real difference**

If a *junction* diode *does* happen to operate in the breakdown region, then it likely won’t likely to operate at all for very long!

Recall that a diode in the breakdown region will *absorb* (and so must dissipate) energy at a high rate.

* A “normal” junction diode is *not* designed to *thermodynamically* handle this event, as its designers assumed breakdown would *not* occur.

* The junction diode in breakdown will likely be *destroyed*!

* However, *Zener* diode manufacturers assume the *opposite*—that the diode will *operate* in the breakdown region.

* As a result, Zener diodes *are designed* to efficiently *dissipate* this heat—thermal equilibrium can be achieved *before* the Zener temperature reaches its *melting* point!
Zener Diode Notation

To distinguish a zener diode from conventional junction diodes, we use a modified diode symbol:

Generally speaking, a zener diode will be operating in either breakdown or reverse bias mode.

For both these two operating regions, the cathode voltage will be greater than the anode voltage, i.e.,:

\[ V_D < 0 \quad \text{(for r.b. and bd)} \]

Likewise, the diode current (although often tiny) will flow from cathode to anode for these two modes:

\[ I_D < 0 \quad \text{(for r.b. and bd)} \]
We're trying to avoid negative numbers

Q: Yikes! Won't the numerical values of both \( i_D \) and \( v_D \) be negative for a zener diode (assuming only rb and b.d. modes).

A: With the standard diode notation, this is true.

Thus, to avoid negative values in our circuit computations, we are going to change the definitions of diode current and voltage!
Stand on your head, and you will see that it is the same plot!

* In other words, for a Zener diode, we denote current flowing from cathode to anode as positive.

* Likewise, we denote diode voltage as the potential at the cathode with respect to the potential at the anode.

Note that each of the above two statements are precisely opposite to the “conventional” junction diode notation that we have used thus far:

\[ v_Z = -v_D \quad \text{and} \quad i_Z = -i_D \]

Two ways of expressing the same junction diode curve!
We’re just rotating the plot
The Zener diode curve—same as the old curve (stand on head)

The $i_Z$ versus $V_Z$ curve for a Zener diode is therefore:

- **Forward bias region**: $V_D$ to $-0.7V$
- **Reverse bias region**: $V_Z > V_{ZK}$
- **Breakdown region**: $V_Z = V_{ZK}$

Diagram:

- Horizontal axis: $V_Z$
- Vertical axis: $i_Z$
- Breakdown point $V_{ZK}$
- Forward bias region from $V_D$ to $-0.7V$
- Reverse bias region for $V_Z > V_{ZK}$
A mathematical statement of each region

Thus, in the forward bias region (as unlikely as this is):

\[ i_Z = -I_s \exp\left(\frac{-V_Z}{nV_T}\right) \]

or approximately:

\[ V_Z \approx -0.7 \text{ V and } i_Z < 0 \]

Likewise, in reverse bias region:

\[ i_Z \approx I_s \quad \text{and} \quad 0 < V_Z < V_{ZK} \]

And finally, for breakdown region:

\[ i_Z > 0 \quad \text{and} \quad V_Z \approx V_{ZK} \]
Zener Diode Models

The conventional diode models we studied earlier were based on junction diode behavior in the **forward** and **reverse** bias regions—they did not “match” the junction diode behavior in **breakdown**!

However, we assume that **Zener** diodes most often operate in **breakdown**—we need **new** diode models!
We need to match in the breakdown and reverse bias regions

Specifically, we need new Zener models, ones that match junction/Zener diode behavior in the reverse bias and breakdown regions.
In terms of Zener notation

Or, “rotating” into the Zener notation:

We will study two important zener diode models, each with familiar names!

1. The Constant Voltage Drop (CVD) Zener model

2. The Piece-Wise Linear (PWL) Zener model
Here’s an accurate model!

Let’s see, we know that a Zener Diode in the reverse bias region can be described as:

\[ i_Z \approx I_s \approx 0 \quad \text{if} \quad v_Z < V_{ZK} \]

Whereas a Zener in the breakdown region is approximately stated as:

\[ i_Z > 0 \quad \text{if} \quad v_Z \approx V_{ZK} \]

Thus, an accurate Zener model would have this “curve”:

\[ i_Z = 0 \quad \text{if} \quad v_Z < V_{ZK} \]

\[ v_Z = V_{ZK} \quad \text{if} \quad i_Z > 0 \]
Déjà vu all over again

**Q:** Hey, this is very similar to the CVD model we studied earlier:

\[ i_D = 0 \quad \text{if} \quad v_D < 0.7 \, V \]

\[ v_D = 0.7 \, V \quad \text{if} \quad i_D > 0 \]

We’ve just sort of changed 0.7 into \( V_{ZK} \)!

*Can we just change the 0.7 V battery in the CVD model into a voltage source of \( V_{ZK} \)?*
**So many lost points because of this**

**A:** That’s exactly correct!

Replace: \( v_Z \) with: \( v_D' \)

\[
Zi \\
\rightarrow Z
\]

\( + \)

\( \downarrow \)

\( i_Z \)

\( \rightarrow \)

\( \downarrow \)

\( i_Z \approx i_D' \)

**Q:** Hey wait!

*The cathode of the Zener diode is oriented upward (i.e., at the purple node) while the ideal diode cathode is oriented downward (i.e., the anode is at the purple node).*

*Is that correct?*

**A:** Absolutely! Make sure **you** get this right!
Ideal is reverse biased, Zener is reverse biased, but...

Analyzing this Zener CVD model, we find that if the model voltage $v_Z$ is less than $V_{ZK}$ (i.e., $v_Z < V_{ZK}$), then:

* the ideal diode will be in reverse bias,

* and thus the model current $i_Z$ will equal zero.

In other words:

\[ i_Z = 0 \quad \text{and} \quad v_Z < V_{ZK} \]

Just like a Zener diode in operating in the reverse bias region!
...if Ideal is *forward* biased, 

Zener is in *breakdown*!

Likewise, we find that if the model current is positive ($i_Z > 0$), then:

* the ideal diode must be *forward* biased, and
* the model voltage must be $v_Z = V_{ZK}$.

In other words:

$$i_Z > 0 \quad \text{and} \quad v_Z = V_{ZK}$$

Just like a Zener diode in operating in the *breakdown* region!

→ Carefully consider this case—if the IDEAL diode is *forward* biased, the model estimates the Zener diode voltage and current when operating in the *breakdown* region—this is *not* a contradiction!
The voltage is not quite constant

Q: So how accurate is the Zener CVD model?

A: Quite accurate! But it is not exact.

Here's the problem: the voltage across a zener diode in breakdown is not exactly equal to $V_{ZK}$ for all $i_z > 0$.

In reality, $v_z$ increases a very small (tiny) amount as $i_z$ increases.

Thus, the CVD model causes a small error, usually acceptable—but for some cases not!

For these cases, we require a better model:

$\rightarrow$ The Zener (PWL) Piece-Wise Linear model.
The Zener Piecewise Linear model

The PWL model includes a very small (e.g., 0.01 Ω) series resistor, such that the voltage across the model \( v_Z \) increases slightly with increasing \( i_Z \).

This small resistance \( R_Z \) is called the dynamic resistance of the Zener diode.

The voltage source \( V_{Z0} \) is not precisely equal to the Zener breakdown voltage \( V_{ZK} \), however, it is typically very close!
Smells like reverse bias region

Analyzing this Zener PWL model, we find that if the model voltage $v_Z$ is less than $V_{Z0}$ (i.e., $v_Z < V_{Z0}$), then:

* the ideal diode will be in reverse bias, and
* the model current $i_Z$ will equal zero.

In other words:

$$i_Z = 0 \quad \text{and} \quad v_Z < V_{Z0} \approx V_{DK}$$

Just like a Zener diode operating in the reverse bias region!
Likewise, we find that if the model current is positive \((i_Z > 0)\), then:

* the ideal diode must be **forward** biased,

* and thus the model **voltage** is:

\[
V_Z = V_{Z0} + i_Z r_Z
\]

Note that the model voltage \(V_Z\) will be near \(V_{ZK}\), but will increase **slightly** as the model current increases.

➔ Just like a Zener diode in **breakdown**!
Thank your middle school math teacher

Q: But how do we construct this PWL model; how do we determine the values of \( V_{z0} \) and \( R_z \)?

A: Typically, the manufacturer will provide two or more test points on the zener diode curve \((V_1, I_1)\) and \((V_2, I_2)\).

We then select \( R_z \) and \( V_{z0} \) so that the PWL model line intersects these two test points (just like the other PWL model!).

Specifically, the dynamic resistance is:

\[
R_z = \frac{V_2 - V_1}{I_2 - I_1} \quad \text{(a very small value!)}
\]

and

\[
V_{z0} = V_1 - I_1 R_z \quad \text{or} \quad V_{z0} = V_2 - I_2 R_z
\]
Example: Fun with Zener Diode Models

Consider this circuit, which includes a Zener diode:

Let’s see if we can determine the voltage across and current through the Zener diode!

First, we must replace the Zener diode with an appropriate model. Assuming that the Zener will either be in breakdown or reverse bias, a good choice would be the Zener CVD model.

Carefully replacing the Zener diode with this model, we find that we are left with an IDEAL diode circuit:
Since this is an IDEAL diode circuit, we know how to analyze it!

**Q:** But wait! The ideal diode in this circuit is part of a Zener diode model. Don’t we need to thus **modify** our ideal diode circuit analysis procedure in some way?

**In order to account for the Zener diode behavior, shouldn’t we **alter** what we assume, or what we enforce, or what we check?**

**A:** NO! There are no Zener diodes in the circuit above! We must analyze this ideal diode circuit in **precisely** the same way as we have always analyzed ideal diode circuits (i.e., section 3.1).
**ASSUME:** Ideal diode is forward biased.

**ENFORCE:** $v_D^i = 0$

**ANALYZE:**

From KVL:

$$17 - v_D^i - 20 - v_2 = 0$$

$$\therefore v_2 = 17 - 0 - 20 = -3.0\, V$$

Likewise from KVL:

$$17 - v_D^i - 20 + v_1 = 10$$

$$v_1 = 10 + 20 + 0 - 17 = 13.0\, V$$

Now from Ohm's Law:

$$i_1 = \frac{v_1}{R_1} = \frac{13}{4} = 3.25\, mA$$

and

$$i_2 = \frac{v_2}{R_2} = -\frac{3}{1} = -3.0\, mA$$
Finally, from \textbf{KCL}:

\[
\begin{align*}
    i'_D &= i'_2 - i'_1 - 5.0 \\
    &= -3.0 - 3.25 - 5.0 = -11.25 \text{ mA}
\end{align*}
\]

\[
\begin{align*}
    &\text{CHECK:} \quad i'_D = -11.25 \text{ mA} \quad \neq 0 \\
    &\text{Yikes! We must } \textbf{MODIFY} \text{ our ideal diode assumption and try again.}
\end{align*}
\]

\textbf{ASSUME}: Ideal diode is \textit{reverse} biased.

\textbf{ENFORCE}: \quad i'_D = 0
ANALYZE:

From KCL:

\[ i_2 = i_1 + 5 + i'_D = i_1 + 5 \]

Now from Ohm's Law:

\[ v_1 = i_1 R_1 = 4 i_1 \quad \text{and} \quad v_2 = i_2 R_2 = (i_1 + 5)(1) = i_1 + 5 \]

From KVL:

\[ 10.0 - v_1 - v_2 = 0 \]

\[ \therefore 10.0 - 4i_1 - (i_1 + 5) = 0 \]

Now solving for \( i_1 \):

\[ i_1 = \frac{10 - 5}{4 + 1} = 1.0 \, mA \]

Therefore:
\[ i_2 = i_1 + 5 = 1 + 5 = 6.0 \text{ mA} \]

And:

\[ \nu_1 = 4i_1 = 4(1) = 4.0 \text{ V} \quad \text{and} \quad \nu_2 = i_1 + 5 = 1 + 5 = 6.0 \text{ V} \]

Now, again using KVL:

\[ 17 - \nu'_D - 20 - \nu_2 = 0 \]

Therefore

\[ \nu'_D = 17 - 20 - \nu_2 \]

\[ = 17 - 20 - 6 \]

\[ = -9.0 \text{ V} \]
CHECK: \( v_D^i = -11.0 \, V \) \( < 0 \)

Q: Our assumption is good!

Since our analysis is complete, can we move on to something else?

A: Not so fast! Remember, we are attempting to find the voltage across, and current through, the Zener diode.

To (approximately) determine these values, we find the voltage across, and current through, the Zener diode model.

So,

\[
\begin{align*}
    v_Z & \approx v_D^i + V_{ZK} \\
    & = -11 + 20 \\
    & = 9.0 \, V
\end{align*}
\]

and

\[
    i_Z = i_D^i = 0
\]

We're done!

Q: Wait! Don't we have to somehow CHECK these values?
A: NO! We assumed nothing about the zener diode, we enforced nothing about the zener diode, and thus there is nothing to explicitly check in regards to the zener diode solutions.

However—like all engineering analysis—we should perform a “sanity check” to see if our answer makes physical sense.

So, let me ask you the question Q: Does this answer make physical sense?

A:
The Shunt Regulator

Say that we have some source voltage $V_s$, but our load requires a lower voltage $V_L$:

Q: Piece of cake! We just insert the proper series resistor $R_S$ to form a voltage divider, right?

A: This in fact is a very bad solution—we need a regulated voltage across the load.
He’s been like that since 1863

A regulated voltage is one that is constant; it never, ever changes—period.

For example, the voltage $V_L$ across the load must be independent of the load itself.

If $R_L$ increases or decreases, then the load current $I_L$ will decrease or increase—but the voltage $V_L$ should remain unchanged.

⇒ This is known as a load regulation.
Or maybe you start watching Netflix

Q: Why would the load current $I_L$ ever change?

A: You must realize that the load resistor $R_L$ simply models a more useful device.

The “load” may in fact be an amplifier, or a component of a cell phone, or a circuit board in a digital computer.

These are all dynamic devices, such that they may require more current at some times than at others (e.g., the computational load increases, or the cell phone begins to transmit).

Thus, modeling the load as a resistor $R_L$ is often not especially accurate; the load instead is defined by the current $I_L$ that it “draws”. 
The source voltage changes

Likewise, a regulated voltage must remain constant, even if the source voltage $V_s$ changes—this is known as line regulation.

Q: Why would the source voltage $V_s$ ever change?

A: There are many reasons why $V_s$ will not be a perfect constant with time.

For example, the energy for an electrical system, such as a “smart phone”, often is supplied by a chemical battery.

As this battery delivers energy, the voltage across it will slowly diminish—it “runs down”.

Yet, we need the regulated voltage to remain constant!
Likewise—as we shall learn later—if the source voltage $V_S$ is the output of an AC to DC power supply, it will likely include an AC component—a time-varying (not constant!) signal that we call the ripple voltage.
Yikes, the load voltage is unregulated

Let's return to original circuit:

From KVL, it circuit analysis, it is apparent that the difference between the source voltage $V_s$ and load voltage $V_L$ is simply the voltage drop across the series resistor $R_S$

$$V_L = V_s - I_L R_S$$

Clearly, the load voltage $V_L$ depends on both source voltage $V_s$ and load current $I_L$—the load voltage $V_L$ will change when $V_s$ and/or $I_L$ change (this is bad!).
Calculus: Is there anything it can’t do?

To quantify this, we simply take the derivative of $V_L$ with respect to $V_S$ and $I_L$:

$$\frac{\partial V_L}{\partial V_S} = \frac{\partial (V_S - I_L R_S)}{\partial V_S} = 1.0$$

$$\frac{\partial V_L}{\partial I_L} = \frac{\partial (V_S - I_L R_S)}{\partial I_L} = -R_S$$

The first of these derivatives quantifies the line regulation.

Since the value is 1.0, the load voltage changes directly with the source voltage (provided load current $I_L$ remains unchanged).
Line regulation: stinky

For example, if the source voltage increases by 100 mV, the load voltage will increase by 100 mV as well (again, if $I_L$ remains constant).

This is very poor line regulation!

Contrast this with the ideal case, wherein the derivative is equal to zero,

This means that a source voltage increase of 100 mV (or any value, for that matter) results in no change in the load voltage.
Load regulation: stinky

The second of the derivatives quantifies load regulation:

\[ \frac{\partial V_L}{\partial I_L} = \frac{\partial (V_S - I_L R_S)}{\partial I_L} = -R_S \]

The minus sign in this case means that an increasing load current \( I_L \) will result in a decreasing load voltage (and vice versa).

For example, if the series resistor \( R_S \) has a value of 500 Ohms \((R_S = 0.5\Omega)\), an increase in the load current of 0.2 mA will cause a 100mV reduction in the load voltage \( V_L \).

This is very poor load regulation!

Contrast this with the ideal case, wherein this derivative is equal to zero, meaning that a load current increase of 0.2 mA (or any value, for that matter) results in no change in the load voltage.
Zener diodes to the rescue!

Q: So what do we do? How do we achieve voltage regulation?

A: One solution is very simple—we just add a shunt Zener diode!

Q: Huh? What good would this do?

A: Let’s replace the Zener with the Zener CVD model and find out!
Be careful!

Now let's:

* ASSUME the ideal diode is forward biased,

* ENFORCE the assumption with a short,

* and then ANALYZE the resulting circuit.
From KVL, we arrive at an important result:

\[ V_L = V_{ZK} \]

\[ \Rightarrow \text{The load voltage is equal to the Zener breakdown voltage!!} \]

**Q:** Wow! Is this *always* true?

**A:** It is *if* the ideal diode forward bias ASSUMPTION is correct!
We need to find the ideal diode current

Q: Is the ASSUMPTION correct?

A: Let's determine the ideal diode current and find out!

First, from KCL:

\[ i_D' = I_S - I_L \]

And from Ohm's Law:

\[ I_S = \frac{V_R}{R_S} \]

Therefore:

\[ i_D' = I_S - I_L = \frac{V_R}{R_S} - I_L \]
We found the ideal diode current!

Finally, from KVL:

\[ V_S - V_R - V_{ZK} = 0 \quad \Rightarrow \quad V_R = V_S - V_{ZK} \]

Therefore, the ideal diode current is:

\[ i_D^i = \frac{V_R}{R_S} - I_L = \frac{V_S - V_{ZK}}{R_S} - I_L = \frac{V_S}{R_S} - \left( \frac{V_{ZK}}{R_S} + I_L \right) \]
Not if, but when!

Q: So how do we CHECK to see if this is positive?

A: Since there are no numeric values given to the variable names, we cannot in this case determine if the current is positive.

Rather, we can determine when the current is positive!

In other words, we can determine the conditions that will make this ideal diode current positive. I.E., if:

\[
\begin{align*}
\frac{i^i_D}{I_L} &= \frac{V_S}{R_S} - \left(\frac{V_{ZK}}{R_S} + I_L\right) > 0 \\
\end{align*}
\]

we find that:

\[
\begin{align*}
\frac{V_S}{R_S} > \frac{V_{ZK}}{R_S} + I_L \\
\Rightarrow V_S > V_{ZK} + R_S I_L
\end{align*}
\]
They better be true...

Thus, the source voltage $V_s$ must be at least $R_s I_L$ larger than the load voltage $V_L = V_{zk}$.

Or rearranging, we equivalently find that the load current $I_L$ has a maximum value—it must be less than the source current $I_s$:

$$I_L < \frac{V_s - V_{zk}}{R_s} = I_s$$

Q: I'm confused; why must these things be true?

A: They must be true in order for the ideal diode current in the Zener CVD model to be positive (i.e., $i_D^i > 0$)—they must be true in order for our ASSUMPTION to be correct.
...or else we have no regulation!

And as a result, they must be true in order for the load voltage $V_L$ to be equal to the Zener breakdown voltage:

$$V_L = V_{ZK} \quad \text{if} \quad V_S > V_{ZK} + R_S I_L \quad !!!!!!$$
Since 1863

Q: Why do you keep putting exclamation points (!!!!!) after $V_L = V_{ZK}$?

Why is this result such a "big deal"?

A: Because the Zener breakdown voltage $V_{ZK}$ is a device parameter—it is a constant.

⇒ It never, ever changes—period.

Thus, since the load voltage $V_L$ is equal to $V_{ZK}$, the load voltage is a constant.

⇒ It never, ever changes—period.
That’s what I call regulation!

The line regulation is thus:

\[ \frac{\partial V_L}{\partial V_S} = \frac{\partial V_{ZK}}{\partial V_S} = 0.0 \]

And the load regulation is:

\[ \frac{\partial V_L}{\partial I_L} = \frac{\partial V_{ZK}}{\partial I_L} = 0.0 \]

They’re both zero!

This means that the load voltage will not change—it will remain at \( V_{ZK} \)—even if the source voltage \( V_S \) and/or load current \( I_L \) change value (provided that \( V_S > V_{ZK} + R_S I_L \)).

➔ Now that’s voltage regulation!
The shunt regulator

This two-port circuit is known as the shunt regulator.

\[
V_s \quad I_s \quad R_s \quad I_l \quad V_l
\]
**A summary**

Summarizing, we find that if:

\[ V_s > V_{zk} + R_s I_L \]

then:

1. The Zener diode will be operating in the **breakdown** region (i.e., \( v_z = V_{zk} \) and \( i_z > 0 \)).

2. The **load voltage** is therefore equal to the Zener breakdown voltage (i.e., \( V_l = v_z = V_{zk} \)).

4. The **source current** \( I_s \) is \( I_s = (V_s - V_{zk})/R_s \).

5. The the Zener diode current is less than the source current (i.e., \( 0 < i_z < I_s \)).
Source current must be larger than the load current

In a shunt regulator, the source current \( I_s = \frac{(V_s - V_{zk})}{R_s} \) must be larger than the load current \( I_L \).

From KCL, we see that the extra source current—the current not “used” by the load—must pass through the Zener diode.

This causes the Zener to operate in the breakdown region.
The Zener takes the extra current...

Now, say the load changes, such that it draws more load current $I_L$ (e.g., the radio transmitter turns on).

Note though, that the source current $I_s$ does not change—if it did, we would see a different load voltage!

Instead, some of the “extra” current from the Zener diode now goes to load—the load current increases, but the source current remains the same ol’ value $I_s = \left(V_s - V_{zk}\right) / R_s$. 

\[
I_s = \frac{V_s - V_{zk}}{R_s}
\]
...but the Zener voltage remains unchanged!

Q: Hey wait!

If there is less current flowing through the Zener diode, won’t the voltage $v_Z = V_L$ drop as well?

A: Not really!

That’s the great thing about the Zener in breakdown—the voltage $V_{ZK}$ across the device is practically independent of the current through it.

⇒ This is why the load regulation is (practically) zero!
The load current can only go so high

Of course, this is true only if the Zener diode is operating in the breakdown region.

Note that as load current \( I_L \) increases to a value equal to the source current, the Zener current will approach zero:

Thus, the Zener will no longer be in breakdown (reverse bias instead!), and so the load voltage will no longer be equal to \( V_{ZK} \). I.E.,:

\[
\text{If } I_L \leq \frac{V_s - V_{ZK}}{R_S} \quad \text{then} \quad V_L \leq V_{ZK}
\]
It also provides line regulation!

Likewise, if the source voltage **increases**, then the **source current** will increase as well.

\[ I_s = \frac{(V_s - V_{Zk})}{R_s} \]

But, the **load current** does **not** increase, as this "extra" source current simply results in greater Zener diode current.

Remember, greater Zener diode current will **not** increase Zener diode voltage →

\[ v_z = V_{Zk} = V_L \]

\[ \Rightarrow \text{This is why the line regulation is (practically) zero!} \]
Example: The Shunt Regulator

Your boss rushes in and excitedly asks/demands that you design a voltage regulator for a new gizmo the company is about to manufacture.

The gizmo requires a voltage of precisely 6.0 Volts, and at a maximum will require power of 1.2 Watts.

The gizmo will require a standard battery(s) as a source (the kind you buy at Quick Trip), with as high a voltage as possible.

However, at least 60% of this battery power must be delivered to the gizmo.

The regulator must likewise be simple and cheap.

You decide to design a shunt regulator to meet these criteria!

This of course means that you must select the Zener diode, the series resistor, and the source (i.e., battery) voltage.

The Zener diode is pretty simple, its breakdown voltage must be equal to the widget voltage of 6.0 volts:
\( V_{ZK} = 6.0 \, V \)

Likewise, the efficiency of this regulator must be greater than 60\%, therefore:

\[
\eta = \frac{V_{ZK}}{V_s} > 0.6 \quad \Rightarrow \quad V_s < \frac{V_{ZK}}{0.6}
\]

Since the battery voltage should also be as high as possible, you decide to make it the largest value that still satisfies the efficiency requirement:

\[
V_s = \frac{V_{ZK}}{\eta} = \frac{6.0}{0.6} = 10.0 \, V
\]

Now, the maximum power that the gizmo will draw is 1.2 Watts, which corresponds to a maximum current of:

\[
I_L^{\text{max}} = \frac{P_L^{\text{max}}}{V_{ZK}} = \frac{1.2}{6.0} = 200 \, mA
\]

The source current should be made equal to this maximum load current value:

\[
I_s = \frac{V_s - V_{ZK}}{R_s} = I_L^{\text{max}}
\]

so that the series resistance should be:

\[
R_s = \frac{V_s - V_{ZK}}{I_L^{\text{max}}} = \frac{10 - 6}{200} = 0.02 \, \Omega = 20 \, \Omega
\]
Thus, your design is complete!

You eagerly present this design to your boss. Your heart thumps as you anticipate both praise and reward. Will you get a raise? a bonus? a promotion? stock options?

Instead, your boss looks at you with derision and disgust.

You palms begin to sweat as you consider what might have gone wrong. You checked all the math; KCL and KVL are verified.

Most of all:

You are sure that you used the “right equations”!

Finally, your boss breaks the awkward silence with this question:
“Where are we—or our customers for that matter—supposed to get a 10 volt battery?”

Yikes! Batteries—at least the kind you can buy at Quick Trip—are typically 1.5 Volts. The standard AAA, AA, B, C, and D-cell batteries are all 1.5 volts.

Of course we can create a higher voltage by “stacking” them in series, but the result would be an integer multiple of 1.5 V.

Of course 10 Volts (i.e., your embarrassing solution) is not an integer multiple of 1.5 Volts!

The closest multiples to 10 Volts are 10.5 Volts and 9.0 Volts.

The problem with the 10.5 Volt solution is that the regulator efficiency would be less than 60%. Thus, the 9.0 Volt solution appears to be the best—6 batteries connected in series.
Therefore, modifying your design for this 9.0 Volt source:

Your job is safe (at least for one more day!).

Be aware: circuit design is not a academic exercise; you must provide a design that is producible and maintainable—you must create a design with components and parts that actually exist!
Regulator Power and Efficiency

Consider now the shunt regulator in terms of power.

The source $V_s$ delivers energy at a rate $P_{in} \ (J/sec)$ to the regulator, and then the regulator in turn delivers energy at a rate $P_L \ (J/sec)$ to the load.

Q: So, is the power delivered by the source equal to the power absorbed by the load?
Power goes to the resistor and diode also

A: Not hardly!

The power delivered by the source is distributed to three devices—the load, the Zener diode, and the series resistor $R_S$.

We can determine the rate at which power is delivered and absorbed, in terms of the device parameters of this regulator circuit—we have four devices, we have four device parameters.
There are four circuit parameters

Specifically, those devices/parameters are:

1. **Source** voltage \( V_s \)
2. **Series** resistance \( R_s \)
3. **Zener** breakdown voltage \( V_{zk} \)
4. **Load** current \( I_L \)

First we determine the rate at which energy is absorbed by the load:

\[
P_L = V_L I_L = V_{zk} I_L
\]

Note, as load current \( I_L \) decreases, the load power likewise decreases.

As \( I_L \) approaches zero (the load is an open circuit), the load power also becomes zero.
The source delivers energy—whether the load uses it or not

Now, energy is delivered by the source at a rate:

\[ P_{in} = V_s I_s = V_s \left( \frac{V_s - V_{zk}}{R_s} \right) \]

Q: Wait! It appears that the input power is independent of the load current \( I_L \)!

Doesn’t that mean that \( P_{in} \) is independent of \( P_L \)?

A: That’s correct!

The energy flowing into the shunt regulator is independent of the rate at which energy is being delivered to the load.
Some goes to resistor, some to the diode

In fact, **even** if the load is drawing no current (i.e., $P_L=0$), the input power is **still** the same value shown above.

**Q:** But **where** does this input power go, if **not** delivered to the load?

**A:** The Zener diode also absorbs energy, at a rate:

$$P_Z = v_Z i_Z = V_{ZK} (I_S - I_L) = V_{ZK} I_S - V_{ZK} I_L = V_{ZK} \left(\frac{V_S - V_{ZK}}{R_S}\right) - V_{ZK} I_L = \frac{V_{ZK}}{R_S} (V_S - V_{ZK}) - V_{ZK} I_L$$

And, the series resistor absorbs energy at a rate:

$$P_R = I_S^2 R_S = \left(\frac{V_S - V_{ZK}}{R_S}\right)^2 R_S = \frac{(V_S - V_{ZK})^2}{R_S}$$
Energy comes from the source, but then is absorbed by the load, resistor and diode.

Remember, the source $V_s$ is an active device—it is a source of energy.

In contrast, the series resistor, Zener diode, and load are passive devices—they can only absorb the energy that the source provides.
Energy is conserved...

By conservation of energy, we conclude then that the power delivered by source must be equal to all absorbed power:

\[ P_{in} = P_R + P_Z + P_L \]

Now, using the results above, this absorbed power can be specified in terms of regulator circuit parameters:

\[
P_R + P_Z + P_L = \left(\frac{V_S - V_{ZK}}{R_S}\right)^2 + \left(\frac{V_{ZK}}{R_S}\right) (V_S - V_{ZK}) - V_{ZK} I_L + V_{ZK} I_L
\]

\[
= \left(\frac{V_S - V_{ZK}}{R_S}\right)^2 + \frac{V_{ZK}}{R_S} (V_S - V_{ZK})
\]

\[
= \frac{V_S - V_{ZK}}{R_S} (V_S - V_{ZK} + V_{ZK})
\]

\[
= \left(\frac{V_S - V_{ZK}}{R_S}\right) V_S
\]
...and it better be!

But of course, we found earlier that:

\[ P_{in} = V_s \left( \frac{V_s - V_{zk}}{R_s} \right) = P_R + P_Z + P_L \]

And so, we have verified that conservation of energy is indeed correct (I can’t begin to tell you what trouble we’d be in if we had determined otherwise!).

Q: So who cares? Why are you telling us this?

A: This analysis shows the relative inefficiency of the shunt regulator.
Let's say the source is powering your phone

Ideally, all the power delivered by the source would be absorbed by the load:

\[ P_L = P_{in} \quad \therefore P_{in} - P_L = 0 \]

But alas, we find for the shunt regulator:

\[ P_{in} - P_L = P_R + P_Z \quad \therefore P_L < P_{in} \]

Q: You’ve not answered my question—again, who cares?

A: Let’s say the source is in fact the battery to your phone—I bet then you would be the one who cares!
**Sarcasm alert!**

The energy *not* delivered to the load (your phone) is *wasted*—it simply *heats* up the series *resistor* and the *Zener diode*.

This *wasted* energy causes your *battery* to “run down” jiffy quick, and your phone soon *stops operating*!

*GrimStarable, Matt Grimmer*

As a result, you might *miss* one of those lucid, erudite, fascinating, and otherwise *life-altering* text/twitter messages that you apparently receive with *astonishing frequency*.

⇒ And wouldn’t that be a horrible tragedy!
Got milk? No! We wasted it all!

Thus, voltage regulators need to be efficient; the difference between $P_{in}$ and $P_L$ should be as small as possible.

Note that the ratio $P_L/P_{in}$ provides the percentage of the source power that is delivered to the load.

For example, if $P_L/P_{in} = 0.6$, then 60% of the source power is delivered to the load—the other 40% is wasted as heat in the regulator circuit.

For the shunt regulator, we find this ratio to be:

$$\frac{P_L}{P_{in}} = \frac{R_S}{V_S} \frac{1}{V_S - V_{ZK}} V_{ZK} I_L = \frac{V_{ZK}}{V_S} \left( \frac{R_S I_L}{V_S - V_{ZK}} \right)$$
Remember, there is a maximum $I_L$

Q: Huh? According to this equation, if the load current is really big, then then 100%—or even more—of the source power is delivered the load!? 

\[ \frac{P_L}{P_{in}} = \frac{V_Z}{V_S} \left( \frac{R_S I_L}{V_S - V_Z} \right) \]

A: But remember, the load current has a maximum possible value:

\[ I_L < I_S = \frac{V_S - V_Z}{R_S} \]

Thus, we can define a maximum load current for a shunt regulator—a upper limit on the load current:

\[ I_L^{\text{max}} = \frac{V_S - V_Z}{R_S} \]

And thus likewise an upper limit on load power:

\[ P_L^{\text{max}} = V_Z I_L^{\text{max}} \]
And bad things happen when unregulated

Q: Why is this a limit?

Why can’t the load current exceed this values?

A: Remember, if the load draws more current than this maximum value, the Zener diode will leave the breakdown region and instead will be reverse biased!

Q: ???

A: As a result, the voltage \( V_Z = V_L \) will be some value less than breakdown voltage \( V_{ZK} \)—the load voltage will be completely unregulated!
**Definition: Regulator efficiency**

The ratio of this **maximum load power** to the **source power** is a specific numerical value known as regulator efficiency $\eta$:

$$\eta = \frac{P_{L}^{\text{max}}}{P_{in}}$$

Obviously(?), regulator efficiency can theoretically be any value between **zero** (the worst) and **one** (the best):

$$0 \leq \eta \leq 1$$

Specifically for the **shunt** regulator, where:

$$I_{L}^{\text{max}} = \frac{V_{S} - V_{ZK}}{R_{S}}$$

we find that regulator efficiency is:

$$\eta = \frac{V_{ZK}}{V_{S}}$$
Efficiency is a problem if
\( V_S \) is way bigger than \( V_L \)

Recall that \( V_S > V_{ZK} > 0 \), so that as we would expect for efficiency:

\[
0 \leq \frac{V_{ZK}}{V_S} \leq 1
\]

Thus, is apparent that the closer the source voltage is to the load voltage \((V_L = V_{ZK})\), the higher the efficiency will be.

But be careful, this likewise will reduce the maximum load current:

\[
I_{L,\text{max}} = \frac{V_S - V_{ZK}}{R_S}
\]
Regulators are either linear or switching

For example, say the source voltage is $V_S = 10\, V$, and the load voltage is $V_L = V_{ZK} = 6.0\, V$.

The efficiency of this shunt regulator is:

$$\eta = \frac{V_{ZK}}{V_S} = \frac{6.0}{10.0} = 0.6$$

This shunt regulator is **60% efficient**!

Q: Doh! That means **40% of the source energy is wasted**; isn’t there some way to do better than that— I might miss an important tweet!

A: Voltage regulators essentially come in two types—the linear regulator and the switching regulator.

→ The shunt regulator is an example of a **linear regulator**.
Switching regulators: complex but efficient

Both the linear and switching regulator exhibit excellent line and load regulation.

But, the linear regulator is typically much less efficient than the switching regulator.

Q: So why don’t we just always use the switching regulator?

A: Switching regulators are typically more complex and costly than linear regulators, and they likewise generate Electromagnetic Interference (EMI).

Linear regulators are typically used if an engineer can stand the relative inefficiency—but often the engineer cannot!

http://electronic-circuit.net/acdc-switching-power-supply/
Linear Voltage Regulators

Note that we can view a shunt regulator as a three-terminal device, inserted between a voltage source and a load:

Integrated circuit technology has resulted in the creation of other three terminal linear voltage regulator designs:
Sometimes you’re just unlucky

These integrated circuit voltage regulators are small and relatively inexpensive.

In addition, these IC regulators typically have better load regulation, line regulation, and/or efficiency than the zener diode shunt regulator!

Q: Wow! The designers of these IC regulators obviously had a much better electronics professor than the dope we got stuck with!

A: Undoubtedly so!
An improved voltage regulator

However, these integrated circuit design engineers did not simply create a better shunt regulator.

Instead, they replaced the entire shunt regulator design with a complex circuit requiring many transistor components.
That op-amp requires many transistors

Integrated circuit technology then allows this complex circuit to be manufactured in a very small space and at very small cost!

The LM341 linear voltage regulator schematic.