

Introducing Computer Theory to Third Graders

Jesse Hamner, PhD

2017

Draft Copy: Do not redistribute or cite.

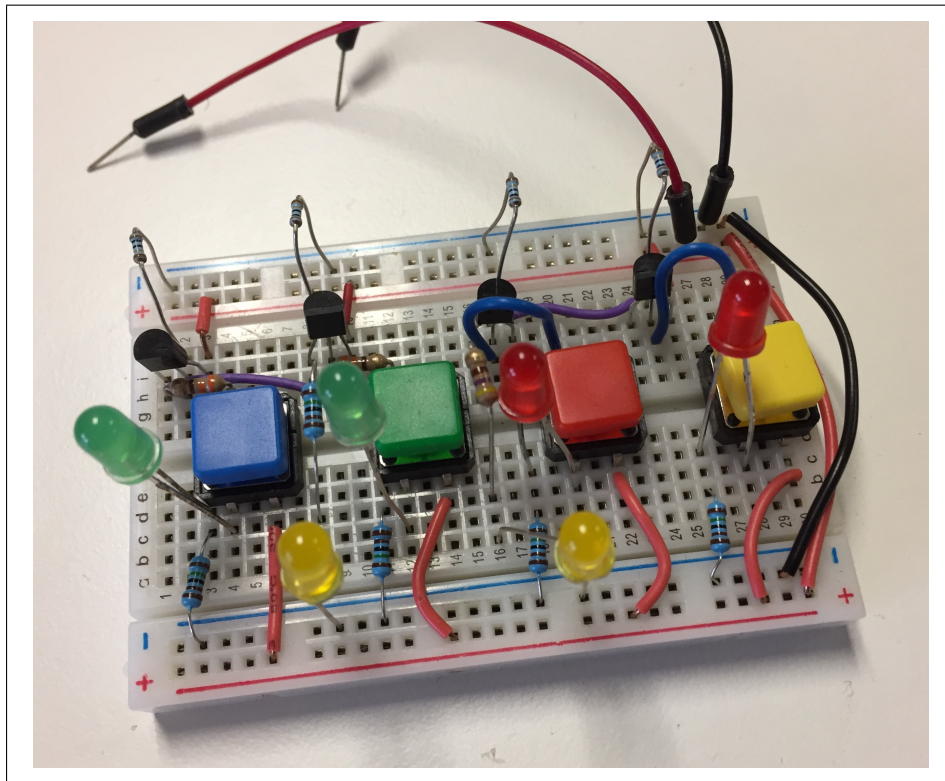
Abstract

Kids are using technology that is intuitive to control, but entirely opaque to the children's understanding. Why do computers work? How do computers count? What makes transistors so important? This workshop aims to produce a short course of study for third to sixth graders, that takes about three afternoons to complete. Children are introduced to math concepts fundamental to the study of computing such as binary counting, exponentiation, basic electronics theory, and computing logic. Several hands-on projects are included in the text and can enhance the students' learning while breaking up the "lecture" format of the workshop.

Keywords: STEM; education; computing theory; logic; electronics theory; elementary education; logic gates; science experiments.

Contents

1	Introduction	3
2	Decimal and Binary Math	4
3	Computer Math	13
4	Electricity	15
5	Basic Electronic Components	18
6	Logic Gates	23
7	Logic Circuits	34



Two logic gates (one OR, one AND) implemented with N-type MOSFET transistors on a solderless breadboard. Note pull-down resistors connecting transistor gate pins to ground.

1 Introduction

This workshop is going to help you learn about computers – but not how to *use* computers. Rather, when we are finished, you will understand a little bit about how and why computers really work. To get to that goal, you’ll need to learn some new topics in math, and learn about electricity. You’ll also need to learn about basic *electronic components*, too. These are the small pieces that enable electricity to do all the things we can make electricity do, like light our homes, play music on the radio, find directions with a GPS, run a gasoline engine, or show you a web page.

Computers are wonderful, multipurpose machines. But computers can’t add or subtract the way you do. They can’t read the way you do. They don’t think. Really, *they only do math*. Everything computers do is controlled by a series of instructions that *all boil down to math and logic*, even if it looks like magic. The way computers do all the things they do can always be described as a series of instructions. Each of these actions can be performed by some combination of special circuit types, and each circuit is made up of basic electronic components. You can learn enough about each basic component to understand what it does, without needing to do a bunch of math. And, you can understand all the individual instructions, even if you don’t want to build your own computer!

Ultimately this workshop will make it possible for us to make, from the most basic electronic components, a simple digital computer that can add two numbers and correctly display the result. Along the way you’ll see how each piece works, and how to put the pieces together.

2 Decimal and Binary Math

The next few sections introduce some math ideas that may or may not be familiar to you. In order to really understand how computers work, you need to understand several math concepts that are not hard, but may be unfamiliar. The first concept is the *exponent*.

Bases and Exponents

Exponentiation is a mathematical operation, written as b^n , involving two numbers, the base, b , and the exponent, n . Exponentiation means “repeated multiplication of the base.”

That is, b^n is the product of multiplying n bases:

$$b^n = \underbrace{b \times \cdots \times b}_n$$

In this case, b^n is called the “ n -th power of b ”, or “ b raised to the power n .”

Put another way: the exponent is a number, smaller and on the upper right hand side of a number, that means “multiplying a number times itself zero times, once, or more than once” depending on whether the number is 0, 1, 2, or another number (written n in the description above). It is possible to use fractions as exponents, but we won’t talk about that here.

The base can be any number, though most people think most easily in base-10. So in base-10, 10 to the power 2, or 10 to the second power, means 10×10 .

Order of Magnitude


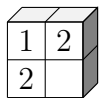
When counting up from zero, in base 10, you eventually get to 9. In order to count any higher, you must “carry the one” over to the next column and reset the counter in the “ones” column (the place that counts by one, from zero to nine). When you move from the “ones” column to the “tens” column, you move to the next *order of magnitude* of the base. “Magnitude” means “size”, and in math, it means, specifically, moving from counting by one number (the base) to counting by the base times itself – first twice, then three times, then more. For base-10, you count from 0 to 9 in the right-most column, then from 10 to 90 in the next column to the left, then from 100 to 900 in the next column to the left, then from 1000 to 9000, and so on. Each time the counter gets full (reaches 9), you cannot represent any more of the quantity being counted without moving to the next largest order of magnitude. That is, if your display only shows two digits, once you count past 99, you have no idea if the number is 0, or 100, or 200, or 900, or ten thousand, or 4 billion.

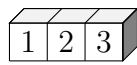
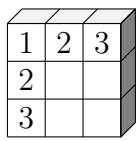
Squares

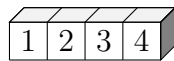
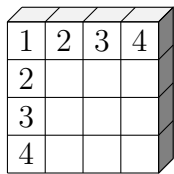
Squares are the same on each of two sides. *Squaring a number* is multiplying the number times itself, just like, in a square, both sides are the same length (number). The most common way you will see a “squared number” described is with a little 2 up above the number, like 2^2 , or 3^2 , or 10^2 . That smaller ‘2’ is the exponent.

Squaring the number 10 gets: $10 \times 10 = 100$

 $1 \times 1 =$  $= 1^2$

 $2 \times 2 =$  $= 2^2$

 $3 \times 3 =$  $= 3^2$

 $4 \times 4 =$  $= 4^2$

Cubes

Cubes are the same length on each of *three* sides. When cubing a number, you are multiplying the number times itself, and then multiplying it times itself *again*, because all three sides are the same (number). The most common way you will see a “cubed number” described is with a little 3 up above the number, like 2^3 , or 3^3 , or 10^3 . That little number is called the “exponent.” The exponent tells you how many times you multiply the number times itself.

Cubing the number 10 gets: $10 \times 10 \times 10 = 1000$



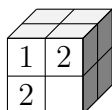
$$1 \times 1 \times 1 =$$



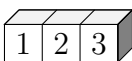
One times itself is just one.



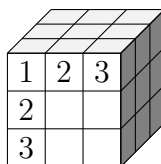
$$2 \times 2 \times 2 =$$



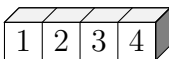
Two sets of four, or $4 + 4$
(that is, $2^2 + 2^2$)



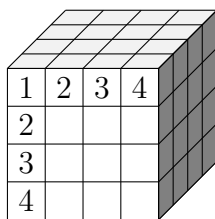
$$3 \times 3 \times 3 =$$



Three sets of nine, or $9 + 9 + 9$
Put another way: $9 \times 3 = 27$



$$4 \times 4 \times 4 =$$



Cubes get big quickly! This
is four sets of 16. $16 + 16 +$
 $16 + 16 = 64$

Higher Order Exponents

You can multiply a number times itself as many times as you want. Understanding a little more about exponents (the number of times you multiply a number times itself) will make understanding the language we use to discuss computers quite a bit easier. Let's do some “powers of two”.

Powers of Two		Notes
2^0	= 1	This one is strange, sort of. But it works! Any number “to the zeroth power” is equal to 1.
2^1	= 2	
2^2	= 4	
2^3	= 8	
2^4	= 16	Since $2^2 = 4$, $2^2 \times 2^2 = 16$
2^5	= 32	This is why you can count to 31 on one hand.
2^6	= 64	
2^7	= 128	
2^8	= 256	8-bit computers were the first really big deals for consumers.
2^9	= 512	
2^{10}	= 1,024	1,000 usually gets the prefix <i>kilo-</i> , like a kilogram is 1000 grams. A <i>kilobyte</i> is 1024 bytes.
2^{16}	= 65,536	
2^{20}	= 1,048,576	2^{10} bytes is a kilobyte; $(2^{10} \times 2^{10})$ bytes is a <i>megabyte</i> .
2^{24}	= 16,777,216	Most computer displays can show up to 16 million colors, using red, green, and blue, all in combination. Each piece of the color can have 256 (2^8) levels, from zero (black) to 255 (100% red, or green, or blue)
2^{30}	= 1,073,741,824	2^{30} bytes is a <i>gibibyte</i> , or a bit more than a <i>gigabyte</i> , which is 1000 megabytes.
2^{32}	= 4,294,967,296	
2^{40}	= 1,099,511,627,776	2^{40} bytes is a <i>tebibyte</i> , or a few percent more than 1000 gigabytes – a <i>terabyte</i> .
2^{50}	= 1,125,899,906,842,624	2^{50} bytes is <i>pebibyte</i> . A petabyte is so large that one petabyte is enough to store the DNA of the entire population of the USA...and then clone them, <i>twice</i> .

Explanation: The reason that any number to the zeroth power is equal to one comes from the way we subtract exponents when dividing. You know that 8 divided by 4 equals 2; written another way, $2^3 \div 2^2 = 2^1$. Notice that the exponents change by subtraction, but the equation is the same! You can *divide* base-exponent numbers by *subtracting* the exponents (*extra-special historical trivia: this is how slide rules work*). And since any number divided by itself equals one, as in $2^3 \div 2^3 = 1$, subtracting the exponents gives 2^0 .

Representing Numbers in Decimal

Decimal means “with tens”. You’ve always been taught to count in the decimal system – by ten. When you are counting by ones, once you get past 9, you reset the right-most number to zero and add that ten to the next column, which is the *tens* column (so you only increment that counter by one, since you are adding *one* “ten” to the total). Once you fill up 9 “tens” (90) and count up past 9 in the “ones” column (that is, you add one to 99), you have to set the ones column and the tens column to zero, and add one to the “hundreds” column. You *carry* the ten to the left from the ones, and carry the hundred to the left, to the next-largest set of tens. One hundred is ten tens, one thousand is 100 tens, and so forth.

It looks like this:

ten thousands	thousands	hundreds	tens	ones	Number
0	0	0	0	0	0
0	0	0	0	1	1
0	0	0	0	2	2
0	0	0	0	3	3
0	0	0	0	4	4
0	0	0	0	5	5
0	0	0	0	6	6
0	0	0	0	7	7
0	0	0	0	8	8
0	0	0	0	9	9
0	0	0	1	0	10

When you get to the bottom of the column, you re-set the counter for that column to zero, and add one to the next column over (carry). So you go from 9 to 10, or 19 to 20, or 99 to 100. So any column can only hold between 0 and 9, before you have to carry over to the next *order of magnitude*, which for a decimal system is *ten times the size of one step in the column*. So when you run out of room for the ones column, you go to the *tens* column, which contains ten times as many units as any single step (from 1 to 2, or from 8 to 9) in the column to the right of it. When you run out of room in the column, you have to carry to the next order of magnitude. You go from the ones, to the tens, to the hundreds (100 is 10×10), to the thousands (1000 is 10×100), and so on.

Here are a few examples of numbers expressed in columns. For each one, you count up how many units in the column exist, then count them for each order of magnitude, and add each value together to get the number being described:

Text Description	ten thousands thousands hundreds tens ones	Number
no ten-thousands, no thousands, no hundreds, no tens, and no ones	0 0 0 0 0	0
no ten-thousands, no thousands, no hundreds, no tens, and one ones	0 0 0 0 1	1
no ten-thousands, no thousands, no hundreds, no tens, and two ones	0 0 0 0 2	2
no ten-thousands, no thousands, no hundreds, three tens, and no ones	0 0 0 3 0	30
no ten-thousands, no thousands, four hundreds, no tens, and no ones	0 0 4 0 0	400
no ten-thousands, no thousands, five hundreds, five tens, and five ones	0 0 5 5 5	555
no ten-thousands, six thousands, five hundreds, no tens, and two ones	0 6 5 0 2	6502
six ten-thousands, eight thousands, no hundreds, three tens, and no ones	6 8 0 3 0	68030

Exercise: What if, in the above table, you counted past 99,999? What number would you see? What happened to the hundred thousand that should be counted? Do you need to know how many hundred thousands are in the number? How would you handle the need for a larger number, if you have it? (Introduces the concept of *overflow*.)

Representing Numbers in Binary

Computers only understand “1” and “0” – because it can sense the electricity in a wire that is either *on* (having a detectable voltage greater than zero) or *off* (having a reference voltage that is basically at ground potential, usually “zero volts”). That means that computers, in order to add, subtract, or store information, has to express *literally everything it can handle* in terms of either ones or zeroes. This system is called *binary*, because computers only understand two “states” – on, or off. Instead of “base-10” counting (where moving to the next column happens when you pass 9), binary is “base-2” counting; the numbers move to the next column when the number passes 1.

In order to add numbers in binary, we can’t count to ten. We must count to one, and then, if the resulting number is greater than one, carry the digit to the next column. Counting in binary is interesting because it is different, and because it introduces us to a new way of doing math.

In order to add two numbers, computers have to do the following:

1. Store each number in a binary format
2. compare each column (ones, twos, fours, eights, etc.) of each number and see if the numbers in that column add up to more than one
3. carry the “one” to the next column (the next *order of magnitude*, which is two times the previous column). That is, if the number goes from one to two, the “one” will go to zero and the “two” will be moved – and added – to the next column over to the left.
4. keep on adding and carrying digits until the addition is complete.
5. count up the values of each order of magnitude (either one, or none) and add them together.
6. report the new number as a [binary] result

Description	eights	fours	twos	ones	base 10
$0 + 0 + 0 + 0$	0	0	0	0	0
no eights, no fours, no twos, one one	0	0	0	1	1
no eights, no fours, one two, no ones	0	0	1	0	2
no eights, no fours, one two, one one	0	0	1	1	3
$0 + 4 + 0 + 0$	0	1	0	0	4
no eights, one four, no twos, one one	0	1	0	1	5
$0 + 4 + 2 + 0$	0	1	1	0	6
no eights, one four, one two, one one	0	1	1	1	7
$8 + 0 + 0 + 0$	1	0	0	0	8
$8 + 0 + 0 + 1$	1	0	0	1	9
$8 + 0 + 2 + 0$	1	0	1	0	10
$8 + 0 + 2 + 1$	1	0	1	1	11
$8 + 4 + 0 + 0$	1	1	0	0	12
$8 + 4 + 0 + 1$	1	1	0	1	13
$8 + 4 + 2 + 0$	1	1	1	0	14
$8 + 4 + 2 + 1$	1	1	1	1	15

Problem 1: what happens in the above table if you add 1 to 15? What number would the computer report?

Problem 2: Let's say the counter runs by itself at some regular speed, like one number (operation) per second. Let's also say you are using the counter as a way to control a single blinking light, since "1" means "there is electricity available to that wire" and so a "1" would turn the light on. Remembering that "a line that has a voltage" is a 1 and "no voltage" is a zero, which line (column) would make the light blink fastest? Which line would blink the slowest?

How Computers Count to 65,535

2^{15}	2^{14}	2^{13}	2^{12}	2^{11}	2^{10}	2^9	2^8	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	zero
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	32,767
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	32,768
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	65,535

If the computer counts up from zero, then once it gets past the 2^{14} column, it jumps to the 2^{15} (32,768) column. If we had told the computer that it was using *signed* numbers, it would count upwards from -32,768, and the first column would be a one, to indicate it

was negative, with the rest of them zeros. That is, the format for storing negative numbers subtracts 32,768 from 0 – the range is still the same (65,536 numbers), but the starting point is different.

Useless skill: Did you know you can count to 31 on one hand? You have five fingers, and each finger can be open or closed, and 2^5 is 32. Use your thumb for the “0 or 1” (2^0 , “two to the zeroth power”) column; the digit to its left represents two to the first power (the “twos digit”); the next digit to the left represents two to the second power (the “fours digit”); and so on.

128	64	32	sixteen	eight	four	two	one	
2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0	Result
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	2 ($2 + 0$)
0	0	0	0	0	0	1	1	3 ($2 + 1$)
0	0	0	0	0	1	0	0	4 ($4 + 0 + 0$)
0	0	0	0	0	1	1	1	7 ($4 + 2 + 1$)
0	0	0	0	1	0	0	0	8 ($8 + 0 + 0 + 0$)
0	0	0	0	0	0	1	0	_____
0	0	0	0	0	1	0	1	_____
0	0	0	0	0	1	1	1	_____
0	0	0	0	1	0	1	0	_____
0	0	0	0	1	1	1	1	_____
0	0	0	1	0	0	0	0	_____
0	0	0	1	1	1	1	1	_____
0	0	1	0	0	0	0	0	_____

Tip: if you have a sequence of all ones, like 7 or 15 or 31, rather than adding up each binary column, you can just subtract one from the next largest column base number. So 0111 equals 1000 minus one.

Joke:

Remember: There are only 10 types of people in the world; those who understand binary, and those who don't.

3 Computer Math

In this section, I introduce the vocabulary terms that help you understand how computers do certain things, and why they matter. For instance, how do computers represent numbers below zero? How do they add?

Sign

Representing *positive* numbers is easy for a computer: it counts upwards from zero. Representing *negative* numbers is harder, because *taking away values by adding them* is a little awkward. To represent a negative, designers use the leading (left-most) bit to declare “positive” (zero) or “negative” (one). Note that the *range* of numbers that can be represented does not change – for a 4-digit binary number, whether counting from zero to 15, or from -8 to 7, the total space used on the number line is still 16 numbers, in order.

Problem: If the computer doesn’t know it is supposed to use that first number to determine whether a number is negative or not, how does it evaluate the number?

Complement

In mathematics, the *complement* of a binary number is

the value obtained by inverting all the bits in the binary representation of the number (swapping 0s for 1s and vice versa). The ones’ complement of the number then behaves like the negative of the original number in some arithmetic operations.*

Here’s an example:

Number	+	–
0	0000	1111
1	0001	1110
2	0010	1101
3	0011	1100
4	0100	1011
5	0101	1010
6	0110	1001
7	0111	1000

*[Wikipedia page on Ones’ Complement](#).

Addition

Adding two values is straightforward. Simply align the values on the least significant bit and add, propagating any carry to the bit one position left. If the carry extends past the end of the word it is said to have “wrapped around”, a condition called an **“end-around carry”**. When this occurs, the bit must be added back in at the right-most bit.

Binary:	Decimal:
0001 0110	22
+ 0000 0011	3
=====	=====
0001 1001	25

Your turn! Add these numbers in binary (then convert them to decimal to check your work)

Binary:	Decimal:
0000 0010	
+ 0000 0011	
=====	=====
-----	-----

Binary:	Decimal:
0000 1010	
+ 0001 1011	
=====	=====
-----	-----

4 Electricity

Electricity is the presence and flow of some amount of electric charge. It is the flow of *electrons* through conductors such as copper wires.

Electricity is a form of energy that comes in positive and negative forms, that occur naturally (as in lightning), or is produced by people (as happens in a generator). It is a form of energy we use to power machines and electrical devices. When the charges are not moving, electricity is called *static electricity* (does that sound familiar? Where else have you heard that term?). Static electricity means a charge is present, but has nowhere to go. When the charges are moving they are an *electric current*, sometimes (but rarely) called 'dynamic electricity'. Lightning is a highly dangerous kind of electricity in nature. Static electricity can cause things to stick together. Electricity can be dangerous, especially around water. Note that *you* are made up of a lot of water.

The experiments we will perform don't use dangerous levels of electricity, so they are safe to handle. You can lick a 9V battery, which feels weird but is not dangerous.

Experiment: make a lemon battery to power an LED with large-gauge copper wire, and large galvanized screws. At least four or six lemon batteries will be required to power even a modest LED. Wire the batteries in series with alligator clips.

Charge

Electric charge is a basic property of electrons and protons. Electrons are negatively charged and protons are positively charged. Things that are negatively charged and things that are positively charged pull on (attract) each other. This makes electrons and protons stick together to form atoms. Things that have the same charge push each other away (they repel each other). This behavior is called the *Law of Charges*. Things that have equal numbers of electrons and protons are neutral. Things that have more electrons than protons are negatively charged, while things with fewer electrons than protons are positively charged.

Voltage

Voltage is a force that makes electricity move through a wire. It is measured in volts. Voltage is also called *electric tension* or *electromotive force* (EMF). It was named after Alessandro Volta.

Though it's not helpful at first, the best definition of voltage is *the difference in electric potential between two points*. Voltage is always measured between two points, for example between the positive and negative ends of a battery, or between a wire and ground.

The most common analogy for electricity is water in pipes. If we think of water in pipes, voltage is the *water pressure* pushing on the water to get it to move through the pipes. Voltage is proportional, meaning that 3 Volts pushes twice as hard on electrons (to get them to move) as 1.5 Volts pushes.

Static shocks, like the kind you get on metal surfaces in the winter when wearing a wool sweater, can be up to 30,000 volts – but it’s not dangerous to you because the amount of electrical charge is so small. The voltage has a very high *potential*, but cannot do much work.

Experiment: Push 9V battery terminals into your arm. Then lick the 9V battery terminals. Why did you feel the electricity on your tongue and not your arm?

Current

An electric current is a flow of electric charge. The SI unit of electric current is the *ampere* (A), almost always shortened to “amp”. This is equal to one coulomb of charge in one second. A coulomb is an exact count of electron charges (6,241,509,629,152,650,000 elementary charges – that’s over six quintillion electrons!). Electrons are very small, and contain only a very, very tiny amount of charge. An amp of current can be a lot or a little, depending on how much voltage is pushing the current. 5 amps at 2,000 volts would kill you. 1 amp at 5 volts can’t even shock you. Smaller currents than one amp are usually measured in *milliamps*, or one-thousandths of an amp. So 500 milliamps is half an amp. 100 milliamps is a tenth of an amp. Milliamps are usually abbreviated “mA”.

Electricity can be dangerous. Never work with electricity unless an adult has shown you how to be safe, and told you the work you are doing is safe.

A little saying to help remember how electricity can be dangerous is:
“volts jolt; mils [milliamps] kill.”

(That is, even high voltages aren’t necessarily dangerous, but even low currents can be deadly.)

Resistance

Resistance is how much a material or component prevents the flow of electricity. Substances with very, very high resistances are *insulators* – they prevent the flow entirely under most conditions. Substances that pass electricity at least some are called *conductors*. Copper, silver, and gold are excellent conductors. Aluminum, zinc, and iron are good conductors too, but not all metals are good conductors. Titanium is a poor conductor, but is not an insulator.

Many resistors are made with some amount of carbon inside them – carbon conducts electricity but not very well. The more carbon between the wires, the higher the resistance.

Modern resistors don't necessarily use carbon, but it is still used. While a resistive material does pass electricity, it also slows down the flow, and prevents some electricity from passing. This electricity that gets "used up" by the resistor is turned into heat.

Experiment: A salt-water resistor. Attach wires to an LED (without the resistor this time, which is normally a bad idea). Hook the positive end to a 3 to 5 volt supply, leaving the negative terminal open. Drop one wire from the negative terminal of the LED into a glass jar half-full of (distilled, if you have it) water. Tape the wire to the lip of the jar, and make sure the wire is submerged. Put the wire from the negative terminal in the jar, under water, but keep the wires pretty far apart. Does the LED come on? Would you expect it to? Now add some salt to the water. What happened? (May need to add more salt).

Ohm's Law

Electricity and Electronics can be understood using many equations, but we'll only look at one of them. Ohm's law says that the voltage, current, and resistance in a circuit are all related, in a way that can be calculated. Ohm's law states that the current through a conductor (wire) between two points is directly proportional to the voltage across the two points.

The law was named after the German physicist Georg Ohm, who, in a treatise published in 1827, described measurements of applied voltage and current through simple electrical circuits containing various lengths of wire.

If we abbreviate "voltage" to V , "resistance" to R , and use I for "current" (the scientist who first understood current pretty well was French, and he referred to current as *intensité de courant*, meaning "current intensity", so we use "I" for current).

$$V = I \times R \tag{1}$$

That is, voltage is equal to however much current is flowing, times how much total resistance is in the circuit. Because the individual components are proportional, the law can be written other ways:

$$I = \frac{V}{R} \tag{2}$$

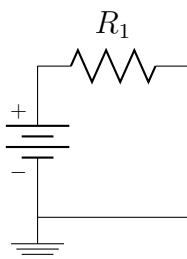
One very practical application of Ohm's law is that it allows an engineer to limit how much current can flow in a circuit by adding a certain level of resistance. Often it would be safer to keep a lot of current from flowing, so even a little extra resistance can make the circuit (and the components in the circuit) safer and more durable. Resistors consume some energy, and turn electricity into heat.

5 Basic Electronic Components

The following sections briefly introduce electronic components. Each of these pieces can be used to make complex, powerful circuits.[†]

Resistors

A resistor limits the electrical current that flows through a circuit. Resistance is the restriction of current. In a resistor, the energy of the electrons that pass through the resistor are changed to heat and/or light. For example, in a light bulb there is a resistor made of tungsten which converts the electrons into light and a great deal of heat.



A very simple circuit. Electricity flows through the resistor (maybe it's a light bulb). The resistor controls how much current flows in the circuit, and heats up.

Capacitors

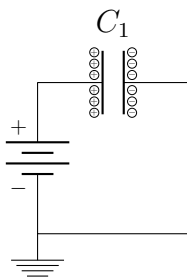
A capacitor (also called condenser, which is the older term) is an electronic device that stores electric energy (a “charge”). It is similar to a battery, but is smaller, lightweight and charges up much quicker. Typically, capacitors hold much less energy than a battery, but can provide almost their entire charge to the circuit quickly. Capacitors are used in many electronic devices today, and can be made out of many different types of material.

The Leyden jar was one of the first capacitors invented: metal foil was placed inside a glass jar, and wrapped around the outside of a glass jar, but neither foil gets near the top of the jar. The glass barrier between the foil sheets (inside and outside) allows a charge to build up between the foil without allowing the electrical charge to move through the glass. Leyden jars didn't hold much charge, but the concept it demonstrated has not changed.

Capacitors are usually made with two metal plates that are on top of each other (or wrapped around each other), but that do not actually touch. When powered, they allow energy to be stored inside an electrical field. Because the plates need a lot of area to store even a small amount of charge, the plates are usually rolled up into some other shape, such as a cylinder. Sometimes, other shapes of capacitors are used for special purposes. A capacitor-like effect can also result just from two conductors being close to each other, whether you want it to

[†]Most of these definitions come from Wikipedia, though they have been edited and augmented by the author in several places.

exist or not— if you’ve ever been shocked by a door knob or metal refrigerator in the winter, you have been one plate of a capacitor, because you’ve been storing a charge!



Open capacitors in a steady state. No current flows, but there is a charge.

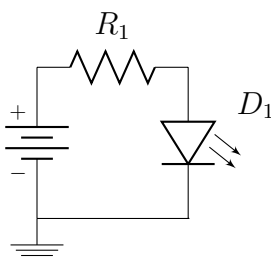
Experiment: make a variable capacitor from aluminum foil, paper, and a paper towel tube. The paper should go around the roll only once. Cut the foil one inch smaller than the paper on all sides. Glue or solder a small wire to one corner of the foil. Tape the foil to the paper. Tape one edge of the paper, foil side down, against the cardboard tube. Make another paper/foil combination. Wrap the paper/foil around the cardboard tube, but tape the paper only to itself, and just loose enough to slide a little.

It is also possible to make a regular capacitor with the paper-foil layers, separated by flat sheets of cardboard. Keep track of “up” and “down” sides! Note that it is pretty easy to gang each “plate” together and make capacitor with a larger value; tying all the anodes together, and all the cathodes together, makes a larger capacitor.

Diodes

A diode is an electronic component with two electrodes (connectors). It allows electricity to go through it only in one direction.

Diodes can be used to convert alternating current to direct current (this circuit is called a diode bridge). They are often used in power supplies and can be used to decode AM (“amplitude modulation”) radio signals (like in a crystal radio). Light-emitting diodes (LEDs) are a type of diode that produce light.



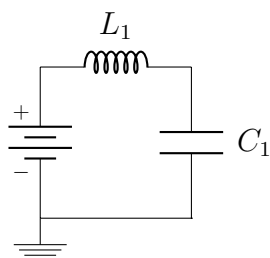
A very simple circuit. Electricity flows through the resistor, that controls how much current flows into the LED. If the battery was put in backwards, no electricity could flow in the opposite direction, so the LED would not light up.

Inductors (Coils)

An inductor, also called a coil or (rarely) a reactor, is a two-terminal electrical component that stores electrical energy in a magnetic field, when electric current is flowing through it. An inductor typically consists of an electric conductor, such as a wire, that is wound into a *coil*. Sometimes inductors include a magnetic bar or ring, around which the wire is wound. Other times the wire is wound with only air in the middle (which still works, because the Earth has a magnetic field).

Inductors do many interesting things. For instance, they are used to block AC while allowing DC to pass; inductors designed for this purpose are called *chokes*. They are also used in electronic filters to separate signals of different frequencies, and in combination with capacitors to make tuned circuits, used to tune radio and TV receivers.

The symbol for an inductor is a coil icon. Inductors are usually abbreviated with an L:



A simple circuit with one coil and one capacitor. This sort of circuit is one way to make an *oscillator*. This kind of circuit is found in radios.

Transistors

A transistor is a *semiconductor* (see below) device used to amplify or switch electronic signals and electrical power. It is composed of semiconductor material usually with at least three terminals for connection to an external circuit. A voltage or current applied to one pair of the transistor's terminals controls the current through another pair of terminals. Because the controlled (output) power can be higher than the controlling (input) power, a transistor can amplify a signal. Today, some transistors are packaged individually, but many more are found embedded in integrated circuits.

The transistor is the fundamental building block of modern electronic devices. It might be the most important – or at least the most influential – engineering discovery, ever.

Why are transistors important? They enable a lot of powerful new devices.

- They can act like switches – with no moving parts, and that can be controlled with a tiny electrical signal
- They can take a small signal and turn it into a powerful signal; they can make sound louder, broadcast radio signals farther, or detect very tiny changes in other materials (thus, you can have electronic thermometers, or radio telescopes)
- When combined into logic gates (more on that below), they enable super-fast computations that have also changed how we do work, science, medicine, and almost every form of research

A Little History

Many years ago, a “computer” meant “the person who did the math problems.” A few people invented adding machines like an abacus, and *slide rules* made even difficult calculations very possible, but they still required a person to do the math. In the same way, people had learned about electricity, but couldn’t do much with it except power light bulbs, street cars, and telegraphs. Talking to someone many miles away meant either using Morse code, or writing a letter.

The first discovery that led to the transistor was the *vacuum tube diode*, which uses a hot wire in a glass enclosure with no air in it to push electrons to another conductor, just like any other diode. Unfortunately, vacuum tube diodes are big, fragile, hard to make, and produce a huge amount of heat. The *vacuum tube triode*, which does everything a transistor does, came next. Since electrons are negatively charged, a slight negative charge on the third wire (between, but not touching, the hot-wire emitter and the collector) permitted people to control the flow of electrons from one side to the other – it acted like a switch.

Vacuum tube triodes have all the same shortcomings as the vacuum diode, and have even more parts to fail. Despite their shortcomings, vacuum tubes changed the world in a *huge* way: they enabled radios, televisions, radar, and the first digital computers. But these computers were the size of a small house, weighed 30 tons, and required powerful cooling.

In the 1940s, three scientists discovered that some substances could either behave like an insulator *or* a conductor, depending on how much electricity or heat was applied to it. Being able to change a slab of glass or silicon from an insulator into a conductor meant they could do everything a vacuum tube could do, except much smaller, cheaper, and cooler. About nine years after they announced the discovery and proved that it works, the three men won the Nobel Prize. They had discovered a *semiconductor*.

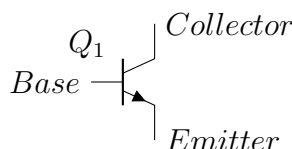
There’s a longer documentary about the transistor on YouTube [here](#).

The world changed amazingly quickly when we started using transistors. Today, there are many different kinds of transistors. The computer in your house has many *billions* of transistors inside it. So does your phone, your car, your microwave, your television...wow!

Some transistors are made especially for certain applications, like making audio louder, or pushing radio signals, or just for doing computations. These are all transistors, but they are different.

Electronics and Transistors

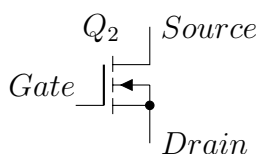
Transistors are usually abbreviated with Q . There are several kinds of transistors, but they work mostly the same.



Most transistors have three pins. For this kind of transistor (called an “NPN”):

- Electricity comes in the *collector*
- Electricity applied to the *base* opens up the flow of electricity to the *emitter*
- Electricity, when able, flows out the emitter and towards a ground.
- Therefore, the base is like a switch. It turns the transistor from an insulator into a conductive wire.

Most computers use a transistor type called a “FET” (*Field Effect Transistor*), that is very efficient and does not require very much electricity to switch from “on” to “off”. They are faster and run cooler than the older types of transistors. Somewhat confusingly, these transistors have different names for their pins, compared to the older “BJT” type transistors, seen above.



For questions of computers, we will treat transistors like a switch and nothing more. That’s easy, right?

But how do transistors do all the cool stuff they do? Why are pocket calculators or iPhones possible? How do collections of electronic components enable us to do math, send text messages, or simulate a nuclear explosion without anything going kaboom?

At the absolute base of the entire electronics world, just above transistors, are *logic gates*. They are the simplest building blocks of digital computers.

A good additional introduction video is [here](https://www.ted.com/talks/ed_tanaka_the_magic_of_the_transistor) (ed.ted.com).

6 Logic Gates

The basic building blocks of computers are called *gates*. Their only function is to report, with some yes-or-no electrical output, the result of their circuit following some input from somewhere else in the circuit. The way to understand how gates make decisions is called “Boolean Algebra”, but really it’s a fairly simple set of decisions that get strung together in useful ways.

Boolean Logic

Boolean logic is the sort of decision making and answers that computers use. You can use it too, of course, and it will help you learn to program computers. The answer to every question in Boolean logic is either “yes” or “no” – “true” or “false”. In computers, that’s a one or a zero, of course. Every operation (decision) is composed of a set of operations: *and*, *or*, and *not*. A British man named George Boole introduced these ideas to the world in 1847[‡] He did not realize he was laying the foundations for computing, but 50 years later another very important mathematician, Claude Shannon, understood that Boole’s system for understanding decisions could be used to design electronic circuits.

It may be helpful to see some of the logic and operations (decisions) made in a graphical form, looking at the overlap between circles, or between a circle and its surrounding area.

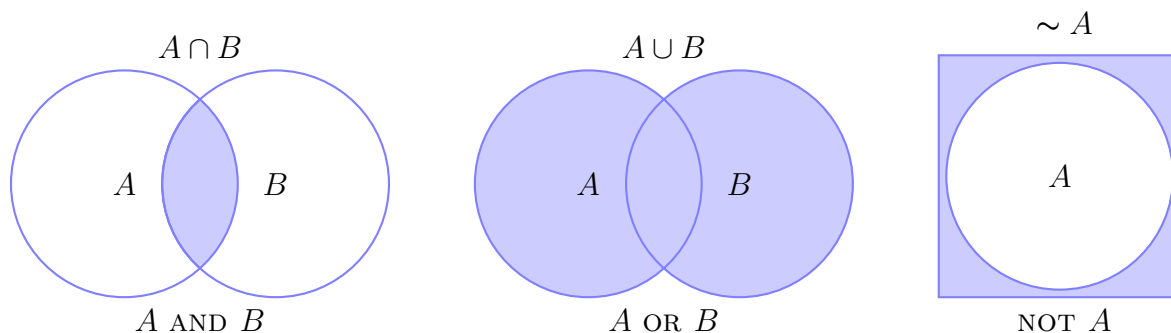


Figure 1: A graphic representing AND, OR, and NOT with overlapping circles.

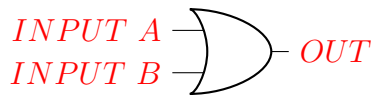
Also note: more complex decisions can be made by combining these operations: NOT + AND, for instance).

The OR Gate

OR gates take two inputs and provide a “yes” if line 1 OR line 2 are equal to 1 – that is, if either line has a voltage level above zero.

[‡] *The Mathematical Analysis of Logic* (1847), and *An Investigation of the Laws of Thought* (1854). For more, see his [Wikipedia entry](#).

The symbol for an OR gate is:



The “Truth Table” (how they behave) is:

OR Gate Truth Table		
A	B	OUT
0	0	0
1	0	1
0	1	1
1	1	1

OR gates are very simple. They are so simple that transistors are not even needed, though they certainly can be used.

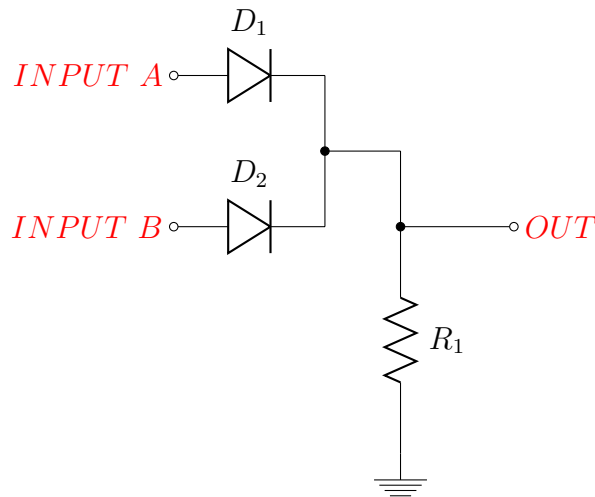


Figure 2: A very simple OR gate schematic. Diodes conduct electricity only in the “forward” direction. So Input A cannot affect Input B. But either way, if there is a signal on Input A or on Input B, OUT will carry a voltage (and therefore, a “1”).

Experiment: Wire up an OR gate on a breadboard, like the one seen in the cover image. Use LEDs as indicators for Input A, Input B, and OUT. Feel free to use diodes, but using transistors is pretty easy. It’s best to use N-type transistors as switches for each LED and for the OR-output LED.

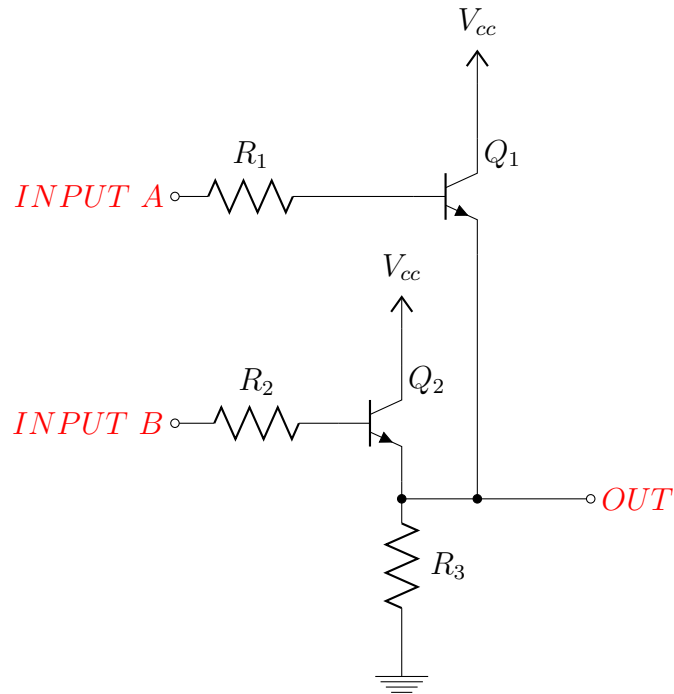


Figure 3: A simple transistor-based OR gate. Transistors are better than diodes in this application because transistors amplify (or at least do not degrade) signals, while diodes lose half a volt or more. Diode *propagation delay* is also slower than transistors, making the circuits without these diodes run faster. In this schematic, each transistor acts as an independent switch for the output – either switch pushes OUT to high.

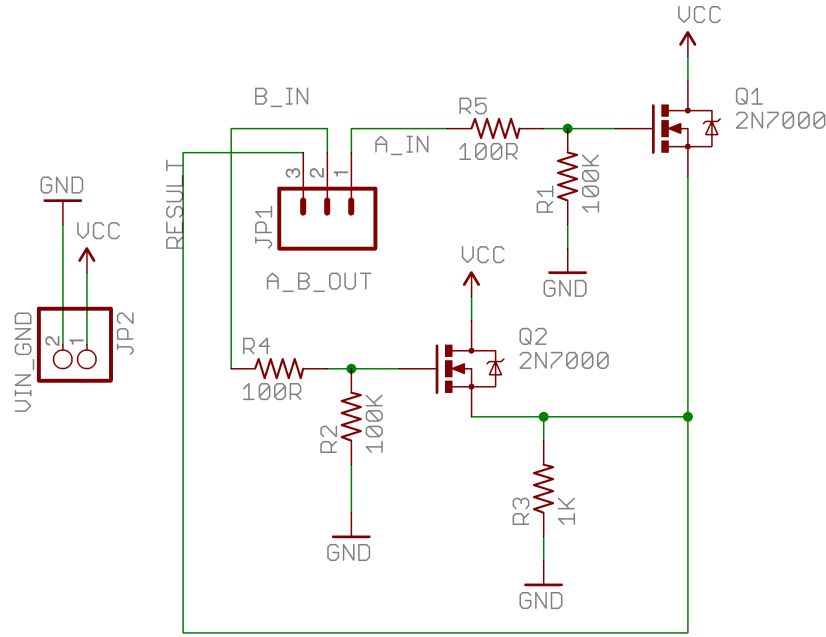


Figure 4: A schematic from a real circuit design program showing a practical, real implementation of the idealized schematic in Figure 3. The wires connecting components are the green lines. Mostly there are more resistors to help keep the transistors behaving well. For instance, the 100K resistors make sure the transistors are always at a low logic level (zero volts) when the inputs are ‘off’.

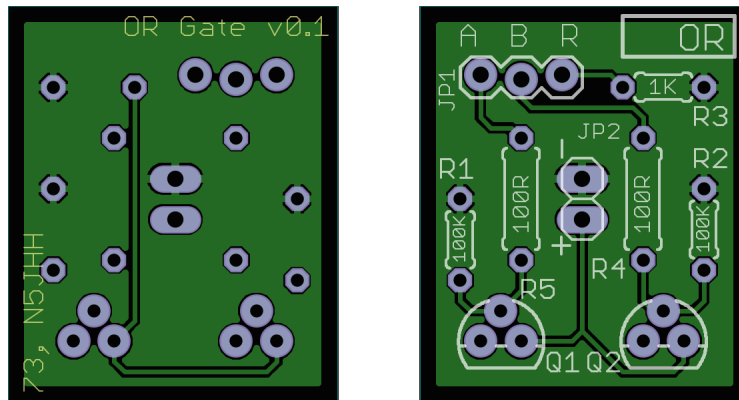
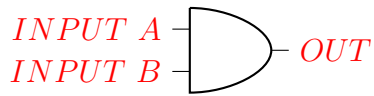


Figure 5: A circuit board designed using the above OR gate schematic. You can solder these without much trouble. You can see the top and bottom of the board here, but all components go on the top. However, both top and bottom have *traces*, that carry electricity instead of using loose/floppy wires. The larger resistors are the lowest value (they’re just there to make the transistors behave a little better): 100R means “100 Ohms”. The smaller resistors are either 100,000 Ohms (that is, 100K Ohms) or 1,000 Ohms (1K Ohms). The OR gate takes two *inputs*, A and B, and produces one *result*, R.

The AND Gate

AND gates take two inputs and provide a “yes” (a one, or a “logic high” signal) only if both line 1 AND line 2 are equal to 1.

The symbol for an AND gate is:



The “Truth Table” (how they behave) is:

AND Gate Truth Table		
A	B	OUT
0	0	0
1	0	0
0	1	0
1	1	1

There are much more complicated AND gates, but this is an easy-to-understand example (and it totally works):

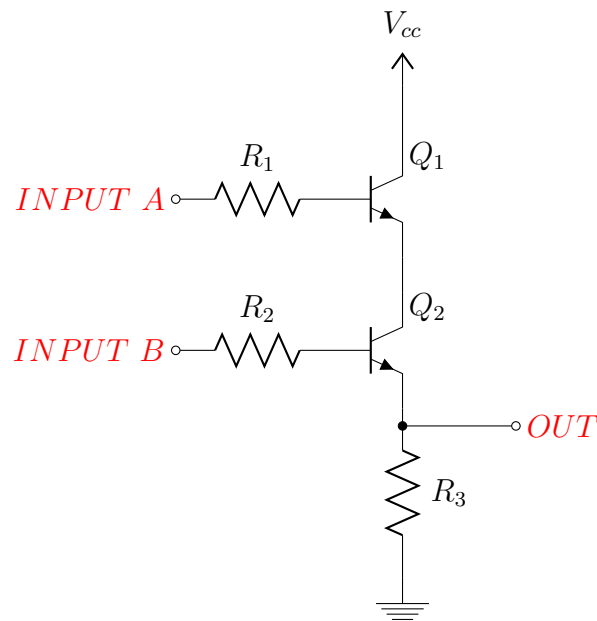


Figure 6: A simple AND gate schematic. When Input A is a zero, no current flows into the “collector” (voltage input wire) of Transistor 2. When Input B is zero, no current can flow to OUT, regardless of whether Input A is 0 or 1. Only when Input A is 1 (current can flow past Q_1) and when Input B is 1 (current is available to Q_2 , and Q_2 is turned “on”, allowing current to pass) does OUT go “high”.

Experiment: Wire up an AND gate on a breadboard. Use LEDs as indicators for Input A, Input B, and OUT.

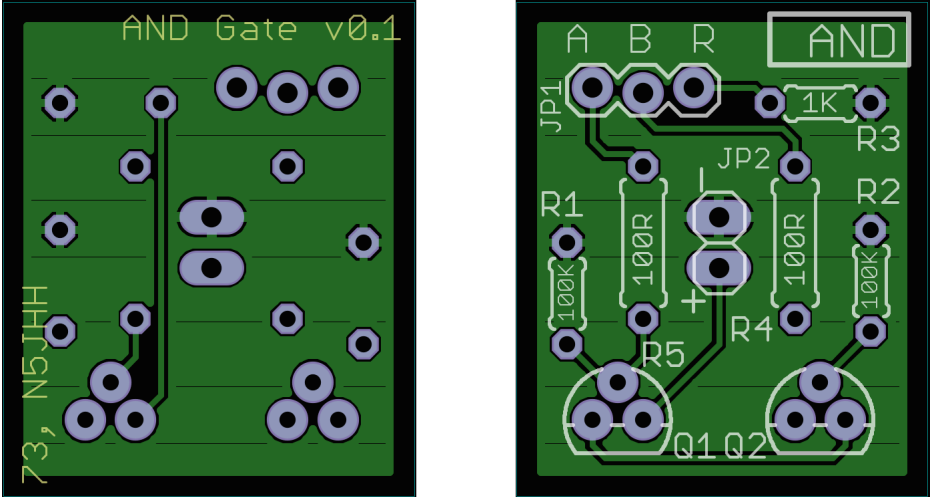
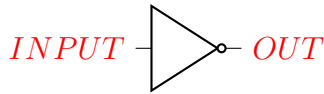


Figure 7: A circuit board designed using the above AND gate schematic. This one looks almost exactly like the OR gate, but the wiring is different! Also, you can see both the front and back sides of the circuit board in this image. The AND gate takes two *inputs*, A and B, and produces one *result*, R.

The NOT Gate

NOT gates take one input and reverse the value of that input. So a 1 becomes a 0, or vice versa. Also note this is the *complement* of the input value, as seen above.

The symbol for a NOT gate is:



The “Truth Table” (how they behave) is:

NOT Gate Truth Table	
IN	OUT
0	1
1	0

A simple NOT circuit is easy to understand, though this example is not how NOT gates are actually implemented, because of the poor efficiency (and wasted energy turns into heat, which then has to be cooled somehow; it is better to not make the heat in the first place).

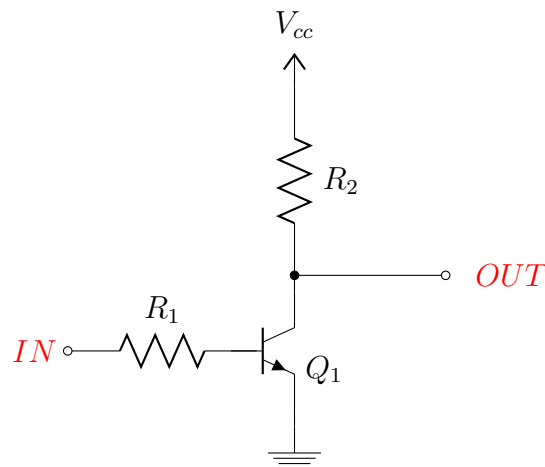


Figure 8: A very simple NOT gate schematic. When IN is logic low (off), electricity cannot pass through to the ground, and so it is available at the OUT terminal. That means OUT is a 1. When IN is carrying current, the transistor opens up and electricity starts flowing to ground – the “path of least resistance”. Then, because R_3 provides a little resistance, electricity flows to ground instead of OUT, meaning OUT has a very, very low voltage available – making it a zero.

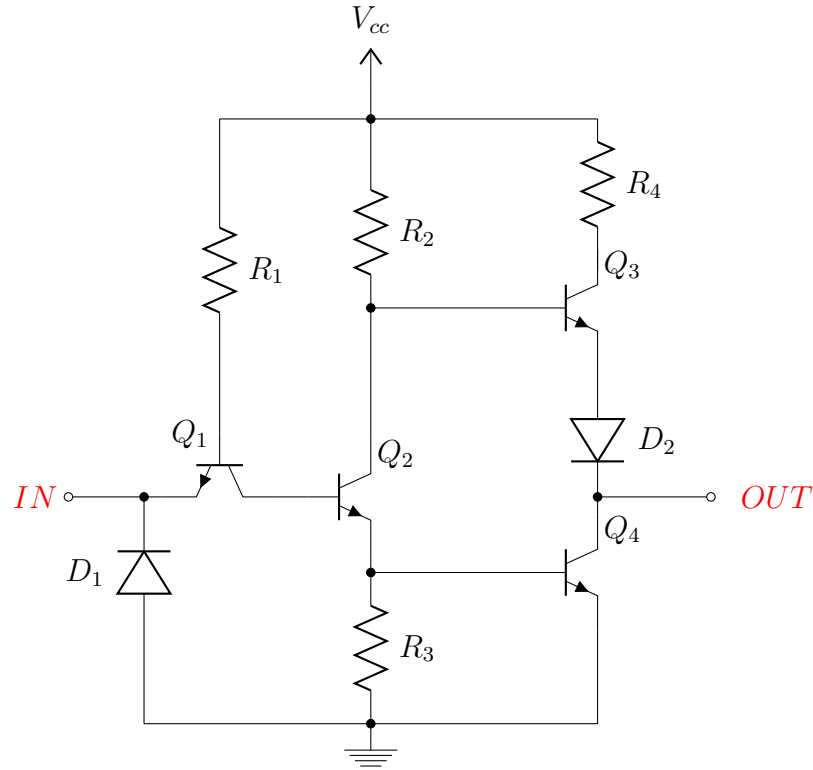


Figure 9: A practical inverter (NOT) circuit. While more complicated than a simple NOT circuit, it is more efficient, wasting much less energy.

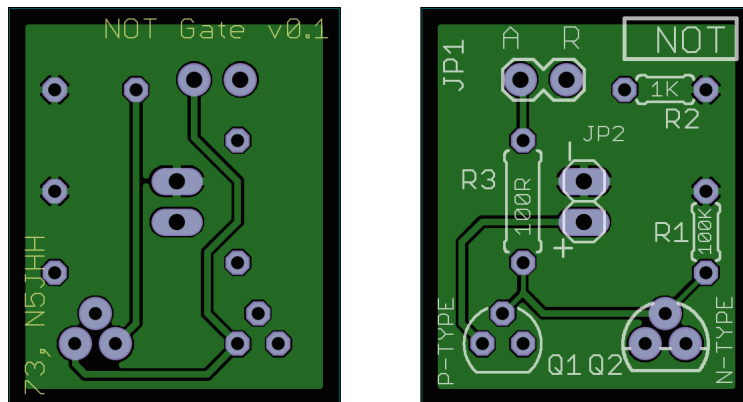


Figure 10: A circuit board designed using a more elegant NOT gate schematic. This one looks almost exactly like the OR gate and the AND gate, but the wiring is different! Also, it has only one input and one output. The NOT gate takes one *input*, A, and produces one *result*, R.

Complex Gates

Combining gates is possible to produce more complex circuits that answer different questions. A “NOT-AND”, or “NAND”, gate just reverses the value obtained from an AND gate – so it has the exact opposite truth table. Since a NAND gate just reverses the output of AND, you can put a NOT gate on the result (output) line of an AND gate and make a NAND gate, just like you were using Legos.

The symbol for a NAND gate is:



(see the circle on the output?
That's the *not* part)

The “Truth Table” (how they behave) is:

NAND Gate Truth Table		
A	B	OUT
0	0	1
1	0	1
0	1	1
1	1	0

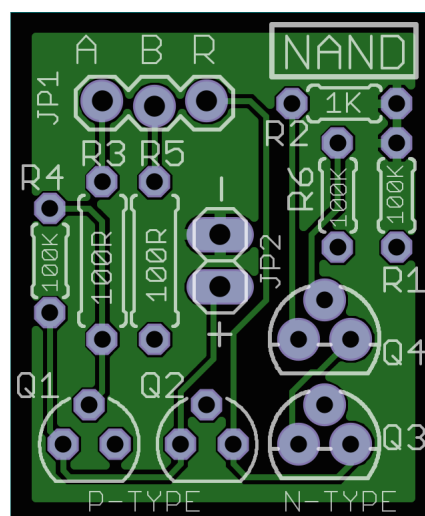
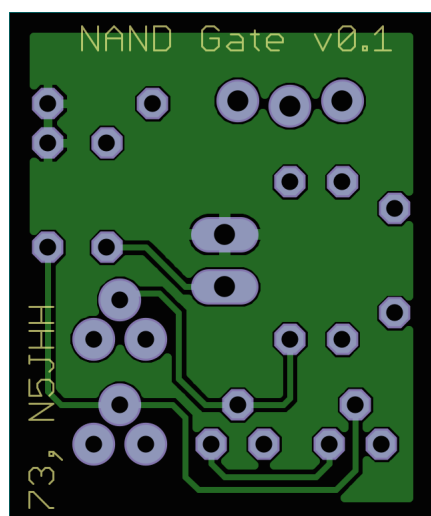


Figure 11: A circuit board designed to make a NAND gate. Note that it uses four transistors.

Not surprisingly, there is also a NOR gate, that has the reverse truth table to an OR gate. NOR gates are easy to implement, though as we have said elsewhere, more complicated NOR gates save power.

The symbol for a NOR gate is:



(see the circle on the output?
That's the *not* part)

The “Truth Table” (how they behave) is:

NOR Gate Truth Table		
A	B	OUT
0	0	1
1	0	0
0	1	0
1	1	0

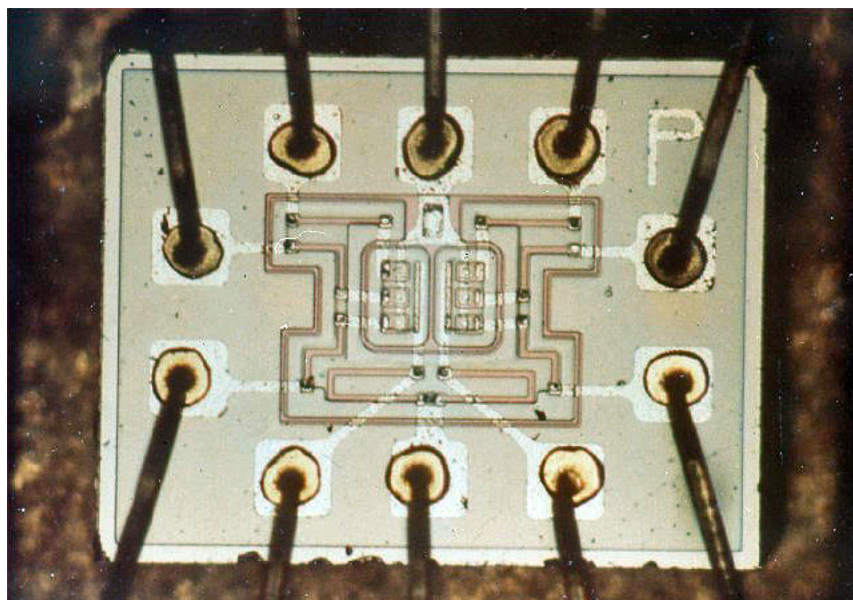


Figure 12: An early integrated circuit NOR gate. It has six transistors, eight resistors, and three pairs of A/B inputs (thus, it has three separate yes/no outputs). This chip was part of the guidance computer that took Americans to the moon in 1969. *Photo credit: Wikimedia Foundation.*

There is also the “eXclusive-OR”, or “XOR”, gate. The XOR gate output (result) is 1 when *the inputs are not the same*. It returns a 1 if, and *only* if, both inputs are different. That is, it provides a one “exclusively if” there is a single “one” (sometimes called “logic high”, for “voltage above zero”) in the two inputs. XOR gates take a lot of transistors to implement. The most common version requires 12, though it is possible to make an XOR gate with 8 transistors.

The symbol for an XOR gate is:



(the curved line behind the OR gate means that it is “exclusive”)

The “Truth Table” (how they behave) is:

XOR Gate Truth Table		
A	B	OUT
0	0	0
1	0	1
0	1	1
1	1	0

7 Logic Circuits

A Half Adder

A *half adder* demonstrates the way in which computer logic gates can correctly add (binary) numbers, though here it can add only two *binary digits* (“bits”). Thus, each half adder can only count to *two*, but that is enough! Long cascades of half adders and *full adders* enable computers to count large numbers. Below, see two examples of half adders. The second example replaces the XOR gate with a series of simpler gates, but the results are the same. Each half adder takes two bits and adds them, passing on the bits for further work or as an answer itself.

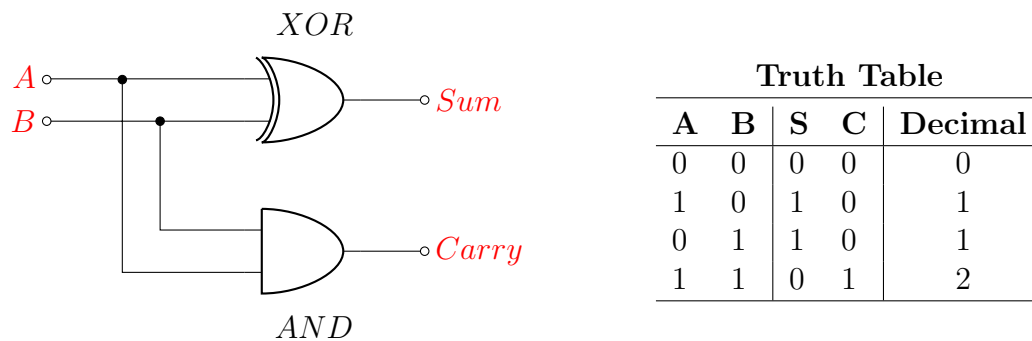


Figure 13: A half-adder circuit. It adds two bits together. If the result is larger than 1, the “carry” signal is high.

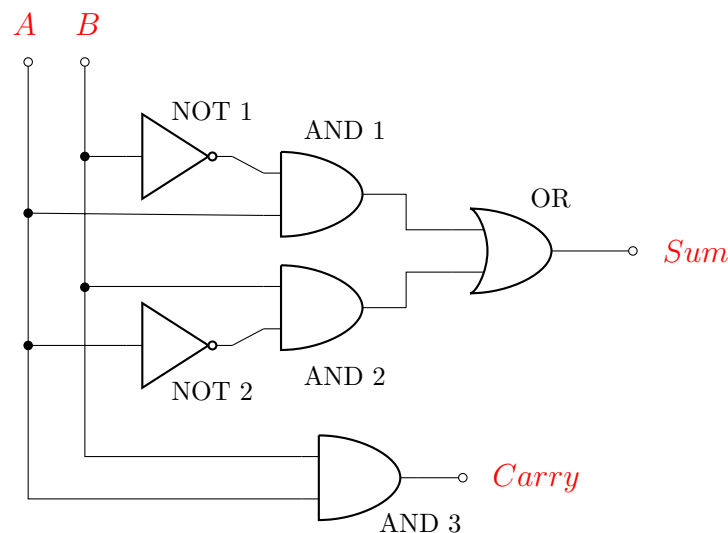


Figure 14: Another half-adder circuit. The XOR gate has been replaced by an equivalent circuit made up of simpler gates.

Full Adder

This type of adder is a little more difficult to implement than a half-adder. The main difference between a half-adder and a full-adder is that the full-adder has *three* inputs and two outputs. The first two inputs are A and B, just like with every half-adder, and the third input is an input carry, designated as C_{in} . Enabling the circuit to handle a carry *in* from another adder makes it possible to add many binary orders of magnitude, since many carry-outs and carry-ins will be needed to add numbers larger than 1. Once we implement a full adder, we can string eight of them together to create a byte-wide adder and cascade the carry bit from one adder to the next.

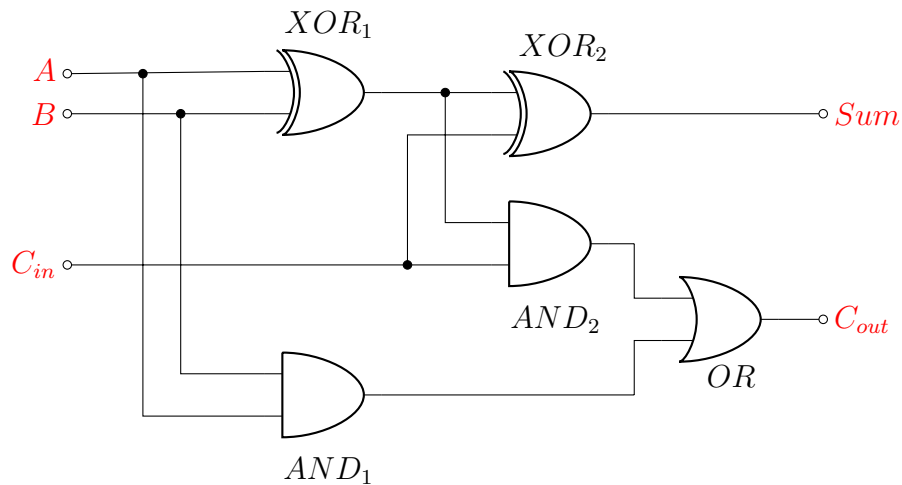


Figure 15: A full-adder circuit, made up of two half-adder circuits. A full-adder enables carry-in as well as carry-out. These units can be linked together to make binary addition work even for very large numbers.

Equality

Comparing two one-bit numbers to see if they are identical is done with XNOR gates. These gates are the same as an XOR gate passing output into a NOT gate.

The symbol for an XNOR gate is:



The “Truth Table” (how they behave) is:

XNOR Gate Truth Table		
A	B	OUT
0	0	1
1	0	0
0	1	0
1	1	1

To compare large numbers, the circuit requires an XNOR gate for each binary digit (bit) in the numbers being compared. If any XNOR gate returns zero, the numbers are not equivalent. To build all of the XNOR gates into a single yes or no answer, each XNOR output could go into a series of AND gates – that is, since AND returns a “yes” only if both inputs are one, a single “no” prevents the AND gate from returning “yes”.

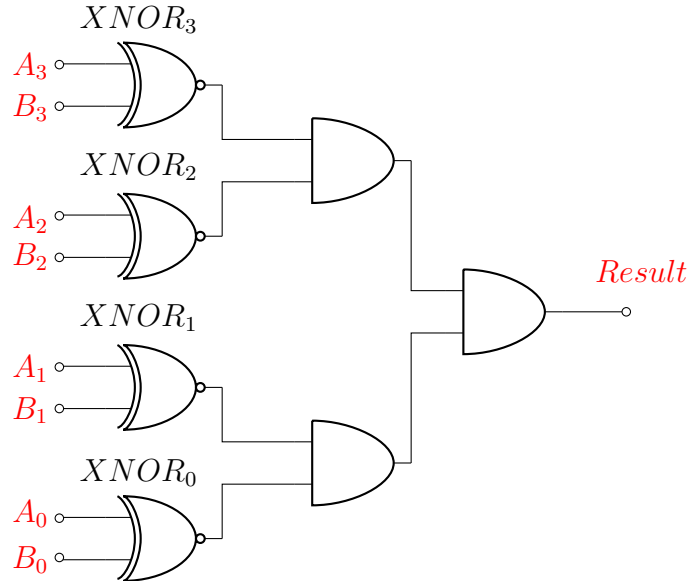


Figure 16: A circuit to provide a yes/no answer to the question, “are two four-bit numbers, A and B, equal?” Each XNOR gate compares the bits in each position of each number, and passes the answer to an AND gate input. The AND results ripple forward to a final answer.

Memory

Logic circuits need to store the numbers they are working with in order to do more than one thing. Central Processing Units (CPUs), Arithmetic Logic Units (ALUs) or Floating-Point Units (FPUs) each work with numbers, and need to fetch them from somewhere and put them somewhere when they are done. The “somewhere” is *memory*. Keeping with the use of binary math and binary logic, a high voltage would be a 1 and a low voltage would be a 0. Through the use of these little stored charges, computers can keep track of many, many pieces of information. No matter what the circuits are doing, everything being stored is composed of some count of zeros and ones.

There have been *many* forms of memory through the years, including punched paper cards, and even a wiggly wire! The most common memory these days is a circuit made up of six transistors, or one transistor and a capacitor or two. Computers read eight, sixteen, 32, or 64 bits at a time from a row of memory and pass the number, or the instructions, from memory on to be processed.

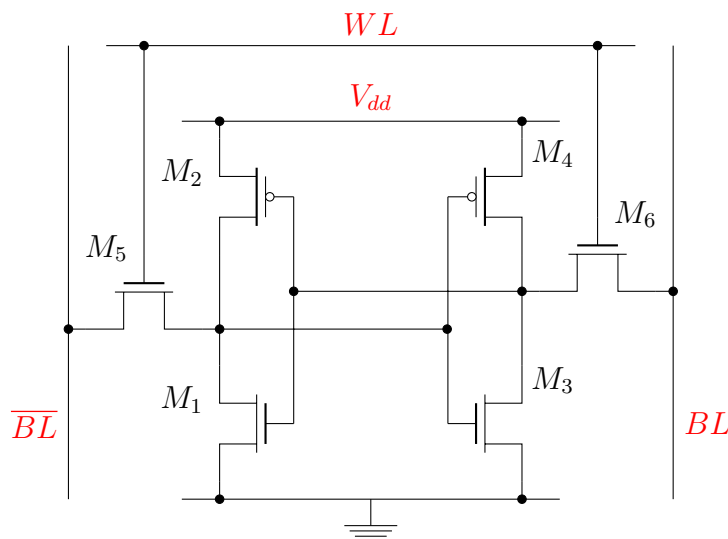


Figure 17: A static ram cell. It uses six transistors per bit instead of one transistor and a capacitor for bit (as used in dynamic ram, the main memory for the computer), so it's expensive and bulky, but it's also faster. Usually SRAM is used very close to the CPU core, so it is available for immediate use by the core processing operations. To keep the diagram simple, no resistors are shown. If you look closely at the M1/M2 and M3/M4 transistors, you can see that they are acting like inverters (that is, each pair makes a NOT gate).