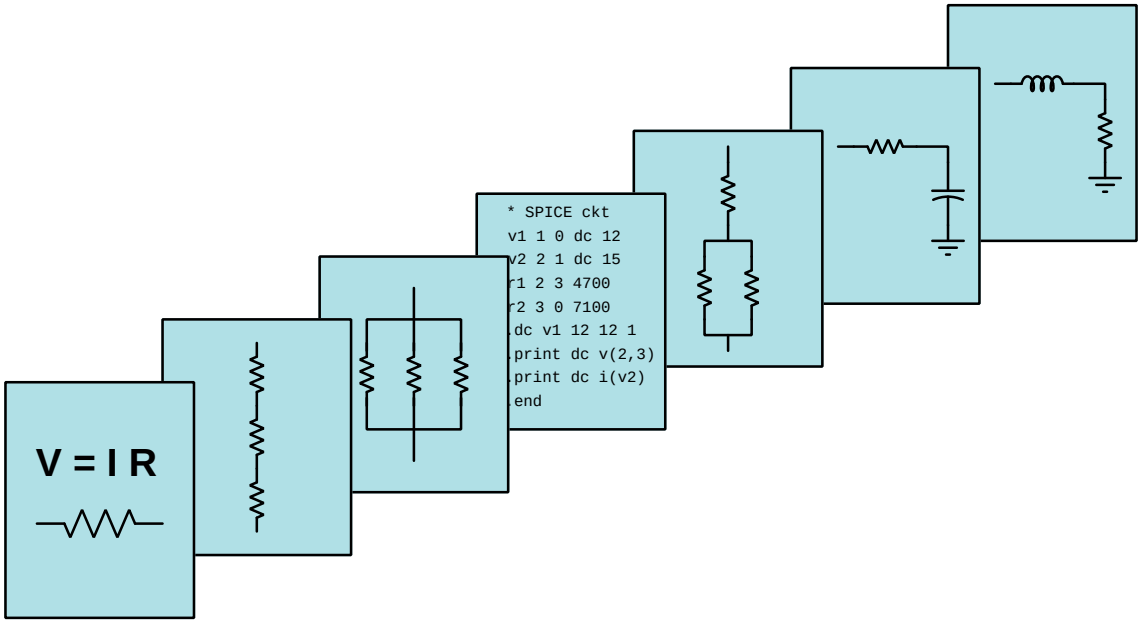
Modular Electronics Learning (ModEL) project



Field-Effect Transistors

© 2019-2025 by Tony R. Kuphaldt – under the terms and conditions of the Creative Commons Attribution 4.0 International Public License

Last update = 29 October 2025

This is a copyrighted work, but licensed under the Creative Commons Attribution 4.0 International Public License. A copy of this license is found in the last Appendix of this document. Alternatively, you may visit `http://creativecommons.org/licenses/by/4.0/` or send a letter to Creative Commons: 171 Second Street, Suite 300, San Francisco, California, 94105, USA. The terms and conditions of this license allow for free copying, distribution, and/or modification of all licensed works by the general public.

Contents

1 Introduction **3**
- 1.1 Recommendations for students . . . 3
- 1.2 Challenging concepts related to field-effect transistors . . . 5
- 1.3 Recommendations for instructors . . . 7

2 Case Tutorial **9**
- 2.1 Example: JFET biasing . . . 10
- 2.2 Example: D-type MOSFET biasing . . . 11
- 2.3 Example: E-type MOSFET biasing . . . 12

3 Simplified Tutorial **13**

4 Full Tutorial **17**
- 4.1 Amplifier concepts . . . 18
- 4.2 Junction field-effect transistors . . . 25
- 4.3 Insulated-gate field-effect transistors . . . 28
- 4.4 FET characteristics . . . 32
- 4.5 Static sensitivity . . . 36

5 Historical References **37**
- 5.1 Early Field-Effect Transistor . . . 38
- 5.2 US Patent for a JFET . . . 42
- 5.3 US Patent for a MOSFET . . . 44

6 Derivations and Technical References **47**
- 6.1 FETs versus BJTs . . . 48
- 6.2 Alternative MOSFET symbols . . . 49
- 6.3 Gauss' Law of Electric Fields . . . 50

7 Animations **55**
- 7.1 Animation of an E-type MOSFET turning on . . . 56

8 Questions **69**
8.1 Conceptual reasoning . . . 73
8.1.1 Reading outline and reflections . . . 74
8.1.2 Foundational concepts . . . 75
8.1.3 JFET depletion regions . . . 78
8.1.4 JFET source and drain terminals . . . 79
8.1.5 Identifying MOSFET states . . . 80
8.1.6 Identifying MOSFET stimuli . . . 83
8.1.7 Sketching MOSFET switch circuit . . . 84
8.1.8 Typical JFET pinch-off voltage values . . . 85
8.1.9 Effect of V_{DS} on depletion region . . . 86
8.1.10 Effects of V_{GS} on MOSFET . . . 89
8.1.11 Interesting transistor symbols . . . 90
8.1.12 Servo motor travel limiting . . . 91
8.2 Quantitative reasoning . . . 92
8.2.1 Miscellaneous physical constants . . . 93
8.2.2 Introduction to spreadsheets . . . 94
8.2.3 Drain current calculation . . . 97
8.2.4 ESD test jig . . . 98
8.3 Diagnostic reasoning . . . 99
8.3.1 Testing a JFET by continuity . . . 100
8.3.2 Testing a JFET by DMM diode check . . . 101
8.3.3 Testing a MOSFET by DMM diode check . . . 102
8.3.4 ESD and MOSFETs . . . 103
8.3.5 Dim lamp . . . 103

9 Projects and Experiments **105**
9.1 Recommended practices . . . 105
9.1.1 Safety first! . . . 106
9.1.2 Other helpful tips . . . 108
9.1.3 Terminal blocks for circuit construction . . . 109
9.1.4 Conducting experiments . . . 113
9.1.5 Constructing projects . . . 117
9.2 Experiment: JFET terminal identification . . . 118
9.3 Experiment: MOSFET terminal identification . . . 119
9.4 Experiment: JFET as a controlled resistance . . . 120
9.5 Experiment: JFET as a controlled current regulator . . . 121
9.6 Experiment: measuring switching time . . . 122

A Problem-Solving Strategies **123**

B Instructional philosophy **125**
B.1 First principles of learning . . . 126
B.2 Proven strategies for instructors . . . 127
B.3 Proven strategies for students . . . 129
B.4 Design of these learning modules . . . 130

C Tools used **133**

D Creative Commons License **137**

E References **145**

F Version history **147**

Index **149**

Chapter 1

Introduction

1.1 Recommendations for students

This module explores the principles of operation, characteristics, and basic uses of field-effect transistors which form the basis of a great many modern electronic circuits.

Important concepts related to bipolar junction transistors (BJTs) include **amplification**, **active** versus **passive** devices, **gain**, **digital** versus **analog** signals, **semiconductors**, **doping**, **depletion regions**, **PN junctions**, forward and reverse **bias**, **transconductance**, **electric fields**, **inversion layers**, **enhancement** versus **depletion**, **threshold voltage**, the mathematical **square law**, **Gauss' Law**, and **electrostatic discharge**.

A useful reading strategy when studying a topic such as this is to write your own summary page of important principles, especially when faced with learning about a number of different component types that are similar (but not identical) to each other. For this module I would recommend writing your own summary page showing every type of field-effect transistor by name and by schematic symbol, and labeling those schematic symbols with voltage polarity and current direction annotations showing how each type of transistor is activated into conduction, as well as the "natural" state of each transistor type.

Here are some good questions to ask of yourself while studying this subject:

- How might an experiment be designed and conducted to measure the threshold voltage for a field-effect transistor? What hypothesis (i.e. prediction) might you pose for that experiment, and what result(s) would either support or disprove that hypothesis?
- How might an experiment be designed and conducted to plot a family of characteristic curves for a field-effect transistor? What hypotheses (i.e. predictions) might you pose for that experiment, and what result(s) would either support or disprove those hypotheses?
- What are some practical applications of field-effect transistors?
- What differentiates an *active* from a *passive* component?

- What is amplification and how does it work?
- How do we calculate gain for any particular device or system?
- Where might we find electromechanical relays used?
- What are vacuum tubes capable of that relays are not?
- Why are vacuum tubes now considered obsolete for most applications?
- How does electrical conduction differ in a semiconductive material compared to a conductive material?
- What role does the depletion region play in a semiconductor diode?
- What role does the depletion region play in a field effect transistor?
- What conditions are necessary to turn a JFET on?
- What conditions are necessary to turn an E-type MOSFET on?
- What conditions are necessary to turn a D-type MOSFET on?
- What is the relationship of collector current to gate voltage in an operating FET?
- What is the relationship of collector current to drain-source voltage in an operating FET?
- If you were to annotate the schematic symbol of an E-type MOSFET, where would the inversion layer be located, and what controlling condition would be necessary to create that inversion layer?
- Why does a parasitic diode exist within a MOSFET, and why do we need to be aware of this fact?
- What is each SPICE simulation attempting to show in the tutorial?
- Under what conditions does a FET act as a resistance, and why does it do so?
- Under what conditions does a FET act to regulate current, and why does it do so?
- What is the significance of a FET's "threshold" voltage?
- What does it mean to say that a FET is a "square-law" device?
- What does Gauss' Law say about electric fields and conductive materials?
- How does Gauss' Law apply to the internal behavior of a MOSFET?
- Why are certain types of FETs susceptible to damage from static electricity?

1.2 Challenging concepts related to field-effect transistors

The following list cites concepts related to this module's topic that are easily misunderstood, along with suggestions for properly understanding them:

- **Necessary conditions for transistor operation** – field effect transistors work on the principle of a transverse electric field affecting the conductivity of a homogenously-doped channel. For insulated-gate FETs (i.e. MOSFETs), the electric field exists across a thin layer of dielectric between the gate terminal and the modulated channel. When that applied gate-channel voltage is of such a polarity as to attract majority charge-carriers into the channel (e.g. electrons into an N-type channel, holes into a P-type channel) then the channel becomes more conductive.

- **Transistor saturation** – a transistor is considered to be *saturated* in its "on" state whenever its conducted current is less than what the transistor is being stimulated to pass; in other words, a "saturated" transistor is turned on to a degree that is in excess of the load current it's actually conducting. Saturation for a field effect transistor is generally a simple matter of exceeding a threshold value for gate-source voltage ($V_{GS(on)}$).

- **Parasitic diode** – MOSFETs are constructed on top of a substrate layer that is of opposite doping to the channel. This creates a PN junction between the channel and the substrate which is called a parasitic diode. This diode points in a direction opposite to the correct drain-source current flow which means it should always be reverse-biased for normal transistor operation. If the drain-source voltage polarity reverses for any reason this diode will essentially short past the transistor and allow current even when the transistor is supposed to be turned off.

- **Schematic annotation as a problem-solving tool** – as an instructor you will find most students new to the study of electronics attempt to solve circuit-analysis problems by inspection, or by plugging given values into familiar equations until something resembling an answer emerges. Such efforts are usually fruitless because it is difficult to track the real-world context of every circuit parameter doing things this way. You will need to demonstrate, encourage, and at times *require* your students to apply the alternative strategy of first sketching a schematic diagram of the circuit and then annotating it with all known values. Once the given information gains context through this annotation, appropriate principles (e.g. Ohm's Law, KVL, KCL) are much easier to identify and then properly apply for that circuit. Then, as calculated values get annotated on the same schematic diagram the next logical steps to follow become easier and easier to identify. I recommend using different colors to label different parameters (e.g. blue for voltage, red for current, black for component values, green for power). Students should have blank "scratch" paper and writing instruments with them at all times, even during discussions, so as to neatly organize information and to articulate and focus their thoughts! Despite introducing some new concepts in the study of transistors, all transistor circuits must still exhibit Kirchhoff's Laws as well as Ohm's Law for its resistors!

- **Relevance of foundational concepts** – this is not a new concept in itself, but rather an approach to learning and to problem-solving that many students struggle with. Students must be encouraged to continually apply trusted concepts such as Ohm's Law, Kirchhoff's Voltage Law, and Kirchhoff's Current Law to every new subject they encounter in Electronics, rather than to ignore these first principles in favor of memorizing new rules. Memorization without

understanding is a recipe for failure. Logically reasoning from trusted laws and principles, however, strengthens recall and hones the ability to tackle novel problems. Within the subject of field-effect transistors, Ohm's Law, KVL, and KCL are ever-relevant and need only be augmented by a few new facts (e.g. channel width modulation, inversion layers, drain-source resistance, etc.) to fully equip oneself for analyzing most any FET circuit.

- **Creative application of foundational concepts** – many students enter college-level study of electronics with an educational background stressing rote memorization at the expense of logical reasoning from trusted principles, and as such tend to find circuit analysis daunting where there is no single procedure or single formula always yielding the correct answer(s). The educational solution to this is for the instructor to *constantly* tie all new circuit concepts back to previously-learned principles and show the logical relationships between them, and then with every new circuit-analysis example challenge students to reason from those concepts on their own. When students construct a mental "toolbox" they may apply in a plurality of ways to any new challenge they encounter, they will be as free as any skilled mechanic who knows how and why each of their hand tools functions to freely select and use whichever of those tools is most appropriate at any given time. A word of caution for new instructors: *be very sparing in your direction on how to use those tools, instead engaging with students Socratically when they are "stuck" and cannot see a path to a solution!*

1.3 Recommendations for instructors

This section lists realistic student learning outcomes supported by the content of the module as well as suggested means of assessing (measuring) student learning. The outcomes state what learners should be able to do, and the assessments are specific challenges to prove students have learned.

- **Outcome** – Demonstrate effective technical reading and writing

  Assessment – Students present their outlines of this module's instructional chapters (e.g. Case Tutorial, Tutorial, Historical References, etc.) ideally as an entry to a larger Journal document chronicling their learning. These outlines should exhibit good-faith effort at summarizing major concepts explained in the text.

  Assessment – Students show how quantitative results were obtained by the author in the Tutorial chapter's examples.

  Assessment – Students present their "Conceptual Toolbox" Theory Journal entries listing important concepts applied in the Tutorial text(s). A "Conceptual Toolbox" serves as inspiration when solving problems, each concept suggesting a potential problem-solving action. Challenge students to collaboratively generate a Conceptual Toolbox for the day's topic during the class session and then use it when solving problems together, thus fostering the habit of reasoning from principles rather than recalling procedures!

- **Outcome** – Properly associate controlling stimuli to effects in transistors

  Assessment – Identify MOSFET transistor states with various input stimuli given; e.g. pose problems in the form of the "Identifying MOSFET states" Conceptual Reasoning question.

  Assessment – Identify proper directions of applied voltage for field effect transistors; e.g. pose problems in the form of the "Identifying MOSFET stimuli" Conceptual Reasoning question.

- **Outcome** – Sketch functional transistor switching circuits

  Assessment – Draw connecting wires between given components to control a MOSFET so that it switches power on and off to a motor; e.g. pose problems in the form of the "Sketching MOSFET switch circuit" Conceptual Reasoning question.

- **Outcome** – Test field-effect transistors with a multimeter

  Assessment – Identify JFET terminals based on multimeter diode-check measurements; e.g. pose problems in the form of the "Testing a JFET by DMM diode check" Diagnostic Reasoning question.

  Assessment – Identify MOSFET terminals based on multimeter diode-check measurements; e.g. pose problems in the form of the "Testing a MOSFET by DMM diode check" Diagnostic Reasoning question.

- **Outcome** – Independent research

  Assessment – Locate MOSFET datasheets and properly interpret some of the information contained in those documents including maximum drain-source resistance when on ($R_{DS(on)}$),

maximum drain current rating, maximum power dissipation rating, gate-source threshold voltage ($V_{GS(on)}$), etc.

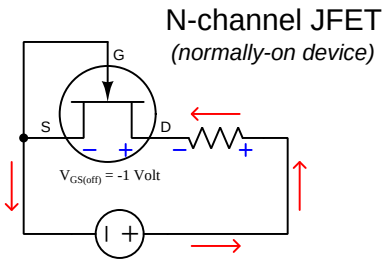
Chapter 2

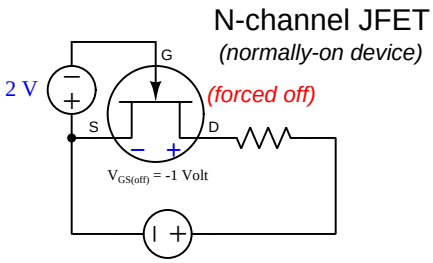
Case Tutorial

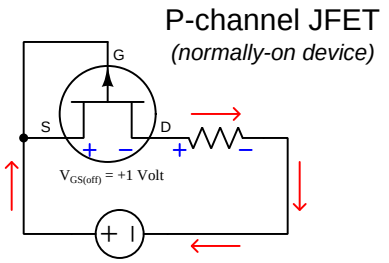
The idea behind a *Case Tutorial* is to explore new concepts by way of example. In this chapter you will read less presentation of theory compared to other Tutorial chapters, but by close observation and comparison of the given examples be able to discern patterns and principles much the same way as a scientific experimenter. Hopefully you will find these cases illuminating, and a good supplement to text-based tutorials.

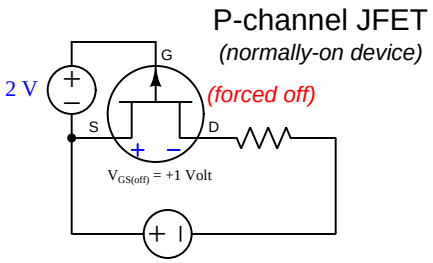
These examples also serve well as challenges following your reading of the other Tutorial(s) in this module – can you explain *why* the circuits behave as they do?

2.1 Example: JFET biasing

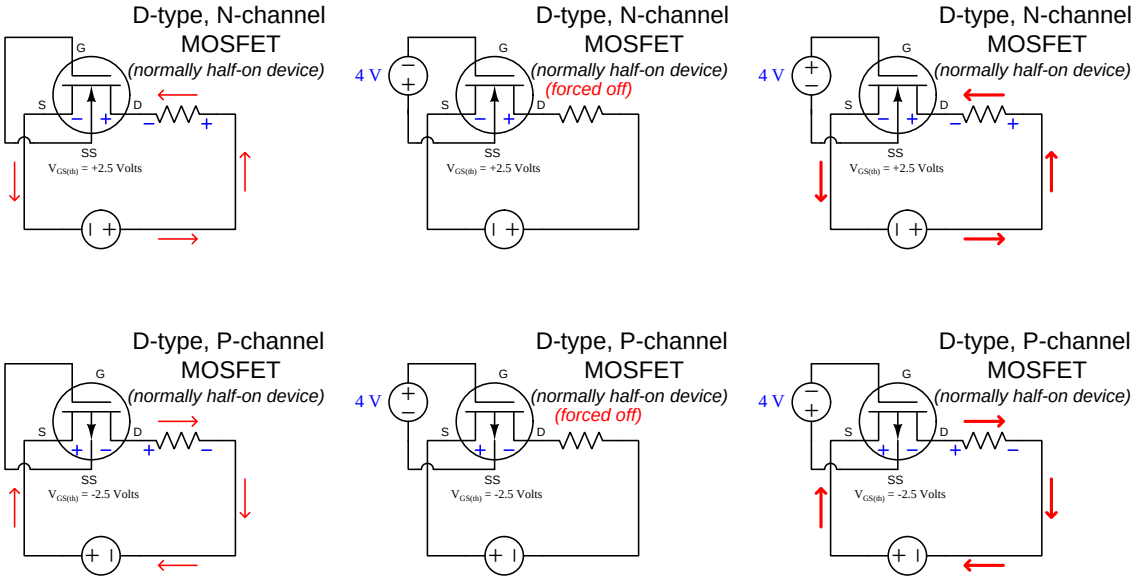




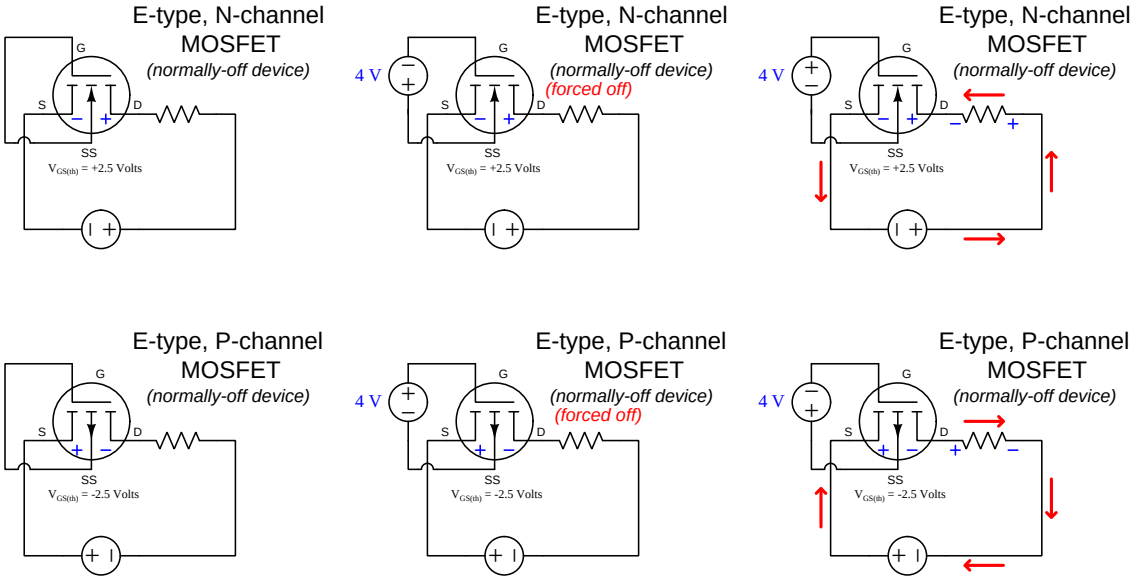




2.2 Example: D-type MOSFET biasing



2.3 Example: E-type MOSFET biasing

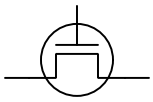


Chapter 3

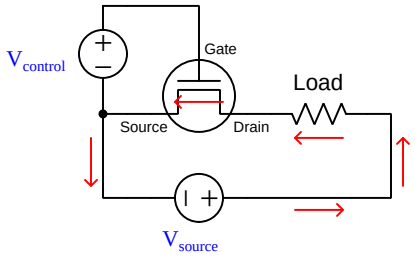
Simplified Tutorial

A *field-effect transistor* is a semiconductor device that may be turned on or off (or made partially conductive) by applying an electric field perpendicular to the direction of the controlled current. The generic symbol for a field-effect transistor, or *FET*, appears below:

FET schematic symbol



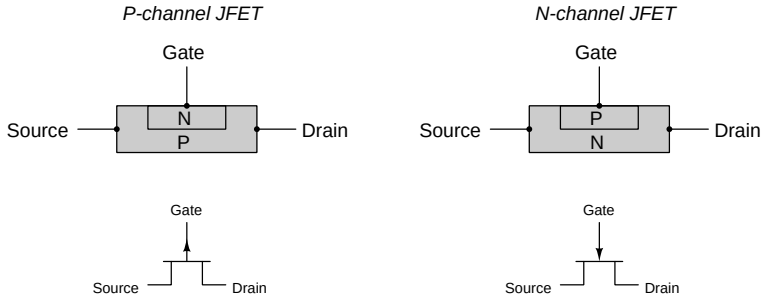
FET operation



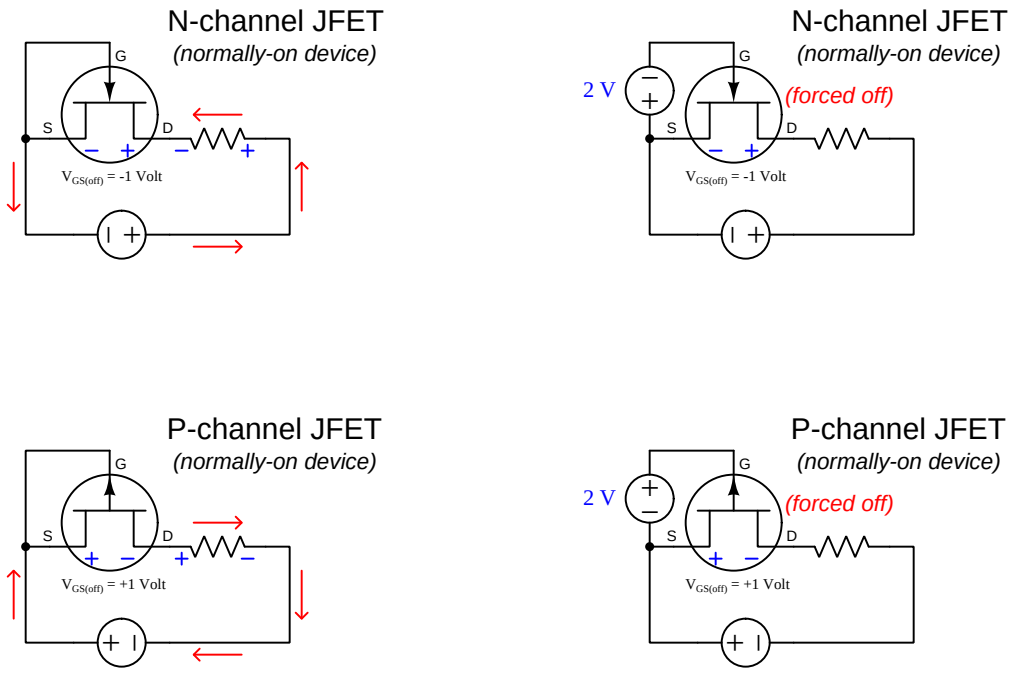
The controlling voltage ($V_{control}$) applied between the FET's *gate* and *source* terminal produces an electric field inside the transistor that is transverse (perpendicular to) the axis of controlled current. Properly operated, there is negligible current flowing through the gate terminal, and therefore the FET appears to the controlling voltage source as a very light (or nonexistent) load. This unique quality gives FETs a very high *current gain*, which means the ratio of controlled current to controlling current is practically infinite. The amount of controlling voltage necessary to either completely turn the transistor on or completely turn it off is called the *threshold voltage*, usually on the order of a few Volts. The turn-off voltage is called the *pinch-off voltage*; the turn-on voltage is simply called *turn-on voltage*.

Perhaps the most confusing aspect of using field-effect transistors is all the different types of FETs there are, and remembering how each of them behave. There exists three major types of FET, and each of these is manufactured in either N-channel or P-channel varieties, referring to the particular doping of the semiconductor "channel" carrying the controlled current. This Simplified Tutorial will focus purely on the behavioral details of FETs and not their internal construction or theory.

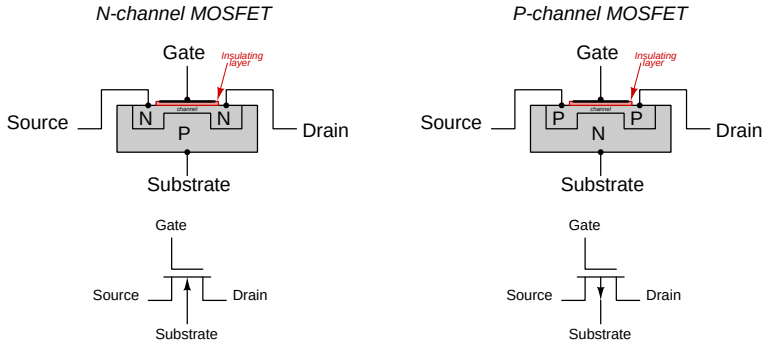
First we have the *Junction Field-Effect Transistor*, or *JFET*, which gets its name from the fact that there is a PN semiconductor junction between the gate terminal and either the source or drain terminals.



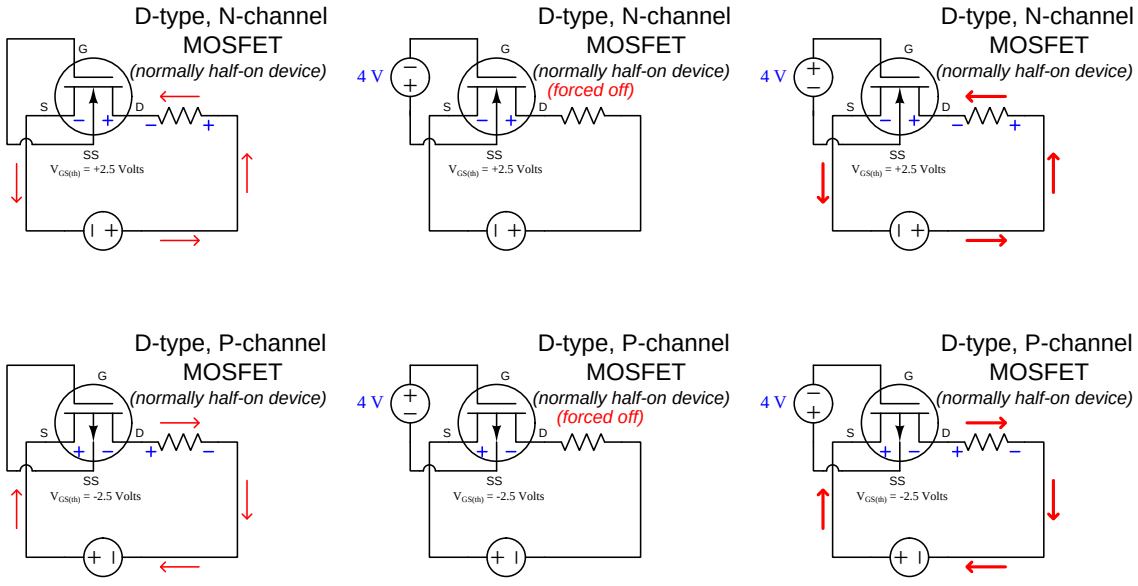
The controlling voltage must be applied with reverse-bias to this PN junction, and doing so has the effect of *turning off* this normally-on transistor. We must be careful never to forward-bias the PN junction between a JFET's gate and channel because that PN junction is very fragile and not intended to carry substantial current:



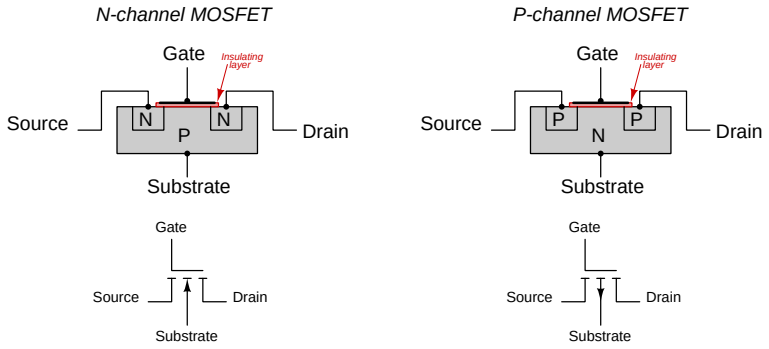
Next we have the *Depletion-type Metal-Oxide Semiconductor Field-Effect Transistor*, or *D-Type MOSFET*, which gets its name from the fact that there is an insulating layer of silicon dioxide separating the metal gate terminal from the "channel" through which current flows between source and drain. The controlling voltage may be of either polarity, one turning the transistor more on (enhancing) and the other turning it more off (depleting). In its natural state the MOSFET is half-way on (i.e. like a resistor):



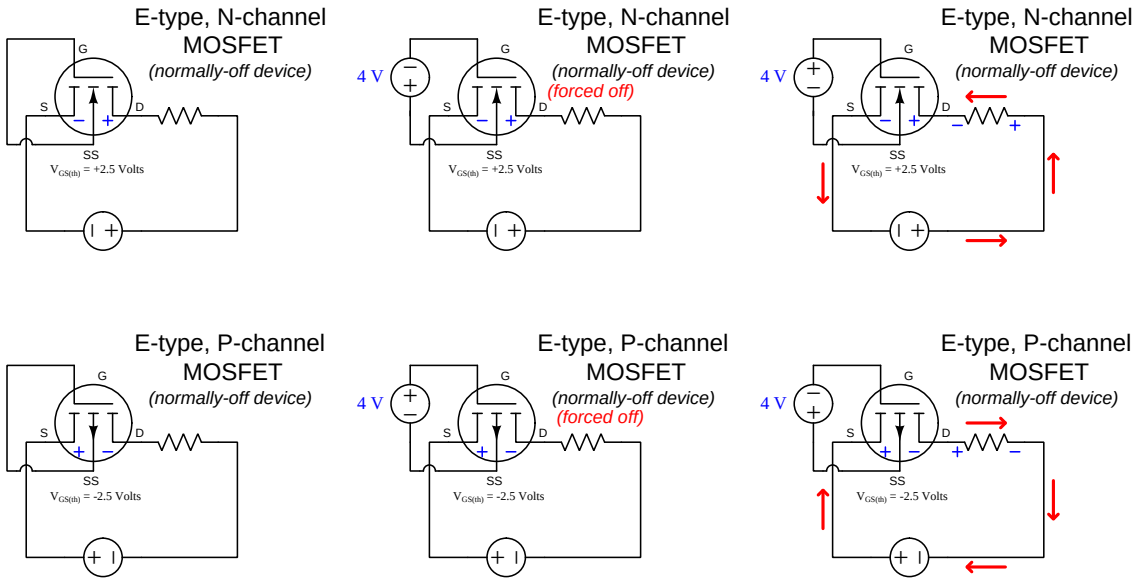
To turn on (enhance) a MOSFET, we need to apply a gate-substrate voltage of such polarity that the electric field will *attract* the channel's majority-type charge carrier toward the gate terminal to make the channel more substantial; e.g. a positive gate for an N-channel and a negative gate for a P-channel. Reversing this polarity will turn off (deplete) the MOSFET:



Last we have the *Enhancement-type Metal-Oxide Semiconductor Field-Effect Transistor*, or *E-Type MOSFET*, which differs from the D-type MOSFET primarily in the fact that it is a normally-off device which may only be turned on by applying the correct polarity of controlling voltage:



Once again, to turn on (enhance) a MOSFET, we need to apply a gate-substrate voltage of such polarity that the electric field will *attract* the channel's majority-type charge carrier toward the gate terminal to make the channel more substantial; e.g. a positive gate for an N-channel and a negative gate for a P-channel. Although it's possible to do, there is no point in applying the reverse of this turn-on polarity because the E-type MOSFET is already a normally-off device and cannot be turned off any more than it is in its natural state:



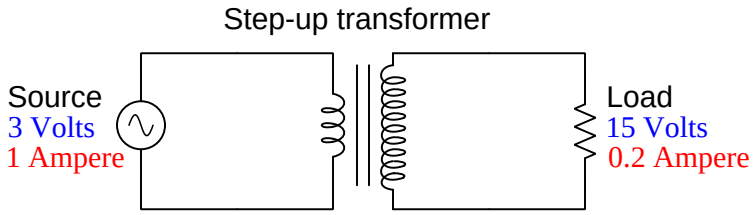
All varieties of MOSFET are susceptible to accidental damage from static electricity, since the high voltages produced by common sources of static electricity can easily puncture the microscopically-thin silicon dioxide insulating layer and thereby render the transistor useless.

Chapter 4

Full Tutorial

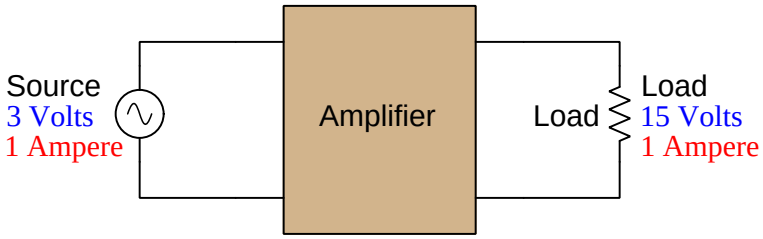
4.1 Amplifier concepts

An *amplifier* is any device allowing a small amount of energy to exert control over a larger amount of energy. It is easy to misunderstand this concept, and so we will begin by exploring what amplification is *not*. Consider a step-up transformer, where an AC source energizes a primary winding, producing a larger voltage at the secondary winding to energize a load:



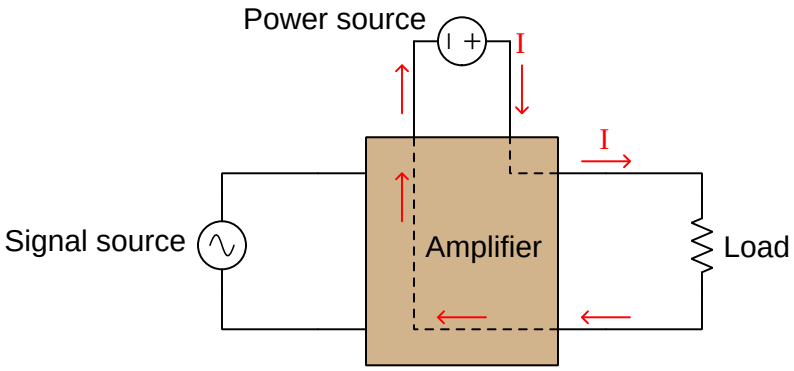
This transformer boosts voltage by a factor of five, but this magnification of voltage is *not* amplification because it comes at the expense of an equivalent reduction in current. Like a mechanical lever used to multiply either force or motion, multiplication of one always leads to division of the other by the same ratio so that energy in equals energy out (in this case, 3 Watts in = 3 Watts out).

An amplifier, then, must be capable of outputting more energy than it receives from its signal source. For example, a circuit boosting voltage while outputting the same current as it inputs would be a true amplifier:



How is this possible, though, without violating the Law of Energy Conservation?

The answer to this very good question is that amplifiers must *also* connect to an additional energy source providing the greater electrical power delivered to the load. The input signal merely controls the flow of electrical energy from source to load. Another way of stating this fact is to say that amplification is a form of *controlled conductivity*, where the input signal determines the degree of conductivity for some other energy source's output to flow to a load:



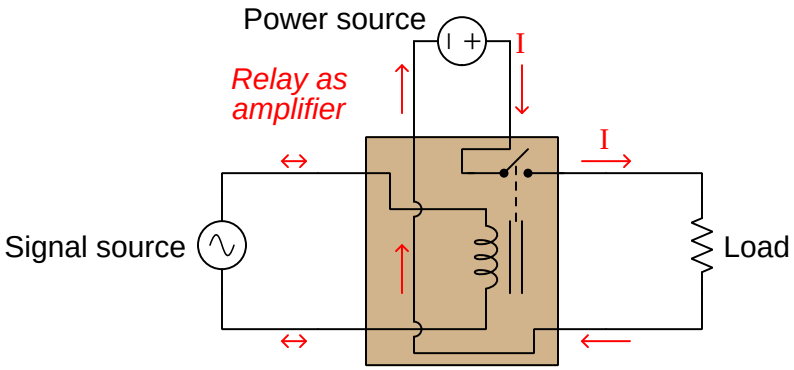
By the same definition, a lever used to magnify force is *not* an amplifying device because you never get more mechanical work out of the lever than what you put into it. However, a motor-driven crane is an amplifying device because a small amount of force and motion applied by a human operator to pushbuttons or levers on the crane *controls the flow of energy from the crane's motor to the load* to lift heavy objects over large distances. The specific component within the crane enabling the operator's weak hand motions to direct many kilowatts' worth of mechanical power to move the load could be a hydraulic valve, a clutch, or some other energy-controlling device. Its specific identity is not important, but its function is essential.

Terms used to distinguish energy-*controlling* devices from other devices, at least in the world of electric circuits, are *active* versus *passive*. Active devices are what make true amplification possible. Components such as resistors, inductors, capacitors, diodes, switches, fuses, and batteries are all passive: they may be sources, they may be loads, and they may alternate roles, or they may be neither, but they do not permit one electrical signal to *control conductivity* for another.

An important way to quantify amplification in any of its forms by a term called *gain*. Simply defined, gain is the ratio of controlled versus controlling quantities, usually symbolized by the variable A. As such, the value of A tells us how many times larger an amplifier's output signal is compared to its input signal. For example, the *power gain* of an amplifier is the ratio of its output signal power divided by the controlling input signal power: $A_P = \frac{P_{out}}{P_{in}}$. The voltage gain of an electronic amplifier is the ratio of output signal voltage to input signal voltage: $A_V = \frac{V_{out}}{V_{in}}$. The current gain of an electronic amplifier is the ratio of output signal current to input signal current: $A_I = \frac{I_{out}}{I_{in}}$. Similarly, we could rate the gain of a motorized crane in terms of force (i.e. the amount of lifting force output divided by the amount of force necessary to push the control lever): $A_F = \frac{F_{lift}}{F_{lever}}$. In every case, gain is a unitless quantity because it is always a ratio between quantities having identical units of measurement (e.g. Watts versus Watts, Volts versus Volts, etc.).

The first active circuit component was the *electromechanical relay* – a combination of solenoid coil and electrical switch. When a power source energizes the solenoid coil, its magnetic field attracts an iron armature which in turn actuates the switch. Relays may be considered active components because the amount of energy conveyed by the switch contact may be much greater than the amount of energy required to energize the coil.

An amplification circuit using a relay is shown in the following schematic diagram:



All energy delivered to the load originates from the power source, not the signal source. All the signal does is energize the relay's coil, to *control* energy sent to the load from the power source. The relay's switch contact then closes to allow the flow of electrical energy to the load, at whatever voltage and current levels the contact is able to handle, which is typically *thousands of times greater*[1] than the energy required to energize the coil.

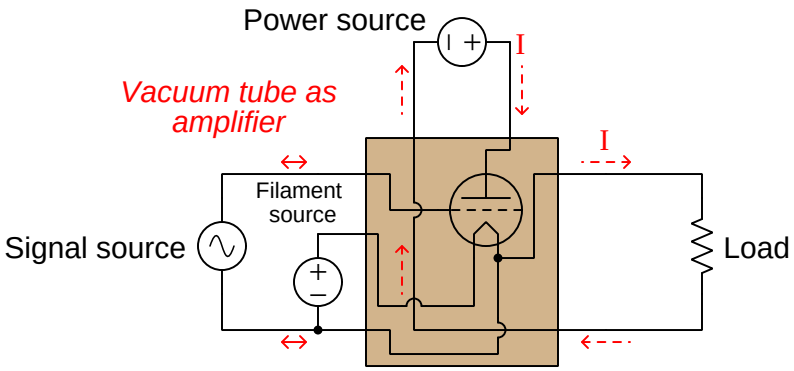
Electromechanical relays revolutionized electric circuits by enabling them to perform functions previously unimaginable. One of these functions is the ability of a fragile switch to exert control over a powerful load: the switch only needs to handle as much current when closed and voltage when open as is necessary to energize the relay coil, while the relay's much larger switch contacts handle the voltage and current requirements of the powerful load. *Telegraph* systems were one early application of relays, where they were used as "repeaters" to boost the power of a telegraph signal weakened by many kilometers of wire resistance, thereby extending the practical range of wired telegraphy. Another function permitted by relays is *automation*, where electrical systems become self-controlling. This is only possible where one electrical signal has the ability to control another, which is the very essence of relay function.

While electromechanical relays made possible such technologies as long-distance telegraph networks, automatic motor control systems, automatic voltage regulators for generators, automatic protection systems for electrical power networks, and even the first digital electric computers, relays do have some significant limitations. Relay contacts are either open or closed, and as such they cannot achieve any states in-between those two. As such, relays are helpless to amplify anything but discrete on-off signals. Amplification of other signal types, such as *analog* signals where information is encoded in the variable magnitude of the signal, is beyond the reach of electromechanical relays.

[1] Thus, we could say that a relay's *power gain* is a unitless value ranging in the thousands.

The next innovation in active devices was the *vacuum tube* or *electron tube*, originally known as the *thermionic valve*. This was a derivative of the incandescent electric lamp, exploiting the fact that a glowing-hot metal surface emits free electrons into the space around it. Consisting of a glass envelope evacuated of air and containing a filament along with one or more metal grids and plates, vacuum tubes used the application of a small voltage between the grid and filament to control the flow of electrons passing from[2] the filament to the plate.

An amplification circuit using a vacuum tube is shown in the following schematic diagram:



Vacuum tubes once again revolutionized electric circuits by enabling them to perform functions previously unimaginable. In fact, this is the innovation that ushered in the era of *electronic* circuits, where the flow of electrons could be precisely controlled by a voltage signal. Unlike relays which are discrete in nature, vacuum tubes are capable of analog operation, gradually restricting the flow of electrons from full rated current to zero, and any level in-between. Vacuum tubes were also much *faster* than relays, allowing a time-varying signal to reproduce its same time-variance at the output. Note in the diagram shown above how the AC signal causes the DC power source to apply a *pulsed* DC signal to the load. It is as though the controlling signal causes an AC-like waveform to be "carved" out of the steady DC supply voltage.

The ability of a vacuum tube to amplify AC signals of arbitrary wave-shape led to the development of new forms of communications including *telephone*, *radio*, and *television*. Prior to vacuum tubes, the only practical form of long-distance electrical communication was the *telegraph*, using discrete on/off states to digitally represent letter and number characters. Vacuum tube electronic circuits also made possible *radar* to detect objects far out of human eyesight, and played a significant role in World War Two. Another significant military application of vacuum tubes was automatic gun direction systems, which when coupled to radar systems made possible automatic defense from aerial attacks.

For all their capabilities, vacuum tubes had significant limitations. Their requirement of a glowing-hot metal filament made them fairly energy-inefficient. They were physically fragile as well, owing to their glass envelope construction. Also, the high-vacuum condition required inside the glass envelope to permit electrons to freely flow between filament and plate was difficult to maintain. Over

[2] Vacuum tubes are one of the few applications where the *electron flow* convention of denoting current makes more sense than *conventional flow*. In the diagram shown here, the red current arrow still points in the direction of conventional flow, even though it is electrons flowing from the "pointed" filament structure to the flat plate structure inside the tube that comprises load current through the tube.

time, ambient air would leak inside the tube and render it useless. Finally, they were limited in current to milliAmperes, requiring high voltage to amplify signals of significant power.

These limitations spurred the invention of active devices based on electricity conducting through solid materials rather than through a vacuum. The most promising of these solid materials was a class of substances called *semiconductors*. A pure "semiconductor" may be defined as a material whose charge carriers inhabit energy states very nearly, but not quite, permitting charge carrier motion within the substance. This stands in contrast to conductors possessing an abundance of free charge carriers, and insulators where the charge carriers are fairly "locked" into place by a lack of available energy states.

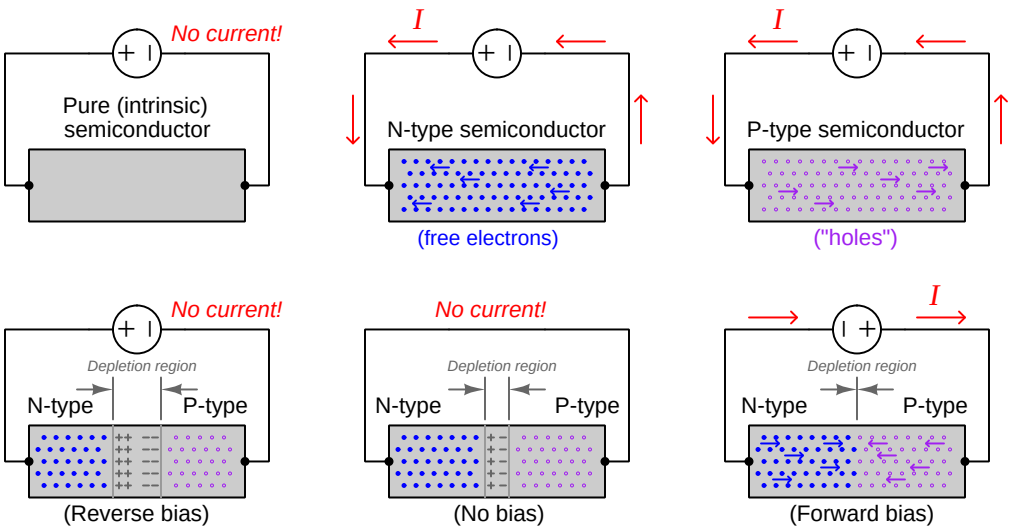
It is not as though a semiconductor is just a poor conductor, either. A conductor of poor quality contains mobile charge carriers, but its charge carriers collide with stationary atoms as they attempt to drift through the atomic lattice of that substance. These collisions convert the charge carriers' kinetic energy into heat, thus constituting electrical resistance. A semiconductor in its pure state, by contrast, contains no free charge carriers at room temperature – i.e. a pure semiconductor is an insulator – but just the right stimulus will *liberate* charge carriers, setting them free to drift through the material's volume. In other words, the phenomenon of *controllable conductivity* is what makes a semiconductor "semi"-conductive.

In a sense, the internal condition of a pure semiconductor is analogous to the internal condition of a vacuum tube: when cool, no free charge carriers exist within the vacuum space of the tube. When stimulated (e.g. heated), however, charge carriers jump off the heated filament and enter the vacuum where they are free to drift, a phenomenon known as *thermionic emission*. In fact, heat is one of the stimuli known to cause semiconductor materials to transition from insulator to conductor, just like a vacuum tube. Light[3] also liberates charge carriers within a pure semiconducting substance.

Through many years of research and experimentation it was discovered that electrical conductivity through a semiconductor could be controlled by carefully alloying ("doping") the semiconducting material with certain types of impurities and applying voltage across junctions of differently-doped semiconductor. The basis of this phenomenon is something called a *depletion region* that forms between differently-doped layers of semiconducting material. Semiconductors doped with impurities placing electrons at energy levels within reach of free motion within the material are called *N-type*. Semiconductors doped with impurities creating vacancies (called *holes*) for low-energy electrons to "hop" into and out of these vacancies from atom to atom are called *P-type*. Both N-type and P-type semiconductors are electrically conductive at room temperature and behave as resistances. Whenever an N-type and a P-type material meet, though, free electrons from the N-type rush to fill vacancies in the P-type, leaving a narrow region at the junction depleted of free charge carriers (i.e. insulating). This so-called *depletion region* may be expanded or collapsed by an applied voltage, making its conduction controllable by a small electrical signal: the key to amplification.

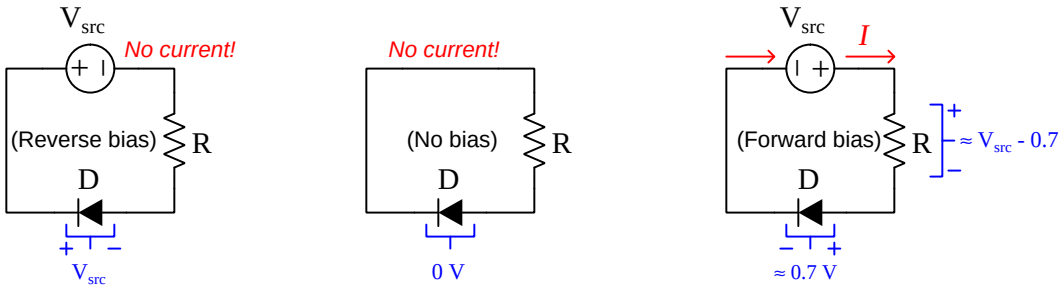
[3] Interestingly, light can also do the same in a vacuum tube, and this is known as the *photoelectric effect*. Not all colors of light will do this, but only those photonic frequencies high enough to meet the *work function* criterion, which is essentially the "escape velocity" energy needed to boost an electron free of a metal surface and into a surrounding vacuum.

A series of illustrations shows how electrical conduction through semiconductors works:



Any section of pure semiconductor, in the absence of heat and light, essentially acts as an insulator and conducts no current. Impurities added to the semiconductor making it either N-type or P-type permits charge carrier motion and thus the semiconductor becomes a conductor. At the junction point of N-type and P-type semiconductors a depletion region forms that is devoid of mobile charge carriers, much like a pure semiconductor, forming a narrow insulating region. However, this depletion region is elastic, and may be expanded or collapsed by the application of an external "bias" voltage.

A single junction formed of N-type and P-type semiconductor alloys (called a *PN junction*) is the basis for solid-state *diodes*. When unbiased, a diode is electrically "open", and remains that way until forward-biased with sufficient voltage. The amount of voltage required to fully collapse the depletion region and make it conductive once again is fairly small, approximately 0.5 to 0.7 Volts for a PN junction made of silicon. It is for this reason that approximate calculations made in semiconductor circuits usually assume a 0.5 to 0.7 Volt forward voltage drop in a conducting diode:

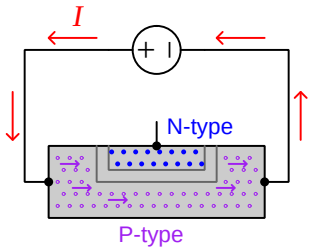


Just like the relay and the vacuum tube, the advent of "solid state" semiconductor electronic devices made possible technologies that were previously unimaginable. Solid-state diodes made conversion of AC power to DC simple, reliable, and power-efficient. *Transistors* which are the solid-state equivalent of vacuum tube amplifiers made possible portable radios and televisions as well as the computer technology you are no doubt using to read this very document!

4.2 Junction field-effect transistors

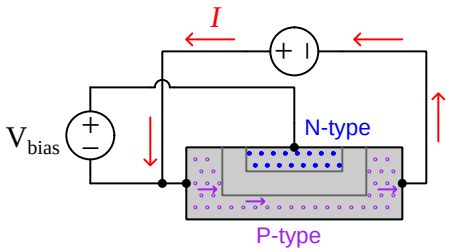
Electrical conduction through a semiconductor diode relies on charge carriers crossing a PN junction, and it is the existence (or collapse) of the depletion region surrounding this junction that either blocks or permits current to pass longitudinally through. However, longitudinal current flow is not the only use for a PN junction and its depletion region; we may also use a PN junction and its depletion region to influence a *transverse* current.

In order to demonstrate this effect, it is helpful to consider a different form of PN semiconductor structure shown in the following illustration. Here, a relatively large block of P-type material has a small inset portion of N-type material. Wire contacts are provided at either end of the P-type block at the N-type insert:



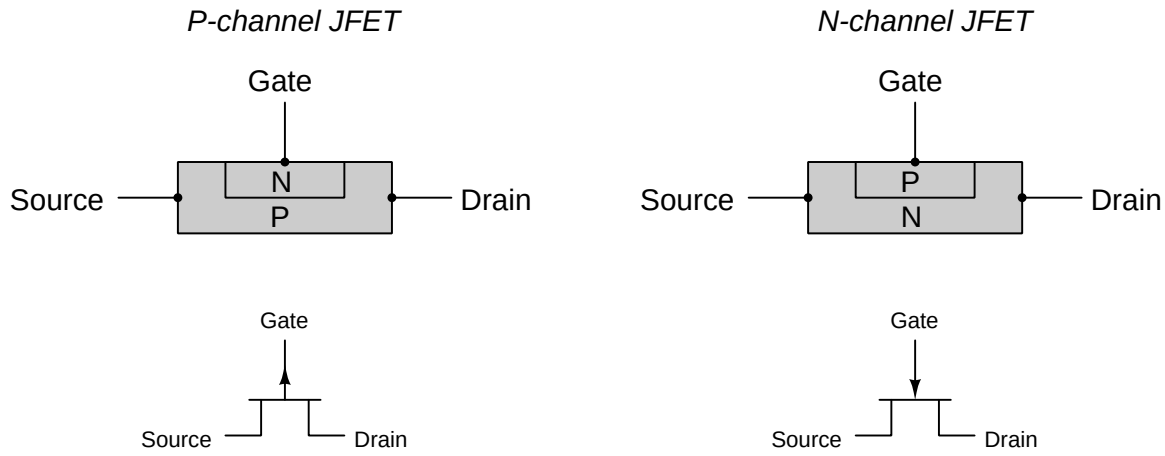
Current through the P-type block, of course, is comprised of *holes* rather than electrons. The isolated N-type insert contains electrons, but these are of no consequence because of the depletion region between the N- and P-type materials. This depletion region narrows the path through which holes must pass, and therefore presents some electrical resistance to those moving holes.

If we intentionally connect a small DC "bias" voltage between the N-type insert and one of the two terminals of the P-type block such that the PN junction is reverse-biased, we will be able to expand the depletion region just as we saw in the case of a reverse-biased diode. Negligible current will pass through the bias source due to the PN junction being reverse-biased, but more importantly the *other* current will be affected by the narrowed "channel" through which it must pass. Thus, we have a method for a small signal (V_{bias}) to influence control over a larger electrical signal:

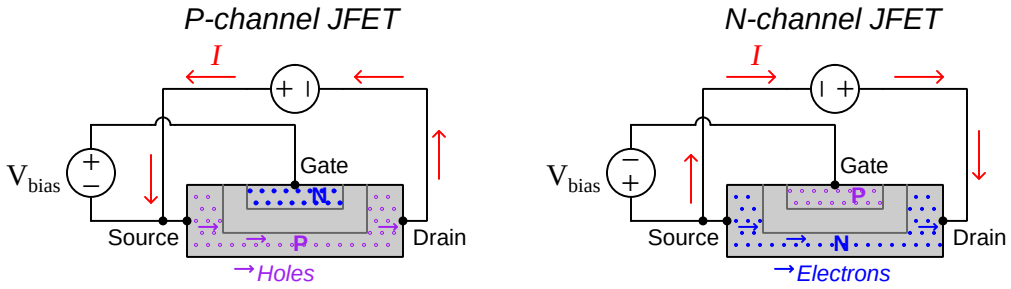


We can imagine increasing V_{bias} until the depletion region expands to fill the P-type block's entire width, at which point the "channel" will be completely *pinched off* and all current will cease. What we have managed to create with this structure is a normally-on *transistor*, turned off by the application of an external bias voltage. More specifically, we call this a *junction field-effect transistor* (*JFET*) because the mechanism of control is the electric field imposed by V_{bias} directly across a PN junction to expand its depletion region.

Field-effect transistors may use P-type "channels" for conduction of the controlled current (as previously shown), or N-type channels. The choice of channel type is largely arbitrary[4] except for the polarity of the controlling voltage signal. The following illustrations show cross-sectional views of the transistor structures (above) and their corresponding schematic symbols (below):



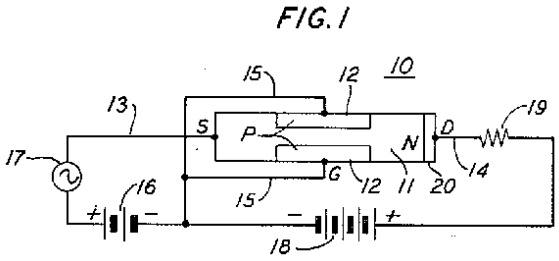
Field-effect transistors are considered *majority-carrier*[5] devices since the controlled current is comprised of charge carriers "native" to the channel material's doping type (e.g. holes flowing through P-type or electrons flowing through N-type). This fact is relevant to the naming of the source and drain terminals: the "source" terminal is the starting point for the majority carriers while the "drain" terminal is the ending point for those majority carriers:



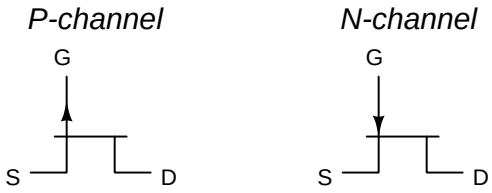
[4] One significant difference is the relative mobility of electrons versus holes. Due to the physics of how each type of charge carrier move, electrons are naturally more mobile than holes. This means that – with all other factors being equal – an N-channel FET conducts electricity better (i.e. has lower resistance) than a P-channel FET. This fact has consequences for high-power transistor circuits, where N-channel transistors are generally preferred over P-channel for their lower "on" resistances and higher current ratings.

[5] By contrast, bipolar junction transistors are considered *minority-carrier* devices since their controlled currents consist of "foreign" charge carriers injected into the base region from the emitter region. For example, an NPN transistor relies on electrons from the N-type emitter being injected and swept across the thickness of the P-type base.

Variations on this design theme are possible when forming JFETs. For example, an illustration taken from the first United States patent for a JFET shows the gate consisting of *two* inserts on either side of the channel (12), those two inserts being made electrically common to each other by a connecting wire (15)[6]:



Additionally, some JFETs are built with the gate(s) positioned closer to the source terminal than to the drain. This design minimizes capacitance[7] between gate and drain terminals. Asymmetric JFETs are sometimes designated in schematic diagrams by symbols showing the gate terminal shifted toward the source and away from the drain:



A common limitation of all JFETs is that their gate terminals are not designed to carry substantial amounts of current. This is because the "transistor action" of a JFET is rooted in the expansion of the depletion region which "pinches off" the channel, and this happens by way of a *reverse-bias* gate-to-channel voltage, not a *forward-bias* voltage. In principle, one could very carefully apply a "forward" gate-channel voltage to a JFET to enhance its channel conductivity, but this would court disaster.

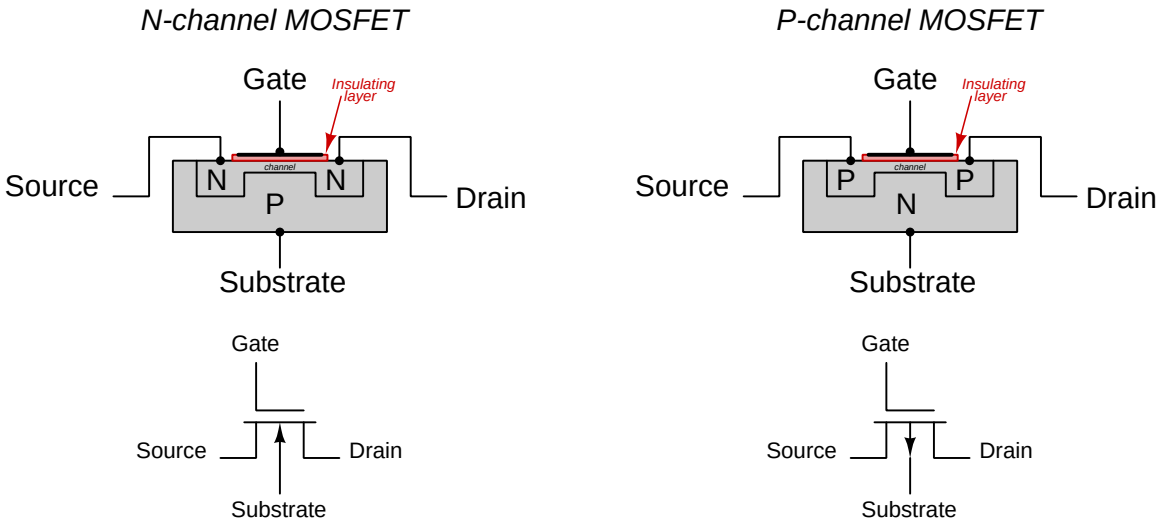
The next step in the evolution of the field-effect transistor eliminated this limitation by making it impossible for a gate current to exist regardless of the applied gate voltage. This was accomplished by *insulating* the gate electrode by a layer of non-conducting material (typically oxidized silicon).

[6] We also see an AC voltage source (17) in this diagram, the purpose being to show how a JFET could be used to amplify AC signals. The DC voltage source numbered 16 is present to properly "bias" the JFET for approximately linear operation so that the AC signal will be reproduced in amplified form at the load resistance (19) with as little distortion as possible.

[7] Recall that the amount of capacitance between any two conductors separated by an insulator is directly proportional to overlapping area and inversely proportional to separation distance ($C = \frac{\epsilon A}{d}$). If the gate insert is positioned closer to the source terminal and farther from the drain terminal, gate-drain capacitance will be less than gate-source capacitance. The rationale for this design feature has to do with the *Miller effect* whereby capacitive coupling between the input and output terminals of an amplifying element such as a transistor tends to degrade high-frequency performance.

4.3 Insulated-gate field-effect transistors

The so-called *Insulated Gate Field-Effect Transistor* (*IGFET*) or *Metal-Oxide-Semiconductor Field-Effect Transistor* (*MOSFET*) utilizes an insulating silicon-dioxide barrier to electrically isolate the metal gate terminal from any other portion of the transistor body. The conductive channel is a thin layer of semiconducting material overlaid on a substrate of the opposite doping:



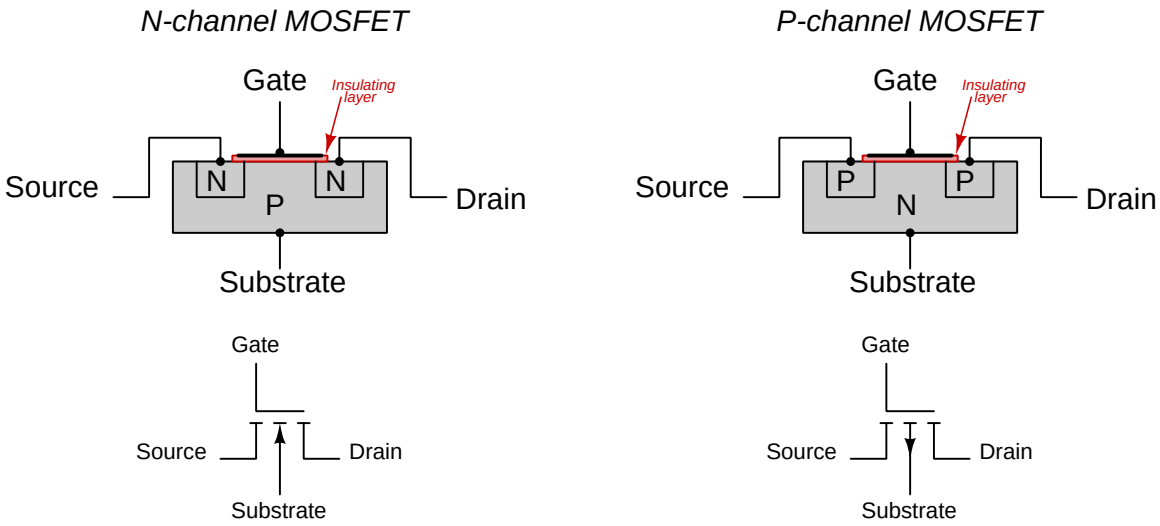
If we apply a voltage between the gate and substrate terminals, an electric field forms between the metal gate terminal and the substrate, passing through the dielectric (insulating) layer as well as the channel and PN junction formed by the channel and substrate. This field redistributes[8] charge carriers directly beneath the insulating layer to alter that channel's electrical conductivity.

Consider the N-channel MOSFET shown on the left-hand side of the above illustration. If the gate is made positive with respect to the substrate, it will attract electrons toward the channel and repel holes into the substrate. This effectively thickens the channel to make it more conductive to electrons (majority carriers for N-type material). Reversing polarity of the gate-substrate voltage has the opposite effect: a negative gate repels electrons and attracts holes, thereby making the N-type channel thinner and therefore less conductive. These same principles are true for the P-channel MOSFET: a *negative* gate terminal attracts holes to the P-type channel to make it more conductive; a *positive* gate terminal repels holes from the P-type channel to make it less conductive.

This is the "field effect" after which the transistor is named: an electric field transverse to a thin channel of semiconductive material affects the number of charge carriers residing in that channel.

[8] The details of this are complicated. According to Gauss's Law, no net electric field can exist within a conductive material, since any net field would motivate charge carriers within that material to move, and as soon as enough charge carriers became displaced they would reach an equilibrium condition where their own self-generated electric field would become equal and opposite to the externally-imposed field. Doped (extrinsic) semiconductors are electrically conductive, of course, and so Gauss's Law tells us any externally-imposed field from the gate terminal will cause electrons and holes to take up new positions within the body of the MOSFET to counteract the externally-imposed electric field. The result is that the placement of electrons and holes within the MOSFET may be manipulated electrostatically by the gate, and this is what gives us control over the MOSFET's source-drain conductivity.

A strong enough electric field between gate and substrate can even force charge carriers to take residence within oppositely-doped regions where they would not naturally exist. This phenomenon is called *inversion*, and it finds application in another style of MOSFET where no channel pre-exists between the source and drain terminals:

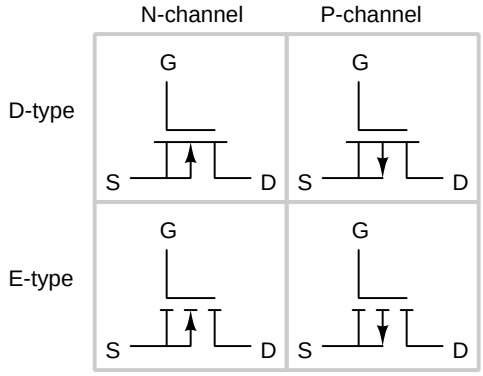


The broken line representing the incomplete channel within these MOSFETs is "made complete" by applying a sufficient gate-substrate voltage with the gate polarized to attract the desired type of charge carriers into the channel. By attracting those carriers toward the gate (mostly from the heavily-doped source and drain regions), some of the semiconductor material lying in between the source and drain regions is actually *forced* to become the other doping type. This "inversion layer" bridges the source and drain together to form a common pathway between those two terminals.

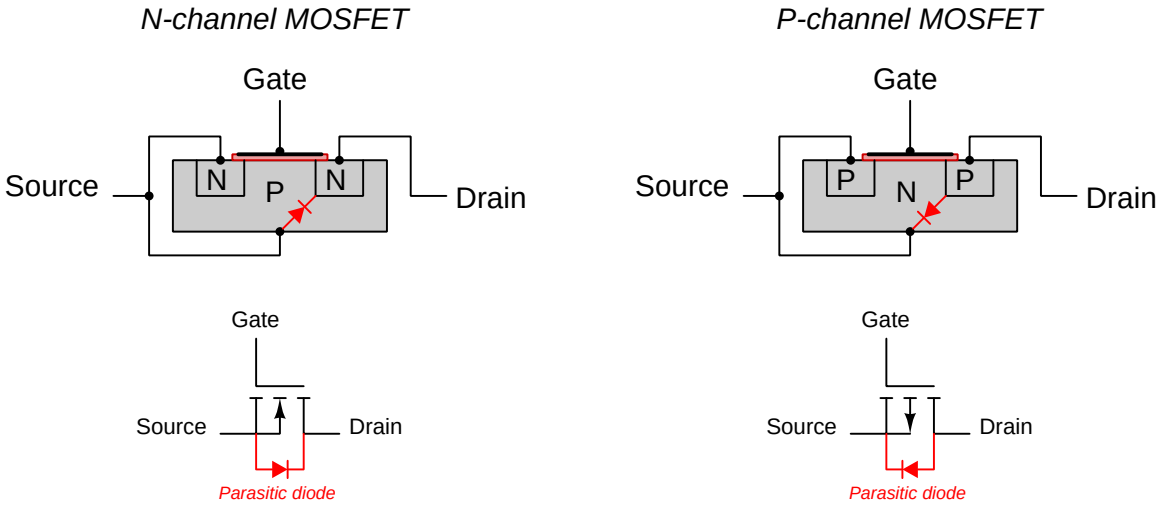
For the N-channel MOSFET the required gate-substrate voltage polarity is a positive gate and negative substrate to attract electrons toward the gate and form an N-type inversion up against the bottom surface of the insulating layer. For the P-channel MOSFET it means a negative gate and positive substrate to attract holes toward the gate to form a P-type inversion layer in the same place. Unless and until a sufficient voltage is applied with the correct polarity to "invert" the channel so it behaves as the opposite doping type, the source and drain regions remain unconnected and the transistor remains in its naturally "off" state.

A "normally-off" MOSFET is technically known as an *enhancement-mode* or *E-type* MOSFET because they must be "enhanced" by an applied gate-substrate voltage in order to conduct at all. The previous style of MOSFET with the pre-existing channel is technically known by the label *depletion-mode* or *D-type* MOSFET, because they are normally conductive and must be "depleted" by an applied voltage in order to turn off.

Although the MOSFET is fundamentally a four-terminal device, they are often constructed with the substrate and source terminals connected by wire, which means the controlling voltage must be impressed between the gate and source terminals:



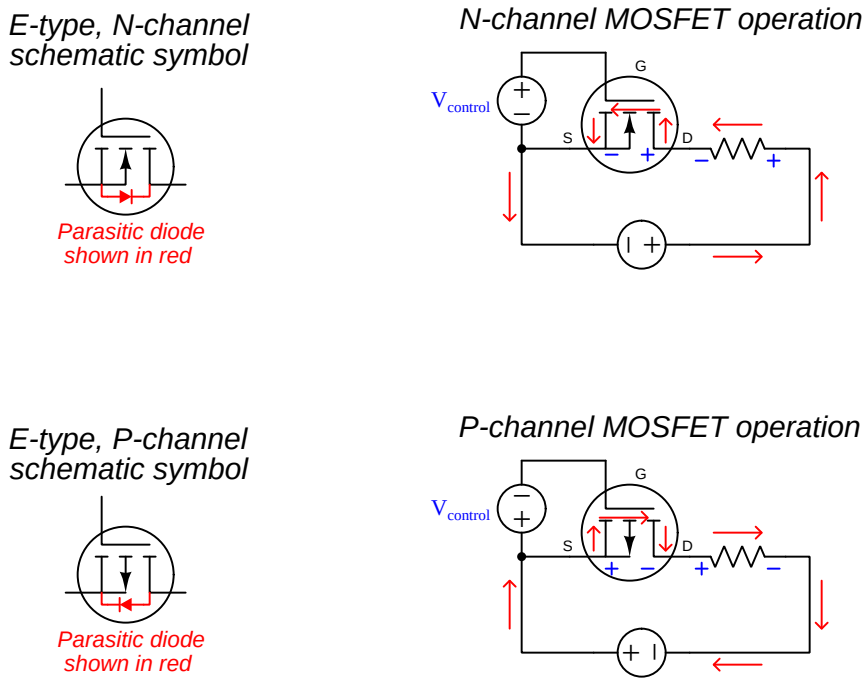
An interesting consequence of this internal connection is that it results in a single PN diode-like parasitic junction connected between the source and drain terminals:



When the source and drain terminals are properly biased[9], this parasitic diode will be reverse-biased and have no effect on the transistor's operation. However, if source-drain polarity ever reverses, this diode will become forward-biased and short past the transistor's channel!

[9]The arrow inside of a MOSFET symbol actually shows you which way this parasitic diode points! For any PN junction the tip of the arrowhead points toward the "N" side and the wide end of the arrowhead is on the "P" side. Note how the wide end of the N-channel MOSFET's arrow connects to the source terminal, as does the anode of the parasitic diode; likewise the tip of the P-channel MOSFET's arrowhead connects to the source terminal, as does the cathode of the parasitic diode. Your task in deciding how to connect any MOSFET into a circuit is to keep this parasitic diode reverse-biased so that the transistor will not be bypassed by it.

Shown below are diagrams of E-type N-channel and P-channel transistors, respectively, being biased into a state of conduction by an applied gate-source voltage labeled as $V_{control}$:



Neither transistor will permit current through its drain or source terminals unless and until this controlling voltage – usually on the order of just a few Volts – is impressed between the gate and substrate, forming an electric field across the insulating oxide layer separating the metal gate from the channel. Note how zero current passes through the gate terminal, nor through the $V_{control}$ voltage source, but only through the source and drain terminals.

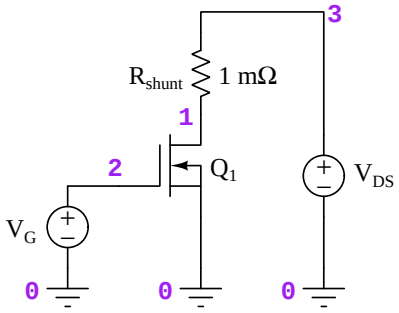
Note also how these three-terminal MOSFETs contain a parasitic diode junction between source and drain terminals[10], and therefore we must purposely arrange the controlled-current circuit such that the direction of current is always *against* this parasitic diode's arrow. This is very important, for if we happened to reverse the polarity of the controlled-circuit source we would find that current would *always* flow in that circuit regardless of the MOSFET's on/off status, the parasitic diode shunting current past the MOSFET channel even when that channel turns off. Remember that the parasitic diode's presence is an unfortunate fact and not an intended feature of the MOSFET's internal structure, which means we must always *work around it* to ensure its presence does not cause unintended problems.

[10]Lest one conclude that the act of making the source and substrate terminals electrically common *creates* this undesired PN junction and therefore we might be better served leaving source and substrate as electrically distinct terminals, consider the fact that even in the four-terminal MOSFET parasitic PN junctions exist: one between substrate and drain, and another between substrate and source. The internal source-substrate jumper actually *bypasses* one of those PN junctions, leaving the other untouched. Parasitic PN junctions are *unavoidable* in MOSFETs!

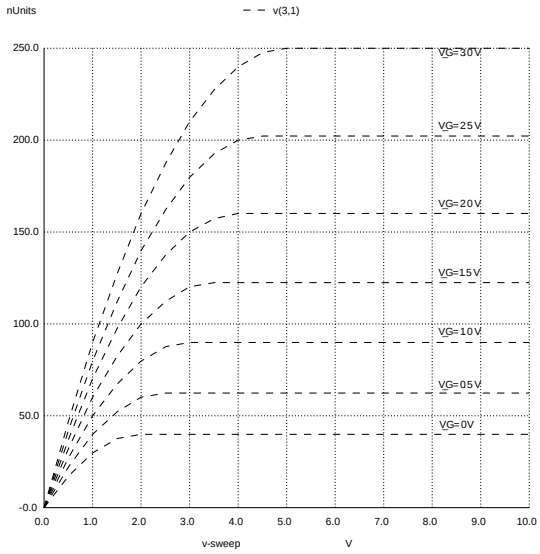
4.4 FET characteristics

One of the most common FET transistor amplifier circuit configurations is the *common-source*, so-called because the source terminal is shared in common by both the DC power supply and the input signal source. This circuit configuration serves well to experimentally explore transistor characteristics, allowing us to vary the input signal and drain-source voltage independently of each other. A shunt resistor connected in series with the transistor's drain terminal allows us to take current measurements using a voltmeter.

If we set the gate voltage value at constant levels while varying the drain-to-source voltage (V_{DS}), we notice some interesting behavior. The transistor's drain current rises rapidly as V_{DS} increases from zero, then the drain current levels off as V_{DS} rises to its full value. It is as though the transistor acts to limit drain current beyond a certain value specified by V_G. In the following SPICE analysis we set V_G to seven different values (0 V, 0.5 V, 1.0 V, 1.5 V, 2.0 V, 2.5 V, and 3.0 V) while varying V_{DS} from 0 to 10 Volts, to achieve a "family" of seven curves representing I_D as a function of V_{DS}:




```
* SPICE netlist
vds 3 0 dc
vg 2 0 dc
rshunt 3 1 1e-3
m1 1 2 0 0 ideal
.model ideal nmos vto=-2
.dc vds 0 10 0.5 vg 0 3 0.5
.plot dc v(3,1)
.end
```

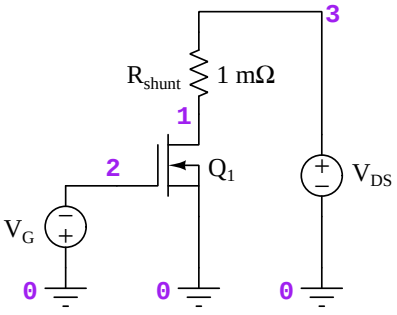


These characteristic curves reveal two different behavioral modes of the FET: one where it acts almost like a *resistor* (very nearly following Ohm's Law) and another where it acts as a *current regulator*. The current-regulating portion of each curve is where it is flat, indicating a constant drain current despite a varying drain-source voltage. The resistive (or "ohmic") portion of each curve is where it resembles an angled line[11], and for most of the seven curves this takes place between 0 and 1 Volt V_{DS}. Both operational modes are useful for different applications: if we wish for the FET to behave as a voltage-controlled resistance, we must ensure it operates with relatively low V_{DS}; if we need the FET to regulate current, we supply it with ample V_{DS}.

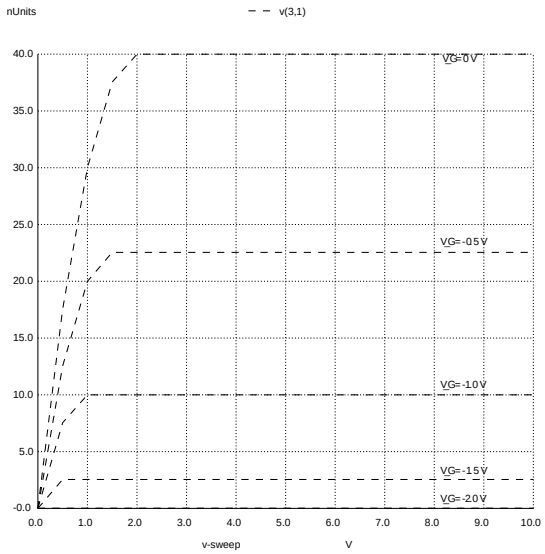
[11] According to Ohm's Law, voltage is proportional to current so long as resistance is constant. Proportional functions are mathematically *linear* and always appear as straight, slanted lines when graphed.

Further examination of the FET's characteristic curve family reveals its normally-half-on nature, as seen on the lowest of the seven curves where $V_G = 0$ Volts. Here we see the transistor able to conduct drain current with no excitation at all applied to the gate terminal. This, of course, is what we would expect of either a JFET or a D-type MOSFET. If we would wish to halt current through the FET using gate voltage, we would have to reverse the polarity of V_G and increase its value until we reached the *pinch-off* voltage (also known as the *threshold voltage*).

This next SPICE simulation tests the same transistor for this pinch-off voltage value. Here, we reverse the polarity of V_G and step it from 0 Volts to -3.0 Volts in -0.5 Volt steps while repeatedly sweeping V_{DS} from 0 to 10 Volts:




```
* SPICE netlist
vds 3 0 dc
vg 0 2 dc
rshunt 3 1 1e-3
m1 1 2 0 0 ideal
.model ideal nmos vto=-2
.dc vds 0 10 0.5 vg 0 3 0.5
.plot dc v(3,1)
.end
```



What we see is another "family" of characteristic curves, this time each curve becoming lower (i.e. less drain current) as V_G intensifies. All curves from $V_G = -2$ V to $V_G = -3$ V are flat at zero current because the transistor has been "pinched off".

An examination of the SPICE netlist code reveals why the computer understands the pinch-off voltage to be at -2 Volts: it is one of the parameters specified in the MOSFET model (`vto = -2`). V_{TO} technically refers to *turn-on voltage* (also known as *threshold voltage*) which is the gate voltage required to force an E-type MOSFET to begin conducting. The developers of SPICE made no distinction between D-type and E-type MOSFETs, so the only difference between these two transistor models is this one parameter. For a D-type MOSFET that is normally-on and requires a "reverse" voltage to turn off, we set V_{TO} to a negative value (i.e. we regard the "turn-on" voltage as the pinch-off voltage). For an E-type MOSFET we would set V_{TO} as a positive value, and this would require a "forward" voltage to get the FET to activate and allow drain current.

This is not just an idiosyncrasy of SPICE, but rather reveals a fundamental characteristic of depletion-mode versus enhancement-mode MOSFETS: they differ behaviorally only in the threshold voltage required to force them to turn off or turn on, respectively. The modulation of a D-type

MOSFET's depletion region and the creation of an E-type MOSFET's inversion region are merely two phases along a continuum of general FET behavior.

The measure of any field-effect transistor's ability to control current by way of a voltage signal is a factor called *transconductance*, labeled g:

$$g = \frac{\Delta I_D}{\Delta V_{GS}}$$

Where,
g = Transconductance, in Siemens
ΔI_D = Change in drain current, Amperes
ΔV_{GS} = Change in gate-source voltage, Volts

Both the name of g ("conductance") and its appropriate unit of measurement (Siemens) reflects the ratio of current to voltage. While voltage divided by current always yields a resistance expressed in *Ohms*, current divided by voltage always yields a conductance[12] expressed in *Siemens*. In this case, the quantity is *trans*conductance because it is a parameter controlled by the transistor.

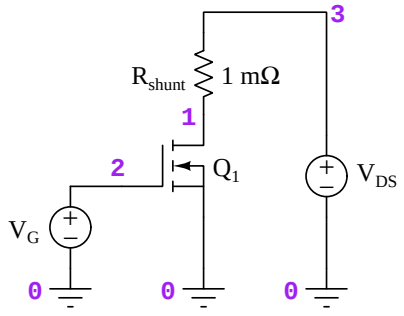
Transconductance is defined by *changes* in I_D and V_{GS} rather than by absolute values of current and voltage because these two variables are not proportional to one another. The reason for this becomes clear when we reflect on the meaning of *pinch-off voltage* and *turn-on voltage*: these are the non-zero values for V_G at which I_D should equal zero. Since one of these two variables is capable of being zero while the other is simultaneously non-zero, they cannot be proportional to each other[13]. Not only is transconductance a non-proportional parameter, but it also varies widely with operating conditions. Datasheets published on field-effect transistors confirms this fact, that transconductance varies widely[14] as a function of gate-source voltage, as a function of temperature, and as a function of drain current.

[12] Transconductance for a field-effect transistor is analogous to β or h_{FE} for a bipolar junction transistor. However, with BJTs we could rightly refer to the ratio of collector current to base current as the transistor's *current gain* whereas with FETs we cannot properly call the ratio of drain current to gate-source voltage as any kind of gain. *Gain* is by definition a unitless ratio, being the quotient of two terms bearing the same units of measurement (e.g. Volts and Volts, Watts and Watts, Amperes and Amperes, etc.). Here, however, the output (controlled) and input (controlling) quantities have different units of measurement and so when we compute their ratio we obtain a value in Siemens rather than a value having no unit. This is also why transconductance must be expressed as g (using the letter G as with any conductance) rather than as A (the standard variable-letter used for gain).

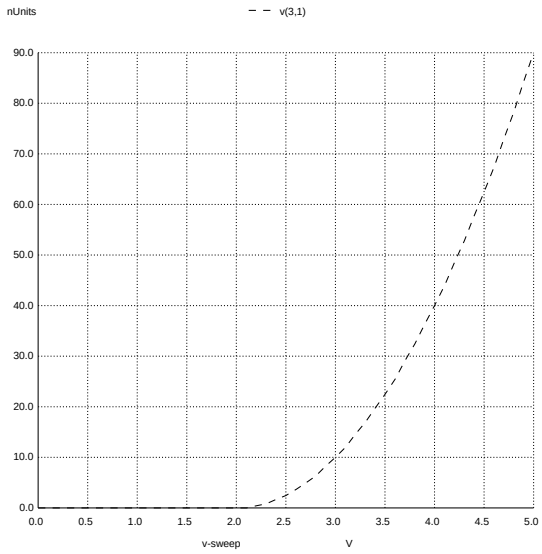
[13] A mathematically rigorous way to say this is that I_D and V_{GS} do not comprise a strictly *linear* function (i.e. $y = mx$), although they could comprise an *affine* function (i.e. $y = mx + b$).

[14] This fact does not, as one might suspect, doom field-effect transistors as dependable amplifying elements. It just means we need to apply a powerful concept called *negative feedback* in our amplifier circuit designs in order to achieve stable behavior.

If transconductance were constant for any FET, we would expect to see the plotted relationship between drain current and gate-source voltage form a straight line. However, when we perform this experiment (shown here as a SPICE simulation where V_{DS} is a constant 15 Volts and V_G "sweeps" from zero Volts to 5 Volts), the result is clearly not linear:




```
* SPICE netlist
vds 3 0 dc 15
vg 2 0 dc
rshunt 3 1 1e-3
m1 1 2 0 0 ideal
.model ideal nmos vto=2
.dc vg 0 5 0.1
.plot dc v(3,1)
.end
```



Drain current is negligible while V_G is less than the turn-off value of 2 Volts, but climbs rapidly as V_G exceeds that threshold value. Examining the plot very closely, we see that when V_G exceeds V_{TO} by 1 Volt the drain current is 10 microAmperes, and that when we double that difference (i.e. V_G exceeds V_{TO} by 2 Volts) the drain current *quadruples* to a value of 40 microAmperes. This is not just nonlinear, but it is specifically a *square* relationship: drain current is proportional to the square of gate-source voltage above the threshold. Represented as a mathematical proportionality:

$$I_D = k(V_G - V_{TO})^2$$

Where,

I_D = Drain current, Amperes
k = Constant of proportionality
V_G = Gate-to-source voltage, Volts
V_{TO} = Threshold ("turn-on") voltage, Volts

For this reason, we say field-effect transistors (of all types) are *square-law* devices. In many applications we would prefer a transistor behave linearly, but in other applications the square-law behavior of FETs is rather useful[15].

[15] For example, some electronic communication systems use FETs to non-linearly "mix" AC signals of different frequencies together in such a way that the result is not just the sum of the two signals but rather contains the *product* of the two signals. This effective multiplication of one signal by the other would be impossible if the transistor behaved linearly.

4.5 Static sensitivity

A very important practical limitation of field-effect transistors with insulated gates (i.e. MOSFETs) is their vulnerability to high voltage impressed across the extremely thin silicon-dioxide insulating layer separating gate and channel. That insulating layer (comprised of oxidized metal) must be thin in order to present a strong electric field[16] across the transistor's channel using modest gate-to-substrate voltage levels. Oxide layer thickness for a typical MOSFET is measured in the *tens of nanometers*. For a sense of perspective, consider the fact that the wavelength of visible light ranges from 400 to 700 nanometers – approximately ten times the thickness of a typical MOSFET's oxide layer. This extreme thinness of the insulating oxide layer makes it easily damaged by excessive voltage[17].

One ubiquitous source of voltage capable of severely damaging MOSFETs is *static electricity*, and the threat to static-sensitive circuits is often referred to as *electrostatic discharge* or *ESD*. For this reason, MOSFETs and MOSFET-based integrated circuits must be handled with care when isolated from circuits[18]. Precautionary measures include *grounding straps* and anti-static smocks worn by people working with such components: the straps make contact with both the person's skin and some Earth-grounded surface, to prevent an electrostatic charge from accumulating[19]. If the work area floor material is not naturally conductive (e.g. concrete), then electrically conductive mats or other surface coverings such as conductive carpet may be applied and bonded to Earth ground. Air nozzles and other machinery capable of accumulating static-electric charges must similarly be bonded to Earth.

[16] The intensity of any electric field (E) is the amount of voltage between the field's end-points (V) divided by the distance between those points (d). Thus, $E = \frac{V}{d}$ and is measured in units of *Volts per meter*. MOSFET conductivity is a function of this electric field strength, and so oxide layer thickness directly affects a MOSFET's threshold voltage.

[17] Absolute maximum gate-source voltage (V_{GS}) ratings for typical discrete MOSFETs is fairly low, with 20 Volts (continuous) being typical.

[18] When installed in circuits, the passive components (resistors, inductors, and diodes in particular) connected to the MOSFET device(s) usually provide enough resistive pathways to shunt any ESD the circuit may be exposed to. When static electricity "discharges" to an object, that object experiences a brief transient of current. So long as that current takes a low-resistance pathway, the voltage it drops along that resistance will not be sufficient to damage the extremely thin insulating layers inside MOSFET devices.

[19] The electrical connection made between the person and Earth ground is not direct, but usually resistive in the mega-Ohm range. Direct low-resistance bonding with Earth would actually pose an electrical hazard to the wearer: an increased probability of dangerous electric shock should that person make contact with an ungrounded ("hot") power conductor. The mega-Ohm resistance pathway offered by an anti-static strap is high enough to limit shock current to levels too low to be threatening to the person, yet low enough to quickly bleed static charges accumulated from foot-shuffling or other bodily motions against insulating surfaces.

Chapter 5

Historical References

This chapter is where you will find references to historical texts and technologies related to the module's topic.

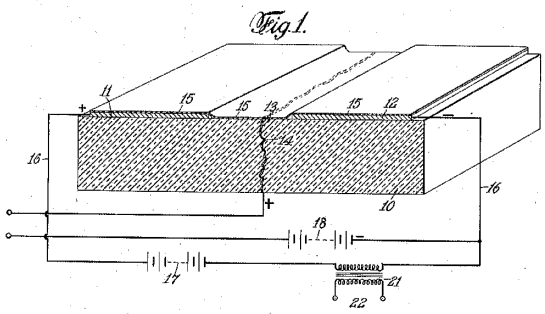
Readers may wonder why historical references might be included in any modern lesson on a subject. Why dwell on old ideas and obsolete technologies? One answer to this question is that the initial discoveries and early applications of scientific principles typically present those principles in forms that are unusually easy to grasp. Anyone who first discovers a new principle must necessarily do so from a perspective of ignorance (i.e. if you truly *discover* something yourself, it means you must have come to that discovery with no prior knowledge of it and no hints from others knowledgeable in it), and in so doing the discoverer lacks any hindsight or advantage that might have otherwise come from a more advanced perspective. Thus, discoverers are forced to think and express themselves in less-advanced terms, and this often makes their explanations more readily accessible to others who, like the discoverer, comes to this idea with no prior knowledge. Furthermore, early discoverers often faced the daunting challenge of explaining their new and complex ideas to a naturally skeptical scientific community, and this pressure incentivized clear and compelling communication. As James Clerk Maxwell eloquently stated in the Preface to his book *A Treatise on Electricity and Magnetism* written in 1873,

> It is of great advantage to the student of any subject to read the original memoirs on that subject, for science is always most completely assimilated when it is in its nascent state . . . `[page xi]`

Furthermore, grasping the historical context of technological discoveries is important for understanding how science intersects with culture and civilization, which is ever important because new discoveries and new applications of existing discoveries will always continue to impact our lives. One will often find themselves impressed by the ingenuity of previous generations, and by the high degree of refinement to which now-obsolete technologies were once raised. There is much to learn and much inspiration to be drawn from the technological past, and to the inquisitive mind these historical references are treasures waiting to be (re)-discovered.

5.1 Early Field-Effect Transistor

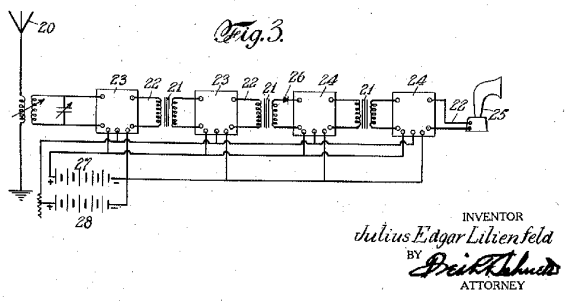
The phenomenon of controlling an electric current through a solid substance by means of a transverse electric field actually was discovered decades before a practical device using that principle could be manufactured. Julius Lilienfeld, an Austro-Hungarian researcher, documented his discovery in a US patent (number 1,745,175) applied for in early October of 1926, with the patent granted late January of 1930. Figure 1 from Lilienfeld's patent shows the construction of his invention:



Within this figure we see a glass substrate (10), two aluminum-foil electrode surfaces (11 and 12), and a thin film comprised of a chemical compound such as copper sulfide (15) which acts as a crude semiconductor. The glass is intentionally fractured (14) and a thin piece of aluminum foil sandwiched between the reassembled glass blocks with its edge adjacent to the copper sulfide film (13). An electric field established between the edge of this foil (13) and the perpendicular copper sulfide film (15) is able to influence the end-to-end resistance of that film as sensed between the end-electrodes (11 and 12).

The amplification circuit shown uses a DC voltage source (17) to provide power, another DC voltage source (18) to provide the necessary bias voltage for successful AC operation, and a transformer (21) to couple the amplified AC signal to a load. The input AC voltage signal is applied to the two terminals on the left-hand edge of the illustration.

Lilienfeld imagined several of these "electric valves" arranged in stages to serve as amplifiers for a radio receiver circuit, as shown in Figure 3 of the patent:



The inventor begins the text of his patent application by contrasting his invention against the common vacuum-tube technology of the time:

> The invention relates to a method of and apparatus for controlling the flow of an electric current between two terminals of an electrically conducting solid by establishing a third potential between said terminals; and is particularly adaptable to the amplification of oscillating currents such as prevail, for example, in radio communication. Heretofore, thermionic tubes or valves have been generally employed for this purpose; and the present invention has for its object to dispense entirely with devices relying upon the transmission of electrons thru an evacuated space and especially to devices of this character wherein the electrons are given off from an incandescent filament. The invention has for a further object a simple, substantial and inexpensive relay or amplifier not involving the use of excessive voltages, and in which no filament or equivalent element is present. More particularly, the invention consists in affecting, as by suitable incoming oscillations, a current in an electrically conducting solid of such characteristics that said current will be affected by and respond to electrostatic changes. Means are associated with the aforesaid conducting solid whereby these electrostatic changes are set up conformably with the incoming oscillations which are thus reproduced greatly magnified in the circuit, suitable means being provided, also, to apply a potential to the said conducting solid portion of the amplifier circuit as well as to maintain the electrostatic producing means at a predetermined potential which is to be substantially in excess of a potential at an intermediate point of said circuit portion. `[page 1]`

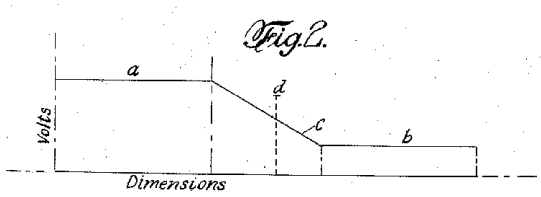
Later on the same page, Lilienfeld begins to describe features of the device in detail:

> Referring to the drawings, 10 designates a base member of suitable insulating material, for example, glass; and upon the upper surface of which is secured transversely thereof and along each side. a pair of conducting members 11 and 12 as a coating of platinum, gold, silver or copper which may be provided over the glass surface by well known methods such as chemical reduction, etc. It is desirable that the juxtaposed edges of the two terminal members 11 and 12 be located as closely as possible to each other; and substantially midway of the same there is provided an electrode member 13, which is of minimum dimensions to reduce capacity effect. This member consists of a suitable metal foil, preferably aluminum foil, and may conveniently be secured in position by providing a transverse fracture 14 in the glass and then reassembling the two pieces to retain between the same the said piece of aluminum foil of a thickness approximating one ten-thousandth part of an inch. The upper edge of this foil is arranged to lie flush with the upper surface of the glass 10.
>
> Over both of the coatings 11 and 12, the intermediate upper surface portion of the glass 10, and the edge of the foil 13 is provided a film or coating 15 of a compound having the property of acting in conjunction with said metal foil electrode as an element of unidirectional conductivity. That is to say, this coating is to be electrically conductive and possess also the property, when associated with other suitable conductors, of establishing at the surface of contact a considerable drop of potential. The thickness of the film, moreover, is minute and of such a degree that the electrical conductivity therethru would be influenced by applying thereto an electrostatic force. A suitable material for this film

and especially suitable in conjunction with aluminum foil, is a compound of copper and sulphur. A convenient way of providing the film over the coatings [page 1]

11 and 12 and the electrode 13 is to spatter metallic copper by heating copper wire with in a vacuum, or by depositing copper from a colloidal suspension, over the entire upper surface and then sulphurizing the deposited copper in sulphur vapor, or by exposure to a suitable gas as hydrogen sulphide or a liquid containing sulphur, as sulphur dissolved in carbon bisulphide. [page 2]

Next, the author describes how his apparatus might be connected to other components to form an amplifying circuit, making reference to Figure 2 (shown below):



To produce the required flow of electrons through the film 15 a substantial potential is applied across the two terminal coatings 11 and 12 as by conductors 16 leading from a battery or like source 17 of direct current. As shown in the diagrammatic view, Fig. 2, the dimensional volt characteristics of the device indicate a substantially steady voltage of value a over the coating 11 and a corresponding steady voltage b of diminished value over the coating 12, while over the portion of the surface between said coatings the voltage in the film 15 will be according to the gradient c. As aforesaid, the electrode 13 is located substantially midway of the inner ends of the terminal coatings 11 and 12 and there is arranged to be supplied thereto a potential indicated by the valued, Fig. 2, and somewhat in excess of the voltage prevailing along the gradient c at this point. This potential may be applied by means of a battery or like source of potential 18, the negative pole of which is connected to the negative pole of the battery 17. In the circuit of the electrode 13 and source of potential 18 is also included some exterior source of oscillating or fluctuating current, which source is indicated, by way of example, in Fig. 3, as the antenna 20 of a radio communication circuit.

The effect of thus providing an excess positive potential in the electrode 13 is to prevent any potential in the oscillating circuit herein before described from rendering said electrode of zero potential or of a negative potential, which would then permit a current to pass from the electrode edge to the film 15; as in the reverse direction where a positive voltage is maintained, the two members – namely electrode and connecting film – act as an electric valve to prevent the flow. Maintaining a positive potential at this point, however, insures that the flow of the electrons from the piece 11 to the piece 12 will be impeded in a predetermined degree, a variation therein being effected conformably to the changing amount of this potential under the influence of the oscillating or fluctuating current introduced. This effect will be repeated on a greatly magnified scale in the circuit of the conducting coatings 11 and 12 and may be reproduced in various circuits or for various purposes as thru a transformer 21, from the secondary of which leads 22 extend

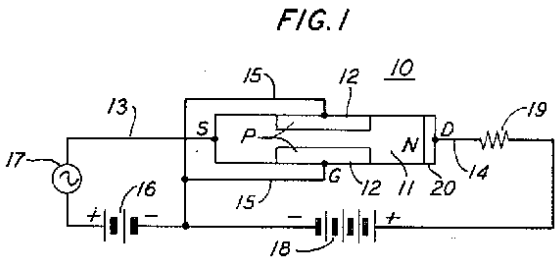
> to any suitable device, which, as shown in Fig. 3, may be further amplifiers of this character as the radio frequency amplifiers 23 and audio frequency amplifiers 24, the last of which is shown connected to a loudspeaker or similar device 25. A current rectifying member 26, however, is necessary where it is desired to convert the radio frequency into audio frequency oscillations. It will be observed that but two sources of potential 27 and 28 – which may be combined into a single, properly tapped source – are required and of potentials approximately 30 and 15 volts respectively for the particular elements employed. [page 2]

It must be understood that at the time of this patent submission, a practical theory of quantum mechanics has not yet been developed to explain electron and hole conduction in solids. As such, Lilienfeld had to speculate as to the principle of his device's operation. The following paragraph is the inventor's attempt to explain *how* his transistor functioned:

> The basis of the invention resides apparently in the fact that the conducting layer at the particular point selected introduces a resistance varying with the electric field at this point; and in this connection it may be assumed that the atoms (or molecules) of a conductor an electron, nature of bipoles. In order for an electron, therefore to travel in the electric field, the bipoles are obliged to become organized in this field substantially with their axes parallel or lying in the field of flow. Any disturbance in this organization, as by heat movement, magnetic field, electrostatic cross-field, etc., will serve to increase the resistance of the conductor; and in the instant case, the conductivity of the layer is influenced by the electric field. Owing to the fact that this layer is extremely thin the field is permitted to penetrate the entire volume thereof and thus will change the conductivity throughout the entire cross-section of this conducting portion. [page 2]

5.2 US Patent for a JFET

George Dacey and Ian Ross were researchers working for Bell Laboratories in the 1950's when they applied for a US patent in late October 1952 entitled "Semiconductor Signal Translating Devices", the patent granted in late January of 1957 (number 2,778,956). This patent is for a junction field-effect transistor (JFET) which we see clearly represented in Figure 1:



Theirs was an N-channel JFET, connected to DC voltage sources, an AC signal source, and a load resistor in such a way as to form a *common-gate* amplifier. The source (S), gate (G), and drain (D) terminals are clearly shown in this figure.

What follows are paragraphs taken from the first page of the patent describing its principle of operation:

> Devices of the type disclosed in the above-identified application comprise, generally, a body of semiconductive material having therein a region or zone of one conductivity type flanked by and contiguous with a pair of zones of the opposite conductivity type. Individual connections, herein termed the source and drain, are made to opposite ends of the first zone, and a third connection, termed the gate herein, is made to the other two zones in common. Both the source and drain are biased relative to the gate so that the PN junctions between contiguous zones are biased in the reverse direction, the potential of the drain, however, being substantially greater than that of the source. Signals are impressed between the source and gate and amplified replicas thereof are obtained in a load or utilization circuit connected between the gate and drain. In effect, the variations in the potential of the gate control the conductivity of the path for the flow of electrical carriers in the intermediate zone from the source to the drain.
>
> One general object of this invention is to enhance the performance of semiconductor signal translating devices of the general type above described. A specific object of this invention is to enable control of the current gain for such devices and, more particularly, to attain current gains greater than unity.
>
> Conductivity in semiconductors such as germanium and silicon is associated with two types of electrical carriers to wit, holes and electrons. In extrinsic material, as is now known, both types of carriers are present but one is in excess of the other. Those in excess are known as majority carriers and those in the minority as minority carriers. In N conductivity type material, the majority carriers are electrons and the minority carriers are holes; in P type material, the majority carriers are holes, and the minority carriers are electrons.

In devices of the type above described, it has been found that the flow of majority carriers from the source to the drain has associated therewith a flow of minority carriers from the drain. It has been found further that the latter is amenable to control and that advantageous results, notably alphas greater than unity and negative resistance characteristics, can be realized by establishing a controlled flow of minority carriers from the drain to the gate. The mechanism involved in minority carrier flow, it has been determined, is dependent upon the drain connection.

In accordance with one broad feature of this invention, in a device of the character under consideration, the drain connection and region are made such as to enhance the flow of minority carriers to the gate. The minority carrier current is controlled by the majority carrier current from the source in such manner that the change in gate voltage to effect an increase in the minority carrier current is of the sign such that the associated dynamic resistance is negative.

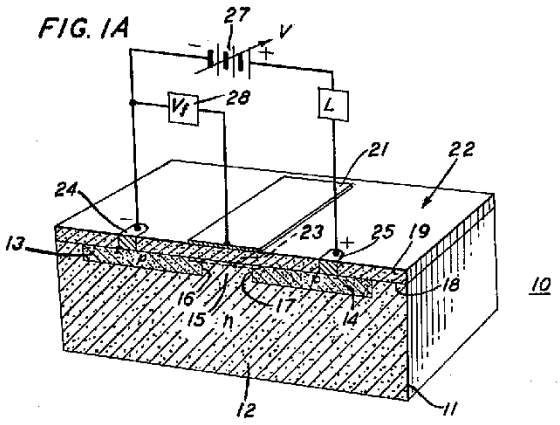
In one illustrative embodiment of this invention, the intermediate zone of the semiconductive body is of N conductivity type and the gate zones are of P type and biased negative relative to both the source and drain. Thus, the majority carriers in the intermediate zone, that is those flowing from source to drain, are electrons, and the minority carriers in this zone are holes. The bias of the gate is such, it will be noted, as to attract the minority carriers thereto from the intermediate zone.

In one specific embodiment, and this illustrates a particular feature of this invention, the drain connection is constructed so as to maintain the minority carrier density in the vicinity thereof at the equilibrium value. This may be effected by establishing a surface or region at or adjacent the drain electrode characterized by low carrier lifetime characteristics, for example by including or introducing into this region an element such as nickel which results in low carrier lifetimes, or producing thereat crystal imperfections as by sandblasting or electron bombardment.

In another specific embodiment, and this illustrates another particular feature of this invention, means are provided to effect injection into the drain region of minority carriers substantially proportional to the majority carriers arriving at the drain. Such injection may be obtained by providing a PN junction at the drain region and controlling the junction bias in such manner that a minority carrier current is injected into the intermediate region, of magnitude dependent upon the drain or load current. For example, the junction may be biased in the forward direction and its potential controlled in accordance with the drop across a resistor in series with the drain. [page 1]

5.3 US Patent for a MOSFET

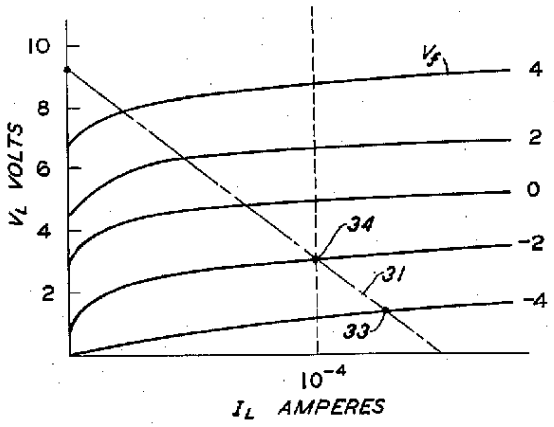
A Korean-born researcher working for Bell Laboratories named Dawon Kahng applied for a patent in late May of 1960 entitled "Electric Field Controlled Semiconductor Device", the patent (number 3,102,230) being granted in late August three years later. In this patent we find a clear description of what we now know as a MOSFET, specifically a depletion-type P-channel MOSFET. Figure 1A of the patent is shown below:



From this illustration we may observe the standard MOSFET features. Built on an N-type semiconductor substrate (12), two thick P-type semiconductor layers comprising source (14) and drain (13) are bridged together by a much thinner layer of P-type material (15). The whole surface of the semiconducting material is covered by an insulating or *dielectric* layer (19), with a metal gate electrode attached on top (21). Two connection points for the source and drain (25 and 24, respectively) penetrate this insulating layer to provide contacts for external source and drain terminals.

The amplifier configuration represented in Figure 1A is *common-drain*, and as a result of this configuration the MOSFET tends to regulate load voltage over a range of load currents (i.e. load voltage remains relatively constant as load resistance varies). This is shown in a set of characteristic curves represented in the patent as Figure 2:

FIG. 2



Note the bias voltage range extending to include values above and below zero, and how load voltage exists even at zero bias. This tells us the transistor conducts current between source and drain in its natural "unstimulated" state, as we would expect for a depletion-type MOSFET.

Introductory paragraphs from the patent explain the operation of this transistor. Note how Kahng credits fellow Bell Laboratories researcher Martin Atalla for developments in applying dielectric layers to semiconductor substrates:

> This invention relates to circuit arrangements including dielectric coated semiconductor devices.
>
> More particularly, the present invention relates to circuit arrangements utilizing a semiconductive device which comprises either a p-n-p or n-p-n wafer having a dielectric film over a portion of the middle zone. Such a device is disclosed in application Serial No. 13,688 of M. M. Atalla, filed March 8, 1960.
>
> In accordance with the present invention useful characteristics are obtained from a device of this type by arranging the associated circuitry to vary an electric field across the oxide in response to variations in voltage across the junctions. In particular, voltage regulation or amplification can be achieved by the invention.
>
> Therefore, a feature of this invention is a novel circuit arrangement which provides across the dielectric layer of the above device an electric field which varies in response to variations in voltage across the p-n junctions.
>
> The invention in its preferred form comprises a semiconductor wafer, typically silicon, including first, second and third regions defining respectively first and second p-n

junctions which intersect a major surface of the wafer. This major surface of the wafer is coated with a suitable dielectric, typically a thermally grown silicon dioxide coating for a silicon wafer, and an electrode is connected to the surface of this oxide coating so as to extend beyond the line of intersection with the surface of the two p-n junctions. A first bias voltage is applied between ohmic contacts to the first and third regions poled to forward bias the first p-n junction and to reverse bias the second. A second bias voltage is applied between the electrode to the oxide coating and the contact to the third region. The electric field across the oxide coating is the result of this second bias and varies in response to variations in the voltage between the two substantially ohmic contacts.

Chapter 6

Derivations and Technical References

This chapter is where you will find mathematical derivations too detailed to include in the tutorial, and/or tables and other technical reference material.

6.1 FETs versus BJTs

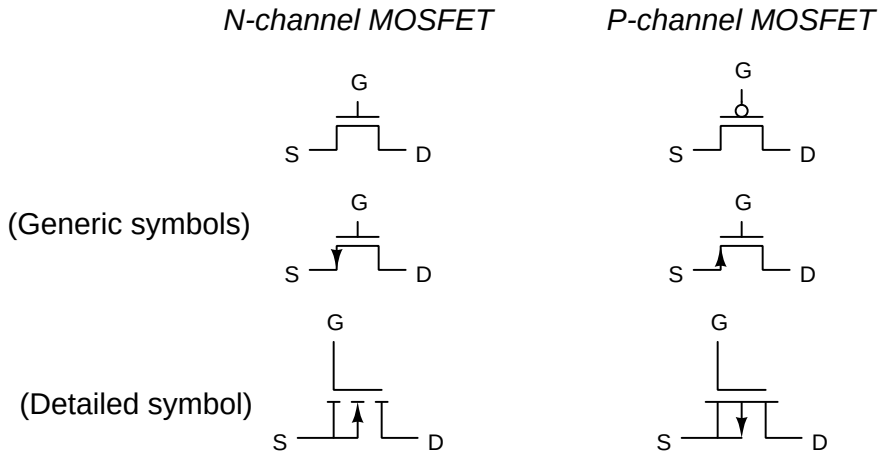
Field-effect transistors (FETs) serve much the same purpose as bipolar junction transistors (BJTs): to act as electrically-controlled throttling elements for electric currents. However, they differ in some significant regards that are worth contrasting:

Characteristic	**FET**	**BJT**
Charge carrier type	Majority	Minority
Principle of operation	Effective width of a continuous channel modulated by voltage	Charge carriers injected into a depleted region
Drive signal	Voltage only (no DC current)	Voltage and current both necessary
Switching time limit	RC time constant of gate capacitance and drive resistance	Stored charges in base waiting to be swept out
Temperature effect	Greater temperature increases channel resistance	Greater temperature decreases effective resistance
Resting status	Normally-on and normally-off types	Normally-off (only)
Gain	Transconductance $g = \frac{\Delta I_{DS}}{\Delta V_{GS}}$	Beta ratio $\beta = \frac{I_C}{I_B}$
Current regulation	Depletion region "pinch" effect intensifies with increasing V_{DS}	Rate of charge injection into base sets a maximum limit on I_C

6.2 Alternative MOSFET symbols

As it so happens, E-type MOSFETs are far more popular than D-type MOSFETs, so much so that MOSFETs are sometimes assumed to be E-type unless otherwise specified. This assumption even extends to symbols used to represent MOSFETs in some schematic diagrams. Confusingly, the E-type is so common that sometimes the broken-line convention is dropped entirely.

Shown below is a set of symbols commonly used to represent E-type MOSFETs. The "generic" symbols are intended to be simple to draw, while the "detailed" symbols are more faithful to the actual construction of the device:

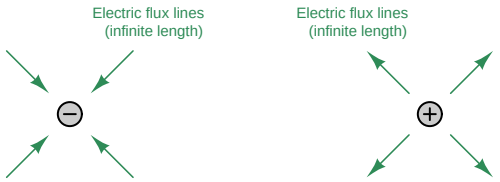


6.3 Gauss' Law of Electric Fields

One of fundamental principles relating to electric fields is *Gauss' Law of Electric Fields*. A common mathematical expression for this law is as follows:

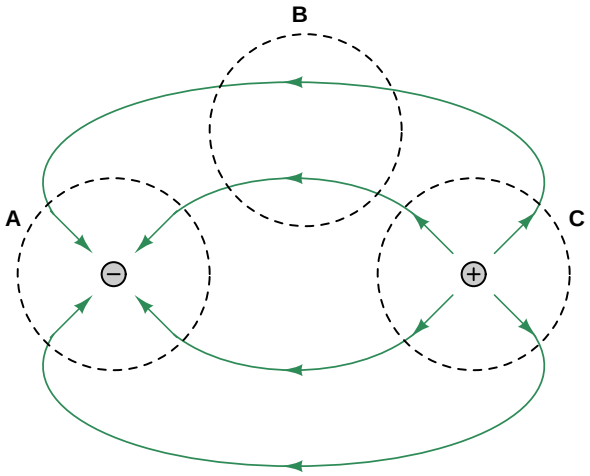
$$\oint \mathbf{E} \cdot d\mathbf{A} = \frac{Q}{\epsilon_0}$$

This equation simply describes electric fields (E) as lines of flux spreading out from a quantity of electric charge (Q). By convention we draw these electric flux lines as radiating away from positive charges and heading toward negative charges as shown below:



If we imagine a three-dimensional surface completely enclosing any electric charge, like a soap bubble surrounding that charge, the total quantity of electric flux lines passing through that surface will be the product of the electric field (also known as the electric flux density, **E**) and the enclosing area (**A**). This "soap bubble" is formally known as a *Gaussian surface*, and we may arbitrarily place that surface anywhere within space for the sake of solving an electric field problem. The major point to understand here is that *any* Gaussian surface drawn will experience an amount of electric flux penetrating its surface proportional to the amount of electric charge enclosed within that surface regardless of the size or shape.

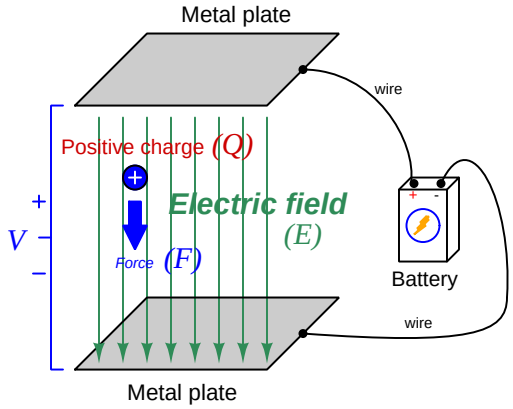
Observe what happens if we connect the electric flux lines together from two opposite charges, and then draw Gaussian surfaces in different regions of space labeled **A**, **B**, and **C**:



Gaussian surface **A** encloses a single negative charge and we see inward-pointing electric flux lines penetrating that surface. Gaussian surface **C** encloses a single positive charge and we see outward-

pointing electric flux lines penetrating that surface. Gaussian surface **B** contains no electric charges at all because it encloses empty space, and we see the number of electric flux lines entering the enclosed volume exactly equaling the number of flux lines leaving it. In other words, for Gaussian surface **B** there is zero net electric flux enclosed within it and also zero *net* penetrations since the number going in balances the number going out.

If we apply an electric field across open space containing a single positive charge, that charge will experience a force proportional to its charge-quantity and the strength of the electric field. The direction of this force will be downward to the negative plate and away from the positive as shown in the following illustration:



$$\vec{F} = Q\vec{E}$$

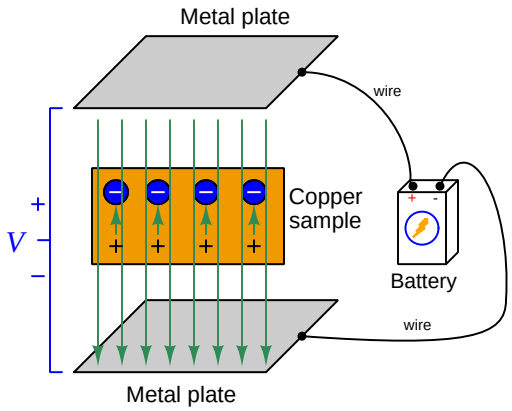
Where,

$\vec{F}$ = Force exerted on the charge (Newtons)
Q = Charge quantity (Coulombs[1])
$\vec{E}$ = Electric field (Newtons per Coulomb)

Unsurprisingly, a negative charge in that same space will experience a force in the opposite direction, upward to the positive plate and away from the negative. With nothing to impede the motion of these charges, they will accelerate as a result of the applied force from the electric field, their acceleration predicted by the familiar equation $F = ma$.

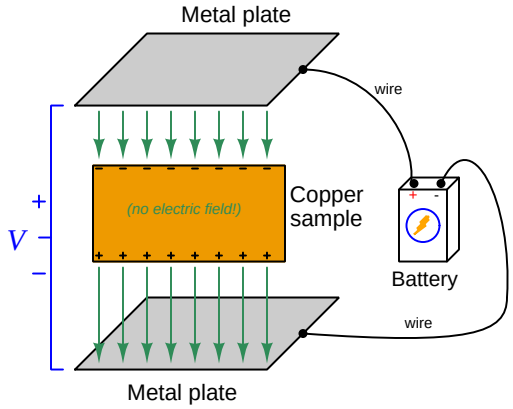
[1] One Coulomb of electric charge is equal to 6.2415×10^{18} electrons.

Now imagine a solid piece of electrically conductive material such as copper metal suspended between the same two metal plates. The outer-most electrons of the copper atoms are free to drift throughout the metal, which is the fact that makes copper metal a conductive material. The effect of the two plates' electric field on the free electrons within the copper sample will be much the same as upon free electrons in a vacuum space between two metal plates: those free electrons will accelerate upward in the direction of the positive plate and away from the negative. This migration of free electrons toward the upper surface of the copper sample leaves positively-charged regions near the lower surface of the copper:



The electrons will migrate as far as they are able, which in this case is the upper surface of the conductive copper sample. The separation of negative and positive charges within the copper sample – the negative charges being mobile electrons and the positive charges being immobile protons in the copper atom nuclei – itself forms an electric field within the copper, pointed in the direction opposite that of the battery's external field. These short vector-arrows pointed upward in the above illustration represent the flux lines from the positive stationary nuclei to the displaced negative electrons.

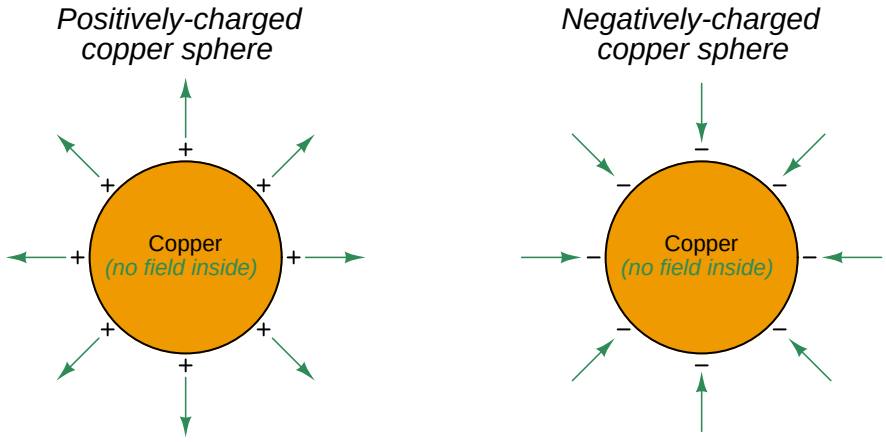
Since these two sets of electric flux lines point in opposite directions, their net effect is to cancel each other out. As more electrons drift upward this cancellation continues until there is zero net electric field within the copper sample. When the net electric field reaches zero, the electrons no longer experience any driving force upward. In this *equilibrium* condition of zero net force on all electric charges, the interior of the copper sample contains no net electric field:



It should not come as a surprise to anyone familiar with basic concepts of electric circuits that this sample of copper naturally excludes any electric field imposed externally upon it, since the mass of copper may be thought of as a large collection of electrically common points, and we know one of the characteristics of electrically common points is equipotentiality. *And, where there is no difference of electrical potential between two points (i.e. no voltage), there can be no electric field between those points.* A conductive object in a state of electrical equilibrium, therefore, contains no internal electric field(s).

This phenomenon of electrical conductors naturally cancelling out internal electric fields finds wide application in electrical and electronic systems. One such application is the *shielding* of sensitive components by surrounding them with a continuous conductive surface. Another application is in insulated-gate field-effect transistors (also known as MOSFETs) where an externally-applied electric field causes electron and hole charge carriers to reposition themselves and in so doing cause a conductive channel transverse to that field to either grow wider or narrower.

An interesting consequence of Gauss' Law of Electric Fields is that any net gain or loss of electric charge by a conductive object results in the unbalanced charges moving to the very exterior surface of the object. For example, if we inject extra electrons into a copper block, those electrons quickly settle into positions on the very outer skin of that block where they remain. Likewise, if electrons are extracted from a copper block, the remaining electrons will migrate within the block so as to leave the resulting positive charges (i.e. electron deficiencies) at the outer surface of that block:



All unbalanced charges migrate to the outside of the object

We may understand this by appealing to the proven fact that a conductive object in a state of electrical equilibrium must have no electric field within its interior. If we imagine drawing a Gaussian surface just beneath the skin of that conductive object, we may conclude from Gauss' Law that the interior of that Gaussian surface must contain no net electric charge (i.e. all positive and negative charges within it are perfectly balanced) because zero net electric field penetrating that Gaussian surface implies zero net electric charge within. This must mean that any and all unbalanced electric charges in that conductive object must reside *outside* the area of no electric field; i.e. outside the Gaussian surface on the conductive object's outer surface. Thus, a Gaussian surface drawn anywhere inside a charged conductive object experiences no electric flux lines but a Gaussian surface drawn outside that same object does.

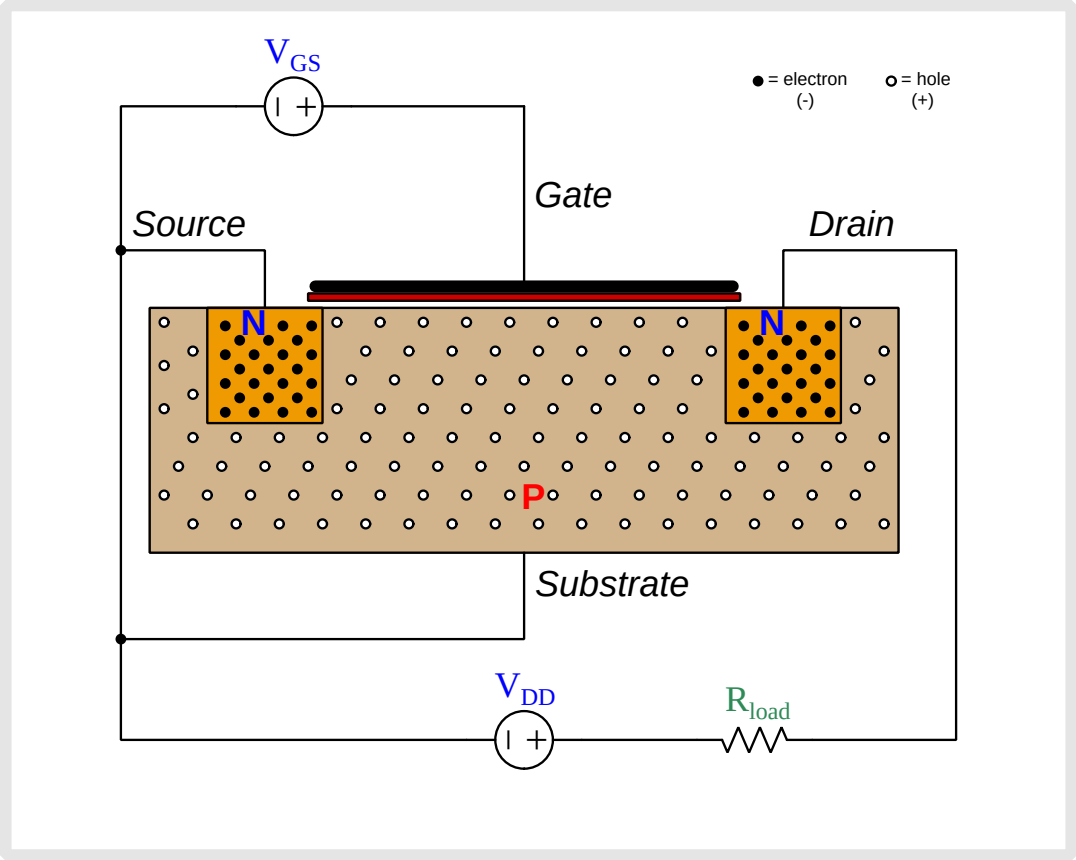
Chapter 7

Animations

Some concepts are much easier to grasp when seen in *action*. A simple yet effective form of animation suitable to an electronic document such as this is a "flip-book" animation where a set of pages in the document show successive frames of a simple animation. Such "flip-book" animations are designed to be viewed by paging forward (and/or back) with the document-reading software application, watching it frame-by-frame. Unlike video which may be difficult to pause at certain moments, "flip-book" animations lend themselves very well to individual frame viewing.

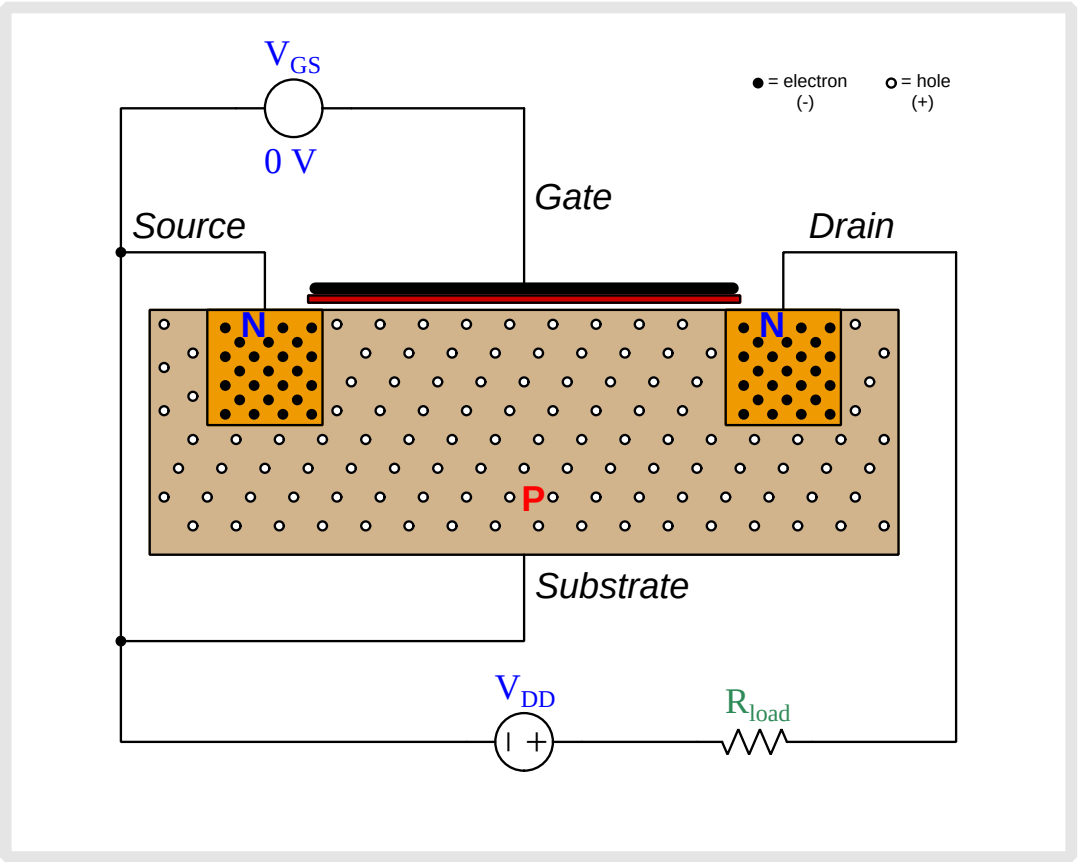
7.1 Animation of an E-type MOSFET turning on

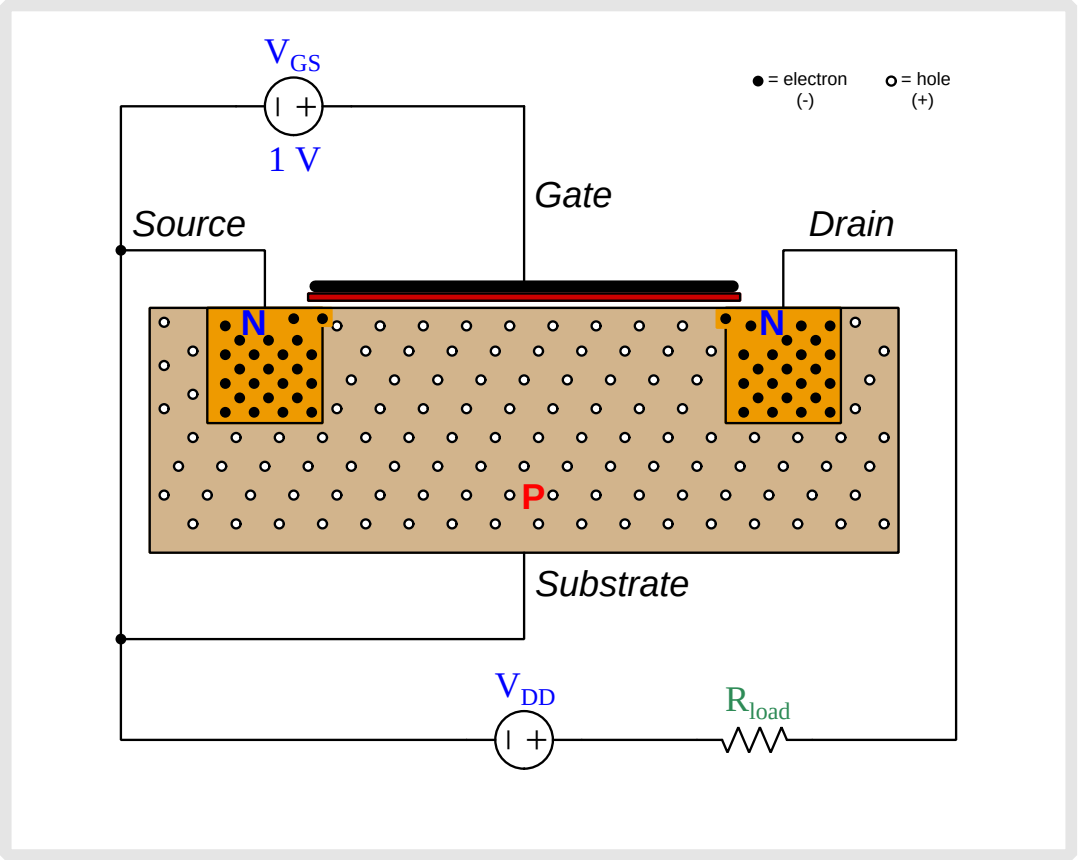
Enhancement-type (E-type) MOSFET transistors are naturally "off" devices, having no continuous channel of conductive semiconductor material joining source and drain by default the way D-type MOSFETs do. Instead the substrate semiconductor material, which is doped oppositely of the source and drain regions, interrupts that path with two series-opposed PN junctions. However, when a transverse electric field is applied between gate and substrate by an external voltage source, the surface-layer of the substrate in direct contact with the insulating metal-oxide layer converts to the opposite type of semiconductor. For an N-channel MOSFET with N-doped source and drain regions and with P-doped substrate, this means the surface of the P-doped substrate just beneath the gate effectively turns into N-type by the force of the electric field drawing free electrons from source and drain (and a few minority electrons existing within the substrate) to that surface region to form a thin film of free electrons capable of supporting source-drain conduction.



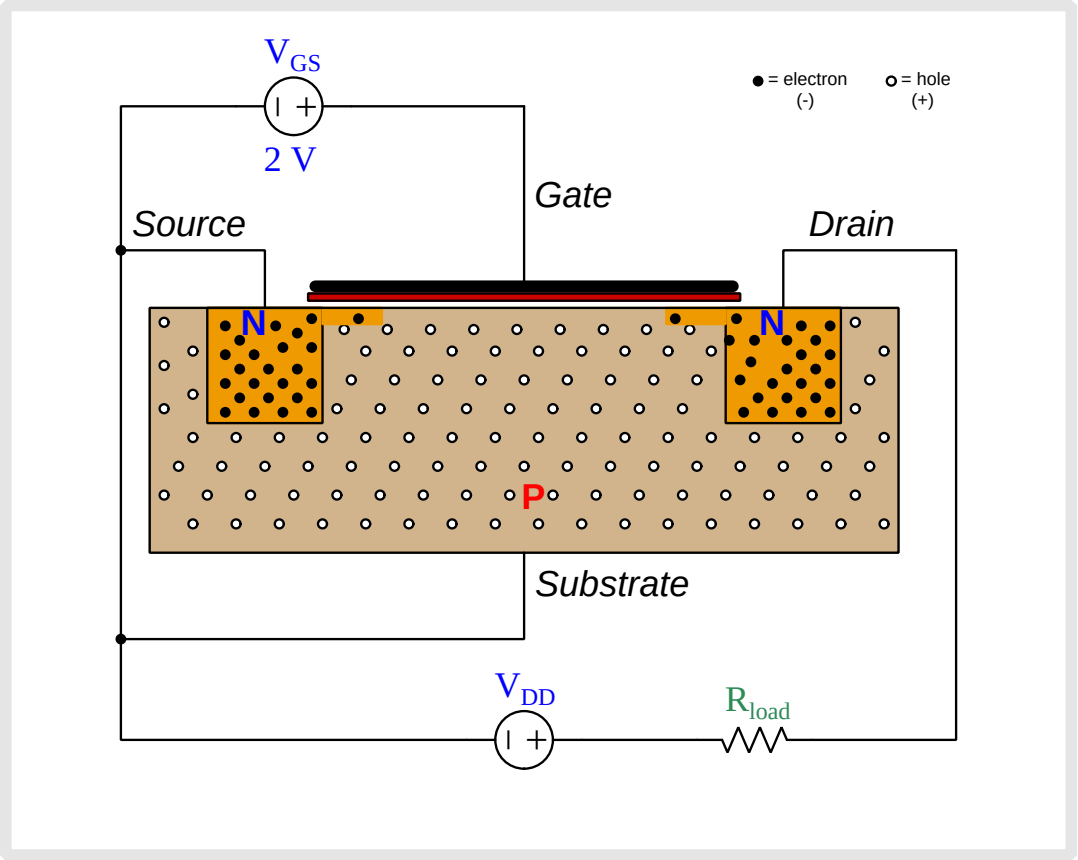
Above we see an N-channel, E-type MOSFET connected to two sources. V_{GS} applies a controlling voltage between the gate and the substrate, with the substrate connected to the source terminal to be one and the same. V_{DD} is the voltage source which will drive current through the load resistor once the transistor begins to conduct.

The N-doped source and drain regions have numerous free electrons (negatively-charged) as mobile charge carriers, while the P-doped substrate has holes (positively-charged). Note that when load current begins to flow, its direction will be shown in *conventional flow notation* which is the same direction in which holes move but is opposite to the direction free electrons move.

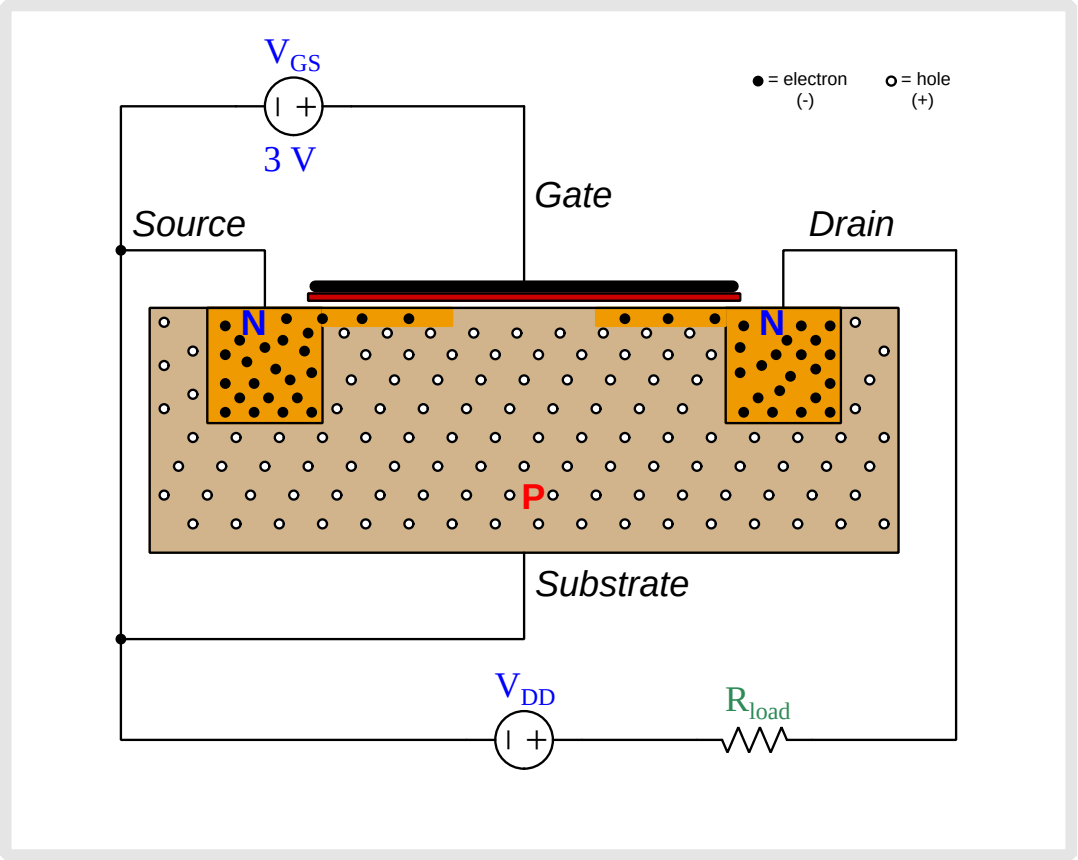

V_GS
0 V
● = electron (-)
○ = hole (+)
Gate
Source
Drain
N
N
P
Substrate
V_DD
R_load



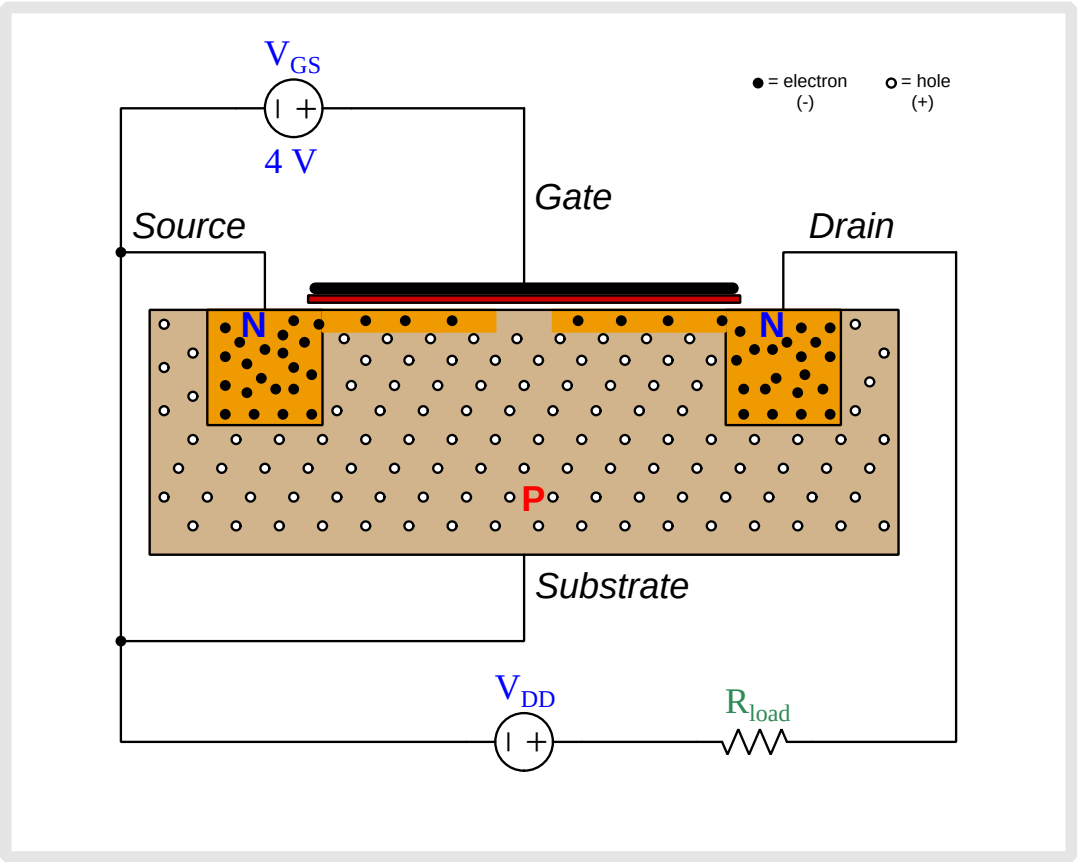
As we begin to apply a gate-to-substrate voltage with the gate positive and substrate negative, negatively-charged free electrons from the source and drain regions are attracted toward the metal gate while positively-charged holes in the substrate are repelled from the gate and driven downward.



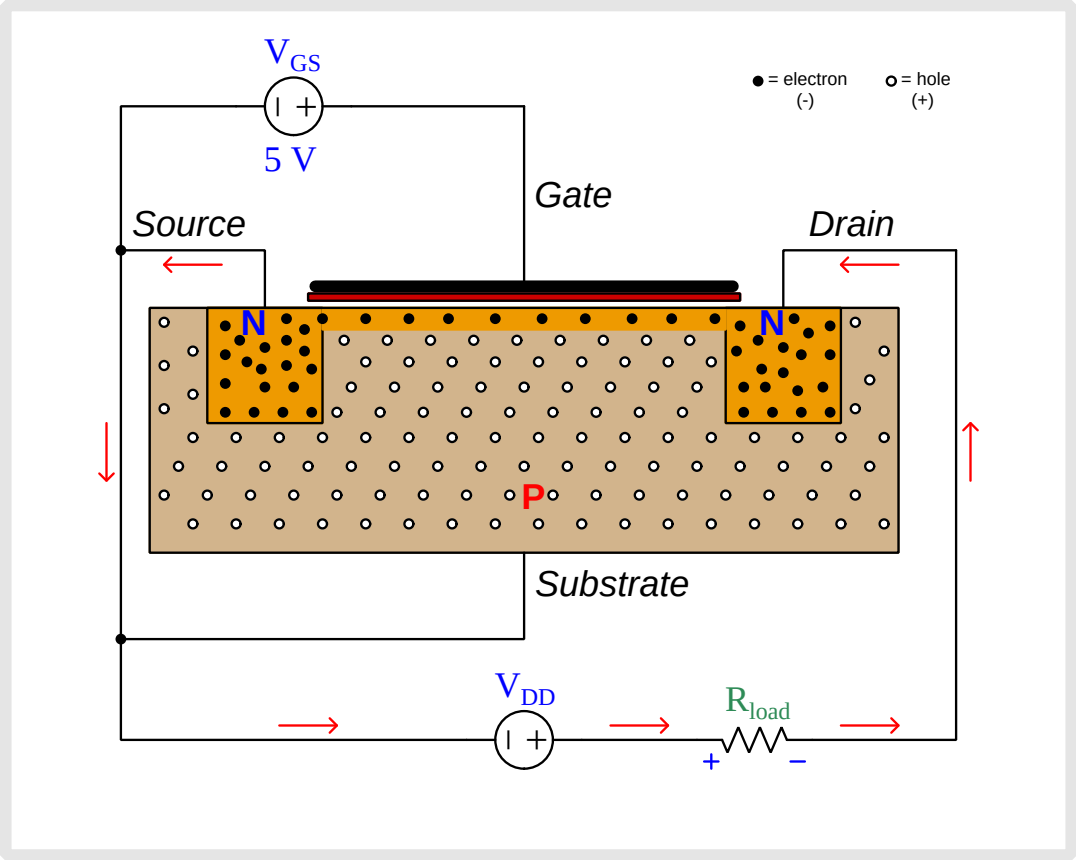
As this V_{GS} voltage increases, a thin film of free electrons begins to form beneath the gate, effectively turning what used to be P-type semiconductor material into N-type. This is called an *inversion layer*.



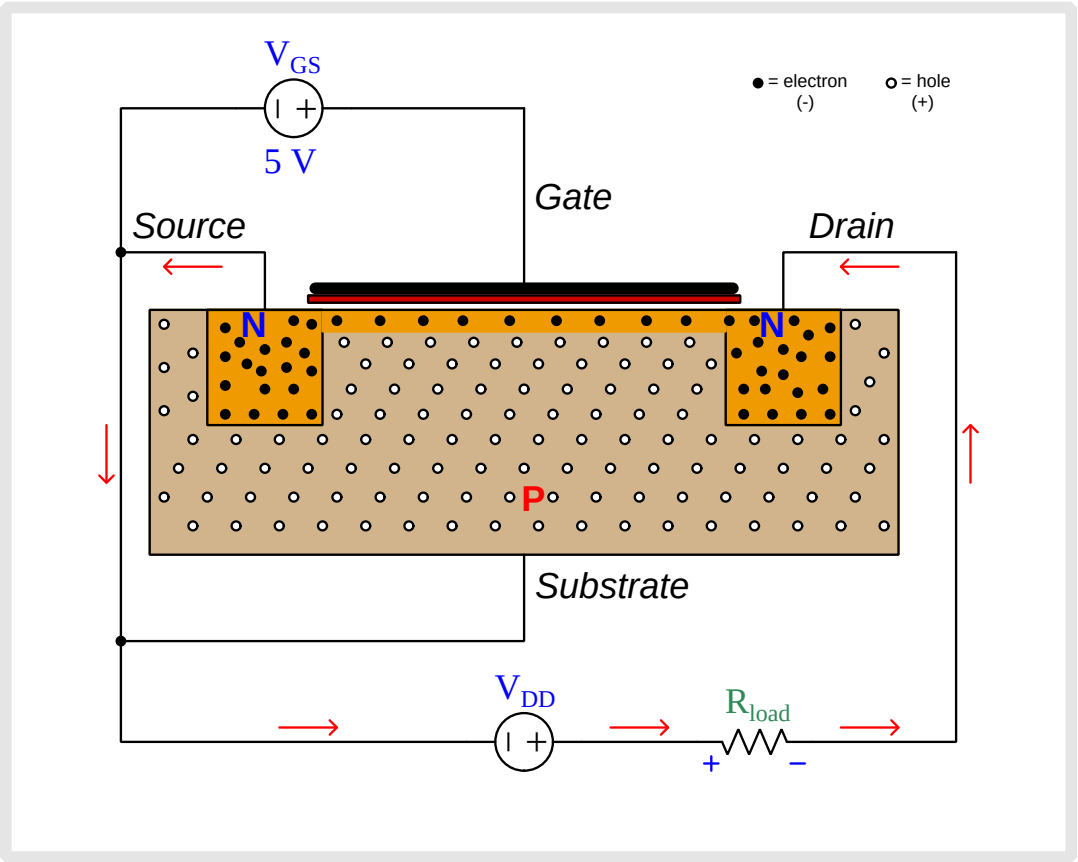
As this V_{GS} voltage increases, a thin film of free electrons begins to form beneath the gate, effectively turning what used to be P-type semiconductor material into N-type. This is called an *inversion layer*.



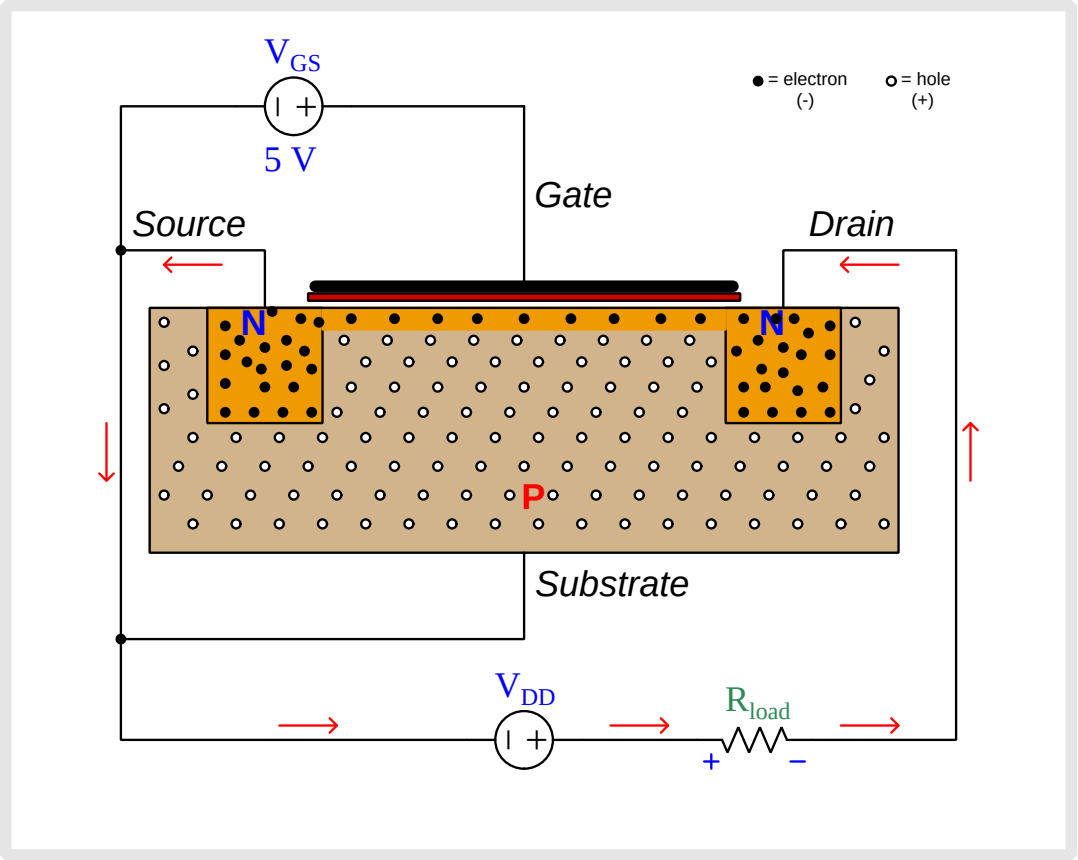
As this V_{GS} voltage increases, a thin film of free electrons begins to form beneath the gate, effectively turning what used to be P-type semiconductor material into N-type. This is called an *inversion layer*.



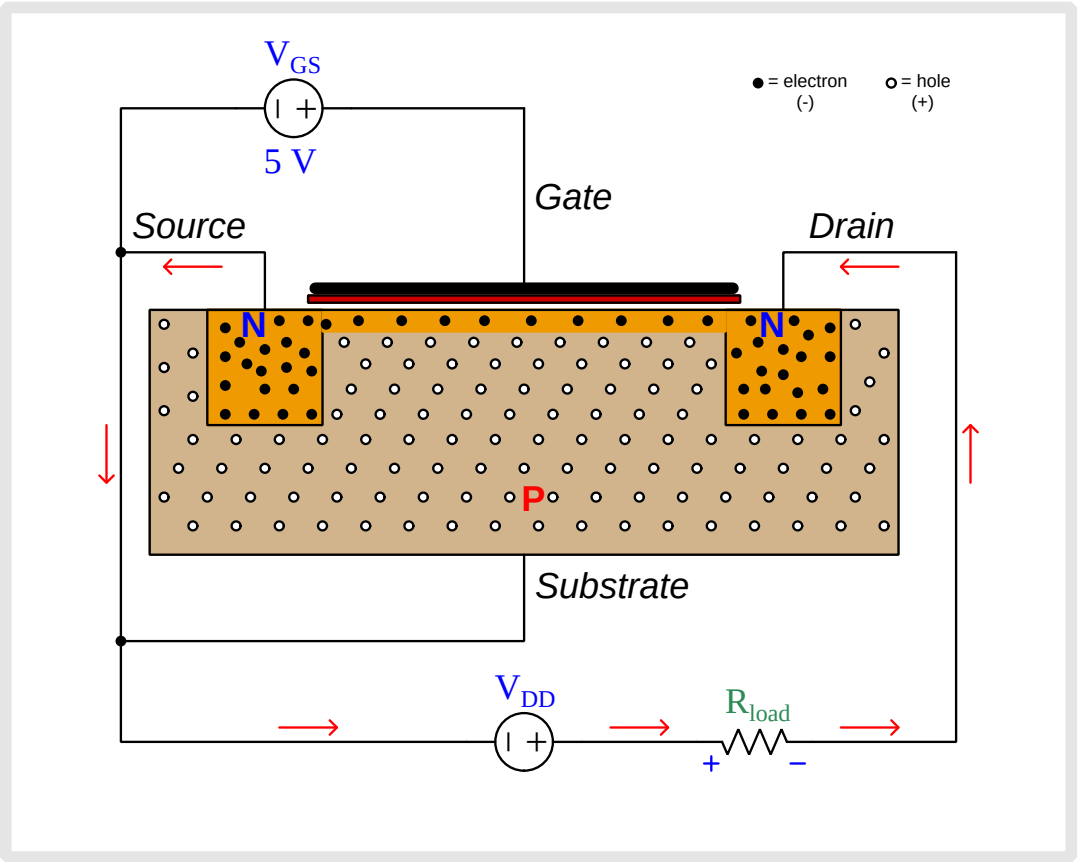
Once V_{GS} reaches the threshold or "turn-on" level, this inversion layer completely connects drain and source to form a new N-type channel permitting conduction between source and drain. The V_{DD} source is now able to promote a current through the MOSFET.



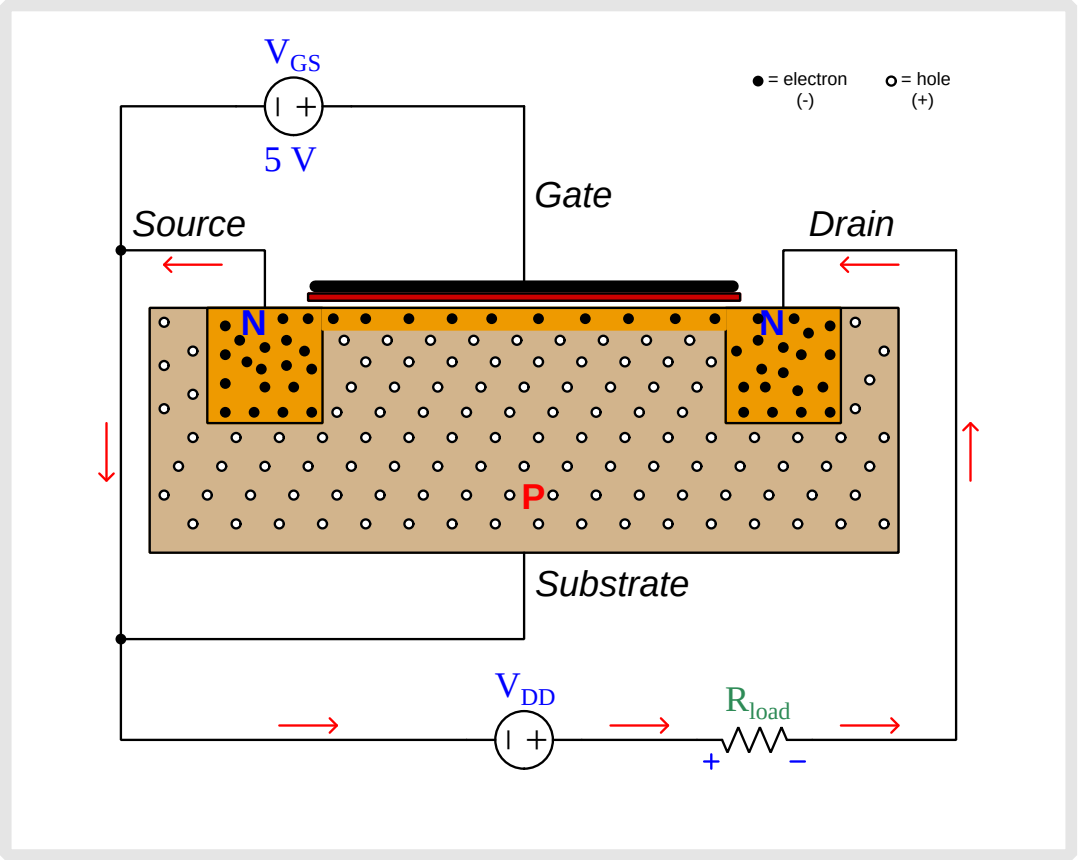
Note how the mobile charge carriers within the inversion layer are electrons, which flow opposite the direction of "conventional flow" as notated by the red arrows. Notice also how new electrons enter the inversion layer from the source region of the MOSFET, which is why that region is so-named (because that is the "source" of fresh charge carriers entering the inversion layer).



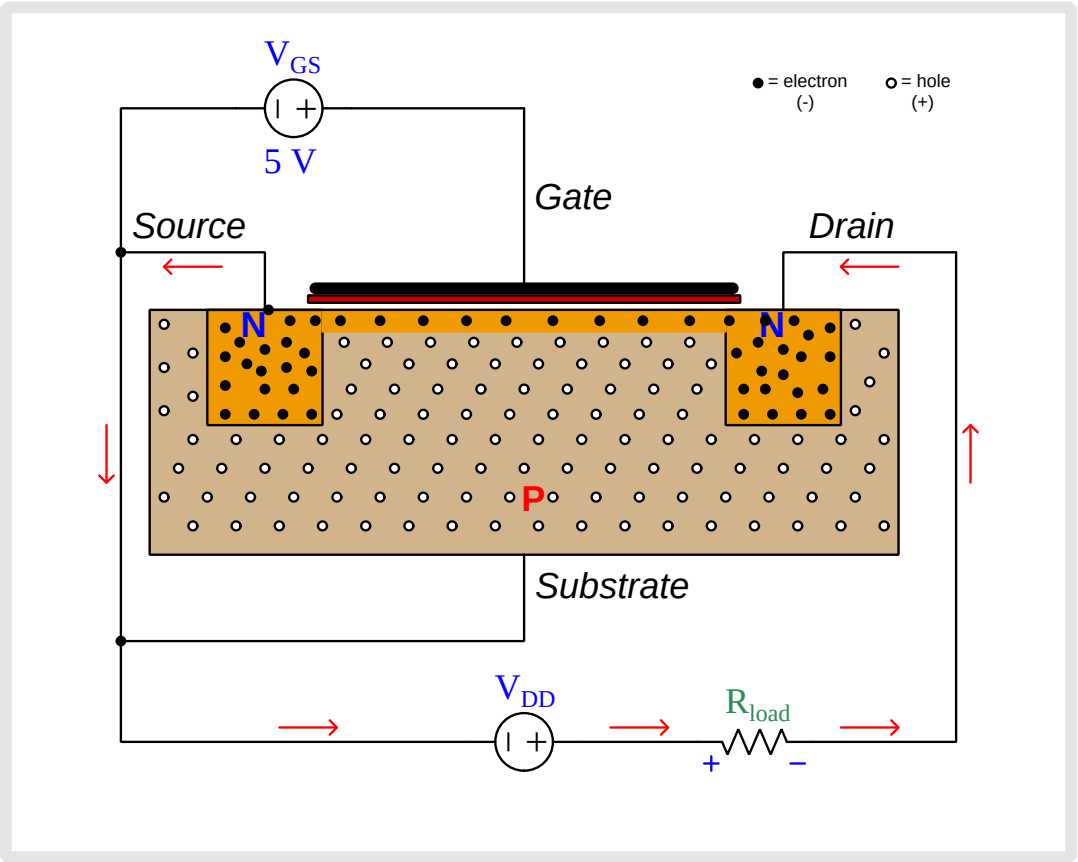
Note how the mobile charge carriers within the inversion layer are electrons, which flow opposite the direction of "conventional flow" as notated by the red arrows. Notice also how new electrons enter the inversion layer from the source region of the MOSFET, which is why that region is so-named (because that is the "source" of fresh charge carriers entering the inversion layer).



Note how the mobile charge carriers within the inversion layer are electrons, which flow opposite the direction of "conventional flow" as notated by the red arrows. Notice also how new electrons enter the inversion layer from the source region of the MOSFET, which is why that region is so-named (because that is the "source" of fresh charge carriers entering the inversion layer).



Note how the mobile charge carriers within the inversion layer are electrons, which flow opposite the direction of "conventional flow" as notated by the red arrows. Notice also how new electrons enter the inversion layer from the source region of the MOSFET, which is why that region is so-named (because that is the "source" of fresh charge carriers entering the inversion layer).



Note how the mobile charge carriers within the inversion layer are electrons, which flow opposite the direction of "conventional flow" as notated by the red arrows. Notice also how new electrons enter the inversion layer from the source region of the MOSFET, which is why that region is so-named (because that is the "source" of fresh charge carriers entering the inversion layer).

Chapter 8

Questions

This learning module, along with all others in the ModEL collection, is designed to be used in an inverted instructional environment where students independently read[1] the tutorials and attempt to answer questions on their own *prior* to the instructor's interaction with them. In place of lecture[2], the instructor engages with students in Socratic-style dialogue, probing and challenging their understanding of the subject matter through inquiry.

Answers are not provided for questions within this chapter, and this is by design. Solved problems may be found in the Tutorial and Derivation chapters, instead. The goal here is *independence*, and this requires students to be challenged in ways where others cannot think for them. Remember that you always have the tools of *experimentation* and *computer simulation* (e.g. SPICE) to explore concepts!

The following lists contain ideas for Socratic-style questions and challenges. Upon inspection, one will notice a strong theme of *metacognition* within these statements: they are designed to foster a regular habit of examining one's own thoughts as a means toward clearer thinking. As such these sample questions are useful both for instructor-led discussions as well as for self-study.

[1] Technical reading is an essential academic skill for any technical practitioner to possess for the simple reason that the most comprehensive, accurate, and useful information to be found for developing technical competence is in textual form. Technical careers in general are characterized by the need for continuous learning to remain current with standards and technology, and therefore any technical practitioner who cannot read well is handicapped in their professional development. An excellent resource for educators on improving students' reading prowess through intentional effort and strategy is the book textitReading For Understanding – How Reading Apprenticeship Improves Disciplinary Learning in Secondary and College Classrooms by Ruth Schoenbach, Cynthia Greenleaf, and Lynn Murphy.

[2] Lecture is popular as a teaching method because it is easy to implement: any reasonably articulate subject matter expert can talk to students, even with little preparation. However, it is also quite problematic. A good lecture always makes complicated concepts seem easier than they are, which is bad for students because it instills a false sense of confidence in their own understanding; reading and re-articulation requires more cognitive effort and serves to verify comprehension. A culture of teaching-by-lecture fosters a debilitating dependence upon direct personal instruction, whereas the challenges of modern life demand independent and critical thought made possible only by gathering information and perspectives from afar. Information presented in a lecture is ephemeral, easily lost to failures of memory and dictation; text is forever, and may be referenced at any time.

General challenges following tutorial reading

- Summarize as much of the text as you can in one paragraph of your own words. A helpful strategy is to explain ideas as you would for an intelligent child: as simple as you can without compromising too much accuracy.
- Simplify a particular section of the text, for example a paragraph or even a single sentence, so as to capture the same fundamental idea in fewer words.
- Where did the text make the most sense to you? What was it about the text's presentation that made it clear?
- Identify where it might be easy for someone to misunderstand the text, and explain why you think it could be confusing.
- Identify any new concept(s) presented in the text, and explain in your own words.
- Identify any familiar concept(s) such as physical laws or principles applied or referenced in the text.
- Devise a proof of concept experiment demonstrating an important principle, physical law, or technical innovation represented in the text.
- Devise an experiment to disprove a plausible misconception.
- Did the text reveal any misconceptions you might have harbored? If so, describe the misconception(s) and the reason(s) why you now know them to be incorrect.
- Describe any useful problem-solving strategies applied in the text.
- Devise a question of your own to challenge a reader's comprehension of the text.

General follow-up challenges for assigned problems

- Identify where any fundamental laws or principles apply to the solution of this problem, especially before applying any mathematical techniques.
- Devise a thought experiment to explore the characteristics of the problem scenario, applying known laws and principles to mentally model its behavior.
- Describe in detail your own strategy for solving this problem. How did you identify and organized the given information? Did you sketch any diagrams to help frame the problem?
- Is there more than one way to solve this problem? Which method seems best to you?
- Show the work you did in solving this problem, even if the solution is incomplete or incorrect.
- What would you say was the most challenging part of this problem, and why was it so?
- Was any important information missing from the problem which you had to research or recall?
- Was there any extraneous information presented within this problem? If so, what was it and why did it not matter?
- Examine someone else's solution to identify where they applied fundamental laws or principles.
- Simplify the problem from its given form and show how to solve this simpler version of it. Examples include eliminating certain variables or conditions, altering values to simpler (usually whole) numbers, applying a limiting case (i.e. altering a variable to some extreme or ultimate value).
- For quantitative problems, identify the real-world meaning of all intermediate calculations: their units of measurement, where they fit into the scenario at hand. Annotate any diagrams or illustrations with these calculated values.
- For quantitative problems, try approaching it qualitatively instead, thinking in terms of "increase" and "decrease" rather than definite values.
- For qualitative problems, try approaching it quantitatively instead, proposing simple numerical values for the variables.
- Were there any assumptions you made while solving this problem? Would your solution change if one of those assumptions were altered?
- Identify where it would be easy for someone to go astray in attempting to solve this problem.
- Formulate your own problem based on what you learned solving this one.

General follow-up challenges for experiments or projects

- In what way(s) was this experiment or project easy to complete?
- Identify some of the challenges you faced in completing this experiment or project.

- Show how thorough documentation assisted in the completion of this experiment or project.
- Which fundamental laws or principles are key to this system's function?
- Identify any way(s) in which one might obtain false or otherwise misleading measurements from test equipment in this system.
- What will happen if (component X) fails (open/shorted/etc.)?
- What would have to occur to make this system unsafe?

8.1 Conceptual reasoning

These questions are designed to stimulate your analytic and synthetic thinking[3]. In a Socratic discussion with your instructor, the goal is for these questions to prompt an extended dialogue where assumptions are revealed, conclusions are tested, and understanding is sharpened. Your instructor may also pose additional questions based on those assigned, in order to further probe and refine your conceptual understanding.

Questions that follow are presented to challenge and probe your understanding of various concepts presented in the tutorial. These questions are intended to serve as a guide for the Socratic dialogue between yourself and the instructor. Your instructor's task is to ensure you have a sound grasp of these concepts, and the questions contained in this document are merely a means to this end. Your instructor may, at his or her discretion, alter or substitute questions for the benefit of tailoring the discussion to each student's needs. The only absolute requirement is that each student is challenged and assessed at a level equal to or greater than that represented by the documented questions.

It is far more important that you convey your *reasoning* than it is to simply convey a correct answer. For this reason, you should refrain from researching other information sources to answer questions. What matters here is that *you* are doing the thinking. If the answer is incorrect, your instructor will work with you to correct it through proper reasoning. A correct answer without an adequate explanation of how you derived that answer is unacceptable, as it does not aid the learning or assessment process.

You will note a conspicuous lack of answers given for these conceptual questions. Unlike standard textbooks where answers to every other question are given somewhere toward the back of the book, here in these learning modules students must rely on other means to check their work. The best way by far is to debate the answers with fellow students and also with the instructor during the Socratic dialogue sessions intended to be used with these learning modules. Reasoning through challenging questions with other people is an excellent tool for developing strong reasoning skills.

Another means of checking your conceptual answers, where applicable, is to use circuit simulation software to explore the effects of changes made to circuits. For example, if one of these conceptual questions challenges you to predict the effects of altering some component parameter in a circuit, you may check the validity of your work by simulating that same parameter change within software and seeing if the results agree.

[3] *Analytical* thinking involves the "disassembly" of an idea into its constituent parts, analogous to dissection. *Synthetic* thinking involves the "assembly" of a new idea comprised of multiple concepts, analogous to construction. Both activities are high-level cognitive skills, extremely important for effective problem-solving, necessitating frequent challenge and regular practice to fully develop.

8.1.1 Reading outline and reflections

"Reading maketh a full man; conference a ready man; and writing an exact man" – Francis Bacon

Francis Bacon's advice is a blueprint for effective education: reading provides the learner with knowledge, writing focuses the learner's thoughts, and critical dialogue equips the learner to confidently communicate and apply their learning. Independent acquisition and application of knowledge is a powerful skill, well worth the effort to cultivate. To this end, students should read these educational resources closely, journal their own reflections on the reading, and discuss in detail their findings with classmates and instructor(s). You should be able to do all of the following after reading any instructional text:

☑ Briefly SUMMARIZE THE TEXT in the form of a journal entry documenting your learning as you progress through the course of study. Share this summary in dialogue with your classmates and instructor. Journaling is an excellent self-test of thorough reading because you cannot clearly express what you have not read or did not comprehend.

☑ Demonstrate ACTIVE READING STRATEGIES, including verbalizing your impressions as you read, simplifying long passages to convey the same ideas using fewer words, annotating text and illustrations with your own interpretations, working through mathematical examples shown in the text, cross-referencing passages with relevant illustrations and/or other passages, identifying problem-solving strategies applied by the author, etc. Technical reading is a special case of problem-solving, and so these strategies work precisely because they help solve any problem: paying attention to your own thoughts (metacognition), eliminating unnecessary complexities, identifying what makes sense, paying close attention to details, drawing connections between separated facts, and noting the successful strategies of others.

☑ Identify IMPORTANT THEMES, especially GENERAL LAWS and PRINCIPLES, expounded in the text and express them in the simplest of terms as though you were teaching an intelligent child. This emphasizes connections between related topics and develops your ability to communicate complex ideas to anyone.

☑ Form YOUR OWN QUESTIONS based on the reading, and then pose them to your instructor and classmates for their consideration. Anticipate both correct and incorrect answers, the incorrect answer(s) assuming one or more plausible misconceptions. This helps you view the subject from different perspectives to grasp it more fully.

☑ Devise EXPERIMENTS to test claims presented in the reading, or to disprove misconceptions. Predict possible outcomes of these experiments, and evaluate their meanings: what result(s) would confirm, and what would constitute disproof? Running mental simulations and evaluating results is essential to scientific and diagnostic reasoning.

☑ Specifically identify any points you found CONFUSING. The reason for doing this is to help diagnose misconceptions and overcome barriers to learning.

8.1.2 Foundational concepts

Correct analysis and diagnosis of electric circuits begins with a proper understanding of some basic concepts. The following is a list of some important concepts referenced in this module's full tutorial. Define each of them in your own words, and be prepared to illustrate each of these concepts with a description of a practical example and/or a live demonstration.

Energy

Conservation of Energy

Amplification

Transformer

Active versus Passive devices

Gain

Relay

Analog signal

Discrete signal

Thermionic emission

Vacuum tube

Doping

N-type semiconductor

P-type semiconductor

Depletion region

Diode

Forward-biasing

Reverse-biasing

Forward voltage drop

Field effect

Majority versus Minority carriers

Junction field-effect transistor

Insulated-gate field-effect transistor

Depletion mode

Enhancement mode

Parasitic effect

Current regulation

Threshold voltage

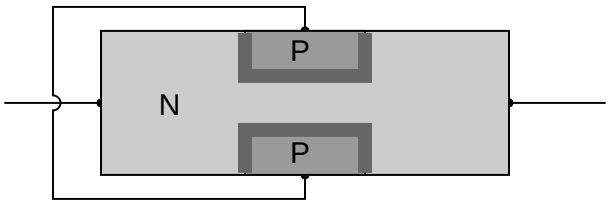
Transconductance

Square-law characteristic

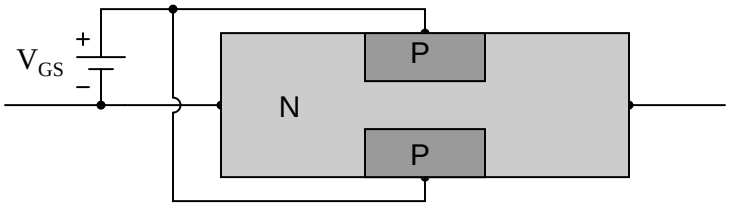
Electrostatic discharge

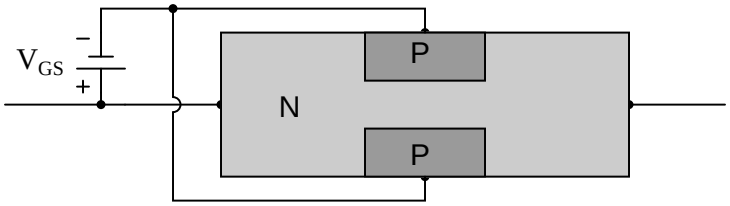
8.1.3 JFET depletion regions

Field-effect transistors (FETs) exhibit depletion regions between the oppositely-doped gate and channel sections, just as diodes have depletion regions between the P and N semiconductor halves. In this illustration, the depletion region appears as a dark, shaded area:



Re-draw the depletion regions for the following scenarios, where an external voltage (V_{GS}) is applied between the gate and channel:



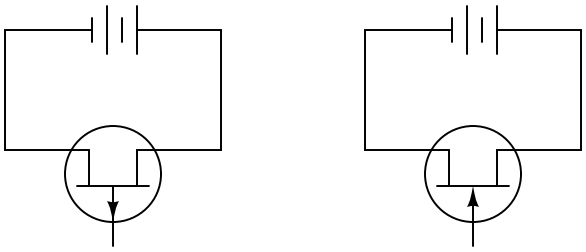


Challenges

- Why do you suppose this type of transistor is called a *field-effect* transistor? What "field" is being referred to in the operation of this device?
- Which of these two demonstrated scenarios is potentially harmful to the transistor, and why?

8.1.4 JFET source and drain terminals

Identify which transistor terminal functions as the *source* and which transistor terminal functions as the *drain* in both of these JFET circuits:



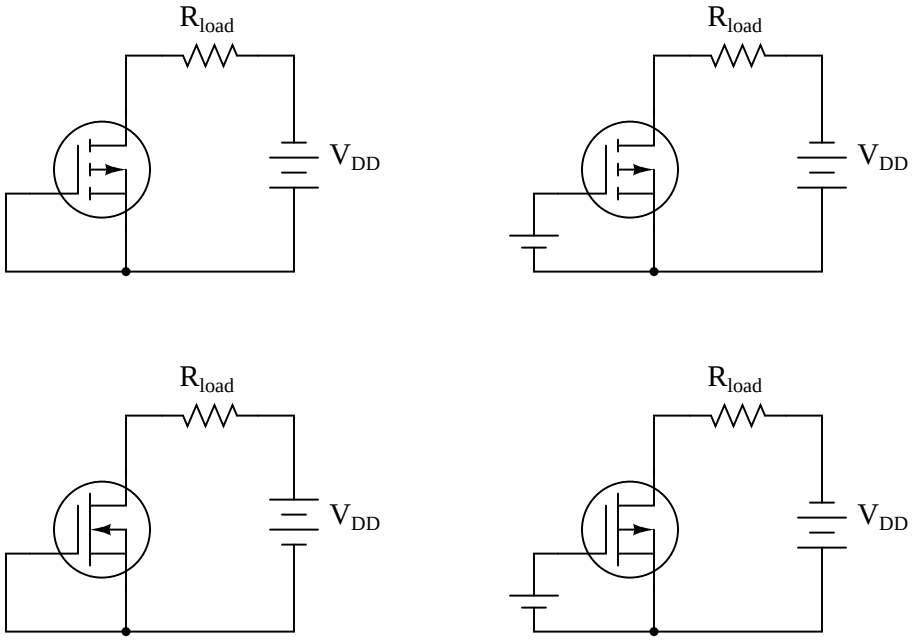
Most importantly, explain *why* we define the terminals as such, given the fact that is often no physical difference between these two terminals of a JFET.

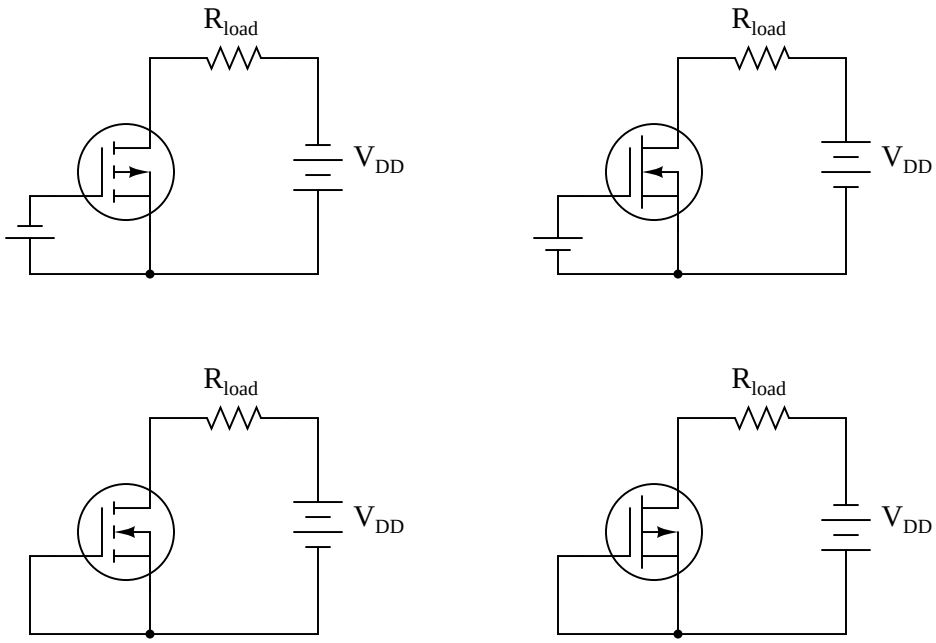
Challenges

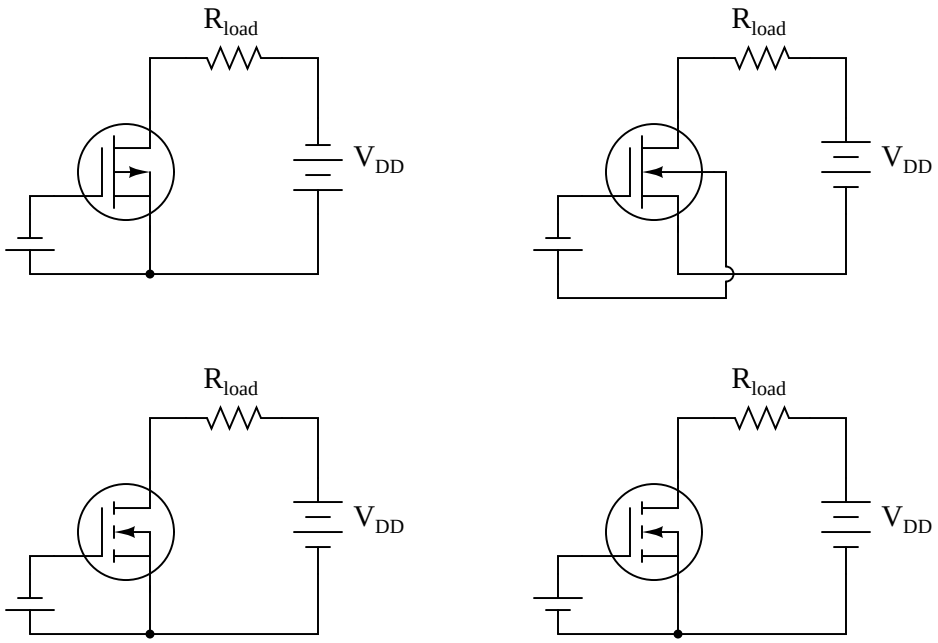
- For JFETs that actually do exhibit a physical distinction between source and drain terminals, what exactly is that distinction?

8.1.5 Identifying MOSFET states

Identify each type of MOSFET (whether it is N-channel or P-channel, D-type or E-type), label the terminals, and determine whether the MOSFET in each of these circuits will be turned *on* or *off*. Additionally, sketch the parasitic diode into each schematic, representing the direction this unintentional PN junction faces:




R load
V DD
R load
V DD
R load
V DD
R load
V DD

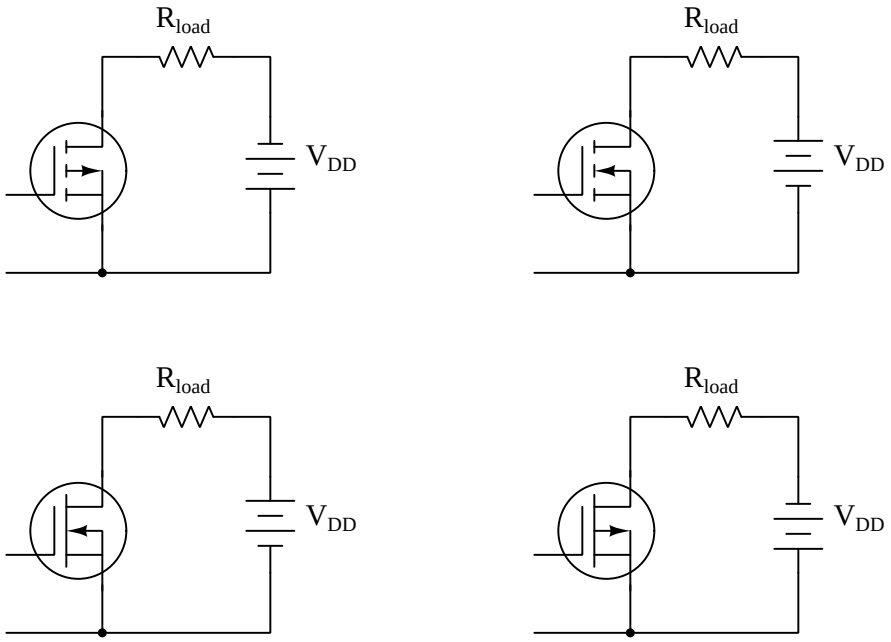


Challenges

- Which of these transistors is *depleted* and which is *enhanced*.

8.1.6 Identifying MOSFET stimuli

Sketch a small voltage source symbol into each of these circuits oriented in such a way as to enhance (turn on) each MOSFET:



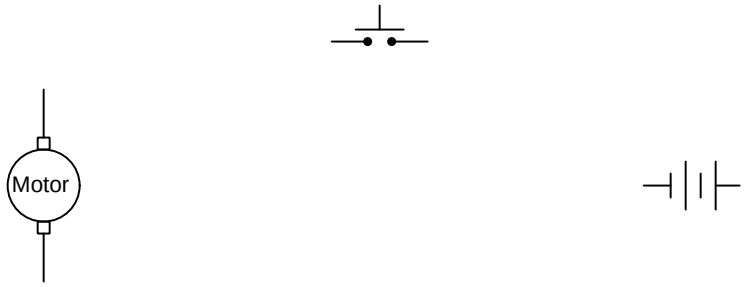
Also, identify what would happen if you simply shorted together the two wire stubs shown in each schematic, rather than connect a small voltage source between them.

Challenges

- Sketch the parasitic diode within each of these transistors.
- Explain how a MOSFET may be damaged by electrostatic potentials.

8.1.7 Sketching MOSFET switch circuit

Sketch a MOSFET into this circuit to control the DC motor at the behest of a pushbutton switch, that switch conducting only as much current as is necessary to activate the transistor which then in turn conducts the motor's current.



Challenges

- Explain why a resistor must be connected between gate and source in your circuit to ensure the MOSFET is able to turn off when the pushbutton is released.

8.1.8 Typical JFET pinch-off voltage values

Explain what *pinch-off voltage* ($V_{GS(off)}$, also known as *threshold voltage*) is for a field-effect transistor. Research the datasheets for some of the following field-effect transistors and determine what their respective cutoff voltages are:

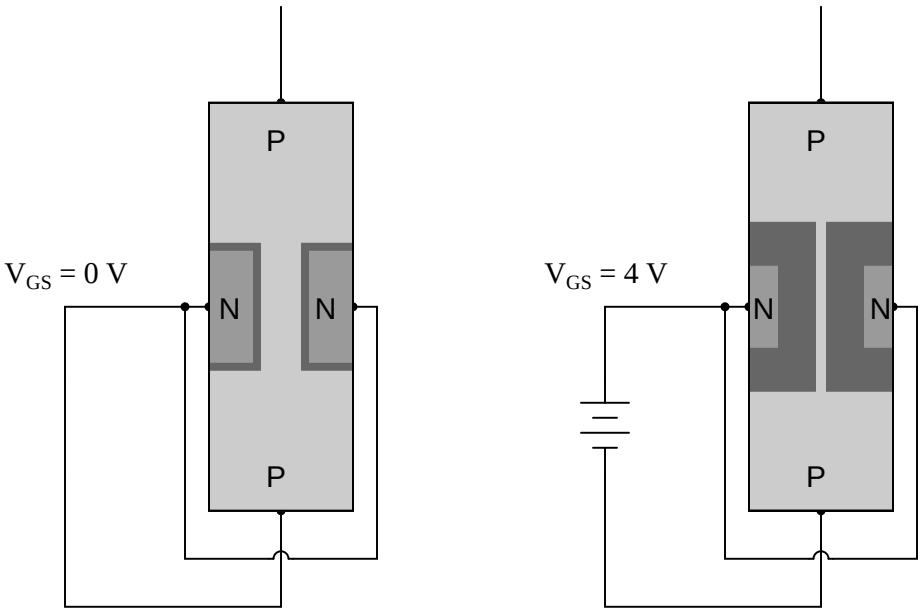
- J110
- J308
- J309
- J310
- 2N5457
- 2N5458

Challenges

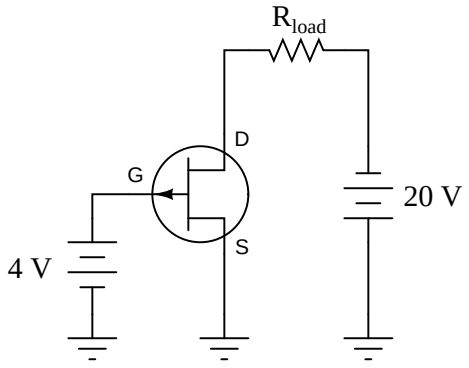
- Based on your research of these datasheets, how constant is $V_{GS(off)}$ between different transistors? In other words, is this a parameter you can accurately predict from the datasheet before purchasing a transistor, or does it vary significantly from transistor to transistor (of the same part number)?

8.1.9 Effect of V_{DS} on depletion region

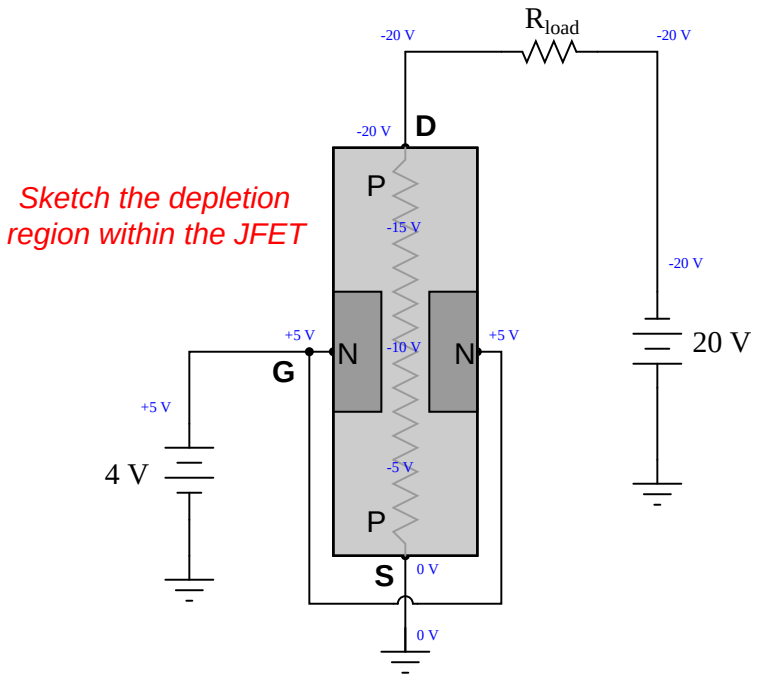
When a reverse-bias voltage is applied between the gate and channel of a JFET, the depletion region within expands. The greater the reverse-bias voltage, the wider the depletion region becomes. With enough applied V_{GS}, this expansion will pinch off the JFET's channel, preventing drain-source current:



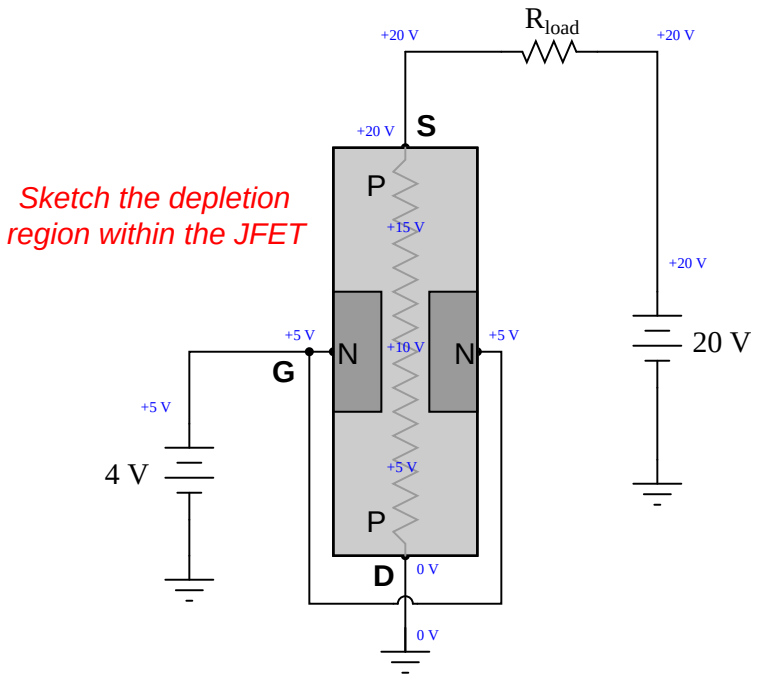
Something not immediately apparent about this effect is that the formation of a wide depletion region necessary for cut-off of a field-effect transistor is *also* affected by the drain-to-source voltage drop (V_{DS}).



If we connect a gate-to-source voltage (V_{GS}) large enough to force the transistor into cutoff mode, the JFET channel will act as a huge resistance. If we look carefully at the voltages measured with reference to ground, we will see that the width of the depletion region must vary within the JFET's channel. Sketch this varying width, given the voltages shown in the illustration:



Something *really* interesting happens if we reverse the polarity of the 20 volt source: the relationship between drain and source changes, and so does the depletion region profile. Sketch the new varying width of the depletion region with the 20 volt source reversed, and comment on what you see:

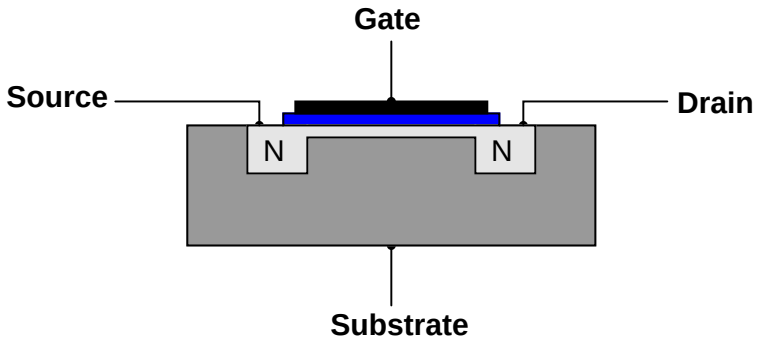


Challenges

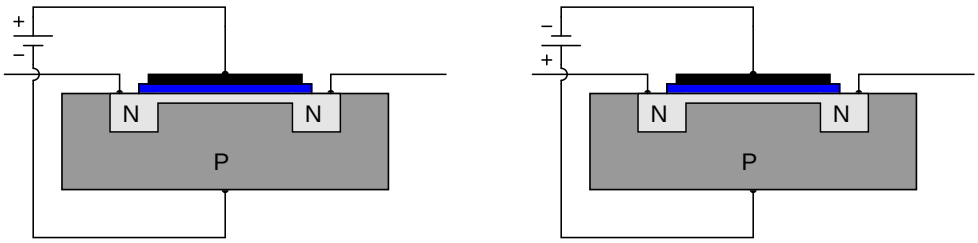
- Will the transistor function equally well with both V_{DS} polarities?

8.1.10 Effects of V_{GS} on MOSFET

Some types of MOSFETs have a source-drain channel already formed with no applied gate voltage:



Explain what happens to source-drain conductivity with each of the following applied gate-to-substrate voltages. Modify the illustrations if necessary:

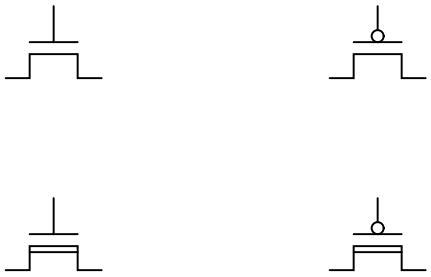


Challenges

- What type of transistor is this? Be as specific as you can.

8.1.11 Interesting transistor symbols

Identify these schematic symbols:

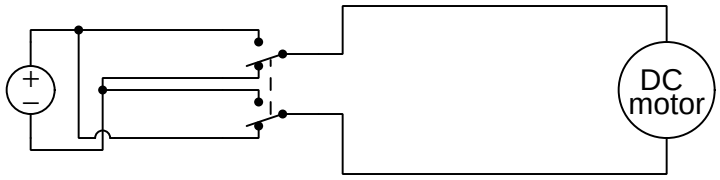


Challenges

- Identify at least one disadvantage of using these "simplified" transistor symbols.

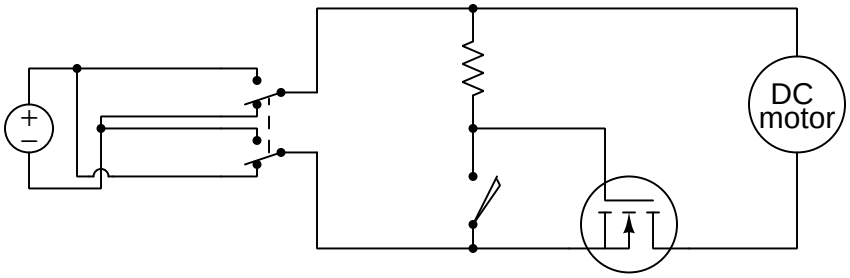
8.1.12 Servo motor travel limiting

The following circuit shows how a DC motor may be controlled by a three-position (center-off) DPDT switch:



Explain how the DC motor will respond to the switch's three positions (up, center, and down).

Suppose this DC motor is used as a *servo* to control the motion of a robot, and we must install a *limit switch* to stop the motor from turning too far in one direction. The following circuit accomplishes this, using the limit switch to control the state of a MOSFET, which in turn constrains the motor's motion:



Answer the following questions about this circuit:

- In which direction will the motor's travel be limited?
- Must the limit switch *close* to stop the motor, or *open* to stop the motor?
- Trace the path of current for both of the "energized" positions of the switch

Challenges

- Explain why the MOSFET's parasitic diode is actually important for the operation of this circuit as designed.
- What factors would determine an appropriate resistance value for the resistor in this circuit?

8.2 Quantitative reasoning

These questions are designed to stimulate your computational thinking. In a Socratic discussion with your instructor, the goal is for these questions to reveal your mathematical approach(es) to problem-solving so that good technique and sound reasoning may be reinforced. Your instructor may also pose additional questions based on those assigned, in order to observe your problem-solving firsthand.

Mental arithmetic and estimations are strongly encouraged for all calculations, because without these abilities you will be unable to readily detect errors caused by calculator misuse (e.g. keystroke errors).

You will note a conspicuous lack of answers given for these quantitative questions. Unlike standard textbooks where answers to every other question are given somewhere toward the back of the book, here in these learning modules students must rely on other means to check their work. My advice is to use circuit simulation software such as SPICE to check the correctness of quantitative answers. Refer to those learning modules within this collection focusing on SPICE to see worked examples which you may use directly as practice problems for your own study, and/or as templates you may modify to run your own analyses and generate your own practice problems.

Completely worked example problems found in the Tutorial may also serve as "test cases[4]" for gaining proficiency in the use of circuit simulation software, and then once that proficiency is gained you will never need to rely[5] on an answer key!

[4] In other words, set up the circuit simulation software to analyze the same circuit examples found in the Tutorial. If the simulated results match the answers shown in the Tutorial, it confirms the simulation has properly run. If the simulated results disagree with the Tutorial's answers, something has been set up incorrectly in the simulation software. Using every Tutorial as practice in this way will quickly develop proficiency in the use of circuit simulation software.

[5] This approach is perfectly in keeping with the instructional philosophy of these learning modules: *teaching students to be self-sufficient thinkers*. Answer keys can be useful, but it is even more useful to your long-term success to have a set of tools on hand for checking your own work, because once you have left school and are on your own, there will no longer be "answer keys" available for the problems you will have to solve.

8.2.1 Miscellaneous physical constants

Note: constants shown in **bold** type are *exact*, not approximations. Values inside of parentheses show one standard deviation (σ) of uncertainty in the final digits: for example, the magnetic permeability of free space value given as $1.25663706212(19) \times 10^{-6}$ H/m represents a center value (i.e. the location parameter) of $1.25663706212 \times 10^{-6}$ Henrys per meter with one standard deviation of uncertainty equal to $0.0000000000019 \times 10^{-6}$ Henrys per meter.

Avogadro's number (N_A) = **6.02214076** $\times 10^{23}$ **per mole** (mol^{-1})

Boltzmann's constant (k) = **1.380649** $\times 10^{-23}$ **Joules per Kelvin** (J/K)

Electronic charge (e) = **1.602176634** $\times 10^{-19}$ **Coulomb** (C)

Faraday constant (F) = **96,485.33212...** $\times 10^{4}$ **Coulombs per mole** (C/mol)

Magnetic permeability of free space (μ_0) = $1.25663706212(19) \times 10^{-6}$ Henrys per meter (H/m)

Electric permittivity of free space (ϵ_0) = $8.8541878128(13) \times 10^{-12}$ Farads per meter (F/m)

Characteristic impedance of free space (Z_0) = 376.730313668(57) Ohms (Ω)

Gravitational constant (G) = $6.67430(15) \times 10^{-11}$ cubic meters per kilogram-seconds squared ($m^3/kg\text{-}s^2$)

Molar gas constant (R) = **8.314462618... Joules per mole-Kelvin** (J/mol-K) = 0.08205746(14) liters-atmospheres per mole-Kelvin

Planck constant (h) = **6.62607015** $\times 10^{-34}$ **joule-seconds** (J-s)

Stefan-Boltzmann constant (σ) = **5.670374419...** $\times 10^{-8}$ **Watts per square meter-Kelvin**^{4} ($W/m^2 \cdot K^4$)

Speed of light in a vacuum (c) = **299,792,458 meters per second** (m/s) = 186282.4 miles per second (mi/s)

Note: All constants taken from NIST data "Fundamental Physical Constants – Complete Listing", from `http://physics.nist.gov/constants`, National Institute of Standards and Technology (NIST), 2018 CODATA Adjustment.

8.2.2 Introduction to spreadsheets

A powerful computational tool you are encouraged to use in your work is a *spreadsheet*. Available on most personal computers (e.g. Microsoft Excel), *spreadsheet* software performs numerical calculations based on number values and formulae entered into cells of a grid. This grid is typically arranged as lettered columns and numbered rows, with each cell of the grid identified by its column/row coordinates (e.g. cell `B3`, cell `A8`). Each cell may contain a string of text, a number value, or a mathematical formula. The spreadsheet automatically updates the results of all mathematical formulae whenever the entered number values are changed. This means it is possible to set up a spreadsheet to perform a series of calculations on entered data, and those calculations will be re-done by the computer any time the data points are edited in any way.

For example, the following spreadsheet calculates average speed based on entered values of distance traveled and time elapsed:

	A	B	C	D
1	Distance traveled	46.9	Kilometers	
2	Time elapsed	1.18	Hours	
3	Average speed	= B1 / B2	km/h	
4				
5				

Text labels contained in cells `A1` through `A3` and cells `C1` through `C3` exist solely for readability and are not involved in any calculations. Cell `B1` contains a sample distance value while cell `B2` contains a sample time value. The formula for computing speed is contained in cell `B3`. Note how this formula begins with an "equals" symbol (`=`), references the values for distance and speed by lettered column and numbered row coordinates (`B1` and `B2`), and uses a forward slash symbol for division (`/`). The coordinates `B1` and `B2` function as *variables*[6] would in an algebraic formula.

When this spreadsheet is executed, the numerical value `39.74576` will appear in cell `B3` rather than the formula `= B1 / B2`, because 39.74576 is the computed speed value given 46.9 kilometers traveled over a period of 1.18 hours. If a different numerical value for distance is entered into cell `B1` or a different value for time is entered into cell `B2`, cell `B3`'s value will automatically update. All you need to do is set up the given values and any formulae into the spreadsheet, and the computer will do all the calculations for you.

Cell `B3` may be referenced by other formulae in the spreadsheet if desired, since it is a variable just like the given values contained in `B1` and `B2`. This means it is possible to set up an entire chain of calculations, one dependent on the result of another, in order to arrive at a final value. The arrangement of the given data and formulae need not follow any pattern on the grid, which means you may place them anywhere.

[6] Spreadsheets may also provide means to attach text labels to cells for use as variable names (Microsoft Excel simply calls these labels "names"), but for simple spreadsheets such as those shown here it's usually easier just to use the standard coordinate naming for each cell.

Common[7] arithmetic operations available for your use in a spreadsheet include the following:

- Addition (`+`)
- Subtraction (`-`)
- Multiplication (`*`)
- Division (`/`)
- Powers (`^`)
- Square roots (`sqrt()`)
- Logarithms (`ln()` , `log10()`)

Parentheses may be used to ensure[8] proper order of operations within a complex formula. Consider this example of a spreadsheet implementing the *quadratic formula*, used to solve for roots of a polynomial expression in the form of $ax^2 + bx + c$:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

	A	B
1	x_1	= (-B4 + sqrt((B4^2) - (4*B3*B5))) / (2*B3)
2	x_2	= (-B4 - sqrt((B4^2) - (4*B3*B5))) / (2*B3)
3	a =	9
4	b =	5
5	c =	-2

This example is configured to compute roots[9] of the polynomial $9x^2 + 5x - 2$ because the values of 9, 5, and -2 have been inserted into cells `B3`, `B4`, and `B5`, respectively. Once this spreadsheet has been built, though, it may be used to calculate the roots of *any* second-degree polynomial expression simply by entering the new a, b, and c coefficients into cells `B3` through `B5`. The numerical values appearing in cells `B1` and `B2` will be automatically updated by the computer immediately following any changes made to the coefficients.

[7] Modern spreadsheet software offers a bewildering array of mathematical functions you may use in your computations. I recommend you consult the documentation for your particular spreadsheet for information on operations other than those listed here.

[8] Spreadsheet programs, like text-based programming languages, are designed to follow standard order of operations by default. However, my personal preference is to use parentheses even where strictly unnecessary just to make it clear to any other person viewing the formula what the intended order of operations is.

[9] Reviewing some algebra here, a *root* is a value for x that yields an overall value of zero for the polynomial. For this polynomial ($9x^2 + 5x - 2$) the two roots happen to be $x = 0.269381$ and $x = -0.82494$, with these values displayed in cells `B1` and `B2`, respectively upon execution of the spreadsheet.

Alternatively, one could break up the long quadratic formula into smaller pieces like this:

$$y = \sqrt{b^2 - 4ac} \qquad z = 2a$$

$$x = \frac{-b \pm y}{z}$$

	A	B	C
1	x_1	= (-B4 + C1) / C2	= sqrt((B4^2) - (4*B3*B5))
2	x_2	= (-B4 - C1) / C2	= 2*B3
3	a =	9	
4	b =	5	
5	c =	-2	

Note how the square-root term (y) is calculated in cell C1, and the denominator term (z) in cell C2. This makes the two final formulae (in cells B1 and B2) simpler to interpret. The positioning of all these cells on the grid is completely arbitrary[10] – all that matters is that they properly reference each other in the formulae.

Spreadsheets are particularly useful for situations where the same set of calculations representing a circuit or other system must be repeated for different initial conditions. The power of a spreadsheet is that it automates what would otherwise be a tedious set of calculations. One specific application of this is to simulate the effects of various components within a circuit failing with abnormal values (e.g. a shorted resistor simulated by making its value nearly zero; an open resistor simulated by making its value extremely large). Another application is analyzing the behavior of a circuit design given new components that are out of specification, and/or aging components experiencing drift over time.

[10] My personal preference is to locate all the "given" data in the upper-left cells of the spreadsheet grid (each data point flanked by a sensible name in the cell to the left and units of measurement in the cell to the right as illustrated in the first distance/time spreadsheet example), sometimes coloring them in order to clearly distinguish which cells contain entered data versus which cells contain computed results from formulae. I like to place all formulae in cells below the given data, and try to arrange them in logical order so that anyone examining my spreadsheet will be able to figure out *how* I constructed a solution. This is a general principle I believe all computer programmers should follow: *document and arrange your code to make it easy for other people to learn from it.*

8.2.3 Drain current calculation

The equation solving for drain current through a JFET is as follows:

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)^2$$

Where,
I_D = Drain current
I_{DSS} = Drain current with the Gate terminal shorted to the Source terminal
V_{GS} = Applied Gate-to-Source voltage
$V_{GS(off)}$ = Gate-to-Source voltage necessary to "pinch off" the JFET

Algebraically manipulate this equation to solve for V_{GS}, and explain why this new equation might be useful to us.

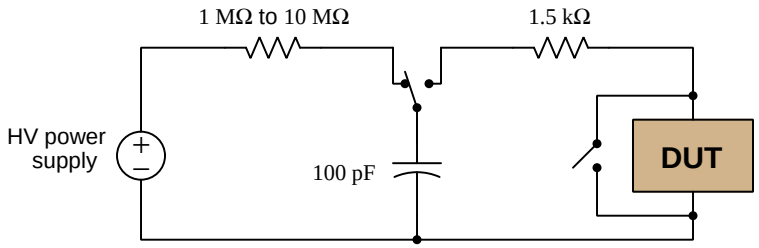
Challenges

- These mathematical functions are clearly nonlinear. Is there any evidence of this nonlinearity in the characteristic curves for a FET?

8.2.4 ESD test jig

Electrostatic discharge, or *ESD*, is a serious concern for microscopic integrated circuits of all kinds, but especially for insulated-gate transistors where the dielectric layer separating gate from channel is extremely thin.

The US Department of Defense "Test Method Standard Microcircuits" (MIL-STD-883E) shows a *human body model* for ESD testing of microelectronic devices as shown in the following diagram. The acronym "DUT" stands for *Device Under Test*:



Calculate the amount of energy delivered to the DUT with a high-voltage power supply of 2000 Volts, assuming the capacitor has reached full energization prior to the SPDT switch being thrown toward the DUT.

Also, identify how some of the foundational concepts you've studied apply to this circuit: **behavior of sources vs. loads**, **Ohm's Law**, **Conservation of Energy**, **Kirchhoff's Voltage Law**, **Kirchhoff's Current Law**, **electric fields**, **time constants**, **properties of opens vs. shorts**. Feel free to include any other foundational concepts not listed here.

Challenges

- Describe the amount of energy represented by a "Joule", in simple terms.
- Describe the purpose of each resistor in the model.
- Describe the purpose of the SPST switch.
- If the DUT had a resistance of 5 kΩ, how long would this circuit's discharge time constant (τ) be?

8.3 Diagnostic reasoning

These questions are designed to stimulate your deductive and inductive thinking, where you must apply general principles to specific scenarios (deductive) and also derive conclusions about the failed circuit from specific details (inductive). In a Socratic discussion with your instructor, the goal is for these questions to reinforce your recall and use of general circuit principles and also challenge your ability to integrate multiple symptoms into a sensible explanation of what's wrong in a circuit. Your instructor may also pose additional questions based on those assigned, in order to further challenge and sharpen your diagnostic abilities.

As always, your goal is to fully *explain* your analysis of each problem. Simply obtaining a correct answer is not good enough – you must also demonstrate sound reasoning in order to successfully complete the assignment. Your instructor's responsibility is to probe and challenge your understanding of the relevant principles and analytical processes in order to ensure you have a strong foundation upon which to build further understanding.

You will note a conspicuous lack of answers given for these diagnostic questions. Unlike standard textbooks where answers to every other question are given somewhere toward the back of the book, here in these learning modules students must rely on other means to check their work. The best way by far is to debate the answers with fellow students and also with the instructor during the Socratic dialogue sessions intended to be used with these learning modules. Reasoning through challenging questions with other people is an excellent tool for developing strong reasoning skills.

Another means of checking your diagnostic answers, where applicable, is to use circuit simulation software to explore the effects of faults placed in circuits. For example, if one of these diagnostic questions requires that you predict the effect of an open or a short in a circuit, you may check the validity of your work by simulating that same fault (substituting a very high resistance in place of that component for an open, and substituting a very low resistance for a short) within software and seeing if the results agree.

8.3.1 Testing a JFET by continuity

Based on these DC continuity tester indications, what type of JFET is this, N-channel or P-channel?

- Resistance with negative test lead on pin 1, positive test lead on pin 2: *no continuity*
- Resistance with negative test lead on pin 1, positive test lead on pin 3: *no continuity*
- Resistance with negative test lead on pin 2, positive test lead on pin 1: **continuity**
- Resistance with negative test lead on pin 2, positive test lead on pin 3: **continuity**
- Resistance with negative test lead on pin 3, positive test lead on pin 1: **continuity**
- Resistance with negative test lead on pin 3, positive test lead on pin 2: **continuity**

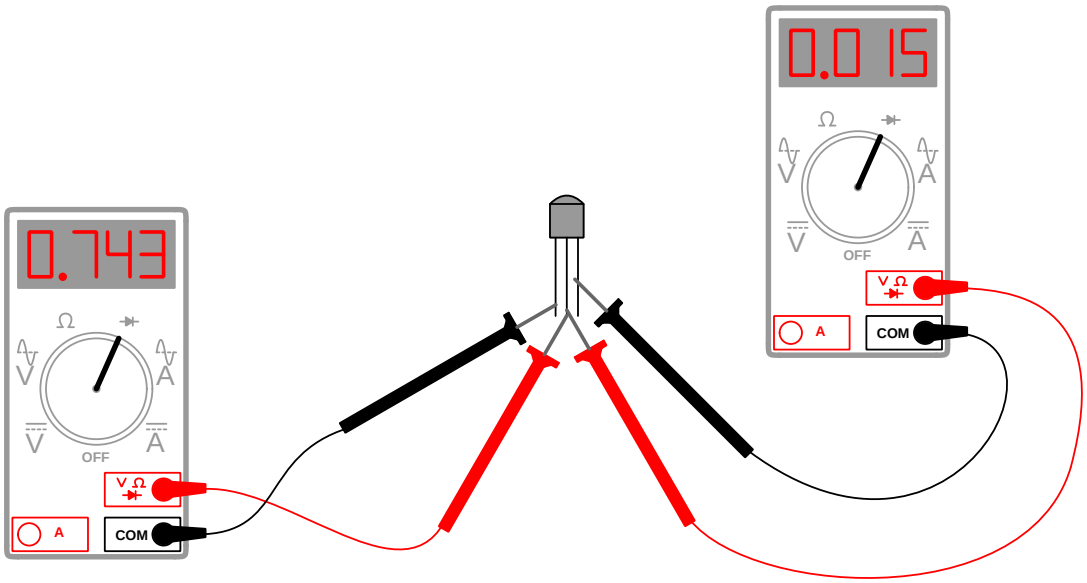
Also, to the best of your ability, identify the transistor's three terminals (source, gate, and drain).

Challenges

- Some analog multimeters switch test lead polarity when in "resistance" or "continuity" mode (i.e. red lead is negative and black lead is positive). How would such a reversal affect a technician's ability to identify these transistor terminals?

8.3.2 Testing a JFET by DMM diode check

From the "diode check" measurements taken with these two meters, identify the terminals on this JFET, and also what type of JFET it is (N-channel or P-channel):



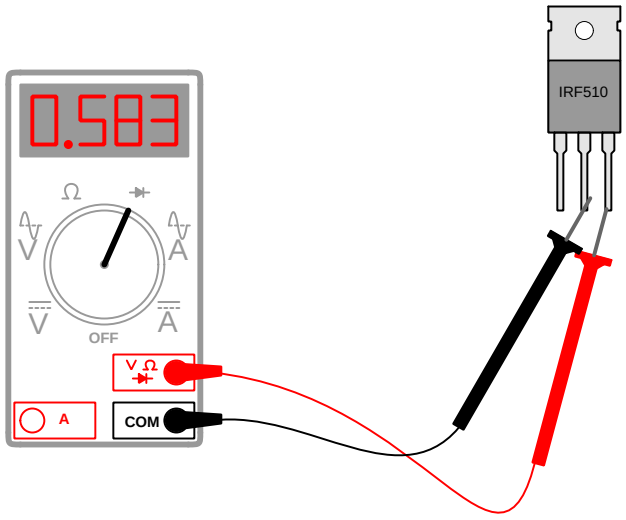
Also, to the best of your ability, identify the transistor's three terminals (source, gate, and drain).

Challenges

- What exactly does the numerical display represent on a DMM set to the "diode check" mode?
- Some analog multimeters switch test lead polarity when in "resistance" or "continuity" mode (i.e. red lead is negative and black lead is positive). How would such a reversal affect a technician's ability to identify these transistor terminals?

8.3.3 Testing a MOSFET by DMM diode check

A technician is using a digital multimeter (with a "diode check" feature) to identify the terminals of a power MOSFET:



The technician obtains the following "diode check" voltage measurements, in this order:

1. Black lead on middle terminal, Red lead on right terminal = 0.583 Volts (*shown in illustration*)
2. Red lead on middle terminal, Black lead on right terminal = O.L. (open)
3. Black lead on middle terminal, Red lead on left terminal = O.L. (open)
4. Black lead on middle terminal, Red lead on right terminal = 0.001 Volts
5. Red lead on middle terminal, Black lead on right terminal = 0.001 Volts

Explain why the fourth and fifth measurements are so different from the first and second, respectively, when they were taken between the same terminals on the MOSFET.

Challenges

- What exactly does the numerical display represent on a DMM set to the "diode check" mode?
- What is the significance of the 0.583 Volt measurement?
- Is it possible to determine the channel type (N- or P-) for this MOSFET strictly by the measurements we take with the multimeter?

8.3.4 ESD and MOSFETs

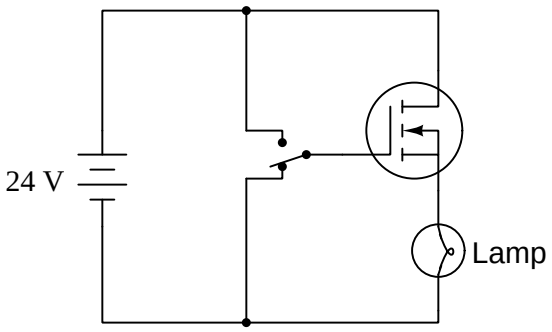
An important consideration when working around circuits containing MOSFETs is *electrostatic discharge*, or *ESD*. Describe what this phenomenon is, and why it is an important consideration for MOSFET circuits.

Challenges

- Identify some common sources of static electricity to be aware of in a laboratory or manufacturing environment.

8.3.5 Dim lamp

Something is wrong with the design of this MOSFET switching circuit: when the SPDT toggle switch is moved to the "on" position, the lamp glows dimly rather than brightly:



First, identify the proper switch positions for "on" and "off".

Explain why the lamp cannot achieve full brightness in this circuit, even when rated for the source's voltage (24 Volts).

Challenges

- Fix this design so that it will work better.

Chapter 9

Projects and Experiments

The following project and experiment descriptions outline things you can build to help you understand circuits. With any real-world project or experiment there exists the potential for physical harm. *Electricity can be very dangerous in certain circumstances, and you should follow proper safety precautions at all times!*

9.1 Recommended practices

This section outlines some recommended practices for all circuits you design and construct.

9.1.1 Safety first!

Electricity, when passed through the human body, causes uncomfortable sensations and in large enough measures[1] will cause muscles to involuntarily contract. The overriding of your nervous system by the passage of electrical current through your body is particularly dangerous in regard to your heart, which is a vital muscle. Very large amounts of current can produce serious internal burns in addition to all the other effects.

Cardio-pulmonary resuscitation (CPR) is the standard first-aid for any victim of electrical shock. This is a very good skill to acquire if you intend to work with others on dangerous electrical circuits. You should never perform tests or work on such circuits unless someone else is present who is proficient in CPR.

As a general rule, any voltage in excess of 30 Volts poses a definitive electric shock hazard, because beyond this level human skin does not have enough resistance to safely limit current through the body. "Live" work of any kind with circuits over 30 volts should be avoided, and if unavoidable should only be done using electrically insulated tools and other protective equipment (e.g. insulating shoes and gloves). If you are unsure of the hazards, or feel unsafe at any time, stop all work and distance yourself from the circuit!

A policy I strongly recommend for students learning about electricity is to *never come into electrical contact*[2] *with an energized conductor, no matter what the circuit's voltage*[3] *level!* Enforcing this policy may seem ridiculous when the circuit in question is powered by a single battery smaller than the palm of your hand, but it is precisely this instilled habit which will save a person from bodily harm when working with more dangerous circuits. Experience has taught me that students who learn early on to be careless with safe circuits have a tendency to be careless later with dangerous circuits!

In addition to the electrical hazards of shock and burns, the construction of projects and running of experiments often poses other hazards such as working with hand and power tools, potential

[1] Professor Charles Dalziel published a research paper in 1961 called "The Deleterious Effects of Electric Shock" detailing the results of electric shock experiments with both human and animal subjects. The threshold of perception for human subjects holding a conductor in their hand was in the range of 1 milliampere of current (less than this for alternating current, and generally less for female subjects than for male). Loss of muscular control was exhibited by half of Dalziel's subjects at less than 10 milliamperes alternating current. Extreme pain, difficulty breathing, and loss of all muscular control occurred for over 99% of his subjects at direct currents less than 100 milliamperes and alternating currents less than 30 milliamperes. In summary, it doesn't require much electric current to induce painful and even life-threatening effects in the human body! Your first and best protection against electric shock is maintaining an insulating barrier between your body and the circuit in question, such that current from that circuit will be unable to flow through your body.

[2] By "electrical contact" I mean either directly touching an energized conductor with any part of your body, or indirectly touching it through a conductive tool. The only physical contact you should ever make with an energized conductor is via an electrically insulated tool, for example a screwdriver with an electrically insulated handle, or an insulated test probe for some instrument.

[3] Another reason for consistently enforcing this policy, even on low-voltage circuits, is due to the dangers that even some low-voltage circuits harbor. A single 12 Volt automobile battery, for example, can cause a surprising amount of damage if short-circuited simply due to the high current levels (i.e. very low internal resistance) it is capable of, even though the voltage level is too low to cause a shock through the skin. Mechanics wearing metal rings, for example, are at risk from severe burns if their rings happen to short-circuit such a battery! Furthermore, even when working on circuits that are simply too low-power (low voltage and low current) to cause any bodily harm, touching them while energized can pose a threat to the circuit components themselves. In summary, it's generally wise (and *always* a good habit to build) to "power down" *any* circuit before making contact between it and your body.

contact with high temperatures, potential chemical exposure, etc. You should never proceed with a project or experiment if you are unaware of proper tool use or lack basic protective measures (e.g. personal protective equipment such as safety glasses) against such hazards.

Some other safety-related practices should be followed as well:

- All power conductors extending outward from the project must be *firmly* strain-relieved (e.g. "cord grips" used on line power cords), so that an accidental tug or drop will not compromise circuit integrity.
- All electrical connections must be sound and appropriately made (e.g. soldered wire joints rather than twisted-and-taped; terminal blocks rather than solderless breadboards for high-current or high-voltage circuits). Use "touch-safe" terminal connections with recessed metal parts to minimize risk of accidental contact.
- Always provide overcurrent protection in any circuit you build. *Always.* This may be in the form of a fuse, a circuit breaker, and/or an electronically current-limited power supply.
- Always ensure circuit conductors are rated for more current than the overcurrent protection limit. *Always.* A fuse does no good if the wire or printed circuit board trace will "blow" before it does!
- Always bond metal enclosures to Earth ground for any line-powered circuit. *Always.* Ensuring an equipotential state between the enclosure and Earth by making the enclosure electrically common with Earth ground ensures no electric shock can occur simply by one's body bridging between the Earth and the enclosure.
- Avoid building a high-energy circuit when a low-energy circuit will suffice. For example, I always recommend beginning students power their first DC resistor circuits using small batteries rather than with line-powered DC power supplies. The intrinsic energy limitations of a dry-cell battery make accidents highly unlikely.
- Use line power receptacles that are GFCI (Ground Fault Current Interrupting) to help avoid electric shock from making accidental contact with a "hot" line conductor.
- Always wear eye protection when working with tools or live systems having the potential to eject material into the air. Examples of such activities include soldering, drilling, grinding, cutting, wire stripping, working on or near energized circuits, etc.
- Always use a step-stool or stepladder to reach high places. Never stand on something not designed to support a human load.
- When in doubt, *ask an expert.* If anything even seems remotely unsafe to you, do not proceed without consulting a trusted person fully knowledgeable in electrical safety.

9.1.2 Other helpful tips

Experience has shown the following practices to be very helpful, especially when students make their own component selections, to ensure the circuits will be well-behaved:

- Avoid resistor values less than 1 kΩ or greater than 100 kΩ, unless such values are definitely necessary[4]. Resistances below 1 kΩ may draw excessive current if directly connected to a voltage source of significant magnitude, and may also complicate the task of accurately measuring current since any ammeter's non-zero resistance inserted in series with a low-value circuit resistor will significantly alter the total resistance and thereby skew the measurement. Resistances above 100 kΩ may complicate the task of measuring voltage since any voltmeter's finite resistance connected in parallel with a high-value circuit resistor will significantly alter the total resistance and thereby skew the measurement. Similarly, AC circuit impedance values should be between 1 kΩ and 100 kΩ, and for all the same reasons.

- Ensure all electrical connections are low-resistance and physically rugged. For this reason, one should avoid *compression splices* (e.g. "butt" connectors), solderless breadboards[5], and wires that are simply twisted together.

- Build your circuit with **testing** in mind. For example, provide convenient connection points for test equipment (e.g. multimeters, oscilloscopes, signal generators, logic probes).

- Design permanent projects with **maintenance** in mind. The more convenient you make maintenance tasks, the more likely they will get done.

- **Always document and save your work**. Circuits lacking schematic diagrams are more difficult to troubleshoot than documented circuits. Similarly, circuit construction is simpler when a schematic diagram precedes construction. Experimental results are easier to interpret when comprehensively recorded. Consider modern videorecording technology for this purpose where appropriate.

- **Record your steps** when troubleshooting. **Talk to yourself** when solving problems. These simple steps clarify thought and simplify identification of errors.

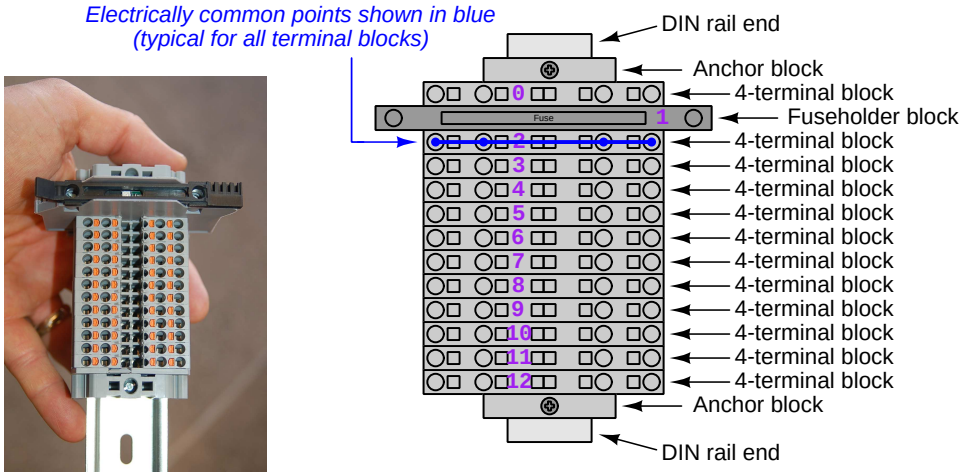
[4] An example of a necessary resistor value much less than 1 kΩ is a *shunt resistor* used to produce a small voltage drop for the purpose of sensing current in a circuit. Such shunt resistors must be low-value in order not to impose an undue load on the rest of the circuit. An example of a necessary resistor value much greater than 100 kΩ is an electrostatic *drain resistor* used to dissipate stored electric charges from body capacitance for the sake of preventing damage to sensitive semiconductor components, while also preventing a path for current that could be dangerous to the person (i.e. shock).

[5] Admittedly, solderless breadboards are very useful for constructing complex electronic circuits with many components, especially DIP-style integrated circuits (ICs), but they tend to give trouble with connection integrity after frequent use. An alternative for projects using low counts of ICs is to solder IC sockets into prototype printed circuit boards (PCBs) and run wires from the soldered pins of the IC sockets to terminal blocks where reliable temporary connections may be made.

9.1.3 Terminal blocks for circuit construction

Terminal blocks are the standard means for making electric circuit connections in industrial systems, and are also quite useful for student-built experiments, providing a robust alternative to solderless breadboards[6]. Terminal blocks provide highly reliable connections rated for significant voltage and current magnitudes, and they force the builder to think carefully about component placement which is an important mental practice. Terminal blocks designed to mount on standard 35 mm DIN rail[7] come in many types and capabilities.

Every student of electricity should build their own terminal block array consisting of several terminal blocks where each block has at least 4 connection points all electrically common to each other[8] and at least one terminal block that is a fuse holder for overcurrent protection. A pair of anchoring blocks hold all terminal blocks securely on the DIN rail, preventing them from sliding off the rail. Each of the terminals should bear a unique number starting from 0. An example is shown in the following photograph and illustration:



Screwless terminal blocks (using internal spring clips to clamp wire and component lead ends) are preferred over screw-based terminal blocks, as they reduce assembly and disassembly time, and also minimize repetitive wrist stress from twisting screwdrivers. Some screwless terminal blocks require the use of a special tool to release the spring clip, while others such as the Phoenix Contact model 3211019 blocks shown in the photograph provide buttons[9] for this task which may be pressed using the tip of any suitable tool such as a small screwdriver or multimeter test probe tip.

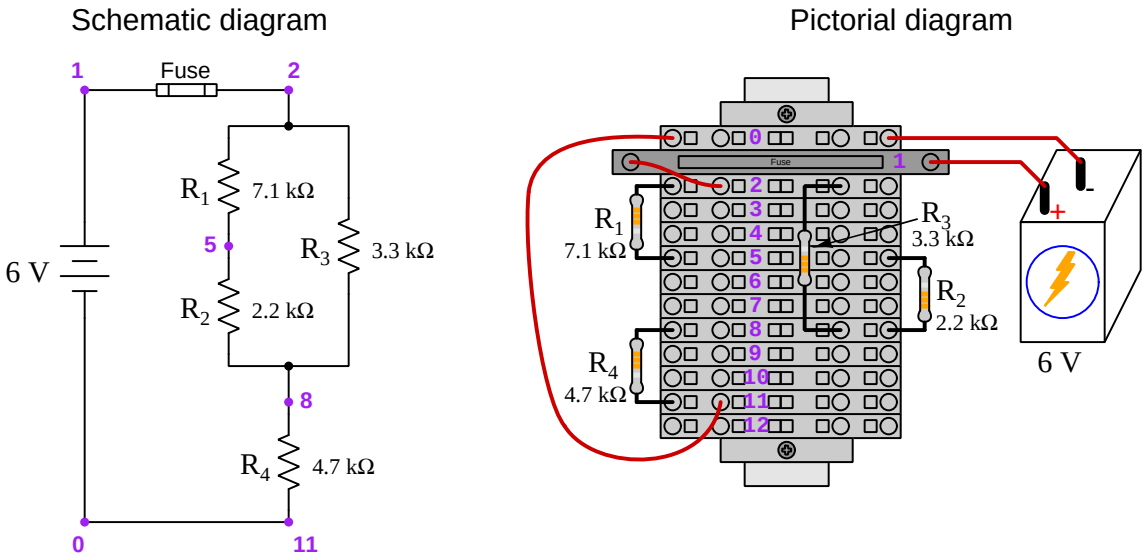
[6] Solderless breadboard are preferable for complicated electronic circuits with multiple integrated "chip" components, but for simpler circuits I find terminal blocks much more practical. An alternative to solderless breadboards for "chip" circuits is to solder chip sockets onto a PCB and then use wires to connect the socket pins to terminal blocks. This also accommodates *surface-mount* components, which solderless breadboards do not.

[7] DIN rail is a metal rail designed to serve as a mounting point for a wide range of electrical and electronic devices such as terminal blocks, fuses, circuit breakers, relay sockets, power supplies, data acquisition hardware, etc.

[8] Sometimes referred to as *equipotential*, *same-potential*, or *potential distribution* terminal blocks.

[9] The small orange-colored squares seen in the above photograph are buttons for this purpose, and may be actuated by pressing with any tool of suitable size.

The following example shows how such a terminal block array might be used to construct a series-parallel resistor circuit consisting of four resistors and a battery:



Numbering on the terminal blocks provides a very natural translation to SPICE[10] netlists, where component connections are identified by terminal number:

```
* Series-parallel resistor circuit
v1 1 0 dc 6
r1 2 5 7100
r2 5 8 2200
r3 2 8 3300
r4 8 11 4700
rfuse 1 2 0.01
rwire 0 11 0.01
.op
.end
```

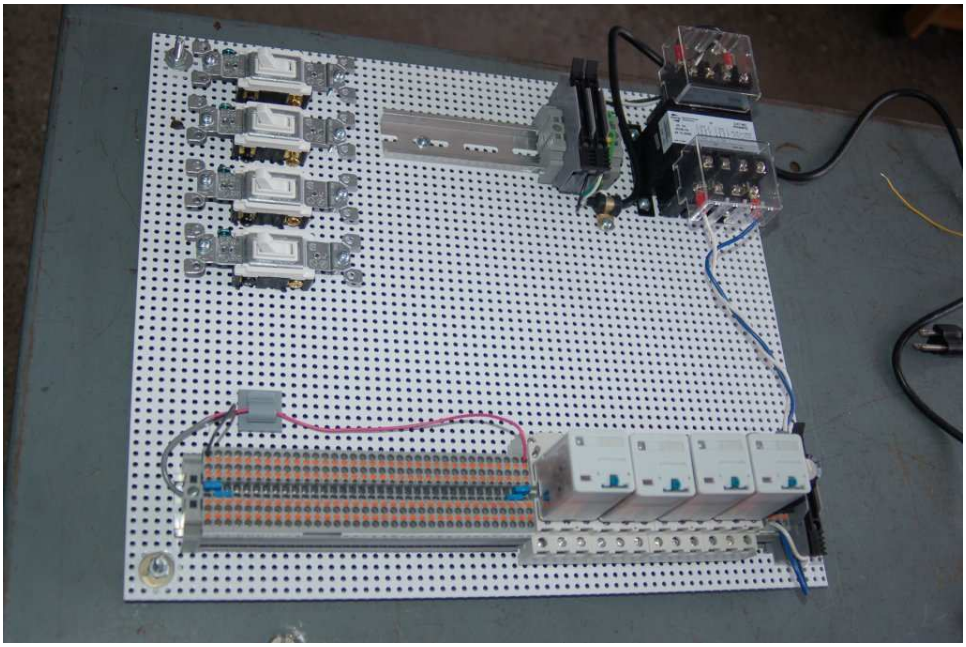
Note the use of two resistances `rfuse` and `rwire` to describe fuse resistance between terminals 1 and 2 and "jumper" wire resistance between terminals 0 and 11, respectively. Being resistances, SPICE requires a resistance value for each, and here we see they have both been set to an arbitrarily low value of 0.01 Ohm appropriate to their real-life resistance values.

Listing all components and wires along with their numbered terminals happens to be a useful documentation method for any circuit built on terminal blocks, independent of SPICE. Such a

[10] SPICE is computer software designed to analyze electrical and electronic circuits. Circuits are described for the computer in the form of *netlists* which are text files listing each component type, connection node numbers, and component values.

"wiring sequence" may be thought of as a *non-graphical description* of an electric circuit, and is exceptionally easy to follow.

An example of a more elaborate terminal block array is shown in the following photograph, with terminal blocks and "ice-cube" style electromechanical relays mounted to DIN rail, which is turn mounted to a perforated subpanel[11]. This "terminal block board" hosts an array of thirty five undedicated terminal block sections, four SPDT toggle switches, four DPDT "ice-cube" relays, a step-down control power transformer, bridge rectifier and filtering capacitor, and several fuses for overcurrent protection:



Four plastic-bottomed "feet" support the subpanel above the benchtop surface, and an unused section of DIN rail stands ready to accept other components. Safety features include electrical bonding of the AC line power cord's ground to the metal subpanel (and all metal DIN rails), mechanical strain relief for the power cord to isolate any cord tension from wire connections, clear plastic finger guards covering the transformer's screw terminals, as well as fused overcurrent protection for the 120 Volt AC line power and the transformer's 12 Volt AC output. The perforated holes happen to be on $\frac{1}{4}$ inch centers with a diameter suitable for tapping with 6-32 machine screw threads, their presence making it very easy to attach other sections of DIN rail, printed circuit boards, or specialized electrical components directly to the grounded metal subpanel. Such a "terminal block board" is an inexpensive[12] yet highly flexible means to construct physically robust circuits using industrial wiring practices.

[11] An electrical *subpanel* is a thin metal plate intended for mounting inside an electrical enclosure. Components are attached to the subpanel, and the subpanel in turn bolts inside the enclosure. Subpanels allow circuit construction outside the confines of the enclosure, which speeds assembly. In this particular usage there is no enclosure, as the subpanel is intended to be used as an open platform for the convenient construction of circuits on a benchtop by students. In essence, this is a modern version of the traditional *breadboard* which was literally a wooden board such as might be used for cutting loaves of bread, but which early electrical and electronic hobbyists used as platforms for the construction of circuits.

[12] At the time of this writing (2019) the cost to build this board is approximately $250 US dollars.

9.1.4 Conducting experiments

An *experiment* is an exploratory act, a test performed for the purpose of assessing some proposition or principle. Experiments are the foundation of the *scientific method*, a process by which careful observation helps guard against errors of speculation. All good experiments begin with an *hypothesis*, defined by the American Heritage Dictionary of the English Language as:

> An assertion subject to verification or proof, as (a) A proposition stated as a basis for argument or reasoning. (b) A premise from which a conclusion is drawn. (c) A conjecture that accounts, within a theory or ideational framework, for a set of facts and that can be used as a basis for further investigation.

Stated plainly, an hypothesis is an *educated guess* about cause and effect. The correctness of this initial guess matters little, because any well-designed experiment will reveal the truth of the matter. In fact, *incorrect* hypotheses are often the most valuable because the experiments they engender lead us to surprising discoveries. One of the beautiful aspects of science is that it is more focused on the process of *learning* than about the status of *being correct*[13]. In order for an hypothesis to be valid, it must be testable[14], which means it must be a claim possible to refute given the right data. Hypotheses impossible to critique are useless.

Once an hypothesis has been formulated, an experiment must be designed to test that hypothesis. A well-designed experiment requires careful regulation of all relevant variables, both for personal safety and for prompting the hypothesized results. If the effects of one particular variable are to be tested, the experiment must be run multiple times with different values of (only) that particular variable. The experiment set up with the "baseline" variable set is called the *control*, while the experiment set up with different value(s) is called the *test* or *experimental*.

For some hypotheses a viable alternative to a physical experiment is a *computer-simulated experiment* or even a *thought experiment*. Simulations performed on a computer test the hypothesis against the physical laws encoded within the computer simulation software, and are particularly useful for students learning new principles for which simulation software is readily available[15].

[13] Science is more about clarifying our view of the universe through a systematic process of error detection than it is about proving oneself to be right. Some scient*ists* may happen to have large egos – and this may have more to do with the ways in which large-scale scientific research is *funded* than anything else – but scientific *method* itself is devoid of ego, and if embraced as a practical philosophy is quite an effective stimulant for humility. Within the education system, scientific method is particularly valuable for helping students break free of the crippling fear of *being wrong*. So much emphasis is placed in formal education on assessing correct retention of facts that many students are fearful of saying or doing anything that might be perceived as a mistake, and of course making mistakes (i.e. having one's hypotheses disproven by experiment) is an indispensable tool for learning. Introducing science in the classroom – *real* science characterized by individuals forming actual hypotheses and testing those hypotheses by experiment – helps students become self-directed learners.

[14] This is the principle of *falsifiability*: that a scientific statement has value only insofar as it is liable to disproof given the requisite experimental evidence. Any claim that is unfalsifiable – that is, a claim which can *never* be disproven by any evidence whatsoever – could be completely wrong and we could never know it.

[15] A very pertinent example of this is learning how to analyze electric circuits using simulation software such as SPICE. A typical experimental cycle would proceed as follows: (1) Find or invent a circuit to analyze; (2) Apply your analytical knowledge to that circuit, predicting all voltages, currents, powers, etc. relevant to the concepts you are striving to master; (3) Run a simulation on that circuit, collecting "data" from the computer when complete; (4) Evaluate whether or not your hypotheses (i.e. predicted voltages, currents, etc.) agree with the computer-generated results; (5) If so, your analyses are (provisionally) correct – if not, examine your analyses and the computer simulation again to determine the source of error; (6) Repeat this process as many times as necessary until you achieve mastery.

Thought experiments are useful for detecting inconsistencies within your own understanding of some subject, rather than testing your understanding against physical reality.

Here are some general guidelines for conducting experiments:

- The clearer and more specific the hypothesis, the better. Vague or unfalsifiable hypotheses are useless because they will fit *any* experimental results, and therefore the experiment cannot teach you anything about the hypothesis.
- Collect as much data (i.e. information, measurements, sensory experiences) generated by an experiment as is practical. This includes the time and date of the experiment, too!
- *Never* discard or modify data gathered from an experiment. If you have reason to believe the data is unreliable, write notes to that effect, but never throw away data just because you think it is untrustworthy. It is quite possible that even "bad" data holds useful information, and that someone else may be able to uncover its value even if you do not.
- Prioritize *quantitative* data over *qualitative* data wherever practical. Quantitative data is more specific than qualitative, less prone to subjective interpretation on the part of the experimenter, and amenable to an arsenal of analytical methods (e.g. statistics).
- Guard against your own bias(es) by making your experimental results available to others. This allows other people to scrutinize your experimental design and collected data, for the purpose of detecting and correcting errors you may have missed. Document your experiment such that others may independently replicate it.
- Always be looking for sources of error. No physical measurement is perfect, and so it is impossible to achieve *exact* values for any variable. Quantify the amount of uncertainty (i.e. the "tolerance" of errors) whenever possible, and be sure your hypothesis does not depend on precision better than this!
- Always remember that scientific confirmation is provisional – no number of "successful" experiments will prove an hypothesis true for all time, but a single experiment can disprove it. Put into simpler terms, *truth is elusive but error is within reach.*
- Remember that scientific method is about *learning*, first and foremost. An unfortunate consequence of scientific triumph in modern society is that science is often viewed by non-practitioners as an unerring source of truth, when in fact science is an ongoing process of challenging existing ideas to probe for errors and oversights. This is why it is perfectly acceptable to have a failed hypothesis, and why the only truly failed experiment is one where nothing was learned.

The following is an example of a well-planned and executed experiment, in this case a physical experiment demonstrating Ohm's Law.

```
Planning Time/Date = 09:30 on 12 February 2019

HYPOTHESIS: the current through any resistor should be exactly proportional
to the voltage impressed across it.

PROCEDURE: connect a resistor rated 1 k Ohm and 1/4 Watt to a variable-voltage
DC power supply.  Use an ammeter in series to measure resistor current and
a voltmeter in parallel to measure resistor voltage.

RISKS AND MITIGATION: excessive power dissipation may harm the resistor and/
or pose a burn hazard, while excessive voltage poses an electric shock hazard.
30 Volts is a safe maximum voltage for laboratory practices, and according to
Joule's Law a 1000 Ohm resistor will dissipate 0.25 Watts at 15.81 Volts
(P = V^2 / R), so I will remain below 15 Volts just to be safe.

Experiment Time/Date = 10:15 on 12 February 2019

DATA COLLECTED:
  (Voltage)    (Current)         (Voltage)   (Current)
   0.000 V   =  0.000 mA          8.100    =  7.812 mA
   2.700 V   =  2.603 mA          10.00 V  =  9.643 mA
   5.400 V   =  5.206 mA          14.00 V  =  13.49 mA

Analysis Time/Date = 10:57 on 12 February 2019

ANALYSIS: current definitely increases with voltage, and although I expected
exactly one milliAmpere per Volt the actual current was usually less than
that.  The voltage/current ratios ranged from a low of 1036.87 (at 8.1 Volts)
to a high of 1037.81 (at 14 Volts), but this represents a variance of only
-0.0365% to +0.0541% from the average, indicating a very consistent
proportionality -- results consistent with Ohm's Law.

ERROR SOURCES: one major source of error is the resistor's value itself.  I
did not measure it, but simply assumed color bands of brown-black-red meant
exactly 1000 Ohms.  Based on the data I think the true resistance is closer
to 1037 Ohms.  Another possible explanation is multimeter calibration error.
However, neither explains the small positive and negative variances from the
average.  This might be due to electrical noise, a good test being to repeat
the same experiment to see if the variances are the same or different.  Noise
should generate slightly different results every time.
```

The following is an example of a well-planned and executed *virtual* experiment, in this case demonstrating Ohm's Law using a computer (SPICE) simulation.

```
Planning Time/Date = 12:32 on 14 February 2019

HYPOTHESIS: for any given resistor, the current through that resistor should be
exactly proportional to the voltage impressed across it.

PROCEDURE: write a SPICE netlist with a single DC voltage source and single
1000 Ohm resistor, then use NGSPICE version 26 to perform a "sweep" analysis
from 0 Volts to 25 Volts in 5 Volt increments.

    * SPICE circuit
    v1 1 0 dc
    r1 1 0 1000
    .dc v1 0 25 5
    .print dc v(1) i(v1)
    .end

RISKS AND MITIGATION: none.

DATA COLLECTED:
       DC transfer characteristic  Thu Feb 14 13:05:08  2019
    ------------------------------------------------------------
    Index   v-sweep         v(1)            v1#branch
    ------------------------------------------------------------
    0       0.000000e+00    0.000000e+00    0.000000e+00
    1       5.000000e+00    5.000000e+00    -5.00000e-03
    2       1.000000e+01    1.000000e+01    -1.00000e-02
    3       1.500000e+01    1.500000e+01    -1.50000e-02
    4       2.000000e+01    2.000000e+01    -2.00000e-02
    5       2.500000e+01    2.500000e+01    -2.50000e-02

Analysis Time/Date = 13:06 on 14 February 2019

ANALYSIS: perfect agreement between data and hypothesis -- current is precisely
1/1000 of the applied voltage for all values.  Anything other than perfect
agreement would have probably meant my netlist was incorrect.  The negative
current values surprised me, but it seems this is just how SPICE interprets
normal current through a DC voltage source.

ERROR SOURCES: none.
```

As gratuitous as it may seem to perform experiments on a physical law as well-established as Ohm's Law, even the examples listed previously demonstrate opportunity for real learning. In the physical experiment example, the student should identify and explain why their data does not perfectly agree with the hypothesis, and this leads them naturally to consider sources of error. In the computer-simulated experiment, the student is struck by SPICE's convention of denoting regular current through a DC voltage source as being *negative* in sign, and this is also useful knowledge for future simulations. Scientific experiments are most interesting when things *do not* go as planned!

Aside from verifying well-established physical laws, simple experiments are extremely useful as educational tools for a wide range of purposes, including:

- Component familiarization (e.g. *Which terminals of this switch connect to the NO versus NC contacts?*)
- System testing (e.g. *How heavy of a load can my AC-DC power supply source before the semiconductor components reach their thermal limits?*)
- Learning programming languages (e.g. *Let's try to set up an "up" counter function in this PLC!*)

Above all, the priority here is to inculcate the habit of hypothesizing, running experiments, and analyzing the results. This experimental cycle not only serves as an excellent method for self-directed learning, but it also works exceptionally well for troubleshooting faults in complex systems, and for these reasons should be a part of every technician's and every engineer's education.

9.1.5 Constructing projects

Designing, constructing, and testing projects is a very effective means of practical education. Within a formal educational setting, projects are generally chosen (or at least vetted) by an instructor to ensure they may be reasonably completed within the allotted time of a course or program of study, and that they sufficiently challenge the student to learn certain important principles. In a self-directed environment, projects are just as useful as a learning tool but there is some risk of unwittingly choosing a project beyond one's abilities, which can lead to frustration.

Here are some general guidelines for managing projects:

- Define your goal(s) before beginning a project: what do you wish to achieve in building it? What, exactly, should the completed project *do*?
- Analyze your project prior to construction. Document it in appropriate forms (e.g. schematic diagrams), predict its functionality, anticipate all associated risks. In other words, *plan ahead.*
- Set a reasonable budget for your project, and stay within it.
- Identify any deadlines, and set reasonable goals to meet those deadlines.
- Beware of *scope creep*: the tendency to modify the project's goals before it is complete.
- Document your progress! An easy way to do this is to use photography or videography: take photos and/or videos of your project as it progresses. Document failures as well as successes, because both are equally valuable from the perspective of learning.

9.2 Experiment: JFET terminal identification

Conduct an experiment to identify the three terminals of a JFET using nothing but an ohmmeter.

EXPERIMENT CHECKLIST:

- Prior to experimentation:
  - ☑ Write an hypothesis (i.e. a detailed description of what you expect will happen) unambiguous enough that it could be disproven given the right data.
  - ☑ Write a procedure to test the hypothesis, complete with adequate controls and documentation (e.g. schematic diagrams, programming code).
  - ☑ Identify any risks (e.g. shock hazard, component damage) and write a mitigation plan based on best practices and component ratings.
- During experimentation:
  - ☑ Safe practices followed at all times (e.g. no contact with energized circuit).
  - ☑ Correct equipment usage according to manufacturer's recommendations.
  - ☑ All data collected, ideally quantitative with full precision (i.e. no rounding).
- After each experimental run:
  - ☑ If the results fail to match the hypothesis, identify the error(s), correct the hypothesis and/or revise the procedure, and re-run the experiment.
  - ☑ Identify any uncontrolled sources of error in the experiment.
- After all experimental re-runs:
  - ☑ Save all data for future reference.
  - ☑ Write an analysis of experimental results and lessons learned.

Challenges

- Science is an *iterative* process, and for this reason is never complete. Following the results of your experiment, what would you propose for your *next* hypothesis and *next* experimental procedure? Hint: if your experiment produced any unexpected results, exploring those unexpected results is often a very good basis for the next experiment!

9.3 Experiment: MOSFET terminal identification

Conduct an experiment to identify the three terminals of an E-type MOSFET using nothing but an ohmmeter.

EXPERIMENT CHECKLIST:

- Prior to experimentation:
  - ☑ Write an hypothesis (i.e. a detailed description of what you expect will happen) unambiguous enough that it could be disproven given the right data.
  - ☑ Write a procedure to test the hypothesis, complete with adequate controls and documentation (e.g. schematic diagrams, programming code).
  - ☑ Identify any risks (e.g. shock hazard, component damage) and write a mitigation plan based on best practices and component ratings.
- During experimentation:
  - ☑ Safe practices followed at all times (e.g. no contact with energized circuit).
  - ☑ Correct equipment usage according to manufacturer's recommendations.
  - ☑ All data collected, ideally quantitative with full precision (i.e. no rounding).
- After each experimental run:
  - ☑ If the results fail to match the hypothesis, identify the error(s), correct the hypothesis and/or revise the procedure, and re-run the experiment.
  - ☑ Identify any uncontrolled sources of error in the experiment.
- After all experimental re-runs:
  - ☑ Save all data for future reference.
  - ☑ Write an analysis of experimental results and lessons learned.

Challenges

- Science is an *iterative* process, and for this reason is never complete. Following the results of your experiment, what would you propose for your *next* hypothesis and *next* experimental procedure? Hint: if your experiment produced any unexpected results, exploring those unexpected results is often a very good basis for the next experiment!

9.4 Experiment: JFET as a controlled resistance

Conduct an experiment to demonstrate a JFET's ability to behave as a voltage-controlled *resistance*.

EXPERIMENT CHECKLIST:

- Prior to experimentation:
  - ☑ Write an hypothesis (i.e. a detailed description of what you expect will happen) unambiguous enough that it could be disproven given the right data.
  - ☑ Write a procedure to test the hypothesis, complete with adequate controls and documentation (e.g. schematic diagrams, programming code).
  - ☑ Identify any risks (e.g. shock hazard, component damage) and write a mitigation plan based on best practices and component ratings.
- During experimentation:
  - ☑ Safe practices followed at all times (e.g. no contact with energized circuit).
  - ☑ Correct equipment usage according to manufacturer's recommendations.
  - ☑ All data collected, ideally quantitative with full precision (i.e. no rounding).
- After each experimental run:
  - ☑ If the results fail to match the hypothesis, identify the error(s), correct the hypothesis and/or revise the procedure, and re-run the experiment.
  - ☑ Identify any uncontrolled sources of error in the experiment.
- After all experimental re-runs:
  - ☑ Save all data for future reference.
  - ☑ Write an analysis of experimental results and lessons learned.

Challenges

- Science is an *iterative* process, and for this reason is never complete. Following the results of your experiment, what would you propose for your *next* hypothesis and *next* experimental procedure? Hint: if your experiment produced any unexpected results, exploring those unexpected results is often a very good basis for the next experiment!

9.5 Experiment: JFET as a controlled current regulator

Conduct an experiment to demonstrate a JFET's ability to behave as a voltage-controlled *current regulator*.

EXPERIMENT CHECKLIST:

- Prior to experimentation:
  - ☑ Write an hypothesis (i.e. a detailed description of what you expect will happen) unambiguous enough that it could be disproven given the right data.
  - ☑ Write a procedure to test the hypothesis, complete with adequate controls and documentation (e.g. schematic diagrams, programming code).
  - ☑ Identify any risks (e.g. shock hazard, component damage) and write a mitigation plan based on best practices and component ratings.
- During experimentation:
  - ☑ Safe practices followed at all times (e.g. no contact with energized circuit).
  - ☑ Correct equipment usage according to manufacturer's recommendations.
  - ☑ All data collected, ideally quantitative with full precision (i.e. no rounding).
- After each experimental run:
  - ☑ If the results fail to match the hypothesis, identify the error(s), correct the hypothesis and/or revise the procedure, and re-run the experiment.
  - ☑ Identify any uncontrolled sources of error in the experiment.
- After all experimental re-runs:
  - ☑ Save all data for future reference.
  - ☑ Write an analysis of experimental results and lessons learned.

Challenges

- Science is an *iterative* process, and for this reason is never complete. Following the results of your experiment, what would you propose for your *next* hypothesis and *next* experimental procedure? Hint: if your experiment produced any unexpected results, exploring those unexpected results is often a very good basis for the next experiment!

9.6 Experiment: measuring switching time

Conduct an experiment to measure how *quickly* a MOSFET is able to turn on and turn off.

EXPERIMENT CHECKLIST:

- Prior to experimentation:
  - ☑ Write an hypothesis (i.e. a detailed description of what you expect will happen) unambiguous enough that it could be disproven given the right data.
  - ☑ Write a procedure to test the hypothesis, complete with adequate controls and documentation (e.g. schematic diagrams, programming code).
  - ☑ Identify any risks (e.g. shock hazard, component damage) and write a mitigation plan based on best practices and component ratings.
- During experimentation:
  - ☑ Safe practices followed at all times (e.g. no contact with energized circuit).
  - ☑ Correct equipment usage according to manufacturer's recommendations.
  - ☑ All data collected, ideally quantitative with full precision (i.e. no rounding).
- After each experimental run:
  - ☑ If the results fail to match the hypothesis, identify the error(s), correct the hypothesis and/or revise the procedure, and re-run the experiment.
  - ☑ Identify any uncontrolled sources of error in the experiment.
- After all experimental re-runs:
  - ☑ Save all data for future reference.
  - ☑ Write an analysis of experimental results and lessons learned.

Challenges

- Science is an *iterative* process, and for this reason is never complete. Following the results of your experiment, what would you propose for your *next* hypothesis and *next* experimental procedure? Hint: if your experiment produced any unexpected results, exploring those unexpected results is often a very good basis for the next experiment!

Appendix A

Problem-Solving Strategies

The ability to solve complex problems is arguably one of the most valuable skills one can possess, and this skill is particularly important in any science-based discipline.

- Study principles, not procedures. Don't be satisfied with merely knowing how to compute solutions – learn *why* those solutions work.
- Identify what it is you need to solve, identify all relevant data, identify all units of measurement, identify any general principles or formulae linking the given information to the solution, and then identify any "missing pieces" to a solution. Annotate all diagrams with this data.
- Sketch a diagram to help visualize the problem. When building a real system, always devise a plan for that system and analyze its function *before* constructing it.
- Follow the units of measurement and meaning of every calculation. If you are ever performing mathematical calculations as part of a problem-solving procedure, and you find yourself unable to apply each and every intermediate result to some aspect of the problem, it means you don't understand what you are doing. Properly done, every mathematical result should have practical meaning for the problem, and not just be an abstract number. You should be able to identify the proper units of measurement for each and every calculated result, and show where that result fits into the problem.
- Perform "thought experiments" to explore the effects of different conditions for theoretical problems. When troubleshooting real systems, perform *diagnostic tests* rather than visually inspecting for faults, the best diagnostic test being the one giving you the most information about the nature and/or location of the fault with the fewest steps.
- Simplify the problem until the solution becomes obvious, and then use that obvious case as a model to follow in solving the more complex version of the problem.
- Check for exceptions to see if your solution is incorrect or incomplete. A good solution will work for *all* known conditions and criteria. A good example of this is the process of testing scientific hypotheses: the task of a scientist is not to find support for a new idea, but rather to *challenge* that new idea to see if it holds up under a battery of tests. The philosophical

principle of *reductio ad absurdum* (i.e. disproving a general idea by finding a specific case where it fails) is useful here.

- Work "backward" from a hypothetical solution to a new set of given conditions.
- Add quantities to problems that are qualitative in nature, because sometimes a little math helps illuminate the scenario.
- Sketch graphs illustrating how variables relate to each other. These may be quantitative (i.e. with realistic number values) or qualitative (i.e. simply showing increases and decreases).
- Treat quantitative problems as qualitative in order to discern the relative magnitudes and/or directions of change of the relevant variables. For example, try determining what happens if a certain variable were to increase or decrease before attempting to precisely calculate quantities: how will each of the dependent variables respond, by increasing, decreasing, or remaining the same as before?
- Consider limiting cases. This works especially well for qualitative problems where you need to determine which direction a variable will change. Take the given condition and magnify that condition to an extreme degree as a way of simplifying the direction of the system's response.
- Check your work. This means regularly testing your conclusions to see if they make sense. This does *not* mean repeating the same steps originally used to obtain the conclusion(s), but rather to use some other means to check validity. Simply repeating procedures often leads to *repeating the same errors* if any were made, which is why alternative paths are better.

Appendix B

Instructional philosophy

B.1 First principles of learning

- **Anyone can learn anything** given appropriate time, effort, resources, challenges, encouragement, and expectations. Dedicating time and investing effort are the student's responsibility; providing resources, challenges, and encouragement are the teacher's responsibility; high expectations are a responsibility shared by both student and teacher.

- **Transfer is not automatic.** The human mind has a natural tendency to compartmentalize information, which means the process of taking knowledge learned in one context and applying it to another usually does not come easy and therefore should never be taken for granted.

- **Learning is iterative.** The human mind rarely learns anything perfectly on the first attempt. Anticipate mistakes and plan for multiple tries to achieve full understanding, using the lessons of those mistakes as feedback to guide future attempts.

- **Information is absorbed, but understanding is created.** Facts and procedures may be memorized easily enough by repeated exposure, but the ability to reliably apply principles to novel scenarios only comes through intense personal effort. This effort is fundamentally creative in nature: casting new concepts in one's own words, running mental experiments to perceive patterns and trends, etc. These acts of making knowledge "one's own" need not be perfect in order to be effective, as the value lies in the activity and not necessarily the finished product. Depth of understanding and permanence of memory are both directly related to mental effort during the learning process – i.e. *easy learning is shallow and fleeting.*

- **Education trumps training.** There is no such thing as an entirely isolated subject, as all fields of knowledge are connected. Training is narrowly-focused and task-oriented. Education is broad-based and principle-oriented. When preparing for a life-long technical career, education beats training every time.

- **Character matters.** Poor habits are more destructive than deficits of knowledge or skill. This is especially true in collective endeavors, where a team's ability to function depends on trust between its members. Simply put, no one wants an untrustworthy person on their team. An essential component of education then, is character development.

- **People learn to be responsible by bearing responsibility.** An irresponsible person is someone who has never *had* to be responsible for anything that mattered enough to them. Just as anyone can learn anything, anyone can become responsible if the personal cost of irresponsibility becomes high enough.

- **What gets measured, gets done.** Accurate and relevant assessment of learning is key to ensuring all students learn. Therefore, it is imperative to measure what matters.

- **Failure is nothing to fear.** Every human being fails, and fails in multiple ways at multiple times. Eventual success only happens when we don't stop trying.

B.2 Proven strategies for instructors

- Assume every student is capable of learning anything they desire given the proper conditions. Treat them as capable adults by granting real responsibility and avoiding artificial incentives such as merit or demerit points.
- Create a consistent culture of high expectations across the entire program of study. Demonstrate and encourage patience, persistence, and a healthy sense of self-skepticism. Anticipate and de-stigmatize error. Teach respect for the capabilities of others as well as respect for one's own fallibility.
- Replace lecture with "inverted" instruction, where students first encounter new concepts through reading and then spend class time in Socratic dialogue with the instructor exploring those concepts and solving problems individually. There is a world of difference between observing someone solve a problem versus actually solving a problem yourself, and so the point of this form of instruction is to place students in a position where they *cannot* passively observe.
- Require students to read extensively, write about what they learn, and dialogue with you and their peers to sharpen their understanding. Apply Francis Bacon's advice that "reading maketh a full man; conference a ready man; and writing an exact man". These are complementary activities helping students expand their confidence and abilities.
- Use artificial intelligence (AI) to challenge student understanding rather than merely provide information. Find productive ways for AI to critique students' clarity of thought and of expression, for example by employing AI as a Socratic-style interlocutor or as a reviewer of students' journals. Properly applied, AI has the ability to expand student access to critical review well outside the bounds of their instructor's reach.
- Build frequent and rapid feedback into the learning process so that students know at all times how well they are learning, to identify problems early and fix them before they grow. Model the intellectual habit of self-assessing and self-correcting your own understanding (i.e. a cognitive *feedback loop*), encouraging students to do the same.
- Use "mastery" as the standard for every assessment, which means the exam or experiment or project must be done with 100% competence in order to pass. Provide students with multiple opportunity for re-tries (different versions of the assessment every time).
- Require students to devise their own hypotheses and procedures on all experiments, so that the process is truly a scientific one. Have students assess their proposed experimental procedures for risk and devise mitigations for those risks. Let nothing be pre-designed about students' experiments other than a stated task (i.e. what principle the experiment shall test) at the start and a set of demonstrable knowledge and skill objectives at the end.
- Have students build as much of their lab equipment as possible: building power sources, building test assemblies[1], and building complete working systems (no kits!). In order to provide

[1] In the program I teach, every student builds their own "Development Board" consisting of a metal chassis with DIN rail, terminal blocks, and an AC-DC power supply of their own making which functions as a portable lab environment they can use at school as well as take home.

this same "ground-up" experience for every new student, this means either previous students take their creations with them, or the systems get disassembled in preparation for the new students, or the systems grow and evolve with each new student group.

- Incorporate external accountability for you and for your students, continuously improving the curriculum and your instructional methods based on proven results. Have students regularly network with active professionals through participation in advisory committee meetings, service projects, tours, jobshadows, internships, etc. Practical suggestions include requiring students to design and build projects for external clients (e.g. community groups, businesses, different departments within the institution), and also requiring students attend all technical advisory committee meetings and to dialogue with the industry representatives at those meetings.
- Repeatedly explore difficult-to-learn concepts across multiple courses, so that students have multiple opportunities to build their understanding.
- Relate all new concepts, whenever possible, to previous concepts and to relevant physical laws. Challenge each and every student, every day, to *reason* from concept to concept and to explain the logical connections between. Challenge students to verify their conclusions by multiple approaches (e.g. double-checking their work using different methods). Ask *"Why?"* often.
- Maintain detailed records on each student's performance and share these records privately with them. These records should include academic performance as well as professionally relevant behavioral tendencies.
- Hold mandatory "check-in" meetings between all program faculty and each new student during their first term. Offer these to all other students as an option, except for any students continuing to manifest unprofessional behaviors, poor academic performance, or who have some other need for a face-to-face meeting with faculty.
- Address problems while they are small, before they grow larger. This is equally true for tutoring technical concepts as it is for helping students build professional habits.
- Build rigorous quality control into the curriculum to ensure every student masters every important concept, and that the mastery is retained over time. This includes (1) review questions added to every exam to re-assess knowledge taught in previous terms, (2) cumulative exams at the end of every term to re-assess all important concepts back to the very beginning of the program, and (3) review assessments in practical (hands-on) coursework to ensure critically-important skills were indeed taught and are still retained. What you will find by doing this is that it actually boosts retention of students by ensuring that important knowledge gets taught and is retained over long spans of time. In the absence of such quality control, student learning and retention tends to be spotty and this contributes to drop-out and failure rates later in their education.
- Finally, *never rush learning.* Education is not a race. Give your students ample time to digest complex ideas, as you continually remind yourself of just how long it took you to achieve mastery! Long-term retention and the consistently correct application of concepts are always the result of *focused effort over long periods of time* which means there are no shortcuts to learning.

B.3 Proven strategies for students

The single most important piece of advice I have for any student of any subject is to take responsibility for your own development in all areas of life including mental development. Expecting others in your life to entirely guide your own development is a recipe for disappointment. This is just as true for students enrolled in formal learning institutions as it is for auto-didacts pursuing learning entirely on their own. Learning to think in new ways is key to being able to gainfully use information, to make informed decisions about your life, and to best serve those you care about. With this in mind, I offer the following advice to students:

- **Approach all learning as valuable**. No matter what course you take, no matter who you learn from, no matter the subject, there is something useful in every learning experience. If you don't see the value of every new experience, you are not looking closely enough!
- **Continually challenge yourself**. Let other people take shortcuts and find easy answers to easy problems. The purpose of education is to stretch your mind, in order to shape it into a more powerful tool. This doesn't come by taking the path of least resistance. An excellent analogy for an empowering education is productive physical exercise: becoming stronger, more flexible, and more persistent only comes through intense personal effort.
- **Master the use of language**. This includes reading extensively, writing every day, listening closely, and speaking articulately. To a great extent language channels and empowers thought, so the better you are at wielding language the better you will be at grasping abstract concepts and articulating them not only for your benefit but for others as well.
- **Do not limit yourself to the resources given to you**. Read books that are not on the reading list. Run experiments that aren't assigned to you. Form study groups outside of class. Take an entrepreneurial approach to your own education, as though it were a business you were building for your future benefit.
- **Express and share what you learn**. Take every opportunity to teach what you have learned to others, as this will not only help them but will also strengthen your own understanding[2].
- Realize that **no one can give you understanding**, just as no one can give you physical fitness. These both must be *built*.
- **Above all, recognize that learning is hard work, and that a certain level of frustration is unavoidable**. There are times when you will struggle to grasp some of these concepts, and that struggle is a natural thing. Take heart that it will yield with persistent and varied[3] effort, and never give up! That concepts don't immediately come to you is not a sign of something wrong, but rather of something right: that you have found a worthy challenge!

[2] On a personal note, I was surprised to learn just how much my own understanding of electronics and related subjects was strengthened by becoming a teacher. When you are tasked every day with helping other people grasp complex topics, it catalyzes your own learning by giving you powerful incentives to study, to articulate your thoughts, and to reflect deeply on the process of learning.

[3] As the old saying goes, "Insanity is trying the same thing over and over again, expecting different results." If you find yourself stumped by something in the text, you should attempt a different approach. Alter the thought experiment, change the mathematical parameters, do whatever you can to see the problem in a slightly different light, and then the solution will often present itself more readily.

B.4 Design of these learning modules

"The unexamined circuit is not worth energizing" – Socrates (if he had taught electricity)

These learning modules, although useful for self-study, were designed to be used in a formal learning environment where a subject-matter expert challenges students to digest the content and exercise their critical thinking abilities in the answering of questions and in the construction and testing of working circuits. Every effort has been made to embed the following instructional and assessment philosophies within:

- The first goal of education is to enhance clear and independent thought, in order that every student reach their fullest potential in a highly complex and inter-dependent world. Robust reasoning is *always* more important than particulars of any subject matter, because its application is universal.
- Literacy is fundamental to independent learning and thought because text continues to be the most efficient way to communicate complex ideas over space and time. Those who cannot read with ease are limited in their ability to acquire knowledge and perspective.
- Articulate communication is fundamental to work that is complex and interdisciplinary.
- Faulty assumptions and poor reasoning are best corrected through challenge, not presentation. The rhetorical technique of *reductio ad absurdum* (disproving an assertion by exposing an absurdity) works well to discipline student's minds, not only to correct the problem at hand but also to learn how to detect and correct future errors.
- Important principles should be repeatedly explored and widely applied throughout a course of study, not only to reinforce their importance and help ensure their mastery, but also to showcase the interconnectedness and utility of knowledge.

These learning modules were expressly designed to be used in an "inverted" teaching environment[4] where students first read the introductory and tutorial chapters on their own, then individually attempt to answer the questions and construct working circuits according to the experiment and project guidelines. The instructor never lectures, but instead meets regularly with each individual student to review their progress, answer questions, identify misconceptions, and challenge the student to new depths of understanding through further questioning. Regular meetings between instructor and student should resemble a Socratic[5] dialogue, where questions serve as scalpels to dissect topics and expose assumptions. The student passes each module only after consistently demonstrating their ability to logically analyze and correctly apply all major concepts in each question or project/experiment. The instructor must be vigilant in probing each student's understanding to ensure they are truly *reasoning* and not just *memorizing*. This is why "Challenge" points appear throughout, as prompts for students to think deeper about topics and as starting points for instructor queries. Sometimes these challenge points require additional knowledge that hasn't been covered in the series to answer in full. This is okay, as the major purpose of the Challenges is to stimulate analysis and synthesis on the part of each student.

The instructor must possess enough mastery of the subject matter and awareness of students' reasoning to generate their own follow-up questions to practically any student response. Even completely correct answers given by the student should be challenged by the instructor for the purpose of having students practice articulating their thoughts and defending their reasoning. Conceptual errors committed by the student should be exposed and corrected not by direct instruction, but rather by reducing the errors to an absurdity[6] through well-chosen questions and thought experiments posed by the instructor. Becoming proficient at this style of instruction requires time and dedication, but the positive effects on critical thinking for both student and instructor are spectacular.

An inspection of these learning modules reveals certain unique characteristics. One of these is a bias toward thorough explanations in the tutorial chapters. Without a live instructor to explain concepts and applications to students, the text itself must fulfill this role. This philosophy results in lengthier explanations than what you might typically find in a textbook, each step of the reasoning process fully explained, including footnotes addressing common questions and concerns students raise while learning these concepts. Each tutorial seeks to not only explain each major concept in sufficient detail, but also to explain the logic of each concept and how each may be developed

[4] In a traditional teaching environment, students first encounter new information via *lecture* from an expert, and then independently apply that information via *homework*. In an "inverted" course of study, students first encounter new information via *homework*, and then independently apply that information under the scrutiny of an expert. The expert's role in lecture is to simply *explain*, but the expert's role in an inverted session is to *challenge*, *critique*, and if necessary *explain* where gaps in understanding still exist.

[5] Socrates is a figure in ancient Greek philosophy famous for his unflinching style of questioning. Although he authored no texts, he appears as a character in Plato's many writings. The essence of Socratic philosophy is to leave no question unexamined and no point of view unchallenged. While purists may argue a topic such as electric circuits is too narrow for a true Socratic-style dialogue, I would argue that the essential thought processes involved with scientific reasoning on *any* topic are not far removed from the Socratic ideal, and that students of electricity and electronics would do very well to challenge assumptions, pose thought experiments, identify fallacies, and otherwise employ the arsenal of critical thinking skills modeled by Socrates.

[6] This rhetorical technique is known by the Latin phrase *reductio ad absurdum*. The concept is to expose errors by counter-example, since only one solid counter-example is necessary to disprove a universal claim. As an example of this, consider the common misconception among beginning students of electricity that voltage cannot exist without current. One way to apply *reductio ad absurdum* to this statement is to ask how much current passes through a fully-charged battery connected to nothing (i.e. a clear example of voltage existing without current).

from "first principles". Again, this reflects the goal of developing clear and independent thought in students' minds, by showing how clear and logical thought was used to forge each concept. Students benefit from witnessing a model of clear thinking in action, and these tutorials strive to be just that.

Another feature of these learning modules is that they do not treat topics in isolation. Rather, important concepts are introduced early in the series, and appear repeatedly as stepping-stones toward other concepts in subsequent modules. This helps to avoid the "compartmentalization" of knowledge, demonstrating the inter-connectedness of concepts and simultaneously reinforcing them. Each module is fairly complete in itself, reserving the beginning of its tutorial to a review of foundational concepts.

To high standards of education,

Tony R. Kuphaldt

Appendix C

Tools used

I am indebted to the developers of many open-source software applications in the creation of these learning modules. The following is a list of these applications with some commentary on each.

You will notice a theme common to many of these applications: a bias toward *code*. Although I am by no means an expert programmer in any computer language, I understand and appreciate the flexibility offered by code-based applications where the user (you) enters commands into a plain ASCII text file, which the software then reads and processes to create the final output. Code-based computer applications are by their very nature *extensible*, while WYSIWYG (What You See Is What You Get) applications are generally limited to whatever user interface the developer makes for you.

The `GNU/Linux` computer operating system

> There is so much to be said about Linus Torvalds' `Linux` and Richard Stallman's `GNU` project. First, to credit just these two individuals is to fail to do justice to the *mob* of passionate volunteers who contributed to make this amazing software a reality. I first learned of `Linux` back in 1996, and have been using this operating system on my personal computers almost exclusively since then. It is *free*, it is completely *configurable*, and it permits the continued use of highly efficient `Unix` applications and scripting languages (e.g. shell scripts, Makefiles, `sed`, `awk`) developed over many decades. `Linux` not only provided me with a powerful computing platform, but its open design served to inspire my life's work of creating open-source educational resources.

Bram Moolenaar's `Vim` text editor

> Writing code for any code-based computer application requires a *text editor*, which may be thought of as a word processor strictly limited to outputting plain-ASCII text files. Many good text editors exist, and one's choice of text editor seems to be a deeply personal matter within the programming world. I prefer `Vim` because it operates very similarly to `vi` which is ubiquitous on `Unix`/`Linux` operating systems, and because it may be entirely operated via keyboard (i.e. no mouse required) which makes it fast to use.

Donald Knuth's TEX typesetting system

Developed in the late 1970's and early 1980's by computer scientist extraordinaire Donald Knuth to typeset his multi-volume magnum opus *The Art of Computer Programming*, this software allows the production of formatted text for screen-viewing or paper printing, all by writing plain-text code to describe how the formatted text is supposed to appear. TEX is not just a markup language for documents, but it is also a Turing-complete programming language in and of itself, allowing useful algorithms to be created to control the production of documents. Simply put, TEX *is a programmer's approach to word processing*. Since TEX is controlled by code written in a plain-text file, this means anyone may read that plain-text file to see exactly how the document was created. This openness afforded by the code-based nature of TEX makes it relatively easy to learn how other people have created their own TEX documents. By contrast, examining a beautiful document created in a conventional WYSIWYG word processor such as Microsoft `Word` suggests nothing to the reader about *how* that document was created, or what the user might do to create something similar. As Mr. Knuth himself once quipped, conventional word processing applications should be called WYSIAYG (What You See Is *All* You Get).

Leslie Lamport's LATEX extensions to TEX

Like all true programming languages, TEX is inherently extensible. So, years after the release of TEX to the public, Leslie Lamport decided to create a massive extension allowing easier compilation of book-length documents. The result was LATEX, which is the markup language used to create all ModEL module documents. You could say that TEX is to LATEX as `C` is to `C++`. This means it is permissible to use any and all TEX commands within LATEX source code, and it all still works. Some of the features offered by LATEX that would be challenging to implement in TEX include automatic index and table-of-content creation.

Tim Edwards' `Xcircuit` drafting program

This wonderful program is what I use to create all the schematic diagrams and illustrations (but not photographic images or mathematical plots) throughout the ModEL project. It natively outputs PostScript format which is a true vector graphic format (this is why the images do not pixellate when you zoom in for a closer view), and it is so simple to use that I have never had to read the manual! Object libraries are easy to create for `Xcircuit`, being plain-text files using PostScript programming conventions. Over the years I have collected a large set of object libraries useful for drawing electrical and electronic schematics, pictorial diagrams, and other technical illustrations.

`Gimp` graphic image manipulation program

Essentially an open-source clone of Adobe's `PhotoShop`, I use `Gimp` to resize, crop, and convert file formats for all of the photographic images appearing in the ModEL modules. Although `Gimp` does offer its own scripting language (called `Script-Fu`), I have never had occasion to use it. Thus, my utilization of `Gimp` to merely crop, resize, and convert graphic images is akin to using a sword to slice bread.

`SPICE` circuit simulation program

`SPICE` is to circuit analysis as TEX is to document creation: it is a form of markup language designed to describe a certain object to be processed in plain-ASCII text. When the plain-text "source file" is compiled by the software, it outputs the final result. More modern circuit analysis tools certainly exist, but I prefer `SPICE` for the following reasons: it is *free*, it is *fast*, it is *reliable*, and it is a fantastic tool for *teaching* students of electricity and electronics how to write simple code. I happen to use rather old versions of `SPICE`, version 2g6 being my "go to" application when I only require text-based output. `NGSPICE` (version 26), which is based on Berkeley `SPICE` version 3f5, is used when I require graphical output for such things as time-domain waveforms and Bode plots. In all `SPICE` example netlists I strive to use coding conventions compatible with all `SPICE` versions.

Andrew D. Hwang's `ePiX` mathematical visualization programming library

This amazing project is a `C++` library you may link to any `C/C++` code for the purpose of generating PostScript graphic images of mathematical functions. As a completely free and open-source project, it does all the plotting I would otherwise use a Computer Algebra System (CAS) such as `Mathematica` or `Maple` to do. It should be said that `ePiX` is *not* a Computer Algebra System like `Mathematica` or `Maple`, but merely a mathematical *visualization* tool. In other words, it won't determine integrals for you (you'll have to implement that in your own `C/C++` code!), but it can graph the results, and it does so beautifully. What I really admire about `ePiX` is that it is a `C++` programming library, which means it builds on the existing power and toolset available with that programming language. Mr. Hwang could have probably developed his own stand-alone application for mathematical plotting, but by creating a `C++` library to do the same thing he accomplished something much greater.

`gnuplot` mathematical visualization software

Another open-source tool for mathematical visualization is `gnuplot`. Interestingly, this tool is *not* part of Richard Stallman's GNU project, its name being a coincidence. For this reason the authors prefer "gnu" *not* be capitalized at all to avoid confusion. This is a much "lighter-weight" alternative to a spreadsheet for plotting tabular data, and the fact that it easily outputs directly to an X11 console or a file in a number of different graphical formats (including PostScript) is very helpful. I typically set my `gnuplot` output format to default (X11 on my Linux PC) for quick viewing while I'm developing a visualization, then switch to PostScript file export once the visual is ready to include in the document(s) I'm writing. As with my use of `Gimp` to do rudimentary image editing, my use of `gnuplot` only scratches the surface of its capabilities, but the important points are that it's *free* and that it *works well*.

`Python` programming language

Both Python and C++ find extensive use in these modules as instructional aids and exercises, but I'm listing Python here as a *tool* for myself because I use it almost daily as a *calculator*. If you open a Python interpreter console and type `from math import *` you can type mathematical expressions and have it return results just as you would on a hand calculator. Complex-number (i.e. *phasor*) arithmetic is similarly supported if you include the complex-math library (`from cmath import *`). Examples of this are shown in the Programming References chapter (if included) in each module. Of course, being a fully-featured programming language, Python also supports conditionals, loops, and other structures useful for calculation of quantities. Also, running in a console environment where all entries and returned values show as text in a chronologically-ordered list makes it easy to copy-and-paste those calculations to document exactly how they were performed.

Appendix D

Creative Commons License

Creative Commons Attribution 4.0 International Public License

By exercising the Licensed Rights (defined below), You accept and agree to be bound by the terms and conditions of this Creative Commons Attribution 4.0 International Public License ("Public License"). To the extent this Public License may be interpreted as a contract, You are granted the Licensed Rights in consideration of Your acceptance of these terms and conditions, and the Licensor grants You such rights in consideration of benefits the Licensor receives from making the Licensed Material available under these terms and conditions.

Section 1 – Definitions.

a. **Adapted Material** means material subject to Copyright and Similar Rights that is derived from or based upon the Licensed Material and in which the Licensed Material is translated, altered, arranged, transformed, or otherwise modified in a manner requiring permission under the Copyright and Similar Rights held by the Licensor. For purposes of this Public License, where the Licensed Material is a musical work, performance, or sound recording, Adapted Material is always produced where the Licensed Material is synched in timed relation with a moving image.

b. **Adapter's License** means the license You apply to Your Copyright and Similar Rights in Your contributions to Adapted Material in accordance with the terms and conditions of this Public License.

c. **Copyright and Similar Rights** means copyright and/or similar rights closely related to copyright including, without limitation, performance, broadcast, sound recording, and Sui Generis Database Rights, without regard to how the rights are labeled or categorized. For purposes of this Public License, the rights specified in Section 2(b)(1)-(2) are not Copyright and Similar Rights.

d. **Effective Technological Measures** means those measures that, in the absence of proper authority, may not be circumvented under laws fulfilling obligations under Article 11 of the WIPO Copyright Treaty adopted on December 20, 1996, and/or similar international agreements.

e. **Exceptions and Limitations** means fair use, fair dealing, and/or any other exception or

limitation to Copyright and Similar Rights that applies to Your use of the Licensed Material.

f. **Licensed Material** means the artistic or literary work, database, or other material to which the Licensor applied this Public License.

g. **Licensed Rights** means the rights granted to You subject to the terms and conditions of this Public License, which are limited to all Copyright and Similar Rights that apply to Your use of the Licensed Material and that the Licensor has authority to license.

h. **Licensor** means the individual(s) or entity(ies) granting rights under this Public License.

i. **Share** means to provide material to the public by any means or process that requires permission under the Licensed Rights, such as reproduction, public display, public performance, distribution, dissemination, communication, or importation, and to make material available to the public including in ways that members of the public may access the material from a place and at a time individually chosen by them.

j. **Sui Generis Database Rights** means rights other than copyright resulting from Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, as amended and/or succeeded, as well as other essentially equivalent rights anywhere in the world.

k. **You** means the individual or entity exercising the Licensed Rights under this Public License. **Your** has a corresponding meaning.

Section 2 – Scope.

a. License grant.

1. Subject to the terms and conditions of this Public License, the Licensor hereby grants You a worldwide, royalty-free, non-sublicensable, non-exclusive, irrevocable license to exercise the Licensed Rights in the Licensed Material to:

A. reproduce and Share the Licensed Material, in whole or in part; and

B. produce, reproduce, and Share Adapted Material.

2. Exceptions and Limitations. For the avoidance of doubt, where Exceptions and Limitations apply to Your use, this Public License does not apply, and You do not need to comply with its terms and conditions.

3. Term. The term of this Public License is specified in Section 6(a).

4. Media and formats; technical modifications allowed. The Licensor authorizes You to exercise the Licensed Rights in all media and formats whether now known or hereafter created, and to make technical modifications necessary to do so. The Licensor waives and/or agrees not to assert any right or authority to forbid You from making technical modifications necessary to exercise the Licensed Rights, including technical modifications necessary to circumvent Effective Technological Measures.

For purposes of this Public License, simply making modifications authorized by this Section 2(a)(4) never produces Adapted Material.

5. Downstream recipients.

A. Offer from the Licensor – Licensed Material. Every recipient of the Licensed Material automatically receives an offer from the Licensor to exercise the Licensed Rights under the terms and conditions of this Public License.

B. No downstream restrictions. You may not offer or impose any additional or different terms or conditions on, or apply any Effective Technological Measures to, the Licensed Material if doing so restricts exercise of the Licensed Rights by any recipient of the Licensed Material.

6. No endorsement. Nothing in this Public License constitutes or may be construed as permission to assert or imply that You are, or that Your use of the Licensed Material is, connected with, or sponsored, endorsed, or granted official status by, the Licensor or others designated to receive attribution as provided in Section 3(a)(1)(A)(i).

b. Other rights.

1. Moral rights, such as the right of integrity, are not licensed under this Public License, nor are publicity, privacy, and/or other similar personality rights; however, to the extent possible, the Licensor waives and/or agrees not to assert any such rights held by the Licensor to the limited extent necessary to allow You to exercise the Licensed Rights, but not otherwise.

2. Patent and trademark rights are not licensed under this Public License.

3. To the extent possible, the Licensor waives any right to collect royalties from You for the exercise of the Licensed Rights, whether directly or through a collecting society under any voluntary or waivable statutory or compulsory licensing scheme. In all other cases the Licensor expressly reserves any right to collect such royalties.

Section 3 – License Conditions.

Your exercise of the Licensed Rights is expressly made subject to the following conditions.

a. Attribution.

1. If You Share the Licensed Material (including in modified form), You must:

A. retain the following if it is supplied by the Licensor with the Licensed Material:

i. identification of the creator(s) of the Licensed Material and any others designated to receive attribution, in any reasonable manner requested by the Licensor (including by pseudonym if designated);

ii. a copyright notice;

iii. a notice that refers to this Public License;

iv. a notice that refers to the disclaimer of warranties;

v. a URI or hyperlink to the Licensed Material to the extent reasonably practicable;

B. indicate if You modified the Licensed Material and retain an indication of any previous modifications; and

C. indicate the Licensed Material is licensed under this Public License, and include the text of, or the URI or hyperlink to, this Public License.

2. You may satisfy the conditions in Section 3(a)(1) in any reasonable manner based on the medium, means, and context in which You Share the Licensed Material. For example, it may be reasonable to satisfy the conditions by providing a URI or hyperlink to a resource that includes the required information.

3. If requested by the Licensor, You must remove any of the information required by Section 3(a)(1)(A) to the extent reasonably practicable.

4. If You Share Adapted Material You produce, the Adapter's License You apply must not prevent recipients of the Adapted Material from complying with this Public License.

Section 4 – Sui Generis Database Rights.

Where the Licensed Rights include Sui Generis Database Rights that apply to Your use of the Licensed Material:

a. for the avoidance of doubt, Section 2(a)(1) grants You the right to extract, reuse, reproduce, and Share all or a substantial portion of the contents of the database;

b. if You include all or a substantial portion of the database contents in a database in which You have Sui Generis Database Rights, then the database in which You have Sui Generis Database Rights (but not its individual contents) is Adapted Material; and

c. You must comply with the conditions in Section 3(a) if You Share all or a substantial portion of the contents of the database.

For the avoidance of doubt, this Section 4 supplements and does not replace Your obligations under this Public License where the Licensed Rights include other Copyright and Similar Rights.

Section 5 – Disclaimer of Warranties and Limitation of Liability.

a. Unless otherwise separately undertaken by the Licensor, to the extent possible, the Licensor offers the Licensed Material as-is and as-available, and makes no representations or warranties of any kind concerning the Licensed Material, whether express, implied, statutory, or other. This includes, without limitation, warranties of title, merchantability, fitness for a particular purpose, non-infringement, absence of latent or other defects, accuracy, or the presence or absence of errors,

whether or not known or discoverable. Where disclaimers of warranties are not allowed in full or in part, this disclaimer may not apply to You.

b. To the extent possible, in no event will the Licensor be liable to You on any legal theory (including, without limitation, negligence) or otherwise for any direct, special, indirect, incidental, consequential, punitive, exemplary, or other losses, costs, expenses, or damages arising out of this Public License or use of the Licensed Material, even if the Licensor has been advised of the possibility of such losses, costs, expenses, or damages. Where a limitation of liability is not allowed in full or in part, this limitation may not apply to You.

c. The disclaimer of warranties and limitation of liability provided above shall be interpreted in a manner that, to the extent possible, most closely approximates an absolute disclaimer and waiver of all liability.

Section 6 – Term and Termination.

a. This Public License applies for the term of the Copyright and Similar Rights licensed here. However, if You fail to comply with this Public License, then Your rights under this Public License terminate automatically.

b. Where Your right to use the Licensed Material has terminated under Section 6(a), it reinstates:

1. automatically as of the date the violation is cured, provided it is cured within 30 days of Your discovery of the violation; or

2. upon express reinstatement by the Licensor.

For the avoidance of doubt, this Section 6(b) does not affect any right the Licensor may have to seek remedies for Your violations of this Public License.

c. For the avoidance of doubt, the Licensor may also offer the Licensed Material under separate terms or conditions or stop distributing the Licensed Material at any time; however, doing so will not terminate this Public License.

d. Sections 1, 5, 6, 7, and 8 survive termination of this Public License.

Section 7 – Other Terms and Conditions.

a. The Licensor shall not be bound by any additional or different terms or conditions communicated by You unless expressly agreed.

b. Any arrangements, understandings, or agreements regarding the Licensed Material not stated herein are separate from and independent of the terms and conditions of this Public License.

Section 8 – Interpretation.

a. For the avoidance of doubt, this Public License does not, and shall not be interpreted to, reduce, limit, restrict, or impose conditions on any use of the Licensed Material that could lawfully

be made without permission under this Public License.

b. To the extent possible, if any provision of this Public License is deemed unenforceable, it shall be automatically reformed to the minimum extent necessary to make it enforceable. If the provision cannot be reformed, it shall be severed from this Public License without affecting the enforceability of the remaining terms and conditions.

c. No term or condition of this Public License will be waived and no failure to comply consented to unless expressly agreed to by the Licensor.

d. Nothing in this Public License constitutes or may be interpreted as a limitation upon, or waiver of, any privileges and immunities that apply to the Licensor or You, including from the legal processes of any jurisdiction or authority.

Creative Commons is not a party to its public licenses. Notwithstanding, Creative Commons may elect to apply one of its public licenses to material it publishes and in those instances will be considered the "Licensor." Except for the limited purpose of indicating that material is shared under a Creative Commons public license or as otherwise permitted by the Creative Commons policies published at `creativecommons.org/policies`, Creative Commons does not authorize the use of the trademark "Creative Commons" or any other trademark or logo of Creative Commons without its prior written consent including, without limitation, in connection with any unauthorized modifications to any of its public licenses or any other arrangements, understandings, or agreements concerning use of licensed material. For the avoidance of doubt, this paragraph does not form part of the public licenses.

Creative Commons may be contacted at `creativecommons.org`.

Appendix E

References

Atalla, Martin, *US Patent 3,206,670*, "Semiconductor Devices Having Dielectric Coatings", application 8 March 1960, patent granted 14 September 1965.

Dacey, George and Ross, Ian *US Patent 2,778,956*, "Semiconductor Signal Translating Devices", application 31 October 1952, patent granted 22 January 1957.

"Department of Defense Test Method Standard Microcircuits", MIL-STD-883E, 31 December 1996.

Haseloff, Eilhard, "Latch-Up, ESD, and Other Phenomena", application report SLYA014A, Texas Instruments Incorporated, May 2000.

Horowitz, Paul and Hill, Winfield, *The Art of Electronics*, Second Edition, Cambridge University Press, 1989.

Kahng, Dawon, *US Patent 3,102,230*, "Electric Field Controlled Semiconductor Device", application 31 May 1960, patent granted 27 August 1963.

Lilienfeld, Julius Edgar, *US Patent 1,745,175*, "Method and Apparatus for Controlling Electric Currents", application 8 October 1926, patent granted 28 January 1930.

Oh, K.S., "MOSFET Basics", Fairchild Semiconductor, Application Note AN9010, July 2000.

Pierret, Robert F., *Semiconductor Device Fundamentals*, Addison-Wesley Publishing Company, 1996.

Timp, G., et. al., "The relentless march of the MOSFET gate oxide thickness to zero", Microelectronics Reliability, Volume 40, Issue 4-5, Elsevier Science Ltd., 2000.

"Understanding Power MOSFETs", Application Note AN-7500, Fairchild Semiconductor, October 1999.

Appendix F

Version history

This is a list showing all significant additions, corrections, and other edits made to this learning module. Each entry is referenced by calendar date in reverse chronological order (newest version first), which appears on the front cover of every learning module for easy reference. Any contributors to this open-source document are listed here as well.

29 October 2025 – corrected a typographical error ("is resembles") courtesy of Wyatt Gleason. Also added some instructor notes.

2 April 2025 – corrected a typographical error ("at at") courtesy of Jason Durrett. Also added another Technical References section, this one on Gauss' Law of Electric Fields.

30 March 2025 – Added another bullet-point to the "Challenging concepts related to field-effect transistors" section of the Introduction chapter.

1 November 2024 – added a new Animations chapter with an animation showing the development of an inversion layer in an E-type MOSFET with applied gate-source voltage.

11 October 2024 – divided the Introduction chapter into sections, one with recommendations for students, one with a listing of challenging concepts, and one with recommendations for instructors.

3 April 2024 – minor edit to one of the footnotes in the Full Tutorial chapter.

4 August 2023 – corrected typographical error in Simplified Tutorial, courtest of Nic Salvadalena.

6 April 2023 – added some Challenge questions.

28 November 2022 – placed questions at the top of the itemized list in the Introduction chapter prompting students to devise experiments related to the tutorial content.

17 November 2022 – added a Conceptual Reasoning question on the use of a MOSFET to limit the travel of a robotic servo motor.

3 November 2022 – added a footnote to the Full Tutorial discussing the unavoidability of parasitic

PN junctions within MOSFETs.

5 September 2022 – edited the Full Tutorial to elaborate more on the effects of V_{GS} on channel conductivity.

10 August 2022 – added a summary page to the end of the "Insulated-gate field-effect transistors" section of the Full Tutorial showing proper E-type MOSFET biasing.

6 April 2022 – added more questions to the Introduction chapter, and also added a requirement to the "Identifying MOSFET states" Conceptual Reasoning question for students to identify the orientation of the parasitic diode within each transistor.

10 December 2021 – edited footnote regarding the parasitic diode inside of an E-type MOSFET, referencing the transistor's arrow symbol as a way of telling which way this diode points.

11 November 2021 – added questions.

8 November 2021 – enhanced the Simplified Tutorial's text, and divided the Full Tutorial into sections.

21 July 2021 – added some Foundational concept application to the "ESD test jig" Quantitative Reasoning question.

10-11 May 2021 – added a Simplified Tutorial, and corrected a silly error where I described the oxide layer as being a metal oxide rather than (correctly) being silicon dioxide. Also, made minor additions to the Introduction chapter and added some instructor notes.

9 May 2021 – commented out or deleted empty chapters.

2 May 2021 – minor edits to the Tutorial to clarify the role of the silicon-dioxide layer.

29 April 2021 – minor edit to image_2225.

18 March 2021 – corrected multiple instances of "volts" that should have been capitalized "Volts".

5 November 2020 – added more instructor comments, and corrected typographical error in the new Case Tutorial section, regarding the mathematical sign of the $V_{GS(th)}$ values.

4 November 2020 – minor additions to the Introduction chapter. Added a Case Tutorial chapter with examples of FETs experiencing different bias conditions.

21 September 2020 – added footnote to Tutorial elaborating on the presence of AC and DC voltage sources in the patent illustration, as well as added text to the Introduction chapter. Also added some instructor notes to a question.

31 August 2020 – significantly edited the Introduction chapter to make it more suitable as a pre-study guide and to provide cues useful to instructors leading "inverted" teaching sessions.

16 April 2020 – changed title of Conceptual Reasoning question from "JFET source and load terminals" to "JFET source and drain terminals". Also clarified the Tutorial's discussion of the inversion layer within an E-type MOSFET.

18 February 2020 – deleted a bullet list in one of the Diagnostic Reasoning questions. Also fixed a typographical error ("reverse-based" instead of "reverse-biased") courtesy of reader Ty Weich.

23 January 2020 – added Foundational Concepts to the list in the Conceptual Reasoning section.

17 December 2019 – added quantitative question about electrostatic discharge (ESD) simulation, based on a model shown in MIL-STD-883E.

21 November 2019 – added content about electrostatic discharge (ESD).

10 November 2019 – minor edit to square-law formula annotation.

24 October 2019 – added square-law demonstration of FET to the tutorial.

20 October 2019 – added the patent numbers to all sections of the Historical reference chapter containing US patents. Also added footnote about gain, in relation to transconductance.

7 October 2019 – added explicit references (and index entries) for *threshold* voltage. Also added another row to the FET versus BJT table in the Technical References chapter.

2 October 2019 – continued writing Tutorial, adding Historical References, and added Experiments.

1 October 2019 – continued writing Tutorial and adding Historical References.

30 September 2019 – document first created.

Index

Active device, 19
Adding quantities to a qualitative problem, 124
Affine function, 34
Alloying, 22
Amplification, 18
Analog, 20
Annotating diagrams, 123
Asymmetric JFET, 27
Atalla, Martin, 45

Bias voltage, 23
Breadboard, solderless, 108, 109
Breadboard, traditional, 112

Capacitance, gate-drain in JFET, 27
Cardio-Pulmonary Resuscitation, 106
Checking for exceptions, 124
Checking your work, 124
Code, computer, 133
Conductance, 34
Conductor, 22
Conservation of Energy, 18
Coulomb, 51
CPR, 106

D-type, 29
Dalziel, Charles, 106
Depletion region, 22, 23
Depletion-mode, 29
Device under test, 98
Device, square-law, 35
Dimensional analysis, 123
DIN rail, 109
Diode, 23
DIP, 108
Discrete, 20
Doping, 22

DUT, 98

E-type, 29
Edwards, Tim, 134
Electric field, 36, 50
Electric shock, 106
Electrically common points, 53, 107
Electron tube, 21
Electrostatic discharge, 36
Electrostatic shielding, 53
Enclosure, electrical, 112
Enhancement-mode, 29
Equilibrium, 53
Equipotential, 53
Equipotential points, 107, 109
ESD, 36
Experiment, 113
Experimental guidelines, 114

Feedback, negative, 34
Field, electric, 36, 50
Field-effect transistor, junction, 25
Function, affine, 34
Function, linear, 34

Gain, 19, 20
Gauss' Law of Electric Fields, 50
Gauss's Law, 28
Gaussian surface, 50
Graph values to solve a problem, 124
Greenleaf, Cynthia, 69

Holes, 22
How to teach with these modules, 131
Hwang, Andrew D., 135

IC, 108
Identify given data, 123

Identify relevant principles, 123
Insulator, 22
Intermediate results, 123
Inversion, 29
Inversion layer, 60
Inverted instruction, 131

JFET, 14, 25
JFET, asymmetric, 27
Junction field-effect transistor, 25

Kahng, Dawon, 44
Knuth, Donald, 134

Lamport, Leslie, 134
Limiting cases, 124
Linear function, 32, 34

Majority, 26
Maxwell, James Clerk, 37
Metacognition, 74
Miller effect, 27
Minority carrier, 26
Moolenaar, Bram, 133
MOSFET, 15, 28, 53
Murphy, Lynn, 69

N-channel versus P-channel FET, 26
Negative feedback, 34
Newton, 51

Ohm's Law, 32
Ohmic behavior of FET, 32
Open-source, 133

P-channel versus N-channel FET, 26
Passive device, 19
Photoelectric effect, 22
Pinch-off voltage, 13, 33, 85
Potential distribution, 109
Problem-solving: annotate diagrams, 123
Problem-solving: check for exceptions, 124
Problem-solving: checking work, 124
Problem-solving: dimensional analysis, 123
Problem-solving: graph values, 124
Problem-solving: identify given data, 123
Problem-solving: identify relevant principles, 123
Problem-solving: interpret intermediate results, 123
Problem-solving: limiting cases, 124
Problem-solving: qualitative to quantitative, 124
Problem-solving: quantitative to qualitative, 124
Problem-solving: reductio ad absurdum, 124
Problem-solving: simplify the system, 123
Problem-solving: thought experiment, 114, 123
Problem-solving: track units of measurement, 123
Problem-solving: visually represent the system, 123
Problem-solving: work in reverse, 124
Project management guidelines, 117

Qualitatively approaching a quantitative problem, 124

Radar, 21
Radio, 21
Reading Apprenticeship, 69
Reductio ad absurdum, 124, 130, 131

Safety, electrical, 106
Schoenbach, Ruth, 69
Scientific method, 74, 113
Scope creep, 117
Semiconductor, 22
Shielding, electrostatic, 53
Shunt resistor, 32, 108
Simplifying a system, 123
Socrates, 130
Socratic dialogue, 131
Solderless breadboard, 108, 109
SPICE, 69, 114
SPICE netlist, 110
SPICE, computer analysis, 32
Square law, 35
Stallman, Richard, 133
Subpanel, 112
Surface mount, 109

Telegraph, 21
Telephone, 21
Television, 21
Terminal block, 107–110, 112

Thermionic emission, 22
Thermionic valve, 21
Thought experiment, 114, 123
Threshold voltage, 13, 33, 85
Torvalds, Linus, 133
Transconductance, 34
Tube, vacuum, 21
Turn-on voltage, 33

Units of measurement, 123

Vacuum tube, 21
Valve, thermionic, 21
Visualizing a system, 123
Voltage, pinch-off, 13, 33, 85
Voltage, threshold, 13, 33, 85
Voltage, turn-on, 33

Wiring sequence, 111
Work function, 22
Work in reverse to solve a problem, 124
WYSIWYG, 133, 134