1

INTRODUCTION

This chapter introduces industrial electronics. It covers some of the key developments and includes definitions for a few of the major terms. It also discusses the skills and knowledge that the industrial technician must possess to work efficiently. Finally, it introduces safety, which is of paramount importance for those preparing to work in the industrial environment.

1-1
PERSPECTIVE

The industrial world is where raw materials are processed, parts are manufactured, and final products are assembled. It provides us with a high standard of living in terms of the goods it offers and the employment it provides. It is central to our economy and to our general well-being. It is a changing world. Many believe that we are now experiencing the second phase of the industrial revolution or a totally new technological revolution. Industry is changing from a labor-intensive sector of our economy to a highly technical and automated environment. This change is impacting the industrial work force, and future industrial workers will work with their brains far more than with their brawn.

The merger of industry and electricity began early in this century, when motors began replacing other energy forms such as steam, water, animals, and humans. The motor was rapidly and widely applied to a variety of industrial tasks. As the applications increased, developing control systems that were more flexible and accurate than the mechanical systems based on clutches, gears, belts, and pulleys became necessary. The Ward-Leonard control system was the first to gain widespread application. In this system, the speed and direction of a direct current (dc) motor were controlled by energizing its armature circuit from a dc generator. By varying the polarity and intensity of the field current in the generator, it was possible to control the direction and speed of the motor efficiently by varying the polarity and magnitude of its armature voltage. The system offered smooth and stepless speed adjustment. When connected to a mechanical system with inertia, the motor was even able to return power to the alternating current (ac) supply upon deceleration. The efficiency and good control characteristics made the Ward-Leonard system very popular even though it is large, expensive, difficult to maintain, and not so efficient by modern standards.

Electronics began to enter the industrial arena during the 1940s. Various control devices were developed to supplement and, in some cases, replace the Ward-Leonard system. The devices of that era included thyratrons, mercury pool rectifiers, and ignitrons. Up until that time, dc motors were chosen for control applications because their torque and speed characteristics could be widely and smoothly controlled with good efficiency. Then, frequency control of ac motors opened up new possibilities. This was an attractive development since an ac motor costs only one-fourth as much as a comparable dc motor. During a later period, solid-state replacements for thyratrons, called *thyristors*, further extended the applications of ac motors. In the 1950s, magnetic and electronic amplifiers (using vacuum tubes) closed the control loop for even more sophisticated applications. These systems used feedback to control positioning automatically and accurately and were called *servomechanisms*. During the late 1950s and early 1960s, solid-state devices made their debut in industry with devices such as the transistor and the thyristor. Robots first appeared in the 1960s and were used in applications such as welding and spray painting and in environments too dangerous for human workers. The operational amplifier made its inroads in the late 1960s, and systems based on digital logic also gained in popularity. The *programmable controller*, also introduced in this period, began to eliminate electromagnetic relay logic in timing and sequencing control. During the 1970s the miracle device, the microprocessor, opened up a new world of sophisticated control for both dc and ac rotating equipment. A *microprocessor* is most of the circuitry of a digital computer reduced to a tiny chip of silicon. Microprocessors quickly boosted the sophistication of automated systems as managers and designers reacted to pressure for increased productivity and quality.

In the modern industrial plant there are a variety of dc motor control systems ranging from fractional horsepower ratings to ratings over 10,000 horsepower (hp). Applications include portable power tools, conveyors, robots, palletizers, hoists, cranes, elevators, rolling mills, tube mills, extrusion mills, and many others. Thanks to high-power thyristors, ac drive systems are available with ratings ranging from fractional horsepower to 20.000 hp and are replacing dc motors in many industrial applications. In addition to motors and motor drives, modern industry uses electronics in many processes and machines. Examples include electronic welding, electroplating, process control, and computer-controlled machines such as lathes, milling machines, drill presses, and robots.

Key Terms in Industrial Electronics

The following list of definitions is brief, but it will help in developing a perspective of modern industrial electronics:

Automation: the implementation of processes by automatic means without the need for human intervention; self-moving, self-adjusting, and self-controlling.

Process control: the control imposed on physical or chemical changes in a material.

Automatic controller: a device that operates automatically to regulate a variable in response to a command and to a feedback signal.

Programmable controller: a device that is configured by software commands (instructions) to sequence and time events.

Robot: a programmable, multifunction manipulator designed to move material, parts, tools, or specialized devices through variable programmed motions for the performance of a variety of tasks (source: Robot Institute of America).

Microprocessor: the entire central processing unit of a computer reduced to a tiny chip of silicon.

Numerical control: a programming system of letters and numbers that regulates an entire machining process and the sequence of operations to machine a work piece.

Flexible manufacturing system: a computerized production line with a high degree of programmability to allow quick and easy changes from one small production batch to another.

CAD/CAM: computer-assisted design/computer-assisted manufacturing.

Computer-integrated manufacturing: an integration of design, engineering, manufacturing, and business functions on a central computer or on a network of computers.

REVIEW QUESTIONS

1. What is the name of the early motor control system that used a dc generator to energize the armature circuit of a dc motor?

2. Thyristors are solid-state replacements for ________.

3. How much does an ac motor cost as compared to a dc motor with similar ratings?

4. A servomechanism uses feedback to automatically control mechanical ________.

1-2
KNOWLEDGE AND SKILLS REQUIRED

The industrial workplace is often a combination of old, newer, and state-of-the-art technology. The worker may be confronted with brand-new, late-technology equipment sitting beside equipment that is 50 years old. This environment demands a thorough understanding of basics. It requires a diligent technician who takes the time and makes the effort to understand all of the equipment in his or her care. Learning cannot be confined to classrooms. The best technicians are always actively involved in learning more about their technology and about their workplace.

Importance of Theory

Basic circuit theory is one of the most valuable tools for the industrial technician. Ohm's law, Kirchhoff's laws, and other basic principles make the logical process of fault isolation possible. This knowledge must be coupled with the ability to use instruments. Meters, oscilloscopes, signal generators, logic probes, logic pulsers, and other test equipment must be used safely, accurately, and effectively. The technician must also understand the theory of operation for the equipment being diagnosed. Some equipment is very complicated, and the overall picture must be clear. If the equipment interacts with other equipment, then the entire system must be understood. A familiarity with electronic circuits and devices is also necessary. An *electronic circuit* is one that contains active components such as transistors and integrated circuits. When troubleshooting at the component level, one must understand the way the active devices function. Then, when they malfunction, it is possible to diagnose their behavior as abnormal and to make the proper replacements. Troubleshooting at the board level is usually less demanding. In *board-level troubleshooting* the fault is traced to one plug-in circuit board or module, and then the entire circuit is replaced. This is fine in those cases in which a spare board is in stock. If a board is not in stock, then loss of production time can cost the company a considerable sum. It is also often wasteful to discard a defective board. Defective boards may be sent back to the manufacturer for repair, or the technician may be expected to repair them.

Importance of Mechanical Skills

In addition to electrical and electronic knowledge, the industrial technician must have considerable mechanical skill. Much of the work will involve mechanisms that interact with circuits. Sometimes it is difficult to determine whether a problem is mechanical, electrical, or both. Mechanical skills are also necessary for the proper disassembly and reassembly of equipment. All parts and fasteners must be reinstalled just as they were. Attitude has a lot to do with this aspect of the job. Sorting parts and fasteners during the disassembly process takes a little more care, but this care usually pays off handsomely during reassembly. The careful technician will also make sketches when tearing down a complex unit, especially the first time.

Interpersonal Relationships

Human skills are sometimes overlooked but are very important. A breakdown in industry can cause a lot of tension. The down time is costly, and there may be considerable pressure on the technician from several sources to get things fixed quickly. Tempers can flare and control is absolutely necessary. This is not a pleasant atmosphere to work in, but it sometimes occurs. Experienced and skillful technicians have the ability to remain calm and communicative under these conditions and are often highly valued for this reason. Another aspect of interpersonal relations is that the technician must communicate with other people who may have more operating experience with the equipment. They may have been there when it broke down or may have noticed some peculiar behavior before it broke down. Obviously, this information is invaluable. The skilled technician who calmly asks a series of logical questions can stabilize a tense situation and get on with the important task at hand.

Another important human skill for a troubleshooter is to listen attentively to everything others have to say but to acknowledge privately the possibility that they may be wrong about some or all of the information they give. First of all, a coworker who did something wrong and damaged the equipment may not be likely to tell the whole truth about it. The person reporting might be confused or biased or might fabricate some of the events for self-protection. This is not to say that people cannot be trusted, but that they do make mistakes and very few like admitting them. The best technicians learn to verify facts. They never assume anything. Once something is verified and noted (mentally or on paper), it is time to move on to the next step. This kind of logical process saves time in the long run. A disorganized technician will wind up going in circles because of the failure to verify and note conditions.

Software problems are another area in which a debate can ensue and tempers can flare. Much modern industrial equipment is programmable. Programs are called *software* and are written by programmers who are sometimes a little too quick to blame the hardware when things don't work as planned. To be fair, there are also hardware technicians who are too quick to blame the software. Software problems can be difficult to solve. A new program that exercises the equipment in ways never tried before often creates a challenging situation, which requires a cooperative effort by both hardware and software people to solve the problem. This is why it is so important to deal skillfully with people; maintaining good working relations pays off handsomely. On the other hand, a rude technician may alienate everyone in the plant and be rendered ineffective because of a lack of communication. Think about it.

Preventive Maintenance

In addition to corrective maintenance, preventive maintenance is an important aspect of industrial electronics. This is often totally ignored, but it shouldn't be. Operation, instruction, and service manuals for the equipment in the plant are invaluable. Periodic lubrication, inspection, cleaning of air filters, running of diagnostic programs, checking for ground faults, and all the other procedures recommended in the equipment manufacturer's literature are absolutely essential for a successful maintenance program.

Organizational skills are invaluable in instituting an effective program. Maintenance logs for every piece of equipment will serve to monitor the program. Logs should contain the date and time and a brief but complete description of every procedure performed and every part replaced. Some technicians start a maintenance log by "fingerprinting" a piece of equipment. This procedure involves a complete operational profile in which all deviations from the original factory specifications are noted. Every effort is then made to restore the equipment to original parameters. Minor problems, such as burned-out indicators and faulty panel switches, are normally corrected immediately. More involved problems should be scheduled for attention as soon as possible, and any parts that are not on hand should be ordered. A maintenance log should also contain an inventory of all spare parts in stock for the equipment. A general rule of thumb is to stock an inventory worth 10 percent of the purchase price of the equipment. The technician's experience and the manufacturer's recommendations will dictate what should be stocked. Some common items are circuit boards, modules, fuses, switches, circuit breakers, panel lamps, resolvers, tachometers, motors, and solid-state parts such as diodes, thyristors, transistors, and integrated circuits.

An operational profile is really a form of preventive maintenance. It is often combined with several procedures that ensure proper operation. The equipment should be given full function and range testing. It may also involve tune-ups of servomechanisms and adjustments of feed rates and spindle speeds. Feedback devices and limit switches are also checked for proper operation. Cleaning may be done at the same time, along with replacement of hydraulic and air filters.

REVIEW QUESTIONS

5. Electronic troubleshooting may take place at the board level or at the ______ level.

6. Which level of troubleshooting demands more knowledge?

7. A good technician has electronic knowledge, mechanical skill, and ______ skills.

8. Programmable equipment may malfunction because of ______ problems in addition to electronic and mechanical faults.

1-3
SAFETY

Safety in any environment involves two major components: knowledge and attitude. Knowledge is the body of information that defines dangerous procedures, conditions, materials, and all other potential hazards. It also includes proper procedures: how and when to use protective equipment, and how to react in an emergency. Technicians should know cardiopulmonary resuscitation (CPR) and first aid treatment for burns, acid contact, and other emergency situations. They must also know the location of all safety devices, emergency showers, fire extinguishers, alarm systems, planned evacuation routes, and shelter areas. Knowledge is absolutely essential but does not guarantee safety. Most industrial workers who are injured or killed have been trained. Some people have a tendency to bypass safe procedures for one reason or another. It is very important to gain knowledge and then make it a way of life. Technicians who have the knowledge and the proper attitudes are seldom injured.

The industrial environment may be replete with many kinds of hazards. It is not possible to deal with all safety areas adequately here. Radiation, loud sounds, rotating machines, pinch points, dangerous chemicals and gases, explosive atmospheres, and lasers are some examples of potential hazards for which specific safety knowledge is acquired on the job. This section will be limited to general rules of electrical and mechanical safety.

Electrical shock can cause falls and other physical reaction injuries, permanent damage to the human body, and even death. Improper procedures may also damage expensive equipment. Moving equipment, such as a robotic arm, can cause serious injury. The following rules will help establish safe procedures and alert the technician to dangerous situations:

1. Circuits that are being serviced must be locked open and tagged. A "buddy" system may prevent any chance of anyone's inadvertently energizing a system before the appropriate time.
2. Dirt and moisture conditions must be noted. They can drastically alter conductivity and may provide unusual current paths.
3. All connections must be checked carefully. Cables that move and cables that are subject to external abuse are especially susceptible to insulation damage and connector damage.
4. Contacts that normally carry heavy current must be inspected periodically. They must be tight and clean. High current can lead to heat, sparks, and flames when it flows through a high-resistance joint or connection.
5. Care must be taken around batteries because some types produce an explosive gas. Open flames or sparks in their vicinity can result in an explosion. Battery spills can be dangerous since the electrolyte is extremely caustic.
6. When working with live line voltage both hands should not touch the equipment simultaneously. Leaning on a cabinet with one hand while working with the other is dangerous. Hand-to-hand shocks can be lethal. High-voltage mats and any other protective equipment that is recommended should be used when working on live circuits. Rings, bracelets, and similar metal items should not be worn around live circuits. Wet clothing and shoes are good conductors of electricity and must not be worn around live circuits.
7. Test equipment must be kept in good working order. Worn leads should be replaced, connectors should be checked regularly, and only exact replacements for broken knobs and probes should be used. Verify that the power cords are in good condition and that grounds are intact. Technicians must always keep their hand tools and soldering equipment clean and in good working order.
8. Manufacturer's service literature should be kept readily available. Follow the recommended procedures for each piece of equipment that is worked on.
9. All work must conform to the current National Electrical Code and any applicable local regulations.
10. Care must be taken when equipment is installed and energized for the first time. A phase error may cause motors to run backward with unpredictable results. Most equipment contains a phase-sequence lockout circuit which will prevent the contactor from closing in the event of a wiring error. Never manually override such devices in such a situation.
11. Never defeat interlocks, limit switches, overcurrent devices, and other protective features. Verify that they all operate properly and have not been tampered with. Do not overrate replacement protective devices.
12. Power devices, such as thyristors, that are subject to large surge currents may physically fail by package rupture and an expulsion of materi-

als. Eye protection should be worn when working around such devices. Electrolytic and tantalum capacitors can also explode when subject to an overvoltage or a reverse bias. A tantalum capacitor may release a liquid electrolyte. Skin, eye, or mouth exposure to this material must be treated promptly. Flush with large amounts of running water and seek medical attention immediately in the case of eye contact or ingestion.

13. The side effects of certain medication may impair judgment or equilibrium or induce drowsiness. Extreme caution must be practiced under such conditions.
14. Equipment grounds must not be removed or circumvented by the use of adaptors. The power cord should be connected before using test probes.
15. Protective clothing and eye protection should be worn when working around high-vacuum devices such as cathode ray tubes.
16. Learn all appropriate hazard symbols and warning signs.
17. Only approved fire extinguishers should be used on electrical equipment. Water is a conductor and should not be used on electrical fires. Carbon dioxide, foam, and halogen types are preferred.
18. Capacitors and hydraulic accumulators store energy after the power has been turned off. A high-voltage capacitor may present a lethal shock hazard even when disconnected from its source.
19. Be wary of moving equipment and pinch points. Loose, floppy clothing invites getting caught in machinery. Do not enter "hard hat areas" without the prescribed protective gear.
20. Use extreme care when working around robots since they can make unpredictable movements. A robot that appears to be "dead" may be awaiting input signals.
21. There are two operating zones, referred to as envelopes, in robotics: the program envelope and the absolute maximum reach envelope, including the reach of any tooling. Learn the maximum envelope and stay clear of it.
22. If servicing is required within a robot's envelope it may be necessary to use blocking posts that are set into the floor to prevent the arm from striking anyone within the envelope.
23. Verify that the robot works properly in the "teach mode." Most control systems allow only slow movements when an operator or technician is in the envelope. Make certain that the controls are working properly.
24. Many robots use an interlocking control system that prevents an automatic cycle from beginning before homing to the start position. This is a safety feature because the home position is usually the fully retracted position. This system must be checked for proper operation.
25. Robot controls may also be interlocked to other pieces of equipment such as conveyors. The entire system must be checked for proper operation to ensure that the entire cycle begins and ends properly, for example, when a part clears a die, a mold, or a fixture.
26. Eye protection is absolutely essential around robots. A welding robot may initiate an arc without any warning. Care must be exercised around lasers. In general, approved safety goggles should be worn in all potentially dangerous situations.
27. The use of vacuum is preferred to the use of compressed air for cleaning electronic equipment. Many of the air filters used in air cooling systems contain fiberglass, which may cause serious eye injury if the glass fibers are blown into the eye.

Be extremely careful with solvents. Many of them are highly volatile. Breathing the fumes can be extremely damaging to the human body. Use of carbon tetrachloride is strictly regulated by law because it is easily absorbed into the skin and causes organ damage. Some solvents should be used only in an area fitted with a flameproof exhaust hood. A respirator and other protective clothing may also be required. Solvents are also known to damage electrolytic capacitors. Electrolytics are especially susceptible to halogenated hydrocarbons, which enter the end seals and cause internal corrosion. Other solvents and degreasers that should be avoided around electrolytics include Freon, chlorine, and fluorine compounds. Spent solvents must be disposed of properly. Because it is usually unacceptable to dump them into a sewer system, they are often stored in holding tanks for eventual removal and processing.

Beware of polyvinyl chloride (PVC) insulation. It is commonly found on wires and is used in heat-shrink tubing. Gaseous hydrogen chloride is released from PVC at a temperature of 230°C. The hydrogen chloride combines with moisture in the air, and a hydrochloric acid mist results. It acts as a primary irritant and corrosive agent to the respiratory system. It is also absorbed on contact with skin. Do not use an open flame to activate heat-shrink tubing. Use a heat gun or other approved technique.

Safety in Tests and Measurements

Safe procedures may not always be obvious. One of the most dangerous and least understood procedures is in the area of floating measurements. A *floating measurement* is taken with the test equipment chassis, for example, of an oscilloscope, at some potential higher than ground. Technicians must sometimes make measurements across two points in a circuit where neither is at ground potential. The common connection for the signal may be hundreds of volts with respect to the line ground. Another related problem is that the green grounding wire on two separate line circuits may be several volts different. Although this condition, known as a *ground loop*, is symptomatic of a poor grounding system, it does cause fault current and interferes with proper test measurements.

Oscilloscopes and other test equipment have a signal common terminal that is connected to the protective ground, which is usually the chassis of the instrument. This protective ground is connected via a green wire to the power line ground when the instrument is plugged into an outlet. This is a safety feature and prevents the chassis, cabinet, connectors, probe ground, and the controls from assuming a dangerous potential with respect to ground. Without this feature, it would be possible to receive a dangerous shock just by touching some part of the instrument. Unfortunately, it also requires that all signals be measured with respect to ground, making it difficult to measure a signal across two points in a circuit when neither point is grounded. Connecting the test equipment in a case like this will ground one of the circuit points and may cause fault currents which could damage the equipment. Even if it does not cause damage, the connection will often inject noise into the circuit, causing abnormal operation and making any meaningful measurement impossible. Too many technicians improperly "float" the test equipment in these cases, creating a potentially dangerous situation. Improper floating defeats the protective ground system by using a three-to-two wire adaptor or by cutting off the ground prong on the power connector. Using an isolation transformer on the oscilloscope is also unacceptable since it also defeats the protective grounding system.

Floating an oscilloscope is dangerous because it allows the cabinet, connectors, and test leads to assume the same potential as that of the probe ground connection. One practice is to rope off the area when floating measurements are being made, to protect personnel from touching the oscilloscope and receiving a dangerous shock. This is a questionable practice, and it does not take into account that the power transformer in the oscilloscope may be unduly stressed, possibly leading to insulation failure. Another problem with this practice is that the capacitive loading effect of the floating instrument exerts a filter action on the circuit point where the probe ground is connected, and this filter action may interfere with proper operation and measurement. For these reasons, floating measurements that defeat the protective ground system are unacceptable even though commonly practiced.

One acceptable method of performing floating measurements is to use an isolation amplifier between the oscilloscope and the circuit under test. The amplifier connects between the oscilloscope and the equipment under test and passes measurement signals across an insulating barrier. Another method that may work in some applications is indirect grounding by using equipment such as a ground isolation monitor. The monitor is connected between the power lines and the test equipment. If the potential of the isolated ground exceeds 40 volts (V) peak, it disconnects power to the test equipment, sounds an alarm, and restores the protective ground circuit.

Another safe technique is to use an oscilloscope with a differential input. Matched probes are used with the negative preamp input connected to the circuit common and the positive preamp input connected to the signal test point. The common mode dynamic range and the common mode rejection ratio are important specifications of such an amplifier since the measured signal may be very small in comparison to the common mode signal. Some technicians forget this point and attempt to use a quasi-differential technique with a dual-trace oscilloscope in the ADD mode. By inverting channel 2, the two signals are electrically subtracted, effectively removing the common mode signal. Unfortunately, the dynamic range of dual-trace amplifiers is too small, and the common mode rejection ratio is only about 100 to 1. The results obtained are often poor.

Another acceptable measurement technique is to use an all-insulated oscilloscope. Battery-operated equipment avoids ground loops and is preferred by many technicians. Finally, some advanced oscilloscopes are capable of storing waveforms. If one of these is available, two separate measurements can be taken and stored; then the common mode signal is mathematically subtracted out. Such oscilloscopes are very expensive.

REVIEW QUESTIONS

9. Safe workers have the required knowledge and the proper ______ regarding their environment and practices.

10. Circuits that are being serviced should be locked open and ______.

11. Dirt and moisture can decrease electrical ______.

12. You should not work on circuits or machines while wearing ______.

CHAPTER REVIEW QUESTIONS

1-1. What is the name given to a chip of silicon containing most of the circuitry of a digital computer?

1-2. Name a device that is configured by software to sequence and time events.

1-3. A programmable, multifunction manipulator designed to move material, parts, or tools through variable programmed motions for the performance of a variety of tasks is commonly called a ______.

1-4. Name the two broad categories of maintenance.

1-5. Which category of maintenance is likely to be ignored?

1-6. An industrial technician should maintain a diary, called a ______, for each piece of equipment.

1-7. According to the rule of thumb, if a piece of equipment cost $50,000 then ______ worth of spare parts should be stocked for that piece of equipment.

1-8. A phase error may cause a piece of equipment to run ______.

1-9. A robot that appears to be "dead" may be awaiting a control ______.

1-10. Heat-shrink tubing should never be activated with a ______.

1-11. What prevents the chassis and connectors of test equipment from assuming a dangerous potential?

1-12. Isolation amplifiers, ground isolation monitors, differential amplifiers, and battery-operated equipment are examples of equipment that can be used to make ______ measurements safely.

ANSWERS TO REVIEW QUESTIONS

1. Ward-Leonard **2.** thyratrons **3.** one-fourth **4.** position **5.** component **6.** component **7.** human **8.** software **9.** attitude **10.** tagged **11.** resistance **12.** metal items and wet clothing

2

SOLID-STATE DEVICES

This chapter investigates solid-state materials and devices. A majority of industrial systems are based on semiconductor technology rather than the older vacuum tube technology. Semiconductors are solid-state crystals that can control and amplify. Every industrial employee with responsibilities for electrical and electronic systems must have a good working knowledge of solid-state devices, including their theory of operation and their physical, electrical, and thermal characteristics.

2-1
SEMICONDUCTOR THEORY

Semiconductors are materials with electrical properties that fit in between the properties of conductors and insulators. Common semiconductor materials include the elements carbon, silicon, germanium, and the compound gallium arsenide. These materials oppose the flow of electricity and can be used to make resistors and control devices.

It will be helpful to review conduction as a prelude to examining semiconduction. Copper is an excellent conductor. It exhibits high conductivity because many, many current carriers are available to support the flow of current. One cubic centimeter (cm^3) of copper has 10^{23} current carriers. Figure 2-1 shows the structure of atomic copper. Note that the outermost electron is called a *valence electron*. The valence electrons are the only ones that can move through the material and support the flow of current. A length of copper wire is held together by metallic bonding. In this situation positive copper ions float in a cloud of valence electrons. Any given electron in the cloud is not associated with any particular copper atom. Electrostatic forces between the positive ions and the negative electrons hold the wire together. Placing a potential difference across the ends of the wire causes a significant current since so many electrons are available to serve as carriers. This type of flow is referred to as a *drift current*. Therefore, conductors such as copper behave as they do because they have large numbers of current carriers.

Semiconductor materials, such as silicon, are held together by *covalent bonding*. In this process atoms share valence electrons with neighboring atoms to gain a total valence count of eight. Figure 2-2 shows the structure. Eight electrons fill the valence orbit and leads to stability. The resulting silicon structure is very stable and is called a *crystal*. Since the valence electrons of the crystal are locked into the covalent bonds, pure silicon conducts very poorly at room temperature. It is not possible to make absolutely pure silicon, but it can be approached with only 1 impurity atom for every 10^{13} silicon atoms. Because of the lack of current carriers, it does not even semiconduct. It is considered an intrinsic semiconductor. The resistivity of intrinsic (pure) silicon at room temperature is 200,000 ohms per cubic cen-

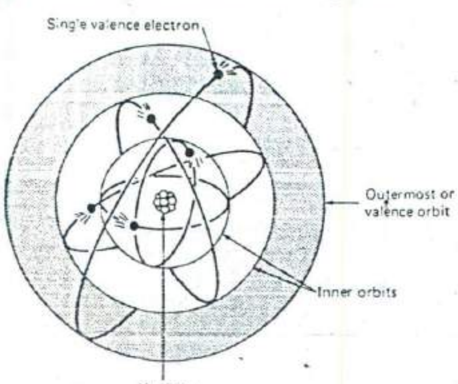


Fig. 2-1 Simplified copper atom.

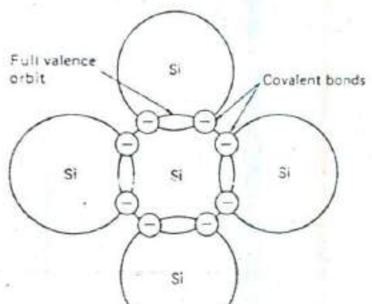


Fig. 2-2 Silicon atoms sharing valence electrons.

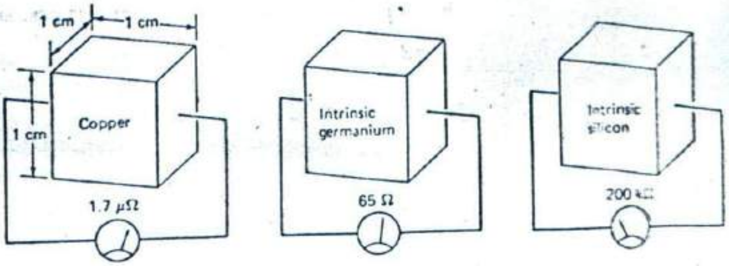


Fig. 2-3 Comparing resistivities of copper, germanium, and silicon.

timeter (Ω/cm^3). Compare this to the resistivity of copper, which is only $1.7\ \mu\Omega/\text{cm}^3$. It should be obvious how few current carriers are present in an intrinsic semiconductor as compared to a conductor. Intrinsic germanium, another semiconductor material, exhibits a resistivity of $65\ \Omega/\text{cm}^3$ at room temperature. Figure 2-3 compares the resistivities of copper, intrinsic germanium, and silicon. It turns out that the higher resistivity of intrinsic silicon is an advantage when building most semiconductor devices and is one of the reasons that silicon is far more popular than germanium.

Temperature can affect the resistivity of semiconductors. When heat enters the crystal, some of the valence electrons will gain enough heat energy to break their covalent bonds. It is interesting to note that electromagnetic radiation (such as light) entering the crystal will also release carriers. This effect is utilized in certain devices and will be covered in the chapter on optoelectronics. The heat-liberated electrons will support the flow of current. Carriers produced in this manner are known as *thermal carriers*. As Fig. 2-4 shows, heat can release an electron from its covalent bond and leave a positive hole behind. The hole also acts as a carrier to support the flow of current. Therefore, we may expect the resistivity of semiconductors to decrease as temperature increases. This relationship is known as a *negative temperature coefficient* and is an important characteristic of solid-state devices. Compare this with conductors which have a positive temperature coefficient. Their resistivity increases with an increase in temperature. This occurs because heat entering a conductor increases the activity of the electron cloud. The increased activity enhances the probability of electron collisions for the drift current. The collisions impede the current, and resistivity increases. This is why conductors become superconductors (no resistance) at absolute zero (−273°C). Conversely, an intrinsic semiconductor insulates at absolute zero since no thermal carriers are present in the crystal.

Doping is the process used to make crystals semiconduct at room temperature. It is possible to diffuse into the crystal impurity atoms that have five valence electrons. These materials, such as arsenic, are called *donor impurities* because they donate extra electrons to the crystal. The donated electrons are not covalently bound and serve as current carriers. Figure 2-5 shows a silicon crystal that is doped with a donor impurity. The doping level is typically very small with something on the order of 1 impurity atom for every 10 million silicon atoms. The doped material is known as *extrinsic silicon* and shows a lower resistivity than intrinsic silicon. Materials doped with donor impurities are known as *N-type semiconductors* since the current carriers are electrons and have a negative charge.

It is also possible to dope the crystal with acceptor impurities. These materials, such as boron, have only three valence electrons. When they enter the crystal, one covalent bond is not satisfied and a hole results. Think of the hole as a position for an electron. These

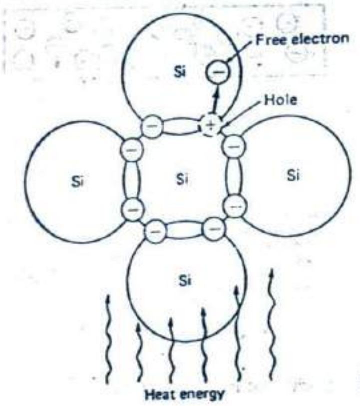


Fig. 2-4 Thermal carriers.

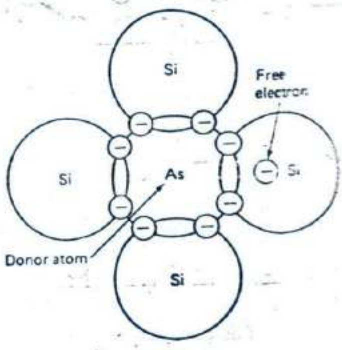


Fig. 2-5 N-type silicon.

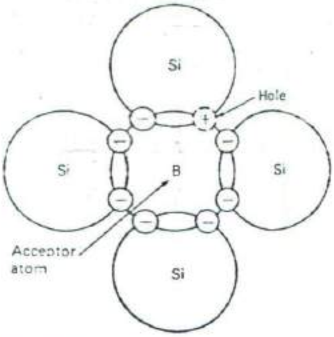


Fig. 2-6 P-type silicon.

positions will accept electrons; for this reason impurities that create holes in the crystal are called *acceptor impurities*. A semiconductor that is doped with acceptor atoms is also extrinsic but is called *P-type*. The structure of P-type silicon is shown in Fig. 2-6. The holes behave as positive charges and support the flow of current in a similar fashion to electrons. However, there are two important differences between electron current and hole current; one of the differences is indicated in Fig. 2-7. Note that at the top of the illustration a hole appears as the electron leaves the crystal at the right. Then that hole is filled, and another appears to its left as the electrons move to the right. It is apparent from Fig. 2-7 that the holes seem to be moving opposite to the direction of the electrons. The second difference is that hole current is based on position swapping, and the apparent motion of the holes is slower than simple electron motion in an N-type crystal. This provides N-type semiconductor material with a comparative speed advantage which is important in some applications.

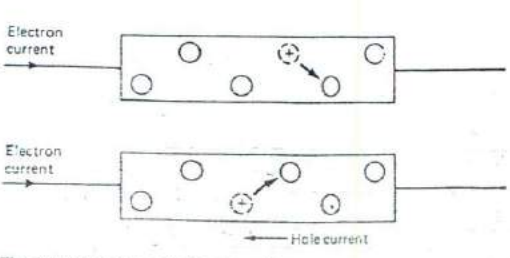


Fig. 2-7 Direction of hole current.

We have discussed two ways to provide crystalline silicon with current carriers. Heat can produce thermal carriers, and impurities can provide electrons or holes to serve as carriers. At room temperature, the extrinsic carriers outnumber the thermal carriers by about 1 million to 1. Thus, the extrinsic conduction is all that concerns us and the thermal carriers are insignificant. However, somewhere in the range of 200 to 400°C, the conduction becomes intrinsic. In *intrinsic conduction* device characteristics are controlled by temperature, and the device becomes useless for most electronic applications. Intrinsic conduction must be avoided, and therefore most semiconductor devices must not exceed a temperature of 200°C. Extreme cold must also be avoided since the crystals can crack. The range of safe operating temperatures varies according to the type of device (diode, transistor, integrated circuit, etc.) and according to its classification. For example, commercial integrated circuits are rated from 0 to 75°C, industrial integrated circuits from −40 to 85°C, and military integrated circuits from −55 to 125°C. Germanium devices exhibit intrinsic conduction at a much lower temperature than silicon devices and must operate over an even more restricted temperature range. This is another reason why silicon is far more useful for most applications. Storage temperatures and soldering temperatures are also specified by device manufacturers and must be observed.

As Fig. 2-8 shows, extrinsic crystals contain majority and minority carriers. The N-type crystal at the top shows a majority of electrons. A few holes can be found and are considered minority carriers. At the bottom, the P-type crystal has a majority of holes. A few electrons act as minority carriers. Thermal carriers come in pairs. Remember, every electron that gains its freedom leaves a hole behind. This means that the number of minority carriers in any extrinsic crystal is directly related to the temperature of the crystal. For example, heating an N-type crystal will cause a few electrons to be freed. These join the other majority carriers. The holes that are created

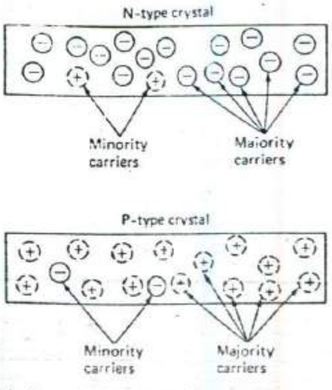


Fig. 2-8 Majority and minority carriers.

at the same time become minority carriers. Heating a P-type crystal also releases a few electrons, which become minority carriers, and the holes join the other majority carriers. It will be seen in subsequent sections of this chapter that minority carriers act to degrade the operation of devices such as diodes and transistors.

REVIEW QUESTIONS

1. *Intrinsic silicon* is another name for ______ silicon.

2. How does the resistivity of intrinsic germanium compare to the resistivity of intrinsic silicon?

3. Semiconductors have a ______ temperature coefficient.

4. Conductors have a ______ temperature coefficient.

5. Extrinsic crystals can be either N-type or ______.

6. Holes, in an N-type crystal, are considered ______ carriers.

2-2

JUNCTION DIODES

Junction diodes are two-terminal devices that utilize both P-type and N-type conduction in a single crystal. They are usually manufactured by a planar diffusion process. By using heat, pressure, and a dopant in gaseous form, it is possible to force atoms into a crystal to create carrier sites. The structure of the typical diffused junction diode is illustrated in Fig. 2-9. The *junction* is the boundary region between the N-type material on the left and the P-type material on the right. It is not a mechanical joint or a seam since a continuous and single crystalline structure is necessary for diode action. The figure also shows that a depletion region forms along both sides of the junction. This region is the result of free electrons from the N material on the left crossing the junction and filling holes in the P material on the right. Notice that there are no carriers in this depletion region. The depletion region behavior is therefore intrinsic. We may expect it to exhibit very high resistance at room temperature. Therefore, we may also expect high lead-to-lead resistance in the diode when the depletion region is present.

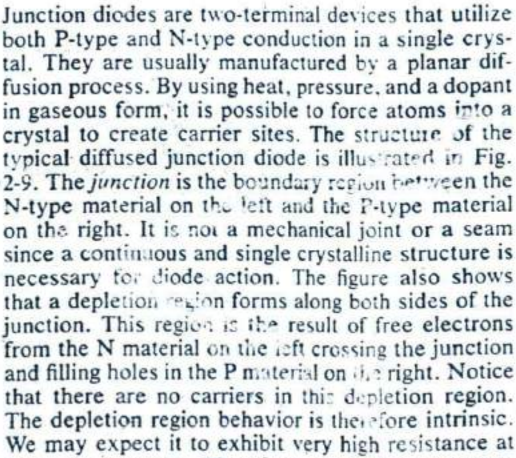


Fig. 2-9 Junction diode.

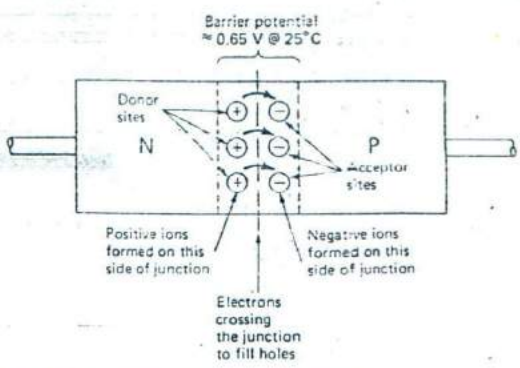


Fig. 2-10 How the barrier potential is formed.

Figure 2-10 shows why all of the electrons from the N material do not cross the junction and fill all of the holes in the P material. As each electron leaves the N side it also leaves its parent donor atom. This donor atom becomes a positive ion since it has lost an electron. Likewise, a negative ion is formed on the P side, where each electron joins the valence orbit of an acceptor atom. These ionized atoms produce a force called the *ionization potential* that prevents any further carrier movement. Note that the negative ions shown in Fig. 2-10 will tend to repel additional electrons that might try to cross the junction to fill holes. The current that flows until the forces go into equilibrium is called a *diffusion current*. This is different from the drift current discussed in the previous section. The ionization potential that develops as a result of the diffusion current is also known as the *barrier potential*. At room temperature, this potential is approximately 0.65 V for silicon junctions and 0.25 V for germanium junctions.

It is possible, by using an external voltage source, to overcome the barrier potential and collapse the depletion region. This is known as *forward bias* and is shown in Fig. 2-11. Note in Fig. 2-11(*b*) that the negative terminal of an external battery has been connected to the N material. Like charges repel, and the electrons are driven toward the junction, thus collapsing the depletion region. Assuming a silicon diode, the external battery must provide more than 0.65 V to produce significant current through the diode since the barrier potential must be overcome. With sufficient forward bias applied, the diode current can be large and some load- or current-limiting device will be required to protect the circuit. The schematic symbol for the diode is given in Fig. 2-11(*a*). The N side is called the *cathode* since it delivers electrons to the P side. The P side is called the *anode*. The direction of conventional current flow is opposite to the electron current; the diode symbol is based on this flow, since the arrow indicates conventional direction when the diode is forward-biased. As a memory aid, note that the anode side of the symbol is shaped like the letter *A*.

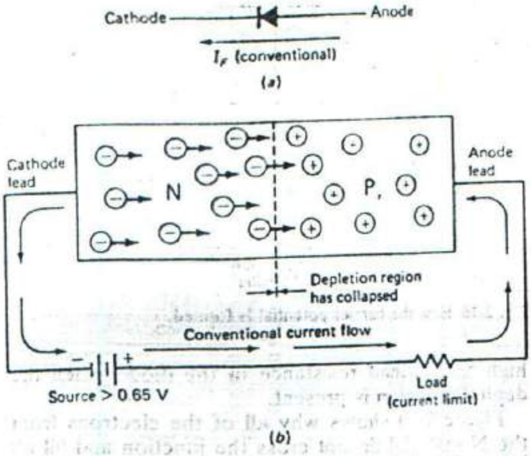


Fig. 2-11 Junction diode with forward bias. (*a*) Schematic symbol of diode. (*b*) Direction of current flow.

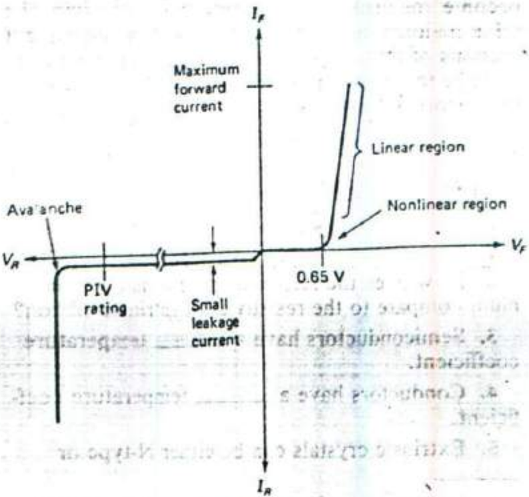


Fig. 2-13 Diode volt-ampere characteristic curve.

Figure 2-12 shows what happens when the diode is reverse-biased. The positive terminal of the external source is now connected to the N side of the diode. Unlike charges attract, and the electrons are moved away from the junction. This makes the depletion region wider. The center of the diode now exhibits intrinsic behavior, and very little current flows at room temperature. There will be some reverse current due to minority carrier action. Minority electrons do exist in the P material. These are driven by the reverse bias source to the junction, where they combine with minority holes in the N material. Since the number of minority carriers is directly related to temperature, reverse (leakage) current increases with temperature. At room temperature, the leakage current in many silicon diodes is so small that it cannot be measured with ordinary meters. It is usually in the nanoampere (nA) or microampere (μA) range. At elevated temperatures, the leakage current can increase to the point where it adversely affects circuit performance. Germanium diodes exhibit much more leakage current at room temperature. It is typically in the microampere range for small devices and in the milliampere (mA) range in large devices. This is another reason why most designs employ silicon devices.

Some important concepts of silicon diode behavior are summarized in Fig. 2-13. It is a volt-ampere graph. It shows diode current for various levels of forward and reverse bias. Forward bias is designated as V_F and ranges from the origin to the right on the horizontal axis. Forward current is designated as I_F and ranges from the origin to the top on the vertical axis. You can see the effect of the barrier potential by examining the forward bias portion of the graph. **There is no current flow until a forward voltage of 0.65 V is reached. At this point the volt-ampere graph is curved, because diode behavior is nonlinear near the turn-on point. This nonlinear region is nearly logarithmic, and some applications utilize this part of the curve when a logarithmic response is needed. As the forward bias is further increased, the graph becomes linear, as it would be for a resistor.** Figure 2-13 also depicts reverse bias (V_R) performance. Notice that as V_R increases from the origin toward the left, a small leakage current flows. This current is small and remains rather constant up to the peak inverse voltage (PIV) rating of the diode. Other designations, such as V_{RRM} (peak repetitive reverse voltage) may be used to specify this point. This maximum reverse bias voltage is typically specified from 50 to 1000 V in general purpose silicon diodes. If the maximum reverse voltage is exceeded, the diode will break down. Breakdown occurs at high potential where the field is strong enough to break covalent bonds. The electrons accelerate and collide with other electrons and knock them loose. **An avalanche of carriers results, and the graph (Fig. 2-13) shows a**

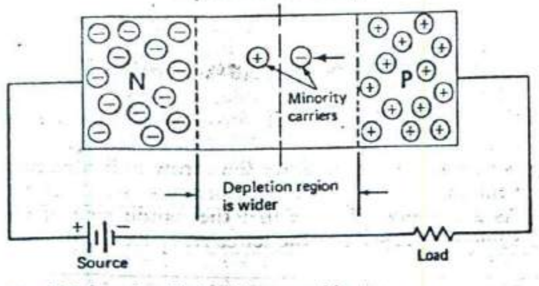


Fig. 2-12 Junction diode with reverse bias.

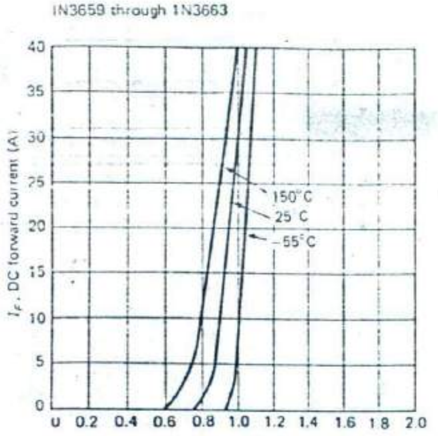


Fig. 2-14 Effect of temperature on forward voltage and conduction.

drastic increase in reverse current flow. This reverse flow, if not controlled, will destroy the diode.

As Fig. 2-14 indicates, volt-ampere characteristic curves can also be used to depict the effect of temperature on diode performance. Note that the curves show only forward bias performance. At 150°C, the diode turns on at 0.6 V and reaches a current of 10 A at 0.8 V. Compare this to the −55° curve, which shows turn on at 0.9 V and a current of 10 A at just over 1 V. This occurs because the barrier potential is temperature-dependent and decreases approximately 2 millivolts per degree Celsius (mV/°C). The part numbers 1N3659 through 1N3663 represent a series of 30 A silicon diffused junction diodes. The 1N3659 is the least expensive part and is rated for V_{RRM} of 50 V. The 1N3663 is the most expensive in the series and is rated for 400 V.

Diodes as Rectifiers

One of the major industrial applications for diodes is in the area of rectification. *Rectification* is a process of changing alternating current to direct current. Since diodes conduct well in one direction and oppose current in the other direction, they make excellent rectifiers. The rectification frequency is typically 60 hertz (Hz) and general purpose diodes are characterized for this frequency. ~~Chapter 6 treats rectification in detail.~~

Some industrial circuits require rectification at higher frequencies. As the frequency goes above 1 kilohertz (kHz), the diode will have to be derated because of the recovery time; this phenomenon is due to stored charges which prevent a diode from turning off instantaneously at the moment of reverse bias. The current that continues to flow when reverse bias is applied causes extra dissipation in the diode; for this reason it must be derated for high-frequency operation. It is known as the *recovery current*, and it also detracts from circuit efficiency. Special fast-recovery diodes have been developed to increase efficiency and reduce diode heating when high frequencies must be rectified. The recovery current flows because a conducting diode injects electrons into the P-type anode. These injected electrons are minority carriers in the P material. At the moment of reverse bias, these minority electrons will support a reverse current flow until they move back to the N-type cathode or recombine with holes. The recovery time is longest when the forward current is high because the number of minority electrons is a function of the forward current. Fast-recovery rectifiers speed up turn off by using gold doping in addition to the ordinary donor and acceptor doping. The gold provides extra sites in the crystal where recombination can occur. This makes it possible for the diode to recover (turn off) faster. These rectifiers recover in dozens or hundreds of nanoseconds and extend rectifier performance to the hundreds of kilohertz.

Schottky Rectifier

The Schottky rectifier is another special diode that extends performance into the megahertz (MHz) region. Schottkys are not PN junction devices but are included in this section because they are used in many of the same application areas (Fig. 2-15). An N-type silicon chip is bonded to a barrier metal (typically platinum). At this semiconductor to metal barrier the diode action is achieved. When the barrier is forward-biased, the electrons from the N-type cathode gain energy to cross the metal barrier. Since the electrons gain energy in order to cross, these devices are also commonly called *hot-carrier diodes*. Once in the metal, the electrons quickly give up this

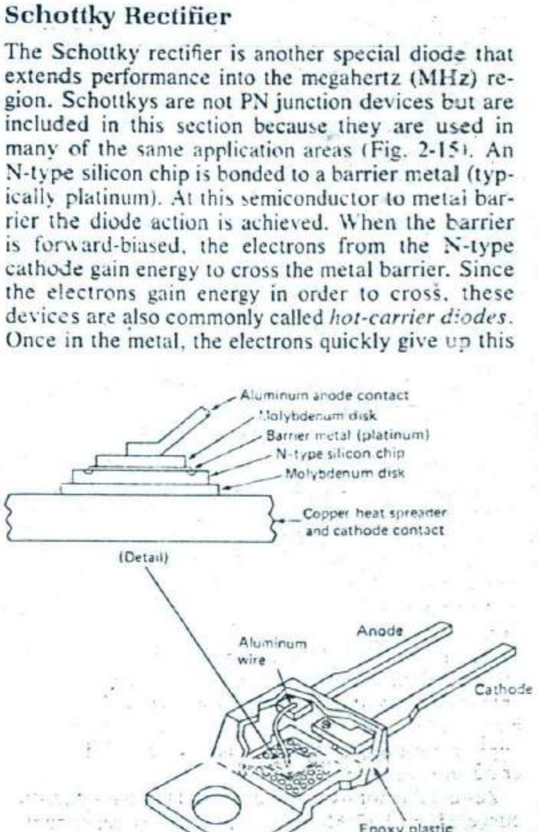


Fig. 2-15 Schottky rectifier T0-220.

extra energy and join the electron cloud associated with metallic conductors. Minority carrier storage is not a problem since the electrons are not minority carriers in the metal anode. Therefore, there is no recombination and no turn-off delay. Unfortunately, Schottky diodes are restricted to low-voltage applications since they cannot block reverse voltages much over 50 V. They are available with current ratings up to around 100 A. The volt-ampere curves in Fig. 2-16 show another interesting characteristic of Schottky diodes: They turn on at a much lower voltage. At room temperature, the turn-on point is 0.35 V, and at 10 A the forward voltage is only 0.63 V. Compare this with Fig. 2-14. You will also notice that the Schottky curves never become linear.

Zener Diodes

It was mentioned that diodes may be destroyed if allowed to avalanche. This is true for rectifiers. However, zener diodes (Fig. 2-17) are normally operated at the avalanche voltage. Fig. 2-17(*b*) shows the characteristic curve for a zener. Note that the *zener knee* is that part of the graph where reverse conduction takes place. The diode is safe if the reverse current is limited to the maximum value. Zeners are rated according to safe power dissipation. The maximum reverse current may be found by dividing the power rating by the zener voltage. For example, a 1-watt (W), 10-volt (V) zener will have a maximum reverse current rating of

$$I_{\text{MAX}} = \frac{1}{10}$$
$$= 0.1 \text{ A}$$

Thus, the diode will be safe at reverse currents up to 0.1 A. However, solid-state devices must be derated for operation above 25°C. This is why diodes are often mounted on thermally conducting assemblies called *heat sinks*.

Zener breakdown is different from avalanche breakdown. Zeners are very heavily doped PN devices. The heavy doping produces a very narrow depletion region, and even a moderate reverse bias produces an intense field across the narrow region. The intense field yields zener emission and the device begins to conduct. Actually, zener emission occurs up to around 4 V of reverse bias. Avalanche conduction is the major mechanism above about 6 V. The region from 4 to 6 V is a combination of avalanche and zener emission. However, zener diodes are available with breakdown ratings up to 200 V. What has happened is that the name *zener* has come to be applied to a broad range of diodes designed to operate in reverse breakover.

Zeners are not used as rectifiers. The characteristic curve (Fig. 2-17[*b*]) shows that ΔV_z is reasonably small from the zener knee to I_{MAX}. The impedance (*Z*) of a zener diode is given by

$$Z = \frac{\Delta V_z}{\Delta I_z}$$

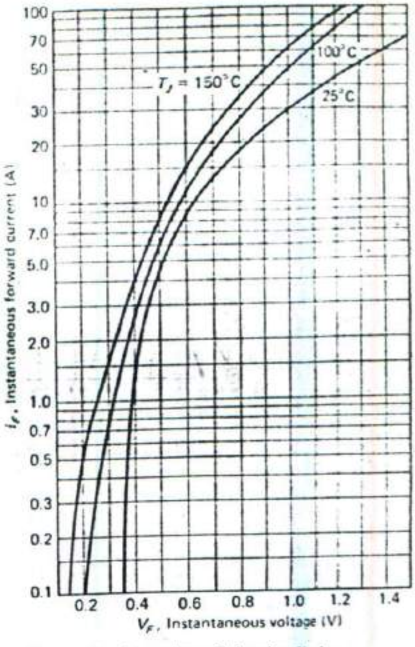


Fig. 2-16 Forward voltage for a Schottky diode.

The ideal zener diode will show no change in V_z and has an impedance of 0 Ω. The best practical zener diodes have a low impedance, which indicates that they approach the ideal. They show only a small voltage change over a range of reverse current. This makes them useful for voltage regulation and reference applications. Zeners are available in voltages from 1.8 to 200 V and in power ratings from 0.25 to 50 W. Figure 2-17(*a*) shows the schematic symbol for

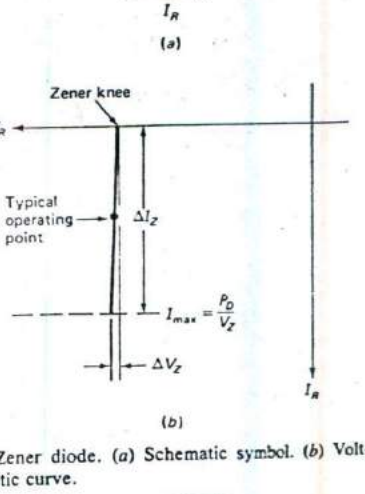


Fig. 2-17 Zener diode. (*a*) Schematic symbol. (*b*) Volt-ampere characteristic curve.

the zener diode. Note that the cathode has extra "wings" that imply the breakdown characteristic curve. Also notice that the conventional current direction is against the arrow when the device is operating in reverse breakover. If the anode of a zener is made 0.65 V positive with respect to the cathode, it will turn on. Zeners have the same forward characteristic curve as is shown in Fig. 2-13 for the rectifier diode.

The zener voltage tends to increase with temperature. Special temperature-compensated zener diodes are available for those applications in which a highly stable voltage reference is required. The temperature compensation is achieved by connecting a second diode in series with the zener. The second diode is forward-biased when the zener is operated in breakover, and it drops less voltage as temperature increases. This tends to cancel the change in zener voltage. These devices are called *reference diodes* or *stabistors*. They operate at around 6 V, and high-precision types are available with a maximum voltage change of only several millivolts (mV) and a temperature coefficient of 0.001 percent. Reference diodes do not have the same forward characteristic curve as ordinary zeners do, because of the second compensating junction.

Other Diode Types

A summary of diode schematic symbols is shown in Fig. 2-18. This illustration will give you an overview of the variety and application areas for diodes. Later chapters will provide more details. The bidirectional breakdown diode is a back-to-back zener device that can be used in transient suppression service or in ac power control. The constant current diode, which is actually a field effect transistor, is covered later in this chapter. The varicap or varactor diode utilizes the depletion region as the dielectric portion of a capacitor. If you refer to Fig. 2-12, you will see that the reverse-biased junction diode has much in common with a capacitor. The width of the depletion region is directly related to the reverse bias voltage. Thus, diodes can serve as voltage-variable capacitors. Light-emitting diodes are also common and make excellent indicators. Photodiodes are useful in sensing and control applications.

Figure 2-19 provides some examples of the way diodes are packaged. The size of the package is related to the device ratings. For example, the DO-15 case is usually restricted to diodes having maximum current ratings of 1 A or less. When the DO-15 case is used to house zener diodes, the maximum rating will be 1 W or less. The TO-3 metal case may be

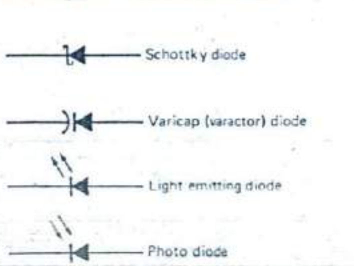


Fig. 2-18 Schematic symbols for various diode types.

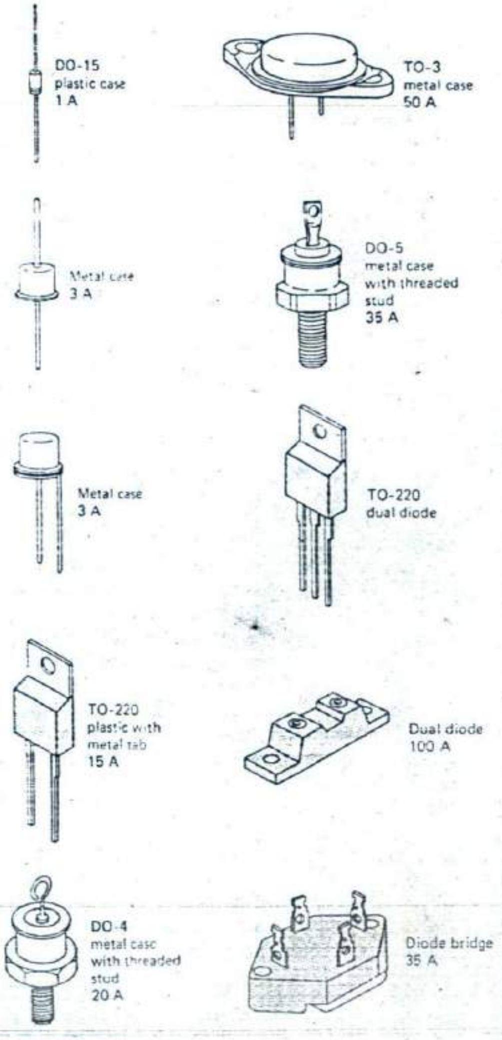


Fig. 2-19 Diode packages.

used to house diodes with current ratings up to 50 A. Current ratings are directly related to the size of the PN structure and its ability to transfer heat to its environment. Therefore, the high-current devices are physically larger and must be mounted in a package that allows efficient heat transfer. For example, the TO-220 case uses a metal tab for conducting heat away from the PN junction. Some packages are designed to house more than one diode. The dual-diode and diode bridge packages shown in Fig. 2-19 are examples. Figure 2-20 shows an elaborate package designed for three-phase service and forced-air cooling. Water cooling is also used in some industrial applications.

REVIEW QUESTIONS

7. A PN junction diode will exhibit a depletion region under conditions of zero bias and ______ bias.

8. Diode leakage current is due to ______ carriers.

9. Diode avalanche occurs at high values of ______ bias.

10. Diodes turn on at lower values of forward bias as temperature ______.

11. When a diode is used to convert alternating current to direct current it is called a ______.

12. Schottky rectifiers turn off very fast because they exhibit no ______ carrier storage.

2-3
JUNCTION TRANSISTORS

Figure 2-21 shows the way the planar diffusion process can be used to build a device with two junctions. Once again, the crystal is continuous, and the junctions represent boundaries between the P- and N-type regions. The structure is called a *bipolar NPN transistor. Bipolar* means that the two polarities of current carriers are present: electrons in the N material and holes in the P material. Unipolar transistors will be covered in the next section.

The schematic symbol for the NPN transistor is also shown in Fig. 2-21(*a*). The leads are identified as emitter, base, and collector. These names tell us something about the functions. The *emitter* emits the carriers, the *base* controls the flow of carriers, and the *collector* collects the carriers coming from the emitter. Note that the physical structure of the transistor is arranged so that the carriers coming from the emitter must pass through the base region to reach the collector. The schematic symbol shows an arrow on the emitter lead. This arrow points in the direction of conventional current flow as it does in the diode symbol.

Another way to build a bipolar junction transistor is illustrated by Fig. 2-22. Here, the base is N material, and the collector and emitter are made up of P material. This structure is called a *PNP transistor*, and the schematic symbol shows the emitter arrow pointing in. Compare Fig. 2-21 with Fig. 2-22. As a memory aid, think of NPN as denoting "Not Pointing iN." Type NPN and PNP transistors are electrical complements. They both can achieve the same basic functions but with opposite emitter-collector flows. They are operated with opposite terminal polarities as well. This feature makes them electrical complements and offers circuit designers some interesting choices.

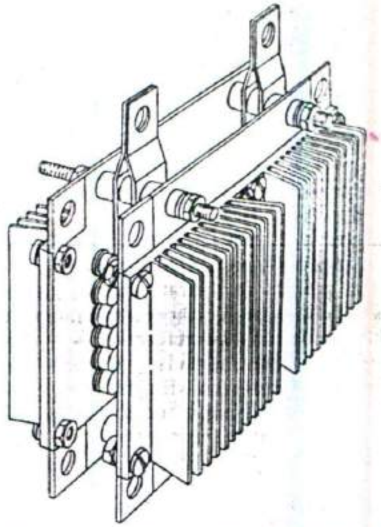


Fig. 2-20 Three-phase rectifier assembly with forced air cooling rated at 650 A.

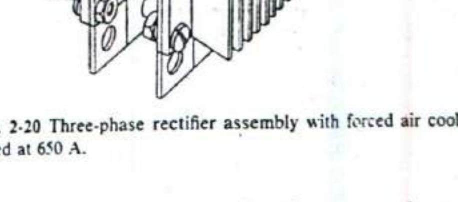


Fig. 2-21 NPN transistor. (*a*) Schematic symbol. (*b*) Physical structure.

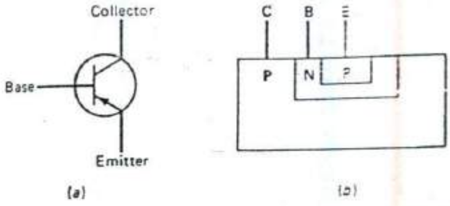


Fig. 2-22 PNP transistor. (*a*) Schematic symbol. (*b*) Physical structure.

Figure 2-23 shows an NPN transistor with both junctions biased. The transistor structure is shown in "sandwich" form to illustrate the circuit action more clearly. Voltage V_{CB} reverse biases the collector-base junction and a depletion region is shown. Voltage V_{BE} forward biases the emitter-base junction, and that depletion region is collapsed. On the basis of what we know about junction behavior, we can expect conventional current to flow from the base to the emitter and zero collector current or a small collector current due to leakage. However, Fig. 2-23 shows that the collector current I_c is substantial. This is so because of the way the transistor is fabricated. The base is a very thin region in the crystal: on the order of 0.025 millimeters (mm) (about a thousandth of an inch). It is moderately doped. The emitter is heavily doped. When the emitter-base junction is forward-biased, the emitter sends large numbers of electrons into the base region. Only a few of the emitter carriers will combine with a hole in the base material because the base does not contain a great number of holes. Most of the emitter carriers will come under the influence of the collector field and be swept through the depletion region and into the collector. Notice that the collector is positive with respect to the base and will attract any electrons that it can get. It is a statistical process whereby the emitter electrons are far less likely to combine with a hole and become base current than they are likely to be swept into the collector. Typically, only 1 or 2 percent of the emitter carriers become base current; the balance, 99 or 98 percent, become collector current.

Even though the base current is only a small percentage of the total current in Fig. 2-23, it is very important because it is a control current. Suppose that V_{BE} is adjusted to 0.3 V. If the transistor is silicon, this voltage is not adequate to collapse the emitter-base depletion region. Now, the emitter will not emit its electrons into the base region. They will never get close enough to the collector field to be swept across the collector-base junction. Therefore, all three currents are now zero. This is a very important concept. A small current (I_B) is controlling much larger currents (I_E and I_C). Another way to demonstrate the control capabilities of the base would be to place a large resistance in series with the base lead in Fig. 2-23. This resistance would reduce the base current, say, to half of what it was. Now we will find the emitter and collector currents about half of what they were before the resistor was added. This shows that a transistor is capable of gradual control: A change in the base circuit will be accompanied by a proportional change in the emitter and collector circuits.

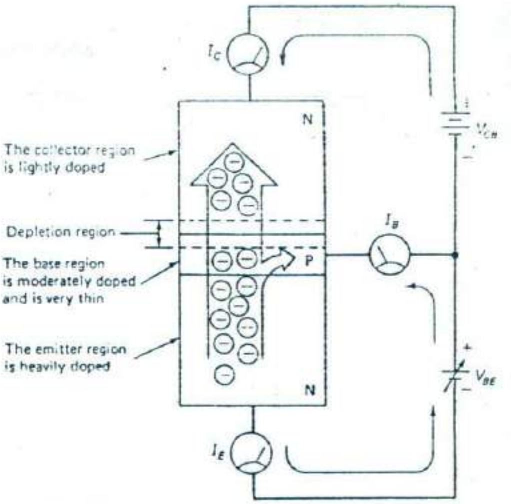


Fig. 2-23 NPN transistor currents.

How is the transistor controlled? We have discussed changing the base-emitter bias voltage. We have also discussed adding resistance to the base circuit to reduce the base current. Both seemed to control emitter and collector behavior. This leads us to the conclusion that either point of view can be taken. In other words, the transistor can be viewed as a voltage-controlled device or as a current-controlled device. Which will be used is a matter of convenience. In Chapter 6, voltage gain in transistor amplifiers is discussed. *Voltage gain* implies that a small voltage change is accompanied by a larger voltage change in another part of the circuit. In this chapter, we are more concerned with the current gain of the bipolar transistor. Even though voltage gain and current gain are both viable points of view, the bipolar transistor is definitely a current-amplifying device. When it is operating as a linear amplifier, any change in base current will cause a corresponding change in emitter and collector currents. The most important idea is this: Base current controls emitter and collector current. There are voltage-controlled transistors, and they will be discussed in the next section.

Bipolar junction transistors are the workhorse of industrial electronics. It is essential that you understand their characteristics. Refer to Fig. 2-24, which shows PNP action, and compare it to Fig. 2-23. The PNP transistor is also reverse-biased from collector to base and forward-biased from base to emitter. Note that both external sources must be reversed to accomplish proper bias for the PNP device. The holes in the P-type emitter are plentiful, and most do not recombine with electrons in the base. The collector is negative, and a majority of the holes are attracted to it. Once again, the base current is very small but critical in that it controls both emitter and collector currents. Both NPN and PNP transistors are capable of current gain, and both are useful as amplifiers and controllers. However, it should be apparent that they are not interchangeable since the applied polarities are reversed.

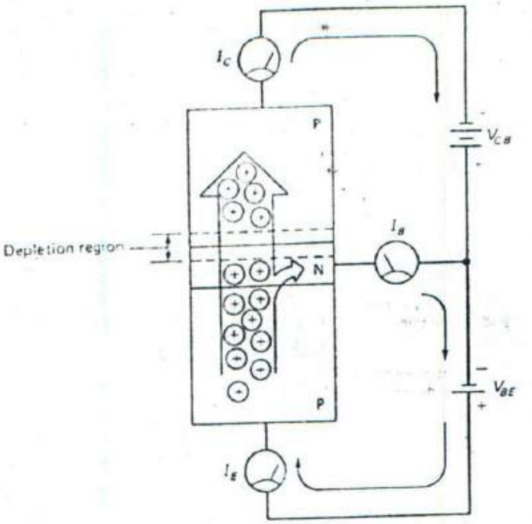


Fig. 2-24 PNP transistor currents.

Figure 2-24 shows holes moving from emitter to collector. Actually, the positions move as discussed in the first section of this chapter. The illustration makes the PNP transistor currents easier to understand. The electrons move from the collector to the emitter by moving from hole to hole. This type of flow is a little slower, and for this reason NPN transistors are better choices for high-frequency applications. The NPN transistor is also more convenient to use in many negative ground circuits. Finally, device manufacturers make a broader line of NPN transistors than of PNP transistors. For these reasons, most circuits use NPN transistors. However, the complementary characteristics of PNP devices make them very useful for certain applications, and many industrial devices will use both types.

Let's summarize some very important ideas. The collector-base junction must be reverse-biased for proper transistor action. The emitter-base junction must be forward-biased for the transistor to be on (to conduct carriers from its emitter to collector). The base current is a small percentage of the total current. The emitter current is the total current ($I_E = I_C + I_B$), and the collector current is almost as large. The base current controls the emitter and collector currents. The NPN and PNP transistors are electrical complements.

Characteristic curves are an important aid to understanding transistor behavior and are also a useful way to present data. Figure 2-25 shows transfer curves for a typical NPN transistor. The curves transfer (relate) the effect of base-emitter voltage to collector current. At room temperature (25°C), the transistor begins conducting at about 0.55 V. At 150°C, the base-emitter turn-on voltage is only about 0.25 V. For a collector current of 40 mA, V_{BE} varies from 0.4 to 0.7 V from the high temperature to the low temperature. The transfer curves show that the transistor is a temperature-sensitive device.

A collector family of characteristic curves is shown in Fig. 2-26. These are not transfer curves since the collector voltage is plotted against the collector current. The graph shows that the collector-to-emitter voltage has little effect on the collector current over most of the graph. For example, inspect the lowest curve in the family. It is labeled for a base current of 0 mA. This curve shows that the collector current is constant at 0 over a range of collector voltage from 0 to 300 V. This is reasonable because with no base current there is no emitter or collector current. What happens as the collector voltage goes over 300 V? The curve shows that the collector current starts to increase. This phenomenon is called *reach-through*. The collector field is now intense enough to reach through the base and into the emitter. Transistors are not normally operated at collector potentials high enough to cause reach-through.

What is controlling collector current in Fig. 2-26 when the collector voltage is normal? The curves provide the answer: base current. Look at the second curve up. It represents collector behavior for a base current of 0.2 mA. The collector current is nearly

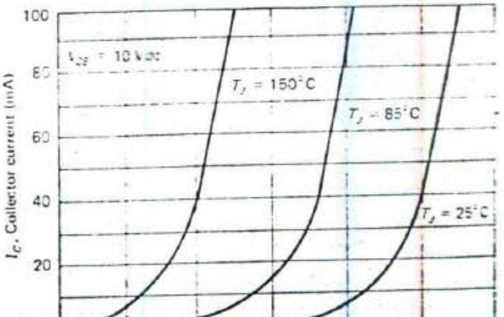


Fig. 2-25 Transfer characteristic curves.

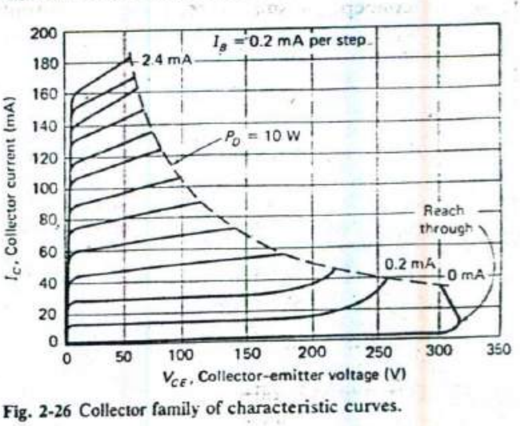


Fig. 2-26 Collector family of characteristic curves.

constant at about 12 mA for a range of collector voltages near 0 to about 200 V. The collector resistance is given by

$$R_c = \frac{\Delta V_{CE}}{\Delta I_c}$$

This equation shows that the collector resistance is high because the change in collector current is small. Once again it must be stressed that base current, not collector voltage, is the major controlling factor. The only two places on the graph where collector voltage does have significant control are near 0 V and at the high end, where reach-through occurs. However, it can be seen that the curves are not horizontal for the higher values of collector current. The slope of these curves indicates that collector voltage is playing a partial role in setting collector current and that the collector resistance is lower for the higher values of collector current.

The dc gain of the transistor can now be investigated with some numbers. *Beta*, or h_{FE}, is defined as the ratio of collector current to base current.

EXAMPLE

Calculate the gain for an operating point of 50 V and a base current of 0.2 mA by using the data from Fig. 2-26.

SOLUTION

$$h_{FE} = \frac{I_C}{I_B}$$

From Fig. 2-26, at $V_{CE} = 50$ V and $I_B = 0.2$ mA, $I_C = 12$ mA.

$$= \frac{12 \times 10^{-3}}{0.2 \times 10^{-3}}$$
$$= 60$$

The ac current gain, h_{fe} (ac beta), is found by

$$h_{fe} = \frac{\Delta I_C}{\Delta I_B} \bigg| V_{CE}$$

The vertical bar in this equation signifies that V_{CE} is to be held constant. Although not plotted, the emitter current can also be found. Remember that the emitter current is the total transistor current. It can be found by

$$I_E = I_B + I_C$$

Using the values from the previous example,

$$I_E = 0.2\text{ mA} + 12\text{ mA}$$
$$= 12.2\text{ mA}$$

The value of h_{FE} varies somewhat as the operating point is changed. The collector family of curves in Fig. 2-26 will now be used to calculate gain for an operating point of 50 V and a base current of 1 mA. These two conditions intersect at a corresponding collector current of 80 mA.

$$h_{FE} = \frac{80\text{ mA}}{1\text{ mA}}$$
$$= 80$$

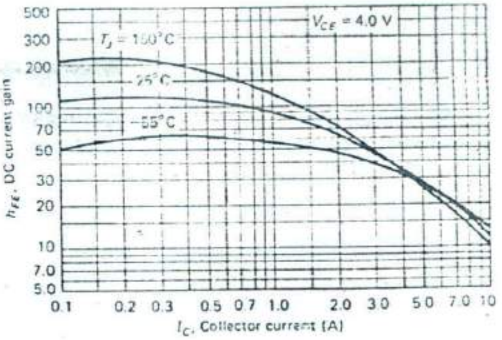


Fig. 2-27 The dc current gain.

Figure 2-27 is another type of transistor graph and shows that h_{FE} varies with collector current and with temperature. Most transistors show best gain at some moderate or intermediate value of collector current. All bipolar transistors show increasing gain as temperature increases.

Refer to Fig. 2-26 to see how a constant power curve can be used to define the safe operating area for a transistor. Transistors are rated according to maximum collector dissipation. Small transistors used to switch and amplify small currents may have a maximum collector dissipation of 100 mW. Large transistors are available with collector dissipation ratings of 300 W. The collector dissipation of a transistor is found by

$$P_C = I_C \times V_{CE}$$

The constant power curve in Fig. 2-26 represents a collector dissipation of 10 W. Any point on the power curve represents a current-voltage product of 10. For example, the curve passes through the 100-mA, 100-V intersection (100 mA × 100 V = 10 W). Since there are an infinite number of operating points, the power curve helps us decide which are safe and which are not. Any point falling on the curve or to the left will not exceed the maximum collector dissipation. Any point falling to the right of the curve will exceed the maximum collector dissipation.

If the maximum collector dissipation is exceeded for any length of time, the transistor will overheat and be damaged or destroyed. This is known as the *thermal limit of the transistor*. There are other failure modes. A high current, even if momentary, can cause a bonding wire within the transistor case to melt (open). Finally, second breakdown may occur. *Second breakdown* is a failure mode peculiar to bipolar power transistors and occurs within the safe operating area defined by a constant power curve. It is associated with high collector voltages which set up fields in the crystal. These fields can cause the current to be focused into a tiny area about the diameter of a human hair. A localized heating, which may melt a hole from the emitter to the collector, shorting out the transistor, results. Figure 2-28 combines the fail-

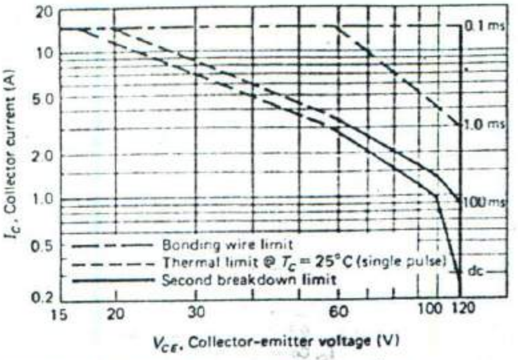


Fig. 2-28 Safe operating area graph for a transistor.

ure modes into one safe operating area graph. Any current greater than 15 A will exceed the bonding wire limit. The dc curve shows the thermal limit as the other controlling factor up to 60 collector-emitter volts. Second breakdown is the limiting factor for collector-emitter voltages from 60 to 120 V, where the maximum current is less than would be allowed by the transistor's dissipation ratings. Figure 2-28 also shows that greater voltages and currents are safe for pulse-mode service. A 0.1 millisecond single pulse will not damage the transistor unless it exceeds the bonding wire limit or involves a collector-emitter voltage greater than 120 V.

Figure 2-29 shows that transistors must be derated above 25°C. For example, at 100°C the thermal derating factor is 0.4. This means that the transistor is capable of dissipating only 40 percent of its rated power at that temperature. It can be seen that second breakdown must also be derated at elevated temperatures, but not as much as the thermal ratings. At 100°C the second breakdown ratings are 80 percent of maximum.

Transistors have three operating modes: cutoff, linear, and saturation. Figure 2-30 depicts the modes and shows an electrical equivalent for each. The transistor circuit is shown at the left and the equivalent emitter to collector circuit at the right. It is important to notice that the base circuit conditions determine which mode is selected. Cutoff (Fig. 2-30[*a*]) is caused by zero base current and results when the base switch is open. The base current is zero, and so are the other currents. All the supply voltage drops across the transistor, which is acting as an open switch from collector to emitter. The load is off for the cutoff mode. The linear mode in Fig. 2-30(*b*) is selected by closing the base switch. The base current is now some moderate value. The collector current is now beta times the base current, and approximately half of the supply voltage drops across the collector-emitter circuit. Notice that the collector-emitter circuit is now acting as a resistance. The load is partially energized in the linear mode. Saturation (Fig. 2-30[*c*]) is caused by the base current's reaching some high value. When one of the resistors has been removed from the base lead this condition results. The collector circuit is now acting as a closed switch. The drop from collector to emitter approaches 0 V, and the current through the load is limited by Ohm's law: the resistance of the load and the value of V_{CC}.

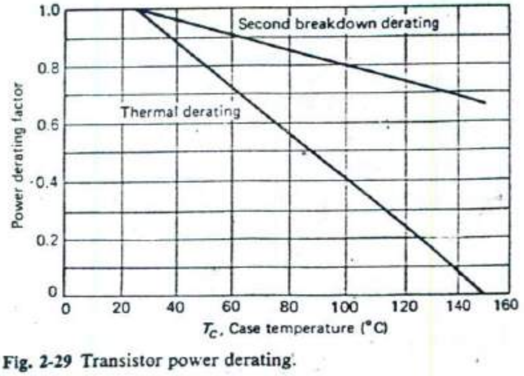


Fig. 2-29 Transistor power derating.

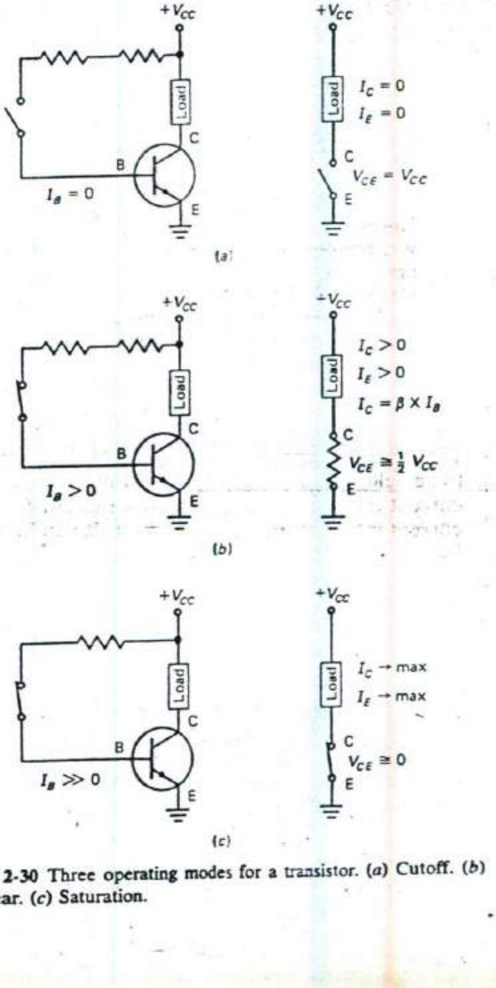


Fig. 2-30 Three operating modes for a transistor. (*a*) Cutoff. (*b*) Linear. (*c*) Saturation.

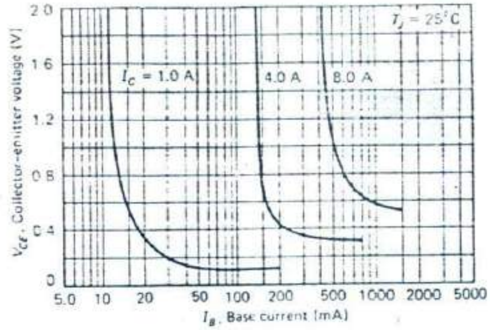


Fig. 2-31 Collector saturation.

The load is fully energized for the saturation mode.

Figure 2-30 also shows that a single source voltage is all that is required to bias both junctions of the transistor properly. When the emitter lead is grounded, a single positive supply will reverse bias the collector and forward bias the base of an NPN transistor. In actual use, the load could represent a motor, a lamp, a heating element, or a relay coil. The linear mode may not be desired in some applications. For example, in digital circuits the load is switched on or off (saturation or cutoff). In other applications, the linear mode may be desired. For example, if the load is a motor, smooth speed control can result from gradually adjusting the base current.

Suppose the circuit of Fig. 2-30 is used to provide on-off control for a dc motor. In this case, the circuit will operate in cutoff or in saturation. When in saturation, the transistor should closely approximate a closed switch from the emitter to the collector. Now look at Fig. 2-31. This graph shows the saturation characteristics of a typical transistor. It shows that perfect switching action is not possible. For a load (collector) current of 1 A, the base current must be at least 100 mA to saturate the transistor fully. Even with this high base current, the graph shows that the collector-to-emitter voltage is a little over 0.1 V. The saturated transistor is approximating a closed switch but not quite attaining it. At high load currents, say 8 A, the graph shows that even more base current is required for saturation and that even more voltage drops from collector to emitter. This is an important factor when transistors are used to switch large loads, as they often must in the industrial environment. Any drop across the transistor when it is saturated detracts from circuit efficiency since some of the circuit energy will heat the transistor, rather than all of it being available for the load. The base losses are also significant in these cases. The graph shows that the base current must be in excess of 1 A to saturate the transistor for an 8-A load current. This represents another loss which detracts from the overall circuit efficiency. The next section of this chapter investigates a newer type of transistor that improves efficiency in these cases.

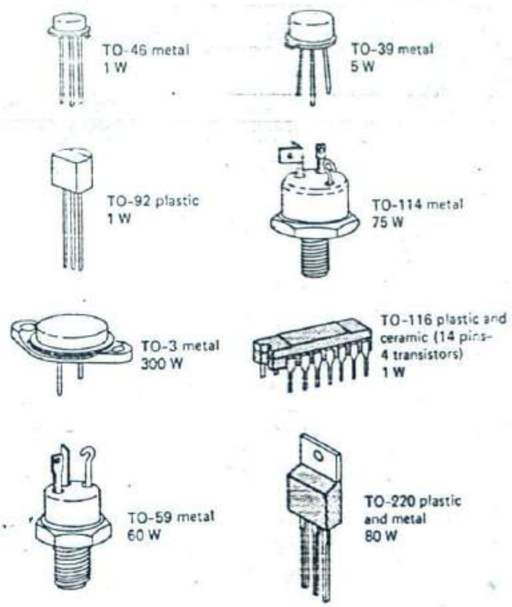


Fig. 2-32 Transistor packages.

The larger the transistor structure, the larger its current and thermal ratings and the better its saturation characteristics. Figure 2-32 shows some popular packages for transistors. The smaller packages are limited in wattage and current ratings. Heat flow is important in large dissipation devices. The large metal packages provide good heat transfer from the crystal to the external heat sink. Some packages house more than one transistor. The TO-116 ceramic package can house four transistors. In that case each transistor will have a maximum dissipation of 250 mW.

REVIEW QUESTIONS

13. On the schematic symbol for an NPN transistor, which lead has an arrow, and how does it point?

14. The base current in a transistor is very important because it _______ the other currents.

15. How must the collector-base junction of a transistor be biased?

16. If a transistor is to be on, how must its base-emitter junction be biased?

17. Refer to Fig. 2-25. If V_{BE} is fixed at 0.6 V and the transistor's temperature is increased from 25 to 85°C, what will happen to the collector current?

18. Use Fig. 2-26 to find the collector current when the collector voltage is 100 V and the base current is 0.6 mA.

2-4
FIELD EFFECT TRANSISTORS

The bipolar junction transistors covered in the last section are current-controlled devices. This section investigates transistors that normally have no input current; they are called ***field effect transistors*** and they are *voltage-controlled*. Field effect transistors (FETs) are unipolar devices. They involve a single conducting channel which can be of either N or P type material. The single conducting channel supports device current with majority carriers only. Bipolar transistors support flow with both majority and minority carriers. For example, in an NPN transistor the emitter injects electrons into a P-type base region. These injected electrons are minority carriers and delay the turn-off of the transistor since they must be swept out of the base (or recombine) before the collector current can cease. The FETs offer advantages in switching service since they do not suffer the delays associated with minority carrier storage. Because they also demand no input current, they are easier to drive. You should recall that the required base current to saturate a power bipolar transistor fully is substantial. The FETs are also less temperature-sensitive and less susceptible to second breakdown in high-power applications. These reasons explain why FETs and power FETs are gaining in popularity.

Figure 2-33 shows the structure and the schematic symbol for an N-channel junction field effect transistor (JFET). The source lead sources the electron current, the drain lead drains it, and the gate lead controls the flow from source to drain. Conventional current flow is from drain to source. The structure shows that the source lead is a contact at one end of a semiconducting N channel, and the drain lead contacts the other end of the same channel. No junctions are crossed, and only majority carriers (electrons, in this case) support the flow. The gate lead is reverse-biased with respect to the N channel. This sets up a depletion region and prevents any gate current from flowing. The negative gate repels the majority electrons in the channel and manages to push some of them down into the substrate, thus increasing the channel resistance and decreasing the source and drain currents. This is known as the *depletion mode of operation*, since the channel carriers are depleted by gate bias. If the gate is made sufficiently negative, all of the electrons are removed from the channel and the drain current ceases. This is called *cutoff*.

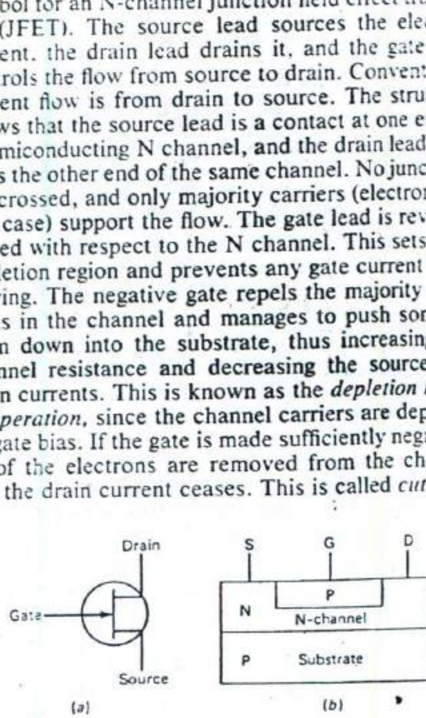


Fig. 2-33 N-channel junction field effect transistor. (*a*) Schematic symbol. (*b*) Structure.

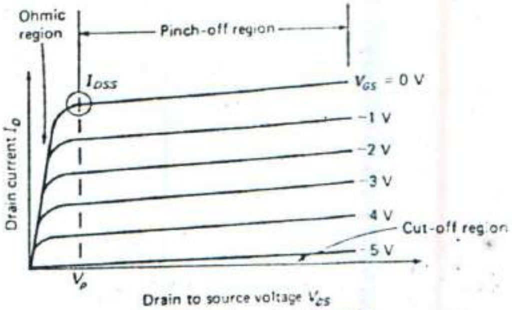


Fig. 2-34 N-channel JFET characteristic curves.

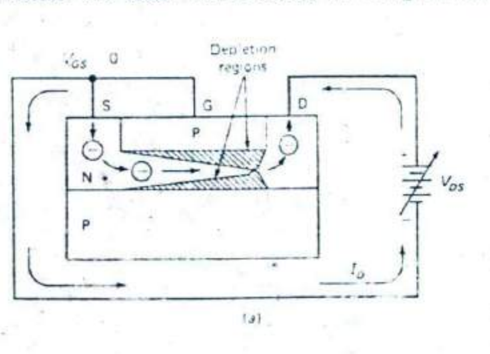
Figure 2-34 shows the drain family of characteristic curves for the N-channel JFET. It can be seen that as the voltage from gate to source (V_{GS}) is made increasingly negative, the drain current decreases. When the gate voltage is more negative than 5 V, the device cuts off and there is no drain current. This is shown as the cut-off region in Fig. 2-34. The graph also shows an ohmic region and a pinch-off region. In the *ohmic region* the drain current is a function of the drain-to-source voltage. It is due to the resistance of the channel, and the graph shows the expected linear volt-ampere relationship. The pinch-off region shows a constant current behavior, where the drain current does not vary over a wide range of drain voltage. The pinch-off voltage is shown as V_P on the graph. Most switching and amplifying operations occur in the pinch-off region of Fig. 2-34. Therefore, these applications require a drain supply greater than V_P.

Pinch-off is explained in Fig. 2-35(*a*). The electron current is shown inside the transistor, and conventional flow is shown outside the transistor. The gate lead is tied to the source lead, and therefore V_{GS} is equal to zero. The drain supply is adjustable. As the drain voltage is gradually increased from 0 V, the transistor initially operates in the ohmic mode. As the drain voltage is increased further, a significant voltage gradient begins to form across the N channel. A gradient is a gradual voltage change along the length of a resistive path that is supporting current flow. The channel current creates a gradually increasing positive voltage from left to right. Therefore, the right end of the channel is more positive with respect to the gate and the substrate than the left end. Or, to say it another way, the gate is more negative with respect to the channel at the right end. This causes a gradually increasing depletion region to form. At pinch-off, the channel is constricted and allows only so much flow regardless of drain voltage. In transistor applications, the pinch-off current is varied by the gate-to-source bias voltage.

The value of drain current that flows during pinch-off when $V_{GS} = 0$ is called I_{DSS} and is shown in Fig. 2-34. It is the maximum current that will flow during normal operation. Every JFET has a specific value of I_{DSS}, and it varies quite a bit from device to device, just as h_{FE} does among bipolar transistors.

The FETs can be used as voltage-controlled resistors. In the ohmic region, shown in Fig. 2-34, the resistance of the drain-source channel is a function of the gate-source voltage alone, and the JFET will behave as an almost pure ohmic resistor. The supply voltage is set at V_P (or less) to operate the device in its ohmic region. With zero gate bias, the current will be equal to I_{DSS}, and the drain-to-source resistance will be at its minimum. As the gate-to-source voltage is made increasingly negative, the channel resistance will increase.

Constant-current or current-limiting diodes are based on pinch-off and are actually FETs with a common gate-source (cathode) connection. Fig. 2-35(*b*) shows a Motorola MC1303 current-limiting diode. Its characteristic curve is shown in Fig. 2-35(*c*). Current I_P is the specified current (3 mA ± 0.6 mA) at the specified test voltage V_T (25 V). Voltage V_{PO} is the maximum voltage that can be applied to the device without exceeding its power or breakdown ratings. Impedance Z_K is the impedance at the limiting voltage ($V_L = 6$ V) and is equal to 50,000 Ω minimum for the diode shown. An ideal current source has an infinite internal impedance, and Z_K is a measure of how closely the actual device approaches the ideal. These diodes are designed for applications requiring a current reference or a constant current over a specified voltage range.

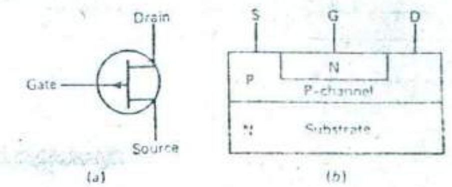


(*a*)

(*b*)

(c)

Fig. 2-35 Pinch-off and a typical field effect diode. (*a*) Pinch-off in an N-channel JFET. (*b*) MC1303 current limiting diode. (*c*) Characteristic curve of MC1303.

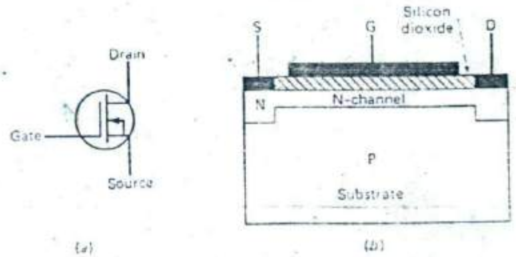


(*a*) (*b*)

Fig. 2-36 P-channel junction field effect transistor. (*a*) Schematic symbol. (*b*) Structure.

The structure and schematic symbol for the P-channel JFET are shown in Fig. 2-36. These devices are the electrical complements of N-channel JFETs. Proper operation for the P-channel transistor includes a positive gate-to-source voltage and a negative drain-to-source voltage. The positive gate voltage will repel the majority holes from the P channel and thereby control drain current. The drain current will be opposite in direction from the N-channel JFET. Note the difference in the schematic symbols. Compare Fig. 2-36 with Fig. 2-33 and look at the arrow on the gate lead. As a memory aid, associate N channel with pointing iN.

All JFETs are operated with reverse bias on their gate leads to prevent gate current. However, there will be some temperature-dependent gate leakage. Also, a large input signal may momentarily overcome the reverse bias and turn on the gate diode. An appreciable amount of gate current will flow, and this will reduce the amplitude of the input signal. These disadvantages are overcome by insulating the gate terminal from the channel with a thin layer of silicon dioxide (metallic oxide). Those FETs that use this technique are known as *metallic-oxide semiconductor field effect transistors*, or MOSFETs. They may also be called *insulated gate field effect transistors* (abbreviated *IGFETs*). Figure 2-37 shows the structure and schematic symbol for an N-channel MOSFET. The structure and schematic symbol show that there is no electrical contact from the gate to the rest of the transistor structure. These devices operate in the depletion mode as do JFETs. A negative voltage applied to the gate terminal will repel electrons from the N channel and reduce drain current. The N-



(*a*) (*b*)

Fig. 2-37 An N-channel MOSFET. (*a*) Schematic symbol. (*b*) Structure.

channel MOSFET characteristic curves will be similar to those shown in Fig. 2-34 for the N-channel JFET.

The MOSFETs may also be operated in the enhancement mode. Notice that the P-type substrate in Fig. 2-38 extends all the way up to the oxide layer. There is no channel connecting the source and the drain. This is a normally off device. Note that the schematic symbol uses a broken line to indicate that there is normally no channel connecting the source to the drain. With zero gate-source bias, the device is cut off and the drain current will be zero. Applying a positive gate voltage causes electrons to be attracted near the gate terminal, and an enhanced N channel is the result. Now drain current will flow. The N-channel characteristic curves are shown in Fig. 2-39. It should be clear that drain current increases with positive gate-source bias. The enhancement mode is only feasible when the gate is insulated from the channel. Any attempt to operate a JFET in the enhancement mode would cause gate current, which is undesirable. In addition, P-channel MOSFETs are available, and they are electrical complements to the N-channel devices. The arrow on the schematic symbols in Figs. 2-37 and 2-38 will be reversed for P-channel transistors.

All of the FETs discussed to this point use a lateral flow of current between source and drain. A newer structure has been developed with the drain connection at the bottom of the device. This allows a vertical flow between source and drain. The vertical format allows a wide, short channel with very low resistance. Thus, vertical FETs (often called *VFETs* or *DMOS* or *VMOS devices*) are very attractive for high-power applications.

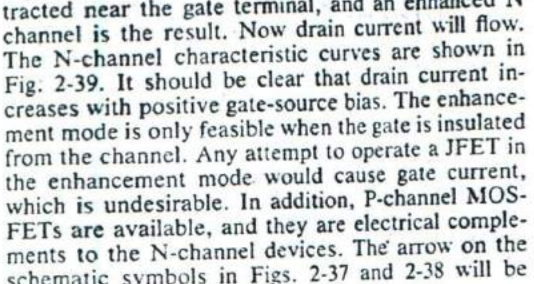


Fig. 2-38 N-channel enhancement MOSFET. (*a*) Schematic symbol. (*b*) Structure.

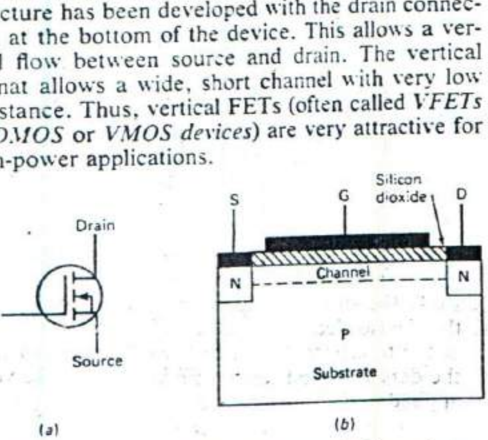


Fig. 2-39 N-channel enhancement mode characteristics.

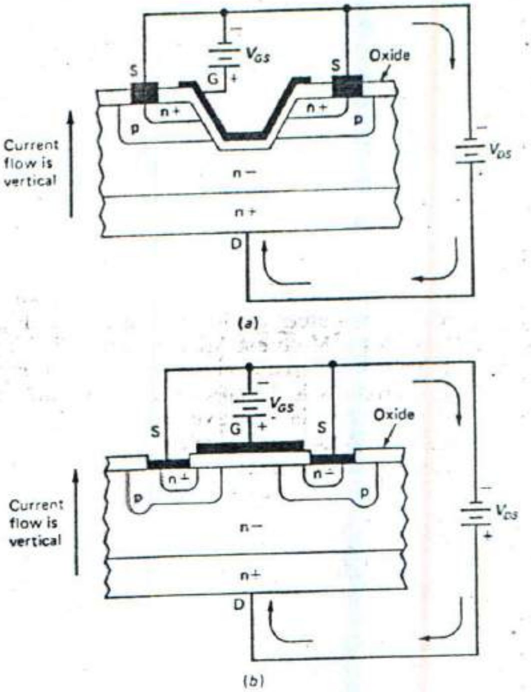


Fig. 2-40 Power FETs. (*a*) VMOS. (*b*) DMOS.

Power FETs are available in two basic structures, as shown in Fig. 2-40. The schematic symbol for either type is the same as shown in Fig. 2-38. They are also available as P-channel devices, in which case the arrow on the symbol will be reversed. The VMOS structure in Fig. 2-40(*a*) uses an etched V groove that extends through the n+ and p and slightly into the n− region. The n+ denotes heavy N-type doping (more than 10^{19} impurity atoms per cubic centimeter), and n− denotes light N-type doping (fewer than 10^{15} impurity atoms per cubic centimeter). There is an oxide layer that insulates the V-shaped metal gate structure from the crystal. The P-type region near each of the drain contacts prevents a continuous N channel between source and drain. The VFETs are enhancement mode devices and are normally off. Applying a positive voltage to the gate causes electrons to be attracted into the P regions, and an enhanced N channel results. Figure 2-40 shows that conventional current flow is vertical from the drain contact on the bottom to the source contacts on the top. The overall N channel is very short and wide, and an extremely low ON resistance can be achieved. These devices are very attractive for switching and controlling large currents in industrial circuitry.

The DMOS shown in Fig. 2-40(*b*) uses the double-diffused structure. The gate metal is in a flat plane and is insulated from the crystal by metallic oxide. It is also an enhancement mode device. A positive gate-to-source bias voltage will attract electrons into

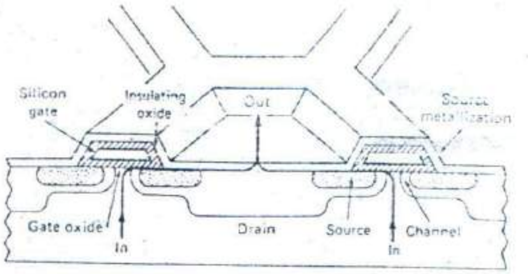


Fig. 2-41 HEXFET® structure (courtesy International Rectifier Company).

the P-type wells and complete the N channel between the source and drain. Once again, current flow is vertical through a wide, short channel for low ON resistance. The International Rectifier Company has registered the name *HEXFET* for their DMOS power FETs. Figure 2-41 shows the HEXFET arrangement. A single HEXFET is made up of hundreds or thousands of hexagonal-shaped source cells. The cells are very dense (over 500,000 per square inch) and act in parallel to provide very low ON resistance. Figure 2-42 shows some typical transfer characteristic curves for 40-, 75-, and 150-W HEXFETs. These curves indicate the high transconductance that power FETs are noted for. Transconductance (g_{fs}) is a transfer characteristic and relates the change in drain current that occurs for a change in gate voltage. The curves for the 150-W transistor in Fig. 2-42 show a change in drain current of 10 A (from 10 to 20 A) when the gate changes 1 V (from 5 to 6 V). The transconductance is

$$
\begin{aligned}
g_{fs} &= \frac{\Delta I_D}{\Delta V_{GS}} \bigg| V_{DS} \\
&= \frac{10}{1} \\
&= 10 \text{ siemens (S)}
\end{aligned}
$$

You should recall that conductance is the reciprocal of resistance and is measured in siemens (S). Transconductance is the figure of merit for power FETs (and small signal FETs) just as h_{FE} is for bipolars.

Bipolar transistors are still very popular; FETs and especially power FETs are newer devices and have been increasing in popularity. Bipolar power transistors have several limitations which detract from their usefulness. They exhibit minority carrier storage, which makes them slower in switching applications. They suffer from second breakdown. Finally, they are very temperature-sensitive. It was shown that beta goes up with temperature. This can lead to a condition known as *thermal runaway*. When a transistor is conducting significant current and dropping significant voltage, it gets hot. This makes its gain increase and it tends to conduct even more current. Now it gets even hotter and the gain increases still more. The transistor may run away and be damaged or destroyed by heat. Power FETs eliminate minority carrier storage problems, do not exhibit thermal runaway tendencies, and have greatly reduced second breakdown problems. Last but not least, they are easy to drive since there is no gate current. In low-power applications, bipolar devices show better gain and usually offer more performance for the money. Thus, the industrial worker will encounter both bipolar and unipolar transistors. It is important to know their schematic symbols and polarities, the way they are controlled, and whether they are normally on or normally off devices.

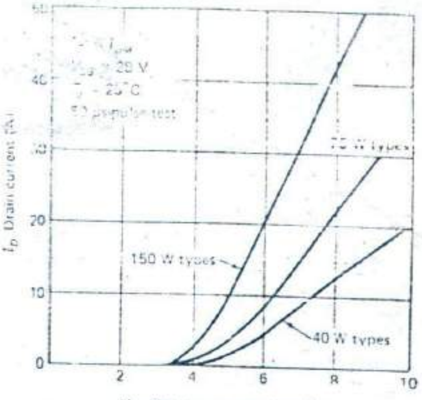


Fig. 2-42 Typical transfer characteristics for N-channel HEXFETs.

Field effect transistors are packaged in many of the same case styles as bipolar transistors. The TO-92 package is popular for low-power devices and the TO-220 and TO-3 packages are often used for high-power devices.

REVIEW QUESTIONS

19. Refer to Fig. 2-33. Does this device normally operate in the depletion mode or in the enhancement mode?

20. The pinch-off region in a FET is also known as the constant _______ region.

21. The P-channel FETs are electrical _______ to N-channel devices.

22. Does the arrow on the schematic symbol point in or out for P-channel FETs?

23. Metallic oxide semiconductor field effect transistors use metallic oxide to _______ the gate from the channel.

24. When the schematic symbol for a MOSFET uses a broken line to represent the drain-to-source channel, it operates in the _______ mode.

25. Enhancement mode transistors are normally _______ devices.

2-5
THYRISTORS

Both the bipolar and unipolar devices studied in the previous sections have a linear mode of operation. This section deals with a family of devices that are not capable of linear operation. Thyristors are strictly switching devices based on a PNPN structure. They are bistable (on or off) and use internal regenerative feedback. They include two, three, and four terminal devices and devices capable of unidirectional (dc) and bidirectional (ac) operation.

The *silicon controlled rectifier* (SCR) is the oldest and most popular thyristor. It is an extremely reliable device and can be expected to deliver billions of operations before failure. It has current ratings that range from 0.25 to several thousand amperes rms. The SCR voltage ratings range up to 5000 V. It is possible to operate SCRs in parallel for even higher current capacity and in series for greater voltage capacity. As a control device, it is one of the most impressive available with microwatt pulses having the ability to switch hundreds of watts. This translates to power gains on the order of 10 million times. Silicon controlled rectifiers can turn on in about 1 microsecond (μs) and turn off in about 10 to 20 μs. They represent an economical solution to many industrial control problems and are especially suited to switching applications up to several kilohertz.

The SCR is sometimes referred to as a *reverse blocking triode thyristor.* Figure 2-43(*a*) shows its structure. Since it is a triode, it has three external connections: an anode (A), a cathode (K), and a gate (G). Figure 2-43(*b*) shows that its PNPN structure can be viewed as a two-transistor structure with two connections between the transistors. Figure 2-43(*c*) shows the equivalent PNP and NPN transistor circuit. This view of the SCR will help you understand its operation. Bipolar transistors are normally off devices. They must be supplied some base current to turn on. Therefore, an SCR is also a normally off device. Notice that the collector lead of the NPN transistor in Fig. 2-43(*c*) is the base current path for the PNP transistor. Also notice that the collector of the PNP transistor is the base path for the NPN transistor. This is the way the regenerative action is achieved. Once the SCR is on, each transistor will hold the other one on. This regenerative latching action means that an SCR must be turned off by some external circuit action. Figure 2-43(*d*) shows the schematic symbol.

Silicon controlled rectifiers can be turned on in five different ways:

1. By avalanche: When the anode is made much more positive than the cathode, forward breakover occurs and latches the device on.
2. By rate of change: If the forward bias voltage across the device increases very quickly, a current will flow to charge the collector-base capacitance of the PNP transistor. This charging current represents base current for the NPN transistor and turns it on.
3. By high temperature: Reverse-biased silicon junctions show a leakage current that approximately doubles for every 8°C temperature rise. At some temperature, the leakage current will reach a level that latches the SCR on.
4. By transistor action: This is the normal mode of operation for all but light-sensitive thyristors. An external gate pulse or signal is used to switch on the NPN transistor and latch the SCR.
5. By light energy: Light entering the junction area will release electron-hole pairs and latch the SCR on. Light-sensitive devices are covered in more detail in the chapter on optoelectronics.

Figure 2-44 shows the volt-ampere characteristic curves for the silicon controlled rectifier. The reverse blocking region is shown in quadrant three of the graph and is similar to that of a silicon rectifier. Normally, the maximum reverse voltage is never exceeded since reverse avalanche could destroy the SCR. It should be emphasized at this point that SCRs are unilateral devices: they normally conduct in one direction only. Forward avalanche is shown in quadrant one. It occurs at zero gate current and at some high value of forward bias. Notice that turn-on occurs at lower values of forward bias as the gate current increases above zero. Current I_{G2} in Fig. 2-44 represents the greatest gate current and ensures turn-on at even a low value of forward bias. This is the normal turn-on mode for the SCR. Once the SCR is on, it drives each of its transistor components into saturation, and its internal resistance drops to a low value. The graph shows that the high conduction

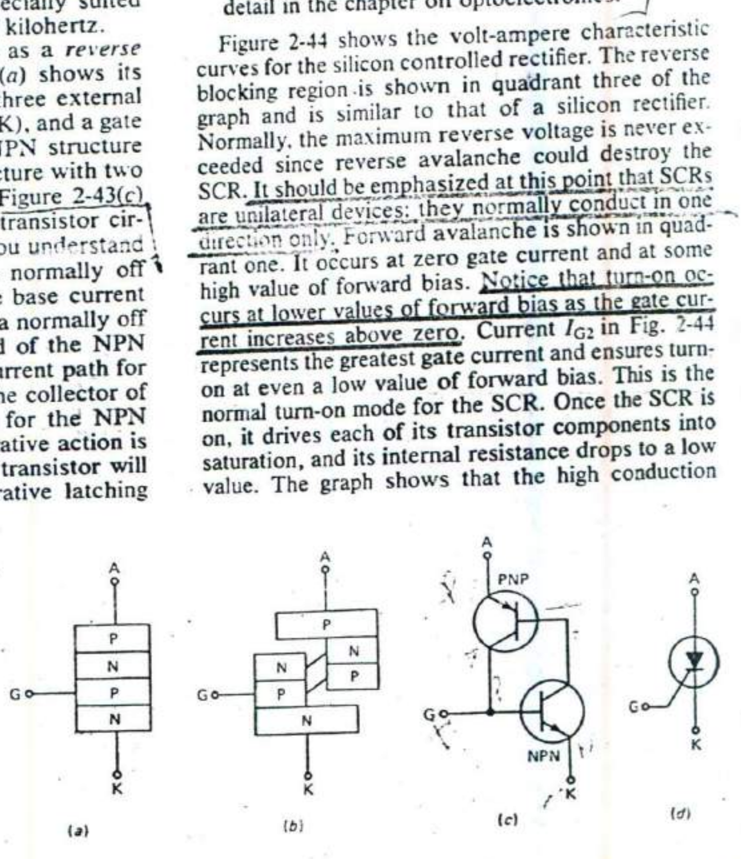


Fig. 2-43 Silicon controlled rectifier. (*a*) Basic structure. (*b*) Viewed as a two-transistor structure. (*c*) Equivalent two-transistor circuit. (*d*) Schematic symbol.

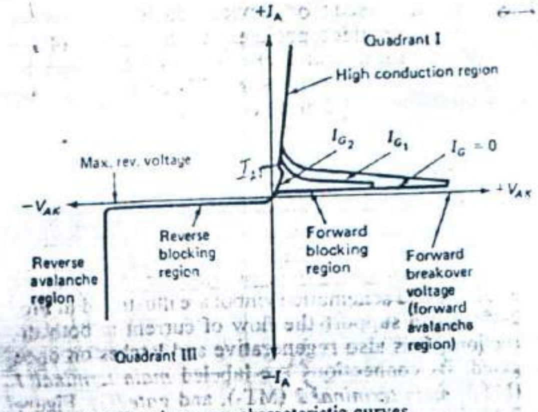


Fig. 2-44 SCR volt-ampere characteristic curves.

region is associated with a low forward voltage drop. This makes the SCR efficient in switching operations.

Why SCR is efficient in switching operation

Figure 2-45 shows a test circuit that demonstrates the latching ability of the SCR. The external battery is arranged to forward bias the SCR. It is not large enough to avalanche the SCR, so there is no current flow. When S_2 is pressed, the gate circuit is completed, and the base of the NPN transistor is now forward-biased. The NPN transistor turns on and supplies current to the PNP base circuit so it also turns on. When S_2 is released, the current continues to flow through the load and the SCR. The SCR has latched on since each transistor is now supplying the base current for the other, and S_1 must now be opened to remove forward bias and turn off the SCR. This is an important concept. Silicon controlled rectifiers can be gated on, but they cannot be gated off. Gate turn-off devices (GTOs) have been developed but are expensive and not very popular. When SCRs are used to control dc power, additional circuitry is required to achieve turn-off. Turn-off is often called *commutation* and is achieved by momentarily zero biasing or reverse biasing the device. For example, one popular commutation circuit uses a second SCR to switch in a charged capacitor across the first SCR to reverse bias it.

[illegible] SCR [illegible] Time

When an SCR is conducting, each junction is forward-biased, and both base regions are heavily saturated with holes and electrons. The saturation is most pronounced at high current levels. To turn an SCR off in a minimum amount of time, a reverse bias must be applied across its anode and cathode terminals. This bias will diffuse the holes and electrons to the end junctions. A reverse current flows while this is happening. After the carriers have been removed, the reverse current will cease, and the two outer junctions will assume a blocking state. Recovery is not complete, however, until the center junction is cleared of its carriers by recombination. If forward voltage is reapplied before the center junction is cleared, the SCR will gate itself back on. The time that elapses between the cessation of reverse current and the point at which forward voltage can safely be reapplied is the turn-off time. It ranges from several microseconds to several hundred microseconds, depending on device design.

In ac power control, commutation is automatic. Figure 2-46 shows the way that an SCR can be used

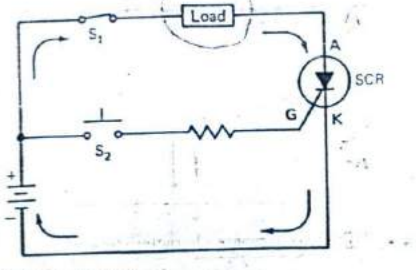


Fig. 2-45 SCR test circuit.

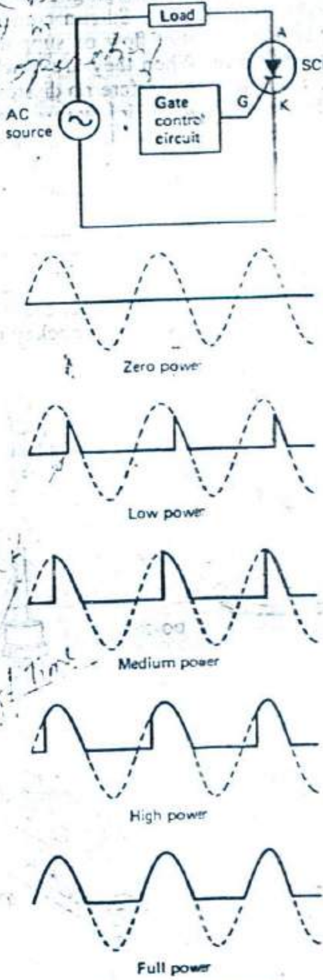


Fig. 2-46 Using an SCR to control ac power.

to control a load efficiently. The waveforms represent load current. The SCR is in series with the load. A gate control circuit is used to pulse the gate lead at the desired moment of turn-on. Turn-off is automatic when the ac line passes through zero. If the SCR is never gated on, the load power is zero. If the SCR is gated on late in the positive alternation, the load power is low. Load power is controlled by conduction angle. With a large conduction angle the circuit is on much of the time and the power will be high. However, since the SCR is a unilateral device only half of the waveform can be utilized. It is possible to achieve full waveform control by using two SCRs connected in inverse parallel or by using rectifier circuits in conjunction with the SCR control. Rectifier circuits are covered in the power sources chapter.

Circuits such as the one shown in Fig. 2-46 are very widely applied in industry. They are commonly used to control motor speed, output from light and heat sources, and battery charging. They are popular because they are efficient. Silicon controlled rectifiers either block current flow or support it with a very low resistance. When they are blocking, there is no current flow and therefore no dissipation in the SCR. When they are on, their low resistance ensures a low voltage drop, and the dissipation in the SCR is small. This means that almost all of the energy is expended in the load and very little in the control device.

Some of the packages used to house silicon controlled rectifiers are shown in Fig. 2-47. As with transistors, the larger devices are mounted in larger packages and are capable of greater dissipation. Notice that the DO-200, or so-called hockey puck package, can be used for devices rated as high as 3000 A. The molded packages at the bottom of the illustration show that more than one SCR can be mounted in a single package. These packages are convenient when bilateral control is needed.

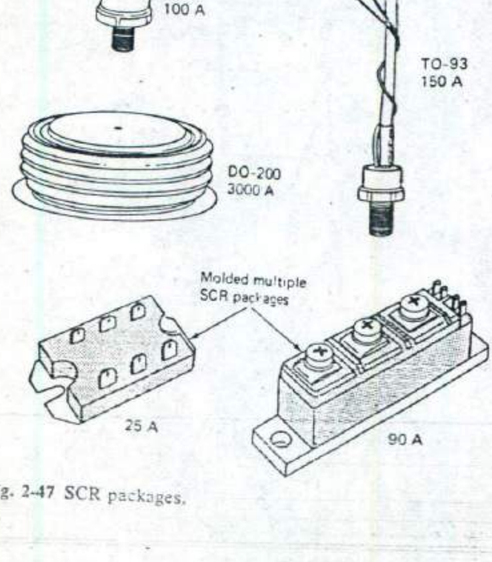


Fig. 2-47 SCR packages.

Triac

Bilateral control may also be achieved with another type of thyristor that is equivalent to two SCRs in a single crystalline structure. It is known as a *triode ac semiconductor switch*, or *triac* for short and its structure and schematic symbol are illustrated in Fig. 2-48. It can support the flow of current in both directions. It is also regenerative and latches on once gated. Its connections are labeled *main terminal 1* (MT_1), *main terminal 2* (MT_2), and *gate* (G). Figure 2-49 shows the triac's characteristic curves. It is a symmetrical device and capable of the same performance in quadrant three of the graph as it is in quadrant one. Current I_H, called the *holding current*, is the minimum value of current required to hold the triac on (keep it latched). SCRs also have some minimum holding current value but only in one quadrant of operation. The graph of Fig. 2-49 does not show it, but the forward and reverse turn-on voltages can

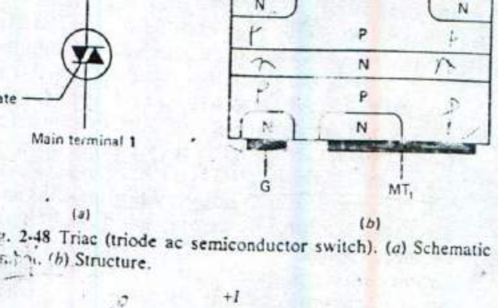


Fig. 2-48 Triac (triode ac semiconductor switch). (*a*) Schematic symbol. (*b*) Structure.

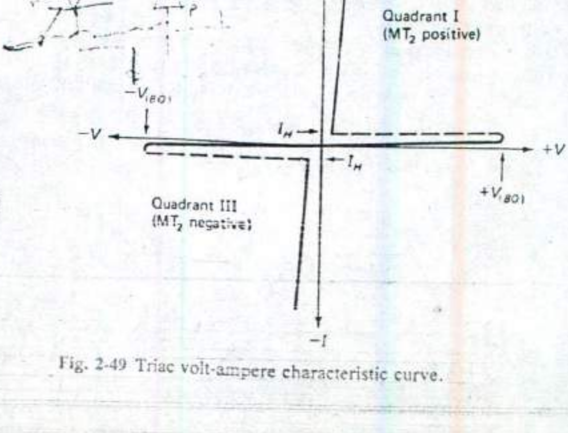


Fig. 2-49 Triac volt-ampere characteristic curve.

be greatly reduced with gate current. As with the SCR, a triac is normally gated on and not operated at breakover.

Figure 2-50 shows the advantage of a triac in ac power control. The circuit uses the main terminals connected in series with the source and the load. A gate control circuit supplies pulses to gate the triac on. The waveforms represent load current and show that conduction angle controls the load power. Compare this illustration with Fig. 2-46. It should be clear that a single triac provides full wave control.

Triacs are very appropriate for some applications but suffer several limitations for others. They are not available with current ratings beyond about 50 A and voltage ratings above 600 V. Also, they have less time to turn off than a pair of SCRs would have in a full waveform control circuit. Each SCR would have an entire half cycle to achieve turn-off if necessary.

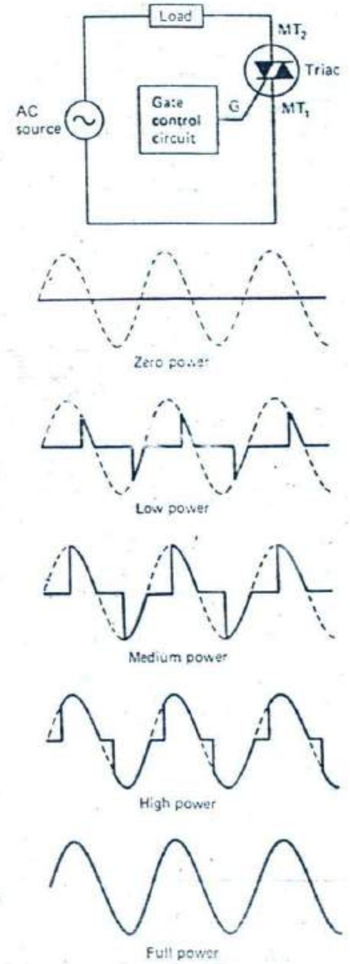


Fig. 2-50 Using a triac to control ac power.

On the other hand, a triac must turn off during the brief moment when the line passes through zero. If the load is inductive (as in a motor), turn-off can be difficult in a triac control circuit. The current lags the line voltage in an inductive circuit. Commutation is attempted when the load current drops below the holding current value. Unfortunately, because of the phase shift, there is a bias voltage across the triac at this time, and the triac recovery current acts as a gate current and tries to turn it back on. The recovery current is due to the recombination of holes and electrons as the device attempts to reestablish its depletion regions and enter the blocking mode. This complicates circuit design. For these reasons, the SCR is still the most popular thyristor for industrial control systems.

Radio Frequency Interference (RFI)

All thyristors cause radio frequency interference (RFI). They can cause a sudden rise in line current at the moment they are gated on. Waveforms with sudden changes in current are rich in harmonic energy. Harmonics are integer multiples of some fundamental frequency. They can extend into the radio frequency spectrum and cause interference with radio receivers, television receivers, logic circuits, and other sensitive equipment. For this reason, thyristor control circuits must include filters to prevent RFI.

REVIEW QUESTIONS

26. Silicon controlled rectifiers are normally turned on by applying a brief pulse to their _______ terminal.

27. Refer to Fig. 2-44. Which value of gate current would be used to ensure device turn-on?

28. Refer to Fig. 2-45. The SCR is gated on by closing switch _______.

29. Refer to Fig. 2-45. The SCR is turned off by opening switch _______.

30. Commutation in an SCR circuit refers to some method of turning the device _______.

31. Commutation is automatic when the power source is _______.

32. Refer to Fig. 2-46. As the SCR is gated on earlier, the power dissipated in the load _______.

2-6
TRIGGER DEVICES

Thyristors can be triggered (gated on) by simple divider circuits consisting of resistors or capacitors across the ac line. One of the divider components can be made adjustable to produce earlier or later firing to accomplish the conduction angle control discussed in the last section. Such simple divider circuits are seldom used, however. They have the disadvantage of temperature instability. Thyristors fire at lower gate currents as their temperature increases. Another problem is that the gate characteristics of the thyristors vary from device to device, even though the part numbers are identical and all have been made by the same manufacturer. Negative resistance devices are generally used to establish a more predictable and stable trigger behavior.

Unijunction Transistor

The *unijunction transistor* (UJT) exhibits negative resistance and is a popular trigger device (Fig. 2-51[*a*]). It is made from a bar of lightly doped N-type silicon with a heavily doped P zone alloyed into the bar. The P zone forms the emitter section of the transistor and its only junction, hence the name *unijunction*. The B_1 and B_2 (base 1 and base 2) contacts at the ends of the bar are ohmic (no diode action). In the equivalent circuit (Fig. 2-51[*b*]) the emitter diode is connected at the junction of two resistors. The bottom resistor is variable. This is where the negative resistance effect occurs. It is called a *negative resistance* because it decreases abruptly when the UJT emitter diode becomes forward-biased with respect to B_1. Complementary UJTs are also available but are less popular. They use a P-type bar and an N-type emitter. The emitter arrow is reversed on the schematic symbol for the complementary UJT.

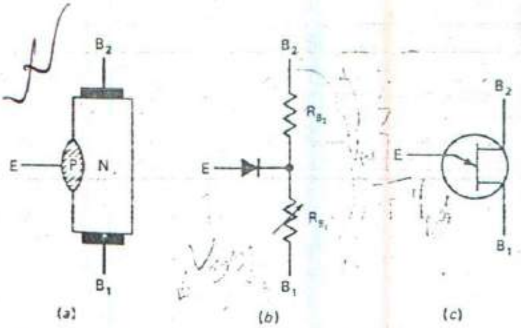


Fig. 2-51 Unijunction transistor (UJT). (*a*) Structure. (*b*) Equivalent circuit. (*c*) Schematic symbol.

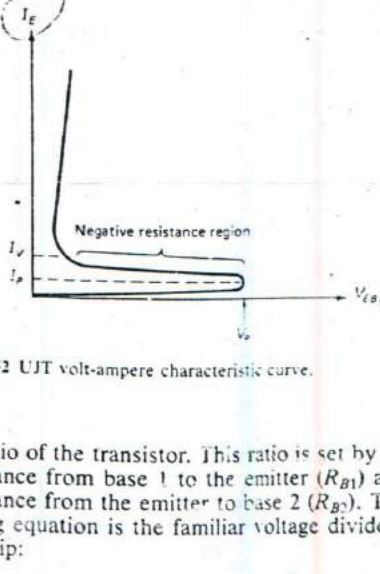


Fig. 2-52 UJT volt-ampere characteristic curve.

The volt-ampere characteristic curve of Fig. 2-52 shows the negative resistance behavior of the UJT. As the forward bias is gradually increased across the emitter to base 1 section, a point, called V_P, is reached; at it the diode becomes forward-biased, and the transistor fires. It then enters its negative resistance region, and the voltage quickly drops to a much lower value. One normally expects the voltage drop to increase as current increases. However, the UJT curve shows a region where voltage decreases as the current increases. This is due to a sudden resistance drop inside the transistor. The curve shows that the emitter current at the firing point is called I_P. It also shows a higher current, called the *valley current* or I_V. This valley current is similar to the holding current in a thyristor. The UJT cannot assume any stable operating point between I_P and I_V. It switches rapidly between the two. This characteristic makes the UJT useful as a trigger device.

The negative resistance behavior of the UJT can be understood by referring to Fig. 2-51. When the emitter diode becomes forward-biased with respect to base 1, it injects minority carriers into the region between the emitter and the base 1 contact. These minority carriers decrease the resistance of the lower part of the bar. This is why the equivalent circuit shows the base 1 resistor as a variable component. When this resistance decreases, the diode forward bias increases, increasing the diode current, and still more minority carriers are injected into the lower part of the bar. This regenerative action rapidly decreases the resistance from the emitter to base 1.

The firing voltage of a UJT is predicted by the voltage across the base leads and the intrinsic standoff ratio of the transistor. This ratio is set by the off resistance from base 1 to the emitter (R_{B1}) and the resistance from the emitter to base 2 (R_{B2}). The following equation is the familiar voltage divider relationship:

$$\text{Intrinsic standoff ratio} = \frac{R_{B1}}{R_{B1} + R_{B2}} = \eta$$

When a supply voltage (V_{BB}) is impressed across the base leads of a UJT, the intrinsic standoff ratio determines the voltage at the cathode of the emitter diode. Since the diode is silicon, an additional 0.6 V is required to turn the diode on. Therefore, the firing voltage, or V_P, is given by

$$V_P = \eta \times V_{BB} + 0.6$$

The intrinsic standoff ratio of UJTs ranges from 0.5 to 0.8

EXAMPLE

The supply is 12 V and the intrinsic standoff ratio is 0.6. Find the firing voltage.

SOLUTION

Use the formula

$$\begin{aligned} V_P &= \eta \times V_{BB} + 0.6 \\ V_P &= 0.6 \times 12 + 0.6 \\ &= 7.8\ \text{V} \end{aligned}$$

W

UJT RELAXATION OSCILLATOR

Figure 2-53(*a*) shows a UJT relaxation oscillator circuit that is useful in many industrial timing and control applications. An *oscillator* is a circuit that changes dc to ac. A relaxation oscillator is one type that uses RC time constants to control the frequency of oscillations. When the supply voltage is applied, the capacitor begins charging through R_1. Eventually, the capacitor voltage reaches the firing point of the UJT. The emitter diode turns on, the internal resistance of the transistor from base 1 to the emitter drops, and the capacitor is rapidly discharged through R_3 (it is usually less than 100 Ω) and the transistor. When the capacitor discharge current reaches I_v, the UJT switches off and the next cycle begins. The waveforms in Fig. 2-53(*b*) and Fig. 2-53(*c*) show an exponential sawtooth at the emitter terminal and a pulse waveform at the B_1 terminal. The period of the waveforms is approximately equal to the R_1C time constant. Since a capacitor reaches 63 percent of its final charge during the first time constant, the approximation is good when the intrinsic standoff ratio is near 0.63.

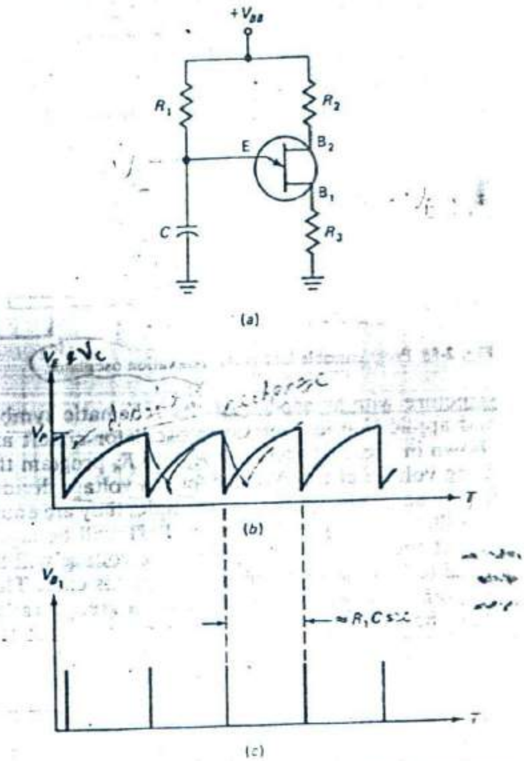


Fig. 2-53 UJT relaxation oscillator and waveforms: (*a*) Circuit (*b*) Sawtooth waveform at the emitter *E*. (*c*) Pulse waveform at base B_1.

EXAMPLE

Suppose R_1 is 100 kΩ and C is 0.1 μF. Find the frequency of oscillation.

SOLUTION

$$\begin{aligned}\text{Period} &= R \times C\\ &= T\\ &= 100 \times 10^3 \times 0.1 \times 10^{-6}\\ T &= 0.01 \text{ s}\end{aligned}$$

The frequency (f) of oscillation is found by the reciprocal of the period:

$$\begin{aligned}f &= \frac{1}{T} = \frac{1}{0.01}\\ &= 100 \text{ Hz}\end{aligned}$$

PHASE CONTROL CKT:

Figure 2-54 indicates one way that a UJT can be used to control an SCR, which, in turn, controls the power delivered to its load. This circuit controls load power by conduction angle. The sooner the SCR is gated on, the larger the conduction angle and the greater the load power. Circuits of this type are called *phase control circuits* since the phase angle of the gating pulse in relation to the source phase determines conduction angle. The zener diode clips the positive peaks of the source at its breakdown voltage. The negative alternations are clipped near 0 V since they forward bias the zener. You should recall from the previous section that the negative alternations are not used by an SCR control circuit of this type, and therefore there is no need to energize the UJT circuit during negative alternations. The breakdown voltage of the zener is reached early during the positive alternation, and the voltage across the UJT circuit is constant for nearly the entire alternation. The capacitor charges through R_2 until the firing voltage of the UJT is reached. When it fires, it develops a pulse across R_4, which gates the SCR on. The phase of the gating pulse can be advanced or retarded by decreasing or increasing R_2. If the period of the UJT oscillator is small, extra gating pulses may be delivered during the positive alternation. These will not cause any effect since the first pulse will gate the SCR on and any subsequent pulses will be ignored.

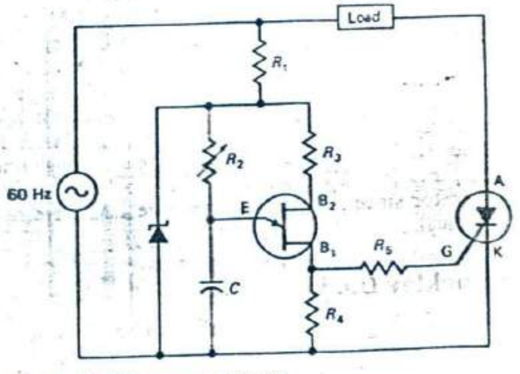


Fig. 2-54 UJT phase control circuit.

Programmable UJT PUT

The *programmable unijunction transistor* (PUT) is an improved trigger device. It is a small thyristor

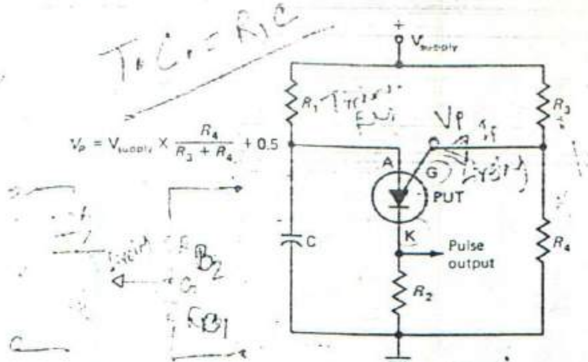


Fig. 2-55 Programmable UJT (PUT) relaxation oscillator.

structure with an anode gate. Its schematic symbol and application in a relaxation oscillator circuit are shown in Fig. 2-55. Resistors R_3 and R_4 program the firing voltage of the PUT and form a voltage divider for the supply voltage. If, for example, they are equal in value, the gate voltage of the PUT will be equal to half the supply voltage. The firing voltage will be equal to half the supply plus 0.5 V in this case. This programming feature gives the PUT a greater range than the UJT and allows the designer to select the desired V_P. It also eliminates the problem with "batch spread" since UJTs of the same manufacturing batch will show a considerable spread in their intrinsic standoff ratios. The PUT can be programmed with precision resistors to eliminate this problem. The peak current I_P is also a function of the programming resistors. By using large resistors, it is possible to obtain a very low peak current in a PUT timing circuit. This makes it possible to increase the size of the timing resistor (R_1 in Fig. 2-55) to a much greater value than it can have in a UJT circuit. The UJT current's being larger is a disadvantage with long time constants since the current will cause a drop across the timing resistor, and the capacitor will never reach the firing voltage if the timing resistor is made too large. The only way to achieve long time constants in a UJT circuit is to use a large timing capacitor since timing resistors above 1 MΩ are not practical.

Shockley Diode

Figure 2-56 shows the schematic symbol for a four-layer diode (also called the *Shockley diode*). It is essentially a miniature SCR with no gate lead. It exhibits negative resistance once its firing voltage is reached. Its volt-ampere characteristic curve is the same as for an SCR, except that there is no possibility of varying the breakover point with gate current. Four-layer diodes are available with breakover ratings from about 10 V to 400 V. They usually exhibit peak currents of about 100 μA and holding currents of 1 mA or so. They are used in some thyristor gating circuits.

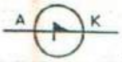


Fig. 2-56 Four-layer diode schematic symbol.

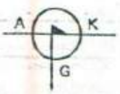


Fig. 2-57 Silicon unilateral switch schematic symbol.

Silicon Unilateral Switch

A *silicon unilateral switch* (SUS) is shown in Fig. 2-57. It is equivalent to a miniature anode gate SCR with a built-in zener diode from its gate to its cathode. This trigger device is more flexible than the four-layer diode because its gate terminal can be used to alter its forward breakover voltage. As all other trigger devices do, it exhibits a negative resistance characteristic. The SUS can be fired at low anode-to-cathode potentials. The major difference between SUS and UJT operations is that the SUS switches at a voltage determined by its internal zener, and the UJT fires at some fraction of its supply voltage. The SUS can also be synchronized or locked out by applying pulse signals or a bias to its gate terminal.

Diac

All the trigger devices investigated to this point are unilateral. They fire in one direction only. As such, they are more appropriately applied in SCR control circuits than they are in triac control circuits. They can be used to gate triacs by adding rectifiers or pulse transformers, but we will investigate bidirectional trigger devices which are well suited to triac control.

The diac (Fig. 2-58) is a transistor-type structure with a bidirectional negative resistance characteristic. Diac breakover current is typically around 100 μA and occurs at approximately 30 V. A diac phase

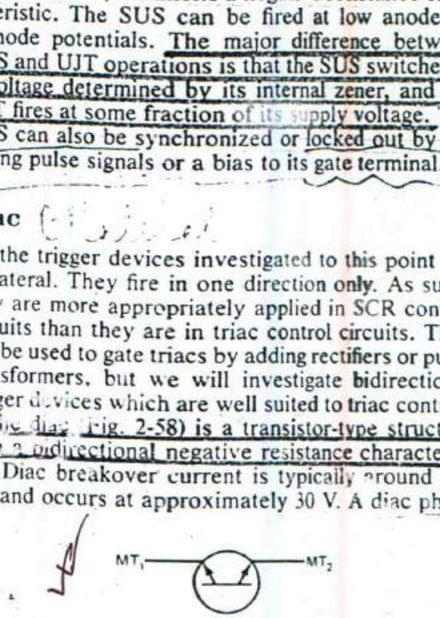


Fig. 2-58 Diac. (a) Schematic symbol. (b) Characteristic curve.

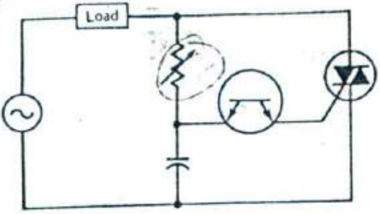


Fig. 2-59 Diac phase control circuit.

control circuit is shown in Fig. 2-59. On either alternation, the capacitor will begin charging through the variable resistor. When the capacitor voltage reaches the breakover potential, the diac fires and gates the triac on. The phase angle can be advanced by decreasing the variable resistor. This shortens the time constant, and the triac gates sooner for a larger conduction angle and greater load power. Increasing the variable resistor will delay firing for a smaller conduction angle and lower load power.

Silicon Bilateral Switch

Another full-wave trigger device, the *silicon bilateral switch* (SBS) is shown in Fig. 2-60. The SBS is equivalent to two SUS devices connected in inverse parallel. It typically exhibits a breakover of around 8 V and has a more pronounced negative resistance region than the diac. It is also more temperature-stable than the diac. The gate lead provides additional capabilities and can be used to alter the breakover characteristics. The gate lead is also useful for eliminating the hysteresis effect, or "snap-on" effect, found in many triac control circuits. This effect is noticed when the phase angle control is slowly advanced from the zero power position. The load "snaps" to some intermediate power level. Then the phase control can be backed off for low-power operation. Any circuit that exhibits a different threshold when a control is moved in one direction than it does when the control is moved in another direction is said to have *hysteresis*. Hysteresis can be very desirable in some applications but is often undesirable in phase control circuits. It is just not possible to smoothly adjust the load for low power when starting from zero power if the circuit has hysteresis.

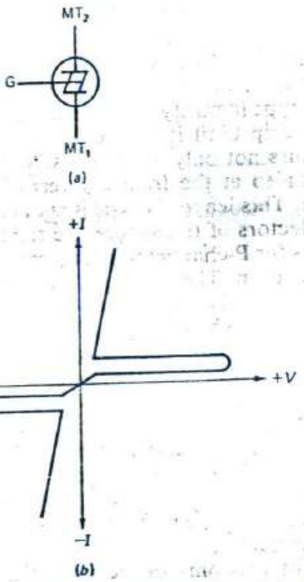


Fig. 2-60 Silicon bilateral switch. (*a*) Schematic symbol. (*b*) Characteristic curve.

The hysteresis effect can be understood by referring to Fig. 2-59. At zero power, the capacitor has been somewhat charged by the prior alternation at the beginning of any positive or negative alternation. This charge is a reverse charge as far as the current alternation is concerned. It takes time to reverse the charge on the capacitor. This time delays the firing of the diac. However, once the diac does fire, it gates on the triac, which in turn tends to drain any residual charge from the capacitor. Therefore, on all subsequent alternations, the reverse charge is absent, and the capacitor charges more quickly to the firing voltage of the diac. The phase angle is now advanced, and the load has snapped to some intermediate power level. It is now necessary to increase the variable resistor to back the power level down to the desired condition.

Asymmetrical ac Trigger

Some phase control circuits take advantage of the gate terminal of the silicon bilateral switch to overcome the hysteresis, but an asymmetrical ac trigger device can also be used (Fig. 2-61). This device presents a forward breakover voltage different from its reverse breakover voltage. It acts as a zener diode in series with an SBS. Note that the schematic symbol includes a zener curve to show this effect. The typical asymmetrical trigger device switches at 8 V in one direction and at 16 V in the other. It will fire first at its lower breakover voltage when being adjusted from zero power. Then it will fire at its higher breakover voltage on the next alternation. This delays firing and tends to offset the tendency for the circuit to snap to an intermediate power level. The asymmetrical trigger is a simple solution to snap-on, but its inherent asymmetry develops a dc component in the load circuit which may not be acceptable in some applications.

Fig. 2-61 Asymetrical ac trigger schematic symbol.

REVIEW QUESTIONS

33. A unijunction transistor has an intrinsic standoff ratio of 0.75. What will its firing voltage be with a 15-V supply?

34. Refer to Fig. 2-53. Assume an intrinsic standoff ratio of 0.63 and calculate the period of oscillation if the timing resistor is 470 kΩ, and the timing capacitor is 0.01 μF.

35. Calculate the frequency of oscillation for question 34.

36. Refer to Fig. 2-54. What happens to the load power when R_2 is adjusted for lower resistance?

37. Refer to Fig. 2-54. What would happen to the load if the period of the UJT oscillator were made longer than 8.33 ms? (Hint: Calculate the period of the positive alternation.)

38. Refer to Fig. 2-55. If the supply is 10 V, R_3 is 220 kΩ, and R_4 is 47 kΩ, what voltage will the capacitor charge to?

2-7
INTEGRATED CIRCUITS

The same planar process that is used to manufacture most of the discrete devices covered in this chapter is also used to manufacture most integrated circuits. The typical integrated circuit uses components that are all formed at the same time and are not individually accessible. This arrangement contrasts to that of a *discrete circuit*, in which individual resistors, diodes, capacitors, transistors, and other components are interconnected to form a working circuit. Discrete circuits require assembly and some system of interconnection such as a printed circuit board. This means that discrete circuits are larger and more costly than an equivalent integrated circuit (IC). The ICs are also usually more reliable and more power-efficient than equivalent discrete circuits. Therefore, designers choose ICs wherever their use is feasible. Many industrial systems are heavily dependent on ICs. They do suffer power-handling limitations, so it is common to find equipment based on both integrated and discrete circuits. The ICs handle most or all of the low-level signal processing, and the discrete circuits use large power devices to control the high-level signals.

The ICs may be classified according to the technology used to manufacture them. Figure 2-62 shows a typical monolithic (*monolithic* means "single stone") IC mounted in an 8 pin package. The *chip* is the monolithic silicon structure that provides the electrical and electronic functions. Some people use the word *chip* when referring to a complete IC package. The silicon chip in Fig. 2-62 is bonded to a plastic or ceramic base. Wire bonds form electrical connections from the silicon chip to the pins. A plastic or ceramic cap completes the assembly. The pins are identified by number. Count counterclockwise when viewing the package from the top. Pin 1 is the first one encountered after the index. The numbers in Fig. 2-62 represent pin numbers.

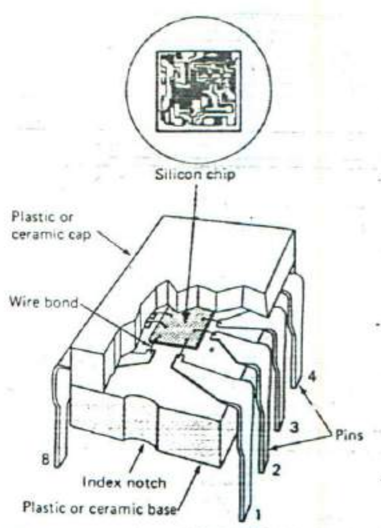


Fig. 2-62 Typical integrated circuit construction.

Figure 2-63 shows an abbreviated cross-sectional view of the chip. Now you can understand why it is considered a monolithic structure. Each component is an area buried in a single silicon structure and has been formed by the planar diffusion process. Note that P-type isolating wells are used to isolate one component function electrically from the next. For example, the NPN bipolar transistor at the left is diode-isolated from the P-channel resistor to its right. In normal operation, the PN isolation diodes are all reverse-biased, and electrical component integrity is maintained. The wells are produced during a manufacturing step called *isolation diffusion*, during which a P-type impurity such as boron is forced to penetrate the chip until it reaches the P-type substrate. This occurs not only at the sides, as shown in Fig. 2-63, but also at the front and rear of every component site. This leaves N-type islands that will become the collectors of transistors, the cathodes of diodes, the sites for P-channel resistors, one plate of a capacitor, and so on. The junction diode shown in the illustration may also be used as a capacitor if it is reverse-biased. Both NPN and PNP transistors are feasible, as are the various unipolar transistors discussed earlier in this chapter. Inductors are not feasible. After all of the components have been formed, evaporated aluminum is deposited onto the surface. The aluminum contacts selected areas of the chip through windows in the silicon dioxide layer. Then, the unwanted aluminum is etched away, leaving aluminum jumpers that interconnect the individual components to form a complete circuit.

The economy of the monolithic IC is now apparent. All of the needed components for a circuit are

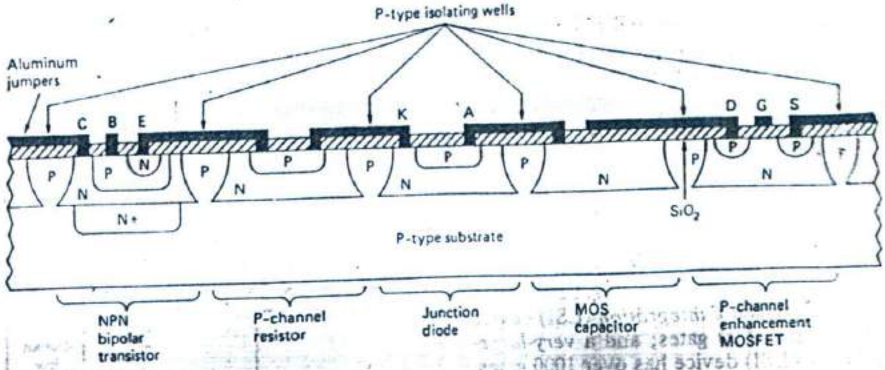


Fig. 2-63 Cross section of a monolithic IC chip.

formed at the same time by the planar diffusion process. Then aluminum interconnects these components to form circuits. Another economical feature is that monolithic ICs are batch-processed. Hundreds of them are processed at the same time on a silicon wafer that is typically several inches in diameter. The completed wafer, including aluminum jumpers, is about 400 micrometers (μm) [0.016 in.] thick. The wafer is scribed, and the individual chips are broken from it. Some complex chips contain thousands of individual components and are only 6250 μm (0.250 in.) by 6250 μm! This miracle of modern technology has allowed room-sized equipment to shrink to desktop size. The price has also decreased dramatically, with some computer-type circuits now costing tens or hundreds of dollars that formerly cost hundreds of thousands of dollars. It is easy to see why monolithic ICs have created a rapid expansion of electronics into every industrial sector. Very sophisticated control systems are now feasible, and they are relatively inexpensive, small, efficient, and reliable.

Most, but not all, ICs are monolithic. Another manufacturing process can combine monolithic ICs with larger component structures to permit larger signals to be controlled. Figure 2-64 shows the structure of a hybrid IC. It can be seen that several types of components are fixed to an insulating substrate of ceramic or glass. Hybrids are either of the thin-film or thick-film variety. A thin-film IC uses very thin films (about 0.3 μm) that are vacuum-deposited on a substrate. Resistors are usually formed by depositing tin oxide, nichrome, or tantalum strips; conductors are made by depositing gold nichrome. The other components, including one or more monolithic ICs, are in chip form and are fastened to the substrate with conductive epoxy. Thick-film ICs use screen-process printing to deposit resistive and conductive patterns onto the substrate. These patterns are much thicker than the vacuum-deposited ones. Hybrid ICs of both types offer several advantages over monolithic ICs. They can take advantage of separate power devices and therefore can handle larger signals (up to several hundred watts). This makes them more flexible and provides a greater range of applications. They have a greater range of available capacitor and resistor values, and precision resistor values can be attained by trimming critical patterns with computer-controlled laser beams. Unfortunately, their higher cost makes circuit designers look to monolithics first, especially for high-volume applications.

Integrated circuits are also identified as being digital or linear. A *digital IC* works with only two circuit conditions: on or off. Chapter 11 deals with digital circuits. A linear IC works with an infinite number of possibilities. For example, the voltage in a linear control circuit might be 8.5 V, or 8.56 V, or 8.567 V, and so on. Several popular ICs are both; that is, they contain both linear and digital circuits.

Monolithic ICs are often further differentiated on the basis of the type of transistor that they use. Bipolar ICs favor NPN transistors since they are easier to fabricate in the chip and have higher performance than integrated PNP transistors. The MOS ICs use metallic oxide semiconductor field effect transistors. The PMOS ICs are based on P-channel transistors, and the NMOS ICs are based on N-channel devices. Also, some manufacturers "invent" terms such as *HMOS* to describe their particular innovative process. The term *HMOS* is used by Motorola, Inc., to identify their high-density NMOS devices. The complementary metallic oxide semiconductor (CMOS) ICs use both N- and P-channel transistors. Some ICs use both bipolar and unipolar transistors and are called *BI-FET* ICs. The field effect

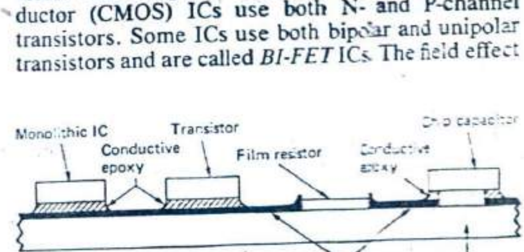


Fig. 2-64 Hybrid integrated circuit construction.

transistors are used in the integrated circuit's input circuits because of their advantages of high impedance, low noise, low leakage currents, and good temperature stability.

Finally, ICs may be identified according to their level of complexity. For example, in the digital world the number of logic gates is a way to categorize the complexity of an IC. A *logic gate* is a single decision-making element, with each gate potentially containing a dozen or so components. A *small-scale integration* (SSI) device will have up to 10 gates; a *medium-scale integration* (MSI) device has between 11 and 100 gates; a *large-scale integration* (LSI) device has between 101 and 1000 gates; and a *very-large-scale integration* (VLSI) device has over 1000 gates. Digital ICs are also identified as *transistor-transistor logic* (TTL), *emitter-coupled logic* (ECL), etc., integrated circuits. These designations will be explained in Chapter 11.

There are thousands and thousands of active IC part numbers, with new ones announced every month. This may be bewildering to a beginner. However, the industrial technician seldom needs to be concerned with the exact circuit located inside a particular IC. You will learn that some very common features which are important to the technician are pretty much the same for many ICs. Along these lines, let's take a look at a few very common IC characteristics. Figure 2-65 shows a totem-pole output stage that is very popular in digital ICs. The transistors form the totem-pole, which is capable of being driven high or low. When Q_1 is off there is no base current path for Q_3, and it is also off. When R_1 provides base current to Q_2 it is on. Therefore, if an external load is connected as shown, the IC will source current to the load. The typical TTL digital IC can source a few milliamperes before its output voltage drops too far below the threshold voltage.

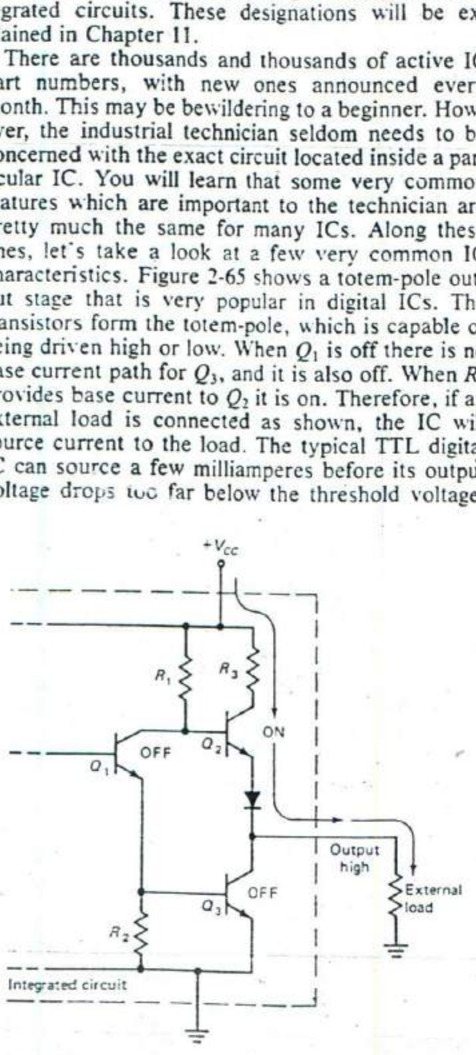


Fig. 2-65 Current source mode for totem pole IC.

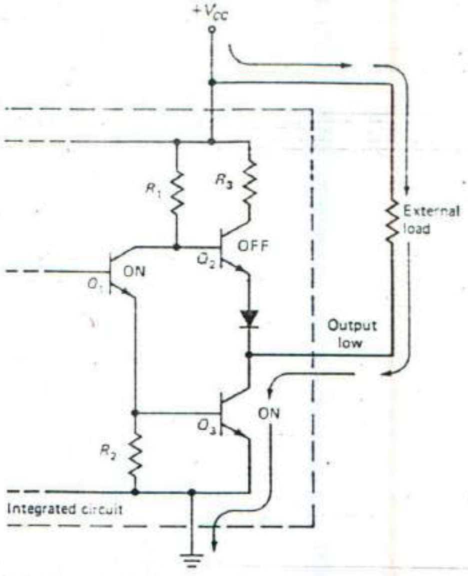


Fig. 2-66 Current sink mode for totem pole IC.

Now look at Fig. 2-66. Here, the totem-pole output is low; Q_1 is saturated on, and the drop across R_1 is large. The base voltage at Q_2 is too low for it to turn on, and Q_3 now has plenty of base current and is on. Note that the external load is now connected to the supply point, and the IC is said to be sinking load current. The typical TTL digital IC can sink several times more current than it can source. Compare Fig. 2-65 and Fig. 2-66 to clarify the difference between current sourcing and current sinking. Please note that either a high or a low output can produce load current, depending on the way the external load is connected. The totem-pole output stage can be redesigned with CMOS output transistors to allow both devices to be off at the same time. Integrated circuits with this feature are called *tri-state devices*. Their outputs can be high, low, or off (high-impedance).

Figure 2-67 shows another popular output stage. It is known as an *open collector IC* because there is no internal connection to the collector of the output transistor. Open collector devices offer the advantage of allowing the output collector circuit to operate at a different voltage level than the rest of the IC circuitry. This is convenient when logic voltages have to be translated from one level to another. Another advantage is that open collector outputs may be tied together without the danger of excessive current in the output circuit. Tying outputs together is usually avoided with totem-pole ICs since a high output would source excessive current to another output that was low. Note that an external pull-up resistor is required to develop an output swing in Fig. 2-67.

Figure 2-68 shows some package styles for ICs. The *dual-inline package* (DIP) is very popular for both digital and linear devices and is made with 14,

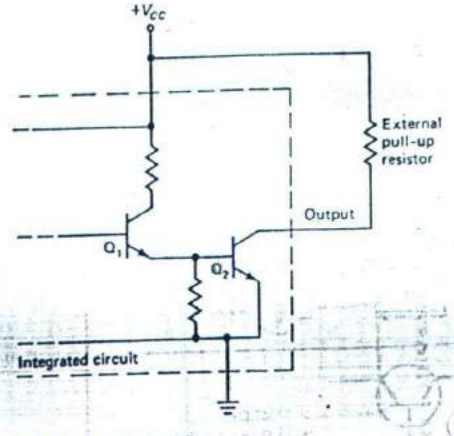


Fig. 2-67 Open collector IC.

16, 18, or 20 pins. The mini-DIP is an 8-pin dual-inline package. The LSI/VLSI package is available with 24, 28, 40, and, less commonly, 64 pins. It is often used for complex ICs such as microprocessors, memory devices, and programmable devices. It is wider than the DIP. The flat pack is used only in compact, low-profile applications and is less popular. The metal packages offer some advantages for heat dissipation and are popular for power applications such as amplifiers and voltage regulators. Hybrid ICs may be housed in any of the packages shown in the left column of Fig. 2-68 or in a wide variety of custom packages.

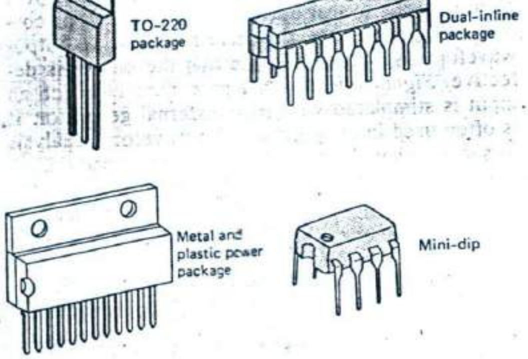


Fig. 2-68 Integrated circuit package styles.

REVIEW QUESTIONS

39. Integrated circuit pins are numbered by counting counterclockwise from the index when viewing the package from the ______.

40. In a monolithic IC, electrical integrity for each component is maintained by reverse-biased ______.

41. Refer to Fig. 2-65. Would the IC source any current if the external load resistor were connected to the supply rather than to ground?

2-8

TROUBLESHOOTING AND MAINTENANCE

Solid-state devices are usually very reliable. Some industrial equipment will operate for years without a single solid-state device failure. However, all of the devices covered in this chapter have two major enemies: heat and transients. In some cases, failures may occur frequently. The industrial technician must become proficient at locating defective devices and replacing them properly. A knowledge of circuit operation is usually mandatory when troubleshooting. This knowledge allows the fault or faults to be isolated. A piece of equipment may contain hundreds or even thousands of electronic components. Obviously, fault isolation is essential if the equipment is to be repaired in a reasonable length of time. Later chapters of this book explain the operation of many circuits and present specific troubleshooting information. This section will be limited to some general ideas concerning locating defective devices and replacing them.

Troubleshooting must always begin by verifying that the equipment is set up and properly connected. In some cases, there is nothing wrong with the equipment itself. A cable may have been pulled out, or a cable may be connected improperly. A control may be set wrong. Always check the obvious things first; it can save a tremendous amount of time.

When it is verified that the equipment is defective, refer to the manufacturer's service literature for the proper tear-down procedure. The literature may also contain specific troubleshooting procedures. Be certain that the power is off before removing panels, covers, or cabinets. Never pull or insert circuit boards with the power on. This can be a danger to you and to the equipment. Once access to the circuitry has been gained, it is time for a thorough visual inspection. Solid-state devices usually do not change in physical appearance when defective, but other components sometimes do. Be sure to inspect fans, air filters, and other cooling components since heat is a killer of electronic equipment.

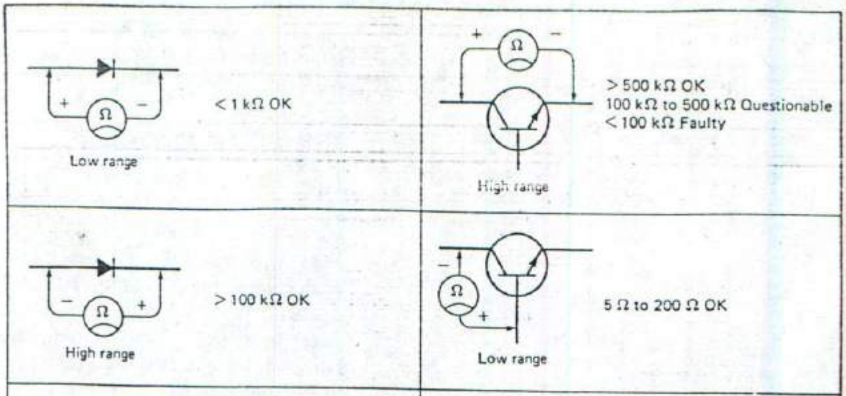


Fig. 2-69 Testing solid-state devices with an ohmmeter.

After any obvious problems have been corrected, it is time to decide how to isolate any remaining faults. Troubleshooting techniques vary from circuit to circuit and from technician to technician. They also depend upon the kinds of test equipment that are available. Whatever you choose, please remember to follow safe procedures. Some of the procedures that can be used to isolate defective solid-state devices include substitution, voltage analysis, waveform analysis, signal injection, and resistance analysis. Substitution is convenient when devices are in sockets. However, remember that a fault somewhere else in the circuit may have damaged the original part and is also likely to damage the substitute. Never change devices with the power on.

Voltage analysis involves comparing actual circuit voltages with voltages specified on a diagram by the equipment manufacturer to determine whether any are out of tolerance. Always verify supply voltages first. Power supply troubleshooting is covered in Chapter 5. A defective device may or may not upset circuit voltages. Waveform analysis uses an oscilloscope to view signals into and out of circuits. If a circuit has the proper operating voltages and the correct input waveform, then a bad or missing output waveform is strong evidence that the circuit is defective. *Signal injection* is a procedure in which an input is stimulated with some external generator. It is often used in conjunction with waveform analysis to see whether the output responds as expected.

Resistance analysis is accomplished with the power off and often with the suspected device isolated from the rest of the circuit. Isolation may be necessary to prevent unwanted paths from producing abnormally low readings. Figure 2-69 shows the way that an ohmmeter can be used to test several devices. Verify your ohmmeter polarity and open circuit voltage. Some multimeters have reverse polarity on the Ohm's function, and others do not use an internal voltage large enough to forward bias junctions on some ranges. Always learn your equipment first. Ohmmeter tests are effective for detecting short-circuited junctions since a very low resistance will be measured in both directions. A very high reading in both directions indicates an open junction. The diode tests in Fig. 2-69 show the expected results for a good diode. It must be emphasized that the readings in this illustration are relative. For example, a diode that measures 200,000 Ω when reverse-biased may have excessive leakage for some circuits. Generally, more leakage is normal in large devices than in small devices. Leakage is also increased when the device is hot. Germanium devices exhibit quite a bit more leakage than silicon devices. You should also know that the forward resistance of any junction will measure differently on different ohmmeter ranges because the volt-ampere characteristic of a diode is nonlinear.

The transistor tests in Fig. 2-69 show the expected ranges of resistance for a silicon transistor. You will note that the collector-to-emitter reverse resistance is not expected to be as high as the collector-to-base reverse resistance. This is because the collector-base leakage is amplified beta times in the collector-emitter test. Your understanding of device theory will make ohmmeter testing more productive. Figure 2-69 also shows that the emitter-base reverse resistance may not test properly because the ohmmeter voltage may be high enough to cause zener breakdown. Most transistor emitter-base junctions show zener action at around 6 V, and many ohmmeters use 9 V or more on their high range. As you gain experience, you will learn the way to test quite a few solid-state devices with an ohmmeter. It is even possible to gate and latch some thyristors by placing the positive lead on the anode and the negative lead on the cathode, and then momentarily short-circuiting the gate to the anode. This will not work with large thyristors since the ohmmeter current, even on the lowest range, is less than the holding current. Field effect transistors are often very sensitive during out-of-circuit tests. The drain-to-source resistance may vary quite a bit as you bring a finger near the gate lead. Metallic oxide semiconductor field effect transistors are susceptible to damage by static discharges and should not be checked this way. Integrated circuits do not usually lend themselves to ohmmeter testing.

Repeated failure of the same device must be investigated. It is probably a thermal problem, a transient problem, a defective power supply, a defective load, or a design flaw. Check to be sure that all cooling components are installed and working properly. Inspect the mounting area on the heat sink. Burrs and peeled-over areas will reduce heat transfer. There should be no paint or heavy oxide on the mounting area. Galvanic action may occur in corrosive atmospheres between copper cases and aluminum heat sinks. Be sure that the correct thermal grease is applied. Everything must be installed correctly to prevent short circuits and excessive device temperature. The insulator may be of the beryllium oxide type. Be careful: Beryllium oxide parts must not be abraded or crushed because the dust is extremely dangerous if inhaled. Do not replace beryllium oxide washers with mica washers, since the heat transfer will be impaired. Do not over- or under-tighten stud mount devices. For example, the proper torques for DO-4 and DO-5 packages (Fig. 2-19) are 15 inch-pounds (in.·lb) and 30 in.·lb, respectively.

Repeated failures may also be caused by the load. Make sure the load is electrically normal. If it is a motor or solenoid, make sure that it is clean and properly lubricated and is operating freely. Also be sure that the mechanical load on the solenoid or motor is working smoothly and not binding. *Transients* are brief periods of overvoltage and may enter the equipment via the power lines. The industrial environment is often replete with transients since large inductive loads such as motors are constantly being turned on and off. It may be necessary to add transient suppression to the supply circuit. Transients are also generated when solid-state devices switch inductive loads within the equipment circuit. Check the suppression networks and devices across relay coils, solenoids, and motors. Suppression is covered in Chapter 4. Repeated failures due to design flaws can be checked by contacting the manufacturer. The engineering staff may supply a circuit modification or a substitute part with better ratings.

Parts Identification

Exact replacement parts are usually the best bet, especially for the technician with limited experience. Solid-state devices often have part numbers on their packages. The manufacturer's literature is also usually helpful for locating part numbers. The Joint Electronic Device Engineering Council (JEDEC) registers part numbers in this country. A registered part has been characterized to meet the specifications listed for that number. Registered solid-state parts (excluding ICs) have numbers prefixed with 1N, 2N, 3N, and 4N. This means that you can buy a 1N5000 rectifier, a 2N690 SCR, a 2N3055 bipolar transistor, a 3N128 field effect transistor, or a 4N32 opto coupler from any of several manufacturers and be reasonably assured that it will work as well as the original device. There are also registered JEDEC package numbers such as TO-3 that specify the physical parameters of devices.

Data manuals, cross-reference guides, and substitution guides are invaluable aids when trying to track down part numbers. These materials contain valuable information concerning physical characteristics, electrical characteristics, and lead identification drawings. You will find many good substitutions in these sources. In some cases the substitutions are not appropriate. It pays to check both electrical and physical parameters. Some companies build quite a few solid-state devices with their own part numbering system. These nonregistered devices can often be substituted for registered devices. An example is a Motorola MJ4502 power transistor, which can be substituted for a 2N5744. The guides usually list both registered and nonregistered device numbers. Some part numbers are proprietary and will not show up in the guides. It may be necessary to buy a replacement from the manufacturer of the equipment in these cases.

Integrated circuit part numbers are also referenced in some guides. The ICs have a part numbering system that can vary considerably from manufacturer to manufacturer. It pays to have a library of data manuals from the various companies that build solid-state devices. Supply catalogs are also helpful in many cases and can often be obtained just by asking for them. It pays to communicate with supply houses and parts jobbers to obtain valuable literature. Integrated circuit part numbers are usually a combination of a prefix, a part number, and a suffix. The *prefix* uses code letters to designate the type of circuit, the *part number* specifies the device type, and the *suffix* code specifies the package type and the temperature range. There are many variations of this basic system, and the manufacturer's data books are usually necessary to decipher all the information contained in the part number. Many IC makers also put date codes on their packages.

Handling Solid-State Devices

The final consideration in this chapter is the safe handling of solid-state devices. Many devices, especially the MOS types, are easily damaged by static discharge. It is a little disconcerting to realize that merely touching an expensive or hard-to-get device can destroy it. The human body can generate thousands of volts through simple movements such as walking, sliding in a chair, or sliding a sleeve across a bench top. These voltages are particularly high in low-humidity conditions. Some workers must wear a conductive wrist strap that is grounded through a high-value resistor to bleed off static charges. Note: This is only practiced in an approved environment and with an approved grounding apparatus. The following guidelines are recommended to prevent static damage of solid-state devices:

1. Work on a metal surface. Plastic laminated table tops are a poor choice for a work surface since static build-up is likely.
2. Do not allow the relative humidity in the work area to go below 50 percent.
3. Do not handle devices any more than is required. They are shipped in protective carriers or pressed into conductive foam and should remain there until it is time to install them.
4. Immediately place removed parts into a protective carrier or conductive foam.
5. Touch the protective package to ground before removing the part.
6. Touch a grounded part of the equipment before removing or installing a part.
7. Use as little motion as possible. Remember, friction generates static electricity.
8. When instruments are connected to circuits, always connect the ground lead first.
9. Use only antistatic spray materials and static-controlled vacuum desoldering equipment.

REVIEW QUESTIONS

42. When testing with an ohmmeter, a reading of 0 Ω in both directions indicates a(n) ______ junction.

43. When testing with an ohmmeter, a reading of infinity ohms in both directions indicates a(n) ______ junction.

44. The number 1N4001 is an example of a(n) ______ part number.

CHAPTER REVIEW QUESTIONS

2-1. Electrons, in a P-type crystal, are considered ______ carriers.

2-2. Calculate the maximum reverse current for a 12-V, 10-W zener.

2-3. Would the zener in question 2-2 be safe when conducting maximum current at a temperature above 25°C?

2-4. Refer to Fig. 2-26 and calculate h_{FE} when V_{CE} is 100 V and the base current is 0.6 mA.

2-5. Calculate emitter current for the conditions of question 2-4.

2-6. Calculate the collector dissipation for the conditions of question 2-4. Is it within the safe thermal operating area?

2-7. What happens to transistor gain as temperature increases?

2-8. When a bipolar power transistor fails at a higher collector voltage and is within the safe thermal area, the failure mode is called ______.

2-9. How do the pulse-mode ratings of transistors compare with their dc ratings?

2-10. Suppose the thermal rating for a transistor is 200 W. Use Fig. 2-29 to determine the maximum thermal dissipation for an operating temperature of 80°C.

2-11. What are the three operating modes for a transistor?

2-12. A perfect switching transistor would show a collector-to-emitter drop of ______V at saturation.

2-13. Power FETs are normally ______ devices.

2-14. Refer to the 40-W transfer curve in Fig.

2-42. What is the change in drain current when the gate-to-source bias increases from 7 to 8 V?

2-15. What is the transconductance for the data given in question 2-14?

2-16. When a bipolar transistor heats, its gain increases, tending to make it conduct more and become even hotter. This is known as ______.

2-17. Refer to Fig. 2-49. What designates the minimum flow to keep the triac on?

2-18. Triac commutation is complicated by ______ loads since the internal recovery current can act to gate the device back on.

2-19. Refer to Fig. 2-59. What happens to the load power as the variable resistance is increased?

2-20. The circuit of Fig. 2-59 cannot achieve smooth control when being adjusted from a ______ power setting.

2-21. The name given to describe the effect of question 2-20 is ______.

2-22. The asymmetrical ac trigger device is designed to eliminate ______.

2-23. Refer to Fig. 2-66. Would the IC sink any current if the external load resistor were connected to ground rather than to the supply?

2-24. Refer to Fig. 2-67 and assume that the external resistor has been removed. With a 5-V supply, what output voltage swing will be developed as Q_1 is turned on and off?

2-25. The number MJ802 is an example of a(n) ______ part number.

2-26. Leakage currents are expected to be ______ in power devices.

ANSWERS TO REVIEW QUESTIONS

1. pure **2.** it is lower **3.** negative **4.** positive **5.** P-type **6.** minority **7.** reverse **8.** minority **9.** reverse **10.** increases **11.** rectifier **12.** minority **13.** emitter, out **14.** controls **15.** reverse **16.** forward **17.** increases from 5 to 85 mA **18.** 50 mA **19.** depletion **20.** current **21.** complements **22.** out **23.** insulate **24.** enhancement **25.** off **26.** gate **27.** I_{G2} **28.** 2 **29.** 1 **30.** off **31.** ac **32.** increases **33.** 11.85 V **34.** 4.7 ms **35.** 213 Hz **36.** it increases **37.** it would be off **38.** 2.26 V **39.** top **40.** diodes or junctions **41.** no **42.** short-circuited **43.** open **44.** registered

3

INTRODUCTION TO MOTOR CONTROLS

This chapter will introduce direct current (dc) and alternating current (ac) motor controls. Of primary interest will be methods of starting and stopping motors and controlling their speed. Stepper motors, used where automatic or computer control is important, are covered, as is the brushless motor. Alternating current motors are very popular in both small and large sizes in a wide range of industrial applications. With the rapid introduction of solid-state control devices they are taking over a growing number of functions previously reserved for dc motors.

3-1

BRAKING DC MOTORS

There are several factors that must be considered in stopping a motor. When a motor is disconnected from the power source it will coast to a stop. Not all machines can be allowed to coast to a stop. When it is necessary to stop a motor quickly, it is accomplished by braking. Braking can be accomplished by several methods, each of which has advantages and disadvantages. There are three basic means of slowing down a motor: friction, dynamic action, and plugging.

Applications of braking vary greatly. For instance, a crane or hoist not only has to stop quickly but also must hold heavy loads. Other motors, such as those controlling machine tools, must stop quickly but do not have to hold a load. Safety braking may be required to protect an operator from injury. In this case, the fastest method is employed with little consideration to the potential damage to the motor or load. When dealing with equipment such as cranes and hoists, we must realize that the load has a tendency to turn the motor. This is known as an *overhauling load*. *Overhung loads* are those loads which are applied in a direction perpendicular to the axis of the shaft. These types of loads are created by supported weights such as those in hoists and elevators.

Friction brakes (also known as *magnetic* or *mechanical brakes*) have been used to stop motors for many years. Their application is similar to the braking of an automobile. The essential parts of the brake are the friction material, shoes, bands or disks, and operating devices. Most friction brakes are electrically released and spring-set, so they will be set in case of a power failure or interruption. Occasionally it is advantageous to have a brake that is electrically set. With this type of brake, it is possible to vary the applied torque by adjusting the voltage to the brake coil.

The *shoe brake* needs only a small movement to release it. For example, a solenoid may have to travel less than 0.10 inches (in.). This short stroke gives fast operation. A drum or wheel is driven by the motor shaft to provide a larger braking surface than that provided by using the motor's shaft alone. Braking torque is directly proportional to the surface area and the spring pressure. Spring pressure is adjustable on most friction brakes.

The *disk brake* is arranged for mounting directly to the motor end bell. The brake lining is a disk, which is supported by a hub keyed to the motor's shaft and rotates with the motor. When the brake is set, a spring pulls the stationary member into contact with the rotating disk. The simplicity of a single moving part, the *pressure plate*, makes for less mechanical maintenance. No linkages or levers are used, as they are in the shoe type.

The *band-type* brake has the friction material fastened to a band of steel which encircles the wheel and may cover as much as 90 percent of the wheel surface. The increased braking surface allows lower pressure per square inch of surface area, and a subsequent reduction in wear of the brake lining. This advantage is offset somewhat by the fact that the braking pressure is not equal over the whole band. The band brake also requires a longer stroke to release it.

The basic material used in all brake linings is asbestos. When servicing any brake parts, do not create dust by using a dry brush or compressed air; use a water-dampened cloth. The asbestos fibers may become airborne if dust is created during servicing.

Breathing dust containing asbestos fibers can cause serious bodily harm. The main disadvantage of any friction-type brake is that it requires more maintenance than other types. The more frequently a motor is stopped by the brake, the more maintenance will be required. The solenoids used to energize most friction brakes are available for either ac or dc operation.

Dynamic action, or *electrical, braking* is accomplished by changing the connections to a motor with or without the use of an auxiliary power source (depending on the type of motor to be braked). If a motor that is still running is reconnected as a generator, the result is dynamic braking. The motor acts as a loaded generator that develops a retarding torque, which stops the motor rapidly. The generator action converts the mechanical energy into electrical energy and dissipates this energy as heat in a resistor. Since dynamic braking is only present when the motor is rotating, a friction-type brake is required to hold any overhauling load after it stops.

The easiest type of motor to brake dynamically is the *shunt-wound* dc motor. While the machine operates as a motor, the counterelectromotive force (cemf) opposes the line voltage and limits the armature current to a value sufficient to provide the output torque requirements. Dynamic braking is accomplished by disconnecting the armature from the line (power source) and placing a current-limiting resistor across the armature terminals while the field remains energized. The armature will be rotating in the magnetic field and will continue to generate a cemf that is proportional to speed and the strength of the field. The armature current flowing through the limiting resistor will be opposite to that produced by the line. This reverse current in the armature will produce a torque opposite to the original motor action and cause the motor to slow down. As the forward speed is reduced, the generated voltage is also reduced. At the zero speed point, the generated voltage is also zero. The motor will stop at this point if no overhauling torque is available to continue the rotation. The connection for dynamic braking of a shunt-wound motor is shown in Fig. 3-1.

The braking resistor is usually selected so that the initial braking current is 150 percent of the normal current. Braking times of milliseconds are not uncommon on fractional horsepower motors. The field winding may or may not be disconnected from the line after the motor has stopped. When a field rheostat is employed, it is customarily short-circuited to increase the field to aid in the braking effect.

Dynamic braking may be used with a series motor, as shown in Fig. 3-2, but the connections are more complicated. If the motor were just disconnected from the line and shunted by a resistance (such as the shunt motor armature), no braking would be obtained. This is because the current would flow through the field in the wrong direction, thereby demagnetizing the field.

With the motor running and current flowing from L1 to L2, the cemf of the armature is in opposition, as shown by the arrow in Fig. 3-2. The field current is flowing from L1 to L2. For braking to occur, the field must be connected in the reverse direction so that the current flows through it in the same direction it does when the motor is across the line. When the switch is in position 2, the resistor is connected in series with the field and the armature, forming a closed loop. Notice that at this time the field connections are reversed so the current flow due to the armature emf is in the same direction as it is when connected as a motor. The energy is quickly dissipated in the resistor, and the armature stops rotating.

A combination of dynamic braking and friction braking can be used when a very large load must be stopped. Because the force of the load would wear out the friction brakes too quickly, dynamic braking can be used to slow down the load, and then the friction brakes can be applied to finally stop and hold the load if necessary.

Dynamic braking of a permanent magnet (PM) motor is accomplished the same way as it is with the shunt-wound motor: the armature is disconnected, and its terminals are shunted by a resistor. A very distinct difference should be noted: with the shunt-wound motor, the motor cannot be dynamically braked if there is a power failure. The field voltage and current must be available to generate the dynamic braking. With the PM motor, the power failure will not affect its braking ability because its field is not affected by any power failures. A normally

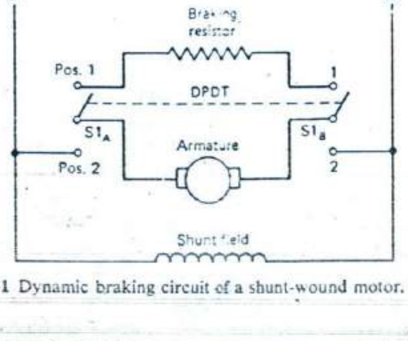


Fig. 3-1 Dynamic braking circuit of a shunt-wound motor.

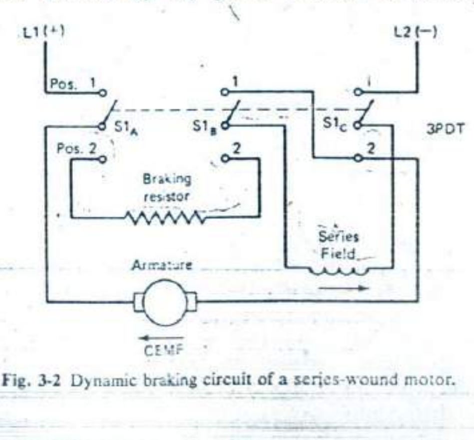


Fig. 3-2 Dynamic braking circuit of a series-wound motor.

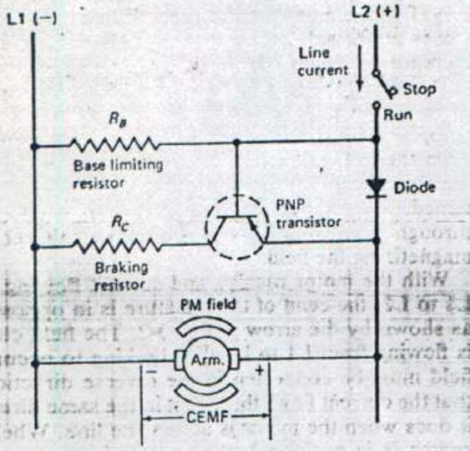


Fig. 3-3 Transistor dynamic braking circuit.

closed relay across the armature will automatically function in case of power failure and load the armature through its braking resistor. This inherent characteristic is very important and useful in many cases. For example it can be used on magnetic tape reel drives to prevent unwanted spillage of tape in the event of a loss of power. Figure 3-3 illustrates the utilization of solid-state electronic components to achieve dynamic braking of a PM motor. The diode drop biases the transistor off when the switch is in the run mode. When the switch is put into the stop position, the armature no longer draws current from the line. This is the *brake mode;* the transistor will conduct because of the polarity of the armature's cemf. The cemf turns on the transistor circuit and R_C acts as the armature load. This circuit will work just as well for a shunt-wound motor, but the field circuit must remain connected to the line until the motor stops.

A *compound-wound motor* is actually both a shunt- and a series-wound motor and can take advantage of either the shunt-braking circuit or the series-dissipating circuit or a combination of the two. However, the slower speed of the compound motor makes shunt-wound braking the preferred method.

Plugging is a way of braking a motor by reversing the power to the armature while the field remains connected as before. Plugging the motor allows for a very rapid and abrupt stop. Plugging can be used to brake a motor if the armature power is removed at the point where the motor speed drops to zero. Otherwise the motor will reverse. Plugging is more severe than the other braking methods mentioned because the voltage across the armature (in the shunt motor) and across the entire motor (in the series motor) is approximately twice its normal value at the instant that braking is initiated. The cemf voltage in the armature is additive to the line voltage (the armature leads have been reversed while the rotation is still in the original direction) until the speed goes to zero. Under normal operating conditions, the cemf generated opposes the line voltage and thereby limits the armature current.

Plugging is not usually recommended because of the high armature currents drawn. Excessive armature heating and brush arcing will result, and brush life may be severely affected. Motors used for plugging are designed for this type of service. In some applications plugging is not permitted, and the motor controls are interlocked to enforce this condition.

Although manual and magnetic starters are used to reverse the direction of a motor, a special plugging switch is typically used in plugging applications. The plugging switch, sometimes called a *zero-speed switch,* is connected to the shaft of the motor or its load through a pulley or a shaft. The rotation of the switch shaft, at a given speed, causes a set of contacts to operate, either by a centrifugal mechanism or by a magnetic induction arrangement. The main function of the plugging switch is to prevent reversal once the countertorque action of plugging has brought the load to a standstill. Without this switch, the motor and load would stop and then run in the opposite direction.

The plugging switch will usually open one set of contacts and close another set of contacts as the shaft speed increases past a preset number of revolutions per minute (rpm). The continuous operating speed of the machine should be many times the speed needed to operate the switch contacts. This will ensure good contact holding force and reduce the possibility of chatter and subsequent false operation of the switch.

Figure 3-4 illustrates the wiring of a motor starter circuit that includes a plugging switch. The normally open (N.O.) contacts of the plugging switch (P) are wired in series with a set of interlock contacts to the reversing starter coil (R). Pushing the start button energizes the forward starter coil (F) and the motor starts forward. The forward coil is held in by the F_1 contacts across the start button now energized (closed). As the motor increases in speed the N.O. contacts on the plugging switch close at some preset rpm. The closing of these N.O. contacts will not energize the reverse coil (R) because of the F_2 con-

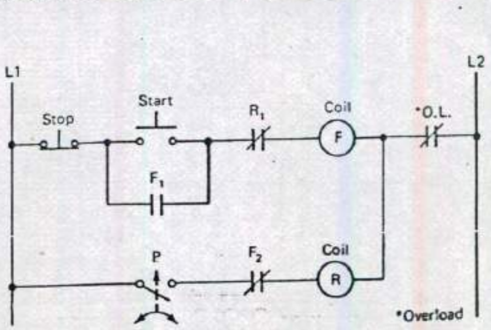


Fig. 3-4 Typical motor plugging circuit.

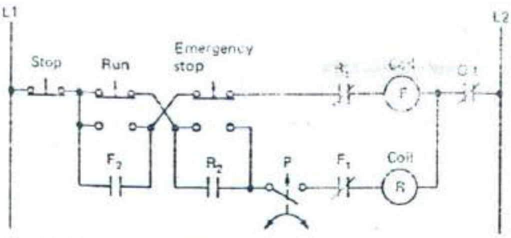


Fig. 3-5 Emergency stop plugging circuit.

tacts, which are now open. Pushing the stop button will drop out the forward coil (F) and cause the F_2 contacts to close, permitting the reversing coil (R) to energize through the still closed plugging switch. The motor connections are now reversed, and the motor's countertorque acts as a braking force. When the motor is stopped the plugging switch opens, disconnecting the reverse coil before the motor can physically reverse its direction.

Plugging is generally used for emergency stopping of a motor. The plugging circuit of Fig. 3-4 would be modified to include the emergency stop circuitry as in Fig. 3-5. The normal stop and run pushbuttons operate as in any standard starter circuit. The important difference is the addition of the emergency stop pushbutton. Pushing the emergency stop pushbutton will deenergize the forward starter (F) and simultaneously energize the reverse starter (R). The reverse starter is held in by the R_2 contacts through the plugging switch (P) and the new closed F_1 contacts of the forward starter coil (F). When the motor speed approaches zero speed, the plugging switch (P) will open to disconnect the starter before the motor reverses direction. There are numerous other types of plugging motors, which use timers, time delay relays, and other special devices.

REVIEW QUESTIONS

1. ______ braking is required to hold an overhauling load.

2. Friction brakes are electrically ______ and spring-set.

3. The ______ type friction brake uses a pressure plate and spring set.

4. The use of ______ in brake linings makes their maintenance a cautious and careful operation.

5. Name two other solid-state devices that could be used instead of the transistor in Fig. 3-3.

6. When plugging a series motor, the voltage across the motor is ______ that of normal operation.

7. Refer to Figs. 3-4 and 3-5. In which figure does the circuit show plugging accomplished at every actuation of the stop buttons?

3-2
SPEED CONTROL OF DC MOTORS

Industrial application of dc motors often requires, in addition to driving loads at a constant speed or torque, the ability to vary the speed of the motor. The speed of most dc motors can be easily adjusted from zero to speeds above the rated value. In addition to good speed control, the dc motor is well suited for applications requiring momentary high-torque outputs. A dc motor can produce torque three to five times its rated value for short durations. Because of their good speed control and high torque, they are used in many industrial applications and also in mining equipment. There are a variety of ways to control the speed of a dc motor, ranging from a simple series rheostat to a modern generation of solid-state devices.

There are three methods by which the speed of a shunt-wound motor can be controlled: field weakening, armature resistance control, and armature voltage control. The speed of a shunt-wound motor can be described by the following equation:

$$\text{rpm} = \frac{V_a - (I_a \times R_a)}{\phi_f}$$

The equation shows that the speed can be made to change by adjusting the variables. V_a, R_a, and ϕ_f (I_a changes proportionally with the load and is not considered a speed-control variable).

If the strength of the magnetic field of the shunt-wound motor is reduced, the motor will speed up. This speed-up occurs with the reduction in the field strength because less cemf is developed by the armature. The difference beween the line voltage and the new cemf produces an increase in armature current, resulting in an increase in output torque and speed. The torque is related to the flux field and armature current by the following equation:

$$T = K \times \phi_f \times I_a$$

The field can be weakened by connecting a rheostat in the field circuit as shown in Fig. 3-6(a). The rpm (speed) equation shows that weakening the field increases the rpm, provided the armature voltage is constant. Standard industrial motors will permit a speed increase of up to 400 percent. This method of speed control is considered efficient, because the power lost in the field rheostat is negligible; however, the field can only be weakened within certain limits. Weakening beyond a limit point can result in excessive speeds and instability. The armature may also overheat. The torque equation shows that a reduction of the field ϕ_f will require an increase in the armature current I_a to maintain a given torque. The torque-speed curves for different values of increasing resistance in the field are shown in Fig. 3-6(b). Field weakening will produce speeds above the normal rated speed. The motor can be overloaded easily because the rated torque drops as the speed in-

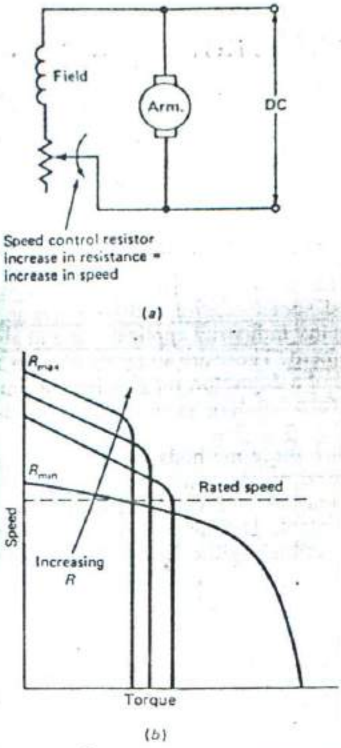


Fig. 3-6 Field weakening speed control of the shunt motor. *(a)* Schematic diagram. *(b)* Torque-speed curves.

creases, as shown in Fig. 3-6(*b*). Therefore, this type of control (field weakening) is limited to cases in which load conditions are very predictable and controlled.

A completely different set of characteristics occurs for a shunt-motor with a rheostat in the armature. In this case of armature resistance control, the field winding is kept at the rated or line voltage. Referring again to the rpm equation, if the armature voltage V_a is reduced (by increasing the resistance), the motor speed will decrease. Therefore, armature resistance control will lower the speed of the shunt-motor below its rated base speed. As indicated by the torque equation, an increase in load will result in an increase in the armature current and a subsequent increase in the voltage drop across the added series rheostat. If the motor is started with very little load and the load is increased, the speed will drop sharply. There will also be a corresponding I^2R (power) loss in the series rheostat.

The resistance methods of field weakening and armature resistance described are very simple and inexpensive and will provide control of a shunt-wound dc motor's speed above or below its rated value. These methods were and are still used in many applications in the industrial environment.

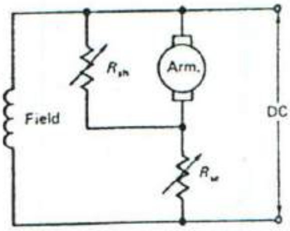


Fig. 3-7 Tandem (series and shunt) armature speed control.

Both series and shunt rheostats may be employed in tandem to improve the speed regulation of the shunt-wound motor by making the operating speed less affected by the changes in load torque. The shunted armature shown in Fig. 3-7 has a variable resistance across (shunt) the armature and acts to increase the current through the series resistance and thereby reduce the difference between no-load and full-load current. The series resistance is used to control the armature voltage, as with the armature control method for the shunt motor.

The resistance methods of speed control just discussed are *open loop* methods (there is no feedback). These methods are still found in many industrial applications. Motors are frequently controlled by varying the supply voltage to the armature. Armature voltage control of shunt-wound motors can be open loop or closed loop (having feedback) to control the speed of the motor. Figure 3-8 shows open loop armature voltage control. This method of motor control has several advantages:

1. Wide speed-range control
2. Speed not appreciably affected by changing load
3. Less power wasted at low speed
4. Ease of interfacing with complex electronic control systems

The nonfeedback controller, shown in Fig. 3-8(*a*), produces the speed/torque curves of Fig. 3-8(*b*). Speed regulation as the motor load varies is essentially the inherent regulation characteristic of the motor as the curves illustrate. The motor will operate along the load line by jumping from one speed/torque curve to another. An infinite number of curves is made available by varying the armature voltage. Armature voltage control may be obtained from one of the simple control circuits shown in Fig. 3-9. These control amplifiers need only a simple reference voltage amplifier and a current (power) amplifier to provide infinite speed control. These circuits are also applicable to PM motors. Though effective, these methods are very inefficient, since substantial power is lost in the controlling device.

Open loop, or nonfeedback, control will be only as stable as the load and the individual components of the system. Figure 3-10 shows silicon controlled rectifier (SCR) motor speed circuits. The field cir-

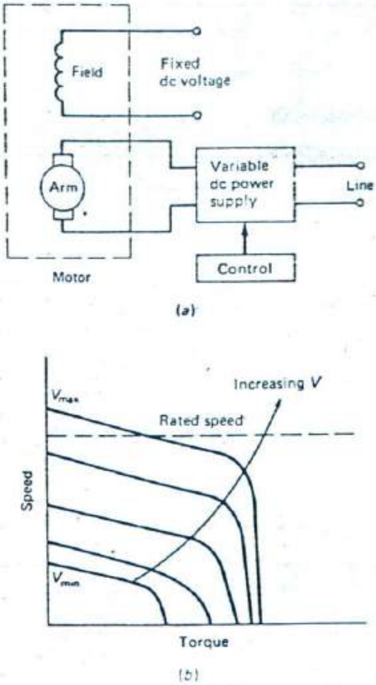


Fig. 3-8 Armature voltage control of speed. *(a)* Schematic diagram. *(b)* Torque-speed curves.

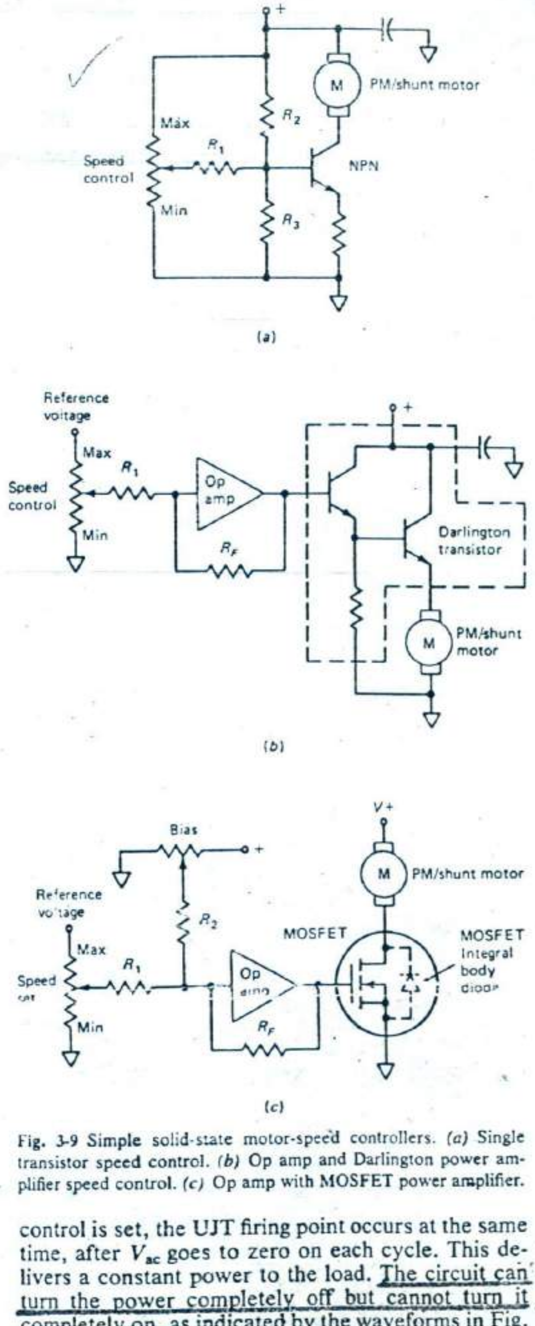
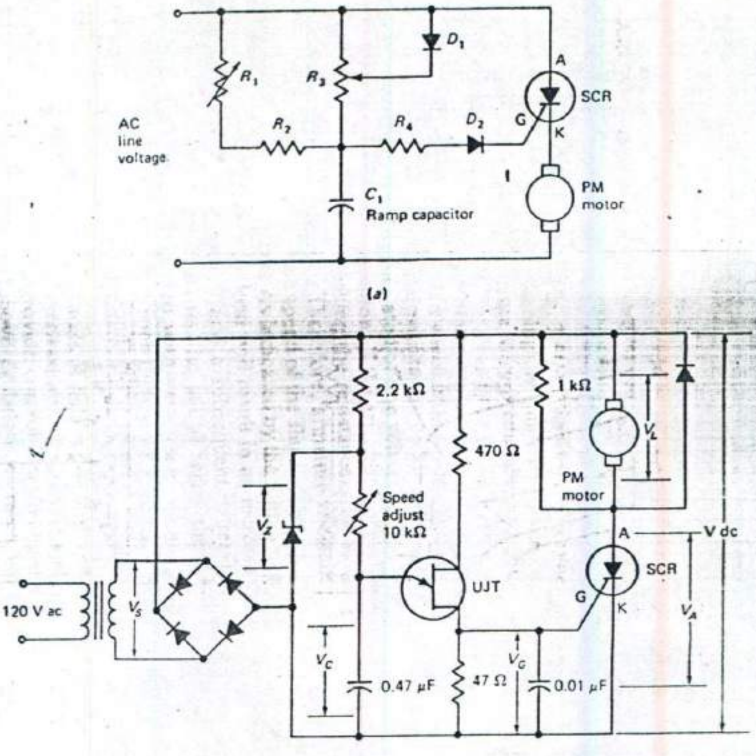


Fig. 3-9 Simple solid-state motor-speed controllers. *(a)* Single transistor speed control. *(b)* Op amp and Darlington power amplifier speed control. *(c)* Op amp with MOSFET power amplifier.

cuits have been omitted for simplicity. The motor may also be a PM motor, as previously mentioned. The circuit in Fig. 3-10(*a*) has a limited speed range but will maintain constant speed under varying load conditions. Though it is not obvious at first glance, this circuit includes a feedback element that enhances its control stability. This type of circuit makes use of the cemf of the armature, which is a function of the motor's speed and, therefore, can be used as an indication of speed changes as the load varies. When a load is applied to the motor, the motor speed will start to decrease and thus reduce the cemf induced in the rotating armature. With this reduced cemf, the SCR will fire earlier in the rectified cycle. The SCR control is less precise than a linear amplifier but offers high efficiency since the SCR is either on or off. The simplicity of the circuitry and low parts count make its use attractive.

The inclusion of the unijunction transistor (UJT) oscillator in Fig. 3-10(*b*) will give a wider range of speed control. It should be noted that in both these circuits, unfiltered direct current is used as the power source. This unfiltered source is necessary in order to commutate the SCR; otherwise it will remain on, and all control will be lost. In Fig. 3-10(*b*) the pulsating direct current also serves to turn off the UJT oscillator, so it starts a new charge cycle on each cycle of the pulsating supply voltage. Once the speed control is set, the UJT firing point occurs at the same time, after V_{ac} goes to zero on each cycle. This delivers a constant power to the load. The circuit can turn the power completely off but cannot turn it completely on, as indicated by the waveforms in Fig. 3-11 (p. 49). Smooth variable control from 5 to 95 percent of available power can be obtained. This method of control is sometimes referred to as a variation of a pulse-width modulation (PWM) speed-control circuit.



(a)

(b)

Fig. 3-10 Motor speed control using SCRs.

With the exception of the circuit in Fig. 3-10(*a*), all the circuits presented thus far have been open loop speed controllers. For applications demanding precise and constant velocity under varying load conditions, a closed loop system is needed. Block diagrams for closed loop controllers are shown in Fig. 3-12 (p. 50). The speed sensor (*tachometer*) generates a voltage proportional to motor speed. This voltage is fed back and compared to a reference voltage. Any error between the two voltages is amplified and corrects the speed of the motor. The extra element in this system that makes it work and closes the feedback loop is the speed sensor (tachometer, in this case). In many cases, a small PM motor is used as a tachometer. Most tachometers develop an output voltage proportional to the shaft speed. The polarity of the output voltage is dependent on the direction of rotation of the shaft. The output voltage from a tachometer may be expressed as

$$V_g = K \times \text{rpm}$$

where V_g is the tachometer output voltage and rpm is the armature shaft speed. A value of 1.0 V/500 rpm is typical. The output increases to 4 V if the tachometer is rotated at 2000 rpm. Alternatives to the generator-type tachometer include inductive pick-ups and optical pick-ups to provide pulses that are integrated (summed) for the speed feedback signal. These pick-ups usually require amplification to be useful.

Figure 3-12(*a*) illustrates a *first quadrant regulation control circuit*, so named because the speed/torque curves are all in the first quadrant of the cartesian coordinate system. That is, they are positive when the shaft is rotating in the forward direction. The motor in Fig. 3-12(*a*) rotates in only one direction (it is *unidirectional*). Its speed will increase or decrease as the comparator turns on or off the input to the amplifier in response to the comparison of the speed (feedback) signal to that of the set point reference. The closed loop system in Fig. 3-12(*b*) is also referred to as a *servosystem*, which can also be used to control the position as well as the velocity of a system, with the addition of a position feedback element. These aspects will be covered in depth in Chapter 10. For now, it can be seen that by proper selection of the references, gains, motor, tachometer, and scaling components, a high-performance speed-control system is obtainable.

The series (universal) motor is designed to operate on either alternating or direct current and is capable of high speeds and starting torques. The speed of the series motor can be changed by changing the voltage across the motor. This is accomplished by three methods: series-resistance control, shunt-resistance control, and variable-voltage control. The

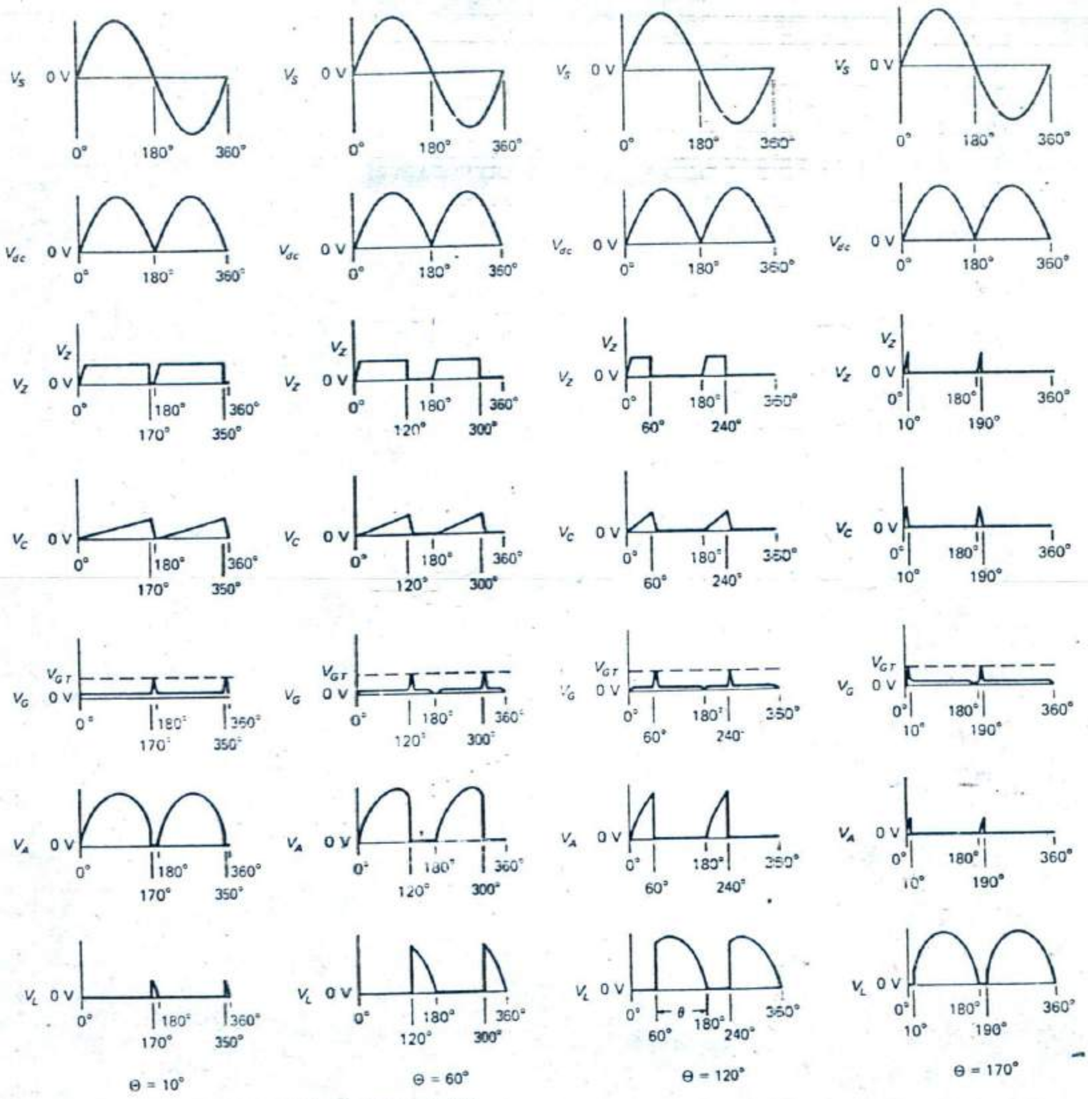


Fig. 3-11 Waveforms for SCR-UJT circuit in Fig. 3-10*b*).

speed/torque curve obtained by using a rheostat in series with the motor (armature and field) is illustrated in Fig. 3-13(*a*). Series-resistance control has good starting characteristics (high torque at low speed), but it is evident that the speed regulation of the motor declines with decreasing speed, thereby making good overall control of the motor's speed difficult. The series resistor produces a voltage drop in the circuit that is proportional to the current in the circuit. This voltage drop across the resistor will increase as the motor is loaded (motor current increases with an increase in load). It can be seen that the voltage across the motor will decrease with the increase in load, and the speed will drop rapidly with an increasing load when a series resistor is used. Also, the higher the resistance, the greater the drop in speed as the load increases. A series resistor or rheostat will have its greatest effect on the starting torque of the motor. Maximum current flows when the motor is started and the resistive drop will limit the motor voltage to its lowest value. The series resistor usually will be adjusted for minimum resistance for starting and then increased as the motor gains speed. In theory, the motor can be adjusted to near standstill (complete stop). However, as a result of reduced inertia, *armature cogging* (the armature speeds up as it enters the flux field and slows down on exiting) is very pronounced at low speed; therefore, the lower limit must be set to a value at which cogging is avoided.

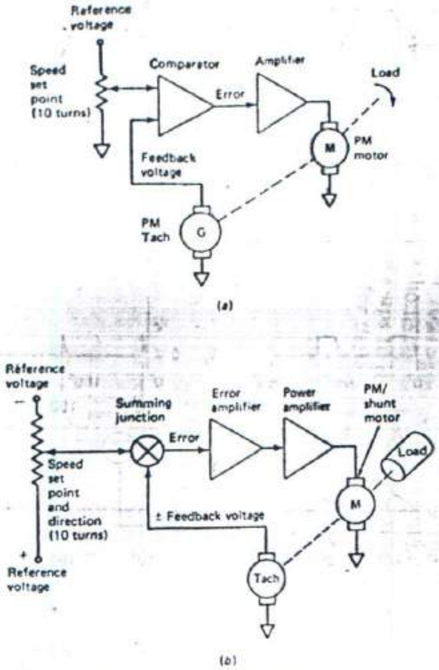


Fig. 3-12 Closed loop speed controllers. *(a)* Unidirectional feedback control. *(b)* Bidirectional feedback speed control.

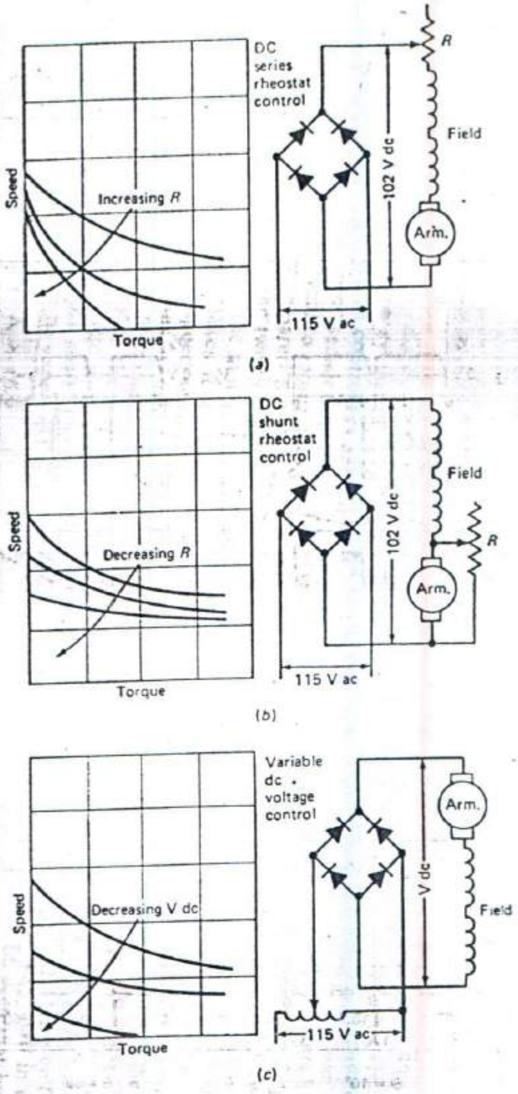


Fig. 3-13 Series (universal) speed control methods.

The *shunt-resistance control* of the series motor, along with its speed/torque curve, is shown in Fig. 3-13(*b*). The series motor can also be controlled by shunting an adjustable resistance across the armature as indicated. The speed control range is usually limited by this method because increased current must pass through the field coils, with a corresponding increase in heating. Although the speed range is limited, this method of control improves the speed regulation, as the curves indicate, while maintaining good starting torques. It is an excellent method for matching the speeds of motors operating in parallel.

Varying the voltage applied to the series motor typically produces a speed change of up to 7:1, depending on the individual motor. Figure 3-13(*c*) shows the circuit, along with the speed/torque curves obtained by *variable-voltage control*. It can be seen that the speed range increases, along with improved regulation and starting torque. The variable voltage can be obtained with the use of an autotransformer and a rectifier assembly. At this point the system may be considered simply as a blackbox: alternating current in, variable direct current out.

Since the series motor can run on alternating current as well as direct current, control for the motor can be half-wave or full-wave. The use of an SCR for control is shown in Fig. 3-14. This is a half-wave device with feedback. The circuit uses the cemf of the motor to vary the firing point, thereby maintaining essentially constant speed control with varying torque requirements. The SCR conducts on the positive half cycles only; therefore the control will provide substantially less than full speed. However, special motors (designed to operate in this circuit) may be used for full-speed operation.

For 60-Hz ac operation, a simple full-wave control such as the one illustrated in Fig. 3-15 can be used. This is an open loop control (no feedback). Varying R_2 varies the time for capacitor C_1 to charge to the diac trigger voltage. When the diac triggers it will turn on the triac, as shown in the waveforms for V_{C1},

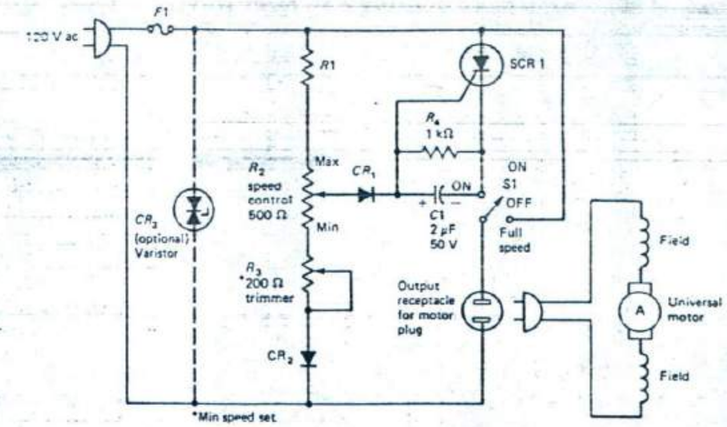


Fig. 3-14 SCR speed control of a universal motor.

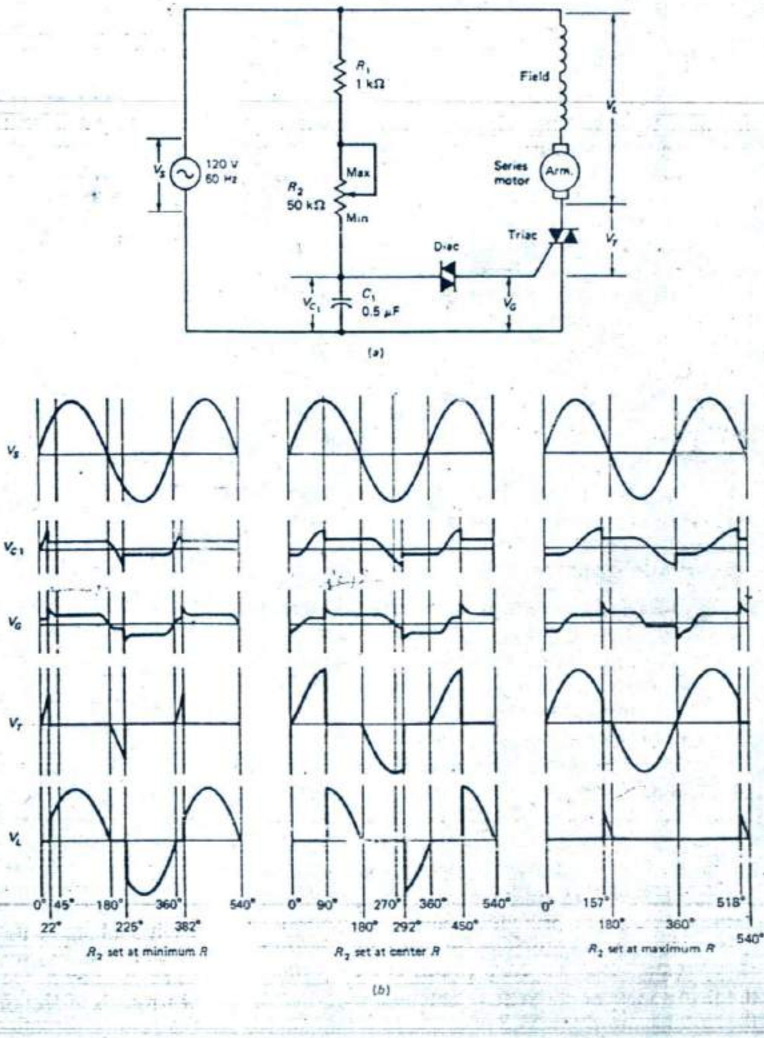


Fig. 3-15 Triac full-wave speed control. *(a)* Schematic diagram. *(b)* Waveforms.

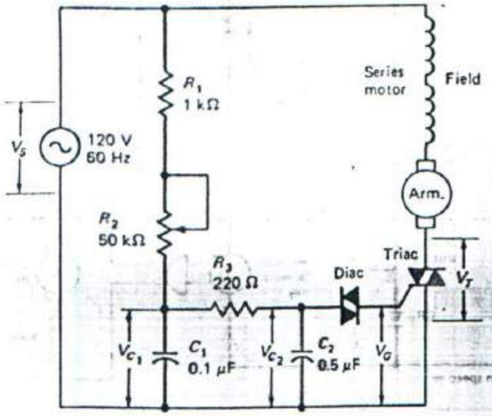


Fig. 3-16 Triac control with hysteresis compensation.

V_G, and V_T in Fig. 3-15(*b*). This will apply the supply voltage V_S, less the drop across the triac (V_T) across the load V_L. When the supply voltage goes through zero, the triac will turn off. It will remain off until capacitor C_1 again charges (in the reverse direction) to the diac trigger voltage. The waveforms show various settings of R_2 used to obtain early firing (22°) for maximum power to late firing (157°) for minimum power. This circuit does exhibit some dissymmetry of the load alternations. This is caused by circuit hysteresis due to capacitor C_1's retaining some charge at the polarity of the initial voltage applied across it. This dissymmetry is apparent by observing the V_L positive and negative waveforms of Fig. 3-15(*b*). The hysteresis can be reduced by using the circuit shown in Fig. 3-16. This circuit is also a popular arrangement for lamp dimmers.

REVIEW QUESTIONS

8. Field weakening of a shunt-wound motor will cause the armature speed to ______.

9. The field weakening method of speed control is reasonably efficient because the field current is ______ in comparison to the armature current.

10. A series rheostat in the armature of the shunt motor will cause its speed to ______ in comparison to base speed.

11. Resistance control of a motor is a form of ______ loop-type control.

12. ______ control of a shunt or PM motor is easily adapted to solid-state or computer-type devices.

13. The unfiltered (pulsating) voltage used in the UJT/SCR control is needed to ______ the SCR.

14. The ______ provides feedback for improved speed regulation in the SCR circuit of Fig. 3-10(*a*).

15. An increase in speed is obtained by ______ the firing angle of the SCR.

3-3

STEPPER MOTORS

The increasing trend toward digital control of machines and process functions has generated a demand for mechanical devices capable of delivering incremental motions of predictable accuracy. The stepper motor is often considered as a digital device which converts electric pulses into proportionate mechanical movement. Each revolution of the stepper motor's shaft is made up of a series of discrete individual steps. The motor usually provides for clockwise (cw) or counterclockwise (ccw) rotation. Therefore, the stepper motor is ideally suited for a wide variety of control and positioning applications in the industrial world. With the rapid growth of solid-state electronics and digital techniques, the stepper motor applications in peripherals, robotics, instrumentation controls, and machine tools have grown rapidly and continue to do so.

Conventional ac and dc motors have a free turning shaft. The stepper motor shaft rotation is incremental. The basic feature of the stepper motor is that upon being energized it will move and come to rest after some number of steps in strict accordance with the digital input commands provided. The stepping motor therefore allows control of the load's velocity, distance, and direction. The *repeatability* (ability to position through the same pattern of movements a number of times) is very good. The only system error introduced by the stepper motor is its single-step error, which is a small percentage of one step and is generally less than 5 percent (0.09°). Most significantly, this error is noncumulative, regardless of the distance traveled or the number of times repositioning takes place.

The stepping motor is an inherently reliable device with bearings being the only part subject to wear. In many applications, a stepping motor can replace shorter-lived, more maintenance-intensive devices such as brakes, clutches, and gears with an overall improvement in reliability.

Stepper motors are divided into three principal types or classes, each with distinct construction and performance characteristics: variable-reluctance (VR), permanent magnet (PM), and PM-hybrid.

A stepper motor's operation is based on the basic magnetic principle: like magnetic poles repel and unlike poles attract. If the stator windings in Fig. 3-17(*a*) are energized so that stator *A* is the north pole, stator *B* is the south pole, and the permanent magnet (PM) rotor is positioned as shown, a torque will be developed to position the rotor 180° from its indicated position. However, it would be impossible to determine the direction of rotation, and, in fact, the rotor may not move at all if the forces are perfectly balanced. If, as indicated in Fig. 3-17(*b*), two additional stator poles *C* and *D* are added and energized as shown, we are able to predict the direction of rotation of the rotor. As indicated in Fig. 3-17(*b*), the rotor's direction of rotation would be counter-

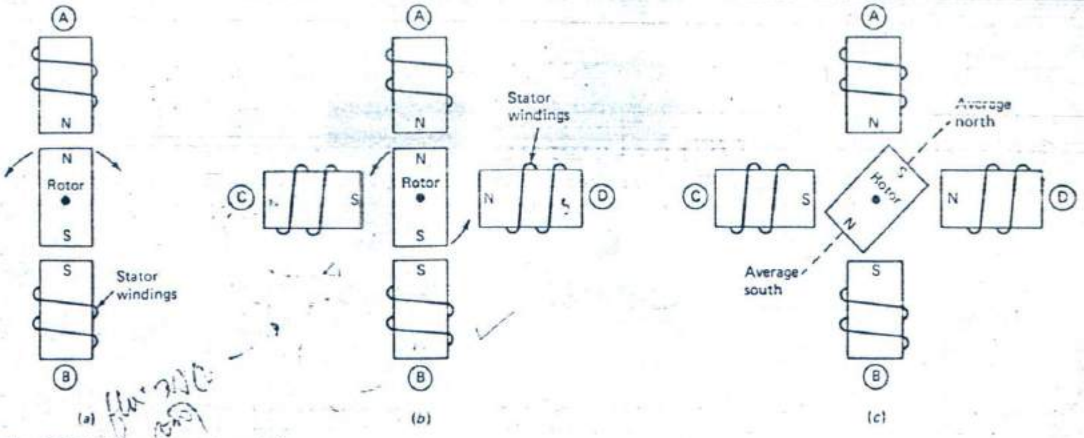


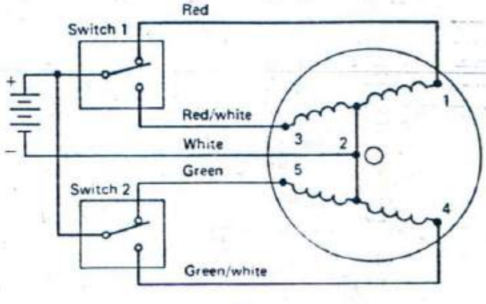
Fig. 3-17 Basic stepper motor rotation.

clockwise with the rotor aligning itself between the "average" south pole and the "average" north pole as in Fig. 3-17(*c*).

The distinguishing feature of the PM stepper motor is the incorporation of a permanent magnet, usually in the rotor assembly. To allow better step resolution, four more stator poles are added, and teeth are machined on each stator pole and also on the rotor. The number of teeth on the rotor and stator determines the step angle that will be obtained each time the polarity of one winding is changed. The stepper motor shaft responds with a specific angular increment each time the winding polarity is changed, moving to the average pole. This specific degree of shaft rotation or increment is known as the *step angle*. The PM stepping motor operates by means of the interactions between the rotor magnet biasing flux and the magnetic forces generated by the stator windings. If the pattern of winding energization is fixed, a series of stable equilibrium points is generated around the motor. If the windings are excited in a particular sequence, the rotor will follow the changing point of equilibrium and rotate in response to the changing pattern, as shown in Fig. 3-18.

By virtue of the rotor's permanent magnet, there is a detent torque developed in the motor even if the stator windings are not excited. The detent torque can be felt by turning a PM stepper by hand. A restoring torque is generated on the rotor whenever the rotor is moved from the position which has minimum reluctance (analogous to resistance in a dc circuit) for the permanent magnet flux. This torque is much lower than the normal energized torque and is typically only a small percentage of the maximum torque.

The variable-reluctance (VR) motor has a stator which has a number of wound poles. The rotor is a cylindrical, toothed unit whose teeth have a relationship to the stator poles and their teeth (the stator may not have teeth). The number of teeth will be

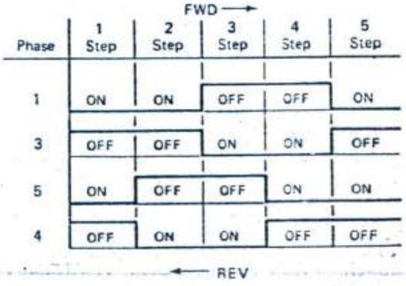


(*a*)

Step	Switch #1	Switch #2
1	1	5
2	1	4
3	3	4
4	3	5
1	1	5

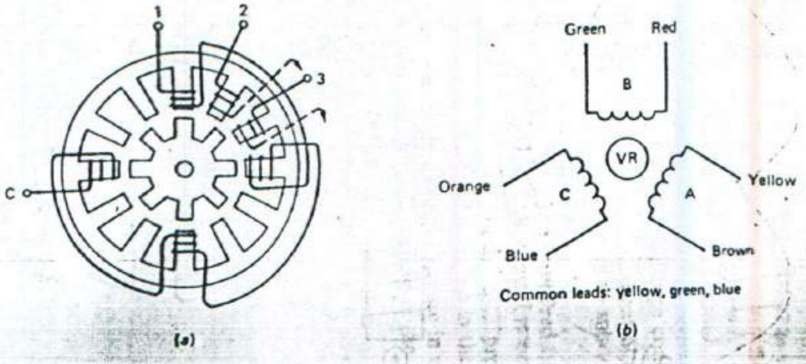
*To reverse direction, read chart up from bottom.

(*b*)



(*c*)

Fig. 3-18 Four-step sequence of a permanent magnet stepper motor. (*a*) Schematic diagram. (*b*) Switching sequence. (*c*) Waveforms.



Step \ Phase	A Brown	B Red	C Orange
1	▨	▨	
2		▨	▨
3			▨
4	▨	▨	

(c)

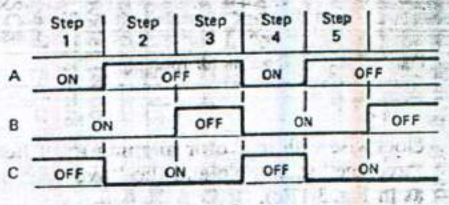


Fig. 3-19 Variable reluctance (VR) stepper motor. *(a)* Cross section of stator showing complete winding of one phase of a three-phase winding. *(b)* Three-phase wiring connection color code. *(c)* Three-phase, dual-excitation chart. *(d)* Stepping waveforms.

determined by the step angle required. A typical VR motor is shown in Fig. 3-19. When a current flows through the proper windings, a torque is developed in such a way as to turn the rotor to a position of minimum magnetic path reluctance. This position will be statically stable in that external torque is required to move the rotor from this stable position. This particular position is not an absolute one. There are many stable positions in the average motor for any given stator energization pattern. When a different set of windings is energized, the stator field changes, causing the rotor to move to a new position. Proper selection of the energizing sequence of the windings allows the stable positions to be made to rotate smoothly around the stator poles, establishing the rotational speed and the direction of the rotor. When the energized pattern is fixed, the rotor position becomes fixed as well. Therefore, the shaft position is stepped by changing the pattern of winding energization. Figure 3-19(*c*) illustrates the standard excitation modes which produce a nominal step angle. The dual excitation (two windings always on) is chosen because of the higher torque available. Unlike the PM stepper, the VR stepper has very little residual magnetism, so there will be no force on the rotor (*detent torque*) when the stator is not energized. The *step angle* (determined by the number of stator and rotor teeth) varies from 7.5 to 30°. The VR steppers exhibit relatively low torque and inertial load capacity. They are, however, reasonably inexpensive and suitable for light computer and industrial instrument applications.

The PM-hybrid stepper rotor combines the rotor construction features of VR and PM types. A cross section of this stepper is shown in Fig. 3-20. The stator is of the wound type. Both the rotor and the wound stator are toothed. This construction gives the PM-hybrid higher torque capacity (50 to 2000+ ounce-inches [oz·in.]) with step accuracies of about ±3 percent and step angles that vary from 0.5 to 15°. PM-hybrid designs offer excellent speed capability: 1000 steps/s and higher can be obtained. Although their cost is relatively high, PM-hybrid designs deliver the best set of performance characteristics for many applications.

Steppers are popular because they can be used in an open loop mode while still offering many of the desirable features of the feedback-type system: feed them a defined number of pulses, and they will position within their step accuracy. The replacement of mechanical parts (which are susceptible to wear),

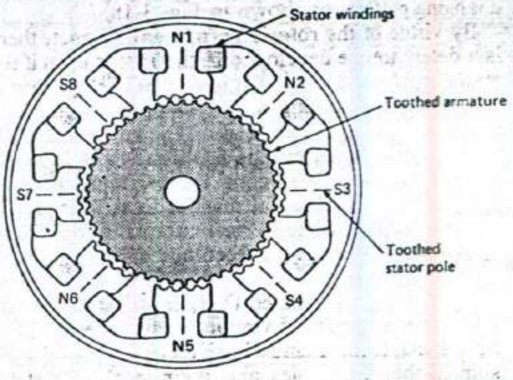


Fig. 3-20 Cross section of PM-hybrid motor with toothed armature and stator.

such as clutches and brakes, is eliminated because stepper motors provide a greater reliability and consistency, reasonable cost, and consistent performance. The stepper motor is an excellent positioning device. On the other hand, stepper motors are not very energy-efficient, but that is the price one must pay to obtain the unique characteristics of the stepper motor. Its limitations include the following: available torque is inversely proportional to speed; speed must increase gradually (if commanded to go from stop to full speed immediately, it will stall); and the stepper motor exhibits a low speed resonance point, where torque is reduced drastically.

Some stepper motors achieve torque amplification by means of a gear system integral with the motor. Methods to achieve torque amplification are the use of planetary gears and the use of a *flexing mechanical spline*. Both methods allow a reduction in speed and an increase in output torque by gearing down. For instance, a stepper that delivers 50 oz·in. of torque at 72 rpm will, with a 4:1 step-down gearing, deliver 200 oz·in. of output torque at a speed of 18 rpm.

To understand the stepping motor's unique characteristics thoroughly, we must examine the stepping motor logic sequencing. A simplified representation of a stepping motor is shown in Fig. 3-21. Initially, poles *A* and *B* are both energized with their north poles up, attracting the rotor's south pole to the position as shown in Fig. 3-21(*a*). Reversing the polarity of pole *A* (Fig. 3-21[*b*]) draws the rotor clockwise 90° to its new position: this is known as a *full step*. If pole *A* had been turned off, instead of being reversed, the rotor would rotate 45° (clockwise) to line up with the field of pole *B*; this is known as a *half step*. The simple stepper motor in Fig. 3-21 would only have four full steps (90°) per revolution, or eight (45°) half steps. Actual stepper motors obtain small angle increments by using large numbers of poles, as shown in Fig. 3-20.

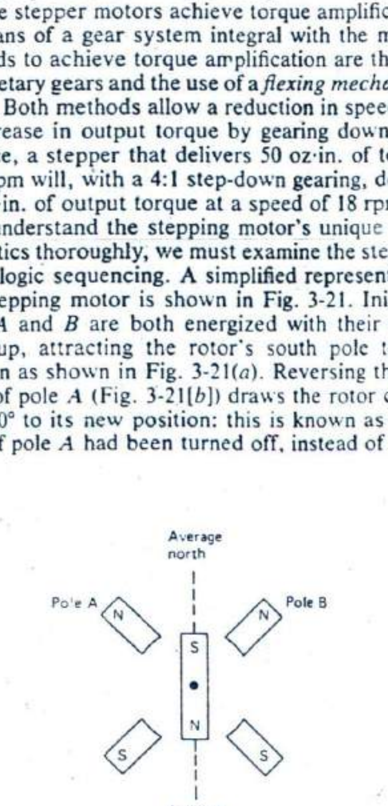


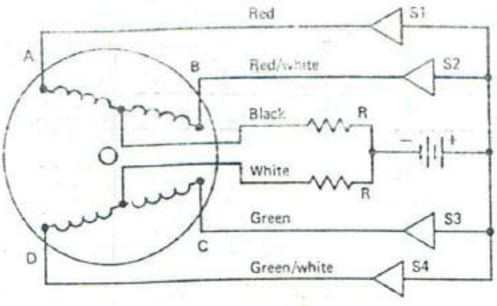
Fig. 3-21 Simplified stepping sequence.

The most common stepper stator windings are center-tapped dual windings known as *bifilar windings*. Bifilar winding eliminates transformer coupling to adjacent windings. The use of bifilar windings on steppers also simplifies the required drive circuitry. Figure 3-22(*a*) shows a bifilar-wound stepper motor, its power supply, and the switching points. Only a single-polarity power supply is needed, whereas the motor of Fig. 3-21 would require a dual power supply for reversal of the poles. Only a single power supply is needed with the center-tapped windings.

The switching sequence shown in Fig. 3-22(*b*) is called a *four-step sequence* (full step). To reverse direction, read the sequencing chart upward from the bottom. Since current is maintained on the motor windings when the motor is not being stepped, a high holding torque results. As a rule of thumb you will find that the power supply is about five times the voltage rating of the motor. Series resistors are used in the common leads to limit the current and improve the inductive/resistive (*L*/*R*) time constant, for better performance.

Figure 3-22(*c*) illustrates the eight-step switching sequence, often called *electronic half stepping*. Under this condition, the rotor moves half its normal distance per step. For example: a 1.8°, 200 step per revolution motor would become a 0.9°, 400 step per revolution motor. Likewise a 0.72°, 500 step per revolution model moves in increments of 0.36° for 1000 steps per revolution. The advantages of operating in this mode include finer resolution and greater speed capability, but with less available torque.

The switching sequence for these motors was originally achieved by mechanical switches, mercury-wetted relays, or a commutator-brush arrangement. These were very awkward and expensive, besides being a maintenance headache. Electronic switching resolved this problem quite easily and efficiently. The circuits of Fig. 3-23 (p. 57) use solid-state devices to drive (switch) the stepper motor windings. These power drivers are used because most steppers require currents from hundreds of milliamperes up to amperes. A 200-oz·in. torque motor operating with 2.5-V windings will require approximately 2.5 A for each winding that is on. This amount of current switching or sinking is too large for logic circuitry to provide, and drivers are required. Metal-oxide semiconductor field effect transistors (MOSFETs) are taking over this type of switching application. An attractive feature of the MOSFET is the built-in diode. The *integral body diode* is so named because it is inherent in the silicon structure of the power MOSFET. When the transistor switches off, a large voltage due to the collapsing field is generated in the



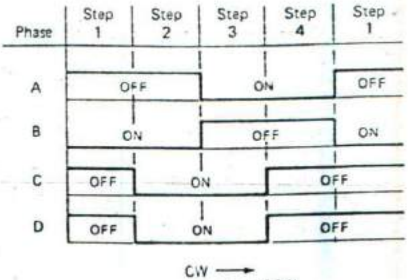
(a)

CW ↓

Step	S1	S2	S3	S4
1	OFF	ON	OFF	ON
2	OFF	ON	ON	OFF
3	ON	OFF	ON	OFF
4	ON	OFF	OFF	ON
1	OFF	ON	OFF	ON

↑ CCW

To reverse direction, read chart up from bottom.



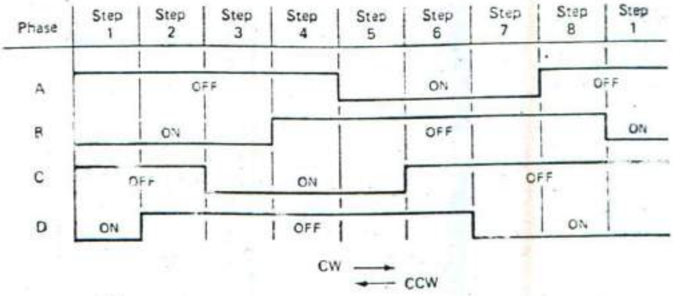
(b)

Half-Step Sequence

CW ↓

Step	S1	S2	S3	S4
1	OFF	ON	OFF	ON
2	OFF	ON	OFF	ON
3	OFF	ON	ON	OFF
4	OFF	OFF	ON	OFF
5	ON	OFF	ON	OFF
6	ON	OFF	OFF	OFF
7	ON	OFF	OFF	ON
8	OFF	OFF	OFF	ON
1	OFF	ON	OFF	ON

↑ CCW



(c)

Fig. 3-22 Bifilar stepper motor. *(a)* Circuit diagram. *(b)* Full-step sequence and waveforms. *(c)* Half-step sequence and waveforms.

motor winding. The voltage forward-biases the body diode, and breakdown is avoided. The MOSFETs will also switch large currents faster than bipolar transistors.

The block diagrams for the logic circuitry required to obtain a full-step sequence for Fig. 3-22(*b*) is shown in Fig. 3-24. (Logic devices are covered in depth in Chapter 11.) Any type of logic family that supports the circuit requirements can be used. Figure 3-24(*b*) shows the typical circuitry required using transistor-transistor logic (TTL). The outputs are then applied to the drivers previously mentioned.

Figure 3-24(*c*) shows the simplicity of using a complementary metallic oxide semiconductor (CMOS) circuit to obtain either the full- or half-step drive for the bifilar motor. It should be noted that the full-step control sequence is two on-time periods followed by two off-time periods. The rate of stepping is determined by the frequency of the applied clock, with each input pulse causing one step (or half step) in the stepper motor shaft. The half-step sequence pattern is three on-time periods followed by five off-time periods; like the full-step control it is very easily obtained by solid-state circuitry.

This type of control leaves a lot to be desired in some modern computerized industrial and robotic

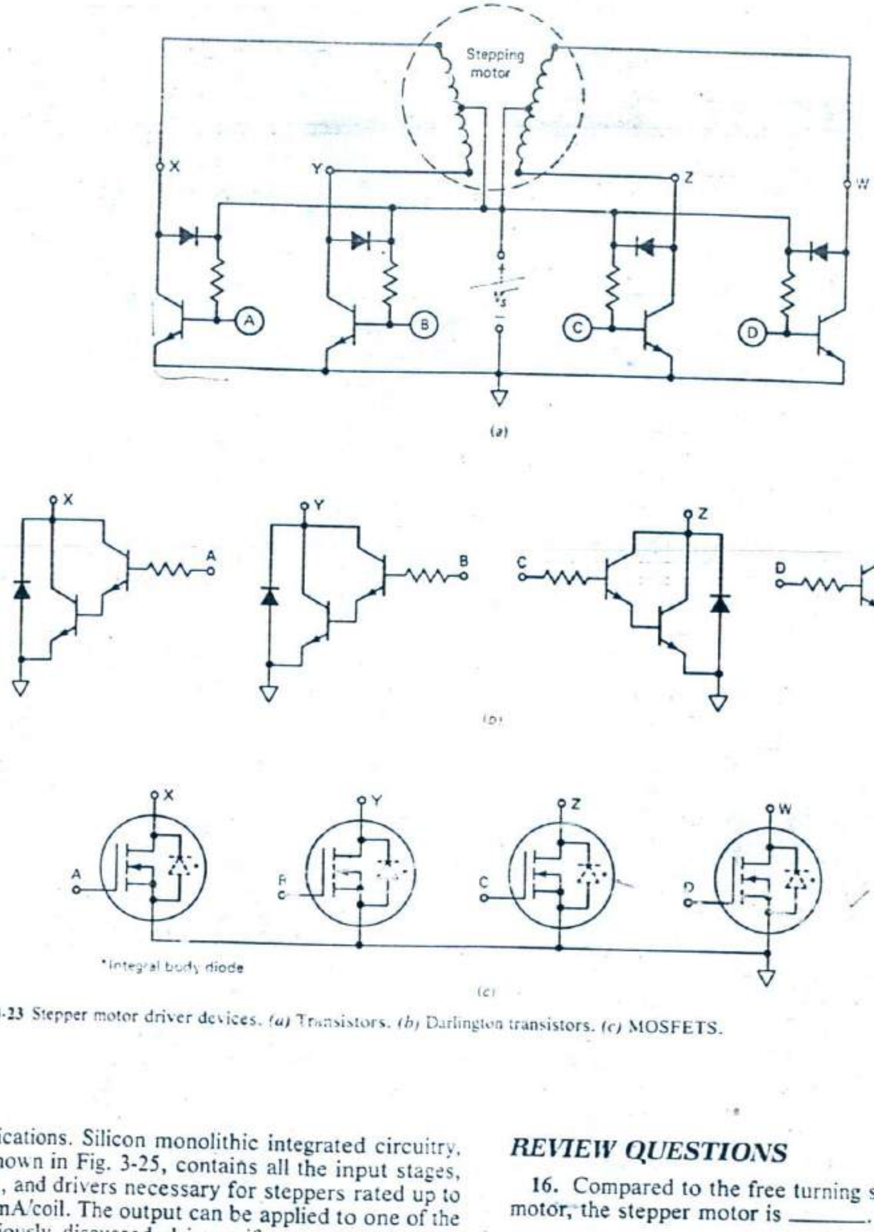


Fig. 3-23 Stepper motor driver devices. *(a)* Transistors. *(b)* Darlington transistors. *(c)* MOSFETS.

applications. Silicon monolithic integrated circuitry, as shown in Fig. 3-25, contains all the input stages, logic, and drivers necessary for steppers rated up to 500 mA/coil. The output can be applied to one of the previously discussed drivers if more power is required. The simplicity of one integrated circuit (IC) is very desirable and cost-effective.

Stepper motors come in a wide variety of sizes, types, and styles. The basic stepping principle for all is the same, and they fall into one of the three types discussed. They may have as little as two windings, or as many as ten phase windings (these take 2500 steps for one revolution).

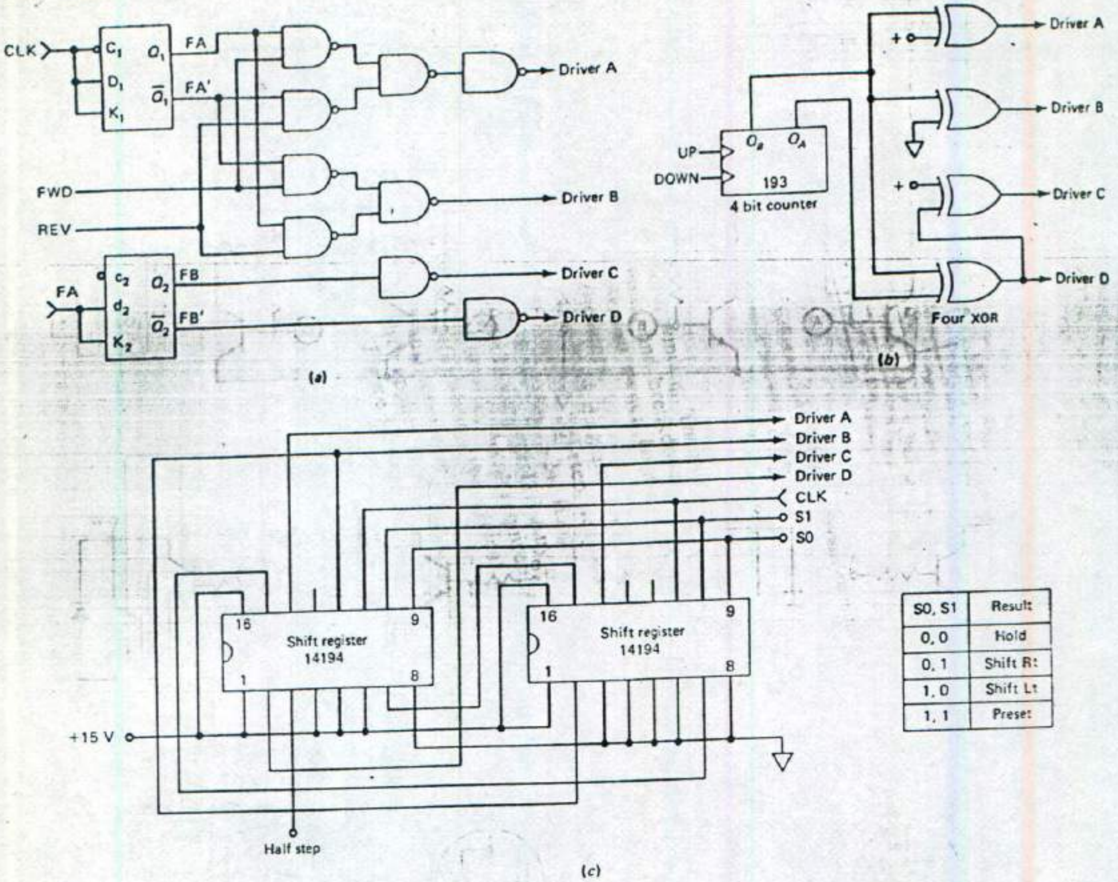
REVIEW QUESTIONS

16. Compared to the free turning shaft of the dc motor, the stepper motor is ______.

17. The stepper motor rotor illustrated in Fig. 3-17(*c*) seeks the ______ pole.

18. The rotor of the PM and PM-hybrid stepper motor is made of what kind of material?

19. The variable-reluctance stepper has ______ residual magnetism.



S0, S1	Result
0, 0	Hold
0, 1	Shift Rt
1, 0	Shift Lt
1, 1	Preset

Fig. 3-24 Stepper control logic. *(a)* Discrete. *(b)* LSI TTL. *(c)* CMOS.

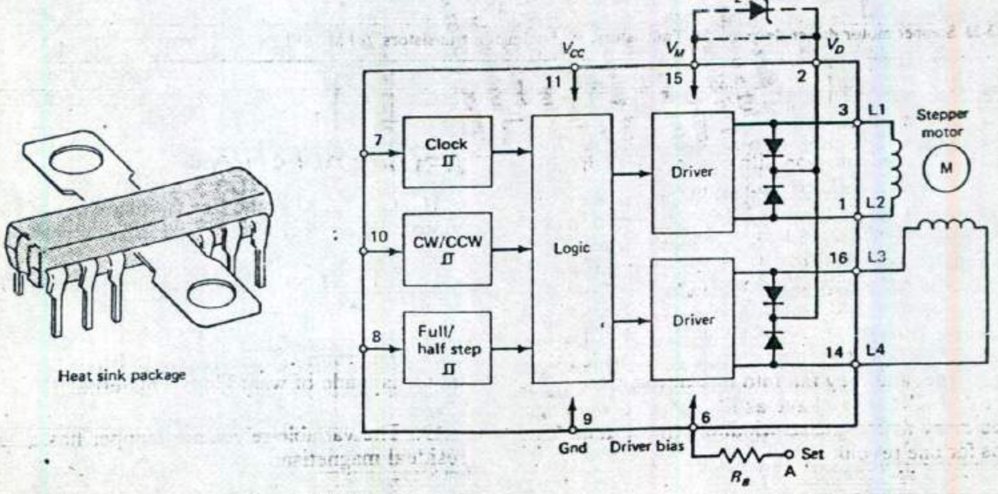


Fig. 3-25 Monolithic stepper motor driver with logic.

3-4
BRAKING AC MOTORS

Induction motors can be dynamically braked to an abrupt stop, just as dc motors can. They are dynamically braked by removing the ac power from the motor and substituting dc power. When this is done, the motor will act the same as the dc shunt motors previously described. The stator of the induction motor, with direct current applied, is similar to the field (stator) of a shunt motor, and the squirrel cage rotor is similar to a shorted armature when dynamically braked. The rotor acts like a dc generator with a shorted armature, with high circulating currents in the rotor bars. This rotational energy is dissipated in the form of heat (in the rotor) when the motor stops abruptly.

The source of direct current for braking may be large batteries or power supplies. The direct current may also be supplied by a charged capacitor. Figure 3-26(*a–d*) shows examples (simple circuits) of ac induction motor braking: Fig. 3-26(*e*) illustrates the *capacitor discharge method*. The capacitor is charged while the motor is running and discharges through the windings when the motor is braked. This arrangement eliminates the necessity for any external power source.

REVIEW QUESTIONS

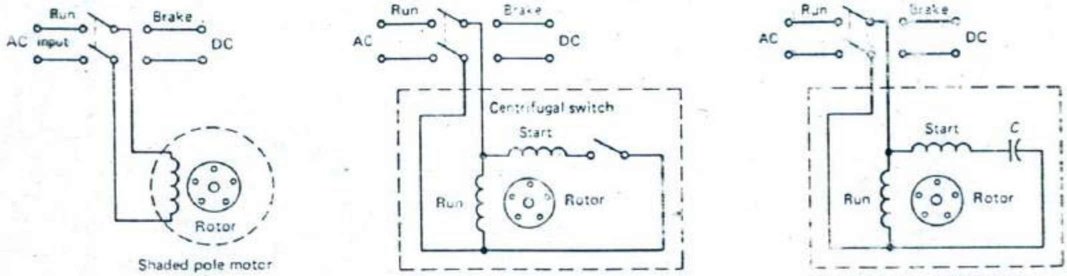
20. Induction motors can be dynamically braked by applying ______ current to their windings.

21. Capacitor discharge braking requires ______ external power source.

3-5
BRUSHLESS MOTORS

The term *brushless* has been applied to a wide variety of electric rotating devices. Most earlier devices were used in applications in which unidirectional operation was acceptable and minimum electronics was a major consideration, as was elimination of the mechanical brush-commutator mechanism. The brushless dc motor uses ac stator voltages of two, three, four, or six phases. Typically, the brushless dc motor has the torque/speed characteristics of the conventional dc PM motor.

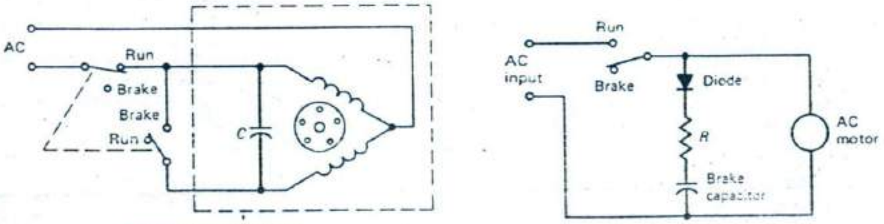
Development of low-cost position sensors (such as Hall effect integrated circuits and switching optical sensors), economies favoring the use of rare earth magnetic materials, and the availability of low-priced semiconductors permit a more competitive position for brushless dc motors. They are especially attrac-



Shaded pole motor (*a*)

Split-phase motor (*b*)

Permanent split capacitor motor (*c*)



Capacitor shorting method

(*d*)

(*e*)

Fig. 3-26 Examples of induction motor braking.

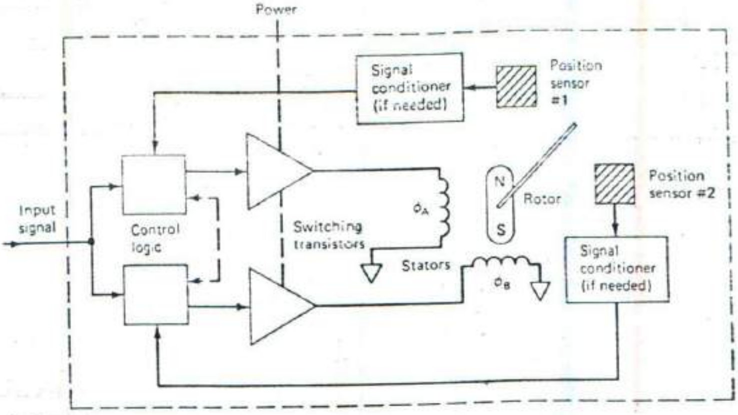


Fig. 3-27 Simplified block diagram of brushless dc motor.

tive in applications requiring tachometer feedback units. The brushless motor has high efficiency, long life, low noise, and low power consumption.

Electronically commutated brushless dc motors operate on the same motor action as the conventional dc motor, except that the supply current switching (in reference to rotor position) is accomplished by solid-state circuitry instead of the mechanical brush-commutator combination used in conventional dc motors. In brushless dc motors transistorized logic senses the position of the PM rotor and controls the distribution of current to the field windings. The field windings are energized in sequence to produce a revolving magnetic field. Rotor position is sensed by solid-state light emitters and sensors, Hall effect devices, or some other means. The sensor feedback signals return to the control unit, which feeds the drivers that turn on transistor switches, thereby delivering the sequentially rotating current to the field coils (stator). Figure 3-27 shows a brushless motor in basic block form. It should be emphasized that the brushless dc motor is not a stepper motor and has smooth continuous shaft rotation and not the fixed-step detents found in the stepper motor. Because of the switching logic and circuitry associated with its control, it could be mistaken for a stepper motor.

The most commonly used methods for angular position sensing are Hall effect sensors and optical sensors. The *Hall effect sensor* detects the magnitude and polarity of a magnetic field. The signals are amplified and processed (within the IC) to form logic-compatible signals. A simplified block diagram of a Hall effect sensor is shown in Fig. 3-28. The Hall element has a constant current (I_H) passed through the element, which is usually indium. A magnetic field with a flux density (B) is applied at right angles to the element and causes the charge carriers to be redistributed within the element, causing a voltage V_H, the Hall voltage, to be induced in a direction perpendicular to the current and magnetic field. The

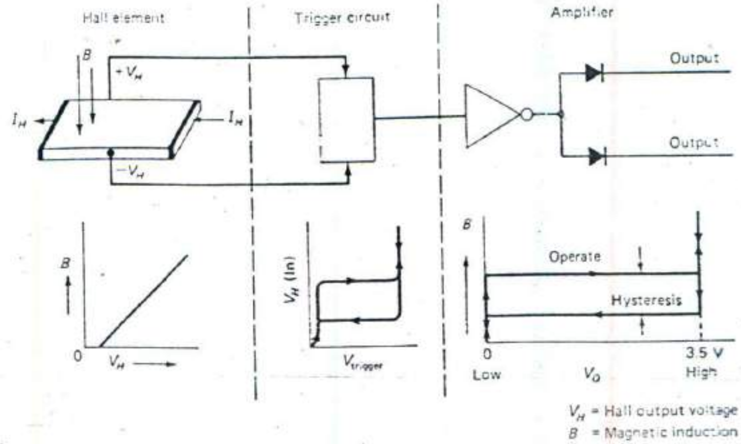


Fig. 3-28 Simplified block diagram of Hall effect sensor.

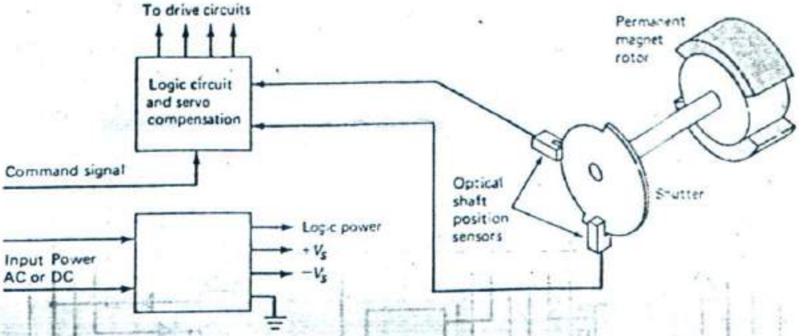


Fig. 3-29 Optical rotor position sensing.

Hall voltage is proportional to I (current) and B (flux density) and is expressed as

$$V_H = \frac{R \times I \times B}{D}$$

where

V_H = Hall voltage, V
B = Flux density, gauss (G)
I = Current, A
D = Element thickness, m
R = Hall constant

In Fig. 3-28, the Hall element exhibits a linear output, but the trigger circuit (comparator) transforms the Hall signal into an ON or OFF signal, and the amplifier shown is the interface to the desired control circuitry.

The other method of sensing rotor position for electronic commutation is the use of photodetectors. The basic concept is illustrated in Fig. 3-29. The stationary light source (solid-state or incandescent) emits a beam that is interrupted by a circular disk chopper mounted on the rotor shaft. The disk chopper will have some number of sectors removed, depending on the number of photodetector sensors and stator phases. A 60° sector would be used with six photodetectors, whereas a 180° sector would be used with two photodetectors (as shown in Fig. 3-29). The photodetectors (covered in Chapter 14) produce an output that is amplified and converted to a drive signal, as with the Hall effect device. These signals are then applied to the control circuitry to produce a rotating field. The electronics, light source, detectors, amplifiers, logic, and drivers are mounted integrally with the motor, as shown in Fig. 3-30. With the advances in solid-state technology, this type of control may eventually be reduced to a single integrated circuit.

The brushless motor stators can be connected in several variations: single-phase, two-phase, three-phase wye, three-phase delta, four-phase delta, and six-phase delta. A rotor having 4 poles and a six-phase delta-connected stator will commutate 12 times per revolution every 30° of the rotor shaft. This arrangement gives the highest power efficiency and peak torque. The major drawback is that it requires 12 switching transistors. By comparison, a four-phase, 4-pole configuration will yield the same peak torque output and requires only eight switching transistors. The three-phase wye and three-phase delta, 4-pole configuration can be full-wave excited by us-

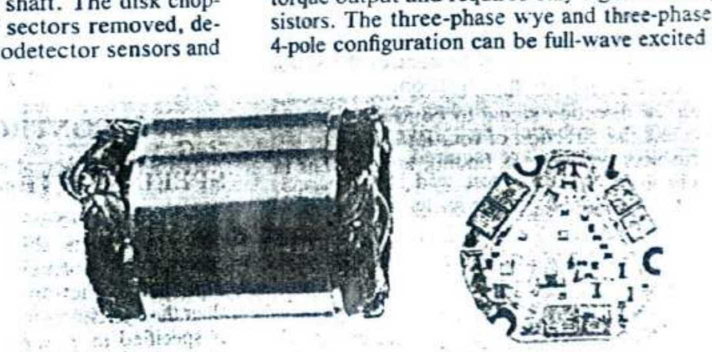


Fig. 3-30 Typical brushless motor and control circuit board.

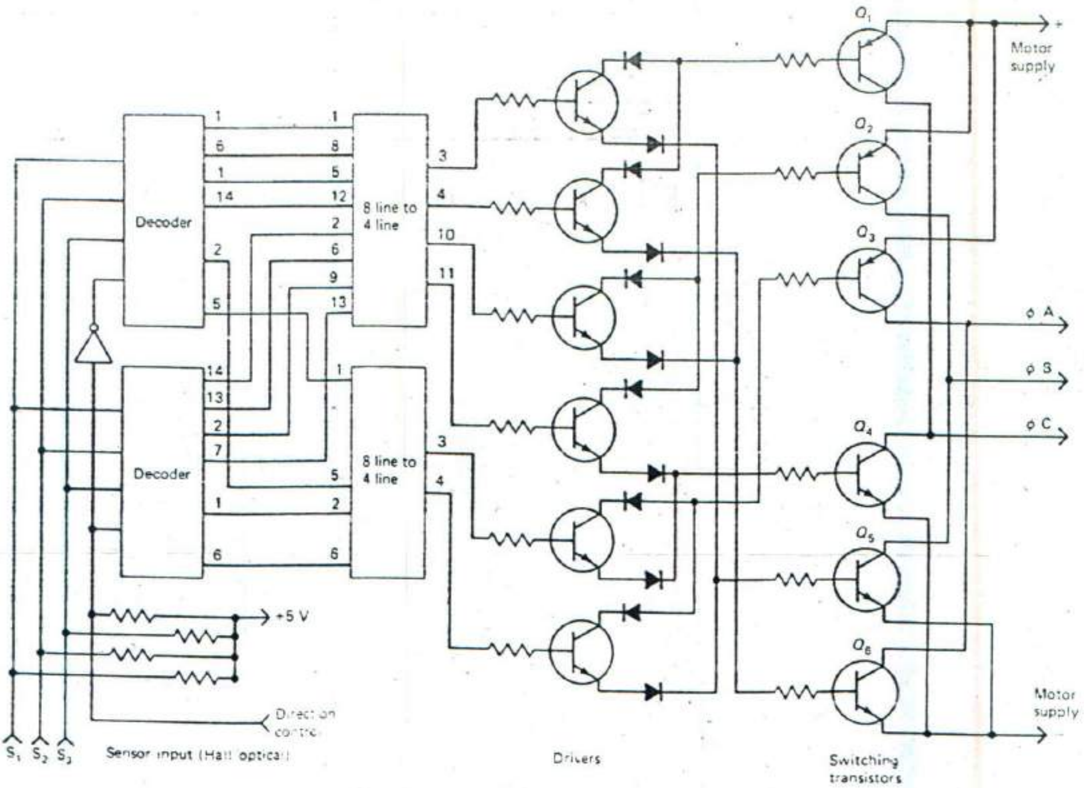


Fig. 3-31 Three-phase commutation circuitry.

ing six transistors with commutation 12 times per revolution.

A schematic for three-phase commutation circuitry is shown in Fig. 3-31. The input from the sensors (omitted for clarity) is connected to the decoders along with the direction signal to control the stator excitation and the direction of rotation.

The brushless dc motor is finding applications in servomechanisms, robotics, disk drives, and wherever direct drive motors are required. Speeds up to 25,000 rpm are not uncommon, and the nonarcing characteristics make it desirable for hazardous areas.

REVIEW QUESTIONS

22. The brushless motor exhibits torque/speed curves similar to those of the dc ______ motor.

23. The stator of the brushless dc motor has alternating or direct current applied to the ______ windings.

24. The brushless motor has a(n) ______ type rotor.

25. Two types of rotor position sensors are light-sensitive and ______ effect type.

26. The brushless motor, like the stepping motor, exhibits detent torque. (true or false)

3-6
SPEED CONTROL OF AC MOTORS

One of the principal characteristics of the ac induction motor is its ability to maintain a near-constant speed under normal load and supply voltage variations. All induction motors have a full speed lower than the synchronous speed. This reduction in speed is specified as *percent slip*. The synchronous speed N is a function of line frequency (f) and number of poles (P):

$$N = \frac{120f}{P}$$

This formula shows that the supply frequency and the number of poles are the factors that determine the speed of the motor. Unlike in the dc motor, the speed of an ac motor is not changed by varying the applied voltage. (One exception is the universal motor covered earlier in this chapter.) Reducing the applied voltage of a large motor in order to reduce its speed could damage the motor. This is due to the excess heat build-up inside the motor. Most ac motors are not designed to have their applied voltage vary more than 10 percent of the nameplate ratings.

Multispeed ac motors, designed to be operated at a constant frequency, are provided with stator windings that can be reconnected to provide a change in the number of poles and thus a change in the speed. These multispeed motors have only a few speeds (usually no more than four), and these speeds are widely separated from each other.

To change the speed of a split-phase motor, it is necessary to change the number of poles. Speed changing a capacitor-start motor is also done by changing the number of poles in the winding. A variable resistor or transformer can be used to vary the voltage across the winding of a permanent-split capacitor motor. These methods are shown in Fig. 3-32. The variable voltage transformer (autotransformer) is preferred to the series resistor. Figure 3-32(*b*) shows this method, which has the advantage of keeping a constant voltage on the motor whether under start or run conditions. There is also a large reduction in heat loss compared to the resistor method of Fig. 3-32(*a*). By reducing the voltage across the run winding only and maintaining full voltage across the capacitor (start) winding, more stable low-speed operation is attained. This method is illustrated in Fig. 3-32(*c*).

The *polyphase* (three-phase) *motor* utilizes two separate and independent windings for each pole. With this arrangement, any desired combination of two, three, or four different speeds is possible. Speed control of ac squirrel cage induction motors can be accomplished if the frequency of the applied voltage to the stator is varied to change the synchronous speed. The change in the synchronous speed (stator rotating field frequency) results in a change in the motor speed. There are two basic methods used to vary the frequency of the applied voltage to the ac motor: one uses an inverter, the other a converter. The inverter is used to change a dc voltage to an ac voltage whose output frequency can be varied. The converter will change the standard 60-Hz line frequency into almost any desired frequency. Both the inverter and converter use solid-state devices (mainly SCRs) for control. Prior to the use of solid-state devices, variable frequency was provided by variable-speed motor-alternator sets. The initial and maintenance costs are high for this approach in comparison to those of the solid-state devices. The solid-state designs are often referred to as *static frequency converters* (no moving parts). The operating frequencies are typically in the range of 10 to 200 Hz. The major applications for variable-frequency drives are textile machinery, machine tools, and steel and paper mills.

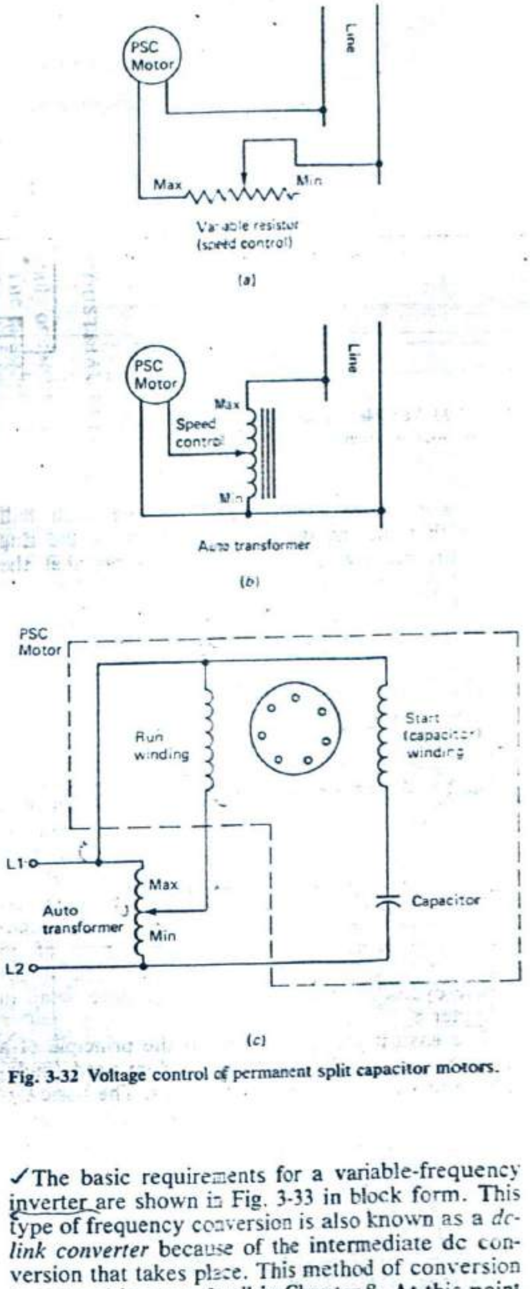


Fig. 3-32 Voltage control of permanent split capacitor motors.

✓The basic requirements for a variable-frequency inverter are shown in Fig. 3-33 in block form. This type of frequency conversion is also known as a *dc-link converter* because of the intermediate dc conversion that takes place. This method of conversion is covered in more detail in Chapter 8. At this point we will examine the system at the block level. The desired speed signal is applied to both the voltage-controlled oscillator and the pulse-width modulator. The three-ring counter (count 1,2,3, 1,2,3, as the name implies), via the logic circuits, controls the distribution of gate pulses to the six-step inverter. It contains six thyristors arranged for a three-phase

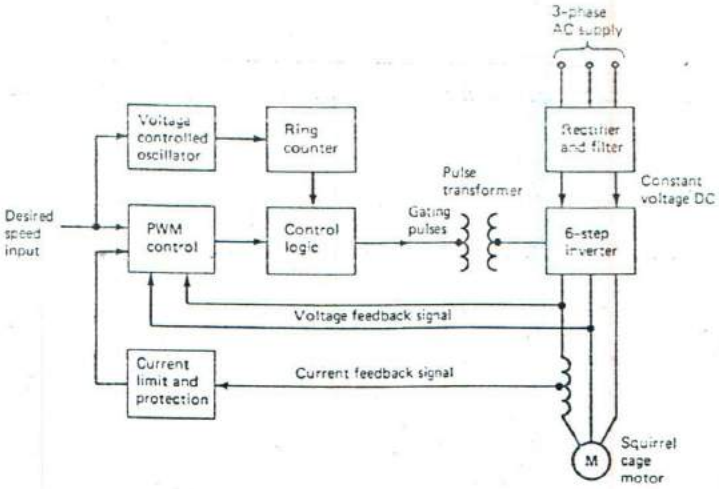


Fig. 3-33 Variable frequency converter block diagram.

full-wave bridge configuration. During each half cycle that the thyristors are gated on by the ring counter, pulse-width modulation ensures that the thyristors are switched on and off at the proper times to achieve voltage control within the inverter. Voltage feedback permits control of constant torque or horsepower as desired. At the same time, stator current control is obtained by means of current feedback to the pulse-width modulator control block.

The other method of static frequency conversion to obtain speed control of ac motors is the *ac-to-ac frequency converter,* which is commonly called a *cycloconverter.* Though inherently more efficient than the dc-link inverter, it has two major disadvantages. First, in order to keep the output harmonics at an acceptable level, it must be operated in a range from 0 Hz to one-third of the ac input source frequency. Second, to obtain bidirectional operation, a three-phase cycloconverter requires a minimum of 36 SCRs with complex control circuitry. The three-phase cycloconverter is presented in more detail in Chapter 8.

The easiest way to understand the principle of a cycloconverter is to study the seldom used single-phase to single-phase cycloconverter. The basic circuit shown in Fig. 3-34 has two two-pulse midpoint phase-controlled converters. One forms the positive group and the other the negative group, which is a dual-converter configuration. The output current of each group flows in only one direction (SCRs are unidirectional devices). To produce an alternating current in the load, the two groups, positive and negative, must be connected in inverse parallel as shown in Fig. 3-34. The positive group (SCR 1 and SCR 3) permit load current to flow only during the positive half of the cycle, when V_{AN} or V_{BN} is positive, and the negative group permits current flow during the negative half cycle, when V_{AN} or V_{BN} is negative.

The waveforms are shown in Fig. 3-35, and an inductive load is assumed. Voltages V_{AN} and V_{BN} are 180° out of phase (*antiphase*). By varying the firing points (t_1 to t_4) of the thyristors, the mean output voltage to the load can be varied as indicated. Varying the firing delay angles about 90° in a sinusoidal manner at the desired output frequency will result in a mean output load voltage that is sinusoidal at the desired frequency, as indicated in Fig. 3-35.

The ac shaded pole motor can be made to vary its speed by varying the voltage across its windings. This is an inefficient method and produces a lot of excess heat. For years these motors (used for fans) had their speed varied by inserting a choke in series with the main winding, which eliminated the *IR* (resistive) losses.

The best passive device, however, is an autotransformer. It has the advantage of maintaining the same voltage under high starting current. There is also

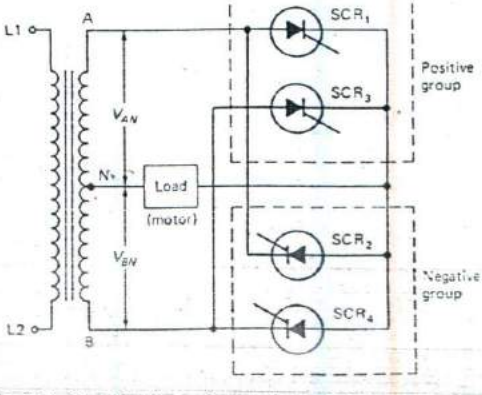


Fig. 3-34 Simplified single-phase cycloconverter.

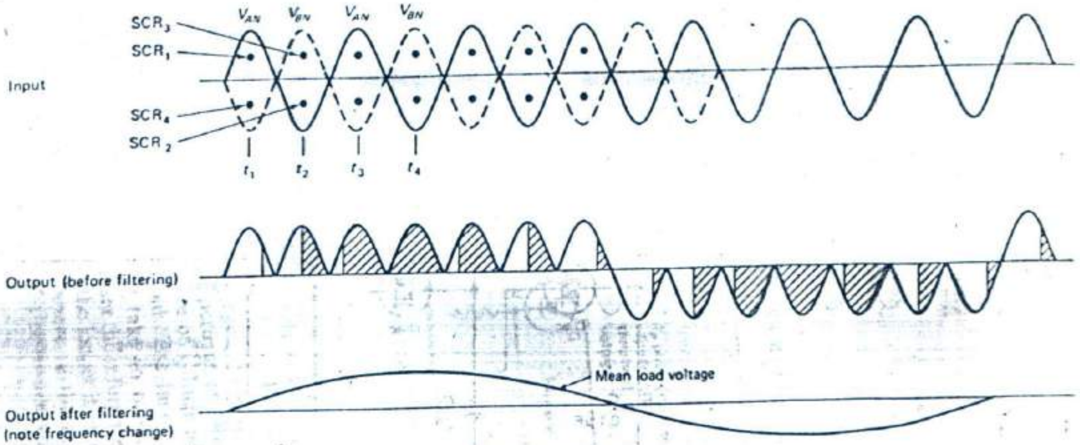


Fig. 3-35 Single-phase cycloconverter waveforms.

very little heat loss compared to the other methods. By reducing the voltage across the run winding (only) of the permanent-split capacitor motor (capacitor run), while maintaining full voltage across the start winding, the speed can be varied by varying the output voltage of an autotransformer.

Most modern variable ac voltage sources are derived from solid-state electronics, which is more cost-effective, reliable, and efficient than resistors, chokes, or transformers for varying ac voltages.

The ac power to a shaded-pole or permanent-capacitor motor is controlled by varying the rms value of the ac voltage using thyristors. The typical method used to vary the voltage is phase control. Unlike brush-type motors, induction motors give no convenient electrical indication of their mechanical speed. This means that direct speed feedback is not usually available. For many applications, such as fixed fan loads, direct voltage adjustment with no feedback (open loop) yields satisfactory performance. An example of this type of circuit is shown in Fig. 3-36(*a*). The single time constant circuit provides satisfactory proportional speed control for shaded-pole and permanent split-capacitor motors when the load is fixed. This type of circuit is best suited for applications that need speed control in the medium- to full-power range. It is especially useful for fan or blower control, in which a small change in motor speed produces a large change in air velocity. Caution must be exercised because the motor may stall if the speed is reduced below the motor's drop-out speed. The single time constant circuit is employed because it cannot provide speed control from maximum to full off. Speed ratios as high as 3:1 can be obtained from the single time constant circuit used with these types of induction motors. As a result of the inductive load and because the triac turns off when the current reaches zero the source voltage may be at a value other than zero. This commutating

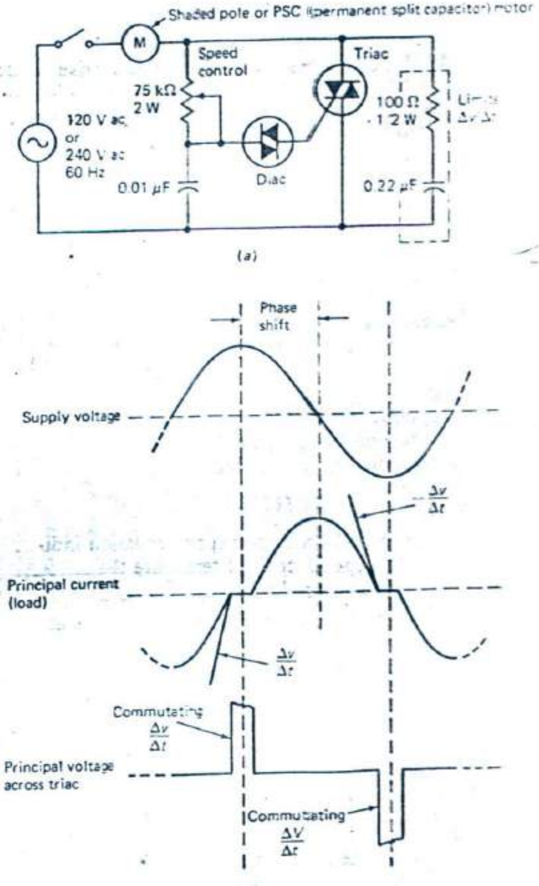


Fig. 3-36 Solid-state speed control. (*a*) Control circuit. (*b*) Voltage and current waveforms.

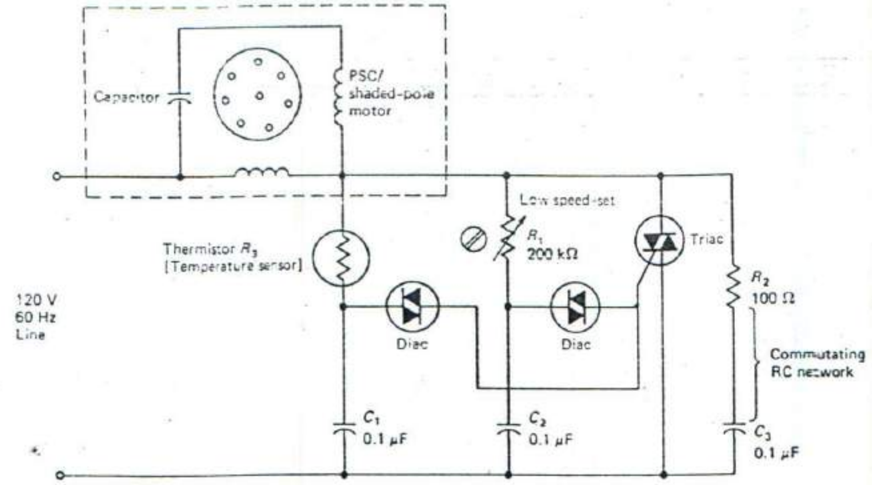


Fig. 3-37 Motor speed controlled by thermistor (temperature).

voltage of the motors may have a rate of rise $\Delta v/\Delta t$ which could retrigger the triac. The commutating $\Delta v/\Delta t$, shown in Fig. 3-36(*b*), can be limited by use of the *RC* network across the device as indicated in Fig. 3-36(*a*).

Figure 3-37 illustrates how this type of circuit could be easily modified to include a feedback signal (closed loop operation) by the addition of a temperature-sensitive resistor (*thermistor*) along with a capacitor and a diac. In this circuit, the thermistor R_3 acts in response to air temperature to control the power to the fan motor. Resistor R_1 and its phase-control network set the minimum fan (blower) speed to provide continuous air circulation and to maintain proper bearing lubrication.

REVIEW QUESTIONS

27. The induction motor speed equation indicates that the speed control terms are the ______ and the number of poles.

28. A split-phase motor with 4 poles runs at 1750 rpm. With 6 poles its approximate speed is ______ rpm.

29. The auxiliary winding in the capacitor-run motor is in series with the ______ winding at reduced speeds.

30. The use of a ______ is preferred to use of a variable resistor for voltage control of a shaded-pole motor.

31. Solid-state frequency changers are also called ______ converters.

32. Besides the inverter, the ______ is a solid-state frequency converter.

3-7
TROUBLESHOOTING AND MAINTENANCE

Preventive maintenance plays a key role in the economics of all industrial installations. In today's highly competitive industry, shutdown of equipment or processes can be very costly. Preventive maintenance, when properly performed, will help prevent troubles before they start. Keeping good records of all maintenance done will establish priorities and give notice of trends. Since motors constitute a major portion of the electrical load in any industrial situation, their maintenance is of extreme importance.

Direct Current Motors

The failure of a dc motor to start may be due to an open in the control circuit, such as a line fuse, low line voltage, or a frozen bearing (this condition will trip the fuses or breakers). Sparking at the brushes may show the need for reseating. The resistance method of controlling the speed of dc motors can also be a source of problems. Resistors can change value, usually because of overheating, so proper ventilation should be confirmed if this type of failure occurs. These resistors, except in the field weakening case, usually decrease the base speed of the motor. The wipers (sliding arm contacts) must be clean and have good contact pressure to ensure proper connections. Some adjustment arms have a small brush-like carbon contact that may need to be seated or replaced. Use the equipment manufacturer's maintenance guide as reference for any adjustments.

The solid-state devices in speed controls may fail by opening or by short-circuiting. The loss of all control in the SCR-type control can be due to a shorted SCR or a UJT that triggers it (if used). The opening of either device will stop the unit from operating or starting. After the power sources and supplies are verified, the solid-state devices can be checked (with power off) by using the methods described in Chapter 2.

The closed loop type of speed control relies on feedback of one sort or another. The tachometer signal, if missing (the line is either short-circuited or opened), would make the system become open loop. With no feedback available, the control would be lost, and the system would run away. In this case, the components and path of the feedback must be checked.

Reduction of control can be due to low voltages, low gain, poor connections, and changes in reference voltage or associated circuits. The mechanical coupling (linkage) in the feedback speed control must be maintained; otherwise, control can be erratic and unpredictable.

The universal motor will run slower on alternating than on direct current so if a rectifier shorts it may cause a drop in speed. The stepper motor will almost always be electronically controlled because of the switching sequence required by its windings. The resistance in series with the power supply affects the slewing (time constant) of the stepper motor. The inductive/resistive (L/R) time constant is inversely proportional to the resistance; therefore, a decrease in resistance will increase the slew speed and vice versa. A large decrease can also cause overheating due to the increase in the holding currents.

In many instances, the logic power is from another source or a separate regulator. These voltages must also be checked. The TTL-type logic families permit only a 250-mV deviation in the supply voltage. Excessive ripple in the power sources can cause false counting or gating of these control circuits and may be suspected in cases of erratic operation. Complementary metallic oxide semiconductor (CMOS) logic is less critical in this respect. Verify power supply parameters early in the troubleshooting process.

Output drivers can fail if they are open or short-circuited. Excessive heat is a major cause of premature failure. All heat sinks must be kept clean, and any air flow restrictions must be corrected. The diodes used for snubbing the inductance of the windings can fail, providing another source of problems. A snubber is a circuit or a component used to absorb the energy produced by the field collapsing in an inductive component.

Stepper Motors

Either a bad driver or its diode (if used) will abort the switching sequence for a stepper motor. In most cases, it will rock back and forth because of the loss of driving torque necessary to overcome the detent torque. In the case of the monolithic or integrated controller, after all input signals have been checked along with the power supplies and the output signals, the chip may have to be replaced. Substitution is sometimes the only practical method available in these cases.

Do not dismantle stepper motors that use a PM rotor. A 5 to 15 percent loss of torque results. Consult the manufacturer's recommendations. A record of failures and maintenance procedures should be kept to alert you to all trends and causes, such as line surges, overheating, and so on. Be sure to replace all shields, fasteners, filters, and grounding straps before returning equipment to service.

Alternating Current Motors

Except small (fractional horsepower) motors, most ac motors are controlled and protected by one or more of the following devices: starters, breakers, fuses, contactors, and thermal overloads. The integrity of these components is the first thing to investigate. If overloads, breakers, or fuses are tripped or blown, you should investigate for overheating, overloading, or loose connections. Interaction with an operator who uses the equipment may yield important information as to the cause of the problem and should not be overlooked or ignored.

Brushless Motors

Modern brushless motors used in the industrial environment eliminate many of the mechanical problems of other types of motors. With no brush or commutation mechanisms, all problems involving the wear associated with these parts are eliminated. The rotors are permanent-magnet and should last indefinitely if not abused. The construction is of a modular form to assist in fast and efficient repair when required. The sensor unit has to be housed in the motor; the commutation logic may or may not be. These devices should be treated like any other semiconductors. The supply voltage to the sensors and logic circuits is critical and must be verified as being within manufacturer's specifications. The sensor and its logic circuits are modular; therefore, replacement with a spare unit is the best solution to a sensor-type problem. The modules can be best serviced at the bench level or returned to the manufacturer for service. As with all solid-state devices, heat is a prime problem. Ventilation systems must be kept clear and clean. A restricted air flow that goes uncorrected will lead to the eventual failure of the replacement part as it did with the original part.

The driver and switching transistors should be tested for short circuits or opens. Remember that there are two switching transistors per phase. Short-circuited transistors may point to overloads, stator coil shorts or grounds, or power supply problems. Power switching devices can be complex, so the manufacturer's manuals are your first source for troubleshooting information. The brushless motor is best serviced by viewing it as a system. The location of a faulty module is achieved by orderly investigation and the process of elimination.

Multispeed Motors

The troubleshooting and maintenance of static inverters are covered in Chapter 8. Speed controls using multispeed motors can encounter problems of loose connections, short circuits, or grounds to the stator coils, especially since multiple windings are involved. The normal testing for these failures is the same as for the other motors, but more elaborate procedures are necessitated by the complexity of the connections. Starters, contactors, and any timers that may be used are also suspect when multispeed motor connections are used. Loose connections are a very common occurrence and should always be checked along with the supply voltage.

REVIEW QUESTIONS

33. Changes in the ______ of the controls for dc motors will affect the speed.

34. If the power source is lost to a disk brake, will the brake remain on or off?

35. If the diode in Fig. 3-3 short-circuits, the transistor and the braking will be ______ all the time.

36. If the grounded end of the bias pots in Fig. 3-9 open, the motors will ______.

37. If C_2 in Fig. 3-16 opens, the hysteresis control will improve. (true or false)

38. Why are the diodes shown inside the symbols of the MOSFET drivers in Fig. 3-23(c)?

39. The brushless motor is usually serviced by replacing a suspect ______.

CHAPTER REVIEW QUESTIONS

3-1. To brake a shunt-wound motor dynamically, the resistor is placed across the field. (true or false)

3-2. Dynamic braking of a series-wound motor requires that the field, armature, and resistor be in a closed ______.

3-3. The ______ motor can be dynamically braked, but with loss of braking power.

3-4. The addition of a ______ in the feedback loop gives linear control of the motor's speed.

3-5. If a tachometer has a K of 20 mV/rpm, at 500 rpm V_g equals ______ V.

3-6. A unidirectional motor control is also called a ______ quadrant control.

3-7. Using resistance control near standstill in a series (universal) motor circuit can cause ______ of the armature.

3-8. ______ resistance control of the series (universal) motor is common when paralleling motors.

3-9. Voltage control of the series motor is the most common method used for speed control. (true or false)

3-10. Series voltage control using a triac is an example of ______ wave control.

3-11. Full-range speed control is usually accomplished with ______ wave-type control.

3-12. All the methods of speed control for the series (universal) motor cause the armature speed to ______.

3-13. The rotor and stator on the PM-hybrid stepper motor are ______ to give a high detent torque.

3-14. What method is used to increase torque in a stepper motor?

3-15. With an 8:1 step-down a 256-rpm stepper ratio, output would be ______.

3-16. Common center-tapped windings are also known as ______ windings.

3-17. The eight-step sequence of a four-winding stepper motor is known as a ______ step.

3-18. The ______ frequency is the prime speed-determining factor of a stepper motor.

3-19. A ______ degree sector would be used with four photodetectors on the rotor positioning mechanism in a brushless motor.

3-20. A three-phase delta stator would require ______ commutations per rotation.

3-21. The three-phase cycloconverter uses at least ______ SCRs.

3-22. The ______ is the solid-state device used for switching alternating current to shaded-pole motors.

3-23. If an induction motor is slowed too much, it may ______.

3-24. A 100 step per revolution stepper motor will take how many steps in the half-step mode for one complete revolution?

3-25. A gear reducer inside a stepper motor will do what to the motor's torque? Its speed?

3-26. Loss of a driver or its signal will cause most stepper motors to ______.

3-27. A three-phase brushless motor will have ______ switching transistors that require testing.

ANSWERS TO REVIEW QUESTIONS

1. friction **2.** released **3.** disk **4.** asbestos **5.** SCR, MOSFET **6.** twice **7.** Fig. 3-4 **8.** increase **9.** small **10.** decrease **11.** open **12.** voltage **13.** commutate **14.** cemf **15.** decrease **16.** locked/stiff **17.** average **18.** magnetic **19.** little **20.** dc **21.** no **22.** permanent magnet **23.** ac **24.** permanent magnet **25.** Hall **26.** false **27.** frequency **28.** 1150 **29.** run **30.** autotransformer **31.** static **32.** cycloconverter **33.** resistance **34.** on **35.** on **36.** rotate constantly **37.** false **38.** they are intrinsic **39.** module

4

BASIC CONTROL DEVICES

This chapter begins with some of the more simple devices of electronics such as switches and fuses, large and small. The wide range of timers from heater types through state of the art electronic types is thoroughly covered. Higher-level switching by relays, contactors, and starters for rotating equipment is presented with actual circuit implications. We will examine all their fundamentals, applications and maintenance requirements.

4-1
SWITCH BASICS

The electric *switch* is a device for making, breaking, or rerouting connections in an electrical circuit. This switching is accomplished by the opening or closing of two metal surfaces. Whichever type is used, the result will be the same. The perfect switch will have 0 Ω resistance when closed, for maximum transfer of power. Conversely, the open switch will have infinite resistance and be capable of withstanding high voltages without arcing. However, any switch tends to arc when it interrupts current flow. The greater the current switched, the hotter the arc will be and the greater the deterioration of the switch contacts. We will look at the types of switches common in today's electronic equipment. With the ever-decreasing size of electronic components it is only natural for switches now to be miniature and subminiature in size. But the current and voltage capacities will also follow the progression: the smaller the switch, the smaller the current and voltage capabilities.

Manual switches are the most widely used today, and each one of us will more than likely operate this type of switch sometime before the day's end. Figure 4-1 shows the basic construction for a manual-type switch. The operator shown in part 4 in Fig. 4-1 can be of a wide variety of styles. A few types of operators are the bat, flat, and baton types. This family of switches is classically referred to as *toggle-type switches*.

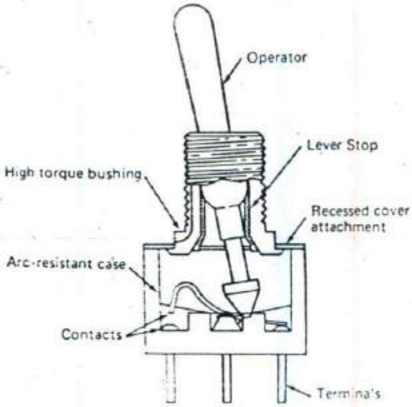


Fig. 4-1 Manual switch construction.

A simple single-pole, single-throw (SPST) switch has a single-pole, single-throw action and is capable of opening or closing a single electric circuit. A double-pole, single-throw (DPST) can open or close two circuits. The DPST switch can be used to open circuits where two "hot" lines are involved, as in 240-V ac equipment. An SPDT has no center position; single-throw switches are either on or off. The three-position toggle switch has two extreme positions plus a center position. The center position is generally off, with the two extreme positions representing on conditions. As illustrated in Fig. 4-2 the B terminal is referred to as the "wiper," or "common" contact of the switch. The center position of the three-position switch is identified by the center OFF notation (c.o.). This notation separates this three-position type of single-pole, double-throw center off (SPDT [c.o.]) switch from a single-pole, double-throw (SPDT).

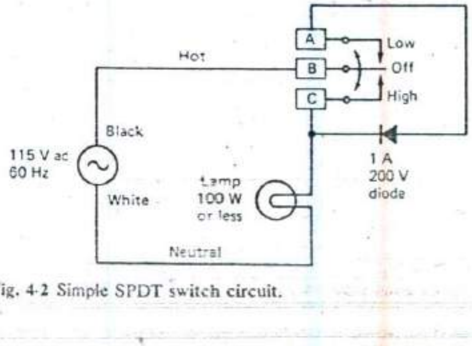


Fig. 4-2 Simple SPDT switch circuit.

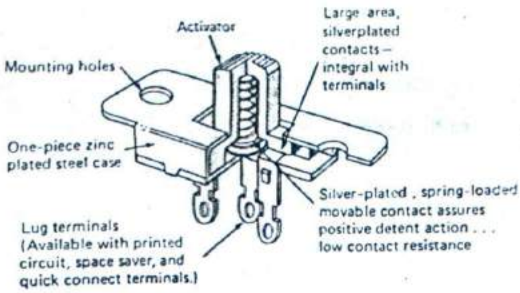


Fig. 4-3 Slide switch construction.

When an operator pushes the button, slide, or lever on a momentary-action switch, the switch contacts move to transfer the circuit to the second set of contacts. When the actuating force is removed, the contacts return to the original position. When an operator actuates a maintained-action switch, the contacts transfer to the second set of contacts. But no change takes place with the contacts when the operator removes the operating force. In some switches the actuator may stay there after removal of the force. Application of a second actuating force is required to return the actuator to its original position. This type of switch is known as *alternate* or *push-push action*.

The slide switch shown in Fig. 4-3 uses a simple slide action to produce the same connections as the toggle switch. Except for the different type of operator action, the switched poles accomplish the same results.

The dual-in-line package (DIP) switch is a miniature form of multipole rocker or slide switch and is made on standard IC socket centers of 0.300 in. The DIP switch can be soldered directly into the printed circuit board (PCB) or placed directly into an IC socket. Most switches are SPST and favor the 14- (seven switches) or 16- (eight switches) pin-package arrangement. Actuation is seldom changed and occurs mainly during installation, testing, and troubleshooting. A ballpoint pen makes a handy actuator for DIP switches. Figure 4-4 shows a 14-pin DIP switch with a typical circuit diagram.

The rotary switch is more indicative of electronic equipment applications. The basic switch structure is shown in Fig. 4-5. The rotary switch is generally a maintained-action switch, but special indexes can be spring-loaded for momentary action. Rotation can be *continuous* (turned through more than one complete circle) or limited to 360° or less. The rotary switch has a variety of positions. If positions are 10° apart, 36 positions are available. The most common angular differences between switch positions are 15, 30, 45, and 90°. One main advantage of the rotary switch is the sequencing ability of its contacts. The rotary switch can also have several banks of switch sections on one shaft, allowing contacts to change simultaneously and in sequence as needed for complex switching applications. Imagine operating an oscilloscope or multimeter without the aid of a rotary switch. Contacts can be short-circuiting or non-short-circuiting. Short-circuiting contacts are also called *make-before-break;* that is, when switching to the next position, the wiper contact will be made to the second position before breaking with the first set of contacts. The non-short-circuiting contacts are called *break-before-make contacts*. When switching from one position to the next, the wiper contacts will break with the first contact before making with the second contact. Applications of the short-circuiting-type contacts include avoiding momentary opening of the feedback elements in a system, dropping out a relay in a control application, and unloading an

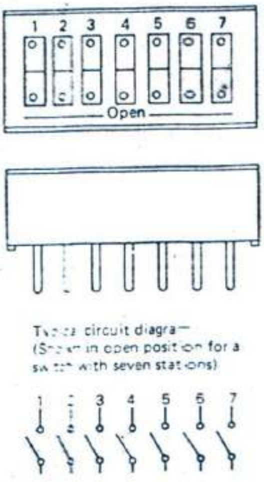
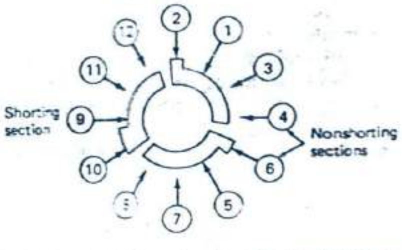


Fig. 4-4 Dip switch (14-pin).



Wafer, 3-pole 3-circuit with 2 nonshorting and 1 shorting moving contacts

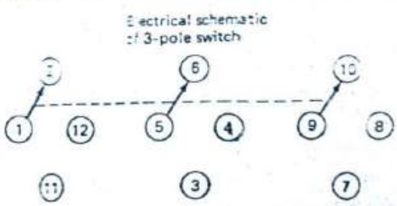


Fig. 4-5 Nomenclature of a standard rotary switch.

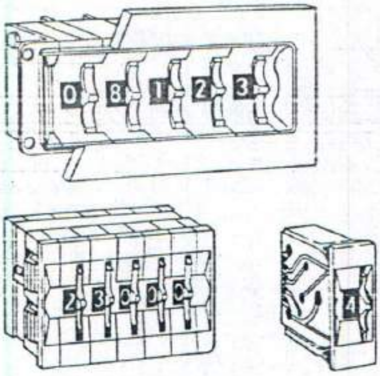


Table 1
10-Position Decimal

Dial Position	Common to:									
	0	1	2	3	4	5	6	7	8	9
0	●									
1		●								
2			●							
3				●						
4					●					
5						●				
6							●			
7								●		
8									●	
9										●

Table 2
8-Position Decimal

Dial Position	Common to:							
	0	1	2	3	4	5	6	7
0	●							
1		●						
2			●					
3				●				
4					●			
5						●		
6							●	
7								●

Table 3
10-Position BCD

Dial Position	Common to:			
	1	2	4	8
0				
1	●			
2		●		
3	●	●		
4			●	
5	●		●	
6		●	●	
7	●	●	●	
8				●
9	●			●

Table 4
10-Position BCD Complement

Dial Position	Common to:			
	$\bar{1}$	$\bar{2}$	$\bar{4}$	$\bar{8}$
0	●	●	●	●
1		●	●	●
2	●		●	●
3			●	●
4	●	●		●
5		●		●
6	●			●
7				●
8	●	●	●	
9		●	●	

Fig. 4-6 Thumbwheel switch and truth tables.

ammeter circuit when switching ranges. Rotary switches are now being made for direct application on printed circuits. They may be smaller than the diameter of a dime, usually have only one deck, and switch low-level signals.

Thumbwheel switches have become more commonplace for today's numerical and computer-control applications. Their switching capability is limited to logic levels, with little or no current. Their specially made decks output binary coded decimal (BCD) (covered in Chapter 11), decimal, or hexadecimal codes (also covered in Chapter 11) necessary to input information for controllers. The switches may also be ganged, as shown in Fig. 4-6.

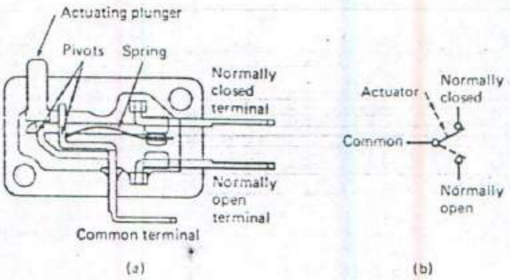


Fig. 4-7 Snap-action switch. *(a)* Operating principle. *(b)* Schematic representation.

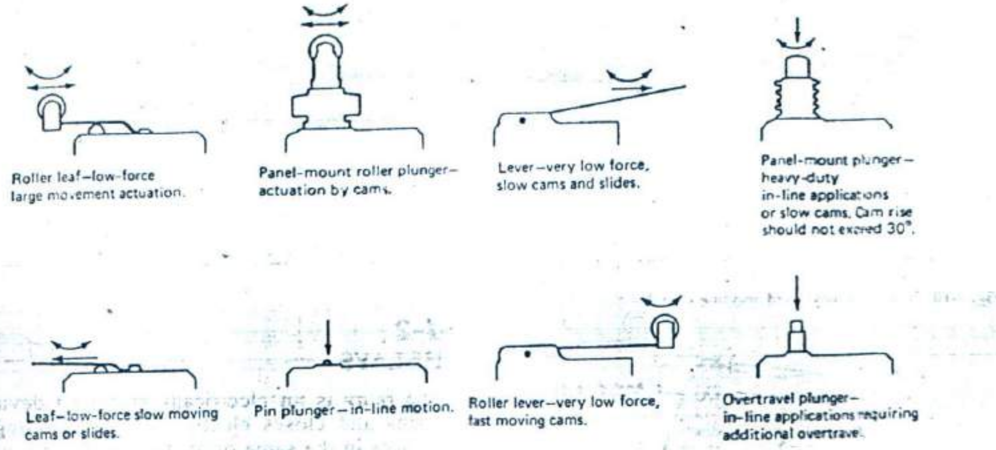
The *position-sensing switch* is typically a snap-acting switch. The switching function is performed by actuation of its input. In Fig. 4-7, the operating principle of the snap-action switch with three electrical contacts is shown in a cutaway view. This type of switch is often called a *microswitch*, after the company that is the world's largest manufacturer of snap-action switches. Figure 4-7 also shows the electrical symbology of this switch in its simplest form, SPDT.

With no force on the plunger, the spring holds the common contact tightly against the normally closed (N.C.) contact. As the plunger is depressed, the force holding the common and N.C. contacts goes through the gap and forces it firmly against the normally open (N.O.) contact. Upon removal of force on the plunger, the N.O. contact's force approaches zero, and the common contact snaps back to its rest, or N.C., position. A few of the actuating methods shown in Fig. 4-8 are the pin, leaf, lever, and plunger. The various methods will fit a variety of different applications.

Snap-action switches boast a low operating force (3 gram-inches [gm · in.] is common), yet ratings of 5 A, 250 V ac are not unusual. The DPDT is the maximum circuit configuration found, but tandem arrangements are possible. This switch is selected not only for its electrical rating, but also its mechanical characteristics, operating force, overtravel, and type of actuator. They are also available in enclosures that can be water- or oil-tight and even explosion-proof. Some environmental needs require these special housings, and nonsparking materials such as nylon may be used for the roller.

A *proximity switch* senses and indicates the presence or absence of an object without requiring physical contact. Simple proximity switches are made up of a sensing head, a receiver circuit for the sense information, and an output circuit including a relay or solid-state switch. The optoelectronic type is covered in Chapter 14.

The *radio frequency (RF) proximity switch* uses an RF oscillator circuit with its coil located in the sensing head. With no metal present, the oscillator output


Roller leaf—low-force large movement actuation.

Panel-mount roller plunger—actuation by cams.

Lever—very low force, slow cams and slides.

Panel-mount plunger—heavy-duty in-line applications or slow cams. Cam rise should not exceed 30°.

Leaf—low-force slow moving cams or slides.

Pin plunger—in-line motion.

Roller lever—very low force, fast moving cams.

Overtravel plunger—in-line applications requiring additional overtravel.

Fig. 4-8 Basic switch actuators.

is normal or logically low. If any metal gets into the field of the sensing head the metal absorbs RF energy. The change in the oscillator is detected, and the logic level goes to the high state. The response can be very rapid as needed for high-speed counting applications. The RF proximity switch can also be used to sense the levels of acid or saline solutions. The operating frequencies are less than 3 megahertz (MHz) and are regulated by the Federal Communications Commission (FCC).

The *ultrasonic beam proximity switch* normally uses two heads: one for transmission and one to receive the signal. The transmitting and receiving heads are matched at a selected resonant frequency above the audio range. The detected output signal is amplified to energize a relay or set a logic device. Placing an object into the field of the beam between transmitter and receiver will lower the output signal amplitude and change the state of the logic or deenergize the relay. The ultrasonic proximity switch overcomes the metal and acidic requirements of the RF proximity switch.

The *Hall-type proximity switch* uses a solid-state structure and requires a magnetic field for actuation. This switch is covered in Chapter 3.

Keyboard switches are of prime importance for operator intervention into a numerical or computerized system. The membrane keyboard is simple and inexpensive and can be sealed. The base may be either the rigid or film type, as shown in Fig. 4-9. The top layer is a flexible polyester membrane, with movable screened contacts. The middle layer is the adhesive spacer that provides the gap between the movable contacts and the stationary switch circuits. The bottom layer is either a polyester film substrate or a PCB, depending on the type of construction preferred. A three-by-four keyboard matrix is popular for touch-tone operations and is indicated as *dual-tone-multifrequency* (DTMF) as required by the phone company.

The 16-position keyboard is often used for hexadecimal input of data or instructions to microprocessors, automatic controllers, and robots. Membrane keyboards are also available in an 8-by-12 matrix configuration and can be interconnected with an encoder (encoders are covered in Chapter 11) integrated circuit to produce the American Standard Code for Information Interchange (ASCII) (covered

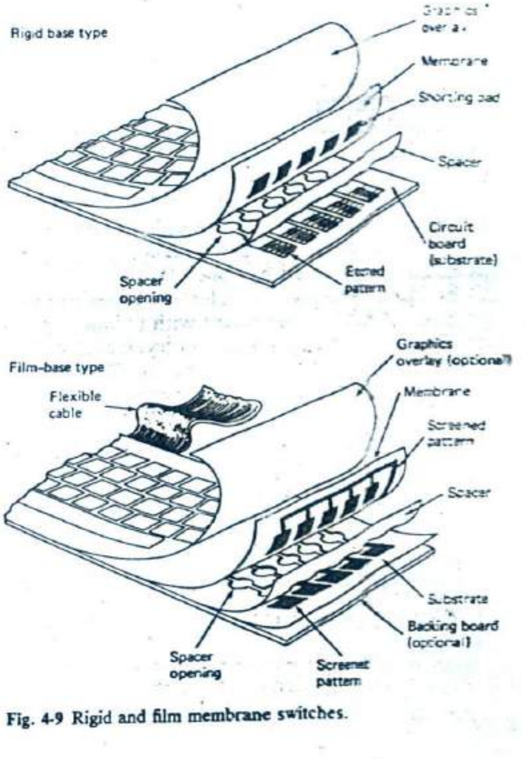


Fig. 4-9 Rigid and film membrane switches.

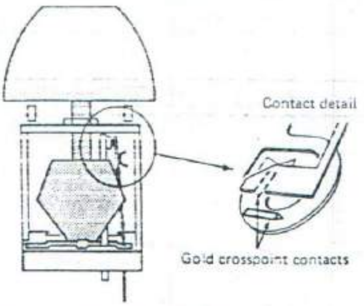


Fig. 4-10 Crosspoint keyboard module switch.

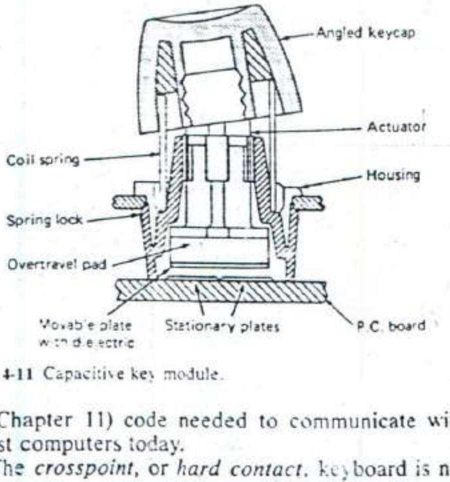


Fig. 4-11 Capacitive key module.

in Chapter 11) code needed to communicate with most computers today.

The *crosspoint*, or *hard contact*, keyboard is not temperature- or humidity-sensitive, is low in cost, and has a low standby power consumption. This type of contact is shown in Fig. 4-10. The switch matrix can be arranged in any of the styles previously mentioned. The contact material is precious metal to provide inertness to chemical action and provides a long life. The moving key action of this type of keyboard is often preferred by touch typists.

Some solid-state keyboards feature capacitive keyswitches to provide a keyboard with unique capabilities. They give high-speed data entry, including key-to-disk, key-to-tape, and other types. With a high-frequency clock (typically 666-kilohertz [kHz] at the input to the key matrix, depression of a key allows capacitive coupling of these pulses to the output of the matrix. Figure 4-11 shows the physical construction of a capacitive key switch. Looking closely at Fig. 4-11, you can see that no physical contact is required for this switch. This contactless style gives very high reliability.

REVIEW QUESTIONS

1. An ideal switch will have _______ ohms when closed and _______ ohms when open.

2. Can an SPDT be used to replace an SPST switch of the same voltage and current ratings?

3. The maximum number of switch positions for a 30° rotary switch is _______.

4. If you do not want to open a circuit momentarily when sequencing with a rotary deck, the switch should be of the _______ variety.

5. A _______ switch is commonly used to input BCD and hexadecimal codes.

6. The _______ action switch is commonly used for position sensing.

4-2
RELAYS

The *relay* is an electrically controlled device that opens and closes electric contacts affecting other devices in the same or another circuit. Basic single-pole relay action is shown in Fig. 4-12 with Fig. 4-12(*a*) being the unenergized state and Fig. 4-12(*b*) the energized state. Without going into the magnetics (to be covered later), when the switch is closed, a north and south pole are produced across the working gap by the solenoid winding of the coil, producing a magnetic flux as shown in Fig. 4-12(*b*). Note that the power source shown is direct current. Operation with alternating current will be discussed shortly. If the current in the coil is slowly increased from zero, a point will be reached where the armature snaps (closing the gap) and closes the normally open (N.O.) contact. This is defined as the *pick-up point* and is specified by current or voltage by the manufacturer. Conversely, if the current in the coil is gradually decreased, a point where the flux is too weak to hold the gap closed is reached and the armature snaps open. This is called the *drop-out current* (or voltage) and is usually considerably less than the pick-up current (or voltage). The only noticeable physical difference between ac and dc relays with the same ratings is the addition of a shaded pole at the end of the core. A shaded pole is a conducting ring that creates a magnetic field that lags the applied field, prohibiting the rapid collapse of the flux across the gap. If this were not the case, the relay would hum or chatter at a 60-Hz rate, and contact positioning would be indeterminate. The shading also causes ac relays to be inherently slower to release than dc relays of the same size. The delay is about 10 ms or longer to prevent contact release at each polarity reversal on 60-Hz operation.

The contacts in an electromagnetic relay make or break the connections in electric circuits. The various contact styles, materials, and ratings that are typical for most relays to be covered in this chapter are presented in Fig. 4-13. High-current contacts are usually of the *single-button* or *bifurcated* style. Noble metal alloys are used for the contact material to reduce oxidation that causes high resistance. The bifurcated contacts have two surfaces for less con-

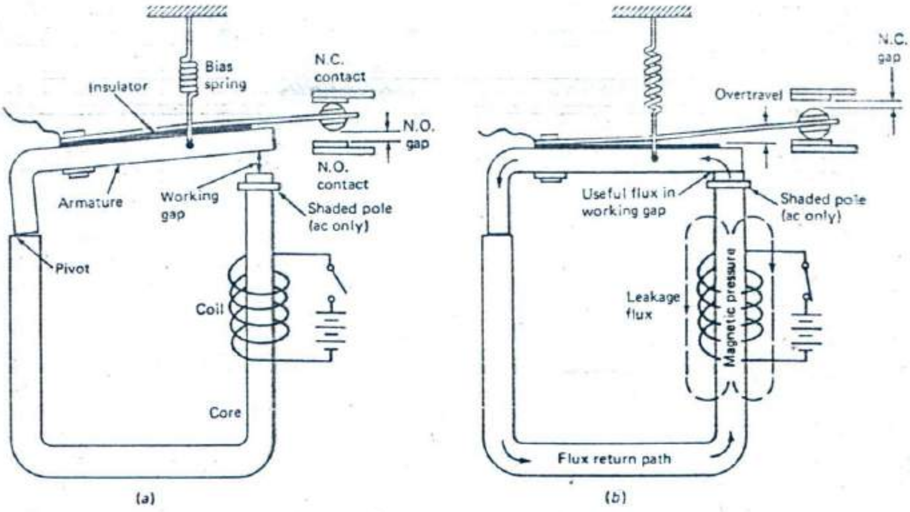


Fig. 4-12 Basic single-pole relay.

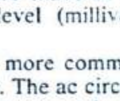
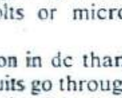
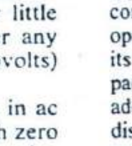
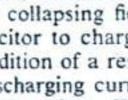
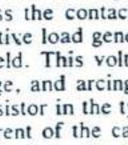
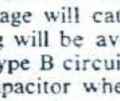
Contact style	Single button	Single button	Bifurcated	Single button	Bifurcated Low Level	Cross Bar Dry Circuit
Resistive load rating @ 28 V dc or 115 V ac	Typ. 7.5 A Max. 10 A Min. 0.500 A	Typ. 5 A Max. 7.5 A Min. 0.500 A	Typ. 5 A Max. 7.5 A Min. 0.500 A	Typ. 2.0 A Max. 3.0 A Min. 0.100 A	Typ. 100 mA Max. 2.0 A Min. 1.0 mA	Typ. 500 μA Max. 250 mA Min. dry circuit
Material	Silver-cadmium oxide	Silver-cadmium oxide	Silver-cadmium oxide	Fine silver and gold diffused	Fine silver and gold diffused	Gold/platinum silver alloy
Inch dimensions	0.125 Dia.	0.100 Dia.	0.100 Dia.	0.078 Dia.	0.062 Dia.	0.017 Dia.
Contact data						

Fig. 4-13 Common relay contacts.

tact resistance at lower actuating force. The crossbar contacts are used for dry circuit (no or little current flow) and are made of gold to hinder any oxidation for low-level (millivolts or microvolts) switching circuits.

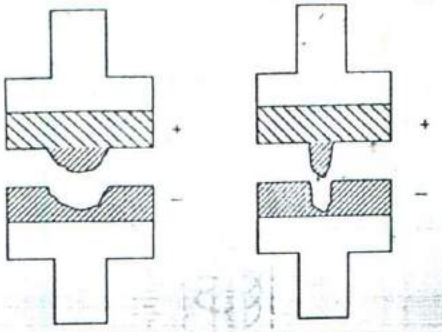
Contact arcing is more common in dc than in ac circuit interruptions. The ac circuits go through zero voltage at each half cycle and extinguish any arcing that occurs. Any metal transfer is generally eliminated, except for a roughening of the contact faces. The dc arcing or spark discharge is damaging and will cause metal to transfer from the negative contact to the positive contact, as shown in Fig. 4-14. The needle-type transfer which occurs at low voltage and low contact force is particularly objectionable, and the needlelike build-up may cause interlocking of the contacts. To prevent this, contact shunting or arc suppression may be used to eliminate or minimize contact arcing when switching inductive loads. The type A circuit shown in Fig. 4-15 uses a capacitor connected across the contacts. When the contacts open, the inductive load generates a voltage due to its collapsing field. This voltage will cause the capacitor to charge, and arcing will be avoided. The addition of a resistor in the type B circuit limits the discharging current of the capacitor when the contacts reclose. The following formulas are applicable for determining *R* (ohms) and *C* (microfarads) for the *RC* suppression network across the contacts:

$$C = \frac{I^2}{10}$$

where

I = circuit current, A
C = capacitance, μF

$$R = \frac{0.1V}{I^x}$$



Dome type transfer (*a*)

Needle type transfer (*b*)

Fig. 4-14 Metal transfer of contacts.

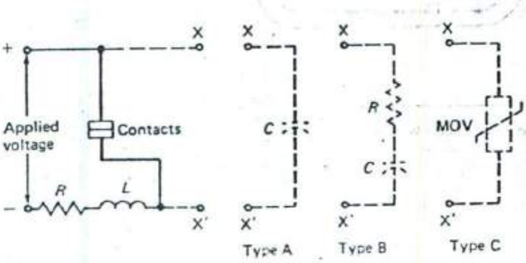


Fig. 4-15 Three types of arc suppression.

where

V = open circuit voltage (induced + supply)

$$X = 1 + \frac{50}{V}$$

Using 0.3 A (I) and 50 V (V), as an example:

$$C = \frac{0.09}{10} = 0.009 \qquad \text{(use 0.01 } \mu\text{F)}$$

$$R = \frac{5}{0.09} = 55.6\ \Omega \qquad \text{(use 56 } \Omega\text{)}$$

The capacitor voltage rating should be greater than the open circuit voltage by at least 20 percent. The use of an oscilloscope will verify operation and permit trimming of the network. The type C network shown in Fig. 4-15 uses a metallic oxide varistor (MOV) for quenching transients (this device is discussed in Chapter 5).

Arcing also causes radio frequency interference (RFI) signals. Because RFI signals can cause problems with the operation of sensitive equipment and instruments, these suppression devices are very important for the role they play in reducing interference. The most effective point of installation is at the source or at the contacts.

Another problem with relays relates to the collapsing magnetic field of the relay coil, generating a transient or cemf voltage expressed as

$$V = \frac{-L di}{dt}\ \text{V}$$

where di is the change in current
dt is the change in time
L is the inductance, in henrys (H)

$$\text{therefore } \frac{di}{dt} = \frac{\text{change in current, A}}{\text{change in time, s}}$$

For example, if the coil inductance (L) is 3 H (typical) and the coil current decreases to zero in 5 ms (dt) when 500 mA (di) was flowing, the induced voltage from the preceding equation is

$$\begin{aligned} V &= \frac{-L di}{dt}\ \text{V} \\ &= \frac{-3 \times 0.5}{0.005} \\ &= \frac{-1.5}{0.5 \times 10^{-3}} \\ &= -300\ \text{V} \end{aligned}$$

If unsuppressed, this voltage will destroy the transistor driver of Fig. 4-16. To minimize this transient, one of three methods shown in Fig. 4-16 can be used. The most effective may be a combination of the types shown. The simple diode solution does reduce transients, but it also delays the drop-out of the coil, and this delay may be undesirable. The diode becomes forward-biased as a result of the induced reverse

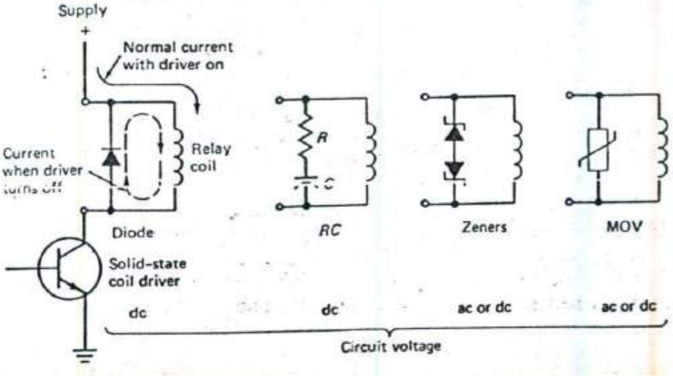


Fig. 4-16 Different types of coil transient suppression.

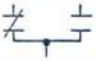
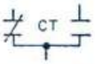
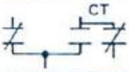
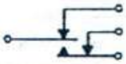
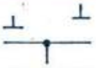
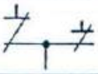
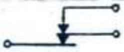
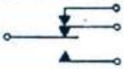
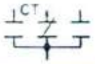
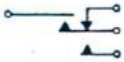
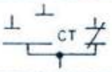
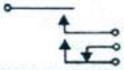
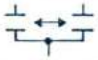
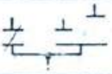
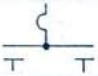
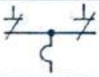
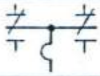
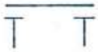
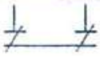
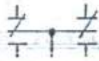
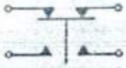
Type A				Type B		
Form	Description	Symbol	Comments			
A	SPNO		Make when energized	A	Make SPST-NO	
B	SPNC		Break when energized	B	Break SPST-NC	
C	SPDT		Transfer when energized	C	Break-make (Transfer) SPDT	
D	SPDT (M-B)		Make before break	D	Make-break (continuity transfer)	
E	SPDT (B-M-B)		Break-make-break	E	Break-make-break	
F	SPST (M-M)		Make-make (sequential)	F	Make-make	
G	SPST (B-B)		Break-break (sequential)	G	Break-break	
H	SPDT (B-B-M)		Break-break-make	H	Break-break-make	
I	SPDT (M-B-M)		Make-break-make	I	Make-break-make	
J	SPDT (M-M-B)		Make-make-break	J	Make-make-break	
K	SPDTNO		Single-pole Double-throw-Center off	K	Center off SPDT	
L	SPDT (B-M-M)		Break-make-make	L	Break-make-make	
U	SPNODM		Make when energized. Includes pigtail.	U	Double make; contact on arm	
V	SPNCDB		Break when energized. Includes pigtail.	V	Double break; contact on arm	
W	SPDT (DB-DM)		Contacts transfer when energized. Includes pigtail.	W	Double break; double make; contact on arm	
X	SPNO (DM)		Make when energized	X	Double make	
Y	SPNC (DB)		Break when energized	Y	Double break	
Z	SPDT (DB-DM)		Transfer when energized	Z	Double make; double break SPDT-DB	

Fig. 4-17 Basic relay contact forms.

voltage of the coil as shown by the previous equation, and the diode remains on until the induced voltage drops to less than 0.6 V (for silicon diodes). The zener diodes and the MOV types shown in Fig. 4-16 preclude this delaying action, by stopping conduction at their higher breakdown potentials, which is far sooner than a diode will. Their breakdown voltage must be greater than the coil operating voltage, but less than the transistor's breakdown voltage.

The basic relay contact forms, of which 18 exist, are shown in Fig. 4-17 (p. 77). The illustration shows that there are two ways to represent relay contacts schematically. Type A and type B are exactly the same electrically. The form identifications are commonly used by various federal, military, and industrial agencies, and familiarity with both types of symbols is necessary. Forms A, B, and C are most commonly encountered in typical relay circuits, as well as in the switches previously discussed as SPST and SPDT types, although form designation is more common for relay contacts.

Figure 4-18 shows contact action that can be observed by an oscilloscope and gives some standard terms used to describe relay action. The contact bounce on opening or closing can create arcing and associated problems. Bounce usually is due to improper armature seating, dirty or oily armatures, or excess ripple in the applied coil voltage.

Relays with misadjusted residual gap screws may suffer contact bounce as well. These screws are shown in Fig. 4-19. They are factory-set, and adjustment in the field should not be attempted. The contact chatter shown in Fig. 4-18 can lead to RFI and may indicate contact spring vibration or that the ac performance of the coil is inadequate.

The general purpose relay shown in Fig. 4-19 typifies a very broad category of relays that finds wide application. They can be designated as almost any type, depending upon the industry in which they are being used. A quick look through any manufacturer's relay catalog will verify this. A general purpose power relay uses the button-type contacts shown in Fig. 4-13 and is capable of switching up to 25 A. This type of power relay switching is common for power supplies, small motors, lights, and auxiliary circuits. A typical circuit for small motor reversing is shown in Fig. 4-20. The closure of S_1 in Fig. 4-20(*a*), or Fig. 4-20(*b*) will reverse the dc motor's direction of rotation. A very subtle situation can be observed by comparing these circuits. In Fig. 4-20(*a*), slow operation of either set of contacts (one contact is still made when the other transfers) will short-circuit the power source and will cause a catastrophic failure. This condition cannot occur in Fig. 4-20(*b*), and this is the preferred circuit.

There are over 20 types of primary relays used today in industry and their basic mechanical structure is similar to that shown in Fig. 4-19. A telephone-type relay (though not confined to the telephone industry) has a long coil compared to its diameter. Both ac and dc types are available, with a

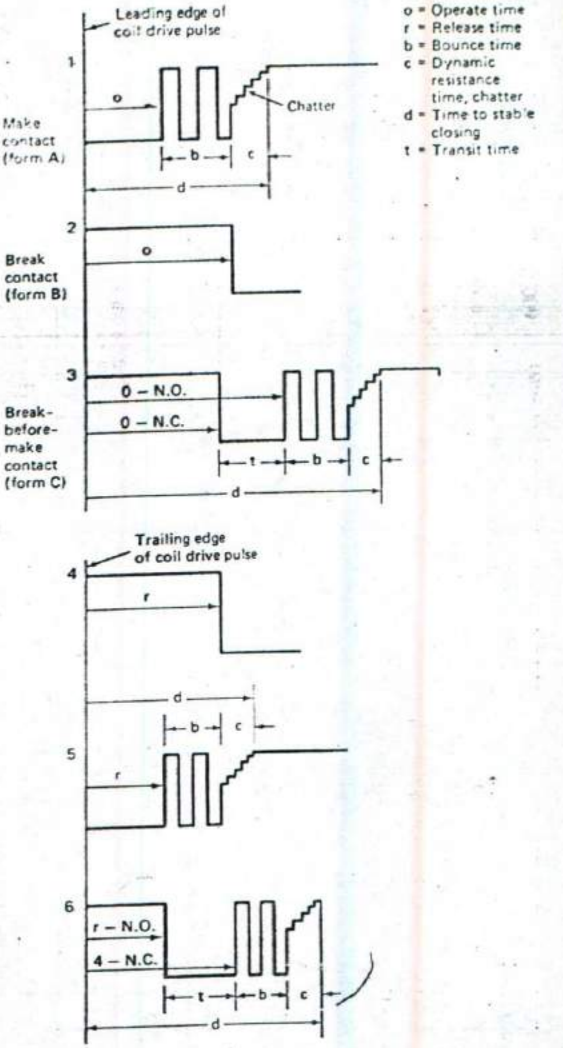


Fig. 4-18 Typical contact response as observed on an oscilloscope.

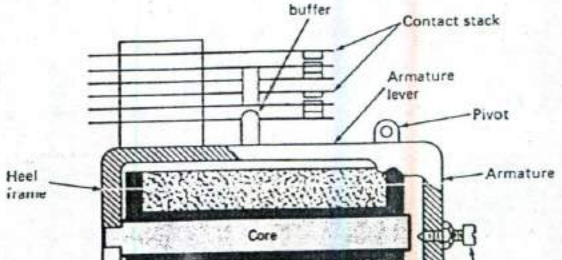


Fig. 4-19 General purpose relay.

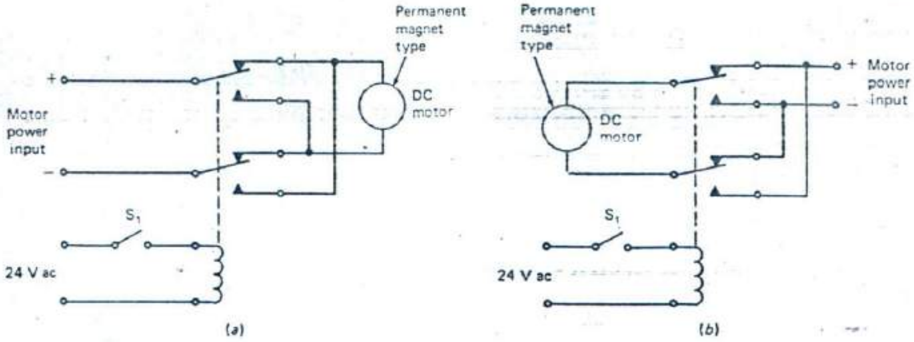


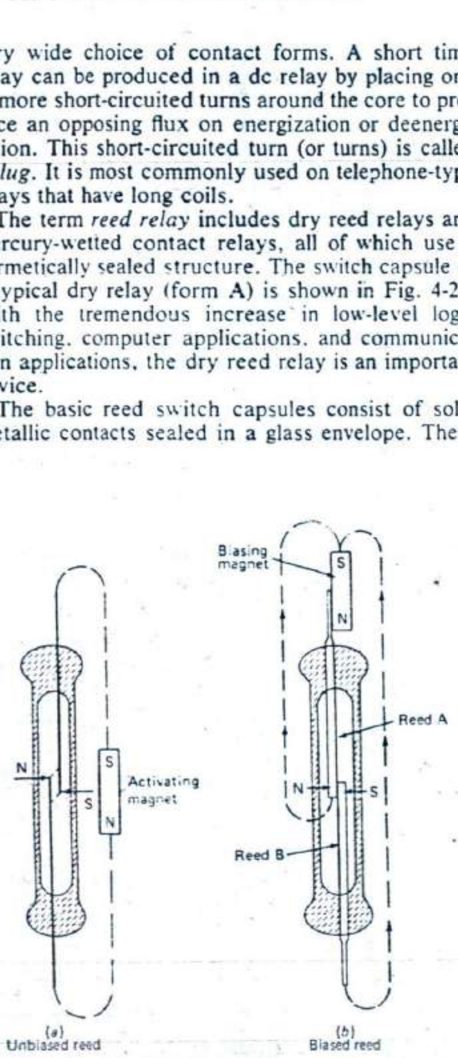
Fig. 4-20 Two different circuits for same function.

very wide choice of contact forms. A short time delay can be produced in a dc relay by placing one or more short-circuited turns around the core to produce an opposing flux on energization or deenergization. This short-circuited turn (or turns) is called a *slug*. It is most commonly used on telephone-type relays that have long coils.

The term *reed relay* includes dry reed relays and mercury-wetted contact relays, all of which use a hermetically sealed structure. The switch capsule of a typical dry relay (form A) is shown in Fig. 4-21. With the tremendous increase in low-level logic switching, computer applications, and communication applications, the dry reed relay is an important device.

The basic reed switch capsules consist of solid metallic contacts sealed in a glass envelope. These flattened ferromagnetic reeds are sealed at each end of the capsule. The reeds are separated by an air gap and overlap inside the tube. The glass tube may contain an inert gas, such as nitrogen, or have a high vacuum for high-voltage switching. The contacts can withstand adverse conditions of humidity, salt spray, and high altitudes. When the capsule is surrounded by a electromagnetic coil of sufficient flux density or is exposed to a magnetic field, the extreme ends of the reeds assume opposite polarity, as shown in Fig. 4-21(*a*). The attraction forces of the opposing magnetic poles overcome the reeds' stiffness, causing them to move toward each other and close. Removal of the magnetic field will return the reeds to their open position. This action can be done at high speed and repeated millions of times. The capsule may contain form B or form C contacts (normally closed) with the normally closed side-biased magnetically with a small magnet, as shown in Fig. 4-21(*b*). The reeds may also be constructed to achieve a normally closed position mechanically. The first reed relays to appear had to be hand-wound, but today they are available with up to 4 poles and with standard coil voltages of 5, 6, 12, and 24 V. Electrostatic shielding is incorporated around the reed capsules to reduce pick-up and cross-talk. These relays are now available in DIP packages for standard IC sockets. High-voltage techniques permit some of them to withstand kilovolts and to have closed-contact resistances of 10 Ω or less.

The original mercury-wetted reed relays were formed with a movable reed having its base in a pool of mercury and the end arranged to move between two sets of stationary contacts. They had many disadvantages such as poor resistance to shock and vibration, and they had to be operated in a vertical position. Recent developments in mercury-wetted contacts have overcome these disadvantages. These relays offer, as their principal advantages, no contact bounce and low, stable contact resistance. Follow Fig. 4-22 as high-speed pictures 1 through 5 show the mercury-wetted contact action for an SPDT bridging (make-before-break) form C. The mercury is shown in black.



(*a*) Unbiased reed (*b*) Biased reed

Fig. 4-21 Magnetic reed switches.

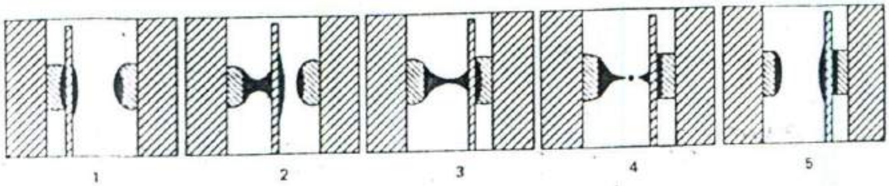


Fig. 4-22 Mercury wetted reed contact action.

The mercury plunger relay has contacts that are hermetically sealed in a glass or metal envelope inside a coil. Both normally open and normally closed contacts are available. These relays can handle up to 50 A. The sealed structure ensures that the contacts are free from dust and air-borne contaminants. This type of relay is position-sensitive and is not useful in vibration environments.

The mercury displacement relay (MDR) shown in Fig. 4-23 provides a self-renewing contact with a single moving part. There are no pins and hinges as in conventional relays, and the liquid mercury provides a new contact surface with every actuation. The MDR is very quiet, considering the current it will handle. (Note: If an MDR breaks, use extreme caution for clean-up because mercury vapor is highly toxic, and its disposal must follow accepted safe practices.)

Latching relays are available in two types: magnetic latching and mechanical latching. They may be of the power, telephone, or even the reed type. Magnetic latching usually employs permanent magnets to make the contacts bistable. Thus, after coil power is removed, the armature is magnetically held in the actuated position. Reset action is accomplished by applying either a voltage of proper polarity to a separate reset coil or a reverse voltage to the actuating coil. Mechanical latching relays are of two basic styles: mechanical reset and electric reset. The *mechanical reset* type employs a coil and armature with a reset mechanism. The mechanism locks the armature in the operated position after the coil has been deenergized. Reset is accomplished by manually tripping the locking mechanism. *Electric reset* types employ a second coil and armature to trip the latch mechanism for resetting to the original position. Some vendors call the latching relay an *impulse*, *alternating*, or *flip-flop* relay for obvious reasons. Solid-state relays will be covered in Chapter 14 on optoelectronics.

A few special relays, though infrequently encountered, are worth mentioning. The *RF*, or *coaxial*, relay is designed to attach directly to coaxial cables and match the cable impedance of 50 to 75 Ω. Radio frequency (RF) losses are minimized by the use of glass silicone and melamine for insulation.

Sensitive relays require only small amounts of power to operate (usually less than 100 mW). They can employ reeds or dry circuit cross-bar contacts. Low-voltage relays with coil resistances around 1kΩ can be directly driven by MOS, TTL, or HTL devices (covered in Chapter 11) without the need for buffers.

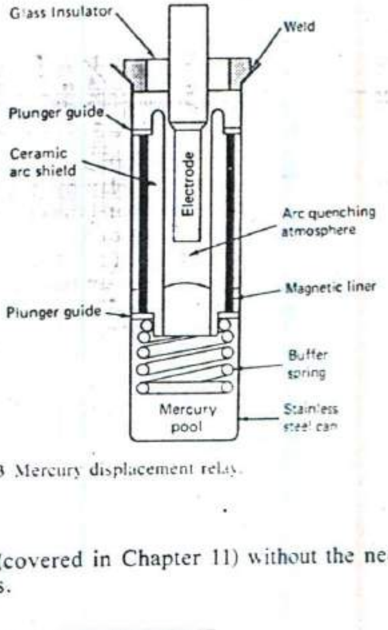


Fig. 4-23 Mercury displacement relay.

REVIEW QUESTIONS

7. Magnetic action of the armature closes the contacts in Fig. 4-12, and the ______ opens the contacts.

8. Is pole shading done on ac or dc relays?

9. Is arcing of contacts more predominant in ac or dc load circuits?

10. Metal transfer not only degrades contacts but can cause them to ______ closed.

11. Which coil suppressors shown in Fig. 4-16 should not be used in circuit Fig. 4-20(*a*) or 4-20(*b*)?

12. By magnetically biasing the reed in Fig. 4-21(*b*), form ______ contacts are realized.

4-3
TIMERS

We will now look at time delay relays (TDRs). Electrothermal or thermostatic delay relays rely on a heating element to provide a temperature differential for mechanical expansion to open or close a contact.

Originally built to a standard tube-type base, they are now available in miniature versions for PCB applications. Though in use for many decades, they still offer several advantages, including operation on alternating or direct current, small size and weight, simple construction, low cost, and long life. A big advantage or disadvantage is the recycle time. The cooling time can be five to ten times the actuating time, which renders them undesirable for use in some circuits. Delays of 2 to 300 s are very common, and some models are adjustable. Time delay relays can also serve as flashers. In a flasher service, the heater circuit is opened with the load circuit, and a 30 percent duty cycle is typical.

The electropneumatic form of TDR (also called the *dashpot time delay relay*) consists of a relay which has a plunger connected to a dashpot. The *dashpot* is a device that employs either a gas (such as air) or liquid to absorb energy and retard the moving parts so as to produce a time delay. When the coil is energized, the solenoid plunger pulls out the dashpot piston at an adjustable rate. At the end of its travel, the plunger completes the magnetic circuit and snaps the contacts closed. The contacts are held by full magnetic power until the coil is deenergized; then drop-out is instantaneous. The gas or liquid used for delaying is usually reused in a closed loop, thereby eliminating outside contaminants.

The addition of a copper collar slug on the telephone-type relay previously discussed delays the energization and sustains deenergization of this type of relay. For maximum pick-up delay, the slug is placed on the armature end of the core. Pull-in delays up to 150 ms and drop-out delays of 500 ms can be obtained. The electromechanical type of TDR is shown in its basic form in Fig. 4-24. A small synchronous ac motor drives a cam through an electromagnetic clutch and gear train. When the cam has rotated far enough, it operates a switch which controls the load. As long as the power-line frequency is constant, the time rotation of the cam will be dependent only on the power-line frequency (not voltage), thereby making the synchronous-motor TDR one of the most stable types. The delay range of the synchronous TDR is of the order of 50 to 1, for a given gear ratio. Delays of 10 s to several days can be obtained, with resolution on the order of 2 percent.

Adjustability is achieved by a mechanism that allows the user to rotate the switch with respect to the cam. Reset is accomplished by releasing the clutch, so a spring mechanism can return the cam to its initial position, and requires 100 ms or more. Some units have automatic reset and will recycle if a sustained circuit is used or wait for another start to begin another cycle. This type is known as the *interval timer*.

The *multicam timer* is an extension of the synchronous-motor TDR, with or without the clutch, but has no reset. A *cam* is a set of circular plates mounted on the shaft coupled to the motor by a gear assembly. Each cam is independently adjustable for

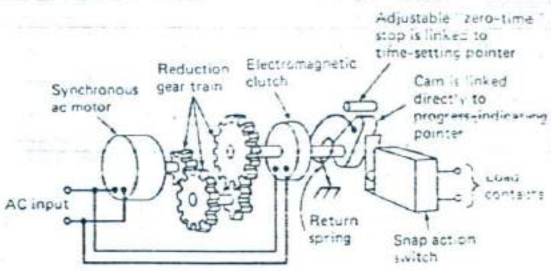
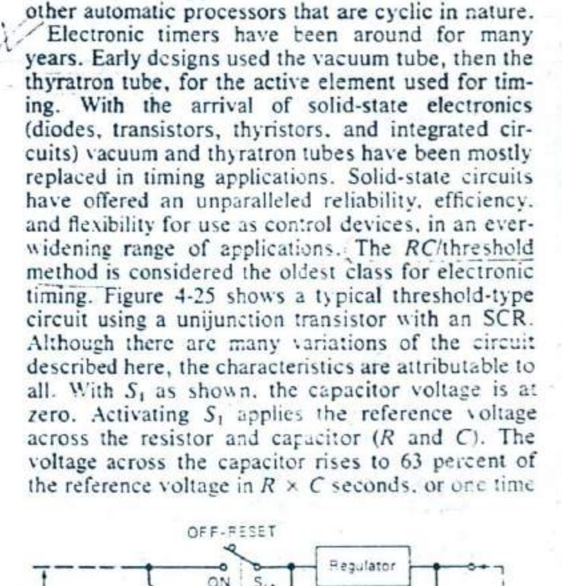


Fig. 4-24 Electromechanical time-delay relay.

2 to 98 percent of the total cycle time. Total revolution time is determined by motor speed and by the gear train assembly. A wide variety of models are available to suit a wide range of applications. A most common type with a push-to-start switch and a friction clutch is found in clothes dryers, dishwashers, and industrial applications such as chemical and other automatic processors that are cyclic in nature.

Electronic timers have been around for many years. Early designs used the vacuum tube, then the thyratron tube, for the active element used for timing. With the arrival of solid-state electronics (diodes, transistors, thyristors, and integrated circuits) vacuum and thyratron tubes have been mostly replaced in timing applications. Solid-state circuits have offered an unparalleled reliability, efficiency, and flexibility for use as control devices, in an ever-widening range of applications. The *RC*/threshold method is considered the oldest class for electronic timing. Figure 4-25 shows a typical threshold-type circuit using a unijunction transistor with an SCR. Although there are many variations of the circuit described here, the characteristics are attributable to all. With S_1 as shown, the capacitor voltage is at zero. Activating S_1 applies the reference voltage across the resistor and capacitor (R and C). The voltage across the capacitor rises to 63 percent of the reference voltage in $R \times C$ seconds, or one time



When $V_C > V_{threshold}$, UJT conducts and triggers SCR

Fig. 4-25 Typical UJT threshold-type circuit.

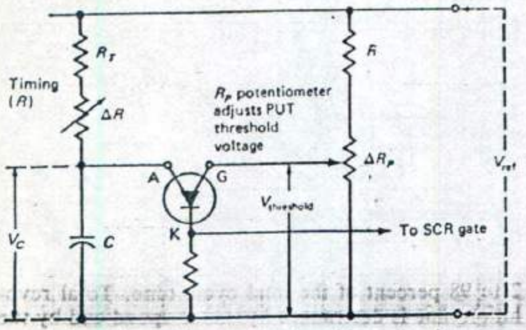


Fig. 4-26 PUT threshold timer circuit.

constant. This product is

$$T \text{ seconds} = R \text{ ohms} \times C \text{ farads}$$
$$1 \text{ s} = 1\ \Omega \times 1\ \text{F}$$

If we wait longer, the voltage will continue to rise higher: to 86 percent at 2*RC* and to 95 percent at 3*RC*. The curve of voltage against time is not linear but exponential. The larger the time constant, the longer it will take to reach the threshold voltage. Also, the closer the reference voltage is to the threshold voltage, the longer it takes to reach it. Let's look at some numbers. If the threshold voltage were 63 percent of the reference voltage, then one *RC* time constant would be required. If we use $R = 1\ \text{M}\Omega$ and $C = 1\ \mu\text{F}$ then the time is

$$\begin{aligned} T &= RC \\ &= (1 \times 10^{6}) \times (1 \times 10^{-6}) \\ &= 1 \text{ s} \end{aligned}$$

What if we want 60 s (1 min)? Then *RC* must be increased 60 times. A 60-MΩ resistor is not practical because of the loading of the UJT on the charging circuit. If we need 1 h (60 min or 3600 s), we must increase *C* to 3600 μF. Such capacitors are large and expensive and have high leakage. We would have to use electrolytic-type capacitors, which vary in value and leakage with temperature changes. With aging, capacitors and resistors change value and drift. All in all, it is not unreasonable to expect a "worst-case" error of 30 percent in *RC*/threshold designs. The unijunction in Fig. 4-25 could be replaced by a programmable unijunction transistor (PUT), permitting a variation of the threshold voltage as shown in Fig. 4-26. About 90 percent of *RC*/threshold timers use either a PUT or UJT.

In the early 1970s, the integrated circuit (IC) timer was introduced; it still enjoys popularity among designers. This timer's success can be attributed to several inherent characteristics; versatility, stability, and low cost. The simplicity of the timer, along with its ability to produce long time delays, has lured designers from using mechanical timers and discrete circuits and into the ranks of IC timer users. The timer consists of two voltage comparators, a bistable flip-flop, a discharge transistor, and a resistor divider. To understand the basic concept, examine Fig. 4-27. This circuit shows the time delay mode, where the output changes state at some design time after a trigger signal is received. An internal resistive divider sets the two comparator levels. Since the resistors are of equal value, the threshold comparator is referenced at two-thirds of the supply voltage and the trigger comparator at one-third of the supply voltage. When the trigger voltage is moved below one-third of the supply, the trigger comparator changes state and sets the flip-flop, driving the output to a high state. An external capacitor can be connected to either the threshold comparator or the trigger comparator, depending on the desired mode of operation. Figure 4-27 shows the IC configured for time delay operation (note that the capacitor is connected to

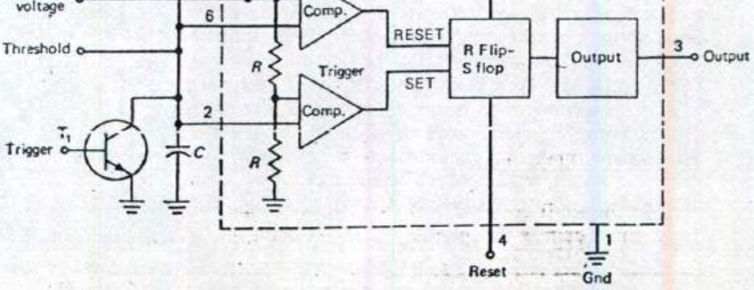


Fig. 4-27 Integrated circuit timer block diagram.

both comparators). The internal discharge transistor is not used in the time delay mode. The operation begins by turning on the external trigger transistor, which grounds the capacitor. The trigger comparator now sees a low state and sets the timer output high. When the transistor is turned off the capacitor begins charging; when it reaches the threshold (two-thirds V_{cc}) the threshold comparator trips and the output switches low and remains low until the external trigger transistor is turned on again. Thus time delay is accomplished by having the output go high, and, after a prescribed time interval, the output goes low.

A few details about a standard IC timer, the 555, are worth mentioning. Because of its circuitry, the 555 IC will only trigger on a negative-going edge of the input pulse. It is necessary that the trigger input be returned to a voltage greater than $1/3V_{cc}$ or that ac coupling into the trigger input be used. The control voltage is brought in at the two-thirds point of the reference divider. Imposing an external voltage at this point can vary the timing. This feature makes it possible to use the timer as a voltage-controlled oscillator or pulse-width modulator. When this control feature is not used, the control voltage input should be bypassed to ground with a ceramic capacitor. The reset overrides all other functions of this timer.

With the arrival of large-scale ICs, mainly because of metallic oxide semiconductor (MOS) technology improvements, IC timers have been improved along with other integrated circuits. The programmable timer/counter of Fig. 4-28 is a typical example. The "programmable" feature increases its value for computer-controlled applications as well as others. The timing cycle is initiated by applying a positive-going trigger pulse to pin 11. This trigger activates the time-base oscillator, enables the counter section, and sets all counter outputs to the "low" state. The time-base oscillator generates pulses with a period equal to RC. These clock pulses are counted by the binary counter section. The timing cycle is completed when a positive-going reset pulse is applied to pin 10. The timing sequence of output waveforms at various circuit terminals after a trigger signal is shown in Fig. 4-28. When in the reset state, both the time-base and the counter sections are disabled, and all counter outputs are at "high" state. In most timing applications, one or more of the counter's outputs are connected back to the reset terminal (S_1 closed in Fig. 4-28). Connected this way, the circuit will start timing when a trigger is applied and will automatically reset itself to complete the timing cycle when a programmed count is complete. If none of the counter outputs is connected back to the reset terminal (S_1 open), the circuit will operate in its astable or free-running mode after an input trigger. The binary counter outputs (pins 1 to 8) are open collector type stages and can be tied together to a common pull-up resistor to form a *wired-or connection*, which simply means that the combined output will be low as long as any one of the outputs is low. This allows the time delay associated with each counter output to be summed by tying them together to a common bus, as shown in Fig. 4-28. For example, if pins 1, 5, and 6 were tied to the output bus, the total time delay would be $(1 + 16 + 32)T = 49T$. In this manner, one can program the timing cycle to be from 1 to $255T$, where $T = RC$.

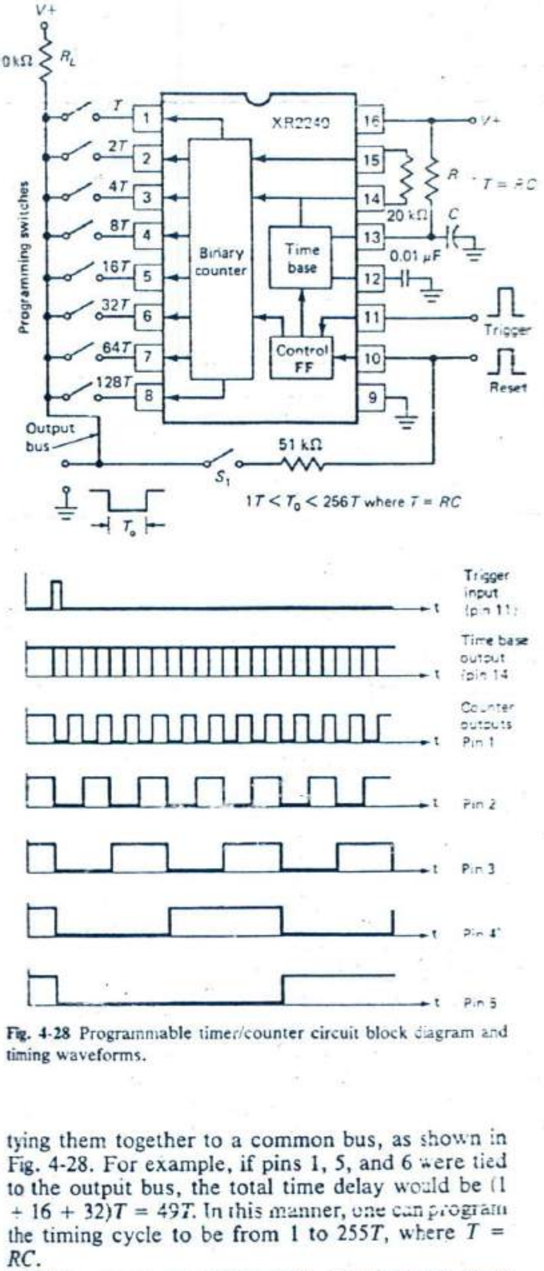


Fig. 4-28 Programmable timer/counter circuit block diagram and timing waveforms.

When power is applied with no trigger or reset inputs, the circuit reverts to its "reset" state. The binary counter outputs are high or in a nonconducting state. Once triggered, the circuit is immune to additional trigger inputs, until the timing cycle is completed or a reset input is applied. During a timing cycle, the outputs change in accordance with the timing diagram of Fig. 4-28. The outputs can be used individually or can be "wired-or" as previously mentioned. The period T of the time-base oscillator can be modulated by applying a dc voltage to pin 13. The time-base oscillator can also be synchronized by applying a sync pulse to pin 12. The time-base period is determined by the external RC network connected to pin 13. The waveform at pin 13 is an exponential ramp with a period $T = RC$. The time-base output (pin 14) is an open collector stage and requires a 20 kΩ pull-up resistor to pin 15 for proper circuit operation. At reset, the time-base output is at high state, and subsequent to triggering, it produces a negative-going pulse train with a period $T = RC$. The time-base output is internally connected to the binary counter and in some applications may be used as an input for externally generated clock pulses.

REVIEW PROBLEMS

13. An electrothermal TDR operates on ac or dc voltage?

14. The dashpot of an electropneumatic relay speeds up actuation. (true or false)

15. Loss of liquid in the dashpot will ______ the delay.

16. To maximize the pick-up delay of a telephone-type relay, the slug is put on the ______ end.

17. What type of timer will wait to start another cycle?

18. What value of resistor is needed for an RC/threshold timer constant of 4 s with a 4 μF capacitor?

4-4

OVERCURRENT PROTECTION

Most electric and electronic circuits need some sort of safety valve for protection from dangerous overloads. The oldest and most common type of protector is the fuse. The common plug-type fuse was originated by Thomas Edison in 1890 and is available at current ratings up to 30 A at 125 V. The high currents that can be present during a short circuit can cause electrical damage to equipment, mechanical damage, and possibly an extremely hazardous fire risk. Therefore, it is extremely important not to bypass, short-circuit, or oversize the replacement of a blown fuse.

The fuse is a device which protects by melting (fusing), thereby opening the circuit in response to an overload or short-circuited current. The typical electronic fuse utilizes silver wire or some other metal link housed in a package that is often cylindrical. All fuses are temperature-sensitive devices and are rated for an ambient temperature of 25°C and must be derated as the temperature increases. This temperature requires good contact connections and air movement to achieve the predicted life.

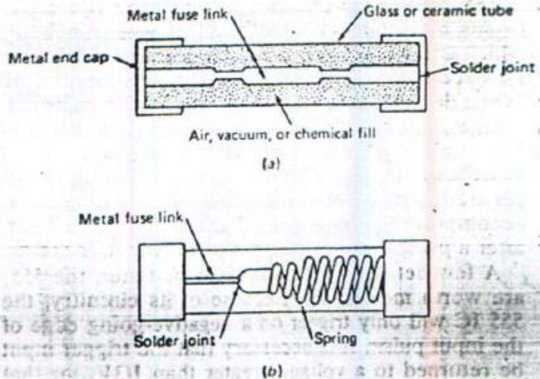


Fig. 4-29 **Fuse construction. (*a*) Fast or medium acting fuse. (*b*) Slow action fuse.**

Fuses have three important specifications—current rating, voltage rating, and fusing characteristics—and are manufactured in a wide variety of styles and shapes. The glass and ceramic tube fuses are popular for protecting electronic equipment and are shown in Fig. 4-29.

The current rating of a fuse is expressed in RMS amperes or fractions of an ampere. Fuse ratings are established by a controlled set of tests referenced by the Underwriters Laboratories, Inc. It is extremely important to remember that fuses are sensitive to current, not voltage. A fuse will not blow (open) until its melting temperature is reached. The current rating is always stamped or otherwise indicated on the fuse body.

The voltage rating of a fuse should be equal to or greater than the circuit voltage. This rating is covered by National Electrical Code (NEC)® regulations and the Underwriters Laboratories (UL) to prevent fire risks. Ratings of 32, 125, and 250 V are common and adequately cover most electronic apparatus. The voltage ratings apply to an interrupting capacity of 10,000 A. This is a large safety factor for the type of short-circuit currents likely to occur in electronic equipment.

The *normal-blo* (standard) fuse is used for resistive or non-surge-type loads and should be replaced with the same type when being renewed. This type of fuse is mainly used for protection against short circuits. The chart in Fig. 4-30 shows the electrical characteristics of normal-blo fuses. Note that the fuses shown can withstand a 10 percent overload for at least 4 h. This style of fuse can have a glass or ceramic body, as shown in Fig. 4-29. Glass fuses give a better visual indication, but the ceramic fuses are more rugged to abuse and will not shatter on extreme overloads.

The second type of fusing characteristic is the *slo-blo*. This type may also be called *time delay* or *dual element*, depending on the manufacturer. These types provide a time delay for overload currents greater than 50 percent. They are useful in protecting surge-type loads such as motors and capacitors. Slow-acting fuses have two blowing mechanisms;

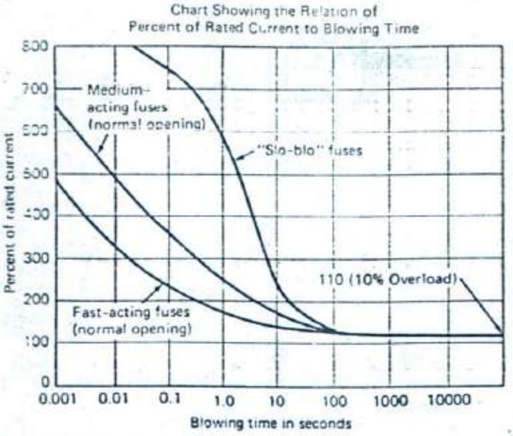


Fig. 4-30 Fusing characteristics.

Fig. 4-29(*b*) shows that a solder joint is under spring tension. A moderate overload will eventually melt the solder, and the joint will pull apart, interrupting the circuit. An extreme overload will melt the metal fuse link and provide a quick response. Compare the difference in Fig. 4-30 for the blowing time between a normal-blo and a slo-blo fuse for the same percentage of overload. For example, the blowing time at 500 percent is 0.01 s for the normal fuse and 2 s for the slo-blo type.

Semiconductor fuses are used to protect diodes, thyristors, and other semiconductors that have little thermal storage capacity and cannot handle heavy overloads for any length of time. This type of fuse is extremely fast-acting compared to normal-blo of the same ratings as indicated in Fig. 4-31, which covers currents up to 10,000 A. It is especially important to use an exact replacement when renewing semiconductor fuses. Note that the chart is a log-log scale so the numbers change rapidly along the time and current axes. The fuses used in some semiconductor circuits may be rated as high as 1000 A. Figure 4-32 shows the typical location of fuses in a solid-state power supply application. Fuses F_1 and F_2 placed in series with semiconductors and the supply lines are a protection against internal faults and overload conditions. Service interruptions cannot be tolerated in large rectifying equipment in some industrial applications such as electrochemical applications. The three-phase bridge of Fig. 4-32 is a usual method of including one or two redundant parallel paths (shown in dashed outline) to maintain uninterrupted operation. If a rectifier fails, the fuse opens, disconnecting the short-circuited rectifier without damage to the equipment or interruption of service. Fuse F_3 is protection against an external short circuit or excessive load currents at the output; F_2 protects against short circuits and overloads.

With international trade more commonplace today, a word about international fuses is appropriate at this time. Most countries have certification by agencies similar to UL here in the United States. The following are acronyms for international, European, and Canadian agencies:

International Electrotechnical Commission (IEC)
Switzerland (SEV)
Verband (West German) (VDE)
Svenska (Sweden) (SEMKO)
Canada (CSA)

The European and Asian fuses are 5 by 20 mm and are not the same as the 1/4 by 1 1/4 in. 3AG fuses that are commonly used in the United States. These fuses are 6.3 by 32 mm. Trying to install a 3AG to renew a 5-mm and vice versa can cause a hazardous situation and should not be attempted. Also note that most foreign fuses are rated at 250 V because that is the primary operating voltage of their main line circuits, just as 120 V is the standard in this country. If foreign equipment is involved, observe fuse sizes and ratings carefully and judiciously.

Circuit breakers are automatic overcurrent protectors designed to open under overload conditions. We will cover the general purpose types applicable to electronic equipment. Circuit breakers are mainly limited to three poles. The simple thermal circuit breaker shown in Fig. 4-33 responds to temperature changes of its internal bimetallic element. This element is in series; if the load current increases to

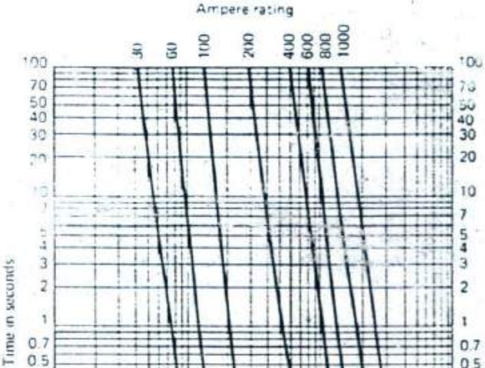


Fig. 4-31 Fuse characteristic curves.

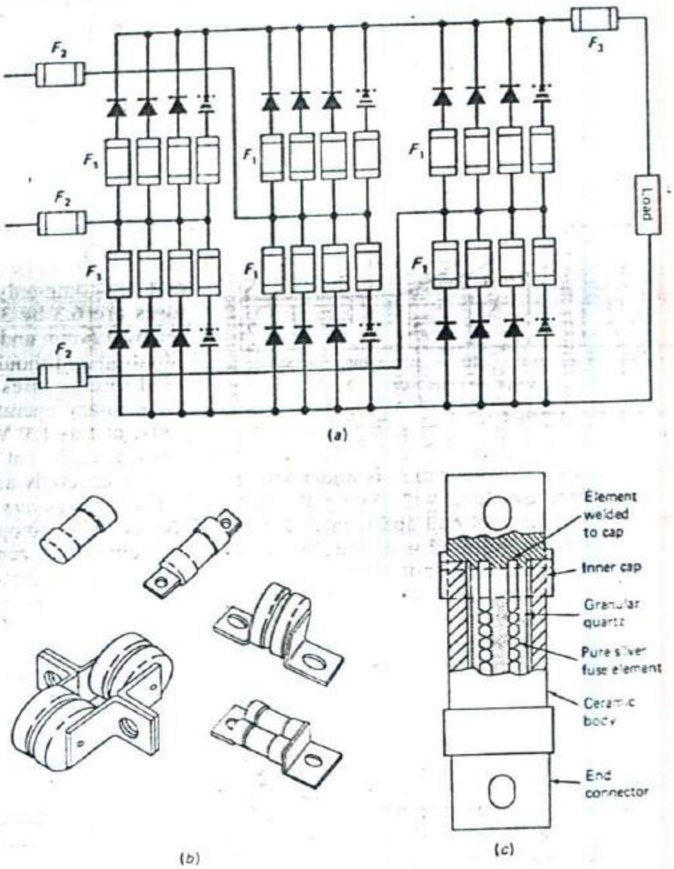


Fig. 4-32 Semiconductor fuses. (*a*) Placement. (*b*) Package. (*c*) Cross-section.

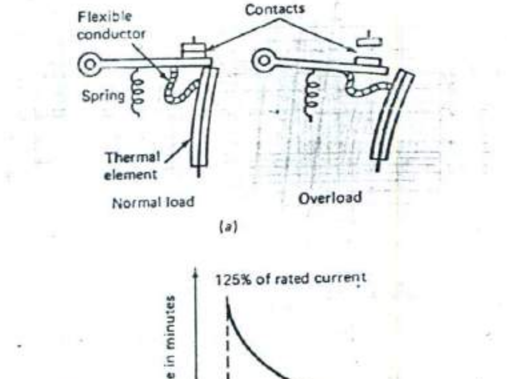


Fig. 4-33 Thermal breaker. (*a*) Action. (*b*) Schematic symbol. (*c*) Characteristic curve.

about 125 percent of the breaker's rating, it bends and the contacts open. To provide additional protection, some types of thermal breakers contain a built-in trip-free mechanism which allows the contacts to open, even if the breaker's lever is manually held reset. The performance curve in Fig. 4-33(*c*) shows that the thermal breaker has no true instantaneous tripping point. This limitation precludes the simple thermal breaker from most applications.

The *thermal-magnetic circuit breaker* and its performance curve are shown in Fig. 4-34. This breaker operates by using the same principle as the simple thermal breaker for moderate overloads. On extreme overloads, the magnetic feature comes into play and takes over to assist in tripping the mechanism. The magnetic field created by the fault current attracts the magnetic plate and opens the contacts. This action provides the thermal-magnetic breaker with an instantaneous trip capability for extreme overloads. The thermal-magnetic types are therefore more desirable for most applications.

The *magnetic-hydraulic circuit breaker* consists of a solenoid with a *dashpot* (piston in fluid). The solenoid coil is wound so that the spring-loaded core

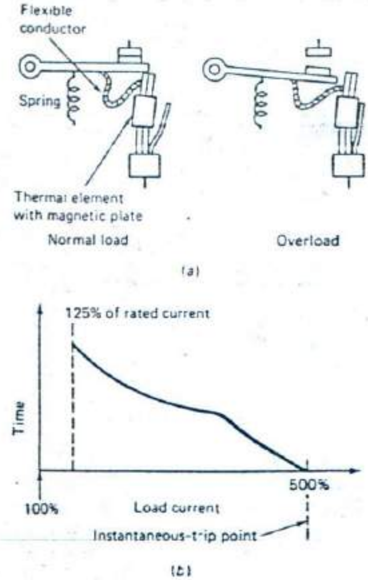


Fig. 4-34 Thermal magnetic breaker. (*a*) Action. (*b*) Characteristic curve.

does not move at normal currents. On heavy overloads (usually ten times the rating), the strength of the magnetic field attracts the armature to a cylinder cap to open the contacts instantaneously. At small, sustained overloads the field of the solenoid pulls the iron plunger into the core. As the core moves into the solenoid, it reduces the gap of the magnetic circuit and in turn increases the magnetic flux. The speed of movement of the piston is determined by the viscosity of the silicone fluid in the dashpot. This style of breaker has a time delay for moderate overloads. If the overload is removed before the piston gets to the top, the spring pushes the piston back to the bottom before the armature can trip the contacts. Since breakers are basically mechanical devices, it is a common practice to manually trip them occasionally to ensure their proper operation.

REVIEW QUESTIONS

19. What other fuse characteristic is important in addition to the current and voltage ratings?

20. Fuse current rating is always given in ______ amperes.

21. Fuses are ______ sensitive devices and therefore require good solid connections.

22. What type of fuse is used for motors or capacitive loads?

23. What is the main criterion of semiconductor fuse usage?

24. What is the major limitation of the simple thermal breaker?

4-5
SOLENOIDS, CONTACTORS, AND STARTERS

A *solenoid* is an electrically energized coil in which the turns of the windings are insulated from each other even if bare wire is used. They are the basis for all electromagnets. We are interested in the solenoid with a moving core called a *plunger*. This device will convert electrical energy into mechanical motion, which may be either linear or rotary. The basic linear solenoids are shown in Fig. 4-35. The plunger is the movable bar of high-permeability steel or soft iron and may be laminated for ac operation to reduce eddy current loss. Referring to Fig. 4-35(*a*), when the coil is energized the plunger becomes magnetized and mutual attraction takes place between the coil and plunger. The traditional solenoid is the iron-clad or box style shown in Fig. 4-35(*b*). The solenoid and plunger are provided with an iron or steel jacket, which is usually laminated for ac operation. The effect of the frame is to increase the magnetic and mechanical forces slightly at the initial position and much more so at or near the final position of the plunger. In the magnetic-cushion type, Fig. 4-35(*b*), the plunger cannot strike the opposite end of the open frame as it passes into the coil, thereby eliminating the hammer blow effect which occurs in iron-clad solenoids. This effect is utilized in electric hammers, but in other applications the constant hammering is detrimental to the solenoid and the noise is objectionable. This effect occurs on the solenoids shown in Fig. 4-35(*c*) and 4-35(*d*). In Fig. 4-35(*c*), the iron-clad solenoid has its plunger travel limited by the closed frame. The type illustrated in Fig. 4-35(*d*) is known as the ***stopped iron-clad solenoid*** because of the iron or steel plug or stop extending downward toward the plunger to increase the pull at or near the final position and also to increase the seated pull.

One of the important factors considered in selecting a solenoid is the work requirement for the solenoid. The force required by the load must not exceed the force developed by the solenoid during any portion of its travel; if it does, the plunger will not pull

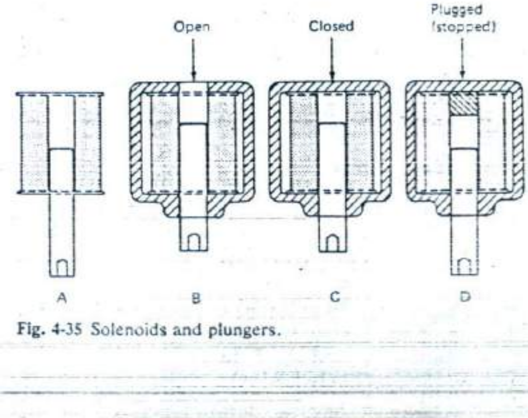


Fig. 4-35 Solenoids and plungers.

Volts—60 Hz	120
Duty	Cont.
Coil resistance (Ohms @ 25°C)	88
Watts seated	9.5
Amps seated	0.24
Amps ⅛"	0.72
Amps ⅜"	1.0
Amps ⅝"	1.22

Fig. 4.36 Sample of solenoid specifications.

in or seat. With an ac solenoid, this could burn out the coil. The coil reactance will not reach its intended maximum because the core (plunger) does not come into position. Excess current will flow in the coil. Never select or replace with an overrated solenoid because it will develop more force than is required by the load. The extra energy must be dissipated by the plunger or field piece on impact, causing premature failure. Proper alignment of plunger and load is also important. Loading which is not centered along the line of plunger travel should be avoided or corrected. If misalignment occurs, plunger wear increases, and with ac solenoids, seating will be impaired. Where an ac laminated plunger is directly linked to a mechanical load, jamming of the plunger could burn out the coil.

With ac solenoids, an inrush current occurs until the plunger is seated. The table in Fig 4-36 gives typical values for an ac solenoid. As can be seen, the longer the stroke the higher the inrush current, and the inrush continues until the plunger is seated. The operating limit of any solenoid is determined by the temperature rating of the insulating materials used in its coil. As long as this maximum is not exceeded, the solenoid's pull characteristics will not be materially degraded. With modern insulation materials, this maximum temperature is usually 110°C. As the coil temperature increases, pull decreases. A widely used rule of thumb is to select a solenoid which will deliver a slightly greater force at 110°C than the load requires for the anticipated coil voltage as shown in Fig. 4-37.

The *duty cycle percentage* is defined as the ratio of solenoid coil ON time to the period of one cycle (ON plus OFF time). For example, if a solenoid is energized for 1 min and deenergized for 4 min the duty cycle will be 20 percent, as follows:

$$\text{Duty cycle, \%} = \frac{\text{ON time} \times 100}{\text{ON time} + \text{OFF time}}$$
$$= \frac{1 \times 100}{1 + 4} = 20\%$$

A majority of intermittent-duty cycles fall into three categories: 50, 25, and 10 percent. Most solenoids operated for up to 30 s ON time fall into the 50 percent category. The 30-s to 3-min range falls into the 25 percent category, and the 3- to 5-min maximum ON time would most likely fall in the 10 percent category.

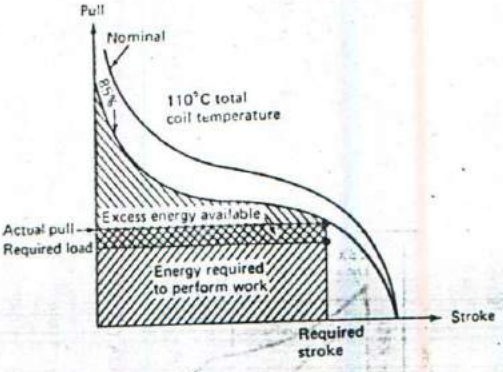


Fig. 4-37 Typical ac solenoid pull-stroke curve.

The continuous-duty solenoid may remain permanently energized at rated voltage without exceeding its maximum temperature rise. A continuous-duty dc solenoid may be subject to continuous ON-OFF cycles without overheating; not so with the ac solenoid because of its inrush current. As Fig. 4-36 shows, high coil current occurs until the plunger seats. The longer the stroke the higher the inrush current. In continuous-duty ac solenoids, subject to ON-OFF cycling, repetitious surges could cause overheating. A reduction in duty cycle or the use of a dc solenoid is necessary. The intermittent-duty solenoid is designed to be energized for a maximum of 5 min without excessive overheating. After 5 min it must be permitted to cool before resuming operation. In some cases thermal cut-out devices are incorporated as a fail-safe measure for excessively high ambient temperatures or overloads which could cause the plunger to stall.

The speed of operation is a function of the applied load and power. An ac solenoid's operating speed will vary, depending on the point of the applied voltage cycle. If a consistent operating speed is required, a dc solenoid should be selected.

Push-type solenoids are usually modified pull types and can use alternating or direct current. They have a nonmagnetic pusher rod projecting through their stops with the plunger protruding out the other end.

We must look at the rotary solenoid with its unique style of operation. Smooth rotary motion is achieved by the use of ball bearings and a varying pitch ball race. The solenoid can turn, step, index, lock, punch, or lift in milliseconds. These units can have right- or left-hand strokes and can weigh up to 5 lb and have starting torques as high as 100 lb · in. for a 25° stroke. In ordinary electromagnets, the magnetic pull increases sharply as the air gap closes, but in the rotary solenoid, this is compensated for by the compound angle of incline of the ball races. The incline of the ball races is steep at the beginning of the rotary stroke and gradually decreases as the balls approach the deep end of the ball races. This arrangement amplifies torque at the start of the rotary stroke, where it is usually needed.

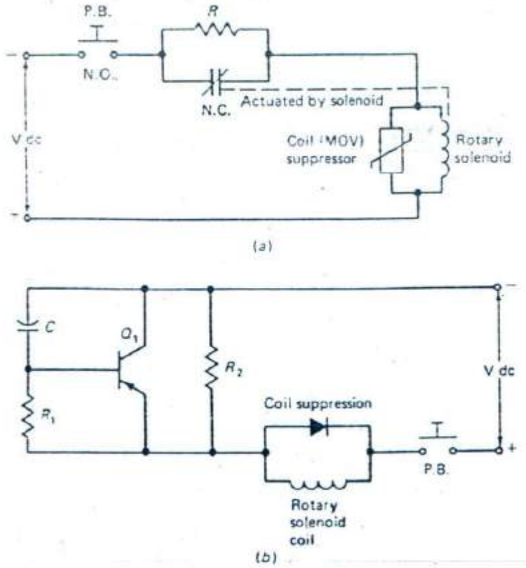


Fig. 4-38 Hold-in current-reduction circuits.

As for all the other electromagnetic devices discussed, the temperature is a very important factor. As the temperature of the coil rises, so does its resistance. Increased resistance reduces current flow with a constant voltage source, consequently decreasing the output torque. Heat can be dissipated by controlling the air flow, by mounting the solenoid on a large surface (heat sinking), or by resorting to duty cycle limiting, which is how they are rated by the manufacturers. One way to decrease the temperature rise in a rotary solenoid is to use a hold-in circuit to reduce current to a point at which torque is sufficient to maintain the solenoid in the energized position. One common method to reduce coil current during hold-in is a normally closed (N.C.) switch in parallel with a hold-in resistor, as shown in Fig. 4-38(*a*). When the pushbutton (PB) closes the circuit, full voltage is impressed across the solenoid coil, resulting from the bypassing of the resistor by the N.C. switch. As the rotary solenoid approaches the end of its stroke, a mechanical connection opens the N.C. contacts. This reduces the solenoid voltage and lowers power dissipation. Figure 4-38(*b*) shows another technique for reducing hold-in current. When the PB switch is first closed, current flows through the base-emitter junction while charging capacitor *C* to the input voltage. This base current turns on the emitter-collector circuit and allows full power to be impressed across the solenoid. The hold-in continuous current is supplied through R_2 thus limiting dissipation in the solenoid. When switch PB is released, the solenoid is released and the capacitor is discharged by R_1 and R_2 to prepare the circuit for the next activation cycle.

TABLE 4-1 CONTRACTOR RATINGS

Motor Rating* 600 V	Resistive Rating* 600 V	Horsepower* 120 V
30 A	40 A	1.5 (1φ2P)
		3 (3φ3P)
40 A	50 A	2 (1φ2P)
		5 (3φ2P)

* At 25°C, 60 Hz.

Contactors are, in general, constructed in the same way as relays. Normally, they are supplied with two, three, or four N.O. contacts. They are used for switching power to control motor loads in refrigerating, air conditioning, heating, lighting, and ventilating systems. Applications also include controlling heating elements and primary power to an assembly such as a copy machine or a computer. Ratings are generally listed by continuous amperes or horsepower. A typical listing will look like the one shown in Table 4-1. The symbol 1φ2P denotes single phase with 2 poles and 3φ3P denotes three-phase with 3 poles. Contacts are usually field-replaceable. As an option, an auxiliary snap-action switch is available for low-power switching or coil interlock and latching. With the use of a control transformer, the contactor coil may be operated at a lower voltage, as shown in Fig. 4-39. The addition of the control transformer is desirable in some instances to keep high voltages off the switches and lights, and away from the operators. This simple circuit provides a means of energizing the contactor. Depressing the momentary start button completes the circuit; the contactor energizes, the light (*R*) lights, and the associated power is applied to the load. The contact's ICR (usually auxiliary) will now close, holding the circuit on when the momentary start button is released. This condition will persist until the operator intervenes by momentarily pushing the stop button. In some cases, a limit or thermal switch may be put in series with the stop button or hold-in contacts (ICR) for protection. One disadvantage of the second arrangement is that an operator can hold the start button in and override the interlocks if they are wired in series with the hold-in contacts.

Contactors can be ac- or dc-operated, that is, for the line or the actuation coils. In dc operation, a magnetic blowout coil wound on a steel core is mounted between pole pieces. The blowout coil is connected in series with the contactor and carries the dc load current with the contactor closed. The current sets up a magnetic field through the core and the pole pieces. When the contacts open, the arc is magnetically forced up and away from the contacts. This lengthening and subsequent extinguishing of the arc is rapid and greatly reduces wear and contact burning. *Contact wear allowance* is the total thickness of material that may be worn away before the contact surfaces become ineffective. The contacts

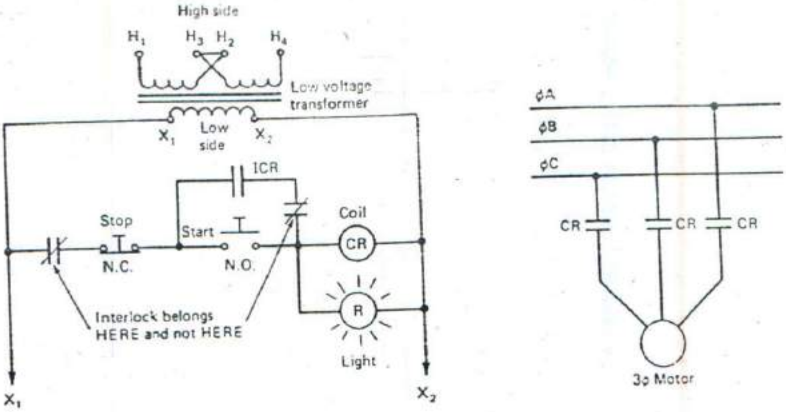


Fig. 4-39 Simple pushbutton contactor circuit.

should be renewed when worn below the amount specified by the manufacturer.

Starters can be of the manual or magnetic type. *Manual contacts* are closed by operator force. Today's machines often require the convenience and safety of remote control. The most common manual type employed today is the drum-type switch. A *drum switch* consists of stationary contact fingers held by spring pressure against contacts on the periphery of a rotating cylinder or sector. Their use is in applications in which the starter is operated frequently, as in a machine shop by a machinist. A reversing drum switch's internal connections are shown in Fig. 4-40(*a*). A stop can be inserted or removed by installing the proper face plates for reversing or ON-OFF operation. By connecting, as shown in Fig. 4-40(*b*), a three-phase, three-wire motor can be reversed. Because no starting windings, capacitors, or centrifugal switches are involved, reversing can be almost instantaneous. When used for trolley- and crane-type operation, dynamic breaking may be included in the starter switch.

The *magnetic starter* is very similar to the magnetic contactor in design and operation. Both have the feature of operating contacts when the coil is energized. The important difference between contactors and starters is the use of overload heater elements in the starter. There are motor starting switches, for controlling small ac or dc motors up to 2 hp, that are equipped with their own overload protective devices. It is not always possible to control the load that is applied to a motor. As a result, the motor may overheat, and serious damage can occur. For this reason, overload heater elements are added in the motor starter. The same current that is causing the motor to overheat is also going through the thermal element. We will now and in the future refer to these elements as *thermal overload relays*, as they are designated by industrial convention. They are available in the bimetallic type and the fusible-alloy type. The *bimetallic* type has two heaters in series with the circuit to be protected, and above these heaters are two bimetallic strips. Heating will activate the bimetallic strips, release, and cause the contacts to open. The *fusible-alloy* type has two heaters, each surrounding a thermal element consisting of a small tube, inside of which is a loose-fitting shaft. The tube and shaft are joined together by a low-melting alloy. On overload, the increased current melts this alloy, allows the shaft to turn, and open the contacts. All overload relays should be *trip-free*. This means that they cannot be held reset, causing damage to the motor. If an overload trips, the cause should be investigated and removed before the reset is actuated to put the starter back into operation. The overloads are selected to have an ultimate tripping current of 15 percent overload on the motor. Figure 4-41 shows a typical three-phase motor starter circuit diagram with thermal overloads included. Pushing the start button energizes the CR coil and closes all CR contacts. The motor runs until the stop button is pushed or until an overload activates the thermal elements in line 1 or line 3, opening one of the overload (O.L.) switches in series with the CR coil. If it is the fusible type, it must be replaced

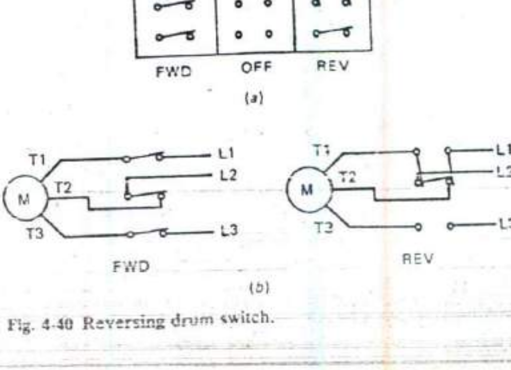


Fig. 4-40 Reversing drum switch.

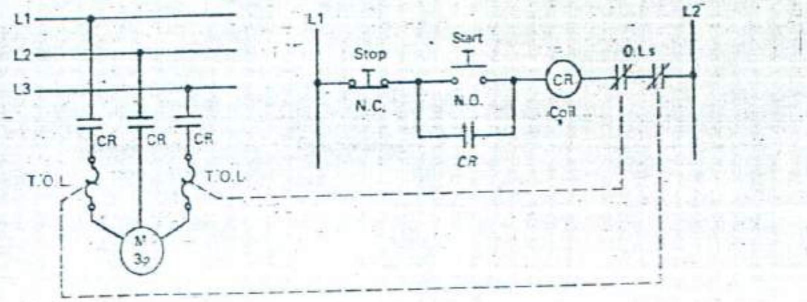


Fig. 4-41 Starter circuit for three-phase motor.

before resetting can be done. The bimetallic type can be reset after it cools and proper investigation of the cause has taken place. A commercial full-voltage starter is shown in Fig. 4-42.

Reduced-voltage starters are divided into several types:

1. Autotransformer
2. Primary resistor
3. Wye-delta (or star-delta)
4. Part winding

Why use a reduced-voltage starter? In connecting a motor directly across the line, the in-rush (starting) current may be on the order of five to ten times the full load current of the motor. This high current often will cause large line disturbances and excess stress on components, connections, and wiring. The basic principle of the reduced-voltage starter is to apply some percentage of the line voltage to start the motor. After the motor starts to rotate, it is switched or incremented to full line voltage. In the *autotransformer* type, starting steps are usually 50, 65, or 80 percent of full voltage. It should be noted that at reduced voltage, the output torque available is also reduced. This effect is important when the motor is starting under a heavy load.

For example.

50 percent voltage tap, the torque is 25 percent.

65 percent voltage tap, the torque is 42 percent.

80 percent voltage tap, the torque is 64 percent.

(Note: Torque varies as the square of the voltage.) The circuit diagram in Fig. 4-43, shows the contacts of a reduced voltage autotransformer starter. The two autotransformers are connected in open delta to provide reduced-voltage starting. The 50 percent taps have been selected for 25 percent starting torque. Five start contacts S and three run contacts R are required. The overloads have been omitted for clarity but would be included and are always necessary. A timing relay is operated by the starter contactor. The motor accelerates at reduced voltage through the S contacts; after a few predetermined seconds, the timer contacts close, deenergizing the starter (S) contactor and energizing the running (R) contactor. This also disconnects the autotransformer and puts full line voltage on the motor.

Figure 4-44 illustrates a *resistor-type starter*. Similar to the autotransformer type previously mentioned, it has the same sequence. After a preset time the run contactor short-circuits the line resistors and applies full line voltage to the motor. This type of starting gives smooth acceleration, because the mo-

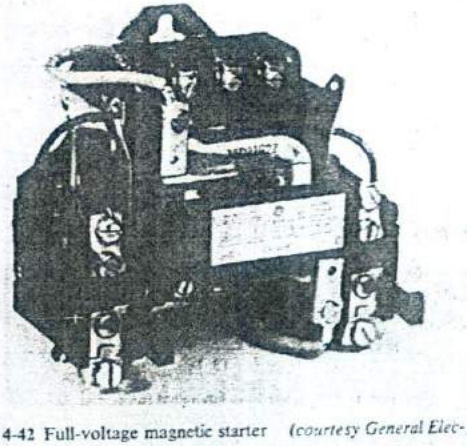


Fig. 4-42 Full-voltage magnetic starter *(courtesy General Electric Co.)*.

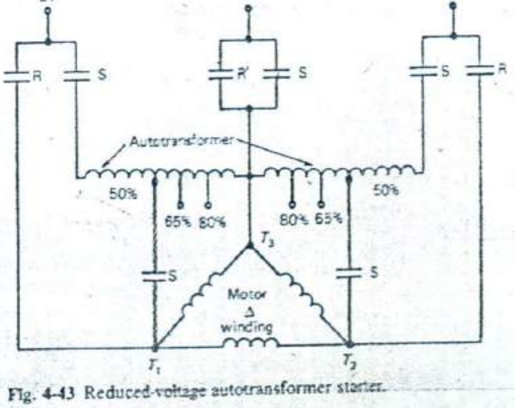


Fig. 4-43 Reduced-voltage autotransformer starter.

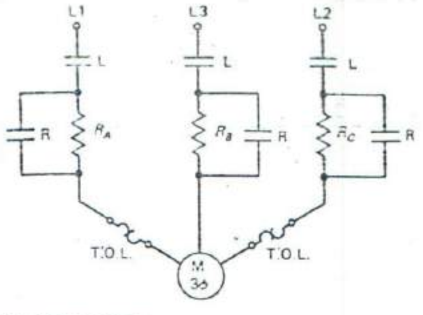


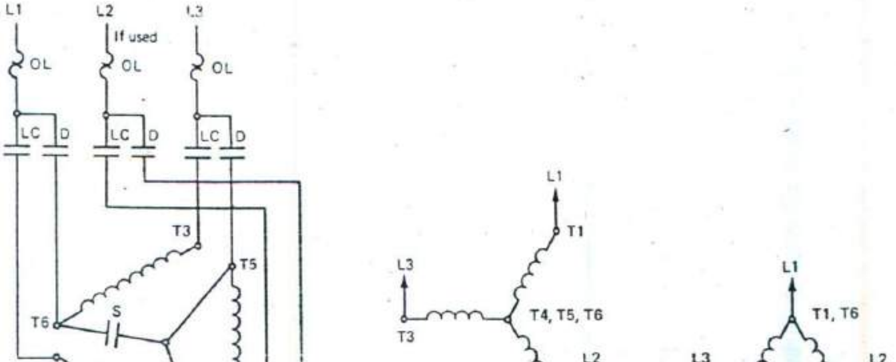
Fig. 4-44 Resistor starter.

tor is never disconnected from the line, but is very inefficient as a result of the high power loss in the resistors during starting.

When using the *wye-delta* (or *star-delta*) ***reduced-voltage types***, all three-phase motor windings (six connections, T_1-T_6) must be available for connecting to the starter as indicated in Fig. 4-45. Figure 4-45(*b*) shows the equivalent start ciruit. This circuit is achieved by closing the S contacts and the LC contacts. After a preset delay, the S contacts open and the D contacts close, and the LC contacts remain closed. Figure 4-45(*c*) shows the equivalent run circuit. The S and D contacts are mechanically interlocked to prevent both from ever activating at the same time. This type of starting requires no extra components as the previously discussed types do, but the motor must be made for this type of starting. That is, all winding connections must be available.

The *part winding starter* also requires that the motor have all windings available (six or nine connections). There are six basic arrangements for this type of starting, and it is best understood by referring to the diagrams and table of Fig. 4-46, in which the run (R) contacts and start (S) contacts are shown. The common two-step circuit which is illustrated provides a starting current equal to about 65 percent of the motor's normal locked rotor current for a starting torque of about 45 percent. The equivalent circuit connections for the table of Fig. 4-46 are shown. The starter is designed so that when the start contacts are energized the part winding of the motor is connected to the input lines. After about 4 s the second set of windings is connected, and the motor develops its normal torque.

The last type of starter we will investigate is the *reversing starter*. The reversing starter is, in effect, two starters, of equal size, used for a given motor application. To reverse direction of any three-phase motor any two line connections are interchanged, as with the drum switch of Fig. 4-40. It is a problem to connect the two starters to the motor properly so that the line feed from one starter is isolated from the other. Either mechanical or electrical or both types of interlocks are employed to prevent both starters from closing their line contacts at the same time. Figure 4-47 (p. 94) shows a commercial reversing magnetic starter. Note in Fig. 4-48 (p. 94) that only one set of overloads is needed to protect both forward and reverse connections. When the forward (F) button is actuated, the normally closed F contacts in the reverse circuit open, disabling the reverse start circuit until the stop button is actuated. And the converse is true for reverse (R) motion; note also the R contacts in the forward circuit. The interlocking described is the electrical type. Catastrophic failure would occur if both sets of contacts were allowed to close at the same time. As a double precaution a mechanically interlocking type of reversing starter is

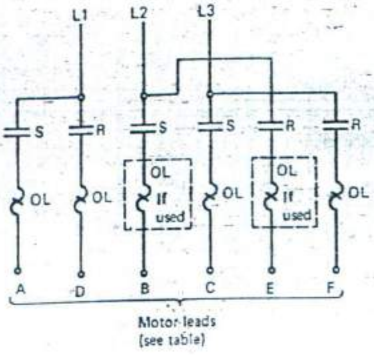


Equivalent start circuit

Equivalent run circuit

(*a*) (*b*) (*c*)

Fig. 4-45 Wye-delta starter. (*a*) Starter circuit. (*b*) Wye connection. (*c*) Delta connection.

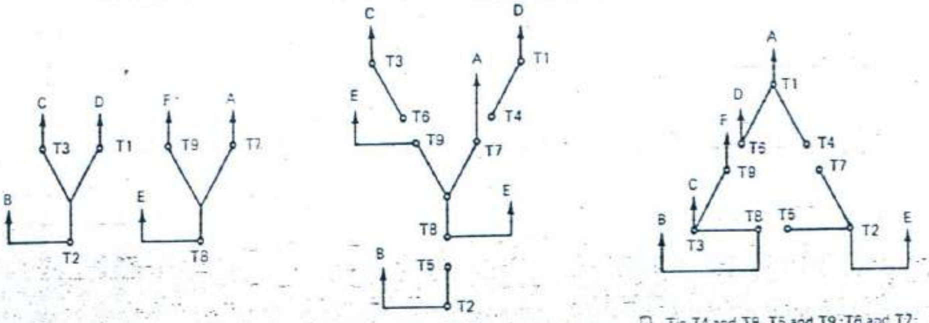


Motor Lead Connections

Part Winding Schemes		Lettered Terminals in Panel					
		A	B	C	D	E	F
1/2 Y or Δ	6 leads	T7	T2	T3	T1	T8	T9
1/2 Y	9 leads ○	T7	T2	T3	T1	T8	T9
1/2 Δ	9 leads ☐	T1	T8	T3	T6	T2	T9
2/3 Y or Δ	6 leads	T9	T8	T1	T3	T2	T7
2/3 Y	9 leads ○	T9	T8	T1	T3	T2	T7
2/3 Δ	9 leads ☐	T1	T4	T9	T6	T2	T3

○ Connect terminals 4, 5, and 6 together at motor terminal box.

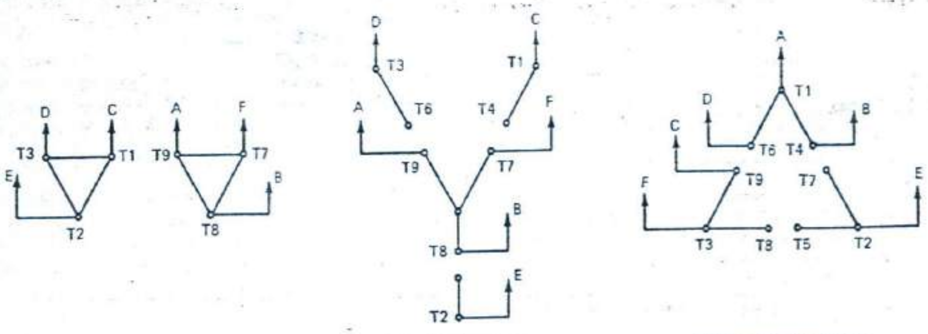
☐ Connect terminals 4 and 8, 5 and 9, 6 and 7 together in three separate pairs at terminal box.



6 Leads-wye
AA

○ Tie T4, T5, and T6 together at box.
9 Leads-wye
BB

☐ Tie T4 and T8, T5 and T9, T6 and T7; together in three separate pairs.
9 Leads-delta
CC



6 Leads-delta
DD

○ Tie T4-T6 as in BB.
9 Leads-wye
EE

☐ Connect pairs as in CC.
9 Leads-delta
FF

Fig. 4-46 Part winding starter.

available and is preferred by many designers and technicians.

REVIEW QUESTIONS

25. The "stopped iron-clad solenoid" shown in Fig. 4-35(*d*) is designed for increased mechanical ________.

26. If misaligned with its load, a solenoid will not pull in or ________.

27. A jammed plunger in an ac solenoid can ________ the coil.

28. What is the maximum ON time for a 33 percent duty cycle if the OFF time is 2 min?

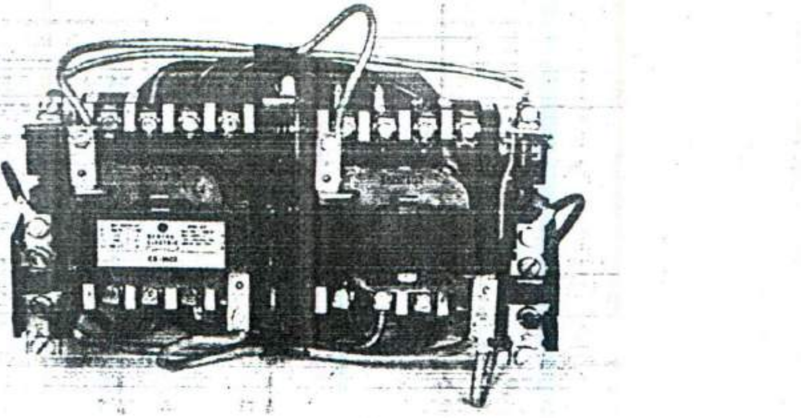


Fig. 4-47 Magnetic reversing starter (courtesy General Electric Co.).

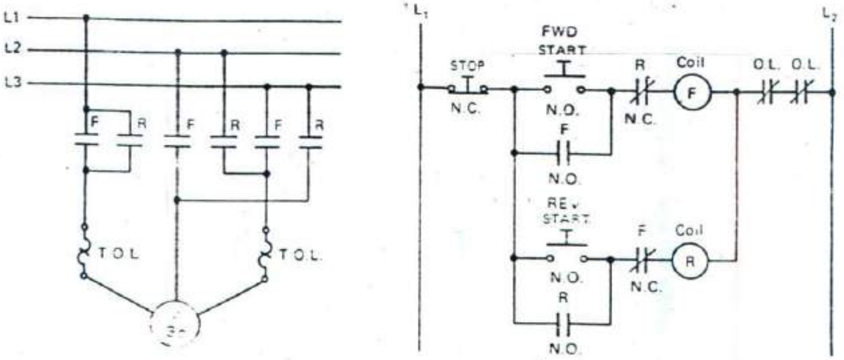


Fig. 4-48 Reversing starter.

4-6
TROUBLESHOOTING AND MAINTENANCE

Basic controls are the primary controls for most industrial equipment. Because most of these devices are line-operated, extreme care and safe troubleshooting techniques are necessary. Because of the nature of these controls the analysis of any problem has to be broken down as to whether it is electrical or mechanical. In some cases, problems may be both. The line voltage may be 110, 220, or 440 V so it is recommended that battery-operated test equipment be used. A buddy system is desirable when controls are remote from the device they are controlling. The best sequence of effective troubleshooting starts with a careful analysis of the problem. If resistance measurements are to be made or if it is necessary to disconnect leads, you must lock out circuits and tag them OFF for maintenance as a standard safety practice. Occasionally, a problem may be due to personnel. As long as people operate controls that operate machines and processes, problems that are not within the normal realm of troubleshooting techniques can arise. They may be due to a misunderstanding or a lack of knowledge of the process. Whatever the problem, it should be recognized and handled carefully, efficiently, and as diplomatically as possible.

Visual inspection of all external controls, lights, fuses, breakers, and connections is the first order of business. Any unusual odors or signs of burning or overheating should be investigated immediately, especially where solenoids, transformers, and relay coils are involved. A blown fuse or tripped circuit breaker indicates the need to investigate, not merely to replace or reset the protection device. Fuses are temperature-sensitive, and therefore poor or loose connections of their holders can cause a failure not related to the existing circuitry. This is also true for circuit breakers and especially thermal overloads used for motor starters. Some overload units may be on the same frame as the motor or processor and thereby subject to constant vibration. This vibration can cause connections to loosen and produce intermittent connections or overheating in high-current circuits. Troubleshooting can be done with the power on (*hot*) or with the power off (*dead*). In the OFF circuit, the ohmmeter is one of the best ways to check continuity of a circuit. Parallel or branch circuits may have to be disconnected to eliminate alternate paths. Most toggle, slide, and snap-action types of switches can be checked this way with good results.

Rotary switches may make at some positions and not at others, so careful testing is necessary. Cleaning can be done with a sharpened pencil eraser or commercial spray cleaners. Be sure to follow the directions on the can. Caution must be observed because many contact cleaners are not safe to use on plastics and should be avoided. Many of today's switches have plastic parts, and instead of a simple cleaning job, a major replacement may be required if the wrong cleaner is used. Most thumbwheel-type switches can be checked with the aid of a schematic or a table such as the one shown in Fig. 4-6.

Keyboard types are best serviced with the use of an oscilloscope since a serial-type code may be produced on actuation. Some keyboards require whole blocks of key replacements, although others allow individual keys to be replaced. The service literature is the best aid in such circumstances.

Electropneumatic TDRs may develop leaks or bind internally and usually require replacement. A shortened time-out can be due to a loss of fluid or gas, as an increase in timing may be caused by dirt or grime on moving parts. If at all possible, remove the unit, if cleaning is indicated, to a vented hood, using the manufacturer's recommendations on cleaning.

Electromechanical TDRs are very susceptible to mechanical wear and fatigue of movable parts and need periodic lubrication. Some have covers to prevent dust, dirt, or vapors from contaminating the enclosure and should be replaced immediately during maintenance.

Most solid-state timers are virtually maintenance-free as far as their mechanical aspects are concerned. But semiconductors do fail and must be replaced. If the frequency of failures increases, the power supply should be investigated. Chapter 5 discusses power supply troubleshooting. Line transients (brief overvoltage condition) are another common cause of solid-state failures and should not be overlooked.

Relays should not hum or chatter when energized. In the case of dc types, the power supply should be checked first. The armature and gap should be free of all dirt and grime. Oil should never be on any part of a relay's magnetic circuit because it attracts and holds dust and dirt, which hinder the magnetic action of the armature.

Most contacts are plated with special materials and should not be cleaned with abrasive-type tools, such as files or emery cloth. Excessive pitting, as in Fig. 4-14, requires replacement of the relay or its contacts. Pitting may be due to failure of the suppression network or a load change from the original design, in which case a different suppression network may be required.

Reed relays are maintenance-free but do fail and need replacement periodically. They are magnetic devices and must not be subject to large stray magnetic fields; they have a low current capacity and will not tolerate overloading. Low supply voltage has a great effect on all relays and should be the first priority when doing checks with power on.

Solenoids are typically on-off, in-out, etc. devices. Any deviation such as noncompletion of a full stroke or sluggish operation should be investigated. Many solenoids are attached to valves that control air, water, oil, or other industrial fluids or gases. They can appear almost any place in a control system and are usually remotely controlled. Alternating current solenoids may hum slightly; however, an excessive hum usually indicates that the plunger is not seated properly. Remember the coil may overheat, as shown in Fig. 4-36, as the result of excessive current flow. Cleanliness when dealing with any electromechanical device cannot be overemphasized. Also remember to check the voltage on the coil because it may indicate a loose or intermittent connection and possibly a low voltage, which can cause poor actuation and result in eventual failure of the unit. A check of the dc resistance of coils is a good practice when looking for an open coil. Remember that ac coils rely on inductive reactance so their dc resistance will be less than that of an equivalent dc type. A current probe or clamp-on ammeter will give the best results to verify the manufacturer's indicated ratings.

Rotary solenoids are intermittent devices; if failures occur too frequently, the current limiting or timing network may be defective. For instance, if the capacitor (C) in Fig. 4-38(b) is leaky, the transistor (Q_1) may shunt the limiting resistor (R_2) and cause the coil to overheat, resulting in premature failure. Or an increase in the resistance of R_2 may cause the rotary solenoid not to hold; a simple ohmmeter test is best to verify this problem.

Common problems encountered with contactors and starters are loose connections and grounds. Most wiring to these devices is through conduits. Nicks and scrapes on wire insulation that occur during installation may produce problems at a later date. Grounding on one lead of a three-phase motor, as in Fig. 4-39, may reduce its speed or torque or even change its rotation. When voltage checking, phase-to-phase voltages, as well as phase-to-neutral measurements, are necessary. Contacts used in starters and contactors are available from the manufacturer and should be replaced when excessive wear is determined. These units should not hum, as discussed in connection with solenoids, and often indicates improper seating. Along with checking the coil voltages, look for loose parts or connections and dirt. Dirt or grime in the armature gap will inhibit proper seating and lead to excessive contact wear due to low pressure. Looking at closed contacts will not provide the necessary information, since contact damage will not be apparent. As mentioned about relays, most contacts are plated and should not be filed. When contacts are pitted or worn, contact replacement is the best solution and may prevent a failure at a future date.

Thermal overloads are temperature-sensitive devices, just as fuses are. Therefore, cleanliness, proper ventilation, and good solid connections are very important for efficient operation. Frequent tripping of an overload may be the first indication of motor damage, a load increase, or some sort of binding. A current measurement of the motor's line current under load will be in order if no short circuits in the wiring or motor are indicated.

Fuses and circuit breakers play an important role in the safe operation of electrical equipment. A violent rupture of a fuse or instaneous tripping of a breaker usually indicates a dead short or breakdown. Checking for short circuits with an ohmmeter is the best procedure before reenergizing the system. If a breakdown is suspected, a Megger (a high-resistance/high-voltage ohmmeter) test is one of the best but should be used only if the suspected device can be isolated from the circuit. Never overrate or override a fuse or circuit breaker. Doing so may result in a fire or other severe equipment damage. The most important aspect of good troubleshooting and maintenance is logical and sensible application of your knowledge.

REVIEW QUESTIONS

29. Basic controls besides having electrical problems may also have ______ problems.

30. A ______ inspection of all external controls is the first approach to solving a problem.

31. In Fig. 4-27, if *C* were short-circuited what symptom would result?

32. Excessive hum on an ac relay or contactor may indicate what?

33. Filing pitted relay contacts is a good maintenance practice. (true or false)

34. Overheating in solenoids is the most common cause of their failures. (true or false)

CHAPTER REVIEW QUESTIONS

4-1. What type of switch matrix is used for the DTMF coding? Hexadecimal?

4-2. Name a contactless type of keyboard switch.

4-3. Calculate the *R* and *C* needed across contacts breaking 1 A at 25 V dc.

4-4. Draw a circuit that makes form C relay contacts by using form A and form B contacts.

4-5. Would you expect more or less contact bounce and chatter from a mercury-type relay?

4-6. What type of relay uses a separate reset coil?

4-7. Caution during cleanup of a broken ______ relay is a necessary safety practice.

4-8. Relay drop-out current is usually considerably ______ than relay pick-up current.

4-9. If the capacitor voltage of Fig. 4-27 is zero, the output is ______.

4-10. If pins 5, 6, and 7 are connected to the output bus on Fig. 4-28, and *RC* equals 10 ms, *T* equals ______.

4-11. To operate in the astable mode, S_1 of Fig. 4-28 is left ______.

4-12. What controls the delay of the magnetic-hydraulic breaker?

4-13. What can happen to a breaker that has not been tripped for a long period?

4-14. If Q_1 short-circuits in Fig. 4-38, the rotary solenoid will ______.

4-15. If you were to put a pilot lamp into the circuit of Fig. 4-41 to indicate control, should it be connected across start, stop, O.L.s, or the coil?

4-16. If the overloads trip in Fig. 4-48 in the forward direction, will the motor stop or reverse direction?

4-17. Increasing a solenoid's stroke will help it to seat better. (true or false)

4-18. Refer to Fig. 4-39. If the 1CR contact failed to close, what would be the symptom?

4-19. What is the purpose of the N.C. R and F contacts in Fig. 4-48?

4-20. Refer to Fig. 4-32. If any of the F_1 fuses opens, a ______ may be short-circuited.

4-21. What is the primary difference between a 3AG and a European fuse?

ANSWERS TO REVIEW QUESTIONS

1. zero, infinite **2.** yes **3.** 12 **4.** short-circuiting **5.** thumbwheel **6.** snap-action **7.** spring action **8.** ac relays **9.** dc loads **10.** stick or lock **11.** single diode, ac power **12.** B **13.** both ac and dc **14.** false, slows actuation **15.** shorten **16.** armature **17.** interval **18.** 1.0 MΩ **19.** fusing characteristics **20.** RMS **21.** temperature **22.** slo-blo or dual element **23.** fast-acting **24.** no instantaneous trip point **25.** pull **26.** seat **27.** overheat/burnout **28.** 1 min **29.** mechanical **30.** visual **31.** cannot trigger, inoperative **32.** dirty magnetic circuit **33.** false **34.** true

5

POWER SOURCES

This chapter deals with one of the more fundamental units in industrial electronics and robotics. Power sources are critical to any industrial system. A malfunctioning source is almost sure to prevent normal system operation. Practically all of the electronic circuits used in industrial systems require direct current (dc) for energization. This chapter examines modern dc sources, their principles and characteristics, and troubleshooting techniques.

5-1
BATTERIES

Cells and batteries have long been important items in industrial systems. Lately, they are increasing in importance because of two major trends. The first is the increasing amount and diversity of portable equipment. Modern components and manufacturing techniques have made it possible to create mobile and portable equipment that formerly would not have been feasible because of size and weight restrictions. This portable equipment must often operate independently of fixed power circuits. The second major trend is uninterruptible circuits and equipment. These devices must continue to operate, or at least retain important information, in the event of a loss of primary power. Such devices are often said to be *battery-backed-up* and are increasing as more digital and computer systems are applied in the industrial environment. A good example is an industrial control system that stores important data in memory circuits as it operates. These memory circuits may be backed up by battery power in case of a failure of the main alternating current (ac) supply. In this case, the control system can quickly resume proper operation when the main supply is restored. Without backup, restoration to normal operation could be very complicated because the system would have no way of determining various data parameters and other control conditions at the time of the interruption.

A battery is made up of series, parallel, or series-parallel combinations of electrochemical cells. Series combinations yield more voltage than a single cell can provide. Parallel combinations yield more ampere-hour (A · h) capacity than a single cell can provide, and series-parallel can increase both voltage and ampere-hour capacity. Cells are divided into two broad categories: primary and secondary. A *primary cell* invokes an irreversible chemical change in the cell structure upon discharge. This means that primary cells cannot be recharged. They are replaced with new cells (or new batteries) when they become discharged. A *secondary cell* or battery also produces chemical change in the cell structure when it is discharging, but the chemical change is reversible. Secondary sources can be restored to a full-charge, or near full-charge, condition by passing a charging current of the proper magnitude and duration through the source in a direction opposite to the discharge current.

Carbon Zinc Cell

Cells and batteries are usually identified by the major materials used to build them. The carbon-zinc cell or battery is a very inexpensive and therefore popular primary type. They are often called *dry cells* since the electrolyte is in paste form to prevent the leaks and spills often associated with *wet cells,* in which the electrolyte is in liquid form. In a carbon-zinc cell (Fig. 5-1) zinc acts as the anode, the electrolyte is ammonium chloride, and a carbon rod serves as the positive contact. The ammonium chloride splits into positive ammonium ions and negative chlorine ions. The zinc anode dissolves in the electrolyte and gives off positive zinc ions, leaving an excess of electrons on the anode and the negative terminal shown in Fig. 5-1. The zinc ions combine with the chlorine ions and form neutral zinc chloride. The ammonium ions are repelled by the zinc ions going into solution and migrate to the carbon electrode, where they pick up an electron and split into ammonium and hydrogen gases. This produces a deficiency of electrons on the carbon electrode and positive terminal shown in Fig. 5-1. When an external load is connected, conventional current flows from the positive terminal, through the load, and into the negative terminal of the cell.

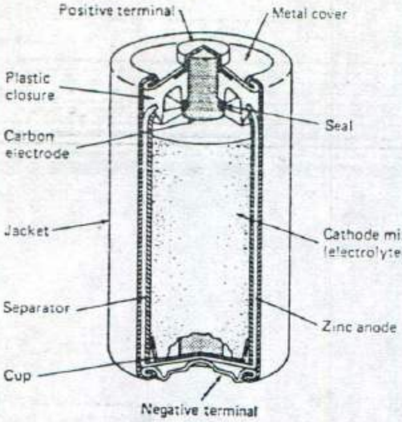


Fig. 5-1 Construction of a carbon-zinc cell.

During high discharge rates, the chemical action of the carbon-zinc cell can cause the cell capacity to fall off. The hydrogen gas that is formed collects on the surface of the carbon rod. Hydrogen gas acts as an insulator, and the resistance of the cell increases. In this condition the cell is said to be *polarized* and cannot deliver adequate current flow. Manganese dioxide is added to the electrolyte to act as a depolarizer. It provides oxygen that combines with the hydrogen gas to form water. However, the depolarizer will not keep up with the hydrogen production at high discharge rates. Therefore, carbon-zinc cells may have to be rested to realize their expected ampere-hour capacity when they are used in high-current applications.

Carbon-zinc cells should not be stored for long periods of time. Storage time is referred to as *shelf life* and should not exceed 1 year. The electrolyte tends to dry out, raising the internal cell resistance. Another problem is that impurities in the zinc anode will set up many *local cells*, and the resulting local action deteriorates and depletes the cell. Shelf life can be improved somewhat by storing carbon-zinc cells and batteries at reduced temperatures.

Alkaline Cell

The alkaline-manganese cell (usually called *alkaline cell* or *alkaline battery*) is constructed inside-out when compared to the carbon-zinc cell and offers considerable improvements. It is also a primary type and cannot be recharged. A zinc rod in the center of the cell serves as the anode, and the electrolyte is an alkaline solution of potassium hydroxide and zinc oxide. The outer, positive electrode is cylindrical manganese dioxide, which provides faster depolarization because of the large surface area of the cylinder. As a result, alkaline cells and batteries are much better suited to high-current applications. They do not require rest periods and can last several times as long as equivalent carbon-zinc types. They generate 1.5 V per cell (as does carbon-zinc) and are built with a button top that contacts the positive electrode and a bottom plate that contacts the zinc rod. This makes their external appearance and polarity compatible with those of carbon-zinc cells. They are more expensive to buy, but they may be more economical in the long run.

EXAMPLE

A size D carbon-zinc cell will provide 50 mA for 15 h. At this point, the cell voltage will have dropped to 1.3 V, which is usually considered the discharged condition. A size D alkaline cell will provide 50 mA for 56 h before reaching the 1.3-V cutoff. What is the capacity of the two cells?

SOLUTION

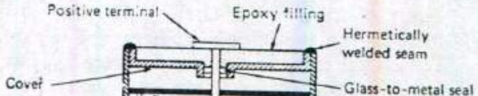
The ampere-hour capacity for the two types can be calculated as follows:

$$\begin{aligned} 50\ \text{mA} \times 15\ \text{h} &= 750\ \text{mA}\cdot\text{h} \\ &= 0.75\ \text{A}\cdot\text{h} \\ 50\ \text{mA} \times 56\ \text{h} &= 2800\ \text{mA}\cdot\text{h} \\ &= 2.8\ \text{A}\cdot\text{h} \end{aligned}$$

It is clear that the alkaline cell provides almost four times the ampere-hour capacity provided by the carbon-zinc cell. Alkaline cells also enjoy a longer shelf life (about 3 years) and are less likely to leak and damage expensive circuits and equipment.

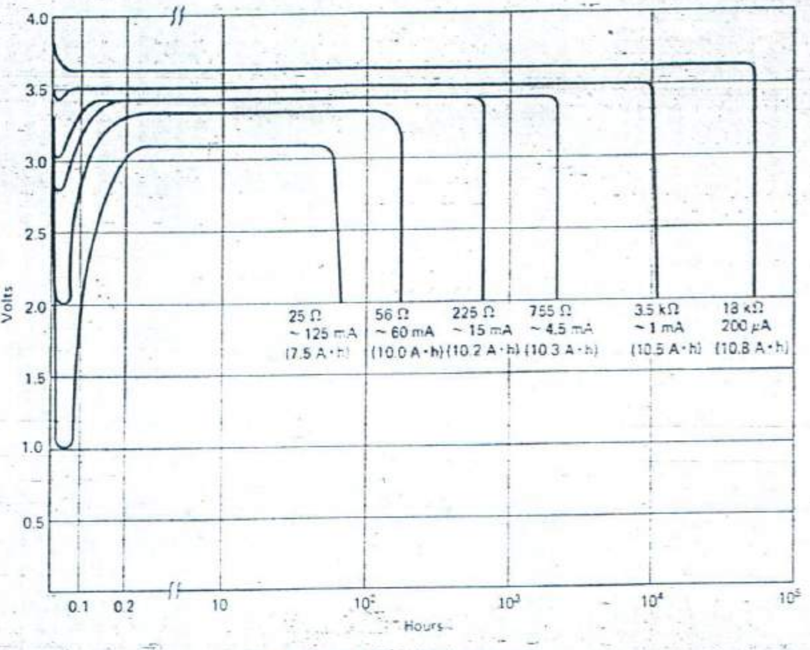
Lithium Cell

Lithium cells are available in several types. The construction of a lithium thionyl chloride cell that boasts the highest energy density available among primary sources is shown in Fig. 5-2. The densities are as high as 420 W · h/Kg and 800 mW · h/cm^3. Lithium cells and batteries can be considered permanent components in some electronic systems. Because of their extremely low self-discharge characteristics and their



Note: The system is immersed in electrolyte.

Fig. 5-2 Construction of a lithium thionyl chloride cell.



Typical discharge 25°C after 1 year storage

Fig. 5-3 Lithium thionyl chloride cell discharge characteristics.

hermetically sealed construction, their projected shelf life is greater than 10 years when they are stored at room temperature. They develop an open circuit potential of 3.68 V and a nominal working 3.4 V.

Figure 5-3 shows the discharge characteristics for a size D lithium thionyl chloride cell. Notice the unusual voltage response for the first 0.1 h of operation. With a 25-Ω load (approximately 125 mA), the cell output drops to 1 V and then recovers to 3.1 V after 0.3 h. This is called the *transition period* and may be an unacceptable characteristic because the transition minimum voltage may not be adequate to operate some systems. This type of cell is not generally used in high-current applications for this reason. Note, from Fig. 5-3, that the transition minimum voltage is about 3.4 V for a load of 3500 Ω (approximately 1 mA). Also note that the cell will deliver 10.5 A · h of energy when loaded this way. This is 14 times the energy that can be realized from the same size carbon-zinc cell! Figure 5-3 also shows that the plateau voltage is extremely flat after the transition period and extends to over 50,000 h for a load current of 200 microamperes (μA). Fifty thousand hours equates to 5.71 years. It should be clear why lithium batteries and cells are attractive sources for low-current applications such as memory backup and as voltage references.

Figure 5-4 shows a typical backup circuit for a complementary metallic oxide semiconductor random access memory (CMOS RAM). When the main 5-V supply is available, diode D_1 is forward-biased, and current is supplied to the memory circuit as shown. Diode D_2 is now reverse-biased, and its leakage is normally in the nanoampere (nA) range and can be ignored. If the main 5-V supply fails, D_1 is reverse-biased, and D_2 becomes forward-biased. A backup current flows as shown in Fig. 5-4. The backup voltage will be approximately 3 V since D_2 will drop 0.6 V when forward-biased. A value of 3 V is adequate for data retention in the CMOS RAM. The resistor is optional and is included to prevent a high charging current through the lithium cell in the event that D_2 short-circuits.

Another variation is the lithium-iodine cell. Its construction can be seen in Fig. 5-5. It uses a lithium anode and an iodine cathode. It is designed to be permanently soldered onto a printed circuit board (PCB). The iodine version does not have the energy density of the type previously discussed. However,

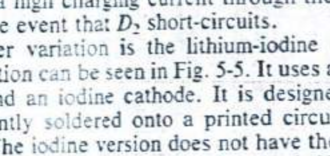


Fig. 5-4 Power back-up for memory chip.

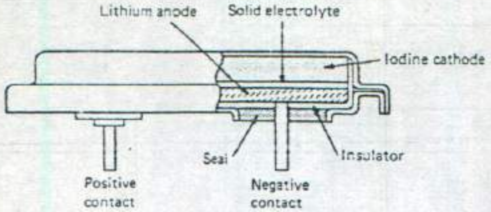


Fig. 5-5 Construction of a lithium-iodine cell.

it is considered a quasisecondary source in that it is capable of some recharging under conditions of low reverse currents. The charging current is typically limited to 1 μA. The circuit of Fig. 5-4 can be modified for a trickle charge of 1 μA by shunting D_2 with a high-value resistor. Since lithium-iodine cells develop an open-circuit voltage of about 2.8 V and since the drop across D_1 can be assumed to equal 0.6 V, the calculation for the resistor is straightforward:

$$\begin{aligned} R &= \frac{V}{I} \\ &= \frac{5 - 0.6 - 2.8}{1\ \mu\text{A}} \\ &= 1.6\ \text{M}\Omega \end{aligned}$$

Figure 5-6 shows that the storage time for a lithium-iodine cell is 100 years at room temperature! Other primary cells with industrial applications include the lithium manganese dioxide, mercury, and silver oxide cells.

Nickel-Cadmium Cell

Secondary cells are often used in those cases in which more current is required than in memory backup and reference applications. The discussion will be limited to two designs: nickel-cadmium cells and gelled electrolyte types.

Nickel-cadmium cells and batteries are often referred to as *ni-cads*. A typical ni-cad cell is shown in Fig. 5-7. It uses a potassium hydroxide electrolyte, a cadmium and iron oxide negative electrode, and a nickel hydroxide and graphite positive electrode. They develop an almost constant 1.25 V per cell over 90 percent of their discharge cycle. The discharge characteristics are shown in Fig. 5-8. Note that the discharge capacity can range from 90 to 120 percent, depending on the rate. A 0.1C rate indicates that the current will be 0.1 times the ampere-hour rating. A size D ni-cad is rated at 4 A · h and will actually deliver 4.8 A · h (120 percent) when it supplies a current of 400 mA (0.1C). The cutoff voltage is usually considered to be 1 V per cell. Ni-cads feature a low internal impedance and are well suited to high-current applications, including pulse applications, where rather high currents must be supplied for short periods. Due to their ability to supply high currents, they must not be short-circuited because cell damage or circuit damage may result. They are usually rated for at least 500 charge-discharge cycles.

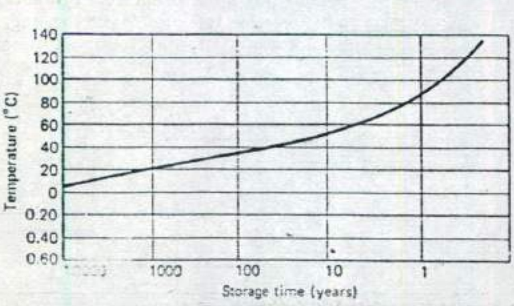


Fig. 5-6 Lithium-iodine cell storage time.

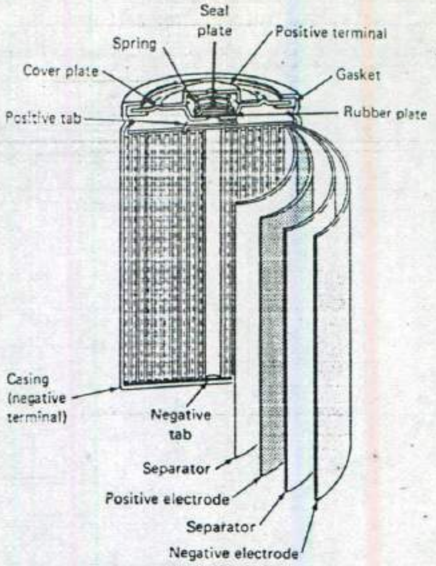


Fig. 5-7 Construction of a nickel cadmium cell.

Ni-cads will retain 50 percent of their capacity when stored for 3 months at 20°C. They may be stored in either a charged or discharged condition. Several cycles of charge-discharge may be required after an extended storage period to restore normal capacity. They may also require several deep cycles

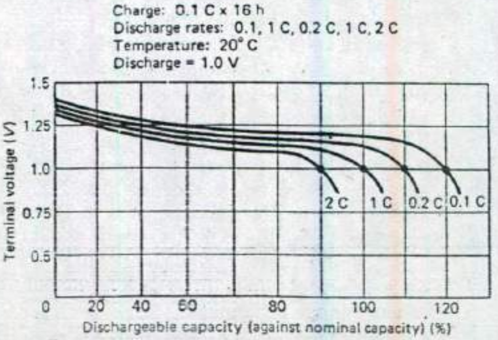


Fig. 5-8 Nickel cadmium cell discharge characteristics.

(drained to 1.0 V per cell and then completely recharged) after an extended number of shallow cycles. A *shallow cycle* is a situation in which the cell is recharged after it has delivered only a portion of its rated capacity. Too many shallow cycles may cause a "memory" effect, and the cell will tend to deliver only that portion of its capacity that it remembers delivering. This effect, although interesting, is often overrated, and ni-cad batteries containing three or more series cells are more often damaged by overly deep discharging. The individual cells are seldom matched in capacity and the weakest cell will eventually drop to 0 V and then become a load. When this occurs, its polarity will reverse and permanent damage often results.

Ni-cads can also be damaged by overcharging and by rapid charging. For this reason, a constant voltage source is not recommended. You may recall that a constant voltage source is characterized by a low internal resistance. Constant current is the preferred charging mode, and a constant current source is characterized by a high internal resistance. A voltage source can approximate a current source simply by addition of a series resistor.

EXAMPLE

It is desired to charge a 12-V ni-cad battery from a 24-V source. Also assume that the charging rate is to be the standard *C/10 rate,* at which the charging current is found by dividing the ampere-hour capacity (C) by 10. Find the value of the series resistor required.

SOLUTION

A C/10 rate will be 0.4 A for a 4-A · h battery. The resistor will have to drop 12 V (24 − 12) at a current of 0.4 A:

$$\begin{aligned} R &= \frac{V}{I} \\ &= \frac{12}{0.4} \\ &= 30\ \Omega \end{aligned}$$

The C/10 rate will require at least 14 h for full charge. The battery will not completely recharge in 10 h since no secondary source is 100 percent efficient. More charging energy must be delivered to the battery than its rated discharge energy. A C/10 rate is considered very safe since the battery may be left on charge indefinitely with no resulting damage due to overcharge. Faster rates are possible, such as a C/4 rate, which will recharge the battery in 5 h. However, it must be noted that a C/4 rate must not be applied to a ni-cad for more than 6 h; if it is, the cell temperature will begin to rise and permanent damage will result.

Special quick-charge ni-cads are available to achieve a full charge in a minimum of 4 h. Special charging circuits are often used with these types to provide automatic cutback to a C/10 rate when full charge is reached. The cutback is usually triggered by a circuit that monitors cell voltage and cell temperature.

Gel Cell

Gelled-electrolyte cells (usually referred to as *gel-cells*) are close cousins to the lead-acid cells used in automobiles. They enjoy a high output of 2.2 V per cell, and their cutoff point is 1.75 V per cell. They can be totally discharged without any danger of damage due to cell reversal. They use lead-calcium electrodes and are completely sealed to operate in any position. However, they do require a one-way safety vent in the event that cell pressure would exceed some safe limit. They are capable of several hundred cycles and should not be charged indefinitely, even at a C/10 rate. They also should not be stored for extended periods in a discharged condition. When this occurs, permanent sulfation of the electrodes will reduce capacity and eventually lead to cell failure. Gel-cells, like ni-cads, have a low internal resistance and should not be short-circuited. They serve well in high-current applications, although the delivered capacity drops in these situations, and rest periods may be required. The major drawback of gel-cells is their tendency to release hydrogen gas. This is especially prevalent at high charging rates. Hydrogen gas is extremely explosive, and this characteristic precludes the use of gel-cells in some industrial environments.

REVIEW QUESTIONS

1. Which broad category of cells undergoes an irreversible chemical change during discharge?

2. Carbon-zinc cells may require rest periods when used at high discharge rates because of cell ________.

3. Calculate how long an alkaline D cell will supply 80 mA before reaching its cutoff voltage.

4. Which type of cell, under a condition of high initial discharge, produces a minimum transition voltage before reaching its plateau voltage?

5. It is desired to trickle-charge a 2.8-V lithium-iodine cell at 2 μA from a 10-V supply. Using Fig. 5-4 as a guide, calculate the value for the resistor that would be connected across D_2.

6. You wish to charge a 6-V, 2-A · h ni-cad battery at the C/10 rate. A 12-V constant-voltage source is available. Calculate the series resistor.

5-2

RECTIFICATION

Rectification is the process of changing alternating to direct current. Solid-state diodes have proved to be very efficient rectifiers. Figure 5-9 shows a half-wave rectifier circuit that uses a diode (*D*) to change alternating current to direct current. An alternating voltage appears across the secondary of the transformer and forward-biases the diode every half cycle. Current flows through the diode and the load resistor only half the time and only in the direction shown.

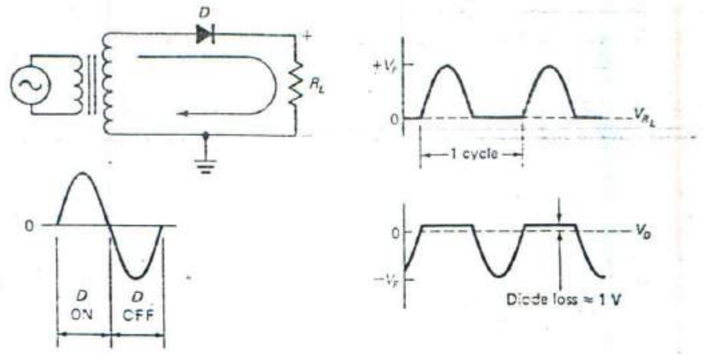


Fig. 5-9 Half-wave rectifier circuit.

Note the waveforms across the load (V_{R_L}) show in Fig. 5-9. This waveform is called *half-wave pulsating direct current* and shows one load pulse for every cycle of the input. The waveform across the diode (V_D) shows the reverse voltage peaks when the diode is off. It also shows the *diode loss*, which is the voltage drop across the diode when it is conducting. For silicon rectifiers, this loss ranges from 0.7 to 1 V and is directly related to load current.

The polarity of the rectified direct current can be predicted by assigning a current direction through the load and then the drop across the load. Figure 5-9 shows conventional current flowing down through the resistor: the resistor polarity is therefore positive at the top. Another way to predict polarity is to observe that the cathode lead of the diode contacts the positive end of the load. This is always the case in any rectifier circuit and is a good way to predict load polarity. If the diode is reversed, the anode lead will contact the top of the load, thus making it negative. Since the bottom of the load resistor is grounded and since ground is the normal reference point, Fig. 5-9 can be considered a positive supply as drawn and a negative supply if the diode is reversed.

Half-wave rectifiers are not time-efficient since they load the source only on every other alternation. The extreme pulsation of load voltage and current is also considered a disadvantage. The pulsations can be removed by filtering, but the process often causes high-heating (root mean square [rms]) currents to flow in the transformer secondary and in the rectifier. Also, the unidirectional current flow in the secondary biases the transformer core with a dc flux, and a larger and heavier core may be required to avoid saturation. Because of their disadvantages, half-wave rectifiers are generally limited to low-power applications. They are more attractive for rectifying high-frequency square waves, for example, in switch-mode power supplies, which are covered in section 5-7.

A full-wave rectifier circuit using a center-tapped transformer is illustrated in Fig. 5-10. It uses two diodes and produces two load pulses for every cycle of the input. The load waveform is called *full-wave pulsating direct current*. When the top of the secondary is positive, D_1 is forward-biased, and load current flows as shown by the solid arrow in Fig. 5-10. Notice that only the top half of the transformer secondary is conducting, since D_2 is now reverse-biased. On the next alternation, D_2 is forward-biased; load current flows as shown by the broken arrow, and only the bottom half of the secondary is conducting. In Fig. 5-10, V_P will be equal to half the peak secondary voltage since only half of the transformer winding conducts at any given time. Note that the top of the load resistor is positive and is in contact with the cathodes of the rectifiers.

Fig. 5-10 Full-wave (center-tap) rectifier circuit.

The bridge rectifier in Fig. 5-11 is full-wave and uses four diodes. It eliminates the need for a center tap. When the top of the secondary is positive, D_1 and D_2 are forward-biased, and current flows as shown by the solid arrow. When the bottom of the secondary is positive, D_3 and D_4 are forward-biased, and current flows as shown by the broken arrow. Once again it is seen that the positive end of the load is in contact with the cathodes of the rectifiers. The load waveform in Fig. 5-11 is drawn as negative-going. Compare it to the load waveform shown in Fig. 5-10. The reason for the difference is that Fig. 5-11 is a negative supply. The positive end of the load is grounded. The circuit can be changed to a positive supply by reversing all four diodes or by changing the ground connection to the negative end of the load. The waveform shown in Fig. 5-11 is what would be seen on an oscilloscope if all conditions were normal. *Normal* means that the instrument ground is connected to the circuit ground, and a negative voltage causes a downward deflection on the oscilloscope screen.

The center-tap and bridge circuits both seem to be capable of the same performance. However, there are several differences. The bridge circuit uses two diodes conducting in series. The diode losses will therefore be twice as great. A typical diode loss in solid-state rectifier circuits is 1 V per diode. Therefore, 2 V will be lost in the bridge circuit and only 1 V in the center-tap circuit. This is significant in low-voltage supplies, and the center-tap circuit may be preferred in those applications. One advantage of the bridge circuit is that it requires only half as many secondary turns to develop any given voltage. Using half as many turns results in a less bulky transformer even though larger wire will be required because the entire secondary must conduct on both alternations. Another advantage of the bridge circuit is that the rectifier diodes are subjected to half the peak inverse voltage (PIV) when compared to the center-tap circuit. Because of these differences, the bridge circuit is generally preferred at higher voltages and the center-tap circuit at lower voltages. However, in practice, both circuits will be found operating over a broad range of voltages.

Figure 5-12 summarizes five important rectifier circuits. It lists their average dc output voltage (V_O), PIV per diode, rms ripple voltage, and ripple frequency and shows their output waveforms. Notice that the average dc output voltage is lowest for the half-wave circuit and is only 45 percent of the rms input voltage. Also notice that the ripple voltage is the highest, at 54 percent of the rms input voltage. This makes the percentage of ac ripple very high for the half-wave supply. The percentage of ripple in the output voltage is given by

$$\begin{aligned}\text{ripple, \%} &= \frac{\text{Ripple voltage as fraction of input voltage}}{\text{average dc output as fraction of input voltage}} \times 100\\ &= \frac{0.54}{0.45} \times 100\\ &= 120\%\end{aligned}$$

This high ripple percentage emphasizes one of the disadvantages of the half-wave rectifier circuit.

Figure 5-12 also shows two three-phase rectifier circuits. These circuits are very popular in industrial equipment and have several advantages over the single-phase circuits already discussed. Note the low rms ripple voltages for the three-phase circuits. For example, in the case of the three-phase bridge circuit:

$$\begin{aligned}\text{ripple, \%} &= \frac{0.057}{1.35} \times 100\\ &= 4.22\%\end{aligned}$$

This low percentage of ripple shows that the output of three-phase rectifiers is a much more pure form of direct current than is provided by single-phase circuits. Figure 5-13 shows why. The waveform for the line-to-neutral circuit is shown in Fig. 5-13(*a*). Three sine waves are drawn 120 electrical degrees apart. The negative alternations are shown as broken lines and are eliminated in the line-to-neutral circuit. The resulting ripple is formed by the positive alternations only. Figure 5-13(*b*) shows that the negative alternations are folded up to the positive part of the graph by the three-phase bridge circuit, resulting in a very small ripple voltage.

Two popular multiphase rectifier circuits are shown in Fig. 5-14. The six-phase star circuit (Fig. 5-14[*a*]) has a small (4.22 percent) ripple content and a ripple frequency equal to six times the line frequency. The graph shows six output pulses for one cycle of input. A six-phase bridge would achieve even less ripple content and a ripple frequency equal to 12 times the input frequency. Figure 5-14(*b*) is a three-phase double wye with an interphase transformer. Again, the ripple percentage is only 4.22 percent, and the ripple frequency is six times the line frequency.

Multiphase rectifiers have low ripple percentages and high ripple frequencies. These are both advantages. Another advantage, especially in high-current supplies, is the low rectifier-cell ratios found in multiphase supplies. The *cell ratio* is the comparison of rectifier current to load current. For example, in a single-phase half-wave circuit the cell ratio is 1.00 since the single diode must conduct the entire load current. A single-phase full-wave circuit shows a cell ratio of 0.5 because any diode is on half the time. Three-phase circuits have cell ratios of only 0.333, and the six-phase circuit shown in Fig. 5-14(*a*) shows a cell ratio of only 0.167. Small cell ratios relax the current ratings for individual diodes used in rectifier service.

Rectifier diodes must be derated according to case temperature and service. The curves in Fig. 5-15 show that these diodes must be derated at temperatures above 150°C. The diodes must also be derated

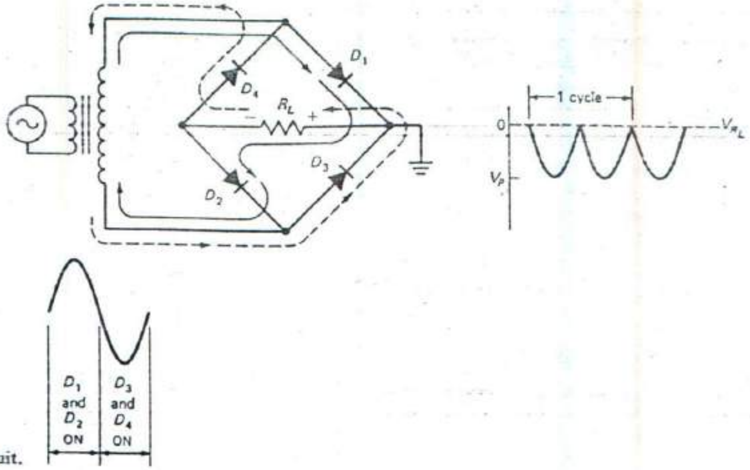
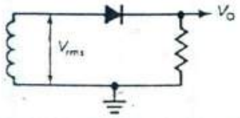
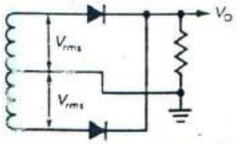
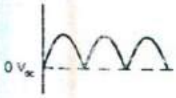
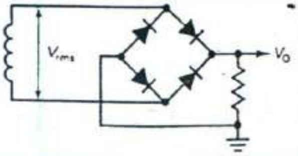
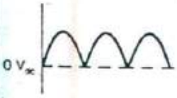
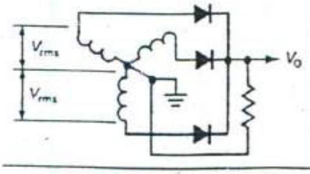
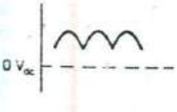
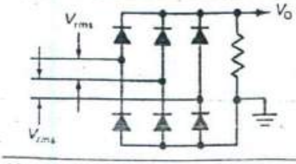
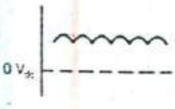
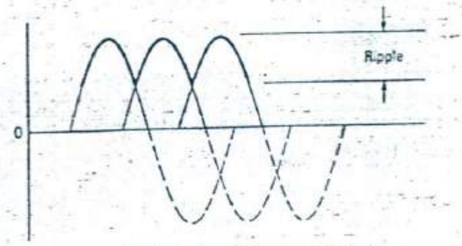


Fig. 5-11 Full-wave (bridge) rectifier circuit.

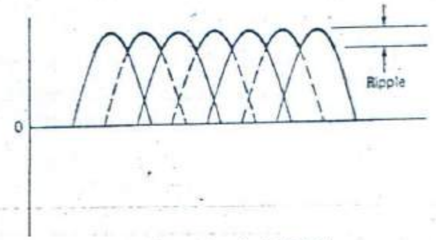
Schematic	Name	V_O (peak)	V_O dc	PIV per Diode	RMS Ripple Voltage	Ripple Frequency	Output Waveform
	Half-wave	1.41 V_{rms}	0.45 V_{rms}	1.41 V_{rms}	0.54 V_{rms}	$1f_L$*	
	Full-wave	1.41 V_{rms}	0.90 V_{rms}	2.82 V_{rms}	0.43 V_{rms}	$2f_L$	
	Bridge (full-wave)	1.41 V_{rms}	0.90 V_{rms}	1.41 V_{rms}	0.43 V_{rms}	$2f_L$	
	Three-phase wye line to neutral (half-wave)	1.41 V_{rms}	1.17 V_{rms}	2.45 V_{rms}	0.21 V_{rms}	$3f_L$	
	(Y or Δ) Three-phase bridge line to line (full-wave)	1.41 V_{rms}	1.35 V_{rms}	2.45 V_{rms}	0.057 V_{rms}	$6f_L$	

*f_L = Line frequency

Fig. 5-12 Summary of rectifier circuits.



(a) Three-phase ripple (line to neutral)



(b) Three-phase ripple (bridge)

Fig. 5-13 Three-phase ripple voltage.

according to the type of circuit in which they are to be used. This is due to the heating effect of pulse current. Power dissipation varies as the square of the current ($P = I^2R$), and the high-current amplitude associated with pulse waveforms creates more heat loss in diodes than is indicated by the average current. A diode is capable of its greatest forward current at direct current (no pulsation) and its least forward current in six-phase star service (the narrowest pulse). The curves show that the same diode can support 12 A in single-phase service and only 8 A in six-phase service up to 150°C. However, do not forget the cell ratios. A single phase full-wave circuit shows a cell ratio of 0.5, and the six-phase star shows a cell ratio of 0.167. This means that six diodes of the type shown in Fig. 5-15 could provide a dc load current of up to 47.9 A [(1/0.167) × 8)] in six-phase star service, and two diodes could supply up to 24 A [(1/0.5) × 12] in single-phase, full-wave service. In actual practice, they would not be called upon to deliver their maximum currents as most circuit designers derate manufacturer's specifications to approximately 70 percent for good reliability.

You may be curious about the dc curve shown in Fig. 5-15. Since we have investigated the use of diodes to rectify (change alternating to direct current), the purpose of this curve may not be readily apparent. In addition to rectification, diodes can be used to isolate one circuit from another. Two dc power supplies can be used in parallel to supply current to a load when one supply cannot provide adequate current. Diodes in the output lines to the load prevent either supply from delivering a current

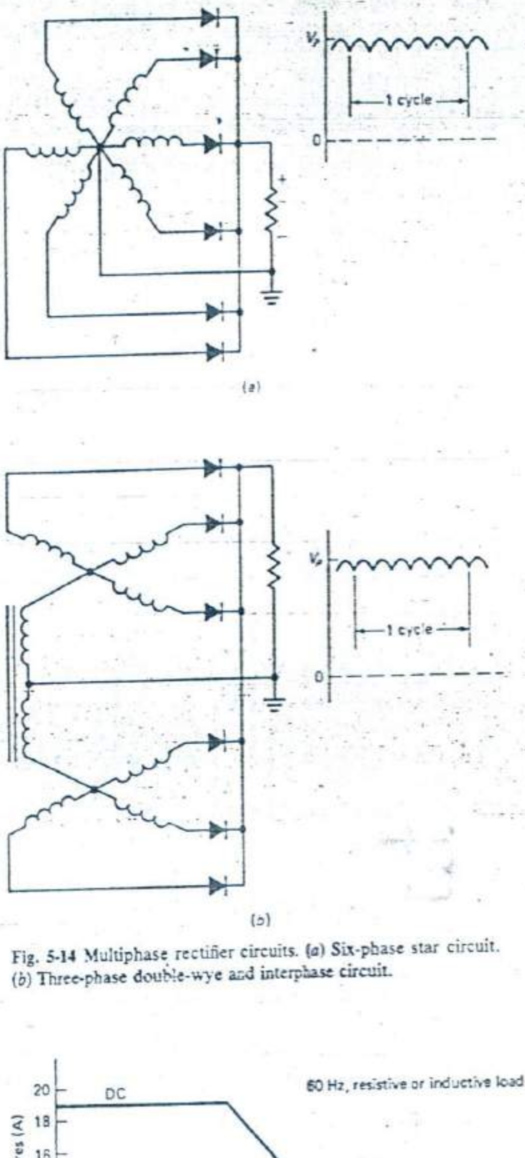


Fig. 5-14 Multiphase rectifier circuits. (a) Six-phase star circuit. (b) Three-phase double-wye and interphase circuit.

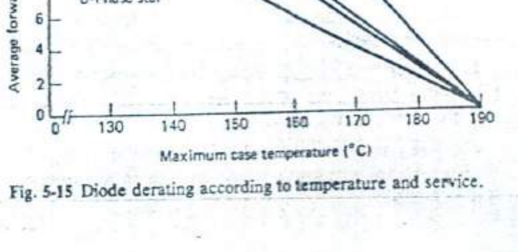


Fig. 5-15 Diode derating according to temperature and service.

to the other supply. This is an example of diode isolation and indicates one example where diode dc ratings would be appropriate.

The two most obvious rectifier diode specifications are their average forward current rating and their PIV rating. Others include the following:

1. Physical characteristics (mounting details, etc.)
2. Thermal characteristics
3. Power dissipation
4. Recovery time
5. Transient voltage rating
6. Avalanche rating
7. Current surge rating

Recovery time is a measure of how quickly a diode can stop conducting when suddenly reverse-biased. It takes time to sweep the carriers from the junction region, and high-frequency rectifiers must have a rapid recovery time or circuit efficiency will suffer and diode heating will be a problem. The *transient voltage* rating is a measure of the amount of nonrepetitive voltage stress a diode can withstand. It is especially useful in industrial environments in which large inductive loads are being switched. The *avalanche characteristic* occurs in controlled-avalanche rectifiers, in which excess voltages can be expected to produce predictable (controlled) results. Noncontrolled avalanche often results in the rectifier's being destroyed. Finally, *current surge ratings* are nonrepetitive current demands and are especially important when capacitor-input filters are used, as discussed in the next section.

REVIEW QUESTIONS

7. Refer to Fig. 5-9. Assume no diode loss and an rms secondary output of 50 V. Calculate the average dc voltage across the resistor. Hint: Refer to Fig. 5-12.

8. Again use Fig. 5-9 and the same assumptions as in question 7. Calculate the ac ripple voltage across the resistor.

9. Is the circuit of Fig. 5-10 a positive supply or a negative supply with respect to ground?

10. Refer to Fig. 5-10. Assume no diode loss and an rms secondary output of 30 V across the entire winding. Calculate the average dc voltage across the resistor.

11. Again use Fig. 5-10 and the same assumptions as in question 10. Calculate the ripple voltage across the resistor and the percentage of ripple.

12. Refer to Fig. 5-11. Assume a per-diode loss of 1 V and a 10-V (rms) secondary. Calculate the average dc output voltage.

13. Why is the diode in Fig. 5-15 rated up to 19 A in dc service and only up to 8 A in six-phase star service?

5-3
FILTERING

Except for applications such as welding, electroplating, battery charging, and motoring, the output of single-phase rectifiers contains too much ac ripple. A filter circuit will be needed to smooth the pulsating waveform and make it more like pure direct current (a straight line). Figure 5-16 shows some power supply filter circuits. They are used between the rectifier output and the load. They are examples of low-pass filters since they are designed to pass direct current (0 Hz) and to block the ac ripple (some multiple of the line frequency).

Figure 5-16 shows that power supply filters may be divided into two broad categories: capacitor input and choke (inductive) input. The capacitive input types are shown in Fig. 5-16(*a*). The rectifier output would be connected at the left of the circuits drawn. The left represents the input side of the filter. Notice that the filter input contains a shunt capacitor. Also notice that in Fig. 5-16(*b*) the inductive input filter types show a series coil (choke) at the filter input. Supply performance is affected by whether the input component is a capacitor or a choke.

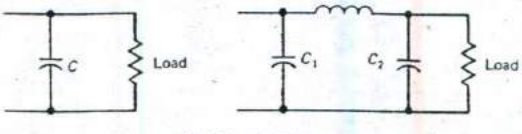
Capacitor input filters draw large current pulses from the rectifiers and from the transformer secondary. They produce a high load voltage when little current is drawn and much less voltage when full load is reached. This makes their voltage regulation poor. Voltage regulation is calculated as:

$$\text{Percent regulation} = \frac{\Delta V}{V_{FL}} \times 100$$

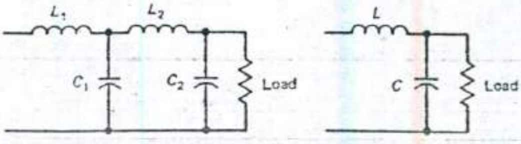
where ΔV = change in voltage from no-load to full-load
V_{FL} = full-load voltage

EXAMPLE

Suppose the output of a power supply drops from 13 V at no-load to 11 V at full-load. Find the percent regulation.



(*a*) Capacitive input



(*b*) Inductive input

Fig. 5-16 Power supply filter circuits.

SOLUTION

$$\text{Percent regulation} = \frac{\Delta V}{V_{\text{FL}}} = \frac{13 - 11}{11}$$
$$= \frac{2}{11} \times 100$$
$$= 18.2\%$$

Choke input filters lengthen the rectifier and transformer conduction time. This decreases the heating effect in these components for any given value of load current. They also show a lower output voltage and better voltage regulation when compared to the capacitor input types. In spite of these facts, the single-capacitor filter, shown in Fig. 5-16, has become very popular. The weight, size, and cost of inductors have eliminated them from most solid-state power supply designs. It is now possible to build supplies with adequate characteristics without using filter inductors because of the excellent current ratings of solid-state diodes, the improvements in electrolytic capacitors, and the performance of modern voltage-regulator circuits. Inductors are more attractive at high frequencies where far less core is required. Switch-mode power supplies use frequencies in the tens and hundreds of kilohertz, and filter inductors are employed in these designs. This topic is covered in section 5-7.

The full-wave, center-tapped circuit in Fig. 5-17(*a*) has a single capacitor filter in parallel with the load. The voltage waveform across this parallel combination (Fig. 5-17[*b*]) shows that the capacitor charges to the peak value of the ac input ($1.41 \times V_{\text{rms}}$). If the capacitor is large enough, it will hold the load voltage near this peak value over the period of the half cycle until the next rectifier pulse comes along. The capacitor has eliminated most of the ac ripple across the load. The remaining ripple is designated as ΔV_o and resembles a sawtooth waveform. The current waveforms (Fig. 5-17[*c*] and [*d*]) indicate that the rectifiers conduct for only short periods of time. Of course, the transformer secondary current is also of a pulse nature. The amplitudes of the current pulses are about five times greater than the load current in a typical power supply of this type. The heating effect is much greater for pulse-type waveforms, and the components must be derated when capacitor input filters are used. For example, the transformer in this circuit should not be expected to deliver more than 80 percent of its rated current.

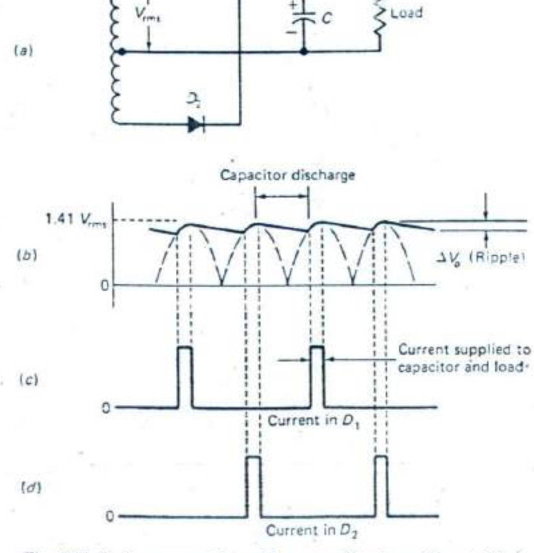


Fig. 5-17 Full-wave rectifier with a capacitive input filter. (*a*) Rectifier circuit. (*b*) Voltage across the capacitor. (*c*) Current in diode D_1. (*d*) Current in diode D_2.

The current pulses can be understood by examining the waveforms in Fig. 5-17. The stored voltage across the capacitor keeps both diodes reverse-biased most of the time. Neither diode will begin conducting until the peak secondary voltage exceeds the capacitor voltage by about 0.6 V. This 0.6-V difference is required to collapse the diode depletion region. At the time of turn-on, the diode supplies charging current to the capacitor and some current to the load. The diode continues to conduct until the secondary voltage reaches peak. It should be clear that the diode current is relatively short in duration under these conditions and that extra stress is placed on some power supply components. Extra voltage stress is also created in three-phase rectifier circuits. The PIV is increased from $2.45 \times V_{\text{rms}}$ to $2.82 \times V_{\text{rms}}$ with a capacitive input filter.

Selecting the capacitance for a simple filter as shown in Fig. 5-17 depends on three factors: ripple frequency, load current, and allowable ripple voltage. It is accomplished by

$$C = \frac{I}{\Delta V} \times T$$

where C = capacitance, F
I = load current, A
ΔV = peak-to-peak ripple, V
T = ripple period, s

EXAMPLE

Suppose a capacitor is needed for a 10-A power supply. Also assume that the supply is full-wave, runs from the 60-Hz single-phase ac line, and may have 1-V peak-to-peak ripple. Find the value of the capacitor.

SOLUTION

The ripple frequency is twice the line frequency in this case, and the period may be found by

$$T = \frac{1}{F}$$
$$= \frac{1}{120}$$
$$= 8.33 \text{ ms}$$

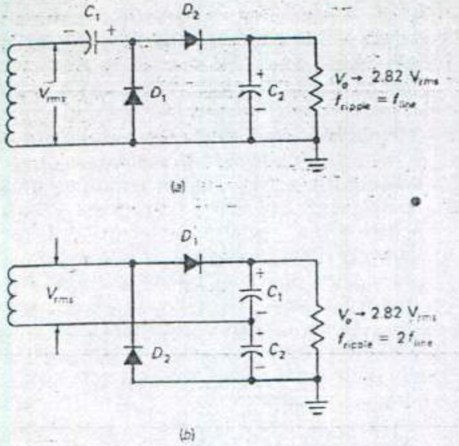


Fig. 5-18 Voltage multipliers. (*a*) Half-wave doubler. (*b*) Full-wave doubler.

Now the capacitance can be found by

$$\begin{aligned} C &= \frac{10}{1} \times 8.33 \times 10^{-3} \\ &= 0.0833\ \text{F} \\ &= 83{,}300\ \mu\text{F} \end{aligned}$$

There are two other important capacitor ratings. The first is the voltage rating. For safety, the voltage rating must be somewhat greater than the peak secondary voltage. The second rating is the current capability of the capacitor. The capacitive ripple current (for an input capacitor) is approximately 2.5 times the load current. This amounts to 25 A for the previous example. This ripple current heats the capacitor. Heating shortens the life of the capacitor, and this factor must be taken into account for good reliability.

Voltage Multipliers

Capacitive filters also lead the way to voltage multipliers. Figure 5-18(*a*) is a half-wave doubler. Assume the first alternation makes the top of the secondary negative. This will forward-bias D_1, and C_1 will be charged to the peak secondary voltage. The next alternation will forward-bias D_2, and the stored charge across C_1 will series add with the secondary voltage. Also, C_2 will be charged to twice the peak secondary voltage. The load will see a peak voltage of $2 \times 1.41 = 2.82\ V_{rms}$. The ripple frequency is equal to the line frequency, since C_2 and the load receive a line pulse once per cycle. The full-wave doubler in Fig. 5-18(*b*) provides two line pulses per cycle. Capacitor C_1 is charged through D_1 to the peak secondary voltage, and C_2 is charged through D_2 to the peak secondary voltage. The capacitors are in series, and the load sees twice the peak secondary voltage. Both capacitors are in the load circuit, and both alternations pulse the load; thus full-wave operation is realized and the ripple frequency is twice the line frequency.

The half-wave doubler in Fig. 5-18(*a*) is safer in some applications because it allows one side of the ac source to be grounded. Neither side of the ac source may be grounded in the full-wave doubler circuit. Grounding the chassis is possible in line-operated equipment (no on-board isolation transformer) with the half-wave doubler and can prevent a "hot" chassis and ground loops. This topic is covered in more detail in a later section of this chapter.

Voltage doublers can provide a dc voltage that is nearly three (2.82) times the ac input. Their output drops rapidly under load, however, and double the rms input is normal under working conditions. They are noted for poor voltage regulation. Figure 5-19 shows that high-order voltage multiplication is also possible; C_1 is charged through D_1 to the peak voltage V. Then C_2 is charged through D_2 with C_1 series aiding the secondary voltage. Now C_3 is charged through D_3, and capacitors C_1 and C_2 series aid the secondary. Thus, a voltage equal to three times V is available at the cathode of D_3. If the circuit shown in Fig. 5-19 is to be used to supply any of the evenly multiplied voltages ($2V$, $4V$, $6V$), the ground will have to be moved to the top of the secondary. In theory, any multiplication factor is possible. In practice, circuit efficiency limits high-order multiplication to very-low-current applications such as in cathode-ray tube supplies.

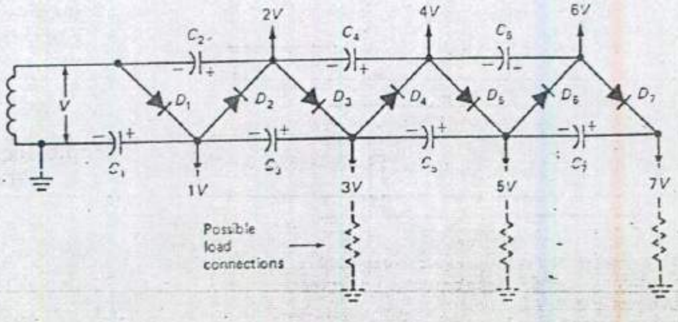


Fig. 5-19 High-order voltage multiplier.

REVIEW QUESTIONS

14. A power supply develops 48 V at no-load and 40 V at full-load. Calculate its regulation.

15. Which broad category of power supply filter develops a high no-load output voltage?

16. Which broad category of power supply filter is noted for poor voltage regulation?

17. Which broad category of power supply filter causes high peak rectifier current?

18. Refer to Fig. 5-17. Assume no diode loss, a very small load current, and a 30-V_{rms} secondary. Calculate the dc load voltage.

19. What will happen to the dc load voltage calculated in question 18 if the load current increases?

20. What will happen to the percentage of ripple in the circuit of question 18 if the load current increases?

5-4

VOLTAGE REGULATION

Voltage regulation is the ability of a power supply to hold its output potential constant under conditions of changing line voltage, temperature, and load current. Some industrial circuits are critical and demand voltage regulation of less than 1 percent. These circuits will require power supplies with voltage regulators.

Line voltage varies with demand. A *brownout* is a condition in which the demand is so high that the power company is forced to reduce the line voltage intentionally to protect its equipment. Industrial customers are not protected from brownouts and may even experience a greater reduction in line voltage than residential consumers.

When line voltage is abnormally low, the output from a power supply will also tend to be low. The ability of a power supply to hold a constant output over a range of line input voltage is termed *line regulation*. One way to achieve line regulation is to use a ferroresonant power transformer. These transformers are designed with separate core windows for the primary and secondary windings. They have two magnetic circuits: the main flux path and the shunt flux path. The main flux path couples the primary and secondary circuits. The secondary circuit is tuned to resonate at the line frequency by connecting a capacitor across a part of or all of the secondary. The value of this capacitor is several microfarads in 60-Hz supplies. The Q of the tuned circuit is high enough to cause large circulating currents that saturate the core in the main flux path. *Saturation* is a decrease or increase in magnetizing force that is not accompanied by a corresponding change in flux density. With the main flux path saturated, line voltage fluctuations will not change the main flux density, and the secondary voltage will remain constant. Of course, if the primary voltage drops too low, the core can come out of saturation, and the secondary voltage will drop.

The shunt magnetic flux path in a ferroresonant transformer is prevented from saturating by air gaps placed in its magnetic circuit. Air has a much higher reluctance than transformer iron, and this characteristic limits the flux in this part of the magnetic circuit. If a greater demand is placed on the secondary circuit, the Q of the secondary tuned circuit is reduced, and therefore the circulating currents are also decreased. Since the shunt path is linear (nonsaturated), it can respond to this change, and a decrease in flux results. With fewer lines of force in the shunt circuit, the main flux density can increase; more energy is transferred from primary to secondary to make up for the increased demand on the secondary circuit. Thus, the ferroresonant transformer also regulates for load changes. Unfortunately, the size, weight, and cost of these transformers eliminate them from many designs, but they are counted among the most reliable voltage regulators available.

A more common way to achieve line and load regulation is to use a separate regulator circuit after the power supply filter. In Fig. 5-20, a simple regulator uses a zener diode (D) in shunt with the load. As long as the zener operates in its constant voltage region, the load voltage will also remain constant. Zener diodes can generate noise (especially when operating near the knee) so that a capacitor is often included in this circuit to bypass the noise from the load. This circuit will operate properly over a range of load and line conditions. If the unregulated input voltage drops too low, the zener will stop conducting and regulation will be lost. If the load demand goes too high, the drop across series resistor R will increase to the point where the zener stops conducting.

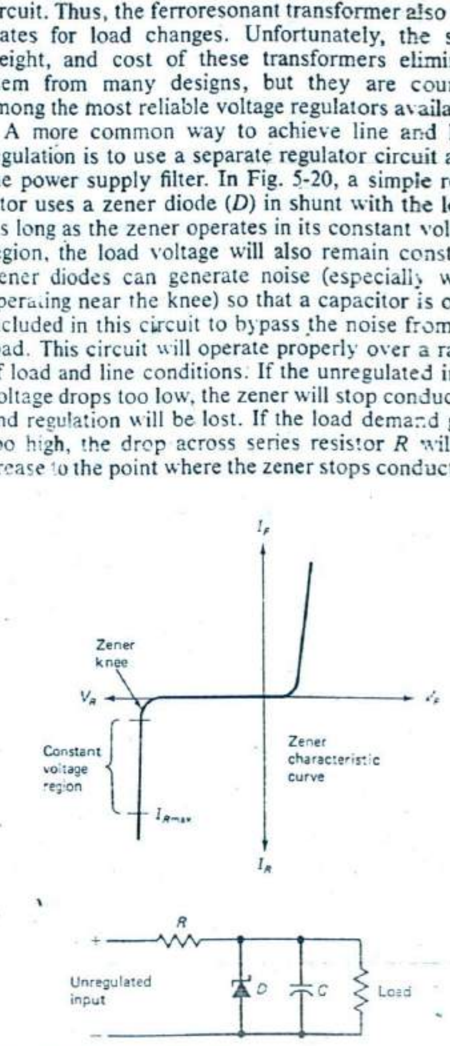


Fig. 5-20 Shunt zener regulator.

and once again regulation will be lost. Another problem is that the safe zener dissipation may be exceeded if the load demand drops to zero or if the unregulated input voltage goes too high.

The regulator circuit of Fig. 5-20 is limited to low-power applications. A few simple calculations will illustrate why.

EXAMPLE

Suppose a regulated 15 V at 2 A is required and the unregulated input is 20 V.

SOLUTION

The zener current is often set at half the load current in this design so it should be 1 A. Ohm's law is used to calculate the required series resistor R. It will conduct the load current plus the zener current $(2 + 1 = 3\ A)$ and will drop the difference between the input voltage and the load voltage $(20 - 15 = 5\ V)$:

$$\begin{aligned} R &= \frac{V}{I} \\ &= \frac{5}{3} \\ &= 1.67\ \Omega \end{aligned}$$

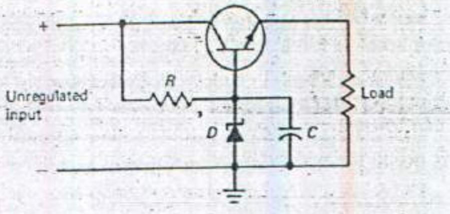
The zener dissipation will be equal to its voltage drop times its current. However, if there is any possibility that the load current can drop to zero the diode will have to conduct all of the current:

$$\begin{aligned} P &= V \times I \\ &= 15 \times 3 \\ &= 45\ W \end{aligned}$$

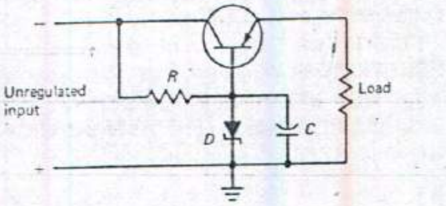
A 45-W zener diode is unacceptable because of its cost, and the efficiency of the circuit is poor when the load current is low. This is why the shunt zener regulator is limited to low-power applications.

Figure 5-21 shows positive and negative regulators that are an improvement over the shunt circuit. In Fig. 5-21(*a*) a zener diode and an NPN transistor form a positive voltage regulator. The transistor is called a *series pass transistor* in this application: It passes the load current from collector to emitter. The zener is used to regulate the base voltage of the transistor. Regulating the base voltage also regulates the emitter voltage if we can assume a constant base-emitter drop in the transistor. If the emitter voltage is regulated, then the load voltage is also regulated in Fig. 5-21. The current gain of the transistor greatly relaxes the zener dissipation. If we assume a gain of 50 from the base to the emitter and a load current of 2 A, the base current will be 40 mA (2/50). Now the zener current can be set to half this value, or 20 mA. If we again assume a 20-V unregulated input and a 15-V zener, resistor R is now

$$\begin{aligned} R &= \frac{5}{0.06} \\ &= 83.3\ \Omega \end{aligned}$$



(*a*) Positive regulator circuit



(*b*) Negative regulator circuit

Fig. 5-21 Series pass circuits.

Notice that the resistor is much larger than in the shunt regulator. Once again, the worst-case zener dissipation occurs if the load current drops to zero. In the case of Fig. 5-21, with no-load current there can be no emitter current and no base current. The zener will now conduct 60 mA for a dissipation of

$$\begin{aligned} P &= 15 \times 0.06 \\ &= 0.9\ W \end{aligned}$$

Now this is far more acceptable. The cost of the zener is reasonable, and circuit efficiency is much better at low load currents. The circuit can also be arranged for regulating negative voltages. This is shown in Fig. 5-21(*b*). Note that the zener is reversed and that the series pass transistor is a *PNP* device.

The shunt circuit of Fig. 5-20 and the series pass circuits of Fig. 5-21 will not develop the same load voltages with a 15-V zener diode. The series pass circuit will have a lower output since the pass transistor will show some drop from base to emitter. This is usually 0.7 V in a silicon transistor that is conducting moderate currents. Therefore, the load voltage can be expected to be 0.7 V less than the zener voltage in Fig. 5-21. However, at high load currents the base-to-emitter drop is going to be greater. A series-pass transistor that is conducting 5 A will show a base-emitter voltage closer to 1.2 V, and the output voltage will then be 1.2 V less than the zener voltage. So the problem with the circuits shown in Fig. 5-21 is that the no-load to full-load voltage will drop a half volt or more in a 5-A supply. This amount of voltage change may not be acceptable in some applications.

A feedback regulator with better performance is shown in Fig. 5-22. The output is sampled by the voltage divider R_2 and R_3, and some portion of the load voltage is fed back to the base of Q_2, which acts as an error amplifier. The feedback voltage is compared to the zener voltage, which acts as a reference. Any error between the two voltages is amplified and used to reduce the error. Suppose the load voltage

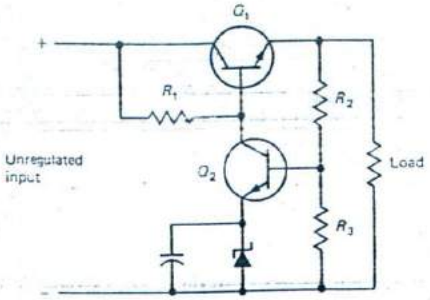


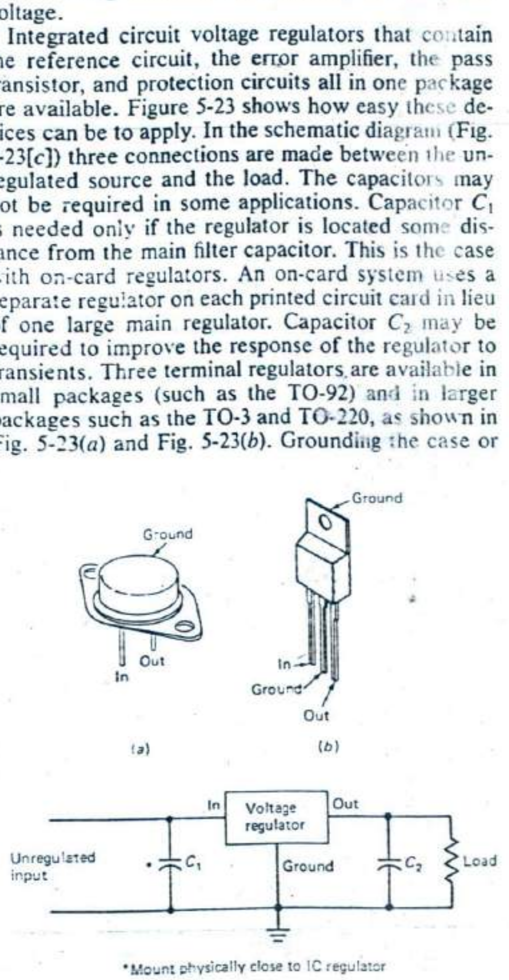
Fig. 5-22 Feedback regulator.

in Fig. 5-22 is 15 V and that $R_2 = R_3$. They will divide the 15 V in half, and therefore the base voltage of Q_2 will be +7.5 V with respect to ground. If the zener is a 6.8-V unit, the emitter of Q_2 will be +6.8 V. The base-emitter bias of Q_2 will therefore be 0.7 V (7.5 − 6.8), and there will be moderate emitter and collector current in Q_2. Resistor R_1 conducts two currents: the base current for Q_1 and the collector current for Q_2. Now, suppose the load demands more current. As always, this will tend to make the output voltage drop. This drop will reduce the base voltage of Q_2. The emitter voltage is constant because of the zener, and now there will be less voltage from base to emitter in Q_2. As V_{be} decreases, so must base current I_b. As I_b decreases, so must collector current I_c. Finally, with Q_2 demanding less collector current, R_1 can supply more current to the base of Q_1 and turn it on harder. When a series-pass transistor is turned on harder it has less resistance and will drop less voltage from collector to emitter. In Fig. 5-22, if V_{ce} in Q_1 drops, more of the unregulated input voltage must drop across the load. Thus, the load voltage has been stabilized by the feedback.

The application of feedback achieves good voltage regulation. It also provides ripple rejection since ripple is also an error in output voltage. Feedback is the basis for most regulator designs. It is based on sampling the output and comparing it to some reference. Any change in output, including ripple, is amplified by an error amplifier. The error amplifier then controls a series-pass transistor to eliminate most of the error. If the output is low, the pass transistor is turned on harder. If the output is high, the pass transistor is turned on less. Using feedback, it is possible to build regulated supplies that show no significant change in output over the full range of load currents and over some range of the unregulated input. It should be emphasized that in circuits such as that in Fig. 5-22 the unregulated input must be at least 2 V greater than the regulated load voltage. Otherwise, the pass transistor will saturate, and control will be lost.

The regulation and ripple rejection of a feedback regulator is related to the gain of the error amplifier. A high-gain error amplifier will provide very good regulation and ripple rejection. However, the accuracy of the output can be no better than the accuracy of the reference voltage. Some power supplies use an integrated circuit zener diode in place of an ordinary zener diode. These integrated circuits are also called *reference diodes* and provide a much more accurate reference voltage. For example, an ordinary zener 6.9-V shunt regulator will show a 17-mV change in output when its input changes 10 percent. An equivalent IC zener, the LM129, will show only a 180-μV change under the same conditions. The IC references are also more temperature-stable than ordinary zener references. The schematic symbol used to represent a reference diode or IC zener is the same as the ordinary zener symbol. The package may be a small type such as the TO-92 with two leads. Thus, it is possible to misidentify an IC reference and replace it with an ordinary zener. This will cause considerable degradation in the accuracy of the output voltage.

Integrated circuit voltage regulators that contain the reference circuit, the error amplifier, the pass transistor, and protection circuits all in one package are available. Figure 5-23 shows how easy these devices can be to apply. In the schematic diagram (Fig. 5-23[*c*]) three connections are made between the unregulated source and the load. The capacitors may not be required in some applications. Capacitor C_1 is needed only if the regulator is located some distance from the main filter capacitor. This is the case with on-card regulators. An on-card system uses a separate regulator on each printed circuit card in lieu of one large main regulator. Capacitor C_2 may be required to improve the response of the regulator to transients. Three terminal regulators are available in small packages (such as the TO-92) and in larger packages such as the TO-3 and TO-220, as shown in Fig. 5-23(*a*) and Fig. 5-23(*b*). Grounding the case or



*Mount physically close to IC regulator

(*c*)

Fig. 5-23 Integrated voltage regulator. (*a*) TO-3 package. (*b*) TO-220 package. (*c*) Regulator circuit.

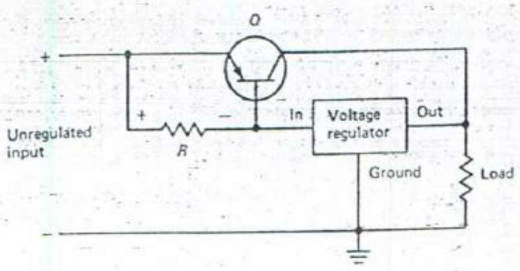


Fig. 5-24 Current boost circuit.

tab of the package is shown only as a general practice since these connections are inputs on some devices. The small package regulators have ratings of around 100 mA, the TO-220 devices as high as 3 A, and some TO-3 devices as high as 10 A. They are available with several fixed voltages such as in the 78XX series. The 7805 is a fixed 5-V regulator, the 7812 regulates at 12 V, the 7815 at 15 V, and so on. They are available as negative regulators in the 79XX series (for example, the 7905 is a negative 5-V regulator). Many integrated voltage regulators are adjustable, or can be configured for adjustment, and some are designed for switch-mode service, as will be seen in section 5-7.

The current of an integrated circuit voltage regulator can be boosted by adding an external PNP pass transistor to a fixed positive regulator as in the circuit of Fig. 5-24. The integrated circuit will supply all the load current up to the point where the current in resistor R is enough to cause a voltage drop of 0.6 V. At this point, the transistor will be turned on, and it too will help support the load current. If R is 4.7 Ω, the turn-on current will be

$$I = \frac{0.6}{4.7}$$
$$= 128 \text{ mA}$$

As the load demands more and more current, the drop across R will continue to increase and turn the pass transistor on harder. Figure 5-24 shows that the polarity of the drop across R is correct for forward-biasing the base-emitter junction of the PNP transistor. A similar circuit arrangement is possible with a negative regulator and an NPN pass transistor.

Fixed integrated voltage regulators can also be used to supply variable output voltages. In this service, the ground terminal is connected to some reference voltage rather than directly to ground. Figure 5-25 shows a dual-complementary supply with adjustable output voltages. The supply is *dual-complementary* since it provides both positive and negative voltages with respect to ground. The rectifier circuit is a combination of two full-wave, center-tapped supplies. This arrangement is also called a *center-tapped bridge rectifier*. Capacitors C_1 and C_2 filter the positive and negative voltages. The 7805 is a fixed positive 5-V integrated regulator, and the 7905 is a fixed negative 5-V regulator. However, the supply is adjustable over a range of 5 to 20 V positive and negative with respect to ground. Notice that the ground terminals of the fixed regulators are driven by triangular symbols. The triangle is commonly used in schematics and block diagrams to represent an amplifier. The inputs of the amplifiers in Fig. 5-25 are at the right and are marked with minus (−) and plus (+) symbols. The minus input is called the *inverting input*. Any positive-going signal applied here will cause the output to go in a negative direction, and any negative-going input will drive the output in a positive direction. Notice that the wiper arm of R_4 supplies the inverting input of the top amplifier. If the wiper arm is moved toward R_5, more of the negative output voltage will be applied to the top circuit. This negative-going signal is inverted by the top amplifier, and the ground lead of the 7805 is driven in a positive direction. The 7805 will develop an output that is 5 V plus the positive voltage sup-

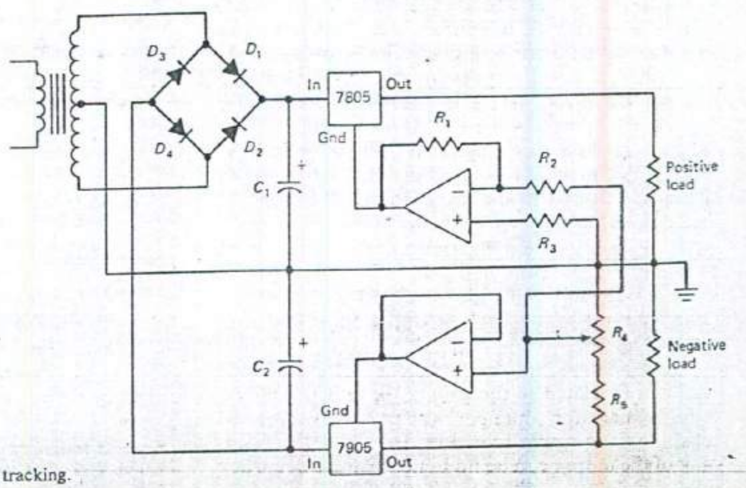


Fig. 5-25 Dual complementary supply with tracking.

plied to its ground lead. This makes the positive supply adjustable by changing the setting of R_4.

What happens to the negative voltage in Fig. 5-25 when R_4 is adjusted? The wiper arm of R_4 also supplies a signal to the bottom amplifier. In this case the signal is supplied to the plus input, which is a noninverting input. When the wiper arm is moved toward R_5 the resultant negative-going signal is amplified and drives the ground lead of the 7905 in a negative direction. The negative output voltage will be -5 V plus the negative voltage supplied to its ground lead. This makes the negative supply voltage adjustable by changing the setting of R_4.

Both outputs shown in Fig. 5-25 are adjustable with R_4. When the wiper arm of R_4 moves toward R_5, the positive output goes more positive and the negative output goes more negative. One control adjusts both supplies. This is known as a *tracking supply*. The positive output voltage tracks the negative output voltage. Supplies can be single or can have more than one output. If they have two outputs, they can be complementary (one positive and one negative). Supplies can be fixed or adjustable. Finally, supplies that have more than one output can be independent or tracking.

REVIEW QUESTIONS

21. Refer to Fig. 5-20. Assume 5 V across the load, a load current of 100 mA, an unregulated input of 8 V, and a zener current of half the load current. Calculate the resistance of R and its power dissipation.

22. Use the data of question 21 and calculate the zener dissipation with the load connected and with the load disconnected.

23. Use the data of question 21. At what load current will the circuit stop regulating? What happens to the zener current at this point? What happens to the output voltage if the load current increases even more?

24. Use the data of question 21. At what unregulated input voltage will the circuit stop regulating? (Hint: Try calculating the load resistance and draw an equivalent circuit.)

25. Refer to Fig. 5-21(b). Assume an unregulated input of 10 V, a current gain from base to emitter of 80, a 5.7-V zener, and a load current of 100 mA. Calculate a value for R that will set the zener current to half the base current.

26. Use the data of question 25. What load voltage can be expected?

5-5
CURRENT REGULATION

Most power supplies operate in the constant-voltage mode. They maintain a fixed load voltage over a range of load currents. A constant-current supply will maintain a fixed load current over a range of load resistances. Constant-current supplies are useful for charging batteries, for supplying bias currents in reference circuits, for energizing electromagnets, and for performing various control applications. Many constant-voltage supplies also have a constant-current mode. This mode is useful to protect the supply and the circuits it energizes in the event of a fault such as a short circuit. A supply that reacts to an overload by changing from constant-voltage to constant-current operation is said to have *automatic crossover*.

Figure 5-26 shows three constant-current circuits. Each will supply a fixed current over a range of load resistance. At some high value of load resistance, the current will fall off as the load voltage cannot exceed the input voltage. The output voltage is called the *compliance voltage*, and the *compliance range* predicts the values of load resistance that will receive a constant current. Figure 5-26(a) is a field effect transistor (FET); the load current will be equal to its I_{DSS} rating. The FETs (constant-current diodes) may be used to supply a constant bias current to zeners or IC zeners for a more stable reference voltage. A transistor circuit in which the drop across the emitter resistor is zener-regulated is depicted in Fig. 5-26(b). The constant voltage across R_1 predicts a constant

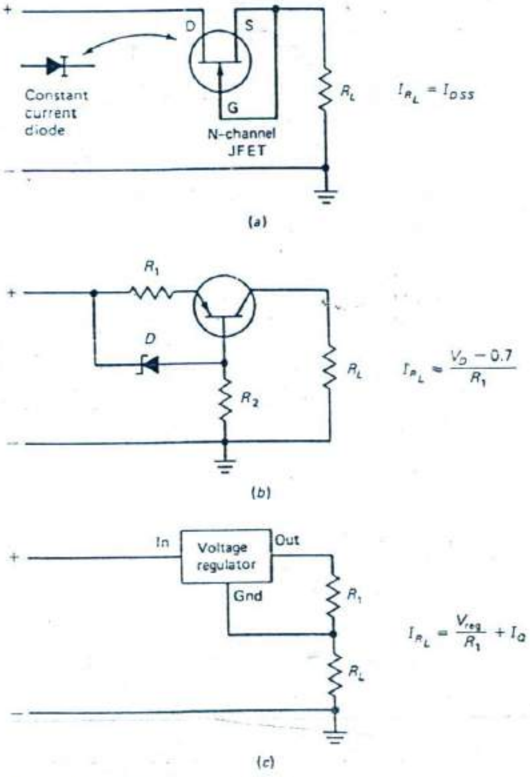


Fig. 5-26 Constant-current circuits.

current in R_1 and also in the emitter and collector circuits of the transistor. Figure 5-26(*c*) shows how an IC voltage regulator can be used to provide a constant current. The load current will be equal to the regulator output voltage divided by R_1 plus the quiescent drain current of the IC.

Figure 5-27 shows a constant-voltage power supply with current limiting. It uses an error amplifier to compare a reference voltage against a sample of the load voltage. Any change in load voltage will cause the amplifier to bias the pass transistor so as to reduce the error. Resistor R_1, diode D, and capacitor C form a stable reference voltage in Fig. 5-27. This reference voltage is fed to the noninverting input of the amplifier. The inverting input of the amplifier is connected to the junction of R_4 and R_5, which divide the load voltage. The amplifier output drives Q_2, which is direct-coupled to Q_1, the pass transistor. The direct connection from the emitter of Q_2 to the base of Q_1 produces a high overall current gain, and the two transistors are usually called a *Darlington pair.* Transistor Q_3 is off when the supply is operating in the constant voltage mode and has no effect on circuit performance. If the load should demand less current, the load voltage will tend to increase. The divided voltage at the inverting input of the error amplifier will also increase. When a positive-going signal appears at the inverting input, the output goes less positive. This means less drive to the Darlington pair, and the pass transistor now drops more voltage. This action decreases the load voltage, and much of the original error is eliminated.

The constant-voltage mode of Fig. 5-27 will range from zero load current to that value of load current which causes enough drop across R_3 to turn on Q_3. Resistor R_3 is in series with the load and can be considered a current-sensing resistor. When the load current goes high enough to drop about 0.6 V across R_3, Q_3 will turn on, assuming it is a silicon transistor. Notice that the polarity across R_3 forward-biases the base-emitter junction of Q_3. When Q_3 comes on, it loads the output of the error amplifier. Current that normally was supplied through R_2 to the base of Q_2 is now supplied to Q_3. This reduces the drive to the Darlington pair, and the output voltage drops. The load resistance can continue to decrease toward short-circuit conditions, and the load current stays reasonably constant. Power supplies that cross over from constant-voltage to constant-current operation at some value of load current employ conventional current limiting.

Conventional current limiting applied to the current boost circuit studied earlier is shown in Fig. 5-28. Most integrated circuit voltage regulators have internal current limiting and protect themselves from overloads. When they are current-boosted, the internal current limiting will not protect the pass transistor. Figure 5-28 shows how two components can be added to provide current limiting for pass transistor Q_1. Resistor R_2 is the current-sensing resistor. It senses that portion of the load current supplied by pass transistor Q_1. When Q_1's current goes high enough, the drop across R_2 reaches 0.6 V and turns on Q_2. When Q_2 is on, it acts in parallel with R_1 to decrease its effective resistance. Less resistance means less voltage drop across R_1 and less bias for pass transistor Q_1. Transistor Q_1 will now drop more voltage from its collector-to-emitter terminal, and the load voltage will begin to drop. The supply has crossed over from constant-voltage to constant-current operation. The maximum load current in this circuit will be equal to the limiting current of the integrated circuit plus the limiting current of the pass transistor.

Conventional current limiting may not always protect the pass transistor and other components from damage. A sustained short circuit will create high-power dissipation in the pass transistor. For example, a 2N3055 pass transistor is rated at 15 A and a collector dissipation of 115 W. If it is operated in a 5-A 12-V power supply with conventional current limiting, it may seem that it should be safe under all load conditions. However, it can be destroyed by excessive collector dissipation. Suppose it is operated with an unregulated input of 18 V. If the load is

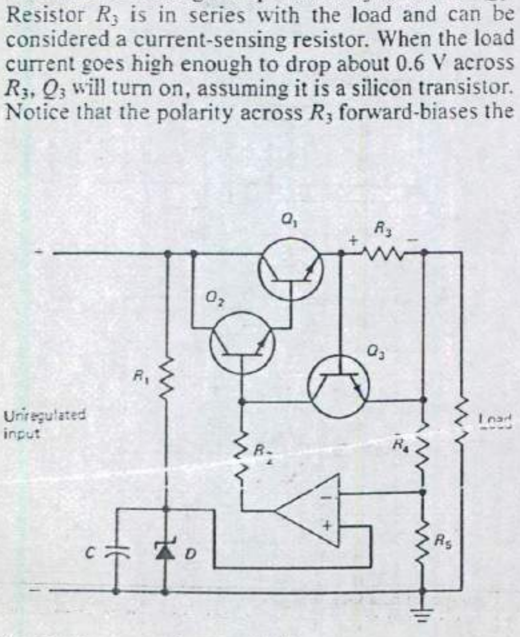


Fig. 5-27 Conventional current limiting.

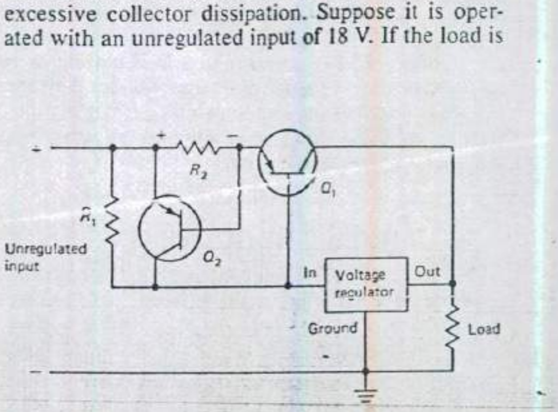


Fig. 5-28 Conventional current limiting added to a current boost circuit.

a short circuit, all of the input will drop across the pass transistor. The transistor dissipation will be

$$\begin{aligned} P_c &= V_{ce} \times I_c \\ &= 18 \times 5 \\ &= 90 \text{ W} \end{aligned}$$

It still seems safe since 90 W is less than its rated 115 W. However, its rated dissipation is for a case temperature of 25°C (77°F). When a transistor is dissipating 90 W it becomes very hot. Figure 5-29 shows the derating curve for the transistor in question. Notice that the curve shows 90 W of dissipation to be maximum at a temperature of 63°C. The transistor may be destroyed if the short circuit lasts long enough for the case to exceed this temperature.

Another form of current limiting may be more desirable, especially when long-term overloads are expected. Foldback current limiting provides better protection for the pass transistor and other circuit components. Figure 5-30 compares the graphs for two types of current limiting. Notice that in Fig. 5-30(*a*) the supply operates at a constant output of 12 V up to load currents just over 5 A. This is the constant-voltage region. As the load resistance drops, the current demand goes beyond 5 A, and the output voltage starts to drop. It continues to drop until it reaches 0 V at short-circuit conditions. There is little current change from the beginning of limiting to the short-circuit condition. In foldback limiting (Fig. 5-30[*b*]) the constant-voltage region is the same. As the load demands more than 5 A, the current starts to fold back (decrease). At short-circuit conditions, the current is 2 A. The dissipation in the pass transistor can now be calculated by using the same conditions as before but with a foldback current of 2 A:

$$P_c = 18 \times 2 = 36 \text{ W}$$

This is a much more reasonable dissipation. According to Fig. 5-29 the transistor is safe up to a case temperature of 140°C. A moderate heat sink will be able to accomplish this, and the pass transistor will be safe even if subjected to a prolonged short circuit.

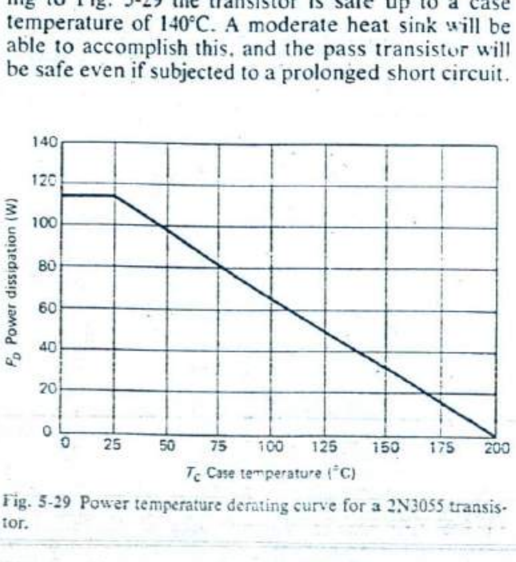


Fig. 5-29 Power temperature derating curve for a 2N3055 transistor.

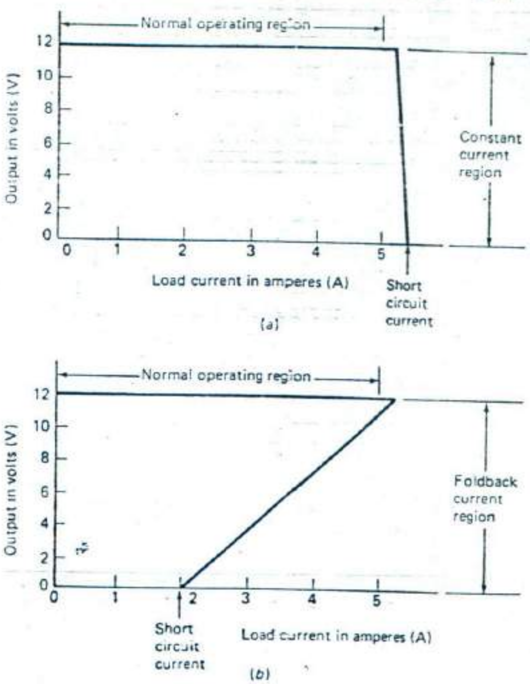


Fig. 5-30 Conventional current limiting compared to foldback current limiting. (*a*) Conventional current limiting. (*b*) Foldback current limiting.

Figure 5-31 shows a circuit that employs foldback current limiting. Transistor Q_3 is off under normal load conditions. Components R_1, D, and C develop a stable reference voltage for the noninverting input of the error amplifier. Resistors R_6 and R_7 divide the

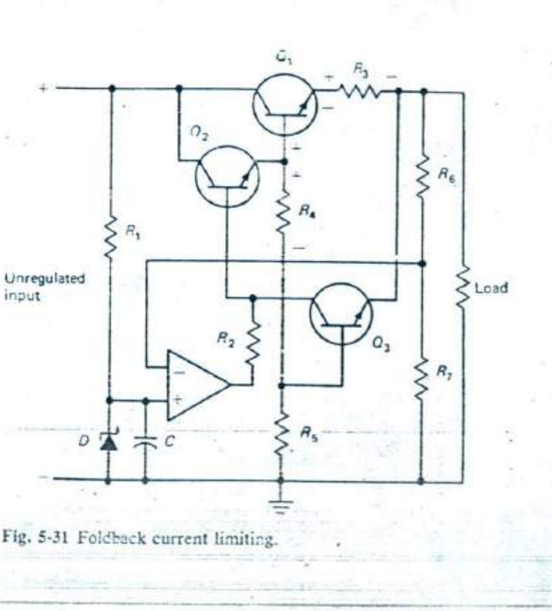


Fig. 5-31 Foldback current limiting.

load voltage for the inverting input. Any error produces more or less drive to the Darlington pair, Q_2 and Q_1, and the supply operates as a constant voltage source. If the load demand continues to increase, the drop across R_3 (the current-sensing resistor) and V_{be} of Q_1 will increase. Note that the polarity of these two drops provides forward bias for the base-emitter circuit of Q_3. Also notice that the drop across R_4 provides a reverse bias for Q_3; Q_3 will not come on until the two forward drops can overcome the drop across R_4 and the required junction voltage of Q_3. When Q_3 comes on, the limiting action begins. Now Q_3 can divert drive current from the pass circuit. This causes the output voltage to begin dropping. It also reduces the voltage across R_4 and R_5 since the base voltage of Q_1 tracks the output. With less drop across R_4, Q_3 turns on harder since the R_4 drop is a reverse-bias source for Q_3. The output is now folding back since Q_3 is conducting more than it did before with a given voltage across the current-sensing resistor. If the load becomes a short circuit, the current in R_4 approaches zero and no longer acts to produce any reverse bias for Q_3. Now, only a fraction of the rated supply current is required in R_3 to keep Q_3 on.

Current-limiting circuits of each type can "latch-up" under certain conditions, although the foldback type is more likely to do so. *Latch-up* occurs when the supply is turned on and the output fails to come up to its full value. It latches at some point on its limiting curve. Some loads present a low resistance at turn-on; for example, an incandescent lamp shows a very low resistance when cold. This type of load can latch the supply at some point on its current-limiting curve. With the current limited, the lamp will never reach its normal operating temperature or its normal operating resistance, and the output will remain latched at some voltage lower than normal. Dual-complementary power supplies may also latch up if one polarity supplies a bias to a circuit energized by the other polarity. At turn-on, the bias is missing, the circuit draws excessive current, and the supply goes into limiting. If the bias side uses voltage from the limiting side, the bias may never reach a normal level, and the limiting action will continue indefinitely.

REVIEW QUESTIONS

27. Refer to Fig. 5-27. Assume R_3 is a 0.15-Ω resistor and Q_3 is a silicon transistor. At what load current will the limiting action begin?

28. Use the same conditions as in question 27. If the output is 12 V, over what range of load resistance will the supply act as a voltage source?

29. Use the same conditions as in question 28. Over what range of load resistance will the supply act as a current source?

30. Refer to Fig. 5-28. Assume that the integrated circuit regulator is internally limited to 1.5 A, that Q_2 is a silicon transistor, and that R_2 is a 0.07-Ω resistor. Calculate the maximum load current.

31. Use the information provided by Fig. 5-29 and Fig. 5-30(*b*). Calculate the maximum safe case temperature if the unregulated input is 15 V and the load is a short circuit.

32. Use Fig. 5-30(*a*) and predict the load current for a 2-Ω load resistor. (Hint: Use Ohm's law and a graphical approach that satisfies a slope of 2 Ω.)

5-6
PROTECTION DEVICES AND CIRCUITS

The current regulation circuits discussed in the previous section provide protection for the power supply and load circuits for certain kinds of faults. This section examines some other devices and circuits that offer additional protection. Fuses and circuit breakers are covered in Chapter 4. You may wish to review the discussion of these devices since they are commonly used to protect power supplies and other industrial circuits.

Line transients are common in the industrial environment. They are caused by lightning, switching of large inductive loads such as motors, and malfunctions in other parts of the industrial plant. They can cause all sorts of damage to wiring and equipment, and semiconductors are especially susceptible. Studies show that one 5000-V transient can be expected per year for every 120-V circuit in this country; lower voltage transients can also be expected even more frequently. Fuses and circuit breakers are too slow-acting to prevent transient damage.

A varistor (Fig. 5-32) can be used to absorb transient energy safely and to protect sensitive diodes, transistors, and integrated circuits. *Varistors* are voltage-dependent resistors. When the line voltage is normal, a varistor has a very high resistance and draws very little current from the line. When a tran-

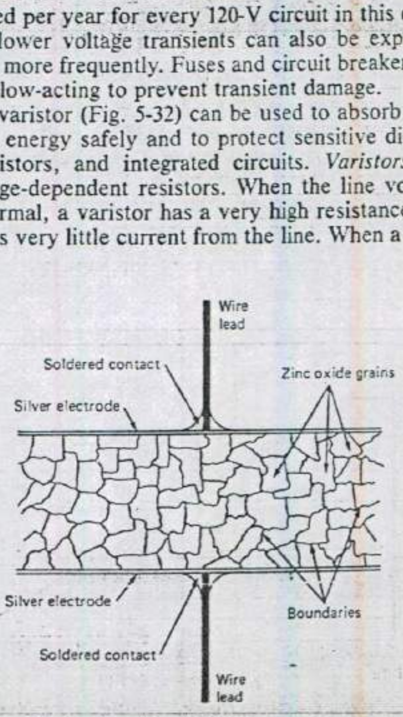


Fig. 5-32 Metal oxide varistor (MOV) structure.

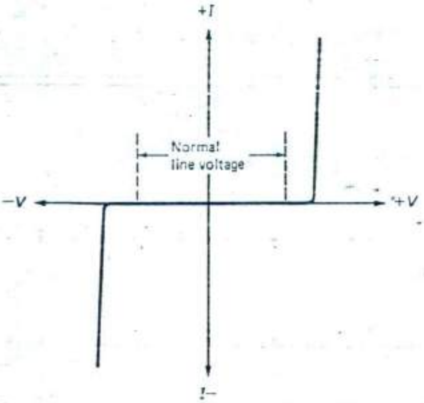
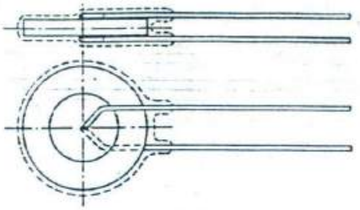
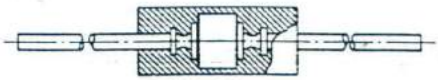


Fig. 5-33 MOV volt-ampere characteristic curve.



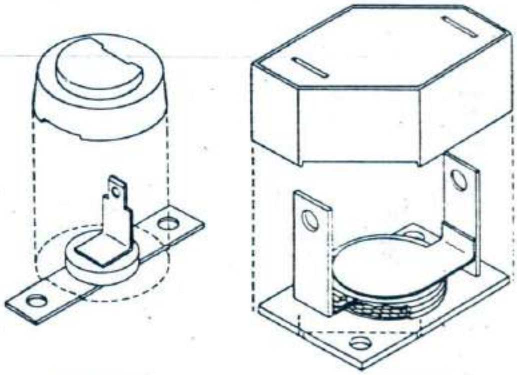
Radial lead package

(a)



Molded axial package

(b)



Power package

(c)

High energy package

(d)

Fig. 5-34 MOV package styles *(Courtesy General Electric Co.)*.

sient comes along, the resistance of the varistor drops sharply. This drop will cause high current in the varistor, and the transient will be safely absorbed. The wafer of zinc oxide in Fig. 5-32 contains boundaries between the grains. These boundaries act as semiconductor junctions, and each requires about 3 V to become forward-biased. The boundaries act in series, and the breakdown characteristics may be controlled by the thickness of the wafer. The metallic oxide varistor (MOV) structure produces a rather high capacitance. This characteristic is usually of no consequence in power line applications.

Figure 5-33 shows the characteristic volt-ampere curve for an MOV. There is no current flow over the normal ac line swing. However, if the line swings positive enough, the positive current in the MOV will increase sharply. The graph shows the same action for positive- and negative-going transients.

Four common packages for MOV devices are shown in Fig. 5-34. The axial devices (Fig. 5-34[*b*]) can absorb 2 joules (J) of transient energy at currents up to 100 A. The high-energy package (Fig. 5-34[*d*]) is rated up to 6500 J and 50,000 A. Fortunately, transients are usually very short in duration.

EXAMPLE

Suppose a transient reaches 5000 V, lasts 10 μs, and causes a current flow of 100 A in an MOV device. What is the absorbed transient energy?

SOLUTION

$$\begin{aligned}\text{Energy} &= V \times I \times t \\ &= 5000 \times 100 \times 10 \times 10^{-6} \\ &= 5\ \text{J}\end{aligned}$$

Five joules is too much for the axial package but could be easily handled by the larger packages shown in Fig. 5-34.

Figure 5-35 shows a typical application of an MOV in a power supply circuit. The varistor is connected in parallel with the transformer primary. Normally, it shows a very high resistance and draws almost no current from the ac line. In the event of a transient, it drops in resistance and absorbs much of the transient energy. It protects the transformer, rectifier, filter, regulator, and load. Depending on the amplitude and duration of the transient, the fuse in Fig. 5-35 may not blow. The MOVs act in nanoseconds and are therefore 100,000 times faster than fuses.

Another type of overvoltage situation can develop in power supply circuits. Many circuits use series pass transistors. These are hardworking devices and are therefore failure prone. Unfortunately, the most common failure mode in a series-pass transistor is an emitter-to-collector short. This fault places the entire unregulated supply voltage across the load. Since the load may contain many sensitive devices such as integrated circuits, extensive damage can result in a piece of equipment if a series-pass transistor short-circuits.

Figure 5-36 shows a crowbar circuit added to a high-current power supply to prevent circuit damage in the event of a regulator failure. The crowbar circuit is made up of D_1, R_9, C_5, and the SCR. With normal load voltage, the zener (D_1) is off. If the load

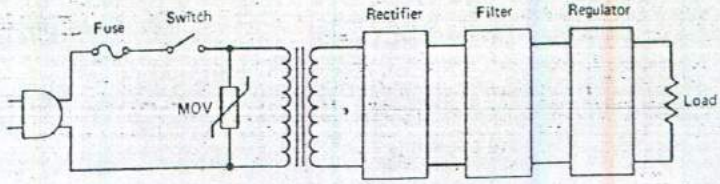


Fig. 5-35 MOV protected power supply.

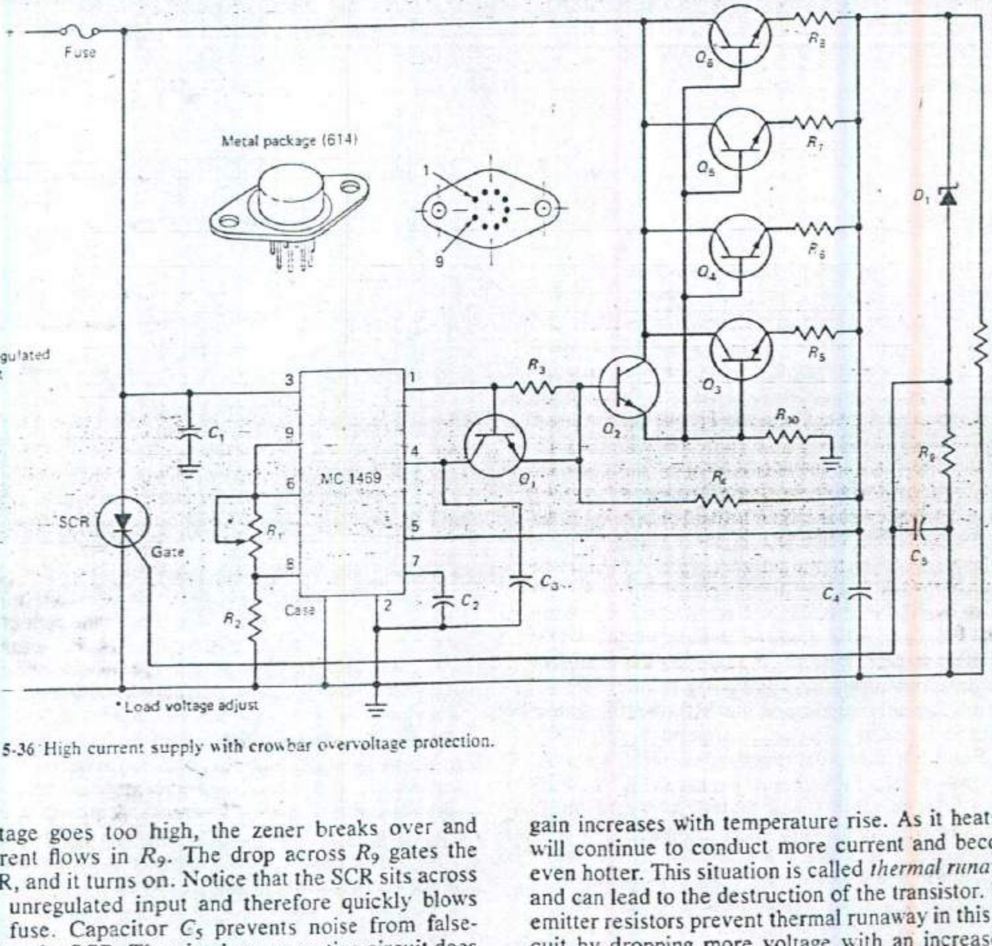


Fig. 5-36 High current supply with crowbar overvoltage protection.

voltage goes too high, the zener breaks over and current flows in R_9. The drop across R_9 gates the SCR, and it turns on. Notice that the SCR sits across the unregulated input and therefore quickly blows the fuse. Capacitor C_5 prevents noise from false-gating the SCR. The simple zener gating circuit does not provide a precise crowbar action. Some circuits are more elaborate. Overvoltage-sensing integrated circuits that provide accurate trip points and programmable delay characteristics are available.

The rest of Fig. 5-36 works as the series-pass circuits already discussed. However, there are a few details worth mentioning. There are four pass transistors operating in parallel. Transistors Q_3 to Q_6 share the load current and give this regulator high current capacity. Resistors R_5 to R_8 help balance the transistor currents. If one transistor has a higher current gain than the others, it will conduct more than its share of the load current. This will make it run hotter than the other three transistors. Transistor gain increases with temperature rise. As it heats, it will continue to conduct more current and become even hotter. This situation is called *thermal runaway* and can lead to the destruction of the transistor. The emitter resistors prevent thermal runaway in this circuit by dropping more voltage with an increase in current. This drop subtracts from the base-emitter bias and decreases the transistor current. Typically, these resistors (R_5 to R_8) are 0.1 Ω. Transistor Q_2 in Fig. 5-36 is a driver for the pass transistors. The MC1469 integrated regulator cannot supply enough drive current for four pass transistors, and Q_2 provides the needed current gain. Transistor Q_1 senses the drop across R_3 and, along with the internal circuits of the MC1469, provides conventional current limiting. Notice that the MC1469 is housed in a 614 metal package with 9 pins.

Thermal protection involves sensing the temperature of some component (usually the pass transistor) and shutting the supply down at some critical temperature. Many of the integrated circuit regulators, such as the 78XX series, employ this technique. A thermal shutdown transistor is located close to the series-pass transistor in the circuit. The shutdown transistor is biased to 0.4 V and remains off at reasonable temperatures; if the shutdown transistor gets too hot (150°C or so), 0.4 V is enough to turn it on. When it comes on, it removes the base drive from the pass transistor, and the output is turned off. The shutdown temperature is several degrees above the temperature at which the regulator will turn back on. The difference between these two temperatures, known as *hysteresis*, prevents the regulator from rapidly *oscillating* (switching back and forth) between on and off conditions.

REVIEW QUESTIONS

33. Is the relationship between voltage and resistance in a metal oxide varistor (MOV) direct or inverse?

34. Is the relationship between the duration of a transient and its energy content direct or inverse?

35. Calculate the energy in joules contained in a 2000-V, 350-A transient that lasts 20 μs.

5-7
SWITCH-MODE SUPPLIES

Almost all of the voltage regulators studied up to this point have used a series-pass transistor. These regulators can provide excellent performance but are relatively inefficient. The pass transistors consume a significant part of the total energy as they conduct the full load current and drop part of the unregulated input voltage. They run hot and require substantial heat sinking in many cases. A design that eliminates series-pass transistors is illustrated in Fig. 5-37. It is not a switch-mode supply but is included in this section because it introduces an important concept.

The bridge section of Fig. 5-37 uses two rectifiers and two silicon-controlled rectifiers. The SCR control circuit gates the bridge into conduction early or later in the half cycle, depending upon demand and line voltage. The waveforms show how the average dc output of the bridge is related to SCR gating time. Early gating produces a high dc average, and late gating produces a low dc average. An error amplifier compares a reference voltage with a sample of the load voltage. Any difference will change the gating of the SCRs and reduce the error.

The SCR-regulated supply is more efficient than any of the series-pass circuits. When an SCR is off, it blocks current. Zero current ensures zero power dissipation in the SCR. When an SCR is on, it drops very little voltage, thus ensuring little power dissipation in the SCR. The point is that the SCR does not dissipate very much power because it is a switching device. It is either off or on. On the other hand, a pass transistor is somewhere between off and on. It operates in the linear region; it acts as a resistor and therefore consumes energy. In switch-mode power supplies the control transistors are not operated in the linear region. This gives this type of supply an efficiency advantage similar to that achieved by the SCR supply and also allows it to be operated at frequencies far beyond the line frequency. The higher frequencies allow a significant reduction in the size and weight of transformers, inductors, and filter capacitors. Switching supplies are about one-third the size and weight of equivalent linear types.

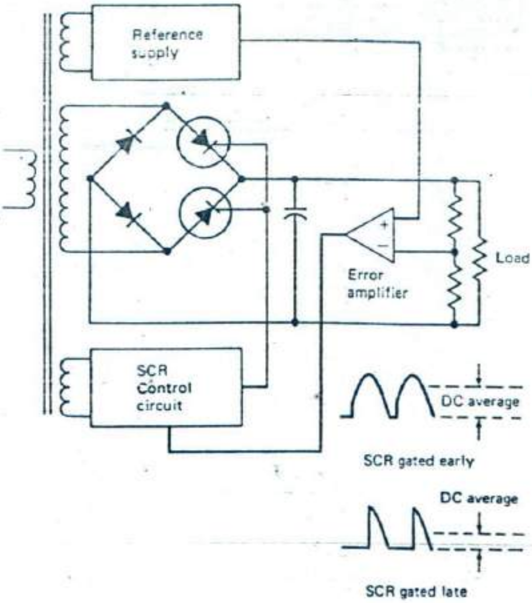


Fig. 5-37 SCR regulated power supply.

Switch-mode operation can achieve regulation by using pulse-width modulation. Figure 5-38(*a*) represents a high-duty cycle since the width of the positive-going pulse is a large percentage of the cycle. In this case it is about 75 percent. A square wave has a duty cycle of 50 percent since the positive pulse is equal to one-half cycle. Assuming an operating fre-

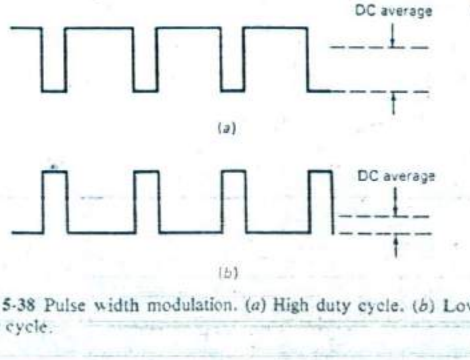


Fig. 5-38 Pulse width modulation. (*a*) High duty cycle. (*b*) Low duty cycle.

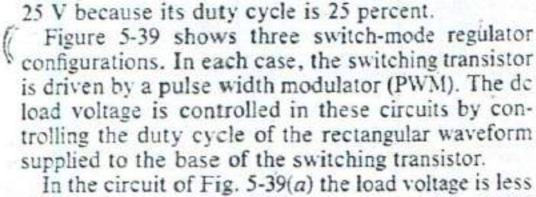
quency of 40 kHz, the calculation would appear as follows:

$$\begin{aligned}\text{Duty cycle, \%} &= \frac{t_{\text{high}}}{T} \times 100\\ &= \frac{18.75 \times 10^{-6}}{25 \times 10^{-6}} \times 100\\ &= 75\%\end{aligned}$$

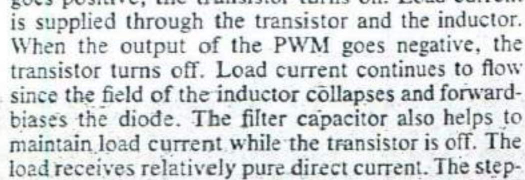
With rectangular waveforms, the dc average is equal to the dc peak times the duty cycle. Assuming peak values of 100 V, the waveform in Fig. 5-38(*a*) would average to 75 V and the waveform in Fig. 5-38(*b*) to 25 V because its duty cycle is 25 percent.

Figure 5-39 shows three switch-mode regulator configurations. In each case, the switching transistor is driven by a pulse width modulator (PWM). The dc load voltage is controlled in these circuits by controlling the duty cycle of the rectangular waveform supplied to the base of the switching transistor.

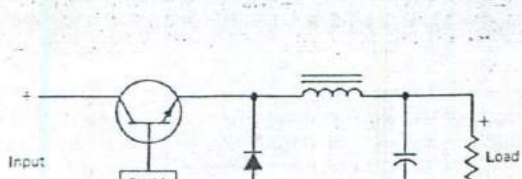
In the circuit of Fig. 5-39(*a*) the load voltage is less than the input voltage. When the output of the PWM goes positive, the transistor turns on. Load current is supplied through the transistor and the inductor. When the output of the PWM goes negative, the transistor turns off. Load current continues to flow since the field of the inductor collapses and forward-biases the diode. The filter capacitor also helps to maintain load current while the transistor is off. The load receives relatively pure direct current. The step-down configuration is much more efficient than its series-pass equivalent, especially in those cases in which the load voltage is considerably less than the input voltage.



(*a*) Step-down configuration



(*b*) Step-up configuration



(*c*) Inverting configuration

Fig. 5-39 Switch-mode regulator configurations.

Not only is the switching mode much more efficient, it also allows regulator configurations not possible with the series-pass arrangement. In Fig. 5-39(*b*) the load voltage is greater than the input voltage. When the PWM goes positive it turns the transistor on and current flows through the transistor and the inductor. When it goes negative, the transistor turns off, and the field in the inductor collapses and generates a voltage that adds in series with the input voltage. The diode is now forward-biased, load current flows, and the filter capacitor is charged. When the transistor turns on again, the diode prevents it from discharging the filter capacitor. The load voltage can be opposite in polarity to the input voltage as shown in Fig. 5-39(*c*). These circuits are handy in cases such as a positive ground system in which one or two negative voltages are required. When the transistor turns off, the inductor discharges through the diode and the load. Since the direction of discharge current must be the same as the direction of charge current, demonstrating that the load polarity is opposite to the source polarity is easy.

Regulation in a switch-mode supply is a matter of comparing the load voltage with a reference voltage and using any error to correct the duty cycle of the output of the PWM. Figure 5-40 shows a step-down regulator based on an integrated circuit specially designed for switch-mode power supplies. The PWM is contained in the chip and consists of an oscillator, AND gate, and a latch. The oscillator produces a rectangular waveform, and its frequency of operation is controlled by external capacitor C_1. The AND *gate* is a circuit that turns on and passes a signal to the latch only when both of its inputs are positive. The *latch* is a storage circuit that remembers a high signal that was applied to its S input and supplies that high signal to its Q output until it is reset at its R input. Transistors Q_1 and Q_2 form a Darlington switch that will be turned on when the Q output of the latch is high. Voltage regulation is achieved in Fig. 5-40 by comparing an internally developed reference voltage with a sample of the load voltage. Any error is compensated for by pulse-width modulation. For example, if the load demands more current, the voltage tends to drop. Dividers R_3 and R_2 will then send a negative-going signal to the error amplifier. It is inverted, and the output of the amplifier goes positive (high), enabling the AND gate. The positive-going oscillator will then set the latch and turn on the Darlington pair. The latch is reset at its R input by the oscillator signal. The duty cycle is a function of the error amplifier output. For example, if the amplifier never goes high, the duty cycle will be 0 percent since the AND gate will never be enabled. On the other hand, if the amplifier output is continuously high, the latch will be immediately set after it is reset and the duty cycle will approach 100 percent. Normal operation falls in between these two extremes.

Figure 5-40 also provides for current limiting. Resistor R_1 senses the current to the Darlington pair.

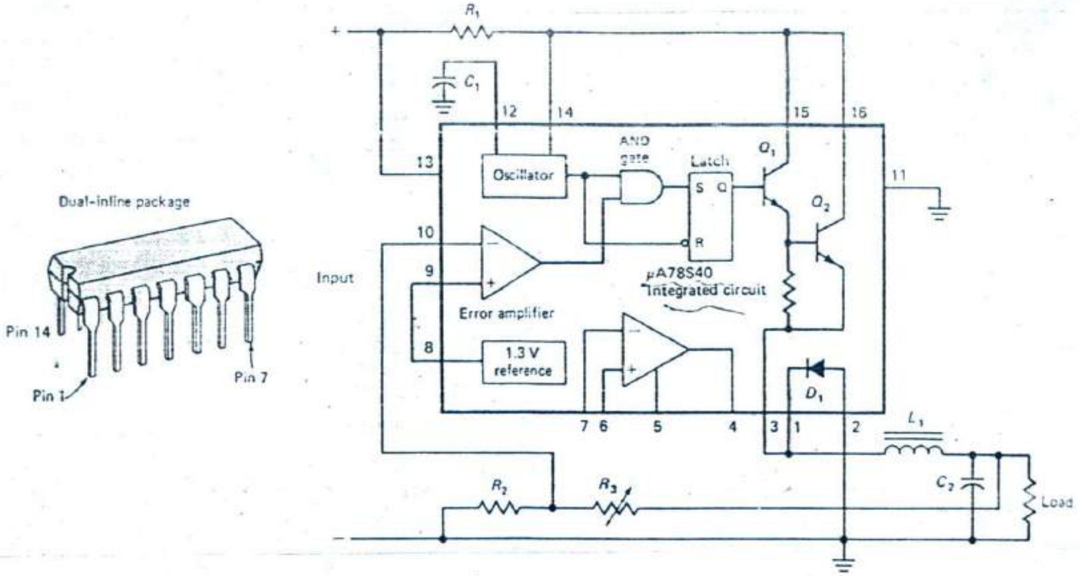


Fig. 5-40 Integrated circuit step-down regulator.

When the drop across R_1 reaches about 0.3 V the oscillator waveform is reduced in duty cycle. Now it is not possible for the Darlington switch to be driven with a duty cycle as high as before, and the output is limited. As shown, the circuit can safely deliver about 1 A of load current. It is possible to use pin 3 of the integrated circuit to drive an external switching transistor for more load current. It is also possible to use an external diode and therefore not use pins 1 and 2 on the device. Pins 4 through 7 access an additional amplifier that can be used for temperature control or some other power supply feature. The integrated circuit can be used in any of the three configurations shown in Fig. 5-39.

Converters are another category of switch-mode supply. A *converter* is a circuit that changes direct current to alternating current and then back to direct current again. This makes it possible to transform one dc voltage to another dc voltage. It also permits transformer isolation at a considerable savings in size and weight because the frequency of operation is much higher than the normal line frequency. Figure 5-41 shows a width-controlled converter with voltage regulation. Transistors Q_1 and Q_2 are driven with out-of-phase rectangular waves. Note that they will alternately conduct and allow primary current in T_1. The load voltage is sampled by divider R_1-R_2 and compared with a reference. Any error is used to modulate the width of the pulses supplied to the base circuits. Increased load demand will be compensated for by increasing the pulse width and the ON time for both transistors. This will increase the energy delivered to the transformer and compensate for the increased load demand.

Diodes D_1 and D_2 in Fig. 5-41 rectify the high-frequency alternating current. Inductor L_1 and capacitor C_1 form a choke-input filter. There are periods of time when both transistors are off, and L_1 will then discharge to maintain the load current. Diode D_3 is forward-biased by L_1 at those times and completes the discharge circuit. The circuit will function without D_3, but then the discharge current is forced to flow through the rectifiers and the secondary of the transformer. This condition is not desirable because it increases the dissipation and lowers circuit efficiency.

Switch-mode regulators and converters are quite a bit more efficient than series-pass arrangements. Better efficiency translates to reduced heat sink requirements, and a smaller supply results. The high frequency of operation reduces the size of filter components and the size of the cores in the inductive components. All in all, it seems that they have every advantage. There is one disadvantage, however. They are noisy. They operate with rectangular waveforms. *Rectangular waveforms* are composed of a fundamental frequency plus a series of harmonically related frequencies. The third harmonic is three times the fundamental; the fifth is five times the fundamental; and so on. Harmonics add to the noise in the output of the supply. They also create another problem, called *electromagnetic interference* (EMI). The energy content of the harmonics falls off at the higher frequencies, and the majority of the noise ranges from 10 to 500 kHz. There is often enough high-frequency energy to cause radiated interference with nearby equipment. Shielding and output filters are required to control radiated EMI.

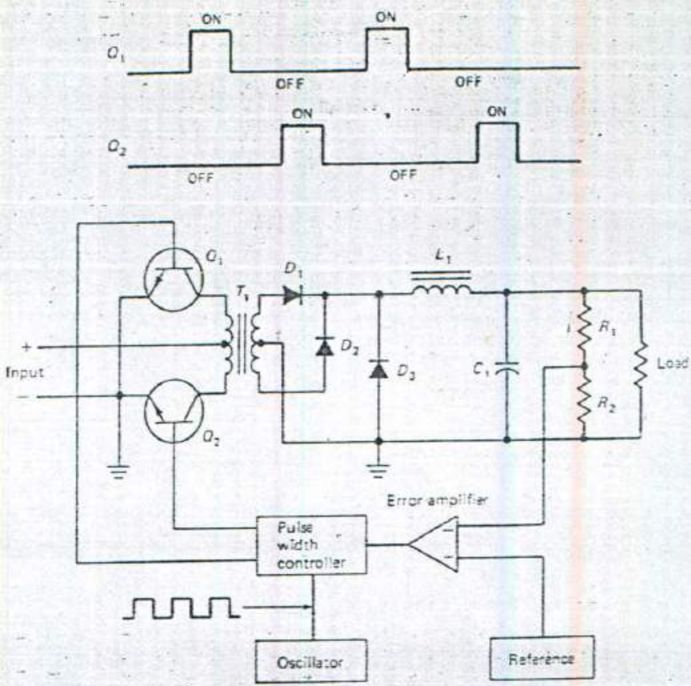


Fig. 5-41 Width-controlled converter/regulator.

Switchers also may create problems with conducted EMI. The location of switching transistors directly off the power line and their high peak currents can create significant line noise. Switchers can place noise on both sides of the line referenced to ground (*common mode noise*) and can also place noise on one side of the line referenced to the other (*differential mode noise*). Line filters are required to attenuate these forms of line noise.

Figure 5-42 shows a frequency-controlled sine wave converter. Sine waves solve some of the noise and EMI problems because they have no harmonic content. The circuit uses power FETs to switch the direct current produced by D_1-D_4. The resulting square wave is converted to a sine wave by a resonant tank circuit formed by L_1 and C_3. These two components are tightly coupled by transformer T_1 and behave as though they were in parallel. The high-frequency sine waves are changed to direct current by Schottky rectifiers D_5 and D_6 and L_2 and C_4 form a choke-input filter.

The circuit of Fig. 5-42 uses frequency modulation rather than pulse-width modulation. The duty cycle of the drive signals supplied to the FETs is fixed at 50 percent. Any change in load voltage will shift the frequency of the drive signal. This is accomplished in the voltage-controlled oscillator (VCO). The tank circuit has some natural resonant frequency established by the inductance of L_1 and the capacitance of C_3. Suppose the VCO is generating a signal above the resonant frequency. A tank circuit shows maximum voltage when driven at its resonant frequency. The circuit Q is such that an octave-frequency increase translates to a 12-decibel (dB) drop in tank voltage. An octave increase means double and a 12-dB drop means one-fourth. Therefore, if the tank voltage were 20 V at a VCO frequency of 150 kHz it would drop to 5 V at a frequency of 300 kHz. Any error in load voltage is corrected by shifting the VCO in the proper direction. Figure 5-42 operates above resonance, and an increased load demand will lower the VCO frequency closer to resonance which will increase the tank voltage. Decreased load demand will raise the VCO frequency further away from resonance and decrease the tank voltage. The slope of the control curve is 12 dB per octave.

The power FETs and Schottky rectifiers in Fig. 5-42 permit good circuit efficiency into the hundreds of kilohertz. This is not possible with bipolar transistors and ordinary rectifiers because of their poor switching performance at high frequencies. They cannot turn off fast enough, because of carrier storage. It takes time to sweep all carriers from their junctions so conduction does not cease immediately when forward bias is removed. Field effect transistors are unipolar and do not exhibit the carrier storage associated with bipolar devices. The Schottky diodes use metal on one side of the junction and doped silicon on the other. This construction technique eliminates the depletion region and the storage problem. Power FETs are also more rugged, in that they do not suffer from *secondary breakdown*, which is a phenomenon suffered in bipolar devices when the crystal develops hot spots. The hot spots are

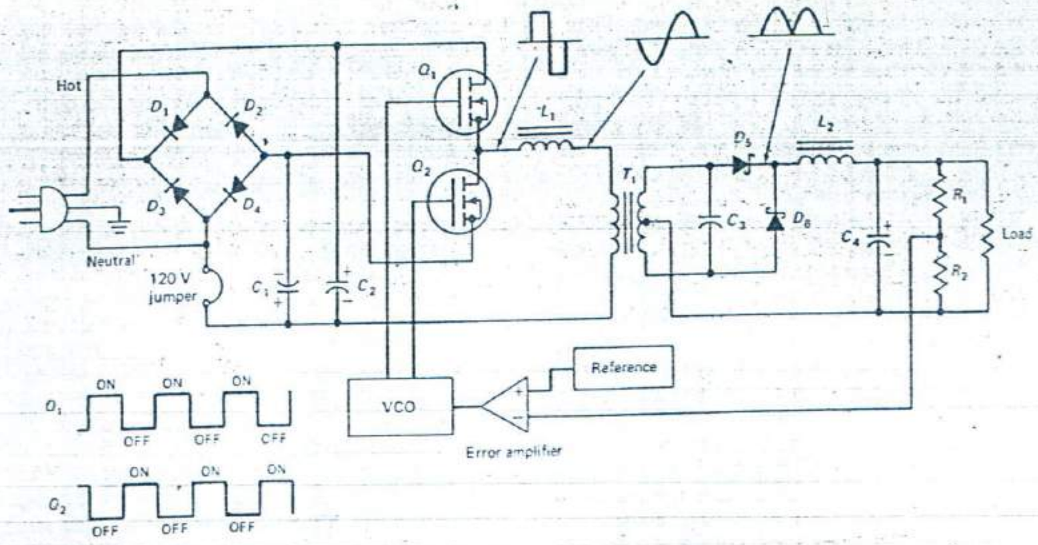


Fig. 5-42 Frequency-controlled sine wave converter.

caused by current crowding associated with the intense fields generated in power devices and can lead to breakdown even though the transistor is operating in its safe area. Primary breakdown is caused by operation outside the safe area.

Diodes D_1 to D_4 and capacitors C_1 and C_2 in Fig. 5-42 make up a bridge-doubler circuit. The circuit acts as a doubler with the 120-V jumper installed; the dc voltage supplied to the switching transistors would be around 240 V with a 120-V ac input. With the jumper removed, the circuit acts as a bridge rectifier and supplies about 240 V to the transistors with a 240-V ac input. Thus, the circuit can be configured for either of two line voltages with the jumper. Line isolation is achieved in T_1. Since the frequency of operation is so high, this transformer is tiny compared to an equivalent line frequency device.

REVIEW QUESTIONS

36. Calculate the efficiency of a series-pass regulator that supplies 5 A at 12 V from an unregulated input of 18 V. (Hint: *efficiency* is found by dividing useful output by total input.)

37. If a switch-mode supply operates at 30 kHz, calculate the duty cycle if the ON time is equal to 12 μs.

38. Assume that you are observing the drive signal to the base of a switching transistor in a power supply. What can you expect to see happen if the load current is suddenly decreased?

39. Calculate the average voltage of a rectangular waveform with a peak value of 25 V and a duty cycle of 85 percent.

40. Can the integrated circuit shown in Fig. 5-40 be used to develop a negative voltage in a positive ground system?

41. What will happen to the load voltage in Fig. 5-40 if R_3 is adjusted for less resistance?

42. Which component in Fig. 5-41 is included to reduce the dissipation in the transformer and the rectifiers?

43. What do schematic symbols D_5 and D_6 in Fig. 5-42 depict?

5-8
TROUBLESHOOTING AND MAINTENANCE

Power sources are considered the heart of electronic equipment. They must function properly for the equipment to work as designed. Technicians must be familiar with power supply operation and know how to verify correct performance. They must be able to diagnose malfunctions and replace components to restore proper and safe operation. Because the power supply can affect all parts of the system, verification of supply voltages, currents, and waveforms must precede any troubleshooting efforts in other parts of the system. More than one technician has spent valuable time troubleshooting a normal circuit that is acting abnormally because of a faulty supply. It is usually easy to verify proper operation in a power supply, and this must be done early in the troubleshooting process.

Safety is the main consideration when troubleshooting. Severe electrical shock, burns, fires, equipment damage, and losses in production time are some of the penalties for improper work procedures.

Grounded test equipment may lead to ground loops. Some parts of the ac line and some parts of the power supply circuitry can be hot with respect to an earth ground. The act of plugging in a piece of test equipment normally connects its case and any ground leads to an earth ground. Therefore, if any of the ground leads is brought into contact with a hot circuit, a ground loop results and very high fault currents will flow. Or the technician's body can become part of the loop if the case of the test equipment and a hot circuit are touched at the same time.

It may be necessary to use battery-operated test equipment when working on some industrial circuits. This permits the instrument case and test leads to remain "floating" with respect to ground. Floating measurements require special safety procedures. For example, the case of a piece of test equipment that is being used in a floating measurement may be hundreds of volts with respect to ground. Touching the case may cause a severe shock if the technician's body has any conductivity to ground. Special test equipment, insulating mats, clothing, and other protective gear are musts when making floating measurements.

Isolation transformers can be used in some cases to prevent ground loops. The power supply can be isolated from the ac line by energizing it from the transformer secondary. This eliminates ground loops but does not eliminate other shock hazards. No matter what the working conditions are, safe practices must be followed. Safe workers have an orderly way of working, regardless of the conditions at the time. They regard all circuits as potentially dangerous and are not lulled into sloppy work habits by terms such as *low voltage*. For example, Fig. 5-42 shows a power supply design that could be used in an industrial control computer to supply 5 V to the logic circuits. What could be safer than 5 V? Well, the power supply does develop 5 V for the load but works initially at 240 V dc in the switching section! Direct current of 240 V dc is potentially very dangerous and must be treated with respect. Figure 5-42 also points out that line isolation is achieved in only part of the power supply circuit. The bridge-doubler runs directly off the ac line, and a ground loop can result during work on this part of the circuit.

The industrial technician must also be aware that the metal case of certain power supply components may be at a high potential with respect to the common ground. For example, refer to the voltage-doubler circuits shown in Fig. 5-18. Depending on capacitor construction, C_1 could be a shock hazard in these circuits. In the half-wave doubler, the case of C_1 would be at the full secondary voltage referenced to ground. In the full-wave doubler, the case of C_1 would be at half the load voltage referenced to ground.

Power supplies are potential shock hazards even when turned off. Today, the large capacitive filter is the most common way to achieve pure direct current. Capacitors can store quite a bit of energy and may store it for long periods of time. Safety demands that the capacitors be discharged before touching parts of the circuit or making some measurements. Short-circuiting capacitors by a hand tool or test lead is not advisable in high-energy supplies. The energy delivered by some capacitors when short-circuited is enough to vaporize an alligator clip! Special discharge rods with internal current-limiting resistors are used to bleed off capacitors in high-energy equipment.

Troubleshooting is a logical procedure. It begins by carefully observing the symptoms and all operating conditions. The logical technician verifies all control settings, external connections and cabling, and power to the unit before tearing it down. Once the technician is sure that there are no obvious external reasons for malfunction, then it is time to power down. In the industrial environment, it is often necessary to lock circuits off and tag them. The tag warns that maintenance is underway. Tear-down procedures are often presented in the equipment manufacturer's literature. So are important safety precautions. Find all relevant literature and use it!

A visual inspection of the inside of the equipment follows tear-down. Look for obvious problems such as foreign objects, leaking batteries and capacitors, broken wires, cracked circuit boards, burned components, components not seated properly in their sockets, circuit boards not seated properly in their connectors, loose connectors, and dirt. The industrial environment is often very dirty, and you may find that the equipment is loaded with grime and dust. This kind of build-up is often conductive and must be cleaned. Be sure to wear a respirator when cleaning up since the dust may be hazardous. Do not forget to check ventilation systems and to clean any air filters. Any liquid material inside the equipment may present a more difficult problem. Try to find out where it came from and what it is.

It may be necessary to remove circuit boards and wash them, in some cases. Check the manufacturer's recommendations before proceeding. Some may recommend a solution of 90 percent ethyl alcohol and 10 percent water. If corrosive materials such as chlorides are on the board, a solution of water and bicarbonate of soda may be used to clean and neutralize the board. The boards must be thoroughly rinsed with deionized water and dried before installation.

Battery and cell maintenance is usually straightforward. A voltage check will verify whether the unit must be replaced. However, do not make the mistake of measuring open-circuit voltage. A weak battery will often show a normal voltage until loaded. Energize the load circuit or select a resistor of the proper resistance and power ratings and make a test under normal load conditions. Some equipment may have a battery test switch for this purpose. If a secondary battery is below the cut-off voltage, verify proper operation of the charging circuit. Keep a maintenance log and always date the replacement of any cell or battery. Many technicians also tag the equipment to indicate the date that the battery was replaced. These habits save time and money. Finally, leaking cells and batteries must be replaced because electrolyte material is highly corrosive and conductive.

The most common failure mode for solid-state rectifiers is a short circuit. Make sure the power is off and that the filters are discharged and then run an ohmmeter check when you suspect a bad rectifier. Use a low range and check for different readings as the ohmmeter polarity is reversed. Do not expect to see infinite reverse resistance when running in-circuit checks. For example, refer to Fig. 5-10. Suppose the ohmmeter positive lead is on the cathode of D_2 and the negative lead is on the anode. This reverse-biases D_2 and D_1, but ohmmeter current will flow through the bottom half of the secondary and the load. If the diode is good, the reading will be a function of the load and secondary resistance. The forward resistance of a solid-state diode is also important. Most ohmmeters do not turn the diode on very hard, and a good diode will show a resistance considerably greater than 0 Ω. A short-circuited diode will show 0 Ω in both directions. Removal of at least one diode lead from the circuit will allow conclusive tests. Silicon rectifiers normally have a reverse resistance higher than the top range of the ohmmeter. Replace any leaky units. Don't forget that at least 0.6 V is necessary to turn on a silicon diode. Some ohmmeters have a special low-voltage ohms function, which cannot be used when testing diodes. Finally, if a rectifier is rated at more than 1000 V it may be a series combination of diodes and cannot be tested with an ordinary ohmmeter. It will test open (infinite resistance) in both directions.

Filter capacitors, especially the electrolytic type, are failure-prone. They may develop excessive leakage and may even short-circuit. They can also dry out and lose much of their capacity. Finally, they can develop a high series resistance which limits their ability to deliver load current. In-circuit testing can be used to find a short-circuited capacitor. Observe polarity and, as in the case of rectifier testing, be aware of other paths for the ohmmeter current. Remove at least one capacitor lead for more conclusive testing. A momentary low resistance followed by increasing resistance is to be expected when ohmmeter-testing large capacitors. Large electrolytics always show some leakage, and the ohmmeter will not reach an infinite reading on its highest range. Testing for excessive leakage is best done at the rated voltage. Also, testing for capacity and series resistance demands a capacitor tester. It may be most effective to try a new capacitor when symptoms such as excessive ripple point to the filter. Electrolytics have a shelf life and the technician should be aware that a "new" capacitor can be defective, especially if it has spent 10 years in storage.

As mentioned before, troubleshooting is a logical process. Good troubleshooters use analysis to limit the possibilities. They can take a set of symptoms and zero in on a set of possible causes. They understand circuit laws and know how circuits operate normally. This knowledge leads them to the answers they are looking for. For example, look at Fig. 5-20. Suppose that the load voltage is zero. First verify that there is some input to the circuit. It is not productive to troubleshoot a regulator circuit until it is verified that the input is normal or at least low. Zero input usually points to a defect in a circuit before the regulator. Low input may point to an overload condition. Suppose the unregulated input in Fig. 5-20 is low or normal. What kinds of faults could cause the load voltage to be zero? Resistor *R* may be open, or the zener, capacitor, or load may be short-circuited. If resistor *R* is open, it will be cold. If there is a short circuit, it will probably be hot. In fact, it may smell and look burned. Use all of your senses but be careful what you touch. You could be burned or shocked.

Suppose the load voltage in Fig. 5-20 is too high. This changes the analysis. Now, the only probable fault is an open zener diode. It is unlikely that *R* is short-circuited because resistors seldom short-circuit; if it did short-circuit, it would probably destroy the diode.

We are dealing with voltage analysis. We are measuring circuit voltages and analyzing possible causes for improper readings. Learn how to take measurements safely when the equipment is on because much troubleshooting must be done this way. Watch where you put your arm, wrist, hand, and fingers. Use insulated probes and never have more than one hand in the circuit at one time. A forearm-to-finger shock can be bad enough, but a hand-to-hand shock can be lethal.

After the voltage analysis, it may be time to go back to resistance analysis. If the load voltage is zero in Fig. 5-20, we know that any of three short circuits and one open circuit are among the possibilities. Power down, discharge the capacitors, and use the ohmmeter to find the problem. Start with an in-circuit check of *R*. Try both polarities even though it is not a diode. Sometimes you can avoid forward-biasing junctions in other parts of the circuit with this technique. The highest reading that you can obtain is closest to the correct reading. If *R* measures too high, you have found the trouble or part of it. There still could be a short circuit that caused excess current in *R* and burned it out. Use the ohmmeter to check; if a short circuit is found, leads will have to be disconnected since the diode, capacitor, and load are connected in parallel.

Look at the series-pass circuit in Fig. 5-21(*a*). Assume a normal unregulated input and zero load voltage. What kinds of component failures are possible? The transistor could be open, resistor *R* could be open, and the diode or the capacitor could be short-circuited. A short-circuited load is another possibility, but it would probably cause the unregulated input voltage to be lower than normal, and the pass transistor would be hot because of the high current flow. Suppose the load voltage in Fig. 5-21(a) were too high. This might be caused by an open diode or a short-circuited pass transistor. A quick voltage reading from ground to the base of the transistor will eliminate one or the other. A high voltage here points to an open zener, and a normal reading indicates that the transistor has short-circuited.

Low output voltage often points to an overload. If a regulator uses current limiting, the output voltage will drop below normal with excessive load current. Even without current limiting, overloads always can be expected to cause voltages to drop below normal. Overloads also make circuits run hot. However, do not jump to conclusions, because some components are quite safe at temperatures that will burn your finger. Sometimes it is necessary to use current analysis. Most technicians avoid this since the circuit has to be broken to insert the ammeter. If a series resistor of known value is available, measure the drop across it and calculate the current with Ohm's law. Don't forget the possibility of latch-up (discussed in the section on current limiting). In some cases, it may be necessary to remove some or all of the load from the supply to determine whether normal operation can be restored.

Another cause of voltage error is a fault in the reference supply. This is easy to verify with a voltage check. Don't forget to check the divider that samples the output voltage. Any problem here will send the wrong voltage to the error amplifier and cause an output error. A current-limit circuit can also cause voltage error if it is defective. Suppose R_3 in Fig. 5-27 increases in value. It will cause the circuit to go into current limiting at less than maximum load current, and the load voltage will be below normal.

Repeated pass transistor failures may indicate a defect in the current-limit circuit. Suppose Q_1 in Fig. 5-31 is replaced and normal operation is restored. It will be wise to check the supply for foldback; otherwise the replacement transistor may fail in a short period of time. If Q_3 is open, the supply will not limit, and the pass transistor will not be protected from overloads. Sometimes the difference between a good technician and a poor technician is that the good technician repairs the equipment once, and the poor technician repairs it once a week.

Blown fuses with glass tubes should be visually inspected. It is easy to tell the difference between a moderate overload and a severe overload. A severe overload often covers the inside of the glass with spattered metal and black smoke. These kinds of clues are valuable. For example, suppose the circuit of Fig. 5-35 is dead. Inspection of the fuse reveals a severe overload. If lightning made the lights flicker when the equipment went dead, it is easy to piece together what happened. If there was no transient, then it is likely that there is a dead short circuit in the equipment, and it will be wise to look for it before energizing the equipment again. A severely blown fuse in Fig. 5-36 indicates that the crowbar circuit was tripped. The pass transistors and Q_2 should be checked for emitter-to-collector short circuits before turning the supply on. If the transistors are good, the integrated circuit may be defective, divider R_1-R_2 may be wrong, D_1 may be short-circuited, the SCR may be short-circuited, or noise may have caused false gating of the SCR.

Switch-mode supplies can exhibit some additional symptoms when compared with linear supplies. They can make sounds such as clicking, chirping, and squealing. They are designed to operate above the limit of human hearing, but defects and overloads may cause the frequency to drop. Overloads can also cause a switcher to shut down and then restart repeatedly. This can create audible clicking and chirping. Always investigate the possibility of an overload when unusual sounds are heard.

An oscilloscope is the preferred instrument for analyzing switch-mode supplies. Rectangular waveforms can cause misleading readings in a voltmeter unless the meter is capable of true rms performance at the frequency of operation. Most meters are not true rms indicators, and some true rms meters are not accurate at switch-mode frequencies. It is also good practice to check the output of linear supplies for ripple and noise with an oscilloscope. The frequency of operation also must be verified in switch-mode supplies. This can be done with sufficient accuracy on a good oscilloscope by measuring the period of the waveform and then deriving the frequency.

The last step is the replacement of defective components. It is mandatory to use exact replacements or substitutions with equal or better specifications. However, this can be a trap for the beginning technician. Replacing a 1-W resistor with a 2-W resistor can cause failure of a more expensive component if the original design intended that the resistor would increase in value under overload conditions. Substituting a larger-capacity filter may cause excessive current peaks in the rectifiers and transformer. Generally, it is safest for the beginner to match replacements to the original parts as closely as possible; this means according to component type, as well. It is poor practice to replace a film resistor with a carbon composition resistor, for example. Such substitutions can even lead to fires. The physical size is also important. Make sure the replacement has the proper lead arrangement and that it will fit in the space available. Make your work look like the factory wiring as much as possible. Replace all components, even seemingly unimportant ones such as ferrite beads. Do not leave long leads on components. For example, unless bypass capacitors such as C_1, C_2, and C_3 in Fig. 5-36 have short leads, the integrated circuit may become unstable.

Finally, it must be emphasized that certain components are special and have rather critical specifications. Rectifiers D_5 and D_6 in Fig. 5-42 must be of the Schottky type. Ordinary rectifiers will not work in this circuit. They will overheat and be destroyed and may cause additional circuit damage. Capacitor C_3 is another example of a special component. It is a special four-lead capacitor designed for low effective series resistance (ESR). An ordinary capacitor has a much higher ESR and would be overheated and destroyed in a circuit of this type. The inductors in switchers and sine wave converters are also special. They are quite often *cup-core* types, where the cores surround the windings or toroid types, where the windings are placed on doughnut-shaped cores.

REVIEW QUESTIONS

44. What must the technician guard against when working on "hot" circuits with grounded test equipment?

45. Which part of a power supply may store a charge for some time?

46. Refer to Fig. 5-11. A technician connects the positive ohmmeter lead to the cathode of D_2 and the negative lead to its anode. Is the technician measuring the forward resistance of D_2? Is the reverse resistance of D_2 being measured?

47. What is being measured in question 46?

48. Refer to Fig. 5-22. Transistor Q_2 is open. What is the symptom?

CHAPTER REVIEW QUESTIONS

5-1. If a 6-V, 2-A · h ni-cad battery is completely discharged, what is the minimum time required to completely restore it at a C/10 charging rate? How long can it be left on charge before it is damaged?

5-2. Why is it poor practice to store discharged gel-cells for an extended period?

5-3. Assume a line frequency of 60 Hz. What is the ripple frequency in a line-to-neutral three-phase wye rectifier?

5-4. What is the rectifier cell ratio for a single-phase, full-wave bridge circuit?

5-5. Refer to Fig. 5-17. Calculate the capacitor value required to keep the ripple voltage at 1 V peak-to-peak. Assume a load current of 10 A and a line frequency of 40 kHz.

5-6. What can you conclude from question 5-5 regarding the size of filter components required in high-frequency switch-mode power supplies?

5-7. Refer to Fig. 5-18(*b*). Assume no diode loss, a light load, and a 120-V rms secondary. What is the dc load voltage?

5-8. Refer to Fig. 5-22. What can be expected to happen to the output voltage if resistor R_2 opens?

5-9. Refer to Fig. 5-24. Assume a pass transistor with a V_{be} of 1.5 V at a collector current of 4 A. Select a value for R that will set the integrated circuit-regulator current at 1 A when the load current is 5 A.

5-10. Use Fig. 5-30(*b*) and predict the load current for a 2-Ω load resistor.

5-11. An MOV is rated at 5 J and 200 A. Assume maximum current flow and calculate the maximum voltage that it can safely withstand for 5 μs.

5-12. Suppose the circuit of Fig. 5-35 is designed to operate on common 120 Vac. Why would it not be possible to use an MOV designed to break over at 150 Vdc?

5-13. Refer to Fig. 5-36. Which component sets the crowbar trip point?

5-14. Refer to Fig. 5-36. Potentiometer R_1 is used to adjust the load voltage. Why must it be adjusted carefully in a circuit of this type?

5-15. Calculate the frequency of the seventh harmonic in a switch-mode supply that operates at 35 kHz.

5-16. What two undesired effects are caused by harmonic energy in switch-mode supplies?

5-17. How much harmonic energy can be found in a sine wave?

5-18. Name two problems or limitations associated with bipolar power transistors that are not associated with power FETs.

5-19. Refer to Fig. 5-22. If R_1 is open, what is the symptom?

5-20. Refer to Fig. 5-27. If R_1 is open, what is the symptom?

5-21. Refer to Fig. 5-36. Transistor Q_2 has a collector-to-emitter short circuit. What is the symptom?

5-22. Refer to Fig. 5-36. Transistor Q_2 is open. What is the symptom?

5-23. Refer to Fig. 5-40. Resistor R_1 has increased in value. What is the symptom?

5-24. Refer to Fig. 5-22. What can be expected to happen to the load voltage if the series-pass transistor develops a collector-to-emitter short?

ANSWERS TO REVIEW QUESTIONS

1. primary **2.** polarization **3.** 35 h **4.** lithium thionyl chloride **5.** 3.3 MΩ **6.** 30 Ω **7.** 22.5 V **8.** 27 V **9.** positive **10.** 13.5 V **11.** 6.45 V; 47.8 percent **12.** 7 V **13.** pulse waveforms show a greater heating effect **14.** 20 percent **15.** capacitor input **16.** capacitor input **17.** capacitor input **18.** 42.3 V **19.** it will decrease **20.** it will increase **21.** 20 Ω; 0.45 W **22.** 0.25 W; 0.75 W **23.** 150 mA; drops to zero; drops below 5 V **24.** 7 V **25.** 2.29 kΩ **26.** 5 V **27.** 4 A **28.** from 3 to infinity Ω **29.** from 0 to 3 Ω **30.** 10.1 A **31.** 150°C **32.** 5.3 A **33.** inverse **34.** direct **35.** 14 **36.** 0.667 or 66.7 percent **37.** 36 percent **38.** the duty cycle will decrease **39.** 21.3 V **40.** yes **41.** it will go down **42.** D_3 **43.** Schottky rectifiers **44.** ground loops and shock **45.** filter capacitor **46.** no; no **47.** the load plus the forward resistance of D_3 (depending on primary resistance, D_1 may also be turned on) **48.** high output and no voltage regulation