2.1 The hemodialysis (HD) machine[1]

The hemodialysis (HD) machine pumps the dialysate as well as the patient's blood through a dialyzer. The blood and dialysate are separated from each other by a semipermeable membrane permitting solute and water transfer as governed by laws of physics. In practice, however, this procedure is somewhat more complex. The operational system of the HD machine represents a complex array of detectors, controllers, monitors, and safety devices to ensure a safe operation. This integrated system allows the operator the ability to control the blood and the dialysate circuits as well as monitor important variables like ultrafiltration (UF) rate, adequacy, dialysate composition, and circuit pressures. Although such advances make patient management somewhat easier for the nephrologist, they do not change the basic tenet of patient carefirst to do no harm. Consequently, it is extremely important for the practicing nephrologist to recognize and understand the terminology, significance, and management of the basic operational mechanics of HD machines.

From a practical point of view, it is often useful to divide the HD process into two main parts, that is, the blood circuit and the dialysate circuit. The standards for HD equipment in the United States are set by the AAMI (Association for the Advancement of Medical Instrumentation). Various ''alarms'' built into the system can signal impending or ongoing system malfunction. Alarms should never be taken lightly and disarming of alarms should never be practiced. The range and sensitivity of the alarms should be internally set as default and the operator should only be able to operate within the set range without being able to alter these settings, especially while HD is in progress. Alarms should be not only visible (2m) but also easily audible (70 dB). All blood alarms [air detector, arterial, venous, blood leak, trans membrane pressure (TMP), blood pump torque] should automatically shut off the blood pump, clamp the venous return line, and stop UF, thus isolating the patient.

Equipment is programmed to automatically switch to ‘‘safe mode,’’ thus essentially isolating the patient from the HD machine. This does not correct the operational characteristics that set off the alarm in the first place, however. Properly trained nurseswhotake active (and proactive) action to correct the malfunction are always the ultimate backup to ensure safety.

2.1.1 THE BLOOD CIRCUIT

The blood circuit consists of the following components:

- Pressure monitors (arterial, prepump; and venous, postdialyzer)
- Blood tubing
- Blood pump
- Heparin pump
- Air leak detector
- Clamps.

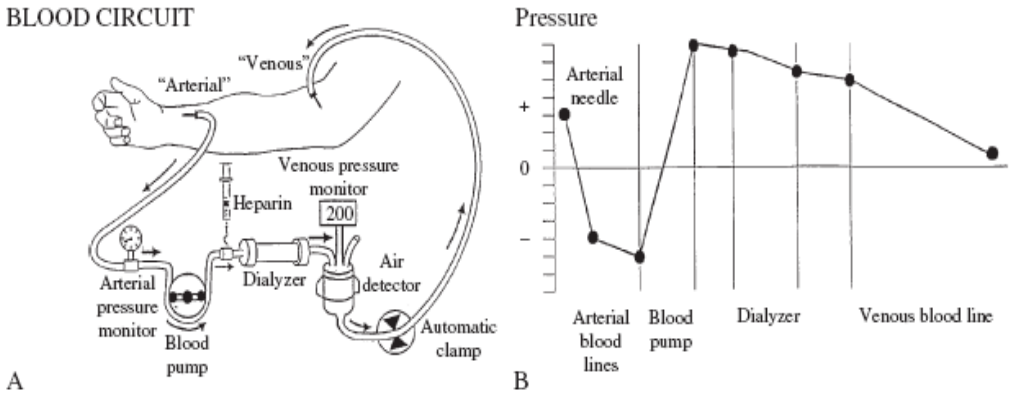


Figure 2-1 (A) The blood circuit. (B) The pressure profile in the blood circuit with an arteriovenous fistula as the blood access. (If a central venous access is used, the pressure profile will reflect the central venous pressures that are close to 0 or even slightly negative.)

2.1.2 Arterial pressure monitor (prepump)

This component monitors the pressure between the blood access and the blood pump. The pressure is negative between the access and the blood pump (Figure. 2-1B) but achieves a high positive range post-blood pump (Figure. 2-1B). The pressure transducer signal is amplified and converted to an electrical signal. Alarms may indicate patient disconnection, separation of blood tubing, or obstruction/kink in the blood circuit. The normal pressure reading in this segment of the blood circuit is negative (subatmospheric). Negative pressure makes this segment prone to entry of air into the bloodstream.

Longer needles with smaller bores increase negative pressurereadings in this segment. Likewise, negative-pressureaugmentation may be seen when longer catheters withsmaller internal diameter bores are used, especially withhigher blood flows. Out-of-range pressures trigger themachine to clamp the blood line and activate the appropriatealarms. Low arterial pressure alarm is caused by:

- Fall in blood pressure.
- Kink between needle and pump.
- Clot (check for air bubbles).
- Suction of vessel wall into the needle.

And high arterial pressure alarm caused by:

- Increase in patient's blood pressure.
- Circuit disruption between access and pump
- Unclamping of saline infusion line.
- Blood pump that has torn the pumping segment (check for blood leak).

2.1.3 Venous pressure monitor (postdialyer)

The venous pressure may build up owing to resistance tovenous return anywhere between the venous drip chamberand the venous needle (together with the access pressure).Venous pressure monitors normally read positive pressures.Out-of-range pressures trigger clamping of theblood line, stopping of the blood pump, and activation ofappropriate alarms, with shutting of the venous return. The low venous pressure alarm is caused by :

- Disruption of connections anywhere downstreamfrom the blood pump to and including the venousneedle and access.
- Low blood flow (upstream of blood pump).

Where a high venous pressure alarm (high venous pressure may rupture the dialyzer membrane!) is caused by:

- Kink in the venous return line;
- Clot in the venous drip chamber; and
- Venous access malfunction.

2.1.4 Blood tubing

Blood tubing is made of biocompatible and nontoxicmaterial. The blood tubing in the pump segment is treatedwith silicone to minimize blood clotting. Because ofits high cost, the use of silicone-treated blood tubing insingle-use systems is uncommon. Leaching of phthalate-(2-ethylhexyl) phthalate (DEHP) from polyvinyl chloride (PVC), a constituent of the blood tubing, may occurinto the blood circulation and lead to liver damage.Phthalate may very rarely lead to anaphylaxis.

2.1.5 Blood pump

Blood is pumped in the circuit by peristaltic action at arate of 200 to 600 mL/min. The pump usually has tworollers (roller rotation compresses the

tubing, thus forcingblood along the tube), operating on a low-voltage motor(less electrical hazard). The blood pump is spring-loadedto prevent under-/overocclusion of the blood tubing(the pump segment of the tubing is made up of thickerand more resilient material). The pump is adaptable todifferent sized tubing if indicated clinically and can beoperated manually in the event of a power loss. It iscalibrated to measure blood flow rate (BFR) dependingon the internal diameter of the tubing:

$$BFR = rpm(measured\ directly) \dots\dots\dots\dots\dots\dots (2-1)$$

$$x\ tubing\ volume(Pi \times r^2 \times 1) \dots\dots\dots\dots\dots\dots (2-2)$$

Where r is the internal radius of the tubing and l is the length of the tubing being compressed between the two rollers. Owing to limited rigidity, the tubing between the two rollers flattens with a high negative pressure and the above formula overestimates the blood flow at high BFR.

2.1.6 Heparin pump

The heparin pump is commonly a syringe pump,although a roller pump may be used. Heparin is infuseddownstream into the positive-pressure segment of theblood circuit (post-blood pump, predialyzer). If infusedprepump in the negative-pressure segment, the risk of airembolism is enhanced.

2.1.7 Air leak detector

The air leak detector is one of the most important features of a HD machine. It is placed distally in the venous blood line and monitors for and prevents air embolus (incidence of major air embolus is approximately in 1:2000 treatments). The usual volume of air needed to result in this complication is 60 to 125 mL (1 mL/kg/min, may vary), especially if rapidly injected. The air presents as foam with micro bubbles.

Likely points of air entry

- Arterial needle;
- Prepump arterial tubing segment;

- Open venous catheter; and
- Empty bags and infusion sets.

Requirements for air detector

- Should preferably be ultrasound (US)-based (detects

change in US frequency caused by air foam).

- Should respond to air in blood, blood and saline, orsaline alone. Because fluids transmit sound moreefficiently, a drop in the intensity of US indicatespresence of air bubbles (rate of transmission ofUS: blood > saline > air).
- Must activate alarm and stop pump.
- Must activate venous line clamp capable of completeocclusion of blood return line to 800mmHg (highcompliance dialyzers will ‘‘squeeze’’ blood into the blood circuit even with the pump stopped).
- Should not be oversensitive (to prevent unnecessaryalarms).

2.1.8 Blood tubing clamps

The blood tubing clamps should be able to withstandpressures up to 800mmHg. They should automaticallyshut if the circuit is broken or electrical power is lost (itshould be possible to open them manually if power is lost).

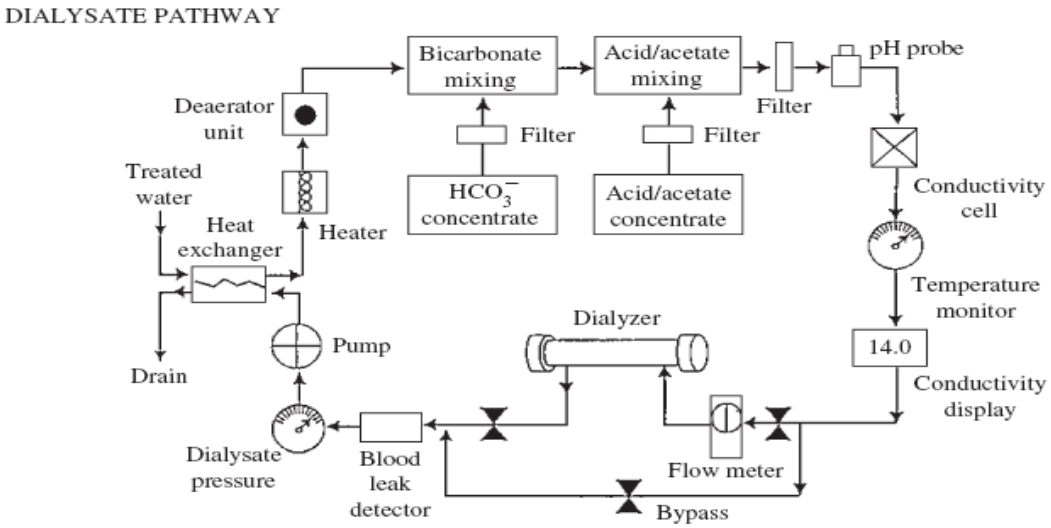


Figure 2-2: The dialysate circuit

2.1.9 Blood leak monitoring

The blood leak monitor allows detection of blood leaks and prevention of dialysate contamination by blood downstream of the dialyzer. The monitor (infrared or photo detector) has a ‘‘flow-through’’ configuration (sensor is at the bottom, and therefore, air bubbles do not interfere) (Fig. 6). Red blood cells present in the dialysate scatter light. The monitor operates by looking for loss of transparency when light is passed through the dialysate column (postdialyzer). Loss of sensitivity may occur owing to biofilm, deposits, or clots. The sensitivity of monitor is 0.25 of 0.35 mL of blood per liter of dialysate.Monitor triggers visual and audible alarms, immediatelydeactivating blood pump.

2.1.10 Dialysate disinfection and rinsing

All parts of the dialysate circuit should be exposed to the disinfectant. Adequate time for disinfection ensures adequate bacterial killing. The machine should be in bypass mode during disinfection with dialysate alarms overridden. The blood pump power supply should be off as a safeguard. The effluent dialysate line should be isolated from the drain with an air break to prevent backflow and siphoning. Heat during disinfection might caramelizedextrose causing malfunction of blood leak detectors and obstruction of valves.

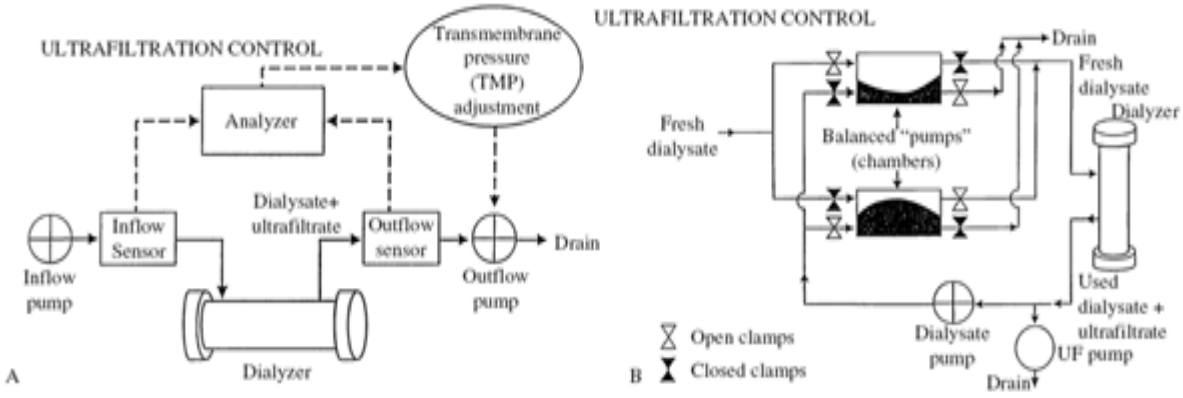


Figure 2-4 (A) flow sensor-based (B)volumetric-based

Possible sources of end toxin/bacterial contamination offinal dialysate

- Contaminated water;
- Back siphon from the drain.
- Dead space in the system.
- Inadequate disinfection.
- Bicarbonate concentrate (aqueous).

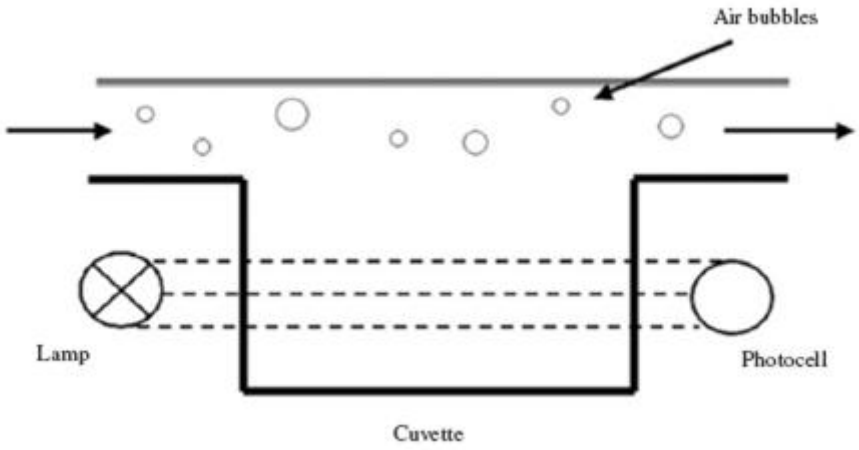


Figure 2-5 blood leak monitor

2.1.11 Power failure

The battery sets off an alarm on the machine. Remember that all systems and monitors are now OFF. The system is no longer FAILSAFE. Do not pump blood from patient into the system. Recirculate blood manually for a maximum of 15 to 30min. The venous clamp should be disconnected to return the blood through the venous line. Heparin should be introduced manually.

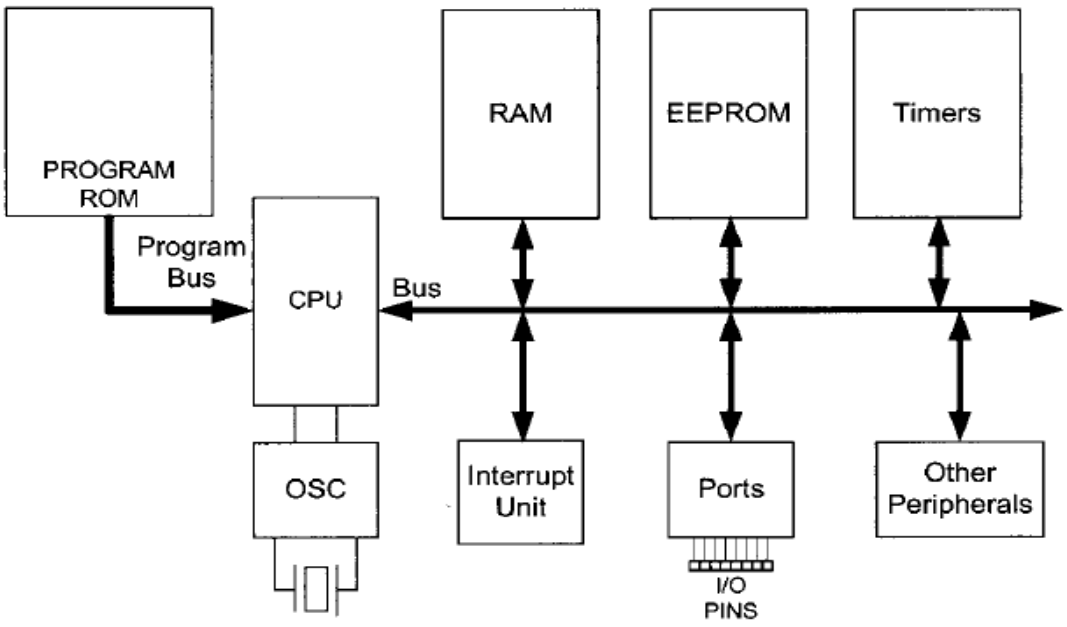
2.2 Microcontroller[4]

The basic architecture of AVR was designed by two student of Norwegian institute of technology (NTH), Alf-EgilBogen and vegradWollan, and then was bought and developed by atmel in 1996. AVR stands for Advanced Virtual RISC, or Alf and Vegrad RISC (the names of the AVR designers)

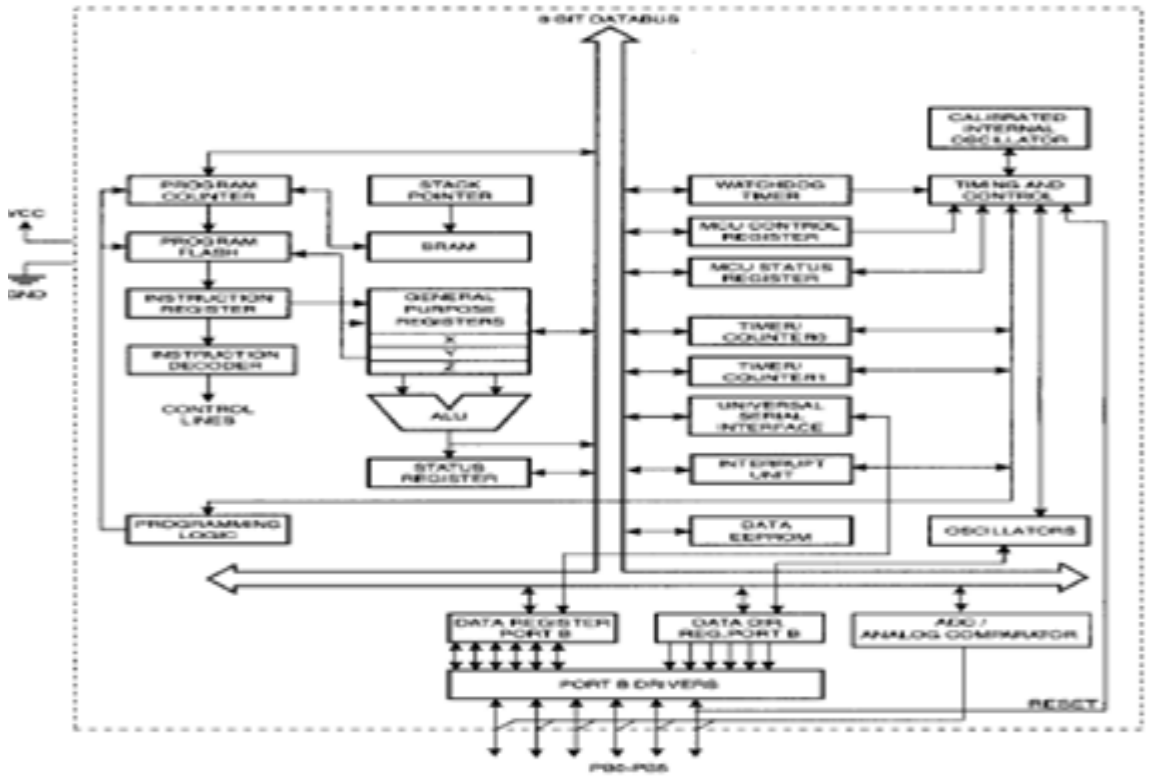
There are many kinds of AVR microcontroller with different properties. Except for AVR32, which is a 32-bit microcontroller, AVRs are all 8-bit microprocessor, meaning that the CPU can work on only 8 bits at a time. Data larger than 8 bits has to be broken into 8-bit pieces to be processed by CPU. One of the problems with AVR microcontrollers is that they are not all 100% compatible in terms of software when going from one family to another family. To run programs written for the ATtiny25 on a ATmega64, we must recompile the program and possibly change some register locations before loading it in to ATmega64. AVRs are generally classified into four broad groups: Mega , Tiny, Special purpose, and Classic.

2.2.1 AVR features

The AVR is an 8-bit RISC single-chip microcontroller with Harvard architecture that comes with some standard features such as on chip program ROM, data RAM, data EEPROM, timers and I/O ports. Most AVRs have some additional features like ADC, PWM, and so on. Figures 2-6 and 2-7 respectively .



Figures 2-6 Architecture of 8-bit AVR microcontroller



Figures 2-7 Block Diagram of ATtiny25

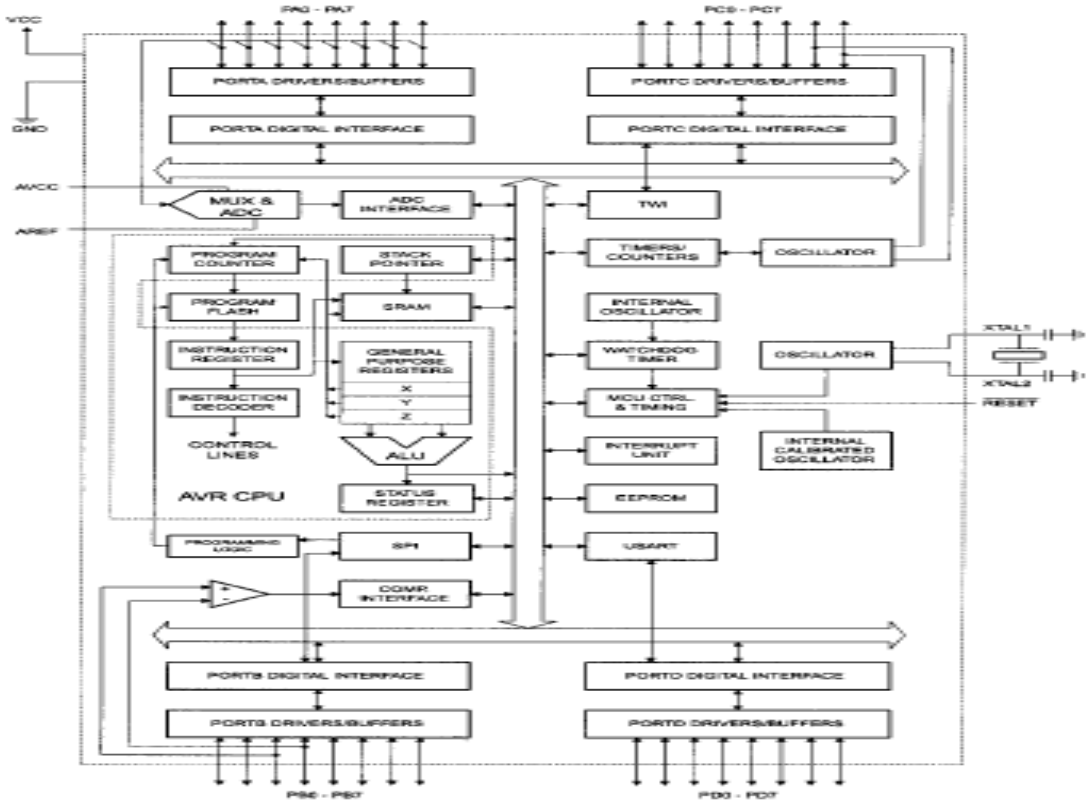


Figure 2-8block diagram of ATtiny25 interconnections

2.2.2 AVR program Rom

In microcontroller, the ROM is used to store programs and for that reason it is called program or code ROM. Although the AVR has 8M (megabytes) of program (code) ROM space, not all family members come with that much ROM writing, depending on the family member. The AVR was one of the first microcontroller to use on-chip Flash memory for program storage. The Flash memory is ideal for fast development because Flash memory can be erased in seconds compared to 20 minutes or more needed for the UV-EPROM.

2.2.3 AVR microcontroller data RAM and EEPROM

While ROM is used to store program (code); the RAM space is for data storage. The AVR has a maximum of 64K bytes of data RAM space. Not all of family members come with that much RAM. The data RAM space has three components: general-purpose registers, I/O memory, and internal SRAM. There are 32 general-purpose registers in all of the AVRs, but the SRAM's size and the I/O memory's size varies from chip to chip. On the Atmel website, whenever the size of RAM is mentioned the internal SRAM size is meant. The internal SRAM space OS used for a read/write scratch pad, in AVR, we also have a small amount of EEPROM to store critical data that does not need to be changed very often.

2.2.4 AVR microcontroller I/O pin

The AVR can have from 3 to 86 pins for I/O. The number of I/O pins depends on the number of pins on package itself. The number of pins for the AVR package from 8 to 100 at this time.In the case of the 8-pin AT90S2323,we have 3 pins for I/O, while in the case of the 100-pin ATmefa1280, we can use up to 86 pins for I/O.

2.2.5 AVR microcontroller peripherals

Most of the AVRs come with ADC (Analog-to-digital convertor),timers, and USART (Universal Synchronous Asynchronous Receiver Transmitter) as standard peripherals. The ADC is 10-bits and thee number of ADC channel in AVR chips varies and can be up to 16, depending on the number of pins in the package. The AVR can have up to 6 times besides the watchdog timer. The USART peripheral allows us to connect the AVR-based system to serial ports such as the COM port of x86 IBM PC, Most AVR family members come with the I2C And SPI buses and some of them have USB or CAN bus as well.

2.2.6 Atmega 16 Microcontroller

The ATmega16 is a low-power CMOS 8-bit microcontroller based on the AVR enhanced RISCarchitecture. By executing powerful instructions in a single clock cycle, the ATmega16 achievesthroughputs approaching 1 MIPS per MHz allowing the system designer to optimize power consumptionversus processing speed.

In order to maximize performance and parallelism, the AVR uses Harvard architecture – withseparate memories and buses for program and data. Instructions in the program memory areexecuted with a single level pipelining. While one instruction is being executed, the next instructionis pre-fetched from the program memory. This concept enables instructions to be executedin every clock cycle. The program memory is In-System Reprogrammable Flash memory.

The fast-access Register File contains 32 × 8-bit general purpose working registers with a singleclock cycle access time. This allows single-cycle Arithmetic Logic Unit (ALU) operation. In a typicalALU operation, two operands are output from the Register File, the operation is executed,and the result is stored back in the Register File – in one clock cycle.Six of the 32 registers can be used as three 16-bit indirect addresses register pointers for

DataSpace addressing – enabling efficient address calculations. One of the these address pointerscan also be used as an address pointer for look up tables in Flash Program memory. These added function registers are the 16-bit X-register, Y-register, and Z-register, described later inthissection.The ALU supports arithmetic and logic operations between registers or between a constant anda register. Single register operations can also be executed in the ALU. After an arithmetic operation,the Status Register is updated to reflect information about the result of the operation.

Program flow is provided by conditional and unconditional jump and call instructions, able todirectly address the whole address space. Most AVR instructions have a single 16-bit word format; every program memory address contains a 16-bit or 32-bit instruction program Flash memory space is divided in two sections, the Boot program section and theApplication Program section. Both sections have dedicated Lock bits for write and read/writeprotection. The SPM instruction that writes into the Application Flash memory section must reside in the Boot Program section.During interrupts and subroutine calls, the return address Program Counter (PC) is stored on theStack. The Stack is effectively allocated in the general data SRAM, and consequently the Stack

size is only limited by the total SRAM size and the usage of the SRAM,all user programs mustinitialize the SP in the reset routine (before subroutines or interrupts are executed). The StackPointer SP is read/write accessible in the I/O space. The data SRAM can easily be accessedthrough the five different addressing modes supported in the AVR architecture.The memory spaces in the AVR architecture are all linear and regular memory maps.A flexible interrupt module has its control registers in the I/O space with an additional globalinterrupt enable bit in the Status Register. All interrupts have a separate interrupt vector in theinterrupt vector table. The interrupts have priority in accordance with their interrupt vector position.The lower the interrupt vector

address, the higher the priority.The I/O memory space contains 64 addresses for CPU peripheral functions as Control Registers,SPI, and other I/O functions. The I/O Memory can be accessed directly, or as the DataSpace locations following those of the Register File, $20 - $5F.

2.2.7 Analog to Digital Convertor

The ATmega16 features a 10-bit successive approximation ADC. The ADC is connected to an8-channel Analog Multiplexer which allows 8 single-ended voltage inputs constructed from thepins of Port A. The single-ended voltage inputs refer to 0V (GND).The device also supports 16 differential voltage input combinations. Two of the differential inputs (ADC1, ADC0 and ADC3, ADC2) are equipped with a programmable gain stage, provide gamplification steps of 0 dB (1x), 20 dB (10x), or 46 dB (200x) on the differential input voltagebefore the A/D conversion. Seven differential analog input channels share a common negativeterminal (ADC1), while any other ADC input can be selected as the positive input terminal. If 1xor 10x gain is used, 8-bit resolution can be expected. If 200 x gainsare used, 7-bit resolution can beexpected.

The ADC contains a Sample and Hold circuit which ensures that the input voltage to the ADC isheld at a constant level during conversion. A block diagram of the ADC is shown in Figure 98.The ADC has a separate analog supply voltage pin, AVCC. AVCC must not differ more than±0.3V from VCC. See the paragraph “ADC Noise Canceler” on page 211 on how to connect this pin.Internal reference voltages of nominally 2.56V or AVCC are provided On-chip. The voltage referencemay be externally decoupled at the AREF pin by a capacitor for better noise performance.

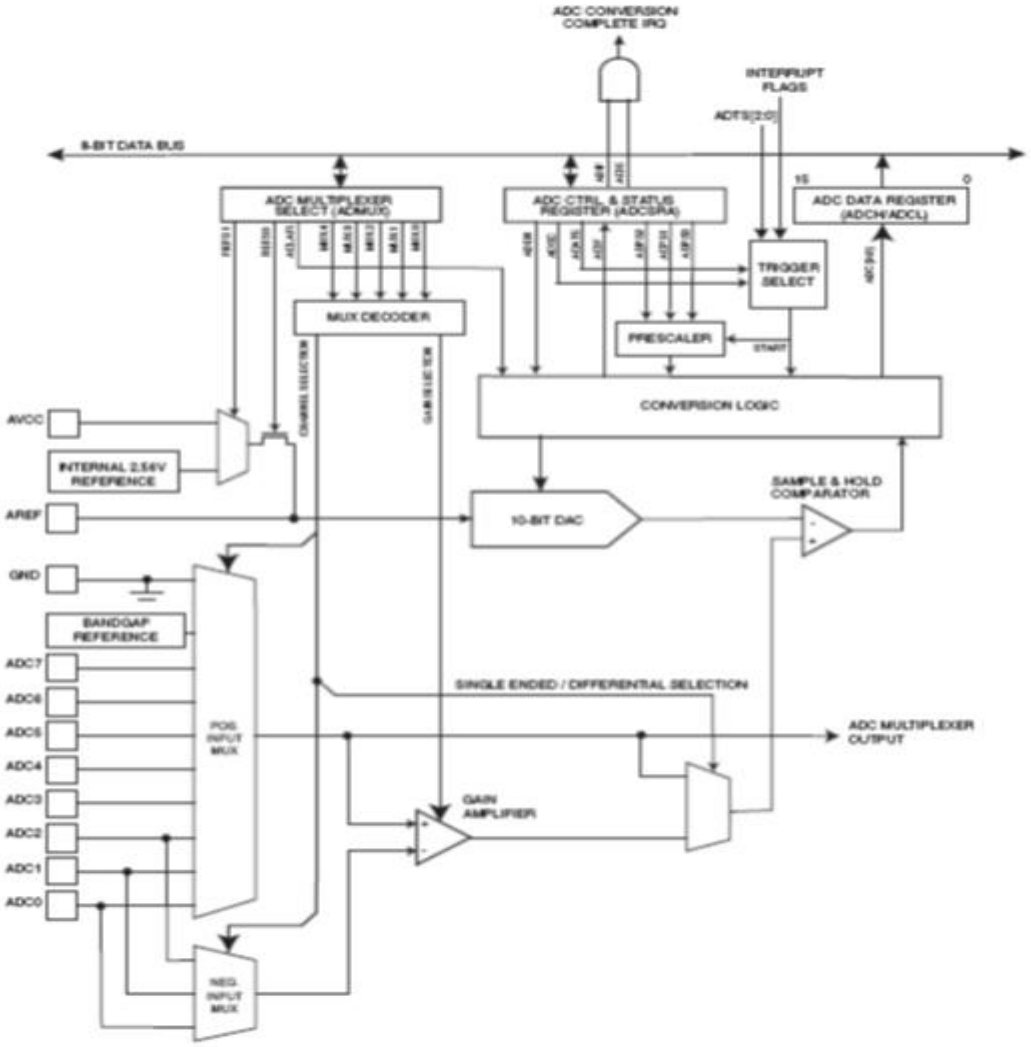


Figure 2-10 analog to digital block schematic

2.2.8 Input And Output ports of Atmega 16

All AVR ports have true Read-Modify-Write functionality when used as general digital I/O ports.This means that the direction of one port pin can be changed without unintentionally changingthe direction of any other pin with the SBI and CBI instructions. The same applies when changingdrive value (if configured as output) or enabling/disabling of pull-up resistors (if configured as input). Each output buffer has symmetrical drive characteristics with both high sink and sourcecapability. The pin driver is strong enough to drive LED displays directly. All port pins have individuallyselectable pull-up resistors with

a supply-voltage invariant resistance. All I/O pins haveprotection diodes to both VCC and Ground as indicated in Figure 3-4.

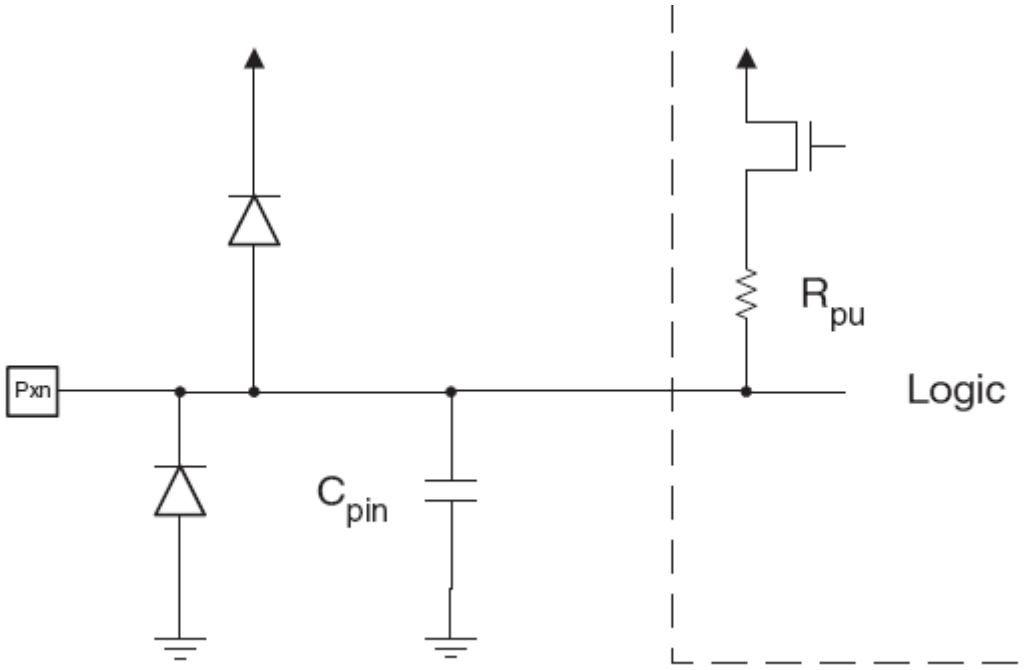


Figure 2-11I/O pin equivalent schematic

Three I/O memory address locations are allocated for each port, one each for the Data Register– PORTx, Data Direction Register – DDRx, and the Port Input Pins – PINx. The Port Input PinsI/O location is read only, while the Data Register and the Data Direction Register are read/write.In addition, the Pull-up Disable – PUD bit in SFIOR disables the pull-up function for all pins in allports when set.

2.3 Sensor Fundamentals[6]

A sensor is a device that converts a physical phenomenon into an electrical signal. As such, sensors represent part of the interface between the physical world and the world of electrical devices, such as computers. The other part of this interface is represented by actuators, which convert electrical signals into physical phenomena.

2.3.1 Importance of this interface

In recent years, enormous capability for information processing has been developed within the electronics industry. The most significant example of this capability is the personal computer. In addition, the availability of inexpensive microprocessors is having a tremendous impact on the design of embedded computing products ranging from automobiles to microwave ovens to toys. In recent years, versions of these products that use microprocessors for control of functionality are becoming widely available. In automobiles, such capability is necessary to achieve compliance with pollution restrictions. In other cases, such capability simply offers an inexpensive performance advantage. All of these microprocessors need electrical input voltages in order to receive instructions and information. So, along with the availability of inexpensive microprocessor shas grown an opportunity for the use of sensors in a wide variety of products.In addition, since the output of the sensor is an electrical signal, sensors tend to be characterized in the same way as electronic devices. The data sheets for many sensors are formatted just like electronic product data sheets.

However, there are many formats in existence, and there is nothing close to an international standard for sensor specifications. The system designer will encounter a variety of interpretations of sensor performance parameters, and it can be confusing. It is important to realize that this confusion is not due to an inability to explain the meaning of the terms—rather it is a result of the fact that

different parts of the sensor community have grown comfortable using these terms differently.

2.3.2 Sensor Data Sheets

It is important to understand the function of the data sheet in order to deal with this variability. The data sheet is primarily a marketing document. It is typically designed to highlight the positive attributes of a particular sensor and emphasize some of the potential uses of the sensor, and might neglect to comment on some of the negative characteristics of the sensor. In many cases, the sensor has been designed to meet a particular performance specification for a specific customer, and the data sheet will concentrate on the performance parameters of greatest interest to this customer. In this case, the vendor and customer might have grown accustomed to unusual definitions for certain sensor performance parameters. Potential new users of such a sensor must recognize this situation and interpret things reasonably. Odd definitions may be encountered here and there, and most sensor data sheets are missing some pieces of information that are of interest to particular applications.

2.3.3 Accuracy or Uncertainty

Uncertainty is generally defined as the largest expected error between actual and ideal output signals. Typical units are kelvin. Sometimes this is quoted as a fraction of the full-scale output or a fraction of the reading. For example, a thermometer might be guaranteed accurate to within 5% of FSO (Full Scale Output). “Accuracy” is generally considered by metrologiest to be a qualitative term, while “uncertainty” is quantitative. For example one sensor might have better accuracy than another if its uncertainty is 1% compared to the other with an uncertainty of 3%.

2.3.4 Sensor Performance Characteristics Definitions

The following are some of the more important sensor characteristics:

2.3.4.1 Transfer Function

The transfer function shows the functional relationship between physical input signal and electrical output signal. Usually, this relationship is represented as a graph showing the relationship between the input and output signal, and the details of this relationship may constitute a complete description of the sensor characteristics. For expensive sensors that are individually calibrated, this might take the form of the certified calibration curve.

2.3.4.2 Sensitivity

The sensitivity is defined in terms of the relationship between input physical signal and output electrical signal. It is generally the ratio between a small change in electrical signal to a small change in physical signal. As such, it may be expressed as the derivative of the transfer function with respect to physical signal. Typical units are volts/kelvin, millivolts/kilopascal, etc.. A thermometer would have "high sensitivity" if a small temperature change resulted in a large voltage change. Span or Dynamic Range The range of input physical signals that may be converted to electrical signals by the sensor is the dynamic range or span. Signals outside of this range are expected to cause unacceptably large inaccuracy. This span or dynamic range is usually specified by the sensor supplier as the range over which other performance characteristics described in the data sheets are expected to apply.Typical units are kelvin, Pascal, newton, etc.

2.3.4.3 Hysteresis

Some sensors do not return to the same output value when the input stimulus is cycled up or down. The width of the expected error in terms of the measured quantity is defined as the hysteresis. Typical units are kelvin or percent of FSO. Nonlinearity (often called Linearity) The maximum deviation from a linear transfer function over the specified dynamic range. There are several measures of this error. The most common compares the actual transfer function with the "best straight line," which lies midway between the two parallel lines that encompass the entire transfer function over the specified dynamic range of the device. This choice of comparison method is popular because it makes most sensors look the best. Other referencelines may be used, so the user should be careful to compare using the same reference.

2.3.4.4 Noise

All sensors produce some output noise in addition to the output signal. In some cases, the noise of the sensor is less than the noise of the next element in the electronics, or less than the fluctuations in the physical signal, in which case it is not important. Many other cases exist in which the noise of the sensor limits the performance of the system based on the sensor. Noise is generally distributed across the frequency spectrum. Many common noise sources produce a white noise distribution, which is to say that the spectral noise density is the same at all frequencies. Johnson noise in a resistor is a good example of such a noise distribution. For white noise, the spectral noise density is characterized in units of volts/Root (Hz). A distribution of this nature adds noise to a measurement with amplitude proportional to the square root of the measurement bandwidth. Since there is an inverse relationship between the bandwidth and measurement time, it can be said that the noise decreases with the square root of the measurement time.

2.3.5 Ultrasonic Sensors

Ultrasonic level sensors emit sound waves, and the liquid surface reflects the sound waves back to the source. The transit time is proportional to the distance between the liquid surface and the transmitter. These sensors are ideal for noncontact level sensing of very viscous fluids such as heavy oil, latex, and slurries. Practically, there are limitations to this method, which include:

- foam on the surface can absorb sound
- speed of sound varies with temperature
- turbulence can cause inaccurate readings.

2.3.6 Photo Sensors

Detection of light is a basic need for everything from devices to plants and animals. In the case of animals, light detection systems are very highly specialized, and often operate very near to thermodynamic limits to detection. Device researchers have worked on techniques for light detection for many years, and have developed devices that offer excellent performance as well.

Clearly, the military has been a major sponsor of light detection device research. Devices for light detection are of fundamental importance throughout military technology, and the maturity and widespread availability of inexpensive photosensors is a direct result of this DOD research investment over many years.Light is a quantum-mechanical phenomena. It comes in discrete particles called photons.Photons have a wavelength λ, a velocity $c = 3 \times 10^8$ m/s , a frequency $\omega = (2 \pi c)/ \lambda$, energy $E = hc / \lambda$ where $h= 6.67\times10^{34}$Js, and even a momentum $P = h /\lambda$. Among all of this, it isimportant to remember the relationship between energy and wavelength. In all cases ،the energy of the photon determines how we detect it.Light detectors may be broken into two basic categories. The so-called quantum detectors all convert incoming radiation directly into an electron in a semiconductor device, and process the resulting

current with electronic circuitry. The thermal detectors simply absorb the energy and operate by measuring the change in temperature with a thermometer.

2.3.7 Quantum Detectors

The quantum detectors, offers the best performance for detection of optical radiation. In all of the quantum detectors, the photon is absorbed and an electron is liberated in the structure with the energy of the photon. It is important to recognize that semiconductors feature the basic property that electrons are allowed to exist only at certain energy levels. If the device being used to detect the radiation does not allow electrons with the energy of the incident photon, the photon will not be absorbed, and there will be no signal.On the other hand, if the photon carries an amount of energy which is "allowed" for an electron in the semiconductor, it can be absorbed. Once it is absorbed, the electron moves freely within the device, subject to electric fields (due to applied voltages) and other effects. Many such devices have a complicated "band structure" in which the allowed energies in the structure change with location in the device. One example of such a "band structure" is that offered by a p-n diode. In a diode, the p-n junction produces a step in the allowed energy levels, resulting in a direction in which currents flow easily and the opposite direction in which current flow is greatly reduced.A photodiode is simply a diode, biased against its easy flow direction ("reverse-biased") so that the current is very low. If a photon is absorbed and an electron is freed, it may pass over the energy barrier if it possesses enough energy. In this respect, the photodiode only produces a current if the absorbed photon has more energy than that needed to traverse the p-n junction. Because of this effect, the p-n photodiode is said to have a cutoff wavelength—photons with wavelength less than the cutoff produce current and are detected, while photons with wavelength greater than the cutoff do not produce current and are not detected.

Photodiodes may be biased and operated in two basic modes: photovoltaic and photoconductive. In the photovoltaic mode, the diode is attached to a virtual ground preamplifier as shown in Figure 2-12, and the arrival of photons causes the generation of a voltage which is amplified by the op-amp. The primary feature of this approach is that there is no dc-bias across the diode, and so there is no basic leakage current across the diode aside from thermally generatedcurrents. This configuration does suffer from slower response because the charge generated must charge the capacitance of the diode, causing an R-C delay.

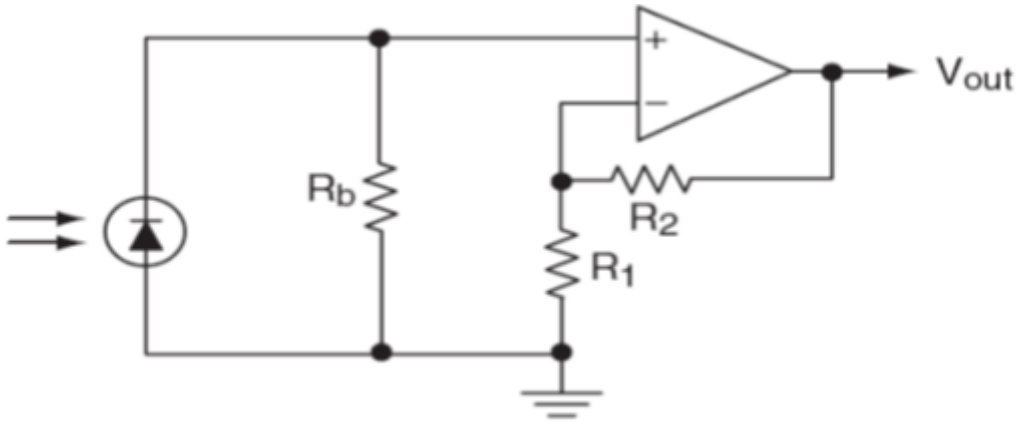


Figure 2-12connection of a photodiodein a photovoltaic mode.

2.3.8 Photoconductive mode

In the photoconductive mode, the diode is biased, and the current flowing across the diode is converted to a voltage (by a resistor), and amplified. A photoconductive circuit is shown in Figure 2-13. The primary advantage of this approach is that the applied bias decreases the effective capacitance of the diode (by widening the depletion region), and allows for faster response. Unfortunately, the dc bias also causes some leakage current, so detection of very small signals is compromised.

In addition to making optical detectors from diodes, it is also possible to construct them from transistors. In this case, the “photocurrent” is deposited in

the base of a bipolar junction transistor. When subjected to a collector-emitter bias (for npn), the current generated by the photons flows from the base to the emitter, and a larger current is caused to flow from the collector to the emitter. For an average transistor, the collector-emitter current is between 10 and 100× larger than the photocurrent, so the phototransistor is fundamentally more sensitive than the diode. Photodiodes and phototransistors are very widely available. Most semiconductor device manufacturers also offer photodiodes and transistors, so there are nearly 100 suppliers. More than 10 manufacturers specialize in photosensors. As a result, optimized photodiodes and transistors are available at very low cost.These devices are also available in packages designed for particular applications.For example, it is common to use a light-emitting diode and a detector mounted in a pair so that passing objects can interrupt the optical beam between them. Opto-interruptors consisting of such emitter-detector pairs are available in a wide variety of configurations. Proximity detectors situated side by side sense the presence of a reflecting surface by causing reflected light to strike the detector.Other applications of optical detector-emitter pairs include measurement of the rotation rate of electric motors. In this case, a disk is mounted on the shaft of the motor with a large number of slits cut through it. The detector emitter pair is mounted so that the slits cause an oscillation in the signaland the rotary position can be determinedby counting the peaks in the signal. This is called an optical encoder, or an incremental encoder, and it is widely used in electric motors, as shown in

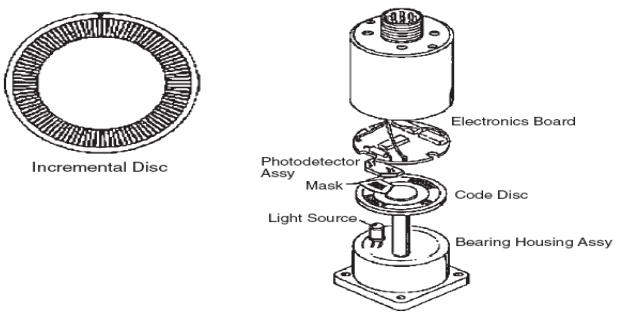


Figure 2-13 Incremental encoders.

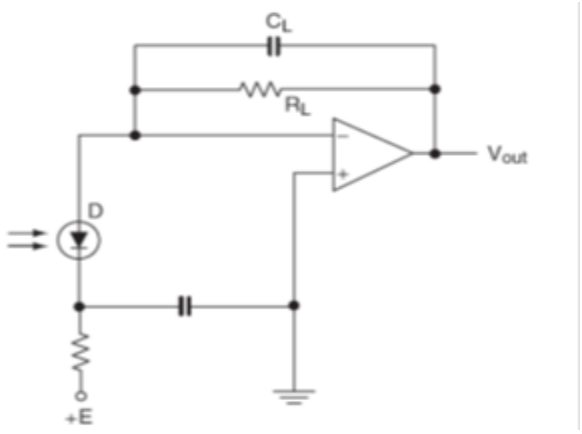


Figure 2-14Photoconductive operating mode

Most phototransistors and photodiodes have their peak sensitivity in the near infrared (see Figure 14.1.4). The peak sensitivity occurs near the cutoff wavelength (near 1 μm) and extends to shorter wavelengths. The location of this peak sensitivity is due tothe energy of the "bandgap" in silicon, and is not easily adjusted.

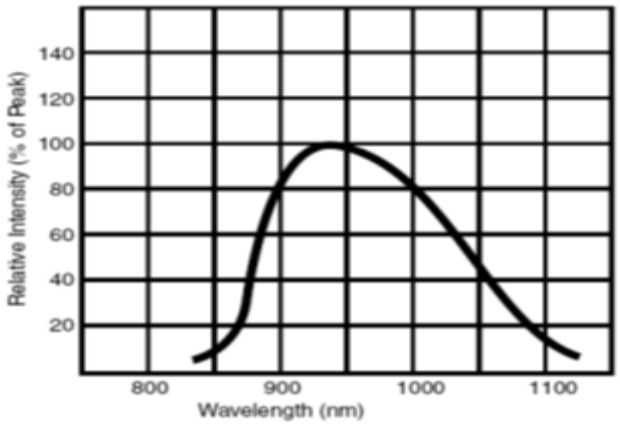


Figure 2-15Typical photo diode spectral response

Photosensors can be made from other electronic materials with different bandgaps, as shown in Table 2-1 None of these materials are as widely available as silicon, and costs for detectors made from InSb can be substantially higher.

Table 2-1 bandgaps of some semiconductors

Material	Bandgap (eV)
ZnS	3.6
CdS	2.41

CdSe	1.8
CdTe	1.5
Si	1.12
Ge	0.67
PbS	0.37
InAs	035
Te	0.33
PbTe	0.3
PbSe	0.27
InSb	018

There is another important consideration to keep in mind when selecting photosensors. In addition to the photocarriers in the device, thermally generated carriers can be produced. The distribution of energies generated by thermal processes is dependent on the thermodynamics of the device, and on the temperature. Because of this relationship, increasing the temperature causes an increase in the number of thermally generated carriers. Conversely, reducing the bandgap of a room-temperature device will also cause an increase in the number of thermally generated carriers. Silicon detectors work well at room temperature, but heating to more than 100°C starts to cause substantial increases in "dark current." Detectors made from materials other than silicon may offer increased cutoff wavelength, but may also require cooling below room temperature.In general, there is a nearly linear relationship between the maximum operating temperature and the cutoff energy for the detector. By selecting a material with a cutoff energy one-fifth that of silicon (such as InSb), it is necessary to cool the device to about one-fifth of the maximum operating temperature of silicon (cooling to 77K is optimal for InSb). This tradeoff

between cutoff and operating temperature imposes severe cost issues for operation of devices at fairly long wavelengths.

If cooling is affordable, a large selection of materials and devices with "engineered band-gaps" is available. The tremendous interest in devices with cutoff wavelengths near 10–20 μm is a direct result of the DOD interest in infrared detectors for night vision. It turns out that the peak of the infrared spectrum for objects at room temperature is in this region, and so the maximum contrast in thermal detection is available by producing devices with sensitivity in this region.

There is a simple relationship between the temperature of an infrared source and the peak wavelength of the blackbody spectrum.

$$\lambda_m = \frac{2898}{T} \dots\dots\dots\dots\dots\dots (2-3)$$

where the wavelength is in microns, and the temperature is in Kelvin. So, for room temperature, the maximum wavelength is near 10 microns.

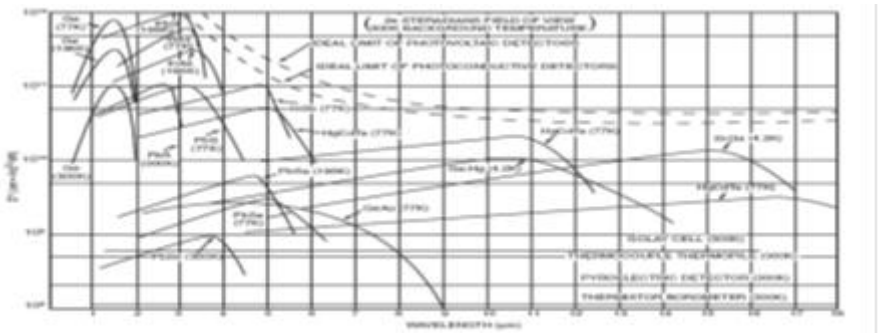


Figure 2-15 typical spectral response of infrared detectors

Of the materials most studied, the clear winner is Mercury Cadmium Telluride (MCT). It may be formulated to have cutoff between 10 and 20 microns, and offers excellent properties for infrared detection. In particular, it offers low dark current, high absorptivity, and low carrier scattering. Unfortunately, it is difficult and expensive to manufacture. As should be expected for anything containing mercury, its fabrication process is an environmental nightmare, and the basic material is not compatible with

electronics. As a result, it is "bump-bonded" onto silicon substrates for readout and signal processing. In addition, it must be operated at or below 77K, which imposes operational complications. A commercial imaging system based on MCT detector arrays generally costs near $100,000 at the present time. Most military applications (ballistic missiles, aircraft imaging systems, satellite systems) can afford this set of costs and complications, but commercial and civilian applications are generally cost-constrained. Therefore, recent research activities have focused on other materials which might be less expensive to make and operate. InSb does not offer sensitivity in the 10–20 μm region, but is more easily made than MCT, is electronics compatible, and can be operated near 100K. Research to extend the operation to higher temperatures is underway throughout academia and industry. Overall, the relationship between cutoff and operating temperature is fairly strict. MCT, which has been the focus of billions of dollars of materials research effort, has only been slightly extended to higher temperatures. There is not tremendous hope that InSb or other materials will benefit from a large change in operating requirements. The other type of optical detector, the thermal detector, does offer some hope for this problem.