Infrared Sensing and Data Transmission Fundamentals

Prepared by: Dave Hyder
Field Applications Engineer

Many applications today benefit greatly from electrical isolation of assemblies, require remote control, or need to sense a position or presence. Infrared light is an excellent solution for these situations due to low cost, ease of use, ready availability of components, and freedom from licensing requirements or interference concerns that may be required by RF techniques. Construction of these systems is not difficult, but many designers are not familiar with the principles involved. The purpose of this application note is to present a "primer" on those techniques and thus speed their implementation.

THE GENERAL PROBLEM

Figure 1 represents a generalized IR system. The transmitting portion presents by far the simplest hurdle. All that needs to be accomplished is to drive the light source such that sufficient power is launched at the intended frequency to produce adequate reception. This is quite easy to do, and specific circuits will be presented later.
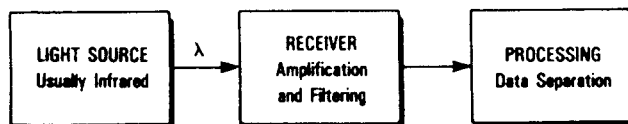

Figure 1. Simplified IR Sensing/Data Transmission System

The bulk of the challenge lies in the receiving area, with several factors to consider. The ambient light environment is a primary concern. Competing with the feeble IR transmitted signal are light sources of relatively high power, such as local incandescent sources, fluorescent lighting, and sunlight.

These contribute to the problem in two ways. First, they produce an ambient level of stimulation to the detector that appears as a dc bias which can cause decreased sensitivity and, worst of all, saturation in some types of detectors. Second, they provide a noise level often 60 dB greater than the desired signal, especially in the form of the 50 or 60 Hz power frequency. Also, recall that the sensitivity of silicon photo detectors extends well into the visible range. This sensitivity, albeit reduced, causes severe interference since the sources in this region are often of significant power, e.g., incandescent lighting and sunlight. In addition to the visible component, both produce large amounts of infrared energy, especially sunlight.

Some IR applications are not exposed to this competition, and for them dc excitation of the source may be adequate. These include some position sensing areas and slow data links over short distances.

But the bulk of IR needs require a distance greater than 30 cm, speeds greater than 300 baud, and exposure to interfering elements. For these needs high-frequency excitation of the source is necessary. This ac drive permits much easier amplification of the detected signal, filtering of lower frequency components, and is not difficult to produce at the driving end. Optical filtering for removal of the visible spectrum is usually required in addition to the electrical, but this too is quite simple.

A WORD ABOUT DETECTORS

Figure 2 shows the three basic detection schemes: a phototransistor, a Darlington phototransistor, and a photodiode. All three produce hole-electron pairs in response to photons striking a junction. This is seen as a current when they are swept across the junction by the bias voltage, but they differ greatly in other respects.
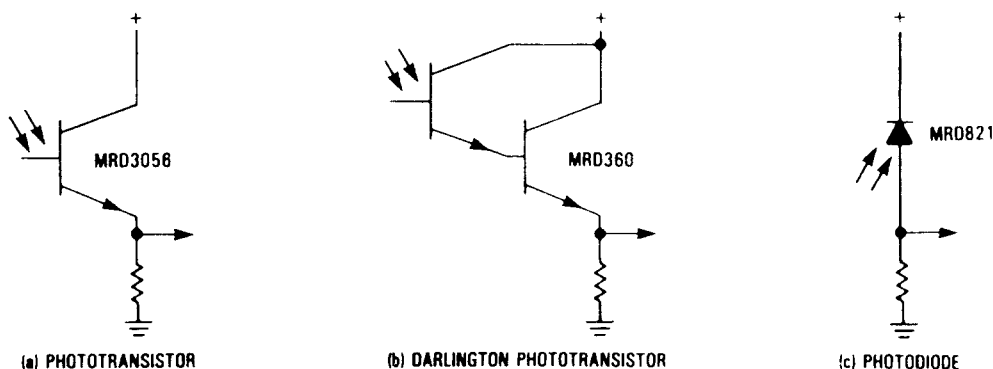

(a) PHOTOTRANSISTOR   (b) DARLINGTON PHOTOTRANSISTOR   (c) PHOTODIODE

Figure 2. The Basic Detectors for IR Photosensing

The most sensitive is the Darlington. The penalties are temperature drift, very-low tolerance to saturation, and speeds, limited to about 5 kHz (usually much less). Next is the single transistor, having similar penalties (but to a lesser degree), with speeds limited to less than 10 kHz. Typically, they are limited to less than half that number. These two detectors normally find their use in enclosed environments, where ample source intensity is available to provide large voltage outputs without much additional circuitry (their prime advantage). Their detection area is almost never exposed to ambient light.

In virtually all remote-control applications (implying distance), the diode is the detector of choice. This is due primarily to its near-freedom from saturation, even in most sunlit environments. The penalty is sensitivity, often in the nanoamp or low microamp region, but balanced by response speed in the nanosecond range. This permits transmission frequencies in the 50-100 kHz area, providing ample data rates, inexpensive amplification, and easy filtering of noise.

For more information on the internal characteristics of these devices, see the appropriate section of the Motorola Optoelectronics data book (#DL118/D).

SHORT DISTANCES

Many applications in position sensing lend themselves well to the sensitive, if slow, nature of phototransistors. When a go, no-go situation exists, these provide a simple solution provided that ambient light is not present at the detector. The designer must ensure that the system operates even if this portion of the equipment is exposed, as by opening a hatch during servicing or final adjustment during production. This is often achieved via covers, tubes limiting light paths, or that enough directionality exists in the basic device construction to provide the needed isolation. Also available for this application are logic-level output devices, usually of the open-collector type, making processor or logic interfacing convenient.

The light source for these uses is chosen primarily by the distance needed. LEDs work well up to about 5 cm. Above this, incandescents are often used due to their high output and ease of drive with low-voltage ac. Fluorescent sources are seldom adequate due to their "cool" color temperature compared to incandescent. That is, not enough output in the near-infrared or infrared portion of the spectrum.

Data can be transmitted in these short distance situations, provided the speeds required are not great. An example is the electrical isolation of two adjacent PC boards in a rack, with IR elements facing each other across the short space. Here the data can be used to drive the LED directly; modulating a high frequency is not necessary.

Speed and sensitivity are the tradeoff. The resistor used to develop a voltage can be made larger to provide increased sensitivity, but speed suffers and tendency toward saturation increases. Values of 50-200 Ω are common, but can be higher.

MODERATE DISTANCES

For the general case of remote control or sensing at distances greater than 30 cm, the vast majority of applications utilize an LED source switched at a carrier frequency of 20 kHz to 50 kHz and a diode detector coupled to ac band-limited amplifiers. Although certainly more complex than the simpler short-distance sensors, today's product offerings make it an easy task

to achieve 10 meters with a data rate of around 5,000 baud at very modest cost.

The transmission end is easily configured. Figure 3 shows a simple IR source capable of 50 kHz transmission. Note that no special techniques are needed to switch the diode at these frequencies. A burst of high frequency is created for each bit time in the data being sent. This mode of gating a carrier on and off is known as CW (continuous wave).
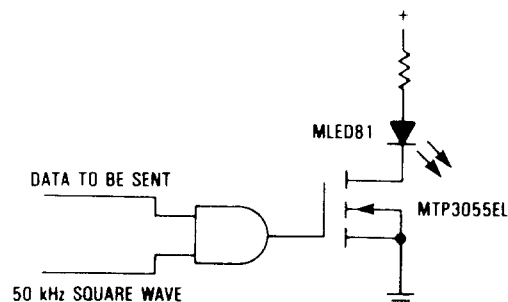

Figure 3. Basic IR Source Drive for CW Operation

The main areas of interest are the switch device and the diode current. Today's IREDs (infrared emitting diodes) are generally capable of around one ampere peak currents, but applications typically limit this to half that value. Most designs that use a 50 percent duty cycle square wave switching waveform have diode currents in the 100-500 mA range. It is important to realize that although IRED output increases linearly with drive current, it drops rapidly with increasing temperature. Therefore, reliability is not the only reason for resisting the temptation to increase range by driving the IRED harder. A diode with a 100 mA continuous rating can be reliably driven with a 200 mA square wave, and so on. It is quite common to use more than one IRED in series for increasing output and range, lowering the current requirements, and increasing reliability of the diodes.

The driver device can be a bipolar transistor or a FET. The bipolar works fine, but requires enough base current for saturation that the driving circuitry often must provide 10-20 mA or more. This may not be available directly from CMOS devices. Darlingtons solve this problem, but are usually much too slow. Another solution is an inexpensive logic-level FET such as the MTP3055EL, its physically smaller cousin, the MTD3055EL, or a MTP4N06L. This provides plenty of speed while being driven directly from any CMOS device, with absolute minimum parts count. A resistor (50-500 Ω) is sometimes used in series with the gate to moderate the very-high switching speed and noise from high frequency oscillations. The resistor is usually not needed if the gate is driven from a medium-speed CMOS gate such as the MC14081B or MC14011UB.

THE RECEIVING PROCESS

At the receiving end, the first item encountered is an IR optical filter as shown in Figure 4. This serves the sole purpose of attenuating the visible portion of the spectrum while leaving the IR intact. It can be a material specifically designed for the purpose, such as the Kodak filter series, but is usually an inexpensive acrylic plastic. This is almost any readily-available red, non-opaque plastic. Suitability is easily proven by inserting a sample between an emitter and detector while observing the

368

detector output. The IR signal should be minimally altered. This filter may be incorporated into the system as a unique piece of the material in front of the detector, or the entire front panel of the product may be made of this plastic. Sometimes lenses are actually molded from it (discussed in a later section).
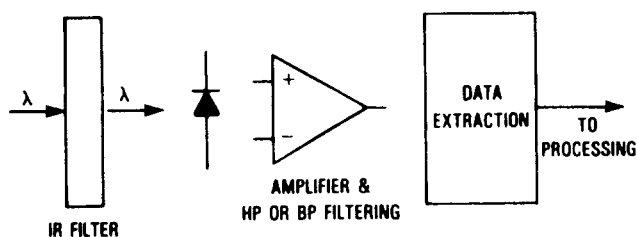

Figure 4. Basic IR Receiver

The detector diode behind the filter is usually constructed as a large-geometry device specifically designed for IR remote control, and presents a large area simply for more IR energy absorption or increased aperture. It is not unusual to find the material used for encapsulation to be red or black, and apparently opaque. The encapsulation serves as an IR filter, as in the case of the MRD821. Even so, an additional one is usually employed as mentioned above, often for the cosmetics of the product.

In addition to visible-light filtering mentioned above, electrical filtering must be applied to greatly attenuate the low-frequency interference present in both the visible spectrum and the IR. This is accomplished by three methods. First, coupling capacitance values are judiciously chosen to begin rolloff just below the transmitted frequency. This is quite effective since the area of interest is usually about a factor of 10^3, or some 9 to 10 octaves above the power-line frequencies.

The second method is to use explicit high-pass filter circuitry, but in practice this is seldom needed due to the effectiveness of the other techniques. A third option is to use a bandpass amplifier, usually with an LC tank. More discussion of this later.

After the signal is brought up to a level sufficient for detection, some method must be employed to extract the data. Most common is a simple peak detector. This detects the presence of the high-frequency pulses, charging a capacitor up to a threshold in a few cycles, at which point a comparator signals the new level. In the absence of a signal (the carrier), the capacitor discharges until the comparator's lower threshold is reached, signifying the opposite logic level. Other techniques are also available, such as the phase-locked loop, whose lock-detect output can be used as the recovered logic-level data.

MORE ON RECEIVING CIRCUITS

Two general methods are used to begin the amplification. First the diode light current (a few microamps or less) may be used to develop a voltage across a series resistance, which is then capacitively coupled to the amplifier using the rolloff of low frequencies mentioned above, as shown in Figure 5a. Second, the current may be driven directly into the amplifier, as in Figure 5b, where the photo current is summed with the feedback current at the amplifier input. Note that in these and other figures, the amplifier symbol does not necessarily denote an actual integrated operational amplifier, but may symbolize a discrete amplifier.

Figure 6 shows an amplifier system coupled to a bandpass amplifier centered about 50 kHz. Here the front end is actually an operational amplifier, used in the mode of Figure 5b. Various choices for operational amplifiers exist; perhaps the first hinges
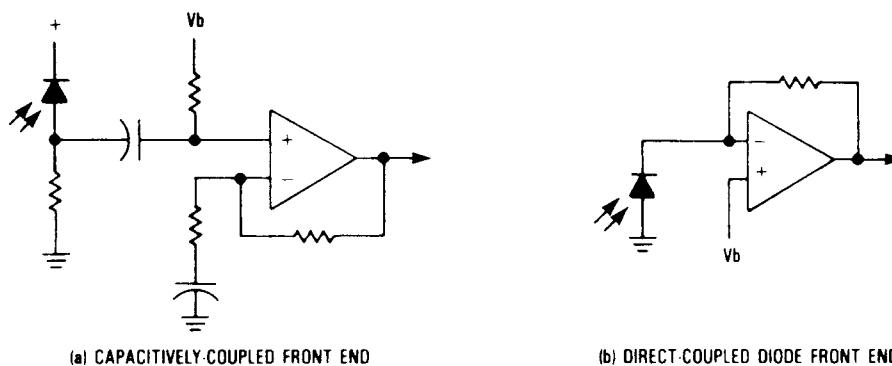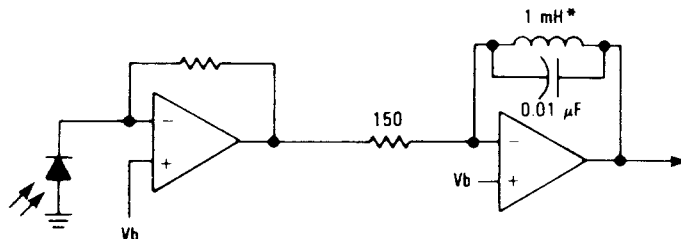

(a) CAPACITIVELY-COUPLED FRONT END

(b) DIRECT-COUPLED DIODE FRONT END

Figure 5. Front-End Amplifier Options

*Toko type 10 PA or equivalent. Available from Digi-Key Corporation, phone (800) 344-4539.

Figure 6. Amplifier Chain Showing 50 kHz Bandpass Filter Second Stage

on the supply voltage. Some recent advances in the technology have greatly increased slew rates and gain-bandwidth products. This has permitted devices that are capable of operation on a single 5 volt supply, yet can be used in the 50 kHz range. An example of this is the MC34072 series, whose input common mode range includes ground, permitting the diode or the other amplifier input to be referenced there. If greater gains are needed, and higher supply rails are available, the MC34082 series provides slew rates of 25 V/μs, or twice that of the MC34083. These operational amplifiers in general do not have the low-noise performance of discrete versions, with the above devices being in the 30 nV/√Hz region. However, the MC33077 provides excellent noise performance of about 4.5 nV/√Hz at a similar slew rate on a 5 volt supply, although its common mode range does not include ground. A simple discrete amplifier example is shown in Figure 7.

Another option that should be considered for data reception is the MC3373 (Figure 8), which integrates many of the functions already described. This device contains the front-end amplifier, a negative-peak detector with comparator, and requires only a few external components. The amplifier may

have the diode directly connected to it, or ac coupled for purposes of rolloff. A tuned circuit can be used for the better noise performance of a band-limited system. Some words of caution: supply bypassing close to the device, particularly at the gain-determining impedance (resistance or tuned circuit), is critical. Without proper bypassing, gain and range suffer. Also, a higher supply voltage of around 12 volts or so assists in greater range performance.

The vast majority of IR links in consumer products (VCRs, TVs) use an LC tank. The inductor is a shielded, adjustable slug type in the 1-5 mH range. Shielding in the form of a metal can usually encloses the entire subassembly, and the designer should expect to employ such shielding in most applications requiring moderate or long distance operation.

Note that in Figures 7, 8, and 9 the bias supply to the receiving diode is heavily decoupled from the supply via an RC. Any noise present at this point directly impacts system noise and sensitivity. Bandwidth is also often limited at the upper end as an aid in overall noise performance as seen in Figures 7 and 9. These amplifiers use small capacitors (33 pF, 10 pF, 100 pF) to roll off frequencies above 100 kHz.
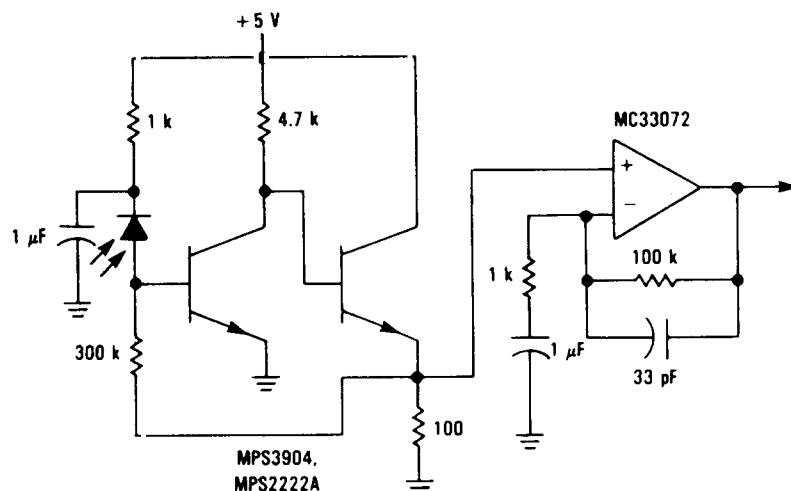

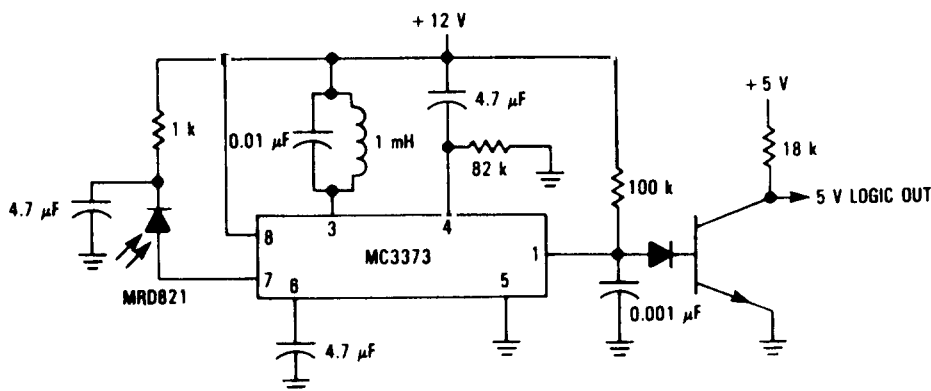Figure 7. Simple Discrete Front End with Op Amp

Figure 8. IR Receiver Using the Integrated MC3373

370

LONG DISTANCES

When the distance to be covered extends beyond 10 meters or so, other methods must be considered. The methods described below have resulted in ranges of 100 meters or more.

At the transmitting end, most of the options available center on increasing the power output. One way is to increase the IRED current, but this is subject to limits as previously discussed. Another solution is to use multiple diodes in series, often three. Note that this does not require additional supply current. Multiple diodes also provide one solution to those applications requiring less directionality, with the IREDs being slightly misaligned from one another.

The diodes can also be driven much harder, and produce proportionally higher instantaneous power, if they are pulsed with a very-short duty cycle. Currents of about an ampere are common, but for only a few microseconds and with a duty cycle of 5 percent or less. This also requires modified receiving techniques.

At the receiving end, most solutions center on increasing the aperture of the system such that simply more energy is gathered. Multiple receiver diodes can be connected in parallel, adding their currents, with the additional possibility of reducing directionality if needed. Another technique is to add a lens, with the diode being placed at the focal point. In higher volume production, this is often molded into a front panel and is usually of the red filtering plastic mentioned earlier. Some systems make use of a flat Fresnel lens, being somewhat more difficult to mount but very effective. They can also be hidden behind a plastic panel.

Front-end amplifiers superior to the simple operational amplifier or discrete versions already mentioned may be found in these highest-performance situations. Such an amplifier is shown in Figure 9, where low-noise transistors are used in a circuit designed specifically for low-noise applications.

When pulsed sources are used, some encoding scheme is normally used to transmit the data. One common technique is to use a single pulse for one edge of a data bit, and two or more closely spaced pulses to signal the opposite edge. These are simply differentiated by some flip-flops and a small amount of timing circuitry. Other schemes use multiple pulses at close intervals to indicate one logic level, and a differing number to denote the other.
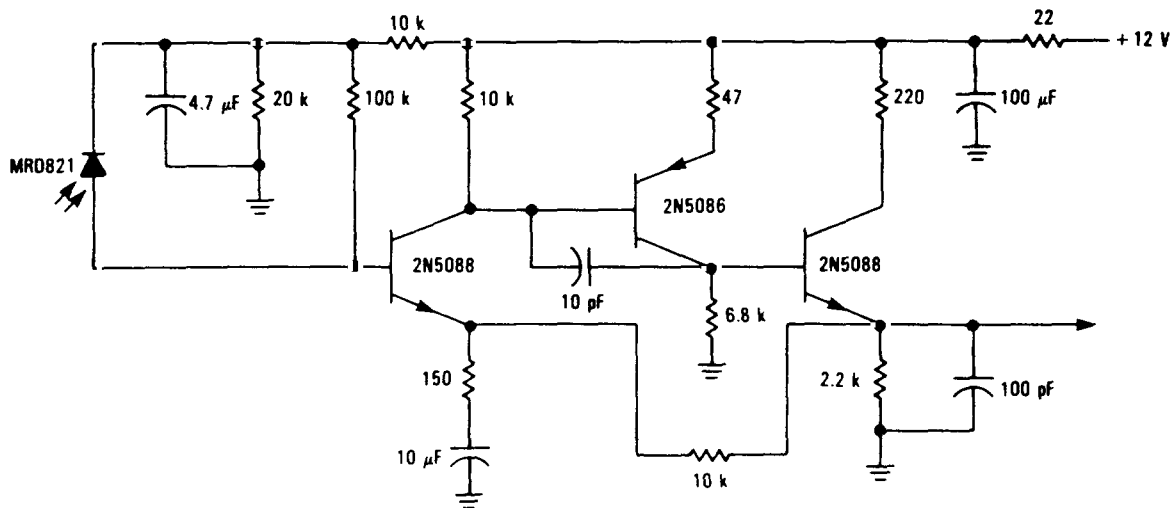

Figure 9. High-Performance Discrete Front-End
Amplifier with Special Attention Paid to Noise

371

One last option is sometimes seen at the end of the amplifier chain and used for the data detection. An analog phase-locked loop circuit can be used to pull a signal from noise and lock to it if appropriate. This lock signal is then used as the recovered data stream. One such device, shown in Figure 10, is the EXAR XR567, a small 8-pin tone decoder with both Type I and Type II phase detectors. It is capable of locking to analog signals in the 25 mV range, and makes/breaks lock at a rate sufficient for about 5,000 baud with 50-100 kHz inputs. The device can be operated up to about 500 kHz.

An advantage of the all-analog system is that the signal never needs to be amplified to the point of rail-to-rail limiting.

Thus, system-wide noise potential is decreased. Back-to-back diodes or similar methods are normally employed ahead of the loop input to hold the signal within a few hundred millivolts to protect against overdrive at close ranges.

CONCLUSION

As can be seen from the above discussion, IR links have become quite easy to implement. With the basic principles in mind, the designer should be able to adapt the techniques mentioned here to his specific system needs.
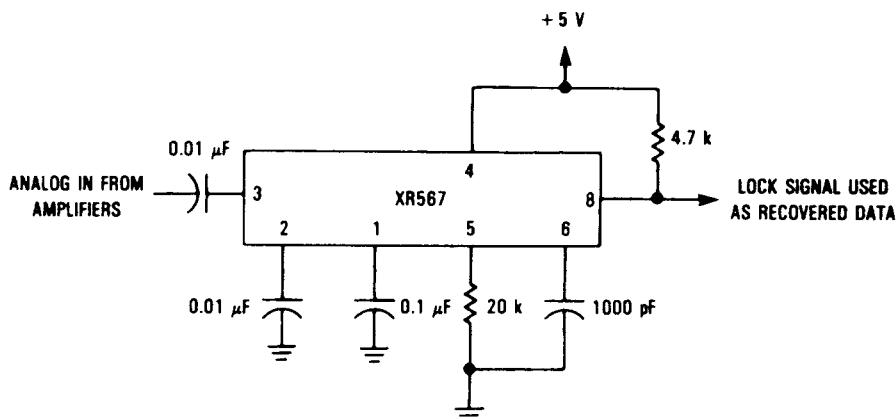

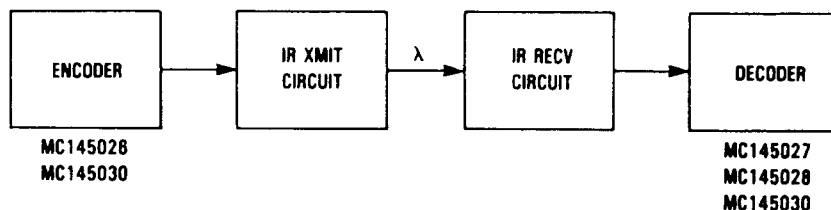**Figure 10. PLL Tone Decoder Used to Recover Data From Analog**

Figure 11. Utilizing Motorola's Encoders and Decoders