Infrared Wireless Links

Part 3

Increasing the distance and using RS-232 ports

This month, I expand upon my previous two articles about basic circuits you can build for wireless communication. This time around, topics include techniques for increasing the range of an infrared link and an infrared link that connects to RS-232 serial ports.

Increasing Distance

In earlier articles, I described an infrared communications link using Motorola's MC145026/27 encoder and decoder chips. Briefly, the encoder reads four bits of data and a five-bit address at its inputs and outputs the information as an encoded stream of serial data. An infrared-emitting diode, or IRED, transmits the encoded data by pulsing at 40-kHz to signify a logic high and remaining off to signify a logic low.

At the receiving end, an infrared receiver module detects the presence or absence of the pulsed infrared energy and converts it back to the original form generated by the encoder. The decoder chip examines the received signal, determines whether the transmitted address matches its own and, if so, latches the four bits of data to its outputs. To guard against errors, the encoder sends each transmission twice, and the decoder must receive two identical transmissions before it latches the data to its outputs.

When you have the link up and running, one of your first challenges is to see how far you can reliably transmit. Two ways to increase the distance of the link are by increasing the power of the

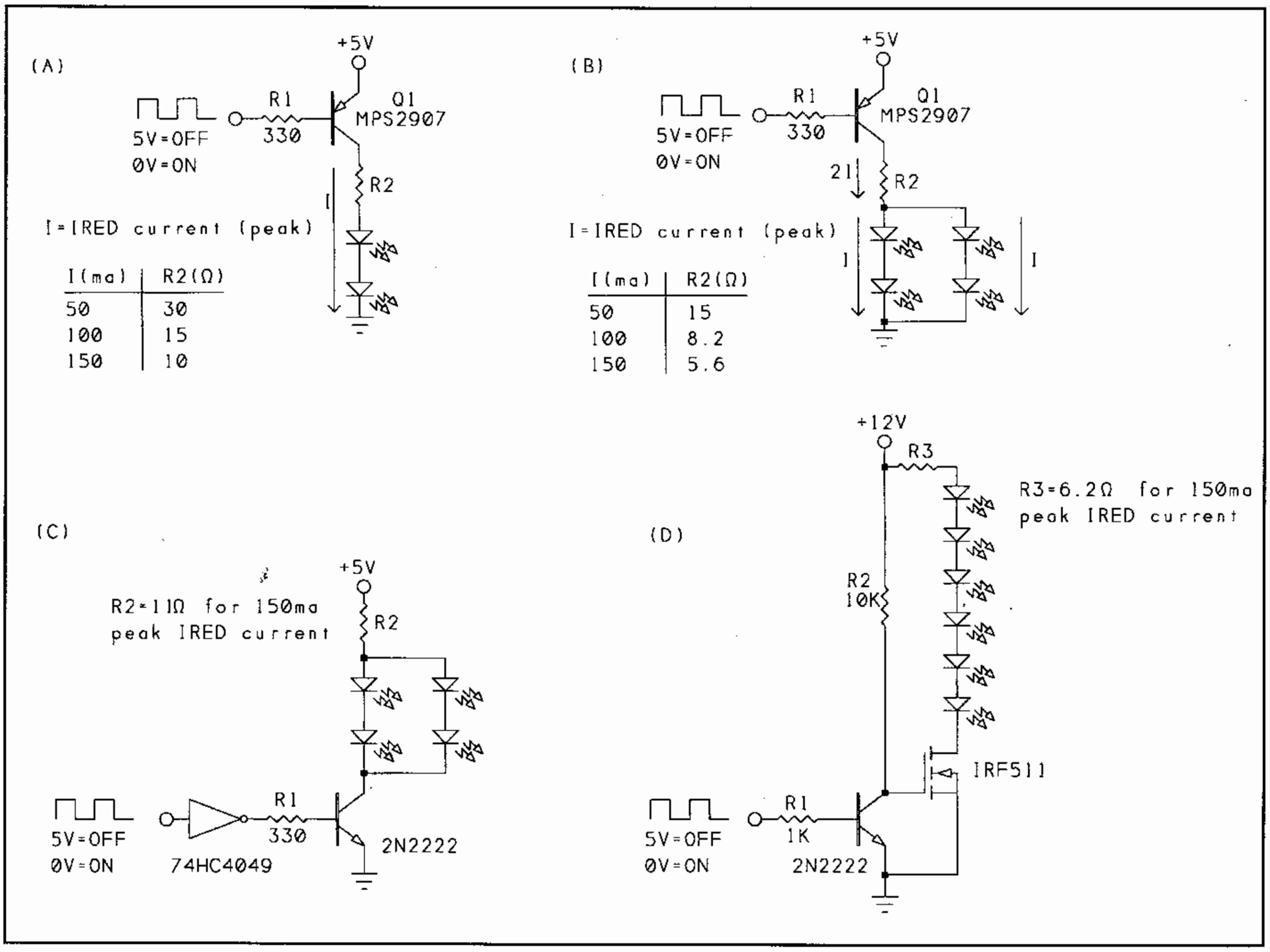

Fig. 1. To increase power of an infrared link, (A) add an IRED in series, (B) add IREDs in parallel, (C) add buffers to drive additional IREDs or (D) drive a MOSFET at 12 volts.

transmitted signal and by focusing the signal more precisely on the receiver.

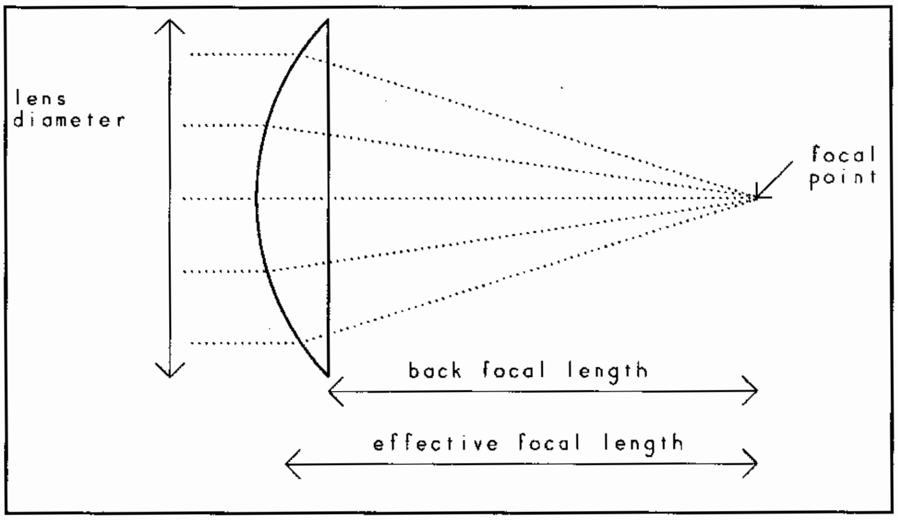

Fig. 2. At transmitter, a lens can focus transmitted infrared energy into a beam, and at receiver, it can focus received energy onto a detector.

Infrared Basics

Like visible light, infrared energy is a form of electromagnetic radiation. "Infra" means below, meaning that infrared frequencies are just below those of red (visible) light. Infrared frequencies are invisible because they're beyond the range that can be detected by the human eye.

Since wavelength is the inverse of frequency, infrared wavelengths are longer than those of visible light. Visible light covers the range 400 to 700 nm (nanometers), while infrared includes 700 nm through 1-million nm (400 nanometers is 0.4 micron, or 4,000 Angstroms, if you prefer these units.)

Infrared-emitting diodes, or IREDs, are low-cost, readily-available sources of infrared energy. An IRED is a semiconductor diode that emits infrared energy when a forward current passes through it. If you're familiar with light-emitting diodes (LEDs), IREDs are similar, except that their composition causes them to emit infrared energy instead of visible light.

An IRED emits energy at a specific wavelength. Two popular types are GaAs (gallium-arsenide), which emits energy at 940 nm, and GaAlAs (gallium-aluminum-arsenide), which emits energy at 880 nm. These are both in the range known as near infrared, signifying that their wavelengths are close to the visible spectrum.

Infrared detectors are also specific in the wavelengths they detect, although most will respond over a range of frequencies. For example, the Sharp GP-1U52X receiver module is most sensitive at 980 nm, but it will also respond to the longer-wavelength emissions from GaAs and GaAlAs IREDs. Although GaAs IREDs are a closer match at 940 nm, the GaAlAs IREDs are generally more efficient. So they may work as well even though 880 nm isn't as good a match with the detector.

You can increase the strength of an infrared signal in either of two ways: by increasing the current through the IREDs, or by increasing the number of IREDs. Figure 1 shows both options in a variety of circuits. All connect to the output of the NAND gate that combines the encoder's output and the 40-kHz oscillator in the circuits presented earlier.

A simple way to double the power is to use two IREDs in series, as shown in Fig. 1(A). With about 1.7 volts across each IRED, the series combination drops 3.4 volts. Instead of wasting energy by dropping 3 volts across a resistor, some of the current does useful work by powering a second IRED.

The maximum possible current through the IREDs is determined by the transistor's base current and gain. Outputs in the 74HC (high-speed CMOS) logic family can sink up to 25 mA (absolute maximum) and are a good choice for driving the base.

Resistor *R2* controls the amount of current through the IREDs. To determine a safe current through an IRED, you need to know the specifications of the IRED you're using as well as how you plan to use the IRED in your circuit. The data sheet for any IRED should include an absolute maximum rating for continuous current. This is the maximum current that the device can withstand without suffering damage. For example, for Harris' F5D1, a typical IRED that's available from Digi-Key, among other suppliers, this value is 100 mA. When the IRED is powered continuously, the current through it shouldn't exceed this value. In fact, since this is an absolute maximum, it's a good idea to stay well below it.

The infrared transmitter doesn't require the IRED to be on continuously. Instead, it pulses the IRED at 40 kHz. In non-continuous, or pulsed, operation, the IRED can handle much greater currents. The amount of allowable current depends on the duty cycle of the pulse, which equals the width of the pulses divided by their rate of repetition. Unfortunately, data sheets are often not too specific about how to determine the limits for a particular pulse width and repetition rate. Occasionally, you get a graph of maximum forward current plotted versus pulse width and duty cycle. Other devices just offer a few examples.

The F5D1's data sheet includes just two ratings for pulsed operation. For 10-s pulses repeating at 100 Hz, the IRED's maximum peak current is 3 amperes, or 30 times the continuous rating. For even shorter 1-s pulses, repeating at 200 Hz, the maximum is 10 amperes. But neither of these describes the situation of the infrared transmitter.

In the infrared link, the amount of time an IRED is on depends on what information it's sending and how often it transmits. When an IRED is pulsed at 40 kHz, it's on for just half of each 25-s cycle. But the IRED pulses only when transmitting logic-high outputs from the encoder. For logic-low outputs, and when no data is transmitting, the IRED is off.

With the encoder chip clocked at 1 kHz, an encoded "1" contains two 3.5-ms high pulses and two 0.5-ms low pulses. This means that the IRED is pulsing almost 90% of the total time. If the 40-kHz oscillator has a 50% duty cycle, the IRED is on for half of the pulsing time, or 45% of each transmission.

If you send a lot of 0s (if the receiver's address is 00, for example), or if you send only occasional short transmissions, the average current will be much less.

In Fig. 1(A), with *R2* at 30 ohms, peak current through the IREDs is about 50 mA, and average current is less than 25 mA, which is well below the 100-mA limit. Even at a peak current of 150 mA, the average over each transmission cycle will be less than 70 mA, which is probably a safe level.

If you do pulse the IRED at 100 mA or more, you must be very careful to design your circuit so that the IRED never comes on continuously. When not transmitting data, the IRED should be off. At greater currents, it's a good idea to use a current-limiting resistor with a 1/2-watt or greater power rating.

To permit greater peak currents, one method sometimes used is transmission of narrower pulses. At the same repetition rate, narrower pulses result in smaller average currents through the IRED.

If you use the GP1U52X infrared receiver module, however, it's probably best to transmit a square wave, with equal, or nearly equal, on and off times for the IRED. While experimenting, I found that the module will detect shorter pulses, up to a limit. But the data sheet shows a square wave as input, although there's no actual specification for the input's duty cycle.

If two IREDs aren't enough, you can add two more in parallel, as illustrated in Fig. 1(B). The current-limiting resistor is smaller in value because it drops the same voltage drop but has twice the current flowing through it.

Figure 1(C) shows four IREDs powered by an npn transistor. A 74HC4049 inverter controls the transistor's base current. With multiples of this circuit, you can have as many IREDs as your power supply can support.

And finally, if you have a 12-volt source available, you can add up to six IREDs in series, as illustrated in Fig. 1(D). The IRF511 MOSFET turns on when a voltage is applied to its gate. To turn on fully, the MOSFET requires a gate drive in excess of 5 volts.

For more-powerful transmissions to a specific receiver, you can mount multiple IREDs in a cluster, all pointing at the receiver. If you want to transmit to multiple receivers, or if a receiver's exact location is unknown, you can mount the IREDs so that they transmit across a wider-dispersed path.

Using Lenses

Another way to increase the range of a link is with optical lenses. A lens is a transparent material with at least one curved surface. A positive, or converging, lens is thicker in the center than at the edges. When parallel rays of energy pass through it, the lens refracts, or bends, them inward, causing the rays to meet at a focal point beyond the lens, as in Fig. 2. In contrast, negative (diverging) lenses are thinner in the center than at the edges, and bend the rays outward.

Positive lenses are useful for focusing the transmitted energy in infrared links. At the receiver, a positive lens can gather the infrared energy that hits it, and focus this energy on the detector. If the lens has a larger diameter than the detector, more energy will strike the detector than would be the case if no lens were used.

At the transmitting end, the lens works in reverse. It intercepts the diverging rays from the transmitter and focuses them into a narrow beam that it focuses on the active element of the detector.

Lenses have several specifications that define their performance. These include diameter, which is the width of the lens measured at right angles to the axis of transmission and effective focal length, which is the distance from the optical center of the lens to its focal point. Back focal length measures from the surface of the lens that's nearer the focal point, rather than from the center.

Many lenses are made from glass, but plastics are also used in their manufacture. For infrared links, you have to be sure that your lens is transparent to infrared frequencies, or nearly so. Although glass absorbs many infrared frequencies, it's transparent to the near-infrared frequencies of IREDs.

Lenses can be useful, but there are limits. If a lens is perfectly focused on the detector, the latter won't see anything at all unless the transmitter is aimed perfectly. You can compromise by focusing slightly beyond the detector, which gives a wider beam at the detector and requires less-critical alignment.

Some IREDs are manufactured with integral lenses that focus the output into a beam. For example, Harris' F5D1 and F5E1 IREDs are identical, except that the F5D1 has a lens that aims the energy in a narrow beam, while the F5E1 has a flat window and wider beam angle. An IRED with an integral lens is an easy, low-cost option, if it can do the job. The flat-window type is useful if you want to add an external lens or if you want a wider beam to reach multiple receivers around a room, for example.

Shown in Fig. 3 are four types of positive lenses. A plano-convex lens has one flat face and one outward-curving face. The flat face makes it easy to mount the lens on a flat surface. A double-convex lens has two outward-curving faces. Compared to a plano-convex lens with the same curvature, a double-convex

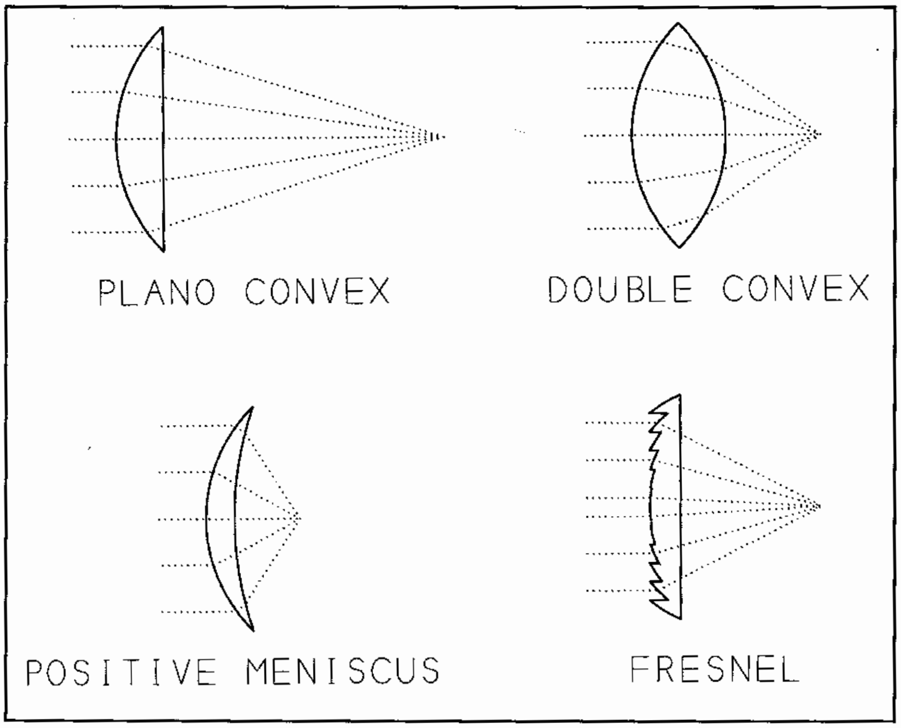

Fig. 3. Shown here are four types of positive lenses.

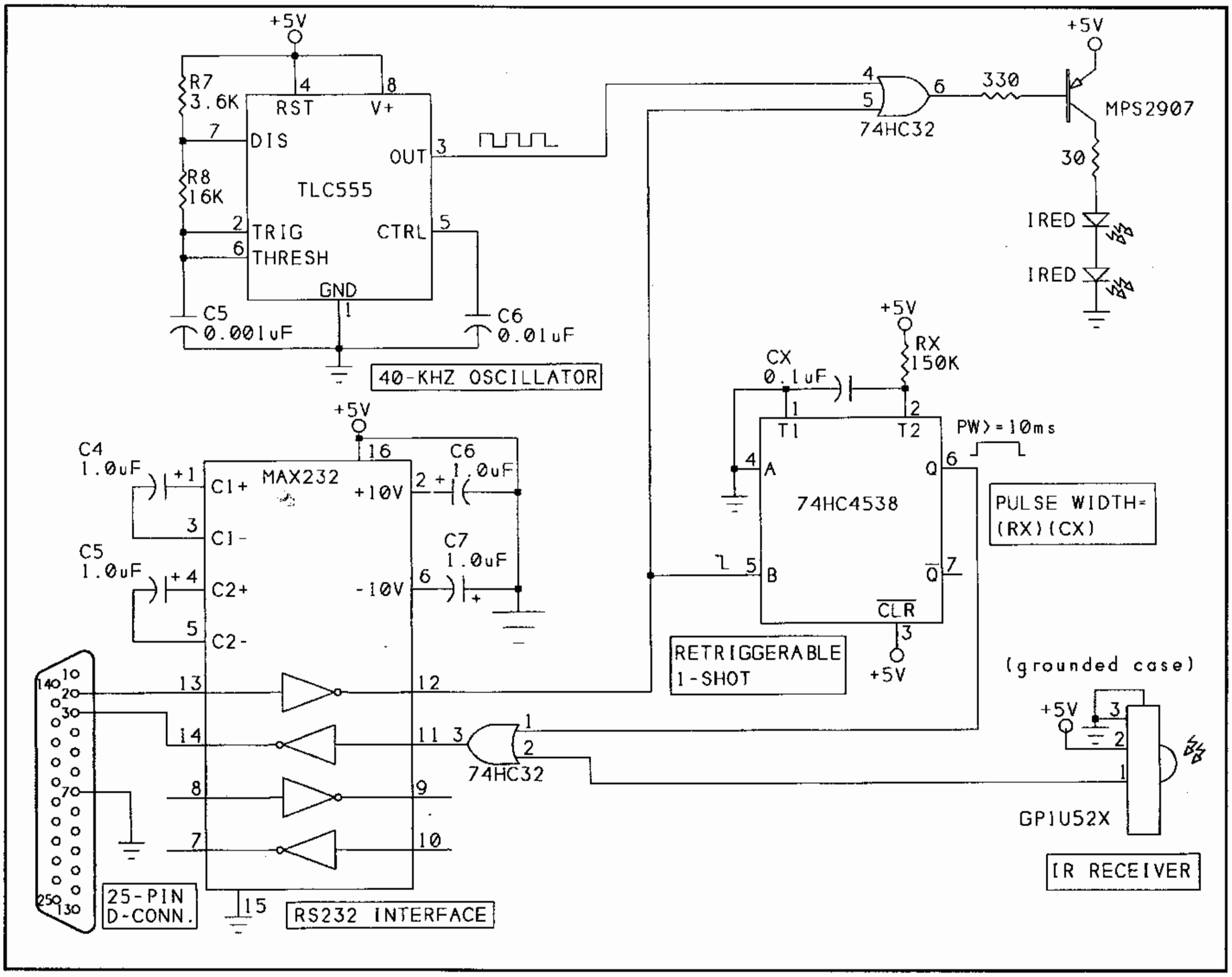

Fig. 4. The above illustrates one end of an infrared link that connects to an RS-232 serial port.

lens has half the focal length. A positive meniscus, or concavo-convex, lens has one outward-curving face and one inward-curving face, with the outward curve sharper than the inward curve. This is the type of lens used to correct for farsightedness in eyeglasses. A Fresnel lens is more complex in structure than the previous three types. It contains of a series of rings of concentric grooves, each of which resembles a section of the surface of a plano-convex lens. It's as if you sliced a series of concentric rings, all of the same height, from a convex lens, and then arranged the rings one inside the other on a flat surface. The outer rings curve more sharply than the inner ones.

The performance of a Fresnel lens is similar to that of a single lens that its rings appear to be cut from. Compared to traditional lenses, Fresnels are thin and light in weight. Many are plastic, with some specified for use with infrared wavelengths. Lighthouses use Fresnel lenses to form a concentrated narrow beam.

If you're interested in experimenting with lenses, Edmund Scientific is one source, with a huge selection, including inexpensive "educational" lenses, lens mounts, optical benches and books on optics.

Although infrared links are most often thought of as simple, line-of-sight paths —for transmitting across a room, for example—optics can extend the reach of a link. With mirrors, for example, you can transmit around corners.

RS-232 links

Links that use the encoder/decoder chips are ideal for sending occasional short bursts of data. If you want to send greater quantities of data or use higher transmission rates, one possibility is to interface infrared transmitters and receivers to a pair of RS-232 serial ports.

The GP1U52X receiver module can receive transmissions at up to 1,200 bits per second (bps). Since each transmitted byte typically includes start and stop bits, the actual transmission rate for your data will be less than this. Still, it's faster than the MC14026's rate of around 20 data bits per second. The tradeoff is that you have to do without the decoder's error-checking and automatic rejection of spurious signals and transmissions intended for other receivers. Through programming, you can add similar features, but this will slow the rate of data transmission. For example, if you send the data twice to verify, you cut data transmission rate in half.

Figure 4 shows one end of a two-way infrared link that interfaces to an RS-232 port. For a complete link, you need one of these circuits at each end. Some of the components, including the receiver module and 40-kHz oscillator, are the same as those used in the encoder/decoder circuits. A MAX232 chip translates between the RS-232 voltages and the 5-volt logic used by the infrared circuit.

The IRED is controlled by the logical OR of the MAX232's data output and a 40-kHz oscillator. When pin 12 of the MAX232 is a logic high, which occurs when the port is idle or transmitting a 1, or "mark," the IRED is off. When pin 13 is a logic low, indicating a 0 or "space," the IRED pulses at 40 kHz.

The IRED is aimed at the receiver at the other end of the link. At this end, a GP1U52X infrared receiver detects the transmitted signal. The receiver's output is low when it detects infrared transmissions at 40 kHz and high when it doesn't.

In a wired RS-232 link, each direction transmits on its own wire. With two wires (and signal ground), you can transmit and receive at the same time. In the infrared link, however, there are no wires from one end to the other, and you need to do something to isolate the two paths from each other. Otherwise, when one end transmits, the receiver at the transmitting end will detect the transmissions from the nearby IRED. This occurs even though the transmissions are aimed at, and intended for, the opposite end. This can cause problems, since everything that you transmit will be received at both ends.

There are several ways you might try to isolate the two directions of transmission. One possibility is optical shielding or placing a physical barrier that keeps the IRED's emissions from reaching the nearby detector. I didn't have much luck with this in my experiments, however.

Another possibility would be to use receivers that are tuned for different modulation frequencies. I gave this a try, using the 40-kHz GP1U52X and Lite-On's LTM-8834-2, which is tuned to detect emissions at 32.7 kHz, and adjusting the oscillators to match. The receivers weren't selective enough, though, and both still detected the off-frequency transmissions from the nearby IREDs.

A third approach is to prevent the computer from seeing any received signals while it's transmitting. This is the approach used in Fig. 4. When pin 12 of the MAX232 goes low, indicating that the IRED will be transmitting, a 74HC-

4538 multivibrator, or one-shot, triggers. The one-shot's output at pin 6 is a high pulse of about 15 ms in duration, which is slightly longer than the transmission time of one byte at 1,200 baud. When pin 6 of the one-shot is high, pin 3 of the 74HC32 OR gate is high and pin 11 of the MAX232 is held high. The result is that any transmissions detected by the receiver are ignored.

Shortly after the IRED finishes transmitting, the one-shot times out. Pin 1 of the OR gate goes low and pin 11 of the MAX232 matches the receiver's output. Since the one-shot is retriggerable, its output remains high if the transmitter sends a series of bytes.

This method of transmitting assumes that each end will know when it's its turn to transmit. If both ends transmit at simultaneously, both receivers will turn off, and nothing, or at best partial transmissions, will be received.

If you're wiring the link directly to a single-board computer, you don't need the MAX232. You can instead connect to a device's 5-volt asynchronous serial input and output. However, since the MAX232 contains inverters, if you don't use it, you need to invert the serial input and output.

You can test the infrared link using two personal computers. Wire the MAX232's input, output and signal ground to the appropriate pins on a 25-pin D-connector or whatever connector your system uses.

At both ends, set up your communication software for matching baud rates of 1,200 or less. Configure the software for half-duplex communications. This means that when you transmit by typing at your keyboard, your software displays what you typed. In full-duplex communication, the display instead shows the echo received from a modem or far end. Since there's no echo when you transmit in one direction at a time, you have to provide your own.

If possible, perform initial tests over a short distance. Align the two ends so that each IRED points at the opposite end's receiver. When you have the basic link working, you can separate the two ends for the distance you desire. Test the link by typing messages at each end.

The messages should appear on the screens at the opposite ends. If you see nothing, use a logic probe or oscilloscope to follow the signal through the link. Type characters and observe the response. Pin 4 of the OR gate should toggle at 40 kHz. When you send a character, you should see toggling at pin 12 of the MAX232, pin 6 of the OR gate and the transistor's collector. At the receiver, you should see toggling at pin 2 of the OR gate and pin 11 of the MAX232.

If you're receiving completely garbled data, be sure your baud rates match. If you receive data that's mostly okay but exhibits occasional bad, missing or extra characters, there could be several things that can cause such a problem. Check the alignment of the IREDs and receivers. Of course, this is difficult to do, except by trial and error, since the transmissions are invisible. Over a distance of several feet, precise alignment shouldn't be necessary.

The one-shot's output should be at least 10 ms in duration. With an oscilloscope, you can set the pulse width precisely by substituting a 500,000-ohm potentiometer for the resistor at pin 2. Connect the center tap to pin 2, and one end to +5 volts. Briefly pulse low pin 5 low and adjust the potentiometer for a 10-ms pulse at pin 6. Alternatively, set the potentiometer to a low value and increase the resistance until you can transmit without also receiving the data at the transmitting end.

If you still have trouble, try using a lower baud rate. The receiver module introduces delays that change the width of the received pulses somewhat, and the delays are proportionately less with a slower baud rate's wider pulses.

You may still find that you occasionally receive spurious characters. This is because the infrared receiver responds to signals it detects in ambient light. A brief pulse may cause the computer to detect a transmission, often a space (20h) or other character, whose binary ASCII code contains mostly 0s. Shielding the receiver from ambient light can help eliminate these.

Remember that you can transmit in only one direction at a time. If you try to transmit before the other end's one-shot has timed out, the data won't be received.

You can use several methods to keep track of which unit's turn it is to transmit. One is to designate one computer the controller, which requests responses from the controlled end. The controlled end transmits only on request.

Many communication programs allow you to use the codes X-ON (Control-Q, or ASCII 11h) and X-OFF (Control-S, or ASCII 13h)) to control transmissions. X-ON means okay to send, and X-

Continued on page 47

"Do you have a faxing section?"

Sources

Digi-Key Corp.
701 Brooks Ave. S.
PO Box 677
Thief River Falls, MN 56701-0677
Tel.: 1-800-344-4539

CIRCLE NO. 120 ON FREE INFORMATION CARD

Edmund Scientific Co.
101 E. Gloucester Pike
Barrington, NJ 08007-1380
Tel.: 609-573-6250

CIRCLE NO. 121 ON FREE INFORMATION CARD

Selected Manufacturers of Infrared Emitters and Detectors

Gilway Technical Lamp
800 W. Cummings Pk.
Woburn, MA 01801
Tel.: 617-935-4442

CIRCLE NO. 122 ON FREE INFORMATION CARD

Harris Semiconductor
PO Box 883
Melbourne, FL 32901
Tel.: 407-724-7800

CIRCLE NO. 123 ON FREE INFORMATION CARD

Lite-On
720 S. Hillview Dr.
Milpitas, CA 95035
Tel.: 408-946-4873

CIRCLE NO. 124 ON FREE INFORMATION CARD

Motorola Semiconductor Products
P.O. Box 20912
Phoenix, AZ 85036
Tel.: 1-800-521-6274

CIRCLE NO. 125 ON FREE INFORMATION CARD

Optek Technology
1215 W. Crosby Rd.
Carrollton, TX 75006
Tel.: 214-323-2200

CIRCLE NO. 126 ON FREE INFORMATION CARD

Sharp Electronics
Microelectronics Group
5700 NW Pacific Rim Blvd., M/S 20
Camas, WA 98607
Tel.: 206-834-2500

CIRCLE NO. 127 ON FREE INFORMATION CARD

Siemens Components
Optoelectronics Div.
19000 Homestead Rd.
Cupertino, CA 95014
Tel.: 408-257-7910

CIRCLE NO. 128 ON FREE INFORMATION CARD

OFF means don't send. You could begin each message with X-OFF to advise the other end not to send anything and end with X-ON to indicate that your message is over and it's okay to transmit. Of course, if both sides by chance try to transmit at the same time, this won't work, since both receivers will turn off.

In addition to typing characters at each end, you should be able to use your communications software to send ASCII files back and forth. This includes files in Intel Hex and Motorola S-record format, which include checksums for verifying the transmissions.

If 1,200 bits per second isn't fast enough for you, see Sharp's Application Notes data book, which includes a description of a 19,200-bps link using Sharp's RY5AR01/AT01/BD01 transmitters and receivers.

Next month, I'll address the subject of radio links.

You can reach me on CompuServe at 71163,3555, on Internet at 71163.3555 @compuserve.com or by mail at Box 3374, Madison, WI 53704-0374. For a personal reply by mail, please include a self-addressed envelope, stamped envelope, if possible. For the first two parts of this series, you can order the October and November 1993 back issues directly from ComputerCraft *for $3.50 each.*

Jan Axelson