High-Speed Nondirective Optical Communication for Wireless Networks

John R. Barry
Joseph M. Kahn
Edward A. Lee
David G. Messerschmitt

Wireless Local Area Networks (LANs) offer a flexible and economical alternative to wired networks, since they avoid the high costs of maintaining and reconfiguring wired networks. There is a further, possibly more important, impetus for wireless LANs, and that is mobility. Portable computers may be a vital component of future office environments, and these computers should have access to all of the services now available to wired networks. Unlike desktop computers, portables have severe power-consumption, size, and weight limitations. The desire for an inexpensive, high-speed data link that meets these constraints has spurred the recent interest in optical wireless communication [1-10].

The first wireless LAN to use infrared radiation was developed over ten years ago [1]. Since then, numerous other infrared LANs have been demonstrated (see Table I). These systems seek to exploit the many advantages enjoyed by infrared over radio as a medium for wireless networks. For example, infrared has an abundance of bandwidth, free from Federal Communications Commission (FCC) regulation, infrared is immune to radio interference, and infrared components are small, inexpensive, and consume little power. Infrared and visible light, being near in wavelength, behave similarly; infrared is absorbed by dark objects, diffusively reflected by light objects, and directionally reflected by mirrors and shiny metals. Infrared light penetrates glass, but not walls, which is an important advantage over radio, because it allows neighboring cells, separated by a wall, to coexist without interference. It also makes infrared radiation a secure medium, confining data to the room in which it originates.

Infrared has disadvantages as well. An infrared link is susceptible to shadowing caused by objects or people positioned between the transmitter and receiver. Shadowing is not catastrophic for nondirective systems because light still reaches the receiver due to reflections from the surrounding environment. As an experiment, form a shadow on a page using your hand: even when your hand is just inches from the page, there is still enough light to read the words below. Similarly, a nondirective link should be robust in the presence of shadowing, although the system may have to operate at a reduced bit rate when the shadowing is severe.

Another disadvantage of infrared is its limited range, which is a result of power limits on infrared transmitters imposed by safety considerations and the fact that infrared light cannot penetrate walls. Thus, a high-speed wireless network using infrared will have small cells, about 5 m in radius. Since the connections between cells will use “wire” (or, more likely, optical fiber), the amount of wire used in an infrared wireless network may not be significantly less than that of a conventional wired network. The wired network of cell base stations that form the backbone of a wireless infrared network, however, will not change once the system is installed. Hence, terminals within such a system can be relocated or moved within the area of coverage without the need for rewiring; this is in stark contrast to conventional wired networks.

Despite the disadvantages of shadowing and small cell size, infrared still shows promise as a medium for wireless networks; in fact, its abundance of unregulated bandwidth appears to make it the only option for a cost-effective high-speed network operating at 100 Mbit/s or higher.

The first infrared LAN [1], which operated at a bit rate of 125 kbit/s, took advantage of the excellent diffusive reflectivity of most ceiling materials, such as plaster and echo tiles. By aiming the transmitters towards the ceiling, infrared radiation was made to fill a large coverage area. Each cell had its own base station, which was fixed on the ceiling and used an array of Light-Emitting Diodes (LEDs) as its optical source. Wired connections were used for communication between the base stations of different cells. Up-link signals received by each base station were broadcast through the wired backbone and re-broadcast from every other base station, so that, from the point of view of the portable terminal, the combination of the various cells appeared as a single shared channel. Up-link collisions were handled by a Carrier Sense Multiple Access Protocol with Collision Detection (CSMA/CD). The up-link and down-link transmissions used different subcarrier frequencies to avoid interference.

The primary advantage of using the diffusive approach is that there is no need for accurate alignment between transmitter and receiver. This is especially important in portable applications, in which readjusting a receiver antenna after every movement is not practical. The main drawback of the diffusive approach is the temporal dispersion caused by reflections from ceilings and walls, which effectively limits the rate of transmission.

Multipath dispersion can be greatly reduced by using direct optical beams, because very little power from a narrow optical beam will fall on an undesired reflecting surface, and the field of view of the receiver can be narrow. Hence, directed-beam systems can achieve high bit rates. For example, one system achieves a bit rate of 1 Mbit/s using 3° beamwidths [6]. Unfortunately, system performance is highly sensitive to pointing accuracy, making a directed-beam system unsuitable for portable computer applications.

This research was supported by the T. J. Watson Research Center of IBM, Hawthorne, NY.
<table>
<thead>
<tr>
<th>Date</th>
<th>Organization</th>
<th>Bit Rate</th>
<th>Directionality</th>
<th>Duplex</th>
<th>Multiple Access</th>
<th>Subcarrier Frequency</th>
<th>Modulation</th>
<th>Wave-length</th>
<th>Power</th>
<th>Area (cm²)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1979-1981</td>
<td>IBM Zurich</td>
<td>64 kbps</td>
<td>Diffuse</td>
<td>Subcarrier</td>
<td>CSMA/CD</td>
<td>0.20/0.40 GHz</td>
<td>Baseband</td>
<td>890 nm</td>
<td>-</td>
<td>-</td>
<td>[1][2]</td>
</tr>
<tr>
<td>1983</td>
<td>Fujitsu</td>
<td>19.2 kbps</td>
<td>LOS, Diffuse</td>
<td>Subcarrier</td>
<td>-</td>
<td>-</td>
<td>1/1.5 MHz</td>
<td>Subcarrier FSK</td>
<td>890 nm</td>
<td>15 mW</td>
<td>0.67 cm²</td>
</tr>
<tr>
<td>1985</td>
<td>Fujitsu</td>
<td>48 kbps</td>
<td>Narrow Up Wide Down</td>
<td>Subcarrier</td>
<td>-</td>
<td>-</td>
<td>1/1.5 MHz</td>
<td>Subcarrier FSK</td>
<td>-</td>
<td>-</td>
<td>[3]</td>
</tr>
<tr>
<td>1984</td>
<td>Hitachi</td>
<td>250 kbps</td>
<td>Narrow Up Wide Down</td>
<td>Subcarrier</td>
<td>CSMA/CD</td>
<td>10 MHz</td>
<td>-</td>
<td>Subcarrier FSK</td>
<td>-</td>
<td>300 mW</td>
<td>[4]</td>
</tr>
<tr>
<td>1985</td>
<td>HP Labs</td>
<td>1 Mbps</td>
<td>Narrow LOS</td>
<td>Wave-length</td>
<td>CSMA/CD</td>
<td>Baseband Manchester</td>
<td>600/880 nm</td>
<td>165 mW</td>
<td>7.6 mm²</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1986</td>
<td>Motorola</td>
<td>50 kbps</td>
<td>Widebeam LOS</td>
<td>TDMA</td>
<td>Baseband</td>
<td>RZ OOK</td>
<td>950 nm</td>
<td>16 mW</td>
<td>-</td>
<td>-</td>
<td>[5]</td>
</tr>
<tr>
<td>1987</td>
<td>Bell Labs</td>
<td>45 Mbps</td>
<td>Narrowbeam LOS</td>
<td>-</td>
<td>-</td>
<td>OOK</td>
<td>800 nm</td>
<td>1 mW</td>
<td>-</td>
<td>-</td>
<td>[6]</td>
</tr>
<tr>
<td>1988</td>
<td>Matsushita</td>
<td>9.2 kbps</td>
<td>Narrow Up Wide Down</td>
<td>Subcarrier</td>
<td>Spatial</td>
<td>500/950 MHz</td>
<td>880 nm</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>[7]</td>
</tr>
</tbody>
</table>

**NOTE:** Directionality refers to a system’s dependence on accurate alignment between transmitter and receiver; diffuse systems have no directionality, while Line-Of-Sight (LOS) narrowbeam systems are highly directional. Duplex Mux specifies how interference between uplink and downlink is avoided. Power refers to the total optical power of the base station, and Area refers to the total area of the receiver photodetectors. All of these systems use Light-Emitting Diode (LED) sources and Silicon (Si)-PIN photodiode receivers, except for [8], which uses laser-diode sources and avalanche photodiode receivers.

In Table I, we present the details of a number of infrared wireless LANs. Most of these systems use directed beams and operate at low to medium bit rates. In this article, we will examine the limitations on speed of a nondirected infrared network, and argue that speeds of 100 Mbps or higher are possible.

Next, we describe a prototyptic infrared system, and in the remainder of this article we discuss system impairments and design strategies. We will limit most of our discussion to the physical-layer issues involved in the design of a single high-speed optical link, and thus only briefly touch on many of the higher-level issues such as multiple access and retransmission protocols. (An excellent discussion of protocols and topologies for an infrared network can be found in [10].) Furthermore, we will concentrate on the downlink (from base station to portable), which is a more challenging problem technically than the uplink for two reasons. First, the complexity of the portable receiver is much more constrained than that of the base station, due to power consumption and cost restrictions. Second, the downlink data rates are likely to be much higher than the uplink rates, because downlink communication will include downloading large executable files and possibly video services, whereas uplink communication will be used for transferring working files and keyboard commands which tend to be much smaller.

**An Infrared LAN**

The transmission-speed capabilities of a wireless infrared LAN are highly dependent on the particulars of the network structure, such as the maximum distance between transmitter and receiver, transmitter power, and width of the optical beam. Therefore, before we discuss speed limitations of a nondirected optical LAN, we must specify precisely the type of network we have in mind.

**Network Description**

The system we envision, similar to that of [1], consists of a wired network of base stations, with one “cell” assigned to each base station (see Figure 1). To increase multiple access and routing efficiencies, we must abandon the simple strategy proposed in [1] of making each base station a repeater. Instead, as suggested in [10], the cellular nature of the network should be exploited. The network should be divided into a hierarchy of two layers, a network layer consisting of the wired network of base stations, and a link layer consisting of the infrared channel in each cell. Due to a tight power budget, the cell radius is limited to about 5 m. Thus, for typical office buildings, each single- or double-occupant office is a separate cell. The office walls prevent interference between cells. In an open office environment, which requires multiple cells per room, some means for

---

![Network Diagram](https://via.placeholder.com/150)

**Fig. 1. An infrared wireless LAN.**
the energy density of incident light onto the retina by factors of 100,000 or more. Therefore, the Maximum Permissible Exposure (MPE) levels are quite small. The cornea, the outer layer of the eye, filters out all wavelengths except those in the visible and near infrared range, roughly from 0.4 µm to 1.4 µm. Light outside this range is absorbed by the cornea and not passed through to the retina. Because of the energy magnification of light passing through the cornea, the MPE for wavelengths within the cornea passband is lower than those outside. This means that the transmitter power in a 1.5 µm system can safely be larger than that in a 850 nm system. However, as mentioned earlier, the availability of inexpensive GaAs lasers and low-noise Si APDs makes wavelengths near 850 nm the better choice.

Various governmental, nongovernmental, and standards organizations worldwide have specified MPEs for laser radiation [11]; they are nearly in consensus. The hazard of lasers is primarily due to their ability to concentrate a monochromatic beam of energy in a small, nearly point-sized image on the retina. However, the optical transmitters used in a nondirective LAN will not have narrow optical beams. Instead, they will be comprised of one or more lasers with diffusive lenses and reflectors to provide a wide and homogeneous optical beam. The lasers will be enclosed within a box in which the laser light undergoes diffusive reflection and only the diffuse radiation escapes. This eliminates the possibility of the emitted radiation being refocused with its original intensity onto the retina, since no combination of optical elements can restore the spatial coherence destroyed by a diffuser reflector.

The MPE for a laser is less than that for an LED; for example, at 900 nm the LED MPE is 10 mW/cm², while the laser MPE is 0.8 mW/cm² (for a viewing time of 8 hours) [12] [13]. The diffuse transmitter described above will be safer than a laser, but in the absence of applicable regulations we can use the laser MPE of 0.8 mW/cm². High-power transmitters will need to be large to meet this requirement. For example, a 1 W optical transmitter will require an aperture area of about 35 cm x 35 cm. The up-link from the portable to the base station will use lower power lasers of about 1 mW, which needs an aperture of only 1 or 2 cm² in area to meet the safety requirement.

### Intensity Modulation

Coherent optical modulation techniques, in which either the phase or the frequency of the roughly 3 x 10¹⁴ Hz optical

---

**Eye Safety**

The eye is more sensitive to infrared radiation than skin, and hence, safety considerations are dictated by eye safety. The human eye is an impressive optical system, capable of focusing
carrier is modulated directly, will not be considered because of the high complexity required to implement a coherent optical receiver [14]. Instead, we will limit consideration to the simpler, and thus more cost-effective, technique of intensity modulation.

Intensity modulation means that data is impressed on the intensity, not the frequency or phase, of the light wave. The intensity of a laser diode or LED can be modulated by varying its bias current. The modulating signal can be recovered at the receiver using a photodetector, which produces an electrical signal proportional to the intensity of the detected lightwave. The composite channel from laser input to photodetector output is a linear channel, suitable for any modulation technique. In Figure 2 we illustrate this idea. The input $u(t)$ modulates the light intensity, and therefore must be positive. The output of the receiver’s photodetector is $y(t)$, which will simply be $u(t)$ corrupted by multipath dispersion and noise. The equivalent baseband channel model, with $u(t)$ as the input and $y(t)$ as the output, is shown in Figure 2b. The model is called baseband because it is centered at dc rather than the actual carrier (about $3 \times 10^{14}$ Hz) of the optical signal. Nevertheless, the bandwidth of this linear channel, which is primarily limited only by the speed of the receiver electronics, can be quite large.

**Link Power Budget**

An important step in the design of a wireless infrared LAN is to determine the required signal irradiance (power per unit area) at the receiver. The result will tell us how much transmit power we need at the base station, as well as how large our cells can be. We can borrow from the extensive body of literature on fiber-optic receiver design to model all of the relevant noises, such as shot noise from the ambient light, thermal noise from the detector’s front-end circuitry, and photodetector dark current noise [15–19]. The signal irradiance at the receiver required to achieve a specific SNR depends on two key parameters: the total area of the receiver’s detector and the receiver bandwidth. Its dependence on bandwidth is due to the fact that receivers with larger bandwidths admit more noise, and hence, require more signal irradiance to maintain a specific SNR.

In Figure 3, we show some results from our link budget analysis, with the required irradiance at the receiver plotted versus detector area. The mean APD gain was chosen at each point on the APD curves optimally (minimizing the required signal irradiance). From Figure 3 we see that a receiver using APDs requires less signal irradiance than a receiver using PIN diodes. (In Photodetector Capacitance and f$^2$ Noise, we explain why the PIN curves flatten at large areas.) Simulation of optical propagation in a 5 m × 5 m room with a 3 m ceiling and a 1 W Lambertian source shows that, at desktop height (0.7 m), signal irradiance ranges from about 1.3 pW/cm² near the corner of the room to 7 μW/cm² in the center of the room [20]. For an APD-based receiver with a 100 MHz bandwidth, inspection of the 100 MHz APD curve in Figure 3 shows that a 1 W source provides adequate coverage for a room this size, provided the total detector area is more than 3 cm². Detectors of this size can be achieved using an array of small detectors. In fact, as we discuss later, the use of a photodetector array is desirable anyway, because it reduces the effects of amplifier noise and eases bandwidth constraints.

**Impairments**

The three major obstacles to achieving high-speed infrared communication indoors are noise from ambient light, high capacitance of large-area photodiodes, and multipath dispersion.

**Ambient Light**

A significant fraction of the light emitted from a lamp in the typical office environment is actually infrared, making ambient light a major source of noise in a wireless optical link. The current from a photodetector in the presence of ambient light with constant power is accurately modeled as a constant plus a zero-mean white Gaussian “shot” noise, with both the constant and noise power-spectral density proportional to the incident optical power. The constant term is not problematic, provided it is small enough to prevent saturation of the preamplifier, but the shot noise is often the dominant noise source at the receiver, because the ambient light power level is often high, typically 20 dB greater than the optical signal power. Of course, this does not imply that the electrical SNR is –20 dB, because the electrical signal power is proportional to the square of the optical signal power, whereas the electrical shot-noise power is only linearly proportional to the optical background power. A comparison of the shot-noise levels in various ambient-light environments is presented in [1], where it is shown that, of the three most common light sources, incandescent light (with tungsten filament) has the largest fraction of its power in the infrared region, followed by sun and fluorescent light.

It may seem fortunate that most office environments use fluorescent rather than incandescent lights, which emit more infrared power. However, unlike incandescent and sun light, the intensity emitted from a fluorescent bulb flickers on and off with the line frequency; this low-frequency intensity-modulation causes spectral lines in the resulting photodetector current at integer multiples of the line frequency (see Figure 4). Interference of these spectral lines, which extend to over 100 kHz, can be avoided by using a line code or modulating the data onto a subcarrier with frequency greater than a few hundred kHz. At 100 Mbps, any subcarrier modulation scheme would require subcarrier frequencies much higher than this, which makes fluorescent light interference a concern only for baseband systems.

Noise from ambient lights can be suppressed by inserting a narrowband optical filter before the photodetector, which ideally would transmit only light whose wavelength is near that of

\[ \text{A Lambertian source has a spherical intensity radiation pattern, so that, at an angle } \theta \text{ from the orientation of the source, the emitted power per solid angle is proportional to } \cos(\theta). \]
the laser source, and reject all other wavelengths. However, implementing such a filter with a wide field of view is difficult. This is because most narrowband optical filters are based on the principle of optical interference, and consist of a stack of thin dielectric slabs. The center wavelength of the passband filter is a strong function of the angle the incident light makes with the plane of the dielectrics; hence, they are highly directional, and ill-suited for nondirected communication. An intriguing, and costly, alternative would be to use nonplanar (e.g., hemispherical) dielectric shells to achieve a wide field-of-view, omnidirectional, narrowband filter.

**Photodetector Capacitance and \( f^2 \) Noise**

There are three primary sources of noise in the electrical signal that follows the photodetector of an indoor infrared system: shot noise \( (\sigma_s) \) from background light, thermal noise \( (\sigma_T) \) from the bias resistance of the detector preamplifiers, and so-called "amplifier" or \( f^2 \) noise \( (\sigma_{f^2}) \). There are other sources of noise, such as photodiode dark-current noise and shot noise from the received signal, but they are negligible in comparison. The total noise variance can be written as

\[
\sigma^2 = \sigma_s^2 + \sigma_T^2 + \sigma_{f^2}^2.
\]  

(1)

An unfortunate byproduct of communicating with nondirected optical beams is a small signal irradiance at the receiver. Therefore, to collect sufficient signal power, the total area of the receiver's photodetector must be large. Unfortunately, the capacitance of a photodetector is proportional to its area, and a large capacitance at the input to an amplifier acts as a low-pass filter, attenuating the high-frequency components of the received signal. This low-pass filter results in a noise penalty, because although the received signal is attenuated by it, any noise that follows the input stage, such as the channel thermal noise of a Field-Effect Transistor (FET)-based preamplifier, is not. When a white noise process following a first-order low-pass filter is referred back to the input of the filter, its power-spectral density becomes quadratic in frequency. For this reason, it is often called \( f^2 \) noise.

There is a way to counter this problem, however. Consider what happens when a single photodetector of area \( A \) is replaced by an array of \( n \) photodetectors, each with area \( A/n \), and each with its own transimpedance preamplifier. The total received signal power remains unchanged, as does the total shot noise. However, because the \( f^2 \) noise variance in each amplifier is proportional to the square of the capacitance—and thus, the square of the area—of each photodetector [18], the \( f^2 \) noise variance in each amplifier is \( \sigma_{f^2}^2/n \). Multiplying by \( n \), the number of amplifiers, yields a total variance of \( \sigma_{f^2}^2/n \). Hence, an array of \( n \) detectors causes a net reduction in \( f^2 \) noise by a factor of \( 1/n \). Of course, \( n \) cannot be made arbitrarily large, because at some point the \( n \)-fold increase of thermal noise, from the load resistance of each detector, will dominate. Symbolically, the total noise variance with an array of \( n \) detectors is

\[
\sigma_{f^2}^2 = n\sigma_s^2 + \sigma_T^2 + \sigma_{f^2}^2/n.
\]  

(2)

This implies the existence of an optimal number of photodiodes, which balances the decreased \( f^2 \) noise with the increased thermal noise: the number of photodiodes \( n \) that

---

**Fig. 5.** The simulated impulse response and corresponding amplitude response for a \( 5 \times 5 \) m room, accounting for up to five reflections [20].

**Fig. 6.** Power penalty due to multipath dispersion as a function of bit rate for unequaled baseband OOK.
minimizes Equation 2 is the nearest integer to $\frac{\delta_1}{\sigma_2}$.

Referring to Figure 3, we see that the 100 MHz PIN curve flattens at large areas, indicating the required signal irradiance for PIN-diode receivers does not decrease once the total detector area has exceeded some critical area. This saturation is because the $f^2$ noise variance, which is proportional to area squared, dominates all other noises at large areas. Therefore, at large areas, the area-squared dependence of signal power in the numerator of the SNR is canceled by the area-squared dependence of the $f^2$ noise variance in the denominator, resulting in an SNR independent of area.

This leads to another way of countering the $f^2$ noise resulting from high-capacitance photodiodes: using APDs rather than PIN diodes. The internal gain of the APD amplifies the power of the electrical signal by a factor of $M^2$, and that of the ambient-light shot noise by a factor of $M^2 P(M)$, where $M$ is the mean APD gain and the noise factor $F(M)$ is greater than 1 for all $M$. To illustrate the overall benefit of an APD, suppose $f^2$ noise is dominant for an APD-based receiver with mean gain $M = 1$. If we increase $M$ by an amount small enough that the $f^2$ noise still dominates, the overall SNR is increased, because the signal power has been increased by a factor $M^2$, but the dominating $f^2$ noise has remained unchanged. Eventually, as $M$ is increased further, the shot noise becomes comparable to the $f^2$ noise. Increasing $M$ even further will only reduce the SNR, because once the shot noise becomes dominant, the $M^2$-fold signal power increase is canceled by the $M^2 F(M)$-fold shot-noise power increase, with a net decrease in SNR by a factor of $1/F(M)$. It can be shown that there is an optimal mean APD gain that maximizes the SNR, and the shot noise at this optimal gain dominates all other noise sources [21].

Once the number of detectors, $n$, is chosen, it is advantageous with respect to noise for each detector to be as large as possible, because the increased shot noise and $f^2$ noise can always be countered by increasing the mean APD gain. Thus, in theory there is no optimal area. In practice, however, area is limited by cost, receiver complexity, bandwidth constraints, and the fact that the APD gain cannot be made arbitrarily large; it too has practical limits.

**Multipath Dispersion**

In an indoor environment, an optical signal in transit from transmitter to receiver undergoes temporal dispersion due to reflections from walls and other reflectors, this dispersion causes multipath fading. Diffuse systems are more prone to multipath effects than directed-beam systems, because their larger beamwidths mean more light hits potential reflectors, and the larger fields of view of their detectors mean more reflected light is detected. In Figure 5a, we show an impulse response for a typical single-occupant office, and in Figure 5b we show the corresponding amplitude response [20]. Viewed in the time domain, the optical channel spreads the transmitted signal, causing Inter-Symbol Interference (ISI). In the frequency domain, the multipath is seen to result in a significant ripple. The channel's amplitude response has a slow frequency roll-off, which is typical of multipath channels with a strong Line-Of-Sight (LOS) path. For example, although the $-3$ dB bandwidth is just 12 MHz, the attenuation at 500 MHz is only 5 dB. To mitigate ISI, the receiver must either signal at a symbol rate less than the $-3$ dB bandwidth or equalize the channel.

The impulse response depicted in Figure 5 was calculated using the technique described in [20]. This article also showed that the dominant source of ISI was from light undergoing two or more reflections. Although the light undergoing multiple reflections is highly attenuated, the increased relative delay of the higher order reflections makes them more troublesome than the lower order reflections. This is illustrated in Figure 5a, in which we see a strong LOS impulse response at time zero, followed by four distinct peaks corresponding to the first-order reflections from each of the four walls, and then a long tail lasting about 60 ns. The long tail, which is a result of light undergoing multiple reflections, causes most of the ISI.

Note that our link power budget analysis assumed that the LOS path was not blocked. In fact, this is required for a nondirected optical link to achieve a 100 Mbit/s bit rate. So unlike radio, which can undergo deep Rayleigh fades of 30 dB or more, the frequency-selective fading of an optical link with an LOS path will always be Ricean, and thus, less severe. When comparing radio multipath to optical multipath, one should also keep in mind that since radio waves pass through walls, reflections can come from far away. In contrast, the optical multipath is due only to reflectors within a room.

To illustrate the adverse effects of multipath when equalization is not performed, consider a simple on-off keyed system with an integrate-and-dump receiver. Suppose the signal at the input to the receiver is just the transmitted signal

![Channel capacity vs. mean APD gain, with and without multipath dispersion, for total detector areas of 5 and 10 cm².](image)

**Fig. 8.** Channel capacity vs. mean APD gain, with and without multipath dispersion, for total detector areas of 5 and 10 cm².
passed through the channel shown in Figure 5, plus additive white Gaussian noise. When we calculate the bit error rate for this system, we find that, when compared to an ideal dispersionless channel, more signal power is required to achieve some bit error rate, say $10^{-3}$. An alternative way of viewing this is that the multipath causes the eye diagram to close, so more transmitter power is required to open it up. In Figure 6, we plot this additional power requirement, or power penalty, versus bit rate. To isolate the penalty due to multipath alone, we assume perfect timing and an optimal decision threshold. The steep increase in penalty for bit rates greater than 12 Mbits—the channel bandwidth—justifies our previous heuristic for mitigating ISI, namely that the signaling rate be less than the $-3$ dB bandwidth. Higher bit rates are possible, at the expense of increased receiver complexity. For example, adaptive equalization can remove most, but not all, of the above ISI penalty. Modulation techniques that have lower symbol rates, such as mult-subcarrier modulation, are also less prone to multipath effects.

**Shannon Capacity**

Using basic principles of information theory, we will show that the nondirected indoor optical channel has a very large Shannon capacity; depending on the particular receiver structure, values of 1 Gbps or higher are possible. This result must be interpreted carefully, since Shannon capacity can never be attained in practice. Furthermore, the capacity results are in fact only estimates, since they are based on estimated system parameters and simulated, not measured, multipath dispersion. Nevertheless, the high capacity estimates do give credence to the feasibility of a low-cost 100 Mbps link.

Since we restrict attention to a single-wavelength system that uses intensity modulation with direct detection, the capacity results are not fundamental limits, i.e., other systems can have higher capacities. Coherent modulation with coherent detection could conceivably increase capacity. More realistically, one can in principle operate several intensity-modulation links simultaneously using $N$ different wavelengths (wavelength-division multiplexing), which will roughly increase the net capacity by a factor of $N$.

**Continuous-Time Capacity**

The mathematical model of the indoor channel we use is illustrated in Figure 7. The input to this system is the intensity of the transmitted optical signal, which is sent through the optical channel (like that in Figure 5). The first noise after the channel, representing the ambient-light shot and preamplifier thermal noise, is modeled as a white Gaussian process with power-spectral density $N_f = n_B A + n N_f$, where $n_B$ is the proportionality constant between the shot noise power-spectral density and area $A$, $n$ is the number of APDs, and $N_f$ is the power-spectral density of the thermal noise at the input to each preamplifier.

The APD gain is a non-Gaussian stationary random process with a skewed probability density function. Nevertheless, to facilitate analysis we will assume it is Gaussian. Numerical analysis reveals that sensitivity calculations using the Gaussian approximation are quite accurate; within 1 dB of the exact value [22]. The transfer function $H(f)$ is the low-pass filter formed by the photodetector capacitance and feedback resistance:

$$H(f) = \frac{g_m R_f}{1 + j2nfR_f cA/n},$$

where $g_m$ is the FET transconductance, $R_f$ is the feedback resistance of each transimpedance amplifier, $c$ is the capacitance per unit area of the photodetectors, $n$ is the number of detectors in the receiver array, and $A$ is the total photodetector area. The second noise source $n_f(t)$, which models the thermal noise in the FET channels, is white and Gaussian with noise density $N_f = n N_f$, where $N_f$ is the power-spectral density of the thermal FET-channel noise in each preamplifier.

This formulation allows us to calculate the theoretical capacity of an indoor infrared channel using Gallager's water-pouring technique [23]. If we first assume the optical channel is

**Fig. 10. Capacity versus symbol rate for baseband digital transmission.**

NOTE: The dashed line represents the capacity of the underlying continuous-time channel. Assuming ideal zero-excess-bandwidth transmit and receive filters, received signal irradiance is 2 mW/cm$^2$, background irradiance is 300 µW/cm$^2$, detector is array of 16 APDs, each with gain 50, total detector area is 8 cm$^2$, detector capacitance is 100 pF/cm$^2$, and each photodetector has its own transimpedance FET amplifier with a 1 kΩ feedback resistor.

---

50 • November 1991 - IEEE Network Magazine
dispersionless, so that the impulse response is a Dirac delta function, there is an analytical closed-form expression for capacity. If the input to our model $x(t)$ is constrained to have power less than $P$, so that

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x^2(t) \, dt \leq P,$$  \hspace{1cm} (4)

then the capacity can be shown to be [24]:

$$C = \frac{1}{\log 2} \left[ C_0 - C_1 \mu n^{-1} (C_0/C_1) \right],$$  \hspace{1cm} (5)

where

$$C_0 = 2 \left[ \frac{3P}{4b} \right]^{1/3},$$  \hspace{1cm} (6a)

$$C_1 = 2 \sqrt{a/b},$$  \hspace{1cm} (6b)

$$a = N_1/A^2 + N_2/(g_m R_n A)^2,$$

$$b = N_2 \left[ \frac{2n}{n g_m} \right]^2.$$  

Applying this result to a receiver with an array of 16 APDs totaling 5 cm$^2$ in area, mean APD gain of 50, capacitance constant of $c = 100$ pF/cm$^2$, feedback resistance of $R_f = 1$ kΩ, received signal power of 10 μW, and background light power of 1.5 mW, the channel capacity is 3.6 Gbps [24].

Accounting for multipath dispersion makes the capacity calculation more difficult. Our approach is to use a numerical water-pouring technique that incorporates the channel frequency response. The resulting capacity is lower than that of Equation 5; for example, using the frequency response shown in Figure 5, the above capacity of 3.6 Gbps reduces to 1.6 Gbps. Consider Figure 8, in which we illustrate the effects of multipath dispersion, photodetector area, and APD gain. We assume that the receiver is an array of 16 APDs, the capacitance constant is $c = 100$ pF/cm$^2$, the feedback resistance of each transimpedance amplifier is $R_f = 1$ kΩ, the received signal irradiance is 2 μW/cm$^2$, and the background light irradiance is 300 μW/cm$^2$. This analysis shows that the capacity of the infrared channel can be increased dramatically by using large-area APDs. Furthermore, it shows that there is an optimal APD gain—which can be 100 or higher, depending on the parameters—that maximizes capacity.

The capacity results also provide insight into the tradeoffs associated with choosing the total area $A$ of the photodetector array. In Figure 8, we observe that a larger area corresponds to a greater channel capacity. In fact, this generalizes so that larger areas always yield higher channel capacities. For the special case of no multipath dispersion, this fact is evident from Equation 6; as $A$ gets large, $C_1 \to 0$; while $C_0$ remains unchanged. Thus, capacity increases monotonically as area increases, and $C \to C_0 \mu n^{-1}$ as $A \to \infty$. This is why large areas mean that a larger fraction of the signal spectrum extends past the −3 dB bandwidth, resulting in more ISI. Thus, although increasing area yields a larger channel capacity, more sophisticated modulation or equalization techniques are required to achieve it.

**Discrete-Time Capacity**

To achieve the continuous-time capacity discussed above, the transmitted signal $x(t)$ must be a zero-mean Gaussian random process with power-spectral density shaped according to a water-pouring procedure. Obviously, this is not very practical. However, we can take one step towards practicality by considering the capacity of a discrete-time transmission system, in which a symbol is transmitted every $T$ seconds. This is how real systems are implemented. It can be shown that the capacity of the discrete-time system is equal to the capacity of the underlying continuous-time channel, provided that the symbol rate $1/T$ is large enough, but the required symbol rate may be prohibitively large, a Gsymbol/s or higher. We now ask the question: what is the capacity of the discrete-time system when smaller, more realistic, symbol rates are used? To answer it, we can use discrete-time water-pouring for the equivalent discrete-time channel of the system illustrated in Figure 9. There, the input $x_k$ is a discrete-time sequence with symbol rate $1/T$. It is fed through a transmit filter before transmission. At the receiver, the output of the photodetector low-pass filter is passed through a receive filter and then sampled at a rate of $1/T$, yielding the discrete-time sequence $y_k$. In Figure 10, we show how the capacity of this discrete-time system varies as a function of the symbol rate.

The continuous-time capacity for this system is a little more than 2 Gbps, and is shown in Figure 10 as a dashed line. We see that the capacity of the discrete-time system is a strong function of symbol rate. For example, if we insist on signaling at a rate of 100 Msymbol/s, from Figure 10 we see that channel capacity is only 350 Mbps. At 200 Msymbol/s, the capacity is 300 Mbps. At 400 Msymbol/s, the symbol rate must be 520 Msymbol/s. To use this capacity curve to determine the required symbol rate for a practical 100 Mbps, we must first quantify what fraction of capacity we hope to achieve. This is a difficult question because it depends on how much complexity
is acceptable. Suppose we measure complexity by the "algorithmic" complexity multiplied by electronics speed, which gives a rough measure of the die area required for implementation, and thus, accounts for cost and power consumption. If complexity is fixed, then it is likely that the fraction of capacity achievable is a function of symbol rate. For a given complexity, systems with lower symbol rates, which use slower electronics, can have higher algorithmic complexity. The tradeoff of increased capacity versus less-sophisticated signal processing is a subtle one and warrants further investigation.

There are more definitive conclusions we can draw from the capacity estimates in Figure 10, however. For example, the capacity estimate at 20 Msymbol/s is 100 Mb/s. This indicates that there is any hope of attaining reliable communication at 100 Mb/s using baseband transmission, the symbol rate must be at least 20 Msymbol/s. For instance, this rules out a 100 Mb/s 64-level pulse-amplitude modulation system, which has a symbol rate of only 16.7 Msymbol/s.

Modulation Techniques

Here, we discuss two modulation schemes suitable for an indoor infrared link. The first is baseband On-Off Keying (OOK), which is a binary signaling scheme and hence, its symbol rate is equal to its bit rate. From Figure 10, we see that, at a symbol rate of 100 Msymbol/s, the capacity is 350 Mb/s. This means that a 100 Mb/s system using OOK would need to operate at 29% capacity, which may be difficult because it would require the implementation of good codes at high speeds. We will therefore consider a second approach, multisubcarrier modulation, in which baseband signaling is foregone altogether and instead multiple symbol streams are modulated onto different subcarrier frequencies.

OOK

The first modulation technique that comes to mind is OOK, in which a pulse of light is sent for a "one" bit and no light is sent for a "zero" bit. Its simplicity makes it the easiest scheme to implement. OOK has a few drawbacks. First, it requires large symbol rates and hence is susceptible to ISI. To mitigate ISI, some type of equalizer is necessary. In particular, a Decision-Feedback Equalizer (DFE) is simple and effective, and hence, is the most likely candidate.

A DFE feeds past decisions through a feedback filter and subtracts the result from the input to the decision device; this nonlinear processing results in less noise enhancement than linear equalization. A DFE is also easier to implement than a linear equalizer, because the input to its feedback filter are symbols, 1 or 0 for OOK. Hence, convolution can be implemented with additions only, no multiplications. Also, implementation of a DFE for impulse responses of the type shown in Figure 5, strong LOS contribution with no leading-edge component, is further simplified because no forward filter is needed. In Figure 11, we show the power penalty of a receiver with a zero-forcing DFE relative to the matched filter-bound, plotted as a function of symbol rate. The multipath dispersion is modeled using the impulse response of Figure 5; other system parameters are listed in the caption. This power penalty is defined by

\[
P_{\text{penalty}} = \frac{P_{\text{DFE}}}{P_{\text{MF}}} \tag{7}
\]

where \(P_{\text{DFE}}\) is the mean-squared error at the slicer input given an ideal zero-forcing DFE, and \(P_{\text{MF}}\) is the mean-squared error at the slicer input given an ideal single-shot matched filter receiver [25]. We see that at 100 Mb/s, the penalty is only 2 dB, about 0.7 dB of which is due to ISI from the low-pass filter created by the feedback resistance and detector capacitance; the rest is from multipath dispersion. This small power penalty indicates that, at this symbol rate, the ISI has little effect on the performance of an ideal zero-forcing DFE. However, as shown previously, restricting the symbol rate to 100 Msymbol/s limits the channel capacity to 350 Mb/s. Furthermore, the tail cancellation performed by a DFE is not information-preserving and hence, it causes an additional, albeit slight, reduction in capacity. Hence, reliable communication at 100 Mb/s using OOK and a DFE may be difficult. To see this, first note that a capacity of 350 Mb/s at a symbol rate of 100 Msymbol/s corresponds to a spectral efficiency of 3.5 bit/symbol. Neglecting the capacity reduction due to the DFE, we can use the Shannon-Hartley law

\[
C = \frac{1}{2} \log_2 (1 + SNR) \tag{8}
\]

to find that \(SNR = 2^C - 1\), or 21 dB, when \(C = 3.5\) bit/symbol. This is only 5.4 dB higher than the SNR required by an ideal OOK system to achieve a bit-error probability of 10^-3. The margin might not be large enough once we account for the ISI power penalty and other inescapable nonidealities, such as imperfect modulation and timing recovery. One possibility for increasing the margin would be to use trellis or convolutional coding with a soft decoder. However, at these symbol rates, the cost of the Viterbi decoder may be prohibitive. A cheaper alternative that offers less margin is error correcting codes, but the digital decoding hardware may still be too costly. Nonetheless, these options warrant further investigation.

Because of the potential difficulties in achieving high speeds using OOK discussed above, we also consider multisubcarrier modulation, in which a larger fraction of the water-pouring bandwidth is used without the need for equalization.

Multisubcarrier Modulation

Optical multisubcarrier modulation, which is gaining popularity in video distribution applications [26], is a promising modulation technique since it can achieve bandwidth efficiency as well as counter the effects of multipath. In a multisubcarrier system, the large water-pouring bandwidth discussed previously is subdivided into a number of subbands, and separate reduced-rate communication links are established across each subband. A raw data stream is divided into N interleaved bit streams. Each substream then modulates its own carrier, and the intensity of the transmitted lightwave is modulated by the sum of the modulated subcarriers. Each substream uses only 1/N of the total bandwidth, and hence, is less susceptible to ISI. The fact that the per-carrier capacity can be run at slower speeds is a further advantage of this approach. The number of carriers, N, should be chosen so that the ISI penalty in each substream is negligible. An additional advantage is gained from the ability of the transmitter to shape the composite signal spectrum according to the channel frequency response and noise spectrum: the number of bits and power per carrier can be adjusted to the SNR expected at each carrier frequency [27], in much the same way as the water-pouring procedure matches a signal spectrum to a channel. A multisubcarrier system will be impacted by any nonlinearities in the optical channel, but we do not believe this problem to be serious due to the ruggedness of the data signal and the relatively small number of carriers that would be required. The effects of nonlinearities can be reduced by using multiple optical sources, modulating only one substream on each source; in fact, it is likely that multiple sources will be required anyway, to meet the high power requirements.

A disadvantage of subcarrier modulation is its inherent power penalty for peak-power-limited transmitters when com-
pared to baseband OOK. For example, a Binary Phase-Shift-Keying (BPSK) single-subcarrier modulation system with 100% modulation suffers a 3 dB power penalty with respect to baseband OOK. Nevertheless, subcarrier modulation is still a promising method because of its advantages cited above. A further advantage of subcarrier modulation is that it moves the signal spectrum away from dc, avoiding the low-frequency harmonics of fluorescent light interference.

**Other Alternatives**

So far we have discussed baseband OOK with a DFE and multisubcarrier modulation as candidate modulation schemes. There are other possibilities, but most can be ruled out because of poor performance or high cost. We now briefly discuss some alternative modulation schemes.

- **Multilevel signaling**—One way of reducing the symbol rate is to use multilevel or M-ary signaling, in which each symbol represents $\log_2 M > 1$ b. For example, a 100 Mbaud data stream with a 16-Differential Phase-Shift-Keying (DPSK) modulation scheme results in a symbol rate of 25 MHz. Unfortunately, multilevel signaling requires a high SNR for proper performance, and is not appropriate for our noisy, power-limited application. Besides, the discrete-time capacity results of Figure 10 indicate that higher, not lower, symbol rates are needed to approach continuous-time capacity.

- **Multimodal modulation**—Multimodal modulation is very similar to the multisubcarrier scheme discussed earlier, except that each data stream modulates a different wavelength (from a different optical source) rather than a different subcarrier. This approach is appealing, but would require an expensive bank of narrowband optical filters at the receiver.

- **Spread-spectrum modulation**—Another way of combating multipath fading is through spread spectrum, in which the transmitted information is spread over a large bandwidth, so that the effects of frequency-selective fading are minimal.

In addition, spread spectrum offers the possibility of code division multiple access. However, the bandwidth of a 100 Mbaud signal is large to begin with, so the speed of the receiver electronics limits any further spreading.

- **Pulse-Position Modulation (PPM)**—In its simplest form, a PPM system with an alphabet size of $M$ will break up the symbol period into $M$ subintervals, and the transmitted signal representing the $k$th symbol in the alphabet is "on" during the $k$th subinterval and "off" everywhere else. Thus, just like multilevel signaling, the symbol rate can be made less than the bit rate. Unlike multilevel signaling, however, the bandwidth in PPM is increased by a factor of $M/\log M$, relative to OOK [28]. The extra noise passed by this larger bandwidth, particularly in the presence of $f^2$ noise, makes PPM a poor choice.

Overall, we view OOK with a DFE and multisubcarrier modulation as the most promising option. The final choice will not be made until after the implementation cost, power consumption, and performance of each are carefully compared.

**Summary and Conclusions**

We have reviewed the three primary obstacles to high-speed communication, namely shot noise from ambient light, high-capacitance photodiodes, and multipath dispersion. We have also presented ways to counter these problems. For example, shot noise from ambient light can be reduced by using a narrowband optical interference filter, and the low-frequency noise from fluorescent lights can be avoided with the use of a line code or a subcarrier to move the signal spectrum away from dc. The problems resulting from high-capacitance photodiodes, such as $f^2$ noise and restricted bandwidth, can be limited by a careful receiver design that incorporates an array of APD-based transimpedance amplifiers. The effects of multipath dispersion, which cause a power penalty if left uncorrected, can be reduced by using a DFE or choosing an appropriate modulation technique.

Our capacity analysis indicates that a truly wideband wireless nondirective optical network is theoretically possible. Of course, it can be difficult to get close to capacity in practice. Nevertheless, a Shannon capacity of greater than 1 Gbaud indicates that 100 Mbaud should be practical, which has exciting implications on future computer environments.

There still remains much work to be done before low-cost and high-speed infrared wireless LANs become practical. In optical devices, two achievements would hasten their arrival: the design of large-area, low-noise, high-capacitance APDs for wavelengths near 850 nm, and the design of narrowband, omnidirectional bandpass filters. There are challenges to be met in electronic circuits as well, such as the design of a large-bandwidth, low-noise amplifier array for the detector frontend, and the design of a low-cost, low-power-consumption, high-speed DFE. More analysis is also required to determine the best modulation scheme. Simulation will be especially valuable in this regard because of the difficulties in analyzing many system impairments such as fluorescent light interference, nonlinear modulation characteristics, and finite-precision and finite-length DFE filters. Finally, additional experimental work is also required to better characterize multipath dispersion and ambient light effects in typical office environments.

**References**

Biography

John R. Barry received his B.S. degree summa cum laude with departmental honors from the State University of New York at Buffalo in June 1986, and his M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley (UC Berkeley), in December 1987, both in electrical engineering. Since 1988, he has held a research assistantship at the Electronics Research Laboratory of UC Berkeley, where he is presently a Ph.D. candidate. His research interests include digital communication theory, fiber and free-space optical communications.

Joseph M. Kahn received his A.B., M.A., and Ph.D. degrees in physics in 1981, 1983, and 1986, respectively, from UC Berkeley. His Ph.D. thesis was entitled "Hydrogen-Related Acceptor Complexes in Germanium." He is Assistant Professor in the Department of Electrical Engineering and Computer Sciences at UC Berkeley. From 1987 to 1990, he was a Member of Technical Staff in the Lightwave Communications Research Department of AT&T Bell Laboratories in Holmdel, NJ, where he performed research on multi-gigabit-per-second coherent optical fiber transmission systems and related device and subsystem technologies. He demonstrated the first BPSK-homodyne optical fiber transmission system, and achieved world records for receiver sensitivity in multi-gigabit-per-second systems. He joined the faculty of UC Berkeley in 1990, where his research interests include optical fiber communication networks and transmission systems, indoor LANs using free-space optical communication, and optical interconnects in digital systems. Mr. Kahn is a recipient of the National Science Foundation (NSF) Presidential Young Investigator award, and is a member of the IEEE.

Edward A. Lee received his B.S. degree from Yale University in 1979, his S.M. degree from the Massachusetts Institute of Technology (MIT) in 1981, and his Ph.D. degree from UC Berkeley in 1986. He is an associate professor in the Electrical Engineering and Computer Science Department at UC Berkeley. His research activities include parallel computation, architecture and software techniques for programmable DSPs, design environments for development of real-time software, and digital communication. He was a recipient of a 1987 NSF Presidential Young Investigator award, an IBM faculty development award, the 1986 Sackrison prize at UC Berkeley for the best thesis in electrical engineering, and a paper award from the IEEE Signal Processing Society. He is coauthor of Digital Communication, with D.G. Messerschmitt, Kluwer Academic Press, 1988, and Digital Signal Processing Experiments with Alan Kamas, Prentice Hall, 1989, as well as numerous technical papers. From 1979 to 1982, he was a member of technical staff at Bell Telephone Laboratories in Holmdel, NJ, in the Advanced Data Communications Laboratory, where he did extensive work with early programmable DSPs, and exploratory work in voiceband data modem techniques and simultaneous voice and data transmission.

David G. Messerschmitt received his B.S. degree from the University of Colorado in 1987, and his M.S. and Ph.D. degrees from the University of Michigan in 1988 and 1989, respectively. He is a Professor of Electrical Engineering and Computer Sciences at UC Berkeley. Prior to 1977, he was at Bell Laboratories in Holmdel, NJ. His current research interests include applications of digital signal processing, digital communications (subscriber loop, fiber optics, and VLSI and digital systems), architectural approaches to dedicated hardware digital signal processing (especially video compression application), video applications of broadband packet networks, and computer-aided design of communications and signal processing systems. He has published over 110 papers, is coauthor of two books, and has 10 patents. He also serves as a consultant to a number of companies. He is a Fellow of the IEEE and a Member of the National Academy of Engineering of the United States. He has served as Editor for Transmission of the Transactions on Communications, and as a member of the Board of Governors of the Communications Society. He has also organized and participated in a number of short courses and seminars devoted to continuing engineering education.

54 • November 1991 - IEEE Network Magazine