High performance receivers for optical communications

The principles underlying optical detection are reviewed and the practical realisation of devices and receivers for digital optical communications is discussed. Some emphasis is placed on the advanced receiver signal processing techniques—both electronic and optoelectronic—that are now coming forward to provide a performance approaching the fundamental limits imposed by quantum mechanical considerations.

by J. J. O'Reilly

1 Introduction

An optical receiver provides optoelectronic conversion together with signal conditioning and processing to facilitate the recovery of an information-bearing message. Present-day practical receivers rely on direct detection; for this we view the incident optical signal as a stream of photons and detection involves photon to electrical-carrier conversion without reference to the phase of the optical signal. A quantum-particulate model for the optical signal is thus appropriate. A direct-detection receiver may be preceded by an optical network to combine the incoming optical signal with the output from an optical local oscillator in such a way that coherent mixing takes place at the surface of the photodetector. This can provide conversion gain enabling improved receiver sensitivity to be obtained. Coherent optical communications is the subject of much current research, not only because of the receiver sensitivity improvements it offers but also because of the novel network architectures it makes possible.[1,2] So far as the receivers themselves are concerned, though, the central requirement is for direct detection as close to the quantum limit as possible. To facilitate this a variety of novel optoelectronic techniques are currently being investigated, including the use of avalanche photodiodes (APDs) incorporating quantum well and superlattice structures[3,4] and the use of semiconductor laser amplifiers to provide optical pre-amplification for a conventional direct detection optical receiver.[5,6]

Accordingly, in this paper we shall review briefly the principles underlying optical detection and examine the practical realisation of devices and basic receiver structures. Consideration will be given to the advanced receiver signal processing techniques—both electronic and optoelectronic—that are now coming forward as a basis for improved performance approaching quantum limited operation.[7,8]

2 The direct-detection process

Direct detection may be viewed as a simple mapping from photons to conduction electrons. Let P be the incident optical-power level and ν be the optical frequency, related to the free-space wavelength λ by $c=\nu\lambda$ with c the velocity of light in vacuum. If h is Planck's constant then the energy per photon is $h\nu=hc/\lambda$ joules and the mean photon arrival rate is given by

$$\frac{P}{h\nu} = \frac{P\lambda}{hc} \text{ photons/s} \quad (1)$$

Each detected photon produces just one external electrical carrier but not all photons are detected. The carrier generation rate is thus lower than the photon arrival rate as determined by the detection probability, known as the quantum efficiency η. The mean electrical carrier generation rate is thus

$$\eta \frac{P\lambda}{hc} \text{ carriers/s} \quad (2)$$

Each electrical carrier transports a unit of electronic charge of q coulombs and so the mean rate of flow of charge, that is the mean photocurrent I, is given by

$$I=\eta \frac{q\lambda}{hc} P \text{ coulombs/s } (\equiv \text{amperes}) \quad (3)$$

The ratio I/P A/W is known as the responsivity R of the detector. Since $\eta \leq 1$ the responsivity is upper-bounded by

$$R \leq \frac{q\lambda}{hc} \text{ A/W} \quad (4)$$

which is proportional to wavelength. This latter feature simply reflects the fact that at longer wavelengths, corresponding to lower optical frequencies, the energy per photon is less and we thus have a higher photon rate for a given optical-power level. Current optical-fibre systems operate in the wavelength range of 850nm to 1550nm and modern detectors can have quantum efficiencies approaching unity. As a consequence, responsivities in the region of 0·5 to 1·0 A/W are typical, depending on the wavelength of operation.

Although the bound of expr. 4 suggests that responsivity increases with λ, such a conclusion is sensitive to the wavelength dependence of η. There comes a point beyond which the photon energy, which decreases with increasing wavelength, is insufficient to provide carrier generation and the quantum

efficiency and responsivity fall essentially to zero. The idealised spectral response thus takes the form shown in Fig. 1*a*, where λ_c denotes the long-wavelength cutoff. The spectral response for various detector materials quantified in terms of the absorption coefficient α—the inverse of the mean photon penetration depth—is shown in Fig. 1*b*.

3 Detection devices

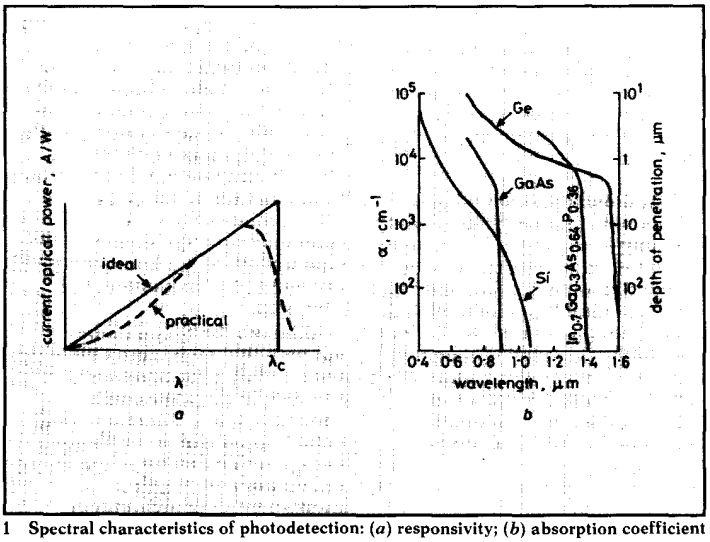
To appreciate more fully the nature of this long-wavelength cutoff, and the other detection limitations, it is appropriate to consider specific detection devices. We shall consider here only semiconductor detectors exploiting photoconduction and first examine briefly this process.

Photoconduction

Consider a sample of intrinsic material with energy bandgap E_g. Provided that $E_g >> k\theta$, where k is Boltzmann's constant and θ is the absolute temperature, the density of free conduction electrons in the conduction band and of holes in the valence band will be negligible. Hence a bias voltage applied to the sample will give rise to negligible current flow. If the sample is now irradiated with light with photon energy $E_p = h\nu > E_g$, photon absorption will occur and electrons will be promoted from the valence band to the conduction band. For each photon absorbed an electron–hole pair is created and these charge carriers can migrate (drift) under the influence of the applied field. Irradiation of the material thus increases its conductivity and we detect this by way of current flow in the external circuit. In this simple picture, if $h\nu < E_g$ then photons cannot be absorbed and we are beyond the long-wavelength cutoff. Since we require $h\nu > E_g$ for photon absorption, corresponding to $hc/\lambda > E_g$, the long-wavelength cutoff is $\lambda_c = hc/E_g$.

We do not, in fact, encounter an abrupt transition from 'photons absorbed' to 'no photons absorbed', especially at room temperature, where the thermal energy must be considered. The more gradual nature of the cutoff encountered in practice relates to the variation of the absorption coefficient with wavelength shown in Fig. 1*b*. To some extent we can enhance the long-wavelength performance by ensuring that there is an adequately long potential absorption path for the photons. Nevertheless, the drop in absorption coefficient in the vicinity of the bandgap is such as to impose severe limitations on the degree of long-wavelength enhancement attainable. A consequence, discernible from Fig. 1*b*, is that silicon detectors cannot be used for wavelengths greater than about 1 μm whereas germanium is useful to beyond 1·6 μm.

Simple photoconductive detectors have not generally been used for optical fibre communications due to various limitations, such as poor speed of response: once the conductivity has been increased by absorption of a pulse of light it remains enhanced for a time determined by the carrier lifetime. This problem is circumvented by the use of semiconductor junction photodetectors.



1 Spectral characteristics of photodetection: (*a*) responsivity; (*b*) absorption coefficient

The pn photodiode

A simple *pn* photodiode is illustrated in Fig. 2*a*. Since we wish to use the device in its photoconductive mode, reverse bias is applied so that a depletion region which is essentially devoid of free carriers is formed in the vicinity of the junction. This depletion region forms an insulating layer such that negligible current flows. The applied voltage appears largely across this narrow high-impedance depletion region and gives rise to a high electric field (*E*-field). If the device is now irradiated and photons are absorbed within the depletion region, hole–electron pairs are created. These carriers drift under the influence of the *E*-field towards their respective 'majority carrier' sides of the junction, the *n*-region for electrons and the *p*-region for holes. This carrier movement gives rise to a displacement current in the external circuit.

The speed of response of the simple *pn* junction photodiode is limited by three factors:

- Transit time: the time taken for carriers entering or produced in the depletion region to cross to their respective majority carrier sides.
- Capacitance: the depletion region is a thin insulating layer whereas the undepleted *n* and *p* regions on either side have relatively high conductivity, the structure of a parallel-plate capacitor. The junction capacitance C taken together with the bulk resistance of the diode and external circuit resistance introduces an RC time constant which imposes a detector bandwidth limitation given by

$$B_D = \frac{1}{2\pi CR} \qquad (5)$$

  For a large detector bandwidth, which corresponds to a short time constant, C and/or R must be small.
- Deep photon penetration: since the depletion region is rather thin, photons absorbed in the undepleted material on either side are likely to be significant. The *E*-field in the undepleted, low-resistivity regions is slight and the dominant mechanism for carrier transport is thus the rather slow process of diffusion. Considering photon absorption in undepleted *n*-type material the hole first diffuses slowly towards the junction. On entering the

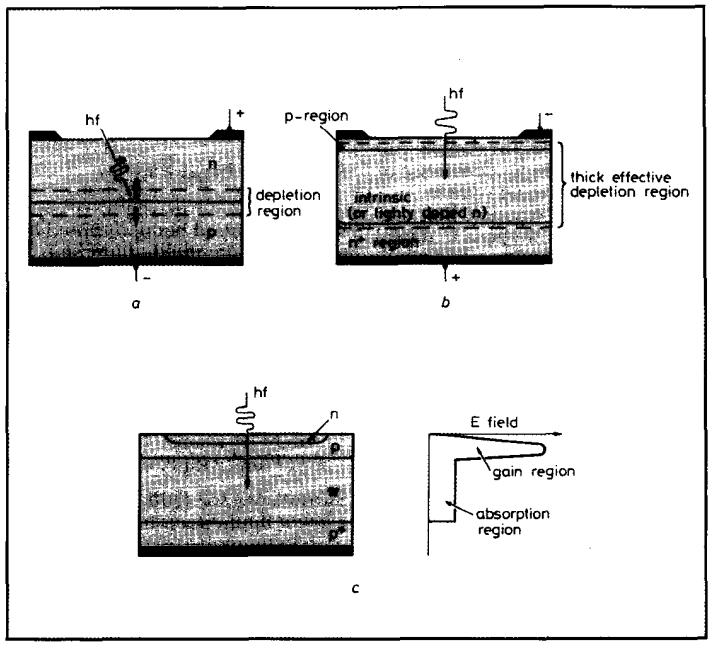
depletion region it comes under the influence of the *E*-field where the drift transport mechanism takes over and it is swept rapidly across to the *p*-type region, where holes are the majority carriers. Hence for a high-speed device we must ensure that photon absorption outside the drift region is negligible.

The pin photodiode

A solution to the speed of response problem is obtained by incorporating a relatively thick 'intrinsic' layer between the *n* and *p* regions to yield a *pin* structure as illustrated in Fig. 2*b*. The intrinsic material is essentially devoid of free carriers and can be thought of as already depleted. In practice the region may be very lightly doped *n*-type with the doping so low that it is considered to correspond to intrinsic material. The application of a reverse bias thus gives rise to a wide 'effective' depletion region comprising the *i*-layer and the depletion regions in the *n* and *p* material on either side. With this much wider layer as the 'dielectric' the reverse-biased device capacitance is much reduced. Similarly, the probability of photon absorption in the high *E*-field region is enhanced since this now includes the relatively thick *i*-region. The only penalty as far as speed is concerned is an increase in transit time, arising because carriers must drift through the *i*-layer as well as the depleted *n* and *p* regions. However, this is by far the least limiting of the factors discussed and we have thus achieved a very desirable tradeoff. Also, in practice the *n*-region is generally very heavily doped to facilitate the production of an ohmic contact, so that any carriers generated within undepleted *n*-type material are likely to recombine before they can enter the high *E*-field region. This gives the device designer a certain degree of freedom: he can choose a long absorption region—essentially the *i*-layer—to maximise quantum efficiency, or he may trade this off in favour of a shorter transit time to improve the speed of response by adopting a shorter absorption region.

The avalanche photodiode

We have already seen that typical responsivities are in the region of 0·5 to 1·0 A/W and we note here that high-performance optical receivers operate at optical-power levels measured in nanowatts. As a result, the primary photocurrent—before any amplification—may be



2 Schematic photodiode structures: (*a*) *pn* photodiode; (*b*) *pin* photodiode; (*c*) 'reachthrough' APD

in the region of nanoamperes. It is all too easy for these very-low-level signal currents to be swamped by noise introduced by the postdetection circuitry. One solution is to provide current gain *within* the detection device and this is the essence of the avalanche photodiode (APD).

Consider the APD structure of Fig. 2*c*. Under reverse bias a spatially concentrated high *E*-field region is produced in the vicinity of the *pn* junction whereas a more modest *E*-field extends throughout the *i*-region to the heavily doped p^+ contact. Incident photons absorbed in the *i*-region generate hole–electron pairs which drift under the influence of the *E*-field, the electrons (in this example) moving towards the high *E*-field region. On entering this region the electrons are rapidly accelerated such that secondary carriers can be produced by impact ionisation. These secondary carriers are also accelerated and can themselves give rise to additional carriers by further ionisations. Consequently, a single photo-induced electron–hole pair can give rise to many secondary pairs and cause a corresponding number of electrons to flow in the external circuitry. The device is said to exhibit avalanche gain, the gain being dependent on the applied voltage.

The increase in effective signal current provided by avalanche gain can be useful in bringing the signal above additive noise, but it is no panacea. The number of secondary pairs produced from a single primary pair is a random number. The mean number corresponds to the avalanche gain, but the variability in the multiplication factor from one detection event to another introduces a new source of noise associated with the avalanche process itself. Hence it does not automatically follow that receivers based on APDs will offer superior performance compared with receivers based on *pin* detectors. Much depends on the relative significance of avalanche noise and other noises added by the receiver circuitry. Accordingly we shall now turn our attention to the question of receiver signal and noise characterisation.

4 Noise in optical receivers

Noise encountered in optical receivers arises from a variety of sources. There is thermal noise due to resistive elements, noise due to the active devices in the postdetection amplifier and shot noise associated with the detector dark current. These are the *additive* noise terms. In addition we encounter signal-dependent noise of quantum origin. In direct detection the particulate character of the optical signal gives rise to shot noise (shot noise associated with the detected signal

photocurrent) and the signal, as mentioned previously, may be further contaminated by avalanche gain noise if an APD is employed. In the limit, even if all *additive* noise sources could be eliminated and if we were to admit no avalanche gain noise, the signal-dependent noise of quantum origin would remain. Consequently we examine first the quantum noise limitations to digital optical communications.

Quantum considerations in direct optical detection

The photon detection events correspond to a nonhomogeneous Poisson process such that the probability distribution for the number of events in a given time interval has the form

$$P(n,\mu) = \frac{\mu^n e^{-\mu}}{n!} \qquad (6)$$

where $\mu = E\{n\}$ is the expected, or mean, number of events. The uncertainty implicit in this probabilistic description corresponds to noise of quantum origin and we wish to assess the influence of this *quantum noise* on system performance.

Consider detection of an optical pulse with expected number of detected photons μ using a decision threshold F. The probability of not detecting such a pulse is given by

$$P[0|1] = \sum_{n=0}^{[F]} P(n,\mu) \geq e^{-\mu} \qquad (7)$$

the lower bound being obtained by selecting a threshold of just *one* detected photon. If we neglect extraneous noise and assume perfect on–off binary signalling then the false alarm probability $P[1|0]$ is zero and the average error probability is given by

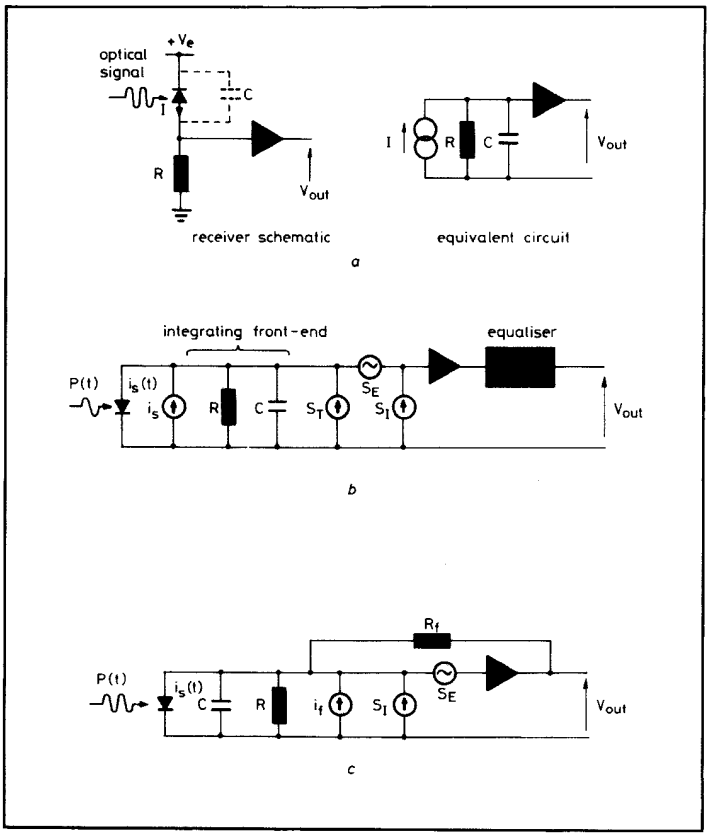
$$\begin{aligned} P_e &= P[0|1]\,P[1] + P[1|0]\,P[0] \\ &= P[0|1]\,P[1] \text{ since } P[1|0] = 0 \\ &= \tfrac{1}{2} P[0|1] \text{ with } P[1] = P[0] = \tfrac{1}{2} \end{aligned} \qquad (8)$$

Substituting for $P[0|1]$, we obtain as a lower bound an error probability

$$P_e \geq \tfrac{1}{2}\, e^{-\mu} \qquad (9)$$

The relationship between this error-probability bound and the received optical-power level is summarised for various bit rates in Table 1.

Practical receivers are, for reasons we shall see shortly,



3 Basic direct-detection receiver structures: (*a*) low-impedance receiver; (*b*) high-impedance receiver or integrating front-end; (*c*) transimpedance receiver

Table 1 Fundamental quantum limit for binary direct detection

P_e	μ	Optical power ($\lambda \approx 1\,\mu$m)		
		8 Mbit/s	140 Mbit/s	1·2 Gbit/s
10^{-5}	12			
10^{-7}	17	≈10 pW	≈100 pW	≈1 nW
10^{-9}	21	≡-80 dBm	≡-70 dBm	≡-60 dBm

typically some 20 dB less sensitive than is indicated by the fundamental limits of Table 1, and require around 2000 photons/bit.

It is instructive, also, to provide a quantum-limited signal-to-noise ratio (SNR) analysis. Here, the quantum noise due to the underlying Poisson process manifests itself as shot noise on the photocurrent. If we measure the signal in terms of the mean photocurrent I given by eqn. 3 then the mean-square shot noise current associated with this signal is given by

$$\overline{i_n^2} = 2qBI \qquad (10)$$

where B is the postdetection noise bandwidth. The signal-to-noise ratio SNR may then be expressed as

$$\text{SNR} = \frac{I^2}{\overline{i_n^2}} = \frac{I^2}{2qBI} = \frac{I}{2qB} = \eta \frac{P}{2hvB} \leq \frac{P}{2hvB} \qquad (11)$$

As an illustration, to obtain 50 dB SNR with a system bandwidth of 5 MHz at an operational wavelength of 1300 nm requires, on purely quantum grounds, a received optical-power level of at least −43 dBm. Once again, practical receivers are somewhat less sensitive than this.

It should be noted that the mean-square shot noise current was estimated above on the basis of the *mean* photocurrent corresponding to the mean received optical power. In fact, the shot noise is dependent on the instantaneous expected value of the photocurrent and this varies with the incident optical signal. We refer to this as signal-dependent noise and note that the simplistic SNR analysis presented here neglects many factors of practical importance. In general we may need to take account of the signal dependence not only of the noise variance but of the probability density function (PDF). For digital communications a higher noise level may be encountered when receiving a 'one' than when receiving a 'zero' and this, together with the detailed PDF, can influence the optimum decision threshold setting F.

Other sources of noise

A variety of other sources of

noise are encountered in practical receivers. These include:

- shot noise on dark current, with spectral density qI_d A²/Hz
- thermal noise due to load and bias resistance R, with spectral density $2k\theta/R$ A²/Hz
- amplifier noise, most conveniently modelled by effective current and voltage noise generators at the input to a hypothetical noise-free amplifier
- avalanche gain noise, associated with shot noise on the signal current—and certain components of dark current experiencing avalanche gain—being increased by a factor $E\{g^2\}$, i.e. the second moment of the gain random variable, whereas (signal current)² is increased by the lesser factor of $(E\{g\})^2$. This implies an avalanche gain noise factor of $E\{g^2\}/(E\{g\})^2 \approx (E\{g\})^x$, with $x \sim 0{\cdot}3$ to $0{\cdot}5$ for a good silicon APD and $x \sim 1$ for a germanium APD or a III-V compound APD.

The performance of a specific optical receiver will depend on the combination of these various noise processes.

5 Direct-detection receiver structures

The overall performance of an optical receiver is critically dependent upon the way in which the photodiode is embedded within the electronic amplification circuitry and also upon the details of the postdetection electrical filtering introduced to limit noise and to shape appropriately the digital signal waveform. We shall consider three basic structures all of which, together with minor variants, find application in optical communications. We shall not concern ourselves with the precise details of amplifier realisation technology but will instead concentrate on the underlying structures.

Low-impedance receiver

The conceptually simplest receiver structure involves terminating the photodiode such that the combination of device capacitance and detector load resistance has a sufficiently short time constant to enable all details of interest in the signal waveform to be resolved. That is the detection system bandwidth B should be at least as great as the signal spectrum bandwidth B_s so that the signal can be passed relatively unimpaired. This requires that the load resistance should be sufficiently small, according to

$$R \leq \frac{1}{2\pi CB_s} \tag{12}$$

Using a low value of passive resistance R introduces a relatively high value for the mean-square noise current, the spectral density being white with double-sided spectral density

$$\frac{2k\theta}{R} \tag{13}$$

so that

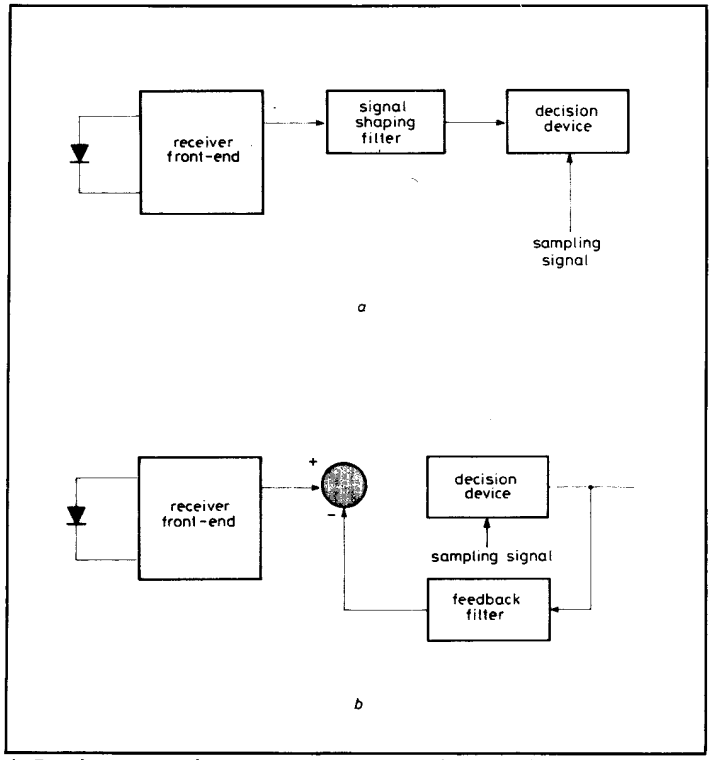
$$\bar{i}_n^2 = \frac{4k\theta B}{R} \tag{14}$$

This noise term generally dominates the performance of a low-impedance receiver, which is thus characterised by a predominantly white noise power spectral density. The receiver schematic and its corresponding noise equivalent circuit are shown in Fig. 3*a*.

High-impedance receiver

In view of the high thermal-noise level encountered with the low-impedance receiver, it is appropriate to consider carefully the implications of letting R be large. The noise spectral density from $2k\theta/R$ is then much reduced but severe signal bandlimiting results due to the CR time constant. This is, though, only a first-order lowpass function and can be compensated by a first-order lead-network equaliser following the amplifier, as shown in Fig. 3*b*. The equaliser corrects for the signal bandlimiting introduced by the front-end circuitry and in so doing renders white the noise contributions associated with current noise components at the front-end, these having been bandlimited by the detector CR circuit and corrected by the equaliser in precisely the same way as has the signal. Any components of amplifier noise, though, corresponding to an equivalent noise *voltage* at the input to the amplifier are enhanced by the equaliser, which imparts a (frequency)² spectral-density dependence. For wideband receivers this f^2 noise tends to dominate but the overall noise level relative to the signal is much reduced compared to that of a low-impedance design.

It should be noted that the high-impedance receiver behaves essentially as an integrator for signal frequencies greater than



4 Post-detection signal processing arrangements: (*a*) linear equalising receiver; (*b*) decision feedback receiver

$1/(2\pi CR)$, which, with R large, means in practice for all signal frequencies of interest. This receiver arrangement is thus often referred to as an integrating front-end. One disadvantageous consequence of this integration action is that rather wide signal voltage excursions must be accommodated and this is exacerbated if a wide range of signal levels must be accommodated. If the mean voltage level varies too widely then this can disturb the bias conditions on the photodiode and amplifier input stage. As a consequence the dynamic range of a high impedance receiver is generally limited.

Transimpedance receiver

It is possible to a certain extent to achieve simultaneously many of the benefits of low noise (as in a high-impedance receiver) and wide unequalised bandwidth (as in a low impedance receiver) by using a transimpedance amplifier configuration as shown in Fig. 3*c*. The *passive* resistance level, dominated in this case by the feedback resistance R_f, is high, so giving a low associated thermal-noise spectral density $2k\theta/R_f$, but the *active* impedance level produced by the action of the amplifier gain operating on R_f in a feedback loop can be low, of the order of $R_f/(1+A)$. It is this latter factor, in combination with the detector and stray capacitance, which determines the unequalised bandwidth. Detailed analysis shows that, once more, the amplifier effective voltage noise dominates and that the action of the feedback loop together with the total shunt input capacitance is to impart a predominantly (frequency)^2 character to this, much as for the high-impedance receiver. The benefit of the transimpedance amplifier, then, is that it does not require a subsequent equaliser and the signal is not integrated by the front-end so that the dynamic range is not limited in the same way as is that for a high-impedance receiver.

6 Receiver signal processing

Let us turn our attention now to the question of postdetection signal processing for binary optical communications. A conventional linear equalising receiver structure is illustrated in Fig. 4*a* and a decision feedback receiver in Fig. 4*b*. The latter has been shown to have advantages for optical-fibre communications when fibre bandwidth is limited[9]—such as may arise with the use of graded-index fibre for long-haul systems—and as a means of combatting the (frequency)^2 noise spectrum associated with high-impedance and transimpedance front-ends.[10]

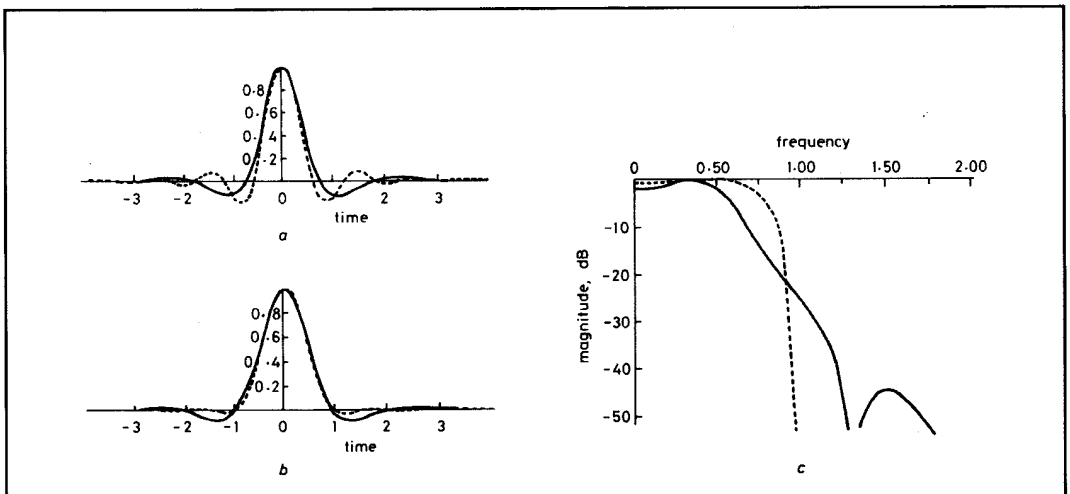
In the main, though, linear equalising receivers have been employed and this is where we concentrate our emphasis. The signal-shaping filter has the task of bandlimiting the signal to limit noise and of shaping the signal such that it is largely concentrated within a single signalling interval so that undue intersymbol interference (ISI) is not experienced. In fact the most common design target is that the equalised signal-element spectrum should have a raised-cosine form:

$$x(t) \leftrightarrow X(f) = \begin{cases} \frac{1}{2}[1+\cos(\pi fT)], & |f|<1/T \\ 0, & \text{elsewhere} \end{cases} \quad (15)$$

where $\leftrightarrow$ denotes a Fourier transform pair. This signal format has zero ISI and yields a system eye diagram which is widely open, both horizontally and vertically. The signal at the output of the shaping filter is passed to a decision device with threshold F; a signal greater than F at the sample time is interpreted as a 'one' whereas a signal less than F indicates a zero. F should clearly be chosen such as to minimise the error probability.

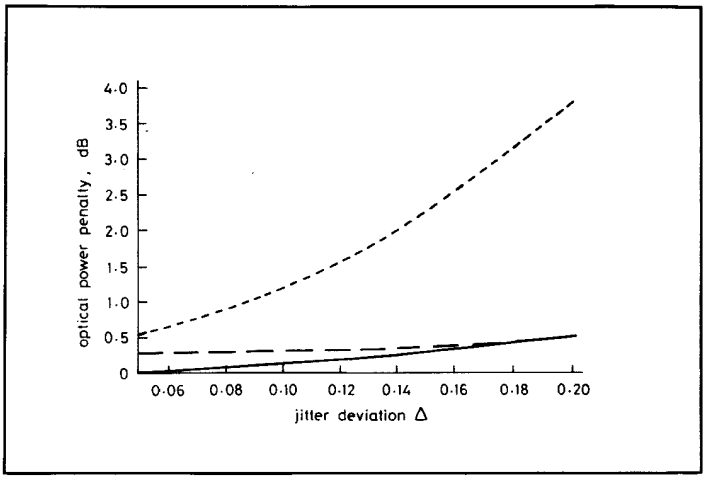
In fact, both the shaping filter and the decision threshold influence the error probability P_e and it is desirable to identify receivers (filters and threshold values) to minimise P_e. To do this formally is a very demanding task due to the wide variety of stochastic impairments which must be accommodated. These are:

- the underlying inhomogeneous Poisson process, which is essentially of quantum origin
- avalanche gain noise, which formally renders this a form of *compound* random process known as a *marked and filtered* Poisson process (MFPP)
- coloured Gaussian additive noise associated with the receiver electronics
- intersymbol interference (ISI)
- jitter, i.e. timing impairments associated with the signal elements and/or the times at which the signal waveform is sampled to extract the data.



5 Comparison of response functions for conventional raised-cosine (---) and Chernoff bound (——) optical receivers: (*a*) impulse response; (*b*) output pulse; (*c*) frequency response

It has been found, though, that a very tight upperbound on P_e, known as the modified Chernoff bound (MCB), can be formulated in such a way that it accommodates all these impairments yet is sufficiently analytically tractable to allow for receiver optimisation using variational calculus.[11] In this way, optimum receivers—filters and decision threshold values—have been identified for direct-detection optical communications in the presence of a comprehensive set of system impairments. Illustrative signal waveforms and frequency response plots for a receiver designed on this basis are shown in Fig. 5.[12] An important feature of this approach is that it accommodates, in a rigorous manner, the critical impairments associated with practical systems. Furthermore, receivers designed to accommodate jitter have been found to be *jitter-tolerant* in that the receiver minimum detectable power for a given error rate is relatively constant over a wide range, and is insensitive to the precise jitter levels, as illustrated in Fig. 6.[13]



6 Power penalty curve for a conventional 100% raised-cosine receiver and a jitter-tolerant receiver optimised under the modified Chernoff bound
---- **raised-cosine receiver**
— — **receiver optimised for Δ = 0·20**
—— **receiver optimised for each Δ**

7 Coherent detection

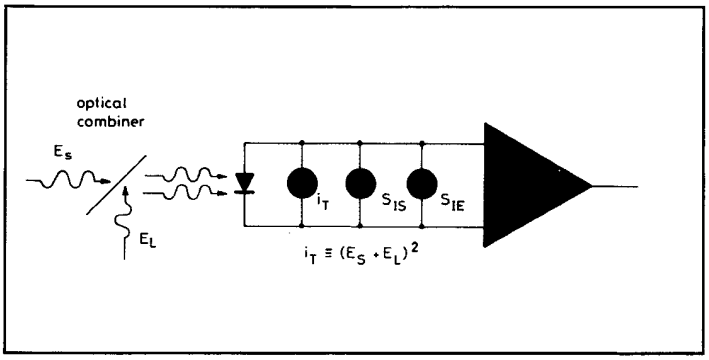
Up to this point we have concerned ourselves solely with direct detection, although it was noted in the introduction that coherent optical detection was possible. For present-day direct-detection receivers, additive receiver noise is a limiting factor. Although avalanche gain can in principle alleviate these difficulties by bringing the signal (and its associated quantum noise) up above the receiver noise level, the gain statistics of practical devices are such as to severely limit the potential of this strategy.

An alternative, which is now the subject of intense investigation in many laboratories, is provided by coherent detection. We shall discuss this only briefly here. The principle may be appreciated with reference to Fig. 7. The incident optical signal, which we denote as $E_s\cos\omega_s t$, is combined with a local oscillator signal $E_L\cos\omega_L t$ so that the input to the photodiode is the sum

$$E_s\cos\omega_s t + E_L\cos\omega_L t \qquad (16)$$

The detector responds to optical power, producing an output current proportional to the total optical power averaged over many cycles of the lightwave signal, so that we have

$$i \equiv \eta \frac{P_{opt}}{h\nu} q \qquad (17)$$



7 Principle of coherent detection

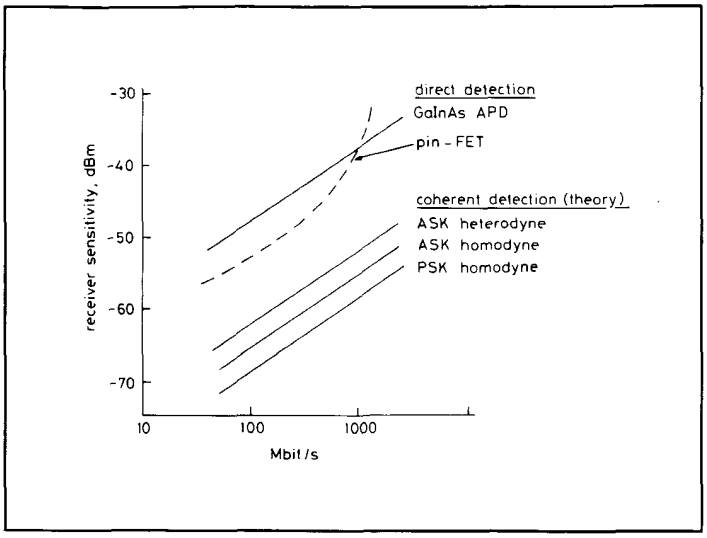
Note that ω_s and ω_L are sufficiently close to one another that, for these purposes, we need not distinguish between them, taking $\omega_s \approx \omega_L \approx 2\pi\nu$. Since optical power is proportional to the square of the electric field, the output current is given by

$$\overline{i(t)} \equiv \overline{(E_s\cos\omega_s t + E_L\cos\omega_L t)^2}$$
$$= \overline{E_s^2\cos^2\omega_s t + E_L^2\cos^2\omega_L t + 2E_sE_L\cos\omega_s t\cos\omega_L t}$$
$$= \tfrac{1}{2}E_s^2 + \tfrac{1}{2}E_L^2 + E_sE_L\cos(\omega_s - \omega_L)t \qquad (18)$$

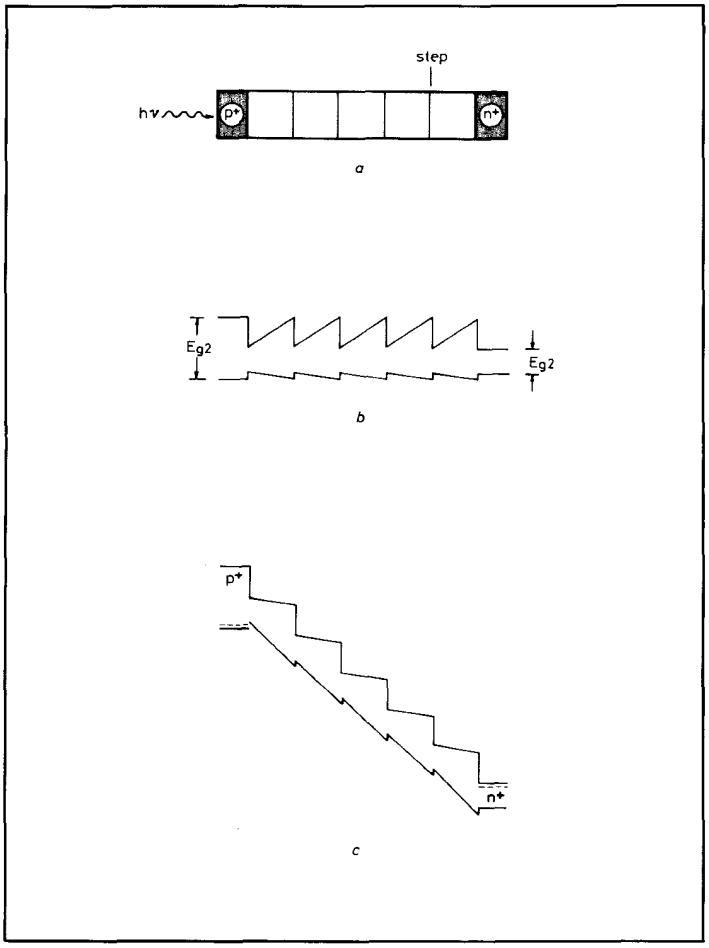
where the overbar represents time averaging over intervals which are large in comparison with the period of the optical wave but short in comparison with the period of the difference frequency $|\omega_s - \omega_L|$ or other signal fluctuations.

With $E_s \ll E_L$ we can neglect the first term in eqn. 18. The second term is just a DC component, whereas the third term conveys information carried in any fluctuations in E_s. Provided the local oscillator power is sufficient we can ensure that the detected signal component $E_sE_L\cos(\omega_s - \omega_L)t$ is large in comparison with noise associated with the receiver amplifier and bias components. It is this conversion gain, realised by E_L being large, which is the main attraction of coherent detection. We should note, though, that this does not mean that *all* noise sources can be rendered negligible. As is shown below, quantum noise, manifesting itself here as the equivalent of shot noise on the total photocurrent, sets a limit to the performance attainable.

There are essentially two classes of coherent detection: homodyne and heterodyne. For homodyne detection the signal and local-oscillator frequencies are equal ($\omega_s = \omega_L$) and the signals are phase synchronous so that signal fluctuations contained in E_s are



8 Receiver sensitivity comparison for (practical) direct-detection and (theoretical) coherent-detection schemes



9 Superlattice avalanche photodiode: (*a*) schematic diagram; (*b*) energy band diagram under zero bias; (*c*) energy band diagram under reverse bias

translated directly down to baseband:

$$i_{AC}(t) = E_L E_s \cos(\omega_s - \omega_L)t \equiv E_L E_s \quad (19)$$

For heterodyne detection $\omega_s \neq \omega_L$ and the electrical detected signal $\equiv E_s E_L \cos(\omega_s - \omega_L)t$ is amplified at the intermediate frequency and then passed to a second, *electrical*, detector.

In principle, homodyne detection offers slightly better noise performance than heterodyne detection. However, this requires that the optical local oscillator be phase locked to the incoming optical signal. This has been achieved very satisfactorily in the laboratory[14] although the complexity is non-negligible.

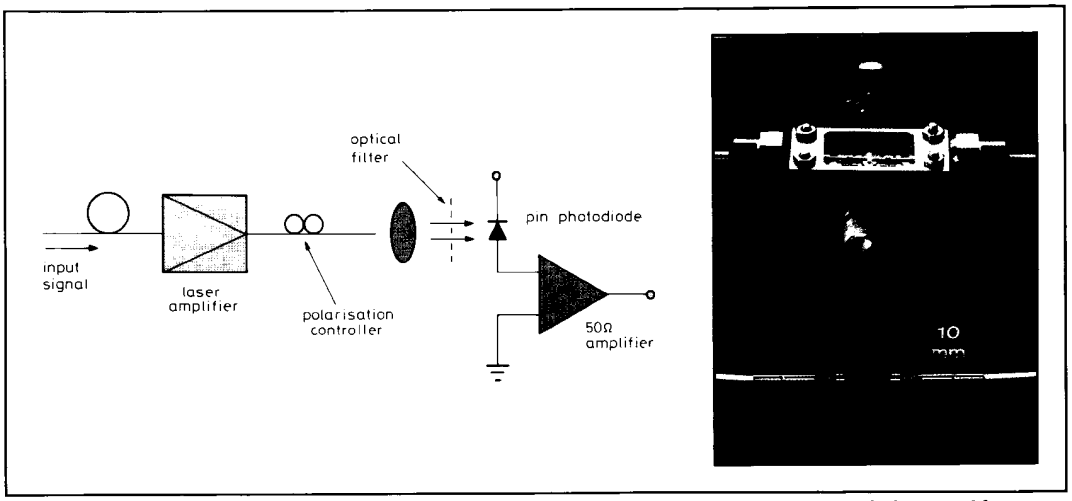
In some circumstances it may be preferable to forego the slight performance advantage of homodyne detection in favour of the simpler realisation involved in phase locking to the electrical intermediate-frequency signal for heterodyne detection. For very wideband applications, though, it should be recognised that the greater front-end receiver bandwidth demanded by heterodyne detection can present difficulties. In this connection the use of partial-response class I signalling has recently been shown to have benefits.[15]

The performance attainable in principle using coherent detection with various signal modulation formats is illustrated in Fig. 8, alongside the typical performance achieved with direct detection.

8 Advanced direct-detection receivers

Given the relative complexity of coherent detection there is clearly merit in investigating the extent to which direct-detection receivers can be refined so as more closely to approach quantum-limited operation. For APD receivers the major limitation here is the gain randomness associated with the avalanche photodiode. Considerable research effort is being directed towards the development and characterisation of advanced APD receivers based on superlattice device structures which offer reduced gain randomness.[3,4] The band diagram for such a 'staircase' APD is illustrated in Fig. 9. Here, carrier multiplication occurs at the well-defined energy band steps, thereby partially removing the randomness associated with the avalanche gain mechanism of a conventional APD.

An alternative approach to improving the performance of



10 Optical receiver using an injection laser pre-amplifier: (*a*) system schematic; (*b*) illustrative realisation of a laser amplifier

direct-detection receivers relies on the use of an injection diode laser as an optical pre-amplifier. The basic arrangement is shown in Fig. 10*a* and a practical realisation in Fig. 10*b*. Theoretical and experimental evidence indicates that a performance comparable with that of coherent-detection receivers is achievable by this means.[16,17] A clear area for further research and development is an examination of the extent to which the novel signal-processing schemes outlined in Section 6 can provide improvements upon the performance achieved to date with these advanced direct-detection receivers.

9 Concluding remarks

This paper has sought to provide a tutorial review of basic optical-receiver principles, ultimately with an emphasis on advanced signal-processing schemes. Included in the latter are optical and optoelectronic techniques—including coherent detection, optical amplification and the use of advanced superlattice optoelectronic detectors—and electronic signal processing which provides receivers able to optimally accommodate the complex stochastic impairments encountered in optical receivers. The judicious marrying of these novel electronic and optoelectronic signal-processing arrangements has been identified as an area deserving further research.

Acknowledgments

Certain of the research reported here has been undertaken at the University of Wales, Bangor, with support from the UK Science and Engineering Research Council. The author is grateful to his colleagues in the Optical Communications Group, Dr. Klaus Schumacher, Raad Fyath, Dr. Henrique Silva and Manuel Monteiro, for useful discussions and for specific illustrative results, and to Dr. M. J. O'Mahony of British Telecom Research Laboratories for permission to reproduce Fig. 10*b*.

References

1 STANLEY, I. W., *et al.*: 'The application of coherent optical techniques to wideband networks', *J. Lightwave Technol.*, 1987, **LT-5**, pp. 439-451
2 O'REILLY, J. J.: 'Optical fibre networks', Chap. 20 *in* GRIFFITHS, J. M. (Ed.): 'Local telecommunications 2', revised edition (Peter Peregrinus, London, 1988)
3 CAPASSO, F.: 'Bandgap engineering via graded gap, superlattice, and periodic doping structures: applications to novel photodetectors and other devices', *J. Vac. Sci. & Technol. B1*, 1983, **2**, pp. 457-461
4 O'REILLY, J. J., and FYATH, R. S.: 'Analysis of the influence of dark current on the performance of optical receivers employing superlattice APDs', *IEE Proc. J, Optoelectron.*, 1988, **135**, pp. 109-118
5 MUKAI, T., YAMAMOTO, Y., and KIMARA, K.: 'S/N and error rate performance in AlGaAs semiconductor laser amplifier and linear repeater systems', *IEEE Trans.*, 1982, **MTT-30**, p. 1548
6 MARSHALL, I.W., and O'MAHONY, M.J.: '10 GHz optical receiver using a travelling-wave semiconductor laser preamplifier', *Electron. Lett.*, 1987, **23**, (20), pp. 1052-1953
7 GARRETT, I.: 'Towards the fundamental limits of optical-fibre communications', *J. Lightwave Technol.*, 1983, **LT-1**, pp. 131-138
8 O'REILLY, J. J.: 'Approaching fundamental limits to digital optical-fibre communications' *in* ZORCOCZKY, P. I. (Ed.): 'Oxford surveys in information technology. Vol. 3' (Oxford University Press, 1986)
9 BALL, P.: 'Extending the range of long wavelength multimode optical fibre transmission using decision feedback', Proc. 8th European Conf. on Optical communications, Cannes, France, September 1982, pp. 406-410
10 O'REILLY, J. J., DUARTE, M., and BLANK, L.: 'On the application of quantised feedback and partial response techniques to high bit-rate optical fibre communications'. Proc. 1982 Allerton Conf., Monticello, Illinois, USA, October 1982, pp. 967-974
11 O'REILLY, J. J., and DA ROCHA, J. R. F.: 'Improved error probability evaluation methods for direct detection optical communication systems', *IEEE Trans.*, 1987, **IT-33**, (6), pp. 839-848
12 DA ROCHA, J. R. F., and O'REILLY, J. J.: 'Linear direct•detection fiber-optic receiver optimization in the presence of intersymbol interference', *IEEE Trans.*, 1986, **COM-34**, (4), pp. 365-374
13 O'REILLY, J. J., DA ROCHA, J. R. F., and SCHUMACHER, K.: 'Optical fiber direct detection receivers optimally tolerant to jitter', *IEEE Trans.*, 1986, **COM-34**, (11), pp. 1141-1147
14 MALYON, D. J.: 'Digital fibre transmission using optical homodyne detection', *Electron. Lett.*, 1984, **20**, pp. 281-283
15 O'REILLY, J. J., and MONTEIRO, M.: 'Combatting limited receiver bandwidth for 4 ϕ-PSK coherent optical transmission—a comparison of full-response and partial response signalling schemes'. Proc. IEEE Int. Conf. on Commun. systems, Singapore, November 1988
16 O'MAHONY, M. J., and MARSHALL, I. W.: 'Wideband optical receivers using travelling wave amplifiers for Gbit/s systems'. Proc. IEEE Globecom. '87, Tokyo, November 1987, pp. 833-836
17 O'MAHONY, M. J., *et al.*: 'Wideband 1·5 μm optical receiver using travelling-wave laser amplifier', *Electron. Lett.*, 1986, **22**, pp. 1238-1240

First received 16th December 1988 and in revised form 30th January 1989

John O'Reilly is Head of the School of Electronic Engineering Science, University College of North Wales, Dean Street, Bangor, Gwynedd LL57 1UT. Professor O'Reilly is an IEE Fellow