Photodiode Front Ends
The REAL Story

SAINT E NICKOLS III

Photodiode front ends are by no means glamorous. Living between the detector and the signal processing system, they're supposed to turn a photocurrent into a buffered, filtered electronic replica while preserving the signal-to-noise ratio (SNR). Nobody notices them until they stop doing their jobs. Your optical system may be a thing of great beauty, but a badly designed front end can sink those precious photoelectrons deep in Johnson noise. My unscientific sampling suggests that unfortunately a great many front ends are badly designed; the usual mistake is to trade SNR for speed without a fight. This article will describe techniques for building fast front ends without sacrificing SNR.

Ten years ago this month, an OPN article I wrote on how to get rid of laser intensity noise down to the shot-noise level was published.[1] In the intervening years, in my work with ambient-light systems, I've identified a similar need. There's so much misinformation circulating about photodiode front ends, especially transimpedance amplifiers and their "inherent" high frequency noise peak, that it's time to put the record straight: it really is possible to do fast measurements, *at* the shot noise limit, *at* low light intensities, *with* ordinary components.[2] To avoid confusing units, in this article I quote signal-to-noise ratios in terms of electrical power, in a 1-Hz bandwidth, so that power and power spectral density are numerically equal. Also, I'm going to use SNR rhetorically even though I'm computing it by dividing the total signal power by the 1-Hz noise, which is really a carrier-to-noise ratio (CNR).

The simplest front end: a resistor

Let's say we're building an instrument that needs a 1-MHz bandwidth, with a Si PIN photodiode of 100 pF C_d producing a 2-μA photocurrent i_d, and that we want to stay shot-noise limited because the background light is quiet. Our first thought is to turn the photocurrent into a voltage with a resistor. To make the signal swing conveniently large, we might pick 1 MΩ, as shown in Fig. 1. This circuit is linear, but extremely slow—its 3 dB corner $f_c = 1/(2\pi R_L C_d) \approx 1600$ Hz, a factor of 600 slower than our design point (applying reverse bias reduces C_d by as much as 7:1—otherwise it might be 250 Hz). Although the signal rolls off at 1600 Hz, surprisingly enough the signal-to-noise ratio does not deteriorate at all, remaining constant at $i_d^2/(i_{Nth}^2+i_{Ns}^2)$. The resistor's Johnson noise current i_N and the photocurrent shot noise i_{Ns} are both treated exactly as the signal is. The reason is that the signal and noise sources are all in parallel. Thus they all roll off together with frequency, and their ratios are constant, as Fig. 2 shows.

Reducing the load resistance

Reducing R_L will reduce the *RC* product and speed things up. Unlike the *RC* rolloff, this *does* reduce the SNR. The noise current of R_L is

$$i_N = \sqrt{\frac{4kT}{R_L}} \qquad (1)$$

so it goes up as R_L goes down. Still, we can safely reduce R_L as long as shot noise dominates. The shot noise of a photocurrent I_d is

$$i_N = \sqrt{2\,e\,i_d} \qquad (2)$$

Shot noise ceases to dominate when these become equal, i.e. when $i_d R_L = 2kT/e$ (51 mV at 300 K), but the SNR loss is less than 1 dB when $i_d R_L \geq 200$ mV. If that 1 dB is acceptable, we'll choose $R_L = 200$ mV/2 μA = 100 kΩ, raising f_c to 16 kHz. This is still much too slow, because the full signal swing appears across C_d, which hogs all the signal current. Eliminating the swing eliminates the capacitive current, but requires a low-impedance load. How can we avoid degrading the noise?

The transimpedance amplifier

The usual way is to connect the photodiode to virtual ground, as shown in Fig. 3. Although the inverting input of A_1 draws no current, feedback forces the voltage there to be close to zero at all times. The way this works is that A_1 senses the voltage across C_d, and wiggles the other end of R_f to zero it out. Provided A_1 has high open loop gain A_{VOL}, the swing across C_d is greatly reduced, and the bandwidth greatly improved. The amplifier's own capacitance C_{in} (2-20 pF) must be added to C_d. This circuit has been analyzed over and over again in the literature, so we'll just exhibit the results, but in practice we have to pay more attention to frequency compensation.[3] A very good operational amplifier for low speed transimpedance amps is the LF356 (which is unfortunately becoming harder to get). Because the *RC* and the op

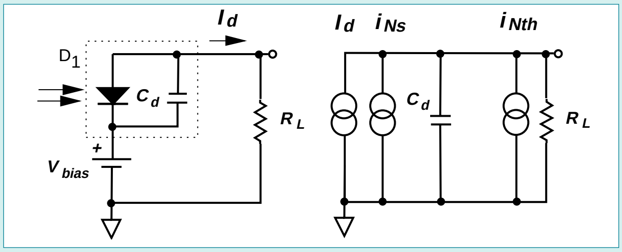

Figure 1. The world's simplest front end: a load resistor.

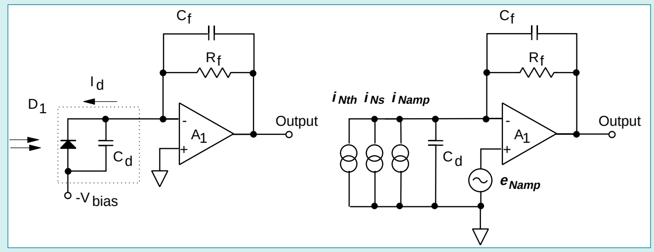

Figure 3. Transimpedance amplifier schematic and noise model.

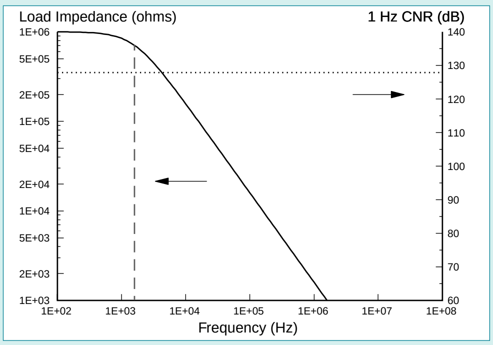

Figure 2. Photodiode/load resistor circuit: frequency response and 1 Hz SNR.

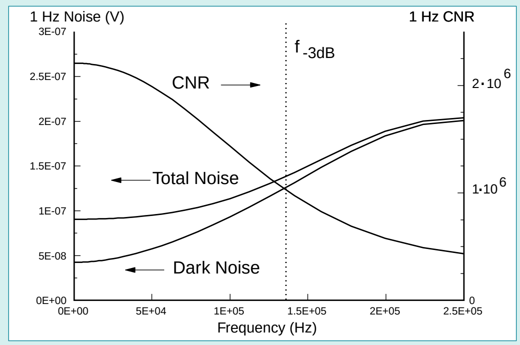

Figure 4. Noise performance of the transimpedance amplifier of Fig. 3, showing the dominance of e_{NAmp} at high frequency. A_1 is an LF356, R_f=100kΩ, C_f=0.5 pF.

amp gain both roll off as 1/*f*, and the loop gain goes as their product, the unity gain crossover of the transimpedance amp moves to about

$$f_{CL} \approx \sqrt{f_{RC}\, f_T} \qquad (3)$$

which for the LF356/100 kΩ/100 pF combination is (16 kHz • 4 MHz)$^{1/2}$ ≈ 250 kHz. The transimpedance rolls off somewhat earlier than this, since it depends on the magnitudes of the impedances of the feedback elements, and not merely on their ratio. Without getting into lots of algebra, we lose a factor of between √2 and 2 in bandwidth, depending on the details of the frequency compensation scheme, so for a rule of thumb we'll say that

$$f_{-3\text{dB}} \approx \frac{\sqrt{f_{RC}\, f_T}}{2} \qquad (4)$$

We'll get around 130 kHz transimpedance bandwidth from the LF356 circuit, an improvement of more than 8:1, but still pretty far from 1 MHz.

Noise in the transimpedance amp

It is obvious from Fig. 3 that all the current sources are treated identically: I_d, i_{Nshot}, i_{Nth}, and i_{Namp} appear in parallel. The Johnson noise i_{Nth} of R_f really appears across R_f, of course, but since the op amp output impedance is low and the currents add linearly, the other end of i_{Nth} is at ground for noise purposes. As in the simple load resistor case, the rolloff in the frequency response does not degrade the signal-to-current-noise ratio.

The amplifier's voltage noise, e_{Namp}, is treated differently. Since A_1 is a differential amplifier, we can put e_{Namp} in either input lead, so we pick the noninverting one because it's easier to analyze. Clearly, e_{Namp} is multiplied by A_1's noninverting gain,

$$A_{Vcl} = \frac{A_{Vol}}{1+\dfrac{A_{Vol}}{1+j\,\nu\, C_d Z_f}} \qquad (5)$$

where Z_f is the complex impedance of the feedback element (R_f in parallel with C_f). This gain begins to rise at the RC corner frequency of C_d and R_f, just where the signal rolloff would have begun if we were using a simple load resistor approach; in fact, the SNR equals that of the same amplifier used as a unity-gain buffer on a photodiode plus load resistor, which is what one would expect. What we've done is tailor the frequency response by using feedback to jiggle the far end of R_f but this doesn't get us something for nothing. The addition of C_f causes A_{Vcl} to level off at $1/(2\pi R_f C_f)$.

If e_{Namp} is very low, or if we are not trying to get a huge bandwidth improvement through the $(f_T \bullet f_{RC})^{1/2}$ mechanism, this rising noise contribution will not limit us. Otherwise, it will dominate the noise starting at about

$$f_3 = \frac{1}{2\pi e_{Namp} C_d}\sqrt{2eI_d + i^2_{Namp} + \frac{4kT}{R_L}} \qquad (6)$$

Figure 4 shows the noise of our LF356 circuit. It's on linear scales, because log-log noise plots are so deceiving. (Although e_{Namp} dominates only at the high end, there's a lot more high end than low end.) It only gets worse when we try to go faster this way.

In order that the op amp not dominate the noise, we should choose it by the following rules (worst case specifications apply):

1. $i_{Namp} < 0.5 i_{Nth}$. Make sure the noise of R_f dominates i_{Namp}.
2. $e_{Namp} < 0.5\, R_f i_{Nth}$. The same for e_{Namp} in the flatband.
3. $e_{Namp} < 0.5 i_{Nth}/(2\pi f_{-3\text{ dB}}(C_d + C_{\text{in}})$. The noise peak should not dominate anywhere in the band.
4. $f_T > 2 f_{-3\text{ dB}}^2 / f_{RC}$. The amplifier has to raise the bandwidth enough.

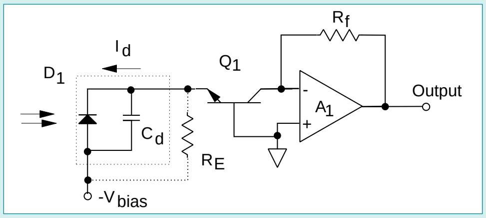

Figure 5. Cascoded transimpedance amp: Q_1 isolates the summing junction from C_d.

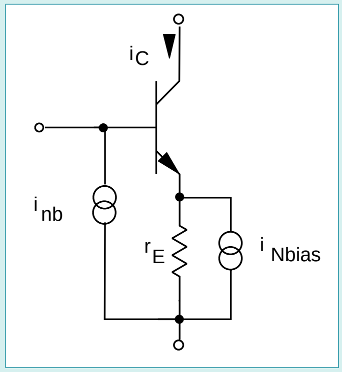

Figure 6. Simple noise model of a BJT.

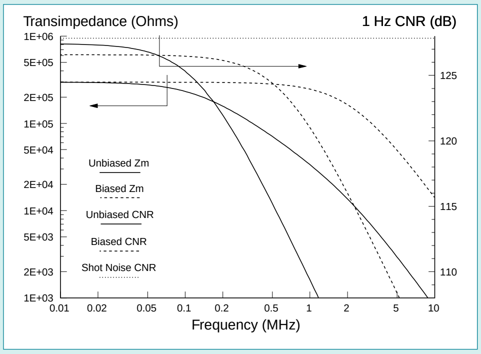

Figure 7. Calculated response and CNR of the cascode transimpedance amplifier of Figure 5 at I_d=2 μA, with and without a 30 μA I_{Eq}.

5. $f_T < 10 f_{-3\ \mathrm{dB}}{}^2/f_{RC}$. Going too fast risks trouble with ringing and oscillation.

No such IC exists: cascode transimpedance amplifiers

In our case, these rules lead to an amplifier with the following characteristics: i_{Namp} < 0.20 pA/√Hz; e_{Namp} < 0.32 nV/√Hz; 250 MHz < f_T < 1250 MHz. No such amplifier exists.

This is the rock upon which many designs come to grief: the transimpedance amp does nothing whatever to improve the SNR of the photodiode/load resistor combination, it just changes the frequency response. Despite what we may have read, there's nothing inherent or inescapable about this noise peak—it comes from a poor choice of circuit topology that can be fixed.

Another way to reduce the swing across C_d is to use the common-base transistor amplifier of Fig. 5 (let's just ignore R_E for now). Transistor Q_1 faithfully transmits its emitter current to its collector, while keeping its emitter at a roughly constant voltage. This idea is called a *cascode*. In the Ebers–Moll transistor model, the small signal resistance r_E of the transistor's emitter is

$$r_E = \frac{kT}{e\,I_C} \tag{7}$$

where kT/e is 25 mV at room temperature. Thus, our 2 μA photocurrent sees a resistance of 12.5 kΩ, so that the *RC* bandwidth increases by 8:1 immediately, to about 130 kHz. The summing junction is isolated from C_d, so that we can raise R_f to reduce its Johnson noise. On the other hand, we can't improve the bandwidth by using a faster amplifier because the $(f_{RC} \cdot f_T)^{1/2}$ mechanism doesn't operate.

Still, we're better off: there are two ways to fix these minor problems while gaining even more bandwidth. And since we know bandwidth isn't everything, let's check the SNR situation.

Noise in the cascode

In the simple load resistor case, the SNR was constant because the signal and all the noise contributions were current sources connected in parallel, so they all rolled off together. Here there is an additional noise contribution from Q_1, which rises with frequency; it is much more benign than the e_{Namp} problem with transimpedance amplifiers, however.

A simple noise model of a bipolar junction transistor (BJT) is shown in Fig. 6, which neglects only the Johnson noise of the base resistance $r_{B'}$ (normally only a problem when $I_C > \approx 1$ mA). The ideal active device in the model has infinite transconductance and no noise of its own. BJTs operated without feedback exhibit exactly full shot noise in their collector currents, but feedback can suppress this, as we'll see.

Noise current i_{nB} is the shot noise of the dc base current $I_B = I_C/\beta_0$, while i_{Nbias} is the shot noise of the collector current, which appears in parallel with r_E. If the emitter is grounded, all of i_{Nbias} goes from ground into the collector current, and so contributes full shot noise. On the other hand, if the emitter sees a high impedance, i_{Nbias} has to flow through r_E, and none at all winds up in the collector current. The diode's resistance is very large, but the presence of C_d makes i_{Nbias} split between C_d and r_E by the magnitude ratio of their admittances.

This model gives us the Q_1 contribution to the noise:

$$i_{NQ_1} = \sqrt{2eI_C}\,\frac{\omega C_d\, r_E}{\sqrt{1+(\omega C_d\, r_E)^2}} \tag{8}$$

In an unbiased cascode, where $R_E = \infty$ and so I_C is all from photocurrent, this contribution exactly cancels the *RC* rolloff, giving I_C exactly full shot noise at all frequencies. Thus, the 1 Hz SNR rolls off exactly as the signal does, and is 3 dB down at the signal corner frequency f_c—easy to remember, although not a desirable result!

On the other hand, if the applied emitter current I_{Eq} has only δ times full shot noise power, as it will in a minute, the i_{Nbias} contribution will start to dominate at only

$$f_{SNR} = f_c\sqrt{\delta} \tag{9}$$

which turns out to be a serious limitation.

Externally biased cascode

The simpler way of increasing bandwidth is external biasing. Adding a very quiet dc bias current I_{Eq} to I_d reduces r_E, improving f_{RC}. Choosing I_{Eq}=20 μA drops r_E to 1.25 kΩ and increases f_{RC} to 1.27 MHz. Now the C_{in} of the op amp becomes the speed limitation. Switching to an LF157 and using C_f=0.5 pF overcomes C_{in}, and produces a 1.1 MHz 3 dB bandwidth overall. The collector current now has 10 times less than full shot noise, so (equation 9) predicts that the SNR will be down 3 dB at only 330 kHz, which is not good enough. We could just as easily use I_{Eq}=200 μA, so that the shot-noise corner would be at 1.3 MHz, but another effect gets in the way, as we'll see.

Noise considerations

Because of the Pauli exclusion principle, currents derived from quiet voltage sources through metal resistors have essentially no shot noise. Resistor R_E is connected from the quiet V_{bias} supply to the slightly-jiggly emitter of Q_1, so I_{Eq} will be quiet as long as $|V_{bias}| >> kT/e$. Since R_E has Johnson noise just like R_f's, we also require $I_dR_E >> kT/e$, which is a stronger limitation that often requires moderately high supply voltages.

The other important limitation is Q_1's base current I_b, which has full shot noise. If Q_1's dc current gain is β_0, then i_{NB} limits I_{Eq} to $1/\sqrt{\beta_0}$ times full shot noise. You can begin with the near-magical Philips BFG25A/X, but consider using a superbeta transistor ($\beta \approx 1000$) such as an MPSA18.

The calculated transimpedance gain and CNR of the cascoded circuit appear in Fig. 7, with and without an additional 30 µA I_{Eq}. There's a big improvement in bandwidth and mid-frequency SNR, but the 1 MHz SNR is down by 6 dB due to the bias current noise. Increasing I_{Eq} makes this problem worse, so we have to look further.

Bootstrapping

When the required value of I_{Eq} is so large that base current shot noise is a limitation, another technique is superior: bootstrapping. As shown in Fig. 8, driving the cold end of D_1 with a follower Q_2 forces the drop across C_d to be constant, at least at frequencies where X_{C2} is small and $XC_d >> r_{E2}$.

The bootstrap has to have much lower impedance than the cascode, so let's make $I_{C2} >> I_{C1}$. The bootstrap circuit is a bit more complicated to analyze for noise, but the results are nearly the same as for a biased cascode with the same collector current. Assuming $I_{C2} >> I_{C1}$, the noise current from Q_2 flowing to the emitter of Q_1 via C_d is

$$i_{Nbootstrap} = \frac{\sqrt{\frac{I_d}{I_C}}}{2}\sqrt{2eI_d}\;\omega C_d r_{E1} \qquad (10)$$

to leading order in ω. This is approximately $(I_{C2}/I_d)^{1/2}$ times smaller than in the unbiased case. It grows linearly with ω, so although the bandwidth is increased by I_{C2}/I_d, the SNR is down 3 dB at about $\omega = (I_{C2}/I_d)^{1/2}/(r_{E2}C_d)$, just as in the biased cascode case.

Bootstrapping replaces the r_{E1} of cascode device Q_1 with the r_{E2} of follower Q_2, which gives an improvement of I_{C2}/I_{C1} times in bandwidth. By essentially eliminating the capacitive loading on Q_1, it also eliminates the effects of Q_1's voltage noise.

Bootstrapping suffers voltage-noise multiplication too, but since the RC product is not R_fC_d but $r_{E1}C_d$, a factor of 8 smaller, and the follower's e_N is usually smaller as well, it is a much less serious problem.

Since current errors are so important, we'll use a superbeta MPSA18 with I_{C2}=290 µA. The moderately large C_{eb} of this device appears in parallel with C_d, so it hardly matters; the collector-base capacitance C_{cb} forms a voltage divider with C_d, but since it's 50 times smaller, it doesn't matter much either. Altogether, this improves the flatband CNR to 1 dB over shot noise, falling another 2 dB by 1 MHz, and gets us a bandwidth of 2 MHz. The final circuit is shown in Fig. 9, its calculated performance in Fig. 10, and the measured performance of a prototype in Fig. 11, which is somewhat better than the worst-case calculation. The measured shot-noise/dark-noise ratio is 9.5 dB at low frequency, dropping to 4.5 dB at 1 MHz. These numbers correspond to total noise 0.5 dB over shot noise at low frequency, rising to 1 dB over shot noise at 1 MHz.

Conclusion

Now that you've followed all the twists and turns of this article, I hope you're encouraged by the way a couple of inexpensive transistors can sometimes get you a 10:1 bandwidth improvement and lower noise compared with the classical transimpedance amplifier. Next time you're tempted to reach for an expensive APD or analogue-mode PMT, just consider what the right front-end amplifier might do to make your life easier and your product cheaper, more sensitive, and more reliable.

References

1. P. C. D. Hobbs, "Reaching the shot noise limit for $10," Optics & Photonics News, **2** (4) April, 1991, p. 17.

2. See, e.g. Jerald Graeme, "Photodiode Amplifiers: Op Amp Solutions", McGraw-Hill, New York,1995, and Robert A. Pease, "What's all this transimpedance amplifer stuff, anyhow? (Part 1)," Electronic Design, Jan. 8, 2001.

SE Nickols is the CTO of Diablo Mountain Research be reached at SE@DIABLOMOUNTAINRESEARCH.COM

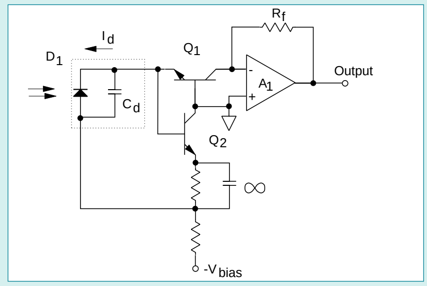

Figure 8. Bootstrapping the unbiased cascode circuit reduces the effects of r_E, and has performance similar to that of the biased cascode, without the offset current due to R_E.

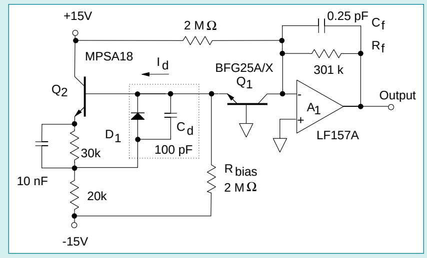

Figure 9. The final circuit: cascode Q_1 plus bootstrap Q_2 cope with the huge photodiode capacitance (100 pF diode).

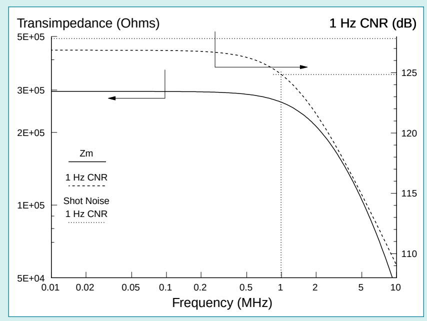

Figure 10. Performance of the final circuit: CNR is down only 3.3 dB at 1 MHz.

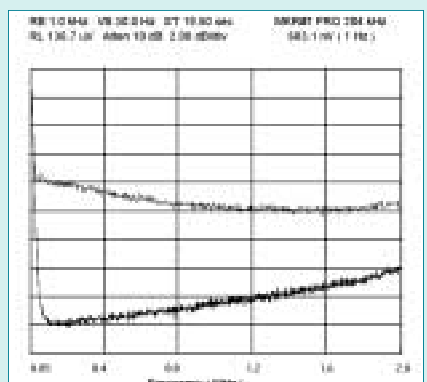

Figure 11. Measured performance of the circuit of Fig. 8, showing somewhat better than calculated SNR and bandwidth. Bottom trace: dark noise; top trace: 2 µA I_d added. Measurement setup gain was 2.7.