THz Receiver Architectures and the Process of Detection

As discussed in the Introduction, an important aspect of all THz sensor design, and all RF systems in general, is receiver architecture. There are two types, incoherent and coherent, both shown schematically in Fig. 1. The incoherent architecture is as old as RF technology itself and the coherent came soon thereafter, dating back to the early part of the 20th century. Like many other system architectures, they persist largely by the ability of engineers to continually improve performance by perfecting the components. In this set of notes, we will address these "canonical" receiver architectures, Then we will look more closely at the concept of "detection", which is so important to understanding how RF systems perform.

Canonical Receiver Architectures

A. Direct-Detection

The block diagram of a generic direct receiver is shown in Fig. 1(a). The incoming radiation from the target, be it thermal emission or transmitted power from the sensor itself, is collected by the receiver where it is rectified from RF (THz) to baseband by a "direct" detector. In most practical cases the baseband is defined by amplitude or frequency modulation of the incoming signal to reduce the effect of gain drifts and 1/f noise that occurs in the THz electronics. The rectified THz signal is then amplified and demodulated down to DC using synchronous detection. For AM modulation the synchronous detection is often carried out using a lock-in amplifier.

In the THz region the direct detector is almost always a power-to-voltage or power-to-current converting device. That is, it is a device that puts out a voltage or current in proportion to the incoming power. There are many examples of such devices, but the most popular are field detectors and bolometers. Field detectors, such as Schottky diodes, respond directly to the THz electric field and generate an output current or voltage through a quadratic term in their current-voltage characteristic. Bolometers are composite devices consisting of a THz absorber and a thermistor. The THz absorber is generally isolated thermally from the environment so that the absorbed THz power raises the temperature both of the absorbing layer and an attached thermistor. The thermistor is, by definition, a device that displays a large change of resistance to a small change of temperature. In some bolometers, such as the composite type, the absorber and thermistor are separate elements. In other bolometers, such as the hot electron type, they are integrated into the same device.

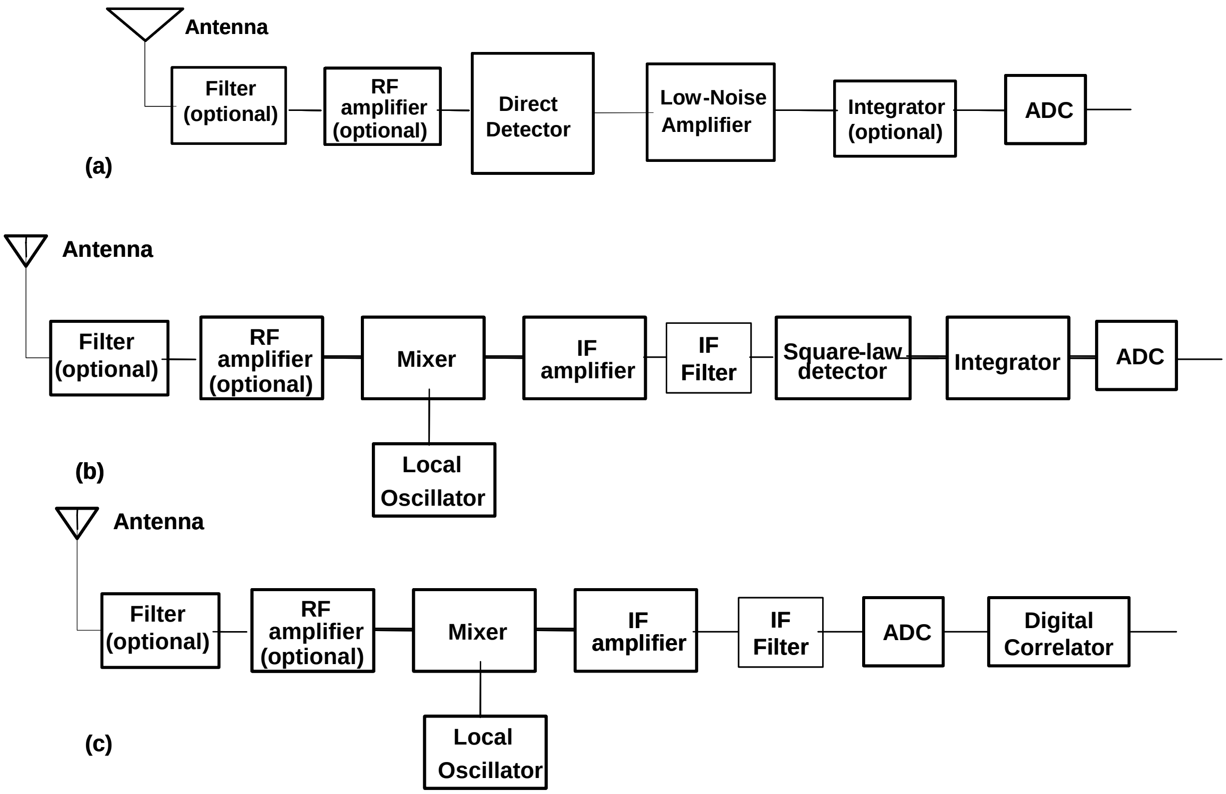

Fig. 1.

A key factor in all direct detectors is spectral bandwidth. As in most analyses of signal processing, we assume here that this is band limited between ν_o and $\nu_o + \Delta\nu$. This can be a real bandwidth defined by a THz bandpass filter, or it might be an approximation to a real spectrum.

B. Pre-Amplified Direct Detection

One of the most successful areas of RF electronics during the past decade has been monolithic microwave integrated circuits (MMICs). By integrating active devices, some passives, and matching circuits on the same semi-insulating substrates, it has become possible to fabricate low-noise amplifiers (LNAs) up to frequencies of 100 GHz and beyond.[1,2] For example, LNAs having a gain of 17 dB, bandwidth of 30 GHz, and a noise figure of 6 dB have been fabricated and tested around 94 GHz. The advantage of an LNA is that if it has adequate gain, it can dominate the noise figure of the following square-law detector, leading to a much lower NEP than can be achieved by direct conversion using the

[1] H. Wang, L. Samoska, T. Gaier, A. Peralta, H. H. Liao, Y. C. Leong, S. Weinreb, Y. C. Chen, M. Nishimoto, and R. Lai, "Power amplifier modules covering 70-113 GHz using MMICs," IEEE Trans. Microwave Theory Tech., vol. 49, pp. 9-16, Jan. 2001.
[2] S. Weinreb, T. Gaier, R. Lai, M. Barsky, Y. C. Leong, and L. Samoska, "High-gain 150-215 GHz MMIC amplifier with integral waveguide transitions," IEEE Microwave Guided Wave Lett., vol. 9, pp. 282-284, July 1999

same square-law detector. As will be shown later, the sensitivity of the preamplified direct receiver can then approach the photon-noise (quantum) limit.

C. Heterodyne

In the heterodyne system of Fig. 1(b), incoming radiation from the target, be it thermal emission or transmitted power from the sensor itself, is combined with power from a local oscillator (LO) on a THz mixer. If the signal and LO frequency are different, there will be a beat-note generated at an intermediate frequency (IF) between the THz signal and baseband. This is called *heterodyne* conversion. It the signal and LO frequency are equal, the beat tone degenerates to dc, and the process is called *homodyne* conversion. Independent of the conversion process, all coherent detectors require a device that can generate an *efficient* conversion of the RF power to the IF band. The most popular mixers are field-type devices having a strong quadratic nonlinearity. Good examples are Schottky diodes, superconductor-insulator-superconductor (SIS) tunnel junctions,[3,4] and superconducting hot-electron bolometers.[5]

Coherent down-conversion has several unique features that distinguish it from direct detection. First, mixing a weak signal with a relatively strong LO effectively amplifies the received signal relative to the receiver noise floor, which can greatly improve the sensitivity compared to direct detection. Second, for the typically weak signals in the THz region, the mixing process is linear. That is, the signal power at the IF frequency is linearly proportional to the signal power at the input. Therefore, the receiver passband can be defined by an IF band pass filter, which is generally much lower in cost and has much higher performance than any THz filter. This feature tends to make coherent receivers the favored approach in applications requiring high spectral resolution, such as molecular spectroscopy.[6] But as will see later, the direct receiver tends to be preferable in wideband applications such as thermal imaging because of its superior spectral bandwidth and simplicity.

A related advantage of heterodyne over direct detection is the preservation of phase information of the received waveform, X. If the IF band is low enough in frequency that analog-to-digital conversion (ADC) can be done, then it is possible to digitize the IF signal and retain phase information assuming that the Nyquist sampling criterion is met. That is, the digitization must be done so that at least two samples are taken per cycle of the highest frequency component in the IF power spectrum. With modern ADC technology, this can be done with high precision (related to the number of bits used in the ADC) and high

[3] A. H. Dayem and R. J. Martin, "Quantum interaction of microwave radiation with tunneling between superconductors," Phys. Rev. Lett., vol. 8, pp. 246-248, Mar. 1962.
[4] G. J. Dolan, T. G. Phillips, and D. P. Woody, "Low noise 115 GHz mixing in superconductor oxide barrier tunnel junctions," Appl. Phys. Lett., vol. 34, pp.347-349, Mar. 1979.
[5] D. E. Prober, "Superconducting terahertz mixer using a transition-edge microbolomoeter," Appl. Phys. Lett., vol. 62, no. 17, pp. 2119-2121, 1993.
[6] J. W. Waters, "Submillimeter-wavelength heterodyne spectroscopy and remote sensing of the upper atmosphere," Proc. IEEE, vol. 80, pp.1679-1701, Nov. 1992.

speed at the same time. For example, it is possible to get ADCs commercially that provide 14-bit accuracy (amplitude range of 16384 decimal, unsigned, or 84 dB) operating at several hundred megasamples per second.

The advantage of maintaining the phase is that one can then carry out the detection process more optimally than by direct detection in terms of signal-to-noise ratio. A straightforward way to achieve optimal detection is with a correlator. And as shown in Fig. 1(c), correlation is best carried out digitally because of the mathematical precision required. In addition, cross correlation can provide other information, such as target range, and clutter information. In any case, the overall process of optimizing the signal-to-noise after detection by processing the phase-coherent waveform optimally is called "matched filtering." We will discuss this much more later on. For now, it is already good to know that optimum performance is possible, but depends on a lot of factors.

D. Pre-Amplified Heterodyne

An intriguing possibility in the millimeter-wave band and lower end of the THz range is a preamplifier feeding a mixer element, as shown optionally in Figs. 1(b) and (c). With recent advances in MMIC solid-state power amplifiers and the possibility of integrating them with high-frequency Schottky mixers monolithically, one can envision a receiver in which an antenna couples radiation to an LNA that, in turn, is coupled to a high-frequency mixer. The mixer then down-converts the radiation to whatever IF band makes sense, be it narrowband for spectroscopic applications, or wider band for thermal imaging applications. Intuitively, one could design the LNA with just enough gain so that the overall receiver sensitivity was not affected significantly by the mixer or following IF electronics. As will be shown later, this provides excellent overall performance if the LNA noise figure is acceptably low. And not surprisingly, it is the same architecture used at lower frequencies in communications and radar receivers alike, perhaps most commonly in the handsets of nearly every mobile telephone made today at PCS wireless frequencies.

The Detection Process

Just as in communications systems, the receiver of an RF sensor must ultimately extract *information* from the received RF power. As we have seen, RF sensors generally utilize different types of information than the time-domain bits of communications systems. There is spatial information in the form of extent and range, there is motional information in the form of Doppler frequency shifts, and there is spectral information in the form of emissive or reflective signatures. But just like communications systems, extracting this information necessarily entails *energy* transfer between the source and the receiver. And energy transfer is always measured as a time-average over the instantaneous power flow, just as occurs in Poynting's theorem.

Mathematically, time averaging is just integration. Electrically, integration in the time domain is equivalent to low-pass filtering in the frequency domain. So given this line of reasoning, the extraction of information is tantamount to low-pass filtering from dc up to some ac frequency dictated by the information bandwidth. In communications systems, this low-frequency range of frequencies is called "baseband". In RF sensors, it is often called the detection band. In either case, it is generally much less in bandwidth than the RF center frequency. And as we will see later, it does not want to be any greater than about one-half of the (Nyquist) sampling frequency required by the information being extracted. The reason for this is related to optimizing the sensor signal-to-noise ratio, as we shall see shortly.

In RF sensors the process of extracting energy from the RF signal is called *detection* – a generic title that sometimes gets confused with the act of electrical rectification. Indeed, rectification such as occurs in Schottky diodes and other nonlinear electrical devices, does convert RF power transfer into a low-frequency (dc) term, usually proportional to the time-averaged power. But detection covers many other possible ways of doing the same thing, two of the most powerful and useful of which are multiplication (e.g., synchronous detection) and *cross-correlation.*

Insight can be gained into all of these detection schemes by analyzing perhaps the simplest imaginable detection scheme, which is square-law detection. A square-law detector is represented electrically by the simple input-output (i.e. transfer function) relation,

$$y = Ax^2$$

It is simple to see how such a transfer function can generate dc information from RF power delivery. It is not so easy to see how this can improve the signal-to-noise ratio, so that becomes a goal of the following analysis. We will derive the effect on the signal-to-noise ratio of a square-law-detector followed by an integrator. We will use three different methods: (1) analog signal processing, (2) digital signal processing, and (3) rigorous probability theory. All three lead to the same result, but provide different insight into the problem and emphasize how important the nonlinear detection process to RF sensors.

A. (Intuitive) Analog Method

A.1. Coherent Signal

We represent the coherent input signal by

$$x(t) = B\cos(\omega t + \phi)$$

which by trigonometry becomes

$$y(t) = (AB^2/2)\,\{1 + \cos[2(\omega t + \phi)]\}$$

So if the integration time of the low-pass filter, t, is >> 1/ω, then its (time-averaged) output becomes

$$\overline{y} = AB^2/2 \equiv \Re\ \overline{P}$$

where $\Re$ is called the responsivity.

A.2. Thermal Signal Input

We assume the thermal input signal is additive white Gaussian noise, band-limited to center frequency ν_0 and bandwidth $\Delta\nu$, as shown in Fig. 2(a). In this case Fourier components at different frequencies are uncorrelated, but any given nonzero component will produce an average output over time by virtue of the quadratic effect of the square law detector. This notion is captured mathematically by

$$\overline{y} = \overline{Ax^2(t)} = A\int_t^{t+\tau} x^2(t)dt \approx A\int_{-\nu_0-\Delta\nu/2}^{-\nu_0+\Delta\nu/2} x^2(\nu)d\nu + A\int_{\nu_0-\Delta\nu/2}^{\nu_0+\Delta\nu/2} x^2(\nu)d\nu$$

where the overbar denotes time averaging, x(ν) is the Fourier transform of x(t), and the last step follows from the Parseval's relation for all Fourier transform pairs. Note that negative frequencies are included here since Gaussian noise has random phase. So there is a second effective passband just like the one shown in Fig. 2(a) but located symmetrically about $\nu = 0$. Neither the time integral nor frequency integral extend to infinity as in Applied Math books since the time averaging process is limited by the integration time τ and the input power spectrum is band-limited to $\Delta\nu$. Since $x^2(\nu)$ is just the power spectral density $S_x(\nu)$, we can write

$$\overline{y} \approx A\int_{-\nu_0-\Delta\nu/2}^{-\nu_0+\Delta\nu/2} S_x(\nu)d\nu + A\int_{\nu_0-\Delta\nu/2}^{\nu_0+\Delta\nu/2} S_x(\nu)d\nu = A < (\Delta x)^2 > 2\cdot\Delta\nu = \Re\cdot\overline{P_{in}}$$

where the second step follows by assuming the power spectrum for x is "white." This is the same result as for a coherent signal. In other words, Gaussian fluctuations deliver an average power through square-law multiplication and integration, just like a coherent signal, but it is only the self-product of individual Fourier components that contribute.

A.3. Noise

Different Fourier components of Gaussian noise at the input to the square law detector produce a zero output after averaging, but still contribute a fluctuation and therefore, output noise. In other words, they create difference-frequency components that should also be zero-mean, *but not necessarily Gaussian.* So we need to calculate the output power spectral density

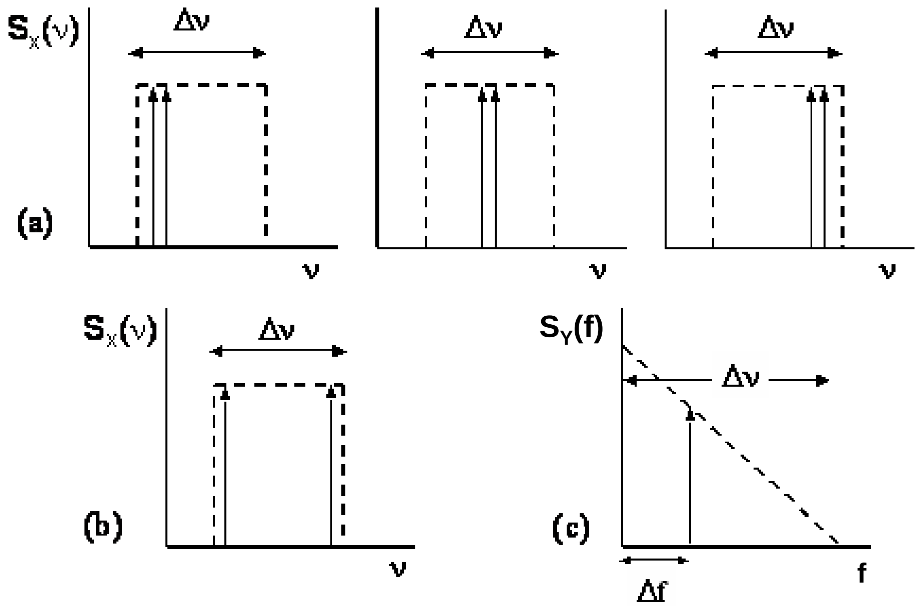

Fig. 2.

$$S_y(f) = \int_t^{t+\tau} y^2(t) e^{-j\omega t} dt = A^2 \int_t^{t+\tau} x^2(t) x^2(t) e^{-j\omega t} dt \, .$$

We now can apply the convolution theorem for Fourier transforms, which states that the transform of a product of Fourier transformed functions is equal to the convolution of the original functions.[7] We apply it by recognizing $x^2(t)$ as the Fourier transform of $S_x(\nu)$, and get the convolution integral

$$S_y(f) = A^2 \int_{-\infty}^{\infty} S_x(\nu) S_x(f-\nu) d\nu$$

$$\approx A^2 \int_{-\nu_0-\Delta\nu/2}^{-\nu_0+\Delta\nu/2} S_x(\nu) S_x(f-\nu) d\nu + A^2 \int_{\nu_0-\Delta\nu/2}^{\nu_0+\Delta\nu/2} S_x(\nu) S_x(f-\nu) d\nu$$

This expression can be understood qualitatively from Fig. 2(a) and (b) which show only the positive-frequency portion and take advantage of the symmetry about $\nu = 0$ by showing the product $S_x(\nu)S_x(\nu\text{-}f)$. If we look at small difference frequencies as in 1(a), there are a maximal number of input Fourier components available, proportional to the input RF bandwidth $\Delta\nu$. But for large difference frequencies, the number of available input Fourier components drops until we reach the minimal condition depicted in Fig. 2(b) where just the extremes of the input spectrum are effective. The resulting output spectrum is given by

$$S_y(f) \approx 2A^2 (S_x)^2 \Delta\nu (1 - f/\Delta\nu)$$

[7] See, for example, "Mathematial Methods of Physics," G. Arfken (Academic, 1970), 2nd Ed., Sec. 15.5.

where S_x is the input power spectral density at any frequency in the passband. Note that as plotted in Fig. 2(c) this is a distinctly-non-white triangular spectrum centered at zero frequency - *our first clue that the output noise statistics are not Gaussian*. But the input noise is "white," so we can estimate it as

$$(S_x)^2 \approx (\Re^2/A^2)<(\Delta P_{in})^2>/<(\Delta\nu)^2>$$

Substitution into the previous expression yields

$$S_y \approx 2\Re^2 (\Delta P_{in})^2 (1 - f/\Delta\nu)/\Delta\nu$$

which is the best form to calculate the quantity that we really want – the output signal-to-noise ratio.

A.4. Output Signal-to-Noise Ratio

Whether Gaussian or not, a good measure of the signal-to-noise ratio, and the one we have adopted exclusively to this point, is the ratio of the average power to power deviation at any point in the receiver. At the output of the *integrator*, this is given by

$$SNR_{out} = \frac{\overline{y^2}}{<(\Delta y)^2>}$$

where $$<(\Delta y)^2> = \int_0^\infty S_y(f)df \approx 2\Re^2(\Delta P_{in})^2 \int_0^{\Delta\nu} \frac{1 - f/\Delta\nu}{\Delta\nu} df$$

Under most conditions, we will want to integrate much longer than $1/\Delta\nu$ (a practice sometimes called "bandwidth narrowing"), which corresponds to limiting the above integral to a range 0 to $f_{max} = \Delta f << \Delta\nu$. Under this condition, the rather slowly varying triangular spectrum can be treated like a constant at the peak value. We then get the result,

$$<(\Delta y)^2> \approx 2\Re^2(\Delta P_{in})^2 \int_0^{\Delta f} \frac{df}{\Delta\nu} \approx 2\Re^2(\Delta P_{in})^2 \frac{\Delta f}{\Delta\nu}$$

The output SNR is then approximated by

$$SNR_{out} = \frac{\overline{y^2}}{<(\Delta y)^2>} \approx \frac{(\Re P_{in})^2}{2\Re^2 <(\Delta P_{in})^2> \Delta f/\Delta\nu} = SNR_{in}^2 \cdot \frac{\Delta\nu}{2\Delta f}$$

showing clearly that the output SNR is improved in proportion to the "bandwidth narrowing" and also as the input SNR squared. As we shall see later, the same thing happens with the passage of a coherent signal embedded in AWGN through a correlator, in which case a factor proportional to $\Delta\nu/\Delta f$ is called the *"processing gain."* But the above derivation shows that the SNR improvement can even happen for AWGN alone when the average power is considered the "signal". This is the case for most passive radiometers collecting thermal radiation in any RF band.

B. (Intuitive) Digital Method

It is interesting to consider the same combination of square-law detector and low-pass filter as above but from a digital sampling perspective. We assume the signal is coherent and is sampled in n_s successive pulses so that the output signal power $S_{out} = (n_s S_{in})^2$. In the same process, the output noise will increase but slower with n_s since each successive pulse is assumed uncorrelated from the rest. It is reasonable to assume that the output noise power $N_{out} = n_s(N_{in})^2$. Hence we can write,

$$SNR_{out} = \frac{(n_s S_{in})^2}{n_s (N_{in})^2} = n_s \cdot SNR_{in}^2$$

Of course the maximum number of samples that can be made in t_s is limited by sampling theory to

$$n_s \approx \Delta\nu \, t_s$$

And for the network at hand, the sampling time is related to the low-pass filter bandwidth by the Nyquist condition

$$t_s = 1/(2\Delta f)$$

So we get once again,

$$SNR_{out} = SNR_{in}^2 \frac{\Delta\nu}{2\Delta f}$$

C. Exact Method (based on probability theory)

We can derive the effect of a square-law detector on the SNR exactly using methods of probability theory. We assume that the noise at the input is AWGN having zero mean, so can be represented by the normalized pdf

$$p_1(x) = (2\pi\sigma^2)^{-1/2} \exp(-x^2/2\sigma^2) \qquad (1)$$

where σ is the standard deviation and σ^2 is the variance. If zero-mean AWGN is put through the square-law detector, the output pdf associated with electrical variable y is found from the transfer relation $y = Ax^2$ and (1) by the general relationship between a pdf and a cumulative distribution function P

$$p_2(y) = (d/dy)P(y)$$

The form of P is given by inspection of the square-law transfer curve and realization that p(x) assumes both negative and positive values of x, but p(y) must only allow for positive values of y:

$$P(y) = \int_{-x(y)}^{+x(y)} p_1(x)dx = \int_{-(y/A)^{1/2}}^{(y/A)^{1/2}} p_1(x)dx \tag{2}$$

$$p_2(y) \equiv \frac{dP}{dy} = \frac{d}{dy}\int_{-(y/A)^{1/2}}^{(y/A)^{1/2}} p_1(x)dx$$

$$= p_1(y/A)^{1/2}[d(y/A)^{1/2})/dy] - p_1(-(y/A)^{1/2})[d(-y/A)^{1/2}/dy] \tag{3}$$

where the last step follows from Leibniz' rule of calculus.[8] Combining (1) and (3) we get

$$p_2(y) = \frac{\exp(-y/2A\sigma^2)}{\sqrt{2\pi\sigma^2 yA}} \tag{4}$$

which is called the "chi-squared" density function plotted in Fig. 3 for A = 1 and σ^2 = 1, 10, and 100. It is a simple exercise to show that it has all the usual properties of a pdf, such as normalization:

$$\int_0^\infty p_2(y)dy = 1$$

One peculiarity is that it has an integrable singularity at y = 0. Physically, this means that low amplitude fluctuations are much more likely than high amplitude ones, which is sometimes called the "smoothing" effect of square-law detectors. Interestingly, for the AWGN input, it yields mean and mean-square values

$$<y> = \int_0^\infty y p_2(y)dy = A\sigma^2 \text{ and } <y^2> = \int_0^\infty y^2 p_2(y)dy = 3A^2\sigma^4$$

So the variance is given by

$$<(\Delta y)^2> = <y^2> - <y>^2 = 2A^2\sigma^4$$

Our goal in this exercise is to rigorously predict the possible improvement in signal-to-noise ratio had by the combination of a square-law detector and a low-pass filter. So now that we know the statistics of the electrical fluctuations after the square-law detector, it is convenient to convert this knowledge into the time domain. The tool for doing this is the autocorrelation function,

$$R_x(t_1,t_2) \equiv <x_1 x_2>$$

[8] A rule often not stated in calculus books but very useful nonetheless:

$$\frac{\partial}{\partial u}\int_{a(u)}^{b(u)} f(u,w)dw = \int_{a(u)}^{b(u)} \frac{\partial f}{\partial u}dw + f(b)\frac{\partial b}{\partial u} - f(a)\frac{\partial a}{\partial u}$$

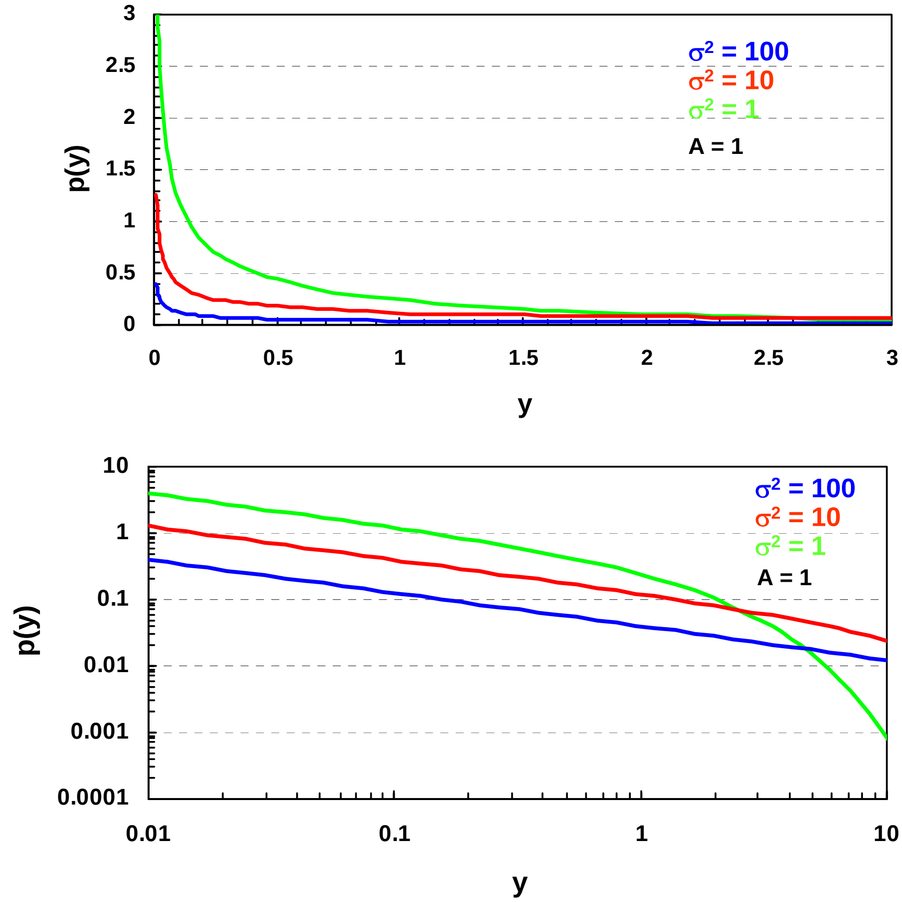

Fig. 3

Where x_1 and x_2 are the possible values of the same electrical variable x at different times t_1 and t_2, respectively. In generally, this transformation from ensemble averages to correlation functions is a difficult one, but becomes tractable in the special case of stationarity. Roughly speaking, this means that the pdf and all ensemble averages (i.e., expectation values) derived from it are all independent of time. It also means that the autocorrelation function is independent of any reference point in time, so that

$$R_x(t_1,t_2) = R_x(t, t + \tau) = R_x(\tau)$$

In other words, the autocorrelation function depends only on a time difference τ, not the absolute time.

After the square law detector, we can write for the autocorrelation function

$$R_y(\tau) = \langle y_1 y_2 \rangle = \langle A(x_1)^2 a(x_2)^2 \rangle = A^2 \langle (x_1)^2 (x_2)^2 \rangle$$

It is well known from probability theory that the ensemble average of the product of the product of two independent random variables is equal to the product of the their averages. Similarly, the mean value of the product $(x_1)^2(x_2)^2$ can be considered as the mean value of $x_1x_1x_2x_2$ which can be shown by permutation

$$< x_1x_1x_2x_2> = < x_1\, x_1><x_2x_2>+ < x_1x_2><x_1x_2> + < x_1x_2><x_1x_2>$$
$$= <(x_1)^2><(x_2)^2> + 2< x_1x_2>^2$$
$$= <(x)^2>^2 + 2[R_x(\tau)]^2$$

Therefore, we can write for the autocorrelation function at the detector output

$$R_y(\tau) = A^2<(x)^2>^2 + 2A^2[R_x(\tau)]^2$$

But if the input pdf $p_1(x)$ is given by (1) (i.e., AWGN), then $<(x)^2> = <(\Delta x)^2> + (<x>)^2 = <(\Delta x)^2> = \sigma^2$. And we get

$$R_y(\tau) = A^2\sigma^4 + 2A^2[R_x(\tau)]^2 \qquad (5)$$

Given this relationship between autocorrelation functions at the input and output, a very useful tool is the Wiener-Kinchine (WK) theorem discussed previously which in general relates the autocorrelation function to the power spectral density by a Fourier transform:

$$S(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} R_x(\tau) \exp(-j\omega\tau) d\tau \qquad (6)$$

or by the inverse transform

$$R_x(\tau) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} S(\omega) \exp(j\omega\tau) d\tau \qquad (7)$$

Upon Fourier transforming (5) using (6), we can write

$$S_y(\omega) = A^2\sigma^4\, \delta(\omega)+ \frac{2A^2}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S_x(\omega') \exp(j\omega'\tau) R_x(\tau) \exp(-j\omega\tau) d\omega' d\tau \qquad (8)$$

Where $\delta(\omega)$ is the Dirac delta function and we have expanded one of the Rx terms in (5) using (7) and the dummy frequency ω' . Separating independent variables, we can write (8) as

$$S_y(\omega) = A^2\sigma^4\, \delta(\omega)+ \frac{2A^2}{2\pi} \int_{-\infty}^{\infty} \{S_x(\omega') \int_{-\infty}^{\infty} R_x(\tau) \exp[-j(\omega-\omega')\tau] d\tau\} d\omega' \qquad (9)$$

$$= A^2\sigma^4\, \delta(\omega)+ \frac{2A^2}{2\pi} \int_{-\infty}^{\infty} \{S_x(\omega') S_x(\omega-\omega') d\omega'$$

or in terms of linear frequency ν,

$$S_y(\nu) = A^2\sigma^4\, \delta(\nu)+ 2A^2 \int_{-\infty}^{\infty} \{S_x(\nu') S_x(\nu-\nu') d\nu' \qquad (10)$$

This is starting to look familiar. The first term is the dc component, and the second is the same convolution of input power spectra that we wrote down earlier in the intuitive analog derivation. In other words, the first term is the signal, and the second term is the noise.

Now that we are in the frequency domain, we can make practical assumptions about the bandwidth at input and output. For the AWGN assumed to this point, S_X is "white" and band-limited over a range from ν_0 to $\nu_0 + \Delta\nu$. Mathematically this can be expressed as the "window" function,

$$S_x(\nu) = S_0\, \theta(\nu - \nu_0)\, \theta(\nu_0 + \Delta\nu - \nu) \tag{11}$$

where θ is the unit step function. Substituting (11) into (10), we get the convolution of two window functions, which intuitively and mathematically leads to a triangle function:

$$\int_{-\infty}^{\infty} \{S_x(\nu')S_x(\nu-\nu')d\nu' = 2S_0^2\Delta\nu(1-\nu/\Delta\nu) \tag{12}$$

The final result of (10) is then

$$S_y(\nu) = A^2\sigma^4\, \delta(\nu) + 2S_0^2\Delta\nu(1-\nu/\Delta\nu) \tag{13}$$

where the noise spectrum after the square-law detector is triangular ! But because the input noise is assumed to be "white", we can re-write it as

$$(S_0)^2 = R^2\, (<(\Delta P)^2>/(\Delta\nu)^2$$

and the output spectrum becomes

$$S_y(\nu) = A^2\sigma^4\, \delta(\nu) + 2\, R^2\, \{(<(\Delta P)^2>/(\Delta\nu)\}\{1 - \nu/\Delta\nu\} \tag{14}$$

At this point we make the same observation as before, namely, that we can take great advantage of the non-white output spectrum by following it with a (linear) low-pass filter whose bandwidth, Δf, is much less than Δν. In this case, the output electrical fluctuations are just

$$<(\Delta y)^2> = \int_0^{\Delta f} S_y(\nu)d\nu \approx 2\, R^2 <(\Delta P)^2>\cdot\Delta f/\Delta\nu$$

Since the input is zero-mean AWGN, $\sigma^2 = <(\Delta x)^2> = <x^2>$. So the ratio of the 1st and 2nd terms in (14) becomes

$$SNR_{out} = <y>^2/<(\Delta y)^2> = A^2<x^2>^2/2\, \Re^2 <(\Delta P)^2>\cdot\Delta f/\Delta\nu$$

$$= \Re^2<P>^2/\{2\, \Re^2 <(\Delta P)^2>\cdot\Delta f/\Delta\nu \}= (\Delta\nu/2\Delta f)\cdot(SNR_{in})^2$$

This is the same result as derived above by the (intuitive) analog and digital-sampling techniques, now justified using rigorous probability theory across the time and frequency domains.