Optimum Pre-Detection Signal Processing
(the "matched filter" concept)

Maximum Signal-to-Noise Ratio (Intuitive Derivation)

As we have seen, THz sensors share many general concepts and characteristics of all RF sensors, particularly those related to electromagnetic coupling and propagation, and signal processing. One of the most important concepts is the existence of signal processing that optimizes the signal-to-noise ratio. The implementation of such signal processing is called "matched filtering". It goes back to the early days of radar (post-WWII), and has been applied universally in radar and communications systems ever since. It pre-supposes we know the waveform of the received signal, meaning that we know its amplitude and phase vs time. In the case of passive detection (i.e., radiometry) where the signal is random noise, a similar concept applies, but is not as powerful. So "matched filtering" is generally limited to active systems, and is best introduced by the following intuitive explanation.

- Detection in the presence of noise has limits imposed by physics (e.g., thermodynamics). To understand these limits, suppose we are attempting to detect the RF pulse plotted in Fig. 1 of waveform $x(t) = A_X \sin(\omega t + \phi)$, carrier frequency $\omega = 2\pi\nu$, and pulse duration T_P .
- Assume that we are trying to detect this pulse in the presence of AWGN having power spectral density S_P such that $<(\Delta P)^2> = S_P\, \Delta\nu$. In this case, the RF signal-to-noise ratio is

$$\left(SNR\right)_{\max} = \frac{\bar{P}}{\sqrt{<\left(\Delta P\right)^2>}} = \frac{\bar{P}\cdot T_p}{S_P\Delta\nu\cdot T_p} \approx \frac{U_p}{S_P\Delta\nu\cdot T_p} \quad (1)$$

Where U_P is the electrical pulse energy. If we assume that the pulse is sampled consistent with the Nyquist condition, then the sample rate f_S should be matched to the pulse width

$$f_s = 1/T_P, \quad (2)$$

and should be twice the twice the RF instantaneous bandwidth

$$f_S = 2\Delta\nu \ . \quad (3)$$

Substitution of these last two into the SNR expression yields

$$\left(SNR\right)_{\max} \approx \frac{2U_p}{S_P} \equiv \frac{2U_p}{N_0} \quad (4)$$

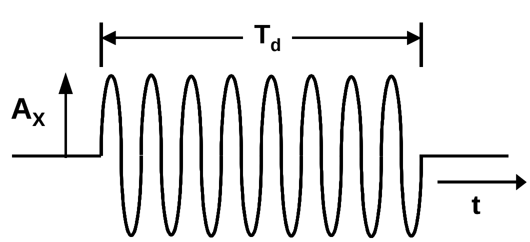

Fig. 1.

where N_0 is another way of writing the power spectral density (following the convention in communications theory). This is a factor of two higher than might be expected intuitively because it assumes implicitly that we have precise knowledge of the phase and amplitude of a signal, not just one or the other.

Maximum Signal-to-Noise Ratio (Rigorous Derivation)

A. Signal

As in the analysis of RF "detection" in the previous section, there are intuitive ways to derive important theorems in RF sensors, and there are more rigorous ways. This Professor believes students should see both because the intuitive ways are the ones easiest to remember and extend (at least qualitatively) to other concepts, but the rigorous ways are more precise and necessary in real systems.

On the subject of optimum signal-to-noise ratio, the rigorous technique is based on methods of linear signal processing. We start by assuming the RF "front-end" of the receiver is a linear network having a known system transfer function $H(\omega)$ and impulse response function $h(t)$ related by

$$\mathrm{H}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} h(t) e^{-j\omega t} dt \tag{5}$$

For a given signal $x(t)$ at the input or output, waveform is related to the Fourier amplitudes $X(\omega)$ of the power spectral density by the Fourier transform pair

$$X(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \quad \text{and} \tag{6}$$

Linear systems theory teaches us that the input and output Fourier amplitudes are related by

$$X_{out}(\omega) = X_{in}(\omega)\cdot H(\omega) \tag{7}$$

Combining the last two

$$x_{out}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X_{out}(\omega) e^{j\omega t} d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} X_{in}(\omega) H(\omega) e^{j\omega t} d\omega \tag{8}$$

But by the convolution theorem, the Fourier transform of the product equals the convolution of the Fourier transform pairs:

$$x_{out}(t) = \int_{-\infty}^{\infty} x_{in}(t-\tau) h(\tau) d\tau \tag{9}$$

In general t can be defined for any time $-\infty < t < \infty$. Here, we want to specify t to a sampling time t_S related to the sampling frequency by $f_S = 1/t_S$. Hence,

$$x_{out}(t_S) = \int_{0}^{t_S} x_{in}(t_S-\tau) h(\tau) d\tau \tag{10}$$

B. Noise

The assumption of a linear network and AWGN noise allows us to write immediately,

$$<(\Delta x)^2> = S_X \Delta v \equiv S_X \Delta v_N \equiv S_X \cdot \frac{1}{2}\int_{-\infty}^{\infty} |H(\omega)|^2 dv \qquad (11)$$

where the factor ½ assumes H (ω) is two sided (i.e., negative frequencies are allowable). and Δv_N is called the noise bandwidth. But by Parseval's theorem, Fourier transform pairs are related by

$$\int_{-\infty}^{\infty} |H(\omega)|^2 dv = \int_0^{\infty} |h(\tau)|^2 d\tau \qquad (12)$$

so that

$$<(\Delta x)^2> = S_x \cdot \frac{1}{2}\int_{-\infty}^{\infty} |h(\tau)|^2 d\tau \qquad (13)$$

C. Signal-to-Noise Ratio

Combining the results of the previous two sections, we have

$$SNR = \frac{|x_{out}(t_S)|^2}{S_x \Delta v_N} = \frac{\left|\int_0^{t_S} x_{in}(t_S - \tau) h(\tau) d\tau\right|^2}{\frac{1}{2} S_x \int_0^{\infty} |h(\tau)|^2 d\tau} \qquad (14)$$

But by the famous Schwarz inequality, we can write

$$\left|\int_0^{t_S} x_{in}(t_S - \tau) h(\tau) d\tau\right|^2 \le \int_0^{t_S} |x_{in}(t_S - \tau)|^2 d\tau \cdot \int_0^{t_S} |h(\tau)|^2 d\tau \qquad (15)$$

Equality occurs if and only if

$$h(\tau) = C \cdot x_{in}(t_S - \tau) \qquad (16)$$

This leads to a very useful bound on the SNR of

$$SNR \le \frac{\int_0^{t_S} |x_{in}(t_S - \tau)|^2 d\tau \int_0^{t_S} |h(\tau)|^2 d\tau}{\frac{1}{2} S_x \int_0^{\infty} |h(\tau)|^2 d\tau} = \frac{\int_0^{t_S} |x_{in}(t_S - \tau)|^2 d\tau}{\frac{1}{2} S_X} \qquad (17)$$

By recognizing the relation $U_P = \int_0^{t_S} \left[x_{in}(t_S - \tau)\right]^2 d\tau$, we get the elegant result (again)

$$SNR \le \frac{2U_X}{S_X} \qquad (18)$$

D. Matched-filter condition

We now recognize (16), $h(t) = C \cdot x_{in}(t_S - t)$, as the definition of the optimum system impulse response function. As written, it is just the time reversal of the input waveform. From linear response theory, we can write the equivalent transfer function

$$H(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} h(t)e^{-j\omega t}dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} C \cdot x_{in}(t_S - t)e^{-j\omega t}dt \tag{19}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} C \cdot x_{in}(t_S - t)e^{-j\omega t}e^{j\omega(t_S - t_S)}dt = C \cdot X^*(\omega)e^{-j\omega t_S}$$

where the last step is recognized as phase conjugation. Eqns (16) and (19) are regarded as equivalent descriptions of optimum linear RF signal processing called "matched filtering." Substitution of the "matched-filter" *h* (*t*) back in to the system transfer convolution leads to

$$x_{out}(t) = \int_0^{\infty} x_{in}(t-\tau)h(\tau)d\tau = C \cdot \int_0^{\infty} x_{in}(t-\tau)x_{in}(t_S - \tau)d\tau \tag{20}$$

This expression suggests a powerful signal processing strategy: multiply the input signal by a time-shifted replica of itself. This is the basis for *correlative* signal processing, now commonly used in radar and communications receivers alike.

Implementation of the Matched-Filter Concept

The conditions for a matched filter in any THz or RF receiver take the form

$$h(\tau) = C \cdot x_{in}(t_S - \tau) \tag{21}$$

and

$$H(\omega) = C \cdot X^*(\omega)e^{-j\omega t_S} \tag{22}$$

$$x_{out}(t) = \int_{-\infty}^{\infty} x_{in}(t-\tau)h(\tau)d\tau = C \cdot \int_{-\infty}^{\infty} x_{in}(t-\tau)x_{in}(t_S - \tau)d\tau \tag{23}$$

where x_{in} is the input electrical signal, $X(\omega)$ is its Fourier spectral component at frequency ω, and the lower limit of the intergration is set to zero to ensure that causality is satisfied – i.e., the output can not occur in time before the input arrives. For students not well-versed in communications or radar theory, these conditions might be esoteric enough to warrant demonstration by a simple example.

Square Sinusoidal Input Pulse

The square sinusoidal input pulse of Fig. 2 can be represented mathematically by the relation

$$x_{in} = A\cos(\omega_0 t + \phi)\, \theta(t + t_p/2)\, \theta(t_p/2 - t) \tag{24}$$

Without loss of generality we can set $\phi = 0$ and obtain

$$X\left(\omega\right)=\frac{1}{2\pi}\int_{-t_p/2}^{t_p/2}A\cos(\omega_0 t)e^{-j\omega t}dt=\frac{1}{2\pi}\int_{-t_p/2}^{t_p/2}\frac{A}{2}(e^{j\omega_0 t}+e^{-j\omega_0 t})\cdot e^{-j\omega t}dt \tag{25}$$

$$X\left(\omega\right)=\frac{A}{2\pi}\left[\frac{2j\sin[(\omega_0-\omega)(t_p/2)]}{j(\omega_0-\omega)}+\frac{2j\sin[(-\omega_0-\omega)(t_p/2)]}{j(-\omega_0-\omega)}\right]$$

$$X\left(\omega\right)=\frac{At_p}{2\pi}\left(\mathrm{sinc}[(\omega_0-\omega)t_p/2]+\mathrm{sinc}[(\omega_0+\omega)t_p/2]\right)$$

From (22) above we can thus write for the matched-filter transfer function

$$H\left(\omega\right)=\frac{CAt_p}{2\pi}\left(\mathrm{sinc}[(\omega-\omega_0)t_p/2]+\mathrm{sinc}[(\omega_0+\omega)t_p/2]\right)e^{-j\omega t_S} \tag{26}$$

a two-sided expression in which negative frequencies are kept as part of the Fourier analysis. One interesting aspect of this power transfer function is that it displays zero crossings along with local minima and maxima on both sides of the peak frequencies. Realizing that any simple receiver will tend to produce $|H(\omega)|^2 > 1$ at all frequencies across a "passband", the matched-filter transfer function will be very difficult to achieve in practice.

The output waveform of the matched filter is given by

$$x_{out}\left(t\right)=C\int_{-\infty}^{\infty}x_{in}\left(t-\tau\right)x_{in}\left(t_S-\tau\right)d\tau \tag{27}$$

Since the sampling time t_s is arbitrary, we can set it to zero without loss of generality. Also, we carefully chose the input waveform to be symmetrical about t = 0, so we can write

$$x_{out}\left(t\right)=C\int_{-\infty}^{\infty}x_{in}\left(t-\tau\right)x_{in}\left(-\tau\right)d\tau=C\int_{-\infty}^{\infty}x_{in}\left(\tau-t\right)x_{in}\left(\tau\right)d\tau \tag{28}$$

Using the specific sinusoidal pulse waveform, we get

$$x_{out}\left(t\right)=CA\int_{-\infty}^{\infty}\cos[\omega_0\left(\tau-t\right)]\cos[\omega_0\left(\tau\right)]\theta(\tau-t+t_p/2)\cdot\theta(t_p/2-\tau+t)\cdot\theta(\tau+t_p/2)\cdot\theta(t_p/2-\tau)d\tau \tag{29}$$

This is not as complicated as if first looks, the four unit step functions being required by the truncated form of the input pulse and by the convolutional nature of the output pulse. After some thought it becomes apparent that the product of the four step functions, Π, has a form that depends on the relationship between the chosen t and the variable of integration τ. For $0 < t < t_p$, we get

$$\Pi = 1 \quad \text{if} \quad -t_p/2 + t < \tau < t_p/2\ ; = 0 \text{ otherwise} \tag{30}$$

For $-t_p < t < 0$ we get

$$\Pi = -t_p/2 < \tau < t + t_p/2\ ;\ = 0 \text{ otherwise} \tag{31}$$

This breaks up x_{out} into two separate integrals. For $0 < t < t_p$

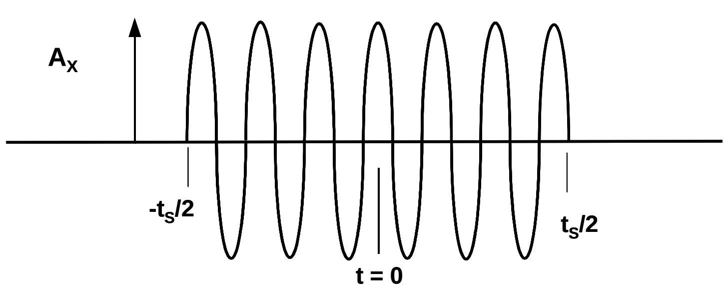

Fig. 1.

$$x_{out}(t) = C \cdot A^2 \int_{t-t_p/2}^{t_p/2} \cos[\omega_0(\tau - t)]\cos[\omega_0(\tau)]d\tau \tag{32}$$

From trigonometry we recall the identity, cos(a)cos(b) = (1/2){cos(a-b) + cos(a+b)}
Application of this to the first integral yields

$$x_{out}(t) = (C/2)A^2 \left\{ \tau\cos(-\omega_0 t) + \frac{\sin[2\omega_0\tau - \omega_0 t]}{2\omega_0} \right\}_{t-t_p/2}^{t_p/2} \tag{33}$$

$$x_{out}(t) = (C/2)A^2 \left\{ (t_p - t)\cos(\omega_0 t) + \frac{\sin[\omega_0 t_p - \omega_0 t]}{2\omega_0} - \frac{\sin[2\omega_0 t - \omega_0 t_p - \omega_0 t]}{2\omega_0} \right\}$$

$$x_{out}(t) = (C/2)A^2 \left\{ (t_p - t)\cos(\omega_0 t) + \frac{\sin[\omega_0 t_p - \omega_0 t]}{\omega_0} \right\}$$

For $\omega_0 >> t_p$ and except where $t \approx t_p$, the second term can be neglected compared to the first and we get

$$x_{out}(t) \approx (C/2)A^2(t_p - t)\cos(\omega_0 t) \quad ; \quad 0 < t < t_p \tag{34}$$

For $-t_p < t < 0$ we get much a very similar analysis

$$x_{out}(t) = C \cdot A^2 \int_{-t_p/2}^{t+t_p/2} \cos[\omega_0(\tau - t)]\cos[\omega_0(\tau)]d\tau \tag{35}$$

$$x_{out}(t) = (C/2)A^2 \left\{ \tau\cos(-\omega_0 t) + \frac{\sin[2\omega_0\tau - \omega_0 t]}{2\omega_0} \right\}_{-t_p/2}^{t+t_p/2} \tag{36}$$

$$x_{out}(t) = (C/2)A^2 \left\{ (t + t_p)\cos(\omega_0 t) - \frac{\sin[-\omega_0 t_p - \omega_0 t]}{2\omega_0} + \frac{\sin[2\omega_0 t + \omega_0 t_p - \omega_0 t]}{2\omega_0} \right\}$$

$$x_{out}(t) = (C/2)\left\{ (t + t_p)\cos(\omega_0 t) - \frac{\sin[-\omega_0 t_p - \omega_0 t]}{2\omega_0} + \frac{\sin[2\omega_0 t + \omega_0 t_p - \omega_0 t]}{2\omega_0} \right\} = \left\{ (t_p - t)\cos(\omega_0 t) + \frac{\sin[\omega_0 t_p + \omega_0 t]}{\omega_0} \right\}$$

$$x_{out}(t) \approx (C/2)(t + t_p)\cos(\omega_0 t) \quad ; \quad -t_p < t < 0. \tag{37}$$

The combination of these can be written

$$x_{out}(t) = (C/2)(t_p - |t|)\cos(\omega_0 t) \tag{38}$$

is nothing more than a cosine with a triangular envelope ! This is plotted in Fig. 3.

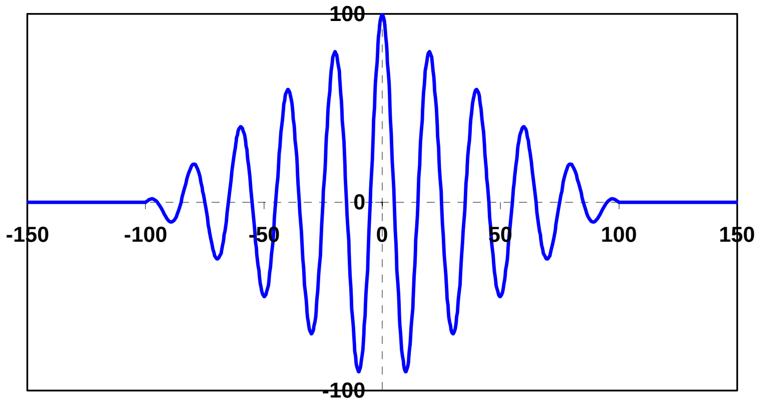

Fig. 3. Output of matched filter for an input signal waveform having the pulsed sinusoid of Fig. 2.

Given AWGN noise at the input, we can now calculate the SNR at the output of the matched filter receiver:

$$\mathrm{SNR} = \frac{|x(t)|^2}{S_x \cdot \Delta\nu} \tag{39}$$

The maximum value occurs at t = 0

$$\mathrm{SNR}_{\max} = = \frac{(C^2 t_p / 2)^2}{S_x \cdot \Delta\nu} \tag{40}$$

By integrating over the triangular pulse and setting $\Delta\nu = 1/t_p$ it is simple to show (see HW problem)

$$\mathrm{SNR}_{\max} = = \frac{2U_p}{S_P} \tag{41}$$

where U_p is the total electrical energy in the pulse.

Physical Interpretation of Matched Filter

This shows the physical effect of the matched filter. A matched filter enhances the "detectability" of a single pulse by raising the instantaneous SNR to the level $2U_p/S_p$ *at the time of sampling* (the peak in the above triangle function for x_{out} occurred at t = 0 since we chose t_s = 0 arbitrarily at the beginning of the calculation). It accomplishes this by tailoring the transfer function to provide the highest system gain at frequencies where the input signal is rich in spectrum, and suppress the gain elsewhere. Some might consider this function just common sense. But a matched filter is doing more than tailoring $H(\omega)$. It is also *tailoring the phase response (or group delay)* through the system so that energy associated with the various spectral components arrives at the output *simultaneously*. This explains the peaked behavior of the output pulses from matched filters of all sorts.