Statistical Detection Theory

We have frequently used the signal-to-noise ratio and its derivatives (e.g., NEP) to characterize the performance of RF and THz systems. Although a very important figure-of-merit in any system, it does not tell us what the system operator ultimately needs to know, which is "what is the *likelihood* that the system will successfully detect a given target in a given setting, and what is the *likelihood* that it will *miss* the given target, or perhaps show the telltale signs of detection when the target is not there at all ?" These questions support the notion that detection in the presence of noise and other (false) targets is always a statistical process. So to properly predict system performance, first a statistical analysis must be made of all the possible detection outcomes, and the various conditions of the system that could have created those outcomes.[1,2,3] It is inherently a reverse statistical process. That is, we know the outcome, but do not know with certainty the conditions that created it. As such, it is guided by a specific sub-field of probability theory that started with Bayes. It requires an accurate model for the system, including the signal-to-noise ratio as the key metric, and a thorough parameterization of the known target and other possible targets or environmental conditions that can create a *detect* for the outcome. This leads to modeling detection as a binary decision in the presence of a "target" return signal buried in additive white Gaussian noise (AWGN) and occasionally accompanied by signal from false targets or "clutter" in the local environment.

Receivers are usually linear up to the detection device, and so linear super-position is assumed to hold. The output signal from the receiver can then be written:

$$X(t) = X_{sig}(t) + X_{noise}(t) \tag{1}$$

where X_{sig} is the expected signal (deterministic) and X_{noise} is due to electronic and thermal noise (random process) as we have discussed in detail earlier. The noise is assumed to simply add to the signal as opposed to multiply against it, non-linearly scale it, etc. Further the above formula exemplifies that now the output is a random process.

[1] R. N. Mcdonough, A. D. Whalen, *Detection of Signals in Noise*, Academic Press, New York, 1995
[2] J. V. DiFranco and W.L. Rubin, *Radar Detection*, SciTech Publishing, Raleigh, NC, 2004
[3] W.B. Davenport and W.L. Root, *An Introduction to the Theory of Random Signals and Noise*, Wiley-IEEE Press, 1987.

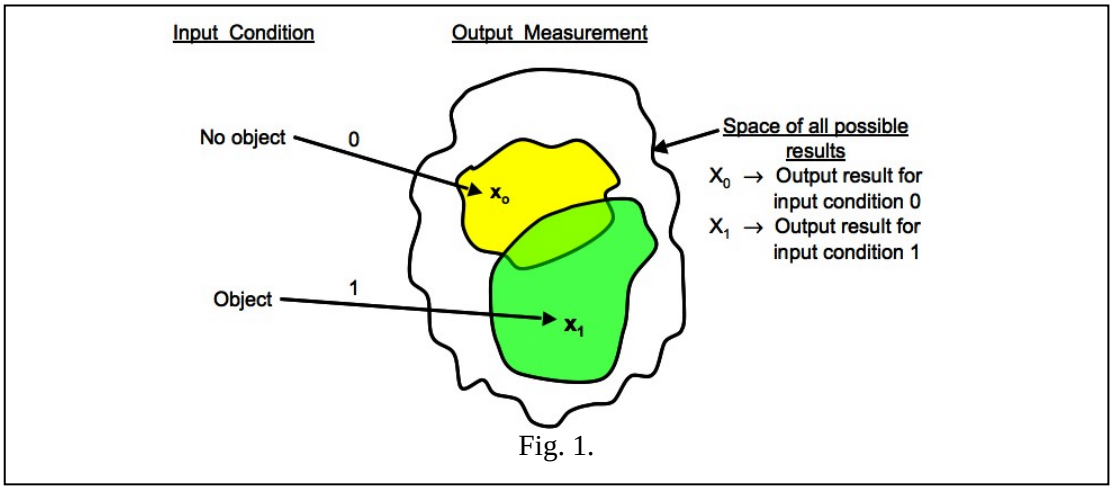

Fig. 1.

And so the SNR as defined is not the final metric, although as will be shown a good indicator and a pivotal measurement to assess the performance of a receiver.

Receiver Analysis: the Bayesian criteria

Since the output signal is a random process, we must apply statistical detection theory, which is relatively simple for RF sensors if they are *fully* binary. That is, there are assumed to be two possible input conditions: (a) target absent ("0" condition), or (b) target present ("1" condition). And there are two possible outputs, now represented by possible *ranges* of the electrical variable X. If a target is absent, we have the range X0, and if the target is present we have the *range* X_1. This mapping between target presence and output range is shown graphically in Fig. 1. And it makes sense for active RF sensors of all types, active (e.g., radar) and passive (e.g., radiometers).

The primary function of the receiver is to determine which of the two input conditions exists based only on measurement of the output signal X_{out} over its entire range. The complication of the receiver function, and the need for a statistical detection theory, is that the ranges represented by X_0 and X_1 generally overlap, creating an ambiguity in the mapping. In other words, there are four possible outcomes of the fully binary target situation, as displayed graphically in Fig. 2: (a) target present and X in unambiguous region of X_1, which is called a *true positive*; (b) target absent and X in unambiguous region of X_0, which is called a *true negative;* (c) target absent, X in ambiguous region of X_0, and mistakenly interpreted as X_1, which is called a false positive (or in radar colloquial, a "false alarm"), and (d) target present, X in ambiguous region of X_1, and mistakenly interpreted as X_0, which is called a false negative (or in radar

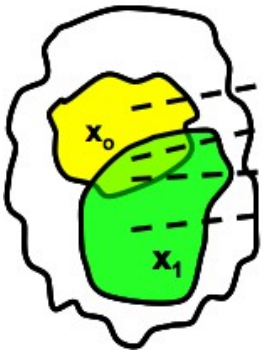

Fig. 2.

colloquial, a "miss"). The goal of the statistical detection theory must be to optimize the accuracy of the receiver by designing it to maximize the occurrence of the first two "correct" outcomes, and minimize the occurrence of the second two "incorrect" ones.

To proceed further, we assume that probability density functions are known to describe the X_0 and X_1 regions of X_{out} space, $p_0(X)$ and $p_1(X)$, respectively. In general, these regions are not bounded as conveniently shown in Fig. 1, but are smeared out to create a continuous overlap over all the X_{out} space. This forces us to seek further information to optimize the receiver accuracy, and led the mathematician Bayes to realize that information must be provided about the input condition and the output result. Ideally, one should know the *a priori* (i.e., beforehand) probabilities π_0 and π_1 for the target being absent or present. And one should also know the *cost*, or loss factor, of each decision made. Specifically, L_{11} is the loss function for a true positive, L_{00} is the loss function for a true negative, L_{10} is the loss function for a false negative, and L_{01} is the loss function for a false positive.[4] Bayes then showed that an optimum receiver function could be made in terms of the "liklihood ratio" L of the *a posteriori* (afterwards) densities p_0 and p_1,

[4] In statistical terminology, L_{01} corresponds to type-I error and L_{10} a type-II error

$$L(X) \equiv \frac{p_0(X)}{p_1(X)} = \frac{\pi_1(L_{10} - L_{11})}{\pi_0(L_{01} - L_{00})} \tag{2}$$

This led to the optimization relations, called the Bayes criteria

$$\text{(a) If } L(X) > \frac{\pi_1(L_{10} - L_{11})}{\pi_0(L_{01} - L_{00})}\text{, then condition 0 exists} \tag{3}$$

$$\text{(b) If } L(X) < \frac{\pi_1(L_{10} - L_{11})}{\pi_0(L_{01} - L_{00})}\text{, then condition 1 exists} \tag{4}$$

It is very important for the reader to realize that these criteria are really a *recipe*. In other words, if the receiver produces an output X at a given decision time, the Bayes criteria tell us how to make the most accurate decision about the input condition that created this X. This can not be a decision that just reduces the likelihood of the two "incorrect" decisions but, rather, one that also enhances the likelihood of the two "correct" decisions. This is a good intuitive way of understanding why the key quantity in the Bayesian technique, and most branches of optimizing probability theory for that matter, is a likelihood *ratio*.

A special case: the Neyman Pearson criterion

In communications systems, one usually has a good understanding of the a priori probabilities and the loss factors by virtue of knowing how the transmitter sends information (e.g., the modulation format) and what the information is used for. But in RF sensing, there are often no *a priori* statistics available on the occurrences of condition 0 vs. condition 1. But one can still define the *a posteriori* probabilities p_0 and p_1 and link them to the hypotheses H_0 and H_1 that the target is absent or present, respectively. An interesting question is then whether or not any optimization of the receiver accuracy can be realized.

In seminal work by Neyman and Pearson,[5] it was found that optimization could still be achieved if the decision space of X (see Fig. 1) was also divided up, the simplest

[5] J. Neyman and E. S. Pearson, Phil. Trans. Royal Soc. London, A231, pp. 289-337 (1933).

division being a single one with respect to a boundary or "threshold" value X_t.[6] The presence of a threshold greatly simplifies the receiver decision-making process. First, it makes the process binary in the sense that any output from the receiver exceeding X_t then supports the H_1 hypothesis about the target, and any output less than X_t supports the H_0 hypothesis. Second, it enables the definition of unique probability distribution functions with respect to $p_0(x)$, $p_1(x)$ and the threshold. In one-to-one correspondence with the previously discussed four possible outcomes to the binary target situation, we have a probability of true positive ("detection") P_d given by

$$P_d \equiv \int_{X_t}^{X_{\max}} p_1(X)dX \ , \tag{5}$$

a probability of true negative P_n given by

$$P_n \equiv \int_{X_{\min}}^{X_t} p_0(X)dX \ , \tag{6}$$

a probability of false positive ("false alarm") P_{fa} given by

$$P_{fa} \equiv \int_{X_t}^{X_{\max}} p_0(X)dX \ , \tag{7}$$

and a probability of false negative ("miss") P_{miss} given by

$$P_{miss} \equiv \int_{X_{\min}}^{X_t} p_1(X)dX \ . \tag{8}$$

Often X_{min} is set to $-\infty$ and X_{max} is set to $+\infty$, but one has to examine carefully the electrical variable and the circuit conditions. For example, all electronic devices are saturable or critical at some X_{max} value, so that close proximity of X_t to X_{max} can significantly impact the probabilities define above. And being legitimate pdfs, p_0 and p_1 are each normalized over the domain $X_{min} < X < X_{max}$ so that we have the important constraints

$$P_d + P_{miss} = 1 \qquad \text{and} \qquad P_n + P_{fa} = 1 \tag{9}$$

[6] This is also a very practical division since threshold circuits are very easy to construct from electronic devices with gain – transistors today, and vacuum tubes in the days of Neyman and Pearson

Given the threshold and its useful implications, the Neyman-Pearson optimization occurs with respect to one or the other of the two types of "incorrect" decisions, now represented by P_{miss} and P_{fa}. If one constrains P_{fa} to be less than or equal to a certain value, $P_{fa,max}$, then there exists a unique threshold $X_{t,0}$ that establishes the maximum possible P_d, $P_{d,max}$ for that P_{fa} constraint. Mathematically, we have

$$P_{fa,\max} \equiv \int_{X_{t,0}}^{X_{\max}} p_0(X)dX \quad \text{and} \quad P_{d,\max} \equiv \int_{X_{t,0}}^{X_{\max}} p_1(X)dX \tag{10}$$

Some might consider this just common sense on the grounds that the threshold should always be set "as low as possible" to maximize P_d while keeping below the maximum tolerable P_{fa}. But there is a deeper meaning. Because $P_{miss} = 1 - P_d$, the maximization of P_d also implies an automatic minimization of P_{miss} subject to the same constraint on P_{fa}. Hence, we have a minimization of "incorrect" decisions. And we are back to the Bayesian thinking once again. Namely, it is the liklihood of *all* incorrect decisions that is affected by any optimization procedure.

To further elucidate the Neyman-Pearson optimization, we can examine the Bayes criteria under special conditions. Assuming the lack of any knowledge of the a priori probabilities, we set them equal and get perfect cancellation in Eqn (2) above. We can also assume that the "cost" or loss caused by false decisions is much greater than the "cost" or loss caused by true ones. This allows us to assume $L_{10} > L_{11}$, and $L_{01} > L_{00}$ in Eqn (2), yielding

$$L(X) \equiv \frac{p_0(X)}{p_1(X)} = \frac{L_{10}}{L_{01}} = \eta \tag{11}$$

where η is a real positive number. The counterparts to the Bayes criteria then become

$$\text{(a) If } L(X) > \eta \text{, then condition 0 exists} \tag{12}$$

$$\text{(b) If } L(X) < \eta \text{, then condition 1 exists} \tag{13}$$

These are usually stated in the reciprocal (Neyman-Pearson) form

$$\text{(a) If } p_1 < p_0/\eta \equiv p_0\,\lambda \text{, then condition 0 exists} \tag{14}$$

$$\text{(b) If } p_1 > p_0/\eta \equiv p_0\lambda \text{, then condition 1 exists} \tag{15}$$

where λ is also a positive integer.

Eqns (14) and (15) can be related uniquely to a threshold X_t when the likelihood ratio p_1/p_0 is a monotonically increasing function of X.[7]. We can then write the implicit definition

$$\left.\frac{p_1}{p_0}\right|_{X=X_t} = \lambda \tag{16}$$

This places the Neyman-Pearson conditions on similar ground as the more general Bayesian criteria, namely, on the basis of a likelihood ratio. But it rarely used in practice simply because of the more practical value of the probability distribution approach outlined above.

Other Criterion for Optimized Reception

For the receiver optimization, there are two techniques in addition to the Bayes and Neyman-Pearson discussed above: the maximum a posteriori probability (MAP) criterion, and the minimax criterion. Each criterion is used in different instances depending on the type and amount of information available at the time of detection. As we have seen, the Bayes criterion uses the loss or cost functions and works to minimize the average cost for each particular decision. The loss function establishes *a priori* the relative cost of for the various outcomes in a decision process. For calculating the cost function the probabilities for each hypothesis are also needed *a priori*.

The MAP (maximum a posteriori probability) criterion chooses the outcome that has the higher probability of having been in effect, given the particular output measurement. The MAP criterion is the same as the Bayes criterion when the cost function for each decision outcome are equal to each other. The minimax criterion uses the cost functions but has no information about the probabilities for the hypotheses a priori. In other words, the minimax criterion is the Bayes criterion when each hypothesis is equally probable. And the Neyman-Pearson criterion uses neither the cost function or prior probabilities. By not making use of any information on each input hypothesis *a*

[7] usually true in RF sensors and communications systems, but not necessarily true in statistical problems in general

priori, the Neyman-Pearson criterion is the most flexible and therefore the most widely used with radar systems.

Implementing the Neyman-Pearson Criterion

As discussed earlier, the Neyman-Pearson conditions are tantamount to maximizing P_d (or minimizing P_{miss}) while maintaining P_{fa} to the "tolerable" level specified by the threshold X_t. It is very helpful to demonstrate this through example.

Example: Coherent detection of signal in Gaussian noise.

In a radar detection problem, let us assume that the two possible target conditions H_0 and H_1 yield the two deterministic output states of electrical variable X shown in Fig. 3.

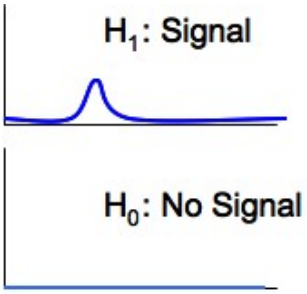

Fig. 3

After some time averaging, X_{sig} will thus always produce zero, or some repeatable "mean" value, X_m. But because we do not know *a priori* the probability for the presence of the target, then X_m is itself a (binary) discrete random variable, $X_{sig} = X_m$ or 0. The added noise results in a the total output electrical random variable $X(t) = X_{sig} + X_{noise}$,. But because X_{sig} is binary, we can write:

$$\begin{aligned} H_0: &\quad X(t) = X_{noise}(t) \\ H_1: &\quad X(t) = X_m + X_{noise} \end{aligned} \tag{17}$$

If the noise is AWGN with variance of σ^2, we can describe these random variable states by two Gaussian distributed random variables, one with a zero mean, and the other with a mean of X_m:

$$p_0 = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left[-\frac{(X)^2}{2\sigma^2} \right] \tag{18}$$

$$p_1 = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left[-\frac{(X - X_m)^2}{2\sigma^2} \right] \tag{19}$$

These *density* functions are plotted in Fig. 4 with the overlap responsible for "incorrect" decisions that will inevitably be made. According to the Neyman-Pearson conditions and the arbitrary threshold shown in Fig. 4, the two primary *distribution* functions are:

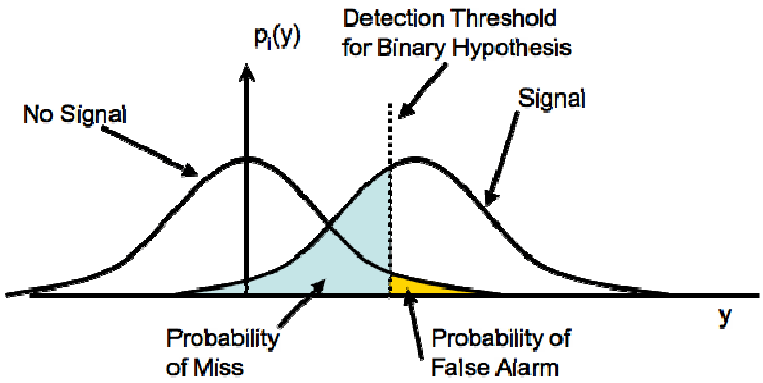

Fig. 4

$$P_d = \frac{1}{\sqrt{2\pi\sigma_n^2}} \int_{X_t}^{\infty} \exp\left[-\frac{(X - X_m)^2}{2\sigma_n^2} \right] dX$$

$$P_{fa} = \frac{1}{\sqrt{2\pi\sigma_n^2}} \int_{X_t}^{\infty} \exp\left[-\frac{(X)^2}{2\sigma_n^2} \right] dX \tag{20}$$

From Eqns (18) and (19), the Neyman-Pearson liklihood ratio test is given as:

$$\lambda(X) = \frac{p_1(X)}{p_0(X)} = \exp\left[-\frac{(X - X_m)^2 - X^2}{2\sigma_n^2} \right] > \lambda_0 \tag{21}$$

By taking the logarithm and re-arranging, we can get a simple threshold test for the measured values $X > X_t$. For example, by setting $\lambda = 1$, we define a condition of "equal errors" (i.e., $P_{fa} = P_{miss}$). Eqn (21) becomes $(X - X_m)^2 = X^2$ which has the obvious solution $X = X_m/2$. This is manifest from Fig. 4. If the X_t is set at this point, then P_d will be maximized with respect to the given P_{fa}. But $P_d = 1 - P_{miss}$, so this threshold also

defines the special condition where $P_d = 1 - P_{fa}$. In other words, P_d also reaches the maximum allowed by probability theory.

For other values of λ it is usually easier to determine the threshold through the distribution functions. We do this by simplifying Eqns (20) through the conventional definitions of the error function and complementary error function

$$\begin{aligned} erf(z) &\equiv \frac{2}{\sqrt{\pi}} \int_0^z \exp(-y^2) dy \\ erfc(z) &\equiv 1 - erf(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty \exp(-y^2) dy \end{aligned} \tag{22}$$

such that each spans over the range 0 to 1 for $0 < z < \infty$, and satisfy erf(z) = 1 – erfc(z). Both are plotted in Fig. 5 where their strong nonlinearity is apparent. The erf(y) function saturates very abruptly just above y = 1.0, and the erfc(y) function falls precipitously from ~1.0 to 0 in the same region.

By substituting $y = X/(2\sigma^2)^{1/2}$ and $z = (X-X_m)/(2\sigma^2)^{1/2}$, we get

$$\begin{aligned} P_d &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{(X_t - X_m)/(\sqrt{2\sigma^2})}^\infty \exp\left[-z^2\right] dz \cdot \sqrt{2\sigma^2} \\ &= \frac{1}{2} erfc\left[(X_t - X_m)/(\sqrt{2\sigma^2}\right] \quad if \;\; X_t > X_m \\ &= \frac{1}{2}\left[1 + erf\,(X_m - X_t)/(\sqrt{2\sigma^2}\right] \quad if \;\; X_t < X_m \end{aligned} \tag{23}$$

$$P_{fa} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{(X_t)/(\sqrt{2\sigma^2})}^\infty \exp\left[-y^2\right] dz \cdot \sqrt{2\sigma^2} = \frac{1}{2} erfc\left[X_t /(\sqrt{2\sigma^2}\right] \tag{24}$$

Note that because we are only considering $X_t > 0$, the maximum of P_{fa} is ½ as one would expect from Fig. 4, but the maximum of P_d can approach 1.0 if $X_m >> X_t$.

By application of the conditions $P_d + P_{miss} = 1$ and $P_n + P_{fa} = 1$, we get trivially,

$$\begin{aligned} P_{miss} &= 1 - P_d = \frac{1}{2}\left[1 + erf\{(X_t - X_m)/(\sqrt{2\sigma^2}\}\right] \quad if \;\; X_t > X_m \\ &= \frac{1}{2}\left[erfc\{(X_m - X_t)/(\sqrt{2\sigma^2}\}\right] \quad if \;\; X_t < X_m \end{aligned} \tag{25}$$

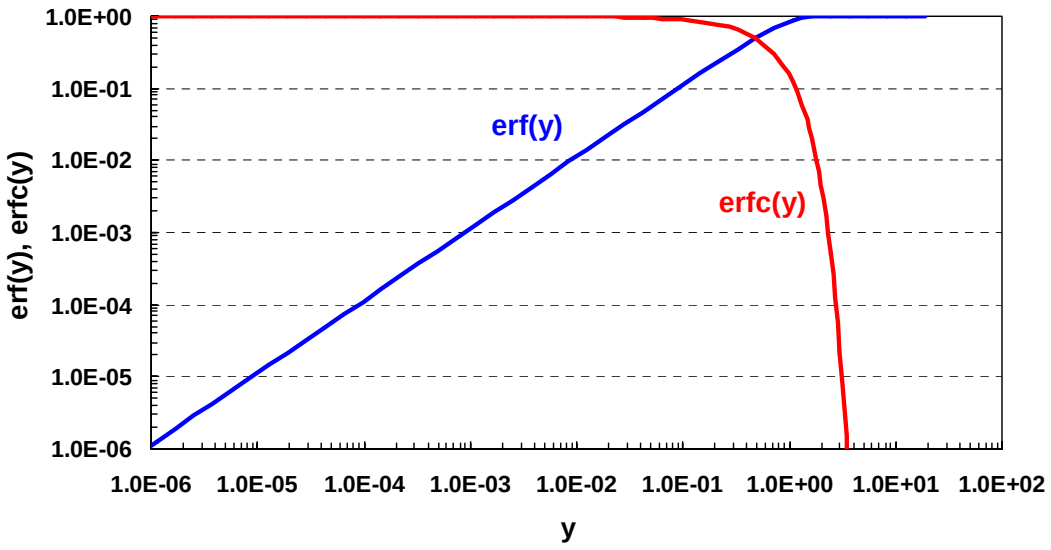

Fig. 5

$$P_n = 1 - P_{fa} = 1 - \frac{1}{2} erfc\left[X_t / (\sqrt{2\sigma^2} \right] = \frac{1}{2}\left[1 + erf\left(X_t / (\sqrt{2\sigma^2}) \right) \right] \qquad (26)$$

From the unique expression for P_{fa}, we can find X_t by inversion

$$X_t = \sqrt{2\sigma^2} \cdot erfc^{-1}(2P_{fa}) \qquad (27)$$

Just 20 years ago, the inverse erfc function was difficult to determine, except by look-up tables. Today it is just a simple library function in Excel, Matlab, or many other PC-based tools !

By dividing both sides by X_m and squaring, we get an interesting relation

$$(X_t / X_m)^2 = (2\sigma^2 / X_m^2) \cdot [erfc^{-1}(2P_{fa})]^2 = (SNR)^{-1}[erfc^{-1}(2P_{fa})]^2 \qquad (28)$$

$$\equiv P_t/P_m$$

where P_t is the threshold power and P_m is mean signal power. This tells us a lot about how to set the threshold to maintain a "tolerable" P_{fa}. And not surprisingly, the proportionality constant is the inverse SNR. In other words, as the SNR goes up, the threshold can be set lower and lower to achieve a given P_{fa}.

Receiver Operating Characteristics

Because of the strong nonlinearity in the probability distributions Eqns (23)-(27) that characterize a receiver according to statistical detection theory, it is helpful to plot them parametrically. For historical and technical reasons related to the operation of radar systems, the most common plot is P_d vs P_{fa} . Because of the relationship SNR = P_m/σ^2 for AWGN and P_m, only two parameters are needed to fully represent Eqns (23) – (27), which are conventionally chosen to be SNR and P_t. Fig. 6 shows such a plot with SNR ranging from 0.1 to 20 (-10 dB to +13 dB), and P_t ranging over four orders of magnitude. There are many interesting aspects of this plot, so many in fact that it gained the title "receiver operating characteristics" or ROC for short. One of most important aspects is the magnitude of SNR needed to achieve "tolerable" P_{fa} and "acceptable" P_d. Because of the anxiety and/or grave consequences caused by false alarms, most radar systems generally require $P_{fa} < 10^{-6}$, many military radar systems requiring much better than that. On the other hand, the "misses" associated with a low P_d ($P_{miss} = 1 - P_d$) can also have grave consequences (e.g., air-traffic control radar). So most radar systems require $P_d >> 0.9$. Inspection of the ROC diagram in Fig. 6 shows quickly that to get into this "ballpark" of operation we must have SNR ≥ 20, no matter what the threshold. This is in stark contrast to the SNR = 1 condition commonly used to determine range and NEP, as we did earlier in the course.

Another important aspect of the ROC diagram is how quickly the P_{fa} drops with increasing SNR at a given P_d, or similarly how quickly P_d rises with increasing SNR at a given P_{fa}. For example Fig. 6 shows that at $P_d = 0.1$, P_{fa} drops ~4 orders of magnitude from 10^{-2} to 10^{-6} as the SNR increases from just 0.5 to 6. And at $P_{fa} = 10^{-6}$, the P_d rises ~90 fold from 0.01 to 0.9 as the SNR increases from about 3 to 20. These are very practical and general results of statistical detection theory and a key reason why radar systems tend to transmit so much power and to put so much emphasis on the signal processing. Every factor-of-two improvement in the SNR makes a big difference !

On computing the ROC plot

The plotting of the ROC curves in Fig. 6 is not so obvious and deserves a quick discussion. It is impossible to write an analytic dependence of P_d on P_{fa} from Eqns (23)

and (24) above. P_d depends on the SNR, which equals $(X_m)^2/(2\sigma^2)$ by definition. But both depend on the threshold to noise ratio TNR $\equiv (X_t)^2(2\sigma^2)$. So as often occurs in numerical analysis, if we vary the TNR continuously over the expected range and vary the SNR parametrically, we can calculate Pd and P_{fa} simultaneously and then "collate" to create the ROC curves.

$$\begin{aligned} P_d &= \frac{1}{2} erfc\left[\sqrt{TNR} - \sqrt{SNR}\right] \quad if \;\; TNR > SNR \\ &= \frac{1}{2}\left[1 + erf\left(\sqrt{SNR} - \sqrt{TNR}\right)\right] \quad if \;\; SNR > TNR \end{aligned} \tag{29}$$

$$P_{fa} = \frac{1}{2} erfc\left[\sqrt{TNR}\right] \tag{30}$$

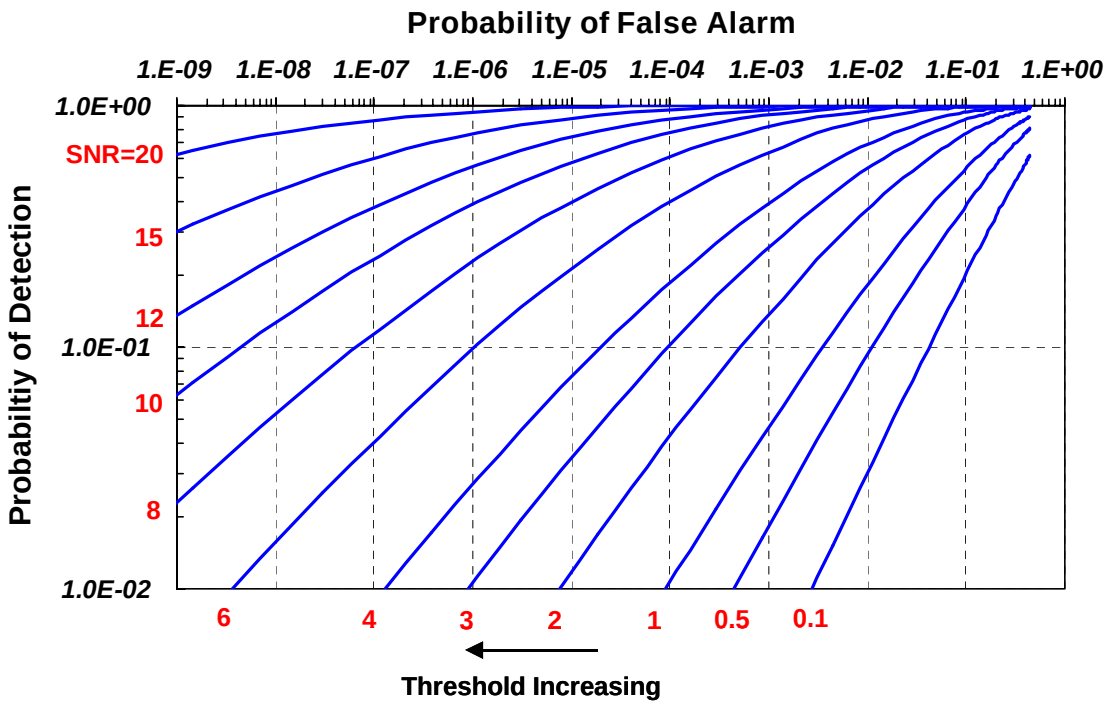

Fig. 6. Universal ROC Curve for Coherent Detection