THE RELATIVE OPERATING CHARACTERISTIC APPROACH TO PROBABILITIES OF DETECTION AND FALSE ALARM

(or, Applying Signal Detection Theory to NDE)

Derek Sturges
GE Aircraft Engines
Engine Support Operation
Cincinnati, Ohio

Tutorial Workshop on Probability of Detection
Spring 1993 ASNT Conference

SIGNAL DETECTION THEORY

- o **RADIO-WAVE DETECTION THEORY**
  - **Application of statistical decision/detection theory**
    - **Probabilistic descriptions of uncertainty about signals and noise**
  - **Gives relationship between Probability of Detection (POD), Probability of False Alarm (PFA), and Signal-to-Noise ratio (S/N)**
  - **Founded on work by North (RCA, 1943), Uhlenbeck (MIT, 1943), Rice (Bell Telephone, 1944), Marcum & Swerling (Rand, 1947), and others**

- o **RADAR/SONAR APPLICATIONS (1943 et seq.)**
  - **Measurement of performance in detection of aircraft or ships**
  - **Improvement in receiver circuitry to optimize signal detection**
  - **Development of autodetection hardware to eliminate the human observer**

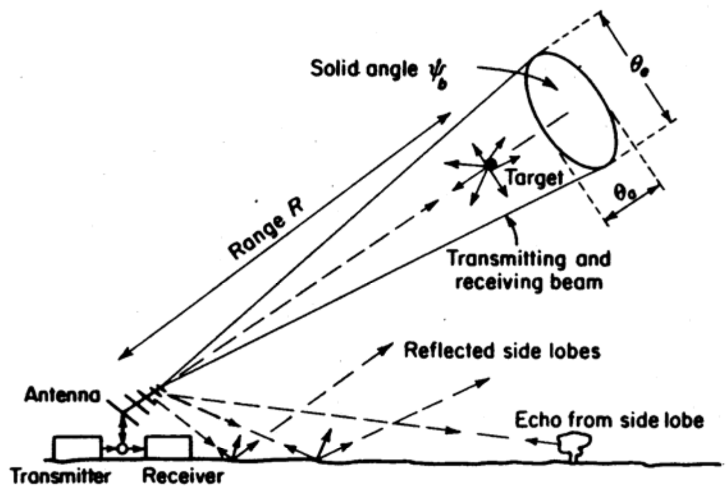
SIGNAL DETECTION THEORY

- o **BASIS**
  - Measurement of signal & noise distributions, or of POD and PFA

- o **ADVANTAGES**
  - A rationale for evaluating and optimizing detection system performance
    - A framework for examining "co-variation" of detection and false alarm

- o **LIMITATIONS**
  - Results depend on the specific signal and noise distributions
    - May be qualitatively - but not quantitatively - transferable
  - Threshold optimization depends on the specific detector (circuitry)
    - Caution: most radar analyses assume time-varying Gaussian noise
    - Caution: radar analyses are based on detector types uncommon in NDE

SIGNAL DETECTION THEORY FOR RADAR SYSTEMS



Typical radar-target geometrical relationships

- **DETECTION GOAL**
  - **Sensing the presence of the sought-after target in the presence of competing indications which arise from background radiation, undesired echoes, or noise generated in the receiver**

SIGNAL DETECTION THEORY FOR RADAR SYSTEMS

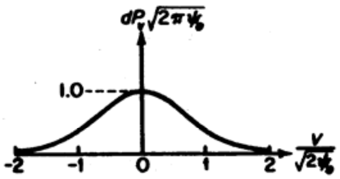
- o **RADAR "SIGNALS"**
  - A delayed series of echo-pulses of radio-frequency energy
    - Affected by beam pattern, target reflectivity & transmission medium
    - Doppler-shifted by radial velocity of target

- o **RADAR "NOISE"**
  - Thermal noise
    - Human-origin, atmospheric, solar or galactic noise
    - Waveguide, duplexer, or receiver noise
  - Non-thermal noise
    - Target signals from side-lobe echo-paths
    - Ground "clutter" - steady (e.g. buildings) or time-varying (e.g. waves)
    - "Jamming" (i.e. deliberate interference)

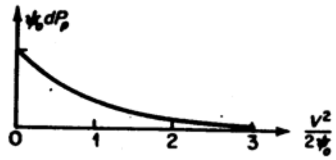
SIGNAL DETECTION THEORY FOR RADAR SYSTEMS

- o **INITIAL GOAL**
  - **Optimizing receiver design for detecting target-reflected signals**
- o **INITIAL APPROACH (D.O.North)**
  - **Analyze probability that a threshold would be exceeded by:**
    - **A single sample of random noise**
    - **A single sample of sinusoidal signal added to random noise**
  - **Assumptions:**
    - **Random noise confined to narrow frequency band by (i-f) filtering**
    - **Envelope detection used (i.e. "video" signals)**

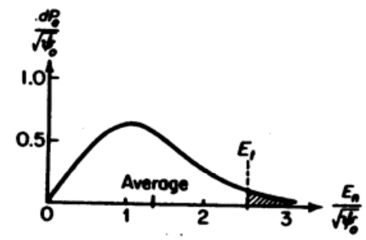
SIGNAL DETECTION THEORY FOR RADAR SYSTEMS



Gaussian distribution of i-f noise



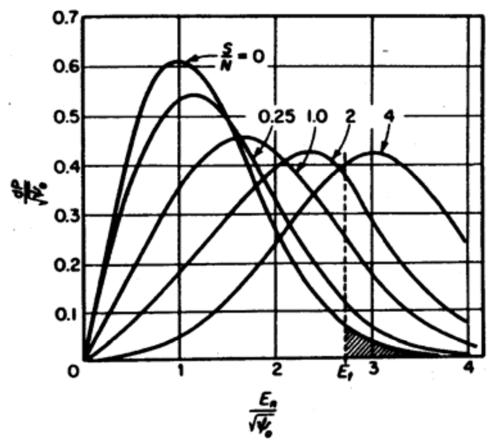
Exponential distribution of i-f noise power



Rayleigh distribution of detected envelope

- **PROBABILITY DISTRIBUTIONS OF NOISE**
  - **Distributions corresponding to North's assumptions**
  - **Probability (PFA) that a single sample of detected noise exceeds threshold level E_t is given by the area under the curve to the right of E_t**

SIGNAL DETECTION THEORY FOR RADAR SYSTEMS



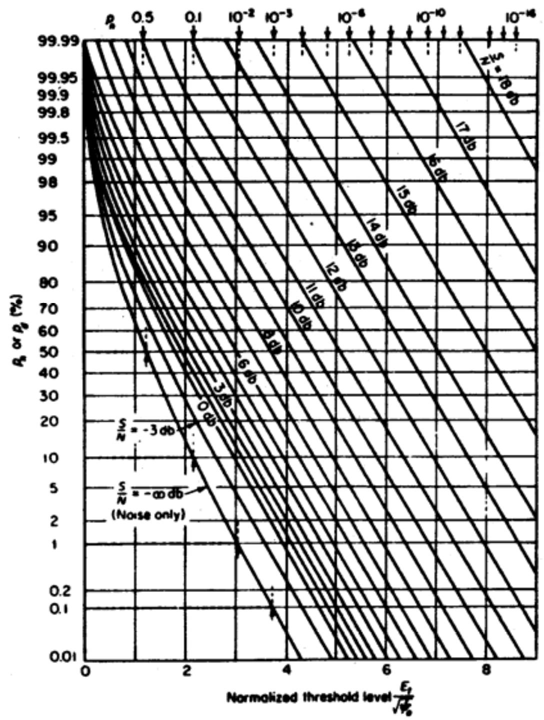
Probability distributions of the envelope of signal-plus-noise

- **DETECTED OUTPUT WITH SIGNAL PRESENT**

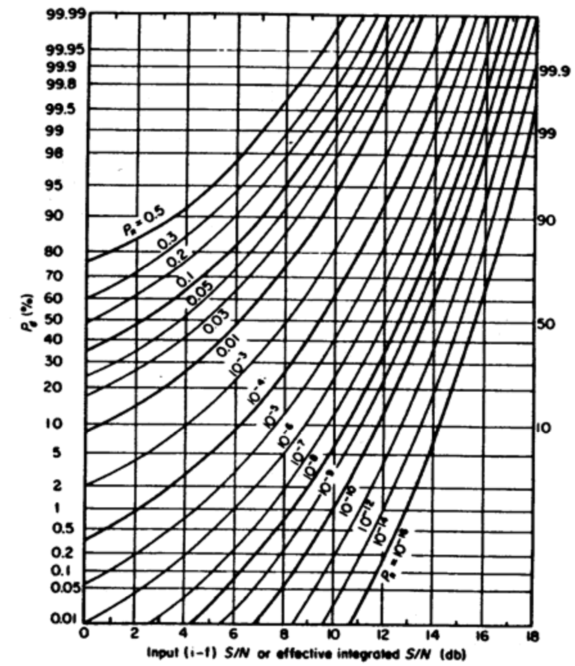
  - **Envelope detection: narrow-band Gaussian noise, with power N**
    - **Steady sinusoidal signal, with received signal power S**

  - **Single-pulse detection probability (POD) is the area to the right of E_t under the curve for each value of S/N**

SIGNAL DETECTION THEORY FOR RADAR SYSTEMS

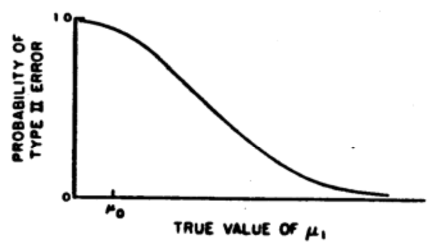


Probability of crossing threshold
(numerical analysis by Rice)

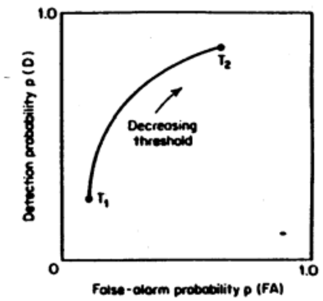


Probability of detection vs. S/N
(single-pulse POD for radar systems)

SIGNAL DETECTION THEORY



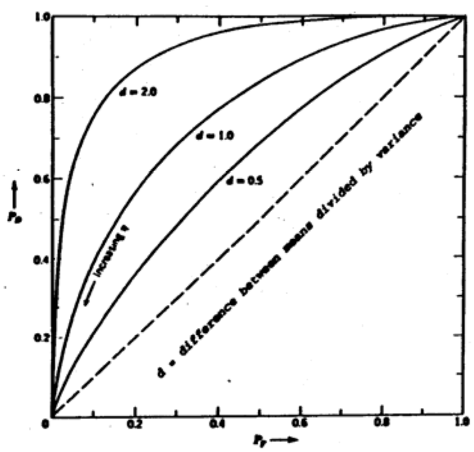
Example of an Operating Characteristic

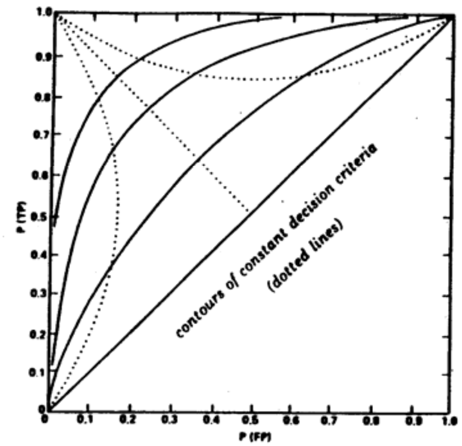


Example of a Relative Operating Characteristic

- **OPERATING CHARACTERISTICS**
  - **Statisticians' designation for specific types of probability plots**
    - **The evaluation component of detection theory**
  - **ROC: a graphic way of comparing two operating characteristics**
    - **For Radar applications, known as Receiver Operating Characteristics**
    - **Later renamed Relative Operating Characteristics**
  - **Merged with aspects of information theory and systems analysis**
    - **Founded on work of Peterson & Birdsall (University of Michigan, 1953), J.A.Swets (Bolt Beranek & Newman, 1954), and others**

SIGNAL DETECTION THEORY FOR RADAR SYSTEMS

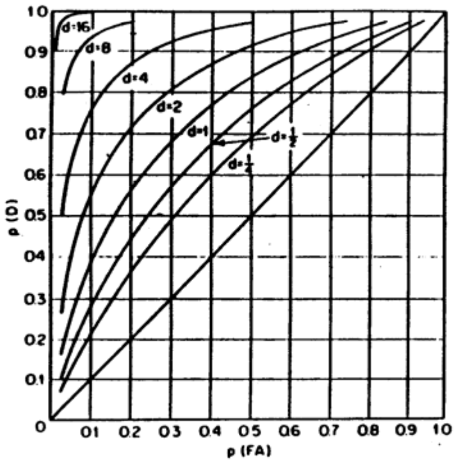


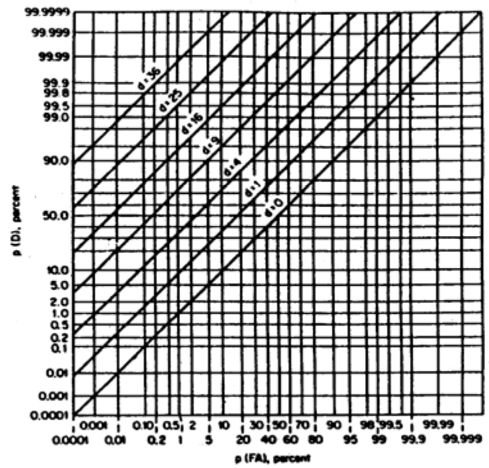


Single-pulse POD curves replotted in ROC format

- **o RECEIVER OPERATING CHARACTERISTIC**
  - **Each curve plots POD against PFA for a single S/N as E_t is varied**
  - **High POD with low PFA is usually the most desirable goal**
    - **Implies high S/N ratio**
    - **Provides a rationale for selecting optimum threshold**

SIGNAL DETECTION THEORY FOR SONAR SYSTEMS



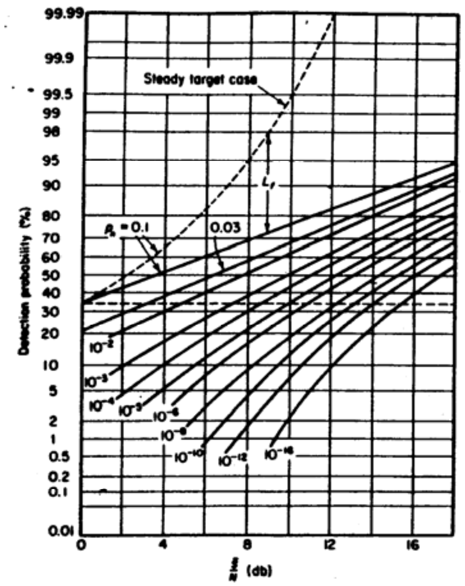


ROC curves reformatted on probability coordinates

- **ROC CURVES IN BINORMAL COORDINATES**
  - **Noise and signal-plus-noise taken as Gaussian, with equal variance**
    - **ROC curves become straight lines of unit slope**
  - **Use is made in Psychoacoustics of a Detection Index, d**
    - **Defined as squared difference between means divided by variance**

SIGNAL DETECTION THEORY FOR RADAR SYSTEMS

Probability of detection for a fluctuating target



- **EFFECT OF VARYING TARGET SIGNAL**
  - **Single-pulse detection probabilities in Gaussian noise**
  - **Effect of a Rayleigh-distributed echo voltage**
  - **Fluctuation raises low POD's and reduces high POD's**

SIGNAL DETECTION THEORY FOR RADAR SYSTEMS

- **INTEGRATION OF PULSE TRAINS**

  - **Radar system detection is rarely based on a single pulse**

  - **Enhanced POD results from efficient use of a train of pulses**
    - **An ideal detector would add energy from n pulses in a matched filter**
      - **Would give integrated S/N ratio n times the single-pulse value**
      - **Practical radars using video integration achieve about $n^{0.6}(S/N)$**

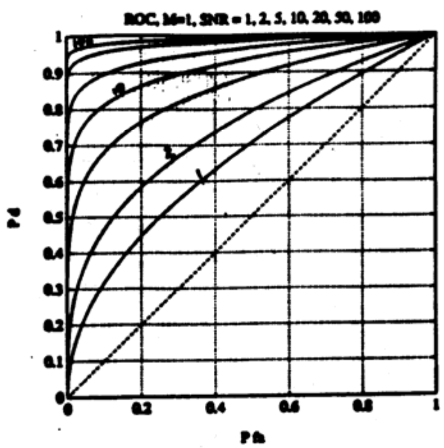
- **SEARCH (SCANNING) RADAR**

  - **Fast scanning increases chance of intercepting target on more than one scan**
    - **Cumulative POD is increased (if scans are statisticaly independent)**
    - $P_n = 1 - (1 - P_i)^n$

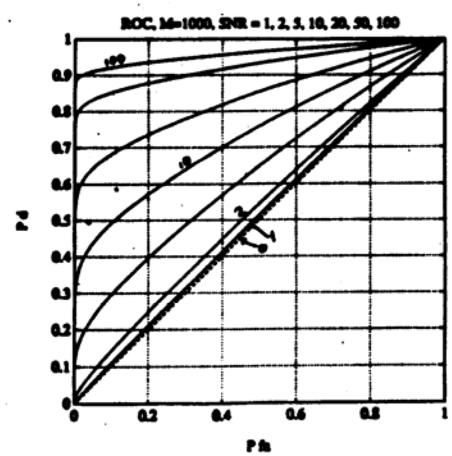
  - **But faster scanning reduces number of pulses integrated per scan**

  - **System parameters may be chosen for optimum net POD**

SIGNAL DETECTION THEORY FOR RADAR SYSTEMS



Inspection of a single cell



Inspection of 1000 cells

- **BROAD AREA SEARCH EFFECTS**
  - Increase the number of area or volume elements searched from 1 to M
    - Chance that noise in one cell will exceed threshold increases
    - E_t must be raised to keep PFA constant
    - Signal-to-noise ratio must be raised to keep POD constant
  - ROC curves tend to assume "peripheral threshold" shape

PSYCHOPHYSICAL APPLICATIONS 0F ROC's

o **TYPICAL HUMAN DISCRIMINATION CHALLENGES**

- **How well do individuals respond to sensory stimuli?**
  - **Brightness, hue, loudness, pitch, taste, smell, etc.**
  - **Size, distance, direction, time, etc.**

- **Is an unbiased measure of cognitive discrimination possible?**
  - **Psychological measurements are plagued by covert discrimination criteria**
    - **Expectations and motivations**
    - **Probabilities and utilities**

o **TWO METHODS HAVE BEEN FOUND TO MINIMIZE BIAS**

- **Present two alternative stimuli, and ask which is A and which is B**

  OR

- **Use single stimuli, and plot data in the ROC format**

PSYCHOPHYSICAL APPLICATIONS 0F ROC's

- o **ANALYSIS OF HUMAN HEARING (1954-57)**
  - An unexpected application of signal detection theory to human perception
  - Traditional view: a sound will always be perceived if it is big enough
  - New view: perception is based in part on the risk of false alarm
    - More sounds are perceived by people willing to accept more false alarms

- o **APPLICATIONS IN OTHER FIELDS (1960's and later)**
  - Clinical Medicine (initial diagnosis)
  - Criminal Justice System (effect of court trials)
  - Library Science (information retrieval)
  - Power Generation industry (evaluation of inspectors)
  - Psychology (memory testing)
  - Radiography (evaluation of film readers)
  - Weather Forecasting

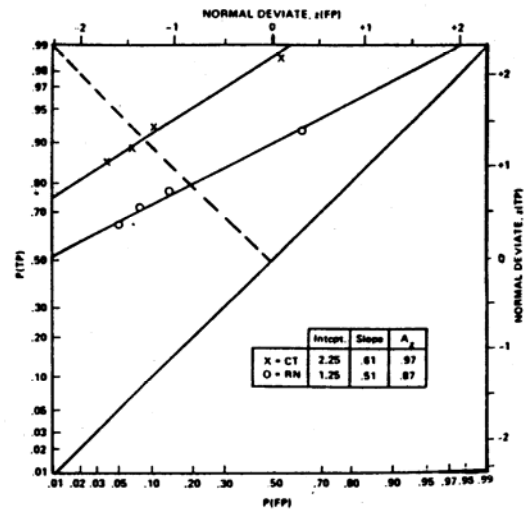
PSYCHOPHYSICAL APPLICATIONS 0F ROC's

- o **SOME CONCLUSIONS ABOUT DETECTION AND RECOGHNITION**
  - **Measures of sensitivity that do not isolate effects of changes in the decision criterion ignore a substantial source of variation**
  - **Measurement of data in terms compatible with signal detection theory reduces test-to-test differences**
  - **ROC analyses help distinguish between causes of change**
    - **Threshold increase is the commonest cause of POD decreasing with time**
      - **May be due to differences between training and actual testing**
      - **The stimulus probability can affect the response probabilities**

- o **RELEVANCE TO NDE**
  - **Likely to be greatest for processes with human sensor/receiver**

<u>**ADDITIONAL ROC EXAMPLES**</u>

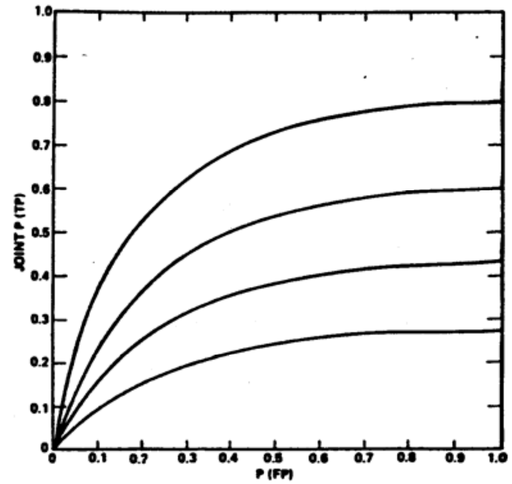
Empirical ROC's from studies of imaging of brain lesions
CT: computed tomography
RN: radionuclide



- o **BINORMAL ROC PLOTS WORK WELL FOR MANY DISTRIBUTIONS**
  - **Some non-normal distributions approximate to straight lines**
    - **e.g. exponential, rectangular, etc.**
  - **Many empirical ROC plots can be fitted by straight lines**
    - **These may be modelled by distributions with different variances**
    - **Slope equals ratio of standard deviations of noise and signal-plus-noise**

ADDITIONAL ROC EXAMPLES

Illustrative joint ROC's:
Probability of detection
and correct classification



- **SIGNAL DETECTION AND CLASSIFICATION**
  - **Joint ROC approach**
    - **Extension of binary ROC to include classification (type, location, etc.)**
    - **Index of classification accuracy independent of detection criterion**
  - **Sequential binary response approach**
    - **Separate binary decisions about detection, type, location, size, etc.)**

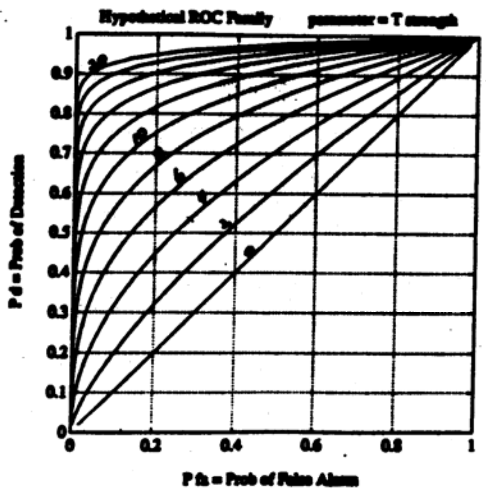
APPLYING SIGNAL DETECTION THEORY TO NDE

- o **PARALLELS BETWEEN RADAR, SONAR, AND ULTRASONIC NDE**
  - **Clearly many similarities**
    - **Pulsed radio-frequency echo detection**
    - **Affected by beam pattern, target reflectivity & transmission medium**
    - **Thermal and non-thermal noise sources**
  - **Some important differences about ultrasonic NDE**
    - **Stationary targets**
    - **Noise usually dominated by time-invariant echoes from microstructure**
- o **TRANSFERABILITY OF RESULTS**
  - **Radar-derived results are of relevance, but must be used with caution**
    - **Offer qualitative insight**
    - **Not quantitatively applicable**
  - **Theories cover only simplified cases (i.e. they only approximate to reality)**

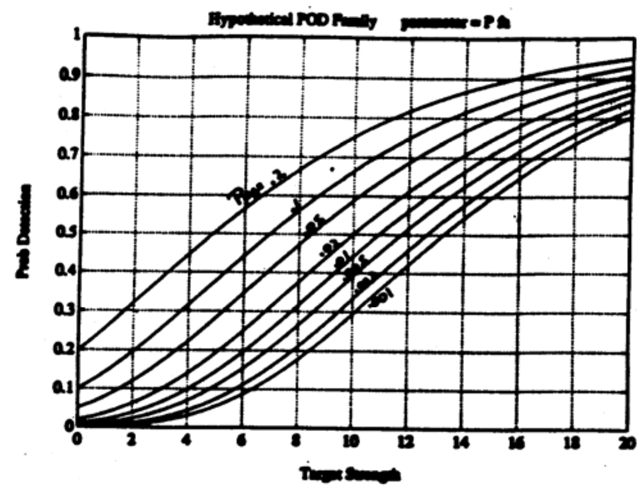
APPLYING SIGNAL DETECTION THEORY TO NDE

- o **GENERAL APPLICATIONS TO NDE**
  - **Signal detection theory applicable to the extent that S & N are measurable**
    - **May be possible for some NDE processes (e.g. eddy-current)**
    - **Likely to be difficult for other processes (e.g. ultrasonics, penetrants)**
  - **Radar targets can be steered into the beam to measure "signal"**
    - **Equivalent may be possible for surface NDE targets, but not subsurface**
  - **Operating Characteristics may be used even without signal distribution**
- o **EXAMPLES (to follow)**
  - **Deriving dependence of POD on defect size from signal theory**
  - **Experimental data displayed in ROC and POD-vs-size formats**
  - **Comparing experimental data with optimized ROC predictions**
  - **Using the ROC format to evaluate inspection/inspector performance**

<u>**AN NDE APPLICATION BASED ON SIGNAL DISTRIBUTIONS**</u>



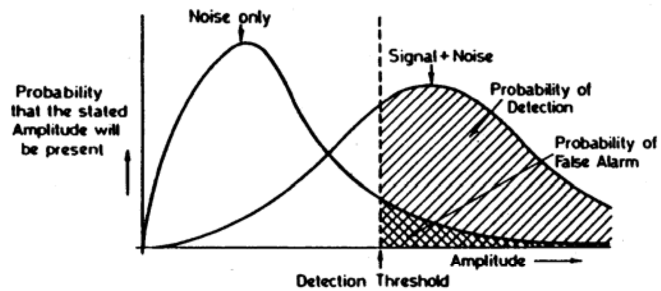
Hypothetical ROC curves

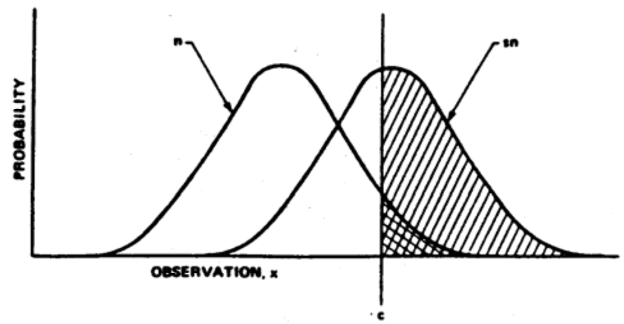


Equivalent plot of POD versus "size"

- o **TAKE S/N AS A MEASURE OF "TARGET STRENGTH"**
  - **Assume only that "target strength" increases as defect size increases**
  - **Select a single Probability of False Alarm**
    - **For low PFA, the result is a typical POD-versus-size plot**

AN NDE APPLICATION BASED ON SIGNAL DISTRIBUTIONS



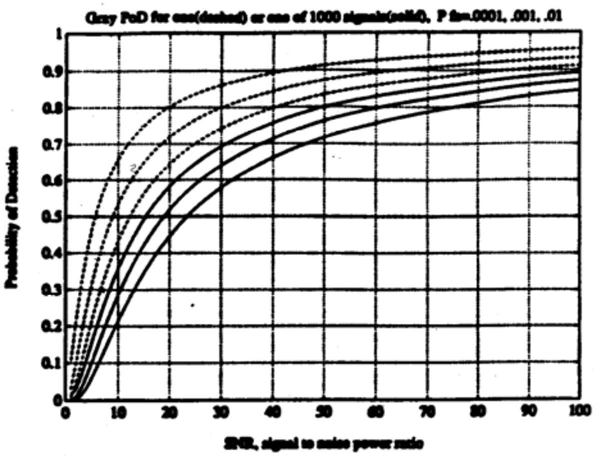


PDF broadened and shifted by signal **PDF shifted without broadening**

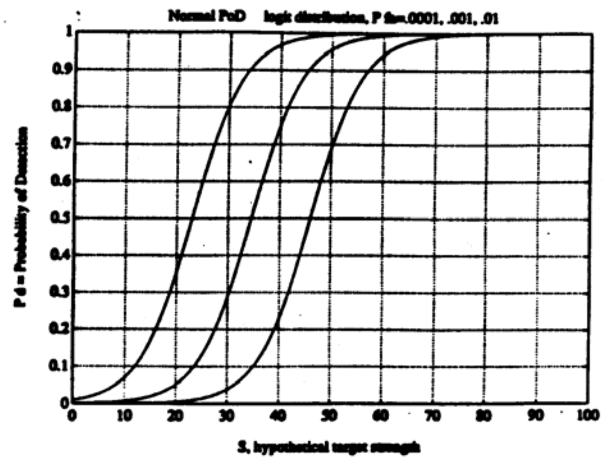
o **DIFFERENCES IN NOISE AND SIGNAL-PLUS-NOISE DISTRIBUTIONS**

- **Distributions depend on phenomena studied and on detection circuitry**
  - **Characterized by shape, mean (location), and variance (scale)**
- **Central limit theorem: large samples often approach Gaussian distributions**

AN NDE APPLICATION BASED ON SIGNAL DISTRIBUTIONS



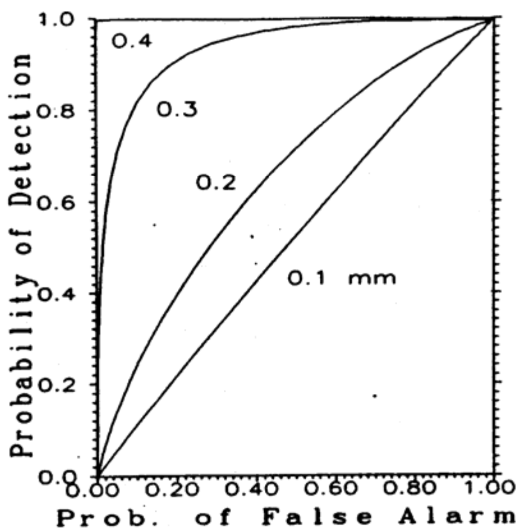
Typical "gray" POD curves



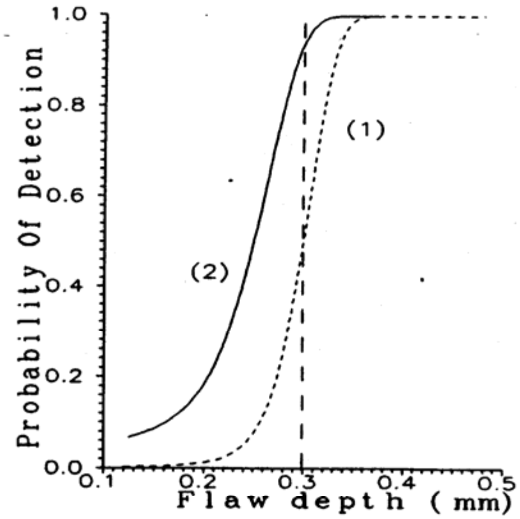
Typical "B&W" POD curves

- **DEPENDENCE ON SHAPE OF S-plus-N DISTRIBUTION**
  - Presence of defect can affect probability density function (pdf) differently
  - "Gray" curves result when the pdf is broadened by presence of defect
  - "Black & white" curves result when pdf is shifted without broadening

AN NDE APPLICATION BASED ON SIGNAL DISTRIBUTIONS



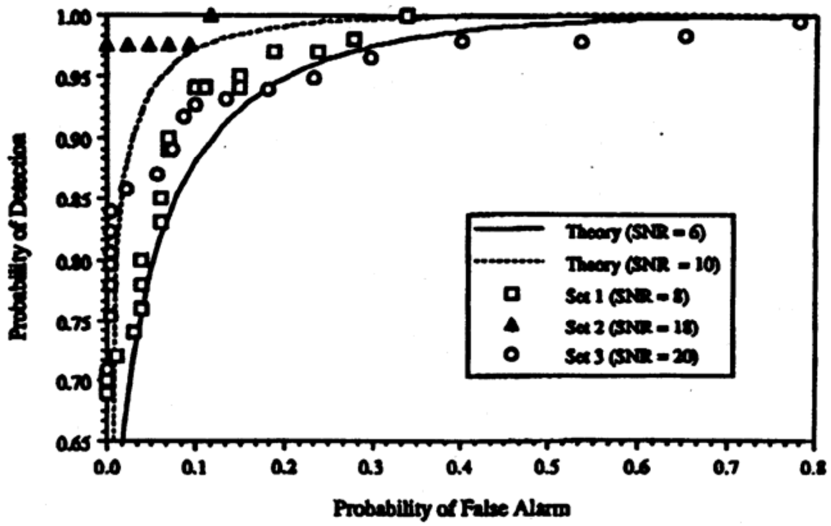
Predicted ROC



Predicted POD

- **PREDICTIONS OF POD AND PFA**
  - Eddy-current inspection of tight cracks
  - Theoretical predictions of defect signals and experimental noise data
  - ROC for various thresholds and four semi-elliptical crack sizes
  - POD for two specific threshold levels

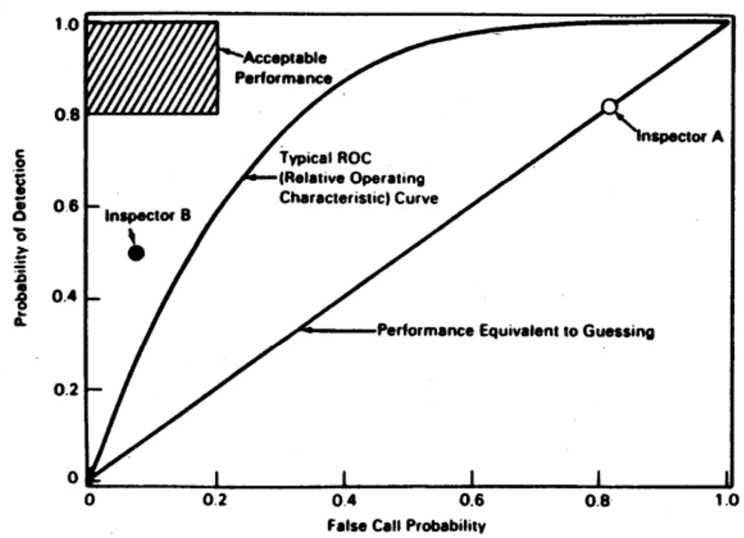
AN NDE APPLICATION BASED ON SIGNAL DISTRIBUTIONS



Comparison of optimized ROC curves with three sets of simulated data

- **OPTIMIZED PREDICTIONS OF POD AND PFA**
  - **Neyman-Pearson likelihood-ratio criterion maximizes POD for given PFA**
- **SIMULATED DEFECT AND NOISE DATA**
  - **Two large sets: simulated spherical defects plus real grain noise data**
  - **One small set (#2): Monte-Carlo simulation of signal and noise**

NDE APPLICATIONS BASED ON POD AND PFA MEASUREMENTS



An Operating Characteristic presentation of POD and PFA

- **POD AND PFA FROM KNOWN-DEFECT SAMPLES**
  - **Diagonal line is equivalent to guessing (target strength = 0)**
  - **The ROC curve indicates the typical effect of changing reporting criteria**
  - **Desirable performance is within the marked box**
    - **Is Inspector A's performance better than Inspector B's?**

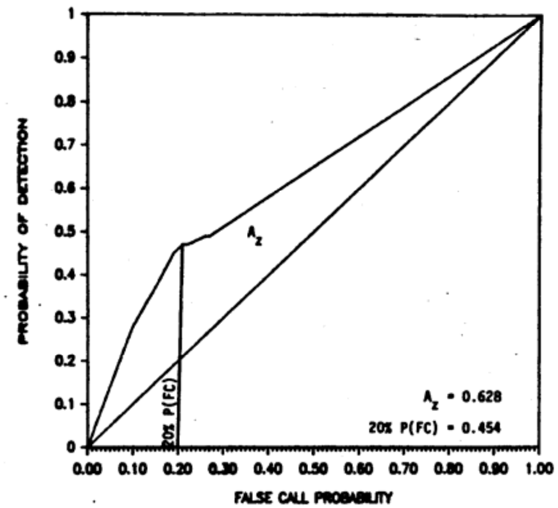
NDE APPLICATIONS BASED ON POD AND PFA MEASUREMENTS

- o **EVALUATION OF A GROUP OF INSPECTORS**
  - **Performance of a group of inspectors should fit an ROC curve**
    - **Approach to the ideal "box" is limited by the sample(s) provided**

- o **EVALUATION OF SINGLE INSPECTORS**
  - **Any well-trained inspector will use consistent reporting criteria**
    - **Performance will not wander along an ROC curve**
    - **Single-point evaluation is appropriate**
    - **Must use composite criteria (such as high POD plus low PFA)**

- o **EVALUATION OF AN INSPECTION PROCEDURE**
  - **An ROC curve presents a more complete description than a single point**
    - **Can be used to determine proper reporting criteria**
  - **Allows comparison of inspection procedures**
    - **The best has highest ROC curve (or bounds the largest area)**

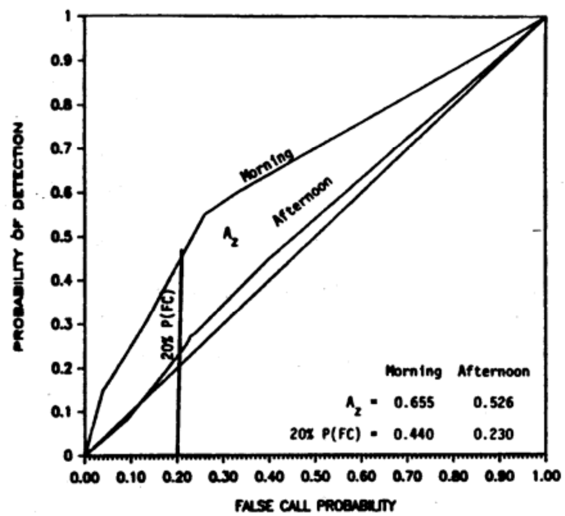
NDE APPLICATIONS BASED ON POD AND PFA MEASUREMENTS



Average inspector performance during an IGSCC round-robin

- **DETECTION AND IDENTIFICATION OF DEFECTS**
  - **12 technicians inspected samples of intergranular stress-corrosion cracks**
  - **Recorded detection and rated certainty of identification of "crack"**
    - **These are "Probability of Detection and Correct Identification" data**
  - **Area A_Z under ROC curve chosen as an evaluation parameter**

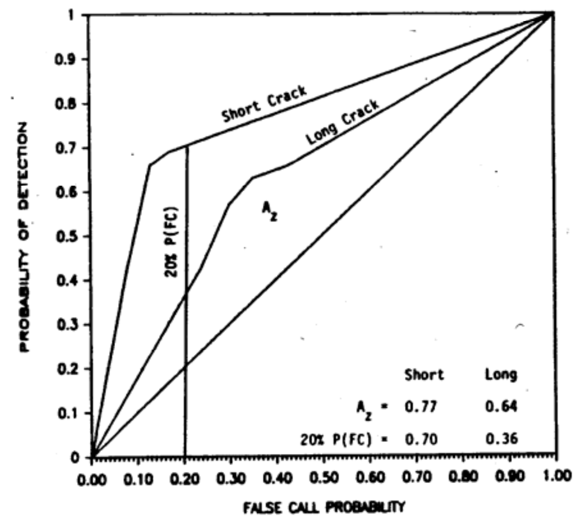
NDE APPLICATIONS BASED ON POD AND PFA MEASUREMENTS



Comparison of performance on far-side short cracks

- **DETRIMENTAL INFLUENCE OF FATIGUE**
  - **Fatigue contributes to incorrect calls**
    - **Long hours (wearing protective clothing) and limited time off**

NDE APPLICATIONS BASED ON POD AND PFA MEASUREMENTS



Comparison of performance on short and long near-side cracks

o **PODCI NOT ALWAYS CORRELATED WITH DEFECT SIZE**

- **Detectability (POD) of long and short cracks was approximately equal**
- **Technicians size predictions were better for short than for long cracks**
  - **PODCI better for short than long cracks**

SUMMARY

- o SIGNAL DETECTION THEORY
  - Useful tool for modelling detectability in communications and in NDE
    - Facilitates studying role of individual inspection parameters
    - Valid only to the extent that the model is valid
  - Extensive work on radar detection readily available
    - Can be applied to NDE with caution (usually not quantitatively useful)
    - Similar studies directly applicable to NDE remain to be done
- o RELATIVE OPERATING CHARACTERISTICS
  - Provide useful framework for displaying parametric POD/PFA relationship
  - Can be used with or without knowledge of signal-plus-noise distributions
  - A different way of looking at NDE capability, not a panacea
    - Problems exist in measuring S/N distributions as well as POD or PFA

SOURCES USED

- D.K.Barton, "Radar Systems Analysis" (Prentice-Hall, 1964)
- T.G.Birdsall, "POD, ROC, Noise and Reality", presentation at the Fall 1992 ASNT meeting
- C-P.Chiou, R.B.Thompson & L.W.Schmerr, "Model-based Signal Processing Techniques for Ultrasonic Flaw Detection: Simulation Studies", Review of Progress in QNDE, Vol.12, (to be published, 1993)
- H.W.Cole, "Understanding Radar" (Collins, 1985)
- W.M.Hall, "Prediction of Pulse Radar Performance" (Proc. IRE 44, 224 (1956)
- P.G.Heasler, D.J.Bates, T.T.Taylor & S.R.Doctor, NUREG/CR-4464 (U.S.Nuclear Regulatory Commission, 1986)
- N.Nakagawa, M.W.Kubovich & J.C.Moulder, "Modelling Inspectability for an Automated Eddy Current Measurement System", Review of Progress in QNDE, Vol.9A, p. 1065 (Plenum Press, 1990)
- D.O.North, "Analysis of Factors Which Determine Signal-to-Noise Discrimination in Radar", Report PTR-6c (RCA Laboratories, 1943)
- J.A.Swets, "The Relative Operating Characteristic in Psychology", Materials Evaluation 41, 1294-1303 (1983)
- J.A.Swets & R.M.Pickett, "Evaluation of Diagnostic Systems: Methods from Signal Detection Theory" (Academic, 1982)
- R.J.Urick, "Principles of Underwater Sound" (McGraw-Hill, 1983)
- W.A.Wheeler, W.L.Rankin, J.C.Spanner, R.V.Badalemente & T.T.Taylor, NUREG/CR-4600 (U.S.Nuclear Regulatory Commission, 1986)

Reliability

- A characteristic that measures our belief that something is worthy of our confidence

Component Reliability

- 1 - F(t), where F(t) is the probability that a component will fail at time t or sooner; implicitly says the probability of failure increases with time beyond a certain point

NDE Reliability

- P(NF | NI), Probability that no defect which threatens loss of structural integrity exists in a component given that NDE says no such defect exists

Inspection

- Does not change the component inspected; Does change our estimate of the reliability of the component to the degree that we have NDE reliability

NDE Reliability and Probability of Detection (POD) are not the same thing

NDE Reliability depends on POD , False Alarm Rate (FAR), the Probability of a critical flaw P(F), knowing where to look and knowing where you can't look

Aspects of Inspection:

Detection	- Identifying conditions that need further evaluation; usually conditions produced by a defect
Characterization	- Describing the nature or "character" of a suspicious set of conditions related to the defect
Sizing	- The quantification of the extent of the detected and characterized defect

EPRI/NPD

per·func·to·ry (pər·fungk′tər·ē) *adj.* Done merely for the sake of getting through; mechanical and without interest; half–hearted; negligent; superficial; careless. [<LL *perfunctorius* negligent < *perfunctor* one who performs an act < *perfungi* get through with < *per-* through + *fungi* perform] — **per·func′to·ri·ly** *adv.* — **per·func′to·ri·ness** *n.*

E&O

Credit for Inspection: (Good NDE !)

- Longer period between inspections
- Modification of defect distribution
- Motivation for improving NDE reliability