Theoretical Limits for Optical Position Estimation Using Imaging Arrays*

Alfred Hero

Dept. of Electrical Engineering and Computer Science
The University of Michigan, Ann Arbor, MI 48109
USA

Résumé: *On présente des résultats qui spécifient les limites fondementales sur la précision des estimées de la position d'une source optique quasimonochromatique vue par un téléscope avec détecteur de photons CCD. Pour un téléscope à petite ouverture, nos résultats indiquent que c'est théoriquement possible de déterminer la position d'une source avec une précision supérieure à 1/50-ième de la limite de diffraction en détectant seulement 1200 photons.*

Abstract: *We investigate fundamental limits on achievable estimation accuracy for estimating the spatial position of a far-field monochromatic optical point source on the basis of diffraction and noise limited spatio-temporal measurements at the output of a light-sensitive imaging (CCD) array. We apply our results to a small aperture CCD telescope and show that it is theoretically possible to resolve point source position to within 1/50-th of the diffraction limit on the basis of fewer than 1200 photons.*

1 Introduction

The problem of localizing the position of one or more optical point sources arises in applications such as galactic astronomy and astrometry, satellite navigation and telemetry, and star tracking systems for global positioning. Lower bounds on achievable localization error are useful in that they characterize the performance limits intrinsic to any estimator which are imposed soley by the structure of the estimation problem. Lower bounds have previously been derived under restrictive assumptions such as: 1) direct noiseless observations of the photon point process incident on the photo-detector array; 2) negligable thermal noise [2,7]. However, while these assumptions may be good approximations for strong optical sources, cryogenically cooled detectors, and integrate-and-dump CCD arrays, they may not be justified for many cases of interest, e.g. weak stellar sources. Indeed, these assumptions are inapplicable for a continuously observed photo-detector output whose single photon-response (SPR) is temporally and spatially bandlimited and contaminated by thermal (electronics) noise. In this paper we derive rate distortion lower bounds on localization MSE which allow us to study the impact of finite detector bandwidth and thermal noise on achievable estimation error. A useful and tractible bound is derived for the special cases of a single point source, low accumulated photon flux, a spherically-symmetric point spread (blur) function, and exponential temporal-decay of the detector SPR. For example, this case arises in laser radar for long distance telemetry and in weak star tracking [1,5,6]. Using the results of this paper we can identify the form of a SNR threshold which separates localization errors into two regimes: the *photon-noise-limited regime*, where estimator performance equal to that achievable for noiseless direct photon detection is possible; and the *thermal-noise-limited regime*, where such direct detection estimator performance is not achievable. We then give numerical results which indicate that for the small aperture telescope studied, fewer than 1200 photons are theoretically required to resolve a point source to within 1/50-th of the diffraction-limit, defined as the $(FWHM)^2$ of the optical blur spot. Furthermore, we show that for this case the rate-distortion bound can be significantly tighter than the classical CR bound.

2 Measurement Model

The measurements are obtained by detecting incoherent quasimonochromatic light on an optical focal plane array of photo detectors. The measurements are distortion and distortion contaminated by optical diffraction, quantum (photon) noise, and thermal (electronics) noise. We model the photo-detector as a fixed disk of radius r_o. Assume a far-field stationary point source generates a symmetric blur function with center of symmetry at detector position $\underline{\tau} = [\tau_1, \tau_2]^T$ relative to the center $[0, 0]$ of the detector surface. We will assume *a priori* that $\underline{\tau}$ is uniformly distributed over a disk $\mathcal{T} = \{u_1, u_2 : u_1^2 + u_2^2 \leq r_1^2\}$ inscribed on the interior of the detector surface $A = \{u_1, u_2 :$

[1]This research was supported in part by the National Cancer Institute under grant R01-CA-46622-01 and the National Science Foundation under grant BCS-9024370.

$u_1^2 + u_2^2 \leq r_o^2\}$, $r_o > r_1$. The output of the detector is observed continuously over an observation interval $[-\frac{T}{2}, \frac{T}{2}]$. Define the spatio-temporal coordinates $(t, \underline{u}) \in I$, $I \stackrel{\text{def}}{=} [-\frac{T}{2}, \frac{T}{2}] \times A$, of photon incidences over the observation interval and over the detector surface. Conditioned on $\underline{\tau}$, let $\lambda = \{\lambda(t, \underline{u} - \underline{\tau}) : (t, \underline{u}) \in I\}$ be the intensity of the incident photon point process $d\mathbf{N} = \{d\mathbf{N}(t, \underline{u}) : (t, \underline{u}) \in I\}$. We assume that the point process $d\mathbf{N}$, equivalently the counting process $\mathbf{N}$, is conditionally Poisson given $\underline{\tau}$. Define $\mathbf{n} = \mathbf{N}(I)$ the total number of incident photons over I. Let $\{(\mathbf{t}_i, \underline{\mathbf{u}}_i)\}_{i=1}^{\mathbf{n}}$ be the $\mathbf{n}$ coordinates of these photons. Let $\mathbf{g}_i$ denote the total induced charge on the photo-detector resulting from a photon interaction at spatial position $\underline{\mathbf{u}}_i$ and at time $\mathbf{t}_i$. Conditioned on $\mathbf{n}$, $\{\mathbf{g}_i\}_{i=1}^{\mathbf{n}}$ is a sequence of i.i.d. random variables with range $\mathcal{G} = \mathbb{R}$ and probability density p_g. The mean μ_g and variance σ_g^2 of the density p_g specify the mean efficiency and energy spread, respectively, of the photon-collection process. We assume that the $\mathbf{g}_i$'s are statistically independent of $\underline{\tau}$ given $\mathbf{n}$.
The sequence $\{(\mathbf{t}_i, \underline{\mathbf{u}}_i), \mathbf{g}_i\}_{i=1}^{\mathbf{n}}$ of photon arrival coordinates $(\mathbf{t}_i, \underline{\mathbf{u}}_i) \in [-\frac{T}{2}, \frac{T}{2}] \times A$ and photo-detector gains $\mathbf{g}_i \in \mathcal{G}$ defines a marked point process $d\mathbf{M}$ with index set $I = [-\frac{T}{2}, \frac{T}{2}] \times A$ and mark space $\mathcal{G}$. In the sequel we refer to the process $\mathbf{M}$ as the *direct photon detection data* or the *photon-noise-limited regime*. By contrast the actual measurement corresponds to the following finite bandwidth and thermal noise contaminated observations over $(t, \underline{u}) \in I$:

$$\begin{aligned} \mathbf{X}(t, \underline{u}) &= \sum_{i=1}^{\mathbf{n}} \mathbf{g}_i p(t - \mathbf{t}_i, \underline{u} - \underline{\mathbf{u}}_i) + \mathbf{w}(t, \underline{u}) \\ &= \iiint_I p(t - v, \underline{u} - \underline{\xi}) d\mathbf{M}(v, \underline{\xi}) + \mathbf{w}(t, \underline{u}), \end{aligned} \tag{1}$$

where $\mathbf{w}$ is a spatio-temporal white Gaussian noise. We assume that the unit energy SPR pulse $p(t, u_1, u_2)$ is mono-exponential over the time variable t and spherically symmetric over the spatial variables u_1, u_2:

$$p(t, \underline{u}) = \begin{cases} \frac{1}{T_p} e^{-t/T_p} \cdot p(\|\underline{u}\|), & t \geq 0 \\ 0, & t < 0 \end{cases}, \tag{2}$$

where $\|\underline{u}\| = \sqrt{u_1^2 + u_2^2}$, $T_p \ll T$, and $\sigma_p \ll r_o$. We will assume that the detector intensity is constant over time and has the form of a spherically symmetric spatial blur function over $\underline{u}$:

$$\lambda(t, \underline{u}) = \frac{1}{T} \lambda(\|\underline{u}\|). \tag{3}$$

The functional independence of λ with respect to t corresponds to imaging a stationary photon source while the symmetry of λ with respect to $\underline{u}$ corresponds to imaging with a symmetric optical aperture. It is also assumed that conditioned on $\underline{\tau}$, the integrated intensity $\Lambda \stackrel{\text{def}}{=} \int_I \lambda = E[\mathbf{n}|\underline{\tau}]$ of $\mathbf{N}$ is functionally independent of $\underline{\tau}$. This assumption, called conditional energy invariance in [3], is appropriate in cases where the average rate Λ is known *a priori* and the spatial support $\{\underline{u} \in \mathbb{R}^2 : \lambda(t, \underline{u} - \underline{\tau}) > 0\}$ is contained in A for all $\underline{\tau}$.
Briefly stated, the estimation objective is to determine the value of the vector parameter $\underline{\tau}$ based on the photo-detector waveform output $\mathbf{X}$ given in (1). A relevant measure for the accuracy of an estimator $\hat{\underline{\tau}}(\mathbf{X})$ of $\underline{\tau}$ is the average of the component MSE's:

$$\overline{\text{MSE}}(\mathbf{X}) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{i=1}^{2} E[(\tau_i - \hat{\tau}_i(\mathbf{X}))^2]. \tag{4}$$

Note that $\overline{\text{MSE}}$ is related to the volume of the concentration ellipsoid $E[(\hat{\underline{\tau}} - \underline{\tau})(\hat{\underline{\tau}} - \underline{\tau})^T]$.

3 Rate Distortion Lower Bounds on MSE

In [4] a rate distortion bound on $\overline{\text{MSE}}$ were derived for a more general parameter estimation problem than the localization problem considered here. This bound is exponentially decreasing in the channel capacity C, associated with the conditional probabilities governing the "channel output" $\mathbf{X}$ given the "channel input" $\underline{\tau}$, and the entropy function $H(\underline{\tau})$ associated with the prior distribution of $\underline{\tau}$. As explained in [4], an exact expression for this bound is intractable, requiring an expression for the joint densities of the process $\mathbf{X}$. To arrive at a tractable bound we used the data processing inequality $C \leq \min\{C_1, C_2\}$, where C_1 and C_2 are, respectively, the channel capacities of a point process channel, mapping the source symbols $\underline{\tau}$ into the direct photon-detection data $\mathbf{M}$, and a continuous waveform channel, mapping $\mathbf{M}$ into the observations $\mathbf{X}$. We then showed that C_1 can be upper bounded by the capacity C_1^* of a Poisson channel with associated intensity λ and mark distribution f_g, and C_2 can be upper bounded by the capacity C_2^* of an additive Gaussian noise channel whose output has the spatio-temporal covariance function $K_X(\underline{z}_1, \underline{z}_2) = K_s(\underline{z}_1, \underline{z}_2) + \frac{N_o}{2}\delta(\underline{z}_1 - \underline{z}_2)$. Here $K_s(\underline{z}_1, \underline{z}_2)$ is the covariance function of the signal component $s(t, \underline{u}) \stackrel{\text{def}}{=} \sum_{i=1}^{\mathbf{n}} \mathbf{g}_i p(t - \mathbf{t}_i, \underline{u} - \underline{\mathbf{u}}_i)$ of the observations $\mathbf{X}$. This gave the following rate distortion lower bound on $\overline{\text{MSE}}$ defined in (4):

$$\overline{\text{MSE}}(\mathbf{X}) \geq B_{rdlb}(\mathbf{X}) \stackrel{\text{def}}{=} \frac{1}{2\pi e} e^{H(\underline{\tau})} e^{-\min\{C_1^*, C_2^*\}}. \tag{5}$$

Application to Point Source Localization

Using the forms (2) and (3) for $p(t, \underline{u})$ and $\lambda(t, \underline{u})$ given in Sec. II, it can be shown [4] that (for $r_o, r_1, T \to \infty$ and $r_1/r_o \to 1$) the covariance function $K_s(\underline{z}_1, \underline{z}_2)$ is homogeneous, i.e. $K_s(\underline{z}_1, \underline{z}_2) = K_s(\underline{z}_1 - \underline{z}_2)$, and isotropic, i.e. $K_s(\underline{z})$ depends only on the magnitude $\|\underline{z}\|$, over the spatio-temporal range of the observation $\underline{z}_1, \underline{z}_2 \in I$ (Note this does not imply the unreasonable properties that the field $\mathbf{X}$ is either conditionally homogeneous or conditionally isotropic given $\underline{\tau}$). For homogeneous K_s the theory of Toeplitz forms provides a spectral domain description of the capacity C_2^* in terms of the limiting capacity-per-unit-volume $\overline{C_2^*} \stackrel{\text{def}}{=} \lim_{r_o, T \to \infty} C_2^*/|I|$, where $|I|$ denotes the volume of the set I. In particular:

$$\overline{C_2^*} = \frac{1}{2} \iiint_{-\infty}^{\infty} df\, d\nu_1 d\nu_2 \ln\left(1 + \frac{2\, G_s(f, \nu_1, \nu_2)}{N_o}\right), \tag{6}$$

where G_s is the power spectral density associated with $K_s(\underline{z})$:

$$G_s(f,\nu_1,\nu_2) = \iiint_{-\infty}^{\infty} dt du_1 du_2 K_s(t,u_1,u_2) e^{-j2\pi(ft+\nu_1 u_1+\nu_2 u_2)}. \tag{7}$$

In the Fourier transform (7) f is temporal frequency and ν_1,ν_2 are spatial frequencies. The limiting form (6) for the normalized capacity $C_2^*/|I|$ gives the large T and large r_o approximation:

$$C_2^* \approx |I|\overline{C_2^*} \tag{8}$$

to be used in the sequel.
Under the assumed form (2) an integral identity can be used to simplify (6) resulting in:

$$\overline{C_2^*} = \frac{\pi}{T_p}\int_0^\infty r\left[\sqrt{1+\tilde{Q}(r)}-1\right]dr, \tag{9}$$

where $\tilde{Q}$ is the spherically-symmetric spatial signal-spectrum computed as the 1-D Hankel transform of $\tilde{K}_s(v) \stackrel{\text{def}}{=} K_s(0,\underline{u})$ and $v \stackrel{\text{def}}{=} \|\underline{u}\|$. For our problem $\tilde{Q}$ has the following form:

$$\tilde{Q}(r) = \gamma|\tilde{P}(r)|^2\left[\Lambda + \frac{\mu_g^2}{\mu_g^2+\sigma_g^2}|\tilde{\Lambda}(r) - \frac{\Lambda}{2\pi r_o r}J_1(2\pi r_o r)|^2\right], \tag{10}$$

In (10) $\tilde{P}$ and $\tilde{\Lambda}$ are the Hankel transforms of $p(\|\underline{u}\|)$ and $\lambda(\|\underline{u}\|)$, $\sqrt{\gamma}$ is the rms single-photon-response-to-noise power ratio (SPRNR):

$$\gamma \stackrel{\text{def}}{=} \frac{2(\mu_g^2+\sigma_g^2)}{N_o},$$

J_1 is the Bessel function of order 1, and, as before, Λ is the integrated intensity.
Using (8) and (9) we thus have:

$$C_2^* \approx \frac{|I|\pi}{T_p}\int_0^\infty r\left[\sqrt{1+\tilde{Q}(r)}-1\right]dr. \tag{11}$$

where $\tilde{Q}$ is given by (10). The approximation (11) is accurate for large T and r_o and $r_o/r_1 \approx 1$.
Furthermore, using results of [3], the capacity C_1^* is equal to the *information divergence* $d(\lambda,\overline{\lambda})$ between two spatial intensities λ and $\overline{\lambda}$:

$$C_1^* = d(\lambda,\overline{\lambda}) \stackrel{\text{def}}{=} \iint_A \lambda(\underline{u})\ln\frac{\lambda(\underline{u})}{\overline{\lambda}(\underline{u})}d\underline{u}, \tag{12}$$

where $\overline{\lambda}(\underline{u})$ in (12) is constant over the spatial domain A:

$$\overline{\lambda}(\underline{u}) = \begin{cases} \frac{\Lambda}{|A|}, & \underline{u}\in A \\ 0, & o.w. \end{cases} \tag{13}$$

Combining all of the above results the rate distortion bound (5) takes the form:

$$B_{rdlb}(\mathbf{X}) = \begin{cases} \frac{|A|}{2\pi e}e^{-\iint_A \lambda(\underline{u})\ln\frac{\lambda(\underline{u})}{\overline{\lambda}(\underline{u})}d\underline{u}}, & \gamma>\gamma_o \\ \frac{|A|}{2\pi e}e^{-\frac{|I|\pi}{T_p}\int_0^\infty r\left[\sqrt{1+\tilde{Q}(r)}-1\right]dr}, & \gamma\le\gamma_o \end{cases}, \tag{14}$$

where γ_o is a SPRNR threshold determined by the condition $C_1^* = C_2^*$ in (5). Specifically, γ_o is the solution $\gamma=\gamma_o$ of the equation:

$$\iint_A \lambda(\underline{u})\ln\frac{\lambda(\underline{u})}{\overline{\lambda}(\underline{u})}d\underline{u} = \frac{|I|\pi}{T_p}\int_0^\infty r\left[\sqrt{1+\tilde{Q}(r)}-1\right]dr, \tag{15}$$

when the solution exists. Note that this solution does not exist when the spatial filter response approaches a delta function, i.e. $T_p,\sigma_p \to 0$, corresponding to perfect observation of the photon arrivals. In this case, since C_2^* becomes unbounded, C_1^* is always less than C_2^*, and hence $\min\{C_1^*,C_2^*\} = C_1^*$, for all $\gamma>0$. Therefore, when perfect photon observations are available, $\gamma_o = 0$ and the lowerbound (14) is identical to the direct detection bound $B_{rdlb}(\mathbf{M}) = |A|/(2\pi e)\exp(-C_1^*)$. Note also that the lower bound (14) separates MSE performance into two SPRNR regimes: the photon-noise-limited regime ($\gamma>\gamma_o$); and the thermal-noise-limited regime ($\gamma\le\gamma_o$). In the photon-noise-limited regime $B_{rdlb}(\mathbf{X})$ decreases exponentially as a function of the information divergence $d(\lambda,\overline{\lambda})$ which measures the difference between the blur function $\lambda(\underline{u})$ and $\overline{\lambda}(\underline{u})$ (13). The closer $\lambda(\underline{u})$ is to the uninformative uniform intensity $\overline{\lambda}(\underline{u})$, the poorer becomes the estimator MSE. On the other hand, in the thermal-noise-limited-regime regime $B_{rdlb}(\mathbf{X})$ decreases exponentially in the rms deviation between the Hankel transform $\tilde{\Lambda}$ of the blur function $\lambda(\|\underline{u}\|)$ and the Hankel transform $\frac{\Lambda}{2\pi r_o r}J_1(2\pi r_o r)$ of the uniform intensity $\overline{\lambda}(\|\underline{u}\|)$. Observe also that the maximum value of $B_{rdlb}(\mathbf{X})$ is the "entropy power" $\frac{1}{2\pi e}e^{H(\underline{\tau})} = |A|/(2\pi e)$ which is approximately equal to the *a priori* variance of τ_1 and τ_2.

Gaussian Beam Model

Finally we specialize to the following spatially symmetric Gaussian blur function and SPR models:

$$\lambda(u_1,u_2) = \frac{\Lambda}{2\pi\sigma_\lambda^2}e^{-\frac{u_1^2+u_2^2}{2\sigma_\lambda^2}}$$

$$p(u_1,u_2) = \frac{1}{2\pi\sigma_p^2}e^{-\frac{u_1^2+u_2^2}{2\sigma_p^2}}.$$

Using (12) it can be shown that:

$$C_1^* = \Lambda\ln\left[\frac{|A|}{2\pi e\sigma_\lambda^2}\right]. \tag{16}$$

Furthermore $\tilde{Q}$ can be shown to have an explicit non-integral form not reproduced here.

4 Numerical Results

It is easily verified that the bound (5) depends on the parameters $r_1, r_o, \sigma_p, \sigma_\lambda, T_p, T, \mu_g, \sigma_g, N_o$ only through the *a priori* variance $|A|/(2\pi e)$ and through the ratios r_1/r_o, r_o/σ_p, σ_λ/σ_p, T/T_p, and $\gamma = 2(\mu^2+\sigma^2)/N_o$. We numerically evaluated the bound for a continuous CCD array with the following parameters: $r_1 = r_o = 16cm$, $\sigma_p = 1cm$,

$\sigma_\lambda = 4cm$, $T_p = 0.5sec$, $T = 1sec$, $\gamma = 0.001$. These parameter values were chosen to correspond to a compact CCD star tracking telescope with small-aperture optics and full-well integrate-and-dump CCD readout.

In Fig. 2 we show a plot of the SPRNR threshold γ_o as a function of the total count rate Λ. Note that γ_o is approximately constant for $\Lambda < 200$ and increases approximately linearly in Λ for $\Lambda > 300$. In Fig. 3 a plot of the rate distortion bound is given as a function of Λ. Also shown for comparison is the classical Cramer-Rao bound which was computed in [2] for the perfectly observable photon counting regime. It was shown in [4] that this CR bound remains a valid lower bound for the imperfect observation model considered in this paper. The MSE curves in Fig. 3 are plotted on a log-linear scale and are normalized by the a priori variance of τ_1. Based on Figure 2 we make the following conclusions. First, unlike B_{rdlb} which converges to the theoretically maximum MSE equal to the a priori variance of τ_1, B_{cr} diverges to infinity as Λ goes to zero. This reflects the fact that the CR bound is not strictly valid for a uniform prior density since such a prior is not differentiable on the boundary of the disk. Second, the rate distortion lower bound is significantly tighter than the CR bound over the range $0 < \Lambda < 1800$. This range is relevant for imaging weak optical sources. Third, our results indicate that only 300 photons are theoretically required to achieve the diffraction limited resolution ($10Log\sigma_\lambda^2 = -15dB$) while fewer than 1200 photons are required to resolve source position to within 1/50-th of the diffraction limit ($-48dB$). An interesting topic for further study is the possible construction of position estimators which can come close to achieving the performance limits predicted in this study.

References

[1] J. Amoss and F. Davidson, "Detection of weak optical images with photon counting techniques, *Applied Optics*, vol. 11, pp. 1793-1799, Aug. 1972.

[2] M. Elbaum, P. Diament, M. King, and W. Edelson, "Maximum angular accuracy of pulsed laser radar in photocounting limit, *Applied Optics*, vol. 16, pp. 1982-1992, July 1977.

[3] A. Hero, "Lower bounds on estimator performance for energy invariant parameters of multi-dimensional random processes, *IEEE Trans. on Info. Theory*, vol. 35, pp. 843-858, July 1989.

[4] A. Hero, "Recovering photon intensity information from continuous photo-detector measurements, *Proc. Conf. on Info. Sci. and Syst.*, Johns Hopkins University, Mar. 1991.

[5] L. Kazovsky, "Beam position estimation by means of detector arrays, *Opt. Quant. Elect.*, vol. 13, pp. 201-208, 1981.

[6] P. Salomon and T. Glavitch, "Image signal processing in sub-pixel accuracy star trackers, *Proc. Soc. Photo. Opt. Instrum. Eng.*, pp 64-74, 1980.

[7] K. Winick, "Cramer-Rao lower bounds on the performance of charge-coupled-device optical position estimators, *Opt. Soc. Am.*, vol. 3, pp. 1809-1815, Nov. 1986.

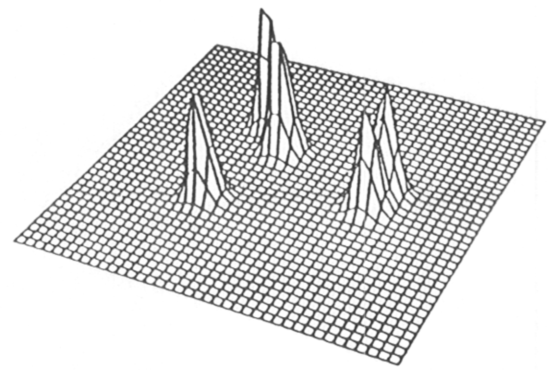

Figure 1: Snapshot of **X** at time $t \in [0, T]$ for low magnitude cylindrical photon intensity λ.

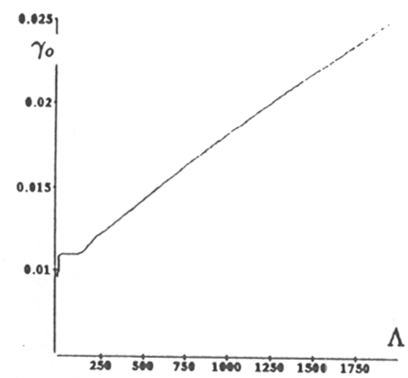

Figure 2: The SPRNR threshold γ_o as a function of count rate Λ

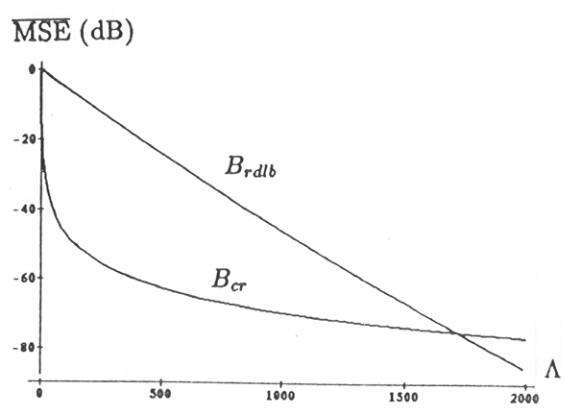

Figure 3: The rate distortion lower bound B_{rdlb} and the Cramer-Rao lower bound B_{cr} on $\overline{\text{MSE}}$