# 1

# A Study of Cramér-Rao-Like Bounds and Their Applications to Wireless Communications

Kao-Peng Chou and Jia-Chin Lin
*Department of Communication Engineering, National Central University*
*Jhongli, Taoyuan,*
*Taiwan*

## 1. Introduction

Estimation theory has been developed over centuries. There are several approaches to utilizing this theory; in this chapter, these approaches are classified into three types. Type I includes the oldest two methods, the least squares (LS) and moment methods; both of these methods are non-optimal estimators. The least squares method was introduced by Carl Friedrich Gauss. Least squares problems fall into linear and non-linear categories. The linear least squares problem is also known as regression analysis in statistics, which have a closed form solution. An important feature of the least squares method is that no probabilistic assumptions of the data are made. Therefore, the linear least squares approach is used for parameter estimation, especially for low complexity design (Lin, 2008; 2009). The design goal of the least squares estimator is to find a linear function of observations whose expectation is a linear function of the unknown parameter with minimum variance. In addition, the least squares method corresponds to the maximum likelihood (ML) criterion if the experimental errors are normally distributed and can also be derived from the moment estimation. As an alternative to the LS method, the moment method is another simple parameter estimation method with probabilistic assumptions of the data. The general moment method was introduced by K. Pearson. The main procedure in the moment method involves equating the unknown parameter to a moment of distribution, then replacing the moment with a sample moment to obtain the moment estimator. Although the moment estimator has no optimal properties, the accuracy can be validated through lengthy data measurements. This is mainly because the estimator based on moment can be maintained to be consistent. Type II includes the methods of minimum variance unbiased estimator (MVUE) and the Bayesian approach, which are both optimal in terms of possible minimum estimation error, i.e., statistical efficiency. MVUE is the best guess of an unknown parameter. The standard MVUE procedure includes two steps. In the first step, the Cramer-Rao lower bound is determined, and the ability of some estimator to approach the bound. In the second step, the Rao-Blackwell-Lehmann-Scheffe (RBLS) theorem is applied. The MVUE can be produced by these two steps. Moreover, a linear MVUE might be found under more restricted conditions.

In the Bayesian method, the Bayesian philosophy begins with the cost function, and the expected cost with respect to the parameter is the risk. The design goal of Bayesian philosophy is to find an estimator that minimizes the average risk (Bayes risk). The most

common cost function is a quadratic function because it measures the performance of the estimator in terms of the square of the estimation error. In this case, the Bayes risk is the mean square error (MSE), and thus, the Bayes estimate is a minimum mean square error (MMSE) estimator. Another common cost function is the absolute function, which regards the absolute estimate error as the Bayes risk. In this case, the Bayes estimate is a minimum mean absolute error (MMAE) estimator. Another estimation, which is not a proper Bayes estimation but fits within Bayes philosophy, is the maximum a posteriori (MAP) estimation. The MAP criterion considers the uniform cost function, and the parameter is discretely, randomly distributed under this assumption. Although this estimate usually only approximates the Bayes estimate for uniform cost, the MAP criterion is widely used for estimator design. Type III includes the maximum likelihood (ML) estimate, which is the most important estimation theory in the 20th century. The ML estimate can be referred to as an alternative MAP without knowledge of apriori probability of the parameters. The ML estimator is the most popular approach for obtaining a practical estimator, which was previously used by Gauss. The general method of estimation was first introduced by R. A. Fisher with the concepts of consistency, efficiency and sufficiency of the estimation function. The ML estimator is required when MVUE does not exist or cannot be found. An advantage of the ML estimator is that a practical estimation is easy to obtain through the prescribed procedures. Another advantage of this approach is that MVUE can be approximated due to its efficiency. Thus, from the theoretical and practical perspectives, the ML approach is the most important and widely used estimation method of this century (Lin, 2003).
Because the ML estimator is essential in estimation theory, the analysis of its performance is a benchmark of estimator design. This benchmark is commonly known as the Cramer-Rao lower bound (CRLB), which is named after Harald Cramer and Calyampudi Radhakrishna Rao. In section 2, the definition of the CRLB is introduced with several examples. A general case of CRLB under two common communication channels is then introduced in section 3. To establish basic knowledge of hybrid parameter estimation, random parameter estimation is presented in section 4. In section 5, Cramer-Rao-like bounds for hybrid parameter estimation are introduced and compared with each other. Lastly, we summarize some practical cases and compare these cases with modified CRB which is most common used Cramer-Rao-like bounds.

## 2. Cramer-Rao lower bound (CRLB)

The Cramer-Rao lower bound (CRLB) is a lower bound on the variance of any unbiased estimator. Many other variance bounds exist, but the CRLB is the easiest one to derive and is thus widely used in many estimation studies. This theory provides a benchmark for examining the performance of novel estimation algorithms and also highlights the impossibility of finding an unbiased estimator with a variance less than this lower bound.
Before introducing the definition of CRLB, there is a simple estimation example that may could help promote understanding of the basic CRLB concept.

Example 2.1

There is a simple signal transmission model with a transmitted signal $s$, a received signal $r[n]$ and an additive white Gaussian noise $w[n]$.

$$r[n] = s + w[n] \tag{1}$$

Here, the index $n$ refers to the $n$'th observation. In this problem, the transmitted signal $s$ is assumed to be an unknown parameter that is deterministic during $n$ observations. The first idea estimate $s$ takes one observation as our estimation, e.g., the $n$'th observation, namely $\hat{s} = r[n]$. To analyze the estimation accuracy, we check the likelihood function of $r[n]$ as shown.

$$p(r[n];s) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(r[n]-s\right)^2\right] \tag{2}$$

Substituting the estimator we chose in this likelihood function yields

$$p(\hat{s};s) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(\hat{s}-s\right)^2\right] \tag{3}$$

Now, the mean value is the target parameter $s$, and the estimation variance is $\sigma^2$. The estimation accuracy can then be determined as

$$\operatorname{var}(\hat{s}) = \sigma^2 = -E\left(\frac{\partial^2 \ln p(r[n];s)}{\partial s^2}\right)^{-1}. \tag{4}$$

Furthermore, we are interested in finding a more accurate estimator by lowering the variance $\sigma^2$. This can be achieved by exploiting multiple observations. Assuming the observation samples are identical independently distributed, the likelihood function for multiple observations is

$$p(\mathbf{r}[n];s) = \frac{1}{\left(2\pi\sigma^2\right)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N-1}\left(r[n]-s\right)^2\right]. \tag{5}$$

A ML estimator can be derived in the same way as for a single observation to yield

$$\hat{s} = \frac{\sum_{n=1}^{N-1} r[n]}{N}, \tag{6}$$

which is an unbiased estimator, namely $E\{\hat{s}\} = s$. We can also find the estimation variance using equation (4); the result is similar to the single observation MLs with a factor $N$ in the denominator:

$$\operatorname{var}(\hat{s}) = \frac{\sigma^2}{N}. \tag{7}$$

An extreme case occurs when $N$ approaches $\infty$, and the process reduces the estimation variance to 0. From this simple example, we can summarize that the ultimate goal of estimator design is to find the minimum variance unbiased estimator (MVUE), and if we wish to illustrate the performance of our estimator, then estimation variance can be found through the likelihood function. Now, we are ready to define the CRLB (Kay, 1998).

*<Theorem>*

*Assume the pdf, $p(r;\theta)$, satisfies the regularity condition*

$$E_{r;\theta}\left[\frac{\partial \ln p(r;\theta)}{\partial \theta}\right]=0 \quad \text{for all } \theta. \tag{8}$$

*Then, the variance of any unbiased estimator $\hat{\theta}$ has a lower limitation*

$$\operatorname{var}(\hat{\theta}) \geq \frac{1}{-E_{r;\theta}\left[\frac{\partial^2 \ln p(r;\theta)}{\partial \theta^2}\right]}. \tag{9}$$

*An unbiased estimator may be found that attains the bound for all $\theta$ if and only if*

$$\frac{\partial \ln p(r;\theta)}{\partial \theta}=I(\theta)(g(r)-\theta) \tag{10}$$

*for some function $I(\theta)$ and $g(r)$. This estimator can be stated as $\hat{\theta}=g(r)$, which is a MVUE with variance $1/I(\theta)$. To attain the variance lower bound, Fisher's information is defined as*

$$I(\theta)=-E_{r;\theta}\left[\frac{\partial^2 \ln p(r;\theta)}{\partial \theta^2}\right], \tag{11}$$

*which is used to calculate the covariance matrices associated with maximum-likelihood estimates.*
An unbiased estimator that achieves the variance lower bound is referred to as "efficient". In other words, an unbiased estimator that achieves the CRLB is an efficient estimator and must be MVUE. Figures 1 and 2 are illustrations of the relationship between a MVU estimator and the CRLB.

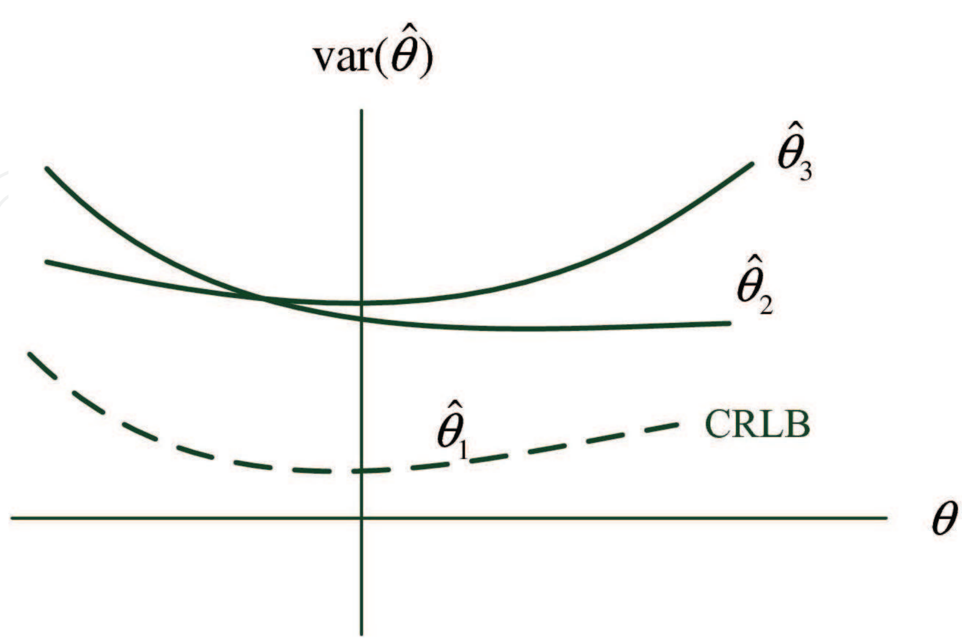



Fig. 1. $\hat{\theta}_1$ MVU and efficient

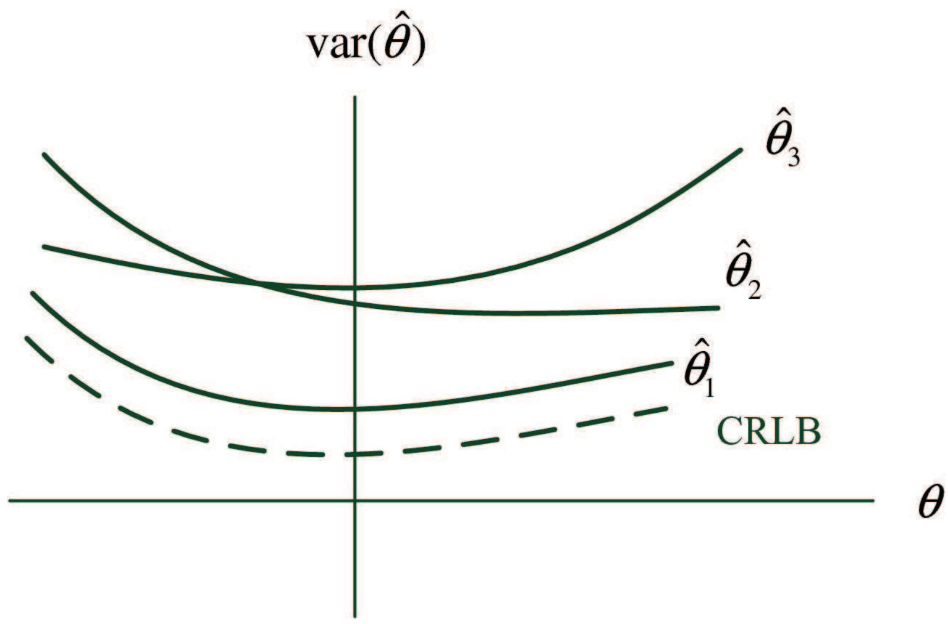



Fig. 2. $\hat{\theta}_1$ MVU and not efficient

Although there are some theories capable of finding MVUE by sufficient statistics and the Rao-Blackwell-Lehmann-Scheffe theorem, we will not introduce the details in this chapter. However, we encourage readers to fully inform themselves concering MVUE from the references in this chapter (Kay, 1998).

A question may be raised concerning why the minimum variance estimator should be an unbiased one. Although the unbiased estimator seems to sucessfully find an perfect estimator $\varphi$ because the expectation value approaches the true parameter i.e., $E[\hat{\theta}] = \theta_0$ , but a biased estimator may outperform than an unbiased one. For example, in some situations, the relationship between a MVUE and a Bayesian MSE estimator may be illustrated in figure 3.

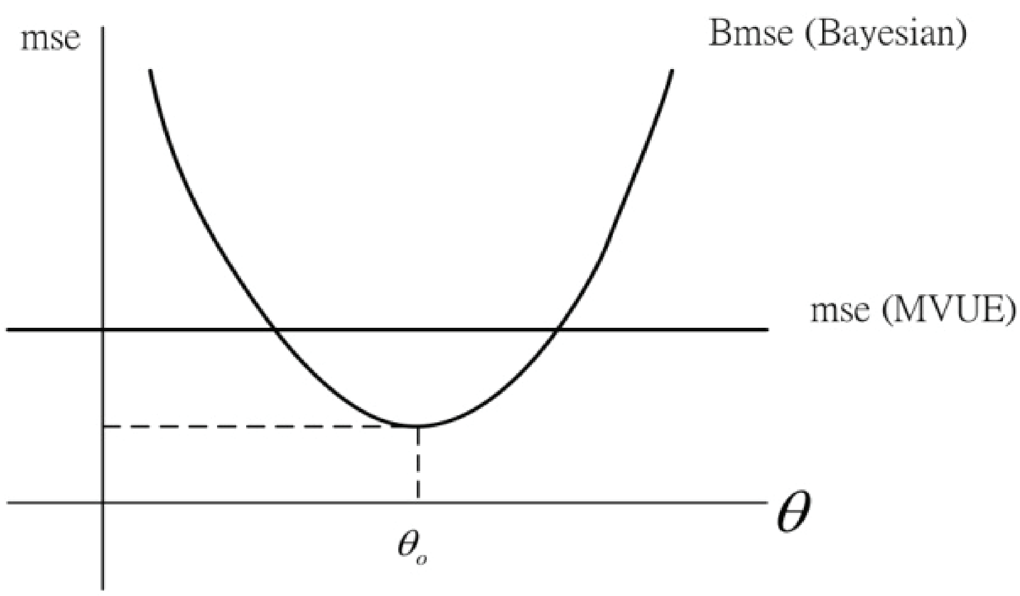



Fig. 3. MVUE vs. Bayesian estimator

In this example, the Bayesian MSE estimator is an unbiased estimator. The performance comparison in figure 3 shows that within a certain parameter interval, the biased Bayesian estimator may have lower estimation variance than MVUE's. However, this comparison also shows that the biased estimator performs terribly outside this interval. Thus, the unbiased estimator has an advantage in terms of consistent performance.

### 2.1 Asymptotic CRLB

For some cases in which the closed form of the CRLB may not be derived, the asymptotic CRLB can be used instead; this form can be attained by assuming that infinite observation samples are available. Under this assumption, we have an observation sample with an infinite signal-to-noise ratio (SNR).

## 3. General case CRLB

### 3.1 Gaussian noise

The AWGN channel is the most common channel model in wireless communication, which was also used in the example in the last section. In example 2.1, we only consider the estimate of symbol $s$. Now, a general form of any parameter $\theta$ is derived.

Example 3.1

Assuming symbol $s$ is transmitted with a general unknown parameter $\theta$ and added with an AWGN $w_n(t)$. The signal model is describe as

$$r_n(t) = s(t;\theta) + w_n(t), \tag{12}$$

where $n$ indicate the $n$ th observation. Following the general CRLB derivation steps, the likelihood function is found first and differentiation with respect to $\theta$ is then performed twice.

$$p(r_n(t);s(t),\theta) = \frac{1}{\left(2\pi\sigma^2\right)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N-1}\left[r_n(t) - s(t;\theta)\right]^2\right] \tag{13}$$

$$\frac{\partial \ln p(r_n(t);s(t),\theta)}{\partial\theta} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left[r_n(t) - s(t;\theta)\right]\frac{\partial s(t;\theta)}{\partial\theta} \tag{14}$$

$$\frac{\partial^2 \ln p(r_n(t);s(t),\theta)}{\partial\theta^2} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(\frac{\partial s(t;\theta)}{\partial\theta}\right)^2 + \left[r_n(t) - s(t;\theta)\right]\frac{\partial^2 s(t;\theta)}{\partial\theta^2} \tag{15}$$

Taking the expectation of $\frac{\partial^2 \ln p(r_n(t);s(t),\theta)}{\partial\theta^2}$ with respect to $p(r;s,\theta)$ into Fisher's information yields

$$I(\theta) = -E_{r;s,\theta}\left\{\frac{\partial^2 \ln p(r_n(t);s(t),\theta)}{\partial\theta^2}\right\} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(\frac{\partial s(t;\theta)}{\partial\theta}\right)^2 \tag{16}$$

Finally, the inverse recipocal of the Fisher's information produced by the CRLB in the AWGN channel.

$$\mathrm{var}(\hat{\theta}) \geq \frac{1}{I(\theta)} = \frac{\sigma^2}{\sum_{n=0}^{N-1}\left(\frac{\partial s(t;\theta)}{\partial\theta}\right)^2} \tag{17}$$

### 3.2 Complex Gaussian channel

Another commonly seen channel is complex Gaussian channel. The mobile communication and wireless communication usually introduce the Rayleigh fading due to multipath delay spread and Doppler shift. In numerical simulation we may use the Jake's (Clarke) model, but in theoretical analysis, complex Gaussian channel is more popular, because it has Rayleigh distributed amplitude with an uniformly distributed phase, which is convenient to use and without loss of generality.

Example 3.2

The signal model can be extended from the general AWGN channel model. We multiply the Rayleigh distributed channel gain $\alpha_0$ and the uniformly distributed channel phase $e^{-j\phi_0}$ with the symbol $s(t;\theta_u)$ .

$$r_n(t) = \alpha_0 e^{-j\phi_0} s(t;\theta_u) + w_n(t) \tag{18}$$

Alternatively, using complex coordinates, i.e., the Gaussian distributed $\alpha_I$ and $\alpha_Q$ with mean $\eta_A$ and variance $\sigma_A^2$ yields

$$r_n(t) = (\alpha_I + j\alpha_Q)s(t;\theta) + w_n(t) \tag{19}$$

Because the $\alpha_I$ , $\alpha_Q$ and $w_n(t)$ terms are Gaussian distributed, the received signal $r_n(t)$ is also Gaussian distributed. To find the joint likelihood function, the mean $m_r$ and variance $\sigma_r^2$ of the received signal should be derived.

$$m_r = \eta_A(1+j)s(t;\theta) \tag{20}$$

$$\sigma_r^2 = 2\sigma_A^2 P_s(t;\theta) + 2\sigma_N^2 \tag{21}$$

Here, $P_s(t;\theta) = s(t;\theta)s(t;\theta)^*$ is the power of the transmitted signal. The joint likelihood function turns out is then described by

$$p_r(r(t);s(t;\theta)) = \frac{1}{\sqrt{2\pi\sigma_r^2}} \exp\left(-\frac{(r(t)-m_r)^2}{2\sigma_r^2}\right) \tag{22}$$

## 4. Random parameter estimation

In previous sections, some basic knowledge of estimation bounds were introduced based on unknown parameters with random interference. These kinds of estimation problems are categorized in the classical estimation approach. Some properties of estimation methods are listed in Table 1.

| | Parameter types | Sample distribution | Parameter distribution |
|---|---|---|---|
| LS | Unknown | Unknown | Non |
| Moment | Unknown | Known | Non |
| MVUE | Unknown | Known | Non |
| Bayesian | Random | Known | Known |
| MAP | Random | Known | Known |
| ML | Both | Known | Uniform |

Table 1. Some estimation properties

Another research area focuses on random parameters estimation, and several approaches, including the Bayesian theorem, MAP and ML, are widely used already. One of the most popular and well-known Bayesian approache is the MMSE estimator. Below, the MMSE will be briefly introduced with an example.

Example 4.1

Assuming that we received signal $r(t)$ that is composed of a random symbol $s$ and white Gaussian noise $w(t)$, the following relationship can be described.

$$r(t) = s + w(t) \tag{23}$$

The conditional pdf of $r(t)$ with a priori information can be stated as

$$p(r(t);s) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left( -\frac{1}{2} \sum_{n=0}^{N-1} (r_n(t) - s)^2 \right) \tag{24}$$

Using Bayes' rule,

$$p(r(t);s) = \frac{p(s;r(t))p(r(t))}{p(s)} \tag{25}$$

After certain computations, the conditional pdf with a posteriori information is obtained as

$$p(s;r(t)) = \frac{1}{\sqrt{2\pi\sigma_{s;r}^2}} \exp\left( -\frac{1}{2\sigma_{s;r}^2} (s - \mu_{s;r})^2 \right), \tag{26}$$

where

$$\sigma_{s;r}^2 = \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\sigma_s^2}}; \tag{27}$$

$$\mu_{s;r} = \left( \frac{N}{\sigma^2} \bar{x} + \frac{\mu_s}{\sigma_s^2} \right) \sigma_{s;r}^2 . \tag{28}$$

The MMSE estimator is then determined as

$$\hat{s} = E\{s \mid r(t)\} = \mu_{s;r} = \alpha \bar{x} + (1 - \alpha)\mu_s \tag{29}$$

where

$$\alpha = \frac{\sigma_s^2}{\sigma_s^2 + \frac{\sigma^2}{N}} \tag{30}$$

The Bayesian mean square error is defined as

$$Bmse(\hat{s}) = E[(s-\hat{s})^2] = \frac{\sigma^2}{N}\left(\frac{\sigma_s^2}{\sigma_s^2 + \frac{\sigma^2}{N}}\right) \le \frac{\sigma^2}{N} \tag{31}$$

As $\sigma_s^2 \to \infty$ i.e., without any information from a prior knowledge, the bound would be the same with the sample mean estimator. This result can be compared with that of the first example in this chapter, and an important concept of Bayesian estimator is revealed: any prior knowledge will result in higher accuracy of the Bayesian estimator.

## 5. Hybrid parameter estimation

In addition to classical estimation and random parameter estimation, there is a more complicated scenario called hybrid parameter estimation. In hybrid parameter estimation, the desired parameter is a vector that is composed of several unknown paramters and random parameters. The parameter vector can be constructed as

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_r^T & \boldsymbol{\theta}_u^T \end{bmatrix}^T , \tag{32}$$

where $\boldsymbol{\theta}_r$ is a random parameter vector and $\boldsymbol{\theta}_u$ is an unknown parameter vector. Because we are considering the random parameters, we assume that we have some prior knowledge of these parameters, such as the probability distribution function. Several techniques for calculating hybrid parameter Cramer-Rao like bounds are described below.

### 5.1 CRLB with nuisance parameter

In our first case, $\boldsymbol{\theta}_r$ is treated as a nuisance parameter, which means that these random parameter are undesired.

Example 5.1

Reformulating the signal model and likelihood function yields

$$r_n(t) = s(t;\boldsymbol{\theta}) + w_n(t) \tag{33}$$

$$p(r_n(t),\boldsymbol{\theta}_r;s(t),\boldsymbol{\theta}_u) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{[r_n(t)-s(t;\boldsymbol{\theta})]^2}{2\sigma^2}\right). \tag{34}$$

Because we assumed that the pdf is well-known and these denoted parameters are unimportant, the marginal likelihood function is derived first, and the nuisance parameters are integrated out of the equation.

$$p(r_n(t);s(t),\boldsymbol{\theta}_u) = \int_{\theta_r} p(r_n(t),\boldsymbol{\theta}_r;s(t),\boldsymbol{\theta}_u)p(\boldsymbol{\theta}_r)d\boldsymbol{\theta}_r \tag{35}$$

Now, the resultant problem becomes a classical estimation problem, and the CRLB can be derived step by step.

1. $$\frac{\partial \ln p(r_n(t);s(t,\boldsymbol{\theta}_u))}{\partial \boldsymbol{\theta}_u} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left[r_n(t)-s(t,\boldsymbol{\theta}_u)\right]\frac{\partial s(t,\boldsymbol{\theta}_u)}{\partial \boldsymbol{\theta}_u} \tag{36}$$

2. $$\frac{\partial^2 \ln p(r_n(t);s(t,\boldsymbol{\theta}_u))}{\partial {\boldsymbol{\theta}_u}^2} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(\frac{\partial s(t,\boldsymbol{\theta}_u)}{\partial \boldsymbol{\theta}_u}\right)^2 + \left[r_n(t)-s(t,\boldsymbol{\theta}_u)\right]\frac{\partial^2 s(t,\boldsymbol{\theta}_u)}{\partial {\boldsymbol{\theta}_u}^2} \tag{37}$$

3. $$I(\boldsymbol{\theta})_{i,j} = E_r\left\{\frac{\partial \ln p(r_n(t);s(t,\boldsymbol{\theta}_u))}{\partial \theta_i}\frac{\partial \ln p(r_n(t);s(t,\boldsymbol{\theta}_u))}{\partial \theta_j}\right\} \tag{38}$$

4. $$CRLB(\hat{\theta}_i) = \left[\frac{1}{I(\boldsymbol{\theta})}\right]_{i,i} \le \mathrm{var}(\hat{\theta}_i) \tag{39}$$

### 5.2 Hybrid CRLB

In some scenarios, the effect of these ramdom parameters cannot be ignored. Another method that considers the joint pdf called joint estimation. The CRLB for this kind of joint estimation is called hybrid Cramer-Rao bound (HCRB). The derivation process is nearly identical to that of ordinary CRLB; the likelihood function is determined, and partial differentiation with respect to the desired parameter is performed twice.

$$r_n(t) = s(t;\boldsymbol{\theta}) + w_n(t) \tag{40}$$

$$p(r_n(t),\boldsymbol{\theta}_r;s(t),\boldsymbol{\theta}_u) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{\left[r_n(t)-s(t;\boldsymbol{\theta})\right]^2}{2\sigma^2}) \tag{41}$$

$$\frac{\partial \ln p(r_n(t),\boldsymbol{\theta}_r;s(t),\boldsymbol{\theta}_u)}{\partial \boldsymbol{\theta}} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left[r_n(t)-s(t;\boldsymbol{\theta})\right]\frac{\partial s(t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{42}$$

$$\frac{\partial^2 \ln p(r_n(t),\boldsymbol{\theta}_r;s(t),\boldsymbol{\theta}_u)}{\partial \boldsymbol{\theta}^2} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(\frac{\partial s(t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2 + \left[r_n(t)-s(t;\boldsymbol{\theta})\right]\frac{\partial^2 s(t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \tag{43}$$

Because the joint pdf is considered, the expection of Fisher's information should be taken with respect to $p(r(t),\theta_r)$

$$I(\boldsymbol{\theta})_{i,j} = E_{r,\boldsymbol{\theta}_r}\left\{\frac{\partial \ln p(r_n(t),\boldsymbol{\theta}_r;s(t),\boldsymbol{\theta}_u)}{\partial \theta_i}\frac{\partial \ln p(r_n(t),\boldsymbol{\theta}_r;s(t),\boldsymbol{\theta}_u)}{\partial \theta_j}\right\} \tag{44}$$

The joint pdf $p(r(t),\theta_r)$ is not easy to determine, and an alternative approach using double layer expectation which computes the expectation with respect to the conditional pdf first. We define the information matrix with respect to the conditional pdf $p(r(t);\theta_r)$ as

$$I(\boldsymbol{\theta}_0)_{i,j} = E_{r;\boldsymbol{\theta}_r}\left\{\frac{\partial \ln p(r_n(t),\boldsymbol{\theta}_r;s(t),\boldsymbol{\theta}_u)}{\partial \theta_i}\frac{\partial \ln p(r_n(t),\boldsymbol{\theta}_r;s(t),\boldsymbol{\theta}_u)}{\partial \theta_j}|\boldsymbol{\theta}_r\right\}. \tag{45}$$

Then expectation is computed with respect to $p(\theta_r)$, and all of the random pararameters are eliminated.

$$\begin{aligned} I(\boldsymbol{\theta})_{i,j} &= E_{\boldsymbol{\theta}_r}\left\{E_{r;\boldsymbol{\theta}_r}\left\{\frac{\partial \ln p(r_n(t),\boldsymbol{\theta}_r;s(t),\boldsymbol{\theta}_u)}{\partial \theta_i}\frac{\partial \ln p(r_n(t),\boldsymbol{\theta}_r;s(t),\boldsymbol{\theta}_u)}{\partial \theta_j}|\boldsymbol{\theta}_r\right\}\right\} \\ &= E_{\boldsymbol{\theta}_r}\left\{I(\boldsymbol{\theta}_0)_{i,j}\right\} \end{aligned} \tag{46}$$

Finally, the HCRB is derived as

$$HCRB(\hat{\theta}_i) = \left[\frac{1}{I(\boldsymbol{\theta})}\right]_{i,i} \le \mathrm{var}(\hat{\theta}_i)\,. \tag{47}$$

### 5.3 Modified CRLB

During the process of deriving the HCRB, an important step involves taking the inverse of the Fisher's information matrix. In some cases, the inverse of the Fisher's information matrix may not exist or cannot be derived into a closed form lower bound. We can then try the modified or simplified bound, such as the MCRB. Instead of taking the inverse of the matrix first, we select the desired estimation element from the information matrix first and then execute the inverse step. After choosing the desired estimation element, the Fisher's information is no longer in a matrix form, and derivation is easier.

$$MCRB(\hat{\theta}_i) = \left[\frac{1}{I(\boldsymbol{\theta})_{i,i}}\right] \le \mathrm{var}(\hat{\theta}_i) \tag{48}$$

An previously reported example can help distinguish the difference between these CR-like bounds (F. Gini, 2000).

Example 5.2

When considering a data-aided joint frequency offset estimation case, the signal model can be described as

$$r_n(t) = Ae^{-j2\pi f_D t}s(t) + w_n(t) \tag{49}$$

Here, $A$ is the complex channel, which can be rewritten as $A = \alpha_0 e^{-j\phi_0} = \alpha_I + j\alpha_Q$, and $e^{-j2\pi f_D t}$ represents the frequency offset. The estimation parameter matrix $\boldsymbol{\theta} = [f_D \quad \alpha_I \quad \alpha_Q]^T$ can be defined. Because this is a data-aided case, $s(t)$ can be a pilot or preamble, and we can assume that $s(t)s(t)^* = 1$ without loss of generality. Then the signal after pilot removal is

$$\begin{aligned} x_n(t) &= r_n(t)s(t)^* \\ &= (\alpha_I + j\alpha_Q)e^{j2\pi f_D t} + v_n(t) \end{aligned} \tag{50}$$

$x_n(t)$ is also Gaussian distributed. Following the derivation of S. M. Kay (1998) and F. Gini (2000), we can find the conditional Fisher's information matrix.

$$I(\theta_0)=\begin{bmatrix} \frac{2\pi^2 N(N-1)(2N-1)}{3\sigma_N^2}(\alpha_I^2+\alpha_Q^2) & -\frac{\pi N(N-1)}{\sigma_N^2}\alpha_Q & \frac{\pi N(N-1)}{\sigma_N^2}\alpha_I \\ -\frac{\pi N(N-1)}{\sigma_N^2}\alpha_Q & \frac{N}{\sigma_N^2}+\frac{(\alpha_I-\eta_A)^2}{\sigma_A^2} & 0 \\ \frac{\pi N(N-1)}{\sigma_N^2}\alpha_I & 0 & \frac{N}{\sigma_N^2}+\frac{(\alpha_Q-\eta_A)^2}{\sigma_A^2} \end{bmatrix} \quad (51)$$

By computing the expectation of $\alpha$, the Fisher's information for the frequency offset is

$$I(f_D)=E_\alpha\{I(\theta_0)\} \quad (52)$$

Then the MCRB is derived as

$$MCRB(f_D)=\frac{1}{\left[I(f_D)\right]_{11}}=\frac{3}{4\pi^2 N(N-1)(2N-1)\rho} \quad (53)$$

where $\rho=(\eta_A^2+\sigma_A^2)/\sigma_N^2$ is the SNR. Now, the difference between the MCRB and the HCRB can be checked. As mentioned previously, the HCRB is

$$\begin{aligned} HCRB(f_D) &= \left[\frac{1}{I(f_D)}\right]_{11} \\ &= \frac{3(K_R+1)(K_R+1+N\rho)}{2\pi^2 N(N-1)}\frac{1}{2(2N-1)(K_R+1)(K_R+1+N\rho)-3N(N-1)\rho K_R} \end{aligned}, \quad (54)$$

where $K_R=\eta_A^2/\sigma_A^2$ is the Rice factor, which is the power ratio between direct path signal and other scatter path signals. A comparison of the HCRB and MCRB can be evaluted as.

$$\frac{HCRB(f_D)}{MCRB(f_D)}=\frac{2(2N-1)(K_R+1)(K_R+1+N\rho)}{2(2N-1)(K_R+1)(K_R+1+N\rho)-3N(N-1)\rho K_R}. \quad (55)$$

Based on the equation above, in the general case, the ratio is always larger than 1, which means that the HCRB is generally a tighter bound than the MCRB. Conversely, when $K_R\to 0$ or $K_R\to\infty$, the ratio of HCRB to MCRB approaches 1. It is interesting that these two bounds only meet for two extreme scenarios, namely the Rayleigh channel and direct path.

### 5.4 Miller Chang bound

The Miller Chang bound (MCB) is proposed by R. W. Miller and C. B. Chang (1978). They state that the MCB can apply to a more restricted class of estimator that is unbiased for each value of the nuisance parameter, which is referred to as locally unbiased, whereas the standard Cramer-Rao bound (CRB) can applies to any estimators that are unbiased over the ensemble. The Miller Chang bound is defined as

$$MCB(\hat{\theta}_i) = E_{\boldsymbol{\theta}_r}\left\{\frac{1}{I(\boldsymbol{\theta}_0)_{i,i}}\right\}. \tag{56}$$

The MCB has a similar form to the MCRB, but the MCB is always tighter than the MCRB. More directly, the MCB applies to more restricted estimators than the CRLB, which implies that the MCB is tighter than CRB, and the MCRB is looser than the CRB, which was derived by A. N. D'Andrea (1994). Therefore, the MCB is tighter than the MCRB. Alternatively, we can also explain this relationship using Jensen's inequality for any convex function $\varphi$ and random variable $x$

$$\varphi\left(E[x]\right) \le E\left[\varphi(x)\right]. \tag{57}$$

In our case, the inverse function for a positive defined matrix is a convex function, so

$$MCRB(\hat{\theta}_i)=\frac{1}{E_{\boldsymbol{\theta}_r}\left\{I(\boldsymbol{\theta}_0)_{i,i}\right\}} \le E_{\boldsymbol{\theta}_r}\left\{\frac{1}{I(\boldsymbol{\theta}_0)_{i,i}}\right\}=MCB(\hat{\theta}_i). \tag{58}$$

Now, from example 5.2 in the MCRB subsection, the MCB of the joint estimated frequency offset is

$$\begin{aligned} MCB(f_D) &= \frac{3\sigma_N^2}{2\pi^2 N(N-1)(2N-1)} E_\alpha\left\{\frac{1}{\alpha_I^2+\alpha_Q^2}\right\} \\ &= MCRB(f_D) E_\alpha\left\{\frac{2(\eta_A^2+\sigma_A^2)}{\alpha_I^2+\alpha_Q^2}\right\} \end{aligned} \tag{59}$$

The final result still remains the expectation term, so it cannot be derived into a closed form. Although the MCB is a tighter bound than the MCRB, the MCRB is more likely to derive into a closed form. In addition, the MCB requires a locally unbiased estimator, which is also a harsh restriction for estimator design, so the MCRB is more popular for theoretical analysis.

### 5.5 Summary of the relationship between Cramer-Rao-like bounds

Some of the relationship between Cramer-Rao-like bounds has been derived previously (Reuven, 1997). In this work, they consider the signal model with Gaussian distributed channel gain and an unknown timing delay. We can also derive this relationship from our examples in subsection 5. Following from example 5.1, if we carry through the calculation to the end, then we will obtain the marginal CRB of the frequency offset $f_D$.

$$CRB(f_D) = \frac{3(K_R+1+N\rho)}{2\pi^2 N(N-1)\rho[N(N+1)\rho+2(2N-1)K_R]} \tag{60}$$

Then, this result is compared with that for the HCRB, which was derived in equation (55).

$$CRB(f_D) = HCRB(f_D)\frac{2(2N-1)(K_R+1)(K_R+1+N\rho)-3N(N-1)\rho K_R}{(K_R+1)[N(N+1)\rho+2(2N-1)K_R]} \tag{61}$$

After calculations, the CRB can be summarized into the HCRB multiplied by a function. We simplified the fraction in equation (62) and found that it is larger than 1 only if $N < 1$. This result implies that $CRB(f_D) \geq HCRB(f_D)$, and the relationship $HCRB(f_D) \geq MCRB(f_D)$ has been proven by equation (56). Another way to prove this is to use a corollary.

"For any positive defined matrix $M$, $\left[M^{-1}\right]_{11} \geq \left[M_{11}\right]^{-1}$, an equal occur if $M$ is diagonal".

Finally, we summarize the relationship between CRB, HCRB and MCRB as

$$CRB(f_D) \geq HCRB(f_D) \geq MCRB(f_D) \tag{62}$$

However, the relationship between the MCB and MCRB was also derived in equations (58-59) using Jensen's inequality. Because the MCRB seems to be a looser bound in the Cramer-Rao-like bounds family, we normalized all other bounds to the MCRB, as shown in figure 4.

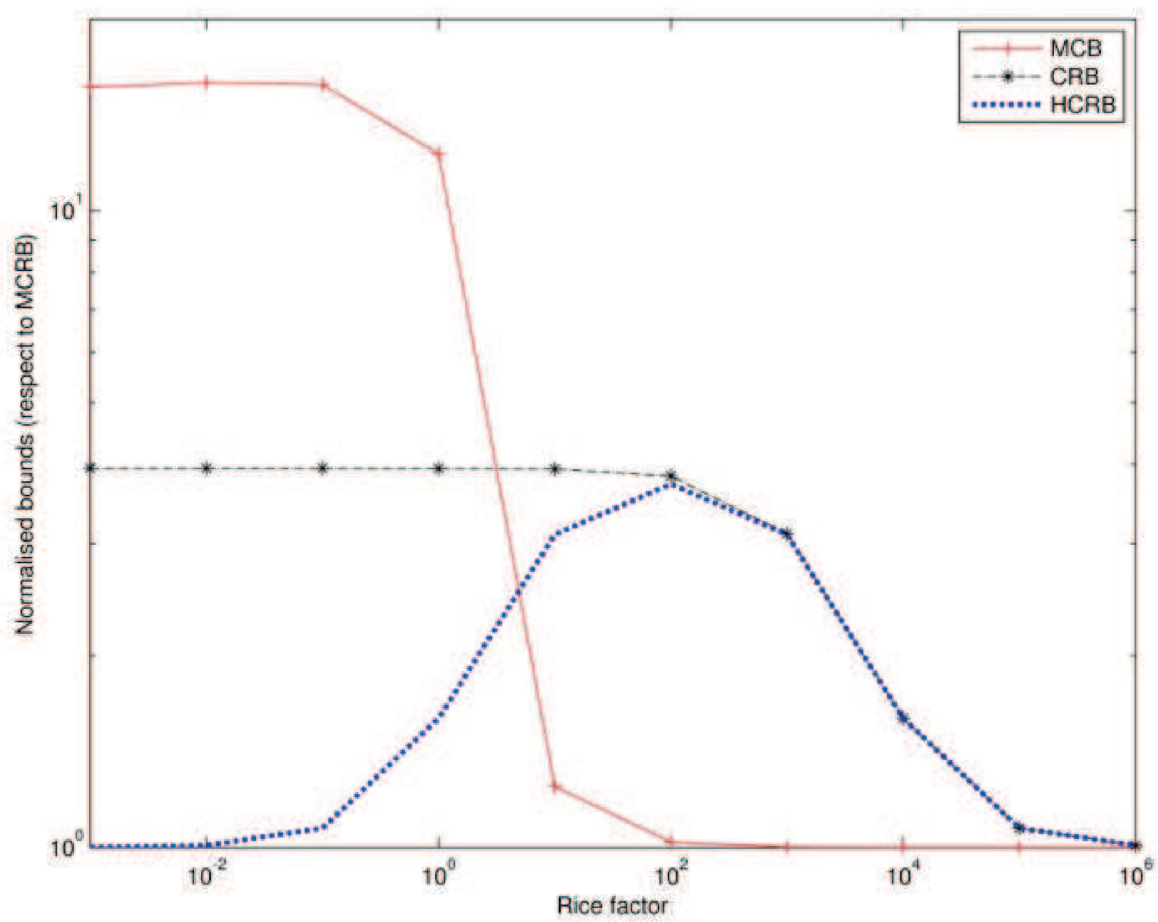



Fig. 4. Normalized bounds versus the Rice factor

From the figure above, the MCB exhibits drastic variation near $K_R = 1$, which indicates that the locally unbiased estimatior of $f_D$ is difficult to find when the power of scatter signal is larger than line-of-sight (LOS) signal. Moreover, when $K_R = 0$ (Rayleigh channel), the ratio of the normalized MCB approaches infinity, which means that no locally unbiased estimator exists. The Rayleigh fading channel is the most frequently used channel model in a wireless communication environment, which is another reason that the MCRB is more popular than the MCB. In multiple parameters estimation, joint estimation techniques have been a popular topics recently. In terms of hybrid parameter joint estimation, the benchmark for comparison with is the HCRB. Based on equation (56) and figure 3, the HCRB has a feature that approaches the MCRB when $K_R \to 0$ or $K_R \to \infty$. As mentioned previously, the scenario $K_R \to 0$ implies the Rayleigh channel. The analysis shows that the MCRB is quite sufficient as a benchmark to design an estimator in the Rayleigh channel environment.

Some prior research has been reported on the relationship among the joint estimate initial phase, timing delay and frequency offset (D'Andrea , 1994). The author summarized and derived some cases in which the CRB is equal to the MCRB.

i. Estimation of $\phi$ when $f_D$, $\tau$ and data are known
ii. Estimation of $\tau$ when $f_D$, $\phi$, and data are known
iii. Estimation of $f_D$ with M-PSK modulation, when $\tau$ and differential data are available but $\phi$ is unknown.

Here, $\phi$, $f_D$ and $\tau$ are the initial phase , frequency offset and timing delay. Other cases may exist in which the CRB is equal to the MCRB, but these cases are difficult to analyze. An important conclusion here is that if an estimator approaches the MCRB, then the MCRB must be closed to the CRB.

## 6. Advanced topics

### 6.1 Carrier phase and clock recovery

As summarized by A. N. D'Andrea (1994), there are several synchronization techniques that can attain or approach the MCRB for a carrier phase $\theta$ and timing $\tau$ estimation. Under the assumption that the frequency offset and timing are known, $MCRB(\theta)$ can be attained using two algorithms.

i. Maximum likelihood decision-directed (ML-DD), proposed in H. Kobayashi (1971)
ii. Ad hoc non-data-aided (ad hoc NDA) method, proposed by A. J. Viterbi (1983).

The $MCRB(\tau)$ can also be attained using the ML-DD algorithm with derivative-matched filters (DMFs); however, the use of DMFs also makes the estimator impractical to implement. Several alternative algorithma have been found that can approach $MCRB(\tau)$ without using DMFs.

i. DD early-late scheme with $T/2$ sample space, proposed by T. Jesupret (1991).
ii. DD scheme, proposed by K. H. Mueller (1976).
iii. NDA scheme, proposed by F. M. Gardner (1986).

Although these alternative algorithms can approach $MCRB(\tau)$ without using DMFs, they are subject to some restrictions that require $\theta$ to be known and a roll-off factor $\alpha$ that should be small.

### 6.2 Frequency offset estimation

In this subsection, three practical carrier frequency estimation techniques are overviewed and compared with the popular MCRB.

**A. NDA loop algorithm**

The first algorithm is a non-data-aided carrier frequency estimation; a block diagram representing this algorithm is shown in figure 5. The received signal $r(t)$ first passes through the matched filter $G^*(f)$ and the so-called "frequency-matched filter" $dG^*(f)/df$. Assuming that the timing is perfectly synchronized, the frequency error is described as

$$e_k = \mathrm{Re}\{x_k y_k^*\}. \tag{63}$$

Then, the frequency error passes through a loop filter and triggers the voltage-control oscillator (VCO) to compensate for the frequency offset. If the loop filter is implemented by a simple digital integrator, then the VCO output can be written as

$$\hat{f}_D(k+1) = \hat{f}_D(k) + \gamma e_k \tag{64}$$

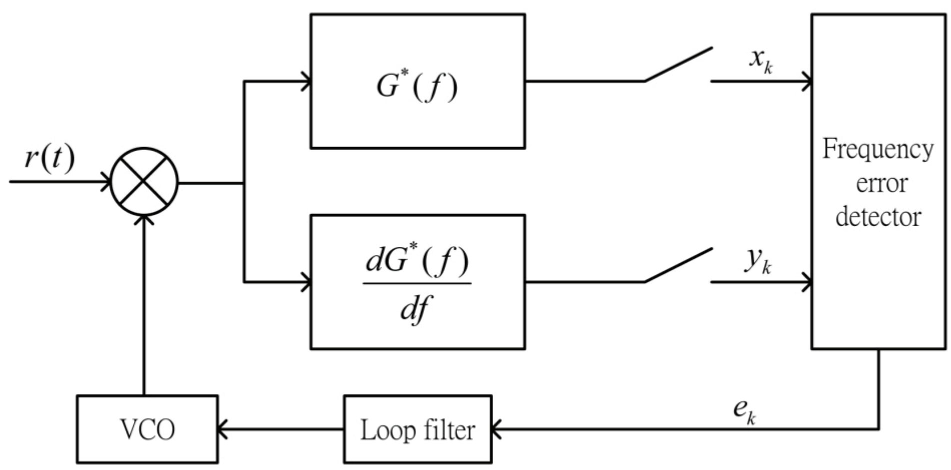



Fig. 5. NDA loop algorithm

The next step is to evaluate the estimation noise performance. There are three assumptions

i. The frequency errors are small as compared to the symbol rate.
ii. The pulse shaping filter $G^*(f)$ is a root-raised cosine function with a roll-off factor $\alpha$.
iii. There is perfect timing delay synchronization

Under these assumptions, the frequency jitter is minimized, and the estimation variance of $f_D$ is derived to be

$$\sigma_{f_D}^2 = \frac{4\alpha B_L T}{\pi^2 T^2} \frac{1}{E_s/N_0}\left(1 + \frac{1}{E_s/N_0}\right), \tag{65}$$

where $B_L$ is the loop noise bandwidth and $T$ is the symbol duration.

**B. Differential decision-directed algorithm**

The second algorithm is a differential decision-directed (DDD) algorithm that is used on PSK signals; the block diagram for this algorithm is shown in figure 5. This algorithm is similar to the NDA algorothm except for the frequency error generator. The assumptions for this algorithm include the following:

i. The frequency errors are small compared to the symbol rate.
ii. $G^*(f)$ is the same as was defined previously.
iii. Timing is perfectly synchronized.

Because we are discussing the M-PSK signal, we can denote our symbol by

$$c_k = \exp(j\varphi_k)\,, \tag{66}$$

where

$$\varphi_k = 2\pi n / M,\ \ n = 1,2...M\,. \tag{67}$$

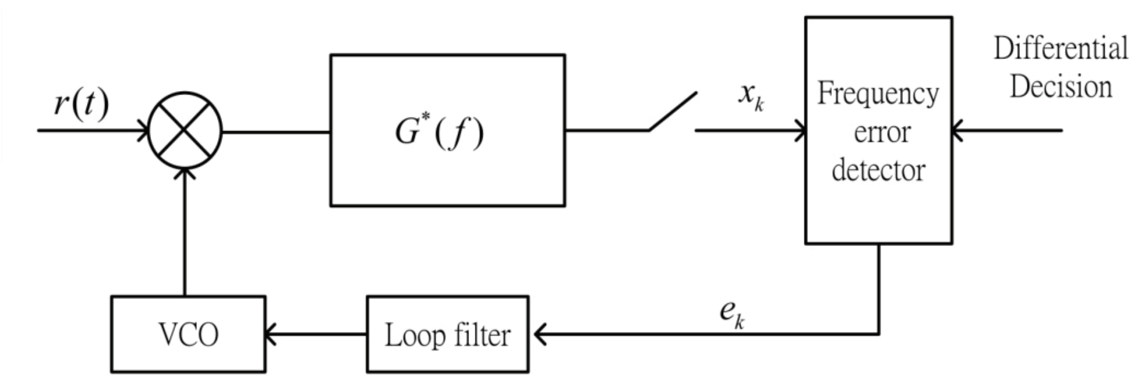



Fig. 6. Differential decision-directed (DDD) algorithm

Then the phase difference between $x_k$ and $x_{k-1}$ will be

$$\Delta\phi_k = \Delta\varphi_k + (f_D - \hat{f}_D)T + \delta_k\,, \tag{68}$$

where $\Delta\varphi_k$ is due to modulation, $(f_D - \hat{f}_D)$ is caused by estimation error, and $\delta_k$ is the phase noise with other interferences, which can be modeled as an uniformly distributed random variable from $-\pi$ to $\pi$. When the difference between $x_k$ and $x_{k-1}$ is correct, perfect $\Delta\hat{\varphi}_k = \Delta\varphi_k$ is obtained. The most important component of this block diagram is the frequency error that is defined as

$$e_k = \mathrm{Im}\left\{x_k x_{k-1}^* \exp(-j\Delta\hat{\varphi}_k)\right\} \tag{69}$$

The performance of the estimator is then

$$\sigma_{f_D}^2 = \frac{B_L T}{\pi^2 T^2}\frac{1}{E_s/N_0}\left(2B_L T + \frac{1}{E_s/N_0}\right) \tag{70}$$

Prior to the simulation, we assume that $B_L T = 5\times10^{-3}$ and the QPSK signals have a roll-off factor $\alpha = 0.5$. The result is compared with the MCRB in figure 7.
As shown in figure 7, these two algorithms yield much greater variance than the MCRB, which indicates that there is still room for improvement.

**C. Feed-forward NDA**

The third algorithm is the feed-forward NDA for M-PSK signal modulation; the block diagram for this algorithm is shown in figure 8. The $F$ function in the middle of the block diagram is a 4th-powered non-linear function. Similar to the previous analyses , the received phase can be separated into three parts:

i. A step-wise increasing quantity $2\pi M f_D kT$ due to the frequency error $f_D$.
ii. A constant initial phase.

iii. A phase noise caused by thermal noise and inter-symbol interference that is uniformly distributed from $-\pi$ to $\pi$.

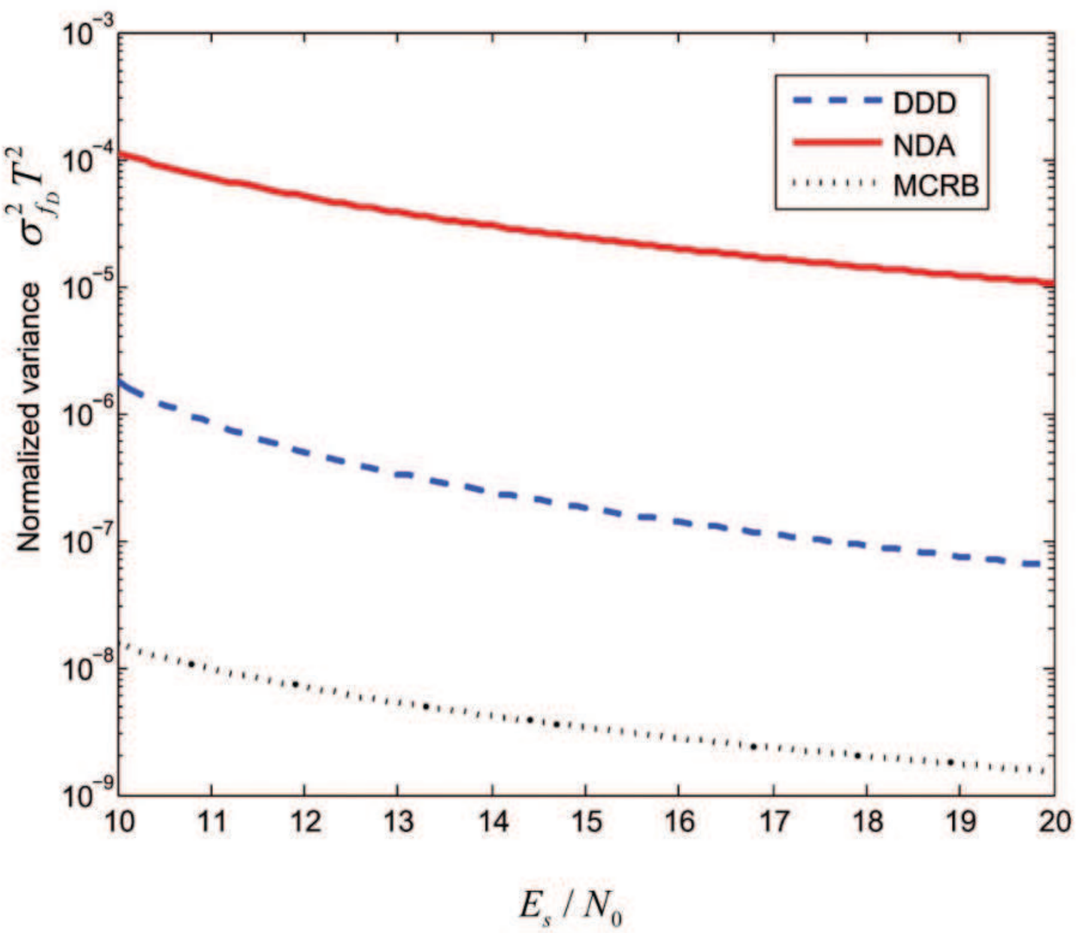



Fig. 7. Comparison of the variance of the two algorithms with that of the MCRB

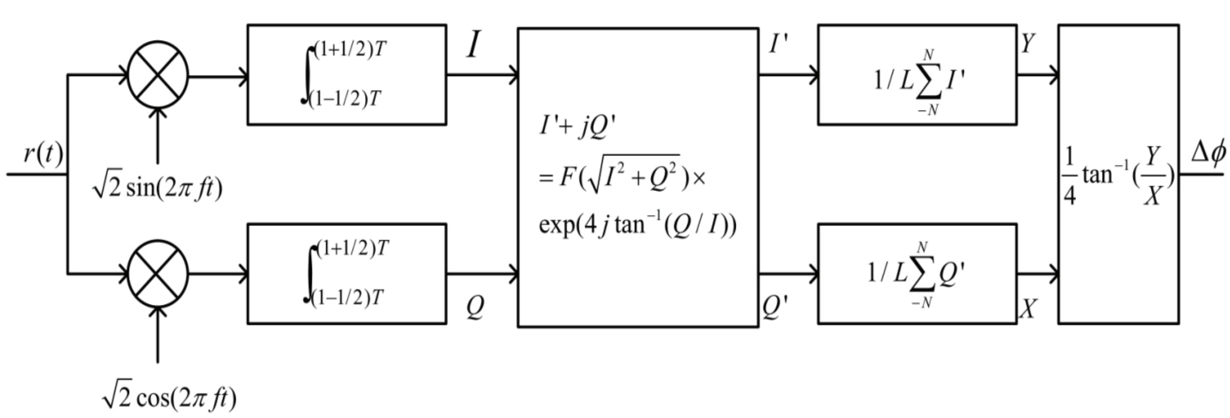



Fig. 8. Feed-forward NDA

The estimation variance has been derived (Bellini, 1990) in a scenario with a very high SNR, the estimation variance can be approached as

$$\sigma_{f_D}^2 \approx \frac{3}{2\pi^2 T^2 L(L^2\text{-}1)m} \frac{1}{E_s/N_0} \tag{71}$$

The MCRB in this case is

$$MCRB(f_D) = \frac{3T}{2\pi^3 (LT)^3} \frac{1}{E_s/N_0} \tag{72}$$

Thus, when $L \gg 1$ and $m = 1$,the algorithm performance will attain the MCRB. However, this result is obtained under very high SNR. Further research is needed to design estimators that can approach or attain the estimation bounds with less restriction.

## 7. References

Bellimi, S., Molinari, C. and Tartara, G. (1990). Digital Frequency Estimation in Burst Mode QPSK Transmission, *IEEE Trans. Commun.*, Vol.38, No.7 , (July 1990), pp. 959-961, ISSN: 0090-6778

Cramer, H. (1946). *Mathematical Method of Statistics*, Princeton University Press, ISBN-13: 978-0691005478, Uppsala, Sweden.

D'Andrea, A. N., Mengali, U. and Reggiannini, R. (1994). The Modified Cramer-Rao Bound and Its Application to Synchronization Problems, *IEEE Trans. Commun.*, Vol.42, No.2/3/4, (Febuary 1994), pp. 1391-1399, ISSN: 0090-6778

Gini, F. and Reggiannini, R. (2000). On the Use of Cramer-Rao-Like Bounds in the Presence of Random Nuisance Parameters, *IEEE Trans. Commun.*, Vol.48, No.12, (December 2000), pp. 2120-2126, ISSN 0090-6778.

Gardner, F. M. (1986). A BPSK/QPSK Timing Error Detecor for Samples Receivers, *IEEE Trans. Commun.*, Vol.34, No.5, (May 1986), pp. 423-429, ISSN: 0090-6778

Jesupret, T., Moeneclaey, M. and Ascheid, G. (1991). Digital Demodulator Synchronization, ESA Draft Final Report, ESTEC No. 8437-89-NL-RE., (Febuary 1991)

Kay, S. M. (1998). *Fundamentals of Statistical Signal Processing,* Prentice Hall, ISBN 0-13-345711-7, Upper Saddle River, New Jersey

Kobayashi, H. (1971). Simultaneous Adaptive Estimation and Decision Algorithm for Carrier Modulated Data Transmission Systems, *IEEE Trans. Commun.*, Vol.19, No.3, (June 1971), pp. 268-280, ISSN: 0018-9332

Kotz, S. and Johnson, N. L. (1993). *Breakthroughs in Statistics: Volume 1: Foundations and Basic Theory*, Springer-Verlag, ISBN: 0387940375, New York.

Lin, J. C. (2003). Maximum-Likelihood Frame Timing Instant and Frequency Offset Estimation for OFDM Communication Over A Fast Rayleigh Fading Channel, *IEEE Trans. Vehic. Technol.,* Vol.52, No.4, (July 2003), pp. 1049-1062.

Lin, J. C. (2008). Least-Squares Channel Estimation for Mobile OFDM Communication on Time-Varying Frequency-Selective Fading Channels, *IEEE Trans. Vehic. Technol.,* Vol.57, No.6, (November 2008), pp. 3538-3550.

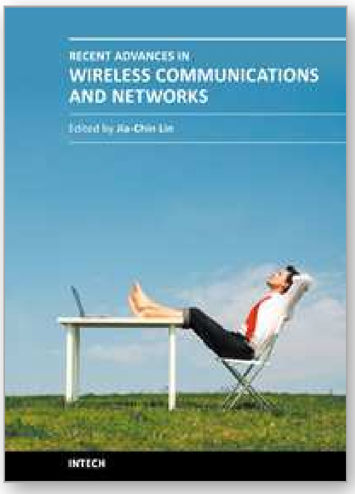
Lin, J. C. (2009). Least-Squares Channel Estimation Assisted by Self-Interference Cancellation for Mobile PRP-OFDM Applications, *IET Commun.,* Vol.3, Iss.12, (December 2009), pp. 1907-1918.

Mueller, K. H. and Muller, M. (1976). Timing Recovery in Digital Synchronous Data Receivers, *IEEE Trans. Commun.*, Vol.24, No.5, (May 1976), pp. 516-530, ISSN: 0090-6778.

Miller, R. W. and Chang, C. B. (1978). A Modified Cramer-Rao Bound and its Applications, *IEEE Trans. On Inform. Throey,* Vol.IT-24, No.3, (May 1978), pp-389-400, ISSN : 0018-9448

Poor, H. V. (1994). *An Introduction to Signal Detection and Estimation*, Springer-Verlag, ISBN: 0-387-94173-8, New York.

Viterbi, A. J. and Viterbi, A. M. (1983). Nonlinear Estimation of PSK-Modulated Carrier Phase with Application to Burst Digital Transmission, *IEEE Trans. Inform. Throey,* Vol.IT-29, No.3, (July 1983), pp. 543-551, ISSN : 0018-9448.




**Recent Advances in Wireless Communications and Networks**
Edited by Prof. Jia-Chin Lin

ISBN 978-953-307-274-6
Hard cover, 454 pages
**Publisher** InTech
**Published online** 23, August, 2011
**Published in print edition** August, 2011

This book focuses on the current hottest issues from the lowest layers to the upper layers of wireless communication networks and provides â€œreal-timeâ€ research progress on these issues. The authors have made every effort to systematically organize the information on these topics to make it easily accessible to readers of any level. This book also maintains the balance between current research results and their theoretical support. In this book, a variety of novel techniques in wireless communications and networks are investigated. The authors attempt to present these topics in detail. Insightful and reader-friendly descriptions are presented to nourish readers of any level, from practicing and knowledgeable communication engineers to beginning or professional researchers. All interested readers can easily find noteworthy materials in much greater detail than in previous publications and in the references cited in these chapters.

**How to reference**
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Kao-Peng Chou and Jia-Chin Lin (2011). A Study of Cramér-Rao-Like Bounds and Their Applications to Wireless Communications, Recent Advances in Wireless Communications and Networks, Prof. Jia-Chin Lin (Ed.), ISBN: 978-953-307-274-6, InTech, Available from: http://www.intechopen.com/books/recent-advances-in-wireless-communications-and-networks/a-study-of-crame-r-rao-like-bounds-and-their-applications-to-wireless-communications

INTECH
open science | open minds