# PhD THESIS

presented in order to obtain the degree of:

## Docteur de l'Ecole Nationale Supérieure des Télécommunications

Speciality : Signal and Images

## Mobile Localization in a GSM Network
Localisation d'un mobile dans un réseau GSM

## Hassan El Nahas El Homsi

**Defended on June 29, 1999, before the committee composed of:**

| | |
|---|---|
| Björn OTTERSTEN | Chairman |
| Jean-Jacques FUCHS | Referees |
| Eric MOULINES | |
| Pierre HUMBLET | Examiners |
| Philippe LOUBATON | |
| Dirk SLOCK | |
| Jean-Louis DORNSTETTER | Guests |
| André TERROM | |

# Acknowledgments

First of all, I would like to thank Pierre Humblet, my thesis supervisor for his continuous help and for motivating me.

I wish to thank Professor Björn Ottersten from KTH for welcoming me in his department in Stockholm. The two weeks of discussions I had with him and his students were very fruitful. I wish also to thank him for accepting to be an examiner of my thesis.

I equally thank Professor Eric Moulines for the several discussions I have had with him during the thesis and for accepting to be an examiner of my thesis.

I also thank Professor Jean-Jacques Fuchs for accepting to be an examiner of my thesis and for his detailed reading of the manuscript and his positive feedback.

I also wish to thank Professor Dirk Slock and Philippe Loubaton for accepting to be on the Jury and for their general interest in my work.

I would like to thank all the people who surrounded me during the thesis in Nortel Networks. A special thank is for Alexandre with whom I have had a lot of discussions, his interest in my work is deeply appreciated. Many thanks to Sophie, Anne, Sarah, Bastien, and Moussa who accepted to check my poor english.

This paragraph would be of course incomplete if I didn't mention Jean-Louis Dornstetter, Nidham Ben Rached, and Pierre Eisenmann who accepted me in the R&D department in Nortel Networks. I also wish to thank Bouygues Telecom for the financial support of this work.

I finally want to express my deep gratitude to my parents for their constant love and encouragement and for all they have sacrificed for the success of their children. It is to them that I dedicate this thesis.

# Abstract

The localization of a subscriber in a radio cellular network has attracted considerable interest since the American Federal Communication Committee (FCC) mandated all operators in the United States to localize their subscribers within 125 meters in 67 per cent of the cases by October 2001. This concerns essentially emergency calls (E 911 calls). In addition to that, the localization technology has several attractive applications such as navigation, home zone billing, fraud detection, and frequency planning enhancement. The North American standardization committee has worked hard towards solving this issue for the various cellular standards. Many proposals have been presented by manufacturers. Their work shows that the natural solution for the GSM standard should be based on the time of arrival technology.

The main obstacle in time of arrival estimation is multipath. The goal is to be able to estimate the time delay of the first path. A class of estimators based on the extraction of the signal or the noise subspace is introduced. These estimators offer almost the same performances as the maximum likelihood with lower complexities. An extension of these algorithms under model errors is introduced.

With at least three time of arrival measurements corresponding to three different base stations, it is possible to locate the handset under some conditions on network synchronization. The maximum likelihood estimator leads to a non linear maximization problem known as hyperbolic trilateration. Suboptimal algorithms are presented offering good results at high signal-to-noise ratios. Complete simulations are conducted in several typical environments such as urban or rural areas incorporating synchronization errors. They show that a root mean square error lower than one hundred meters is achievable in most cases.

# Résumé

La localisation de mobiles dans un réseau radio cellulaire a reçu un intérêt considérable depuis que le comité fédéral américain de communication a demandé aux opérateurs nord-américains de localiser leurs abonnés avec une précision de 125 mètres dans 67 pour cent des cas. Ceci concerne essentiellement les appels d'urgence. Outre les appels d'urgence, il existe d'autres applications comme la navigation, la gestion de la taxation, la détection de fraudes et la planification cellulaire. Les travaux conduits par les comités américains de normalisation ont privilégié les solutions basées sur l'estimation du temps d'arrivée.

L'obstacle principal à l'estimation du temps d'arrivée est le trajet multiple. Il faut pouvoir estimer le retard temporel du premier trajet. Une classe d'estimateurs basée sur l'extraction des sous-espaces signal ou bruit est introduite. Ces estimateurs offrent quasiment les même performances que l'estimateur du maximum de vraisemblance tout en ayant une complexité moindre. Une extension de ces estimateurs en présence d'erreurs de modèle est présentée.

Avec un minimum de trois mesures de temps d'arrivée relatives à trois différentes stations de base, il est possible de localiser le mobile sous certaines conditions de synchronisation du réseau. L'estimateur du maximum de vraisemblance conduit alors à un problème de maximisation non linéaire connu sous le nom de triangulation hyperbolique. Des algorithmes sous-optimaux sont présentés montrant d'excellents résultats lorsque le rapport signal sur bruit est suffisamment élevé. Des simulations exhaustives sont présentées dans différents environnements typiques comme les zones urbaines ou rurales en incluant des erreurs de synchronisation. Elles montrent qu'une erreur quadratique moyenne inférieure à cent mètres est possible dans la plupart des cas.

# Contents

x

# List of Figures

xiii

# List of Tables

# Chapter 1

# Introduction

Since the cellular concept was introduced in the 60's, wireless technologies have known a fast development. The main enhancement was the introduction of digital communications instead of the classical analog communications. This is mainly due to Shannon's work on information theory. Many digital standards exist nowadays. They represent the second generation standards, mainly DAMPS, GSM, and IS-95 CDMA.

The digital technology has permitted a rapid increase in the performances of cellular systems. We can distinguish now between two kinds of standards: the first is based on Time Division Multiple Access (TDMA) and the second is based on the Code Division Multiple Access (CDMA). The GSM system is the most widely used system nowadays and it is most likely that it has several years left before the third generation emerges.

A public land mobile network (PLMN) is made up of a large number of relay stations called base stations (BS). These relays have the role of covering the territory with radio resources. The territory is divided into small areas called cells. One relay may cover more than one cell depending on its configuration. It can be omnidirectional (one cell), bisectorized (two cells), trisectorized (three cells) or even more. Base stations are gathered in groups, so that all base stations that are in one group communicate with one Base Station Controller (BSC). One BSC and its base stations constitute one Base Station Subsystem (BSS). Mobile services Switching Centers (MSC) similar to the ones of a Public Switched Telephone Network (PSTN) are connected to all the BSCs. They centralize all the traffic coming from the BSCs and insure the interconnection with the PSTN. These MSCs constitute the network subsystem (NSS). MSCs are usually connected to a huge database called Home Location Register (HLR). This database contains information concerning subscribers. One or more MSCs are gateway MSCs (GMSC): they are connected to the PSTN (see Figure 1.1).
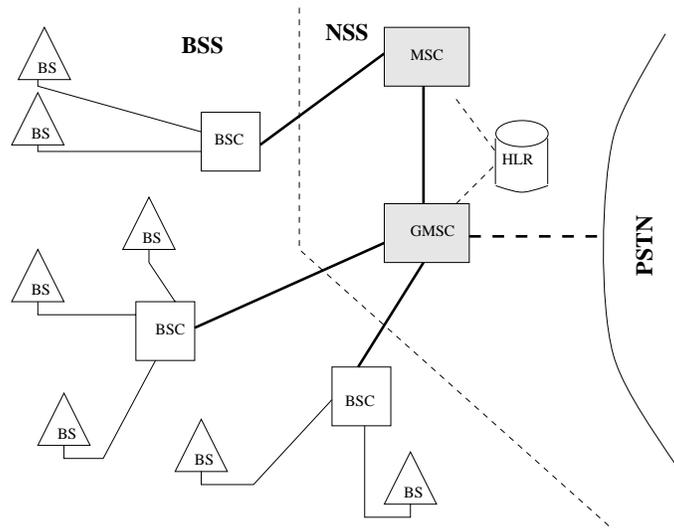
Figure 1.1: PLMN architecture

Two sets of frequencies are allocated in a Frequency Division Duplex (FDD) cellular radio system such as GSM: one for uplink transmissions (mobile to base station) and the other for downlink transmissions (base station to mobile).

There are three versions of the GSM standard according to the allocated frequency bandwidth:

- GSM 900: 890-915 MHz (uplink), 935-960 MHz (downlink).

- DCS 1800: 1710-1785 MHz (uplink), 1805-1880 MHz (downlink).

- PCS 1900: 1850-1910 MHz (uplink), 1930-1990 MHz (downlink).

Transmission in GSM is made by Time Division Multiple Access (TDMA). Several mobiles can communicate simultaneously in the same area. Each one occupies a time division allocated to it in one TDMA cycle. In addition to TDMA, there is also Frequency Division Multiple Access (FDMA). Each cell is assigned a set of frequencies.

As described in the specifications [1], the GSM system offers speech, data, and short messages transmission services. A subscriber must be able to call and to be called. In addition, the mobile must be able to communicate while moving.

## 1.1   Locating a handset

In June 1996, the American Federal Communication Committee (FCC) requested all North American mobile cellular networks to meet some requirements on the location of emergency calls by October 2001. These operators must be able to locate an emergency call made by one of their subscribers within 125 meters in 67 per cent of cases.

The localization technology has several additional applications. These applications are mainly accident reports, navigation, home zone billing, fraud detection, and statistical measurements (handover failure, traffic location).

Locating a handset is often associated to GPS, which is a well-known satellite-based technology for locating a compatible handset everywhere on the Earth in three dimensions. However, this attractive system has several drawbacks. Although it has a quite complete coverage of the Earth, it cannot work in general in some difficult environments such as urban environments or indoor environments since it requires the visibility of at least four satellites. It is possible to offer location services by using the already installed infra-structures of cellular networks whenever the mobile handset is in a covered area. The accuracy of the estimated position can be significantly better than with GPS. The aim of this dissertation is to show that it is possible to locate a mobile handset using the network's own capability.

The location computation itself is done in a entity called Mobile Location Center (MLC). This entity can be implemented in the mobile, in the network, or even somewhere else. The Mobile Location Center (MLC) needs all collected measurements performed at the handset or in the network and some other data related to the network architecture. There exist many possible techniques to locate a handset:

- Geometric approaches based on measurements of time of arrival, distance, angle of arrival, or signal level strength.

- Pattern matching approaches based on the analysis of some control parameters by matching them on prediction maps previously computed.

Synchronization is also a topic related to location. It is fully required in all methods based on time of arrival measurements.

## 1.2   Dissertation overview

This work addresses the problem of estimating a mobile handset position in a GSM radio network using time of arrival estimation.

Chapter 2 is a description of the physical layer of the GSM standard. Different logical channels are described. The different kinds of bursts used are presented too. The GMSK modulation used in GSM is depicted; it is shown to be almost linear though it is a phase modulation.

Chapter 3 discusses the different ways of locating a handset. Time of arrival, angle of arrival, signal strength, and pattern matching approaches are discussed. The time of arrival approach will be retained as being acceptable in terms of accuracy and complexity.

Chapter 4 discusses the problem of time of arrival estimation. The environment is assumed to exhibit discrete (specular) multipath components. No diffuse paths are present. The maximum likelihood estimator is used and shown to be more accurate than the classical estimator known as the matched filter (cross-correlator). Two sets of algorithms are presented in the time and the frequency domain. A link with array processing techniques is established. An extension to the case of an unknown modulation pulse is discussed.

Chapter 5 is a study of the hyperbolic trilateration procedure that is used to estimate the mobile position with at least three time of arrival measurements. It is shown that the lack of knowledge of a time reference leads to a hyperbolic trilateration made from differences of times of arrival so that the time reference contribution disappears.

Chapter 6 is a brief discussion on the synchronization issue: the problems of pseudo-synchronization and absolute synchronization are discussed separately.

Chapter 7 is the chapter of simulations. The channel model of the American normalization committee is used in order to perform simulations of both the time of arrival estimation and the hyperbolic trilateration on realistic environments.

Chapter 8 is a conclusion. It discusses also some possible future directions.

## 1.3   Contributions

This dissertation treats the important aspects of a cellular localization system:

- Basic principles: this localization system is based on time of arrival estimation followed by a hyperbolic trilateration in a synchronized or pseudo-synchronized network.

- Time of arrival estimation: an analogy with array processing is presented. A rigorous development shows the possibility of substituting the signal samples with the least squares estimate of the channel impulse response (achieved by means of the training sequence) without any loss in performances. Many algorithms are adapted to the particular problem of time of arrival estimation. Extension to the case of unknown modulation pulse is analyzed: an original ESPRIT-like algorithm is presented.

- Hyperbolic trilateration: Cramer-Rao computation with and without knowledge of the timing advance. Several algorithms are simulated and compared. The Taylor expansion algorithm shows ideal results at low signal-to-noise ratios.

- Synchronization: description of a general algorithm for the estimation of the difference of transmission times between base stations by exploiting noisy measurements performed on the radio side of the network.

In addition to these theoretical studies, several contributions have been presented to the American standardization committee T1P1.5 on the elaboration of a common channel model. Simulations have been conducted based on this model. A standard has since been adopted.

# Chapter 2

# Wireless communications

In the following, we describe the problems caused by the radio propagation environment and the techniques used for protecting communications. We describe the GSM system as a typical radio cellular network and discuss its radio modulation in details. We detail the equalization procedure or more precisely the channel impulse response estimation when a given training sequence is available.

## 2.1 Radio propagation model

The radio interface is the most difficult part in the cellular concept. This is mainly due to the channel propagation characteristics that corrupt the communications. For this purpose, a permanent signaling dialog exists between the mobile and the network whether the mobile is in communication or not.

### 2.1.1 Path loss

As the radio signal propagates, its intensity decreases. Indeed, the signal energy is distributed on a spherical front. In free space, the path loss is proportional to $d^{-2}$ where $d$ refers to the distance of propagation between the transmitter and the receiver. In typical environments such as urban or rural areas, there exist many empirical models that compute the path loss according to some parameters such as the distance or the height of the antenna. It is common to state that the path loss is proportional to $d^{-\frac{\gamma}{10}}$ where $\gamma \in [20, 40]$.

The Okumura-Hata formula [2] is perhaps the best known empirical formula used to compute the received power:

$$P_r = P_t + g_a - L_p - \gamma \log(d) \tag{2.1}$$

where $P_t$ is the transmitted power, $g_a$ is the antenna gain in the signal direction, $L_p$ is the path loss at one km. This equation is expressed in dB. The

received power $P_r$ represents a mean value, it is subject to variations around its mean value due to slow fading.

## 2.1.2   Slow and fast fading, coherence time

As the radio signal propagates, it is subject to some fluctuations. Slow fading is a fluctuation of the local mean power of the signal due to shadowing; as the mobile moves, it is subject to obstruction from many objects such as buildings or trees. This fading is usually modeled by a log-normal variable that is added to the mean received power (2.1). Its variance $\sigma_f^2$ depends on the environment.

Fast fading is caused by the reflection of the signal on an object. Indeed, the reflected signal is made up of a large number of partial waves with random phases and amplitudes. It results in quick fluctuations around the local mean power. Fast fading is widely related to the coherence time $\tau_c$ which is defined by taking the auto-correlation function of the channel impulse response $R(t)$ as:

$$\frac{R(\tau_c)}{R(0)} = 0.5 \tag{2.2}$$

This definition is valid for narrow band signal where the auto-correlation function is supposed to be constant on the signal bandwidth. The coherence time decreases when the mobile handset speed increases. At 900 Mhz, typical values of the coherence time vary from 1 ms at high speed to 100 ms for non moving handsets.

## 2.1.3   Multipath and delay spread

This is our main concern as we are interested in estimating the time of arrival. As described in the previous section, the signal is subject to some reflections. These reflections spread the signal in time. Each reflection is a copy of the transmitted signal and is subject to fading.

The amplitude $x$ of a reflected path is usually modeled by a complex Gaussian variable with zero mean. The signal envelope, $r = |x|$, follows a Rayleigh distribution:

$$p(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \tag{2.3}$$

where $\sigma^2 = \mathrm{E}\left[|x|^2\right]$. On the contrary, if the path is the direct line of sight (LOS) path, then its amplitude is not zero mean. Its envelope is modeled by a Ricean distribution given by:

$$p(r) = \frac{r}{\sigma^2} e^{-\frac{r^2+m^2}{2\sigma^2}} I_0 \left(\frac{rm}{\sigma^2}\right) \tag{2.4}$$

Where $m^2 = |\mathrm{E}[x]|^2$, $\sigma^2 = \mathrm{E}\left[|x - \mathrm{E}[x]|^2\right]$, and $I_0$ is the modified Bessel function of first kind.

The delay spread is an important characteristic of the channel impulse response. Supposing that the $i$-th path arrives at $\tau_i$ and has a power attenuation of $a_i$, and defining the $p$-th moment of the signal by:

$$m_p = \frac{\sum_i a_i \tau_i^p}{\sum_i a_i}$$

the delay spread is given by:

$$\xi_1 = \sqrt{m_2 - m_1^2} \qquad (2.5)$$

Another possible definition of the delay spread is given by:

$$\xi_2 = \max_i \tau_i - \min_i \tau_i \qquad (2.6)$$

It is important here to mention the dependency of the delay spread on the distance $d$ between the mobile and the base station the mobile is connected to. Yuanking has shown in [3] that the delay spread given by the second definition does not depend on $d$. On the contrary, Greenstein et. al. have shown in [4] that the delay spread given by the first definition depends on $d$. They proposed the following density distribution which we will refer to as the Greenstein model:

$$\xi_1 \sim T_1 d^\epsilon y \qquad (2.7)$$

where, $y$ is a log-normal variable, ($y = 10^{\frac{x}{10}}$ where $x$ is Gaussian zero mean and variance $\sigma_y^2$). $\sigma_y^2$, $T_1$ and $\epsilon$ are environment dependent constants. This model states that the delay spread rises in general when the mobile moves away from the base station.

## 2.1.4   Interferences

In any cellular system like GSM, the resource which is extremely scarce is the radio spectrum. It must be shared by all base stations and mobiles. This induces a lot of interferences that must be minimized while doing the network frequency planning.

We will make the assumption that the interferences are Gaussian, uncorrelated with the signal of interest (AWGN approximation). This assumption is true if there is a large number of interferers. It gives results that are optimistic but simplifies drastically the amount of computations needed for simulations; the interferers will not be generated individually but will be generated as one single Gaussian variable.

### 2.1.5    Doppler spread

As the mobile moves, the signal is subject to the Doppler effect that shifts it in frequency. This shift becomes more and more important as speed increases. The Doppler shift is given by $\frac{v}{\lambda}\cos(\theta)$, where $\theta$ is the angle between the signal direction of arrival and the mobile direction. At 900 MHz, the maximum shift, $f_d = \frac{v}{\lambda}$, is about one Hertz per km/h.

## 2.2    Protection techniques

- Channel coding and interleaving: channel coding is a powerful technique to secure the transmitted bits against fading. It consists in adding redundant information to the source data. The codes used in GSM are some convolutional codes and one Fire code. Depending on the transmitted information, some bits are protected more than others. Interleaving consists in shuffling the bits before putting them in bursts. The consequence is better protection against block fading.

- Discontinuous transmission (DTX) consists in transmitting at a reduced rate during a voice communication when nothing is said and silence remains. In GSM, the reduced rate is about 12 % of the normal rate. The main advantage of DTX is that it reduces the average interference and increases the handset battery life.

- Power control: like DTX, power control is a technique that limits the average interference and saves the handset battery life. When the communication quality is good enough, it is worthless to transmit at a high signal level. In this case, the mobile is requested to transmit at a lower level.

- Slow frequency hopping: Slow frequency hopping used in TDMA systems consists in changing the frequency carrier at regular intervals. Since the fading is frequency selective, the signals sent are independent even at low speed. This property is sometimes called frequency diversity. Fast frequency hopping, where the frequency changes at the modulation rate, is not used in GSM.

- Space diversity: this technique is used nowadays for uplink receivers only, but it could be used for downlink receivers in the near future. At the reception end, the base station has several antennae. The base station thus receives several copies of the same signal. The gain obtained from diversity is proportional to the number of antennae. Usually, two antennae are used providing a gain greater than 3 dB.

- Training sequences: this technique consists in sending a known data sequence for the receiver to be able to estimate the channel impulse response.

## 2.3    GSM system overview

The GSM standard is a Frequency Division Duplex (FDD) system that uses a combination of two techniques: Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA). A set of frequencies is allocated for each cell. The same frequencies are used by several geographically spaced cells according to a previously defined reuse pattern. The time domain is divided into small windows called slots. These slots are organized in cycles of 8 slots called TDMA frames. The slot duration is 156.25 bit periods which represents about 577 $\mu$s. Since the coherence time is longer than that (see Section 2.1.2), we can make the assumption that the channel impulse response is constant within a slot of one TDMA frame.

### 2.3.1    The duplex physical channel

A traffic channel occupies simultaneously two links; uplink (from the mobile to the base station) and downlink (from the base station to the mobile). These two links are separated in time and frequency:

- Three slots in time: the mobile cannot send and receive at the same time.

- The duplex shift in frequency $\Delta W$: a duplex physical channel is composed of two simple physical channels (downlink and uplink). A pair of frequencies is associated to every duplex physical channel; $f_u$ for the uplink channel and $f_d$ for the downlink channel, so that:

$$f_u = f_d - \Delta W$$

This physical channel in GSM is divided into eight subchannels, each one occupying one eighth of the time. A traffic communication uses one of these subchannels. All mobiles transmit only during the length of their respective subchannel. Frequency hopping may be used in GSM so that one given mobile can use slots on different frequencies according to a previously defined hopping sequence. Power control is used in GSM on traffic channels.

A slot hosts a sequence of modulated bits that represents the information to be sent. This sequence is called burst. There are four kinds of burst in GSM, their durations are smaller than the slot duration in all cases, the duration difference is a guard time at the beginning and at the end of the burst. Normal bursts contain 116 information bits. Eight different training sequences are defined for normal bursts.

**Normal bursts**

The normal burst is the most common burst. It is used for traffic channels. It contains 148 bits as shown in Figure 2.1.

| Tail | Information | Training sequence | Information | Tail |
|------|-------------|-------------------|-------------|------|
| 3 | 58 | 26 | 58 | 3 |

Figure 2.1: Normal burst format

**Access bursts**

The access burst is the first burst sent by a mobile when it wants to communicate with the network. It is used on the uplink channel for establishing a dialog with the base station. This burst is short compared to a normal burst but has a longer training sequence (41 bits) to enhance the receiver detection. An access burst contains 87 bits as shown in Figure 2.2.

| Tail | Training sequence | Information | Tail |
|------|-------------------|-------------|------|
| 7 | 41 | 36 | 3 |

Figure 2.2: Access burst format

**Synchronization bursts**

The synchronization burst is the first burst decoded by the mobile. It is used in the downlink direction only. This burst gives the information needed by the mobile to communicate in the current cell. It contains 148 bits and has the largest training sequence (64 bits) as shown in Figure 2.3.

| Tail | Information | Training sequence | Information | Tail |
|------|-------------|-------------------|-------------|------|
| 3 | 39 | 64 | 39 | 3 |

Figure 2.3: Synchronization burst format

**Frequency correction bursts**

This burst is used on the downlink channel for correcting the local oscillator of the mobile so that it can easily decode the synchronization burst. All its 148 bits are null, the modulated signal is just a pure sine wave at $1/4T = 1625/24 \approx 67.71$ kHz higher than the carrier central frequency.

## 2.3.2   Logical channels

The slots are organized in cycles of 26 frames for traffic channels or 51 frames for signaling channels. A superframe represents $26 * 51$ frames, it is a common cycle for both traffic and signaling channels. A hyperframe has a length of 3 h 28 min 53 s 760 ms and corresponds to 2048 superframes. An index called FN (Frame Number) characterizes every frame in the hyperframe.

The slots belong to different logical channels. Each logical channel has specific functions to accomplish. There are two groups of logical channels:

- Common channels: these channels are used in one link, up or down. The main channels here are: BCCH, SCH, and FCCH on the downlink; RACH on the uplink. No power control or frequency hopping is applied on these channels: they transmit the information at full power so that the mobile or the base station can detect them more easily. They are multiplexed in the 51-multiframe.

- Dedicated channels: these are duplex channels that occupy two simultaneous slots in both up and down links. The main channels here are: TCH and SACCH. They are multiplexed in the 26-multiframe.

Here is a quick description of these channels:

**BCCH**

For a mobile to be able to detect the nearest base station, each base station in the network broadcasts permanently information on its identity on the BCCH channel. The transmitted data are for example the cell identity, the set of used carriers, the frame number, and the carriers of the neighboring cells. The BCCH channel of one cell is always sent on the same frequency called BCCH frequency. It occupies the first slot of the TDMA frame and is not allowed to hop.

**SCH and FCCH**

These channels host respectively the synchronization and frequency correction bursts described above.

**RACH and the Timing Advance parameter (TA)**

This channel hosts the access burst. An access burst is sent whenever the mobile requests a connection.  As access to this channel is random by definition, it is subject to collision when two mobiles use the same slot. In this case, the mobile reiterates its request after a random delay. The protocol used for re-transmission is inspired from the Aloha protocol.

The propagation delay of one burst is not negligible compared to the bit period ($48/13 \approx 3.69\,\mu s$). One bit period corresponds to 1108 meters of propagation. In normal bursts, the 8.25 guard bit periods (156.25 - 148 ) may not be enough and the burst could overlap next slot. Bursts in GSM are transmitted in advance so that they arrive inside the corresponding slot. The advance in time, also called timing advance (TA), is computed using access bursts. These burst are short enough so that they cannot overlap the next slot.  The time of arrival of this burst at the base station corresponds to twice the propagation time between the base station and the mobile. The TA parameter is sent to the mobile so that it anticipates its transmission. It is coded in 6 bits and has 64 possible values that correspond to how many half bit periods of propagation separate the base station from the mobile. The precision of this value is therefore equal to one quarter of a bit period. This corresponds to 277 meters of wave propagation. Although this precision is low, it is quite enough to avoid any problem of overlapping between two consecutive bursts of different users.  This parameter can also be used to enhance the location procedure since it provides a distance information; this will be discussed later.

**TCH and SACCH**

The TCH channel is the channel that transmits the traffic data on both links at 13.6 kb/s. The SACCH channel is the signaling channel that is associated to one TCH channel.

## 2.3.3   Location Area Code (LAC)

When the mobile is requested by the network, a search procedure is operated on a set of cells called location area. At any time, the network knows in which LAC the mobile is. Indeed, the mobile sends this information to the network at regular intervals and every time it enters a new LAC. This procedure is called location update. The LAC informations of all subscribers are stored in huge databases called Home Location Register (HLR) and Visitor Location Register (VLR).

### 2.3.4   Handover

Handover is the main feature in cellular systems. Its goal is to allow the mobile
to keep its communication while moving from one cell to another. A handover
is anticipated by the network after analyzing the measurements reports sent by
the mobile. These reports contain each at most the best six BCCH signal levels
from neighboring cells. A measurements report is sent every 104 bursts (480
ms). When a handover is decided, the old base station sends to the mobile the
information that concerns the new base station (frequency carrier, slot number,
. . . ).

A handover can be intra BS, intra BSC, intra MSC, or inter MSC depending
on the configuration of the two base stations that are involved in the handover
procedure.

Network synchronization can enhance the handover procedure. In a synchro-
nized network, the mobile can be informed in advance of the new timing advance
to apply; this is not possible in a non synchronized network where the mobile
has to send an access burst to get this information. Thus there are two kinds of
handover: synchronous handover and asynchronous handover.

It is important to note that access bursts do not hop during an initial access,
but can hop in case of a handover since they are sent on a traffic channel. This fact
will be used later to justify the uncorrelation between successive access bursts.

### 2.3.5   The BSIC

In some cases, the mobile is able to listen to two base stations that use the same
BCCH frequency. The BSIC (Base Station Identity Code) is a parameter that
differentiates cells that use the same BCCH frequency.

The BSIC determines also the training sequence used in normal bursts.

### 2.3.6   The GMSK modulation

The modulation used in GSM is a constant envelope phase modulation called
Gaussian Minimum Shift Keying (GMSK). It can be described as a modified
Minimum Shift Keying (MSK) modulation. Each bit introduces a phase shift of
$\pm\frac{\pi}{2}$.

The modulated signal can be written in the following way:

$$s(t) = \sqrt{\frac{2E_b}{T}} e^{j[2\pi f_0 t + \theta(t)]} \tag{2.8}$$

where $E_b$ is the bit energy, $T$ the bit period, and $f_0$ the carrier frequency. The phase variation $\theta(t)$ can be written as:

$$\theta(t) = \theta_0 + \sum_i b_i \phi(t - iT)$$

$b_i$ are the transmitted bits after differential coding:

$$b_i = d_i d_{i-1}$$

and $\phi(t)$ the modulation function:

$$\phi(t) = \frac{\pi}{2} \int_{-\infty}^{t} s_e(\tau) d\tau$$

In the MSK, the function $s_e(t)$ is a rectangular window. The phase variation is continuous but not smooth enough. In the GMSK modulation, to reduce the spectrum occupancy, this window is filtered by a Gaussian function:

$$s_e(t) = \text{rect}\left[\frac{t}{T}\right] \star \frac{1}{\sqrt{2\pi}\sigma T} e^{-\frac{t^2}{2\sigma^2 T^2}} \tag{2.9}$$

with $\sigma = \frac{\sqrt{\ln 2}}{2\pi BT}$ and $BT = 0.3$.

$\phi(t)$ can be written more simply with one integral by noticing that [5]:

$$\frac{d\phi(x)}{dx} = \frac{\pi}{2} \int_{x-T/2}^{x+T/2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} dt$$

We obtain therefore:

$$\begin{cases} \phi(x) = \frac{\pi}{2} \left[ \psi(x + T/2) - \psi(x - T/2) \right] \\ \\ \psi(x) = \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} + x \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} dt \end{cases} \tag{2.10}$$

The drawback of the GMSK modulation compared to the MSK is that it introduces more inter symbol interference (ISI): the equalization is then a little more complex.

All phase modulations with differential coding can be approximated to amplitude modulated pulses (AMP) [6]. This approximation is very good in the GMSK modulation case. For more details see Appendix A.

The transmitted burst can be written in the following way:

$$s(t) \approx e^{j\phi_0} \sum_k j^k d_k f(t - kT) \tag{2.11}$$

with $\phi_0$ is an initial phase and $f(t)$ the main pulse in the decomposition. It is a decomposition in amplitude modulated pulse (AMP) by the transmitted bits with no differential coding.
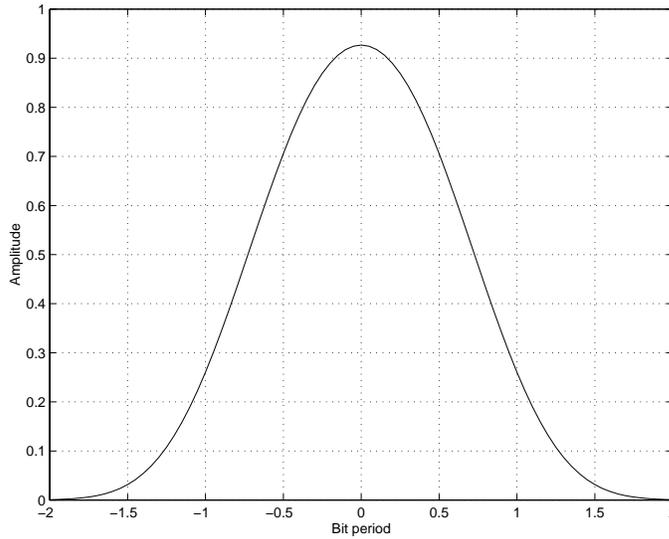


Figure 2.4: GMSK main pulse

By doing a demodulation at $\frac{1}{4T}$, we can get rid of the term $j^k$:

$$s(t)e^{-j\frac{\pi t}{2T}} \approx e^{j\phi_0} \sum_k d_k a(t - kT) \tag{2.12}$$

with

$$a(t) = f(t)e^{-j\frac{\pi t}{2T}}$$

From now on, we will assume a demodulation at $\frac{1}{4T}$, and a null initial phase so that the received signal will be written in the following way:

$$s(t) = \sum_k d_k a(t - kT) \tag{2.13}$$

## 2.4   Channel equalization

The goal of equalization is to reduce the inter symbol interference (ISI) in order to estimate the transmitted bit. Indeed, since the GMSK main pulse duration is four times the bit period, inter symbol interference is unavoidable.

At the reception, the received signal is the convolution of the transmitted signals with a cascade of filters which are:

- the modulation pulse shape $a(t)$,

- the transmission filter $F_e(t)$ that fits the signal to the desired bandwidth,

- the filter $C(t)$ that corresponds to the propagation over the air,

- The reception filter $F_r(t)$ that translates the signal to its base-band version.

The channel impulse response is defined as the cascade of these filters. The transmission scheme can then be written in the following way:

$$
\begin{cases}
y(t) = s(t) \star h(t) + b(t) \\
s(t) = \sum_k d_k \delta(t - kT) \\
h(t) = a(t) \star F_e(t) \star C(t) \star F_r(t)
\end{cases}
\tag{2.14}
$$

where $y(t)$ is the received signal, $s(t)$ is the transmitted signal, $\delta(t)$ the Dirac function, and $b(t)$ an additive complex noise that is usually assumed to be zero mean and uncorrelated with $s(t)$:

$$
\mathrm{E}[s(t)b(\tau)] = 0
$$

$h(t)$ is the channel impulse response. For convenience, $h(t)$ is assumed to have a finite duration of $p$ bit periods. $p$ is the memory of the channel. One consequence of Shannon's work is the sampling theorem which states that it is possible to reconstruct any finite bandwidth signal from its samples at a rate that is greater or equal to twice its highest frequency.

In the GSM case, the bandwidth of each carrier is 200 kHz, so that the complex envelope is limited to 100 kHz. The bit rate, that is approximately equal to 271 kHz is then enough to satisfy the Shannon's criterion. From now on, all signals are sampled at the bit rate. The sampled signal is:

$$
y(jT) = \sum_{k=0}^{p-1} h(kT) d_{j-k} + b(jT) \qquad p \le j \le N
\tag{2.15}
$$

The last equation can then be written in a more compact formula using matrices:

$$\mathbf{Y} = \mathbf{D(d)H} + \mathbf{B} \tag{2.16}$$

where $\mathbf{d} = [d_1 \cdots d_q]^T$ refers to the bits of the training sequence,

$$\mathbf{H} = [h(0) \cdots h((p-1)T)]^T$$

$$\mathbf{Y} = [y(pT) \cdots y(qT)]^T$$

$$\mathbf{B} = [b(pT) \cdots b(qT)]^T \qquad p < q$$

and:

$$\mathbf{D(d)} = \begin{pmatrix} d_p & d_{p-2} & \cdots & d_1 \\ d_{p+1} & d_p & \cdots & d_2 \\ \vdots & \vdots & & \vdots \\ d_q & d_{q-1} & \cdots & d_m \end{pmatrix} \tag{2.17}$$

The noise samples are assumed to be uncorrelated between each other, i.e. $\mathrm{E}[\mathbf{BB}^*] = \sigma^2 \mathbf{I}$.

One way of estimating the bits is the maximum likelihood (ML) detector. This detector is obtained by maximizing the likelihood of the observation with respect to the parameter to be estimated or equivalently by maximizing the log-likelihood of the observation:

$$\hat{\theta} = \arg\max \ln p(y|\theta)$$

In our case, $\theta$ is a vector that contains the channel impulse response coefficients and the unknown bits. In the context of Gaussian noise, the log-likelihood can be written in the following way:

$$\ln p(\mathbf{Y}|\mathbf{H}, \mathbf{d}) = -\frac{1}{\sigma^2} \|\mathbf{Y} - \mathbf{D(d)H}\|^2 \tag{2.18}$$

Several approaches are possible to solve the equation. The channel impulse response is unknown. It is possible to estimate it directly without any prior knowledge of the bits: this is called blind equalization. We will focus on the GSM case where a training sequence consisting of known bits is available in all bursts.

We will solve the system (2.18) in two steps: in the first step, we assume that $\mathbf{d}$ refers to the training sequence and estimate the channel impulse response accordingly. In the second step, $\mathbf{d}$ refers to the unknown bits and the channel impulse response estimated from the first step is used.

It should be noted that this scenario is not optimal since the estimations of the unknown bits and the channel impulse response are not performed jointly. However, the low gain in performances given by the joint estimation does not justify its huge complexity increase.

We suppose that the channel impulse response is constant within one burst. This approximation is usually verified at normal mobile speed. The estimation of the channel impulse response is straightforward using the training sequence:

$$\hat{\mathbf{H}} = \mathbf{D}^{\dagger}(\mathbf{d})\mathbf{Y} \tag{2.19}$$

where $\mathbf{D}^{\dagger} = (\mathbf{D}^{*}\mathbf{D})^{-1}\mathbf{D}^{*}$ refers to the Moore-Penrose pseudo-inverse of $\mathbf{D}$. Note that the matrix $\mathbf{D}^{\dagger}(\mathbf{d})$ can be precomputed and stored as it only depends on the training sequence.

The error covariance of this estimator is:

$$\mathrm{E}\left[\left(\hat{\mathbf{H}} - \mathbf{H}\right)\left(\hat{\mathbf{H}} - \mathbf{H}\right)^{*}\right] = \sigma^{2}\left[\mathbf{D}^{*}(\mathbf{d})\mathbf{D}(\mathbf{d})\right]^{-1}$$

The training sequences are chosen for their good auto and inter correlation properties, i.e. they are almost uncorrelated with theirs own shifted versions:

$$\mathbf{D}^{*}(\mathbf{d})\mathbf{D}(\mathbf{d}) \approx m\mathbf{I}$$

This shows that the variance error is reduced by a factor $m = q - p + 1$.

$$\sigma_{r}^{2} = \frac{\sigma^{2}}{m}$$

In the case we have some a priori information on the parameter to be estimated, the Maximum A Posteriori (MAP) estimator can be used successfully to minimize the covariance error $\mathrm{E}\left[\left(\hat{\mathbf{H}} - \mathbf{H}\right)\left(\hat{\mathbf{H}} - \mathbf{H}\right)^{*}\right]$. Supposing that the channel impulse response is Gaussian zero mean with covariance matrix $\mathbf{K}_{h} = \mathrm{E}[\mathbf{H}\mathbf{H}^{*}]$, the MAP estimator is (see [7] for example):

$$\begin{aligned}
\hat{\mathbf{H}} &= \left[\mathbf{D}^{*}(\mathbf{d})\mathbf{D}(\mathbf{d}) + \sigma^{2}\mathbf{K}_{h}^{-1}\right]^{-1}\mathbf{D}^{*}(\mathbf{d})\mathbf{Y} \\
&= \mathbf{K}_{h}\mathbf{D}^{*}(\mathbf{d})\left[\mathbf{D}(\mathbf{d})\mathbf{K}_{h}\mathbf{D}^{*}(\mathbf{d}) + \sigma^{2}\mathbf{I}\right]^{-1}\mathbf{Y} \tag{2.20}
\end{aligned}$$

The second equation is obtained using the inverse matrix lemma. It avoids the inversion of the covariance matrix $\mathbf{K}_{h}$. Note that equation (2.19) is a special case of this equation when no information a priori is available, i.e. $\mathbf{K}_{h} = \infty$.

If the noise is not white, its covariance $\mathbf{K}_{b} = \mathrm{E}[\mathbf{B}\mathbf{B}^{*}]$ can be incorporated:

$$\begin{aligned}
\hat{\mathbf{H}} &= \left[\mathbf{D}^{*}(\mathbf{d})\mathbf{K}_{b}^{-1}\mathbf{D}(\mathbf{d}) + \mathbf{K}_{h}^{-1}\right]^{-1}\mathbf{D}^{*}(\mathbf{d})\mathbf{K}_{b}^{-1}\mathbf{Y} \\
&= \mathbf{K}_{h}\mathbf{D}^{*}(\mathbf{d})\left[\mathbf{D}(\mathbf{d})\mathbf{K}_{h}\mathbf{D}^{*}(\mathbf{d}) + \mathbf{K}_{b}\right]^{-1}\mathbf{Y} \tag{2.21}
\end{aligned}$$

Once the channel impulse response is estimated, we estimate the unknown bits by minimizing:

$$\ln p(\mathbf{Y}|\mathbf{d}) \approx \|\mathbf{Y} - \mathbf{D}(\mathbf{d})\hat{\mathbf{H}}\|^2$$

where $\mathbf{d}$ represents in this case the unknown bits. The well known Viterbi algorithm [8] is usually used here to recursively estimate the unknown bits. This algorithm can also provide soft confidence values for each bit enabling soft decoding.

# Chapter 3

# Technologies for location

This chapter presents a description of possible technologies that might be used for locating a mobile handset in a radio cellular network. It explains why the best technology in the actual context is based on time of arrival.

In general, the idea is to gather as many measurements as possible and to exploit a large number of relevant observations. Each observation contributes to add some more information and enhances therefore the estimation accuracy. The set of relevant measurements can be obtained at the handset from signals coming from different base stations or can be measured from the signal coming from the mobile and impinging on several base stations. Accumulating both measurements (i.e. on downlink and uplink channels) enhances the final estimation accuracy.

## 3.1   Distance estimation from signal strength

This is the most intuitive technique to locate a handset. The mobile measures the signal strength from several base stations. A good candidate is the BCCH channel since it is transmitted at full power. The mobile can compute the distance that separates it from the base station by means of an appropriate path loss empirical model. In general, the path loss is proportional to the distance between the mobile and the base station :

$$\frac{P_r}{P_t} = cd^{-\gamma} \tag{3.1}$$

Two distances are required for locating a handset by a simple circular trilateration[1]. This methodology has been analyzed in [9]. The main advantage of this method is that it can easily be implemented in GSM[2]. Indeed, the mobile sends a mea-

---

[1]The ambiguity can be solved by selecting the intersection that is inside the area of interest.
[2]Some operators have already implemented this technology for some commercial applications such as home zone billing.

surements report every 0.48 s that includes the downlink signal levels of the best six base stations. The signal power level is expressed on a discrete scale running from 0 to 63.

The main drawback of this method is its low precision due mainly to shadowing. However it may be assisted by a prediction map in order to correct the estimation resulting from the trilateration alone. Such prediction maps are available in the database of an operator. They are generally obtained by simulation.

Another way of assisting the method is to provide collected field measurements as training data; this is subject of next section.

## 3.2   Pattern matching based on training data

This technique can be compared to pattern recognition in the sense that the mobile position is assigned to a small area (smaller than a cell). This assignment decision is made by matching some observations to some prediction maps. The area of interest is divided into a large number of small areas of location. This methodology requires a large amount of collected data, which are used to build a decision model based on some information criterion. This technique is in contrast with the others in the sense that the location computation is not performed by trilateration; it is purely statistical. The handset is assigned an area where its probability of presence is high enough.

Many methods exist to build the decision model based on which the location area is decided. Decision trees [10] represent one powerful tool for this purpose.

## 3.3   Angle of Arrival (AoA) estimation

It is perhaps one of the most famous techniques for source localization. This topic received a lot of interest in the last two decades [11]. The research led to some well-known algorithms. The problem of interest is to locate multiple coherent or non coherent sources that impinge on an array consisting of multiple sensors.

Estimating the angle of arrival of a signal requires antenna array. This equipment is expensive and current GSM operators are generally not willing to use such installations. For this reason, this technique will be omitted for further considerations.

# 3.4   Time of Arrival (ToA) estimation

The estimation of the time of arrival (ToA) parameter is an old technique used in various applications such as radar and sonar data processing, geological acoustic sounding, and medical imaging processing. In the GSM context, the ToA estimation can be achieved by means of the training sequence. Two scenarios are possible, uplink or downlink:

- The mobile estimates the ToA of the training sequence from bursts coming from different base stations. Synchronization bursts are good candidates since they have the longest training sequence (64 bits) and are sent at full power. Moreover, consecutive synchronization bursts are spaced by 10 TDMA frames offering better uncorrelation between bursts. The time needed to observe 20 synchronization is 0.96 s. This method requires modifications to actual mobile handsets so that they are able to estimate the ToA parameter with a better accuracy.

- Several base stations measure the same signal coming from the mobile handset. Since traffic bursts are subject to power control, access bursts are the best candidates in this approach. They are always transmitted at full power and have a long training sequence (41 bits). However, access bursts are sent only at the beginning of a communication or whenever a handover occurs[3]. A scenario suggested by Ericsson is to force the mobile to perform an intra cell handover so that the mobile send an access burst to retrieve the new link information such as power control and timing advance. The base station does not respond to this handover and the mobile reiterates its demand. Overall, the mobile sends a high number of access bursts (exactly 70 bursts within 0.32 s) that are measured at several base stations. This method does not require any modification to the mobile handset[4]. It makes use of diversity to increase the number of samples according to the diversity order. The main drawback is its huge complexity on the network side and that it cannot work in idle mode. Note that in this special case, access bursts are allowed to hop since they are sent on a traffic channel. They are then assumed to be uncorrelated between each other.

In both scenarios, the main source of error is multipath. The goal of time of arrival estimation is to detect a path that is as close as possible to the direct line of sight path. It is however impossible to avoid bias in a time of arrival estimation when the direct line of sight path is not present.

At least three times of arrival are required if the location is done in two dimensions. The unknown variables are $x$, $y$, and $t_{\mathrm{r}}$ which is an unknown reference

---

[3]In the case of asynchronous handover only.

[4]This is not really true since some handsets does not support intra cell handover.

time that will be discussed in Chapter 5. The trilateration is called hyperbolic because it is based on Difference Times of Arrival (DToA) which corresponds to the equation of a hyperbola. The DToA technology has the advantage of eliminating constant biases that may be present in the ToA estimation.

Base station synchronization is always required for the location to be possible. The differences in transmission times between al base stations (also called RTD for Real Time Difference) involved in the procedure are required. This will be discussed in Chapter 6.

As the main source of problems in the ToA estimation is multipath, the performances are slightly sensitive to the signal-to-noise ratio. A ToA can be measured even at low signal-to-noise ratios once the training sequence is located. In the downlink scenario, the detection probability can be enhanced by communicating the handset prior information on the synchronization state of the neighboring base stations. This can give some compensation to the limitation of the handset compared to the base station sensibility and the advantage that it avoids for the handset the BSIC decoding (see Section 2.3.5) that is required to identify the base station the handset is measuring.

## 3.5   Hybrid methods - Joint Angle and Delay Estimation (JADE)

The method is a combination of both ToA and AoA estimation. The estimation is performed jointly. The joint estimation enhances significantly the delays estimation. Indeed, two closely spaced (in time) paths may have two different angles and therefore be easily resolved [12–14].

## 3.6   Conclusion

The discussed approaches are shown briefly in Table 3.1. The discussion on accuracy might be subjective since the approaches have not been tested at all until now (except the signal level approach). Our preference goes to the ToA approaches (downlink or uplink) since it seems to give an acceptable accuracy with some few modifications to the mobile handsets and the network.

From now on, we will not distinguish between the downlink or the uplink time of arrival technologies since these two approaches are strictly identical from a calculation point of view. Three base stations at least must be involved in the location procedure in order to obtain three independent times of arrival. Assuming that the differences of transmission times of the base stations are known, the

| Method | Link | Mode | Synchronization | Accuracy |
|---|---|---|---|---|
| Signal level | Down | Idle | Not required | Poor |
| Uplink ToA | Up | Dedicated | Required | Good |
| Downlink ToA | Down | Idle | Required | Good |
| AoA | Up | Dedicated | Not required | Good |
| JADE | Up | Dedicated | Required | Excellent |
| Pattern Matching | Up or/and down | - | Not required | Excellent |

Table 3.1: Summary of the different technologies

system becomes identifiable and the handset can be located.

# Chapter 4

# Time of arrival estimation

In this chapter, array processing techniques are applied to the general problem of temporal analysis of a received narrow band signal consisting of replicas of a known shape. This signal is received on one single sensor. We are in particular interested in the estimation of the time of arrival (ToA), which is by definition the time delay of the first path. For this purpose, we apply high resolution methods for estimating all delays and then decide to retain the first one as the time of arrival. We will apply the results to GSM.

The goal is to determine the time delay that is as close as possible to the line of sight (LOS) path. The main obstacle to time of arrival estimation is multipath. When the paths are closely spaced, it is hard to resolve them with a wide pulse shape such as the GMSK pulse. Moreover the line of sight path may not exist and it is impossible in this case to avoid a bias in the estimation. In [15], it has been noticed that allowing biases in the estimator can reduce consequently the variance error.

## 4.1   Notations and assumptions

In wireless communications, the channel impulse response varies in time due to mobility. In a Time Division Multiple Access (TDMA) system with short bursts such as GSM, we can assume that the channel impulse response is constant within one burst for normal mobile speed since the coherence time is longer than the burst duration (in GSM 577 $\mu$s). In addition, the bursts are uncorrelated from one burst to another. This is due to frequency hopping[1] and, for high mobile speed, to the coherence time that is shorter than the length of one TDMA frame (4.615 ms in GSM).

---

[1]On traffic channels only.

The channel impulse response is estimated by means of the known training sequence (TSC). In GSM, such training sequences are composed of 26 bits in normal bursts, 64 bits for synchronization bursts, and 41 bits for access bursts (see Section 2.3.1). They are located in the middle of the burst.

We use the following constants:

- $q$ refers to the length of the training sequence,

- $p$ refers to the channel impulse response length,

- $L$ refers to the number of successive observations[2].

Supposing $d$ discrete specular paths exist, the channel filter that corresponds to the propagation of the $j$-th burst is:

$$C_j(t) = \sum_{i=1}^{d} s_{ij}\delta(t - \tau_i) \qquad 1 \leq j \leq L \qquad (4.1)$$

$\{\tau_i\}_{1 \leq i \leq d}$ are the delays to be estimated, and $\{s_{ij}\}_{1 \leq i \leq d}$ the gains (fading) of all paths. The channel impulse response corresponding to one given burst $j$ can be written by:

$$h_j(t) = \sum_{i=1}^{d} s_{ij}a(t - \tau_i) \qquad 1 \leq j \leq L \qquad (4.2)$$

with $a(t)$ the known modulation pulse function. Our goal is to estimate the times of arrival of the different delays, or at least the dominant ones and then define the first one as the time of arrival (ToA). For this purpose, we receive $L$ consecutive bursts. The part of the $j$-th burst that corresponds to the TSC is:

$$y_j(t) = \sum_{i=1}^{q} d_i h_j(t - iT) + n_j(t) \qquad 1 \leq j \leq L \qquad (4.3)$$

where $\{d_i\}_{1 \leq i \leq q}$ is the training sequence, $T$ the bit period, and $n_j$ a white complex Gaussian noise with variance $\sigma^2$. The pulse $a(t)$ is supposed to be continuous and has a finite duration $S_a$. We suppose that the delays are constant during the observations of the $L$ bursts. This is true at usual mobile speeds, i.e. less than 150 km/h. The complex gains are not constant as they vary from one burst to another; they are however constant within one burst.

---

[2]In case of diversity, this number increases according to the diversity order (number of antennae) if we assumed that the delays are identical on the antennae and that the antennae are spaced enough from each other so that they are uncorrelated.

All signals are sampled at the bit rate which satisfies the Nyquist/Shannon criterion[3].

In general, the number of delays is unknown but it can be estimated by several techniques as shown in [16–22]. These techniques are mainly the AIC and MDL criteria.

We will use the following notations:

$\mathbf{a}(\tau_i) = [a(-\tau_i) \cdots a((p-1)T - \tau_i)]^T$ sampled pulse shifted by $\tau_i$,
$\tau = (\tau_1 \cdots \tau_d)$ delays to be estimated,
$\mathbf{A}(\tau) = [\mathbf{a}(\tau_1) \cdots \mathbf{a}(\tau_d)]$ rectangular $p \times d$ matrix,
$\mathbf{Y}_j = [y_j(pT) \cdots y_j(qT)]^T$ observations in one burst,
$\mathbf{H}_j = [h_j(0) \cdots h_j((p-1)T)]^T$ channel impulse response,
$\mathbf{S}_j = [s_{1j} \cdots s_{dj}]^T$ gains in burst $j$,
$\mathbf{N}_j = [n_j(pT) \cdots n_j(qT)]^T$ noise vector,
$\mathbf{K}_y = \mathrm{E}\left[\mathbf{Y}_j \mathbf{Y}_j^*\right]$ signal covariance matrix,
$\mathbf{K}_s = \mathrm{E}\left[\mathbf{S}_j \mathbf{S}_j^*\right]$ gain covariance matrix,
$\mathbf{K}_h = \mathrm{E}\left[\mathbf{H}_j \mathbf{H}_j^*\right]$ channel covariance matrix.

We note $m = q - p + 1$ the number of signal samples of $y_j(t)$. The pulse is normalized, i.e. $\mathbf{a}^*(\tau)\mathbf{a}(\tau) = 1$. The gains are assumed to be Gaussian with zero mean and uncorrelated in time (from one burst to another), but might be correlated in space (between the gains of various delays in the same observation). That is to say:

$$\mathrm{E}\left[\mathbf{S}_j \mathbf{S}_k^*\right] = \delta_{jk}\mathbf{K}_s \tag{4.4}$$

Consequently, the envelope of each gain follows a Rayleigh distribution[4]. The uncorrelation in time is well justified if slow frequency hopping is used. The noise is Gaussian with zero mean, uncorrelated in time and space:

$$\mathrm{E}\left[\mathbf{N}_j \mathbf{N}_k^*\right] = \delta_{jk}\sigma^2 \mathbf{I} \tag{4.5}$$

Let $\mathbf{D}(\mathbf{d})$ be the matrix defined in (2.17) where $\mathbf{d}$ refers to the set of bits belonging to the training sequence. The observations can be rewritten as:

$$\begin{cases} \mathbf{Y}_j = \mathbf{D}(\mathbf{d})\mathbf{H}_j + \mathbf{N}_j \\ \mathbf{H}_j = \mathbf{A}(\tau)\mathbf{S}_j \end{cases} \qquad 1 \leq j \leq L \tag{4.6}$$

---

[3]In GSM the bit rate is 270 kHz which is greater than $2 \times 100 = 200$ kHz.

[4]If the line of sight path follows a Ricean distribution then the time of arrival estimation accuracy is greatly enhanced. We are therefore considering the worst case.

We can rewrite the observations in a more compact formula:

$$\mathbf{Y}_j = \mathbf{D}(\mathbf{d})\mathbf{A}(\tau)\mathbf{S}_j + \mathbf{N}_j \qquad 1 \leq j \leq L \tag{4.7}$$

The problem of estimating the delays is then equivalent to the joint estimation of the angles of arrival of $d$ unknown sources impinging on an array of $m$ sensors [11, 23, 24]:

$$\check{\mathbf{Y}}_j = \check{\mathbf{A}}(\theta)\check{\mathbf{S}}_j + \check{\mathbf{N}}_j \qquad 1 \leq j \leq L \tag{4.8}$$

where $\check{\mathbf{A}}(\theta) = [\check{\mathbf{a}}(\theta_1) \cdots \check{\mathbf{a}}(\theta_d)]$, $\theta$ denotes the set of the $d$ angles of arrival to be estimated. The $m$ dimension vector $\check{\mathbf{a}}(\theta)$ is the array spatial signature for a source coming from direction $\theta$. It is also called steering vector. In our case the steering vector is $\mathbf{D}(\mathbf{d})\mathbf{a}(\tau)$.

The differences between (4.7) and (4.8) are:

- $m$ is the number of samples instead of the number of sensors,

- the observations $\mathbf{Y}_j$ are the consecutive samples on one single antenna instead of the array output $\check{\mathbf{Y}}_j$,

- the matrix $\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)$ depends on the time delays while $\check{\mathbf{A}}(\theta)$ depends on the angles of arrival. Note that the rank of $\check{\mathbf{A}}(\theta)$ is $d$ and the rank of $\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)$ is equal to $\min(d, p, q - p)$. However, this minimum is in general equal to the number of delays $d$.

- the random gains $\check{\mathbf{S}}_j$ play the role of the sources instead of the gains $\mathbf{S}_j$.

The following definitions are useful to evaluate an estimator:

## Consistency

An estimator is consistent when it is asymptotically unbiased, i.e. the estimator tends to the true value when the number of observations tends to infinity.

## Efficiency

An estimator is efficient if it reaches the Cramer-Rao bound that provides a lower bound on the variance error of any unbiased estimator. An estimator may not be efficient but asymptotically efficient, i.e. when the number of observations tends to infinity, the variance error tends to the asymptotic Cramer-Rao bound.

## 4.2   Temporal approach

This approach deals directly with the samples in the time domain while the second approach developed in section 4.3 deals with the Fourier transform of the observations. All results obtained in this section are valid in the frequency domain too. As we will see, the frequency approach offers new perspectives.

We analyze the general case where the gain covariance matrix $\mathbf{K}_s$ is assumed to be definite positive. It is possible to suppose that this matrix is diagonal, i.e. uncorrelated gains, as done in [25].

It is common to distinguish two maximum likelihood estimators [26]:

### 4.2.1   The Deterministic Maximum Likelihood (DML)

In this approach, the unknown gains $\mathbf{S}_j$ are estimated in addition to the $d$ delays. We must then perform a joint estimation of the gains in each observation and the common delays. This approach is interesting if, for some reason, we are not sure that the gains are Gaussian.

The likelihood to be maximized is:

$$\prod_{j=1}^{L} p(\mathbf{Y}_j | \tau, \mathbf{S}_j, \sigma^2) = \frac{1}{\pi^L \sigma^{2Lm}} e^{-\frac{1}{\sigma^2} \sum_{j=1}^{L} \|\mathbf{Y}_j - \mathbf{D}(\mathbf{d})\mathbf{A}(\tau)\mathbf{S}_j\|^2} \tag{4.9}$$

We then have to solve the following problem:

$$(\hat{\tau}, \hat{\mathbf{S}}_1, \cdots, \hat{\mathbf{S}}_L, \hat{\sigma}^2) = \arg\min_{\tau} \frac{1}{\sigma^2} \sum_{j=1}^{L} \|\mathbf{Y}_j - \mathbf{D}(\mathbf{d})\mathbf{A}(\tau)\mathbf{S}_j\|^2 + Lm \ln \sigma^2 \tag{4.10}$$

A derivation with respect to $\sigma^2$ gives:

$$\hat{\sigma}^2(\tau) = \frac{1}{Lm} \sum_{j=1}^{L} \|\mathbf{Y}_j - \mathbf{D}(\mathbf{d})\mathbf{A}(\tau)\mathbf{S}_j\|^2$$

Hence, the estimator is given by:

$$(\hat{\tau}, \hat{\mathbf{S}}_1, \cdots, \hat{\mathbf{S}}_L) = \arg\min_{\tau} \sum_{j=1}^{L} \|\mathbf{Y}_j - \mathbf{D}(\mathbf{d})\mathbf{A}(\tau)\mathbf{S}_j\|^2 \tag{4.11}$$

We then have the well-known solution:

$$\hat{\mathbf{S}}_j(\tau) = [\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)]^{\dagger} \mathbf{Y}_j \tag{4.12}$$

$$\hat{\sigma}^2(\tau) = \frac{\text{Tr}(\mathbf{\Pi}^{\perp}_{\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)}\hat{\mathbf{K}}_y)}{m} \tag{4.13}$$

$\mathbf{\Pi}^{\perp}_{\mathbf{A}}$ refers to the projection matrix onto the null space of $\mathbf{A}$. $\hat{\mathbf{K}}_y$ is the estimated covariance matrix by averaging the data:

$$\hat{\mathbf{K}}_y = \frac{1}{L} \sum_{j=1}^{L} \mathbf{Y}_j \mathbf{Y}_j^* \tag{4.14}$$

The next step is to maximize with respect to $\tau$:

$$
\begin{aligned}
\sum_{j=1}^{L} \|\mathbf{Y}_j - \mathbf{D}(\mathbf{d})\mathbf{A}(\tau)\left[\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)\right]^{\dagger}\mathbf{Y}_j\|^2 &= \sum_{j=1}^{L} \|\mathbf{\Pi}^{\perp}_{\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)}\mathbf{Y}_j\|^2 \\
&= \sum_{j=1}^{L} \text{Tr}\left(\mathbf{Y}_j^*\mathbf{\Pi}^{\perp}_{\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)}\mathbf{Y}_j\right) \\
&= L\,\text{Tr}\left[\mathbf{\Pi}^{\perp}_{\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)}\hat{\mathbf{K}}_y\right]
\end{aligned}
$$

We have used the fact that $\mathbf{\Pi}^{\perp}_{\mathbf{A}}\mathbf{\Pi}^{\perp}_{\mathbf{A}} = \mathbf{\Pi}^{\perp}_{\mathbf{A}}$ and that $\text{Tr}(\mathbf{AB}) = \text{Tr}(\mathbf{BA})$.

$\hat{\mathbf{K}}_y$ is a sufficient statistic for the estimation of the delays. The maximum likelihood estimate of $\tau$ is obtained by minimizing the following concentrated function:

$$V_{\text{DML}}(\tau) = \text{Tr}\left[\mathbf{\Pi}^{\perp}_{\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)}\hat{\mathbf{K}}_y\right] \tag{4.15}$$

This estimator is known to be inefficient [27], in the sense that it does not reach the Cramer-Rao bound. This is not surprising since the number of parameters to be estimated increases with the number of realizations. For $L$ observations, we have to estimate $L(d+1)$ parameters. It is straightforward to see, for example, that $\hat{\sigma}^2(\tau)$ in (4.13) is biased since:

$$\text{E}[\hat{\sigma}^2(\tau)] = \frac{m-d}{m}\sigma^2$$

## 4.2.2 The Stochastic Maximum Likelihood (SML)

In this approach, we estimate $\tau$ without estimating the gains. $\mathbf{Y}_j$ is Gaussian as it is a linear sum of Gaussian variables. The quantity to be maximized is:

$$\prod_{j=1}^{L} p(\mathbf{Y}_j|\tau, \mathbf{K}_s, \sigma^2) = \frac{\exp\left[-L\text{Tr}(\hat{\mathbf{K}}_y\mathbf{K}_y^{-1})\right]}{\pi^{Lm}\,|\mathbf{K}_y|^L} \tag{4.16}$$

with:

$$\mathbf{K}_y = \mathbf{A}(\tau)\mathbf{K}_s\mathbf{A}^*(\tau) + \sigma^2\mathbf{I} \tag{4.17}$$

In this case, $\hat{\mathbf{K}}_y$ is a sufficient statistic for the estimation of both $\tau$ and $\mathbf{K}_s$.

The stochastic maximum likelihood estimator is given by:

$$(\hat{\tau}, \hat{\mathbf{K}}_s, \hat{\sigma}^2) = \arg \max_{\tau, \mathbf{K}_s, \sigma^2} \mathrm{Tr}\left[\hat{\mathbf{K}}_y \mathbf{K}_y^{-1}\right] + \ln|\mathbf{K}_y| \tag{4.18}$$

$\mathbf{K}_s$ and $\sigma^2$ can be eliminated and $\tau$ is obtained by minimizing the following cost function [28]:

$$V_{\mathrm{SML}}(\tau) = \ln\left|\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)\hat{\mathbf{K}}_s(\tau)\left[\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)\right]^* + \hat{\sigma}^2(\tau)\mathbf{I}\right| \tag{4.19}$$

with,

$$\hat{\mathbf{K}}_s(\tau) = \left[\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)\right]^\dagger \left(\hat{\mathbf{K}}_y - \hat{\sigma}^2(\tau)\mathbf{I}\right)\left[\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)\right]^{\dagger*} \tag{4.20}$$

$$\hat{\sigma}^2(\tau) = \frac{\mathrm{Tr}\left(\mathbf{\Pi}^\perp_{\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)}\hat{\mathbf{K}}_y\right)}{m - d} \tag{4.21}$$

Unlike DML, SML is known to be consistent and asymptotically efficient [26]. It is usually called unconstrainted maximum likelihood because the estimation of $\mathbf{K}_s$ has been performed without the constraint that it must be positive definite. In some cases $\hat{\mathbf{K}}_s(\tau)$ given by (4.20) may be indefinite. Taking into account such constraint is complex [29]. Moreover, the estimator that takes this constraint into account gives asymptotically (i.e. $L \to \infty$) the same performances. Again, this is due to the efficiency of SML.

## 4.2.3   Channel impulse response estimation

According to (4.6), the least squares estimate of $\mathbf{H}_j$ is straightforward and given by:

$$\hat{\mathbf{H}}_j = \mathbf{D}^\dagger(\mathbf{d})\mathbf{Y}_j \tag{4.22}$$

The training sequences were designed to have very good autocorrelation properties, i.e:

$$\mathbf{D}^*(\mathbf{d})\mathbf{D}(\mathbf{d}) \approx m\mathbf{I} \tag{4.23}$$

If we assume (4.23) to hold with equality[5], the least squares estimation of the channel impulse response has a simple formula:

$$\hat{\mathbf{H}}_j = \frac{\mathbf{D}^*(\mathbf{d})\mathbf{Y}_j}{m} \tag{4.24}$$

We then have a noisy version of the $j$-th channel impulse response $h_j(t)$:

$$\hat{h}_j(t) = \sum_{i=1}^{d} s_{ij}a(t - \tau_i) + w_j(t) \qquad 1 \le j \le L \tag{4.25}$$

---

[5]In fact, this is impossible for a finite sequence length, however we will accept this approximation for the training sequences of access and synchronization bursts.

We note $\sigma_r^2 = \frac{\sigma^2}{m}$ the reduced variance of the noise $w$.  Obviously, $\hat{\mathbf{H}}_j$ is Gaussian complex with zero mean and covariance given by:

$$\mathbf{K}_{\hat{h}} = \mathbf{K}_h + \sigma_r^2\mathbf{I} = \mathbf{A}(\tau)\mathbf{K}_s\mathbf{A}^*(\tau) + \sigma_r^2\mathbf{I} \tag{4.26}$$

If we note $\hat{\mathbf{K}}_{\hat{h}} = \frac{1}{L}\sum_{j=1}^{L}\hat{\mathbf{H}}_j\hat{\mathbf{H}}_j^*$, then it is possible to rewrite the deterministic and stochastic ML with respect to $\hat{\mathbf{K}}_{\hat{h}}$ instead of $\hat{\mathbf{K}}_y$ by:

$$
\begin{aligned}
V_{\mathrm{DML}}(\tau) &= \mathrm{Tr}\left[\mathbf{\Pi}_{\mathbf{A}}^{\perp}(\tau)\hat{\mathbf{K}}_{\hat{h}}\right] & (4.27)\\
V_{\mathrm{SML}}(\tau) &= \ln\left|\mathbf{A}(\tau)\hat{\mathbf{K}}_s(\tau)\mathbf{A}^*(\tau) + \hat{\sigma}_r^2(\tau)\mathbf{I}\right| & (4.28)
\end{aligned}
$$

with,

$$\hat{\mathbf{K}}_s(\tau) = \mathbf{A}^{\dagger}(\tau)\left(\hat{\mathbf{K}}_{\hat{h}} - \hat{\sigma}_r^2(\tau)\mathbf{I}\right)\mathbf{A}^{\dagger *}(\tau)$$

$$
\begin{aligned}
\hat{\sigma}_r^2(\tau) &= \frac{\mathrm{Tr}\left(\mathbf{\Pi}_{\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)}^{\perp}\hat{\mathbf{K}}_y\right)}{m(m-d)} & (4.29)\\
&= \frac{\mathrm{Tr}\left(\mathbf{\Pi}_{\mathbf{A}}^{\perp}(\tau)\hat{\mathbf{K}}_{\hat{h}}\right)}{p-d} + O(1/\sqrt{L})\\
&= \frac{\mathrm{Tr}\left(\mathbf{\Pi}_{\mathbf{D}}^{\perp}(\mathbf{d})\hat{\mathbf{K}}_y\right)}{m(m-p)} + O(1/\sqrt{L})
\end{aligned}
$$

For the proof of (4.29) see Appendix B.  The interpretation of this result is that the signal samples can be replaced with the least squares estimate of the channel impulse response without any loss, if and only if (4.23) holds with equality.  $\hat{\mathbf{K}}_{\hat{h}}$ is a sufficient statistic for the estimation of $\tau$. It is of course not sufficient for the estimation of the gains as we need all the individual observations for this purpose (4.12).  $\hat{\mathbf{K}}_{\hat{h}}$ appears to be a more compact statistic than $\hat{\mathbf{K}}_y$ under the following conditions:

- Channel impulse response duration $\max_i\tau_i - \min_i\tau_i < (p-S_a)T$, ($S_a$ is the pulse duration).

- The training sequence must be located: this implies that the propagation conditions are good enough so that it can be detected.

- $p > d$. This condition is required to pseudo-inverse $\mathbf{A}$, in other words $\mathbf{A}$ must be full column rank.

- $\mathbf{D}^*(\mathbf{d})\mathbf{D}(\mathbf{d}) = m\mathbf{I}$

From now on, we will use $\hat{\mathbf{K}}_{\hat{h}}$ instead of $\hat{\mathbf{K}}_y$ as a more compact sufficient statistic. We call the column vector $\mathbf{a}(\tau)$ the steering vector by analogy with the steering vector of an antenna array $\check{\mathbf{a}}(\theta)$ (4.8).

The conclusion is that the variance error of the channel impulse response is reduced by the factor $m$ compared to the noise variance. The gain is $10 \log m$. For an estimation of $p = 10$ coefficients it represents $(m = q - p + 1)$:

- 17.40 dB for synchronization bursts,

- 15.05 dB for access bursts,

- 12.30 dB for traffic bursts.

For this reason we will not worry about noise properties. For instance, the assumption of white noise may be not satisfied if it is due to one strong interferer.

### 4.2.4  Cramer-Rao Bound (CRB)

The Cramer-Rao bound provides a lower bound on the variance error of any unbiased estimator [30]. Therefore it is interesting to compare an estimator with its Cramer-Rao bound. It should be noted that this bound is not reachable in all cases. This bound can be written in the following way:

$$E[(\theta - \hat{\theta})(\theta - \hat{\theta})^*] \geq \mathbf{CRB}(\theta) \tag{4.30}$$

The Cramer-Rao bound is given by the inverse of the Fisher information matrix defined by:

$$\mathbf{FIM}(\theta) = \mathrm{E}\left[\frac{\partial p(y|\theta)}{\partial \theta}\frac{\partial p(y|\theta)}{\partial \theta^T}\right] \tag{4.31}$$

where $p(y|\theta)$ refers to the conditional probability density with the knowledge of $\theta$. In our case $\theta = \tau$, relation (4.30) shows in particular that the diagonal elements of the Cramer-Rao bound represent the minimum variance error of the estimated delays respectively.

The asymptotic, i.e. when $L \to \infty$, deterministic Cramer-Rao bound on $\tau$ is given by [27]:

$$\mathbf{CRB}_{\mathrm{DET}}(\tau) = \frac{\sigma_r^2}{2L}\left[\mathrm{Re}\left(\mathbf{R}(\tau) \odot \mathbf{K}_s^T\right)\right]^{-1} \tag{4.32}$$

$\odot$ refers to the Schur product matrix (the element by element product) and:

$$\mathbf{R}(\tau) = \frac{d\mathbf{A}^*(\tau)}{d\tau}\mathbf{\Pi}_{\mathbf{A}}^{\perp}(\tau)\frac{d\mathbf{A}(\tau)}{d\tau} \tag{4.33}$$

This bound cannot be reached. This is a direct consequence of the inefficiency of DML as discussed in section 4.2.1.

The asymptotic stochastic Cramer-Rao bound on $\tau$ is [26]:

$$\mathbf{CRB}_{\text{STO}}(\tau) = \frac{\sigma_r^2}{2L} \left[ \text{Re} \left( \mathbf{R}(\tau) \odot \mathbf{U}^T(\tau) \right) \right]^{-1} \tag{4.34}$$

with $\mathbf{R}(\tau)$ defined in (4.33) and,

$$\begin{aligned} \mathbf{U}(\tau) &= \mathbf{K}_s \mathbf{A}^*(\tau) \mathbf{K}_{\hat{h}}^{-1} \mathbf{A}(\tau) \mathbf{K}_s \\ &= \left[ \mathbf{K}_s^{-1} + \sigma_r^2 \mathbf{K}_s^{-1} (\mathbf{A}^*(\tau) \mathbf{A}(\tau))^{-1} \mathbf{K}_s^{-1} \right]^{-1} \end{aligned}$$

Unlike the deterministic Cramer-Rao bound, the stochastic Cramer-Rao bound is reachable, which is a direct consequence of the efficiency of the SML.

## 4.2.5   Some well-known estimators

Maximum likelihood estimators require a multi-dimensional search that is difficult to achieve. In general, local extrema may appear if a descent type method is used. Newton iterative algorithms have been studied in [31]. There are also some other iterative algorithms that maximize the likelihood function such as the alternating projection approach [32] and the estimation-maximization algorithm [33–35].

The best well-known suboptimal estimator is obtained by matched filtering the received signal. In other words, cross-correlating the channel impulse response with the GMSK pulse. With our notations, that corresponds to:

$$\hat{\tau} = \arg \max_{\tau} \frac{1}{L} \sum_{j=1}^{L} |\mathbf{a}^*(\tau) \hat{\mathbf{H}}_j|^2 = \arg \max_{\tau} \mathbf{a}^T(\tau) \hat{\mathbf{K}}_{\hat{h}} \mathbf{a}(\tau) \tag{4.35}$$

For two closely spaced delays with variance $a_1$ and $a_2$, it is straightforward to show that this function produces only one maximum at:

$$\hat{\tau} = \frac{a_1 \tau_1 + a_2 \tau_2}{a_1 + a_2}$$

Another classical estimator is the Capon's beamformer [36] that can be written as:

$$\hat{\tau} = \arg \max_{\tau} \frac{1}{\mathbf{a}^*(\tau) \hat{\mathbf{K}}_{\hat{h}}^{-1} \mathbf{a}(\tau)} \tag{4.36}$$

These estimators (matched filter and Capon's beamformer) are equivalent to the maximum likelihood estimator when the delays are sufficiently spaced and at

high signal-to-noise ratios. In other cases, i.e. when the delays are closely spaced or the signal-to-noise ratio is low, they are strongly biased and fail to resolve the delays. Moreover, the Capon's beamformer requires a matrix inversion that may be ill conditioned.

This is the main reason why we prefer the algorithms described in the following. They are approximations of the maximum likelihood estimator under some conditions. They are computationally simple and have good asymptotic properties.

An eigenvalue decomposition (EVD) of $\mathbf{K}_{\hat{h}}$ will be useful:

$$\mathbf{K}_{\hat{h}} = \mathbf{E}_s \mathbf{\Lambda}_s \mathbf{E}_s^* + \sigma_r^2 \mathbf{E}_n \mathbf{E}_n^* = \mathbf{E}_s (\mathbf{\Lambda}_s - \sigma_r^2 \mathbf{I}) \mathbf{E}_s^* + \sigma_r^2 \mathbf{I} \qquad (4.37)$$

with $\mathbf{\Lambda}_s = \mathrm{diag}(\lambda_1, \cdots, \lambda_d)$ the strongest $d$ eigenvalues, $\mathbf{E}_s$ the rectangular matrix composed from the corresponding $d$ normalized eigenvectors, and $\mathbf{E}_n$ the rectangular matrix composed from the other $p - d$ normalized eigenvectors. It is common to say that $\mathbf{E}_s$ spans the signal subspace and $\mathbf{E}_n$ the noise subspace. $\mathbf{E}_s$ and $\mathbf{E}_n$ are orthogonal.

Note that the noise estimated by the equations (4.21) or (4.29) is just the mean value of the last eigenvalues that correspond to the noise subspace.

### Multiple SIgnal Classification (MUSIC)

In large samples approximation, i.e $L \to \infty$, another way of writing the deterministic maximum likelihood is the following (proof in [27]):

$$\hat{\tau} = \arg \min_{\tau} \mathrm{Tr} \left( \mathbf{A}^*(\tau) \hat{\mathbf{E}}_n \hat{\mathbf{E}}_n^* \mathbf{A}(\tau) \mathbf{K}_s \right) \qquad (4.38)$$

Thus, if $\mathbf{K}_s$ is diagonal, i.e. the gains are uncorrelated, the trace is minimized by minimizing each term of the sum (represented by the trace) separately. We obtain the well-known MUSIC estimator. MUSIC is thus asymptotically equivalent to the deterministic maximum likelihood when the delays are uncorrelated. The $d$ dimensional search is reduced to a one dimensional search of $d$ minimum values that are the $d$ lowest values of $\|\hat{\mathbf{E}}_n^* \mathbf{a}(\tau)\|^2$.

Note that the computation of the eigenvectors associated to the $d$ strongest eigenvalues is easier as the corresponding eigenvalues are in general very different[6]. Moreover if $2d < p$, it is more convenient to compute $\mathbf{E}_s$. $\mathbf{E}_n \mathbf{E}_n^*$ is obtained by

---

[6]The convergence speed of an iterative algorithm for eigenvalues decomposition is in general proportional to the ratios between consecutive eigenvalues.

$\mathbf{I} - \mathbf{E}_s\mathbf{E}_s^*$. It is common to search for the $d$ peaks of the following function, also called MUSIC spectrum:

$$\Phi(\tau) = \frac{1}{\mathbf{a}^*(\tau)(\mathbf{I} - \hat{\mathbf{E}}_s\hat{\mathbf{E}}_s^*)\mathbf{a}(\tau)} \qquad (4.39)$$

MUSIC has been widely studied in the literature, see [27,37,38].

**Weight Subspace Fitting (WSF)**

This estimator is based on the fact that the range space of $\mathbf{A}(\tau)$ is the same as $\mathbf{E}_s$. This implies that there is one unique full rank matrix $\mathbf{T}$ so that: $\mathbf{E}_s = \mathbf{A}(\tau)\mathbf{T}$. In other words, $\tau$ is the unique vector that satisfies this relation. As $\hat{\mathbf{E}}_s$ is corrupted by noise, finding the estimate can be formulated in the following way [39–41]:

$$[\hat{\tau}, \hat{\mathbf{T}}] = \arg\min_{\tau,\mathbf{T}} \|\hat{\mathbf{E}}_s\mathbf{W}^{\frac{1}{2}} - \mathbf{A}(\tau)\mathbf{T}\|^2 \qquad (4.40)$$

where $\mathbf{W}$ is a weighting matrix for the signal subspace.

$\mathbf{T}$ can be eliminated easily and the WSF estimator becomes:

$$\hat{\tau} = \arg\min_{\tau} \mathrm{Tr}\left(\mathbf{\Pi}_{\mathbf{A}}^{\perp}(\tau)\hat{\mathbf{E}}_s\mathbf{W}\hat{\mathbf{E}}_s^*\right) \qquad (4.41)$$

There exists an optimum weighting matrix in the sense that it minimizes the asymptotic covariance error [39]:

$$\hat{\tau} = \arg\min_{\tau} \mathrm{Tr}\left(\mathbf{\Pi}_{\mathbf{A}}^{\perp}(\tau)\hat{\mathbf{E}}_s\mathbf{W}_{\mathrm{OPT}}\hat{\mathbf{E}}_s^*\right) \qquad (4.42)$$

with,

$$\mathbf{W}_{\mathrm{OPT}} = (\mathbf{\Lambda}_s - \sigma_r^2\mathbf{I})^2\mathbf{\Lambda}_s^{-1} \qquad (4.43)$$

This matrix can be replaced in practice with a consistent estimate without affecting the asymptotic properties of the estimator. The corresponding estimator is a large samples approximation of the stochastic maximum likelihood estimation [26]. It is efficient in the sense that its asymptotic covariance corresponds to the stochastic Cramer-Rao bound (4.34). Note that at high signal-to-noise ratios and large samples, the deterministic and stochastic maximum likelihood estimators coincide.

# 4.3   Frequency approach

## 4.3.1   Formulation

In this case, we perform exactly the same computations as before but on the Fourier transform of the observations. We then have from (4.25):

$$\bar{\tilde{h}}_j(f) = \tilde{h}_j(f) = \bar{a}(f)\sum_{i=1}^{n} s_{ij}e^{-j2\pi\tau_i f} + \bar{w}_j(f) \qquad 1 \le j \le L \qquad (4.44)$$

$\bar{(.)}$ denotes the Fourier transform of $(.)$, $w$ is a complex Gaussian noise with reduced variance $\sigma_r^2 = \frac{\sigma^2}{m}$.

We take the same number of samples as in the temporal domain, i.e. $p$ samples. Let the set of frequencies of these samples be $f = (f_1, \cdots, f_p)$ on the nonzero support of $\bar{a}(f)$. We can then express the observations in the following way:

$$\bar{\tilde{\mathbf{H}}}_j = \tilde{\mathbf{H}}_j = \mathbf{\Gamma}(\gamma)\tilde{\mathbf{A}}(\tau)\mathbf{S}_j + \bar{\mathbf{W}}_j \qquad 1 \le j \le L \qquad (4.45)$$

with $\mathbf{\Gamma}(\gamma) = \mathrm{diag}(\gamma)$, $\gamma = [\bar{a}(f_1)\cdots\bar{a}(f_p)]^T$ and:

$$\tilde{\mathbf{A}}(\tau) = \begin{pmatrix} e^{-2\pi j f_1 \tau_1} & \cdots & e^{-2\pi j f_1 \tau_d} \\ \vdots & & \vdots \\ e^{-2\pi j f_p \tau_1} & \cdots & e^{-2\pi j f_p \tau_d} \end{pmatrix}$$

The advantage of this new representation is that it corresponds exactly to the estimation of the angles of arrival of $d$ sources impinging on a uniform linear array (ULA) of $p$ sensors (4.8) with different sensor gains [42, 43].

Assuming that the Nyquist/Shannon criterion is satisfied at the bit rate, the Fourier transform can be written in the following way:

$$\bar{\tilde{\mathbf{H}}}_j = \tilde{\mathbf{H}}_j = \mathcal{F}\hat{\mathbf{H}}_j \qquad (4.46)$$

with $\mathcal{F}$ being the following $p \times p$ matrix:

$$\mathcal{F} = T \begin{pmatrix} e^{-2\pi j f_1 T} & \cdots & e^{-2\pi j p f_1 T} \\ \vdots & & \vdots \\ e^{-2\pi j f_p T} & \cdots & e^{-2\pi j p f_p T} \end{pmatrix}$$

$T$ is the bit period. Equation (4.45) is similar to (4.7). The noise remains Gaussian. It is white if $\mathcal{F}$ is orthonormal, which is the case if an FFT is used. It is worth noting here that there are no edge effects while performing the FFT on $\hat{\mathbf{H}}_j$ since in practice $p = 10$ and $pT$ is about 37 $\mu$s which is longer than typical durations ($< 15\mu$s).

The new covariance is:

$$\mathbf{K}_{\tilde{h}} = \mathcal{F}\mathbf{K}_{\hat{h}}\mathcal{F}^* \qquad (4.47)$$

which gives directly the new covariance $\mathbf{K}_{\tilde{h}}$ from the latter one $\mathbf{K}_{\hat{h}}$.

The noise covariance becomes:

$$\mathbf{K}_{\bar{w}} = \sigma_r^2 \mathcal{F}\mathcal{F}^* \qquad (4.48)$$

If the sample frequencies are chosen uniformly, i.e. $f_i = f_1 + (i-1)\Delta$, the matrix $\tilde{\mathbf{A}}(\tau)$ has a Vandermonde structure. For practical reasons, we choose $\Delta = \frac{1}{pT}$ so that $\mathcal{F}$ becomes an FFT matrix and the noise $\bar{\mathbf{W}}_j$ remains white. $f_1$ is chosen so that the samples are centered, we then have:

- $f_1 = -f_p = -\frac{p-1}{2pT}$.

- $f_i = \frac{i - \frac{p+1}{2}}{pT}$

- $\mathcal{F}\mathcal{F}^* = \mathcal{F}^*\mathcal{F} = I$

In this case, the steering vector is $\boldsymbol{\Gamma}(\gamma)\tilde{\mathbf{a}}(\tau)$ with:

$$\tilde{\mathbf{a}}(\tau) = [e^{-2\pi j f_1 \tau}, \cdots, e^{-2\pi j f_p \tau}]^T = e^{-2\pi j f_1 \tau}[1, e^{-2\pi j \Delta \tau}, \cdots, e^{-2\pi j p \Delta \tau}]^T \qquad (4.49)$$

It should be noted here that there is no ambiguity on the possible value of $\tau$ if and only if the phase range does not exceed $2\pi$, in other words:

$$0 \leq 2\pi\Delta\tau \leq 2\pi \Longrightarrow 0 \leq \tau \leq \frac{1}{\Delta} = pT$$

This condition is supposed to be verified by the assumptions shown in Section 4.2.3.

All algorithms depicted in section 4.2.5 can be applied here too. We can also use some elegant algorithms taking advantage of the Vandermonde structure of the matrix $\tilde{\mathbf{A}}(\tau)$:

- Root MUSIC: the estimates are obtained by mean of roots of a polynomial that is related to the noise subspace,

- IQML: an iterative algorithm of successive least squares estimations without eigen-decomposition,

- Root WSF: the estimates are obtained by means of roots of a polynomial that will be estimated,

- ESPRIT: this estimator reduces the computations in the sense that it avoids the search for the maximum of a function (like the MUSIC spectrum function); the delays are obtained explicitly without search.

These methods are known as parametric methods. They are computationally attractive because they exploit the Vandermonde structure of $\tilde{\mathbf{A}}(\tau)$. In statistical array processing, this corresponds to the case of uniform linear arrays (ULA) with known calibration given by the elements of the diagonal matrix $\boldsymbol{\Gamma}(\gamma)$.

We note in the following $\tilde{\mathbf{E}}_s$ and $\tilde{\mathbf{E}}_n$ the new signal and noise subspace:

$$\mathbf{K}_{\tilde{h}} = \tilde{\mathbf{E}}_s \tilde{\boldsymbol{\Lambda}}_s \tilde{\mathbf{E}}_s^* + \tilde{\mathbf{E}}_n \tilde{\boldsymbol{\Lambda}}_n \tilde{\mathbf{E}}_n^*$$

### 4.3.2   Some well-known estimators

**Root MUSIC**

This algorithm is a variant of MUSIC and can be seen as an extension to the Pisarenko's approach [44].

MUSIC is defined by the $d$ values of $\tau$ that minimizes the following cost function:

$$\|\hat{\tilde{\mathbf{E}}}_n^* \mathbf{\Gamma}(\gamma)\tilde{\mathbf{a}}(\tau)\|^2 = \|\hat{\tilde{\mathbf{E}}}_n^* \mathbf{\Gamma}(\gamma)[1, z, \cdots, z^p]^T\|^2 \tag{4.50}$$

with $z = z(\tau) = e^{-j2\pi\tau\Delta}$, the right side following from (4.49). It is clear that in a noise free model these values are the $d$ roots of this function that lie on the unit circle:

$$z_i = e^{-j2\pi\tau_i\Delta} \qquad 1 \leq i \leq d$$

The delays are then obtained by means of the argument of these roots:

$$\tau_i = \frac{-1}{2\pi\Delta} \arg z_i$$

Since only the values of $z$ on the unit circle are of interest, we can substitute $z^*$ with $\frac{1}{z}$. The resulting function is a polynomial that can be written in the following way using the estimated noise subspace:

$$H(z) = \sum_{k=0}^{2(p-1)} \mathrm{Tr}_{k-p} \left[ \mathbf{\Gamma}^*(\gamma)\hat{\tilde{\mathbf{E}}}_n \hat{\tilde{\mathbf{E}}}_n^* \mathbf{\Gamma}(\gamma) \right] z^k \tag{4.51}$$

where $\mathrm{Tr}_k(.)$ denotes the trace of the $k$-th diagonal of $(.)$, in particular $\mathrm{Tr}_0(.)$ is the trace of $(.)$. In this case the matrix is hermitian and we have $\mathrm{Tr}_k(.) = \mathrm{Tr}_{-k}^*(.)$.

The Root MUSIC estimator is given by taking the argument of the $n$ closest roots to the unit circle of $H(z)$ that are inside or on the unit circle. Note that MUSIC is given by the minimum of (4.50) on the unit circle and is therefore not equivalent to Root MUSIC. It has been shown in [45] that Root MUSIC is asymptotically equivalent to MUSIC but as all parametric methods, it has a better threshold at low signal-to-noise ratios.

**Iterative Quadratic Maximum likelihood (IQML) and Root WSF**

These two algorithms are based on a different parameterization. Let $\mathbf{B}(\mathbf{b})$ be the following $p \times (p - d)$ matrix [46, 47]:

$$\mathbf{B}^*(\mathbf{b}) = \begin{pmatrix} b_d & \cdots & b_0 & 0 & \cdots & 0 \\ 0 & b_d & \cdots & b_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & b_d & \cdots & b_0 \end{pmatrix} \tag{4.52}$$

where $\mathbf{b} = [b_0 \cdots b_d]^T$ are the set of coefficients of the following polynomial:

$$\sum_{i=0}^{d} b_i z^{d-i} = b_0 \prod_{i=1}^{d} (z - e^{-j2\pi\tau_i\Delta}) \tag{4.53}$$

From the definition of $\mathbf{B}(\mathbf{b})$, $\mathbf{B}^*(\mathbf{b})\tilde{\mathbf{A}}(\tau) = 0$, and since the rank of $\mathbf{B}(\mathbf{b})$ is by construction $p - d$, it follows that its columns span the null subspace of $\tilde{\mathbf{A}}(\tau)$:

$$\mathbf{\Pi_B}(\mathbf{b}) = \mathbf{\Pi}_{\tilde{\mathbf{A}}}^{\perp}(\tau) \tag{4.54}$$

and,

$$\mathbf{\Pi}_{\mathbf{\Gamma}^{-*}\mathbf{B}}(\mathbf{b}) = \mathbf{\Pi}_{\mathbf{\Gamma}\tilde{\mathbf{A}}}^{\perp}(\tau) \tag{4.55}$$

It is equivalent to estimate $\mathbf{b}$ or $\tau$ as they are directly related. It is however easier to estimate $\mathbf{b}$ if an iterative approach is used. From definition (4.52), the DML criterion (4.27) can be written as:

$$V_{\mathrm{DML}}(\mathbf{b}) = \mathrm{Tr}\left[\mathbf{\Pi}_{\mathbf{\Gamma}^{-*}\mathbf{B}}(\mathbf{b})\hat{\mathbf{K}}_{\tilde{h}}\right] \tag{4.56}$$

The IQML algorithm is an iterative algorithm that tries to minimize this cost function iteratively by minimizing the following linear problem at iteration $k$:

$$\mathbf{b}_k = \arg\min_{\mathbf{b}} \mathrm{Tr}\left[\mathbf{B}(\mathbf{b})\left[\mathbf{B}^*(\mathbf{b}_{k-1})\mathbf{\Gamma}^{-1}\mathbf{\Gamma}^{-*}\mathbf{B}(\mathbf{b}_{k-1})\right]^{-1}\mathbf{B}^*(\mathbf{b})\mathbf{\Gamma}^{-1}\hat{\mathbf{K}}_{\tilde{h}}\mathbf{\Gamma}^{-*}\right] \tag{4.57}$$

$\mathbf{b}$ can be initialized so that $\mathbf{B}^*(\mathbf{b_0})\mathbf{\Gamma}^{-1}\mathbf{\Gamma}^{-*}\mathbf{B}(\mathbf{b_0}) = \mathbf{I}$. Each iteration is a simple least squares estimation. To avoid the trivial solution $\mathbf{b} = 0$, one can add the constraint that $\|\mathbf{b}\| = 1$ by normalizing $\mathbf{b}_k$ at each iteration. One can also use the fact that the roots lie on the unit circle so that the elements of $\mathbf{b}$ must satisfy the conjugate symmetry constraint:

$$b_i = b_{d-i}^* \quad 0 \le i \le d \qquad \Longleftrightarrow \qquad \mathbf{Jb} = \mathbf{b}^c \tag{4.58}$$

This reduces the number of parameters to be estimated. Moreover $\mathbf{b}$ should be constrainted to ensure that the corresponding polynomials has all its zeroes on the unit circle. However, the gain in performances obtained by taking into account this constraint is negligible [47].

The advantage of IQML is that it does not require any eigenvalue decomposition. However, it is known to be inefficient and leads to biased estimates [48]. Recently a modified IQML approach has been proposed [49] which overcomes the bias of the basic version and is asymptotically efficient.

Root WSF is a derivation of the WSF estimator. Introducing the matrix $\mathbf{B}(\mathbf{b})$, the WSF criterion becomes:

$$\hat{\mathbf{b}} = \arg\min_{\mathbf{b}} \mathrm{Tr}\left(\mathbf{\Pi}_{\Gamma^{-*}\mathbf{B}}(\mathbf{b})\hat{\tilde{\mathbf{E}}}_s\tilde{\mathbf{W}}_{\mathrm{OPT}}\hat{\tilde{\mathbf{E}}}_s^*\right) \qquad (4.59)$$

with $\tilde{\mathbf{W}}_{\mathrm{OPT}}$ defined as in (4.43) by:

$$\tilde{\mathbf{W}}_{\mathrm{OPT}} = (\tilde{\mathbf{\Lambda}}_s - \hat{\sigma}_r^2\mathbf{I})^2\tilde{\mathbf{\Lambda}}_s^{-1}$$

The same algorithm as for IQML is used. However, in this special case, two iterations are enough [11]. This is due to the fact that in this case $[\mathbf{B}^*(\mathbf{b})\mathbf{\Gamma}^{-1}\mathbf{\Gamma}^{-*}\mathbf{B}(\mathbf{b})]^{-1}$ can be replaced with a consistent estimate without affecting the asymptotic performances [48].

For minimizing each step, one can use the following formula:

$$\mathrm{Tr}(\mathbf{ABCD}) = \mathrm{Vec}^*(\mathbf{B}^*)(\mathbf{A}^T \otimes \mathbf{C})\mathrm{Vec}(\mathbf{D})$$

Thus the step described by (4.57) reduces to the minimization of $\mathbf{b}^*\mathcal{L}\mathbf{b}$ with a constraint on $\mathbf{b}$. The solution is the eigenvector of $\mathcal{L}$ associated to the smallest eigenvalue.

### Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT)

The algorithm exploits the Vandermonde structure of $\tilde{\mathbf{A}}(\tau)$. $\tilde{\mathbf{A}}(\tau)$ satisfies the following invariance properties [50]:

$$\mathbf{J}_1\tilde{\mathbf{A}}(\tau) = \mathbf{J}_2\tilde{\mathbf{A}}(\tau)\mathbf{\Phi}(\tau) \qquad \mathbf{\Phi}(\tau) = \mathrm{diag}(e^{2\pi j\Delta\tau_1}, \cdots, e^{2\pi j\Delta\tau_d}) \qquad (4.60)$$

where $\mathbf{J}_1$ and $\mathbf{J}_2$ are selection matrices of the first and the last $p-1$ rows, i.e.:

$$\mathbf{J}_1 = [\mathbf{I}_{p-1} \ 0] \qquad \mathbf{J}_2 = [0 \ \mathbf{I}_{p-1}]$$

Our goal is to determine the matrix $\mathbf{\Phi}(\tau)$ which gives explicitly the different delays. Using the fact that the range space of $\tilde{\mathbf{E}}_s$ is the same as the range space of $\mathbf{\Gamma}(\gamma)\tilde{\mathbf{A}}(\tau)$, there exists one unique full rank matrix $\mathbf{T}$ so that:

$$\mathbf{J}_1\tilde{\mathbf{E}}_s = \mathbf{J}_1\mathbf{\Gamma}(\gamma)\tilde{\mathbf{A}}(\tau)\mathbf{T} \qquad \mathbf{J}_2\tilde{\mathbf{E}}_s = \mathbf{J}_2\mathbf{\Gamma}(\gamma)\tilde{\mathbf{A}}(\tau)\mathbf{T}$$

Noting $\mathbf{\Gamma}_1$ the matrix $\mathbf{\Gamma}$ after deleting its last row and last column, and $\mathbf{\Gamma}_2$ the matrix $\mathbf{\Gamma}$ after deleting its first row and first column, one can write:

$$\mathbf{J}_1\tilde{\mathbf{E}}_s = \mathbf{\Gamma}_1\mathbf{J}_1\tilde{\mathbf{A}}(\tau)\mathbf{T} \qquad \mathbf{J}_2\mathbf{E}_s = \mathbf{\Gamma}_2\mathbf{J}_2\tilde{\mathbf{A}}(\tau)\mathbf{T}$$

Using the invariance properties shown by (4.60), we get:

$$\mathbf{J}_1\tilde{\mathbf{E}}_s = \mathbf{\Gamma}_1\mathbf{J}_2\tilde{\mathbf{A}}(\tau)\mathbf{\Phi}(\tau)\mathbf{T} \qquad \mathbf{J}_2\tilde{\mathbf{E}}_s = \mathbf{\Gamma}_2\mathbf{J}_2\tilde{\mathbf{A}}(\tau)\mathbf{T}$$

It follows that:

$$\boldsymbol{\Gamma}_2\boldsymbol{\Gamma}_1^{-1}\mathbf{J}_1\tilde{\mathbf{E}}_s = \mathbf{J}_2\tilde{\mathbf{E}}_s\boldsymbol{\Psi}(t) \tag{4.61}$$

With $\boldsymbol{\Psi}(t) = \mathbf{T}^{-1}\boldsymbol{\Phi}(\tau)\mathbf{T}$. The eigenvalues of $\boldsymbol{\Psi}(t)$ directly yield the desired delays. We have to estimate this matrix. One way of doing that is by least squares estimation:

$$\hat{\boldsymbol{\Psi}}(\tau) = \left(\mathbf{J}_2\hat{\tilde{\mathbf{E}}}_s\right)^{\dagger}\boldsymbol{\Gamma}_2\boldsymbol{\Gamma}_1^{-1}\mathbf{J}_1\hat{\tilde{\mathbf{E}}}_s$$

Some algorithms may perform better than least squares: total least squares ES-PRIT [51], weighted ESPRIT [52], and multiple invariance ESPRIT [53].

### 4.3.3 Forward-Backward averaging

Forward-Backward averaging is a technique used to virtually double the available samples. It is based on the following equation which is a direct consequence of the Vandermonde structure of $\mathbf{A}(\tau)$:

$$\mathbf{J}\tilde{\mathbf{A}}^c(\tau) = \tilde{\mathbf{A}}(\tau)\phi$$

with $\phi = \mathrm{diag}(e^{2\pi(m-1)\tau_1}, \cdots, e^{2\pi(m-1)\tau_d})$ and $\mathbf{J}$ the anti-identity matrix. We can define a modified covariance matrix in the following way:

$$\frac{1}{2}(\mathbf{K}_{\tilde{h}} + \mathbf{J}\mathbf{K}_{\tilde{h}}^c\mathbf{J}) = \frac{1}{2}(\tilde{\mathbf{A}}(\mathbf{K}_{\tilde{h}} + \phi\mathbf{K}_{\tilde{h}}^c\phi^*)\tilde{\mathbf{A}}^*) \tag{4.62}$$

The reason for the appellation "forward-backward" is that while the indices of $\tilde{\mathbf{H}}_j$ run forward $(1, 2, \cdots, m)$ those of $\mathbf{J}\tilde{\mathbf{H}}_j$ run backward $(m, m-1, \cdots, 1)$. The consequence is that the extraction of the signal subspace from this new matrix correspond to the space spanned by $\tilde{\mathbf{A}}(\tau)$.

While Forward-Backward averaging is known to enhance the performance of some algorithms such as Root MUSIC or ESPRIT [54], it is however shown that it must not be used for a statistically efficient method such as Root WSF [55].

## 4.4 Simulations

In the following, simulations are presented for the simple case of a channel impulse response consisting of two delays equally powered. Synchronization bursts are used (downlink scenario). The GMSK pulse is filtered so that it fits the 200 kHz bandwidth around its carrier. We filtered it at $1/T$, i.e. between $-1/2T$ and $+1/2T \approx 135$ kHz. In Figure 4.1 the GMSK pulse and two filtered versions are displayed at $1/T$ and at $0.8/T$. It is shown that while the perfect GMSK pulse has a duration of 4 bit periods, the filtered versions duration is longer (8 bit periods at $1/T$). From now on, we will assume in simulations that the pulse
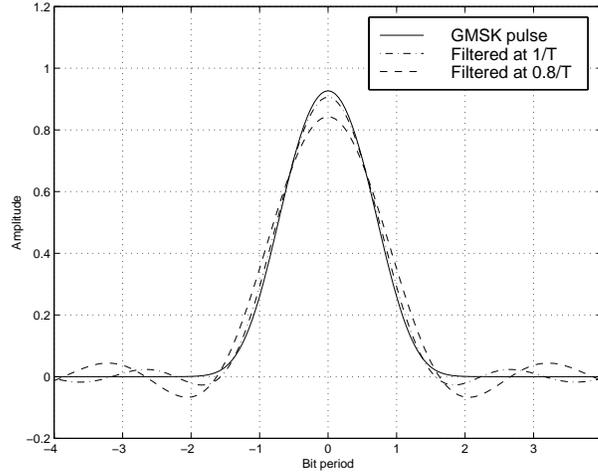
Figure 4.1: The GMSK main pulse and its filtered versions at $0.8/T$ and $1/T$.

filtered at $1/T$ is the transmitted one.

Both delays are assumed to be correlated. In the general case, the covariance matrix of the fading can be written as:

$$\mathbf{K}_s = 10^{\frac{\text{SNR}}{10}} \begin{pmatrix} 1 & \rho \\ \rho^* & 1 \end{pmatrix} \qquad (4.63)$$

with $\rho$ being the correlation factor between the two paths, and SNR the signal-to-noise ratio expressed in dB.

In Figure 4.2, simulations are done with temporal MUSIC (4.39) and the matched filter (4.35) assuming no correlation and a signal-to-noise ratio of 10 dB. There is one bit period separation between both delays. It is clearly shown that while the temporal MUSIC algorithm resolves the two delays (the two clear peaks), the matched filter has only one peak and is therefore unable to deduce any relevant information.

The following simulations are performed with the GMSK pulse function filtered at $1/T$, i.e. between $-1/2T$ and $+1/2T$ (see Figure 4.1). Since the duration of this pulse is about 8 bit periods, $p = 10$ coefficients are used for the channel impulse response estimation. The normalized Root Mean Square error (RMS) is drawn for the first delay (results for the second delay are similar):

- In Figure 4.3, simulations with respect to the SNR assuming uncorrelation
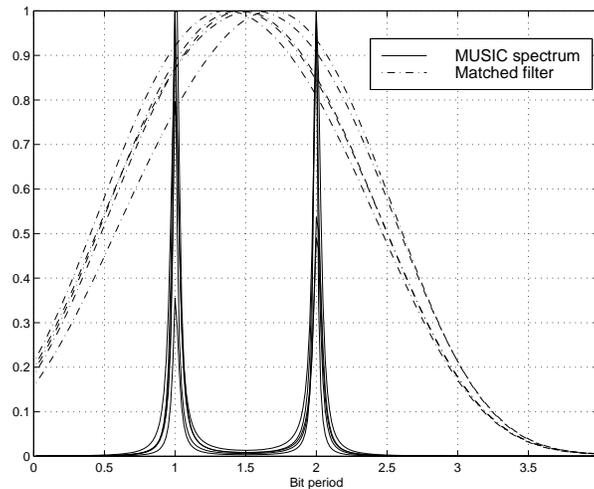
Figure 4.2: Temporal MUSIC and matched filtering (cross-correlation) at 10 dB, 5 independent trials are run: the matched filter is unable to resolve two delays equally powered spaced by one bit period.

$(\rho = 0)$.

- In Figure 4.4, simulations with respect to the SNR assuming a correlation factor of $\rho =$0.99.

- In Figure 4.5, simulations with respect to the number of bursts assuming uncorrelation $(\rho = 0)$.

- In Figure 4.6, simulations with respect to the delay separation assuming no correlation $(\rho = 0)$.

- In Figure 4.7, simulations on the Forward-Backward averaging effect on ESPRIT.

Note that a RMS error higher than 1 bit period indicates a failure in resolving both delays. As expected, simulations show that the best candidate is Root WSF since it is efficient even when the delays are correlated. Forward-Backward averaging is shown to be useful for suboptimal methods especially in critical conditions (when few observations are available). However it should not be used for statistically efficient methods such as Root WSF [55].
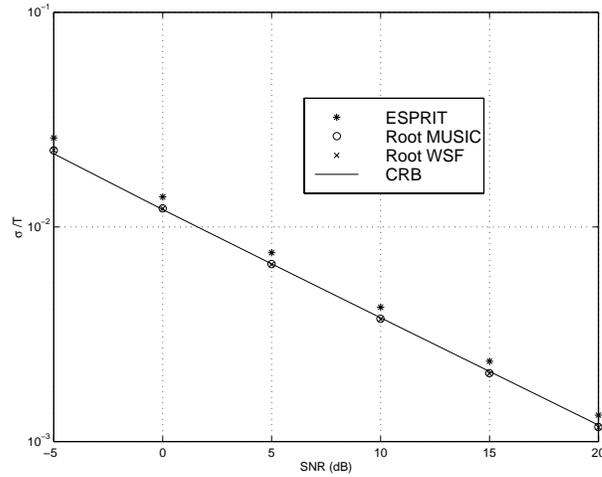
Figure 4.3: Algorithm performance with respect to the SNR, $L = 100$ synch. bursts are used, $\Delta\tau = T$, and $\rho = 0$: MUSIC and Root WSF are efficient.
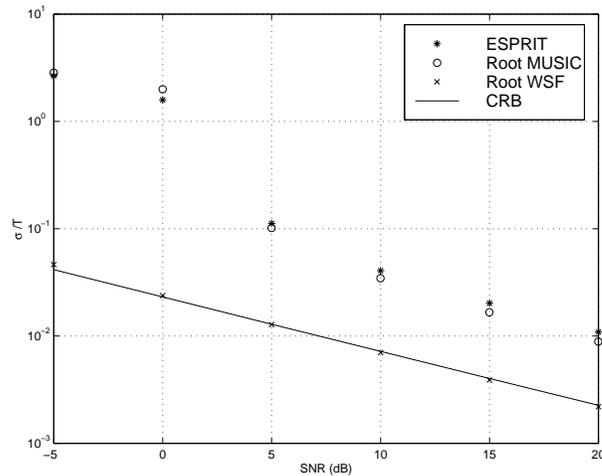


Figure 4.4: Algorithm performance with respect to the SNR, $L = 100$ synch. bursts are used, $\Delta\tau = T$, $\rho = 0.99$: only Root WSF is efficient.

## 4.5   Extension to an unknown modulation pulse

The subspace fitting algorithms described in previous sections assume perfect knowledge of the modulation pulse shape. However, these algorithms are very
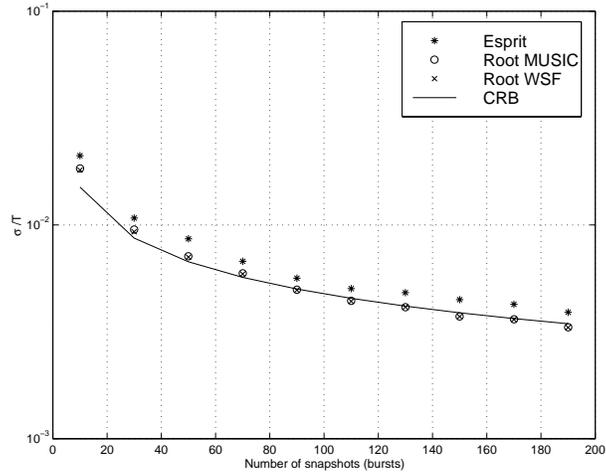
Figure 4.5: Algorithm performance with respect to the number of bursts at 8 dB, $\Delta\tau = T$, $\rho = 0$.
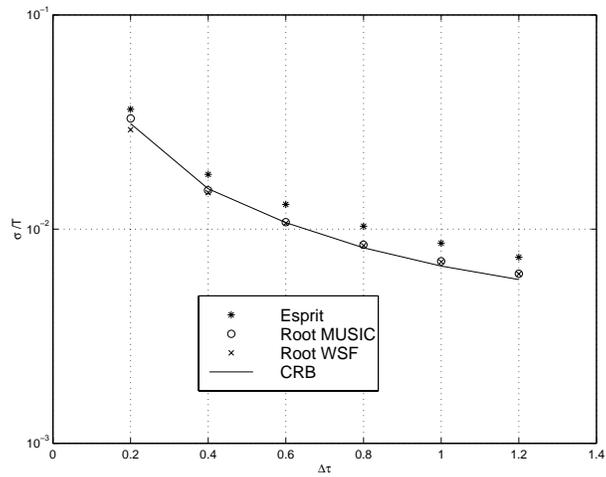


Figure 4.6: Algorithm performance with respect to the delays separation at 8 dB, $L = 100$ synch. bursts used, $\rho = 0$.

sensitive to the lack of information in the modulation pulse function. In Figure 4.8, simulations are done using the pulse filtered at $1/T$ and the delays estimator assumes a wrong pulse (the one filtered at $0.8/T$, see Figure 4.1). It results
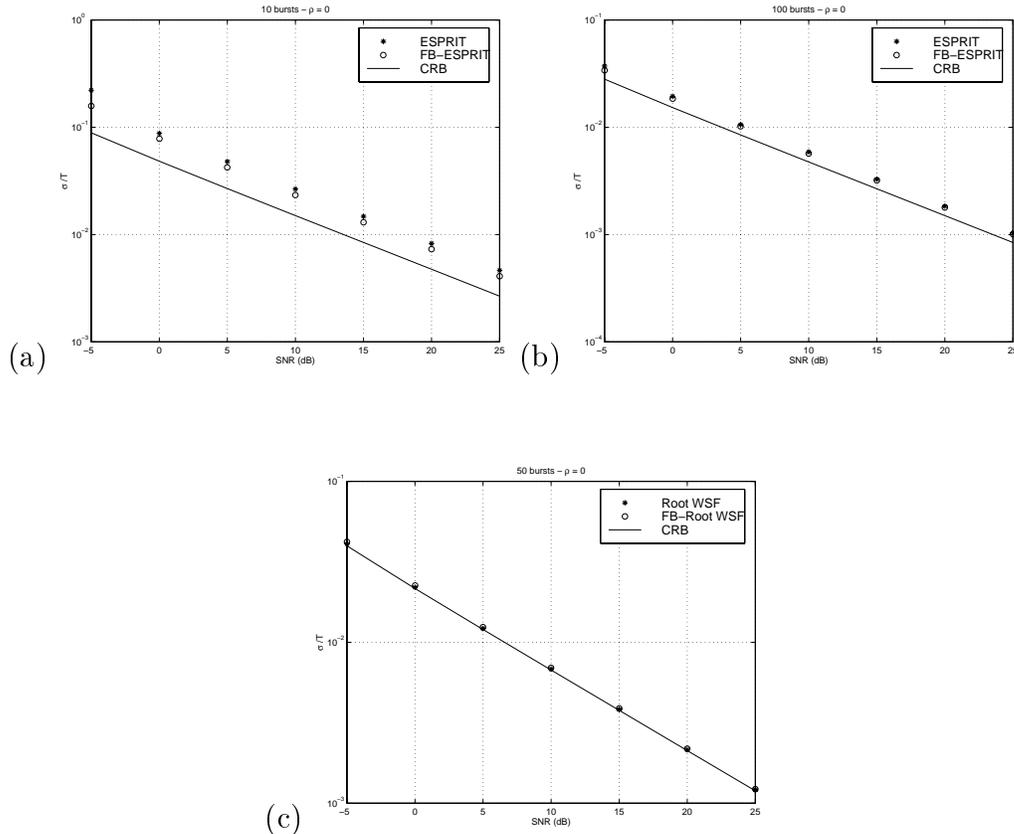
Figure 4.7: The Effect of forward-backward averaging: $\Delta\tau = 1T$, $\rho = 0$. (a) with ESPRIT using 10 synch bursts, (b) with ESPRIT using 100 synch. bursts, (c) with Root WSF using 50 bursts. Forward-Backward averaging is useful for suboptimal methods especially if few number of bursts are available. However, it should not be used for optimal methods such as Root WSF, Figure (c) shows no improvement in performance when using forward-backward averaging.

in biases in the estimation. In practice, it is hard to have a perfect knowledge of the pulse shape since it is modified by the transmission and reception filters which may differ between manufacturers. Depending on the degree to which the wrong pulse deviates from the true one, serious degradations may result.

In [56,57], the performances in presence of model errors are analyzed. In [58], such errors are assumed to be random and a new corrected version of the subspace fitting approach is derived. A new weighting matrix $\mathbf{W}$ (4.41) is found to include the uncertainty on the model. Such approach is not valid in our case

since the pulse shape is purely deterministic and does not vary in time.

We show in this section that under some conditions, the delays can be estimated without knowledge of the pulse shape. We perform a joint estimation of the pulse shape and the delays making use of the uncorrelation of successive observations.
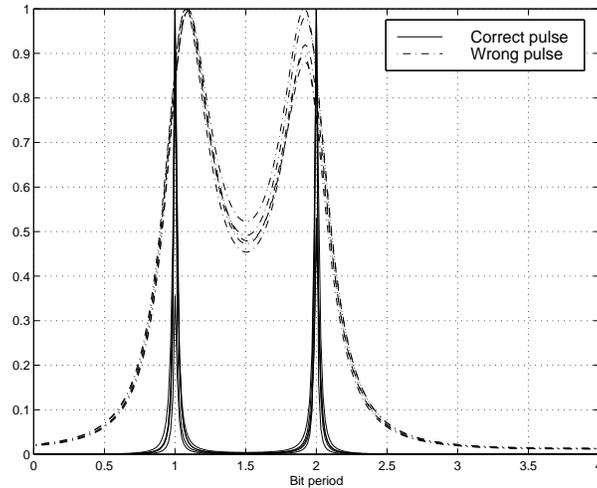


Figure 4.8: Temporal MUSIC and matched filtering (cross-correlation): 5 independent trials are run at 10 dB and using 100 synch. bursts showing that a bias appears when a wrong pulse is assumed. The wrong pulse here is the GMSK main pulse filtered at 0.8/T.

The problem to be solved is the one shown in (4.45) assuming an unknown pulse shape, in other words $\gamma$ is unknown. In fact, this problem is equivalent to the problem of the estimation of $d$ Gaussian sources that impinge on $m$ antennae with unknown calibration. It is shown in [59] that the problem is identifiable under some conditions. For the particular case of Uniform Linear Array (ULA), the problem is not identifiable in the sense that the sources can be located to within an unknown initial phase. In other words, it is possible to estimate only the difference of angles of arrival of two given paths.

In our case, adding some assumptions on the pulse shape removes this uncertainty. The assumptions we take are the facts that the pulse is real and symmetric. For the particular case of Difference Times of Arrival (DToA) computation, this uncertainty is not an issue since it results in an unknown common time refer-

ence for all times of arrival. This reference will be removed when computing the differences.

## 4.5.1   Iterative approach

Iterative algorithms have been proposed in the literature, see [42, 43]. These algorithms jointly estimate the diagonal elements of the matrix $\boldsymbol{\Gamma}(\gamma)$ and the delays $\tau$ (see equation 4.45). They proposed a general procedure that uses any algorithm that works assuming known matrix $\boldsymbol{\Gamma}$. It is based on the following equation that stems from the fact that $\boldsymbol{\Gamma}(\gamma)\tilde{\mathbf{A}}(\tau)$ lies in the signal subspace represented by $\tilde{\mathbf{E}}_s$, i.e:

$$[\hat{\gamma} \ \hat{\tau}]^T = \arg \max_{\gamma,\tau} \sum_{i=1}^{d} \|\hat{\tilde{\mathbf{E}}}_s^* \boldsymbol{\Gamma}(\gamma)\tilde{\mathbf{a}}(\tau_i)\|^2 \tag{4.64}$$

This is equivalent to:

$$[\hat{\gamma} \ \hat{\tau}]^T = \arg \max_{\gamma,\tau} \gamma^* \sum_{i=1}^{d} \mathbf{D}_i^* \hat{\tilde{\mathbf{E}}}_s \hat{\tilde{\mathbf{E}}}_s^* \mathbf{D}_i \gamma \tag{4.65}$$

where $\mathbf{D}_i = \text{diag}[\tilde{\mathbf{a}}(\tau_i)]$. In other words, $\gamma$ is the eigenvector associated to the maximum eigenvalue of the matrix $\sum_{i=1}^{d} \mathbf{D}_i^* \hat{\tilde{\mathbf{E}}}_s \hat{\tilde{\mathbf{E}}}_s^* \mathbf{D}_i$. The proposed algorithm is similar to the one in [43].

Once this eigenvector is computed, we eliminate the uncertainty on $\gamma$ by incorporating the constraint that the pulse is symmetric and real (i.e. its Fourier transform is real and symmetrical):

- First of all, the delays are estimated with a standard algorithm assuming that $\boldsymbol{\Gamma}(\gamma)$ is the identity, or using $\boldsymbol{\Gamma}(\gamma)$ that corresponds to the original GMSK pulse.

- $\gamma$ is then computed using (4.65) by taking the elements of the eigenvector associated to the highest eigenvalue $\mathbf{V}_{\text{MAX}} = [v_1 \cdots v_m]$:

$$\gamma_i = \sqrt{|v_i||v_{m+1-i}|} \qquad 1 \le i \le m$$

- The delays are computed again using $\gamma$ found in previous step and so on until some convergence criterion is achieved.

Note that to obtain an estimation of the pulse shape, $\gamma$ should be normalized so that $\gamma^* \gamma = 1$.

## 4.5.2   A modified ESPRIT algorithm

In the following, a non-iterative ESPRIT-like algorithm is derived in contrast to the iterative approach described in previous section where the standard ESPRIT can be used to compute the delays at each iteration. We note $\mathbf{D}(\eta) = \text{diag}(\eta)$ the diagonal matrix $\mathbf{\Gamma}_2\mathbf{\Gamma}_1^{-1}$, and $\mathbf{\Psi}(\tau) = \mathbf{T}^{-1}\mathbf{\Phi}(\tau)\mathbf{T}$. Note that $\eta$ is function of $\gamma$ and represents some unknown vector related to the pulse shape. From equation (4.61) it appears that the problem to be solved is the following:

$$[\hat{\eta}\ \hat{\mathbf{\Psi}}(\tau)]^T = \arg\min_{\eta\,\Psi} \|\mathbf{D}(\eta)\mathbf{J}_1\hat{\tilde{\mathbf{E}}}_s - \mathbf{J}_2\hat{\tilde{\mathbf{E}}}_s\mathbf{\Psi}(\tau)\|^2 \tag{4.66}$$

with the following constraints:

$$\begin{cases} \text{Im}(\eta) = 0 \\ \mathbf{J}\mathbf{D}(\eta)\mathbf{J}\mathbf{D}(\eta) = \mathbf{I} \end{cases} \tag{4.67}$$

These constraints are direct consequences of the assumptions we made on the pulse shape. In particular, we have:

$$\det[\mathbf{D}(\eta)] = 1 \tag{4.68}$$

The delays to be estimated are obtained from the phases of the eigenvalues of $\mathbf{\Psi}$ since $\mathbf{\Phi}$ is diagonal. $\hat{\mathbf{E}}_s$ is obtained from the estimated covariance matrix $\hat{\mathbf{K}}_{\tilde{h}} = \frac{1}{L}\sum_{j=1}^{L}\tilde{\mathbf{H}}_j\tilde{\mathbf{H}}_j^*$.

The solution to the problem shown in (4.66) without incorporation of the constraints (4.67) is given by [60]:

$$\hat{\mathbf{\Psi}}(\tau) = \left(\mathbf{J}_2\hat{\tilde{\mathbf{E}}}_s\right)^{\dagger}\mathbf{D}(\hat{\eta})\mathbf{J}_1\hat{\tilde{\mathbf{E}}}_s$$

Substituting this expression in 4.66, we get:

$$\begin{aligned} \hat{\eta} &= \arg\min_{\eta} \text{Tr}\left[\mathbf{D}(\eta)\mathbf{J}_1\hat{\tilde{\mathbf{E}}}_s\hat{\tilde{\mathbf{E}}}_s^*\mathbf{J}_1^T\mathbf{D}^*(\eta)\mathbf{\Pi}_{\mathbf{J}_2\hat{\tilde{\mathbf{E}}}_s}^{\perp}\right] \\ &= \arg\min_{\eta} \eta^*\left[\mathbf{\Pi}_{\mathbf{J}_2\hat{\tilde{\mathbf{E}}}_s}^{\perp} \odot \mathbf{J}_1(\hat{\tilde{\mathbf{E}}}_s\hat{\tilde{\mathbf{E}}}_s^*)^T\mathbf{J}_1^T\right]\eta \end{aligned}$$

In other words:

$$\hat{\eta} = \alpha\mathbf{V}_{\text{MIN}}$$

where $\mathbf{V}_{\text{MIN}} = [v_1\cdots v_{m-1}]$ is the eigenvector of $\mathbf{\Pi}_{\mathbf{J}_2\hat{\tilde{\mathbf{E}}}_s}^{\perp} \odot \mathbf{J}_1(\hat{\tilde{\mathbf{E}}}_s\hat{\tilde{\mathbf{E}}}_s^*)^T\mathbf{J}_1^T$ associated to the smallest eigenvalue and $\alpha$ a scale factor. The delays estimation are related to the phases of the eigenvalues of $\mathbf{\Psi}(\tau)$ and are then indifferent to the magnitude of $\alpha$. This is a set of solutions determined by an initial phase that is

related to the argument of $\alpha$.

Adding the constraint that the pulse is real and symmetric forces this solution to be unique. We propose the following solution: $\hat{\eta} = [\hat{\eta}_1 \cdots \hat{\eta}_{m-1}]$ with $\hat{\eta}_i = \sqrt{\frac{|v_i|}{|v_{m-i}|}}$ so that the constraints shown in (4.67) are satisfied.

It should be noted that the solution shown here is not optimal since the constraints have been applied on the unconstrainted solution.

The pulse can be deduced from $\hat{\eta}$ by:

$$\hat{\gamma}_i = \hat{a}(f_i) = \frac{\prod_{j=1}^{i-1} \frac{1}{\hat{\eta}_j}}{\sum_{k=1}^{m} \prod_{j=1}^{k-1} \frac{1}{\hat{\eta}_j^2}} \qquad 1 \leq i \leq m \qquad (4.69)$$

where we assumed that $\eta_0 = 1$.

### 4.5.3  The DToA special case

As discussed in Section 3.4, the localization procedure makes use of Differences Times of Arrival (DToA). In other words, it computes the differences of times of arrival calculated on several links between the mobile and the base stations.

In that case, the discussion on identifiability is no more relevant since the problem is related to the determination of a scale factor of $\gamma$ or $\eta$. Such a scale factor introduces a shift to all delays and does not affect the differences of times of arrival.

We can generalize the algorithms described above by combining all available links. The non-iterative approach will be written as:

$$[\hat{\eta}, \hat{\boldsymbol{\Psi}}_1, \hat{\boldsymbol{\Psi}}_2, \cdots]^T = \arg \min_{\eta, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \cdots} \sum_i \alpha_i \|\mathbf{D}(\eta)\mathbf{J}_1\hat{\bar{\mathbf{E}}}_{s_i} - \mathbf{J}_2\hat{\bar{\mathbf{E}}}_{s_i}\boldsymbol{\Psi}_i\|^2 \qquad (4.70)$$

where $\boldsymbol{\Psi}_i$ is the matrix $\boldsymbol{\Psi}(\tau)$ that correspond to the $i$-th link, $\eta$ the column vector that contains the pulse information that is shared on all links, $\hat{\bar{\mathbf{E}}}_{s_i}$ the estimated signal subspace on the $i$-th link, and $\alpha_i$ a weighting factor; it can be a function of the signal-to-interference ratio on the $i$-th link. It is however hard to set an optimal value to these coefficients. This problem remains an open question.

The solution is:

$$\hat{\bar{\boldsymbol{\Psi}}}_i = \left(\mathbf{J}_2\hat{\bar{\mathbf{E}}}_{s_i}\right)^\dagger \mathbf{D}(\hat{\eta})\mathbf{J}_1\hat{\bar{\mathbf{E}}}_{s_i}$$

and:

$$\hat{\eta} = \arg\min_{\eta} \eta^* \left[ \sum_i \alpha_i \left( \mathbf{\Pi}^{\perp}_{\mathbf{J}_2 \hat{\tilde{\mathbf{E}}}_{s_i}} \odot \mathbf{J}_1 (\hat{\tilde{\mathbf{E}}}_{s_i} \hat{\tilde{\mathbf{E}}}^*_{s_i})^T \mathbf{J}_1^T \right) \right] \eta$$

$\eta$ is computed from the eigenvector associated to the lowest eigenvalue $\hat{\eta} = \alpha \mathbf{V}_{\mathrm{MIN}}$.

## 4.5.4   Simulations

Simulations are performed here with the same assumptions as in 4.4. No correlation is assumed. In Figure 4.9 a comparison is done between the new ESPRIT-like approach assuming unknown pulse and the standard ESPRIT algorithm. The normalized standard deviation of the estimation of the first delay is drawn with respect to the SNR (the results on the second delay are the same). It is shown that the loss in performance due to the lack of knowledge of the pulse is around 2 dB only.
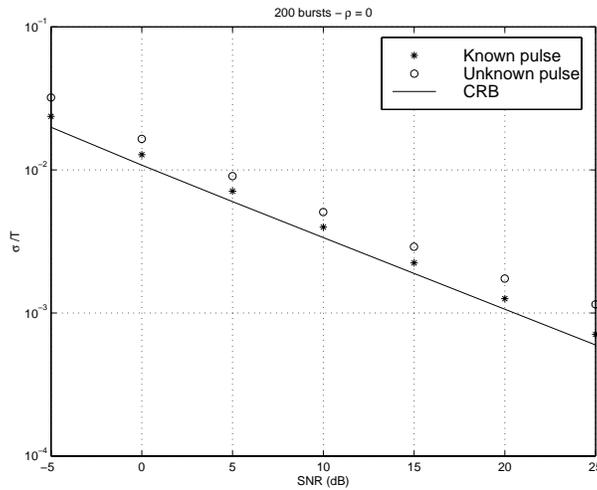


Figure 4.9: ESPRIT with and without the knowledge of the pulse, $\Delta\tau = T$, $L = 200$ synch. bursts are used.

In Figure 4.10 a comparison is done between the iterative approach done with Root MUSIC and the non-iterative ESPRIT-like proposed algorithm. The iterative approach appears to be more accurate, it is however far more complex.

As expected, the results are not as good as those done with the knowledge of the pulse. Such algorithms should be run if the propagation conditions are good
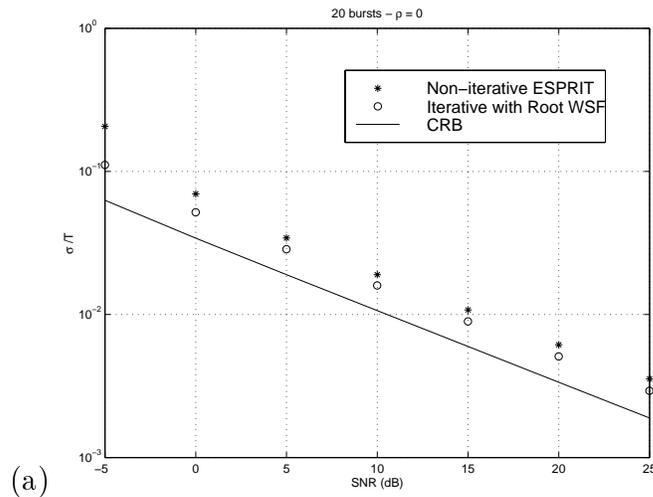
Figure 4.10: Algorithm performance with an unknown pulse, $\Delta\tau = T$, $\rho = 0$, '*' refers to the non-iterative ESPRIT-like's approach and 'o' refers to the iterative approach performed using Root WSF.

enough (few paths, high signal-to-interference ratio), so that the prior information on the pulse is easily corrected. Once a good estimation of the pulse is available, standard subspace algorithms should be used.

## 4.6  On complexity

We presented in this chapter a set of algorithms based on the extraction of the signal or the noise subspace. Finding a subspace from a covariance matrix requires a large amount of computation. However, as we are just interested in finding few eigenvectors, some algorithms such as the power iteration method [61] might be useful.

Recently, there has been some studies on subspace tracking. Basically, the goal is to be able to update quickly the signal subspace whenever a new observation is available [62,63], see also [64]. Moreover, performing an exponential weighting of the data enables tracking the moving delays. The projection approximation subspace tracking (PAST) [65] is a simple algorithm based on a recursive least squares estimation approach. It appears to be a promising technique.

# Chapter 5

# Hyperbolic trilateration

In this chapter, we deal with the problem of estimating the coordinates of the handset once enough times of arrival are available. This procedure called trilateration or data fusion is possible if information on the network synchronization is available. The problem reduced then to a non linear maximization.

## 5.1 Problem statement

We have $n$ times of arrival observations and $m$ differences of synchronization observations. All these observations are corrupted with noise.

The times of arrival can be written in the following way:

$$f_i = \frac{r_i}{c} + t_i - q_1 + b_i \qquad 1 \leq i \leq n \tag{5.1}$$

with,

- $r_i$ the distance between the mobile and the $i$-th base station,

- $c \approx 3.10^8$ m/s the speed of light,

- $t_i$ the transmission time of the $i$-th transmitter,

- $q_1$ a reference time[1],

- $f_i$ the estimated times of arrival,

- $b_i$ an additive white Gaussian noise.

---

[1]It should be noted that this reference has a different interpretation according to the location system. For instance in the GPS system [66] this will be the handset reference.

The $n$ base stations involved in the localization procedure are assumed to have different coordinates. Indeed the mobile can hear more cells that sites (in the case of trisectorized sites for instance) and the cells that belong to the same base station provide little extra information.

The synchronization information is expressed by the following $n$ equations:

$$\bar{t}_i = t_i - q_2 + w_i \qquad 1 \le i \le n \qquad (5.2)$$

Such measurements can be simply computed by measuring the times of arrival of different signals coming from different base stations to a known point as will be discussed in next chapter. $q_2$ is another reference time, and $w_i$ an additive white Gaussian noise.

We can rewrite these equations in a more compact way:

$$\begin{cases} \tilde{\mathbf{F}} = \mathbf{A}(\theta) + \mathbf{T} - q_1 \mathbf{1} + \mathbf{B}_f \\ \bar{\mathbf{T}} = \mathbf{T} - q_2 \mathbf{1} + \mathbf{B}_t \end{cases} \qquad (5.3)$$

with,

- $\theta$ the column vector of the coordinates to be estimated (in two dimensions $\theta = [x \ y]^T$),

- $\tilde{\mathbf{F}} = [f_1 \cdots f_n]^T$,

- $\mathbf{A}(\theta) = \frac{1}{c}[r_1 \cdots r_n]^T$,

- $\mathbf{T} = [t_1 \cdots t_n]^T$,

- $\mathbf{B}_f = [b_1 \cdots b_n]^T$,

- $\mathbf{B}_t = [w_1 \cdots w_m]^T$,

- $\mathbf{1} = [1 \cdots 1]^T$.

We assume that the noise vectors $\mathbf{B}_f$ and $\mathbf{B}_t$ are zero mean Gaussian variables with the following covariances:

- $\mathrm{E}[\mathbf{B}_f \mathbf{B}_f^T] = \mathbf{Q}_f$

- $\mathrm{E}[\mathbf{B}_t \mathbf{B}_t^T] = \mathbf{Q}_t$

- $\mathrm{E}[\mathbf{B}_f \mathbf{B}_t^T] = 0$

Let $\mathbf{Q} = \mathbf{Q}_t + \mathbf{Q}_f$. This covariance matrix contains the confidence values of the time of arrival measurements after incorporation of the synchronization information uncertainties.

Obviously the system (5.3) is equivalent to:

$$\tilde{\mathbf{F}} = \mathbf{A}(\theta) + \bar{\mathbf{T}} - t_r\mathbf{1} + \mathbf{B}$$

where $\mathbf{B} = \mathbf{B}_f + \mathbf{B}_t$ and $t_r = q_1 + q_2$. Incorporating $\bar{\mathbf{T}}$ in $\tilde{\mathbf{F}}$ the system becomes:

$$\mathbf{F} = \mathbf{A}(\theta) - t_r\mathbf{1} + \mathbf{B} \tag{5.4}$$

where $\mathbf{F} = \tilde{\mathbf{F}} - \bar{\mathbf{T}}$. These equation leads to a hyperbolic trilateration as in GPS [66]. $n$ must be strictly greater than two if the position estimation is required in two dimensions since three variables need to be estimated: $x$, $y$, and $t_r$. This is required for the system to be identifiable, but of course, the greater is $n$, the better is the accuracy. In the specific case of radio propagation, it should be noted that the times of arrival (i.e. $\tilde{\mathbf{F}}$) are biased in general because of multipath propagation. For instance, the direct line of sight may not exist at all.

We perform a joint estimation of $\theta$ and $t_r$. From (5.4) the maximum likelihood estimator is given by:

$$[\hat{\theta} \ \hat{t}_r] = \arg\max_{\theta,t_r} [\mathbf{F} - \mathbf{A}(\theta) - t_r\mathbf{1}]^T \mathbf{Q}^{-1} [\mathbf{F} - \mathbf{A}(\theta) - t_r\mathbf{1}] \tag{5.5}$$

Derivating with respect to $t_r$, we obtain:

$$\hat{t}_r(\theta) = \frac{\mathbf{1}^T\mathbf{Q}^{-1}[\mathbf{F} - \mathbf{A}(\theta)]}{\mathbf{1}^T\mathbf{Q}^{-1}\mathbf{1}} \tag{5.6}$$

Substituting in (5.5), we obtain a concentrate function to be minimized with respect to $\theta$:

$$\hat{\theta} = \arg\max_{\theta} [\mathbf{F} - \mathbf{A}(\theta)]^T \tilde{\mathbf{Q}} [\mathbf{F} - \mathbf{A}(\theta)] \tag{5.7}$$

with:

$$\tilde{\mathbf{Q}} = \mathbf{Q}^{-1} - \frac{\mathbf{Q}^{-1}\mathbf{1}\mathbf{1}^T\mathbf{Q}^{-1}}{\mathbf{1}^T\mathbf{Q}^{-1}\mathbf{1}}$$

Note that $\tilde{\mathbf{Q}}$ is rank deficient since $\tilde{\mathbf{Q}}\mathbf{1} = 0$. That is the reason why the time reference has been eliminated in (5.7) compared to (5.5).

## 5.2    Cramer-Rao bound

The Cramer-Rao bound on $\theta$ is given by (see definition in Section 4.2.4):

$$\mathbf{CRB}^{-1}(\theta) = \left[\frac{\partial \mathbf{A}(\theta)}{\partial \theta}\right]^T \tilde{\mathbf{Q}} \frac{\partial \mathbf{A}(\theta)}{\partial \theta} \tag{5.8}$$

Proof in Appendix C. In two dimensions we have:

$$\frac{\partial \mathbf{A}(\theta)}{\partial \theta} = \frac{1}{c} \begin{bmatrix} \frac{x-x_1}{r_1} & \frac{y-y_1}{r_1} \\ \frac{x-x_2}{r_2} & \frac{y-y_2}{r_2} \\ \vdots & \vdots \\ \frac{x-x_n}{r_n} & \frac{y-y_n}{r_n} \end{bmatrix} \tag{5.9}$$

where $x_i$ and $y_i$ are the coordinates of the $i$-th transmitter. Therefore the root mean square of any given estimator is bounded by:

$$\mathbf{RMS}(\theta) \geq \sqrt{\mathrm{Tr}(\mathbf{CRB}(\theta))}$$

This bound is clearly $\theta$ dependent. This fact is important and means that the performance of the estimator depends on the problem geometry. This phenomenon is called Geometric Dilution of Precision (GDoP); it indicates how much the geometry influences the final location error [67].

In simulations, we will take a mean value for the RMS:

$$\mathbf{RMS} \geq \sqrt{\frac{\int_\mathcal{S} \mathrm{Tr}\left[\mathbf{CRB}(\theta)\right] d\theta}{\int_\mathcal{S} d\theta}} \tag{5.10}$$

In the GSM standard, an additional interesting piece of data is available: the timing advance (see Section 2.3.2). It is an approximation of the propagation time between the mobile and the base station it is connected to. Taking this information into account the trilateration is the intersection of hyperbolas with a circle centered on the base station and with a radius equal to the timing advance parameter expressed in meters. For simplicity we will take the base station the mobile is connected to as a reference so that the timing advance can be written as:

$$\mathrm{TA} = |\theta| + b_{\mathrm{TA}}$$

$b_{\mathrm{TA}}$ refers to the perturbation on the timing advance, we will assume it as being Gaussian with zero mean and variance $\sigma_{\mathrm{TA}}$. We will note $\mathbf{CRB}_{\mathrm{TA}}$ the Cramer-Rao bound taking into account this new information. Using the formula: $\frac{\partial |\theta|}{\partial \theta} = \frac{\theta}{|\theta|}$, the new Fisher information matrix is:

$$\mathbf{FIM}_{\mathrm{TA}}(\theta) = \mathbf{FIM}(\theta) + \frac{1}{\sigma_{\mathrm{TA}}^2} \frac{\theta \theta^T}{|\theta|^2} \tag{5.11}$$

The inverse matrix lemma yields:

$$\mathbf{CRB}_{\text{TA}}(\theta) = \mathbf{CRB}(\theta) \left[ \mathbf{I} - \frac{\theta \theta^T \mathbf{CRB}(\theta)}{\sigma_{\text{TA}}^2 |\theta|^2 + \theta^T \mathbf{CRB}(\theta)\theta} \right] \qquad (5.12)$$

It is interesting to note two special cases:

- $\sigma_{\text{TA}}^2 \to \infty$, i.e. no information available on the timing advance, obviously implies

$$\mathbf{CRB}_{\text{TA}}(\theta) \to \mathbf{CRB}(\theta)$$

- $\sigma_{\text{TA}}^2 \to 0$, i.e. exact knowledge of the timing advance, implies

$$\mathbf{CRB}_{\text{TA}}(\theta) \to \mathbf{CRB}(\theta) \left[ \mathbf{I} - \frac{\theta \theta^T \mathbf{CRB}(\theta)}{\theta^T \mathbf{CRB}(\theta)\theta} \right]$$

# 5.3 Algorithms for hyperbolic trilateration

## 5.3.1 Taylor expansion (scoring method)

This is an iterative algorithm for directly minimizing the maximum likelihood criterion (5.7) by performing a Taylor expansion of $\mathbf{A}(\theta)$:

$$\mathbf{A}(\theta_n) \approx \mathbf{A}(\theta_{n-1}) + \frac{\partial \mathbf{A}(\theta_{n-1})}{\partial \theta} [\theta_n - \theta_{n-1}] \qquad (5.13)$$

Replacing this equation in (5.7) gives:

$$\theta_n = \arg\max_\theta \left[ \mathbf{F} - \mathbf{A}(\theta_{n-1}) - \frac{\partial \mathbf{A}(\theta_{n-1})}{\partial \theta}(\theta - \theta_{n-1}) \right]^T \tilde{\mathbf{Q}} \qquad (5.14)$$

$$\left[ \mathbf{F} - \mathbf{A}(\theta_{n-1}) - \frac{\partial \mathbf{A}(\theta_{n-1})}{\partial \theta}(\theta - \theta_{n-1}) \right]$$

And each iteration will be written as follows [68]:

$$\theta_n = \theta_{n-1} + \left[ \left( \frac{\partial \mathbf{A}(\theta_{n-1})}{\partial \theta} \right)^T \tilde{\mathbf{Q}} \frac{\partial \mathbf{A}(\theta_{n-1})}{\partial \theta} \right]^{-1} \left( \frac{\partial \mathbf{A}(\theta_{n-1})}{\partial \theta} \right)^T \tilde{\mathbf{Q}} [\mathbf{F} - \mathbf{A}(\theta_{n-1})]$$

$$(5.15)$$

The algorithm converges to the value that minimizes the log-likelihood whenever a proper initialization is available. This algorithm is also called the scoring method: it is the Newton-Raphson method [69] in which the Hessian has been replaced by its expected value. This expected value is nothing but the Fisher information matrix. Note that this matrix has the same dimension of $\theta$ (two or three) and is therefore very simple to invert. Note also that the base stations must not be arranged linearly so that the Fisher Information Matrix is full column rank.

## 5.3.2   Least squares

This approach gives a non iterative approximate solution. It is not based on the maximum likelihood criterion (5.7). We assume again that the base stations are not arranged linearly. Let us write the following $n - 1$ equations:

$$(r_i - r_1)(r_i + r_1) = r_i^2 - r_1^2 \qquad 2 \le i \le n$$

Note that $r_1$ is the distance between the handset and the base station it is connected to (the nearest base station). Noting $\mathbf{H}$ the following $(n - 1) \times n$ matrix:

$$\mathbf{H} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -1 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{5.16}$$

These equations can be expressed by the following compact formula:

$$[\mathbf{HA}(\theta)] \odot [(\mathbf{H} \odot \mathbf{H})\mathbf{A}(\theta)] = -2\mathbf{HJ}\theta + \mathbf{HK} \tag{5.17}$$

with,

$$\mathbf{J} = \frac{1}{c^2} \begin{bmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix} \qquad \mathbf{K} = \frac{1}{c^2} \begin{bmatrix} x_1^2 + y_1^2 \\ \vdots \\ x_n^2 + y_n^2 \end{bmatrix} \tag{5.18}$$

Note that the elements of $\mathbf{HA}(\theta)$ are nothing but the differences times of arrival between the base station the mobile is connected to and all other base stations. We can rewrite the equation in the following way:

$$[\mathbf{HA}(\theta)] \odot [\mathbf{HA}(\theta)] + [\mathbf{HA}(\theta)] \odot [(\mathbf{H} \odot \mathbf{H} - \mathbf{H})\mathbf{A}(\theta)] - \mathbf{HK} = -2\mathbf{HJ}\theta$$

After simplification, we obtain:

$$[\mathbf{HA}(\theta)] \odot [\mathbf{HA}(\theta)] - \mathbf{HK} = -2\mathbf{HJ}\theta - 2\frac{r_1}{c}\mathbf{HA}(\theta) \tag{5.19}$$

Without loss of generality, the coordinates of the first transmitter can be set at the origin. Thus, $r_1 = |\theta|$ and the equation to be solved is:

$$\frac{1}{2}\mathbf{HK} - \frac{1}{2}[\mathbf{HF}] \odot [\mathbf{HF}] = \mathbf{HJ}\theta + \frac{|\theta|}{c}\mathbf{HF} + \mathbf{B}_1 \tag{5.20}$$

where $\mathbf{B}_1 = -\frac{1}{2}\mathbf{HB} \odot \mathbf{HB} - \mathbf{RHB}$ and $\mathbf{R} = \frac{1}{c}\mathrm{diag}(r_2, \cdots, r_n)$. Supposing that the signal-to-noise ratio is high enough, the term $\mathbf{HB} \odot \mathbf{HB}$ can be neglected. $\mathbf{B}_1$ is then Gaussian zero mean with covariance matrix given by:

$$\mathbf{Q}_1 = \mathbf{RH}^T\mathbf{QHR} \tag{5.21}$$

The problem to be solved is then equivalent at high signal-to-noise ratios to the following non-linear problem, where $\mathbf{M}$ and $\mathbf{V}$ are known in advance:

$$\mathbf{Y} = \mathbf{M}\theta + |\theta|\mathbf{V} + \mathbf{B}_1 \tag{5.22}$$

where,

$$\mathbf{Y} = \frac{1}{2}\mathbf{HK} - \frac{1}{2}[\mathbf{HF}] \odot [\mathbf{HF}]$$

and,

$$\mathbf{M} = \mathbf{HJ} \qquad \mathbf{V} = \frac{1}{c}\mathbf{HF}$$

Two cases are considered here:

- $n$ is equal to the minimum required number of base stations ($n = 3$ if $\theta = [x\ y]^T$), there is no extra information. We can then solve the system explicitly by computing $\theta$ in term of $|\theta|$ from (5.22). Substituting this value in $r_1^2 = |\theta|^2$ gives a quadratic or cubic expression in $r_1$. The solution in the area of interest is selected as the final position.

- $n$ is strictly greater than the minimum required then there are extra information and the set equations (5.22) will not meet at the same point. We have then to estimate the point that minimize the variance error.

From now on, we will suppose that $n$ is strictly greater than the minimum required.

## Simple suboptimal estimation

A sub-optimal solution consists in estimating $\theta$ and $|\theta|$ in equation (5.22) as if they were independent:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{M} & \mathbf{V} \end{bmatrix} \begin{bmatrix} \theta \\ |\theta| \end{bmatrix} + \mathbf{B}_1 \triangleq \mathbf{G}_1\phi_1 + \mathbf{B}_1 \tag{5.23}$$

We have then the following solution:

$$\hat{\phi}_1 = \left[\mathbf{G}_1^T\mathbf{Q}_1^{-1}\mathbf{G}_1\right]^{-1}\mathbf{G}_1^T\mathbf{Q}_1^{-1}\mathbf{Y} \tag{5.24}$$

$\hat{\theta}$ is obtained from $\hat{\phi}_1$ after deleting the last element of $\hat{\phi}_1$ that corresponds to the estimation of $|\theta|$. If the estimation of $|\theta|$ is negative then the handset position is considered to be located at the base station itself. Note that $\mathbf{Q}_1$ is $\theta$ dependent and is then unknown. One might perform two iterations. First, $\mathbf{Q}_1$ is replaced by the identity matrix and then it is replaced by its value using the last estimation of $\theta$.

**Chan's approach**

Once an estimation of $\theta$ and $|\theta|$ is available, one can use the fact that $\theta^T \theta = |\theta|^2$ to enhance the estimation [70]:

$$\hat{\phi}_1 \odot \hat{\phi}_1 = \left[ \begin{array}{c} \mathbf{I} \\ \mathbf{1}^T \end{array} \right] \mathrm{diag}(\theta \odot \theta) + \mathbf{B}_2 \triangleq \mathbf{G}_2 \phi_2 + \mathbf{B}_2 \qquad (5.25)$$

If the location is in two dimensions, this last equation becomes:

$$\hat{\phi}_1 \odot \hat{\phi}_1 = \left[ \begin{array}{c} x^2 \\ y^2 \\ x^2 + y^2 \end{array} \right] + \mathbf{B}_2 = \left[ \begin{array}{cc} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{array} \right] \left[ \begin{array}{cc} x^2 & 0 \\ 0 & y^2 \end{array} \right] + \mathbf{B}_2$$

Neglecting the noise second-order terms, the covariance of $\mathbf{B}_2$ is:

$$\mathbf{Q}_2 = 4 \, \mathrm{diag}(x, y, r_1) \mathbf{Q}_1 \, \mathrm{diag}(x, y, r_1)$$

As $\mathbf{Q}_2$ is $\theta$ dependent, one can use the value obtained in the first step.

The least squares estimation of $\phi_2$ is given by:

$$\hat{\phi}_2 = \left[ \mathbf{G}_2^T \mathbf{Q}_2^{-1} \mathbf{G}_2 \right]^{-1} \mathbf{G}_2^T \mathbf{Q}_2^{-1} \left[ \hat{\phi}_1 \odot \hat{\phi}_1 \right] \qquad (5.26)$$

Once an estimation of $\phi_2 = \theta \odot \theta$ is available, we have to root $\hat{\phi}_2$ to obtain $\theta$. If the location is in two dimensions we will have four possibilities. Amongst all the possibilities we choose the point that is the closest to the initial estimation of $\theta$. If one element is negative then its root does not exist, we choose in this case to force it to be zero. In other words we use the corresponding coordinates of the current base station (i.e. the base station the mobile is connected to). This is the reason why this algorithm performs better than the CRB at poor signal-to-noise ratios.

This algorithm is still suboptimal as some approximations on noise have been done. However, at good signal-to-noise ratios, this algorithm is known to be efficient [70] as it reaches the Cramer-Rao bound (5.10). We will use this algorithm as a good initial estimation for the iterative Taylor algorithm.

## 5.4   Simulations and results

In simulations, a perfect hexagonal network is assumed. The base stations are equally spaced. To avoid border effects all mobiles are generated in the center of the network as shown in Figure 5.1. The distance between a base station and each of its six nearest neighbors is 10 km. Five base stations are involved in
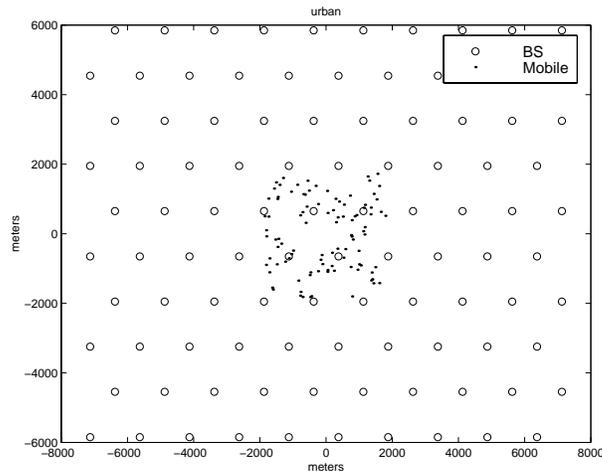
Figure 5.1: Network architecture: 'o' represents a base station and the dots refer to several handset positions.

each trilateration procedure. 200 handsets are generated, and for each one 20 trials are performed which gives a total of 4000 Monte Carlo simulations. The noise resulting for the time of arrival estimation and the synchronization error of the corresponding base station is assumed to be white gaussian with covariance matrix given by:

$$\mathbf{Q} = \sigma_{\mathrm{TOA}}^2 \mathbf{I}$$

Three algorithms are tested:

- The suboptimal least squares algorithm (5.3.2),

- Chan's algorithm (5.3.2),

- Taylor (5.3.1) with initialization given by Chan's algorithm and with two iterations.

Cramer-Rao bounds, with and without the knowledge of the timing advance, have been drawn in Figure 5.2. It is clear that for usual values of $\sigma_{\mathrm{TA}}$, i.e. greater than 500 meters, the gain is quite negligible.

The Cramer-Rao bound without any knowledge of the timing advance is used according to (5.10). The Taylor expansion algorithm seems to be the best one (Figure 5.3).

Note that Chan's approach performs better than the CRB at poor signal-to-noise ratios: indeed this estimator uses the a priori information that the mobile is in the serving cell as discussed ealier.
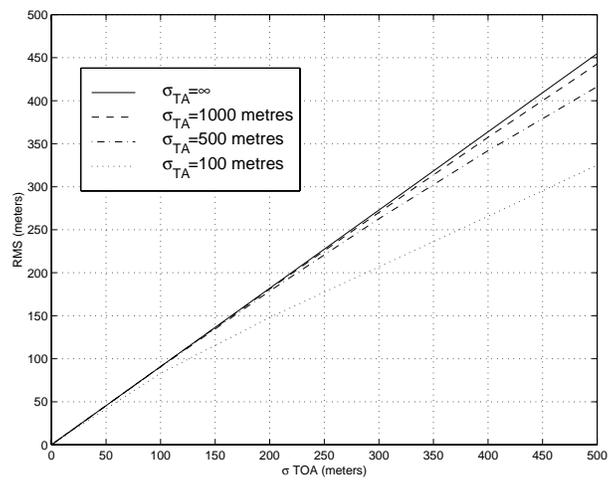
Figure 5.2: CRB with and without the knowledge of the timing advance, five base stations are involved in the trilateration.
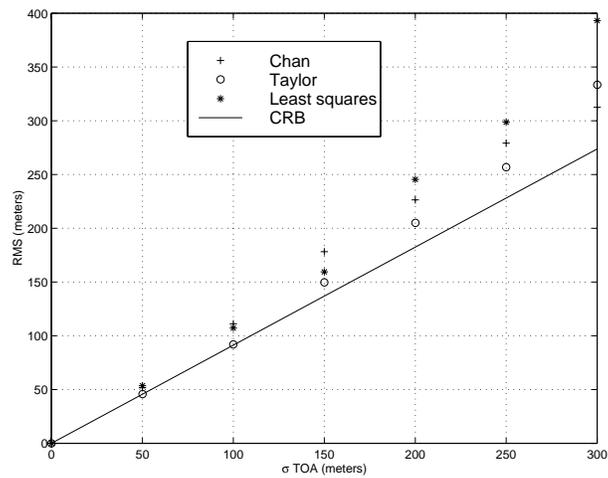


Figure 5.3: Algorithms simulations
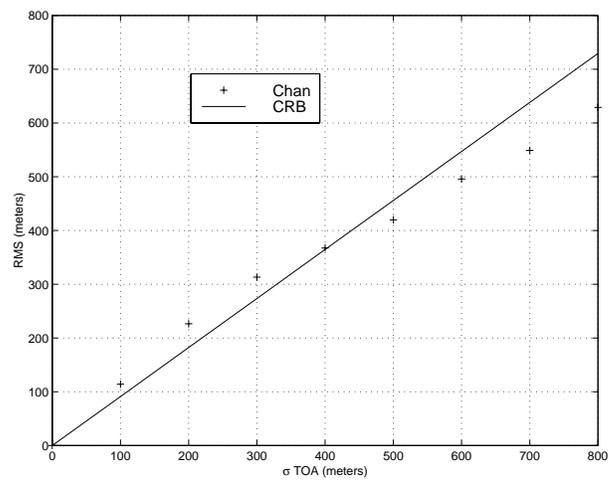
Figure 5.4: Chan's approach vs. the CRB.

# Chapter 6

# On network synchronization

Network synchronization refers in this context to the knowledge of the TDMA frame transmission times or more precisely, the differences of transmission times, also called Real Times Difference (RTD), between all base stations. There are two levels of synchronization:

- Network pseudo-synchronization: in this case the transmission times of the TDMA frames are still uncorrelated and randomly distributed. However, the exact differences of transmission times are known.

- Network absolute synchronization: in this case all the base stations share the same clock. The TDMA frames are sent at the same moment.

As explained in the previous sections, the time of arrival technology for mobile positioning requires only pseudo-synchronization. Absolute synchronization is not really required though it can accelerate the procedure. Absolute synchronization is not a new interest, it has several advantages [71]:

- Acceleration of the handover procedure since synchronous handover replaces the asynchronous one: there is no more need to send an access burst.

- Better equalization: the bursts are completely overlapped[1] by the interfering burst, the channel impulse response estimation will be true then to the whole part of the burst.

- The possibility of defining orthogonal frequency hopping laws among base stations towards a dynamical resources allocation.

- The possibility of performing dynamical training sequences allocation.

---

[1]We are neglecting here the fact that the time of propagation to the mobile is different for two different bursts: however this propagation time is still small compared to the burst period, and the statement here is approximately valid.

71

Note that to get these advantages, a synchronization error below one bit period is enough. However, this is not enough for location purposes since one bit period is equivalent to 1108 meters of propagation. Moreover it is hard to maintain this degree of synchronization since the base stations keep drifting randomly due to the differences of their local oscillatories. For these reasons, we prefer the pseudo-synchronization scenario.

The basic idea to get the synchronization information is in the exploitation of the location information; to get the handset position, synchronization information is required, so if the position is known in advance then it is easy to get the synchronization information. In other words, dedicated mobiles (also called LMU for Location Mobile Unit) can be used successfully at known positions to get the synchronization information. This information can be reported periodically (one report per minute). Assuming that the base stations drift slowly, it is possible to interpolate between two consecutive reports. In fact, there are some requirements on base station drifts; they are not allowed to exceed 50 ns/s.

Of course the use of an external device such as the GPS system to get the synchronization information is possible but it is not the subject of this dissertation since we are concerned with a system where the network locates its subscribers by its own capability.

## 6.1   The pseudo-synchronization problem

As shown in Figure 6.1, the pseudo-synchronization problem may be defined as to estimate the differences in TDMA transmission times according to a set of observations of these differences. If we refer to each real times difference observation by a connection that links the two concerned base stations, we can then define a graph composed from the base stations and these links. The pseudo-synchronization problem is identifiable if and only if the graph is connected, in other words all base stations must be connected with at least one link.

The $m$ real times difference noisy observations are put into one single column vector $\mathbf{Y}$:

$$\mathbf{Y} = \mathbf{MX} + \mathbf{B} \tag{6.1}$$

$\mathbf{X}$ is a vector with dimension equal to the number of observed links in the network $a$. The number of unobserved links is then $\frac{n(n-1)}{2} - a$. $\mathbf{B}$ is a Gaussian noise with zero mean and covariance $\mathbf{Q}$. Note that $\mathbf{M}$ is full column rank. Some RTDs might be measured more than one time. The unknown RTDs are computed by taking into account the constraint that the sum on any cycle in the graph is null
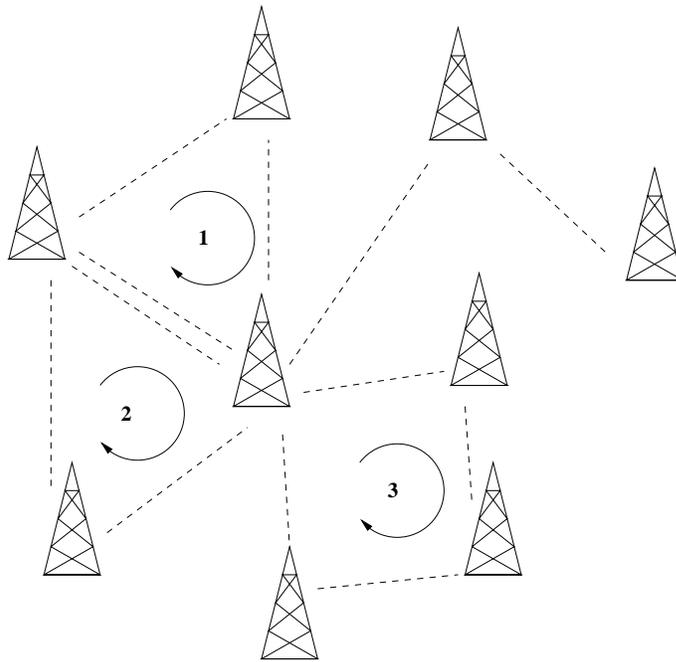
Figure 6.1: The pseudo-synchronization problem. There are three cycles in this network, the sum of the RTDs on each of them is by definition null. All base stations are connected to the network.

by definition. The only condition for identifiability is the network connectivity. This condition requires that $a \geq n - 1$. We then get:

$$(n-1) \leq a \leq \min(\frac{n(n-1)}{2}, m) \qquad (6.2)$$

The constraints are the fact that the sum of the RTDs on any cycle are null: for that purpose, we can retain a base of cycles. The number of cycles in a minimal base of cycles is given by a known theorem. This theorem gives also a method for building such a base (see [72] for example):

**Theorem 1** *The number of cycles of any minimal base in a connected network with $n$ base stations and $a$ links is $N = a - n + 1$.*

The unobserved RTDs can then be computed from the $\frac{n(n-1)}{2} - a$ cycles of the base. Since the network is connected it is possible to estimate all RTDs. The constraint related to a cycle can be written in the following way:

$$\mathbf{C}_i^T \mathbf{X} = 0 \qquad 1 \leq i \leq N$$

where $\mathbf{C}_i$ is a vector that contains 1 or -1 on the elements of the cycle and 0 otherwise.

The system to be solved is a least squares problem with constraints:

$$\begin{cases} \mathbf{Y} = \mathbf{M}\mathbf{X} + \mathbf{B} \\ \mathbf{C}^T\mathbf{X} = 0 \end{cases} \tag{6.3}$$

where $\mathbf{C} = [\mathbf{C}_1 \cdots \mathbf{C}_N]$ is full column rank (independent cycles).

This system can be easily solved using Lagrange multipliers:

$$J(\mathbf{X}) = \frac{1}{2}\left[\mathbf{Y} - \mathbf{M}\mathbf{X}\right]^T \mathbf{Q}^{-1}\left[\mathbf{Y} - \mathbf{M}\mathbf{X}\right] + \mu^T\mathbf{C}^T\mathbf{X} \tag{6.4}$$

Derivating $J$ with respect to $\mathbf{X}$ will give:

$$\frac{\partial J(\mathbf{X})}{\partial \mathbf{X}} = -\mathbf{M}^T\mathbf{Q}^{-1}\left[\mathbf{Y} - \mathbf{M}\mathbf{X}\right] + \mathbf{C}\mu \tag{6.5}$$

We note $\tilde{\mathbf{X}} = \left[\mathbf{M}^T\mathbf{Q}^{-1}\mathbf{M}\right]^{-1}\mathbf{M}^T\mathbf{Q}^{-1}\mathbf{Y}$ the weighted least squares estimation of the RTDs from the observations without any constraint, and $\mathbf{K} = \left[\mathbf{M}^T\mathbf{Q}^{-1}\mathbf{M}\right]^{-1}$ its error covariance.

Setting the derivative with respect to $\mathbf{X}$ to zero, we get the following:

$$\mu = \left(\mathbf{C}^T\mathbf{K}\mathbf{C}\right)^{-1}\mathbf{C}^T\tilde{\mathbf{X}}$$

The final solution can be expressed in the following way:

$$\hat{\mathbf{X}} = \left[\mathbf{I} - \mathbf{K}\mathbf{C}\left(\mathbf{C}^T\mathbf{K}\mathbf{C}\right)^{-1}\mathbf{C}^T\right]\tilde{\mathbf{X}} \tag{6.6}$$

The interpretation is simple. It is a weighted projection matrix. Supposing the noise is white and that each RTD is observed once, i.e. $\mathbf{K} = \sigma^2\mathbf{I}$, equation (6.6) becomes:

$$\hat{\mathbf{X}} = \mathbf{\Pi}_{\mathbf{C}}^{\perp}\tilde{\mathbf{X}} \tag{6.7}$$

Incorporating a constraint is equivalent to project the estimated RTDs (without constraints) onto the null space of $\mathbf{C}$. In the case the cycles do not share any element, i.e. $\mathbf{C}_i^T\mathbf{C}_j = \mathbf{C}_i^T\mathbf{C}_i\delta_{ij}$, equation (6.7) consists in removing the mean RTD in a cycle from each of its element.

Simple algebra manipulations show that:

$$\hat{\mathbf{X}} = \mathbf{X} + \left[\mathbf{I} - \mathbf{K}\mathbf{C}\left(\mathbf{C}^T\mathbf{K}\mathbf{C}\right)^{-1}\mathbf{C}^T\right]\mathbf{K}\mathbf{M}^T\mathbf{Q}^{-1}\mathbf{B} \tag{6.8}$$

This proves that our estimator is unbiased with covariance matrix given by:

$$\mathbf{K}_{\hat{\mathbf{X}}} = \left[ \mathbf{I} - \mathbf{K}\mathbf{C} \left( \mathbf{C}^T \mathbf{K}\mathbf{C} \right)^{-1} \mathbf{C}^T \right] \mathbf{K}$$

Again, if $\mathbf{K} = \sigma^2 \mathbf{I}$, then we simply have $\mathbf{K}_{\hat{\mathbf{X}}} = \sigma^2 \mathbf{\Pi}_{\mathbf{C}}^T$. In other words, the noise variance is reduced by:

$$\frac{\text{Tr}(\mathbf{\Pi}_{\mathbf{C}}^T)}{\text{Tr}(\mathbf{Q})} = \frac{a - N}{a} = \frac{n - 1}{a} \tag{6.9}$$

Taking into account relation (6.2) the gain is somewhere between 1 and $\frac{2}{n}$. If we suppose that $L$ realizations of $\mathbf{Y}$ are available, then we can use the same results substituting $\mathbf{Y}$ with $\frac{1}{L} \sum_{i=1}^{L} \mathbf{Y}_i$. The resulting noise variance will be reduced by $L$.

It is possible to extend the algorithm in presence of drifts. If we suppose that the base stations are drifting slowly with different speeds, we can use a Kalman filter to estimate jointly drift and offset of each base station. $\mathbf{X}_i$ must then be estimated from the following system:

$$\begin{cases} \mathbf{Y}_i = \mathbf{M}\mathbf{X}_i + \mathbf{B}_i \\ \mathbf{X}_i = \mathbf{X}_{i-1} + \mathbf{M}\mathbf{V}\Delta t + \mathbf{W}_i \\ \mathbf{C}^T \mathbf{X}_i = 0 \end{cases} \tag{6.10}$$

where $\mathbf{V}$ is a vector that contains the drift differences, and $\Delta t$ the period between two consecutive measurements.

## 6.2   Absolute synchronization

This approach is different from previous one in the sense that each base station corrects iteratively its time offset until a global synchronized state is reached. Each correction is made by monitoring the neighboring base stations. The algorithm is iterative since the base stations are not allowed to modify their time offset abruptly more than a certain limit. The algorithm presented in the following insures an autonomous synchronization; the time offset corrections are not centralized but each base station correct its time offset independently.

Before presenting the algorithm, we introduce the Perron-Frobenius theorem.

**Definition 1** *A square matrix $\mathbf{A}$ is said to be reducible if and only if the linear operator defined by $\mathbf{A}$ contains an invariant subspace. If such a condition is not verified, the matrix is said to be irreducible.*

A reducible matrix $\mathbf{A}$ can be written via some permutations under the following way:

$$\mathbf{A}' = \begin{pmatrix} \mathbf{B} & 0 \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \tag{6.11}$$

where $\mathbf{B}$ and $\mathbf{D}$ are two square matrices.

The following theorem is a simplified version of the Perron-Frobenius theorem [73, 74]:

**Theorem 2** *If all elements $a_{ij}$ of an irreducible matrix $\mathbf{A}$ are non negative then it has one simple real strictly positive maximal eigenvalue $r$. Noting $s_i = \sum_{k=1}^{n} a_{ik}$ the sum of the elements in each row, then:*

$$\min_{1 \leq i \leq n} s_i \leq r \leq \max_{1 \leq i \leq n} s_i \tag{6.12}$$

*Moreover all eigenvalues lies in the disk $|z| \leq \max_{1 \leq i \leq n} s_i$.*

The bounds shown in this version of the theorem are not the best achievable bounds. Nevertheless, they will be enough for our purposes.

The problem to be solved is to synchronize a set of transmitters $\{A_i\}_{0 \leq i \leq N-1}$. The exact positions of all transmitters are known. We note $t_i$ the time offset of the $i$-th transmitter. We suppose that at the beginning, $t_i$ is a random variable uniformly distributed on $[0, 1]$. For the problem to be identifiable, the network must be connected. We affect to each point in the network a reference time[2].

All transmitters are listening to theirs own neighbors. All transmitters have at least one neighbor since the network is non-oriented and connected. At each iteration, the time offset of the $i$-th transmitter is modified as [75]:

$$t_i^{(n+1)} = t_i^{(n)} + \alpha \frac{\sum_{j=1}^{N} p_{ij}(t_j^{(n)} - t_i^{(n)})}{\sum_{j=1}^{N} p_{ij}} \tag{6.13}$$

$p_{ij}$ is a weighting coefficient, it is a confidence value that evaluates the link between the $i$-th and $j$-th transmitters. Note that $p_{ii} = 0$. $\alpha$ is a constant that fixes the convergence speed of the algorithm, $0 < \alpha \leq \alpha_0 < 1$ ($\alpha_0$ is the drift limitation). We are not considering the noise here. Indeed, since the algorithm requires a high number of successive observations, we assume that the resulting noise variance converges to zero.

---

[2]This assumes that the absolute reference 0 is known and shared by all transmitters. We can avoid such assumption if the network is roughly synchronized. For example, all TDMA frames are synchronized to within a multi-frame.

Noting $\mathbf{X}_n = [t_1 \cdots t_N]^T$ the state of the network at step $n$ and $\mathbf{P}$ the matrix made from $\frac{p_{ij}}{\sum_{j=1}^{N} p_{ij}}$, we have:

$$\mathbf{X}_n = \mathbf{X}_{n-1} + \alpha(\mathbf{P} - \mathbf{I})\mathbf{X}_{n-1} = [(1 - \alpha)\mathbf{I} + \alpha\mathbf{P}]\,\mathbf{X}_{n-1} = \mathbf{M}\mathbf{X}_{n-1} \qquad (6.14)$$

which finally gives,

$$\mathbf{X}_n = \mathbf{M}^n \mathbf{X}_0 \qquad (6.15)$$

Note that $\mathbf{P}$ is not necessarily symmetric since all base stations have not the same number of neighbors.

**Theorem 3** *The algorithm described by equation (6.15) always converges in probability to the synchronized state, in other words:*

$$\lim_{n \to \infty} \mathbf{M}^n v = \beta \mathbf{1}$$

*where $v$ is a randomly initialized vector and $\beta$ a scale factor.*

**Proof 1** *It is straightforward to see that $\mathbf{P}$ is an irreducible matrix since the network is assumed to be connected. And since all its element are positive, the Perron-Frobenius theorem states that the maximal eigenvalue of $\mathbf{P}$ is unique and real. Since the sums of the elements of all rows in $\mathbf{P}$ are all equal to one, it follows that this maximal eigenvalue is one. The eigenvector associated to this eigenvalue is $\frac{1}{\sqrt{n}}\mathbf{1}$. All other eigenvalues have a module that is less or equal to 1. The eigenvalues of $\mathbf{M}$ follows from those of $\mathbf{P}$. All these eigenvalues except the maximum one that is equal to one are inside the disk $|z| < 1$. The algorithm converges then in probability to the synchronized state whatever the initial state is.*

The convergence speed of this algorithm is proportional to the ratio between the two consecutive strongest eigenvalues, $\frac{1}{\lambda(\alpha)}$, where:

$$\lambda(\alpha) = \max_{1 \leq i \leq n} \left( |1 - \alpha + \alpha\mu_i| < 1 \right)$$

$\mu_i$ refers to the eigenvalues of $\mathbf{P}$. $\alpha$ must be chosen properly to minimize $\lambda(\alpha)$:

$$\hat{\alpha} = \min \left( \arg \min_{\alpha} \lambda(\alpha), \alpha_0 \right)$$

$\alpha_0$ is the drift limitation previously defined and it is most likely that it is the fundamental limitation since its value is very low (50 ns/s).

In Figure 6.2, simulations are done in a network consisting of fifty base stations. We suppose that each base station is connected to the same number of
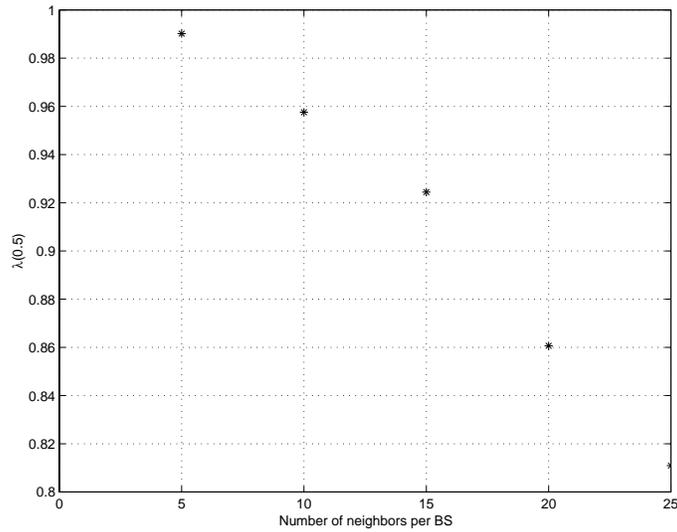
Figure 6.2: $\lambda(\alpha)$ variations according to the number of neighbors per base station, $\alpha = 0.5$.

neighbors. All connected links are equally weighted. The simulation shows that $\lambda(\alpha)$ decreases when the number of neighbors increases allowing faster convergence of the algorithm ($\alpha = 0.5$).

It should be noted that the convergence speed could be enhanced using few master timing references perfectly synchronized among the base stations as shown in [75]. However, those masters should take their reference from an external device such as GPS.

# Chapter 7

# Global simulations and final results

In previous sections, simulations were performed independently on time of arrival estimation and on the trilateration procedure. In this chapter, we combine both procedures to perform the complete simulation process of a mobile localization.

The performance highly depends on the profile of the environment the mobile is located in. This profile determines the parameters related to the channel impulse response such as the number of paths, the delay spread, or the existence of the line of sight path. It also determines the path loss that gives the signal-to-interference ratio.

## 7.1 Channel impulse response model

In this section, we investigate the way of generating a channel impulse response represented by a delay spread, a number of paths, and for each one, its time delay, its mean power, and its amplitude probability distribution.

It is common to say that each path in a channel impulse response is the result of the superposition of a large number of partial waves, each with a delay time that corresponds to the delay of this path [76,77]. The amplitude of a given path is in consequence that of the resulting sum of these partial waves:

$$E(t) = \sum_{i=1}^{n} E_i \exp\left[ j(\phi_0 + \frac{2\pi}{\lambda} vt \cos \alpha_i) \right] \tag{7.1}$$

$\alpha_i$ refers to the angle of incidence of the $i$-th partial wave and $\phi_0$ the initial phase. $v$ refers to the mobile velocity.

In the case the path does not correspond to the direct line of sight path, these partial waves are uncorrelated and have a phase uniformly distributed. The resulting amplitude is Gaussian complex with zero mean, its envelope follows a Rayleigh distribution. If a strong constant amplitude is present as in the case of the line of sight path, the resulting amplitude has a non-zero mean, and the distribution of the envelope is more likely to be Ricean.

The amplitude time correlation function is shown to be (see [77] for example):

$$R(\tau) = \int_0^{2\pi} p(\alpha) \exp\left[j\frac{2\pi}{\lambda}v\tau\cos\alpha\right]d\alpha \tag{7.2}$$

$p(\alpha)$ refers to the angular distribution of the partial waves. Usually $p(\alpha)$ is assumed to be uniformly distributed; this assumption leads to the well-known Jakes' model:

$$R(\tau) = J_0(\frac{2\pi}{\lambda}v\tau) \quad \Leftrightarrow \quad \bar{R}(f) = \frac{1}{\sqrt{1 - \left(\frac{f}{f_d}\right)^2}} \tag{7.3}$$

where $\bar{R}(f)$ refers to the Doppler spectrum and $f_d = \frac{v}{\lambda}$ refers to the Doppler frequency. In our simulations, we will assume that $p(\alpha)$ is Gaussian with a constant mean value generated randomly and a standard deviation of 0.15 radians. This model corresponds to the case where the partial waves are concentrated in a narrow beam. In general, the higher the velocity is, the less correlated the fading is.

The fading simulation is done by generating the partial waves randomly according to (7.1). If the path is assumed to be purely diffused, i.e. no deterministic components present, then all amplitudes are randomly generated and the resulting amplitude follows a Rayleigh distribution. In contrast, if a strong main component is present, then the fading is more likely to be Ricean; we therefore need to select one partial wave among the $n$ to have a stronger amplitude. It is common to use the Nakagami distribution [78] to model the fading behaviour:

$$p(r) = \frac{2m^m r^{2m-1}}{\Gamma(m)\Omega^m} \exp(-\frac{m}{\Omega}r^2) \tag{7.4}$$

The constant $m$ is related to the importance of the main component: the larger is $m$, the more important is this component. $\Omega$ refers to the path energy. We have three possibilities (see Figure 7.1):

- $m = 1$: we simply have the Rayleigh distribution.

- $m > 1$: this is an approximation of the Ricean distribution though it is not exactly that Ricean distribution (see [79]).

- $m = \infty$: there is one lonely deterministic strong component, the distribution is a Dirac function.
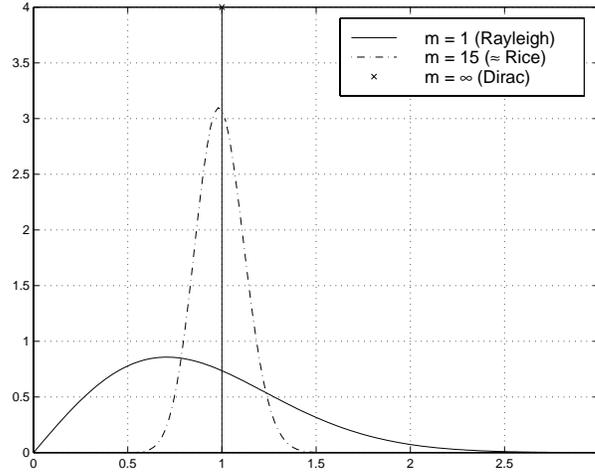


Figure 7.1: Nakagami Distribution

The partial waves generation is done using the two following relations [77]:

$$A_1 = \sqrt{\Omega\sqrt{1 - \frac{1}{m}}}$$

$$\sigma^2 = \frac{\Omega}{n}(1 - \sqrt{1 - \frac{1}{m}})$$

where $A_1$ is the amplitude of the main component, and $\sigma^2$ the power of other partial waves.

## 7.2  Environment profiles

Several profiles are chosen for simulations. They represent a large set of typical environments. The selected profiles are:

- Urban (indoor at 3km/h and outdoor at 3 and 50 km/h): populated areas, multi-story buildings, and city centers.

- Suburban outdoor (at 3 and 50 km/h): residential houses, suburbs, and villages.

- Rural outdoor (at 3 and 100 km/h): inhabited areas, highways, fields, and forests.

The main difference between urban and other profiles is in the presence of line of sight path.  No line of sight is assumed in the urban whereas a strong line of sight path is present in the rural and suburban profiles.  Moreover, the number of scatterers is higher in the urban profile and the delay spread is also longer.  These facts lead unavoidably to poorer results in the urban case.  The assumptions made on each of these profiles are based on the common channel model that had been established in T1P1.5 meetings [80] for each company to simulate its own location algorithm.

## 7.3    Time of Arrival simulations

### 7.3.1    Assumptions

We are concerned in this section with the channel impulse response model assumptions.

**Delay-Spread model**

We take the Greenstein model previously discussed (2.7) and set the three parameters of this model $\sigma_y$, $T_1$, and $\epsilon$ according to each profile.

**Fading model**

We use the Nakagami fading model according to the description made in (7.1). The distribution of the angles of incidence of the partial waves is supposed to be Gaussian with an initial phase randomly generated and a standard deviation of 0.15 radians.

The power generation is exponential for the urban profile; the power of each delay is weighted with $\exp(-6\tau_i)$ where $\tau_i$ is generated uniformly on $[0, 1]$.

Once the powers and delays are generated, powers are normalized so that their sum is equal to one and the delays are adjusted so that the desired delay spread (generated according to the Greenstein model) is met:

$$a_i \Leftarrow \frac{a_i}{\sum_{i=1}^{d} a_i} \qquad \tau_i \Leftarrow \tau_i \frac{\xi_{\text{new}}}{\xi_{\text{old}}}$$

| Parameters | Urban | Suburban | Rural |
|---|---|---|---|
| $T_1$ ($\mu$s) | 0.4 | 0.3 | 0.1 |
| $\epsilon$ | 0.5 | 0.3 | 0.3 |
| $\sigma_y$ | 4 | 4 | 4 |
| Mobile speed (km/h) | 3/50 | 3/50 | 3/100 |
| LOS presence | no | yes | yes |
| Number of delays | 20 | 6 | 6 |
| Delays generation | $\{\tau_i\}_{1 \leq i \leq 20} \sim \mathcal{U}[0,1]$ | $\tau_1 = 0$ $\{\tau_i\}_{2 \leq i \leq 6} \sim \mathcal{U}[0,1]$ | $\tau_1 = 0$ $\{\tau_i\}_{2 \leq i \leq 6} \sim \mathcal{U}[0,1]$ |
| Power generation | $a_i \sim \mathcal{U}[0.5, 1.5]$. $\exp(-6\tau_i)$ | $a_1 = 4.3$ $\{a_i\}_{2 \leq i \leq 6} \sim \mathcal{U}[0.5, 1.5]$ | $a_1 = 4.3$ $\{a_i\}_{2 \leq i \leq 6} \sim \mathcal{U}[0.5, 1.5]$ |
| Nakagami $m$ parameter | $\{m_i\}_{1 \leq i \leq 20} = 1$ | $m_1 = 15$ $\{m_i\}_{2 \leq i \leq 6} \sim \mathcal{U}[1,5]$ | $m_1 = 15$ $\{m_i\}_{2 \leq i \leq 6} \sim \mathcal{U}[1,5]$ |
| Number of partial waves per delay | 100 | 100 | 100 |
| Initial phase $\phi_0$ | $\mathcal{U}[0, 2\pi]$ | $\mathcal{U}[0, 2\pi]$ | $\mathcal{U}[0, 2\pi]$ |
| $\alpha_j$ | $\mathcal{N}(0, 0.15)$ | $\mathcal{N}(0, 0.15)$ | $\mathcal{N}(0, 0.15)$ |

Table 7.1: Time of arrival simulations assumptions according to each profile.

All parameters are shown in Table 7.1. In Figure 7.2 and 7.3 examples of channel generation are shown for the urban and the rural profiles, the delay 0 refers to the line of sight path.

## 7.3.2   Simulations

The difficulty encountered is the determination of the signal subspace dimension. Theoretically, it is equal to the number of paths under the assumption that this number is lower than the channel impulse response length. This condition is not met for the urban environment. However, some paths have a relative low power and can be neglected, others are so close to each other that they could be merged into one single path. For these reasons, the dimension of the signal subspace is lower than the number of paths in practice, and few eigenvectors from the channel covariance matrix are selected as belonging to the signal subspace. We call this subspace the quasi signal subspace and the range space of other eigenvectors the quasi noise subspace.

In general, the quasi signal subspace dimension rises when the number of paths, the speed, or the delay spread increase. Since the delay spread is statistically related to the distance between the handset and the base station according to the Greenstein model, the dimension rises when the distance increases too.
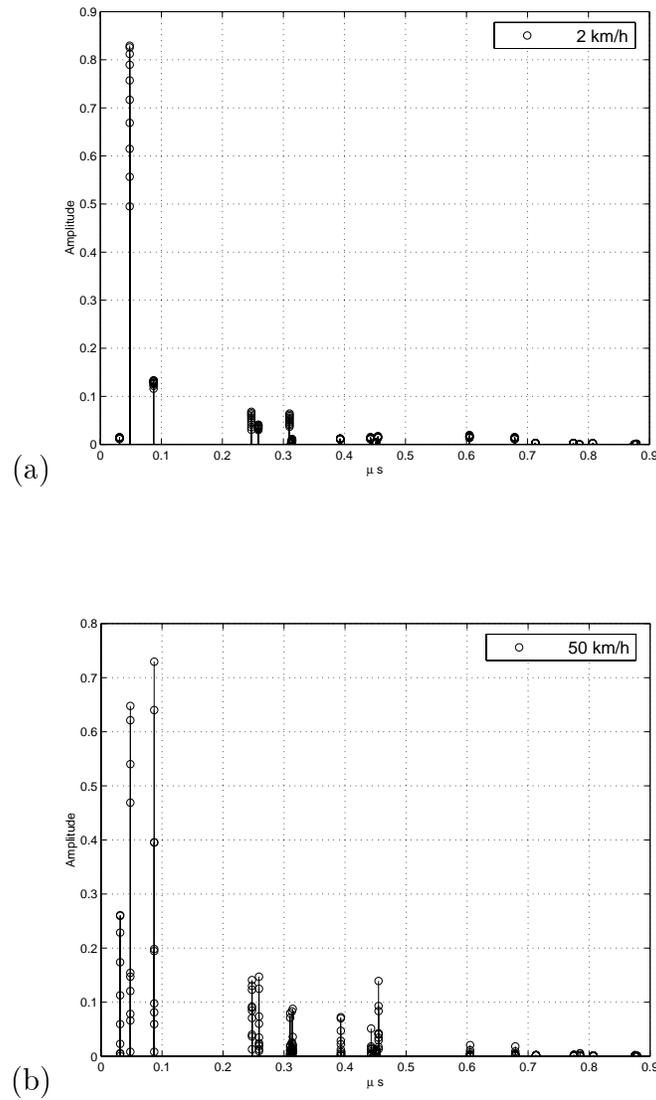
Figure 7.2: Urban profile example, 20 consecutive bursts are generated: (a) at 2 km/h, (b) at 50 km/h. At high speed, the fading is less correlated in time. There is no line of sight path and a bias in the ToA estimation is unavoidable.

The impact of the distance and the speed on the eigenvalues is depicted in Figures 7.4 and 7.5. The mean values of the strongest three eigenvalues are drawn. It is shown that in most cases the first eigenvalue concentrates more than 90% of the channel impulse response energy.

(a)



(b)
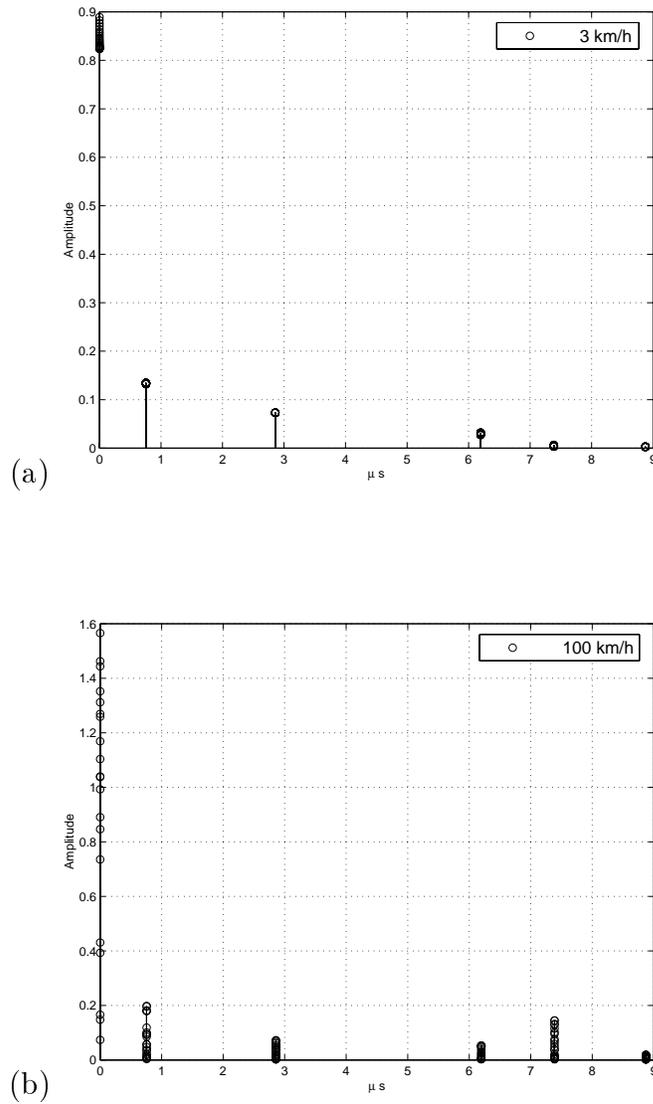
Figure 7.3: Rural profile example, 20 consecutive bursts are generated: (a) at 3 km/h, (b) at 100 km/h. A strong line of sight path is always present.

The number of these eigenvalues represents the dimension of the quasi signal subspace. We choose to take the Minimum Description Length criterion (MDL) to estimate the dimension of the quasi subspace under the constraint that this dimension should not exceed a given dimension $d_{\max}$:
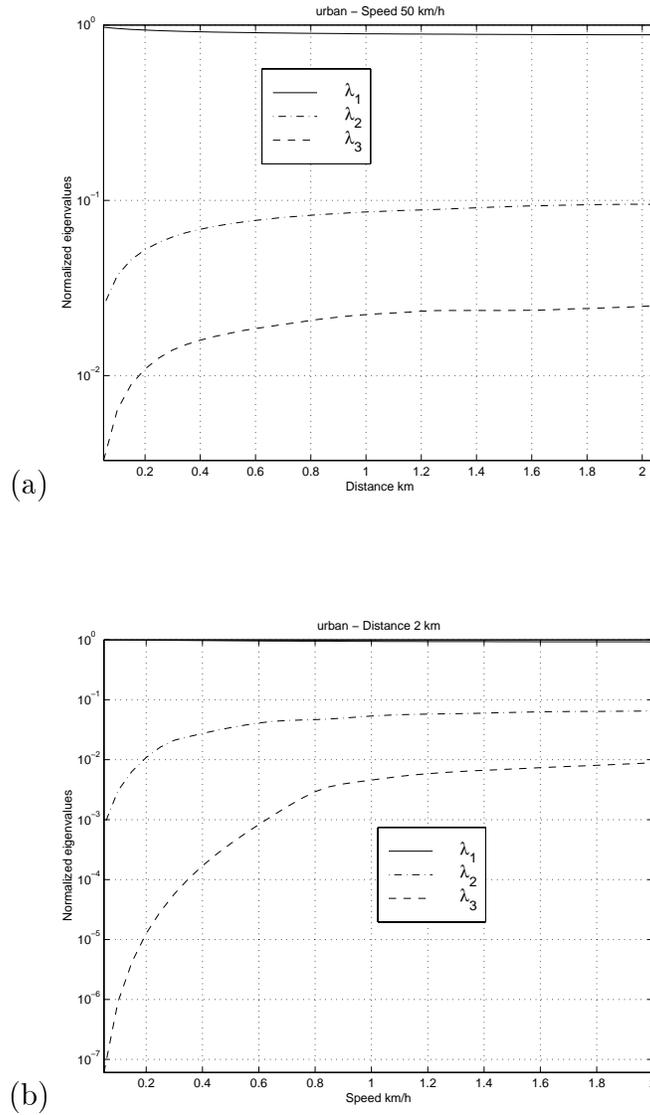
Figure 7.4: Eigenvalues behaviour in urban environment according to: (a) the distance, (b) mobile speed.

$$\hat{d} = \min_{j}(\hat{d}_{\mathrm{MDL}}, d_{\max}) \qquad (7.5)$$

The constraint that the dimension should not exceed $d_{\max}$ comes from the fact that the MDL criterion may overestimate the quasi signal subspace dimension.
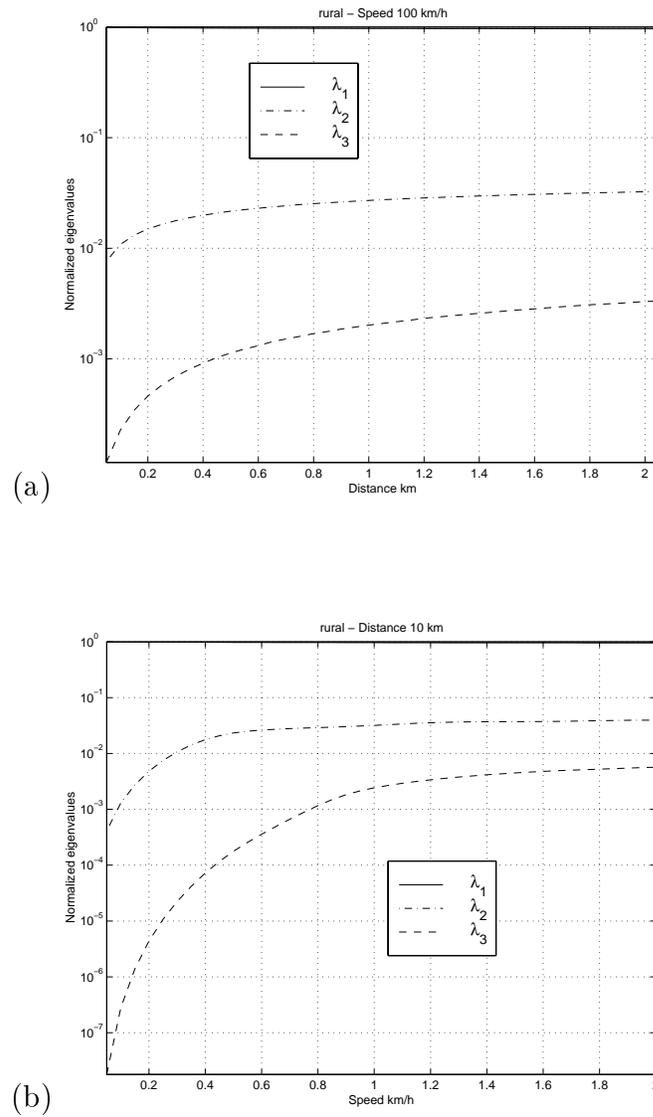
Figure 7.5: Eigenvalues behaviour in rural environment according to: (a) the distance, (b) mobile speed.

In practice, we choose $d_{\max} = 4$.

$\hat{d}_{\mathrm{MDL}}$ is given by [17]:

$$\hat{d}_{\mathrm{MDL}} = \arg \min_{0 \le k \le p-1} L \log \left[ \frac{\left( \frac{1}{p-k} \sum_{i=k+1}^{p} \lambda_i \right)^{p-k}}{\prod_{i=k+1}^{p} \lambda_i} \right] + \frac{1}{2} k(2p - k) \log L$$

$\{\lambda_i\}_{1 \le i \le p}$ are the set of eigenvalues of the channel impulse response covariance matrix, $p$ is the channel impulse response length, and $L$ is the number of observed bursts.

Since the delays are assumed to be independent (uncorrelated), a good candidate algorithm for time of arrival estimation is Root MUSIC described in 4.3.2. We take into account the relevant roots only, i.e. the roots that are closest to the unit circle. The time of arrival is defined as the minimum time delay that is obtained from the root with the lowest argument.

In Table 7.2, results of simulations are presented for all environments described in last section at 10 dB. The RMS error in meters is computed for the best 90% results in order to eliminate the results that are too bad: these results can be corrected by taking the timing advance information (i.e. the position of the training sequence). Simulations clearly show that Root MUSIC is better than the simple poor matched filter in all cases (see Table 7.2). Moreover histograms of results are drawn for the rural and urban profiles at 3 km/h showing a better distribution while using Root MUSIC (see Figure 7.6).

Figures 7.7, 7.8, and 7.9 shows that the ToA estimation accuracy is improved when:

- the speed increases due to better uncorrelation of successive observations,

- the number of observations increases due to a better knowledge of the signal subspace,

- the distance MS-BS decreases due to a lower delay spread,

- the maximum signal subspace dimension $d_{\max}$ (see (7.5)) increases due to a better characterization of the signal subspace. However as discussed earlier, this value should not be very high.

The results show as expected that the ToA estimation problem depends slightly on the signal-to-noise ratio value. The main issue is nothing but multipath. Even in a no noise environment (SNR $= \infty$) the error is not negligible for urban cases ($> 100$ m).

(a)

(b)
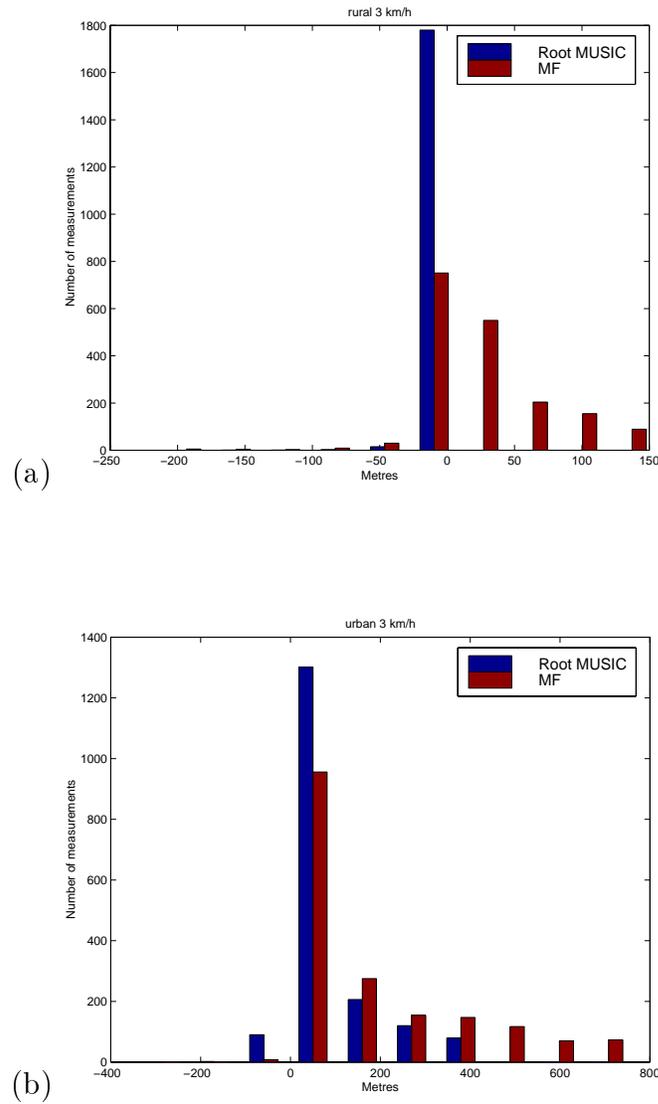
Figure 7.6: ToA estimation error histogram at 3 km/h and 10 dB using 20 synch. bursts for: (a) rural environment at 10 km, (b) urban environment at 1 km.

## 7.4    Location simulations

In this section, we combine the times of arrival estimation with the trilateration procedure to get the final location.

| Profiles | Root MUSIC | | | Matched filter | | |
|---|---|---|---|---|---|---|
| | 0 dB | 10 dB | $\infty$ dB | 0 dB | 10 dB | $\infty$ dB |
| Urban 3 km/h | 311 | 231 | 210 | 323 | 323 | 323 |
| Urban 50 km/h | 156 | 135 | 134 | 294 | 294 | 294 |
| Suburban 3 km/h | 26 | 20 | 19 | 76 | 74 | 74 |
| Suburban 50 km/h | 10 | 6 | 4 | 47 | 47 | 47 |
| Rural 3 km/h | 18 | 16 | 12 | 60 | 58 | 58 |
| Rural 100 km/h | 5 | 3 | 3 | 44 | 44 | 44 |

Table 7.2: RMS error (in meters) for the best 90% results of ToA simulation results at 10 dB and using 20 synch. bursts.

## 7.4.1   Assumptions

**Network model**

We take a perfect hexagonal network with omnidirectional base stations. The distance between two base stations varies according to the environment. This distance is equal to $D = R\sqrt{3}$ where $R$ is the cell radius.

We choose a reuse pattern with $K = 7$. The minimal distance between two base stations sharing the same set of frequencies is [81]: $L = D\sqrt{K} = D\sqrt{7}$, see Figure 7.10.

**Path-loss model**

We will use the Okumura-Hata formula previously discussed (2.1). The parameters $D$, $L_p$, and the fading standard deviation are set according to each environment (see Table 7.3). All mobiles are generated randomly inside the network so that border effects are neglected, see Figure 7.11.

The uplink and downlink budget are assumed to be equivalent so that one can use either the uplink or the downlink budget. We choose to use the uplink budget. The noise variance is related to the base station sensibility, its value is set to -118 dBm.

All simulation parameters are shown in Table 7.3. The C/I CDF are shown in Figure 7.12 for the best four base stations.

## 7.4.2   Results of simulations

For simplicity the interferences are approximated as AWGN. This assumption leads to optimistic results. The difference in performance is however negligible.
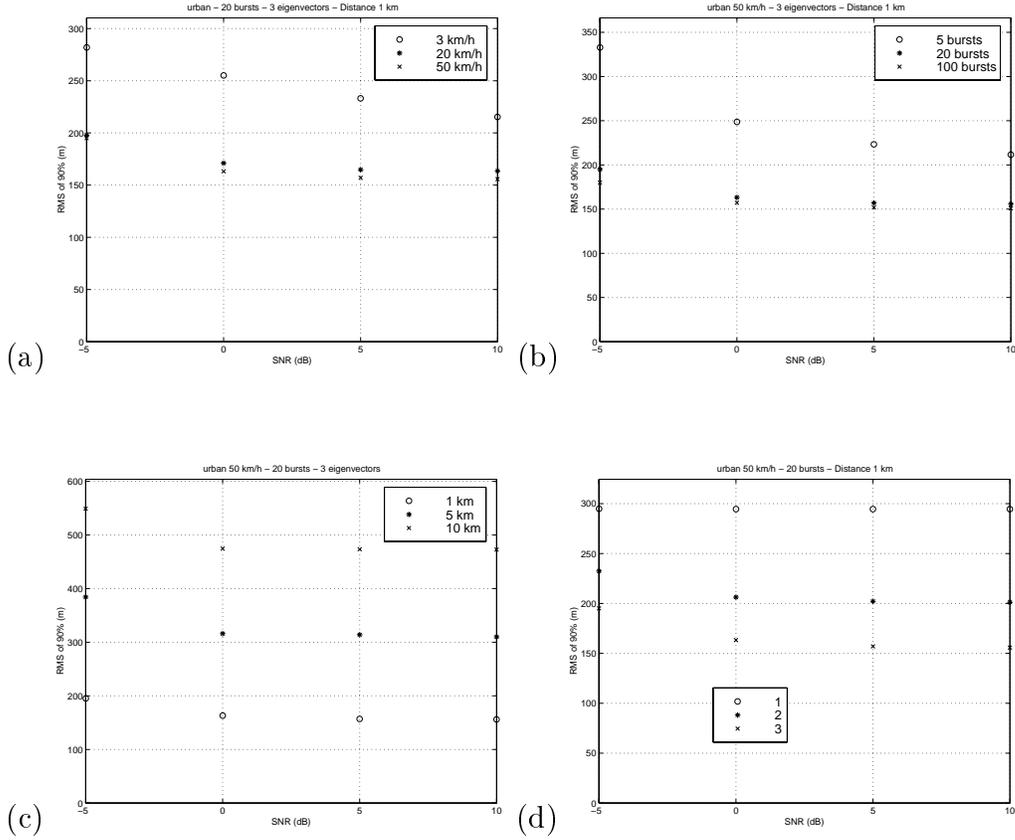
Figure 7.7: ToA estimation in the urban environment: (a) for different mobile speeds, (b) for different number of bursts, (c) for different distances MS-BS, (d) for different maximum signal subspace dimensions.

| Parameters | Urban | Urban indoor | Suburban | Rural |
|---|---|---|---|---|
| $D$ (meters) | 1500 | 1500 | 4500 | 10000 |
| $L_p$ (dB) | 126 | 139.5 | 116 | 98 |
| Fading std. (dB) | 6 | 8.5 | 6 | 6 |
| Mobile speed (km/h) | 3/50 | 3 | 3/50 | 3/100 |
| Sensibility | -118 dBm | -118 dBm | -118 dBm | -118 dBm |

Table 7.3: Location simulations assumptions according to each profile.

We did not consider the mobile speed is the position, in other words the mobile is assumed to be fixed. The speed parameter is only taken into account in the
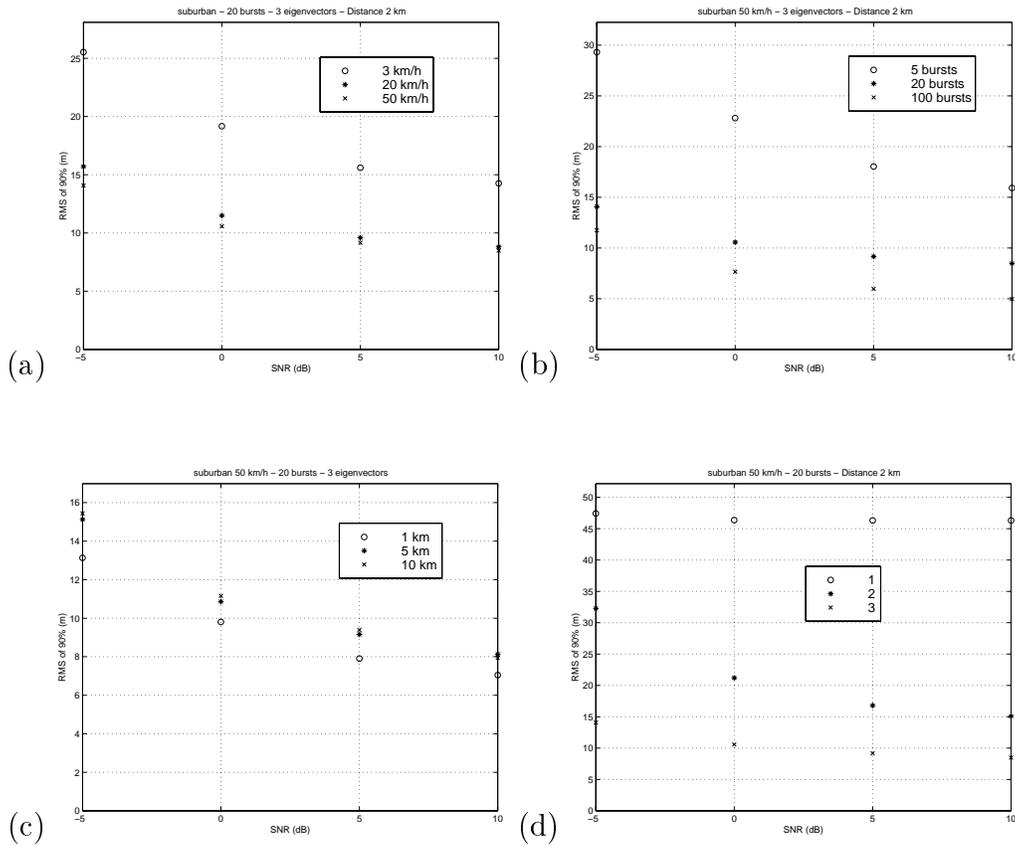
Figure 7.8: ToA estimation in the suburban environment: (a) for different mobile speeds, (b) for different number of bursts, (c) for different distances MS-BS, (d) for different maximum signal subspace dimensions.

correlation properties of successive bursts; at 50 km/h in one second, the mobile can move only 14 meters, this effect is therefore not considered. 20 synchronization bursts are used per link. Three, four, and five base stations are involved in the simulations.

The selected algorithms are:

- Root MUSIC for the time of arrival estimation as discussed in Section 7.3.2.

- The Taylor algorithm for the trilateration procedure with two iterations and initialized with the Chan's algorithm (see Chapter 5). It should be mentioned here that no weighting values are associated to each time of arrival involved in the trilateration. In other words, the error covariance
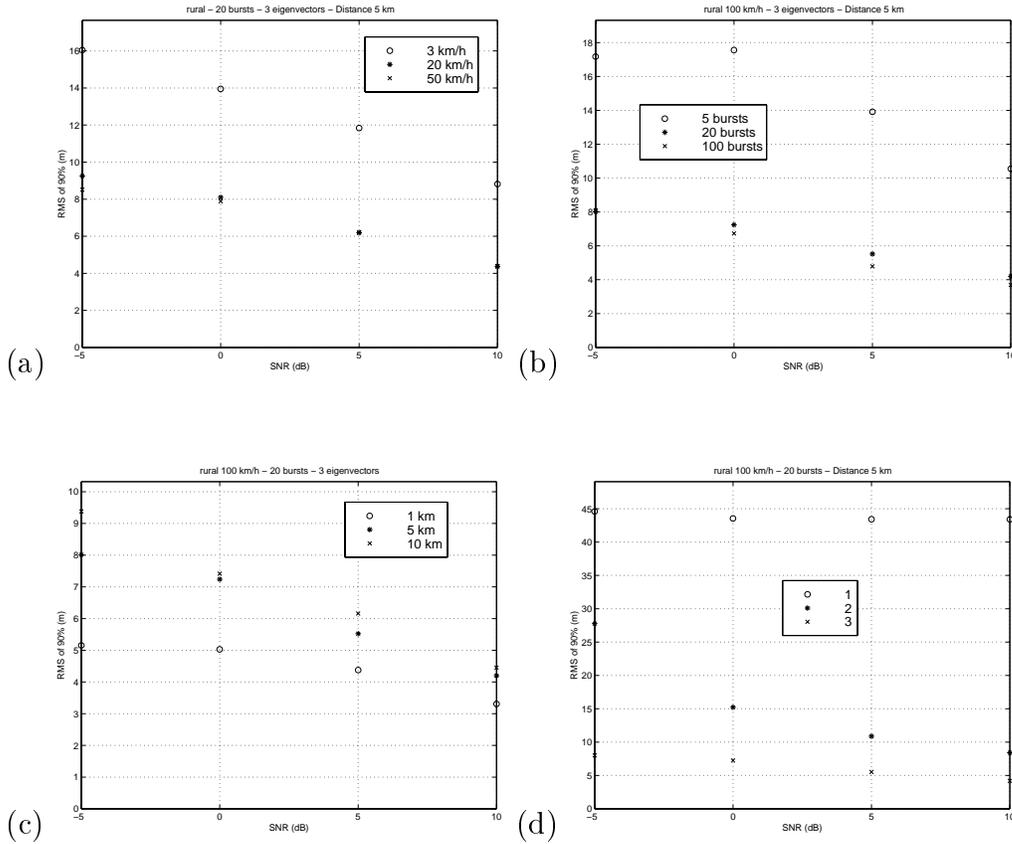
Figure 7.9: ToA estimation in the rural environment: (a) for different mobile speeds, (b) for different number of bursts, (c) for different distances MS-BS, (d) for different maximum signal subspace dimensions.

matrix of the estimated times of arrival $\mathbf{Q}_f$ is proportional to the identity matrix. This leads to a loss in the performances; affecting weighting values to each time of arrival is not an easy task. Indeed, it is not correct to choose this value according to the C/(I+N) ratio, since the time of arrival accuracy depends more on the channel impulse response structure than on the C/(I+N) ratio as discussed in Section 7.3.2.

In the following simulations, 500 handset positions were generated and for each one 20 independent trials were run. That gives a total of 10000 Monte Carlo simulations that is high enough in order to obtain reliable results. Two sets of simulations are done, the first assumes a perfect network synchronization and the other assumes synchronization errors. The numerical results are shown
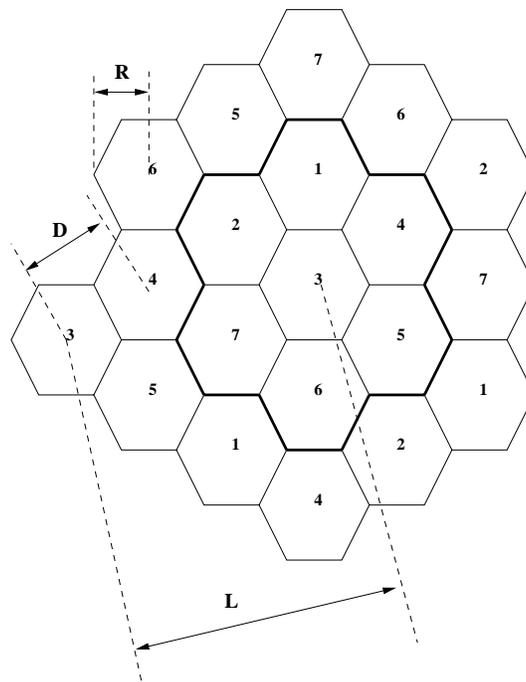
Figure 7.10: Network model with a reuse pattern of 7 cells.
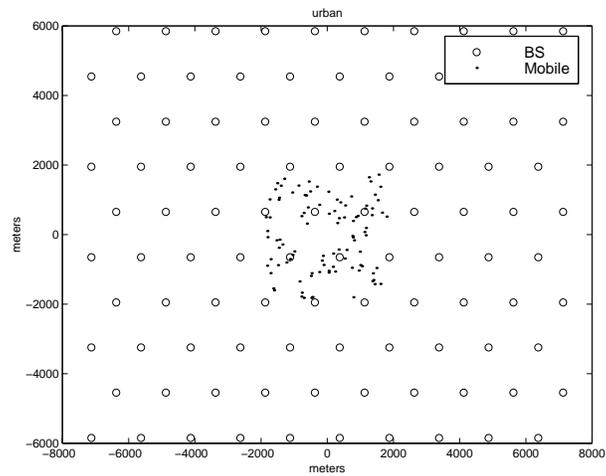


Figure 7.11: Network model, urban profile, all mobiles are generated in the middle of the network to avoid border effects.

in terms of:

- maximum error in meters of the best 67% of results,
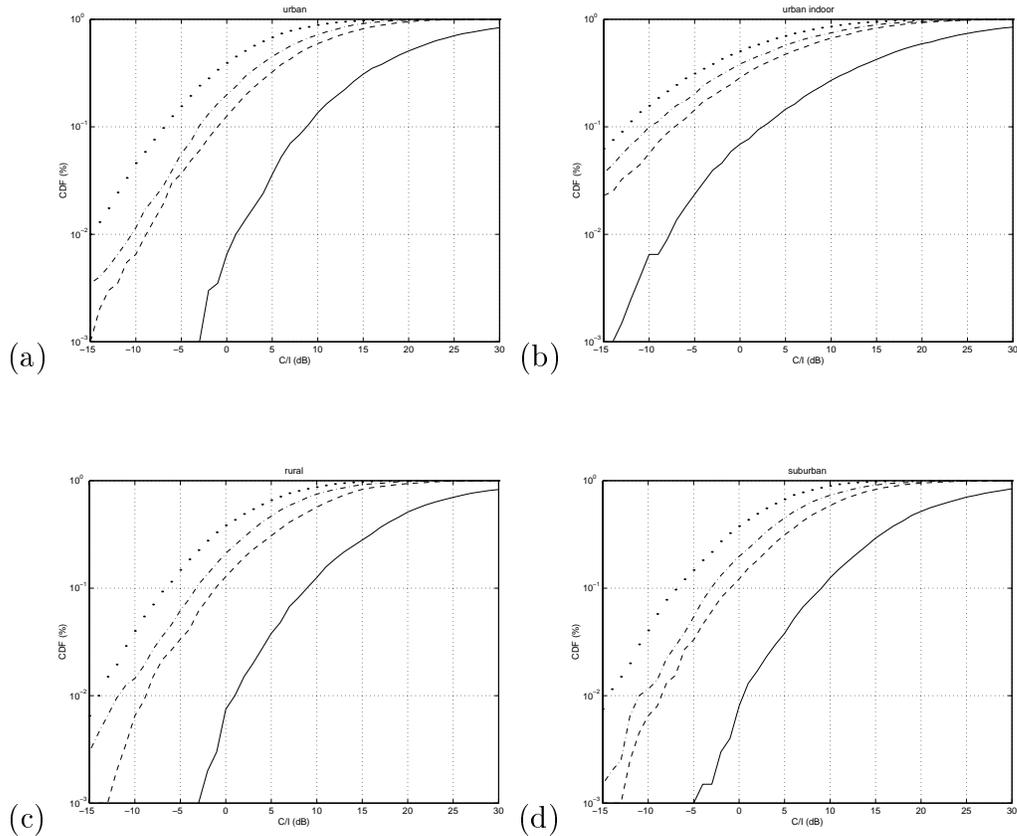
Figure 7.12: C/I CDF for the BCCH channel for the serving cell and the nearest three base stations: urban profiles are interference limited while rural and suburban profiles are noise limited.

- percentage of results with an error lower than 125 meters,

- root mean square error of the best 90 % of results.

**Perfect knowledge of the RTDs**

Perfect synchronization is assumed in these simulations. The results are depicted in Table 7.4. In general, an improvement is observed when the number of base stations involved in the procedure rises. There is an anomaly in the urban case; the results using five base stations (shown in bold) are slightly worse than with four. The reason for that is related to the wrong weighting of the times of arrival before data fusion. The improvement is very significant with four base stations compared to three. Indeed, with three base stations there is no extra information

since three is the minimum number of base stations required.

| Profiles | 67 % in meters | | | Percentage < 125 m | | | RMS of 90% | | |
|---|---|---|---|---|---|---|---|---|---|
| | 3 BS | 4 BS | 5 BS | 3 BS | 4 BS | 5 BS | 3 BS | 4 BS | 5 BS |
| Urban out. 3 km/h | 104 | 93 | 93 | 72 | 76 | 78 | 115 | 84 | 80 |
| Urban out. 50 km/h | 95 | 83 | **84** | 74 | 79 | 82 | 104 | 76 | 72 |
| Urban in. 3 km/h | 108 | 94 | **97** | 71 | 75 | 77 | 119 | 86 | 82 |
| Suburban 3 km/h | 17 | 12 | 12 | 91 | 98 | 98 | 20 | 12 | 12 |
| Suburban 50 km/h | 12 | 8 | 8 | 93 | 100 | 100 | 12 | 7 | 7 |
| Rural 3 km/h | 6 | 5 | 5 | 90 | 100 | 100 | 10 | 4 | 4 |
| Rural 100 km/h | 4 | 3 | 3 | 91 | 100 | 100 | 5 | 3 | 3 |

Table 7.4: Location simulations results with three, four, and five base stations. Network perfectly synchronized.

## Synchronization errors

The relative offset of each base station is assumed to be corrupted with a Gaussian noise with zero mean and a standard variation of 100 ns which corresponds to approximately 30 meters. Results are shown in Table 7.5 and shown in term of CDF curves for all specified environments in Figures 7.13, 7.14, 7.16, and 7.15.

| Profiles | 67 % in meters | | | Percentage < 125 m | | | RMS of 90% | | |
|---|---|---|---|---|---|---|---|---|---|
| | 3 BS | 4 BS | 5 BS | 3 BS | 4 BS | 5 BS | 3 BS | 4 BS | 5 BS |
| Urban out. 3 km/h | 116 | 99 | 99 | 69 | 75 | 76 | 124 | 90 | 85 |
| Urban out. 50 km/h | 106 | 91 | 89 | 72 | 78 | 80 | 111 | 81 | 77 |
| Urban in. 3 km/h | 120 | 103 | 100 | 69 | 74 | 76 | 130 | 92 | 86 |
| Suburban 3 km/h | 49 | 38 | 34 | 91 | 98 | 98 | 42 | 31 | 28 |
| Suburban 50 km/h | 46 | 35 | 31 | 93 | 100 | 100 | 39 | 28 | 25 |
| Rural 3 km/h | 47 | 34 | 30 | 90 | 100 | 100 | 42 | 28 | 25 |
| Rural 100 km/h | 45 | 33 | 29 | 90 | 100 | 100 | 39 | 27 | 24 |

Table 7.5: Location simulations results with three, four, and five base stations. Synchronization errors are included.

(a)



(b)
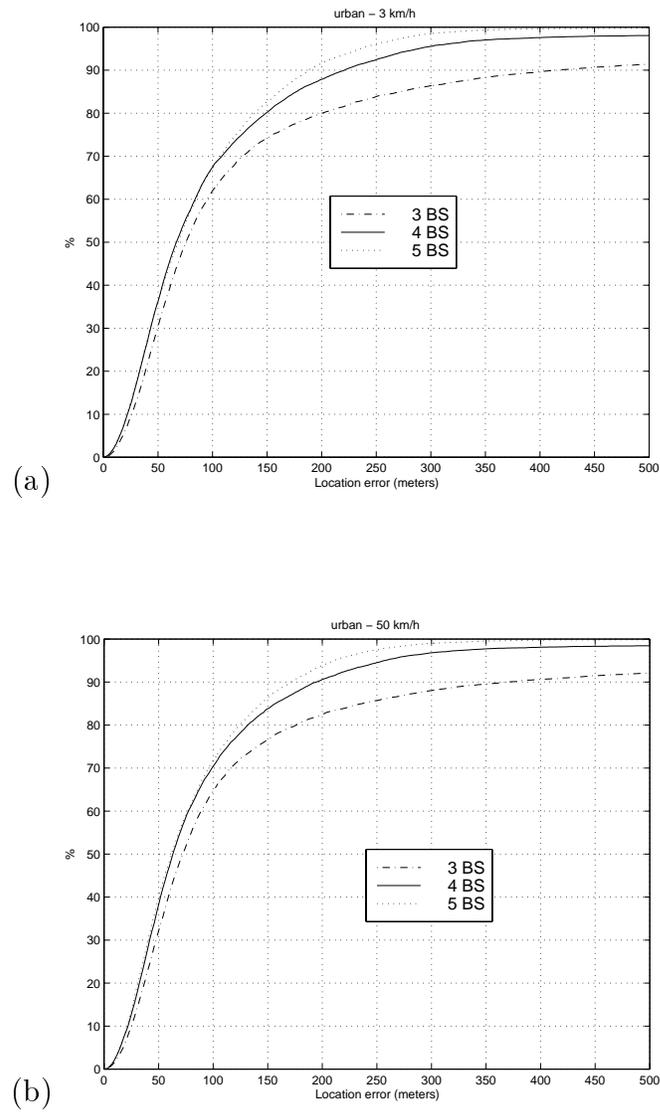
Figure 7.13: Location estimation in the urban environment with synchronization errors: (a) at 3 km/h, (b) at 50 km/h.
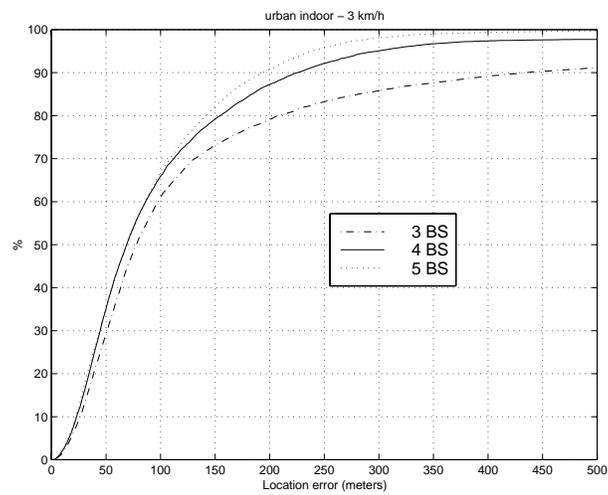
Figure 7.14: Location estimation in the urban indoor environment with synchronization errors: 3 km/h.

(a)



(b)
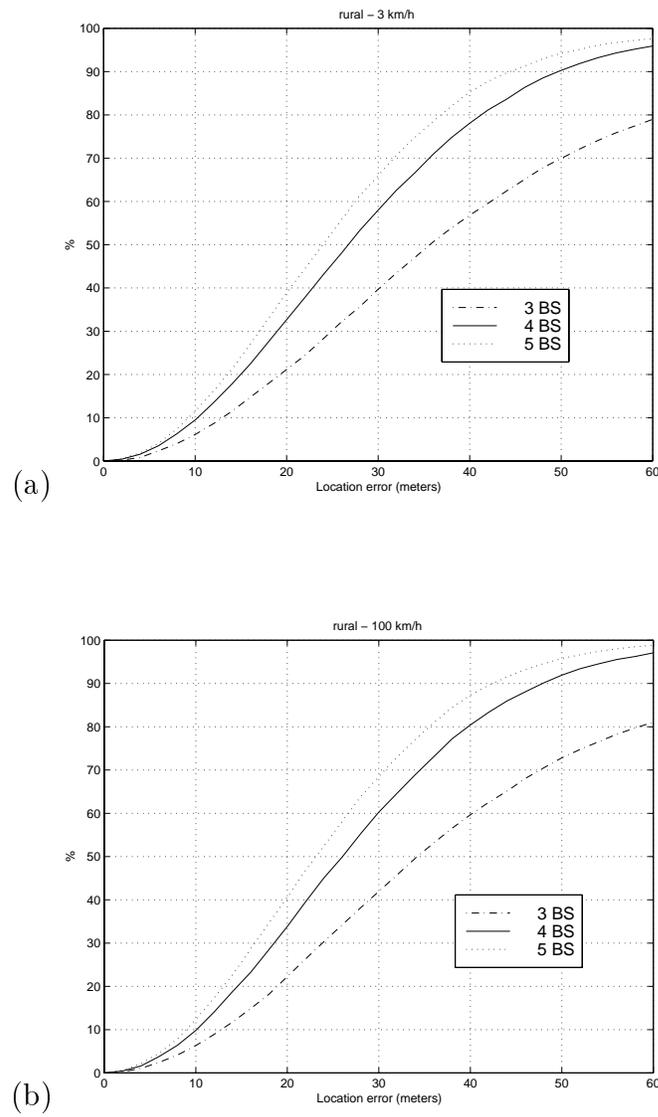
Figure 7.15: Location estimation in the rural environment with synchronization errors: (a) at 3 km/h, (b) at 100 km/h.

(a)



(b)
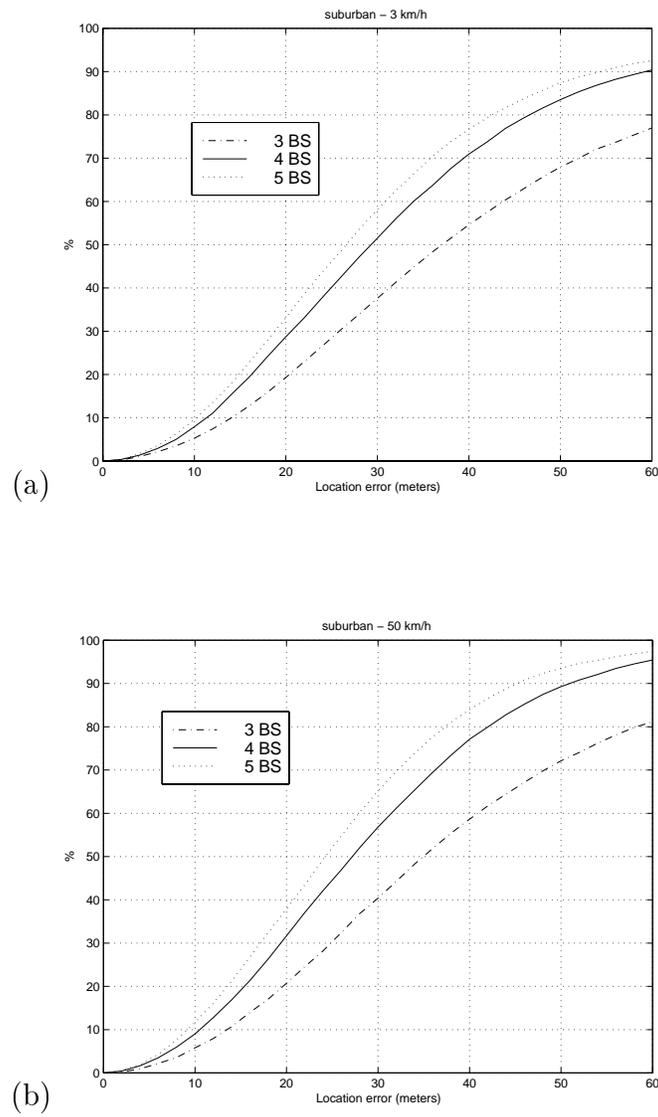
Figure 7.16: Location estimation in the suburban environment with synchronization errors: (a) at 3 km/h, (b) at 50 km/h.

# Chapter 8

# Conclusion and further research

In this dissertation, we have presented a global system for locating a mobile handset in a GSM network. This system makes use of the own capability of the network.

We have shown that the best technology for this purpose is based on time of arrival estimation either at the handset, at the base station, or even both for more accuracy. The main source of problems in time of arrival estimation is multipath. Based on the maximum likelihood estimator, subspace fitting algorithms are applied to estimate the time delays of the dominant paths. The time of arrival is simply the path that has the lowest time delay. An extension is presented to the case of an unknown modulation pulse shape. Two approaches are considered: the first one is iterative, it begins with a first initialization of the pulse shape and converges iteratively to the true one. The second approach is a variant of the well-known ESPRIT algorithm, it is non iterative and has a lower complexity.

Hyperbolic trilateration is then analyzed and appears to be a non linear maximization problem. The scoring approach seems to offer optimal performances at high signal-to-noise ratios and good results at lowest ratios whenever an appropriate initialization is available.

Network synchronization is depicted. A distinction is made between pseudo and absolute synchronization. An algorithm is presented for each approach.

The last chapter presented the global system taking the channel model established in the T1P1.5 committee.

Some aspects have not been treated in this manuscript and might be important to complete the study:

**Diffuse paths**

In this work, we assume that each path arrived at one discrete time delay (specular path). An interesting extension is the case of diffuse paths. In this scenario the partial waves of each path arrive around a mean value.

The correct model for the channel impulse response must be modified as follows:

$$h_j(t) = \sum_{i=1}^{d} \sum_{k=1}^{K_i} s_{ijk} a(t - \tau_i - \tilde{\tau}_{ik}) \qquad 1 \leq j \leq L \tag{8.1}$$

This topic has been studied recently, two main approaches exist:

- Taylor expansion [82, 83].

- The spreading in time is assumed to follow a certain distribution (mainly uniform or Gaussian) centered on the delay and a joint estimation of the distribution mean and variance is performed [84–86].

**CRB for the unknown pulse shape case**

When we estimate the delays assuming an unknown modulation pulse, the comparison of the performances has been done using the standard CRB. Computing the CRB for an unknown modulation pulse shape is difficult to achieve but might be interesting.

**Optimal weighting**

No weighting factor or covariance matrix were applied to the times of arrival while performing the trilateration. It is of course possible to use empirical weighting based on the signal-to-noise ratio or the channel impulse response delay spread. Finding such values remains an open question.

**CDMA systems - UMTS**

Localization is a requirement is future standards such as the Universal Mobile Telecommunications System (UMTS). Unlike GSM, UMTS is a wideband system based on Direct-Sequence Code Division Multiple Access (DS CDMA).

# Chapter 9

# Résumé détaillé en français

## 9.1 Introduction

Depuis que le concept cellulaire fut introduit dans les années soixante, les communications sans fil ont connu un développement sans précédent. Il existe actuellement différents standards. Le système européen GSM est sans doute le standard le plus répandu de nos jours. Ce système est basé sur l'accès multiple des utilisateurs à la fois en temps et en fréquence. Plusieurs utilisateurs peuvent communiquer depuis la même zone sur la même fréquence en se partageant équitablement le temps d'émission.

En juin 1996, le comité fédéral américain de communication (FCC) a demandé aux opérateurs américains de localiser les appels d'urgence émanant de leurs abonnés avec une précision de 125 mètres dans 67 pour cent des cas dès octobre 2001.

La localisation est souvent associée au GPS. Ce système offre une couverture complète du globe terrestre mais demeure peu fiable en zone urbaine. L'objet de cette thèse est de montrer qu'il est possible d'offrir un service de localisation fiable en utilisant les infra-structures déjà existantes des réseaux GSM. L'étude porte sur la localisation par estimation du temps d'arrivée d'au moins trois canaux correspondant à différentes stations de base. L'estimation de la position s'effectue alors moyennant des données de synchronisation des différentes stations de base impliquées.

## 9.2 Communications sans fil

### 9.2.1 Modèle de propagation radio

- Modèle d'atténuation : au fur et à mesure que les ondes radio se propagent, elles perdent de l'intensité. Ainsi dans l'espace libre, la décroissance de l'énergie est inversement proportionnelle au carré de la distance parcourue. Il existe de nombreux modèles empiriques de propagation. Ceux-ci reposent en général sur des mesures de terrain. Le modèle d'Okumura-Hata est le modèle le plus couramment utilisé.

- Evanouissements lent et rapide : outre l'atténuation due à la distance parcourue, le signal radio est sujet à d'autres fluctuations ; fluctuations lentes dues aux masquages provoqués par des obstacles physiques et fluctuations rapides dues à la réflexion du signal sur une surface. En effet, le signal réfléchi provient de la contribution de plusieurs réflexions locales d'amplitudes et de phases différentes. Ces réflexions peuvent ainsi s'ajouter ou s'annuler provoquant ainsi des fluctuations rapides de l'enveloppe du signal reçu.

- Trajet multiple et étalement temporel : le signal radio subit des réflexions multiples. Chaque réflexion se caractérise par une énergie moyenne et une phase. Le signal reçu se compose ainsi d'une multitude de copies. L'enveloppe $r$ de chaque copie est modélisée par une loi de distribution de Rayleigh :

$$p(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \tag{9.1}$$

$\sigma^2 = \mathrm{E}\left[|x|^2\right]$ désigne l'énergie moyenne de l'écho. En revanche, le trajet direct, s'il existe, n'est pas de moyenne nulle, son enveloppe suit la loi de distribution de Rice :

$$p(r) = \frac{r}{\sigma^2} e^{-\frac{r^2+m^2}{2\sigma^2}} I_0\left(\frac{rm}{\sigma^2}\right) \tag{9.2}$$

où $m^2 = |\mathrm{E}[x]|^2$ et $\sigma^2 = \mathrm{E}\left[|x - \mathrm{E}[x]|^2\right]$.

L'étalement temporel est un paramètre important qui caractérise l'environnement, il est donné par :

$$\xi = \sqrt{m_2 - m_1^2} \tag{9.3}$$

où $m_p = \frac{\sum_i a_i \tau_i^p}{\sum_i a_i}$ est le moment d'ordre $p$ du signal. $a_i$ et $\tau_i$ sont respectivement l'énergie moyenne et le retard du $i$-ième trajet. Yuanking [3] a montré par des mesures de terrain que cette quantité suit une loi de distribution aléatoire qui dépend de la distance parcourue :

$$\xi \sim T_1 d^\epsilon y \tag{9.4}$$

$T_1$ et $\epsilon$ sont deux constantes qui dépendent de l'environnement. $d$ est la distance parcourue et $y$ est une variable log-normale.

- Interférences : dans un système de communication cellulaire, la ressource radio est rare. Elle est ainsi partagée et réutilisée. Ceci introduit inévitablement des interférences qu'il faut minimiser lors de la plannification cellulaire. Nous ferons l'approximation que les interférences sont gaussiennes (approximation AWGN). Cette hypothèse est justifiée si le nombre d'interféreurs est suffisamment élevé. Cette hypothèse simplifie énormément les simulations car les interféreurs ne seront pas générés individuellement mais l'ensemble sera généré par une seule variable gaussienne.

- Ecart Doppler : lorsque le mobile bouge, le signal radio est sujet à l'effet Doppler qui le translate en fréquence. Cet écart est d'autant plus important que la vitesse est grande. A 900 Mhz, l'écart maximal est à peu près égal à un Hertz par km/h.

Il existe différents moyens de protection comme le codage de canal, l'entrelacement, la transmission discontinue de parole (DTX), le contrôle de puissance, le saut de fréquence, la diversité spatiale et l'utilisation de séquences d'apprentissage.

## 9.2.2   Aperçu de la norme GSM

Le système GSM est basé sur la combinaison de deux technologies ; l'accès multiple par répartition à la fois en temps et en fréquence. Un groupe de fréquences porteuses est alloué à chaque cellule. Plusieurs mobiles peuvent communiquer sur la même fréquence en utilisant chacun une portion de temps. Le domaine temporel est ainsi découpé en cycles de communication ou trames. Chaque trame se compose de huit portions de temps dédiées chacune à un utilisateur. Une portion de temps est de longueur 577 $\mu$s. En général, le temps de cohérence (ou mémoire) du canal radio est plus long. Cette hypothèse justifie le fait que la réponse impulsionnelle du canal est constante pendant la durée d'une portion de temps.

### Le canal physique duplex

Les fronts de communication montant (mobile vers station de base) et descendant (station de base vers mobile) sont séparés en temps et en fréquence. Il y a trois portions de temps de décalage et un écart fréquentiel constant entre ces deux fronts. Un canal physique GSM représente l'association de deux portions de temps sur les deux fronts. On parle donc de canal physique duplex.

Chaque portion de temps accueille une séquence de bits d'information appelée burst. Celle-ci a une durée inférieure à la durée de la portion de temps

pour éviter tout chevauchement d'un burst sur le suivant. Le burst représente l'information véhiculée et il existe sous différents formats. Il contient toujours une séquence d'apprentissage de bits prédéfinie et une séquence de bits d'information. La séquence d'apprentissage sert à aider le récepteur à identifier le canal radio pour faciliter la récupération des données. Les bursts normaux ont une séquence d'apprentissage de 26 bits, les bursts d'accés 48 bits et les burst de synchronisation 64 bits.

**Les canaux logiques**

Les trames de huit portions de temps sont organisées en multi-trames de 26 ou 51 trames selon la nature de l'information qu'elles véhiculent : trafic ou signalisation. On parle ainsi de canaux logiques dont on distingue les canaux dédiés qui fonctionnent en mode duplex (trafic) et les canaux de signalisation qui ne fonctionnent que sur un seul front montant ou descendant.

**La modulation GMSK**

La modulation utilisée en GSM est une modulation de phase à enveloppe constante appelée GMSK. Cependant, il est toujours possible d'approximer une modulation de phase à codage différentiel par une modulation linéaire [6]. Dans ce cas, le signal reçu s'exprime de manière simple :

$$s(t) \approx \sum_k d_k a(t - kT) \tag{9.5}$$

où $d_k$ désigne la séquence de bits émise et $a(t)$ l'impulsion principale de la modulation GMSK. Cette impulsion s'étale sur quatre temps bits.

## 9.2.3   Egalisation

L'égalisation consiste à démoduler le signal pour en déduire une estimation des bits d'information émis. A la réception, le signal résulte de la convolution du signal d'émission par le filtre d'émission et le filtre représenté par la réponse impulsionnelle du canal, le tout étant noyé dans du bruit blanc supposé gaussien si bien qu'à l'arrivée le signal s'écrit sous la forme suivante :

$$y(t) = \sum_{k=0}^{p-1} d_k h(t - kT) + b(t) \tag{9.6}$$

La fonction $h(t)$ est communément appelée réponse impulsionnelle du canal même si elle intègre tous les filtres mis en jeu (modulation, émission, canal et réception). Il est important de pouvoir estimer la réponse impulsionnelle pour pouvoir égaliser. Cette estimation s'effectue par le biais de la séquence
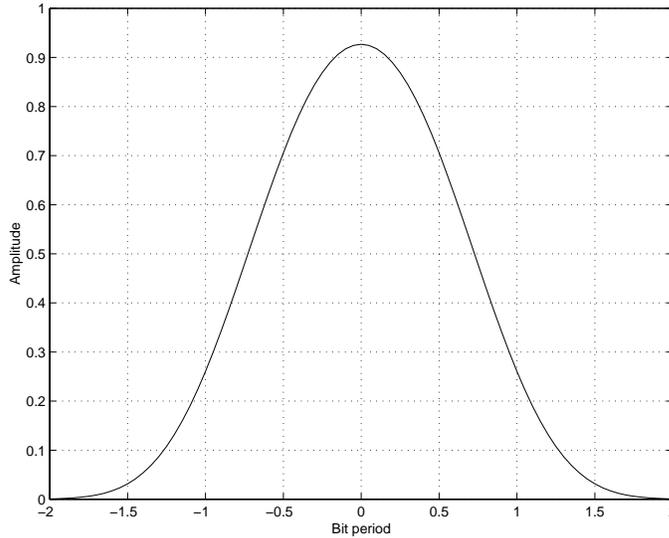
Figure 9.1: Impulsion GMSK principale

d'apprentissage. Ainsi en échantillonnant le signal reçu au rythme du temps bit qui est supposé satisfaire le critère de Nyquist/Shannon, on peut écrire :

$$\mathbf{Y} = \mathbf{D}(\mathbf{d})\mathbf{H} + \mathbf{B} \tag{9.7}$$

où $\mathbf{d}$ représente la séquence de bit émise , $\mathbf{H}$ et $\mathbf{Y}$ représentent respectivement la réponse impulsionnelle et le signal échantillonné. $\mathbf{D}(\mathbf{d})$ s'écrit sous la forme :

$$\mathbf{D}(\mathbf{d}) = \begin{pmatrix} d_p & d_{p-2} & \cdots & d_1 \\ d_{p+1} & d_p & \cdots & d_2 \\ \vdots & \vdots & & \vdots \end{pmatrix} \tag{9.8}$$

Le bruit $\mathbf{B}$ est supposé blanc gaussien décorrélé de moyenne nulle, i.e. $E[\mathbf{B}\mathbf{B}^*] = \sigma^2 \mathbf{I}$.

La stratégie classique mais sous-optimale consiste à estimer la réponse impulsionnelle grâce à la séquence d'apprentissage puis d'estimer les bits d'information en utilisant cette estimation de la réponse impulsionnelle. En général, cette dernière étape s'effectue avec l'algorithme de Viterbi.

En supposant que $\mathbf{d}$ représente la séquence d'apprentissage, l'estimation de la réponse impulsionnelle aux moindres carrés est immédiate et est égale à :

$$\hat{\mathbf{H}} = \mathbf{D}^{\dagger}(\mathbf{d})\mathbf{Y} \qquad (9.9)$$

$\mathbf{D}^{\dagger} = \left[\mathbf{D}^{*}\mathbf{D}\right]^{-1}\mathbf{D}^{*}$ désigne la pseudo-inverse de $\mathbf{D}$.

## 9.3 Technologies pour la localisation

Dans cette partie, différentes technologies sont présentées pour localiser un mobile. Nous nous efforcerons d'expliquer pourquoi l'approche basée sur l'estimation du temps d'arrivée semble être la plus plausible.

### Estimation de la distance

C'est sans doute la méthode la plus intuitive et la plus directe. Il s'agit d'estimer la distance parcourue par le signal reçu en mesurant son intensité à la réception. La puissance à l'émission est connue à travers le paramètre de contrôle de puissance. En utilisant une formule simple de propagation il est possible d'avoir une estimation de la distance. En répétant le même principe sur deux stations de base, trouver la position du mobile devient possible en calculant l'intersection de deux cercles.

### Reconnaissance à base d'apprentissage

Cette méthode ressemble au problème de la reconnaissance de caractère où, à partir de mesures diverses, on essaie de trouver la zone de localisation la plus probable dans laquelle le mobile est susceptible de se trouver. Cette décision s'effectue moyennant une base d'apprentissage composée de mesures recueillies sur le terrain.

### Angle d'arrivée

C'est une technique connue qui a suscité de nombreuses publications ces dernières décennies. Il s'agit d'estimer l'angle d'arrivée d'un signal sur un réseau de plusieurs capteurs espacés d'une longueur d'onde.

### Temps d'arrivée

L'estimation du temps d'arrivée est une ancienne technique utilisée dans de nombreuses applications. Dans le contexte GSM, cette estimation s'effectue grâce à la séquence d'apprentissage. Deux scénarii sont possibles selon que les mesures s'effectuent sur le front montant ou descendant.

## 9.4    Estimation du temps d'arrivée

### 9.4.1    Notations et hypothèses

Nous partons du principe que la réponse impulsionnelle du canal se compose de
$d$ réflexions discrètes. Chaque réflexion est caractérisée par un retard et une
amplitude complexe :

$$h_j(t) = \sum_{i=1}^{d} s_{ij}a(t - \tau_i) \qquad 1 \le j \le L \tag{9.10}$$

L'indice $j$ porte sur la $j$-ième réalisation. En effet la réponse impulsionnelle
du canal, supposée constante pendant la durée d'un burst varie d'un burst à
l'autre. Pour obtenir la meilleure estimation possible, il faut accumuler un max-
imum d'observations. $a(t)$ est l'impulsion GMSK principale. $s_{ij}$ est l'amplitude
complexe associée au $i$-ième retard lors de l'émission du $j$-ième burst. Seuls les
retards $\tau_i$ sont supposés constants pendant la durée d'observation des $L$ bursts.

Le trajet direct n'est pas toujours présent. Le but est d'estimer le retard du
premier trajet d'énergie suffisamment élevée pour qu'il puisse être détecté. La
stratégie choisie pour cela est d'estimer la totalité des retards puis de choisir le
premier comme étant le temps d'arrivée.

### 9.4.2    Approche temporelle

Cette approche consiste à traiter le signal reçu dans le domaine temporel. En
échantillonnant au rythme du temps bit, celui-ci s'écrit sous la forme suivante :

$$\mathbf{Y}_j = \mathbf{D}(\mathbf{d})\mathbf{A}(\tau)\mathbf{S}_j + \mathbf{N}_j \qquad 1 \le j \le L \tag{9.11}$$

Il existe deux versions du maximum de vraisemblance selon la connaissance
que l'on a sur le vecteur des amplitudes complexes $\mathbf{S}_j$. En effet, si ce vecteur
ne suit pas une loi de distribution aléatoire ou que sa loi est tout simplement
inconnue, alors il faut estimer les amplitudes instantanées conjointement avec les
retards, ceci donne le maximum de vraisemblance déterministe dont l'expression
réduite qu'il faut maximiser par rapport à $\tau$ s'écrit :

$$V_{\mathrm{DML}}(\tau) = \mathrm{Tr}\left[\boldsymbol{\Pi}_{\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)}^{\perp} \hat{\mathbf{K}}_y\right] \tag{9.12}$$

En revanche si l'on suppose que ces amplitudes sont gaussiennes de moyennes
nulles alors il faudra estimer leur matrice de covariance $\mathbf{K}_s$ conjointement avec les
retards. Ceci donne le maximum de vraisemblance stochastique dont l'expression
réduite qu'il faut maximiser par rapport à $\tau$ est :

$$V_{\mathrm{SML}}(\tau) = \ln\left|\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)\hat{\mathbf{K}}_s(\tau)\left[\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)\right]^* + \hat{\sigma}^2(\tau)\mathbf{I}\right| \tag{9.13}$$

avec,
$$\hat{\mathbf{K}}_s(\tau) = [\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)]^\dagger \left(\hat{\mathbf{K}}_y - \hat{\sigma}^2(\tau)\mathbf{I}\right) [\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)]^{\dagger*} \tag{9.14}$$

et,
$$\hat{\sigma}^2(\tau) = \frac{\mathrm{Tr}\left(\mathbf{\Pi}^\perp_{\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)}\hat{\mathbf{K}}_y\right)}{m - d} \tag{9.15}$$

Dans les deux cas de figure, déterministe ou stochastique, la statistique suffisante est la matrice de covariance du signal reçu. Il est possible de réduire ces expressions du maximum de vraisemblance en supposant que les séquences d'apprentissage sont parfaites dans le sens où chaque séquence est orthogonale à ses versions décalées, i.e. :
$$\mathbf{D}^*(\mathbf{d})\mathbf{D}(\mathbf{d}) = \mathbf{I} \tag{9.16}$$
Cette égalité est impossible étant donné que les séquences d'apprentissage sont finies. Néanmoins, c'est une bonne approximation pour les longues séquences (burst d'accès et de synchronisation). Les expressions du maximum de vraisemblance deviennent :

$$V_{\mathrm{DML}}(\tau) = \mathrm{Tr}\left[\mathbf{\Pi}^\perp_{\mathbf{A}}(\tau)\hat{\mathbf{K}}_{\hat{h}}\right] \tag{9.17}$$

$$V_{\mathrm{SML}}(\tau) = \ln\left|\mathbf{A}(\tau)\hat{\mathbf{K}}_s(\tau)\mathbf{A}^*(\tau) + \hat{\sigma}_r^2(\tau)\mathbf{I}\right| \tag{9.18}$$

avec,
$$\hat{\mathbf{K}}_s(\tau) = \mathbf{A}^\dagger(\tau)\left(\hat{\mathbf{K}}_{\hat{h}} - \hat{\sigma}_r^2(\tau)\mathbf{I}\right)\mathbf{A}^{\dagger*}(\tau)$$

et,
$$\begin{aligned} \hat{\sigma}_r^2(\tau) &= \frac{\mathrm{Tr}\left(\mathbf{\Pi}^\perp_{\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)}\hat{\mathbf{K}}_y\right)}{m(m - d)} \\ &= \frac{\mathrm{Tr}\left(\mathbf{\Pi}^\perp_{\mathbf{A}}(\tau)\hat{\mathbf{K}}_{\hat{h}}\right)}{p - d} + O(1/\sqrt{L}) \end{aligned} \tag{9.19}$$

Ces formules, d'une complexité moindre, démontrent que la matrice de covariance de l'estimation aux moindres carrés de la réponse impulsionnelle est une statistique suffisante :

$$\hat{\mathbf{K}}_{\hat{h}} = \sum_{j=1}^L \hat{\mathbf{H}}_j\hat{\mathbf{H}}_j^*$$

Les estimateurs au maximum de vraisemblance s'avèrent difficiles à mettre en œuvre car ils exigent une recherche multi dimensionnelle d'un maximum. On leur préfère les estimateurs suivants basés sur l'extraction des sous-espaces signal ou bruit représentés respectivement par les $d$ vecteurs propres associés aux $d$ plus fortes valeurs propres et $p - d$ vecteurs propres associés aux $p - d$ plus faibles valeurs propres de la matrice de covariance :

$$\mathbf{K}_{\hat{h}} = \mathbf{E}_s \boldsymbol{\Lambda}_s \mathbf{E}_s^* + \sigma_r^2 \mathbf{E}_n \mathbf{E}_n^* \tag{9.20}$$

$\mathbf{E}_s$ est appelé sous-espace signal car son image est identique à l'espace linéaire généré par les impulsions GMSK décalées des différents retards. $\mathbf{E}_n$ est orthogonal à $\mathbf{E}_s$.

**MUSIC**

L'estimateur MUSIC exploite l'othogonalité des sous-espaces signal et bruit. Ainsi, les délais sont donnés par les $d$ pics de la fonction suivante appelée aussi spectre MUSIC :

$$\Phi(\tau) = \frac{1}{\mathbf{a}^*(\tau)(\mathbf{I} - \hat{\mathbf{E}}_s \hat{\mathbf{E}}_s^*)\mathbf{a}(\tau)} \tag{9.21}$$

**WSF**

$$[\hat{\tau}, \hat{\mathbf{T}}] = \arg \min_{\tau, \mathbf{T}} \|\hat{\mathbf{E}}_s \mathbf{W}^{\frac{1}{2}} - \mathbf{A}(\tau)\mathbf{T}\|^2 \tag{9.22}$$

$\mathbf{W}$ est une matrice de pondération du sous-espace signal. On peut éliminer $\mathbf{T}$ facilement, on obtient alors :

$$\hat{\tau} = \arg \min_{\tau} \mathrm{Tr} \left( \boldsymbol{\Pi}_{\mathbf{A}}^{\perp}(\tau) \hat{\mathbf{E}}_s \mathbf{W} \hat{\mathbf{E}}_s^* \right) \tag{9.23}$$

Il existe une matrice de pondération optimale qui minimise l'erreur de covariance asymptotique [39] :

$$\mathbf{W} = \mathbf{W}_{\mathrm{OPT}} = (\boldsymbol{\Lambda}_s - \sigma_r^2 \mathbf{I})^2 \boldsymbol{\Lambda}_s^{-1} \tag{9.24}$$

En pratique, cette matrice peut être remplacée par un estimateur consistant sans affecter les propriétés asymptotiques de l'estimateur. On démontre alors que l'estimateur réalise les mêmes performances que le maximum de vraisemblance stochastique lorsque le nombre d'observations $L$ est suffisamment élevé. En ce sens, l'estimateur WSF réalise la meilleur performance et atteint l'erreur quadratique minimale connue sous le nom de borne de Cramer-Rao [26].

### 9.4.3   Approche fréquentielle

Dans cette approche, on traite le signal dans le domaine fréquentiel. La tranformée de Fourier de la réponse impulsionnelle du canal s'écrit :

$$\bar{\tilde{h}}_j(f) = \tilde{h}_j(f) = \bar{a}(f) \sum_{i=1}^{n} s_{ij} e^{-j2\pi\tau_i f} + \bar{w}_j(f) \qquad 1 \leq j \leq L \tag{9.25}$$

$\bar{(.)}$ représente la transformée de Fourier de $(.)$, $\bar{w}$ est la transformée de Fourier du bruit complexe de variance réduite $\sigma_r^2$.

Nous prenons le même nombre d'échantillons que dans le domaine temporel. Les fréquences d'échantillonnage $f = (f_1, \cdots, f_p)$ sont prises sur la partie non nulle de $\bar{a}(f)$. Les observations s'écrivent :

$$\bar{\tilde{\mathbf{H}}}_j = \tilde{\mathbf{H}}_j = \mathbf{\Gamma}(\gamma)\tilde{\mathbf{A}}(\tau)\mathbf{S}_j + \bar{\mathbf{W}}_j \qquad 1 \le j \le L \qquad (9.26)$$

avec $\mathbf{\Gamma}(\gamma) = \mathrm{diag}(\gamma)$, $\gamma = [\bar{a}(f_1)\cdots\bar{a}(f_p)]^T$ et :

$$\tilde{\mathbf{A}}(\tau) = \begin{pmatrix} e^{-2\pi j f_1 \tau_1} & \cdots & e^{-2\pi j f_1 \tau_d} \\ \vdots & & \vdots \\ e^{-2\pi j f_p \tau_1} & \cdots & e^{-2\pi j f_p \tau_d} \end{pmatrix}$$

La différence fondamentale de cette nouvelle représentation par rapport à l'ancienne réside dans la séparation de l'information des retards de celle de l'impulsion ; la matrice $\tilde{\mathbf{A}}(\tau)$ dépend maintenant uniquement des retards et l'information de l'impulsion se trouve désormais dans la matrice $\mathbf{\Gamma}(\gamma)$. De plus, le fait que les fréquences d'échantillonnage soient uniformément espacées fait que la matrice $\tilde{\mathbf{A}}(\tau)$ a une structure de Vandermonde. De nouvaux algorithmes plus élégants peuvent alors être appliqués.

## Root MUSIC

Dans un environnement sans bruit, la fonction suivante s'annule lorsque $\tau$ est égal à l'un des retards.

$$\|\hat{\bar{\mathbf{E}}}_n^* \mathbf{\Gamma}(\gamma)\tilde{\mathbf{a}}(\tau)\|^2 = \|\hat{\bar{\mathbf{E}}}_n^* \mathbf{\Gamma}(\gamma)[1, z, \cdots, z^p]^T\|^2 \qquad (9.27)$$

avec $z = z(\tau) = e^{-j2\pi\tau\Delta}$ et $\Delta$ l'espacement entre deux fréquences consécutives. Etant donné que $z$ est sur le cercle unité, on peut substituer $z$ par $\frac{1}{z^*}$. Il s'ensuit que la fonction précédente devient un polynôme. Root MUSIC revient à calculer les $d$ racines de module inférieur à un et les plus proches du cercle unité. Les retards sont alors calculés à partir des arguments de ces racines complexes.

## IQML et Root WSF

Ces deux algorithmes sont basés sur une paramétrisation différente. En notant :

$$\mathbf{B}^*(\mathbf{b}) = \begin{pmatrix} b_d & \cdots & b_0 & 0 & \cdots & 0 \\ 0 & b_d & \cdots & b_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & b_d & \cdots & b_0 \end{pmatrix} \qquad (9.28)$$

où $\mathbf{b} = [b_0 \cdots b_d]^T$ sont les coefficients du polynôme suivant :

$$\sum_{i=0}^{d} b_i z^{d-i} = b_0 \prod_{i=1}^{d} (z - e^{-j2\pi\tau_i \Delta}) \tag{9.29}$$

Il découle de cette définition que $\mathbf{B}^*(\mathbf{b})\tilde{\mathbf{A}}(\tau) = 0$. Or, comme le rang de $\mathbf{B}(\mathbf{b})$ est par construction égale à $p - d$, il vient :

$$\mathbf{\Pi_B}(\mathbf{b}) = \mathbf{\Pi}_{\tilde{\mathbf{A}}}^{\perp}(\tau) \tag{9.30}$$

De même,

$$\mathbf{\Pi_{\Gamma^{-*}B}}(\mathbf{b}) = \mathbf{\Pi}_{\mathbf{\Gamma}\tilde{\mathbf{A}}}^{\perp}(\tau) \tag{9.31}$$

Il est équivalent d'estimer $\mathbf{b}$ ou $\tau$. Ils sont directement liés. Il est cependant plus facile d'estimer $\mathbf{b}$. Le critère du maximum de vraisemblance déterministe s'écrit :

$$V_{\text{DML}}(\mathbf{b}) = \text{Tr}\left[\mathbf{\Pi_{\Gamma^{-*}B}}(\mathbf{b})\hat{\mathbf{K}}_{\tilde{h}}\right] \tag{9.32}$$

L'algorithme IQML est un algorithme itératif qui minimise cette fonction. Ainsi à chaque itération, on a :

$$\mathbf{b}_k = \arg\min_{\mathbf{b}} \text{Tr}\left[\mathbf{B}(\mathbf{b})\left[\mathbf{B}^*(\mathbf{b}_{k-1})\mathbf{\Gamma}^{-1}\mathbf{\Gamma}^{-*}\mathbf{B}(\mathbf{b}_{k-1})\right]^{-1}\mathbf{B}^*(\mathbf{b})\mathbf{\Gamma}^{-1}\hat{\mathbf{K}}_{\tilde{h}}\mathbf{\Gamma}^{-*}\right] \tag{9.33}$$

On peut initialiser $\mathbf{b}$ de telle sorte que $\mathbf{B}^*(\mathbf{b_0})\mathbf{\Gamma}^{-1}\mathbf{\Gamma}^{-*}\mathbf{B}(\mathbf{b_0}) = \mathbf{I}$. L'avantage est que chaque itération est une simple minimisation d'une forme quadratique. Pour éviter la solution triviale $\mathbf{b} = 0$, on peut ajouter la contrainte $\|\mathbf{b}\| = 1$ en normalisant $\mathbf{b}_k$ à chaque itération.

Root WSF est obtenu en remplaçant simplement la matrice $\hat{\mathbf{K}}_{\tilde{h}}$ par $\tilde{\mathbf{E}}_s \tilde{\mathbf{W}}_{\text{OPT}} \tilde{\mathbf{E}}_s^*$.

## ESPRIT

Cet algorithme exploite la structure Vandermonde de $\tilde{\mathbf{A}}(\tau)$. En effet, $\tilde{\mathbf{A}}(\tau)$ satisfait la propriété d'invariance suivante [50] :

$$\mathbf{J}_1\tilde{\mathbf{A}}(\tau) = \mathbf{J}_2\tilde{\mathbf{A}}(\tau)\mathbf{\Phi}(\tau) \qquad \mathbf{\Phi}(\tau) = \text{diag}(e^{2\pi j \Delta \tau_1}, \cdots, e^{2\pi j \Delta \tau_d}) \tag{9.34}$$

où $\mathbf{J}_1$ and $\mathbf{J}_2$ sont respectivement les matrices de sélection des premières et dernières $p - 1$ lignes, i.e. :

$$\mathbf{J}_1 = [\mathbf{I}_{p-1} \; 0] \qquad \mathbf{J}_2 = [0 \; \mathbf{I}_{p-1}]$$

Le but est de déterminer la matrice $\mathbf{\Phi}(\tau)$ qui donne directement les différents retards. Étant donné que l'espace engendré par $\tilde{\mathbf{E}}_s$ est identique à celui engendré par $\mathbf{\Gamma}(\gamma)\tilde{\mathbf{A}}(\tau)$, il existe une matrice unique de rang plein $\mathbf{T}$ telle que :

$$\mathbf{J}_1\tilde{\mathbf{E}}_s = \mathbf{J}_1\boldsymbol{\Gamma}(\gamma)\tilde{\mathbf{A}}(\tau)\mathbf{T} \qquad \mathbf{J}_2\tilde{\mathbf{E}}_s = \mathbf{J}_2\boldsymbol{\Gamma}(\gamma)\tilde{\mathbf{A}}(\tau)\mathbf{T}$$

En notant $\boldsymbol{\Gamma}_1$, la matrice $\boldsymbol{\Gamma}$ après élimination de sa dernière ligne et dernière colonne, et $\boldsymbol{\Gamma}_2$, la matrice $\boldsymbol{\Gamma}$ après élimination de sa première ligne et première colonne, on peut écrire :

$$\mathbf{J}_1\tilde{\mathbf{E}}_s = \boldsymbol{\Gamma}_1\mathbf{J}_1\tilde{\mathbf{A}}(\tau)\mathbf{T} \qquad \mathbf{J}_2\mathbf{E}_s = \boldsymbol{\Gamma}_2\mathbf{J}_2\tilde{\mathbf{A}}(\tau)\mathbf{T}$$

La propriété d'invariance donne alors :

$$\mathbf{J}_1\tilde{\mathbf{E}}_s = \boldsymbol{\Gamma}_1\mathbf{J}_2\tilde{\mathbf{A}}(\tau)\boldsymbol{\Phi}(\tau)\mathbf{T} \qquad \mathbf{J}_2\tilde{\mathbf{E}}_s = \boldsymbol{\Gamma}_2\mathbf{J}_2\tilde{\mathbf{A}}(\tau)\mathbf{T}$$

soit :

$$\boldsymbol{\Gamma}_2\boldsymbol{\Gamma}_1^{-1}\mathbf{J}_1\tilde{\mathbf{E}}_s = \mathbf{J}_2\tilde{\mathbf{E}}_s\boldsymbol{\Psi}(t) \qquad (9.35)$$

Avec $\boldsymbol{\Psi}(t) = \mathbf{T}^{-1}\boldsymbol{\Phi}(\tau)\mathbf{T}$. Les valeurs propres de $\boldsymbol{\Psi}(t)$ sont identiques à celles de $\boldsymbol{\Phi}$ et donnent donc les retards. Pour estimer $\boldsymbol{\Psi}(\mathrm{t})$, on peut utiliser une simple estimation aux moindres carrés :

$$\hat{\boldsymbol{\Psi}}(\tau) = \left(\mathbf{J}_2\hat{\tilde{\mathbf{E}}}_s\right)^{\dagger}\boldsymbol{\Gamma}_2\boldsymbol{\Gamma}_1^{-1}\mathbf{J}_1\hat{\tilde{\mathbf{E}}}_s$$

### 9.4.4 Extension au cas d'une modulation inconnue

Les algorithmes présentés dans les sections précédentes supposent une connaissance parfaite de la forme de l'impulsion de modulation. Ces algorithmes sont cependant sensibles à une déformation de cette impulsion. Une erreur, même minime, sur la forme de cette impulsion peut engendrer un biais important dans les résultats d'estimation. Il est difficile en pratique d'avoir une connaissance parfaite car cette impulsion est modifiée par les filtres d'émission et de réception, ceux-ci pouvant être différents d'un constructeur à l'autre.

Nous proposons ici une extension dans le cas où l'impulsion est totalement inconnue. En effet, il est possible d'estimer conjointement l'impulsion et les retards à une référence temporelle absolue près, i.e. on est capable d'estimer les différences entre les différents retards sans pouvoir toutefois en déduire leur valeur. Ceci ne pose pas de problème pour la localisation étant donné que seules les différences de temps d'arrivée nous interesse. Il est néanmoins possible d'estimer le temps absolu en utilisant une simple information a priori sur l'impulsion. On utilisera l'information que l'impulsion de modulation est réelle et symétrique pour se débarasser de cette ambiguïté.

## Approche itérative

La méthode la plus directe est de reprendre l'idée que le sous-espace signal est engendré par $\tilde{\mathbf{E}}_s$.

$$[\hat{\gamma} \ \hat{\tau}]^T = \arg\max_{\gamma,\tau} \sum_{i=1}^{d} \|\hat{\tilde{\mathbf{E}}}_s^* \mathbf{\Gamma}(\gamma)\tilde{\mathbf{a}}(\tau_i)\|^2 \tag{9.36}$$

qui est équivalent à :

$$[\hat{\gamma} \ \hat{\tau}]^T = \arg\max_{\gamma,\tau} \gamma^* \sum_{i=1}^{d} \mathbf{D}_i^* \hat{\tilde{\mathbf{E}}}_s \hat{\tilde{\mathbf{E}}}_s^* \mathbf{D}_i \gamma \tag{9.37}$$

où $\mathbf{D}_i = \text{diag}[\tilde{\mathbf{a}}(\tau_i)]$. $\gamma$ est donc le vecteur propre associé à la valeur propre maximale de la matrice $\sum_{i=1}^{d} \mathbf{D}_i^* \hat{\tilde{\mathbf{E}}}_s \hat{\tilde{\mathbf{E}}}_s^* \mathbf{D}_i$ à un facteur multiplicatif près. Ce facteur multiplicatif peut être déterminé grâce aux hypothèses faites sur l'impulsion. L'idée est donc de partir d'une estimation initiale des retards et d'en déduire ainsi la forme de l'impulsion représentée par le vecteur $\gamma$. Une fois cette estimation disponible, on estime à nouveau les retards en utilisant un algorithme quelconque et ainsi de suite jusqu'à ce qu'une convergence soit atteinte.

## Approche non itérative

En reprenant la même idée que l'estimateur ESPRIT et en notant $\mathbf{D}(\eta) = \text{diag}(\eta)$ la matrice diagonale $\mathbf{\Gamma}_2\mathbf{\Gamma}_1^{-1}$, et $\mathbf{\Psi}(\tau) = \mathbf{T}^{-1}\mathbf{\Phi}(\tau)\mathbf{T}$, l'estimateur conjoint s'écrit de la manière suivante :

$$[\hat{\eta} \ \hat{\mathbf{\Psi}}(\tau)]^T = \arg\min_{\eta \ \Psi} \|\mathbf{D}(\eta)\mathbf{J}_1\hat{\mathbf{E}}_s - \mathbf{J}_2\hat{\mathbf{E}}_s\mathbf{\Psi}(\tau)\|^2 \tag{9.38}$$

La solution de ce problème est donnée par :

$$\hat{\mathbf{\Psi}}(\tau) = \left(\mathbf{J}_2\hat{\mathbf{E}}_s\right)^{\dagger} \mathbf{D}(\hat{\eta})\mathbf{J}_1\hat{\mathbf{E}}_s$$

En incorporant cette equation dans (9.38), on obtient [60] :

$$\hat{\eta} = \arg\min_{\eta} \text{Tr}\left[\mathbf{D}(\eta)\mathbf{J}_1\hat{\mathbf{E}}_s\hat{\mathbf{E}}_s^*\mathbf{J}_1^T\mathbf{D}^*(\eta)\mathbf{\Pi}_{\mathbf{J}_2\hat{\mathbf{E}}_s}^{\perp}\right] = \arg\min_{\eta} \eta^*\left[\mathbf{\Pi}_{\mathbf{J}_2\hat{\mathbf{E}}_s}^{\perp} \odot \mathbf{J}_1(\hat{\mathbf{E}}_s\hat{\mathbf{E}}_s^*)^T\mathbf{J}_1^T\right]\eta$$

$\eta$ est égal au vecteur propre associé à la valeur propre minimale de la matrice $\mathbf{\Pi}_{\mathbf{J}_2\hat{\mathbf{E}}_s}^{\perp} \odot \mathbf{J}_1(\hat{\mathbf{E}}_s\hat{\mathbf{E}}_s^*)^T\mathbf{J}_1^T$ à un facteur multiplicatif près. Ce facteur multiplicatif peut être déterminé grâce aux hypothèses faites sur l'impulsion.

## Différence de temps d'arrivée

Dans les deux algorithmes décrits précédemment, un facteur multiplicatif inconnu persiste si aucune hypothèse sur l'impulsion n'est disponible. En fait dans les deux cas, seule la phase de ce facteur a une influence sur l'estimation des

retards. Cette influence se traduit par un décalage temporel de tous les retards. Pour pouvoir localiser un mobile, seules les différences de temps d'arrivée sont utilisées. Un tel décalage n'a aucune incidence sur cette différence de temps du moment qu'il est identique sur tous les liens. En résumé, il suffit de choisir la même phase du coefficient multiplicatif (par exemple nulle).

Il est possible d'affiner davantage la précision finale en supposant que l'impulsion est identique sur les différents liens de communication. L'approche non itérative devient alors :

$$[\hat{\eta}, \hat{\mathbf{\Psi}}_1, \hat{\mathbf{\Psi}}_2, \cdots]^T = \arg \min_{\eta, \mathbf{\Psi}_1, \mathbf{\Psi}_2, \cdots} \sum_i \alpha_i \|\mathbf{D}(\eta)\mathbf{J}_1\hat{\mathbf{E}}_{s_i} - \mathbf{J}_2\hat{\mathbf{E}}_{s_i}\mathbf{\Psi}_i\|^2 \qquad (9.39)$$

où $\mathbf{\Psi}_i$ est la matrice $\mathbf{\Psi}(\tau)$ qui correspond au $i$-ième lien, $\eta$ le vecteur colonne qui contient l'information de l'impulsion identique et partagée sur les différents liens. $\hat{\mathbf{E}}_{s_i}$ est le sous-espace signal estimé du $i$-ième lien et $\alpha_i$ un facteur de pondération qui reflète le degré de confiance de chaque lien. Il peut être fonction du rapport signal-sur-bruit ou de l'étalement temporel. Il est cependant difficile d'affecter une valeur optimale à ces valeurs. La solution est :

$$\hat{\mathbf{\Psi}}_i = \left(\mathbf{J}_2\hat{\mathbf{E}}_{s_i}\right)^{\dagger}\mathbf{D}(\hat{\eta})\mathbf{J}_1\hat{\mathbf{E}}_{s_i}$$

et :

$$\hat{\eta} = \arg \min_{\eta} \eta^* \left[\sum_i \alpha_i \left(\mathbf{\Pi}^{\perp}_{\mathbf{J}_2\hat{\mathbf{E}}_{s_i}} \odot \mathbf{J}_1(\hat{\mathbf{E}}_{s_i}\hat{\mathbf{E}}^*_{s_i})^T\mathbf{J}_1^T\right)\right]\eta$$

### 9.4.5 Simulations

Le modèle de propagation utilisé est simple avec deux échos espacés d'un temps bit et de même puissance moyenne. L'impulsion de modulation est l'impulsion GMSK filtrée à $1/T$ et échantillonnée au temps bit. Dix échantillons sont utilisés, i.e. $p = 10$.

#### Performances des algorithmes classiques : corrélation nulle

Lorsque le nombre d'observations est suffisamment élevé, Root MUSIC et Root WSF sont optimaux car ils atteignent la borne de Cramer-Rao. En revanche, ESPRIT leur est légèrement inférieur (Figure 9.2).

#### Performances des algorithmes classiques : corrélation élevée

Lorsque les échos sont corrélés, seul Root WSF est optimal (Figure 9.3).
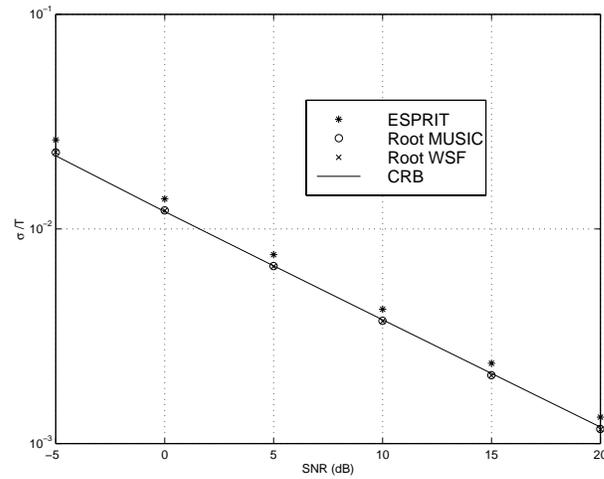
Figure 9.2: Performances des algorithmes en fonction du rapport signal sur bruit, $L = 100$ bursts de synchronisation sont utilisés, $\Delta\tau = 1T$ et $\rho = 0$.
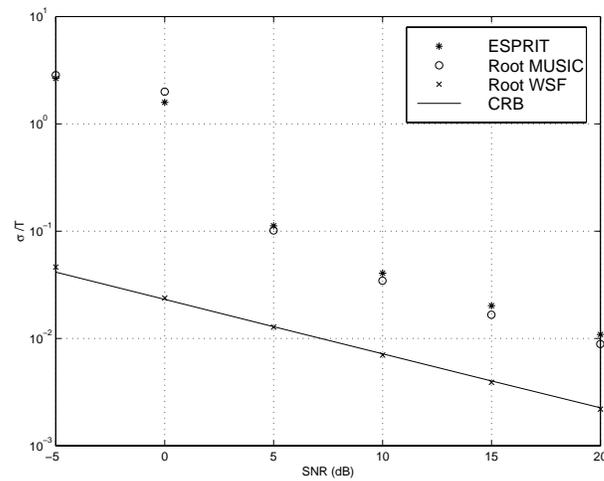


Figure 9.3: Performances des algorithmes en fonction du rapport signal sur bruit, $L = 100$ bursts de synchronisation sont utilisés, $\Delta\tau = 1T$ et $\rho = 0.99$.

## Performances d'ESPRIT et d'ESPRIT modifié

Schéma identique avec une corrélation nulle. L'algorithme ESPRIT modifié est légèrement inférieur, ce qui est normal vu qu'il ne prend pas en compte la con-

naissance de l'impulsion (Figure 9.4.5).
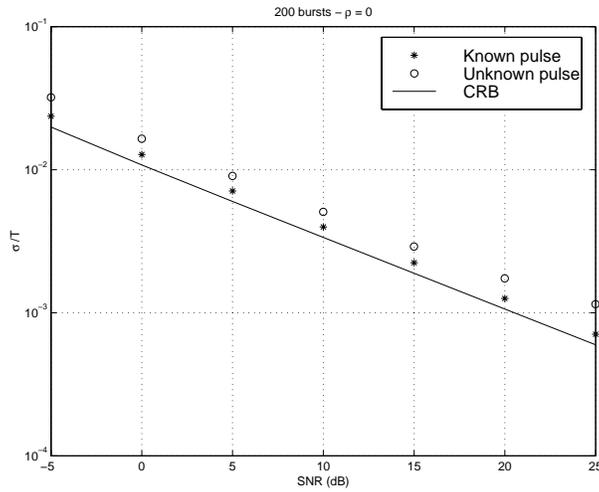


Figure 9.4: ESPRIT et ESPRIT modifié, $L = 200$ bursts de synchronisation sont utilisés.

## 9.5  Triangulation hyperbolique

Nous nous intéressons ici à la fusion des données représentées par les estimations des temps d'arrivée. Nous prendrons en compte une éventuelle erreur de synchronisation des stations de base. Le système à résoudre s'écrit de la manière suivante :

$$\begin{cases} \tilde{\mathbf{F}} = \mathbf{A}(\theta) + \mathbf{T} - q_1\mathbf{1} + \mathbf{B}_f \\ \bar{\mathbf{T}} = \mathbf{T} - q_2\mathbf{1} + \mathbf{B}_t \end{cases} \tag{9.40}$$

où,

- $\theta$ représente les coordonnées à estimer (en deux dimensions $\theta = [x \ y]^T$),

- $\tilde{\mathbf{F}}$ représente les temps d'arrivée estimés,

- $\mathbf{A}(\theta)$ représente les distances qui séparent les $n$ stations de base du mobile, ce vecteur dépend donc de $\theta$,

- $\mathbf{T}$ représente les temps d'émission,

- $\mathbf{B}_f$ et $\mathbf{B}_t$ sont des vecteurs de bruit,

- $q_1$ et $q_2$ sont deux temps de référence inconnus,

- $\mathbf{1} = [1 \cdots 1]^T$.

On supposera que les deux vecteurs de bruit $\mathbf{B}_f$ and $\mathbf{B}_t$ sont gaussiens de moyenne nulle :

- $\mathrm{E}[\mathbf{B}_f \mathbf{B}_f^T] = \mathbf{Q}_f$

- $\mathrm{E}[\mathbf{B}_t \mathbf{B}_t^T] = \mathbf{Q}_t$

- $\mathrm{E}[\mathbf{B}_f \mathbf{B}_t^T] = 0$

Il est possible d'exprimer ce système sous forme compacte par :

$$\mathbf{F} = \mathbf{A}(\theta) - t_{\mathrm{r}}\mathbf{1} + \mathbf{B} \tag{9.41}$$

où $\mathbf{F} = \tilde{\mathbf{F}} - \bar{\mathbf{T}}$. Cette équation aboutit à une triangulation hyperbolique. Le nombre de stations de base $n$ doit être strictement supérieur à deux si on localise en deux dimensions. Il y a trois variables à estimer ; les coordonnées et le temps de référence $t_{\mathrm{r}}$. Bien entendu, lorsque $n$ augmente, l'erreur d'estimation se trouve réduite.

L'estimateur du maximum de vraisemblance s'écrit :

$$[\hat{\theta} \ \hat{t}_{\mathrm{r}}] = \arg \max_{\theta, t_{\mathrm{r}}} [\mathbf{F} - \mathbf{A}(\theta) - t_{\mathrm{r}}\mathbf{1}]^T \mathbf{Q}^{-1} [\mathbf{F} - \mathbf{A}(\theta) - t_{\mathrm{r}}\mathbf{1}] \tag{9.42}$$

En éliminant $t_{\mathrm{r}}$, on obtient une expression plus compacte :

$$\hat{\theta} = \arg \max_{\theta} [\mathbf{F} - \mathbf{A}(\theta)]^T \tilde{\mathbf{Q}} [\mathbf{F} - \mathbf{A}(\theta)] \tag{9.43}$$

avec,

$$\tilde{\mathbf{Q}} = \mathbf{Q}^{-1} - \frac{\mathbf{Q}^{-1}\mathbf{1}\mathbf{1}^T\mathbf{Q}^{-1}}{\mathbf{1}^T\mathbf{Q}^{-1}\mathbf{1}}$$

### Développement en série de Taylor

C'est un algorithme itératif qui cherche à minimiser directement le critère du maximum de vraisemblance. Chaque itération s'écrit de la manière suivante :

$$\theta_n = \theta_{n-1} + \left[ \left( \frac{\partial \mathbf{A}(\theta_{n-1})}{\partial \theta} \right)^T \tilde{\mathbf{Q}} \frac{\partial \mathbf{A}(\theta_{n-1})}{\partial \theta} \right]^{-1} \left( \frac{\partial \mathbf{A}(\theta_{n-1})}{\partial \theta} \right)^T \tilde{\mathbf{Q}} [\mathbf{F} - \mathbf{A}(\theta_{n-1})]$$
$$\tag{9.44}$$

La convergence de cet algorithme dépend en grande partie de son initialisation.

**Méthode des moindres carrés**

Cette méthode est basée sur une reformulation des équations. Ainsi en écrivant les $n-1$ équations suivantes :

$$(r_i - r_1)(r_i + r_1) = r_i^2 - r_1^2 \qquad 2 \leq i \leq n$$

où $r_i$ désigne la distance entre le mobile et la $i$-ième station de base. On peut obtenir le système compact suivant :

$$[\mathbf{HA}(\theta)] \odot [(\mathbf{H} \odot \mathbf{H})\mathbf{A}(\theta)] = -2\mathbf{HJ}\theta + \mathbf{HK} \qquad (9.45)$$

où $\mathbf{H}$ désigne la matrice $(n-1) \times n$ suivante :

$$\mathbf{H} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -1 & 0 & \cdots & 0 & 1 \end{bmatrix} \qquad (9.46)$$

et,

$$\mathbf{J} = \frac{1}{c^2} \begin{bmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix} \qquad \mathbf{K} = \frac{1}{c^2} \begin{bmatrix} x_1^2 + y_1^2 \\ \vdots \\ x_n^2 + y_n^2 \end{bmatrix} \qquad (9.47)$$

Notons que les éléments de $\mathbf{HA}(\theta)$ sont les différences de temps d'arrivée (DToA) entre la $i$-ième station de base et la station de base courante à laquelle le mobile est connecté. Soit, après quelques manipulations algébriques :

$$[\mathbf{HA}(\theta)] \odot [\mathbf{HA}(\theta)] - \mathbf{HK} = -2\mathbf{HJ}\theta - 2\frac{r_1}{c}\mathbf{HA}(\theta) \qquad (9.48)$$

Sans perte de généralités, nous pouvons considérer que la station de base courante est à l'origine, soit de manière équivalente $r_1 = |\theta|$. L'équation prend alors la forme suivante :

$$\frac{1}{2}\mathbf{HK} - \frac{1}{2}[\mathbf{HF}] \odot [\mathbf{HF}] = \mathbf{HJ}\theta + \frac{|\theta|}{c}\mathbf{HF} + \mathbf{B}_1 \qquad (9.49)$$

A fort rapport signal sur bruit, on peut négliger les termes d'ordre deux du bruit, et $\mathbf{B}_1$ demeure gaussien de moyenne nulle :

$$\mathbf{B}_1 \approx -\mathbf{RHB}$$

où $\mathbf{R} = \frac{1}{c}\text{diag}(r_2, \cdots, r_n)$. Sa covariance s'écrit :

$$\mathbf{Q}_1 = \mathbf{RH}^T\mathbf{QHR} \qquad (9.50)$$

Finalement, on tombe sur le problème non linéaire suivant :

$$\mathbf{Y} = \mathbf{M}\theta + |\theta|\mathbf{V} + \mathbf{B}_1 \tag{9.51}$$

où,

$$\mathbf{Y} = \frac{1}{2}\mathbf{H}\mathbf{K} - \frac{1}{2}\left[\mathbf{H}\mathbf{F}\right] \odot \left[\mathbf{H}\mathbf{F}\right]$$

et,

$$\mathbf{M} = \mathbf{H}\mathbf{J} \qquad \mathbf{V} = \frac{1}{c}\mathbf{H}\mathbf{F}$$

L'estimation se fait alors en deux étapes. On estime dans un premier temps $\theta$ et $|\theta|$ aux moindres carrés comme s'il s'agissait de deux variables indépendantes, on appellera cette estimateur l'estimateur des moindres carrés. Dans un deuxième temps, on inclura la contrainte entre ces deux variables, cet estimateur est dû à Chan [70].

## Simulations

Les simulations ont été effectuées sur un réseau hexagonal composé de cent stations de base espacées les unes des autres de dix km. 200 mobiles sont générés aléatoirement au milieu du réseau afin d'éviter les effets de bord. (Figure 9.5).
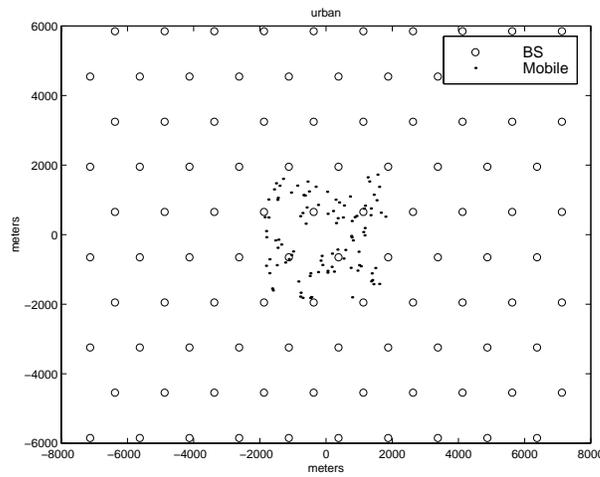


Figure 9.5: Modèle : 'o' représente une station de base et les points différentes positions de mobiles.

Cinq stations de base sont concernées par la localisation. 20 simulations de Monte Carlo sont effectuées par mobile, soit un total de 4000 simulations. Les

résultats (Figure 9.6) montrent que l'algorithme de Taylor initialisé par l'approche de Chan et avec deux itérations seulement semble offrir de bonnes performances à bon rapport signal sur bruit.
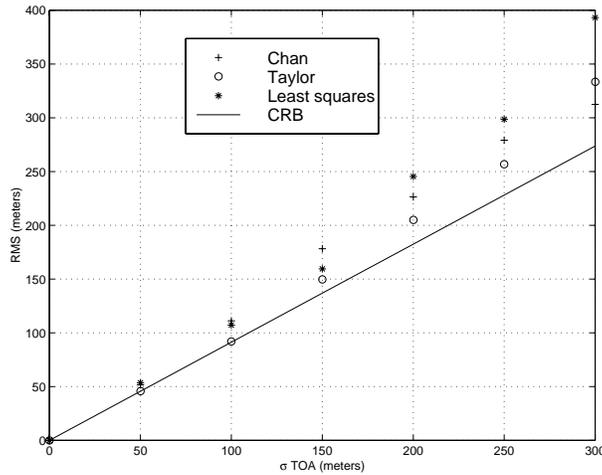


Figure 9.6: Résultats de simulation

## 9.6 Synchronisation du réseau

On entend par synchronisation du réseau la connaissance des temps d'émission des trames TDMA de toutes les stations de base du réseau, ou plus précisemment des différences de temps d'émission (RTD) étant donné qu'il est difficile d'avoir une référence absolue partagée par toutes les stations. Il existe deux degrés de synchronisation :

- Pseudo-synchronisation : dans ce cas les différences de transmission sont connues avec exactitude.

- Synchronisation absolue : dans ce cas les temps d'émission sont identiques et les différences sont donc nulles.

Pour pouvoir localiser, la pseudo-synchronisation est suffisante mais la synchronisation offre d'autres avantages. Nous discuterons ci-dessous des deux problèmes séparément.

**Pseudo-synchronisation**

Il s'agit d'estimer au maximum de vraisemblance les différences de temps d'émission à partir d'observations bruitées des différences de transmission. Toutes les différences ne sont peut-être pas observées et certaines sont peut-être observées plusieurs fois. La seule condition pour pouvoir identifier toutes les différences est que le réseau des observations soit connexe (Figure 9.7).
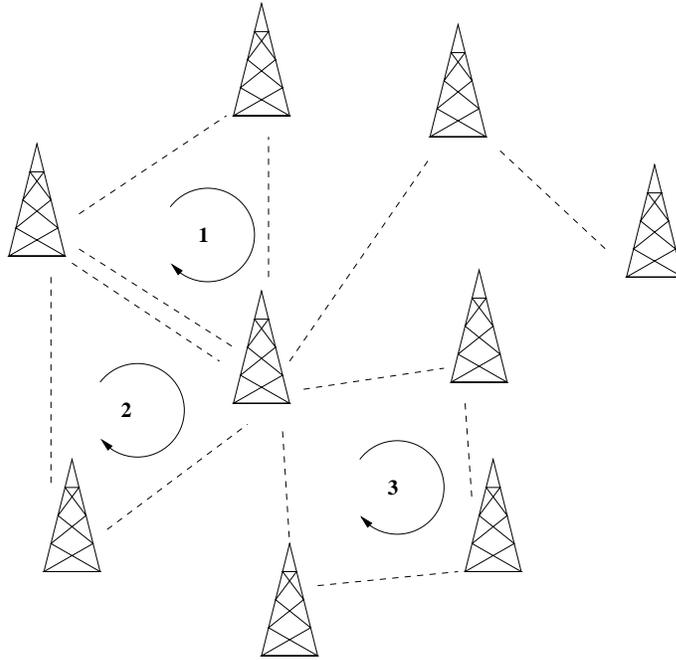


Figure 9.7: Le problème de la pseudo-synchronisation : il y a trois cycles dans ce réseau, la somme des RTD sur chacun d'eux est par définition nulle.

Les $m$ observations bruitées des différences sont représentées par le vecteur colonne $\mathbf{Y}$ :

$$\mathbf{Y} = \mathbf{M}\mathbf{X} + \mathbf{B} \tag{9.52}$$

$\mathbf{X}$ est un vecteur de dimension égale au nombre de liens observés, elle est donc inférieure ou égale à m. Les contraintes à prendre en compte sont le fait que les sommes des différences sont nulles sur tous les cycles :

$$\mathbf{C}_i^T \mathbf{X} = 0$$

où $\mathbf{C}_i$ est un vecteur colonne qui contient des 1, des -1, ou des 0 selon l'ordre du parcours du $i$-ième cycle.

Le problème à résoudre est celui d'une estimation linéaire sous contrainte :

$$\begin{cases} \mathbf{Y} = \mathbf{MX} + \mathbf{B} \\ \mathbf{C}^T\mathbf{X} = 0 \end{cases} \tag{9.53}$$

où $\mathbf{C} = [\mathbf{C}_1 \cdots \mathbf{C}_N]$ est de rang plein (cycles indépendants). La solution est donnée par :

$$\hat{\mathbf{X}} = \left[\mathbf{I} - \mathbf{KC}\left(\mathbf{C}^T\mathbf{KC}\right)^{-1}\mathbf{C}^T\right]\tilde{\mathbf{X}} \tag{9.54}$$

avec $\mathbf{K} = \left[\mathbf{M}^T\mathbf{Q}^{-1}\mathbf{M}\right]^{-1}$ où $\mathbf{Q}$ est la matrice de covariance du bruit $\mathbf{B}$.

### Synchronisation absolue

Chaque station de base corrige son temps d'émission en fonction de l'observation du temps d'arrivée de ses voisines :

$$t'_i = t_i + \alpha\frac{\sum_{j\neq i} p_{ij}(t_j - t_i)}{\sum_{j\neq i} p_{ij}} \tag{9.55}$$

$p_{ij}$ est un coefficient de pondération. En notant $\mathbf{X}_n = [t_1 \cdots t_N]^T$ l'état du réseau à l'étape $n$ et $\mathbf{P}$ la matrice fabriquée à partir des $\frac{p_{ij}}{\sum_{j\neq i} p_{ij}}$, on obtient :

$$\mathbf{X}_n = \mathbf{X}_{n-1} + \alpha(\mathbf{P} - \mathbf{I})\mathbf{X}_{n-1} = [(1-\alpha)\mathbf{I} + \alpha\mathbf{P}]\mathbf{X}_{n-1} = \mathbf{MX}_{n-1} \tag{9.56}$$

Soit :

$$\mathbf{X}_n = \mathbf{M}^n\mathbf{X}_0 \tag{9.57}$$

Notons que $\mathbf{P}$ n'est pas forcément symétrique (toutes les stations de base n'ont pas le même nombre de voisines). On peut démontrer que ce système converge en probabilité vers l'état synchrone :

$$\lim_{n\to\infty} \mathbf{M}^n v = \beta\mathbf{1}$$

où $v$ est initialisé aléatoirement et $\beta$ un facteur quelconque.

## 9.7 Simulations globales

Dans les sections précédentes, nous avons traité séparément le problème de l'estimation du temps d'arrivée et celui de la fusion de données. Nous nous proposons ici d'utiliser les deux étapes consécutivement dans des environnements concrets pour pouvoir en déduire les performances que l'on peut attendre d'un tel système de localisation.

## 9.7.1  Modèle de canal

Nous examinons ici la manière de générer une réponse impulsionnelle du canal
représentée par un étalement temporel, un nombre d'échos et pour chaque écho
son retard, sa puissance moyenne et la distribution de probabilité de son ampli-
tude.

On conviendra que chaque écho résulte de la superposition de plusieurs ondes
qui arrivent toutes au même moment [76, 77] :

$$E(t) = \sum_{i=1}^{n} E_i \exp\left[j(\phi_0 + \frac{2\pi}{\lambda}vt\cos\alpha_i)\right] \tag{9.58}$$

où $\alpha_i$ désigne l'angle d'incidence de la $i$-ième onde, $\phi_0$ désigne la phase initiale
et $v$ désigne la vitesse du mobile. L'amplitude résultante est donc la somme des
différentes contributions. Si l'écho correspondant n'est pas un trajet direct, les
phases de ces ondes sont uniformes, l'amplitude est alors gaussienne de moyenne
nulle et le module suit une loi de distribution de Rayleigh. En revanche, lorsqu'il
s'agit d'un trajet direct, les phases ne sont pas tout a fait uniformes et l'amplitude
résultante est gaussienne de moyenne non nulle.  Son module suit une loi de
distribution de Rice.  La corrélation dans le temps de l'amplitude est donnée
par :

$$R(\tau) = \int_0^{2\pi} p(\alpha) \exp\left[j\frac{2\pi}{\lambda}v\tau\cos\alpha\right] d\alpha \tag{9.59}$$

$p(\alpha)$ désigne la distribution de probabilité angulaire des ondes.  Si celles-ci
sont uniformes, on obtient alors le célèbre modèle de Jakes :

$$R(\tau) = J_0(\frac{2\pi}{\lambda}v\tau) \quad \Leftrightarrow \quad \bar{R}(f) = \frac{1}{\sqrt{1 - \left(\frac{f}{f_d}\right)^2}} \tag{9.60}$$

avec $\bar{R}(f)$ le spectre Doppler et $f_d = \frac{v}{\lambda}$ la fréquence Doppler.  Dans nos
simulations, $\alpha$ sera supposé gaussien de moyenne constante et d'écart-type 0.15
radians.  Ce modèle correspond au fait que les ondes arrivent concentrées dans
un faisceau étroit.  Dans tous les cas de figure, plus la vitesse est grande plus
l'amplitude tend à être moins corrélée dans le temps.  La génération de l'amplitude
sera faite selon la formule (9.58).  Si parmi les différentes ondes aucune n'est
déterministe, on obtient simplement une loi de distribution de Rayleigh.  En
revanche si une onde est déterministe, la moyenne statistique n'est pas nulle,
et on obtient une loi de distribution de Rice. Il est fréquent d'utiliser la loi de
distribution de Nakagami pour modéliser ces différences :

$$p(r) = \frac{2m^m r^{2m-1}}{\Gamma(m)\Omega^m} \exp(-\frac{m}{\Omega}r^2) \tag{9.61}$$

La variable $m$ reflète l'importance de la composante déterministe : plus $m$ est grand, plus cette composante est importante. $\Omega$ désigne la puissance moyenne de l'écho. Nous avons trois possibilités :

- $m = 1$ : distribution de Rayleigh.

- $m > 1$ : approximation de la loi de Rice.

- $m = \infty$ : distribution de Dirac, l'écho est purement déterministe.

La gérération des ondes s'effectue grâce aux deux relations suivantes [77] :

$$A_1 = \sqrt{\Omega\sqrt{1 - \frac{1}{m}}}$$

$$\sigma^2 = \frac{\Omega}{n}(1 - \sqrt{1 - \frac{1}{m}})$$

avec $A_1$ l'amplitude de la composante déterministe et $\sigma^2$ la puissance moyenne des autres ondes.

Plusieurs profils ont été sélectionnés :

- Urbain (en indoor à 3km/h et outdoor à 3 and 50 km/h) : zones denses, grands immeubles et centre ville.

- Banlieue outdoor (à 3 et 50 km/h) : zones résidentielles, banlieues et villages.

- Rural (à 3 et 100 km/h) : zones inhabitées, autoroutes, champs et forêts.

La différence majeure entre le profil urbain et les autres profils réside dans l'existence du trajet direct. Il n'y a pas de trajet direct dans les zones urbaines alors qu'un puissant trajet direct existe dans les autres profils.

Ces hypothèses sont basées sur le modèle de canal établi lors des réunions du comité de normalisation américain T1P1.5 [80].

## 9.7.2 Estimation du temps d'arrivée

Le modèle de Greenstein est choisi pour générer un étalement temporel et les paramètres des différentes constantes sont fixés pour chaque profil. La génération des puissances moyennes des différents échos est exponentiellement décroissante pour le profil urbain. Les valeurs des différents paramètres figurent dans la Table 9.1.

| Parameters | Urban | Suburban | Rural |
|---|---|---|---|
| $T_1$ ($\mu$s) | 0.4 | 0.3 | 0.1 |
| $\epsilon$ | 0.5 | 0.3 | 0.3 |
| $\sigma_y$ | 4 | 4 | 4 |
| Mobile speed (km/h) | 3/50 | 3/50 | 3/100 |
| LOS presence | no | yes | yes |
| Number of delays | 20 | 6 | 6 |
| Delays generation | $\{\tau_i\}_{1 \leq i \leq 20} \sim \mathcal{U}[0,1]$ | $\tau_1 = 0$ $\{\tau_i\}_{2 \leq i \leq 6} \sim \mathcal{U}[0,1]$ | $\tau_1 = 0$ $\{\tau_i\}_{2 \leq i \leq 6} \sim \mathcal{U}[0,1]$ |
| Power generation | $a_i \sim \mathcal{U}[0.5, 1.5]$. $\exp(-6\tau_i)$ | $a_1 = 4.3$ $\{a_i\}_{2 \leq i \leq 6} \sim \mathcal{U}[0.5, 1.5]$ | $a_1 = 4.3$ $\{a_i\}_{2 \leq i \leq 6} \sim \mathcal{U}[0.5, 1.5]$ |
| Nakagami $m$ parameter | $\{m_i\}_{1 \leq i \leq 20} = 1$ | $m_1 = 15$ $\{m_i\}_{2 \leq i \leq 6} \sim \mathcal{U}[1,5]$ | $m_1 = 15$ $\{m_i\}_{2 \leq i \leq 6} \sim \mathcal{U}[1,5]$ |
| Number of partial waves per delay | 100 | 100 | 100 |
| Initial phase $\phi_0$ | $\mathcal{U}[0, 2\pi]$ | $\mathcal{U}[0, 2\pi]$ | $\mathcal{U}[0, 2\pi]$ |
| $\alpha_j$ | $\mathcal{N}(0, 0.15)$ | $\mathcal{N}(0, 0.15)$ | $\mathcal{N}(0, 0.15)$ |

Table 9.1: Paramètres de simulations du temps d'arrivée pour les différents profils.

La difficulté rencontrée réside dans la détermination de la dimension du sous-espace signal. En théorie, ce nombre est égal au nombre d'échos. Cependant, il est parfois difficile de distinguer deux échos très proches l'un de l'autre et certains échos ont une énergie moyenne relativement faible. En conséquence, la dimension du sous-espace signal est en pratique inférieure au nombre d'échos. Pour l'estimer, le critère de description minimale (MDL) [17] est utilisé en lui imposant une limite maximale :

$$\hat{d} = \min_j (\hat{d}_{\mathrm{MDL}}, d_{\max}) \qquad (9.62)$$

La contrainte de limite maximale vient du fait que le critère MDL pourrait surestimer la dimension. En pratique, $d_{\max} = 4$.

$\hat{d}_{\mathrm{MDL}}$ est donné par :

$$\hat{d}_{\mathrm{MDL}} = \arg \min_{k=0\cdots p-1} L \log \left[ \frac{\left( \frac{1}{p-k} \sum_{i=k+1}^{p} \lambda_i \right)^{p-k}}{\prod_{i=k+1}^{p} \lambda_i} \right] + \frac{1}{2} k(2p-k) \log L$$

$\{\lambda_i\}_{1 \leq i \leq p}$ sont les valeurs propres de la matrice de covariance de l'estimation de la réponse impulsionnelle et $L$ le nombre d'observations (bursts). Les délais

sont supposés décorrélés entre eux. On utilisera Root MUSIC comme estimateur du temps d'arrivée. Dans la table 9.2, les résultats des simulations sont présentés à 10 dB. Les résultats sont présentés en terme de la racine de l'erreur quadratique moyenne des 90% meilleurs résultats. La raison à cela est que certains mauvais résultats dégradent cette erreur et il est possible de les corriger en prenant l'information de l'avance en temps (i.e. la position de la séquence d'apprentissage).

20 bursts de synchronisation sont utilisés. Les simulations montrent clairement que Root MUSIC est supérieur au simple filtrage adapté (Table 9.2).

| Profiles | Root MUSIC | | | Matched filter | | |
|---|---|---|---|---|---|---|
| **SNR** | 0 dB | 10 dB | $\infty$ dB | 0 dB | 10 dB | $\infty$ dB |
| Urban 3 km/h | 311 | 231 | 210 | 323 | 323 | 323 |
| Urban 50 km/h | 156 | 135 | 134 | 294 | 294 | 294 |
| Suburban 3 km/h | 26 | 20 | 19 | 76 | 74 | 74 |
| Suburban 50 km/h | 10 | 6 | 4 | 47 | 47 | 47 |
| Rural 3 km/h | 18 | 16 | 12 | 60 | 58 | 58 |
| Rural 100 km/h | 5 | 3 | 3 | 44 | 44 | 44 |

Table 9.2: Racine de l'erreur quadratique moyenne (en mètres) pour les 90% meilleurs résultats en utilisant 20 bursts de synchronisation.

Les résultats montrent que l'estimation du temps d'arrivée dépend légèrement du rapport signal sur bruit. Le problème majeur est bien le trajet multiple. Même dans un environnement sans bruit (rapport signal sur bruit infini) l'erreur n'est pas nulle.

### 9.7.3 Estimation de la position

Cette section présente les simulations globales dans un réseau hexagonal parfait avec des antennes omnidirectionnelles. La distance minimale entre les stations de base varie selon l'environnement. Nous avons choisi un motif de réutilisation de 7 stations de base. Nous avons pris le modèle de propagation d'Okumura-Hata. Les paramètres $D$, $L_p$, et l'écart-type de l'évanouissement ainsi que les autres paramètres sont donnés dans la Table 9.3. Les budgets de puissance montant et descendant sont supposés être équilibrés. Nous avons choisi le front montant. La sensibilité de la station de base détermine le niveau du bruit, elle est fixée à -118 dBm.

Par souci de simplification, les interférences sont supposées gaussiennes (approximation AWGN). La vitesse est prise en compte pour la génération des am-

| Parameters | Urban | Urban indoor | Suburban | Rural |
|---|---|---|---|---|
| $D$ (meters) | 1500 | 1500 | 4500 | 10000 |
| $L_p$ (dB) | 126 | 139.5 | 116 | 98 |
| Fading std. (dB) | 6 | 8.5 | 6 | 6 |
| Mobile speed (km/h) | 3/50 | 3 | 3/50 | 3/100 |
| Sensibility | -118 dBm | -118 dBm | -118 dBm | -118 dBm |

Table 9.3: Paramètres des simulations de la localisation pour les différents profils.

plitudes et n'intervient pas dans le calcul de la position, i.e. le mobile est supposé immobile pendant la durée des observations (à 50 km/h le mobile bouge de 14 mètres seulement). 20 bursts de synchronisation sont utilisés. Trois, quatre, et cinq stations de base sont impliquées dans chaque triangulation.

Les algorithmes utilisés sont :

- Root MUSIC pour le temps d'arrivée.

- Le développement de Taylor, initialisé par l'algorithme de Chan avec deux itérations, pour la triangulation. Les temps d'arrivée sont pondérés de la même manière.

Les simulations utilisent 500 mobiles générés aléatoirement au milieu du réseau et pour chacun, 20 simulations de Monte Carlo sont effectuées, soit un total de 10000 simulations de Monte Carlo. Les résultats sont exprimés en termes de :

- L'erreur maximale des meilleurs 67% résultats,

- Pourcentage des résultats ayant une erreur inférieure à 125 mètres,

- La racine de l'erreur quadratique moyenne des 90 % meilleurs résultats.

Deux séries de simulations sont effectuées dans les cas d'une synchronisation parfaite et imparfaite.

**Synchronisation parfaite**

Les résultats sont donnés dans la Table 9.4.

**Erreur de synchronisation**

Le temps d'émission relatif à chaque trame TDMA est perturbé par un bruit gaussien de moyenne nulle et d'écart-type 100 ns (ce qui correspond à 30 mètres de propagation environ). A titre indicatif, les résultats sont également donnés en terme de densité cumulée pour le profil urbain à 3 et 50 km/h (Figure 9.8).

| Profiles | 67 % in meters | | | Percentage < 125 m | | | RMS of 90% | | |
|---|---|---|---|---|---|---|---|---|---|
| | 3 BS | 4 BS | 5 BS | 3 BS | 4 BS | 5 BS | 3 BS | 4 BS | 5 BS |
| Urban out. 3 km/h | 104 | 93 | 93 | 72 | 76 | 78 | 115 | 84 | 80 |
| Urban out. 50 km/h | 95 | 83 | **84** | 74 | 79 | 82 | 104 | 76 | 72 |
| Urban in. 3 km/h | 108 | 94 | **97** | 71 | 75 | 77 | 119 | 86 | 82 |
| Suburban 3 km/h | 17 | 12 | 12 | 91 | 98 | 98 | 20 | 12 | 12 |
| Suburban 50 km/h | 12 | 8 | 8 | 93 | 100 | 100 | 12 | 7 | 7 |
| Rural 3 km/h | 6 | 5 | 5 | 90 | 100 | 100 | 10 | 4 | 4 |
| Rural 100 km/h | 4 | 3 | 3 | 91 | 100 | 100 | 5 | 3 | 3 |

Table 9.4: Résultats des simulations avec trois, quatre et cinq stations de base. Synchronisation parfaite.

| Profiles | 67 % in meters | | | Percentage < 125 m | | | RMS of 90% | | |
|---|---|---|---|---|---|---|---|---|---|
| | 3 BS | 4 BS | 5 BS | 3 BS | 4 BS | 5 BS | 3 BS | 4 BS | 5 BS |
| Urban out. 3 km/h | 116 | 99 | 99 | 69 | 75 | 76 | 124 | 90 | 85 |
| Urban out. 50 km/h | 106 | 91 | 89 | 72 | 78 | 80 | 111 | 81 | 77 |
| Urban in. 3 km/h | 120 | 103 | 100 | 69 | 74 | 76 | 130 | 92 | 86 |
| Suburban 3 km/h | 49 | 38 | 34 | 91 | 98 | 98 | 42 | 31 | 28 |
| Suburban 50 km/h | 46 | 35 | 31 | 93 | 100 | 100 | 39 | 28 | 25 |
| Rural 3 km/h | 47 | 34 | 30 | 90 | 100 | 100 | 42 | 28 | 25 |
| Rural 100 km/h | 45 | 33 | 29 | 90 | 100 | 100 | 39 | 27 | 24 |

Table 9.5: Résultats des simulations avec trois, quatre et cinq stations de base. Les erreurs de synchronisation sont incluses.

## 9.8 Conclusions et directions futures

Dans cette thèse, nous avons présenté un système de simulation global pour localiser un mobile dans un réseau GSM. Ce système utilise les propres possibilités du réseau sans faire intervenir un quelconque moyen extérieur.

Nous nous sommes basés sur l'estimation du temps d'arrivée dans ses deux versions front montant ou front descendant. Le problème majeur est le trajet multiple. A partir de l'estimateur du maximum de vraisemblance, des algorithmes basés sur l'extraction des sous-espaces signal ou bruit ont été introduits pour l'estimation de l'ensemble des retards des différents échos. Le temps d'arrivée est défini comme étant le minimum de ces retards. Une extension de ces algorithmes est présentée lorsque l'impulsion de modulation est inconnue.

(a)



(b)
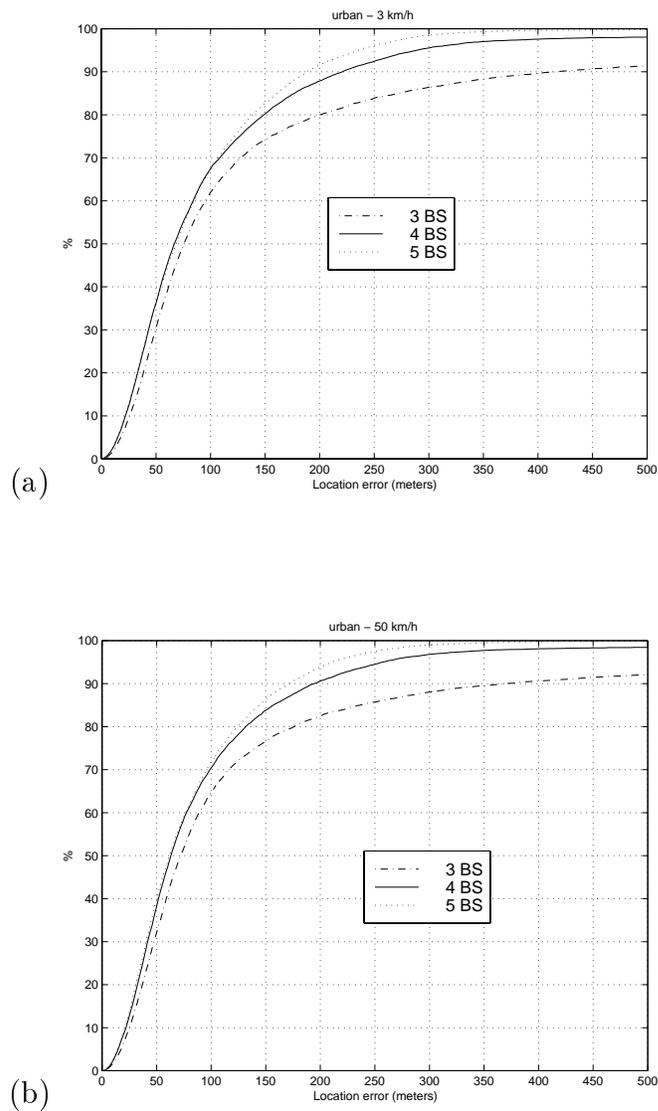
Figure 9.8: CDF pour le profil urbain : (a) à 3 km/h, (b) à 50 km/h.

La fusion des données est ensuite analysée et aboutit à une triangulation hyperbolique. Le développement en série de Taylor semble offrir de bons résultats lorsqu'il est bien initialisé.

La synchronisation est ensuite abordée dans ses deux versions pseudo et absolue. Enfin, des simulations globales dans des environnements réalistes sont

présentées. Les résultats semblent encourageants.

Les sujets non abordés ou à compléter sont les suivants :

### Echos diffus

Nous n'avons considéré que les échos discrets ayant un retard bien défini. Une extension possible serait le cas des échos diffus où les ondes arrivent autour d'une valeur moyenne. Le modèle de la réponse impulsionnelle du canal est modifié :

$$h_j(t) = \sum_{i=1}^{d} \sum_{k=1}^{K_i} s_{ijk} a(t - \tau_i - \tilde{\tau}_{ik}) \qquad 1 \le j \le L \tag{9.63}$$

Deux approches sont possibles :

- Développement en série de Taylor [82,83].

- Estimation conjointe du retard moyen et de l'écart-type autour de la moyenne. [84–86].

### Borne de Cramer-Rao dans le cas d'un impulsion inconnue

Nous n'avons pas calculé cette borne. Sa valeur pourrait être utile pour évaluer les résultats des algorithmes qui estiment conjointement l'impulsion et les retards.

### Pondération optimale des temps d'arrivée

Nous n'avons pas pondéré les temps d'arrivée lors de la triangulation. Il est en effet difficile de trouver de telles valeurs. On pourrait éventuellement se contenter de valeurs empiriques en fonction du rapport signal-sur-bruit ou de l'étalement temporel.

### Systèmes CDMA - UMTS

La localisation est requise pour les futurs standards. Ces standards, comme l'UMTS (Universal Mobile Telecommunications System), seront basés sur la technologie à répartition en codes (CDMA).

# Appendix A

# GMSK pulse

A phase modulation has the following general reduced expression:

$$s(t) = e^{j \sum_k b_k \phi(t - kT)} \tag{A.1}$$

$T$ refers to the bit period. $\phi(t)$ is the modulation function, for a modulation index of 0.5, it can be written as follows:

$$\phi(t) = \begin{cases} 0 & \text{if } t < 0 \\ \frac{\pi}{2} & \text{if } t \geq LT \end{cases} \tag{A.2}$$

$L$ is the called the memory of the modulation. In practice, $\phi(t)$ is obtained by integrating a known signal $s_e(t)$:

$$\phi(t) = \frac{\pi}{2} \int_{-\infty}^{t - LT} s_e(\tau) d\tau \tag{A.3}$$

The GMSK modulation specified in [1] is given by the following function:

$$s_e(t) = \text{rect}\left[\frac{t}{T}\right] \star \frac{1}{\sqrt{2\pi}\sigma T} e^{-\frac{t^2}{2\sigma^2 T^2}} \tag{A.4}$$

with $\sigma = \frac{\sqrt{\ln 2}}{2\pi BT}$ et $BT = 0.3$.

In reality $L$ is very high, but $L = 3$ is a good approximation for the GMSK as shown in Figure A.1. Referring to [6], a phase modulation can be decomposed into amplitude modulated pulses (AMP). Assuming differential coding, i.e. $b_i = d_i d_{i-1}$, and taking the first component of the decomposition called main pulse, $s(t)$ can be rewritten in the following way:

$$s(t) = \sum_k d_k a(t - kT) \tag{A.5}$$
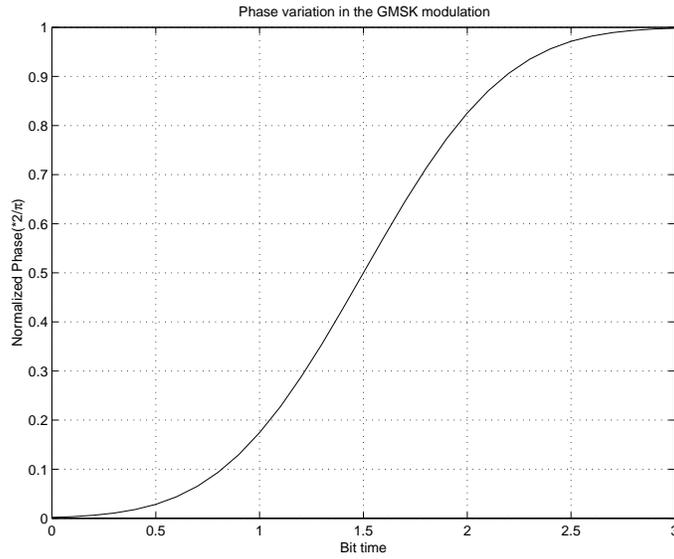
Figure A.1: The modulation function in the GMSK.

with,

$$a(t) = e^{-j\frac{\pi t}{2T}} \prod_{i=0}^{L-1} \sin[\phi(t + iT)] \cos[\phi(t + (i - L)T)] \tag{A.6}$$

Other components appear to be negligible; the second component is given by:

$$b(t) = \frac{a(t)a(t + 2T)}{a(t + T)}$$

which has an energy 24.35 dB lower than the first pulse. The main pulse $a(t)$ and its secondary pulse are shown in Figure A.2.

$a(t)$ represents the main pulse. In the GMSK modulation, this pulse concentrates about 99.6% of the modulation power. In other words, it is a good approximation to state that the GMSK modulation is linear. The GMSK pulse has a length of four bit periods, but since it is filtered to match the 200 kHz bandwidth, its length may exceed eight bit periods.

Figure A.3 shows the main pulse normalized spectrum.
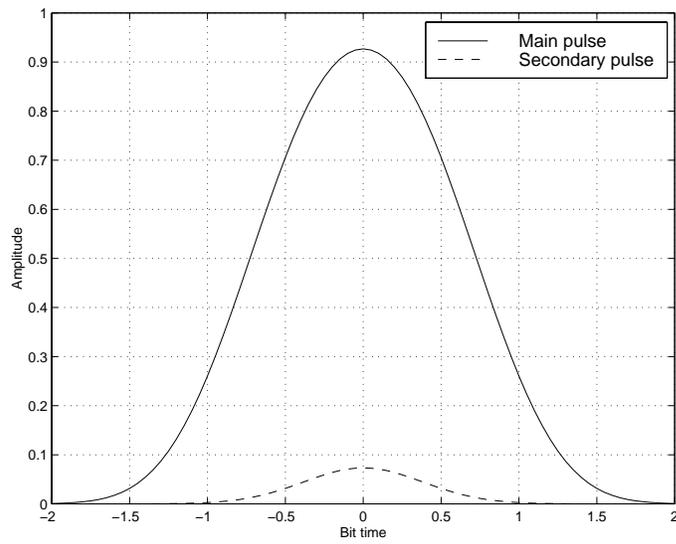
Figure A.2: GMSK main and secondary pulses.
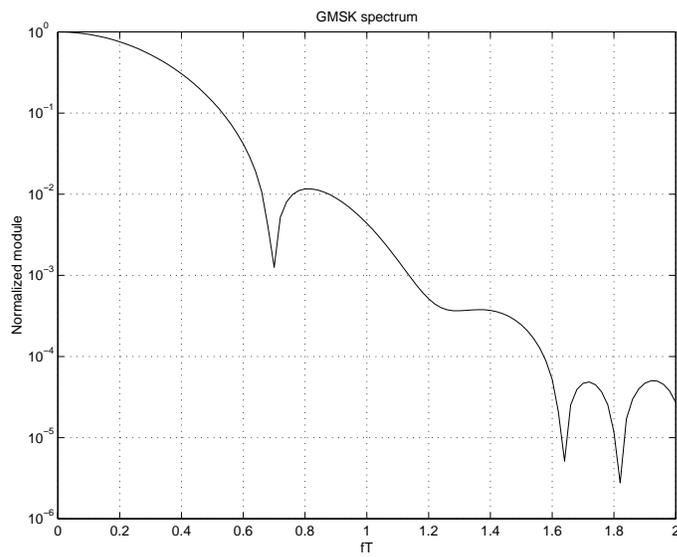


Figure A.3: GMSK spectrum.

# Appendix B

# Proof for the variance formula

$$
\begin{aligned}
\hat{\sigma}_r^2(\tau) &= \frac{\text{Tr}\left(\mathbf{\Pi}_{\mathbf{D}(\mathbf{d})\mathbf{A}(\tau)}^{\perp}\hat{\mathbf{K}}_y\right)}{m(m-d)} \\
&= \frac{\text{Tr}\left(\mathbf{\Pi}_{\mathbf{A}}^{\perp}(\tau)\hat{\mathbf{K}}_{\hat{h}}\right)}{p-d} + O(1/\sqrt{L}) \\
&= \frac{\text{Tr}\left(\mathbf{\Pi}_{\mathbf{D}}^{\perp}(\mathbf{d})\hat{\mathbf{K}}_y\right)}{m(m-p)} + O(1/\sqrt{L})
\end{aligned}
$$

Asymptotically, we have the following relations:

$$
\lim_{L\to\infty} \hat{\mathbf{K}}_y = \mathbf{K}_y = \mathbf{D}(\mathbf{d})\mathbf{A}(\theta)\mathbf{K}_s\left[\mathbf{D}(\mathbf{d})\mathbf{A}(\theta)\right]^* + \sigma^2\mathbf{I}
$$

$$
\lim_{L\to\infty} \hat{\mathbf{K}}_{\hat{h}} = \mathbf{K}_{\hat{h}} = \mathbf{A}(\theta)\mathbf{K}_s\mathbf{A}^*(\theta) + \sigma_r^2\mathbf{I}
$$

The result is straightforward by noticing that for any full column rank $m \times d$ matrix $\mathbf{A}$:

$$
\mathbf{\Pi}_{\mathbf{A}}^{\perp}\mathbf{A} = 0 \qquad \text{Tr}(\mathbf{\Pi}_{\mathbf{A}}^{\perp}) = m - d
$$

For example:

$$
\text{Tr}\left(\mathbf{\Pi}_{\mathbf{A}}^{\perp}(\tau)\mathbf{K}_{\hat{h}}\right) = \sigma_r^2\text{Tr}\left(\mathbf{\Pi}_{\mathbf{A}}^{\perp}(\tau)\right) = (m-d)\sigma_r^2
$$

# Appendix C

# Derivation of the Cramer-Rao bound

The Cramer-Rao bound is given by the inverse of the Fisher information matrix (4.31). The log-likelihood is given by:

$$\log p(\mathbf{F}|\theta, t_{\mathrm{r}}) = \frac{1}{2} \left[\mathbf{F} - \mathbf{A}(\theta) - t_{\mathrm{r}}\mathbf{1}\right]^T \mathbf{Q}^{-1} \left[\mathbf{F} - \mathbf{A}(\theta) - t_{\mathrm{r}}\mathbf{1}\right]$$

And the following results are straightforward:

- $\mathrm{E}\left[\frac{\partial p(\mathbf{F}|\theta, t_{\mathrm{r}})}{\partial \theta} \frac{\partial p(\mathbf{F}|\theta, t_{\mathrm{r}})}{\partial \theta^T}\right] = \left(\frac{\partial \mathbf{A}(\theta)}{\partial \theta}\right)^T \mathbf{Q} \frac{\partial \mathbf{A}(\theta)}{\partial \theta}$

- $\mathrm{E}\left[\frac{\partial p(\mathbf{F}|\theta, t_{\mathrm{r}})}{\partial t_{\mathrm{r}}} \frac{\partial p(\mathbf{F}|\theta, t_{\mathrm{r}})}{\partial t_{\mathrm{r}}}\right] = \mathbf{1}^T \mathbf{Q} \mathbf{1}$

- $\mathrm{E}\left[\frac{\partial p(\mathbf{F}|\theta, t_{\mathrm{r}})}{\partial \theta} \frac{\partial p(\mathbf{F}|\theta, t_{\mathrm{r}})}{\partial t_{\mathrm{r}}}\right] = -\left(\frac{\partial \mathbf{A}(\theta)}{\partial \theta}\right)^T \mathbf{Q} \mathbf{1}$

The Fisher information matrix is:

$$\mathbf{FIM}(\theta, t_{\mathrm{r}}) = \begin{bmatrix} \left(\frac{\partial \mathbf{A}(\theta)}{\partial \theta}\right)^T \mathbf{Q} \frac{\partial \mathbf{A}(\theta)}{\partial \theta} & -\left(\frac{\partial \mathbf{A}(\theta)}{\partial \theta}\right)^T \mathbf{Q} \mathbf{1} \\ -\mathbf{1}^T \mathbf{Q} \frac{\partial \mathbf{A}(\theta)}{\partial \theta} & \mathbf{1}^T \mathbf{Q} \mathbf{1} \end{bmatrix}$$

The Cramer-Rao bound on $\theta$ is given by the upper left side of $\mathbf{FIM}^{-1}$. Using block matrix inversion we obtain:

$$\mathbf{CRB}^{-1}(\theta) = \left[\frac{\partial \mathbf{A}(\theta)}{\partial \theta}\right]^T \tilde{\mathbf{Q}} \frac{\partial \mathbf{A}(\theta)}{\partial \theta}$$

with:

$$\tilde{\mathbf{Q}} = \mathbf{Q}^{-1} - \frac{\mathbf{Q}^{-1}\mathbf{1}\mathbf{1}^T\mathbf{Q}^{-1}}{\mathbf{1}^T\mathbf{Q}^{-1}\mathbf{1}}$$

# Appendix D

# Mathematical notations

| | |
|---|---|
| $\mathbf{A}^\dagger = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$ | Moore-Penrose pseudo-inverse of $\mathbf{A}$ |
| $\mathbf{A} \geq \mathbf{B}$ | matrix comparison $\Leftrightarrow (\forall x, x^*\mathbf{A}x \geq x^*\mathbf{B}x)$ |
| $\mathbf{A}^T$ | transpose of $\mathbf{A}$ |
| $\mathbf{A}^c$ | conjugate of $\mathbf{A}$ |
| $\mathbf{A}^*$ | transpose conjugate of $\mathbf{A}$ |
| $\mathbf{\Pi_A} = \mathbf{AA}^\dagger$ | projection onto the range space of $\mathbf{A}$ |
| $\mathbf{\Pi_A^\perp} = \mathbf{I} - \mathbf{\Pi_A}$ | projection onto the null space of $\mathbf{A}$ |
| $\mathbf{A} \odot \mathbf{B}$ | Schur product of $\mathbf{A}$ and $\mathbf{B}$, i.e. the element by element product |
| $\star$ | convolution operator |
| $\mathbf{A} \otimes \mathbf{B}$ | Kronecker product of $\mathbf{A}$ and $\mathbf{B}$ |
| $\mathrm{Tr}(\mathbf{A})$ | trace of $\mathbf{A}$ |
| $\|\mathbf{A}\|^2 = \mathrm{Tr}(\mathbf{A}^*\mathbf{A})$ | Frobenius norm of $\mathbf{A}$ |
| $|\mathbf{A}|$ | determinant of $\mathbf{A}$ |
| $\mathrm{Vec}(\mathbf{A})$ | the column vector made by stacking the columns of $\mathbf{A}$ |
| $\mathrm{Re}(.)$ | real part |
| $\mathrm{Im}(.)$ | imaginary part |
| $\mathrm{E}[.]$ | mathematical expectation |
| $\delta_{ij}$ | Kronecker symbol, 1 if $i = j$, 0 otherwise |
| $\delta(t)$ | Dirac delta fucntion |
| $\mathbf{I}$ | identity matrix with the appropriate dimension |
| $\mathbf{J}$ | anti-identity matrix with the appropriate dimension |
| $j^2 = -1$ | |
| $I_0$ and $J_0$ | the modified and normal Bessel function of order 0 |
| $\Gamma$ | the gamma function |
| $\mathcal{N}(m, \sigma^2)$ | normal variable with mean $m$ and variance $\sigma^2$ |
| $\mathcal{U}[a, b]$ | uniform variable on $[a, b]$ |

# Appendix E

# Abbreviations

AMP         Amplitude Modulated Pulse
AWGN        Additive White Gaussian Noise
BCCH        Broadcast Common CHannel
BER         Bit Error Rate
BSIC        Base Station Identity Code
BSC         Base Station Controller
BSS         Base Station Subsystem
BS          Base Station
CDMA        Code Division Multiple Access
CRB         Cramer-Rao Bound
DoA         Direction of Arrival
DToA        Difference Times of Arrival
DTX         Discontinuous Transmission
ETSI        European Telecommunications Standards Institute
ESPRIT      Estimation of Signal Parameters via Rotational Invariance Techniques
FCCH        Frequency Correction Channel
FDD         Frequency Division Duplex
FDMA        Frequency Division Multiple Access
FFT         Fast Fourier Transform
FH          Frequency Hopping
FN          Frame Number
GDoP        Geometric Dilution of Precision
GMSK        Gaussian Minimum Shift Keying
GSM         Global System for Mobile communications
HLR         Home Location Register
HO          HandOver
IQML        Iterative Quadratic Maximum Likelihood
ISI         Inter Symbol Interference
JADE        Joint Angle and Delay Estimation
LAC         Location Area Code

| | |
|---|---|
| LMU | Location Mobile Unit |
| MLC | Mobile Location Center |
| MSC | Mobile services Switching Center |
| MSK | Minimum Shift Keying |
| MUSIC | MUltiple SIgnal Characterization |
| PLMN | Public Land Mobile Network |
| PSTN | Public Switched Telephone Network |
| RACH | Random Access CHannel |
| RTD | Real Times Difference |
| SCH | Synchronization CHannel |
| SIM | Subscriber Identity Member |
| SMS | Short Message Service |
| TA | Timing Advance |
| TDMA | Time Division Multiple Access |
| ToA | Time of Arrival |
| UMTS | Universal Mobile Telecommunications System |
| ULA | Uniform Linear Array |
| VLR | Visitor Location Register |
| WSF | Weight Subspace Fitting |

# Bibliography

[1] European Telecommunication Standards Institute, "Digital cellular telecommunication system (phase 2+)," 1997.

[2] M. Hata, "Empirical formula for propagation loss in land mobile radio service," *IEEE Trans. Vehicular Technology*, vol. 29, no. 3, pp. 317–325, August 1980.

[3] L. Yuanqing, "A theoretical formulation for the distribution density of multipath delay spread in a land mobile radio environment," *IEEE Trans. Vehicular Technology*, vol. 43, no. 2, pp. 379–388, May 1994.

[4] L. J. Greenstein, V. Erceg, Y. S. Yeh, and M. V. Clark, "A new path-gain/delay-spread propagation model for digital cellular channels," *IEEE Trans. Vehicular Technology*, vol. 46, no. 2, pp. 477–485, May 1997.

[5] M. Mouly and M.-B. Pautet, *The GSM system for mobile communications*. Europe Media, 1992.

[6] P. A. Laurent, "Exact and approximate construction of digital phase modulations by superposition of amplitude modulated pulses (AMP)," *IEEE Trans. Communications*, vol. 34, no. 2, pp. 150–160, February 1986.

[7] G. J. Bierman, *Factorization methods for discrete sequential estimation*. Academic Press, 1977.

[8] G. D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Information Theory*, vol. 18, no. 3, pp. 363–378, May 1972.

[9] M. Hellebrandt, R. Mathar, and M. Scheibenbogen, "Estimating position and velocity of mobiles in a cellular radio network," *IEEE Trans. Vehicular Technology*, vol. 46, no. 1, pp. 65–71, February 1997.

[10] L. Breiman, *Classification and regression trees*. Monterey, California: Wadsworth, 1984.

[11] H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Signal Processing Magazine*, pp. 67–94, July 1996.

[12] M. Wax and A. Leshem, "Joint estimation of time delays and directions of arrival of multiple reflections of a known signal," *IEEE Trans. Signal Processing*, vol. 45, no. 10, pp. 2477–2484, October 1997.

[13] A.-J. Van Der Veen, M. C. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift-invariance techniques," *IEEE Trans. Signal Processing*, vol. 46, no. 2, pp. 405–418, February 1998.

[14] G. G. Raleigh and T. Boros, "Joint space-time parameter estimation for wireless communication channels," *IEEE Trans. Signal Processing*, vol. 46, no. 5, pp. 1333–1343, May 1998.

[15] A. Zeira and P. M. Schultheiss, "Time delay estimation for closely spaced echoes," *Proc. ICASSP*, pp. 2763–2766, 1990.

[16] M. Wax, T.-J. Shan, and T. Kailath, "Spatio-temporal spectral analysis by eigenstructure methods," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 8, pp. 1190–1196, August 1984.

[17] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, no. 2, pp. 387–392, April 1985.

[18] Y. Q. Yin and P. R. Krishnaiah, "On some nonparametric methods for detection of the number of signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 11, pp. 1533–1538, November 1987.

[19] M. Wax and I. Ziskind, "Detection of the number of coherent signals by the MDL principle," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 8, pp. 1190–1196, August 1989.

[20] K. M. Wong, Q.-T. Zhang, J. P. Reilly, and P. C. Yip, "On information theoretic criteria for determining the number of signals in high resolution array processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 11, pp. 1959–1971, November 1990.

[21] H. Lee and F. Li, "An eigenvector technique for detecting the number of emitters in a cluster," *IEEE Trans. Signal Processing*, vol. 42, no. 9, pp. 2380–2388, September 1994.

[22] G. Xu, R. H. Roy, and T. Kailath, "Detection of number of sources via exploitation of centro-symmetry property," *IEEE Trans. Signal Processing*, vol. 42, no. 1, pp. 102–112, January 1994.

[23] A. Bruckstein, T.-J. Shan, and T. Kailath, "The resolution of overlapping echos," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, no. 6, pp. 1357–1367, December 1985.

[24] M. C. Vanderveen, A.-J. Van Der Veen, and A. Paulraj, "Estimation of multipath parameters in wireless communications," *IEEE Trans. Signal Processing*, vol. 46, no. 3, pp. 682–690, March 1998.

[25] B. Göransson and B. Ottersten, "Efficient direction estimation of uncorrelated sources using polynomial rooting," *Proceedings DSP 97*, 1997.

[26] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 10, pp. 1783–1795, October 1990.

[27] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 5, pp. 720–741, May 1989.

[28] J. F. B. Böhme, "Estimation of spectral parameters of correlated signals in wavefields," *Signal Processing*, vol. 11, pp. 329–337, 1986.

[29] Y. Bresler and A. Macovski, "Maximum likelihood estimation of linearly structured covariance with application to antenna array processing," *Proc. ICASSP*, pp. 172–175, 1986.

[30] H. L. Van Trees, *Detection, estimation and modulation theory, part I and III*. New York: Wiley, 1971.

[31] D. Starer and A. Nehorai, "Newton algorithms for conditional and unconditional maximum likelihood estimation of the parameters of exponential signals in noise," *IEEE Trans. Signal Processing*, vol. 40, no. 6, pp. 1528–1534, June 1992.

[32] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 10, pp. 1553–1560, October 1988.

[33] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the em algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 4, pp. 477–489, April 1988.

[34] M. I. Miller and D. R. Fuhrmann, "Maximum-Likelihood narrow-band direction finding and the EM algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 9, pp. 1560–1577, September 1990.

[35] M. Segal, E. Weinstein, and B. R. Musicus, "Estimate-Maximize algorithms for multichannel time delay and signal estimation," *IEEE Trans. Signal Processing*, vol. 39, no. 1, pp. 1–16, January 1991.

[36] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 2408–2418, 1969.

[37] B. Friedlander, "A sensitivity analysis of the MUSIC algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 10, pp. 1740–1751, October 1990.

[38] P. Stoica and A. Nehorai, "Music, maximum likelihood, and Cramér-Rao bound: further results and comparisons," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 12, pp. 2140–2150, December 1990.

[39] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, no. 5, pp. 1110–1121, May 1991.

[40] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, no. 11, pp. 2436–2249, November 1991.

[41] B. Ottersten, M. Viberg, and T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," *IEEE Trans. Signal Processing*, vol. 40, no. 3, pp. 590–600, March 1992.

[42] B. Friedlander and A. Weiss, "Eigenstructure methods for direction finding with sensor gain and phase uncertainties," *Proc. ICASSP*, pp. 2681–2684, 1988.

[43] V. C. Soon, L. Tong, H. Y. F., and L. R., "A subspace method for estimating sensor gains and phases," *IEEE Trans. Signal Processing*, vol. 42, no. 4, pp. 1973–1976, April 1994.

[44] V. F. Pisarenko, "The retrieval of harmonics from covariance functions," *Geophys. J. Roy. Astronom. Soc.*, pp. 347–366, 1972.

[45] B. D. Rao and K. V. Hari, "Performance analysis of root-Music," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 12, pp. 1939–1949, December 1989.

[46] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 5, pp. 1081–1089, October 1986.

[47] P. Stoica and K. C. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 7, pp. 1132–1143, July 1990.

[48] P. Stoica, J. Li, and T. Söderström, "On the inconsistency of IQML," *Signal Processing*, vol. 56, no. 2, pp. 185–190, January 1997.

[49] M. Kristensson, M. Jansson, and B. Ottersten, "Modified IQML and weighted subspace fitting without eigendecomposition," *to appear in Signal Processing*, 1999.

[50] R. Roy and T. Kailath, "ESPRIT - Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 7, pp. 984–995, July 1989.

[51] B. Ottersten, M. Viberg, and T. Kailath, "Performance analysis of the total least squares ESPRIT algorithm," *IEEE Trans. Signal Processing*, vol. 39, no. 5, pp. 1122–1135, May 1991.

[52] A. Eriksson and P. Stoica, "Optimally weighted ESPRIT for direction estimation," *Signal Processing*, vol. 38, pp. 223–229, 1994.

[53] A. L. Swindlehurst, B. Ottersten, R. Roy, and T. Kailath, "Multiple invariance ESPRIT," *IEEE Trans. Signal Processing*, vol. 40, no. 4, pp. 867–881, April 1992.

[54] R. Bachl, "The forward-backward averaging technique applied to TLS-ESPRIT processing," *IEEE Trans. Signal Processing*, vol. 43, no. 11, pp. 2691–2699, November 1995.

[55] M. Jansson, *On subspace methods in system identification and sensor array signal processing*. PhD thesis, Royal institute of Technology, Department of Signals, Sensors and Systems, Stockholm, Sweden, 1997.

[56] A. L. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in the presence of model errors, part I: the MUSIC algorithm," *IEEE Trans. Signal Processing*, vol. 40, no. 7, pp. 1758–1774, July 1992.

[57] A. L. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in the presence of model errors, part II: multidimensional algorithms," *IEEE Trans. Signal Processing*, vol. 41, no. 9, pp. 2882–2890, September 1993.

[58] M. Jansson, A. L. Swindlehurst, and B. Ottersten, "Weighted subspace fitting for general array error models," *IEEE Trans. Signal Processing*, vol. 46, no. 9, pp. 2484–2498, September 1998.

[59] A. Paulraj and T. Kailath, "Direction of arrival estimation by eigenstructure methods with unknown sensor gains and phases," *Proc. ICASSP*, pp. 640–643, 1985.

[60] D. Asztély, *Spatial and spatio-temporal processing with antenna arrays in wireless systems*. PhD thesis, Royal institute of Technology, Department of Signals, Sensors and Systems, Stockholm, Sweden, 1999.

[61] G. G. H. and V. Loan, *Matrix computations*. Baltimore: Johns Hopkins University Press, 1990.

[62] G. W. Stewart, "An updating algorithm for subspace tracking," *IEEE Trans. Signal Processing*, vol. 40, no. 6, pp. 1535–1541, June 1992.

[63] G. Xu, R. H. Roy, and T. Kailath, "Fast subspace decomposition," *IEEE Trans. Signal Processing*, vol. 42, no. 3, pp. 539–551, March 1994.

[64] S. Marcos, *Les méthodes à haute résolution*. Paris: Hermès, 1998.

[65] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Processing*, vol. 43, no. 1, pp. 95–107, January 1995.

[66] E. D. Kaplan, *Understanding GPS, principles and applications*. Boston: Artech House, 1996.

[67] D. J. Torrieri, "Statistical theory of passive location systems," *IEEE Trans. Aeros. Electronic Systems*, vol. 20, pp. 183–198, March 1984.

[68] W. H. Foy, "Position-location solutions by Taylor-series estimation," *IEEE Trans. Aeros. Electronic Systems*, vol. 12, no. 2, pp. 187–194, March 1976.

[69] M. Wax and T. Kailath, "Optimum localization of multiple sources by passive arrays," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, no. 5, pp. 1210–1218, October 1983.

[70] Y. T. Chan and K. C. Ho, "A simple and efficient estimator for hyperbolic location," *IEEE Trans. Signal Processing*, vol. 42, no. 8, pp. 1905–1915, August 1994.

[71] L. Le Bris, *Allocation de resources radio*. PhD thesis, Université Paris VII, UFR d'informatique, Paris, France, 1999.

[72] M. Gondran and M. Minoux, *Graphes et algorithmes*. Eyrolles, 1995.

[73] F. R. Gantmacher, *Théorie des matrices, tomes I et II*. Dunod, 1966.

[74] I. S. Gradshteyn and I. M. Ryzhik, *Tables of integrals, series, and products*. Academic Press, 1979.

[75] J. C.-I. Chuang, "Autonomous time synchronization among radio ports in wireless personal communications," *IEEE Trans. Vehicular Technology*, vol. 43, no. 1, pp. 27–32, February 1994.

[76] W. R. Braun, "A physical mobile radio channel model," *IEEE Trans. Vehicular Technology*, vol. 40, no. 2, pp. 472–482, May 1991.

[77] U. Dersch and R. J. Rüegg, "Simulations of the time and frequency selective outdoor mobile radio channel," *IEEE Trans. Vehicular Technology*, vol. 42, no. 3, pp. 338–344, 1993.

[78] M. Nakagami, *The m distribution– a general formula of intensity distribution of rapid fading*, pp. 3–36. New York: Pergamon Press, 1958.

[79] S. Aun Abbas and A. U. Sheikh, "A geometric theory of Nakagami fading multipath mobile radio channel with physical interpretations," *Proc. VTC*, pp. 637–641, 1996.

[80] Ericsson, "Evaluation of positioning measurement systems," *T1P1.5/98-110, http://www.t1.org*, 1998.

[81] X. Lagrange, P. Godlewski, and S. Tabbane, *Réseaux GSM-DCS*. Paris: Hermès, 1995.

[82] D. Asztély and B. Ottersten, "The effects of local scaterring on direction of arrival estimation with MUSIC and ESPRIT," *Proc. ICASSP*, 1998.

[83] D. Asztély and B. Ottersten, "A generalized array manifold model for local scaterring in wireless communications," *Proc. ICASSP*, 1997.

[84] T. Trump and B. Ottersten, "Estimation of nominal direction of arrival and angular spread using an array of sensors," *Signal Processing*, vol. 50, 1996.

[85] Y. Meng, P. Stoica, and K. M. Wong, "Estimation of the directions of arrival of spatially dispersed signals in array processing," *IEE Proc.-Radar, Sonar Navig.*, vol. 143, no. 1, pp. 1–9, February 1996.

[86] M. Bengtsson and B. Ottersten, "Rooting techniques for estimation of angular spread with an antenna array," *Proc. VTC*, 1996.