Probabilistic Techniques for Equalization of the Mobile Radio Channel in the Presence of Co-Channel Interference

Carlo Luschi

A thesis submitted for the degree of Doctor of Philosophy.
The University of Edinburgh.
February 2002

Abstract

An important issue for present and future wireless cellular systems is the capability of delivering differentiated quality of service through the nonstationary and dispersive wireless channel. Therefore, in conjunction with countermeasures against multipath distortion, powerful interleaving and forward error correction coding schemes are usually employed to obtain adequate error performance. In wireless receivers employing the concatenation of an equalizer and a channel decoder, the overall performance can be improved by *soft-output equalization*, which provides the decoder with data reliability information to be used in *soft-decision decoding*. The theory of soft value processing has been well established for the case of the Gaussian dispersive channel. However, co-channel and adjacent-channel interference are often one of the dominant impairments in wireless networks, and one of the major factors that limit the capacity of cellular systems. This thesis studies the problem of soft-output equalization of the mobile radio channel in interference-limited environments, where it is often difficult to obtain an accurate statistical model of the (non-Gaussian) disturbance.

The first part of the thesis proposes a new technique for single-channel MAP trellis equalization in the presence of multipath and non-Gaussian interference. The approach is based on the *non-parametric* estimation of the density function of the overall disturbance by means of *kernel smoothing*. The work considers the problem of density estimation with limited volume of data, and addresses the use of a whitening filter in the presence of coloured interference. As an application, simulation results are provided for the GSM system, showing a significant performance improvement with respect to the trellis equalizer based on the Gaussian assumption.

The second part of the thesis considers the case of an antenna array receiver, and studies a simple method to derive the reliability information at the output of a deterministic decision-feedback *least-squares* space-time equalizer. Computer simulations for the Enhanced Data Rates for GSM Evolution (EDGE)/Enhanced General Packet Radio Service (EGPRS) system show that the receiver performance can be significantly improved by a soft-output calculation based on short-term statistics of the equalizer output error. The thesis also addresses the additional use of soft-decision feedback, which provides further robustness to the proposed soft-output equalizer.

The study shows the relevance of probabilistic processing for robust equalization of the wireless channel in the presence of non-Gaussian interference, and emphasizes the advantages of strategies that do not rely on a statistical model of the disturbance.

Declaration of originality

I hereby declare that the research recorded in this thesis and the thesis itself was composed and originated entirely by myself in the Department of Electronics and Electrical Engineering at The University of Edinburgh and in the Global Wireless Systems Research Department at Bell Laboratories, Lucent Technologies. Some of the simulation results presented in this thesis was obtained by using GSM and EDGE/EGPRS Matlab simulators developed at Bell Laboratories, Lucent Technologies, with the contribution also of Magnus Sandell, Paul Strauch and Alexandr M. Kuzminkiy.

(signed)

Carlo Luschi

Acknowledgements

I would like to thank my supervisor, Prof. Bernard Mulgrew, for his always valuable comments and suggestions, and for his constant encouragement and understanding of my situation of part-time student.

I am also indebted to my manager at Bell Labs Research, Dr. Ran Yan, who have proposed and sponsored my part-time postgraduates studies at the University of Edinburgh.

Finally, I would like to aknowledge useful discussions with my Bell Labs colleagues Alexandr M. Kuzminskiy and Paul Strauch.

Contents

List of figures

Acronyms and abbreviations

ACI	adjacent channel interference
AMC	adaptive modulation and coding
APP	*a posteriori* probability
ARQ	automatic repeat request
AWGN	additive white Gaussian noise
BER	bit-error rate
BLER	block-error rate
CCI	co-channel interference
CM	constant modulus
DFE	decision-feedback equalizer
DF-LS	decision-feedback least-squares
ECSD	enhanced circuit switched data
EDGE	enhanced data rates for GSM evolution
EGPRS	enhanced general packet radio service
FA	finite alphabet
GMSK	Gaussian minimum-shift keying
GPRS	general packet radio service
GSM	global system for mobile communications
HT	hilly terrain
ISI	intersymbol interference
L-LS	linear least-squares
LPE	linear prediction-error
LS	least-squares
MAP	maximum *a posteriori* probability
MCS	modulation and coding scheme
MISE	mean integrated square error
ML	maximum likelihood
MLSE	maximum likelihood sequence estimation
MMSE	minimum mean-square error

MS	mean-square
MSE	mean-square error
MVU	minimum variance unbiased
pdf	probability density function
PSK	phase-shift keying
QoS	quality of service
RA	rural area
RBF	radial basis function
SIR	signal-to-interference ratio
SNR	signal-to-noise ratio
TDMA	time division multiple access
TS	training sequence
TU	typical urban
WSSUS	wide sense stationary uncorrelated scatterers

Nomenclature

a_ℓ	prediction-error filter coefficients
$\mathbf{a}$	prediction-error filter vector
b_k	transmitted (coded) bits
B	frequency bandwidth
B_c	channel coherence bandwidth
B_D	Doppler spread
c	multipath channel impulse response
c_n	multipath channel tap-gain
c_ℓ	space-time equalizer coefficients
$\mathbf{c}$	space-time equalizer vector
d_k	transmitted symbols
$\mathbf{d}$	transmitted symbols vector
D	delay parameter
e_k	error signal samples
$\mathbf{e}$	error signal vector
$E\{.\}$	expected value
$\mathcal{E}$	mean-square cost
f	receive filter impulse response
f_c	carrier frequency
f_D	maximum Doppler shift
F	receive filter frequency response
$\mathcal{F}$	receiver noise figure
g	transmit impulse response
h	equivalent channel impulse response
h_ℓ	equivalent discrete-time channel tap-gain
$\mathbf{I}$	identity matrix
J	least-squares cost
J'	regularized least-squares cost
k_0	kernel width parameter

$K(.,.)$	kernel function
L	channel length
$L(.)$	log-likelihood ratio
$\mathcal{M}$	mapping rule
n_k	noise-plus-interference samples
$\mathbf{n}$	noise-plus-interference vector
N	noise power spectral density
$\mathcal{N}(.)$	Gaussian density
$p(.)$	probability density function
$\mathbf{p}$	time averaged cross-correlation vector
P	signal power
r	received signal
r_k	received signal samples
$\mathbf{r}$	received signal vector
$\mathbf{R}$	time averaged correlation matrix
s	transmitted signal
S_c	Doppler spectrum
T	symbol interval
T_c	channel coherence time
T_s	sampling period
u_k	uncoded bits
w	thermal noise
w_k	thermal noise samples
z_k	matched filter output samples
α	array response
δ	Dirac's delta function
η	mean value
θ	angle of arrival
λ	regularization coefficient
λ_c	carrier wavelength
$\lambda(.)$	metric increment
$\Lambda(.)$	accumulated metric
μ_i	i-th order moment

μ_k	trellis state
ξ_k	trellis state transition
σ_0	kernel width
σ^2	variance
τ	time delay
Φ_c	power delay profile

Chapter 1
Introduction

Channel equalization dates back to the early work of Lucky [1], Proakis and Miller [2], and others (see, e.g., [3]-[6] and references therein), who established the theory of *adaptive transversal* (or *tapped-delay-line*) *equalizers*, adjusted by the *zero-forcing* or the *minimum mean-square error* (MMSE) criterion. This early work aimed almost entirely at the telephone channel, which can be essentially characterized as a linear time-invariant, intersymbol interference (ISI) channel. Later work was related to the line-of-sight microwave channel, which may be considered as a very slowly time-varying channel [3], and to the time-varying ionospheric and tropospheric channels [5]. From the point of view of the equalizer structure, it was soon recognized that over highly dispersive channels the performance of linear receivers can be considerably improved by nonlinear schemes. Nonlinear receivers employ *decision-feedback equalizers* [3], [7], [8] or *trellis equalizers* based on *symbol-by-symbol maximum a posteriori probability* (MAP) *estimation* [9]-[13] or *maximum likelihood sequence estimation* (MLSE) [17]-[21]. In particular, among trellis based receivers, Forney's MLSE receiver for finite-length ISI channels [17], [19] has gained large popularity due to the efficient implementation of the trellis processor by means of the Viterbi algorithm [18], [22].

Equalization of the mobile digital radio channel has recently presented some new challenges, mainly due to the effect of the relative motion between the transmitter and receiver [5], [23], [24], [27], [28]. In the case of *time division multiple access* (TDMA) wireless systems, user mobility coupled with multipath propagation results in frequency selective time-varying fading, which causes severe ISI [5]. Due to these channel characteristics, practical TDMA mobile radio receivers often employ trellis equalization techniques [29]-[33]. The operation of these MLSE or MAP equalizers relies on the estimation of the channel impulse response, which is usually accomplished using a sequence of known training symbols transmitted within each TDMA data burst [32], [33].

Another important aspect of wireless transmission is that the receiver is generally affected by interference from other communications operating in the same frequency band [27]. In modern cellular systems, the reuse of the same carrier frequencies in different cells (*frequency reuse*)

causes *co-channel interference* (CCI). Frequency reuse also introduces adjacent channel interference (ACI), when neighbouring cells use frequencies that are spectrally adjacent to each other. In current TDMA systems like the Global System for Mobile Communications (GSM) [28], improved spectral efficiencies are obtained by either reducing the cell size or lowering the frequency reuse factor. This also applies to next generation TDMA standards like the General Packet Radio Service (GPRS) and Enhanced Data Rates for GSM Evolution (EDGE) systems [34], [35]. However, reduced cell sizes cause frequent handoffs, while lower frequency reuse factors result in higher CCI [27]. Therefore, the receiver performance in the presence of interference is of primary importance for achieving a higher system capacity. Channel equalization in the presence of CCI and/or ACI has been studied, e.g., in [36], [39]-[43]. In interference-limited scenarios, the performance of trellis equalizers critically depends on the validity of the statistical model of the disturbance resulting from the sum of interference and thermal noise, and on the quality of the channel estimate that represents the ISI. Under these conditions, in the case of multiple antenna receivers, interference cancellation by means of deterministic *space-time* processing [44]-[47] can often provide an effective alternative solution, with advantages in terms of robustness and implementation complexity [48]. Multiple antennas can give additional degrees of freedom for suppressing ISI, CCI, and ACI [49]. In general, the spatial dimension allows signal separation and unwanted signal suppression in multiple-access communication systems. The use of multiple antennas also provides signal diversity and thus reduces the effect of signal fading [51], [52].

An important issue for present and future cellular systems is the capability of delivering quality of service (QoS) in terms of delay, throughput and error performance guarantees over the nonstationary and dispersive wireless channel. Therefore, in conjunction with countermeasures against multipath and interference, powerful interleaving and forward error correction coding schemes are usually employed in order to obtain acceptable transmission quality [53]. In receivers employing the concatenation of an equalizer and a channel decoder, the overall performance can be improved by *soft-decision decoding* [56]-[58], [13], [14]. This requires a *reliability information* associated with the equalizer output data. In this respect, the optimum equalizer is the symbol-by-symbol MAP estimator, which has been employed as a building block for *iterative (turbo) channel estimation, equalization and decoding* (see, e.g., [59] and references therein). These techniques have been recently the object of extensive research. However, earlier work has been mainly devoted to improving the error performance of receivers in noise-limited scenarios [53]. This thesis studies the application of *soft-value* processing to a

probabilistic receiver in the presence of (*non-Gaussian*) interference.

In the presence of ISI, CCI, and/or ACI, the optimum trellis equalizer is based on *joint detection* of the desired and interfering signals [62]-[64]. Given the knowledge of the channel response of desired signal and interferers, joint MLSE and joint MAP equalization are optimal. However, their implementation cost can be prohibitive, since the computational complexity increases exponentially with the sum of the channel length of all the co-channel signals.

The rest of the thesis is organized as follows. Chapter 2 introduces the system model used in the thesis and discusses the basic issues of equalization of the mobile channel in interference-limited environment. The material presented in this chapter includes a review of the wireless propagation channel model, and of the effect of noise and interference on the design of a cellular system. A basic problem of channel equalization is that of deriving a set of sufficient statistics for data estimation, which requires an appropriate design of the receiver front-end. The chapter provides a discussion on the design of the front-end filter in the presence of additive non-Gaussian disturbance. The second part of the chapter reviews the principles of *soft-in/soft-out* processing and its application to iterative techniques for equalization and decoding, and discusses the use of trellis equalization and linear or decision feedback space-time equalization in the presence of non-Gaussian disturbance.

Chapter 3 studies a novel trellis processor for single-channel MAP equalization in the presence of ISI and non-Gaussian interference [65], [66]. Conventional trellis equalizers assume additive white Gaussian noise (AWGN), where in the presence of CCI and/or ACI the input disturbance is generally non Gaussian. In order to correctly set the problem of trellis data estimation, a proper statistical characterization of the disturbance is required. The proposed approach is based on the *non-parametric* estimation of the density function of CCI plus noise by *kernel smoothing* [67]-[70]. Here, the term *non-parametric* is used to signify that no prior assumption is made on the statistical model of the noise-plus-interference. The work considers the problem of density estimation with a limited volume of training data, and identifies symmetry conditions that allow to effectively double the size of the training set. The chapter also addresses the use of a whitening filter in the presence of temporally coloured disturbance. Contributions include the study of the optimum smoothing parameter for the case of complex densities, and the analysis of the density estimator performance in terms of mean integrated square error (MISE). The analysis is used to derive a simple proof that matched filtering does not provide sufficient statistics for data estimation in the presence of additive non-Gaussian

disturbance, in the case of a Gaussian mixture density model of the noise-plus-CCI. The implementation complexity of the proposed equalizer is analysed using explicit complexity formulas as a function of the algorithm parameters. As an application, simulation results are provided for the GSM system, showing a significant performance improvement with respect to the equalizer based on the Gaussian assumption and improved quality of the output reliability information for soft-decision decoding.

Chapter 4 considers the case of an antenna array receiver, and studies the performance of a simple method to derive the reliability information at the output of a deterministic *least-squares* (LS) *space-time equalizer* [72], [73]. Linear or decision-feedback LS equalization has the advantage of estimating directly the space-time filter coefficients, without requiring the estimate of the channel. The *a posteriori* probability (APP) calculation is based on the computation of the *short-term* statistics of the output disturbance, under the assumption that the decision-feedback space-time processor whitens the sum of CCI and noise both spatially and temporally. Contributions include here the analysis of the performance of different error variance estimators, and the study of a novel variance estimator for PSK signals. Computer simulations for the EDGE/EGPRS system show that the attainable receiver block-error rate performance largely depends on the quality of the reliability information delivered to the channel decoder. The same chapter also studies the use of *soft-decision feedback* [74], [59]. This technique, based on the theory of *nonlinear mean-square* (MS) estimation [75], is employed in conjunction with the above soft-output calculator to increase the robustness of decision feedback soft-out space-time equalizers.

Finally, Chapter 5 summarizes the results of the thesis and addresses areas for future research.

Chapter 2
Equalization of the mobile radio channel

Wireless communication systems generally require channel equalization to combat intersymbol interference (ISI) due to multipath propagation and interference from other users and other cells. In mobile cellular systems, given the need to operate at low values of signal-to-noise and signal-to-interference power ratios, interleaving and forward error correction coding are usually employed in conjunction with equalization, in order to provide adequate error performance. The performance of these receivers can be improved by means of iterative processing based on a probabilistic use of *soft* decisions in channel estimation, equalization, and decoding. As discussed in Chapter 1, soft-value processing has been the subject of extensive research over the last decade. Most of earlier work has considered soft-output equalization in the presence of additive Gaussian disturbance. This chapter discusses the basic issues related to the problem of soft-output equalization of the time division multiple access (TDMA) mobile radio channel in interference-limited scenarios. The chapter is organized as follows. Section 2.1 reviews the main features of the wireless propagation channel, including the characterization of the signal variations due to multipath fading, path loss, and shadowing, and the effect of the signal and interference parameters on the design of a cellular system. Section 2.2 introduces the signal model used in this thesis and discusses the design of the receiver front-end in the presence of non-Gaussian noise. Section 2.3 addresses the use of soft decisions in a probabilistic receiver with interleaving and channel decoding, and provides some background on iterative processing based on the calculation of the *a posteriori* probability. Finally, Section 2.4 discusses advantages and disadvantages of trellis processing and linear or decision-feedback space-time processing for soft-out equalization in the presence of interference.

2.1 Wireless propagation channel

In the mobile radio environment, the transmitted signal is modified by the time-varying propagation channel. In this respect, three physical mechanisms can be identified [5], [23], [27], [28]:

a) long-term signal fading or *path loss*; *b) medium-term* signal variations, due to *shadowing*; *c) short-term* signal fading, due to *multipath* propagation. Each of these phenomena is caused by a different underlying physical principle and must be taken into account, at different levels, in the design of a radio receiver.

Path loss or *inverse distance power loss* accounts for slow variations of average signal strength caused by varying distances between transmitter and receiver. Signal variations due to shadowing are often modelled as log-normal fading. Path loss and shadowing information is essential in determining the size of the coverage area for radio communications and in selecting optimum locations for base antennas [27].

Multipath fading results in rapid variations in the envelope of the received signal. Typically, the received envelope can vary by as much as 30-40 dB over a fraction of a wavelength [23]. In wireless transmission systems, multipath propagation often causes *frequency selective* distortion, which results in ISI. Over the typical duration of one burst in a wireless system, both inverse power loss and shadowing correspond to a simple attenuation of the received signal. Therefore, the receiver design and the choice of the proper digital signal processing algorithms depend essentially on the characterization of the multipath process.

In a multipath propagation channel, due to reflections and diffractions by scatterers, the transmitted signal usually reaches the receiver's antenna via several paths. The superposition of the arriving paths at a given value of delay induces destructive and constructive interference, which varies as a function of the position. As the antenna moves through this interference pattern, its spatial variation appears as a time-variation in the received signal. In addition, due to the motion of the antenna, the signal on each path undergoes a Doppler frequency shift that depends on the path arrival angle α. The maximum (one-sided) Doppler shift f_D equals the relative speed v of the transmitter, channel scatterers and receiver, divided by the carrier wavelength $\lambda_c = c/f_c$ (where f_c is the carrier frequency and c is the velocity of light). Consequently, the received signal is the sum of many Doppler shifted, scaled and delayed versions of the transmitted signal. The fading multipath channel is thus generally a doubly spread channel in time and frequency [5], [23]-[27]. At a given time t, the channel impulse response can be modelled as a densely tapped delay line with delay index τ [5]

$$c(t,\tau) = \sum_{n=0}^{N_c-1} c_n(t)\,\delta(\tau - \tau_n)\,. \tag{2.1}$$

In terms of the various paths, $c(t,\tau)$ is the sum of the complex gains of all the paths with delay τ, measured at the current location of the receiver's antenna. Under the assumption of *wide sense stationary uncorrelated scatterers* (WSSUS) [23], [26], from the *central limit theorem* one can assume Gaussian statistics. Therefore, the channel is characterized by the mean and correlation of the time-varying response $c(t,\tau)$. When, for a given delay τ_n, the complex gains due to different scatterers have similar amplitudes (*diffuse* scattering), the real and imaginary parts of $c(t,\tau)$ are zero-mean stationary independent coloured Gaussian processes, and their envelope follows a *Rayleigh* distribution [5], [23], [27], [28]. In this case, the tap-gain processes $c_n(t)$ are expressed as

$$c_n(t) = \sum_{k=0}^{K_n-1} \varrho_k \, e^{j(2\pi f_{Dk} t + \varphi_k)} \,, \tag{2.2}$$

which represents a cluster of incoherent rays arriving with (approximately) equal delay. In (2.2), the gain factors ϱ_k, the Doppler shifts $f_{Dk} = f_D \cos\alpha_k$, and the phase shifts φ_k may be assumed fixed during very short time intervals. Differently, in the presence of a dominant path (line-of-sight or *specular* reflection), the corresponding cluster can be modelled essentially as a single coherent path

$$c_n(t) = \varrho_n \, e^{j(2\pi f_{Dn} t + \varphi_n)} \,. \tag{2.3}$$

In this case $c(t,\tau)$ has a non-zero mean and the envelope has a *Rician* distribution [23], [28].

As a first-order statistical description, $c(t,\tau)$ can be decomposed into a *specular* and a *diffuse* component. The specular component, relative to the dominant path, is defined by $c_s(t,\tau) = E\{c(t,\tau)\}$ and is known as the *channel mean*. The diffuse component is given by $c_d(t,\tau) = c(t,\tau) - E\{c(t,\tau)\}$ and is *Rayleigh* distributed. A sufficient second order statistical description of the process $c(t,\tau)$ is given by the *tap-gain cross-correlation function*

$$R_{cc}(t_1,\tau_1,t_2,\tau_2) = E\{c(t_1,\tau_1)c^*(t_2,\tau_2)\}$$

$$= \sum_{n_1=0}^{N_c-1} \sum_{n_2=0}^{N_c-1} E\{c_{n_1}(t_1)c^*_{n_2}(t_2)\}\delta(\tau_1 - \tau_{n_1})\delta(\tau_2 - \tau_{n_2}) \,. \tag{2.4}$$

Under the WSSUS assumption and in the case of Rayleigh distribution ($E\{c_n(t)\} = 0$) one has

$$R_{cc}(t_1, \tau_1, t_2, \tau_2) = R_{cc}(t_1, t_2, \tau_1)\delta(\tau_1 - \tau_2)\,,$$

where

$$R_{cc}(t_1, t_2, \tau) = \sum_{n=0}^{N_c-1} E\{c_n(t_1)c_n^*(t_2)\}\delta(\tau - \tau_n)$$

$$= R_{cc}(t_1 - t_2, \tau) = R_c(t_1 - t_2)\Phi_c(\tau)\,. \tag{2.5}$$

In equation (2.5), the quantity $R_c(t)$ is the *normalized autocorrelation function*, $R_c(0) = 1$, and

$$\Phi_c(\tau) = \sum_{n=0}^{N_c-1} E\{|c_n(t)|^2\}\delta(\tau - \tau_n) \tag{2.6}$$

is the *delay power density profile* [5]. The function $\Phi_c(\tau) = R_{cc}(0, \tau)$ is proportional to the average power received from scatterers at delay τ, and the range of values of τ over which $\Phi_c(\tau)$ is essentially non-zero is the *multipath delay spread* of the channel.

The Fourier transform of $R_{cc}(t, \tau)$ with respect to the correlation lag t is the *scattering function*

$$S_{cc}(f, \tau) = \int_{-\infty}^{\infty} R_{cc}(t, \tau)e^{-j2\pi ft}dt\,. \tag{2.7}$$

For a WSSUS channel, $S_{cc}(f, \tau)$ is proportional to the power scattered by the medium at delays $(\tau, \tau + d\tau)$ in the Doppler shift interval $(f, f + df)$. The Fourier transform $S_c(f)$ of the normalized autocorrelation function $R_c(t)$ is the *Doppler spectrum* of the fading channel process [23]-[25], and the range of values over which $S_c(f)$ is non-zero is the *Doppler spread* of the channel, $B_D = 2f_D$.

Each resolvable multipath component may be then characterized by its own appropriate Doppler spectrum and corresponding Doppler spread. Under the assumption of isotropic scattering, $R_c(t)$ is commonly modelled as [23], [24]

$$R_c(t) = J_0(2\pi f_D t)\,, \tag{2.8}$$

where $J_0(x)$ is the zero-order Bessel function. The resulting fading power spectrum is the *classical* Doppler spectrum [23], [25]

$$S_c(f) = \begin{cases} 1/[\,\pi f_D \sqrt{1-(f/f_D)^2}\,] & |f| \leq f_D \\ 0 & |f| > f_D \end{cases} . \tag{2.9}$$

The value of the Doppler spread B_D provides a measure of how rapidly the channel impulse response varies in time. Since $S_c(f)$ is related to $R_c(t)$ by the Fourier transform, the reciprocal of B_D is a measure of the *channel coherence time* T_c [5]. Hence, a fast fading channel corresponds to a small coherence time or, equivalently, a large Doppler spread.

In a similar way, the reciprocal of the multipath spread is defined as the *channel coherence bandwidth* B_c, which provides a measure of the frequency interval where the fading is correlated [5]. When the signal bandwidth B is greater than B_c, different frequency components are subjected to different attenuations and phase shifts, and the channel is said to be *frequency selective.*

For mobile radio channels, the multipath delay spread ranges from a few microseconds up to tens of microseconds. Therefore, signal bandwidths in the order of 200 kHz result in a frequency selective channel. Cellular systems like GSM and EDGE specify the multipath channel models for the different propagation environments, in order to provide appropriate test conditions for different implementations of the receiver [76]. For each propagation condition, the GSM multipath channel is described by the time-varying impulse response (2.1), with complex tap gains $c_n(t)$ of a given mean power and delay τ_n [28], [76]. In the GSM channel profiles, the presence of a line-of-sight or specular component is specified only for the first tap of the model for rural area conditions (RA). In all other cases, including the model for typical urban area (TU), the stochastic processes are characterized by a Rayleigh distribution with the classical Doppler spectrum (2.9). In terms of multipath delay spread, the rural area response decays fast within one bit interval. The hilly terrain (HT) model has a long-delay part around 15-20 μsec due to distant reflections. The typical urban (TU) impulse response spreads over a delay interval of 5 μsec, which corresponds to almost two 3.69 μsec bit interval duration. This symbol duration results much smaller than the channel coherence time relative to mobile speeds v of up to 100 km/h. Therefore, the GSM channel can be characterized as a slowly time-varying fading channel.

The above channel model with the Doppler spectrum (2.9) is based on the assumption that the received signal angle of arrival is uniformly distributed in the interval $(0, 2\pi)$ [23]. This assumption is reasonable in the case of a mobile station, where the scatterers surrounding the receiver are about the same height than the mobile. However, in the case of a base station, the receive antenna is usually higher than the surrounding scatterers (especially in a macrocell environment), and the distribution of the angle of arrival is often restricted to a small angular region. In the case of a multiple-antenna receiver, it is important to include in the channel model spatial information on angle of arrival and antenna array geometries, in order to take into account the effect on the correlation between signals received at distinct antenna elements [23], [77]. For a multiple-antenna receiver, the spatial information is usually added to each multipath component of (2.2), which can be rewritten for the generic m-th antenna array element as [78]

$$c_{n,m}(t) = \sum_{k=0}^{K_{n,m}-1} \varrho_k \, e^{j(2\pi f_{Dk} t + \varphi_k)} \alpha_m(\theta_{n,k}) \, . \tag{2.10}$$

In (2.10), $\alpha_m(\theta_{n,k})$ denotes the m-th component of the *array response vector*, which is a function of the array geometry and angle of arrival $\theta_{n,k}$ [78], [45]. For a linear array one has

$$\alpha_m(\theta_{n,k}) = e^{j2\pi m \frac{x}{\lambda_c} \sin\theta_{n,k}} \, , \tag{2.11}$$

where x represents the distance between antenna array elements and λ_c is the carrier wavelength. With this model, the multipath is characterized in terms of Doppler spread, delay spread, and *angle spread.*

Besides multipath distortion, wireless cellular systems are affected by noise and interference. The thermal noise power received at the antenna depends on the temperature and bandwidth of the system. Denoting by T_0 the noise temperature in degrees Kelvin, the input noise spectral density at the receive antenna is given by $N = \mathcal{K}T_0$ Watt/Hz, where $\mathcal{K}$ is the Boltzmann's constant. Additionally, circuit noise is generated within the RF and IF stages of the receiver. The equivalent input noise density N_F that accounts for the noise generated within the receiver is conventionally represented by the receiver *noise figure* $\mathcal{F} \triangleq (N + N_F)/N$. Hence, the total noise power in the receiver bandwidth B_F results $(N + N_F) \cdot B_F = N\mathcal{F}B_F = N_0 B_F$, where $N_0 = N\mathcal{F}$ indicates the total noise density at the receiver input.

In mobile cellular systems, frequency reuse causes co-channel interference (CCI), due to the use of the same carrier frequencies in spatially separated cells, and adjacent-channel interference

(ACI), when spectrally adjacent carrier frequencies are used by neighbouring cells. Therefore, interference is one of the major factors that limit the capacity of cellular systems [27] and often one of the dominant impairments in wireless networks. The presence of interference results in a multiple-input channel model, where each interferer is usually characterized by independent multipath fading. In the presence of multiple receive antennas, this model produces a spatially and temporally coloured non-Gaussian disturbance.

It is important to note that in general the time-slot structure of the interfering signals is not synchronized with that of the user of interest. With such an *asynchronous* interference model, different portions of a received burst of the signal of interest are randomly affected by independent interfering signals. This thesis will also consider the case of *synchronous* interference, where the interfering slots are (approximately) time aligned with that of the desired user, which is a reasonably accurate model for a cellular system with synchronized base stations and small cell size.

While the multipath fading model characterizes the short-term or *local* variation of the co-channel signals, the long-term or *global* statistics are described by path loss models, which account for the reduction of received signal power with the distance from the transmitter. The classical free space model predicts that the received power P decays with the square of the radio path length d,

$$P(d) = (\frac{\lambda_c}{4\pi d})^2 P_T \,, \tag{2.12}$$

where λ_c is the carrier wavelength and P_T is the transmitted power [23], [5]. Free space propagation does not apply in a mobile radio environment and the propagation path loss depends not only on the distance and wavelength, but also on the antenna heights of the mobile terminal and base station and on the local radio environment [23]. The simplest path loss model assumes that the received power is given by

$$P_{\mathrm{dB}}(d) = P_{\mathrm{dB}}(d_0) - 10\beta \log_{10}(\frac{d}{d_0}) + \epsilon_{\mathrm{dB}} \quad \mathrm{dBm} \tag{2.13}$$

where the term $P_{\mathrm{dB}}(d_0)$ gives the received signal power (in dBm) at a known reference distance from the transmitting antenna [23], [27]. The parameter β is the *path loss exponent*, which depends on the cell size and the characteristics of the local environment. β ranges from 3 to 4 for a typical urban macrocellular environment, and from 2 to 8 for a microcellular environment

[23], [27]. The parameter ϵ_{dB} in (2.13) is a zero-mean random variable that represents the error between the actual and estimated path loss. This statistical variation of $P_{\mathrm{dB}}(d)$ is caused by shadowing, and is generally modelled by a log-normal distribution with mean η_P

$$\eta_P(d) = E\{P_{\mathrm{dB}}(d)\} = \eta_P(d_0) - 10\beta \log_{10}(\frac{d}{d_0}) \quad \mathrm{dBm} \tag{2.14}$$

and variance σ_P^2. For macrocells, σ_P^2 usually ranges from 5 to 12 dB, with $\sigma_P^2 = 8$ dB being a typical value. σ_P^2 has been observed to be nearly independent on the radio path length d [27].

Path loss models are usually employed in the design of a cellular system to derive the required transmitter power and *frequency reuse factor*, which is defined as the ratio of the co-channel reuse distance between cells using the same set of carrier frequencies and the radius of the cells. The minimum transmitter power is determined on the basis of the required outage probability due to thermal noise, defined as the probability that the signal-to-noise power ratio (SNR) is below a predefined threshold. The above outage probability can be computed from the propagation path loss model and transmitted power, given the spatial distribution of the mobile receivers within the cell. Similarly, the required cellular frequency reuse distance is determined from the system outage probability due to interference, defined as the probability that the average signal-to-interference power ratio (SIR) is lower than a predefined threshold. For a given frequency reuse pattern and channel utilization, the outage probability due to interference can be obtained by computing the SIR distribution as a function of the transmitted power and path loss model parameters [79], [80], [27]. The outage probability thresholds depend in general on the transmission quality requirements, the modulation and coding scheme, and the signal processing algorithms implemented at the receiver. Current cellular systems are usually deployed to provide 90-95 percent coverage for voice service. As a result, SNRs and SIRs that are much higher than the target are achieved over a significant portion of the cell area. For packet data service, this margin can be used to provide higher data rates by means of *rate adaptation* techniques such as *adaptive modulation and coding* [35], [81].

2.2 Receiver design and signal model

This section introduces the signal model assumed in the thesis and discusses the design of the receiver front-end in the presence of additive non-Gaussian disturbance. For simplicity, we limit the analysis to the case of a single-channel receiver. The model can be directly extended

to the case of multiple-antenna receiver considered in Chapter 4.

2.2.1 Single-channel receiver

The system block diagram for single-channel reception is given in Figure 2.1. We consider a linearly modulated signal, expressed in complex baseband notation as

$$s(t) = \sum_i b_i g(t - iT) \,, \tag{2.15}$$

where b_i denotes the i-th transmitted symbol, $1/T$ is the symbol rate, and $g(t)$ is the transmit impulse response or pulse shape. The symbols are assumed independent, identically distributed (i.i.d.), with values $\{+1, -1\}$ (the analysis can be easily generalized to the case of symbols d_i taken from an M-ary constellation). The signal (2.15) is transmitted over a multipath fading channel with impulse response $c(t, \tau)$, which produces the channel output

$$y_c(t) = \int_{-\infty}^{\infty} s(t - x)c(t, x)dx = \sum_i b_i h_c(t, t - iT) \,, \tag{2.16}$$

with

$$h_c(t, \tau) \triangleq g(\tau) \otimes c(t, \tau) = \int_{-\infty}^{\infty} g(\tau - x)c(t, x)dx \,, \tag{2.17}$$

where the symbol "$\otimes$" denotes the convolution operator. In addition to channel distortion, the receiver is affected by co-channel interference and thermal noise. Figure 2.1 assumes for simplicity the presence of one dominant co-channel interferer, which is modelled as

$$y'_c(t) = \sum_i b'_i h'_c(t, t - iT) \,. \tag{2.18}$$

The thermal noise is represented by an additive white complex Gaussian process $w_c(t)$, with zero mean and double-sided power spectral density $N_0 = 2\sigma^2$.

The resulting signal is passed through a front-end receive filter with impulse response $f(t)$. Letting $y(t) \triangleq y_c(t) \otimes f(t)$, $y'(t) \triangleq y'_c(t) \otimes f(t)$ and $w(t) \triangleq w_c(t) \otimes f(t)$, the filter output is

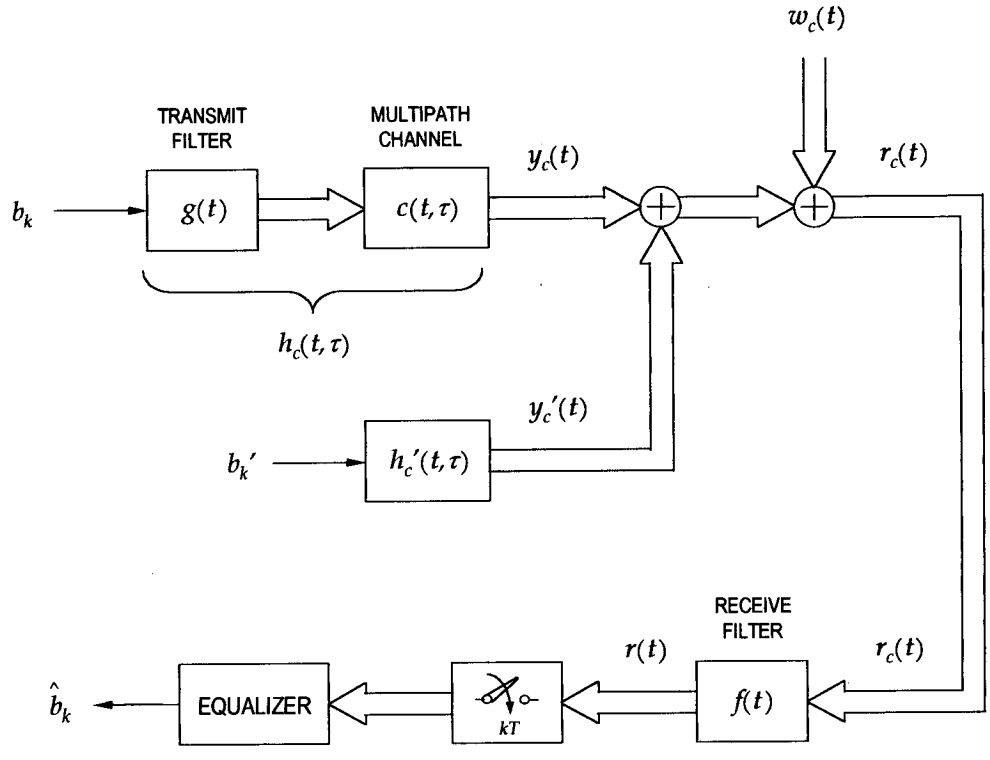

Figure 2.1: *System model for the single-channel receiver.*

expressed as

$$r(t) = y(t) + y'(t) + w(t)$$

$$= \sum_i b_i h(t, t - iT) + \sum_i b'_i h'(t, t - iT) + w(t)\,, \tag{2.19}$$

where $h(t,\tau)$ and $h'(t,\tau)$ denote the equivalent channel impulse responses of the desired and CCI signals

$$h(t,\tau) \triangleq \int_{-\infty}^{\infty} f(x) h_c(t - x, \tau - x) dx\,, \tag{2.20}$$

$$h'(t,\tau) \triangleq \int_{-\infty}^{\infty} f(x) h'_c(t - x, \tau - x) dx\,. \tag{2.21}$$

Using (2.1) and (2.17), (2.20) can be rewritten in terms of the multipath channel taps $c_n(t)$ as

$$h(t,\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x) g(y - x) c(t - x, \tau - y) dx dy$$

$$= \sum_{n=0}^{N_c-1} \int_{-\infty}^{\infty} f(x) c_n(t - x) g(\tau - \tau_n - x) dx\,. \tag{2.22}$$

The receiver model assumes that sufficient statistics for data estimation are obtained by symbol-rate sampling at the output of the analog front-end filter. In particular, if the signal $y(t)$ is bandlimited with bandwidth $B = 1/2T$, and $f(t)$ is the ideal low-pass filter with Fourier transform

$$F(f) = \begin{cases} \sqrt{T} & |f| \le 1/2T \\ 0 & |f| > 1/2T \end{cases}\,, \tag{2.23}$$

it is $y(t) = y_c(t)$, $y'(t) = y'_c(t)$, and the noise $w(t)$ has variance $E\{|w(t)|^2\} = 2\sigma^2$.

Under the above assumption, the taps of the equivalent discrete-time channel and co-channel impulse responses at the input of the equalizer are defined as $h_\ell^{(k)} = h(t = kT, \tau = \ell T)$ and $h_\ell^{\prime(k)} = h'(t = kT, \tau = \ell T)$, respectively. The received signal samples at the equalizer input $r_k = r(kT)$ are then written as

$$r_k = y_k + y'_k + w_k$$

$$= \sum_{\ell=0}^{L-1} b_{k-\ell} h_\ell^{(k)} + \sum_{\ell=0}^{L'-1} b'_{k-\ell} h_\ell^{\prime(k)} + w_k \,, \tag{2.24}$$

where $y_k = y(kT)$, $y'_k = y'(kT)$, and $w_k = w(kT)$. Using the filter (2.23) (or, more in general, if $F(f)$ is the square root of a Nyquist filter [5]), $E\{w_k w^*_{k-\ell}\} = 2\sigma^2 \delta_\ell$, and the process w_k is white.

As an application, we consider the case of the GSM system. In this case, the transmitted GMSK signal is given by

$$s(t) = \exp[j\frac{\pi}{2} \sum_{i=0}^{N-1} \xi_i q(t - iT)] \,, \tag{2.25}$$

where $\xi_i \in \{+1, -1\}$, and $q(t)$ is the phase response of the continuous-phase modulation, which results from the integration of the Gaussian pulse response [82]

$$q(t) = \int_{-\infty}^{t} u(\tau) d\tau, \tag{2.26}$$

$$u(t) = B\sqrt{\frac{\pi}{2\log 2}} \exp[-\frac{\pi^2 B^2 (t - 2T)^2}{2\log 2}] \,. \tag{2.27}$$

For GSM, $BT = 0.3$ is chosen [28], which corresponds to

$$q(t) = \begin{cases} 0 & t \leq 0 \\ 1 & t \geq 4T \end{cases} \,. \tag{2.28}$$

It is possible to show that (2.25) can be approximated by the linear model [83]

$$s(t) \cong \sum_i j^i b_i g(t - iT) \,, \tag{2.29}$$

with b_i binary symbols determined by the recursion $b_i = \xi_i b_{i-1}$. Hence, the corresponding received samples result

$$r_k = \sum_{\ell=0}^{L-1} j^{k-\ell} b_{k-\ell} h_\ell^{(k)} + \sum_{\ell=0}^{L'-1} j^{k-\ell} b'_{k-\ell} h_\ell^{\prime(k)} + w_k \,. \tag{2.30}$$

However, a binary modulation can be recovered from the samples (2.30) by applying the derotation [84]

$$j^{-k} r_k = \sum_{\ell=0}^{L-1} b_{k-\ell} [j^{-\ell} h_\ell^{(k)}] + \sum_{\ell=0}^{L'-1} b'_{k-\ell} [j^{-\ell} h_\ell^{\prime(k)}] + j^{-k} w_k$$

$$= \sum_{\ell=0}^{L-1} b_{k-\ell} \tilde{h}_\ell^{(k)} + \sum_{\ell=0}^{L'-1} b'_{k-\ell} \tilde{h}_\ell^{\prime(k)} + \tilde{w}_k \,. \tag{2.31}$$

Therefore, the received signal samples at the input of the equalizer can still be expressed by (2.24), by defining the taps of the discrete-time equivalent channels as

$$\tilde{h}_\ell^{(k)} = j^{-\ell} \cdot h(t = kT, \tau = \ell T) \,, \tag{2.32}$$

$$\tilde{h}_\ell^{\prime(k)} = j^{-\ell} \cdot h'(t = kT, \tau = \ell T) \,. \tag{2.33}$$

Observing that

$$j^{-n} = \cos(\frac{\pi n}{2}) - j \sin(\frac{\pi n}{2}) = e^{-j2\pi n (1/4T) T} \,,$$

one has that the derotation produces the equivalent channel

$$\tilde{h}(t, \tau) = h(t, \tau) \cdot e^{-j2\pi f_d \tau}, \qquad f_d = \frac{1}{4T} \tag{2.34}$$

with the corresponding discrete-time representation $\tilde{h}_\ell^{(k)} = \tilde{h}(t = kT, \tau = \ell T)$.

2.2.2 Receiver front-end and sufficient statistics

The discrete-time signal model introduced in the previous section assumes that a set of sufficient statistics for estimating the transmitted symbol sequence is derived by Baud-spaced sampling at the output of the front-end filter. As will be further discussed in Chapter 3, in the case of additive white Gaussian disturbance, the optimum front-end processor is given by the *whitened matched filter*, which consists of the continuous-time filter matched to the received pulse, followed by a noise whitening filter [17], [21]. If the input noise is white and Gaussian, the output of the above processor is a set of sufficient statistics modelled by an equivalent discrete-time, symbol-spaced ISI signal in white Gaussian noise. A first difficulty with the analog matched filter approach arises when the channel is *a priori* unknown, as discussed, e.g., in [85], [86]. Clearly, if the received signal impulse response $h_c(t,\tau)$ is not known and possibly time-varying, the matched filter can not be identified. Even if one happens to know h_c precisely, it may be difficult to design and accurately implement an adaptive analog matched filter. Besides practical implementation issues, the use of an analog matched filter with *a priori* unknown channel implies an imprecise notion of the front-end processing for the discrete-time model (2.24) [86]. Moreover, the matched filter is generally no longer optimum in the presence of non-Gaussian noise. In this respect, Chapter 3 will give a simple proof that matched filtering does not provide a sufficient statistic for data estimation under the non-Gaussian noise model of (2.19), (2.24). Generalized forms of the conventional matched filter for detection in non-Gaussian noise have been proposed (see, e.g., [87]), obtained from the ordinary matched filter by replacing multipliers with non-linearities. If the useful signal is strictly bandlimited with bandwidth B, a set of sufficient statistics can be obtained by sampling the output of a fixed analog filter $F(f)$ with bandwidth $B_F > B$, provided that

$$F(f) = \begin{cases} \text{arbitrary but } \neq 0 & |f| < B \\ \text{arbitrary} & B \leq |f| < B_F \\ 0 & \text{elsewhere} \end{cases}, \tag{2.35}$$

and the output is sampled at rate $1/T_s \geq B + B_F$ [88]. In fact, since $F(f) \neq 0$ for $|f| < B$, it is possible to show that the above processing is reversible [88]. The above conditions require that the sampling theorem applies to the useful signal, but tolerate an undersampling of the noise process as long as there is no aliasing of the disturbance into the passband $|f| < B$. One advantage of the latter solution is its simplicity. Notice that, compared to the matched filter of the Gaussian noise case, it does not reduce the dimensionality of the observation space [21].

In TDMA wireless systems, the transmitted signal is often substantially bandlimited with bandwidth $B \cong 1/2T$. In these conditions, from the above discussion follows that sufficient statistics can be obtained by symbol-rate sampling at the output of a low-pass filter $f(t)$ with bandwidth $B_F = 1/2T$. This motivates the assumption behind the discrete-time model (2.24). The approach can be extended to include the case of non-zero excess bandwidth by introducing oversampling and fractionally spaced equalization [61].

It is worth mentioning that, from a practical point of view, when $B > 1/2T$ the design of the receiver front-end implies a trade-off between sampling rate and analog prefilter complexity. In general, oversampling relaxes the requirements on the filter roll-off, thus leading to a simple analog filter implementation. Suboptimum receiver front-ends that employ Baud-rate sampling with excess bandwidth signals have been studied in [86], [85]. In this case, the optimum front-end filter followed by symbol-rate sampling may be viewed as a technique that sacrifices signal-to-noise ratio for reduced signal distortion [86]. The opposite effect is obtained by employing a low pass filter of bandwidth $1/2T$ followed by symbol-rate sampling [85].

2.3 Equalization and decoding in modern probabilistic receivers

TDMA digital cellular systems require channel equalization as a countermeasure against the time-varying multipath propagation. As discussed in Section 2.1, the *multipath delay spread* of the GSM mobile channel ranges from two to five symbol intervals [28], [76]. With these channel characteristics linear equalizers often lead to poor performance, resulting in noise enhancement at the frequencies corresponding to a spectral null of the channel response [5], [29]. On the other hand decision feedback equalizers, although attractive from the implementation point of view, generally suffer from error propagation when operating at low signal-to-noise ratios [5], [8]. For this reason, TDMA cellular receivers conventionally resort to trellis equalization [29]-[33].

Among trellis equalizers, maximum-likelihood (ML) sequence estimation implemented by the Viterbi algorithm is the optimum sequence detector. It is widely used in digital mobile receivers for processing both the ISI trellis (equalization) and the channel code trellis (channel decoding). However, MLSE equalizers provide only *hard* decision on the received symbol sequence. In modern digital wireless receivers employing the concatenation of an equalizer and a channel decoder, the performance can be further improved by *soft-decision decoding* [56], [13], [14],

[57], which requires a *reliability information* (*soft* decision) about the equalizer output data. Furthermore, in some advanced schemes implementing *iterative equalization and decoding* (*turbo equalization*) [89]-[92], the channel decoder must be able to provide soft outputs for both the coded bits and the information bits [57], [55], [15], [16], [93], [59]. As pointed out in [57], there are several advantages in maintaining soft values as long as possible in a digital receiver.

In terms of bit error probability, the optimum algorithm for *soft-in/soft-out* equalization and decoding is the symbol-by-symbol maximum *a posteriori* probability (MAP) algorithm. Originally proposed by Bahl *et al.* [9] for decoding of convolutional codes, the MAP algorithm initially received very little attention. In fact, it provides a minimal advantage in BER performance over ML decoders, at a higher implementation cost. However, being an APP calculator, it has the advantage of intrinsically providing optimal soft output values. Differences and similarities between ML sequence detection and symbol-by-symbol MAP detection have been discussed, e.g., in [11] and [12]. In particular, in [12] it has been shown that the logarithmic version of the MAP algorithm is equivalent to a combination of a forward and backward Viterbi processors, coupled by a dual maxima computation.

As an APP calculator, the MAP algorithm is now often employed as the optimum *soft-in/soft-out module* for iterative processing [16], [93], [53]. The idea of iterative or *turbo* processing, first introduced by Berrou *et al.* in 1993 with the so called *turbo codes* [54], has been recently extensively applied to improve the performance of equalization and decoding for a variety of communication systems. Of particular relevance is the mentioned application of the *turbo* concept to iterative equalization and decoding [89]-[92], where the dispersive transmission channel is regarded as an (unknown) rate 1 convolutional code concatenated with a forward error correction code, so that the receiver operation is treated as the decoding of a serially concatenated code. A block diagram illustrating the concept of iterative equalization and decoding is shown in Figure 2.2. Consider the case where the transmitted binary symbols b_i are obtained from the uncoded bits u_j by means of rate k/n channel encoding and interleaving. The *soft-in/soft-out* equalizer receives the channel output sequence $r_0, \ldots, r_{N-1}$, and delivers soft values on the encoded bits b_i in the form of log-likelihood values

$$L(b_i|r_0, \ldots, r_{N-1}) \triangleq \log \frac{\Pr(b_i = +1|r_0, \ldots, r_{N-1})}{\Pr(b_i = -1|r_0, \ldots, r_{N-1})} . \tag{2.36}$$

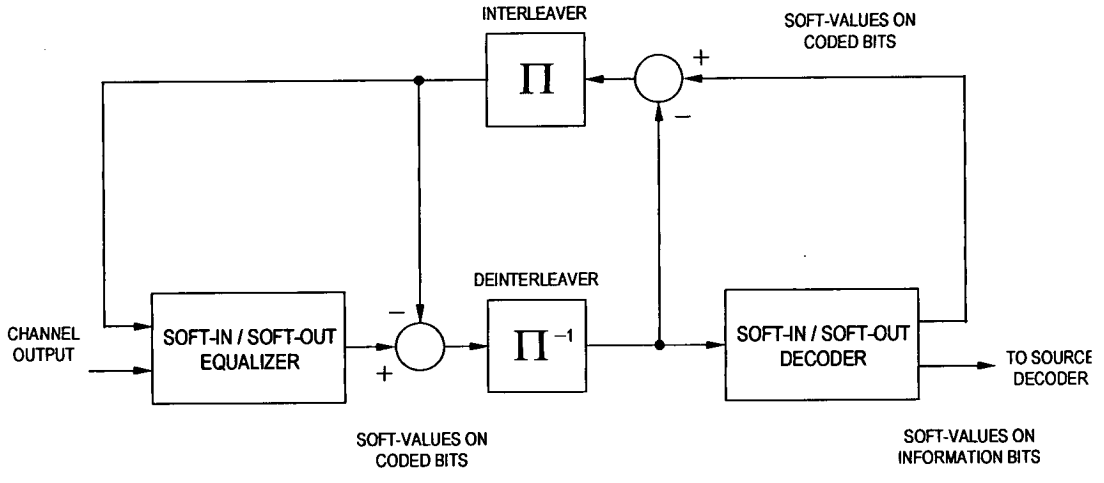

Figure 2.2: *Iterative equalization and decoding scheme.*

This values are deinterleaved and passed to the channel decoder. The *soft-in/soft-out* decoder acts as a nonlinear filter, which uses the dependence between the bits to produce less noisy estimates of the data. The output of the decoder is given by the log-likelihood values on the coded bits, and can be decomposed into two parts: *a) channel information* and *b) extrinsic information*. The former is the input to the decoder, while the latter is the incremental information about the bit b_i that derives from all other bits excluding itself. Hence, it represents the new information obtained by the decoder using the dependence between the bits introduced by the code. The extrinsic information is calculated by simply subtracting the decoder input from the decoder output. In an iterative equalization and decoding scheme, these new L-values are interleaved and fed back to the equalizer as *a priori* information $L(b_i)$ about the coded bit b_i. The equalizer now performs a second-pass equalization of the channel output sequence, taking into account the *a priori* knowledge $L(b_i)$. The extrinsic information from the equalizer is then obtained by subtracting the input *a priori* knowledge from the equalizer output. This iterative process can be performed a number of times, and at the last iteration the decoder produces the

final (hard or soft) estimates of the information bits u_j.

The operation of an APP trellis equalizer is based on the knowledge of the channel impulse response. Conventional TDMA receivers estimate the channel at the equalizer start-up using a known training sequence transmitted within each burst. In GSM receivers, this initial channel estimation is usually obtained by correlating the received signal r_k with the training sequence bits (*correlative channel sounding*) [32], [33], [59]. Given the good autocorrelation properties of the GSM training sequences, this corresponds to least-squares channel estimation [59]. The performance of conventional channel estimators can be improved by iterating the equalization and channel estimation procedures on a burst-by-burst basis [59]. With this *iterative channel estimation* approach, the first pass performs conventional channel estimation by training. The resulting channel estimate is then used to estimate the data in the trellis equalizer. After that, one or more iterations can be performed, where hard or soft symbol decisions at the equalizer output are fed back, and used together with the original training symbols to obtain an extended training sequence for a new channel estimation. This longer sequence results in a smaller channel estimation error, which in turn improves the performance of the trellis processor. However, the gain is clearly reduced in the presence of decision errors at the equalizer output. As is intuitive, soft-decision feedback mitigates the effect of error propagation, and provides robustness at low signal-to-noise ratios [94], [59]. In a receiver implementing iterative equalization and decoding, iterative channel estimation can be performed using the feedback of the channel decoder output. The latter solution provides an additional performance gain, since the channel decoder reduces the number of decision errors in the feedback sequence [59].

The use of soft decisions in wireless communication systems is not limited to iterative processing for equalization and channel decoding, but has also been successfully applied to *joint source and channel decoding* [95], [96], [53], and to *soft packet combining* to reduce the frequency of retransmission in hybrid *automatic repeat request* (ARQ) protocols [97], [53].

Most of earlier work on soft-value processing for equalization and decoding has been devoted to improving the error performance in noise-limited environments, where the above techniques have contributed to reach error rates close to the Shannon limit [53]. Relatively little research has addressed the use of reliability information for equalization and decoding in the presence of CCI and ACI. Under interference-limited conditions, conventional decoding algorithms based on the assumption of Gaussian disturbance are often no longer adequate. In a scenario where the dominant disturbance is constituted by structured and possibly nonstationary interference,

the application of a probabilistic trellis equalizer requires proper modelling of the disturbance, with the characteristic of robustness with respect to the structure of the interference.

The next section, after briefly reviewing recent advances in MAP trellis equalization of the additive Gaussian noise channel, will introduce two different novel approaches for soft-output equalization in the presence of *non-Gaussian* interference.

2.4 Trellis equalization and space-time equalization

As mentioned in the previous section, historically the main disadvantage for the use of the MAP algorithm with respect to ML sequence estimation has been its increased implementation cost. However, over the last decade several suboptimum soft-output trellis equalizers have been proposed, which provide nearly optimum performance at a significantly reduced implementation complexity. Among them, the *soft-output Viterbi algorithm* (SOVA) [56] is based on the idea of processing only one competing path per decoding stage (the survivor of the Viterbi algorithm). Its soft output is therefore noisier than with the optimum MAP algorithm. Following a different approach, the implementation of MAP processors has been significantly simplified by operating in the logarithmic domain (*log-MAP algorithm*), which avoids the need of calculating summations of exponential terms [14], [15]. Further simplification has been obtained by the *max-log-MAP algorithm* studied in [13]-[15]. Finally, the complexity penalty associated with the presence of both the forward and backward recursion has been eliminated by a suboptimal algorithm where processing is limited to the forward recursion, with the introduction of a decision delay depending on the code/channel memory [13], [14]. This *forward max-log MAP algorithm* has been shown to provide a negligible degradation with respect to the optimum strategy for MAP equalization of the GSM channel [13].

The above equalizers provide excellent performance in noise-limited environments [59]. However, in the presence of interference, a basic problem with all these probabilistic trellis processors is that their performance largely depends on the validity of the statistical model of the disturbance, and on the quality of the channel estimate that represents the ISI code. As already observed, while conventional trellis equalizers assume additive white Gaussian noise (AWGN), in the presence of CCI and/or ACI the total input disturbance is generally non-Gaussian. Moreover, the temporal colour of the CCI is *a priori* unknown. In this situation, the optimum trellis equalizer based on Gaussian statistics performs *joint detection* of all

the co-channel signals. Joint ML and MAP detection have been studied in [60]-[64]. Denoting by L_m the channel length of the m-th co-channel signal, $m = 1, \ldots, M$, and by $\mu_{k,m} \triangleq (b_{k-1,m}, \cdots, b_{k-L_m+1,m})$ the corresponding ISI state at time k, joint detection is based on the definition of the *joint state* $(\mu_{k,1}, \mu_{k,2}, \cdots, \mu_{k,M})$, and the number of states required to implement a binary trellis results $2^{\sum_{m=1}^{M}(L_m-1)}$. Therefore, although joint ML and joint MAP trellis processors are optimal, they can be prohibitively expensive even in the presence of moderate multipath delay spread. Reduced-state techniques have been addressed, e.g., in [60], [61], and [98]. An additional problem of trellis equalizers based on joint detection is that their operation relies on an estimate of the channel impulse response of all co-channel signals, which requires the knowledge of the training sequence of each interferer.

In the presence of *a priori* unknown interference, probabilistic techniques can be employed to cope with the difficulty of accurately modelling the disturbance. The new trellis processor studied in Chapter 3 is based on the estimation of the (generally multimodal) probability density function of the interference plus noise. The fundamental issue of density estimation with the availability of a limited volume of training data is addressed here by applying *kernel smoothing* techniques [67]-[70]. As observed in [71], the design of any decision rule should be based on a consistent estimate of the decision densities, which, in turn, is based on smoothing. In the case of correlated disturbance, the trellis branch metrics are computed at the output of a *linear prediction-error* filter, which (approximately) whitens the sum of noise and interference. Major advantages of the proposed strategy are its intrinsic robustness and general applicability to those cases where accurate modelling of the interference is difficult or a model is not available. The implementation complexity of the proposed equalizer results lower than with joint detection in the presence of one dominant co-channel signal, and does not increase with an increasing number of interferers. A further cost reduction may be achieved by preclustering of the training data for density estimation [99], [100].

It is worth emphasizing the close relationship of the above *non-parametric* trellis equalizer with *radial basis function* (RBF) neural networks [101]-[103]. It has been shown that the optimum symbol-by-symbol detector can be implemented by a recurrent version of the RBF network [104]. In fact, the RBF network, originally conceived as a general method for approximating nonlinear mapping, can itself be considered as a non-parametric technique for estimation of *a posteriori* probabilities [105] and can be seen as a method to realize the Bayesian decision function [102], [107], [108], [39], [40] . In this view, the present study implicitly proposes

a strategy for the application of neural network techniques to Bayesian channel equalization based on the ISI trellis.

In the case of multichannel reception, an effective alternative to trellis equalization is given by interference cancellation by linear or decision-feedback *spatio-temporal* processing [44]-[47], [110], [111]. Spatio-temporal filters operate simultaneously on all the receive antennas, processing signals in both space and time. Exploiting the spatial dimension provides CCI suppression, noise reduction, and spatial diversity against fading. Space-time *minimum mean-square error* (MMSE) receivers have been shown to provide good performance against CCI and ISI [45], [46], [111]. Space-time equalizers based on the deterministic *least-square* (LS) criterion have the further advantage of not relying on an estimate of the channel response, or on particular assumptions on the statistical model of the interference [110], [48]. In fact, in a scheme that relies on the estimation of the channel impulse response for the signal of interest, the available gain may be in practice reduced by the poor performance of conventional channel estimators. However, to enable the use of deterministic space-time equalizers in receivers based on soft-in/soft-out processing, one needs to provide means to use input *a priori* probability on the coded bits, and to generate reliability information for soft-decision decoding by estimating the *a posteriori* probability on the coded bits at the equalizer output.

The use of *a priori information* in a deterministic equalizer has been recently addressed in [92] for an MMSE block decision-feedback equalizer. A similar approach can be applied to a LS processor. To solve the problem of estimating the *a posteriori* probability at the equalizer output would in general require a MAP trellis processor. A hybrid receiver structure constituted by a space-time equalizer followed by a trellis equalizer has been recently proposed in [111], [112] and [113] (although not based on the motivation of providing soft-output values for channel decoding, see also [44]). However, it has been shown in [46] that the asymptotic performance of a space-time receiver incorporating a full-state trellis processor is only marginally superior to that of a decision-feedback space-time equalizer. Chapter 4 considers a regularized LS decision-feedback space-time receiver. In this case, assuming that the feedforward filters approximately whiten both spatially and temporally the sum of noise and CCI [46], we propose a simple method to derive the reliability information for soft-decision decoding. The technique relies only on the estimation of the output error variance, which can be periodically updated within each burst to cope with the nonstationarity of the error statistics due to asynchronous interference and/or feedback of possible decision errors. The robustness of the above decision-

feedback receiver is further improved by using the soft-output calculator in conjunction with soft-decision feedback based on nonlinear MS estimation.

Chapter 3
Non-parametric trellis equalization in the presence of non-Gaussian interference

In time division multiple access (TDMA) radio communications, the received signal is affected by co-channel interference (CCI) and intersymbol interference (ISI) resulting from multipath propagation. Channel equalizers often employed in practical receivers perform maximum likelihood (ML) [17]-[21] or maximum *a posteriori* probability (MAP) data estimation [9]-[13] on the ISI trellis. ML sequence estimation (MLSE) using the Viterbi algorithm [18] is well known as the optimum detection technique for signals corrupted by finite-length (deterministic) ISI and additive white Gaussian noise (AWGN), in the sense that it minimizes the probability of a sequence error. The symbol-by-symbol MAP algorithm, proposed over two decades ago by Bahl *et al.* [9] for decoding of convolutional codes, has recently received renewed interest as a *soft-in/soft out* decoder for iterative decoding of parallel or serially concatenated codes [54], [55], [14]-[16]. As a trellis equalizer, the MAP algorithm is optimum in the sense that it minimizes the probability of symbol error. In receivers employing the concatenation of an equalizer and a channel decoder, the performance is improved by soft-decision decoding and iterative equalization and decoding (*turbo* equalization) [89]-[91], [59]. In this respect, being an *a posteriori* probability (APP) calculator, the MAP algorithm has the advantage of intrinsically providing optimal *a posteriori* probability as a soft-output value.

In addition to ISI and thermal noise, radio receivers are affected by interference from other communications operating in the same frequency band. This chapter considers the problem of equalization of the ISI channel with single channel reception, in the presence of thermal noise and co-channel interference (CCI) or adjacent-channel interference (ACI). The optimum trellis equalizer in the presence of ISI, CCI, and AWGN, is based on joint detection of the co-channel signals [60]-[64]. As already mentioned in Chapter 2, although joint ML and joint MAP detection are optimal, their implementation complexity increases exponentially with the sum of the channel lengths of the desired and CCI signals, and can become prohibitive in

the presence of more than one interferer. In addition, the estimation of the channel impulse response of all co-channel signals requires the knowledge of the training sequence of each interferer. Even assuming the knowledge of the training sequences, the estimation of the co-channel responses will usually be inaccurate for relatively low received power levels of the CCI. On the other hand, conventional receivers employ a trellis equalizer which treats the sum of noise and CCI or ACI as additive, white, Gaussian noise. In reality, in many situations of interest, the sum of noise and interference is a non-Gaussian random process, and the above suboptimum approach corresponds to a degradation of the error performance.

The chapter studies a novel *non-parametric* trellis equalizer based on the estimation of the probability density function of the noise-plus-interference. Given the limited volume of training data, the work is based on the application of density estimation by *kernel smoothing* [67]-[69]. The temporal colour of the CCI is taken into account by a whitening filter.

Section 3.1 gives the system model and formalizes the operation of MAP and ML trellis equalizers for finite-length ISI in the presence of additive disturbance. Section 3.2 presents the new trellis equalizer based on non-parametric density estimation, and studies the design of the optimum kernel width or smoothing parameter. The section also provides the analysis of the density estimator performance in terms of mean integrated square error (MISE). The application of the proposed technique to a wireless receiver is discussed in Section 3.3. The same section addresses the design of the whitening filter for the disturbance, and discusses the implementation complexity of the non-parametric trellis processor. Simulation results are presented in Section 3.4 for the GSM transmission system. Finally, conclusions are drawn in Section 3.5.

3.1 MAP and ML trellis equalization

3.1.1 System model

Consider the received signal

$$r_k = y_k + n_k = \sum_{\ell=0}^{L-1} b_{k-\ell} h_\ell^{(k)} + n_k \,, \tag{3.1}$$

where $b_k \in \{+1, -1\}$ are the transmitted symbols, the L complex tap-gains $h_\ell^{(k)}$ represent the samples of the equivalent channel impulse response at time k, and $n_k = y'_k + w_k$ indicates the

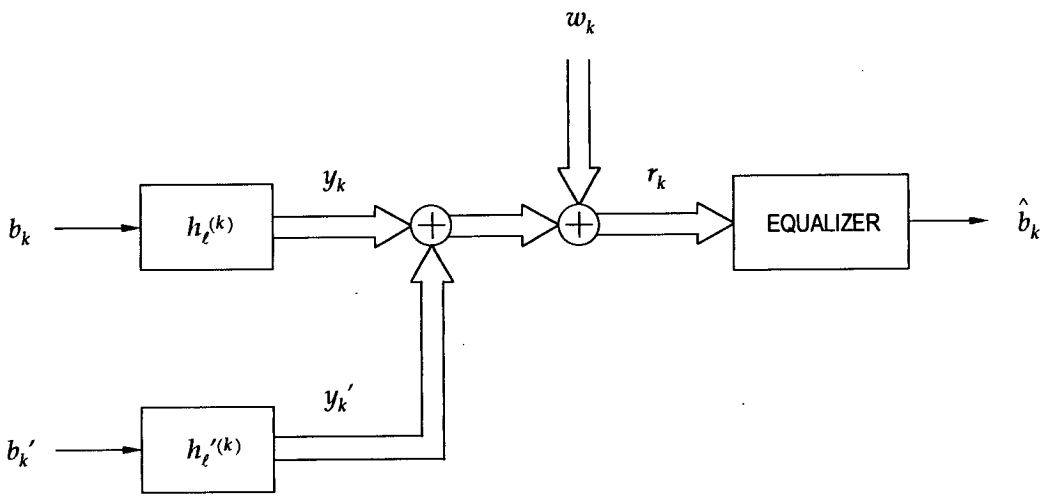

Figure 3.1: *Discrete-time signal model.*

sum of co-channel interference and thermal noise. As discussed in Chapter 2, the discrete-time model (3.1) assumes that sufficient statistics for data estimation are obtained by symbol-rate sampling at the output of a front-end receive filter. In the AWGN case, the optimum receive filter consists of Forney's whitened matched filter followed by symbol-rate sampling [17], [21] (or, equivalently, the matched filter followed by symbol-rate sampling and a noise whitening filter). The MLSE receiver proposed by Ungerboeck [19] does not require the whitening filter, and a sufficient statistic for data detection employing a modified metric is given by the matched filter output sampled at the symbol rate (see Section 3.1.2). Optimality considerations on the receiver structure developed for the case of MLSE in AWGN can also be applied to symbol-by-symbol MAP receivers. However, if the noise is not Gaussian the matched filter is no longer optimum. As already pointed out, in the case where the useful signal is strictly bandlimited with bandwidth B, a sufficient statistic for data estimation can be obtained by a fixed analog filter $F(f)$ with bandwidth $B_F > B$, whose output is sampled at a rate $1/T_s \geq B + B_F$ [88]. In fact, provided that $F(f) \neq 0$ for $|f| < B$, the above processing is reversible, and from

the *factorization theorem* [114], [115] follows that any invertible transformation of a sufficient statistic produces a sufficient statistic (see also Section 3.2). According to Chapter 2, here we assume that the transmitted signal is bandlimited with bandwidth $B \cong 1/2T$, and that (3.1) is obtained by symbol-rate sampling at the output of an ideal low-pass filter.

This section considers the CCI samples as independent complex non-Gaussian random variables. In the presence of dispersive (temporally coloured) interference, a suitable temporal prewhitening is assumed to produce approximately independent non-Gaussian disturbance. The validity of this assumption will be discussed in Section 3.2.

3.1.2 Symbol-by-symbol MAP algorithm for finite-length ISI and additive independent disturbance

Suppose that the symbols b_k are transmitted in finite blocks of length N. Assuming the knowledge of the channel impulse response, a *soft-output* symbol-by-symbol MAP equalizer computes the *a posteriori* log-likelihood ratio

$$L(b_k|r_0,\dots,r_{N-1}) \triangleq \log \frac{\Pr(b_k=+1|r_0,\dots,r_{N-1})}{\Pr(b_k=-1|r_0,\dots,r_{N-1})}, \qquad 0 \le k \le N-1\,. \tag{3.2}$$

Let $\mu_k \triangleq (b_{k-1},\dots,b_{k-L+1})$ denote the generic ISI state at time k, and $S(b_k=\hat{b}_k)$ be the set of states corresponding to the transmitted symbol $b_k=\hat{b}_k$. Indicating by ξ_k the transition from the state μ_k to μ_{k+1}, the MAP algorithm results in a *forward* and *backward* recursions with the transition metric $\lambda(\xi_k)$, coupled by a dual-maxima operation [9], [18], [12] (see Appendix A)

$$L(b_k|r_0,\dots,r_{N-1}) = \max'_{\mu_{k+1}\in S(b_k=+1)} \Lambda(\mu_{k+1}) - \max'_{\mu_{k+1}\in S(b_k=-1)} \Lambda(\mu_{k+1}) \tag{3.3}$$

$$\Lambda(\mu_{k+1}) = \Lambda^f(\mu_k) - \lambda(\xi_k) + \Lambda^b(\mu_{k+1})\,, \tag{3.4}$$

where $\Lambda(\mu_k)$ is the overall accumulated metric for the state μ_k, Λ^f and Λ^b are the accumulated metrics in the forward and backward recursions

$$\Lambda^f(\mu_k) = \max'_{\mu_{k-1}} \left\{ \Lambda^f(\mu_{k-1}) - \lambda(\xi_{k-1}) \right\} \tag{3.5}$$

$$\Lambda^b(\mu_{k+1}) = \max_{\mu_{k+2}}{}' \left\{ \Lambda^b(\mu_{k+2}) - \lambda(\xi_{k+1}) \right\} , \tag{3.6}$$

and $\max'\{x,y\} \triangleq \max\{x,y\} + \log(1 + e^{-|x-y|})$ [13], [15], [12]. Using $\max'\{x,y\} \approx \max\{x,y\}$ corresponds to the *max-log-MAP* algorithm [13]-[15] (see also [59]). Taking into account that $p(r_k|b_k, \ldots, b_{k-L+1}) = p_n(r_k - \sum_{\ell=0}^{L-1} b_{k-\ell} h_\ell^{(k)})$, the metric increment $\lambda(\xi_k)$ in (3.4)-(3.6) results

$$\lambda(\xi_k) = -\log p_n(r_k - \sum_{\ell=0}^{L-1} b_{k-\ell} h_\ell^{(k)}) - \log \Pr(b_k)$$

$$= -\log p_n(r_k - \sum_{\ell=0}^{L-1} b_{k-\ell} h_\ell^{(k)}) - \frac{1}{2} b_k L(b_k) . \tag{3.7}$$

When the equalizer receives some *a priori* information the term $-(1/2)b_k L(b_k)$ in (3.7) has a fundamental role in deriving a *soft-in/soft-out* MAP equalizer [16], [90], to be used in an iterative equalization and decoding (or *turbo* equalization) scheme. Observe that the above derivation relies on the assumption of known channel. In practice the channel is usually estimated using a known training sequence at the equalizer start-up.

In the case where n_k is modelled as AWGN, from (3.7) one obtains the conventional Euclidean distance metric

$$\lambda(\xi_k) = \frac{1}{2\sigma_n^2} \, | \, r_k - \sum_{\ell=0}^{L-1} b_{k-\ell} h_\ell^{(k)} \, |^2 - \frac{1}{2} b_k L(b_k) . \tag{3.8}$$

The block diagram of the corresponding *parametric* equalizer is shown in Figure 3.2. The equivalent optimum receiver proposed by Ungerboeck [19] uses the metric

$$\lambda(\xi_k) = -\frac{1}{2\sigma_n^2} \, \mathrm{Re}\{ \, b_k^* [\, 2z_k - b_k R_0 - 2 \sum_{\ell=1}^{L-1} b_{k-\ell} R_\ell \,] \, \} - \frac{1}{2} b_k L(b_k) , \tag{3.9}$$

where z_k denotes the matched filter output, and R_k is the channel deterministic autocorrelation. (3.9) can be viewed as the result of partitioning Forney's Euclidean distance metric into a filtering operation and a modified metric, where the filtering exactly inverts the noise whitening operation [20].

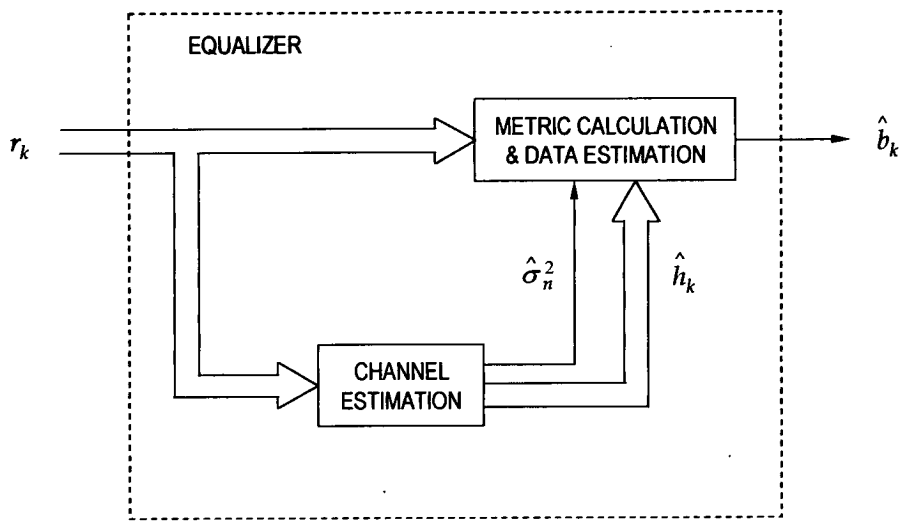

Figure 3.2: *Block diagram of the conventional parametric trellis equalizer.*

3.1.3 ML sequence estimation for finite-length ISI and additive independent disturbance

As already pointed out, the logarithmic version of the MAP algorithm is equivalent to a combination of forward and backward Viterbi processors, coupled by a dual-maxima computation [12]. Moreover, the metric (3.7) corresponds to the branch metric for MLSE implemented by the Viterbi algorithm [18]. In particular, the MLSE equalizer is obtained by retaining the forward recursion of the max-log MAP algorithm, and substituting the calculation of the log-likelihoods (3.3) by a trace-back of the symbols on the survivor path, after reaching the termination of the trellis [18]. From an intuitive point of view, this simply corresponds to replacing the MAP backward recursion with a decision delay.

3.2 Trellis equalization by non-parametric density estimation

Conventional digital receivers usually employ a trellis equalizer which treats co-channel and/or adjacent channel interference as additive, white, Gaussian noise. As discussed in the previous section, this results in the Euclidean metric (3.8) or in the modified metric (3.9), and leads to an efficient implementation of the trellis processor. In practice, however, CCI and ACI are often coloured non-Gaussian processes. Figure 3.3 shows an example of the density function of the noise-plus-CCI n_k for the case of the GSM channel. The plot has been obtained by a histogram of the data in 2000 bursts, considering one dominant interferer under stationary propagation conditions. From Figure 3.3, it is apparent that the sum of additive noise and CCI can not be realistically modelled by a Gaussian distribution. In this respect, a possible approach is to try to estimate the probability density function of the noise plus interference, instead of relying on the Gaussian assumption. With this *non-parametric* equalizer, the density function is estimated from a given number of observation data. To this purpose, one can use the known training data usually sent for the estimation of channel parameters. However, the available volume of training data is usually very limited. In this situation histogram methods generally lead to poor results. A possible approach to circumvent this problem is to resort to *kernel smoothing* techniques [67]-[69] .

3.2.1 Density estimation by kernel smoothing

3.2.1.1 Parzen estimator

An estimate of the probability density function p_X of a complex random variable X can be built from a set of data X_i, $i = 1, \ldots, n$, by means of a *smoothing function* or *kernel function* $K(x, X_i)$ (see [68], [69] and references therein). With the classical method proposed by Parzen [67], an estimate of the unknown probability density function is given by

$$\hat{p}_X(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} K(x, X_i) \,. \tag{3.10}$$

A possible choice for the function $K(x, X_i)$ is the *Gaussian kernel* of fixed width σ_0

$$K(x, X_i) = \frac{1}{2\pi\sigma_0^2} \, e^{-|x - X_i|^2 / 2\sigma_0^2} \,. \tag{3.11}$$

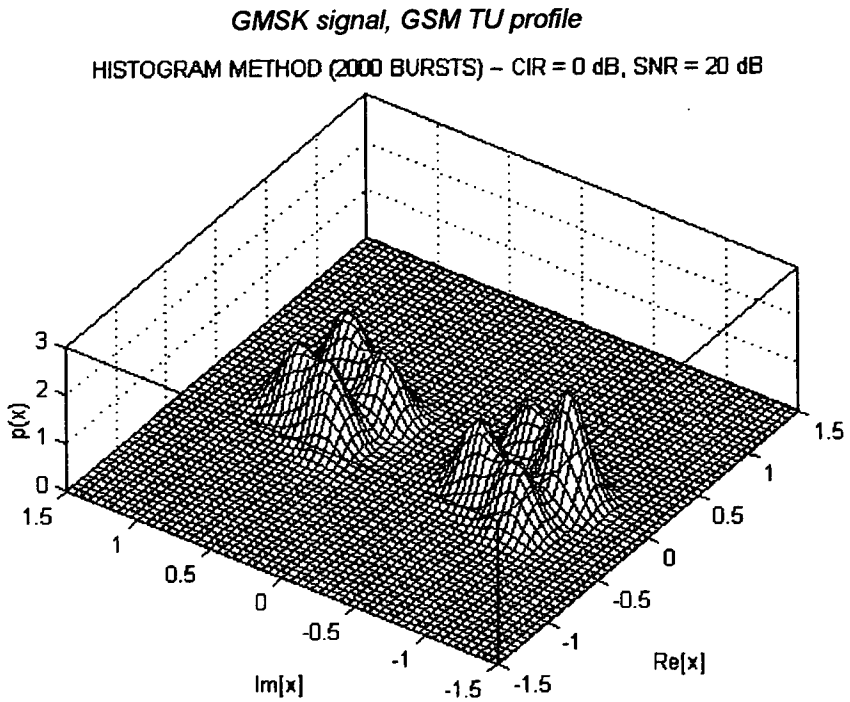

Figure 3.3: *Example of the density function of CCI (derotated GMSK signal) plus AWGN for a GSM receiver.*

The estimator (3.10) with kernel (3.11) is (asymptotically) *unbiased* (i.e., $\lim_{n\to\infty} E\{\hat{p}_X(x)\} = p_X(x)$), provided that $\sigma_0 = \sigma_0(n)$ is chosen as a function of n such that $\lim_{n\to\infty} \sigma_0(n) = 0$. Moreover, the estimator is *consistent* (i.e., $\lim_{n\to\infty} E\{|p_X(x) - \hat{p}_X(x)|^2\} = 0$), provided that $\lim_{n\to\infty} n\sigma_0(n) = \infty$ [67].

3.2.1.2 Transition metrics for non-parametric trellis equalization

In the case of a Bayesian trellis equalizer, the random variable X represents one realization of the process of noise-plus-interference corresponding to a given received burst. Consider the received signal (3.1), and assume that the channel is approximately constant within the burst duration. Then, once the channel taps h_ℓ are estimated using the N_t training symbols $\tilde{b}_i$, they can be used to derive the set of observations X_i, $i = 1, \ldots, n = N_t - L$ of the random

disturbance X according to

$$X_i = \hat{n}_i = r_i - \sum_{\ell=0}^{L-1} \tilde{b}_{i-\ell}\hat{h}_\ell \,, \tag{3.12}$$

where "^" denotes the estimated value. At this point we recall that the transition metric (3.7) of the optimum symbol-by-symbol MAP algorithm results

$$\lambda(\xi_k) = -\log \hat{p}_n(r_k - \sum_{\ell=0}^{L-1} b_{k-\ell}h_\ell) - \frac{1}{2}b_k L(b_k) \,. \tag{3.13}$$

Therefore, using (3.10) and (3.11) one can directly estimate the quantity $\log \hat{p}_X(x) = \log \hat{p}_n(x)$ for

$$x = \hat{n}_k = r_k - \sum_{\ell=0}^{L-1} b_{k-\ell}\hat{h}_\ell \,, \tag{3.14}$$

and obtain

$$\lambda(\xi_k) = -\log \hat{p}_n(x) - \frac{1}{2}b_k L(b_k) \,. \tag{3.15}$$

The conceptual difference between the operation of the proposed equalizer and the conventional parametric equalizer is illustrated in Figure 3.5. In the equalizer based on the Gaussian assumption, the observations (3.12) are used to produce an estimate of the variance which parametrizes the Gaussian density, while the proposed approach uses the same data to estimate the density function. The block diagram of the resulting equalizer is shown in Figure 3.4. From the implementation point of view, the value of the density $\log \hat{p}_n(x)$ at the point (3.14) can be computed separately for each trellis branch at a given time k. Alternatively, the density can be precomputed for a finite number of values x (e.g., at the points of an appropriate grid), and stored in a look-up-table before starting the trellis processing. Clearly, the former approach is generally preferable both in terms of complexity and accuracy. Moreover, from a conceptual point of view, it seems advisable to concentrate the effort on the particular problem of estimating the density at the points of interest, rather than trying to solve the more general problem of estimating the entire density. With respect to the computational complexity associated with the estimator (3.10), observe that other kernels than the Gaussian density (3.11) may be considered as alternatives for a practical implementation. Specifically, the use of the *Epanechnikov kernel*

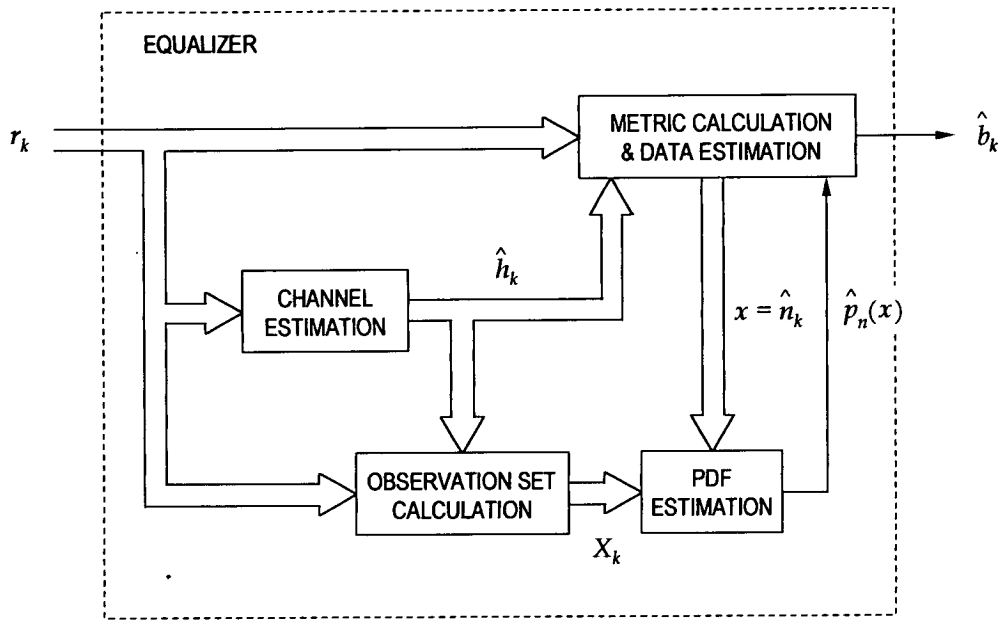

Figure 3.4: *Block diagram of the non-parametric trellis equalizer.*

[68]

$$K_E(x, X_i) = \begin{cases} \frac{2}{\pi\sigma_0^2}(1 - \frac{|x - X_i|^2}{\sigma_0^2}) & \frac{|x - X_i|^2}{\sigma_0^2} < 1 \\ 0 & \text{elsewhere} \end{cases} \tag{3.16}$$

would avoid the need of computing n exponential functions at each evaluation of (3.10). On the other hand, note that using the Gaussian kernel the quantity $\log \hat{p}_n(x)$, involving the logarithm of a sum of exponentials, can be efficiently computed using the *max-log* approximation of Section 3.1.2. Alternatively, the exponential and logarithmic functions may be implemented by means of a look-up table. Further discussion on the implementation cost of the proposed algorithm will be provided in Section 3.3.

It has to be emphasized the fact that the above technique deals with the statistical model of a random variable, obtained as the realization of the noise-plus-interference process at a given time instant. With a proper adaptive procedure, the approach can be extended to those cases

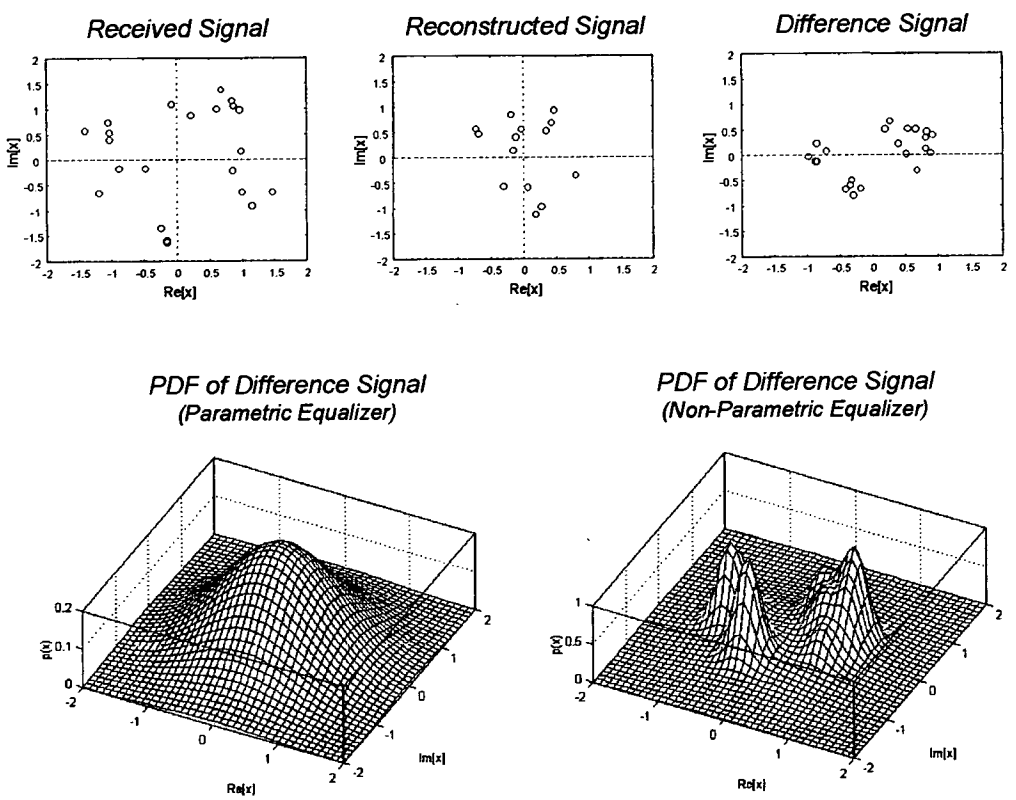

Figure 3.5: *Illustration of the operation of the non-parametric and the conventional parametric equalizer.*

where the interference can not be considered approximately stationary within the burst.

3.2.2 Gaussian mixture model of the noise-plus-interference

Although the proposed non-parametric equalizer has general applicability, and is suitable for situations where a model of the additive disturbance is not available, in this section we discuss the special case where the CCI signals have the same structure of the signal of interest. This interference model will be used to get insight in the operation of the estimator, and to obtain a specific equalizer design for the application discussed in Section 3.4, under the assumption of an (unknown) deterministic finite-state machine model for the co-channel signals. For simplicity, the analysis considers one dominant CCI signal. The result is easily extended to the general case of more than one interferer.

Consider the received signal (3.1). The sum of noise and CCI at time k can be expressed as

$$n_k = y'_k + w_k = \sum_{\ell=0}^{L'-1} b'_{k-\ell} h'^{(k)}_{\ell} + w_k \,, \tag{3.17}$$

where $b'_k \in \{+1, -1\}$ are the co-channel symbols, $h'^{(k)}_{\ell}$, $0 \leq \ell \leq L'-1$ denote the taps of the co-channel impulse response, and w_k is white Gaussian noise with zero mean and variance $2\sigma^2$, which is assumed independent of y'_k. The co-channel taps $h'^{(k)}_{\ell}$ at time k are regarded as an unknown, but deterministic mapping from $(b'_k, \ldots, b'_{k-L'+1})$ to y'_k. Hence, at a given time k, the distribution of the random variable n_k can be derived from those of b'_k and w_k. Given a generic binary quantity β, we define

$$\eta_i = \eta_{i,1} + j\eta_{i,2} \triangleq \sum_{\ell=0}^{L'-1} \beta_{i,\ell} h'_{\ell} \,, \qquad 1 \leq i \leq 2^{L'} \,, \tag{3.18}$$

where $\beta_i = \{\beta_{i,\ell}\}_{\ell=0}^{L'-1}$ denotes one of the $2^{L'}$ distinct sequences of elements $\beta_{i,\ell} \in \{+1, -1\}$. Then, the expression of the density function of n_k results

$$p_n(x) = \frac{1}{2^{L'}} \sum_{i=1}^{2^{L'}} p_w(x - \eta_i) \,, \tag{3.19}$$

where $p_w(x)$ is the complex Gaussian density with variance $2\sigma^2$. In fact, from (3.17), n_k is the sum of the two random variables y'_k and w_k, and its density $p_n(x)$ is given by the convolution of

the marginal densities $p_{y'}(x)$ and $p_w(x)$ of y'_k and w_k, respectively [75]. $p_w(x)$ is the complex Gaussian density. Hence, to obtain $p_n(x)$ one has to derive the density function $p_{y'}(x)$ of the co-channel interference. But y'_k is itself modelled as the sum of the L' independent random variables $\theta_{k,\ell} = b'_{k-\ell}h'_\ell = b'_{k-\ell}(h'_{\ell,1} + jh'_{\ell,2})$. Since the symbols b'_k are assumed to be independent, identically distributed with density $p_{b'}(x) = (1/2)[\delta(x+1) + \delta(x-1)]$, the density of $\theta_{k,\ell}$ results

$$p_{\theta_\ell}(x_1, x_2) = \frac{1}{2}[\delta(x_1 + h'_{\ell,1}) \cdot \delta(x_2 + h'_{\ell,2}) + \delta(x_1 - h'_{\ell,1}) \cdot \delta(x_2 - h'_{\ell,2})]\,, \tag{3.20}$$

where x_1 and x_2 denote the real and imaginary part of x, respectively. Then, straightforward calculation yields

$$p_{y'}(x) = p_{y'}(x_1, x_2) = \frac{1}{2^{L'}} \sum_{i=1}^{2^{L'}} \delta(x_1 - \eta_{i,1}) \cdot \delta(x_2 - \eta_{i,2})\,, \tag{3.21}$$

with $\eta_i = \eta_{i,1} + j\eta_{i,2}$ defined according to (3.18). Finally, the convolution between $p_{y'}(x)$ in (3.21) and $p_w(x)$ gives (3.19).

From (3.19), the density of noise-plus-interference is given by a number of symmetric Gaussian kernels, whose centres are the points of the hypothetical scatter diagram obtained in the absence of thermal noise. Comparison of (3.19) and (3.10) reveals the strong connection between the structure of the Parzen estimator and the true density of the disturbance (3.17). In particular, for $\sigma^2 \to 0$, the observations X_i in (3.10) correspond to the points of the complex plane defined by (3.18), with the binary parameters $\beta_{i,\ell}$ replaced by the co-channel symbols $b'_{k-\ell}$. Therefore, the estimator defined by (3.10) and (3.11) will approach the true density (3.19) as soon as the dimension of the training data is large enough to represent the $2^{L'}$ equiprobable sequences $\beta_i = \{\beta_{i,\ell}\}_{\ell=0}^{L'-1}$. From this point of view, in the presence of the Gaussian mixture noise model (3.19), the use of the Parzen estimator with Gaussian kernel could be interpreted as one way of using *a priori* knowledge on the noise distribution.

We observe that in (3.18) for each index $i = i'$ corresponding to the binary sequence $\beta_{i'} = \{\beta_{i',\ell}\}_{\ell=0}^{L'-1}$ there is an index $i = i''$ with $\beta_{i''} = \{-\beta_{i',\ell}\}_{\ell=0}^{L'-1} = -\beta_{i'}$. This means that for each i' there is an i'' such that $\eta_{i'} = -\eta_{i''}$. Hence, exchanging each pair of indexes i' and i'' in the

sum (3.19) and taking into account the symmetry of the Gaussian density $p_w(x)$ gives

$$p_n(-x) = \frac{1}{2^{L'}} \sum_{i=1}^{2^{L'}} p_w(-x+\eta_i) = \frac{1}{2^{L'}} \sum_{i=1}^{2^{L'}} p_w(x-\eta_i) = p_n(x) \,. \tag{3.22}$$

It is easy to see that an analogous result is obtained if the model (3.17) includes more than one interferer. The importance of this fact is that it effectively doubles the available volume of training data in the density estimator (3.10). Indeed, it implies that, if $\{X_i\}$ are values assumed by the random variable n_k, then the set $\{-X_i\}$ contains values assumed by n_k with the same probability. Therefore, together with each outcome X_i one can additionally consider $-X_i$ as if it was the result of a parallel experiment. This leads to the enlarged data set $\{X_i, -X_i\}$.

3.2.3 Choice of the smoothing parameter

The kernel width or smoothing parameter σ_0 in equation (3.11) can be chosen in order to minimize a given measure of inaccuracy of the density estimator. The most widely used measure of global accuracy is the mean integrated square error (MISE), defined as [68], [69]

$$\mathrm{MISE}(\hat{p}_X) \triangleq E\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [p_X(x) - \hat{p}_X(x)]^2 dx_1 dx_2 \} \,, \tag{3.23}$$

where x_1 and x_2 denote the real and imaginary part of x. Optimal smoothing in the minimum MISE sense has been studied in [68], [69]. Following a similar approach, this section provides the analysis of the MISE performance of the kernel density estimator and derives the optimum smoothing parameter for complex densities. In the case of complex data, by standard elementary properties of mean and variance, (3.23) can be rewritten as

$$\mathrm{MISE}(\hat{p}_X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [E\{\hat{p}_X(x)\} - p_X(x)]^2 dx_1 dx_2$$

$$+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E\{[\hat{p}_X(x) - E\{\hat{p}_X(x)\}]^2\} \, dx_1 dx_2 \,. \tag{3.24}$$

Hence, the MISE of the estimator is given by the sum of the integrated square bias and the integrated variance. It is convenient here to parametrize the kernel, introducing the notation

$(1/\sigma_0^2)K_1((x - X_i)/\sigma_0) = K(x, X_i)$, where $K_1(x)$ is a symmetric function defined as the product of the univariate kernels $K_1(x_1)$ and $K_1(x_2)$, and σ_0 denotes the width of the univariate kernel components. For simplicity, it is also assumed that the kernel $K_1(x)$ has mean

$$\mu_1 \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 K_1(y)dy_1dy_2 = \int_{-\infty}^{\infty} y_1 K_1(y)dy_1 = 0, \tag{3.25}$$

and variance[5]

$$\mu_2 \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1^2 K_1(y)dy_1dy_2 = \int_{-\infty}^{\infty} y_1^2 K_1(y)dy_1 \neq 0. \tag{3.26}$$

At this point, a classical approach [68], [69] is to approximate the bias and variance using Taylor series expansion, which gives

$$E\{\hat{p}_X(x)\} - p_X(x) \approx \frac{1}{2}\sigma_0^2\mu_2 [\frac{\partial^2 p_X(x)}{\partial x_1^2} + \frac{\partial^2 p_X(x)}{\partial x_2^2}] \tag{3.27}$$

$$E\{[\hat{p}_X(x) - E\{\hat{p}_X(x)\}]^2\} \approx \frac{1}{n\sigma_0^2} p_X(x) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_1^2(y)dy_1dy_2 \,. \tag{3.28}$$

To derive (3.27) one observes that the bias can be written as

$$E\{\hat{p}_X(x)\} - p_X(x) = \frac{1}{n}\sum_{i=1}^{n} E\{K(x, X_i)\} - p_X(x)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{\sigma_0^2} K_1(\frac{x-\xi}{\sigma_0}) p_X(\xi)d\xi_1d\xi_2 - p_X(x)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_1(y)[p_X(x - \sigma_0 y) - p_X(x)]dy_1dy_2 \,, \tag{3.29}$$

[5] μ_1 and μ_2 are the mean and variance of a random variable x with density function $K_1(x)$.

having taken into account that K_1 integrates to unity. Then, using the multidimensional form of Taylor's theorem, we can approximate

$$p_X(x-\sigma_0 y) \approx p_X(x) - \sigma_0 \sum_i \frac{\partial p_X(x)}{\partial x_i} y_i + \frac{1}{2}\sigma_0^2 \sum_{i,j} \frac{\partial^2 p_X(x)}{\partial x_i \partial x_j} y_i y_j \,, \tag{3.30}$$

and from (3.29)

$$E\{\hat{p}_X(x)\} - p_X(x) = -\sigma_0 \frac{\partial p_X(x)}{\partial x_1} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 K_1(y) dy_1 dy_2$$

$$- \sigma_0 \frac{\partial p_X(x)}{\partial x_2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_2 K_1(y) dy_1 dy_2 + \frac{1}{2}\sigma_0^2 \frac{\partial^2 p_X(x)}{\partial x_1^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1^2 K_1(y) dy_1 dy_2$$

$$+ \sigma_0^2 \frac{\partial^2 p_X(x)}{\partial x_1 \partial x_2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 y_2 K_1(y) dy_1 dy_2 + \frac{1}{2}\sigma_0^2 \frac{\partial^2 p_X(x)}{\partial x_2^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_2^2 K_1(y) dy_1 dy_2$$

$$= \frac{1}{2}\sigma_0^2 \mu_2 \left[\frac{\partial^2 p_X(x)}{\partial x_1^2} + \frac{\partial^2 p_X(x)}{\partial x_2^2} \right].$$

For the variance, assuming independent X_i one gets

$$E\{[\hat{p}_X(x) - E\{\hat{p}_X(x)\}]^2\} = E\{\hat{p}_X^2(x)\} - E^2\{\hat{p}_X(x)\}$$

$$= \frac{1}{n\sigma_0^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_1^2(y) p_X(x-\sigma_0 y) dy_1 dy_2 - \frac{1}{n}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_1(y) p_X(x-\sigma_0 y) dy_1 dy_2 \right]^2 \tag{3.31}$$

and approximating $p_X(x-\sigma_0 y)$ by the first term of the Taylor expansion (3.30)

$$E\{[\hat{p}_X(x) - E\{\hat{p}_X(x)\}]^2\} \approx \frac{1}{n\sigma_0^2} p_X(x) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_1^2(y) dy_1 dy_2$$

$$-\frac{1}{n}\,[\,p_X(x)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K_1(y)dy_1dy_2\,]^2$$

$$=\frac{1}{n\sigma_0^2}\,p_X(x)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K_1^2(y)dy_1dy_2-\frac{1}{n}\,p_X^2(x)$$

$$\approx\frac{1}{n\sigma_0^2}\,p_X(x)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K_1^2(y)dy_1dy_2\,.$$

From (3.28) one has that the variance is approximately proportional to the height of the true density function. Both bias and variance are affected by the smoothing parameter σ_0. As σ_0 decreases, the bias diminishes while the variance increases, and the opposite occurs as σ_0 increases [69].

Substituting now (3.27) and (3.28) in (3.24) the MISE is approximated as

$$\mathrm{MISE}(\hat{p}_X)\approx\frac{1}{4}\sigma_0^4\mu_2^2\,\Gamma(p_X)+\frac{1}{n\sigma_0^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K_1^2(y)dy_1dy_2\,, \tag{3.32}$$

where

$$\Gamma(p_X)\triangleq\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\,\frac{\partial^2 p_X(x)}{\partial x_1^2}+\frac{\partial^2 p_X(x)}{\partial x_2^2}\,]^2dx_1dx_2\,. \tag{3.33}$$

Therefore, minimizing the MISE (3.32) with respect to σ_0 yields the approximate optimal width

$$\sigma_{0(opt)}=\left(\frac{2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K_1^2(y)dy_1dy_2}{n\,\mu_2^2\Gamma(p_X)}\right)^{1/6}\,. \tag{3.34}$$

If $K_1(x)$ is the Gaussian kernel with $\mu_2=1$, one computes

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K_1^2(y)dy_1dy_2=(\,\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-y^2}dy\,)^2=\frac{1}{4\pi}\,. \tag{3.35}$$

Hence, substituting (3.35) in (3.34) we restate for the complex case the result that minimizing (3.23) with the Gaussian kernel (3.11) yields the optimum width

$$\sigma_{0(opt)} = \left(\frac{1}{2\pi n \Gamma(p_x)}\right)^{1/6} \tag{3.36}$$

with $\Gamma(p_x)$ defined in (3.33).

If $p_x(x)$ is the complex Gaussian density p_w with variance $2\sigma^2$, $\Gamma(p_x)$ results

$$\Gamma(p_w) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\, \frac{1}{\sigma^2}(\frac{x_1^2}{\sigma^2} + \frac{x_2^2}{\sigma^2} - 2)\, p_w(x)\,]^2 dx_1 dx_2 = \frac{1}{2\pi\sigma^6}\,, \tag{3.37}$$

and using (3.37) in (3.36) one obtains

$$\sigma_{0(opt)} = \left(\frac{1}{n}\right)^{1/6} \sigma\,. \tag{3.38}$$

Finally, consider the density $p_n(x)$ of the noise-plus-interference (3.17). Now, using (3.19) and applying Cauchy's inequality we derive

$$\Gamma(p_n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{\, \frac{1}{2^{L'}} \sum_{i=1}^{2^{L'}} [\, \frac{\partial^2 p_w(x-\eta_i)}{\partial x_1^2} + \frac{\partial^2 p_w(x-\eta_i)}{\partial x_2^2}\,]\, \}^2 dx_1 dx_2$$

$$\leq \frac{1}{2^{L'}} \sum_{i=1}^{2^{L'}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\, \frac{\partial^2 p_w(x-\eta_i)}{\partial x_1^2} + \frac{\partial^2 p_w(x-\eta_i)}{\partial x_2^2}\,]^2 dx_1 dx_2$$

$$= \frac{1}{2^{L'}} \sum_{i=1}^{2^{L'}} \Gamma(p_w) = \frac{1}{2\pi\sigma^6}\,. \tag{3.39}$$

Hence, from (3.36) and (3.39) follows

$$\sigma_{0(opt)} \geq \left(\frac{1}{n}\right)^{1/6} \sigma\,. \tag{3.40}$$

Note that choosing

$$\sigma_0 = k_0 \cdot \left(\frac{1}{n}\right)^{1/6} \sigma\,, \qquad 1 \leq k_0 < \infty \tag{3.41}$$

satisfies the conditions of Section 3.2.1 for unbiasedness and consistency of the estimator. For a specific application, the optimum parameter $k_0 = k_{0(opt)}$ in (3.41) can be determined if the true density function p_n is known or can be obtained numerically. Then, with a given volume n of training data, the kernel width can be selected from (3.41) according to the value of the noise variance σ^2. In a practical receiver, an estimate of σ^2 can be derived by the training symbols, taking into account the estimated channel response and the measure of the received signal level.

Using (3.36) and (3.39) in (3.32) also gives the following approximate upper bound to the minimum MISE

$$\begin{aligned} \mathrm{MISE}_{\min} &\approx \frac{1}{4}\sigma_{0(opt)}^4\,\Gamma(p_n) + \frac{1}{4\pi n\sigma_{0(opt)}^2} \\ &= \frac{3}{4}\frac{\Gamma(p_n)^{1/3}}{(2\pi n)^{2/3}} \leq \frac{3}{8\pi n^{2/3}\sigma^2}\,, \end{aligned} \tag{3.42}$$

i.e.,

$$\mathrm{MISE}_{\min} \leq \mathrm{MISE}_0 = \frac{3}{8\pi n^{2/3}\sigma^2}\,. \tag{3.43}$$

When an estimate of the optimum parameter $k_{0(opt)}$ is available, observing that from (3.36) $\Gamma(p_n) = 1/(2\pi n\sigma_{0(opt)}^6)$, the expression of the corresponding minimum MISE results

$$\mathrm{MISE}_{\min} \approx \frac{1}{4}\sigma_{0(opt)}^4\,\frac{1}{2\pi n\sigma_{0(opt)}^6} + \frac{1}{4\pi n\sigma_{0(opt)}^2} = \frac{3}{8\pi n\sigma_{0(opt)}^2}\,, \tag{3.44}$$

or, in terms of the parameter k_0 of (3.41),

$$\mathrm{MISE}_{\min} \approx \frac{1}{k_{0(opt)}^2} \cdot \mathrm{MISE}_0\,. \tag{3.45}$$

The above analysis refers the case of fixed-width kernel. Adaptive kernel density estimators which use kernels of different width for each training sample have also been proposed [68], [70]. For a study of the optimum smoothing parameter $\sigma(x)$ for adaptive kernel methods see,

e.g., [68].

3.2.4 Sufficient statistics

The analysis of Section 3.2.2 can be used to derive a simple proof that matched filtering does not provide sufficient statistics for data detection under the non-Gaussian noise model (3.17).

Consider the random variables (sequence of random variables) b, and the corresponding observation set r. If b is estimated using some function, or statistic, $\gamma(r)$ of the measurement set r, then the statistic γ is said to be *sufficient* for b if $p(b|r,\gamma) = p(b|\gamma)$, or equivalently if $p(r|b,\gamma) = p(r|\gamma)$. A way to analyse sufficiency is given by the *factorization theorem* [21], [114], [115].

Theorem 1 (Factorization Theorem) *$\gamma = \gamma(r)$ is a sufficient statistic for b if and only if*

$$p(r|b) = f(r,\gamma) \cdot g(b,\gamma)\,. \tag{3.46}$$

From the factorization theorem directly follows that, if γ is a sufficient statistic, then an invertible transformation $\psi = \psi(\gamma)$ is also a sufficient statistic [21], [114].

Example 1 (Additive Gaussian Noise) *Consider the signal model (3.1), where b_k are the data to be estimated, and n_k is an additive white Gaussian disturbance with variance $2\sigma^2$. We have*

$$p(r_0,\ldots,r_{N-1}|b_0,\ldots,b_{N-1}) = \prod_k p_n(r_k - \sum_{\ell=0}^{L-1} b_{k-\ell}h_\ell)$$

$$= \frac{1}{2\pi\sigma^2}\prod_k \exp(-\frac{1}{2\sigma^2}\,|\,r_k - \sum_{\ell=0}^{L-1} b_{k-\ell}h_\ell\,|^2)$$

$$= [\,\frac{1}{2\pi\sigma^2}\exp(-\frac{1}{2\sigma^2}\sum_k |r_k|^2)\,] \cdot [\exp(-\frac{1}{2\sigma^2}\sum_n\sum_m b_n b_m^* \sum_{\ell=0}^{L-1} h_\ell h_{\ell+n-m}^*)$$

$$\cdot \exp(\frac{1}{\sigma^2}\,\mathrm{Re}\{\sum_n b_n \sum_m r_m h_{m-n}^*\})\,]\,. \tag{3.47}$$

In (3.47), the quantities $z_k = \sum_m r_m h^*_{m-k}$, $0 \le k \le N-1$ *represent the* matched filter *output. Then, letting* $\gamma = \gamma(r_0, \dots, r_{N-1}) = \{z_k\}_{k=0}^{N-1}$, *from (3.47) we can write*

$$p(r_0, \dots, r_{N-1} | b_0, \dots, b_{N-1}) = f(r_0, \dots, r_{N-1}, \gamma) \cdot g(b_0, \dots, b_{N-1}, \gamma) \tag{3.48}$$

and from the factorization theorem we obtain the well known result that the matched filter output is a sufficient statistic for $b_0, \dots, b_{N-1}$.

Let now consider the problem of estimating the data $b_0, \dots, b_{N-1}$ from the received signal (3.1), where n_k is the non-Gaussian disturbance defined by (3.17). Assuming for simplicity $L' = 1$ in (3.17)-(3.19) (non-dispersive interference), one obtains

$$p(r_0, \dots, r_{N-1} | b_0, \dots, b_{N-1}) = \prod_k p_n(r_k - \sum_{\ell=0}^{L-1} b_{k-\ell} h_\ell)$$

$$= \frac{1}{4\pi\sigma^2} \prod_k [\exp(-\frac{1}{2\sigma^2} | r_k - \sum_{\ell=0}^{L-1} b_{k-\ell} h_\ell + h'_0 |^2)$$

$$+ \exp(-\frac{1}{2\sigma^2} | r_k - \sum_{\ell=0}^{L-1} b_{k-\ell} h_\ell - h'_0 |^2)]$$

$$= \frac{1}{2\pi\sigma^2} \exp(-\frac{1}{2\sigma^2} \sum_k |r_k|^2) \cdot \exp(-\frac{1}{2\sigma^2} \sum_n \sum_m b_n b^*_m \sum_{\ell=0}^{L-1} h_\ell h^*_{\ell+n-m})$$

$$\cdot \exp(\frac{1}{\sigma^2} \mathrm{Re}\{\sum_n b_n \sum_m r_m h^*_{m-n}\}) \cdot \exp(-\frac{1}{2\sigma^2} N |h'_0|^2)$$

$$\cdot \prod_k \cosh(\frac{1}{\sigma^2} \mathrm{Re}\{(r_k - \sum_{\ell=0}^{L-1} b_{k-\ell} h_\ell) h'^*_0\}) \,. \tag{3.49}$$

Inspection of (3.49) shows that the statistics $\gamma = \gamma(r_0, \dots, r_{N-1}) = \{z_k = \sum_m r_m h^*_{m-k}\}_{k=0}^{N-1}$ no longer allow the factorization (3.48). Therefore, from the factorization theorem, we conclude that for the non-Gaussian problem in consideration the matched filter output is not a sufficient

statistic for $b_0, \ldots, b_{N-1}$.

3.2.5 Temporal whitening

The trellis equalizer with branch metric (3.15) is based on the assumption that the samples n_k at the output of the receive filter are independent. Given a temporally coloured CCI, we consider the use of a whitening filter of the disturbance before the trellis processor. MAP and ML equalization in the presence of ISI and correlated noise has been studied e.g. in [116], by modelling the disturbance as a finite-order Markov process. The trellis branch metrics derived in [116] in the case of (signal-dependent) additive Gauss-Markov noise correspond to first uncorrelating the noise with a finite impulse response filter, and then applying the Euclidean metric to the filter output. Here, let $R_{n(\ell)} \triangleq E\{n_k n^*_{k-\ell}\} = R^*_{n(-\ell)}$ denote the autocorrelation of n_k. If the z-transform of $R_{n(\ell)}$ admits the spectral factorization $S_n(z) = S^2_{n0} \cdot G(z)G^*(1/z^*)$, with $G(z)$ causal, monic, and minimum-phase (*canonical* filter response), one can define the whitening filter $A(z) \triangleq S_{n0} \cdot 1/G(z)$ [17] (see also [117]). Since $G(z)$ is causal and minimum-phase, $A(z)$ is causal and stable. Being $A(z)$ an invertible transformation, there is no loss of information, i.e. the output of the whitening filter is still a sufficient statistic. In practice, the whitening filter $A(z)$ can be implemented by a *linear prediction-error* (LPE) filter. We point out that the LPE filter will ideally produce uncorrelated CCI-plus-noise samples [119], but this does not necessarily imply independence, since the process continues in general to be non-Gaussian. However, independent noise is often assumed for signal detection in uncorrelated non-Gaussian noise, and in certain cases this results a good approximation [87]. The use of a whitening filter is also supported by the observation that reducing the correlation between the samples will certainly reduce their *dependence*. Note that in some particular cases the whitened disturbance turns out to be actually independent. As an example, this happens when the variance of the thermal noise tends to zero and the co-channel is minimum-phase (in fact, in this case the ideal LPE filter inverts the co-channel). From the implementation point of view, it should be taken into account that a whitening filter for the disturbance will inevitably increase the channel memory for the desired signal. Hence in order not to increase the number of states of the equalizer, the number of taps of the LPE filter has to be kept small. In other terms, the problem is that of a trade-off between reducing the colour of the input disturbance and limiting the complexity of the ISI trellis.

3.3 Application to a TDMA cellular radio receiver

This section considers the application of non-parametric trellis equalization to a TDMA radio receiver. Numerical results will be provided in the next section for the GSM system. GSM is a TDMA system with 200 kHz carrier spacings, where a TDMA frame on a given channel is divided into eight time slots, each one reserved to one user [120]. In TDMA cellular systems, the encoded and modulated data are transmitted within each time slot according to a specified burst format, containing a training sequence midamble [121]. The transmitted signal is often substantially bandlimited with bandwidth $B \cong 1/2T$. As an example, the GMSK signal of GSM has almost zero excess bandwidth [76]. Therefore, sufficient statistics can be obtained by Baud-rate sampling at the output of a low-pass filter with bandwidth $B_F = 1/2T$. As discussed in Section 2.2, in the case of the GSM receiver, we consider the linearized model of the GMSK signal, and assume that (3.1) is obtained by derotation of the received samples.

With the typical multipath fading of mobile communications, the discrete-time process y'_k of (3.17) is generally coloured [76]. Therefore, the receiver should include a digital LPE filter to whiten the disturbance at the input of the equalizer. In a conventional TDMA receiver, the channel is estimated by correlating the received signal r_k with the known training sequence bits [33], [59]. Here, we implement a joint ML estimation of the LPE filter coefficients vector $\mathbf{a} \triangleq [a_0 = 1 \; a_1 = -\alpha_1 \cdots \; a_P = -\alpha_P]$ and the equivalent channel taps vector $\boldsymbol{\zeta} \triangleq [\zeta_0 \; \zeta_1 \cdots \; \zeta_{L+P-1}]$, where $\zeta_\ell \triangleq \sum_i a_i h_{\ell-i}$ (see, e.g., [113]). Denoting by r_k the received signal samples at the input of the LPE filter, the approach assumes the model

$$\sum_{i=0}^{P} a_i r_{k-i} = \sum_{\ell=0}^{L+P-1} b_{k-\ell}\zeta_\ell + \varepsilon_k \,, \tag{3.50}$$

where the term $\sum_{i=0}^{P} a_i r_{k-i}$ denotes the LPE filter response to the received signal, while $\sum_{\ell=0}^{L+P-1} b_{k-\ell}\zeta_\ell$ indicates the channel plus LPE filter response to the transmitted symbols b_k, and ε_k is a zero-mean white Gaussian process with variance $2\sigma^2$. Since from (3.1) $r_k = y_k + n_k$, the model (3.50) corresponds to assuming

$$\sum_{i=0}^{P} a_i y_{k-i} = \sum_{i=0}^{P} a_i \sum_{\ell=0}^{L-1} b_{k-i-\ell} h_\ell = \sum_{\ell=0}^{L+P-1} b_{k-\ell}\zeta_\ell \tag{3.51}$$

and

$$\sum_{i=0}^{P} a_i n_{k-i} = \varepsilon_k \,, \tag{3.52}$$

with ε_k white (see Figure 3.6). Therefore, the effect of the LPE filter is that of whitening the additive disturbance n_k. The formulation (3.50) permits the description of the channel plus whitening filter as a vector inner product, which in turn allows the simultaneous estimation of the LPE coefficients and the equivalent channel taps at the output of the LPE filter [113]. In fact, letting $\boldsymbol{\alpha} \triangleq [\alpha_1 \cdots \alpha_P]$, (3.50) can be rewritten as

$$\begin{bmatrix} 1 & -\alpha_1 & \cdots & -\alpha_P \end{bmatrix} \begin{bmatrix} r_k \\ r_{k-1} \\ \vdots \\ r_{k-P} \end{bmatrix} = \begin{bmatrix} \zeta_0 & \zeta_1 & \cdots & \zeta_{L+P-1} \end{bmatrix} \begin{bmatrix} b_k \\ b_{k-1} \\ \vdots \\ b_{k-L-P+1} \end{bmatrix} + \varepsilon_k \,,$$

or equivalently

$$r_k = \begin{bmatrix} \alpha_1 & \cdots & \alpha_P \end{bmatrix} \begin{bmatrix} r_{k-1} \\ \vdots \\ r_{k-P} \end{bmatrix} + \begin{bmatrix} \zeta_0 & \zeta_1 & \cdots & \zeta_{L+P-1} \end{bmatrix} \begin{bmatrix} b_k \\ b_{k-1} \\ \vdots \\ b_{k-L-P+1} \end{bmatrix} + \varepsilon_k$$

$$= [\boldsymbol{\alpha} \;\; \boldsymbol{\zeta}]\, \mathbf{v}_k + \varepsilon_k \,, \tag{3.53}$$

where $\mathbf{v}_k \triangleq [r_{k-1} \cdots r_{k-P} \; b_{k-1} \cdots b_{k-L-P+1}]^T$. Given the Gaussian assumption on ε_k, the likelihood function of $r_0, \ldots, r_{N-1}$ given $[\boldsymbol{\alpha} \;\; \boldsymbol{\zeta}]$ and $\mathbf{v}_0, \ldots, \mathbf{v}_{N-1}$ results

$$-\log p(r_0, \ldots, r_{N-1} \,|\, [\boldsymbol{\alpha} \;\; \boldsymbol{\zeta}], \mathbf{v}_0, \ldots, \mathbf{v}_{N-1}) = -\log p(\varepsilon_0, \ldots, \varepsilon_{N-1})$$

$$= \frac{1}{2\sigma^2} \sum_{k=0}^{N-1} |r_k - [\boldsymbol{\alpha} \;\; \boldsymbol{\zeta}]\, \mathbf{v}_k|^2 \tag{3.54}$$

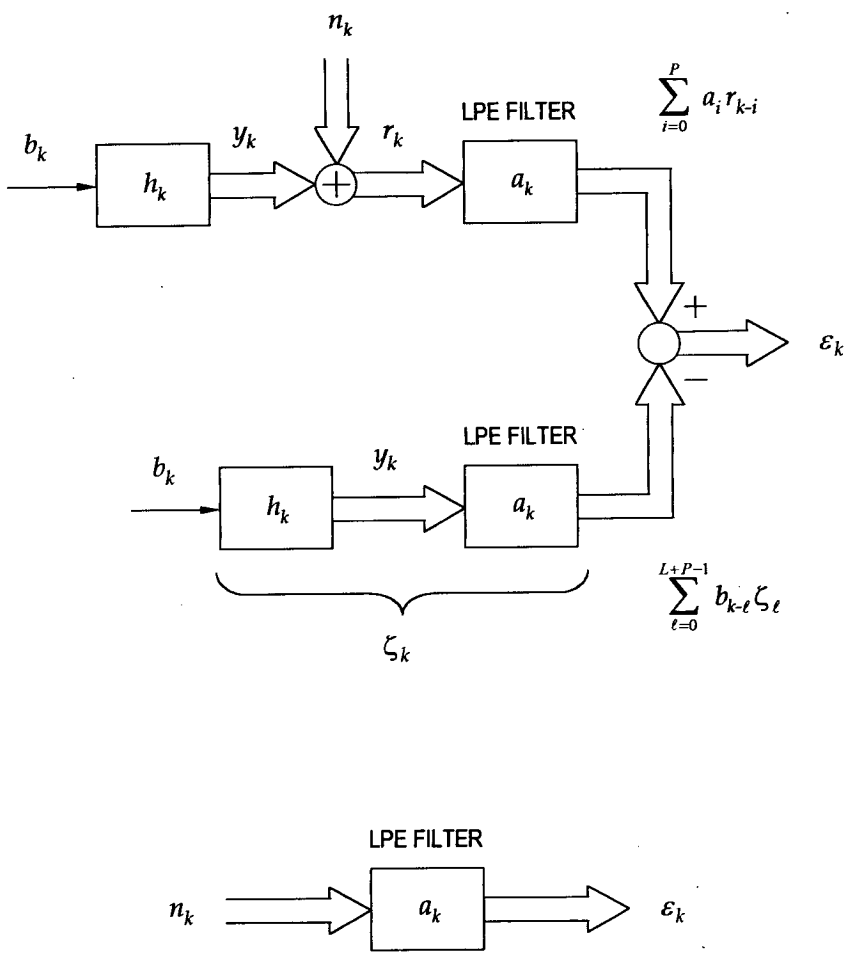

Figure 3.6: *Equivalent representations of the operation of the whitening filter.*

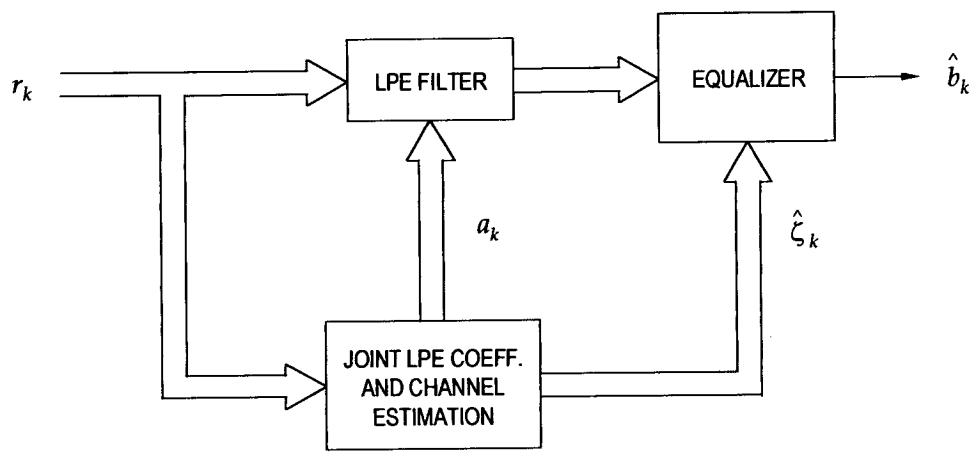

Figure 3.7: *Structure of the receiver with prewhitening filter for the interference plus thermal noise.*

(having neglected the constant term $N \log 2\pi\sigma^2$). Therefore, assuming the knowledge of $\mathbf{v}_0$, $\ldots$, $\mathbf{v}_{N-1}$, the ML estimate $[\hat{\alpha}\ \hat{\zeta}]$ of the vector $[\alpha\ \zeta]$ can be obtained by minimizing (3.54) with respect to $[\alpha\ \zeta]$. This corresponds to the least-squares estimation of the unknown parameters α and ζ. Defining

$$\mathbf{R} \triangleq \frac{1}{N} \sum_{i=0}^{N-1} \mathbf{v}_i^* \mathbf{v}_i^T \tag{3.55}$$

$$\mathbf{p} \triangleq \frac{1}{N} \sum_{i=0}^{N-1} \mathbf{v}_i^* r_i\,, \tag{3.56}$$

and introducing a regularization term with *regularization coefficient* $\lambda \geq 0$ [122] (see also

Section 4.1.2), one obtains

$$[\hat{\alpha}\ \hat{\zeta}]^T = (\mathbf{R} + \lambda \mathbf{I})^{-1}\mathbf{p}\,. \tag{3.57}$$

In the receiver under study, the conventional equalizer then uses the LPE filter output and the estimated channel $\hat{\zeta}$ derived from (3.57) to compute the metrics (3.8) or (3.9). The complete algorithm with the non-parametric equalizer of Section 3.2 is summarized in Table 3.1. In this case, the filter output and the channel estimate are employed to derive the observation set (3.12) for kernel density estimation. For a GSM receiver, given the $N_t = 26$ training bits, the procedure results in $n = 22$ training data for the density estimator. This data set can be doubled using the result of Section 3.2.2.

It is worth pointing out that an enlarged training set increases the implementation complexity of the non-parametric equalizer. In the following, we discuss the complexity associated with the calculation of the log-MAP branch metrics for equalization of a GSM burst, and compare the computational cost of the proposed algorithm with the conventional parametric equalizer and the optimum equalizer based on joint detection. The complexity of the metric calculations for the parametric equalizer is written as $C = C_0\, 2^L$, where C_0 is an implementation dependent quantity. In the presence of M co-channel GSM signals with channel length L, the corresponding cost for the joint-MAP equalizer results approximately proportional to $MC_0\, 2^{ML} = M2^{(M-1)L}\, C$. For the non-parametric equalizer with Gaussian kernel, given the estimated channel, kernel width, and observation set (3.12) (which are derived from the burst midamble before starting the trellis processing), the computation of the quantities $|x - X_i|^2 / 2\sigma_0^2$ in (3.11) requires $2L + 3$ multiplications for each trellis branch. In the case where the exponential in the kernel function (3.11) and the logarithm in the metric (3.15) are implemented by a look-up table, the metrics $-\log \hat{p}_n(x)$ of (3.15) require therefore $n\,(2L+3)\,2^L$ multiplications per trellis stage. Taking into account also operations of table look-up and additions, the overall metrics complexity for one burst can be estimated in $\kappa \cdot n\, C_0\, 2^L = \kappa \cdot nC$, with $\kappa \approx 0.5 - 1$.

From the above estimates, the metrics complexity for the proposed equalizer is lower than for joint detection if $n < M2^{(M-1)L}$. Furthermore, the joint detection approach requires an additional cost for the calculation of the accumulated metrics and the soft-output values, which increases exponentially with M.

TABLE 3.1

SUMMARY OF THE NON-PARAMETRIC ALGORITHM WITH COLOURED DISTURBANCE

NOISE PREWHITENING AND CHANNEL ESTIMATION

STEP 1: compute matrix (3.55) and vector (3.56) using received signal r_k and known training symbols

STEP 2: jointly estimate LPE filter coefficients a_k and equivalent channel taps ζ_k using (3.57)

STEP 3: compute LPE filter output $\sum_{i=0}^{P} a_i r_{k-i}$

MAP TRELLIS EQUALIZATION

STEP 4: compute observation set (3.12) for kernel density estimation using signal from STEP 3, channel estimate from STEP 2 and known training symbols

STEP 5: for each trellis branch, evaluate (3.14) using signal from STEP 3 and channel estimate from STEP 2

STEP 6: compute metric increments (3.15) using (3.10), and store accumulated metrics

STEP 7: compute *a posteriori* probabilities (3.2) using (3.3)-(3.6)

The complexity of the optimal joint-MAP equalizer may be significantly reduced by suboptimal approaches based on a reduced-state trellis. State reduction has been achieved by channel memory truncation [98], or by the *multi-stage* joint detection scheme of [61], based on successive interference cancellation. With the latter technique, the complexity grows linearly instead of exponentially with M. However, its performance is affected by error propagation in the case of comparable received power of the different co-channel signals [61]. As already observed, a reduced complexity non-parametric equalizer may be obtained by using a kernel function different from (3.11). A cost reduction may also be achieved by performing preclustering of the training data (which is equivalent to exploring the multimodality of the estimated density) [99], [100] (see also [123]).

3.4 Simulation results

The effectiveness of the strategy based on density estimation by kernel smoothing has been assessed by computer simulation for the case of a GSM receiver with single channel reception. The GMSK transmitted symbols are obtained from the source bits by rate 1/2 convolutional encoding and interleaving, according to the GSM specifications for the full-rate speech traffic channel [125]. The simulator includes the frequency selective, multipath fading channel with the classical Doppler spectrum [76], CCI, and white Gaussian noise. Ideal frequency hopping is implemented. One dominant co-channel interferer is assumed, characterized by an independent fading process. The noise and interference powers are identified by the signal-to-noise ratio (SNR) and the signal-to-interference ratio (SIR), respectively. At the receiver, the soft-output data produced by a 16-states log-MAP equalizer are deinterleaved and decoded by a Viterbi convolutional channel decoder.

To establish the ultimate performance of the proposed equalizer, we first consider the ideal case of known channel and relative speed 0 km/h. The GSM typical urban area (TU) multipath profile [76] is assumed for both co-channel signals. The accuracy of the density estimator (3.10) in terms of minimum MISE is studied in Figures 3.8 and 3.9. The MISE curves for different SIRs suggest that a proper choice of the kernel width is obtained by taking $k_0 \in (1.5, 3.5)$ in (3.41). In particular, for a fixed SNR the minimum MISE is given by values of k_0 which decrease with increasing SIRs. This fact is in agreement with the analysis of Section 3.2.3. In fact, for SIR $\rightarrow \infty$ the noise-plus-CCI approaches the Gaussian distribution, and from (3.36), (3.39), and (3.41) follows that $k_{0(opt)} = 1/(2\pi\sigma^6\Gamma(p_n))^{1/6} \rightarrow 1$. Figures 3.8 and 3.9 also

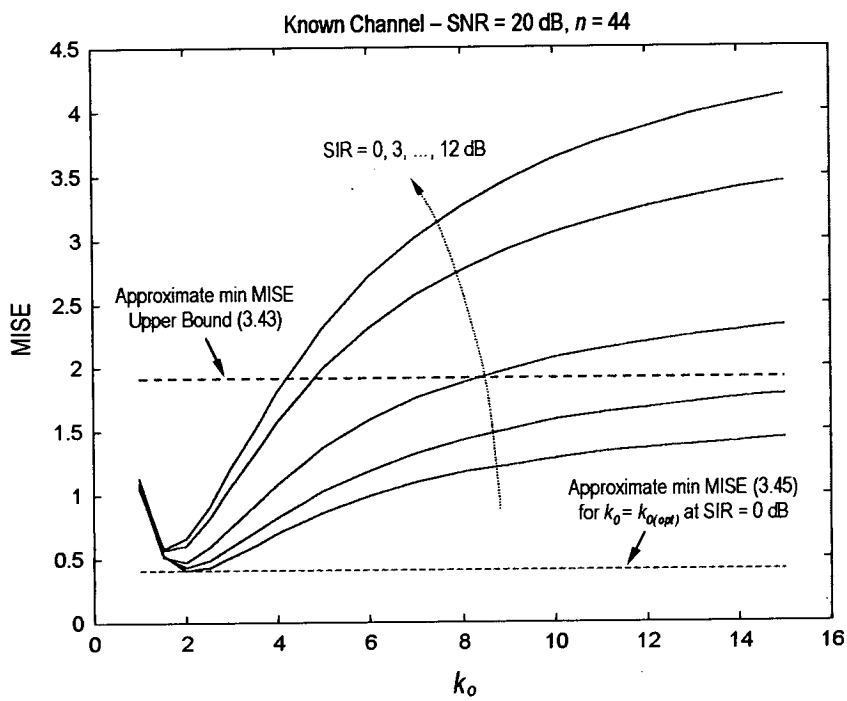

Figure 3.8: *Simulated MISE of the estimated density in the case of known channel, as a function of the kernel width parameter k_0 of equation (3.41). GSM TU0 profile, SNR = 20 dB.*

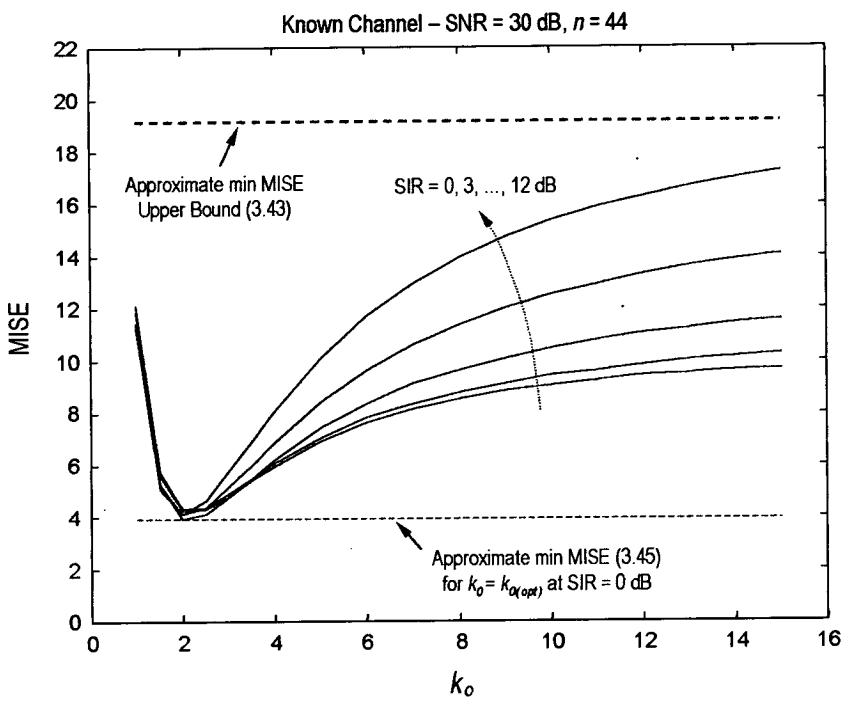

Figure 3.9: *Simulated MISE of the estimated density in the case of known channel, as a function of the kernel width parameter k_0 of equation (3.41). GSM TU0 profile, SNR = 30 dB.*

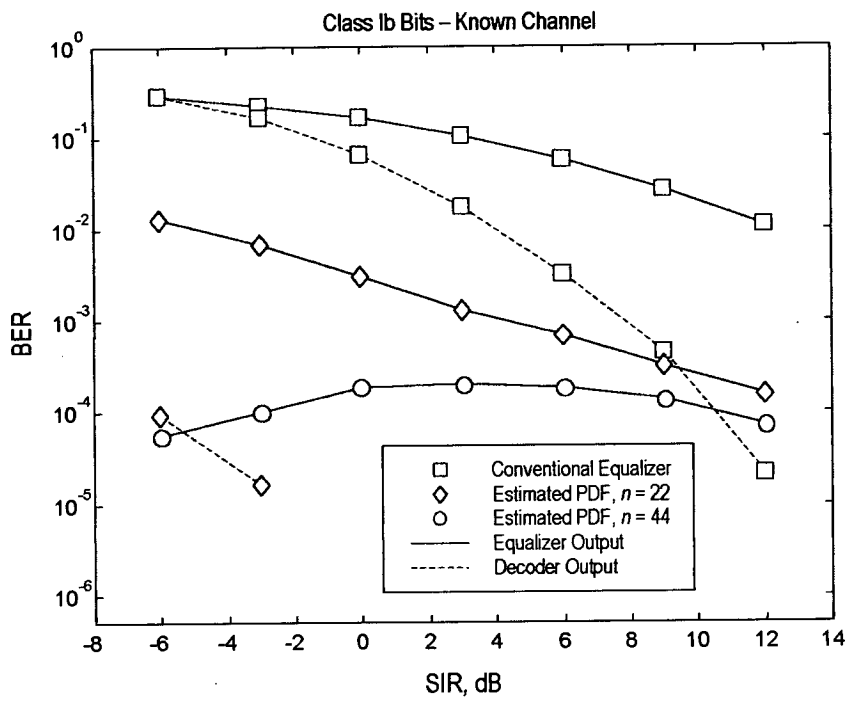

Figure 3.10: *Error performance in the case of known channel. GSM TU0 profile, SNR* = 30 *dB.*

report the minimum MISE calculated from (3.45) with the estimated $k_{0(opt)}$ at SIR = 0 dB. The calculated values are seen to be in good agreement with the minimum MISE obtained by simulation. Figure 3.10 shows the bit-error rate (BER) performance of the proposed receiver in the case of known channel for the signal of interest and SNR = 30 dB. The MAP non-parametric equalizer is compared with the MAP trellis processor that assumes additive white Gaussian disturbance. The parametric equalizer uses the metric (3.9), and operates directly on the output of the matched filter. The figure also addresses the effect of doubling the data set for density estimation, as discussed in Section 3.2. The error rate curve for $n = 44$ is seen to increase around the singular point SIR = 3 dB. This is due to the fact that at this particular level of interference the CCI-plus-noise densities conditioned to different transmitted symbols may partially overlap, depending on the phase difference between the desired and co-channel signals. This situation corresponds to the occurrence of comparable metric increments for different branches of the ISI trellis. The results of Figure 3.10 indicate that the non-parametric equalizer offers a potential improvement of more than two orders of magnitude in terms of BER

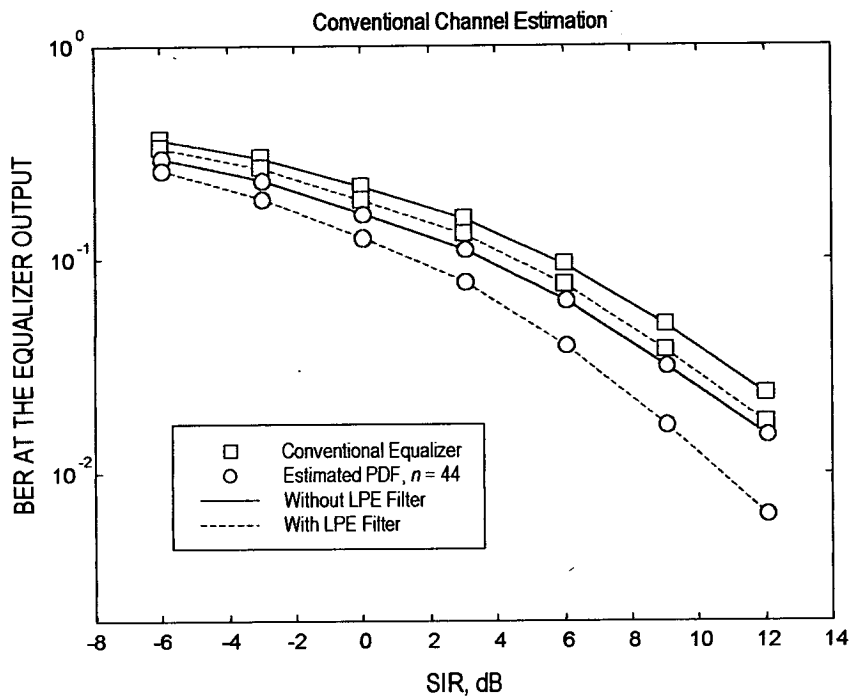

Figure 3.11: *Error performance in the case of estimated channel. GSM TU0 profile, SNR* $= 30$ *dB. Density estimator with fixed kernel width* $\sigma_0 = 0.05$.

at the equalizer output. For the non-parametric equalizer with $n = 44$, from the simulation of 16000 bursts per BER point no error has been detected at the corresponding output of the channel decoder.

Figures 3.11 to 3.14 illustrate the equalizer and decoder performance when the channel of the signal of interest is estimated from the training symbols. In these simulations, we also introduce an LPE filter for prewhitening of the coloured disturbance. As discussed in Section 3.2, choosing the prediction order involves a trade-off between performance and complexity. In the figures, a 16-states trellis and a 2-taps LPE filter are used for both the parametric and non-parametric equalizer (see also [112]). The regularization coefficient λ of (3.57) is fixed to 10^{-2}. The plots refer to the case $n = 44$, which consistently provides a gain of about 0.5 dB with respect to using the original training set size $n = 22$. Finally, Figures 3.13 and 3.14 include the performance obtained by iterative channel estimation [59]. In this case, after the equalization of the entire burst, the data decisions are fed back to produce an improved channel estimate,

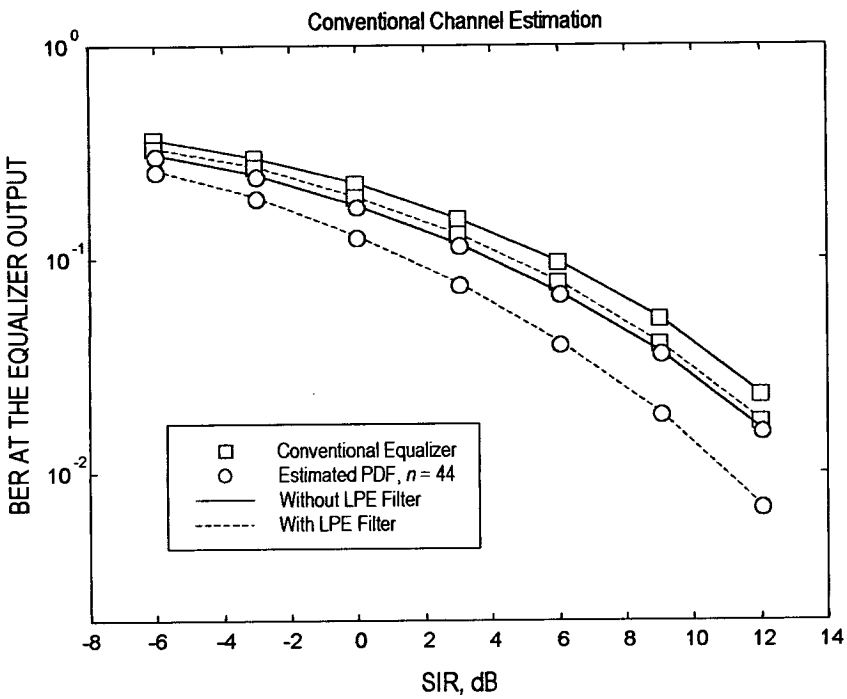

Figure 3.12: *Error performance in the case of estimated channel. GSM TU50 profile, SNR* $= 30$ *dB. Density estimator with fixed kernel width* $\sigma_0 = 0.05$.

which is used in a second pass equalization. One observes that at low SIRs the performance of all the equalizers is affected by the error associated with the channel estimate. However, the collected results show that the non-parametric approach still provides a performance gain of about 4 dB at a BER = 10^{-2} at the output of the channel decoder.

Notice that the performance of both the channel and density estimators may be further improved by the use of decision-feedback from the decoder output. The latter approach results particularly attractive for receivers implementing iterative equalization and decoding [59]. As discussed in [59], the above iterative channel estimation schemes imply an increased implementation complexity. In particular, the use of feedback from the output of the channel decoder may cause latency problems with the 8-burst interleaving scheme of GSM [125]. This problem is mitigated in third-generation TDMA cellular standards, which employ a reduced interleaving depth [59]. Further study should investigate improved channel estimation techniques in the presence of non-Gaussian noise.

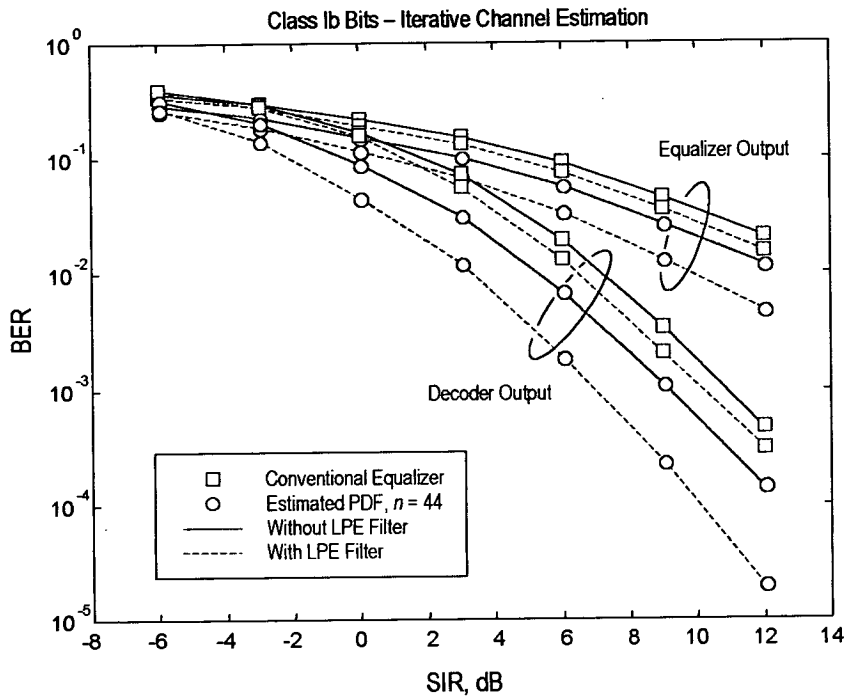

Figure 3.13: *Error performance with iterative channel estimation. GSM TU0 profile, SNR* $= 30$ *dB. Density estimator with fixed kernel width* $\sigma_0 = 0.05$.

The above simulations refer to a synchronous interference scenario. This model is reasonably accurate in some practical situations, such as a GSM system with synchronized base stations and small cell size. Simulation with asynchronous interference shows that the proposed non-parametric equalizer still outperforms the conventional trellis processor. However, in the presence of asynchronous CCI, the proper approach consists in introducing an adaptation of the estimated density. In this respect, further research should address the study of the performance/complexity trade-off of schemes based on adaptive density estimation.

3.5 Concluding remarks

A probabilistic trellis processor has been studied for trellis equalization of the ISI channel in the presence of non-Gaussian interference. The proposed approach is based on the *non-parametric* estimation of the density function of CCI-plus-noise by kernel smoothing techniques. Major

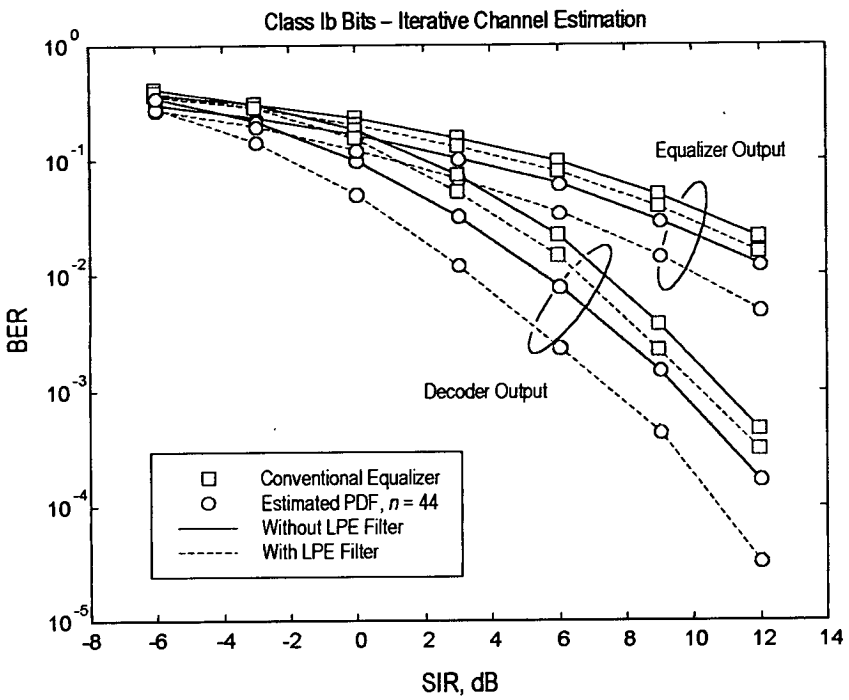

Figure 3.14: *Error performance with iterative channel estimation. GSM TU50 profile, SNR* $= 30$ *dB. Density estimator with fixed kernel width* $\sigma_0 = 0.05$.

advantages of this technique are its intrinsic robustness and general applicability to those cases where there are difficulties in accurate modelling the interference. The chapter has considered the design of the optimum smoothing parameter for complex densities, and provided an analysis of the MISE performance of the density estimator. With respect to the problem of density estimation with the availability of a small volume of data, symmetry conditions have been identified that allow to actually double the size of the training set. The problem of deriving sufficient statistics for data detection in the presence of non-Gaussian correlated noise has also been discussed, and the use of a whitening filter for the disturbance has been studied. The implementation cost of the proposed equalizer has been compared to that of the conventional parametric equalizer and the optimum equalizer based on joint detection, using explicit complexity formulas as a function of the algorithm parameters.

Simulation results have been reported for the GSM system, showing that despite the limited volume of training data, the proposed strategy can provide a significant improvement of the er-

ror performance with respect to the receiver that assumes Gaussian disturbance. The proposed *non-parametric* technique can be applied to trellis processors implementing symbol-by-symbol MAP estimation or ML sequence estimation. In the case of a MAP equalizer, the collected data show that estimating the actual statistics of the non-Gaussian disturbance allows for an additional gain in terms of improved quality of the output reliability information for soft-decision decoding. The above MAP equalizer is therefore particularly suitable for application as an inner *soft-in/soft-out* module in iterative equalization and decoding.

Chapter 4
Soft-output space-time equalization

The use of antenna arrays can help to combat co-channel interference (CCI) in wireless digital mobile systems. In interference-limited scenarios, *spatio-temporal filtering* techniques [44]-[47], [110] can be competitive with respect to trellis based algorithms, because of the difficulties of accurately modelling the space-time characteristics of the interference, especially for asynchronous cellular systems [48].

This chapter considers a diversity receiver for the Enhanced Data rates for GSM Evolution (EDGE) system, which is being standardized as an high-data rate evolution of GSM and TDMA/136. The new system will provide data rates significantly higher than GSM through the use of an 8-PSK modulation scheme instead of GMSK, and link adaptation implemented by *adaptive modulation and coding* (AMC). AMC transmission will support different modulation and coding schemes (with the GMSK format retained for the lower rate transmissions), dynamically selecting the modulation and code rate to match the varying link quality [35], [81]. The EDGE standardization activities within the European Telecommunications Standards Institute (ETSI) are considering both packet- and circuit-switched based systems: Enhanced General Packet Radio Service (EGPRS) and Enhanced Circuit Switched Data (ECSD) [35]. For EDGE in 8-PSK modes, due to the increased dimensionality of the modulation symbol alphabet and the corresponding exponential increase of the number of states of the ISI trellis, the implementation of optimum trellis equalizers becomes unrealistic. Therefore, linear or decision feedback filtering techniques, or reduced-state trellis equalizers have to be considered.

In digital receivers employing the concatenation of an equalizer and a channel decoder, the performance is improved by soft-decision decoding and iterative equalization and decoding [57], [59]. In this respect, as we have seen in the previous chapters, MAP trellis equalizers [9] have the advantage of intrinsically providing optimal *a posteriori* probability as a soft-output value delivering reliability information to the channel decoder. On the other hand, in a diversity receiver employing (nonprobabilistic) linear or decision-feedback space-time filtering, this reliability information is not available, and has to be computed by a suitable post-processing.

This fact, not often addressed in previous studies, has a significant impact on the decoder performance.

Space-time equalization for multiple antenna receivers in the presence of CCI and intersymbol interference (ISI) have been widely studied in the recent literature (see, e.g., [44], [45] and references therein). In particular, [46] provides the analysis of receiver structures based on space-time linear filtering followed by a decision-feedback section or a (single-input) trellis processor. In the latter case, the space time front-end filtering aims to maximize the signal-to-CCI plus noise ratio at the input of the trellis equalizer (without suppressing ISI), while whitening the CCI and noise components [46]. The asymptotic performance (for infinite-length filters) of the space-time receiver that uses a full state trellis processor after space-time filtering is only marginally superior to that of an MMSE decision-feedback equalizer (DFE) [46]. A diversity receiver for joint convolutional coding and decision-feedback equalization has been recently proposed [126], which employs a space-time DFE with soft decision feedback. However, in [126] the DFE output is directly fed to the channel decoder.

In order to fully exploit the potential advantage of the coding scheme, the decoder should be provided with (an approximation to) the *a posteriori* probability on the coded bits, derived from the output of the space time equalizer. The present chapter studies 8-PSK EGPRS transmission schemes with interleaving and convolutional coding [125], and with a diversity receiver employing linear and decision-feedback space-time filtering based on the deterministic *least-squares* (LS) criterion. The chapter discusses and quantifies by computer simulation the effect on the receiver performance of the quality of the soft-output values (reliability information on the coded bits) delivered by the space-time equalizer. In particular, the study compares the performance of soft values based on *burst statistics* and *short-term statistics* of the equalizer output error. This work considers the implementation of regularized training-based algorithms for space-time equalization. However, the soft output calculation can be applied to enhanced space-time processors based on a semi-blind or training-like approach (see [110], [48] and references therein).

The chapter is organized as follows. Section 4.1 introduces the system model and formalizes the problem of regularized least-squares space-time equalization. Section 4.2 studies the derivation of the reliability information at the output of the deterministic equalizer. The method is based on the computation of the sample variance of the output disturbance. The same section also addresses the use of the above soft-output calculator in conjunction with soft-decision feed-

back [74], [128], [129] and derives the optimum (mean-square) soft-decision device for 8-PSK symbols. The system performance is analysed by computer simulation in Section 4.3. Finally, conclusions are drawn in Section 4.4.

4.1 Space-time equalizer

4.1.1 System model

The EGPRS system reuses the GSM carrier spacing of 200 kHz and the symbol rate of 271 kHz. The transmitted data are protected by a rate 1/3, constraint length 7 punctured convolutional code, with block interleaving over 4 bursts (20 msec) [125]. As in GSM, the burst format includes a 26 symbols training sequence midamble, 2×58 data symbols, 2×3 tail and 8.25 guard symbols.

We assume that $P+1$ co-channel signals are transmitted over independently fading multipath channels to the M-branch diversity receiver shown in Figure 4.1. The received signal at the m-th antenna, $1 \leq m \leq M$, can be expressed as

$$r_m(k) = x_m(k) + n_m(k) = \sum_{\ell=0}^{\nu-1} d^{(0)}(k-\ell) h_m^{(0)}(k,\ell) + n_m(k)\,, \tag{4.1}$$

where $x_m(k)$ denotes the signal of interest with data symbols $d(k) = d^{(0)}(k)$, and

$$n_m(k) = \sum_{p=1}^{P} \sum_{\ell=0}^{\nu-1} d^{(p)}(k-\ell) h_m^{(p)}(k,\ell) + w_m(k) \tag{4.2}$$

represents the sum of noise and CCI. The noise $w_m(k)$ is modelled as a complex zero mean Gaussian random process with variance σ_w^2. This chapter considers the case where the transmitted data $d^{(p)}(k)$, $0 \leq p \leq P$ for the p-th co-channel user are taken from an 8-PSK symbol alphabet. The complex tap-gains $h_m^{(p)}(k,\ell)$, $0 \leq \ell \leq \nu-1$ denote the equivalent discrete-time impulse response for the p-th co-channel. It is assumed that each burst of data is received under quasi-stationary propagation conditions, so that $h_m^{(p)}(k,\ell) \approx h_m^{(p)}(\ell)$. Both synchronous and asynchronous interference scenarios are considered, with co-channel signals characterized by independent multipath fading processes.

Indicating by L the time span of the space-time receiver in number of symbol intervals, one

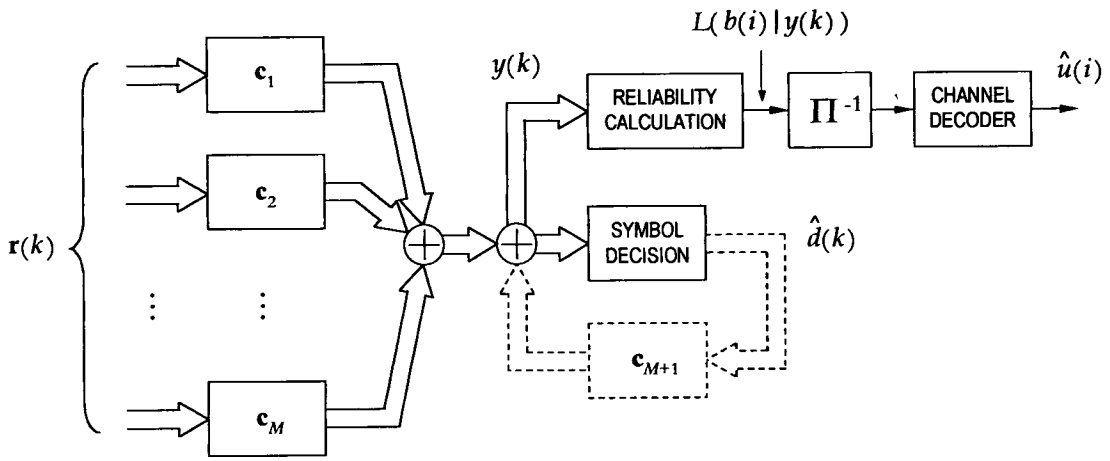

Figure 4.1: *Linear/decision-feedback space-time receiver.*

also has in vector notation

$$\mathbf{r}_m(k) = \mathbf{x}_m(k) + \mathbf{n}_m(k)\,, \tag{4.3}$$

where $\mathbf{r}_m(k) \triangleq [r_m(k) \cdots r_m(k-L+1)]^T$, $\mathbf{x}_m(k) \triangleq [x_m(k) \cdots x_m(k-L+1)]^T$, and $\mathbf{n}_m(k) \triangleq [n_m(k) \cdots n_m(k-L+1)]^T$.

4.1.2 Regularized linear LS equalization

The output of the space-time equalizer is written as

$$y(k) = \sum_{m=1}^{M} y_m(k) \tag{4.4}$$

with

$$y_m(k) = \sum_{\ell=0}^{L-1} c_m(\ell) r_m(k-\ell), \quad 1 \le m \le M \,. \tag{4.5}$$

Therefore, letting

$$\mathbf{r}(k) = \mathbf{x}(k) + \mathbf{n}(k) \triangleq [\ \mathbf{r}_1^T(k) \ \cdots \ \mathbf{r}_M^T(k)\ ]^T \tag{4.6}$$

and

$$\mathbf{c} \triangleq [\ \mathbf{c}_1^T \ \cdots \ \mathbf{c}_M^T\ ]^T \,, \tag{4.7}$$

where $\mathbf{c}_m \triangleq [c_m(0) \cdots c_m(L-1)]^T$, one gets

$$y(k) = \sum_{m=1}^{M} \mathbf{r}_m^T(k)\, \mathbf{c}_m = \mathbf{r}^T(k)\, \mathbf{c} \,. \tag{4.8}$$

Consider the residual disturbance at the equalizer output, given by the error signal

$$e(k) \triangleq y(k) - d(k) \,. \tag{4.9}$$

Introducing the vectors $\mathbf{y} \triangleq [y(L) \cdots y(N)]^T$, $\mathbf{d} \triangleq [d(L) \cdots d(N)]^T$, and $\mathbf{e} \triangleq \mathbf{y} - \mathbf{d} = [e(L) \cdots e(N)]^T$, we can express in compact form

$$\mathbf{y} = \mathbf{A}\mathbf{c} \tag{4.10}$$

with the $(N-L+1) \times ML$ matrix $\mathbf{A} \triangleq [\mathbf{r}(L) \cdots \mathbf{r}(N)]^T$. The training-based linear LS algorithm minimizes the cost

$$J(\mathbf{c}) \triangleq ||\mathbf{e}||^2 = \sum_{i=L}^{N} |e(i)|^2 \tag{4.11}$$

over the training interval. Given the training data sequence $\{\tilde{d}(k)\}_{k=1}^{N}$, we have $J(\tilde{\mathbf{c}}) = (\tilde{\mathbf{A}}\tilde{\mathbf{c}} - \tilde{\mathbf{d}})^H (\tilde{\mathbf{A}}\tilde{\mathbf{c}} - \tilde{\mathbf{d}}) = \tilde{\mathbf{c}}^H \tilde{\mathbf{A}}^H \tilde{\mathbf{A}} \tilde{\mathbf{c}} - 2\,\mathrm{Re}\{\tilde{\mathbf{c}}^H \tilde{\mathbf{A}}^H \tilde{\mathbf{d}}\} + \tilde{\mathbf{d}}^H \tilde{\mathbf{d}}$, where tilde indicates quantities relative to the burst midamble. Hence, provided that $\mathbf{A}$ is full rank[3], one obtains the

[3]The full column rank requirement implies that the matrix $\mathbf{A}$ has at least as many rows as column, that is $N - L + 1 \ge ML$.

least-squares solution for the space-time filter coefficients [119]

$$\tilde{\mathbf{c}}_o \triangleq \arg\min_{\tilde{\mathbf{c}}} J(\tilde{\mathbf{c}}) = \tilde{\mathbf{R}}^{-1}\tilde{\mathbf{p}} \tag{4.12}$$

where

$$\mathbf{R} \triangleq \frac{1}{N-L+1}\,\mathbf{A}^H\mathbf{A} = \frac{1}{N-L+1}\sum_{i=L}^{N}\mathbf{r}^*(i)\,\mathbf{r}^T(i) \tag{4.13}$$

is the *time-averaged correlation matrix* of the equalizer input, and

$$\mathbf{p} \triangleq \frac{1}{N-L+1}\,\mathbf{A}^H\mathbf{d} = \frac{1}{N-L+1}\sum_{i=L}^{N}\mathbf{r}^*(i)\,d(i) \tag{4.14}$$

is the *time-averaged cross-correlation vector* between the equalizer input and the transmitted data. Correspondingly, the equalizer output results $y(k) = \mathbf{r}^T(k)\,\tilde{\mathbf{R}}^{-1}\tilde{\mathbf{p}}$. It is worth emphasizing that during the equalizer operation on the payload, the received vector $\mathbf{r}(k)$ and the corresponding matrix $\mathbf{A}$ will be generally different from those used to estimate the coefficient vector $\tilde{\mathbf{c}}$. Denoting by $\mathbf{A}^+ \triangleq (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H$ the *Moore-Penrose pseudoinverse* of the matrix $\mathbf{A}$ [119], [130], in (4.12) it is $\mathbf{R}^{-1}\mathbf{p} = \mathbf{A}^+\mathbf{d}$, and the equalizer output can also be rewritten as $y(k) = \mathbf{r}^T(k)\,\mathbf{A}^+\mathbf{d} \Rightarrow \mathbf{y} = \mathbf{A}\mathbf{A}^+\mathbf{d}$. At the burst midamble one has $\tilde{\mathbf{y}} = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^+\tilde{\mathbf{d}}$, which corresponds to the orthogonal projection of $\tilde{\mathbf{d}}$ onto the range space of $\tilde{\mathbf{A}}$ [119]. Hence, $\tilde{\mathbf{e}} = \tilde{\mathbf{y}} - \tilde{\mathbf{d}} = (\tilde{\mathbf{A}}\tilde{\mathbf{A}}^+ - \mathbf{I})\tilde{\mathbf{d}}$, where $\mathbf{I}$ denotes the $(N-L+1)\times(N-L+1)$ identity matrix. In this case we say that the equalizer is in its *least-squares conditions*, and $\tilde{\mathbf{A}}^H\tilde{\mathbf{e}} = (\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}\tilde{\mathbf{A}}^+ - \tilde{\mathbf{A}}^H)\tilde{\mathbf{d}} = \mathbf{0}$, i.e., the error vector is orthogonal to the range of the matrix $\tilde{\mathbf{A}}$ (*principle of orthogonality*) [119].

A means of mitigating a possible ill-conditioning of the least-square problem (4.11) (i.e., a nearly rank deficient matrix $\mathbf{A}$) is to include in the cost a *stabilizing functional* or *regularizing operator* $\mathbf{c}^H\Lambda\,\mathbf{c}$ where Λ is a positive-definite matrix. This gives the *Tichonov functional* [122]

$$J'(\mathbf{c}) \triangleq ||\mathbf{e}||^2 + \mathbf{c}^H\Lambda\,\mathbf{c}\,. \tag{4.15}$$

Choosing $\Lambda = \lambda\mathbf{I}$, with $\mathbf{I}$ denoting the $ML \times ML$ identity matrix, the cost function (4.15) becomes

$$J'(\mathbf{c}) \triangleq ||\mathbf{e}||^2 + \lambda\,||\mathbf{c}||^2 \tag{4.16}$$

where $\lambda \geq 0$ is a *regularization coefficient*. For small λ, (4.16) corresponds to a minimization subject to the constraint that the solution c_o has minimum norm. The regularization coefficient relaxes the full-rank condition for the solution of the LS problem (see [122] and references in [48] and [122]). In this case, the solution (4.12) becomes

$$\tilde{\mathbf{c}}_o \triangleq \arg\min_{\tilde{\mathbf{c}}} J'(\tilde{\mathbf{c}}) = (\tilde{\mathbf{R}} + \lambda \mathbf{I})^{-1} \tilde{\mathbf{p}} \,, \tag{4.17}$$

and

$$y(k) = \mathbf{r}^T(k) \, (\tilde{\mathbf{R}} + \lambda \mathbf{I})^{-1} \tilde{\mathbf{p}} \,. \tag{4.18}$$

4.1.3 Regularized LS decision-feedback equalization

With a decision-feedback space-time equalizer, the output signal can be expressed as

$$y(k) = \sum_{m=1}^{M+1} y_m(k) = \sum_{m=1}^{M+1} \mathbf{r}_m^T(k) \, \mathbf{c}_m \,. \tag{4.19}$$

In (4.19), the vectors $\mathbf{r}_m(k)$ and $\mathbf{c}_m$, $1 \leq m \leq M$, are defined as in the previous section, and

$$y_{M+1}(k) = \mathbf{r}_{M+1}^T(k) \, \mathbf{c}_{M+1} \tag{4.20}$$

where, denoting by $L_f = L$ and L_b the number of forward and backward filter coefficients,

$$\mathbf{c}_{M+1} \triangleq [\; c_{M+1}(0) \;\; \cdots \;\; c_{M+1}(L_b - 1) \;]^T \tag{4.21}$$

$$\mathbf{r}_{M+1}(k) \triangleq [\; \hat{d}(k - D) \;\; \cdots \;\; \hat{d}(k - D - L_b + 1) \;]^T \,, \tag{4.22}$$

with D delay parameter. As shown in Figure 1, the symbols $\hat{d}(k)$ in (4.22) represent the hard or soft symbol decisions[4] at the output of the feedforward space-time filter [74], [126]. Therefore, letting

$$\mathbf{c} \triangleq [\; \mathbf{c}_1^T \;\; \mathbf{c}_2^T \;\; \cdots \;\; \mathbf{c}_{M+1}^T \;]^T \tag{4.23}$$

[4] $\hat{d}(k) = \tilde{d}(k)$ in correspondence to the training sequence midamble.

$$\mathbf{r}(k) \triangleq [\ \mathbf{r}_1^T(k) \quad \mathbf{r}_2^T(k) \quad \cdots \quad \mathbf{r}_{M+1}^T(k)\ ]^T , \tag{4.24}$$

one has

$$y(k) = \mathbf{r}^T(k)\,\mathbf{c} . \tag{4.25}$$

Hence, one still writes (4.10), where the matrix $\mathbf{A}$ is now $(N-L_f-D-L_b+1)\times(ML_f+L_b)$. With the above notation, the decision-feedback LS and regularized LS cost functions can still be formalized according to (4.11) and (4.16), with the respective solutions given by (4.12) and (4.17).

4.2 Calculation of the soft-output values

According to the notation introduced in Section 4.1, the equalizer output is expressed as

$$y(k) = d(k) + e(k) , \tag{4.26}$$

where $d(k) \in \{d^{(q)} = e^{j2\pi q/Q}, 1 \leq q \leq Q = 8\}$. One also writes $d(k) = \mathcal{M}(b_1(k)\, b_2(k) \cdots b_{n_Q}(k))$, where $\mathcal{M}$ denotes the mapping rule of the coded bits onto the modulated PSK symbols, and $n_Q = \log_2 Q$. From the *central limit theorem*, the residual disturbance $e(k)$ can be modelled as a Gaussian noise process. We further assume the disturbance $e(k)$ to be white. This is justified by taking into account that the feedforward part of the optimum space-time decision-feedback equalizer whitens the sum of CCI and noise both spatially and temporally [46] (observe that this does not hold for the linear space-time equalizer [46], where this assumption is expected to penalize to a certain extent the decoder performance). The bit log-likelihoods for channel decoding will then be computed by treating the error $e(k)$ as a white Gaussian noise process, which is characterized by its mean and variance, $e(k) = \mathcal{N}(\eta, \sigma^2)$.

4.2.1 Bit log-likelihoods

In order to perform soft-decision decoding, the channel decoder must be provided with *reliability information* on the coded bits. This reliability measure can be given in terms of the bit

log-likelihoods

$$L(b_i(k)|y(k)) = \log \frac{\Pr(b_i(k) = +1|y(k))}{\Pr(b_i(k) = -1|y(k))} . \tag{4.27}$$

Define by $S(b_i = +1)$ and $S(b_i = -1)$ the subsets of symbol indexes corresponding to $b_i = +1$ and $b_i = -1$, respectively

$$S(b_i = +1) \triangleq \{q : d^{(q)} = \mathcal{M}(b_1^{(q)}\, b_2^{(q)} \cdots b_i^{(q)} = +1 \cdots b_{n_Q}^{(q)})\} , \tag{4.28}$$

$$S(b_i = -1) \triangleq \{q : d^{(q)} = \mathcal{M}(b_1^{(q)}\, b_2^{(q)} \cdots b_i^{(q)} = -1 \cdots b_{n_Q}^{(q)})\} . \tag{4.29}$$

Then

$$\Pr(b_i(k) = +1|y(k)) = \sum_{q \in S(b_i=+1)} \Pr(d(k) = d^{(q)}|y(k)) \tag{4.30}$$

$$\Pr(b_i(k) = -1|y(k)) = \sum_{q \in S(b_i=-1)} \Pr(d(k) = d^{(q)}|y(k)) \tag{4.31}$$

where

$$\Pr(d(k) = d^{(q)}|y(k)) = \frac{\Pr(d^{(q)}, y(k))}{p(y(k))} = \frac{p(y(k)|d^{(q)}) \Pr(d^{(q)})}{\sum_{q=1}^{Q} p(y(k)|d^{(q)}) \Pr(d^{(q)})} . \tag{4.32}$$

Hence, using (4.30)-(4.32) in (4.27) gives

$$L(b_i(k)|y(k)) = \log \frac{\sum_{q \in S(b_i=+1)} p(y(k)|d^{(q)}) \Pr(d^{(q)})}{\sum_{q \in S(b_i=-1)} p(y(k)|d^{(q)}) \Pr(d^{(q)})} \tag{4.33}$$

and for $\Pr(d^{(q)}) = 1/Q$, $1 \leq q \leq Q$

$$L(b_i(k)|y(k)) = \log \frac{\sum_{q \in S(b_i=+1)} p(y(k)|d^{(q)})}{\sum_{q \in S(b_i=-1)} p(y(k)|d^{(q)})} . \tag{4.34}$$

From the assumptions on $e(k) = y(k) - d(k)$, we have $p(y(k)|d(k)) = \mathcal{N}(d(k) + \eta, \sigma^2)$

$$p(y(k)|d(k)) = \frac{1}{\pi\sigma^2}\, e^{-|y(k)-d(k)-\eta|^2/\sigma^2} . \tag{4.35}$$

4.2.2 Implementation in the logarithmic domain

The derivation of the log-likelihoods $L(b_i(k)|y(k))$ can be carried out without the need of computing the logarithm of a sum of probabilities.

In fact, from (4.33) one can write [13]

$$L(b_i(k)|y(k)) = \log \frac{\sum_{q \in S(b_i=+1)} p(y(k)|d^{(q)}) \Pr(d^{(q)})}{\sum_{q \in S(b_i=-1)} p(y(k)|d^{(q)}) \Pr(d^{(q)})}$$

$$= \max'_{q \in S(b_i=+1)} \{ \log p(y(k)|d^{(q)}) + \log \Pr(d^{(q)}) \} - \max'_{q \in S(b_i=-1)} \{ \log p(y(k)|d^{(q)}) + \log \Pr(d^{(q)}) \} \tag{4.36}$$

where $\max'\{x,y\} \triangleq \max\{x,y\} + \log(1 + e^{-|x-y|})$, and

$$\log p(y(k)|d^{(q)}) = -\log(\pi\sigma^2) - \frac{1}{\sigma^2} |y(k) - d^{(q)} - \eta|^2 . \tag{4.37}$$

Further simplification is obtained by approximating $\max'\{x,y\} \approx \max\{x,y\}$ (*max-log* approximation) [13], [92]. For equiprobable symbols we finally have

$$L(b_i(k)|y(k)) = \frac{1}{\sigma^2} \{ \min_{q \in S(b_i=+1)} |y(k) - d^{(q)} - \eta|^2 - \min_{q \in S(b_i=-1)} |y(k) - d^{(q)} - \eta|^2 \} . \tag{4.38}$$

The above formulation also allows to take into account *a priori* information on the coded bits, which can be included in the terms $\log \Pr(d^{(q)})$ in (4.36). At this point, in order to compute the log-likelihood values for soft-decision decoding by means of (4.34) and (4.35) or (4.38), what is required is an estimate of the mean η and variance σ^2 of the output error $e(k)$. It is worth noting that operating in the logarithmic domain, besides reducing the implementation complexity of the APP calculator, has the additional advantage of avoiding possible numerical problems for low values of σ^2.

4.2.3 Variance estimators

This section considers some practical estimators for the variance of the disturbance $e(k) = \mathcal{N}(\eta, \sigma^2)$ at the output of the space-time equalizer. The task of estimating the output variance

is complicated by the small amount of available data. This restriction is necessary to ensure that the quantity to be estimated does not appreciably change over the duration of the observation interval.

We assume an ergodic process, so that the ensemble averages are approximated by time averages over the observation interval. Therefore, given the data values $X_1, \ldots, X_n$, we will compute the sample mean

$$\hat{\eta} = \frac{1}{n} \sum_{i=1}^{n} X_i \,, \tag{4.39}$$

and the sample variance

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} |X_i - \hat{\eta}|^2 \,. \tag{4.40}$$

Although mathematically equivalent, the *two-pass algorithm* (4.40) is numerically preferable with respect to the familiar *textbook algorithm* $\hat{\sigma}^2 = (1/(n-1))\,[\sum_{i=1}^{n} |X_i|^2 - n\,|\hat{\eta}|^2\,]$ [131], [122]. As is known, the estimates (4.39) and (4.40) are independent, $\hat{\eta} = \mathcal{N}(\eta, \sigma^2/n)$, and $2(n-1)\,\hat{\sigma}^2/\sigma^2 = \chi^2_{2(n-1)}$, where χ^2_n denotes the chi-squared distribution with n degrees of freedom [114]. Hence

$$\mathrm{Var}(\hat{\sigma}^2) = \sigma^4/(n-1) \,. \tag{4.41}$$

Moreover, the estimator $[\hat{\eta}\ \ \hat{\sigma}^2]$ given by (4.39), (4.40) is the *minimum-variance unbiased* (MVU) estimator of $[\eta\ \ \sigma^2]$ [114].

In the presence of rounding errors in the calculation of (4.39), (4.40) from quantized data, the accuracy of the above estimator can generally be improved by the *corrected two-pass algorithm* [131], [122]. In this case, after calculation of the sample mean (4.39), the sample variance is obtained as

$$\hat{\sigma}^2 = \frac{1}{n-1}\,\Big[\,\sum_{i=1}^{n} |X_i - \hat{\eta}|^2 - \frac{1}{n}\,\Big|\sum_{i=1}^{n} (X_i - \hat{\eta})\,\Big|^2\,\Big] \,, \tag{4.42}$$

where the second term in the RHS of (4.42) corrects the roundoff error of the first term [131].

4.2.3.1 Estimator based on the known training symbols (TS)

Under the assumption that the equalizer operates in its least-squares conditions, an estimate of the variance of the output disturbance for a given transmitted burst can be obtained by using the training sequence midamble. In this case, considering that $E\{\mathbf{r}^*(k)\mathbf{r}^T(k)\} \approx \mathbf{R}$ and $E\{\mathbf{r}^*(k)d(k)\} \approx \mathbf{p}$, one has the mean square error [119]

$$E\{|\tilde{e}(k)|^2\} = \mathcal{E}_{\min} \approx \sigma_d^2 + \tilde{\mathbf{c}}_o^H \tilde{\mathbf{R}}\, \tilde{\mathbf{c}}_o - 2\,\mathrm{Re}\{\tilde{\mathbf{c}}_o^H \mathbf{p}\}\,, \tag{4.43}$$

with $\tilde{\mathbf{c}}_o$ given by (4.17). Alternatively, the variance can be obtained by taking the true error $\tilde{e}(k)$ and computing

$$\hat{\sigma}^2 = E\{|\tilde{e}(k) - E\{\tilde{e}(k)\}|^2\}\,, \tag{4.44}$$

where the ensemble averages are then approximated by time averages over the burst midamble. Although considering the ensemble averages it is $E\{\tilde{e}(k)\} = E\{\tilde{y}(k)\} - E\{\tilde{d}(k)\} = E\{\tilde{\mathbf{r}}^T(k)\}$ $\tilde{\mathbf{c}}_o - E\{\tilde{d}(k)\} = 0$, when using local error statistics we will compute the actual mean value of the error over the observation interval. In this case, using (4.39) and (4.40), the variance of $\hat{\eta}$ is σ^2/n, while the variance of $\hat{\sigma}^2$ is $\sigma^4/(n-1)$.

4.2.3.2 Estimator based on the distance from the finite alphabet (FA)

One problem with estimating the variance from the training symbols at the burst midamble is that it is not guaranteed that the space-time equalizer operates under least-squares conditions over the entire burst. In fact, even when the propagation channels can be considered quasi-stationary, the statistics of the output error can abruptly change in the presence of asynchronous interference. Moreover, in the case of decision-feedback equalization, the error statistics depend on possible decision errors. One way to alleviate this problem is to compute the error statistics taking into account the data from the payload. In this respect, a possible approach is to compute

$$\hat{\sigma}^2 = E\{|\hat{e}(k) - E\{\hat{e}(k)\}|^2\}\,, \tag{4.45}$$

where $\hat{e}(k)$ is the estimated error at the output of the space-time equalizer, and with the ensemble averages approximated by time averages over a given time interval using (4.39) and (4.40).

As an estimate $\hat{e}(k)$ of the error for the payload symbols we consider the distance from the finite alphabet at the equalizer output [127]. Denoting by $\hat{d}(k)$ the estimated symbols obtained by hard-decision, we have

$$\hat{e}(k) = y(k) - \hat{d}(k) \,. \tag{4.46}$$

Of course, the accuracy of the above estimator is guaranteed only in the absence of decision errors.

4.2.3.3 Estimator based on the signal constant modulus property (CM)

This section introduces an alternative approach for estimating the error variance from the payload symbols. To the best of our knowledge, this technique has not been previously reported in the literature.

Taking into account the circularly symmetric distribution of $e(k)$, one has $E\{(\mathrm{Re}\{e(k)\} - E\{\mathrm{Re}\{e(k)\}\})^2\} = E\{(\mathrm{Im}\{e(k)\} - E\{\mathrm{Im}\{e(k)\}\})^2\} = \sigma^2/2$. Therefore, one can approximate the variance $\sigma^2/2$ by a measure of the fluctuation of the envelope at the equalizer output around the unit circle. To prove this, we compute the variance of the quantity $\delta(k) \triangleq |y(k)| - |d(k)|$, with $y(k) = d(k) + e(k)$. We start by observing that

$$E\{(\delta(k) - E\{\delta(k)\})^2\} = \sum_{q=1}^{Q} E\{(\delta(k) - E\{\delta(k)\})^2 | d(k) = d^{(q)}\}$$

$$\cdot \Pr(d(k) = d^{(q)})$$

$$= E\{(\delta(k) - E\{\delta(k)\})^2 | d(k) = d^{(q)}\} \,. \tag{4.47}$$

Similarly

$$E\{\delta(k)\} = \sum_{q=1}^{Q} E\{\delta(k) | d(k) = d^{(q)}\} \Pr(d(k) = d^{(q)})$$

$$= E\{\delta(k)|d(k) = d^{(q)}\} . \tag{4.48}$$

Hence

$$E\{(\delta(k) - E\{\delta(k)\})^2\} = E\{(\delta(k) - E\{\delta(k)|d(k)\})^2|d(k)\} . \tag{4.49}$$

Letting $x \triangleq \mathrm{Re}\{e(k)\}$ and $y \triangleq \mathrm{Im}\{e(k)\}$, and taking $d(k) = 1 + j0$ one writes

$$\delta(k)|d(k) = \sqrt{(1+x)^2 + y^2} - 1 . \tag{4.50}$$

The function $g(x,y) = \sqrt{(1+x^2)+y^2} - 1$ can be approximated by its Taylor series about the point $(x = x_0, y = y_0)$

$$g(x,y) \approx g(x_0,y_0) + \left.\frac{\partial g}{\partial x}\right|_{x_0,y_0} x + \left.\frac{\partial g}{\partial y}\right|_{x_0,y_0} y + \frac{1}{2}\Big(\left.\frac{\partial^2 g}{\partial x^2}\right|_{x_0,y_0} x^2$$

$$+ \left.\frac{\partial^2 g}{\partial x \partial y}\right|_{x_0,y_0} xy + \left.\frac{\partial^2 g}{\partial y^2}\right|_{x_0,y_0} y^2 \Big) . \tag{4.51}$$

Then, with x and y independent Gaussian random variables with zero mean and variance $\sigma_x^2 = \sigma_y^2 = \sigma^2/2$, using (4.51) for $x_0 = 0, y_0 = 0$ yields

$$E\{\delta(k)|d(k)\} = E\{g(x,y)\} \approx g(0,0) + \frac{1}{4}\left.\frac{\partial^2 g}{\partial x^2}\right|_{0,0} \sigma^2 + \frac{1}{4}\left.\frac{\partial^2 g}{\partial y^2}\right|_{0,0} \sigma^2 . \tag{4.52}$$

Therefore, from (4.51), (4.52), and neglecting moments of order higher than 2 [75]

$$E\{(\delta(k) - E\{\delta(k)|d(k)\})^2|d(k)\} = E\{(g(x,y) - E\{g(x,y)\})^2\}$$

$$\approx \frac{1}{2}\Big(\left.\frac{\partial g}{\partial x}\right|_{0,0}\Big)^2\sigma^2 + \frac{1}{2}\Big(\left.\frac{\partial g}{\partial y}\right|_{0,0}\Big)^2\sigma^2 . \tag{4.53}$$

Finally, taking into account that $\partial g/\partial x|_{0,0} = 1$ and $\partial g/\partial y|_{0,0} = 0$, and using (4.53) in (4.49), one obtains

$$E\{(\delta(k) - E\{\delta(k)\})^2\} \approx \sigma^2/2 . \tag{4.54}$$

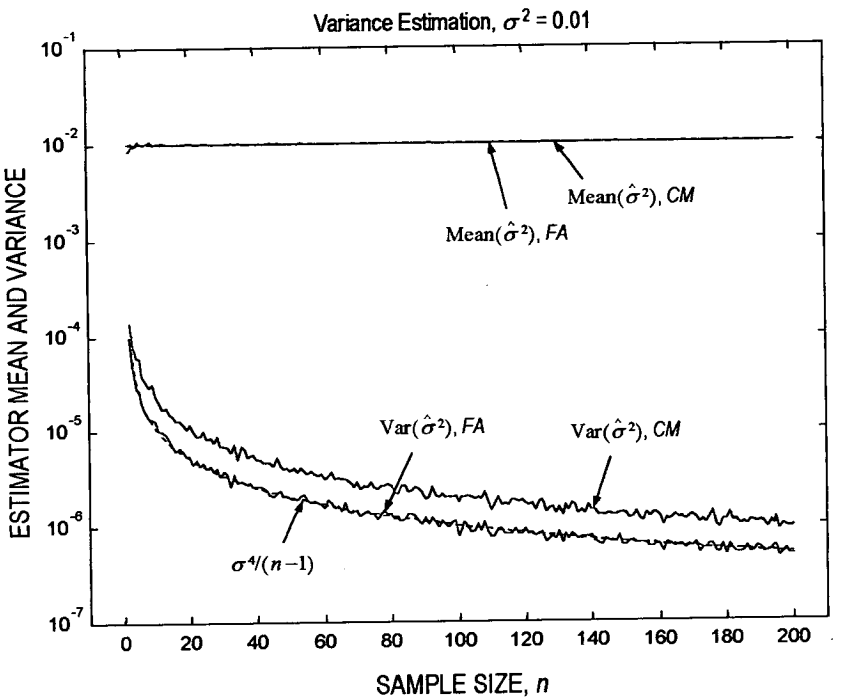

Figure 4.2: *Bias and variance of different variance estimators as a function of the sample size, for the zero-mean Gaussian disturbance model ($\sigma^2 = 0.01$). Averages computed over 400 independent trials.*

Therefore, we take the variance estimate

$$\hat{\sigma}^2 = 2E\{(\delta(k) - E\{\delta(k)\})^2\}\,, \tag{4.55}$$

where again the ensemble averages are approximated by time averages over a given interval. The performance of the estimator using (4.39) and (4.40) has been estimated numerically by generating directly the signal (4.26) with $e(k) = \mathcal{N}(0, \sigma^2)$. The results reported in Figures 4.2 to 4.6 are obtained by averaging over 400 independent simulation runs. The estimator (4.55) based on the signal constant modulus property (CM) is compared with the estimator of (4.45), (4.46), based on the distance from the finite alphabet (FA). Figures 4.2 to 4.4 show the mean and variance of the estimators as a function of the sample size n, for values of the true variance $\sigma^2 = 0.01$, $\sigma^2 = 0.1$, and $\sigma^2 = 1$. Figures 4.5 and 4.6 give the behaviour of the mean and variance of the estimators as a function of the true variance, for sample size $n = 15$

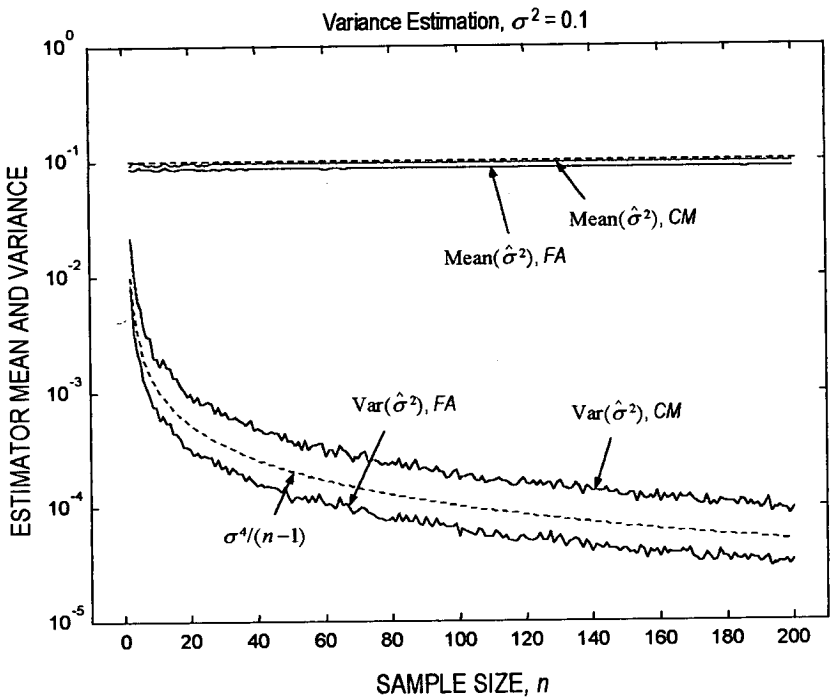

Figure 4.3: *Bias and variance of different variance estimators as a function of the sample size, for the zero-mean Gaussian disturbance model ($\sigma^2 = 0.1$). Averages computed over 400 independent trials.*

and $n = 100$, respectively. As shown in the figures, the estimator based on the distance from the finite alphabet has a smaller variance than the estimator (4.55). The variance of the FA estimator results equal to $\sigma^4/(n-1)$ for low values of σ^2, i.e., in the absence of decision errors. However, the estimator (4.55) based on the constant modulus property provides a lower bias for values of the true variance $\sigma^2 > 0.1$.

4.2.3.4 Burst statistics versus short-term statistics

An important design parameter for the variance estimation is the length of the observation interval. For the estimator based on the training symbols, the observation interval is given by the training sequence midamble. However, as already pointed out, computing the variance using the data from the payload, one has to consider the effect of abrupt changes of the error statistics due to asynchronous interference and/or feedback of possible decision errors. Therefore, the

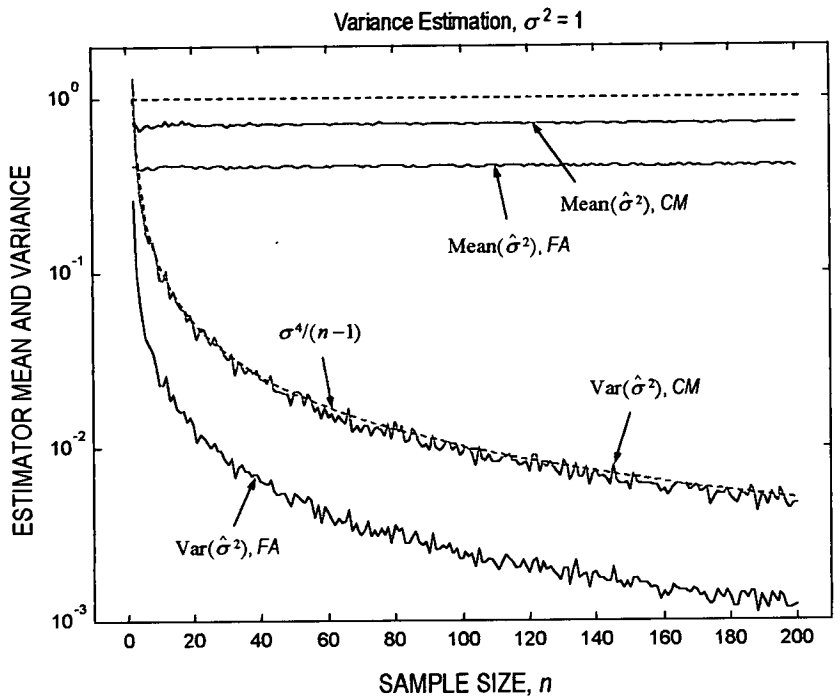

Figure 4.4: *Bias and variance of different variance estimators as a function of the sample size, for the zero-mean Gaussian disturbance model ($\sigma^2 = 1$). Averages computed over 400 independent trials.*

simulation study will compare for the different scenarios the performance of variance estimators that compute *burst statistics* – i.e. perform time averages using the data of the entire burst – with that of estimators based on *short-term statistics*, obtained on observation intervals given by a sliding window of a predetermined length [127].

Note that, since variance estimation is performed after equalization, computing short-term or burst statistics does not affect the operation of the space-time equalizer. Clearly, the variance estimators that make use of the payload symbols imply a (moderate) increase of the receiver complexity with respect to the case of training-based estimator.

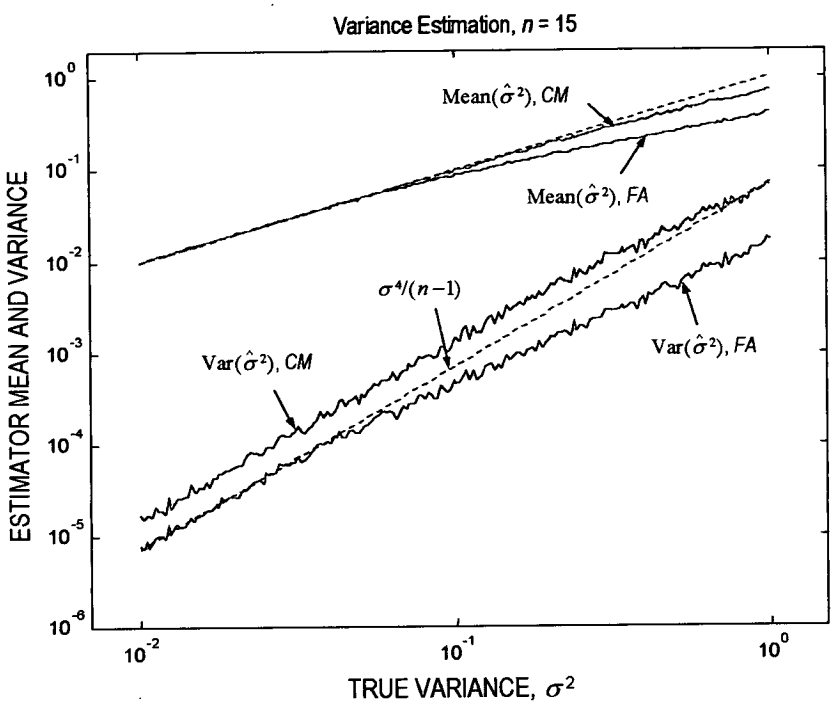

Figure 4.5: *Bias and variance of different variance estimators with sample size $n = 15$, for the zero-mean Gaussian disturbance model. Averages computed over 400 independent trials.*

4.3 Soft-decision feedback by nonlinear mean-square estimation

The robustness of the space-time receiver studied in the previous sections can be further improved by the additional use of soft values for the local feedback of tentative decisions in the decision-feedback equalizer.

The technique of *soft-decision feedback* has recently found wide application, due to its inherent capability to reduce error propagation in decision-feedback equalization [128], [129], reduced-state trellis equalization [133], [74], and iterative channel estimation [94], [59]. Soft decision of binary data has been studied, e.g., in [74], [128], [129] and [94], while the case of QPSK symbols has been addressed in [126], [134]. This section derives the nonlinearity for optimal (mean-square) soft decision of 8-PSK symbols, under the assumption that the residual disturbance is modelled as additive, white, Gaussian noise (AWGN).

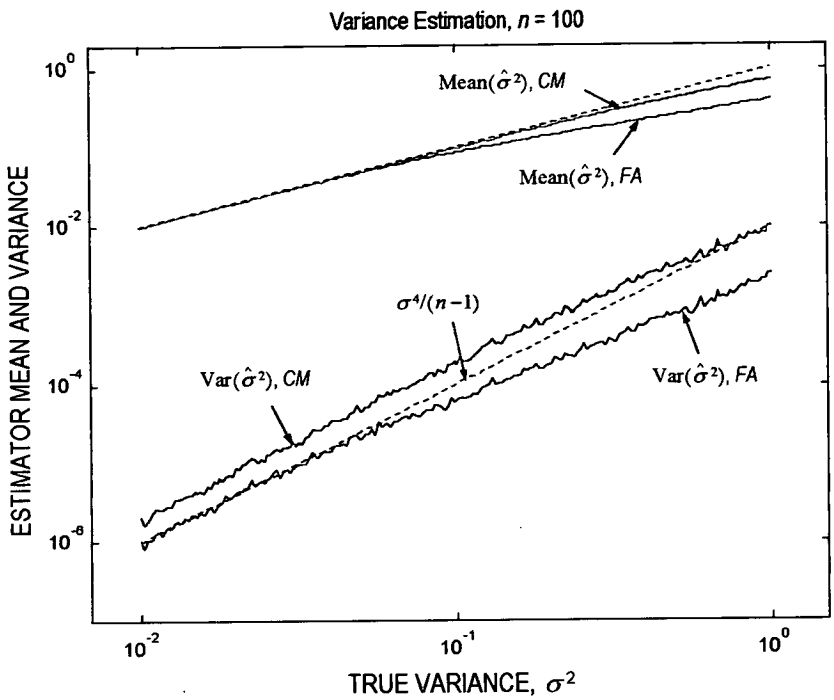

Figure 4.6: *Bias and variance of different variance estimators with sample size $n = 100$, for the zero-mean Gaussian disturbance model. Averages computed over 400 independent trials.*

We assume the equalizer output signal model (4.26) with $e(k) = \mathcal{N}(0, \sigma^2)$. Optimum soft decision on the symbols $d(k)$ can be obtained by nonlinear mean-square estimation according to (see Appendix B)

$$\hat{d}(k) = \arg \min_{f(y(k))} E\{|\, d(k) - f(y(k))\,|^2\} = E\{d(k)|y(k)\}$$

$$= \sum_{q=1}^{Q} d^{(q)} \Pr(d(k) = d^{(q)}|y(k)) \,. \tag{4.56}$$

Observe that the term $\Pr(d(k)|y(k))$ in (4.56) can be expressed as

$$\Pr(d(k)|y(k)) = \frac{\Pr(d(k), y(k))}{\Pr(y(k))}$$

$$= \frac{\Pr(y(k)|d(k))\Pr(d(k))}{\sum_{q=1}^{Q}\Pr(y(k)|d(k)=d^{(q)})\Pr(d(k)=d^{(q)})} . \tag{4.57}$$

Then, since for a fixed time k, $e(k)$ is a Gaussian random variable

$$\Pr(y(k)|d(k)) = p(y(k)-d(k)) = \frac{1}{\pi\sigma^2}e^{-|y(k)-d(k)|^2/\sigma^2} , \tag{4.58}$$

assuming equiprobable symbols (4.57) gives

$$\Pr(d(k)|y(k)) = \frac{e^{2\,\mathrm{Re}\{y^*(k)d(k)\}/\sigma^2}}{\sum_{q=1}^{Q} e^{2\,\mathrm{Re}\{y^*(k)d^{(q)}\}/\sigma^2}} . \tag{4.59}$$

It is easy to see that in the case of binary symbols substituting (4.59) in (4.56) gives a nonlinear characteristic expressed by the familiar sigmoid function [74], [128], [129], [94]

$$\hat{d}(k) = \arg\min_{f(y(k))} E\{|\, d(k) - f(y(k))\,|^2\}$$

$$= \tanh(\frac{2\,\mathrm{Re}\{y(k)\}}{\sigma^2}) . \tag{4.60}$$

Analogously, for a 4-PSK constellation from (4.56), (4.59) one obtains [126]

$$\hat{d}(k) = \arg\min_{f(y(k))} E\{|\, d(k) - f(y(k))\,|^2\}$$

$$= \frac{1}{\sqrt{2}}[\tanh(\frac{\sqrt{2}\,\mathrm{Re}\{y(k)\}}{\sigma^2}) + j\tanh(\frac{\sqrt{2}\,\mathrm{Im}\{y(k)\}}{\sigma^2})] . \tag{4.61}$$

In the case where the transmitted data $d(k)$ are taken from an 8-PSK symbol alphabet, $d(k) \in \{e^{j2\pi q/Q}, 1 \le q \le Q = 8\}$, and letting $y_1 \triangleq \mathrm{Re}\{y(k)\}$, $y_2 \triangleq \mathrm{Im}\{y(k)\}$, one has

$$\sum_{q=1}^{Q} e^{2\,\mathrm{Re}\{y^*(k)d^{(q)}\}/\sigma^2} = e^{2y_1/\sigma^2} + e^{2y_2/\sigma^2} + e^{-2y_1/\sigma^2} + e^{-2y_2/\sigma^2}$$

$$+ e^{\sqrt{2}(y_1+y_2)/\sigma^2} + e^{\sqrt{2}(-y_1+y_2)/\sigma^2} + e^{\sqrt{2}(-y_1-y_2)/\sigma^2} + e^{\sqrt{2}(y_1-y_2)/\sigma^2}$$

$$= 2\cosh(\frac{2y_1}{\sigma^2}) + 2\cosh(\frac{2y_2}{\sigma^2}) + 4\cosh(\frac{\sqrt{2}y_1}{\sigma^2})\cosh(\frac{\sqrt{2}y_2}{\sigma^2}) , \tag{4.62}$$

and

$$\sum_{q=1}^{Q} d^{(q)} e^{2\,\mathrm{Re}\{y^*(k)d^{(q)}\}/\sigma^2} = e^{2y_1/\sigma^2} + je^{2y_2/\sigma^2} - e^{-2y_1/\sigma^2} - je^{-2y_2/\sigma^2}$$

$$+ \frac{1}{\sqrt{2}}(1+j)e^{\sqrt{2}(y_1+y_2)/\sigma^2} + \frac{1}{\sqrt{2}}(-1+j)e^{\sqrt{2}(-y_1+y_2)/\sigma^2}$$

$$+ \frac{1}{\sqrt{2}}(-1-j)e^{\sqrt{2}(-y_1-y_2)/\sigma^2} + \frac{1}{\sqrt{2}}(1-j)e^{\sqrt{2}(y_1-y_2)/\sigma^2}$$

$$= 2\sinh(\frac{2y_1}{\sigma^2}) + j2\sinh(\frac{2y_2}{\sigma^2}) + 2\sqrt{2}\sinh(\frac{\sqrt{2}y_1}{\sigma^2})\cosh(\frac{\sqrt{2}y_2}{\sigma^2})$$

$$+ j2\sqrt{2}\cosh(\frac{\sqrt{2}y_1}{\sigma^2})\sinh(\frac{\sqrt{2}y_2}{\sigma^2}) . \tag{4.63}$$

Hence, substituting (4.59) in (4.56) and taking into account (4.62) and (4.63) yields

$$\hat{d}(k) = \arg \min_{f(y(k))} E\{|\, d(k) - f(y(k))\,|^2\} = \alpha(y(k))/\beta(y(k)) \tag{4.64}$$

where

$$\alpha(y) \triangleq \sinh(\frac{2\,\mathrm{Re}\{y\}}{\sigma^2}) + \sqrt{2}\sinh(\frac{\sqrt{2}\,\mathrm{Re}\{y\}}{\sigma^2})\cosh(\frac{\sqrt{2}\,\mathrm{Im}\{y\}}{\sigma^2})$$

$$+ j[\sinh(\frac{2\,\mathrm{Im}\{y\}}{\sigma^2}) + \sqrt{2}\cosh(\frac{\sqrt{2}\,\mathrm{Re}\{y\}}{\sigma^2})\sinh(\frac{\sqrt{2}\,\mathrm{Im}\{y\}}{\sigma^2})] , \tag{4.65}$$

and

$$\beta(y) \triangleq \cosh(\frac{2\,\mathrm{Re}\{y\}}{\sigma^2}) + \cosh(\frac{2\,\mathrm{Im}\{y\}}{\sigma^2})$$

$$+2\cosh(\frac{\sqrt{2}\,\mathrm{Re}\{y\}}{\sigma^2})\cosh(\frac{\sqrt{2}\,\mathrm{Im}\{y\}}{\sigma^2})\,. \tag{4.66}$$

The nonlinearity (4.64) can be easily implemented in a look-up table. In this case, the only computational cost derives from the calculation of the quantities $\mathrm{Re}\{y(k)\}/\sigma^2$ and $\mathrm{Im}\{y(k)\}/\sigma^2$, where the variance of the output error σ^2 can be estimated with the techniques of Section 4.2.3.

Note that when the samples $y(k)$ are correlated, (4.64) corresponds to the simplifying assumption that the *a posteriori* probability of $d(k)$ is a function only of the current observation $y(k)$. In this case the above approach, although suboptimum, has the advantage of simplicity and low implementation cost with respect to the use of a trellis *a posteriori* probability (APP) calculator (see, e.g., [74], [126]).

It is worth mentioning that the above approach with a variance estimate based on time average over a sliding window differs from some previously proposed soft-decision schemes (see e.g. [126]). The use of a short-term variance results in an adaptive nonlinear device, which can provide an advantage in the presence of nonstationary output statistics.

4.4 Numerical results

The performance of the EGPRS receiver with soft-output space-time equalization has been estimated through computer simulation of the MCS-5 coding scheme [125]. The results refer to a simulation length of 10^5 coded data blocks. According to the system model described in Section 4.1, the simulator includes an M-dimensional frequency selective multipath channel, CCI, and additive temporally and spatially white Gaussian noise. One dominant synchronous or asynchronous interfering cell is assumed, with the co-channel users transmitting independent 8-PSK signals. Ideal frequency hopping is implemented. The noise and interference powers are identified by the signal-to-noise ratio (SNR) and the signal-to-interference ratio (SIR), respectively. Unless stated otherwise, in the simulations the SNR is kept fixed to 35 dB. The channel is modelled according to the GSM typical urban (TU) profile with classical Doppler spectrum [76]. The receive antenna array is modelled as a linear array with M elements spaced by $3\lambda_c/M$, where λ_c denotes the carrier wavelength. The corresponding array response is obtained by assuming a uniformly distributed angle of arrival within a 90 deg sector.

The receiver includes a square-root raised cosine filter with roll-off 0.12. We simulate the regularized linear LS (L-LS) spatio-temporal equalizer with a time span of $L = L_f = 5$

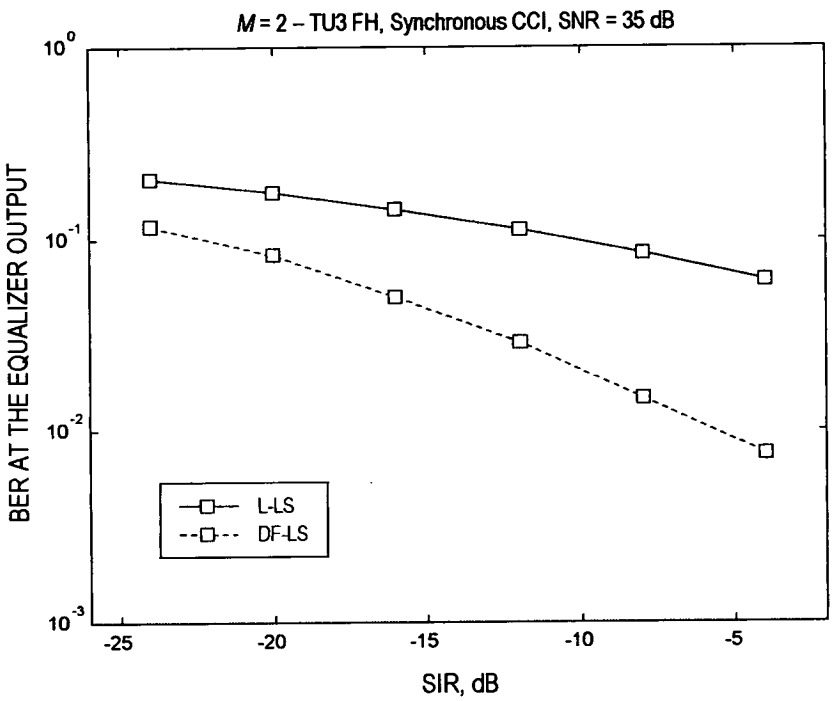

Figure 4.7: *Bit-error rate performance of the linear (L-LS) and decision-feedback (DF-LS) equalizers with $M = 2$ receive antennas and synchronous CCI.*

coefficients, and the regularized decision-feedback LS (DF-LS) spatio-temporal equalizer with $L_f = 5$ coefficients for the feedforward filters and $L_b = 3$ coefficients for the feedback filter. In all cases the regularization coefficient is chosen equal to $\lambda = 10^{-4}$. The calculation of the soft-output values at the equalizer output is implemented in the logarithmic domain using the max-log approximation (4.38).

4.4.1 Synchronous CCI

In the case of synchronous CCI, the simulation considers the use of $M = 2$ receive antennas. Figure 4.7 shows the bit-error rate (BER) performance of the linear LS and decision-feedback LS equalizers for the TU multipath profile with a mobile speed of 3 km/h (TU3). As exemplified in Figure 4.8, in the case of linear equalization the output error distribution remains almost stationary within the time slot. In this situation, the variance estimators based on the burst statistics have the advantage of a larger amount of observation data over the estimators based

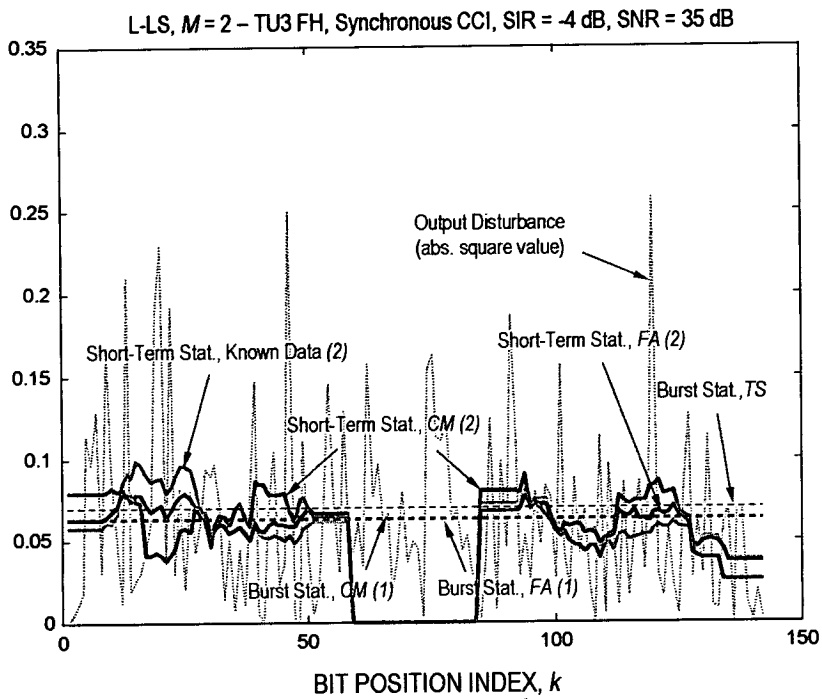

Figure 4.8: *Example of the variance estimate of one burst obtained (1) from burst statistics, and (2) from short-term statistics using a sliding window of length $W = 15$. Linear equalizer with $M = 2$ receive antennas and synchronous CCI.*

on short-term statistics. Reasonably accurate estimates are also obtained using the training sequence symbols. The total block-error rate (BLER) performance [76] reported in Figure 4.9 shows that for this scenario there is no substantial difference between computing the reliability information from the training sequence (TS estimator) and from the data of the entire burst (FA and CM estimators). Here, the schemes based on short-term statistics are penalized by the higher estimation variance. In fact, as can also be seen from Figures 4.2 to 4.6, the estimation variance increases monotonically by reducing the sample size n.

The results are significantly different for the decision-feedback equalizer. In this case, as illustrated in Figure 4.10, the error statistics is often highly nonuniform due to the presence of decision errors and error propagation. Correspondingly, one expects a performance advantage for the estimators based on short-term statistics. The behaviour of the different variance estimators is quantified in Figure 4.11 in terms of mean-square error (MSE) with respect to the sample

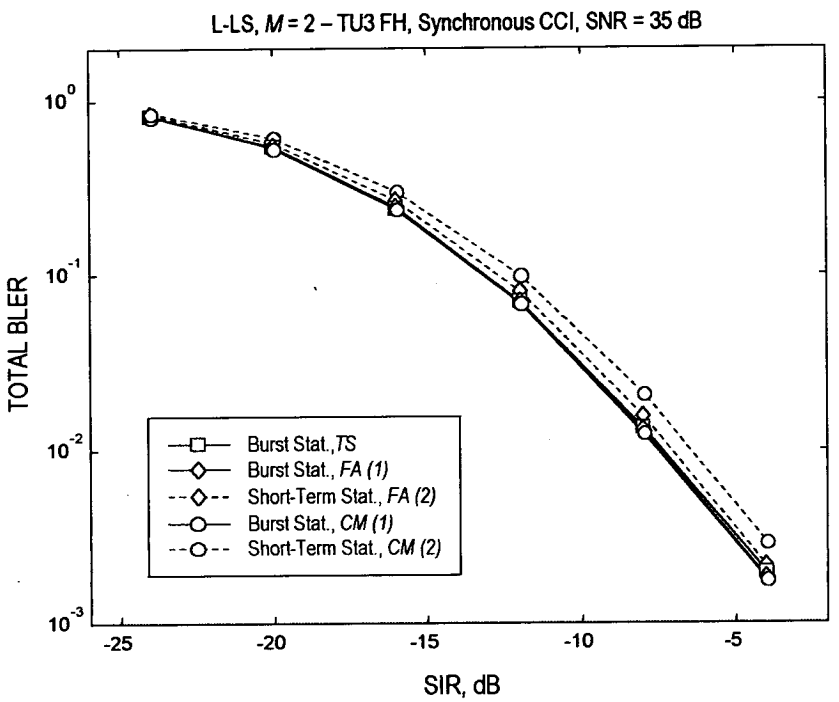

Figure 4.9: *Block-error rate performance of the soft-output linear equalizer with $M = 2$ receive antennas and synchronous CCI.*

variance based on the known symbols, computed over a sliding window of length $W = 15$. The resulting BLER performance is shown in Figure 4.12. The plot indicates that calculating the soft-output values based on a window of length $W = 15$ produces about one order of magnitude reduction in BLER with respect to using the statistics obtained from the training sequence midamble. One also observes that the estimators based on the distance from the finite alphabet (FA) and on the deviation of the signal modulus (CM) give substantially the same BLER performance.

4.4.2 Asynchronous CCI

In the presence of asynchronous CCI, we consider the use of $M = 4$ receive antennas. The BER performance of the L-LS and DF-LS equalizers for the TU50 multipath profile is shown in Figure 4.13. In this scenario, different parts of the single time slot of the signal of interest are affected by (a portion of) independent asynchronous co-channel bursts from an interfering

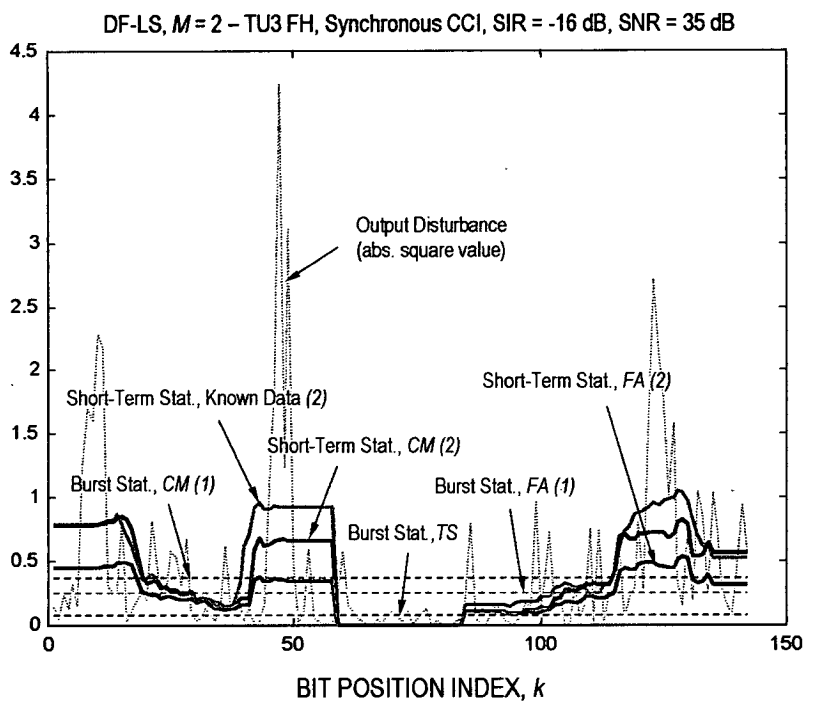

Figure 4.10: *Example of the variance estimate of one burst obtained (1) from burst statistics, and (2) from short-term statistics using a sliding window of length* $W = 15$. *Decision-feedback equalizer with* $M = 2$ *receive antennas and synchronous CCI.*

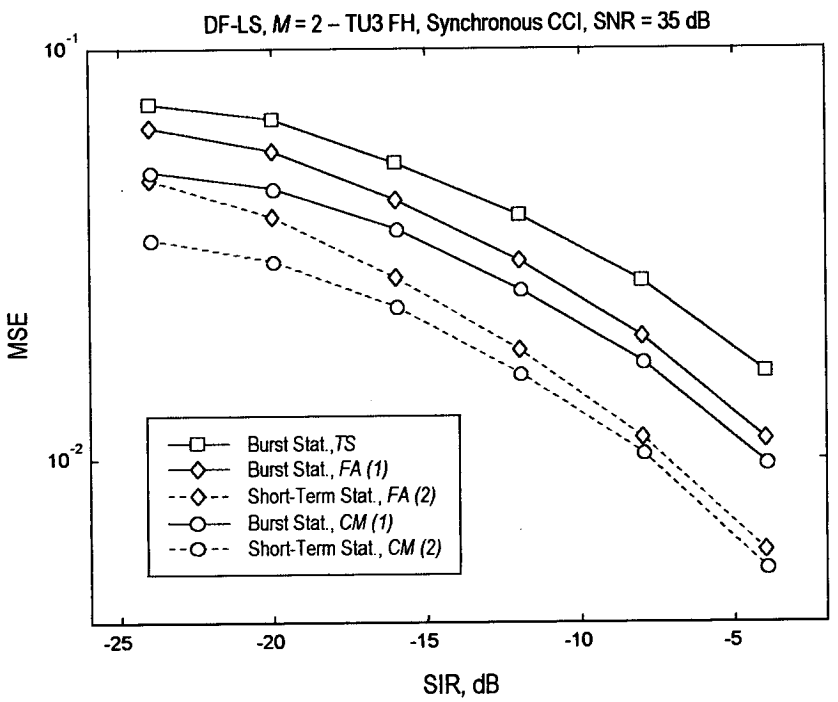

Figure 4.11: *Mean-square error of the variance estimates obtained (1) from burst statistics, and (2) from short-term statistics using a sliding window of length $W = 15$. Decision-feedback equalizer with $M = 2$ receive antennas and synchronous CCI.*

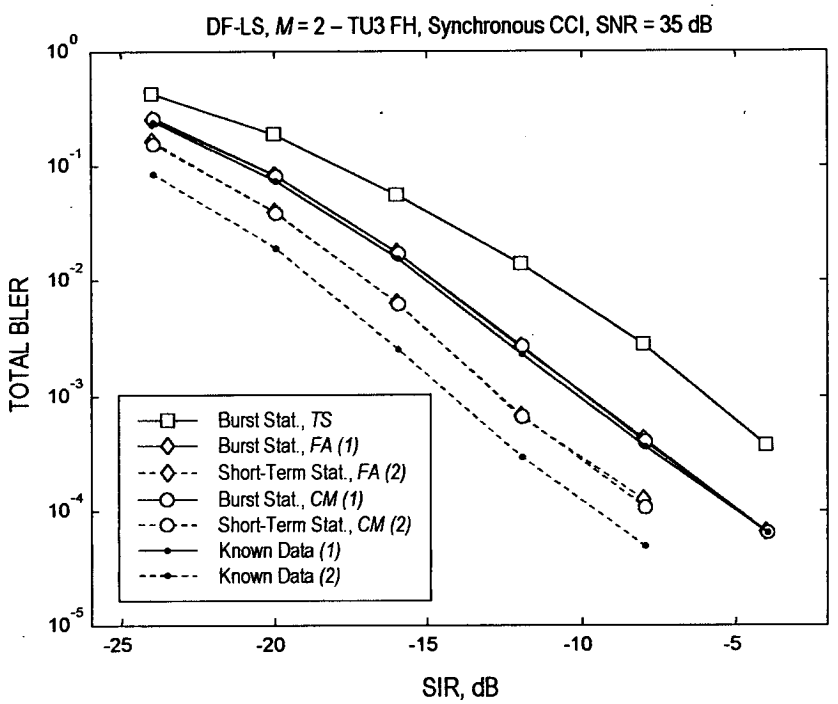

Figure 4.12: *Block-error rate performance of the soft-output decision-feedback equalizer with $M = 2$ receive antennas and synchronous CCI.*

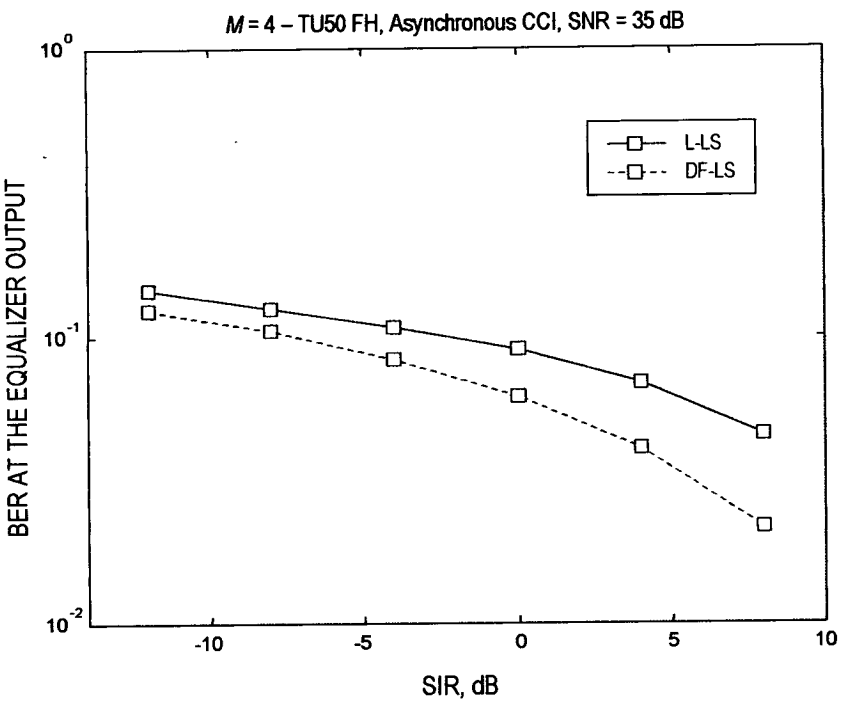

Figure 4.13: *Bit-error rate performance of the linear (L-LS) and decision-feedback (DF-LS) equalizers with $M = 4$ receive antennas and asynchronous CCI.*

cell. Since the space-time equalizer is designed on the basis of the CCI affecting the training sequence midamble, the output error statistics will present a step behaviour at the point where the CCI changes. Therefore, as shown in Figure 4.14, the BLER performance of the linear equalizer are now significantly improved by computing the reliability information based on the short-term error statistics. The results indicate a BLER reduction of two to four orders of magnitude with respect to the technique based on the training sequence midamble.

An illustration of the error distribution for the decision-feedback equalizer is given in Figure 4.15, which clearly shows the step behaviour of the error statistics. Under these conditions, a major impact on the decoder performance derives from the capability of discriminating the quality of bits pertaining to the different portions of the burst. The total BLER performance reported in Figure 4.16 shows for both FA and CM estimators a significant BLER reduction of two to four orders of magnitude (corresponding to a SIR gain of about 20 dB) with respect to the receiver employing a TS-based soft-output.

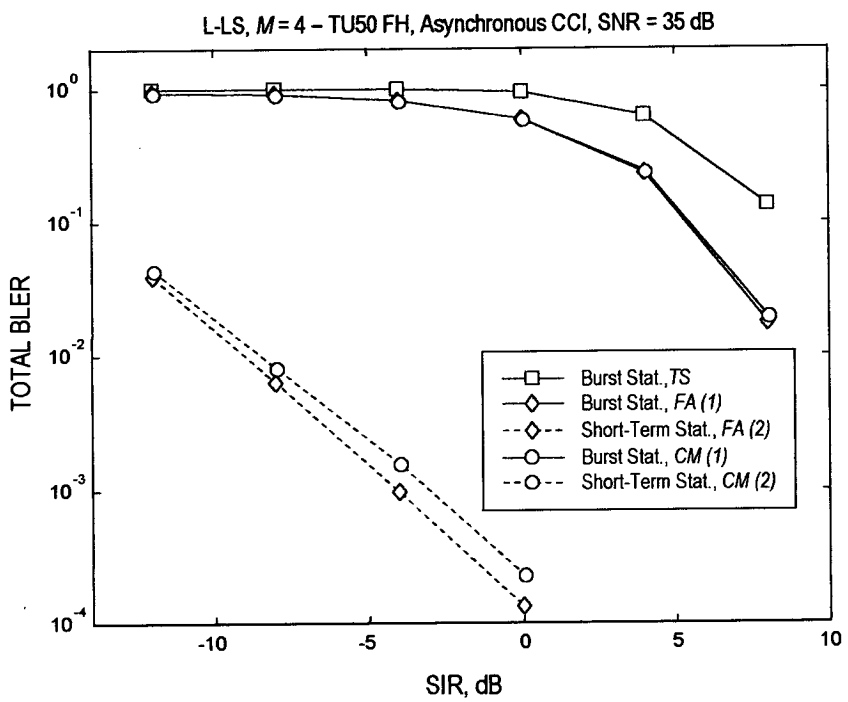

Figure 4.14: *Block-error rate performance of the soft-output linear equalizer with* $M = 4$ *receive antennas and asynchronous CCI.*

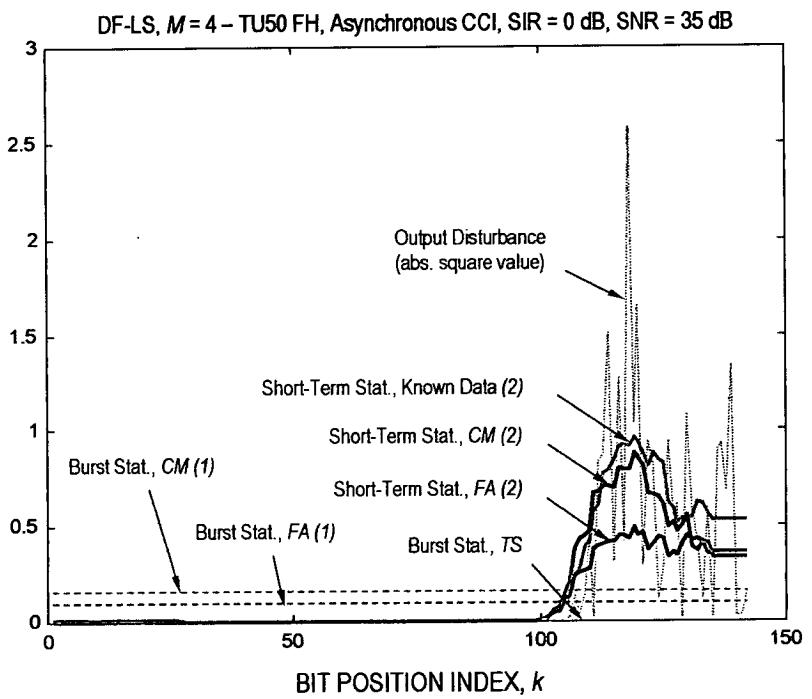

Figure 4.15: *Example of the variance estimate of one burst obtained (1) from burst statistics, and (2) from short-term statistics using a sliding window of length* $W = 15$. *Decision-feedback equalizer with* $M = 4$ *receive antennas and asynchronous CCI.*

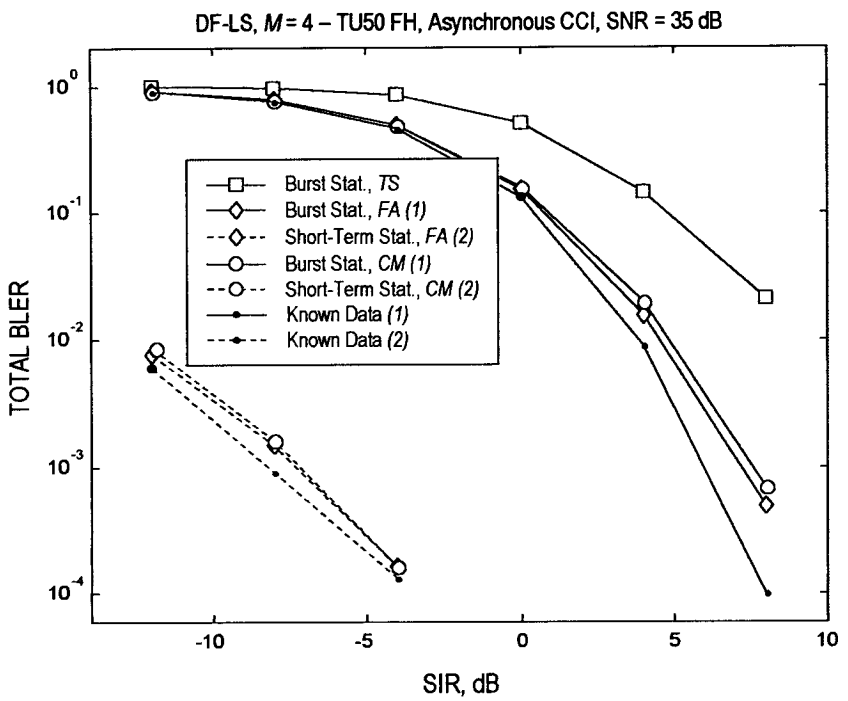

Figure 4.16: *Block-error rate performance of the soft-output decision-feedback equalizer with $M = 4$ receive antennas and asynchronous CCI.*

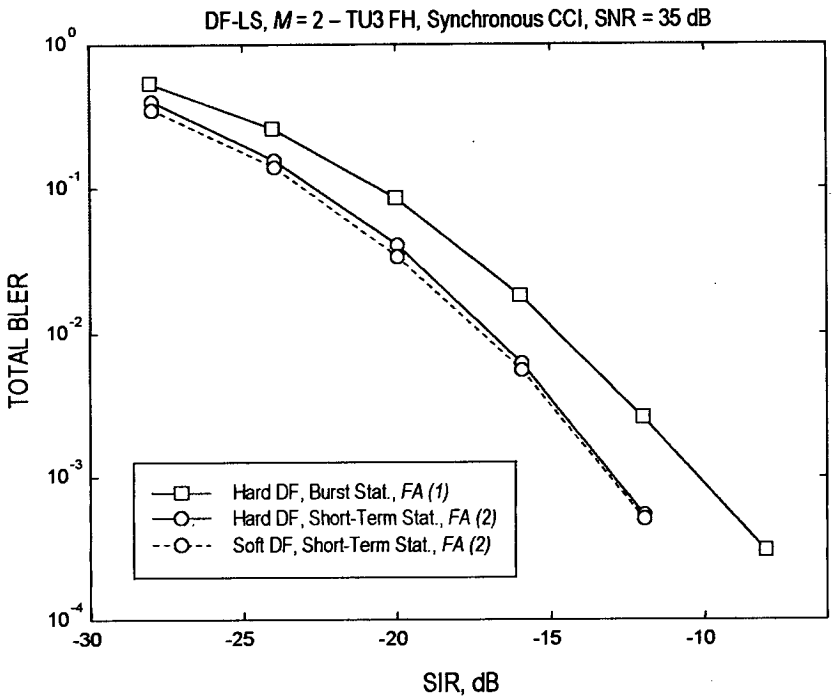

Figure 4.17: *Block-error rate performance of the soft-output space-time equalizer with hard and soft decision feedback.* $M = 2$ *receive antennas and synchronous CCI. SNR* $= 35$ *dB.*

4.4.3 Soft -decision feedback

The results of the previous sections refer to the use of probabilistic information on the quality of the output of the deterministic space-time equalizer. This section addresses the additional use of soft values for local feedback of the tentative decision in the DF-LS equalizer.

The performance of the nonlinearity (4.64) for soft decision-feedback soft-output space-time equalization has been analysed by computer simulations and compared to the case of hard-decision feedback. The results refer to the EGPRS coding schemes MCS-5 and the synchronous CCI scenario of Section 4.4.1 with $M = 2$ receive antennas and TU3 propagation. In the simulations, hard or soft feedback is used in conjunction with the soft-output calculation based on the distance from the finite alphabet (FA variance estimator). Depending on the signal-to-distortion ratio at the equalizer output, soft-decision feedback can provide an improved BER performance by reducing the effect of error propagation [128]. From our point of view, the

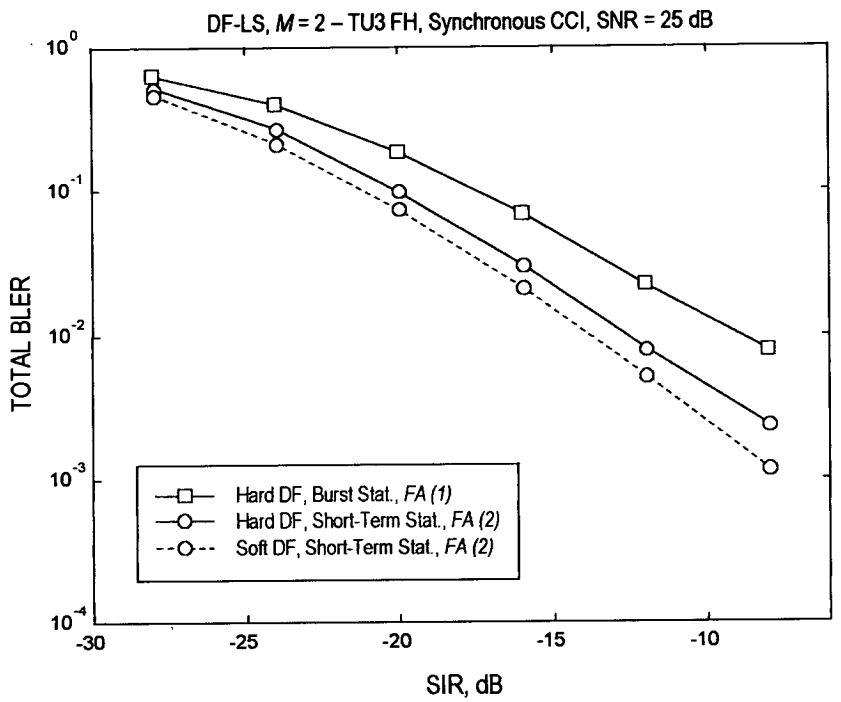

Figure 4.18: *Block-error rate performance of the soft-output space-time equalizer with hard and soft decision feedback. $M = 2$ receive antennas and synchronous CCI. SNR $= 25$ dB.*

substantial value is here the improved quality of the reliability information at the equalizer output, which can produce an additional gain at the output of the channel decoder. Figures 4.17 and 4.18 report the simulated BLER performance corresponding to an SNR of 35 dB and 25 dB, respectively. In Figure 4.17, the space-time equalizer is in the condition to effectively reject the interference and correspondingly the use of soft-decision produces only a marginal gain (see also [128]). However, comparison with Figure 4.18 shows that the soft-decision feedback scheme provides improved robustness in those cases where the equalizer operation is affected by additional impairments like a reduced input SNR.

4.5 Summary

This chapter has studied the performance of channel decoding for an antenna array receiver employing a deterministic spatio-temporal equalizer in an interference-limited scenario. A simple

method has been proposed to derive the reliability information on the coded bits at the equalizer output, which is based on the estimation of the output error variance over a given time window. In particular, the study has investigated different variance estimators based on burst statistics and short-term statistics of the output error. The simulation results obtained for EGPRS MCS-5 show that the receiver block-error rate can be significantly reduced by the selection of a proper rule to derive the *a posteriori* probability on the coded bits. In the presence of nonstationary disturbance, the performance is largely dominated by the capability of tracking the variations of the output error statistics by using an APP calculator based on a short-term variance estimator, with a gain of up to 20 dB in terms of SIR.

The chapter has also addressed the use of the above soft-output calculator in conjunction with soft-decision feedback based on nonlinear MS estimation. The optimum MS soft-decision device for 8-PSK symbols has been derived, and its effectiveness has been assessed by simulation of the EGPRS MCS-5 coding scheme. The collected data show that, although most of the gain often derives from the use of the appropriate output APP estimator, soft-decision feedback provides additional robustness to the soft-output LS space-time equalizer.

The above results show that probabilistic techniques can be effectively applied to a receiver implementing a deterministic space-time filtering algorithm. Furthermore, the design of probabilistic processing can be as critical to the overall receiver performance as the actual space-time algorithm, and results essential to the development of an effective and robust equalizer for interference-limited scenarios.

Chapter 5
Conclusions

This thesis has considered the application of probabilistic techniques to equalization of the mobile radio channel in the presence of ISI, noise and interference. In particular, dealing with receivers that implement the concatenation of an equalizer and a channel decoder, the work has emphasized the importance of delivering probabilistic information on the quality of the equalizer output in interference-limited scenarios, where it is often difficult to accurately model the disturbance.

In the case of MAP trellis equalization in the presence of non-Gaussian interference, improved output reliability information can be provided by a novel technique based on *non-parametric* estimation of the density function of the interference plus noise. In the proposed approach, the fundamental issue of density estimation with a limited volume of training data has been addressed by applying kernel smoothing techniques. In this respect, symmetry conditions have been identified, which effectively double the size of the training set. The thesis has considered the design of the optimum smoothing parameter for complex densities, and provided an analysis of the MISE performance of the density estimator. The study includes a discussion on the problem of deriving sufficient statistics for data detection in the presence of non-Gaussian correlated noise. Major advantages of the proposed strategy are its intrinsic robustness and general applicability. Simulation results are provided for the GSM system, showing significantly improved error performance with respect to the conventional trellis equalizer based on the Gaussian assumption, and improved quality of the output reliability information for soft-decision decoding. The collected data refer to a synchronous interference scenario. However, the approach may be extended to the case of asynchronous interference by introducing an adaptation of the estimated density. Further work should also investigate the proposed equalizer in conjunction with improved channel estimation techniques for non-Gaussian noise.

For multiple-antenna receivers, interference cancellation by means of deterministic space-time filtering techniques can often provide an effective alternative to trellis equalization, with advantages in terms of robustness and implementation cost. Linear or decision-feedback space-time equalizers based on the deterministic LS criterion have the further advantage of not relying on

an estimate of the channel response, or on particular assumptions on the statistical model of the interference. To enable the use of deterministic space-time equalizers in receivers based on soft-value processing, a simple method has been proposed for deriving the reliability information on the coded bits at the equalizer output. The technique is based on the assumption that the feedforward space-time filters produce an (approximately) uncorrelated output error. The approach relies on the estimation of the output error variance, which can be periodically updated within each burst. Simulation results for the EDGE/EGPRS system show that the receiver performance can be significantly improved by tracking the variations of the output error statistics by a soft-output calculator based on a short-term variance estimator, with performance gains of up to 20 dB. The thesis has also addressed the use of soft-decision feedback, where the analysis includes the derivation of a nonlinearity for soft decision of 8-PSK symbols. Computer simulations show that soft-decision feedback provides additional robustness to the proposed soft-output LS space-time equalizer.

The above results show the relevance of probabilistic processing for equalization of the mobile radio channel in the presence of *a priori* unknown non-Gaussian interference, and demonstrate the advantages of techniques that do not rely on a statistical model of the input disturbance for application to modern TDMA cellular systems.

Appendix A
Log-MAP algorithm

For completeness, this appendix reports the derivation of equations (3.3), (3.4) of the log-MAP trellis equalizer. Since $\Pr(b_k, r_0, \ldots, r_{N-1}) = \Pr(b_k | r_0, \ldots, r_{N-1}) \cdot \Pr(r_0, \ldots, r_{N-1})$, the log-likelihood ratio (3.2) can be expressed as

$$L(b_k | r_0, \ldots, r_{N-1}) = \log \sum_{\mu_{k+1} \in S(b_k=+1)} \Pr(\mu_{k+1}, r_0, \ldots, r_{N-1})$$

$$- \log \sum_{\mu_{k+1} \in S(b_k=-1)} \Pr(\mu_{k+1}, r_0, \ldots, r_{N-1}) , \qquad \text{(A.1)}$$

where $\mu_k \triangleq (b_{k-1}, \ldots, b_{k-L+1})$ denotes the generic ISI state at time k and $S(b_k = \hat{b}_k)$ is the set of trellis states corresponding to the transmitted symbol $b_k = \hat{b}_k$. In (A.1), the quantity $\Pr(\mu_k, r_0, \ldots, r_{N-1})$ can be factorized as

$$\Pr(\mu_k, r_0, \ldots, r_{N-1}) = \Pr(\mu_k, r_0, \ldots, r_{k-1}) \cdot \Pr(r_k, \ldots, r_{N-1} | \mu_k, r_0, \ldots, r_{k-1}) . \qquad \text{(A.2)}$$

But, given μ_k and the independence of the samples n_k, the outputs $r_k, \ldots, r_{N-1}$ are independent of $r_0, \ldots, r_{k-1}$, which implies

$$\Pr(\mu_k, r_0, \ldots, r_{N-1}) = \Pr(\mu_k, r_0, \ldots, r_{k-1}) \cdot \Pr(r_k, \ldots, r_{N-1} | \mu_k) . \qquad \text{(A.3)}$$

Following [18], one obtains that each term of (A.3) can be solved recursively,

$$\Pr(\mu_k, r_0, \ldots, r_{N-1}) = [\sum_{\mu_{k-1}} \Pr(\mu_{k-1}, r_0, \ldots, r_{k-2}) \cdot e^{-\lambda(\xi_{k-1})}]$$

$$\cdot [\sum_{\mu_{k+1}} \Pr(r_{k+1}, \ldots, r_{N-1} | \mu_{k+1}) \cdot e^{-\lambda(\xi_k)}] , \qquad \text{(A.4)}$$

where

$$e^{-\lambda(\xi_k)} \triangleq \Pr(\mu_{k+1}|\mu_k) \cdot \Pr(r_k|\xi_k) \tag{A.5}$$

is the *exponentiated metric* increment for the transition $\xi_k \triangleq (\mu_k \rightarrow \mu_{k+1})$, and $\Pr(r_k|\xi_k) = \Pr(r_k|\mu_{k+1}, \mu_k)$. In (A.4), the first term represents the *forward recursion*, where the probability of the state μ_k is given by the probabilities of the previous states μ_{k-1} weighted with the transition metric $e^{-\lambda(\xi_{k-1})}$. The second factor is time-reversed and represents the *backward recursion.*

The logarithm of a sum of M exponential terms e^{α_m} can be efficiently computed as [13], [15], [12]

$$\log \sum_{m=0}^{M-1} e^{\alpha_m} = \max'\{\alpha_0, \ldots, \alpha_{M-1}\}, \tag{A.6}$$

with $\max'\{x, y\} \triangleq \max\{x, y\} + \log(1 + e^{-|x-y|})$. In fact, letting

$$f_2 \triangleq \max'\{\alpha_0, \alpha_1\} = \log(e^{\alpha_0} + e^{\alpha_1}) = \max\{\alpha_0, \alpha_1\} + \log(1 + e^{-|\alpha_0 - \alpha_1|}),$$

(A.6) can be derived recursively as [15]

$$f_3 \triangleq \max'\{\alpha_0, \alpha_1, \alpha_2\} = \log(e^{f_2} + e^{\alpha_2}) = \max\{f_2, \alpha_2\} + \log(1 + e^{-|f_2 - \alpha_2|})$$

$$\vdots$$

$$f_M \triangleq \max'\{\alpha_0, \ldots, \alpha_{M-1}\} = \log(e^{f_{M-1}} + e^{\alpha_{M-1}}) = \max\{f_{M-1}, \alpha_{M-1}\}$$

$$+ \log(1 + e^{-|f_{M-1} - \alpha_{M-1}|}).$$

Using (A.6), (A.1) can be rewritten as

$$L(b_k|r_0, \ldots, r_{N-1}) = \max'_{\mu_{k+1} \in S(b_k=+1)} \log \Pr(\mu_{k+1}, r_0, \ldots, r_{N-1})$$

$$- \max_{\mu_{k+1} \in S(b_k=-1)}{}' \log \Pr(\mu_{k+1}, r_0, \ldots, r_{N-1}) , \tag{A.7}$$

where from (A.4)

$$\log \Pr(\mu_k, r_0, \ldots, r_{N-1}) = \max_{\mu_{k-1}}{}' \{ \log \Pr(\mu_{k-1}, r_0, \ldots, r_{k-2}) - \lambda(\xi_{k-1}) \}$$

$$+ \max_{\mu_{k+1}}{}' \{ \log \Pr(r_{k+1}, \ldots, r_{N-1} | \mu_{k+1}) - \lambda(\xi_k) \} . \tag{A.8}$$

In (A.7) and (A.8), the quantity

$$\Lambda(\mu_k) \triangleq \log \Pr(\mu_k, r_0, \ldots, r_{N-1})$$

is the overall accumulated metric for the state μ_k at time k. Defining now the accumulated metrics in the forward and backward recursions as

$$\Lambda^f(\mu_k) \triangleq \log \Pr(\mu_k, r_0, \ldots, r_{k-1})$$

$$\Lambda^b(\mu_k) \triangleq \log \Pr(r_k, \ldots, r_{N-1} | \mu_k) ,$$

and using (A.3) and (A.4), one has

$$L(b_k | r_0, \ldots, r_{N-1}) = \max_{\mu_{k+1} \in S(b_k=+1)}{}' \Lambda(\mu_{k+1}) - \max_{\mu_{k+1} \in S(b_k=-1)}{}' \Lambda(\mu_{k+1}) \tag{A.9}$$

$$\Lambda(\mu_{k+1} \in S(b_k)) = \Lambda^f(\mu_k) - \lambda(\hat{\xi}_k) + \Lambda^b(\mu_{k+1} \in S(b_k)) \tag{A.10}$$

with $\hat{\xi}_k = (\mu_k \rightarrow \mu_{k+1} \in S(b_k))$, and

$$\Lambda^f(\mu_k) = \max_{\mu_{k-1}}{}' \left\{ \Lambda^f(\mu_{k-1}) - \lambda(\xi_{k-1}) \right\} , \tag{A.11}$$

$$\Lambda^b(\mu_{k+1}) = \max_{\mu_{k+2}}{}' \left\{ \Lambda^b(\mu_{k+2}) - \lambda(\xi_{k+1}) \right\} . \tag{A.12}$$

Therefore the algorithm consists of computing and storing the metrics $\Lambda^f(\mu_k)$ and $\Lambda^b(\mu_{k+1})$

corresponding to a forward and backward Viterbi algorithms, and then processing the trellis according to (A.10) and (A.9), which provides the equalizer soft-output.

For the metric increment $\lambda(\xi_k)$, taking into account that $p(r_k|b_k,\ldots,b_{k-L+1}) = p_n(r_k - \sum_{\ell=0}^{L-1} b_{k-\ell}h_\ell^{(k)})$, from (A.5) one obtains

$$\lambda(\xi_k) = -\log[\,\Pr(\mu_{k+1}|\mu_k)\cdot\Pr(r_k|\xi_k)\,]$$

$$= -\log\Pr(b_k) - \log p(r_k|b_k,\ldots,b_{k-L+1})$$

$$= -\log p_n(r_k - \sum_{\ell=0}^{L-1} b_{k-\ell}h_\ell^{(k)}) - \log\Pr(b_k)\,. \tag{A.13}$$

In a *turbo* receiver, the *a priori* information $\Pr(b_k)$ is generally provided as an *L-value*

$$L(b_k) = \log\frac{\Pr(b_k=+1)}{\Pr(b_k=-1)} = \log\frac{\Pr(b_k=+1)}{1-\Pr(b_k=+1)}\,.$$

Observing that

$$b_k\cdot L(b_k) = \log\frac{\Pr(b_k)}{1-\Pr(b_k)}\,, \tag{A.14}$$

the term $\log\Pr(b_k)$ in (A.13) can be written as

$$\log\Pr(b_k) = \frac{1}{2}b_kL(b_k) - \log(e^{L(b_k)/2} + e^{-L(b_k)/2})\,,$$

where the term $\log(e^{L(b_k)/2} + e^{-L(b_k)/2})$ does not depend on b_k being $+1$ or -1, and can be neglected. Therefore, (A.13) becomes

$$\lambda(\xi_k) = -\log p_n(r_k - \sum_{\ell=0}^{L-1} b_{k-\ell}h_\ell^{(k)}) - \frac{1}{2}b_kL(b_k)\,. \tag{A.15}$$

Appendix B
Nonlinear mean-square estimation

The general problem of *mean square* (MS) estimation of a complex random variable d in terms of another complex random variable y consists of finding the function $f(y)$ that minimizes the MS error [75]

$$\mathcal{E} \triangleq E\{|d - f(y)|^2\} . \tag{B.1}$$

(B.1) can be equivalently rewritten as

$$\mathcal{E} = E\{E\{|d - f(y)|^2|y\}\} . \tag{B.2}$$

Since the quantities $E\{|d - f(y)|^2|y\}$ are positive, one has that $\mathcal{E}$ is minimum if $E\{|d - f(y)|^2|y\}$ is minimum for every y. But

$$E\{|d - f(y)|^2|y\} = E\{|d|^2|y\} + |f(y)|^2 - f(y)E\{d^*|y\} - f^*(y)E\{d|y\} , \tag{B.3}$$

which is minimum for [114], [75]

$$f(y) = E\{d|y\} . \tag{B.4}$$

Therefore $E\{d|y\}$, the *conditional mean* of d given y, is the optimum nonlinear estimator of d (in the MS sense). The corresponding minimum value of (B.1) results

$$\mathcal{E}_{\min} = E\{E\{|d - E\{d|y\}|^2|y\}\} , \tag{B.5}$$

which is the conditional variance of d given y, averaged over all observations y.

Letting $\epsilon = d - f(y)$ with $f(y) = E\{d|y\}$, one derives

$$E\{\epsilon|y\} = E\{d|y\} - f(y) = 0 \tag{B.6}$$

and

$$E\{\epsilon f^*(y)\} = E\{E\{\epsilon f^*(y)|y\}\}$$

$$= E\{f^*(y) \cdot E\{\epsilon|y\}\} = 0\,, \tag{B.7}$$

i.e., the error ϵ is uncorrelated with $E\{d|y\}$. This implies that all the available information about d has been encoded in $E\{d|y\}$.

Since $f(y) = E\{d|y\}$ minimizes the risk $\mathcal{E} = E\{|d - f(y)|^2\}$, according to Bayes estimation theory $E\{d|y\}$ is the *Bayes estimate* of d for the cost function $C(d - f(y)) = |d - f(y)|^2$. It is possible to show [135] that the conditional mean is also the Bayes estimate for any symmetric and convex cost C, provided that the *a posteriori* density $p(d|y)$ is symmetric (about the conditional mean).

Appendix C
List of publications

C. Luschi and B. Mulgrew, "Non-parametric trellis equalization in the presence of non-Gaussian interference", in *Proc. 10th IEEE Workshop on Statistical Signal and Array Processing*, (Pocono Manor, PA, USA), pp. 201–205, Aug. 2000.

C. Luschi and B. Mulgrew, "Non-parametric trellis equalization in the presence of non-Gaussian interference", *IEEE Trans. Commun.*, accepted for publication, Aug. 2001.

C. Luschi, A. M. Kuzminskiy, P. Strauch, and B. Mulgrew, "Performance of soft-output space-time equalization for EGPRS", in *Proc. IEEE Veh. Technol. Conf.*, (Boston, MA, USA), pp. 2325–2332, Sept. 2000.

C. Luschi, A. M. Kuzminskiy, P. Strauch, and B. Mulgrew, "Soft-output space-time equalization for EGPRS", submitted to *IEEE J. Select. Areas Commun.*, Mar. 2001.

Related publications:

C. Luschi, M. Sandell, P. Strauch, and R.-H. Yan, "Iterative channel estimation, equalization, and decoding for GSM receivers", *Bell Labs Technical Memorandum*, Feb. 1999.

C. Luschi *et al.*, "Advanced signal processing algorithms for energy-efficient wireless communications", *Proc. IEEE*, vol. 88, no. 10, pp. 1633–1650, Oct. 2000.

Bibliography

[1] R. W. Lucky, "Automatic equalization for digital communication," *Bell Syst. Tech. J.*, vol. 44, pp. 547–588, Apr. 1965.

[2] J. G. Proakis and J. H. Miller, "An adaptive receiver for digital signaling through channels with intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-15, no. 4, pp. 484–497, July 1969.

[3] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, no. 9, pp. 1349–1387, Sept. 1985.

[4] D. P. Taylor, G. M. Vitetta, B. D. Hart, and A. Mämmelä, "Wireless channel equalisation," *Eur. Trans. Telecommun.*, vol. 9, no. 2, pp. 117–143, Mar. 1998.

[5] J. G. Proakis, *Digital Communications*. New York: Mc Graw-Hill, 3rd ed., 1995.

[6] S. Benedetto, E. Biglieri, and V. Castellani, *Digital Transmission Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[7] J. Salz, "Optimum mean square decision feedback equalization," *Bell Syst. Tech. J.*, vol. 52, pp. 1341–1373, Oct. 1973.

[8] C. A. Belfiore and J. H. Park, Jr., "Decision feedback equalization," *Proc. IEEE*, vol. 67, no. 8, pp. 1143–1156, Aug. 1979.

[9] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 284–287, Mar. 1974.

[10] K. Abend and B. D. Fritchman, "Statistical detection for communication channels with intersymbol interference," *Proc. IEEE*, vol. 58, no. 5, pp. 779–785, May 1970.

[11] J. F. Hayes, T. M. Cover, and J. B. Riera, "Optimal sequence detection and optimal symbol-by-symbol detection: Similar algorithms," *IEEE Trans. Commun.*, vol. COM-30, no. 1, pp. 152–157, Jan. 1982.

[12] A. J. Viterbi, "An intuitive justification and a simplified implementation of the MAP decoder for convolutional codes," *IEEE J. Select. Areas Commun.*, vol. 16, no. 2, pp. 260–264, Feb. 1998.

[13] W. Koch and A. Baier, "Optimum and sub-optimum detection of coded data disturbed by time-varying intersymbol interference," in *Proc. IEEE Globecom*, (San Diego, CA, USA), pp. 807.5.1–807.5.6, Dec. 1990.

[14] P. Höher, "TCM on frequency-selective fading channels: a comparison of soft-output probabilistic equalizers," in *Proc. IEEE Globecom*, (San Diego, CA, USA), pp. 376–381, Dec. 1990.

[15] P. Robertsson, P. Höher, and E. Villebrun, "Optimal and suboptimal maximum a posteriori algorithms suitable for turbo decoding," *Eur. Trans. Telecommun.*, vol. 8, no. 2, pp. 199–125, Mar. 1997.

[16] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "A soft-input soft-output APP module for iterative decoding of concatenated codes," *IEEE Commun. Letters*, vol. 1, no. 1, pp. 22–24, Jan. 1997.

[17] G. D. Forney, Jr., "Maximum likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 3, pp. 363–378, May 1972.

[18] G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.

[19] G. Ungerboeck, "Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems," *IEEE Trans. Commun.*, vol. COM-22, no. 5, pp. 624–636, May 1974.

[20] G. E. Bottomley and S. Chennakeshu, "Unification of MLSE receivers and extention to time-varying channels," *IEEE Trans. Commun.*, vol. 46, no. 4, pp. 464–472, Apr. 1998.

[21] L. C. Barbosa, "Maximum likelihood sequence estimators: A geometric view," *IEEE Trans. Inform. Theory*, vol. 35, no. 2, pp. 419–427, Mar. 1989.

[22] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 260–269, Apr. 1967.

[23] W. C. Jakes, *Microwave Mobile Communications*. Piscateway, NJ: IEEE Press, 1993.

[24] R. H. Clarke, "A statistical theory of mobile-radio reception," *Bell Syst. Tech. J.*, vol. 47, pp. 957–1000, July 1968.

[25] M. J. Gans, "A power spectral theory of propagation in the mobile radio environment," *IEEE Trans. Veh. Technol.*, vol. VT-21, pp. 27–38, Feb. 1972.

[26] P. Höher, "A statistical discrete-time model for the WSSUS multipath channel," *IEEE Trans. Veh. Technol.*, vol. 41, no. 4, pp. 461–468, Nov. 1992.

[27] G. L. Stüber, *Principles of Mobile Communication.* Boston: Kluwer Academic Publisher, 1996.

[28] R. Steele, ed., *Mobile Radio Communications*. Piscateway, NJ: IEEE Press, 1996.

[29] J. G. Proakis, "Adaptive equalization for TDMA digital mobile radio," *IEEE Trans. Veh. Technol.*, vol. 40, no. 2, pp. 333–341, May 1991.

[30] R. D'Avella, L. Moreno, and M. Sant'Agostino, "An adaptive MLSE receiver for TDMA digital mobile radio," *IEEE Trans. Commun.*, vol. 7, no. 1, pp. 122–129, Jan. 1989.

[31] G. Castellini, F. Conti, E. D. Re, and L. Pierucci, "A continuously adaptive MLSE receiver for mobile communications: Algorithm and performance," *IEEE Trans. Commun.*, vol. 45, no. 1, pp. 80–89, Jan. 1997.

[32] A. Baier, G. Heinrich, and U. Wellens, "Bit synchronization and timing sensitivity in adaptive Viterbi equalizers for narrowband TDMA digital mobile radio systems," in *Proc. IEEE Veh. Technol. Conf.*, pp. 377–384, June 1988.

[33] A. Baier, "Correlative and iterative channel estimation in adaptive Viterbi equalizers for TDMA mobile radio," in *Proc. ITG*, no. 107, pp. 363–368, VDE Verlag, Apr. 1989.

[34] J. Cai and D. J. Goodman, "General packet radio service in GSM," *IEEE Commun. Mag.*, pp. 122–131, Oct. 1997.

[35] A. Furuskär, S. Mazur, F. Müller, and H. Olofsson, "EDGE: Enhanced data rates for GSM and TDMA/136 evolution," *IEEE Pers. Commun.*, pp. 56–66, June 1999.

[36] B. R. Petersen and D. D. Falconer, "Suppression of adjacent-channel, cochannel, and intersymbol interference by equalizers and linear combiners," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3109–3117, Dec. 1994.

[37] S. Chen, B. Mulgrew, and S. McLaughlin, "Adaptive Bayesian equalizer with decision feedback," *IEEE Trans. Signal Proc.*, vol. 41, no. 9, pp. 2918–2927, Sept. 1993.

[38] S. Chen, S. McLaughlin, B. Mulgrew, and P. M. Grant, "Adaptive Bayesian decision feedback equalizer for dispersive mobile radio channels," *IEEE Trans. Commun.*, vol. 43, no. 5, pp. 1937–1946, May 1995.

[39] S. Chen and B. Mulgrew, "Overcoming co-channel interference using an adaptive radial basis function equalizer," *Signal Proc.*, vol. 28, no. 1, pp. 91–107, July 1992.

[40] S. Chen, S. McLaughlin, B. Mulgrew, and P. M. Grant, "Bayesian decision feedback equaliser for overcoming co-channel interference," *IEE Proc. Commun.*, vol. 143, no. 4, pp. 219–225, Aug. 1996.

[41] N. W. K. Lo, D. D. Falconer, and A. U. H. Sheikh, "Adaptive equalization for co-channel interference in a multipath fading environment," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1441–1453, Feb. 1995.

[42] G. E. Bottomley and K. Jamal, "Adaptive arrays and MLSE equalization," in *Proc. IEEE Veh. Technol. Conf.*, (Chicago, IL, USA), pp. 50–54, July 1995.

[43] G. E. Bottomley, K. J. Molnar, and S. Channakeshu, "Interference cancellation with an array processing MLSE receiver," *IEEE Trans. Veh. Technol.*, vol. 48, no. 5, pp. 1321–1331, Sept. 1999.

[44] A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Proc. Mag.*, pp. 49–83, Nov. 1997.

[45] A. J. Paulraj and B. C. Ng, "Space-time modems for wireless personal communications," *IEEE Pers. Commun.*, vol. 5, no. 1, pp. 36–48, Feb. 1998.

[46] S. L. Ariyavisitakul, J. H. Winters, and I. Lee, "Optimum space-time processors with dispersive interference: Unified analysis and required filter span," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1073–1083, July 1999.

[47] E. Lindskog, A. Ahlén, and M. Sternad, "Combined spatial and temporal equalization using an adaptive antenna array and a decision feedback equalization scheme," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 2, (Detroit, USA), pp. 1189–1192, May 1995.

[48] A. M. Kuzminskiy, C. Luschi, and P. Strauch, "Comparison of linear and MLSE spatio-temporal interference rejection combining with an antenna array in a GSM system," in *Proc. IEEE Veh. Technol. Conf.*, (Tokyo, Japan), pp. 172–176, May 2000.

[49] J. H. Winters, "Optimum combining in digital mobile radio with cochannel interference," *IEEE Trans. Veh. Technol.*, vol. VT-33, no. 3, pp. 144–155, Aug. 1984.

[50] M. V. Clark, L. J. Greenstein, W. K. Kennedy, and M. Shafi, "Optimum linear diversity receivers for mobile communications," *IEEE Trans. Veh. Technol.*, vol. 43, no. 1, pp. 47–56, Feb. 1994.

[51] P. Balaban and J. Salz, "Optimum diversity combining and equalization in digital data transmission with applications to cellular mobile radio – Part I: Theoretical considerations," *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 885–894, May 1992.

[52] J. H. Winters, J. Salz, and R. D. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 1740–1751, Feb. 1994.

[53] D. J. Costello, J. Hagenauer, H. Imai, and S. B. Wicker, "Applications of error-control coding," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2531–2560, Oct. 1998.

[54] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes," in *Proc. IEEE Int. Conf. Commun.*, (Geneva, Switzerland), pp. 1064–1070, May 1993.

[55] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. Inform. Theory*, vol. 42, no. 2, pp. 429–445, Mar. 1996.

[56] J. Hagenauer and P. Höher, "A Viterbi algorithm with soft-decision outputs and its applications," in *Proc. IEEE Globecom*, (Dallas, TX, USA), pp. 47.1.1–47.1.7, Nov. 1989.

[57] J. Hagenauer, "Soft-in/soft-out: The benefits of using soft decisions in all stages of digital receivers," in *Proc. Int. Workshop DSP Techniques Appl. Space Commun.*, (ESTEC Noordwijk, The Netherlands), Sept. 1992.

[58] A. R. Calderbank, "The art of signaling: Fifty years of coding theory," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2561–2595, Oct. 1998.

[59] C. Luschi, M. Sandell, P. Strauch, J.-J. Wu, C. Ilas, P.-W. Ong, R. Baeriswyl, F. Battaglia, S. Karageorgis, and R.-H. Yan, "Advanced signal processing algorithms for energy-efficient wireless communications," *Proc. IEEE*, vol. 88, no. 10, pp. 1633–1650, Oct. 2000.

[60] K. Giridhar, S. Chari, J. J. Shynk, and R. P. Gooch, "Joint demodulation of cochannel signals using MLSE and MAPSD algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. IV, (Minneapolis, MN, USA), pp. 160–163, Apr. 1993.

[61] K. Giridhar, J. J. Shynk, A. Mathur, S. Chari, and R. P. Gooch, "Nonlinear techniques for the joint estimation of cochannel signals," *IEEE Trans. Commun.*, vol. 45, no. 4, pp. 473–484, Apr. 1997.

[62] H. Yoshino, K. Fukawa, and H. Suzuki, "Interference cancelling equalizer (ICE) for mobile radio communication," *IEEE Trans. Veh. Technol.*, vol. 40, no. 2, pp. 333–341, May 1991.

[63] S. W. Wales, "Technique for cochannel interference suppression in TDMA mobile radio systems," *IEE Proc. Commun.*, vol. 142, no. 2, pp. 106–114, Apr. 1995.

[64] P. A. Ranta, A. Hottinen, and Z.-C. Honkasalo, "Co-channel interference cancelling receiver for TDMA mobile systems," in *Proc. IEEE Int. Conf. Commun.*, (Seattle, WA, USA), pp. 17–21, June 1995.

[65] C. Luschi and B. Mulgrew, "Non-parametric trellis equalization in the presence of non-Gaussian interference," in *Proc. 10th IEEE Workshop on Statistical Signal and Array Processing*, (Pocono Manor, PA, USA), pp. 201–205, Aug. 2000.

[66] C. Luschi and B. Mulgrew, "Non-parametric trellis equalization in the presence of non-Gaussian interference." *IEEE Trans. Commun.*, accepted for publication, Aug. 2001.

[67] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Statist.*, vol. 33, pp. 1065–1076, 1962.

[68] B. W. Silverman, *Density Estimation*. London: Chapman & Hall, 1996.

[69] A. W. Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis*. Oxford: Oxford University Press, 1997.

[70] J.-N. Hwang, S.-R. Lay, and A. Lippman, "Nonparametric multivariate density estimation: A comparative study," *IEEE Trans. Signal Proc.*, vol. 42, no. 10, pp. 2795–2810, Oct. 1994.

[71] C. Diamantini and A. Spalvieri, "Quantizing for minimum average misclassification risk," *IEEE Trans. Neural Networks*, vol. 9, no. 1, pp. 174–182, Jan. 1998.

[72] C. Luschi, A. M. Kuzminskiy, P. Strauch, and B. Mulgrew, "Performance of soft-output space-time equalization for EGPRS," in *Proc. IEEE Veh. Technol. Conf.*, (Boston, MA, USA), pp. 2325–2332, Sept. 2000.

[73] C. Luschi, A. M. Kuzminskiy, P. Strauch, and B. Mulgrew, "Soft-output space-time equalization for EGPRS." submitted to *IEEE J. Select. Areas Commun.*, Mar. 2001.

[74] S. H. Müller, W. H. Gerstacker, and J. B. Huber, "Reduced-state soft-output trellis equalization incorporating soft feedback," in *Proc. IEEE Globecom*, (London, UK), pp. 95–100, Nov. 1996.

[75] A. Papoulis, *Probability, Random Variables and Stochastic Processes.* New York: McGraw-Hill, 3rd ed., 1991.

[76] GSM-05.05, *Digital Cellular Telecommunications System (Phase 2+): Radio Transmission and Reception.* ETSI, 8.4.0 ed., 1999.

[77] J. Salz and J. H. Winters, "Effect of fading correlation on adaptive arrays in digital mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 4, pp. 1049–1057, Nov. 1994.

[78] R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed, "Overview of spatial channel models for antenna array communication systems," *IEEE Pers. Commun.*, vol. 5, no. 1, pp. 10–22, Feb. 1998.

[79] K. Samaras, C. Demetrescu, C. Luschi, and R.-H. Yan, "Capacity calculation of a packet switched voice cellular network," in *Proc. IEEE Veh. Technol. Conf.*, (Tokyo, Japan), pp. 26–30, May 2000.

[80] C.-C. Lee and R. Steele, "Signal-to-interference calculations for modern TDMA cellular communication systems," *IEE Proc. Commun.*, vol. 142, no. 1, pp. 21–30, Feb. 1995.

[81] S. Nanda, K. Balachandran, and S. Kumar, "Adaptation techniques in wireless packet data services," *IEEE Commun. Mag.*, pp. 54–64, Jan. 2000.

[82] J. Anderson, T. Aulin, and C.-E. Sundberg, *Digital Phase Modulation*. New York: Plenum Press, 1986.

[83] P. Laurent, "Exact and approximate construction of digital phase modulations by superposition of amplitude modulated pulses (AMP)," *IEEE Trans. Commun.*, vol. 34, no. 2, pp. 150–160, Feb. 1986.

[84] A. Baier, "Derotation techniques in receivers for MSK-type CPM signals," in *Proc. Eusipco*, (Barcelona, Spain), Sept. 1990.

[85] G. M. Vachula and F. S. Hill, Jr., "On optimal detection of band-limited PAM signals with excess bandwidth," *IEEE Trans. Commun.*, vol. COM-29, no. 6, pp. 886–890, June 1981.

[86] K. M. Chugg and A. Polydoros, "MLSE for an unknown channel – Part I: Optimality considerations," *IEEE Trans. Commun.*, vol. 44, no. 7, pp. 836–846, July 1996.

[87] J. O. Gustavsson, *Detection of Signals Corrupted by Non-Gaussian Disturbances*. PhD thesis, Lulea University of Technology, Lulea, Sweden, Apr. 1997.

[88] H. Meyr, M. Oerder, and A. Polydoros, "On sampling rate, analog prefiltering, and sufficient statistics for digital receivers," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3208–3214, Dec. 1994.

[89] A. Glavieux, C. Laot, and J. Labat, "Turbo equalization over a frequency selective channel," in *Proc. Int. Symposium on Turbo Codes and Related Topics*, (Brest, France), pp. 96–102, Aug. 1997.

[90] G. Bauch, H. Khorram, and J. Hagenauer, "Iterative equalization and decoding in mobile communications systems," in *Proc. Eur. Pers. Mobile Commun. Conf.*, (Bonn, Germany), pp. 307–312, Oct. 1997.

[91] G. Bauch and V. Franz, "Iterative equalization and decoding for the GSM-system," in *Proc. IEEE Veh. Technol. Conf.*, vol. 3, (Ottawa, Canada), pp. 2262–2266, May 1998.

[92] P. Strauch, C. Luschi, M. Sandell, and R.-H. Yan, "Turbo equalization for an 8-PSK modulation scheme in a mobile TDMA communication system," in *Proc. IEEE Veh. Technol. Conf.*, (Amsterdam, The Netherlands), pp. 1605–1609, Sept. 1999.

[93] J. Lodge, R. Young, P. Höher, and J. Hagenauer, "Separable MAP filters for the decoding of product and concatenated codes," in *Proc. IEEE Int. Conf. Commun.*, (Geneva, Switzerland), pp. 1740–1745, May 1993.

[94] M. Sandell, C. Luschi, P. Strauch, and R.-H. Yan, "Iterative channel estimation using soft decision feedback," in *Proc. IEEE Globecom*, (Sydney, Australia), pp. 3728–3733, Nov. 1998.

[95] J. Hagenauer, "Source-controlled channel decoding," *IEEE Trans. Commun.*, vol. 43, no. 9, pp. 2449–2457, Sept. 1995.

[96] P. Strauch, C. Luschi, M. Sandell, and R.-H. Yan, "Low complexity source controlled channel decoding in a GSM system," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, (Phoenix, AZ, USA), pp. 2571–2574, Mar. 1999.

[97] D. Chase, "Code combining – A maximum likelihood decoding approach for combining an arbitrary number of noisy packets," *IEEE Trans. Commun.*, vol. COM-33, pp. 385–393, May 1985.

[98] J.-T. Chen, J.-W. Liang, H.-S. Tsai, and Y.-K. Chen, "Joint MLSE receiver with dynamic channel description," *IEEE J. Select. Areas Commun.*, vol. 16, no. 9, pp. 1604–1615, Dec. 1998.

[99] K. Fukunaga and R. R. Hayes, "The reduced Parzen classifier," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 4, pp. 423–425, Apr. 1989.

[100] B. Jeon and D. A. Landgrebe, "Fast Parzen density estimation using clustering-based branch and bound," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 9, pp. 950–954, Sept. 1994.

[101] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ: Prentice-Hall, 2nd ed., 1999.

[102] B. Mulgrew, "Applying radial basis functions," *IEEE Signal Proc. Mag.*, vol. 13, no. 2, pp. 50–65, Mar. 1996.

[103] I. Cha and S. A. Kassam, "Interference cancellation using radial basis function networks," *Signal Proc.*, vol. 47, pp. 247–268, 1995.

[104] J. Cid-Sueiro, A. Artez-Rodriguez, and A. R. Figueiras-Vidal, "Recurrent radial basis function networks for optimal symbol-by-symbol equalization," *Signal Proc.*, vol. 40, pp. 53–63, 1994.

[105] E. A. Wan, "Neural network classification: A Bayesian interpretation," *IEEE Trans. Neural Networks*, vol. 1, no. 4, pp. 303–305, Dec. 1990.

[106] D. F. Specht, "Probabilistic neural networks for classification, mapping, or associative memory," in *Proc. IEEE Int. Conf. on Neural Networks*, vol. 1, pp. 525–532, July 1988.

[107] B. Mulgrew, "Nonlinear signal processing for adaptive equalization and multi-user detection," in *Proc. Eusipco*, (Rhodes, Greece), Sept. 1998.

[108] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," *IEEE Trans. Neural Networks*, vol. 4, no. 4, pp. 570–579, July 1993.

[109] S. K. Patra and B. Mulgrew, "Subset centre selection with fuzzy implemented radial basis function equalizer design," in *Proc. First Int. Symp. on Commun. Sys. and Digital Signal Proc.*, pp. 21–25, Apr. 1998.

[110] A. M. Kuzminskiy, "Finite amount of data effects in spatio-temporal filtering for equalization and interference rejection in short burst wireless communications," *Signal Proc.*, vol. 80, pp. 1987–1997, 2000.

[111] S. Ratnavel, A. Paulraj, and A. G. Constantinides, "MMSE space-time equalization for GSM cellular systems," in *Proc. IEEE Veh. Technol. Conf.*, vol. 1, (Atlanta, USA), pp. 331–335, Apr. 1996.

[112] J.-W. Liang, J.-T. Chen, and A. J. Paulraj, "A two-stage hybrid approach for CCI/ISI reduction with space-time processing," *IEEE Commun. Letters*, vol. 1, no. 6, pp. 163–165, Nov. 1997.

[113] D. Astély and B. Ottersten, "MLSE and spatio-temporal interference rejection combining with antenna arrays," in *Proc. Eusipco*, (Rhodes, Greece), Sept. 1998.

[114] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. I: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[115] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991.

[116] A. Kavcic and J. M. F. Moura, "The Viterbi algorithm and Markov noise memory," *IEEE Trans. Inform. Theory*, vol. 46, no. 1, pp. 291–301, Jan. 2000.

[117] D. Middleton, "Threshold detection in correlated non-Gaussian noise fields," *IEEE Trans. Inform. Theory*, vol. 41, no. 4, pp. 976–1000, July 1995.

[118] L. Izzo and M. Tanda, "Asymptotically optimum diversity detection in correlated non-Gaussian noise," *IEEE Trans. Commun.*, vol. 44, no. 5, pp. 542–545, May 1996.

[119] S. Haykin, *Adaptive Filter Theory*. Upper Saddle River, NJ: Prentice-Hall, 3rd ed., 1996.

[120] GSM-05.01, *Digital Cellular Telecommunications System (Phase 2+): Physical Layer on the Radio Path; General Description*. ETSI, 5.3.0 ed., June 1997.

[121] GSM-05.02, *Digital Cellular Telecommunications System (Phase 2+): Multiplexing and Multiple Access on the Radio Path*. ETSI, 5.2.0 ed., Nov. 1996.

[122] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*. Cambridge, MA: Cambridge University Press, 2nd ed., 1994.

[123] K. Georgoulakis and S. Theodoridis, "Efficient clustering techniques for channel equalization in hostile environments," *Signal Proc.*, vol. 58, pp. 153–164, 1997.

[124] S. Theodoridis, C. F. N. Cowan, C. P. Callender, and C. M. S. See, "Schemes for equalization of communication channels with nonlinear impairments," *IEE Proc. Commun.*, vol. 142, no. 3, pp. 165–171, June 1995.

[125] GSM-05.03, *Digital Cellular Telecommunications System (Phase 2+): Channel Coding*. ETSI, 8.4.0 ed., 1999.

[126] S. L. Ariyavisitakul and Y. G. Li, "Joint coding and decision feedback equalization for broadband wireless channels," *IEEE J. Select. Areas Commun.*, vol. 16, no. 9, pp. 1670–1678, Dec. 1998.

[127] Y.-J. Liu, M. Wallace, and J. W. Ketchum, "A soft-output bidirectional decision feedback equalization technique for TDMA cellular radio," *IEEE J. Select. Areas Commun.*, vol. 11, no. 7, pp. 1034–1045, Sept. 1993.

[128] T. Frey and M. Reinhardt, "Signal estimation for interference cancellation and decision feedback equalization," in *Proc. IEEE Veh. Technol. Conf.*, (Phoenix, AZ, USA), pp. 155–159, May 1997.

[129] E. de Carvalho and D. T. M. Slock, "Burst mode non-causal decision-feedback equalizer based on soft-decisions," in *Proc. IEEE Veh. Technol. Conf.*, (Ottawa, Canada), pp. 414–418, May 1998.

[130] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Baltimore: The Johns Hopkins University Press, 3rd ed., 1996.

[131] T. F. Chan, G. H. Golub, and R. J. LeVeque, "Algorithms for computing the sample variance: Analysis and recommendations," *The American Statistician*, vol. 37, no. 3, pp. 242–247, Aug. 1983.

[132] M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions*. New York: Dover, 1970.

[133] M. V. Eyuboglu and S. U. H. Qureshi, "Reduced-state sequence estimation with set partitioning and decision-feedback," *IEEE Trans. Commun.*, vol. 36, no. 1, pp. 13–20, Jan. 1988.

[134] S. L. Ariyavisitakul and G. M. Durant, "A broadband wireless packet technique based on coding, diversity, and equalization," *IEEE Commun. Mag.*, pp. 110–115, July 1998.

[135] H. L. V. Tree, *Detection, Estimation, and Modulation Theory*. John Wiley & Sons, 1968.