SCVT : IEEE symposium on communications and vehicular technology in the Benelux : proceedings, 3rd, Eindhoven, October 25-26 1995

Citation for published version (APA):

Smulders, P. F. M., & Meerendonk, van den, H. J. A. (Eds.) (1995). *SCVT : IEEE symposium on communications and vehicular technology in the Benelux : proceedings, 3rd, Eindhoven, October 25-26 1995.* Technische Universiteit Eindhoven.

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

PROCEEDINGS IEEE Third Symposium

on Communications and Vehicular

Technology in the Benelux

October 25-26, 1995

Eindhoven, The Netherlands

Editors

Peter Smulders and Harald van den Meerendonk

Telecommunications Division

Eindhoven University of Technology, The Netherlands

Organization

IEEE Vehicular Technology/Communications Society

Joint Chapter in the Benelux Section

and

Telecommunications Division

Eindhoven University of Technology, The Netherlands

IEEE
Benelux Section

Technische Universiteit **tu** Eindhoven
Telecommunications Division

IEEE THIRD SYMPOSIUM
ON COMMUNICATIONS AND VEHICULAR TECHNOLOGY
IN THE BENELUX

EXECUTIVE COMMITTEE MEMBERS

Prof.Dr. Ramjee Prasad

Prof.Dr. August Laloux

Mr. John Farserotu

Ir. Kiwi Smit

Prof.Dr. Luc Vandendorpe

Dr. Stan Baggen

Dr. Riccardo de Gaudenzi

Dr. Marc Durvaux

Ir. Gerard J.M. Janssen

Dr. A.S. Krishnakumar

Prof.Dr.Ir. Marc Moeneclaey

Ir. Peter F.M. Smulders

OPENING ADDRESS SPEECH
R. Prasad, TU Delft

SESSION CHAIRMEN

Session 1: TDMA systems

P. Delogne, UCL Louvain-la-Neuve

Session 2: Radio Channel Modelling

M. Herben, TU Eindhoven

Session 3: Spread Spectrum

L. Vandendorpe, UCL Louvain-la-Neuve

Session 4: Modulation & Coding

F. Willems, TU Eindhoven

Session 5: Network & Switching Architectures

J. de Stigter, TU Eindhoven

Session 6: Acquisition & Synchronization

M. Moeneclaey, RU Gent

Session 7: Optical Fiber Transmission

G. Khoe, TU Eindhoven

Session 8: Equalization & Diversity

G. Janssen, TU Delft

SYMPOSIUM CHAIRMAN'S MESSAGE

On behalf of the Executive Committee of the IEEE Vehicular Technology / Communications Society Joint Chapter, Benelux Section, I heartily welcome the participants, speakers, and session chairmen to the Third IEEE Symposium on Communications and Vehicular Technology in the Benelux (SCVT '95).

The objective of SCVT '95 is to focus the various research activities in the field of Telecommunications, namely TDMA Systems, Radio Channel Modelling, Spread Spectrum, Modulation & Coding, Network & Switching Architectures, Acquisition & Synchronization, Optical Fiber Transmission, and Equalization & Diversity.

There are eight oral sessions and one poster session. A poster session is planned to accomodate a large number of papers, to provide enough time for the oral session and to avoid parallel sessions. It is worth mentioning that there is no difference in the quality of the papers whether they belong to oral and poster sessions.

As it is customary, students will be given the best paper award in the evening of Wednesday 25 October 1995, during the Symposium Dinner.

The organisation of SCVT '95 has required significant effort and time of Peter Smulders of Eindhoven University of Technology. I must express my sincere thanks to him. Also, I would like to acknowledge the authorities of the Eindhoven University of Technology for giving the permission to organize the SCVT '95 in its campus. And of course we are all indebted to the symposium authors whose willingness to share their knowledge with us will make attendance at the SCVT '95 a very rewarding experience.

I look forward to meeting you all in Eindhoven.

Ramjee Prasad
Chairman, IEEE Vehicular Technology / Communications Society Joint Chapter in the Benelux

Contents

POSTER SESSION

Session 4:   Modulation & Coding

Session 5: Network & Switching Architectures

Session 6: Acquisition & Synchronization

Session 7: Optical Fiber Transmission

Session 8:   Equalization & Diversity

List of Authors

Capacity Investigation of an Advanced TDMA Air-Interface Concept for a 3rd Generation System

Michael Schönborn, Klaus David
DeTeMobil, P.O.Box 8865, D-48047 Münster, Germany
E-mail: schoenbo@DeTeMobil.de, kdavid@DeTeMobil.de

Abstract: After an overview of the Advanced TDMA (ATDMA) system concept, an analytical approach for the assessment of capacity is given by this paper. This is done by using the novel ATDMA burst and frame structure and comparing it against the 2nd generation Global System for Mobile Communication (GSM). This shows under which conditions a capacity increase is feasible. The capacity is defined in various ways i.e.: cell capacity [channels/cell/MHz], traffic capacity [Erlang/cell/MHz] and information capacity [Mbit/s/cell/MHz], each emphasising on a different aspect of capacity [1].

I. INTRODUCTION: THE ATDMA SYSTEM CONCEPT

Within the European RACE ATDMA project an advanced air interface and system concept has been designed by a consortium of leading manufactures, operators, and research institutes [2]. Based on this publication this section gives a brief overview over the ATDMA system design.

Besides the adaptation to different cell types in order to match the propagation characteristics in macro, micro or pico cellular environments, the ATDMA system offers several advanced control techniques. Furthermore a novel burst and frame structure has been developed which supports a variety of services such as: high quality speech with only 30ms end to end delay, data with long or short delay up to bitrates of 2Mbit/s.

All this should also allow an improved quality of service (QoS) and a larger capacity both being major potential key bottlenecks of a cellular system.

A. Transport

The transport chain for the ATDMA air interface [1] has been designed based on speech requirements and the desire to support an ISDN-B channel on a single carrier in macro cells. In addition, to benefit from relaxed propagation conditions in smaller cells, the key physical layer radio interface parameters (carrier spacing, baud rate, modulation type, etc.) have been optimised for each of the four supported

cell types. Table 1 depicts the transport parameters for this so-called static adaptation.

Cell type	long-Macro	short Macro	Micro	Pico
Modulation	GMSK	Binary Offset QAM		
Carrier Symbol Rates (kbaud)	360	450	1800	
Minimum Carrier Separation (kHz)	276.92		1107.69 = 4 x 276.92	
Frame Duration (ms)	5			
Slots per Frame	15	18	72	
Payload (symbol)	76			96
Training Sequence (symbol)	23	29		15
Tail Bits (symbol)	8			6
Guard Time (symbol)	13	12		8
Slot Length (symbol)	120	125		
Frames per multiframe	128			

Table 1: Transport parameters for static adaptation

Within cells, transport parameters can be dynamically varied to adapt to both slow and fast variations in propagation and interference conditions. For each service type, this has lead to the definition of so-called *transport modes*. A transport mode is characterised by a certain configuration of the set {modulation, error control code, amount of radio resources} which guarantees a given performance for a given level of signal-to-noise + interference. It is selected by the link adaptation process which, based on on-air parameter measurements, will trigger air interface changes from one mode to another so as to maintain the target bearer quality of service. Table 2 summarises the global channel code rates of transport modes for two example data services supported by the ATDMA system design, 9.6kbit/s and 64kbit/s respectively.

Mode number	9.6kbit/s global code rate	64kbit/s global code rate
1	1/2	1/2
2	3/8	3/8
3	1/4	1/4

Table 2: Channel code rates of transport modes for example data services

For completeness it should be mentioned that these data services satisfy different requirements concerning the delay through the radio access system. The listed transport modes for 9.6kbit/s and 64kbit/s are designed for a delay below 30ms and 300ms, respectively.

For voice services it is assumed that a UMTS speech codec to be supported by ATDMA might offer fixed network speech quality with transport modes using a gross rate of 13kbit/s and 26kbit/s, respectively. The radio access system delay is considered to be below 30ms.

B. Control

In order to cope with the current environmental conditions and user needs and to ensure high quality with optimum use of resources, several advanced control techniques form the ATDMA radio access control system [2].

These techniques range from a quality based Power Control, ensuring the desired target quality for each traffic slot independently, over Link Adaptation, a method where the coding and modulation is adapted to the radio channel characteristics, and Packet Access which allows a flexible bitrate allocation for different services to Dynamic Channel Allocation offering quite some capacity improvement for specific situations like hot spots and avoiding detailed frequency planning.

Selective repeat ARQ, Handover and Macro Diversity complete this list of control techniques.

II. CAPACITY COMPARISON

The analytical comparison in this paper is carried out by using the specified or recommended data in terms of burst and frame structure and available services in each system.

We assume various capacity definitions: cell capacity [channels/cell/MHz], traffic capacity [Erlang/cell/MHz] and information capacity [Mbit/s/cell/MHz], each emphasising on a different aspect of capacity [1]

As the allocated bandwidth has a great impact on the capacity of any cellular system it is necessary to find a way of eliminating this dependence. In this comparison we assume for each system a common bandwidth of 2 x 25MHz. The advantage of this procedure is that almost any calculated figure can directly be compared between the different systems. On the other hand it has to be kept in mind that the Erlang B formula used to calculate the traffic load for voice services is not a linear function of available

channels per cell. Hence the allocated bandwidth has some impact on the normalised traffic capacity.

A. Definition of Parameters

In this section a short definition is given for some parameters used in this comparison. They are introduced in the order of their appearance in the sections below.

No. of speech duplex channels: To evaluate the number of speech duplex channels equation (1) is used.

$$\textit{no. of speech duplex channels} = \frac{25\,MHz}{carrier\ separation\ [MHz]} \cdot \frac{slots}{carrier} \quad (1)$$

baud rate: This is the reciprocal value of the bit duration.

Theor. maximum information capacity: This capacity figure is described by fraction (2)

$$\frac{baud\ rate}{carrier\ separation} \quad in \left[\frac{Mb\,/\,s}{MHz}\right] \quad (2)$$

Information capacity for data services: This capacity figure is described by fraction (3)

$$\frac{net\ or\ gross\ data\ rate\ users\,/\,carrier}{carrier\ separation} \quad in \left[\frac{Mb\,/\,s}{MHz}\right] (3)$$

Theor. maximum traffic capacity for voice services: This capacity figure is described by fraction (4)

$$\frac{slots\,/\,carrier}{carrier\ separation} \quad in \left[\frac{users}{MHz}\right] \quad (4)$$

Theoretical minimum cluster size: Cluster size, considering path loss slope γ, shadowing margin σ and minimum C/I+N (in [dB]) calculated as given in equation (5) [3].

$$K_{theor} = \frac{1}{3} \cdot \left(6 \cdot 10^{\frac{\sigma + min\,C/(I+N)}{10}}\right)^{\frac{2}{\gamma}} \quad (5)$$

Actual minimum cluster size: Mapped theoretical minimum cluster size onto the possible cluster sizes $K = i^2 + ij + j^2$ (i, j positive integers and 0); i.e. 1, 3, 4, 7, 9 ... .

Channels / cell: If the total amount of speech duplex channels are evenly spread over a cluster with the cluster size K each cell (theoretically) offers

$$\frac{channels}{cell} = \frac{no.\ of\ speech\ duplex\ channels}{K} \quad (6)$$

Note that non-DCA systems, like GSM, can offer only integer multiples of a set of speech duplex channels, e.g. 8, 16, 24 ... , in a cell. Anyway, the theoretical figure is used throughout this comparison.

Normalised traffic capacity [Erlang/cell/MHz]: This is the max. traffic load divided by 25 MHz. While the max. traffic load is evaluated by the Erlang B formula.

B. System Presentation

In this section the burst and frame structure of the systems considered are briefly described. Also various capacity figures for an assumed spectrum of 50MHz are given.

1. ATDMA (short macro cell)

A complete description of this system is given in [2]. Some key parameters are shown in table 3. As already stated in the description of the ATDMA system design there are four supported cell types. Out of these the short macro cell is selected as GSM is often used as a macro cellular system, too.

carrier separation	276.92
modulation scheme	Binary Offset QAM
channels / carrier	18
no. of speech duplex channels	1625
selected bit rates: speech (gross) data (net / gross) baud rate	13kbit/s 9.6kbit/s / 26.4kbit/s 450kbit/s
theor. maximum information capacity [Mbit/s / MHz]	1.625
information capacity for data services [Mbit/s / MHz]:	(9.6/25.6 data) 0.312 / 0.858

Table 3: ATDMA parameters
(short macro and micro cell)

2. GSM

A complete description of this system is given in [4]. Some key parameters are shown in table 4.

carrier separation	200kHz
modulation scheme	GMSK
channels / carrier	8
no. of speech duplex channels	1000
selected bit rates: speech (gross) data (net / gross) baud rate	22.8kbit/s 9.6kbit/s / 22.8kbit/s 270.8kbit/s
theor. maximum information capacity [Mb/s / MHz]	1.354
information capacity for data services [Mb/s / MHz]:	(9.6/22.8 data) 0.384 / 0.912

Table 4: GSM parameters

C. Theoretical Capacity Limits

1. ATDMA (short macro cell) vs. GSM

	ATDMA (short macro)	GSM
data service [kbit/s]	9.6 / 26.4	9.6 / 22.8
theor. maximum information capacity [Mb/s / MHz]	1.625	1.354
theor. maximum information capacity for data services [Mb/s / MHz]:	0.312 / 0.858	0.384 / 0.912
speech service gross [kbit/s]	13	22.8 (11.4)
theor. maximum traffic capacity [users / MHz]	65.0	40.0 (80.0)

Table 5: Theoretical capacity limits for
ATDMA and GSM

D. Capacity limits concerning quality requirements

These numbers are calculated for a path loss slope of $\gamma = 3.6$ and a shadowing margin $\sigma = 6.6$dB in equation (5).

3

	ATDMA (short macro)	GSM
voice service (gross) [kbit/s]	13 .. 26	22.8 (11.4)
min. C/(I+N) [dB]	3.1	9
theor. minimum cluster size:	3.12	6.64
actual minimum cluster size:	4	7
normalised cell capacity [channels/cell / MHz]	16.24	5.72 (11.44)
normalised traffic capacity [Erlang/cell / MHz]	≥ 15.2	5.0 (10.5)

Table 6: ATDMA vs. GSM concerning voice service

III. DISCUSSION AND CONCLUSION

After a brief presentation of the various transport and control features of the ATDMA system concept the investigation of this paper focused on a capacity comparison of ATDMA with GSM. This was based purely on the burst and frame structures, not considering the capacity effect of, for example, APC.

The results are:

* For the same cluster size ATDMA can offer 1.62 times more speech channels/cell than GSM

* As the consideration of the minimum C/I values show, there is an indication that ATDMA can be operated one cluster size smaller than GSM. This could result in a capacity advantage of about 3 for ATDMA ($K=4$) compared to GSM ($K=7$).

For the interpretation of these and other results, it is very important to be aware that:

* ATDMA offers these capacity increases in combination with increases QoS.

* The ATDMA system concept provides a whole range of techniques which will further increase this capacity advantage. For this study the project developed various powerful testbeds [5].

* The exact values of these capacity figures are very dependant on the assumptions made and on the scenarios chosen.

REFERENCES

[1] K. David, T. Kassing, D. Florack, "ATDMA Test Objectives", Proceedings of 5th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications; pp. 691 - 695; The Hague, 18 - 22 September 1994.

[2] A. Urie, M. Streeton, C. Mourot, "An Advanced TDMA Mobile Access System for UMTS", IEEE Personal Communications Special Issue; Vol.2 No.1; pp. 38 - 47; February 1995.

[3] Lee, W.C.Y., "Mobile Cellular Telecommunications Systems", McGrawHill; New York 1989.

[4] ETSI, "GSM Recommendations", GSM 01.02 - 12.21.

[5] D. Grillo, N. Metzner, and E.D. Murray, "Testbeds for assessing the Performance of a TDMA based Radio Access Design for UMTS", IEEE PCM, Vol. 2, No. 2, pp.36-45, April 95

A new method for reducing the power consumption of portable handsets in TDMA mobile systems : Conditional Equalization

Lionel Husson
Husson@supelec.fr

Jean-Claude Dany
Jean-Claude.Dany@supelec.fr

Supélec, Service Radioélectricité et Électronique
Plateau de Moulon, 91192 Gif sur Yvette Cédex FRANCE

Abstract :

In TDMA mobile radio systems, the constraint of the size and the autonomy of the portable handset becomes a major interest. In this paper, we propose a new method, conditional equalization, which minimizes the power consumption due to equalization. Its principle is to resort to the processing of equalization only when it is mostly needed and to abstain from using it when the distorsions due to the propagation channel are weak.

The theoretical performances of our method are analysed for a 2-path Rayleigh fading channel. At each time-slot, the receiver estimates the need of equalization with a deterministic criterion. The criteria proposed in this paper resort to thresholds chosen to provide a good compromise between degradation and percentage of spared equalization.

The performances are also analysed for the recommended COST 207 channels. The extensive simulations we carried out show that conditional equalization can ensure, at 9 dB, a degradation inferior to 0.5 dB while providing a respective gain of 43% spared equalization for Bad Urban, 62 % for Hilly Terrain and 79% for Typical Urban models.

Conditional equalization is shown to greatly improve the autonomy of the receiver by drastically reducing the power consumption due to equalization without a noticeable degradation of the performances.

I. INTRODUCTION

In mobile communication systems, the propagation channel imposes perturbation effects on the transmission link (attenuation, distortions, and fluctuations). It has direct consequences on the functionalities which are needed in the transmitter and the receiver. Estimation of the channel and equalization are a means of obtaining the performances required in mobile radio TDMA systems.

With the development of mobile systems, a major constraint is the desired size and autonomy of the portable handset. Some techniques are used to minimize the power consumption such as control of power or vocal activity detection. In this article we propose a device to limit the consumption due to equalization. The idea is to adapt the power resources of the handset depending on the quality of the received signal. Its principle is to resort to the processing only when it is mostly needed, and to abstain from using it when the distortions due to the channel are weak.

The paper is organised as follows. In section II, we present how is lead the evaluation of the performances of the receiver in terms of probability of error. In section III, we explain the principle of conditional equalization. Sections IV describes the used criteria and displays the theoretical performances of this device over a 2-path Rayleigh fading channel. Simulation results are the core of section V. The method is used over the channels recommended by the COST 207 final report. We show its applicability and its interest for reducing the power consumption of portable handsets in TDMA systems. Finally, section VI contains the concluding remarks.

II. EVALUATION OF ERROR PROBABILITIES IN THE CASE OF MULTIPATH CHANNELS

In this section, we present how are computed the error probabilities. The whole analytical computation will only be expressed for a Rayleigh fading channel with two paths of equal energies, which is a typical case [3].

a) notations and hypotheses

The received signal at time $t_n = n.T_s$ (T_s denotes the symbol period.) is :

$$r_n = s_n.c_0 + \sum_{i=1}^{N} c_i . s_{n-i} + \eta_n$$

where s_i is the ith symbol and η_n an additive noise.

$H(z)$ denotes the discrete response of the channel

$$H(z) = \sum_{i=0}^{N} c_i . z^{-i}$$

$$c_i = |c_i|. \exp(j\theta_i)$$

We will only consider binary modulation where symbols s_i take with equal probabilities the values $+A$ and $-A$. We define γ_i as :

$$\gamma_i = \frac{A^2}{\sigma_\eta^2} . |c_i|^2$$

The received bit-energy-to-noise ratio is

$$\gamma_b = \sum_{i=0}^{N} \gamma_i = \frac{E_b}{N_0}$$

E_b is the energy per bit of the received signal. η is the baseband equivalent of an additive white gaussian noise of double-sided spectral density $N_0/2$.

b) Optimum performances

b-1) fixed channel

For a binary phase shift keying transmission and for any fixed channel with additive white gaussian

noise, the matched filter bit error probability is the limit of achievable performances. This bit error probability is given by :

$$P_e(\gamma_b) = f(\gamma_0, \gamma_1, ..., \gamma_N) = \frac{1}{2} \, \text{erfc}(\sqrt{\gamma_0 + \gamma_1 + ... + \gamma_N}) \quad (1) \quad {}^{1}$$

b-2) Rayleigh fading channel

In the case of an independent multipath Rayleigh fading channel, the received energy-to-noise ratio γ_b fluctuates. Each γ_i can be considered as a random variable. The bit-error probability function must be averaged over the possible channel responses.

$$P_e = \int_{\gamma_0} \int_{\gamma_1} .. \int_{\gamma_N} f(\gamma_0, \gamma_1, ..., \gamma_N) \, p(\gamma_0) \, p(\gamma_1) \, .. \, p(\gamma_N)$$
$$d\gamma_0 \, d\gamma_1 \, .. \, d\gamma_N$$

Each of the $\{\gamma_i\}$ is distributed according to a chi-squared distribution with 2 degrees of freedom.

$$p(\gamma_i) = \frac{1}{b_i} \cdot \exp(-\gamma_i / b_i)$$

where b_i is the average signal-to-noise ratio (SNR) for the ith path.

The average SNR is defined by $\Gamma_b = E[\gamma_b] = \sum_{i=0}^{N} b_i$, where $E[\ ]$ represents expectation.

b-3) 2-path Rayleigh fading channel

If we consider a Rayleigh fading channel with 2 paths with equal average SNR b, the above formula becomes :

$$P_e = \int_{\gamma_0} \int_{\gamma_1} f(\gamma_0, \gamma_1) \, p(\gamma_0) \, p(\gamma_1) \, d\gamma_0 \, d\gamma_1$$
with
$$p(\gamma_i) = \frac{1}{b} \cdot \exp(-\gamma_i / b)$$

$$P_e = \frac{1}{2} - \frac{3+2b}{4(b+1)} \cdot \sqrt{\frac{b}{b+1}} \qquad (\Gamma_b = 2b).$$

c) Performances in absence of equalization

c-1) fixed channel

The channel is estimated by the receiver. The estimation of the channel enables us to decide which path has the greatest energy, $|c_{max}|^2$. The reception system takes a decision relatively to this path. In the

following, we consider a perfect estimation of the channel.

The decision is done on the real part of r_n / c_{max}. The contribution of the other paths is their projection on the direction defined by the greatest energy path. Let x denote the residual intersymbol interference (ISI) term.

$$x = \sum_{i \neq i_{max}} c_i \cdot s_{n-i} \cdot \cos(\theta_i - \theta_{max})$$

where i_{max} corresponds to the path having the greatest energy.
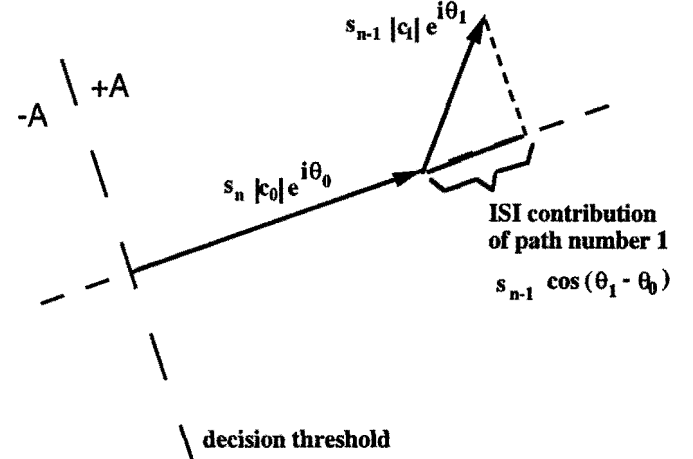
Fig. 1. effects of intersymbol interference example : 2-path channel with $s_n = +A$ and $s_{n-1} = +A$

For a fixed x, the error probability is given by :

$$p_e(x) = \frac{1}{2} \, \text{Prob}(\, \text{Real}(\, A + x / c_{max} + \eta_n / c_{max} \,) < 0)$$
$$+ \frac{1}{2} \, \text{Prob}(\, \text{Real}(\, -A + x / c_{max} + \eta_n / c_{max} \,) > 0)$$

$$p_e(x) = \frac{1}{4} \, \text{erfc}(\frac{c_{max} A + x}{\sigma_\eta}) + \frac{1}{4} \, \text{erfc}(\frac{c_{max} A - x}{\sigma_\eta})$$

It follows that the error probability is :
$$P_e = E_x[\, p_e(x) \,]$$

c-2) 2-path fixed channel

We consider a 2-path channel and we determine which path has the greatest energy. If $\gamma_0 > \gamma_1$, the above error probability is simply expressed as :

$$P_e = \frac{1}{4} \text{erfc}(\sqrt{\gamma_0} + |\cos(\theta)| \sqrt{\gamma_1}) + \frac{1}{4} \text{erfc}(\sqrt{\gamma_0} - |\cos(\theta)| \sqrt{\gamma_1}) \quad (2)$$

where $\theta = \theta_1 - \theta_0$

We denote in the following this function by $f_{01}(\gamma_0, \gamma_1, \theta)$. Symmetrically, when $\gamma_1 > \gamma_0$ the error probability is denoted $f_{10}(\gamma_0, \gamma_1, \theta)$.

c-3) 2-path Rayleigh fading channel

We consider a Rayleigh fading channel with 2 independent paths. The bit error probability must be averaged over all the channel realizations. Each of the

---

^{1} $\quad \text{erfc}(x) = 1 - \text{erf}(x) \text{ and } \text{erf}(x) = \frac{2}{\sqrt{\Pi}} \int_0^x e^{-t^2} \, dt$

6

$\{\gamma_i\}$ is distributed according to a chi-squared distribution with 2 degrees of freedom. $\theta = \theta_1 - \theta_0$ is distributed according to a uniform distribution on $[0, 2\Pi]$. Considering the symmetries of the formulation, the integration can be led on $[0, \Pi/2]$.

$$P_e = \int_{\theta=0}^{\frac{\Pi}{2}} \int_{\gamma_0=0}^{\infty} \int_{\gamma_1=0}^{\gamma_0} f_{01}(\gamma_0, \gamma_1, \theta) \, p(\gamma_0)p(\gamma_1)p(\theta) \, d\gamma_0 d\gamma_1 d\theta$$

$$+ \int_{\theta=0}^{\frac{\Pi}{2}} \int_{\gamma_1=0}^{\infty} \int_{\gamma_0=0}^{\gamma_1} f_{10}(\gamma_0, \gamma_1, \theta) \, p(\gamma_0)p(\gamma_1)p(\theta) \, d\gamma_0 d\gamma_1 d\theta$$

Figure 2 represents the bit error rate (BER) with optimum equalization and without any equalization in the case of a Rayleigh fading channel with 2 independent paths of equal average energies.
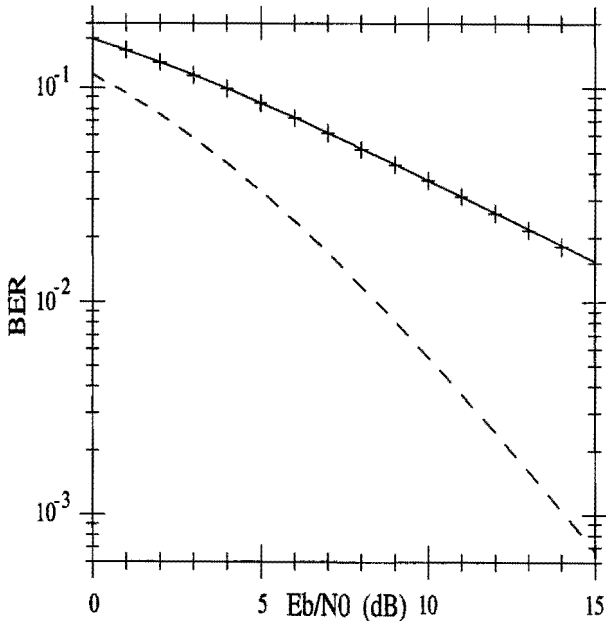
Fig. 2. BER for Rayleigh fading channel with 2 paths of equal average energies
(- - -) optimum : analytical matched filter bound
(---) no equalization : analytical computation
(+) no equalization : Monte Carlo simulation

III. CONDITIONAL EQUALIZATION

The purpose of conditional equalization is to spare the handset resources, by resorting to the complete processing of the equalizer only in the most severe distortion cases which are those who imply the degradation of the performances.

At each time-slot, the receiver estimates, by his own, the need of resorting to equalization. The proposed method does not need any additive signalling on the network. When the channel is considered as weakly distortive by the receiver, the processing is limited to a division of the demodulated

signal by the greater energy path coefficient c_{max}, followed by a decision on the real part.
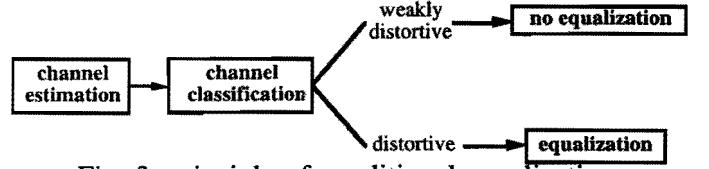
Fig. 3. principle of conditional equalization.

This device implies the definition of a criterion to make a partition between "weakly distortive channels" and "distortive channels". We propose deterministic criteria which enable the receiver to decide at each time-slot if equalization will be used or not. In the following section, we describe the two proposed criteria, and analyse their analitycal performances in the case of a 2-path Rayleigh fading channel.

IV. ANALYTICAL PERFORMANCES OF CONDITIONAL EQUALIZATION

A. Definition of a criterion : C1

We define a criterion (denoted by C1) based on the paths' energies : If the energy of a path is far superior than the total energy of the other paths, no equalization will be led.
For a 2-path channel this criterion is expressed by :

$$\boxed{\gamma_0 < \alpha \, \gamma_1 \text{ or } \gamma_1 < \alpha \, \gamma_0} \Rightarrow \text{no equalization}$$

$\boxed{\text{else}} \Rightarrow$ equalization
with $\alpha < 1$

a) Error probability

We consider a Rayleigh fading channel with 2 independent paths. Equalization is optimal equalization conditionally to criterion C1.
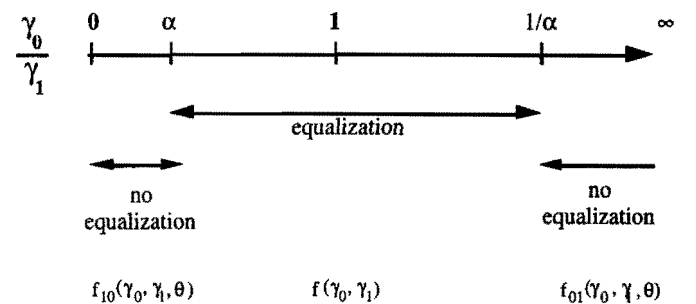
Fig. 4. domains for the computation of the error probability when using criterion C1

To evaluate the error probabilily, we lead the integration of the conditional error probability on γ_0, γ_1 and θ. The part of the computation corresponding to the "no equalization" case is the integration of the functions $f_{01}(\gamma_0, \gamma_1, \theta)$ and $f_{10}(\gamma_0, \gamma_1, \theta)$ defined by formula (2), which are θ-dependent. The other part, corresponding to the equalization case is obtained by integration of the function $f(\gamma_0, \gamma_1)$ (1), which is independent of θ. We can represent the error probability computation relatively to this criterion on a

7

γ_0/γ_1 axis (Fig. 4.). The error probability can be expressed using the functions defined above as :

$$P_e = \int_{\theta=0}^{\frac{\Pi}{2}} F_{01}(b_0,b_1|\cos(\theta)|^2,\alpha|\cos(\theta)|^2)\, p(\theta)\, d\theta$$
$$+ \int_{\theta=0}^{\frac{\Pi}{2}} F_{10}(b_0|\cos(\theta)|^2,b_1,\alpha|\cos(\theta)|^2)\, p(\theta)\, d\theta$$
$$+ E(b_0,b_1,\alpha)$$

where $F_{01}(b_0,b_1,\alpha)$ (and its symmetrical $F_{10}(b_0,b_1,\alpha)$) and $E(b_0,b_1,\alpha)$, representing the integration of the error probability on the two parameters γ_0, and γ_1 with a convenient bound, are defined by :

$$F_{01}(b_0,b_1,\alpha) = \int_{\gamma_0=0}^{\infty} \int_{\gamma_1=0}^{\alpha\gamma_0} [\frac{1}{4}\mathrm{erfc}(\sqrt{\gamma_0}+\sqrt{\gamma_1})+\frac{1}{4}\mathrm{erfc}(\sqrt{\gamma_0}-\sqrt{\gamma_1})]$$
$$p(\gamma_0)p(\gamma_1)\, d\gamma_0 d\gamma_1$$

$$E(b_0,b_1,\alpha) = \int_{\gamma_0=0}^{\infty} \int_{\gamma_1=\alpha\gamma_0}^{\gamma_0/\alpha} \frac{1}{2}\mathrm{erfc}(\sqrt{\gamma_0+\gamma_1})\, p(\gamma_0)p(\gamma_1)\, d\gamma_0 d\gamma_1$$

γ_0 and γ_1 are distributed according to a chi-squared distribution with 2 degrees of freedom and θ is uniformly distributed over $[0, \frac{\Pi}{2}]$

b) Correspondence between the choice of α and the percentage of equalization

The percentage of "no equalization" is related to the probability that either $\gamma_0<\alpha\,\gamma_1$ or $\gamma_1<\alpha\,\gamma_0$.
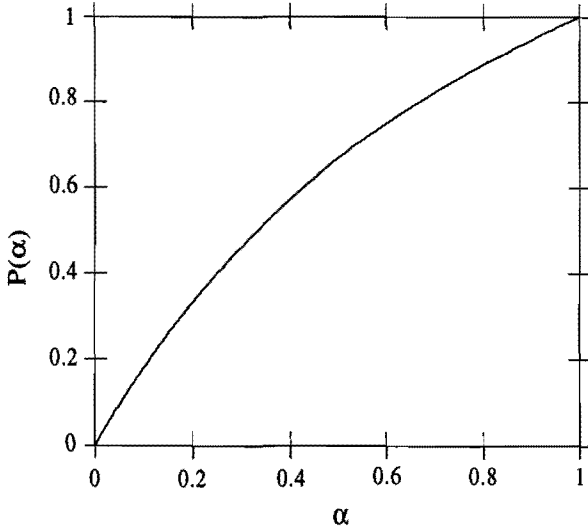
Fig. 5. Percentage of no equalization $P(\alpha)$ in the case of a Rayleigh fading channel with 2 paths of equal average energies with criterion C1.

In the case of Rayleigh fading, a simple computation leads to the percentage of no equalization which is :

$$P(\alpha) = \frac{\alpha}{\alpha+\dfrac{b_0}{b_1}} + \frac{\alpha}{\alpha+\dfrac{b_1}{b_0}}$$

c) Performances

Figure 6 shows the theoretical performances of an optimum equalizer with conditional equalization relatively to criterion C1, in the case of a Rayleigh fading channel with 2 paths of equal average energies.
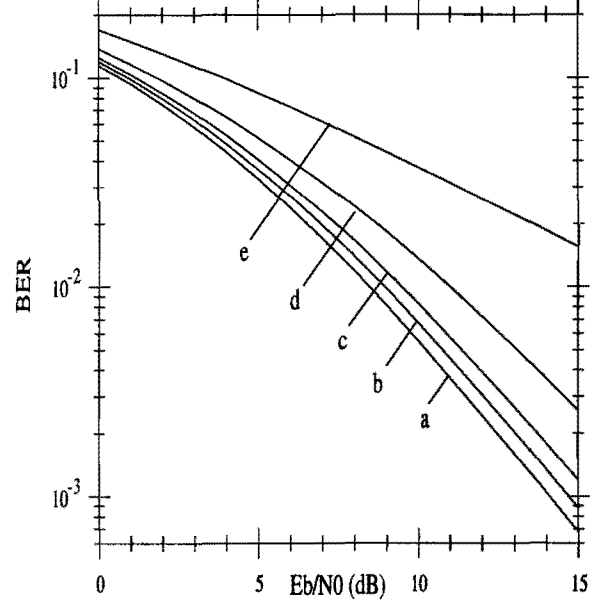
Fig. 6. performances of conditional equalization with criterion C1 and optimal equalization.
optimum equalization used (a) 100 % of time [α=0], (b) 66 % of time [α=0.2], (c) 50 % of time [α=0.3], (d) 33 % of time [α=0.5], (e) 0 % of time [α=1].

The degradation at 9 dB, when using the equalizer 50 % of the time, for instance, corresponds to 1.1 dB. This result shows that it is possible to spare a percentage of the power consumption of the equalizer without a great damage to the BER performances.

B. Definition of a second criterion : C2

We define a second criterion (denoted by C2) based on the ISI contribution : We determine which path has the greatest energy. Then, if the projection of the other paths on the axis defined by this greatest energy path is far inferior, the intersymbol interference is considered as negligible and no equalization is led. In the case of 2-path channel, this criterion is expressed by :

$$\left\{ \begin{array}{c} \gamma_0<\gamma_1 \\ \& \\ |\cos(\theta)|^2\,\gamma_0<\alpha\,\gamma_1 \end{array} \right. \text{or} \left\{ \begin{array}{c} \gamma_1<\gamma_0 \\ \& \\ |\cos(\theta)|^2\,\gamma_1<\alpha\,\gamma_0 \end{array} \right.$$

$\Rightarrow$ no equalization

$\boxed{\text{else}} \Rightarrow$ equalization

with $\alpha<1$ and $\theta= \theta_1-\theta_0$

This criterion explicitly depends on θ.

8

We must analyse two cases to evaluate the error probabilities :

i. $\theta \in [0, \arccos(\sqrt{\alpha})]$

then $\alpha / |\cos(\theta)|^2 < 1$ and the criterion becomes :

$$\boxed{|\cos(\theta)|^2 \gamma_0 < \alpha \gamma_1 \text{ or } |\cos(\theta)|^2 \gamma_1 < \alpha \gamma_0}$$
$\Rightarrow$ no equalization
$\boxed{\text{else}} \Rightarrow$ equalization

If we define $\alpha' = \alpha/|\cos(\theta)|^2$, then we can represent the computation related to this case on the γ_0/γ_1 axis :
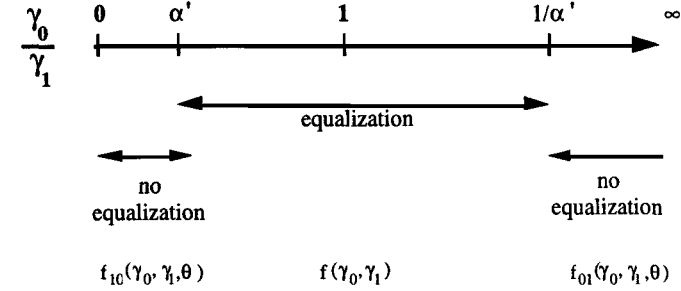
Fig. 7. domains of computation for the error probability when $\alpha / |\cos(\theta)|^2 < 1$

ii. $\theta \in [\arccos(\sqrt{\alpha}), \Pi/2]$

then $\alpha / |\cos(\theta)|^2 > 1$ and the criterion becomes :

$$\boxed{\gamma_0 < \gamma_1 \text{ or } \gamma_1 < \gamma_0} \Rightarrow \text{no equalization}$$
$\boxed{\text{else}} \Rightarrow$ equalization
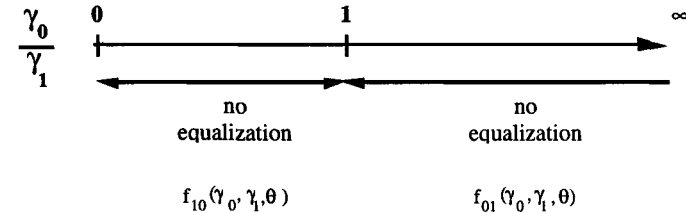
Fig. 8. domains of computation for the error probability when $\alpha / |\cos(\theta)|^2 > 1$

The condition of "no equalization" is always true.

a) Error probability

We consider a Rayleigh fading channel with 2 paths and an optimal equalization conditionally to criterion C2. To evaluate the bit error probability, the integration on θ is led considering the two cases. The resulting error probability can be expressed by using the previous defined functions F_{01}, F_{10} and E.

b) Correspondence between α and percentage of equalization

Figure 9 represents the percentage of spared equalization for a Rayleigh fading channel with 2 paths of equal average energies
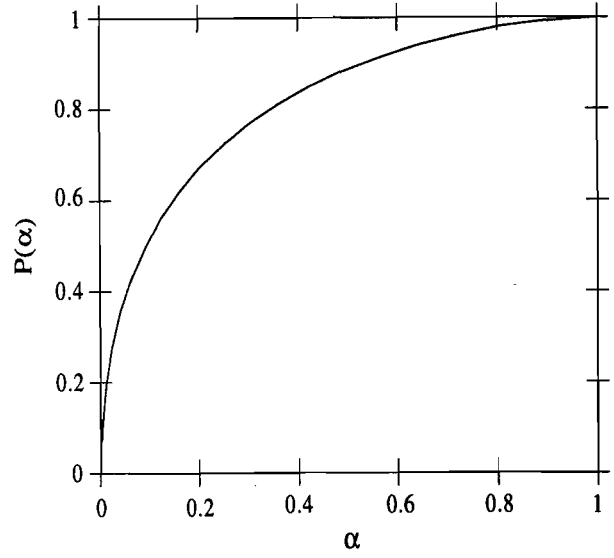
Fig. 9. Percentage of spared equalization $P(\alpha)$, Rayleigh fading channel with 2 paths of equal average energies with criterion C2.

c) Performances

BER performances for a conditional optimum equalization using criterion C2, for a Rayleigh fading channel with 2 paths of equal energies are displayed on figure 10.
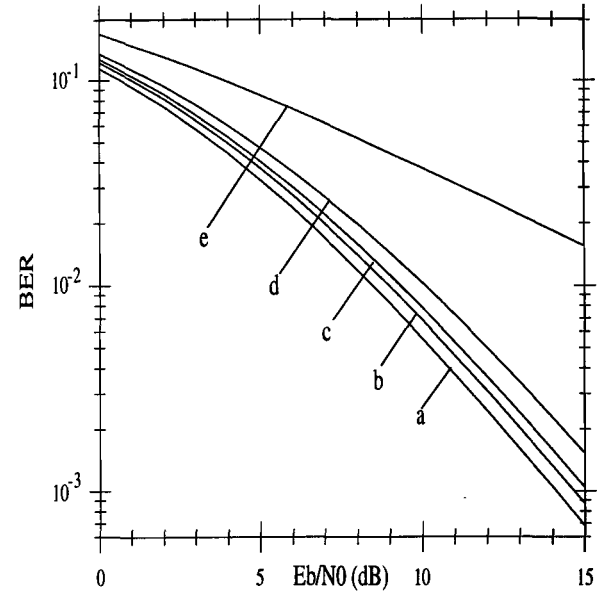
Fig. 10. performances of conditional equalization with criterion C2 and optimal equalization.
optimum equalization used (a) 100 % of time [$\alpha=0$], (b) 65 % of time [$\alpha=0.05$], (c) 50 % of time [$\alpha=0.1$], (d) 35 % of time [$\alpha=0.2$], (e) 0 % of time [$\alpha=1$].

The criterion C2, using more information about the channel, is more efficient than criteria C1. When using the equalizer 50 % of the time the loss of energy at 9 dB is close to 0.8 dB, which id inferior to the degradation obtained with criterion C1.

Figure 11 explicitly shows that criterion C2 enables to reach better performances than criterion C1 for a given percentage of no equalization. Besides it shows than under 50 % of spared equalization, the degradation is relatively low.
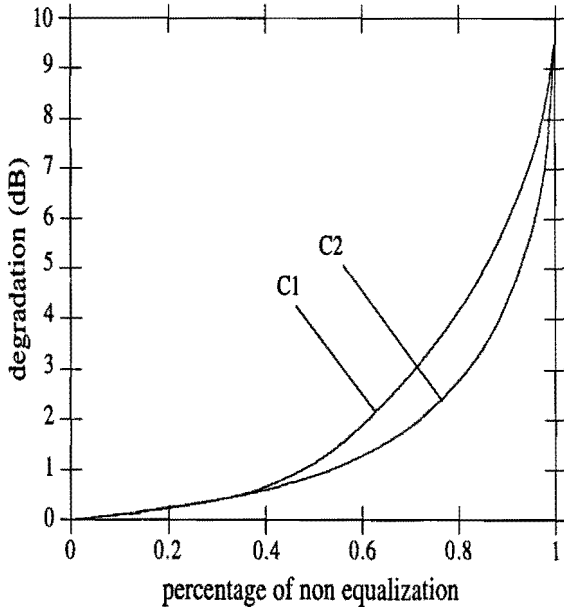
Fig. 11. comparison of the degradation of performances for conditional optimal equalization with criteria C1 an C2 relatively to the performances at E_b/N_0 = 9 dB for unconditional optimal equalization

These theoretical results show that, when using an relevant criterion, conditional equalization enables to spare a great percentage of power consumption and time spent in the equalizer with a limited damage of the bit error rate performance. Consequently, conditional equalization appears as a promising technique.

V. SIMULATIONS
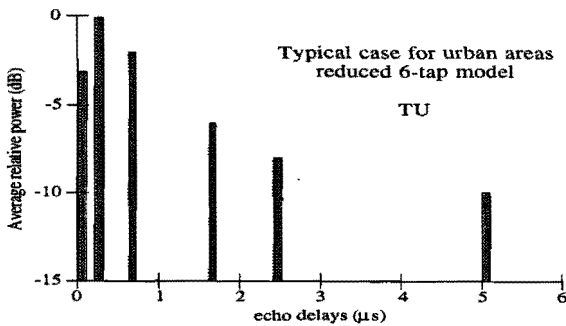
Fig. 12. profile for the Typical Urban case, 6-tap model

We carried out simulations of Conditional Equalization (DFE) using the most efficient criterion, C2, for the channels proposed by the COST 207 [6]. Used modulation is 2PSK. Figures 12 to 14 display the profile of the considered channels : Typical Urban (TU), Bad Urban (BU) and Hilly Terrain (HT) models.
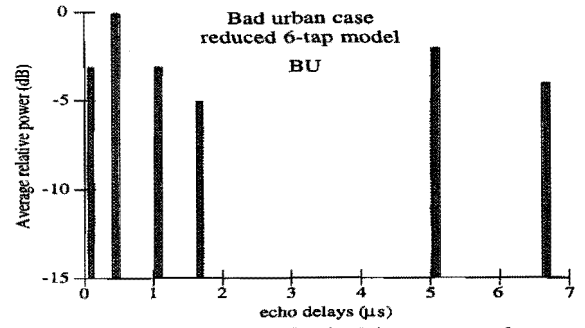
Fig. 13. profile for the Bad Urban case, 6-tap model
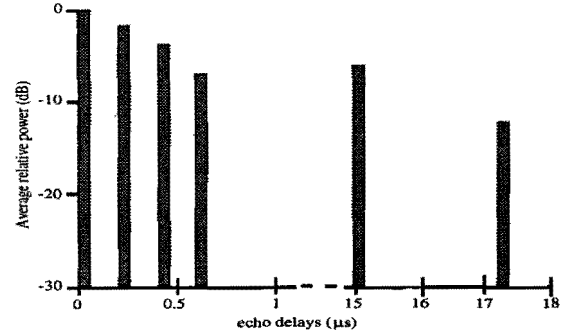
Fig. 14. profile for the Hilly Terrain case, 6-tap model
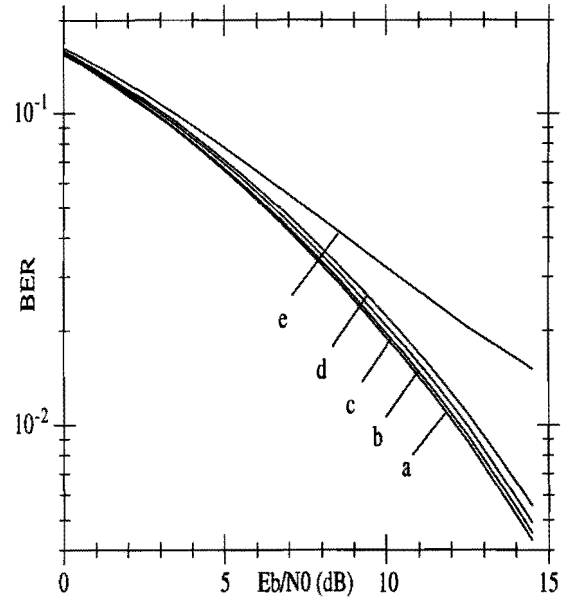
Fig. 15. performances of conditional equalization in the Typical Urban case:
DFE equalizer used (a) : 100% of time [α=0],(b) : 49% of time [α=0.1], (c) : 27% of time [α=0.2], (d) : 17% of time [α=0.3], (e) : 0% of time [α=∞].

The simulations have been carried out using Monte Carlo method over 7500 channel realizations, estimation of the channels being supposed perfect. Figures 15 to 17 display the result of the use of a conditional DFE over the considered channels.
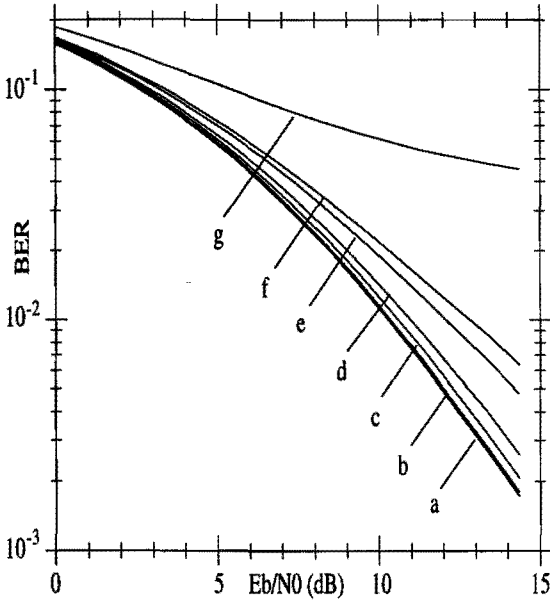
Fig. 16. performances of conditional equalization in the Bad Urban case :
DFE equalizer used (a) : 100% of time [α=0], (b) : 82% of time [α=0.1], (c) : 65% of time [α=0.2], (d) : 53% of time [α=0.3], (e) 36% of the time [α=0.5], (f) 30% of the time [α=0.6], (g) : 0% of time [α=∞].
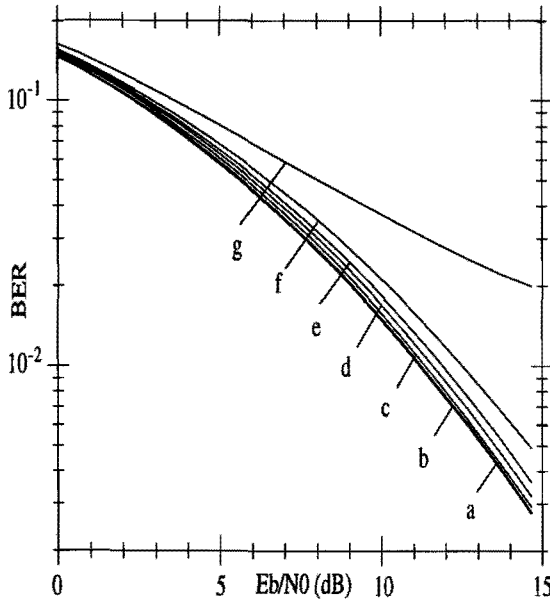
Fig. 17. performances of conditional equalization in the Hilly Terrain case :
DFE equalizer used (a) : 100% of time [α=0], (b) : 66% of time [α=0.1], (c) : 45% of time [α=0.2], (d) : 35% of time [α=0.3], (e) 26% of the time [α=0.4], (f) 18% of the time [α=0.6], (g) : 0% of time [α=∞].

It appears that, for mobile radio channels, conditional equalization enable to spare a great percentage of equalization without an important loss of performances. However, due to the different profiles of the considered channels (e.g. delay spread [7]), conditional equalization does not enable to reach for the Bad Urban channel a compromise between spared equalization and degradation as good as for the Bad Urban channel than for Hilly Terrain channel and Typical Urban channel (avoiding half of equalizations leads, at 9 dB, to a respective degradation of 0.6 dB, 0.2 dB and 0.1 dB).
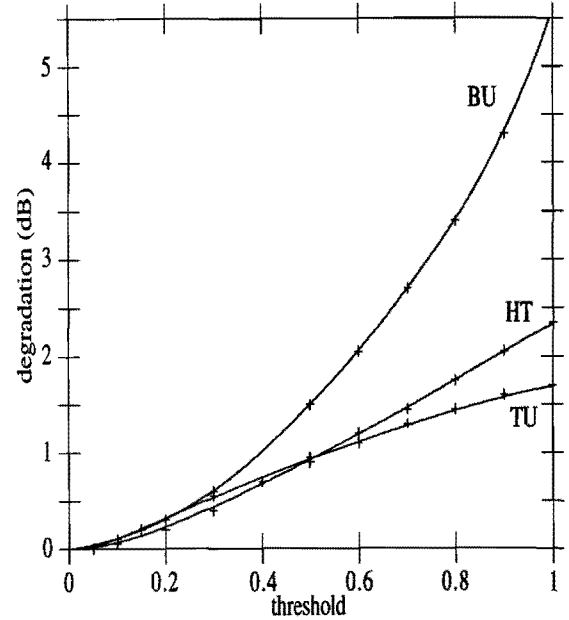
Fig. 18. degradation of the performances for conditional DFE (using C2) relatively to the performances at 9 dB for unconditional DFE. Considered channels : TU, BU and HT COST 207 models.
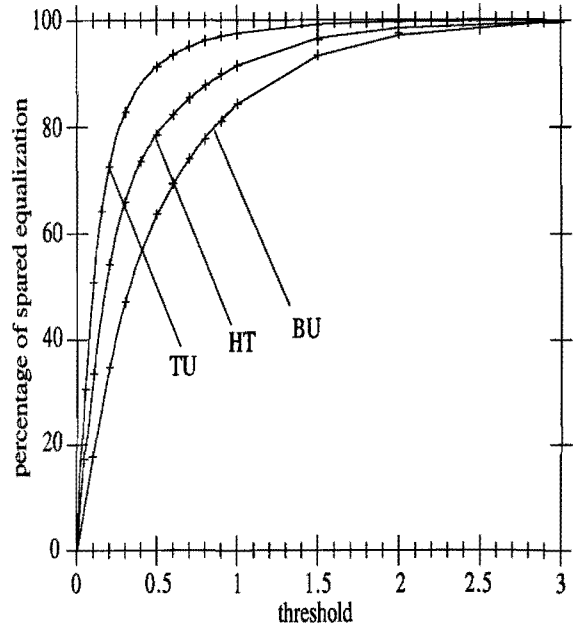
Fig. 19. relation between the chosen threshold and the percentage of spared equalization, for the considered COST 207 models, TU, BU and HT.

In order to use conditional equalization in practice , the threshold used in the decision criterion must be preliminary set to a fixed value. Figure 18

11

displays the degradation, at 9 dB, as a function of this threshold for the TU, BU and HT channels. Figure 19 displays the percentage of spared equalization as a function of the threshold for the same channels. This figures enable to make a good compromise between the gain (spared equalization) and the loss (degradation) inherent to conditional equalization.

A majoring of the degradation for all channels can be ensured by the determination of the threshold in the worst case (Bad Urban) using figure 18. Thus, setting the threshold to 0.27 leads to a degradation inferior to 0.5 dB in all the considered cases. The gain of spared equalization is of 43 % for Bad Urban, 62 % for Hilly Terrain and 79 % for Typical Urban channels (see figure 19). The power consumption due to equalization is reduced in the same proportions.

VI. CONCLUSION AND FUTURE WORK

We propose a new method for increasing the autonomy of the portable handset in TDMA mobile radio systems. It minimizes the power consumption due to equalization by resorting to equalization only when it is mostly needed and abstaining from equalizing when the distortions due to the propagation channel are weak. By adapting the resources of the receiver to its real needs, a great percentage of useless equalization can be spared without a noticeable degradation of the performances.

At each time-slot, the receiver autonomously decides if equalization is needed or not according to a relevant criterion. The proposed method does not need any additive signalization on the network. Two criteria have been theoretically analysed in the case of 2PSK transmission through a 2-path Rayleigh fading channel. Extensive simulations for 2PSK transmission over the recommended COST 207 channels show the efficiency of conditional equalization for mobile channels.

The most efficient criterion enables to have, at 9 dB, a degradation inferior to 0.5 dB, while providing a respective gain of 43 % of spared equalization for Bad Urban channel, 62 % for Hilly Terrain channel and 79 % for Typical Urban channel. This results appeal conditional equalization to be a promising method to improve the autonomy of portable handset in mobile systems.

Encouraged by these results, our work goes on in the analysis of the sensitivity of the method to the accuracy of the channel estimate and the search of other criteria taking into account the estimation of the power of noise and interferences. An evaluation of the performances of our method for other modulations such as 4PSK, MSK and GMSK is also in the scope of our future research work.

REFERENCES
[1]   J. G. Proakis, *Digital communications*, Mc Graw Hill, 1989.

[2]   A. J. Levy, "Fast error rate evaluation in the presence of intersymbol interferences", *IEEE tr. on comm.*, vol. COM-33, no. 5, pp. 479-481, May 1985.

[3]   G. D'Aria, V. Zingarelli, "Fast adaptive equalizers for narrow-band TDMA mobile radio", *IEEE tr. on veh. tech.*, vol. 40, pp. 392-404, May 1991.

[4]   J. C. Bic, D. Duponteil, J. C. Imbeaux, *Éléments de comunications numériques*, Dunod, 1986.

[5]   A. Wautier, "Influence de l'estimation du canal sur les performances d'un égaliseur dans le cadre des radiocommunications avec les mobiles", *P h D dissertation, Université de Paris XI, France*, Dec. 1992.

[6]   M. Failli (chairman), "Digital land mobile radio communications", COST 207 final report, *CIC Inf. Technol. and Sciences, Brussels*, pp.135-166,1989.

[7]   J. C-I. Chuang, "The effects of time delay spread on portable radio communications channels with digital modulation", *IEEE jr. on Sel. Areas in Comm.*, vol. SAC-5, no. 5, pp. 879-889, June 1987.

Autonomous Slot Assignment Schemes for PRMA++ Third Generation TDMA Systems

Thorsten Benkner* and Klaus David**
*University of Siegen, Dep. for Electrical Engineering&Informatics, D-57068 Siegen,Germany
**DeTeMobil GmbH, Postfach 8865, D-48047 Münster, Germany
e-mail: benkner@nue.et-inf.uni-siegen.de

Abstract—Some traffic and interference adaptive dynamic slot assignment strategies for the potential medium access control mechanism PRMA++ of 3rd generation TDMA proposals like the RACE ATDMA system are investigated in microcellular environments. Frequency planning in microcellular systems is extremly difficult or can not be afforded at all. Furthermore the interference situation can change very rapidly, which is to a large extend caused by the relatively high user mobility compared with macro-cellular structures. So only full adaptive slot assignment schemes that are adaptive to both traffic and interference can be used if capacity degradations shall be avoided. Also a decentralized mode of operation is useful to reduce the amount of control and administration in the face of a high number of small cell sites. In this paper some dynamic slot assignment schemes that fulfill the above requirements and which are based on interference measurements and/or adaptive learning of allocation patterns are compared. Furthermore the novell scheme CS+ is presented which allows to reduce the level of interference and to further increase the capacity of PRMA++ cellular mobile radio systems.

I. INTRODUCTION

Third generation cellular mobile radio systems will offer a large variety of different teleservices to a rapidly increasing number of subscribers. On a medium-term scale about 100 million subscribers are expected only in Europe. To cope with this huge number a lot of measures will have to come together like new medium access control mechanisms, new cellular structures and intelligent allocation of network resources. However the most efficient method to increase the network capacity is the introduction of microcells or even picocells with cell radii of about 1 km down to 100 m and below.

With the introduction of such small microcells a lot of severe problems occur. At first frequency planning is very difficult or can not be afforded at all because the interference situation can change very rapidly during the operation time. Especially in picocells, base station antennas will probably be mounted in street lamp elevations. So the electromagnetic wave propagation will be influenced very much by medium-term alterations in the neighbourhood of the base station locations. But the worse and much faster interference variations come from the user mobility itself. This, because of the small cell radii, high relative mobility causes temporarily very high spatial traffic inhomogeneties that fixed channel assignment schemes (FCA) are not able to cope with, even if a frequency preplanning could be afforded. That is the reason why the introduction of microcells implicitly requires dynamic channel assignment schemes (DCA) which are able to adapt to such variations. A lot of such schemes have been proposed in the technical literature but only very few are suitable for microcellular operation because many of them are not able to adapt to varying interference situations and/or they need a high degree of centralized control mechanisms which is a disadvantage in microcellular systems with a huge number of base stations. So for microcells only autonomous and interference adaptive schemes, so called full adaptive schemes [1] are really useful.

A flexible medium access scheme is very desirable for third generation cellular networks because of the high flexibility and variety of new teleservices. One class of protocols that are suited for such tasks are packet access schemes which can easily offer variable bitrates and carry also bursty traffic. Within

the RACE ADTMA project [2] the PRMA++ medium access protocol has been proposed [3] which has been further developed from PRMA [4] to fulfill the special requirements of ATDMA.

In this paper some channel/slot assignment schemes are fitted for PRMA++ and comparatively investigated in a planar microcell environment. In section II a brief description of the PRMA++ protocol is given. Section III presents the analyzed scenario. The investigated DCA schemes are described and their behaviour in the scenario is presented in section IV. We conclude in section V.

II. BRIEF DESCRIPTION OF PRMA++

PRMA++ is a packet reservation medium access mechanism proposed for operation within the RACE ATDMA project for third generation wireless cellular networks [2]. For microcell operation the TDMA frames of 5 ms length will have 72 time slots (fig. 1) of 66 payload bits and 59 bits for burst overhead including guard bits, training sequence, and inband signalling (fig. 2). The carrier bit rate is 1.8 Mbit/s.

| R | I | I | I | .... | R | I | I | fak |
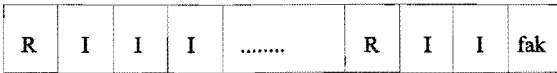
Fig. 1. PRMA++ uplink frame structure

| H 4 | DATA 33 | Si. 5 | Training Sequence 29 | Si. 5 | DATA 33 | T 4 | Guard 12 |
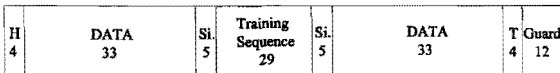
Fig. 2. PRMA++ slot structure [3]

A PRMA++ uplink frame consists of reservation slots (R), informaton slots (I) and at least one fast acknowledgement slot (fak). The downlink frame structure is similar to the uplink case, but here we have instead of R slots corresponding delayed acknowledgement slots (A slots) and instead of the fak-slot it has a fast paging slot (FP).

A mobile that wants to gain access to one or more traffic slots (I slots) transmits a reservation packet in form of an Air-Interface Channel Identifier (ACI) and the number of desired I slots in the next R slot.

In case of a successful reservation the base station transmits the mobile's address on the corresponding downlink A slot and also the numbers of the assigned slots if there are enough free I slots available. In case of no free I slots the mobile is queued and will get an assignment if there are free slots again. Mobiles reserve on a talkspurt basis, this means that only when the voice coder delivers speech packets transmission capacity is reserved (discontinous transmission, DTX). In case of voice transmissions the speech packets will not be queued, they will be dropped if the access delay becomes larger than the maximum packet holding time (10 ms). At the end of a talkspurt the I slot is released by transmitting a corresponding release message to the base station. Fast paging slots are used to inform the mobile where to find paging information. The corresponding fast acknowledgement slot on the uplink confirms a paging message. Further details of PRMA++ can be found in [3].

III. ANALYZED SCENARIO

The interaction of PRMA++ and the, to this access scheme suitably adapted channel assignment schemes FCA, DINF (see IV.), channel segregation (CSEG) [5] and a modified segregation approach (CS+, see IV.B), are investigated by computer simulation in a planar microcellular layout of 28 hexagonal cells [7] arranged as a "cluster of clusters". The cell radius is 1 km. For the FCA case the cluster size is 4 and every cell of the central cluster has exactly 6 potential cochannel interference cells. Results were only taken from the central cluster cells, so border effects can be neglected. The whole system has four carriers available. The number of I slots was restricted to 23 to reduce simulation time. It was found out that one R slot is sufficient to supply enough reservation capacity for this number of traffic slots because in the interesting load range the R access time is always below the PRMA++ framelength of 5 ms (see also [3]). The acknowledgement delay is 3 time slots. These investigations are restricted to the uplink case because this is the most critical one as only here a contention process occurs.

A mobility model was included because user mobility has a very high influence on the performance of the investigated dynamic channel assign-

ment schemes. Mobiles choose a random speed in the range of 5 to 60 km/h and a random start direction. Speed and direction can be randomly changed after all two seconds in the range of +/- 20km/h and +/- 90 degrees. User mobility implicitly implies intercell handover, so an optimized handover at the cell border was included too. An intercell handover call is treated like a new call in the destination cell.

The included channel model considers a pathloss slope γ equal to 4 and lognormal distributed shadow fading $s(\sigma)$ with a standard deviation of $\sigma = 4$ dB. So the carrier level C is equal to

$$C = Ar^{-\gamma} s(\sigma) \qquad (1)$$

where A is a constant depending on transmission power and antenna gain. Fastfading was not explicitly modelled because it is assumed that micro diversity, coordinated slow frequency hopping (SFH) and interleaving (depth 4) is used. The minimum required cochannel interference ratio CIR_{min} for a sufficent speech quality was assumed to be 8 dB.

The traffic load model consists of poisson distributed call arrivals, negative exponential distributed call durations of 120 s mean and also includes slow DTX with negative exponential talkspurts of mean activity periods of 1.41 s and mean silence periods of 1.74 s. Microcells are characterized by a very inhomogeneous load distribution. Here this is accounted for by the relatively high grade of user mobility for this cell radius and by a hotspot cell per cluster in which an additional traffic load is increased from 0 % to 100 % in steps of 20 % over a base load of 35 Erlangs. An intracell handover was not included in the simulation environment because of the relatively short talkspurt duration of 1.41 s. Simulation results have shown that the inclusion of intracell handover during talkspurts does not improve very much the performance of the system. Anyway the next talkspurt will need a new reservation and gets a new slot assignment.

IV. DCA SCHEMES FOR PRMA++ OPERATION

This section describes the behaviour of the investigated fixed and dynamic slot assignment

strategies. At first we will resort to a (on the average) homogeneous load situation wich will be influenced by a rather high grade of user mobility (see sec. III). The FCA and DINF strategies will be investigated in this environment at first.

A. Homogeneous Load Situation

In this paper three full adaptive DCA schemes are investigated in a PRMA++ environment. The first one (DINF) assigns slots according to carrier to interference (CIR) measurements. In this case the base stations monitor the interference level on all available I slots. An I slot is assigned if the calculated CIR from the measured receive power level during the R access period and the I slot interference level is above a defined setup treshold. The difference between the setup treshold and CIR_{min} is called setup hysteresis CIR_s. It has a very large influence on the performance of the system as will be shown later. The operation principle is shown in fig. 3 in form of a flowchart.
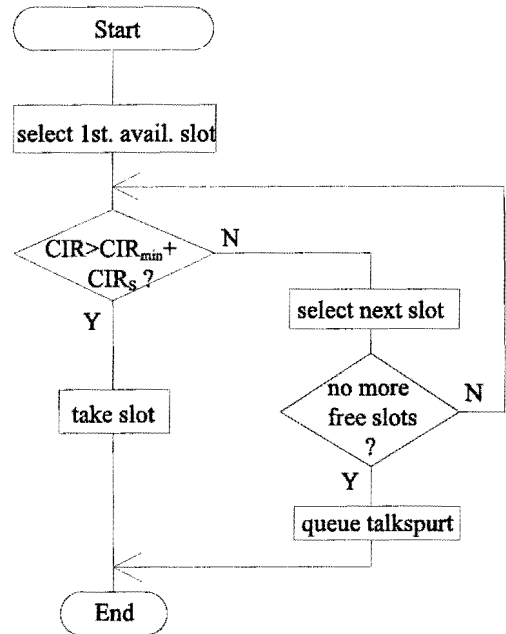
Fig. 3. DINF flowchart

At first we consider the R slot access time which is mainly attributed to the contention process for the FCA and DINF schemes in a homogeneous load distribution between 30 and 55 Erlangs per cell. In the case of the DINF assignment strategy three different setup hysteresis levels (6, 12 and 14 dB) are investigated. Fig. 4 shows that the average R slot

access time for the FCA case is in the order of the framelength (5 ms) whereas the DINF schemes can provide shorter access times due to the higher number of available carriers and thus R slots per cell. R slot access delays are always between 3 and 4 ms for all investigated DINF schemes.

The total I slot access time from the beginning of a talkspurt to the successful assignment of the traffic slot is even more interesting. If we impose a limit of 20 ms as acceptable I slot access delay it is possible to accommodate a load of about 37 Erl. with the static FCA strategy (s. fig. 5). With the DINF strategy and a setup hysteresis of 14 dB a cell can carry a load of about 40 Erl. in a homogeneous load scenario. If we decrease the setup hysteresis the I access delay will become very low. But on the other side, as can be seen from fig. 6, the probability that a talkspurt will be interfered takes on unacceptable values.

If we claim that the interference probability of the dynamic slot assignment scheme shall not be worse than the results of the corresponding FCA scheme (about 5 %) we will have to introduce a rather high hysteresis level of about 14 dB. With such a high setup hysteresis only very small capacity improvements of about 5 % can be reached. But on the other side we still have the advantage that no frequency preplanning is required.

Fig. 5. Mean I slot access time
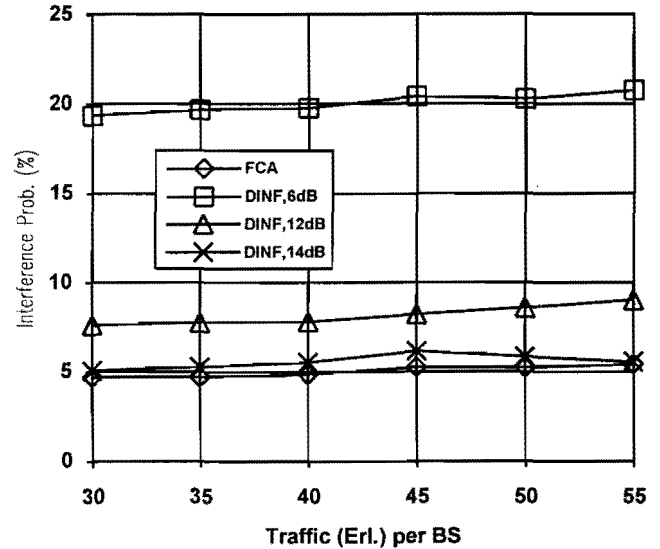
Fig. 6. Probability that a talkspurt will be interfered.

B. Inhomogenous Load Distribution

As already mentioned in section III. microcells are characterized by spatially and temporarily very inhomogeneous traffic load distributions. We will account for this by introducing a hotspot cell per cluster in which the load is increased from 35 Erl. to 70 Erl. in steps of 20 %. In this environment also the channel segregation strategie [5] and the newly introduced, improved segregation like strategy CS+ will be investigated.

The channel segregation strategy that is applied here is slightly modified for PRMA++ operation in the

Fig. 4. Mean R slot access delay attributed to the contention process.

sense that every base station keeps a list of slot usage priorities for every carrier which will be updated on a talkspurt basis. The free slot with the highest priority will be used next. The priority of the used slot will be modified after each talkspurt according to the rule:

$$P_{new} = \frac{P_{old} \cdot N + 1}{N + 1} \qquad (2)$$

if the talkspurt was not interfered and

$$P_{new} = \frac{P_{old} \cdot N}{N + 1} \qquad (3)$$

if it was interfered. N is the number of slot usage <u>attempts</u> (not necessarily slot usages !) and will be incremented by one after each attempt. A slot will be considered free if the measured interference level on it is low enough that a sufficiently low interference probability can be expected. Fig. 7 shows the working principle of CSEG in form of a flowchart.

The original channel segregation strategy of [5] only adapts to medium-term interference/traffic variations. Channel reusability will not be significantly improved as CSEG is a kind of 'adaptive FCA' [1]. Improvements of CSEG can be obtained if not the establishment of a quasi static reuse pattern will be the premier target of the algorithm but a higher flexibility. In the following an improved version of CSEG is introduced which establishes a 'weaker' reuse pattern and uses priority dependent C/I values to decide if a slot is free or not.

For this purpose CS+ firstly omits the priority decrease function when the considered slot with the currently highest available priority is interfered (see dashed box in fig. 7). This results in a somewhat 'weaker' reuse pattern. CSEG uses always the same setup treshold even if the observered slot has a very high priority. But the probability that a high priority slot will be interfered is usually lower than that of a low priority slot because the high priority slots form a very stable reuse pattern. These slots are used much more frequent than low priority slots and can be used to cover the base traffic load of the system.

In contrast to the CSEG scheme which shows nearly no dynamic behaviour when the reuse pattern is once established the CS+ strategy allows also the use of slots with low priorities when they supply a

CIR which is substantially higher than CIR_{min}. By use of such enhanced safety margins we have the possibility to use additional slots when no more high priority slots are available. Therefore a priority dependent CIR treshold is introduced:

$$CIR_{tresh} = CIR_{min} + CIR_S - log(P) \qquad (4)$$

P is the current slot priority and CIR_S is a constant minimum hysteresis.
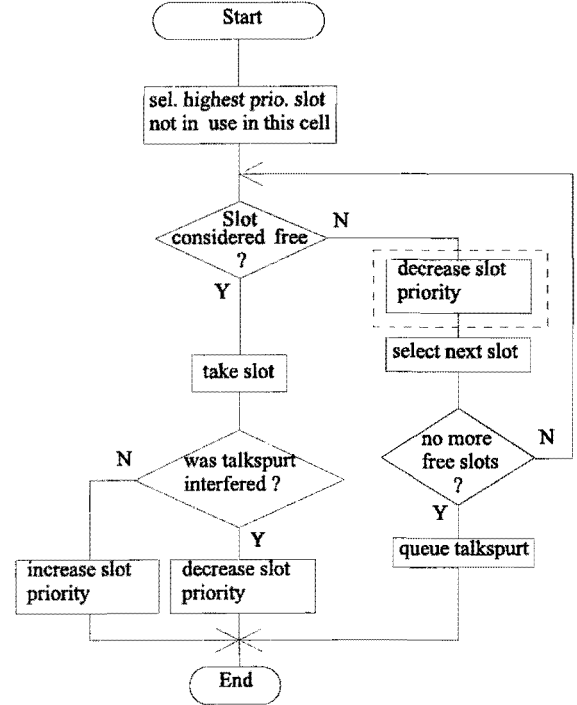
Fig. 7. CSEG flowchart

This dynamic component leads to enhanced performance especially in heterogeneous and time variant load distributions as the following results show. At first the I slot access delay as a function of the load increase in the hotspot cell is shown (fig. 8). As expected the mean I slot access delay is drastically increased by higher hotspot loads. Imposing a 20 ms access delay limit we are able to carry about 15 % load increase with the FCA scheme. The CSEG strategy which can adapt much better to traffic inhomogenities reaches about 55 % load increase. Even better performs the DINF scheme with about 85 % load increase. But the best result can be obtained with the new CS+ strategy (CIR_S = 6 dB) which is always below the 20 ms margin.

We have to look also at the interference probability to get a more detailed view into the system behaviour. Fig. 9 shows the probability that a talkspurt will be interfered during transmission.
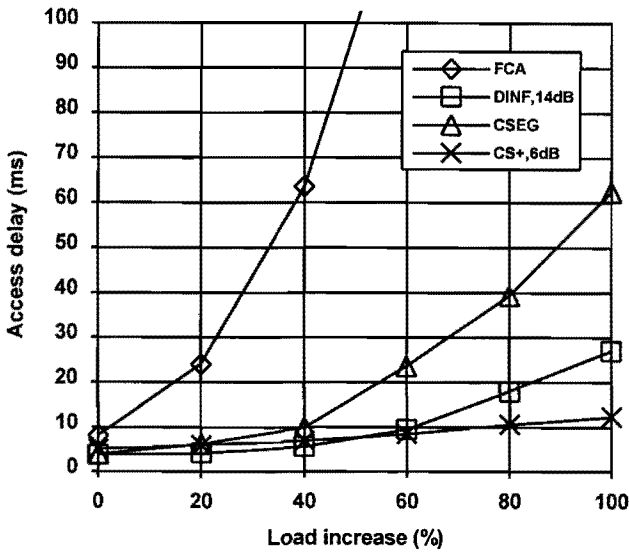
Fig. 8. Mean I slot access delay in the hotspot cell (base load = 35 Erl.)
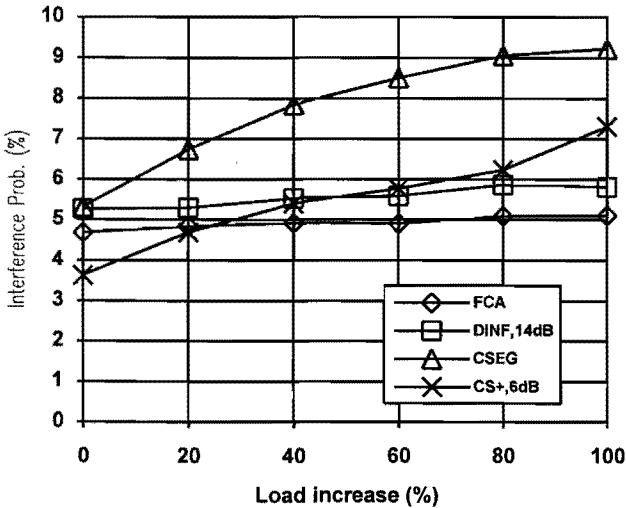
Fig. 9. Average interference probability in the central cluster.

In the most interesting range between 20 and 80 % load increase the interference probabilities of DINF and CS+ are comparable. But the I slot access delay of the CS+ scheme is lower than that of the DINF scheme especially when the hotspot load increase becomes higher.

We see that the constant setup threshold of the CSEG scheme is still too low for a good quality transmission. CSEG suffers mainly from the relatively high user mobility. But also with this too low threshold the I access delay of the CSEG scheme is higher than that of the DINF or the CS+ scheme. Increasing the setup threshold will further worsen the I access behaviour. The CS+ strategy offers about 2 % less interference probability and a much better I access performance than the CSEG scheme.

V. CONCLUSIONS

Four channel assignment strategies for the potential medium access control mechanism PRMA++ of 3rd generation advanced TDMA systems have been investigated. It was found that the statistical multiplex gain of discontinous transmission can be exploited to a large extend with PRMA++. The fixed carrier assignment suffers from severe capacity degradations especially in spatially and temporarily heterogeneous traffic distributions which are typical for microcellular cell structures. The channel segregation approach can improve this situation as long as the temporal load distribution does not change too rapidly. As channel segregation is in principle an adaptive FCA scheme we can not get a significant DCA capacity gain. The investigated interference adaptive slot assignment scheme DINF is able to follow such strong traffic variations but requires very high setup treshold values if the interference level shall not become too high. This can limit the potential DCA gain very much. A novel approach is the CS+ scheme which is able to reduce the interference level and can also offer a DCA capacity gain. All the investigated adaptive schemes are fully decentralized and need no frequency preplanning which is a major requirement for microcell operation.

18

REFERENCES

[1] Beck R., Panzer H.: "Strategies for handover and dynamic channel allocation in microcellular mobile radio systems", 39th IEEE Vehicular Technology Conference VTC'89, San Francisco, USA, 1989

[2] Urie A.: "Advanced TDMA Mobile Access", ICUPC'93, Ottawa, Oct. 13-15, 1993.

[3] Dunlop J., Cosimini P., Robertson D. : "Optimization of Packet Access Mechanisms for Advanced Time Division Multiple Access", 7th IEE Conf. on Mobile Personal Communications, Brighton, UK, Dec. 13-15, 1993.

[4] Goodman D.J., Wei S.X. : "Efficiency of Packet Reservation Multiple Access", IEEE Transactions on Vehicular Technology, Vol. 40, No.1, Feb. 1991.

[5] Furuya Y., Akaiwa Y.: "Channel Segregation, A Distributed Adaptive Channel Allocation Scheme for Mobile Communications Systems", DMR II, Stockholm, 1987.

[6] Frullone M., Riva G., Grazioso P., Carciofi C.: "Self-Adaptive channel allocation strategies in Cellular Environments with PRMA", 44th IEEE Vehicular Technology Conference VTC'94, Stockholm, June 1994

[7] MacDonald, V.H. : "The Cellular Concept", BSTJ, Vol.58, No.1, Jan 1978.

Request-TDMA: A Multiple-Access Protocol for Wireless Multimedia Networks

George R.J. Linnenbank, Paul J.M. Havinga, Sape J. Mullender, Gerard J.M. Smit

University of Twente
Department of Computer Science
P.O.box 217, 7500 AE Enschede, the Netherlands
linnenba@cs.utwente.nl

Abstract

In this paper we propose a multiple-access protocol that is well suited for a limited-bandwidth (1-10 Mbps) single-cell wireless network for indoor multimedia communication. We do not consider interference between cells. The proposed protocol exploits the available bandwidth fully and supports real-time and non-real-time communication. The throughput per mobile[1] station is higher compared to other multiple access protocols, it offers low latency for both real-time and non-real-time communication and the reserved but unused bandwidth is reused for non-real-time communication. Furthermore, the throughput and latency remain stable under high loads.

1 Introduction

In the current wireless communication networks, bandwidth is a scarce resource. Therefore it has to be exploited as efficiently as possible. Considerations of cost and saving power dictate that mobile wireless stations must have as few electronic components as possible. Therefore, mechanisms to exploit the bandwidth efficiently need to be of low complexity.

Many multiple-access protocols have been suggested to access the available network bandwidth. Most of these multiple-access mechanisms rely on a contention approach. Contention makes these systems unstable under high loads resulting in inefficient bandwidth exploitation and high latency. Therefore, these multiple-access protocols are not well suited for multimedia communication. Examples of these protocols are Aloha [1], ISMA [2] [3], R-ISMA [4] [5] and PRMA [6] [7].

Other systems divide the available bandwidth in channels that are separated in time, frequency or code (TDMA [8], FDMA [8], CDMA [9]). A drawback of the FDMA and CDMA systems is that, due to parallel channel separation, a mobile station usually can only use one channel at a time. The maximum bandwidth for

a mobile station is therefore merely the bandwidth of a single channel. This can only be helped by adding hardware (more transceivers).

Since in TDMA channels are sequential, multiple channels can be used to obtain more bandwidth. However, once a channel has been allocated to a mobile station the bandwidth is lost when that station has an inactive period (in CDMA this will lead to a lower bit-error rate for the other mobile stations).

In our system we focus on a single, small cell (3-5 m radius) in which only a few mobile stations need to be served. In this paper, we ignore interference between cells; we will address this issue in a forthcoming paper. We use near-field radio [10] as the communication medium. The available network bandwidth is limited (1-10 Mbps) and must support a mixture of real-time and non-real-time communication for multimedia applications. When a single mobile station is active, its obtainable bandwidth must approach the available network bandwidth as close as possible. This in not possible to achieve with FDMA and CDMA. When more mobile stations are active, the aggregate bandwidth of these mobile stations must also approach the available network bandwidth as close as possible. This is not achievable with random-access protocols as contention results in bandwidth loss. Finally, the protocol must support dynamic bandwidth allocation and communication priorities. This is not supported by any of the protocols named above.

We propose a multiple-access protocol that will satisfy these requirements in a better way than previously proposed multiple-access protocols. In Section 2 we describe the near-field wireless network system for which we intend to use the multiple-access protocol. Then we describe the multiple-access protocol in detail in Section 3. In Section 4 we present and discuss the results of an extensive simulation of the protocol. We compare the proposed protocol with several similar protocols in Section 5 and explain why our protocol is better. The future work and a summary of the conclusions

---

1. In this paper we consider a mobile station to be a wireless station that is not a base station. Since we only consider single cell environments, mobility itself is not an issue here.

are given in Section 6 and Section 7.

2 System description

2.1 System Architecture

Our goal is to design and implement a wireless network for an office environment with one base station per room, see Figure 1. Each base station is connected to a backbone ATM network. In each room a limited number of mobile multimedia stations can be present. The communication of a mobile station is temporarily suspended when it moves from one room to another and continues when a new connection has been made to a new base station. Through the new base station an obsolete connection to the previous base station can be closed.

2.2 Uplink communication

The wireless network consists of independent cells. Each cell is served by a base station. The down-link communication (base to mobile) is fully controlled by the base station. Down-link communication is recognized to be relatively straightforward to implement. More difficult to implement is a multiple-access protocol for the up-link communication (mobile to base). Since the knowledge of the activity of the mobile stations is distributed over the mobile stations, it is difficult to allocate bandwidth to the mobile stations efficiently. In this paper we focus on the up-link communication.

2.3 Multimedia application requirements

Multimedia applications have properties that cannot be supported by many of the known multiple-access protocols. Some applications require *real-time behaviour*. Random-access protocols [1]-[7] suffer from unbounded statistical latency under high loads which is intolerable in a multimedia environment where the network load can be close to the network bandwidth. The throughput of these random-access protocols drops rapidly when the aggregated load generated by the mobile

stations approached the network bandwidth. Therefore, random-access must be avoided.

In a multimedia environment, the information of one application can have a higher *priority* than the information of another application. For example, delaying a file transfer will not be hazardous (therefore a file transfer is of lower priority), but it is desirable that the transfer of an audio packet is not delayed (which therefore is of higher priority). Few wireless multiple-access protocols (like the one described in [12]) have successfully included priorities in the wireless multiple-access protocols. We must include priorities in the multiple-access protocol to support multimedia applications successfully.

Multimedia applications may also generate *variable amounts of data* in time. Reservation mechanisms such as CDMA [9], FDMA and TDMA [8], allow applications to reserve bandwidth. For bursty real-time applications, such as variable-bit-rate video, reservations must be made for the peak bandwidth so that a significant portion of the reserved bandwidth will not be used. Therefore, the utilization of the network bandwidth is not optimal. Since the available bandwidth in our wireless network is limited to 1-10 Mbps, we must be able to reuse the reserved but unused bandwidth for non-real-time applications.

Summarising, a protocol is needed that is stable regarding throughput and latency under high loads. This protocol must also support priorities (it must at least support real-time and non-real-time communication) and it must be able to dynamically allocate bandwidth.

2.4 Single-frequency near-field radio

A single mobile station in a cell must be able to obtain all the bandwidth for its communication. When a multiple-frequency system is used, multiple transmitters and receivers are needed to allow the mobile station to transmit and receive on all channels to obtain all of the cell's bandwidth. This makes the hardware complex (splitting
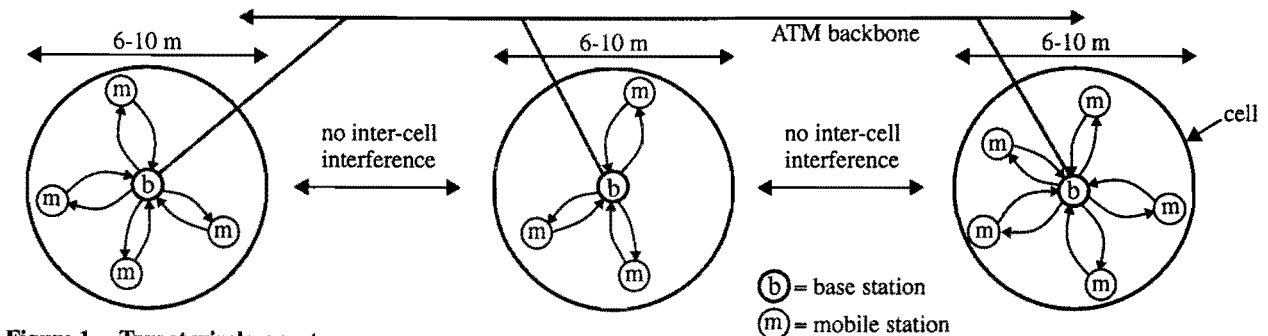
Figure 1. Target wireless system

21

and mixing data streams) and expensive. In order to keep the transceiver design simple and inexpensive. we use a single-frequency system.

In our system we want to have one base station per room without introducing inter-cell interference. As the power of a near-field radio signal drops an order of magnitude more rapidly than a far-field radio signal, it offers a good separation between cells [10]. Cells can be placed close together without introducing much inter-cell interference. In an office environment, each room can be provided with the full bandwidth of a base station, giving a high bandwidth density (Mbit/m²).

For our prototype network, we use a small near-field radio spectrum (7.9 - 8.9 MHz). As there is little noise in this frequency band, bandwidths up to 10 Mbit/s can be achieved. In our prototype system we will provide a bandwidth of 1 Mbps using binary-PSK.

2.5 Mobility

Since we intend to avoid overlapping cells, continuous mobile communication cannot be supported. The communication will be a form of connect and disconnect also called *nomadic* communication. In this paper, the functional requirements for connect/disconnect, handover, etc., is not an issue since these functions must be implemented at a higher level in the protocol stack.

3 Protocol description

3.1 Transmission units

We use ATM cell transmissions over the wireless link, partly because the network connecting the base stations is also an ATM network. Using ATM cells keeps the base station architecture simple since no conversion is needed between the wireless channel and the B-ISDN network.

Since each transmitter-receiver pair will have different distances, the propagation delays between them will also be different. Therefore, synchronisation bytes are used with each transmission to synchronise transmitter and receiver.

The prototype transceiver runs at 1 Mbps. The basic unit of transmission is a byte. This means that we can exchange very short messages between base and mobile stations. Since messages are preceded by 2 synchronisation bytes, it will take 16+8 μs to send a byte and 16+53*8 μs to send an ATM cell.

The error rates of the wireless link still have to be investigated. When the error rates are too high, an extra error detection code must be appended to the ATM cell in order to avoid unnecessary processing of corrupted packets.

3.2 Connections

Mobile stations can have multiple connections. In our protocol, a *connection* is a communication agreement between base and mobile station. For each connection a number of time slots can be reserved during connection set-up, giving the connection a guaranteed minimal amount of bandwidth. A connection that reserves bandwidth is called a *real-time connection* and a connection with no bandwidth guarantees is called a *non-real-time connection*. It depends on the activity of all other connections, including the other non-real-time connections, how much bandwidth can be obtained by a non-real-time connection. In Section 3.3.1 we describe how bandwidth can be requested. How the bandwidth is allocated will be discussed in Section 3.5.

The control and data messages of a multimedia application can be transmitted over separate connections. However, it is not necessary to do so. Control messages can be inserted in the data traffic since the ATM cell header will indicate that the cell contains control information instead of data. In this paper, we do not make a distinction between control messages and data messages. We will focus on using the bandwidth as efficiently as possible while keeping the latency limited.
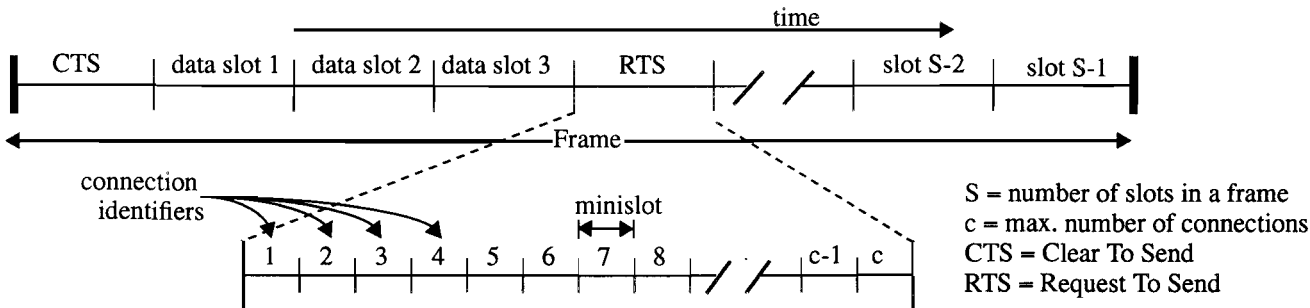
Figure 2. The R-TDMA frame structure

22

3.3 Up-link frame structure

In the R-TDMA protocol, time is divided into frames. Each *frame* contains S uplink slots for uplink communication. The frame structure is shown in Figure 2. Two slots in the frame have a special function. The first slot in the frame is the *Clear To Send* (CTS) slot. The base station uses this slot to inform the mobile stations which slots they may use in the next S uplink slots. The second special slot is the *Request To Send* (RTS) slot, where mobile stations can request slots for the next frame. As we will show later, the location of the RTS slot influences the latency and it determines the available scheduling time. Therefore, the location of the RTS slot in the frame is an important design parameter. The remaining $S-2$ uplink slots are data slots which can be used by the mobile stations for uplink communication.

Obviously, the frame is not complete. Downlink data slots are also needed to obtain full-duplex wireless communication. Although downlink slots are not difficult to add to the frame, it is not in the scope of this paper. We will address this subject in a forthcoming paper.

3.3.1 Request To Send slot

The Request To Send slot is divided into *minislots*. Each connection uses 1 minislot. During connection set-up, the base station will assign minislot i to connection i. A connection uses its minislot to inform the base station how many ATM cells are waiting in its message queue. The base station uses this information and Quality of Service (QoS) agreements, see Section 3.4, to compute a slot allocation. A minislot must be large enough to send one byte (the basic unit of transmission), including a synchronisation period and a gap for switching one transmitter off and another on.

A connection must send a request on a per frame basis. This means that it requests at most $S-2$ slots and new requests must be made in every next frame until all the messages have been transmitted. The base station will compute a slot allocation based on the most recent requests. Requested slots that were not allocated in the previous frame are not taken into account.

When the frame size is large enough and the number of connections can become large, then it is possible to extend the RTS period over multiple slots, according to the need for more minislots. However, since we design a wireless network with only few mobile stations in a cell, we will not address this option.

3.3.2 CTS slot

The Clear To Send slot is used to send a Clear To Send

message which is a normal ATM cell that is always transmitted in the first slot of the frame. It has two primary functions. First it informs every mobile station in reach that a base station is present and what its name is. This information is needed since the mobile stations must know that they are in reach of a base station in order to set up a connection.

The other function is to broadcast the slot allocation for the current frame to the active mobile stations that reside in the base station's cell. Slots that have not been allocated to a connection may be used for connection set-up. Usually free slots are available during low loads. Under high loads a slot must be kept free regularly by the base station to enable new mobile stations to set up a connection.

3.3.3 Data slots

The data slots are used to send packets containing application data. In our system we use ATM cells in order to simplify the translation to the fixed ATM network.

In B-ISDN networks the ATM header requires 5 bytes to detect header errors and to identify all types of communication from all connections on a single link. But, since in our network an ATM packet is only transmitted after a slot request followed by a slot allocation, the base station already knows the source connection of the packet. Hence, less header information is needed. A header of one byte is sufficient to identify 256 special events and/or message types. Identifier i $(0 \leq i \leq c-1)$ of a connection can be considered part of the virtual circuit identifier, giving an address space of $c*256$ different headers per base station or 256 per connection. Including a single CRC byte that is computed over the full message, the total cell overhead can be reduced from 10% (5/48) to 4% (2/48). The impact of this choice is discussed in Section 4.3.1.

3.4 Connection set-up

Connection set-up is implemented in a very simple way. The CTS message includes information about slots that have not been allocated to real-time or non-real-time traffic. These slots can be used for connection set-up.

When a mobile station wants to set up a connection it waits for the first CTS message that indicates which slots can be used for connection set-up. It will then try to send a connection request using one of the available slots and it will use exponential back-off when no response from the base station is received within a timeout period. When no response is received, either a collision has occurred or no bandwidth was available for

23

transmitting a response.

Connection set-up is the only contention phase in the protocol. It cannot be avoided since new mobile stations can enter the cell at any time.

During connection set-up a mobile station can negotiate over a range of Quality of Service parameters. These parameters influence the behaviour of the slot scheduler, giving more or less priority to requests based on these agreements. In our prototype implementation, only an agreement on the number of reserved slots can be made.

3.5 Slot scheduler

Base stations have a slot scheduler that allocates data slots in the frame to the connections. The scheduler allocates the slots based on the requested number of slots and the bandwidth guarantees of the connections.

Real-time connections will always get allocated a number of slots up to the number of slots that was reserved during connection set-up. Additional real-time data packets for which slots are requested will be treated as non-real-time packets.

The remaining slots are allocated round-robin to the non-real-time connections. Each non-real-time connection that requested slots will get one slot per turn. When few non-real-time connections are active, connections may get multiple turns per frame.

4 Protocol simulation

In this section we describe the simulation of the protocol and the results we obtained. We used a simulator that can simulate a MAC protocol for multiple mobile stations and multiple non-interfering base stations. Its basic transmission unit is a message. Signal effects such as multi-path, fading, noise are not included in the simulator, however collisions are simulated by making packet contents corrupt.

4.1 Frame implementation

In our simulations we used frames of 20 slots giving 18 data slots per frame. The first slot in the frames is the CTS slot and the 11th slot is used for the RTS slot. Packets are 53*8 bits large and a *gap* of 16 bits is used to space the messages. The available bandwidth for data communication can be expressed as

$$\frac{Slotsize}{Slotsize + Gap} \times \frac{S-2}{S} \qquad \text{Eq. 1}$$

In the simulation setup this gives us 87% of the link bandwidth for data communication. A base station can support up to 16 connections. Therefore, the RTS slot consists of 16 minislots. A minislot is 26 bits long and contains a single information byte with a start and a stop bit (10 bits). Mobile stations can inform the base station that they have up to 255 ATM cells to transmit. The remaining 2 byte times are used for synchronisation of base and mobile station, and leaves time for one mobile station to switch its transmitter off while another mobile station switches its transmitter on. The frame structure used for the simulation is shown in Figure 3.

4.2 Configurations

We simulated the protocol in various configurations. In each configuration there was a single base station. The number of mobile stations (N) and the generated traffic p (exponentially distributed), were varied. Each mobile station has a real-time connection and a non-real-time connection, each generating a load of p.

During connection set-up, real-time connections will request the base station to reserve $\lceil S*p \rceil$ slots, which is
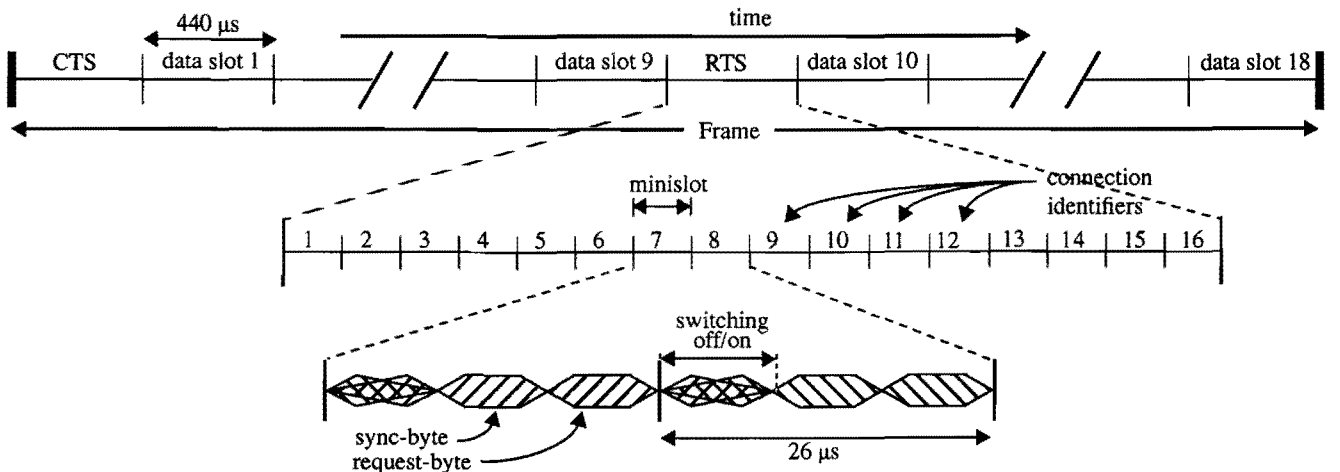
Figure 3.   Simulated frame structure
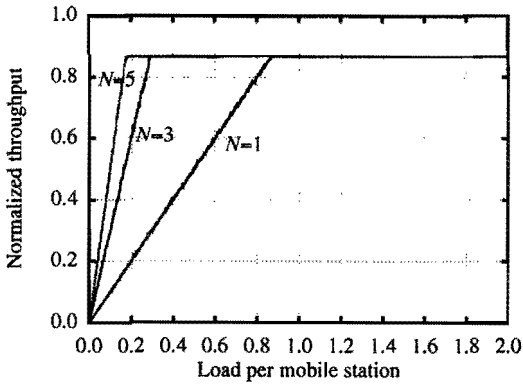
Figure 4. Simulated throughput as a function of the load of 1, 3 and 5 mobile stations

Figure 5. Simulated throughput per priority as a function of connection load with 1 and 4 mobile stations

Figure 6. Simulated real-time latency as a function of load for 1, 2, 3 & 5 real-time connections
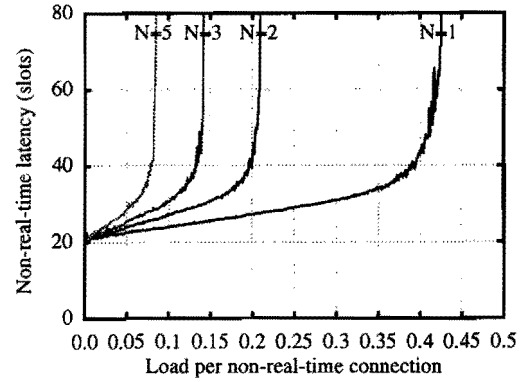
Figure 7. Simulated non-real-time latency as a function of load for 1, 2, 3 & 5 non-real-time connections

the average number of slots that is generated per frame by a real-time connection. The connection is refused when the base station cannot reserve this number of slots. No attempt for reserving fewer slots is made since a shortage of bandwidth will results in increasing latency which is intolerable for real-time communication.

4.3 Results

The results of the simulations are given in Figure 4 through Figure 7. We will briefly discuss these figures.

4.3.1 Throughput

From Figure 4 we conclude that the bandwidth is used with a maximum efficiency. The throughput is identical to the load as long as the network is not saturated. When the load is higher than the available network bandwidth, all bandwidth is used.

Figure 5 shows that non-real-time data throughput will decrease when the aggregated load exceeds the network capacity. Due to bandwidth reservation real-time connections can send data at the cost of non-real-time data. The figure also shows that the real-time throughput remains stable after network saturation.

4.3.2 Latency

From Figure 6 and Figure 7, we can conclude that the average latency is very limited. For real-time traffic the average latency exceeds two frames (40 slots) only when the aggregated real-time load approaches the network capacity very closely. The non-real-time latency behaves roughly the same, exceeding two frames when the aggregated load of real-time and non-real-time traffic approached the network capacity.

Note that the simulation results of the latency for 5 real-time connections stops at a load of roughly 0.175 per connection. When the load is larger than 0.175 per connection, each real-time connection wants to reserve 4 or more slots (20 or more in total) while only 18 data slots are available in a frame. The base station will refuse the fifth real-time connection. It is therefore not possible to obtain results for 5 real-time connections with a higher load than 0.175 per connection.

Also note that the real-time latency results contain peaks at regular intervals. This can be explained as follows: When a real-time connection generates exactly n (n=1,2,...) packets per frame it will reserve n slots per frame. These reserved slots do not give capacity to handle a small burst which can easily occur with exponen-

tially distributed data generation. Therefore the latency will be relatively high. However, when the connection generates only a slight amount of data more per frame, it will reserve a full additional slot. This suddenly gives some overcapacity resulting in a sudden drop of the latency. In a forthcoming paper we will give a mathematical analysis of this phenomenon.

5 Related work

There are several protocols that resemble our protocol. We will point out the improvements we made compared to those protocols.

5.1 The bit-map protocol

In the bit-map protocol [8] a station that wants to transmit data must indicate this in a special slot. Each station has a single slot where it can indicate its intention to transmit. After all stations made a request during the request period, the stations that indicated a desire to transmit will start transmission one after the other. Since the order in which the stations transmit their packets is fixed, the average latency is not equal for each station. Hence, the protocol is not fair concerning delay.

The protocol assumes a fixed number of stations. Clearly this is not the case in a dynamic wireless environment where stations can come and go. However, this can be solved by using a cycle where the frame size is large enough to allow new stations to set up a connection, just as we proposed in our protocol.

Another restriction is that a station can only request one slot per cycle. It is not possible for a single station to obtain all of the available bandwidth (i.e. obtaining all slots in a cycle). When few stations are active the overhead of the request period becomes a larger portion of consumed bandwidth.

Finally, the most important assumption is that all stations can receive all requests since every station needs to know which other stations will be transmitting first. This cannot be realised in a wireless network due to hidden stations. This can be solved by having the base station collecting all request and broadcasting the requests back to the mobile stations. We used a delayed broadcast to allow the base station to allocate the slots more efficiently.

5.2 A wireless MAC with client-server traffic and capture

The multiple-access protocol proposed in [11] also uses minislots to optimize bandwidth requests. However, the

requests are performed based on an Aloha approach which means that the load must be limited. In order to allocate enough slots under high loads, the request period needs to be increased since many minislots are lost because of collisions. This results in a high request overhead. Since the minislot access is random, no throughput and latency guarantees can be given and therefore high-speed real-time communication cannot be supported.

5.3 ATM-based transport architecture for multiservices wireless PCN

In [12] a protocol is proposed that is very similar to our protocol. However, it uses fixed allocation of slots together with random access of request slots. The fixed allocation means that potentially throughput is waisted since it is not likely that a fixed allocation CBR connection always exactly uses the same amount of bandwidth. The most commonly known CBR communication is voice transmission. For this type of communication silence period detection is used to prevent unnecessary transmissions saving both battery power and bandwidth when this bandwidth can be reused.

For the remaining unallocated slots an Aloha access mechanism is used comparable to [11]. This means that under high loads some slots will remain empty since collision will occur in the request period. The request period cannot be increased to allow more successful Aloha accesses to succeed since that also increases the frame size which cannot be tolerated concerning the CBR connections. Therefore, bandwidth cannot be optimally used.

Finally, the only guaranteed allocated slots are those that were reserved for CBR connections. VBR connections, e.g. video, must use random access to obtain slots. That means that under high loads, video can not be supported.

6 Future work

Currently we are designing a hardware implementation of the R-TDMA protocol. This implementation will be tested at our university.

Our next step will be to apply the R-TDMA protocol to a multi-cell environment where inter-cell interference can occur. The current implementation will not function in a multi-cell environment with inter-cell interference. Channel allocation must be added to the protocol to minimize inter-cell interference.

We will test a variety of slot scheduling algorithms that

can support a variety of QoS properties. Some examples were given earlier in this paper.

7 Conclusions

In R-TDMA the bandwidth is optimally used. The throughput remains high under high loads and does not result in threshing.

Real-time connections have a guaranteed throughput with limited latency. Non-real-time connections share the bandwidth in a best effort way. All the bandwidth that is not used by real-time connections can be used by the non-real-time connections. When the network bandwidth is not saturated, non-real-time latency remains low.

The frame size can be increased in order to increase throughput. Decreasing the frame size leads to lower latency for real-time traffic. A significant part of the fixed initial latency is limited by the time required to schedule the slots. Therefore, the latency can be smaller when a faster scheduling algorithm is used.

The use of a slot scheduler makes our multiple access protocol very flexible. Not only does it deal with real-time and non-real-time connections, the scheduler can also take more priorities into account. In our system we used a round robin scheduler with two priorities, but many other well known scheduling mechanism can be used such as earliest deadline first, first-come first-serve, etc. However, the more complex the scheduling algorithm, the more computation time is needed and the longer the RTS - CTS time will become, thus increasing the average latency significantly.

References

[1] Abramson, N., Development of the ALOHANET, IEEE Transactions on Information Theory, vol. IT-31, pp. 119-123, March 1995.

[2] Mukumoto, K., Fukuda, A., Idle-Signal Multiple-Access (ISMA) Scheme for Terrestrial Packet Radio Networks, Electronics and Communications in Japan, Vol. 64-B, No. 10, pp. 66 - 74, 1981.

[3] Murase, A., Imamura, K., Idle-Signal Casting Multiple Access with Collision Detection (ICMA-CD) for Land Mobile Radio, IEEE Transactions on Vehicular Technology, Vol. VT-36, No. 2, pp. 45 - 50, May 1987.

[4] Murase, A., Imamura, K., Idle-Signal Casting Multiple Access with Data Slot Reservation (ICMA-DR) for Packet Radio Communications, IEEE Transactions on Vehicular Technology, Vol. 38, No. 2, pp. 45 - 50, May 1989.

[5] Wu, G., Mukumoto, K., Fukuda, A., Performance Evaluation of Reserved Idle Signal Multiple-Access Scheme for Wireless Communication Networks, IEEE Transactions on Vehicular Technology, Vol. 43, No. 3, pp. 653 - 658, Aug. 1994.

[6] Goodman, D.J., Wei, S.X., Efficiency of Packet Reservation Multiple Access, IEEE Transactions on Vehicular Technology, Vol. 49, No. 1, pp. 170 - 176, Feb. 1991.

[7] Wen, J.-H., Wang, J.-W., Throughput Analysis of Packet Reservation Multiple Access Protocol for Wireless Communications, Proceedings PIMRC'94/WCN94, pp. 1242 - 1246, Sept. 1994.

[8] Tanenbaum, A.S., Computer Networks, Second Edition, Prentice Hall, Englewood Cliffs, New Jersey, 1989.

[9] Lee, W.C.Y., Overview of Cellular CDMA, IEEE Transactions on Vehicular Technology, Vol. 40, No. 2, pp. 291 - 302, May 1991.

[10] Demers, A., Elrod, S., Kantarjiev, C., A Nano-Cellular Local Area Network Using Near-Field RF Coupling, Virginia Tech's Fourth Symposium on Wireless Personal Communications, 1994.

[11] LaMaire, R.O., Krishna, A., Ahmadi, H., Analysis of a Wireless MAC Protocol with Client-Server Traffic and Capture, IEEE Journal on Selected Areas in Communications, Vol. 12, No. 8, pp. 1299 - 1313, Oct. 1994.

[12] Raychaudhuri, D., Wilson, N.D., ATM-Based Transport Architecture for Multiservices Wireless Personal Communication Networks, IEEE Journal on Selected Areas in Communications, Vol. 12, No. 8, pp. 1401 - 1414, Oct. 1994.

AN INTEGRAL EQUATION APPROACH TO THE PREDICTION OF WAVE PROPAGATION IN AN INDOOR ENVIRONMENT

B. De Backer[†]   *F. Olyslager*[‡]   *D. De Zutter*[‡]

Department of Information Technology (INTEC), Ghent University
Sint-Pietersnieuwstraat, 41, 9000 Ghent, Belgium
Tel : 32 9 264 3354, Fax : 32 9 264 3593
email : bernard.debacker@intec.rug.ac.be

Abstract. A thorough understanding of the mechanisms which govern wave propagation in an indoor environment is needed for the design and performance analysis of the next generation of mobile communication systems (PCS). In this paper a novel technique, based on the rigorous Moment Method (MoM), is introduced for the prediction of indoor wave-propagation. To this end two numerical methods are presented which significantly reduce the memory and computational requirements, associated with the use of the MoM, and the technique is illustrated by an example.

I. INTRODUCTION

Over the last few years significant research efforts have been spent on the development of reliable and accurate wave-propagation models for the indoor environment in the UHF-region (0.3-3 GHz). Given the complexity of electromagnetic wave behaviour in this environment, such models are highly desirable as design tools in the network optimisation stage of cellullar based systems, where a balance needs to be found in the drive for both maximal capacity, minimal co-channel interference, and minimal cost.

For the main part these efforts have been directed towards the development of ray-tracing based models, which essentialy lead to a high-frequency approximation of Maxwell's equations. In this approximation rays, which represent tubes of energy, are launched from the transmitter location. The interaction of these rays with partitions within the builing are then followed up until they reach the receiver location or until their associated energy drops below a given threshold value. A number of authors [1-3] demonstrated the viability of such an approach, using simple geometrical optics (GO) rules of reflection and transmission to model the mentioned interaction. While their results were certainly encouraging, the same authors recognised that in an indoor environment a significant amount of energy can also reach the receiver by means of diffraction at wall, door or window edges. This is well explained by the fact that in most cases, a line of sight (LOS) path between transmitter and receiver will not exist. In future systems this path might even be purposefully obstructed to achieve a uniform cell coverage and a reduced dynamic range of the field strength within this cell.

To include diffraction in the ray-tracing approach, Keller's geometrical theory of diffraction (GTD), or one of it's derivatives, has been used [4,5]. The applicability of these theories is however restricted to objects of larger dimensions (a few wavelengths or more) and of specific geometry (wedge- or cone-like structures). Add to this the significant increase in computational complexity if used in ray-tracing (each point of diffraction introduces an additional source), to conclude that this method leaves a lot to be desired.

Until recently no attempts were made at solving Maxwell's equations in their entirety for the indoor propagation problem, given the large geometrical dimensions of the considered space. In [6,7] however, a two dimensional FDTD scheme was used. In [7] it was shown that reducing the propagation problem to two dimensions is acceptable for power level predictions, leading to significantly reduced memory and computational requirements.

In this paper we present another rigorous approach for solving Maxwell's equations for the indoor propagation problem. In section II a set of boundary integal equations will be formulated. These equations will be solved using the Moment Method (MoM). To improve the efficiency of this method for problems of large dimensions, two numerical methods were developed. These are presented in sections III and IV. Finally, in section V the technique will be illustrated by an example.

II. INTEGRAL EQUATIONS FORMULATION

In this section we show how a set of coupled Boundary Integral Equations (BIE) can be formulated which, in combination with the Moment Method, can be used for the field strength predic-

tion in an indoor environment. This should merely be seen as an introduction to a technique which has been widely and succesfully used in many other areas. For a more detailed discussion on BIE we for example refer to [8,9].

A. Geometry of the problem.

In our approach the indoor environment is modelled as a two dimensional configuration. A general geometry of this configuration is shown in figure 1. The walls and other partitions in the environment are modelled as homogeneous and possibly lossy isotropic materials, characterised by an arbitrary relative complex permittivity $\epsilon_{r,j}$ and complex permeability $\mu_{r,j}$ (j=1 ...L). The number, size, or position of these materials is arbitrary as well, opening the possibility to model the environment in any given detail.
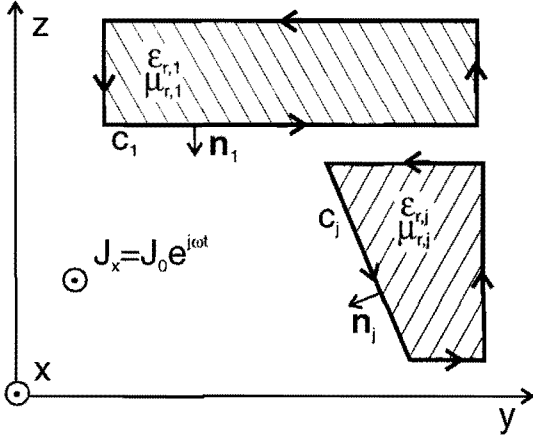
Figure 1 - General geometry.

In this configuration a two dimensional line source is introduced. The surface current along this source is fixed at $J_x = J_0 exp(j\omega t)$, where J_0 is constant and ω is the angular frequency. Due to the two dimensional nature of the problem this source will excite a TE-field in the modelled indoor environment. This means that the electromagnetic field will be composed of an E_x-field component perpendicular to figure 1, and an H_t-field component in the plane of this figure.

B. A set of integral equations.

Starting from Maxwell's equations, it can be shown, using Green's identity, that these E_x and H_t field components in any point **r** inside dielectric structure j (j=1 ...L) can be expressed as a contour integral along the boundary c_j of this structure (the time dependence $exp(j\omega t)$ is suppressed in these, and in the following equations) :

$$E_x(\mathbf{r}) = \oint_{c_j} \left(\frac{\partial G_j}{\partial n'} \cdot E_x(\mathbf{r}') - \frac{jk_j^2}{\omega\epsilon_j} \cdot G_j \cdot H_t(\mathbf{r}') \right) \cdot dc' \quad (1)$$

$$H_t(\mathbf{r}) = \oint_{c_j} \left(-\frac{j\omega\epsilon_j}{k_j^2} \cdot \frac{\partial^2 G_j}{\partial n \partial n'} \cdot E_x(\mathbf{r}') - \frac{\partial G_j}{\partial n} \cdot H_t(\mathbf{r}') \right) \cdot dc' \quad (2)$$

with $k_j = \omega\sqrt{\epsilon_j\mu_j} = 2\pi/\lambda_j$, λ_j the wavelength in dielectric j, and with G_j the 2-D homogeneous space Green's function of the Helmholtz equation:

$$G_j(\mathbf{r}|\mathbf{r}') = \frac{j}{4}H_0^{(2)}(k_j|\mathbf{r} - \mathbf{r}'|) \quad (3)$$

$H_n^{(2)}(z) = J_n(z) - jY_n(z)$ is the n-th order Hankel function of the second kind.

Expressions, similar to (1) and (2) can be derived for the field in the free space surrounding the various structures. In this case the contour of integration comprises of all individual contours c_j, taken in opposite directions. Given these expressions, the E_x and H_t field components can be determined in any point in space if their boundary values are known on all the interfaces c_j of the respective structures. By imposing continuity of expressions (1), (2) and those for the free-space area at these boundaries, a set of coupled integral equations is obtained [8], from which these boundary values can be extracted. This is done using the Moment Method.

C. The Moment Method.

Through the use of this method the coupled integral equations are transformed to a discretised set of linear equations. This is done by introducing a set of linear segments along the given boundaries c_j, and by expanding the unknown boundary field components H_t and E_x along these segments :

$$H_t = \sum_i H_{t,i}p_i \qquad E_x = \sum_i E_{x,i}t_i \qquad i = 1\dots N \quad (4)$$

H_t is approximated in a piecewise constant way (p_i are pulse functions) and, for reasons of continuity, E_x is approximated in a piecewise linear way (t_i are overlapping triangle functions). This is shown in figure 2 for the first dielectric structure of figure 1.
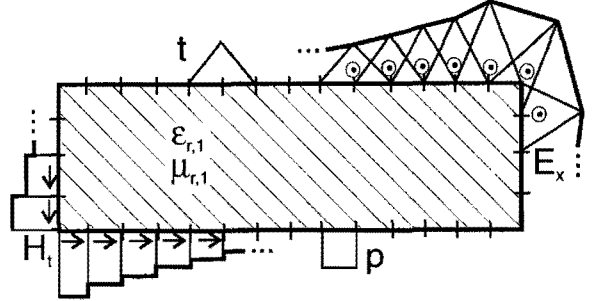
Figure 2 - Expansion of the boundary fields.

The expansion coefficients $H_{t,i}$ and $E_{x,i}$ in expressions (4) are determined by imposing the integral equations in a weighted sense. To do so, Galerkin testing is used where the E_x-equations are weighted with pulse functions and the H_t-equations with triangle functions. Testing these expressions along the boundaries leads to a system of linear equations:

$$S \cdot X = Y \quad (5)$$

the matrix S in this representation is composed of four N-dimensional square submatrices, each representing one out of the four interactions between the

29

N pulse and N triangle functions, calculated with the appropriate integral kernel from (1-2). The vector X contains the 2N unknown expansion coefficients $E_{x,i}$ and $H_{t,i}$. In vector Y the interactions are found which correspond to the H_t-pulse functions along the boundary of the source structure. The expansion coefficients, associated with these functions are known, since along this boundary $H_t = J_0$ (note that the E_x-field along this boundary is not fixed by the source condition).

D. Requirements of the classical Moment Method.

Traditionally, the use of the Moment Method has been restricted to structures of up to a few wavelengths or less. Due to the computational requirements ($2N(2N + 1)$ elements need to be calculated and a $2N \times 2N$ system needs to be solved), and the memory requirements ($2N(2N + 1)$ elements need to be stored), only powerfull supercomputers were capable of solving larger problems. In the next two sections we present two numerical methods which significantly ease these restrictions, making it possible to perform indoor wave propagation simulations on an average workstation.

III. THE INTERACTION INTEGRALS

The calculation of the interaction matrix S in (5) requires the evaluation of four types of double integrals. One integration relates to the expansion function of the unknown field component, the other to the testing function. If we represent these functions by respectively b_e and b_o (e stands for 'excitation', o for 'observation'), these integrals can be grouped as follows:

$$\int_{-o1}^{o1} do \int_{-e1}^{e1} de \left[b_o(o) b_e(e) K(\mathbf{r}|\mathbf{r'}) \right] \qquad (6)$$

In this expression K represents the appropriate integral kernel from (1-2). Local coordinates e and o were introduced along the excitation and observation cells as shown in figure 3. This figure also shows $\mathbf{r}$ and $\mathbf{r'}$.
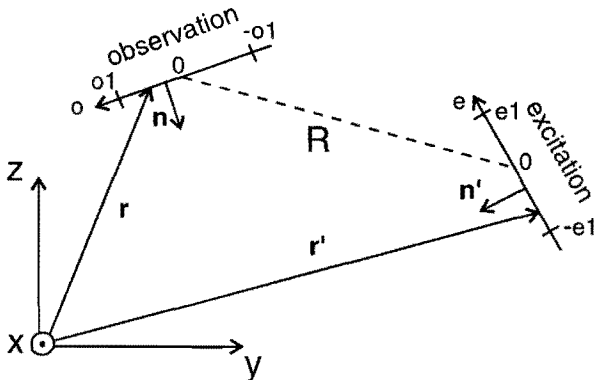
Figure 3 - Coordinate systems.

A. Gaussian quadrature formulas.

Classically simple Gaussian quadrature formulas are used to perform the double integration in (6). In this case the dominant time factor in this integration

will by far be the recurrent evaluation of the Hankel function, $H_0^{(2)}(k_j|\mathbf{r} - \mathbf{r'}|)$ and/or, resulting from its derivative, $H_1^{(2)}(k_j|\mathbf{r} - \mathbf{r'}|)$, contained in the kernel K. If, as common, a 10 point quadrature formula is used, these complex functions need to be evaluated 100 times each, for each interaction integral. As the cell separation R (see figure 3) increases, such complexity however rapidly becomes unwarranted, as shown next.

B. Expansion of the integrand.

If the excitation and observation cells are separated by a few wavelengths or more, the argument of both the Hankel functions will show only small variations. This observation lead us to expand these Hankel functions in a Taylor series, centered around $k_j R$, the argument corresponding to the distance between both cell centers (see figure 3).

$$H_n^{(2)}(k_j|\mathbf{r} - \mathbf{r'}|) \approx \sum_k \frac{1}{k!} \cdot \left. \frac{d^k H_n^{(2)}(z)}{dz^k} \right|_{z=k_j R} (\alpha)^k \qquad (7)$$

with $n = 0$ or 1 and $\alpha = k_j(|\mathbf{r} - \mathbf{r'}| - R)$. We found that for cell sizes less than $\lambda_j/8$ and cell separations greater than $\lambda_j/2$, a relative accuracy of 0.5% can be obtained if only three terms of (7) are used in the integration. Such an accuracy proved more than adequate for the considered scattering problem. Futhermore, since the derivatives in (7) can be expressed in terms of $H_0^{(2)}(z)$ and $H_1^{(2)}(z)$, only two Hankel functions need to be evaluated per integration, leading to very small computation times.

C. Asymptotic approximation of the integrand.

For even larger cell separations, the Hankel functions in the integrand can be approximated by their asymptotic expansions :

$$\lim_{\beta \gg |n|} H_n^{(2)}(\beta) = \sqrt{\frac{2}{\pi\beta}} e^{-j\beta + j\pi/4(2n+1)} \qquad (8)$$

with $\beta = k_j|\mathbf{r} - \mathbf{r'}|$. To obtain an analytically integrable form, further far-field approximations can be made of the distance related factors in the integrand. We found that these asymptotic expansions can be used with the same accuracy as above, for far-field cell separations starting from two wavelengths for the Hankel function itself, to fifteen wavelengths for its derivatives.

D. Simulation of larger problems.

Through the use of these two integral approximations, computation times for larger problems can be kept to an acceptable level. Given that these approximations are generally 50 to 200 times faster in calculation than Gaussian integration, the near-field interactions will take up most of the time in the computation of the system matrix S in (5). If we assume an average of 20 near-field cells, and an average speedup factor of 100 as compared to Gaussian integration, the computational complexity will not rise as $2N(2N + 1)$, but rather as $(2N)^2/100 + 20(2N)$.

30

IV. Sparse Matrix Transformation

As mentioned in section II, a second major impediment to using the MoM in the simulation of larger problems lies in the high memory requirements, associated with the storage of the $2N(2N+1)$ interaction integrals. Even though the use of high-frequency approximations like ray-tracing might in this case not be entirely justified, their very existence suggests the possibility of a significant reduction in the complexity of the description of the high-frequency interaction between the respective structures in the indoor environment.

A. Discrete transformation.

To achieve such a reduction in the presented MoM-formalism, we reconsider the system of linear equations (5), where the vector of unknowns X is written as $X = A \cdot U$, and each side of the system is multiplied by a matrix B to obtain :

$$B \cdot S \cdot A \cdot U = B \cdot Y \qquad (9)$$

or

$$T \cdot U = W \qquad (10)$$

with $T = B \cdot S \cdot A$, $W = B \cdot Y$, and where A and B are $2N \times 2N$ matrices. These equations show how the original vector of unknowns X can be found by first solving for U in the alternate form (10), followed by the transformation $X = A \cdot U$. Our motivation to introduce (9) lies in finding transformation matrices A and B, which render T sparse, that is which introduce a significant number of near-zero elements in T, which can be discarded in the solving of system (10).

B. A new set of basis functions.

As shown in [10] these matrices A and B can be seen as perfoming a discrete transformation to a new set of respectively expansion and testing functions. Therefore our problem could alternatively be formulated as the search for a new set of basis functions, which reduce, or localize the far-field interactions in the system matrix T.

It is now a well known fact in electromagnetism that the far-field properties of a given source distribution are directly related to its spatial spectral content [11]. Since the basis functions, presented in section II, show no, or only linear spatial variations, and since they were defined along cells, no longer than $\lambda_j/8$, the spectral content of these functions will be very broad, and they will therefore radiate in all directions, leading to a dense interaction matrix S. The localized extent of these functions on the other hand offers the opportunity to construct, by linear combination, a new, and much improved set of basis functions.

To this end the segments along the straight sections of each dielectric structure are grouped in a number of blocks. Assuming a block of M segments, new basis functions $s0_j$ and $s1_j$ ($j = 1...M$) are defined along this block as:

$$s0_j = \sum_{k=1}^{M} p_k sin(\pi(j - 0.5)(k - 0.5)/M) \qquad (11)$$

$$s1_j = \sum_{k=1}^{M} t_k sin(\pi(j - 0.5)(k - 0.5)/M) \qquad (12)$$
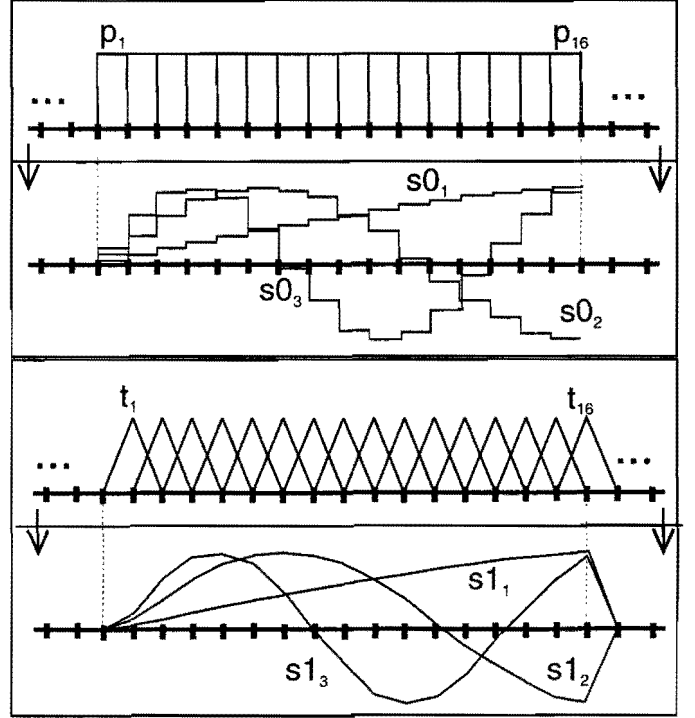
Figure 4 - A new set of basis functions.

with ($k = 1...M$), and where t_j is the rising triangle function, defined over cell k. This is shown in figure 4. For reasons of clarity only three of both the $s0$ and $s1$ functions are shown. Also, for the same reason, a block of 16 segments was assumed ($M = 16$). In reality, even larger blocks are taken, extending over a few wavelengths of the dielectric structure. The resulting basis functions $s0_j$ and $s1_j$ will have a much reduced spatial spectral content, and will therefore only radiate in specific directions, leading to a reduced range of interaction, and significantly fewer elements in the interaction matrix T. Further improvements are achieved by using additional taper functions to suppress possible radiation sidelobes. Depending on the size and geometry of the problem, from 80 to 95% of all elements in T can in this way be discarded. An example of this will be given in the presentation.

C. Simulation of larger problems.

A final note on the transformations performed by matrices A and B. For our choice of new basis functions, (11) and (12), these matrices perform a two dimensional discrete sine transformation (on a block by block basis) of the interaction matrix S. By using fast fourier algorithms in the implementation of this transformation, computation times can be kept at very acceptable levels, even for large problems.

V. Examples

In this section an example is presented which illustrates the outlined technique. In the choice of this example we have tried to emphasize the improvements on a ray-tracing approach. A comparison with

31

ray-tracing or measurements however is not included. This issue is addressed by another paper [12].
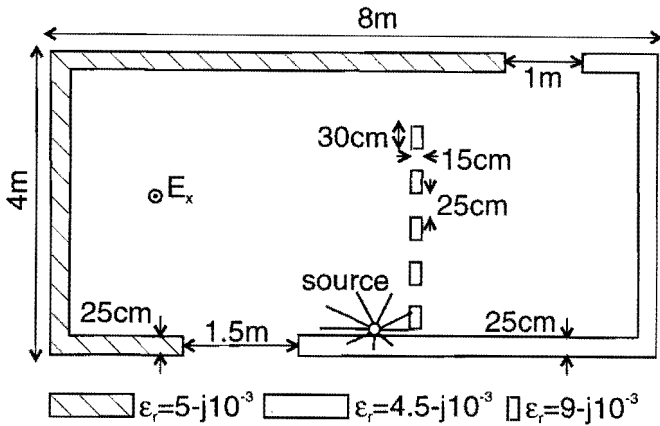
Figure 5 - The considered indoor environment.

Figure 5 shows the indoor environment under consideration. Two adjacent rooms are represented, in which an electromagnetic field is induced at 0.9 GHz. The outside walls of this configuration are characterized by a slightly different permittivity, as shown in the figure. In addition, both rooms are separated by a partitioning wall, composed of a number of smaller elements of higher permittivity. In the outside walls two apertures are found, one of 1m wide (representing for example an open door), another of 1.5m (a window, for example).

In figure 6 the induced electromagnetic field is represented. Of this TE-field, only the E_x-field component is shown, since it is this component that will be of importance at the receiving end of the mobile link. This plot shows an overall field distribution, in dB, in which the respective parts of the environment are still clearly visible. It is clear that those parts in the immediate vicinity of the source, strongly influ-ence the overall field distribution. This influence is clearly a near-field effect, and can therefore not be accurately modelled by a ray-tracing approach.

A further improvement on the ray-tracing approach is clearly the inclusion, in the MoM-formalism, of all relevant propagation mechanisms. This means that the transmission through the partitioning section, and the diffraction at the wall edges, are accurately modelled with no additional complexity. The effect of this diffraction is particulary clear at the corners on the right of the configuration.

The field distribution, represented in figure 6, is shown in its full detail. That is, no filters were used to spatially average this result, as is commonly done in a ray-tracing approach to take account of its high-frequency nature [3]. This of course does not mean that we pretend our result to be accurate down to a wavelength scale. The accuracy of the solution might however be sufficient to recognise those areas in the indoor environment where the fading is strongest, both in level and in spatial distribution.

Finally, a note on the calculation of the plot, shown in figure 6. The used MoM offers the advantage that, once the field values are known on the boundaries of the respective structures in the indoor environment, the calculation of such an overall field plot requires relatively little additional computation time. This is not the case for the ray-tracing approach where the problem needs to be resolved for each field location in this plot. From such a field plot, the important wave-propagation mechanisms for the specific indoor environment can easily be identified, leading to an important insight in the overall propagation problem, which is what will be required in future mobile network design.
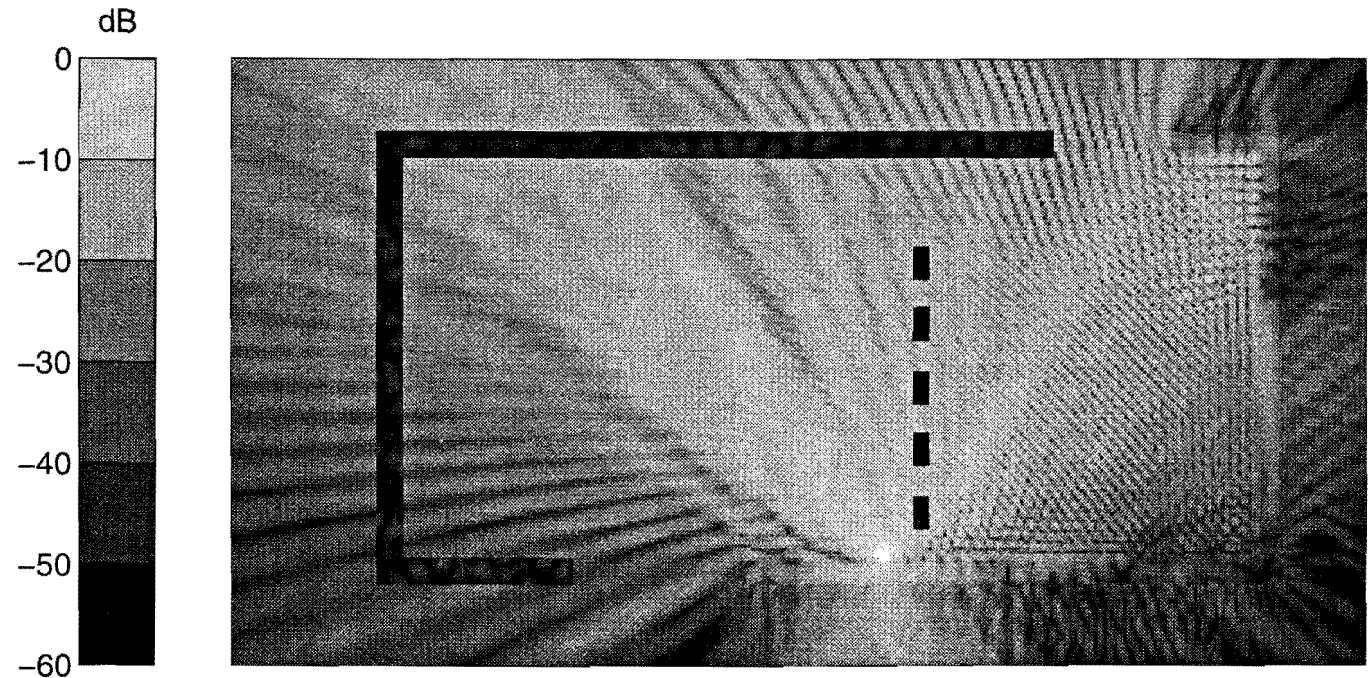
Figure 6 - Induced E_x-field distribution.

VII. References

[1] S.Y. Seidel, S. Rappaport *A ray tracing technique to predict path loss and delay spread inside buildings.* IEEE Globecom '92, pp. 1825–1929.

[2] J.W. McKown, R.L. Hamilton *Ray tracing as a design tool for radio networks.* IEEE Network Magazine, Nov. 1991, pp. 27–30.

[3] W. Honcharenko, H.L. Bertoni, J.L. Dailing, J. Qian, H.D. Yee *Mechanisms governing UHF propagation on single floors in modern office builings.* IEEE Trans. on Vehic. Techn., Vol. 41, No. 4, Nov. 1992, pp 496–504.

[4] S.Y. Seidel, S. Rappaport *Site-specific propagation prediction for wireless in-building personal communication system design.* IEEE Trans. on Vehic. Techn., Vol. 43, No. 4, Nov. 1994, pp. 879–891.

[5] P. Kreuzgruber, T. Bründl, W. Kuran, R. Gahleitner *Prediction of indoor radio propagation with the ray splitting model including edge diffraction and rough surfaces.* IEEE Vehic. Techn. Conf. '94, pp. 878–882.

[6] A. Lauer, A. Bahr, I. Wolff *FDTD simulations of indoor propagation.* IEEE Vehic. Techn. Conf '94, pp. 883–886.

[7] G. Yang, S. Li, J-F. Lee, K. Pahlavan *Computer simulation of indoor radio propagation.* PIMRC '93, pp.10–14.

[8] F. Olyslager, D. De Zutter *Rigorous boundary integral equation solution for general isotropic and uniaxial anisotropic dielectric waveguides in multilayered media including losses, gain and leakage.* IEEE Trans. Microwave Theory Tech., vol. MTT-41, No. 8, Aug. 1993, pp. 1385–1392.

[9] F. Olyslager, D. De Zutter, K. Blomme *Rigorous analysis of the propagation characteristics of general lossless and lossy multiconductor transmission lines in multilayered media.* IEEE Trans. Microwave Theory Tech., Vol. MTT-41, No.1, Jan. 1993, pp. 1385–1392.

[10] F.X. Canning *Transformations that produce a sparse moment method matrix* J. Electromagnetic Waves Appl., Vol. 4, no. 9, 1990, pp. 893–913.

[11] J. Van Bladel *Electromagnetic Fields.* Springer-Verlag, Berlin, 1985.

[12] B. De Backer, H. Börjeson, F. Olyslager, D. De Zutter *Electromagnetic study of wave propagation through a windowed wall at 1.8 GHz.* Submitted to the '96 IEEE Vehic. Techn. Conf.

A FIELD STRENGTH PREDICTION MODEL FOR SMALL CELLS MOBILE SYSTEMS USING A 3-D BUILDING DATABASE

Maria Perucca

CSELT, via G. Reiss Romoli, 274
10148 TORINO-ITALY
Tel: +39 11 2287260
Fax: +39 11 2285577
E-mail: maria.perucca@cselt.stet.it

Abstract

Cells with reduced size (few km) require propagation models that take into account the building structures.
In case base station antennas are installed just above the roof-top of a building or, more generally, at the same height of the surrounding buildings ("small cells"), the path loss is determined mainly by diffraction and scattering phenomena, occurring at the roof-tops encountered by the propagating wave.
A new model for the computation of the field strength in a small cells environment is presented. It is based on a simplified evaluation of the building diffraction: buildings are considered as knife edges and Deygout's method is utilized for the diffraction loss evaluation. Despite this simplification, with respect to other models, which are based on average parameters and do not consider the real building structures, the new model is easily applicabile to any urban structure, provided that a suitable 3-D topographical data base (with adequate horizontal and vertical resolution, 1 to 2m) is available.

1. Introduction.

Future cellular systems have to provide a higher capacity than actual systems, in order to cope with the ever increasing demand of personal communications. The reduction of cell dimensions, in order to decrease the reuse distance and, consequently, to increase the system capacity, appears as a suitable solution.
Because of the reduced cell sizes, usual statistical macrocellular propagation prediction procedures are no longer applicable, and new models are needed, taking into account the real building structures.
When the cell reduction is obtained by positioning the base station (BS) well below the roof-top heights of the surrounding buildings ("microcells"), 3÷10m height from ground level, the waves mainly propagate in street canyons, guided by buildings. Studies regarding statistical [1] and rigorous [2] (based on ray tracing or ray launching techniques using GTD or UTD approaches) models are in progress.
In "small cells" (BS height just above roof-tops). however, the pathloss is determined mainly by diffraction and scattering at roof-tops, i.e. the main rays propagate above the roof-tops. Such a scenario makes the definition of a propagation model quite critical: although some models based on ray tracing techniques and UTD [3] have been investigated, the complexity of the environment (involving many scattering surfaces and multipath propagation) makes it difficult to apply such models to a real scenario: on the other hand, simplified models[4], requiring only a few parameters, provide good predictions only in case of large areas

described with a poor spatial resolution, and do not appear sufficiently detailed in this context of interest.

Moreover, the application of these models to a real scenario is also a critical point: that is how to extract the required propagation parameters required from maps or topographical database (provided that such maps are available).

In the following, a new model is presented together with a measurement campaign results, performed in Turin. Preliminary comparisons between measured and simulated values are reported.

2. The new approach.

From the previous considerations, it comes out that a model for small cells, in order to be an effective tool in planning such systems, has to satisfy the following two main requirements:

- it has to retain the real buildings structure information, avoiding averaging procedures of the geometrical features of the area under examination;

- the application to a real zone of an urban area, with the use of topographical maps or data base, has to be of easy achievement.

Despite a simplified diffracted field calculation, the new model developed in CSELT has proven to fulfill the previous conditions.

2.1 Outline of the model.

In order to evaluate the buildings attenuation by diffraction, the model makes use of the approximated solutions based on Huygens-Fresnel theory.

A single obstacle (building) is assumed to be a perfectly absorbing half-plane screen (knife-edge approach), having no thickness along the propagation direction and an infinite extension in the transversal section. This approach is correct as long as the signal wavelenght is negligible with respect to the obstacle size and to the distances between obstacles or between the Tx or Rx and the obstacle; this condition is satisfied in UHF band and upper bands.

A rigorous solution can be easily obtained only in the case of a single knife edge [5], whereas an extension to several obstacles is a more complex task. Therefore, approximations that try to take into account the interactions among consecutive obstacles by means of a proper combinations of single obstacle contributions are made. The differences among the various models lie in the different methodology used to evaluate the interaction phenomena [6],[7].

In the case of above roof-tops BS and mobile terminal at street level, the last building generally obstructs deeply the direct radio path, and can be considered as the main obstacle affecting the link. This consideration

justifies the choice of Deygout's approach [7] in respect to other ones, because it leads (as a general opinion) to a more accurate extimation if a single main obstacle exists. The selection of the main obstacle is made in terms of the greatest $\dfrac{h}{\rho_1}$ ratio (ρ_1 is the first Fresnel zone ray of the considered path and h the clearence), and the corresponding diffraction loss is computed as if it were alone. The same procedure is then recursively applied to the two path sections generated by the main obstacle until all obstacles are considered. For each main obstacle, the values of d_1, d_2, h for the parameter $v_o = \sqrt{2}\,\dfrac{h}{\rho_1}$ are computed and the field attenuation is given by:

$$\frac{E}{E_o} = \frac{1+j}{2}\int_{v_o}^{\infty} e^{-j\frac{\pi}{2}v^2}\,dv$$

The loss contribution L_i due to one screen is:

$$L_i = -20\log\left|\frac{E_i}{E_o}\right| =$$
$$= 6 - 10\log\left[1 + 2C^2(v_o) + 2S^2(v_o) \pm 2(C(v_o) + S(v_o))\right]$$

where $C(x), S(x)$ are the Fresnel integrals and the sign - or + refers to positive or negative values of v_o respectively; the total path loss is then given by:

$$L = L_o + \sum_i L_i$$

where L_o refers to free space loss. The contribution due to a non isotropic antenna pattern can be added to this basic loss equation.

2.2 Application to a 3-D data base of Turin: simulation results.

The above model has been applied to a zone of Turin, for which digitized maps were available, in order to obtain field coverage and "best server" maps.

The basic cartography is related to the central part of Turin; the scale of each map is 1:2000 and its size is 1800m x 1200m. A spatial resolution of 1m has been considered as the best choice, in order to avoid inaccuracy errors. Terrain data have been digitized and stored in two data bases: the former includes the ground height above the mean sea level (in this case the term "ground" includes also the buildings); the latter provides, for each surface elements of 1mx1m, a binary information on the presence (or not) of a building. The elevation accuracy has been estimated to be around 2m. Data are organized in 1800 x 1200 elements arrays.

Fig.1 shows the map considered for the calculations: it represents the buildings height in a densely built up zone in the center of Turin. Since the terrain is nearly flat, the utilized data base contains actually only the height of buildings above the ground. For each configuration of BS and mobile terminal, the direct radio path is considered and the corresponding building height profile (with a 0.5m resolution step) is obtained from the data base. Then, each building is substituded by a screen with the same height and located in the middle of the building, giving rise to the screens series, to which the Deygout's method is applied.

Simulations have been performed for three different BS's heights 30m, 38m and 48m (2m, 10m and 20m above the roof-top of the building, respectively) and four pointing directions (with respect to North: 90°, 30°, 330° and 270°), positioning the BS on a TELECOM ITALIA mobile radio site (in Tab.1 geographical and electric parameter of the BS are reported).

In Figg.2a, 2b the field coverages for two BS heights (30m and 48m) and one pointing direction (330°) are depicted: field values are reported for each element of the map within the main lobe, with a colour code corresponding to different field levels (30, 40, 50, 65 and 80dB (μV/m)). The color of the buildings is black irrespectively of their height.

Some preliminary considerations can be inferred by the simulated coverages: the values found in computations appear to agree with expected field pattern, from both qualitative and quantitative points of view, and the loss pattern seems to be consistent with the building structure.

In general, it can be observed that the shape of a cell is very irregular: high field strengths can be found even far from the transmitting antenna in LOS streets or in open areas, whereas greater loss values are generated by high buildings also in the vicinity of the BS. Strong dependency on the street orientation is also found. Moreover, lower propagation loss is encountered when the BS height is increased.

3. Measurement campaign and experimental results.

A measurement campaign has been carried out in the center of Turin. The transmitting antenna was positioned on a TELECOM ITALIA site (Fig.1 shows the exact position on the map), at three different heights and four pointing directions. The receiver was mounted on a van. In Tab.1 the characteristics of transmitter and receiver are described.

Transmitting antenna	
Antenna characteristics	
type:	eight dipole array
gain:	18 dB
3dB lobe H plane:	60°
3dB lobe E plane:	8°
Transmitting parameters	
freq.:	1795 MHz
transmitting power:	1 W
heights:	30m, 38m, 48m
pointing directions:	90°, 30°, 330°, 270°
Receiver:	
dipole $\lambda/2$	
height:	2m

Tab.1: Antennas' characteristics and parameters.

Electrical field values were authomatically aquired and visualized in real time during the measurements. In particular Rx positions (usually at a street crossing, or at a change of direction), markers were memorized, in order to authomatically localize the measured routes on the digitized maps. The spatial sampling distance of the measurements was set to 5cm.

In order to compare measured and simulated values, the following processing has been carried out: measurements routes have been digitized and coordinates have been associated to each measured value, using the markers' information, afterwards, this values have been assigned to the corresponding pixel of the digitized map, the resulting electrical field is derived as the average over all the measured values laying on the same pixel (usually 5÷40).

From a qualitative point of view, some considerations can be drawn, which are in common with simulation results:
- received field shows a strong dependence on the building lay-out, high values are found at high distances from BS in particular cases of street orientations or open areas, whereas strong attenuation occurs in narrow streets among high buildings;
- generally, field values increase when the BS height increases.

4. Preliminary comparison between measurements and simulations.

One BS configuration (height: 38m; pointing direction: 330°) is chosen and reported as an example in Figg.3a, 3b and 3c. Field coverage (Fig.3a) and measured values (Fig.3b) have been compared: in Fig.3c the errors (differences between simulated and measured values) are depicted, in order to emphasize the zones where measured values are over- or under-estimated.

	BS HEIGHT		
Pointing direction	30m N mean err. st. dev	38m N mean err. st. dev	48m N mean err. st. dev
90°	13757 -9.24 15.85	13727 -0.39 16.03	13772 -4.24 15.04
30°	18065 -3.51 15.62	18103 8.03 14.56	18088 -1.38 11.80
330°	26526 -1.89 14.69	25038 -2.06 11.8	26569 -6.04 11.29
270°	16680 -4.95 15.19	16695 -3.03 13.94	15700 -5.95 13.46
Total	75028 -4.31 15.25	73563 0.58 13.84	74129 -4.55 12.65

Tab.2: Comparison results.

A preliminary estimate of the model quality can be obtained from the mean value and standard deviation of the errors between simulated and measured values. In Tab.2 results for the different pointing directions and BS heights are reported separately: the number of the considered pixels N, the mean error, and its standard deviation can be found. In the last row, the mean values for each BS height is reported.

The results show that, on the average, the model is not strong biased but, on the other hand, quite high values of standard deviation are found. Usually, lower values of standard deviation are obtained for the heigher BS height. This fact points out that BS antennas positioned nearly at the same height of the roof-tops is, as expected, a critical situation.

At this point, some considerations can be drawn. Since the model is not statistic, but it is based on a deterministic approach, mean error values averaged over all the points do not describe satisfactorly the model's performance. A better way to validate the model is to group into classes the different situations and analyse each of them separately. In this way, particular situations in which simulated values are affected by large errors can be individuated and, consequently, adequate solutions can be investigated in more detail.

As an example of a typical situation that leads to larger errors is a high isolated building with finite width causing in its shadowed zone simulated attenuation values higher then measurements. This fact can be noted in Fig.3, where a high building causes a deep shadow zone in simulations, whereas measured values do not seem to be affected by such a heigh attenuation. A better investigation of the performance of the model can suggest, as a possible improvement, a better characterization of the main obstacle.

5. Further improvements.

The above model is an initial attempt to find a suitable tool for small cell systems radioplanning. Its formulation can be further improved when a cause of error is individuated and isolated.

First of all, when a heigh isolated building is encountered, a further contribution can be introduced to take into account the diffraction occurring at the building's sides, since knife-edge approach assumes an infinite extension in the transversal section.

Moreover, diffraction phenomena by a thick screen can be studied in order to better characterize the last (or main) screen.

Another possible improvement can be achieved by a better geometrical description of the street in which the mobile terminal is located: the last diffracting screen can be considered as a radio source generating direct and reflected (from the street wall) rays.

6. Conclusions.

The propagation channel model in a small cells environment has been developed; a deterministic approach, based on the Deygout's solution of the multi-screen diffraction, is adopted. It takes into account the effective geometrical structure of the buildings and, moreover, appears to be easily applicable to different areas, jointly with a proper topographical data base with adequate resolution.

Measurements have also been performed in the center of Turin, in order to validate the method.

Preliminary comparisons between measurements and simulations have been performed: quite good agreement is found in some situations, whereas discrepancies in specific situations suggest that possible improvement should be done, in order to achieve a better global performance of the model.

References.

[1] J. E. Berg, R. Bownds, F. Lotse, "Path Loss and Fading Models for Microcells at 900MHz", IEEE 1992.

[2] S. Y. Tan, H. S. Tan, "UTD Propagation Model in an Urban Street Scene for Microcellular Communications", IEEE Trans. El. Comp., Vol.35, No.4, November 1993.

[3] D.J. Chicon, T. Kürner, and W. Wiesbck, "Modellierung der Wellenausbreitung in urbanem Gelände", *FREQUENZ*, Band 47, Heft 1-2, 1993, pp. 2-11.

[4] COST 231 subgroup on propagation models, "Urban propagation models for small cell and micro cell mobile radio in the 900 and 1800 MHz bands", COST 231 TD(90) 119.

[5] D. C. Livingstone, "The physics of microwave propagation", Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1970.

[6] J. Epstein, D. W. Peterson, "An experimental study of wave propagation at 850 Mc/s", Proc. IRE, Vol.41, No.5, May 1953.

[7] J. Deygout, "Multiple knife-edge diffraction of microwaves", IEEE Trans. AP, Vol.14, No.2, 1966.

Colour map:

BUILDING HEIGHTS

Area: t33Ba

UTM: 394100 4991680

X axis [m]: 1800
Y axis [m]: 1200

up to [m]: 10
up to [m]: 20
up to [m]: 25
up to [m]: 30
up to [m]: 35
> [m]: 35
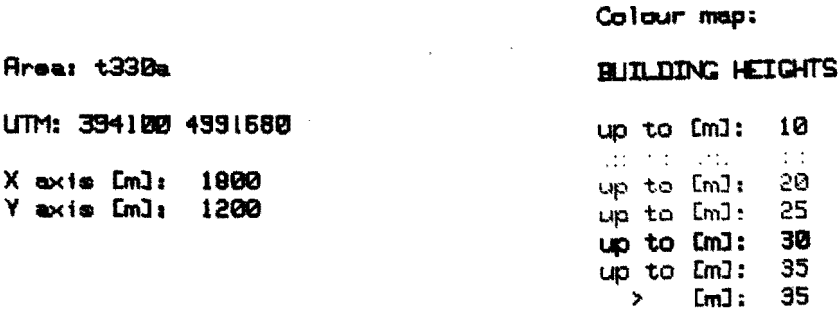
Fig.1 Building heights representation for the area of interest.
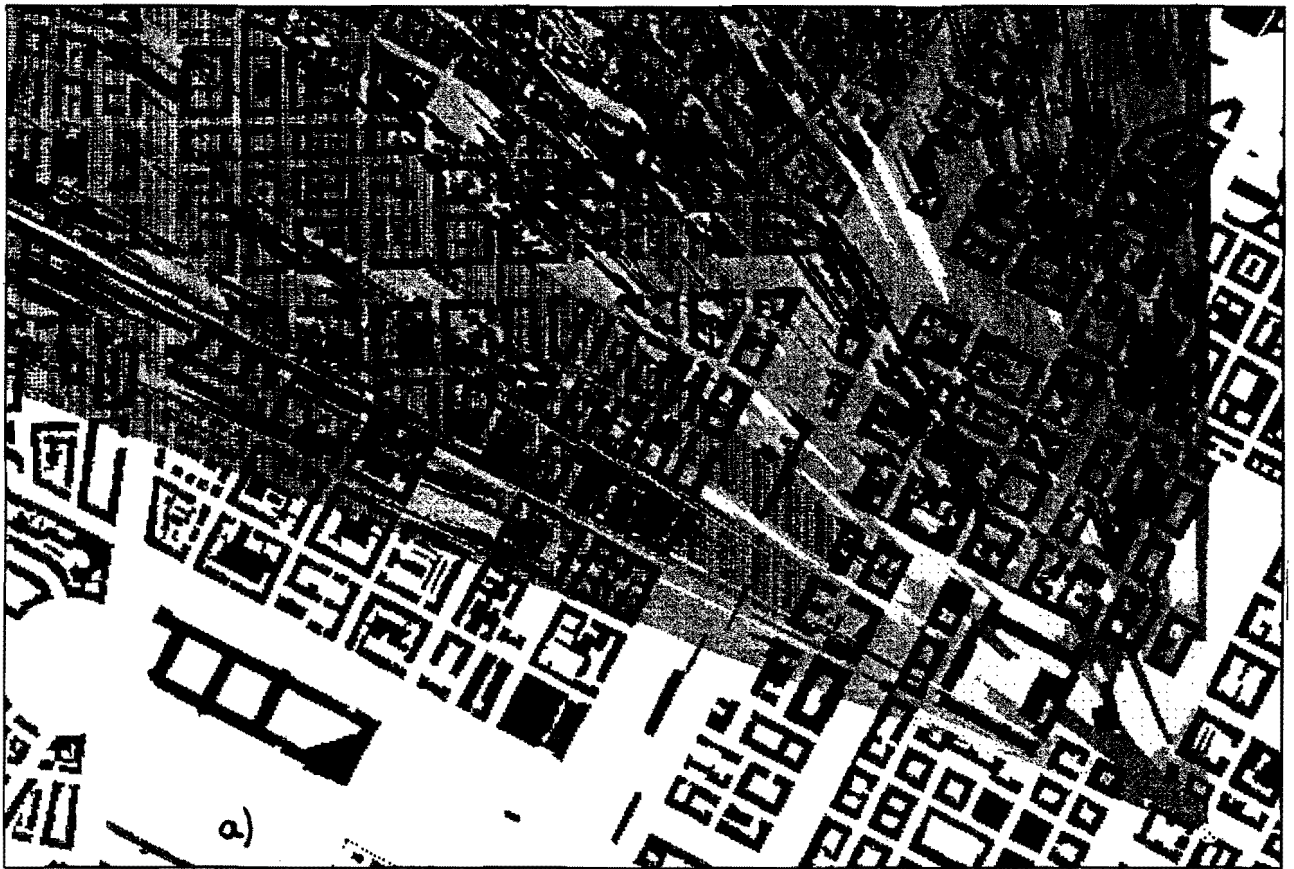
Area: t330   UTM: 394100 4991680

X axis [m]: 1000
Y axis [m]: 1200

EL. FIELD [dBuV/m]:   E = 80   E = 40
                                          E>= 30
                          E = 50   E< 30

Fig. 2: Field coverages. BS height: a)30m; b) 48m.

Fig.3: a) field coverage; b) measured values.

Area: t330

UTM: 394100 4391680

X axis [m]: 1800
Y axis [m]: 1200

Colour map:

ERROR [dBuV/m]:

E<-20
E<-10
E<= 0
E<= 10

E> 20

Fig.3: c) errors.

ON THE MODELLING OF ELECTROMAGNETIC WAVE SCATTERING BY A ROW OF CYLINDERS

H.J.F.G. Govaerts*, G.A.J. van Dooren**, and M.H.A.J. Herben

Eindhoven University of Technology, Faculty of Electrical Engineering P.O. Box 513,

5600 MB Eindhoven, The Netherlands, e-mail: M.H.A.J.Herben@ele.tue.nl

* now with Ericsson Telecommunicatie B.V., P.O. Box 8,

5120 AA Rijen, The Netherlands

** now with the European Space Agency ESTEC, P.O. Box 299,

2200 AG Noordwijk, The Netherlands

Abstract

The scattering of electromagnetic (EM) waves incident upon a row of metallic circular cylinders is analyzed. Experimental results are presented and compared with computer simulations based on the Uniform Theory of Diffraction (UTD). It is demonstrated that the single-cylinder UTD-model presented in [1] is an efficient tool to predict the fields scattered by a row of cylinders, e.g. street lamp-posts, and can therefore be used for the modelling of EM-wave propagation in urban environments.

1 Introduction

The present-day development of mobile-communication systems requires a reliable model to predict the EM-field strength distribution in urban environments. At Eindhoven University of Technology (EUT) the deterministic prediction tool *FiPre* (*Field Prediction*) is under development [2, 3]. The present version of *FiPre* uses the Uniform Theory of Diffraction (UTD) to model EM-wave diffraction at block-shaped obstacles. The advantage of this asymptotic technique for scattered-field computations is that it does not need large computer memory capacity and runtime, so that it can be implemented on a low-cost Personal Computer (PC).

The deterministic model *FiPre* is best suited for the design of Land Mobile Satellite (LMS) and micro-cellular terrestrial systems, because in those geometries only a limited number of objects surrounding the mobile are of importance in the determination of the field strength at the location of the mobile. To include the ability to accurately calculate the EM-wave scattering from cylinder structures, such as lamp-posts and chimneys, a three-dimensional UTD-based model accounting for the computation of EM-wave scattering by a straight circular cylinder was presented recently [1].

The presence of multiple cylinders on the radio wave propagation path gives also rise to multiply-reflected and multiply-diffracted field contributions, which will be referred to as 'interaction terms'. Elsherbeni and Kishk

presented a method by which the scattered field from multiple parallel cylinders can be calculated including the interaction terms [4]. They applied the boundary value method with plane-wave illumination, and found that for large separations between the cylinders the interaction terms can be neglected.

Our UTD-based method also demonstrates that the amplitude of these terms decreases with increasing separation of the cylinders [1]. This is due to the diverging character of a wave after being reflected or diffracted by the convex surface of a cylinder. Thus, it seems possible that these terms may be neglected for our specific application.

Since accounting for these interaction terms imposes higher complexity and longer runtime of the field strength prediction tool, it is the aim of this paper to get an impression of the applicability of the efficient single-cylinder UTD-algorithm to a multiple-cylinder configuration which frequently occurs in urban environments, i.e. a row of lamp-posts. Therefore, results of some well-defined experiments will be presented and they will be compared with predictions obtained from UTD-based computer simulations.

2   Experimental setup

At EUT we have the ability to perform indoor bi-static scattering measurements at 50 GHz for the determination of the shielding properties of various obstacles. A schematic diagram and the complete description of the measurement setup can be found in [5]. The information necessary to understand the experimental results to be presented later on is shown in Fig-

ure 1. An array of four equidistantly spaced finite-length cylinders was mounted on a rotating table along with the transmitter probe at T. The following non-variable parameters were measured with an accuracy of ± 0.5 mm: $d_T =$ 452 mm, $d_{TR} =$ 699 mm, $h_T =$ 1498 mm, $h_R =$ 1443 mm, and $d_c =$ 178 mm. The angle span of the receiver probe at P' was taken $\psi \in [-12°, 12°]$, with an angular step of 0.1° and an accuracy of $\pm 0.05°$. The secondary effect of diffraction at the ends of the cylinders was considered not to be important.

The 1.5 GHz mobile-communication frequency was scaled to the 50 GHz frequency used in the experiments. The cylinders used were made of brass and all have identical radii of 2.5 mm, corresponding to 8.3 cm at 1.5 GHz, which is near the radius of commonly used lamp- posts. The separation between the cylinder axes of symmetry was chosen to be 40 mm (approximately 7 wavelengths), which corresponds to 1.3 m at 1.5 GHz. Note that this is much worse than in practical situations, since the separation of lamp-posts is generally substantially larger.

3   Experimental results

The measurements were carried out as a function of the azimuth angle ψ, for various orientations of the array, indicated by the angle in Figure 1. Vertical polarization was used because it was demonstrated in [1] that then the scattered field components are the largest. The received-power results relative to the free-space power are depicted in Figure 2 by the solid curves. The dashed curves are obtained with UTD by simply adding the

individual scattered fields from each cylinder, and thus excluding the interaction terms. Except for the symmetric situation with $\kappa = 0°$, the asymmetry of the curves is clearly visible. For angles κ up to 60°, the UTD-simulation curves correspond quite well with the measured curves. At $\kappa = 80°$, the curves also follow similar courses with coinciding minima and maxima, but large differences in power level are present in the deep shadow region. These are readily explained by the fact that the interaction terms become more significant as the angle between the propagation direction of the incident wave and the longitudinal direction of the linear array decreases.

The measurement with $\kappa = 80°$ was repeated once, without touching any part of the measurement setup. The second measurement result, represented by the dash-dotted curve in Figure 2(e), only slightly differs from the first measurement. This demonstrates clearly the accuracy of the measured results.

4   Conclusions

The single-cylinder UTD-algorithm presented in [1] can also be applied in configurations involving multiple cylinders, provided that the separation between the cylinders is sufficiently large. It was demonstrated that for a separation of 7 wavelengths well-acceptable predictions can be obtained with the single-cylinder approach. Since the separation in a row of lamp-posts or chimneys generally exceeds 7 wavelengths, the single-cylinder UTD-method suffices for the modelling of EM-wave propagation in urban environments, although the scarcely occurring situation with a nearly longitudinally incident wave should be approached very carefully.

References

[1] H.J.F.G. Govaerts, G.A.J. van Dooren, and M.H.A.J. Herben, "Theoretical and experimental validation of UTD applied to electromagnetic-wave scattering by circular cylinders", *International Journal of Infrared and Millimeter Waves*, vol. 16, no. 6, pp. 1065–1081, 1995.

[2] G.A.J. van Dooren and M. Sforza, "FiPre: A prediction tool for the planning of mobile and fixed satellite communication services", *Proc. 1st IEEE Symp. on Comm. and Vehicular Techn.*, Delft, pp. 6.6.1–6.6.8, Benelux Chapter on Communications and Vehicular Technology, 1993.

[3] G.A.J. van Dooren, *"A deterministic approach to the modelling of electromagnetic wave propagation in urban environments"*, Ph.D. thesis, Eindhoven University of Technology, 1994.

[4] A.Z. Elsherbeni and A.A. Kishk, "Modeling of cylindrical objects by circular dielectric and conducting cylinders", *IEEE Trans. Ant. Prop.*, vol. AP-40, no. 1, pp. 96–99, 1992.

[5] G.A.J. van Dooren, M.G.J.J. Klaassen, and M.H.A.J. Herben, "Large bandwidth diffraction measurements at 54 GHz using both time-domain filtering and frequency smoothing", *Proc. 14th AMTA symposium*, Columbus, Ohio, vol. 1, pp. 13.9–13.14, 1992.
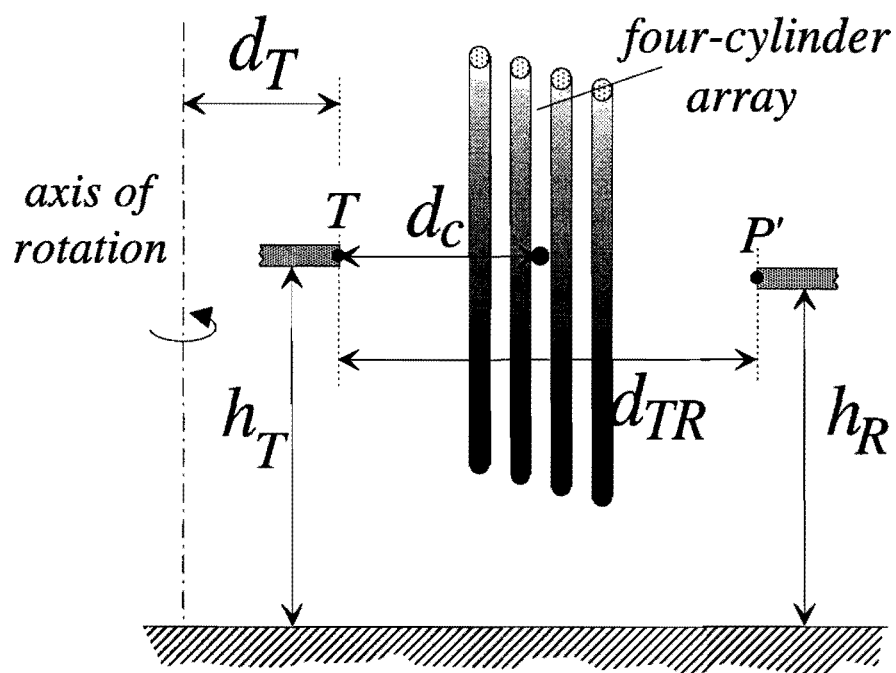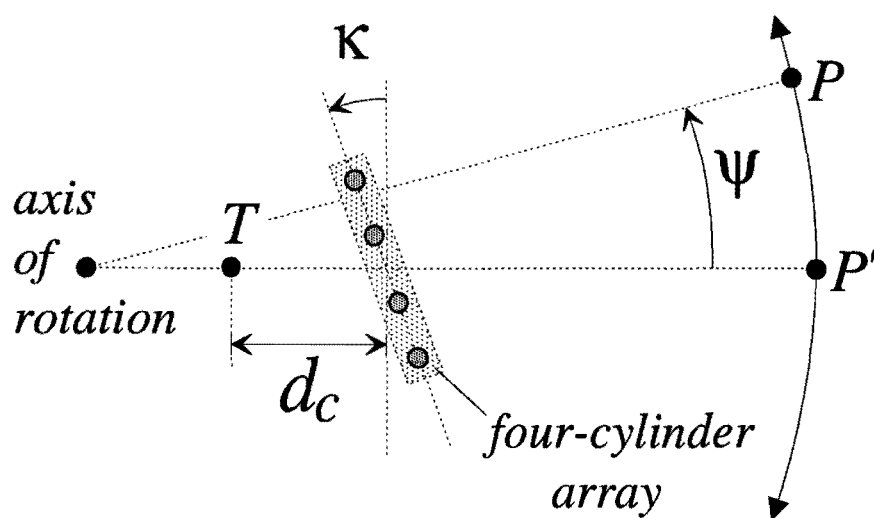
Figure 1: (a) Side view of the measurement setup; (b) Schematic top view of the corresponding model.
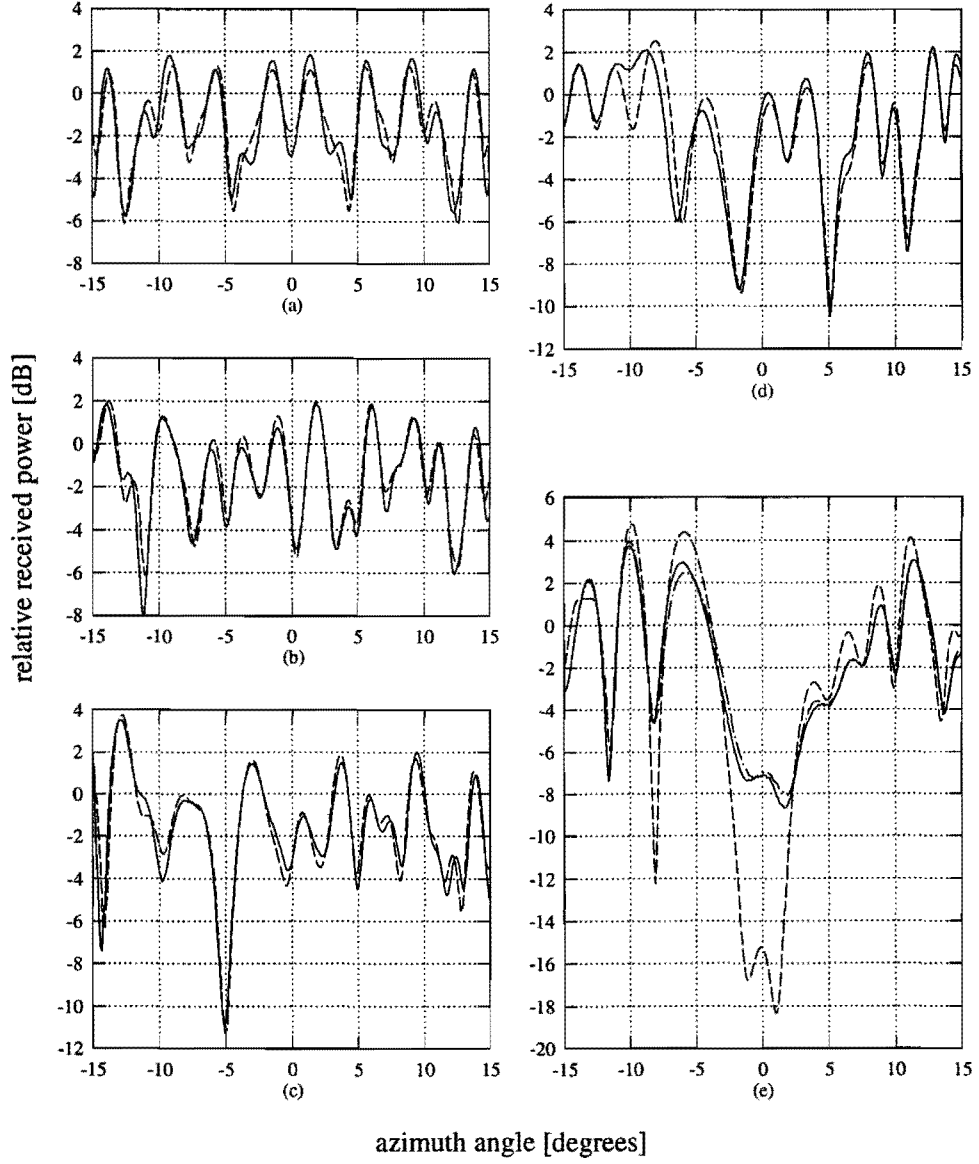
Figure 2: *Total received power relative to the free-space power as a function of the azimuth angle ψ for four parallel cylinders at (a) κ = 0°, (b) κ = 20°, (c) κ = 40°, (d) κ = 60°, and (e) κ = 80°. ————/— · — · —measurements; — — — — UTD-simulations.*

Performance of Correlated Diversity techniques on Rayleigh fading channels

Pieter Kempeneers   Marc Moeneclaey

Commun. Engineering Lab. University of Ghent

Sint–Pietersnieuwstraat 41 B–9000 GENT, BELGIUM

Fax: +32 9 2644295

E–mail: Pieter.Kempeneers@lci.rug.ac.be, Marc.Moeneclaey@lci.rug.ac.be

Abstract

Rayleigh fading channels are encountered in various applications, such as mobile radio, indoor radio and HF communications [1]–[3]. Several diversity techniques can be used in order to improve performance. Usually, the BER performance of a receiver using diversity combining is analyzed under the assumption of uncorrelated diversity signals. In this contribution we investigate the effect of correlation between the diversity signals. We show that correlation increases the BER as compared to the case of uncorrelated diversity signals.

1 Introduction

Some commonly used diversity techniques for fading channels are multipath diversity, antenna diversity, frequency diversity and time diversity.

The multipath fading channel is often modeled as a tapped delay line with uncorrelated tap gains. Using DS/SS communication over such a channel allows to resolve about T_d/T_c paths, where T_d and T_c denote the delay spread of the channel and the chip duration, respectively. Appropriate combining of the resolved paths in a RAKE receiver provides diversity [4],[5]. However, when the chip rate is increased, the relative delay between resolvable paths decreases so that the paths are likely to be correlated.

With antenna diversity, different signals are captured by the various antennas. Those that are lined up near to each other do not receive uncorrelated information: it is likely that signals captured by two nearby antennas are affected in a similar way by the channel.

When the frequency spacing between the subsequent carriers of a frequency diversity system is smaller than the coherence bandwidth of the channel, will the diversity signals will be correlated.

The time spacing of a time diversity system must be larger then the coherence time of the channel, otherwise no independent information can be obtained from the separated time signals.

In section 2 the BER of the correlated diversity receiver is analyzed. Numerical results are provided in section 3. Conclusions are drawn in section 4.

2 Performance analysis of the correlated diversity receiver

We assume that each diversity signal is subject to flat Rayleigh fading. We consider BPSK transmission, but the results remain valid for for QPSK symbols. Using a diversity technique of "order" D, the D received diversity signals can be represented in general as:

$$r_{1,k} = \sqrt{E_b}a_k c_{1,k} + n_{1,k}$$
$$\vdots \qquad\qquad\qquad (1)$$
$$r_{D,k} = \sqrt{E_b}a_k c_{D,k} + n_{D,k},$$

where E_b denotes the bit energy, a_k is the k-th transmitted symbol and $c_{l,k}$ is the instantaneous complex-valued gain of the l-th diversity signal; $n_{l,k}$ are the white Gaussian noise terms for each diversity signal.

We will make no limitations on the number of diversity signals. The only restriction that is made on the complex-valued gains of the diversity signals is that they are jointly Gaussian with zero mean (i.e. Rayleigh fading), and that the average fading gain equals 1, i.e.

$$\sum_{l=1}^{D} E[\mid c_{l,k} \mid^2] = 1. \qquad (2)$$

A Maximum Likelihood decoder should choose $\hat{a}_k$ to minimize

$$\sum_{l=1}^{D} |r_{l,k} - \sqrt{E_b}\hat{a}_k c_{l,k}|^2 \qquad (3)$$

For given values of the channel gains, the conditional bit error rate for the k-th symbol is then given by the expression:

$$\mathrm{BER}(\underline{c}) = Q(\sqrt{\frac{2E_b}{N_0}}\sqrt{\sum_{l=1}^{D}|c_{l,k}|^2}), \qquad (4)$$

where $\underline{c}$ is the $(D\times1)$ gain vector with components $c_{l,k}$. We can write $\mathrm{BER}(\underline{c})$ as:

$$\mathrm{BER}(\underline{c}) = Q(\sqrt{\frac{2E_b}{N_0}}\sqrt{\underline{c}^T\underline{c}^*}) \qquad (5)$$

If we use the bound $Q(x) \le \frac{1}{2}\exp{(-x^2/2)}$, BER is upper bounded by:

$$\mathrm{BER}(\underline{c}) \le \frac{1}{2}\exp(-\frac{E_b}{N_0}\underline{c}^T\underline{c}^*) \qquad (6)$$

The real and imaginary parts $\underline{c}_R$ and $\underline{c}_I$ of $\underline{c}$ are identically and independently distributed. We

therefore can split the exponential factor in the upper bound and write

$$\mathrm{BER}(\underline{c}) \le \frac{1}{2}\exp(-\frac{E_b}{N_0}\underline{c}_R^T\underline{c}_R + \underline{c}_I^T\underline{c}_I) \qquad (7)$$

Averaging $\mathrm{BER}(\underline{c})$ over all values of $\underline{c}$, we obtain:

$$\mathrm{BER} = E[\mathrm{BER}(\underline{c})] \le \frac{1}{2}\frac{1}{\det[I + \frac{2E_b}{N_0}R]}, \qquad (8)$$

where R is the covariance matrix of the vectors $\underline{c}_R$ or $\underline{c}_I$.

3 Numerical results

In order to obtain numerical results, we consider the covariance matrix R, given by

$$R = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{D-1} \\ \rho & 1 & \rho & \cdots & \rho^{D-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{D-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{D-1} & \rho^{D-2} & \rho^{D-3} & \cdots & 1 \end{bmatrix} \qquad (9)$$

where ρ is the correlation factor and $\sigma^2 = E[c_{R,l,k}^2] = E[c_{I,l,k}^2] = 1/(2D)$. With $\rho < 1$, "near[1]" diversity signals will be more correlated than those which are more separated from each other. This is intuitively pleasing and justifies more or less our choice for this specific covariance matrix.

No correlation

When $\rho = 0$, the diversity signals are uncorrelated, and we obtain,

$$\mathrm{BER} \le \frac{1}{2}(1 + \frac{E_b}{N_0}\frac{1}{D})^{-D}. \qquad (10)$$

When the number of diversity signals grows to infinity, the BER approaches the Gaussian case without fading:

$$\mathrm{BER} \le \frac{1}{2}e^{-E_b/N_0} \qquad (11)$$

---

[1]"near" should be interpreted in a wider sense: it could be in time, space or frequency, depending on the diversity technique

Identical diversity signals

When $\rho = 1$, all $c_{l,k}$ are identical, and we obtain, independently of D,

$$\text{BER} \leq \frac{1}{2}(1 + \frac{E_b}{N_0})^{-1} \qquad (12)$$

which is the same expression as for a single diversity signal. Hence, for $\rho = 1$ diversity does not improve the performance as compared to no diversity.

Asymptotic BER

When $E_b/N_0 \gg 1$, and R is nonsingular,

$$\text{BER} \approx \frac{1}{2}(\frac{N_0}{2E_b})^D \frac{1}{det(R)}. \qquad (13)$$

The determinant of the covariance matrix R from (9) is given by

$$det(R) = \sigma^{2D}(1 - \rho^2)^{D-1}. \qquad (14)$$

Finally, we obtain for the asymptotic BER:

$$\text{BER} \approx \frac{1}{2}(\frac{E_b}{N_0}\frac{1}{D})^{-D}(1 - \rho^2)^{1-D}. \qquad (15)$$

From (13) we observe that the asymptotic BER is proportional to $(E_b/N_0)^{-D}$, irrespective of the covariance matrix R (provided that R is nonsingular). Correlation only affects the constant of proportionality. When the covariance matrix is given by (9), we see from (15) that the asymptotic BER increases by a factor $(1-\rho^2)^{1-D}$ as compared to the case of uncorrelated diversity signals. The asymptotic correlation loss in E_b/N_0, defined as the increase in E_b/N_0 (in dB) needed to maintain the same asymptotic BER as for uncorrelated diversity signals, is given by

$$L(\rho, D) = -10\frac{D-1}{D}\log(1 - \rho^2) \, [dB]. \qquad (16)$$

For $D \gg 1$, $L(\rho, D)$ approaches $-10\log(1 - \rho^2)$.

Assuming that R is given by (9), the BER expression (8) is evaluated numerically for various values of ρ and D, and the results are shown in figures 1–3. For $\rho = 1$, the BER is independent of D, and, for large E_b/N_0, proportional to $(E_b/N_0)^{-1}$. For $\rho \neq 1$, the use of diversity (i.e. $D > 1$) yields a noticeable performance gain as compared to the case $D = 1$, even for correlation as large as $\rho = 0.99$.

The asymptotic correlation loss $L(\rho, D)$ from (16) is shown in figure 4. For large D, the loss corresponding to $\rho = 0.9$ and $\rho = 0.99$ amounts to 7.2 dB and 17 dB, respectively.

4   Conclusions

In this contribution we have investigated the effect of correlation on the BER performance of a receiver that uses diversity combining of Rayleigh fading signals.

When the covariance matrix of the fading gains of the D diversity signals is nonsingular, the asymptotic BER is proportional to $(E_b/N_0)^{-D}$. The correlation matrix affects only the constant of proportionality. The presence of correlation increases the BER as compared to uncorrelated diversity signals. When the D diversity signals have identical fading gains (i.e. singular covariance matrix), the BER is independent of the number D of diversity signals, and is for large E_b/N_0 proportional to $(E_b/N_0)^{-1}$. Hence, even for highly correlated (but non identical) fading gains, it is advantageous to use diversity combining.

Aknowledgement

References

[1] Lee, W. C. Y. "Mobile communications enge-

neering". Published by Mc Graw Hill, 1982.

[2] A. S. Acampora and J. N. Winters, "A wireless network for wideband indoor communications," IEEE J. Select. Areas Commun., vol. SAC-5, pp. 796-805, June, 1987

[3] Clark C. Watterson, J. R. Juroshek, and W. D. Bensema: "Experimental Confirmation of an HF Channel Model", IEEE Trans. on Commun. Techn., vol. COM-18, no. 6, December 1970.

[4] Proakis, J. G.: "Digital Communications (2nd edition)", McGraw-Hill, New York, 1989.

[5] Wuyts, K. and Moeneclaey, M.: "A reduced Complexity Rake Receiver with DFE for DS/SS Comm. over Multipath Fading Channels." Proceedings ICC '95, Seattle, pp 1359-1363 June 1995.

Figure 2: The BER for 5 diversity signals with $\rho = 0, \rho = 0.9, \rho = 0.99, \rho = 1$

Figure 1: The BER for 2 diversity signals with $\rho = 0, \rho = 0.9, \rho = 0.99, \rho = 1$
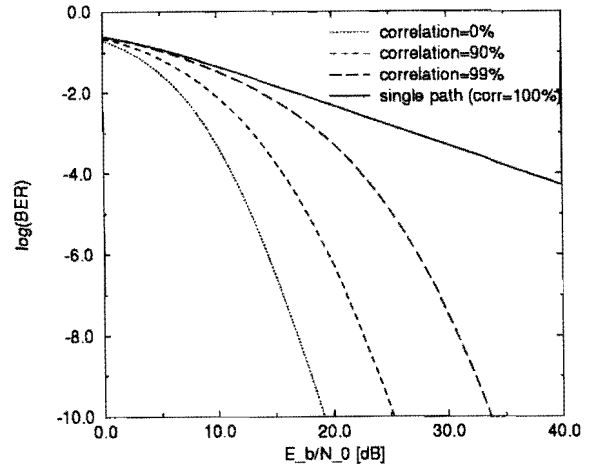
Figure 3: The BER for 10 diversity signals with $\rho = 0, \rho = 0.9, \rho = 0.99, \rho = 1$
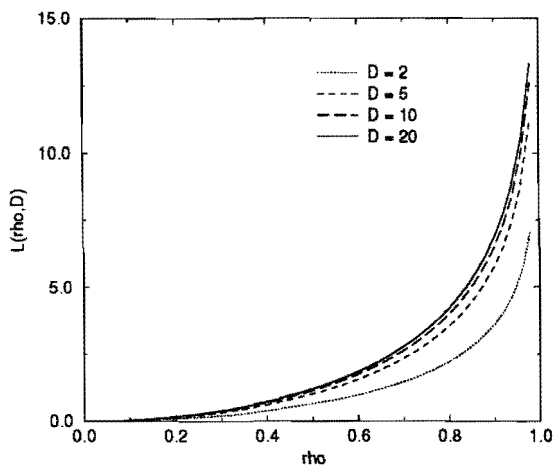
Figure 4: The correlation loss $L(\rho, D)$ for $D = 2, 5, 10, 20$

Fractionally spaced linear joint detection for Multitone CDMA Systems

O. van de Wiel[*] and L. Vandendorpe[†]
Communications and Remote Sensing Lab.
Université Catholique de Louvain
Place du Levant, 2 - Bâtiment Stévin
B- 1348 Louvain-La-Neuve
E-mail : vdd@tele.ucl.ac.be

Abstract

In this paper, the principles used in Fractionally Spaced Equalizers (FSE) are extended for the joint detection of the signals produced by different users using multitone CDMA (MT-CDMA). An explicit solution is provided for the coefficients of the joint device optimized for an MMSE criterion. The performance of such detectors is then investigated by means of the error probability whose exact derivation is provided for the steady-state. The influence of the number of tones and the oversampling factor is analyzed for simple channels. The effect of timing errors is also investigated, as well as the convergence properties of the associated adaptive filters.

1 Introduction

The combination of multitone modulation with CDMA has been introduced in [5]. It has been shown that the larger symbol duration associated with multitone systems was favorable to combat the effect of large delay spreads. The main feature of this system is however that for a constant bandwidth, the ratio between the number of chips per symbol and the number of tones has to be constant. Hence when the number of tones increases, the number of chips per symbol does as well and longer spreading sequences can be used. Consequently, more users can potentially be accommodated because of the enlarged sequence set (e.g. Gold sequences).

When multitone CDMA signals are transmitted through a linear frequency selective channel, intersymbol and inter-carrier interferences arise. Moreover, when several users are transmitting in the same bandwidth, non-zero cross-correlations between spreading sequences produce multi-user interferences. These interferences limit the global performance, but linear adaptive filters can be used to build improved decision variables [2, 4]. The steady state performance of MMSE driven linear and decision feedback (DF) equalizers for the purpose of minimizing the self-interference of a single user have been studied in [6, 9] respectively. In [8], the performance of different adaptive algorithms (LMS,RLS) able to cope with self-

interference has been studied. It was assumed that the filters were matched to both the symbols and the channel. Similarly linear devices performing joint detection for multitone CDMA systems have been proposed in [7]. Again it was assumed that the filters were matched to both the symbols and the channels. Channel matched filtering requires channel estimation and hence an additional computation power.

In the approach presented in the present paper, filtering matched to the symbols only is performed but the multi-user detector works at a higher rate compared to the symbol rate and hence a fractionally spaced joint detector is proposed. The continuous output of the matched filters is oversampled and a linear filter optimally combines the samples in order to produce the best possible estimates of the symbols that have been transmitted on all tones of all users. Such detectors have been investigated in [3]. Classical CDMA was considered and the performance was demonstrated for an AWGN channel by computer simulations and an LMS device. In the present paper we investigate the fractionally spaced joint detection (FSJD) for MT-CDMA. Moreover it is shown how to obtain the coefficients of the device by a set of linear equations. On the other hand it is explained how to compute the exact bit error probability in a steady state scenario. Finally the performance of FSJD for MT-CDMA is investigated for two-path asynchronous channels.

The paper is organized as follows : in section 2, the CDMA scenario, the modulation scheme and the receiver imperfections considered in this study are defined, together with the notations used throughout the paper. Section 3 is devoted to the concept of FSJD and the derivation of the coefficients, while the performance of FSJD for MT-CDMA signals is determined in section 4. The performance is measured by means of the error probability whose exact expression is computed thanks to an approach based on the characteristic functions. The computation results are presented in section 5.

2 Scenario and Notations

The transmitter of an MT-CDMA system is depicted in figure 1. It can be seen that the transmitted signal associated with a particular user is obtained by the modulation of N_t tones in parallel first and then spreading of the multitone signal by a user specific sequence having N_c chips per symbol duration. The constant bandwidth requirement means that the ratio between the number of tones and the number of chips per symbol has to be constant.

The scenario under consideration is asynchronous CDMA, with N_u users transmitting in the same bandwidth. Timing errors are taken into account in the receiver. It is also assumed that the shape $u(t)$ of the transmitted symbols is rectangular. Unless otherwise stated as a subscript, the duration of this pulse is the symbol duration, namely T.

Resulting from these assumptions, if we denote by $I_p^k(n)$ the complex symbol sent by user k on its p-th carrier at logical time n, by P the transmitting power, the signal emitted by user k can be expressed as :

$$x^k(t) = \sqrt{\frac{2P}{N_t}} \sum_{p=0}^{N_t-1} \sum_{n=-\infty}^{\infty} I_p^k(n) u_p^k(t-nT) \quad (1)$$

where the spreading pulse is, if T_c is the chip time $(N_c T_c = T)$,

$$u_p^k(t) = \sum_{i=0}^{N_c-1} a_i^k u_{T_c}(t-iT_c) \exp\left(j2\pi pt/T\right) \quad (2)$$

The most general case for a multi-user scenario is when the channels $c^k(t)$ of the different users are different. This corresponds to an up-link. The analysis of the down-link is easier, because the channels are all equal.

The signal received at the base station antenna is obtained by the summation of all the emitted signals arriving through their respective channels and unpredictable additive gaussian noise :

$$r(t) = \sum_{k=0}^{N_u-1} x^k(t) \otimes c^k(t) \otimes \delta(t-\tau^k) + n(t)$$

$$= \sum_{k=0}^{N_u-1} \sqrt{\frac{2P}{N_t}} \sum_{n=-\infty}^{\infty} \sum_{p=0}^{N_t-1} I_p^k(n) h_p^k(t-nT-\tau^k)$$
$$+n(t) \qquad (3)$$

where $h_p^k(t) = u_p^k(t) \otimes c^k(t)$ and $\otimes$ denotes convolution.

Because no channel estimation is performed, the only knowledge of the user signals which is available is the emitted pulse shape. We will assume that the time delay is imperfectly estimated, with an error ϵ_τ^k.

At the receiver end, the filters are matched to the emitted pulses corresponding to the different tones of the different users. We will compute the continuous output of this bank of filters, in order to provide an easy way to develop the oversampled version. The continuous output of the filter matched to carrier p of user k is, assuming all users have the same reception power (i.e. P is constant)

$$y_p^k(t) = \frac{r(t) \otimes \left[(u_p^k(-t))\right]^\star}{T\sqrt{2P}}$$
$$= \frac{1}{T\sqrt{N_t}} \sum_{l=0}^{N_u-1} \sum_{n=-\infty}^{\infty} \sum_{q=0}^{N_t-1} I_q^l(n)$$
$$w_{q,p}^{l,k}(t-nT-\tau^l) + \frac{\nu_p^k(t)}{T\sqrt{2P}}$$

where we have assumed the notation $w_{q,p}^{l,k}(t) = h_q^l(t) \otimes \left[u_p^k(-t)\right]^\star$ and $\nu_p^k(t) = n(t) \otimes \left[u_p^k(-t)\right]^\star$, $()^\star$ standing for complex conjugation.

These filter outputs will be sampled, taking into account the imperfectly estimated delay and the oversampling factor N with respect to the symbol rate, at discrete instants given by $t_m^k = m\frac{T}{N} + \tau^k + \epsilon_\tau^k$, so that the sampled values can be expressed as $z_p^k(m) = y_p^k(t_m^k)$, and

$$z_p^k(m) = \frac{1}{T\sqrt{N_t}} \sum_{l=0}^{N_u-1} \sum_{n=-\infty}^{\infty} \sum_{q=0}^{N_t-1} I_q^l(n)$$
$$w_{q,p}^{l,k}\left[(m-nN)T/N + \tau^k - \tau^l + \epsilon_\tau^k\right]$$
$$+ \frac{\nu_p^k\left(mT/N + \tau^k + \epsilon_\tau^k\right)}{T\sqrt{2P}} \qquad (4)$$

3 Fractionally spaced linear joint detection (FSJD)

The FSJD device makes a linear combination of the samples obtained by oversampling the outputs of all filters matched to all tones of all users. This is illustrated by Fig 2.

Considering the symbol emitted on carrier s of user j, at logical time m, a prediction of this symbol will be built in the following way :

$$\hat{I}_s^j(m) = \sum_{k=0}^{N_u-1} \sum_{p=0}^{N_t-1} \sum_{n_1=-K}^{K} c_{n_1}^{k,j,p,s} z_p^k(mN+n_1) \quad (5)$$

This expression gives rise to a total number of coefficients equal to $N_u^2 N_t^2(2K+1)$, e.g. $N_u N_t(2K+1)$ per filter.

If the design of the device is made for an MMSE criterion, the optimal coefficients are the solution of the problem :

$$\tilde{c}_{n_1}^{k,j,p,s} = \arg\min_c J_{MSE}$$
$$\text{with } J_{MSE} = \sum_{j=0}^{N_u-1} \sum_{s=0}^{N_t-1} \left[\frac{E\left|\epsilon_s^j(m)\right|^2}{N_u N_t}\right]$$
$$= \sum_{j=0}^{N_u-1} \sum_{s=0}^{N_t-1} \left[\frac{E\left|I_s^j(m) - \hat{I}_s^j(m)\right|^2}{N_u N_t}\right]$$

The orthogonality principle can be extended naturally here [1]. It is now expressed as the orthogonality between the prediction errors and the input samples involved in the prediction operation, as follows :

$$E\left[\epsilon_s^j(m) \left(z_p^k(mN+n_1)\right)^\star\right] = 0 \qquad (6)$$

This condition has to be verified for any j,k,s,p, and n_1, the expectation being taken on m. Equation (6) implicitly defines the set of equations that the optimal coefficients have to fulfill.

$$E\left[I_s^j(m) \left(z_p^k(mN+n_1)\right)^\star\right] =$$
$$\sum_{k'=0}^{N_u-1} \sum_{p'=0}^{N_t-1} \sum_{n_1'=-K}^{K} c_{n_1'}^{k',j,p',s} \qquad (7)$$
$$E\left[z_{p'}^{k'}(mN+n_1') \left(z_p^k(mN+n_1)\right)^\star\right]$$

With $\mu_{n_1,n_1'}^{k,k',p,p'} = \int_{-\infty}^{\infty} u_{p'}^{k'}(t-t_{n_1'}^{k'})\left[u_p^k(t-t_{n_1}^k)\right]^* dt$ and N_0 is the one-sided power spectral density of the AWGN, this leads to

$$\frac{1}{T\sqrt{N_t}}\left[w_{s,p}^{j,k}\left(n_1 T/N + \tau^k - \tau^j + \epsilon_\tau^k\right)\right]^*$$
$$= \sum_{k'=0}^{N_u-1}\sum_{p'=0}^{N_t-1}\sum_{n_1'=-K}^{K} c_{n_1'}^{k',j,p',s}\theta_{n_1,n_1'}^{k,k',p,p'} \qquad (8)$$

that has to be verified for every j,k,s,p and n_1. This is thus a concise representation of a linear systems consisting of $N_u^2 N_t^2 (2K+1)$ equations with as many unknowns.

4 Bit error Probability

The expression of the prediction of the symbol emitted at logical time m on carrier s of user j can be restated as

$$\hat{I}_s^j(m) = \sum_{l=0}^{N_u-1}\sum_{q=0}^{N_t-1}\sum_{n=-\infty}^{\infty} I_q^l(n) r_{m-n}^{l,j,q,s}$$
$$+\frac{1}{T\sqrt{2P}}\sum_{k=0}^{N_u-1}\sum_{p=0}^{N_t-1}\sum_{n_1=-K}^{K} c_{n_1}^{k,j,p,s}$$
$$\nu_p^k\left[(mN+n_1)T/N + \tau^k + \epsilon_\tau^k\right]$$

where $r_m^{l,j,q,s} = \frac{1}{T\sqrt{N_t}}\sum_{k=0}^{N_u-1}\sum_{p=0}^{N_t-1}\sum_{n_1=-K}^{K} c_{n_1}^{k,j,p,s}$
$w_{q,p}^{l,k}\left[(mN+n_1)T/N + \tau^k - \tau^l + \epsilon_\tau^k\right]$

On the basis of this expression, we intend to compute the bit error probability by means of an approach based on the characteristic functions. Assuming BPSK modulation, we have that

$$P_{e,j,s} = \Pr\left[\Re\left(\hat{I}_s^j(m)\right) \leq 0 | I_s^j(m) = 1\right] \qquad (9)$$

In order to achieve this computation, we have to notice that the covariance between two noise samples can be expressed as $E\left[\nu_p^k(n_1 T/N + \tau_k)\left(\nu_{p'}^{k'}(n_1' T/N + \tau_{k'})\right)^*\right] = 2N_0\mu_{n_1,n_1'}^{k,k',p,p'}$. Consequently, the noise term in equation (9) has a variance given by

$$\sigma_{N,j,s}^2 = \frac{N_0}{TN_t E_b}\sum_{k=0}^{N_u-1}\sum_{k'=0}^{N_u-1}\sum_{p=0}^{N_t-1}\sum_{p'=0}^{N_t-1}\sum_{n_1=-K}^{K}\sum_{n_1'=-K}^{K}$$
$$c_{n_1}^{k,j,p,s}\left(c_{n_1'}^{k',j,p',s}\right)^* \mu_{n_1,n_1'}^{k,k',p,p'}$$

The BPSK modulated symbols $I_s^j(m)$ take the values $+1$ and -1 with probability of $1/2$. Hence each interfering term, which is of the form $(x+\jmath y)I$, has a characteristic function given by $\psi(\omega_1,\omega_2) = \cos(x\omega_1 + y\omega_2)$.

Hence, the characteristic function of the estimated symbol is

$$\psi_{\hat{I}_m^j} = \exp\left(\jmath(\omega_1\Re(r_0^{j,j,s,s}) + \omega_2\Im(r_0^{j,j,s,s}))\right)$$
$$\times \exp\left(\frac{-(\omega_1^2+\omega_2^2)\sigma_{N,j,s}^2}{4}\right)$$
$$\times \prod_{k=0}^{N_u-1}\prod_{p=0}^{N_t-1}\prod_{n=-\infty}^{\infty}\cos\left(\omega_1\Re(r_{m-n}^{k,j,p,s}) + \omega_2\Im(r_{m-n}^{k,j,p,s})\right)$$

In the double product, the value for $k = j$ and $p = s$ and $m = n$ does not have to be taken into account, because its value is assumed. This characteristic function contains all the information needed to evaluate the error probability. Indeed, its inverse transformation gives the joint probability density function, the marginal with respect to the imaginary part of which has to be integrated from $-\infty$ to 0 to obtain the probability that the decision variable is negative. We thus have

$$P_{e,j,s} = \frac{1}{2} - \frac{1}{2\pi\jmath}\int_{-\infty}^{\infty}\frac{\psi_{\hat{I}_s^j(m)}(\omega_1,0)}{\omega_1}d\omega_1 \qquad (10)$$

Considering the expression of $\psi_{\hat{I}_s^j(m)}$, we have

$$P_{e,j,s} = \frac{1}{2} - \frac{1}{\pi}\int_0^{\infty}\frac{\sin\left(\omega_1\Re(r^{j,j,s,s})\right)}{\omega_1}$$
$$\times \exp\left(\frac{-\omega_1^2\sigma_{N,j,s}^2}{4}\right)$$
$$\times \prod_{k=0}^{N_u-1}\prod_{p=0}^{N_t-1}\prod_{n=-\infty}^{\infty}\cos\left[\omega_1\Re(r_{m-n}^{k,j,p,s})\right]d\omega_1$$

The same remark concerning the term to be forgotten in the double product is valid here again.

5 Computation Results

The bandwidth expansion factor, sometimes called the processing gain in CDMA systems is taken to be 7. Because the possible length of the chosen Gold sequences are only of the form $2^n - 1$, the constant bandwidth requirement can only be

roughly reached. Because of exponentially growing complexity, we are only considering low numbers of tones, namely 1,2,4 and 8, the corresponding number of chips per symbol being respectively 7,15,31 and 63.

A system with four users is considered. The scenario being asynchronous, we have taken the relative delays between user uniformly distributed on a single-tone symbol duration. The channels experienced by the different users are two-paths, with the power equally distributed between the paths. This particular choice allows to define the remaining unknowns in the model, namely the partiel function $w_{q,p}^{l,k}(t)$. The delays of the second path are fixed and can take values greater than the symbol duration. The channels are not time-varying.

Fig. 3 shows the performance of FSJD for a single carrier system, which is nothing but *usual* CDMA. The average BER among all users is provided. We can clearly see that the fractionally spaced approach is effective, the positive effect of an increased oversampling factor, and the slight saturation for large values of this factor. The filter length, expressed in symbol duration span, is kept constant. When the oversampling factor goes up, the improvement is due to the fact that the FSJD device can render the effect of channel matched filtering with more accuracy. The performance achieved over an AWGN channel is also given.

The performance of a symbol spaced JD, particular case of FSJD when $N = 1$ is presented in figure 4, the BERs being averaged over all tones of all users. The *multitone* effect, i.e. the positive effect of an enlarged number of tones is illustrated. After FSJD filtering, the results are comparable when the absolute time span of the filters is constant ($N_t = 2, K = 1$ and $N_t = 1, K = 2$) and the BER is improved when this span is increased ($N_t = 4, K = 1$).

Fig. 5 presents the performance of a system using 2 carriers, for different oversampling factors. Here again, the positive effect of an increased oversampling factor is demonstrated.

Fig. 6 shows the performance for MT-CDMA systems with 4. The same effect of an increased oversampling factor is present. It is worth mentioning that a pure interference canceler ($K = 0$) already shows a performance improvement.

The influence of the timing error has also been studied. Timing errors between 0 and one tenth of the symbol duration for a single carrier system have been considered. The sensitivity of FSJD to timing errors has been investigated for different configurations. Fig. 7 gives the BER as a function of the timing error for $E_b/N_0 = 9dB$. From this figure, it is clear that the sensitivity to timing error decreases as the oversampling factor of the FSIC increases, independently of the number of carriers considered. Furthermore, the case without FSJD shows that the system becomes rapidly unusable, even for a small timing error.

The BER performance studied in this paper is the steady state performance. From the spectrum of the correlation matrix used for the optimal filter coefficient computation, it is possible [10] to derive an estimate of the time constant that would be required by the adaptive process used to implement the optimal filtering process. This estimate is given, in term of symbol duration, by

$$\tau = -\frac{0.5}{\ln\left(1 - \frac{2}{\chi+1}\right)} \quad (11)$$

$$\chi = \frac{\lambda_{max}}{\lambda_{min}} \quad (12)$$

and the results are gathered in fig. 8. The time constants have been corrected by a factor of N_t, because the estimate is expressed in term of symbol duration, and the multitone modulation has the property of increasing the symbol duration. This results in an increased convergence phase. If we had considered the problem in term of symbol duration, the results would have been inverted, with respect to the number of tones. Anyway, a potential problem appears from this graph, the time constant being an increasing function of the oversampling factor, independently of the number of carriers considered.

6 Conclusions

In the present paper, fractionally spaced linear joint detection has been investigated for MT-CDMA systems. The coefficients of the device have been obtained in close form for an MMSE criterion. The performance of the device in steady-state conditions has been investigated. It has been shown how an exact expression of the BER could be obtained. The effectiveness of the device has been demonstrated for two-paths channels.

The advantage of FSJD is clearly that no explicit channel estimation has to be performed. Hence it can be concluded that MT-CDMA with FSJD reception is an attractive solution. As a matter of fact, MT-CDMA benefits from the feature that when a larger number of tones are used, longer codes can be used for the same bandwidth and hence more users can be accommodated. FSJD reception ensures that the level of performance will be sufficient.

References

[1] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall International, 1991.

[2] A. Kajiwara and M. Nakagawa, "Microcellular CDMA System with a Linear Multiuser Interference Canceler", *IEEE Journal on Selected Areas in Communications*, Vol. 12, No. 4, May 1994, pp. 605-611.

[3] P. N. Monogioudis, R. Tafazolli and B. G. Evans, "Linear adaptive fractionally spaced equalization of CDMA multiple access interference", *Electronics Letters*, vol. 29, No. 21, October 1993, pp. 1823-1824.

[4] M. Rupf, F. Tarkoy and J.L. Massey, "User-Separating Demodulation for Code Division Multiple Access Systems", *IEEE Journal on Selected Areas in Communications*, Vol. 12, No. 5, June 1994, pp. 786-795.

[5] L. Vandendorpe, "Multitone Spread Spectrum Multiple Access Communications System in a Multipath Rician Fading Channel", *IEEE Transactions on Vehicular Technology*, Vol. 44, No. 2, May 1995, pp. 327-337.

[6] L. Vandendorpe and O. van de Wiel, "Analysis of residual interference after MSE Linear equalization of Multitone spread spectrum signals", *IEEE International Conference on Communications*, Seattle, June 1995, pp. 1673-1677.

[7] L. Vandendorpe and O. Van de Wiel, "Performance Analysis of Linear Joint Equalization and Multiple Access Cancelation for Multitone CDMA", Accepted for publication in *Wireless Personal Communications*.

[8] O. van de Wiel and L. Vandendorpe, "A Comparison between RLS and LMS linear equalization for OFDM/DS transmission in an indoor environment", *IEEE Third Communication Theory Mini-Conference in conjunction with GLOBECOM'94*, San Francisco, November 1994, pp. 71-75.

[9] O. van de Wiel and L. Vandendorpe, "Analysis of residual interference after MSE DFE equalization of multitone spread spectrum signals", *IEEE Vehicular Technology Conference*, Chicago, July 1995, pp. 394-397.

[10] O. Macchi, "Adaptive Processing, The Least Mean Squares Approach with Applications in Transmission", *John Wiley and Sons*, Chichester, England, 1995.
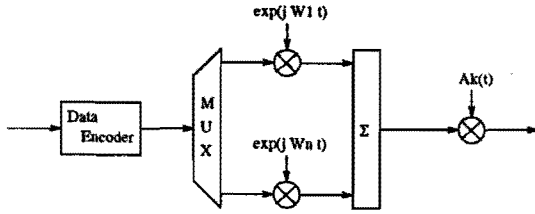
Figure 1: MT-CDMA Transmitter
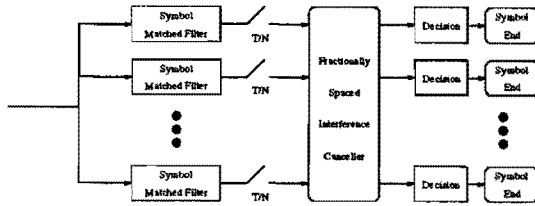
Figure 2: FSJD receiver structure. The number of branches is equal to $N_u N_t$

Figure 4: Symbol spaced (FS)JD, for N_t equal to 1,2,4 and 8, without and with JD. K=1 for all cases. (dotted line is AWGN performance)

Figure 3: FSJD, single carrier, different oversampling factors N (1,2,4,8), constant span filter, dotted line being AWGN reference channel
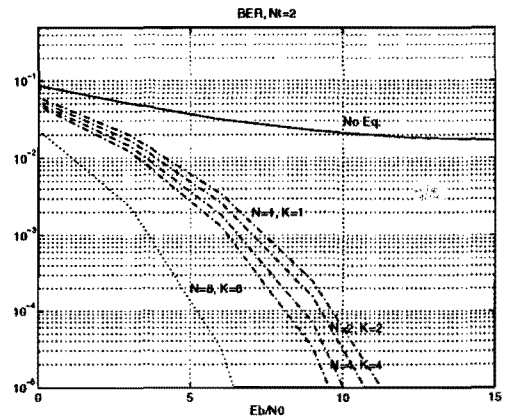
Figure 5: BER for Nt=2, without and with FSJD, for various (N,K), dotted line is AWGN performance
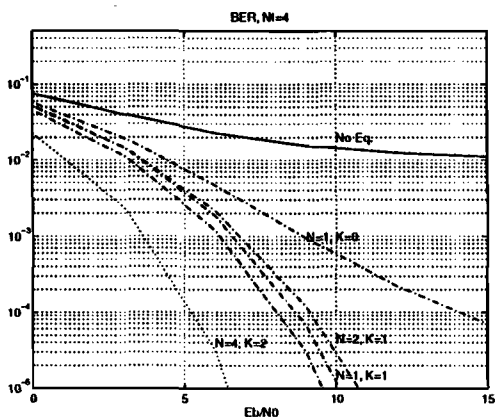
Figure 6: BER for Nt=4, without and with FSJD, for various (N,K), dotted line is AWGN performance.
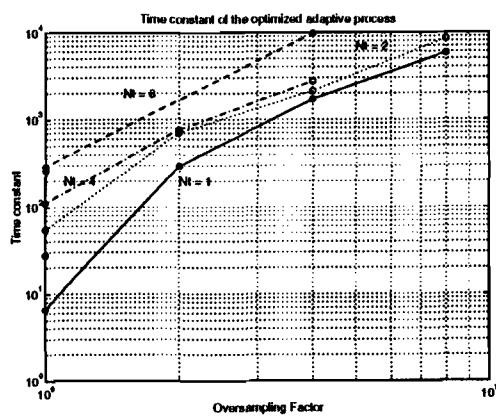
Figure 8: Time constant of an adaptive process implementing FSJD, for different numbers of carriers
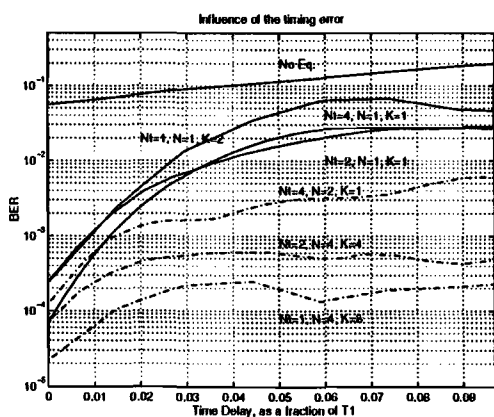
Figure 7: BER for a timing error ranging from 0 to T1/10, for a system with no FSJD, and for 1 and 2 tones systems

59

Theoretical Analysis and Simulation of Burst-and-Random-Error-Prone Direct Sequence CDMA Channels.

Mqhele E. Dlodlo, *Member, IEEE* and Shinsuke Hara, *Member, IEEE.*

Telecommunication and Traffic-Control Systems Group
Delft University of Technology
P.O. Box 5031, 2600 GA Delft, The Netherlands
fax: 015-281774
email: M.E.H.Dlodlo@et.tudelft.nl
S.Hara@et.tudelft.nl

Abstract: **For those channels dominated by bursty interference, in which long burst-error correcting codes would, however, be expensive to implement, the designer can interleave a short random-error correcting code. Alternatively, a channel-state-sensing concatenated coding scheme can be used.**

In this paper we analytically estimate the duration of the burst so as to design the interleaving degree. Code interleaving is applied to plain burst-error-correcting codes and the performance compared with that of channel-state-sensing concatenated coding using separate decoding. The results are confirmed using Monte Carlo simulation.

It is found that channel-state-sensing concatenated coding outperforms code interleaving in terms of both throughput and average delay.

I. Introduction

In a land-mobile satellite environment where a number of systems are competing for the same bandwidth, problems with the additive white Gaussian noise can sometimes be overshadowed by bursty interference from both inter- and intra-system sources. Burst-error correcting codes such as Burton, Fire, Reed-Solomon and Product codes can effectively combat this kind of interference. Their shortcoming, however, is that they are inefficient under random-error conditions, such as on an open highway through the savanna (with its sparse foliage). In this experiment we study the concept of estimating the duration of a fade so as to design the interleaving degree of a code interleaver. We then contrast the technique with a concatenated coding scheme applied to the same block code as that used in interleaving. In addition, we study the effect of interleaving a concatenated code.

The techniques under study are not in themselves new. They have been used successfully in other fading environments but to the best of the authors' knowledge, their performance still needs to be compared. This study is particularly focused on the land-mobile satellite (LMS) channels. What follows in this section is a brief overview of each technique, and then an outline of the paper.

One of the main problems for the designer of an interleaver is to estimate the duration of the burst, should it occur, in order to design an adequate interleaving degree. In direct sequence CDMA, code interleaving as a method of making the successive symbols span a large enough interval to encompass independent fades is practicable. It is most suited to a rapidly moving vehicle such as on a tarred rural highway if the interleaving degree is to be large enough [1]. The interleaving degree is significantly influenced by the expected burst length. Fade statistics are analytically derived as in [2] and used to design the interleaving degree. Interleaving to combat the bursts due to shadowing and interference is combined with a short random-error correcting code to produce results comparable to those expected from a long code.

The other technique under study is concatenated coding which matches two shorter codes to produce a longer equivalent with a lower code rate and a significantly higher error-correcting capability. In concatenated coding, a moderately efficient random-error correcting code is used as the inner code and the outer code is a Reed-Solomon code, which is burst-error correcting. The key design consideration here is the matching of the two codes in such a way that the symbols of the inner code fit in a precise way into those of the outer code. A secondary consideration is that the inner code need not be so

efficient that the outer code is rendered inefficient under random-error conditions. As an additional measure, in this paper, we require the loop to sense the random-error state and suspend the outer code for the duration of the good state. A state-sensing concatenated code can be used to great effect with a variable-parity RS code. This latter aspect, however, is beyond the scope of the current analysis.

Section II briefly discusses the dervation of the interleaving degree and then outlines the three error-control strategies in a little more detail. Section III presents the relevant system model and performance measures for error control. The results and conclusions follow in sections IV and V, respectively.

II. Code Interleaving and Concatenated Coding

A. Code Interleaving

It is desired to design an interleaver with an encoding delay, τ_i, and storage capacities, S_i and S_d already minimised. The interleaving degree, λ, is determined both analytically and by use of a multi-tone model of the shadowed Ricean channel. The interleaver can be either convolutional or of block form for use with block codes [3]. In either case, a code of length n containing k_b information bits per code word is transformed by the interleaving degree from the (n, k_b) to the $(\lambda n, \lambda k_b)$ form. In other words, the original t_{fec}-error-correcting code becomes a t_{fec}-burst-correcting code [4]. That is, the code will now correct up to t_{fec} bursts of length λ or less. Block codes are preferred to convolutional codes when interleaving is used because convolutional codes require a fairly long tail of the message in order to converge [5]. The objective here is to scramble a short code and use it where a long one would have been needed. For success, λ must be long enough for the product, λn to completely engulf the length of the burst. The estimation of the value of the probability density function of the burst in real time is a daunting task. To get round the problem, we only estimate its statistics under fading conditions as done by [2]. The technique is well illustrated in that reference for the Gaussian spectrum. We highlight key aspects of it here with respect to the Doppler spectrum which is more appropriate for

the monopole antenna used by vehicle-mounted and hand-held transceivers.

The Level Crossing Rate (LCR)

For the channel model of the envelope

$$re^{j\theta} = ze^{j\phi_0} + we^{j\phi}, \quad z, w > 0 \qquad (1)$$

where the phases are uniformly distributed in $[0, 2\pi]$, z is lognormal and w, Rayleigh. These distributions belong to the Gaussian family of processes. The conditional probability density function of r given z is Ricean and the consequence of these observations leads to the now familiar shadowed Ricean distribution of land mobile satellite communication systems

$$p(r) = \frac{r}{b_0\sqrt{2\pi d_0}} \cdot \int_0^\infty \frac{I_0(rz/b_0)}{z} \cdot e^{-(\ln z - \mu)^2/2d_0} \cdot e^{-(r^2+z^2)/2b_0} dz \qquad (2)$$

where the received signal r is much smaller than the average multipath scatter power b_0. The line-of-sight component z is lognormally distributed due to shadowing, and its mean value and variance are μ_0 and d_0, respectively. The interest is in calculating the number of times per second that the envelope crosses the fading threshold R with either a positive or a negative slope, i.e. the level-crossing rate (LCR). The general formula for the LCR as proposed by Rice [6] is

$$N_R = \int_0^\infty \dot{r}\, p(R, \dot{r}) d\dot{r} \qquad (3)$$

where the rate is with respect to time. By assuming statistical independence between the envelope and its rate of change, as in [6], then one can solve for the multipath and the line-of-sight components separately in dr/dt. Finally the LCR is found as a function of the channel's fade statistics, i.e.

$$N_R = \frac{b_2\sqrt{1-\rho^2}}{\sqrt{2\pi}} \cdot \frac{\sqrt{b_2 + 2\rho\sqrt{b_2 d_2} + d_2}}{b_2(1-\rho^2) + 4\rho\sqrt{b_2 d_2}} \cdot p(R) \qquad (4)$$

where b_m and d_m are the moments of the multipath and the direct components respectively. The moments are given by

$$b_m = (2\pi)^m \int_0^\infty w(f)(f-f_c)^m df, \qquad (5)$$

$$m = 0, 1, 2, \cdots.$$

where f_c is the carrier frequency and the average multipath (scattered component) power is

$$b_0 = \int_0^\infty w(f)df = N_0.$$ The derivation is developed more fully in Appendix II of [6].

At this point one notes that the moments in (4) depend on the power density spectrum of the antenna used. Let us assume that the mobile unit uses a vertical omnidirectional monopole. We can safely further assume that a large number of signals with similar characteristics impinge on such a receiving antenna in the mobile system environment. The resulting Doppler power spectral density due to shadowing and fading takes the form

$$P(f) = \begin{cases} \dfrac{b_0}{\pi f_D} \cdot \dfrac{1}{\sqrt{1-(f/f_D)^2}}, & -f_D \le f-f_c \le f_D \\ 0 & \text{otherwise} \end{cases} \qquad (6)$$

where $f_D = (v/c)f_c$ is the maximum Doppler frequency spread and $f = f_c + (1/2\pi)(d\theta/dt)$ [2]. Thus, the vehicular speed, v, relative to the speed of light, c, determines the maximum Doppler spread. The moments b_2 and d_2 can then be determined from (6) and substituted for in (4) as follows:

$$b_2 = (2\pi)^2 \int_{-f_D}^{f_D} f^2 P(f)df \text{ and } d_2 = (2\pi)^2 \int_{-f_D}^{f_D} f^2 P(f)df.$$

Solving for b_2 first we find that

$$b_2 = (2\pi)^2 \cdot \frac{b_0 f_D}{\pi} \cdot \int_{-f_D}^{f_D} \frac{(f/f_D)^2}{\sqrt{1-(f/f_D)^2}} df \qquad (7)$$

$$= 2\pi b_0 f_D$$

Similarly for d_2 we get

$$d_2 = 2\pi d_0 f_D. \qquad (8)$$

Substituting (7) and (8) in (4) gives $p(\dot{r})$ in closed form such that the LCR becomes

$$N_{Rnorm} = \frac{b_0\sqrt{1-\rho^2}\sqrt{\pi\left(b_0+2\rho\sqrt{b_0 d_0}+d_0\right)}}{b_0(1-\rho^2)+4\rho\sqrt{b_0 d_0}} \cdot p(R) \qquad (9)$$

Equation (9) is already normalised with respect to the Doppler frequency so as to make it independent of the vehicle speed.

The AFD and the Interleaving Degree

In calculating the AFD, N_R is divided by the Doppler Frequency, a step which effectively makes $f_{Dnorm} = 1$ Hz for any given speed. If we note that all the other values in equation (9) are constants, then $N_R = f(f_D, p(r))|_{r=R}$ and $N_{Rnorm} = f(p(r))|_{r=R}$ only. A practical implementation, on the other hand, would require a consideration of the varying vehicle speed and the consequent adaptive control of the interleaving degree.

As in [7], using the normalised LCR, we express the average fade duration (AFD) in the form

$$T_N(f) = \frac{1}{N_{Rnorm}} \int_0^R p(r)dr. \qquad (10)$$

Based on these approximations, the minimum interleaving degree can then be conditioned on both the AFD and the packet duration, $N_p T_b$. In order to express the interleaving degree, λ, in terms of the code length, we use the relationship

$$\lambda \cdot n \cdot T_b = T_N(f) \qquad (11)$$

where T_b is the bit duration and n, the code length. Other interleaving design considerations can be found in [3].

B. Concatenated Coding

As shown in Figure 1, the outer code can be viewed as acting on a random-error superchannel inside which there is the inner code and possibly an interleaver as well. The design of concatenated coding is well described in [8]. In this design, we will use the Reed-Solomon code as the outer code assuming it can be easily adjusted to different message lengths and symbol sizes, and is well matched to other BCH codes with their efficient decoding algorithms in the bounded-distance decoding class. We match the codes by letting the length of the inner block code provide

the power, m in the Galois Field, GF(2^m), of the outer code. For ease of comparison with interleaving, we use a BCH inner code, even though a convolutional code is normally preferred in efficient concatenated coding schemes. Separate decoding will be performed to allow for flexibility in automatically controlling the characteristics of the outer code relative to those of the inner code with changing channel conditions. This is important for maintaining a high coding gain. The overall length of the concatenated and interleaved codes is $N_\lambda = \lambda mnN_b$ bits and the overall information content per word is $K_\lambda = \lambda mk_b K_b$. Thus, the effective code rate becomes $R_\lambda = k_b K_b / nN_b = r_b R_b$. In separate decoding, the inner decoder sees the (N_b, K_b) t_{fec}-error-correcting code, while the outer decoder handles the (n, k_b) t_{rs}-error-correcting code. The inner code has the characteristics of the primitive BCH codes as summarised in [9]. In the outer RS code, $n = 2^m - 1$ GF(2^m) symbols, with $n - k_b = 2t_{rs}$ parity-check symbols per code word. Its designed minimum distance is $d_{min} = n - k_b + 1$. Note that $m = N_b$, the code length of the inner code for proper matching.

III. System Model and Error Control with Validation

A. The System

The system model for the land-mobile satellite communication channel is widely discussed [2],[9],[10],[11],[12] and its highlights are presented here for ease of review. The system under consideration is assumed capable of handling up to K simultaneous users in a slotted direct-sequence code division multiple access scheme (s-DS/CDMA), at a bit rate $R_b = 1/T_b$, where T_b is the bit duration. The CDMA spreading code is approximately random with a chip duration T_c where the code length is $N = T_b/T_c$. The offered traffic rate, G, (in packets per slot) is Poisson-distributed for the mobile satellite channel such that the generation probability of the kth packet is

$$P_{tr}(k) = \frac{G^k}{k!} \exp(-G),$$

$$k = [0, K], \qquad K >> 1 \qquad (12)$$

as in [9]. The channel is represented by the shadowed Ricean distribution of equation (2) which degenerates to approximately the Rayleigh model under heavy shadowing. For spread spectrum, the scatter power in path m is given by

$$b_{0m} = b_0 \left(1 - \exp10\left(-T_c/T_s\right)\right).\exp10\left(-(m-1)\left(T_c/T_s\right)\right)$$

where T_s is the delay spread as depicted in [13] and $exp10(x)$ is the base-10 antilogarithm. The channel parameters are measured for each type of shadowed locality.

If each of the K users transmits the signal $s(t) = Aa_k(t)b_k(t)\cos(\omega_c t + \phi_k)$ then the received signal can be depicted as

$$r(t) = \sum_{k=1}^{K} \sum_{m=1}^{M} A\beta_{mk} \, a_k(t - \tau_{mk})b_k(t - \tau_{mk}).$$
$$\cos\left[(\omega_c + \omega_{mk})t + \phi_{mk}\right] + n(t) \qquad (13)$$

where m and k denote the path and user number, respectively. A is the transmitted signal amplitude, which is assumed to be constant and identical for all users. The assumption is straightforward for the downlink but requires on-board power control for
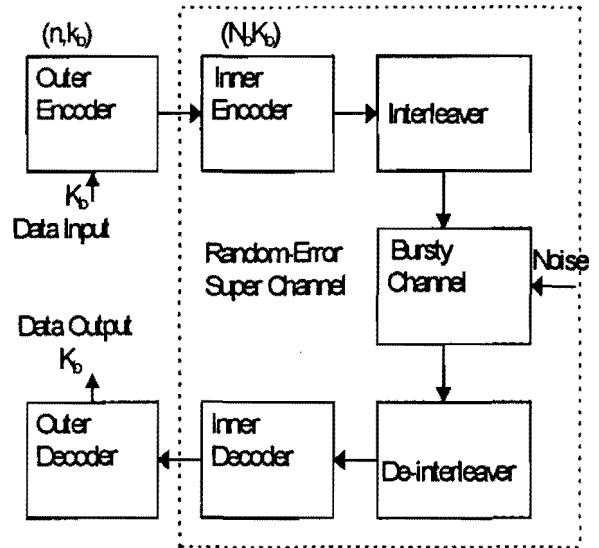
Figure 1. The concatenated interleaving scheme for burst-and-random-error containment.

the uplink. For user k, $\{a_k\}$ is the spread spectrum code, $\{b_k\}$ is the data sequence, $\omega_c + \omega_{mk}$ is the carrier plus Doppler angular frequency, ϕ_{mk} is the carrier phase, τ_{mk} is the time delay and $n(t)$ is white Gaussian noise with two-sided spectral density $N_0/2$. The instantaneous path amplitude gain is denoted as β_{mk}. The decision statistic from

the correlation operation under perfect synchronisation is

$$u = A\beta_{ij}b_i^0 + \sum_{k=1}^{K_{max}} I_k + \eta_i \qquad (14)$$

where η_i is a zero-mean Gaussian variable with variance N_0T_b, T_b is the bit duration, I_k denotes the cross-correlation products from interfering users and multipath signals [14], and b_i^0 is the data bit. This u is the signal seen by the baseband demodulator which then discriminates between the desired signal $A\beta_{ij}b_i^0$ and the cross-correlation products along with noise to the right.

B. Error-Control Measures

We measure the performance of the interleaved concatenated code using throughput as a function of offerd traffic and success probability

$$S = \frac{k_b}{n} \sum_{k=1}^{K_{max}} k p_{tr}(k) P_s(k) \qquad (15)$$

where the packet success probability is bounded by

$$P_s \le \frac{2^{(k_b-1)}}{2^{k_b}-1} \cdot \sum_{j=0}^{t_{rs}} \frac{j+t_{rs}}{n} \cdot \binom{n}{j} p_e^j (1-p_e)^{n-j} \qquad (16)$$

in the fast fading case [4] and

$$P_s \le \int_0^\infty \frac{2^{(k_b-1)}}{2^{k_b}-1} \cdot \sum_{j=0}^{t_{rs}} \frac{j+t_{rs}}{n} \cdot \binom{n}{j} p_e^j (1-p_e)^{n-j} \cdot p_{rss}(r) dr \qquad (17)$$

in slow fading, where n is the outer code length which we arbitrarily equate with the packet length[15][8]. The factor $2^{(k_b-1)}/2^{k_b}-1$ accounts for the average number of information bit errors per symbol error. The bit error probability is respectively

$$p_e(r) = \frac{1}{2}\int_0^\infty \text{erfc}\left(\frac{rA}{\sigma\sqrt{2}}\right)p_{rss}(r)dr \qquad (18)$$

and

$$p_e(r) = \frac{1}{2}\text{erfc}\left(\frac{rA}{\sigma\sqrt{2}}\right) \qquad (19)$$

were the noise variance is

$$\sigma^2 = N_0T_b + \sigma_i^2 \qquad (20)$$

and its interference component is

$$\sigma_i^2 = \frac{2KA^2}{3N}\left(b_0 + \frac{\exp(2\mu_0 + 2d_0)}{2}\right) \qquad (21)$$

where the bit energy, $E_b = A^2T_b/2$ [15].

The second measure of performance is the average delay. The development closely follows that of [16] but additionally caters for the encoding delay of the interleaver. Let us assume constant length packets, with a very large packet transmission period, T_p, and a transmitter-receiver inter-antenna propagation delay, τ_p. The time from the generation of a packet to the start of its transmission is, on average, half the period T_p of the packet currently under way so that the time taken by a successful packet without re-transmission is $T_p/2 + T_p$. The average re-transmission delay is $3/2T_p$. Note that this differs from model descriptions based on the start of transmission rather than the generation time. In the interleaved case, there is an additional encoding delay, $\tau_i = \lambda n(n-1)T_b$. If a packet is not successfully received, then the request for re-transmission travels T_a seconds making the total waiting time for receiving positive acknowledgement $(T_p + \tau_i + \tau_p) + (T_a + \tau_p) = T_p + T_a + \tau_i + 2\tau_p$. Processing time has been assumed negligible overall and transmitter buffering is assumed to be located after the interleaver to cut down on encoding delay. On receipt of acknowledgement, then either transmission of the failed packet resumes or new packets continue undisturbed. Since G is the offered traffic and S the throughput, the average number of re-transmissions is $(G - S)/S$. Finally, the average packet delay can be found as

$$D_a = \frac{3}{2}T_p + \tau_i + \tau_p + \left(\frac{G}{S}-1\right)\left(T_a + T_p + 2\tau_i + 2\tau_p\right) \qquad (22)$$

seconds, which can be normalised by the packet duration, T_p, into

$$D = 1.5 + T_i + d + \left(\frac{G}{S}-1\right)\left(\frac{T_a}{T_p} + 1 + 2T_i + 2d\right) \qquad (23)$$

time slots. If we regard the number of time slots of $N_T = T_a/T_p$ as uniformly distributed in the interval $[1, N_{Tm}]$ with the probability of selection being $1/N_{Tm}$ then the expected value is $N_{Tm}/2$ such that

$$D = 1.5 + T_i + d + \left(\frac{G}{S}-1\right)\left(\frac{1}{2}N_{Tm} + 1 + 2T_i + 2d\right) \qquad (24)$$

where $d = \tau_p / T_p$ and T_i is the encoding delay. Evidently, an increase in throughput should make D approach $1.5 + T_i + d$ as S approaches G in the limit.

In general, compound channels, where burst- and radom-errors co-exist, burst-and-random-error control codes are used [4].

64

Nevertheless, with interleaving, it is not necessary to resort to these codes at all if the average fade duration is accurately predicted. Similarly, well-designed concatenated codes will obviate long codes.

C. Validation.

In this experiment, we validate the results of code interleaving using Monte Carlo simulation of the shadowed Ricean channel model. The fading simulation model assumes that there is, at any instance, only one direct component in the received signal, that the direct component is from all directions above the horizon with equal probability and that its envelope is lognormally distributed, while its phase is uniform. We choose the envelope of the direct component using the independent lognormal random sequence generator, and its angle of incidence using the independent uniform random sequence generator operating over $[0, 2\pi]$. At this point we revert to the multi-tone model to select the envelope and phase of the multipath component.

The multi-tone model could not be used in the case of the lognormal component because the sum of lognormal distributions is not itself a lognormal distribution. A simulation run is then carried out for each codeword.

IV. Results.

Let us recall that the object here is to be able to efficiently control burst-and-random errors using any combination of the techniques just described. In order to demonstrate the relative strengths of the techniques, the following assumptions were made. The average fade duration was calculated over the range of car speeds from 5 through 100 km/h. The correlation coefficient, ρ was set at 0.7 as suggested by Loo's measurements taken in a densely wooded area, where the LOS component is only residual [2].

The speed of light was taken to be 2.997925×10^8 ms^{-1}. For performance calculations, the satellite is assumed to be in geo-synchronous orbit. The other settings are: $T_b = 12.7~\mu s$, $N = 127$, $n = 127$, $N_b = 7$, $E_b/N_0 \in [20, 30, 40]$ dB, $d = 74$ slots, $N_{Tm} = 3$, $K_{max} = 2N$, $T_s = 6.5T_c$. N_p equals the code length in effect under each scheme.
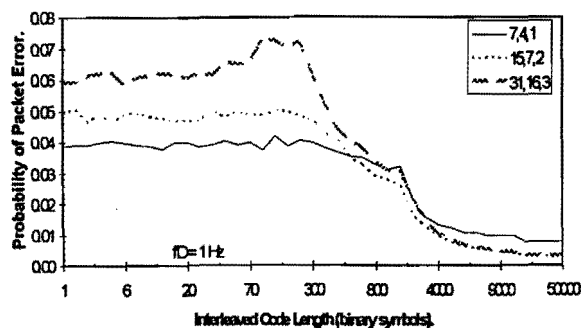
Figure 2. Simulation of the effect of the interleaving degree on the packet error probability in the shadowed Ricean channel with zero correlation between the envelope and the rate of change of the signal envelope.

The calculated value of the AFD under slow fading, heavily shadowed conditions is 19 884 bits. According to Figure 2, the irreducible probability of packet error is reached when the interleaved code reaches between 13 817 and 21 213 symbols in length which is consistent with the calculation. The agreement with Kenneth Brayer's field tests on various adaptive convolutional and concatenated cyclical codes [17] is remarkable as well. The rest of the tests depend on this result for the product in equation (11).

Figure 3 contrasts a plain RS (127, 123, 2) with the BCH (7, 4, 1) in concatenation with various $GF(128)$ RS codes. Curve e, for instance is equivalent to a (889, 492, 7) binary code whose code rate is superior to that of either of its constituent codes. The strong point of concatenated coding is its ability to lower the overall code rate. On the other hand, interleaving shows a factor of eight improvement over the plain (127, 123, 2) code. This is evident from Figures 3 and 4. Code interleaving simply improves the performance of the same code by converting the channel relative to the code length so that the duration of the fade looks short compared to the duration of the code itself. The difference, then, with concatenated coding is that the latter not only lengthens the code but also lowers the code rate producing a more powerful code from two moderately efficient codes.

Another observation is that the Reed-Solomon code is an excellent burst-error-correcting code with moderate random-error capabilities. This is amply illustrated in Figures 3 and 5. Note that the concatenated code is dominated by the power of the outer RS code. In

65
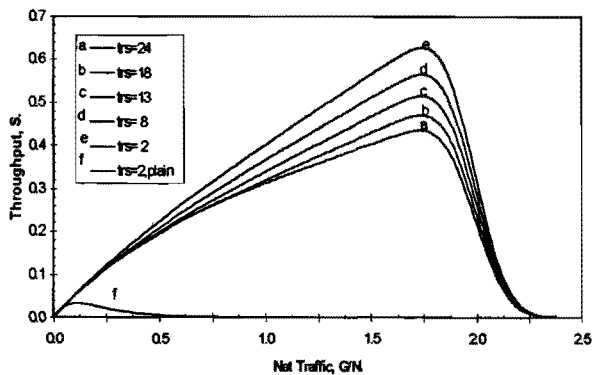
Figure 3. The throughput of concatenated block codes in a heavily shadowed, slow fading channel.

Figure 5, interleaving converts the channel such that it looks random-error-like. Hence, the performance of the concatenated code drops by a factor of two.
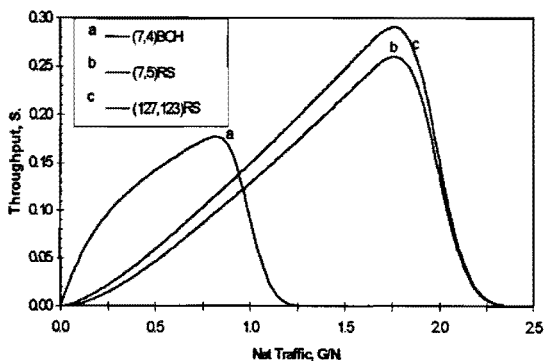
Figure 4. The throughput of the BCH (7,4), RS (7,5) and RS (127,123) codes interleaved to degree 19 884 bits.
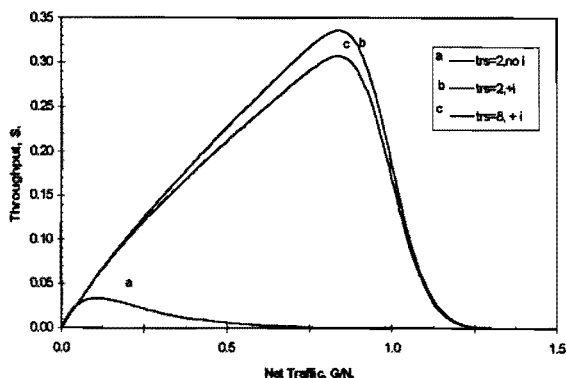
Figure 5. The throughput of the (889,444) and the (889,492) codes interleaved to degree 23 packet lenghts versus that of the plain (127,123) code.

Curves *b* and *c* in Figure 6 show a rather insignificant contribution of the encoding delay of the interleaver to the overall average delay due to the concatenated code. This is due to the special switching design of the synchronous interleaver which greatly reduces T_i. What is more significant is that both techniques notably reduce the average delay by boosting throughput.
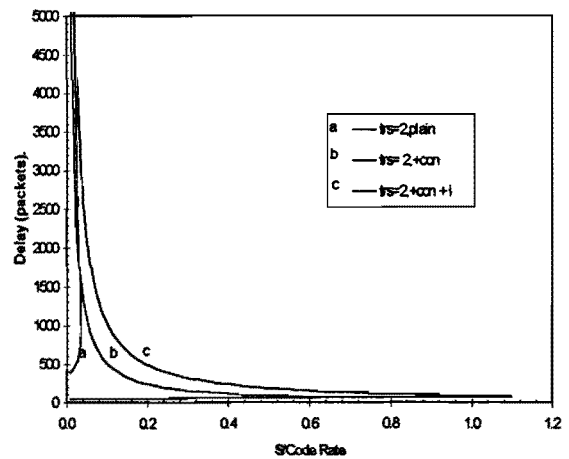
Figure 6. The average delay performance of the various coding schemes using the RS (127,123,2) code.

V. Conclusions

Three approaches to burst-error control in the heavily shadowed, slow fading channel have been compared. One is the use of a plain burst-error correcting RS code. The second involves interleaving either a random-error correcting or a burst-error correcting code. A third alternative is channel-state-sensing concatenated coding which can also be used in hybrid fashion with interleaving.

Channel-state-sensing concatenated coding is by far the most efficient of the three in terms of both throughput and average delay. Its combination with separate decoding significantly simplifies decoder design while retaining the power of the scheme.

For the best results, however, the designer needs to pay particular attention to a couple of points. It is important that the original codes be matched and efficient enough for the probability-of-error requirements at hand. In practice, the more powerful convolutional codes will be used for the inner code.

VI. References

1. Marvin K. Simon, Jim K. Omura, Robert A. Scholtz, Barry K. Levitt; *Spread Spectrum Communications Handbook, rev. ed.;* 1994.

2. Chun Loo; "A Statistical Model for a Land Mobile Satellite Link"; *IEEE Trans., VT*, vol. VT-34, No. 3, August 1985.

3. John L. Ramsey; "Realization of Optimum Interleavers"; IEEE Trasn. on Information Theory, Vol. IT-16, No. 2, May 1970. Pp. 338 - 345.

4. Shu Lin / Daniel Costello, Jr.; *Error-Control Coding: Fundamentals and Applications*; Prentice Hall Series in Computer Science Applications in Electrical Engineering, 1983. Chapters 9 & 15.

5. Vijay K. Bhargava, David Haccoun, Robert Matyas, Peter P. Nuspl; *Digital Communications by Satellite: Modulation, Multiple Access and Coding*; Wiley-Interscience, N.Y., 1981. Chapter 12.

6. S. O. Rice; "Statistical properties of a sine wave plus random noise", *Bell Syst. Tech. J.*, vol. 27, Jan. 1948, pp. 109-117.

7. William C. Jakes, Jr. (ed.); *Microwave Mobile Communications*; John Wiley, 1974. Chapter 1.

8. Arnold M. Michelson and Allen H. Levesque; *Error-Control Techniques for Digital Communication*; Wiley-Interscience, NY, 1985. Chapter 11.

9. Mqhele E. Dlodlo, Rogier N. van Wolfswinkel, Richard D. van Nee; "Performance Analysis of Hybrid Forward Error Correction Schemes in a Fast and Slow Rician Fading Wideband Land-Mobile Satellite Channel with BPSK and Path Diversity"; *Proceedings, IEEE PIMRC'94*, vol. II, The Hague, The Netherlands. pp. 579-583.

10. R.D.J. van Nee, Howard S. Misser and R. Prasad; "Direct-Sequence Spread Spectrum in a Shadowed Rician Fading Land-Mobile Satellite Channel"; *IEEE J. Select. Areas Commun.*, Vol. 10, No. 2, Feb. 1992.

11. R.D.J. van Nee and R. Prasad, "Spread-spectrum path diversity in a shadowed Rician fading land-mobile satellite channel", *IEEE Veh. Technol.*, vol. 42, pp. 131-135, May 1993.

12. R.D.J. van Nee; *Multipath and Multi-Transmitter Interference in Spread Spectrum Communication and Navigation Systems*; Delft University Press, 1995. Chapter 2.

13. Jan van Rees; "Measurements of the Wide-Band Radio Channel Characteristics for Rural, Residential, and Suburban Areas"; IEEE Trans. on Veh. Technol., vol. VT-36, no. 1, Feb. 1987, pp. 2 - 6.

14. Michael B. Pursley, Dilip V. Sarwate; "Error Probability for Direct-Sequence Spread-Spectrum Multiple-Access Communications — Part I: Upper and Lower Bounds"; *IEEE Trans. Commun.*, vol. COM-30, no. 5, May 1982.

15. Richard D. van Nee, Rogier N. van Wolfswinkel and Ramjee Prasad; "Slotted ALOHA and Code Division Multiple Access Techniques for Land-Mobile Satellite Personal Communications"; *IEEE J. Select. Areas Commun.*, vol. 13, no. 2, Feb. 1995, pp. 382-388.

16. Leonard Kleinrock and Fouad A. Tobagi; "Packet-Switching in Radio Channels: Part I — Carrier Sense Multiple-Access Modes and Their Throughput-Delay Characteristics"; IEEE Trans. on Commun., vol. COM-23, no. 12, Dec. 1975.

17. Kenneth Brayer; "Error Correction Code Performance on HF, Troposcatter, and Satellite Channels"; IEEE Trans. on Commun. Technol., Vol. COM-19, No. 5, October 1971. Pp. 781 - 789.

Unslotted Hybrid CDMA/ISMA Protocol for Indoor Wireless Computer Communications

Huy Linh Anh Le, *Student Member, IEEE**
Huub van Roosmalen****
Jos Nijhof, *Member, IEEE**
Ramjee Prasad, *Senior Member, IEEE**

*Telecommunications and Traffic Control Systems Group
Delft University of Technology
P.O. Box 5031, 2600 GA Delft
The Netherlands
Tel.: +31 15-78.23.86 (before 10-10-'95)
+31 15-278.24.17 (after 10-10-'95)
Fax: +31 15-78.17.74
E-Mail: Huy@Octopus.et.tudelft.nl

**Holland Institute of Traffic Technology B.V.
P.O. Box 245
7300 AE Apeldoorn
The Netherlands
Fax: +31 55-432553

ABSTRACT - A hybrid Code Division Multiple Access / Inhibit Sense Multiple Access (CDMA/ISMA) protocol has been proposed as an effective multiple access scheme for Indoor Wireless Computer Communications. This new protocol combines the advantages of both CDMA and ISMA into one protocol. On the one hand the ISMA protocol introduces a limitation to the number of simultaneous accesses to the transmission channel. On the other hand the CDMA protocol introduces an improvement to the packet survival chance. *Slotted* hybrid CDMA/ISMA protocol has been reported in [1]. It is shown that the performance of the hybrid protocol is indeed better than CDMA only. In addition, code sharing can be applied to reduce hardware cost. This paper presents the performance analysis of the *unslotted* CDMA/ISMA protocol in order to take more advantage of the strength of the hybrid protocol. The performance comparison between the slotted and unslotted hybrid CDMA/ISMA protocol is evaluated in terms of throughput and delay using computer simulation and mathematical analysis.

I. INTRODUCTION

Indoor wireless office communication is our main research field. The office system we focus on consists of a building in which users work together in groups. The participants generate terminal traffic. Terminals communicate with each other by radio transmission using a random access protocol. Terminals might not detect each other's transmission in radio communications. It can easily happen that two users are hidden from each other by some obstacle, in which case severe performance degradation results. This is called the *hidden terminal problem*. The introduction of a central base station can alleviate this problem by instructing it to send a busy tone to all participating terminals to forbid new transmissions when a transmission is going on. Still, a situation can occur in which two or more terminals simultaneously start their transmission, resulting in a collision. However, a great reduction in the number of simultaneous transmissions is achieved by the introduction of a central base station. This concept is called ISMA [1]-[4].

68

If we could somehow increase the survival chances of colliding packets, we could improve protocol performance. The near-far effect is one way to achieve this. A packet may 'capture' the receiver if it is much stronger than its competitors. This can happen when terminals use the same transmission power, but at different distances from the receiver. Because the 'near terminals' have better performance compared to the 'far terminals' (due to the better survival chance of the packets), the near-far effect introduces an unfair element among the terminals. Perfect power control, in which the transmitted power of the terminals are adjusted such that their received powers are all equal, can eliminate the near-far effect described above. The performance of CDMA has been reported in a number of publications, e.g. [5]-[10].

The hybrid CDMA/ISMA protocol combines the advantages of both CDMA and ISMA into one protocol. The advantages of CDMA and ISMA are the improvement of the survival chance of packets and the limitation of contention in the channel. Code sharing can also be applied. This is an important aspect because the number of distinct useful transmission codes is limited, especially for short code length.

In [11] the hybrid protocol combines Direct Sequence CDMA with *slotted* p-persistent ISMA. In this paper, we have investigated the performance of the hybrid CDMA/ISMA protocol using the *unslotted* non-persistent ISMA scheme. It is expected that for low traffic, the delay of the unslotted protocol will improve because when a terminal has a data packet to send, it does not have to wait until the start of the next time slot. In addition, even when data packets collide, there is a probability that the data is received correctly due to the use of CDMA. This is the strength of the hybrid CDMA/ISMA protocol. By using the *unslotted non-persistent* ISMA scheme, it is expected that we take more advantage of this strength.

This paper is organized as follows. Section II gives the protocol description of the unslotted hybrid CDMA/ISMA protocol. The performance analysis is then given in Section III. Section IV shows the assumptions used. Results are discussed in Section V. Finally, conclusions can be found in Section VI.

II. PROTOCOL DESCRIPTION

Fig. 1 depicts an example of the 16-terminal network configuration of the protocol. There is one central base station that is connected to several transceivers by wire. The base station controls the traffic flow with a busy tone that can be detected by all participating terminals.

Because all terminals can detect the busy tone, the hidden terminal problem is solved. Several terminals together with one transceiver form a group. Wireless communication is considered to take place between the transceiver and the terminals. The terminals around each transceiver share the same code. In this way the number of codes can be reduced. When a data packet is ready for transmission, a terminal transmits the packet to its transceiver according to the chosen ISMA scheme.
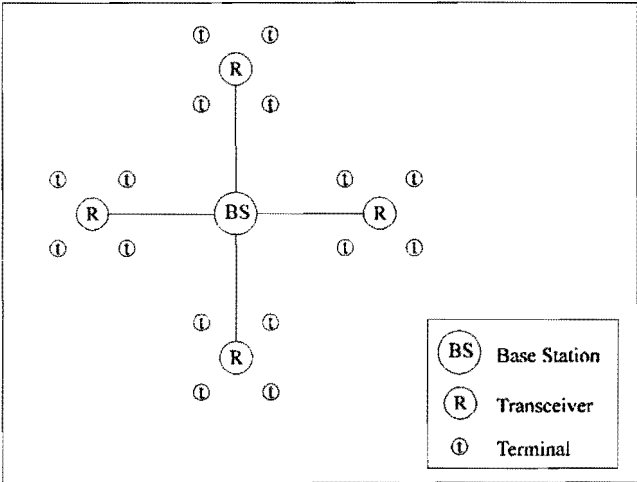
Fig. 1: 16-terminal network configuration

After the arrival of the data packet, the transceiver then simply forwards this packet to the base station which forwards it to all terminals via the transceivers. The destination then receives the packet and decides whether the packet is errorless or not. In case the packet was erroneous, a retransmission must take place. The return channel is not included in our analysis because only the base station makes use of this return channel and therefore contention is not a concern. The state of the terminals can either be free or blocked. At the beginning, the terminals are in the free state. If a packet arrives at a free terminal, the terminal jumps into the blocked state. In the blocked state, the terminal takes care that the arriving data packet is serviced successfully. In the mean time the blocked terminal ignores all incoming packets. This is a consequence of the assumption that the terminals do not have buffers for the incoming data packets.

The unslotted non-persistent ISMA protocol
Users behind terminals generate data packets. The data packets arrive at terminals, which will take care of the correct delivery of the packets. We assume that the arrivals of these packets are generated by a Poisson process. The SDL diagram of the unslotted non-persistent ISMA protocol is shown in Fig. 2. Before transmitting the data packet to the receiver, the terminal first listens to the channel to detect whether there is a busy tone going on. The base station broadcasts a busy

tone to signal that the channel is busy because there is a transmission in progress. If the channel is busy, the terminal waits a random delay before it can try again. Otherwise, the packet will be transmitted immediately. If a collision occurs from the transmission, the collided packets have to wait a random delay before they are allowed to try again (Fig. 2).
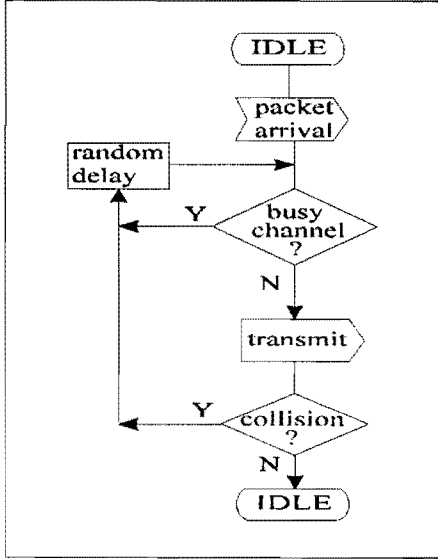
Fig. 2: Unslotted non-persistent ISMA

Although the number of collisions is greatly reduced with ISMA, collisions still can occur. In this unslotted ISMA scheme, collisions may occur due to multiple terminals transmitting packets during an interval called the *inhibit delay fraction d*. This interval d is necessary to switch from 'idle' to 'busy'. The reverse interval from 'busy' to 'idle' is denoted by d'. The inhibit delay fraction is normalized to the packet length, resulting in $0 \leq d < 1$ (Fig. 3). In this paper we assume $d = d'$.
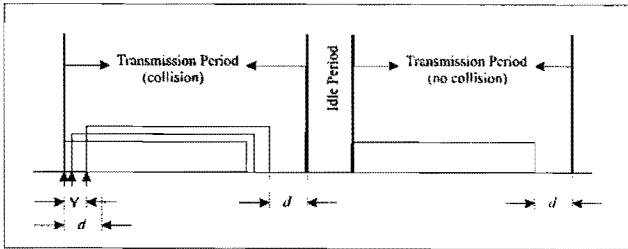
Fig. 3: Data Cycle

III. PERFORMANCE ANALYSIS

The protocol performance has been measured in terms of throughput S and delay D in relationship with the offered traffic G. The throughput measures the efficiency of the protocol and is defined as the fraction of time in which correct data packets are received. The delay is the time between the arrival of the first bit of a data packet at a terminal and its arrival of the last bit at the destination.

A. Throughput
The performance analysis of the hybrid unslotted non-persistent CDMA/ISMA will be explained from Fig. 3. The time axis in Fig. 3 is normalized to the packet duration. If a packet arrives and the terminal senses the channel free then the packet is sent immediately. Suppose that this time instant is t. It takes d, the inhibit delay fraction, before a busy tone is heard by all terminals. This moment occurs at the same time at all terminals because of the symmetric network configuration. Any other packet arriving between t and $t+d$ will sense the channel free (because the busy tone has not arrived yet) and will be transmitted resulting in a conflict. Because a CDMA scheme is also used, the conflict does not necessarily result in the loss of all packets. There is a probability that the first arriving packet can still be recovered successfully. If no other terminal transmits a packet during this period d, then no conflict occurs. We model the channel between the terminals and the corresponding transceivers as a multipath Rayleigh-fading channel. So, when no conflicts occur, there is still a probability that the packet *cannot* be recovered successfully.

Let $t+Y$ be the time of occurrence of the last packet arriving between t and $t+d$. This obviously means that Y must be between zero (the first transmitted packet is the only packet in the transmission period) and d (the end of the vulnerable period). The transmission of all packets arriving in $(t, t+Y)$ will be completed at $t+Y+1$. As noted before, the channel is sensed unused only a period d later. So, any terminal becoming ready between $t+d$ and $t+Y+1+d$ will sense the channel busy and hence will reschedule its packet. The interval between t and $t+Y+1+d$ is called a *transmission period* (*TP*). There will be *at most* one successful transmission during a *TP*. The *idle period* (*IP*) is defined as the period of time between two consecutive *TP*'s. A transmission period plus the following idle period constitute a *cycle*. Let $\overline{TP}$ be the expected duration of the transmission period, $\overline{I}$ the expected duration of the idle period, and the average cycle time can be written as:

$$\overline{t_c} = \overline{TP} + \overline{I} \tag{1}$$

Let U denote the time during a cycle in which the inbound channel (towards base station) is used to carry a successful packet transmission and $\overline{U}$ the corresponding average value, then we can write the throughput as [12]:

$$S = \frac{\overline{U}}{\overline{t_c}} \qquad (2)$$

The expected useful time $\overline{U}$ can easily be computed. When a packet is successful, the channel carries useful information for a duration of T_{pd}, the packet duration. In the unsuccessful case no useful information is carried at all or, in other words:

$$U = \begin{cases} T_{pd} & Successful \quad period \\ 0 & Unsuccessfull \quad period \end{cases} \qquad (3)$$

If $P_{success}$ denotes the probability that a transmitted packet is successful then

$$\overline{U} = T_{pd} \cdot P_{success} \qquad (4)$$

As noted before, the time is normalized to T_{pd} (the packet duration), and therefore T_{pd} equals to one. So, this gives us:

$$\overline{U} = P_{success} \qquad (5)$$

To calculate the average idle period we make the assumption that the total rate at which users schedule new and retransmitted packets forms a Poisson process with parameter G. So, new and rescheduled packets arrive at a rate of G packets per unit time. In literature, G is also called the offered channel traffic.

The probability of the idle time being greater than some value t is the probability that no packets are scheduled within a time interval of duration t and with the assumed Poisson packet scheduling process this probability becomes e^{-Gt}. Therefore the average value of I can be expressed as:

$$\overline{I} = \frac{1}{G} \qquad (6)$$

The average duration of a transmission period equals to:

$$\overline{TP} = 1 + \overline{Y} + d \qquad (7)$$

where $\overline{Y}$ is the expected value of Y. Since Y denotes the time at which the last interfering packet is scheduled, the probability of Y being smaller than some time y is the probability that no other packets (either new or retransmissions) are scheduled for transmission in an interval of duration d-y. With Poisson arrivals, the distribution for Y is $e^{-G(d-y)}$. The average of Y is therefore given by:

$$\overline{Y} = d - \frac{1}{G}\left(1 - e^{-Gd}\right) \qquad (8)$$

Applying the formulas obtained above we get:

$$\overline{t_c} = \overline{TP} + \overline{I} = 1 + \overline{Y} + d + \overline{I}$$
$$= 1 + \left[d - \frac{1}{G}\left(1 - e^{-Gd}\right)\right] + d + \frac{1}{G} \qquad (9)$$

$$S = \frac{\overline{U}}{\overline{t_c}} = \frac{P_{success}}{1 + \left[d - \frac{1}{G}\left(1 - e^{-Gd}\right)\right] + d + \frac{1}{G}}$$
$$= \frac{G \cdot P_{success}}{G(1 + 2d) + e^{-Gd}} \qquad (10)$$

The term $P_{success}$ in (10) is the only term that needs to be specified. Herefore, we distinguish between four types of transmissions during a transmission period:

1. Successful transmission without conflict
2. Unsuccessful transmission without conflict
3. Successful transmission with conflict
4. Unsuccessful transmission with conflict

If the channel is used for the delivering of a successful packet during a TP then this is denoted by a successful transmission (situation 1 or 3). The difference between a transmission with or without conflict can be found in the number of packet transmissions during a TP. In case there is more than one packet transmission during a TP, we speak of a transmission with conflict. If there is only one packet transmission then the transmission is conflict-free. Keeping this in mind, we can divide $P_{success}$ into two parts:

$$P_{success} = P_{success|noconflict} + P_{success|conflict} \qquad (11)$$

In which $P_{success|noconflict}$ corresponds to situation 1 and $P_{success|conflict}$ to situation 3.

$$P_{success|noconflict} = e^{-Gd} P_{ps} \qquad (12)$$

This is equal to the probability that no terminal transmits during the inhibit delay fraction d multiplied by the packet success probability P_{ps}. The multiplication with P_{ps} is necessary because of the assumption of the Rayleigh fading channel.

$$P_{ps} = \left(1 - P_{be|x=0}\right)^{L_p} \qquad (13)$$

71

$P_{be|x=0}$ is the bit error probability in Rayleigh fading channel using the CDMA scheme and is caused by self interference due to multipath. L_p denotes the packet length. We assume that during the packet duration, the bit error rate remains constant. The calculation of $P_{success|noconflict}$ is not very complicated. However, this is not the case for $P_{success|conflict}$ which is the probability that the first packet has arrived successfully given a collision (between two or more packets).

$$P_{success|conflict} = \sum_{x=1}^{K-1} P_x(d) \cdot P\{first\_packet\_OK|x\}$$
(14)

in which:
K is the number of active users in the system;
$P_x(d)$ is the probability of x arrivals during d for a Poisson distribution;
x denotes the number of arrivals during d.

Consider a reference terminal T_R. This terminal sends its data packet at an idle period first. Other terminals sending within the inhibit delay fraction interfere with the reception at this reference terminal T_R. The bit error rate depends on the configuration. So, if we try to take the different configurations into account, we have to calculate the bit error probability for all possible configurations and then average.

$$P\{first\_packet\_OK|x\} = E\left[1 - P_{be|\bar{x}}\right]^{L_p}$$
(15)

$E[.]$ denotes the expectation value, and $\bar{x}$ a certain configuration of x. The total number of combinations for $\bar{x}$ equals $\binom{K-1}{x}$. Because we assume an uniform distribution for all possible configurations we can write (K is the total number of active users in the system):

$$E\left[1 - P_{be|\bar{x}}\right]^{L_p} = \frac{1}{\binom{K-1}{x}} \cdot \sum_{\bar{x}} \left(1 - P_{be|\bar{x}}\right)^{L_p}$$
(16)

$$P_{success|conflict} = \sum_{x=1}^{K-1} P_x(d) \cdot \frac{1}{\binom{K-1}{x}} \cdot \sum_{\bar{x}} \left(1 - P_{be|\bar{x}}\right)^{L_p}$$
(17)

The formula for the bit error probability with DPSK modulation in Rayleigh fading channel is given by Kavehrad and Ramamurthi (formula 13a in [13]) in which the signal to noise ratio (SNR) is taken to be 20 dB. Combining the above results, the throughput can easily be calculated.

B. Delay

The average packet delay is defined as the duration between the transmission of the first bit till the correct reception of the last bit. For the calculation of the packet delay of the unslotted hybrid CDMA/ISMA protocol, we use the block diagram depicted in Fig. 4.
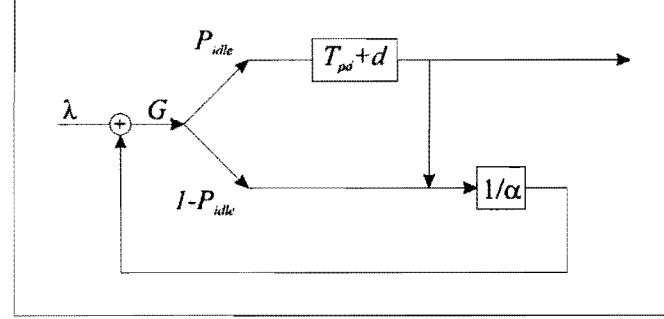
Fig. 4: Delay in the unslotted non-persistent hybrid CDMA/ISMA protocol

If a new packet arrives, the terminal immediately senses the channel to decide if it can transmit the packet. In case the channel is sensed busy (probability $1-P_{idle}$), the packet is rescheduled for a later time instant. This random delay process is assumed to be negative exponentially distributed with parameter α. The average random delay time is then $1/\alpha$. The average number of times the packet has to suffer this delay is $G(1 - P_{idle})/S$. In which G, S and P_{idle} are defined as before.

On the other hand, if the channel is sensed idle (probability P_{idle}), the packet will be transmitted immediately. This transmission will take a packet duration T_{pd} plus the inhibit delay fraction d before the knowledge about the outcome of the transmission is available (see also Fig. 3). So, the total delay until the terminal knows if the transmission was successful or not is $\left(T_{pd} + d\right)$. The transmission of a data packet does not necessarily result in a successful transmission; there are two possibilities: a successful or a failed transmission. If the transmission was successful, the packet leaves the system. The delay this packet suffers is $\left(T_{pd} + d\right)$. If the transmission was a failure, the packet again has to wait a random delay. In this case, the average delay is $\left(T_{pd} + d + \frac{1}{\alpha}\right)$ and the average number of schedulings is $\left(\frac{GP_{idle}}{S} - 1\right)$. Combining the above results, the average delay is finally given by formula (18).

$$D = \left(\frac{GP_{idle}}{S} - 1 \right) \left[T_{pd} + d + \frac{1}{\alpha} \right]$$
$$+ \frac{G(1 - P_{idle})}{S} \cdot \frac{1}{\alpha} + \left(T_{pd} + d \right) \qquad (18)$$

IV. ASSUMPTIONS

After DPSK modulation the signal is spread with a Gold code. The length of these codes are chosen to be 31. The signal is then transmitted, modulated on a 1.7 GHz carrier. The data rate is arbitrarily chosen to be 256·1024 b/s (0.26 Mb/s). The packet length is 64 bits. The delay spread is supposed to be 100 ns. Perfect power control is assumed to assure that signals from terminals within the same group arrive at the transceiver with the same power.

The far field model [14] is chosen to describe the signal attenuation. Given the transmitted power P_T, the received power P_R can be expressed by the following equation:

$$P_R = \frac{g_T g_R}{\left(\frac{4\pi f l}{c} \right)^\alpha} P_T \qquad (19)$$

where g_T and g_R are respectively the transmitter and the receiver gains, f the signal frequency, l the distance between the transmitter and the receiver, c the speed of light and α is the attenuation parameter. This attenuation parameter is chosen equal to 2, corresponding to free space propagation.

Fading is the result of the propagation of the transmitted signal through several paths. The channel is modeled as a Rayleigh fading channel. The Rayleigh fading channel model is valid when each path contributes the same amount of energy to the composite received signal. A Line of Sight (LOS) path is therefore assumed to be absent.

An example of a 16-terminal network configuration that we adopted in this paper is shown in Fig. 2. The terminals are clustered around distributed transceivers at a fixed distance (5 m). The distributed transceivers are clustered around the base station at a fixed distance (30 m). Those distances are kept fixed for all simulations. The number of terminals within a transceiver group (also called group size) is chosen as a power of two. For a certain fixed number of participated terminals, we have to halve the group size if we want to double the number of codes (the number of codes is also the number of transceivers). The comparison is fair in this manner, because when we want to investigate the effect of the number of codes on the performance, we

have to keep the number of participated terminals fixed. The number of terminals is 32 unless stated otherwise.

V. RESULTS

Fig. 5 a and b depict the simulation comparison between the slotted and unslotted hybrid protocol.
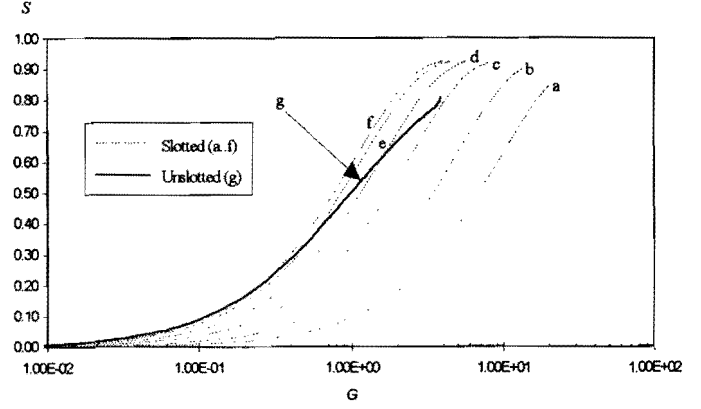
Fig. 5a: Throughput versus offered traffic using simulation of slotted (a..f: P_{tr} =0.1, 0.2, 0.4, 0.6, 0.8, 0.9) and unslotted (g: d=0.01) hybrid protocol; 8 codes; α=0.1
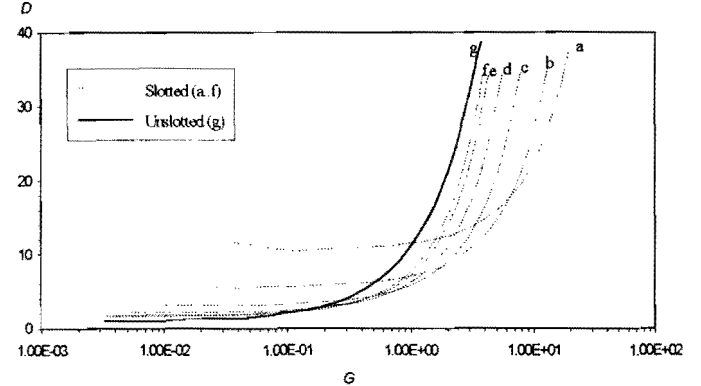
Fig. 5b: Delay versus offered traffic using simulation of slotted (a..f: P_{tr} =0.1, 0.2, 0.4, 0.6, 0.8, 0.9) and unslotted (g: d=0.01) hybrid protocol; 8 codes; α=0.1

In [11] P_{tr} is defined for the slotted protocol as the probability that the packet will actually be transmitted in case the channel is sensed *free*. For low traffic, the delay of the unslotted protocol is, as expected, better than the slotted protocol. For high traffic the situation is the other way round. The slotted protocol does not differ much from the unslotted protocol for high values of P_{tr}.

Fig. 6 represents the effect of d on the performance of the unslotted protocol. For high values of d, the performance degrades tremendously. This is due to the increase of the number of data packets colliding during

this period. This is why the ISMA protocol is not useful for environment in which the propagation delay is high.
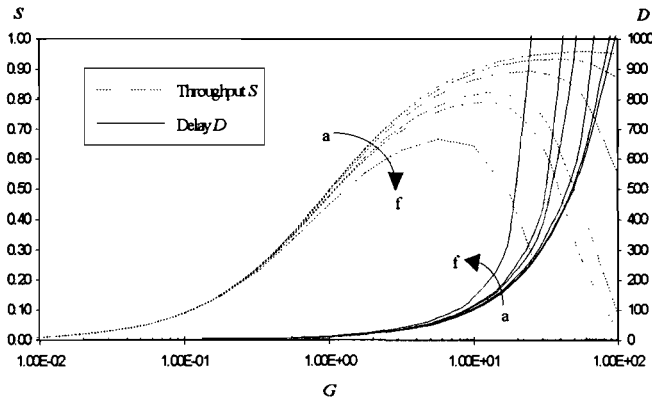
Fig. 6: Throughput and delay versus offered traffic using mathematical analysis of unslotted protocol, varying the inhibit delay fraction d. a..f: d=0.01, 0.02, 0.04, 0.08, 0.1 and 0.2; 8 codes; α=0.1
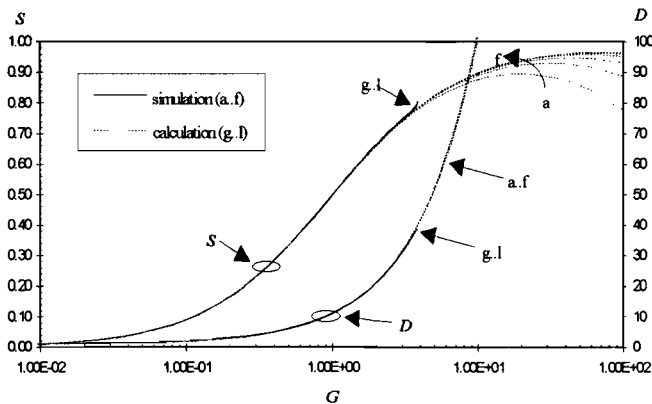
Fig. 7: Comparison between simulation and mathematical analysis for the unslotted protocol. a..f, g..l: 1, 2, 4, 8, 16, 32 codes; d=0.01; α=0.1

The comparison between the simulation and mathematical model for the unslotted protocol is given in Fig. 7. In case of simulation, the arrival rate at terminals is the input parameter and G is one of the output parameters. In addition, due to the assumption that terminals does not have buffers the performance of simulation does not exceed $G \approx 4$ packets/(packet duration). The simulation and mathematical model does quite agree with each other. This graph also shows the effect of code sharing. For a 32-terminal network, it is sufficient to use only four codes.

VI. CONCLUSIONS

The unslotted hybrid CDMA/ISMA protocol has been investigated. The protocol performance is measured in terms of throughput S and delay D in relationship with the offered traffic G. Herefore, we have been able to derive a close form formula for the throughput and delay. Also a simulation program has been developed to compare the results. The results from simulation and mathematical model does quite agree with each other.

Comparison between the slotted and unslotted protocol shows that for low offered traffic the delay for the unslotted protocol is better than the slotted protocol. For high traffic it is the other way round.

Furthermore it is concluded that the proposed hybrid protocol performs very well in case the propagation delay is not a concern. If the propagation delay is too high the performance drops quickly.

REFERENCES

[1]   R. Prasad,
"Performance Analysis of Mobile Packet Radio Networks in Real Channels with Inhibit-Sense Multiple Access",
IEE Proc. I, Vol. 138, pp. 458-464, October 1991.

[2]   I. Widipangestu, A.J. 't Jong and R. Prasad,
"Capture Probability and Throughput Analysis of Slotted ALOHA and Unslotted np-ISMA in a Rician/Rayleigh Environment",
IEEE Transactions on Vehicular Technology, Vol. 43, No. 3, pp. 457-465, August 1994.

[3]   K.J. Zdunek, D.R. Ucci and J.L. Locicero,
"Throughput of Nonpersistent Inhibit Sense Multiple Access with Capture",
Electronics Letters, Vol. 25, No. 1, pp. 30-31, 5th January 1989.

[4]   R. Prasad, and C.Y. Liu,
"Throughput Analysis of Some Mobile Packet Radio Protocols in Rician Fading Channels",
IEE Proceedings-I, Vol. 139, No. 3, pp. 297-302, June 1992.

[5]   M.G. Jansen, and R. Prasad,
"Capacity, Throughput, and Delay Analysis of a Cellular DS CDMA System With Imperfect Power Control and Imperfect Sectorization",
IEEE Transactions on Vehicular Technology, Vol. 44, No. 1, pp. 67-75, February 1995.

[6]   R.A. Scholtz,
"The Origins of Spread Spectrum Communications",
IEEE Transactions on Communications, Vol. COM-30, No. 5, pp. 822-854, May 1982.

[7] W.C.Y. Lee,
"Overview of Cellular CDMA",
IEEE Transactions on Vehicular Technology,
Vol. 40, No. 2, pp. 291-302, May 1991.

[8] M.B. Pursley,
"Performance Evaluation of Phase Coded
Spread Spectrum Multiple Access
Communication - Part I: System Analysis",
IEEE Transactions on Communications, Vol.
COM-25, No. 8, pp. 795-799, August 1977.

[9] R.L. Pickholtz, D.L. Schilling and L.B. Milstein
"Theory of Spread Spectrum Communications -
A Tutorial",
IEEE Transactions on Communications, Vol.
COM-30, No. 5, pp. 855-884, May 1982.

[10] K.S. Gilhousen, I.M. Jacobs, R. Padovani, A.J.
Viterbi, L.A. Weaver, Jr. and C.E.Wheatley III,
"On the Capacity of a Cellular CDMA System",
IEEE Transactions on Vehicular Technology,
Vol. 40, No. 2, pp. 303-312, May 1991.

[11] H. van Roosmalen, J. Nijhof and R. Prasad,
"Performance Analysis of a Hybrid
CDMA/ISMA Protocol for Indoor Wireless
Computer Communications",
IEEE Journal on Selected Areas in
Communications, Vol. 12, No. 5, pp. 909-916,
June 1994.

[12] L. Kleinrock and F.A. Tobagi,
"Packet Switching in Radio Channels: Part I-
Carrier Sense Multiple Access and their
Throughput-Delay Characteristics",
IEEE Trans. on Commun., Vol. COM-23, No.
12, pp. 1400-1416, December 1975.

[13] M. Kavehrad and B. Ramamurthi,
"Direct -Sequence Spread Spectrum with DPSK
Modulation and Diversity for Indoor Wireless
Communications",
IEEE Transactions on Communications, Vol.
COM-35, No. 2, pp. 224-236, February 1987.

[14] A. B. Carlson,
Communication systems,
Mc. Graw Hill Book Company, 3rd edition,
1988, pp. 92.

Maximum Likelihood Joint Phase Estimators in CDMA Communication Systems

L. Schumacher* and L. Vandendorpe[†]
Communications and Remote Sensing Lab.
Université catholique de Louvain
Place du Levant, 2 - Bâtiment Stévin
B- 1348 Louvain-La-Neuve
E-mail : vdd@tele.ucl.ac.be

September 15, 1995

Abstract

In this paper, the problem of carrier phase estimation in a Multi User CDMA context is considered. Regarding the phases as deterministic but unknown parameters, maximum-likelihood (ML) estimation is applied in both data-aided (DA) and non-data aided (NDA) contexts, leading to analytical expressions of ML phase estimators. Variances of these estimators are estimated in a BPSK modulation scheme and compared with respect to various benchmarks.

1   Introduction

Efficient digital communications require the determination of various parameters characterizing the transmission link : timing, channel responses, signal power,... and phase for coherent detection. Moreover, in a CDMA system, parameter estimation is complicated by the interference due to other active users. Interference between users, or Multiple Access Interference (MAI), results in biased parameters and/or important jitter.

This paper presents analytical expressions of Maximum-Likelihood joint phase estimators to be used in a CDMA communication system. After a brief system description in Section 2, Data-Aided ML joint phase estimation is studied in Section 3. An estimator is derived by linearizing the ML equation. The quality of this ML DA estimator is estimated by computing its variance and comparing it with respect to the Cramer-Rao bound and the variance of a conventional estimator. In Section 4, a Non-Data Aided joint phase estimator is derived, following a method described in [1]. The variance of this ML NDA estimator is compared to that corresponding to the Cramer-Rao bound.

---

2 System description

The lowpass equivalent signal transmitted by user k is given by

$$x_k(t) = \sqrt{2\,P_k} \sum_{n=-\infty}^{\infty} I_k^n a_k(t) u(t-nT) \tag{1}$$

where I_k^n are the data produced by user k, $a_k(t)$ is the waveform associated with the periodical pseudo-noise sequence multiplying the signal for user k and $u(t)$ is the symbol shape.

Assuming that the channel between user k and the receiver is a linear channel with equivalent lowpass impulse response $c_k(t)$, the received signal from N_u active users is given by

$$r(t) = \sum_{k=1}^{N_u} \sqrt{2\,P_k}\; e^{-j\phi_k} \sum_{n=-\infty}^{\infty} I_k^n h_k(t-nT) + n(t) \tag{2}$$

where $h_k(t) = [a_k(t)u(t)] \otimes c_k(t)$, ϕ_k is the phase parameter for user k, and n(t) is the additive white gaussian noise (AWGN) with one-sided power spectral density of N_0.

The parameters to be estimated are regarded as deterministic but unknown. This leads thus to maximum-likelihood estimation [2]. The logarithm of the likelihood function $\Lambda_L(r)$ writes

$$
\begin{aligned}
\Lambda_L(r) \;=\; & Cst - \frac{1}{N_0} \int_{-\infty}^{\infty} |r(t)|^2 + \frac{2}{N_0} \Re \left[\sum_{k=1}^{N_u} 2\,P_k\, e^{j\phi_k} \sum_{n=-\infty}^{\infty} (I_k^n)^\star y_k^n \right] \\
& - \frac{1}{N_0} \sum_{k=1}^{N_u} 2\,P_k \sum_{n=-\infty}^{\infty} |I_k^n|^2 x_{k,k}^{n,n} - \frac{1}{N_0} \sum_{k=1}^{N_u} 2\,P_k \sum_{n=-\infty}^{\infty} \sum_{\substack{m=-\infty \\ m \neq n}}^{\infty} I_k^n (I_k^m)^\star x_{k,k}^{n,m} \\
& - \frac{2}{N_0} \sum_{k=1}^{N_u} \sum_{\substack{l=1 \\ l \neq k}}^{N_u} \sqrt{P_k\,P_l}\; e^{j(\phi_l - \phi_k)} \sum_{n=-\infty}^{\infty} \sum_{\substack{m=-\infty \\ m \neq n}}^{\infty} I_k^n (I_l^m)^\star x_{k,l}^{n,m}
\end{aligned}
\tag{3}
$$

where y_k^n represents the normalized matched filter output and $x_{k,l}^{n,m}$ the normalized channel correlation coefficients

$$y_k^n \;=\; \frac{1}{\sqrt{2\,P_k}\,T} \int_{-\infty}^{\infty} r(t) h_k^\star(t-nT) dt \tag{4}$$

$$x_{k,l}^{n,m} \;=\; \frac{1}{T} \int_{-\infty}^{\infty} h_k(t-nT) h_l^\star(t-mT) dt \tag{5}$$

The last two terms of the expression (3) hereabove may be interpreted as interference. The former represents the self-interference (ISI), while the latter is associated with multiple access interference (MAI).

3 Data-Aided Joint Phase Estimation Estimation

3.1 Derivation of the estimator

The ML estimators are theoretically derived from the maximization of the likelihood function. A necessary but not sufficient condition for this maximum can be obtained by differentiating $\Lambda(r)$ with respect to the unknown parameters Φ and setting the result equal to zero [2].

$$\left.\frac{\partial \Lambda(r)}{\partial \Phi}\right|_{\Phi = \hat{\Phi}} = 0 \qquad (6)$$

Moreover, in the present case, $e^{j\hat{\phi}_u}$ is linearized around the true value of the parameter to be estimated, such that

$$e^{j\hat{\phi}_{u,DA}} = e^{j\phi_u}\left(1 + j(\hat{\phi}_{u,DA} - \phi_u)\right) \qquad (7)$$

With such a linearization, the following developments fall under the hypothesis of high SNR. Using (6) and (7), the first derivative of (3) leads to the following condition

$$\Im\left[2\,P_u\,T\,e^{j\phi_u}\left(1 + j(\hat{\phi}_{u,DA} - \phi_u)\right)\sum_{n=-\infty}^{\infty}(I_u^n)^\star y_u^n \right.$$
$$-2\,T\sqrt{P_u\,P_k}\,e^{j\phi_u}\left(1 + j(\hat{\phi}_{u,DA} - \phi_u)\right)$$
$$\left.\sum_{k=1}^{N_u} e^{-j\phi_k}\left(1 - j(\hat{\phi}_{k,DA} - \phi_k)\right)\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}(I_u^n)^\star I_k^m x_{k,u}^{m,n}\right] = 0 \qquad (8)$$

Solving (8) gives

$$(\hat{\phi}_{u,DA} - \phi_u) = -\frac{\Im\left[A_u\right]}{\Re\left[A_u\right]} \qquad (9)$$

where

$$A_u = 2\,P_u\,T\,e^{j\phi_u}\sum_{n=-\infty}^{\infty}(I_u^n)^\star y_u^n$$
$$-2\,T\sqrt{P_u P_k}\sum_{k=1}^{N_u}(1 - j(\hat{\phi}_{k,DA} - \phi_k))\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}(I_u^n)^\star I_k^m x_{k,u}^{m,n} \qquad (10)$$

One can note in (9) that the linearization of $e^{j\hat{\phi}_{u,DA}}$ has led to a result similar to classical ones [3], but where the highly nonlinear arctan function has disappeared.

The following developments are limited to the 2-user case. Expanding y_u^n, direct $(x_{u,u}^{n,n})$ and ISI $(x_{u,u}^{m,n}, m \neq n)$ terms vanish from the numerator as real terms. Moreover, regarding MAI and noise contributions in the denominator as negligible, the ML DA joint phase estimators are of the type

$$(\hat{\phi}_{u,DA} - \phi_u) = \frac{\Im[Noise_u]\,\Re[Direct_v + ISI_v] + \Re[MAI_{v,u}]\,\Im[Noise_v]}{\Re[Direct_u + ISI_u]\,\Re[Direct_v + ISI_v] - \Re[MAI_{v,u}]\,\Re[MAI_{u,v}]} \qquad (11)$$

where

$$Direct_u = \sum_{n=-\infty}^{\infty} |I_u^n|^2 x_{u,u}^{n,n} \qquad (12)$$

$$ISI_u = \sum_{n=-\infty}^{\infty} \sum_{\substack{m=-\infty \\ m \neq n}}^{\infty} (I_u^n)^\star I_u^m x_{u,u}^{m,n} \qquad (13)$$

$$MAI_{u,v} = 2\sqrt{\frac{P_u}{P_v}} e^{(\phi_v - \phi_u)} \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} (I_v^n)^\star I_u^m x_{u,v}^{m,n} \qquad (14)$$

$$Noise_u = \sqrt{\frac{2}{P_v}} \frac{e^{j\phi_u}}{T} \sum_{n=-\infty}^{\infty} (I_u^n)^\star \int_{-\infty}^{+\infty} n(t) h_u^\star(t - nT) dt \qquad (15)$$

3.2   Variance of the estimator - Computational results

To be able to compute the variance of (11), the denominator is replaced by its mathematical expectation. Expression of the variance is then derived in the case of a BPSK modulation. It depends on the structure of the correlation factors $x_{u,v}^{m,n}$, the phase difference between users, and the size of the averaging window.

This expression of the variance has been computed in different scenarii. The results presented here (Fig. 1) are related to a situation where the coupling between users is as high as 20 % of the direct ray, without spreading around it ($x_{u,v}^{n,n} = 0.2\ x_{u,u}^{n,n}$; $x_{u,v}^{m,n} \equiv x_{u,u}^{m,n} \equiv 0$ for $m \neq n$). Different phase offsets are tested, and the resistance to the Near-Far effect is also shown. Variance of the ML DA estimators is presented with respect to the Cramer-Rao bound and the variance of a conventional estimator. By conventional estimator, one should understand a process not modelizing the possible coupling, and dealing with the matched filters output y_u^n as if they were corresponding to a single active user, although several are interfering through MAI.

As $\frac{E_b}{N_0}$ grows, ML DA estimators appear (Fig. 1, left column) to be performing better than the conventional ones, excepted when the phase difference reaches zero. These estimators are not limited by the threshold effect visible on the performance of the conventional estimator used in a multiple access system. Asymptotically, their variance tends to the Cramer-Rao bound.

This independence of the treshold effect is especially interesting when estimation has to face a Near-Far effect (Fig. 1, right column. Near-Far power ratio is 1:10). The Near-Far effect increases the value of the treshold limiting the performance of the conventional estimator in the multiple access system. Joint estimation outperforms then such an estimator.

4   Non Data-Aided Joint Phase Estimation

4.1   Derivation of the estimator

The NDA maximum-likelihood estimators result from the maximization of the likelihood function $\Lambda(r)$ averaged over the statistical symbol distribution.
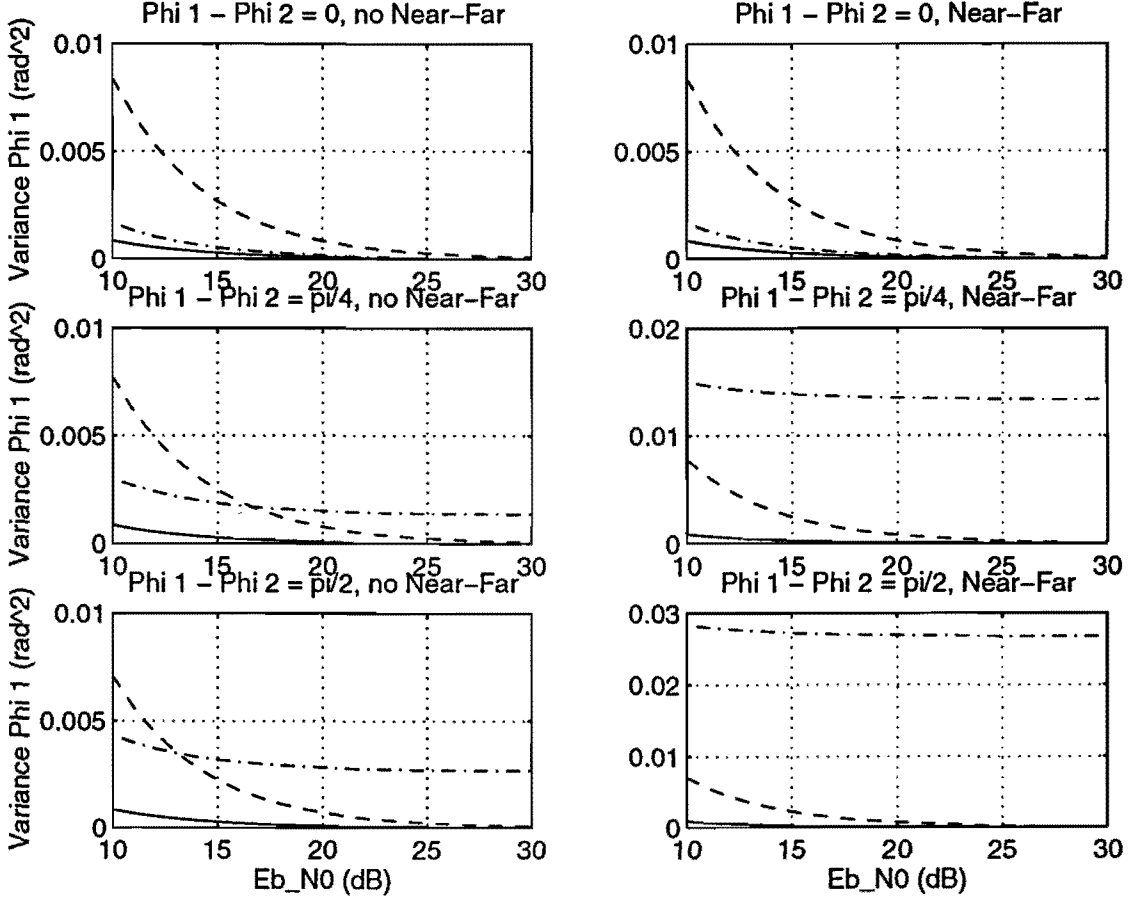
Figure 1: Variance of DA ML estimators (Joint - -, Cramer-Rao —, Conventional - .)

$$\bar{\Lambda}(r) = \int\limits_{-\infty}^{+\infty} \Lambda(r) \, p(I_k^n) \, dI_k^n \tag{16}$$

In this case, symbols are regarded as independent, identically distributed zero-mean binary data. The averaging of the likelihood function $\Lambda(r)$ over the pdf of I_k^n is hardly analytically manageable, due to the presence of the exponential function. Moeneclaey and de Jonghe proposed in [1] a way to derive the NDA likelihood function $\Lambda(r)$, based on the expansion of the exponential function into a power series. Applying their method, $\bar{\Lambda}(r)$ writes

$$\begin{aligned}
\bar{\Lambda}(r) &= \frac{1}{2} \sum_{k=1}^{N_u} \left(\frac{4 \, P_k \, T}{N_0} \right)^2 \sum_{n=-\infty}^{+\infty} \Re \left[e^{j\phi_k} \, (I_k^n)^\star \, y_k^n \right]^2 \\
&\quad + \frac{1}{2} \sum_{k=1}^{N_u} \sum_{l=1}^{N_u} \frac{4 \, P_k \, P_l}{N_0^2} \sum_{p=-\infty}^{+\infty} \sum_{q=-\infty}^{+\infty} e^{2j(\phi_l - \phi_k)} \left[I_k^p \, (I_l^q)^\star \, x_{k,l}^{p,q} \right]^2
\end{aligned} \tag{17}$$

To notice is the NDA-typical quadratic operator applying to each term of this expression.

In practice, the maximization of the NDA likelihood function (17) is realized, similarly to the DA case, by founding parameter values zero forcing the first derivative of $\bar{\Lambda}(r)$ with respect to them. Considering BPSK modulation, conditions to satisfy are of the form

$$\Im \left[e^{2j\phi_u} \sum_{n=-\infty}^{+\infty} (y_u^n)^2 + e^{2j\phi_u} \sum_{k=1}^{N_u} e^{-2j\phi_k} \sum_{p=-\infty}^{+\infty} \sum_{q=-\infty}^{+\infty} \left(x_{k,u}^{p,q} \right)^2 \right] \Bigg|_{\Phi = \hat{\Phi}_{NDA}} = 0 \tag{18}$$

leading to the following NDA ML phase estimator

$$\hat{\phi}_{u,NDA} = -\frac{1}{2} \tan^{-1} \frac{\Im \left[\sum\limits_{n=-\infty}^{+\infty} (y_u^n)^2 + \sum\limits_{\substack{k=1 \\ k \neq u}}^{N_u} exp \left(-2j\hat{\phi}_{k,NDA} \right) \sum\limits_{p=-\infty}^{+\infty} \sum\limits_{q=-\infty}^{+\infty} \left(x_{k,u}^{p,q} \right)^2 \right]}{\Re \left[\sum\limits_{n=-\infty}^{+\infty} (y_u^n)^2 + \sum\limits_{\substack{k=1 \\ k \neq u}}^{N_u} exp \left(-2j\hat{\phi}_{k,NDA} \right) \sum\limits_{p=-\infty}^{+\infty} \sum\limits_{q=-\infty}^{+\infty} \left(x_{k,u}^{p,q} \right)^2 \right]} \tag{19}$$

Like with the DA estimator, coupling effects appear through correlation factors $x_{k,u}^{p,q}$.

4.2 Variance of the estimator - Computational results

In order to determine a more realistic lower bound on the phase variance than the Cramer-Rao bound, the method outlined in [4] is applied. After heavy calculations, an analytical expression of the variance is finally derived for a 2-user system.

To allow some comparison with similar performance in mono-user systems [5], the variance expression is simulated in an asynchronous system $(x_{u;v}^{n;n} = 0.1x_{u;u}^{n;n})$ (Fig.2). These simulations are lead regardless of possible improvement due to code cross-correlation properties. The width of the observation window is the same as that mentioned in [5].
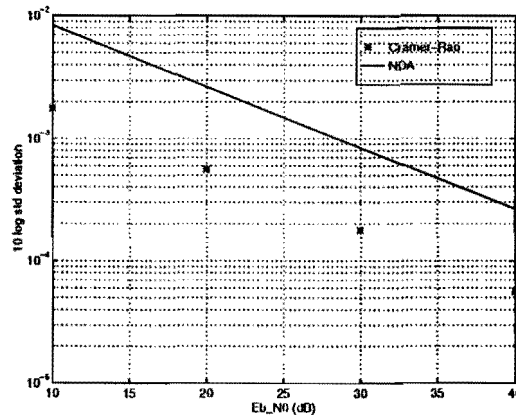
Figure 2: Lower bound on the performance

Moreover, the resistance in a Near-Far context is tested (Fig.3). The channel is the same as that used in the previous simulation, but the observation window is 10 times narrower. Regarding the Cramer-Rao bound as the ultimate performance level, reachable in a AWGN context, it logically appears that the estimator is badly performing when the power of the interfering user grows.
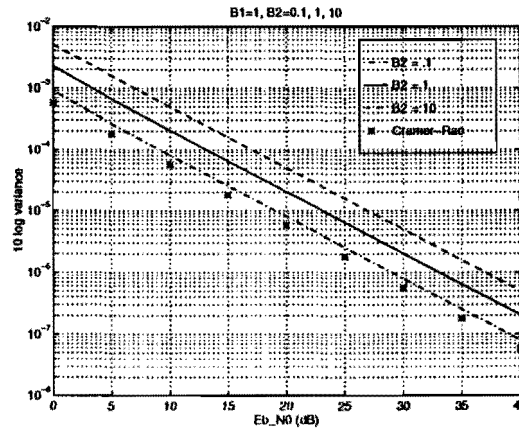
Figure 3: Near-Far effect

5   Conclusions

Maximum-Likelihood joint phase estimators for CDMA multiple access communication systems were presented, in both Data-Aided and Non Data-Aided contexts. Their analytical expression were established. Performance of these estimators were demonstrated by computing their variance and comparing them with other variances and the Cramer-Rao bound.

References

[1] G. de Jonghe M. Moeneclaey. Tracking Performance Comparison of Two Feedforward ML-Oriented Carrier-Independant NDA Symbol Synchronizers. *IEEE Transactions on Communications*, 40:1423–1425, September 1992.

[2] H.L. Van Trees. *Detection, Estimation and Modulation Theory*. John Wiley and Sons, Inc., 1968.

[3] J.G. Proakis. *Digital Communications*. Mc Graw-Hill International Editions, 1989.

[4] L. E. Franks. Carrier and Bit Synchronization in Data Communication - A Tutorial Review. *IEEE Transactions on Communications*, 28(8):1107–1120, August 1980.

[5] L. E. Franks M. H. Meyers. Joint Carrier Phase and Symbol Timing Recovery for PAM Systems. *IEEE Transactions on Communications*, 28(8):1121–1129, August 1980.

All Optical Signal Processing Based on Biphase Pulse Coding

TIOS
Tele-Informatics
& Open Systems

J.M. Rijnders and A.C. van Bochove
Dept. of Electrical Engineering, University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands.

University
of Twente

Abstract

A novel optical signal processing (OSP) method is proposed which is able to manipulate data in the optical domain. It uses biphase pulse line-coding and time-slot shifts of the input signal in order to enable signal manipulation.
It features wavelength transparency, a constant processing delay, signal regeneration in case of on-off keying and suits most modulation formats. Although it can be used in many applications it is presented here as an ATM cell header replacement module.

Possible Application

In ATM networks cells are being routed through the network on basis of the information contained within the header (see figure 1). Optical transparency can be achieved when the main functions of the switching nodes can be performed in the optical domain. These functions are:

* **header replacement of ATM-cells at switch input.**

* **routing of ATM-cells from switch input to output.**

figure 1: simplified ATM network example

Line-code requirement

Based on an electrical interface at the sender (Tx) and receiver (Rx) side, figure 2 gives the required biphase pulse encoding, decoding and electro-optic (E/O) conversion steps. Moreover, only the data to be manipulated has to be encoded.

figure 2: encoding and decoding.

OSP operation

1. A part of the input is used for table-lookup and switch management (indicated by 'electronics').
2. The input signal is copied and delayed by 0, 1 and 2 time-slots (=bits), respectively
3. When the input consists of biphase pulses, appropriate adjustment of the electro-optic switches (sw1, sw2 and sw3) enables us to select a '1' or a '0' in every time-slot and forward it to the output (see example below).

—— optical path
—— electrical path

figure 3: OSP configuration

Example

logical representation: 1 1 0 1 0
biphase representation: 0101100110

signal
delay   10 9 8 7 6 5 4 3 2 1 0 time-slots
0T.........0101100110
1T.........0101100110
2T........0101100110

Conclusions

Optically transparent signal processing, based on biphase pulse line-coding, has been presented.
The advantages are:
* Independent of modulation format since the input bits are only rearranged to obtain the output bits.
* Transparent to wavelength.
* High speed operation.
* Possibility to integrate.

Possible improvements:
* Use of Runlength Limited line-coding to improve code rate and to reduce the required hardware.

Other application:
* All optical Add/Drop Node

83

All Optical Signal Processing Based on Biphase Pulse Coding

J.M. Rijnders and A.C. van Bochove
Dept. of Electrical Engineering, University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands.

Abstract

A novel optical signal processing (OSP) method is proposed which is able to manipulate data in the optical domain. It uses biphase pulse line-coding and time-slot shifts of the input signal in order to enable signal manipulation.

It features wavelength transparency, a constant processing delay, signal regeneration in case of on-off keying and suits most modulation formats. Although it can be used in many applications it is presented here as an ATM cell header replacement module.

Possible Application

In ATM networks cells are being routed through the network on basis of the information contained within the header (see figure 1). Optical transparency can be achieved when the main functions of the switching nodes can be performed in the optical domain. These functions are:

* header replacement of ATM-cells at switch input.

* routing of ATM-cells from switch input to output.

figure 1: simplified ATM network example

Line-code requirement

Based on an electrical interface at the sender (Tx) and receiver (Rx) side, figure 2 gives the required biphase pulse encoding, decoding and electro-optic (E/O) conversion steps. Moreover, only the data to be manipulated has to be encoded.

figure 2: encoding and decoding.

OSP operation

1. A part of the input is used for table-lookup and switch management (indicated by 'electronics').
2. The input signal is copied and delayed by 0, 1 and 2 time-slots (=bits), respectively
3. When the input consists of biphase pulses, appropriate adjustment of the electro-optic switches (sw1, sw2 and sw3) enables us to select a '1' or a '0' in every time-slot and forward it to the output (see example below).

figure 3: OSP configuration

Example

logical representation: 1 1 0 1 0
biphase representation: 0101100110

signal
delay
         |10|9|8|7|6|5|4|3|2|1|0| time-slots
0T.........0101100110
1T.........0101100110
2T........0101100110

Conclusions

**Optically transparent signal processing, based on biphase pulse line-coding, has been presented.
The advantages are:**

- Independent of modulation format since the input bits are only rearranged to obtain the output bits.
- Transparent to wavelength.
- High speed operation.
- Possibility to integrate.

Possible improvements:

- Use of Runlength Limited line-coding to improve code rate and to reduce the required hardware.

Other application:

- All optical Add/Drop Node

A Software Implemented Spread Spectrum Modem based on two TMS320C50 DSPs

Eric Kooistra

Eindhoven University of Technology, Faculty of Electrical Engineering, Telecommunications Division, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, fax: +31-(0)40-2455197

Abstract

A BPSK/DS-CDMA modem is described. Many functions of this spread spectrum modem are implemented in software on a DSP board. Constraints for proper IF receiver operation are given, plus BER measurement results for a prototype and a frequency-synthesizer resolution enhancement algorithm.

Brief description of the poster session

The poster session describes the design of a direct-sequence spread spectrum modem based on a signal processing board that contains a D/A-converter, an A/D-converter, and two digital signal processors (DSP) [2,3]. The spread spectrum signal is generated at baseband by multiplying each bit with the pseudo-noise code, so $T_b=N_cT_c$ where N_c is the code length [chips]. The modulation scheme applied is BPSK with square-root raised-cosine pulse shaping of the chips (p1). The IF receiver demodulates the input signal at an intermediate frequency (e.g. f_{IF}=128 kHz) through coherent subsampling, using a software Costas loop for carrier tracking. This subsampling creates a baseband replica by spectrum folding (p2). The despreading also occurs in software by means of a serial search code acquisition scheme, and a delay-lock loop (DLL) for code tracking [6]. The proper operation of the DLL and Costas loop is determined by three constraints that concern the resolution of the sample frequency f_{fs} and chip rate f_c, the loop bandwidth B_L and the loop sample frequency f_s [1,4,5] (p3). A prototype version of the modem was built using one TMS320C25 DSP. This prototype can achieve 16 kchip/s and measurement results are presented of its bit error rate performance (p4). A new design of the modem uses two TMS320C50 DSPs. On this new modem a multiple correlator scheme can be implemented, in order to increase the sample frequencies of the loops, so $f_s\gg f_b$ instead of $f_s=f_b$. Furthermore, thanks to a smooth resolution enhancement algorithm, both the sampling frequency and the chip rate can be generated by means of the timers of the DSPs. The algorithm has a recursive structure and is implemented by means of a look-up table. The 0's and 1's in this table state whether the next timer period has to be T or $T+1$ [ΔT_{min}], and the size of the table determines the resolution enhancement factor (e.g. a factor 7). Together the two DSPs are expected to be sufficiently powerful to implement in software all functions of the baseband transmitter and IF receiver for chip rates up to 64 kchip/s (p5).

References

[1] Best, E.B., *Phase-locked loops, theory, design and applications*, second edition, McGraw-Hill, 1993

[2] Kamperman, F.L.A.J., *Design study for a spread-spectrum picoterminal satellite network and the realization of a modem*, IVO Report EUT, Telecommunications Division, 1993

[3] Kooistra, E., *On the development of a spread spectrum transceiver based on a digital signal processor*, IVO Report EUT, Telecommunications Division, 1995

[4] Lindsey, W.C., C.M. Chie, *A survey of digital phase-locked loops*, Proceedings IEEE, Vol. 69, pp. 410-431, Apr. 1981

[5] Natali, F.D., *All-digital coherent demodulator techniques*, Proc. Int. Telemetering Conf., Vol. VIII, pp. 89-108, Oct. 1972

[6] Ziemer, R.E. and R.L. Peterson, *Digital communications and spread spectrum systems*, Macmillan Publishing Company, 1985

The BPSK / DS-CDMA transceiver

■ The baseband transmitter and modulator

$$v_o(t) = Ad(t)p(t) \cos(2\pi f_{IF}t)$$

data d_i f_b → **Data signal spreader** | code p_j f_c → **Pulse shaping filter** f_{DAC} → **DAC** → **LPF$_T$** → **BPSK modulator** →

p_j (into Data signal spreader)

Transmit clock (into DAC)

f_{IF} → **IF carrier oscillator** (into BPSK modulator)

C25-transceiver   *analog*

■ The IF receiver

To analog amplifiers in the receiver

AGC

Frequency synthesizer

f_{fs}

Receiver control

NCO $2f_c$

→ **BPF$_R$** v_i f_{IF} → **ADC** → **Digital I,Q down conversion** → **Pulse shaping $H_R(z)$** → **Despreading** → data f_b

C25-transceiver

Tasks:
 − Demodulation
 − Despreading

→ data bits

Demodulation by subsampling

- ## Digital I,Q downconversion, $\quad f_{fs} = \dfrac{4}{2m+1} f_{IF}$

Time domain:

$\bullet$ = sample moment

T_{fs}

T_{fs}

T_{IF}

t

$T_{fs} = \dfrac{T_{IF}}{4}$

$T_{fs} = 3 \dfrac{T_{IF}}{4}$

Frequency domain:

$-f_{fs}$ \quad 0 \quad f_{fs} \quad $2f_{fs}$ \quad nf_{fs}

f

m=0

m=1

m=2

Example

m=3

m=4

m=5

Baseband replica

93
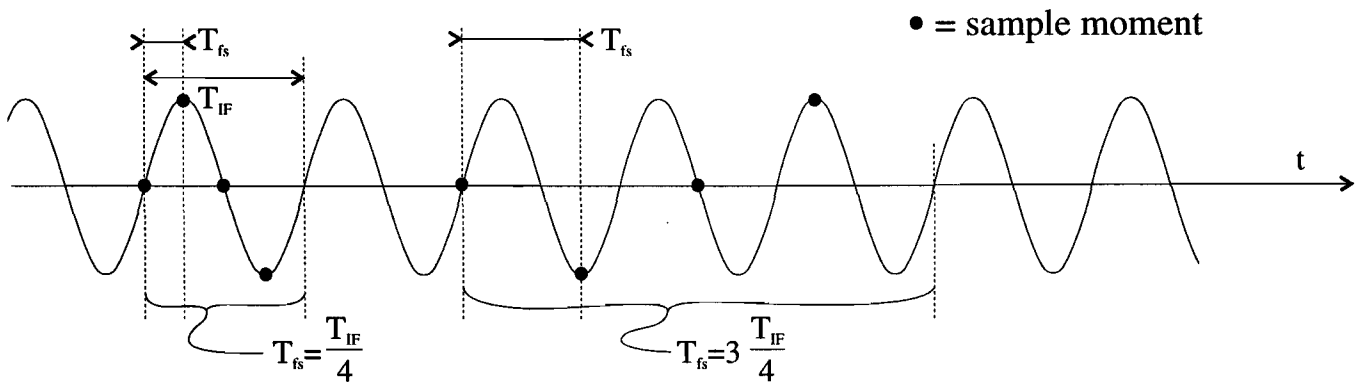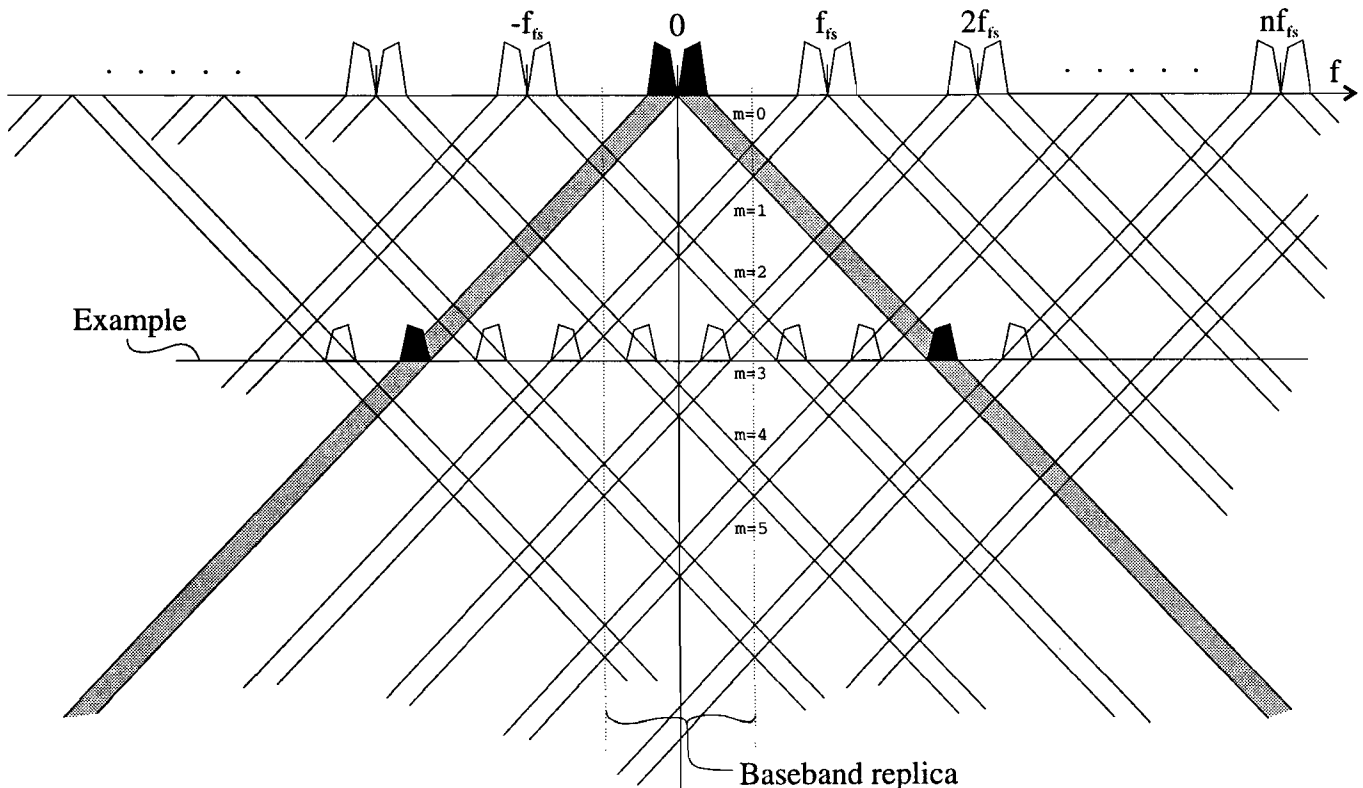
Despreading by correlating with the code

- **Functions:**
    - 1   Code acquisition (coarse)
    - 2a   Code tracking delay-lock loop, DLL (fine)
    - 2b   IF carrier tracking Costas loop

- **Constraints for the loops:**
    - DLL: $\Delta T_{min} < \dfrac{1}{50 N_d f_c}$, $\dfrac{\Delta T_{min}}{T_c^2} < 0.5 B_L$, $B_L < \dfrac{f_s}{10}$
    - Costas: $\Delta f_{min} < \dfrac{f_s}{50}$, $\Delta f_{min} < 0.5 B_L$, $B_L < \dfrac{f_s}{10}$

Performance of the C25-transceiver

- **BER measurement results:**
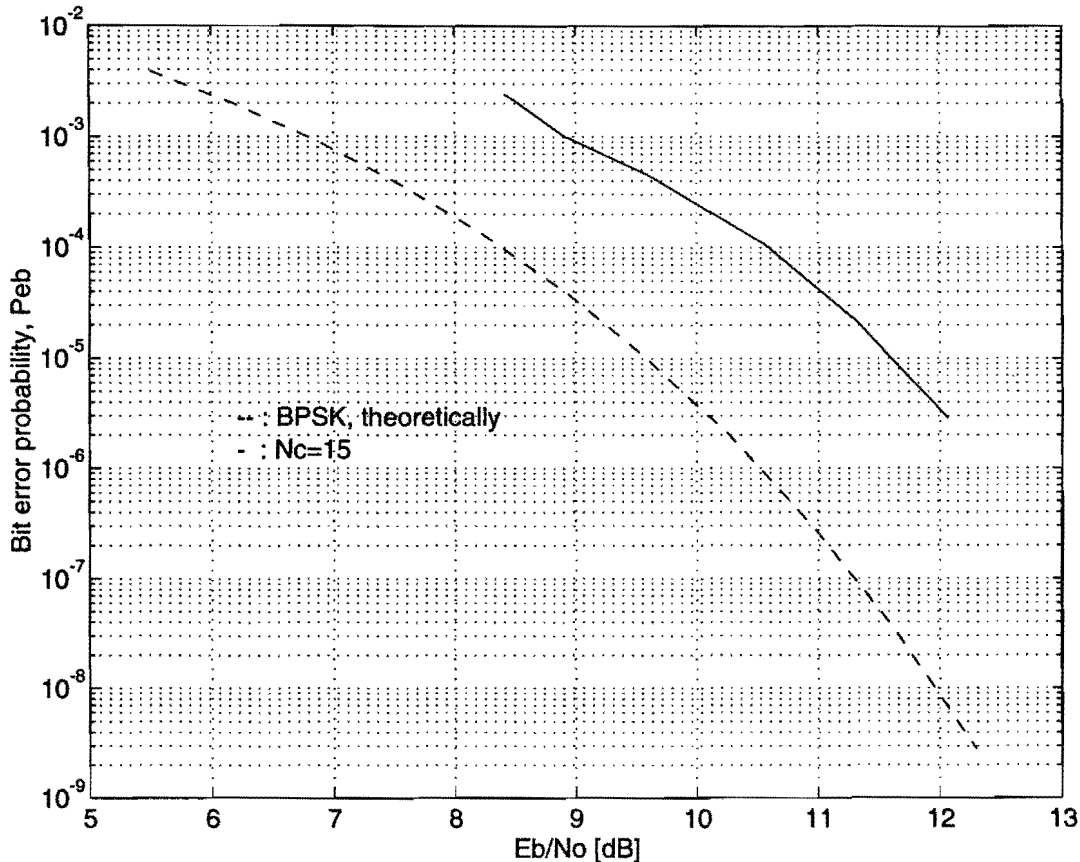
- **Chip rate :** $f_c \leq 16 \text{ kchip/s}$

   (due to limited processor capabilities)

- **Bit rate :** $f_b \geq 250 \text{ bit/s}$

   (due to the minimum update frequency f_s of the control loops)

   $\longrightarrow$ code length : $N_c = 63$ chips

The C50-transceiver

■ Schematic diagram of the DSP board:

Receiving   Transmitting

ADC   DAC

f_{fs}

timer   timer

TMS320 C50   TMS320 C50
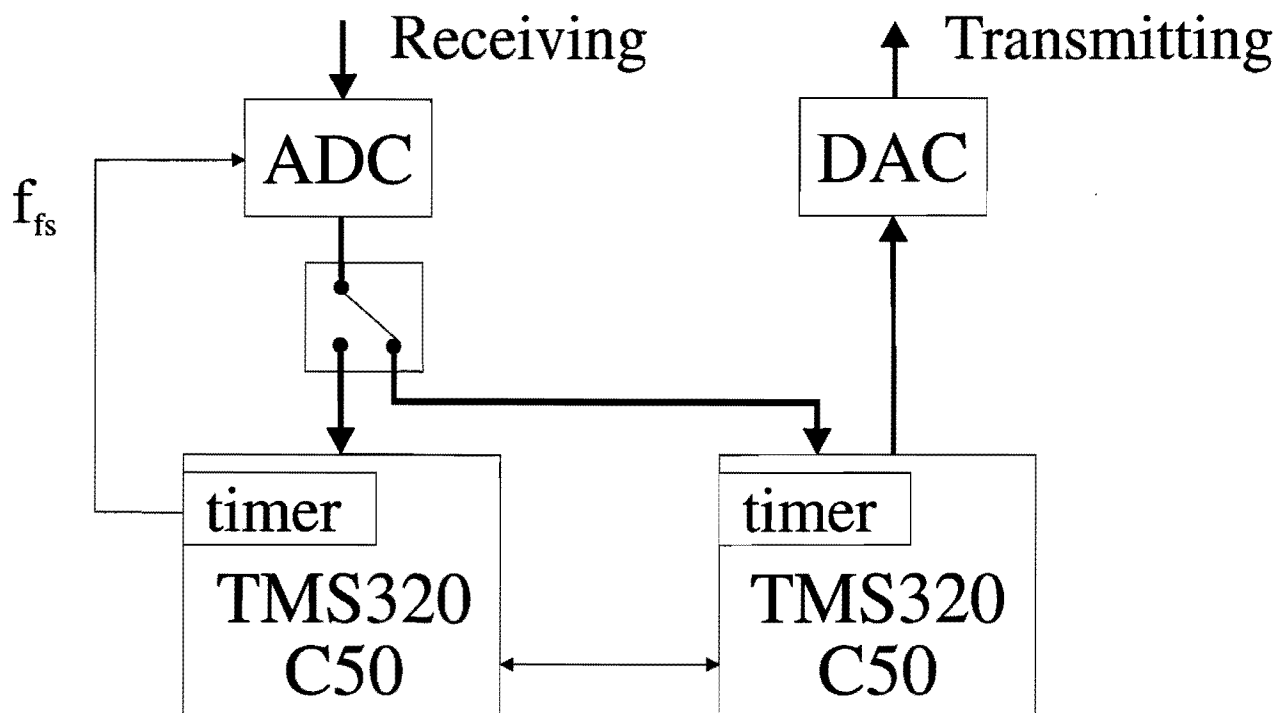
■ Expected performance:
  − $f_c = 64$ kchip/s
  − Increased f_s by means of multiple correlators

■ Particularities:
  − ADC samples switch
  − Two timers for generating T_c and T_{fs}

  → Resolution enhancement algorithm, e.g.:

$$
\begin{array}{ccccccc}
0, & 0, & 0, & 0, & 0, & 0, & 0 \\
0, & 0, & 0, & 0, & 0, & 0, & 1 \\
0, & 0, & 1, & 0, & 0, & 0, & 1 \\
0, & 1, & 0, & 1, & 0, & 0, & 1 \\
0, & 1, & 1, & 0, & 1, & 0, & 1 \\
0, & 1, & 1, & 1, & 0, & 1, & 1 \\
0, & 1, & 1, & 1, & 1, & 1, & 1 \\
1, & 1, & 1, & 1, & 1, & 1, & 1 \\
\end{array}
$$

Performance analysis of a Pico-Cellular CDMA network based on Hybrid DS/SFH, DS and SFH in a Shadowed Rician Fading channel

Omar Fatah, Student Member, IEEE*
René Rooimans, Member, IEEE**
Ramjee Prasad, Senior Member, IEEE*

*Telecommunications and Traffic Control Systems Group
Delft University of Technology
P.O. Box 5031, 2600 GA Delft
The Netherlands
Fax: +31-15-781774

**Electriciteitsbedrijf Zuid-Holland
Telecommunications & Information Systems
P.O. Box 909, 2270 AX Voorburg
The Netherlands
Fax +31-70-3811906

ABSTRACT - The performance analysis in terms of bit error rate, throughput and delay of a hybrid Direct Sequence (*DS*)/Slow Frequency Hopping (*SFH*), *DS* and *SFH* is analyzed for a Rician and Lognormal fading channel. Expressions for Bit Error Rate (BER), throughput and delay are derived for *BPSK* and *QPSK* modulations. The effect of Selection Diversity (*SD*) and Forward Error Correction (*FEC*) coding on the performance is also investigated.

The performance of hybrid *DS/SFH* has been compared with that of pure *DS* and *SFH* for a given delay spread T_m and fixed bandwidth B_T. Multipath, multiple access and thermal noise are considered here as the only sources of interference. Also, a comparison in terms of bit error rate has been made between the shadowed Rician fading channel and the Rician fading channel without shadowing.

1. INTRODUCTION

In recent years there has been a steadily growing interest in wireless communication systems. The main causes of this rapid expansion are the development of affordable wireless technologies.

Beside applications of these wireless services in an outdoor environment, there is an increasing development of wireless systems in an indoor environment. Especially in the office environment the application of these systems, such as wireless computer networks and cordless telephones, offers the advantages of flexibility and saving of costs compared to the cable systems.

Spread Spectrum (*SS*) multiple access techniques can be used in order to provide the possibility for users to transmit simultaneously. Code Division Multiple Access (*CDMA*), of which Direct Sequence (*DS*) and Slow Frequency Hopping (*SFH*) systems are two implements, provide a good answer to the problem of multiple access communication systems in fading environments.

DS is a good mean to combat the multipath fading [1]. However, *DS/SS* is known to suffer from the near-far effect. On the other hand, *SFH* is a good solution to cope with the multiple access interference. Besides, it is less sensitive to the near-far effect [2]. As a consequence, the combination of *DS* with *SFH* in a Hybrid system is an attractive solution for combating the multipath effect, the multiple access interference and reducing the near-far problem [1,2].

The objective of this paper is two fold. Firstly the performance of a hybrid DS/SFH is investigated in a pico-cellular environment which can be characterized by a Rician and Lognormal fading channel. Secondly, a comparison has been made between hybrid DS/SFH, DS and SFH CDMA.

This paper is organized as follows. Section 2 gives the system description. The Bit Error Rate (BER) is studied in section 3. Section 4 shows the numerical results for the BER. Throughput and delay are discussed in section 5. In section 6 we find the numerical results for the throughput and delay. Finally, conclusions are given in section 7.

2. SYSTEM DESCRIPTION

The system is a star connected network with K users. All of them are at the same distance from the base station. We study the performance of the system when the base station is in receiving mode.

2.1 Transmitter model

For the transmitter model, we follow the same methodology as the one given in [1]. After appropriate filtering, transmitted signal for *QPSK* becomes:

$$s_k^Q(t) = \sqrt{2P}a_{c,k}(t)b_{c,k}(t)$$
$$\cdot \cos\left[\omega_c t + \omega_k(t)t + \alpha_k(t) + \theta_k\right] \quad (1)$$
$$+ \sqrt{2P}a_{s,k}(t)b_{s,k}(t)$$
$$\cdot \sin\left[\omega_c t + \omega_k(t)t + \alpha_k(t) + \theta_k\right]$$

where P is the transmitted power and $a_{x,k}$ $(x=c,s)$ is the spreading sequence. The quantity $\alpha_k(t)$ represents the phase shift introduced by the frequency hopper when it switches from one frequency to another. The results obtained in case of QPSK can also be used in case of *BPSK* modulation by setting $b_{s,k}(t) = 0$, $b_{c,k}(t) = b_k(t)$ and $a_{c,k}(t) = a_k(t)$ in equation (1).

2.2 Channel model

The link between the *kth* user and the base station is characterised by a lowpass equivalent impulse response given by :

$$h_k(t) = \sum_{l=1}^{L} \beta_{kl}\delta(t - \tau_{kl})exp(j\gamma_{kl}) \quad (2)$$

where *kl* refers to path *l* and user *k*. We assume there are *L* resolvable paths associated to each user. The *lth* path of the *kth* user is characterized by three random variables : the gain β_{kl}, delay τ_{kl} and the phase γ_{kl}. We make the following assumptions concerning the channel:

- The path gain β_{kl} is Rician and Log-normal distributed;
- The delay τ_{kl} is assumed to be uniformly distributed random variable over [0,T];
- The phase factor γ_{kl} is uniformly distributed over $[0,2\pi]$;
- The random variables β_{kl}, γ_{kl} and τ_{kl} are independent for different values of *k* and *l*;
- The channel is assumed to be of the slow fading type, which means that the different random variables do not change considerably over the duration of one bit.
- The *rms* delay spread T_m is less than the bit duration *T* in order to avoid intersymbol interference;

In a radio environment fast and slow variations of the received signal strength occur. The fast variations are described by a Rician *PDF* [3]. The *PDF* of the signal amplitude conditional on P_0 ($P_0 \approx (1/2)A^2$) is written as:

$$P_\beta(r|P_0) = \frac{r}{\sigma_r^2}exp\left[-\frac{r^2 + 2P_0}{2\sigma_r^2}\right] \quad r \geq 0, \ A \geq 0 \ (3)$$
$$\cdot I_0\left[\frac{\sqrt{2P_0}r}{\sigma_r^2}\right]$$

where *A* and σ_r represent the Line Of Sight (*LOS*) and scattered components of the radio signal, respectively. *R* is called the Rician factor and is defined as the ratio of the *LOS* power $A^2/2$ and the scattered power σ_r^2. The slow fluctuations in the received mean area power P_0 caused by signal obstruction, can be modelled by lognormal *PDF*:

$$f_L(P_0) = \frac{1}{\sqrt{2\pi} \cdot \sigma \cdot P_0}exp\left\{-\frac{[ln(P_0) - m]^2}{2\sigma^2}\right\} \quad (4)$$

where σ and $\xi=exp(m)$ are the logarithmic variance and area mean power, respectively. The value of σ is the most of time given in *dB*, where :

$$\sigma_{Nat} = \sigma_{dB}\frac{ln(10)}{10} = 0.23\sigma_{dB}$$. The unconditional *PDF* can be obtained from equations (3) and (4) [4]:

$$P_\beta(r) = \int_0^\infty \frac{r}{\sigma_r^2}exp\left[-\frac{r^2 + 2P_o}{2\sigma_r^2}\right] \cdot I_0\left[\frac{r \cdot \sqrt{2P_0}}{\sigma_r^2}\right]$$
$$\cdot \frac{1}{\sqrt{2\pi}\sigma P_0}exp\left[-\frac{(ln(P_0) - m)^2}{2\sigma^2}\right]dP_0 \quad (5)$$

2.3 Receiver model

We assume that a received signal is composed of the contributions of the different users, their different paths and additive white Gaussian noise. Therefore, we have for *QPSK* the following received signal:

$$r(t) = \sqrt{2P}\sum_k\sum_l \beta_{kl} a_{c,k}(t-\tau_{kl})b_{c,k}(t-\tau_{kl})$$
$$\cdot\cos\left[\omega_c t + \omega_k(t-\tau_{kl})t + \alpha_k + \theta_k + \gamma_{kl}\right]$$
$$+ \sqrt{2P}\sum_k\sum_l \beta_{kl} a_{s,k}(t-\tau_{kl})b_{s,k}(t-\tau_{kl}) \qquad (6)$$
$$\cdot\sin\left[\omega_c t + \omega_k(t-\tau_{kl})t + \alpha_k + \theta_k + \gamma_{kl}\right]$$
$$+ n(t)$$

where $n(t)$ is white Gaussian noise with single-sided power spectral density N_0 [W/Hz].
we define the following correlation functions [1,5]:

$$R_{lk}^{xy}(\tau_{kl}) = \int_0^{\tau_{kl}} a_{x,l}(t)a_{y,k}(t-\tau_{kl})dt \text{ and }$$

$$\hat{R}_{lk}^{xy}(\tau_{kl}) = \int_{\tau_{kl}}^T a_{x,l}(t)a_{y,k}(t-\tau_{kl})dt \text{ in which x and y}$$

may be both c or s. In these definitions the periodicity has been taken into account. We also define the following parameters:

$$X_k^{xy} = \sum_l \beta_{kl} f[\Psi_{kl}]R_{lk}^{xy}(\tau_{kl}) \text{ and}$$

$$\hat{X}_k^{xy} = \sum_l \beta_{kl} f[\Psi_{kl}]\hat{R}_{lk}^{xy}(\tau_{kl}). \text{ When } x=y, f(\cdot) \text{ is a co-}$$

sine function else it is a sine function. If $k=l$ we take $l \neq j$. So, after dehopping and despreading, we have for the in-phase component:

$$z_{0,c} = \sqrt{\frac{P}{8}}T\beta_{1j}b_{c,1}^0$$
$$+ \sqrt{\frac{P}{8}}\sum_k\left[d_1\left(b_{c,k}^{-1}\right)b_{c,k}^{-1}X_k^{cc} + d_2\left(b_{c,k}^0\right)b_{c,k}^0\hat{X}_k^{cc}\right. \qquad (7)$$
$$\left. + d_1\left(b_{s,k}^{-1}\right)b_{s,k}^{-1}X_k^{cs} + d_2\left(b_{s,k}^0\right)b_{s,k}^0\hat{X}_k^{cs}\right] + v_c$$

where $b_{c,k}^{-1}$ and $b_{c,k}^0$ are the previous and the current bit of in-phase data stream of the kth user. In case of $BPSK$ receiver, we do not have a quadrature component ($b_{s,k}(t)=0$).

3. STUDY OF THE BER

We follow the similar methodology as the one presented in [2], [6] and [7]. The conditional bit error probability, which is computed as the mean of several situations, corresponding to the possible hit situations produced by the multiple access interferers, is then given by:

$$P_e(k|\beta) = \sum_{n_i=0}^{k-1} P_e(n_i|\beta)P(n_i) \qquad (8)$$

where $P_e(n_i|\beta)$ is the conditional bit error rate assuming there are n_i active interferers out of $k-1$, and $P(n_i)$ is the probability of having n_i active interferers out of $k-1$. For random hopping patterns with a number of frequencies equal to q, the probability that any two users use the same frequency is given by $1/q$. So, $P(n_i)$ is given by

$$P(n_i) = \begin{bmatrix} k-1 \\ n_i \end{bmatrix}\left[\frac{1}{q}\right]^{n_i}\left[1-\frac{1}{q}\right]^{k-1-n_i} \text{ which is a binomial}$$

distribution. The probability P_e, assuming n_i interferers is defined as the probability that the decision variable is lower than 0 assuming a "1" is transmitted (or conversely). It is the same for each of the two components (in-phase and quadrature) in case of $QPSK$. We get then:

$$P_e^Q = P\left(z_{0,c} < 0|b_{c,1}^0 = 1\right) = P\left(z_{0,s} < 0|b_{s,1}^0 = 1\right) \text{ where}$$

$z_{0,c}$ is given by:

$$z_{0,c} = A_0\beta_{1j}b_{c,1}^0 T + I_{c,mp} + I_{c,ma} + v_c \qquad (9)$$

in which $I_{c,mp}$ and $I_{c,ma}$ are the multipath and the multiple access interferences of the in-phase component. These interference parameters are modelled as Gaussian random variables with zero mean and variances $\sigma_{c,mp}^2$ and $\sigma_{c,ma}^2$, respectively. The bit error rate P_e for $BPSK$ is defined as:

$$P_e^B = P\left(z_0 < 0|b_1^0 = 1\right) \text{ where } z_0 \text{ is given by:}$$

$$z_0 = A_0\beta_{1j}b_1^0 T + I_{mp} + I_{ma} + v \cdot$$

The power of the multipath of the in-phase and quadrature components is given by:

$$\sigma_{c,mp}^2 = \sigma_{s,mp}^2 = \frac{P}{8}(L-1)\left(\sigma_r^2 + exp\left(m+\frac{\sigma^2}{2}\right)\right)\frac{2T^2}{3N} \qquad (10)$$

and the power of the multi-user interference is given by:

$$\sigma_{c,ma}^2 = \sigma_{s,ma}^2 = \frac{P}{8}n_i L\left(\sigma_r^2 + exp\left(m+\frac{\sigma^2}{2}\right)\right)\frac{2T^2}{3N} \qquad (11)$$

for both QPSK and BPSK. The bit error rate can be computed as a function of the bit energy E_b to total noise power spectral density N_0, assuming the value of β_{1j}. The bit error rate has to be averaged over the distribution of β_{1j}. If we assume that the total noise is Gaussian with zero-mean and variance $\sigma_{tot}^2 = \sigma_{mp}^2 + \sigma_{ma}^2 + \sigma_v^2$, we can use the bit error rate given in [8] as:

$$P_e = \frac{1}{2}erfc\left[\sqrt{\frac{E}{N_T}}\right] = \frac{1}{2}erfc\left[\frac{A_0 r T_b}{\sqrt{2}\sigma_{tot}}\right] \qquad (12)$$

We define the instantaneous received power P_r over one period of the carrier as follows: $P_r = \frac{r^2}{2}$.

We assume that the multipath interference, associated with the reference user, is due to L paths instead of $L-1$. So, in case of $BPSK$, we can write the conditional bit error rate given a number of n_i active interferers as:

$$P_e^B(n_i|P_r) = \frac{1}{2}erfc\left\{\left[\left[\frac{E_b}{N_0}\right]^{-1}\cdot\frac{1}{2P_r} + \frac{2L}{3NP_r}\right.\right.$$
$$\left.\left.(1+n_i)\left(\sigma_r^2 + exp\left(m+\frac{\sigma^2}{2}\right)\right)\right]^{0.5}\right\} \qquad (13)$$

99

Assuming that $N_q = 2N_b$ we find for *QPSK* the following expression:

$$P_e^Q(n_i | P_r) = \frac{1}{2} \text{erfc}\left\{ \left[\left[\frac{E_b}{N_0} \right]^{-1} \cdot \frac{1}{2P_r} + \frac{L}{3NP_r} \right] \right. \quad (14)$$
$$\left. \cdot (1 + n_i) \left(\sigma_r^2 + \exp\left(m + \frac{\sigma^2}{2} \right) \right) \right]^{0.5} \right\}$$

The bit error rate is given as a function of the mean received bit energy E_b to noise N_0 ratio.

3.1 BER in case of SD

The process of choosing the strongest signal from a set of M signals carrying the same information, is called selection diversity (SD) of order M. The M signals are assumed to be equally distributed and mutually independent.
The Cumulative Density Function (*CDF*) of the M signals subject to both Rician fading and Log-Normal fading is given in [4] by:

$$P_\beta(r) = \int_0^\infty \left(1 - Q\left(\frac{\sqrt{2P_0}}{\sigma_r}, \frac{r}{\sigma_r} \right) \right) \frac{1}{\sqrt{2\pi}\sigma P_0} \quad (15)$$
$$\cdot \exp\left(-\frac{(\ln(P_0) - m)^2}{2\sigma^2} \right) dP_0$$

The *CDF* of β_{max} is the *CDF* of equation (15) raised to the power of M. The *PDF* of β_{max} can now be obtained by differentiating the *CDF* with respect to r:

$$P_{\beta_{max}}(r) = M\left[\int_0^\infty \left(1 - Q\left(\frac{\sqrt{2P_0}}{\sigma_r}, \frac{r}{\sigma_r} \right) \right) \frac{1}{\sqrt{2\pi}\sigma P_0} \exp\left(-\frac{(\ln(P_0) - m)^2}{2\sigma^2} \right) dP_0 \right]^{M-1}$$
$$\cdot \int_0^\infty \frac{r}{\sigma_r^2} \exp\left[-\frac{r^2 + 2P_0}{2\sigma_r^2} \right] \cdot I_0\left[\frac{r\sqrt{2P_0}}{\sigma_r^2} \right]$$
$$\cdot \frac{1}{\sqrt{2\pi}\sigma P_0} \exp\left[-\frac{(\ln(P_0) - m)^2}{2\sigma^2} \right] dP_0 \quad (16)$$

the average bit error rate is given as a function of the mean energy to white noise ratio by:

$$P_e = \int_0^\infty \sum_{n_i=0}^{k-1} P_e^{B,Q}(n_i | P_r) P(n_i) p_{\beta_{max}}(P_r) dP_r \quad (17)$$

where $P_r = r^2/2$ and we used the fact that $p_{\beta_{max}}(P_r)dP_r = p_{\beta_{max}}(r)dr$.

3.2 BER in case of FEC coding

In this subsection the bit error rate is considered when Forward Error Correction (FEC) block codes are used. An *(n,k) FEC* block code is a *FEC* code which encodes k data bits into n encoded bits called a block. The code can correct $t = (d-1)/2$ errors per block [9], if d is the Hamming distance. A specific code used here is the *(23,12) Golay* code. The Hamming distance of this specific code

is given by $d_{GC} = 7$ which can correct $t = 3$ errors per block.
The probability of having m errors in a block of n bits is given by $P(m,n) = \binom{n}{m} \cdot P_e^m (1 - P_e)^{n-m}$ where P_e $(=P_e(n_i|P_r))$ is the conditional bit error rate at the hard decision output. The probability of having more than t errors in a code block of n bits is $P_{ec} = \sum_{m=t+1}^{n} P(m,n)$. An approximation for the bit error rate after decoding, P_{ec1}, is given in [8] by:

$$P_{ec1} = \frac{1}{n} \sum_{m=t+1}^{n} m \binom{n}{m} \cdot P_e^m (1 - P_e)^{n-m} \quad (18)$$

With the substitution of equation (18) in equation (8), we get the average bit error probability:

$$P_e = \int_0^\infty \sum_{n_i=0}^{k-1} P_{ec1}(n_i | P_r) P(n_i) p_{\beta_{max}}(P_r) dP_r \quad (19)$$

4. BER RESULTS

4.1 Assumptions

In this paper we present the numerical results of the average bit error rate in a *single Pico-Cellular* environment. The performance is given in terms of *BER* of hybrid *DS/SFH*, *DS* and *SFH* techniques with selection diversity (*SD*) and a (23,12) Golay coding. In order to get a good idea about the performance of *BPSK* compared to that of *QPSK* modulation scheme we will keep the bandwidth, the bit rate and the *rms* delay spread fixed. This means that qNR_b has to be the same for both modulation techniques. Since the bit rate in both systems should be the same, we will have $T_q = 2T_b$, where T_b and T_q are the bit duration for *BPSK* and *QPSK* respectively. Assuming the same number of frequencies q for these two modulation types, we finally have: $N_q = 2N_b$. In this considered *Pico-Cellular* system, the standard deviation of the Log-Normal distribution is taken equal to 4 *dB* and the Rice factor R equal to 6.8 dB [9]. The fixed bandwidth is taken equal to 80 *MHz* ($B_T = 80$ *MHz*) and the *rms* delay spread T_m equal to 150 *ns*. We will consider a number of users $K=15$, unless stated otherwise. For a hybrid *DS/SFH* technique, the number of frequencies in the hopping pattern q is taken equal to 10, the bit rate R_b is 64 *kbits/s* (in case of *FEC* coding, the bit rate R_b is 128 *kbits/s*), the number of resolvable paths L is 2 and the code sequence period N is 127 and 255 in case of *BPSK* and *QPSK* (in case of *FEC* coding, it is 63 and 127) respectively. In case of pure *DS* technique, the number of frequencies is reduced to 1, the bit rate is the same as in case of hybrid

DS/SFH, the number of resolvable paths is 13 and the code sequence period is 1270 and 2550 for *BPSK* and *QPSK* (in case of *FEC* coding, it is 630 and 1270) respectively. In case of pure *SFH*, the code sequence period is reduced to 1, the number of resolvable paths is also reduced to 1, the bit rate is the same as in case of hybrid *DS/SFH* and the number of frequency hopping is 1270 and 2550 for *BPSK* and *QPSK* (in case of *FEC* coding, it is 630 and 1270) respectively.

4.2 Results
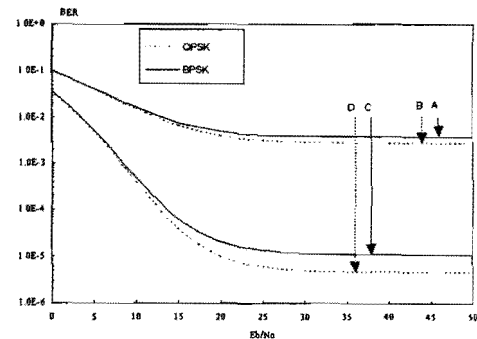
Figure 1: *Effect of SD on the BER of hybrid DS/SFH system for BPSK and QPSK for $\sigma=4dB$. A,B:M=1; C,D:M=3*

In figure 1, we can see in one plot the effect of *SD* (*M*=3) on the performance of the *BER* compared to that without selection diversity (*M*=1). Using *SD* means that we select the largest signal of all the signals we receive, which means that this signal has the largest power and that is why the performance with *SD* is much better than that without *SD*.
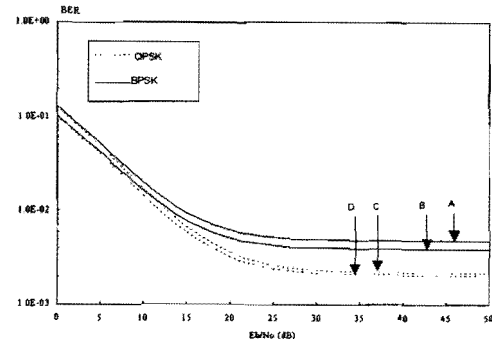
Figure 2: *Effect of Golay coding on the BER of hybrid DS/SFH system for BPSK and QPSK modulation without SD (M=1), for $\sigma=4$ dB; A,C: Golay coding; B,D: no coding*

The effect of *FEC* coding on the performance of the hybrid *DS/SFH* system for *BPSK* and *QPSK* without selection diversity is shown in figure 2, again with the constraint of a fixed bandwidth and

as a function of the ratio of the source bit energy and the spectral density N_0. It is seen that in figure 2, the performance with coding and without *SD* is not improving. This is due to the value of the spreading sequence period of $N=63$ in case of *BPSK* and $N=127$ in case of *QPSK*, which has worse correlation properties than $N=127$ and $N=255$ for *BPSK* and *QPSK* respectively.
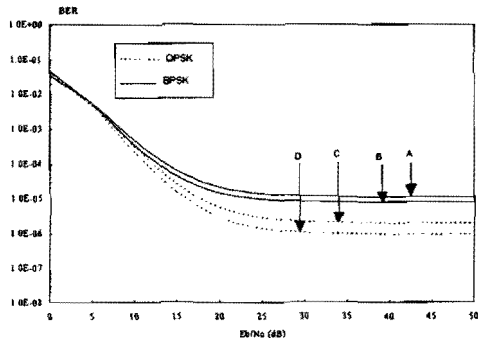
Figure 3: *Effect of SD (M=3) and FEC coding on the BER of hybrid DS/SFH system for BPSK and QPSK modulation for $\sigma=4$ dB; A,C: no coding; B,D: Golay coding*

In figure 3 is the effect of *FEC* coding on the performance of the hybrid *DS/SFH* system for *BPSK* and *QPSK* with selection diversity presented. With the constraint of a fixed bandwidth and as a function of the ratio of the source bit energy and the spectral density N_0. We can see that using *FEC* coding with *SD* in figure 3 improves the *BER*. We can also see that *QPSK* has a better performance than *BPSK* in case of fixed bandwidth, bit rate and delay spread. This is due to the larger spreading sequence in case of *QPSK* because of the bandwidth and bit rate constraint mentioned before. As a result of this larger spreading sequence there is less multipath and multi-user interference with *QPSK*. This is reflected in a lower bit error rate.
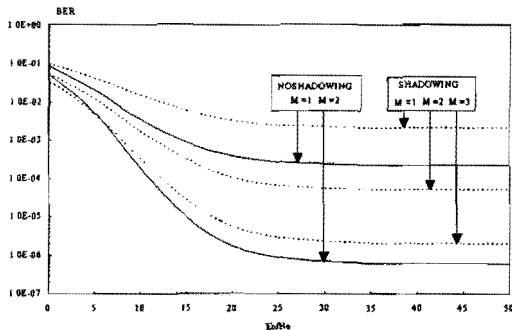
Figure 4: *Effect of SD on the BER of hybrid DS/SFH with ($\sigma=4$ dB) and without shadowing using QPSK as modulation technique*

101

In figure 4, the effect of *SD* in terms of the *BER* is given for selection diversity of order one, two and three in case of shadowing. The performance is given for a hybrid *DS/SFH* with shadowing compared to that without shadowing and without *SD*, using *QPSK* as modulation technique. We can see that without use of *SD* (*M=1*), the performance of the system without shadowing is better than that with shadowing. It is also seen that the influence of shadowing on the performance can be compensated with use of selection diversity. This is due to the fact that in case of shadowing with *SD*, we select the largest path of all paths in all antennas used. The received power in case of shadowing is less than that without shadowing and this is the result of the obstacles we find between sender and receiver.
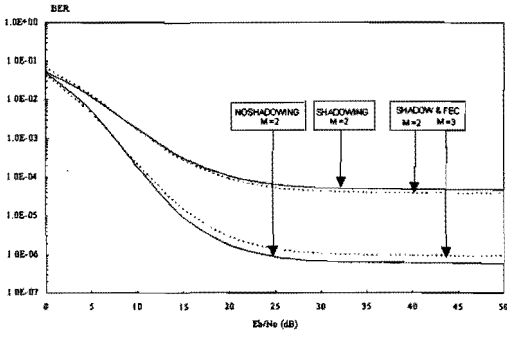
Figure 5: *Effect of SD and Golay coding on the BER of hybrid DS/SFH with (σ=4 dB) and without shadowing for QPSK*

In figure 5, the effect of *FEC* coding with *SD* is given in terms of the *BER*. The performance is done for a hybrid *DS/SFH* with *(σ=4 dB)* and without shadowing, using *QPSK* as modulation technique. It is seen that the performance of this system without shadowing is better than that with shadowing, when the same order of diversity is used. The power in case of shadowing is less than in case without shadowing because of the obstacles we have between sender and receiver. We can also see that the performance in case of shadowing is improving with use of *SD* and *FEC* coding and it is better than using only *FEC* coding. This is due to the selection of the largest path of all the paths, which means that there are less errors. Some of this errors will be corrected by use of *FEC* coding.

5. THROUGHPUT AND DELAY

The throughput is defined as the average number of successfully received packets per time slot. The throughput *S* for a slow fading channel is defined in [10] as:

$$S = \sum_{k=1}^{c} k \cdot P_k \cdot P'_{sk}(k) \text{ [packets/time slot]} \quad (20)$$

where:

$$P'_{sk}(k) = \int_{\beta_{max}} [1 - P_e(k|\beta_{max})]^{N_P} f(\beta_{max}) d\beta_{max} \text{ is the}$$

probability of one successful packet given *k* simultaneous transmitting users, and

$$P_k = \binom{C}{k} \cdot (P_{nt})^k \cdot (1 - P_{nt})^{C-k} \text{ is the probability that } k$$

packets are generated during a time slot. The probability that a user has a packet (new or retransmission packet) is given by $P_{nt} = \frac{G}{C}$, where the offered traffic *G* is defined as the average number of transmissions (new packets plus retransmitted packets) per time slot by *k* users.

The average delay is defined as the average number of time slots it takes for a packet to be successfully received. The probability of success per transmitted packet can be defined as *P(suc. per packet)* = $\frac{S}{G}$. The average delay *D* in an indoor environment, assuming negligible propagation delay can be obtained on the same basis as was done in [11,12]:

$$D = 1.5 T_p + \left(\frac{G}{S} - 1\right) \cdot \left(T_p + T_a + T_r\right) \quad (21)$$

where *(G/S - 1)* is the average number of retransmissions for a packet to be successfully received. The normalised average delay D_n is given by:

$$D_n = 1.5 + \left(\frac{G}{S} - 1\right) \cdot \left(1 + \frac{T_a}{T_p} + \frac{T_r}{T_p}\right) \quad (22)$$

6. THROUGHPUT AND DELAY RESULTS

6.1 Assumptions

In this subsection we use the same assumptions as in subsection 5.1.

6.2 Results

In figure 6 and figure 7 respectively, the throughput and delay of the hybrid *DS/SFH* is shown. The influence of the order of diversity *M* for a packet size *Np* is presented. An increase of the order of diversity *M* results in a higher throughput and lower delay. *QPSK* has a better performance than *BPSK* for fixed bandwidth, bit rate and delay spread. This is due to the fact that the multipath and multiple access terms are smaller in case of *QPSK* modulation as can be seen from equations (17) and (18).

102
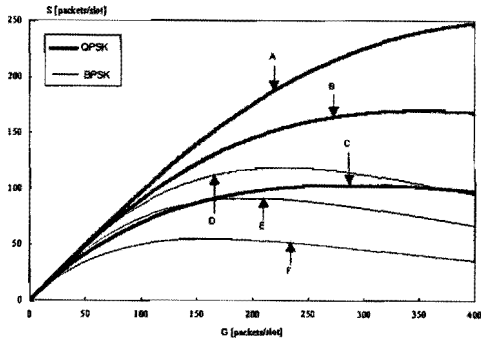
Figure 6: *Effect of SD on the throughput of hybrid DS/SFH for σ=4 dB, using QPSK and BPSK as modulation techniques. A,D:M=3; B,E:M=2; C,F:M=1*

Figure 7: *Effect of SD on the delay of hybrid DS/SFH for σ=4 dB, using QPSK and BPSK as modulation techniques. A,D:M=3; B,E:M=2; C,F:M=1*

Figure 8: *Effect of FEC coding with SD (M=3) on the throughput of hybrid DS/SFH for σ=4 dB, using QPSK and BPSK as modulation. A,C: no coding; B,D: Golay coding*

Figure 9: *Effect of FEC coding with SD (M=3) on the delay of hybrid DS/SFH for σ=4 dB, using QPSK and BPSK as modulation techniques. A,C: no coding; B,D: Golay coding*

Figure 8 and figure 9 respectively, show the comparison of the throughput and delay with and without coding for a hybrid *DS/SFH* system with selection diversity of order 3. *SD* with *FEC* coding yields the best result, with respect to the throughput and delay.

Figure 11: *Comparison in terms of the throughput between hybrid DS/SFH, DS and SFH for σ=4 dB, using QPSK and BPSK as modulation techniques with SD (M=3) and Golay coding. A,B: Hybrid; C,D: DS; E,F: SFH*
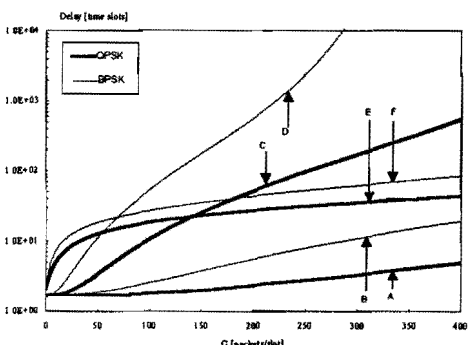
Figure 12: *Comparison in terms of delay between hybrid DS/SFH, DS and SFH for σ=4 dB, using QPSK and BPSK as modulation techniques with SD (M=3) and Golay coding. A,B: Hybrid; C,D: DS; E,F: SFH*

103

In figure 11 and figure 12 respectively, the throughput and delay of the hybrid *DS/SFH* technique is compared with the throughput and delay of the pure *DS* and *SFH* techniques, with *FEC* coding and selection diversity of order 3. The hybrid *DS/SFH* technique has the highest throughput and lowest delay performance. The maximum throughput for both *SFH* and *DS* is almost equal, however the throughput for *SFH* does not vary much over the large range of the offered traffic *G*.

7. CONCLUSIONS

The performance in terms of bit error rate, throughput and delay has been investigated for a *single Pico-Cellular* radio channel using hybrid *DS/SFH*, *DS* and *SFH CDMA* with *QPSK* and *BPSK* modulation. The performance has been evaluated using an analytical uniform delay profile. The radio channel in the *Pico-Cellular* environment is characterized by a shadowed Rician slow fading type.

From the performance results we have seen that for a fixed bandwidth B_T and fixed *rms* delay spread T_m the hybrid *DS/SFH CDMA* technique has a better performance than the pure *DS* and *SFH CDMA* techniques. We have also seen that under the assumption of fixed bandwidth and bit rate, *QPSK* yields a better performance than *BPSK*. Increasing the order of diversity *M* and using forward error correction (*FEC*) coding, both yield a better performance for all three techniques.

After comparing the performance of a shadowed Rician fading channel with that of a Rician fading channel in a *single Pico-Cellular* environment, it is observed that due to shadowing, the performance in terms of *BER*, throughput and delay is worse than that without shadowing. This effect can be compensated with use of *SD* and *FEC* coding.

The results presented in this report are valid under the assumption of uniform power delay profile, perfect power control and perfect synchronization and tracking at the receiver.

REFERENCES

[1]   R. Prasad, R. Rooimans and L. Vandendorpe, "Hybrid Slow Frequency Hopping/Direct Sequence Spread Spectrum Communications Systems with B- and QPSK Modulation in an indoor Wireless Environment", Proceedings of the fourth International Symposium on Personal, Indoor and Mobile Radio Communications, Yokohama, Japan, September 9-11 1993.

[2]   J. Wang and M. Moeneclaey,"Hybrid DS/SFH-SSMA with Predetection Diversity and Coding Over Indoor Radio Multipath Rician-Fading Channels", IEEE Transactions on Communications, Vol. 40, No. 10, October 1992.

[3]   J.T.S. Chia, "Radiowave propagation and handover criteria for microcells", British Telecom Technical Journal, vol.8, pp.50-61, October 1990.

[4]   H.S. Misser and R. Prasad, "Bit error probability evaluation of a microcellular spread-spectrum multiple access system in a Shadowed Rician fading channel", IEEE Transactions on VT, 1992.

[5]   M. Kavehrad and B. Ramamurthi, "Direct-Sequence Spread Spectrum with DPSK Modulation and Diversity for Indoor Wireless Communications", IEEE Transactions on Communications, vol. COM-35, February 1987, pp. 224-236, 11 refs.

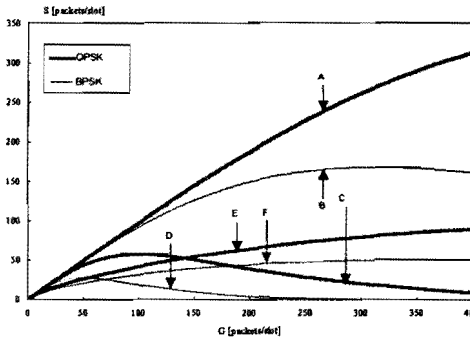[6]   J. Wang and M. Moeneclaey, "Hybrid DS/SFH Spread-Spectrum Multiple Access with Predetection Diversity and Coding for Indoor Radio", IEEE Journal on Selected Areas in Communication, VOL. 10, No. 4 May 1992.

[7]   E.A. Geraniotis, "Noncoherent Hybrid DS/SFH Spread-Spectrum Multiple-Access Communications" IEEE Transactions on Communications, Vol. Com-34, No. 9, September 1986.

[8]   J.G. Proakis, "Digital communications", second edition. McGraw-Hill International Editions, Computer Science Series, 1989.

[9]   H. S. Misser, "Bit error probability evaluation of a microcellular spread-spectrum multiple access system in a shadowed Rician fading channel", Vehicular Technology Society 42nd VTS conference frontiers of Technology, May 10-13, 1992, Denver, Colorado, pp.439-442.

[10]   C.A.F.J. Wijffels, H.S. Misser and R. Prasad, "A micro-cellular CDMA System over Slow and Fast Rician Radio Channels with Forward Error Correcting Coding and Diversity", IEEE Transactions on VT, vol.42, No.4, November 1993.

[11]   R. Prasad, C.A.F.J. Wijffels and K.L.A. Sastry, "Performance analysis of slotted CDMA with DPSK modulation diversity and BCH-coding in indoor radio channels", Archiv für Electonik und Uebertragungstechnik, vol.46, No.6, November 1992, pp.375-382.

[12]   L. Kleinrock and F.A. Tobagi, "Packet Switching in Radio Channels: Part I-Carrier Sense Multiple-Access Modes and their Throughput-Delay Characteristics", IEEE Transactions on Communications, vol. Com-23, No.12, December 1975, pp.1400-1416.

Slotted ALOHA and Unslotted Non-persistent ISMA for Indoor Pico Cellular Communication

Neeli R. Prasad, Student Member, IEEE
Jos Nijhof, Member, IEEE

Telecommunications and Traffic Control Systems group
Delft University of Technology
P.O. Box 5031, 2600 GA Delft
The Netherlands
Fax: +31-15-781774

ABSTRACT

The paper presents the throughput of Slotted ALOHA and unslotted non-persistent ISMA and spectrum efficiency of a single cell data network. These have been analysed using interference model in Rician fading environment and capture effect in a single cell structure. The results maybe of importance for the future Indoor Wireless communications (IWC).

1. Introduction

In the last few years the demand of IWC has grown significantly. This fast growing numbers of users demands efficient use of scarce frequency spectrum. Cellular radio systems has been classified into three categories: i) marcocells of 2 to 20 km diameter, ii) microcells of 0.4 to 2 km and iii) picocells of much smaller size (20 to 400 m) especially suited for indoor radio communications (e.g. offices, research laboratories, hospitals modern factories, university campuses, etc.).

This paper considers the situation where terminals communicating with a centrally located base station using packet radio. During this communication, terminals compete with each other to get access to the common channel, thus giving a reason to investigate, which multiple access protocols are suited for this required situation. Two protocols have been investigated for the performance in pico cellular systems for a single cell structure.

The rooms of 19th and 20th floors of Electrical Engineering, Delft University of Technology, The Netherlands, has been considered for this investigation [1]. The rooms are located on the both sides of 65 m long and 2 m wide, straight corridor. Each room is 4 m wide, 6 m long and 2.80 m high. These rooms are separated by a wall made of plaster board, rockwool and steel frames. The separating wall of corridor and rooms is made of bricks. The floor is made of inforced concrete of 50 cm and above the doors of each room, at the height of 2.10 m, a glass window runs up to the ceiling. There is an open stairway on the both end of the corridor leading to the upper and lower floors. Each room has been considered a cell, with a centrally located base station using packet radio. The investigation includes the channel where the test terminal and the interfering terminals are within a small cell, so the signals coming from both the test terminal and the interfering terminals are Rician distributed. Spectrum efficiency has also been looked into.

2. Indoor wireless Communication Channel

The terminals are assumed to be in one room of the 19th or 20th floor with base station in the centre. In this way, the amplitude of the signals can be modelled by Rician pdf (probability density function) [1]

$$f(A) = \frac{A}{\sigma^2} exp(-\frac{A^2 + s^2}{2\sigma^2})I_0(\frac{As}{\sigma^2}) \quad (1)$$

where A is the signal amplitude, σ^2 is the average fading power, s is the peak amplitude of the dominant received multipath component and $I_n(\cdot)$ is the modified Bessel function of the first kind and

n^{th} order. From (1) the pdf of the fast varying instantaneous power, $p = \frac{A^2}{2}$, is found as

$$f(p) = \frac{K+1}{p_0} \exp\left(-\frac{K+1}{p_0}p - K\right)$$
$$\times I_0\left(2\sqrt{\frac{K^2+K}{p_0}p}\right) \qquad (2)$$

where $p_0 = \frac{s^2 + 2\sigma^2}{2}$, is the local mean power and

K the Rician factor , which is defined as the ratio of the average power of the dominant multipath component and the average fading power receiver over the non-dominant paths

$$K = \frac{s^2}{2*\sigma^2} \qquad (3)$$

The values of K are 8.2, 7.6 and 3.8 dB.

3. Performance Analysis

The protocol performance has been measured in terms of throughput S in relationship with the offered traffic G. The efficiency of the protocol can be measured by the throughput and is defined as the fraction of time in which correct data packets are received.

3.1 Slotted ALOHA

All active terminals are assumed to transmit their messages to a single receiver over a common channel in packets of duration τ , regardless of the results of the competing terminals. An unsuccessful packet will be retransmitted after waiting a random number of slots. The channel is memoryless, i.e., a retransmitted packet experiences collisions uncorrelated with its previous attempts to capture the receiver [2].

If the number of packets generated in the network (for messages plus retransmissions) are binomial distributed, with mean generation rate λ packets per second, the mean offered channel traffic expressed in packets per time is

$$G = \lambda * \tau \qquad (4)$$

The probability of a test signal being overlapped by n other packets is then

$$Pr[n] = \binom{N}{n}\left[\frac{G}{N}\right]^n\left[1 - \frac{G}{N}\right]^{N-n} \qquad (5)$$

where G is the offered traffic as the average number of transmissions per time slot, n is the active number of users and N is maximum number of users in a cell. With capture effect

$$P_{cap}(z_0) = 1 - \sum_{n=1}^{N} Pr[n] Prob\{P_s / P_n < z_0\} \qquad (6)$$

The channel throughput S is given [2] by

$$S = G\left[1 - \sum_{n=1}^{N} Pr[n]F_{zn}(z_0)\right] \qquad (7)$$

with

$$F_{zn}(z_0) = \int_0^{z_0} dz \int_0^{\infty} f_{ps}(zw)f_{pnr}(w)wdw \qquad (8)$$

as distribution function to calculate the throughput in the presence of Rician interfering signals.

3.2 Unslotted Non-persistent ISMA

In Unslotted np-ISMA, the transmission only occurs when the terminal gets permission to transmit, i.e., if an idle signal is received on the signalling channel. If there is a busy signal, the transmission attempt will be unsuccessful. Such packets are rescheduled for later by putting them into retransmitting buffer. We assume that the arrivals of these packets are generated by a binomial process.

The throughput is given by using eq.(5) and (6)

$$S = \frac{exp(-dG) + \sum_{n=1}^{N} \frac{(n+1)}{\varepsilon^{n-1}} Pr[n] \exp\left[-\frac{\varepsilon^2}{2}\right] \int_0^{\infty} x^m \exp\left[-\frac{x^2}{2}\right] I_{n-1}[\varepsilon x] Q\left(\alpha, \sqrt{2z_0}x\right) dx}{1 + 2d + \frac{1}{G} exp(-dG)}$$

$$(9)$$

where d is the inhibit delay fraction [3,4].

4. Spectrum Efficiency

This section describes the indoor wireless data network structure for the 19th floor of the Electrical Engineering Department of the Delft University of Technology by evaluating the spectrum efficiency.

We know from [1] that the maximum failure probability for the 19th floor network is 1.46%. Therefore the capture probability can be obtained by

$$P_{cap}(z_0) = 1 - P_{fail}(z_0) \qquad (10)$$

So, $P_{cap}(z_0) = 98.54\%$ from eq. (10). Any network should provide the maximum channel capacity and the highest spectrum efficiency. We know from eq. (7) and (9) that the throughput is maximal for the maximum success probability. Therefore for the both protocols, $p_{cap}(z_0) = 98.54\%$, provides the maximum throughput.

Spectrum efficiency can be obtained by [6]

$$SE = \gamma \frac{P_{cap}(z_0)}{sC_u} \cdot \frac{L-H}{L}$$
$$\left[\frac{bits}{sMHz - km^2} \right] \qquad (11)$$

where the modulation speed γ represents the bit rate (R_b) to bandwidth (B_T) ratio, $\gamma = \frac{R_b}{B_T}$ with $B_T = W$, s is the cell area, L is the packet length and H is the length of the header. The number of user data bits per packet is L-H. The $P_{cap}(z_0)$ depends on certain values of the receiver threshold. Different values of threshold corresponds to a certain type of modulation. For example, with modulation type FSK $\left(f_b = \frac{R_b}{2} \right)$ and DPSK (M=8), γ amounts to 1 and 3 respectively.
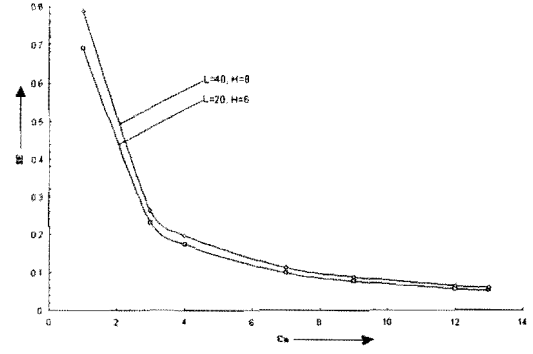
Figure 2 : Spectrum Efficiency SE versus cluster size C_u for $\gamma = 1$ and $s = 1 \ km^2$

Fig. 2 shows the plot of spectrum efficiency versus cluster size (Cu) and reuse distance (R_u). Cluster size and reuse distance are related as [5]

$$R_u = \sqrt{3C_u} \qquad (12)$$

the cluster size is given by [5]

$$C_u = i^2 + ij + j^2, \quad i, j \geq 0 \qquad (13)$$

with integer i and j. Eq.(13) gives $C_u = 1,3,4,7,9,12,13.....$

It can also be seen from fig. 2 that the spectrum efficiency is maximum for $C_u = 1$ and minimum reuse distance.

Fig. 3 shows the plot for spectrum efficiency versus packet length (L) for header length (H) as a parameter
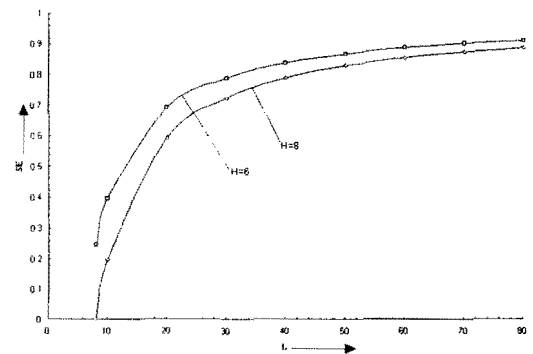
Figure 3: Spectrum efficiency SE versus packet packet length L for H=6 and 8 bytes in case of $C_u = 1$, $\gamma = 1$ and $s = 1 \ km^2$

Thus we know that at maximum spectrum efficiency highest throughput can be obtained for unit cluster size.

107

4. Conclusions

The throughput of the Slotted ALOHA and Unslotted Non-persistent ISMA, protocols have been analysed by considering a Rician distribution with capture effect.

Further study has been done to spectrum efficiency for indoor wireless data network structure for 19th floor.

When comparing spectrum efficiency of packet length of 20 and 40 bytes with header length 6 and 8 bytes respectively, the latter system offers higher maximum spectrum efficiency at cluster size 1, thus giving the highest throughput.

The present study will be extend by developing a model to Rician fading interfering signals and influence of multiple cell for indoor wireless communications, taking into consideration protocols like Unslotted ALOHA, Unslotted Non-persistent ISMA, Slotted Non-persistent ISMA and p-persistent ISMA.

References

[1]   F. van der Wijk, A. Kegel and R. Prasad, "Assessment of a Pico-Cellular System Using Propagation Measurements at 1.9 Ghz for Indoor Wireless Communications", IEEE Transactions on Vehicular Technology, vol 44, pp. 155-162, February 1995.

[2]   J.C. Arnbak and W.van Blitterswijk, "Capacity of Slotted ALOHA in Rayleigh-Fading Channels", IEEE Journal on Selected Areas in Communications, Vol. SAC-5, pp. 261-269, February 1987.

[3]   R. Prasad, "Performance analysis of Mobile Packet Radio Networks in Real Channels with Inhibit-Sense Multiple Access", IEE Proceedings-I, vol. 138, pp. 458-464, October 1991.

[4]   R. Prasad and C.Y Liu, "Throughput Analysis of some Mobile packet Radio protocols in Rician fading Channels", IEE Proceedings-I, vol. 139, pp. 297-299, June 1992.

[5]   R. Prasad and A. Kegel, "Improved Assessment of Interference Limits in Cellular Radio Performance", IEEE Transactions on Vehicular Technology, vol. 40, pp. 412-419, May 1991.

[6]   J. P. M. G. Linnartz, "Effects of Fading and Interference in Narrowband Land-Mobile Networks", Ph.D. Thesis, Delft University of Technology, The Netherlands, 1991.

Non-coherent detection of MACPM modulation techniques in mobile radio channel environment

Tomaž Javornik, Gorazd Kandus, Miha Mohorčič
Jožef Stefan Institute
Ljubljana, Slovenia
fax:+386 61 219 385
E-mail:tomaz.javornik@ijs.si

Abstract

In the following paper the simple non-coherent differential detection of multi-amplitude continuous modulation scheme is analysed in mobile radio channel. The proposed MACPM differential non-coherent detector consists of two parts: frequency discrimination detector and differential envelope detector. BER is simulated and plotted taking into consideration various distortions caused by mobile radio channel such as: Rayleigh fading with Doppler frequency shift and Gaussian noise. It was found out that non-coherently detected MACPM signals have worse performance than non-coherently detected CPM signals even though MACPM signals have higher bandwidth efficiency.

1. Introduction

In the second generation mobile radio systems the modulation techniques (GMSK and π/4 QPSK) with relatively low bandwidth and high power efficiency are implemented. The allocated frequency band and expected bit rates for future mobile radio stimulate mobile system designers to investigate the modulation techniques with higher bandwidth efficiency.

The bandwidth efficiency can be improved:
- by increasing the modulation dimensionality: Two dimensional schemes are usually implemented in mobile communication systems. If orthogonal data shaping and orthogonal carriers are used, the signal dimension is increased by the factor of two. The Q^{2}PSK modulation is an example of four dimensional modulation scheme, where two orthogonal carriers (sine and cosine) and two orthogonal data shaping pulses (sine and cosine) are used to increase the signal bandwidth efficiency.
- by increasing the modulation level: In the past mostly binary modulation schemes were used (BPSK), where one bit is transmitted per symbol interval. In the case that two or more bits per symbol interval are transmitted, the bandwidth efficiency of the signal is increased. Examples of multi-level modulation schemes are: M-ary PSK and M-ary QAM.
- by superposition: If two modulated signals with constant envelope and with different amplitudes are summed up, the non-constant envelope signal with higher bandwidth efficiency is obtained. An example of so improved bandwidth efficiency is multi-amplitude continuous phase modulation MACPM scheme.

By increased bandwidth the power efficiency of the signal is usually decreased. The power efficiency can be increased again by adding an appropriate encoder to the system.

Bandwidth efficient multi-level, multi-dimensional and superimposed modulation techniques require receivers with a high degree of complexity. The coherent detection with perfect carrier recovery outperforms non-coherent detection of the signal in the Gaussian channel. The complexity of coherent receiver, which requires the implementation of carrier recovery circuit, is higher than the complexity of non-coherent receiver. However, in urban and rural mobile radio environment, where the main source of errors is due to movement of mobiles, the coherent detection is inferior to the non-coherent differential detection, therefore the differential receivers are usually implemented in the second generation mobile communication systems.

In the paper we are analysing a non-coherent differential detection of multi-amplitude continuous phase modulation (MACPM) scheme for different

modulation indices. The MACPM signal is obtained by a superposition of two continuous phase modulated (CPM) signals. By MACPM signal a compromise is reached among power efficiency, bandwidth efficiency and implementation complexity.

At the beginning of the paper the MACPM signal is described. After that, the non-coherent differential receiver for MACPM signal is introduced and analysed. After the description of simulation model the results for non-coherently and differentially detected MACPM and CPM signals are computed, plotted and analysed. At the end a comparison between the complexity of coherent and non-coherent detection is made and discussed.

2. Multi-amplitude CPM signal

Multi-amplitude CPM (MACPM) signal is defined as a superposition of two CPM signals with different amplitudes

$$s(t,\phi_1,\phi_2)= K\cos(2\pi f_c t +\phi_1(t))+$$
$$K\Lambda\cos(2\pi f_c t +\phi_2(t)) \tag{1}$$

where

$$K = \sqrt{\frac{2E}{T}\frac{1}{(2\Lambda+1)}}$$

is the average signal energy, E is the symbol energy, T is the symbol interval, f_c is the carrier frequency, h is the modulation index and $q(t)$ is the phase pulse given by

$$q(t)= \int_{-\infty}^{t_0} g(\tau)d\tau ,$$

where $g(t)$ is the frequency pulse. The CPM signal is obtained when the parameter $\Lambda=0$, while the optimal value of parameter $\Lambda=2$ for MACPM signals.

The information carrying phase is defined as

$$\phi_j = 2\pi h\sum_{i=-\infty}^{n}\alpha_j^{(i)}q(t-iT)+\phi_{j0} \qquad j = 1,2$$

where $\alpha_j^{(i)}$ is the symbol value the of i-th symbol interval chosen from following set of values

$$\alpha^{(i)} \in \{\pm1,\dots\pm(M-1)\}.$$

Various MACPM modulation schemes are obtained by the choice of modulation index h, maximum symbol value M, pulse shaping function and duration. By the increased value of modulation index and maximum symbol value M, the signal bandwidth is increased as well. The power efficiency of the MACPM signal is proportional to the value of modulation index. Pulse

shaping and pulse duration have also influence on power and bandwidth efficiency. For example, if the raised cosine pulse shaping is used instead of the rectangular pulse shaping, the signal bandwidth is decreased and power efficiency remains nearly the same. The comments about MACPM signals are not valid for weak modulation indices.

Our analysis is limited to the rectangular pulse shapes and pulse duration of one symbol interval. We took into account the following modulation indices: $h=0.5$, $h=0.25$ and $h=0.125$. If the modulation index is not the rational number, the signal has infinite number of phase states. Only rational modulation indices are convenient for hardware realisation.
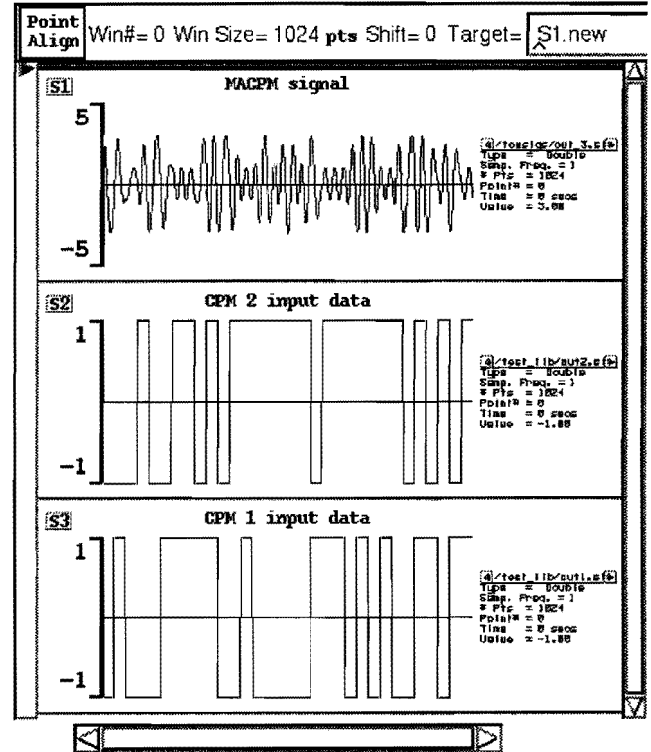
Figure 1: MACPM signal for $h=0.5$ and $f_c=2/T$ and input data for CPM components

The MACPM signal for $h=0.5$ and carrier frequency $f_c=2/T$ is plotted in Figure 1-S1. The corresponding input data $\alpha_1^{(i)}$ for CPM component with low amplitude and input data $\alpha_2^{(i)}$ for CPM component with high amplitude are shown in Figure 1-S3 and Figure 1-S2. T denotes the bit interval. The signal phase is continuous and the signal envelope is variable. If the data inputs are compared with the MACPM signal it could be found out that $\alpha_2^{(i)}$ define the frequency or phase changes of the output signal. If $\alpha_2^{(i)}=+1$ the signal frequency is

$$f_s = f_c +\frac{1}{2}\frac{1}{2T}$$

and if the $\alpha_2^{(i)}=-1$ the signal frequency is

$$f_s = f_c - \frac{1}{2}\frac{1}{2T}.$$

The phase change in one symbol interval is $+\pi/2$ for $\alpha_2^{(i)}=+1$ and $-\pi/2$ for $\alpha_2^{(i)}=-1$. The signal envelope depends on both input data. The relations between symbol envelope and input date are shown in Table 1.

$\alpha_1^{(i)}$	$\alpha_2^{(i)}$	envelope(i)	envelope(i+1)
+1	+1	1	1
+1	-1	1	3
-1	+1	1	3
-1	-1	1	1
+1	+1	3	3
+1	-1	3	1
-1	+1	3	1
-1	-1	3	3

Table 1: Relations between MACPM signal envelope and input data.

It is evident form Table 1 that the signal envelope does not change in the next symbol interval, if both input signals $\alpha_1^{(i)}$ and $\alpha_2^{(i)}$ have the same value. If the signal values $\alpha_1^{(i)}$ and $\alpha_2^{(i)}$ are different, the signal envelope is changed from one symbol interval to another. The above mentioned signal envelope property can be applied in differential amplitude detection of MACPM signals.

The signal envelope and the phase changes can be clearly shown in the constellation diagrams. The constellation diagrams of MACPM signals for $h=0.5$, $h=0.25$ and $h=0.125$ are plotted in Figure 2.
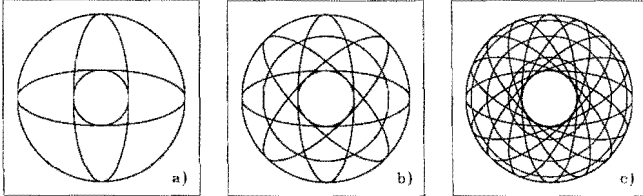
Figure 2: MACPM signal constellation diagrams for a) $h=0.5$, b) $h=0.25$ and c) $h=0.125$

3. MACPM non-coherent differential receiver

The non-coherent MACPM receiver, based on the assumption that MACPM modulation is a combination of frequency and amplitude modulation, is plotted in Figure 3.

After IF filtering the signal is processed in automatic gain control circuit (AGC). After AGC the receiver is divided in two parts. The upper part is simple FM demodulator based on discrimination detection. It is used to detect data carried by the high amplitude CPM component. The lower part is differential envelope

demodulator which detects data carried by the low amplitude CPM component.
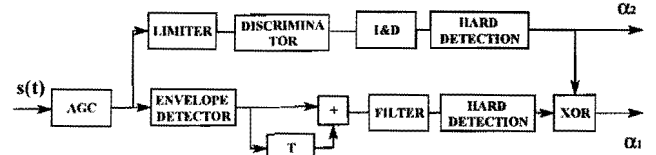
Figure 3: Non-coherent MACPM receiver for MACPM signals

The MACPM signal written in Equation 1 can be modified and written in the following form

$$s(t,\phi_1,\phi_2)= \sqrt{5+4\cos(\Delta\phi)}$$
$$\cos\left(2\pi f_c t +\phi_2 + \arctan\left(\frac{\sin(\Delta\phi)}{2+\cos(\Delta\phi)}\right)\right) \quad (2)$$

where $\Delta\phi$ is the phase difference $\Delta\phi=\phi_1-\phi_2$. In Equation 2 the ratio of CPM component amplitudes $\Lambda=2$. In this case the MACPM signal has the form of FM modulated signal with amplitude fluctuation. If the condition

$$\phi_2 > \arctan\left(\frac{\sin(\Delta\phi)}{2+\cos(\Delta\phi)}\right)$$

is satisfied, the non-coherent detection of MACPM signal can be implemented.

The received and amplified MACPM signal is then processed in the limiter-discriminator block. The limiter-discriminator block performs as a detector. Its output is smoothed by an integrate-sample-and-dump (ISD) filter with the integration time T. The output of the discriminator is the derivative of the received signal phase. The integrate-sample-and-dump filter re-integrates the signal phase producing the phase difference in each sample interval

$$\Delta\psi = \Delta\phi_2 + \Delta\eta$$

where $\Delta\phi_2$ is the phase difference of high amplitude CPM signal component and $\Delta\eta$ is the phase noise caused by fluctuation of signal envelope and channel imperfections.

The AGC output signal is processed by envelope detector in the lower part of the receiver. The envelope change in symbol interval is detected by simple differential decoder. After differential decoding the low pass filter filters out the higher frequency noise. The filtered signal is hard detected. The symbol carried by the small amplitude CPM component is obtained by the XOR operation on FM discriminator and envelope detector outputs of the receiver.

4. Simulation model

The simulation model for MAMSK signal analysis consists of

- two random binary sources
- MACPM modulator
- mobile channel
- MACPM demodulator and
- error counter

Two random bit generators generate two pseudo random bit streams. The bit generation frequency is normalised to 1Hz in each bit stream.

Each CPM component modulator is implemented as a voltage controlled oscillator (VCO). The VCO gain factor depends on modulation index. The output signal of one CPM component modulator is amplified by 3dB and added to the other non-amplified CPM signal component of the other CPM modulator. The simulations are done in baseband so the signal carrier frequency is $f_c = 0$.

The mobile radio channel distorts the MAMSK signal. The mobile channel model takes into consideration following distortions:

- Rayleigh fading
- Doppler frequency shift and
- white Gaussian noise.

The fast Rayleigh fading and the Doppler frequency shift are simulated by multiplying the MACPM signal by a single complex time-varying weight. The weight is generated by passing complex white Gaussian noise through a fading filter, and interpolating the output of the fading filter afterwards. The fading filter is based on the Jake's model, which has the following frequency response

$$H(f) = \begin{cases} \dfrac{A}{\sqrt{1-\left(\dfrac{f}{f_d}\right)^2}} & for \quad |f| \le f_d \\ 0 & for \quad |f| > f_d \end{cases}$$

where f_d is the Doppler frequency shift and A is a constant.

In the model the signal power, the signal-to-noise ratio and the bandwidth of the signal are calculated. The adequate amount of Gaussian noise is added to the signal.

The MACPM non-coherent receiver is used in the model according to the description in the previous section.

BER is obtained by counting the errors. At least 50 errors should be counted out to calculate BER and at least 1000 bits should be processed in simulation run to find out the value for low signal to noise ratio.

5. Simulation results

The MACPM and CPM signals in Gaussian noise environment are simulated at the beginning. The BER simulation results for MACPM signal and for CPM signals are plotted in Figure 4 and Figure 5 respectively. The bit error rate (BER) is in inverse proportion to the value of modulation index. The CPM signals achieve lower BER at the same signal to noise ratios than the MACPM signal, because of the better minimum Euclidean distance.
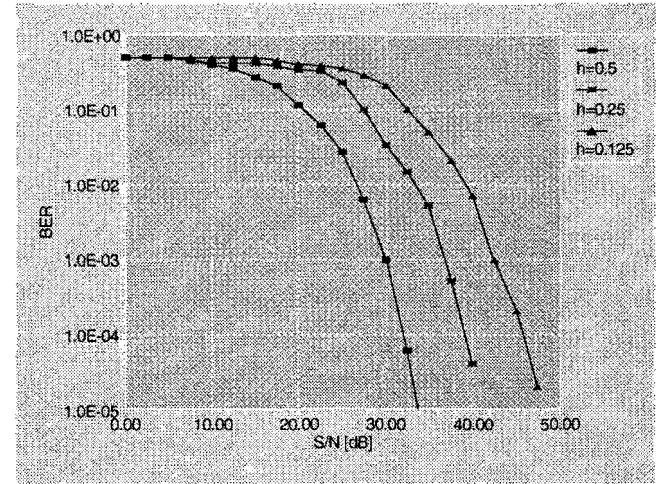
Figure 4: BER for MACPM signals *(h=0.5, 0.25, 0.125)* in Gaussian channel
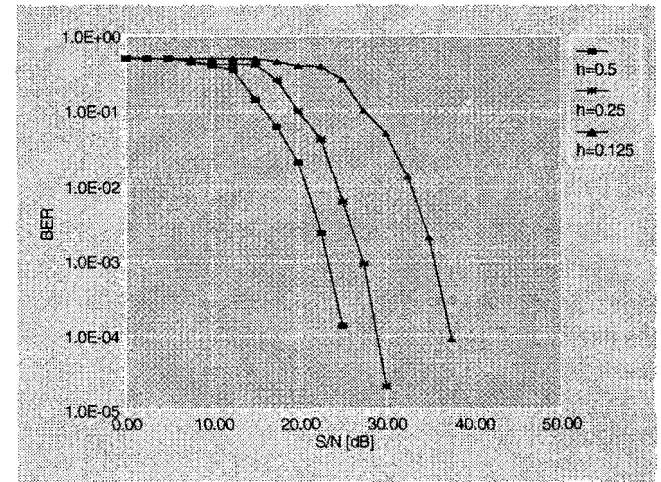
Figure 5: BER for CPM signals *(h=0.5, 0.25, 0.125)* in Gaussian noise channel

In the next simulation run CPM and MACPM signals are analysed in a typical mobile radio communication environment characterised by Rayleigh fading and Doppler frequency shift. Rayleigh fading channel models are used in mobile communication systems to

112

simulate the effects of multiple point scatters in the neighbourhood of the moving receiver. The Doppler frequency shift is caused by random motion of mobile terminal. It depends on the mobile terminal velocity and on the angle between the mobile motion direction and the incoming radio wave direction. The maximum frequency shift is achieved in the case of direct approaching or moving away. BER for CPM and MACPM signals caused by Rayleigh fading and Doppler frequency shift is plotted in Figure 6. The Doppler frequency shift is normalised by the bit interval T. The curves for MACPM signals are close together, because the majority of errors are caused by differential amplitude detection which is independent of the modulation index value. For CPM signals the influence of modulation index on Doppler frequency shift is greater. Small modulation indices are less resistant to Rayleigh fading and Doppler frequency shift.
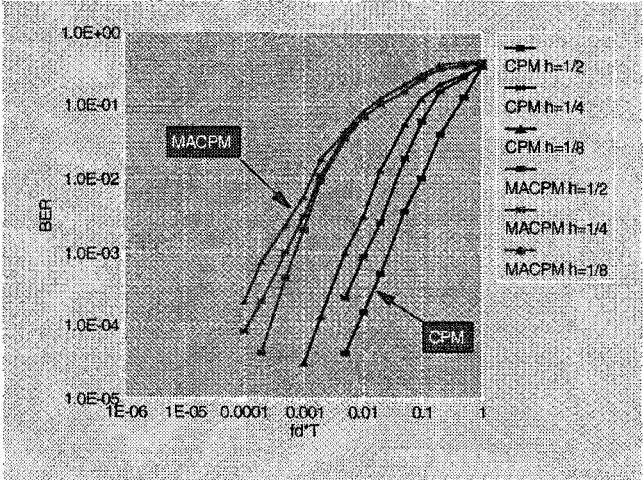
Figure 6: BER for CPM and MACPM signals in Rayleigh fading channel with Doppler frequency shift
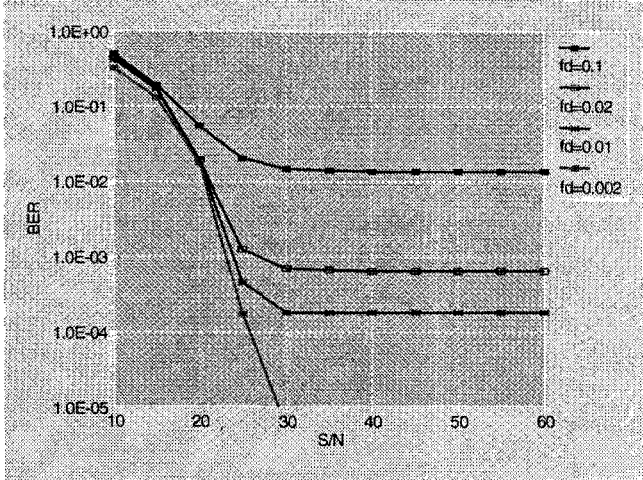
Figure 7: Non-coherently detected CPM signal with h=0.5 in Gaussian noise and Rayleigh fading with Doppler frequency shift channel.
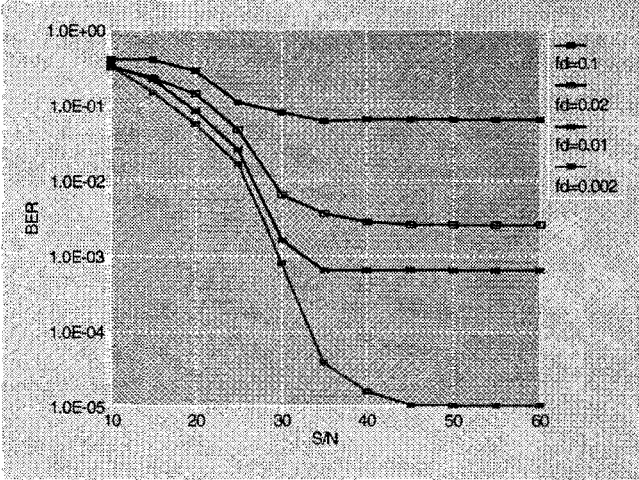
Figure 8: Non-coherently detected CPM signal with h=0.25 in Gaussian noise and Rayleigh fading with Doppler frequency shift channel.
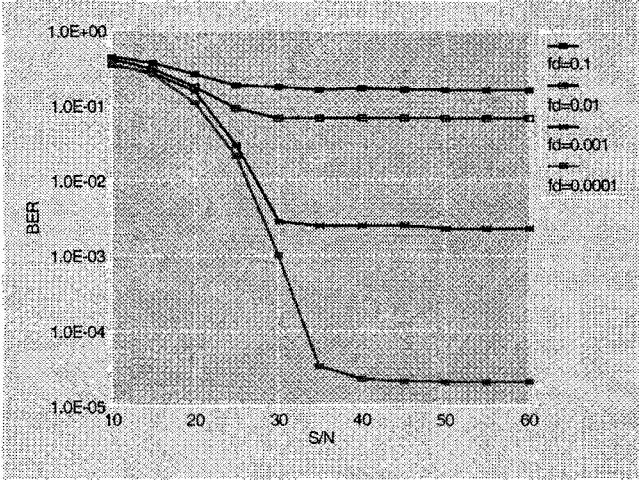
Figure 9: Non-coherently detected MACPM signal with h=0.5 in Gaussian noise and Rayleigh fading with Doppler frequency shift channel.
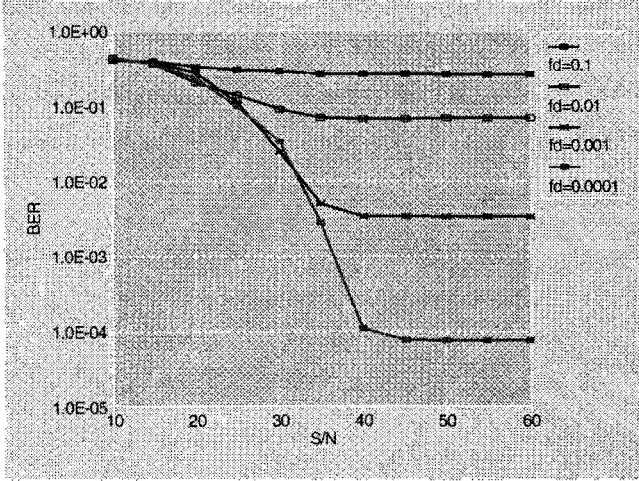
Figure 10: Non-coherently detected MACPM signal with h=0.25 in Gaussian noise and Rayleigh fading with Doppler frequency shift channel.

In Figures 7, 8, 9 and 10 the BER diagrams for CPM and MACPM signals distorted by Rayleigh fading, Doppler frequency shift and additive Gaussian noise

are shown. At lower values of signal to noise ratio (S/N) the dominant distortion is Gaussian noise. At higher S/N values the Rayleigh fading and the Doppler frequency shift are the main source of errors.

5. Conclusion

The non-coherent detection of MACPM signals is analysed in the paper. The proposed receiver structure is easy to implement, because no carrier recovery block is needed.

If the results for coherently detected MACPM signals with ideal carrier recovery are compared with the results for non-coherently differential detected MACPM signals, the degradation of the system performance is noticed only for Gaussian channel model. The non-coherent MACPM signal detection performance in Rayleigh fading channel with Doppler frequency shift is similar to the performance of the coherent signal detection. In the channel taking into account Gaussian noise, Rayleigh fading and Doppler frequency shift the degradation of few dB can be observed for non-coherent detection comparing to coherent detection.

If we compare non-coherent MACPM detection to non-coherent CPM detection we can conclude that CPM signals with non-coherent detection have better performance than MACPM signals in all analysed types of channels, although their bandwidth efficiency is substantially lower. Due to their high bandwidth efficiency the MACPM signals could be the choice for future mobile communication systems.

Low power efficiency of the MACPM signal can be improved by convolutional encoding.

References

[1.] D.Saha, T.G.Bridsall, *Quadrature-Quadrature Phase-Shift Keying*, IEEE Trans. on Comm. Vol. 37, No.5, May 1989, pp. 437-448

[2.] K.Feher, *Advanced Digital Communications*, Prentice-Hall, Inc., 1987.

[3.] Proakis, *Digital Communications*, McGraw-Hill, New York, 1983.

[4.] Jake, *Microwave Mobile Communication*, J.Wiley, 1974

[5.] J. B. Anderson, T. Aulin, C. E. Sundberg, *Digital Phase Modulation*, Plenum press, New York, 1986

A Dual-Signal Receiver for Narrowband DPSK Modulation with Co-channel Interference Cancellation

Gerard J.M. Janssen
Telecommunications and Traffic Control Systems Group
Delft University of Technology
Delft, The Netherlands
Telephone +31-15-786736; Fax +31-15-781774
E-mail: g.janssen@et.tudelft.nl

Abstract: A dual-signal receiver (DSR) is proposed using differentially coherent detectors for simultaneous reception of two Differential Phase Shift Keying (DPSK) modulated co-channel signals. Differentially coherent detection does not require phase reference recovery, which makes this dual-signal receiver suitable for application in the fading mobile channel. In this paper the DPSK DSR is analysed for the non-fading AWGN channel. Computational and simulation results for the bit-error-rate (BER) are presented. It is shown that the DPSK DSR is near-far resistant. For BER = 10^{-4}, the DPSK DSR shows a degradation of 2 dB for the large signal and less than 1 dB for the small signal at SNR = 12 dB when compared to the coherent BPSK DSR.

1. Introduction

The demand for frequency bandwidth for mobile radio and wireless indoor communications is increasing dramatically. The expected dense use of the spectrum requires that the scarcely available bandwidth must be shared by the users as efficiently as possible. Protocols, modulation types and coding schemes have been developed to achieve efficient usage of the available bandwidth in a multi-user environment, without destroying each others transmissions. In cellular networks, co-channel interference is a main capacity limiting factor. Further enhancement of frequency efficiency can be achieved by improving the match of the receiver to the multi-signal environment.

In reference [1-3], a dual-signal receiver (DSR) was proposed for simultaneous reception of two Binary Phase Shift Keying (BPSK) modulated co-channel signals. The receiver consists of two coherent BPSK detectors in succession, where the first detector demodulates the large signal by exploiting the capture effect. The fact that both data signals are uncorrelated was used to separate both signals by remodulation of the input signal with the detected symbols of the large signal and suppress this signal. The small signal is then demodulated by the second detector. The BPSK DSR needs an accurate carrier phase reference to perform coherent detection and signal suppression. While every symbol is detected by comparing it to the reference carrier, errors are made

independent of previous symbols, and signal suppression can be implemented by subtraction of the detected symbol from the input signal. Therefore, in addition to an accurate phase reference, also an accurate amplitude estimate is required.

The requirement of accurate phase and amplitude references is difficult to fulfil in a mobile fading channel where the received signal varies rapidly. For differentially coherent detection, the phase of the previously received symbol is used as a reference, which requires a phase correlation time of only two symbols. Therefore, differentially coherent detection is well suited for use in the mobile fading channel.

In this paper a receiver structure is presented for receiving two DPSK modulated co-channel signals simultaneously, using two differentially coherent detectors in succession. Because carrier recovery is not needed, this receiver requires little synchronisation time and is also cheaper to implement.

The fact that no accurate phase reference is available makes the differentially coherent detector more sensitive to co-channel interference [4]. The interference not only deteriorates the actual symbol to be detected, but also the reference symbol. Furthermore, it is impossible to do straight cancellation by subtraction of the estimated signal from the input signal, like in coherent BPSK.

The contents of this paper is as follows. In section 2 the DPSK DSR is introduced. The differences with the coherent BPSK DSR are discussed. In section 3, BER expressions for the large and small signal are derived. Computational BER results and simulation results for the DPSK DSR are given in section 4, and compared to the results for the coherent BPSK DSR. In section 5, conclusions are given.

2. DPSK dual-signal receiver

The composite received signal $r(t)$, consists of two 2-DPSK modulated signals s_i and s_c of different strength and AWGN $n(t)$:

$$r(t) = A_i b_i \cos(\omega_0 t + \phi_i) + A_c b_c \cos(\omega_0 t + \phi_c) + n(t) \tag{1}$$

Here, A denotes signal amplitude, ω_0 is the carrier frequency and ϕ indicates carrier phase. The signal parameters of the large and small signal are indicated by the subscripts i and c, respectively. $b_{i,c}$ are asynchronous random antipodal differentially encoded symbols with rectangular pulse shape and duration T_s. The differential encoding operation is given by:

$$d_{code,k} = d_{code,k-1} \oplus d_k \qquad (2)$$

where $d_{code,k}$ is the kth differential encoded bit, d_k is the kth input data bit, and $d \in \{0,1\}$. $\oplus$ indicates the exclusive-or operation. The transmitted symbol $b_k = 2d_{code,k} - 1$, $b_k \in \{-1,+1\}$ with equal probability of occurrence.

In figure 1, the block diagram of the DPSK DSR is shown.
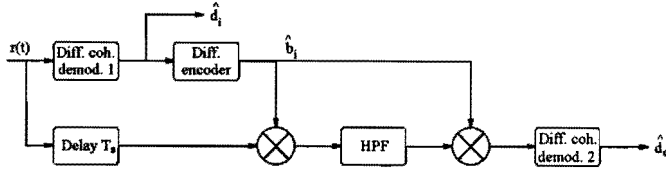
Figure 1: Dual-signal receiver structure for 2-DPSK modulated co-channel signals.

The DSR consists of two demodulators in succession, and is equivalent with the structure of the DSR for coherent BPSK signals [1]. Both demodulators are identical with complex equivalent baseband block diagram as shown in figure 2.
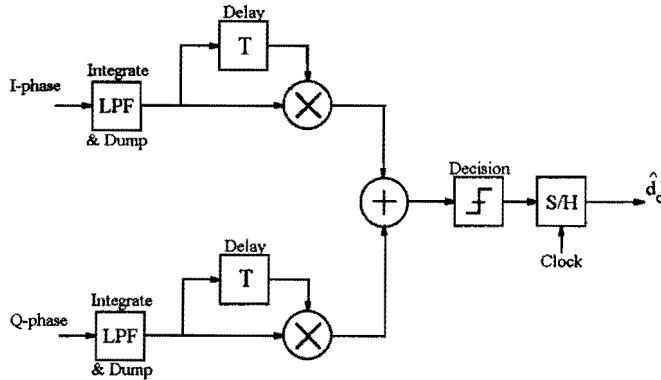
Figure 2: Complex baseband equivalent block diagram of the optimum 2-DPSK demodulator.
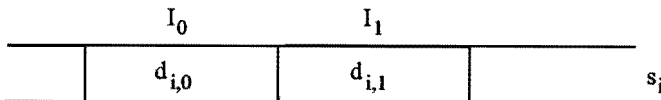
Figure 3: Signal constellation with overlapping large and small signals.

In figure 3, a simple signal constellation is shown, where $I_0 = Ref_I + j\,Ref_Q$ is the reference symbol and $I_1 = Sig_I + j\,Sig_Q$ the symbol to be detected. The decision making by the detector is performed as:

$$\hat{d}_k = 1 \quad if \quad Sig_I Ref_I + Sig_Q Ref_Q > 0$$
$$\hat{d}_k = 0 \quad if \quad Sig_I Ref_I + Sig_Q Ref_Q < 0 \qquad (3)$$

The operation of the DSR is as follows. The first detector is captured by the large signal s_i and determines an estimate $\hat{d}_i$ of the transmitted data signal d_i. This estimate $\hat{d}_i$ is used to cancel s_i in a delayed version of the input signal $r(t)$. The delay T_s is required to compensate for the integration time in the detector.

The first step in the cancellation process is to make both signals s_i and s_c distinguishable. This is done by remodulation of the delayed input signal with $\hat{b}_i$, which is the differentially encoded estimate $\hat{d}_i$. $\hat{b}_i$ is correlated with b_i but fully uncorrelated with b_c because both data signals d_i and d_c are independent. The effect of the remodulation is that the signal power spectrum is concentrated in a bandwidth much less than the original bandwidth of $1/T_s$, whereas the spectral bandwidth BW of $\hat{b}_i s_c$ increases to $1/T_s < BW < 2/T_s$ depending on the symbol timing difference between $\hat{b}_i$ and b_c. So after remodulation with $\hat{b}_i$ both signal spectra have become distinct because of the large difference in bandwidth.

Unlike for coherent BPSK, the detected estimate $\hat{d}_i$ is uncorrelated with b_i, but requires differential encoding to become correlated. Remodulation with $\hat{b}_i$ results in the term $\hat{b}_i b_i$. For coherent BPSK, the corresponding term is $\hat{d}_i d_i$. At first glance both results look identical, however, they are not because the differentially encoded variable $\hat{b}_i$ contains phase memory. The changes of sign of the terms $\hat{b}_i b_i$ and $\hat{d}_i d_i$ for both modulation techniques can be modelled by a Markov state diagram as shown in figure 4.
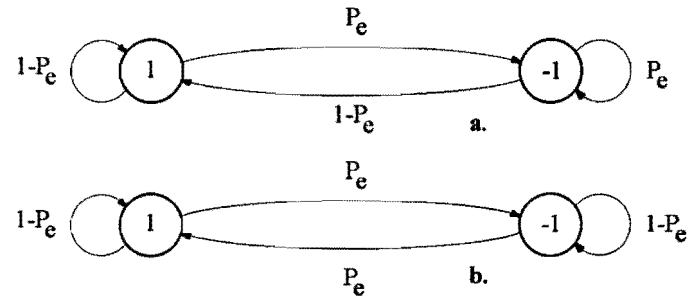
Figure 4.: Phase state diagram for the reverse modulated signal with a). Coherent BPSK modulation and b.) Differentially coherent DPSK modulation.

This clearly illustrates the difference. For coherent BPSK the remodulated signal $\hat{d}_i d_i = 1$ in case no error is made

116

in the detection of $\hat{d}_i$ and $\hat{d}_i d_i = -1$ in case of an error. So the average value of $E[\hat{d}_i d_i] > 0$. This DC component results in a delta function at $f = 0$ in the frequency spectrum, [1]. However, for 2-DPSK, the remodulated signal $|\hat{b}_i b_i| = 1$, but changes sign every time an error occurs in $\hat{d}_i$, and therefore the average value of $E[\hat{b}_i b_i] = 0$. The average sign change rate is $R_{sign} = P_{ei} / T_s$ and depends on P_{ei}, and therefore, also the bandwidth of the remodulated signal depends on P_{ei}. The remodulated signal $\hat{b}_i b_i$ can be seen as a digital signal with amplitude values $\pm A_i$ and a random time between sign changes. The probability of a sign duration of kT_s is $P_{ei}(1 - P_{ei})^{k-1}$. For low P_{ei}, the distribution of the sign changes can be approximated by the continuous Poisson distribution with average change rate $\mu = P_{ei} / T_s$. The signal power spectrum of the signal $\hat{b}_i b_i$ is then found by taking the Fourier transform of the autocorrelation function $R_{\hat{b}b}(\tau) = A^2 \exp(-2\mu|\tau|)$:

$$G_{\hat{b}b}(f) = \frac{A^2}{\mu\left[1 + (\pi f / \mu)^2\right]} = \frac{A^2 T_s}{P_{ei}\left[1 + (\pi f T_s / P_{ei})^2\right]} \quad (4)$$

The spectrum of $\hat{b}_i b_i$ is shown in figure 5. for several values of P_{ei}. It is clearly seen that the amount of decrease of the bandwidth is determined by P_{ei}.
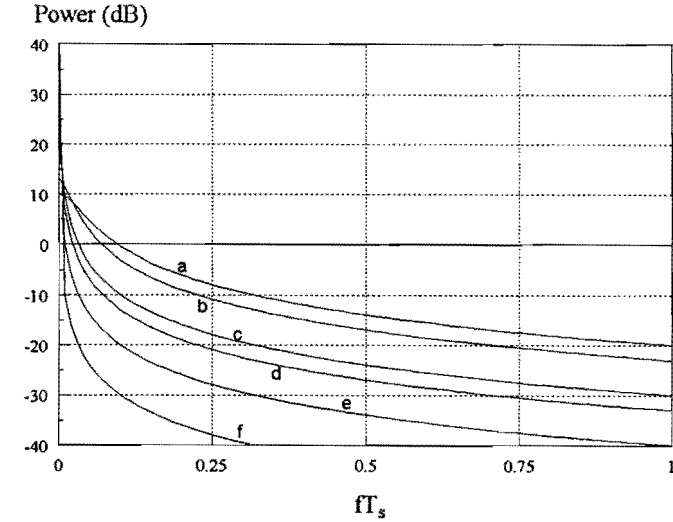
Figure 5: Frequency spectrum after remodulation for different BER: a.) $P_{ei} = 10^{-1}$; b.) $P_{ei} = 5 \cdot 10^{-2}$; c.) $P_{ei} = 10^{-2}$; d.) $P_{ei} = 5 \cdot 10^{-3}$; e.) $P_{ei} = 10^{-3}$; f.) $P_{ei} = 10^{-4}$.

The power of the large signal is removed by filtering the remodulated signal with a high pass filter (HPF).
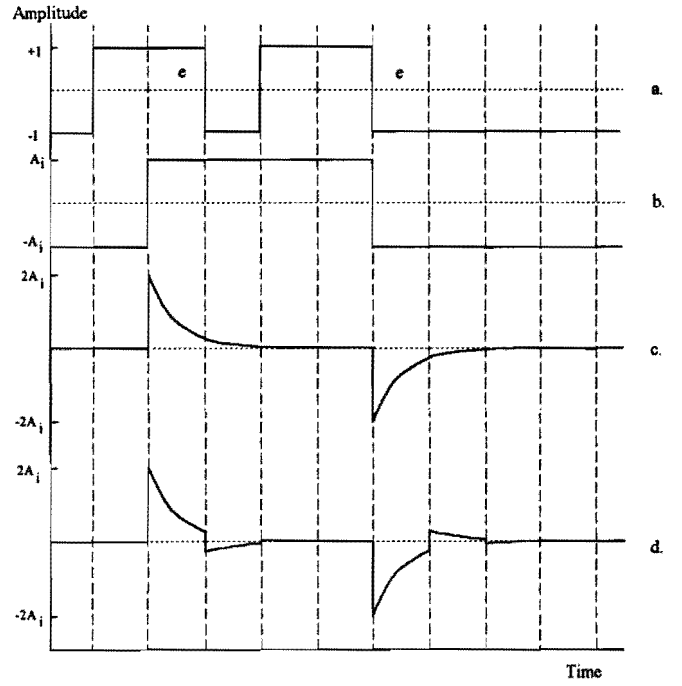
Figure 6: Time domain sample of the baseband signal of s_i without timing error: a.) estimated data $\hat{d}_i$, ε indicates an error; b.) signal after the 1st remodulation with $\hat{b}_i$; c.) remodulated signal after filtering; d.) signal Δs_i after the 2nd remodulation with $\hat{b}_i$.

The filter bandwidth is a compromise between: i.) filtering out as much of the large signal power as possible. This requires a HPF with a high cut-off frequency. ii.) leaving the weaker reverse modulated signal $\hat{b}_i s_c$ undistorted. This requires a very low cut-off frequency. Here, a first-order high-pass RC filter (HPF) with time constant τ is used. Now, the remaining interference of the large signal is an exponentially decaying voltage which starts with amplitude $2A_i$. This is shown in figure 6. After the second reverse modulation with $\hat{b}_i$ which is required to restore s_c, the remaining interference of the large signal is an exponentially decaying burst of symbols which starts each time an error has occurred in the first demodulator. It can be shown that the amplitude of the interfering bits has dropped to A_{sup} after N symbols with N given by

$$N = \left\lceil 1 - \frac{\tau}{2T_s} \ln\left(\frac{A_{\text{sup}} T_s}{\tau(1 - \exp(-2T_s / \tau))} \right) \right\rceil \quad (5)$$

For $\tau = 1.59 T_s$, the 3 dB bandwidth is $0.1/T_s$ and the value for $N = 4$ with $A_{\text{sup}} = -30$ dB. A small τ will result in less interference caused by the remaining of the large signal due to errors made in the first demodulator, however, it will also cause more distortion to the small signal. For large τ, the opposite is the case. There exists an optimum τ as a function of P_{ei}. Optimisation of the suppression filter is subject of further research.

3. BER calculations

BER for the large signal. For the calculation of the BER of the large signal, we start with an arbitrary sample of the signals as shown in figure 7. Both signals are assumed asynchronous, and receiver timing with respect to the large signal s_i is assumed perfect.
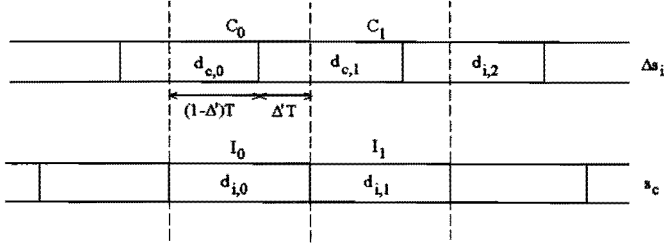
Figure 7: Symbol constellation for the large signal.

For every symbol I_j of s_i to be detected, three symbols of s_c are involved. Every different combination of these symbols results in a different BER for the symbol $\hat{d}_i$. The symbols of s_c that influence the detection of symbol I_1 of s_i can be written as $s_c = A_c d_{c,j}$, with $j \in \{0, 1, 2\}$. The effective signal C_j in one bit interval of s_i is $C_j' = A_c\{(1-\Delta')d_{c,j} + \Delta' d_{c,j+1}\} = A_c \kappa_{c,j}$. Let the in-phase and quadrature-phase signal- and reference samples be defined as

$$Sig_{I,Q} = I_{1_{I,Q}} + C_{1_{I,Q}}' + n_{1_{I,Q}}$$
$$Ref_{I,Q} = I_{0_{I,Q}} + C_{0_{I,Q}}' + n_{0_{I,Q}} \qquad (6)$$

where $I_{0,1}$ and $C_{0,1}$ are indicated in figure 7, and $n_{0_{I,Q}}$, $n_{1_{I,Q}}$ are statistically independent AWGN noise samples. Now the conditional error probability is defined as

$$P_{ei}(\varepsilon|\Delta',\phi',d_{i,1}=1) = Pr(Sig_I Ref_I + Sig_Q Ref_Q < 0) \quad (7)$$

with ϕ' is the phase difference between s_i and s_c, and $\Delta' T_s$ is the timing difference between b_i and b_c. For high signal-to-noise ratio, where the influence of the quadratic noise terms $n_{0_I}n_{1_I}$ and $n_{0_Q}n_{1_Q}$ can be neglected [5, 6], it can be shown that the conditional BER is of the form:

$$P_{ei}(\varepsilon|\Delta',\phi') = 0.5\exp(-\mu\gamma_i) \qquad (8)$$

with, $\gamma_i = E_{bi}/N_0$, and μ is determined by the symbol constellation of s_i and s_c:

$$\mu(m,\kappa_{c,0},\kappa_{c,1},\psi_i,\phi') = \frac{1 + \dfrac{m\kappa_{c,0}\kappa_{c,1}}{\psi_i} + \dfrac{m\kappa_{c,0}+\kappa_{c,1}}{\sqrt{\psi_i}}\cos\phi'}{1 + \dfrac{\kappa_{c,0}^2+\kappa_{c,1}^2}{2\psi_i} + \dfrac{m\kappa_{c,0}+\kappa_{c,1}}{\sqrt{\psi_i}}\sin\phi'}$$

$$(9)$$

Here, $\psi_i = A_i^2/A_c^2$ and $m = d_{i,0}d_{i,1} \in \{-1,+1\}$. Because three interfering symbols are involved, the average conditional BER for $\hat{d}_i$ is given by

$$P(\varepsilon|\phi',\Delta') = \frac{1}{2^4}\sum_{\text{All } m}\sum_{2^3 \text{ strings}} 0.5\cdot\exp(-\mu\gamma_i) \qquad (10)$$

The average P_{ei} is calculated by integration of (10) over Δ' and ϕ'.

BER for the small signal. The small signal is detected after suppression of the large signal. For the model used to calculate the BER P_{ec} of the small signal we make the following assumptions: *i.)* The exponentially decaying interference of Δs_i becomes zero after N correctly detected symbols following an error. Here, $A_{\sup} = -30\,dB$ is taken. *ii.)* When an error occurs, the interference amplitude is reset to the initial condition of $2A_i$. So the interference is always due to one error, and not the sum of the influence of more that one error within N symbols.
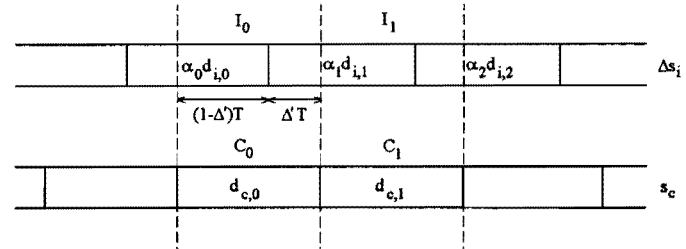
Figure 8: Symbol constellation for the small signal.

Now, the symbols of Δs_i that influence the detection of symbol C_1 of s_c, have an exponentially decaying amplitude. The amplitude of the jth symbol of Δs_i can be written as $A_i\alpha_j b_{i,j}$, where $j \in \{0, 1, 2\}$, and

$$\alpha_j = \frac{1}{T_s}\int_{(j-1)T_s}^{jT_s} 2\exp\left(\frac{-t}{\tau}\right)dt$$
$$= \frac{2\tau}{T_s}\exp\left(\frac{-(j-1)T_s}{\tau}\right)\left(1-\exp\left(\frac{-T_s}{\tau}\right)\right) \qquad (11)$$

The effective amplitude I_j' in one bit interval of s_c is

$$I_j' = A_i\{\alpha_j(1-\Delta')b_{i,j} + \alpha_{j+1}\Delta' b_{i,j+1}\} = A_i\kappa_{i,j}, \qquad \text{with}$$

$$\kappa_{i,j} = \frac{2\tau}{T_s}\exp\left(\frac{-jT_s}{\tau}\right)\left\{b_j\left(\exp\left(\frac{(1-\Delta)T_s}{\tau}\right)-1\right)\right.$$
$$\left. +b_{j+1}\left(1-\exp\left(\frac{-\Delta T_s}{\tau}\right)\right)\right\} \qquad (12)$$

The BER for three interfering symbols, as shown in figure 8 is calculated with (8) and (9) by replacing:

118

$\gamma_i \rightarrow \gamma_c = E_{bc}/N_0$, $\psi_i \rightarrow \psi_c = 1/\psi_i$, $\kappa_{c,j} \rightarrow \kappa_{i,j}$, and $m = d_{c,0}d_{c,1} \in \{-1,+1\}$.

For the average BER, the probability of all possible amplitude levels in the interfering string caused by the remaining of the large signal Δs_i, have to be taken into account. These probabilities Pr(*string*) are given in Appendix A.

Now the average conditional BER for the small signal is given by

$$P(\varepsilon|\phi', \Delta') = \frac{1}{2^4} \sum_{\text{All strings}} \Pr(string) \sum_{\text{All } m} \sum_{\substack{\text{All } d_{i0}, \\ d_{i,1}, d_{i,2}}} 0.5 \cdot \exp(-\mu\gamma_c)$$

(13)

Again, P_{ec} is calculated by integration of (13) over Δ' and ϕ'.

4. Computational results

In this section different performance results for the DPSK DSR are presented. For the computational results the HPF time constant $\tau = 15.9 T_s$ has been taken, which corresponds to a -3 dB bandwidth of $0.01/T_s$. For the coherent BPSK DSR a suppression error of -30 dB is assumed due to inaccurate amplitude estimation.

Bit Error Rate
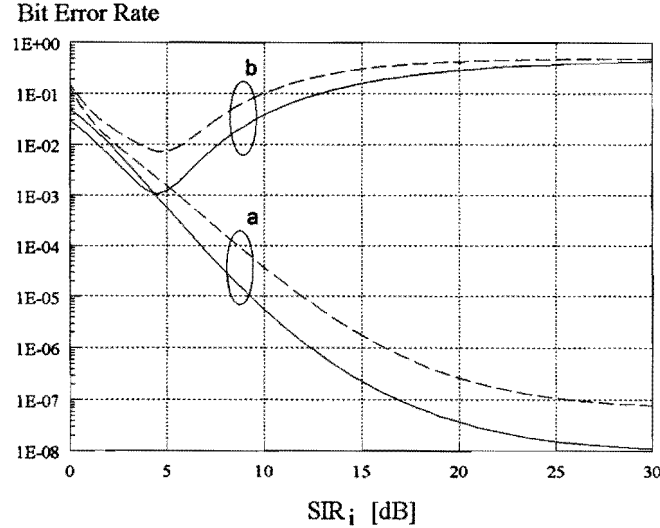
Figure 9: BER P_{ei} (a) and P_{ec} (b) as a function of SIR_i, for the DPSK DSR (- - - -) and the coherent BPSK DSR (———), for SNR_i = 12 dB.

In figure 9, the BER performance of the DPSK and the coherent BPSK DSRs are compared. Computational results are given for P_{ei} and P_{ec} as a function of SIR_i with constant $SNR_i = 12\,dB$ for the large signal. In figure 10, the same results are given for $SNR_i = 18\,dB$.

Bit Error Rate
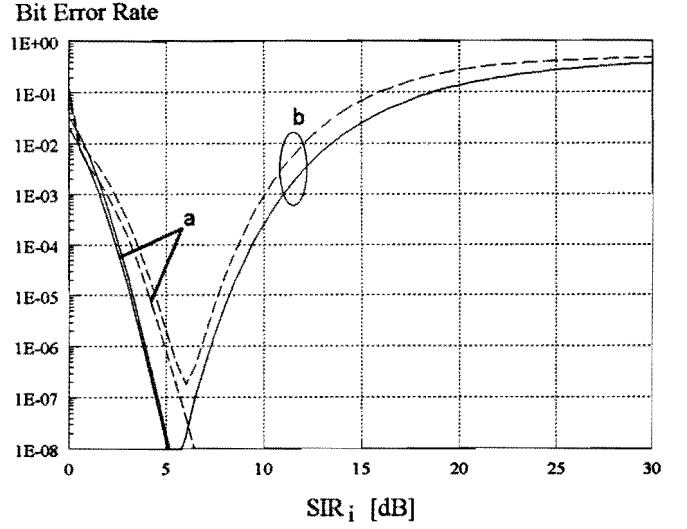
Figure 10: BER P_{ei} (a) and P_{ec} (b) as a function of SIR_i, for the DPSK DSR (- - - -) and the coherent BPSK DSR (———), for SNR_i = 18 dB.

For small SIR_i, the errors made by the second demodulator are highly correlated with the errors made in the first demodulator, which detects the large signal. This is because of the high interference levels caused by these errors. The small signal BER P_{ei} reaches a minimum where the BERs due to large signal interference and noise are balanced. For larger SIR_i the decreasing SNR_c is the main cause of the increasing P_{ec}. The performance of the DPSK DSR for the large signal at $P_{ei} = 10^{-4}$, shows a degradation of 2 dB (SNR_i = 12 dB) and 1 dB (SNR_i = 18 dB) compared to the BPSK DSR. The detection performance of the small signal shows a degradation of 1 - 2 dB compared to the BPSK DSR.

In the figures 11 and 12, simulation results are compared to the computational results for the worst case of synchronous symbol transmission ($\Delta' = 0$) for the large and the small signal, respectively. It is seen that the simulation results for the large signal are in good correspondence with the computational results. For small SIR_i < 2 dB, the theoretical model is not accurate yet. In the simulation results for the small signal, as shown in figure 12, also the effect of the HPF bandwidth BW_{HPF} has been taken into account. The computational results are for BW_{HPF} = $0.01/T_s$. Simulation results are given for BW_{HPF} = $0.02/T_s$, $0.01/T_s$ and $0.005/T_s$. From these results it is clear that a single bandwidth for the HPF will not give optimum performance. For high P_{ei}, which results in a wider large signal spectrum after reverse modulation as shown in figure 5, a larger filter bandwidth is required to achieve optimum performance than for low P_{ei}, where the HPF distortion of the small signal is the dominant deteriorating factor.

119
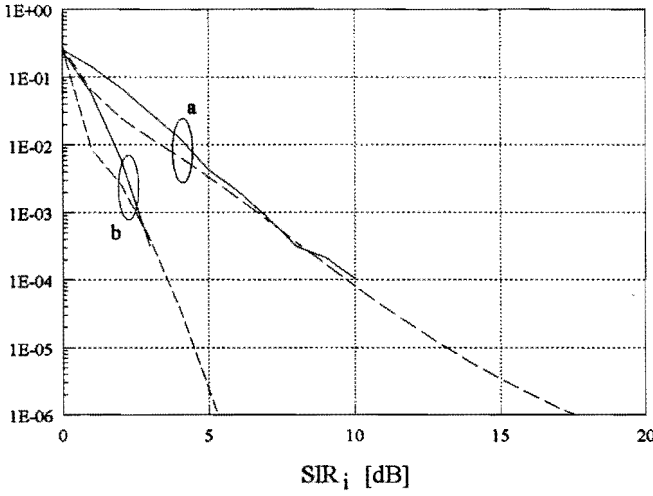
Bit Error Rate

Figure 11: Computational (- - - -) and simulation (——)
BER results for the large signal as a function of SIR_i for
the DPSK DSR for: a.) SNR_i = 12 dB and b.) SNR_i = 18 dB.

Bit Error Rate
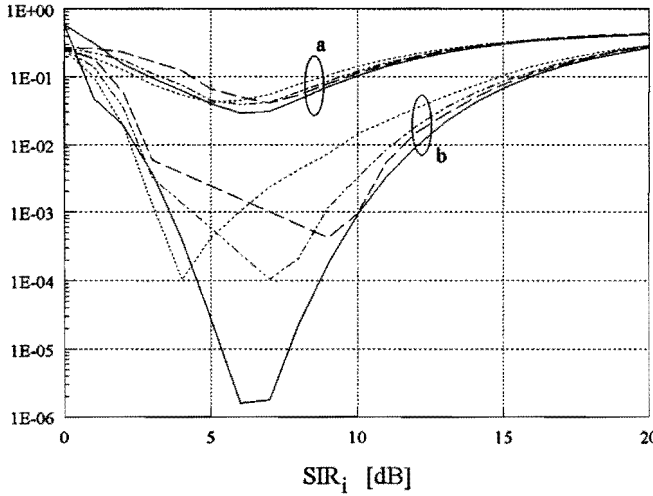
Figure 12: Computational (——) and simulation BER results
for the small signal of the DPSK DSR as a function of SIR_i
for: a.) SNR_i = 12 dB and b.) SNR_i = 18 dB;
BW_{HPF} = 0.02/T_s (------), BW_{HPF} = 0.01/T_s (_.._.._..) and
BW_{HPF} = 0.005/T_s (_ _ _ _).

In figure 13, the performance of a desired signal is shown
as a function of SIR. The SNR of the signal is constant
SNR = 12 dB and 18 dB. For positive SIR the desired
signal is the large one which is demodulated by the first
demodulator. For negative SIR, the desired signal is
demodulated by the second demodulator. The results are
compared to the performance of the BPSK DSR. Again
is clear that the difference in performance between the
DPSK and BPSK DSR is not very large. The
deterioration of the BPSK DSR for the small signal for
very small SIR is caused by the assumption of the non-
perfect amplitude estimate. Note that the desired signal
can be well demodulated for SIR << 0 dB. So, the DPSK
DSR does not suffer from the near-far effect. The

challenging area for further research is of course the area
of SIR close to 0 dB.

Bit Error Rate
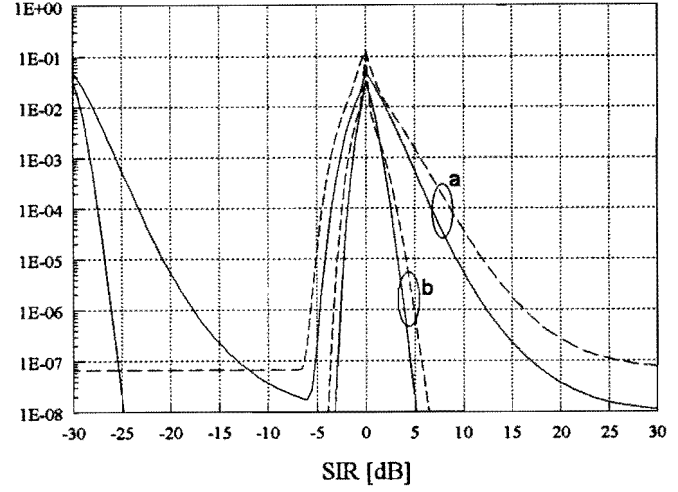
Figure 13: BER of the desired signal as a function of SIR, for
the DPSK DSR (- - - -) and the coherent BPSK DSR
(——) for: a.) SNR = 12 dB and b.) SNR = 18 dB.

6. Conclusions

In this paper, the BER performance of a dual-signal
receiver (DSR) concept using differentially coherent
demodulators for simultaneous reception of two DPSK
modulated co-channel signals, has been analytically
evaluated for both, the large and small signal. Because no
phase reference needs to be recovered, which requires a
stable received signal phase, the DPSK DSR is well
suited for use in the mobile fading channel.
The demodulation performance for the large signal is the
same as is achieved with a conventional differentially
coherent demodulator. After cancellation of the large
signal, a desired signal can be well demodulated for SIR
<< 0 dB. So, the DPSK DSR does not suffer from the
near-far effect. The performance of the DPSK DSR
shows a degradation of only a few dB compared to the
coherent BPSK DSR.

Acknowledgement

Appendix A: String probabilities

Let an erroneously detected symbol of the large signal be
indicated by E. In case that N or more successive
symbols of the large signal were detected correctly, than
the interfering amplitude has decayed to zero, and is
indicated by η. Symbols transmitted less than NT_b after
an error show decaying amplitude and are indicated by J

120

with $J \in \{2, .. ,N\}$. Now the probabilities of occurrence of the different string types are given by

$$\Pr(\eta,\eta,\eta) = (1 - P_{ei})^{N+2}$$

$$\Pr(\eta,\eta,E) = \Pr(\eta,E,2) = \Pr(N,\eta,\eta) = P_{ei}(1 - P_{ei})^{N+1}$$

$$\Pr(\eta,E,E) = P_{ei}^2(1 - P_{ei})^{N}$$

$$\Pr(E,E,E) = P_{ei}^3$$

$$\Pr(N - 1,N,\eta) = P_{ei}(1 - P_{ei})^{N}$$

$$\Pr(J,J+1,J+2) = P_{ei}(1 - P_{ei})^{J+1} \quad \forall J \in \{2,..,N-2\}$$

$$\Pr(J,J+1,E) = P_{ei}^2(1 - P_{ei})^{J} \quad \forall J \in \{2,..,N-1\}$$

$$\Pr(J,E,2) = P_{ei}^2(1 - P_{ei})^{J} \quad \forall J \in \{2,..,N\}$$

$$\Pr(J,E,E) = P_{ei}^3(1 - P_{ei})^{J-1} \quad \forall J \in \{2,..,N\}$$

$$\Pr(E,2,3) = P_{ei}(1 - P_{ei})^{2}$$

$$(A.1)$$

References

[1]   G.J.M. Janssen, "Dual-signal Receiver Structures for Simultaneous Reception of Two BPSK Modulated Co-channel Signals Using Signal Cancellation", *Wireless Personal Communications*, Vol. 1, No. 1, pp. 43-59, Kluwer, December 1994.

[2]   G.J.M. Janssen, "Receiver Structure for Simultaneous Reception of Two BPSK Modulated Cochannel Signals", *Electronics Letters*, Vol. 29, No. 12, pp. 1095 - 1097, June 1993.

[3]   G.J.M. Janssen, "BER and Outage Performance of a Dual-Signal Receiver for Narrowband BPSK Modulated Co-channel Signals in a Rician Fading Channel", IEEE Proc. PIMRC'94, pp. 601 - 606, The Hague, The Netherlands, September 1994.

[4]   A.S. Rosenbaum, "Binary PSK Error Probabilities with Multiple Cochannel Interferences", IEEE Trans. on Commun., VOL. COM-18, No. 3, pp. 241 - 253, June 1970.

[5]   J.H. Park, Jr., Correspondence: "On Binary DPSK Detection", IEEE Trans. on Comm. Techn., Vol. Com-26, pp. 484 - 486, April 1978

[6]   J.G. Proakis, "Digital Communications", 2nd edition, McGraw-Hill Book Inc., 1989.

Performance Assessment of Guided Scrambling

Levente Pátrovics and Kees A. Schouhamer Immink
Philips Research Laboratories, 5656 AA Eindhoven, The Netherlands
Fax: +31 40 274 46 48
E-mail: patro@prl.philips.nl, immink@prl.philips.nl

Abstract— Guided scrambling, a line coding technique introduced recently, can be used to generate DC-free codes. This algorithm is compared to traditional high-rate DC-free codes based on a simple efficiency criterion using the product of the variance of the running digital sum and the redundancy of the code. During the analysis a random model of guided scrambling is introduced. Encoder complexities are considered.

I. INTRODUCTION

In optical and magnetic recording the channel code has to conform to a variety of constraints. One of these constraints is the DC-free (also called balanced) property, which means that the spectral density of the coded sequence should be zero at spectral frequency zero. The purpose of this is to prevent the accumulation of charge in the receiver circuits and to allow transmission of a pilot tone or some other low-frequency signal without interference.

In considering DC-free sequences the term *disparity* plays an important role. The disparity of a binary sequence gives the number of ones minus the number of zeros in the sequence. The disparity of the coded sequence from a given instant ($t = 0$) to the current position is called the *running digital sum* (RDS) of the coded sequence.

From a channel coding point of view the DC-free property can be assured if the RDS of the coded sequence remains bounded, which means that in the long term the numbers of zeros and ones transmitted are approximately equal.

Figure 1 demonstrates the RDS: the transmission of a one or a zero respectively increases or decreases the RDS by one.

The DC-free property is not sufficient to ensure attractive features: the spectrum of the code should also remain low in the low-frequency interval. To measure low-frequency suppression we introduce the *cut-off frequency*, ω_0, which is defined as follows:

L. Pátrovics is a Ph.D. student at the Technical University of Budapest, Hungary. He is working as a trainee at Philips Research Labs until December 1995. After this period he can be contacted at the following E-mail address: patro@fix.fsz.bme.hu.

Fig. 1. Illustration to the RDS.

$$H(\omega_0) = \frac{1}{2},$$

where $H(\omega)$ is the spectral density function of the coded sequence. For a general code the value of ω_0 is hard to calculate, but Justesen [4] has found a useful relation between the sum variance and the cut-off frequency of a DC-balanced code. That relation is

$$\omega_0 \approx \frac{1}{2s^2},$$

where the sum variance s^2 is defined by the following expectation:

$$s^2 = E\left\{Z_i^2\right\} \tag{1}$$

and Z_i is the running digital sum at instant i.

If we consider codes taking at most N different RDS values, we can see that the channel capacity, $C(N)$, of these codes is quite close to unity even for small values of N (e.g. for $N = 11$, $C(N) \approx 0.95$).

The traditional measure of the efficiency of a code is the ratio of the code rate to the channel capacity of the code. A new measure that takes also the sum variance of the code into account was introduced by Immink [2]. The efficiency of a balanced code taking at most N different RDS values is

$$E = \frac{(1 - C(N))\sigma^2(N)}{(1 - R)s^2},$$

where $C(N)$ and $\sigma^2(N)$ are the channel capacity and the sum variance of the maxentropic code [2, p. 178], and R is the rate of the implemented code. According to the new efficiency measure a code is more efficient

if its redundancy is lower (or, equivalently, its code rate is higher) or its sum variance is lower. Several high rate codes and their efficiencies are described in the next section.

The product $(1 - C(N))\sigma^2(N)$ converges to the value 0.2326 very quickly, so that the new efficiency can be estimated from

$$E = \frac{0.2326}{(1-R)s^2}. \qquad (2)$$

II. High-rate DC-free codes

As mentioned above, the channel capacity of DC-free codes is close to unity for large values of N, so codes with a high rate can be constructed. As we are primarily interested in high-rate codes (codes having rate greater than 8/9), we do not consider traditional, low-rate channel codes but concentrate on codes whose rate approaches unity asymptotically.

In the sections which follow we describe several high-rate coding algorithms and examine their efficiency on the basis of Immink [2]. The efficiency of the algorithms will serve for the purpose of comparison with the guided scrambling algorithm.

High-rate codes using enumeration are not covered in what follow as we have found no results relating their efficiency in the literature.

A. Polarity-bit code

This algorithm has very low redundancy: only one redundant bit is sufficient. For each source word $\mathbf{x} = (x_1, \ldots, x_m)$ we have two codewords: $\mathbf{c} = (1, x_1, \ldots, x_m)$ and $\overline{\mathbf{c}} = (0, \overline{x}_1, \ldots, \overline{x}_m)$, where $\overline{x}_i$ is the complement of the bit x_i.

Denote the RDS after transmission of the previous codeword by z and the disparity of $\mathbf{c}$ by d_c. The principle of operation is as follows:

1. If $zd_c = 0$, then we choose between $\mathbf{c}$ and $\overline{\mathbf{c}}$ randomly.
2. If $zd_c < 0$, then we transmit $\mathbf{c}$.
3. If $zd_c > 0$, then we transmit $\overline{\mathbf{c}}$.

Figure 2 illustrates the method: in the case shown, $zd_c > 0$, so we select $\overline{\mathbf{c}}$ for transmission.
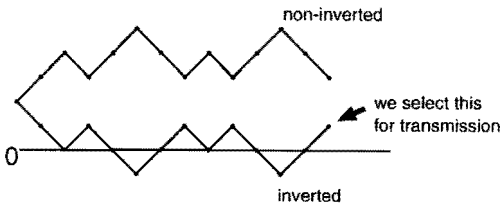
Fig. 2. Illustration of the polarity-bit algorithm.

The rate of the polarity-bit code is

$$R = \frac{n-1}{n} = 1 - \frac{1}{n}.$$

Immink calculated the sum variance of the algorithm as [2, p. 203]

$$s^2 = \frac{2n-1}{3}$$

so that

$$(1-R)s^2 = \frac{2n-1}{3n}.$$

From (2) the asymptotic efficiency is

$$E = 0.2326 : \frac{2}{3} \approx 0.3489. \qquad (3)$$

B. Zero-disparity codes

The RDS bounds are automatically satisfied if only zero-disparity codewords are used, i.e. codewords having equal numbers of zeros and ones. A unique zero-disparity codeword is assigned to each source word. If we use codewords of length n, the number of such codewords is:

$$M = \binom{n}{n/2},$$

so the code rate is

$$R = \frac{\log M}{n} = \frac{1}{n} \log \binom{n}{n/2}.$$

Using Stirling's formula, $n! \approx (n/e)^n \sqrt{2\pi n}$, we get:

$$R \approx 1 - \frac{\log n\pi - 1}{2n}.$$

As the sum variance of the code is $(n+1)/6$, the redundancy-sum variance product

$$(1-R)s^2 \approx \frac{(\log n\pi - 1)(n+1)}{12n} = \left(1 + \frac{1}{n}\right)\frac{\log n\pi - 1}{12}$$

tends to infinity as $n \to \infty$, so the efficiency of the code is asymptotically 0. For smaller codeword lengths ($n < 160$), however, zero-disparity codes are more efficient than the polarity bit algorithm.

C. Low-disparity codes

Low-disparity codes use codewords of disparity $0, \pm 1, \ldots, \pm K$. The zero-disparity codewords are assigned uniquely to source words, and the complement codeword pairs of non-zero disparity also belong to a source word. In the latter case the RDS determines which of the two codewords is transmitted, as in the case of the polarity-bit algorithm.

Calculations of the efficiency have shown [3] that the asymptotic efficiency of low-disparity codes is maximal for $K \approx 0.47\sqrt{n}$, and here $E \approx 0.56$.

III. GUIDED SCRAMBLING

The polarity-bit principle can be generalized as follows: for every codeword $\mathbf{x}$ we have a set of codewords $\mathcal{C}_x = \{\mathbf{c}_1, \ldots, \mathbf{c}_L\}$, and we choose the one with the best properties for transmission. In the case of balanced codes the term "best" denotes the codeword that minimizes the absolute value of the RDS at the codeword end.

To implement the above algorithm we need a simple method providing a one-to-L invertible mapping between $\mathbf{x}$ and $\mathcal{C}_x$. The method called *guided scrambling*, suggested by Fair *et al.* [1], has selection sets of size $L = 2^r$ (r is the number of redundant bits) and can be summarized as follows.

1. Given the source word $\mathbf{x}$, generate the set $\mathcal{B}_x = \{\mathbf{b}_1, \ldots, \mathbf{b}_L\}$ by preceding $\mathbf{x}$ by all the possible binary sequences of length r. Hence:

$$\mathbf{b}_1 = (0, 0, \ldots, 0, x_1, \ldots, x_m), \quad \ldots,$$

$$\mathbf{b}_L = (1, 1, \ldots, 1, x_1, \ldots, x_m).$$

(This process is called *augmenting*.)

2. All the vectors in $\mathcal{B}_x$ have to be run through a self-synchronizing scrambler (see below) to obtain the elements of $\mathcal{C}_x$.

3. Select from $\mathcal{C}_x$ the best codeword for transmission.

4. At the receiver end, descramble the codeword and remove the first r bits. We then have the original source word $\mathbf{x}$.

A. Self-synchronizing scrambler

A self-synchronizing scrambler of length s performs the following operation on its input $\mathbf{u} = (u_1, \ldots, u_n)$ to produce $\mathbf{v} = f(\mathbf{u})$ (see Figure 3a):

$$v_i = u_i + \sum_{k=1}^{s} a_k v_{i-k}, \qquad (4)$$

where $v_i = 0$ ($i = -1, .., -k + 1$) in the case we are examining (so-called *block guided scrambling*), which means that the contents of the scrambler before scrambling consists entirely of zeros.

At the receiver end, $\mathbf{u}$ can be obtained from:

$$u_i = v_i + \sum_{k=1}^{s} a_k v_{i-k}.$$

The structure of the descrambler is illustrated in Figure 3b.

The scrambler can be briefly characterized by its *characteristic polynomial*:

$$s(x) = 1 + \sum_{k=1}^{s} a_k x^k.$$

a. Scrambler
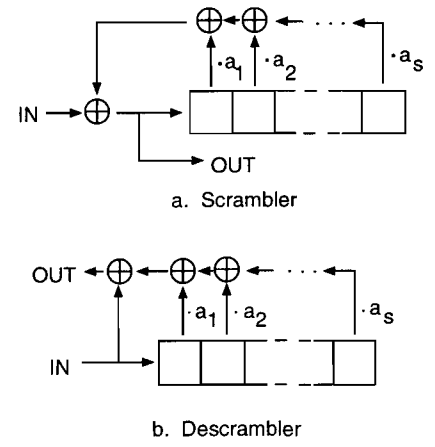
b. Descrambler

Fig. 3. Self-synchronizing scrambler.

B. Augmenting

To understand the effect of augmenting, several properties of scramblers should be noted. First, the scrambler is linear in the following sense: if $f(\mathbf{u}_1) = \mathbf{v}_1$ and $f(\mathbf{u}_2) = \mathbf{v}_2$, then:

$$f(\mathbf{u}_1 + \mathbf{u}_2) = \mathbf{v}_1 + \mathbf{v}_2. \qquad (5)$$

Considering the above, complementing the ith bit of $\mathbf{u}$ means adding to $\mathbf{u}$ the ith unity vector defined as $\mathbf{e}_i = (0, 0, \ldots, 1, \ldots, 0, 0)$, where the single one is at position i. Let us call $\mathbf{s}_i = f(\mathbf{e}_i)$ the ith *characteristic sequence* of the scrambler.

If we use $\mathbf{x}'$ to denote the vector $(0, \ldots, 0, x_1, \ldots, x_m)$ ($\mathbf{x}$ preceded by r zeros) and $\mathbf{c}' = f(\mathbf{x}')$, then the process of augmenting can be considered as follows: if $\mathbf{x}$ is preceded by $(d_1, \ldots, d_r)$, then, according to (5), the corresponding element of $\mathcal{C}_x$ will be:

$$f\left(\mathbf{x}' + \sum_{i=1}^{r} d_i \mathbf{e}_i\right) = \mathbf{c}' + \sum_{i=1}^{r} d_i \mathbf{s}_i.$$

A scrambler can be considered "good" if adding the linear combinations of $\{\mathbf{s}_i\}$ ($i = 1, \ldots, r$) produces sufficiently different codewords, providing a broad enough selection set. A sufficient condition to ensure that balanced transmission is possible even under worst case conditions of the source input is that the all-ones vector is a member of the linear combinations of the characteristic sequences.

C. Scrambler examples

Fair *et al.* [1] suggested several scrambling polynomials, two of which will be described in this section.

C.1 The $1 + x$ scrambler

This scrambler uses one redundant bit, and the scrambled sequence obtained with $v_i = u_i + v_{i-1}$ actually performs discrete integration of the input sequence. The vector $\mathbf{s}_1$ is the all-ones sequence, so

124

the augmenting inverts or non-inverts the vector $\mathbf{c}'$. It can be seen that this is essentially the polarity-bit algorithm, and if binary one and zero have the same probability in the input sequence, the statistical properties are also the same.

C.2 The $1 + x^a$ scrambler

This scrambler uses a redundant bits. The algorithm can be considered as an "interleaved polarity bit" method: setting d_i to one means inverting every dth bit of $\mathbf{x}'$, beginning with the ith bit. Setting all d_i's to one inverts the whole sequence, so that this algorithm also produces a balanced code.

D. Analysis of guided scrambling

To determine the efficiency of the algorithm we need the sum variance of the code sequence. By constructing the finite state automaton model of the encoder it turned out that general results cannot easily be given. The state transition matrix depends greatly on the length and feedback points of the scrambler.

During the analysis we used the following simplified model of the algorithm. For each source block $\mathbf{x}$ we obtained the set $\mathcal{C}_x$ by *randomly* drawing L from the 2^n possible codewords independently and we transmitted the "best" of them. As we concentrated on high rate codes, the length n was assumed to be very great, e.g. $n > 50$.

One difficulty with the model is that its state space is infinite: it might happen that only codewords of positive (or negative) disparity are drawn, resulting in a monotonic increasing (or decreasing) RDS. The probability of this, however, is negligible in the circumstances considered. It can be shown that the transmission is almost certain to be balanced. To obtain a finite number of states we truncated the state space by omitting those states which can be reached from the RDS=0 state with a probability of less than 10^{-6}. (In fact, such states do not contribute significantly to the sum variance.)

D.1 Transition probabilities of the finite state machine

Let $z_i, i = 0, 1, \ldots$, denote the RDS sum value after transmission of the ith codeword (note the difference from Z_i in (1) which was defined at the bit level). We will also use z_i for the encoder state after the ith codeword has been transmitted.

In this section we will calculate the following probabilities:

$$D(z'') = \mathcal{P}\left(z_{i+1} = z'' | z_i = z'\right)$$

(for convenience's sake we will omit the condition $z_i = z'$, which is always assumed).

We make the following remarks concerning the analysis:

1. For the sake of simplicity only codes using codewords of even length are considered.
2. For reasons of symmetry we only need to calculate the probabilities for $z' \geq 0$. The results can be easily extended for $z' < 0$.
3. As already stated, we only consider states that can be reached from $z' = 0$ with a probability greater than 10^{-6}. Since the absolute value of the codeword disparity is at most n, we are not interested in $z' > n$.

We now define several quantities to make the equations shorter.

Note that the probability of drawing a codeword of disparity d from the 2^n possible codewords of length n is

$$\left(\begin{array}{c} n \\ \frac{n-d}{2} \end{array} \right) 2^{-n}. \qquad (6)$$

Using (6), the probability of the next state being $z^* = z' + d$ is

$$p_{z^*} = \left(\begin{array}{c} n \\ \frac{n-(z^*-z')}{2} \end{array} \right) 2^{-n}.$$

For states z^* where $|z^* - z'|$ is odd or greater than n, let $p_{z^*} = 0$.

Denote by $z_j^*, j = 1, \ldots, L$ the state drawn in the jth draw. The state finally chosen will be $z'' = z_k^*$ if $|z_k^*| \leq |z_j^*|$ for all j. If k_1 and k_2 such that $z_{k_1}^* = -z_{k_2}^*$ and $|z_{k_1}^*| = |z_{k_2}^*| \leq |z_j^*|$ for all j exist, we choose between $z_{k_1}^*$ and $z_{k_2}^*$ randomly.

We introduce the quantity $R(a, b)$ for the probability that during the L draws no state is drawn from the interval $[a, b]$.

$$R(a, b) = \left(1 - \sum_{i=a}^{b} p_i \right)^L,$$

and we use $R(-\infty, b)$ for the probability of drawing only states having RDS greater than b.

According to the value of z'' we can distinguish two cases:

1. If $z'' > n - z'$, then $p_{-z''} = 0$ because $|-z'' - z'| > n$. In this case we call z'' *single*.
2. If $|z''| \leq n - z'$, then $p_{-z''} > 0$ and z'' is *double*.

In the single case $D(z'')$ is easy to calculate:

$$D(z'') = R(-\infty, z'' - 1) - R(-\infty, z''). \qquad (7)$$

If z'' is double, we define the following probabilities for a *non-negative* z'':

$$D_1(z'') = \mathcal{P}\{z'' \text{ is minimal and } -z'' \text{ was not drawn}\}$$

$$= R(-z'', z''-1) - R(-z'', z'').$$

Similarly,

$$D_1(-z'') = \mathcal{P}\{-z'' \text{ is minimal and } z'' \text{ was not drawn}\}$$

$$= R(-z''+1, z'') - R(-z'', z'').$$

We define $D_2(z'')$ as the probability of cases in which among the states drawn both a negative and a positive state have minimal $|RDS|$:

$$D_2(z'') = \mathcal{P}\{z'' \text{ is minimal and } -z'' \text{ was drawn}\}$$

$$= R(-z''+1, z''-1) - R(-z'', z'') - D_1(z'') - D_1(-z'').$$

As mentioned above, if both z'' and $-z''$ were drawn, we choose between them randomly. Hence:

$$D(z'') = D_1(z'') + \frac{1}{2}D_2(z'') =$$

$$\frac{1}{2}\left[R(-z''+1, z''-1) + R(-z''+1, z'') - R(-z'', z''-1) - R(-z'', z'')\right]. \quad (8)$$

Similarly,

$$D(-z'') = D_1(-z'') + \frac{1}{2}D_2(z'') =$$

$$\frac{1}{2}\left[R(-z''+1, z''-1) + R(-z'', z''-1) - R(-z''-1, z'') - R(-z'', z'')\right]. \quad (9)$$

D.2 Calculating the sum variance

Using equations (7), (8), and (9), the transition probabilities for each pair of RDS states can be determined. Let us denote the largest significant state by K. After omitting the non-significant states we get the truncated transition probability matrix Q with elements $q_{i,j}$ $(-K \le i \le K, -K \le j \le K)$.

The sum variance can be obtained from [2, p. 199]:

$$s^2 = E\{Z_i^2\} = E\{z_i^2\} - \frac{1}{6}(n^2-1)r_0, \quad (10)$$

where r_0 is the correlation between two different positions of the same codeword.

The stationary probabilities of the RDS states, π_i, and the value of $E\{z_i^2\}$ can be calculated on the basis of the state transition matrix Q:

$$E\{z_i^2\} = \sum_{j=-K}^{K} \pi_j j^2. \quad (11)$$

Now only the correlation remains to be determined:

$$r_0 = E\{c_{j_1} c_{j_2}\}, \quad j_1 \ne j_2.$$

We know that [2]

$$r(k) = E\{c_{j_1} c_{j_2} | z_{i+1} = z_i + 2k\} = \frac{1}{n-1}\left(\frac{4}{n}k^2 - 1\right).$$

Using the above, we obtain that

$$r_0 = \sum_{j=-K}^{K} \pi_j \sum_{l=-K}^{K} Pr\{z_{i+1} = l | z_i = j\} r\left(\frac{l-j}{2}\right)$$

$$= \sum_{j=-K}^{K} \pi_j \sum_{l=-K}^{K} q_{j,l} \frac{1}{n-1}\left[\frac{4}{n}\left(\frac{l-j}{2}\right)^2 - 1\right]$$

$$= \frac{1}{n-1}\left\{\frac{1}{n}\left[\sum_{j=-K}^{K} \pi_j \sum_{l=-K}^{K} q_{j,l}(l-j)^2\right] - 1\right\}. \quad (12)$$

The sum variance can be determined using (10) and (12):

$$s^2 = \sum_{j=-K}^{K} \pi_j j^2 - \frac{1}{6}(n^2-1)r_0 = \sum_{j=-K}^{K} \pi_j j^2 -$$

$$-\frac{n+1}{6n}\left[\sum_{j=-K}^{K} \pi_j \sum_{l=-K}^{K} q_{j,l}(l-j)^2\right] + \frac{n+1}{6}. \quad (13)$$

D.3 Results of the analysis

On the basis of (13) we calculated the efficiency of the random drawing algorithm. Figure 4 shows the efficiencies for different codeword lengths and redundancies. The points connected have the same redundancy R, and the ith point on a curve belongs to a code with i redundant bits and codeword length $i\frac{1}{R}$. For comparison purposes we have also plotted the asymptotic efficiency of the polarity bit algorithm (see (3)) in the figure.
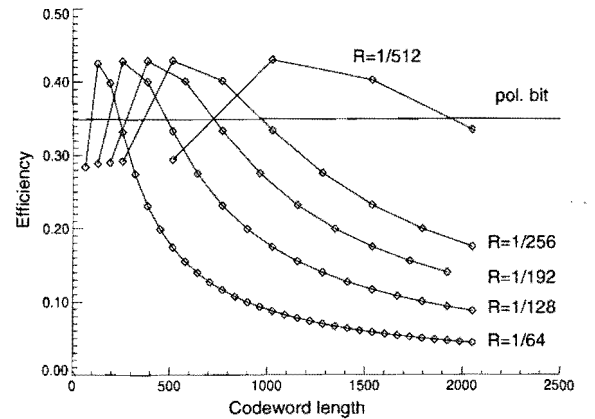
Fig. 4. Efficiency of random drawing algorithm.

We can see from the curves that the efficiency of a code is essentially determined by the number of redundant bits used. If the number of redundant bits

is fixed, the code efficiency is practically independent of the codeword length. Codes having two redundant bits (or, equivalently, $L = 2^2 = 4$ random draws) have the greatest efficiency. Codes using three redundant bits are also efficient.

With increasing number of redundant bits the efficiency tends to zero. This can be explained by the fact that two or three redundant bits provide a broad enough selection set to have a codeword with small sum variance. Increasing the redundancy further does not decrease the sum variance significantly, hence the efficiency decreases.

The use of two redundant bits means that four different codewords have to be compared before transmission. This requires a more complex encoder than the polarity bit algorithm, but the encoder complexity still remains relatively low.

To check the validity of the calculations, we performed simulations on the guided scrambling algorithm using several different scrambling polynomials. The efficiencies obtained for pseudorandom scramblers corresponded well with the analytical results. The $1 + x^a$ scramblers, however, showed smaller efficiencies than the calculated values. The $1 + x^2$ scrambler still produced a slightly better efficiency than the polarity-bit algorithm, but the $1 + x^3$ scrambler was inferior to it. Finding the reasons of lower efficiencies of the latter scramblers can be subject of further research.

IV. CONCLUSIONS

We used a new efficiency criterion to examine the guided scrambling coding algorithm. We have found that if two redundant bits are used the guided scrambling algorithm is more efficient than the polarity bit algorithm and requires only relatively low encoder complexity.

REFERENCES

[1] I. J. Fair, W. D. Grover, W. A. Krzymien, and R. I. MacDonald, "Guided Scrambling: A New Line Coding Technique for High Bit Rate Fiber Optic Transmission Systems", *IEEE Trans. Commun.*, vol. 39, no. 2, Feb. 1991.

[2] K. A. S. Immink, *Coding Techniques for Digital Recorders*, Prentice-Hall, Inc., Englewood Cliffs, 1991.

[3] K. A. S. Immink, *Properties and Constructions of Binary Channel Codes*, Ph.D. thesis, Eindhoven University of Technology, 1985.

[4] J. Justesen, "Information Rates and Power Spectra of Digital Codes", *IEEE Trans. Inf. Theory*, vol. IT-28, no.3, May 1982.

Design of distributed restoration algorithms for ATM meshed networks

Kris Struyve, Piet Demeester
University of Gent - IMEC
Department of Information Technology (INTEC)
Sint-Pietersnieuwstraat 41, 9000 Gent, Belgium
Tel: +32.9.264.33.16, Fax: +32.9.264.35.93
Email: kris.struyve@intec.rug.ac.be

Leo Nederlof, Luc Van Hauwermeiren
Alcatel Corporate Research Centre, Antwerp, Belgium

Abstract

This paper presents two dynamic distributed restoration algorithms, i.e. LINKRES and PATHRES, to restore disrupted traffic in meshed broadband ATM networks due to single or multiple link failures or node failures. The performances of these algorithms are analysed on two test networks. The results are compared with other distributed restoration algorithms published in literature and with graph theoretical algorithms. It is shown that LINKRES and PATHRES achieve fairly high restoration ratios in seconds, though they are less efficient in terms of resource utilisation compared to the graph theoretical algorithms.

1.   Introduction

Network failures in high-speed and highly integrated broadband public networks can cause huge loss of service. This results in serious revenue loss for business customers, causes inconveniences for residential customers and can be critical for human life-savings in case of emergency situations. Hence, survivable network design ensuring service continuity at an affordable cost becomes increasingly important[1].

A number of survivability techniques have been proposed to protect traffic against network failures. Usually these techniques are engineered for a specific transport architecture, i.e. Wavelength Division Multiplexing (WDM), Synchronous Digital Hierarchy (SDH)[2,3,4,5,6], Asynchronous Transfer Mode (ATM)[7,8,9], and a specific network topology, i.e. ring or mesh. Survivability techniques are further classified[12] based on their control mechanism, i.e. centralised or distributed, their type of alternate routing, i.e. link or path, and their route calculation process, i.e. preplanned or dynamic.

In case of *centralised* survivability techniques, a central operations system (OS) calculates the alternate routes to reroute the affected traffic due to the network failure. In case of *distributed* techniques, the reconfigurable cross-connects (CC) themselves calculate the alternate routes in a parallel manner. Techniques with a combined centralised and distributed control are also possible. Centralised techniques use resources more efficiently since the OS has a global view of the network status - in terms of connectivity and spare resources - at the time of the failure. Distributed techniques in contrast have only local information and are likely to be less optimal. Though, centralised techniques require reliable communciation means between the OS and the CCs and extra care to maintain a consistent and up-to-date view of the network. Furthermore, distributed techniques achieve much higher restoration speeds than centralised techniques.

Survivability techniques using a *link* or *local* type of alternate routing reroute affected connections between the nodes adjacent to the failure. Techniques using a *path* or *source-destination* type of alternate routing reroute affected connections between their end nodes. Link restoration techniques tend to be faster and easier to implement, whereas path restoration techniques tend to be more efficient in terms of resource utilisation.

Preplanned survivability techniques make use of preassigned spare stand-by resources at the time of the failure, e.g. Automatic Protection Switching (APS 1:N) and Self Healing Rings (SHR). *Dynamic* techniques on the other hand make use of spare resources which are available at the time of the failure, e.g. dynamic distributed restoration algorithms. Preplanned techniques are extremely fast - 100 ms - but are economically less attractive due to the deployment of dedicated spare resources. Moreover, dynamic techniques are more flexible to

cope with multiple failures whilst preplanned techniques are designed for a specific limited set of possible failure conditions.

This paper presents two dynamic distributed restoration algorithms for meshed public broadband networks, i.e. LINKRES and PATHRES. The approach taken in LINKRES is based on an initial distributed link restoration algorithm[10]. The most significant difference is the extension of the confirm phase to optimise the restoration of node failures. PATHRES is derived from LINKRES, but uses path restoration.

The rest of the paper is organised as follows: section 2 discusses two dynamic distributed restoration algorithms LINKRES and PATHRES. Section 3 analyses the performances of the algorithms on two test networks and compares these results with results of other distributed restoration algorithms published in literature. Finally, conclusions are presented in section 4.

2. Link versus path restoration

The first algorithm, LINKRES, uses a link restoration approach (Figures 1 and 2). Whereas most distributed restoration algorithms using link restoration presented in literature are inadequate for multiple link failures and node failures, LINKRES is able to recover from single and multiple link failures as well as node failures. Komine's algorithm[5] handles node failures, but assumes unidirectional trails. However, in order to comply with ITU recommendations[11], ATM trails are to be bidirectional and both directions should follow the same route. Hence, the designation of sender and chooser roles to the nodes adjacent to the failure, based on upstream and downstream positioning, is not free of ambiguity. If, on the other hand, the designation is based on an arbitration convention such as node identifier (ID), deadlock occurs if the node with the lowest ID fails. Indeed, the neighbour nodes, having higher IDs, cannot distinguish a link failure from a node failure. They play the role of choosers and await in vain request messages coming from the failed sender node.

The second algorithm, PATHRES, on the other hand, uses a path restoration approach (Figure 3). As any other path restoration technique, PATHRES is also able to recover from single as well as multiple link and node failures. A major drawback of dynamic distributed path restoration algorithms is that for every failed connection a restoration process is invoked. So, many more restoration messages are

generated which slows down the overall restoration speed. However, PATHRES is to be used in combination with rapid back-up VPC restoration techniques[8,9]. Hence, during the restoration phase some of the disrupted VP connections are recovered via their preassigned back-up VPCs while the other VP connections, which do not have preassigned back-up VPCs, are recovered using PATHRES. Next, during the post-restoration phase PATHRES is used to reconstruct back-up VPCs. In order to further reduce the number of messages, failed connections between the same pair of end nodes may be grouped into VP groups (VPG).
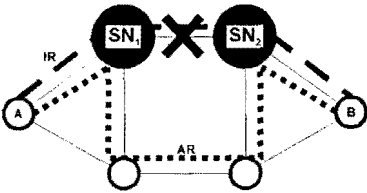
Figure 1   In case of link restoration, the two end nodes of the failing link are designated sender nodes. The initial route (IR) is rerouted to the alternative route (AR) between these sender nodes.
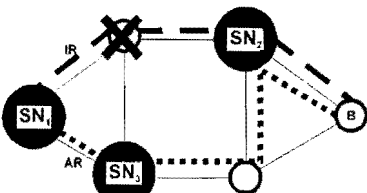
Figure 2   In case of link restoration, the nodes adjacent to the failing node are designated sender nodes. The initial route (IR) is rerouted to the alternatieve route (AR) between the matching sender nodes SN_1 and SN_2.
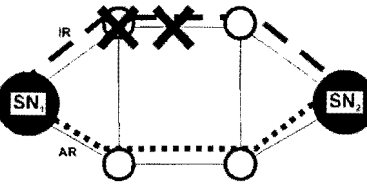
Figure 3   In case of path restoration, the end nodes of each failing restoration unit (due to a link or a node failure) are designated sender nodes. The initial route (IR) is rerouted to the alternative route (AR) between these end nodes.

Both LINKRES and PATHRES assume bidirectional VP connections. Since all restoration messages express bandwidth information in pairs, the algorithms can handle symmetrical as well as asymmetrical connections. Furthermore, whereas conventional SENDER - CHOOSER algorithms flood request messages from one single sender node all the way to the chooser node, LINKRES and PATHRES flood request messages from multiple

sender nodes towards each other. Hence, these algorithms can set the hop count limit to almost half of the conventional algorithms which reduces the size of the restoration area. The algorithms are designed for ATM networks, but with some minor modifications they can also be applied in SDH networks.

2.1. Link restoration algorithm LINKRES

LINKRES utilizes four phases (Figure 4) to restore disrupted traffic between the nodes adjacent to the failure, i.e. a request phase, a confirm and decision phase, a connect phase and a release phase. In the following paragraphs we will discuss each phase of the algorithm in turn.

The nodes adjacent to the network failure, called *sender nodes*, detect the failure and invoke the *request phase*. In case of a link failure (Figure 1), the two end nodes of the link act as sender nodes, while in case of a node failure (Figure 2), the neighbour nodes of the failing node act as sender nodes. During the request phase, nodes search for candidate alternative routes by selectively flooding *request messages*. Each sender node selectively floods request messages on all of its outgoing links excluding the link on which it detected the alarm. *Tandem nodes* store and selectively rebroadcast updated copies of received request messages. Hence, logically, N trees TR_n (n=1,...,N) are constructed branch-per-branch originating from the N sender nodes SN_n. Nodes explicitly reserve spare capacity requested in the propagating request messages. If the sum of the requested capacity of all forwarded request messages is less than the requested capacity of the corresponding received request message, this received request is stored in a *hold buffer*. Later on, if reserved spare capacity is released, the requests in the hold buffer are further rebroadcast. In addition to the requested capacity, request messages contain a *sender ID* SN_i to identify the tree TR_i to which the request belongs. *Signatures* are used to distinguish between different requests on the same link belonging to the same tree. A *hop count* is included to limit the extent of the restoration area. While expanding, branches of different trees, e.g. TR_i and TR_j, meet at *collision nodes* CN_{ij}^k (k=1,...,K). A collision identifies a candidate alternative route CAR_{ij}^k and triggers the next phase of the algorithm, i.e. the confirm and decision phase. A tandem node determines whether two requests collide based on the contents of their *list of other sender nodes*. Each sender node initializes this list of a request message with the IDs of all other sender nodes which are one

or two hops away and between which at least one connection is disrupted.
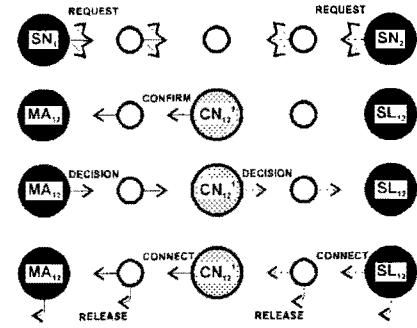
Figure 4   Overview of the execution phases of LINKRES.

The *confirm and decision phase* is necessary to resolve so called *over-requests*. An over-request occurs during the request phase if the sum of the requested capacity of all forwarded requests is greater than the requested capacity of the matching received request. In order to prevent looping and enable localisation of over-requests, request messages contain a *route* field. The collision node CN_{ij}^k assigns an unique key to the candidate alternative route CAR_{ij}^k, designates the two originating sender nodes SN_i and SN_j of the interfering trees TR_i and TR_j as *master node* MA_{ij} (e.g. SN_i) and *slave node* SL_{ij} (e.g. SN_j) - based on an arbitration convention (e.g. node ID) - and forwards a *confirm message* towards the master node MA_{ij}. The confirm message is a combination of the two colliding request messages and informs the master node concerning the other end node (i.e. SL_{ij}), the capacity and the route of CAR_{ij}^k. Upon receiving the confirm message, the master node MA_{ij} detects and resolves all over-requests in the *master part* and some of the over-requests in the *slave part* of the candidate route. Hence, the master node forwards its partial decision by means of a *decision message* to the slave node SL_{ij} via the candidate alternative route CAR_{ij}^k. Finally, upon receiving the decision message, the slave node is able to determine the actual available capacity of the route. As such, the CAR_{ij}^k becomes permanent and the *connect phase* is instigated.

The slave node SL_{ij} allocates disrupted connections to the alternative route AR_{ij}^k and transmits a *connect message* towards the master node along AR_{ij}^k. Nodes, upon processing a connect message, perform the necessary cross-connections, mark the requested capacity as permanently connected and forward *release messages* wherever over-requests occured to break down obsolete parts of the request trees. Releasing a branch of a request tree TR_i stimulates

130

the elaboration of another tree TR_j and thus the continuation of the search process.

2.2. Path restoration algorithm PATHRES

Though PATHRES is similar in approach to LINKRES, there are some major differences which simplify the algorithm (Figure 5). Suppose that each VPC or VPG - from now on shortly denoted as restoration unit (RU) - has an unique key known by both of its end nodes. Upon detecting alarm signals, the end nodes selectively flood request messages per failed RU. Unlike LINKRES where a sender node SN_i floods one type of request messages which collide with request messages originating from other different sender nodes SN_j $(j=1,...,i-1,i+1,...N)$ dependent on their list of other sender nodes, in this algorithm different trees originate at a sender node each representing one RU_{ij}. Furthermore, a tree TR_{ij} only collides with its complementary tree TR_{ji} originating from the other end node SN_j. The flooding of request messages in tandem nodes is even more restricted, i.e. apart from hop count limitation and loop prevention, only the first request message of a tree TR_{ij} is flooded. All following request messages belonging to TR_{ij} are released. A request message is flooded on a link if, and only if, the unreserved capacity of the link equals at least the capacity of RU_{ij}. Over-requests are still possible, though only one request message of TR_{ij} is ever flooded on a link, even if reserved spare capacity is released.
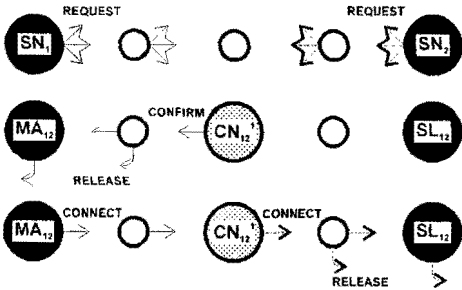
Figure 5   Overview of the execution phases of PATHRES.

Trees TR_{ij} and TR_{ji} meeting at intermediate nodes CN_{ij}^{k} identify a CAR_{ij}^{k}. To retain a single alternative route from the multiple collisions between TR_{ij} and TR_{ji}, the collision node transmits a confirm message towards the master node MA_{ij}. The alternative route is selected based on a first-come-first-served basis, i.e. the first confirm message arriving at MA_{ij} determines the CAR_{ij}^{k} which is used to reroute RU_{ij}. A tandem node propagates a confirm message towards the master node if, and only if, previously it has not received any confirm messages related to other candidate alternative routes between MA_{ij} and

SL_{ij}. If so, the node releases also the corresponding outgoing branches, with exception of the branches which are part of CAR_{ij}^{k}. The first confirm message arriving at the master node is made permanent by issueing a connect message towards the slave node. Upon processing a connect message, the node performs the cross-connection. While obsolete parts of the (master) tree TR_{ij} are released during the confirm phase, the obsolete parts of the (slave) tree TR_{ji} are released whenever a node in the slave part receives a connect message. Upon receiving the connect message at the slave node, the rerouting of the affected RU_{ij} is completed.

3.   Simulation results

In this section we will analyse LINKRES and PATHRES on two different test networks, i.e. the New Jersey LATA network (Figure 6) and the CATS network (Figure 7), using an object-oriented discrete-event simulator ATMSIM. Each node in the network is equiped with a single-server-FIFO controller to process restoration messages. We adopted following assumptions to be inline with the RREACT paper[4]: all messages are 64 bytes long, all messages have equal priority, it requires 10 ms to service any arriving message and an additional 10 ms to generate each forwarded message, the propagation delay on any link is set to 0.5 ms in the New Jersey LATA network and to 1 ms in the CATS network, the signal transmission rate is set to 64 kbit/s, bandwidth is expressed in *equivalent units*, the bandwitdth of any VPC equals one unit, all VPCs are symmetric and fault detection times are neglected.

Figure 6   New Jersey LATA network (11 nodes, 23 links)

In the following simulations we have also solved the restoration problems using graph theoretical algorithms, i.e. Busacker-Gowen and Dijkstra (not

taking time aspects into consideration nor any hop limit (HL) restrictions). This allows for a resource utilisation and restoration ratio comparison between dynamic distributed algorithms and centralised algorithms. It turns out that distributed restoration algorithms are suboptimal, b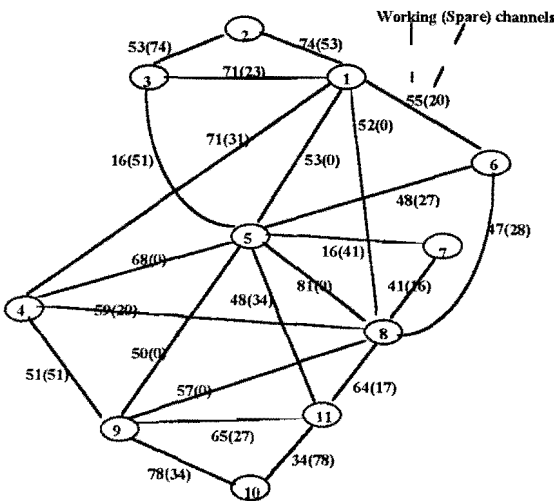ut the restoration speed of distributed algorithms is shown to be on the order of seconds, whereas the restoration speed of centralised algorithms is expected to be on the order of minutes or more[12].
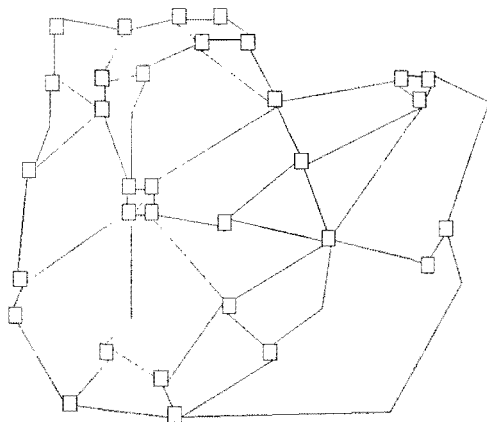
Figure 7   CATS test network (32 nodes, 54 links, 340 VPCs)

3.1. Evaluations on the New Jersey LATA network

The RREACT paper presents simulation results of link failures in the New Jersey LATA network using different link restoration algorithms, i.e. FITNESS[6], Two-Prong[3] and RREACT[4]. Table 1 provides a comparison of the restoration performances of these three algorithms and of LINKRES for a particular example link failure. These results show that LINKRES attains full restoration and is as cost efficient as RREACT and Two-Prong. Though the restoration speed is lower compared to Two-Prong, LINKRES outperforms FITNESS by far. From Table 2 we can conclude that an average restoration ratio of 98% is achieved within the typical 2 second restoration requirement. LINKRES assigns affected VPCs to alternative routes on a first-come-first-served basis and hence, as argued by Anderson et al[13], is likely to be suboptimal. The optimal solution for the single-commodity minimum-cost restoration problem, which we have calculated with the Busacker-Gowen algorithm (Table 3), confirms this statement. The evaluations on the New Jersey LATA network show that the performance of LINKRES is comparable to RREACT and Two-Prong while offering the extra asset of restoring node failures. This is demonstrated on the CATS network.

Restoration algorithm	link failure N05-N08 (81 working channels lost)		
	Restoration ratio (%)	Restoration resource usage	Restoration time (ms)
RREACT	100	204	402
FITNESS	100	289	1756
Two-Prong	100	210	273
LINKRES	100	204	647

Table 1 Comparison of four dynamic distributed link restoration algorithms for a particular example single link failure in the New Jersey LATA network.

Average restoration ratio (%)	Restoration time (s) (LINKRES,HL = 5)
85	0.60
90	0.75
95	1.15
100	5.00

Table 2 Restoration times necessary to achieve specific average restoration ratio thresholds using LINKRES in the New Jersey LATA network.

	LINKRES (HL = 5)	BUSACKER-GOWEN
Average restoration ratio (%)	100	100
Restoration time (s)	5.00	-
Total resource usage of restored connections	4322	4182

Table 3 Restoration performance of LINKRES for all possible single link failures in the New Jersey LATA network (the total resource usage of affected connections equals 1252). The restoration performance achieved with Busacker-Gowen is also included.

3.2. Evaluations on the CATS network

We will first consider single link failures and next node failures. For the evaluation of PATHRES, we aggregate all VP connections between the same pair of end nodes in one VPG and use these VPGs as restoration units. Moreover, we assume that before the PATHRES algorithm is invoked all affected VPCs are torn down. So, more spare capacity is available for restoration.

3.2.1. Single link failures

Tables 4 and 5 as well as Figure 8 show that, for the given default assumptions, LINKRES restores all link failures within 1.4 s, whereas PATHRES attains full average restoration within 7.1 s. PATHRES is clearly much slower than LINKRES. The worse time performance of PATHRES is due to the higher number of restoration processes invoked (i.e. summed over all link failures: 54 for LINKRES versus 836 for PATHRES) and consequently the larger amount of restoration messages generated. Furthermore, in case of PATHRES the restoration

132

information has to travel between the end nodes of the VPGs, whilst in case of LINKRES the information travels between the nodes adjacent to the failure. As a result of the differing restoration approach the hop limit (HL) for LINKRES is set to 3 whereas for PATHRES it is set to 6.

	Link restoration		Path restoration	
	LINKRES (HL = 3)	BUSACKE R-GOWEN	PATHRES (HL = 6)	DIJK-STRA
Average restoration ratio (%)	100	100	100	100
Restoration time (s)	1.40	-	7.10	-
Total resource usage of restored connections	6719	6679	6017	5111

Table 4 Restoration performance of LINKRES versus PATHRES for all possible single link failures in the CATS network (the total resource usage of affected connections equals 3824). The performances achieved with Busacker-Gowen and Dijkstra for respectively link and path restoration are also included.

	Link restoration	Path restoration
Average restoration ratio (%)	Restoration time (s) (LINKRES,HL = 3)	Restoration time (s) (PATHRES,HL=6)
85	0.60	2.90
90	0.70	3.40
95	0.85	4.20
100	1.40	7.10

Table 5 Restoration times necessary to achieve specific average restoration ratio thresholds using LINKRES versus PATHRES in the CATS network.

Average restoration ratio (%)	Restoration time t(s) (SER = 10 ms, GEN = 10 ms)	Restoration time t(s) (SER = 10 ms, GEN = 0.1 ms)	Restoration time t(s) (SER = 100 ms, GEN = 0.1 ms)
85	0.60	0.35	2.15
90	0.70	0.40	2.50
95	0.85	0.55	3.05
100	1.40	0.85	5.05

Table 6 Restoration times necessary to achieve specific average restoration ratio thresholds using LINKRES in the CATS network for different service and generation delays.

Table 4 also indicates the optimal solution obtained with the Busacker-Gowen algorithm for link restoration. For path restoration we sequentially searched for all affected VPGs the shortest (in terms of cost) alternative path using the Dijkstra algorithm, each time updating the spare capacity of the links in the network. We cannot apply Busacker-Gowen as this algorithm possibly splits up the affected flow into multiple alternative flows. It should be stressed that the iterative Dijkstra restoration solution[14] does

not provide us with the optimal minimum-cost solution, i.e. the outcome depends on the order in which the VPGs are restored. Path restoration is more cost efficient than link restoration, i.e. path restoration uses less spare capacity to reroute all the affected connections. Moreover, LINKRES and PATHRES are shown to be less cost efficient compared to respectively Busacker-Gowen and Dijkstra.

Table 6 illustrates the influence of the service and generation delays of restoration messages in the controllers on the overal restoration time performance. Neglecting the delays associated with the generation of forwarded new restoration messages, reduces the restoration time with almost 35%. Detailed studies of restoration message processing in ATM cross-connect are ongoing. Preliminary simulation results indicate that service delays measure approximately 100 ms and that the generation delays are indeed neglectable. Hence, full average restoration takes 5 s instead of 1.4 s.

Figure 8 Average restoration ratio in function of restoration time to restore single link failures in the CATS network using LINKRES versus PATHRES.

3.2.2. Node failures

Evaluation results of node failures using LINKRES and PATHRES are summarized in Tables 7 and 8 and Figure 9. Note that for the calculation of the metrics only restorable connections - connections not terminating at the failing node - are taken into account. PATHRES restores node failures more adequately than LINKRES, i.e. LINKRES restores less than 90%, whereas PATHRES easily restores 95%. Again we note that link restoration is remarkably faster than path restoration: LINKRES reaches an average restoration ratio of 85% after 1.8

s whilst PATHRES needs 3.5 s Similar conclusions as for link failures can be drawn: PATHRES is more efficient than LINKRES, while the graph algorithms achieve better performances.

	Link restoration		Path restoration	
	LINKRES (HL = 5)	BUSACKE R-GOWEN	PATHRES (HL=7)	DIJK-STRA
Average restoration ratio (%)	88.67	97.79	97.53	99.05
Restoration time (s)	3.75	-	8.20	-
Total resource usage of rerouted connections	3728	4443	4026	3812

Table 7 Restoration performance of LINKRES versus PATHRES for all possible single node failures in the CATS network (the total resource usage of affected connections equals 2802). The performances achieved with Busacker-Gowen and Dijkstra for respectively link and path restoration are also included.

	Link restoration	Path restoration
Average restoration ratio (%)	Restoration time (s) (LINKRES, HL = 5)	Restoration time (s) (PATHRES,HL=7)
80	1.30	3.00
85	1.85	3.50
90	-	4.00
95	-	5.10
100	-	-

Table 8 Restoration times necessary to achieve specific average restoration ratio thresholds using LINKRES versus PATHRES in the CATS network.

Whereas link restoration of a single link failure is a single-commodity flow rerouting problem - just one pair of nodes adjacent to the link failure - link restoration of a node failure is a multi-commodity flow rerouting problem as there are multiple pairs of nodes adjacent to the node failure between which affected flow needs to be restored. Hence, Busacker-Gowen does not provide us with the optimal minimum-cost solution. Moreover, we applied Busacker-Gowen iteratively for the different flows. Neither Busacker-Gowen (link restoration) nor Dijkstra (path restoration) succeed full restoration of all node failures, which gives us confidence on the restoration performance of LINKRES and PATHRES. Remark that PATHRES performs nearly as well as Dijkstra.

4. Summary and conclusion

Two distributed dynamic restoration algorithms were presented. Both algorithms restore bidirectional (symmetrical and asymmetrical) VPCs in meshed ATM networks which have failed due to link or node

failures. LINKRES is a link restoration algorithm, whilst PATHRES is a path restoration algorithm. In the first phase of the algorithms, restoration messages are flooded to search for candidate alternative routes. The second phase selects alternative routes amongst different candidates. Finally, during the third phase affected connections are allocated to the established alternative routes.
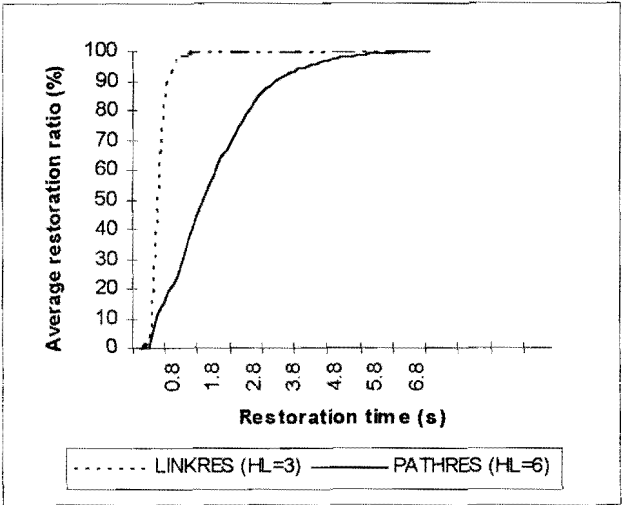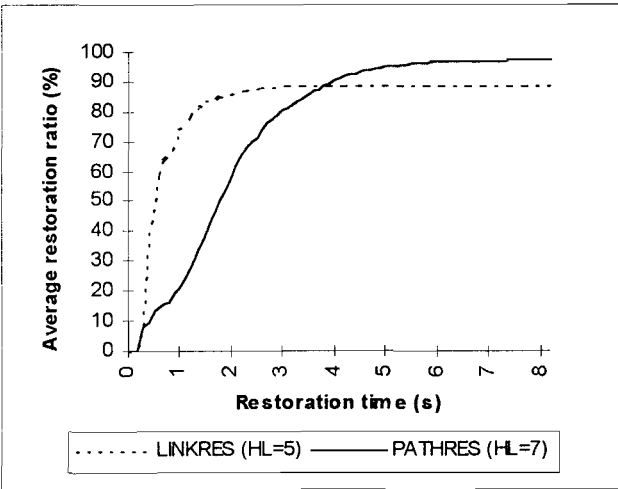

Figure 9 Average restoration ratio in function of restoration time to restore single node failures in the CATS network using LINKRES versus PATHRES.

The algorithms were analysed by computer simulations on two test networks. Simulations of single link failures on the first test network revealed that the restoration performance of the LINKRES algorithm matches the performance of the RREACT and the Two-Prong algorithms, and outperforms the FITNESS algorithm. The restoration of single link failures and node failures was simulated on the second network using LINKRES versus PATHRES. Hence, PATHRES demonstrated to achieve higher average restoration ratios for node failures than LINKRES. However, LINKRES proved to be much faster. The processing delay of a restoration message is shown to have a severe influence on the overall restoration speed performance. Concerning future work, the obtained results encourage us to enhance the restoration of link and node failures using LINKRES, and to evaluate the performance of PATHRES in co-operation with backup-VPC restoration techniques.

Acknowledgement

134

1   Leo Nederlof, Kris Struyve, Chris O'Shea, Howard Misser, Yonggang Du, Braulio Tamayo, "End-to-end Survivable Broadband Networks", IEEE Communications Magazine, September 1995

2   W.D. Grover, "The Self-Healing Network: a fast distributed restoration technique for networks using digital cross-connect machines", proceedings of Globecom '87, pp. 28.2.1-28.2.6, 1987

3   Chow, J. Bicknell, S. McCaughy, "A fast distributed network restoration algorithm", International Phoenix Conference on Computer and Communications, March 22-26, 1993, Tempe, Az.

4   Chow, S. McCaughey, S. Syed, "RREACT: A distributed network restoration protocol for rapid restoration of active communication trunks", proceedings of 2nd IEEE Network Management and Control Workshop, Westchester, 1993

5   H.Komine, T. Chujo, T. Ogura, K. Miyazaki, T. Soejima, "A distributed restoration algorithm for multiple link and node failures of transport netorks", proceedings of Globecom '90, pp. 403.4.1-5, December 1990

6   C.H. Yang, S. Hasegawa, "FITNESS: Failure Immunization Technology for Network Service Survivability", proceedings of Globecom '88, pp 47.3.1-47.3.6, November 1988

7   H.Fujii, N. Yoshikai, "Restoration Message Transfer mechanism and Restoration Characteristics of the Double-Search Self-Healing ATM network", IEEE JSAC, vol. 12, no.1, pp. 149-158, January 1994

8   R. Kawamura, K. Sato, I. Tokizawa, "Self-healing ATM network techniques utilizing VPs", 5th International Network Planning Symposium, Kobe, Japan, May 1992

9   P.A. Veitch, D.G. Smith, I. Hawker, "A distributed protocol for fast and robust VP restoration", 12th IEE UK Teletraffic Symposium, Windsor, March 1995

10 Leo Nederlof, Hans Vanderstraeten, Patrick Vankwikelberge, "A new distributed restoration algorithm to protect ATM meshed networks against link and node failures", proceedings of ISS '95, April 1995

11 ITU-T Recommendation M.3100, "Generic network information model", Geneva 1992

12 T-H. WU, "Fiber Network Service Survivability", Artech House, May 1992

13 J. Anderson, B. Doshi, S. Dravida, P. Harshavardhana, "Fast Restoration of ATM Networks", IEEE-JSAC, vol 12, no 1, January 1994

14 D. Vercauteren, P. Demeester, J. Luystermans, E. Houtrelle, "Availability analysis of multi-layer networks", 3rd International Conference on Telecommunication Systems, March 16-19, 1995

Developing Location and Tracing Methods for Fast Moving Portables.

A.C.Papavramidis, member IEEE, G.S.Tombras[†], member IEEE, G.E.Alexakis[†]

EUROCONTROL/DEI.1, Rue de la Fusee 96, B-1130 Brussels, BELGIUM
TEL.: +322 729 30 35, FAX: +322 729 90 64, E-mail: papavramides.thanos@eurocontrol.be
†Laboratory of Electronics, Dept. of Physics, University of Athens, TYPA Buildings, GR 157 71 Athens, GREECE

Abstract
Locating, tracing and addressing, are key functions in a universal PCS. While for the general case there exist recommendations and standards, fast moving subscribers (such as air travellers), represent an open challenge.
Following recent developments and trends, the paper proposes a methodology to develop and establish effective techniques for locating, tracing and addressing very fast moving portables.
In doing so the paper identifies key requirements to achieve a given set of objectives as well as critical areas where necessary work has to probe further.
The proposed feasible techniques are also treated for optimisation concerning the bulk of addressing overheads, adaptive routing and call set-up delays.

1. Introduction

A future Universal PCS should be able to provide reliable (at least) toll level quality communication (voice/data) "from any place to anybody, anytime, anywhere...". All these services should be "personal", meaning that are addressed to the individual rather than to a certain location.

These conditions cannot be met when a subscriber moves with exceptionally high speed (of the order of 1000 km/h) in a non-ordinary RF environment. For this work, fast moving portables are those that have been carried into a commercial aircraft. This results to a loss of communication at least for two reasons: i) electromagnetic isolation and ii) very high speed that is beyond the limit covered by land mobile systems (cellular, telepoint, etc.). For the case of the air traveller concern has been already expressed by international standardisation bodies and forums[1,2].

There have also been efforts that try to alleviate the problem by offering voice (and in the near future fax/data) services to air passengers by means of the so called Aeronautical Passenger Communications (APC), which uses mainly downlink VHF or satellite communications. APC and similar future systems (like the Terrestrial Flight Telephone System developed in Europe by ETSI) do not meet the basic PCS requirements, as passengers can only originate calls, when over certain areas, while themselves cannot be accessed by an external calling party.

In past work[3] we had examined the possibilities and drawn proposals for the necessary measures and adaptations to portables of current (e.g. DECT) and future mobile systems (such as FLMPTS and UMTS) that will enable them to interoperate with subnetworks within an aircraft. In this paper we assume that the integration of the portable into the aircraft's subnetworks has been achieved and we shall be dealing with the question of locating, tracing and addressing the active airborne portable.

The actual conventional methods[4,5] for tracing moving portables within a land mobile network is paging and/or registration (spontaneous or intentional), followed by authentication. Considering the situation with the airborne portable, neither technique can be applied. The airborne portable is actually *disconnected* as the aircraft moves up and over uncovered areas. Currently the use of all portables within aircraft is simply forbidden, mainly due to uncertainty concerning harmful effects on the avionics.

The paper approaches the problem by defining a domain of suitable quality and properties. With respect to this domain the active airborne portable is virtually still and perfectly accessible. It is envisaged that this special domain should have practically zero relative speed with respect to the portable, thus representing a "moving cell" (MC), around it.

The paper further describes the mechanism that allows the portable to perform a controlled transit to this MC through service access points, which could be nodes of a world-wide intelligent network [6]. Since these points represent a gateway from one state to an other, we shall call them "migration points" (MP). Thus the problem of locating, tracing and addressing the portable, collapses to the question of locating, tracing and addressing its associated "moving cell"(MC). The question of accessing the portable within the MC was addressed in [3].

The paper follows a top-down approach for a proposed system concept. In this system context, some system segments have been already developed, others are under development and others are completely open. The paper examines the various components necessary for

system integration as well as the work needed to implement the missing or open system segments.

The text generally follows the terminology established by ITU in [4].

2. The Aeronautical Telecommunication Network

How an MC (ie. a parked or flying aircraft) can be accessed by any calling party in the world and how the calling party could "know" that the called entity is actually served by the particular MC?. To answer these questions, we must first explain briefly about the Aeronautical Telecommunication Network (ATN).

The ATN[7] development is the result of work initiated by SICASP (Secondary Surveillance Radar Improvement and Collision Avoidance System Panel) of International Civil Aviation Organisation, to support Communications, Navigation and Surveillance services to aircraft. It mainly envisaged as a global data transfer switching. During the development it was decided that the network should better be of open architecture able to interconnect a number of systems (and eventually subnetworks) using standard protocols from the standard ISO/OSI reference stack) and international communication links

The general frame of the ATN architecture encompasses at least three classes of subnetworks:

- the ground subnetworks
- the airborne subnetworks
- the mobile subnetworks

The **ground subnetworks** have all their nodes on the ground and are normally dedicated to Air Transport - Air Traffic Management Services. They are open to external access (PSTN/ISDN and other) via their own nodes.

The **airborne subnetworks** are distinguished in three categories:

The *avionics subnetwork* servicing airborne CNS applications which interwork rather automatically with the related ground based ones.

The *Flight Deck subnetwork* providing access to ATN for pilots and cockpit crew and the *cabin management subnetwork* connecting passengers and cabin crew with the ground and is mostly suitable for supporting APC services.

The **mobile subnetworks** in the frame of ATN, provide connectivity between ground subnetworks and airborne ones. Due to the openness of the ATN architecture there is virtually no restriction in the use of these subnetworks. Some examples are:

- Gatelinks: A data/voice, link, based on infrared or fibre optic carriers, which provides connectivity when the aircraft is gate-parked at the airport.

- VHF Voice/Data Communications with two elements, one strictly dedicated to Air Traffic Control services and the other intended to serve APC.

- Satellite Communications (Voice & Data): currently based exclusively on the INMARSAT system offering APC and Automatic Dependent Surveillance (ADS).

The generic structure of ATN architecture is given in fig. 8. Its main element is the ATN Router which interconnects subnetworks at the network level and supports interworking for different Physical and Datalink Layers.

The organisational unit of the ATN, is the Router Domain (RD), with its Domain Main Router and possibly additional ones. An RD communicates with the "rest of ATN" via special routers which act as Boundary Intermediate Systems (BIS).

By using BIS to interconnect Router Domains (RD), the ATN exhibits both a hierarchical and a distributed structure (fig. 1).
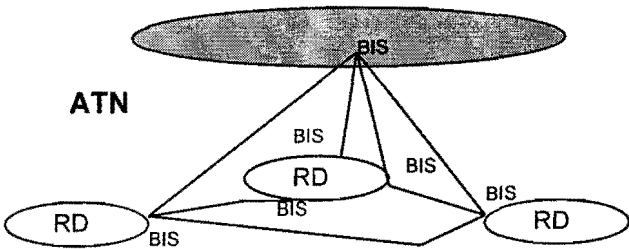
Fig. 1: Example of ATN structure

Standard ISO protocols are used at each layer level (for peer to peer communication and addressing) and special functions and applications are under specification development.

ICAO standards for the ATN air-ground (mobile) subnetworks are in the final stage of development. It has to be noted that this development has not so far took into account any potential use of these subnetworks to support PCS mobility.

3. Aspects of Addressing within ATN

As mentioned already, ATN standards follow ISO Reference Model. Hence the corresponding addressing plan has been developed accordingly to provide the network layer with addressing information to be utilised by ATN End Systems (ES), and ATN Intermediate Systems (IS) located at the boundaries of avionics, mobile and ground subnetworks. Furthermore the addressing plan of the ATN Network Service Access Point (NSAP) specifies a hierarchical, name-based format and is applicable for either fixed or mobile end and intermediate systems. (see fig. 3). Particular subnetwork addressing format was avoided since ATN ES and IS were specified to perform

137

translation between the ATN NSAP (ISO) format and possibly private format required on the next-hop sub-network. This was designed as to provide interoperability and interworking between different (existing or private) subnetworks within the ATN domain space.

IDP: Initial Domain Part
DOMAIN: Domain Identifier
LOC-AREA: Local Area Identifier
ID: System Identifier
SEL: NSAP Selector

Fig. 2: ATN NSAP address structure (from [7])

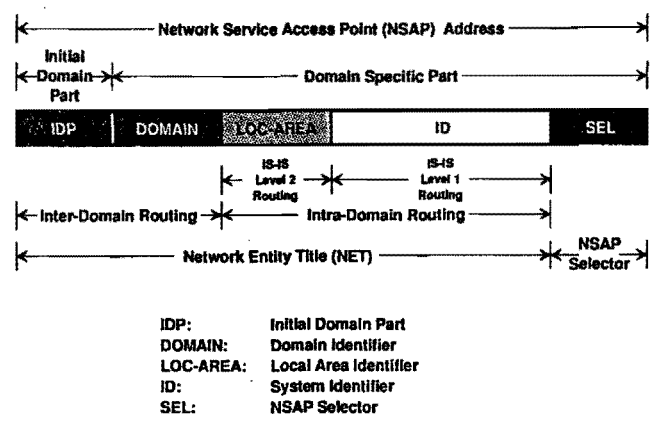The Moving Cell (ie. the aircraft) is provided with a global unique address identifier (ie. the 24 bits address as defined by ICAO Annex 10, Vol. I, Part I, Chapter 2, 2.5.6.1). This is a critically important property that actually enables the location and tracing (and eventually addressing) of an active portable within the aircraft's subnetwork. Thus every MC possesses a globally unique address and an infrastructure exists to support call routing to it.

4.   The Proposed Method
As it was already mentioned the initial concept was to assign to the fast moving portable a new domain (the moving cell) and thus converting the problem of locating and tracing the portable into locating and tracing the moving cell. We have also discussed that using the advent of ATN and its internal addressing functions together with the important fact that every aircraft in the world will have a unique ID and address within ATN, we may access a particular aircraft (ie. a moving cell) from an external to ATN network (ie. PSTN/ISDN, etc.). This possible via the ATN Boundary Intermediate Systems (ie. Routers), which can be specially designed and specified to act intelligent interfaces between ATN and PSTN/ISDN or the global AIN in the future (WIN). In the following paragraphs we shall propose a working philosophy and analyse in some detail how this can be achieved.

5.   A Case to Case Approach
To facilitate the analysis the proposed working philosophy is decomposed into individual cases. This is done with respect to factors defining the status of the mobile when a call set up is requested.

A set of key definitions is necessary to support the breakdown of the proposed methodology.
Local calls : those originated within the area served by subscriber's home Mobile Switching Centre (MSC). This covers calls originated either by a mobile or fixed terminal.
Remote call : any other call.
Local MP : any MP within the area served by the subscriber's MSC.
Remote MP : all other MPs.
As the MC moves towards its destination is sequentially served (accessed) by consecutive MPs
Entry MP : the first in the sequence (could be either *local* or *remote*).
Interim MP (*local* or *remote*) : any other MP.

The subscriber's MC can be accessible (ie. served) by either the entry MP (ie. during the first segment of the trip) or by any other MP, along its route. As a result calls are grouped into two main categories with respect to entry MP status: Group L: when the entry MP is local, and Group R: when the entry MP is remote (ie. the portable enters an ATN domain while being in a roaming condition). Each Group can be further subdivided into four cases depending whether the call is local or remote and the MC is served by the entry MP or an interim one. Table I illustrates all possible combinations.

Table I: Routing cases
for accessing the airborne portable

Group	Serving MP	CALL
L (Entry MP Local)	Entry	Local
		Remote
	Interim	Local
		Remote
R (Entry MP Remote)	Entry	Local
		Remote
	Interim	Local
		Remote

Following the designation in Table I, we may establish a code name for each case as follows: LEL case is the one for which the entry MP is local, the MC is served by entry MP and the call is local. The rest of the code names are derived accordingly.

6.   General Considerations
As the portable enters an MC (within an ATN domain), it performs registration (followed by authentication) to the associated VLR Data Base (functioning within the entry MP). Since several routers can co-

exist within the same ATN domain, they may act (following standardisation arrangements) as Base Station Systems (BSS) for the MC, but the domain's MSC shall be the entry MP. This is because MSC and HLR/VLR functions should be performed by an ATN Boundary Intermediate System (BIS), thus ensuring an effective interface between the ATN domain and the rest of WIN (or AIN).

MC's unique global address and ID (the so called Mode S address [10]) is known by default to the entry MP and thus can be associated (following addressing arrangements; see also para. 3) as a prefix to the portable's (or subscriber's) unique global address. MP-MC connection is realised via the optimum available ATN subnetwork (gatelink, VHF, satellite, etc.). Dynamic adaptive routing is part of ATN standardisation. Effective call set-up procedures within the aircraft are presented in [3] and partly in [9] and they may cover current systems (ie. GSM, DECT) or future ones (FLMTS, UMTS).

In analysing the various cases, due consideration is paid to the possibilities offered by dynamic routing and the optimised combination of hierarchy and distributed intelligent in an Advanced Intelligent Network[8].

7. Case by Case Analysis
7.1. LEL Case
The call is straightforwardly directed to the Home MSC and after the associated HLR has performed the necessary bookkeeping, the call is forwarded to the entry MP, consequently to the MC and to the active portable, thus establishing a call set-up. Fig. 3 .illustrates the proposed procedure for case LEL.

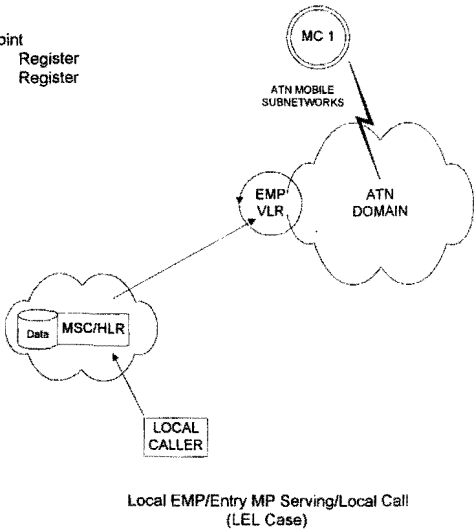Local EMP/Entry MP Serving/Local Call
(LEL Case)

Fig. 3: LEL Case

7.2. LER Case
The only difference with respect to case LEL, is that the call is originated from an area not covered by the Home MSC. After reaching the MSC the call is routed as in case LEL. Fig. 4 .illustrates the proposed procedure for case LER.

7.3. LIL Case
Here the MC has moved into another ATN domain and eventually is served by an interim MP following a handover from the entry MP. There are two possibilities for the entry MP: i) to remain active in the loop receiving and rewriting all the calls to the interim MP or ii) to relay all the routing information to the home MSC/ HLR and then withdraw from the loop. The second possibility is the optimum one since it is foreseen that remote ATN domain shall be connected via "ordinary" PSTN/ISDN links. Fig. 5 illustrates the case with the two options.

Local EMP/Entry MP Serving/RemoteCall
(LER Case)
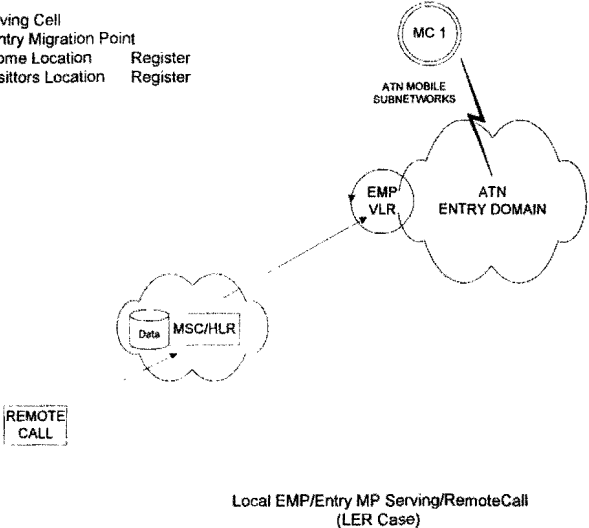
Fig. 4: LER Case

7.4 LIR Case
It could be identical to case LIL with a remotely originated call, but the advent of dynamic routing and distributed intelligent multi-layered AIN[7] offers a much more efficient alternative:

The home MSC/HLR receives updates for the new MC/portable location and address (ie. interim MP), and continuously broadcasts these data to the nearer higher layer AIN nodes having DB management capabilities.

The data are further forwarded to higher layer nodes possibly up to the highest AIN level. As a result the remote call is dynamically routed following the optimum route (concerning AIN situation and traffic conditions) to the interim MP. This location/tracing strategy constitutes a deviation from [4] and [5] recommendations and standards and is analysed

(including its benefits, requirements and performances) in [7].

The implications and the impact on Data Bases performance and capacity requirements will be discussed later. Fig. 6 depicts the case with the two alternatives.

Local EMP/Interim MP Serving/Local Call, with two options

Fig.5: LIL Case

7.5 RIR Case

This case was chosen from the total of Group R because it represents extreme conditions. The benefits of a multi-layered, distributed intelligent and quasi-hierarchical AIN are shown better. In this case the portable has entered an ATN domain while already in roaming condition. After that its MC was handed-over to an interim MP and then call was originated from a remote are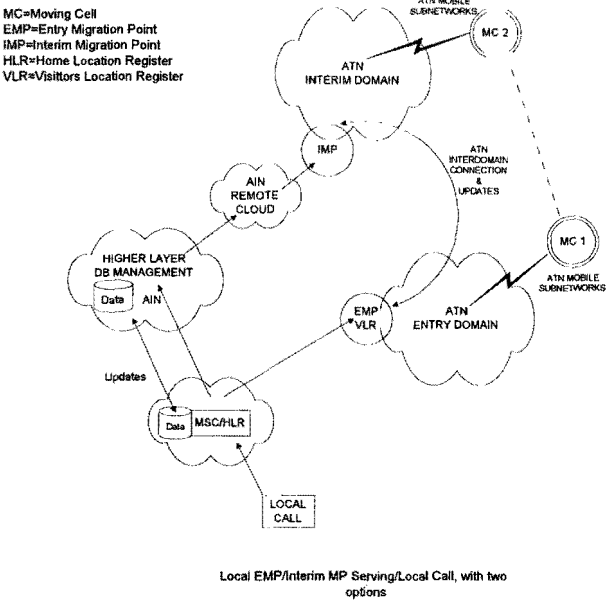a (ie. outside the Home MSC/HLR coverage). Following the recommendations in [4] and [5], the call should be routed to the portable's MSC/HLR where a bookkeeping is performed concerning the portable's whereabouts and consequently the call is forwarded to the entry MP, or to the interim MP (following the options described in para. 6.3). In the alternative proposed by [7], the call, after reaching the first higher layer node containing the necessary addressing data in its DB, is dynamically routed to its final destination (fig.7).

To better visualise the case let us imagine the following scenario: A subscriber in the Netherlands calls a mobile in Athens Greece, while the mobile itself is about half an hour before landing in Stockholm, after having started an air trip from Rome. Conventionally the call is addressed to MSC/HLR in Greece, then is forwarded to Rome and after that following ATN and/or PSTN routes reaches the interim MP in Stockholm from where is relayed to its MC and hence to the active mobile. Instead and following the proposed in [7] location/tracing strategy, the call after reaching the

nearest higher layer node (of a future AIN), is dynamically routed to Stockholm's MP and then to the mobile.

To achieve this the overall system design and specifications (in the form of a standard) should allow for a minimum defined performance.
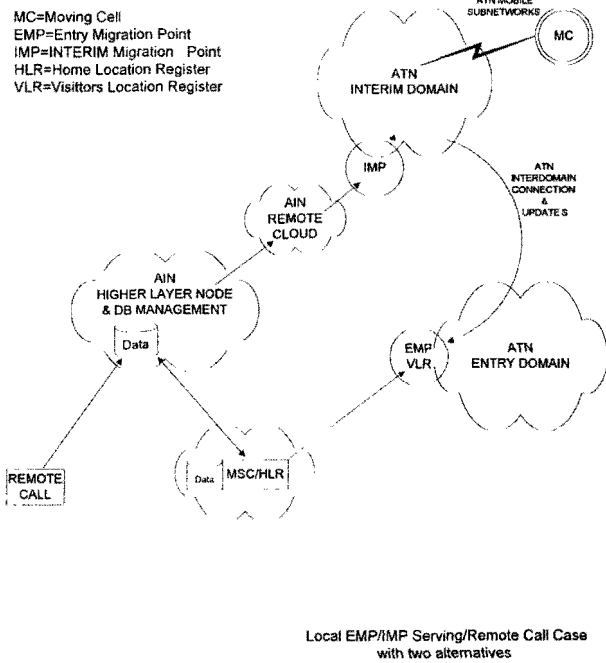

Local EMP/IMP Serving/Remote Call Case with two alternatives

Fig. 6: LIR Case

For the rest of the cases of Group R, the reader is invited to extrapolate the corresponding procedures.

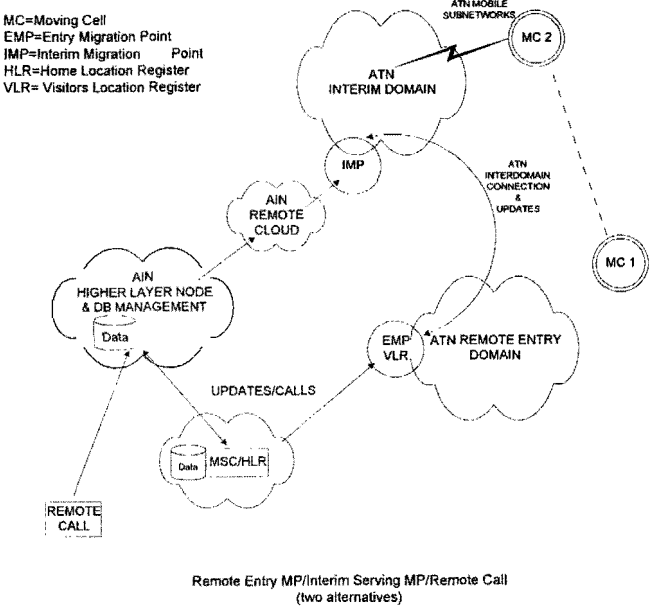Remote Entry MP/Interim Serving MP/Remote Call (two alternatives)

Fig. 7: RIR Case

8. Impact on distributed Data Bases capacity and other network performance requirements.

The proposed scheme should be specified and designed as to:

- cope with required data flow rates associated with Data Bases (eg. queries, updates, system messages, etc.)
- exhibit minimum performance (eg. delays, availability, resilience, etc.)

In order to do so it is critical to know or to estimate/model, the nature of data to be carried, the expected traffic and its projections, the Data Base system performance minimum requirements including overload control, etc.

A number of important services related to AIN have also to be taken onto account:

- authentication and verification
- encryption/decryption
- special services (call waiting, voice mail, etc.)
- billing

Effort on these issues should be co-ordinated with the development of the World-wide (Advanced) Intelligent Network.

9. Proposals for Future Action

A brief summary of proposed future action is given below:

Short term actions: Research should be carried out (as to conduct simulations and measurements) to reveal potential problem of Base Terminal Station and Portables interworking with avionics.
ATN standards to be adapted accordingly to allow for PCS services.

Medium term actions: To design suitable BTS for GSM (including its DCS1800 version) and optionally DECT for operation within aircraft (to achieve low power/weight equipment and connectivity to the cabin management subnetwork).

Long term actions: ICAO, ITU and ETSI to cooperate as to develop standards for their ATN, FLMTS and UPTS so that systems can interwork in the future. Foreseen multimode portables to be able to operate in low power and different frequency bands.

Furthermore work has to be carried out on MP nodes in order to develop system requirements for DB structure, management and (simulated) traffic load

10. Conclusions

The paper described a methodology to cover air traveller increasing need to remain connected while travel-ling long distances in high speeds. Developments in different areas (ie. Future WIN/AIN, ATN ISO/OSI network interworking , 3rd generation PCS, etc.) were conceptually brought together in a system context, points open to further development were identified and the work to be done was outlined. In addition issues concerning air interface compatibility and system interworking (such as PCS and avionics) were addressed.

The novel features of the proposed methodology are the use of ATN infrastructure and , capabilities together with the introduction of the Moving Cell and Migration Point concepts.

The proposed location and tracing methodology if implemented should allow location and tracing of very fast moving active subscribers while minimising addressing overheads and call set-up delays.

11. References

[1] CCIR, Draft Rev. of Question 74-1/8: *"Public Mobile telephone Service with Aircraft"*.

[2] ITU-R 73-3 *"Public Mobile telephone Service with Aircraft"*

[3] Papavramidis A.C., Tombras G.S., Doukas D.: "Adaptation of Land Mobile Systems for Onboard Operation" *Proc. of the 2nd IEEE International Conference on Universal Personal Communications*, Ottawa, Oct. 1993, p.p. 258-263.

[4] CCITT, *"Blue Book Rec. Q.1000 Series,* "Public land mobile network interworking with ISDN and PSTN," Nov. 1988.

[5] EIA/TIA/IS-41, "Cellular radio telecommunications intersystem operations: Functional overview", Jan. 1991

[6] G.R. Ash *et al.* "Robust Design and Planning of a World-wide Intelligent Network", *IEEE JSAC,* vol. 7, no 8, Oct. 1989.

[7] ICAO Doc 9578-AN/935 *"Manual of the Aeronautical Telecommunication Network"*.

[8] Wang J. Z. *"A Fully Distributed Location Registration Strategy for Universal Personal Communication Systems".* IEEE Journal on Selected Areas in Communications, Vol.11 No 6 August 1993.

[9] Papavramidis A.C.: *"Using ATN to Support Onboard Personal Communication Services"* NAVAIR, Official Journal of the International Federation of Air Traffic Safety Electronics Associations (IFATSEA), Vol. 16, Autumn 1993.

[10] ICAO/ANNEX 10, Amendment 67 "Mode S specifications and recommendations".

Fig. 8: ATN generic structure

The diagram contains the following labels:

To Other Domains

Aircraft Domain

ATC
WX
AAC
AOC

Aeronautical Comm Service Provider Domain

To Other Domains

Air Traffic Service Provider Domain

ATC
WX
AAC
AOC

WX ATC

WX

ATC

ATC
WX

To Other Domains

AAC

AOC

Aeronautical Operating Agency (e.g. Airline) Domain

AOC AAC

To Other Domains

LEGEND:
- ◎ Intra-Domain Router
- ◉ Inter-Domain Router
- ▭ Host Computer (End-System)
- ⟋ Inter-Domain Mobile Link
- ─── Intra-Domain Fixed Link
- ━━━ Inter-Domain Fixed Link
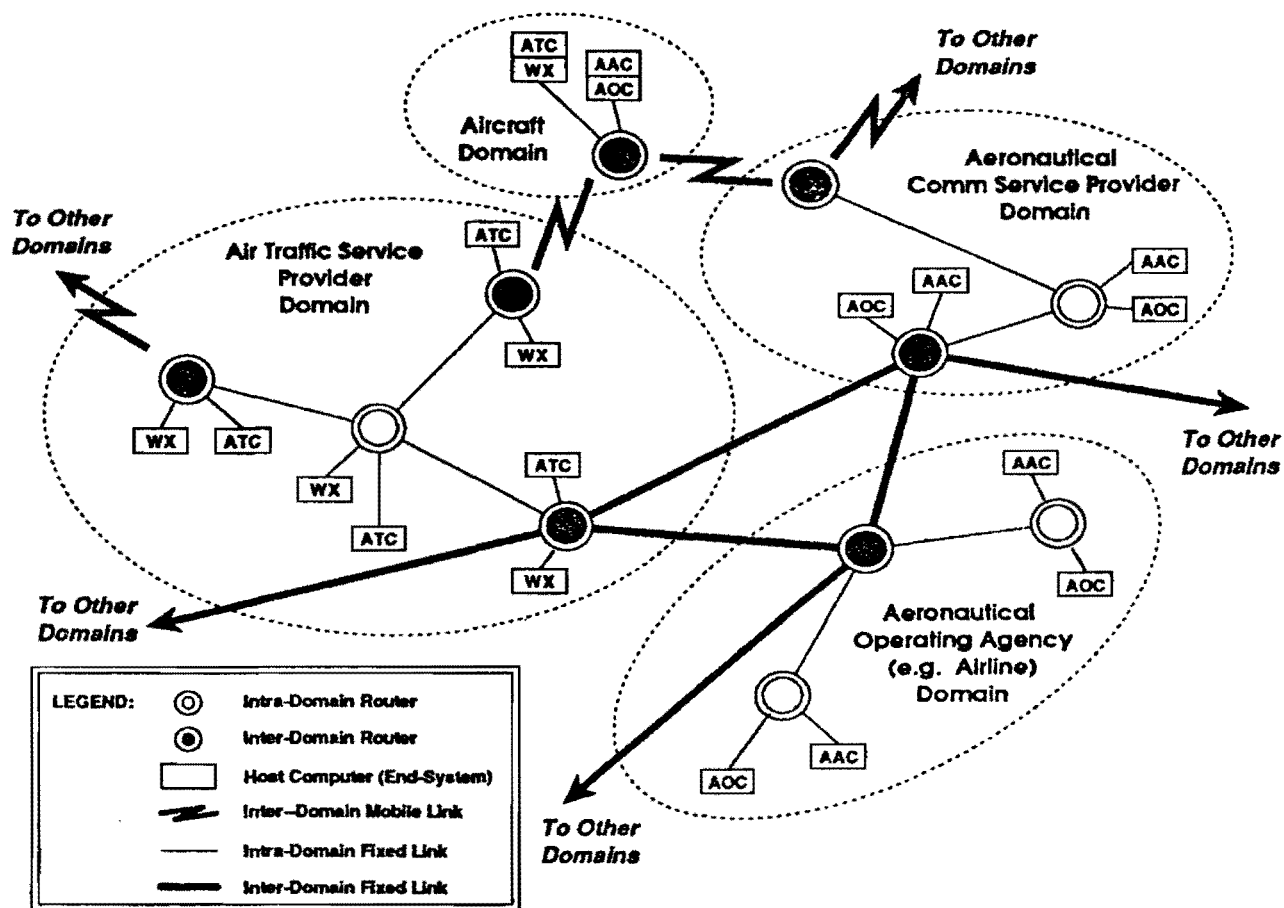
Asynchronous Transfer Mode Based Mobile Broadband System

Anand R. Prasad, *IEEE Student Member*
Telecommunications and Traffic-Control Systems Group
Department of Electrical Engineering
Delft University of Technology
P.O. Box 5031, 2600 GA Delft
The Netherlands
Fax: +31-15-781774
E-Mail: Anandr @ Octopus.ET.TuDelft.NL

Abstract- This paper proposes a wireless asynchronous transfer mode (ATM) based mobile broadband system (MBS). The system is proposed for integrated services mobile broadband network that provides an extension of fibre optics based ATM network capabilities. This means the proposed system must support reasonable number of service classes, bit rates and quality of service (QoS) levels associated with ATM.
Different modulation techniques, protocols, medium access control, error control and handover issues required are finally presented in the paper.

1. Introduction

Recently study of MBS, ATM and ATM oriented MBS has drawn the attention of several researchers [1]-[24].
With enormous complexity of managing and operating the many different types of networks that are now in use, the door is open for finding a common platform - a network on which all established services can be supported and which will allow new services to be introduced without needing new networks to run them on. The answer seems to be asynchronous transfer mode (ATM). This is the technology being defined and standardised for broadband integrated services digital network (B-ISDN). Thus, ATM, when adequately modified, is also an answer for the future mobile broadband systems.
The aim of this paper is to give a proposal for the use of mobile broadband system (MBS) with ATM. Different modulation techniques, protocols etc. necessary to be implemented for the use of ATM in MBS will be finally presented
The goal of MBS is to make B-ISDN services available for mobile users. Thus it has to be able to transmit most of the broadband services and conventional narrowband services over the air interface without a great loss of quality. The high data rate envisaged for MBS require operation at higher frequencies, currently taken to be the 60 GHZ band.

Asynchronous Transfer Mode (ATM) has to be adapted so as to reach this aim. The basic idea of ATM is the separation of services, their data and data rates. Thus an ATM-Net does not have to make available service specific channels. Because of great number of different services this leads to a flexible, service independent and uniform network structure.
ATM is a packet oriented transmission scheme. The transmission path of the packets of constant length, the so called ATM-Cells, will be established during connection set-up between the two end-points by assigning a virtual channel. At this time, the necessary resources are provided and the logical channels are assigned. All packets of a virtual channel are carried over the same path. The transmission capacity of virtual channel are characterised by the parameters mean bit rate and peak bit rate during connection set-up. ATM-Cells are generated according to the need of data source. Thus, ATM is a very good method to meet dynamic requirements of connections with variable data rates.

MBS will be the interface between the fixed ATM-Net at the base station side and mobile ATM-Net at the mobile station side. Normally, the ATM-Net at the mobile station side will only consist of one end system. For every end station it is possible to operate several virtual channels with different data rates at the same time.

2. Criteria for System Design

The main MBS concept is to provide mobile users, access to all range of services which is or will be available for fixed users of IBCN. Multimedia requirements being the driving force behind the development.
Thus the main features of the system level design can be summarised as below:

- Capability of working with a variety of services including voice, data and multimedia.

- Good quality of service (QoS) for various service types.
- Compatibility with ATM/BISDN (future broadband networks).
- Low terminal cost/complexity/power consumption.
- Efficient, scaleable and moderate cost network architecture.
- Compliance with regulatory constraints (in terms of operating bandwidth, power levels etc.)

3. MBS Transport Architecture

This section is focused on the issues related to achieving ATM-type broadband communications capabilities. A conceptual view of such a network is shown in Figure 1.
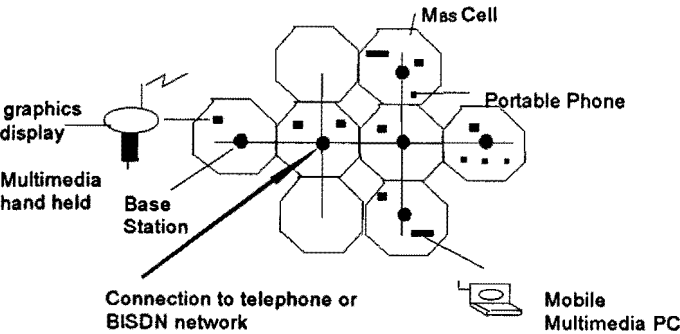

Figure 1: Conceptual view of MBS

3.1 ATM-Based Wireless Transport

Most important benefit of ATM is its flexibility, it can be used for the new high bit rate services which are either variable bit rate (VBR) or burst traffic. Several factors that tend to favour the use of ATM cell transport in MBS, can be given as follows:

- flexible bandwidth allocation and service type selection for a range of applications,
- efficient multiplexing of traffic from bursty data/multimedia sources,
- end-to-end provisioning of broadband services over wireline and wireless networks,
- suitability of available ATM switching for inter-cell switching,
- improved service reliability with packet switching techniques,
- ease of interfacing with wired BISDN systems.

Taking the above considerations, adoption of ATM compatible fixed length cell-relay format for MBS can be recommended. A possible ATM-compatible MBS approach is given in Figure 2.

With this approach the 48 byte ATM cell payload will become the basic unit of data within the MBS network. Within MBS, specific protocol layers (e.g., data-link and medium access control layer) will be added to the ATM payload as required, and replaced by ATM headers before entering the fixed network [18]. The use of ATM switching for intercell traffic also avoids the crucial problem of developing a new backbone network with sufficient throughput to support intercommunication among large numbers of small cells. ATM multiplexers may be used to combine traffic from several base stations into a single ATM port.
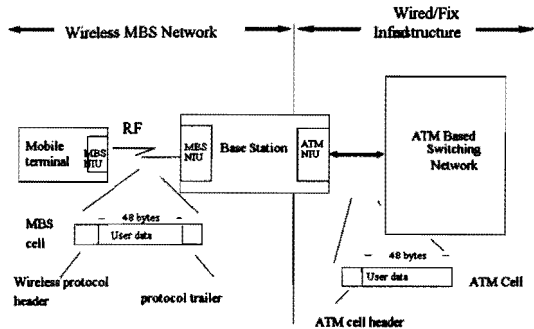
Figure 2: ATM-compatible MBS approach

3.2 Protocol Layering

For a seamless internetworking mechanism with the wired broadband network it is vital to have the MBS protocol layering harmonised with the ATM stack. In Figure 3 a protocol reference model is proposed. In this approach new wireless channel specific physical, medium access control and data-link layers are added below the ATM network layer. This means that regular network layer and control services such as call set-up, virtual channel identifier/virtual path identifier (VCI/VPI) addressing, cell prioritisation and flow-control indication will continue to be used by mobile services. While, the baseline ATM network and

Figure 3: Relation of wireless network protocol layers

144

signalling protocol will have to be specified to specific mobility related functions such as address registration (roaming), broadcasting, hand-off etc.
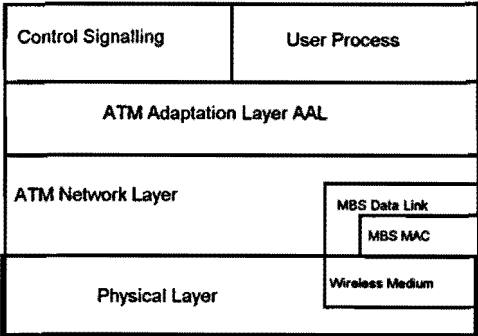
In the following sections, issues related to design of the wireless medium specific layers (i.e. physical layer, medium access control and data-link) are discussed.

3.3 Physical Layer

Right kind of modulation method and bit rate selection for macro, pico and micro cell is an important issue for the system design of MBS. The actual bit-rates to be used will depend upon the balance between service requirements and signal quality, bandwidth and power constraints.

After studying different techniques from the reference it is found that modulation technique based on coherent orthogonal frequency division multiplexing (COFDM) of multiple (N>>1) quadrature amplitude modulation (QAM) carriers (currently being used for digital audio broadcast in Europe and is also under consideration for digital high definition television (HDTV)), can achieve bit-rates up to 155.5 Mb/s. Which is necessary for broadband services, thus COFDM of multiple QAM carrier is proposed for MBS. This modulation technique is relatively impervious to interference. OFDM permits any subset of transmitters to compatibly operate at a peak bit-rate lower than the nominal channel speed.

3.4 Medium Access Control

To extend services integrating fixed networks to mobile users (MBS - B-ISDN) an adaptation of the MAC schemes specialised in transmission of constant data rates of speech services to the variable and high data rates of broadband services is necessary.
This section presents consideration on the MAC scheme for MBS which has to be adapted to the transmission virtual channels of ATM based Networks like B-ISDN.

The uplink of MBS air interface is the critical part of MAC scheme. As depicted in Figure 4 it consists of a system of mobile stations each having a different number of virtual channels (VC) as message sources and one base station as message sink.

The generation times of ATM-Cells of virtual channels can only be estimated by using stochastic processes characterised by mean bit rate and peak bit rate. As cells of different virtual channels of a mobile station have to be handled with different priorities due to

different data rates, there has to be a separate queue for each virtual channel as shown in Figure 4.

Thus, as the uplink structure is described, it is the task of the MAC scheme to assign transmission slots to virtual channels in such a way that the Quality of Service (QoS) parameter of the ATM-Net and the desired transmission rates of virtual channels are guaranteed. MAC scheme has to keep the delays at a minimum as ATM-Net will remove the cells with a delay greater than a specific limit. As well, it has to maximise throughput because of the limited channel capacity. Thus, it is necessary to find a optimum value between transmission delay and throughput in a MAC protocol so as to reach the QoS parameter as close as possible.
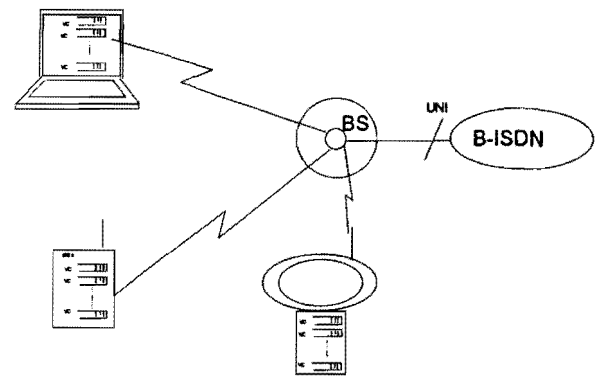

Figure 4: VC queuing in mobile stations

Due to variable cell generation rates the MAC scheme has to adjust the allocated transmission capacity to the varying short-term requirements.

Work done in [9] shows that a minimum transmission delay in combination with maximum throughput will be reached using Dynamic Slot Allocation (DSA) protocol. Sophisticated algorithms can be used to lower the delay furthermore. As concluded in [24] dynamic addressing reduces addressing overhead by a factor 8. This results in performance improvements ranging from 5% to 33%. Besides this dynamic addressing has two important features, real-time support and security. Considering these factors dynamic addressing is proposed for MBS system.

3.5 Data Link Layer

From Figure 3 it can be seen that wireless medium will generally require a custom data-link protocol that interfaces with ATM transport layer in relatively seamless manner. This layer is required to insulate the ATM-type service from wireless medium specific problems caused by high error rate at the physical level and/or delay/blocking at medium access control level.

At the same time, it is desirable to keep the additional MBS data-link functionality to a minimum in order to limit the amount of gateway processing required at the base stations.

From Figure 2 it can be seen that, the baseline approach for maximum ATM compatibility is to define MBS data-link packet as 53 byte ATM cell encapsulated by an additional MBS header. As per [17], so as to avoid high protocol overheads, it may be appropriate to compress essential ATM header information into an abbreviated network layer header. To this header then a wireless medium-specific data-link header is added. An example of MBS data-link packet format can be seen in Figure 5.

Functions as shown in Figure 5 are described in the following:

Service Type Definition

A suitable field in the MBS header may be provided to indicate whether a packet is of type supervisory/control, CBR, VBR, data, burst etc.
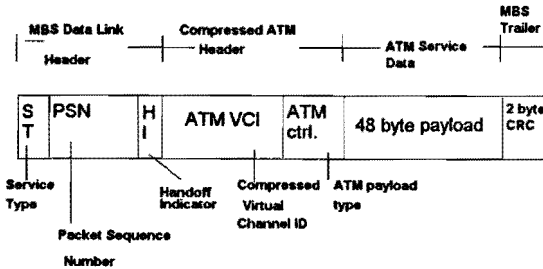

Figure 5: Example MBS data-link packet format.

Error Control

The ATM Cell Header is coded only by the cyclic error correction (CRC), used for single error correction and multiple error detection. In MBS there is a different scenario. If there is a line of sight path, the link quality will be affected by fast Raleigh fading processes. Obstacles like buildings, trucks or others will cause interruption of direct path. This may also occur in indoor environment due to moving people or other obstacles.

The following proposal is based on a scenario including one link between mobile station and base station. Furthermore it is assumed that fast handover processes are possible and variable delay is not critical.

The truncated ARQ type-II scheme is applicable to MBS, if fast handover process is possible and variable delay is not critical. If the above assumption is correct the truncated ARQ type-II with FEC as C_0 is proposed for MBS.

Hand-off/handover Support

It is very important for MBS to define a sophisticated handover scheme, because the severe propagation conditions at 60 GHz have a great effect on handover and system performance. Due to the fact that LOS path is required for a communication link, it will be mandatory to have a very fast handover decision scheme, which takes into account that total number of handover has to be as small as possible, which will be much higher as in other cellular radio networks, because due to propagation at 60 GHz the cell sizes are much smaller than the conventional networks.

In the following an outline of a handover scheme is presented.

Handover

A lot of reasons can request handover. The main reasons are classified as follows:

Radio Reasons: This can be either an unacceptable radio transmission quality like a high error rate or a low signal level.

Operation and Management (O&M) Reasons: This includes normal maintenance of equipment, equipment failure and channel rearrangement.

Traffic Reasons: If the traffic load is distributed uneven over the base station some mobiles at the border of cell can be allocated to an adjacent cell.

Handover is performed in three steps. First, measurements are made to get reliable information on the current situation in the scenario. Second, measurement value has to be processed in an appropriate manner to decide which base station is the most suitable. Finally, the execution of the handover, where the connection is switched to another base-station.

For MBS a Mobile Controlled Handover (MCHO) scheme is recommended in [3]. The main advantage of MCHO scheme is the low time elapsing for a handover.

The decision for handover are based on the measurement performed by the stations which are evaluated by the handover algorithm in an appropriate manner.

Handover Execution

In general two different handover procedure exist, the forward and backward handover. The backward handover performs signalling completely via the current base station and the mobile terminal changes its point of attachment to the fixed network after the connection of the new point has been established. The forward handover establishes a signalling link to the target base station before releasing the communication link with the current base station. This kind of handover is most suited to 60 GHz transmission, because if the signalling link fails for any reason, possibly an obstructed LOS a call could be established.

4. Conclusions and Recommendations

ATM compatible MBS approach has been presented. Asynchronous Transfer Mode can be used for bit rates needed for MBS.

For the use of ATM with MBS a new architecture of ATM cell has been described. Specific layers (e.g. data-link and media access control layers) should be added to the ATM payload.

After studies COFDM modulation technique has been proposed, DSA protocol has been chosen for MBS and MCHO scheme has been chosen for handover techniques.

At present a lot of work has to be done in the area of millimetric waves. For application of MBS the proposed ATM architecture can be used, but, yet it has to be tested. Further studies have to be done to find the appropriate kind of modulation technique, protocols and handover techniques etc. For millimetric (mm) waves, thus, technological developments remains to be done.

Biological effect of mm waves has yet to be studied.

References

[1] A. Gusmao and M. Prögler, "Physical channel design for mobile broadband system", RACE Mobile Telecommunications Workshop, pp. 134 - 136, Metz, 16-18 June 1993.

[2] R. Prasad and L. Vandendorpe, "Cost 231 Project: Performance evaluation of a millimetric-wave indoor wireless communications system", RACE Mobile Telecommunications Workshop, pp. 137 - 144, Metz, 16-18 June 1993.

[3] H. Hussman, "Investigation on MBS Handover Algorithm Based on Raytracing Modelling", RACE Mobile Telecommunications Workshop, pp. 145 - 148, Metz, 16-18 June 1993.

[4] J. Brazio, C. Belo, Stein Svaet, B. Langen, D. Plassmann, P. Roman, A. Cimmino, "MBS Scenario", RACE Mobile Telecommunications Workshop, pp. 194 - 197, Metz, 16-18 June 1993.

[5] Richard Carver, "MBS Services and Applications" RACE Mobile Telecommunications Workshop, pp. 227 - 231, Metz, 16-18 June 1993.

[6] Geir Løvnes, "Work on propagation aspects within MBS", RACE Mobile Telecommunications Workshop, pp. 394 - 396, Metz, 16-18 June 1993.

[7] Holger Hußmann, "Simulation of propagation using raytracing technique", RACE Mobile Telecommunications Workshop, pp. 397 - 400, Metz, 16-18 June 1993.

[8] D. Plassman, N. Vohn and P.Decker, "Suitable ARQ/FEC Methods for MBS", RACE Mobile Telecommunications Workshop, pp. 459 - 464, Metz, 16-18 June 1993.

[9] D. Petras, "Performance evaluation of Medium Access Control Scheme for MBS", RACE Mobile Telecommunications Workshop, vol. 1, pp. 52 - 57, Metz, 17-19, June 1994.

[10] W. Fischer, E. Wallmeier, T. Woster, S. Davis and A. Hayter, "Data Communications using ATM: Architectures, Protocols and Resource Management" IEEE Communications Magazine, vol. 32, no. 8, pp. 24-33, August 1994.

[11] M. Listani and R. Winkler, "Addressing and control issues for movability features in the B-ISDN connectionless service", Proceedings of the IFIP TC6 Workshop on Broadband Communications, pp. 23-64Estoril Portugal, 20-22 January 1992.

[12] R. Händel and M.N. Huber, ATM Networks Concepts, Protocols, Applications, Addison Wesley, June 1991.

[13] J. Walrand, Communication Networks: A First Course, Aksen Associates, 1991.

[14] K. Zauner, "On the ATM road to broadband", Telecom report international, No. 3, 1993, pp. 26-29.

[15] A. Tubtiang, H.I. Kwon and G. Pujolle, "A Simple ATM Switching Architecture for Broadband-ISDN and Its Performance", Modelling and Performance Evaluation of ATM Technology (C-15), IFIP 1993, pp. 361-367.

[16] W. Stalings, Advances in ISDN and Broadband ISDN IEEE Computer Society Press, 1992.

[17] I. M. Leslie, D. R. McAuley and D. L. Tennenhouse, "ATM Everywhere?" IEEE Network, pp. 40-46, March 1993.

[18] M.J. McTiffin, A.P. Hulbert, T.J. Ketseoglou, W. Heimsch and G. Crisp, "Mobile Access to an ATM Network Using a CDMA Air Interface", IEEE Journal on Selected Areas in Communications, Vol. 12, No. 5, pp. 900-908, June 1994.

[19] D. Raychaudhri, "ATM-Based Transport Architecture for Multiservices Wireless Personal Communication Networks", IEEE Journal on Selected Areas in Communications, Vol. 12, No. 8, pp. 1401-1414, October 1994.

[20] R. R. Geij, "Mobile Multimedia Scenario Using ATM and Microcellular Technologies", Vehicular Technology, Vol. 43, No. 3, pp. 699-703, August 1994, pp. 699-703.

[21] E. Skatun "Broadband integrated services digital network" Telektronikk, vol. 89, iss. 1, pp. 123-130, 1993.

[22] P.F.M. Smulders, A.G.Wagemans, "Wideband indoor radio propagation measurements at 58 GHz", Electron. lett., 1992, 28, pp. 1270-1271.

[23] P.F.M.Smulders, A.G.Wagemans, "Wideband measurements of mm-wave indoor radio channels", Proc. 3rd Int. Symp. on Personal, Indoor and Mobile Radio Commun., pp. 329-333, October 1992.

[24] M. Prögler, "MBS Air Interface", RACE Mobile Telecommunications Workshop, vol. 1, pp. 133 - 139, Metz, 17-19 June 1994.

[24] V. Bhargavan, "A Dynamic Addressing Scheme for Wireless Media Access", ICC'95, Seattle, pp. 756 - 760.

148

Shared Mobility with GSM

M. Engels, D. Peeters, W. Philibert, A. Van de Capelle

K.U.Leuven, Department of Electrical Engineering (ESAT),
Kard. Mercierlaan 94, 3001 Heverlee, Belgium.
Fax.: +32-16-32.19.86
E-mail: Marc.Engels@esat.kuleuven.ac.be

Abstract

The car traffic has reached its saturation point. Ridesharing offers an alternative to maintain or even increase our mobility, on condition that rides can be offered in a flexible way. In this paper a GSM-based service, Global Shared Mobility with GSM or GSM^2, is described that realizes the above-stated objective. In GSM^2 three subservices are distinguished: hitch-hiking, transport on demand, and home-to-work traffic. The description of these GSM^2 services and their components are further worked out in the paper. Implementation based on an intelligent network (IN) strategy is proposed. For the home-to-work subservice a simulator has been developed.

1. Introduction.

In the industrialized world, a continuous increase of mobility is experienced. Figure 1 illustrates this evolution for the past 20 years.

The higher mobility is reflected in an increased road traffic, as the public transportation has even decreased over the same period of time. The popularity of the car can also be concluded from figure 2, which gives the distribution of the various means of transport for home to work traffic. It shows that 63% of the

people uses the car, of which only 8% as a passenger. As a consequence, most cars are occupied by only one person [2].

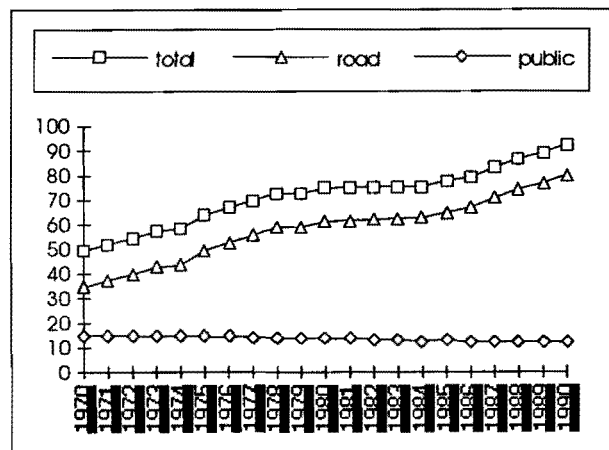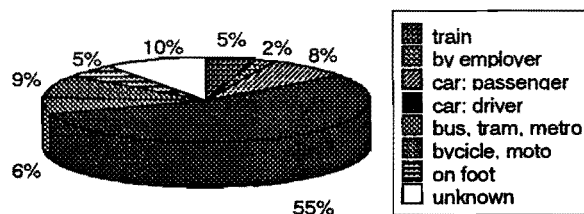Figure 1: Mobility in Belgium from 1970 to 1990 (from [1]).

Figure 2: transportation means for home-to-work traffic (from [3]).

This situation causes frequent traffic jams, which cost in Belgium 20 billion Belgian francs a year, according to the calculations of the "Federatie van Belgische Importeurs en Automobielconstructeurs (FEBIAC)"[4] .

To solve (or reduce) this problem, we can do three things: invest in public

transport, construct more roads, or promote ridesharing. Although the service of the public transport can be improved dramatically, the small market share limits the impact of these measures. Therefore an overall solution should mainly focus on the road traffic. The construction of extra roads is an expensive solution, against which a lot of resistance exists. This is certainly true in a country like Belgium, which already has a dense network of highways and almost no open space left.

As a consequence, we believe that the only viable solution is the more efficient utilization of the present transportation capacity by means of ridesharing. In this paper, we interpret ridesharing as all possible ways too share a car, including hitch-hiking and carpooling.

If we want to increase ridesharing we first must understand why it is not widely spread nowadays. The results of such a study are summarized in table 1.

not often possible	20.0 %
not flexible	15.0 %
does not know a partner	12.8 %
prefers to drive alone	10.5 %
not thought about it	8.3 %
too complicated	7.1 %
too many roundabouts	6.4 %
not reliable	3.6 %
no financial profit	2.6 %
too dangerous	2.1 %
extends the driving time	1.9 %
other reasons	9.7 %

Table 1: Primary reason for not using ridesharing (source [5]).

As can be seen, the majority of these reasons have to do with the way ridesharing is currently organized. If we would have a centralized ridesharing system, which is flexible and easy in use, about 60% of these obstacles could be solved. An additional 10% can be tackled by a good marketing strategy or an extra financial incentive.

In this paper, we present a centralized ridesharing system based on GSM. The remainder of this text is organized as follows. Section 2 introduces the GSM^2 service. Next, we propose an approach to implement this service. In section 4 we present the GSM^2 simulator. In the last section we discuss the future of GSM^2. Finally we draw some conclusions.

2. The GSM^2 Services.

In GSM^2 three subservices can be distinguished, depending on the wanted flexibility and the exploited transport capacity: hitch-hiking, transport on demand, and home-to-work traffic. Each of these subservices will be described more in detail in the following subsections.

2.1 Hitch-hiking with GSM^2

This first subservice adds the possibility of identification and accounting to the plain old hitch-hiking. The aforementioned identification increases the safety of passengers as well as drivers. Upon embarking user's identification is send to the GSM^2 exchange, where verification and accounting takes place.

Following components are involved in the system:

- A GSM phone.
- A passenger identification number.
- A passenger signalisation sign.
- A driver identification number.
- A driver signalisation sign.
- The GSM^2 exchange, an intelligent switch.

A passenger wishing to drive along makes his way to a place with sufficient potential transport capacity and makes his wish clear with his signalisation sign. When a driver who stops has a driver sign on his car, the passenger knows that he is registered with the GSM^2 service.

Before starting off, a data-message is send to the GSM^2 exchange via the GSM

phone. This message contains the driver- and the passenger identification and the kilometer reading. The GSM^2 exchange immediately confirms whether both parties are conform to the regulations of the service.

The advantage of such an identification is the higher security because one is sure that:

- the driver has a valid car-insurance and drivers license
- the driver is free of serious infringements to the traffic code or public law, and has not caused a serious accident.
- the passenger is not a criminal and has an account that guarantees the payment of the traffic costs.

At the destination, there is again a data communication where the driver identification and the kilometer reading are passed-on to the GSM^2 exchange. Accounting takes than place automatically.

To limit the manual actions, a further improvement could be the incorporation of a card reader in the system. The identification number can then be stored on personalized cards and need no longer be typed in via the GSM phone. Also an automatic reading of the kilometer counter would reduce the required actions. Unfortunately both options seem to be quite expensive and thus are currently considered unrealistic.

2.2 Transport on demand with GSM^2

The second GSM^2 service is a superset of the previous one. Indeed, besides identification and automatic accounting, the service brings together passengers and drivers. For this purpose the service works with a network of fixed stop-places (e.g. bus stops). It hence offers the facilities of a bus in a flexible way,

making use of the available transport capacity in private cars.

To operate this service, following equipment is needed:

- A GSM phone for the driver.
- A network of stop-places. Every stop-place is numbered and equipped with a phone.
- A road-map with the numbered stop-places.
- The GSM^2 exchange.

A potential passenger goes to a stop-place and sends the GSM^2 exchange a data message by aid of the stop-place's phone. This message contains the passenger's identification number and the number of the start and destination place.

The GSM^2 exchange thereupon broadcasts this request in the relevant cell (cells). The driver willing to grant the request for mobility notifies the GSM^2 exchange of his engagement by sending his identification number. If several drivers react, one will be selected by the GSM^2 exchange. The exchange sends the driver and eventually the passenger a confirmation.

The driver makes his way to the start-place, using the numbered road-map, and picks up the passenger. The actual begin and end of the ride are notified to the GSM^2 exchange by sending the passenger identification number and position of the kilometer counter. Accounting can hence take place automatically.

In the case of a lack of drivers, a taxi could be contacted to guarantee the passengers a transport certainty. As passengers are very sensitive to reliability this would be a great advantage of the service. However, it is only feasible if it will be needed exceptionally.

A useful addition to the transport on demand service would be offered by the

Global Positioning System (GPS). Indeed, potential drivers can then be addressed more selectively and vehicles can be guided to the stop-place. Because a fast breakthrough is expected for vehicle guidance systems based on GPS, this option might be feasible in a near future.

2.3 Home-to-work traffic with GSM²

The last GSM² subservice aims at utilizing the regular traffic flow from residences to working places and back more efficiently as transport capacity. Unlike the previous subservices, the GSM² exchange will actively match the passengers with the drivers, using a database which contains the mobility profile (the daily trajectory) of the subscribed drivers and passengers. Identification and automatic accounting is also possible with this subservice.

Following components are needed:

- A GSM phone for the driver.
- A phone (GSM or fixed) for the passenger.
- The GSM² exchange.
- The GSM² database with the mobility-profile of the users (passengers as well as drivers).

The GSM² database makes a distinction between users that want to be passenger, driver, or both. A further classification is based on the available data about their daily trajectory (whether or not start time, start coordinates and destination-coordinates are known). The data is collected with the subscription form to become member of the GSM² service. Subsequent changes can be made via a new form or directly via a PC with modem.

With his phone, the passenger passes on his identification number to the GSM²-exchange. If he uses a GSM, the identification can be based on his smart card. If the start coordinates and stop coordinates of the requested ride are not available in the database the exchange asks for the missing data. The mechanism to pass the coordinates can be based on a fixed set of stop places, the location of fixed telephones, or the zip-codes.

The GSM² exchange interprets the incoming call and start a matching procedure. It returns a list of potential drivers, whose trajectory and start-time correspond with the asked ride. The GSM² exchange sends a data message, that contains the start and destination coordinates to the first driver on the list, who gets the GSM² message on his display.

If the driver accepts the call, he is put through to the passenger so that both parties can arrange the details of the ride, for instance the exact pick-up place. At the end of the conversation the GSM² exchange is notified that an agreement is reached and the identification number of the driver is registered.

The match can also be unsuccessful: the driver can reject the call, for instance because he does not follow his standard trajectory, or ignore it. It is also possible that no agreement is reached during the conversation. In all these cases, the GSM² exchange calls the next driver on the list.

If after several efforts no driver has been found, the GSM² exchange broadcasts a paging message to all GSM users within a certain cell. The further operation is then very similar to the previous subservice, except that the passenger and the driver have a conversation to arrange the details.

Also similar extensions, i.e. taxi company and GPS, can be made to the service. In addition, GPS would allow to widen the set of possible drivers with ones that have neither a fixed departure time nor a fixed trajectory.

Concluding, the last two subservices search at every instant of time the best combination of offered transport capacity and potential users. They differ in the a priori knowledge and the offered flexibility.

3. The GSM^2 Exchange.

As can be seen in the previous section, the GSM^2 exchange plays a crucial role in the service. It interrogates the passenger about his requested ride, it performs a matching procedure in close cooperation with the GSM^2 database and it sets up a connection between the passenger and the driver.

To realize such complex functionality in the telephone network the intelligent network (IN) concept seems to be the most appropriate. In IN [6], the control functions are separated from the switching functions and brought together in an Service Control Point (SCP). If specialized resources are needed, for instance to interact with the user, they are located in a separate device, the Intelligent Peripheral (IP). The detection of IN-calls and the controlled switching takes place in the Service Switching Point (SSP).

Figure 3 shows how GSM^2 can be realized as a IN service. The SSP detects that the user calls the GSM^2 service. It then gives control to the SCP via the signaling network (dashed lines). The SCP instructs the IP, which has text-to-speech capabilities, to ask the appropriate questions and to return the answers. Consequently the SCP terminates the call and starts a matching procedure in the GSM^2 database. Because this functionality requires complicated actions, the matching will be performed on a separate computer that is connected to the SCP. When a match is found the

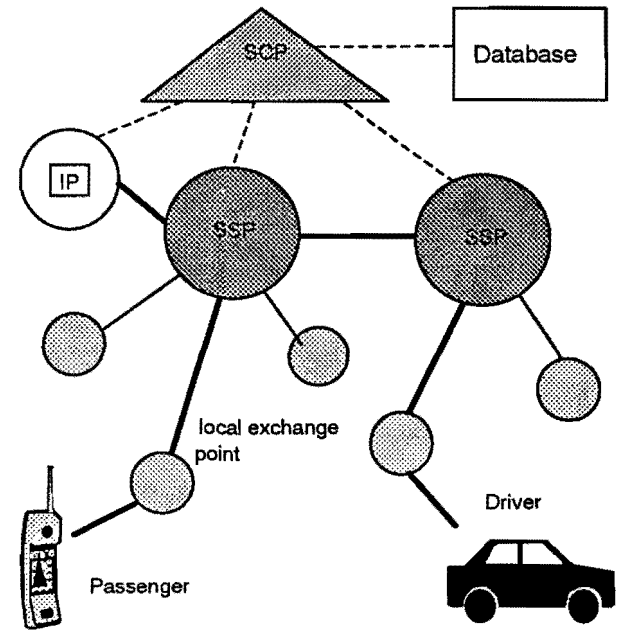SCP sets-up the connection between the selected driver and the passenger.

Figure 3: GSM^2 as a IN-service.

4. The GSM^2 Simulator.

To prove the feasibility of the proposed GSM^2 service we developed a simulator, of which a screen dump can be found in figure 4. The simulator was developed in C++ with a portable graphics library XVT. As a consequence, it can be run both on a PC as well as on a Sun workstation.

The GSM^2 simulator focuses on the functionality of the service from a user's point of view. It can hence be used to get feedback from a wide audience of potential customers. Especially, the mechanisms to communicate start- and destination coordinates to the GSM^2 service must be further evaluated.

Also the IN concept is already introduced in the GSM^2 simulator. Indeed, the IP and SCP entities are present and perform their respective functions. However, the GSM^2 database is still situated in the SCP. In the next step this database will be transferred to a separate entity.
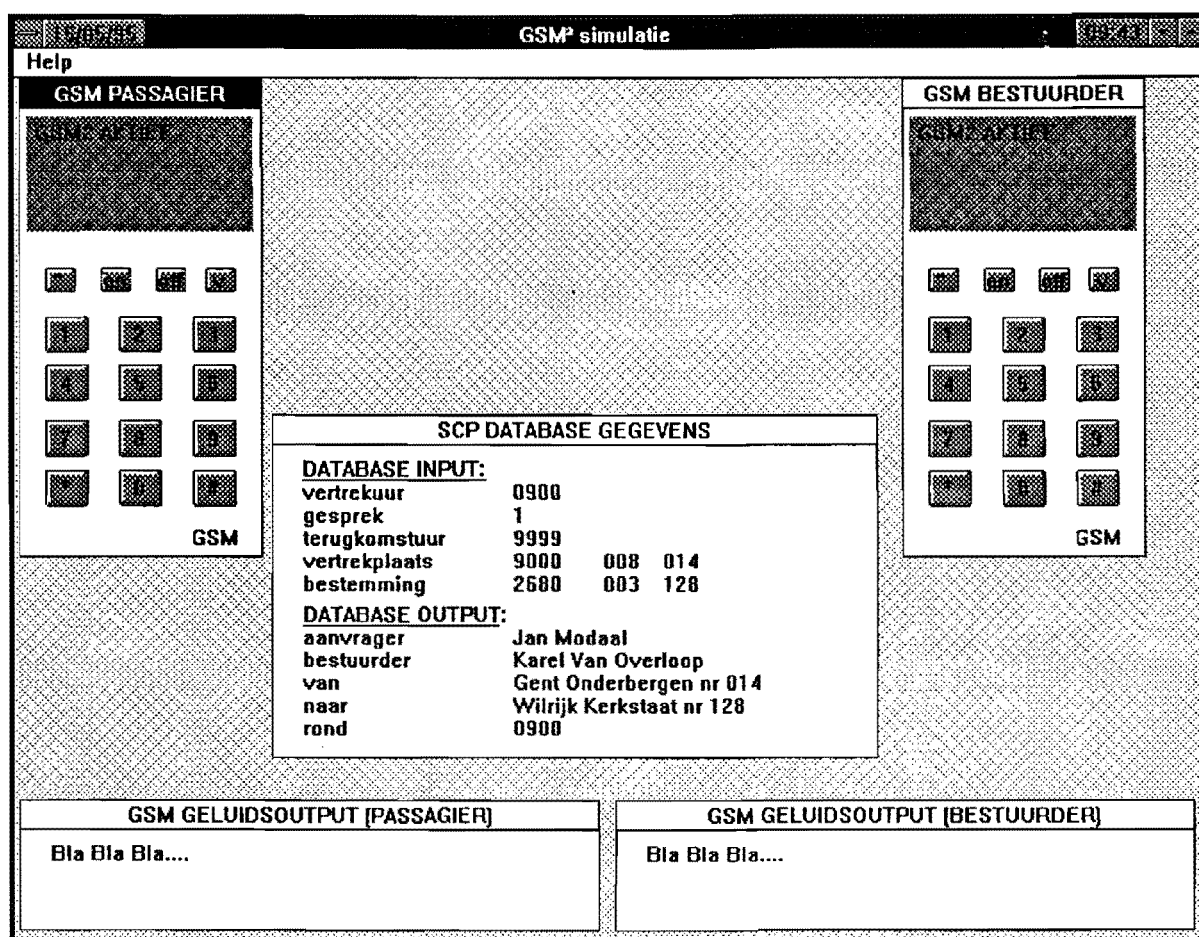
Figure 4: Screen dump of the GSM² simulator.

5. The future of GSM².

In the near future, we will further investigate the technical feasibility of GSM². In the simulator we will move the database to a separate entity and incorporate the feedback on the service functionality. Later on, we plan to make a prototype of the system on a real SCP.

Other aspects, however, are also crucial for a successful introduction of GSM². First of all, the GSM² system supposes already from the start a sufficient number of members. As a consequence, a wide proliferation of GSM (higher than 10%) and a marketing approach on a massive scale are prerequisites for success. Other promoting factors are a high fuel cost and persistent traffic-jams. Also government initiatives that reward shared mobility, e.g. by car-pool lanes or financial incentives, will be advantages for the GSM² service.

Finally the main question remains: who is going to exploit the GSM² service, the government or a private company? In the latter case rendability is an important issue. To this end, the service can generate income by means of membership fees, overhead on the price per kilometer which is transferred from passenger to driver, or charges on the telephone costs to the GSM² exchange (cfr. service lines).

Socio-economic studies have to be performed to evaluate the market potential and economic rendability of the GSM² service. Given the above constraints it should not be expected that the GSM² service is operational before 2000.

6. Conclusions.

In this paper we discussed a new GSM-based service, that organizes rideshiring in a more flexible and efficient way. For

the implementation of this GSM^2 service, we proposed an implementation based on an intelligent network (IN) strategy. We also presented our functional simulator of the service. Further work will concentrate on the technical feasibility and socio-economic factors.

Acknowledgments

References

[1] N.C.B. Economisch Departement, Verkeer in het jaar 2000, mobiliteit of chaos.

[2] P. Darge, "Dossier files", Knack magazine, january 1990, pp. 52-58.

[3] Ministerie van Economische Zaken, Nationaal Instituut voor de Statistiek, "Voorlopige resultaten van de volkstelling van 1990: vervoersstatistieken", inlichtingen- en documentatiedienst, Rubenslei 2, 2018 Antwerpen.

[4] Ministerie van de Vlaamse Gemeenschap, Departement Leefmilieu en Infrastructuur, Handboek: bedrijfsvervoerplan.

[5] Vansevenant P., "Enquete naar het gebruik van vervoermiddelen bij bedrijven en instellingen in Antwerpen in het kader van DRIVE V1069 Carpooling System Management".

[6] J. J. Garrahan, P. A. Russo, K. Kitami, R. Kung, "Intelligent Network Overview", IEEE Communications Magazine, Vol. 31, No. 3, March 1993, pp. 30-37.

DIGITAL LOCK DETECTORS ASSOCIATED WITH THE GARDNER TIMING ERROR DETECTOR

G. Karam [1], V. Paxal [1], and M. Moeneclaey [2]

[1] SAT, Telecommunications Division
75626 Paris Cedex 13, France
[2] University of Ghent
B-9000 Ghent, Belgium

ABSTRACT

Two lock detectors for clock recovery circuit are presented in this paper. Both of them are associated with the Gardner timing error detector (GA-TED) which requires two samples per symbol period, T. The first one (Detector A) requires two additional samples and hence, an oversampling frequency of 4/T is needed. The second (Detector B) is more attractive in practice. It is based on the same signal samples required for the GA-TED and thus an oversampling frequency of 2/T is sufficient. The paper describes the basic principle of these lock detectors and concentrate on the analysis of their performances. We show that the performance of Detector A is completely independent of channel distortion. And, Detector B becomes sensitive to channel distortion when the hermetian symmetry of the channel transfer function is destroyed. However, based on a computer simulation study it is shown that performance degradation remains negligible even for severe channel distortion. In addition, we discuss practical implementation of these lock detectors and analyze its consequence on the system performance.

I - INTRODUCTION

Usually, in digital communication systems the symbol rate is accurately known and the clock recovery circuit implemented in the receiver does not require any acquisition-aiding technique. However, in some applications [1, 2] the symbol rate can be variable in a large range and the receiver should make use of such techniques to be able to recover the symbol rate used by the transmitter. Thus, the clock recovery circuit takes the form of a phase lock loop (PLL) using an acquisition-aiding technique, e.g., sweeping, which only operates during the out-of-lock mode of the loop. Consequently, a reliable lock detector is needed to indicate whether the PLL is in the in-lock or in the out-of-lock mode and to switch off or on, respectively, the acquisition-aiding technique.

One can imagine a lock detector for clock synchronizer based on the analysis of the received signal constellation. Assuming a noiseless ideal channel, it is known that the received signal, sampled at symbol rate, is always different from zero if the clock synchronizer is locked. Thus, by defining a window around the signal constellation origin, a lock detector can be obtained by counting the number of times the received signal point is within the window, and compare the obtained result to a predefined threshold. When the threshold is exceeded the clock synchronizer is considered unlocked, otherwise it is locked. In practice, the implementation of this detector is very simple, but its performance is very sensitive to signal-to-noise ratio (SNR) and channel distortion.

In this paper we investigate two lock detectors which exploit the correlation property of the received signal. Both detectors are associated with the Gardner timing error detector (GA-TED) and are quasi-insensitive to SNR and channel distortion. The GA-TED has been introduced in [3]. It requires two samples per symbol period and its S-curve has a sinusoidal shape. The lock detectors presented in this study lead to S-curve whose shape is cosinusoidal in comparison to that of the GA-TED. Thus, when the GA-TED is operating at its stable equilibrium point, i.e., the clock synchronizer is locked, the lock detector produces its maximum average output. Conversely, when the clock synchronizer is unlocked, the timing error is an increasing (or decreasing) function of time and the lock detector average output is zero. Hence, the in-lock and out-of-lock modes of the clock synchronizer can be easily distinguished by measuring the lock detector average output.

The first lock detector we present requires two samples per symbol period which are located halfway between the samples needed by the GA-TED and thus, four samples per symbol period are needed in total. We will show that its performance is completely independent of the shape of the received baseband pulse. In other words, the optimal sampling instant given by the GA-TED is also optimal to the lock detector and leads at its output, to the maximum average signal. Hence, this lock detector is "matched" to the GA-TED.

The second lock detector requires the same two samples per symbol, as needed by the GA-TED. It is not matched to the GA-TED, i.e., its performance depends on the channel distortion and the optimal instant given by the GA-TED does not correspond to the maximum of the lock detector S-curve. However, by means of analysis and computer simulations we will show that its performance degradation is negligibly small even for severe channel distortion, which makes it very attractive in practice.

The paper is organized as follows. The basic principle of both lock detectors is described in the next section. In section III, we derive some of their characteristics. Practical implementation issues are discussed in section IV, and finally section V is dedicated for conclusions.

II - BASIC PRINCIPLE

Denoting by x(t) the received baseband complex signal, we can write:

$$x(t) = \sum_i a_i \, g(t - iT + \tau) + n(t), \qquad (1)$$

where g(t) represents the channel impulse response including the pulse shaping of the transmit and receive filters, the a_k's the data symbols, T the symbol period, τ the phase of the symbol clock, and n(t) the equivalent baseband noise.

The GA-TED implementation requires two samples per T and its output signal at the decision instants $kT + \hat{\tau}$, where $\hat{\tau}$ represents the estimated phase of the symbol clock delivered by the clock synchronizer, can be written as:

$$\varepsilon_k = \text{Re}\left[x_k \left(x^*_{k+\frac{1}{2}} - x^*_{k-\frac{1}{2}} \right) \right] \qquad (2)$$

where Re(.) is the real part and $x_\ell = x(\ell T + \hat{\tau})$.

In what follows, we present two lock detectors which are associated with the GA-TED. The input-output relationship of Detector A is given by:

$$u_k = \text{Re}\left[x_{k+\frac{3}{4}} \left(x^*_{k+\frac{5}{4}} - x^*_{k+\frac{1}{4}} \right) \right] \qquad (3)$$

The analysis of (2) and (3) shows that Detector A requires two samples per symbol that are located halfway between the samples needed by the GA-TED. Hence, four samples per T are needed in total and the receiver implementation requires an oversampling frequency of 4/T.

The second lock detector (Detector B) requires the same two samples per T, as needed by the GA-TED. Thus, an oversampling factor of 2, in comparison to the symbol rate, is sufficient to implement the system. The output of detector B can be written as:

$$v_k = \text{Re}(x_k x^*_k - x_{k-1/2} x^*_{k-1/2}). \qquad (4)$$

The principle of these lock detectors is based on the fact that the derivation of their S-curves, leads to a cosine waveform vs. timing error $\tau_e = \tau - \hat{\tau}$, whereas the S-curve of the GA-TED has a sinusoidal shape. The lock detector system operates in parallel to the clock synchronizer which takes the form of a PLL incorporating the GA-TED. When the PLL is locked, the average of the GA-TED output is zero, and the lock detector average output is at its maximum. Whereas, when the PLL operates in the out-of-lock mode, τ_e varies continuously with time and the average of the lock detector output can be obtained by integrating its S-curve which leads to zero. Consequently, the block diagram of the lock detector system can be sketched as in Fig. 1. The output signal of the lock detector is first averaged. Next, a decision device compares the averaged signal to a positive threshold to discriminate the in-lock and out-of-lock modes of the PLL.
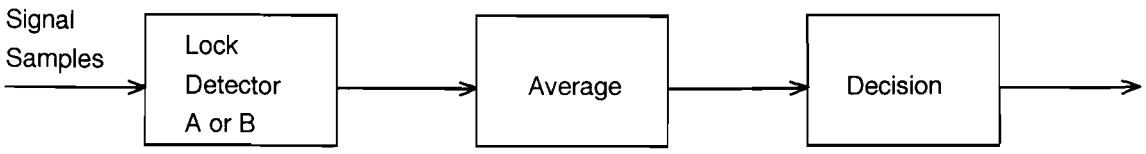
157

Fig. 1: Block diagram of the lock detector system.

III - DETECTOR CHARACTERISTICS

A. S-curves derivation

To analyze the performance of the lock detectors presented, we derive their S-curves and compare them to that of the GA-TED. The S-curve is obtained by computing the average of the detector output signal.

Considering detector A, its S-curve is given by $E(u_k)$, where $E(.)$ represents the expectation function. Using (3) we can write:

$$E(u_k) = Re\left[E\left(x_{k+\frac{3}{4}} x^*_{k+\frac{5}{4}} \right) - E\left(x_{k+\frac{3}{4}} x^*_{k+\frac{1}{4}} \right) \right] \quad (5)$$

Assuming uncorrelated data symbols with a variance σ_a^2 and a white noise, applying (1) into (5) leads to

$$E(u_k) = \sigma_a^2 Re\left[\sum_i g\left(\tau_e - \frac{3}{4} T - iT \right) g^*\left(\tau_e - \frac{5}{4} T - iT \right) \right.$$
$$\left. - \sum_j g\left(\tau_e - \frac{3}{4} T - jT \right) g^*\left(\tau_e - \frac{1}{4} T - jT \right) \right] \quad (6)$$

where it can be noted that $E(u_k)$ is independent of noise.

Using Poisson's formula and assuming that $G(f)$, which is the Fourier transform of g(t), vanishes outside the bandwidth $[-1/T, 1/T]$ we can rewrite after a long derivation (avoided here for brevity):

$$E(u_k) = 4 |\gamma| \cos\left(2\pi \frac{\tau_e}{T} + \theta \right), \quad (7)$$

where $|\gamma|$ and θ are the magnitude and phase, respectively, of a complex term γ given by :

$$\gamma = |\gamma| e^{j\theta}$$
$$= \frac{\sigma_a^2}{T} \int_{-\infty}^{+\infty} G^*(-f) \, G\left(\frac{1}{T} - f \right) \sin\pi fT . df \quad (8)$$

Similarly, the derivation of the GA-TED S-curve leads to:

$$E(\varepsilon_k) = 4|\gamma| \sin\left(2\pi \frac{\tau_e}{T} + \theta \right). \quad (9)$$

Thus, comparison of (7) and (9) shows that the lock Detector A has an S-curve identical to the S-curve of the GA-TED shifted by a phase $\pi/2$. Therefore, when the S-curve of the GA-TED crosses zero, the S-curve of the lock Detector A reaches its maximum. In other words, when the clock synchronizer is locked, the GA-TED operates at its stable lock point which corresponds to $E(\varepsilon_k) = 0$, given by:

$$\tau_e^o = -\frac{T}{2\pi} \theta. \quad (10)$$

Hence, the average of the lock detector output is:

$$E(u_k) \big|_{\tau_e^o} = 4 |\gamma|, \quad (11)$$

which corresponds to the maximum value. Note that this property is completely independent of the shape of g(t). Hence, Detector A is matched to the GA-TED.

When the clock synchronizer operates in the out-of-lock mode, τ_e varies with time and the average of the lock detector output is obtained by averaging (7) versus τ_e which leads to zero. Consequently, the two modes: in-lock and out-of-lock of the synchronizer can be easily discriminated by comparing the average detector output of the lock detector to a threshold between zero and $4|\gamma|$.

The S-curve of Detector B can be obtained by averaging v_k given by (4)

$$E(v_k) = Re\left[E\left(x_k x^*_k \right) - E\left(x_{k-\frac{1}{2}} x^*_{k-\frac{1}{2}} \right) \right] \quad (12)$$

Using (1) and the correlation property of data and noise lead to:

$$E(v_k) = \sigma_a^2 Re\left[\sum_i g\left(\tau_e - iT\right) g^*\left(\tau_e - iT\right)\right.$$
$$\left. - \sum_j g\left(\tau_e - \frac{T}{2} - jT\right) g^*\left(\tau_e - \frac{T}{2} - jT\right)\right] \quad (13)$$

Finally, we obtain:

$$E(v_k) = 4|\rho|\cos(2\pi\frac{\tau_e}{T} + \psi), \quad (14)$$

where $|\rho|$ and θ are the magnitude and phase, respectively, of the complex constant

$$\rho = |\rho|e^{j\psi} = \frac{\sigma_a^2}{T}\int_{-\infty}^{+\infty} G^*(-f) G\left(\frac{1}{T} - f\right) df \quad (15)$$

$E(v_k)$ has its maximum at the timing phase:

$$\tau_e^1 = -\frac{T}{2\pi}\psi \quad (16)$$

The analysis of (8), (10), (15) and (16) shows that unless $G(f)$ has hermetian symmetry, Detector B produces its maximum average value at a timing phase which does not correspond to the stable equilibrium point of the GA-TED. Hence, Detector B is not matched to the GA-TED when the channel distortion distroys the hermetian symmetry of $G(f)$. However, if we assume that $\tau_e^0 - \tau_e^1$ is small, the signal loss (in comparison to the maximum value) at the lock detector output for $\tau_e = \tau_e^0$ (stable lock point) becomes negligible. This point is investigated in the next subsection.

B. Effect of channel distortion

In the previous subsection, we have shown that Detector A is matched to the GA-TED and its performance is completely independent of channel distortion. Here, we investigate the effect of channel distortion on Detector B which is not matched to the GA-TED when the hermetian symmetry of $G(f)$ is distroyed. For this purpose, we consider a channel introducing one echo in addition to the direct path. Its impulse response is given by:

$$g(t) = h(t) + \beta\, h(t - \Delta) \quad (17)$$

where β is a complex coefficient, Δ a delay, and $h(t)$ the impulse response of a raised cosine Nyquist filter with a roll-off factor α.

We fix $\alpha = 0.5$ and $\Delta = 0.4\,T$ and we compute $\tau_e^0 - \tau_e^1$ for different values of β. It appears that in all cases, $\tau_e^0 - \tau_e^1 \cong 0$ and the signal loss at the output of Detector B is negligible. For example, with $|\beta| = -5\,dB$ and $Arg(\beta) = 0.62832\,rd$ we obtain:

$$\tau_e^0 = -0.1259\ xT$$

and

$$\tau_e^1 = -0.1258\ xT$$

which lead to $\tau_e^0 - \tau_e^1 = 10^{-4} x\ T$. Considering a QPSK modulation, Fig. 2 shows the corresponding S-curves of the GA-TED and the lock detectors A and B. It can be seen that the S-curve of Detector B is very close to that of Detector A which matches the S-curve of the GA-TED. Note also, that the maximum amplitude of the S-curve of Detector B is greater than that of Detector A.
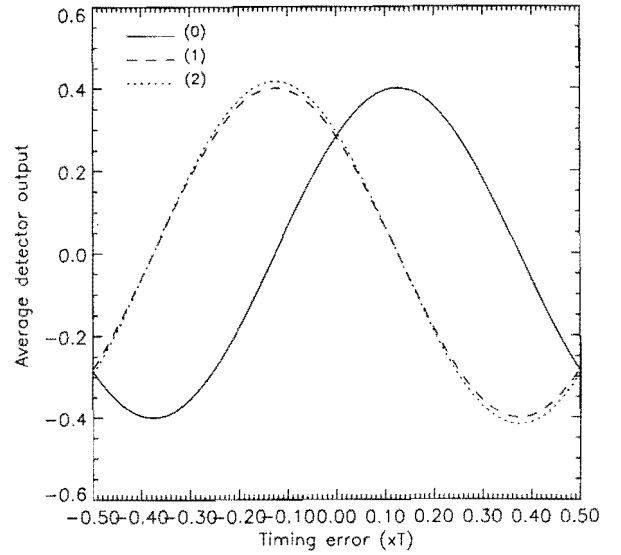

Fig. 2: S-curves with |β| = - 5 dB
  (0): GA-TED
  (1): Lock Detector A
  (2): Lock Detector B.

In conclusion, Detector B is to be preferred over Detector A in practice, because of its similar performance at a much lower complexity (two samples per symbol versus four samples per symbol).

159

IV - PRACTICAL IMPLEMENTATION

In the previous sections, we assume an ideal average operator $E(.)$ at the lock detector output. But in practice, this operator is approximated using an estimator given by:

$$Z_k = \frac{1}{N} \sum_{i=(k-1)N}^{k.N} z_i \qquad (18)$$

where z_i represents the lock detector output (u_i for Detector A and v_i for Detector B), N the number of z_i values used for one estimation of the average value and k a positive integer. Note that the estimator gives an output value each N x T period.

The reliability of this estimator depends strongly upon the value chosen for N. Furthermore, it depends on noise variance (or SNR) and channel distortion. Thus, the performance of the lock detectors A and B becomes sensitive to noise and channel distortion.

In addition, when the clock synchronizer is unlocked, the lock detectors performance depends on the value of the frequency offset Δf between the symbol clocks used at the transmitter and receiver. For a given N value, large Δf corresponds to a rapid evolution of τ_e with time. Hence, the integration vs. τ_e of the lock detector S-curve is well approximated using (18) and $Z_k \approx 0$ for any k. Inversely, small Δf values lead to a slow evolution of τ_e with time. Thus, only a small part of the S-curve is integrated and Z_k becomes variable with time and different from zero.

In the rest of this section, we concentrate on the lock detector B and investigate its performance by means of computer simulations to analyze the effects of N, Δf, SNR and channel distortion when using (18) to approximate the operator $E(.)$.

The effect of Δf is first studied by considering an ideal noiseless channel with a QPSK modulation. In Fig. 3 the computed Z_k values are shown vs. time. When the clock synchronizer is locked, $\tau_e = 0$ and Z_k corresponds to the maximum value of the lock detector S-curve which is theoretically $E(u_k) = 0.5$. We remark that:

$$Z_k = E(u_k) + B_k \qquad (19)$$

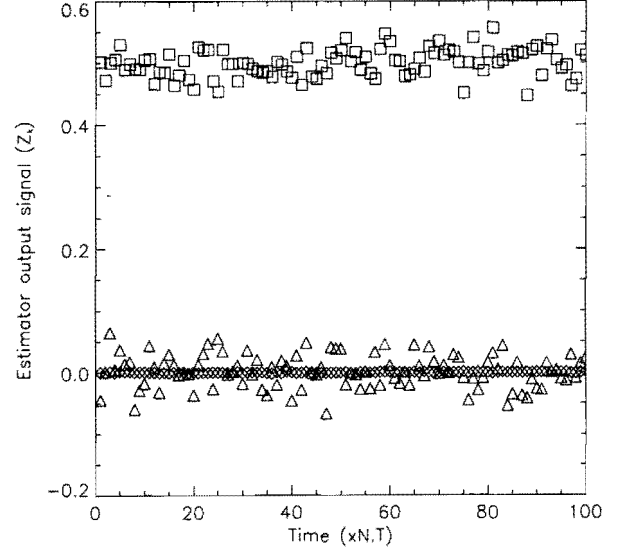where B_k is a noise caused by the estimation given in (18), with a variance σ^2.

Fig. 3: Estimator output vs. time with N = 1000:
Diamond points: unlocked with $\Delta f = \Delta f_1$
Triangular points: unlocked with $\Delta f = \Delta f_2$
Square points: locked.

Considering the synchronizer in the out-of-lock mode, two frequency offsets have been used in our simulations $\Delta f_1 = 5.7 / T$ and $\Delta f_2 = 8.10^{-3} / T$. The results in Fig. 3 show that in this case $E(u_k) = 0$, and (19) still valid. Note that when the synchronizer is unlocked, σ^2 increases when Δf decreases. However, in this example there is no difficulty to discriminate the in-lock and the out-of-lock modes of the synchronizer. To highlight this point, we consider N = 100 in Fig. 4. It can be seen that the estimator noise variance increases tremendously when the synchronizer is unlocked for $\Delta f = \Delta f_2$. Thus, it is no longer possible to discriminate the two modes of the synchronizer when Δf is small.

In order to evaluate the performance of the lock detector we compute the estimator noise variance σ^2 when the synchronizer is locked and unlocked with $\Delta f = \Delta f_1$ (large frequency offset) and $\Delta f = \Delta f_2$ (small frequency offset). Table I gives the result for N = 100, 500 and 1000 considering an ideal noiseless Nyquist channel with a roll-off factor $\alpha = 0.5$ and a QPSK modulation. In addition, we compute a normalized ratio σ / Δ_{max}, where Δ_{max} represents the maximum amplitude of the S-curve. Also, Δ_{max} represents the ideal distance between

160

Z_k in the in-lock mode and out-of-lock mode of the synchronizer. Table I shows clearly that decreasing N increases σ^2 if the synchronizer is unlocked with small frequency offsets
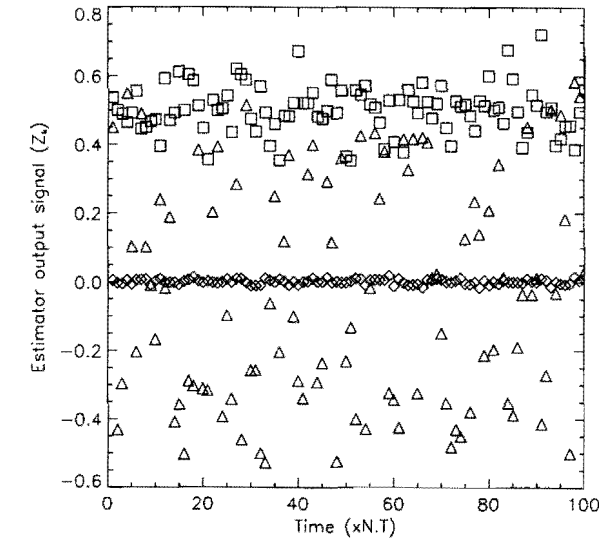

Fig. 4: Same as Fig. 3 with N = 100.

N	Synchronizer mode	σ^2	σ^2/Δ^2_{max}
1000	Locked	4.7×10^{-4}	1.9×10^{-3}
	Unlocked, Δf_1	3.5×10^{-7}	1.4×10^{-6}
	Unlocked, Δf_2	2.5×10^{-4}	1.0×10^{-3}
500	Locked	9.8×10^{-4}	3.9×10^{-3}
	Unlocked, Δf_1	2.5×10^{-6}	1.0×10^{-5}
	Unlocked, Δf_2	5.3×10^{-4}	2.1×10^{-3}
100	Locked	5.6×10^{-3}	2.2×10^{-2}
	Unlocked, Δf_1	4.8×10^{-5}	1.9×10^{-4}
	Unlocked, Δf_2	0.11	0.44

Table I: Performance of Detector B in the presence of a Nyquist noiseless channel.

Table II gives the effect of noise. In this table the same channel is considered but the SNR = 7 dB. In Table III, in addition to the noise, we consider the effect of channel distortion and take g(t) as given in (17) and use the same parameters considered in subsection III-B.

N	Synchronizer mode	σ^2	σ^2/Δ^2_{max}
1000	Locked	9.7×10^{-4}	3.9×10^{-3}
	Unlocked, Δf_1	7.6×10^{-5}	3.0×10^{-4}
	Unlocked, Δf_2	7.3×10^{-4}	2.9×10^{-3}
500	Locked	2.1×10^{-3}	8.4×10^{-3}
	Unlocked, Δf_1	1.8×10^{-4}	7.2×10^{-4}
	Unlocked, Δf_2	1.5×10^{-3}	6.0×10^{-3}
100	Locked	1.0×10^{-2}	4.0×10^{-2}
	Unlocked, Δf_1	1.0×10^{-3}	4.0×10^{-3}
	Unlocked, Δf_2	0.12	0.48

Table II: Performance of Detector B with an ideal Nyquist channel and SNR = 7 dB.

N	Synchronizer mode	σ^2	σ^2/Δ^2_{max}
1000	Locked	7.7×10^{-4}	4.4×10^{-3}
	Unlocked, Δf_1	8.0×10^{-5}	4.6×10^{-4}
	Unlocked, Δf_2	5.8×10^{-4}	3.3×10^{-3}
500	Locked	1.4×10^{-3}	8.1×10^{-3}
	Unlocked, Δf_1	1.7×10^{-4}	9.8×10^{-4}
	Unlocked, Δf_2	1.1×10^{-3}	6.3×10^{-3}
100	Locked	7.0×10^{-3}	4.0×10^{-2}
	Unlocked, Δf_1	9.8×10^{-4}	5.6×10^{-3}
	Unlocked, Δf_2	6.4×10^{-2}	0.37

Table III: Performance of Detector B with channel distortion and SNR = 7 dB.

The results show that noise and channel distortion do not seem to have significant influence on the lock detector performance. However, a particular care must be taken when dimensioning the value of N, to avoid the degradation caused by the effect of small frequency offsets when the synchronizer is unlocked.

161

V - CONCLUSIONS

Two lock detectors associated with the GA-TED have been described in this paper. Detector A requires an oversampling frequency of 4/T at the receiver, but its performance is independent of the shape of the channel impulse response and thus it matches to the GA-TED. Detector B uses the same signal samples needed for the GA-TED, i.e., the required oversampling frequency is 2/T, but it does not match to the GA-TED. However, we have shown by means of computer simulations study that Detector B is quasi-independent of channel distortion.

In addition, we have discussed some of the implementation issues and studied the effect of SNR, channel distortion and frequency offset between the symbol clocks used at the transmitter and receiver.

VI - REFERENCES

[1]   G. Karam, V. Paxal and H. Sari, "A variable-rate QPSK demodulator for digital satellite TV broadcasting," IBC '94 Conf. Rec., pp. 646-650, September 1994, Amsterdam.

[2]   G. Karam, K. Maalej, V. Paxal and H. Sari, "Design and performance of a variable-rate QAM modem for digital cable television," to be presented at IBC '95, September 1995, Amsterdam.

[3]   F.M. Gardner, "A BPSK/QPSK timing-error detector for sampled receivers," IEEE Trans. Commun., vol. COM-34, pp. 423-429, May 1986.

JITTER ACCUMULATION IN CASCADED REGENERATORS[*]

Katrien Bucket and Marc Moeneclaey [**]

Communication Engineering Laboratory
University of Ghent
Sint-Pietersnieuwstraat 41
B-9000 Gent
BELGIUM.

Phone : + 32-9-264 34 13
Fax : + 32-9-264 42 95
E-mail : Marc.Moeneclaey@lci.rug.ac.be

Abstract - In this contribution we investigate the jitter accumulation in a long chain of identical regenerators that perform timing recovery by means of a nonlinearity, followed by either a PLL or a bandpass filter for extracting the spectral line at the symbol rate. Assuming jitter peaking, we derive a simple inequality in terms of the jitter transfer function that determines whether the random jitter or the systematic jitter accumulates faster. From this inequality, we show that a second-order PLL yields systematic jitter that accumulates faster than random jitter, whereas the opposite holds for a symmetrical bandpass filter with second-order equivalent lowpass transfer function.

I. INTRODUCTION

For transmitting digital information over very long distances (e.g. a transoceanic fiber optical link), a cascade of regenerators is often used. Each regenerator detects the digital information transmitted by its upstream regenerator, and sends the detected information to its downstream regenerator. The regenerator's symbol clock needed for the detection and transmission of the digital information is recovered from the signal received by the considered regenerator.

In many cases, timing extraction involves passing the received signal through a nonlinearity, in order to generate a sinusoidal component at the symbol rate. This sinusoidal component at the output of the nonlinearity is extracted by a narrowband bandpass filter tuned to the symbol rate, or by a phase-locked loop. Because of additive noise and the randomness of the data (self-noise), the symbol clock derived from the extracted sinusoid is affected by jitter [1]. The jitter components caused by additive noise and self-noise are called random jitter and systematic jitter, respectively.

When the jitter transfer function (to be dicussed in section II) exhibits peaking, both the random and the systematic jitter accumulate exponentially along the regenerator chain. Some authors find that systematic jitter accumulates faster than random jitter [2], while others have observed the opposite phenomenon [3].

In this contribution we investigate the accumulation of the random and the systematic jitter. We derive a very simple condition, involving the magnitude and the phase of the jitter transfer

---

function, which determines which type of jitter accumulates faster. Our main conclusion is that random jitter accumulates faster than systematic jitter when timing extraction is performed by means of a typical bandpass filter, and the opposite occurs when a typical PLL is used for timing extraction.

II. JITTER ANALYSIS

The jitter in a chain of identical regenerators is investigated by means of the Chapman model, shown in Fig.1. In this model, $\theta_k(t)$ denotes the jitter at the output of the k-th regenerator, measured with respect to a jitter-free clock. The noise terms $n_k(t)$ can be decomposed into a random component $n_{r,k}(t)$ caused by additive noise, and a systematic component $n_{s,k}(t)$ caused by self-noise. The random components $n_{r,k}(t)$ are uncorrelated, and have a constant power spectral density S_r. The systematic components are identical ($n_{s,k}(t) = n_s(t)$) because each regenerator transmits the same data sequence, and have a constant power spectral density S_s. The quantity H(f) is the jitter transfer function of the regenerator. In the case of timing extraction by means of a PLL, H(f) equals the closed loop transfer function of the PLL. In the case of timing extraction by means of a symmetrical bandpass filter, H(f) equals $H_E(f)/H_E(0)$, where $H_E(f)$ is the transfer function of the equivalent lowpass filter. In both cases (PLL or bandpass filter), H(0)=1.

We restrict ourselves to the following type of jitter. The jitter $\theta_k(t)$ is the jitter at the output of the k-th regenerator, measured with respect to a jitter-free clock. This jitter has a random and a systematic component. The variances of these components at the output of the k-th regenerator are denoted by $\sigma_{t,r}^2(k), \sigma_{t,s}^2(k)$ where the subscript t refers to jitter, while the subscripts r and s refer to the random and the systematic component of the jitter.

From Fig.1, it can be derived that

$$\frac{\sigma_{t,j}^2(N)}{\sigma_{t,j}^2(1)} = \frac{1}{B_n} \int_0^\infty T_{t,j}(f,N) \, df \qquad (1)$$

where j=r or j=s, B_n is the noise bandwidth of the jitter transfer function, i.e.

$$B_n = \int_0^\infty |H(f)|^2 \, df \qquad (2)$$

and ([2]-[6])

$$T_{t,r}(f,N) = |H(f)|^2 \frac{|H(f)|^{2N} - 1}{|H(f)|^2 - 1} \qquad (3)$$

$$T_{t,s}(f,N) = |H(f)|^2 \frac{|H^N(f) - 1|^2}{|H(f) - 1|^2} \qquad (4)$$

We will investigate the jitter accumulation in the case of a jitter transfer function given by

$$H(f) = \frac{1 + 2j\alpha\zeta(f/f_o)}{1 + 2j\zeta(f/f_o) - (f/f_o)^2} \qquad (5)$$

where f_o and ζ denote the natural frequency and the damping factor. For $\alpha=0$, H(f) is the jitter transfer function corresponding to a second-order symmetric bandpass filter. For $\alpha=1$, H(f) is the jitter transfer function corresponding to a second-order PLL.

When $|H(f_m)|^2 = \max |H(f)|^2$ exceeds one, jitter peaking occurs. In this case, the variances of the considered jitter components will increase exponentially with N, because the frequency components near $f=f_m$ are amplified by each regenerator in the cascade. For $\alpha=0$ and $\alpha=1$, jitter peaking occurs for $\zeta \leq \sqrt{2}/2$ and irrespective of ζ, respectively. The amount of jitter peaking in dB is expressed as $20 \log |H(f_m)|$.

In the case of jitter peaking and a large number N of regenerators, the integral in (1) is dominated by the contribution from the integrand near $f=f_m$. Using a similar approach as in [5], the following approximations are obtained near $f=f_m$:

$$T_{t,r}(f,N) \approx \frac{|H(f_m)|^2}{|H(f_m)|^2 - 1} |H(f_m)|^{2N}$$

$$\exp(-\frac{N}{2}L_1(f_m)(f - f_m)^2) \qquad (6)$$

$$T_{t,s}(f,N) \approx \frac{|H(f_m)|^2}{|H(f_m)-1|^2}|H(f_m)|^{2N}$$

$$\exp(-\frac{N}{2}L_2(f_m)(f-f_m)^2) \tag{7}$$

where

$$L_2(f_m) = -\frac{d^2}{df^2}\Big[\ln|H(f)|^2\Big]_{f=f_m} \tag{8}$$

This yields [5], [6]

$$\frac{\sigma_{t,r}^2(N)}{\sigma_{t,r}^2(1)} \cong \frac{1}{B_n}\frac{|H(f_m)|^2}{|H(f_m)|^2-1}\sqrt{\frac{2\pi}{NL_2(f_m)}}|H(f_m)|^{2N} \tag{9}$$

$$\frac{\sigma_{t,s}^2(N)}{\sigma_{t,s}^2(1)} \cong \frac{1}{B_n}\frac{|H(f_m)|^2}{|H(f_m)-1|^2}\sqrt{\frac{2\pi}{NL_2(f_m)}}|H(f_m)|^{2N} \tag{10}$$

From the above, it follows that for large N, the dependence on N is the same for all considered jitter variances : the jitter variances are proportional to $|H(f_m)|^{2N}/\sqrt{N}$. It is the constant of proportionality that determines whether the increase of the random jitter is larger than the increase of the systematic jitter. From (9)-(10), we obtain for large N :

$$\frac{\sigma_{t,r}^2(N)/\sigma_{t,r}^2(1)}{\sigma_{t,s}^2(N)/\sigma_{t,s}^2(1)} = \frac{|H(f_m)-1|^2}{|H(f_m)|^2-1} \tag{11}$$

Hence, when the ratio R, given by

$$R = \frac{|H(f_m)-1|^2}{|H(f_m)|^2-1} \tag{12}$$

is larger (smaller) than one, it is the random (the systematic) jitter that exhibits the larger increase, both for the jitter and the alignment jitter.

The condition R>1 is easily shown to be equivalent with

$$Re[H(f_m)] = |H(f_m)|\cos(\psi(f_m)) < 1 \tag{13}$$

where $\psi(f)$ is the phase of the jitter transfer function H(f).

The condition (13) can be interpreted in the following way. The random jitter at the N-th regenerator consists of contributions caused by $n_{r,k}(t)$ for k=1,...,N; these contributions are uncorrelated. The systematic jitter at the N-th regenerator consists of contributions caused by $n_{s,k}(t)$ for k=1,...,N; the contributions from $n_{s,k}(t)$ and $n_{s,l}(t)$ are correlated; their correlation coefficient evaluated at the jitter peaking frequency is $\cos((k-l)\psi(f_m))$. Hence, when $\psi(f_m)$ is sufficiently small, the systematic contributions are essentially in-phase : most of the correlation coefficients are large, so that the sum of the correlated contributions (i.e. the systematic jitter) yields a larger variance than the sum of the uncorrelated contributions (i.e. the random jitter). On the other hand, when $\psi(f_m)$ is sufficiently large, many of the correlations $\cos((k-l)\psi(f_m))$ are negative, because the systematic contributions are not well aligned in phase; consequently, the sum of the correlated contributions (i.e. the systematic jitter) yields a smaller variance than the sum of the uncorrelated terms (i.e. the random jitter).

In the case of timing extraction by means of a second-order PLL, the ratio R is given by

$$R = \frac{(f_m/f_o)^2}{2-(f_m/f_o)^2} \tag{14}$$

where

$$(f_m/f_o) = \frac{\sqrt{1+8\zeta^2}-1}{4\zeta^2} \tag{15}$$

It is easily verified that R<1 for $0<\zeta<1$, which indicates that systematic jitter accumulates faster than random jitter, irrespective of the amount of peaking.

When timing extraction is performed by means of a symmetrical second-order bandpass filter, the ratio R is given by

$$R = \frac{1+2\zeta^2}{1-2\zeta^2} \tag{16}$$

provided that $\zeta^2<1/2$ (no peaking occurs for $\zeta^2>1/2$). We observe that R>1 for $0<\zeta^2<1/2$, so that random jitter accumulates faster than systematic jitter.

III. RESULTS

In Fig.2, the ratio R is shown as a function of the jitter peaking, for timing extraction by means of a bandpass filter (i.e. $\alpha=0$ in (7)) and by means of a PLL (i.e. $\alpha=1$ in (7)). We observe that R>1 when a bandpass filter is used, and R<1 when a PLL is used. Hence, for a long chain of regenerators, the random jitter exhibits a larger (smaller) increase than the systematic jitter, when timing extraction is performed by means of a bandpass filter (by means of a PLL). This is confirmed by Figs. 3 and 4, which compare the increase of the random and systematic components of the jitter, for both the bandpass filter and the PLL timing extraction. Figs. 3 and 4 show the correct results (1) obtained by numerical integration along with the asymptotic (large N) results (9)-(10). In the case of a large amount of jitter peaking and a large number N of regenerators in the considered chain, the correct and asymptotic results match quite well; this indicates that the ratio of random to systematic jitter accumulation is close to the ratio R, given by (12) and shown in Fig.2. Note that for a jitter peaking of about 0.15 dB, the random jitter accumulation is about 10 dB larger (smaller) than the systematic jitter accumulation when timing extraction is performed by means of the bandpass filter (by means of the PLL).

IV. CONCLUSIONS

In this paper we have shown that, in the case of jitter peaking and a large number of regenerators, the increase of the random components of the jitter and the alignment jitter is larger (smaller) than the increase of the systematic components of the jitter and the alignment jitter, provided that the real part of the jitter transfer function (evaluated at the frequency where the peaking occurs) is smaller (larger) than one.

When timing extraction is performed by means of a symmetrical second-order bandpass filter, the above condition indicates that random jitter accumulates faster than systematic jitter (as observed in [3]). On the other hand, when a PLL is used for timing extraction, the above condition indicates that systematic jitter accumulates faster than random jitter (as observed in [2]).

References

[1]   E.O. Sunde, "Self-Timing Regenerative Repeaters", Bell System Technical Journal, Vol. 36, July 1957.

[2]   T. Shimamura and I. Eguchi, "Analysis of Jitter in a Chain of PLL Timing Recovery Circuits", IEEE Transactions on Communications, Vol. COM-25, No. 9, September, 1977.

[3]   P.R. Trischitta and E.C. Varma, "Jitter in Digital Transmission Systems", Artech House, 1989.

[4]   H. Meyr, L. Popken and H.R. Mueller, " Synchronization Failures in a Chain of PLL Synchronizers", IEEE Transactions on Communications, Vol. COM-34, No. 5, May 1986.

[5]   B.R.L. Duttweiler, "The Jitter Performance of phase-locked loops extracting timing from baseband data waveforms", Bell System Technical Journal, Vol. 95, Jan 1976.

[6]   U. Mengali and G. Pirani, " Jitter Accumulation in PAM Systems", IEEE Transactions on Communications, Vol. COM-28, No. 8, August 1980.
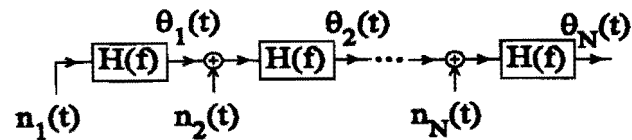
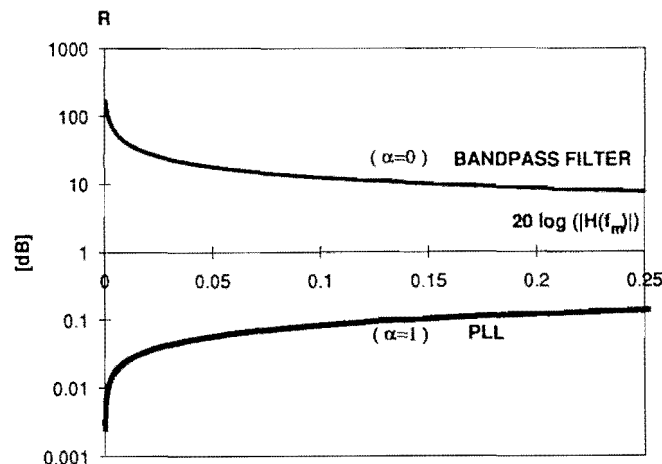Figure 1 Jitter accumulation along a chain of cascaded regenerators
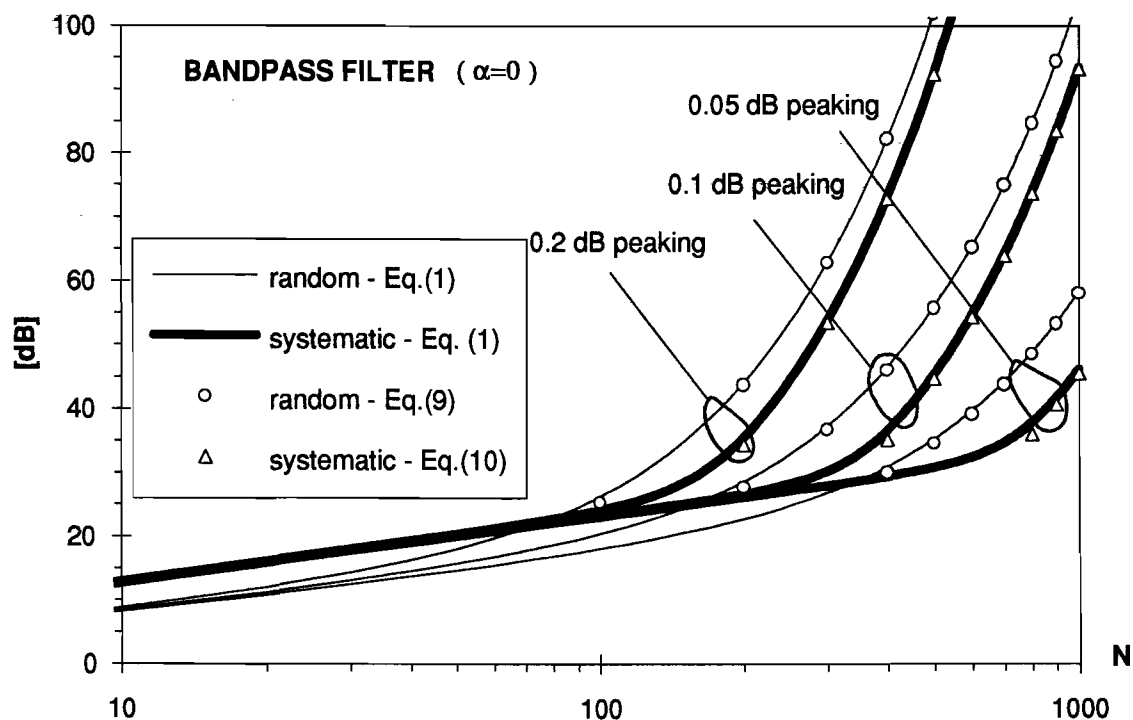

Figure 2 Ratio R as a function of $20 \log|H(f_m)|$

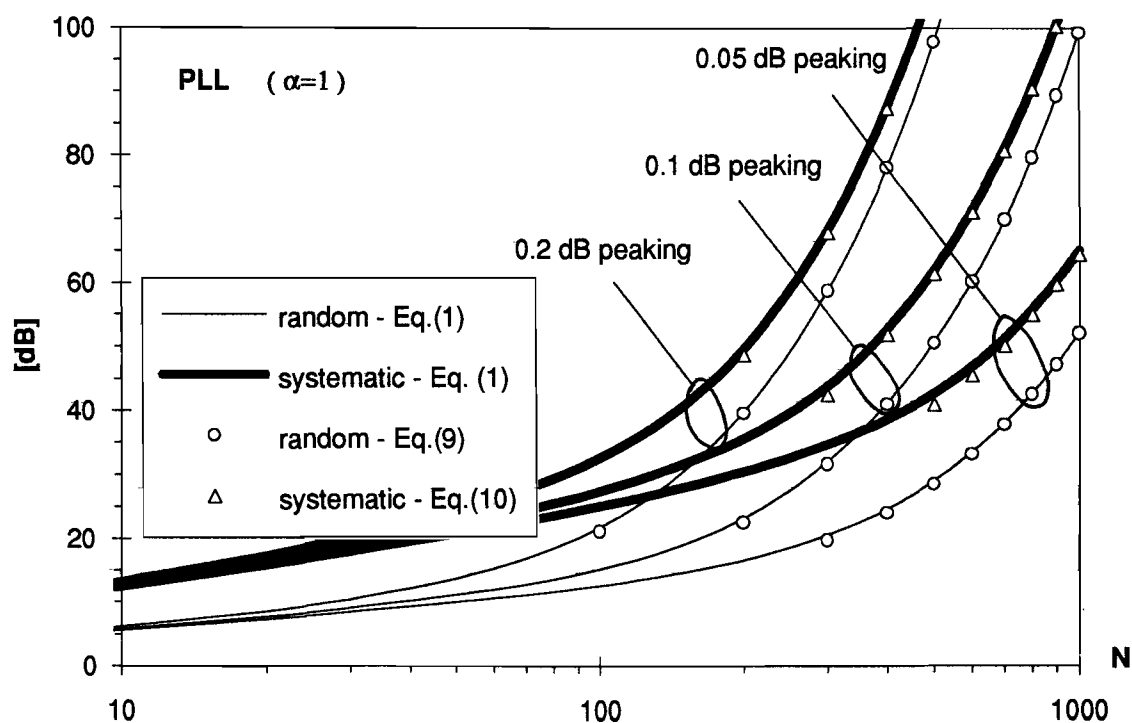Figure 3 Jitter accumulation for a bandpass filter ($\alpha = 0$)

Figure 4 Jitter accumulation for a PLL ($\alpha=1$)

AN ALL DIGITAL DEMODULATOR FOR TELEMETRY SIGNALS

Marie-Laure BOUCHERET[1], Ivar MORTENSEN[1]
Guy LESTHIEVENT[2]

(1) Ecole Nationale Supérieure des Télécommunications (TELECOM Paris), site de Toulouse , 10 avenue E Belin, 31028 Toulouse Cédex
E-mail : bouchere@enst-tlse.fr, tel: (33) 62 17 29 87, mortense@enst-tlse.fr, tel : (33) 62 17 29 83

(2) Centre National d'Etudes Spatiales (CNES), 18 avenue E Belin, 31055 Toulouse Cédex
E-mail : g.lesthievent@bach.cnes.fr, tel : (33) 61 27 32 92

ABSTRACT

There is a growing use of Trellis Coded Modulations for systems in which a gain in terms of power or bandwidth is required. The receiver has two major functions : carrier and symbol synchronisation and decoding/demodulation. This paper deals with the synchronisation part of an all digital demodulator, with input signals being telemetry data received from observation satellites or microsatellites. This study has been performed under CNES/DGA funding [CNE-94].

1 INTRODUCTION :

With the growing interest in Earth Observation, the X band (8.025 to 8.4 GHz), allocated to payload telemetry for this service, is becoming more and more crowded. The use of quasi-polar circular Low Earth Orbit satellites, of higher precision instruments (and the coupling between them : optical with radar for instance) increases the overall data rate to several hundreds of megabits per second. Moreover, the transmission scheme used for telemetry must also comply with the maximum authorized power flux density on earth for those bands. So, there is a need for high overall spectral efficiency.

In the same way, micro or small satellites need high power efficient modulations for their telemetry even if the rates are under the megabits per second. As the S telemetry band (2.208 to 2.29 GHz) is also

becoming crowded, there is again a need for high spectral efficiency with good BER performances.

In both cases, the use of Trellis Coded Modulations (TCM) allows the achievement of high spectral efficiency:
- X Band : the system specifications (10^{-6} BER at Eb/No=10dB , spectral efficiency 50% higher than the one of QPSK) leads to the use of 16-ary modulation. A 4*16PSK TCM has been selected bacause of the non-linearity of the satellite channel.
- S band : A 4*8PSK TCM has been selected to comply with system specifications (spectral efficiency of QPSK, 10^{-5} BER at Eb/No=6dB).

Moreover, both MCT have to be rotationally invariant.

The paper will be restricted to the synchronisation aspects of the demodulator.

Main specifications are the following:
* modulation : 16PSK (X band), 8PSK (S band)
* maximum frequency deviation : 0.24% Rs (X band), 20% Rs (S band) where Rs is the symbol rate.
* maximum frequency ramp : -10^{-4} Rs/s (S band), $-7 \cdot 10^{-3}$ Rs/s (S band)
* cycle slipping rate (CSR): $\leq 10^{-7}$ (X band), $\leq 10^{-6}$ (S band).This very tight specification is due to the use of rotationally invariant TCM (each cycle slip generates about 2 errors)

The main differences with former studies are the following:
- large amount of frequency deviation (up to 20 % of the symbol rate for S-band)

168

- low Eb/No
- use of higher order modulations (M=8 or 16, most of the former results being for M=4).

2 RECEIVER ARCHITECTURE :

2 1 General:

The functions to be implemented are the following:

- baseband conversion and generation of in-phase (I) and in-quadrature (Q) channels : In an analog receiver, this function is linked to the phase recovery In a digital receiver, it is a specific circuitry called DAF (Digital Anti-aliasing Filter).

- carrier phase recovery : this is the main difficulty because of the very low specified CSR. Feedforward structures cannot be used because their noise bandwidth is much larger than those of feedback structures. A PLL is necessary.

- carrier frequency recovery :
* 16PSK: With a specified frequency deviation equal to 0.24 % Rs, it is not worth implementing a specific circuitry for frequency correction. This function is performed by the PLL.
* 8PSK : a digital PLL cannot recover a frequency deviation of 20% Rs. A specific circuitry is necessary. In a noiseless environment, a frequency estimator can cope with a frequency deviation up to 6.25% Rs for 8PSK. Our specifications (20% Rs frequency deviation, low Eb/No) leads to the use of a feedback structure.

- symbol timing recovery : This function is not critical, as very robust timing recovery schemes are available. Both feedforward and feedback structures can be used. The main advantage of feedforward structure (no acquisition time) is not important because of the use of a PLL for phase recovery : the acquisition time of the timing loop is negligible wrt the acquisition time of the PLL.

2 2 Architectures:

* 16PSK :

The architecture is shown in figure 2-1. It has the main characteristics of a digital receiver with small frequency deviation wrt the symbol rate:

- first stage is downconversion and I/Q channels generation (DAF)
- matched filtering is performed before frequency/phase correction (small frequency deviation).
- timing recovery is performed prior to phase recovery

* 8PSK:

The architecture is shown in figure 2-2. The differences with the 16PSK case comes from the specified frequency deviation:
- matched filtering is performed after frequency correction.
- frequency/phase loops switching : as will be pointed out later, the frequency loop has to be desactivated before the phase loop starts because frequency variation generated by the frequency loop would drive the PLL out of lock.
- matched filtering is performed after phase correction because the PLL performs frequency tracking (necessary because of the specified frequency ramp) when the frequency loop is switched off.

3 DAF AND SYNCHRONISATION ALGORITHMS:

3 1 Digital Antialiasing Filter :

* 16PSK :

The input signal, centered at FI=B, is sampled at Fe=4 B, where B is the bandwidth occupied by the signal. The signal is then frequency translated by -FI, filtered by a half-band filter H_{PB} and downsampled by a factor 2 (figure below).
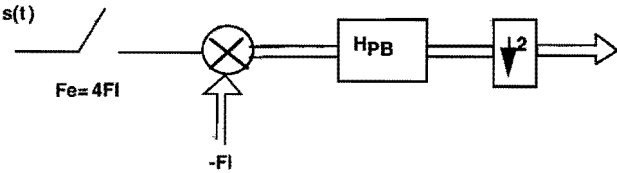
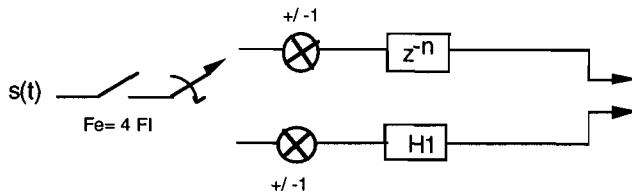
Implementation is very simple because:
- frequency translation by -FI is a simple multiplication by $(-j)^n$
- half of the filter coefficients are equal to zero (property of half-band filters).

- the half-band filter is polyphased in two paths : one is a simple delay, the other is a filter H1 with the odd coefficients of H_{PB}.

The resulting architecture is given hereafter:

* 8PSK :

Two possibilities have been investigated:
- classical DAF (as for 16PSK case)
- a simplified DAF, described hereafter.

When a frequency loop is necessary, it can be seen that frequency correction can be used to perform downconversion, further filtering being performed by the matched filter.

In this scheme, the input signal centered at FI=B/2 is sampled at Fe = 2B. Then, it is frequency translated by -FI (which is a simple multiplication by $(-j)^n$). Fine downconversion is performed by the frequency loop, I and Q channels are generated at the matched filter output.

frequency correction
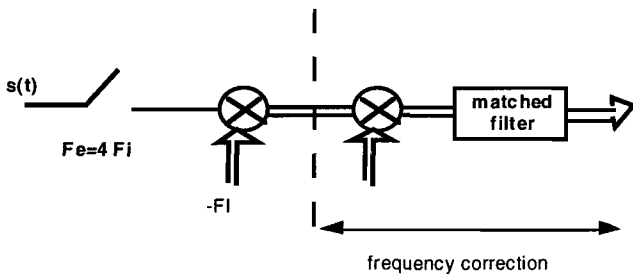
The classical DAF has been selected because it allows the use of IF filters with broader transition bandwidths.

3 2 Carrier frequency recovery (8PSK):

Numerous detectors are available in the litterature. In [MOE-94], it is shown that they are all related to the detector derived from the Maximum Likelihood (ML) criterion [GAR-90] (called ML detector in the following).

The ML detector output is:

$$e(k)=Im\{x(t_0+kT)y^*(t_0+kT)\}$$

where:
- t_0 is the timing error,
- T the symbol period
- x(t) the matched filter (g(t)) output
- y(t) the frequency matched filter (gf(t)) output.

gf(t) is given by : $gf(t)= -2j\pi \, t \, g(-t)$.

An extra filter is necessary, which leads to an increased complexity.

A simplified algorithm (called SML in the following) which does not require an extra filter has been proposed in [KAR-92]. The frequency matched filter is approximated by a filter with impulse response : $gs(t) = -j \, sgn(t) \, g(-t)$. Its output is easily derived from the matched filter output. Frequency responses of $gs(t)$, gf(t) and g(t) and given hereafter (with Rs=0.25 Hz).

It can be noticed than $gs(t)$ is less selective than gf(t).
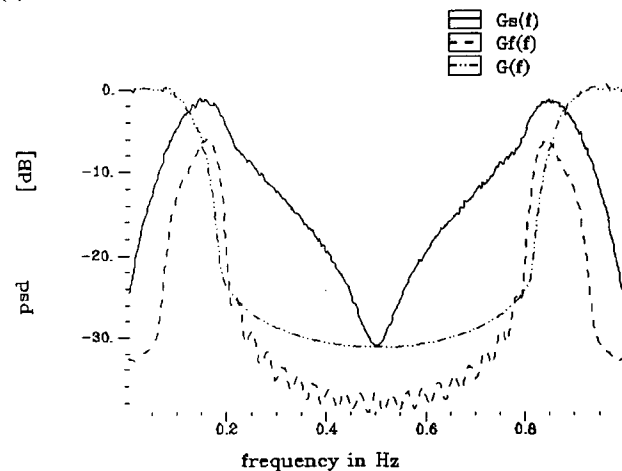
Another problem is relative to the dependence on timing error. The slope at the origin for the ML detector is given by :

$$A(t_0) \propto \frac{1}{\alpha} cos^2\left(\frac{\pi t_0}{T}\right)$$

where α is the rolloff.

With 1 sample/symbol, all the loop parameters are timing-dependent. It can be easily shown that 2 samples/symbol are sufficient to get a timing-independent frequency loop.

Two types of loops can be defined:
- timing independent "asynchronous" loops with 2 samples/symbol

- timing dependent "synchronous" loops with 1 sample/symbol

The synchronous loops leads to lower variance and has been selected.

Next step is the detector selection. The following table gives the frequency error standard deviation for both detectors at Es/N0=10 dB. Bl is the loop bandwidth.

	ML detector	SML detector
BlT=0.0001	0.23 % Rs	0.3 % Rs
BlT=0.001	0.55 % Rs	0.6 % Rs
BlT=0.01	1.7 % Rs	2.8 % Rs

For Es/No=10 dB, performances are similar. The SML detector is selected because of its lower complexity.

The loop bandwidth value Bl results from a trade-off between acquisition time and steady state jitter. Loop bandwitdh switching between acquisition and tracking is necessary to reduce acquisition time.

3-3 Carrier phase recovery:

The use of a feedback structure is mandatory because of the specified cycle slipping rate. It is important to note that the frequency loop has to be switched off before the phase loop starts : in effect, the frequency variation generated by the frequency loop would drive the PLL out of lock.

Various algorithms for MPSK are available [MOE-94-2], [GAR-88].

* *detector selection for MPSK (M=8 or 16):*

Two detectors have been studied:

- detector derived from the ML criterion :

$$u_\phi(k)=\text{Im}\{p^M(k)\}$$

with $p(k)=\rho(k)\exp\{j\psi(k)\}$ the matched filter output at time kT.

- Viterbi et Viterbi [VIT-83] (or NDAVVFB) algorithm :

$$\hat{\varphi}=\frac{1}{M}\text{Arctan}\left\{\frac{\sum\limits_k \text{Im}(\gamma(k))}{\sum\limits_k \text{Re}(\gamma(k))}\right\}$$

$\gamma(k)=F[\rho(n)]\exp[jM\psi(n)]$, where $\rho(n)$ and $\psi(n)$ are respectively the modulus and the phase of $p(n)$ and $F[\rho(n)]$ is a non-linear function (ρ^m). The optimum value of m depends on the SNR.

Both detectors have similar performances in terms of variance and leads to hardware complexity of the same order.

* *Acquisition time :*

* *8PSK :* Initial frequency deviation is given by the frequency standard deviation (0.3% Rs) in the frequency loop. Bandwidth switching is necessary to achieve an acceptable acquisition time. It should be noted that acquisition fails with too large bandwidths because of the low Es/N0.

* *16PSK :* Initial frequency deviation is equal to 0.24%Rs. Same remarks for 8PSK apply.

* *Cycle slipping rate (CSR) :*

This parameter is very important when rotationally invariant MCT are used. An upper bound is given in [JON-93]:

$$\text{CSR}=\exp\left\{-\frac{2}{M^2\sigma_\varphi^2}\right\}$$

where σ_φ^2 is the PLL closed loop variance.

With BlT=0.5 %, the cycle slipping rate is below the specified values for both 8PSK and 16PSK.

3 4 Timing recovery:

This function is the less critical one.

Two possibilities have been investigated:
- feedforward structures, based on the use of a estimator (such as [OER-88]). The main advantage of this structure is that there is no acquisition time.
- feedback structure : a carrier independant detector is necessary. The classical Gardner detector [GAR-86] with 2 samples per symbol has been considered. Correction is performed by interpolating

the matched filter output (this function is efficiently implemented by polyphasing the matched filter).

Acquisition time of the timing loop is not critical. It is in fact negligible compared to the acquisition time of the PLL. Moreover, we have checked by exhaustive simulation that, in the case of 8PSK, timing and frequency loops have a correct behaviour during the acquisition time.

So, a feedback structure based on the Gardner detector has been selected.

4 RESULTING ARCHITECTURES AND PERFORMANCES :

This paragraph sums up the results of the previous ones.

- 8PSK (figure 4-1):
* Frequency/phase correction is performed before the matched filter
 * frequency loop :
 - detector : SML
 - loop bandwidth (acquisition): 10^{-3} Rs
 - acquisition time : 6000 symbols
 - loop bandwidth (tracking): 10^{-4} Rs
 - frequency jitter at Es/No=10dB: 0.3% Rs
 * phase locked loop:
 - Viterbi detector
 - loop bandwidth (acquisition): $4\ 10^{-2}$ Rs
 - acquisition time : tens of thousands of symbols
 - loop bandwidth (tracking): $5\ 10^{-3}$ Rs
 - phase jitter at Es/No=10 dB : 2.5°
 - CSR : $2\ 10^{-7}$
 * timing recovery:
 - Gardner loop

- 16PSK (figure 4-2):
* phase correction is performed after the matched filter
 * phase locked loop:
 - Viterbi detector
 - noise bandwidth : 10^{-2} Rs(acquisition), $5\ 10^{-3}$ Rs (tracking)
 - acquisition time : tens of thousands of symbols
 - phase jitter at Es/No=16 dB : 0.8°
 - CSR : $2\ 10^{-17}$
 * timing recovery:
 - Gardner loop

5 CONCLUSION

A all-digital demodulator for rotationally invariant TCM based on 8PSK and 16PSK has been proposed. The major difficulty is relative to the use of a PLL, necessary to comply with the specified CSR.

6 REFERENCES

[CNE-94] "Étude de définition de deux TCM et des algorithmes de synchronisation de porteuse et de rythme associés", contrat CNES n° 962/94/CNES/131/00

[GAR-86] F M Gardner "A BPSK/QPSK timing error detector for sampled receivers", IEEE Transactions on Communications, may 1986

[GAR-88] F M Gardner "Demodulator reference techniques suited for digital implementation", final report, ESTEC contract n° 6847/86/NL/DG, august 1988

[GAR-90] FM GARDNER "Frequency detectors for digital demodulators via Maximum-Likelihood derivation", final report, ESTEC contract n° 8022/88/NL/DG, part 2, juin 90

[JON-93] G DE JONGHE et M MOENECLAEY "Asymptotic cycle slip probability expression for the NDA feedforward carrier synchronization for M-PSK", ICC 93

[KAR-92] G Karam, F Daffara, H Sari, "Simplified versions of the maximum likelihood frequency detector",
GLOBECOM 92, Orlando, USA

[MOE-94] M Moeneclaey "Overview of digital algorithms for carrier frequency synchronisation", proceedings of DSP 94, London, 26-28 september 1994

[MOE-94-2] M Moeneclaey et al."ML oriented NDA carrier synchronization for general rotationally symetric signal constellations", IEEE on COM , august 94

[OER-88] Oerder, Meyr "Digital filter and square timing recovery", IEEE transactions on Communications, COM-36, may 1988

[VIT-83] Viterbi AJ et Viterbi AM "Non linear estimation of PSK-modulated carrier phase with application to burst digital transmission", IEEE on IT, vol IT-29, 1983
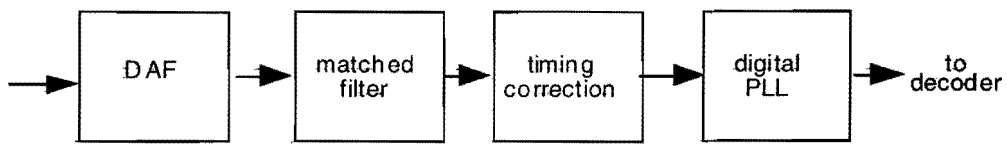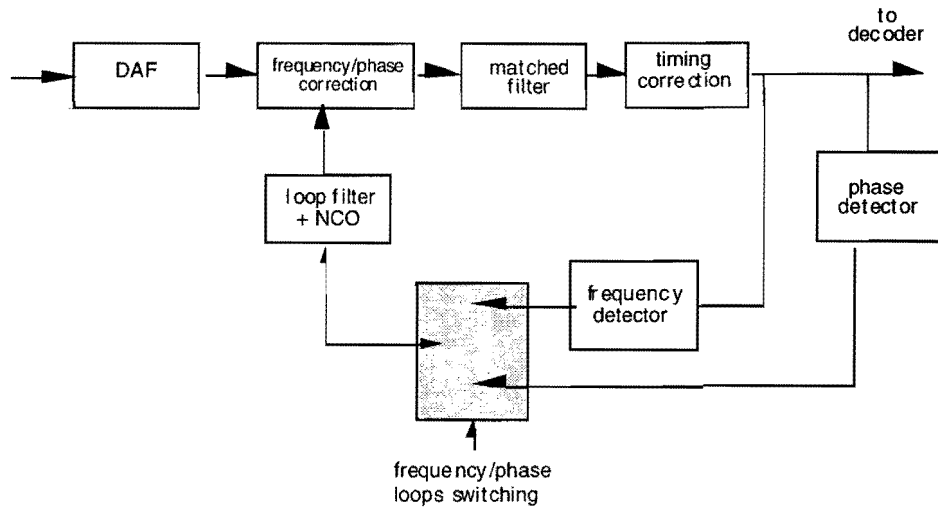
figure 2-1 : 16PSK case

frequency/phase
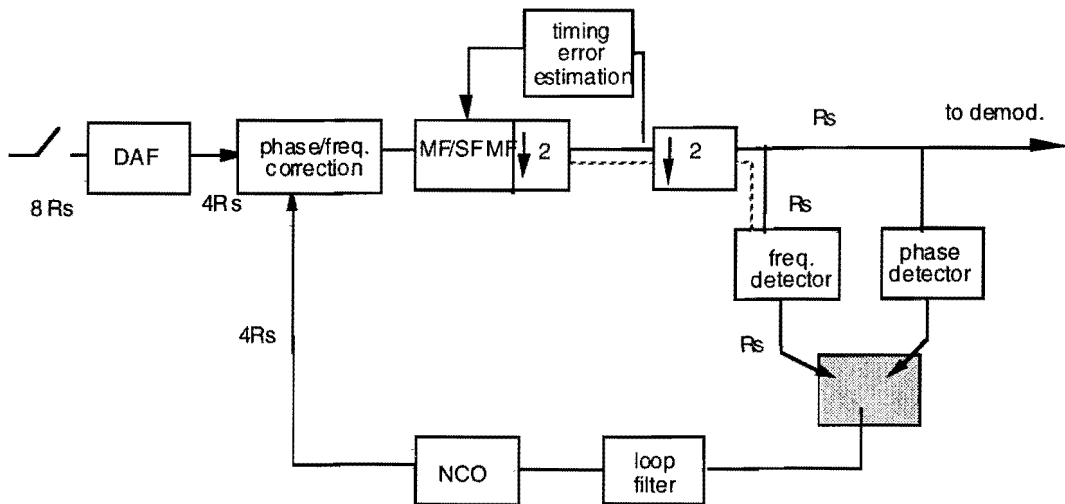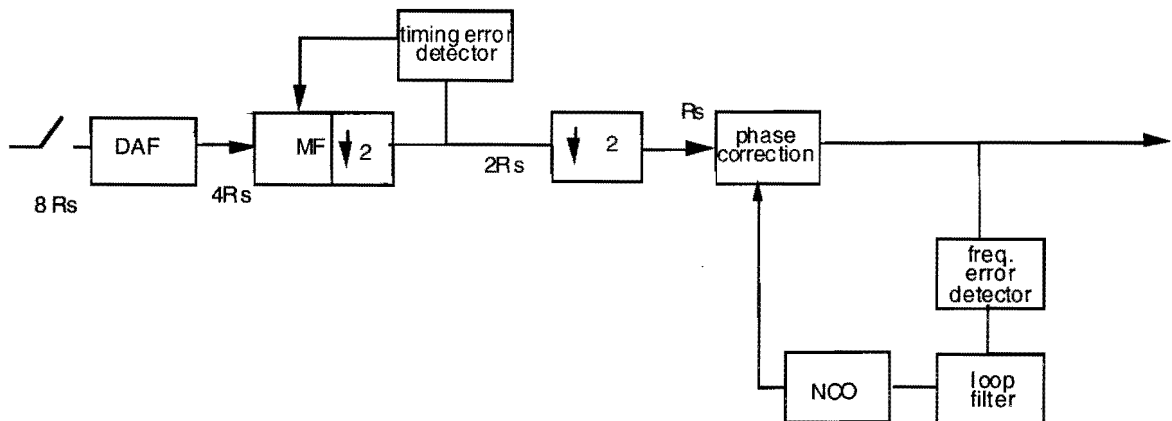loops switching

figure 2-2 : 8PSK case

figure 4 1 : 8PSK : final architecture

4 2 : 16PSK : final architecture

173

BER calculation of the optical coherent DPOLSK receiver based on Gaussian approximation of the pdf.'s.

Henk Tattje & Wim van Etten
Teleinformatics and Open Systems group
Twente University of Technology
POB 217
7500 AE Enschede
Fax : +3153333815
E-mail : tattje@cs.utwente.nl
          etten@cs.utwente.nl

In the DPOLSK transmission system the state of polarisation (SOP) of the lightwave is the information carrying parameter. The receiver calculates the Stokes parameters of the incomming lightwave, which completely characterize the SOP. Exact expressions for the pdf's of the relevant parameters are derived implicitely. Explicit evaluation of those expressions is quite cumbersome and gives little insight in the analysis process. It is shown that for SNR's which are not too small, all relevant variables may well be approximated by gaussian distributions. It is shown how the non-linear operations in the receiver change the signal and noise spectra, in this way enabling to study the influence of phase noise on the performance of the receiver as well. The results of the analysis show that describing the pdf's by gaussian distributions approximates the BER's within tens of a dB compared with the exact analysis. Our method is much less involved and gains more insight in the signal and noise processing.

The DPOLSK receiver

The DPOLSK receiver is outlined in fig 1. The received optical signal is split in the polarization beam splitter into two orthogonal polarized signals. The local laser signal is split likewise into two orthogonal polarized components, which are added to the corresponding polarized input signal components.
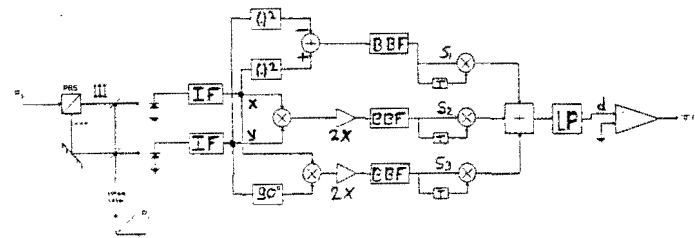

Fig. 1. The DPOLSK receiver

These optical signals are then converted to two electrical signals in the photodiodes. The signals from the photodiodes are filtered by two IF filters.

The information is differentally encoded by flipping the polarization of the transmitted signal. The state of polarisation at the receiver is random because of the transmission through the fiber. Whatever the change in polarisation caused by the fiber, the polarisation states of the received signal will correspond with two antipodal points on the Poincaré sphere.[1] These points can be found by computing the Stokes parameters of the received signal.

The relevant Stokes parameters are:
  (x and y denoting the IF signals)
  S1 : $x.y$
  S2 : $x.H\{y\}$; $H\{y\}$ denoting the Hilbert transform of the signal y.
  S3 : $x^2 - y^2$

S1, S2 and S3 are the x, y and z coordinates respectively of a point S on the Poincaré sphere.

To recover the information sent, we must take the product of consecutively computed Stokes parameters and evaluate the sum of these products.

Analysis

To determine the Stokes parameters we process the signals from the fotodiodes in three subsystems:
Box 1: multiply
Box 2: multiply
Box 3: square-and-difference

These three signal are lowpass filtered and then fed to the delay-and-multiply boxes.
The output signals are added and after lowpass filtering fed to a comparator where the decision 0 or 1 is made.

The analysis is straightforward if the signals in the receiver are all modelled as Gaussian processes.

We shall discuss the actuall distributions and compare them with gaussian distributions.

Without loss of generality, we will assume that the received signal is linearily polarised and all the power is in the x-branch. [2]
This implies that the signal in the x-branch has mean: μ_x and the y-branch has mean :zero.

Our second assumption is that the IF signals x and y are gaussian distributed, which is valid if the power of the local oscillator is large enough.[3]

Since we want to study the high SNR case, we assume that $\mu_x > \sigma_x$ .The signals x and y are gaussian distributed and uncorrelated .

$$f_x(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x - \mu_x}{\sigma_x}\right)^2}$$

$$f_y(y) = \frac{1}{\sigma_y \sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{y}{\sigma_y}\right)^2}$$

The power spectral density of x and y in the passband of the IF filter is given by

$$S_{xx}(f) = S_{yy}(f) = eRB_N \frac{1}{2} P_L$$

The bandwidth B_N of the IF filter equals 1/T with T equal to the symbol time.
R is the responsivity of the photodiodes, P_L is the local laser power.
From this it follows :

$$\sigma_x^2 = \sigma_y^2 = 2eRB_N \frac{1}{2} P_L$$

Box 1: multiply

z=x.y;
x and y are gaussian and independent

$$x = N(\mu_x, \sigma_x^2) \quad y = N(0, \sigma_y^2)$$

175

Now the autocorrelation of z is given by

$$R_{zz}(\tau) = E\{x(t)x(t+\tau)y(t)y(t+\tau)\}$$

which will be approximated by a gaussian pdf

$$f_z(z) = N(0, \mu_x^2 \cdot \sigma_y^2)$$

Since x and y are independent it follows:

$$R_{zz}(\tau) = E\{x(t)x(t+\tau)\}E\{y(t)y(t+\tau)\}$$
$$R_{zz}(\tau) = R_{xx}(\tau) \cdot R_{yy}(\tau)$$
$$R_{zz}(\tau) = (C_{xx}(\tau) + \mu_x^2) \cdot R_{yy}(\tau)$$

Since we are interested in the tail of the pdf we will compare the distributions. In fig 2 the exact distribution of z and the gaussian approximation are shown. At high SNR the approximation is good.

while $\quad |C_{xx}(\tau)| < \sigma_x^2$
and assuming $\quad \mu_x^2 > \sigma_x^2$

$$R_{zz}(\tau) = \mu_x^2 \cdot R_{yy}(\tau)$$
$$S_{zz}(f) = \mu_x^2 \cdot S_{yy}(f)$$
$$\mu_z = 0$$
$$\sigma_z^2 = \mu_x^2 \cdot \sigma_y^2$$

Fig. 2. Comparison of distributions

The exact pdf of z is given by:

$$f_z(z) = \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \frac{1}{|x|} e^{-\frac{1}{2\sigma^2} \cdot \left(\frac{z^2}{x^2} + (x-\mu_x)^2\right)^2} dx$$

with $\quad \sigma^2 = \sigma_x^2 = \sigma_y^2$
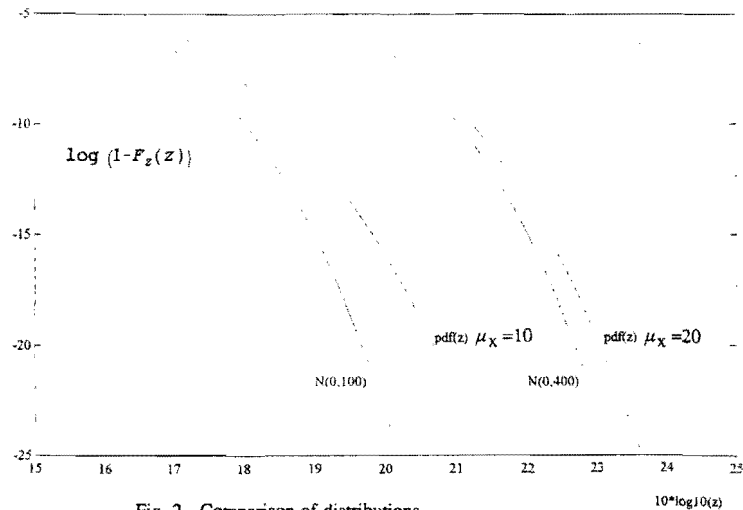
Interesting is that in the high SNR case the term

$$\mu_x^2 \cdot S_{yy}(f)$$

is dominant, so the spectrum of z is the same as the spectrum of y,
there is no doubling of the input spectrum.

Box 2: multiply

$z = x.H\{y\}$;
x and y are gaussian and independent

$$x = N(\mu_x, \sigma_x^2) \quad y = N(0, \sigma_y^2)$$

analysing as in box 1 we find essentially the same result as for box 1:

$$R_{zz}(\tau) = \mu_x^2 . R_{yy}(\tau)$$
$$S_{zz}(f) = \mu_x^2 . S_{yy}(f)$$
$$\mu_z = 0$$
$$\sigma_z^2 = \mu_x^2 . \sigma_y^2$$

The exact pdf of z is given by:

$$f_z(z) = \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \frac{1}{|x|} e^{-\frac{1}{2\sigma^2} \cdot \left(\frac{z^2}{x^2} + (x - \mu_x)^2\right)^2} dx$$

which will be approximated by a gaussian pdf

$$f_z(z) = N(0, \mu_x^2 . \sigma_y^2)$$

Box 3: square and difference

$z = x.x - y.y$
x and y gaussian distributed and independent

$$x = N(\mu_x, \sigma_x^2) \quad y = N(0, \sigma_y^2)$$

Now the autocorrelation of z is given by

$$R_{zz}(\tau) =$$
$$E\{x(t)x(t)x(t+\tau)x(t+\tau)\} +$$
$$E\{y(t)y(t)y(t+\tau)y(t+\tau)\}$$

We know that for gaussian processes with zero mean:

$$E\{x_1 x_2 x_3 x_4\} =$$
$$E\{x_1 x_2\} . E\{x_3 x_4\} +$$
$$E\{x_1 x_3\} . E\{x_2 x_4\} +$$
$$E\{x_1 x_4\} . E\{x_2 x_3\}$$

with μ_x unequal to zero we find:

$$R_{zz}(\tau) =$$
$$2\mu_x^2 . \sigma_x^2 + 4\mu_x^2 . C_{xx}(\tau) + 2C_{xx}^2(\tau) +$$
$$\sigma_x^4 + \mu_x^4 + 2C_{yy}^2(\tau) + \sigma_y^4$$

with $\mu_x^2 > \sigma_x^2$ it follows :

$$R_{zz}(\tau) = C_{xx}^2(\tau) + 4\mu_x^2 . C_{xx}(\tau) + \mu_x^4$$
$$\mu_z = \mu_x^2$$
$$\sigma_z^2 = 4\mu_x^2 . \sigma_x^2$$

The exact pdf of z is given by:

$$f_z(z) = \frac{1}{2} . e^{-\frac{1}{2} . (\mu_x^2 - z)} . I_0(\mu_x . \sqrt{z})$$

Which will be approximated by a gaussian pdf

$$f_z(z) = N(\mu_x^2, 4\mu_x^2 . \sigma_x^2)$$

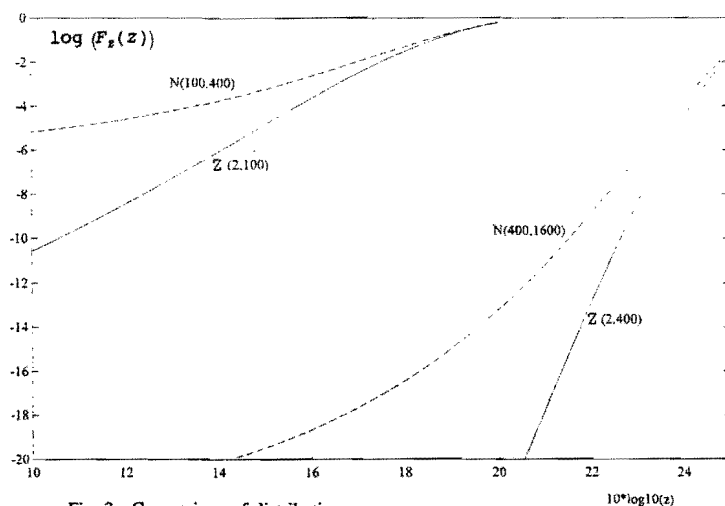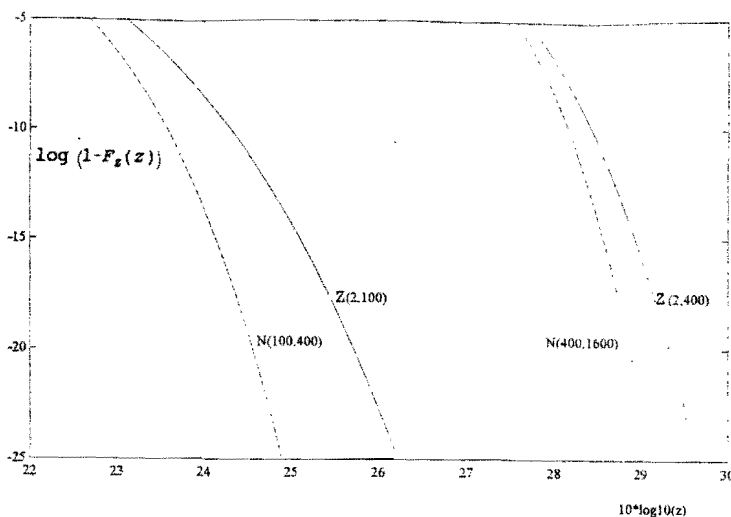In fig 3 the exact distribution and the gaussian approximation are shown. At high SNR the approximation is good.

Fig. 3. Comparison of distributions

Multiply by 2 box.

The output has double the mean and 4 times the variance of the input.

Delay and multiply box.

$z(t) = x(t).x(t-T)$; T = symboltime

$$x = N(\mu_x, \sigma_x^2)$$
$$C_{xx}(T) = 0$$

Now the autocorrelation of z is given by :

$$R_{zz}(\tau) = E\{x(t)x(t-T)x(t+\tau)x(t+\tau-T)\}$$

with $C_{xx}(T) = 0$ and $\mu_x^2 > \sigma_x^2$ it follows :

$$\mu_z = \mu_x^2$$
$$\sigma_z^2 = 4\mu_x^2 . \sigma_x^2$$

Baseband filter

The basebandfilter is a lowpassfilter with bandwidth 1/2T with T equal to the symbol time.

$$H_{bf}(f) = 1; \quad \frac{-1}{2T} < f < \frac{1}{2T}$$
$$H_{bf}(f) = 0; \quad elsewhere$$

We observe that the bandwidth of this filter is half the width of the spectrum of the dominant signal coming from the square-and-difference box.
The power spectral density of this signal is flat, so the variance of the output signal is half the variance of the incoming signal.

while if $\mu_x = 0$ then

$$\mu_z = 0$$
$$\sigma_z^2 = 4\sigma_x^4$$

The exact pdf of z is given by

$$f_z(z) = \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \frac{1}{|x|} e^{-\frac{1}{2\sigma^2} \cdot \left((\frac{z}{x} - \mu_x)^2 + (x - \mu_x)^2 \right)^2} dx$$

178

which will be approximated by a gaussian pdf :

if $\mu_x = 0$ then $f_z(z) = N(0, 4\sigma_x^4)$,
otherwise $f_z(z) = N(\mu_x^2, 4\mu_x^2 \cdot \sigma_x^2)$

Predetection filter

The predetectionfilter is a lowpass filter with bandwith 1/2T with T equal to the symbol time.

$$H_{pf}(f) = 1; \quad \frac{-1}{2T} < f < \frac{1}{2T}$$
$$H_{pf}(f) = 0; \quad elsewhere$$

The input signal from the square and difference box is dominant. The spectrum of this signal is confined to the same bandwidth as the predection filter, so this filter does not change the variance of this signal.

Results

Inspecting the signals coming out of the predection filter we observe that the signal coming from the square-and- difference box is dominant. We find that the output of the predection filter has:

$$\mu_{pf} = \mu_x^4$$
$$\sigma_{pf}^2 = 4\mu_x^6 \sigma_x^2$$
$$f_{pf} = N(\mu_x^4, 4\mu_x^6\sigma_x^2)$$

To compute the BER we must evaluate

$$Q\left(\frac{\mu_{pf}}{\sigma_{pf}}\right) = Q\left(\frac{\mu_x^4}{2 \cdot \mu_x^3 \cdot \sigma_x}\right) = Q\left(\frac{\mu_x}{2\sigma_x}\right)$$

with μ_x originating from the signal power.

In fig 4 this result and the original result found by Benedetto[2] are shown.
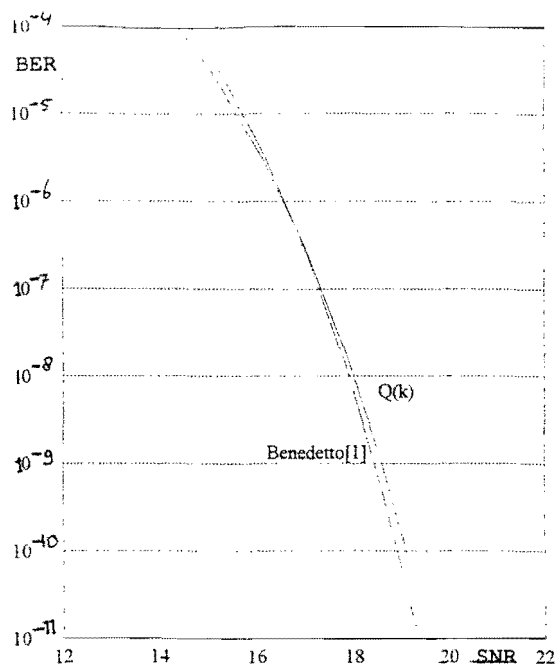

Fig. 4. Comparison of BER vs. SNR (dB)

Conclusion

Comparing the results obtained by Benedetto and from this analysis we conclude that the Gaussian approximation gives satisfying results.

References

[1] Born, M & Wolf, E
PRINCIPLES OF OPTICS
Pergamon Press, 1975

[2] Benedetto, S. et al
THEORY OF POLARISATION
SHIFT KEYING MODULATION
IEEE transactions on communications, Vol. 40, No. 4, April 1992, pp. 708-721

[3] Garrett, I & Jacobsen, G.
THEORY OF OPTICAL HETERO-
DYNE NARROW DEVIATION
FSK RECEIVERS WITH DELAY
DEMODULATION
Journal of Lightwave Technology, Vol. 6, No. 9, September 1988, pp. 1415-1423

179

BIT-ERROR-RATE COMPUTATION IN OPTICAL FIBRE LINKS
BY GRAM-CHARLIER SERIES EXPANSION

Patrice Mégret, Michael Vandroogenbroek, Jean-Christophe Froidure, Véronique Moeyaert, Michel Lamquin and Michel Blondel

Faculté Polytechnique de Mons
Service d'Electromagétisme et de Télécommunications
Boulevard Dolez, 31
B-7000 Mons, Belgium
fax: +32 (0) 65 37 41 99
Email: Megret@fpms.fpms.ac.be

Abstract

We present simulation results for the computation of the bit-error-rate in optical communication systems. The methods we use to solve this problem are the gaussian approximation and the Gram-Charlier series expansion.

1. Introduction

Bit-error-rate computation in optical fibre links is a difficult task because the quantum nature of the light has to be taken into account. This property introduces a fundamental noise called quantum noise [1] and is responsible of the fact that all the noise parameters (mean, variance, statistical distributions, ...) are correlated to the optical digital signal transmitted along the fibre which, in turn, leads to very evolved computational effort.

In order to compute the bit-error-rate, several techniques exist and we use the gaussian approximation [2] to have a quick answer and the Gram-Charlier series expansion [3] to compute with more accuracy. The mathematical theory is also generalised to the case of multilevel optical communication systems [4].

The first method is very simple and can be used to obtain analytical expressions or rapid numerical solutions which allow us to quickly study the influence of the link parameters. Of course this gaussian approximation neglects some fundamental properties of the link and its accuracy strongly depends on the validity of the assumptions used to compute the bit-error-rate.

The second method is an efficient and rapid numerical way to compute the bit-error-rate in an optical fibre link and can take into account a lot of parameters such as the quantum noise, the receiver noises, the fibre dispersion and attenuation, ... The idea is to multiply the gaussian distributions by appropriate correction polynomial functions whose coefficients are easily computed from the link parameters. This method is well suited to study the effects of optical noises (modal noise [5], partition noise [6], chirping noise [7], ...) on the link performances.

2. Modelling

Figure 1 shows a typical optical receiver which is composed of an optical to electrical O/E converter (PIN or APD) followed by appropriate equalisation and amplification circuits.

The incident power P(t) is converted into the electrical signal v(t) and the receiver noises n(t) simply add to v(t) to form the signal v'(t) from which clock is recovered and decision is taken to regenerate the emitted signal.
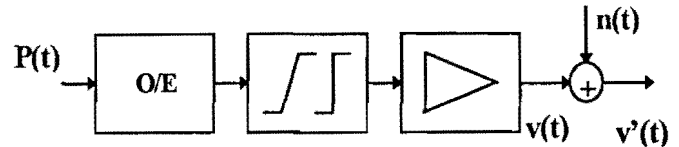

Figure 1: typical digital optical receiver scheme

The incident power P(t) has the expression:

$$P(t) = \sum_{\rho=-\infty}^{+\infty} a_\rho r(t-\rho T)$$

with $a_\rho \in \{\alpha_i\}$ $i = 0,1,\ldots,m-1$ $\qquad$ (1)

where r(t) is the elementary optical signal at the fibre output, a_ρ is the emitted symbol, T is the symbol duration and m is the number of levels (for binary communications m=2).

In practice, the signal r(t) is the convolution of the signal emitted by the laser and the fibre impulse response which allows us to take into account the fibre attenuation and dispersion.

Due to the quantum nature of the light and the optical detection process, it can be shown that the voltage v(t) is stochastic and comes from a filtered randomly multiplied Poisson process [8]. The statistical behaviour of v'(t) can be analyzed through its cumulants of order n which are given by [3]:

$$\lambda_{v'|A,n}(t) = \frac{\eta G_n}{h\nu} P(t) * h^n(t) + \lambda_{n,n} \qquad (2)$$

where η is the detector quantum efficiency, $h\nu$ is the photon energy, G_n is the nth avalanche gain moment, h(t) is the receiver impulse response, P(t) is the incident power given by (1) and $\lambda_{n,n}$ is the nth cumulant of the noise n(t).

The coefficients G_n can be found in [3] and are given by the following recursive relations:

$$G_n = \sum_{m=1}^{2n-1} C_{n,m} G_1^m$$

$$C_{n+1,m} = -m(1-k)C_{n,m} + (m-2)kC_{n,m-2} +$$
$$[1 + (m-1)(1-2k)]C_{n,m-1} \qquad (3)$$

$$C_{n,m} = 0 \quad if \ m \le 0 \ or \ m \ge 2n$$
$$C_{1,1} = 1$$

where k is the ionization ratio of the APD photodetector. In the case of a PIN photodetector, the relation (2) is correct with $G_n = 1$ for all n.

Equation (2) is the key relation and shows that the statistical properties (mean, variance, ...) of v'(t), given the particular emitted symbol $a_p=\alpha_i$, are dependent on the emitted sequence A={$a_{-\infty}$, ..., a_{p-1}, a_p, a_{p+1}, ..., $a_{+\infty}$} which means that there is intersymbol interference on an infinite number of elements. (in (2) | A reminds that these quantities are sequence dependent).

In practice, we shall assume stationarity so the bit-error-rate is independent of the time and can be computed on any transmitted a_p (i.e. a_0) and limit the number of interfering symbols to the p preceding symbols and the s succeeding symbols around a_0.

The bit-error-rate is finally computed from the following relation:

$$BER = \sum_{i=0}^{m-1} \Pr[a_0 = \alpha_i] \Pr[b_0 \ne a_0 | a_0 = \alpha_i] \qquad (4)$$

where Pr[A] means the probability of the event A and b_0 is the regenerated symbol.

The computation of the bit-error-rate by (4) is numerically difficult because we only know an infinite series of cumulants (2) from which we have to build the m probability functions Pr[$b_0 \ne a_0$ | $a_0 = \alpha_i$] of detecting a wrong symbol. Relation (2) also shows that these functions are different and do not necessarily follow gaussian distributions.

To solve this problem, several approximations exist and two are used here:

a) the *gaussian approximation* which only takes the first two cumulants and thus neglects the higher order cumulants. This approach is equivalent to model the probability functions by m different gaussian distributions.

b) the *Gram-Charlier series expansion* in which the m gaussian distributions of point a) are multiplied by correcting polynomial functions whose coefficients are related to the cumulants (2).

3. Gaussian approximation

In this approximation, we suppose that all the cumulants for n > 2 are equal to zero which implies that v'(t) is a gaussian variable with mean and variance at the decision time t_0 for a_0:

$$\mu_{v'|A} = \lambda_{v'|A,1}(t_0) = \frac{\eta G_1}{h\nu} P(t) * h^1(t) + \lambda_{n,1}$$
$$\sigma_{v'|A}^2 = \lambda_{v'|A,2}(t_0) = \frac{\eta G_2}{h\nu} P(t) * h^2(t) + \lambda_{n,2} \qquad (5)$$

From these two relations, we can compute, for each $a_0 = \alpha_i$, Pr[$b_0 \ne a_0$ | $a_0 = \alpha_i$] by calculating the mean, over the sequence A={a_{-p}, ..., a_{-1}, a_0, a_1, ..., a_s}, of the gaussian distributions defined by (5) i.e.:

$$\Pr[v'(t_0)|A] = \frac{1}{\sqrt{2\pi}\sigma_{v'|A}} \exp\left(\frac{v'(t_0) - \mu_{v'|A}}{2\sigma_{v'|A}}\right)^2 \qquad (6)$$

So we have:

$$\Pr\left[v'(t_0)\big|a_0 = \alpha_i\right] = E_A\left[\Pr\left[v'(t_0)\big|A\right]\right] \qquad (7)$$

As (7) is also gaussian, the quantities $\Pr[b_0 \neq a_0 \mid a_0 = \alpha_i]$ are easy to compute by classical integration [11] and the bit-error-rate is finally obtained by (4).

4. Gram-Charlier series expansion

The idea behind this technique is to express the unknown distributions $\Pr[v'(t_0) \mid A]$ as a function of a well-known probability distribution $p(v'(t_0))$ and its successive derivatives $p^{(n)}(v'(t_0))$:

$$\Pr\left[v'(t_0)\big|A\right] = \sum_{n=0}^{+\infty} d_{|A,n}\, p^{(n)}(v'(t_0)) \qquad (8)$$

where:

$$p^{(n)}(v'(t_0)) = \frac{d^n p(v'(t_0))}{dv'(t_0)^n} \qquad (9)$$

It is important to note that only the coefficients $d_{|A,n}$ depend on the sequence A. Now, the goal is to choose the distribution $p(v'(t_0))$ and to find the coefficients $d_{|A,n}$ as a function of the cumulants (2). In practice, one takes for the well-known distribution the normalized gaussian variable $Z(x)$:

$$Z(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-x/2\right)^2 \qquad (10)$$

with:

$$x = \frac{v'(t_0) - \mu_{v'|A}}{\sigma_{v'|A}} \qquad (11)$$

and the coefficients $d_{|A,n}$ are obtained from the following recursive relation [3,9]:

$$d_{|A,n} = (-1)^n \frac{\chi_{|A,n}}{n!} \qquad (12)$$

with:

$$\chi_{|A,0} = 1$$
$$\chi_{|A,1} = \chi_{|A,2} = 0 \qquad (13)$$
$$\chi_{|A,n+1} = \sum_{i=2}^{n}\binom{n}{i}\frac{\lambda_{v'|A,i+1}(t_0)}{\lambda_{v'|A,2}^{(i+1)/2}(t_0)}\chi_{|A,n-i}$$

To compute de bit-error-rate, one must proceed as follow:

1°) compute the coefficients $d_{|A,n}$ for each sequence A={a_{-p}, ..., a_{-1}, a_0, a_1, ..., a_s} by relations (12) and (13).

2°) compute $\Pr[v'(t_0) \mid A]$ by (8), (9) and (10).
3°) follow the same procedure as for the gaussian approximation, i.e. compute (7) and so on. In fact, in (8) the only stochastic terms are the coefficients $d_{|A,n}$, so it is equivalent to compute:

$$d_{|a_0,n} = E_A\left[d_{|A,n}\right] \qquad (14)$$

and write:

$$\Pr\left[v'(t_0)\big|a_0 = \alpha_i\right] = \sum_{n=0}^{+\infty} d_{|a_0,n}\, p^{(n)}(v'(t_0)) \,(15)$$

In practice, to use the Gram-Charlier series expansion, one has to limit the number of terms to N because relations (15) and (8) hold an infinite number of components.

By manipulating the recursive relation (12) and using the properties of the gaussian variable and its derivatives [11], it can be shown [4,10] that relations (8) and (15) are equivalent to multiply $Z(x)$ by a correcting polynomial $R(x)$.

5. Results

The top of the figure 2 shows the probability density functions computed in the binary case by the gaussian approximation (GA) and by the Gram-Charlier series expansion with 10 terms (GCA) whereas the bottom of this figure shows the bit-error-rate versus the decision threshold. In any case, the right part of this figure gives a zoom of the intersection of the two probability density functions of the left part. We can see that the difference between GA et GCA is in the tails of the distributions and that the best decision threshold is slightly different.

Figure 3 shows a comparison between the probability density functions computed by the GA and by the GCA. As we can see, the Gram-Charlier method is equivalent to multiply the gaussian approximation by correction polynomials which depend on the link properties. In this simulation, we can see that the correction polynomials are quite flat in the region where the curves intersect, so the difference between GA and GCA is quite low. This is due to the fact that no optical noise has been taken into consideration.

Figure 4 compares the two methods to compute the bit-error-rate versus the mean received optical power for a link working at 1310 nm, with an APD gain of 40, at 2.4 Gbit/s and with modal noise (or any additive zero mean gaussian noise). In each point the decision threshold is optimized to

minimize the BER and we can see that the GCA method gives a slightly lower BER than the GA method when the power is high and that the BER tends to saturate. So we can say that the very fast GA method gives an upper bound on the BER. Figure 5 shows the results for 2.4 Gbit/s link working at 1550 nm on a fibre optimized for 1300

nm. In that case, the dispersion is taken into account by introducing the fibre impulse response into the computation of the received signal r(t) and assuming that the optical signals along the fiber are gaussian. We can clearly see that the dispersion degrades the BER.
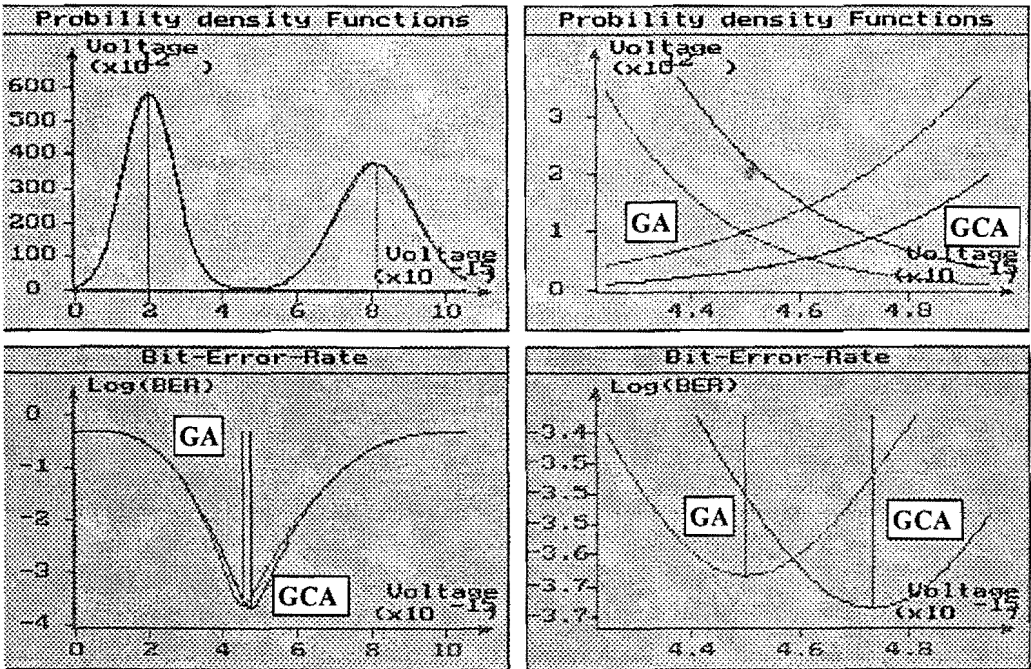
Figure 2: probability density functions and bit-error-rate versus decision threshold
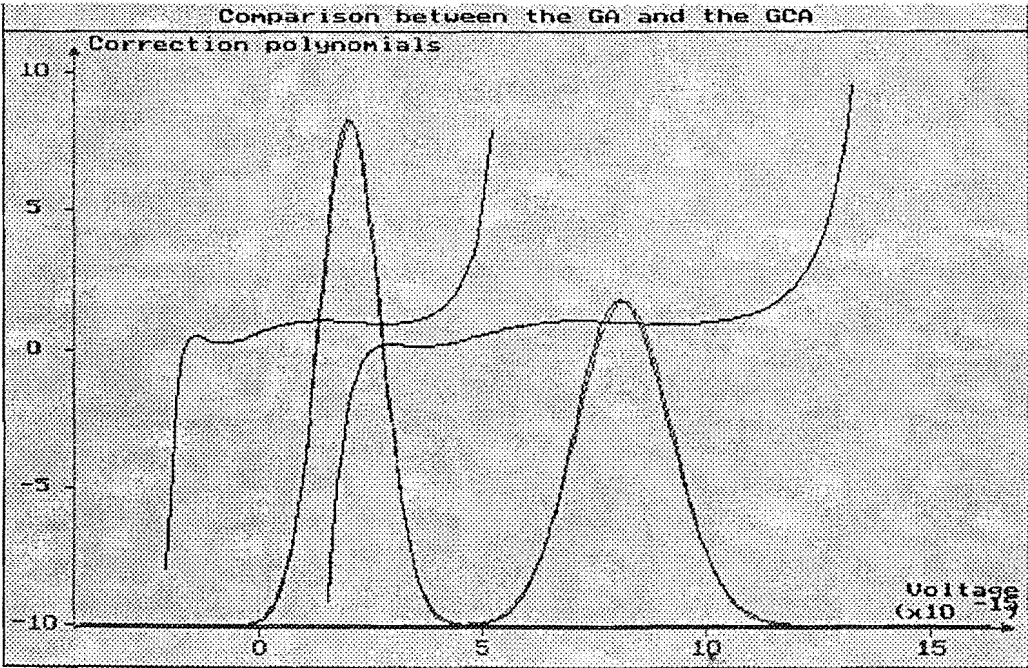
Figure 3: probability density functions
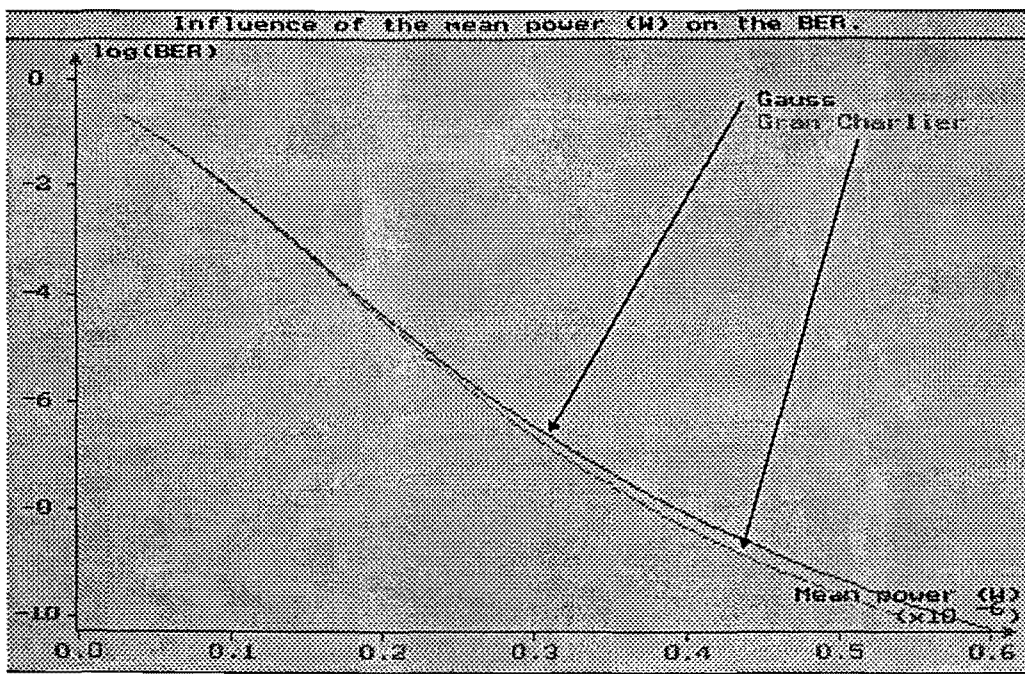
Figure 4: comparison of the bit-error-rate at 2.4 Gbit/s computed by GA and GCA

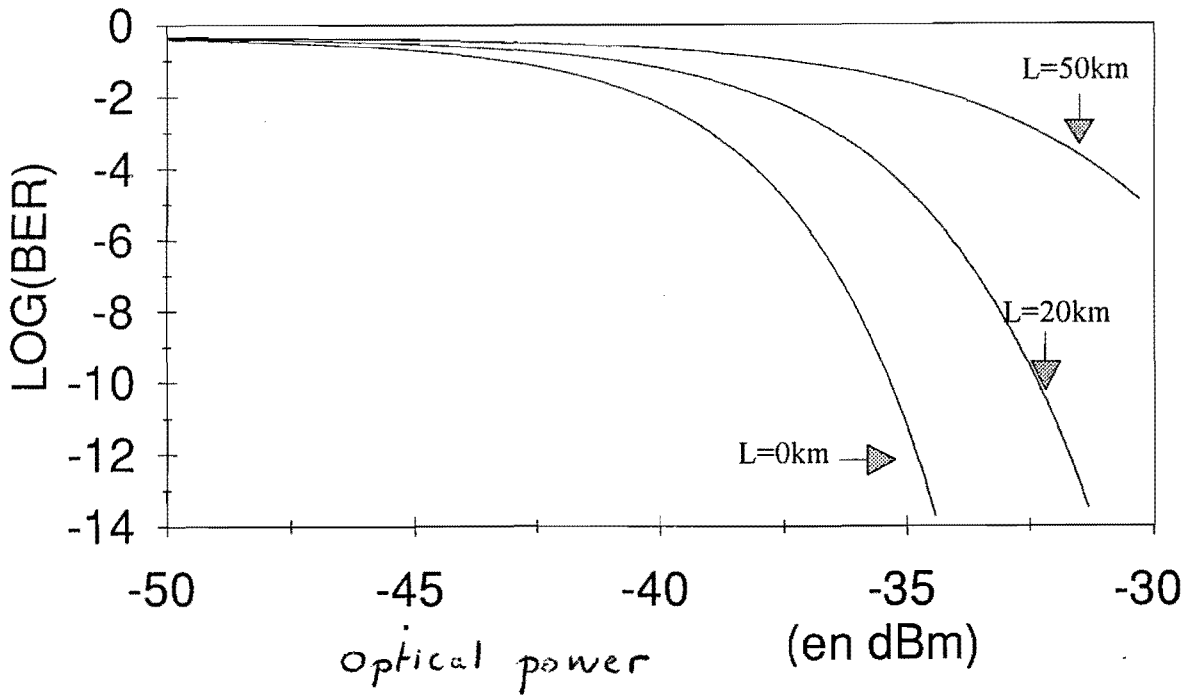APD, G1=50, λ=1550nm, B=2,488Gbit/s, ε=0,1, D=17ps/(km.nm), σ_λ=1nm

Figure 5: bit-error-rate in a link where there is dispersion
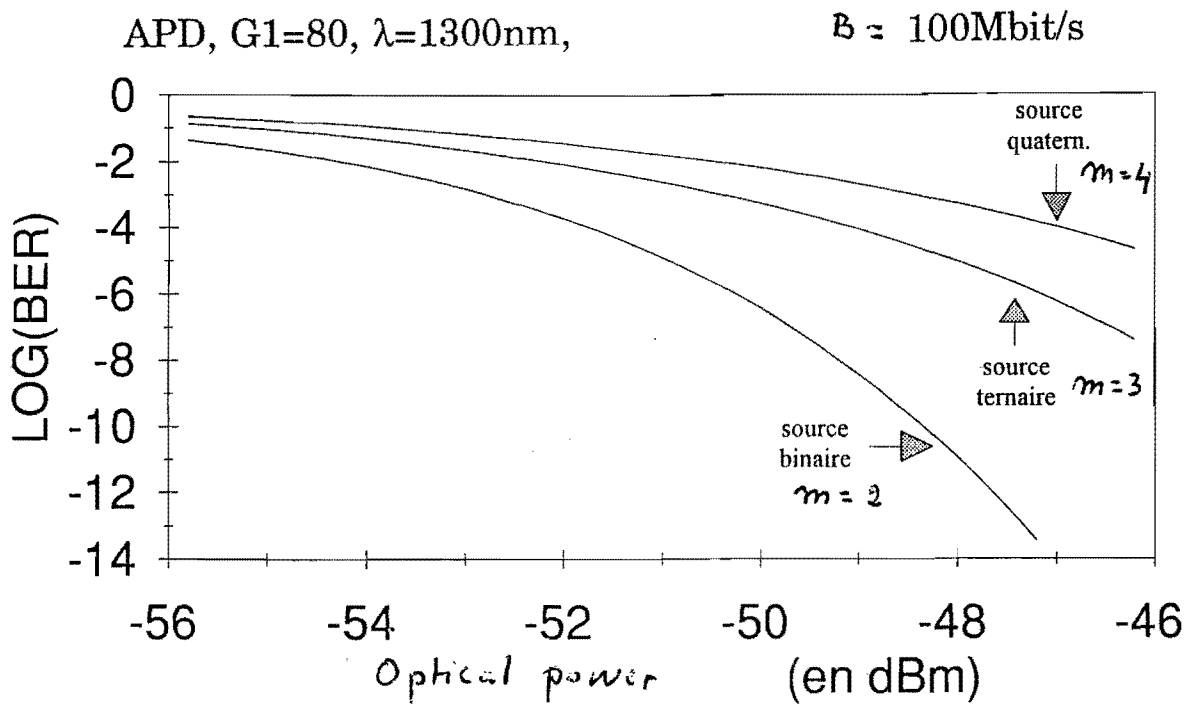
Figure 6: bit-error-rate in m-ary link

In the case of multilevel transmission systems, similar results can be obtained. In particular, m gaussian probability density functions and m correction polynomials are computed and the graphs of figure 2 and 3 simply show m functions. As an example, figure 6 compares the results for binary, ternary and quaternary transmission at 1300 nm and with an APD of mean gain of 80. Of course, when the number m of levels increases, the mean power needed to achieve the same quality increases also.

6. Conclusions

We have compared the gaussian approximation and the Gram-Charlier series expansion in the problem of computing the bit-error-rate of an optical link. The results show that the Gram-Charlier method is equivalent to multiply the gaussian approximation by appropriate correction polynomials whose coefficients can be computed from the link parameters.

It has been shown that the gaussian approximation is faster than the Gram-Charlier method and gives an upper bound on the bit-error-rate.

7. References

[1] " Fundamentals of Photonics "
B.E.A. Saleh, M.C. Teich
Wiley, 1991

[2] " Error Probabilities in Digital Fiber Optic Communication Systems "
G.L. Cariolaro
IEEE Transactions on Information Theory, vol. IT-24, n°2, March 1978, pp. 213-221

[3] " Fiber Optics Receiver Error Rate Prediction Using the Gram-Charlier Series "
M. Mansuripur, J.W. Goodman, E.G. Rawson, R.E. Norton
IEEE Transactions on Communications, vol. COM-28, n°3, March 1980, pp.402-407

[4] " Etude théorique et expérimentale des bruits d'origine optique dans les liaisons à fibres optiques; influence sur les performances; optimisation des paramètres des liaisons "
P. Mégret

PhD Thesis, Faculté Polytechnique de Mons, 1993

[5] " The Phenomenon of Modal Noise in
Analogue and Digital Optical Fibre Systems "
R.E. Epworth
Proceedings of ECOC 1978, pp. 492-501

[6] " Threshold Gain Difference Requirements for
Nearly Single-Longitudinal-Mode Lasers "
J. C. Cartledge, A.F. Elrefaie
IEEE Journal of Lightwave Technology,
vol.8, n°5, May 1990, pp. 704-715

[7] " The Effect of Laser Chirping on Lightwave
System Performances "
J. C. Cartledge, G.S. Burley
IEEE Journal of Lightwave Technology,
vol.7, n°3, May 1989, pp. 568-573

[8] " Random Point Processes in Time and Space "
D. L. Snyder, M.I. Miller
Springer-Verlag, 1991

[9] " Recursive Methods for Numerical
Computation of the Coefficients of a
Generalization of the Gram-Charlier Series "
L. C. Calvez, R. Genin
Proceedings of the IEEE, Auguest 1976

[10] " Le taux d'erreur dans les liaisons optiques
monomodes à hauts débits et longues
distances : étude théorique, analyse statistique
des mesures " O. Ducarme, Travail de fin
d'études, Faculté Polytechnique de Mons,
1992

[11] " Handbook of Mathematical Functions "
M. Abramowitz, I.A. Stegun
Dover publications, 1972

10 Gbit/s Long Haul Soliton versus NRZ Optical Transmission in the 1300nm Window

Robert C.J. Smets, Jean G.L. Jennen

Eindhoven University of Technology, Building EH 12.25, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Abstract—We have compared the operation and applicability of soliton transmission systems versus NRZ (Non Return to Zero) transmission systems for the 1300nm optical window. For a different number of in-line amplifiers the BER (Bit Error Rate) curves of both transmission systems have been simulated. For both systems the same fibre dispersion and attenuation has been assumed and all amplifier gains have been chosen so that the fibre losses are compensated for. In case of the soliton system an extra dispersion shifted fibre has been located directly after the laser diode to unchirp and compress the laserpulses. With respect to the BER simulations, soliton and NRZ transmission systems are expected to be competitive. Transmission exhibiting an error probability below 10^{-9} is attainable for both systems. Under equal circumstances, we conclude from our simulations that soliton systems show better BER performance than NRZ systems beyond 450 km. Owing to the dispersion compensating properties soliton transmission systems are advantageous when long haul transmission is considered. However, NRZ transmission is both a well established and relatively simple technique and can be more readily implemented.

I. INTRODUCTION

Since the introduction of standard single mode fibre (SSMF) more than 55 million kilometers has been installed. At present 2.5 Gbit/s data transmission has become a commercial standard. In the near future the increasing demand for more network capacity can be satisfied by introducing 10 Gbit/s transmision systems. The very low dispersion in the 1300 nm window of SSMF makes a 10 Gbit/s system attractive. Considering the recently achieved progress in the development and use of quantum-well laser amplifiers (QWLAs) [1], [2], [3], attenuation of over 50 km of SSMF can be easily compensated for, making repeaters superfluous. Recently, more and more promising experimental results are being published on NRZ [4] [5] as well as soliton transmission. Even 20 Gbit/s soliton transmission over 200 km [6] has been realised as well as 2 × 10 Gbit/s wavelength division multiplexed NRZ transmission over 63.5 km [5]. As solitons can be multiplexed easily in the time domain it is more likely that for higher bitrates (i.e. ≥ 40 Gbit/s) soliton systems will be prefered to NRZ systems whose maximum bitrate is limited by the maximum achievable electrical bandwidth in current electronics. Although the literature available on soliton as well as NRZ systems is extensive, theoretical study on long
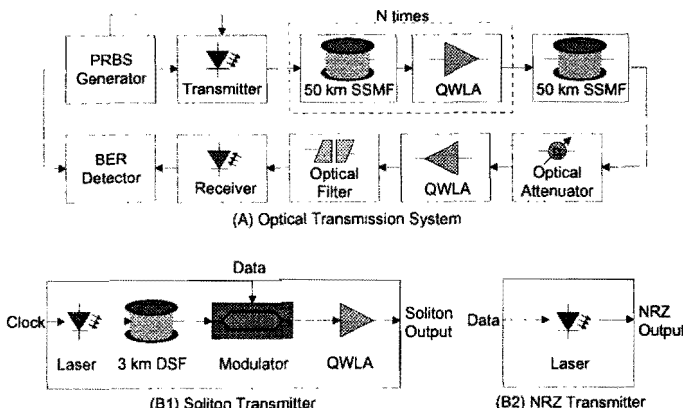

Figure 1: Optical Transmission System (A), Soliton Transmitter (B1), NRZ Transmitter (B2)

haul transmission systems with cascaded QWLAs is not yet complete. The effects of amplifier saturation, ASE and timing jitter on system performance still need investigation. In this paper we compare the performance between a 10 Gbit/s NRZ system [4], [5] and a 10 Gbit/s soliton system theoretically by means of computing bit error rate (BER) curves.

II. TRANSMISSION MODEL

Figure 1 shows the NRZ and soliton transmission system. In the case of NRZ transmission the laser diode is directly modulated with the PRBS signal. The soliton system requires a more complex transmitter. Chirped near-soliton pulses are generated by gain switching of the laser diode and unchirped and compressed in the dispersion shifted fibre after which the optical modulator modulates the data on the soliton train. Finally, the power is boosted by a QWLA up to a level where nonlinear effects compensate dispersion effects in the SSMF. Both soliton and NRZ transmission systems are built up out of eight different components figure 1, namely QWLAs, fibres, a laserdiode, a photodiode, a PRBS generator, a BER detector, a filter and an attenuator. The soliton system needs an additional modulator. Therefore nine models are needed for a theoretical description of both systems.

QWLA model

The model used to describe the QWLAs is based upon the rate equations in [7] and [8], where the z-dependency has been eliminated by integration over the amplifier length L. The ASE is wavelength dependent [4], namely:

$$P_{out}(\tau, \lambda) = P_{in}(\tau, \lambda)e^{h(\tau, \lambda)} + h\frac{c^2}{\lambda^3}\frac{NF}{2}e^{h(\tau, \lambda)}\mathrm{d}\lambda \quad (1)$$

with

$$h(\tau, \lambda) = \int_0^L g(z, \tau, \lambda)\mathrm{d}z , \quad (2)$$

$$\varphi_{out}(\tau, \lambda) = \varphi_{in}(\tau, \lambda) - \frac{1}{2}\alpha_H h(\tau, \lambda) \quad (3)$$

and

$$\frac{\partial h(\tau, \lambda)}{\partial \tau} = -\frac{h(\tau, \lambda) - g_0(\lambda)L}{\tau_c}$$
$$-\frac{\sum_{\lambda_i}\frac{\lambda_i}{\lambda}P_{in}(\tau, \lambda_i)/P_{sat}(\lambda)}{\tau_c}\cdot\left(e^{h(\tau, \lambda)} - 1\right) , \quad (4)$$

where τ_c is the carrier lifetime, $P_{sat}(\lambda)$ represents the $1/e$ saturation output power, $g_0(\lambda)L$ represents the small signal amplifier gain, $h(\tau, \lambda)$ is the time dependent amplifier gain, α_H is the linewidth enhancement factor, $d\lambda$ is the wavelength discretisation stepsize, h is Planck's constant, c stands for the velocity of light, NF is the fibre coupled noise figure, and $P_{in}(\tau, \lambda_i)$ represents the total inputpower to each amplifier, i.e. the sum of input signal power and ASE originating from the previous amplifiers. τ is the reduced time with respect to a reference plane moving with the signal ($\tau = t - \frac{z}{v_g}$, with v_g the group velocity). The amplifier is assumed to have zero reflecting input/output facets and to be polarization independent. The gain is assumed to have a Gaussian shaped spectrum. The equations are numerically solved using *Euler forward* iteration.

Fibre model
Propagation in the fibre is discribed by the nonlinear Schrödinger equation [9],

$$\frac{\partial A}{\partial z} = -\frac{\alpha}{2}A - \frac{i}{2}\beta_2\frac{\partial^2 A}{\partial \tau^2} + i\gamma|A|^2 A, \quad (5)$$

where A is the slowly varying envelope of the optical signal, α represents the losses in the fibre, γ is the nonlinearity parameter and $\beta_2 = -D\frac{\lambda^2}{2\pi c}$ is the dispersion parameter. D is the dispersion. Eq. 5 is numericaly solved using the *split-step Fourier* method.

Laser diode model
The model employed is a simple Fabry Perot model

where the z-dependence has been eliminated [10]. Laser pulses are generated by gain switching of the laser diode. After transmission through a dispersion shifted fibre, the pulses closely resemble soliton pulses. In the case of NRZ modulation the signal shows an extinction ratio and slopes equal to experimentally observed values.

Photodiode model
The optical signal is converted to the electrical domain by the photodiode as discribed by [11]. A quantum efficiency of 0.8 is taken. The model includes shot noise, signal-spontaneous beat noise, spontaneous-spontaneous beat noise and thermal noise. The electrical bandwidth is 10.2 Ghz and the thermal noise is adjusted, so that the receiver has a senstivity of -13.7 dBm at 10 Gbit/s NRZ and a BER of 10^{-9}, in agreement with measured data (figure 2).

Optical filter
The optical filter is assumed to be square and sufficiently broad with respect to the signals bandwidth. The optical filter reduces the ASE by allowing only a small fraction of the ASE spectrum to pass. The effective bandwidth of the filter is 2 nm and the filter introduces a loss of 2 dB.

Optical attenuator
The optical attenuator is used primarily to attenuate the signal in order to compute BER values for different receiver input powers.

Optical modulator
The optical modulator has only functionallity in the soliton system. Its purpose is to modulate the soliton train. When the modulator is excited by a logical one, a raised-cosine window is opened letting a soliton pass. When a logical zero is applied, a window of smaller amplitude is opened, attenuating the soliton. The optical modulator introduces a loss of 5 dB. The roll-off factor of the raised cosine function is $\beta = T_b/2$, with T_b the bit time. The extinction ratio i.e. the ratio between the peak power of a logical one and a logical zero is equal to $ER = 20$.

PRBS generator and BER detector
The PRBS generator generates a $2^7 - 1$ pseudo-random bit sequence. The BER detector computes the optimum BER for a number of detection thresholds and sampling times, assuming a gaussian distribution for all noise sources. The BER is obtained by averaging over all $2^7 - 1$ bits. We emphasize that the clock of the PRBS generator is used to sample the received data and that there is no clock recovery at the receiver.
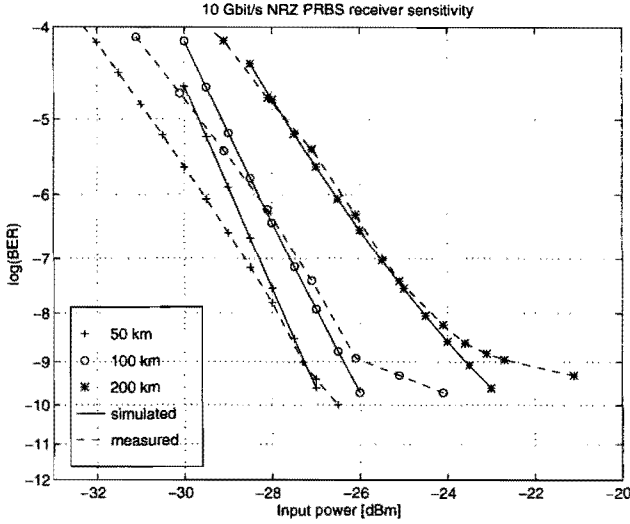
Figure 2: Measured and simulated BER curves in absence of clock recovery, $ER = 6$
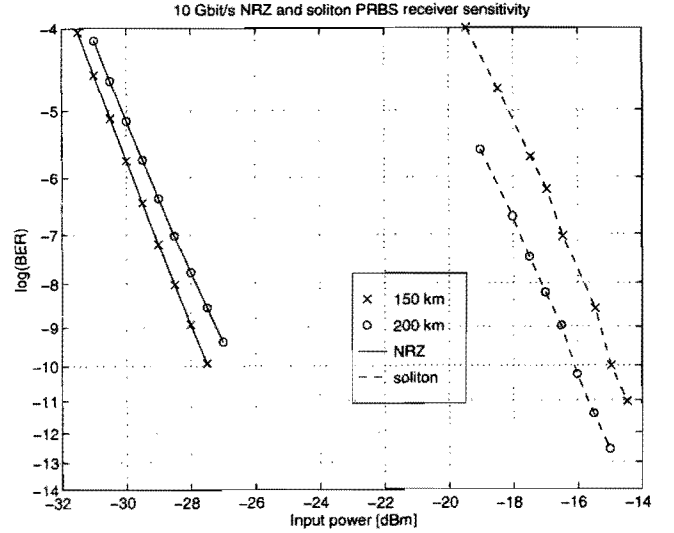

Figure 3: Simulated NRZ and soliton BER curves in absence of clock recovery, $ER = 20$
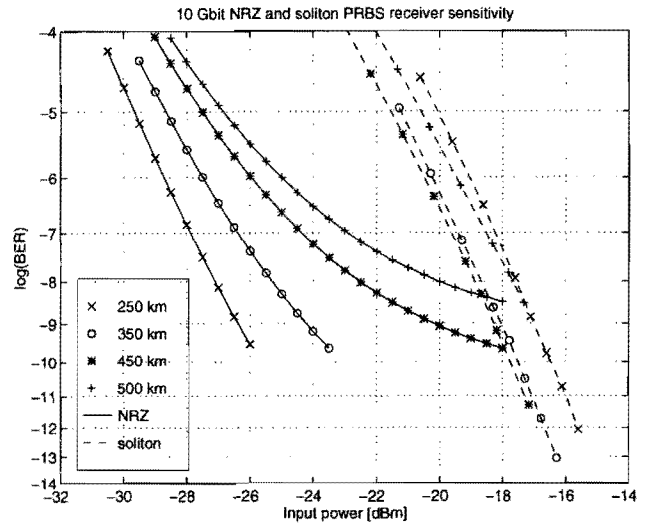
Figure 4: Simulated NRZ and soliton BER curves in absence of clock recovery, $ER = 20$

III. COMPUTER SIMULATIONS

At 1300 nm the 50 km SSMF sections exibit a loss of $\alpha = 0.4$ dB/km and a dispersion of $D = 0.6$ ps/km/nm and a nonlinearity parameter of $\gamma = 2.1$ W/km. The 3 km dispersion shifted fibre exibits a loss of $\alpha = 0.4$ dB/km and a dispersion of $D = -17$ ps/km/nm. The QWLAs are adjusted to a gain of 22.5 dB at the signal wavelength of $\lambda = 1310$ nm, almost equal to the loss of the 50 km fibre span. A 60 nm gain bandwidth was assumed. The fibre to fibre noise figure of the amplifier was estimated at 9 dB. For the 3 dB saturation output power $P_{sat,3dB} = 10$ dBm was assumed. The linewidth enhancement factor was estimated at $\alpha_H = 5$. For the gain-recovery time $\tau_c = 200$ ps was taken. Using the NRZ or soliton transmitter a power of 2 dBm or 6 dBm respectively, is launched into the first fibre section. All computer simulations were carried out with the same algorithms. The only difference between the NRZ and soliton simulations is the number of sample points per bit. Because the phase of the solitons is of major importance, 256 sample points per bit were used for soliton simulations while 20 sample points sufficed in the case of NRZ simulations. Figure 2 shows measured and simulated NRZ BER-curves with an extinction ratio of $ER = 6$. The good agreement between the measured and theoretical curves validates our model.

IV. DISCUSSION

Figure 3 depicts the simulated BER-curves for NRZ and soliton transmission over 150 km and 200 km. The higher receiver input power of the soliton signal is due to the amplifier in the soliton transmitter,

therefore better BER performance is expected. Additional simulations reveal that this different behaviour can be explained by the absence of clock recovery. As solitons are sensitive to timing jitter caused by saturation and soliton interaction, timing jitter is expected to be the most likely cause. In contrary to NRZ transmission it can be clearly seen that in the case of soliton transmission at 200 km a better BER is obtained with less input power than at 150 km. The same result is obtained for longer transmission links, presented in figure 4. However, after 450 km no further improvement is observed. Additional simulations show that the improvement of the soliton signal could be explained by lesser non-linear effects in the fibre caused by saturation of the amplifiers resulting in a broadened pulse which is less sensitive to timing jit-
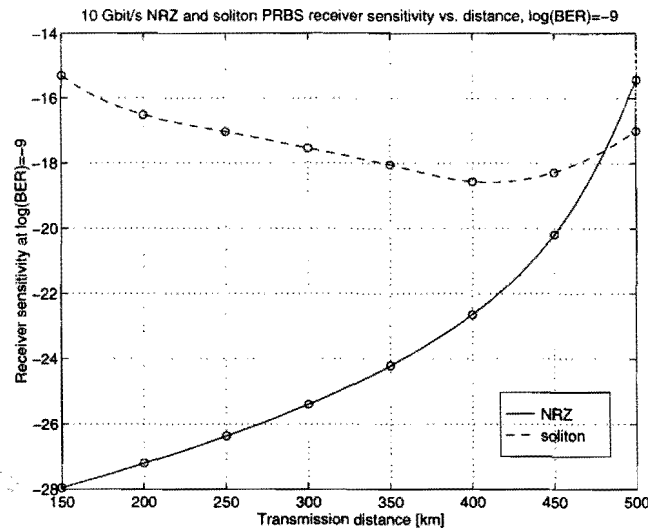
Figure 5: Receiver sensitivity versus transmission distance at $BER = 10^{-9}$

ter, resulting in a better BER. The degradation after 450 km is probably caused by less dispersion compensation due to diminishing output power caused by saturation of the amplifier by ASE. By placing an additional ASE-filter at 450 km we expect to increase the maximum transmission distance at a $BER = 10^{-9}$. In contrary to NRZ transmission soliton transmission simulations did not reveal BER floors however. Figure 5 shows that the receiver sensitivity of the NRZ system at $BER = 10^{-9}$ in contrary to the soliton system is strongly dependent on the transmission distance due to built up ASE. Only after 450 km the soliton system shows a better sensitivity. When a BER of 10^{-12} is required, a soliton system is prefered when more than 350 km needs to be bridged.

V. CONCLUSION

A transmission model has been presented for simulation of NRZ and soliton systems. With respect to the simulated BER curves we expect soliton transmission to be superiour to NRZ when long haul transmission systems beyond 450 km are considered. Timing jitter caused by soliton interaction and saturation of the QWLAs is expected to seriously degrade the performance of soliton transmission systems and therefore needs further investigation. The relatively complex transmitter makes NRZ systems preferable to soliton systems.

VI. ACKNOWLEDGEMENT

REFERENCES

[1] L.F. Tiemeijer, et al., "Polarization Insensitive Multiple Quantum Well Lase Amplifiers for the 1300 nm Window," *Applied Physiscs Letters*, vol. 8, pp. 826–828, February 1993.

[2] L.F. Tiemeijer, et al., "33 dB Fiber to Fiber Gain +13 dBm Fiber Saturation Power Polarization Independent 1310 nm MQW Laser Amplifiers," in *Technical Digest Topical Meeting on Optical Amplifiers and Applications, Post-deadline*, pp. 1–5, June 1995.

[3] L.F. Tiemeijer, *Optical Properties of Semiconductor Lasers and Laser Amplifiers for Fiber Optical Communication*. PhD thesis, Delft University of Technology, The Netherlands, December 1992.

[4] J.G.L Jennen et al., "Performance of Cascaded 1300 nm QW Laser Amplifiers in 10 Gbit/s Long Haul NRZ Transmission," in *Technical Digest Topical Meeting on Optical Amplifiers and Applications*, pp. 123–126, June 1995.

[5] H. de Waardt, *High Capacity 1300 nanometer Optical Transmission*. PhD thesis, Delft University of Technology, The Netherlands, September 1995.

[6] J.J.E. Reid, C.T.H.F. Liedenbaum et al., "Realisation of 20 Gbit/s Long Haul Soliton Transmission at 1300 nm on Standard Single Mode Optical Fibre," in *Proceedings 20th European Conference on Optical Communication / Post-deadline Papers*, pp. 61–64, September 1994.

[7] G.P. Agrawal and N.A. Olsson, "Self-Phase Modulation and Spectral Broadening of Optical Pulses in Semiconductor Laser Amplifiers," *IEEE Journal of Quantum Electronics*, vol. 25, pp. 2297–2306, November 1989.

[8] G.P. Agrawal and N.A. Olsson, "Amplification and Compression of Weak Pico Second Optical Pulses by Using Semiconductor-Laser Amplifiers," *Optics Letters*, vol. 14, pp. 500–502, May 1989.

[9] G.P. Agrawal, *Nonlinear Fiber Optics*. Academic Press, Inc., 1989. ISBN 0-12-045140-9.

[10] G.P. Agrawal and N.K. Dutta, *Long-Wavelength Semiconductor Lasers*. Van Nostrand Reinhold Company Inc., 1986. ISBN 0-442-20995-9.

[11] N.A. Olsson, "Lightwave Systems with Optical Amplifiers," *IEEE Journal of Lightwave Technology*, vol. 7, pp. 1071–1082, July 1989.

190

Comparison between semiconductor and fibre optical amplifiers operating in the 1300 nm telecommunications window

Jean G.L. Jennen, Twan W.H. van Osch
Eindhoven University of Technology, Building EH 12.25, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Dennis R. Simons
Eindhoven University of Technology, Building TNO 0.11, P.O. Box 595, 5600 AN Eindhoven, The Netherlands

Abstract— We have compared the operation of semiconductor and fibre optical amplifiers for the 1300 nm optical window. The key parameters are considered and evaluated with regard to different amplifier applications in telecommunications systems. We have studied representative data of both amplifier types, i.e. characteristics of indium gallium arsenide strained quantum well laser amplifiers (QWLA's) and Pr^{3+}-doped germanium gallium sulphide fibre amplifiers. Concerning the QWLA's, use of both compressively and tensile strained quantum wells allow for a high polarisation insensitivity while maintaining a high gain. For these amplifiers experimental as well as modelling data are available. Looking at the fibre amplifiers, Ge-Ga sulphide glass is a promising host glass material for a state-of-the-art device due to its low phonon energy. Optical properties of bulk sulphide glass already allow for modelling of the fibre amplifier characteristics.

In general, semiconductor and fibre amplifiers exhibit competitive performances in terms of gain, saturation power and polarisation insensitivity. In WDM and analogue CATV systems implementation of fibre amplifiers is expected to be more attractive than semiconductor amplifiers, mainly due to the low carrier lifetime ($\approx$200 ps) of the latter amplifier type. Nevertheless, this application is still under investigation. On the other hand, the low carrier lifetime provides the opportunity of wavelength conversion within the semiconductor amplifier.

I. INTRODUCTION

At present more than 90% of the terrestrial fibre network is based on standard single mode fibre (SSMF), showing zero dispersion at a wavelength of about 1310 nm. In order to protect the huge investments already made an upgrade of the existing systems to higher bitrates (10 Gbit/s and beyond) is desirable. When operating in the 1300 nm window and applying SSMF, a high bitrate × distance product could be achieved, if it were not for the relatively large attenuation of the fibre. This draw-back can be overcome by the implementation of optical amplifiers in future networks.

As opposed to the development in the 1550 nm range, where the Erbium Doped Fibre Amplifier (EDFA) clearly has left the competition far behind, the competition between semiconductor and fibre amplifiers in the 1300 nm window still has not been resolved. Therefore, a comparison between state-of-the-art devices of both types is helpful in determining their applicability. Questions that arise are which type of amplifier will show the best performance in due course and does it depend on application. In order to give the initial impetus to the answers we firstly describe the different amplifier types and corresponding characteristics in section II. Then, the system performance of the optical amplifiers when applied in various systems is discussed in section III. In Section IV we will come to some answers concerning the questions mentioned above. Finally, conclusions are drawn in section V.

II. DEVICE CHARACTERISTICS

In this section we take a closer look at the fundamental parameters of both amplifier types. We have extracted our data from literature, simulations as well as experimental results. The following subsections will give an insight in the currently potential amplifier performance of each type separately. In order to obtain a clear picture an overview is presented at the end of this section.

A. Quantum well laser amplifiers

The operation of QWLA's is described by the rate equations presented in [1]. After elimination of z by integration over the amplifier length L and introduction of wavelength dependency we arrive at the following reduced rate equations, namely

$$P_{out}(\tau, \lambda) = P_{in}(\tau, \lambda)e^{h(\tau,\lambda)} +$$

$$+ h\frac{c^2}{\lambda^3}\frac{NF}{2}e^{h(\tau,\lambda)}\mathrm{d}\lambda \qquad (1)$$

with $h(\tau, \lambda) = \int_0^L g(z, \tau, \lambda)\mathrm{d}z$,

$$\varphi_{out}(\tau, \lambda) = \varphi_{in}(\tau, \lambda) - \frac{1}{2}\alpha_H h(\tau, \lambda) \quad \text{and} \quad (2)$$

$$\frac{\partial h(\tau, \lambda)}{\partial \tau} = -\frac{h(\tau, \lambda) - g_0(\lambda)L}{\tau_c} +$$

$$- \frac{\sum_{\lambda_i}\frac{\lambda_i}{\lambda}P_{in}(\tau, \lambda_i)/P_{sat}(\lambda)}{\tau_c}\left(e^{h(\tau,\lambda)} - 1\right), \qquad (3)$$

where τ_c is the carrier lifetime, $P_{sat}(\lambda)$ represents the saturation output power, $g_0(\lambda)L$ stands for the small signal amplifier gain, $h(\tau, \lambda)$ is the time dependent amplifier gain, α_H is the linewidth enhancement factor, $\mathrm{d}\lambda$ represents the wavelength discretisation stepsize, h is the Planck constant, c stands for the velocity of light, NF is the fibre coupled noise figure, $\varphi_{in}(\tau, \lambda)$ and $\varphi_{out}(\tau, \lambda)$ represent the phase of the input and output signal respectively, $P_{out}(\tau, \lambda)$ is the amplifier output power, and $P_{in}(\tau, \lambda_i)$ represents the total input power to each amplifier, i.e. the sum of input signal power and ASE originating from the previous amplifiers. When near zero facet reflectivity is assumed the gain may be represented as a Gaussian shaped spectrum.

$$g_0(\lambda)L = \frac{\ln 10}{10}\left\{g_p - 12\left(\frac{\lambda - \lambda_p}{\Delta\lambda}\right)^2\right\} , \qquad (4)$$

where g_p represents the peak gain in dB at wavelength λ_p and $\Delta\lambda$ stands for the amplifier 3 dB bandwidth.

In order to solve these equations numerically measured or estimated values for small signal gain, peak wavelength, saturation output power, optical bandwidth, noise figure, and carrier lifetime have to be provided. Residual polarisation sensitivity can be taken into account by solving

the Rate Equations 1, 2 and 3 applying different parameter sets for TE and TM gain respectively, assuming all gains are saturated equally by TE and TM intensities.

From literature [2, 3, 4] and additional measurements we have derived up to date values for the fundamental parameters mentioned above. Using pump currents of 400-500 mA fibre coupled gains up to 33 dB have been recorded together with 3 dB saturation output powers of more than 13 dBm. Furthermore, saturated output powers of 16 dBm are demonstrated. When no high gain is required 3 dB saturation output powers of even 18 dBm can be achieved, whereas fibre coupled output powers exceed 20 dBm [5]. The wavelength at which the peak gain is experienced can vary within the range of 1280-1330 nm. The amplifier modules exhibit a 3 dB optical bandwidth of approximately 60 nm and a carrier lifetime of about 200 ps. Residual polarisation sensitivity is reduced to 0.5 dB. Noise figures as low as 6.5 dB have been shown, however at record gain a noise figure of 9 dB is expected.

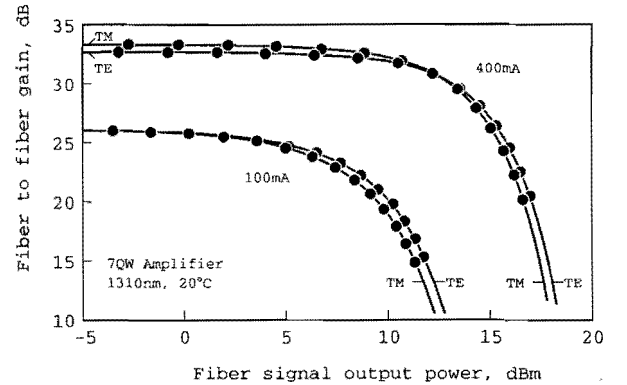Some of the amplifier characteristics mentioned above are illustrated by Figures 1 and 2.

Figure 1: Polarisation resolved fibre to fibre gain versus output power, at 400 mA 33 dB fibre to fibre gain is obtained for both polarisations (after ref. [5]).

B. Pr^{3+}-doped fibre amplifiers

Until today, most Pr^{3+}-doped fibres were based on ZBLAN fluoride glasses. Unfortunately, these glasses have a limited quantum efficiency. An efficiency of 3.4 % was reported in [6]. Therefore, there is considerable interest in the development
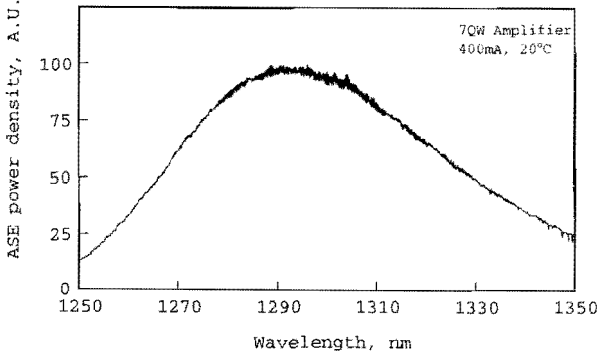
Figure 2: Amplified Spontaneous Emission spectrum recorded at 400 mA drive current (after ref. [5]).

of low-phonon energy Pr^{3+}-doped glasses for improved efficiency. We show that germanium gallium sulphide is a very promising host glass for efficient Pr^{3+}-doped fibre amplifiers.
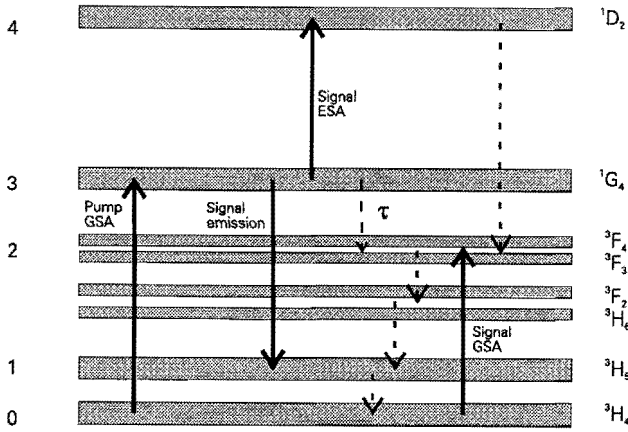

Figure 3: Energy levels of Pr^{3+} in a glass host.

The energy level diagram of Pr^{3+} in a host material is depicted in Figure 3. Various electron transitions must be considered to understand the operation of the PDFA. Some of these transitions are required to obtain gain, while others are undesirable. Firstly, the amplifier must be pumped by a laser, exciting electrons from the $^3H_4$ ground state to the $^1G_4$ meta-stable state. The optimum pump wavelength is about 1020 nm, depending on the glass composition. The $^1G_4$-$^3H_5$ transition can be induced by signal photons in the 1300 nm wavelength region. This stimulated emission is responsible for the gain. From the $^3H_5$ level the electrons rapidly

fall non-radiatively to the $^3H_4$ ground state.

The main problem of PDFA's used to be the limited spontaneous emission lifetime τ of the meta-stable level. In glasses with a relatively high phonon-energy (ZBLAN) the majority of pumped electrons will be lost by multi-phonon relaxation. This results in a spontaneous emission lifetime as low as 110 μs in ZBLAN. Further losses are due to signal ground-state absorption (GSA) and signal excited-state absorption (ESA). These two mechanisms are most harmful at the long-wavelength tail of the 1300 nm window.

To obtain accurate results in all operating regimes of a PDFA we must resort to a numerical model which evaluates both forward and backward propagating ASE across the whole 1300 nm window. In order to find the optimum fibre geometry, the profiles of the optical modes must be included in the calculation. The transition rate W_{IK} between any two levels I and K can be expressed as [7]

$$W_{IK}(r,z) = \int_0^\infty \sigma_{IK}(\nu)\frac{I(\nu,r)}{h\nu}P(\nu,z)\mathrm{d}\nu , \quad (5)$$

where h is the Planck constant and $\sigma_{IK}(\nu)$ is the cross-section that is involved. $P(\nu,z)$ is the optical power spectral density in W/Hz over the fibre cross-section. $I(\nu,r)$ is the transverse intensity distribution of the optical mode at frequency ν, normalised to

$$2\pi \int I(\nu,r) r \mathrm{d}r = 1 . \quad (6)$$

Equation 5 is used to evaluate the pump absorption rate W_{03}, signal excited-state absorption rate W_{34} and the signal stimulated emission rate W_{31}.

Once these transition rates are known, the steady state population densities can be calculated using

$$\eta_3(r,z) = \rho(r) \times$$

$$\times \frac{W_{03}(r,z)}{W_{03}(r,z) + W_{34}(r,z) + W_{31}(r,z) + \frac{1}{\tau}} \quad (7)$$

193

and

$$\eta_0(r, z) = \rho(r) - \eta_3(r, z) , \qquad (8)$$

where $\eta_3(r, z)$ and $\eta_0(r, z)$ are the densities of Pr^{3+}-ions in the ${}^1G_4$ and ${}^3H_4$ states, respectively. $\rho(r)$ is the total density of Pr^{3+}-ions.

In [8] propagation equations are given for PDFA's. The propagation of the pump, signal and ASE beams is expressed as

$$\frac{d P_p^{\pm}(z)}{d z} = \mp [g_p(z) + \alpha(\nu_p)] P_p^{\pm}(z) , \qquad (9)$$

$$\frac{d P_s(z)}{d z} = [g_e(z, \nu_s) - g_a(z, \nu_s) + \\ -g_g(z, \nu_s) - \alpha(\nu_s)] P_s(z) \qquad (10)$$

and

$$\frac{d P_{ase}^{\pm}(z, \nu)}{d z} = \pm [g_e(z, \nu) - g_a(z, \nu) + \\ -g_g(z, \nu) - \alpha(\nu)] P_{ase}^{\pm}(z, \nu) \qquad (11)$$

$$\pm 2h\nu\Delta\nu g_e(z, \nu) ,$$

where $\Delta\nu$ is the bandwidth of the wavelength section being considered. The pump ground-state absorption, signal/ASE stimulated emission, signal/ASE excited-state absorption and signal/ASE ground-state absorption factors are given by

$$g_p(z) = 2\pi\sigma_{03}(\nu_p) \int_0^a \eta_0(r, z) I(r, \nu_p) r d r , \qquad (12)$$

$$g_e(z, \nu) = 2\pi\sigma_{31}(\nu) \int_0^a \eta_3(r, z) I(r, \nu) r d r , \qquad (13)$$

$$g_a(z, \nu) = 2\pi\sigma_{34}(\nu) \int_0^a \eta_3(r, z) I(r, \nu) r d r \qquad (14)$$

and

$$g_g(z, \nu) = 2\pi\sigma_{02}(\nu) \int_0^a \eta_0(r, z) I(r, \nu) r d r , \qquad (15)$$

where a is the fibre core radius. The background loss $\alpha(\nu)$ is expressed in m^{-1}.

Using the optical properties of bulk glasses (cross-sections and lifetime) [9], the optimum structure of Pr^{3+}-doped sulphide fibres was determined [8]. A cut-off wavelength of 800 nm was chosen and a numerical aperture of 0.3. Using these optimised parameters the amplifier gain under various circumstances has been evaluated.
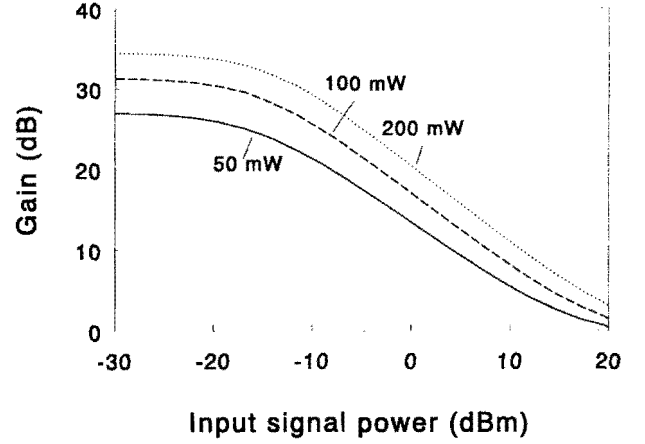

Figure 4: Gain versus input signal power for different pump powers.

In Figure 4 the gain versus input signal power is shown. We assumed a 500 ppmwt Pr^{3+}-doped fibre with a background loss of 0.1 dB/m. The fibre length is optimised for each pump power. The signal wavelength is 1310 nm. A gain of more than 25 dB is expected at 50 mW pump power. This pump power can be delivered by a single laser diode. For even higher gains a bidirectional pump configuration with up to 4 pump lasers can be used. Depending on pump power 3 dB saturation output powers of 10-18 dBm and saturated output powers of 13-20 dBm are obtained.

No noise figure data on sulphide fibre is available yet, but it is expected that it will be comparable to ZBLAN fibres, where values as low as 3.5 dB have been reported [6].
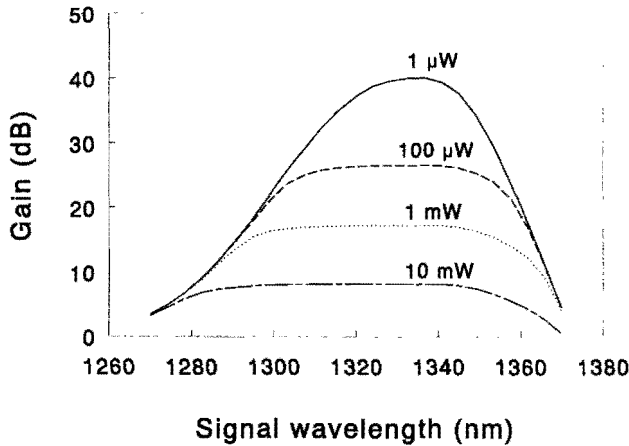
Figure 5: Gain versus input signal wavelength for different input signal powers.

In Figure 5 the gain versus signal wavelength is depicted for input signal powers ranging from 1 μW up to 10 mW using a pump power of 100 mW. The maximum small-signal gain is obtained at approximately 1335 nm. If the input signal power to the amplifier is increased, the gain curve becomes increasingly flat.

C. Overview

The information presented in the previous subsection is summarized in Table 1. Concerning the PDFA data one should note that most of the values given result from reliable simulations. Although the proposed PDFA is not commercially available yet, a lot of effort is put into the development of a method to draw a fibre from the promising host material.

III. SYSTEM PERFORMANCE

A diversity of applications is distinguished when the usage of optical amplifiers is concerned. In this paper we consider a limited subset of applications in both digital and analogue networks. We also show that amplifier nonlinearities can be essential assets for certain applications.

A. Long haul transmission

Short-term applications of 1300 nm optical amplifiers can be found in high bitrate long haul transmission systems. We distinguish three possible locations of the amplifiers in the link as can

Table 1: Key parameters of quantum well laser amplifiers and Pr^{3+}-doped fibre amplifiers.

Parameter	QWLA	PDFA
Small signal peak gain	> 30 dB	> 40 dB
Peak gain wavelength	1280-1330 nm	$\approx$ 1335 nm
Optical bandwidth	$\approx$ 60 nm	$\approx$ 25 nm
Small signal gain at 1310 nm	> 30 dB	> 30 dB
3 dB Saturation output power	13-18 dBm	10-18 dBm
Saturated output power	16-20 dBm	13-20 dBm
Noise Figure	6.5-9 dB	$\approx$ 4 dB
Inversion lifetime	$\approx$ 200 ps	$\approx$ 350 μs

be seen in Figure 6 where an example of a transmission link is presented. The optical amplifiers can either be used as booster, e.g. in a soliton transmitter, as preamplifier in order to increase the receiver sensitivity, or as in-line amplifier to compensate the fibre losses of the previous section.

The feasibility of 10 Gbit/s NRZ optical transmission over 200 km of fibre using QWLA repeaters has been demonstrated in [10]. Depending on the extinction ratio of the input signal a transmission length of even 450 km is attainable. When we take into account the information of Table 1 a better performance is expected with the implementation of PDFA's instead of QWLA's owing to the smaller noise figure. When only considering application as booster no substantial difference in performance is expected between the two types of amplifiers, because the
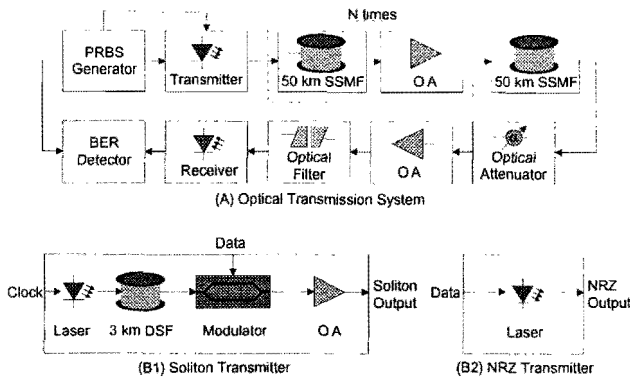
195

Figure 6: Long haul transmission system (A), soliton transmitter (B1) and NRZ transmitter (B2).

noise figure is of less importance in that case.

In WDM systems implementation of fibre amplifiers is favoured since no cross-gain modulation penalties will occur contrary to the use of semiconductor amplifiers. Due to the fact that the inversion lifetime of the latter amplifiers is in the same order of magnitude as the bit period of the optical signals, modulation of the gain spectrum will occur. Thus, interaction between any two signals will cause additional penalties in receiver sensitivity. On the other hand, the optical bandwidth and the peak gain wavelength of QWLA's are better matched with WDM system requirements, i.e. channel wavelengths around 1307 nm and a flat gain spectrum over a large bandwidth.

B. Analogue CATV distribution

In CATV distribution systems optical fibre is used for the transmission of analogue subcarrier-modulated video signals. Great interest is shown in optical amplifiers for the 1300 nm window to upgrade these systems and to allow for high splitting ratios. Optical amplifiers must be highly linear for this application to avoid distortion. Due to the low carrier lifetime QWLA's will introduce distortion when they are driven into saturation. The spontaneous emission lifetime of the PDFA is large enough to avoid this undesired effect, making it the preferred amplifier for analogue CATV systems. The tilt in the gain versus wavelength characteristics can however lead to nonlinear distortion due to the chirping of the laser.

C. Wavelength conversion

Nonlinearities like four wave mixing and cross-gain modulation can be exploited in order to accomplish wavelength conversion. The effect of four wave mixing occurs in both amplifier types when two or more optical carriers are launched into the amplifier at sufficiently high power levels. In case two input signals at the wavelengths λ_1 and λ_2 are used the optical output spectrum also shows signal components at the wavelengths $2\lambda_1 - \lambda_2$ and $2\lambda_2 - \lambda_1$. Thus, up as well as down conversion is possible using the effect of four wave mixing. However, employment of this effect is only suitable for conversion over a few nanometers and input powers of more than 100 μW are required [11].

Another nonlinear effect that can be used to achieve wavelength conversion is cross-gain modulation. Since practical signal rates are well above 5 kbit/s this effect is only noticeable in semiconductor optical amplifiers owing to the relatively small inversion lifetime of about 200 ps. In order to obtain signal transformation of one wavelength to another a continuous wave signal at the target wavelength together with the original input signal are coupled into a semiconductor amplifier. By means of cross-gain modulation the input signal information is copied to the target wavelength. Advantages of this method over four wave mixing are the smaller input powers required (10-20 μW) and the possibility of wavelength conversion over the entire optical bandwidth of the amplifier.

IV. DISCUSSION

According to Table 1 no striking differences in performance between the two amplifier types are expected when considering operation at practical wavelengths near 1310 nm. The only remarkable distinction is found in the inversion lifetime of both types. Therefore, this parameter will constitute a crucial factor in the determination of the best amplifier performance in system environments. As is shown in the previous section a relatively small inversion lifetime of QWLA's in combination with the effect of gain saturation can indeed deteriorate optical signals in certain applications like WDM transmission

196

and analogue CATV distribution. On the other hand, exploitation of the occurring dynamic gain modulation to our advantage leads to a simple design for wavelength conversion.

Clearly, a standard semiconductor optical amplifier is unsuitable for application in analogue AM-CATV transmission systems. However, the development of a so-called gain clamped laser amplifier [12] offers new possibilities for highly linear amplification of intensity modulated signals in the currently used CATV band.

At the moment, the PDFA described in this paper has not reached the stage of becoming commercially available yet. So, the QWLA has taken a perfect head start over the fibre amplifier competition. Nevertheless, ongoing research and developments can certainly challenge the present position of QWLA's. We expect that in the near future both amplifier types will coexist each with their specific applications.

V. CONCLUSIONS

In conclusion, we have shown that in general semiconductor and fibre amplifiers exhibit competitive performances in terms of gain, saturation power and polarisation insensitivity. Furthermore, we have demonstrated that the choice whether to apply a semiconductor or fibre amplifier in the 1300 nm window strongly depends on the intended application. Although the semiconductor amplifier may not be the most suitable choice in all possible fields of operation, commercial considerations like costs and present availability of the amplifiers can tip the balance in favour of the QWLA. The main advantages of QWLA's over PDFA's are their compactness and simplicity (only one current source required), their cost-effectiveness (no external pump lasers and expensive fibre necessary), and their availability (no Ge-Ga sulphide fibre could be drawn yet). However, when a breakthrough in the Ge-Ga sulphide fibre draw technology is attained serious competition of the PDFA is to be expected, especially in the area of analogue CATV distribution.

REFERENCES

[1] G.P. Agrawal and N.A. Olsson, "Amplification and Compression of Weak Picosecond Optical Pulses by using Semiconductor-Laser Amplifiers," *Optics Letters*, vol. 14, no. 10, pp. 500–502, May 1989.

[2] L.F. Tiemeijer, et al., "Polarization Insensitive Multiple Quantum Well Laser Amplifiers for the 1300 nm Window," *Applied Physics Letters*, vol. 62, no. 8, pp. 826–828, February 1993.

[3] L.F. Tiemeijer, et al., "27 dB Gain Unidirectional 1300 nm Polarization-Insensitive Multiple Quantum Well Laser Amplifier Module," *IEEE Photonics Technology Letters*, vol. 6, no. 12, pp. 1430–1432, December 1994.

[4] L.F. Tiemeijer, et al., "33 dB Fiber to Fiber Gain +13 dBm Fiber Saturation Power Polarization Independent 1310 nm MQW Laser Amplifiers," in *Technical Digest Topical Meeting on Optical Amplifiers and Applications / Post-deadline Papers*, (Davos, Switzerland), pp. PD1_1–PD1_5, June 1995.

[5] L.F. Tiemeijer, "High Performance MQW Laser Amplifiers for Transmisssion Systems Operating in the 1310 nm Window at Bitrates of 10 Gbit/s and Beyond," in *Proceedings 21th European Conference on Optical Communication / Invited Paper Tu.B.2.1*, (Brussels, Belgium), September 1995.

[6] Y. Ohishi, et al., "Praseodymium-Doped Fiber Amplifiers at 1.3 μm," *IEICE Transactions on Communications*, vol. E77-B, no. 4, pp. 421–439, 1994.

[7] M. Artiglia, et al., "Optical Fibre Amplifiers: Physical Model and Design Issues," *Optical and Quantum Electronics*, vol. 26, pp. 585–608, 1994.

[8] A.W.H. van Osch, *Modelling of Praseodymium-Doped Fluoride and Sulfide Fibre Amplifiers for the 1.3 μm Wavelength Region*. Eindhoven University of Technology, 1995. ISBN 90-6144-294-X.

[9] D.R. Simons, et al., "Pr^{3+}-doped GeS_x-based Glasses for Fiber Amplifiers at 1.3 μm," *Optics Letters*, vol. 20, no. 5, pp. 468–470, 1995.

[10] J.G.L. Jennen, et al., "Performance of Cascaded 1300 nm QW Laser Amplifiers in 10 Gbit/s Long Haul NRZ Transmission," in *Technical Digest Topical Meeting on Optical Amplifiers and Applications*, (Davos, Switzerland), pp. 123–126, June 1995.

[11] L.F. Tiemeijer, "Effects of Nonlinear Gain on Four-Wave Mixing and Asymmetric Gain Saturation in a Semiconductor Laser Amplifier," *Applied Physics Letters*, vol. 59, no. 5, pp. 499–501, July 1991.

[12] L.F. Tiemeijer, et al., "Reduced Intermodulation Distortion in 1300 nm Gain-Clamped MQW Laser Amplifiers," *IEEE Photonics Technology Letters*, vol. 7, no. 3, pp. 284–286, March 1995.

Performance of Time–domain Equalized Multicarrier Systems on ADSL Loops

Mark Van Bladel and Marc Moeneclaey

Communication Engineering Lab., University of Ghent
Sint–Pietersnieuwstraat 41, B–9000 Ghent, Belgium

Abstract — Multicarrier communication, without time–domain equalizer (TEQ), over a channel whose impulse response duration is of the same order as the inverse carrier spacing, requires a long guard interval yielding a considerable loss in both power and bandwidth efficiency. A smaller guard interval (yielding a smaller loss) can be allowed when using at the receiver a TEQ that shortens the impulse response of the cascade of channel plus TEQ.

In this contribution the impulse response of a FIR TEQ is determined according to a MMSE criterion, resulting in an eigenvalue problem. Numerical performance results as a function of TEQ complexity are presented, assuming both upstream and downstream Asymmetric Digital Subscriber Line (ADSL) communication over twisted pair cable.

1. Introduction

The last few years multicarrier systems, consisting of a large number of orthogonal, QAM modulated carriers, have been suggested for various applications [1–3]. For example, in [4] multicarrier communication was proposed for terrestrial and satellite broadcasting of digital audio (DAB) and in [5] for digital terrestrial television. Another main application is in high–rate communication over twisted–pair cable [6]. High rate digital subscriber line (HDSL) services can provide full duplex transmission at 1.5 to 2 Mbit/s over a distance of about 3.5 km. Asymmetric digital subscriber lines (ADSL) can offer a data rate of 1.5 to 6 Mbit/s from the central office to the subscriber over a distance of 3.0 to 5.5 km.

Multicarrier systems have some specific assets as compared to single–carrier communication. They are very robust for communication over dispersive channels : because of the insertion of a guard time whose duration exceeds the channel impulse response duration, no time domain equalization is needed at the receiver [1–2]. Further, modulation and demodulation can be implemented efficiently by means of a Fast Fourier Transform (FFT). This is in contrast with single–carrier communication, where a long time–domain equalizer is required (high complexity, slow convergence).

However, because of the guard–time insertion, a multicarrier system suffers a loss of bandwidth efficiency and a loss of power efficiency. A measure for this loss is given by $1 + T_g F_c$, where T_g denotes the guard time interval and F_c the frequency spacing between individual carriers. In the case of ADSL the carrier spacing is usually 3 kHz and a typical value for the impulse response duration is 250 μsec, so that the bandwidth and power efficiency are reduced by a factor 1.75 or 2.43 dB. In order to avoid this large loss, a time–domain equalizer (TEQ) can be introduced at the receiver side, such that the cascade of channel and TEQ yields a sufficiently short impulse response which allows a short guard time and a correspondingly small loss of efficiency [7].

In this contribution the TEQ impulse response is derived from a minimum mean–square error criterion. For a typical channel impulse response, the performance of the TEQ is investigated as a function of the number of TEQ taps.

2. Transceiver description

A guard time or cyclic prefix is introduced in multicarrier transceivers in order to eliminate ISI [1,2,3]. If, in a discrete–time representation, the channel impulse response has a duration of $\nu+1$ samples (obtained by sampling at rate $F_s = N.F_c$, where N is the size of the FFT and F_c the carrier spacing), each block of N samples is prefixed by a repeat of the ν last samples. Note that if we use a N–point FFT block, we can transmit at most N/2 carriers, i.e. at frequencies $0, F_c, ..., (N/2-1) F_c$. For the transmitted signal samples s_k we can write:

$$s_k = \mathrm{Re}\left\{ \sum_{n=0}^{N/2-1} a_{mn} \exp(\frac{j\,2\pi\,nk}{N}) \right\}$$

where a_{mn} denotes the m–th symbol on the n–th carrier and k falls within the range $[m(N+\nu)-\nu, m(N+\nu)+N-1]$. As shown in Figure 1, at the receiver side the cyclic prefix is first removed and a serial–to–parallel conversion is performed. After that follows a FFT. The output samples of the FFT are then fed into a frequency domain equalizer, which scales and rotates the data symbols in order to compensate for the attenuation and phase shift introduced by the channel.

The introduction of a guard time of ν samples yields a reduction of both the bandwidth and power efficiency by a factor of $(1+\nu/N)$. When the condition $\nu << N$ is not satisfied, the occurrence of a large loss can be avoided by using a linear time domain equalizer (TEQ) at the receiver. The TEQ is a filter with n_w taps, that should be selected in such a way that the cascade of the channel and the TEQ yields an impulse response whose duration is essentially limited to n_b samples, with $n_b << N$. In the next section the determination of the TEQ will be stated as an eigenvalue problem.

3. The eigenvalue problem

Consider the discrete equalizer model of Figure 2. Assume that $\underline{h} = [h_0 \ldots h_{M-1}]^T$ represents the original channel impulse response, $\underline{w} = [w_0 \ldots w_{n_w-1}]^T$ the TEQ and $\underline{b} = [b_0 \ldots b_{n_b-1}]^T$ the (short) desired impulse response of the cascade of $\underline{h}$ and $\underline{w}$. Let $\underline{v}$ denote a vector of n_w received samples and $\underline{x}_b$ an array of n_b samples of the transmitted sequence, delayed over Δ samples. Our aim is to determine $\underline{w}$, $\underline{b}$ and Δ such that the mean square error between $\underline{w}^T.\underline{v}$ and $\underline{b}^T.\underline{x}_b$ is minimum.

For the mean square error we can write:

$$E\{e^2\} = E\{(\underline{w}^T.\underline{v} - \underline{b}^T.\underline{x}_b)^2\} =$$

$$\underline{w}^T.R_{vv}.\underline{w} + \underline{b}^T.R_{xx}.\underline{b} - 2\underline{w}^T.R_{vx}.\underline{b} \quad (1)$$

with $R_{vv} = E\{\underline{v}.\underline{v}^T\}$, $R_{vx} = E\{\underline{v}.\underline{x}_b^T\}$ and $R_{xx} = E\{\underline{x}_b.\underline{x}_b^T\}$. Note that R_{vx} and R_{xx} both depend on the delay Δ.

For a given Δ, the optimal solution for the taps of the TEQ, $\underline{w}_{opt}$, can be found by putting $\partial(E\{e^2\})/\partial\underline{w} = 0$, which leads to

$$\underline{w}_{opt} = R_{vv}^{-1}.R_{vx}.\underline{b} \quad (2)$$

Now (2) is an expression for the optimal $\underline{w}$ as a function of $\underline{b}$. Substituting this expression in (1) gives

$$E\{e^2\} = \underline{b}^T.(R_{xx} - R_{vx}^T(R_{vv}^{-1})^T R_{vx}).\underline{b} = \underline{b}^T.O.\underline{b}$$

Minimizing this expression we see that $\underline{b}_{opt}$ can be found as the eigenvector corresponding to the smallest eigenvalue of the matrix O. Moreover we will choose $\underline{b}_{opt}$ such that $\underline{b}_{opt}^T.\underline{b}_{opt} = 1$. The minimum eigenvalue λ_{min} is considered as the performance measure, because it equals the minimum value of the power of the error signal e. We will calculate λ_{min} for various values of Δ, in order to find the minimum over Δ. When this globally minimum eigenvalue λ_{opt} and the corresponding eigenvector $\underline{b}_{opt}$ are found, the optimum tap values for the TEQ can be calculated.

4. Numerical results

In order to obtain the numerical results presented in this section we considered carrier–serving–area (CSA) loop no.6 (9000 feet length − 26 gauge). Figure 3 shows the transfer characteristics (including transmit and receive filters) for downstream and upstream communication, which occupy the frequency intervals 144 kHz − 768 kHz and 0 − 144 kHz respectively. Figure 4 gives the corresponding impulse responses. Note the much longer duration of the impulse response in the case of upstream communication. The carrier spacing is 3 kHz and the sampling frequency is 1.536 MHz; this corresponds to a 512–point FFT. In figures 5a and 5b the minimum eigenvalue is given as a function of the delay Δ, applied to the input signal, and this for a desired impulse response of $n_b = 32$ taps and various lengths n_w of the TEQ. As can be seen from Fig. 5a, there is a large range of delay values Δ that yield a mean square error of less than 0.01 %, even for a TEQ with only 16 taps. The results in Fig. 5b show a much worse performance on the upstream channel : for $n_b = 32$ the mean square error is still a few percent.

In Fig. 6a–6b the energy spectrum is shown of the error signal between the resulting eigenvector and $\underline{w}_{opt} * \underline{h} = (R_{vv}^{-1}).R_{vx}.\underline{b}_{opt} * \underline{h}$. Fig. 6a and 6b are for downstream and upstream communication, respectively. Δ is taken to be 255. We see that the carriers in the transition zone between downstream and upstream frequencies are the main contributors to the error signal energy.

5. Conclusions

In this paper we presented a method to calculate the optimum filter tap coefficients for a time–domain equalizer that is used in order to shorten the channel impulse response. Performance results on a typical channel transfer function were calculated. Communication over the upstream channel proved to be much more critical than over the downstream channel.

References

[1] J. Bingham, "Multicarrier modulation for datatransmission : an idea whose time has come", IEEE Commun. Mag., Vol. 28, May 1990, pp. 5–14

[2] S. Weinstein and P. Ebert, "Data transmission by frequency–division multiplexing using the discrete Fourier transform", IEEE Trans. Commun. Tech., Vol. 29, Oct. 1971, pp. 628–634

[3] M. Alard and R. Lassalle, " Principles of modulation and channel coding for digital broadcasting for mobile receivers", EBU Review – Technical No. 224, August 1987

[4] P. Höher, J. Hagenauer, E. Offer, C. Rapp and H. Schulze, "Performance of an PCPC–coded OFDM–based digital audio broadcasting (DAB) system", Proc. Globecom '91, Phoenix, Arizona, Dec. 1991, pp. 40–46

[5] P. de Bot and C. Baggen, "Reed–Solomon codes for OFDM broadcasting over frequency selective channels using error and erasure coding", 6th International Workshop on Digital Communications, Tirrenia, Italy, 5–9 September 1993, pp. 203–214

[6] High–speed digital subscriber lines, IEEE J. Sel. Ar. in Comm., vol. 9, no.6, August 1991

[7] J. Chow, J. Cioffi and J. Bingham, " Equalizer training algorithms for multicarrier modulation systems", International Conference on Communications, Geneva, Switzerland, pp. 761–765, May 1993
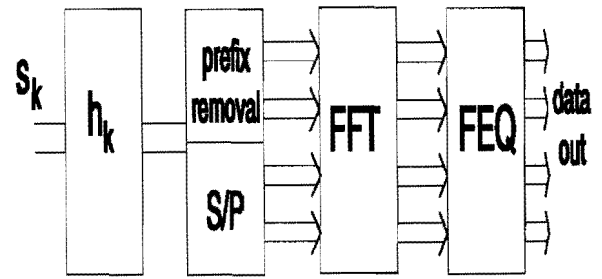
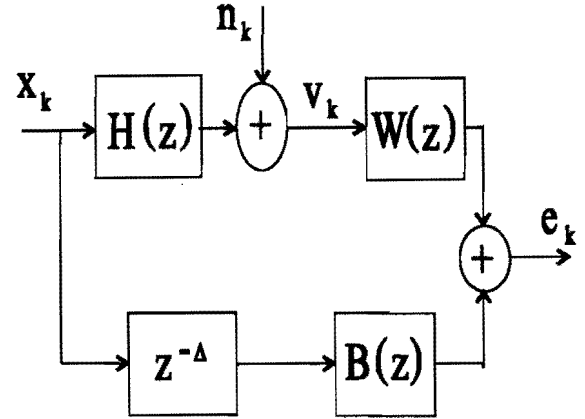Fig. 1 : Structure of an OFDM receiver
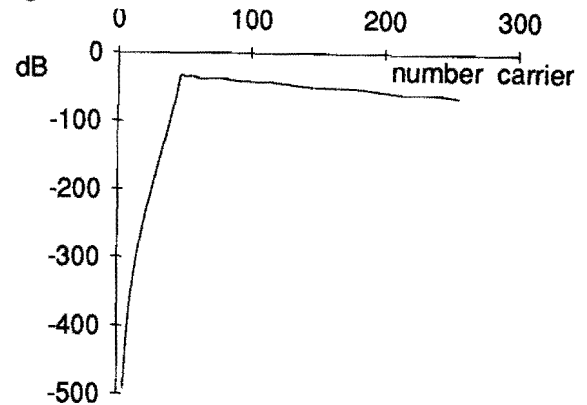

Fig. 2 : Discrete equalizer model

Fig. 3a : Downstream transfer characteristic

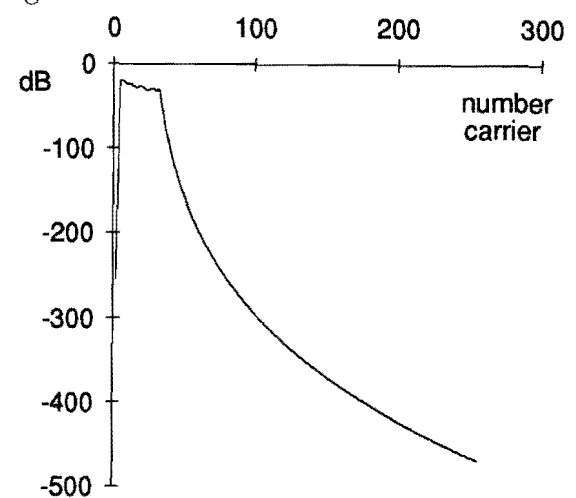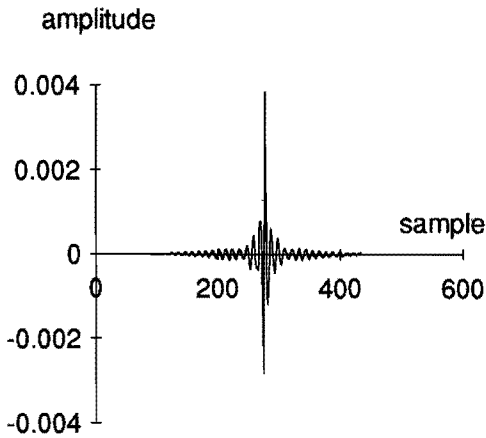Fig. 3b : Upstream transfer characteristic
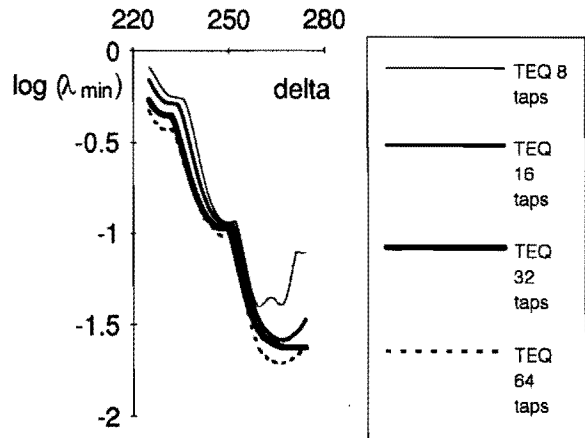
Fig. 4a : Downstream impulse response

Fig. 5b : $\lambda_{min}(\Delta)$ for upstream communication

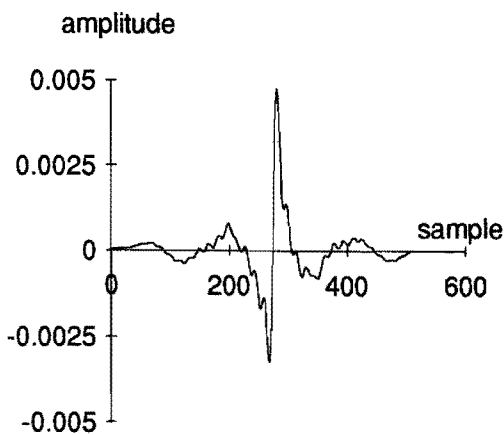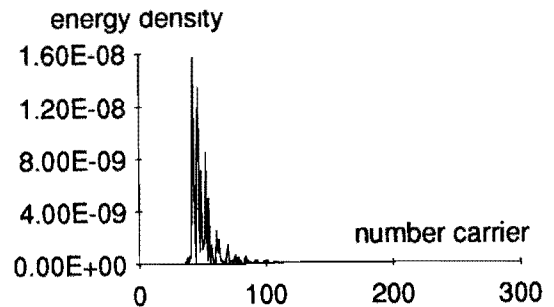Fig. 4b : Upstream impulse response

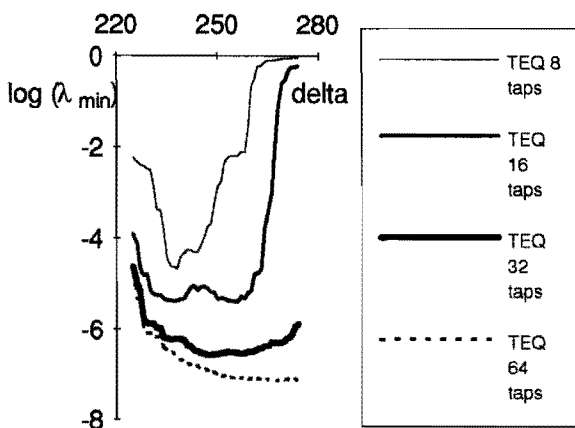Fig. 6a : Energy density spectrum of error signal (downstream)

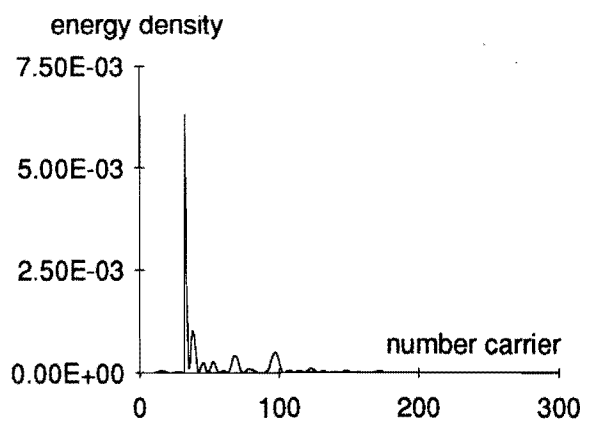Fig. 5a : $\lambda_{min}(\Delta)$ for downstream communication

Fig. 6b : Energy density spectrum of error signal (upstream)

Performance of Quasi-Coherent Signal Combining in case of Co-Channel Interference for Microwave Indoor Communication Systems

Bas W. 't Hart, Gerard J.M. Janssen

Telecommunications and Traffic Control Systems Group, Delft University of Technology
P.O. Box 5031, 2600 GA Delft, The Netherlands
Tel: +31-15-786826, Fax: +31-15-781774
E-mail: b.thart@et.tudelft.nl, g.janssen@et.tudelft.nl

Abstract: In this paper the performance enhancement for combining multiple antenna signals is analysed in case of co-channel interference. The proposed combining technique, is based on quasi-coherent combining of the dominant paths of the signals received by a number of closely spaced antenna elements.

Besides the influence of multipath interference, co-channel interference is taken into account, which degrades system performance severely, especially for small SIR's. In the model, the required phase shifts of the antenna signals are determined in two ways: with and without decision feedback.

The BER is evaluated for coherent Binary Phase Shift Keying in the Rician fading indoor multipath channel. Computational results show, that quasi-coherent signal combining yields a major improvement of the BER performance compared to the one antenna case without requiring a priori knowledge of the multipath channel. The obtained BER results show that the combining method using remodulation with decision feedback results in a much better performance than the simple method without decision feedback in the case of small SIR.

I INTRODUCTION

Frequency bands in the microwave and millimetre wave region will be used for indoor wireless personal communications in future systems. Advantages of the higher frequency bands are: i.) availability of large bandwidths; ii.) better shielding by concrete, so that cell sizes can be smaller; iii.) antenna dimensions can be small. In office environments, applications are foreseen to interconnect personal computer systems at high bit rates > 20 Mbit/s.

However, the indoor channel is an adverse frequency selective multipath channel, causing severe signal dispersion which limits the maximum usable symbol rate. To improve transmission performance in a multipath channel, different antenna combining techniques have been proposed, however, which require a priori channel knowledge, [1,2].

In [3], a new antenna combining technique was proposed. This technique of combining multiple antenna signals to achieve better transmission performance can be applied in the situation of a multipath frequency selective communication channel. The technique is based on quasi-coherent (QC) combining of the dominant path of the signals received by a number of closely spaced antenna elements. In case of small array dimensions, the channel power delay profile (PDP) is nearly equal for all antennae. The signals of the antennae are phase shifted so that the phases of the dominant path of all antennae become nearly equal. Summation of the signals results in nearly coherent addition of the dominant paths of the PDP's. Signals which arrive from other directions, will add incoherently. Therefore, the dominant path is strongly enhanced compared to the other paths in the resulting PDP, which shows less time delay spread and consequently has a more equalised frequency response.

Besides the influence of multipath interference, which has been discussed in [3,4], in this paper co-channel interference is taken into account, which degrades system performance severely, especially for small SIR's. Another extension is made in the way the phase shift of the antenna signals is determined in the model. Instead of using a simple method without decision feedback, remodulation is applied to obtain the phase difference.

In the next section, the indoor multipath channel is modelled. The QC signal combining technique is explained in section III. Section IV discusses the applied BER model for coherent Binary Phase Shift Keying (BPSK) modulated signals. Computational results on the BER performance for QC-combining in case of co-channel interference, obtained for both the simple method and the remodulation method, are given in section V. Section VI contains the conclusions.

II INDOOR MULTIPATH CHANNEL MODEL

In a multipath channel, multiple replicas of the signal simultaneously arrive at the receiver antenna. The received signal consists of one line-of-sight (LOS)

component or otherwise dominant component, and a number of reflected components, each having different amplitude, phase and time delay. This multipath effect causes frequency selectivity and dispersion of the signal. The multipath channel is modelled as, [5]:

$$h(t) = \sum_{k=0}^{N} \beta_k \exp(j\theta_k)\delta(t - \tau_k) \qquad (1)$$

where, β_k is the path gain, θ_k is the phase shift and τ_k is the time delay of the kth path ($k = 0$ for the dominant signal and $\tau_0 = 0$), and $\delta(.)$ represents the Dirac delta function. We assume that the channel is time invariant. From the impulse response, the power delay profile (PDP) is derived as

$$PDP(t) = |h(t)|^2 \qquad (2)$$

An important characteristic parameter of the multipath channel is the Root Mean Square delay spread τ_{RMS}. The discrete version can be written as

$$\tau_{RMS} = \sqrt{\frac{M_2}{M_0} - \left(\frac{M_1}{M_0}\right)^2} \qquad (3)$$

with $M_n = \sum_{k=0}^{N} \beta_k^2 \tau_k^n$.

The LOS or dominant path can be modelled as a Rician fading channel where the distribution of the instantaneous received power P_i, is given by

$$f_{Rice}(P_i | K, \overline{P}) = \frac{1+K}{\overline{P}} \exp\left(-\frac{(1+K)P_i + K\overline{P}}{\overline{P}}\right) \cdot I_0\left(2\sqrt{\frac{K(1+K)P_i}{\overline{P}}}\right) \qquad (4)$$

Here, $\overline{P}$ = M_0 is the average received power, and K is the Rice factor which is defined by

$$K \triangleq \frac{P_{LOS}}{P_{Reflected}} = \frac{\beta_0^2}{\sum_{k=1}^{N} \beta_k^2} \qquad (5)$$

III QUASI-COHERENT COMBINING OF MULTIPLE ANTENNA SIGNALS

To reduce the multipath effects a signal combining technique is applied. From the results presented in [4] and [6], it can be concluded that for antennae that are separated by a few wavelengths, the amplitudes of identical paths in different PDP's are highly correlated, especially for LOS situations. The aim of the signal combining technique described here, is to sum the antenna signals in such a way that the dominant paths of the M

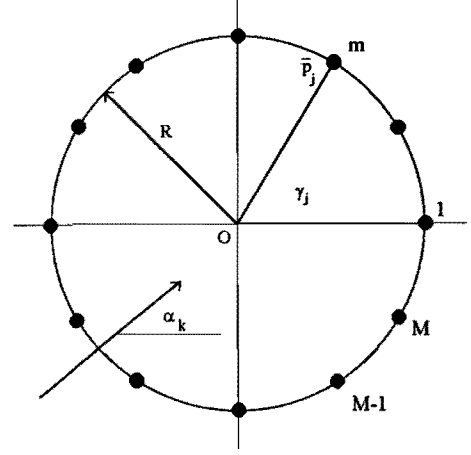antennae add coherently, while the reflected paths add incoherently.

Figure 1: Structure of the M-element circular antenna array.

If a circular antenna array is used, as shown in figure 1, where O is the origin of the array, $\bar{p}_j$ is the position of the jth antenna, γ_j is the direction of the jth antenna with respect to O, α_k is the direction of arrival of the kth path, and $\bar{n}_k$ is the unit vector with direction α_k, the phase difference between two antennae#m,n is given by:

$$\Delta\phi_{m,n} = \frac{2\pi}{\lambda}\left[(\bar{p}_m \cdot \bar{n}_k) - (\bar{p}_n \cdot \bar{n}_k)\right]$$
$$= \frac{2\pi R}{\lambda}\left[\cos(\alpha_k - \gamma_m) - \cos(\alpha_k - \gamma_n)\right] \qquad (6)$$

Let the dominant path arrive at the array with reference direction α_R, and let us take the signal phase of antenna#1 as reference phase, with $\gamma_1 = 0$. For coherent summation of the reference path, the M antenna signals are phase shifted by

$$\Delta\phi_{m,1} = \frac{2\pi R}{\lambda}\left[\cos(\alpha_R - \gamma_m) - \cos\alpha_R\right] \qquad (7)$$

for $m \in \{2,...M\}$. However, other paths arriving from different directions are also phase shifted at antenna m with $\Delta\phi_{m,1}$, which results in a random phase for R/λ large enough. The resulting phase θ_m of antenna m for a path with direction of arrival α_k is

$$\theta_m = \phi + \frac{2\pi R}{\lambda}\left[\cos(\alpha_k - \gamma_m) - \cos(\alpha_R - \gamma_m) + \cos\alpha_R\right] \qquad (8)$$

with ϕ a random phase. As can be seen in [4], for array radius $R > (M\lambda/4\pi)$ good decorrelation between the antenna signals is achieved, which is equivalent with antenna distance $d > \lambda/2$. This signal combining

technique results in beam forming of the array in the direction of arrival α_R of the dominant path.

Applying the signal combining technique results in a new Rician channel with improved performance, if the PDP is the same for all elements. The improvement is due to the following factors which are discussed in detail in [4]. First, in case of coherent combining of the dominant signals of an M-element array with sufficiently large R, the expected value of the Rice factor $\overline{K}_{sum}$ of the new channel becomes

$$\overline{K}_{SUM} = \frac{M^2\beta_0^2}{M\sum_{k=1}^{L}\beta_k^2} = MK \qquad (9)$$

This means that the interference power from the weaker paths, and therefore also the Inter-Symbol-Interference (ISI) caused by these signals, decreases with $1/M$ relative to the dominant path. For large M and for a large Rice factor, the expected value of the delay spread $\overline{\tau}_{RMS}$ of the new channel can be approximated by

$$\overline{\tau}_{RMS} = \frac{\tau_{RMS}(1)}{\sqrt{M}} \qquad (10)$$

where, $\tau_{RMS}(1)$ is τ_{RMS} for the signal received by one antenna. A third factor is the increase of signal-to-noise ratio (SNR). At high frequencies the received noise power is mainly generated in the receiver front-end. In our case the noise is generated in the M front-ends after each antenna element. In the combining process, M independent noise variables of equal power P_n are summed, resulting in the final output noise power MP_n. The ratio of the signal power after combining, $S(M)$, and the power of one antenna, $S(1)$, can be approximated for large K by

$$\frac{S(M)}{S(1)} = \frac{M^2\beta_0^2(1+1/MK)}{\beta_0^2(1+1/K)} \approx M^2 \qquad (11)$$

So $SNR(M)$ after combining increases approximately with $M \cdot SNR(1)$.

For achieving this performance improvement, it is crucial that the phase of the dominant path at the outputs of the array elements is estimated with sufficient accuracy. In our case the phase of the total signal is taken as the estimate of the phase of the dominant path. In [4], it is shown that for small values of $K > 0$ dB, which are often found in indoor channels, large probability is found that the phase difference φ between the resulting phase of the total signal and the phase of the dominant component, is less than $\pi/4$, which means a reasonable correlation of the dominant signal components after phase shifting. Because the determined phase shift does not exactly result in

coherent combining of the dominant paths, the term quasi-coherent instead of coherent signal combining is used.

IV BER MODEL

In this section, the BER model as given in [7] is adapted for the multipath channel with interference. The input of the receiver matched filter in case of multipath is given by

$$r_{dem}(t) = \sum_{k=0}^{N}\beta_k d(t-\tau_k)\cos(\theta_k - \Phi_c) \qquad (12)$$

where Φ_c is the required phase shift.

This signal consists of the contributions of $N+1$ paths. The maximum number of bits which are simultaneously active in the channel during one bit time T is $L+2$, with $L = \lceil (\tau_N-\tau_0)/T \rceil = \lceil \tau_N/T \rceil$ ($\lceil x \rceil$ indicates the integer value larger or equal to x) which is sketched in figure 2. From these active bits, the bit which has maximum average energy is selected as the desired bit, which has to be detected.
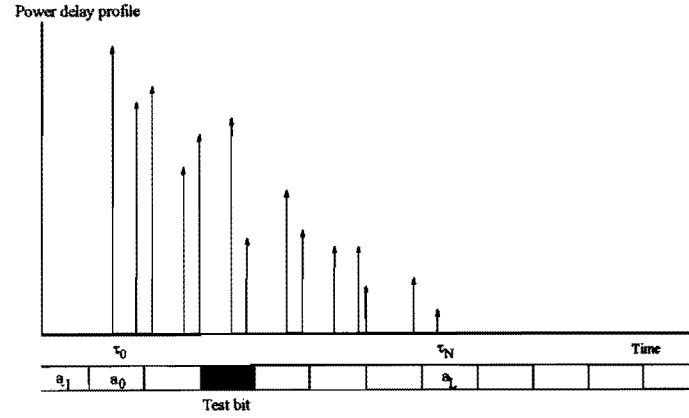

Figure 2: Schematic representation of the channel impulse response used to compute BER values.

The received power for a single bit is determined by transmitting a test bit over the multipath channel as depicted in figure 2. Let the test bit be defined as $p_{test}(t) = U(t) - U(t-T)$, then the output signal $s_{test}(\tau_s)$ after demodulation for given sample time τ_s is

$$s_{test}(\tau_s) = \int_{\tau_s-T}^{\tau_s}\sum_{k=0}^{N}\beta_k p_{test}(t-\tau_k)\cos(\theta_k - \Phi_c)dt \qquad (13)$$

Here, τ_s is the sample time referred to the delay of the paths in the channel impulse response $h(t)$, $\tau_s \in [T, \tau_N+T)$. The sign of s_{test} is determined by the phases of the contributing paths.

When a data signal is transmitted, a sequence of bits is active in the channel. The sequence of active bits can be written in vector notation as $\underline{a} = (a_{-1}, a_0, a_1, ..., a_{d-1}, a_d, a_{d+1}, ..., a_L)$, with $a_j \in \{-1,+1\}$, and the desired bit is

204

indicated as a_d. The position of a_d in $\underline{a}$ depends on $h(t)$, and the sample time that maximises the average energy. The index d is equal to $d = \lceil \tau_s/T \rceil$. In the integration interval $[\tau_s - T, \tau_s)$ in general two bits, a_{i-1} and $a_i \in \underline{a}$, are active in path k. The index i is given by $i = \lceil \tau_k/T \rceil$. The contribution of path k to the output signal of the matched filter is given by

$$s_k = \beta_k T \cos(\theta_k - \Phi_c)\left[(1 - \Delta_k)a_{i-1} + \Delta_k a_i\right]$$
$$= a_{i-1}\kappa_{k,i-1} + a_i\kappa_{k,i} \tag{14}$$

with $\kappa_{k,i-1} = \beta_k(1 - \Delta_k)T\cos(\theta_k - \Phi_c)$, and $\kappa_{k,i} = \beta_k \Delta_k T \cos(\theta_k - \Phi_c)$. Δ_k indicates the fraction of bit a_i in the integration interval for path k:

$$\Delta_k = \mathrm{Frac}\left(\frac{\tau_k - \tau_s}{T}\right) \quad \text{if } \tau_s < \tau_k$$
$$= \mathrm{Frac}\left(\frac{\tau_s - \tau_k}{T}\right) \quad \text{if } \tau_s > \tau_k \tag{15}$$

The output signal of the matched filter for all $N+1$ paths is given by

$$s_{out}(\underline{a}, \tau_s) = \underline{a} \cdot \mathrm{K}(\tau_s) \tag{16}$$

where, $\mathrm{K}(\tau_s)$ is the $L+2 \times N+1$ path matrix for sample time τ_s.

$$\mathrm{K}(\tau_s) = \begin{pmatrix} \kappa_{0,-1} & 0 & 0 & \dots & 0 & 0 \\ \kappa_{0,0} & \kappa_{1,0} & \kappa_{2,0} & \dots & \dots & \dots \\ 0 & \kappa_{1,1} & \kappa_{2,1} & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \kappa_{N-1,L-2} & \dots \\ \dots & \dots & \dots & \dots & \kappa_{N-1,L-1} & \kappa_{N,L-1} \\ 0 & 0 & 0 & \dots & 0 & \kappa_{N,L} \end{pmatrix} \tag{17}$$

Note that $\mathrm{K}(\tau)$ is periodic in T. Row l of K, with $l \in \{-1, 0, \dots, L\}$, contains the contribution of bit a_l, the column $\mathrm{K}_k = (0, 0, \dots, \kappa_{k,i-1}, \kappa_{k,i}, 0, \dots, 0)^T$ contains the contribution of path k. Now the received power S_i from bit i in a bit interval T, can be written as

$$S_i(\tau_s) = \left|\sum_{k=0}^{N}\kappa_{k,i}(\tau_s)\right|^2 \tag{18}$$

The received ISI power S_{ISI}, is the average received power due to bits that are uncorrelated with the desired bit

$$S_{ISI} = \sum_{k=0}^{N}\sum_{\substack{l=-1 \\ l \neq d}}^{L}\kappa_{k,l}^2 \tag{19}$$

So far, only the reception of one signal has been discussed. In the situation of co-channel interference, other signals, also composed of a dominant component and a number of reflected components, are received. In case of one interfering signal, the total output signal of the matched filter becomes:

$$s_{out}(\underline{a}, \underline{a}^*, \tau_s) = \underline{a} \cdot \mathrm{K}(\tau_s) + \underline{a}^* \cdot \mathrm{K}^*(\tau_s) \tag{20}$$

where * denotes the undesired signal. The matrices K and K^* belong to different PDP's. The data sequence $\underline{a}$ and the data sequence of the interfering signal $\underline{a}^*$ are uncorrelated. Further, the assumption is made that the data sequences $\underline{a}$ and $\underline{a}^*$ are time synchronous, so the worst case situation is considered.

The BER can be calculated as the average BER for all possible combinations of the values of the active bits in the channels of both the desired and the undesired signal. If we let the desired bit have the constant value $a_d = +1$, then the vector $\underline{a}$ takes on $V = 2^{L+1}$ different values. Because there is no desired bit in the interfering signal, the vector $\underline{a}^*$ takes on $W = 2^{L^*+2}$ different values. This results in $V \cdot W$ possible output signals $s_{out}(\underline{a}_y, \underline{a}_z^* | a_d = +1)$, with $\underline{a}_y \in \{\underline{a}_1, \dots, \underline{a}_V\}$ and $\underline{a}_z^* \in \{\underline{a}_1^*, \dots, \underline{a}_W^*\}$. The error probability, P_{eyz}, for the combination of $\underline{a}_y$ and $\underline{a}_z^*$ is given by

$$P_{eyz} = P\left(\varepsilon | a_d = +1, \underline{a}_y, \underline{a}_z^*\right)$$
$$= P\left(sign\left(s_{out}(\underline{a}_y, \underline{a}_z^* | a_d = +1)\right) = -sign(s_d)\right)$$
$$= \begin{cases} Q\left(\frac{\left|s_{out}(\underline{a}_y, \underline{a}_z^* | a_d = +1)\right|}{\sigma_n}\right), \\ \qquad \text{if } sign(s_{out}) = sign(s_d) \\ 1 - Q\left(\frac{\left|s_{out}(\underline{a}_y, \underline{a}_z^* | a_d = +1)\right|}{\sigma_n}\right), \\ \qquad \text{if } sign(s_{out}) = -sign(s_d) \end{cases} \tag{21}$$

with $Q(z) = \int_z^{\infty} e^{-\lambda^2/2}d\lambda$ and σ_n^2 the received noise power. $Sign(s_d)$ is determined with (13). The BER is determined by averaging P_{eyz} over all $V \cdot W$ possible sequences $(a_{-1}, a_0, \dots, a_d, \dots, a_L)$, $(a_{-1}^*, a_0^*, \dots, a_{L^*}^*)$:

$$P(\varepsilon | a_d = +1) = \frac{1}{V \cdot W}\sum_{y=1}^{V}\sum_{z=1}^{W}P_{eyz} \tag{22}$$

σ_n^2 is related to the Signal-to-Noise Ratio (SNR) as $\sigma_n^2 = S_{av}/SNR$ with S_{av} is the average received power of the desired signal

$$S_{av} = \sum_{k=0}^{N}\beta_k^2 \tag{23}$$

V RECEIVER MODEL

The required phase shift Φ_c (see eq. (12)) is calculated in two ways. The first method applied is based on remodulation. In figure 3, the scheme of a coherent BPSK receiver with QC-combining of multiple antenna signals is shown for the case of two antennae. Remodulation is applied for phase estimation, by multiplying the delayed input signal with the estimated data. The delay of the input signal is introduced to take into account the time required for optimum detection of $d(t)$. In case no errors are made, remodulation fully removes the modulation, resulting in a clean carrier signal from which the reference phase is determined. In case of errors, the fraction $(1-2P_e)^2$ of the total signal power is concentrated in the recovered carrier, with P_e is BER, [8].
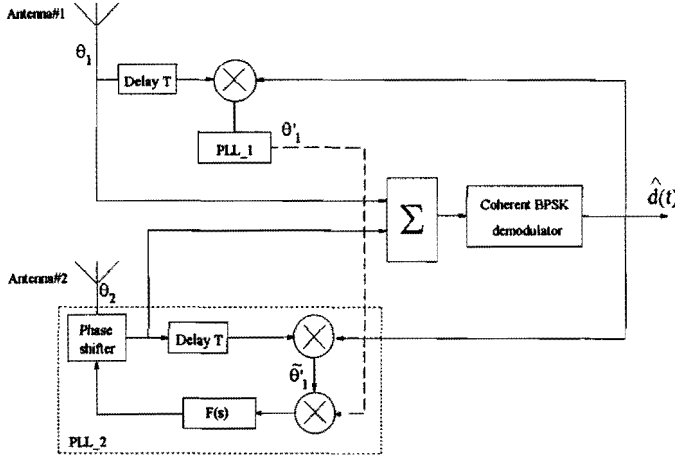
Figure 3: Scheme of coherent BPSK receiver with QC-combining of multiple antenna signals with decision feedback.

In a multipath environment, however, ISI components due to reflections are present. Remodulation of a multipath signal results in a carrier signal with additional components which are uncorrelated or only partly correlated with $\hat{d}(t)$.

The complex low-pass equivalent multipath signal can be written as

$$r_{lp}(t) = \sum_{k=0}^{N} \beta_k d(t - \tau_k) \exp(j\theta_k) \qquad (24)$$

If we assume that the phase locked loop (PLL) has infinitely small bandwidth, rejecting phase-jitter due to noise and time variant ISI, the phase recovered after remodulation is given by

$$\Phi_c = \arg\left(E\left[\sum_{k=0}^{N} \beta_k d(t - \tau_k) \hat{d}(t - \tau_s) \exp(j\theta_k) \right] \right)$$

(25)

In case of low BER (25) becomes

$$\Phi_c = \arg\left(\sum_{k=0}^{N} \beta_k R_d(\tau_k - \tau_s) \exp(j\theta_k) \right) \qquad (26)$$

where $R_d(\tau)$ is the autocorrelation function of $d(t)$, and τ_s is the optimum sample time. Only the paths with a time difference less than one bit time with respect to τ_s contribute to Φ_c. Interference with uncorrelated data $d^*(t)$ will not affect the phase estimation process. The phase shifting of the signal of antenna#m over $\Phi_{c,1} - \Phi_{c,m}$, is performed by a PLL controlled phase shifter. After summation, the resulting signal is demodulated by the BPSK demodulator which recovers its phase reference from the combined signal in the same way as described by (26).

This method using decision feedback, will be called the remodulation method. Here, only the paths of the desired signal that contribute to the desired bit contribute to the phase shift Φ_{c_s}.

The second method can use a less complex receiver structure, as shown in figure 4.
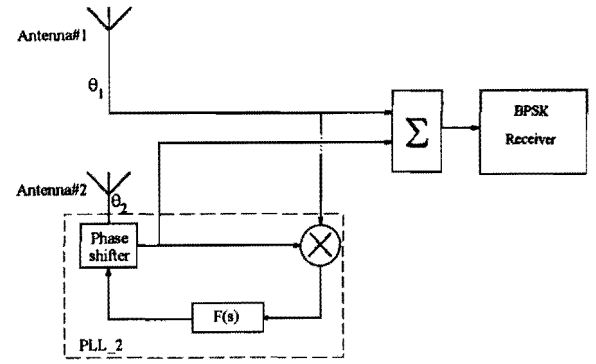

Figure 4: Scheme of coherent BPSK receiver with QC-combining using the multiplying technique for phase shift estimation.

The phase difference between two array elements is determined from multiplying the reference signal with the complex conjugate of the signal under consideration:

$$\Phi_{c,1} - \Phi_{c,m} = \arg\left(E\left[(r_1 + r_1^*) \cdot (r_m + r_m^*)' \right] \right) \qquad (27)$$

where r_m represents the signal of antenna#m, and ' denotes the complex conjugate. Using a PLL with an infinitely small bandwidth, the recovered phase difference is given by

$$\Phi_{c,1} - \Phi_{c,m} =$$

$$\arg\left(\sum_{p_1=0}^{N} \sum_{p_m=0}^{N} \beta_{p_1} \beta_{p_m} R_d(\tau_{p_1} - \tau_{p_m}) \exp(j(\theta_{p_1} - \theta_{p_m})) + \right.$$

$$\left. \sum_{q_1=0}^{N'} \sum_{q_m=0}^{N'} \beta_{q_1}^* \beta_{q_m}^* R_d^*(\tau_{q_1}^* - \tau_{q_m}^*) \exp(j(\theta_{q_1}^* - \theta_{q_m}^*)) \right) \qquad (28)$$

206

where $R_d(\tau)$ and R_d^* are the autocorrelation functions of $d(t)$ and $d^*(t)$, respectively. Notice that all paths of both the desired and the interfering PDP now contribute to the estimated phase shift. In the following, this method without decision feedback will be called the simple method.

VI COMPUTATIONAL RESULTS

The BER results of quasi coherent signal combining in case of one interfering signal are calculated for both phase estimation methods. The cumulative BER different distributions (CBD's) are computed from 500 different generated channel impulse responses with an exponential power delay profile for both the desired and the interfering signal. This is done for different number of antennae, different Rice factor, different bit time, and SIR, with SNR = 10 dB. The maximum number of active bits that are taken into account to calculate the BER is 7. In order to compare the performance of different bitrates R_b for the same τ_{RMS}, the normalised bitrate R_{norm}, defined as

$$R_{norm} \triangleq R_b \cdot \tau_{RMS} = \frac{\tau_{RMS}}{T_b} \qquad (29)$$
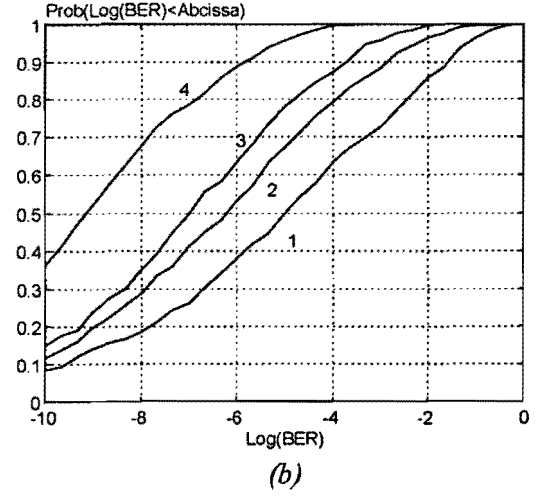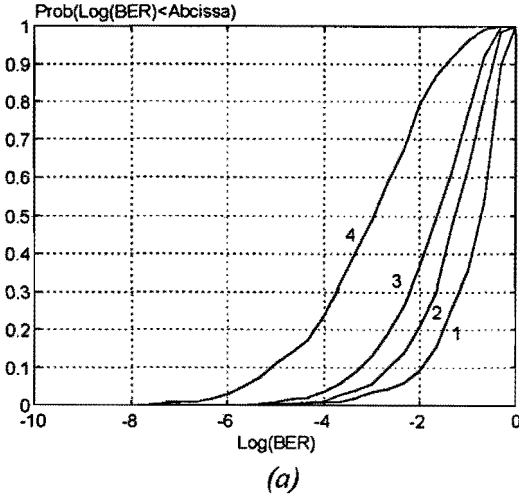
is used.

Figure 5: Cumulative distribution function for the BER (remodulation method) with $R_{norm} = 0.01$, and $K = 6$ dB, for: (a) $M = 1$, (b) $M = 4$; 1.) SIR = 0 dB; 2.) SIR = 3 dB; 3.) SIR = 6 dB; 4.) SIR = ∞;
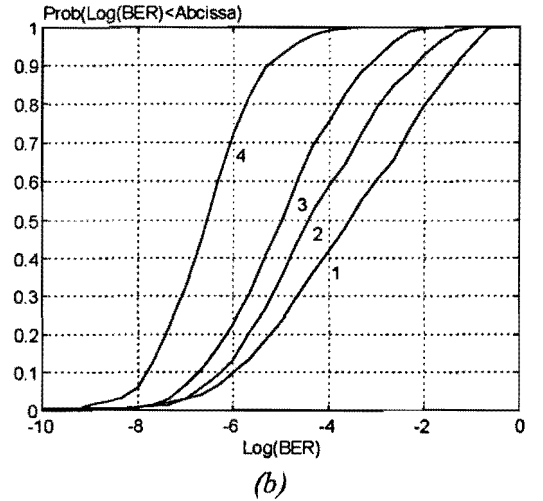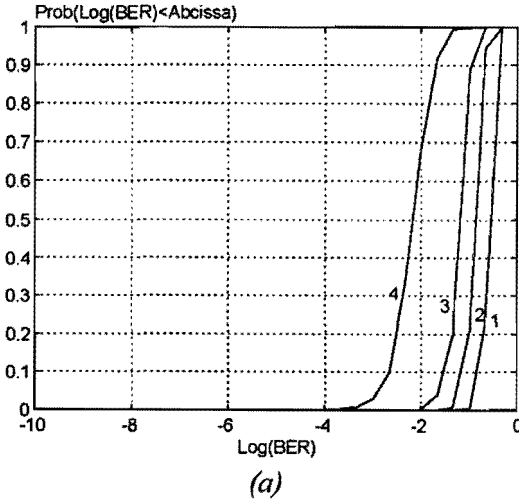

Figure 6: Cumulative distribution function for the BER (remodulation method) with $R_{norm} = 2$, and $K = 6$ dB, for: (a) $M = 1$ (b) $M = 4$; 1.) SIR = 0 dB; 2.) SIR = 3 dB; 3.) SIR = 6 dB; 4.) SIR = ∞;

Effect of quasi coherent signal combining. First the effect of QC-combining is evaluated by comparing the BER results for $M = 1$ and $M = 4$. CBD's are presented for four different Signal-to-Interference ratio's, SIR = 0 dB, 3 dB, 6 dB, and ∞ (no co-channel interference), and for two values of the normalised bitrate, $R_{norm} = 0.01$ (narrowband case), and $R_{norm} = 2$ (heavy ISI). To avoid that multipath interference is dominant over the co-channel interference, a Rice factor of K = 6 dB is used. Figure 5 shows the results for $R_{norm} = 0.01$ without and without QC-combining ($M = 1$ and $M = 4$), figure 6 shows the results for $R_{norm} = 2$. The results are obtained using the remodulation method. It is obvious that QC-combining results in superior performance compared to the one antenna case. The results also show that the influence of co-channel interference is considerable. It

should be noted however, that these results represent a worst case situation since the data sequences of the desired and the interfering signal are assumed synchronous.

Effect of remodulation. The effect of remodulation in QC-combining is evaluated by comparing the BER results obtained for the simple method and for the remodulation method. The CBD's are shown for SIR = 0 dB, 3 dB, 6 dB, and ∞ (no co-channel interference), and for two values of the normalised bitrate, $R_{norm} = 0.01$ (narrowband case), and 2, and $M = 4$. A Rice factor of K = 3 dB is used. Figure 7 shows the results for $R_{norm} = 0.01$ obtained for the simple method and with decision feedback. Figure 8 shows the same results for $R_{norm} = 2$.
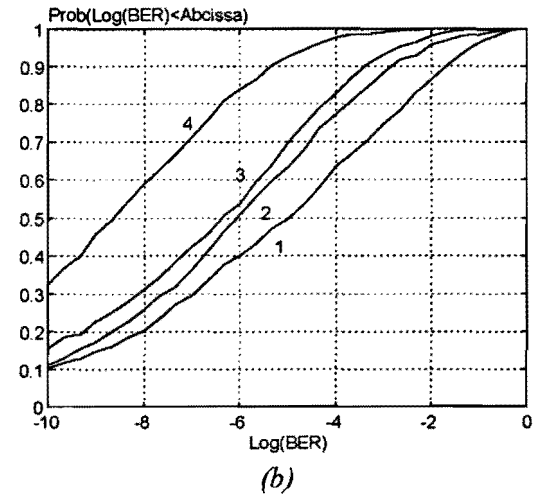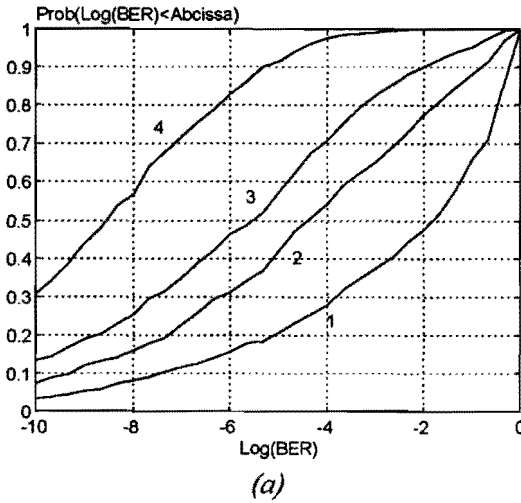

Figure 7: Cumulative distribution function for the BER with M = 4, and K = 3 dB, obtained for $R_{norm} = 0.01$ using: (a) simple method, (b) remodulation method; 1.) SIR = 0 dB; 2.) SIR = 3 dB; 3.) SIR = 6 dB; 4.) SIR = ∞;
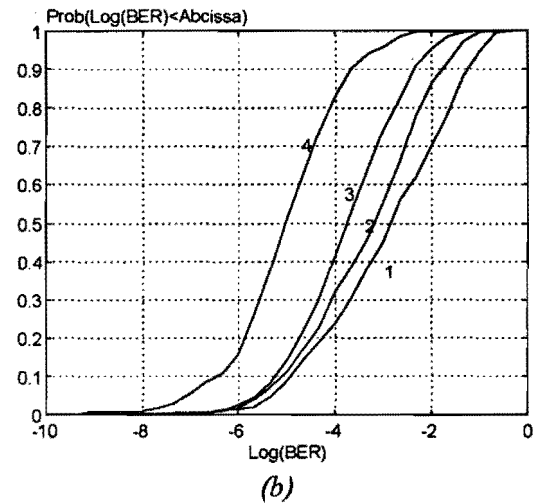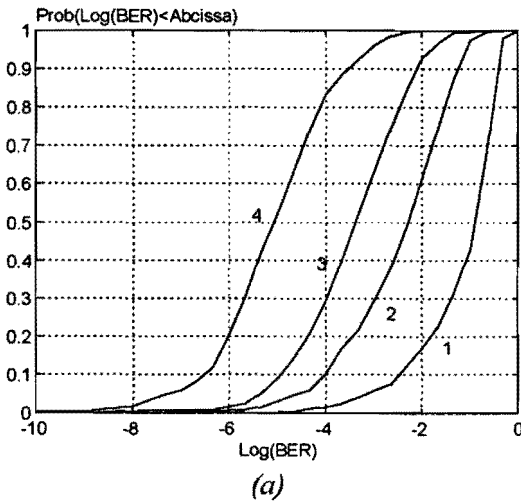

Figure 8: Cumulative distribution function for the BER with M = 4, and K = 3 dB, obtained for $R_{norm} = 2$ using: (a) simple method, (b) remodulation method; 1.) SIR = 0 dB; 2.) SIR = 3 dB; 3.) SIR = 6 dB; 4.) SIR = ∞;

The difference in BER results for the remodulation method and the simple method is clearly shown in figure 9. In this figure the CBD's for $R_{norm} = 0.01$, $K = 6$ dB, and $M = 4$, are depicted for both methods.
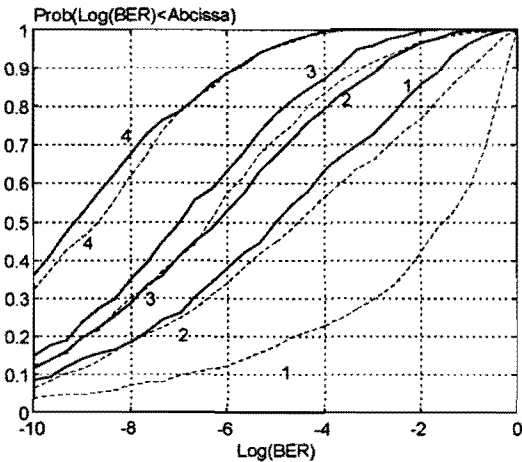

Figure 9: Cumulative distribution function for the BER with M = 4, R_{norm} = 0.01, and K = 6 dB, obtained with the simple method - - - and the remodulation method —— for: 1.) SIR = 0 dB; 2.) SIR = 3 dB; 3.) SIR = 6 dB; 4.) SIR = ∞;

VI CONCLUSIONS

Quasi-coherent signal combining yields a major improvement of the BER performance compared to the one antenna case. The obtained BER results also show that the influence of co-channel interference on the BER is considerable. It should be noted however, that the results in this report are obtained for the worst case situation since the data sequences of the desired and the interfering signal are assumed synchronous.

BER performance in case of co-channel interference is significantly better for the method using remodulation than for the simple method. This is cause by the fact that remodulation selects only the signal which correlates with the estimated data for phase estimation. Especially for a small SIR the difference is remarkable.

For $SIR = \infty$, the results obtained for the remodulation method are only slightly better than those obtained for the simple method. So, in absence of co-channel interference, the simple method is an appropriate method for phase estimation in QC-combining.

REFERENCES

[1] J.H. Winters, "Optimum Combining in Digital Mobile Radio with Cochannel Interference", *IEEE JSAC*, Vol. SAC-2, No. 4, pp. 528 - 539, July 1984.

[2] M.V. Clark, L.J. Greenstein, W.K. Kennedy, M. Shafi, "MMSE diversity combining for wideband digital cellular radio", *IEEE Trans. on Comm.*, Vol. 40, No. 6, pp. 1128 - 1135, June 1992.

[3] G.J.M. Janssen, W.R. Bredero, and R. Prasad, "Performance Enhancement for Microwave and Millimetre Wave Indoor Communication Systems by Quasi-Coherent Combining of Multiple Antenna Signals", *Proc. IEEE PIMRC'94*, The Hague, Sept. 1994, pp. 730-735.

[4] G.J.M. Janssen, L. Maranatha, and R. Prasad, "Evaluation of BER Performance for Measured Indoor Microwave Channels Using Quasi-Coherent Combining of Multiple Antenna Signals", *COST, ?*

[5] A.A.M. Saleh, R.A. Valenzuela, "A statistical model for indoor multipath propagation", *IEEE JSAC*, Vol. SAC-5, No. 2, pp. 128-137, Feb. 1987.

[6] H. Hashemi, "Impulse response modelling of indoor radio propagation channels", *IEEE JSAC*, Vol. SAC-11, pp. 967-978, Sept. 1993.

[7] G.J.M. Janssen, B.C. v. Lieshout, R. Prasad, BER Performance of Millimetre Wave Indoor Communication Sytems Using Multiple Antenna Signals", IEEE Proc. Communication Theory Miniconference, pp. 105 - 109, Globecom'94, San Francisco, December 1994.

[8] G.J.M. Janssen, "Dual-signal Receiver Structures for Simultaneous Reception of Two BPSK Modulated Co-channel Signals Using Signal Cancellation", *Wireless Personal Communications*, Vol. 1, No. 1, pp. 43 - 59, Kluwer Academic Publishers, December 1994.